Des relevés spectroscopiques de galaxies à raies d'émission aux contraintes cosmologiques : d'eBOSS à DESI

Arnaud De Mattia

université PARIS-SACLAY

From emission line galaxy spectroscopic surveys to cosmological constraints: from eBOSS to DESI

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 576, Particules, Hadrons, Électrons, Noyaux, Ingénierie, Imagerie, Cosmologie et Spatial (Pheniics)
Spécialité de doctorat : Astroparticules & Cosmologie
Unité de recherche : Université Paris-Saclay, CEA, Département de Physique des Particules, 91191, Gif-sur-Yvette, France
Référent : Faculté des sciences d'Orsay

Thèse présentée et soutenue à Gif-sur-Yvette, le 21 septembre 2020, par

Arnaud de Mattia

Composition du jury:

Réza Ansari Professeur, Université Paris-Saclay, IJCLab	Président
Christophe Balland Professeur, Sorbonne Université, LPNHE	Rapporteur & examinateur
Julien Guy Directeur de recherche, Lawrence Berkeley National Laboratory, USA	Rapporteur & examinateur
Filippo Vernizzi Directeur de recherche, Université Paris-Saclay, CEA IPhT	Rapporteur & examinateur
Francis Bernardeau Directeur de recherche, CNRS & Sorbonne Université, IAP	Examinateur
Sylvain de la Torre Astronome-adjoint, Université Aix-Marseille, LAM	Examinateur
Vanina Ruhlmann-Kleider Directrice de recherche, Université Paris-Saclay, CEA DPhP	Directrice de thèse

Thèse de doctorat

NNT: 2020UPASP010

Contents

Remerciements

Ce travail de thèse a été effectué au sein du groupe de cosmologie du Département de Physique des Particules (DPhP) de l'Institut de Recherche sur les lois Fondamentales de l'Univers du CEA Saclay. Ma bourse de thèse a été alimentée conjointement par le labex P2IO (ANR-10-LABX-0038) et par le DPhP. La participation du groupe de cosmologie à eBOSS a été financée par l'Agence nationale de la recherche (ANR-16-CE31-0021).

Je remercie le président et les examinateurs de mon jury d'avoir accepté d'évaluer mon travail de thèse et d'avoir fait le déplacement à Saclay pour ma soutenance malgré les conditions sanitaires difficiles. Merci pour leur bienveillance et leurs questions fort pertinentes. Je remercie mes rapporteurs d'avoir lu ce manuscrit en profondeur et de ne pas m'avoir tenu rigueur des nombreuses fautes de frappe qui ponctuaient la première version. Leurs suggestions et corrections ont été extrêmement bienvenues, et j'espère ne pas en avoir oublié. Malgré les multiples relectures de ce manuscrit, il est probable que des erreurs s'y trouvent encore ; je présente par avance mes excuses à mes futurs lecteurs et lectrices qui sauront, je n'en doute pas, les déceler et me les signaler.

Je remercie ma directrice de thèse, Vanina Ruhlmann-Kleider, pour son encadrement pendant mon stage de fin d'études et la thèse qui s'en est suivie. Avec force patience, diplomatie et détermination, elle m'a notamment appris à présenter mon travail à un public non-initié. Je lui suis infiniment reconnaissant pour ses précieux conseils lors de ma candidature au poste de chercheur au DPhP. Bien sûr, beaucoup m'ont aidé dans la préparation de l'après-thèse ; ils et elles se reconnaîtront.

Je remercie le personnel du département pour son accueil, et particulièrement Pascal Debu et Bertrand Vallage de m'avoir conforté dans mon goût pour la recherche. Ce fut un plaisir de travailler dans le groupe de cosmologie du DPhP. Doctorants, post-doctorants, chercheuses et chercheurs permanents sont considérés sur un pied d'égalité. Bien que nos sujets de recherche soient variés, les repas du midi et les réunions du vendredi sont garants d'une cohésion que beaucoup pourraient nous envier. Merci donc à tous les membres du groupe que j'ai côtoyés, qui sont partis, partent, restent quelques années ou pour plus longtemps, que je suis heureux

d'avoir pour collègues.

J'ai eu la chance de travailler dans la collaboration internationale eBOSS. Je remercie tous ses membres pour leur bienveillance, leur soutien et leur intégrité scientifique. J'ai beaucoup appris à leurs côtés. Je suis extrêmement reconnaissant envers ceux qui ont soutenu ma candidature à des contrats post-doctoraux et permanents, et suis très heureux à l'idée de travailler avec eux dans l'expérience DESI.

Je remercie mes proches pour leur soutien indéfectible. Je suis en particulier infiniment reconnaissant envers mes parents de m'avoir transmis ce goût pour les humanités et les sciences.

Pour finir, je souhaite plein de courage à mes collègues doctorants et post-doctorants dans la suite de leur carrière.

Introduction

Astrophysicists reveal the biggest 3D map of the Universe. On this day of July 20, 2020, numerous newspapers have devoted an article to the press release of the eBOSS collaboration, accompanying the submission of its last papers. This marks the end of 20 years of spectroscopic observations by the Sloan Digital Sky Survey (SDSS). This is also a turning point in the size of spectroscopic galaxy surveys: in a few months, the Dark Energy Spectroscopic Instrument will start regular operations, gathering ten times more galaxy redshifts than ever assembled.

Building such a gigantic map of the Universe requires a tremendous instrumental and observational effort. First, the sky is pictured, and target galaxies are identified in the resulting images with colour selections. Then, optical fibres are assigned to each of these target to guide the light from the focal plane of the telescope down to the spectrographs, which decompose galaxy light into spectra. From these spectra, redshifts are measured, which, in combination with galaxy angular positions, are turned in 3-dimensional coordinates.

However inspiring this 3D map is, putting the fragility of our blue marble in perspective to the immensity of space, it also contains a wealth of cosmological information. Indeed, studying the statistical properties of the distribution of galaxies provides insights into one of the greatest mystery of the Universe: the acceleration of its expansion, which was evidenced by the distance measurement of a few dozens of distant supernovae in 1998. The so-called clustering analyses of galaxy redshift surveys rely on two physical effects: baryon acoustic oscillations and redshift-space distortions.

As a result of the conjoint action of gravity and radiation pressure in the primordial Universe, acoustic waves propagate into the plasma of baryons and photons. With the Universe cooling down, photons are released, weaving their path trough the structure of the Universe before reaching our instruments. They form a picture of the Universe, called the cosmic microwave background, where the pattern of these acoustic waves can be clearly seen. Freezing when photons decouple, these so-called baryon acoustic oscillations are imprinted in the large scale galaxy distribution. This standard ruler, measured at different redshifts by the SDSS, allows constraints to be put on the expansion rate of the Universe, and hence on the dark energy

equation of state.

In addition, under the action of gravity, galaxies tend to fall in overdensities, such that their motion is coherent on large scales — and appear incoherent on small scales, where galaxies live in gravitationally bound systems. As galaxy velocities change their measured redshift, they impact the observed spatial distribution of galaxies. These so-called redshift space distortions grant access to the growth rate of structure — namely how fast structures grow in the Universe. Such a measurement provides constraints on the dark energy equation of state or helps test general relativity on large scales.

In particular, in this manuscript, we perform the clustering analysis of the eBOSS sample of emission line galaxies (ELGs), located at a redshift between 0.6 and 1.1, when the Universe was half younger than today. These ELGs are dense at these redshifts and exhibit strong emission features in their spectra, making them ideal for clustering studies. This new sample in SDSS benefited from various updates to best pave the way for DESI. The cosmological measurement was hindered by significant observational systematics, which we aimed at correcting for. We also dedicated significant work to the development and improvement of the analysis pipeline, modelling analysis systematic effects which were unaccounted for so far.

This manuscript is divided into six chapters, as follows. In Chapter 1, we expose the concordance model of cosmology, which we assume throughout the rest of the manuscript, with an emphasis on baryon acoustic oscillations. Structure formation is described in Chapter 2, where we specify the theoretical model to be confronted to observations. We then take a short break to describe the construction of a galaxy spectroscopic survey in Chapter 3, which eventually results in a catalogue of galaxy angular position and redshifts. In Chapter 4 we explain how to compress and exploit this amount of data in a summary statistics, called the galaxy power spectrum, and how the survey geometry impacts its measurement. Also discussed is the impact of inferring this survey geometry from the observed data itself. Given these propaedeutic studies, we perform the clustering analysis of the eBOSS ELG sample in Chapter 5, and review cosmological implications of the SDSS galaxy surveys. Chapter 6 exposes two side-projects complementing the ELG analysis, and briefly discusses DESI prospects.

The concordance model of cosmology

A study of the large scale distribution of galaxies — the purpose of this manuscript — surely requires a propaedeutic on the structure of the Universe. We will first describe how to measure cosmological distances and how there are related to the energy content of the Universe. Next, we will explain how tiny, quantum fluctuations grew and propagated in the early Universe to imprint on the large scale matter distribution.

We do not aim at giving a complete description of modern cosmology — many dedicated textbooks are available — but to include the minimal notions to make this manuscript somewhat self-consistent.

1.1 The homogeneous Universe

A cornerstone of the concordance model of cosmology ΛCDM is general relativity, which we will very briefly present hereafter. The large scale, global geometry of the Universe will be determined by invoking the cosmological principle. This will allow us to define cosmological distances and explain how they are related to the energy content of the Universe, which we will also review. This section is mainly based on Dodelson (2003) and Peter and Uzan (2012) textbooks and Bernardeau (1998) course.

1.1.1 General relativity

Any metric theory of gravitation relies on Einstein's equivalence principle, which consists in:

1. the weak equivalence principle: the trajectory of any test particle is independent of its internal structure or composition;
2. local position invariance: the result of any non-gravitational experiment is independent of its space-time position;
3. local Lorentz invariance: the result of any non-gravitational experiment is independent of the laboratory's motion if in free fall.

Under these assumptions, gravitation stems from a curved space-time, whose geometry is described by a metric $\mathbf{g}$. The length element between two space-time events separated by dx^μ is:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \tag{1.1}$$

where we sum over repeated indices, and Greek indices run from 0 to 3. On the contrary, Latin indices will run from 1 to 3. $g_{\mu\nu}$ is a symmetric tensor, of inverse $g^{\mu\nu}$ and determinant g. Particles only submitted to gravitation follow the geodesics of this 4-dimensional manifold, i.e. their trajectory $x^\mu(\tau)$ (with τ a scalar parameter) maximises the action:

$$S = \int g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} d\tau. \tag{1.2}$$

Maximising S, using the Euler-Lagrange equation, gives the geodesic equation:

$$\frac{d^2 x^\alpha}{d\tau^2} + \Gamma^\alpha_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0, \tag{1.3}$$

with the Christoffel symbols:

$$\Gamma^\alpha_{\mu\nu} = \frac{1}{2} g^{\lambda\alpha} \left(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu} \right) \tag{1.4}$$

where we noted the partial derivative $\partial_\mu = \partial/\partial x^\mu$. The geodesic equation can be recast in the form:

$$u^\mu u^\alpha_{;\mu} \qquad \text{with} \qquad u^\mu = \frac{dx^\mu}{d\tau}. \tag{1.5}$$

Here $;\nu$ is the covariant derivative, which takes into account the variation in the coordinate system. The covariant derivative is defined for a vector T^α as:

$$T^\alpha_{;\mu} = \partial_\mu T^\alpha + \Gamma^\alpha_{\mu\nu} T^\nu, \tag{1.6}$$

and is extended to any tensor $T^{\mu_1\cdots\mu_p}_{\nu_1\cdots\nu_q}$ following:

$$\begin{aligned} T^{\mu_1\cdots\mu_p}_{\nu_1\cdots\nu_q;\alpha} = \partial_\alpha T^{\mu_1\cdots\mu_p}_{\nu_1\cdots\nu_q} + \Gamma^{\mu_1}_{\alpha\lambda_1} T^{\lambda_1\cdots\mu_p}_{\nu_1\cdots\nu_q} + \cdots + \Gamma^{\mu_p}_{\alpha\lambda_p} T^{\mu_1\cdots\mu_{p-1}\lambda_p}_{\nu_1\cdots\nu_q} \\ - \Gamma^{\lambda_1}_{\alpha\nu_1} T^{\mu_1\cdots\mu_p}_{\lambda_1\cdots\nu_q} - \cdots - \Gamma^{\lambda_q}_{\alpha\nu_q} T^{\mu_1\cdots\mu_p}_{\nu_1\cdots\nu_{q-1}\lambda_q}. \end{aligned} \tag{1.7}$$

An observer four-velocity is defined as $u^\mu = dx^\mu/d\tau$, with the proper time $d\tau = \sqrt{-g_{\mu\nu}dx^\mu dx^\nu}$. Note that we set $c = \hbar = 1$ in all this manuscript. Therefore, $u^\mu u_\mu = -1$. Photons follow light geodesics, $ds^2 = -d\tau^2 = 0$. The Riemann tensor measures the non-commutativity of the covariant derivative due to space-time curvature:

$$T^\mu_{;\alpha;\beta} - T^\mu_{;\beta;\alpha} = R^\mu{}_{\nu\alpha\beta} T^\nu. \tag{1.8}$$

In terms of Christoffel symbols:

$$R^\mu{}_{\nu\alpha\beta} = \partial_\alpha \Gamma^\mu_{\nu\beta} - \partial_\beta \Gamma^\mu_{\nu\alpha} + \Gamma^\mu_{\sigma\alpha}\Gamma^\sigma_{\nu\beta} - \Gamma^\mu_{\sigma\beta}\Gamma^\sigma_{\nu\alpha}. \tag{1.9}$$

The only contraction leading to a tensor is the Ricci tensor:

$$R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu} \tag{1.10}$$

whose trace $R = g^{\mu\nu}R_{\mu\nu}$ is dubbed curvature or Ricci scalar.

The equations governing the evolution of $g_{\mu\nu}$ can be derived following the principle of least action. Einstein and Hilbert proposed the scalar curvature R as Lagrangian:

$$\mathcal{S}_{\rm EH} = \frac{c^4}{16\pi G_{\rm N}} \int d^4x \sqrt{-g} R \tag{1.11}$$

but there are many other possibilities — leading to modified gravity theories. We further introduce the cosmological constant Λ and add the Lagrangian of matter $\mathcal{L}_m$, such that the total action reads:

$$\boxed{\mathcal{S} = \frac{c^4}{16\pi G_{\rm N}} \int d^4x \sqrt{-g}\,(R - 2\Lambda) + \int d^4x \sqrt{-g}\mathcal{L}_m} \tag{1.12}$$

where $G_{\rm N}$ is the Newton constant. Varying the action ($\delta\mathcal{S} = 0$), one obtains the Einstein field equation:

$$\boxed{G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G_{\rm N}}{c^4} T_{\mu\nu}} \tag{1.13}$$

with:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R, \tag{1.14}$$

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta\sqrt{-g}\mathcal{L}_m}{\delta g^{\mu\nu}} = -2\frac{\delta\mathcal{L}_m}{\delta g^{\mu\nu}} + g_{\mu\nu}\mathcal{L}_m. \tag{1.15}$$

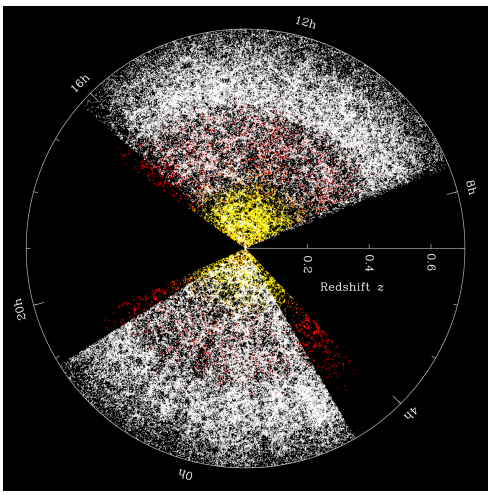

Figure 1.1 – The distribution of galaxies probed by the SDSS/BOSS survey. The SDSS Main Galaxy Sample is shown in yellow, the BOSS LOWZ and CMASS samples are displayed in red and white, respectively. Credits: SDSS.

are the Einstein tensor and momentum-energy tensor, respectively. In addition, since $g_{\mu\nu;\alpha} = 0$ and $G^{\mu\nu}_{\ ;\nu} = 0$, Eq. (1.13) ensures the conservation of the total energy-momentum tensor:

$$\boxed{T^{\mu\nu}_{\ ;\nu} = 0} \tag{1.16}$$

In ΛCDM general relativity is applied to cosmological scales, which requires some assumptions on the large scale geometry of the Universe.

1.1.2 Cosmological principle

Cosmology consists in the study of the past and the future of the Universe. Though we can only observe an infinitely small portion (our past light-cone, defined as $ds^2 = 0$) of one single universe, we claim to have a theory that describes its evolution. This cannot go without some assumptions on its nature.

On scales below 100 Mpc, the Universe is anisotropic: we observe large structures probed by large galaxy surveys, such as the SDSS (see Figure 1.1). However, on scales larger than 200 Mpc, the Universe appears relatively isotropic. One also notes the remarkable isotropy of the temperature of the cosmic microwave background (CMB), a picture of the sky at redshift $z_\star \simeq 1090$: subtracting the dipole due to the proper motion of our galaxy (inducing relative temperature fluctuations of 10^{-3}), remaining relative temperature fluctuations are only of the order of 10^{-5}. Hence, we observe the Universe to be isotropic on large scales.

In addition, the Copernican principle states that we do not occupy a special place in the Universe. Further assuming that the Universe is isotropic (even beyond the scales probed by observations) implies the cosmological principle: the Universe is spatially isotropic and homogeneous.

As such, and within general relativity, the cosmological principle enables us to derive equations governing the dynamics of the smooth, *background* Universe. Observed inhomogeneities are described as perturbations over this background.

1.1.3 Cosmological distances

The Friedmann-Lemaître-Robertson-Walker (FLRW) metric is the generic metric that satisfies the cosmological principle:

$$\boxed{ds^2 = -dt^2 + a^2(t)\left(\frac{dx^2}{1-Kx^2} + x^2 d\Omega^2\right)} \tag{1.17}$$

where t is the cosmic time, $a(t)$ is the scale factor, x is a radial coordinate and $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ is the infinitesimal solid angle ($\phi \in [0, 2\pi]$ and $\theta \in [-\pi, \pi]$). K is the curvature term. If $K = 0$, the spatial part of the metric is flat; if $K > 0$, it is said to be closed, and if $K < 0$, to be open.

In the limit of small separations, the physical distance between two objects is given by $\mathbf{r} = a\mathbf{x}$. Therefore, its derivative gives the relative velocity of the two bodies:

$$\mathbf{v} = \dot{a}\mathbf{x} + a\dot{\mathbf{x}} = H\mathbf{r} + a\dot{\mathbf{x}}, \tag{1.18}$$

where we used the Hubble parameter $H = \dot{a}/a$. $\dot{}$ denotes a derivative with respect to the cosmic time t. $a\dot{\mathbf{x}}$ is due to the galaxy proper velocities; it is zero in average. Thus: $\langle\mathbf{v}\rangle = H\mathbf{r}$. Actually, the simple observation that $\langle\nabla\cdot\mathbf{v}\rangle$ is a constant in space (following from the cosmological principle) leads to the same conclusion. This demonstrates Hubble-Lemaître law (Hubble, 1929): the velocity between two objects (far enough that space can be considered homogeneous, but close enough that H can be considered constant) is proportional to their distance.

In practice, we do not measure velocities, but redshifts z. Indeed, the absorption or emission lines of astrophysical objects are shifted to larger — redder — wavelength, as the Universe expands. Let us consider a galaxy, located at x_e, emitting a photon at t_e that the observer receives at t_0 (in $x_0 = 0$). The photon follows the geodesic $ds^2 = 0$; therefore:

$$\int_{t_e}^{t_0} \frac{dt}{a(t)} = \int_{x_0}^{x_e} \frac{dx}{\sqrt{1-Kx^2}}. \tag{1.19}$$

If a second photon is emitted at $t_e + \delta t_e$, with $\delta t_e \ll 1/H(t_e)$ it will be received at $t_0 + \delta t_0$ such that:

$$\int_{t_e}^{t_0} \frac{dt}{a(t)} = \int_{t_e+\delta t_e}^{t_0+\delta t_0} \frac{dt}{a(t)}. \tag{1.20}$$

Then:

$$\frac{\delta t_e}{a(t_e)} = \frac{\delta t_0}{a(t_0)}. \tag{1.21}$$

Therefore, the wavelength $\lambda_e \propto \delta t_e$ of a photon emitted at t_e will be stretched by the factor:

$$\boxed{1 + z(t_e) = \frac{\lambda_0}{\lambda_e} = \frac{a(t_0)}{a(t_e)}} \tag{1.22}$$

thus defining the redshift $z(t_e)$ of the galaxy at t_e. In the following, we naturally set $a(t_0) = 1$.

Let us now define the different cosmological distances which will be used to locate objects in space-time. First, let us recast the metric (1.17) into the more convenient form:

$$ds^2 = -dt^2 + a^2(t)\gamma_{ij}dx^i dx^j = a^2(\eta)\left(-d\eta^2 + \gamma_{ij}dx^i dx^j\right) \tag{1.23}$$

where η ($d\eta = dt/a$) is called the conformal time and:

$$\gamma_{ij} dx^i dx^j = d\chi^2 + f_K^2(\chi) d\Omega^2 \tag{1.24}$$

with:

$$f_K(\chi) = \begin{cases} K^{-1/2} \sin\left(\sqrt{K}\chi\right) & K > 0 \\ \chi & K = 0 \\ (-K)^{1/2} \sinh\left(\sqrt{-K}\chi\right) & K < 0. \end{cases} \tag{1.25}$$

The *radial comoving distance* χ of an object emitting at time t_e and observed at t_0 in $\chi(t_0) = 0$ is obtained by integration along a radial light geodesic:

$$\chi(t_e) = \int_{t_e}^{t_0} \frac{dt}{a(t)}. \tag{1.26}$$

Noting that $dt = da/(aH) = -dz/\left((1+z)H(z)\right)$, the radial comoving distance can be recast as:

$$\boxed{\chi(z_e) = \int_0^{z_e} \frac{dz}{H(z)}} \tag{1.27}$$

The *comoving angular distance* D_M relates the transverse comoving size of an object (comoving surface area dS_e) to the corresponding solid angle $d\Omega_0$ seen by an observer, namely:

$$dS_e = D_\mathrm{M}^2 d\Omega_0^2. \tag{1.28}$$

The solution is read directly on Eq. (1.24):

$$\boxed{D_\mathrm{M}(z) = f_K(\chi(z))} \tag{1.29}$$

The *angular diameter distance* D_A relates the transverse physical of size of an object to the corresponding solid angle seen by an observer. Since the physical surface area is $dS_p = a(t_e)^2 dS_e$, we have:

$$\boxed{D_\mathrm{A}(z) = \frac{D_\mathrm{M}(z)}{1+z}} \tag{1.30}$$

The *luminosity distance* D_L relates the luminosity L_e of a source located at the radial comoving distance χ to the flux ϕ_0 measured by the observer:

$$\phi_0 = \frac{L_e}{4\pi D_\mathrm{L}^2}. \tag{1.31}$$

The luminosity ($\propto dE/dt$) decreases as a^2, as the time interval dt ($\propto a d\eta$) scales as a and the photon energy dE ($\propto \lambda_e^{-1}$) as a^{-1}. Therefore:

$$\phi_0 = \frac{L_e\,(1+z)^{-2}}{4\pi f_K^2(\chi)} \tag{1.32}$$

and:

$$\boxed{D_\mathrm{L}(z) = (1+z)\, f_K(\chi(z))} \tag{1.33}$$

So far, all distances are expressed as a function of H and K — which will be related to the energy content of the Universe in the following. However, we will first discuss the measurement of the Hubble constant $H_0 = H(z=0)$, the value of H at the present epoch, as it does not require any further assumption.

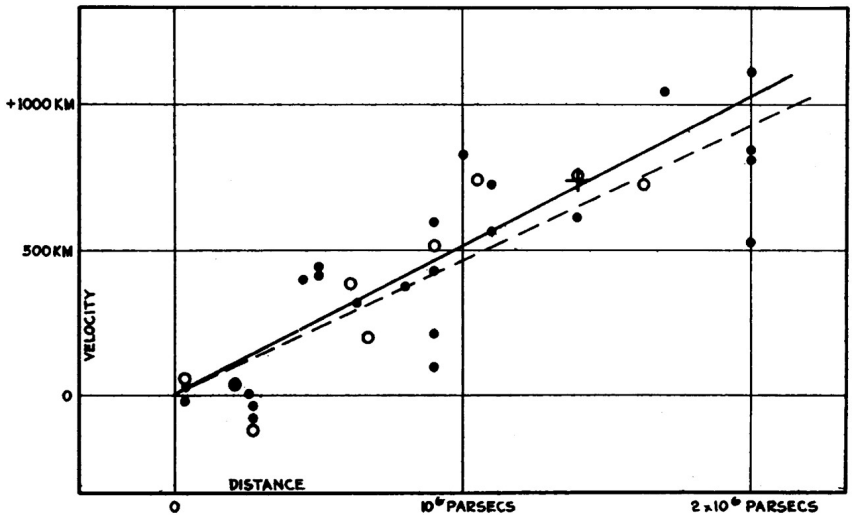

Figure 1.2 – The distance - redshift relation of nearby galaxies measured by Edwin Hubble in 1929. The continuous line represents the fit from galaxies taken separately (black disks) and the dashed lines from galaxies combined into groups (white circles). Taken from Hubble (1929).

1.1.4 The Universe expansion

At first order in $z \ll 1$, the luminosity distance (1.33) is:

$$D_{\rm L} = \frac{cz}{H_0} = \frac{v}{H_0}. \tag{1.34}$$

Therefore, H_0 can be determined by measuring redshifts and luminosity distances independently. This method — the so-called Hubble diagram — is hence the most direct evidence for the expansion of the Universe. In his seminal paper of 1929, Hubble measured $H_0 = 500\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, 7 times higher than the most recent measurements (see Figure 1.2). One should however target galaxies far enough that typical galaxy velocities $\sim 300\,\mathrm{km\,s^{-1}}$ are negligible compared to the Hubble expansion. A fractional contribution of galaxy proper velocities of 3% to the observed redshifts requires a Hubble flow of $v \sim 1 \times 10^4\,\mathrm{km\,s^{-1}}$, thus one has to observe galaxies at $z \sim 0.03$.

The most direct measurement relies on parallaxes (e.g. Luri et al., 2018), but can only be used for stars close to the Milky Way (up to a few kpc). Otherwise, one can focus on different classes of astrophysical objects of known or standardisable luminosity. These are called standard candles:

1. Cepheids: variable stars with a period of 2 to 100 days. This period has been found to correlate with luminosity with a scatter of 20% in luminosity. This relation can be calibrated with parallax measurements (e.g. Riess et al., 2018). Given their luminosity, Cepheids can only be targeted up to 20 Mpc. However, these Cepheids can be used to calibrate the magnitude of other astrophysical objects found in the same galaxy.
2. supernovae (SN) Ia: they consist in the stellar explosion of a white dwarf in a binary system. Their maximal intrinsic luminosity, comparable to that of an entire galaxy, is correlated with the characteristic time of their light-curve, with a scatter of 12%. Their absolute magnitude can be calibrated with Cepheids (e.g. Dhawan et al., 2018).
3. tip of the red giant branch (TGRB): the luminosity of brightest stars in the red giant branch has been shown to be somewhat insensitive to their composition or their mass (e.g. Freedman et al., 2019).

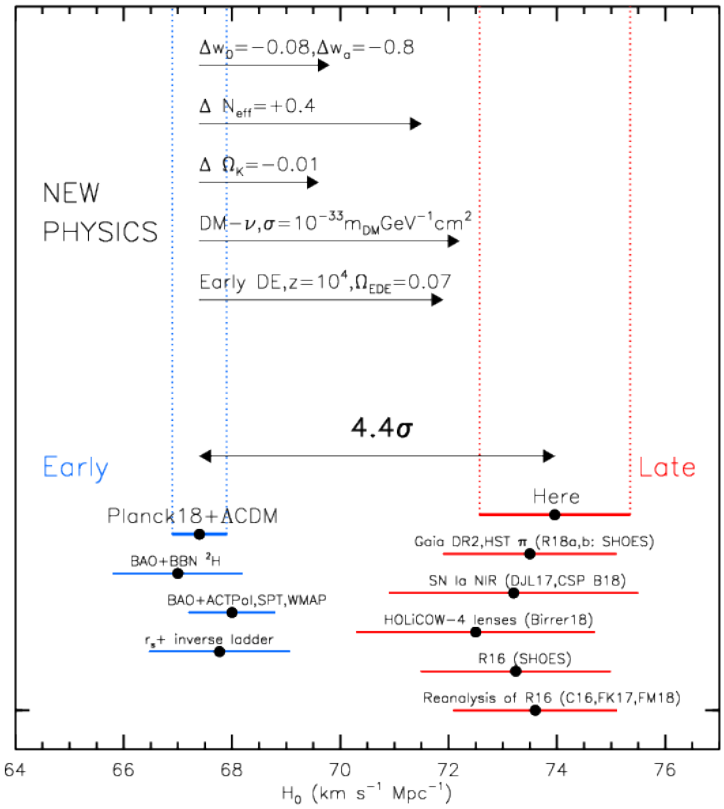

Figure 1.3 – Comparison of the early and late type measurements of H_0. Taken from Riess et al. (2019).

4. galaxies: their luminosity is correlated to their rotation velocities (Tully-Fisher relation), their velocity dispersion, or the surface brightness fluctuations (e.g. Ajhar et al., 2001).

Other methods, which do not rely on distance ladder, include:

1. SN II photosphere: after their explosion, the photosphere of SN II propagates at $v/c \sim 0.01$. One can obtain their angular diameter distance by measuring the angular diameter θ of this photoshere at a given time after the explosion (e.g. Schmidt, 1993).
2. time delays due to gravitational lensing: H_0 can be measured by comparing the time difference between two images of a quasar lensed by a foreground galaxy, if the mass distribution of that galaxy is known (e.g. Birrer et al., 2019).
3. Sunyaev-Zel'dovich effect: the CMB black body spectrum is distorted by the inverse Compton diffusion by the ionised gas along the line of sight. Measuring this distortion, and the temperature and density map of the X emission of the gas, one can infer its distance (e.g. Bonamente et al., 2006).

Today, as can be seen on Figure 1.3, there is a strong tension between the local (late) measurements of H_0 and those relying on the physics in the early Universe — CMB or Big Bang nucleosynthesis (BBN) with baryon acoustic oscillations (BAO). The latest results lead to a 5.3σ discrepancy (Wong et al., 2019).

1.1.5 Friedmann equations

The time dependence of the scale factor a is related to the energy-momentum tensor $T_{\mu\nu}$ by the Einstein equation (1.13). The most general form of $T_{\mu\nu}$ in a homogeneous and isotropic

Universe is:

$$T_{\mu\nu} = \rho u_\mu u_\nu + P\hat{\gamma}_{\mu\nu}, \tag{1.35}$$

with the four-velocity $u_\mu = \delta^\mu_0$. $\hat{\gamma}_{\mu\nu} = g_{\mu\nu} + u_\mu u_\nu$ is the projector onto the spatial part of the metric. For a comoving observer of four-velocity u_μ, $\rho = T_{\mu\nu}u^\mu u^\nu$ is the fluid energy density and $P = T_{\mu\nu}\hat{\gamma}^{\mu\nu}/3$ its pressure, which only depend on time. This corresponds to a perfect fluid (no shear).

Then, the Einstein equation (1.13) reduces to the Friedmann equations:

$$\boxed{H^2 = \frac{8\pi G_\mathrm{N}}{3}\rho - \frac{K}{a^2} + \frac{\Lambda}{3}} \tag{1.36}$$

$$\boxed{\frac{\ddot{a}}{a} = -\frac{8\pi G_\mathrm{N}}{6}(\rho + 3P) + \frac{\Lambda}{3}} \tag{1.37}$$

where we recall $\dot{}$ denotes a derivative with respect to the cosmic time t. Combining the two equations above (deriving Eq. (1.36), noting that $\dot{H} = \ddot{a}/a - H^2$, to be replaced by Eq. (1.36) and Eq. (1.37)), or, equivalently, using the conservation equation (1.16) gives:

$$\dot{\rho} + 3H\left(\rho + P\right) = 0. \tag{1.38}$$

We obtain two independent equations and three unknowns: a, ρ and P. To close the system, another equation is required, namely the fluid equation of state, which takes the form:

$$\boxed{P = w\rho} \tag{1.39}$$

Injecting this formula into Eq. (1.38) implies that, for w constant (which we will use throughout this manuscript):

$$\rho \propto a^{-3(1+w)}. \tag{1.40}$$

Already, one can note from Eq. (1.36) that the cosmological constant is equivalent to a fluid of equation of state $w = -1$ and the curvature term K to $w = -1/3$. For pressureless matter, $w = 0$ and for relativistic species, $w = 1/3$. Injecting Eq. (1.40) into Eq. (1.36), we find:

type	equation of state w	energy density ρ	scale factor a	conformal time η
cosmological constant	-1	$\propto a^0$	$\propto e^{Ht}$	$\eta \propto e^{-Ht}$
curvature	$-1/3$	$\propto a^{-2}$	$\propto t$	$\propto \ln t$
pressureless matter	0	$\propto a^{-3}$	$\propto t^{2/3}$	$\propto t^{1/3}$
relativistic species	$1/3$	$\propto a^{-4}$	$\propto t^{1/2}$	$\propto t^{1/2}$

Table 1.1 – Dynamics of the Universe as a function of the energy content.

We usually define the critical density as:

$$\rho_c = \frac{3H^2}{8\pi G_\mathrm{N}}. \tag{1.41}$$

Then, we define the density parameter of fluid X with energy density ρ_X by $\Omega_X = \rho_X/\rho_c$. In particular:

$$\rho_\Lambda = \frac{\Lambda}{8\pi G_\mathrm{N}} \qquad \rho_K = -\frac{3K}{8\pi G_\mathrm{N} a^2}. \tag{1.42}$$

The first Friedmann equation can be rewritten:

$$\sum_X \Omega_X + \Omega_\Lambda + \Omega_K = 1. \tag{1.43}$$

Let us note $\Omega_{X,0} = \Omega_X(z=0)$, and write the equation of state of fluid fluid X w_X. Assuming these fluids to couple only through gravitation (such that Eq. (1.40) is verified independently for each of them):

$$\Omega_X = \Omega_{X,0}(1+z)^{3(1+w_X)} \left[\frac{H_0}{H(z)}\right]^2, \tag{1.44}$$

such that:

$$\boxed{H^2(z) = H_0^2 \left[\sum_X \Omega_{X,0}(1+z)^{3(1+w_X)} + \Omega_{K,0}(1+z)^2 + \Omega_{\Lambda,0}\right]} \tag{1.45}$$

We can therefore compute distances (1.27), (1.29) if we specify $\Omega_{X,0}$, $\Omega_{K,0}$, $\Omega_{\Lambda,0}$ and H_0. We adopt the usual convention $H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, and will often use the reduced (or *physical*) density parameters $\omega_{X,0} = \Omega_{X,0}h^2$, which are proportional to the physical energy density today (since $\rho_X = \Omega_X \rho_c$ with ρ_c given by Eq. (1.41)).

1.2 Energy content

In the ΛCDM model, the following energy content of the Universe is assumed: dark energy (through the cosmological constant), radiation (photons and neutrinos), baryons and cold dark matter. In general, best constraints on energy densities come from the CMB, whose phenomenology will be presented in Section 1.3. We thus provide parameter densities today in Table 1.2, as measured by Planck Collaboration et al. (2018), combining CMB (TT, TE, EE, lowE, lensing) and BAO measurements and assuming a flat ΛCDM model with two massless and one neutrino of mass $m_\nu = 0.06\,\mathrm{eV}$. Table 1.2 also contains other parameters discussed in this chapter. In the following, we review the different components mentioned hereabove, emphasising some alternate methods to measure the corresponding density parameters.

Letting curvature free, Planck Collaboration et al. (2018) measured its density parameter to be $\Omega_{K,0} = 0.0007 \pm 0.0019$, combining CMB (TT, TE, EE, lowE, lensing) and BAO measurements. Such a low (absolute) value of curvature density will be explained in the last part of this section, while we will assume a flat ΛCDM model elsewhere. Within this model, the evolution of density parameters is shown as a function of redshift in Figure 1.4.

1.2.1 Dark energy

The scale factor a can be Taylor expanded following:

$$a(t) = 1 + H_0\,(t-t_0) - \frac{1}{2} q_0 H_0^2\,(t-t_0)^2 + \dots \tag{1.46}$$

with:

$$q_0 = -\frac{\ddot{a}(t_0)}{H_0^2}. \tag{1.47}$$

reduced Hubble constant	h	0.6766 ± 0.0042
baryon density	$\omega_{b,0}$	$0.022\,42 \pm 0.000\,14$
cold dark matter density	$\omega_{c,0}$	$0.119\,33 \pm 0.000\,91$
photon density	$\omega_{\gamma,0}$	$(2.4729 \pm 0.0022) \times 10^{-5}$
dark energy density	$\Omega_{\Lambda,0}$	0.6889 ± 0.0056
optical depth at reionisation	$\tau_{\rm rei}$	0.0561 ± 0.0071
redshift of reionisation	$z_{\rm rei}$	7.82 ± 0.71
index of the primordial power spectrum	n_s	0.9665 ± 0.0038
amplitude of the primordial power spectrum	$10^9 A_s$	2.105 ± 0.030
normalisation of the matter power spectrum	$\sigma_{8,0}$	0.8102 ± 0.0060
redshift of matter-radiation equality	$z_{\rm eq}$	3387 ± 21
last scattering redshift	$z_\star$	1089.80 ± 0.21
drag redshift	$z_{\rm drag}$	1060.01 ± 0.29
sound horizon at the drag epoch	$r_{\rm drag}$	$147.21 \pm 0.23\,{\rm Mpc}$

Table 1.2 – Cosmological parameters as measured by Planck Collaboration et al. (2018) (TT, TE, EE, lowE, lensing, BAO) within the flat ΛCDM 6 parameter model, i.e. varying $\omega_{b,0}$, $\omega_{c,0}$, $\theta_{\rm MC}$ (equivalent of h), τ, $\ln\left(10^{10} A_s\right)$ and n_s. A_s is provided at the pivot scale $k_0 = 0.05\,{\rm Mpc}^{-1}$. $\Omega_{\Lambda,0}$ is imposed by Eq. (1.43) assuming flatness, but we quote it for convenience. The photon density is measured by COBE/FIRAS (Fixsen, 2009). The sum of neutrino masses is fixed to 0.06 eV, yielding $\omega_{\nu,0} = 6.4 \times 10^{-4}$.

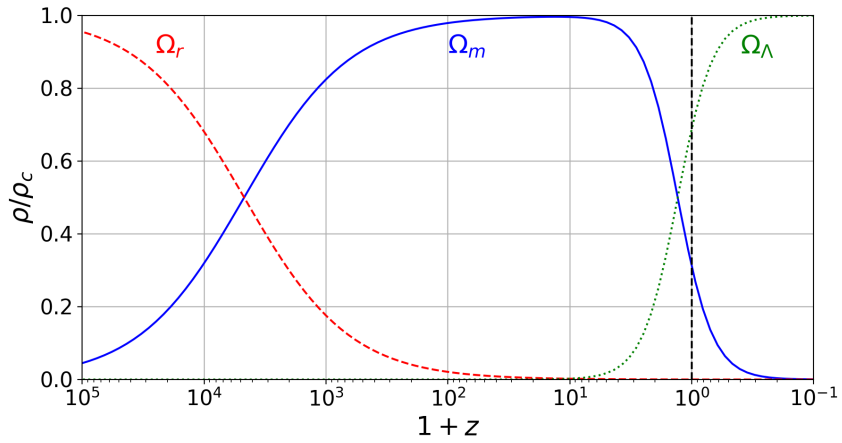

Figure 1.4 – The density parameters as a function of redshift, with the values provided in Table 1.2: $\Omega_m(z) = \Omega_c(z) + \Omega_b(z)$ and $\Omega_r(z) = \Omega_\gamma(z) + \Omega_\nu(z)$ (neutrinos are considered massless for simplicity).

The sign of q_0 stems from the original expectation that the Universe is filled with matter and radiation only and hence its expansion should decelerate. Expanding $H = \dot{a}/a$ at first order we find:

$$H(z) = H_0 \left[1 + (q_0 + 1)\, z\right] + \mathcal{O}(z^2), \tag{1.48}$$

and for the luminosity distance:

$$D_{\rm L}(z) = H_0^{-1} \left[1 - \frac{1}{2}\,(q_0 - 1)\, z\right] z + \mathcal{O}(z^3). \tag{1.49}$$

Therefore, when $z \ll 1$, the Hubble diagram is only sensitive to H_0, then it depends on q_0 for $z \lesssim 0.5$, and finally on all density parameters as $z \gtrsim 1$. In 1998, two different collaborations, the High-Z Supernova Search Team (HZT Riess et al., 1998) and the Supernova Cosmology

Project (SCP Perlmutter et al., 1999) measured $q_0 < 0$ based on the Hubble diagram of 50 and 60 SN Ia, respectively, thus providing evidence for the acceleration of the Universe expansion. The energy component that creates this acceleration is generically called *dark energy*.

Using the second Friedmann equation (1.37) and the equation of state (1.39),

$$q_0 = \frac{1}{2}\sum_X (1 + 3w_X)\,\Omega_{X,0}. \tag{1.50}$$

Therefore, for q_0 to be negative, there should exist at least one component with $w_X < -1/3$. In a ΛCDM Universe and neglecting radiation today (we will see this is legitimate in Section 1.2.3):

$$q_0 = \frac{1}{2}\Omega_{m,0} - \Omega_{\Lambda,0}, \tag{1.51}$$

$q_0 < 0$ if and only if $\Omega_{\Lambda,0} > \Omega_{m,0}/2$. Note that this result is independent of the curvature $\Omega_{K,0}$.

In 1998 $\Omega_{\Lambda,0}$ was found to be strictly positive at 99% confidence level. Today's best constraint from SN Ia alone is:

$$\Omega_{\Lambda,0} = 0.702 \pm 0.022 \tag{1.52}$$

within the flat ΛCDM model, using the Pantheon sample (Scolnic et al., 2018).

As stated above, the cosmological constant can be interpreted as a perfect fluid of density $\rho_\Lambda = \Lambda/(8\pi G_\mathrm{N})$ and equation of state $w_\Lambda = -1$. Though such a fluid cannot be found in classical physical systems, it is expected in quantum physics. Applying the cosmological principle, the energy-momentum tensor for vacuum energy is indeed:

$$T_{\mu\nu} = -\rho_v g_{\mu\nu}, \tag{1.53}$$

which is similar to that of Λ. However, predictions for the vacuum energy based on quantum field theory exceed the observed energy density associated to the cosmological constant by 60 to 120 orders of magnitude. This is known as the cosmological constant problem.

Hence, giving up on dark energy to match vacuum energy, on can explore deviations to a pure cosmological constant $w = -1$. Letting w free, using CMB TT, TE, EE, lowE + lensing, BAO and SN Ia data within a flat CDM model Planck Collaboration et al. (2018) measure $w = -1.028 \pm 0.032$, compatible with a cosmological constant ($w = -1$). Further allowing for a time variation of w, the Chevallier-Polarski-Linder (CPL) parametrisation (Chevallier and Polarski, 2001; Linder, 2003) reads:

$$w(a) = w_0 + (1 - a)\,w_a. \tag{1.54}$$

With the same data set and model, Planck Collaboration et al. (2018) find $w_0 = -0.961 \pm 0.077$ and $w_a = -0.28^{+0.31}_{-0.27}$, again fully compatible with a cosmological constant ($w_0 = -1$ and $w_a = 0$).

Note however that Eq. (1.54) does not provide a physical explanation for dark energy. The most natural one is a quintessence field ϕ evolving in a potential $V(\phi)$, such that Eq. (1.12) becomes:

$$\mathcal{S} = \frac{1}{16\pi G_\mathrm{N}}\int d^4x\sqrt{-g}R - \frac{1}{16\pi G_\mathrm{N}}\int d^4x\,[\partial_\mu\phi\partial^\mu\phi + 2V(\phi)] + \int d^4x\sqrt{-g}\mathcal{L}_m. \tag{1.55}$$

In this case, the equation of state of the homogeneous scalar field is given by:

$$w_\phi = \frac{\dot\phi^2 - 2V(\phi)}{\dot\phi^2 + 2V(\phi)}. \tag{1.56}$$

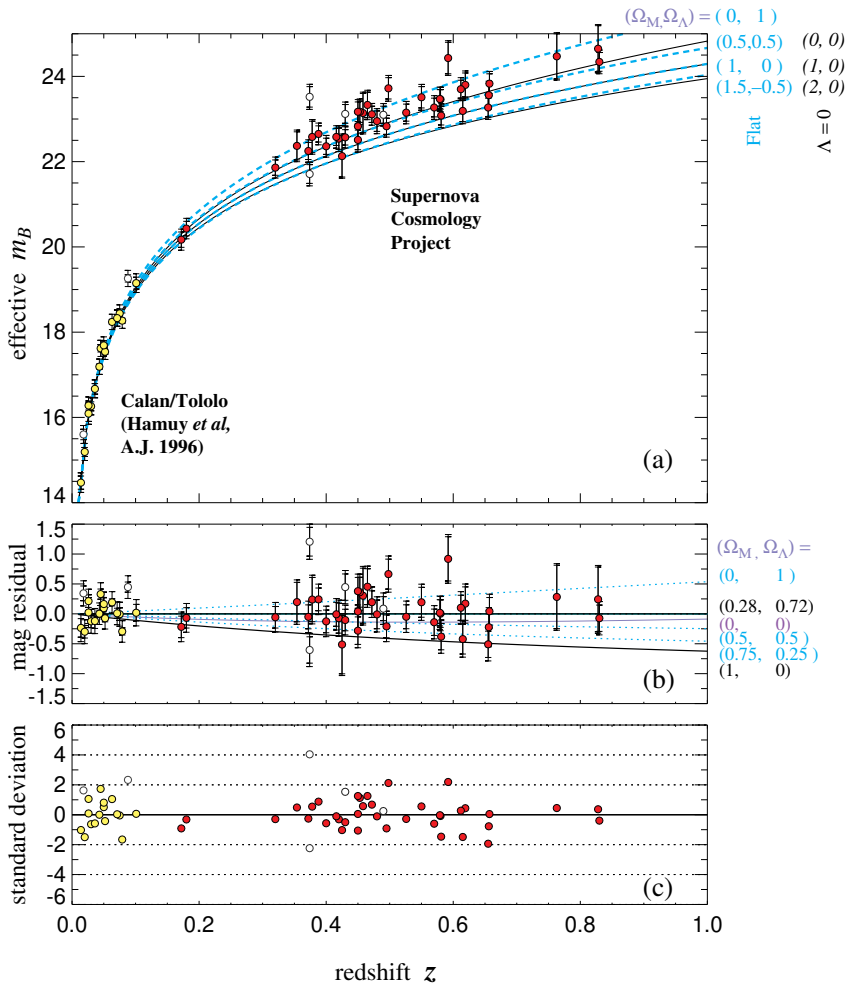

Figure 1.5 – Hubble diagram of 42 high redshift SN Ia from the Supernova Cosmology Project, and 18 low redshift supernovae from the Cálan/Tololo Supernova Survey. ΛCDM predictions are plotted with different $(\Omega_{m,0}, \Omega_{\Lambda,0})$ values. Taken from Perlmutter et al. (1999).

To obtain $w_\phi \simeq -1$, as favoured by current observations, the field potential energy $V(\phi)$ must dominate over the kinetic term.

However, neither a cosmological constant nor a more general scalar field explain why dark energy just starts to dominate (see Figure 1.4). This question is even more striking if we note that, assuming a cosmological constant, the expansion of the Universe will eventually be exponential:

$$a(t) \propto e^{Ht} \qquad \text{with} \qquad H = \sqrt{\frac{\Lambda}{3}}. \tag{1.57}$$

One would note that the advent of dark energy is somewhat concomitant to the emergence of non-linear structure formation, as will be presented in Chapter 2. In addition, it is not clear whether Friedmann equations are a correct description of the average dynamics of the structured Universe. Therefore, some tried to explain the recent acceleration of the expansion by the feedback due to non-linear structure formation. Though the idea appears extremely appealing, simulations have shown this backreaction to be too small to explain the observed acceleration of the expansion of the Universe (Schwarz, 2010).

The alternative to a dark energy fluid is a modification of gravity itself, just as Λ can be thought of as a natural addition to the Einstein-Hilbert action (1.11). Many modified gravity models have been proposed, including:

1. scalar theories: a scalar field interacts with the metric. As an example, $f(R)$ theories replace the Ricci tensor R of the Einstein-Hilbert action (1.11) by a function of it, $f(R)$. f is conveniently chosen such that the effect of modified gravity is suppressed in the large or rapidly changing curvature limit, through the so-called chameleon mechanism.
2. extra-dimension theories: extra dimensions are introduced, such that the acceleration of the expansion is driven by the extra dimensions, and standard general relativity is recovered at small scales, through the so-called Vainshtein mechanism.
3. massive gravity: the graviton is given a mass. Again, the Vainshtein mechanism ensures general relativity is recovered on small scales. However, these theories are prone to instabilities.
4. Horndeski theories: most general theories of gravity involving a scalar field and a metric tensor in 4 dimensions, such that equations of motion are second order.

Many modified gravity models were ruled out by the constraint that the speed of gravitational waves should (almost) equate the speed of light, as imposed by the recent observation of a gravitational wave with electromagnetic counterpart (LIGO Scientific Collaboration and Virgo Collaboration, 2017).

Therefore, the cosmological constant Λ appears as the simplest parametrisation satisfying current observations. We will thus consider the ΛCDM model in the rest of this manuscript (except otherwise stated).

1.2.2 Baryonic matter

Baryonic matter is a synonym for regular matter. Today, baryonic matter resides in gravitationally bound systems such as galaxies, stars, and in the circumgalactic medium.

In earlier times, when $T \gg 1\,\mathrm{MeV}$, the Universe is filled with relativistic particles at equilibrium: electrons, positrons, neutrinos and photons, and non-relativistic neutrons and protons. The ratio of neutron to proton densities is:

$$\left(\frac{n}{p}\right)_{\mathrm{eq}} \simeq \exp\left[-\frac{Q}{T}\right], \tag{1.58}$$

with $Q = m_n - m_p \simeq 1.293\,\mathrm{MeV}$ the mass difference between neutrons and protons.

Around $T \simeq 0.8\,\mathrm{MeV}$, the weak interaction cannot maintain equilibrium anymore; neutrinos decouple and the neutron to proton ratio is of order $(n/p)_{\mathrm{f}} \simeq \exp(-1.293/0.8) \simeq 0.2$. Then, neutrons decay into protons:

$$n \to p + e + \bar{\nu}_e \tag{1.59}$$

with a lifetime $\tau_n \simeq 880\,\mathrm{s}$.

Nuclei can almost only be formed through two-body reactions due to the low density of neutrons and protons preventing higher order interactions. Then, deuterium must be formed first to allow the production of heavier elements. Because of its low binding energy, deuterium is easily photodissociated and its production starts only around $T_{\mathrm{nuc}} \simeq 0.066\,\mathrm{MeV}$. At that time, namely $t_{\mathrm{nuc}} \simeq 300\,\mathrm{s}$, $(n/p)_{\mathrm{nuc}} = (n/p)_{\mathrm{f}} \exp(-t_{\mathrm{nuc}}/\tau_n) \simeq 0.14$. Then, deuterium combine rapidly

in helium-4, which is much favoured due to its higher binding energy. Since helium-4 nuclei contain virtually all neutrons, their mass abundance is:

$$Y_p \simeq \left(\frac{2n}{n+p}\right)_{\rm nuc} = 2\frac{(n/p)_{\rm nuc}}{1+(n/p)_{\rm nuc}} \simeq 0.25. \tag{1.60}$$

The decrease of temperature and density, the absence of stable nuclei with masses $A = 5$ and $A = 8$ and the increased Coulomb repulsion stop the primordial nucleosynthesis after lithium-7. At the end of Big Bang nucleosynthesis, $T \simeq 0.03\,\mathrm{MeV}$ ($t \simeq 1000\,\mathrm{s}$), apart from helium-4, traces of deuterium (mass abundance of $\sim 10^{-4}$), helium-3 (mass abundance of $\sim 10^{-4}$, including tritium (vanishing through beta decay), and lithium-7 ($\sim 10^{-9}$, including beryllium-7) remain. Unlike helium-4, the relative abundances of these primordial nuclei depend heavily on the ratio of baryon-to-photon densities $\eta_{b\gamma} = n_b/n_\gamma$.

Therefore, a measurement of the energy density of baryons can be obtained by probing the concentration of the different light nuclei produced by the Big Bang nucleosynthesis. At 95% confidence level $\omega_{b,0}$ is found to be (Tanabashi et al., 2018):

$$0.021 < \omega_{b,0} < 0.024. \tag{1.61}$$

Heavier nuclei are then produced in the first stars through stellar nucleosynthesis.

1.2.3 Radiation

Radiation comprises all relativistic species at a given redshift z. The only relativistic species today are the photons from the cosmic microwave background, relativistic neutrinos and any other possible light thermal relic.

1.2.3.1 Photons

In the early Universe ($T \sim 100\,\mathrm{eV}$), the interaction rate Γ between photons and electrons of:

$$p + e \longleftrightarrow H + \gamma \tag{1.62}$$

is much larger than the expansion rate H. Therefore, protons, electrons, hydrogen and photons are in thermodynamic equilibrium and treated as Fermi-Dirac and Bose-Einstein gases. Their chemical potentials obey:

$$\mu_p + \mu_e = \mu_H + \mu_\gamma \tag{1.63}$$

with $\mu_\gamma = 0$. At that time, all baryons are non-relativistic. In this limit, their Fermi-Dirac distribution can be integrated over their momentum to yield their density:

$$n_X = g_X \left(\frac{m_X T}{2\pi}\right)^{3/2} e^{(\mu_X - m_X)/T} \qquad X \in \{e, p, H\}\,, \tag{1.64}$$

where g_X is the degeneracy. Let us define the ionisation fraction $X_e = n_e/\left(n_p + n_H\right) = n_e/n_b$. We also have $n_p = n_e$, following from electric neutrality. Then, taking $m_H = m_p$ in the prefactor of Eq. (1.64), Eq. (1.63) can be recast into the Saha equation:

$$\frac{X_e^2}{1 - X_e} = \left(\frac{m_e T}{2\pi}\right)^{3/2} \frac{e^{-E_I/T}}{n_b}, \tag{1.65}$$

where $E_I = m_e + m_p - m_H = 13.6\,\mathrm{eV}$ is the hydrogen ionisation energy. The number density of baryons is related to that of photons through the baryon-to-photon ratio $\eta_{b\gamma} = n_b/n_\gamma$. The

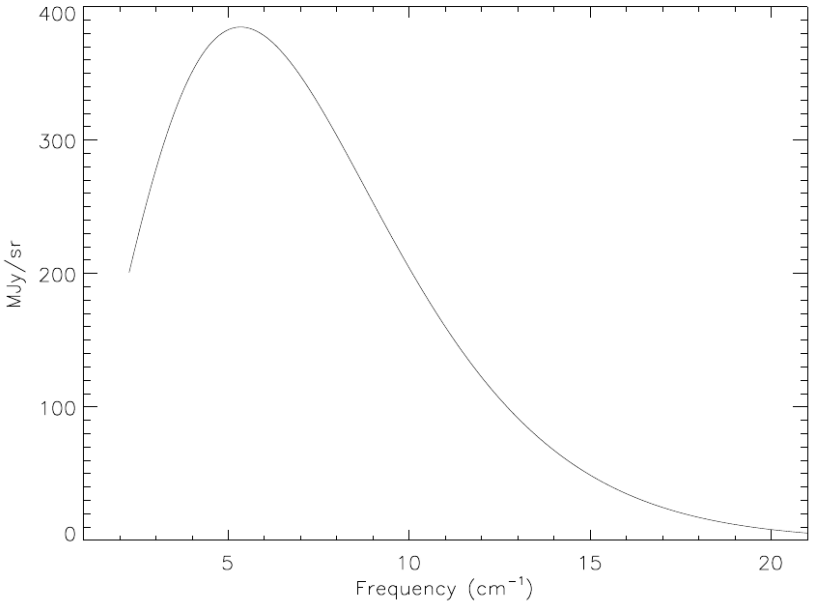

Figure 1.6 – Measurement of the microwave radiation spectrum by the instrument FIRAS of the COBE satellite, and the black body spectrum best fit. Error bars are hidden in the thickness of the line. Taken from Fixsen et al. (1996).

photon density n_γ follows from integrating the Bose-Einstein distribution function ($2\zeta(3)/\pi^2 \simeq 0.2436$):

$$n_\gamma = 2\frac{\zeta(3)}{\pi^2}T^3. \tag{1.66}$$

The baryon-to-photon ratio has been measured to be $\eta_{b\gamma} \simeq 5 \times 10^{-10}$ from the relative fraction of nuclei produced during the Big Bang nucleosynthesis (see Section 1.2.2). Therefore, defining the recombination by $X_e = 1/2$, we find $T_{\rm rec} \simeq 3500\,{\rm K}$. So far we have assumed all gases to be in thermodynamic equilibrium; actually the Boltzmann equation should be solved to take into account the dynamic process of recombination of both helium and hydrogen atoms. At the end of recombination, the electron density falls rapidly; and so does the photon-electron scattering rate, until photons decouple from the baryon plasma — around the last scattering redshift $z_\star = 1089.80 \pm 0.21$ as measured by Planck Collaboration et al. (2018) (see Table 1.2).

Photons then free-stream, forming the cosmic microwave background, with a spectral energy distribution following a black body spectrum. The current CMB temperature was measured to a very high precision by COBE/FIRAS (see Figure 1.6): $2.7255 \pm 0.0006\,{\rm K}$ (Fixsen, 2009).

From the photon energy density:

$$\rho_\gamma = 2\frac{\pi^2}{30}T^4 \tag{1.67}$$

and Eq. (1.41), we find:

$$\omega_{\gamma,0} = (2.4729 \pm 0.0022) \times 10^{-5}. \tag{1.68}$$

Apart from foregrounds (galaxy emission, dust emission), small deviations to the black body spectrum (μ, y distortions) could be expected, but have not yet been detected. However, some angular anisotropies have been found; first, a dipole ($\delta T/T \sim 10^{-3}$), compatible with our motion with respect to the CMB rest-frame, and smaller fluctuations, $\delta T/T \sim 10^{-5}$, which will be described in the next section.

1.2.3.2 Neutrinos

Very similarly to the cosmic microwave background, neutrinos decouple from electrons when the rate of weak interactions becomes smaller than the expansion rate (when $T \sim 1\,\mathrm{MeV}$). Because of entropy conservation, photons get reheated after neutrino decoupling. Their temperature T_γ increases by the cubic root of the ratios of the effective number of relativistic degrees of freedom before and after decoupling, $(11/4)^{1/3}$. Assuming neutrino temperature was not changed by decoupling, their temperature today is thus $T_\nu = (4/11)^{1/3} T_\gamma$. Hence, the energy density per family of massless neutrinos (particles and antiparticles) is:

$$\rho_\nu = \frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\rho_\gamma, \tag{1.69}$$

the factor 7/8 arising due to the fermionic nature of neutrinos. Therefore, the parameter density of massless ($m_\nu \lesssim 10^{-4}\,\mathrm{eV}$) neutrinos is:

$$\omega_{\nu,0} = (5.6161 \pm 0.0049) \times 10^{-6} \times N_\nu \tag{1.70}$$

with N_ν the number of neutrino families; 3, in the standard model. However, neutrino decoupling is not instantaneous, which is accounted for by taking the effective number of neutrinos to be $N_{\mathrm{eff}} = 3.046$ (Mangano et al., 2002; de Salas and Pastor, 2016). This number would increase if there existed other thermal relics. The effective number of degrees of freedom in non-photon radiation density was measured to be $N_{\mathrm{eff}} = 2.99 \pm 0.17$, including CMB and BAO data, by Planck Collaboration et al. (2018), consistent with the 3.046 value. In this manuscript we thus do not consider any other thermal relics. Then, Eq. (1.69) would hold as long as neutrinos are relativistic. From the neutrino oscillations detected by the Super-Kamiokande and Sudbury Neutrino Observatory experiments (Kajita, 1999; Ahmad et al., 2001), we know that at least one of the neutrino mass eigenstates is non-zero. Therefore, the neutrino parameter density today is (if neutrino masses are almost degenerate, Mangano et al. 2005):

$$\omega_{\nu,0} = \frac{\sum_\nu m_\nu}{93.14\,\mathrm{eV}}. \tag{1.71}$$

Constraints on neutrinos masses can be inferred from large scale structure analyses. Based on the Lyman-α absorption in quasar spectra Palanque-Delabrouille et al. (2019) find $\sum_\nu m_\nu < 0.71\,\mathrm{eV}$ at the 68% confidence level. Combining with CMB, lensing and BAO data, they find $\sum_\nu m_\nu < 0.113\,\mathrm{eV}$ at the 68% confidence level.

The best limit on the sum of neutrino masses from particle physics experiments is $\sum m_\nu < 1.1\,\mathrm{eV}$ at 90% confidence level, as recently determined by Katrin (Aker et al., 2019) from tritium beta decay.

In the rest of this manuscript, we fix $\sum_\nu m_\nu$ to its lower limit of 0.06 eV — except otherwise stated, and consider one massive and two massless neutrinos, as in Planck Collaboration et al. (2018).

1.2.4 Dark matter

Dark matter was introduced as an answer to the hereafter described observations.

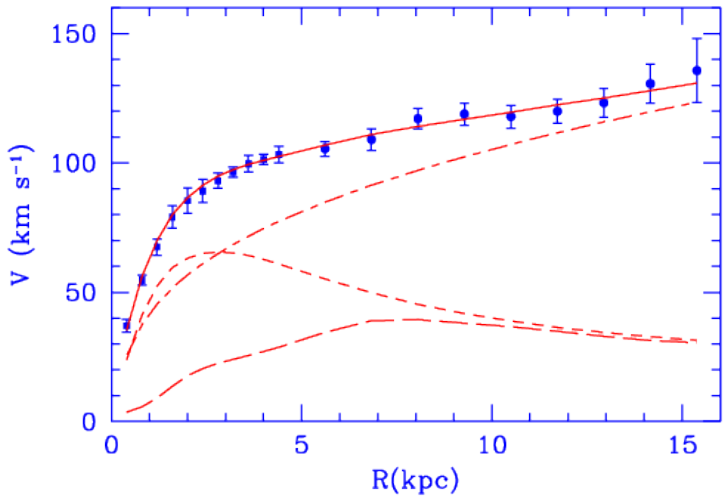

Figure 1.7 – Rotation curve of galaxy M33 (points with error bars) inferred from observations of the 21 cm line emission of HI. The best fit model is shown in continuous line. The stellar, gas and dark matter contribution are shown in short dashed, long dashed and dashed dotted lines, respectively. Taken from Corbelli and Salucci (2000).

1.2.4.1 Galaxies

Rotation curves represent the orbital velocity of all galaxy components (stars, gas...) as a function of the distance to the centre of a galaxy.

Following Newtonian dynamics (legitimate in the weak field and non-relativistic limit), the Keplerian velocity is related to the mass inside the circular orbit following:

$$\frac{v^2(r)}{r} = \frac{G_{\rm N} M(<r)}{r^2}. \tag{1.72}$$

The luminosity of spiral galaxies exponentially decreases with the distance to the centre, with a typical scale R_d. If galaxy luminosity traced all the matter contained in the galaxy, $M(<r)$ should be constant further away than a few times the radius of the optical disk R_d. Therefore,

$$v(r) \propto \frac{1}{\sqrt{r}} \qquad r > R_d, \tag{1.73}$$

which is not what is observed: rotation curves of spiral galaxies tend to flatten to a velocity v_∞, as exemplified in Figure 1.7. The missing mass is attributed to dark matter. Typical density profiles of dark matter halos follow $\rho_{\rm DM} \propto r^{-2}$ at large radius, though this trend varies from one galaxy to another (e.g., for Corbelli and Salucci (2000), $\rho_{\rm DM} \propto r^{-1.3}$).

Apart from rotation curves, the total mass in galaxies can also be probed by strong gravitational lensing. Of course, galaxy baryonic mass is not only located in stars but also in diffuse gas, which can be detected through e.g. X emission. Gas and stars are found to represent $\simeq 15\%$ of total galaxy masses (e.g. Tortora et al., 2019).

Hence, if Newtonian gravity holds on galaxy scales, most of the mass in galaxies is under the form of dark matter.

1.2.4.2 Galaxy clusters

Zwicky (1937) found that the velocity dispersion of the Coma cluster suggested a much larger mass than was present in its luminous component.

Clusters are gravitationally bound and thus can be considered in steady state. The virial theorem (see Section 2.2.2.1) allows the gravitational mass to be related to the galaxy velocity dispersion $\langle v^2 \rangle$ and R_h, the radius containing half of the mass:

$$M_{\rm vir} = \frac{\langle v^2 \rangle R_h}{\alpha G_{\rm N}}, \tag{1.74}$$

with α a coefficient depending on the density profile, typically 0.4 for galaxy clusters. Another method to measure the total mass involves strong gravitational lensing in the centre region of the cluster ($10\,{\rm kpc} \lesssim r \lesssim 200\,{\rm kpc}$) and weak gravitational lensing in the outer regions ($r \lesssim 1\,{\rm Mpc}$). Measuring the density profile and temperature of hot gas, captured in the gravitational potential, through X emission also allows constraints to be put on the total mass. Typically, stellar and gas masses are found to represent 5% and 12% of the cluster total (virial) mass, and thus the ratio $\omega_{b,0}/\omega_{c,0}$ is expected to be $\geq 17\%$ (Bahcall, 1996). Hence, given the ω_b value measured by Big Bang nucleosynthesis (1.61) we find that $\Omega_{m,0} \simeq 0.25$.

We will see in the next section that baryon acoustic oscillations can be used to put a strong constraint on the ratio $\omega_{b,0}/\omega_{m,0}$. Altogether, observations of the CMB, measurements of the acceleration of the expansion through SN Ia, gravitational weak lensing, and galaxy clustering analyses provide cosmological constraints on the density of total matter, which is of order 30%.

1.2.4.3 Cold dark matter

Dark matter is said to be cold or hot depending on whether it was relativistic when it decoupled from the primordial plasma.

Cold dark matter (CDM) particles are massive particles weakly interacting with themselves and baryons, radiation and dark energy and are generically called WIMP (Weakly Interacting Massive Particle). We have already noticed that the temperature and thus baryon density fluctuations were of order 10^{-5} at the decoupling (at $a \simeq 1 \times 10^{-3}$); we will see in the next section that these fluctuations grow as a in the matter dominated epoch, thus making it hard to form a density contrast of 1 as is typical today. Cold dark matter alleviates this issue as it forms potential wells earlier than baryonic matter since it is not coupled to radiation. In addition, the standard model of particle physics provides numerous candidates for cold dark matter particles, whose cross-section is found to be typical of the weak interaction.

Within the cold dark matter paradigm structure form hierarchically: small structures form before larger ones. The density profile $\rho(r)$ of dark matter halos has been extensively studied with N-body simulations. These helped determined analytic forms for $\rho(r)$, the most common being the Navarro-Frenk-White profile (Navarro et al., 1996). Note however that the shape of the profile in the smallest scales ($r \sim 0.1\,{\rm kpc}$) is still not very well constrained by simulations. There are also a number of unresolved issues, among them:

1. the number of satellite galaxies in our local group is expected to be of the order of 1000, based on numerical simulations. However, we only observe 36 galaxies within 0.5 Mpc.
2. density profiles are more peaked at $r \to 0$ than what observations suggest (de Blok, 2010).

These two issues may be solved by baryonic feedback (involving the dynamics of the gas). To tackle the apparent issues of CDM, many other dark matter models have been proposed: hot dark matter, warm dark matter, fuzzy dark matter, decaying dark matter, black holes, etc.

Neutrinos would have been a good candidate for dark matter, since they interact very weakly with other particles. However, as in any other hot dark matter model, structure formation is

suppressed below the free-streaming scale of neutrinos, thus provoking a cut-off in the power spectrum of density fluctuations. Given current cosmological constraints on their masses (see Section 1.2.3), neutrinos cannot represent the total dark matter [1].

In the rest of this manuscript we will only consider weakly interacting cold dark matter.

1.2.5 Inflation

We have already noted that curvature density today is found to be small, $|\Omega_{K,0}| \lesssim 0.002$ at 1σ, combining CMB and BAO measurements (Planck Collaboration et al., 2018). However, its value must have been even lower in the past. Indeed, using Eq. (1.44) and Eq. (1.45):

$$\Omega_K = \Omega_{K,0}\left[\sum_X \Omega_{X,0}(1+z)^{1+3w_X} + \Omega_{K,0} + \Omega_{\Lambda,0}(1+z)^{-2}\right]^{-1}. \tag{1.75}$$

For a ΛCDM Universe and neglecting Ω_Λ since we work at high redshift:

$$\Omega_K = \Omega_{K,0}\left[\Omega_{m,0}\,(1+z) + \Omega_{r,0}\,(1+z)^2 + \Omega_{K,0}\right]^{-1}. \tag{1.76}$$

Therefore, the curvature density becomes increasingly small as $z \to \infty$. At the redshift of equality $z_{\rm eq} \simeq 3400$ (see Table 1.2) when the radiation energy density equates that of matter:

$$\Omega_{m,0}\,(1+z_{\rm eq})^3 = \Omega_{r,0}\,(1+z_{\rm eq})^4\,, \tag{1.77}$$

we find $|\Omega_K(z_{\rm eq})| < 5\times10^{-6}$. At Planck scale $z_{\rm Pl} \simeq 5\times10^{31}$, below which a quantum description of gravity is required, we find $|\Omega_K| < 2\times10^{-62}$. Such constraints are considered as fine tuning and thus call for a mechanism — dubbed *inflation* — to make the early value of $|\Omega_K|$ more natural. Based on Eq. (1.75), a simple answer is that the energy content of the early Universe is dominated by a fluid with $w < -1/3$. Let us write t_i, t_f the cosmic time at the beginning and end of inflation. We define the number of e-folds as:

$$N = \ln\left[\frac{a(t_f)}{a(t_i)}\right]. \tag{1.78}$$

Using Friedmann equations (1.36 - 1.37), during inflation (neglecting curvature and excluding $w < -1$):

$$\dot{H} = \frac{\ddot{a}}{a} - H^2 \simeq -4\pi G_{\rm N}\,(\rho + P) \lesssim 0. \tag{1.79}$$

Hence $H(t_f) \lesssim H(t_i)$; using Eq. (1.44) and the number of e-folds (1.78):

$$\left|\frac{\Omega_K(t_f)}{\Omega_K(t_i)}\right| = \left[\frac{a(t_f)}{a(t_i)}\frac{H(t_f)}{H(t_i)}\right]^{-2} \gtrsim \left[\frac{a(t_f)}{a(t_i)}\right]^{-2} = e^{-2N}. \tag{1.80}$$

If we ask for $|\Omega_K(t_f)| < 10^{-60}$ at the end of the inflation, while allowing a natural $|\Omega_K(t_i)| = \mathcal{O}(1)$, then we need $N \gtrsim 70$.

Another apparent issue is the remarkable homogeneity of the CMB. The last scattering surface is measured to be at $z_\star = 1089.80 \pm 0.21$ (see Table 1.2). The particle horizon, which delineates the regions of the Universe which have been in causal contact, is, at this redshift:

$$\eta(z_\star) = \int_{z_\star}^{\infty} \frac{dz}{H(z)}. \tag{1.81}$$

1. For an (old) review of the debates on dark matter, listen to http://www.astronomy.ohio-state.edu/~dhw/Rap/index.html.

Assuming only matter and radiation we find $\eta(z_\star) \simeq 3 \times 10^2\,\mathrm{Mpc}$. Dividing by $D_\mathrm{M}(z_\star)$ leads to an angular scale of $\sim 1^\circ$ on the sky. That the CMB temperature appears so homogeneous on much larger scales is puzzling. The same issue applies to temperature perturbations, measured in the CMB to be correlated on scales much larger than 1° and, equally, to density perturbations, which we see today below the particle horizon in the large scale structure, but were above horizon in the past. Therefore, a mechanism is required to causally connect these regions, which can be achieved through an early phase of accelerated expansion.

1.2.5.1 Scalar field

We have already seen in Section 1.2.1 that such an accelerated expansion can be achieved with a scalar field ϕ — dubbed the inflaton. Let us determine the shape of its potential $V(\phi)$. Varying the metric and ϕ in the action (1.55) (neglecting the matter Lagrangian) yields the Friedmann and Klein-Gordon equations (Peter and Uzan, 2012):

$$H^2 = \frac{8\pi G_\mathrm{N}}{3}\left(\frac{1}{2}\dot\phi^2 + V(\phi)\right) - \frac{K}{a^2} \tag{1.82}$$

$$\frac{\ddot a}{a} = \frac{8\pi G_\mathrm{N}}{3}\left(V(\phi) - \dot\phi^2\right) \tag{1.83}$$

$$0 = \ddot\phi + 3H\dot\phi + V_{,\phi}(\phi) \tag{1.84}$$

with $V_{,\phi}$ the derivative of V with respect to ϕ. Neglecting curvature since it will eventually be negligible, Friedman equations (1.82) and (1.83) give:

$$\dot H = \frac{\ddot a}{a} - H^2 = -4\pi G_\mathrm{N}\dot\phi^2. \tag{1.85}$$

Requiring an accelerated expansion, i.e. $\ddot a/a$ to be large, the field ϕ should not vary much during inflation (see Eq. (1.83)):

$$\dot\phi^2 \ll V(\phi) \qquad \ddot\phi \ll 3H\dot\phi. \tag{1.86}$$

This phase is dubbed as *slow-roll*. Then Eq. (1.82) and Eq. (1.84) yield:

$$H^2 = \frac{8\pi G_\mathrm{N}}{3}V \qquad 3H\dot\phi = -V_{,\phi}. \tag{1.87}$$

Thus, from the first equation and Eq. (1.85):

$$\left|\dot H\right| \ll H^2 \tag{1.88}$$

and using both equations (1.87):

$$\ddot\phi = -\frac{d}{dt}\left(\frac{V_{,\phi}}{3H}\right) \simeq -\frac{V_{,\phi,\phi}\dot\phi}{3H} = \frac{1}{24\pi G_\mathrm{N}}\frac{V_{,\phi}V_{,\phi,\phi}}{V}, \tag{1.89}$$

such that conditions (1.86) translate into:

$$\left(\frac{V_{,\phi}}{V}\right)^2 \ll 24\pi G_\mathrm{N} \qquad \frac{|V_{,\phi,\phi}|}{V} \ll 24\pi G_\mathrm{N}. \tag{1.90}$$

The potential must be very flat. Hence, the potential can be Taylor expanded to study the early dynamics of inflation.

1.2.5.2 Initial conditions

For the inflation to be predictive, the sensitivity of the Universe expansion with respect to initial conditions should be small. As long as its evolution is monotonous, ϕ can be used as time variable such that Eq. (1.85) is re-expressed as:

$$H_{,\phi} = -4\pi G_{\rm N}\dot{\phi}, \tag{1.91}$$

and Eq. (1.82) becomes:

$$H_{,\phi}(\phi)^2 - 12\pi G_{\rm N} H^2(\phi) = -32\pi^2 G_{\rm N}^2 V(\phi). \tag{1.92}$$

Perturbing $H(\phi) = H_{sr}(\phi) + \delta H(\phi)$, with $H_{sr}(\phi)$ the reference slow-roll solution of Eq. (1.92), $\delta H(\phi)$ satisfies the linearised Eq. (1.92):

$$H_{sr,\phi}(\phi)\delta H_{,\phi}(\phi) = 12\pi G_{\rm N} H_{sr}(\phi)\delta H(\phi) \tag{1.93}$$

whose solution is given by:

$$\delta H(\phi) = \delta H(\phi_i)\exp\left[12\pi G_{\rm N}\int_{\phi_i}^{\phi} d\phi \frac{H_{sr}(\phi)}{H_{sr,\phi}(\phi)}\right]. \tag{1.94}$$

with ϕ_i the initial value of the field. The number of e-folds can be written as:

$$N = \int_{t_i}^{t} dt H(t) = \int_{\phi_i}^{\phi} d\phi \frac{H(\phi)}{\dot{\phi}} = -4\pi G_{\rm N}\int_{\phi_i}^{\phi} d\phi \frac{H(\phi)}{H_{,\phi}(\phi)}. \tag{1.95}$$

Therefore, the initial perturbation $\delta H(\phi_i)$ is exponentially suppressed with the number of e-folds:

$$\delta H(\phi) = \delta H(\phi_i) e^{-3N(\phi)}. \tag{1.96}$$

The final value of H is thus insensitive to that at the beginning of inflation.

1.2.5.3 Perturbations

At first order, the initial quantum perturbations $\delta\phi$ around the homogenous solution ϕ (see Eq. (1.84)) fluctuate following the Klein-Gordon equation; in Fourier space:

$$\ddot{\delta\phi} + 3H\dot{\delta\phi} + \frac{k^2}{a^2}\delta\phi + V_{,\phi,\phi}(\delta\phi) = 0 \tag{1.97}$$

with k the wavenumber, corresponding to the inverse of the perturbation wavelength. This wavelength is usually compared to the comoving Hubble radius $1/(aH) = 1/\mathcal{H}$ (with $\mathcal{H} = aH$) to identify two regimes: super-Hubble ($k \ll \mathcal{H}$) and sub-Hubble ($k \gg \mathcal{H}$). In the sub-Hubble regime, k^{-1} the oscillation time of a mode k is small compared to the Hubble rate, which can typically be neglected in the equations. Conversely, in the super-Hubble regime, Hubble friction terms dominate the field dynamics.

This is the case during inflation for the super-Hubble modes, which freeze out in classical perturbations due to the (dominant) friction term $3H\dot{\delta\phi}$ of Eq. (1.97). These perturbations are adiabatic, (almost) Gaussian distributed with a nearly scale invariant spectrum (see next Section 1.3.1 for the definition). Namely, the dimensionless power spectrum of the spatial curvature perturbations $\mathcal{R}$ (gravitational potential on hypersurfaces $\delta\phi = 0$) is parametrised as:

$$\boxed{\frac{k^3}{2\pi^2}P_{\mathcal{R}}(k) = A_s\left(\frac{k}{k_0}\right)^{n_s - 1 + \frac{1}{2}\alpha_s \ln(k/k_0)}} \tag{1.98}$$

with A_s the amplitude of the power spectrum (at pivot scale k_0) and n_s the spectral index. A_s (at $k_0 = 0.05\,\mathrm{Mpc}^{-1}$) and n_s were precisely measured in the CMB by Planck Collaboration et al. (2018), see Table 1.2. In particular, they find $n_s = 0.9665 \pm 0.0036$, showing a neat deviation to scale invariance ($n_s = 1$). Precisely measuring n_s is of great importance, as for power law potentials $V \sim \phi^p$, n_s is related to the number of e-folds of inflation N through $1 - n_s = (p+2)\,/\,(2N)$. In addition, α_s can be non-zero: this is called the running of the spectral index. In standard single-field slow-roll models of inflation, α_s is of the order of $(n_s - 1)^2 \sim 10^{-3}$ (given $n_s \simeq 0.96$); any deviation would exclude these simplest models. Current constraints from Planck Collaboration et al. (2018) — $\alpha_s = -0.0041 \pm 0.0067$ using CMB and BAO data — do not show any departure from a power-law power spectrum; we therefore take $\alpha_s = 0$ in the following.

No vector mode is produced, but inflation generates primordial gravitational waves. The tensor-to-scalar ratio r is constrained by CMB observations through the additional Sachs-Wolfe effect induced by tensor modes (see next Section 1.3.4). Figure 1.8 shows the current best constraints on n_s and $r_{0.002}$ (r given at the pivot scale $k = 0.002\,\mathrm{Mpc}^{-1}$) from Planck Collaboration et al. (2018). They find a small tensor-to-scalar ratio $r_{0.002} < 0.058$ at the 95% confidence level, using Planck (`Plick` likelihood), BICEP2/Keck (BICEP2 Collaboration et al., 2016) and BAO data. The detection of primordial gravitational waves would be a major discovery in cosmology and physics in general.

In the single-field slow-roll scenario discussed so far, some non-Gaussianities can be produced at the second order in the perturbations and are thus expected to be small. However, many inflationary models (multifield, topological defects, curvaton...) predict levels of non-Gaussianity which could be detectable. Detection (or non-detection) of deviations to Gaussianity on large scales would thus be a signature of the dynamics and the field content during inflation. Non-Gaussianities are usually parametrised through a quadratic term in the Bardeen potential Ψ (see next Section 1.3.2 for the definition):

$$\Psi(\mathbf{k}) = \Psi_{\rm lin}(\mathbf{k}) + \int \frac{d^3k_1 d^3k_2}{(2\pi)^3} f_{\rm NL}^{\Psi}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}) \Psi_{\rm lin}(\mathbf{k}_1) \Psi_{\rm lin}(\mathbf{k}_2) \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2). \tag{1.99}$$

Therefore, non-Gaussianities are most naturally probed by the bispectrum, given (at tree level), by:

$$\langle \Psi(\mathbf{k}_1)\Psi(\mathbf{k}_2)\Psi(\mathbf{k}_3)\rangle = (2\pi)^3\, \delta_D^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \Big[2 f_{\rm NL}^{\Psi}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) P_{\Psi}^{\rm lin}(k_1) P_{\Psi}^{\rm lin}(k_2) \tag{1.100}$$

$$+ (1 \to 2 \to 3) + (1 \to 3 \to 2) \Big]\,. \tag{1.101}$$

The most simple form for $f_{\rm NL}^{\Psi}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ is a constant, $f_{\rm NL}^{\rm local}$; but other shapes exist (equilateral, orthogonal, enfolded...) to probe different signatures of inflation. Latest measurements by Planck Collaboration et al. (2019) provide the most stringent constraints on non-Gaussianities, with $f_{\rm NL}^{\rm local} = -0.9 \pm 5.1$ at 68% confidence level. The best constraints on $f_{\rm NL}^{\rm local}$ from large-scale structure data are of the order of 35 at 95% confidence level (Slosar et al., 2008; Ross et al., 2013; Castorina et al., 2019). CMB measurements are intrinsically limited by cosmic variance to $\sigma(f_{\rm NL}^{\rm local}) \sim 5$, a precision that future large scale structure surveys may outperform.

1.3 Linear perturbations and baryon acoustic oscillations

This section is devoted to the linear growth of perturbations left by inflation in the primordial Universe. We first emphasise that density fluctuations are described by stochastic fields

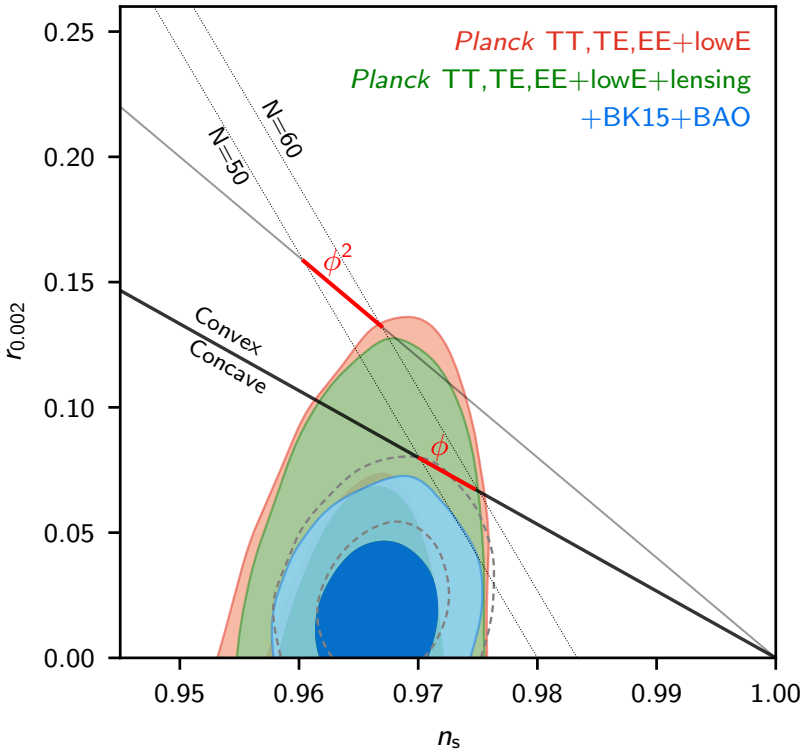

Figure 1.8 – Constraints in the plane $n_s - r_{0.002}$ from the Planck mission alone (red and green) and adding BICEP2/Keck (BICEP2 Collaboration et al., 2016) and BAO constraints in blue. Dashed lines are the contours obtained using `CamSpec` instead of `Plick` for the high-ℓ (small scales) Planck likelihood. The solid lines show the $n_s - r$ relation for V linear (separating concave and convex potentials) and quadratic in ϕ, to first order in slow-roll. The dotted lines correspond to $N = 50$ and $N = 60$ e-folds, assuming single-field inflation with a power law potential $V \sim \phi^p$. Taken from Planck Collaboration et al. (2018).

(Section 1.3.1), then review the evolution of perturbations in the radiation and matter dominated era, depending on their wavelength (Section 1.3.2), and explain how the coupling between radiation and baryons led to baryon acoustic oscillations (Section 1.3.3), visible in the cosmic microwave background (Section 1.3.4) and in the matter power spectrum (Section 1.3.5). This section is mainly based on Peter and Uzan (2012) textbook and Hu and Sugiyama (1996).

1.3.1 Stochastic fields

Owing to their quantum origin, fluctuations left by inflation cannot be predicted — nor measured. Therefore, fluctuations will be studied in a statistical sense, such that the evolution of their distribution can indeed be predicted — which is the aim of the next sections.

Let us for example consider the density contrast:

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}} \tag{1.102}$$

of the fluid of *background* density $\bar{\rho}$. In all the following, we will apply the ergodic principle, which states that volume-averaged quantities are equal to their expectation value, denoted by

$\langle\rangle$. Then, $\langle\delta\rangle$ is 0. This relates to the cosmological principle, if we suppose the large scale homogeneity is a result of statistical homogeneity and ergodicity.

One can also study the second moment of δ, i.e. its 2-point correlation function:

$$\xi(\mathbf{s}) = \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{s})\rangle \tag{1.103}$$

ξ is correctly defined (i.e. does not depend on $\mathbf{x}$) as we assume the Universe to be statistically homogeneous. If we further assume statistical isotropy, ξ depends only on s the norm of $\mathbf{s}$.

Let us define the Fourier transform of $\delta(\mathbf{x})$:

$$\delta(\mathbf{k}) = \int d^3x \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \tag{1.104}$$

which is the convention we will use throughout this manuscript. Its conjugate is $\delta(\mathbf{k})^\star = \delta(-\mathbf{k})$ since $\delta(\mathbf{x})$ is a real field. We have:

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle = \int d^3x \int d^3s \, \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{s})\rangle \, e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x} - i\mathbf{k}'\cdot\mathbf{s}}. \tag{1.105}$$

Following from statistical homogeneity:

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle = (2\pi)^3 \, \delta_D^{(3)}(\mathbf{k}+\mathbf{k}') P(\mathbf{k}) \tag{1.106}$$

where $\delta_D^{(3)}$ is the Dirac distribution and the power spectrum $P(\mathbf{k})$ is related to the correlation function by:

$$P(\mathbf{k}) = \int d^3s \xi(\mathbf{s}) e^{-i\mathbf{k}\cdot\mathbf{s}}. \tag{1.107}$$

Again, assuming isotropy, $P(\mathbf{k})$ depends only on k the norm of $\mathbf{k}$. If δ follows a Gaussian distribution (as is almost the case in the simplest models of inflation — single-field slow-roll — see Section 1.2.5), all information is contained in its power spectrum. In the following, we will tell the story of a small perturbation δ of cold dark matter and baryons, from the end of inflation to after the CMB emission. This will enable us to infer the evolution of its power spectrum.

1.3.2 Linear growth of perturbations

The purpose of the following section is to understand the overall shape of the matter power spectrum, which (as we will see in the following) can be achieved by considering two fluids, matter and radiation, interacting only through gravitation. The description of baryon acoustic oscillations (with a focus on the CMB), is deferred to Section 1.3.3.

As will be seen in the following, sub-Hubble fluid perturbations can be well described by Newtonian gravity. However, large scale perturbations must be described within general relativity, which requires perturbing the FLRW metric (1.23). In practice, up to the CMB emission, just like fluid perturbations, metric fluctuations are very small ($\mathcal{O}(10^{-5})$). Hence, the FLRW metric (1.23) is linearly perturbed following (Peter and Uzan, 2012):

$$ds^2 = a^2(\eta)\left[-(1+2A)\,d\eta^2 + 2B_i dx^i d\eta + (\gamma_{ij}+h_{ij})\,dx^i dx^j\right], \tag{1.108}$$

where we recall that η ($d\eta = dt/a$) is the conformal time and γ_{ij} is the spatial part of the FLRW metric (see Eq. (1.24)). Since the metric is a 4×4 symmetric tensor, there are $4\times(4+1)/2 = 10$ degrees of freedom:

— 1 for the scalar A;

— 3 for the vector B_i;

— 6 for the 3×3 symmetric tensor h_{ij}.

However, 4 of these degrees of freedom are identical to a choice of coordinates; this is known as gauge freedom. Here we choose the Newtonian gauge, where the scalar part of the metric is diagonal. We focus on scalar modes only; they evolve independently from vector modes (which can be shown to decay as a^{-2} in absence of stress vector modes) and tensor modes. Then, the perturbed metric can be written as:

$$ds^2 = a^2(\eta)\left[-(1+2\Phi)\,d\eta^2 + (1-2\Psi)\,\gamma_{ij}dx^i dx^j\right]. \tag{1.109}$$

where Φ and Ψ are the so-called Bardeen potentials in the Newtonian gauge. The scalar perturbations of the energy-momentum tensor of a fluid of background density $\bar{\rho}$ and pressure $\bar{P}$ (see Eq. (1.35)) can be written in the Newtonian gauge as:

$$\delta T_{00} = \bar{\rho}a^2\,(\delta + 2\Phi) \tag{1.110}$$

$$\delta T_{0i} = -\bar{\rho}a^2\,(1+w)\,D_i V \tag{1.111}$$

$$\delta T_{ij} = \bar{P}a^2\left(\frac{\delta P}{\bar{P}}\gamma_{ij} - 2\Psi\gamma_{ij} + \Delta_{ij}\bar{\pi}\right), \tag{1.112}$$

where D_i is the covariant derivative associated to γ_{ij}, $\Delta_{ij} = D_iD_j - 1/3\gamma_{ij}\Delta$ and $\Delta = D_iD^i$. δ and $\delta P/\bar{P}$ are the relative perturbations of $\rho = \bar{\rho} + \delta\bar{\rho}$ and $P = \bar{P} + \delta P$, respectively. D_iV and $\Delta_{ij}\bar{\pi}$ are the scalar parts of the fluid four-velocity and the anisotropic stress tensor, respectively. We also define the sound speed c_s and the entropy perturbation Γ as: $\delta P = c_s^2\delta\rho + \bar{P}\Gamma$. Then, Einstein equations (1.13) give:

$$(\Delta + 3K)\,\Psi = 4\pi G_{\mathrm{N}}a^2\bar{\rho}\,[\delta - 3\mathcal{H}\,(1+w)\,V] \tag{1.113}$$

$$\Psi - \Phi = 8\pi G_{\mathrm{N}}a^2\bar{P}\bar{\pi} \tag{1.114}$$

$$\Psi' + \mathcal{H}\Phi = -4\pi G_{\mathrm{N}}a^2\bar{\rho}\,(1+w)\,V \tag{1.115}$$

$$\begin{aligned}&\Psi'' + 3\mathcal{H}\left(1+c_s^2\right)\Psi' + \left[2\mathcal{H}' + \left(\mathcal{H}^2 - K\right)\left(1+3c_s^2\right)\right]\Psi - c_s^2\Delta\Psi \\ &= -\left(\mathcal{H}^2 + 2\mathcal{H}' + K\right)\left[\frac{\Gamma}{2} + \left(3\mathcal{H}^2 + 2\mathcal{H}'\right)\bar{\pi} + \mathcal{H}\bar{\pi}' + \frac{1}{3}\Delta\bar{\pi}\right] - 9c_s^2\mathcal{H}^2\left(\mathcal{H}^2 + K\right)\bar{\pi},\end{aligned} \tag{1.116}$$

where all derivatives $'$ are with respect to the conformal time and we use $\mathcal{H} = a'/a = aH$. We also define the density contrast in the comoving gauge $\delta^C = \delta - 3\mathcal{H}\,(1+w)\,V$, for which Eq. (1.113) looks like the Poisson equation.

Then, the conservation of the energy momentum tensor $\delta T^{\mu}_{\nu;\mu} = 0$ (see Eq. (1.16)) yields the following conservation equations:

$$\delta' + 3\mathcal{H}\left(c_s^2 - w\right)\delta = -\,(1+w)\,(\Delta V - 3\Psi') - 3\mathcal{H}w\Gamma \tag{1.117}$$

$$V' + \mathcal{H}\left(1 - 3c_s^2\right)V = -\Phi - \frac{c_s^2}{1+w}\delta - \frac{w}{1+w}\left[\Gamma + \frac{2}{3}\,(\Delta + 3K)\,\bar{\pi}\right]. \tag{1.118}$$

In the following we restrict to a flat Universe: the curvature term K is 0.

On would note that in principle, all species (photons, neutrinos, baryons, cold dark matter) should be considered to conclude on e.g. the dark matter perturbations. However, it has been shown by Weinberg (2002); Voruz et al. (2014) that up to the CMB fluid perturbations are

effectively the sum of a slow (of the order of the Hubble time $1/\mathcal{H}$) and a fast (of the order of $1/(kc_s)$) solution. The slow solution of dark matter perturbations dominates over the fast mode. On the contrary, the fast solutions of baryons, photons, and neutrinos dominate over the slow modes. We are primarily interested in the matter power spectrum. Since most of matter is in the form of dark matter, we first neglect baryons. Hence, we consider two fluids, radiation and (dark) matter, in the slow mode. The fast mode, involving baryons and photons (neglecting neutrinos) will be discussed in Section 1.3.3.

Let us introduce the scale factor normalised at equality (see Eq. (1.77)):

$$y = \frac{a}{a_{\rm eq}} \qquad \text{where} \qquad \Omega_{m,0} a_{\rm eq}^{-3} = \Omega_{r,0} a_{\rm eq}^{-4}. \tag{1.119}$$

Using this new variable, the first Friedmann equation (1.45) describing the background evolution is recast into:

$$H^2 = H_{\rm eq}^2 \frac{1+y}{2y^4} \qquad \text{with} \qquad H_{\rm eq} = H(a_{\rm eq}) = H_0 \sqrt{2\Omega_{m,0} a_{\rm eq}^{-3}} \tag{1.120}$$

$$\mathcal{H}^2 = \mathcal{H}_{\rm eq}^2 \frac{1+y}{2y^2} \qquad \text{with} \qquad \mathcal{H}_{\rm eq} = k_{\rm eq} = \mathcal{H}(a_{\rm eq}) = H_0 \sqrt{2\Omega_{m,0} a_{\rm eq}^{-1}}, \tag{1.121}$$

where we introduced $k_{\rm eq}$ the wavenumber corresponding to the Hubble radius at equality. The matter and radiation density parameters are thus given by (see Eq. (1.44)):

$$\Omega_m = \frac{y}{1+y} \qquad \Omega_r = \frac{1}{1+y}. \tag{1.122}$$

We assume radiation and matter interact only through gravity; therefore, the conservation equations are satisfied for each fluid. In addition, matter (subscript m) and radiation (subscript r) are assumed to be perfect fluids; $\Gamma_a = 0$, $\bar{\pi}_a = 0$, $a \in \{m, r\}$. Hence, the sound speed is $c_a^2 = \delta P_a / \delta\rho$ and using Eq. (1.114), $\Psi = \Phi$. At the time of interest, matter is non-relativistic, so $w_m = c_m^2 = 0$. The equation of state of radiation is constant $w_r = c_r^2 = 1/3$. Therefore, the conservation equations (1.117) and (1.118) give, in Fourier space and for each fluid:

$$\delta_m' = k^2 V_m + 3\Phi' \tag{1.123}$$

$$\delta_r' = \frac{4}{3} k^2 V_r + 4\Phi' \tag{1.124}$$

$$V_m' = -\mathcal{H} V_m - \Phi \tag{1.125}$$

$$V_r' = -\Phi - \frac{1}{4}\delta_r. \tag{1.126}$$

To close the system above, we need an additional equation for Φ, obtained from Einstein equations. Those apply to the mass weighted average of the two fluids: the total density contrast is $\Omega\delta = \Omega_m \delta_m + \Omega_r \delta_r$ and total four velocity is $\Omega\,(1+w)\,V = \Omega_m\,(1+w_m)\,\Omega_m V_m + \Omega_r\,(1+w_r)\,\Omega_r V_r$ (with $\Omega = \Omega_m + \Omega_r = 1$ and $\Omega w = \Omega_m w_m + \Omega_r w_r$). In the following, we will use Eq. (1.113) or (1.115), which give (using Eq. (1.36) to replace the background density $\bar{\rho}$):

$$-k^2\Phi = \frac{3}{2}\mathcal{H}^2 \left[\Omega_m \delta_m + \Omega_r \delta_r - 3\mathcal{H}\left(\Omega_m V_m + \frac{4}{3}\Omega_r V_r\right)\right] = \frac{3}{2}\mathcal{H}^2 \delta^C \tag{1.127}$$

$$\Phi' + \mathcal{H}\Phi = -\frac{3}{2}\mathcal{H}^2 \left(\Omega_m V_m + \frac{4}{3}\Omega_r V_r\right) = -\frac{3}{2}\mathcal{H}^2 \left(1 + \frac{1}{3}\Omega_r\right) V. \tag{1.128}$$

One can note that there is no mode coupling in the system (1.123) - (1.128): each Fourier mode k will evolve independently. Note also that all fields are a function of the wavevector $\mathbf{k}$, which we drop for conciseness in the following.

1.3.2.1 Initial conditions

To solve the above system of equations, one has to specify a set of initial conditions, which are a combination of:

— adiabatic initial conditions: the entropy perturbation $\Gamma = 1/(\Omega w)\sum_{a\in\{m,r\}}\left(c_s^2 - c_a^2\right)\Omega_a\delta_a$ (with $\Omega c_s^2 = \sum_a (1+w_a)/(1+w)\,\Omega_a c_a^2$) is set to zero. Hence, the relative density contrast $S = \delta_m/(1+w_m) - \delta_r/(1+w_r)$ is zero,

— isocurvature initial conditions: the gravitational potential $\Psi = \Phi$ is set to zero, and from the Poisson equation $\delta^C = \sum_{a\in\{m,r\}}\Omega_a\delta_a^C = 0$.

Because of the system linearity, any solution will be the combination of solutions obtained with either pure adiabatic or pure isocurvature initial conditions. Here we choose pure adiabatic initial conditions (as produced by the simplest models of inflation — single-field slow-roll). Then, we impose:

$$\Phi = \Phi_i \qquad \Phi' = 0 \qquad S = 0 \qquad S' = 0 \tag{1.129}$$

deep in the radiation era ($y = a/a_{\rm eq} \ll 1$), such that modes are super-Hubble ($k/\mathcal{H} \ll 1$). From Eq. (1.123) and Eq. (1.124) we have:

$$S' = k^2\left(V_m - V_r\right) = 0. \tag{1.130}$$

Using Eq. (1.128) one obtains:

$$V_i = -\frac{\Phi_i}{2\mathcal{H}_i} \qquad V_{m,i} = V_{r,i} = V_i. \tag{1.131}$$

With the Poisson equation (1.127) and since $\Omega_m = y/(1+y) \ll 1$ and $\Omega_r = 1/(1+y) \simeq 1$:

$$\delta_i^C = -\frac{2}{3}\frac{k^2}{\mathcal{H}_i^2}\Phi_i \qquad \delta_{r,i}^C = \delta_i^C \qquad \delta_{m,i}^C = \frac{3}{4}\delta_i^C, \tag{1.132}$$

$$\delta_i = -2\Phi_i \qquad \delta_{r,i} = \delta_i \qquad \delta_{m,i} = \frac{3}{4}\delta_i. \tag{1.133}$$

In addition, the spatial curvature perturbations $\mathcal{R}$ introduced in Section 1.2.5.3 read $\mathcal{R} = \Psi - \delta\rho/[3(\rho+P)]$ in the Newtonian gauge. Hence, still in the radiation era, the initial $\mathcal{R}$ is:

$$\mathcal{R}_i = \Phi_i - \frac{1}{4}\delta_{r,i} = \frac{3}{2}\Phi_i. \tag{1.134}$$

Hence, given Eq. (1.98) and within pure adiabatic initial conditions, the power spectrum of δ_i is proportional to k^{n_s-4} and the power spectrum of δ_i^C (or $\delta_{m,i}$, $\delta_{r,i}$) is proportional to k^{n_s}.

In the following, we follow the evolution of a perturbation of wavenumber k in this system of two fluids in purely gravitational interaction in an expanding Universe. This evolution will depend on whether k is super-Hubble $k \ll \mathcal{H}$ (Hubble friction term dominates) or sub-Hubble $k \gg \mathcal{H}$ (Hubble expansion can be neglected). As $\mathcal{H}$ increases with time, an initially super-Hubble mode will eventually become sub-Hubble. As we will see in the following, the evolution of sub-Hubble modes will depend on whether radiation or matter dominates the energy content of the Universe. Hence, the evolution of a mode k depends on the time it becomes sub-Hubble, i.e. on its size. We will first study the evolution of super-Hubble modes, then sub-Hubble modes in the radiation and matter dominated era.

1.3.2.2 Super-Hubble modes

Super-Hubble modes in the radiation era correspond to the initial conditions described above. Here we detail the evolution of modes remaining super-Hubble in the matter dominated era. In practice, we neglect k^2 terms compared to (conformal) time derivatives. Therefore Eq. (1.130) ensures that S remains constant, equal to its initial value $S = 0$. Hence:

$$\delta_m = \frac{3}{4}\delta_r. \tag{1.135}$$

Integrating Eq. (1.123) and Eq. (1.124) while neglecting the k^2 terms compared to the conformal time derivative yields:

$$\delta_m = 3\Phi - \frac{9}{2}\Phi_i \qquad \delta_r = 4\Phi - 6\Phi_i, \tag{1.136}$$

where initial conditions are set in the radiation dominated era by Eq. (1.133). Based on Eq. (1.116) and Eq. (1.121), one can show that, for $k \ll k_{\rm eq}$, the gravitational potential reads (Peter and Uzan, 2012):

$$\Phi = \frac{\Phi_i}{10y^3}\left[16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16\right]. \tag{1.137}$$

Hence, in the matter era ($y \gg 1$):

$$\Phi = \frac{9}{10}\Phi_i \qquad \delta_r = -\frac{12}{5}\Phi_i \qquad \delta_m = -\frac{9}{5}\Phi_i. \tag{1.138}$$

1.3.2.3 Sub-Hubble modes in the radiation-dominated era

For sub-Hubble modes in the radiation-dominated era the potential Φ is determined by the fluctuations δ_r, which can be considered as an external driving force. Inserting Eq. (1.123) and its derivative into Eq. (1.125), we find:

$$\delta_m'' + \mathcal{H}\delta_m' = 3\Phi'' + 3\mathcal{H}\Phi' - k^2\Phi. \tag{1.139}$$

Since in the radiation-dominated era $\eta = \mathcal{H}^{-1} \propto a$ (see Table 1.1), the homogeneous equation has two base solutions, $\delta_m \propto 1$ and $\delta_m \propto \ln a$. Using Green's method, the full solution reads:

$$\delta_m = A + B\ln a + \int_0^\eta du\,[\ln a(u) - \ln a(\eta)]\,\mathcal{H}^{-1}(u)\left(k^2\Phi - 3\Phi'' - 3\mathcal{H}\Phi'\right). \tag{1.140}$$

In the regime $y \ll 1$, we have determined that $\delta_m = -3\Phi_i/2$ (see Eq. (1.133)), such that $A = 3\Phi_i/2$, $B = 0$. Using numerical integration, the full solution reads (Hu and Sugiyama, 1996):

$$\delta_m(\mathbf{k}, a) = I_1\Phi_i(\mathbf{k})\ln\left(I_2\frac{a}{a_\mathcal{H}(k)}\right) \qquad a_\mathcal{H}(k) \ll a \ll a_{\rm eq}, \tag{1.141}$$

with $I_1 \simeq 9$ and $I_2 \simeq 0.6$. $a_\mathcal{H}(k) = y_\mathcal{H}(k)a_{\rm eq}$ is the scale factor at which mode k crosses the Hubble radius, that is: $k = \mathcal{H} = k_{\rm eq}\sqrt{(1+y_\mathcal{H})/2}/y_\mathcal{H}$ (see Eq. (1.121)), i.e.:

$$y_\mathcal{H} = \frac{a_\mathcal{H}}{a_{\rm eq}} = \frac{1+\sqrt{1+8\,(k/k_{\rm eq})^2}}{4\,(k/k_{\rm eq})^2}. \tag{1.142}$$

In conclusion, for a sub-Hubble mode in the radiation era:

$$\delta_m \propto \begin{cases} \ln a & k\eta \gg 1 \\ 1 & k\eta \ll 1. \end{cases} \tag{1.143}$$

1.3.2.4 Transition to the matter-dominated era

In this section we describe the transition of a sub-Hubble mode in the radiation-dominated era to the matter-dominated era. We have just seen that in the radiation-dominated era radiation perturbations remain roughly constant while small scale matter perturbations grew logarithmically as $\ln y$ (see Eq. (1.143)). In addition $\Omega_m/\Omega_r = y$ so the contribution of matter density fluctuations to the total density contrast $\delta = \Omega_m\delta_m + \Omega_r\delta_r$ eventually dominates, which is the regime we now consider. The total density contrast is then given by:

$$\delta \simeq \Omega_m\delta_m = \frac{y}{1+y}\delta_m. \tag{1.144}$$

We are interested in sub-Hubble modes $y \gg y_H$ i.e. $k/\mathcal{H} \gg 1$ and thus neglect velocity contributions in the Poisson equation (1.127), and more generally time derivatives with respect to spatial derivatives. In this regime, injecting the Poisson equation into Eq. (1.139):

$$\delta_m'' + \mathcal{H}\delta_m' - \frac{3}{2}\mathcal{H}^2\Omega_m\delta_m = 0. \tag{1.145}$$

Next, we note that $\mathcal{H}^2\Omega_m = k_{\rm eq}^2/(2y)$ (see Eq. (1.121) and Eq. (1.122)) and we trade derivatives with respect to η for derivatives with respect to y: $\delta_m' = y' d\delta_m/dy$. Using $\mathcal{H} = y'/y$ and $y'' = (y\mathcal{H})' = y\mathcal{H}^2 + yy' d\mathcal{H}/dy = \{y - (2+y)/[2(1+y)y]\}\mathcal{H}^2$ we find the Meszaros equation (Meszaros, 1974):

$$\frac{d^2\delta_m}{d^2y} + \frac{2+3y}{2y(y+1)}\frac{d\delta_m}{dy} - \frac{3}{2y(y+1)}\delta_m = 0. \tag{1.146}$$

The growing and decaying solutions to this equation are:

$$D_1(y) = y + \frac{2}{3} \qquad D_2(y) = \frac{15}{8}(2+3y)\ln\left(\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right) - \frac{45}{4}\sqrt{1+y}, \tag{1.147}$$

such that the full solution reads:

$$\delta_m(\mathbf{k}, y) = A_1(\mathbf{k})D_1(y) + A_2(\mathbf{k})D_2(y) \qquad y \gg y_H, \tag{1.148}$$

with A_1, A_2 constant in time. These constants can be determined by joining the solution above to that obtained deep in the radiation dominated era, Eq. (1.141), in the limit $y \ll 1$ (but still with $y \gg y_H$):

$$\delta_m(\mathbf{k}, y \ll 1) = I_1\Phi_i(\mathbf{k})\ln\left(I_2\frac{a_{\rm eq}}{a_{\mathcal{H}}}\right) + I_1\Phi_i(\mathbf{k})\ln y \tag{1.149}$$

$$= \frac{2}{3}A_1(\mathbf{k}) + \left(\frac{15}{4}\ln 4 - \frac{45}{4}\right)A_2(\mathbf{k}) - \frac{15}{4}A_2(\mathbf{k})\ln y. \tag{1.150}$$

Thus, by identification of A_1 and A_2, Eq. (1.148) reads:

$$\delta_m(\mathbf{k}, y) = I_1\Phi_i(\mathbf{k})\left[\frac{3}{2}\ln\left(4I_2e^{-3}\frac{a_{\rm eq}}{a_{\mathcal{H}}}\right)D_1(y) - \frac{4}{15}D_2(y)\right]. \tag{1.151}$$

In the matter-dominated era ($y \gg 1$) solutions (1.147) behave as:

$$D_1(a) \propto a \qquad D_2(a) \propto a^{-3/2}. \tag{1.152}$$

Hence, the logarithmic growth of sub-Hubble matter perturbations in the radiation-dominated era becomes linear in the matter domination.

1.3.2.5 Sub-Hubble modes in the late Universe

In this manuscript, we will focus on sub-Hubble modes in the matter or dark energy dominated era, where density fluctuations are described by Eq. (1.145). Usually, one trades derivatives with respect to η for derivatives in a, such that:

$$\frac{d^2\delta_m}{da^2} + \left(\frac{1}{H}\frac{dH}{da} + \frac{3}{a}\right)\frac{d\delta_m}{da} - \frac{3}{2}\frac{\Omega_{m,0}}{a^5}\left(\frac{H_0}{H}\right)^2\delta_m = 0 \tag{1.153}$$

where we used $\Omega_m = \Omega_{m,0}a^{-3}\left(H_0/H\right)^2$ (see Eq. (1.44)). The full solution is a linear combination of the growing and decaying modes:

$$\delta_m(\mathbf{k}, a) = A_+(\mathbf{k})D_+(a) + A_-(\mathbf{k})D_-(a) \tag{1.154}$$

with:

$$D_+(a) = \frac{5}{2}\frac{H(a)}{H_0^4}\Omega_{m,0}\int_0^a \frac{du}{\left(uH(u)\right)^3} \qquad D_-(a) = H(a). \tag{1.155}$$

Matter density perturbations grow, eventually becoming non-linear ($\delta_m \not\ll 1$). This late time evolution of matter density perturbations will be described in Chapter 2.

1.3.2.6 Transfer function

The imprint of the physics in the early Universe on the density perturbations δ (of any species — cold dark matter, radiation, baryons...) can be encoded in the transfer function, defined as:

$$\boxed{T(k) = \frac{\delta^C(\mathbf{k}, a=1)\delta_i^C(\mathbf{0})}{\delta_i^C(\mathbf{k})\delta^C(0, a=1)}} \tag{1.156}$$

where δ_i^C are initial conditions, given in Section 1.3.2.1. Just as evolution equations above, $T(k)$ only depends on the wavenumber k. By construction, $T(k \to 0) = 1$. Other definitions of the transfer function exist; in the Boltzmann code `CLASS` (Blas et al., 2011), initial conditions are not δ_i^C but the spatial curvature perturbations $\mathcal{R}_i$.

The global shape of the matter transfer function, displayed on Figure 1.9 can be inferred from the above study:

— for a mode $k \ll k_{\rm eq} = \mathcal{H}_{\rm eq}$, as discussed in Section 1.3.2.2 when k is super-Hubble Φ is constant (except the 9/10 step in the radiation to matter transition, see Eq. (1.138)), hence using the Poisson equation (1.127), $\delta_m^C \propto \mathcal{H}^{-2}$. After k enters the Hubble radius in the matter-dominated era, as shown in Section 1.3.2.4 and Section 1.3.2.5 it also grows as $a \propto \mathcal{H}^{-2}$ (using Table 1.1);

— for a mode $k \gg k_{\rm eq} = \mathcal{H}_{\rm eq}$, as seen previously $\delta_m^C \propto \mathcal{H}^{-2}$ while the mode remains super-Hubble. After k enters the Hubble radius (when $\mathcal{H} \equiv \mathcal{H}_k = k$) in the radiation-dominated era, δ^C remains constant (or logarithmically growing), till matter-radiation equality (when $\mathcal{H} = \mathcal{H}_{\rm eq}$), after which it grows again as $\mathcal{H}^{-2}$, as seen previously. During the time k is sub-Hubble in the radiation-dominated era, δ_m^C of super-Hubble modes grows by a factor $(\mathcal{H}_k/\mathcal{H}_{\rm eq})^2 = (k/k_{\rm eq})^2$. The missing growth factor for modes $k \gg k_{\rm eq}$ is thus $(k/k_{\rm eq})^2$.

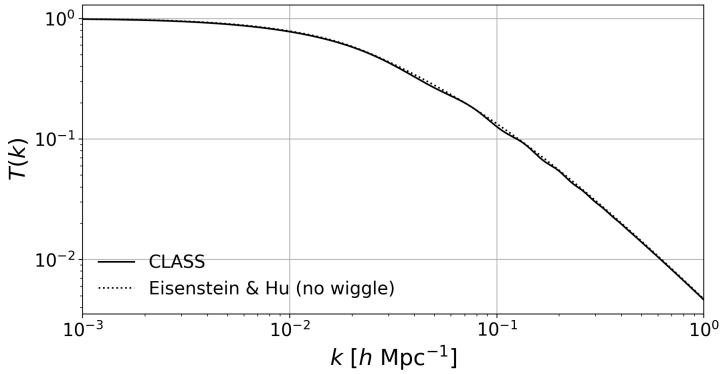

Figure 1.9 – Transfer function computed with the cosmological parameters of Planck Collaboration et al. (2018) (TT, TE, EE, lowE, lensing, BAO) using the Boltzmann code `CLASS` of Blas et al. (2011) (continuous line). The *no-wiggle* transfer function (dotted line) is based on the analytic formula of Eisenstein and Hu (1998).

Therefore, the overall shape of the matter transfer function is given by:

$$T_m(k) \sim \begin{cases} 1 & k \ll k_{\rm eq} \\ (k_{\rm eq}/k)^2 & k \gg k_{\rm eq} \end{cases} \tag{1.157}$$

The power spectrum of initial curvature perturbations $\mathcal{R}_i$ is set by inflation to be $\propto k^{n_s-4}$ (see Eq. (1.98)). As seen in Section 1.3.2.1, within adiabatic initial conditions, the power spectrum of $\delta^C_{m,i}$ goes as k^{n_s}. Therefore, the shape of the linear matter power spectrum at $a \gg a_{\rm eq}$ reads:

$$P_m^{\rm lin}(k) \propto \begin{cases} k^{n_s} & k \ll k_{\rm eq} \\ k^{n_s-4} & k \gg k_{\rm eq} \end{cases} \tag{1.158}$$

Usually in large scale structure analyses, instead of A_s, to fix the normalisation of the power spectrum we define the root mean square of density fluctuations in a sphere of $8\,{\rm Mpc}\,h^{-1}$, i.e.:

$$\sigma_8(z) = \sqrt{\frac{1}{2\pi^2}\int k^2 dk P_m^{\rm lin}(k,z) W^2(rk)} \qquad r = 8\,{\rm Mpc}\,h^{-1} \tag{1.159}$$

with $W(x) = 3\left(\sin x - x\cos x\right)/x^3$ the Fourier transform of the unit top-hat window. We note $\sigma_{8,0} = \sigma_8(z=0)$; see Table 1.2 for its value measured by Planck Collaboration et al. 2018.

In the following we will explain the small wiggles seen in Figure 1.9 around the smooth transfer function of Eisenstein and Hu (1998).

1.3.3 Propagation of baryon acoustic waves in the early Universe

The system (1.123) - (1.126) and (1.127) - (1.128) is correct for a dark matter fluid, but is an approximation for baryons and photons. First, Thomson diffusion couples baryons and photons. Second, the sound speed of baryons is non-zero and varies with temperature (and slightly with time). In the following we still assume baryons have zero sound speed.

To account for Thomson diffusion, let us introduce the interaction interaction term $\tau'/R\,(V_\gamma - V_b)$ in Eq. (1.123) and Eq. (1.125):

$$\delta_b' = k^2 V_b + 3\Phi' \tag{1.160}$$

$$V_b' = -\mathcal{H}V_b - \Phi + \frac{\tau'}{R}\,(V_\gamma - V_b) \tag{1.161}$$

where we introduced $\tau' = a n_e \sigma_T$ the inverse of the photon mean free path (n_e the free electron density, σ_T the Thomson cross-section) and $R(\eta) = 3\Omega_b(\eta)/\left(4\Omega_\gamma(\eta)\right)$ is the baryon-to-photon energy density ratio. A correct modelling of radiation requires a kinetic approach, using a so-called Boltzmann hierarchy: the phase-space distribution function is perturbed and its dependence on the cosine angle between the wavenumber $\mathbf{k}$ and the fluid momentum is expanded onto the basis of Legendre polynomials. Here, however, we keep the same formalism and introduce the baryon coupling term $\tau'\,(V_b - V_\gamma)$:

$$\delta_\gamma' = \frac{4}{3}k^2 V_\gamma + 4\Phi' \tag{1.162}$$

$$V_\gamma' = -\Phi - \frac{1}{4}\delta_\gamma + \tau'\,(V_b - V_\gamma)\,. \tag{1.163}$$

This approach is equivalent to the first two multipoles of the Boltzmann hierarchy ($\theta_0 = \delta_\gamma/4$, $\theta_1 = kV_\gamma$).

Let us consider scales smaller than the Hubble radius $k \gg \mathcal{H}$, but larger than the mean free path, $k \ll \tau'$. In this so-called *tight coupling* regime, V_b and V_γ are (almost) similar. At zeroth order, $V_b - V_\gamma = 0$. Injecting this estimate into the baryon Euler equation (1.161) we find $\tau'\,(V_\gamma - V_b) = \left(V_\gamma' + \mathcal{H}V_\gamma + \Phi\right) R$ at first order. Using this expression for Eq. (1.163), and noting that $R' = \mathcal{H}R$ (since $R \propto a^3/a^4 = a^{-1}$, see Table 1.1), we have:

$$(1+R)\,V_\gamma' + R'V_\gamma = -\frac{1}{4}\delta_\gamma - (1+R)\,\Phi. \tag{1.164}$$

Replacing V_γ and V_γ' using Eq. (1.162) one finds:

$$\delta_\gamma'' + \frac{R'}{1+R}\delta_\gamma' + k^2 c_s^2 \delta_\gamma = 4\left[\Phi'' + \frac{R'}{1+R}\Phi' - \frac{1}{3}k^2\Phi\right], \tag{1.165}$$

where we used the sound speed of the baryon and photon fluid $c_s = 1/\sqrt{3\,(1+R)}$. Let us introduce the proper Sachs-Wolfe term $\theta_{\rm SW} = \delta_\gamma/4 + \Phi$, a main ingredient of the temperature fluctuations of the CMB which will be explained in the following. In the limit where R and Φ are constant in time, one obtains:

$$(1+R)\,\theta_{\rm SW}'' + \frac{k^2}{3}\theta_{\rm SW} = -\frac{k^2}{3}R\Phi, \tag{1.166}$$

whose solution reads:

$$\theta_{\rm SW} = C_A \cos\left(kc_s\eta\right) + C_I \sin\left(kc_s\eta\right) - R\Phi. \tag{1.167}$$

This can be seen as an oscillator with effective mass $m_{\rm eff} = 1 + R$. The zero shift $-R\Phi$ is due to photons being dragged into potential wells created by baryons.

Note however that the variations of the effective mass $m_{\rm eff} = 1+R$ or the potential Φ cannot be ignored over several periods of oscillations. Exploiting the fact that variations on one period

are small ($\mathcal{H} = R'/R \ll k$), one can show that the above solution becomes (Peter and Uzan, 2012):

$$\theta_{\rm SW} = C_A \frac{\cos(kr_s)}{(1+R)^{1/4}} + C_I \frac{\sin(kr_s)}{(1+R)^{1/4}} - R\Phi \tag{1.168}$$

with the sound horizon:

$$r_s(\eta) = \int_0^\eta c_s(u)du = \frac{2}{3k_{\rm eq}}\sqrt{\frac{6}{R_{\rm eq}}} \ln\left[\frac{\sqrt{1+R(\eta)}+\sqrt{R(\eta)+R_{\rm eq}}}{1+\sqrt{R_{\rm eq}}}\right], \tag{1.169}$$

$R_{\rm eq} = R(\eta_{\rm eq})$ being the value of R at equality.

Therefore, acoustic waves propagate in the fluid of baryons and photons at the variable sound speed c_s. These are called baryon acoustic oscillations (BAO).

In addition, the tight coupling described above is not perfect, leading to the so-called Silk damping, which tends to smooth out the density fluctuations on scales $k > k_{\rm D}$, such that:

$$\theta_{\rm SW} = e^{-k^2/k_{\rm D}^2}\left[C_A \frac{\cos(kr_s)}{(1+R)^{1/4}} + C_I \frac{\sin(kr_s)}{(1+R)^{1/4}}\right] - R\Phi \tag{1.170}$$

with:

$$k_D^{-2} = \frac{1}{6}\int_0^\eta \frac{du}{\tau'}\left[\frac{1}{1+R}\left(\frac{4}{5}+\frac{R^2}{1+R}\right)\right] \simeq \frac{1}{6}\frac{\eta}{\tau'}, \tag{1.171}$$

where the last approximation is legitimate since the term in brackets remains very close to 1 (4/5 for $R = 0$ and 1 for $R \to \infty$). Taking into account the anisotropic stress tensor of radiation which we neglected so far, the 4/5 term in Eq. (1.171) must be replaced by 16/15. $1/k_{\rm D}(\eta)$ may be estimated by the length of the typical photon random walk of $N \sim \eta\tau'$ steps, $1/k_{\rm D}(\eta) \sim \sqrt{N}/\tau' \sim \sqrt{\eta/\tau'}$. Diffusion damping leaves the term $-R\Phi$ unaffected.

Let us now specify C_A and C_I. Imposing adiabatic initial conditions, $\delta_{\gamma,i} = -2\Phi_i$ (see Eq. (1.133)). Neglecting V_γ in Eq. (1.162), $\delta'_\gamma = 4\Phi'$, thus $\delta_\gamma = 4\Phi - 6\Phi_i$. Here we focus on small scales, that enter the Hubble radius well in the radiation dominated era; for those, when k enters the sound horizon, $\Phi \simeq 0$ and then $\delta_\gamma = -6\Phi_i$, $\delta'_\gamma \simeq 0$. Therefore:

$$C_A = -\frac{3}{2}\Phi_i \qquad C_I = 0 \tag{1.172}$$

so:

$$\boxed{\theta_{\rm SW} = -\frac{3}{2}\Phi_i e^{-k^2/k_{\rm D}^2}\cos(kr_s(\eta))} \tag{1.173}$$

Had we chosen isocurvature initial conditions instead, the sine term would have been excited.

For super-Hubble modes in the matter-dominated era, using Eq. (1.138), we have $\theta_{\rm SW} = 1/4(-12/5)\Phi_i + 9/10\Phi_i = 3/10\Phi_i$. In terms of $\Phi = 9/10\Phi_i$ the potential at the last scattering surface, we find:

$$\boxed{\theta_{\rm SW} = \frac{1}{3}\Phi} \tag{1.174}$$

i.e. $\theta_{\rm SW}$ shows a plateau on large scales. For isocurvature initial conditions, we have instead $\theta_{\rm SW} = 2\Phi$.

As mentioned in Section 1.2.3.1, after recombination, the free electron density and thus the Thomson scattering rate τ' drop rapidly. Once decoupled from the primordial plasma, photons free stream and form the cosmic microwave background. The above results will be used to qualitatively describe the shape of the temperature fluctuations of the CMB in the next section.

1.3.4 The cosmic microwave background

In this section we comment on the CMB temperature power spectrum measured by Planck Collaboration et al. (2018) and shown in Figure 1.11. We start by introducing the angular power spectrum. Then, we explain how density, velocity and potential perturbations are related to the observed temperature fluctuations and qualitatively discuss the shape of their power spectrum.

1.3.4.1 The angular power spectrum

Today's photon temperature fluctuations θ are analysed through their correlation function:

$$C(\vartheta) = \langle \theta(\mathbf{e}_1)\theta(\mathbf{e}_2)\rangle \tag{1.175}$$

where we recall the brackets denote an ensemble average. From isotropy, the correlation function only depends on the cosine angle $\cos(\vartheta)$ between the two lines of sight $\mathbf{e}_1$ and $\mathbf{e}_2$, and thus can be expanded onto the basis of Legendre polynomials $\mathcal{L}_\ell$:

$$C(\vartheta) = \langle \theta(\mathbf{e}_1)\theta(\mathbf{e}_2)\rangle = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell \mathcal{L}_\ell(\mathbf{e}_1\cdot\mathbf{e}_2). \tag{1.176}$$

This defines the angular power spectrum C_ℓ. θ can be developed onto the basis of spherical harmonics $Y_{\ell m}(\mathbf{e})$:

$$\theta(\mathbf{e}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\mathbf{e}) \tag{1.177}$$

which is inverted into:

$$a_{\ell m} = \int d\mathbf{e}\, \theta(\mathbf{e}) Y^\star_{\ell m}(\mathbf{e}). \tag{1.178}$$

The temperature fluctuations are zero in average:

$$\langle a_{\ell m}\rangle = 0. \tag{1.179}$$

Inserting the expansion:

$$\frac{2\ell+1}{4\pi}\mathcal{L}_\ell(\mathbf{e}_1\cdot\mathbf{e}_2) = \sum_{m=-\ell}^{\ell} Y_{\ell m}(\mathbf{e}_1) Y^\star_{\ell m}(\mathbf{e}_2) \tag{1.180}$$

into Eq. (1.176), and using twice the orthonormality of spherical harmonics yields:

$$\int d\mathbf{e}_1 \int d\mathbf{e}_2 \,\langle \theta(\mathbf{e}_1)\theta(\mathbf{e}_2)\rangle\, Y^\star_{\ell m}(\mathbf{e}_1) Y_{\ell' m'}(\mathbf{e}_2) = \langle a_{\ell m} a^\star_{\ell' m'}\rangle = C_\ell \delta_{\ell\ell'}\delta_{mm'}. \tag{1.181}$$

A natural estimator for C_ℓ is thus:

$$\hat{C}_\ell = \frac{1}{2\ell+1}\sum_{m=-\ell}^{\ell} |a_{\ell m}|^2 . \tag{1.182}$$

In the simplest models of inflation (single-field slow-roll, see Section 1.2.5), primordial potentials follow (almost) Gaussian statistics and so do the perturbations in the linear regime. Therefore, $a_{\ell m}$ are Gaussian distributed with variance C_ℓ, which thus fully describes the statistics of the temperature fluctuations. Then, if Eq. (1.182) is taken as estimator for C_ℓ, $(2\ell+1)\,\hat{C}_\ell/C_\ell$ follows a χ^2 distribution with $2\ell+1$ degrees of freedom. In the high ℓ limit, by virtue of the central limit theorem, $\hat{C}_\ell$ is Gaussian distributed. Its variance is:

$$\left\langle \left(\hat{C}_\ell - \left\langle \hat{C}_\ell\right\rangle\right)^2 \right\rangle = \frac{2}{2\ell+1} C_\ell^2, \tag{1.183}$$

as there are only $2\ell+1$ independent modes. This lower bound on the variance (independent of any experiment) is called *cosmic variance*, and limits the measurement precision on large scales.

1.3.4.2 The Sachs-Wolfe formula

Temperature fluctuations θ can be related to those at the decoupling epoch by the Sachs-Wolfe formula:

$$\theta(\mathbf{x}_0, \eta_0, \eta_e) = \theta_{\rm SW}(\mathbf{x}_e, \eta_e) + \theta_{\rm Doppler}(\mathbf{x}_e, \eta_e) + \theta_{\rm ISW}(\mathbf{x}_e, \eta_e, \eta_0) \tag{1.184}$$

$$\theta_{\rm SW}(\mathbf{x}_e, \eta_e) = \left(\frac{1}{4}\delta_\gamma + \Phi\right)(\mathbf{x}_e, \eta_e) \tag{1.185}$$

$$\theta_{\rm Doppler}(\mathbf{x}_e, \eta_e) = -e^i \left(D_i V_b + \bar{V}_{bi} + \bar{E}_i\right)(\mathbf{x}_e, \eta_e) \tag{1.186}$$

$$\theta_{\rm ISW}(\mathbf{x}_e, \eta_e, \eta_0) = \int_{\eta_e}^{\eta_0} \left[(\Phi' + \Psi') - e^i \bar{E}_i' - e^i e^j \bar{E}_{ij}'\right] \tag{1.187}$$

where η_e is the (average) conformal time at decoupling and $\mathbf{x}_e = \mathbf{x}_0 + \mathbf{e}\,(\eta_0 - \eta_e)$ with $\mathbf{x}_0$ and η_0 the observer's position and conformal time, $\mathbf{e}$ the line of sight. $\bar{V}_{bi}$ are vector perturbations to the baryon four-velocity, while $\bar{E}_i$ and $\bar{E}_{ij}$ are vector and tensor perturbations in h_{ij} (see Eq. (1.108)).

Let us discuss the different pieces of this formula.

Proper Sachs-Wolfe effect The $\theta_{\rm SW}$ term is called proper Sachs-Wolfe effect, and was derived (approximately) in Section 1.3.3 for modes $\mathcal{H} \ll k \ll \tau'$. The density term (coming from Stefan-Boltzmann law) means that denser regions are hotter. The potential term Φ accounts for the Einstein effect: a photon in a potential well is redshifted. For adiabatic initial conditions $\theta_{\rm SW}$ is maximum for (see Eq. (1.173)):

$$k_{(p)} = p\frac{\pi}{r_s(z_\star)}, \qquad p \in \mathbb{N}^\star \tag{1.188}$$

where r_s is given by Eq. (1.169) and we recall $z_\star$ is the redshift of the last scattering surface (see Section 1.2.3.1). This corresponds to an angular scale:

$$\gamma_{(p)} \simeq \frac{\pi}{k_{(p)} D_{\rm M}(z_\star)} = \frac{r_s(z_\star)}{p D_{\rm M}(z_\star)}, \tag{1.189}$$

therefore, to a multipole:

$$\boxed{\ell_{(p)} \simeq k_{(p)} D_{\rm M}(z_\star) = p\pi \frac{D_{\rm M}(z_\star)}{r_s(z_\star)}} \tag{1.190}$$

For $p = 1$, we find $\gamma \simeq 0.01\,$rad, $\ell \simeq 300$ in the flat ΛCDM model with parameters of Table 1.2. With isocurvature initial conditions, one should replace p by $p - 1/2$ in the formula above. The zero shift $-R\Phi$ in Eq. (1.167) is responsible for the even BAO peaks being higher than the odd ones.

On large scales, based on Eq. (1.174) one can show that $\ell(\ell+1)C_\ell \sim \ell^{n_s - 1}$.

Doppler Sachs-Wolfe effect The $\theta_{\rm Doppler}$ term corresponds to the Doppler effect, due to the emitter's (last scattering) velocity with respect to the observer. This term tends to counterbalance the BAO induced by the proper Sachs-Wolfe effect. Indeed, from Eq. (1.162), in the limit Φ is constant:

$$V_\gamma = \frac{3}{k^2}\theta_{\rm SW}'. \tag{1.191}$$

Hence, V_b, which is tied to V_γ in the strong coupling limit, is in quadrature with $\theta_{\rm SW}$, and thus tends to smooth the density fluctuations. This effect is smaller at high baryon density (large R), since $V_\gamma \propto c_s \propto (1+R)^{-1/2}$.

Integrated Sachs-Wolfe effect The integrated Sachs-Wolfe effect $\theta_{\rm ISW}$ has two contributions, the early and late Sachs-Wolfe effects. The early Sachs-Wolfe effect is due to the change in Bardeen potentials at the radiation to matter era transition. It tends to increase the amplitude of the power spectrum around the first acoustic peak. The late Sachs-Wolfe effect is caused by curvature and early dark energy. Both tend to increase the Sachs-Wolfe plateau, at scales larger than the first peak ($\ell \lesssim 30$).

1.3.4.3 Width of the last scattering surface

Decoupling is not instantaneous. Let us define the visibility function, $g(\eta) = \tau' \exp(-\tau(\eta))$, where $\tau(\eta) = \int_\eta^{\eta_0} \tau'(u)du$ (with η_0 the present conformal time) is the optical depth. $g(\eta)d\eta$ is the probability for a photon to last scatter between η and $\eta + d\eta$. The redshift of the last scattering surface $z_\star$ is defined as the redshift at which the photon optical depth equals unity, i.e. $\tau(z_\star) = 1$ (see Table 1.2 for its value measured by Planck Collaboration et al. 2018). The full width at half maximum of the visibility function is of order $\Delta z \sim 200$.

Therefore, the temperature fluctuations are convolved with the visibility function, which can be roughly modelled as a Gaussian centred on the last scattering surface. Adding the Silk damping leads to the acoustic visibility function $\hat{g} = ge^{-k^2/k_D^2}$, which best describes how oscillations are frozen into the CMB. The Silk damping actually shifts the effective visibility function to slightly higher redshift, such that most of the temperature fluctuations in the CMB come from an epoch where the tight coupling regime is legitimate (Hu and Sugiyama, 1996).

1.3.4.4 Reionisation

First stars emit UV photons which ionise the neutral atoms formed by recombination at a redshift $z_{\rm rei} \simeq 7-8$ (see Table 1.2). Therefore, a fraction $1 - e^{-\tau_{\rm rei}}$ (with $\tau_{\rm rei} = \tau(\eta_{\rm rei})$) of CMB photons scatter on the free electrons, which averages the temperature fluctuations along different lines of sight. The main effect is a damping of scales that are sub-Hubble at the time of reionisation ($\ell \gtrsim 10$), by a factor $e^{-2\tau_{\rm rei}}$.

1.3.4.5 Sunyaev-Zel'dovich effect

Clusters contain hot gas at temperature $\sim 10^7 - 10^8\,$K. The highly energetic electrons can increase the CMB photon energy through inverse Compton scattering. The thermal Sunyaev-Zel'dovich effect therefore redistributes the spectral energy distribution from the large to the small wavelengths, which makes it a powerful tool to detect clusters. If the gas moves with respect to the CMB, the induced Doppler effect is dubbed kinetic Sunyaev-Zel'dovich.

1.3.4.6 Sensitivity to cosmological parameters

Variations of cosmological parameters impact the position and the relative amplitude of the BAO peaks of the CMB temperature power spectrum. When the baryon density $\omega_{b,0} = \Omega_{b,0}h^2$ increases, the amplitude of the oscillations increases, and c_s decreases, such that the peak positions are shifted to higher multipoles. When the matter density $\omega_{m,0}$ increases (at fixed

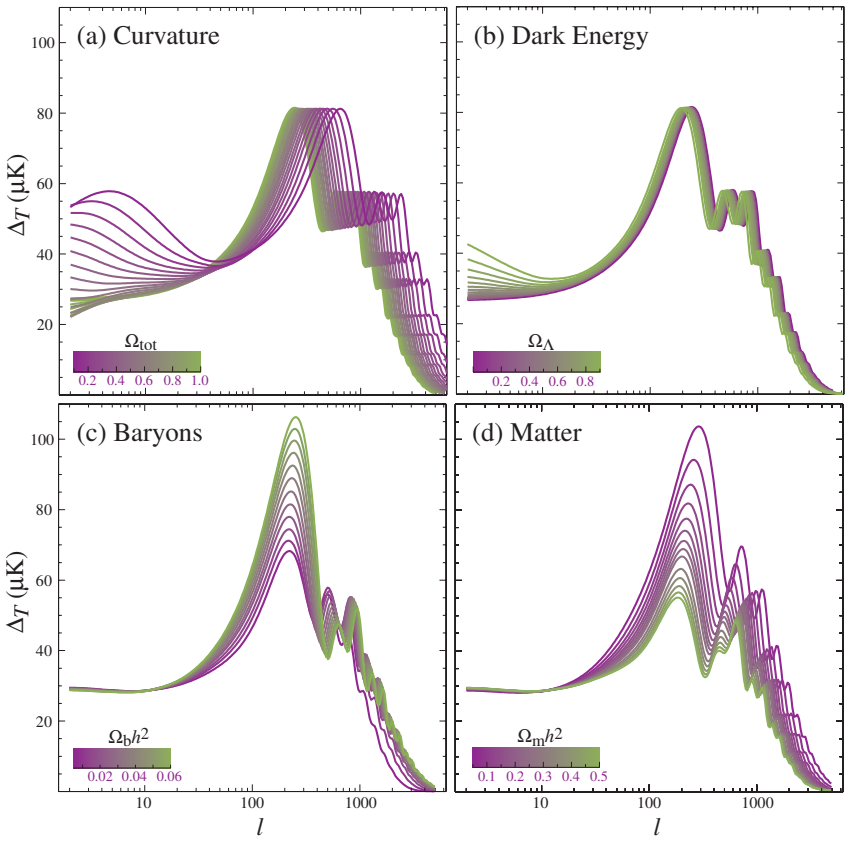

Figure 1.10 – Impact on density parameters (top left: $1 - \Omega_{K,0}$, top right: $\Omega_{\Lambda,0}$, bottom left: $\omega_{b,0}$, bottom right: $\omega_{m,0}$) on the (square root) of power spectrum of CMB temperature anisotropies $\Delta_T = T\sqrt{\ell\,(\ell+1)\,C_\ell/\,(2\pi)}$. Taken from Hu and Dodelson (2002).

baryon density), the overall amplitude of the peaks decreases, due to reduced radiation driving, while the relative heights of the even and odd peaks increases. The curvature and cosmological constant both increase the Sachs-Wolfe plateau ($\ell \lesssim 30$) and change the position of the acoustic peaks due to the distance - redshift relation $D_{\rm M}(z) = f_K(\chi(z))$ (see Eq. (1.190) and Eq. (1.29)).

The parameters of the primordial fluctuations also impact the shape of the CMB power spectrum, as can be seen in Figure 1.10. A higher scalar index n_s increases the small scales relative to the large scales, and so does the tensor index n_T. Gravitational waves (if large enough) increase the Sachs-Wolfe plateau compared to the (purely scalar) oscillations. Finally, the initial conditions (adiabatic or isocurvature) impact both the position of the acoustic peaks and the normalisation of the power spectrum.

The CMB temperature power spectrum, measured by Planck Collaboration et al. (2018) is shown in Figure 1.11. In addition to the temperature fluctuations, Planck Collaboration et al. (2018) measured the CMB polarisation projected on the E and B modes. In the CMB, E modes are produced by scalar and tensor modes, while B modes are only caused by tensor modes, i.e. gravitational waves. Primordial B modes, if detected, would give major insights into inflation. Then, three power spectra are formed, TT, TE, EE. In addition, structures lens the CMB backlight, an effect which was detected at 40σ by Planck Collaboration et al. (2018). The CMB lensing power spectrum is sensitive to the background evolution in the more recent $z < 2$ epoch, which makes it a powerful probe for dark energy.

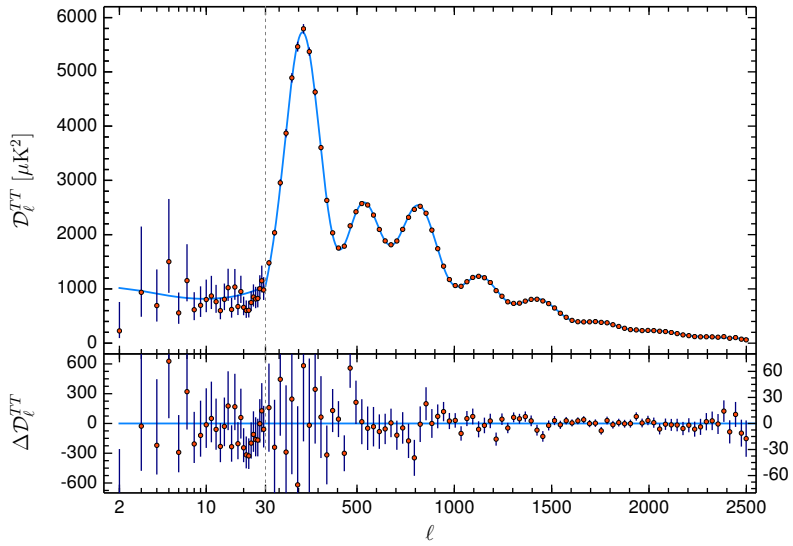

Figure 1.11 – The CMB temperature power spectrum. The horizontal axis switches from logarithmic to linear scale at $\ell = 30$. Foreground parameters are fixed to their best fit values to position the data measurements. The plotted best fit model (blue curve) uses TT, TE, EE, lowE and lensing data. Taken from Planck Collaboration et al. (2018).

1.3.5 Baryon acoustic oscillations in the matter power spectrum

Let us now describe how baryon acoustic oscillations impact the power spectrum of baryons and dark matter. Contrary to the case of the CMB where BAO manifest as fluctuations in the photon temperature, BAO are imprinted in the matter distribution by potential wells formed by baryons, which we will first discuss. Then, we will describe dark matter fluctuations and connect them to baryons to model the total matter transfer function.

1.3.5.1 Decoupling of baryons

In the tight coupling limit discussed so far, baryons oscillate with photons. Compton scattering leads to an equilibrium between photon and baryon velocities through an exchange of momentum. However, the baryon momentum is $\rho_b V_b$, while the photon momentum is $4/3\rho_\gamma V_\gamma$. Thus, by virtue of momentum conservation, the rate of change in baryon velocity due to Compton drag is scaled by $R^{-1} = 4/3\rho_\gamma/\rho_b$ compared to the photon case. Hence, baryons see the inverse mean free path $\tau_d' = \tau'/R$. This can be shown more formally by integrating the Euler equation (1.161):

$$V_b = \frac{1}{a}\int_0^\eta du\,(\tau_d' V_\gamma - \Phi)\,a(u)e^{-\tau_d(u)}, \tag{1.192}$$

with $\tau_d(u) = \int_u^\eta dv\tau_d'(v)$, and $\tau_d' = \tau'/R$ as found previously. The $1/a$ front factor just accounts for the natural V_b scaling as a^{-1} in the absence of source terms.

This expression is very similar to the previously discussed convolution of the photon temperature with the visibility function due the width of the last scattering surface (see Section 1.3.4.3). Here, the baryon drag visibility function is $g_b \propto a\tau_d' e^{-\tau_d}$ (normalised such that its integral is 1). Similarly to the redshift of the last scattering surface $z_\star$ (see Section 1.3.4.3), the drag redshift $z_{\rm drag}$ is defined as the redshift at which the baryon drag optical depth equals unity, i.e. $\tau_d(z_{\rm drag}) = 1$. We note $r_{\rm drag} = r_s(z_{\rm drag})$ the value of the sound horizon at the drag epoch and

report the reader to Table 1.2 for the values of $z_{\rm drag}$ and $r_{\rm drag}$ measured by Planck Collaboration et al. (2018).

In the tight coupling limit, Φ is neglected compared to $\tau_d' V_\gamma$, V_γ is given by Eq. (1.191), and $\delta_b = 3/4\delta_\gamma$. Then, at the drag epoch, we have approximately:

$$V_b(\mathbf{k}, \eta_d) = V_\gamma(\mathbf{k}, \eta_d) D_b(k) \tag{1.193}$$

$$\delta_b(\mathbf{k}, \eta_d) = \frac{3}{4}\delta_\gamma(\mathbf{k}, \eta_d) D_b(k) \tag{1.194}$$

with:

$$D_b(k) = \int_0^{\eta_0} d\eta g_b(\eta) e^{-(k/k_D(\eta))^2} \simeq e^{-(k/k_{\rm Silk})^{m_{\rm Silk}}} \tag{1.195}$$

where $k_{\rm Silk}$ is the Silk damping length and $m_{\rm Silk}$ the Silk damping steepness. As we will see in the following, at the drag epoch the baryon velocity $V_b(\mathbf{k}, \eta_d)$ dominates over the acoustic density $\delta_b(\mathbf{k}, \eta_d)$; therefore, baryons move and form new perturbations, roughly 1/2 out of phase with respect to the CMB (see Eq. (1.188)) in the small scale ($kr_{\rm drag} \gg 1$) regime. For adiabatic initial conditions (Hu and Sugiyama, 1996):

$$\boxed{k_{(p)} = \left(p - \frac{1}{2}\right)\frac{\pi}{r_{\rm drag}}, \qquad p \in \mathbb{N}^\star} \tag{1.196}$$

This so-called velocity overshoot effect is in fact less prominent on large scales $kr_{\rm drag} \lesssim 10$, thus inducing a phase shift with respect to the above solution (Eisenstein and Hu, 1998).

1.3.5.2 Dark matter fluctuations

In the absence of baryons, dark matter fluctuations are described by Eq. (1.148). Introducing baryons as an additional matter component (uncoupled to dark matter):

$$\frac{d^2\delta_c}{d^2 y} + \frac{2+3y}{2y\,(y+1)}\frac{d\delta_c}{dy} - \frac{3}{2y\,(y+1)}\frac{\Omega_c}{\Omega_m}\delta_c = 0 \tag{1.197}$$

where indices c and m refer to cold dark matter and total matter, respectively.

The two solutions to this equation ($i \in \{1, 2\}$) are expressed with Gauss hypergeometric function F through:

$$U_i = (1+y)^{-\alpha_i} F\left(\alpha_i, \alpha_i + \frac{1}{2}, 2\alpha_i + \frac{1}{2}, \frac{1}{1+y}\right) \qquad \alpha_i = \frac{1 + (-1)^i\sqrt{1 + 24\Omega_c/\Omega_m}}{4}. \tag{1.198}$$

In the limit $y \ll 1$, $U_i(y \ll 1) = y^{-\alpha_i}$. Therefore, baryons tend to slow the growth of dark matter perturbations. One can equally match the obtained general solution

$$\delta_c(\mathbf{k}, \eta) = I_1 \Phi_i(\mathbf{k})\left[A_1(k)U_1(\eta) + A_2(k)U_2(\eta)\right] \tag{1.199}$$

to that obtained in the radiation dominated era, Eq. (1.141) in the limit $y \ll 1$, thereby determining A_1 and A_2:

$$\begin{aligned} A_1(k) = &\frac{\Gamma(\alpha_1)\Gamma(\alpha_1 + 1/2)}{\Gamma(2\alpha_1 + 1/2)\left[d\Gamma(\alpha_1 + 1/2) + d\Gamma(\alpha_1) - d\Gamma(\alpha_2 + 1/2) - d\Gamma(\alpha_2)\right]} \\ &\times \left[\ln\left(I_2 \frac{a_{\rm eq}}{a_{\mathcal{H}}(k)}\right) + 2d\Gamma(1) - d\Gamma(\alpha_2) - d\Gamma(\alpha_2 + 1/2)\right]. \end{aligned} \tag{1.200}$$

and similarly ($1 \leftrightarrow 2$) for A_2. Here, $\Gamma(x)$ and $d\Gamma(x) = \Gamma'(x)/\Gamma(x)$ are gamma and digamma functions, respectively.

1.3.5.3 Total matter transfer function

Now that we have a full description of both baryon and dark matter fluctuations up to the drag epoch, we can describe their evolution through the present, in the linear regime, on sub-Hubble modes.

The combined fluid:

$$\begin{aligned}\delta_m &= \frac{\Omega_{b,0}}{\Omega_{m,0}}\delta_b + \left(1 - \frac{\Omega_{b,0}}{\Omega_{m,0}}\right)\delta_c \\ V_m &= \frac{\Omega_{b,0}}{\Omega_{m,0}}V_b + \left(1 - \frac{\Omega_{b,0}}{\Omega_{m,0}}\right)V_c\end{aligned} \tag{1.201}$$

follows the solution (1.148), and, in the matter or dark energy dominated era Eq. (1.154), i.e.:

$$\delta_m(\mathbf{k}, \eta) = C_1(\mathbf{k})D_1(\eta) + C_2(\mathbf{k})D_2(\eta). \tag{1.202}$$

C_1 and C_2 can be determined by matching δ_m and $\delta'_m = C_1 D'_1 + C_2 D'_2$ to their values at the drag epoch ($\delta'_m(\mathbf{k}, \eta_d) \simeq k^2 V_m(\mathbf{k}, \eta_d)$). We find, for the growing mode:

$$C_1(\mathbf{k}) = G_\delta(\eta_d)\delta_m(\mathbf{k}, \eta_d) + G_V(\eta_d)k^2 V_m(\mathbf{k}, \eta_d), \tag{1.203}$$

with:

$$G_\delta = \frac{D'_2}{D_1 D'_2 - D'_1 D_2} \qquad G_V = -\frac{D_2}{D_1 D'_2 - D'_1 D_2} \tag{1.204}$$

and similarly for the decaying mode ($1 \leftrightarrow 2$). If $a_d \gg a_{\rm eq}$ we can take $y \propto \eta^2$ (see Table 1.1) and the limit $y \gg 1$, such that the above Eq. (1.203) gives:

$$\delta_m(\mathbf{k}, \eta) = \frac{a}{a_d}\left[\frac{3}{5}\delta_m(\mathbf{k}, \eta_d) + \frac{1}{5}\eta_d k^2 V_m(\mathbf{k}, \eta_d)\right]. \tag{1.205}$$

Hence, for scales much below the horizon at the drag epoch ($k\eta_d \gg 1$), the velocity term $kV_m(\mathbf{k}, \eta_d)$, if of the same order as $\delta_m(\mathbf{k}, \eta_d)$, dominates the growing mode: this is the velocity overshoot effect that we previously mentioned.

As in Eq. (1.201), the matter transfer function is split up between baryons (T_b) and cold dark matter (T_c):

$$T(k) = \frac{\Omega_{b,0}}{\Omega_{m,0}}T_b(k) + \left(1 - \frac{\Omega_{b,0}}{\Omega_{m,0}}\right)T_c(k). \tag{1.206}$$

Based on previous results (using Eq. (1.157), (1.168), (1.172), (1.191), (1.193), (1.194), (1.203), (1.204)), we can determine:

$$T_b(k) \propto \left(\frac{k_{\rm eq}}{k}\right)^2 D_b(k)\left(1 + R(\eta_d)\right)^{-1/4}\left[\cos\left(kr_{\rm drag}\right) - \frac{D_2(\eta_d)}{D'_2(\eta_d)}kc_s(\eta_d)\sin\left(kr_{\rm drag}\right)\right]G_\delta(\eta_d) \tag{1.207}$$

and (using Eq. (1.157), (1.199), (1.200), (1.203), (1.204)):

$$T_c(k) \propto \left(\frac{k_{\rm eq}}{k}\right)^2 \left\{G_\delta\left[A_1 U_1 + A_2 U_2\right] - G_V\left[A_1 U'_1 + A_2 U'_2\right]\right\}. \tag{1.208}$$

The baryon and cold dark matter transfer functions computed using Eisenstein and Hu (1998) analytic formula are shown in Figure 1.12. In practice, the transfer function is computed using Boltzmann codes such as CAMB (Lewis et al., 2000; Lewis and Challinor, 2011) or CLASS (Blas

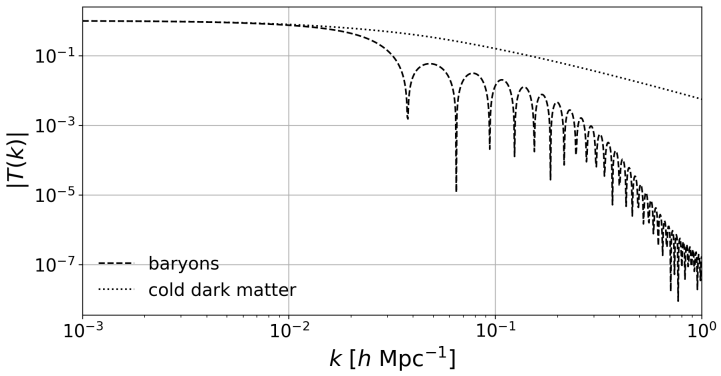

Figure 1.12 – Matter transfer function computed with the analytic formula of Eisenstein and Hu (1998) using the cosmological parameters of Planck Collaboration et al. (2018) (TT, TE, EE, lowE, lensing, BAO). The baryon and cold dark matter contributions are shown in dashed and dotted lines, respectively.

et al., 2011) — which also give predictions for the CMB power spectra, solving the full Boltzmann hierarchy.

After the drag epoch, baryons and cold dark matter both satisfy the following equation (Eq. (1.139) in the sub-Hubble regime and in the matter domination era, $\mathcal{H} = 2/\eta$):

$$\delta_a'' + \frac{2}{\eta}\delta_a' = -k^2\Phi, \qquad a \in \{b, c\}\,. \tag{1.209}$$

$\Phi \propto \mathcal{H}^2\delta \propto \eta^{-2}\delta$ (using the Poisson equation (1.127) in the sub-Hubble regime) is constant in the matter-dominated era since the growing mode is $\delta = \Omega_c\delta_c + \Omega_b\delta_b \propto a \propto \eta^2$ (using Eq. (1.209)). Hence, at late times, baryon and cold dark matter density contrasts converge to the same value $\delta_b = \delta_c = -k^2\eta^2\Phi/6$: gravitational evolution drives baryons and cold dark matter together.

We will see in Chapter 2 that galaxies trace the matter distribution. Therefore, the baryon acoustic oscillations imprinted in the matter distribution can be seen in the galaxy distribution. BAO were first detected in the galaxy correlation function (Eisenstein et al., 2005) of the SDSS and power spectrum (Cole et al., 2005) of the 2dF galaxy survey in 2005. A recent measure of BAO in the SDSS/BOSS galaxy sample is shown in Figure 1.13, which corresponds to a 7σ detection. Comparing the measured $r_{\rm drag}$ in the distribution of galaxies to its expected value helps constrain cosmological distances at the effective redshift of the galaxy sample, through the Alcock-Paczynski test which will be described in Section 4.2.1. However, in this manuscript we aim at going beyond this simple (but robust) measurement and describe the full shape of the galaxy power spectrum — which is the purpose of the next chapter.

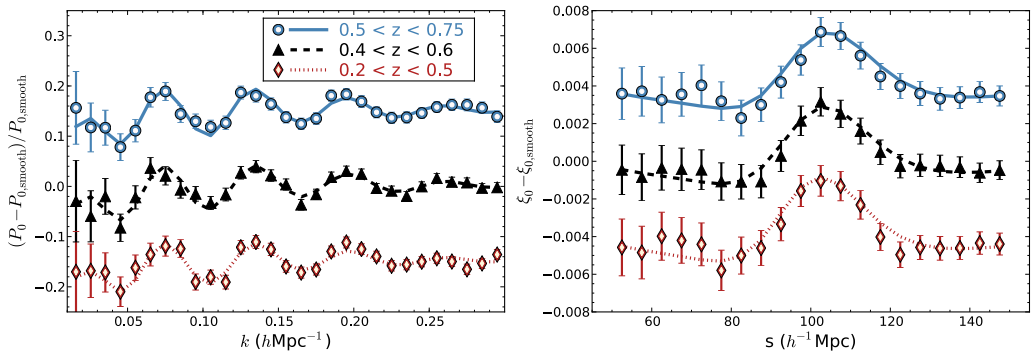

Figure 1.13 – BAO wiggles in the galaxy power spectrum and correlation function of SDSS/BOSS data, in three redshift bins: $0.2 < z < 0.5$, $0.4 < z < 0.6$ and $0.5 < z < 0.75$. Left: data points represent the measured power spectrum, divided by a smooth model without wiggles, and shifted by ± 0.15 for the lower and higher redshift bins. Right: data points represent the measured correlation function difference to a smooth model without the BAO peak, shifted by ± 0.004 for the lower and higher redshift bins. Taken from Alam et al. (2017).

Bibliography

Ahmad, Q. R., R. C. Allen, T. C. Andersen, J. D. Anglin, G. Bühler, et al. (2001). Measurement of the Rate of ν_e + d –> p + p + e^- Interactions Produced by ^{8}B Solar Neutrinos at the Sudbury Neutrino Observatory. Phys. Rev. Lett. *87*(7), 071301. arXiv:nucl-ex/0106015.

Ajhar, E. A., J. L. Tonry, J. P. Blakeslee, A. G. Riess, and B. P. Schmidt (2001). Reconciliation of the Surface Brightness Fluctuation and Type Ia Supernova Distance Scales. ApJ *559*(2), 584–591. arXiv:astro-ph/0105366.

Aker, M., K. Altenmüller, M. Arenz, M. Babutzka, J. Barrett, et al. (2019). Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN. Phys. Rev. Lett. *123*(22), 221802. arXiv:1909.06048.

Alam, S., M. Ata, S. Bailey, F. Beutler, D. Bizyaev, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. MNRAS *470*(3), 2617–2652. arXiv:1607.03155.

Bahcall, N. A. (1996). Clusters and superclusters of galaxies. *arXiv e-prints*, astro–ph/9611148. arXiv:astro-ph/9611148.

Bernardeau, F. (1998). Cosmologie, la formation des grandes structures de l'Univers. *arXiv e-prints*, astro–ph/9804230. arXiv:astro-ph/9804230.

BICEP2 Collaboration, Keck Array Collaboration, P. A. R. Ade, Z. Ahmed, R. W. Aikin, et al. (2016). Improved Constraints on Cosmology and Foregrounds from BICEP2 and Keck Array Cosmic Microwave Background Data with Inclusion of 95 GHz Band. Phys. Rev. Lett. *116*(3), 031302. arXiv:1510.09217.

Birrer, S., T. Treu, C. E. Rusu, V. Bonvin, C. D. Fassnacht, et al. (2019). H0LiCOW - IX. Cosmographic analysis of the doubly imaged quasar SDSS 1206+4332 and a new measurement of the Hubble constant. MNRAS *484*(4), 4726–4753. arXiv:1809.01274.

Blas, D., J. Lesgourgues, and T. Tram (2011). The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. J. Cosmology Astropart. Phys. *2011*(7), 034. arXiv:1104.2933.

Bonamente, M., M. K. Joy, S. J. LaRoque, J. E. Carlstrom, E. D. Reese, et al. (2006). Determination of the Cosmic Distance Scale from Sunyaev-Zel'dovich Effect and Chandra X-Ray Measurements of High-Redshift Galaxy Clusters. ApJ *647*(1), 25–54. arXiv:astro-ph/0512349.

Castorina, E., N. Hand, U. Seljak, F. Beutler, C.-H. Chuang, et al. (2019). Redshift-weighted constraints on primordial non-Gaussianity from the clustering of the eBOSS DR14 quasars in Fourier space. J. Cosmology Astropart. Phys. *2019*(9), 010. arXiv:1904.08859.

Chevallier, M. and D. Polarski (2001). Accelerating Universes with Scaling Dark Matter. *International Journal of Modern Physics D 10*(2), 213–223. arXiv:gr-qc/0009008.

Cole, S., W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, et al. (2005). The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. MNRAS *362*(2), 505–534. arXiv:astro-ph/0501174.

Corbelli, E. and P. Salucci (2000). The extended rotation curve and the dark matter halo of M33. MNRAS *311*(2), 441–447. arXiv:astro-ph/9909252.

de Blok, W. J. G. (2010). The Core-Cusp Problem. *Advances in Astronomy 2010*, 789293. arXiv:0910.3538.

de Salas, P. F. and S. Pastor (2016). Relic neutrino decoupling with flavour oscillations revisited. J. Cosmology Astropart. Phys. *2016*(7), 051. arXiv:hep-ph/1606.06986.

Dhawan, S., S. W. Jha, and B. Leibundgut (2018). Measuring the Hubble constant with Type Ia supernovae as near-infrared standard candles. A&A *609*, A72. arXiv:1707.00715.

Dodelson, S. (2003). *Modern cosmology*. Burlington: Academic Press.

Eisenstein, D. J. and W. Hu (1998). Baryonic Features in the Matter Transfer Function. ApJ *496*(2), 605–614. arXiv:astro-ph/9709112.

Eisenstein, D. J., I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, et al. (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. ApJ *633*(2), 560–574. arXiv:astro-ph/0501171.

Fixsen, D. J. (2009). The Temperature of the Cosmic Microwave Background. ApJ *707*(2), 916–920. arXiv:0911.1955.

Fixsen, D. J., E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, et al. (1996). The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set. ApJ *473*, 576. arXiv:astro-ph/9605054.

Freedman, W. L., B. F. Madore, D. Hatt, T. J. Hoyt, I. S. Jang, et al. (2019). The Carnegie-Chicago Hubble Program. VIII. An Independent Determination of the Hubble Constant Based on the Tip of the Red Giant Branch. ApJ *882*(1), 34. arXiv:1907.05922.

Hu, W. and S. Dodelson (2002). Cosmic Microwave Background Anisotropies. ARA&A *40*, 171–216. arXiv:astro-ph/0110414.

Hu, W. and N. Sugiyama (1996). Small-Scale Cosmological Perturbations: an Analytic Approach. ApJ *471*, 542. arXiv:astro-ph/9510117.

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

Kajita, T. (1999). Atmospheric neutrino results from Super-Kamiokande and Kamiokande - Evidence for ν_μ oscillations. *Nuclear Physics B Proceedings Supplements 77*(1-3), 123–132. arXiv:hep-ex/9810001.

Lewis, A. and A. Challinor (2011). CAMB: Code for Anisotropies in the Microwave Background.

Lewis, A., A. Challinor, and A. Lasenby (2000). Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. ApJ *538*(2), 473–476. arXiv:astro-ph/9911177.

LIGO Scientific Collaboration and Virgo Collaboration (2017). GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. Phys. Rev. Lett. *119*(16), 161101. arXiv:1710.05832.

Linder, E. V. (2003). Exploring the Expansion History of the Universe. Phys. Rev. Lett. *90*(9), 091301. arXiv:astro-ph/0208512.

Luri, X., A. G. A. Brown, L. M. Sarro, F. Arenou, C. A. L. Bailer-Jones, et al. (2018). Gaia Data Release 2. Using Gaia parallaxes. A&A *616*, A9. arXiv:1804.09376.

Mangano, G., G. Miele, S. Pastor, and M. Peloso (2002). A precision calculation of the effective number of cosmological neutrinos. *Physics Letters B 534*(1-4), 8–16. arXiv:astro-ph/0111408.

Mangano, G., G. Miele, S. Pastor, T. Pinto, O. Pisanti, et al. (2005). Relic neutrino decoupling including flavour oscillations. *Nuclear Physics B 729*(1-2), 221–234. arXiv:hep-ph/0506164.

Meszaros, P. (1974). The behaviour of point masses in an expanding cosmological substratum. A&A *37*(2), 225–228.

Navarro, J. F., C. S. Frenk, and S. D. M. White (1996). The Structure of Cold Dark Matter Halos. ApJ *462*, 563. arXiv:astro-ph/9508025.

Palanque-Delabrouille, N., C. Yèche, N. Schöneberg, J. Lesgourgues, M. Walther, et al. (2019). Hints, neutrino bounds and WDM constraints from SDSS DR14 Lyman-α and Planck full-survey data. *arXiv e-prints*, arXiv:1911.09073. arXiv:1911.09073.

Perlmutter, S., G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, et al. (1999). Measurements of Ω and Λ from 42 High-Redshift Supernovae. ApJ *517*(2), 565–586. arXiv:astro-ph/9812133.

Peter, P. and J.-P. Uzan (2012). *Cosmologie primordiale*. Belin.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2018). Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, arXiv:1807.06209. arXiv:1807.06209.

Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, et al. (2019). Planck 2018 results. IX. Constraints on primordial non-Gaussianity. *arXiv e-prints*, arXiv:1905.05697. arXiv:1905.05697.

Riess, A. G., S. Casertano, W. Yuan, L. Macri, B. Bucciarelli, et al. (2018). Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant. ApJ *861*(2), 126. arXiv:1804.10655.

Riess, A. G., S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic (2019). Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM. ApJ *876*(1), 85. arXiv:1903.07603.

Riess, A. G., A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, et al. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. AJ *116*(3), 1009–1038. arXiv:astro-ph/9805201.

Ross, A. J., W. J. Percival, A. Carnero, G.-b. Zhao, M. Manera, et al. (2013). The clustering of galaxies in the SDSS-III DR9 Baryon Oscillation Spectroscopic Survey: constraints on primordial non-Gaussianity. MNRAS *428*(2), 1116–1127. arXiv:1208.1491.

Schmidt, B. P. (1993). *Type II Supernovae, Expanding Photospheres, and the Extragalactic Distance Scale*. Ph. D. thesis, HARVARD UNIVERSITY.

Schwarz, D. J. (2010). Cosmological backreaction. *arXiv e-prints*, arXiv:1003.3026. arXiv:1003.3026.

Scolnic, D. M., D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, et al. (2018). The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. ApJ *859*(2), 101. arXiv:1710.00845.

Slosar, A., C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan (2008). Constraints on local primordial non-Gaussianity from large scale structure. J. Cosmology Astropart. Phys. *2008*(8), 031. arXiv:astro-ph/0805.3580.

Tanabashi, M., K. Hagiwara, K. Hikasa, K. Nakamura, Y. Sumino, et al. (2018). Review of Particle Physics*. Phys. Rev. D *98*(3), 030001.

Tortora, C., L. Posti, L. V. E. Koopmans, and N. R. Napolitano (2019). The dichotomy of dark matter fraction and total mass density slope of galaxies over five dex in mass. MNRAS *489*(4), 5483–5493. arXiv:1902.10158.

Voruz, L., J. Lesgourgues, and T. Tram (2014). The effective gravitational decoupling between dark matter and the CMB. J. Cosmology Astropart. Phys. *2014*(3), 004. arXiv:1312.5301.

Weinberg, S. (2002). Cosmological Fluctuations of Small Wavelength. ApJ *581*(2), 810–816. arXiv:astro-ph/0207375.

Wong, K. C., S. H. Suyu, G. C. F. Chen, C. E. Rusu, M. Millon, et al. (2019). H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes. *arXiv e-prints*, arXiv:1907.04869. arXiv:1907.04869.

Zwicky, F. (1937). On the Masses of Nebulae and of Clusters of Nebulae. ApJ *86*, 217.

Large scale structure

So far we have considered perturbations in the linear regime only. While we could restrict analyses of the large scale structure to linear scales (typically $k \lesssim 0.1\,h\,\mathrm{Mpc}^{-1}$ at $z = 1$), the amount of information (for a Gaussian field) scales as k^3 — that is, increasing the maximum wavenumber used in the analysis by a factor of 2 corresponds to a spectroscopic survey roughly 8 times bigger — hence a precision increased by $\sqrt{8}$. Apart from the theoretical motivation, there is therefore much practical interest in describing the evolution of density perturbations in the quasi-linear regime — though it comes at the cost of introducing new parameters to encompass our limited knowledge of the small scale physics (thereby reducing the effective gain in precision).

Theoretical literature on large scale structure is quite substantial. Instead of trying to give a general overview of its wonderful developments, we rather focus on those used in the clustering analysis we performed — while briefly commenting on alternate approaches.

2.1 Perturbation theory

After the baryon drag epoch, baryons and dark matter are driven through the effect of gravitation towards a single fluid of matter (see Section 1.3.5.3). As we focus on sub-Hubble scales, relativistic terms in Einstein equations can be dropped (see Section 1.3.2.5), and the dynamics be described with Newtonian gravity (in an expanding Universe). This is the approach followed in this section, which relies on the review Bernardeau et al. (2002) and on papers Crocce and Scoccimarro (2006a,b); Bernardeau et al. (2012); Taruya et al. (2012).

2.1.1 Vlasov-Poisson equations

We describe matter as a collisionless fluid that evolves only through gravitation, in an expanding Universe. Phase-space conservation of the matter distribution function $f(\mathbf{x}, \mathbf{p}, \eta)$ implies (Liouville theorem):

$$\frac{df}{d\eta} = \frac{\partial f}{\partial \eta} + \frac{d\mathbf{x}}{d\eta} \cdot \frac{\partial f}{\partial \mathbf{x}} + \frac{d\mathbf{p}}{d\eta} \cdot \frac{\partial f}{\partial \mathbf{p}} = 0. \tag{2.1}$$

We recall η is the conformal time, related to the cosmic time t by $dt = a d\eta$. $\mathbf{x}$ and $\mathbf{p}$ are the fluid comoving position and momentum. All spatial derivatives will be taken with respect to comoving coordinates. We describe deviations to the smooth background expanding Universe. Hence, the momentum is related to the (physical) peculiar velocity $\mathbf{u} = a d\mathbf{x}/dt = d\mathbf{x}/d\eta$ as:

$$\mathbf{p} = am\mathbf{u} \tag{2.2}$$

with m the particle mass. The equation of motion reads:

$$\frac{dp}{d\eta} = -am\nabla\Phi. \tag{2.3}$$

Hence, Eq. (2.1) becomes the Vlasov equation:

$$\frac{\partial f}{\partial \eta} + \frac{\mathbf{p}}{ma} \cdot \frac{\partial f}{\partial \mathbf{x}} - am\nabla\Phi \cdot \frac{\partial f}{\partial \mathbf{p}} = 0 \tag{2.4}$$

and we consider the Poisson equation:

$$\Delta\Phi = a^2 4\pi G_\mathrm{N} \bar{\rho}\delta = \frac{3}{2}\mathcal{H}^2\Omega_m\delta \tag{2.5}$$

where we used $\Delta = \partial_i \partial_i$ and replaced the background matter density $\bar{\rho}$ using the critical density (1.41). We recall that $\delta = \rho/\bar{\rho} - 1$ is the fluid density contrast. Eq. (2.5) corresponds to its counterpart in general relativity (1.113), if, in the latter:

— we neglect curvature ($K = 0$), which is legitimate at the scales of interest and since the Universe is nearly flat;

— we neglect velocity (V), which is legitimate in the non-relativistic limit;

— we take $\Phi = \Psi$, i.e. $\bar{P}\bar{\pi}$ in Eq. (1.114) is negligible, which is legitimate in the non-relativistic limit ($\bar{P} = 0$).

The first three moments of the phase-space distribution function are, by definition:

$$\int d^3p f(\mathbf{x}, \mathbf{p}, \eta) = \rho(\mathbf{x}, \eta) \tag{2.6}$$

$$\int d^3p \frac{\mathbf{p}}{am} f(\mathbf{x}, \mathbf{p}, \eta) = \rho(\mathbf{x}, \eta)\mathbf{u}(\mathbf{x}, \eta) \tag{2.7}$$

$$\int d^3p \frac{p_i p_j}{a^2 m^2} f(\mathbf{x}, \mathbf{p}, \eta) = \rho(\mathbf{x}, \eta) \left[u_i(\mathbf{x}, \eta) u_j(\mathbf{x}, \eta) + \sigma_{ij}(\mathbf{x}, \eta)\right] \tag{2.8}$$

with σ_{ij} the stress tensor. Taking the zeroth moment of Eq. (2.4), one finds the continuity equation:

$$\boxed{\delta' + \partial_i \left[(1 + \delta)\, u_i\right] = 0} \tag{2.9}$$

and taking the first moment, subtracting the above continuity equation gives the Euler equation:

$$\boxed{u_i' + \mathcal{H} u_i + u_j \partial_j u_i = -\partial_i \Phi - \frac{1}{\rho} \partial_j \left(\rho \sigma_{ij}\right)} \tag{2.10}$$

We recall $'$ denotes a derivative with respect to the conformal time η. The single stream evolution is the regime where the fluid evolves in a single coherent flow; at any position $\mathbf{x}$ (and time η) the fluid has a unique velocity $\mathbf{u}(\mathbf{x}, \eta)$, $f(\mathbf{x}, \mathbf{p}, \eta) = \rho(\mathbf{x}, \eta) \delta_D^{(3)}(\mathbf{p} - am\mathbf{u})$. The stress tensor σ_{ij} represents deviations to this regime. In the mildly non-linear regime, before structures collapse and virialise, $\sigma_{ij} = 0$ to a good approximation (e.g. Bernardeau et al., 2002), which we adopt in the following.

2.1.2 Linear regime

Let us start with the linear regime: $\delta, u_i \ll 1$, i.e. the density contrast and velocities (as well as their gradients) are small. We define $\theta = \nabla \cdot \mathbf{u}$ the divergence of the velocity field and $\mathbf{w} = \nabla \times \mathbf{u}$ its vorticity. From Eq. (2.9) and taking the divergence and vorticity of Eq. (2.10), we obtain:

$$\delta' + \theta = 0 \tag{2.11}$$

$$\theta' + \mathcal{H}\theta = -\Delta\Phi \tag{2.12}$$

$$\mathbf{w}' + \mathcal{H}\mathbf{w} = \mathbf{0}. \tag{2.13}$$

The last equation ensures that $w \propto a^{-1}$, i.e. vorticity is diluted in expansion. We will thus neglect it in the following. Deriving the first equation, injecting the second and using Eq. (2.5) to replace the potential, one finds:

$$\delta'' + \mathcal{H}\delta' - \frac{3}{2}\mathcal{H}^2 \Omega_m \delta = 0 \tag{2.14}$$

which we already found and solved in Section 1.3.2.5 (see Eq. (1.153)). The general solution is given in Eq. (1.154). We simply recall that in an Einstein de Sitter Universe (i.e. flat Universe with $\Omega_m = 1$), the growing and decaying solutions read:

$$D_+(a) \propto a \qquad D_-(a) \propto a^{-3/2}. \tag{2.15}$$

The logarithmic growth rate of linear perturbations is defined as the logarithmic derivative of the linear growth factor D_+:

$$\boxed{f = \frac{d\ln D_+}{d\ln a}} \tag{2.16}$$

Then, for the growing mode $\delta \propto D_+$:

$$\delta' = \frac{D'_+}{D_+}\delta = f\mathcal{H}\delta \tag{2.17}$$

and from Eq. (2.11) we get, in the linear regime:

$$\boxed{\theta = -f\mathcal{H}\delta} \tag{2.18}$$

f can be very well parametrised by Ω_m^γ. As an example, Linder (2005) find:

$$f \simeq \Omega_m^{0.55+0.05[1+w(z=1)]} \qquad w = -\frac{1}{3}\frac{d\ln\left[\Omega_m(a)^{-1}-1\right]}{d\ln a} \tag{2.19}$$

is accurate to a fraction of a percent for typical CDM cosmologies with additional dark energy fluid. In particular, $f \simeq \Omega_m^{0.55}$ to better than 0.05% if $\Omega_m \in [0.22, 1]$ in a flat ΛCDM Universe.

Turning the linear system of Eq. (2.5) and Eq. (2.11) - (2.12) in Fourier space ($\Delta \to -k^2$) we see that $\mathbf{k}$ modes are independent of each other. Any coupling between different wavenumbers in the equations below stems from non-linearity.

We finally derive a useful expression for the linear order velocity in Fourier space. Taking $\mathbf{u}$ to be curl-free ($\mathbf{w} = \mathbf{0}$), as argued is legitimate at the beginning of the section, we have $\mathbf{u} = -i\theta\mathbf{k}/k^2$, and using Eq. (2.18) we find:

$$\boxed{\mathbf{u}(\mathbf{k}) = if\mathcal{H}\frac{\mathbf{k}}{k^2}\delta(\mathbf{k})} \tag{2.20}$$

2.1.3 Beyond the linear regime

As we mentioned previously, to both improve cosmological constraints and understand early structure formation we wish to extend the description of the power spectrum beyond the linear regime. In this section, we will assume that vorticity remains zero. Therefore, the velocity field is fully described by its divergence θ. In the following, we will work in Fourier space. In this space, Eq. (2.9) becomes:

$$\delta'(\mathbf{k},\eta) + \theta(\mathbf{k},\eta) = -\int \frac{d^3k_1}{(2\pi)^3}\frac{\mathbf{k}\cdot\mathbf{k}_1}{k_1^2}\theta(\mathbf{k}_1,\eta)\delta(\mathbf{k}-\mathbf{k}_1,\eta). \tag{2.21}$$

Defining the kernel:

$$\alpha(\mathbf{k}_1,\mathbf{k}_2) = \frac{\mathbf{k}_{12}\cdot\mathbf{k}_1}{k_1^2} \qquad \mathbf{k}_{12} = \mathbf{k}_1 + \mathbf{k}_2, \tag{2.22}$$

the continuity equation can be rewritten:

$$\delta'(\mathbf{k},\eta)+\theta(\mathbf{k},\eta)=-\int\frac{d^3k_1d^3k_2}{(2\pi)^3}\delta_D^{(3)}(\mathbf{k}-\mathbf{k}_{12})\alpha(\mathbf{k}_1,\mathbf{k}_2)\theta(\mathbf{k}_1,\eta)\delta(\mathbf{k}_2,\eta) \tag{2.23}$$

Similarly, taking the divergence of Eq. (2.10) and replacing the potential with Eq. (2.5) gives:

$$\theta'(\mathbf{k},\eta)+\mathcal{H}\theta(\mathbf{k},\eta)+\frac{3}{2}\Omega_m\mathcal{H}^2\delta(\mathbf{k},\eta)=-\int\frac{d^3k_1d^3k_2}{(2\pi)^3}\delta_D^{(3)}(\mathbf{k}-\mathbf{k}_{12})\beta(\mathbf{k}_1,\mathbf{k}_2)\theta(\mathbf{k}_1,\eta)\theta(\mathbf{k}_2,\eta) \tag{2.24}$$

with:

$$\beta(\mathbf{k}_1,\mathbf{k}_2)=\frac{(\mathbf{k}_{12}\cdot\mathbf{k}_2)\,(\mathbf{k}_1\cdot\mathbf{k}_2)}{k_1^2k_2^2}=\frac{k_{12}^2\mathbf{k}_1\cdot\mathbf{k}_2}{2k_1^2k_2^2} \tag{2.25}$$

where the least equality is obtained by symmetrisation $1\leftrightarrow 2$.

Let us define the vector:

$$\mathbf{\Psi}=\begin{pmatrix}\delta\\-\frac{\theta}{f\mathcal{H}}\end{pmatrix} \tag{2.26}$$

and rewrite the time derivatives with respect to:

$$u=\ln D_+, \tag{2.27}$$

the number of e-folds of expansion. The continuity and Euler equations give:

$$\frac{\partial\Psi_1(\mathbf{k},u)}{\partial u}-\Psi_2(\mathbf{k},u)=\int\frac{d^3k_1d^3k_2}{(2\pi)^3}\delta_D^{(3)}(\mathbf{k}-\mathbf{k}_{12})\alpha(\mathbf{k}_1,\mathbf{k}_2)\Psi_2(\mathbf{k}_1,u)\Psi_1(\mathbf{k}_2,u) \tag{2.28}$$

$$\frac{\partial\Psi_2(\mathbf{k},u)}{\partial u}+\left(\frac{3\Omega_m}{2f^2}-1\right)\Psi_2(\mathbf{k},u)-\frac{3\Omega_m}{2f^2}\Psi_1(\mathbf{k},u)=\int\frac{d^3k_1d^3k_2}{(2\pi)^3}\delta_D^{(3)}(\mathbf{k}-\mathbf{k}_{12})\beta(\mathbf{k}_1,\mathbf{k}_2)\Psi_2(\mathbf{k}_1,u)\Psi_2(\mathbf{k}_2,u) \tag{2.29}$$

which can be rewritten in matrix form:

$$\left[\delta_{ab}\frac{\partial}{\partial u}+\Omega_{ab}\right]\Psi_b(\mathbf{k},u)=\int\frac{d^3k_1d^3k_2}{(2\pi)^3}\delta_D^{(3)}(\mathbf{k}-\mathbf{k}_{12})\gamma_{abc}(\mathbf{k}_1,\mathbf{k}_2)\Psi_b(\mathbf{k}_1,u)\Psi_c(\mathbf{k}_2,u) \tag{2.30}$$

where δ_{ab} is the Kronecker delta ($\delta_{ab}=1$ if $a=b$, else 0) and:

$$\Omega_{ab}=\begin{pmatrix}0 & -1\\ -\frac{3}{2f^2}\Omega_m & \frac{3}{2f^2}\Omega_m-1\end{pmatrix}\qquad\gamma_{abc}(\mathbf{k}_1,\mathbf{k}_2)=\begin{cases}\frac{1}{2}\alpha(\mathbf{k}_1,\mathbf{k}_2) & (a,b,c)=(1,2,1)\\ \frac{1}{2}\alpha(\mathbf{k}_2,\mathbf{k}_1) & (a,b,c)=(1,1,2)\\ \beta(\mathbf{k}_1,\mathbf{k}_2) & (a,b,c)=(2,2,2)\\ 0 & \text{otherwise}\end{cases} \tag{2.31}$$

where we symmetrised $(a,b,c)=(1,2,1)$ and $(a,b,c)=(1,1,2)$. One would note that the time evolution can be fully decoupled from the spatial evolution if $f(a)=\Omega_m(a)^{0.5}$, which is a very good approximation, as described in the previous section (see Eq. (2.19)), especially since most of the structure growth takes place when $\Omega_m(a)\simeq 1$ (e.g. Bernardeau et al., 2002). Therefore, perturbation theory (PT) results will be very well approximated by considering $f(a)=\Omega_m(a)^{0.5}$ — which is assumed in the following, except otherwise stated — or, equivalently, performing the

calculation in an Einstein de Sitter Universe ($D_+(a) = a$) and replace a by the growth factor of the desired ΛCDM cosmology.

Let us search for solutions $\mathbf{\Psi}$ using the perturbative expansion:

$$\Psi_a(\mathbf{k}, u) = \sum_{n=1}^{+\infty} e^{nu} \psi_a^{(n)}(\mathbf{k}) \tag{2.32}$$

where we keep the fastest growing mode $\propto e^{nu} = D_+^n$ for the term of order n. Note that $\psi_1^{(1)} = \psi_2^{(1)} = \delta_0$ from linear theory (see Eq. (2.18)). The recursion relation reads:

$$\sigma_{ab}^{-1}(n)\psi_b^{(n)}(\mathbf{k}) = \int \frac{d^3k_1 d^3k_2}{(2\pi)^3} \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_{12}) \gamma_{abc}(\mathbf{k}_1, \mathbf{k}_2) \sum_{m=1}^{n-1} \psi_b^{(m)}(\mathbf{k}_1) \psi_c^{(n-m)}(\mathbf{k}_2). \tag{2.33}$$

The inverse of $\sigma_{ab}^{-1}(n) = n\delta_{ab} + \Omega_{ab}$ is computed as:

$$\sigma_{ab}(n) = \frac{1}{(2n+3)(n-1)} \begin{pmatrix} 2n+1 & 2 \\ 3 & 2n \end{pmatrix}. \tag{2.34}$$

$\psi_a^{(n)}(\mathbf{k})$ is an expression of order n in δ_0, hence injecting

$$\psi_a^{(n)}(\mathbf{k}) = \int \frac{d^3q_1 \cdots d^3q_n}{(2\pi)^{3(n-1)}} \delta_D^{(3)}(\mathbf{k} - \mathbf{q}_1 - \cdots - \mathbf{q}_n) F_a^{(n)}(\mathbf{q}_1, \cdots \mathbf{q}_n) \delta_0(\mathbf{q}_1) \cdots \delta_0(\mathbf{q}_n) \tag{2.35}$$

into Eq. (2.33), which must be verified for any initial field δ_0 leads to the recursion relation for the perturbation theory kernels:

$$F_a^{(n)}(\mathbf{q}_1, \cdots \mathbf{q}_n) = \sigma_{ab}(n) \sum_{m=1}^{n-1} \gamma_{bcd}(\mathbf{q}_1 + \cdots + \mathbf{q}_m, \mathbf{q}_{m+1} + \cdots + \mathbf{q}_n) F_c^{(m)}(\mathbf{q}_1, \cdots \mathbf{q}_m) F_d^{(n-m)}(\mathbf{q}_{m+1}, \cdots \mathbf{q}_n) \tag{2.36}$$

and $F_1^{(1)} = F_2^{(1)} = 1$. We will note $F_{a,s}^{(n)}$ the symmetrised version of these kernels, i.e. their average over all permutations of the wavevectors $\mathbf{q}_i$. In particular, the symmetrised second order kernels read:

$$F_{1,s}^{(2)}(\mathbf{q}_1, \mathbf{q}_2) = \frac{5}{7} + \frac{1}{2} \frac{\mathbf{q}_1 \cdot \mathbf{q}_2}{q_1 q_2} \left(\frac{q_1}{q_2} + \frac{q_2}{q_1} \right) + \frac{2}{7} \left(\frac{\mathbf{q}_1 \cdot \mathbf{q}_2}{q_1 q_2} \right)^2 \tag{2.37}$$

$$F_{2,s}^{(2)}(\mathbf{q}_1, \mathbf{q}_2) = \frac{3}{7} + \frac{1}{2} \frac{\mathbf{q}_1 \cdot \mathbf{q}_2}{q_1 q_2} \left(\frac{q_1}{q_2} + \frac{q_2}{q_1} \right) + \frac{4}{7} \left(\frac{\mathbf{q}_1 \cdot \mathbf{q}_2}{q_1 q_2} \right)^2. \tag{2.38}$$

So far, we have only considered the fastest growing modes $\propto e^{nu} = D_+^n$. Note however that initially growing modes can couple to decaying modes. Let us reintroduce the full time dependence of the PT solutions. Taking the Laplace transform $\mathcal{L}$ (with respect to u) of Eq. (2.30):

$$\sigma_{ab}^{-1}(\omega)\Psi_b(\mathbf{k}, \omega) = \phi_a(\mathbf{k}) + \int \frac{d^3k_1 d^3k_2}{(2\pi)^3} \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_{12}) \gamma_{abc}(\mathbf{k}_1, \mathbf{k}_2) \mathcal{L}\left(\Psi_b(\mathbf{k}_1, u)\Psi_c(\mathbf{k}_2, u)\right)(\omega) \tag{2.39}$$

where $\phi_a(\mathbf{k}) = \Psi_a(\mathbf{k}, u = 0)$ is the initial condition. Multiplying both sides by $\sigma_{ab}(\omega)$ and taking the inverse Laplace transform gives:

$$\boxed{\Psi_a(\mathbf{k}, u) = g_{ab}(u)\phi_b(\mathbf{k}) + \int_0^u dv g_{ab}(u - v) \int \frac{d^3k_1 d^3k_2}{(2\pi)^3} \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_{12}) \gamma_{bcd}(\mathbf{k}_1, \mathbf{k}_2) \Psi_c(\mathbf{k}_1, v) \Psi_d(\mathbf{k}_2, v)} \tag{2.40}$$

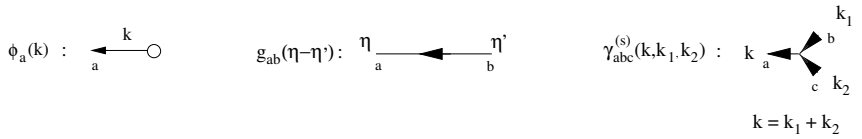

Figure 2.1 – The diagrammatic representation of Eq. (2.40) building blocks: the initial field (left), the propagator (middle) and the vertex (right). Here and in all diagrams in this section, time u is denoted η. Taken from Crocce and Scoccimarro (2006b).

where $g_{ab}(u)$, the Laplace transform of $\sigma_{ab}(\omega)$ is called *linear propagator*. Causality ensures that $g_{ab}(u)$ is non-zero for $u > 0$ only. The expression above holds in any cosmology (assuming Newtonian dynamics), provided the correct Ω_{ab} is used. If $f = \Omega_m^{0.5}$, we have:

$$g_{ab}(u) = \frac{e^u}{5}\begin{pmatrix} 3 & 2 \\ 3 & 2 \end{pmatrix} - \frac{e^{-3u/2}}{5}\begin{pmatrix} -2 & 2 \\ 3 & -3 \end{pmatrix}. \tag{2.41}$$

The first term of Eq. (2.40) simply represents the linear evolution of the initial conditions through time encoded in the linear propagator. Conversely, the second part represents mode coupling: the two fields $\Psi_c(\mathbf{k}_1, v)$ and $\Psi_d(\mathbf{k}_2, v)$ interact at time $v \leq u$ through the vertex $\gamma_{bcd}(\mathbf{k}_1, \mathbf{k}_2)$, with the conservation of wavenumber (outgoing $\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2$) due to statistical homogeneity ensured by $\delta_D^{(3)}(\mathbf{k} - \mathbf{k}_{12})$. The result of this interaction is then linearly evolved up to u with $g_{ab}(u - v)$. Again, we seek a solution to Eq. (2.40) with the series:

$$\Psi_a(\mathbf{k}, u) = \sum_{n=1}^{+\infty} \Psi_a^{(n)}(\mathbf{k}, u) \tag{2.42}$$

with $\Psi_a^{(n)}$ written as:

$$\Psi_a^{(n)}(\mathbf{k}, u) = \int \frac{d^3q_1 \cdots d^3q_n}{(2\pi)^{3(n-1)}} \delta_D^{(3)}(\mathbf{k} - \mathbf{q}_1 - \cdots \mathbf{q}_n) \mathcal{F}_a^{(n)}(\mathbf{q}_1, \cdots \mathbf{q}_n, u) \delta_0(\mathbf{q}_1) \cdots \delta_0(\mathbf{q}_n). \tag{2.43}$$

$\mathcal{F}_a^{(n)}$ kernels are obtained via a recursion relation, similar to Eq. (2.36):

$$\mathcal{F}_a^{(n)}(\mathbf{q}_1, \cdots \mathbf{q}_n, u) = \int_0^u dv g_{ab}(u - v) \sum_{m=1}^{n-1} \gamma_{bcd}(\mathbf{q}_1 + \cdots \mathbf{q}_m, \mathbf{q}_{m+1} + \cdots \mathbf{q}_n) \\ \mathcal{F}_c^{(m)}(\mathbf{q}_1, \cdots \mathbf{q}_m, v) \mathcal{F}_d^{(n-m)}(\mathbf{q}_{m+1}, \cdots \mathbf{q}_n, v). \tag{2.44}$$

We will note $\mathcal{F}_{a,s}^{(n)}$ their symmetrised version.

Eq. (2.40) can be represented in terms of diagrams. There are three ingredients, represented in Figure 2.1: the initial field ϕ_a, the linear propagator g_{ab} and the vertex γ_{abc}.

The n-th order term $\Psi_a^{(n)}$ of the expansion 2.42 can also be obtained with a diagrammatic representation (see Figure 2.2). The n-th order diagram, with $n-1$ vertices connecting n initial fields (represented with circles) is obtained with the following procedure. Starting from a leaf at time u, going back in time leads to either an initial field (for $\Psi_a^{(1)}(\mathbf{k})$) or a vertex at intermediate time s, opening up two branches, themselves leading to an initial field (for $\Psi_a^{(2)}(\mathbf{k})$) or another vertex. Then, each vertex is attributed a factor of 2 if asymmetric. The result of the procedure for terms $n = 1$ to $n = 4$ can be seen in Figure 2.2. It is straightforward to build up the n-th order solution from these diagrams. Oriented lines (in increasing time) represent the linear propagator g_{ab}. Any pair of fields entering a vertex are multiplied (with their weight) to obtain

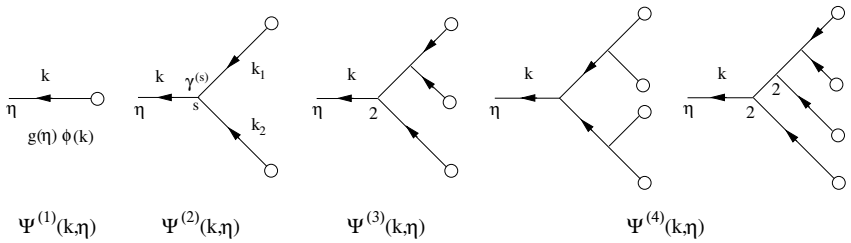

Figure 2.2 – The diagrammatic representation of the n-th order term $\Psi_a^{(n)}(\mathbf{k}, u)$, from $n = 1$ to $n = 4$. Taken from Crocce and Scoccimarro (2006b).

a new field with a wavevector equal to the sum of the incoming ones at a time v such that $0 \leq v \leq u$. All intermediate wavevectors and time are integrated over.

For example, the $n = 2$ and $n = 3$ terms are given by:

$$\Psi_a^{(2)}(\mathbf{k}, u) = \int_0^u dv g_{ab}(u - v) \int \frac{d^3k_1 d^3k_2}{(2\pi)^3} \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) \gamma_{bcd}(\mathbf{k}_1, \mathbf{k}_2) g_{ce}(v) \phi_e(\mathbf{k}_1) g_{df}(v) \phi_f(\mathbf{k}_2) \tag{2.45}$$

$$\Psi_a^{(3)}(\mathbf{k}, u) = 2 \int_0^u dv g_{ab}(u - v) \int \frac{d^3k_1 d^3k_2}{(2\pi)^3} \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) \gamma_{bcd}(\mathbf{k}_1, \mathbf{k}_2) g_{ce}(v) \phi_e(\mathbf{k}_1) \Psi_d^{(2)}(\mathbf{k}_2, v). \tag{2.46}$$

Initial conditions will be taken proportional to the same field δ_0:

$$\phi_a(\mathbf{k}) = \varepsilon_a \delta_0(\mathbf{k}). \tag{2.47}$$

In the purely growing mode $\varepsilon_1 = \varepsilon_2 = 1$. In this case, using Eq. (2.41), in Eq. (2.40) we have $g_{ab}(u)\phi_b(\mathbf{k}) = e^u \delta_0(\mathbf{k})$, i.e. $\mathcal{F}_{a,s}^{(1)} = e^u F_{a,s}^{(1)}$. Then, from Eq. (2.44), setting initial conditions infinitely far away in the past, i.e. the integration lower bound to $-\infty$, such that negative exponentials in Eq. (2.41) vanish, recursively we recover standard PT kernels, i.e. $\mathcal{F}_{a,s}^{(n)} = e^{nu} F_{a,s}^{(n)}$.

2.1.4 Lagrangian approach

Before we utilise the previous results to compute the PT prediction for the power spectrum, for completeness let us briefly introduce another approach to the Eulerian description adopted so far.

In the Eulerian picture, we described the fluid through its phase-space distribution (or its momenta, density and velocity) at a given position $\mathbf{x}$ and conformal time η. The Lagrangian approach instead follows the particle positions. Namely, the position at any time η is:

$$\mathbf{x}(\eta) = \mathbf{q} + \mathbf{\Psi}(\mathbf{q}, \eta) \tag{2.48}$$

where $\mathbf{q}$ is the particle initial position and $\mathbf{\Psi}(\mathbf{q}, \eta)$ the displacement field. Mass conservation ensures that $(1 + \delta(\mathbf{x}))\, d^3x = d^3q$. Hence, the density can be related to the displacement field:

$$1 + \delta(\mathbf{x}, \eta) = \frac{1}{|\mathbf{M}(\mathbf{x}, \eta)|} = \frac{1}{J(\mathbf{x}, \eta)} \qquad \text{with} \qquad M_{ij} = \delta_{ij} + \frac{\partial \Psi_i}{\partial q_j}. \tag{2.49}$$

Note that the above relation breaks down when $J(\mathbf{x}, \eta)$ vanishes. This happens when particles at different initial positions $\mathbf{q}$ end up at the same position $\mathbf{x}$, which is called shell-crossing. The

equation of motion reads:

$$\frac{d^2\mathbf{x}}{d\eta^2} + \mathcal{H}\frac{d\mathbf{x}}{d\eta} = -\nabla\Phi. \tag{2.50}$$

Taking its divergence and replacing the potential with Eq. (2.5) yields:

$$J(\mathbf{x},\eta)\mathbf{M}^{-1}_{ij}\frac{\partial}{\partial q_j}\left(\frac{d^2\Psi_i}{d\eta^2} + \mathcal{H}\frac{d\Psi_i}{d\eta}\right) = \frac{3}{2}\Omega_m\mathcal{H}^2\left[J(\mathbf{q},\eta) - 1\right]. \tag{2.51}$$

This non-linear equation can be solved perturbatively for $\mathbf{\Psi}$.

A useful approximation was introduced by Zel'dovich (Zel'dovich, 1970). The particle displacement is taken to be the (irrotational) solution in linear theory:

$$\mathbf{\Psi}(\mathbf{q},\eta) = iD_+(\eta)\int\frac{d^3k}{(2\pi)^3}\frac{\mathbf{k}}{k^2}\delta_0(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{q}} \tag{2.52}$$

which is simply $\mathbf{u}/(f\mathcal{H})$ with $\mathbf{u}$ the linear order velocity as given by Eq. (2.20). The density field is deduced from mass conservation ($(1+\delta(\mathbf{x}))\,d^3x = d^3q$):

$$1 + \delta(\mathbf{x},\eta) = \frac{1}{\left[1-\lambda_1 D_+(\eta)\right]\left[1-\lambda_2 D_+(\eta)\right]\left[1-\lambda_3 D_+(\eta)\right]} \tag{2.53}$$

where λ_i, $i \in [1,3]$ are the eigenvalues of the tensor $\partial\Psi_i/\partial q_j$. Gravitational collapse happens when (at least) one $\lambda_i D_+(\eta) \to 1$. As D_+ increases, depending on the eigenvalues, different structures will form:

— if $\lambda_1 \ll \lambda_2, \lambda_3$ and $\lambda_1 > 0$ (or similar permutation): a pancake;

— if $\lambda_1 \sim \lambda_2 \ll \lambda_3$ and $\lambda_1, \lambda_2 > 0$ (or similar permutation): a filament;

— if $\lambda_1 \sim \lambda_2 \sim \lambda_3 > 0$: a peak;

— if $\lambda_1, \lambda_2, \lambda_3 < 0$: an underdense region (void).

Let us come back to the Eulerian description and derive the PT prediction for the power spectrum.

2.1.5 The power spectrum

So far we have described the non-linear evolution of the density contrast. To have access to its statistical properties, e.g. its variance given by the power spectrum, we will use the well-known Wick theorem. Given a Gaussian field δ of mean 0, for any integer n:

$$\begin{aligned}\langle\delta(\mathbf{k}_1)\cdots\delta(\mathbf{k}_{2n+1})\rangle &= 0\\ \langle\delta(\mathbf{k}_1)\cdots\delta(\mathbf{k}_{2n})\rangle &= \sum_{\text{combinations}}\prod_{(i,j)}\langle\delta(\mathbf{k}_i)\delta(\mathbf{k}_j)\rangle\end{aligned} \tag{2.54}$$

In the second case, the sum has $(2n-1)!!$ (double factorial) terms, corresponding to all the possibilities of pairing $2n$ fields. In all the following, we will assume that initial fields ϕ_a follow Gaussian statistics, as is (almost) the case by the end of single-field slow-roll inflation (see Section 1.2.5). Any n-point correlation of initial fields can therefore be expressed as a sum of products of 2-point correlations. Assuming Eq. (2.47), these are related to the power spectrum $P_0(\mathbf{k})$ of the initial field δ_0 by:

$$\langle\phi_a(\mathbf{k})\phi_b(\mathbf{k}')\rangle = (2\pi)^3\,\delta_D^{(3)}(\mathbf{k}+\mathbf{k}')\varepsilon_a\varepsilon_b P_0(\mathbf{k}). \tag{2.55}$$

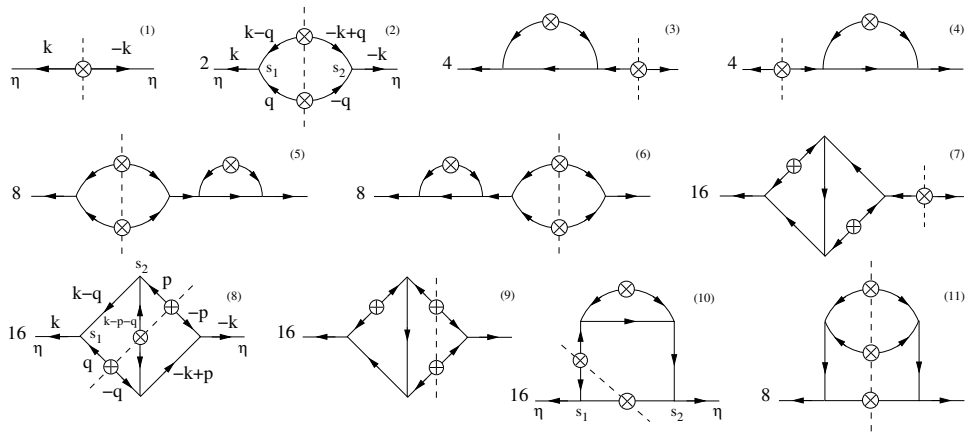

Figure 2.3 – The diagrammatic representation of the power spectrum at tree order (top left), the 3 1-loop diagrams (top right) and 7 out of the 29 2-loop diagrams (bottom). Dashed lines symbolise the connection between the diagrams of $\Psi_a^{(n)}(\mathbf{k}, u)$ (see Figure 2.2). Taken from Crocce and Scoccimarro (2006b).

The power spectrum of the evolved (density or velocity divergence) fields:

$$\left\langle \Psi_a(\mathbf{k}, u)\Psi_b(\mathbf{k}', u)\right\rangle = (2\pi)^3\,\delta_D^{(3)}(\mathbf{k}+\mathbf{k}')P_{ab}(\mathbf{k}, u) \tag{2.56}$$

is obtained by replacing $\Psi_a(\mathbf{k}, u)$ by their series expansion (2.42). One obtains:

$$P_{ab}(\mathbf{k}, u) = \sum_{l=0}^{+\infty} P_{ab}^{(l)}(\mathbf{k}, u) \tag{2.57}$$

where the l-loop term is given by:

$$(2\pi)^3\,\delta_D^{(3)}(\mathbf{k}+\mathbf{k}')P_{ab}^{(l)}(\mathbf{k}, u) = \sum_{m=1}^{2l+1} \left\langle \Psi_a^{(m)}(\mathbf{k}, u)\Psi_b^{(2l+2-m)}(\mathbf{k}', u)\right\rangle . \tag{2.58}$$

The l-loop contribution to the power spectrum correlates $2l+2$ initial fields — correlations of odd n initial fields vanishing following the above Wick theorem.

A diagrammatic representation of this equation is given in Figure 2.3. For each $m \in [1, 2l+1]$, trees of $\Psi_a^{(m)}(\mathbf{k}, u)$ and $\Psi_b^{(2l+2-m)}(\mathbf{k}', u)$ are positioned opposite to each other, with their initial fields face-to-face. Then, the initial fields ($\phi_c(\mathbf{k})$, $\phi_d(\mathbf{k}')$) are paired and replaced by their power spectrum $(2\pi)^3\,\delta_D^{(3)}(\mathbf{k}+\mathbf{k}')\varepsilon_c\varepsilon_d P_0(\mathbf{k})$, symbolised by a cross in Figure 2.3. Diagrams are weighted by the factor coming from the different tree combinations of Figure 2.2, and by the number of pair combinations.

The first diagram of Figure 2.3 gives the power spectrum at tree level:

$$P_{ab}^{(0)}(\mathbf{k}) = g_{ac}(u)g_{bd}(u)\varepsilon_c\varepsilon_d P_0(\mathbf{k}). \tag{2.59}$$

The second diagram is a 1-loop term:

$$P_{ab}^{(1)}(\mathbf{k}) \ni \int d^3q dv_1 dv_2 g_{ac}(u-v_1)\gamma_{cde}(\mathbf{k}, \mathbf{q}, \mathbf{k}-\mathbf{q})g_{df}(v_1)\varepsilon_f g_{eg}(v_1)\varepsilon_g P_0(q)P_0(|\mathbf{k}-\mathbf{q}|) \\ \times\, g_{bh}(u-v_2)\gamma_{hij}(-\mathbf{k}, -\mathbf{q}, \mathbf{q}-\mathbf{k})g_{ik}(v_2)\varepsilon_k g_{jl}(v_2)\varepsilon_l. \tag{2.60}$$

In fact, loop corrections of standard perturbation theory (SPT) which we used so far are ill-behaved on small scales: when approaching non-linear scales, high loop contributions are of the same order of magnitude with alternative signs. To remedy this lack of convergence, resummation schemes have been proposed.

2.1.6 Gamma expansion

Gamma expansion is an alternative to the SPT treatment described above, which allows loop corrections to be resummed. We recall that the linear propagator g_{ab} links the linearly time-evolved field to its initial condition. Let us introduce its generalisation to the non-linear case, the 2-point propagator G_{ab}, defined as (Crocce and Scoccimarro, 2006a):

$$\left\langle \frac{\delta \Psi_a(\mathbf{k}, u_f)}{\delta \phi_b(\mathbf{k}')} \right\rangle = \delta_D^{(3)}(\mathbf{k} - \mathbf{k}') G_{ab}(\mathbf{k}, u_f, u_i), \tag{2.61}$$

which relates the non-linear evolved field $\Psi_a(\mathbf{k}, u_f)$ to the initial conditions $\phi_b(\mathbf{k}_1)$ at u_i. The condition $\delta_D^{(3)}(\mathbf{k} - \mathbf{k}')$ stems from statistical homogeneity. $G_{ab}(\mathbf{k}, u_f, u_i)$ can be expanded in series of loop contributions:

$$G_{ab}(\mathbf{k}, u_f, u_i) = g_{ab}(u_f - u_i) + \sum_{l=1}^{+\infty} G_{ab}^{(l)}(\mathbf{k}, u_f, u_i), \tag{2.62}$$

which can be calculated within SPT. It is expected that on large scales $G_{ab}(\mathbf{k} \to 0, u_f, u_i) = \delta_{ab} g_{ab}(u_f - u_i)$ while on small scales the evolved density field $\Psi_a(\mathbf{k}, u_f)$ decorrelates from the initial field and then $G_{ab}(\mathbf{k} \to +\infty, u_f, u_i) = 0$. In fact, by summing the dominant diagrams (with loops connected to a *principal path*) of each loop l, $G_{ab}(\mathbf{k}, u)$ can be shown to be exponentially suppressed (Crocce and Scoccimarro, 2006a):

$$G_{ab}(\mathbf{k}, u) \xrightarrow[k \to +\infty]{} g_{ab}(u_f - u_i) \exp\left(-\frac{1}{2} k^2 \sigma_d^2(u_f, u_i) \right) \tag{2.63}$$

with $\sigma_d^2(u_f, u_i)$ the variance of the velocity $\mathbf{u} / (f\mathcal{H})$ along one direction, integrated from time u_i to u_f; in linear theory:

$$\sigma_d^2(u_f, u_i) = \left(e^{u_f} - e^{u_i}\right)^2 \sigma_d^2 = \frac{1}{3} \left(e^{u_f} - e^{u_i}\right)^2 \int \frac{d^3 q}{(2\pi)^3} \frac{P_0(q)}{q^2}. \tag{2.64}$$

In the following, we attempt to give a flavour of the Renormalised Perturbation Theory (RPT) presented in Crocce and Scoccimarro (2006b,a) and Regularised Perturbation Theory of Bernardeau et al. (2012); Taruya et al. (2012), without providing formal demonstration.

Let us consider an Einstein de Sitter Universe, $u = \ln a$, and focus on $(a, b) \in \{(1,1), (1,2)\}$, i.e. on the response of the evolved density field Ψ_1 to the initial conditions ϕ_1 and ϕ_2 (following arguments are similar for $(a, b) \in \{(2,1), (2,2)\}$). We recall Eq. (2.41):

$$\begin{aligned} g_{11}(a) &= \frac{3}{5} a + \frac{2}{5} a^{-3/2} \\ g_{12}(a) &= \frac{2}{5} a - \frac{2}{5} a^{-3/2}. \end{aligned} \tag{2.65}$$

The 1-loop contribution to G_{ab} was calculated in Crocce and Scoccimarro (2006a):

$$\begin{aligned} G_{11}^{(1)}(k, a) &= \frac{3}{5} \left[\alpha(a) f(k) - \beta(a) i(k)\right] + \frac{2}{5} \left[\delta(a) g(k) - \gamma(a) h(k)\right] \\ G_{12}^{(1)}(k, a) &= \frac{2}{5} \left[\alpha(a) f(k) - \beta(a) h(k)\right] - \frac{2}{5} \left[\delta(a) f(k) - \gamma(a) h(k)\right] \end{aligned} \tag{2.66}$$

where:

$$\begin{aligned} \alpha(a) &= a^3 - \frac{7}{5} a^2 + \frac{2}{5} a^{-1/2} & \beta(a) &= \frac{3}{5} a^2 - a + \frac{2}{5} a^{-1/2} \\ \gamma(a) &= \frac{2}{5} a^2 - a^{1/2} + \frac{3}{5} a^{-1/2} & \delta(a) &= \frac{3}{5} a^2 - \frac{7}{5} a^{-1/2} + a^{-3/2}. \end{aligned} \tag{2.67}$$

The four functions of k ($f(k)$, $g(k)$, $h(k)$, $i(k)$) in Eq. (2.66) are given in Crocce and Scoccimarro (2006a). They all behave as $-k^2\sigma_d^2/2$ when k is large, and then:

$$G_{ab}^{(1)}(k,u_f,u_i) \xrightarrow[k\to+\infty]{} -\frac{1}{2}k^2\sigma_d^2(u_f,u_i)g_{ab}(u_f-u_i). \tag{2.68}$$

Similarly, higher order loop corrections are expected to behave as $\to (-1)^l/\left(2^l l!\right) k^{2l}\sigma_d^{2l}(u_f,u_i)$ at large k, in agreement with Eq. (2.63).

Crocce and Scoccimarro (2006a) noticed that the expression of G_{ab} up to 1-loop (sum of Eq. (2.65) and Eq. (2.66)) can be regarded as a first order expansion of:

$$\begin{aligned} G_{11}^{\rm RPT}(k,a) &= \frac{3}{5}ae^{\alpha_g(a)f(k)-\beta_g(a)i(k)} + \frac{2}{5}a^{-3/2}e^{\delta_d(a)g(k)-\gamma_d(a)h(k)} \\ G_{12}^{\rm RPT}(k,a) &= \frac{2}{5}ae^{\alpha_g(a)f(k)-\beta_g(a)h(k)} - \frac{2}{5}a^{-3/2}e^{\delta_d(a)f(k)-\gamma_d(a)h(k)} \end{aligned} \tag{2.69}$$

where contributions are split between the leading growing mode $\propto a$ (indices g) and the leading decaying one $\propto a^{-3/2}$ (indices d) and with:

$$\begin{aligned} \alpha_g(a) &= \alpha(a)a^{-1} \qquad & \beta_g(a) &= \beta(a)a^{-1} \\ \delta_d(a) &= \delta(a)a^{3/2} \qquad & \gamma_d(a) &= \gamma(a)a^{3/2}. \end{aligned} \tag{2.70}$$

These are the expressions of $G_{ab}(k,a)$ in the RPT formalism. Note that there are no free parameters. The result in any ΛCDM scenario is obtained by replacing a by $D_+ = e^u$. The 2-point propagator can be computed in simulations by cross-correlating final with initial density fields. The resummation presented above was shown to be in very good agreement with such estimates from simulations.

The $(p+1)$-point propagator $\Gamma_{ac_1\cdots c_p}$, introduced in Bernardeau et al. (2008), is a natural extension of the 2-point propagator:

$$\frac{1}{p!}\left\langle \frac{\delta^p\Psi_a(\mathbf{k},u_f)}{\delta\phi_{c_1}(\mathbf{k}_1)\cdots\delta\phi_{c_p}(\mathbf{k}_p)}\right\rangle = \frac{\delta_D^{(3)}(\mathbf{k}-\mathbf{k}_1-\cdots\mathbf{k}_p)}{(2\pi)^{3(p-1)}}\Gamma_{ac_1\cdots c_p}(\mathbf{k}_1\cdots\mathbf{k}_p,u_f,u_i) \tag{2.71}$$

It was shown in Bernardeau et al. (2008) that these propagators can be used to express the non-linear power spectrum:

$$\begin{aligned} P_{ab}(k,u_f) = \sum_{p=1}^{+\infty} p! \int \frac{d^3q_1\cdots d^3q_p}{(2\pi)^{3(p-1)}}\delta_D^{(3)}(\mathbf{k}-\mathbf{q}_1-\cdots\mathbf{q}_p)\Gamma_{a,(p)}(\mathbf{q}_1\cdots\mathbf{q}_p,u_f,u_i) \\ \Gamma_{b,(p)}(\mathbf{q}_1\cdots\mathbf{q}_p,u_f,u_i)P_0(q_1)\cdots P_0(q_p) \end{aligned} \tag{2.72}$$

where we introduced the short-hand notation: $\Gamma_{a,(p)} = \Gamma_{ac_1\cdots c_p}\varepsilon_{c_1}\cdots\varepsilon_{c_p}$. The expression above remains valid under any initial conditions, providing $P_0(q_i)$ is replaced by the power spectrum of the initial fields ϕ_{a_i} and ϕ_{b_i} at q_i. The diagrammatic representation of the power spectrum up to 2-loop order is depicted in Figure 2.4.

Just as for the 2-point propagator, the general multi-point propagator is expanded following:

$$\Gamma_{ab_1\cdots b_p}(\mathbf{k}_1,\cdots\mathbf{k}_p,u_f,u_i) = \sum_{l=1}^{+\infty}\Gamma_{ab_1\cdots b_p}^{(l)}(\mathbf{k}_1,\cdots\mathbf{k}_p,u_f,u_i). \tag{2.73}$$

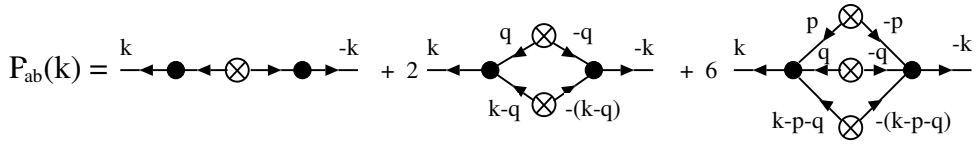

Figure 2.4 – The diagrammatic representation of the power spectrum up to 2-loop using the gamma expansion. Non-linear multipoint propagators, whose resummation is symbolised by the black dots, are linked together at the crosses where the initial power spectrum P_0 is injected. Taken from Taruya et al. (2012).

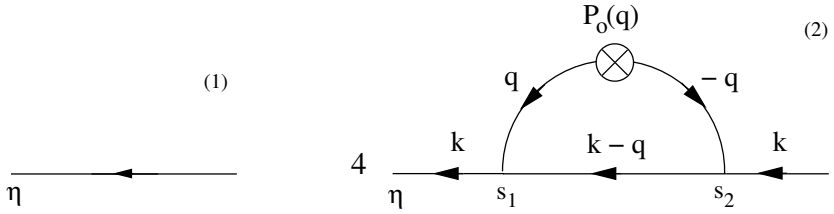

Figure 2.5 – The diagrammatic representation of the 2-point propagator up to 1-loop. Taken from Crocce and Scoccimarro (2006a).

Again, by summing the dominant diagrams (with loops connected to a *principal tree*) of each loop l, the general multi-point propagator is shown to be exponentially suppressed at high k (Bernardeau et al., 2008):

$$\Gamma_{ab_1\cdots b_p}(\mathbf{k}_1,\cdots\mathbf{k}_p,u_f,u_i)\xrightarrow[k\to+\infty]{}\exp\left(-\frac{1}{2}k^2\sigma_d^2(u_f,u_i)\right)\Gamma^{(0)}_{ab_1\cdots b_p}(\mathbf{k}_1,\cdots\mathbf{k}_p,u_f,u_i) \tag{2.74}$$

where $k=|\mathbf{k}_1+\cdots\mathbf{k}_p|$.

In general, $\Gamma_{ab_1\cdots b_p}$ is expected to be well-described at low k by a truncation of Eq. (2.73), using PT to compute $\Gamma^{(l)}_{ab_1\cdots b_p}$, and at high k by Eq. (2.74). The RegPT scheme aims at interpolating between the two. Let us consider as an example the late time behaviour of the two point propagator G_{ab}, in an Einstein de Sitter Universe ($u=\ln a$) and with $\varepsilon_a=\varepsilon_b=1$. The response of the evolved density field to the initial one δ_0 is, up to 1-loop:

$$G_1(k,a)=G_{11}(k,a)+G_{12}(k,a)=g_{11}(a)+g_{12}(a)+G^{(1)}_{11}(k,a)+G^{(1)}_{12}(k,a)\xrightarrow[a\to+\infty]{}a+a^3f(k), \tag{2.75}$$

which, within RPT, is resummed following (see Eq. (2.69)):

$$G_1^{\rm RPT}(k,a)=a\exp\left[a^2f(k)\right]. \tag{2.76}$$

Rather, in the RegPT scheme:

$$G_1^{\rm RegPT}(k,a)=a\left(1+a^2f(k)+\frac{1}{2}k^2\sigma_d^2a^2\right)e^{-\frac{1}{2}k^2\sigma_d^2a^2}. \tag{2.77}$$

This amounts to subtracting the high k limit $-k^2\sigma_d^2a^2/2$ from the exponent in (2.76), which is then re-expanded at first order. In other words, the RegPT scheme interpolates between the low k behaviour where standard loop corrections are expected to be accurate and the high k exponential damping.

This scheme is straightforwardly extended to any multipoint propagator and to any order in loop corrections, e.g., up to 2-loop:

$$\Gamma^{\rm RegPT}_{ac_1\cdots c_p}(u_f,u_i) = \left\{\left[1+\frac{1}{2}k^2\sigma_d^2(u_f,u_i)+\frac{1}{2}\left(\frac{1}{2}k^2\sigma_d^2(u_f,u_i)\right)^2\right]\Gamma^{(0)}_{ac_1\cdots c_p}(u_f,u_i)+\right.$$
$$\left.\left[1+\frac{1}{2}k^2\sigma_d^2(u_f,u_i)\right]\Gamma^{(1)}_{ac_1\cdots c_p}(u_f,u_i)+\Gamma^{(2)}_{ac_1\cdots c_p}(u_f,u_i)\right\}\exp\left[-\frac{1}{2}k^2\sigma_d^2(u_f,u_i)\right] \quad (2.78)$$

where we dropped the $\mathbf{k}_1\cdots\mathbf{k}_p$ variables for conciseness.

The difference between the RPT and RegPT schemes for G_{ab} only matters when $k \gtrsim 1\,h\,\mathrm{Mpc}^{-1}$, where the exponential damping is very strong (Bernardeau et al., 2012). Hence, both RPT and RegPT schemes reproduce well G_{ab} computed in numerical simulations.

The RegPT predictions for the 3-point propagator Γ_{111} are shown in Figure 2.6, for different values of wavenumbers k_1, k_2, k_3, with $\mathbf{k}_3$ the outgoing momentum, $\mathbf{k}_3 = \mathbf{k}_1 + \mathbf{k}_2$. Measurements from numerical simulations are in very good agreement with the 1-loop predictions (when including binning effects in the measurements of Γ_{111}).

2.1.7 The RegPT prescription for the power spectrum

From now on we set initial conditions in the growing mode $\varepsilon_a = \varepsilon_b = 1$ in the limit $u_i \to -\infty$. Hence, only the fastest growing mode dominates. In this case, the tree-order propagator reads (Taruya et al., 2012):

$$\Gamma^{(0)}_{a,(p)}(\mathbf{k}_1,\cdots\mathbf{k}_p,u) = e^{pu}F^{(p)}_{a,s}(\mathbf{k}_1,\cdots\mathbf{k}_p) \quad (2.79)$$

and at l-loop order:

$$\Gamma^{(l)}_{a,(p)}(\mathbf{k}_1,\cdots\mathbf{k}_l,u) = e^{(2l+p)u}c^{(p),(l)}\int\frac{d^3q_1\cdots d^3q_p}{(2\pi)^{3l}}F^{(2l+p)}_{a,s}(\mathbf{q}_1,-\mathbf{q}_1,\cdots\mathbf{q}_l,-\mathbf{q}_l,\mathbf{k}_1,\cdots\mathbf{k}_p)P_0(q_1)\cdots P_0(q_l) \quad (2.80)$$

with $c^{(p),(l)} = \binom{2l+p}{p}(2l-1)!!$ and $F^{(p)}_{a,s}$ the symmetrised standard PT kernels. The regularised propagator $\Gamma^{\rm RegPT}_{a,(p)}$ is then given by Eq. (2.78) at 2-loop, and by similar formulae at l-loop. In the latter formula, since the fastest growing mode dominates, $\sigma_d^2(u) = e^{2u}\sigma_d^2$ (see Eq. (2.64)). We recall that σ_d emerges from the resummation of the high k end. Hence, following Bernardeau et al. (2012), Taruya et al. (2012) argue that the integration upper bound $k_\Lambda(k)$ in:

$$\sigma_d^2(k) = \int_0^{k_\Lambda(k)}\frac{dq}{6\pi^2}P_0(q) \quad (2.81)$$

should depend on k. Taruya et al. (2012) found that $k_\Lambda(k) = k/2$ reproduces well the power spectrum measurements from numerical simulations.

Obtaining the 2-loop power spectrum is now straightforward using Eq. (2.72):

$$P_{ab}(k,u) = \Gamma^{\rm RegPT}_{a,(1)}(k,u)\Gamma^{\rm RegPT}_{b,(1)}(k,u)P_0(k) \quad (2.82)$$
$$+2\int\frac{d^3q}{(2\pi)^3}\Gamma^{\rm RegPT}_{a,(2)}(\mathbf{q},\mathbf{k}-\mathbf{q},u)\Gamma^{\rm RegPT}_{b,(2)}(\mathbf{q},\mathbf{k}-\mathbf{q},u)P_0(q)P_0(|\mathbf{k}-\mathbf{q}|) \quad (2.83)$$
$$+6\int\frac{d^3pd^3q}{(2\pi)^6}\Gamma^{\rm RegPT}_{a,(3)}(\mathbf{p},\mathbf{q},\mathbf{k}-\mathbf{p}-\mathbf{q},u)\Gamma^{\rm RegPT}_{b,(3)}(\mathbf{p},\mathbf{q},\mathbf{k}-\mathbf{p}-\mathbf{q},u)P_0(p)P_0(q)P_0(|\mathbf{k}-\mathbf{p}-\mathbf{q}|), \quad (2.84)$$

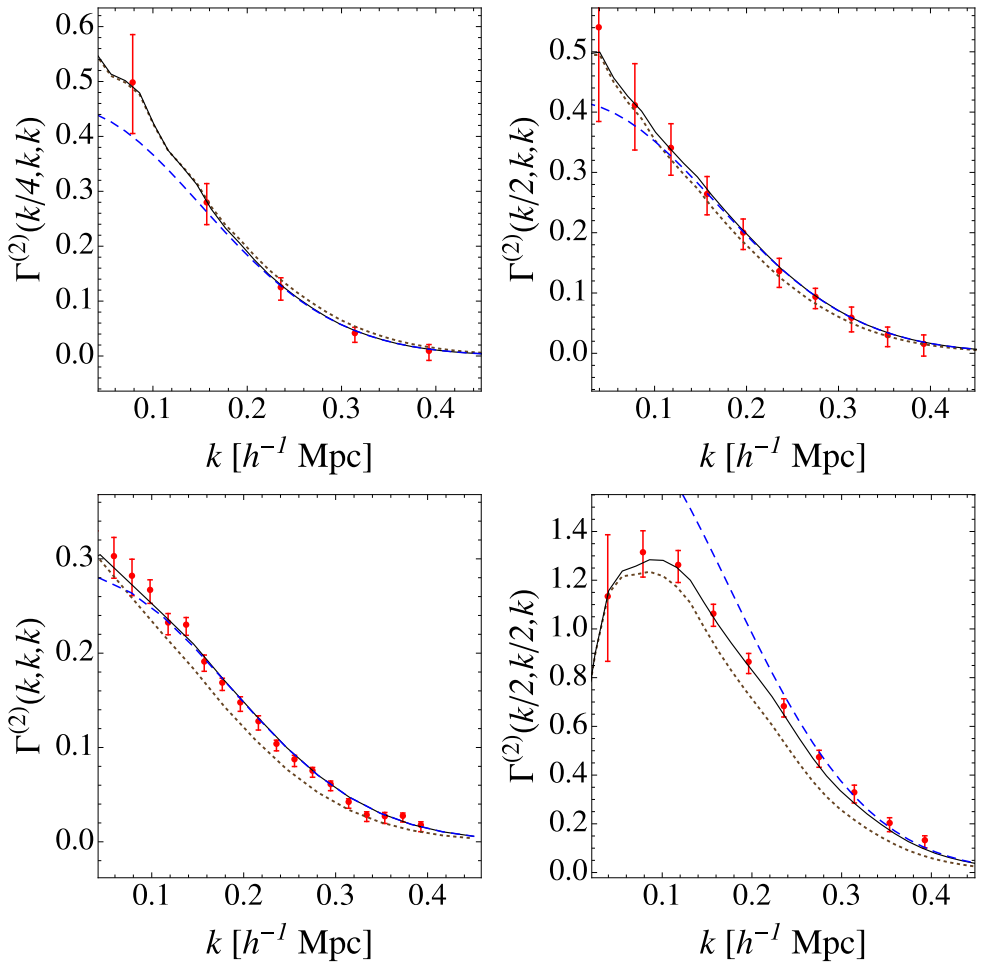

Figure 2.6 – Comparison of the 1-loop 3-point propagator (continuous black lines) against numerical simulations (points with error bars) in four configurations (upper left: squeezed $k_1 = k/4; k_2 = k_3 = k$, upper right: elongated $k_1 = k/2; k_2 = k_3 = k$, bottom left: equilateral $k_1 = k_2 = k_3 = k$, bottom right: colinear $k_1 = k_2 = k/2; k_3 = k$). The dashed blue line is the prediction which takes no account of the binning effect in the measurement of Γ_{111}. The dotted line is the prediction at tree order. Taken from Bernardeau et al. (2012).

where $\Gamma^{\mathrm{RegPT}}_{a,(1)}$, $\Gamma^{\mathrm{RegPT}}_{a,(2)}$ and $\Gamma^{\mathrm{RegPT}}_{a,(3)}$ are computed at 2-loop, 1-loop and tree order, respectively. Their respective contributions and their sum are displayed in Figure 2.7. One would note that higher order contributions are consistently shifted to higher k.

Numerically speaking, $\Gamma^{(1)}_{a,(1)}$, $\Gamma^{(0)}_{a,(2)}$, $\Gamma^{(1)}_{a,(2)}$ $\Gamma^{(0)}_{a,(3)}$, entering $\Gamma^{\mathrm{RegPT}}_{a,(1)}$, $\Gamma^{\mathrm{RegPT}}_{a,(2)}$ and $\Gamma^{\mathrm{RegPT}}_{a,(3)}$ are calculated without approximation and a fitting function is used for $\Gamma^{(2)}_{a,(1)}$. The low dimensional (≤ 3) integrals are computed using the Gauss-Legendre method. $\Gamma^{(3)}_{a,(1)}$ is numerically evaluated with a Monte Carlo technique, using the library `CUBA` (Hahn, 2005). Though a `Fortran` code was available [1], we recoded the RegPT prescription in `C`, wrapped in `Python`, allowing for parallelisation and tuning of $k_\Lambda(k)$, integration ranges and precision in package called `pyregpt` [2]. In practice, given our measurement precision, we kept the same integration ranges and $k_\Lambda(k)$ as proposed in Taruya et al. (2012). We also utilised this framework to implement bias and RSD terms (as described in next sections).

1. at http://www2.yukawa.kyoto-u.ac.jp/~atsushi.taruya/regpt_code.html
2. at https://github.com/adematti/pyregpt

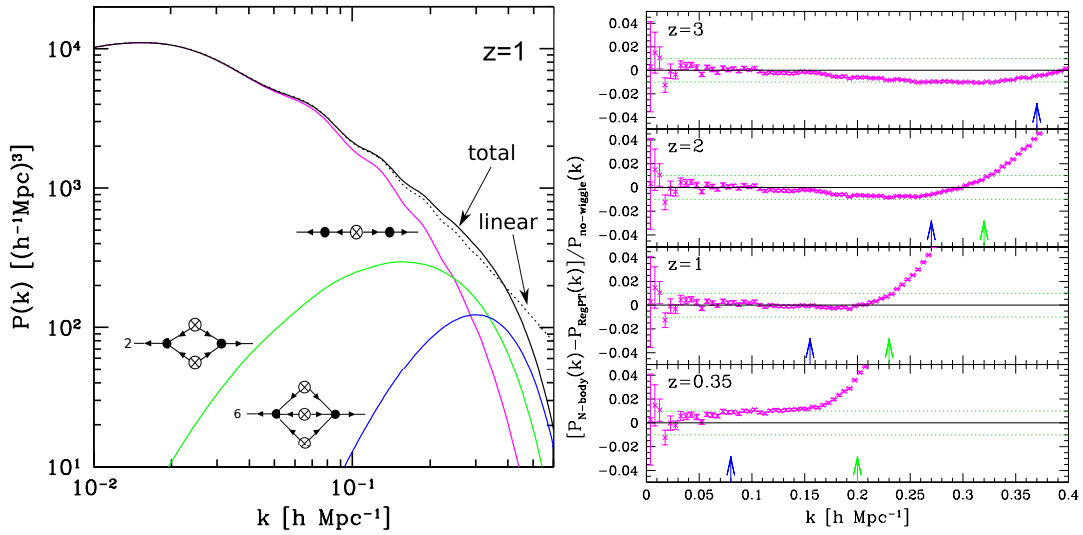

Figure 2.7 – Left: contributions of the first three terms of the gamma expansion (Eq. (2.82) in pink, Eq. (2.83) in green and Eq. (2.84) in blue) to the total power spectrum (continuous line) at 2-loop, at $z = 1$. For reference, the linear power spectrum $e^{2u}P_0(k)$ is shown in dashed lines. Right: relative difference between the power spectrum measured from N-body simulations to the RegPT prediction. The blue and green arrows show the 1% validity range achieved at 2-loop order for the Lagrangian resummation theory (LRT) and closure theory, respectively. At $z = 3$, the validity range of closure theory is larger than the maximum wavenumber of the figure ($k = 0.4\,h\,\mathrm{Mpc}^{-1}$). Taken from Taruya et al. (2012).

Comparison of the RegPT prediction to numerical simulations is shown in Figure 2.7 (right panel). The RegPT prediction is in general good agreement with simulations, with a percent accuracy achieved up to $k \simeq 0.15,\ 0.23,\ 0.33\,h\,\mathrm{Mpc}^{-1}$ at $z = 0.35,\ 1,\ 2$ respectively. For comparison, pure linear PT breaks down at $k \simeq 0.1\,h\,\mathrm{Mpc}^{-1}$ at $z = 1$. One would note however that the RegPT power spectrum is slightly underestimated for the lower redshift $z = 0.35$, attributed to a poor convergence of $\Gamma^{\mathrm{RegPT}}_{(1)}$. At high redshift ($z = 3$), the power spectrum seems overestimated around $k \simeq 0.3\,h\,\mathrm{Mpc}^{-1}$ but Taruya et al. (2012) attribute this effect to the simulation. The agreement between the prediction for the correlation function, given by the Hankel transform:

$$\xi(r) = \int \frac{dk k^2}{2\pi^2} P(k) \frac{\sin(kr)}{kr} \tag{2.85}$$

and numerical simulations was found to be even better in Taruya et al. (2012). RegPT only breaks down around $s = 25\,\mathrm{Mpc}\,h^{-1}$ at $z = 0.35$ and $s = 20\,\mathrm{Mpc}\,h^{-1}$ at $z = 1$.

Also indicated in Figure 2.7 (right panel) is the validity range (for 1% precision) of two other resummation schemes, Lagrangian resummation theory (LRT, Matsubara, 2008; Okamura et al., 2011) and closure theory (Taruya and Hiramatsu, 2008; Taruya et al., 2009) at 2-loop order. Overall, RegPT performs at least as well as the two other schemes. An alternative prescription for the power spectrum relying on gamma expansion is called MPTbreeze (Crocce et al., 2012). It extrapolates the result obtained at 1-loop in Eq. (2.76) to the multipoint case. The calculation in this scheme is faster, and less sensitive to the UV cutoff $k_\Lambda(k)$, but less accurate on intermediate scales.

Note that a faster RegPT scheme, RegPT-fast, was developed. In this scheme, the evolved power spectrum is Taylor expanded as a function of the initial one (using the explicit dependence in Eq. (2.72)), around a set of fiducial cosmologies. Then, predictions for any cosmology close enough to the fiducial one can be obtained within seconds with 1-dimensional integrals. Since we kept the fiducial cosmology fixed, we only used the standard RegPT scheme.

Finally, let us recall that we have neglected the anisotropic stress tensor σ_{ij} in the Euler equation (2.10) to perform standard PT and further RegPT. Another approach called effective field theory of large scales structures (EFTofLSS, Carrasco et al. 2012), does not make this assumption. Perturbations are split into large and small wavelength modes, the latter contributing to σ_{ij}. These small scale fluctuations source counterterms used to regularise the perturbation theory calculations which involve integrals of the power spectrum up to non-linear scales. Note however that such a procedure requires measuring parameters encapsulating small scale physics, from e.g. numerical simulations.

2.2 Numerical simulations

Theoretical calculations are key to understand the non-linear formation of structure. Their reliability and validity range is tested against numerical simulations, which provide an accurate picture of non-linear structure formation.

We will first review N-body simulations, evolving collisionless dark matter through gravity only. The connection between the outcome of these simulations and galaxy distribution will then be discussed. This section is partly inspired from Codis-Decara (2015); Wechsler and Tinker (2018).

2.2.1 N-body simulations

In practice, cosmological N-body simulations do not account for collisions between particles. This in contrast to collisional N-body codes, which allow momentum exchange between particles, and are required to described small systems, such as stellar clusters and galaxy centres. The Vlasov-Poisson equations (2.4) and (2.5) describe the evolution of the phase-space distribution function $f(\mathbf{x}, \mathbf{u}, t)$, which operates in a 6-dimensional space (plus time). Solving it directly using a 6-dimensional grid would be computationally very costly and inefficient, as f gets more and more localised with non-linear structure formation. Therefore, in N-body simulations, the distribution function $f(\mathbf{x}, \mathbf{u}, t)$ is sampled by a great number (typically 10^{10-12} today) of particles.

2.2.1.1 Initial conditions

N-body simulations evolve particles through gravity, starting from initial conditions that are specified at a redshift $z \sim 100$. Various schemes exist to generate the initial distribution of particles. The first is random uniform sampling, but it would generate undesirable structure formation from sampling noise even when no other perturbations are imposed. The second solution is a regular grid of particles, but it leads to preferred directions and scales (the initial particle separations). To overcome these issues, White (1994) suggested the glass-like distribution. Starting from random uniform sampling, particles are evolved through the gravity solver (described hereafter), taking the opposite of the gravitational force, such that particles tend to repulse each other. This reverse evolution typically takes 150 integration steps and spans

an expansion factor of 10^6. It yields a particle distribution that does not show any preferred direction nor scale and does not lead to structure formation if subsequently evolved through gravity for expansion factors as large as 30. The second step consists in imprinting the desired initial fluctuations in this particle distribution.

A density contrast field δ is generated following a Gaussian distribution, with variance the linear power spectrum at the initial redshift, as provided by Boltzmann codes such as CAMB (Lewis et al., 2000; Lewis and Challinor, 2011) or CLASS (Blas et al., 2011). Based on the generated density field, particle positions are shifted and are assigned velocities. This step is typically performed using the Zel'dovich approximation or second order Lagrangian perturbation theory, which allows starting the simulation at lower redshift. Then, multiple techniques exist to evolve the particles through gravitational interaction.

2.2.1.2 The particle-particle method

After the pioneering work of Holmberg (1941) modelling nebulae with light bulbs, the particle-particle (PP) method was used as soon as the early sixties, with up to 100 particles (e.g. Aarseth, 1963). In this method, the peculiar velocity of each particle i is calculated using:

$$\frac{d\mathbf{u}^{(i)}}{dt} + H\mathbf{u}^{(i)} = \frac{1}{m^{(i)}} \sum_{j \neq i} \mathbf{F}_j^{(i)} \tag{2.86}$$

where the gravitational force of particle j on i is given by:

$$\mathbf{F}_j^{(i)} = G_\mathrm{N} m^{(i)} m^{(j)} \frac{\mathbf{r}^{(j)} - \mathbf{r}^{(i)}}{\left[\left|\mathbf{r}^{(j)} - \mathbf{r}^{(i)}\right|^2 + \epsilon^2\right]^{3/2}}. \tag{2.87}$$

ϵ is a small, softening parameter. Here we have considered the standard Plummer softening (Aarseth, 1963), but other kernels exist (Dehnen, 2001), which can be dynamically adapted. Since particles are considered as a (necessarily incomplete) Monte-Carlo sampling of the distribution function, softening corresponds to the assumption that the mass density is smooth on small scales, i.e. it reduces shot noise. In practice, softening avoids divergence at null separation and thus eases numerical integration, allowing for an increase in the number of particles by a few orders of magnitude (Dehnen, 2001). Of course, ϵ introduces a bias with respect to pure Newtonian dynamics. An optimum can be found between shot noise and bias.

The accuracy of the PP method depends on the (adaptive) time step. Note also that this method scales as $\mathcal{O}\left(N^2\right)$, with N the number of particles. Hence, faster techniques have been developed.

2.2.1.3 Particle-mesh methods

In the particle-mesh (PM) method, the particle density is evaluated on a grid using an interpolation kernel, typically cloud-in-cell or triangular shaped cloud (Eastwood and Hockney (1974), see Section 4.1.7). The gravitational potential is computed with the Poisson equation (2.5), typically in Fourier space, using Fast Fourier Transforms (FFT). The potential is thus softened at the typical grid cell size. Particle motions are computed from this gravitational potential, using the same interpolation kernel. This method scales as FFT algorithms, $\mathcal{O}\left(N_m \log N_m\right)$ with N_m the total mesh size, plus the interpolation with scales as $\mathcal{O}\left(N\right)$ (with N the number of particles). Hence, this method is fast, but inaccurate below distances of several

cell sizes. To increase the small scale precision, one can resort to adaptive mesh refinement (AMR) methods, which dynamically increase the grid resolution in regions of high density (or regions where the coarser grid leads to too large errors). Another solution is to enhance PM with PP at small scales (i.e. of the order of a few cell sizes), leading to hybrid P^3M codes. However, these codes are limited by the integration in regions of very dense clustering, which can be solved by using AMR — leading to AP^3M codes (Couchman, 1991).

2.2.1.4 Tree codes

In tree codes (PPTree) particles are grouped together into smaller and smaller (in size) cells, following a tree structure — which does not need to be recomputed at each time step. Then, the gravitational potential of groups of particles is expanded into multipoles around the group centre of mass. This multipole expansion is used to compute the force applied to another group of particles. Smaller and smaller cells are considered until the opening angle (size over distance) is less than a threshold parameter. The complexity of such an algorithm is $\mathcal{O}(N \log N)$. These codes (an example of which is Dehnen 2000) can be coupled to PM codes. Such hybrid Tree-PM codes are the most efficient algorithms for cosmological purposes (e.g. GADGET-2, Springel, 2005).

Note however that numerical simulations solving the Vlasov-Poisson equations in phase space, hence without resorting to particle sampling, have been recently developed (Hahn and Angulo, 2016; Sousbie and Colombi, 2016). These techniques eliminate the sampling noise and force softening of the N-body simulations described previously.

2.2.1.5 N-body simulations used in the analysis of the eBOSS ELG sample

In this manuscript, we make use of two N-body simulations. We first consider the MultiDark simulation MDPL2 (Klypin et al., 2016), of volume $1\,\mathrm{Gpc}^3\,h^{-3}$, using $3,840^3$ dark matter particles of mass $1.51 \times 10^9\,\mathrm{M}_\odot\,h^{-1}$. The GADGET-2 code was run with the flat ΛCDM cosmology [3]:

$$h = 0.6777,\ \Omega_{m,0} = 0.307115,\ \Omega_{b,0} = 0.048206,\\ \sigma_{8,0} = 0.8228,\ n_s = 0.9611. \tag{2.88}$$

The simulation was started at $z = 100$ from Gaussian initial conditions and with velocities assigned through the Zel'dovich approximation.

As the MDPL2 simulation lacked of cosmological volume, we then focused on the OuterRim simulation (Heitmann et al., 2019) of volume $27\,\mathrm{Gpc}^3\,h^{-3}$ and $10\,240^3$ dark matter particles of mass $1.85 \times 10^9\,\mathrm{M}_\odot\,h^{-1}$. The simulation was run with the HACC (Hardware/Hybrid Accelerated Cosmology Code), described in Habib et al. (2016). This code includes an architecture-independent PM method for large scales (covering 4 orders of magnitude in scale) and an architecture-tunable part for the more time-consuming resolution of small scales (2 orders of magnitude), based on the PP or PPTree method. The simulation was run with the following flat ΛCDM cosmology:

$$h = 0.71,\ \omega_{c,0} = 0.1109,\ \omega_{b,0} = 0.02258,\\ \sigma_{8,0} = 0.8,\ n_s = 0.963. \tag{2.89}$$

3. https://www.cosmosim.org/cms/simulations/mdpl2/

The simulation was started at $z = 200$ from Gaussian initial conditions and with velocities assigned through the Zel'dovich approximation. The $z = 0$ snapshot of this simulation is shown in Figure 2.8.

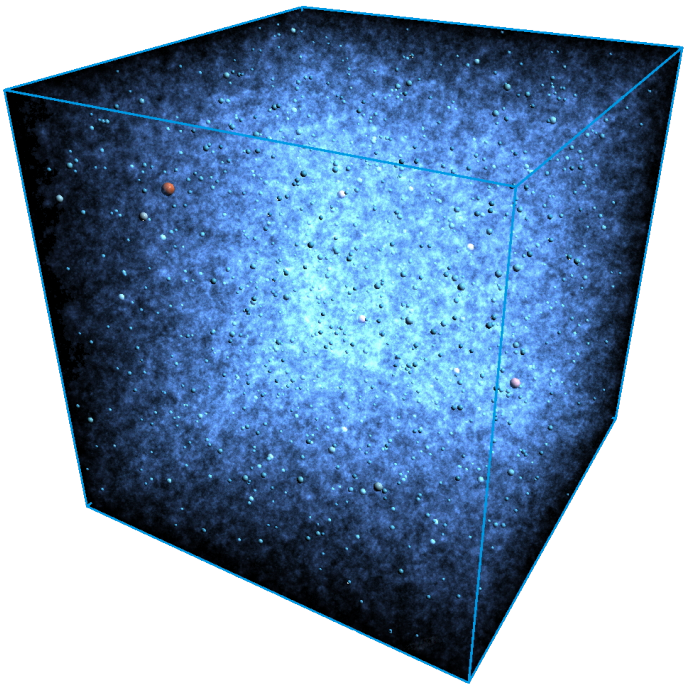

Figure 2.8 – Snapshot of the OuterRim simulation at $z = 0$. Halos above $1.8 \times 10^{15}\,\mathrm{M}_\odot$ are shown as spheres (red means more massive), while halos down to $\sim 5 \times 10^{10}\,\mathrm{M}_\odot$ are represented as blue splats. Taken from Heitmann et al. (2019).

2.2.2 Connection to tracer density

The outcome of a N-body simulation is a collection of dark matter particle positions and velocities, from the initial redshift to the final one. White and Rees (1978) argued that galaxy formation can be described as a two-step process. First, dark matter forms halos, defined as overdense gravitationally bound systems. This process can be understood within the spherical collapse model which we briefly review hereafter. Then, baryons aggregate into denser units due to gas dynamical dissipative processes and form galaxies. Hence, halos appear as the first building block connecting the dark matter density field to galaxies. At small scales, baryonic matter cannot be considered a collisionless fluid, hence requiring hydrodynamical codes to enhance N-body simulations with baryonic processes. However, hydrodynamical simulations are computationally costly. Complementary approaches include semi-analytical models or prescriptions (abundance matching or halo occupation distribution — HOD) to populate dark matter halos with galaxies. We review these different techniques in the following.

2.2.2.1 Spherical collapse model

The spherical collapse model describes the evolution of an initially spherical overdensity through gravitation. Applying Birkhoff theorem to this initial perturbation, in an Einstein de

Sitter Universe:

$$\ddot{R} = -\frac{G_{\rm N} M}{R^2} \qquad M = \frac{4\pi}{3} R^3 \rho \tag{2.90}$$

where ρ is the density, and $\dot{}$ denotes a derivative with respect to the cosmic time. This equation can be integrated into $E = \dot{R}^2/2 - GM/R$. The system is bound when the energy E is negative, in which case the solution can be parametrised as follows:

$$\begin{aligned} R &= \frac{R_m}{2} \left(1 - \cos\zeta\right) \\ t &= \frac{t_m}{\pi} \left(\zeta - \sin\zeta\right), \end{aligned} \tag{2.91}$$

with $R_m = -G_{\rm N} M/E$ and $t_m = \pi G_{\rm N} M \left(-E\right)^{-3/2} / \sqrt{8}$. The radius grows until $R = R_m$ at $t = t_m$ ($\zeta = \pi$) and then shrinks until $R = 0$ at $t = 2t_m$ ($\zeta = 2\pi$). The density contrast reads:

$$\delta_{sc} = \frac{\rho}{\bar{\rho}} - 1 = \frac{9 M G_{\rm N} t^2}{2R^3} - 1 \tag{2.92}$$

where we used the mean Universe density $\bar{\rho} = 1/\left(6\pi G_{\rm N} t^2\right)$ (see Eq. (1.36) and Table 1.1). The above solution (2.91) can be expanded at linear order ($\zeta \ll 1$) to yield:

$$R^{\rm lin}(t) = \frac{R_m}{4} \left(\frac{6\pi t}{t_m}\right)^{2/3} \left[1 - \frac{1}{20} \left(\frac{6\pi t}{t_m}\right)^{2/3} + o\left[\left(\frac{t}{t_m}\right)^{2/3}\right]\right]. \tag{2.93}$$

The spherical overdensity evolves in three phases:

— growing R: initially ($t \ll t_m$), $\delta_{sc} \propto t^{2/3}$, as expected (from Eq. (2.15) and Table 1.1, $D_+ \propto a \propto t^{2/3}$). When $t = t_m$, R reaches its maximum R_m; at that time $\delta_{sc}(t_m) = 9\pi^2/16 - 1 \simeq 4.55$. Using the expansion at linear order, $\delta_{sc}^{\rm lin}(t_m) = 3/20\,(6\pi)^{2/3} \simeq 1.06$. Since $\dot{R}(t_m) = 0$ the energy is fully potential $E = U(R_m) = -GM/R_m$.

— collapse: the overdensity decouples from the expansion of the Universe and collapses at $t = 2t_m$, when the spherical model predicts an infinite density. Using the expansion at linear order, $\delta_c = \delta_{sc}^{\rm lin}(2t_m) = 3/20\,(12\pi)^{2/3} \simeq 1.686$, which is called the *critical density contrast*. Variations of δ_c with cosmology are rather modest (Bernardeau et al., 2002).

— virialisation: in practice the overdensity does not reach an infinite density but rather achieves an equilibrium, when the total energy $E = K + U = 1/2\dot{R}^2 - GM/R$ is half the potential energy, $E = U(R_{\rm vir})/2$. Since $E = U(R_m)$, the radius at virialisation is $R_{\rm vir} = R_m/2$ and the corresponding overdensity (at $t = 2t_m$) is $\delta_{\rm vir} = (6\pi)^2/2 - 1 \simeq 177$. In a ΛCDM Universe, the expansion rate is increased, therefore the background density is smaller at virialisation, so $\delta_{\rm vir}$ is larger.

We note that the spherical collapse model has interesting cosmological applications. It can for example be used to predict the density distribution within spheres, using large deviation theory (Uhlemann et al., 2016).

2.2.2.2 Dark matter halos

Dark matter forms halos, defined as overdense gravitationally bound systems. One would note however that halos can themselves host other, smaller, gravitationally bound structures evolving in the host halo potential — those are called subhalos. The most physical definition of halo boundaries remains under debate. The virial radius $R_{\rm vir}$ of halos is defined as the radius

of the spherical volume within which the mean halo density is $\Delta_{\rm vir}$ times the matter density at that redshift. The virial mass is then $M_{\rm vir} = 4\pi/3 R_{\rm vir}^3 \Delta_{\rm vir} \bar{\rho}$ with $\bar{\rho}$ the matter density (other definitions rely on the critical density). Different values are taken for $\Delta_{\rm vir}$; in the matter dominated area (Einstein de Sitter Universe), for the spherical collapse, we have just seen that $\Delta_{\rm vir} = 1 + \delta_{\rm vir} \simeq 178$. Note that with this definition, $R_{\rm vir}$ will naturally increase with time (even though the halo size remains static in physical units) due to the decrease of $\bar{\rho}$ (the so-called pseudoevolution, see Diemer et al. 2013). Hence, More et al. (2015) introduced the splashback radius, defined as the apocentre of the first orbit of particles after collapse into the halo; in spherical symmetry, this radius separates infalling from orbiting material.

Halos can be found in dark matter simulations using different techniques. Historically, the first two halo finders were the spherical overdensity (SO) method of Press and Schechter (1974) and the Friends-of-Friends (FOF) algorithm of Davis et al. (1985). The SO method locates density peaks and grows spherical shells until the overdensity drops below a threshold value, typically given by the spherical collapse model. Conversely, the FOF algorithm identifies particles which are closer than a specific linking length. This FOF approach can be extended to using the velocity information. The halo centres are determined afterwards, and can e.g. be defined as the halo centre of mass. The so-called halo merger trees store halo merging history, i.e. they provide all the progenitors of a given halo backwards in time. Overall, halo properties (bulk velocity, virial mass, peak of the rotation curve) are found to agree well between different halo finders. Some differences are found in the identification of subhalos and their properties. In particular, phase-space finders are more effective in finding low-mass subhalos. However, the definition of the mass of a subhalo is somewhat ambiguous, and the peak of the rotation curve $v_{\rm max}$ is often used as a proxy for the original infall mass (Knebe et al., 2011).

The halo finder applied on the OuterRim simulation presented in Section 2.2.1.5 is a simple FOF algorithm, with linking length 0.168 times the mean inter-particle separation. The halo centre is given at the minimum of the gravitational potential. Halos in the OuterRim simulation are shown in Figure 2.8. The Rockstar (robust overdensity calculation using k-space topologically adaptive refinement, Behroozi et al. 2013) algorithm uses the particle positions and velocities (and time component) to identify halos, halo substructures and halo merger trees. This algorithm was applied on the MultiDark simulation presented in Section 2.2.1.5.

2.2.2.3 Hydrodynamical simulations

Galaxy formation is the result of baryonic processes. One can describe baryons as a fluid following continuity and Euler equations, with a pressure P and energy U linked through the first law of thermodynamics. Two types of codes exist: smooth particle hydrodynamical (SPH) codes, that sample the fluid with particles (e.g. GADGET-2, Springel, 2005) or Euler codes that solve the fluid equation on a grid (e.g. RAMSES, Teyssier, 2002). In addition to this simple gas dynamics, other baryonic processes are considered: radiative gas cooling or photoheating, star formation, stellar feedback (including supernovae explosions), active galactic nuclei (AGN) feedback, magnetic fields, etc. Note that a number of these processes takes place at very small scale — below the simulation resolution — and hence requires effective models, tuned directly on observations or through semi-analytical models, as described below. Galaxies can then be identified based on the density of stellar particles (of typical mass $\sim 10^6\,{\rm M}_\odot$, e.g. Aubert et al. 2004; Dubois et al. 2016). Though hydrodynamical simulations are a major step towards understanding the complex process of galaxy formation and its relation to the environment, results

differ significantly depending on the feedback implementation (Scannapieco et al., 2012). In addition, hydrodynamical simulations are numerically very expensive and limited to a few Mpc in size and are thus improper for direct large scale clustering predictions (typically $\sim 100\,\mathrm{Mpc}$ scales). However, they can be used to specifically simulate galaxy formation over a large dynamical range of scales starting from patches of a large initial N-body simulation (e.g. Crain et al., 2009) and infer prescriptions for HOD models, as discussed below.

2.2.2.4 Semi-analytical models

The star formation efficiency and hence galaxy formation depends on the ability of the gas to cool down. In turn, gas is reheated by feedback processes. The stellar-to-halo mass function peaks around $10^{12}\,\mathrm{M}_\odot$; above this pivot mass, AGN feedback heats the gas of the surrounding interstellar medium (ISM) (Silk and Rees, 1998). At lower halo mass, massive stars ($\gtrsim 5-8\mathrm{M}_\odot$) end their life into type II supernovae, ejecting energy and momentum in the ISM or even outside of the halo. A semi-analytical model of galaxy formation puts into equation the flows between stars, hot gas and cold gas reservoirs, and predicts galaxy star formation history and metallicity enrichment (Baugh, 2013). Prescriptions of semi-analytical models are based on simplifying assumptions that must be tested against high-resolution hydrodynamical simulations or observed data. They typically rely on 10 to 30 parameters which can be measured observationally, by studying the relations between galaxy colours, magnitude, gas fraction, at different redshifts (e.g. Henriques et al., 2015). When applied on merger trees of N-body simulations, semi-analytical models can predict the number of galaxies per halo and hence provide groundings for HOD prescriptions. Contreras et al. (2013) compared the results obtained from different semi-analytical galaxy formation models run on the Millenium simulation (Springel et al., 2005) and found very good agreement between semi-analytical models for samples selected by stellar mass, as shown in Figure 2.9. However, they differ significantly (though their overall shape agree) for samples selected by cold gas mass or star formation rate — owing to the different prescriptions for gas-reheating and gas-stripping processes.

We have mentioned theoretical uncertainties in the way galaxies populate halos. These are to be constrained with observed data, such as galaxy clustering. The two following, simple, schemes make a direct connection between halos and galaxies such that their (possible) parameters can be easily tuned on observed clustering measurements.

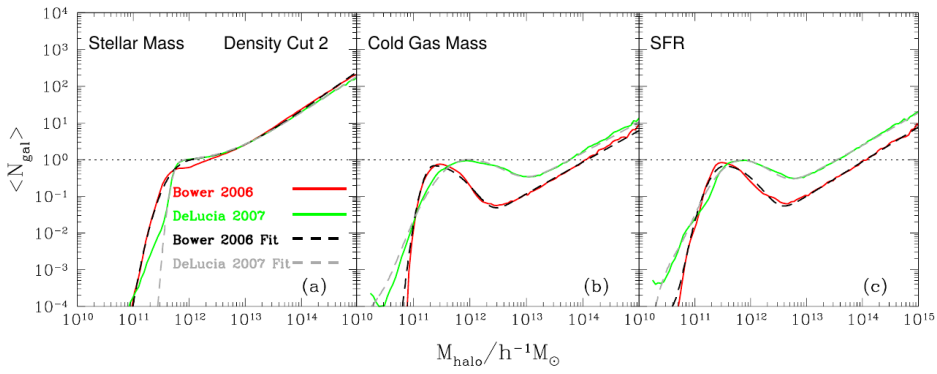

Figure 2.9 – HOD measured on the outputs of two different semi-analytical models (GALFORM and LGALAXIES) run on the Millennium simulation. Galaxy samples are selected by stellar mass (left), cold gas mass (middle) and star formation rate (right). HOD (introduced in Section 2.2.2.6) fits are shown in dashed lines. Taken from Contreras et al. (2013).

2.2.2.5 Abundance matching

The simplest way to associate galaxies with halos is to assume a monotonic relation between a galaxy property (e.g. stellar mass, luminosity for a volume-limited galaxy sample) and the property of its host halo (e.g. its mass). Both halos and galaxies are ranked (following the given property) and associated one-to-one in descending order.

Thanks to the increased resolution of numerical simulations, it has become possible to identify subhalos in halos. Subhalo abundance matching (SHAM) associates galaxies to these halos and subhalos. Of course, in such a model, one has to specify which galaxy and halo properties to associate, and possibly allow for some scatter in this relation. For example, subhalos are rapidly stripped off their outer material after accretion (by 30% to 70% depending on the location in the host halo, Nagai and Kravtsov 2005), while tidal stripping starts much later for galaxies (selected by stellar mass). Hence, galaxy properties are often matched to properties (e.g. mass) of the subhalo at the time it accreted into its host halo, or quantities more robust to stripping, such as the peak of their rotation curve $v_{\rm max}$. SHAM was found to be quite successful at describing observational samples (e.g Moster et al., 2010). However, the environment property is still expected to have some impact on the galaxy properties; for example, Simha et al. (2012) found that SHAM does not account for the stellar mass loss of satellites found in hydrodynamical SPH simulations.

2.2.2.6 Halo occupation distribution

The HOD specifies the probability distribution $P(N|M)$ of the number of galaxies (selected according to some criterion, e.g. stellar mass or star formation rate) conditioned on the halo mass M. Contributions to $P(N|M)$ are typically split between central galaxies in the halo and satellite galaxies orbiting within the halo. In essence, the HOD is a description of the way galaxies populate halos, so it is not predictive. Its parameters can be fitted to semi-analytical models. An example is given in Figure 2.9, where 5 parameters are required to model $\langle N(M)\rangle$ in the case of the stellar mass selected sample and 9 in the case of the selections based on the cold gas mass or on the star formation rate. HODs can also be constrained by galaxy clustering; below $1\,{\rm Mpc}\,h^{-1}$ (the 1-halo term) the 2-point correlation function is sensitive to the fraction of satellite galaxies as the number of galaxy pairs scales as the square of the number of satellites. On larger scales, galaxy clustering is more sensitive to the halo mass or the scatter between halo mass and the property used to select galaxies (Wechsler and Tinker, 2018). Note that one has also to specify the full probability law (Poisson, Bernouilli, negative binomial...) where to place central and satellite galaxies within the halo (e.g. Jiménez et al., 2019). In the standard HOD approach, the galaxy content of a halo of a given mass is statistically independent of its larger scale environment. However, it has been shown that clustering of halos depends on other, secondary, properties than their mass, e.g. formation time, concentration (that we will define in Section 2.3.1) or spin often correlated with the assembly history of the dark matter halo — a phenomenon called *halo assembly bias* (Wechsler et al., 2002; Gao et al., 2005). Similarly, the galaxy properties or number of galaxies per halo can depend on these secondary halo properties — which is called *galaxy assembly bias*. In the HOD framework, the mean number of galaxies per halo can be changed following some secondary property, e.g. concentration for Hearin et al. (2016) decorated HODs, which do not require the full halo merger tree (that is computationally demanding to generate for large simulations).

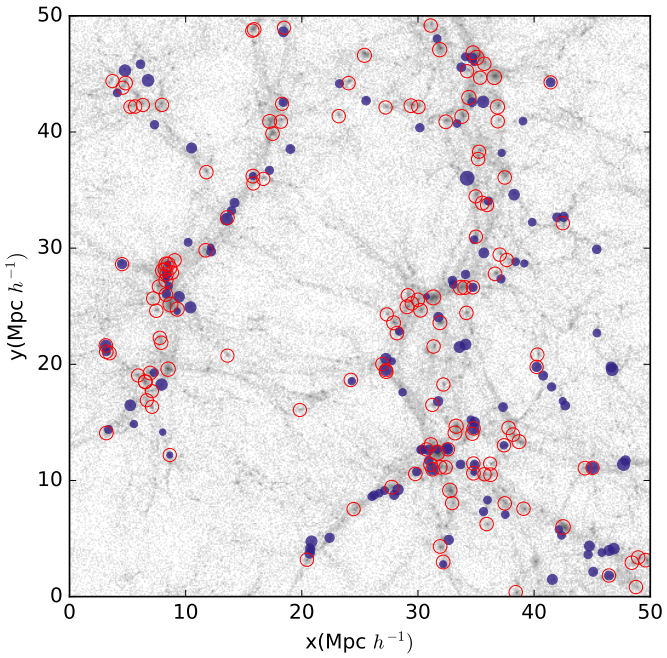

Figure 2.10 – Simulated ELGs (blue filled circles) and dark matter halos above $10^{11.8}\,\mathrm{M}_{\odot}\,h^{-1}$ (open red circles) on top of the dark matter distribution (grey) at $z = 1$. Taken from Gonzalez-Perez et al. (2018).

2.2.3 The case of eBOSS ELGs

In the following we specifically review the small scale clustering properties of the eBOSS emission line galaxy (ELG) sample which we will use in our clustering analysis. As we will see in the next chapter, eBOSS ELGs are selected through their [OII] flux, which is highly correlated to the star formation rate. Note however that [OII] flux is reduced due to dust attenuation in massive galaxies, as pointed out by Gonzalez-Perez et al. (2018).

For galaxy samples selected by luminosity or stellar mass, such as luminous red galaxies (LRGs), the average number of central galaxies $\langle N_{\mathrm{cent}}(M)\rangle$ can be described by a step function, eventually reaching 1 at high mass halos (see Figure 2.9); this is however not the case for galaxies selected through their [OII] flux as are eBOSS ELGs. This can be seen in Figure 2.12 (top panel, continuous lines), which displays the central and satellite HODs. The central HOD resembles a Gaussian plus a step function — or power law. Gonzalez-Perez et al. (2018) found that the HOD of central disks looks like a Gaussian, while the distribution of central spheroids, which have higher stellar masses, is closer to a step function. On the other hand, the predicted mean number of satellite galaxies $\langle N_{\mathrm{sat}}(M)\rangle$ follows a usual power law, with a satellite fraction of $\simeq 2\%$.

Running a semi-analytical model of galaxy formation (GALFORM) on the merger history of N-body simulations (Millenium suite, Springel et al., 2005; Boylan-Kolchin et al., 2009) Guo et al. (2013) and Gonzalez-Perez et al. (2018) found that ELGs reside rather in filaments than in the density peaks (as would be the case of LRGs), as can be seen in Figure 2.10.

Again based on a semi-analytical model (LGALAXIES) run on a N-body simulation (MXXL, Angulo et al., 2012), Orsi and Angulo (2018) claimed that the quenching of star formation rate induced by gas-stripping processes lowers the fraction of satellite ELGs in the inner part of

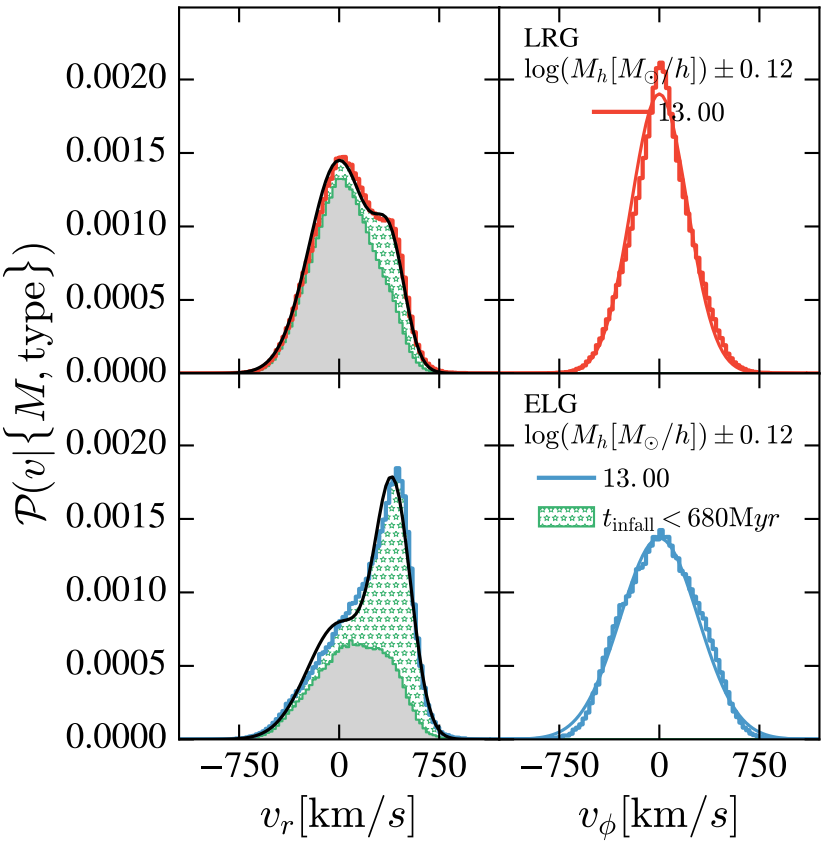

Figure 2.11 – Intra-halo radial (left panels) and tangential (right panels) velocity distributions, for LRG (top) and ELG (bottom) satellites, as predicted by GALFORM, in halos of $10^{13\pm0.125}{\rm M}_\odot\, h^{-1}$. Infall radial velocities are positive. The black continuous line (left) is a double Gaussian fit, while the red and blue solid lines (right) are single Gaussian fits. The dotted green regions (left) show the contribution from galaxies recently accreted, within the last 680 Myr. Taken from Orsi and Angulo (2018).

halos, hence favouring recently accreting galaxies, located in the halo outskirts, with net infall velocities. On the contrary, LRGs are distributed more evenly in the halos, with a much less pronounced infall velocity. This can be observed in Figure 2.11, displaying the tangential and radial velocities of LRGs and ELGs, specifically emphasising the contribution from recently infalling galaxies ($t_{\rm infall} < 680\,{\rm Myr}$): the infall contribution from these recently accreted galaxies is much higher in the case of ELGs.

Guo et al. (2019) used the eBOSS ELG sample to simultaneously constrain, in 4 redshift bins ($0.7 < z < 0.8$, $0.8 < z < 0.9$, $0.9 < z < 1.0$, $1.0 < z < 1.2$), the stellar-to-halo mass relation, the survey completeness (in terms of stellar mass) and the quenched galaxy fraction, which can be combined to provide HOD predictions. For this, they used galaxy stellar masses inferred from eBOSS ELG photometric measurements in g, r and z bands (see Section 3.2), and the small scale 2-point correlation function of the eBOSS ELG sample, which is modelled using the MultiDark simulation. In the low mass end, their HOD predictions match those of Gonzalez-Perez et al. (2018); Griffin et al. (2019) very well, as seen in Figure 2.12 (top panel). However, in massive halos, their central HOD is higher which they claim may be due to the treatment of dust (and thus line extinction) by GALFORM. They find a peak halo mass of $\simeq 10^{12}\,{\rm M}_\odot$ for centrals and $\simeq 10^{12.6}\,{\rm M}_\odot$ for satellites. This is in line with previous measurements by Favole et al. (2016), which rely on photometric, spectroscopic and lensing data, mapped onto N-body simulations

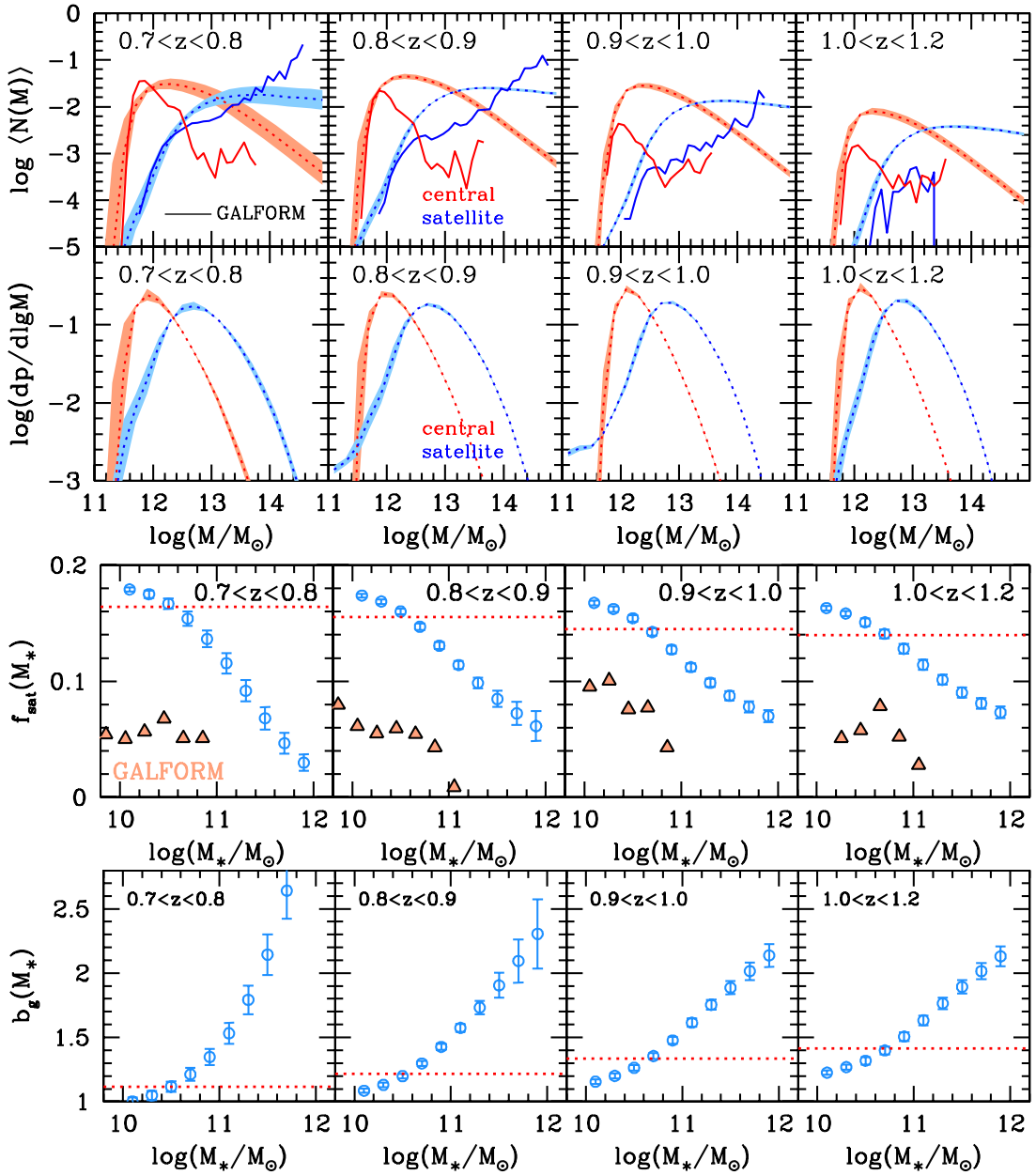

Figure 2.12 – Top: HOD model of eBOSS ELG central (red) and satellite (blue) galaxies, in 4 redshift bins. GALFORM predictions (as described in Gonzalez-Perez et al. (2018); Griffin et al. (2019), run on the MultiDark simulation) are shown in continuous lines. The second line of plots shows the probability distribution of host halo masses, for centrals (red) and satellites (blue). Middle: satellite fraction as a function of stellar mass; dotted lines are the mean and GALFORM predictions are shown with triangles. Bottom: galaxy bias and its mean (dotted lines) as a function of stellar mass. Taken from Guo et al. (2019).

using SHAM to account for survey incompleteness. Guo et al. (2019) also find a higher satellite fraction, $\simeq 15\%$ versus $\simeq 5\%$ (Figure 2.12, middle panel), indicating that the star formation of satellites may be too suppressed by gas-stripping processes in GALFORM. This is consistent with the value of 22% found by Favole et al. (2016). Finally, the bias measurement ranges from 1.1 at $z = 0.7$ to 1.4 at $z = 1.2$ (Figure 2.12, bottom panel). This is higher than Gonzalez-Perez et al. (2018), which can be explained by the higher halo mass and satellite fraction.

Finally, within HOD modelling tuned to reproduce predictions from a version of LGALAXIES, Jiménez et al. (2019) found that the scatter of the HOD for satellite galaxies has a significant impact on the small scale clustering (one halo term) of galaxies selected through their star formation rate, favouring a negative binomial distribution rather than a Poisson distribution as usually assumed in HOD prescriptions.

These findings will be used in Section 5.2 to design the HOD prescriptions for the mock challenge organised in eBOSS to test the reliability of our theoretical model.

2.3 Tracer bias

With Section 2.1 we are able to describe the evolved, mildly non-linear, matter power spectrum. However, we do not observe all matter, but objects emitting or absorbing light: galaxies, quasars, gas, etc. A key hypothesis in clustering analyses is that the large scale density fluctuations $\delta_g(\mathbf{x})$ of these objects can be related to the properties of the total matter field $\delta(\mathbf{x})$, namely:

$$\delta_g(\mathbf{x}) = \mathcal{F}(\delta(\mathbf{x})). \tag{2.94}$$

Such objects are called *tracers* of the large scale mass density field. One may consider two approaches to model this relation. One can relate the density of galaxies to the local mass density field and describe how the latter connects to the large scales. For example, the galaxy - halo connection discussed in the previous section can be associated to a model of the clustering of halos to obtain the relation between the galaxy and matter power spectrum. This is the approach we briefly recap in the first section. The other way consists in expanding the relation between δ_g and the mass density field, with coefficients that are free parameters of the model — which is the option considered in the second section and in our clustering analysis.

2.3.1 Halo models

A complementary approach to perturbation theory is the so-called halo model. Within this formalism, we first derive the non-linear dark matter power spectrum. The calculation is straightforwardly extended to the power spectrum of galaxies, given a HOD prescription. This section follows Seljak (2000).

2.3.1.1 Dark matter power spectrum

All matter is assumed to reside in halos with a mass M and density profile $\rho(r)$, typically of the (spherical averaged) form:

$$\rho(r) = \frac{\rho_s}{(r/R_s)^\gamma \left[1 + (r/R_s)^\alpha\right]^{(\beta-\gamma)/\alpha}}. \tag{2.95}$$

The profile goes as $(r/R_s)^{-\gamma}$ in the halo inner part and as $(r/R_s)^{\gamma-\beta}$ in the outer part. For the Navarro-Frenk-White profile (Navarro et al., 1996), for example, $(\alpha, \beta, \gamma) = (1, 3, 1)$, and for the Hernquist profile (Hernquist, 1990), $(\alpha, \beta, \gamma) = (1, 4, 1)$. This model is assumed to be universal in units of the scale radius R_s. One can define the concentration:

$$c = \frac{R_{\rm vir}}{R_s} \tag{2.96}$$

with $R_{\rm vir}$ the virial radius; higher concentration means the halo is denser in its centre. The halo mass M is simply the integral of Eq. (2.95), such that the parametrisation (ρ_s, R_s) can be recast

into (M, c). The halo mass function dn/dM describes the number density of halos as a function of mass. It can be expressed as:

$$\frac{dn}{dM}dM = \frac{\bar{\rho}}{M}f(\nu)d\nu \tag{2.97}$$

where the function $f(\nu)$ depends only on (the square of) the peak height $\nu = (\delta_c/\sigma(M))^2$. δ_c is the critical density contrast in the spherical collapse model (see Section 2.2.2.1). $\sigma^2(M)$ is the variance of the linear density fluctuations in a sphere of radius $R(M)$, with $M = 4\pi/3R^3\bar{\rho}$ (see e.g. Eq. (1.159) for a radius of $8\,\mathrm{Mpc}\,h^{-1}$):

$$\sigma^2(M) = \frac{1}{2\pi^2}\int k^2 dk P_m^{\mathrm{lin}}(k)W^2(kR) \tag{2.98}$$

with $W(x) = 3\left(\sin x - x\cos x\right)/x^3$ the Fourier transform of the unit top-hat window and $P_m^{\mathrm{lin}}(k)$ the linear matter power spectrum at the redshift of interest ($P_m^{\mathrm{lin}}(k) = D_+^2 P_0(k)$). Press and Schechter (1974) (PS) were the first to propose an analytic prediction for f: $\nu f(\nu) = \sqrt{\nu/(2\pi)}e^{-\nu/2}$, based on the postulate that the mass fraction contained in halos of mass $> M$ is equal to twice [4] the probability that the average density contrast in a sphere of radius $R(M)$ is above δ_c. However, the PS mass function was shown to overpredict the density of halos at low masses ($M < M_\star$, with $M_\star$ the non-linear mass scale defined by $\sigma(M_\star) = 1$), triggering both theoretical (e.g. Bond et al., 1991) and phenomenological developments. The obtained prescriptions can be generically written as:

$$\nu f(\nu) \propto \left(1 + (a\nu)^{-p}\right)(a\nu)^{1/2}e^{-a\nu/2}. \tag{2.99}$$

For example, based on N-body simulations Sheth and Tormen (1999) proposed $(a, p) = (0.707, 0.3)$ (while $(a, p) = (1, 0)$ for the PS mass function). The scaling factor is given by mass conservation:

$$\int d\nu f(\nu) = 1, \tag{2.100}$$

since we assume all matter is contained in halos. The final ingredient to predict the non-linear matter power spectrum is the halo bias $b(\nu)$, which Sheth and Tormen (1999) proposed to be:

$$b(\nu) = 1 + \frac{a\nu - 1}{\delta_c} + \frac{2p}{\delta_c\left[1 + (a\nu)^p\right]}. \tag{2.101}$$

The non-linear dark matter power spectrum $P_{dm}(k)$ is split in a one halo term $P_{1h}(k)$, accounting for intra-halo correlation (since halos are not point-like) and a two halo term $P_{2h}(k)$, describing the correlation between two halos:

$$P_{dm}(k) = P_{1h}(k) + P_{2h}(k). \tag{2.102}$$

The halo - halo correlation function is obtained by convolving the density profiles of halos. It is thus more convenient to work in Fourier space, in which the halo density profile (normalised to the halo mass) reads:

$$y(k, M) = \frac{4\pi}{M}\int dr r^2 \rho(r, M)\frac{\sin(kr)}{kr} \tag{2.103}$$

where by definition $y(0, M) = 1$. The two halo power spectrum is simply obtained by integration over the mass function, weighted by the bias and halo density profile (Seljak, 2000):

$$P_{2h}(k) = P_m^{\mathrm{lin}}(k)\left[\int d\nu f(\nu)b(\nu)y(k, M(\nu))\right]^2. \tag{2.104}$$

4. The fudge factor 2 was introduced in Press and Schechter (1974) to ensure mass conservation (see Eq. (2.100)). A demonstration based on the theory of excursion sets was provided in Bond et al. (1991).

Hence the two halo term is simply the linear matter power spectrum $P_m^{\rm lin}(k)$ times a k-dependent function.

The one halo term is the convolution of the density profile, integrated over the mass function:

$$P_{1h}(k) = \int d\nu f(\nu) \frac{M(\nu)}{\bar{\rho}} \left[y(k, M(\nu))\right]^2 . \tag{2.105}$$

The term $M(\nu)$ makes the relative contribution of high mass halos to the one halo power spectrum larger than to the two halo term. In the limit $k \to 0$ and $y \to 1$ the one halo term reduces to Poisson shot noise. Finally, Seljak (2000) specify the halo mass dependence of $\rho(r, M)$ (and thus $y(k, M)$) assuming the concentration c follows a (negative) power law in mass. They find their dark matter power spectrum prediction to be in very good agreement with a fitting formula tuned on numerical simulations (Peacock and Dodds, 1996).

As can be seen in Figure 2.13, on large scales, the one halo term of the power spectrum is dominated by massive halos ($M \simeq 10^{14}\,{\rm M}_\odot$), contrary to very small scales ($k \gtrsim 100\,h\,{\rm Mpc}^{-1}$), where $y(k, M)$ significantly departs from 1, faster for most massive halos as they are typically less concentrated. In fact, Seljak (2000) argue that the non-linear dark matter power spectrum is mostly sensitive to the typical scale where $y(k, M)$ significantly departs from 1 (rather than to the complete inner halo profile) and to the halo mass function — which we recall depends on the linear matter power spectrum through $\sigma(M)$. Hence, valuable cosmological information may still be present on highly non-linear scales.

2.3.1.2 Galaxy power spectrum

To extend the previous formalism to the galaxy power spectrum, one requires the probability of N galaxies to reside in a halo of mass M, i.e. a HOD prescription. The mean density $\bar{n}$ of galaxies in the sample verifies:

$$\int \frac{\langle N \rangle}{M} f(\nu) d\nu = \frac{\bar{n}}{\bar{\rho}}. \tag{2.106}$$

Let us further assume that galaxies follow the dark matter profile — a condition that can be relaxed by changing $y(k, M(\nu))$ below. Then, the two halo term gives:

$$P_{g,2h}(k) = P_m^{\rm lin}(k) \left[\frac{\bar{\rho}}{\bar{n}} \int d\nu f(\nu) b(\nu) \frac{\langle N \rangle}{M} y(k, M(\nu))\right]^2 . \tag{2.107}$$

Note that this last formula is not correct for central galaxies, as by definition they do not contribute a $y(k, M(\nu))$ term. In practice, this only matters on small scales where the two halo term is anyway subdominant compared to the one halo term.

On large scales, the two halo galaxy power spectrum behaves as:

$$P_{g,2h}(k) = \langle b \rangle^2 P_m^{\rm lin}(k) \tag{2.108}$$

where the average galaxy bias is obtained by integration over the mass function, weighted by the mean number of galaxies:

$$\langle b \rangle = \frac{\bar{\rho}}{\bar{n}} \int d\nu f(\nu) b(\nu) \frac{\langle N \rangle}{M}. \tag{2.109}$$

The one halo term is given by (Seljak, 2000):

$$P_{g,1h}(k) = \frac{\bar{\rho}^2}{\bar{n}^2} \int d\nu f(\nu) \frac{M(\nu)}{\bar{\rho}} \frac{\langle N(N-1) \rangle}{M(\nu)^2} \left[y(k, M(\nu))\right]^p . \tag{2.110}$$

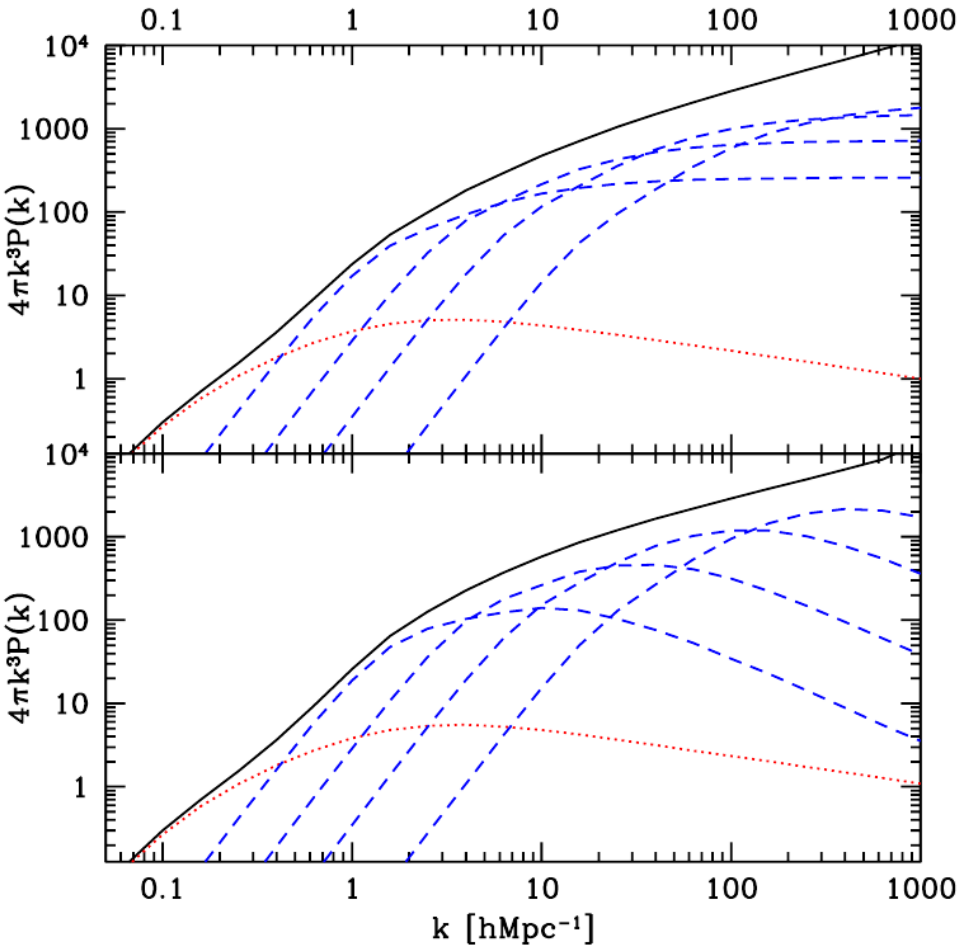

Figure 2.13 – Contributions to the one halo matter power spectrum from different halo mass intervals. Dashed lines, from left to right: $M > 10^{14}\,\mathrm{M}_\odot\,h^{-1}$, $10^{14}\,\mathrm{M}_\odot\,h^{-1} > M > 10^{13}\,\mathrm{M}_\odot\,h^{-1}$, $10^{13}\,\mathrm{M}_\odot\,h^{-1} > M > 10^{12}\,\mathrm{M}_\odot\,h^{-1}$, $10^{12}\,\mathrm{M}_\odot\,h^{-1} > M > 10^{11}\,\mathrm{M}_\odot\,h^{-1}$. The total non-linear dark matter power spectrum is shown in solid lines, the two halo term in dotted lines. Top and bottom panels show different choices in terms of halo mass function and concentration models.

where p is taken to be 2 when $\langle N(N-1)\rangle > 1$ and 1 when $\langle N(N-1)\rangle < 1$. Indeed, in the former case, the one halo term is dominated by satellite galaxies. In the latter case, there are virtually only satellite-central pairs, where centrals contribute a $y(k, M(\nu))$ term. The expressions (2.107) and (2.110) for the galaxy power spectrum reduce to the dark matter power spectrum (2.104) and (2.105) when both $\langle N\rangle\,(M)$ and $\langle N(N-1)\rangle\,(M)$ are independent of M, that is, galaxies follow the same distribution as dark matter halos. Seljak (2000) found that this model reproduces well the results obtained with semi-analytical models run on N-body simulations and with observations, as can be seen in Figure 2.14. In particular, as seen on this figure, the model accounts for the small-scale power-law scaling of the galaxy clustering, as opposed to the dark matter halo clustering. The one halo term of the galaxy power spectrum is indeed dominant at smaller scales than that of the dark matter power spectrum, as $\langle N(N-1)\rangle\,(M) < \langle N\rangle\,(M)$ for halos of mass $\lesssim 10^{14}\,\mathrm{M}_\odot\,h^{-1}$. Note that a possible environment dependence of $\langle N(N-1)\rangle$ would not affect the one halo term if the average is taken over all the environments covered by the galaxy sample — however, such a dependence would impact the two halo term. Here

we took the galaxy density profile within halos to be the same as dark matter, but considering another one is straightforward.

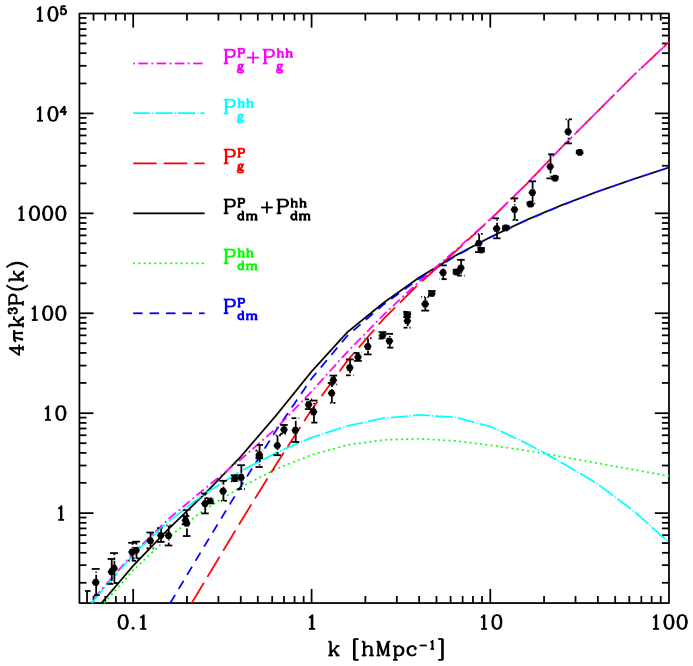

Figure 2.14 – Comparison of the dark matter and galaxy halo models to the observed galaxy power spectrum of APM and IRAS surveys (concatenated in Peacock and Dodds, 1994), displayed as points with error bars. The dark matter one halo term is shown as dashed blue lines, the two halo term with green dotted lines, the total non-linear dark matter power spectrum in black continuous lines. The galaxy one halo term is shown in dashed red, the two halo term in light blue and the total galaxy power spectrum in magenta. Taken from Seljak (2000).

This approach is very interesting, as it provides insights in the different contributions to the highly non-linear power spectrum as a function of scale and halo mass. However, predicting the galaxy power spectrum requires a HOD prescription, which we have seen depends on many parameters to be marginalised over. Also, we will see in Section 2.4 that another, non-linear effect called redshift space distortions must be accounted for, which is a theoretical bottleneck to probing very small scales with 3-dimensional clustering.

We note however the work of Hand et al. (2017), extending the analysis of the galaxy power spectrum up to $k = 0.4\,h\,\mathrm{Mpc}^{-1}$ in the halo model approach, with 9 free parameters in addition to the cosmological parameters. The improvement in cosmological constraints compared to the perturbative approach described below appeared marginal (5 to 10% reduction in the uncertainty).

2.3.2 Perturbative approach

Let us extend the perturbative approach for the matter power spectrum of Section 2.1 with a bias development. We follow the calculation of McDonald and Roy (2009).

First, we need to specify the quantities the galaxy density field δ_g can depend on. Of course, δ_g is related to the matter density contrast δ and velocity divergence θ. We further consider the velocity $\mathbf{u}$ and the potential Φ, to introduce some dependence on the path followed by the

fluid to reach its density contrast δ and velocity divergence θ at the redshift of interest. Neither a global change in the potential Φ nor its first derivative (the gravitational field) should be observable. Hence, the relevant variable to account for the δ_g dependence in Φ is:

$$s_{ij}(\mathbf{x}) = \partial_i\partial_j\Phi(\mathbf{x}) - \frac{1}{3}\delta_{ij}\delta(\mathbf{x}) = \gamma_{ij}\delta(\mathbf{x}), \tag{2.111}$$

where we implicitly rescaled Φ by $a^2 4\pi G_{\rm N}\bar{\rho}$ and introduced, following the Poisson equation (2.5):

$$\gamma_{ij} = \partial_i\partial_j\Delta^{-1} - \frac{1}{3}\delta_{ij}. \tag{2.112}$$

The trace of $\partial_i\partial_j\Phi(\mathbf{x})$ is removed to avoid redundancy with δ. Similarly, a global change in the velocity field is not observable, hence we work with $\partial_i u_j = \partial_i\partial_j\Delta^{-1}\theta$. We implicitly rescaled both the peculiar velocity $\mathbf{u}$ and its divergence θ by $-f\mathcal{H}$ such that $\delta = \theta$ at linear order (see Eq. (2.18)). Hence, it is more convenient to work with (not to be confused with the conformal time):

$$\eta = \theta - \delta, \tag{2.113}$$

which is non-zero only above linear order. Then, we build the quantity:

$$t_{ij} = \partial_i u_j - \frac{1}{3}\delta_{ij}\theta - s_{ij} = \gamma_{ij}\eta. \tag{2.114}$$

Therefore, the galaxy density will be a function of δ, η, s_{ij}, t_{ij}. However, because of isotropy and homogeneity, the galaxy density can only depend upon scalar variables. Up to third order in the initial density field, we can build the products (remembering that t_{ij} is already second order):

$$s^2 = s_{ij}s_{ij} \qquad st = s_{ij}t_{ij} \qquad s^3 = s_{ij}s_{jk}s_{kl}. \tag{2.115}$$

Let us define the Fourier-space kernel S related to the operator γ_{ij}:

$$\gamma_{ij}(\mathbf{k}_1)\gamma_{ij}(\mathbf{k}_2) = S(\mathbf{k}_1,\mathbf{k}_2) = \left(\frac{\mathbf{k}_1\cdot\mathbf{k}_2}{k_1k_2}\right)^2 - \frac{1}{3}. \tag{2.116}$$

Using Eq. (2.35), at second order in the initial field δ_0:

$$\eta(\mathbf{k}) = D_+^2\int\frac{d^3q}{(2\pi)^3}\delta_0(\mathbf{q})\delta_0(\mathbf{k}-\mathbf{q})\left[F_{2,s}^{(2)}(\mathbf{q},\mathbf{k}-\mathbf{q}) - F_{1,s}^{(2)}(\mathbf{q},\mathbf{k}-\mathbf{q})\right] \tag{2.117}$$

$$= D_+^2\int\frac{d^3q}{(2\pi)^3}\delta_0(\mathbf{q})\delta_0(\mathbf{k}-\mathbf{q})D^{(2)}(\mathbf{k}_1,\mathbf{k}_2) \tag{2.118}$$

where we recall $F_{a,s}^{(2)}$ are the symmetrised PT kernels, given by Eq. (2.37) and Eq. (2.38), and D_+ is the linear growth factor. In the second line we introduced:

$$D^{(2)}(\mathbf{k}_1,\mathbf{k}_2) = F_{2,s}^{(2)}(\mathbf{k}_1,\mathbf{k}_2) - F_{1,s}^{(2)}(\mathbf{k}_1,\mathbf{k}_2) = \frac{2}{7}\left[S(\mathbf{k}_1,\mathbf{k}_2) - \frac{2}{3}\right]. \tag{2.119}$$

Hence, in real space, at second order, $\eta = 2/7s^2 - 4/21\delta^2$. Therefore, in the following, we work with the variable:

$$\nu = \eta - \frac{2}{7}s^2 + \frac{4}{21}\delta^2 \tag{2.120}$$

which is non-zero at third order only. Altogether we come up with the following set of variables:

first order	δ
second order	δ^2, s^2
third order	δ^3, δs^2, s^3, st, ν

We recover the well-known result that at linear (first) order the galaxy density is proportional to the matter density, with a coefficient called *linear bias* — up to some shot noise. Following McDonald and Roy (2009), we include a mean-zero Gaussian variable ϵ, of same magnitude as δ to account for the stochasticity of the galaxy - matter density relation. A Taylor expansion then provides the galaxy density up to third order:

$$\rho_g = p_0 \tag{2.121}$$
$$+ p_\delta \delta + p_\epsilon \epsilon \tag{2.122}$$
$$+ \frac{1}{2} p_{\delta^2} \delta^2 + \frac{1}{2} p_{s^2} s^2 + \frac{1}{2} p_{\epsilon^2} \epsilon^2 + p_{\delta\epsilon} \delta\epsilon \tag{2.123}$$
$$+ \frac{1}{6} p_{\delta^3} \delta^3 + \frac{1}{2} p_{\delta s^2} \delta s^2 + \frac{1}{6} p_{s^3} s^3 + p_{st} st + p_\nu \nu + \frac{1}{6} p_{\epsilon^3} \epsilon^3 + \frac{1}{2} p_{\delta^2\epsilon} \delta^2 \epsilon + \frac{1}{2} p_{\delta\epsilon^2} \delta\epsilon^2 + \frac{1}{2} p_{s^2\epsilon} s^2 \epsilon. \tag{2.124}$$

With this expansion, the expected value of ρ_g is (odd order terms vanish due to the Gaussianity of δ_0):

$$\langle \rho_g \rangle = p_0 + \frac{1}{2} p_{\delta^2} \sigma^2 + \frac{1}{3} p_{s^2} \sigma^2 + \frac{1}{2} p_{\epsilon^2} \sigma_\epsilon^2 \tag{2.125}$$

with $\langle \delta^2 \rangle = \sigma^2$, $\langle s^2 \rangle = S(\mathbf{q}, -\mathbf{q})\sigma^2 = 2/3\sigma^2$ and $\langle \epsilon^2 \rangle = \sigma_\epsilon^2$. Then, the galaxy density contrast $\delta_g = \rho_g / \langle \rho_g \rangle - 1$ reads at third order:

$$\delta_g = c_\delta \delta + c_\epsilon \epsilon \tag{2.126}$$
$$+ \frac{1}{2} c_{\delta^2} \left(\delta^2 - \sigma^2 \right) + \frac{1}{2} c_{s^2} \left(s^2 - \frac{2}{3} \sigma^2 \right) + \frac{1}{2} c_{\epsilon^2} \left(\epsilon^2 - \sigma_\epsilon^2 \right) + c_{\delta\epsilon} \delta\epsilon \tag{2.127}$$
$$+ \frac{1}{6} c_{\delta^3} \delta^3 + \frac{1}{2} c_{\delta s^2} \delta s^2 + \frac{1}{6} c_{s^3} s^3 + c_{st} st + c_\nu \nu + \frac{1}{6} c_{\epsilon^3} \epsilon^3 + \frac{1}{2} c_{\delta^2\epsilon} \delta^2 \epsilon + \frac{1}{2} c_{\delta\epsilon^2} \delta\epsilon^2 + \frac{1}{2} c_{s^2\epsilon} s^2 \epsilon \tag{2.128}$$

where c's are functions of the p's, σ and σ_ϵ. The galaxy - matter cross power spectrum is calculated at 1-loop by cross-correlating δ_g above with $\delta = \delta^{(1)} + \delta^{(2)} + \delta^{(3)}$ (where $\delta^{(i)}$ are given by Eq. (2.35) times D_+^i), remembering that odd order terms in δ_0 vanish (e.g. terms (2.127) must be correlated with $\delta^{(2)}$ only), and using Wick theorem (2.54) to break 4-tuples of δ_0 into

two power spectra P_0. We find:

$$P_{mg}(k) = c_\delta P_{\delta\delta}(k) \tag{2.129}$$
$$+ c_{\delta^2} \int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) F_{1,s}^{(2)}(\mathbf{q}, \mathbf{k}-\mathbf{q}) + \frac{34}{21} c_{\delta^2} \sigma^2 P_m^{\rm lin}(k) \tag{2.130}$$
$$+ c_{s^2} \int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) F_{1,s}^{(2)}(\mathbf{q}, \mathbf{k}-\mathbf{q}) S(\mathbf{q}, \mathbf{k}-\mathbf{q}) \tag{2.131}$$
$$+ 2c_{s^2} P_m^{\rm lin}(k) \int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) F_{1,s}^{(2)}(-\mathbf{q}, \mathbf{k}) S(\mathbf{q}, \mathbf{k}-\mathbf{q}) \tag{2.132}$$
$$+ \frac{1}{2} c_{\delta^3} \sigma^2 P_m^{\rm lin}(k) + \frac{1}{3} c_{\delta s^2} \sigma^2 P_m^{\rm lin}(k) \tag{2.133}$$
$$+ 2c_\nu P_m^{\rm lin}(k) \int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) \left[\frac{3}{2} D^{(3)}(\mathbf{q}, -\mathbf{q}, -\mathbf{k}) - 2F_{1,s}^{(2)}(-\mathbf{q}, \mathbf{k}) D^{(2)}(\mathbf{q}, \mathbf{k}-\mathbf{q}) \right] \tag{2.134}$$
$$+ 2c_{st} P_m^{\rm lin}(k) \int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) D^{(2)}(-\mathbf{q}, \mathbf{k}) S(\mathbf{q}, \mathbf{k}-\mathbf{q}) \tag{2.135}$$
$$+ \frac{1}{2} c_{\delta\epsilon^2} \sigma_\epsilon^2 P_m^{\rm lin}(k), \tag{2.136}$$

where $D^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = F_{2,s}^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) - F_{1,s}^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$, $P_m^{\rm lin}(k) = D_+^2 P_0(k)$ and $P_{\delta\delta}(k)$ is the 1-loop matter density - density power spectrum. Let us reorder these terms. Terms $\propto \sigma^2 P_m^{\rm lin}(k)$ (c_{δ^2}, c_{δ^3}, $c_{\delta s^2}$) and the shot noise contribution $c_{\delta\epsilon^2}$ ($\propto \sigma_\epsilon^2 P_m^{\rm lin}(k)$) are included in a redefinition of the linear bias c_δ dubbed b_1. Also, the c_{st} term gives, for $k \to 0$:

$$\int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) D^{(2)}(-\mathbf{q}, \mathbf{k}) S(\mathbf{q}, \mathbf{k}-\mathbf{q}) \xrightarrow[k\to 0]{} -\frac{8}{63}\sigma^2 \tag{2.137}$$

Hence, the kernel of the c_{st} term is redefined as:

$$D^{(2)}(-\mathbf{q}, \mathbf{k}) S(\mathbf{q}, \mathbf{k}-\mathbf{q}) \to D^{(2)}(-\mathbf{q}, \mathbf{k}) S(\mathbf{q}, \mathbf{k}-\mathbf{q}) + \frac{8}{63} \tag{2.138}$$

and the counter term, $-16/63 c_{st} P_m^{\rm lin}(k)\sigma^2$ is absorbed into b_1. Similarly, the kernel of the c_{s^2} term is redefined as:

$$F_{1,s}^{(2)}(-\mathbf{q}, \mathbf{k}) S(\mathbf{q}, \mathbf{k}-\mathbf{q}) \to F_{1,s}^{(2)}(-\mathbf{q}, \mathbf{k}) S(\mathbf{q}, \mathbf{k}-\mathbf{q}) - \frac{68}{63} \tag{2.139}$$

and the counter term, $68/63 c_{s^2} P_m^{\rm lin}(k)\sigma^2$ is absorbed into the linear bias b_1. With the above renormalisation, b_1 is eventually expressed as:

$$b_1 = c_\delta + \left(\frac{34}{21} c_{\delta^2} + \frac{68}{63} c_{s^2} + \frac{1}{2} c_{\delta^3} + \frac{1}{3} c_{\delta s^2} - \frac{16}{63} c_{st} \right) \sigma^2 + \frac{1}{2} c_{\delta\epsilon^2} \sigma_\epsilon^2. \tag{2.140}$$

Finally, after angle integration, c_{s^2}, c_ν and c_{st} terms are found to be $\propto \sigma_3^2(k) P_m^{\rm lin}(k)$, with:

$$\sigma_3^2(k) = \frac{105}{16} \int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) \left[D^{(2)}(-\mathbf{q}, \mathbf{k}) S(\mathbf{q}, \mathbf{k}-\mathbf{q}) - \frac{8}{63} \right]. \tag{2.141}$$

The 1-loop galaxy - matter cross power spectrum eventually comprises 4 terms only:

$$P_{gm}(k) = b_1 P_{\delta\delta}(k) + b_2 P_{b2,\delta} + b_{s2} P_{bs2,\delta} + b_{\rm 3nl} \sigma_3^2(k) P_m^{\rm lin}(k) \tag{2.142}$$

with:

$$P_{b2,\delta}(k) = \int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) F_{1,s}^{(2)}(\mathbf{q}, \mathbf{k}-\mathbf{q}) \tag{2.143}$$

$$P_{bs2,\delta}(k) = \int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) F_{1,s}^{(2)}(\mathbf{q}, \mathbf{k}-\mathbf{q}) S(\mathbf{q}, \mathbf{k}-\mathbf{q}). \tag{2.144}$$

Let us move to the galaxy auto power spectrum:

$$P_{gg}(k) = b_1^2 P_{\delta\delta}(k) + 2b_1 b_2 P_{b2,\delta} + 2b_1 b_{s2} P_{bs2,\delta} + 2b_1 b_{3\rm nl}\sigma_3^2(k) P_m^{\rm lin}(k) \tag{2.145}$$

$$+ \frac{1}{2}\int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) \left[c_{\delta^2}^2 + 2c_{\delta^2} c_{s^2} S(\mathbf{q}, \mathbf{k}-\mathbf{q}) + c_{s^2}^2 S^2(\mathbf{q}, \mathbf{k}-\mathbf{q})\right] + N_\epsilon. \tag{2.146}$$

The first line corresponds to the first - second order bias cross terms. The second line contains the auto-correlation of second order terms, and all the shot noise like terms generated by ϵ. Again, the $c_{\delta^2}^2$, $c_{\delta^2} c_{s^2}$ and $c_{s^2}^2$ terms have the low k limit:

$$\int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) \xrightarrow[k\to 0]{} \int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q)^2 \tag{2.147}$$

$$\int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) S(\mathbf{q}, \mathbf{k}-\mathbf{q}) \xrightarrow[k\to 0]{} \frac{2}{3}\int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q)^2 \tag{2.148}$$

$$\int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) S^2(\mathbf{q}, \mathbf{k}-\mathbf{q}) \xrightarrow[k\to 0]{} \frac{4}{9}\int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q)^2 \tag{2.149}$$

which are removed from these terms and absorbed in the shot noise like term. Note that this latter term does not necessarily correspond to a pure Poisson shot noise and should be treated as a free stochastic bias parameter. Hence, after renormalisation, the galaxy auto power spectrum reads:

$$\boxed{\begin{aligned} P_{gg}(k) = {} & b_1^2 P_{\delta\delta}(k) + 2b_1 b_2 P_{b2,\delta} + 2b_1 b_{s2} P_{bs2,\delta} + 2b_1 b_{3\rm nl}\sigma_3^2(k) P_m^{\rm lin}(k) \\ & + b_2^2 P_{b22}(k) + 2b_2 b_{s2} P_{b2s2}(k) + b_{s2}^2 P_{bs22}(k) + N_g \end{aligned}} \tag{2.150}$$

where:

$$P_{b22}(k) = \frac{1}{2}\int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) \left[P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) - P_m^{\rm lin}(q)\right] \tag{2.151}$$

$$P_{b2s2}(k) = \frac{1}{2}\int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) \left[P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) S(\mathbf{q}, \mathbf{k}-\mathbf{q}) - \frac{2}{3} P_m^{\rm lin}(q)\right] \tag{2.152}$$

$$P_{bs22}(k) = \frac{1}{2}\int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) \left[P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) S^2(\mathbf{q}, \mathbf{k}-\mathbf{q}) - \frac{4}{9} P_m^{\rm lin}(q)\right]. \tag{2.153}$$

The bias prescription for the galaxy - velocity cross power spectrum, assuming no velocity bias (i.e. galaxy velocities are the same as the matter fluid velocity), is similar to the galaxy - matter cross power spectrum (2.142), except that the density kernel $F_{1,s}^{(2)}$ in Eq. (2.143) and Eq. (2.144) should be replaced by that for the velocity divergence $F_{2,s}^{(2)}$. Namely:

$$\boxed{P_{g\theta}(k) = b_1 P_{\delta\theta}(k) + b_2 P_{b2,\theta} + b_{s2} P_{bs2,\theta} + b_{3\rm nl}\sigma_3^2(k) P_m^{\rm lin}(k)} \tag{2.154}$$

with $P_{\delta\theta}$ the matter density - velocity cross power spectrum and:

$$P_{b2,\theta}(k) = \int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) F_{2,s}^{(2)}(\mathbf{q}, \mathbf{k}-\mathbf{q}) \tag{2.155}$$

$$P_{bs2,\theta}(k) = \int \frac{d^3q}{(2\pi)^3} P_m^{\rm lin}(q) P_m^{\rm lin}(|\mathbf{k}-\mathbf{q}|) F_{2,s}^{(2)}(\mathbf{q}, \mathbf{k}-\mathbf{q}) S(\mathbf{q}, \mathbf{k}-\mathbf{q}). \tag{2.156}$$

The bias terms of the galaxy - galaxy and galaxy - velocity power spectra are shown in Figure 2.15. We see that bias terms have different scale dependence, e.g. $\sigma_3^2(k) P_m^{\rm lin}(k)$ is higher at large scales ($0.05\,h\,\mathrm{Mpc}^{-1} < k < 0.1\,h\,\mathrm{Mpc}^{-1}$) than at smaller scales, relative to the other terms.

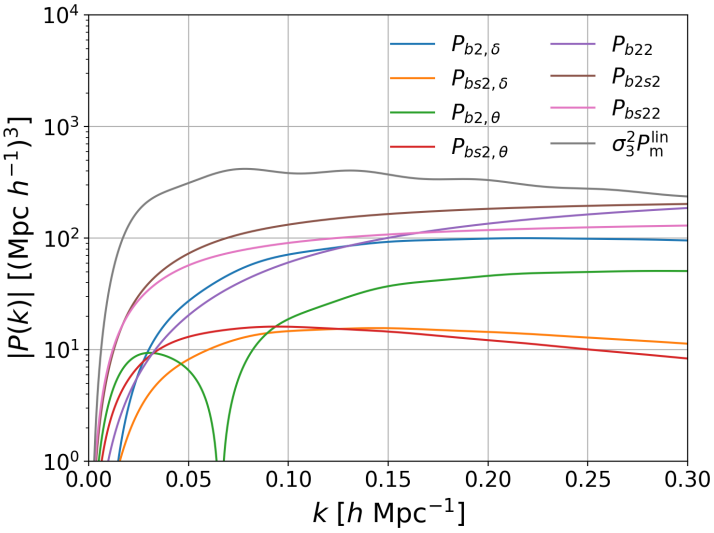

Figure 2.15 – Bias terms of the galaxy - galaxy and galaxy - velocity power spectra up to 1-loop.

The bias expansion described above results in 4 bias coefficients (b_1, b_2, b_{s2}, $b_{\rm 3nl}$), plus the shot noise term N_g to be marginalised over.

Note that halos (or galaxies) can also be depicted as a fluid gravitationally coupled to dark matter. In this so-called coevolution picture, assuming no velocity bias, and an initial local bias in Lagrangian space, Chan et al. (2012); Baldauf et al. (2012); Saito et al. (2014) found the following identification of the b_{s2} and $b_{\rm 3nl}$ bias coefficients:

$$b_{s2} = -\frac{4}{7}(b_1 - 1)\,, \tag{2.157}$$

$$b_{\rm 3nl} = \frac{32}{315}(b_1 - 1)\,. \tag{2.158}$$

These relations were found to be in relatively good (but not perfect) agreement with dark matter halo bias measurements in N-body simulations by Saito et al. (2014). By default, as in Gil-Marín et al. (2015); Beutler et al. (2017) clustering analyses, we adopt the above prescription, which we will show in Section 5.2.2 does not impact our cosmological measurements.

All bias terms are computed within our package `pyregpt` [5].

5. at https://github.com/adematti/pyregpt

2.4 Redshift space distortions

We have seen how to describe galaxy clustering in real space, i.e. as we knew their true space-time position. However, the redshift of a galaxy is due to both the Hubble flow and its peculiar velocity (and gravitational redshift, which we neglect here). When transforming a galaxy redshift into a distance (with Eq. (1.27)), distortions appear in the so-called *redshift space*. The derivations of this section follow Taruya et al. (2010). For conciseness bias terms are dropped, but will be reintroduced in Section 2.4.3.

2.4.1 Linear theory

The redshift space position $\mathbf{s}$ is related to the real space position $\mathbf{r}$ through:

$$\boxed{\mathbf{s} = \mathbf{r} - f v_z \hat{\mathbf{z}}} \tag{2.159}$$

with $v_z = -\mathbf{u}\cdot\hat{\mathbf{z}}/(f\mathcal{H})$, where we recall $\mathbf{u}$ is the physical peculiar velocity and f the growth rate of structure given by Eq. (2.16). $\hat{\mathbf{z}}$ is the line of sight. Let δ_s be the matter density contrast in redshift space. Mass conservation implies that $[1+\delta_s(\mathbf{s})]\,d^3s = [1+\delta(\mathbf{r})]\,d^3r$, hence:

$$\delta_s(\mathbf{s}) = \left|\frac{\partial \mathbf{s}}{\partial \mathbf{r}}\right|^{-1} [1+\delta(\mathbf{r})] - 1. \tag{2.160}$$

Its Fourier transform reads:

$$\delta_s(\mathbf{k}) = \int d^3r \left\{1 + \delta(\mathbf{r}) - \left|\frac{\partial \mathbf{s}}{\partial \mathbf{r}}\right|\right\} e^{-i\mathbf{k}\cdot\mathbf{s}}. \tag{2.161}$$

With Eq. (2.159), noting that $\left|\frac{\partial \mathbf{s}}{\partial \mathbf{r}}\right| = 1 - f\partial_z v_z$ ($|f\partial_z v_z| < 1$ at the scales of interest), we find:

$$\delta_s(\mathbf{k}) = \int d^3r \left\{\delta(\mathbf{r}) + f\partial_z v_z\right\} e^{i(fk\mu v_z - \mathbf{k}\cdot\mathbf{r})} \tag{2.162}$$

where μ denotes the cosine angle $\mu = \hat{\mathbf{k}}\cdot\hat{\mathbf{z}}$. When $|fkv_z| \ll 1$, and using that, at linear order, $ik_z v_z = \mu^2\delta$ (see Eq. (2.20)), we find the Kaiser formula (Kaiser, 1987):

$$\boxed{\delta_s(\mathbf{k}) = \left(1 + f\mu^2\right)\delta(\mathbf{k})} \tag{2.163}$$

Hence, the redshift space linear power spectrum reads:

$$\boxed{P_s(k,\mu) = \left(1 + f\mu^2\right)^2 P_m^{\rm lin}(k)} \tag{2.164}$$

with $P_m^{\rm lin}(k)$ the linear matter power spectrum at the considered redshift. The power spectrum in redshift space is enhanced with respect to real space, due to coherent motions on large scales. This enhancement increases with $f \simeq \Omega_m^\gamma$ (see Eq. (2.19)). Hence, redshift space distortions are a powerful tool to probe the energy content of the Universe (through Ω_m) or modified gravity (through γ).

2.4.2 Beyond the linear regime

In all generality, the power spectrum in redshift space reads (using Eq. (2.162)):

$$P_s(\mathbf{k}) = \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \left\langle e^{-ifk\mu\Delta v_z} \left[\delta(\mathbf{r}) + f\partial_z v_z(\mathbf{r})\right] \left[\delta(\mathbf{r}+\mathbf{x}) + f\partial_z v_z(\mathbf{r}+\mathbf{x})\right] \right\rangle \tag{2.165}$$

with $\Delta v_z = v_z(\mathbf{r}) - v_z(\mathbf{r}+\mathbf{x})$. This is an exact expression, which reveals two competing effects:

1. the Kaiser effect: at linear order, the term in braces correlates Eq. (2.163), and thus enhances the power spectrum on large scales with respect to real space, as discussed previously.
2. the Finger-of-God effect (Jackson, 1972): at small scales, as perturbations become non-linear, velocities become out of phase and the oscillations of the exponential term $e^{-ifk\mu\Delta v_z}$ (in the brackets) damp the redshift space power spectrum. This typically happens when $k \gtrsim 1/\sigma_v$, with (see Eq. (2.64)):

$$\sigma_v^2 = f^2 \left\langle v_z^2(0) \right\rangle = f^2\sigma_d^2 = \frac{f^2}{3} \int \frac{d^3q}{(2\pi)^3} \frac{P_m^{\rm lin}(q)}{q^2}. \tag{2.166}$$

Most contributions to the Finger-of-God term come from satellites within the same halo — though central galaxies may still not be at rest with respect to the halo centre and hence contribute to the Finger-of-God effect.

A simple, phenomenological model consists in splitting the Finger-of-God correction from the Kaiser term:

$$P_s(k,\mu) = D_{\rm FoG}(k,\mu,\sigma_v) \left[P_{\delta\delta}(k) + 2f\mu^2 P_{\delta\theta}(k) + f^2\mu^4 P_{\theta\theta}(k)\right], \tag{2.167}$$

where we recall $P_{\delta\delta}(k)$, $P_{\delta\theta}(k)$ and $P_{\theta\theta}(k)$ are the density - density, density - velocity and velocity - velocity power spectra ($\theta = \partial_i v_i$). Eq. (2.167) can be compared with Eq. (2.164), valid at large scales. However, this model ignores the coupling between the Finger-of-God and the Kaiser terms at intermediate scales. Let us rewrite Eq. (2.165) in the more compact form:

$$P_s(\mathbf{k}) = \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \left\langle e^{j_1 A_1} A_2 A_3 \right\rangle \tag{2.168}$$

where:

$$j_1 = -ik\mu f \tag{2.169}$$
$$A_1 = \Delta v_z = v_z(\mathbf{r}) - v_z(\mathbf{r}+\mathbf{x}) \tag{2.170}$$
$$A_2 = \delta(\mathbf{r}) + f\partial_z v_z(\mathbf{r}) \tag{2.171}$$
$$A_2 = \delta(\mathbf{r}+\mathbf{x}) + f\partial_z v_z(\mathbf{r}+\mathbf{x}). \tag{2.172}$$

Let us introduce the vectors $\mathbf{A} = (A_1, A_2, A_3)$ and $\mathbf{j} = (j_1, 0, 0)$. By definition of the cumulant generating function $\left\langle e^{\mathbf{j}\cdot\mathbf{A}} \right\rangle_c$:

$$\left\langle e^{\mathbf{j}\cdot\mathbf{A}} \right\rangle = \exp \left\langle e^{\mathbf{j}\cdot\mathbf{A}} \right\rangle_c. \tag{2.173}$$

Deriving both terms of the equality with respect to j_2 and j_3 and taking $j_2 = j_3 = 0$ yields:

$$\left\langle e^{j_1 A_1} A_2 A_3 \right\rangle = \exp \left\langle e^{j_1 A_1} \right\rangle_c \left[\left\langle e^{j_1 A_1} A_2 A_3 \right\rangle_c + \left\langle e^{j_1 A_1} A_2 \right\rangle_c \left\langle e^{j_1 A_1} A_3 \right\rangle_c \right]. \tag{2.174}$$

The first exponential term is essentially due to virialised motions inside halos and is hence difficult to treat perturbatively. Therefore, it is modelled as a damping term $D_{\rm FoG}(k,\mu,\sigma_v)$,

ignoring the spatial correlation of velocities in A_1. Next, we expand the term in brackets up to second order in j_1 (recalling that $\langle A_2\rangle = \langle A_3\rangle = 0$):

$$\left\langle e^{j_1 A_1} A_2 A_3 \right\rangle_c + \left\langle e^{j_1 A_1} A_2 \right\rangle_c \left\langle e^{j_1 A_1} A_3 \right\rangle_c \simeq \langle A_2 A_3\rangle_c + j_1 \langle A_1 A_2 A_3\rangle_c + j_1^2 \tag{2.175}$$

$$\left[\frac{1}{2}\left\langle A_1^2 A_2 A_3 \right\rangle_c + \langle A_1 A_2\rangle_c \langle A_1 A_3\rangle_c\right]. \tag{2.176}$$

The term $\langle A_1^2 A_2 A_3\rangle_c$ involves the trispectrum, which at tree-order is $\mathcal{O}(P_0^3)$. We ignore it in the following. The term $\langle A_2 A_3\rangle_c$ simply develops into the Kaiser term. Hence, we are left with:

$$P_s(k,\mu) = D_{\rm FoG}(k,\mu,\sigma_v)\left[P_{\delta\delta}(k) + 2f\mu^2 P_{\delta\theta}(k) + f^2\mu^4 P_{\theta\theta}(k) + A(k,\mu,f) + B(k,\mu,f)\right]. \tag{2.177}$$

The A term reads:

$$A(\mathbf{k}) = j_1 \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \langle A_1 A_2 A_3\rangle_c \tag{2.178}$$

$$= -ik\mu f \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \langle [v_z(\mathbf{r}) - v_z(\mathbf{r}+\mathbf{x})]\,[\delta(\mathbf{r}) + f\partial_z v_z(\mathbf{r})]\,[\delta(\mathbf{r}+\mathbf{x}) + f\partial_z v_z(\mathbf{r}+\mathbf{x})]\rangle_c\,. \tag{2.179}$$

We have (using homogeneity, and taking velocity to be curl-free):

$$\langle v_z(\mathbf{r})\,[\delta(\mathbf{r}) + f\partial_z v_z(\mathbf{r})]\,[\delta(\mathbf{r}+\mathbf{x}) + f\partial_z v_z(\mathbf{r}+\mathbf{x})]\rangle_c \tag{2.180}$$

$$= \int \frac{d^3p}{(2\pi)^3}\frac{d^3k}{(2\pi)^3}\frac{d^3q}{(2\pi)^3} \langle v_z(\mathbf{p})\,[\delta(\mathbf{k}) + ifk_z v_z(\mathbf{k})]\,[\delta(\mathbf{q}) + ifq_z v_z(\mathbf{q})]\rangle\, e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.181}$$

$$= -i\int \frac{d^3p}{(2\pi)^3}\frac{d^3k}{(2\pi)^3}\frac{d^3q}{(2\pi)^3}\frac{p_z}{p^2}\left\langle \theta(\mathbf{p})\left[\delta(\mathbf{k}) + f\frac{k_z^2}{k^2}\theta(\mathbf{k})\right]\left[\delta(\mathbf{q}) + f\frac{q_z^2}{q^2}\theta(\mathbf{q})\right]\right\rangle e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.182}$$

$$= -i\int \frac{d^3k}{(2\pi)^3}\frac{d^3p}{(2\pi)^3}\frac{p_z}{p^2} B_\sigma(\mathbf{p},\mathbf{k},-\mathbf{k}-\mathbf{p}) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.183}$$

where:

$$(2\pi)^3\,\delta_D^{(3)}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3) B_\sigma(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = \left\langle \theta(\mathbf{k}_1)\left[\delta(\mathbf{k}_2) + f\frac{k_{2z}}{k_2^2}\theta(\mathbf{k}_2)\right]\left[\delta(\mathbf{k}_3) + f\frac{k_{3z}}{k_3^2}\theta(\mathbf{k}_3)\right]\right\rangle. \tag{2.184}$$

Similarly:

$$\langle v_z(\mathbf{r}+\mathbf{x})\,[\delta(\mathbf{r}) + f\partial_z v_z(\mathbf{r})]\,[\delta(\mathbf{r}+\mathbf{x}) + f\partial_z v_z(\mathbf{r}+\mathbf{x})]\rangle_c = -i\int \frac{d^3k}{(2\pi)^3}\frac{d^3p}{(2\pi)^3}\frac{p_z}{p^2} B_\sigma(\mathbf{p},\mathbf{k}-\mathbf{p},-\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{2.185}$$

Hence:

$$A(\mathbf{k}) = k\mu f \int \frac{d^3p}{(2\pi)^3}\frac{p_z}{p^2}\left[B_\sigma(\mathbf{p},\mathbf{k}-\mathbf{p},-\mathbf{k}) - B_\sigma(\mathbf{p},\mathbf{k},-\mathbf{k}-\mathbf{p})\right]. \tag{2.186}$$

The B term reads:

$$B(\mathbf{k}) = j_1^2 \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \langle A_1 A_2\rangle_c \langle A_1 A_3\rangle_c \tag{2.187}$$

$$= -(k\mu f)^2 \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \langle [v_z(\mathbf{r}) - v_z(\mathbf{r}+\mathbf{x})]\,[\delta(\mathbf{r}) + f\partial_z v_z(\mathbf{r})]\rangle_c \tag{2.188}$$

$$\times \langle [v_z(\mathbf{r}) - v_z(\mathbf{r}+\mathbf{x})]\,[\delta(\mathbf{r}+\mathbf{x}) + f\partial_z v_z(\mathbf{r}+\mathbf{x})]\rangle_c \tag{2.189}$$

We have:

$$\langle v_z(\mathbf{r}+\mathbf{x})\left[\delta(\mathbf{r})+f\partial_z v_z(\mathbf{r})\right]\rangle = \int \frac{d^3k}{(2\pi)^3}\frac{d^3p}{(2\pi)^3}\langle v_z(\mathbf{k})\left[\delta(\mathbf{p})+ifp_z v_z(\mathbf{p})\right]\rangle e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.190}$$

$$= \int \frac{d^3k}{(2\pi)^3}F(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.191}$$

with:

$$F(\mathbf{k}) = \frac{k_z}{k^2}\left[P_{\delta\theta}(k)+f\frac{k_z^2}{k^2}P_{\theta\theta}(k)\right]. \tag{2.192}$$

In addition:

$$\langle v_z(\mathbf{r})\left[\delta(\mathbf{r})+f\partial_z v_z(\mathbf{r})\right]\rangle = \int \frac{d^3k}{(2\pi)^3}F(\mathbf{k}) \tag{2.193}$$

$$= 0 \tag{2.194}$$

since $F(\mathbf{k})$ is odd. The product $\langle A_1A_2\rangle_c\langle A_1A_3\rangle_c$ eventually yields a convolution in Fourier space:

$$B(\mathbf{k}) = (k\mu f)^2\int\frac{d^3p}{(2\pi)^3}F(\mathbf{p})F(\mathbf{k}-\mathbf{p}). \tag{2.195}$$

A and B terms are calculated at 2-loop order using the RegPT scheme, based on the 1-loop bispectrum for A and the 1-loop power spectrum for B.

We usually project the μ dependence onto the basis of Legendre polynomials:

$$P_\ell(k) = \frac{2\ell+1}{2}\int_{-1}^{1}d\mu P(k,\mu)\mathcal{L}_\ell(\mu). \tag{2.196}$$

The A and B term multipoles (monopole $\ell=0$ and quadrupole $\ell=2$) calculated at 1- and 2-loop order using the RegPT scheme are shown in Figure 2.16. One can see that the magnitude of the 2-loop prediction for A is larger than that of the 1-loop results, and baryon acoustic oscillations are significantly damped in the 2-loop compared to the 1-loop case.

Figure 2.17 compares the redshift space power spectrum model of Eq. (2.177) to N-body simulations. We see that linear theory breaks early ($k \lesssim 0.05\,h\,\mathrm{Mpc}^{-1}$), even at $z=1$. The 1-loop monopole prediction already provides percent accurate results up to $k \simeq 0.14\,h\,\mathrm{Mpc}^{-1}$ at $z=1$, a limit pushed to $k \simeq 0.23\,h\,\mathrm{Mpc}^{-1}$ at 2-loop. Two Finger-of-God damping terms are considered:

$$D_{\mathrm{FoG}}(k,\mu,\sigma_v) = \begin{cases} e^{-k^2\mu^2\sigma_v^2} & \text{Gaussian} \quad (2.197a) \\ \left[1+\frac{(k\mu\sigma_v)^2}{2}\right]^{-2} & \text{Lorentzian .} \quad (2.197b) \end{cases}$$

Both give similar results at high redshift; at $z=1$, the Lorentzian form (Cole et al., 1995) (thick lines) seems slightly better. This form, which was shown to better reflect the pairwise velocity distribution seen in the simulations than the standard Gaussian, is the one we adopt in the rest of this manuscript.

Note that at 2-loop, we should not ignore the $\langle A_1^2A_2A_3\rangle_c$ term in Eq. (2.176). Actually, Taruya et al. (2013) showed that the contribution of this term is subdominant and can safely be neglected — which we do in the rest of this manuscript.

As in Section 2.1.7, though a `Fortran` code was available [6], we recoded the A and B calculations in `C`, wrapped in `Python` as part of our `pyregpt` [7] package.

6. at http://www2.yukawa.kyoto-u.ac.jp/~atsushi.taruya/regpt_code.html
7. at https://github.com/adematti/pyregpt

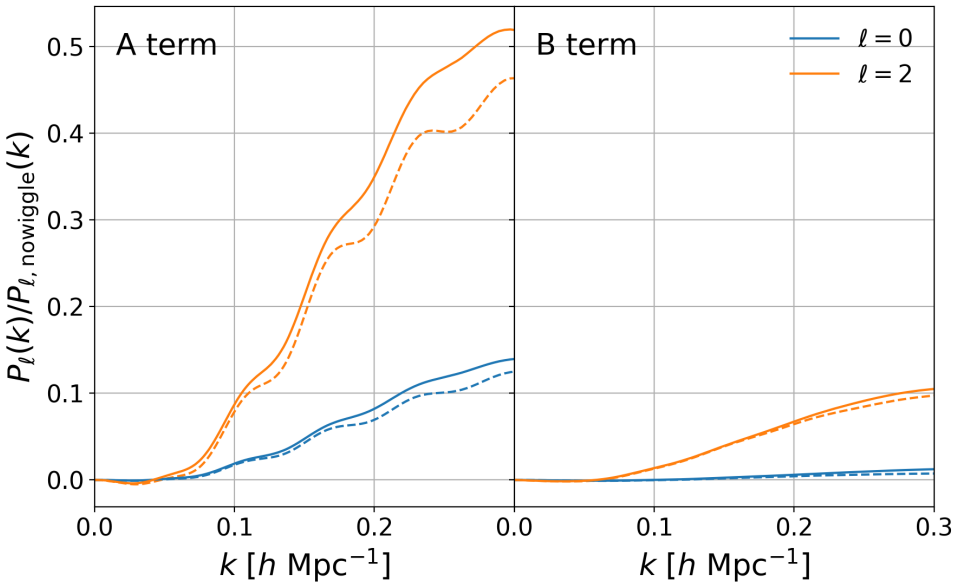

Figure 2.16 – Left: ratio of the A term multipoles (blue: monopole, orange: quadrupole) computed following the RegPT scheme to the no-wiggle power spectrum multipoles using the formula of Eisenstein and Hu (1998). Right: same, for the B term.

2.4.3 Final theoretical model

Let us recap the theoretical model for the redshift space galaxy power spectrum. In the previous subsection we did not consider galaxy bias. Writing that at linear order $\delta = \delta_g/b_1$ and assuming no velocity bias, b_1 can be factored out by replacing f by $\beta = f/b_1$. Hence the contributions of the non-linear RSD terms to the galaxy power spectrum are written: $b_1^3 A(k,\mu,\beta) + b_1^4 B(k,\mu,\beta)$. The final power spectrum model is obtained by collecting the RSD terms and the bias terms:

$$P_g(k,\mu) = D_{\mathrm{FoG}}(k,\mu,\sigma_v)\left[P_{g,\delta\delta}(k) + 2f\mu^2 P_{g,\delta\theta}(k) + f^2\mu^4 P_{\theta\theta}(k) + b_1^3 A(k,\mu,\beta) + b_1^4 B(k,\mu,\beta)\right] \tag{2.198}$$

Galaxy - galaxy and galaxy - velocity power spectra $P_{g,\delta\delta}$ and $P_{g,\delta\theta}$ are given by Eq. (2.150) and Eq. (2.154), respectively. In these formulae, $P_{\delta\delta}$, $P_{\delta\theta}$ and $P_{\theta\theta}$ (in Eq. (2.198)) are calculated at 2-loop order with the RegPT scheme (they correspond to P_{11}, P_{12} and P_{22} in the notations of Section 2.1.7), as is also the case of A and B terms in Eq. (2.198).

We adopt a Lorentzian form for the Finger-of-God effect (2.197b):

$$D_{\mathrm{FoG}}(k,\mu,\sigma_v) = \left[1 + \frac{(k\mu\sigma_v)^2}{2}\right]^{-2}, \tag{2.199}$$

with σ_v the velocity dispersion, which is left as a free parameter.

By default, as discussed at the end of Section 2.3.2, we fix:

$$b_{s2} = -\frac{4}{7}(b_1 - 1), \tag{2.200}$$

$$b_{3\mathrm{nl}} = \frac{32}{315}(b_1 - 1). \tag{2.201}$$

The other bias coefficients, b_1, b_2 and N_g are left free in the fits. Note that at linear order b_1 multiplies the matter power spectrum, and hence is fully degenerate with its normalisation

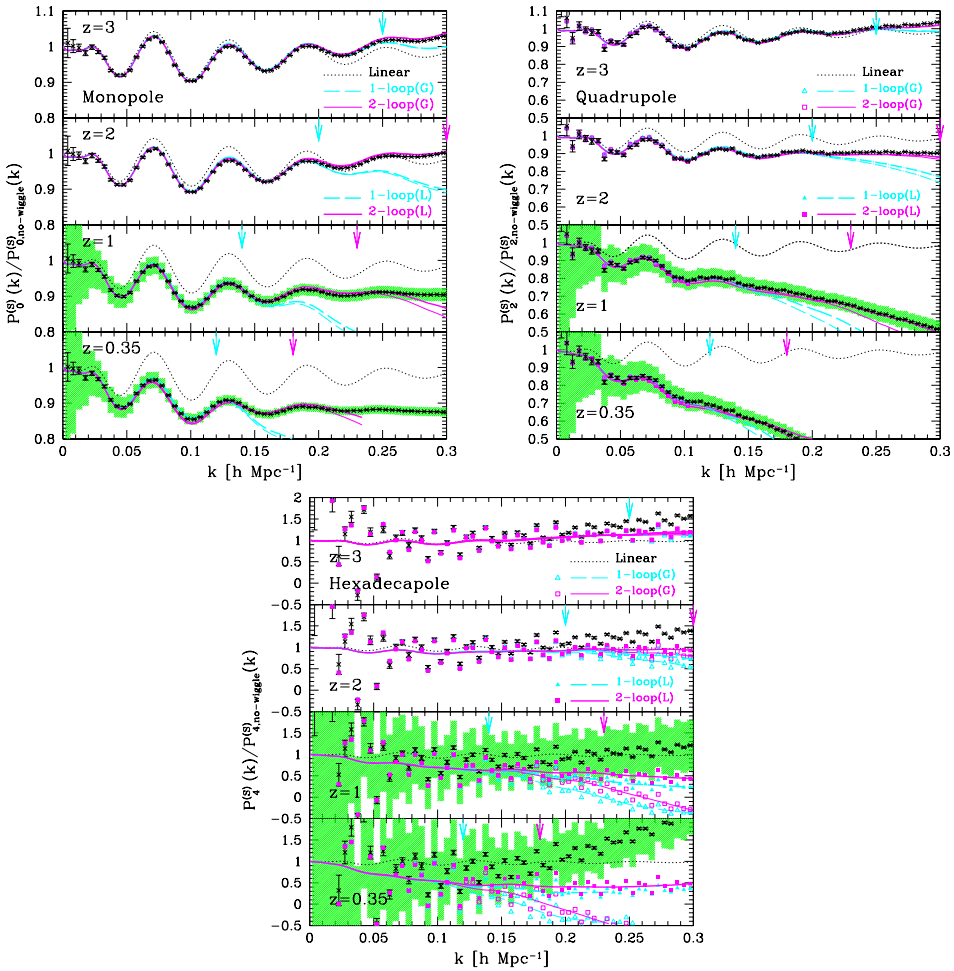

Figure 2.17 – Ratio of the redshift space power spectrum multipoles (left: monopole, right: quadrupole, bottom: hexadecapole) from model (2.177) to the Kaiser formula (2.164) based on the no-wiggle power spectrum of Eisenstein and Hu (1998), for 4 different redshifts (bottom to top: $z = 0.35$, $z = 1$, $z = 2$, $z = 3$). The Kaiser linear power spectrum (2.164) is shown as black dotted lines, 1-loop and 2-loop RegPT-based power spectra (2.177) in dashed blue and continuous magenta lines, respectively. Thin lines use a Gaussian form for the Finger-of-God term while thick lines rely on a Lorentzian function. Points with error bars are measurements from 60 independent N-body simulations of 2048 Mpc h^{-1} box side. Blue (magenta) points in the quadrupole and hexadecapole panels represent the 1(2)-loop model predictions, taking into account the finite grid size used to estimate the power spectrum. Arrows denote the maximum wavenumber where the monopoles of model predictions and N-body measurements are within 1%. The green shaded areas correspond to the expected error bars of a galaxy survey with a volume of 5 Gpc^3 h^{-3}, a density of $5 \times 10^{-4}\, h^3\,$Mpc^{-3} and linear bias of 1 (approximately 1/3 of the statistical uncertainty expected with the eBOSS ELG sample). Taken from Taruya et al. (2013).

σ_8 (see Eq. (1.159)). The same remark applies to f (see Eq. (2.164)) and, to a lesser extent, to b_2. Hence, in the rest of this manuscript, we will quote the combinations $f\sigma_8$, $b_1\sigma_8$ and $b_2\sigma_8$.

Let us finally recall that all perturbation theory calculations (2-loop power spectra and A and B terms, and 1-loop bias terms) are consistently performed with our package `pyregpt` [8]. The input linear matter power spectrum $P_m^{\rm lin}(k)$ is computed with the Boltzmann code CLASS (Blas et al., 2011).

8. at https://github.com/adematti/pyregpt

The above theory model power spectrum is still not quite ready to be compared to the observed data power spectrum — apart from the estimation of the power spectrum itself, other effects, due to the survey geometry, have to be considered as will be discussed in Section 4.2. However, let us first detail in the following chapter what we mean by *survey geometry*, i.e. how we select and observe galaxies in a spectroscopic redshift survey.

Bibliography

Aarseth, S. J. (1963). Dynamical evolution of clusters of galaxies, I. MNRAS *126*, 223.

Angulo, R. E., V. Springel, S. D. M. White, A. Jenkins, C. M. Baugh, et al. (2012). Scaling relations for galaxy clusters in the Millennium-XXL simulation. MNRAS *426*(3), 2046–2062. arXiv:1203.3216.

Aubert, D., C. Pichon, and S. Colombi (2004). The origin and implications of dark matter anisotropic cosmic infall on $\sim L_*$ haloes. MNRAS *352*(2), 376–398. arXiv:astro-ph/0402405.

Baldauf, T., U. Seljak, V. Desjacques, and P. McDonald (2012). Evidence for quadratic tidal tensor bias from the halo bispectrum. Phys. Rev. D *86*(8), 083540. arXiv:1201.4827.

Baugh, C. M. (2013). Luminosity Bias: From Haloes to Galaxies. PASA *30*, e030. arXiv:1302.2768.

Behroozi, P. S., R. H. Wechsler, and H.-Y. Wu (2013). The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. ApJ *762*(2), 109. arXiv:1110.4372.

Bernardeau, F., S. Colombi, E. Gaztañaga, and R. Scoccimarro (2002). Large-scale structure of the Universe and cosmological perturbation theory. Phys. Rep. *367*(1-3), 1–248. arXiv:astro-ph/0112551.

Bernardeau, F., M. Crocce, and R. Scoccimarro (2008). Multipoint propagators in cosmological gravitational instability. Phys. Rev. D *78*(10), 103521. arXiv:astro-ph/0806.2334.

Bernardeau, F., N. van de Rijt, and F. Vernizzi (2012). Resummed propagators in multicomponent cosmic fluids with the eikonal approximation. Phys. Rev. D *85*(6), 063509. arXiv:1109.3400.

Beutler, F., H.-J. Seo, S. Saito, C.-H. Chuang, A. J. Cuesta, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: anisotropic galaxy clustering in Fourier space. MNRAS *466*, 2242–2260. arXiv:1607.03150.

Blas, D., J. Lesgourgues, and T. Tram (2011). The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. J. Cosmology Astropart. Phys. *2011*(7), 034. arXiv:1104.2933.

Bond, J. R., S. Cole, G. Efstathiou, and N. Kaiser (1991). Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations. ApJ *379*, 440.

Boylan-Kolchin, M., V. Springel, S. D. M. White, A. Jenkins, and G. Lemson (2009). Resolving cosmic structure formation with the Millennium-II Simulation. MNRAS *398*(3), 1150–1164. arXiv:0903.3041.

Carrasco, J. J. M., M. P. Hertzberg, and L. Senatore (2012). The effective field theory of cosmological large scale structures. *Journal of High Energy Physics 2012*, 82. arXiv:1206.2926.

Chan, K. C., R. Scoccimarro, and R. K. Sheth (2012). Gravity and large-scale nonlocal bias. Phys. Rev. D *85*(8), 083509. arXiv:1201.3614.

Codis-Decara, S. (2015). *From cosmology to galaxy formation : what can we learn from the large-scale structure of the Universe ?* Theses, Université Pierre et Marie Curie - Paris VI.

Cole, S., K. B. Fisher, and D. H. Weinberg (1995). Constraints on Omega from the IRAS redshift surveys. MNRAS *275*(2), 515–526. arXiv:astro-ph/9412062.

Contreras, S., C. M. Baugh, P. Norberg, and N. Padilla (2013). How robust are predictions of galaxy clustering? MNRAS *432*(4), 2717–2730. arXiv:1301.3497.

Couchman, H. M. P. (1991). Mesh-refined P 3M: A Fast Adaptive N-Body Algorithm. ApJ *368*, L23.

Crain, R. A., T. Theuns, C. Dalla Vecchia, V. R. Eke, C. S. Frenk, et al. (2009). Galaxies-intergalactic medium interaction calculation - I. Galaxy formation as a function of large-scale environment. MNRAS *399*(4), 1773–1794. arXiv:0906.4350.

Crocce, M. and R. Scoccimarro (2006a). Memory of initial conditions in gravitational clustering. Phys. Rev. D *73*(6), 063520. arXiv:astro-ph/0509419.

Crocce, M. and R. Scoccimarro (2006b). Renormalized cosmological perturbation theory. Phys. Rev. D *73*(6), 063519. arXiv:astro-ph/0509418.

Crocce, M., R. Scoccimarro, and F. Bernardeau (2012). MPTBREEZE: a fast renormalized perturbative scheme. MNRAS *427*(3), 2537–2551. arXiv:1207.1465.

Davis, M., G. Efstathiou, C. S. Frenk, and S. D. M. White (1985). The evolution of large-scale structure in a universe dominated by cold dark matter. ApJ *292*, 371–394.

Dehnen, W. (2000). A Very Fast and Momentum-conserving Tree Code. ApJ *536*(1), L39–L42. arXiv:astro-ph/0003209.

Dehnen, W. (2001). Towards optimal softening in three-dimensional N-body codes - I. Minimizing the force error. MNRAS *324*(2), 273–291. arXiv:astro-ph/0011568.

Diemer, B., S. More, and A. V. Kravtsov (2013). The Pseudo-evolution of Halo Mass. ApJ *766*(1), 25. arXiv:1207.0816.

Dubois, Y., S. Peirani, C. Pichon, J. Devriendt, R. Gavazzi, et al. (2016). The HORIZON-AGN simulation: morphological diversity of galaxies promoted by AGN feedback. MNRAS *463*(4), 3948–3964. arXiv:1606.03086.

Eastwood, J. W. and R. W. Hockney (1974). Shaping the force law in two-dimensional particle-mesh models. *Journal of Computational Physics 16*(4), 342–359.

Eisenstein, D. J. and W. Hu (1998). Baryonic Features in the Matter Transfer Function. ApJ *496*(2), 605–614. arXiv:astro-ph/9709112.

Favole, G., J. Comparat, F. Prada, G. Yepes, E. Jullo, et al. (2016). Clustering properties of g-selected galaxies at z $\sim$ 0.8. MNRAS *461*(4), 3421–3431. arXiv:1507.04356.

Gao, L., V. Springel, and S. D. M. White (2005). The age dependence of halo clustering. MNRAS *363*(1), L66–L70. arXiv:astro-ph/0506510.

Gil-Marín, H., J. Noreña, L. Verde, W. J. Percival, C. Wagner, et al. (2015). The power spectrum and bispectrum of SDSS DR11 BOSS galaxies - I. Bias and gravity. MNRAS *451*(1), 539–580. arXiv:1407.5668.

Gonzalez-Perez, V., J. Comparat, P. Norberg, C. M. Baugh, S. Contreras, et al. (2018). The host dark matter haloes of [O II] emitters at 0.5 < z < 1.5. MNRAS *474*(3), 4024–4038. arXiv:1708.07628.

Griffin, A. J., C. G. Lacey, V. Gonzalez-Perez, C. d. P. Lagos, C. M. Baugh, et al. (2019). The evolution of SMBH spin and AGN luminosities for z < 6 within a semi-analytic model of galaxy formation. MNRAS *487*(1), 198–227. arXiv:1806.08370.

Guo, H., X. Yang, A. Raichoor, Z. Zheng, J. Comparat, et al. (2019). Evolution of Star-forming Galaxies from z = 0.7 to 1.2 with eBOSS Emission-line Galaxies. ApJ *871*(2), 147. arXiv:1810.05318.

Guo, Q., S. White, R. E. Angulo, B. Henriques, G. Lemson, et al. (2013). Galaxy formation in WMAP1 and WMAP7 cosmologies. MNRAS *428*(2), 1351–1365. arXiv:1206.0052.

Habib, S., A. Pope, H. Finkel, N. Frontiere, K. Heitmann, et al. (2016). HACC: Simulating sky surveys on state-of-the-art supercomputing architectures. New A *42*, 49–65. arXiv:1410.2805.

Hahn, O. and R. E. Angulo (2016). An adaptively refined phase-space element method for cosmological simulations and collisionless dynamics. MNRAS *455*(1), 1115–1133. arXiv:1501.01959.

Hahn, T. (2005). CUBA—a library for multidimensional numerical integration. *Computer Physics Communications 168*(2), 78–95. arXiv:hep-ph/0404043.

Hand, N., U. Seljak, F. Beutler, and Z. Vlah (2017). Extending the modeling of the anisotropic galaxy power spectrum to k = 0.4 hMpc^{-1}. J. Cosmology Astropart. Phys. *2017*(10), 009. arXiv:1706.02362.

Hearin, A. P., A. R. Zentner, F. C. van den Bosch, D. Campbell, and E. Tollerud (2016). Introducing decorated HODs: modelling assembly bias in the galaxy-halo connection. MNRAS *460*(3), 2552–2570. arXiv:1512.03050.

Heitmann, K., H. Finkel, A. Pope, V. Morozov, N. Frontiere, et al. (2019). The Outer Rim Simulation: A Path to Many-core Supercomputers. ApJS *245*(1), 16. arXiv:1904.11970.

Henriques, B. M. B., S. D. M. White, P. A. Thomas, R. Angulo, Q. Guo, et al. (2015). Galaxy formation in the Planck cosmology - I. Matching the observed evolution of star formation rates, colours and stellar masses. MNRAS *451*(3), 2663–2680. arXiv:1410.0365.

Hernquist, L. (1990). An Analytical Model for Spherical Galaxies and Bulges. ApJ *356*, 359.

Holmberg, E. (1941). On the Clustering Tendencies among the Nebulae. II. a Study of Encounters Between Laboratory Models of Stellar Systems by a New Integration Procedure. ApJ *94*, 385.

Jackson, J. C. (1972). A critique of Rees's theory of primordial gravitational radiation. MNRAS *156*, 1P. arXiv:astro-ph/0810.3908.

Jiménez, E., S. Contreras, N. Padilla, I. Zehavi, C. M. Baugh, et al. (2019). Extensions to the halo occupation distribution model for more accurate clustering predictions. MNRAS *490*(3), 3532–3544. arXiv:1906.04298.

Kaiser, N. (1987). Clustering in real space and in redshift space. MNRAS *227*, 1–21.

Klypin, A., G. Yepes, S. Gottlöber, F. Prada, and S. Heß (2016). MultiDark simulations: the story of dark matter halo concentrations and density profiles. MNRAS *457*(4), 4340–4359. arXiv:1411.4001.

Knebe, A., S. R. Knollmann, S. I. Muldrew, F. R. Pearce, M. A. Aragon-Calvo, et al. (2011). Haloes gone MAD: The Halo-Finder Comparison Project. MNRAS *415*(3), 2293–2318. arXiv:1104.0949.

Lewis, A. and A. Challinor (2011). CAMB: Code for Anisotropies in the Microwave Background.

Lewis, A., A. Challinor, and A. Lasenby (2000). Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. ApJ *538*(2), 473–476. arXiv:astro-ph/9911177.

Linder, E. V. (2005). Cosmic growth history and expansion history. Phys. Rev. D *72*(4), 043529. arXiv:astro-ph/0507263.

Matsubara, T. (2008). Resumming cosmological perturbations via the Lagrangian picture: One-loop results in real space and in redshift space. Phys. Rev. D *77*(6), 063530. arXiv:astro-ph/0711.2521.

McDonald, P. and A. Roy (2009). Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS. J. Cosmology Astropart. Phys. *2009*(8), 020. arXiv:0902.0991.

More, S., B. Diemer, and A. V. Kravtsov (2015). The Splashback Radius as a Physical Halo Boundary and the Growth of Halo Mass. ApJ *810*(1), 36. arXiv:1504.05591.

Moster, B. P., R. S. Somerville, C. Maulbetsch, F. C. van den Bosch, A. V. Macciò, et al. (2010). Constraints on the Relationship between Stellar Mass and Halo Mass at Low and High Redshift. ApJ *710*(2), 903–923. arXiv:0903.4682.

Nagai, D. and A. V. Kravtsov (2005). The Radial Distribution of Galaxies in Λ Cold Dark Matter Clusters. ApJ *618*(2), 557–568. arXiv:astro-ph/0408273.

Navarro, J. F., C. S. Frenk, and S. D. M. White (1996). The Structure of Cold Dark Matter Halos. ApJ *462*, 563. arXiv:astro-ph/9508025.

Okamura, T., A. Taruya, and T. Matsubara (2011). Next-to-leading resummation of cosmological perturbations via the Lagrangian picture: 2-loop correction in real and redshift spaces. J. Cosmology Astropart. Phys. *2011*(8), 012. arXiv:1105.1491.

Orsi, Á. A. and R. E. Angulo (2018). The impact of galaxy formation on satellite kinematics and redshift-space distortions. MNRAS *475*(2), 2530–2544. arXiv:1708.00956.

Peacock, J. A. and S. J. Dodds (1994). Reconstructing the Linear Power Spectrum of Cosmological Mass Fluctuations. MNRAS *267*, 1020. arXiv:astro-ph/9311057.

Peacock, J. A. and S. J. Dodds (1996). Non-linear evolution of cosmological power spectra. MNRAS *280*(3), L19–L26. arXiv:astro-ph/9603031.

Press, W. H. and P. Schechter (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. ApJ *187*, 425–438.

Saito, S., T. Baldauf, Z. Vlah, U. Seljak, T. Okumura, et al. (2014). Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum. Phys. Rev. D *90*(12), 123522. arXiv:1405.1447.

Scannapieco, C., M. Wadepuhl, O. H. Parry, J. F. Navarro, A. Jenkins, et al. (2012). The Aquila comparison project: the effects of feedback and numerical methods on simulations of galaxy formation. MNRAS *423*(2), 1726–1749. arXiv:1112.0315.

Seljak, U. (2000). Analytic model for galaxy and dark matter clustering. MNRAS *318*(1), 203–213. arXiv:astro-ph/0001493.

Sheth, R. K. and G. Tormen (1999). Large-scale bias and the peak background split. MNRAS *308*(1), 119–126. arXiv:astro-ph/9901122.

Silk, J. and M. J. Rees (1998). Quasars and galaxy formation. A&A *331*, L1–L4. arXiv:astro-ph/9801013.

Simha, V., D. H. Weinberg, R. Davé, M. Fardal, N. Katz, et al. (2012). Testing subhalo abundance matching in cosmological smoothed particle hydrodynamics simulations. MNRAS *423*(4), 3458–3473. arXiv:1011.4964.

Sousbie, T. and S. Colombi (2016). ColDICE: A parallel Vlasov-Poisson solver using moving adaptive simplicial tessellation. *Journal of Computational Physics 321*, 644–697. arXiv:physics.comp-ph/1509.07720.

Springel, V. (2005). The cosmological simulation code GADGET-2. MNRAS *364*(4), 1105–1134. arXiv:astro-ph/0505010.

Springel, V., S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, et al. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. Nature *435*(7042), 629–636. arXiv:astro-ph/0504097.

Taruya, A., F. Bernardeau, T. Nishimichi, and S. Codis (2012). Direct and fast calculation of regularized cosmological power spectrum at two-loop order. Phys. Rev. D *86*(10), 103528. arXiv:1208.1191.

Taruya, A. and T. Hiramatsu (2008). A Closure Theory for Nonlinear Evolution of Cosmological Power Spectra. ApJ *674*(2), 617–635. arXiv:astro-ph/0708.1367.

Taruya, A., T. Nishimichi, and F. Bernardeau (2013). Precision modeling of redshift-space distortions from a multipoint propagator expansion. Phys. Rev. D *87*(8), 083509. arXiv:1301.3624.

Taruya, A., T. Nishimichi, and S. Saito (2010). Baryon acoustic oscillations in 2D: Modeling redshift-space power spectrum from perturbation theory. Phys. Rev. D *82*(6), 063522. arXiv:1006.0699.

Taruya, A., T. Nishimichi, S. Saito, and T. Hiramatsu (2009). Nonlinear evolution of baryon acoustic oscillations from improved perturbation theory in real and redshift spaces. Phys. Rev. D *80*(12), 123503. arXiv:0906.0507.

Teyssier, R. (2002). Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. A&A *385*, 337–364. arXiv:astro-ph/0111367.

Uhlemann, C., S. Codis, C. Pichon, F. Bernardeau, and P. Reimberg (2016). Back in the saddle: large-deviation statistics of the cosmic log-density field. MNRAS *460*(2), 1529–1541. arXiv:1512.05793.

Wechsler, R. H., J. S. Bullock, J. R. Primack, A. V. Kravtsov, and A. Dekel (2002). Concentrations of Dark Halos from Their Assembly Histories. ApJ *568*(1), 52–70. arXiv:astro-ph/0108151.

Wechsler, R. H. and J. L. Tinker (2018). The Connection Between Galaxies and Their Dark Matter Halos. ARA&A *56*, 435–487. arXiv:1804.03097.

White, S. D. M. (1994). Formation and Evolution of Galaxies: Les Houches Lectures. *arXiv e-prints*, 9410043. arXiv:astro-ph/9410043.

White, S. D. M. and M. J. Rees (1978). Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. MNRAS *183*, 341–358.

Zel'dovich, Y. B. (1970). Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations. A&A *500*, 13–18.

The survey design

Galaxy redshift surveys of the SDSS proceed in two steps: first, *targets* are selected from a *photometric survey*; then spectra of these targets are observed as part of a *spectroscopic survey*. Both photometric and spectroscopic surveys are conducted using the SDSS telescope, equipped with either cameras (with different filters) for the photometric part or spectrographs for the spectroscopic operations. We first describe the SDSS telescope, camera and eBOSS spectrographs in Section 3.1. Contrary to other eBOSS spectroscopic samples, the ELG sample considered in this manuscript is based on a specific photometric survey, DECaLS, which we review, as well as the target selection, in Section 3.2. Finally, the spectroscopic operations, using the SDSS telescope equipped with the eBOSS spectrographs, are described in Section 3.3.

3.1 The SDSS instrument

The Sloan Digital Sky Survey (SDSS, York et al., 2000) was designed in the mid 1980's, as a wide-area multi-band imaging and spectroscopic survey, with the primary purpose to understand the large scale structure of the Universe. This project had been made possible by the recent developments in CCD image sensors, computational processing and instrument control. Covering a fourth of the celestial sphere ($10\,000\,\mathrm{deg}^2$), the SDSS was planned to outperform existing spectroscopic surveys by one magnitude, allowing redshifts of 10^6 galaxies and 10^5 quasars to be measured. The SDSS commissioning phase started in 1998, and survey operations began in May 2000.

Let us first briefly recap in Section 3.1.1 the SDSS telescope and setup for both photometric and spectroscopic observations, as it serves as a basis for the spectroscopic operations of the eBOSS program as described in Section 3.1.2.

This section mainly relies on York et al. (2000), Gunn et al. (2006) and Smee et al. (2013).

3.1.1 The SDSS setup

A dedicated Ritchey-Chrétien, 2.5 m telescope (Gunn et al., 2006) was built by the SDSS collaboration to perform both the imaging and spectroscopic surveys. Located at Apache Point Observatory (APO), Sunspot, New Mexico, its peculiar design, with its rolling enclosure and baffle system, is displayed in Figure 3.1 (left). Its large field of view (3° diameter) was chosen to cover the entire survey in a reasonable amount of time. The imaging camera is mounted at the Cassegrain focus, and can be easily replaced by a fibre plug plate linked to two fibre-fed spectrographs for spectroscopic observations. The best observing conditions are reserved for the imaging survey, while less ideal (moonless) conditions are dedicated to the spectroscopic survey. Besides the main telescope, a 0.5 m Photometric Telescope (PT) is used to calibrate the imaging survey, while a seeing monitor and cloud scanner record observing conditions.

3.1.1.1 Photometric setup

The imaging camera (shown in Figure 3.1, right) uses an array of CCDs of 2048×2048 24 µm-wide pixels. The CCDs are organised in 5 rows and 6 columns. Each row is assigned a different filter: r, i, u, z, g, of effective central wavelengths 3590 Å, 4810 Å, 6230 Å, 7640 Å and 9060 Å. The imaging survey was conducted following a drift-scan strategy: the imaging camera sweeps the sky in the direction parallel to the CCD columns, such that one point on the sky is observed successively through the 5 filters of the 5 CCD rows. The effective integration time for the 5 CCD rows, scanning a stripe of $2.33° \times 0.22°$ on the sky, is 5.7 minutes. A second scan,

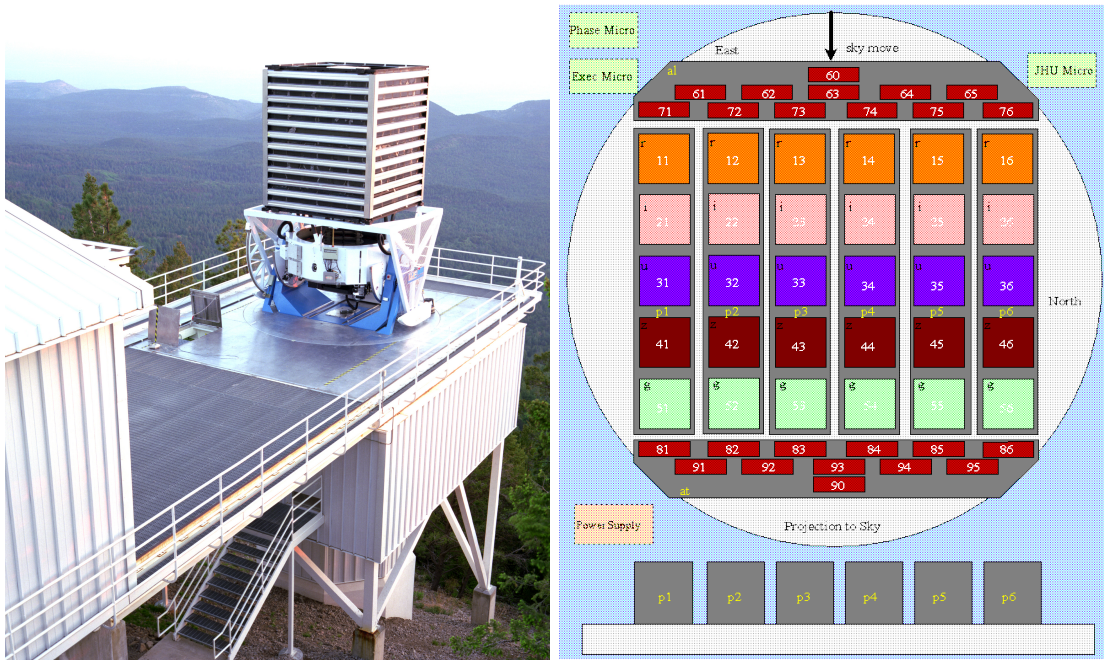

Figure 3.1 – Left: SDSS telescope, at Apache Point Observatory, New Mexico. The rolling enclosure is on the left; the wind baffle system is mounted onto the telescope on the right. Right: sketch of the SDSS camera: the 6 columns and 5 rows (with filters r, i, u, z, g) of 2048×2048 imaging CDDs. Narrow astrometric 128×2048 CCDs (in red) are located above and below photometric CCDs. Credits: http://skyserver.sdss.org.

shifted by 0.21° in the row direction, fills the unobserved stripes left by the 5.8′ gaps between CCD columns. Narrow r-filtered 128×2048 CCDs located around the photometric CCDs are used to perform astrometric calibration, covering the dynamic range ($r \simeq 8.5 - 16.8$) between stars of the fundamental astrometric catalogues (Hipparcos and Tycho, ESA, 1997) and the brightest unsaturated stars of the photometric survey. The PT provides photometric calibration patches, themselves calibrated with respect to 157 primary standards of the United States Naval Observatory 1 m telescope.

3.1.1.2 Spectroscopic setup

SDSS spectroscopic observations were conducted using two multi-object fibre-fed twin spectrographs measuring spectra from the near ultraviolet to the near infrared (Smee et al., 2013). Light was guided from the focal plane to the two spectrographs by 640 fibre optic strands, 180 µm in diameter (3″ on the sky). The fibre diameter was chosen to maximise the signal-to-noise ratio for galaxies at the typical redshift $z = 0.1$ probed by the SDSS. The number of fibres, and thus the number of spectra measured during a single exposure, was determined based on the target density ($120\,\mathrm{deg}^{-2}$) and the covered field-of-view ($5\,\mathrm{deg}^2$), keeping spare fibres for calibration purposes or ancillary programs. The number of fibres was limited by the number of pixels of the spectrograph CCDs (2048×2048), ensuring that each spectrum can cover 3 pixels, with additional 3 pixel gaps to avoid crosstalk between different spectra. The silica fibres are plugged into an aluminium plate, 3.2 mm thick and 800 mm in diameter, to be positioned at the focal plane of the telescope. Bending momenta are applied at the edges of the plate and through its centre post to make it match the slightly curved (by up to 2.12°) best-focus surface. The plate is

thick enough to ensure fibres stay orthogonal to the plug plate, and thus aligned with incoming light rays. Fibres are grouped by harnesses of 20 fibres, terminated by a v-groove block. The 16 v-groove blocks (holding the $16 \times 20 = 320$ fibres for each spectrograph) are glued together to an aluminium slitplate with a 640 mm radius of curvature which enters the spectrograph.

The two spectrographs are directly mounted on the rear of the telescope — instead of being installed on a separate, fixed bench — to keep fibres short for maximising throughput, minimising its variations due to repeated stress and fibre motions and keeping the focal ratio degradation low. The plug plates, fibre harnesses and slitheads are mounted on an aluminium cartridge, that can be installed and removed by a single operator in a few minutes (see Figure 3.2, left). Eight cartridges were fabricated by the SDSS; they are prepared during the day for each pointing of the following night of observing. Such a preparation includes plugging fibres by hand in holes of the aluminium plates within predefined area. A fibre mapper then illuminates sequentially the different fibres at the slithead end and record their positions on the focal plane through a narrow-band camera. This operation, which takes about 5 minutes, is also useful to identify broken or unplugged fibres.

3.1.1.3 The spectrograph design

The spectroscopic resolving power R is defined as the wavelength of interest divided by the full width at half maximum (FWHM) of the fibre image on the CCD (in wavelength units). The minimum resolving power was set such that the redshift accuracy is limited by the typical galaxy velocity dispersion $\simeq 100$ to $200\,\mathrm{km\,s^{-1}}$, corresponding do $R \simeq 1500 - 3000$. In parallel, the spectrograph wavelength range was set to $3900\,\text{Å} - 9100\,\text{Å}$ to encompass observations of the H and K absorption lines of CaII down to $z = 0$ and that of the [OII] doublet from $z > 0.05$ and to allow the Hα emission line to be observed to a redshift of $z > 0.2$ and quasars up to $z \sim 5$. The extension of the wavelength range deeper in the NIR allowed to push up the limit for redshift measurements of LRGs using the 4000 Å break, serendipitously allowing the first observation of the baryon acoustic oscillation feature by the SDSS (Eisenstein et al., 2005). Given the spectrograph 2048×2048 CCDs (with 24 µm-wide pixels), it appeared necessary to use two channels, blue ($\lambda < 6000\,\text{Å}$) and red ($\lambda > 6000\,\text{Å}$) for each spectrograph. As shown in Figure 3.2 (right), light enters each spectrograph through the fibres glued on the slitplate concentric to a mirror which reflects the light in a collimated beam. A dichroic beam splitter reflects the blue part ($\lambda < 6000\,\text{Å}$) of the collimated beam (with efficiency 98%) and transmits the red light (with efficiency 94%). The two beams are dispersed through a grism with ruling densities of 640 and 440 lines $\mathrm{mm^{-1}}$ for the blue and red channels, respectively, before entering the two cameras. The cameras are cooled down by a dewar system, including 0.6 L reservoirs mounted on the telescope, refilled every hour by 10 L intermediate reservoirs, themselves connected during the day to a 180 L dewar installed on the telescope platform.

3.1.1.4 Performance of the spectroscopic system

The simulated resolving power is shown in Figure 3.3. A circular source (with diameter equal to that of the optical fibres) was placed in the slitplate and the resolution was measured as the FWHM of the image captured by the CCD, assuming perfect optics. Note however that raw spectra are $\simeq 3$ pixel wide in the spatial direction, and are thus collapsed in one-dimensional spectra for further analysis (see Section 3.3.3.1). The true resolving power, with unperfect optics but based on collapsed spectra, is thus expected to be slightly higher. In order to ensure reliable

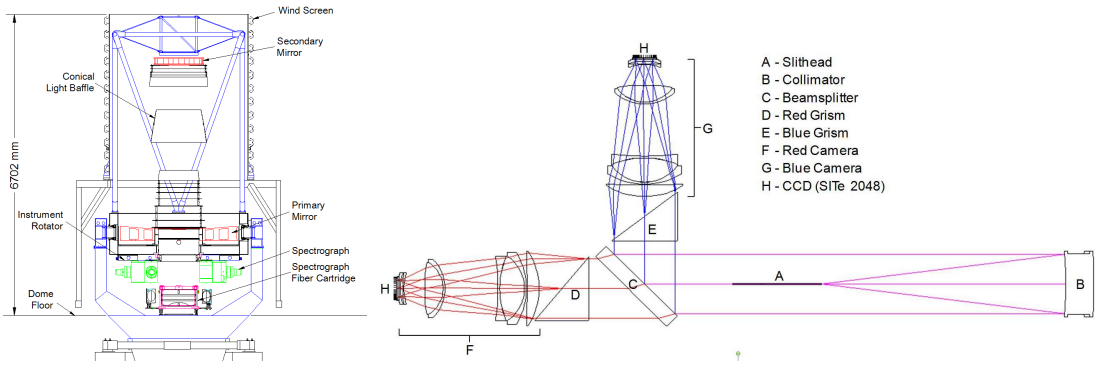

Figure 3.2 – Left: sketch of the SDSS telescope, with the two 320 kg spectrographs mounted at the back of the telescope rotator. The fibre cartridge, here shown dismounted, is placed between the two spectrographs in operation. Right: optical layout of the twin SDSS spectrographs. The light beam, guided by optical fibres, enters through the slithead (A), is reflected by the collimator (B), split between the red and blue channels by the dichroic (C), and dispersed by the red (D) or blue (E) grism before entering the cameras (F and G). Taken from Smee et al. (2013).

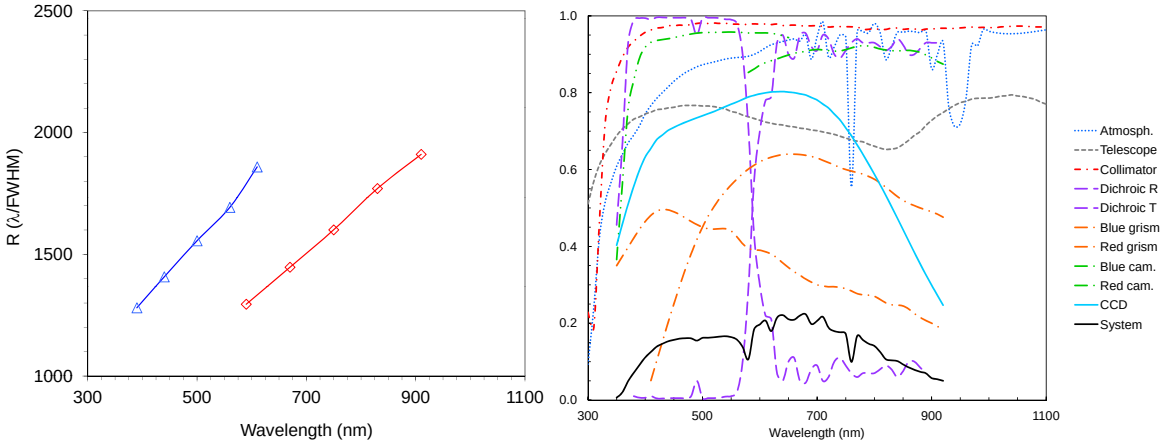

Figure 3.3 – Left: simulated resolving power R of the blue and red channels of the SDSS spectrographs. R is expected to be slightly underestimated. Right: predicted throughput of the full SDSS installation, including sky extinction. Taken from Smee et al. (2013).

redshift measurements even for galaxy spectra lacking strong absorption and emission lines, the required signal-to-noise was set to $15\,\text{Å}^{-1/2}$ over the whole wavelength range. This translates into a lower bound for the total throughput from 10% to 17% depending on the wavelength. Figure 3.3 (left) displays the predicted throughput, from the atmospheric extinction to the CCD. The (measured) transmission of the blue and red grisms is the main driver to the total throughput, which meets the requirements.

The SDSS (phase I), and its 3-year extension SDSS-II, ended in 2008. The Astrophysical Research Consortium, owning and operating APO, called for projects and awarded the Baryon Oscillation Spectroscopic Survey (BOSS, Dawson et al., 2013) five years of dark observing time, starting in 2009.

3.1.2 The BOSS/eBOSS spectrographs

BOSS primary purpose was to provide percent level measurement of the BAO signal by probing 1.5 million luminous galaxies at redshift $z < 0.7$. In addition, BOSS pioneered a new technique to map large scale structure by probing neutral hydrogen absorption in quasar spectra — the so-called Lyman-α forest. To this end, BOSS measured more than 150 000 quasar spectra in the redshift range $2.15 < z < 3.5$.

3.1.2.1 Technical requirements

To meet these scientific requirements in a 5-year program, the BOSS spectrograph had to measure 35% more spectra, per unit time, of objects that are 1 magnitude fainter than in SDSS-I and II. Therefore, both the number of fibres and the total throughput had to be increased. The number of fibres was increased to 1000 (500 per spectrograph), and their diameter reduced to 120 µm, corresponding to the typical size of a $z \simeq 0.7$ galaxy image on the focal plane ($2''$). The minimum wavelength was decreased to 3560 Å to measure half of the Lyman-α forest between the Lyman-α and Lyman-β transitions (1216 Å and 1026 Å) at $z \gtrsim 2.2$. The redder wavelength was increased to 10 400 Å to include the CaII K and H absorption lines, which mainly drive the redshift determination of luminous red galaxies, up to redshift $z \simeq 0.8$. The resolving power was set to 1400 in the range $3800\,\text{Å} < \lambda < 4900\,\text{Å}$ to ensure reliable classification of calibration stars through the Balmer series. In the other part of the wavelength range, the required resolving power was set to 1000, corresponding to redshift errors of $300\,\text{km}\,\text{s}^{-1}$.

3.1.2.2 Instrument upgrades

The increase of the number of fibres and wavelength range was made possible by the upgrade of the SDSS spectrograph CCDs to 4096×4096 CCD arrays of 15 µm pixels. The beamsplitter was changed to cover the extended passband. Based on previous SDSS experience, a signal-to-noise ratio of $3-4$ per angstrom was required for the faintest targets. To meet this requirement with one hour exposure, the peak throughput had to be doubled with respect to SDSS. This was enabled by the higher quantum CCD efficiency and by upgrading the grism, responsible for most of the throughput loss in SDSS-I and SDSS-II (see Figure 3.3, right) to volume-phase holographic (VPH) gratings (with ruling densities of 520 and 400 lines mm^{-1} for the blue and red channels, respectively).

3.1.2.3 Performance of the spectroscopic system

Figure 3.4 shows the measured variations in the resolving power as a function of wavelength. The trough around $\simeq 6000$ Å reveals the region where the dichroic separates the blue and red channels. In this region, the resolving power is taken as the mean of the blue and red cameras, weighted by their respective throughput. The resolving power is given for central fibres (in the slitplate and corresponding location on the CCDs) in green, and fibres located on the edges in grey. The measured resolving power meets the requirements (shown in dashed lines) in the case of central fibres, but for edge fibres the resolving power significantly decreases in the blue end, especially for spectrograph 1. These variations are understood as the consequence of differences in the optical alignment, which are enhanced at blue wavelengths and large field angles between the two spectrographs.

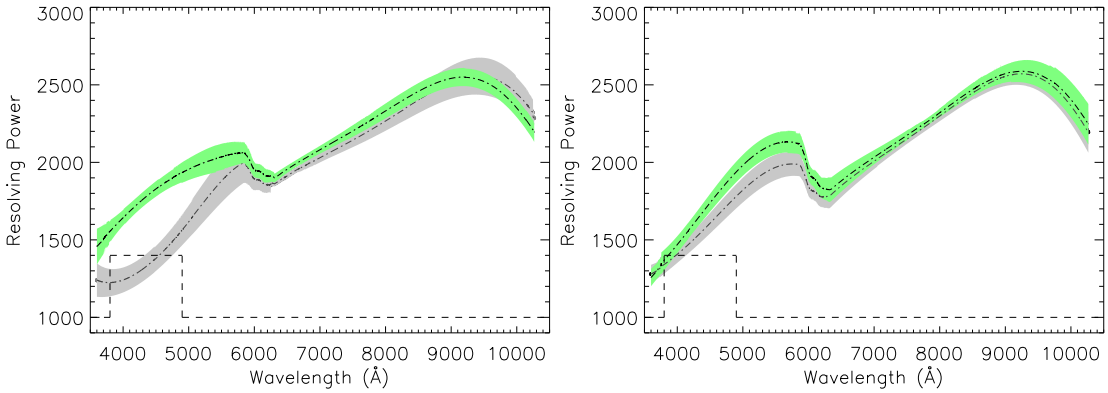

Figure 3.4 – Left: measured resolving power in BOSS spectrograph 1 for central fibres (green, representative of 80% of the fibres) and edge fibres (grey). Right: BOSS spectrograph 2. Requirement is in dashed lines. Taken from Smee et al. (2013).

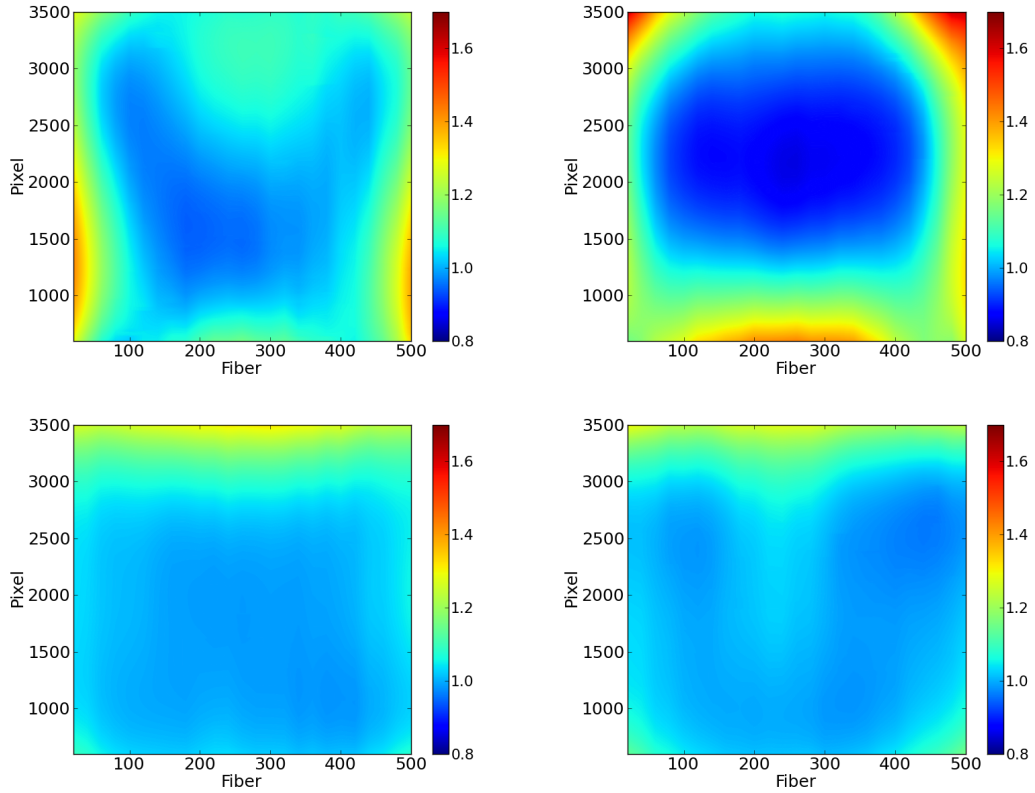

Figure 3.5 – The pixel dispersion (RMS) of the line spread function as a function of the position on the CDD (top: blue channels, bottom: red channels, left: spectrograph 1, right: spectrograph 2). Higher wavelength are at the top of the CCDs. Taken from Smee et al. (2013).

Corresponding trends are visible in Figure 3.5, which displays the measured variations of the dispersion σ_p of the line spread function in terms of pixels as a function of the position on the CCDs. The resolution degrades noticeably on the edges of the blue CCDs; indeed, these regions receive light from fibres which are mostly plugged on the edges of the focal plane, where deviation to flatness is the most important. In addition, the redder wavelengths in the red CCDs suffer from the increased path length in the optical fibres.

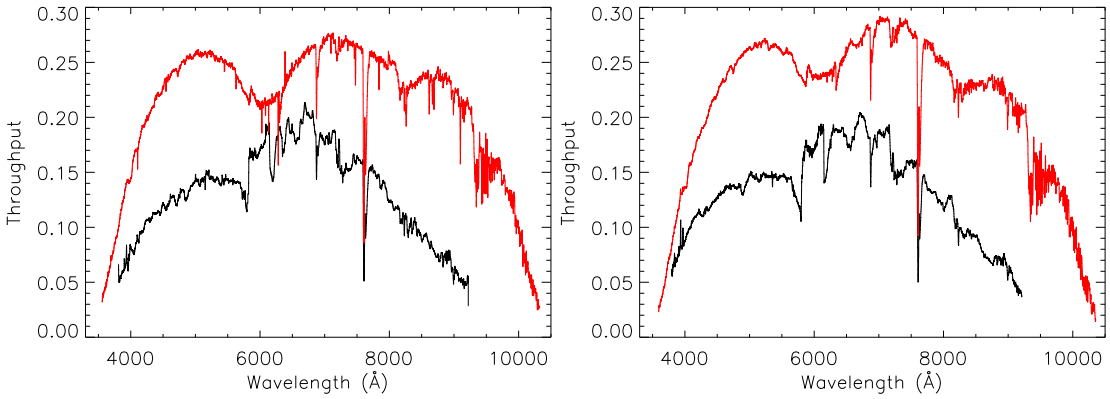

Figure 3.6 – Throughput of spectrograph 1 (left) and 2 (right) for SDSS (black) and BOSS (red). Taken from Smee et al. (2013).

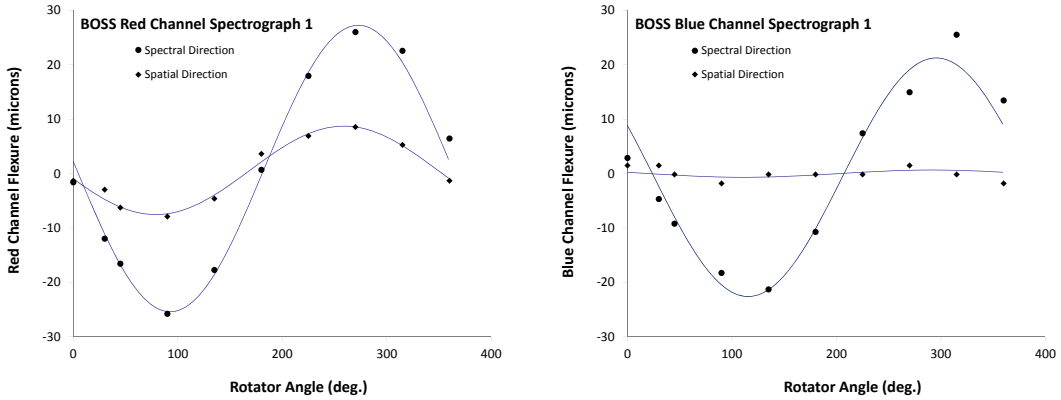

Figure 3.7 – Image shifts due to varying gravity-induced flexure as a function of the rotator angle, in the red (left) and blue (right) channels of spectrograph 1, for a fixed zenith angle of 72°. A similar behaviour is seen for spectrograph 2. Taken from Smee et al. (2013).

The throughput of the BOSS instrument is much higher than that of SDSS, as shown in Figure 3.6; the improvement is a factor of two at peak throughput. One can notice, again, the trough around the dichroic at $\simeq 6000$ Å. Spectrograph 2 shows a slightly higher throughput than spectrograph 1, especially in the dichroic region (15%), while the difference in peak throughput is $\simeq 5\%$. Not shown here are variations of the per-fibre throughput: consistently with previous remarks on the resolving power, fibres near the edges of the spectrographs receive slightly less throughput due to the lower optical quality at large field angle.

The gravity-induced deformation of the instrument changes with the orientation of the telescope during pointing (due to rotating Earth). This leads to shifts in the image on CCDs in both the spatial and spectral directions. Motions in the spatial direction increase cross-talk between different spectra; in the spectral direction, they degrade the resolution. As shown in Figure 3.7, the variations in the spatial direction are much smaller, as a consequence of the rectangular shape of the collimator (taller than wide) and its installation. The rotation of the telescope in an one-hour exposure is 15°, leading to maximum flexures of 6.2 µm and 6.6 µm along the spectral direction for spectrograph 1 and 2, respectively; given the pixel size (15 µm), these correspond to 0.41 and 0.44 pixels, respectively, lower than the requirement of 0.5 pixels.

BOSS spectrographs were reused by eBOSS (Dawson et al., 2016), the 5-year extension of BOSS. eBOSS was designed to allow clustering measurements at higher redshifts, with 280 000

new LRGs in the redshift range $0.6 < z < 1.0$, 170 000 ELG redshifts in $0.6 < z < 1.1$ and 330 000 quasars (QSOs) in $0.8 < z < 2.2$. The number of high redshift $z > 2.1$ Lyman-α quasars was also increased by 60 000. The final data release (DR) of eBOSS, which the ELG sample discussed in this manuscript is part of, is DR16.

3.2 Photometric survey

The BOSS/eBOSS spectrographs, installed on the SDSS telescope were used to conduct the spectroscopic observations of the eBOSS ELG sample. However, ELG targets for spectroscopic follow-up were not selected on images taken by the SDSS camera but came from the DECaLS survey, based on images taken by DECam, which we describe below. Note that apart from selecting targets for spectroscopic follow-up, deep imaging surveys can be used in multiple science cases, e.g. study of clusters, of galaxy properties (luminosity, colour, stellar mass, star formation rate...) and their impact on transient events such as supernovae, study of the Milky Way and its halo. Before heading to the details of the DECaLS survey in Section 3.2.2 and the target selection in Section 3.2.3, let us first give a few definitions of photometric observables.

This section borrows a lot from Dey et al. (2019).

3.2.1 A few definitions

The magnitude is the unitless measure of the object luminosity in a given passband. The difference of magnitude between two objects of brightness I_1 and I_2 is:

$$m_2 - m_1 = -2.5 \log_{10} \left(\frac{I_2}{I_1} \right). \tag{3.1}$$

AB magnitudes are an absolute reference magnitude system. Denoting f_ν the spectral flux density at frequency ν, the corresponding monochromatic AB magnitude is defined as:

$$m_{\rm AB} = -2.5 \log_{10} \left(\frac{f_\nu}{1\,{\rm erg\,s^{-1}\,Hz^{-1}\,cm^{-2}}} \right) - 48.60 \simeq -2.5 \log_{10} \left(\frac{f_\nu}{3631\,{\rm Jy}} \right), \tag{3.2}$$

where $1\,{\rm erg} = 10^{-7}\,{\rm W\,s}$ and $1\,{\rm Jy} = 10^{-26}\,{\rm W\,Hz^{-1}\,m^{-2}}$. In a frequency range, the bandpass AB magnitude is defined as:

$$m_{\rm AB} = -2.5 \log_{10} \left(\frac{\int f_\nu \, (h\nu)^{-1} \, e(\nu) d\nu}{\int (h\nu)^{-1} \, e(\nu) d\nu \times {\rm erg\,s^{-1}\,Hz^{-1}\,cm^{-2}}} \right) - 48.60, \tag{3.3}$$

where h is the Planck constant, $e(\nu)$ is the equal energy filter response function.

The galactic extinction modifies the apparent magnitude of an extragalactic object depending on its angular position on the sky. Therefore, magnitudes are often corrected by the galactic extinction — as is the case in this manuscript. The correction is based on the galactic E(B-V) extinction map of Schlegel et al. (1998), which is propagated through the DECam filters by the coefficients of Schlafly and Finkbeiner (2011).

We express the imaging depth as the AB magnitude of a point source object detected at 5σ. In practice, it is derived from the errors measured by the source detection pipeline called `The Tractor`. As in this manuscript, it is often corrected for the Galactic extinction.

The seeing is the FWHM of the point spread function (PSF) of a point source object, e.g. a distant star.

3.2.2 DECaLS

Contrary to other BOSS and eBOSS surveys making use of the SDSS-I-II-III optical imaging data (and near-infrared data from the Wide-field Infrared Survey Explorer (WISE), Wright et al. 2010), eBOSS ELG targets were selected from data releases 3 and 5 of the Dark Energy Camera Legacy Survey (DECaLS, Dey et al., 2019). This choice was motivated by the DECaLS photometry being at least one magnitude deeper than the SDSS imaging in all g, r, z bands, which allowed selecting fainter targets. The final DECaLS release covers 9000 deg^2 in the North (NGC) and South (SGC) galactic caps, based on images captured by the Dark Energy Camera (DECam) mounted on the Blanco telescope at the Cerro Tololo Inter-American Observatory (CTIO), Chile. DECam is also used for the DES imaging survey (The Dark Energy Survey Collaboration, 2005), covering 5000 deg^2 in the SGC. Most of the DECaLS footprint is covered by images taken as part of the "The DECam Legacy Survey of the SDSS Equatorial Sky" program (PI: D. Schlegel and A. Dey) including 64 observing nights over 3 years, later extended to 157 scheduled nights. In addition, a 1130 deg^2 part of the SGC DECaLS footprint (including the eBOSS ELG SGC footprint) uses DES raw images.

3.2.2.1 Technical requirements

DECaLS is part of the Legacy Surveys which are used to select targets for the next generation survey built on the Dark Energy Spectroscopic Instrument (DESI, DESI Collaboration et al., 2016). Other Legacy Surveys are BASS (Zou et al., 2017) in the g and r bands and MzLS in the z band, using images respectively taken by the Bok 90-inch and the Mayall 4 m telescope at Kitt Peak, Arizona, in the North galactic sky (5000 deg^2, Dec. $> 32°$). These surveys are completed by WISE imaging in the near-infrared 3.4 µm (W1) and 4.6 µm (W2) filters, critical to the LRG and QSO DESI target selections. The imaging depth requirements are 5σ detections of a fiducial $g = 24.0$, $r = 23.4$ and $z = 22.5$ AB mag galaxy with an exponential light profile of half-light radius $r_{\rm half} = 0.45''$. The seeing must be better than $1.5''$ FWHM in the z band. The systematic errors in astrometric calibration must be less than 30 milliarcseconds and the random errors less than 95 milliarcseconds, so as to ensure a good positioning of spectroscopic fibres of the DESI instrument. Astrometric calibration is controlled with Gaia (Gaia Collaboration et al., 2016). Coverage requirements include a fill factor (ratio of the full depth coverage to the total footprint) of 90% and non-photometric observations (i.e. when the root mean square of magnitude measurements is larger than 1% in the g and r bands and 2% in the z band) cover regions smaller than 3° in diameter each.

3.2.2.2 DECam

DECam was built by the Dark Energy Survey collaboration (Flaugher et al., 2015) and mounted at CTIO in 2011 and 2012, with first light in September 2012. DECam uses a 5 fused silica lens optical system, to ensure a good imaging quality over the wide 3.18 deg^2 field-of-view and high throughput over the wavelength range 400−1000 nm. DECam includes 8 filters (see Figure 3.8, left), with very high transmission and tight uniformity over their 620 mm diameter. The detector consists in 62 imaging CCDs (59 operational) of 2048×4096 pixels arranged in a roughly hexagonal shape, and 12 2048×2048 CCDs for guiding and focus (see Figure 3.8, right). CCDs are back-illuminated, with 250 µm thick pixels of size 15 µm ($0.262''$ on the sky), resulting in a high quantum efficiency of 70%, 90%, 90% and 75% in the g, r, i and z bands.

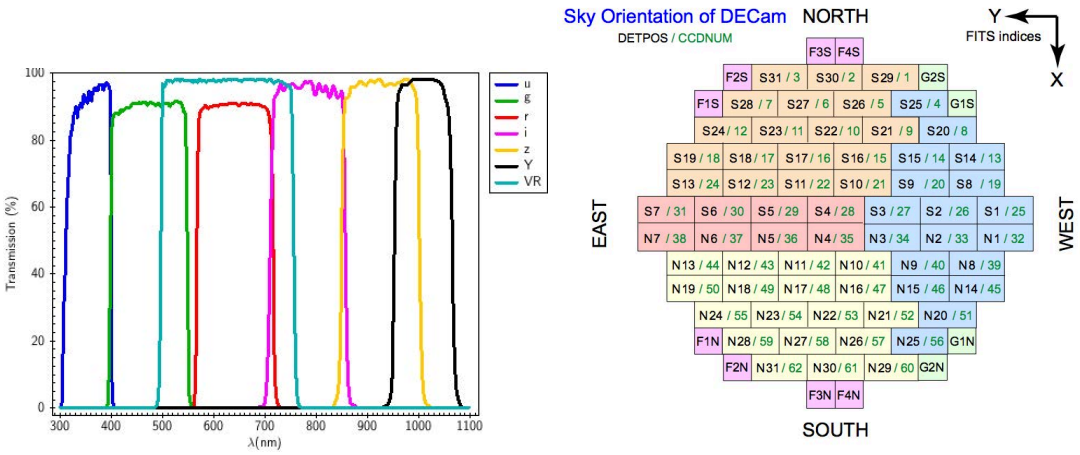

Figure 3.8 – Left: the DECam filters. Right: sketch of the DECam CCD arrangement. The 2k × 2k CCDs dedicated to guiding and focus are labelled G (green) and F (pink), respectively. The CCD colours indicate their respective readout electronics. Taken from Flaugher et al. (2015).

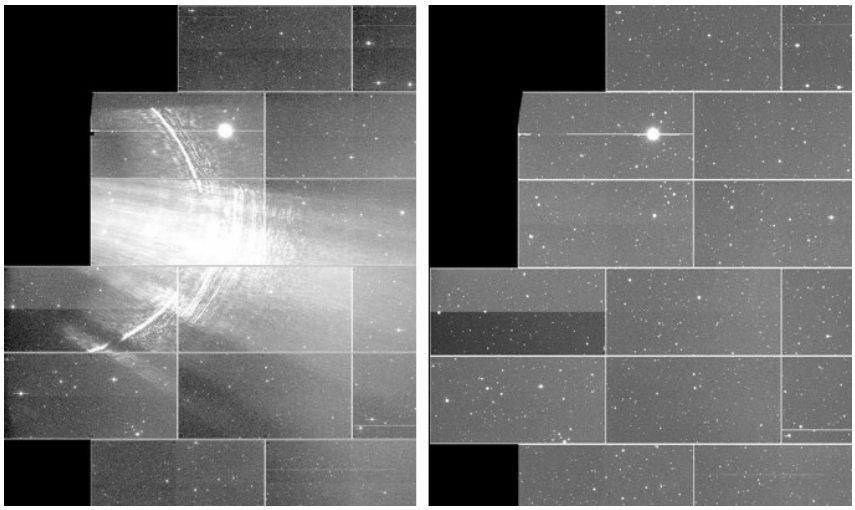

Figure 3.9 – Image taken ≃ 1.5° away from the bright star Canopus, before (left) and after (right) application of anti-reflective paint. Taken from Flaugher et al. (2015).

The CCDs are read out and cleared in 20 s with 6−9 electrons readout noise. They are cooled down to −100 °C by a liquid nitrogen dewar. A two blade shutter ensures an exposure time uniformity of 10 ms: at the start of the exposure, the masking blade is moved in the direction away from the stored blade; at the end of the exposure, the stored blade masks the aperture (in a 1.1 s move). Some stray light was noticed when a very bright star was ≃ 1.5° off-axis, due to reflections on the edges of the filter changer and shutter (see Figure 3.9, left). This was solved by mounting additional baffles and applying an anti-reflective paint in mid-March 2014 (see Figure 3.9, right).

3.2.2.3 Observing strategy

Contrary to the SDSS imaging survey drift-scan strategy, DECam points a patch of the sky using one filter for each pass. In addition to increasing the signal-to-noise, several observations of the same patch of sky allow to remove detector artefacts or particle events. The second and third

passes are respectively shifted by $(\Delta\text{R.A.}, \Delta\text{Dec.}) = (0.2917°, 0.0833°)$ and $(\Delta\text{R.A.}, \Delta\text{Dec.}) = (0.5861°, 0.1333°)$ in order to fill in the gaps left between the CCDs. The pointings were chosen following Hardin, Sloane and Smith [1] approach to cover a sphere with a fixed number of points. This ensures a coverage of 98.88% for one exposure and 98.01% for two exposures.

Observations in the g and r bands are performed when the moon is under the horizon; z band is imaged during the morning and evening twilight and when the moon is in the sky. Observing conditions are defined photometric if two criteria are met: (i) a clear, 90 % transparent sky and (ii) seeing better than 1.3″ FWHM. Pass 1 is performed when both (i) and (ii) are satisfied; pass 2 when only one of the two is fulfilled and pass 3 when none of the above criteria are met. In any case, pointings with airmass (the light path length through the atmosphere compared to the vertical) greater than 2.4 or directed to less than 40 to 50° from the Moon are excluded, and pointings less than 1.2° from bright planets (Mars, Neptune) are avoided. The exposure time is adapted on-the-fly to the observing conditions in order to ensure a roughly uniform imaging depth, while avoiding overheads due to reobserving too shallow exposures. DECaLS and MzLS are the first surveys to use dynamic observing time (Burleigh et al., 2020); DESI will also apply this technique to spectroscopic observations.

The central 1000×1000 pixels of a DECam CCD (N4) are analysed on-the-fly and a `Python` software measures the seeing, sky brightness, atmosphere transparency and telescope pointing error. The raw image is first corrected by the bias and gain of the amplifier. The sky brightness is estimated by sigma clipping (removing outliers, due to e.g. bright sources). The root mean square of sky counts in a pixel is calculated as $\sigma_{\rm sky}$. The image is then convolved with a Gaussian of FWHM 5 pixels, and pixels with a signal $\geq 20\sigma_{\rm sky}$ are selected as potential source candidates. The source magnitude is determined by comparing the total number of electrons collected in a disk of aperture diameter 7″ to the mode of the number of electrons in pixels located within an annulus of 14−20″ times the aperture area. Sources too close to CCD edges, bad pixels and other sources are discarded. The seeing is obtained as the median of the FWHM of the circular 2D Gaussian fitted to sources with $20 < S/N < 100$. The photometric zero point is estimated on-the-fly by comparing the aperture magnitudes previously estimated to a catalogue of the Panoramic Survey Telescope and Rapid Response System 1 (Pan-STARRS or PS1) Data Release 1 (Chambers et al., 2016) matched to Gaia DR1, which is used to measure astrometric offsets. The relative atmospheric transparency is then calculated from the photometric zero point and the airmass. The 5σ AB magnitude depth is computed based on the photometric zero point, galactic extinction and $\sigma_{\rm sky}$. These estimates are used to update the exposure time for the next observation (clipped between $[56, 200]$ s, $[40, 175]$ s and $[80, 250]$ s for g, r and z respectively), and suggest the new pass number to the observer. This image processing takes 10 s. However, given the time to read and write images on disk, the exposure time can only be updated after two pointings. Estimated exposure time based on the conditions $\simeq 3$ min prior to the observations is found to be a general improvement over a fixed exposure time, especially in poor observing conditions (passes 2 and 3) (Burleigh et al., 2020). The telescope pointing offset must be corrected manually by the observers. The median seeing is 1.3, 1.2 and 1.1″ in the g, r and z bands respectively.

1. http://neilsloane.com/icosahedral.codes/

3.2.2.4 Data reduction

These imaging data are processed through the NOAO (National Optical Astronomy Observatory) Community Pipelines (CPs) — one for each instrument. CPs perform instrumental calibration: CCD corrections, such as bias and flat fielding subtraction, identification and removal of artifacts such as pixel cross-talks or cosmic rays, astrometric and photometric calibration. The CP provides a function mapping each pixel to celestial coordinates, known as a world coordinate system (WCS). Astrometric calibration is performed by least square fitting a continuous, smooth function (TPV: tangent plane projection with polynomial distortions) to the centroid of known, detected sources. The 2CMASS catalogue (Skrutskie et al., 2006) was utilised as a reference for early DECam data (from 2013 to 2014), before using Gaia DR1 (Gaia Collaboration et al., 2016). This calibration takes into account the mean atmospheric distortions in each filter as it is performed on individual exposures. The TPV however does not account for small scale pixel distortions, nor star proper motions. The CP removes spatial variations of the CCD (i.e. flattens the CCD), provides the conversion from analog digital units to photoelectrons, data quality masks and weight maps. DECaLS (as well as other Legacy Surveys) is meant to be intra-calibrated by comparing different photometric exposures of the same stars to infer the system throughput and transparency of the atmosphere. This calibration is further propagated to non-photometric observations. However, the photometric calibration was tied to PS1 DR1 up to Legacy Surveys DR6, due to their limited coverage. The CP processing is imperfect (bad background subtraction around large, bright objects, moving objects not removed). The CP outputs (calibrated individual images, data quality masks and weight maps) and WISE data are transferred to the National Energy Research Scientific Computing Center (NERSC), where further processing is done: improved astrometric and photometric calibrations, source extraction and catalogue creation.

3.2.2.5 The `legacypipe`

The source catalogues are produced by a pipeline called `legacypipe`, wrapping `The Tractor`, a forward modelling approach to extract and fit sources in pixel-level data. The footprint is divided into $0.25° \times 0.25°$ bricks. The first step of `The Tractor` is source detection. For this, the sky background is computed in a two-step process: the median pixel signal is measured in sliding windows over each CCD, fitted with a dimensional spline; then pixels which are more than 3σ away from the spline prediction are masked and the background is again fitted by a 2D spline. The PSF is estimated on the background-subtracted CCD images (using PSFex, Bertin, 2011), which are then convolved with their own PSF to facilitate source detection. Five stacks are created from the obtained images, one in each of the g, r and z bands, one "flat" weighted combination of all three bands such that the AB colour mag is 0, and another "red" combination such that $g - r = 1$ and $r - z = 1$. Sources are detected with a $6\sigma_{\text{sky}}$ threshold (σ_{sky} the sky noise level) in each of the five stacks.

The second step of `The Tractor` is source fitting. For each detected source, a model is fitted simultaneously to all the CP pixel data containing the source. Four models are considered: a delta function (for point source), a de Vaucouleurs profile $\propto r^{-1/4}$, an exponential or a composite — de Vaucouleurs and exponential — profile. The same model (convolved with the PSF of each exposure) is fitted in all three g, r and z bands, resulting in measurements of the source position, shape parameters and photometry (magnitude). The source model of WISE pixel data is forced to be the same as that for the grz bands, but convolved with the WISE PSF. This allows to

Figure 3.10 – From left to right: a `legacypipe` $0.25° \times 0.25°$ grz brick image (2212p085), model from `The Tractor`, and residuals. Taken from Dey et al. (2019).

use WISE data for sources which are otherwise undetected or blended in the WISE catalogues. An example of `The Tractor` source fitting is shown for a brick in Figure 3.10. Astrometric calibration is performed with PS1 DR1 stars, their position being taken from Gaia DR1 after the Legacy Surveys DR3 and from Gaia DR2 (Gaia Collaboration et al., 2018) after the Legacy Surveys DR7. This results in a scatter in astrometric calibration of $\simeq 20$ milliarcsecond.

DECaLS source catalogues produced by the `legacypipe` contain different types of objects; stars, galaxies, quasars, etc. The ELG target selection described below aims at selecting an homogeneous subsample of these targets, in the desired redshift range, for spectroscopic follow-up.

3.2.3 eBOSS ELG target selection

eBOSS ELG targets were selected in the imaging catalogues with the following requirements (Dawson et al., 2016; Raichoor et al., 2017): a minimum surface density of $170\,\mathrm{deg}^{-2}$, a relative variation of target density below 15% with respect to imaging depth, galactic extinction, stellar density and uncertainties in the imaging zeropoint. Preliminary ELG target selections relied on the well-understood, homogeneous SDSS imaging (Raichoor et al., 2016; Delubac et al., 2017). However, with the first DECaLS releases (DR3 and DR5) which were already deeper than SDSS, it became possible to push the target selection to higher redshift. This is the reason why the eBOSS collaboration moved to DECaLS imaging (Raichoor et al., 2017).

The ELG footprint is divided into its NGC and SGC components (see Figure 3.11). The ELG NGC footprint covers $\simeq 554\,\mathrm{deg}^2$, between $126° < \mathrm{RA} < 169°$ and $14° < \mathrm{Dec} < 29°$. The SGC covers $\simeq 616\,\mathrm{deg}^2$ with two adjacent regions; one between $317° < \mathrm{RA} < 360°$ and $-2° < \mathrm{Dec} < 2°$ and another between $0° < \mathrm{RA} < 45°$ and $-5° < \mathrm{Dec} < 5°$. The SGC raw images come from the DES collaboration. As can been noted in Figure 3.11, the SGC benefits from a much higher number of observations. This is confirmed by Figure 3.12, showing a higher depth (by $\simeq 0.5\,\mathrm{mag}$) in the SGC. This figure also shows that DECaLS photometry, even in the early data releases is 1 to 2 magnitudes deeper than the SDSS one. Note however that SDSS depth is less scattered than that in DECaLS, as a result from the SDSS drift-scan strategy, and DECaLS observations being not completed. The photometric properties of the eBOSS ELG target sample are described in Table 3.1. As the DECaLS/DR3 release was not public before the eBOSS target catalogues were required for the spectroscopic follow-up, raw images were processed through the DECaLS/DR3 pipeline by the eBOSS ELG team. The target selection was performed later on a part of the NGC footprint (chunk `eboss25`) for which a modified version of the DECaLS/DR5 pipeline (still using PS1-based astrometric calibration) was used.

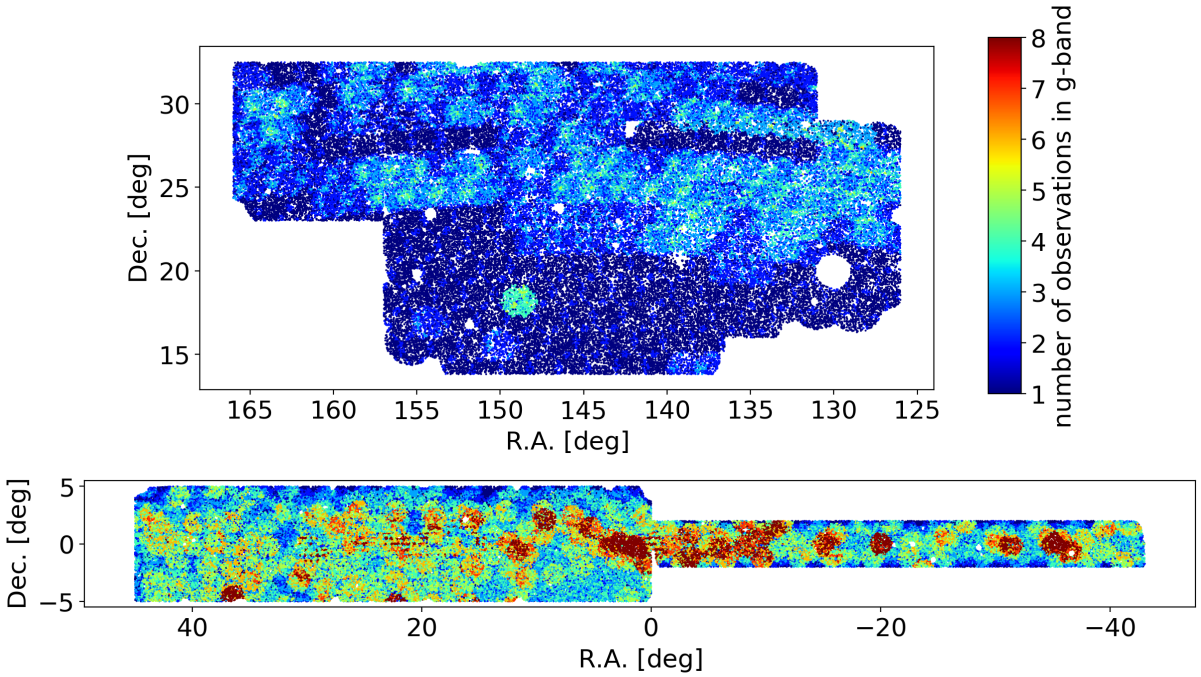

Figure 3.11 – eBOSS ELG targets, as characterised by their number of observations in g-band (top: NGC, bottom: SGC). The number of observations is higher than 8 (maximum displayed here) for 4% of ELG targets.

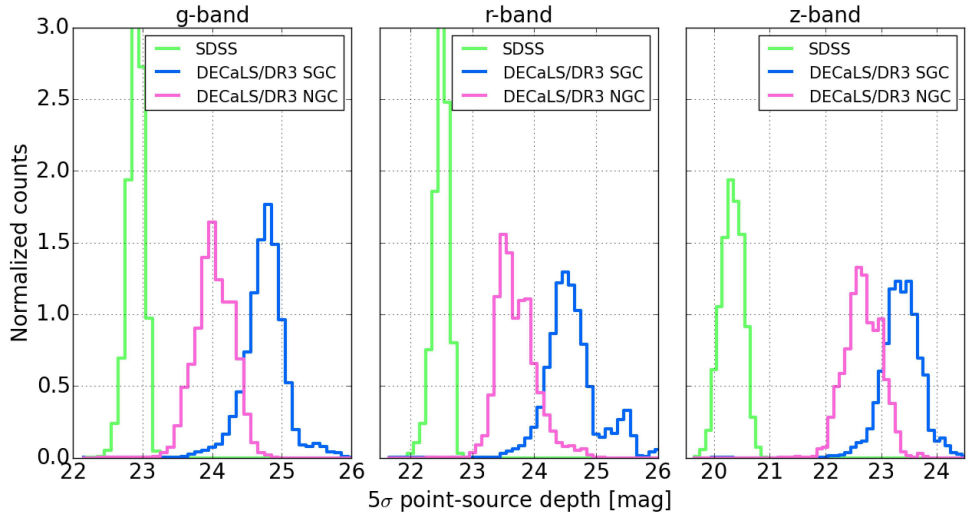

Figure 3.12 – Depth of SDSS and DECaLS imaging surveys, in the NGC and SGC. Taken from Raichoor et al. (2017).

The eBOSS ELG targets were selected according to the following rationale: [OII] emitters are selected through a cut in g magnitude and galaxies are selected in the desired redshift range ($0.6 < z < 1.1$) through a box cut in the $g - r$, $r - z$ colour - colour diagram. It was indeed shown by Comparat et al. (2015) that the g-band magnitude correlates best with [OII] flux in the redshift range $0.7 < z < 1.1$ (see Figure 3.13, left), because in this redshift range the g band corresponds to the rest-frame UV, and thus traces the amount of ionising photons that can eventually lead to the nebulous emission lines (see Section 3.3.2).

region	area [deg^2]	observations	filter	number of exposures	seeing [arcsecond]	depth [mag]
NGC	554	DECaLS	g	2	1.6	23.8
			r	2	1.5	22.5
			z	2	1.3	22.5
SGC	616	DES	g	4	1.6	24.5
			r	3	1.2	22.9
			z	3	1.1	22.9

Table 3.1 – Median photometric properties of the eBOSS ELG target sample.

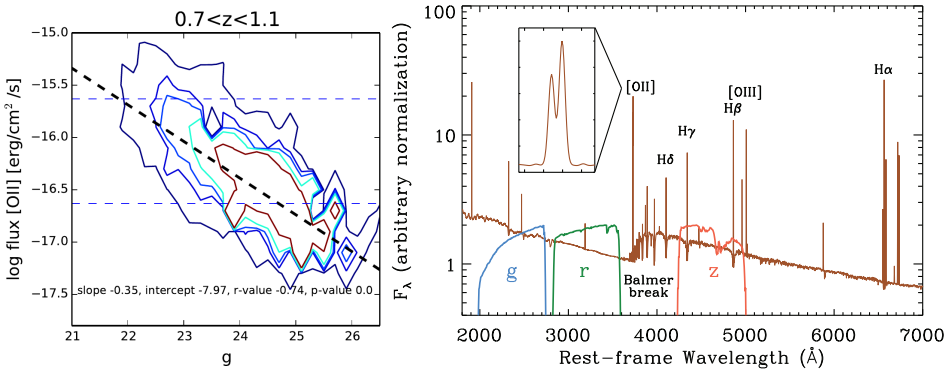

Figure 3.13 – Left: correlation between g-band magnitude and [OII] flux, obtained from various data sets, including GAMA (Driver et al., 2011), VVDS (Le Fèvre et al., 2013), zCOSMOS (Scoville et al., 2007) and the SEQUELS (Alam et al., 2015) programs. Taken from Comparat et al. (2015). Right: ELG rest-frame spectrum, and the position of DECam filters for a galaxy at redshift $z = 1$. Taken from DESI Collaboration et al. (2016).

The colour cuts can be understood looking at Figure 3.13 (right), showing a rest-frame ELG spectrum, and the position of DECam filters for a galaxy at redshift $z = 1$. Going to lower redshift, the filters shift to the right and $r - z$ increases (bluer); going to higher redshifts, $g - r$ and $r - z$ increase.

Figure 3.14 shows the colour selections on a test sample to assess the final target selection. Plotted objects come from the CFHTLS/W4 (Gwyn, 2012) region ($\simeq 20\,\mathrm{deg}^2$ at RA $= 333°$ and Dec $= 2°$) covered by the DES deep fields, with photometric redshifts of Coupon et al. (2009). The grz magnitudes are taken from DECaLS/DR3. In order to test the effect of the target selection on the eBOSS footprint, the latter magnitudes are slightly scattered to account for photometric scatter due to the lower depth in the eBOSS ELG NGC/SGC regions (Raichoor et al., 2017). Note that $0.7 < z < 1.1$ objects are more scattered in the NGC (left panel) than in the SGC (right panel), due to the shallower NGC photometry. Photometric scatter lets low redshift objects enter in the blue end of the selection box. Since the density of such objects is higher than redder ones, we expect the density of low redshifts to be higher in regions of shallower imaging after target selection. For this reason, the NGC selection box was reduced to redder colours, in order to avoid too many low redshift objects to scatter in the target selection box. We will see in Section 5.3.1.6 that this effect is however present in the data redshift distribution as a function of the imaging depth. The colour cuts are summarised in Table 3.2 and provide a list of 269 718 targets.

In addition to the above cuts, several masks were applied to ensure a clean photometry (Raichoor et al., 2020). They are displayed in Figure 3.15 and the area and number of ELG targets they

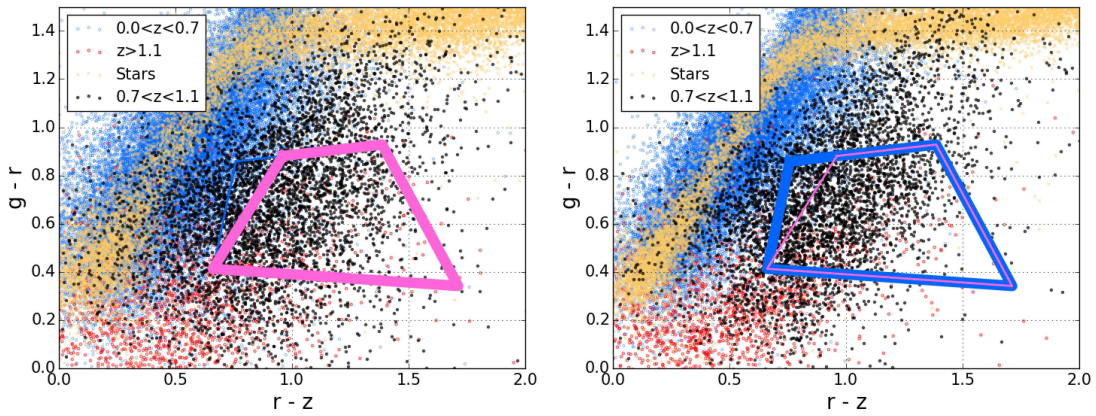

Figure 3.14 – DECaLS/DR3 objects matched to CFHTLS/W4, with the NGC (pink) and SGC (blue) target selection box. Redshifts are photometric, from CFHTLS (Coupon et al., 2009). See text for details. Taken from Raichoor et al. (2017).

region	[OII] emitters	redshift range
NGC	$21.825 < g < 22.9$	$-0.68 \times (r-z) + 0.457 < g - r < 0.112 \times (r-z) + 0.773$ $0.637 \times (g-r) + 0.399 < r - z < -0.555 \times (g-r) + 1.901$
SGC	$21.825 < g < 22.825$	$-0.68 \times (r-z) + 0.457 < g - r < 0.112 \times (r-z) + 0.773$ $0.218 \times (g-r) + 0.571 < r - z < -0.555 \times (g-r) + 1.901$

Table 3.2 – eBOSS ELG target selection: magnitude and colour cuts.

remove are presented in Table 3.3. Bits 1 to 5 described below were applied to the initial target catalogues; the few targets removed by these bits which are mentioned in Table 3.3 are due to a slight change in the implementation. Other masks were applied a posteriori. All masks are defined at the brick level ($0.25° \times 0.25°$ with 3600×3600 pixels), except for bits 8 and 11.

not $g+r+z$ (bit 1) excludes regions without imaging in either the g, r and z bands. This mask is obtained from the `legacypipe` depth images.

xy bug (bit 2) a coding error was present in the target selection scripts, applying the not grz mask in reverted coordinates. To reproduce this effect, an additional mask, that is recovered from the `legacypipe` depth images is introduced.

`decam_anymask` (bit 3) this mask rejects sources if their centre lies in a defective pixel in any of DECam images, and is often activated near CCD edges. DECaLS DR3 `legacypipe` version stored the mask value for each detected object. However, this mask must be known at any location of the eBOSS footprint (as part of the survey selection function, see Section 4.2.2). This has been achieved by running the DECaLS DR7 pipeline (which stores this information at the brick level) on the eBOSS ELG images.

imprecise bit 3 (bit 8) due to differences in pipeline versions, the recovered `decam_anymask` (bit 3) does not fully match the one of DECaLS/DR3; therefore, the footprint was divided into small Healpix (Górski et al., 2005) pixels of 11 arcmin^2 (nside = 1024), and those (37) for which more that 10% of DECaLS/DR3 objects have a `decam_anymask` flag different than recovered from DECaLS/DR7 were flagged.

`tycho2inblob` (bit 4) removes sources whose profile overlaps Tycho-2 stars (Høg et al., 2000), as stored by the `legacypipe` pipeline.

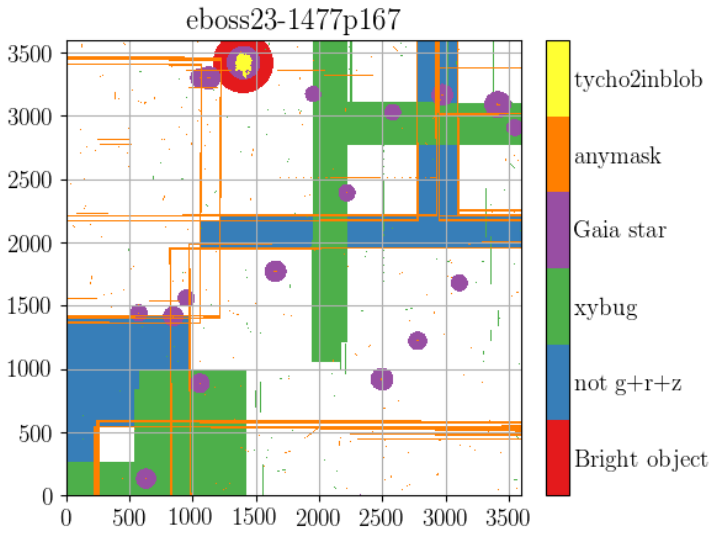

Figure 3.15 – Photometric masks on a `legacypipe` $0.25° \times 0.25°$ brick. Taken from Raichoor et al. (2020).

bright objects (bit 5) removes SDSS bright objects [2] and sources within a larger radius around Tycho-2 stars ($10^{3.5-0.15V}$ arcsecond, with V the Tycho-2 star `MAG_VT` quantity)

Gaia stars (bit 6) removes sources within a radius $10^{2.32-0.07G}$ arcsecond around a Gaia DR2 (Gaia Collaboration et al., 2018) star of magnitude $0 < G < 16$, which complete the Tycho-2 star sample used in `tycho2inblob`.

Mira star (bit 7) removes sources close to the Mira star located at (R.A., Dec.) = $(34.84°, -2.98°)$. Since this star is variable, its magnitude in the Tycho-2 catalogue is not representative of that at the time of DECam observations. A circular mask of radius 2° is applied around the star location.

bad photometric calibration (bit 11) DECaLS/DR3 data included images from numerous surveys. DECaLS DR8 restricts to images from DES and DECaLS, and shows an improved photometric calibration. Regions where the source magnitude estimation in the DECaLS DR3 and DR5 catalogues used for the eBOSS ELG target selection differ by more than $\simeq 10$ mmag from DECaLS/DR8 were masked, using the `mangle` software (Hamilton and Tegmark, 2004).

As we will see in Chapter 5, variations in the imaging quality imprints fluctuations in the density of targets, leading to *photometric systematics*, which we will have to mitigate.

3.3 Spectroscopic operations

Spectroscopic observations are carried out onto the catalogue of photometric targets described in the previous section. As explained in Section 3.1, photometric targets must be assigned spectroscopic fibres to measure their spectrum. This fibre assignment will be discussed in Section 3.3.1. Then, each spectrum will be used to produce a redshift measurement, following a method presented in Section 3.3.3, after a brief description of eBOSS ELG spectra in Section 3.3.2.

2. https://data.sdss.org/sas/dr10/boss/lss/reject_mask/

bit	mask	removed area [deg^2]	removed targets
1	not *grz*	67.2	27
2	*xy* bug	49.7	0
3	recovered `decam_anymask`	210.1	142
4	`tycho2inblob`	4.7	0
5	bright objects	57.6	7
6	Gaia stars	54.0	17456
7	Mira star	12.5	3555
8	imprecise bit 3	0.1	15
11	bad photometric calibration	72.7	16325

Table 3.3 – Angular photometric veto masks applied to the eBOSS ELG catalogues. Masks 1 to 5 were applied prior to the fibre assignment described below, while all other masks were applied after spectroscopic observations. Extracted from Raichoor et al. (2020).

3.3.1 Fibre assignment

Fibre assignment consists in assigning spectroscopic fibres to targets (Blanton et al., 2003), so as to maximise the fraction of targets that receive a fibre, the so-called *tiling success rate*. This implies to optimally assign targets to a tile (corresponding to a plug plate of Section 3.1.1) and place the tiles on the sky. The footprint is divided into a set of rectangles in celestial coordinates (the so-called tiling chunks) where fibre assignment is performed independently. Fibres of a same tile cannot be placed closer than 62″, due to the physical size of the housing of the optical fibre. Note that tile overlaps allow the intersecting region to be observed several times, such that two targets closer than 62″ may be assigned a fibre in these locations.

The first problem consists in optimally allocating targets to each tile (placed at a given location). This is performed in two steps: fibres are first assigned to the *decollided set* of targets, and then to *collided* targets.

3.3.1.1 Decollided set of targets

First, the set of decollided targets, i.e. the largest set containing targets more than 62″ apart is found. For this, a Friends-of-Friends algorithm is run with a linking length of 62″ to obtain the collision groups, i.e. groups of (at least one) targets within the linking length. Then, the multiplicity (i.e. the number of targets) of collision groups being usually small enough, all possible fibre-to-target assignments are explored to find the solution maximising the number of decollided targets. If two solutions are equivalent (e.g. a collision group of 2 targets), one is chosen randomly. An example of this procedure is shown in Figure 3.16 (top left): decollided targets are displayed with continuous circles, collided ones with dashed circles. Note however that targets may have different priorities. In this case, the algorithm maximises the number of fibres allocated to higher priority targets. The obtain decollided set of targets is then assigned to tiles using a network flow formalism (see Figure 3.16, bottom left). Fibres flow from the sources to the sink; they can either go through the overflow arc, at an arbitrarily high cost (in the figure 1000), or be allocated to the 11 decollided targets (filled circles), and get assigned to one of the two tiles, each with a capacity of 592 fibres (the number of science fibres per plate in SDSS-I and-II), with zero cost. This kind of problem is very generic; its best solution (i.e. the one that minimises the total cost) can be found in polynomial time (Goldberg, 1997). This

whole procedure maximises the number of targets which are assigned a fibre in regions where tiles do not overlap.

3.3.1.2 Collided set of targets

The second step consists in allocating remaining fibres to collided targets in tile overlaps. Groups covered by one tile only are not considered since the best fibre assignment has already been achieved for them in the first step. Hence, for this second step only targets within groups which contain at least one target in a tile overlap are considered. These are allowed to be allocated to another tile, with the constraint that all targets which were assigned a fibre in the first step still receive one at the end of the second step. This step can also be recast as a network flow problem for collision groups with multiplicity ≤ 3 (see Figure 3.16, upper right). The minimum capacity (in parentheses) for arcs connecting sources to collision groups (in filled circles) is set to the number of targets which were already allocated fibres in the first step. The maximum capacity is set to the maximum number of targets which can be assigned a fibre (best scenario), given the number of overlapping tiles (regardless of the number of free fibres in the tiles). Arcs flowing from the collision groups to the tiles are given as maximum capacity the maximum number of targets of the group which would go to the given tile in the best scenario. The minimum capacity is set to the minimum number of targets that can be assigned to the tile (here, 1 in all cases). Finally, arcs flowing from the tiles to the sink are given the number of fibres still available after targets within groups covered by one tile only have been allocated fibres (in the first step). This network flow solution is optimal in nearly all cases; very rare complications arise when tile boundaries cross collision groups. Targets of collision groups with multiplicity > 3 are assigned to tiles according to the best of all possible target-to-tile assignments of their specific collision group. Again, target priorities are accounted for in this second step.

The two-step process described above (assignment of decollided and collided targets to tiles) is nearly optimal given fixed tile positions. However, with uniform plate coverage, in regions of high target density the fraction of targets which are assigned a fibre will decrease significantly. For example, the left panel of Figure 3.17 shows a simulated angular distribution of targets, which are assigned fibres following the steps mentioned above. In the middle panel, squares symbolise decollided targets which could not be assigned a fibre; these are concentrated in high density regions, where the fraction of decollided targets which are allocated a fibre can be as low as 10%. Therefore, tile centres are moved to assign a maximum number of fibres to the decollided set of targets, following the heuristic described hereafter.

3.3.1.3 Tiling

The tile initial positions are determined either based on the sphere coverage by Hardin, Sloane and Smith [3] (for large chunks) or by simply aligning tiles in rows (for small chunks). Then, the decollided set of targets is assigned to tiles (step 1 above). To introduce a preference in the tile positions, fibres are allowed to be allocated to tiles up to $2.5R_{\rm tile}$ away from their initial tile center, with a cost:

$$c = \begin{cases} 0 & \text{if } r < R_{\rm tile} \\ c_{\rm max}\frac{(r/R_{\rm tile})^{\alpha}-1}{2.5^{\alpha}-1} & \text{if } r \geq R_{\rm tile} \end{cases}. \tag{3.4}$$

3. http://neilsloane.com/icosahedral.codes/

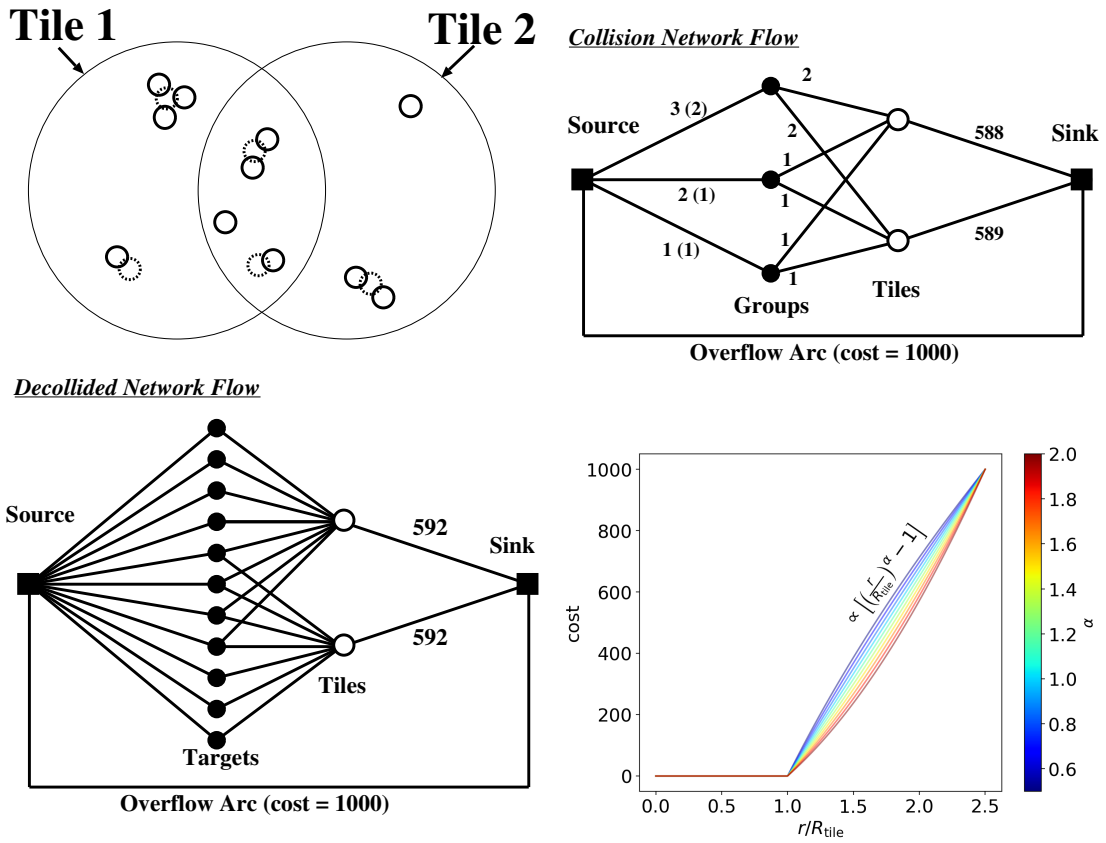

Figure 3.16 – Upper left: an example of set of targets (decollided: continuous circles, collided: dashed circles), to be assigned to two tiles. Bottom left: the assignment of the decollided set of targets to tiles is described as a network flow problem. Top right: collided and decollided targets of collision groups (of multiplicity ≤ 3) in tile overlaps are allocated to tiles using the network flow formalism (unmarked arcs have a capacity of 1). Taken from Blanton et al. (2003). Bottom right: the cost function of the target-to-tile assignment used when tile positions are perturbed.

$c_{\max}$ is the maximum cost, corresponding to not assigning the fibre (1000 in Figure 3.16), and α is a tuning parameter, typically ranging between 0.5 and 2. Then, tiles are moved to minimise the total cost, and the procedure is iterated until a (local) minimum is found. The α parameter is used to tune the tile motions. Higher α values decrease the cost when $r/R_{\text{tile}} \leq 2.5$ (see Figure 3.16, bottom right panel) and thus favor more change in tile positions, which is preferred in large chunks. Finally, the number of tiles is found to match SDSS requirements (that $> 99\%$ of decollided targets are assigned fibres) following a binary search, rerunning the tile positioning and the decollided target-to-tile assignments described above for each number of tiles.

Following the procedure described above, the final tile positions are shown for the simulated set of targets in Figure 3.17 (right): the distribution of decollided targets which are assigned a fibre is much more isotropic; only chunk edges are badly covered, but they may be provided fibres by another overlapping chunks. Note however, that inefficiencies may arise when the number of tiles is too low to ensure complete coverage in all parts of the chunk. One would also note that plate positions vary depending on the target density, and thus may introduce large scale correlations between the set of targets that are assigned a fibre and the underlying clustering; these correlations are expected to be small when almost targets are assigned a fibre — which is the case of the eBOSS ELG sample.

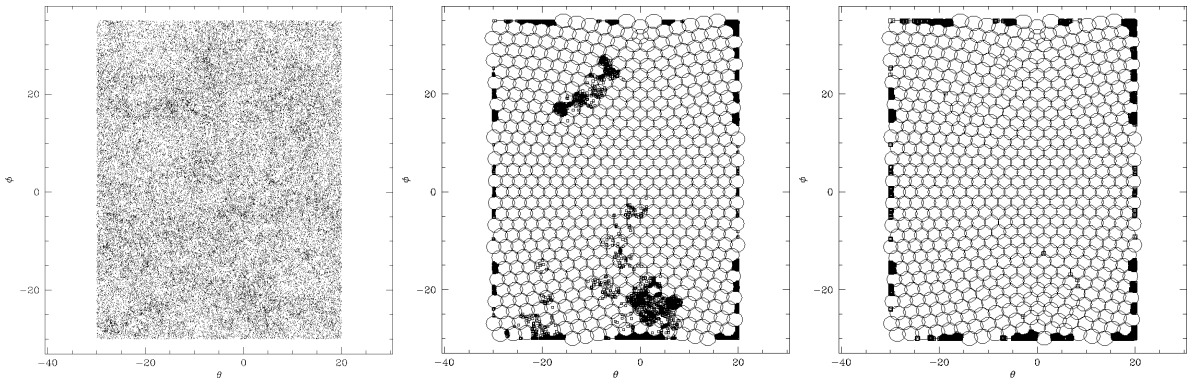

Figure 3.17 – Left: simulated set of targets. Middle: fibre assignment with fixed tile positions; black squares display decollided targets which are not assigned a fibre. Right: same, optimising the tile positions.

bit	mask	removed area [deg^2]	removed targets
9	centerpost	0.6	166
10	TDSS_FES targets	1.3	308

Table 3.4 – Angular tiling veto masks applied to the eBOSS ELG catalogues. Extracted from Raichoor et al. (2020).

3.3.1.4 eBOSS ELG tiling

The eBOSS ELG programme has been allocated 305 plates with $\simeq$ 850 fibres dedicated to the ELG spectra. The eBOSS ELG footprint is split in four tiling chunks: two in the NGC (`eboss23` and `eboss25`) and two in the SGC (`eboss21` and `eboss22`). Some targets of the Time Domain Spectroscopic Survey (TDSS, Ruan et al., 2016) ($\simeq$ 50 per tile) were tiled at the same time as ELGs. A subset of them, TDSS "FES" targets (206, 451, 469, 280 in chunks `eboss21`, `eboss22`, `eboss23` and `eboss25`, respectively) had the same priority (0) as eBOSS ELGs, while other "RQS" targets were targeted with lower priorities (1, 2 and 3). To account for possible knockouts of eBOSS ELG targets by TDSS FES targets, we applied a mask of radius 62″ around each of them (bit 10). Also, fibres of a given plate cannot be placed within 92″ of the plate centerpost (see Section 3.1.1). Since the ELG tile density is high, the centerpost of a given plate can be covered by another one. For simplicity we apply a veto mask (bit 9) to all centerpost, which does not remove any fiber-assigned target. The effect of these two `mangle` masks — TDSS FES targets and centerposts — is given in Table 3.4.

Once targets are assigned fibres, their spectrum can be measured with the spectroscopic setup described in Section 3.1.2. Let us first present eBOSS ELG spectra.

3.3.2 eBOSS ELG spectra

As their name suggests, emission line galaxies show emission lines in their spectra. Most of these lines, called nebular lines, are emitted by the gas distributed throughout the galaxy, which is photo-excited by ultraviolet light produced by hot stars with short lifetimes. Therefore, these emission lines are a probe of the galaxy star formation rate, used for example to constrain the stellar mass function (SMF), and study the stellar-to-halo mass relations (SHMRs) (Guo et al., 2019). Note that emission lines can also be produced by excited gas around a central black hole.

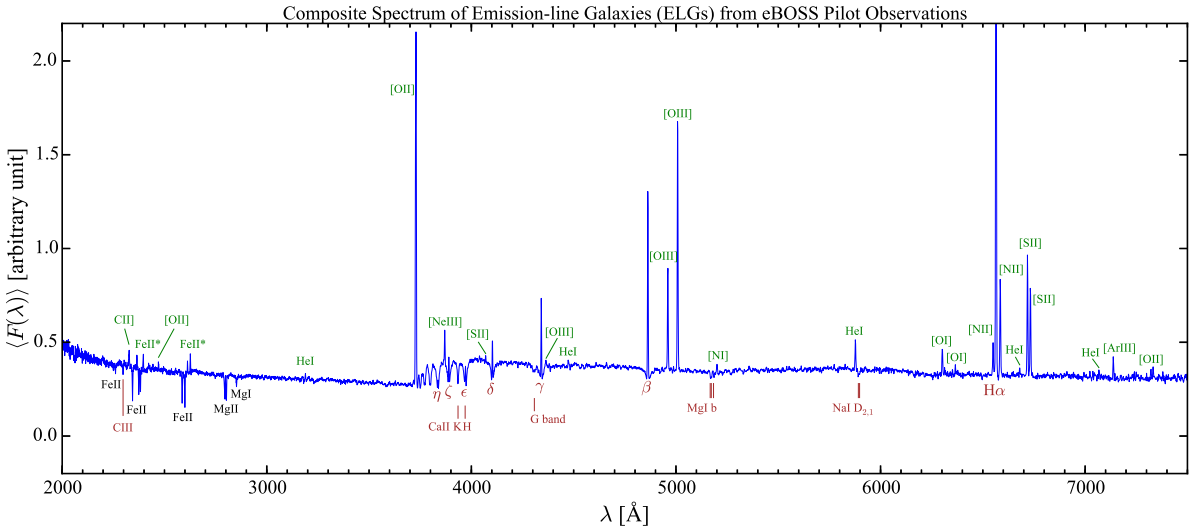

Figure 3.18 – Stacked rest-frame spectrum of the 12 000 ELGs from SEQUELS and eBOSS pilot observations. Emission lines are labelled in green, stellar absorption lines in red and interstellar/circumgalactic medium absorption lines in black. Taken from Zhu et al. (2015).

Quasars, which were targeted by the eBOSS programme, constitute a subclass of these active galaxy nuclei (AGN).

A stacked spectrum of SEQUELS (Sloan Extended Quasar, ELG and LRG Survey, Alam et al., 2015) and eBOSS ELG pilot observations is shown in Figure 3.18. One of the ELG most prominent nebular line is the [OII]($\lambda\lambda 3727, 3730$) doublet: II denotes that it is a transition of singly-ionised oxygen, and the brackets stand for a forbidden transition — a transition that is not allowed by usual quantum selection rules, but still happens at higher order approximation, at a very low rate. $\lambda 3727$ and $\lambda 3730$ are the rest-frame wavelengths in angstrom. Other forbidden lines include metals like the doubly-ionised oxygen [OIII]($\lambda 5008$), neutral oxygen [OI]($\lambda 6300$), singly-ionised nitrogen [NII]($\lambda 6584$).

Forbidden emission lines correspond to deexcitations from meta-stable states, which almost always yield a photon in nebulae. Indeed, contrary to laboratory experiments, the nebular gas is of very low density, such that the probability of collisional deexcitation is tiny. In contrast, permitted nebular emission lines, such as HeI, Hα($\lambda 6563$), are emitted through radiative transitions from a photo-excited state.

The strong continuum at ~ 2000 Å is compatible with emission from hot massive (O/B) stars in the UV. Blueward of 2900 Å are strong absorption lines, similar to those found in quasar absorption line systems, due to gas in the interstellar or circumgalactic medium. One can also notice the Balmer discontinuity at 3646 Å, caused by electrons being ejected from the second energy level of hydrogen atoms. Galaxies with high stellar formation have a decreased Balmer discontinuity due to high ultraviolet light emitted by young hot stars.

Other features, usually most prominent in early-type, red galaxies are still visible in the stacked ELG spectrum in Figure 3.18. First, from the break at 3646 Å to 6563 Å we find the Balmer series (η, ζ, ϵ, δ, γ, β, Hα). Second, the atmosphere of cool red giant stars is responsible for e.g. the MgI b($\lambda 5175$), NaI D($\lambda 5892$), CaII K($\lambda 3933$) and CaII H($\lambda 3968$) absorption lines — though less pronounced than in red galaxies.

Star-forming ELGs are an interesting tracer for clustering analyses. Indeed, their distribution follows the star formation rate, which increases up to $z \simeq 2$, where red galaxies are rarer. Also,

the strong emission lines ease the redshift measurement — thus allowing reduced observing time. In particular, the [OII] doublet can be resolved with a spectrograph resolving power of 4000 (as is the case of DESI), thus making the redshift measurement extremely robust. Both the high density and fast redshift measurement make ELGs an ideal tracer for clustering analysis at redshift $1 \lesssim z \lesssim 2$.

Let us now review how eBOSS ELG redshifts are measured.

3.3.3 Redshift determination pipeline

The eBOSS ELG spectroscopic observations were conducted between September 2016 and February 2018. Each plate was observed with 15 min exposures until the median squared signal-to-noise ratio in the red channel rSN^2 reached 22; this threshold was reached with 4.7 exposures in average. First observations, covering half of the chunk `eboss21` footprint, used a threshold of 40. When the signal-to-noise ratio could not be reached before a plate was unplugged, the plate was reobserved later, with different (physical) fibre-to-target assignments. These repeat observations are used to tune the quality criteria of the redshift pipeline and assess the reliability of the redshift measurement (see Section 3.3.3.3). Other repeat observations come from the allocation of remaining, free fibres to eBOSS ELG targets in tile overlaps.

3.3.3.1 Spectroscopic data reduction

The first step of the redshift measurement, performed by the `IDLSPEC2D` software, is the reduction of raw 2D (spatial, spectral) images obtained at different exposures into single 1D spectra. As discussed in Section 3.1.2, the fibres entering the slithead are grouped by bundles of 20 fibres. The gap between fibres is 260 µm, while the separation between bundles is 624 µm, thus avoiding contamination between bundles, which are therefore treated separately. We recall that spectra are aligned (though slightly curved) along the CCD columns, with a typical width (along the CCD rows) of 3 pixels. Pixel signals are first debiased, converted to electron counts, and flat fielded. Scattered light is estimated in the space between fibre bundles and interpolated over the spectral traces. Let us call D_i the post-processed pixel values. Then, the following cost function is minimised for each row of pixels (i.e. for each wavelength) to extract the number of collected electrons $\hat{N}_f$ from each fibre f of the bundle:

$$\chi^2_{\text{row}} = \sum_{\text{pixel } i} w_i \left(D_i - \sum_{\text{fibre} f} \hat{N}_f \hat{P}_{f,i} \right)^2 . \tag{3.5}$$

The sum $\sum_{\text{pixel } i}$ runs over the spatial extension (along the fibre spread) of a fibre bundle. $\hat{P}_{f,i}$ is a Gaussian modelling the fibre dispersion in the spatial direction (typically one pixel wide) and w_i are weights. To minimise the variance of $\hat{N}_f$, w_i should be inverse-variance weights, $w_i = var(D_i)^{-1}$. In the BOSS DR12 pipeline, $var(D_i)$ was directly taken to be $var(\text{readout}) + \max(D_i, 0)$, with $var(\text{readout})$ the readout noise and D_i assumed to be Poisson. However, estimating the inverse-variance weights from the data itself leads to biased estimates of the signal, with a signal-dependent bias. Therefore, since DR13, w_i is taken to be $var(\text{readout})^{-1}$ (which varies because of the flat field correction), at the cost of a slightly increased (up to 10% in the bright end) variance of $\hat{N}_f$ (Bautista et al., 2017). To summarise, the above procedure collapses the 2D images in 1D information as it yields, for each fibre, a single electron count $\hat{N}_f$ for each pixel wavelength.

Subsequent steps involve correcting for the fibre flat field, subtracting the sky background model and performing the photometric calibration. These operations require averaging over several spectra, with inverse variance weights determined by fitting a spline to the spectra to avoid biases due to flux - pixel variance weight correlations. The fibre flat field (fiberflat) is performed by comparing the flux of different fibres equally illuminated by a flat lamp spectrum. The sky model (skymodel) is obtained from the flat fielded sky fibres. The flux is calibrated (calib) with respect to standard star spectra, whose synthetic templates used in eBOSS DR16 were developed for DESI (Ahumada et al., 2020). In addition, flux correction vectors (fluxcorr) are estimated for each fibre through a low order polynomial fit to correct for miscalibrations between different exposures, and flux distortion vectors (fluxdistort) model the throughput variations depending on the position in the focal plane. After flat fielding and sky subtraction, the flux is:

$$F_e = \hat{N}/\text{fiberflat} - \text{skymodel}. \tag{3.6}$$

where we drop the fibre dependence for simplicity, and the calibrated flux is:

$$F = F_e/\text{calib} \cdot \text{fluxcorr} \cdot \text{fluxdistort}. \tag{3.7}$$

3.3.3.2 Redshift measurement

The redshift measurement and classification pipeline used in BOSS, `IDLSPEC1D` (Bolton et al., 2012), found the best redshift estimate based on a χ^2 minimisation approach. The measured 1D spectrum (output by `IDLSPEC2D`) is compared to a linear combination of a set of spectral templates and a low-order polynomial to marginalise over galactic and intrinsic extinction as well as remaining calibration errors. The spectral templates are constructed for three different classes, `STAR`, `QSO` or `GALAXY`, using a rest-frame principal component analysis (PCA) applied to a training sample based on observed spectra from stars, quasars, and galaxies, respectively. Quasar and galaxy spectra are extracted from SDSS early data while star spectra are constructed from external data sets. Then, each observed spectrum is fitted by the different template classes [4]; the best measurement (redshift, redshift error and target class) is taken to be that with the minimum (reduced) χ^2 among the three classes. The χ^2 difference between the first and the second best fit solutions is used to derive the confidence of the redshift measurement (`ZWARNING` flag).

Note that in practice, some spectra are unusable, due to hardware issues, e.g. unplugged fibre, too small wavelength coverage, spectrograph issues, as flagged by `ZWARNING` [5].

`IDLSPEC1D` provided good redshift measurements for 99% of the BOSS CMASS and LOWZ samples of bright galaxies. eBOSS, however, targeted fainter LRGs and ELGs for which `IDLSPEC1D` performed poorly (finding $\simeq$ 70% of good spectra), while eBOSS requirements aimed at 90% good redshifts and classification (Dawson et al., 2016). In addition, the redshift accuracy should remain below $300\,\mathrm{km\,s^{-1}}$ for $z < 1.5$, with fewer than 1% catastrophic redshifts, i.e. redshifts classified as good by the pipeline but which actually differ from the truth by more than $1000\,\mathrm{km\,s^{-1}}$ for ELGs and LRGs, and $3000\,\mathrm{km\,s^{-1}}$ for QSOs.

Therefore, a new redshift pipeline, `REDROCK`, was used. This pipeline, originally developed for DESI, is based on a combination of the `IDLSPEC1D` PCA approach and the archetype method of Hutchinson et al. (2016).

4. For BOSS LOWZ and CMASS galaxy samples, `QSO` templates were discarded.
5. https://www.sdss.org/dr16/algorithms/bitmasks/#ZWARNING

The first step of `REDROCK` is similar to the `IDLSPEC1D` pipeline: spectra are fitted with a linear combination of PCA spectral templates of stars, galaxies and quasars — without the polynomial marginalisation over extinction and calibration. Here, however, new templates were produced for galaxies and stars at a higher resolution than that of the BOSS spectrograph (Ross et al., 2020). Instead of the BOSS empirical templates, 10 new galaxy PCA eigenspectra templates were obtained from 20 000 physically-motivated theoretical galaxy spectra spanning stellar age, metallicity, and star formation rate. Similarly, new stellar templates, based on 30 000 theoretical spectra of stars of different masses and evolutionary stages were produced. The higher resolution templates were then convolved by the instrument resolution as a function of wavelength and position in the focal plane.

Then, in a second step, a fraction of the templates described above was used to fit the spectra, similar to Hutchinson et al. (2016). Their method consists in fitting observed spectra with single spectrum templates (in possible combination with lower order polynomial nuisance terms), in this case called *archetypes*, instead of a linear combination of several templates, as in the first step. The purpose of this additional step is to remove solutions involving a non-physical combination of PCA templates. In practice, this fitting procedure is performed in the vicinity of at most 3 best fit solutions obtained from the first step in each class, using 110 galaxy, 64 quasar and 40 star archetypes in combination with the first three Legendre polynomials to marginalise over the broadband spectrum. Then, similarly to `IDLSPEC1D`, the measurement (redshift, redshift error and target class) with the lowest χ^2 is retained, and the χ^2 difference (hereafter called $\Delta\chi^2$) to the second best fit is used as a reliability criterion.

3.3.3.3 Performance of the redshift pipeline

Reliability criteria were determined for the redshift pipeline to maximise the fraction of good redshift measurements while keeping the rate of catastrophic redshifts low. A first tuning was performed with repeat observations of LRG spectra. Spectra of unplugged or sky fibres, with insufficient wavelength coverage, or with a best χ^2 at the limit of the redshift fitting range were discarded. The threshold $\Delta\chi^2 \geq 9$ was chosen such that 93% of the 11 556 pairs were classified with good redshifts, 0.5% of these being catastrophic redshifts (Ross et al., 2020). Applying only the $\Delta\chi^2 \geq 9$ criterion, 10% of sky spectra were considered to have good redshifts — while only a very small fraction of them is expected to correspond, by chance, to real astrophysical objects. Adding the constraint that the model coefficient for the best fit archetype is positive (for physicality), the fraction of sky spectra with good redshifts dropped to 4.4%. In practice, only 0.04% of LRG redshift measurements satisfying the $\Delta\chi^2 = 9$ criterion failed to satisfy the positiveness requirement. The quality cuts mentioned so far (spectrum measurement, $\Delta\chi^2$ and positiveness of the archetype coefficient) are encompassed in the cut:

$$\text{ZWARNING} = 0 \tag{3.8}$$

To fulfill eBOSS ELG requirements, the following additional quality cuts were applied to `REDROCK` redshift estimates:

$$\text{SN\_MEDIAN[i]} > 0.5 \quad \text{or} \quad \text{SN\_MEDIAN[z]} > 0.5 \tag{3.9}$$

$$\text{zQ} \geq 1 \quad \text{or} \quad \text{zCont} \geq 2.5. \tag{3.10}$$

The first criterion imposes a minimum median signal-to-noise of the good pixels in the red (i and z-band) part of the spectrum (in units of $\mathrm{erg\,s^{-1}\,cm^{-2}\,\AA^{-1}}$), where the [OII] line of $z \simeq 1$

galaxies is expected to be observed. Note that the [OII] doublet cannot be clearly resolved by the eBOSS spectrographs given their resolving power (see Figure 3.4). The second criterion utilises the a posteriori flags zQ and zCont developed by Comparat et al. (2016). The zQ flag (where Q stands for quality) quantifies the strength of the emission lines; zQ ≥ 1 means there is at least one line at signal-to-noise above 3. The zCont flag quantifies the strength of the continuum; zCont ≥ 2.5 means that one of the u, g, i or z magnitudes is below 19.5. The redshift measurements satisfying the quality cuts (3.8), (3.9), (3.10) are qualified as *reliable*.

While `IDLSPEC1D` (augmented by a zQ, zCont criterion, Eq. (1) of Raichoor et al. 2017) provided a reliable redshift between 0.6 and 1.1 for 74.0% of objects (not classified as stars), `REDROCK`, with the same criteria, increased this fraction to 80.7% (Raichoor et al., 2020), as reported in Table 3.5. The catastrophic failure rate, i.e. the fraction of reliable redshifts differing by more than $1000\,\mathrm{km\,s^{-1}}$, was estimated from the $\simeq 13\,000$ repeat observations passing the quality cuts above. This rate dropped from 0.5% with `IDLSPEC1D` to 0.3% with `REDROCK`, well below the eBOSS requirement.

In addition, visual inspections were carried out on one plate of the eBOSS ELG survey and two plates of the ELG pilot program (Comparat et al., 2016). Table 3.6 reports the results on the ~ 1900 reliable redshift measurements between 0.6 and 1.1, sorted in 4 confidence levels:

0. no information, useless spectrum;
1. information in the spectrum, but the redshift is a guess;
2. features are visible and the redshift is likely to be correct;
3. definitely correct.

The pipeline and visual inspections are in very good agreement for a confidence above 2. Conservatively assuming the pipeline is wrong when the confidence is 0 (1.6% of the sample), the pipeline provides a redshift precision better than $300\,\mathrm{km\,s^{-1}}$ for 98.1% of the reliable redshifts in $0.6 < z < 1.1$, with a fraction of catastrophic redshifts of $\simeq 1.8\%$.

Lastly, eBOSS ELG redshift measurements were compared to external data sets: DEEP2 (Newman et al., 2013) using the DEIMOS spectrograph (of resolving power $R \sim 5900$) on the Keck II 10 m telescope, and WiggleZ (Drinkwater et al., 2010) using the AAOmega spectrograph (of resolving power $R \sim 1300$) on the AAT 3.9 m telescope. Restricting to reliable eBOSS ELG redshifts within $0.6 < z < 1.1$, the 146 matches with DEEP2 were found to have a mean velocity difference with respect to eBOSS of $16 \pm 39\,\mathrm{km\,s^{-1}}$ and only 1 redshift was found to differ by more than $1000\,\mathrm{km\,s^{-1}}$. The 216 matches with WigleZ were found to have a mean velocity difference with respect to eBOSS of $-2 \pm 47\,\mathrm{km\,s^{-1}}$ with 2 redshifts found to differ by more than $1000\,\mathrm{km\,s^{-1}}$ (Raichoor et al., 2020).

Based on the analysis of repeat observations and visual inspection, confirmed by comparison to external data sets, $\simeq 99\%$ of reliable `REDROCK` redshifts have a precision better than $300\,\mathrm{km\,s^{-1}}$, with a $\simeq 1\%$ catastrophic failure rate, which fulfills eBOSS requirements.

The obtained catalogue of photometric targets and redshifts cannot be readily used for clustering analyses. One has to first specify the survey selection function, i.e. the probability to measure a galaxy redshift in any location of the sky. This will be done in Chapter 5, but first, let us discuss the estimation of the observed galaxy power spectrum and understand the impact of the aforementioned selection function on 2-point correlation function measurements.

redshift fitter	criterion	reliable z	reliable $0.6 < z < 1.1$	catastrophic z	catastrophic $0.6 < z < 1.1$
`IDLSPEC1D`	Eq. (1) of Raichoor et al. (2017)	83.1%	74.0%	0.5%	0.5%
`REDROCK`	Eq. (3.8)	93.0%	82.0%	0.7%	0.6%
`REDROCK`	Eq. (3.8) & Eq. (3.9)	91.8%	81.3%	0.6%	0.6%
`REDROCK`	Eq. (3.8) & Eq. (3.9) & Eq. (3.10)	90.6%	80.7%	0.3%	0.3%

Table 3.5 – Redshift statistics for various criteria. The last line criterion is used in eBOSS. Estimates of catastrophic rates are computed from repeat observations. Taken from Raichoor et al. (2020).

conf. flag	fraction	$\lvert\Delta v\rvert < 300\,\mathrm{km\,s^{-1}}$	$\lvert\Delta v\rvert < 1000\,\mathrm{km\,s^{-1}}$
3	71.5%	99.9%	99.9%
2	24.0%	99.3%	99.6%
1	2.9%	94.5%	96.3%
0	1.6%	6.5%	6.5%
all	100%	98.1%	98.2%

Table 3.6 – Assessment of the `REDROCK` redshift measurement from visual inspection of three plates, containing $\sim$ 1900 ELGs with reliable redshifts in $0.6 < z < 1.1$. See text for details. Taken from Raichoor et al. (2020).

Bibliography

(1997). *The HIPPARCOS and TYCHO catalogues. Astrometric and photometric star catalogues derived from the ESA HIPPARCOS Space Astrometry Mission*, Volume 1200 of *ESA Special Publication*.

Ahumada, R., C. A. Prieto, A. Almeida, F. Anders, S. F. Anderson, et al. (2020). The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra. ApJS *249*(1), 3. arXiv:1912.02905.

Alam, S., F. D. Albareti, C. Allende Prieto, F. Anders, S. F. Anderson, et al. (2015). The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III. ApJS *219*(1), 12. arXiv:1501.00963.

Bautista, J. E., N. G. Busca, J. Guy, J. Rich, M. Blomqvist, et al. (2017). Measurement of baryon acoustic oscillation correlations at z = 2.3 with SDSS DR12 Lyα-Forests. A&A *603*, A12. arXiv:1702.00176.

Bertin, E. (2011). *Automated Morphometry with SExtractor and PSFEx*, Volume 442 of *Astronomical Society of the Pacific Conference Series*, pp. 435.

Blanton, M. R., H. Lin, R. H. Lupton, F. M. Maley, N. Young, et al. (2003). An Efficient Targeting Strategy for Multiobject Spectrograph Surveys: the Sloan Digital Sky Survey "Tiling" Algorithm. AJ *125*(4), 2276–2286. arXiv:astro-ph/0105535.

Bolton, A. S., D. J. Schlegel, É. Aubourg, S. Bailey, V. Bhardwaj, et al. (2012). Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey. AJ *144*(5), 144. arXiv:1207.7326.

Burleigh, K. J., M. Landriau, A. Dey, D. Lang, D. J. Schlegel, et al. (2020). Observing Strategy for the Legacy Surveys. *arXiv e-prints*, arXiv:2002.05828. arXiv:2002.05828.

Chambers, K. C., E. A. Magnier, N. Metcalfe, H. A. Flewelling, M. E. Huber, et al. (2016). The Pan-STARRS1 Surveys. *arXiv e-prints*, arXiv:1612.05560. arXiv:1612.05560.

Comparat, J., T. Delubac, S. Jouvel, A. Raichoor, J. P. Kneib, et al. (2016). SDSS-IV eBOSS emission-line galaxy pilot survey. A&A *592*, A121. arXiv:1509.05045.

Comparat, J., J. Richard, J.-P. Kneib, O. Ilbert, V. Gonzalez-Perez, et al. (2015). The 0.1 <z < 1.65 evolution of the bright end of the [O ii] luminosity function. A&A *575*, A40. arXiv:1408.1523.

Coupon, J., O. Ilbert, M. Kilbinger, H. J. McCracken, Y. Mellier, et al. (2009). Photometric redshifts for the CFHTLS T0004 deep and wide fields. A&A *500*(3), 981–998. arXiv:astro-ph/0811.3326.

Dawson, K. S., J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, et al. (2016). The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data. AJ *151*(2), 44. arXiv:1508.04473.

Dawson, K. S., D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, et al. (2013). The Baryon Oscillation Spectroscopic Survey of SDSS-III. AJ *145*(1), 10. arXiv:1208.0022.

Delubac, T., A. Raichoor, J. Comparat, S. Jouvel, J. P. Kneib, et al. (2017). The SDSS-IV eBOSS: emission line galaxy catalogues at z $\approx$ 0.8 and study of systematic errors in the angular clustering. MNRAS *465*(2), 1831–1846. arXiv:1611.06934.

DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, et al. (2016). The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, arXiv:1611.00036. arXiv:1611.00036.

Dey, A., D. J. Schlegel, D. Lang, R. Blum, K. Burleigh, et al. (2019). Overview of the DESI Legacy Imaging Surveys. AJ *157*(5), 168. arXiv:1804.08657.

Drinkwater, M. J., R. J. Jurek, C. Blake, D. Woods, K. A. Pimbblet, et al. (2010). The WiggleZ Dark Energy Survey: survey design and first data release. MNRAS *401*(3), 1429–1452. arXiv:astro-ph/0911.4246.

Driver, S. P., D. T. Hill, L. S. Kelvin, A. S. G. Robotham, J. Liske, et al. (2011). Galaxy and Mass Assembly (GAMA): survey diagnostics and core data release. MNRAS *413*(2), 971–995. arXiv:1009.0614.

Eisenstein, D. J., I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, et al. (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. ApJ *633*(2), 560–574. arXiv:astro-ph/0501171.

Flaugher, B., H. T. Diehl, K. Honscheid, T. M. C. Abbott, O. Alvarez, et al. (2015). The Dark Energy Camera. AJ *150*(5), 150. arXiv:1504.02900.

Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, et al. (2018). Gaia Data Release 2. Summary of the contents and survey properties. A&A *616*, A1. arXiv:1804.09365.

Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, et al. (2016). Gaia Data Release 1. Summary of the astrometric, photometric, and survey properties. A&A *595*, A2. arXiv:1609.04172.

Goldberg, A. V. (1997). An efficient implementation of a scaling minimum-cost flow algorithm. *Journal of Algorithms 22*(1), 1 – 29.

Górski, K. M., E. Hivon, A. J. Banday, B. D. Wand elt, F. K. Hansen, et al. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ *622*(2), 759–771. arXiv:astro-ph/0409513.

Gunn, J. E., W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, et al. (2006). The 2.5 m Telescope of the Sloan Digital Sky Survey. AJ *131*(4), 2332–2359. arXiv:astro-ph/0602326.

Guo, H., X. Yang, A. Raichoor, Z. Zheng, J. Comparat, et al. (2019). Evolution of Star-forming Galaxies from z = 0.7 to 1.2 with eBOSS Emission-line Galaxies. ApJ *871*(2), 147. arXiv:1810.05318.

Gwyn, S. D. J. (2012). The Canada-France-Hawaii Telescope Legacy Survey: Stacked Images and Catalogs. AJ *143*(2), 38. arXiv:1101.1084.

Hamilton, A. J. S. and M. Tegmark (2004). A scheme to deal accurately and efficiently with complex angular masks in galaxy surveys. MNRAS *349*(1), 115–128. arXiv:astro-ph/0306324.

Høg, E., C. Fabricius, V. V. Makarov, S. Urban, T. Corbin, et al. (2000). The Tycho-2 catalogue of the 2.5 million brightest stars. A&A *355*, L27–L30.

Hutchinson, T. A., A. S. Bolton, K. S. Dawson, C. Allende Prieto, S. Bailey, et al. (2016). Redshift Measurement and Spectral Classification for eBOSS Galaxies with the redmonster Software. AJ *152*(6), 205. arXiv:1607.02432.

Le Fèvre, O., P. Cassata, O. Cucciati, B. Garilli, O. Ilbert, et al. (2013). The VIMOS VLT Deep Survey final data release: a spectroscopic sample of 35 016 galaxies and AGN out to z ∼6.7 selected with $17.5 \leq i_{AB} \leq 24.75$. A&A *559*, A14. arXiv:1307.0545.

Newman, J. A., M. C. Cooper, M. Davis, S. M. Faber, A. L. Coil, et al. (2013). The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts. ApJS *208*(1), 5. arXiv:1203.3192.

Raichoor, A., J. Comparat, T. Delubac, J. P. Kneib, C. Yèche, et al. (2017). The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: final emission line galaxy target selection. MNRAS *471*(4), 3955–3973. arXiv:1704.00338.

Raichoor, A., J. Comparat, T. Delubac, J. P. Kneib, C. Yèche, et al. (2016). The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: selecting emission line galaxies using the Fisher discriminant. A&A *585*, A50. arXiv:1505.01797.

Raichoor, A., A. de Mattia, A. J. Ross, C. Zhao, S. Alam, et al. (2020). The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Large-scale Structure Catalogues and Measurement of the isotropic BAO between redshift 0.6 and 1.1 for the Emission Line Galaxy Sample. *arXiv e-prints*, arXiv:2007.09007. arXiv:2007.09007.

Ross, A. J., J. Bautista, R. Tojeiro, S. Alam, S. Bailey, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Large-scale structure catalogues for cosmological analysis. MNRAS *498*(2), 2354–2371. arXiv:2007.09000.

Ruan, J. J., S. F. Anderson, P. J. Green, E. Morganson, M. Eracleous, et al. (2016). The Time-Domain Spectroscopic Survey: Understanding the Optically Variable Sky with SEQUELS in SDSS-III. ApJ *825*(2), 137. arXiv:1602.02752.

Schlafly, E. F. and D. P. Finkbeiner (2011). Measuring Reddening with Sloan Digital Sky Survey Stellar Spectra and Recalibrating SFD. ApJ *737*(2), 103. arXiv:1012.4804.

Schlegel, D. J., D. P. Finkbeiner, and M. Davis (1998). Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. ApJ *500*(2), 525–553. arXiv:astro-ph/9710327.

Scoville, N., H. Aussel, M. Brusa, P. Capak, C. M. Carollo, et al. (2007). The Cosmic Evolution Survey (COSMOS): Overview. ApJS *172*(1), 1–8. arXiv:astro-ph/0612305.

Skrutskie, M. F., R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, et al. (2006). The Two Micron All Sky Survey (2MASS). AJ *131*(2), 1163–1183.

Smee, S. A., J. E. Gunn, A. Uomoto, N. Roe, D. Schlegel, et al. (2013). The Multi-object, Fiber-fed Spectrographs for the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey. AJ *146*(2), 32. arXiv:1208.2233.

The Dark Energy Survey Collaboration (2005). The Dark Energy Survey. *arXiv e-prints*, astro–ph/0510346. arXiv:astro-ph/0510346.

Wright, E. L., P. R. M. Eisenhardt, A. K. Mainzer, M. E. Ressler, R. M. Cutri, et al. (2010). The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. AJ *140*(6), 1868–1881. arXiv:1008.0031.

York, D. G., J. Adelman, J. Anderson, John E., S. F. Anderson, J. Annis, et al. (2000). The Sloan Digital Sky Survey: Technical Summary. AJ *120*(3), 1579–1587. arXiv:astro-ph/0006396.

Zhu, G. B., J. Comparat, J.-P. Kneib, T. Delubac, A. Raichoor, et al. (2015). Near-ultraviolet Spectroscopy of Star-forming Galaxies from eBOSS: Signatures of Ubiquitous Galactic-scale Outflows. ApJ *815*(1), 48. arXiv:1507.07979.

Zou, H., X. Zhou, X. Fan, T. Zhang, Z. Zhou, et al. (2017). Project Overview of the Beijing-Arizona Sky Survey. PASP *129*(976), 064101. arXiv:1702.03653.

Measurement of the galaxy power spectrum

A galaxy redshift survey results in a catalogue of galaxy angular positions and redshifts (and possibly weights to correct for systematic effects, as we will see in Chapter 5). In Section 4.1 we detail the estimation of the galaxy power spectrum from these data catalogues and discuss the optimality of such a compression. However, the expectation value of the derived estimator does not perfectly match the pure theoretical model presented in Section 2.4.3 due to geometry effects discussed in Section 4.2. Some of them stem from the survey selection function (the expected galaxy density in the absence of clustering) which modulates the measured galaxy density. In practice, this survey selection function is partly inferred from the observed data, leading to integral constraints which we discuss in Section 4.3.

4.1 Estimation of the galaxy power spectrum

We could have followed the seminal paper Feldman et al. (1994) or Yamamoto et al. (2006) which derive the optimal weighting scheme (the so-called FKP weights, see below) from a variational principle, but we rather take a more general approach of optimal quadratic estimator. The beginning of this section is inspired from Tegmark et al. (1998) and simplifies the calculation of Abramo et al. (2016) to the single tracer case we are interested in.

4.1.1 Data compression

From the set of positions $\mathbf{X} = (\mathbf{x}_i)_{i\in[1,N_g]}$ (and weights) provided by galaxy redshift surveys we wish to measure cosmological parameters $\mathbf{p}$ with the best accuracy possible. A usual approach to parameter estimation is based on the likelihood function $L(\mathbf{X}|\mathbf{p})$, the probability to observe the data vector $\mathbf{X}$ if the true value of the parameters is $\mathbf{p}$. $\mathbf{p}$ are often estimated through the maximisation of L. In a Bayesian approach, taking the product of $L(\mathbf{X}|\mathbf{p})$ with some prior on the parameters $\pi(\mathbf{p})$ gives (up to some normalisation) the posterior, i.e. the probability distribution of the cosmological parameters.

Note that $L(\mathbf{X}|\mathbf{p})$ is a *joint* probability function of the N_g galaxy positions $\mathbf{x}_i$. These galaxies are clustered — which is precisely what we want to probe — and hence $L(\mathbf{X}|\mathbf{p})$ cannot be split into a product of one dimensional probabilities as would be possible if the galaxy positions were not correlated.

Hence, we want to *compress* the data vector $\mathbf{X}$ in a vector $\mathbf{d}$ of smaller size in such a way that (i) the likelihood $L(\mathbf{d}|\mathbf{p})$ is easy to compute (both theoretically and numerically) and (ii) is lossless. Let us define the Fisher matrix (Fisher, 1935):

$$F_{\mu\nu} = \left\langle \frac{\partial \ln L}{\partial p_\mu} \frac{\partial \ln L}{\partial p_\nu} \right\rangle \tag{4.1}$$

The Cramér-Rao bound ensures that no unbiased method can measure parameters $\mathbf{p}$ with a covariance $\mathbf{C}^p$ less than [1] $\mathbf{F}^{-1}$. If $F_{\mu\nu}$ remains the same when $\mathbf{X}$ is compressed into $\mathbf{d}$ then the compression is lossless.

Let us first examine the statistical properties of the galaxy density field $n(\mathbf{x})$, which can be built from the galaxy positions $\mathbf{x}_i$ following $n(\mathbf{x}) = \sum_{i\in[1,N_g]} \delta_D^{(3)}(\mathbf{x} - \mathbf{x}_i)$.

1. i.e. $\mathbf{C}^p - \mathbf{F}^{-1}$ is positive semidefinite.

4.1.2 The galaxy density field

In the following we assume that galaxy positions are a random Poisson point-process realisation of an underlying Gaussian field $1+\delta(\mathbf{x})$ modulated by the survey selection function $\bar{n}(\mathbf{x})$ providing the expected density of galaxies at position $\mathbf{x}$ in the absence of clustering.

Let us show that the galaxy density field n obeys (Feldman et al., 1994):

$$\boxed{\langle n(\mathbf{x})n(\mathbf{x}')\rangle = \bar{n}(\mathbf{x})\bar{n}(\mathbf{x}')\left[1+\xi(\mathbf{x},\mathbf{x}')\right] + \bar{n}(\mathbf{x})\delta_D^{(3)}(\mathbf{x}-\mathbf{x}')} \tag{4.2}$$

with $\xi(\mathbf{x},\mathbf{x}') = \langle\delta(\mathbf{x})\delta(\mathbf{x}')\rangle$ the 2-point correlation function of the density contrast $\delta(\mathbf{x})$.

Let $g(\mathbf{x},\mathbf{x}')$ be an arbitrary continuous function. We have:

$$\left\langle \int d^3\mathbf{x}\int d^3\mathbf{x}' g(\mathbf{x},\mathbf{x}')n(\mathbf{x})n(\mathbf{x}')\right\rangle = \int d^3\mathbf{x}\int d^3\mathbf{x}' g(\mathbf{x},\mathbf{x}')\left\langle n(\mathbf{x})n(\mathbf{x}')\right\rangle. \tag{4.3}$$

Let us define a regular grid of cells with volume δV such that each cell i has an occupation number $n_i \in \{0,1\}$. Let us denote the cell coordinates $\mathbf{x}_i$. We can write:

$$\int d^3\mathbf{x}\int d^3\mathbf{x}' g(\mathbf{x},\mathbf{x}')\left\langle n(\mathbf{x})n(\mathbf{x}')\right\rangle = \sum_i\sum_j g(\mathbf{x}_i,\mathbf{x}_j)\left\langle n_i n_j\right\rangle. \tag{4.4}$$

By definition:

$$\langle n_i\rangle = \bar{n}(\mathbf{x}_i)\delta V. \tag{4.5}$$

Given $i\neq j$, then $\langle n_i n_j\rangle = \bar{n}(\mathbf{x}_i)\bar{n}(\mathbf{x}_j)\delta V^2\left\langle\left[1+\delta(\mathbf{x}_i)\right]\left[1+\delta(\mathbf{x}_j)\right]\right\rangle$. Since $\langle\delta(\mathbf{x})\rangle = 0$, we have:

$$\langle n_i n_j\rangle = \bar{n}(\mathbf{x}_i)\bar{n}(\mathbf{x}_j)\delta V^2\left[1+\xi(\mathbf{x}_i,\mathbf{x}_j)\right]. \tag{4.6}$$

If $i=j$, since $n_i\in\{0,1\}$, $\langle n_i^2\rangle = \langle n_i\rangle$, therefore:

$$\left\langle n_i^2\right\rangle = \bar{n}(\mathbf{x}_i)\delta V. \tag{4.7}$$

Hence:

$$\begin{aligned}
\int d^3\mathbf{x}\int d^3\mathbf{x}' g(\mathbf{x},\mathbf{x}')\left\langle n(\mathbf{x})n(\mathbf{x}')\right\rangle &= \sum_i\sum_{j\neq i} g(\mathbf{x}_i,\mathbf{x}_j)\bar{n}(\mathbf{x}_i)\bar{n}(\mathbf{x}_j)\left[1+\xi(\mathbf{x}_i,\mathbf{x}_j)\right]\delta V^2 + \sum_i g(\mathbf{x}_i,\mathbf{x}_i)\bar{n}(\mathbf{x}_i)\delta V \\
&= \int d^3\mathbf{x}\int d^3\mathbf{x}' g(\mathbf{x},\mathbf{x}')\bar{n}(\mathbf{x})\bar{n}(\mathbf{x}')\left[1+\xi(\mathbf{x},\mathbf{x}')\right] + \int d^3\mathbf{x} g(\mathbf{x},\mathbf{x})\bar{n}(\mathbf{x}) \\
&= \int d^3\mathbf{x}\int d^3\mathbf{x}' g(\mathbf{x},\mathbf{x}')\left\{\bar{n}(\mathbf{x})\bar{n}(\mathbf{x}')\left[1+\xi(\mathbf{x},\mathbf{x}')\right] + \bar{n}(\mathbf{x})\delta_D^{(3)}(\mathbf{x}-\mathbf{x}')\right\}.
\end{aligned} \tag{4.8}$$

Eq. (4.8) holds for any function g, which proves Eq. (4.3). $\bar{n}(\mathbf{x})\delta_D^{(3)}(\mathbf{x}-\mathbf{x}')$ is the Poisson shot noise term. We are actually interested in the correlation ξ, so we define the fluctuation field as:

$$\mathcal{F}(\mathbf{x}) = n(\mathbf{x}) - \bar{n}(\mathbf{x}) \tag{4.9}$$

which is a Poisson realisation of $\bar{n}(\mathbf{x})\delta(\mathbf{x})$ and for which we have:

$$\langle\mathcal{F}(\mathbf{x})\mathcal{F}(\mathbf{x}')\rangle = \bar{n}(\mathbf{x})\bar{n}(\mathbf{x}')\xi(\mathbf{x},\mathbf{x}') + \bar{n}(\mathbf{x})\delta_D^{(3)}(\mathbf{x}-\mathbf{x}'). \tag{4.10}$$

4.1.3 Optimal quadratic estimator

The fluctuation field $\mathcal{F}(\mathbf{x})$ of Eq. (4.9) contains cosmological information that we want to extract, e.g. cosmological parameters $\mathbf{p}$. Let us call $\mathcal{F} = (\mathcal{F}_i)_{i\in[1,N]}$ a spatially discretised version of $\mathcal{F}(\mathbf{x})$.

Let us assume that $\mathcal{F}$ follows Gaussian statistics. By construction its mean is $\mathbf{0}$, and we dub its covariance $\mathbf{C}$, which depends upon cosmological parameters $\mathbf{p}$. Then, the likelihood of $\mathcal{F}$ given $\mathbf{p}$ reads:

$$L(\mathcal{F}|\mathbf{p}) = \frac{1}{(2\pi)^{N/2}\,|\mathbf{C}|^{1/2}} e^{-\frac{1}{2}\mathcal{F}^T\mathbf{C}^{-1}\mathcal{F}}. \tag{4.11}$$

Let us compute the Fisher information matrix (4.1). Calling $\mathbf{Z} = \mathcal{F}\mathcal{F}^T$ and $\mathbf{A}_\mu = \mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial p_\mu}$:

$$\frac{\partial \ln L}{\partial p_\mu} = -\frac{1}{2}\frac{\partial \ln|\mathbf{C}|}{\partial p_\mu} - \frac{1}{2}\,\mathrm{Tr}\left(\frac{\partial \mathbf{C}^{-1}}{\partial p_\mu}\mathbf{Z}\right) \tag{4.12}$$

$$= -\frac{1}{2}\,\mathrm{Tr}\,(\mathbf{A}_\mu) + \frac{1}{2}\,\mathrm{Tr}\left(\mathbf{A}_\mu\mathbf{C}^{-1}\mathbf{Z}\right). \tag{4.13}$$

Since $\langle\mathbf{Z}\rangle = \mathbf{C}$, $\mathrm{Tr}\,\langle\mathbf{A}_\mu\mathbf{C}^{-1}\mathbf{Z}\rangle = \mathrm{Tr}\,(\mathbf{A}_\mu)$ and we have:

$$F_{\mu\nu} = \left\langle \frac{\partial \ln L}{\partial p_\mu}\frac{\partial \ln L}{\partial p_\nu}\right\rangle \tag{4.14}$$

$$= \frac{1}{4}\left[\left\langle \mathcal{F}^T\mathbf{A}_\mu\mathbf{C}^{-1}\mathcal{F}\mathcal{F}^T\mathbf{A}_\nu\mathbf{C}^{-1}\mathcal{F}\right\rangle - \mathrm{Tr}\,(\mathbf{A}_\mu)\,\mathrm{Tr}\,(\mathbf{A}_\nu)\right]. \tag{4.15}$$

The first term yields factors of the kind $\langle\mathcal{F}_i\mathcal{F}_j\mathcal{F}_k\mathcal{F}_l\rangle = C_{ij}C_{kl} + C_{ik}C_{jl} + C_{il}C_{jk}$ as given by the Wick theorem (2.54). The term involving $C_{ij}C_{kl}$ cancels $\mathrm{Tr}\,(\mathbf{A}_\mu)\,\mathrm{Tr}\,(\mathbf{A}_\nu)$ while $C_{ik}C_{jl} + C_{il}C_{jk}$ provides:

$$F_{\mu\nu} = \frac{1}{2}\,\mathrm{Tr}\left[\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial p_\mu}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial p_\nu}\right]. \tag{4.16}$$

Another derivation can be found in Appendix A of Vogeley and Szalay (1996).

If we define the quadratic form:

$$\hat{q}_\mu = \frac{1}{2}C^{-1}_{ik}\frac{\partial C_{kl}}{\partial p_\mu}C^{-1}_{lj}\mathcal{F}_i\mathcal{F}_j - \Delta_\mu \tag{4.17}$$

where Δ_μ is some offset which we will fix in the following, we find (using again Wick theorem) that $\mathrm{Cov}(\hat{q}_\mu, \hat{q}_\nu) = F_{\mu\nu}$. We want our estimator to have variance $F^{-1}_{\mu\nu}$ so we naturally define it as [2]:

$$\hat{p}_\mu = F^{-1}_{\mu\nu}\hat{q}_\nu \tag{4.18}$$

and the offset Δ_μ is given by:

$$\Delta_\mu = \frac{1}{2}C^{-1}_{ik}\frac{\partial C_{kl}}{\partial p_\mu} - F_{ij}\,\langle p_j\rangle\,. \tag{4.19}$$

Up to now we have called $\mathbf{p}$ *cosmological parameters*, but the above formalism holds for any quantity to be estimated from the measurements $\mathcal{F}$, such as the bandpower galaxy power spectrum. Namely, p_μ are $P(\mathbf{k}_\mu)$, the average power spectrum in a $\mathbf{k}$-volume $V_{\mathbf{k}_\mu}$:

$$P(\mathbf{k}_\mu) = \frac{1}{V_{\mathbf{k}_\mu}}\int_{V_{\mathbf{k}_\mu}} d^3k P(\mathbf{k}). \tag{4.20}$$

We derive the optimal quadratic estimator for $P(\mathbf{k}_\mu)$ in the following section.

2. In the following $\hat{}$ denotes an estimator.

4.1.4 The power spectrum estimator

Let us start by computing the Fisher information of the power spectrum. We adopt the continuum limit of the density field, such that Eq. (4.16) becomes:

$$F_{\mu\nu} = \frac{1}{2}\int d^3x_1 d^3x_2 d^3x_3 d^3x_4 C^{-1}(\mathbf{x}_1,\mathbf{x}_2)\frac{\partial C(\mathbf{x}_2,\mathbf{x}_3)}{\partial P(\mathbf{k}_\mu)}C^{-1}(\mathbf{x}_3,\mathbf{x}_4)\frac{\partial C(\mathbf{x}_4,\mathbf{x}_1)}{\partial P(\mathbf{k}_\nu)}. \tag{4.21}$$

Using Eq. (4.10), we have:

$$C(\mathbf{x}_1,\mathbf{x}_2) = \bar{n}(\mathbf{x}_1)\bar{n}(\mathbf{x}_2)\xi(\mathbf{x}_1,\mathbf{x}_2) + \bar{n}(\mathbf{x}_1)\delta_D^{(3)}(\mathbf{x}_2-\mathbf{x}_1) \tag{4.22}$$

$$= \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)}\left[\bar{n}(\mathbf{x}_1)\bar{n}(\mathbf{x}_2)P(\mathbf{k}) + \bar{n}(\mathbf{x}_1)\right]. \tag{4.23}$$

However, this expression is too complex to be inverted. Let us assume that the power spectrum is roughly constant on the range of interest, $P(\mathbf{k}) \simeq P(\mathbf{k}_\mu)$. Then $\int d^3k e^{i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)}/(2\pi)^3$ can be traded for a Dirac distribution $\delta_D^{(3)}(\mathbf{x}_2-\mathbf{x}_1)$. By definition, the inverse matrix C^{-1} verifies:

$$\int d^3x_2 C^{-1}(\mathbf{x}_1,\mathbf{x}_2)C(\mathbf{x}_2,\mathbf{x}_3) = \int d^3x_2 C(\mathbf{x}_1,\mathbf{x}_2)C^{-1}(\mathbf{x}_2,\mathbf{x}_3) = \delta_D^{(3)}(\mathbf{x}_3-\mathbf{x}_1) \tag{4.24}$$

Noting that $\int d^3x_2 \delta_D^{(3)}(\mathbf{x}_2-\mathbf{x}_1)\delta_D^{(3)}(\mathbf{x}_3-\mathbf{x}_2) = \delta_D^{(3)}(\mathbf{x}_3-\mathbf{x}_1)$ it is clear than one has just to take the inverse of the term in brackets, i.e.:

$$C^{-1}(\mathbf{x}_1,\mathbf{x}_2) \simeq \left[\bar{n}(\mathbf{x}_1)^2 P(\mathbf{k}_\mu) + \bar{n}(\mathbf{x}_1)\right]^{-1}\delta_D^{(3)}(\mathbf{x}_2-\mathbf{x}_1) = \frac{w_{\rm FKP}(\mathbf{x}_1,\mathbf{k}_\mu)}{\bar{n}(\mathbf{x}_1)}\delta_D^{(3)}(\mathbf{x}_2-\mathbf{x}_1) \tag{4.25}$$

where we used the FKP weight (Feldman et al., 1994):

$$w_{\rm FKP}(\mathbf{x},\mathbf{k}) = \frac{1}{1+\bar{n}(\mathbf{x})P(\mathbf{k})}. \tag{4.26}$$

Let us move to the functional derivative. We have:

$$\frac{\partial P(\mathbf{k})}{\partial P(\mathbf{k}_\mu)} = \begin{cases} 1 & \mathbf{k}\in V_{\mathbf{k}_\mu} \\ 0 & \text{else,} \end{cases} \tag{4.27}$$

thus:

$$\frac{\partial C(\mathbf{x}_1,\mathbf{x}_2)}{\partial P(\mathbf{k}_\mu)} = \int_{V_{\mathbf{k}_\mu}} \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)}\bar{n}(\mathbf{x}_1)\bar{n}(\mathbf{x}_2). \tag{4.28}$$

The Fisher information then reads:

$$F_{\mu\nu} = \frac{1}{2}\int d^3x_1 d^3x_3 \int_{V_{\mathbf{k}_\mu}}\frac{d^3k_1}{(2\pi)^3}\int_{V_{\mathbf{k}_\nu}}\frac{d^3k_3}{(2\pi)^3} e^{i(\mathbf{k}_1-\mathbf{k}_3)\cdot(\mathbf{x}_3-\mathbf{x}_1)}\bar{n}(\mathbf{x}_1)\bar{n}(\mathbf{x}_3)w_{\rm FKP}(\mathbf{x}_1,\mathbf{k}_\mu)w_{\rm FKP}(\mathbf{x}_3,\mathbf{k}_\nu) \tag{4.29}$$

Inverting the spatial and Fourier space integrals, we make appear Fourier transforms of the survey selection function:

$$W(\mathbf{x},\mathbf{k}_\mu) = \bar{n}(\mathbf{x})w_{\rm FKP}(\mathbf{x},\mathbf{k}_\mu), \tag{4.30}$$

which characterises the survey footprint. Hence:

$$F_{\mu\nu} = \frac{1}{2}\int_{V_{\mathbf{k}_\mu}}\frac{d^3k_1}{(2\pi)^3}\int_{V_{\mathbf{k}_\nu}}\frac{d^3k_3}{(2\pi)^3}W(\mathbf{k}_3-\mathbf{k}_1,\mathbf{k}_\mu)W(\mathbf{k}_1-\mathbf{k}_3,\mathbf{k}_\nu) \tag{4.31}$$

$$\simeq \frac{\delta_{\mu\nu}V_{\mathbf{k}_\mu}}{2\,(2\pi)^3}\int d^3x \bar{n}(\mathbf{x})^2 w_{\rm FKP}(\mathbf{x},\mathbf{k}_\mu)^2. \tag{4.32}$$

The second line follows from the assumption that the scales probed are much smaller that the survey typical size, i.e.:

$$W(\mathbf{k}_3-\mathbf{k}_1,\mathbf{k}_\mu)W(\mathbf{k}_1-\mathbf{k}_3,\mathbf{k}_\nu) \to (2\pi)^3 \int d^3x \bar{n}(\mathbf{x})^2 w_{\rm FKP}(\mathbf{x},\mathbf{k}_\mu) w_{\rm FKP}(\mathbf{x},\mathbf{k}_\nu)\delta_D^{(3)}(\mathbf{k}_3-\mathbf{k}_1). \quad (4.33)$$

Let us define the effective volume:

$$V_{\rm eff} = \int d^3x \bar{n}(\mathbf{x})^2 w_{\rm FKP}(\mathbf{x},\mathbf{k}_\mu)^2 P(\mathbf{k}_\mu)^2. \quad (4.34)$$

The relative uncertainty in the bandpowers of the power spectrum is hence:

$$\frac{\sigma^2_{\hat{P}(\mathbf{k}_\mu)}}{P(\mathbf{k}_\mu)^2} = \frac{F^{-1}_{\mu\mu}}{P(\mathbf{k}_\mu)^2} = \frac{2\,(2\pi)^3}{V_{\mathbf{k}_\mu} V_{\rm eff}}. \quad (4.35)$$

Of course, the fractional uncertainty reduces as the $\mathbf{k}$-space voxel increases. Also, in the case $\bar{n}P(\mathbf{k}_\mu) \gg 1$, $V_{\rm eff} \to V$ the survey volume, and:

$$\frac{\sigma^2_{\hat{P}(\mathbf{k}_\mu)}}{P(\mathbf{k}_\mu)^2} \to \frac{2\,(2\pi)^3}{V_{\mathbf{k}_\mu} V}, \quad (4.36)$$

a limit known as *cosmic variance*.

Let us now compute the quadratic form Eq. (4.17):

$$\hat{q}_\mu = \frac{1}{2}\int d^3x_1 d^3x_2 d^3x_3 d^3x_4 C(\mathbf{x}_1,\mathbf{x}_2)^{-1} \frac{\partial C(\mathbf{x}_2,\mathbf{x}_3)}{\partial P(\mathbf{k}_\mu)} C(\mathbf{x}_3,\mathbf{x}_4)^{-1} \mathcal{F}(\mathbf{x}_1)\mathcal{F}(\mathbf{x}_4) - \Delta_\mu. \quad (4.37)$$

Using Eq. (4.25) and Eq. (4.28), we have:

$$\hat{q}_\mu = \frac{1}{2}\int d^3x_1 d^3x_3 \int_{V_{\mathbf{k}_\mu}} \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot(\mathbf{x}_3-\mathbf{x}_1)} w_{\rm FKP}(\mathbf{x}_1,\mathbf{k}_\mu) w_{\rm FKP}(\mathbf{x}_3,\mathbf{k}_\mu)\mathcal{F}(\mathbf{x}_1)\mathcal{F}(\mathbf{x}_3) - \Delta_\mu. \quad (4.38)$$

Using Eq. (4.32) and Eq. (4.18) gives (in the diagonal limit):

$$\hat{P}(\mathbf{k}_\mu) = \frac{1}{AV_{\mathbf{k}_\mu}} \int_{V_{\mathbf{k}_\mu}} d^3k \int d^3x_1 d^3x_3 e^{i\mathbf{k}\cdot(\mathbf{x}_3-\mathbf{x}_1)} w_{\rm FKP}(\mathbf{x}_1,\mathbf{k}_\mu) w_{\rm FKP}(\mathbf{x}_3,\mathbf{k}_\mu)\mathcal{F}(\mathbf{x}_1)\mathcal{F}(\mathbf{x}_3) - \frac{\Delta_\mu}{F_{\mu\mu}} \quad (4.39)$$

where $A = \int d^3x \bar{n}(\mathbf{x})^2 w_{\rm FKP}(\mathbf{x},\mathbf{k}_\mu)^2$.

Let us estimate Δ_μ as given in Eq. (4.19):

$$\Delta_\mu = \frac{1}{2}\int d^3x_1 d^3x_2 C(\mathbf{x}_1,\mathbf{x}_2)^{-1} \frac{\partial C(\mathbf{x}_2,\mathbf{x}_1)}{\partial P(\mathbf{k}_\mu)} - F_{\mu\nu}P(\mathbf{k}_\nu) \quad (4.40)$$

which reduces to (using Eq. (4.25), Eq. (4.28) and Eq. (4.32)):

$$\Delta_\mu = \frac{1}{2}\int_{V_{\mathbf{k}_\mu}} \frac{d^3k}{(2\pi)^3} \int d^3x \bar{n}(\mathbf{x}) w_{\rm FKP}(\mathbf{x},\mathbf{k}_\mu) - F_{\mu\nu}P(\mathbf{k}_\nu) \quad (4.41)$$

$$= \frac{1}{2}\int_{V_{\mathbf{k}_\mu}} \frac{d^3k}{(2\pi)^3} \int d^3x \left[\bar{n}(\mathbf{x}) w_{\rm FKP}(\mathbf{x},\mathbf{k}_\mu) - \bar{n}(\mathbf{x})^2 w_{\rm FKP}(\mathbf{x},\mathbf{k}_\mu)^2 P(\mathbf{k}_\mu)\right] \quad (4.42)$$

$$= \frac{V_{\mathbf{k}_\mu}}{2\,(2\pi)^3} \int d^3x \bar{n}(\mathbf{x}) w_{\rm FKP}(\mathbf{x},\mathbf{k}_\mu)^2. \quad (4.43)$$

Therefore, Eq. (4.39) gives the FKP estimator (Feldman et al., 1994):

$$\hat{P}(\mathbf{k}_\mu) = \frac{1}{AV_{\mathbf{k}_\mu}} \int_{V_{\mathbf{k}_\mu}} d^3k \left[\int d^3x_1 d^3x_3 e^{i\mathbf{k}\cdot(\mathbf{x}_3-\mathbf{x}_1)} w_{\rm FKP}(\mathbf{x}_1,\mathbf{k}_\mu) w_{\rm FKP}(\mathbf{x}_3,\mathbf{k}_\mu) \mathcal{F}(\mathbf{x}_1)\mathcal{F}(\mathbf{x}_3) - \int d^3x \bar{n}(\mathbf{x}) w_{\rm FKP}(\mathbf{x},\mathbf{k}_\mu)^2 \right] \tag{4.44}$$

The second term subtracts the shot noise. Sources of suboptimality in this estimator are (i) $C(\mathbf{x}_1,\mathbf{x}_2)^{-1}$ is not diagonal and (ii) in practice, the galaxy power spectrum evolves over the footprint. Moreover, this estimator is biased since we took the window function $|W(\mathbf{k})|^2$ to be localised in $\mathbf{k}$ to state that $F_{\mu\nu}$ is diagonal in Eq. (4.39). In fact, Eq. (4.44) estimates the convolution $(P * |W|)(\mathbf{k})$ — of which the multiplication by the matrix $F^{-1}_{\mu\nu}$ (see Eq. (4.18)) would have been a discrete matrix deconvolution. This window function effect will be discussed in Section 4.2.2.

We are actually interested in the anisotropy of the power spectrum with respect to the line of sight $\hat{\eta}$, which can be probed with the multipoles of the power spectrum:

$$P_\ell(k) = \frac{2\ell+1}{2} \int_{-1}^{1} d(\hat{\mathbf{k}}\cdot\hat{\eta}) P(k, \hat{\mathbf{k}}\cdot\hat{\eta}) \mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\eta}) \tag{4.45}$$

for which Yamamoto et al. (2006) proposed the following estimator:

$$\hat{P}_\ell(k_\mu) = \frac{2\ell+1}{AV_{k_\mu}} \int_{V_{k_\mu}} d^3k \int d^3x_1 d^3x_2 e^{i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)} w_{\rm FKP}(\mathbf{x}_1,k_\mu) w_{\rm FKP}(\mathbf{x}_2,k_\mu) \mathcal{F}(\mathbf{x}_1)\mathcal{F}(\mathbf{x}_2) \mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}_m) - P_\ell^{\rm noise}(k_\mu) \tag{4.46}$$

where we take the so-called *midpoint* line of sight $\mathbf{x}_m = (\mathbf{x}_1+\mathbf{x}_2)/2$. Note that the definition of the line of sight of the pair of galaxies at position $\mathbf{x}_1$ and $\mathbf{x}_2$ is only clear in the limit where the galaxies are very far from the observer, $|\mathbf{x}_2-\mathbf{x}_1| / |\mathbf{x}_m| \ll 1$: the so-called *local plane-parallel* approximation. In Eq. (4.46) V_{k_μ} is a spherical shell of radius k_μ, and for simplicity we neglected the variation of $w_{\rm FKP}(\mathbf{x},\mathbf{k})$ with the direction of $\mathbf{k}$ (though we would expect some due to redshift space distortions).

The shot noise term reads:

$$P_\ell^{\rm noise}(k_\mu) = \frac{2\ell+1}{AV_{k_\mu}} \int_{V_{k_\mu}} d^3k \int d^3x \bar{n}(\mathbf{x}) w_{\rm FKP}(\mathbf{x},k_\mu)^2 \mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}) \tag{4.47}$$

$$= \frac{\delta_{\ell 0}}{A} \int d^3x \bar{n}(\mathbf{x}) w_{\rm FKP}(\mathbf{x},k_\mu)^2. \tag{4.48}$$

Note that the integration in Eq. (4.48) is not fully correct if performed on a finite $\mathbf{k}$-space grid, but a fair approximation if the sampling is high enough, which we will make sure is the case in the following.

4.1.5 An optimal estimator for the growth rate

A careful reader may point out that we have not answered the fundamental question of Section 4.1.1: is the power spectrum compression lossless for the parameter we aim at measuring, e.g. the growth rate of structure f?

For this, let us build the optimal quadratic estimator for the growth rate of structure f. We follow the same reasoning as Castorina et al. (2019) did for the amplitude of primordial non-Gaussianity $f_{\rm NL}^{\rm local}$. Using the Kaiser formula (2.164), the power spectrum of the density contrasts at redshift z_1 et z_2 reads (in redshift space):

$$P(k,\mu,z_1,z_2) = \left(b(z_1)+f\mu^2\right)D(z_1)\left(b(z_2)+f\mu^2\right)D(z_2)P_m^{\rm lin}(k) \tag{4.49}$$

where b is the linear bias, D the linear growth rate (normalised such that $D(0)=1$) and $P_m^{\rm lin}(k)$ is the linear power spectrum taken at $z=0$. We take $\mu = \hat{\mathbf{k}}\cdot\hat{\mathbf{x}}_m$, but the choice of line of sight will not impact our conclusions. For simplicity we assumed f to be constant over the redshift range. Then:

$$\frac{\partial P(k,\mu,z_1,z_2)}{\partial f} = \left(b(z_1)+f\mu^2\right)D(z_1)\mu^2 D(z_2)P_m^{\rm lin}(k) + 1 \leftrightarrow 2 \tag{4.50}$$

Hence:

$$\frac{\partial C(\mathbf{x}_1,\mathbf{x}_2)}{\partial f} = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)}\bar{n}(\mathbf{x}_1)\bar{n}(\mathbf{x}_2)\left[\left(b(z_1)\mu^2+f\mu^4\right)D(z_1)D(z_2)P_m^{\rm lin}(k) + 1\leftrightarrow 2\right] \tag{4.51}$$

Assuming again that $C(\mathbf{x}_1,\mathbf{x}_2)^{-1}$ is diagonal, we have:

$$\hat{q}_f = \frac{1}{2}\int\frac{d^3k}{(2\pi)^3}P_m^{\rm lin}(k)\int d^3x_1 d^3x_3 w_{\rm FKP}(\mathbf{x}_1,\mathbf{k})w_{\rm FKP}(\mathbf{x}_3,\mathbf{k})\mathcal{F}(\mathbf{x}_1)\mathcal{F}(\mathbf{x}_3) \left[\left(b(z_1)\mu^2+f\mu^4\right)D(z_1)D(z_3)+1\leftrightarrow 3\right]e^{i\mathbf{k}\cdot(\mathbf{x}_3-\mathbf{x}_1)} - \Delta_f. \tag{4.52}$$

Then, noticing that:

$$\mu^2 = \frac{1}{3}\mathcal{L}_0(\mu) + \frac{2}{3}\mathcal{L}_2(\mu) \tag{4.53}$$

$$\mu^4 = \frac{1}{5}\mathcal{L}_0(\mu) + \frac{4}{7}\mathcal{L}_2(\mu) + \frac{8}{35}\mathcal{L}_4(\mu) \tag{4.54}$$

we find:

$$\hat{q}_f = \frac{1}{2}\int\frac{d^3k}{(2\pi)^3}P_m^{\rm lin}(k)\sum_{\ell\in\{0,2,4\}}\int d^3x_1 d^3x_2 e^{i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)}w_{\rm FKP}(\mathbf{x}_1,\mathbf{k})w_{\rm FKP}(\mathbf{x}_2,\mathbf{k})\mathcal{F}(\mathbf{x}_1)\mathcal{F}(\mathbf{x}_2) \left[w(z_1)w_\ell(z_2)+1\leftrightarrow 2\right]\mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}_m) - \Delta_f. \tag{4.55}$$

with:

$$w(z) = D(z) \tag{4.56}$$

$$w_0(z) = \left[\frac{1}{3}b(z)+\frac{1}{5}f\right]D(z) \tag{4.57}$$

$$w_2(z) = \left[\frac{2}{3}b(z)+\frac{4}{7}f\right]D(z) \tag{4.58}$$

$$w_4(z) = \frac{8}{35}fD(z) \tag{4.59}$$

If we had taken f to be z-dependent, we would have wanted to measure its average value over a redshift bin, leading to the additional constraint that spatial integrals should be performed within that redshift bin — just as with the constraint (4.27). Up to a normalisation factor (given by the Fisher information) and the constant bias Δ_f, Eq. (4.55) is the quadratic optimal estimator for f. If we neglect the variations of b, D over the redshift range, we see that the building block is just the Yamamoto estimator (4.46). Hence, the compression brought by the latter is lossless for f. Let us recall the key assumptions that guarantee this optimality:

1. the galaxy density field follows Gaussian statistics
2. the covariance of the galaxy density field is taken to be diagonal in redshift space to compute FKP weights
3. the local plane-parallel approximation i.e. $|\mathbf{x}_2 - \mathbf{x}_1| \,/\, |\mathbf{x}_m| \ll 1$, holds
4. there is no redshift evolution of the redshift space galaxy power spectrum

For an approach taking into account the redshift evolution of the power spectrum over the survey footprint to optimally measure redshift space distortions, we refer the reader to e.g. Ruggeri et al. (2019).

4.1.6 Practical computation

In practice, we take FKP weights to be scale independent:

$$\boxed{w_{\rm FKP}(\mathbf{x}) \equiv \frac{1}{1+\bar{n}(\mathbf{x})P_0}} \tag{4.60}$$

where P_0 is the typical value of the power spectrum at the scales of interest.

Building the fluctuation field (4.9) requires the survey selection function i.e. the expected (weighted) number density of galaxies in the absence of clustering, $\bar{n}(\mathbf{x}) = \langle n(\mathbf{x})\rangle$. In practice, it is Poisson-sampled by a synthetic catalogue[3]. Let us introduce galaxy weights w_g and random weights w_s which include FKP weights and possibly other weights to e.g. correct for observational systematics. Let us denote the random weighted density n_s, rename the galaxy weighted density n_g, and dub its expected value the survey selection function $W(\mathbf{x}) = \langle n_g(\mathbf{x})\rangle$. We call:

$$\alpha = \frac{\int d^3x W(\mathbf{x})}{\int d^3x n_s(\mathbf{x})} \tag{4.61}$$

the ratio of the (ensemble- and volume-averaged) density of galaxies to random objects. The fluctuation field (4.9) is thus recast into the FKP field (Feldman et al., 1994):

$$\boxed{F(\mathbf{x}) = n_g(\mathbf{x}) - \alpha n_s(\mathbf{x})} \tag{4.62}$$

Eq. (4.2) ignores galaxy weights. In fact, each weighted galaxy contributes a total $w_g^2(\mathbf{x})$ to the tiny cell it belongs to, yielding a shot noise contribution $W(\mathbf{x})w_g(\mathbf{x})\delta_D^{(3)}(\mathbf{x}-\mathbf{x}')$ in Eq. (4.2):

$$\langle n_g(\mathbf{x})n_g(\mathbf{x}')\rangle = W(\mathbf{x})W(\mathbf{x}')\left[1+\xi(\mathbf{x},\mathbf{x}')\right] + W(\mathbf{x})w_g(\mathbf{x})\delta_D^{(3)}(\mathbf{x}-\mathbf{x}'). \tag{4.63}$$

3. We assume throughout this manuscript that the density of this synthetic catalogue is sufficiently high to avoid systematic bias from undersampling the survey selection function.

Applying the same equation to n_s (in which case $\xi = 0$):

$$\alpha^2 \langle n_s(\mathbf{x}) n_s(\mathbf{x}') \rangle = W(\mathbf{x}) W(\mathbf{x}') + \alpha W(\mathbf{x}) w_s(\mathbf{x}) \delta_D^{(3)}(\mathbf{x} - \mathbf{x}'). \tag{4.64}$$

Finally, the galaxy - random correlation is shot noise free, $\alpha \langle n_g(\mathbf{x}) n_s(\mathbf{x}') \rangle = W(\mathbf{x}) W(\mathbf{x}')$. Hence:

$$\langle F(\mathbf{x}) F(\mathbf{x}') \rangle = W(\mathbf{x}) W(\mathbf{x}') \left[1 + \xi(\mathbf{x}, \mathbf{x}')\right] + W(\mathbf{x}) \left[w_g(\mathbf{x}) + \alpha w_s(\mathbf{x})\right] \delta_D^{(3)}(\mathbf{x} - \mathbf{x}'), \tag{4.65}$$

which is similar to Eq. (4.10) except for the shot noise term. Thus, the total shot noise term (4.48) reads:

$$P_\ell^{\rm noise}(k_\mu) = \frac{\delta_{\ell 0}}{A} \int d^3x W(\mathbf{x}) \left[w_g(\mathbf{x}) + \alpha w_s(\mathbf{x})\right] \tag{4.66}$$

$$= \frac{\delta_{\ell 0}}{A} \int d^3x \left[W(\mathbf{x}) w_g(\mathbf{x}) + \alpha^2 n_s(\mathbf{x}) w_s(\mathbf{x})\right]. \tag{4.67}$$

In order to keep the Poisson shot noise as low as possible, we will typically have $n_s \sim 20 - 50 n_g$. The normalisation factor A is given by:

$$A = \alpha^2 \int d^3r n_s^2(\mathbf{r}). \tag{4.68}$$

In practice, volume integrals are replaced by discrete sums following:

$$\int d^3x n_g(\mathbf{x}) \cdots \to \sum_{i=1}^{N_g} w_{g,i} \cdots \qquad \int d^3x n_s(\mathbf{x}) \cdots \to \sum_{i=1}^{N_s} w_{s,i} \cdots \tag{4.69}$$

where N_g and N_s denote the number of galaxies and randoms, respectively. In this case,

$$P_\ell^{\rm noise}(k_\mu) = \frac{\delta_{\ell 0}}{A} \left[\sum_{i=1}^{N_g} w_{g,i}^2 + \alpha^2 \sum_{i=1}^{N_s} w_{s,i}^2\right] \qquad A = \alpha^2 \sum_{i=1}^{N_s} w_{s,i} n_{s,i}, \tag{4.70}$$

where we took $\int d^3x W(\mathbf{x}) \simeq \int d^3x n_g(\mathbf{x})$ in the first shot noise term.

Note that Eq. (4.46) involves a double sum, which is non-separable due to the choice of line of sight $\hat{\mathbf{x}}_m$. Therefore, the complexity scales as $\mathcal{O}(N_k \times N_s^2)$ where N_k is the number of Fourier modes to be evaluated in bin k. Complexity is dominated by the random objects as $N_s \gg N_g$. Hence, as suggested in Yamamoto et al. (2006) we rather choose one of the galaxies of each pair as line of sight, such that the double sum can be split into:

$$\boxed{\hat{P}_\ell(k_\mu) = \frac{2\ell + 1}{A V_{k_\mu}} \int_{V_{k_\mu}} d^3k F_0(\mathbf{k}) F_\ell(-\mathbf{k}) - P_\ell^{\rm noise}(k_\mu)} \tag{4.71}$$

where we used the quantities:

$$F_\ell(\mathbf{k}) = \int d^3x e^{i\mathbf{k}\cdot\mathbf{x}} F(\mathbf{x}) \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{x}}). \tag{4.72}$$

With this scheme, complexity drops to $\mathcal{O}(N_k \times N_s)$. In practice, however, correctly sampling the $\mathbf{k}$-space requires N_k to be quite large (typically $\gtrsim 10^3$) and N_s is typically of order 10^6, so that we still want to speed up our algorithm.

One would note that $F_0(\mathbf{k})$ is a simple Fourier transform, and therefore can be computed with a Fast Fourier Transform (FFT) if F_0 is interpolated on a regular Cartesian mesh. Several

techniques exist to recast Eq. (4.72) in a simple Fourier transform form. Bianchi et al. (2015); Scoccimarro (2015) first noticed that the scalar product is separable into a configuration space and Fourier space components:

$$\hat{\mathbf{x}} \cdot \hat{\mathbf{k}} = \frac{\mathbf{x}_i}{x} \frac{\mathbf{k}_i}{k} \tag{4.73}$$

and so can be the polynomial in $\hat{\mathbf{k}} \cdot \hat{\mathbf{x}}$ $\mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{x}})$. For example, for $\ell = 2$, $\mathcal{L}_2(x) = 5/2\left[3x^2 - 1\right]$ i.e.:

$$F_2(\mathbf{k}) = \frac{5}{2}\left[3A_2(\mathbf{k}) - F_0(\mathbf{k})\right] \tag{4.74}$$

with:

$$A_2(\mathbf{k}) = \sum_{i,j \in [1,3]^2} \frac{k_i k_j}{k^2} B_{ij}(\mathbf{k}) \tag{4.75}$$

where:

$$B_{ij}(\mathbf{k}) = \int d^3x e^{i\mathbf{k}\cdot\mathbf{x}} F(\mathbf{x}) \frac{x_i x_j}{x^2} \tag{4.76}$$

is a simple Fourier transform. The complexity of FFTs is $\mathcal{O}(N_m \log N_m)$ where N_m is the total mesh size. Many libraries exist to calculate FFTs, an example of which is `FFTW` (Frigo and Johnson, 2005). However, with the above implementation, the number of FFTs required to estimate the power spectrum multipole of order ℓ is $(\ell + 1)(\ell + 2)/2$. Though we implemented our own power spectrum estimator following the prescription of Bianchi et al. (2015); Scoccimarro (2015) we eventually used the `nbodykit` implementation of Hand et al. (2018). It relies on the observation that (Hand et al., 2017):

$$\mathcal{L}_\ell(\hat{\mathbf{x}} \cdot \hat{\mathbf{k}}) = \frac{4\pi}{2\ell + 1} \sum_{m=-\ell}^{m=\ell} Y_{\ell m}(\hat{\mathbf{x}}) Y^\star_{\ell m}(\hat{\mathbf{k}}) \tag{4.77}$$

where $Y_{\ell m}$ are the spherical harmonics, such that:

$$F_\ell(\mathbf{k}) = \frac{4\pi}{2\ell + 1} \sum_{m=-\ell}^{m=\ell} Y^\star_{\ell m}(\hat{\mathbf{k}}) \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} F(\mathbf{x}) Y_{\ell m}(\hat{\mathbf{x}}) \tag{4.78}$$

In this case, the calculation of any multipole only requires $2\ell + 1$ FFTs.

As mentioned previously, the FKP field F must be interpolated on a mesh, introducing aliasing effects to be accounted for.

4.1.7 Aliasing and interlacing

Let us write the FKP field F:

$$F(\mathbf{x}) = \sum_{i=1}^{N} F_i \delta_D^{(3)}(\mathbf{x} - \mathbf{x}_i) \tag{4.79}$$

where $N = N_g + N_s$ is the total number of galaxies (observed and random), F_i is the weight of galaxies or randoms. In practice, F_i can also include spherical harmonics at $\hat{\mathbf{x}}_i$ to compute $F_\ell(\mathbf{k})$ as mentioned previously. The interpolation of F on the mesh corresponds to the evaluation on the mesh of the continuous function:

$$\widetilde{F}(\mathbf{x}) = \int d^3\mathbf{x}' F(\mathbf{x}') W(\mathbf{x} - \mathbf{x}'), \tag{4.80}$$

where W is a window function that describes the scheme to assign the weights F_i to the mesh. For ease of computation this assignment scheme is usually based on a separable piecewise polynomial, i.e.:

$$W(\mathbf{r}) = W^{(p)}\left(\frac{r_x}{H}\right) W^{(p)}\left(\frac{r_y}{H}\right) W^{(p)}\left(\frac{r_z}{H}\right), \tag{4.81}$$

where H is the cell size, which we take to be the same in each of the x, y and z axes (without loss of generality). p, the order of W, corresponds to the number of mesh nodes (along each axis) where a weight F_i is allocated.

We can for example take $W^{(p)}$ to be the $(p-1)$-th order convolution of the top hat function $W^{(1)} = \mathbb{1}_{\left[-\frac{1}{2},\frac{1}{2}\right]}$ (see NGP below) with itself. Here are the expressions up to fourth order, given as a function of s, the distance between galaxy i and the mesh node divided by the cell size H:

Order 1: nearest grid point (NGP)

$$W^{(1)}(s) = \begin{cases} 1 & \text{if } |s| < \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{4.82}$$

Order 2: cloud-in-cell (CIC)

$$W^{(2)}(s) = \begin{cases} 1 - |s| & \text{if } |s| < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.83}$$

Order 3: triangular shaped cloud (TSC)

$$W^{(3)}(s) = \begin{cases} \frac{3}{4} - |s|^2 & \text{if } |s| < \frac{1}{2} \\ \frac{1}{2}\left(\frac{3}{2} - |s|^2\right)^2 & \text{if } \frac{1}{2} \leqslant |s| < \frac{3}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{4.84}$$

Ordre 4 : piecewise cubic spline (PCS)

$$W^{(4)}(s) = \begin{cases} \frac{1}{6}\left(4 - 6|s|^2 + 3|s|^3\right) & \text{if } |s| < 1 \\ \frac{1}{6}\left(2 - |s|\right)^3 & \text{if } 1 \leqslant |s| < 2 \\ 0 & \text{otherwise.} \end{cases} \tag{4.85}$$

The field $\widetilde{F}$ is evaluated at the mesh nodes $\mathbf{r}_m = \mathbf{m}H$ with $\mathbf{m}$ an integer vector:

$$\widetilde{F}^M(\mathbf{r}) = \begin{cases} \widetilde{F}(\mathbf{r}_m) & \text{if } \mathbf{r} = \mathbf{r}_m \\ 0 & \text{otherwise.} \end{cases} \tag{4.86}$$

FFTs in fact compute the discrete Fourier transform $\widetilde{F}^M(\mathbf{k})$ of $\widetilde{F}^M(\mathbf{r})$. Let us specify how $\widetilde{F}^M(\mathbf{k})$ relates to $\widetilde{F}(\mathbf{k})$ and then $F(\mathbf{k})$. For this we define the sampling function Ш as the sum over the integer vector $\mathbf{m}$:

$$\text{Ш}(\mathbf{r}) = \sum_{\mathbf{m}\in\mathbb{Z}^3} \delta_D^{(3)}(\mathbf{r} - \mathbf{m}H). \tag{4.87}$$

Therefore, by definition of $\widetilde{F}^M(\mathbf{k})$:

$$\widetilde{F}^M(\mathbf{k}) = \sum_{\mathbf{m}\in\mathbb{Z}^3} \widetilde{F}^M(\mathbf{r}_m) e^{-i\mathbf{r}_m\cdot\mathbf{k}} \tag{4.88}$$

$$= \sum_{\mathbf{m}\in\mathbb{Z}^3} \widetilde{F}(\mathbf{r}_m) e^{-i\mathbf{r}_m\cdot\mathbf{k}} \tag{4.89}$$

$$= \int d^3\mathbf{r}\,\text{Ш}(\mathbf{r})\widetilde{F}(\mathbf{r}) e^{-i\mathbf{r}\cdot\mathbf{k}}. \tag{4.90}$$

From this equality, the convolution theorem gives:

$$\widetilde{F}^M(\mathbf{k}) = \frac{1}{(2\pi)^3}\int d^3\mathbf{k}'\text{Ш}(\mathbf{k}')\widetilde{F}(\mathbf{k}-\mathbf{k}'). \tag{4.91}$$

The Fourier transform of a Dirac comb is a Dirac comb:

$$\text{Ш}(\mathbf{k}) = \frac{(2\pi)^3}{H^3}\sum_{\mathbf{n}\in\mathbb{Z}^3}\delta_D(\mathbf{k}-2k_N\mathbf{n}), \tag{4.92}$$

were the sum runs on $\mathbf{n}$ integer and $k_N = \pi/H$ is the Nyquist wavenumber. Therefore:

$$\widetilde{F}^M(\mathbf{k}) = \frac{1}{H^3}\sum_{\mathbf{n}\in\mathbb{Z}^3}\widetilde{F}(\mathbf{k}-2k_N\mathbf{n}). \tag{4.93}$$

We have thus expressed $\widetilde{F}^M(\mathbf{k})$ as a function of the field interpolated on the mesh, $\widetilde{F}(\mathbf{k})$. Recalling Eq. (4.80) and using the convolution theorem once again:

$$\widetilde{F}^M(\mathbf{k}) = \frac{1}{H^3}\sum_{\mathbf{n}\in\mathbb{Z}^3}W(\mathbf{k}-2k_N\mathbf{n})F(\mathbf{k}-2k_N\mathbf{n}). \tag{4.94}$$

Hence, every wavenumber $\mathbf{k}-2k_N\mathbf{n}$ of the quantity F — which we want to evaluate — contributes to the quantity $\widetilde{F}^M(\mathbf{k})$ which we compute in practice: this effect is called aliasing. However, the sum is weighted by the function $W(\mathbf{k})$. If W is localised in $\mathbf{k}$-space, then only the first term, $\mathbf{n}=\mathbf{0}$, will remain, and dividing $\widetilde{F}^M(\mathbf{k})$ by $W(\mathbf{k})$ will yield the correct result $F(\mathbf{k})$. However, for W to decrease rapidly in $\mathbf{k}$-space we need $W(\mathbf{r})$ to decrease slowly, i.e. each weight F_i must be assigned to a large number of nodes — which can become prohibitively expansive, more than the FFTs themselves. Let us review the performance of the functions $W^{(p)}(s)$ described previously:

$$\int drW^{(1)}(r/H)e^{-irk} = H\int duW^{(1)}(u)e^{-iHuk} \tag{4.95}$$

$$= H\int_{-\frac{1}{2}}^{\frac{1}{2}} due^{-i\frac{\pi uk}{k_N}} \tag{4.96}$$

$$= H\,\text{sinc}\left(\frac{\pi k}{2k_N}\right), \tag{4.97}$$

where sinc is the cardinal sinus. As $W^{(p)}(s)$ is the $(p-1)$ convolution of $W^{(1)}(s)$:

$$\int drW^{(p)}(r/H)e^{-irk} = H\,\text{sinc}^p\left(\frac{\pi k}{2k_N}\right). \tag{4.98}$$

Finally, as W is separable (see Eq. (4.81)):

$$W(\mathbf{k}) = H^3\,\text{sinc}^p\left(\frac{\pi k_x}{2k_N}\right)\text{sinc}^p\left(\frac{\pi k_y}{2k_N}\right)\text{sinc}^p\left(\frac{\pi k_z}{2k_N}\right). \tag{4.99}$$

These different windows are shown in real and Fourier space in Figure 4.1. We see that as p increases, $W(\mathbf{k})$ gets more peaked, such that the contributions from the undesirable wavenumbers become less important in the sum (4.94). In practice, in this manuscript, we consider the TSC assignment scheme (order $p=3$).

Even with this quite high order scheme, $F(\mathbf{k})$ is not well recovered close to the Nyquist frequency. Jing (2005) proposed an iterative scheme to reconstruct the true field F, which we

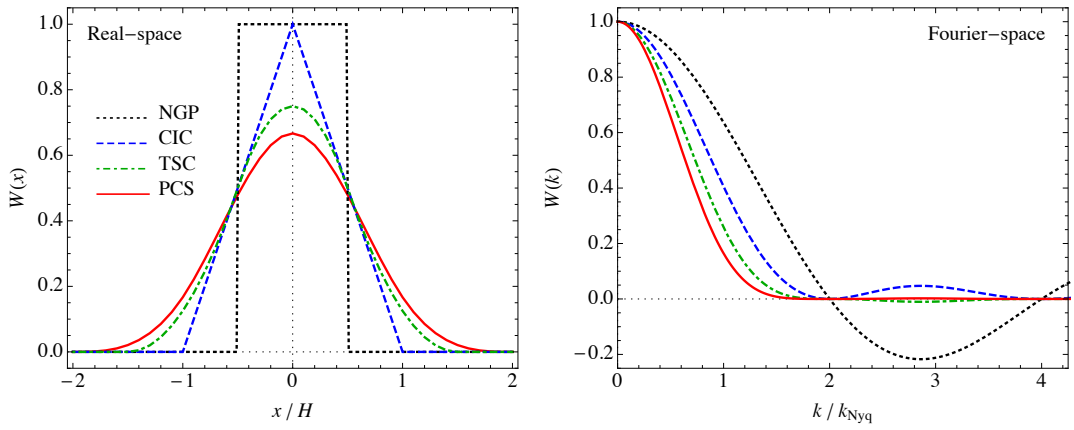

Figure 4.1 – Left: window function of the first four order assignment schemes. Right: same, in Fourier space. Taken from Sefusatti et al. (2016).

considered first before choosing the method of interlacing (Sefusatti et al., 2016) that we describe hereafter.

Let us first recall that:

$$\widetilde{F}^M(\mathbf{k}) = \frac{1}{H^3} \sum_{\mathbf{n}\in\mathbb{Z}^3} \widetilde{F}(\mathbf{k} - 2k_N\mathbf{n}). \tag{4.100}$$

If we shift the mesh by $H/2$ along each axis, the Fourier transform of $\widetilde{F}(\mathbf{r})$ becomes (with $\mathbf{H}$ the vector of components H):

$$\widetilde{F}^{m,1/2}(\mathbf{k}) = \int d^3\mathbf{r}\text{Ш}(\mathbf{r} + \mathbf{H}/2)\widetilde{F}(\mathbf{r})e^{-i\mathbf{r}\cdot\mathbf{k}}. \tag{4.101}$$

A translation of $\text{Ш}(\mathbf{r})$ is a phase shift in Fourier space, so using the convolution theorem:

$$\widetilde{F}^{m,1/2}(\mathbf{k}) = \frac{1}{(2\pi)^3} \int d^3\mathbf{k}'\text{Ш}(\mathbf{k}')\widetilde{F}(\mathbf{k} - \mathbf{k}')e^{-i\mathbf{k}'\cdot\mathbf{H}/2} \tag{4.102}$$

$$= \frac{1}{H^3} \sum_{\mathbf{n}\in\mathbb{Z}^3} \widetilde{F}(\mathbf{k} - 2k_N\mathbf{n})e^{-i\pi(n_x+n_y+n_z)} \tag{4.103}$$

$$= \frac{1}{H^3} \sum_{\mathbf{n}\in\mathbb{Z}^3} (-1)^{n_x+n_y+n_z} W(\mathbf{k} - 2k_N\mathbf{n})F(\mathbf{k} - 2k_N\mathbf{n}). \tag{4.104}$$

Computing the average of $\widetilde{F}^M(\mathbf{k})$ and $\widetilde{F}^{m,1/2}(\mathbf{k})$:

$$\frac{\widetilde{F}^M(\mathbf{k}) + \widetilde{F}^{m,1/2}(\mathbf{k})}{2} = \frac{1}{H^3} \sum_{\mathbf{n}\in\mathbb{Z}^3} \left[1 + (-1)^{n_x+n_y+n_z}\right] W(\mathbf{k} - 2k_N\mathbf{n})F(\mathbf{k} - 2k_N\mathbf{n}). \tag{4.105}$$

The fundamental mode ($n_x + n_y + n_z = 0$) is preserved while every odd $n_x + n_y + n_z$ combination (including the major contribution $|n_x + n_y + n_z| = 1$) is cancelled, therefore reducing the effect of aliasing by a factor of $\simeq 2$. One could go further and average the FFTs computed on meshes shifted by $s/a\mathbf{H}$ ($s \in [0, a-1]$), such that aliasing is reduced by a factor a. However, taking $a = 2$ yields results satisfactory enough for our purpose. Of course, the obtained Fourier transform must sill be divided by $W(\mathbf{k})$. The efficiency of the interlacing method is shown in Figure 4.2, in comparison to the direct summation method, where Fourier transforms are computed exactly. The typical improvement brought by interlacing is of $\simeq 10^2$ and increases with the order of the interpolation kernel. Using the TSC scheme in combination of aliasing ensures relative errors of $\simeq 10^{-4}$ all the way up to the Nyquist frequency.

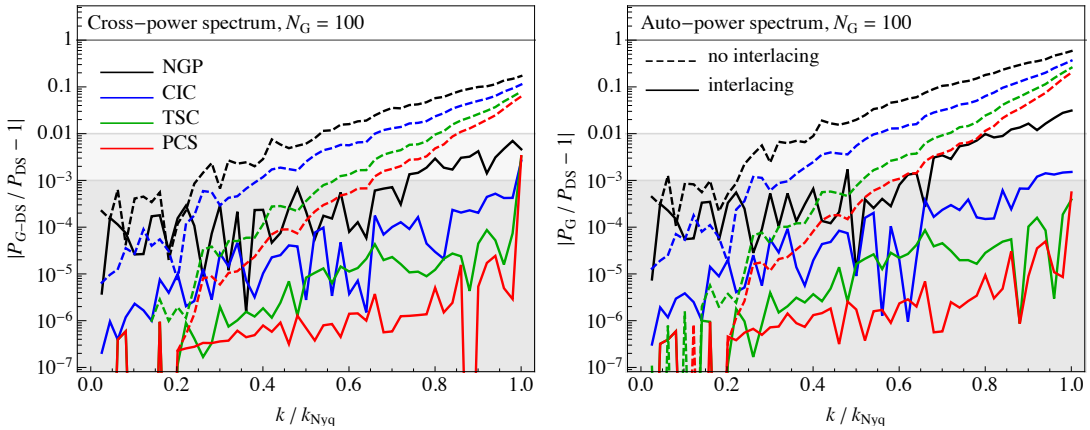

Figure 4.2 – Left: relative difference of the cross power spectrum between the mesh (of size 100) and direct summation method to the auto power spectrum using direct summation, for different interpolation schemes (NGC to PCS), with (dashed lines) and without (continuous lines) interlacing. Right: relative difference of the auto power spectrum calculation with the mesh method divided by the direct summation result. Measurements are performed using the N-body simulation of Colombi et al. (2009). Taken from Sefusatti et al. (2016).

4.2 Geometry effects

In the previous section we built an estimator for the power spectrum of the spatial distribution of galaxies; however, one has only access to galaxy angular positions and redshifts. These coordinates are in fact converted into comoving coordinates assuming a *fiducial* cosmology, whose impact on the power spectrum measurement is discussed in Section 4.2.1. Also, we noticed in the previous section that the estimated power spectrum results from a convolution with the survey window function, which we detail in Section 4.2.2. Finally, the impact of the line of sight definition is discussed in Section 4.2.3 and the sampling of the mesh used to compute the FFTs in Section 4.2.4.

4.2.1 Alcock-Paczynski effect

The galaxy power spectrum is calculated in Cartesian space, while spectroscopic redshift surveys provide a catalogue of angular positions and redshifts. Angular positions and redshifts are therefore converted into comoving distances, assuming a *fiducial* cosmology. This *fiducial* cosmology does not necessarily match the true cosmology of the observed data, yielding distortions known as the Alcock-Paczynski effect (Alcock and Paczynski, 1979). By definition of the radial and angular comoving distances (see Eq. (1.27) and Eq. (1.29)), the comoving length element between two galaxies (at redshift $\simeq z_{\rm eff}$) separated by a redshift dz and an angle $d\theta$ reads:

$$dl_{\parallel} = \left.\frac{d\chi}{dz}\right|_{z=z_{\rm eff}} dz = \frac{dz}{H(z_{\rm eff})} = D_{\rm H}(z_{\rm eff})dz \tag{4.106}$$

$$dl_{\perp} = D_{\rm M}(z_{\rm eff})d\theta, \tag{4.107}$$

where we defined $D_{\rm H}(z) = 1/H(z)$. Denoting by the subscript fid values in the fiducial cosmology, we define the scaling parameters $q_\parallel$, $q_\perp$:

$$q_\parallel = \frac{dl_\parallel}{dl_\parallel^{\rm fid}} = \frac{D_{\rm H}(z_{\rm eff})}{D_{\rm H}^{\rm fid}(z_{\rm eff})} \tag{4.108}$$

$$q_\perp = \frac{dl_\perp}{dl_\perp^{\rm fid}} = \frac{D_{\rm M}(z_{\rm eff})}{D_{\rm M}^{\rm fid}(z_{\rm eff})}. \tag{4.109}$$

Then, the true wavenumbers parallel and perpendicular to the line of sight must be multiplied by $q_\parallel$ and $q_\perp$, respectively, to obtain their values in the fiducial cosmology. Conversely, the true wavenumbers are related to the fiducial ones through $k_\perp = k_\perp^{\rm fid}/q_\perp$ and $k_\parallel = k_\parallel^{\rm fid}/q_\parallel$. With $k = \sqrt{k_\parallel^2 + k_\perp^2}$, $\mu = k_\parallel/k$, in the (k, μ) space (Ballinger et al., 1996):

$$k = \frac{k^{\rm fid}}{q_\perp}\left[1 + \left(\mu^{\rm fid}\right)^2\left(\frac{q_\perp^2}{q_\parallel^2} - 1\right)\right]^{1/2} \tag{4.110}$$

$$\mu = \frac{\mu^{\rm fid} q_\perp}{q_\parallel}\left[1 + \left(\mu^{\rm fid}\right)^2\left(\frac{q_\perp^2}{q_\parallel^2} - 1\right)\right]^{-1/2} \tag{4.111}$$

such that the power spectrum multipoles $P_\ell^{\rm fid}(k^{\rm fid})$ in the *fiducial* space are related to the power spectrum in the true space $P(k, \mu)$ through:

$$P_\ell^{\rm fid}(k^{\rm fid}) = \frac{2\ell + 1}{2 q_\parallel q_\perp^2}\int_{-1}^{1} d\mu^{\rm fid} P(k(k^{\rm fid}, \mu^{\rm fid}), \mu(\mu^{\rm fid}))\mathcal{L}_\ell(\mu^{\rm fid}). \tag{4.112}$$

Eq. (4.112) is the transform to be applied to the model prediction $P(k, \mu)$ to convert it in fiducial coordinates. The resulting $P_\ell^{\rm fid}(k^{\rm fid})$ can be compared to the power spectrum measurements for which the fiducial cosmology fid is used to convert angular positions and redshifts into distances; see e.g. d'Amico et al. (2020); Ivanov et al. (2020).

In this thesis and as in other BOSS and eBOSS analyses (e.g. Beutler et al., 2017), to save computing time the model prediction $P(k, \mu)$ is calculated assuming a fixed template cosmology, often chosen to match the fiducial cosmology. In return, $q_\parallel$ and $q_\perp$ are left free in the cosmological fits. These parameters are mostly constrained through the BAO feature [4] in the power spectrum or 2-point correlation function. The BAO scale is provided by $r_{\rm drag}$, the comoving sound horizon at the redshift at which the baryon-drag optical depth equals unity (see Section 1.3.5). Hence, $q_\parallel$ and $q_\perp$ are effectively sensitive to $\alpha_\parallel$ and $\alpha_\perp$, respectively:

$$\begin{aligned} \alpha_\parallel &= \frac{D_{\rm H}(z_{\rm eff}) r_{\rm drag}^{\rm fid}}{D_{\rm H}^{\rm fid}(z_{\rm eff}) r_{\rm drag}} \\ \alpha_\perp &= \frac{D_{\rm M}(z_{\rm eff}) r_{\rm drag}^{\rm fid}}{D_{\rm M}^{\rm fid}(z_{\rm eff}) r_{\rm drag}} \end{aligned} \tag{4.113}$$

Therefore, varying the scaling parameters as part of the cosmological inference allows constraints to be put on the Hubble parameter and comoving angular distance, divided by $r_{\rm drag}$, at the effective redshift $z_{\rm eff}$ of the spectroscopic sample.

4. While this is fully true for BAO analyses, it is a bit less clear for RSD ones due to the possible influence of the broadband.

4.2.2 Survey geometry

As stated in the previous section, the observed galaxy power spectrum is modulated by the window function. This effect is usually included in the theoretical model prediction. First BOSS clustering analyses performed the convolution of the power spectrum model with the window function in Fourier space (e.g. Beutler et al., 2014), which was quite time-consuming. Instead, Wilson et al. (2017) suggested to Fourier transform the model power spectrum and simply multiply the correlation function with the window function in configuration space — which is indeed much faster. This is the method we develop in the following, but we do not follow the derivation of Wilson et al. (2017). We rather start directly from the expression of the power spectrum estimator, as it makes the derivation more natural and accurate: though this is a quite general statement that the observed power spectrum is convolved with the survey window function, details (such as the normalisation or choice of line of sight) depend on the specific estimator used.

4.2.2.1 Derivation of the window function effect

Let us recall the Yamamoto estimator (4.71) (where we take the limit of an infinitely thin $\mathbf{k}$-shell Ω_k):

$$\hat{P}_\ell(k) = \frac{2\ell+1}{A}\int\frac{d\Omega_k}{4\pi}\int d^3x_1 d^3x_2 e^{i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)}F(\mathbf{x}_1)F(\mathbf{x}_2)\mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}_2) - P_\ell^{\rm noise}(\mathbf{k}) \tag{4.114}$$

with F the FKP field (4.62). Using Eq. (4.65), we find:

$$\left\langle\hat{P}_\ell(k)\right\rangle = \frac{2\ell+1}{A}\int\frac{d\Omega_k}{4\pi}\int d^3x_1 d^3x_2 e^{i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)}W(\mathbf{x}_1)W(\mathbf{x}_2)\xi(\mathbf{x}_1,\mathbf{x}_2)\mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}_2) \tag{4.115}$$

Let us define $\mathbf{s} = \mathbf{x}_1 - \mathbf{x}_2$, and (in the local plane-parallel approximation limit $s \ll x_1, x_2$):

$$\xi(\mathbf{x}_1,\mathbf{x}_2) = \sum_p \xi_p(s)\mathcal{L}_p(\hat{\mathbf{x}}_1\cdot\hat{\mathbf{s}}). \tag{4.116}$$

Substituting into Eq. (4.115):

$$\left\langle\hat{P}_\ell(k)\right\rangle = \frac{2\ell+1}{A}\sum_p\int\frac{d\Omega_k}{4\pi}\int d^3x\int d^3s e^{-i\mathbf{k}\cdot\mathbf{s}}W(\mathbf{x})W(\mathbf{x}-\mathbf{s})\xi_p(s)\mathcal{L}_p(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}})\mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}) \tag{4.117}$$

Using the Rayleigh plane wave expansion:

$$e^{-i\mathbf{k}\cdot\mathbf{s}} = \sum_{q=0}^{+\infty}(-i)^q(2q+1)j_q(ks)\mathcal{L}_q(\hat{\mathbf{k}}\cdot\hat{\mathbf{s}}) \tag{4.118}$$

and

$$\int\frac{d\Omega_k}{4\pi}\mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}})\mathcal{L}_q(\hat{\mathbf{k}}\cdot\hat{\mathbf{s}}) = \frac{\delta_{\ell q}}{2\ell+1}\mathcal{L}_\ell(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}}), \tag{4.119}$$

we find:

$$\left\langle\hat{P}_\ell(k)\right\rangle = (-i)^\ell\frac{2\ell+1}{A}\sum_p\int d^3s j_\ell(ks)\xi_p(s)\int d^3x W(\mathbf{x})W(\mathbf{x}-\mathbf{s})\mathcal{L}_p(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}})\mathcal{L}_\ell(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}}). \tag{4.120}$$

Let us define the coefficient $A_{\ell p}^q$ by:

$$\mathcal{L}_\ell(\mu)\mathcal{L}_p(\mu) = \sum_{q=0}^{\ell+p}A_{\ell p}^q\mathcal{L}_q(\mu) \tag{4.121}$$

the window function multipoles:

$$\mathcal{W}_\ell^{\delta,\delta}(s) = \frac{2\ell+1}{4\pi A}\int d\Omega_s \int d^3x W(\mathbf{x})W(\mathbf{x}-\mathbf{s})\mathcal{L}_\ell(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}}) \tag{4.122}$$

and the *window-convolved* correlation function (this appellation is not strictly correct as the following formula is a product and not a convolution as in Fourier space):

$$\xi_\ell^c(s) = \sum_{p,q} A_{\ell p}^q \frac{2\ell+1}{2q+1}\xi_p(s)\mathcal{W}_q^{\delta,\delta}(s). \tag{4.123}$$

We eventually find that the expected value of the power spectrum estimator reads:

$$\left\langle \hat{P}_\ell(k)\right\rangle = 4\pi\,(-i)^\ell \int s^2 ds j_\ell(ks)\xi_\ell^c(s) \tag{4.124}$$

which is the formula we use to account for the window function effect in the model. The correlation function multipoles entering Eq. (4.123) can be obtained by a simple inverse Hankel transform of the theory power spectrum $P_\ell(k)$:

$$\xi_\ell(s) = \frac{i^\ell}{2\pi^2}\int k^2 dk j_\ell(ks)P_\ell(k). \tag{4.125}$$

4.2.2.2 Practical calculation

Hankel transforms Eq. (4.124) and Eq. (4.125) can be performed with a $\mathcal{O}(N\log N)$ complexity, with N the number of bins in the integral, using algorithms such as `FFTLog` (Hamilton, 2000) which we employ in this manuscript. This is opposed to the previous technique performing a convolution in Fourier space, of complexity $\mathcal{O}(N^2)$.

In practice, the infinite sum of Eq. (4.123) should be truncated. As in Beutler et al. (2017) we only consider correlation function multipoles up to $\ell = 4$, thereby limiting window function multipoles to $\ell = 8$. In this case:

$$\xi_0^c(s) = \xi_0(s)\mathcal{W}_0^{\delta,\delta}(s) + \frac{1}{5}\xi_2(s)\mathcal{W}_2^{\delta,\delta}(s) + \frac{1}{9}\xi_4(s)\mathcal{W}_4^{\delta,\delta}(s) \tag{4.126}$$

$$\begin{aligned}\xi_2^c(s) = {}& \xi_0(s)\mathcal{W}_2^{\delta,\delta}(s) + \xi_2(s)\left[\mathcal{W}_0^{\delta,\delta}(s) + \frac{2}{7}\mathcal{W}_2^{\delta,\delta}(s) + \frac{2}{7}\mathcal{W}_4^{\delta,\delta}(s)\right] \\ & + \xi_4(s)\left[\frac{2}{7}\mathcal{W}_2^{\delta,\delta}(s) + \frac{100}{693}\mathcal{W}_4^{\delta,\delta}(s) + \frac{25}{143}\mathcal{W}_6^{\delta,\delta}(s)\right]\end{aligned} \tag{4.127}$$

$$\begin{aligned}\xi_4^c(s) = {}& \xi_0(s)\mathcal{W}_4^{\delta,\delta}(s) + \xi_2(s)\left[\frac{18}{35}\mathcal{W}_2^{\delta,\delta}(s) + \frac{20}{77}\mathcal{W}_4^{\delta,\delta}(s) + \frac{45}{143}\mathcal{W}_6^{\delta,\delta}(s)\right] \\ & + \xi_4(s)\left[\mathcal{W}_0^{\delta,\delta}(s) + \frac{20}{77}\mathcal{W}_2^{\delta,\delta}(s) + \frac{162}{1001}\mathcal{W}_4^{\delta,\delta}(s) + \frac{20}{143}\mathcal{W}_6^{\delta,\delta}(s) + \frac{490}{2431}\mathcal{W}_8^{\delta,\delta}(s)\right].\end{aligned} \tag{4.128}$$

We checked that adding $\xi_6(s)$ has a completely negligible impact on the model prediction in practice.

Usually the survey selection function is not analytic and is instead sampled with a synthetic catalogue, as mentioned in Section 4.1.6. Hence, the 2-point window function $\mathcal{W}_\ell^{\delta,\delta}$ is generally estimated by computing the correlation function of this synthetic catalogue, as will be detailed in Section 4.3.5.3.

4.2.2.3 Normalisation

In the above derivation we have made clear that window functions $\mathcal{W}_\ell^{\delta,\delta}(s)$ (see Eq. (4.122)) should be normalised by A, the quantity that scales the power spectrum estimation. Instead, previous BOSS and eBOSS analyses followed the suggestion of Wilson et al. (2017) who proposed to normalise the 2-point window function in the limit $s \to 0$, so as to recover the true (unconvolved) power spectrum at small scales.However, the low s limit of $\mathcal{W}_\ell^{\delta,\delta}(s)$ may be ill-defined in practice. Indeed, usually the survey selection function is described by a synthetic catalogue, which should be very finely sampled to correctly estimate the window function at small separation. In addition, with a synthetic catalogue of density n_s, following the notations in Section 4.1.6, the A term reads:

$$A = \alpha^2 \int d^3r n_s^2(\mathbf{r}) = \alpha^2 \sum_{i=1}^{N_s} w_{s,i} n_{s,i}. \tag{4.129}$$

The estimation of the density $n_s(\mathbf{r})$ in Eq. (4.129) is in practice non-trivial when accounting for various survey selection effects. For example, $n_s(\mathbf{r})$ is commonly taken to be the redshift density $n(z)$, computed by binning (weighted) data in redshift slices (as we will see in Section 5.3.1.7), while it can also be a function of the angular position on the sky.

For these reasons, normalising $\mathcal{W}_\ell^{\delta,\delta}(s)$ to be 1 in the $s \to 0$ limit, independently of the normalisation A used in the power spectrum estimate would have led to errors of a few percent in the cosmological parameter inference in eBOSS. In contrast, using term (4.129) in the normalisation of window functions $\mathcal{W}_\ell^{\delta,\delta}(s)$ and $\mathcal{W}_{\ell p}^{i,j}(s,\Delta)$, A terms divide both the power spectrum measurements and model and compensate. Therefore, the estimation of A does not impact the estimation of cosmological parameters.

4.2.3 Wide-angle effects

In Eq. (4.116) we have used the plane-parallel approximation $s \ll x_1, x_2$. This approximation breaks down for wide-angle spectroscopic surveys, for which the maximum separation between a pair of galaxies can be of the order of their distance to the observer, yielding distortions in the power spectrum at large scales (with respect to the local plane-parallel limit).

Since we search for distortions on large scales, we can assume linear perturbation theory. Then, the galaxy density contrast in redshift space $\delta_s(\mathbf{r})$ is related to that in real space $\delta(\mathbf{r})$ through (Szalay et al., 1998; Castorina and White, 2018):

$$\delta_s(\mathbf{r}) = b\left(\delta(\mathbf{r}) + \beta \frac{\partial v_r(\mathbf{r})}{\partial r}\right) + f\alpha(r) v_r(\mathbf{r}) \tag{4.130}$$

where $v_r = -\mathbf{u}\cdot\hat{\mathbf{r}}/(f\mathcal{H})$ and $\beta = f/b$. Note that there is an additional term with respect to Eq. (2.163): $f\alpha(r)v_r(\mathbf{r})$, where $\alpha(r) = (2 + \partial \ln W(r)/\partial \ln r)/r$, with W the survey selection function. This term arises because mass conservation is to be applied to the density $W(1+\delta)$, while in Eq. (2.160) we applied mass conservation to $1+\delta$, assuming W to be constant. In this manuscript we drop this term, though its impact should be remembered for future studies of galaxy clustering on very large scales, e.g. focusing on primordial non-Gaussianity.

The following calculation requires to specify the triangle formed by a pair of galaxies and the observer. This triangle is fully described by 3 scalar quantities, which can be the separation s of the pair of galaxies, and two angles, θ and γ, as depicted in Figure 4.3.

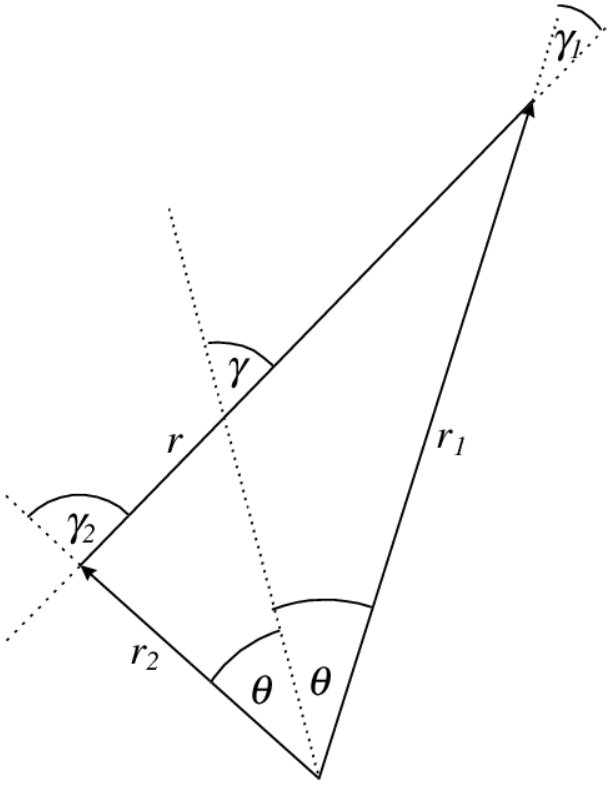

Figure 4.3 – Configuration of the triangle formed by the observer and a pair of galaxies, which we specify with the separation ($s = r$) of the two galaxies, and the two angles, θ and γ. Taken from Szalay et al. (1998).

In their pioneering work, Szalay et al. (1998) showed that the correlation function of δ_s reads:

$$\xi(s,\theta,\gamma) = c_0(\theta,\gamma)\Xi_0(s) + c_2(\theta,\gamma)\Xi_2(s) + c_4(\theta,\gamma)\Xi_4(s) \tag{4.131}$$

where:

$$\Xi_\ell(s) = \frac{b^2 i^\ell}{2\pi^2}\int k^2 dk j_\ell(ks) P_m^{\rm lin}(k) \tag{4.132}$$

with $P_m^{\rm lin}(k)$ the linear matter power spectrum and:

$$c_0(\theta,\gamma) = 1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2 - \frac{4}{15}\beta^2\cos^2\theta\sin^2\theta \tag{4.133}$$

$$c_2(\theta,\gamma) = \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right)\cos 2\theta\mathcal{L}_2(\cos\gamma) + \frac{2}{3}\left(\beta - \frac{1}{7}\beta^2 + \frac{4}{7}\beta^2\sin^2\theta\right)\sin^2\theta \tag{4.134}$$

$$c_4(\theta,\gamma) = \frac{8}{35}\beta^2\mathcal{L}_4(\cos\gamma) - \frac{4}{21}\beta^2\sin^2\theta\mathcal{L}_2(\cos\gamma) - \frac{1}{5}\beta^2\left(\frac{4}{21} - \frac{3}{7}\sin^2\theta\right)\sin^2\theta. \tag{4.135}$$

However, expressing ξ in redshift space as a function of three variables (separation and two angles) instead of two (separation and one angle) as in Eq. (4.116) significantly slows down practical computation. Hence, Reimberg et al. (2016) and Castorina and White (2018) proposed to develop the θ dependence (which is expected to be small in practical applications) in powers of the ratio of the pair separation to the distance. Namely, for a separation $\mathbf{s} = \mathbf{x}_1 - \mathbf{x}_2$ and a generic line of sight $\mathbf{d}$:

$$\xi(\mathbf{x}_1,\mathbf{x}_2) = \sum_{p,n}\left(\frac{s}{d}\right)^n \xi_p^{(n)}(s)\mathcal{L}_p(\hat{\mathbf{d}}\cdot\hat{\mathbf{s}}) \tag{4.136}$$

of which Eq. (4.116) is the limit as $s/d \to 0$. In this framework, Eq. (4.123) simply rewrites:

$$\xi_\ell^c(s) = \sum_{p,q,n} A_{\ell p}^q \frac{2\ell+1}{2q+1} s^n \xi_p^{(n)}(s) \mathcal{W}_q^{\delta,\delta,(n)}(s) \tag{4.137}$$

where:

$$\mathcal{W}_\ell^{\delta,\delta,(n)}(s) = \frac{2\ell+1}{4\pi A} \int d\Omega_s \int d^3x d^{-n} W(\mathbf{x}) W(\mathbf{x}-\mathbf{s}) \mathcal{L}_\ell(\hat{\mathbf{d}}\cdot\hat{\mathbf{s}}) \tag{4.138}$$

In the case of the bisector line of sight, as shown with a dashed line in Figure 4.3, one can show that (Castorina and White, 2018):

$$\cos\theta = 1 - \frac{(1-\mu_b^2)\,\epsilon_b^2}{8} + o(\epsilon_b^2) \tag{4.139}$$

where $\epsilon_b = s/d$ and $\mu_b = \cos\gamma$. Substituting this expression into Eq. (4.131) which we project onto the Legendre multipoles yields:

$$\xi_0(s,d) = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right)\Xi_0(s) + \left(\frac{s}{d}\right)^2 \left[-\frac{4}{45}\beta^2\Xi_0(s) + \left(\frac{1}{5}\beta + \frac{1}{45}\beta^2\right)\Xi_2(s)\right] \tag{4.140}$$

$$\xi_2(s,d) = \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right)\Xi_2(s) + \left(\frac{s}{d}\right)^2 \left[\frac{4}{45}\beta^2\Xi_0(s) + \left(-\frac{3}{7}\beta - \frac{53}{441}\beta^2\right)\Xi_2(s) - \frac{4}{245}\beta^2\Xi_4(s)\right] \tag{4.141}$$

$$\xi_4(s,d) = \frac{8}{35}\beta^2\Xi_4(s) + \left(\frac{s}{d}\right)^2 \left[\left(\frac{8}{35}\beta + \frac{24}{245}\beta^2\right)\Xi_2(s) + \frac{4}{245}\beta^2\Xi_4(s)\right], \tag{4.142}$$

all other multipoles being zero. Of course, the zeroth order terms correspond to the linear perturbation theory prediction for the multipoles of the correlation function in the local plane-parallel approximation. The identification of $\xi_p^{(n)}(s)$ of Eq. (4.136) in the above formulae is straightforward. Note that first corrections to the local plane-parallel approximation are second order and no odd multipoles are generated (as expected by symmetry of the line of sight definition with respect to the pair of galaxies). We could do the same exercise with the midpoint line of sight where $\mathbf{d} = \mathbf{x}_m = (\mathbf{r}_1 + \mathbf{r}_2)/2$ in Figure 4.3. In this case, ϵ_m and μ_m are related to those of the bisector line of sight through (Castorina and White, 2018):

$$\epsilon_b = \epsilon_m \left[1 + \frac{\mu_m^2 \epsilon_m^2}{4} + \cdots\right] \tag{4.143}$$

$$\mu_b = \mu_m \left[1 + \frac{\epsilon_m^2}{4}\left(1-\mu_m^2\right) + \cdots\right] \tag{4.144}$$

In this case again, first corrections to the local plane-parallel approximation are second order and no odd multipoles are generated. In the case of the endpoint line of sight ($\mathbf{d} = \mathbf{r}_1$ in Figure 4.3):

$$\epsilon_b = \epsilon_e \left[1 + \frac{\mu_e \epsilon_e}{2} + \cdots\right] \tag{4.145}$$

$$\mu_b = \mu_e - \frac{r_1}{2}\left(1-\mu_e^2\right) + \cdots \tag{4.146}$$

which yields odd multipoles:

$$\xi_0(s,d) = \xi_0^{(0)}(s) + \left(\frac{s}{d}\right)^2 \left[\frac{4}{15}\beta + \frac{4}{35}\beta^2\right] \Xi_2(s) \tag{4.147}$$

$$\xi_1(s,d) = \left(\frac{s}{d}\right) \left[-\frac{4}{5}\beta - \frac{12}{35}\beta^2\right] \Xi_2(s) \tag{4.148}$$

$$\xi_2(s,d) = \xi_2^{(0)}(s) + \left(\frac{s}{d}\right)^2 \left[\left(-\frac{8}{21}\beta - \frac{8}{49}\beta^2\right) \Xi_2(s) + \frac{8}{49}\beta^2 \Xi_4(s)\right] \tag{4.149}$$

$$\xi_3(s,d) = \left(\frac{s}{d}\right) \left[\left(\frac{4}{5}\beta + \frac{12}{35}\beta^2\right) \Xi_2(s) - \frac{16}{63}\beta^2 \Xi_4(s)\right] \tag{4.150}$$

$$\xi_4(s,d) = \xi_4^{(0)}(s) + \left(\frac{s}{d}\right)^2 \left[\left(\frac{4}{35}\beta + \frac{12}{245}\beta^2\right) \Xi_2(s) + \frac{144}{539}\beta^2 \Xi_4(s)\right] \tag{4.151}$$

Corrections to odd multipoles are first order in s/d while corrections to even multipoles remain at second order in s/d.

As demonstrated in Beutler et al. (2019), the above odd multipoles of the correlation function couple to the odd multipoles of the window function, leading to first order corrections to the even multipoles of the *window-convolved* correlation function (4.137).

4.2.4 Irregular μ sampling

As mentioned in Section 4.1.6, power spectrum multipoles are calculated on a discrete **k**-space mesh, making the angular modes distribution irregular at large scales (low k). We account for this effect in the model using the technique employed in Beutler et al. (2017) which weights each (k,μ) mode according to its sampling rate $N(k,\mu)$ in the **k**-space mesh $\mathbf{M}$:

$$P^M(k,\mu) = \frac{2N(k,\mu)}{\int_{-1}^{1} d\mu N(k,\mu)} P(k,\mu) \tag{4.152}$$

with the piecewise function:

$$N(k,\mu) = \left|\{k_i(k) \leq m < k_{i+1}(k)\,\&\,\mu_i(\mu) \leq \hat{\mathbf{m}}\cdot\hat{\eta} < \mu_{i+1}(\mu) | \mathbf{m} \in \mathbf{M}\}\right|, \tag{4.153}$$

where $k_i(k)$ and $k_{i+1}(k)$ are the lower and upper bounds of a k-bin containing k (similarly for μ), $\mathbf{m}$ is a vector from the mesh $\mathbf{M}$ and $\hat{\eta}$ is the considered line of sight. In the case of a varying line of sight as in Eq. (4.71), we average $N(k,\mu)$ over the lines of sight of all the objects in the synthetic catalogue sampling the survey selection function. Though this correction should be applied after the convolution by the window function, as in Beutler et al. (2017), for the sake of computing time we include it when integrating the galaxy power spectrum over the Legendre polynomials in Eq. (4.112).

4.2.5 Complete model for the measured power spectrum

Let us recap the full theoretical model, including all geometry corrections.

As detailed in Section 2.4.3, the redshift space power spectrum is given by Eq. (2.198). However, the galaxy power spectrum is measured assuming a fiducial cosmology which does not necessarily match the true one, as explained in Section 4.2.1. The power spectrum prediction in the fiducial space is hence given by the AP transform (4.112), where we include the effect of the irregular μ sampling as described in Section 4.2.4. Since the power spectra in the model

prediction are calculated once, within the fiducial cosmology, we leave the scaling parameters $q_\parallel$ and $q_\perp$ free. These parameters are sensitive to $\alpha_\parallel$ and $\alpha_\perp$, as given by Eq. (4.113). Finally, the measured power spectrum is modulated by the survey selection function, which we take into account in the model using Eq. (4.124). In this equation, $\xi_\ell(s)$ is the Hankel transform of the theoretical power spectrum, after AP transform and inclusion of the irregular μ sampling. We also implemented wide-angle effects, though we did not take them into account in practice, as we found their impact to be negligible for our analyses.

Fitted cosmological parameters are the logarithmic growth rate of structure f and the scaling parameters $\alpha_\parallel$ and $\alpha_\perp$. As explained in Section 2.4.3, we consider 4 nuisance parameters: the linear and second order biases b_1 and b_2, the velocity dispersion σ_v and $A_g = N_g/P_0^{\rm noise}$, with N_g the constant galaxy stochastic term (see Eq. (2.150)) and $P_0^{\rm noise}$ the measured Poisson shot noise (see Eq. (4.67)). We quote the combinations $f\sigma_8$, $b_1\sigma_8$ and $b_2\sigma_8$, as mentioned in Section 2.4.3.

We implemented this complete model in a `Python` package `pyspectrum`[5].

The model is compared to the data using a Gaussian likelihood (except otherwise stated). Minimisations are performed using the algorithm `Minuit` (James and Roos, 1975; iminuit team, iminuit team), taking large variation intervals for all parameters. We systematically check that the fitted parameters do not reach the input boundaries. In Chapter 5 we also run Monte Carlo Markov Chains (MCMC) to sample the posterior distribution with the package `emcee` (Foreman-Mackey et al., 2013). We implemented these options in a coherent framework, `montelss`[6]. `pyspectrum` and `montelss` were used in eBOSS ELG (de Mattia et al., 2020) and QSO (Neveux et al., 2020) Fourier space analyses.

4.3 Integral constraints

As detailed in the previous section, spectroscopic surveys are not exhaustive and are characterised by a selection function $W(\mathbf{r})$, giving the expected density of observed galaxies at any redshift space position $\mathbf{r}$ in the absence of clustering. In case one has a full knowledge of this selection function, density fluctuations can be fairly estimated by the difference of the observed density of galaxies to $W(\mathbf{r})$. In particular, density fluctuations averaged on the whole survey footprint can be non-zero, due to large scale clustering modes.

However, the true survey selection function is difficult to determine in practice. Its norm, i.e. the expected mean density of galaxies, may be unknown. It is commonly taken to be the actually observed mean data density. Thus, the integral of the inferred density fluctuations over the whole survey footprint is set to zero, leading to a so-called integral constraint (IC) (Peacock and Nicholson, 1991; Wilson et al., 2017), referred to as global IC in the following. Similarly, the radial survey selection function may be estimated from the data itself, as in the clustering analyses of the BOSS (e.g. Gil-Marín et al., 2016; Beutler et al., 2017; Alam et al., 2017) and eBOSS (e.g. Ata et al., 2018; Zarrouk et al., 2018) surveys. We propose to treat this effect as another integral constraint, dubbed radial IC, to be included in the model power spectrum.

This section relies on our original work de Mattia and Ruhlmann-Kleider (2019).

5. at https://github.com/adematti/montelss/tree/master/pyspectrum
6. at https://github.com/adematti/montelss

4.3.1 Impact of the global integral constraint on the observed density fluctuations

The power spectrum estimator is based on the FKP field (4.62):

$$F(\mathbf{r}) = n_g(\mathbf{r}) - \alpha n_s(\mathbf{r}) \tag{4.154}$$

where we recall that $n_g(\mathbf{r})$ and $n_s(\mathbf{r})$ denote respectively the density of observed and random galaxies, including weighting, e.g. corrections for systematics effects (see Chapter 5) or a redshift weighting scheme, such as FKP weights. The observed galaxy density is $n_g(\mathbf{r}) = W(\mathbf{r})\{1+\delta(\mathbf{r})\}$, with $W(\mathbf{r})$ the survey selection function and $\delta(\mathbf{r})$ the density contrast. We recall that $W(\mathbf{r})$ describes the expected density of observed, possibly weighted, galaxies in the absence of clustering. We assume the shape of the survey selection function is known, and sampled by the (weighted) synthetic catalogue: $n_s(\mathbf{r}) \propto W(\mathbf{r})$.

In Section 4.1.6 we defined α as the expected value (see Eq. (4.61)):

$$\alpha = \frac{\left\langle \int d^3x n_g(\mathbf{x}) \right\rangle}{\int d^3x n_s(\mathbf{x})}. \tag{4.155}$$

However, the total number of galaxies $\left\langle \int d^3x n_g(\mathbf{x}) \right\rangle$ is very difficult to predict as it depends on the luminosity function, the target selection and the full instrumental setup. Therefore, in practice, the scaling α is taken to be the volume average:

$$\alpha = \frac{\int d^3x n_g(\mathbf{x})}{\int d^3x n_s(\mathbf{x})} = \frac{\sum_{i=1}^{N_g} w_{g,i}}{\sum_{i=1}^{N_s} w_{s,i}}. \tag{4.156}$$

with N_g, N_s and $w_{g,i}$, $w_{s,i}$ the number and weights of observed and random galaxies, respectively. Then, the observed window-convolved, integral-constraint-corrected (subscript cic) density fluctuations are (Peacock and Nicholson, 1991; Beutler et al., 2014):

$$\begin{aligned}
\delta^{\rm cic}(\mathbf{r}) &= n_g(\mathbf{r}) - \alpha n_s(\mathbf{r}) \\
&= W(\mathbf{r})\{1+\delta(\mathbf{r})\} - W(\mathbf{r})\frac{\int d^3x W(\mathbf{x})\{1+\delta(\mathbf{x})\}}{\int d^3x W(\mathbf{x})} \\
&= W(\mathbf{r})\left\{\delta(\mathbf{r}) - \int d^3x W_{\rm glo}(\mathbf{x})\delta(\mathbf{x})\right\},
\end{aligned} \tag{4.157}$$

with $W_{\rm glo}(\mathbf{r}) = \frac{W(\mathbf{r})}{\int d^3x W(\mathbf{x})}$.

We find two terms, $\delta^{\rm c}(\mathbf{r}) = W(\mathbf{r})\delta(\mathbf{r})$ corresponding to the density contrast weighted by the selection function, and the integral constraint term $\int d^3r W_{\rm glo}(\mathbf{r})\delta(\mathbf{r})$. The normalisation of $W_{\rm glo}(\mathbf{r})$ ensures that $\int d^3r \delta_{\rm cic}(\mathbf{r}) = 0$ over the entire footprint: modes larger than the survey size are suppressed.

4.3.2 Extension to the radial integral constraint

The true radial selection function of a spectroscopic survey is often a complex function of the luminosity function, sky lines, spectrograph efficiency and redshift determination algorithm (Blake et al., 2010). As an example, in BOSS, the radial distribution of the synthetic catalogue is directly inferred from the data. Various techniques exist: random redshifts can be picked from the whole data redshift distribution (the so-called *shuffled* scheme, see Samushia

et al. 2012; Ross et al. 2012; Reid et al. 2016), or, assuming the true radial selection function should be somewhat smooth, it can be fitted by a spline from which random redshifts are drawn (Samushia et al., 2012; Ross et al., 2012). A third possibility (*binned* scheme) is to weight an arbitrary initial random redshift distribution by $\alpha(r)$ (with $r = |\mathbf{r}|$) to match the data radial density in redshift or comoving distance bins of size δr, below which variations of $W(r)$ are neglected. Here, $\alpha(r)$ is given by:

$$\alpha(r) = \frac{\int d^3x n_g(\mathbf{x})\epsilon_{\rm rad}(r,x)}{\int d^3x n_s(\mathbf{x})\epsilon_{\rm rad}(r,x)}, \tag{4.158}$$

where $\epsilon_{\rm rad}(r,x)$ is 1 if r and x belong to the same radial bin, 0 otherwise. In the limit $\delta r \to 0$ could take $\epsilon_{\rm rad}(r,x) = \delta_D^{(1)}(r-x)$. By construction, the radially-normalised random density and true selection function match:

$$\frac{n_s(\mathbf{r})}{\int d^3x n_s(\mathbf{x})\epsilon_{\rm rad}(r,x)} = \frac{W(\mathbf{r})}{\int d^3x W(\mathbf{x})\epsilon_{\rm rad}(r,x)}. \tag{4.159}$$

Then, the following density fluctuations are observed:

$$\begin{aligned}\delta^{\rm cic}(\mathbf{r}) &= n_g(\mathbf{r}) - \alpha(r) n_s(\mathbf{r}) \\ &= W(\mathbf{r})\left\{1+\delta(\mathbf{r})\right\} - W(\mathbf{r})\frac{\int d^3x W(\mathbf{x})\left\{1+\delta(\mathbf{x})\right\}\epsilon_{\rm rad}(r,x)}{\int d^3x W(\mathbf{x})\epsilon_{\rm rad}(r,x)} \\ &= W(\mathbf{r})\left\{\delta(\mathbf{r}) - \int d^3x W_{\rm rad}(\mathbf{x})\delta(\mathbf{x})\epsilon_{\rm rad}(r,x)\right\},\end{aligned} \tag{4.160}$$

with $W_{\rm rad}(\mathbf{r}) = \frac{W(\mathbf{r})}{\int d^3x W(\mathbf{x})\epsilon_{\rm rad}(r,x)}$. Note that the global integral constraint is also automatically imposed.

Eq. (4.157) for the global IC and Eq. (4.160) for the radial IC can be recast in the following form:

$$\delta^{\rm cic}(\mathbf{r}) = W(\mathbf{r})\left\{\delta(\mathbf{r}) - \int d^3x W_{\rm ic}(\mathbf{x})\delta(\mathbf{x})\epsilon_{\rm ic}(\mathbf{r},\mathbf{x})\right\}, \tag{4.161}$$

with $\epsilon_{\rm ic}(\mathbf{r},\mathbf{x})$ some generic window function ($\epsilon_{\rm glo}(\mathbf{r},\mathbf{x}) = 1$ for the global integral constraint, $\epsilon_{\rm rad}(r,x)$ for the radial one), and:

$$W_{\rm ic}(\mathbf{r}) = \frac{W(\mathbf{r})}{\int d^3x W(\mathbf{x})\epsilon_{\rm ic}(\mathbf{r},\mathbf{x})}. \tag{4.162}$$

The general expression for the observed density fluctuations (4.161) will be used in the following section to calculate the effect of integral constraints on 2-point statistics.

4.3.3 Impact on the 2-point statistics

As a warm-up let us first compute the correlation function $\xi^{\rm cic}(\mathbf{s})$ of the observed density fluctuations (4.161). As $W(\mathbf{r})$ is uncorrelated with $\delta(\mathbf{r})$, $\xi^{\rm cic}(\mathbf{s})$ can be expressed as (dropping

the normalisation factor A (4.68)):

$$\xi^{\rm cic}(\mathbf{s}) = \int d^3x W(\mathbf{x})W(\mathbf{x}-\mathbf{s})\xi(\mathbf{s}) \tag{4.163}$$

$$- \int d^3\Delta\xi(\mathbf{\Delta}) \int d^3x W(\mathbf{x})W(\mathbf{x}-\mathbf{s})W_{\rm ic}(\mathbf{x}-\mathbf{\Delta})\epsilon_{\rm ic}(\mathbf{x}-\mathbf{s},\mathbf{x}-\mathbf{\Delta}) \tag{4.164}$$

$$- \int d^3\Delta\xi(\mathbf{\Delta}) \int d^3x W(\mathbf{x})W(\mathbf{x}+\mathbf{s})W_{\rm ic}(\mathbf{x}-\mathbf{\Delta})\epsilon_{\rm ic}(\mathbf{x}+\mathbf{s},\mathbf{x}-\mathbf{\Delta}) \tag{4.165}$$

$$+ \int d^3\Delta\xi(\mathbf{\Delta}) \int d^3x W(\mathbf{x})W(\mathbf{x}-\mathbf{s})$$
$$\int d^3y W_{\rm ic}(\mathbf{y})W_{\rm ic}(\mathbf{y}-\mathbf{\Delta})\epsilon_{\rm ic}(\mathbf{x},\mathbf{y})\epsilon_{\rm ic}(\mathbf{x}-\mathbf{s},\mathbf{y}-\mathbf{\Delta}). \tag{4.166}$$

Term (4.163) is the true correlation function $\xi(\mathbf{s})$ multiplied by the window function, which we described in Section 4.2.2, specifically Eq. (4.123). Cross-terms (4.164) and (4.165) account for the correlation between the density field and the integral constraint term of Eq. (4.161). Finally, term (4.166) accounts for the auto-correlation of the integral term in Eq. (4.161).

To go further one has to specify the line of sight definition. Therefore, let us directly work with the power spectrum estimator. As shown in the previous section, the first term (4.163) yields Eq. (4.124) with the *window-convolved* correlation function as given by Eq. (4.137) when using the endpoint line of sight. The first cross-term Eq. (4.164) reads (in Fourier space and with the endpoint line of sight $\hat{\mathbf{r}}_1$):

$$IC_\ell^{\delta,{\rm ic}}(k) = \frac{2\ell+1}{4\pi A}\int d\Omega_k \int d^3r_1 \int d^3r_2 \int d^3r_3 e^{-i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} \langle\delta(\mathbf{r}_1)\delta(\mathbf{r}_3)\rangle$$
$$W(\mathbf{r}_1)W(\mathbf{r}_2)W_{\rm ic}(\mathbf{r}_3)\mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}_1)\epsilon_{\rm ic}(\mathbf{r}_2,\mathbf{r}_3) \tag{4.167}$$

where the normalisation factor A is reintroduced. Defining $\mathbf{\Delta} = \mathbf{r}_1 - \mathbf{r}_3$ and $\mathbf{s} = \mathbf{r}_1 - \mathbf{r}_2$, we can write (see Eq. (4.136)):

$$\langle\delta(\mathbf{r}_1)\delta(\mathbf{r}_3)\rangle = \sum_{n,p}\left(\frac{\Delta}{r_1}\right)^n \xi_p^{{\rm ep},(n)}(\Delta)\mathcal{L}_p(\hat{\mathbf{r}}_1\cdot\hat{\mathbf{\Delta}}) \tag{4.168}$$

where we recall the sum over n corresponds to wide-angle corrections. Using the Rayleigh plane wave expansion Eq. (4.118) and Eq. (4.119) we find:

$$IC_\ell^{\delta,{\rm ic}}(k) = \frac{2\ell+1}{A}(-i)^\ell \int d^3s j_\ell(ks) \sum_{p,n}\int d^3\Delta\Delta^n\xi_p^{{\rm ep},(n)}(\Delta)$$
$$\int d^3r_1 r_1^{-n} W(\mathbf{r}_1)W(\mathbf{r}_1-\mathbf{s})W_{\rm ic}(\mathbf{r}_1-\mathbf{\Delta})\mathcal{L}_\ell(\hat{\mathbf{r}}_1\cdot\hat{\mathbf{s}})\mathcal{L}_p(\hat{\mathbf{r}}_1\cdot\hat{\mathbf{\Delta}})\epsilon_{\rm ic}(\mathbf{r}_1-\mathbf{s},\mathbf{r}_1-\mathbf{\Delta}). \tag{4.169}$$

$IC_\ell^{\delta,{\rm ic}}(k)$ is simply equal to the Hankel transform:

$$IC_\ell^{i,j}(k) = 4\pi(-i)^\ell \int s^2 ds j_\ell(ks)\int\Delta^2 d\Delta \sum_{p,n}\frac{4\pi}{2p+1}\Delta^n\xi_p^{{\rm ep},(n)}(\Delta)\mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta) \tag{4.170}$$

with $(i,j) = (\delta,{\rm ic})$, where we used:

$$\mathcal{W}_{\ell p}^{\delta,{\rm ic},(n)}(s,\Delta) = \frac{(2\ell+1)(2p+1)}{(4\pi)^2 A}\int d\Omega_s \int d\Omega_\Delta$$
$$\int d^3x x^{-n} W(\mathbf{x})W(\mathbf{x}-\mathbf{s})W_{\rm ic}(\mathbf{x}-\mathbf{\Delta})\mathcal{L}_\ell(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}})\mathcal{L}_p(\hat{\mathbf{x}}\cdot\hat{\mathbf{\Delta}})\epsilon_{\rm ic}(\mathbf{x}-\mathbf{s},\mathbf{x}-\mathbf{\Delta}). \tag{4.171}$$

By symmetry, the even multipoles of the second cross-term (ic,δ) would be equal to those of the first one in the case of the midpoint line of sight. This is not the case with the endpoint line of sight. Indeed:

$$IC_\ell^{\mathrm{ic},\delta}(k) = \frac{2\ell+1}{4\pi A}\int d\Omega_k \int d^3r_1 \int d^3r_2 \int d^3r_3 e^{-i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} \langle\delta(\mathbf{r}_2)\delta(\mathbf{r}_3)\rangle \\ W(\mathbf{r}_1)W(\mathbf{r}_2)W_{\mathrm{ic}}(\mathbf{r}_3)\mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}_1)\epsilon_{\mathrm{ic}}(\mathbf{r}_1,\mathbf{r}_3) \quad (4.172)$$

gives:

$$IC_\ell^{\mathrm{ic},\delta}(k) = \frac{2\ell+1}{A}(-i)^\ell \int d^3s j_\ell(ks)\sum_{p,n}\int d^3\Delta\Delta^n \xi_p^{\mathrm{ep},(n)}(\Delta) \\ \int d^3r_2 r_2^{-n} W(\mathbf{r}_2+\mathbf{s})W(\mathbf{r}_2)W_{\mathrm{ic}}(\mathbf{r}_2-\mathbf{\Delta})\mathcal{L}_\ell(\hat{\mathbf{r}}_1\cdot\hat{\mathbf{s}})\mathcal{L}_p(\hat{\mathbf{r}}_2\cdot\hat{\mathbf{\Delta}})\epsilon_{\mathrm{ic}}(\mathbf{r}_2+\mathbf{s},\mathbf{r}_2-\mathbf{\Delta}) \quad (4.173)$$

with $\mathbf{s}=\mathbf{r}_1-\mathbf{r}_2$, $\mathbf{\Delta}=\mathbf{r}_2-\mathbf{r}_3$. Taking the opposite $\mathbf{s}\to-\mathbf{s}$ and defining:

$$\mathcal{W}_{\ell p}^{\mathrm{ic},\delta,(n)}(s,\Delta) = \frac{(2\ell+1)(2p+1)}{(4\pi)^2 A}\int d\Omega_s \int d\Omega_\Delta \\ \int d^3x x^{-n} W(\mathbf{x})W(\mathbf{x}-\mathbf{s})W_{\mathrm{ic}}(\mathbf{x}-\mathbf{\Delta})\mathcal{L}_\ell(\widehat{\mathbf{x}-\mathbf{s}}\cdot(-\hat{\mathbf{s}}))\mathcal{L}_p(\hat{\mathbf{x}}\cdot\hat{\mathbf{\Delta}})\epsilon_{\mathrm{ic}}(\mathbf{x}-\mathbf{s},\mathbf{x}-\mathbf{\Delta}) \quad (4.174)$$

we obtain the integral constraint correction (4.170), with $(i,j)=(\mathrm{ic},\delta)$. Let us move to the last term (4.166):

$$IC_\ell^{\mathrm{ic},\mathrm{ic}}(k) = \frac{2\ell+1}{4\pi A}\int d\Omega_k \int d^3r_1 \int d^3r_2 \int d^3r_3 \int d^3r_4 e^{-i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} \langle\delta(\mathbf{r}_3)\delta(\mathbf{r}_4)\rangle \\ W(\mathbf{r}_1)W(\mathbf{r}_2)W_{\mathrm{ic}}(\mathbf{r}_3)W_{\mathrm{ic}}(\mathbf{r}_4)\mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}_1)\epsilon_{\mathrm{ic}}(\mathbf{r}_1,\mathbf{r}_3)\epsilon_{\mathrm{ic}}(\mathbf{r}_2,\mathbf{r}_4) \quad (4.175)$$

gives:

$$IC_\ell^{\mathrm{ic},\mathrm{ic}}(k) = \frac{2\ell+1}{4\pi A}(-i)^\ell \int d^3s j_\ell(ks)\sum_{p,n}\int d^3\Delta\Delta^n \xi_p^{\mathrm{ep},(n)}(\Delta)\int d^3r_1 W(\mathbf{r}_1)W(\mathbf{r}_1-\mathbf{s}) \\ \int d^3r_3 r_3^{-n} W_{\mathrm{ic}}(\mathbf{r}_3)W_{\mathrm{ic}}(\mathbf{r}_3-\mathbf{\Delta})\mathcal{L}_\ell(\hat{\mathbf{r}}_1\cdot\hat{\mathbf{s}})\mathcal{L}_p(\hat{\mathbf{r}}_3\cdot\hat{\mathbf{\Delta}})\epsilon_{\mathrm{ic}}(\mathbf{r}_1,\mathbf{r}_3)\epsilon_{\mathrm{ic}}(\mathbf{r}_1-\mathbf{s},\mathbf{r}_3-\mathbf{\Delta}) \quad (4.176)$$

with $\mathbf{s}=\mathbf{r}_1-\mathbf{r}_2$, $\mathbf{\Delta}=\mathbf{r}_3-\mathbf{r}_4$. This is Eq. (4.170) with $(i,j)=(\mathrm{ic},\mathrm{ic})$ if we define:

$$\mathcal{W}_{\ell p}^{\mathrm{ic},\mathrm{ic},(n)}(s,\Delta) = \frac{(2\ell+1)(2p+1)}{(4\pi)^2 A}\int d\Omega_s \int d\Omega_\Delta \int d^3y W(\mathbf{y})W(\mathbf{y}-\mathbf{s}) \\ \int d^3x x^{-n} W_{\mathrm{ic}}(\mathbf{x})W_{\mathrm{ic}}(\mathbf{x}-\mathbf{\Delta})\mathcal{L}_\ell(\hat{\mathbf{y}}\cdot\hat{\mathbf{s}})\mathcal{L}_p(\hat{\mathbf{x}}\cdot\hat{\mathbf{\Delta}})\epsilon_{\mathrm{ic}}(\mathbf{y},\mathbf{x})\epsilon_{\mathrm{ic}}(\mathbf{y}-\mathbf{s},\mathbf{x}-\mathbf{\Delta}). \quad (4.177)$$

To summarise, the window-convolved, integral-constraint corrected power spectrum reads:

$$P_\ell^{\mathrm{cic}}(k) = 4\pi(-i)^\ell \int s^2 ds j_\ell(ks)\xi_\ell^{\mathrm{cic}}(s) \quad (4.178)$$

with ξ_ℓ^{cic} given by:

$$\boxed{\xi_\ell^{\mathrm{cic}}(s) = \xi_\ell^{\mathrm{c}}(s) - IC_\ell^{\delta,\mathrm{ic}}(s) - IC_\ell^{\mathrm{ic},\delta}(s) + IC_\ell^{\mathrm{ic},\mathrm{ic}}(s)} \quad (4.179)$$

In this equation, we recall that (see Eq. (4.137)):

$$\xi_\ell^{\rm c}(s) = \sum_{p,q,n} A_{\ell p}^q \frac{2\ell+1}{2q+1} s^n \xi_p^{(n)}(s) \mathcal{W}_q^{\delta,\delta,(n)}(s), \tag{4.180}$$

and:

$$IC_\ell^{i,j}(s) = \int \Delta^2 d\Delta \sum_{p,n} \frac{4\pi}{2p+1} \Delta^n \xi_p^{{\rm ep},(n)}(\Delta) \mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta). \tag{4.181}$$

To speed up practical calculations, since $IC_\ell^{i,j}(s)$ are linear in $\xi_p^{{\rm ep},(n)}$, $\mathcal{W}_{\ell p}^{i,j,(n)}$ can be directly summed over $(i,j) \in \{(\delta,{\rm ic}),({\rm ic},\delta),({\rm ic},{\rm ic})\}$ to yield a single $IC_\ell(s)$ term. In the specific case of the global integral constraint ($\epsilon_{\rm ic}(\mathbf{r},\mathbf{x}) = 1$), the window function (4.177) is separable and:

$$IC_\ell^{\rm glo,glo}(k) = \frac{\int \Delta^2 d\Delta \xi_0^{\rm c}(\Delta)}{\int \Delta^2 d\Delta \mathcal{W}_0^{\delta,\delta,(0)}(\Delta)} \mathcal{W}_\ell^{\delta,\delta,(0)}(k). \tag{4.182}$$

with $\mathcal{W}_\ell^{\delta,\delta,(0)}(k)$ the Hankel transform of $\mathcal{W}_\ell^{\delta,\delta,(0)}(s)$ introduced in Eq. (4.138). As noted by Beutler et al. (2019), this is correct at any order n included in the calculation of $\xi_0^{\rm c}(\Delta)$.

The line of sight definition in $\mathcal{W}_\ell^{\delta,\delta,(n)}(s)$ of Eq. (4.138), and the first line of sight in $\mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta)$ of Eq. (4.171), (4.177) and (4.177), used to define the cosine angle with the separation vector $\mathbf{s}$, should be the same as in the power spectrum (or correlation function) estimator. Conversely, the line of sight connecting $\xi_p^{{\rm ep},(n)}(\Delta)$ to $\mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta)$, defining the cosine angle with the separation vector $\mathbf{\Delta}$, is a purely practical choice. Our calculations use the endpoint line of sight, but the derivation with any line of sight $\mathbf{d}$ is straightforward by replacing x^{-n} in Eq. (4.171), (4.174) and (4.177) by d^{-n}, changing the arguments of Legendre polynomials accordingly and taking the corresponding $\xi_p^{\mathbf{d},(n)}(\Delta)$ in Eq. (4.170). In particular, taking the midpoint line of sight makes first order ($n = 1$) wide-angle corrections vanish and makes Eq. (4.171) and Eq. (4.174) equal.

We have used the endpoint line of sight as second line of sight in $\mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta)$. With this choice, by definition, the integral constraint corrections completely cancel with the convolved power spectrum monopole $P_0^{\rm c}(k)$ on large scales. This would only be asymptotically true (as the wide-angle correction order $n \to +\infty$) if a different line of sight definition was used in $\mathcal{W}_\ell^{\delta,\delta,(n)}(s)$ and $\mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta)$. As mentioned in Section 4.2.5, we restrict ourselves to the zeroth order ($n = 0$) wide-angle expansion as we will show it to be a fair approximation for a typical BOSS-like survey. Let us see how the above results compare to previous ones in the next section.

4.3.4 Discussion

The global integral constraint was first discussed in Peacock and Nicholson (1991); Beutler et al. (2014); Wilson et al. (2017), who expressed the observed window-convolved, integral-constraint-corrected power spectrum as:

$$P^{\rm cic}(\mathbf{k}) = P^{\rm c}(\mathbf{k}) - P^{\rm c}(\mathbf{0}) \, |W(\mathbf{k})|^2 \tag{4.183}$$

with $P^{\rm c}(\mathbf{k})$ the window-convolved power spectrum. $W(\mathbf{k})$ is the (Fourier transform of) the survey selection function, rescaled by imposing $|W(\mathbf{k})|^2 = 1$ when $k \to 0$, so that $P^{\rm cic}(k) \to 0$ as $k \to 0$. Only one correction term is assumed, $P^{\rm c}(\mathbf{0})|W(\mathbf{k})|^2$, while our complete derivation yields 3 terms, see Eq. (4.179) with ic = glo.

The multipoles of the integral constraint correction $P^{\rm c}(\mathbf{0})|W(\mathbf{k})|^2$ of Eq. (4.183) are actually exactly equal to our last correction term, $IC_\ell^{\rm glo,glo}(k)$ (4.182). However, Eq. (4.183) introduces a

minus sign in front of $P^{\rm c}(\mathbf{0})|W(\mathbf{k})|^2$ to obtain a damping of the power spectrum at large scales. Instead, in our Eq. (4.179), $IC_\ell^{\rm glo,glo}(s)$, which accounts for the auto-correlation of the density contrast integral over the footprint, leads to an increase of power at large scales. Damping of power comes from the two other terms $-IC_\ell^{\delta,\rm glo}(s) - IC_\ell^{\rm glo,\delta}(s)$ which correspond to the cross-correlations between the density contrast and its integral over the footprint, respectively. Finally, note that our full derivation of the integral constraint effect makes it clear that the correction $P^{\rm c}(\mathbf{0})|W(\mathbf{k})|^2$, with W describing the full survey footprint, is no longer valid if this footprint is composed of different chunks with separate normalisation factors (4.156). Indeed, in this case, $\epsilon_{\rm glo}(\mathbf{r},\mathbf{x})$ is 1 if $\mathbf{r}$ and $\mathbf{x}$ belong to the same chunk and integrals of term (4.177) cannot be split as in Eq. (4.182); the resulting global IC effect is then larger.

We detail the different terms contributing to the global integral constraint on Figure 4.4 (left). As will be described in Section 4.3.6, in this figure, global and radial integral constraints are modelled for a BOSS CMASS-like survey footprint divided in 6 chunks (which is encoded in $\epsilon_{\rm ic}(\mathbf{r},\mathbf{x})$). We also include the shot noise correction as will be explained in Section 4.3.5.2. As expected, on large scales, the window-convolved, integral-constraint-corrected monopole reaches 0. The effect of the complete global integral constraint is negligible at the scales involved in a RSD analysis ($k \gtrsim 0.01\,h\,\mathrm{Mpc}^{-1}$), but is significant at large scales. As also shown in this figure, taking only $-IC_\ell^{\rm glo,glo}$, as in Eq. (4.183), for the global integral constraint correction (4.179) appears to be a very legitimate approximation in the illustrated survey case [7].

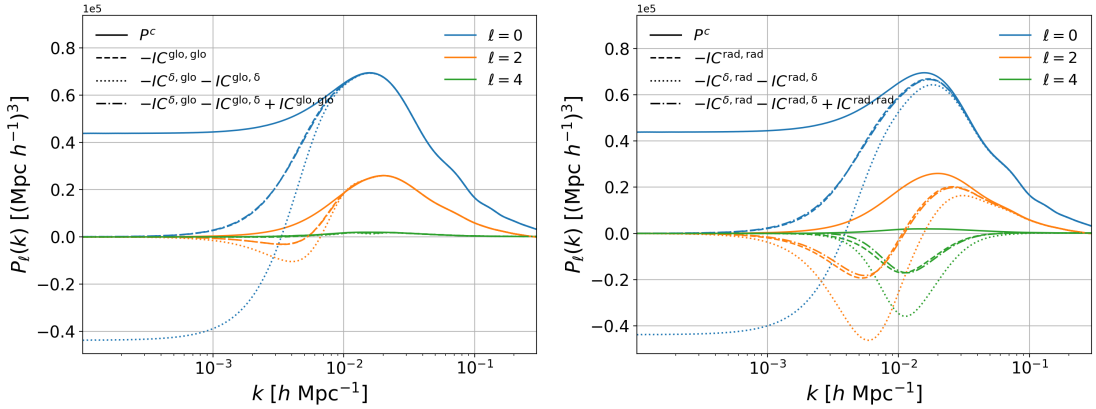

Figure 4.4 – Left: power spectrum multipoles (blue: monopole, orange: quadrupole, green: hexadecapole) including the different contributions to the global integral constraint (see text). The complete result ($-IC_\ell^{\delta,\rm glo} - IC_\ell^{\rm glo,\delta}$, dot-dashed curve) cannot be distinguished from the partial correction ($-IC_\ell^{\rm glo,\delta}$, dashed). Right: same, with the radial integral constraint. See Section 4.3.6 for more details about the survey configuration assumed in this figure.

Figure 4.4 (right) displays the different contributions to the radial integral constraint. Compared to the global IC, the radial IC has a larger effect on a broader range of wavenumbers, especially in the quadrupole and hexadecapole. We may thus expect a non-negligible impact of the radial integral constraint on clustering measurements in the illustrated case.

7. We checked that this remains a good approximation for the radial integral constraint, as can be inferred from Figure 4.4 (right, dot-dashed and dashed curves). However, we observed that this approximation fails to account for the angular integral constraint described in Section 4.3.7.

4.3.5 Computing window functions

This section is devoted to the practical computation of window functions $\mathcal{W}_\ell^{\delta,\delta,(n)}(s)$ and $\mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta)$ ($(i,j)\in\{(\delta,\mathrm{ic}),(\mathrm{ic},\delta),(\mathrm{ic},\mathrm{ic})\}$) required in Section 4.3.3. We first argue that a reasonable approximation for the true survey selection function can be inferred from data. We then detail the contribution from Poisson shot noise. Finally, we present algorithms required to compute these window functions from a synthetic catalogue sampling the survey selection function.

4.3.5.1 Approaching the true survey selection function

Our derivation of integral constraint corrections makes use of the true, underlying, survey selection function $W(\mathbf{r})$. However, as we mentioned previously, one has often only access to a synthetic catalogue whose density $n_s(\mathbf{r})$ is tuned to match the actual data in certain regions (chunks, or radial bins for the radial IC). Thus, clustering modes inprinted in the data distribution are propagated to this synthetic catalogue. Using $n_s(\mathbf{r})$ for $W(\mathbf{r})$ does not change the value of $W_{\mathrm{ic}}(\mathbf{r})$ as it is normalised in regions where the integral constraint is imposed, i.e. where the expected mean of $n_s(\mathbf{r})$ is unknown. However, if that tuned $n_s(\mathbf{r})$ is taken for $W(\mathbf{r})$, the calculated window effect and integral constraint corrections statistically differ from the truth.

We can provide some estimate for this bias in the dominant term (4.123). The density $n_s(\mathbf{r})$ tuned on the actual data can be expressed as (see Eq. (4.157) and (4.160)):

$$n_s(\mathbf{r}) = W(\mathbf{r})\left\{1+\int d^3x W_{\mathrm{ic}}(\mathbf{x})\delta(\mathbf{x})\epsilon_{\mathrm{ic}}(\mathbf{r},\mathbf{x})\right\}, \tag{4.184}$$

where $W_{\mathrm{ic}}(\mathbf{r})$ can be described by $n_{s,\mathrm{ic}}(\mathbf{r}) = \frac{n_s(\mathbf{r})}{\int d^3x n_s(\mathbf{x})\epsilon_{\mathrm{ic}}(\mathbf{r},\mathbf{x})}$. Then, the multipoles of the 2-point window function estimated from $n_s(\mathbf{r})$ can be written:

$$\mathcal{S}_\ell^{\delta,\delta,(0)}(s) = \mathcal{W}_\ell^{\delta,\delta,(0)}(s) + IC_\ell^{\mathrm{ic},\mathrm{ic}}(s). \tag{4.185}$$

Taking $n_s(\mathbf{r})$ for $W(\mathbf{r})$ in $IC_\ell^{\mathrm{ic},\mathrm{ic}}(s)$, Eq. (4.185) provides an estimate for $\mathcal{W}_\ell^{\delta,\delta,(0)}(s)$ at first order in the integral constraint corrections. A relation similar to Eq. (4.185) holds for any wide-angle correction order m, $\mathcal{S}_\ell^{\delta,\delta,(m)}(s) = \mathcal{W}_\ell^{\delta,\delta,(m)}(s) + IC_\ell^{\mathrm{ic},\mathrm{ic}}(s)$, provided $\mathcal{W}_{\ell p}^{\mathrm{ic},\mathrm{ic},(n)}$ (see Eq. (4.177)) entering $IC_\ell^{\mathrm{ic},\mathrm{ic}}(s)$ (see Eq. (4.176)) is replaced by:

$$\begin{aligned}\mathcal{W}_{\ell p}^{\mathrm{ic},\mathrm{ic},(m),(n)}(s,\Delta) = \frac{(2\ell+1)(2p+1)}{(4\pi)^2 A}\int d\Omega_s\int d\Omega_\Delta\int d^3y\, y^{-m} W(\mathbf{y})W(\mathbf{y}-\mathbf{s})\\ \int d^3x\, x^{-n} W_{\mathrm{ic}}(\mathbf{x})W_{\mathrm{ic}}(\mathbf{x}-\mathbf{\Delta})\mathcal{L}_\ell(\hat{\mathbf{y}}\cdot\hat{\mathbf{s}})\mathcal{L}_p(\hat{\mathbf{x}}\cdot\hat{\mathbf{\Delta}})\epsilon_{\mathrm{ic}}(\mathbf{y},\mathbf{x})\epsilon_{\mathrm{ic}}(\mathbf{y}-\mathbf{s},\mathbf{x}-\mathbf{\Delta})\end{aligned} \tag{4.186}$$

(and similarly for other line of sight definitions).

Note that such a dependence of the estimation of window effects and integral constraint corrections in the observed clustering also yields a change in the variance of cosmological measurements, whose study is beyond the scope of this manuscript.

4.3.5.2 Shot noise

In principle, formulae given in Section 4.3.3 (and Eq. (4.185)) fully describe integral constraint corrections. However, they are derived in real space, where accounting for a Poisson shot noise term (a spike at $s=0$) is numerically challenging.

The shot noise contribution to the correlation function weighted by the window function only shows up in the monopole (see Eq. (4.65)):

$$\xi_0^{\mathrm{c}}(s) \ni \int d^3x W(\mathbf{x}) w(\mathbf{x}) \delta_D^{(3)}(s) \tag{4.187}$$

where $w(\mathbf{x})$ is the total (data and random [8]) weight at position $\mathbf{x}$. In Fourier space, this translates in a simple offset in the monopole. However, the shot noise contribution to the integral constraint correction leaks on all scales of all multipoles:

$$IC_\ell^{i,j}(s) \ni SN_\ell^{i,j}(s) \qquad (i,j) \in \{(\delta,\mathrm{ic}),(\mathrm{ic},\delta),(\mathrm{ic},\mathrm{ic})\}, \tag{4.188}$$

with:

$$SN_\ell^{\mathrm{ic},\delta}(s) = \frac{2\ell+1}{4\pi} \int d\Omega_s \int d^3x W(\mathbf{x}) w_{\mathrm{ic}}(\mathbf{x}) W(\mathbf{x}+\mathbf{s}) \mathcal{L}_\ell(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}}) \epsilon_{\mathrm{ic}}(\mathbf{x},\mathbf{x}+\mathbf{s}) \tag{4.189}$$

and

$$SN_\ell^{\mathrm{ic},\mathrm{ic}}(s) = \frac{2\ell+1}{4\pi} \int d\Omega_s \int d^3x W(\mathbf{x}) W(\mathbf{x}+\mathbf{s}) \int d^3y W_{\mathrm{ic}}(\mathbf{y}) w_{\mathrm{ic}}(\mathbf{y}) \mathcal{L}_\ell(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}}) \epsilon_{\mathrm{ic}}(\mathbf{x},\mathbf{y}) \epsilon_{\mathrm{ic}}(\mathbf{x}+\mathbf{s},\mathbf{y}), \tag{4.190}$$

where we used $w_{\mathrm{ic}}(\mathbf{r}) = \frac{w(\mathbf{r})}{\int d^3x W(\mathbf{x}) \epsilon_{\mathrm{ic}}(\mathbf{r},\mathbf{x})}$ and considered the endpoint line of sight $\hat{\mathbf{x}}$.

Integrals over $\mathbf{s}$ of shot noise contributions to $\xi_0^{\mathrm{c}}(s)$ and $IC_\ell^{i,j}(s)$ are equal, such that total shot noise contributions vanish as $k \to 0$, as expected. In practice, we choose to normalise terms (4.187) and (4.188) by their integral over $\mathbf{s}$ [9], and multiply them by the shot noise measured by the Yamamoto estimator (4.67).

The impact of the global and radial integral constraints on shot noise is shown in Figure 4.5. By definition, the total shot noise reaches 0 as $k \to 0$ in the monopole. The impact of the RIC on the quadrupole and hexadecapole shot noise should not be ignored.

Note that a common practice of Fourier space clustering analyses is to remove the shot noise contribution from the power spectrum monopole measurement (see Eq. (4.67)). However, as shown above, the shot noise term is not a simple constant anymore once integral constraints are accounted for. Therefore, in this manuscript, we compare the power spectrum measurements with their shot noise and add up in the model the shot noise as described above.

4.3.5.3 Calculation of window functions from a synthetic catalogue

The survey selection function entering window functions $\mathcal{W}_\ell^{\delta,\delta,(n)}(s)$ and $\mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta)$ ($(i,j) \in \{(\delta,\mathrm{ic}),(\mathrm{ic},\delta),(\mathrm{ic},\mathrm{ic})\}$) required in Section 4.3.3 can be randomly sampled by a synthetic catalogue. We implement a classic pair-count algorithm to compute the anisotropic 2-point correlation and the algorithm from Slepian and Eisenstein (2018) to compute the anisotropic 3-point correlation of the synthetic catalogue. Lines of sight are defined according to the Yamamoto power spectrum estimator which we use for clustering measurements.

8. The total shot noise estimate (Eq. (4.187) and (4.188)) is valid in the limit of infinite random density, but appears to be a legitimate approximation if the random density is high enough, as shown by our cosmological fits in Section 4.3.6.3, 4.3.7.2, and 4.3.7.3.

9. In the case W is sampled by a synthetic catalogue, the normalisation for $SN_\ell^{\mathrm{ic},\delta}(s)$ and $SN_\ell^{\mathrm{ic},\mathrm{ic}}(s)$ is the (weighted) number of correlated pairs.

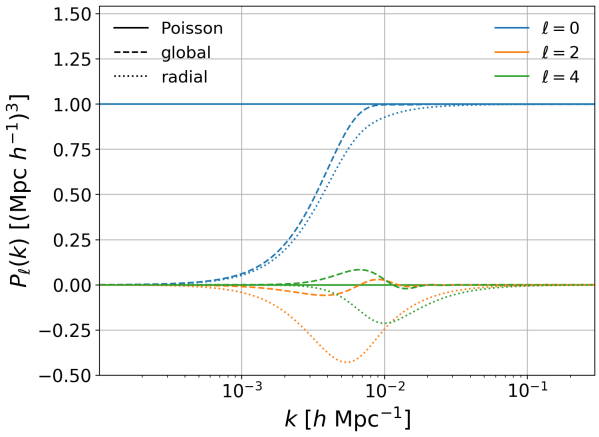

Figure 4.5 – Shot noise with the global (dashed lines) and radial (dotted lines) integral constraints (blue: monopole, orange: quadrupole, green: hexadecapole), relative to the standard Poisson shot noise $P_0^{\rm noise}$ of Eq. (4.67), which we normalised to 1. As expected, the standard, constant, shot noise term is cancelled by the integral constraint contribution at large scales. The same survey case as in Figure 4.4 is assumed.

Pairs (or triplets) of random objects are binned in comoving distance bins of typical size $1\,{\rm Mpc}\,h^{-1}$ (for $\mathcal{W}_\ell^{\delta,\delta,(n)}(s)$) to $4\,{\rm Mpc}\,h^{-1}$ (for $\mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta)$); the result is divided by the comoving volume of each distance bin to recover density, renormalised to data counts (by multiplying by α^2), and divided by the normalisation factor A.

Window function calculations $\mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta)$ involving $\epsilon_{\rm ic}(\mathbf{r},\mathbf{x})$ are performed by correlating random objects at position $\mathbf{r}$ with random objects at position $\mathbf{x}$ satisfying $\epsilon_{\rm ic}(\mathbf{r},\mathbf{x}) \neq 0$, i.e. residing in the same spatial bin.

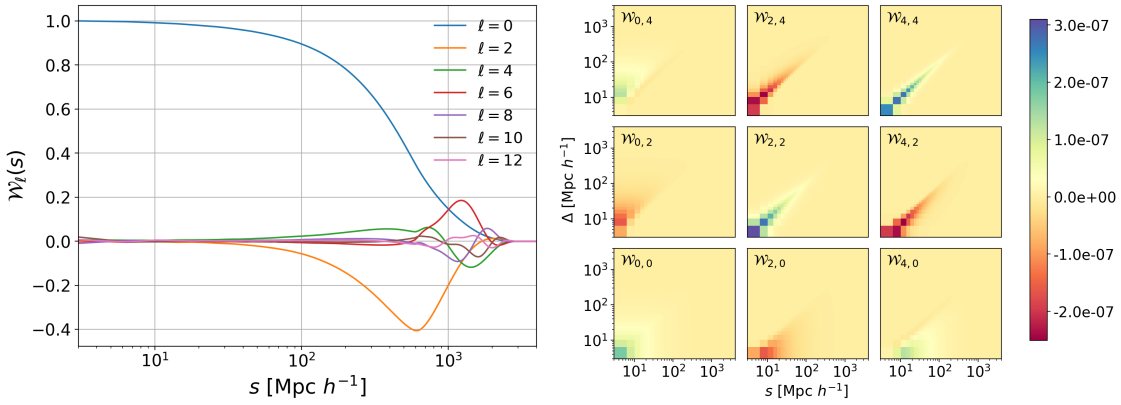

Figure 4.6 – Left: the window function multipoles $\mathcal{W}_\ell^{\delta,\delta,(0)}(s)$ convolving (in Fourier space) the power spectrum. Right: the window function multipoles $\mathcal{W}_{\ell p}^{\rm rad,rad,(0)}(s,\Delta)$ entering the integral constraints.

We show examples of $\mathcal{W}_\ell^{\delta,\delta,(0)}(s)$ and $\mathcal{W}_{\ell p}^{\rm rad,rad,(0)}(s,\Delta)$ in Figure 4.6. All window functions (for the global and radial integral constraints) could be accurately computed in $\simeq 1500$ CPU hours.

We can also take advantage of the usual feature of survey selection functions $W(\mathbf{x})$, namely that they are separable into an angular and a radial parts. Then, we can compute the angular 2 or 3-point correlation function of the survey selection function and perform the integration over the footprint depth. This (less noisy) method yields similar results as the one described above and can be faster in the case of the wide-angle expansion, where the window functions

$\mathcal{W}_\ell^{\delta,\delta,(n)}(s)$ and $\mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta)$ of different orders n can be calculated from the same angular correlation functions.

We implemented all the algorithms required to compute the 2- and 3-point correlation functions of the survey selection function in a coherent `C` code, wrapped into `Python`, dubbed `pycute` [10].

4.3.6 Radial integral constraint in RSD analyses

In this section we discuss the impact of the radial integral constraint on Fourier-space RSD measurements, and show that Eq. (4.179), along with practical details developed in Section 4.3.5, helps us recover an unbiased estimation of cosmological parameters. We consider three sets of mock catalogues: the *baseline* mocks for measurements including only the global IC, and the *binned* and *shuffled* mocks for the radial one. These mocks are described in Section 4.3.6.1, while details about power spectrum measurements and the RSD modelling are given in Section 4.3.6.2. Cosmological fits are performed in Section 4.3.6.3. We discuss the potential impact of the radial integral constraint on other clustering analyses in Section 4.3.6.4.

4.3.6.1 Mock catalogues

We work on a set of 84 high fidelity N-series mocks used for the LRG clustering mock challenge (Tinker, 2016) of the BOSS Data Release 12 (DR12) (Alam et al., 2017). These mocks were built from the 3 projections of 7 independent, periodic box realisations of side $2600\,\mathrm{Mpc}\,h^{-1}$ at redshift $z = 0.5$. The simulated cosmology is a flat ΛCDM model with:

$$\begin{aligned} h = 0.7,\ \Omega_{m,0} = 0.286,\ \omega_{b,0} = 0.02303, \\ \sigma_{8,0} = 0.82,\ n_s = 0.96. \end{aligned} \tag{4.191}$$

Simulations were run with the GADGET2 code (Springel, 2005), and a HOD modelling was used to populate dark matter halos with galaxies, mimicking the observed clustering in data.

The covariance matrix required for cosmological fits is built from 2048 Multidark Patchy mocks provided by the BOSS collaboration. These approximate mocks were calibrated on BigMultiDark simulations (Klypin et al., 2016); halo abundance matching was applied to reproduce the 2- and 3-point clustering measurements and mocks at different redshifts were combined into light cones (Kitaura et al., 2016). We use V6C catalogues, which were adjusted to reproduce the data clustering measurements. These mocks are also used to compute the standard deviation (blue shaded area) shown in Figure 4.8, 4.13 and the error bars plotted in Figure 4.11.

Both N-series and Multidark Patchy mocks were trimmed following the DR12 CMASS NGC (hereafter abbreviated by CMASS) selection function using the `make_survey` software (White et al., 2014). We impose integral constraints consistently on N-series and Multidark Patchy mocks in each series of cosmological fits. *Baseline* mocks are obtained by using a random catalogue sampling the true survey selection function, normalised according to Eq. (4.156), so that those mocks are impacted by the global integral constraint only. The *binned* or *shuffled* schemes are used to impose the radial integral constraint. In the *shuffled* scheme, random redshifts are drawn from each mock data realisation. In the *binned* scheme, random galaxies are weighted to reproduce the mock data radial distribution in comoving distance bins of size $\delta r = 2\,\mathrm{Mpc}\,h^{-1}$.

10. at https://github.com/adematti/pycute

To enhance the radial integral constraint effect coming from the *binned* or *shuffled* schemes, we divide the CMASS footprint ($\simeq 7420\,\mathrm{deg}^2$) in 6 smaller chunks of size ranging from $\simeq 980\,\mathrm{deg}^2$ to $\simeq 1570\,\mathrm{deg}^2$ (see Figure 4.7), areas representative of the 4 chunks of the eBOSS ELG survey (Raichoor et al., 2017). The *binned* or *shuffled* schemes are applied to the 6 chunks separately, before these are recombined in a single catalogue used for the power spectrum measurement. This procedure, though not representative of the real BOSS DR12 analysis, allows us to test our modelling of the radial integral constraint in stringent conditions.

One can appreciate the difference between the true survey selection function (continuous black curve in Figure 4.7) and those measured from one mock realisation (dashed black curve for full CMASS and colored dashed curves for each chunk). Future spectroscopic surveys like DESI will be very wide. Their angular selection function may hence vary on the sky, depending on e.g. the photometric depth achieved. Modelling these variations may indeed require to combine small patches with locally constant radial selection functions.

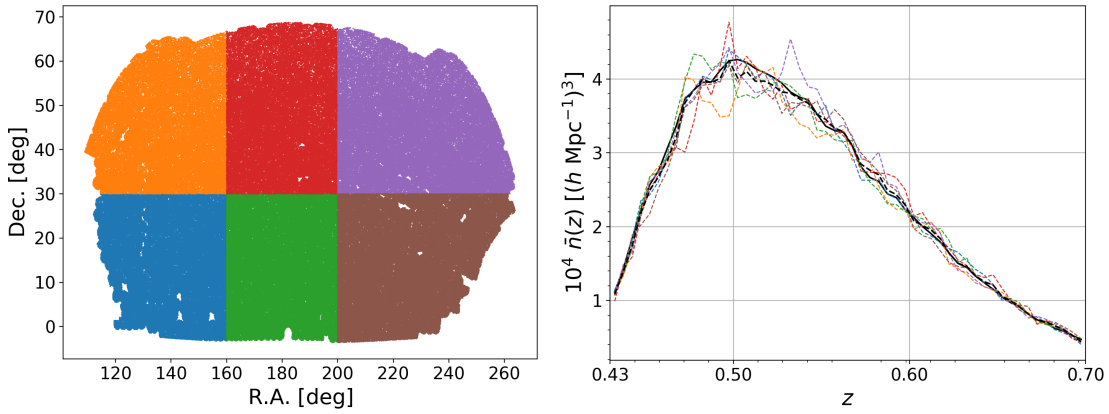

Figure 4.7 – Left: the CMASS footprint, divided in 6 different chunks. Right: different estimates of the redshift density $n(z)$, using bins $\Delta z = 0.005$: the redshift density of the true survey selection function (continuous black curve), and, for one mock data realisation, $n(z)$ measured in the whole CMASS footprint (dashed black curve) and in the 6 chunks individually. The scatter between the $n(z)$ estimates is due to noise and clustering.

We apply FKP weights $w_{\mathrm{FKP}}(z) = 1/\left(1 + n(z)P_0\right)$ to both mock data and randoms using $P_0 = 20\,000\,\mathrm{Mpc}^3\,h^{-3}$. The redshift density $n(z)$, which also enters the normalisation of power spectrum measurements (see Eq. (4.68), with $n_{s,i} = n(z_i)$), is calculated in bins of $\Delta z = 0.005$.

The redshift density is computed according to the type of mocks. When the radial IC is imposed (*binned* or *shuffled* mocks), the redshift density $n(z)$ is computed for each CMASS mock data realisation, as is done in an actual data analysis (see Section 5.3.1.7). As we divide the CMASS footprint into 6 chunks, one may want to measure $n(z)$ in the 6 chunks separately (see Figure 4.7, right). However, we will show in Section 4.3.8 that doing so significantly biases clustering measurements as FKP weights, using local $n(z)$, smooth out clustering. We therefore choose to measure $n(z)$ from the whole CMASS sample, making the bias due to FKP weights negligible.

Similarly, when the global IC (*baseline* mocks) is imposed, we compute $n(z)$ on the random catalogue accounting for the true survey selection function (see the continuous black curve on Figure 4.7, right) but rescale it according to the weighted number of data for each mock realisation.

4.3.6.2 Analysis methods

As mentioned in Section 4.1.6, we use the implementation of the Yamamoto estimator in the Python toolkit `nbodykit` (Hand et al., 2018) to compute (mock) data power spectra. We take a large box size of $4000\,\mathrm{Mpc}^3\,h^{-3}$ to reduce sampling effects in low k-bins. The FKP field is interpolated on a 512^3 mesh following the triangular shaped cloud scheme (see Section 4.1.7). The Nyquist frequency is thus $k \simeq 0.4\,h\,\mathrm{Mpc}^{-1}$, more than twice larger than the maximum wavenumber used in our analysis. We employ the interlacing technique discussed in Section 4.1.7 to mitigate aliasing effects. Power spectrum multipoles are measured in bins of $\Delta k = 0.01\,h\,\mathrm{Mpc}^{-1}$, starting from $k = 0\,h\,\mathrm{Mpc}^{-1}$. The discrete $\mathbf{k}$-space mesh makes the angular mode distribution irregular at large scales, an effect which we account for in the model following Section 4.2.4 (with a varying line of sight).

We use the model described in Section 4.2.5. Power spectrum monopole, quadrupole and hexadecapole are fitted from $0.01\,h\,\mathrm{Mpc}^{-1}$ to $0.15\,h\,\mathrm{Mpc}^{-1}$.

We use the fiducial BOSS DR12 cosmology in our analysis:

$$\begin{aligned} h = 0.676,\ \Omega_{m,0} = 0.31,\ \Omega_{\Lambda,0} = 0.69,\ \omega_{b,0} = 0.022, \\ \sigma_{8,0} = 0.80,\ n_s = 0.97,\ \sum m_\nu = 0.06\,\mathrm{eV}. \end{aligned} \tag{4.192}$$

In all figures showing the power spectrum model alone (4.4, 4.9), we use $f = 0.75$, $b_1 = 2$, $b_2 = 1$, $\sigma_v = 4\,\mathrm{Mpc}\,h^{-1}$.

The radial integral constraint correction is consistently calculated using in $\epsilon_{\mathrm{rad}}(\mathbf{r}, \mathbf{x})$ the same comoving distance bins of size $\delta r = 2\,\mathrm{Mpc}\,h^{-1}$ as used in the *binned* scheme. We account for the chunk-splitting of the CMASS footprint by adding the condition that $\mathbf{r}$ and $\mathbf{x}$ should belong to the same chunk for $\epsilon_{\mathrm{rad}}(\mathbf{r}, \mathbf{x})$ to be non-zero. The infinite sum over p in Eq. (4.181) should be truncated; as in Section 4.2.2.2 we use a maximum correlation function multipole $\ell = 4$.

Window function calculations are based on different estimates of the survey selection function, according to the type of mocks.

In the radial IC case (*binned* or *shuffled* mocks), to mimic an actual data analysis, the radial part of $W(\mathbf{r})$ should in principle be estimated on each realisation of the mocks, for which specific window function and integral constraint corrections should be derived. However, this would require a large computation time. Thus, in this work, window function estimations for the radial IC case are based on the selection function $n_s(\mathbf{r})$ calculated with a synthetic catalogue tuned to match the radial distribution (using the *binned* scheme) in each chunk of one realisation of mock data. All illustrations of window functions (Figure 4.6) and integral constraint corrections (Figure 4.4, 4.5, 4.9, 4.14) are provided for this estimation of the survey selection function.

Similarly, in the global IC case (*baseline* mocks), $W(\mathbf{r})$ entering window function calculations is taken to be the true selection function, which should be normalised using Eq. (4.156) for each realisation of the mocks, as done in a real data analysis. However, as for the radial IC case, we normalise the selection function on one realisation of the mocks only (the same realisation as in the radial IC case).

We checked that using in the model the estimate provided by Eq. (4.185) (with ic = glo in the *baseline* case, ic = rad in the *binned* or *shuffled* cases) for $\mathcal{W}_\ell^{\delta,\delta,(0)}(s)$ has a negligible impact on the measurement of cosmological parameters in our analysis. Hence, we do not include this correction in our cosmological fits.

4.3.6.3 Cosmological fits with the radial integral constraint

Figure 4.8 illustrates the effect of tuning the radial selection function on data. Both the *binned* scheme (using comoving distance bins of size $\delta r = 2\,\mathrm{Mpc}\,h^{-1}$) and the *shuffled* scheme result in a loss of power parallel to the line of sight, thus mostly affecting the power spectrum quadrupole and hexadecapole, compared to the *baseline* relying on the true survey selection function. The effect of the radial integral constraint is significantly increased when it is imposed to the 6 chunks separately (orange and red curves), which is the case considered in the following cosmological fits. We thus expect that neglecting the radial integral constraint would lead to a significant bias in the fitted cosmological parameters. Figure 4.8 also shows that power spectrum measurements obtained with the *binned* and *shuffled* scheme are very similar. It is thus fair to account for e.g. the *shuffled* scheme with the *binned* scheme, for which we computed window functions in Section 4.3.5.3. We also checked that differences between these two schemes in the measurement of cosmological parameters are negligible (i.e. small compared to the uncertainty on the mean of the mocks).

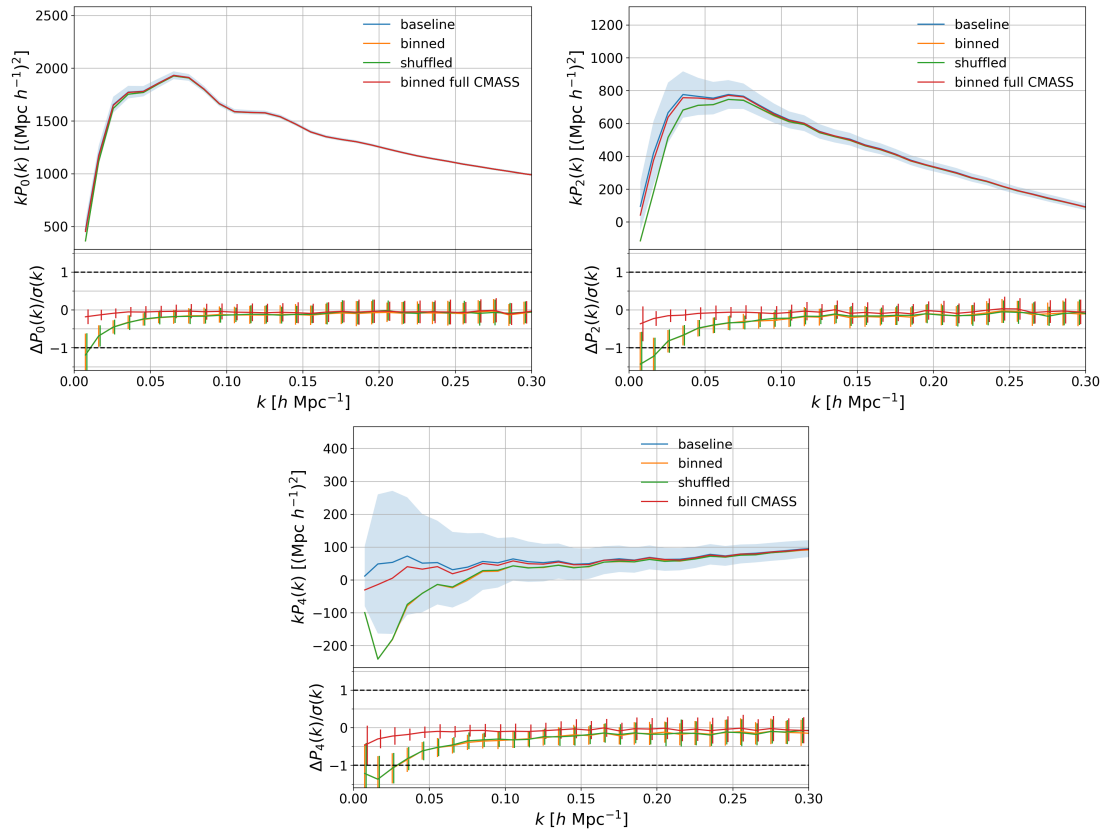

Figure 4.8 – Top panels: power spectrum multipoles (upper left: monopole, upper right: quadrupole, bottom: hexadecapole) measured from the 84 N-series mocks (see Section 4.3.6.1), using three different ways to model the redshift distribution in the random catalogues. The true selection function is used in the *baseline* (blue). The random redshift distribution of the *binned* (orange) and *shuffled* (green) schemes is inferred in 6 separate chunks of each mock data realisation, as described in the text. For comparison, the red curve shows the effect of the *binned* scheme applied to the full CMASS footprint. The blue shaded area represents the standard deviation of the mocks. Bottom panels: difference of the *shuffled* and *binned* schemes to the *baseline*, with the standard deviation of the difference given by the error bars, normalised by the standard deviation of the mocks.

Figure 4.9 displays the model for the global and radial integral constraint corrections corresponding to the CMASS footprint divided in 6 chunks. By construction, the power spectrum monopoles converge to zero at large scales when the global or radial IC is applied. We stress again that window functions are normalised according to Section 4.2.2.3, without using any low k nor low s limit. In the specific, illustrated case, the effect of the radial integral constraint is found to be large (comparable to the window function effect alone) in both Fourier and configuration space correlation functions. Note however that the latter prediction cannot be directly compared to configuration space measurements using e.g. the Landy-Szalay estimator (Landy and Szalay, 1993) from which the window function effect is already removed. We will come back to this point in Section 4.3.9.

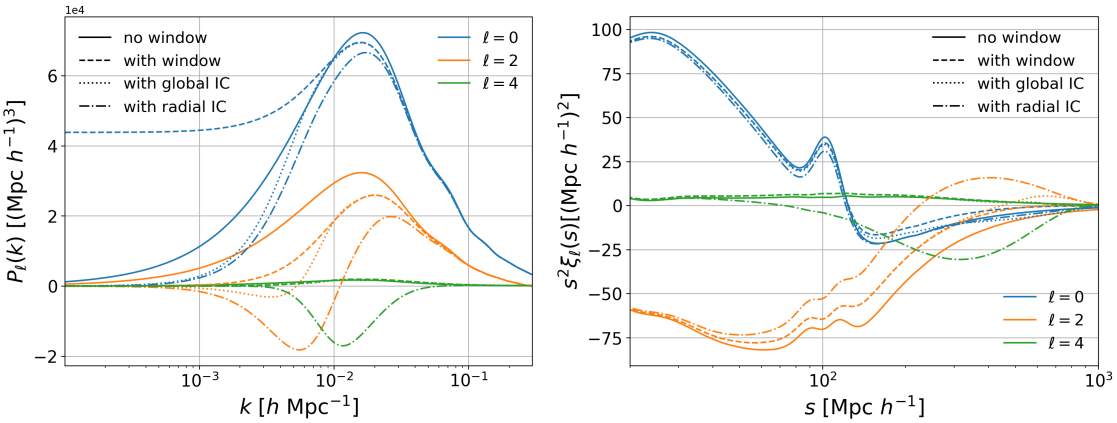

Figure 4.9 – Left: power spectrum multipoles (blue: monopole, orange: quadrupole, green: hexadecapole) without (continuous line) and with the window function effect only (dashed lines, first term in Eq. (4.179)), with the global integral constraint (dotted lines) and with the radial integral constraint (dash-dotted lines). Integral constraints ensure that the power spectrum monopoles reach zero on large scales. Right: same, in configuration space. The correlation function multipoles are multiplied by the window function $\mathcal{W}_\ell^{\delta,\delta,(0)}(s)$, i.e. we plot Eq. (4.179).

We show wide-angle contributions to the convolved power spectrum multipoles and the radial integral constraint in Figure 4.10. They are significant for $k \lesssim 1 \times 10^{-2}\,h\,\mathrm{Mpc}^{-1}$. However, they remain small compared to the radial integral constraint correction. Then, for simplicity, we did not include these wide-angle corrections ($n \geq 1$) in our analysis. However, wide-angle contributions may dominate over radial integral constraint corrections for a large survey with a constant radial selection function.

As a preview, Figure 4.11 illustrates the agreement between the mean of the 84 N-series mocks and the appropriate model in the *baseline* (left) and the *binned* scheme (right). In the *baseline* case, we fit the RSD model including the global integral constraint. The model performs well enough, given the high measurement precision (reduced $\chi^2 = 77.4/(42-7) = 2.21$). Note that same cosmological and nuisance parameters are kept for the model in the right-hand plot, where the *binned* scheme is applied to the mocks, and the radial integral constraint is included in the model. Here again, the agreement between the model and the mean of the mocks is correct (reduced $\chi^2 = 82.2/(42-7) = 2.35$). We thus anticipate that our radial integral constraint correction will very well account for the *binned* and *shuffled* schemes in the cosmological fits.

Figure 4.12 and Table 4.1 present the cosmological fit results obtained in three different cases.

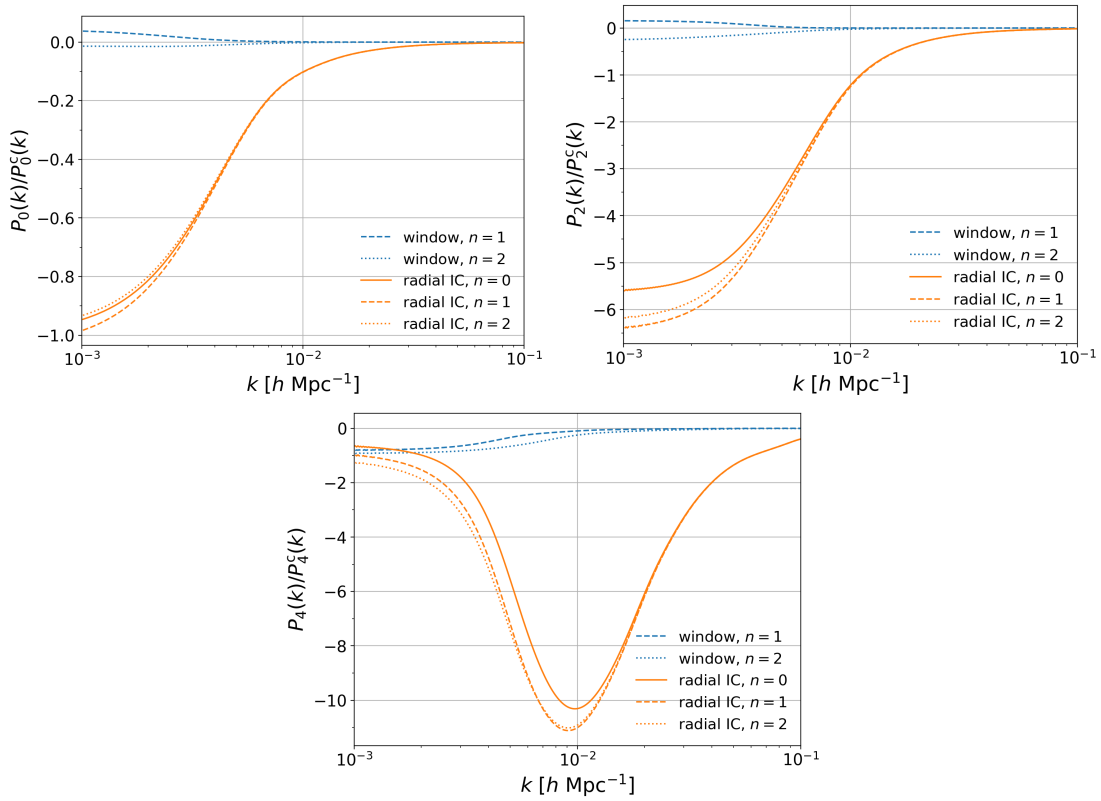

Figure 4.10 – Ratio of wide-angle corrections to the convolved power spectrum multipoles $P_\ell^c(k)$ (upper left: monopole, upper right: quadrupole, bottom: hexadecapole) at zeroth order. In blue are shown wide-angle contributions to the convolved power spectrum monopole up to order $n = 1, 2$ (dashed and dotted lines respectively). Corrections for the radial integral constraint up to order $n = 0, 1, 2$ (continuous, dashed and dotted lines respectively) are plotted in orange.

The *baseline* cosmological fits (column 1 in Table 4.1) are obtained with the power spectrum measurements on the *baseline* mocks and with the global integral constraint applied in the model. The expected parameter values (column 4 in Table 4.1), predicted from the N-series and fiducial cosmologies (Eq. (4.191) and Eq. (4.192)), are recovered to the statistical uncertainty on the mean of the mocks.

Applying the *binned* scheme to the mocks (see the orange curve in Figure 4.8) while modelling the global integral constraint only results in a bias (compared to the *baseline* case) on all cosmological parameters of roughly 30% of the statistical error on one realisation (column 2 in Table 4.1). The goodness-of-fit (probed by the χ^2 distribution, Figure 4.12) is significantly degraded ($\Delta\chi^2 \simeq 5$).

The modelling of the radial integral constraint successfully removes the bias to better than the statistical uncertainty on the mean of the mocks (column 3 in Table 4.1), and the goodness-of-fit is well recovered. No increase on cosmological parameter errors could be detected. Similar results (not reported here) were obtained with the *shuffled* scheme.

4.3.6.4 Discussion

We showed that the tuning of the radial selection function on data may significantly bias RSD measurements. We recall that the radial IC effect has been purposely enhanced to test the robustness of our predictions by applying the *binned* scheme separately in 6 small chunks cut

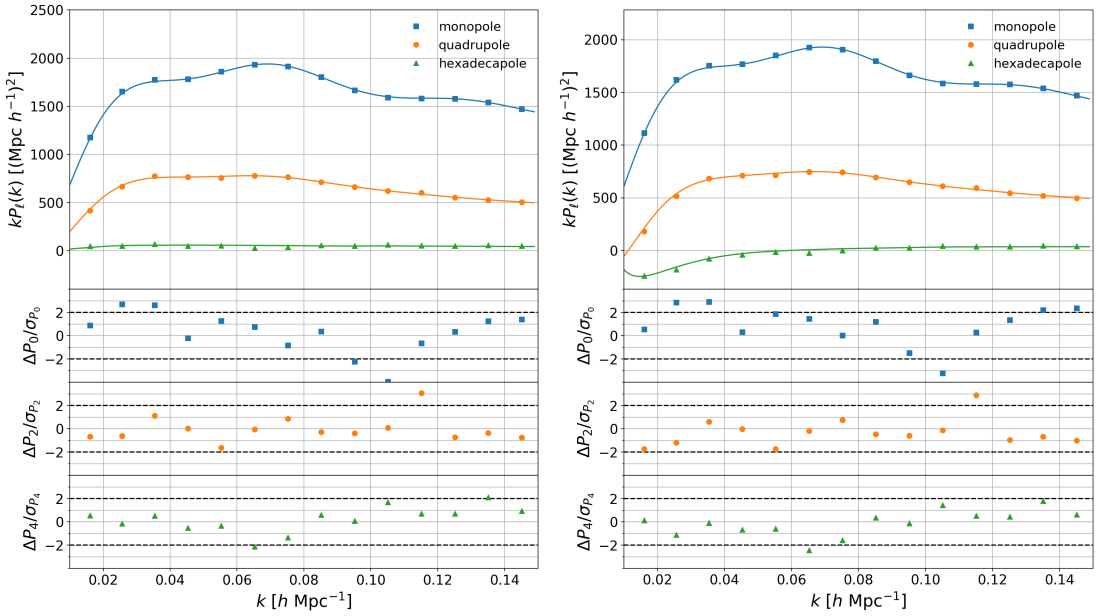

Figure 4.11 – Top panels: mean of the power spectrum multipoles (blue: monopole, orange: quadrupole, green: hexadecapole) measured from the 84 N-series mocks (data points, including error bars) and the corresponding model predictions (continuous lines). Error bars give the uncertainty on the mean of the 84 mocks (standard deviation of the MultiDark Patchy mocks divided by $\sqrt{84}$) but are below the marker size. Bottom panels: residuals (difference of the measurements to the model divided by error bars). Left: the model is fitted to the mean of the *baseline* mocks (see Figure 4.8), including the global integral constraint only. Right: the *binned* scheme is applied to the mocks, and the radial integral constraint is included in the model (without refitting).

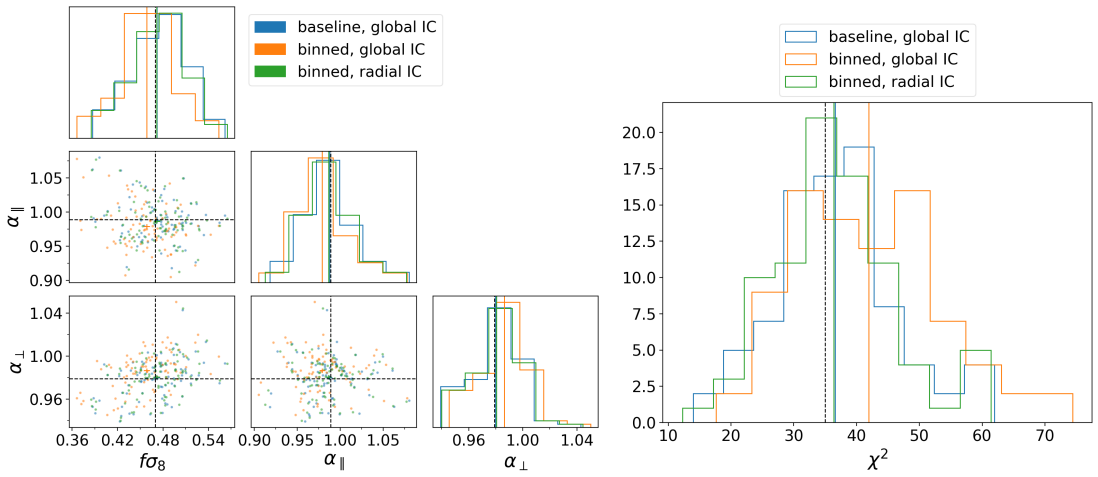

Figure 4.12 – Left: distributions of the cosmological parameters $f\sigma_8$, $\alpha_\parallel$, $\alpha_\perp$ measured on the 84 N-series mocks. The *baseline* (blue) uses the true selection function of the whole CMASS footprint. In orange, the *binned* scheme is applied to the mocks. In green, the radial integral constraint is added to the model. Continuous lines give the mean of the 84 best fits; the size of the cross in the 2D plots is the standard deviation of the best fits divided by $\sqrt{84}$. Dashed lines show the expected values from the mock cosmology. Right: the corresponding χ^2 distributions. The vertical dashed line shows the number of degrees of freedom ($42 - 7 = 35$).

into the CMASS footprint. After modelling of the induced radial integral constraint, the bias on $f\sigma_8$ and on scaling parameters is well below the uncertainty on the mean of the mocks. More

	Baseline global IC	Binned global IC	Binned radial IC	Expected
$\alpha_\parallel$	0.988 ± 0.031	0.979 ± 0.032	0.987 ± 0.031	0.989
$\alpha_\perp$	0.980 ± 0.018	0.986 ± 0.019	0.980 ± 0.018	0.979
$f\sigma_8$	0.472 ± 0.037	0.459 ± 0.037	0.472 ± 0.037	0.470
χ^2	36.6 ± 9.7	42 ± 11	36.4 ± 9.6	ndof = 35

Table 4.1 – Mean and standard deviation of the cosmological parameters fitted on the 84 N-series mocks, corresponding to Figure 4.12. Error bars should be divided by $\sqrt{84} \sim 10$ to obtain errors on the mean of the mocks.

statistics would be required to test our modelling further, but we note that the obtained bias is already below the expected $\simeq 1\%$ statistical uncertainty on $f\sigma_8$ expected from future galaxy surveys like DESI and Euclid (Laureijs, 2009).

The red curve in Figure 4.8 shows that the *shuffled* (or *binned*) scheme applied to the CMASS NGC footprint as a whole (i.e. without the division in 6 chunks) has a small effect on clustering measurements, as already noted on SDSS DR7 and DR9 by Samushia et al. (2012); Ross et al. (2012). We checked that not accounting for the radial IC in the model results in a negligible shift of cosmological parameters (compared to the statistical uncertainty) $\Delta\alpha_\parallel = -0.003$, $\Delta\alpha_\perp = 0.001$ and $\Delta f\sigma_8 = -0.001$. Note that, contrary to our test case, the BOSS DR12 analysis used the NGC and SGC CMASS and LOWZ samples. These were divided in three overlapping redshift slices (Alam et al., 2017), thus probably decreasing the radial integral constraint correction with respect to the pure window function effect. We thus infer that conclusions of the BOSS DR12 analysis are likely unaffected by the radial integral constraint effect.

Nonetheless, we will see in Chapter 5 that the radial integral constraint has a large effect on the eBOSS ELG clustering measurements, as its angular footprint is divided into small chunks to account for the angular variations of the selection function. In general, we expect the radial integral constraint to be non-negligible for deep arrow or pencil-like surveys. Also, analyses focusing on large scales, aiming at e.g. setting constraints on primordial non-Gaussianity (Ross et al., 2013; Castorina et al., 2019) may benefit from accounting for the radial IC.

4.3.7 Angular integral constraint to mitigate angular systematics

In this section we show that potential unknown angular systematics can be mitigated using an angular integral constraint which can be combined with the radial one. Though we note that only the additive part of systematics is removed in this way, we emphasise that the technique described hereafter may be used as a consistency check in clustering analyses.

4.3.7.1 Problem statement

As the radial selection function, the angular selection function of a spectroscopic survey can be difficult to evaluate because of residual photometric calibration errors or other potential photometric systematics. We consider these systematics to be completely unknown; that is, they are not and cannot be corrected by any photometric template in the analysis.

As a case study, we inject photometric systematics into our mocks as a function of right ascension (R.A.) and declination (Dec.), using a weight:

$$w_{\rm sys}(\text{R.A.}, \text{Dec.}) = 1 + 0.2 \sin\left(\frac{2\pi}{10(\text{deg})}\text{R.A.}(\text{deg})\right) \sin\left(\frac{2\pi}{5(\text{deg})}\text{Dec.}(\text{deg})\right). \tag{4.193}$$

The amplitude of the systematics ($\pm 20\%$) is very large compared to the typical requirements of a target selection (e.g. $\pm 7.5\%$ in Prakash et al. 2016). $w_{\rm sys}$ has a drastic impact on power spectrum measurements as can be seen on Figure 4.13 (left column, orange and blue curves). A standard RSD analysis, using scales $0.01\,h\,\text{Mpc}^{-1} \lesssim k \lesssim 0.15\,h\,\text{Mpc}^{-1}$ would be impossible. We will see that these systematics can be strongly reduced using a similar procedure as in Section 4.3.6.

4.3.7.2 Angular integral constraint

We suggest to remove the contaminated modes by weighting randoms from the synthetic catalogue by $\sum_{i\in\text{pixel}} w_{g,i} / \sum_{i\in\text{pixel}} w_{s,i}$ in pixels, thus nulling the density fluctuations in each pixel. We dub such a technique the *pixelated* scheme. We use a `HEALPix` [11] (Górski et al., 2005) map with nside $= 64$ (pixel area of $\simeq 0.84\,\text{deg}^2$), for which the contaminated and uncontaminated mocks look similar, as shown by the red and green curves in Figure 4.13 (left column). The *pixelated* scheme completely mitigates angular systematics in the quadrupole and hexadecapole. However, a difference close to a scale factor and statistically significant at small scales remains in the monopole.

We model the *pixelated* scheme by an angular integral constraint. A similar idea is developed in Burden et al. (2017) to mitigate the impact of the DESI fibre assignment.

Formulae are directly deduced from Section 4.3.3 by replacing $\epsilon_{\rm ic}(\mathbf{x}, \mathbf{y})$ by $\epsilon_{\rm ang}(\mathbf{x}, \mathbf{y})$, being non-zero if $\mathbf{x}$ and $\mathbf{y}$ lie within the same pixel. Contrary to the radial IC case (*binned* or *shuffled* schemes), for the angular IC (*pixelated* scheme) one has access to an estimate of the survey selection function accounting for known systematics (weighted synthetic catalogue), independent of the observed data. Thus, in the following, window function calculations are based on the true survey selection function (without accounting for angular variations of Eq. (4.193)), as in the *baseline* case. The angular IC correction is calculated from the same window functions for both uncontaminated and contaminated *pixelated* mocks, though in the latter case the normalisation of the survey selection function would be inferred from contaminated data in a practical analysis. However, as introduced systematics are purely angular and independent of the actual clustering, the normalisations estimated with Eq. (4.68) ($n_{s,i} = n(z_i)$, as explained in Section 4.3.6.1) are statistically equivalent.

The angular integral constraint has a very large impact on the quadrupole and hexadecapole shot noise, as can be seen in Figure 4.14.

In all this section, we use the same model and fitting k-range as in Section 4.3.6. Applying the *pixelated* scheme to uncontaminated mocks, cosmological parameters are well recovered when modelling the angular integral constraint: as shown in Figure 4.15 and Table 4.2, differences with the *baseline* case (same as in Section 4.3.6) are within the uncertainty on the mean of the mocks (column 2 in Table 4.2). The error bar on $f\sigma_8$ increases by 22%.

The magnitude of angular systematics makes it impossible to perform any relevant standard cosmological fit: we measure biases of $\Delta\alpha_\parallel = 0.043$, $\Delta\alpha_\perp = -0.037$ and $\Delta f\sigma_8 = -0.17$. On the

11. http://healpix.jpl.nasa.gov/

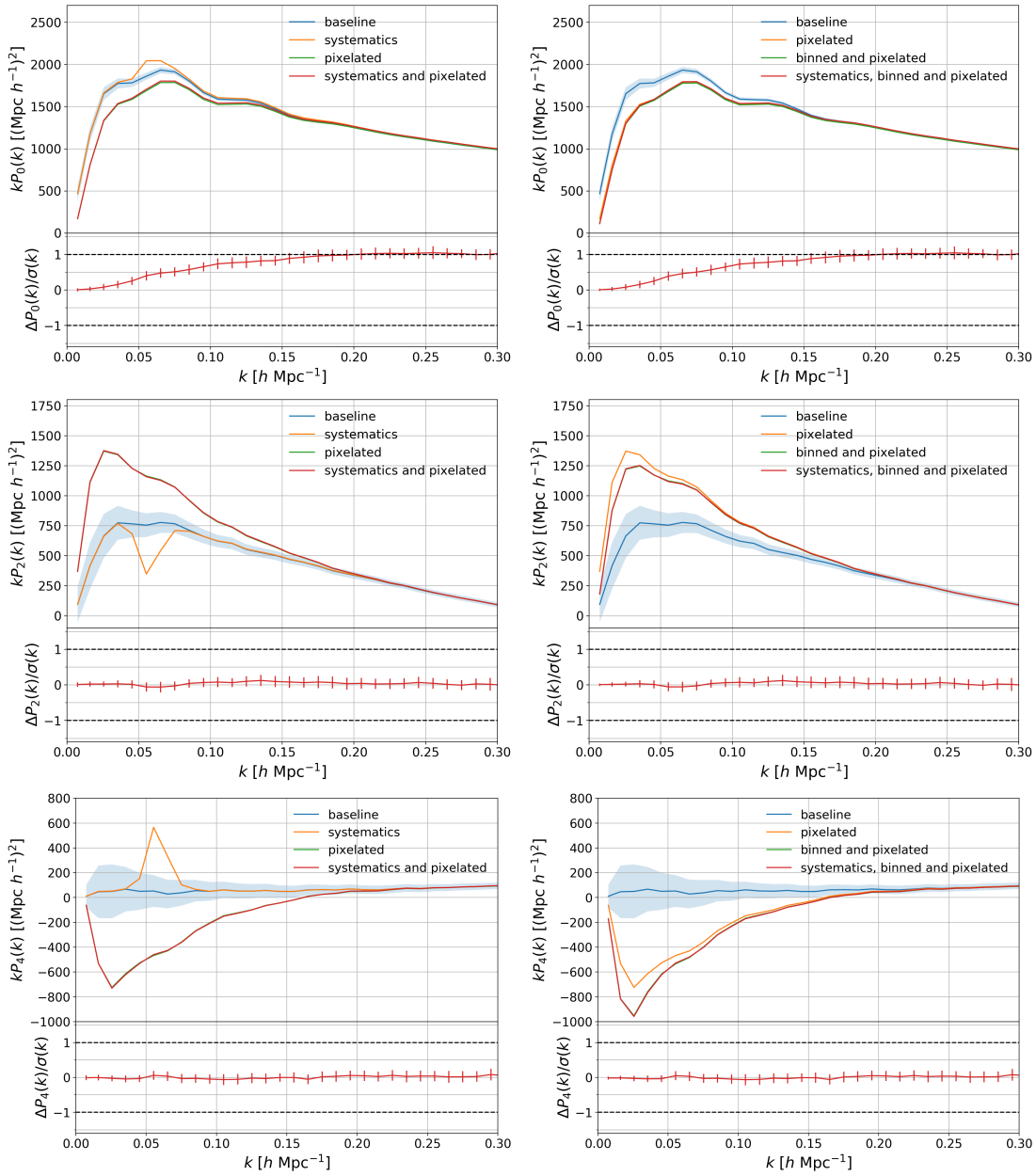

Figure 4.13 – Left column, top panels: power spectrum multipoles (top: monopole, middle: quadrupole, bottom: hexadecapole) measured from the 84 N-series mocks: *baseline* measurements (blue), with angular systematics (orange, Eq. (4.193)), using the *pixelated* scheme applied to uncontaminated (green) and contaminated (red) mocks (see text for details). Left column, bottom panels: difference of the *pixelated* scheme with and without angular systematics, with the standard deviation of the difference given by the error bars, normalised by the standard deviation of the mocks. Right column, top panels: *baseline* (blue), *pixelated* scheme (orange, same as green in the left column), *binned* and *pixelated* schemes applied to uncontaminated (green) and contaminated (red) mocks. Right column, bottom panels: normalised difference of the *binned* and *pixelated* scheme with and without angular systematics. In both columns, the blue shaded area represents the standard deviation of the mocks.

contrary, the cosmological analysis remains possible when applying the *pixelated* scheme to the contaminated mocks and the angular integral constraint in the model (column 3 in Table 4.2). Scaling parameters are recovered within the statistical uncertainty on the mean of the mocks,

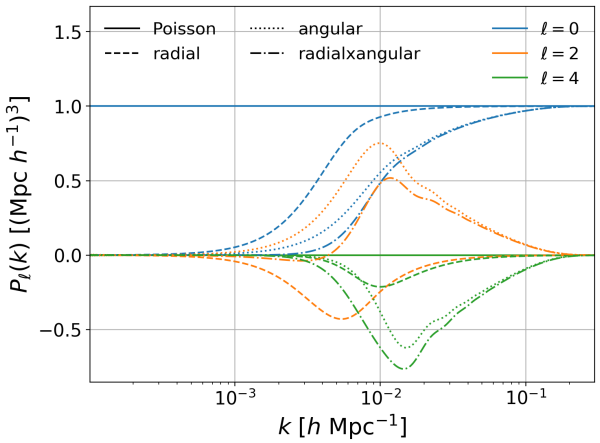

Figure 4.14 – Normalised shot noise with the radial, angular, and combined (radial x angular) integral constraints (blue: monopole, orange: quadrupole, green: hexadecapole).

and a bias of $\simeq 20\%$ of the error on a single realisation can be seen on $f\sigma_8$, which will be further discussed in Section 4.3.7.4. Note that a lower bias on the growth rate would be expected with smaller, more realistic angular systematics.

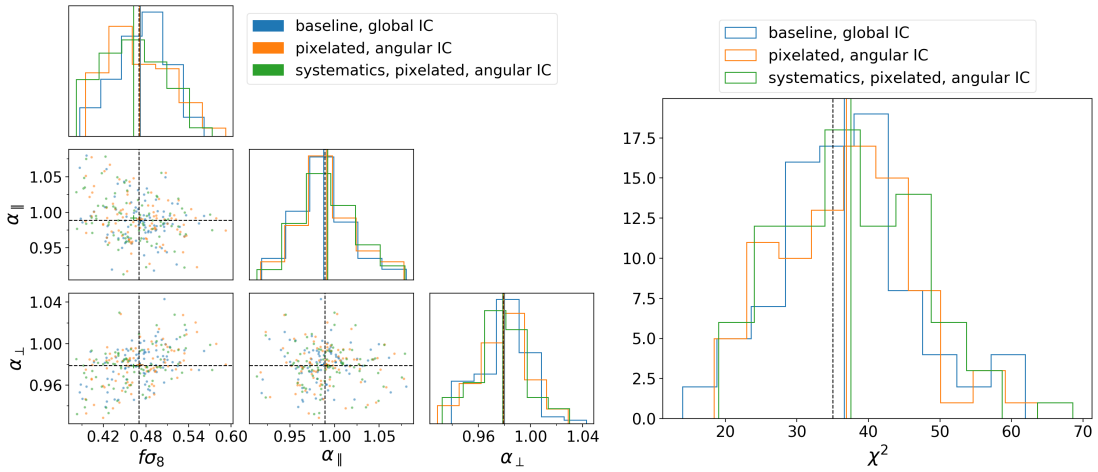

Figure 4.15 – Left: distributions of the cosmological parameters $f\sigma_8$, $\alpha_\parallel$, $\alpha_\perp$ measured on the 84 N-series mocks. The *baseline* (blue) uses the true selection function of the whole CMASS footprint. In orange, the *pixelated* scheme is applied on the mocks and the angular integral constraint is used in the model. In green, angular systematics (see Eq. (4.193)) are added onto the mocks. Right: the corresponding χ^2 distributions.

4.3.7.3 Combining radial and angular integral constraints

One would probably like to combine the radial and angular integral constraints, to account for both unknown radial and angular selection functions. A prerequisite to applying the radial integral constraint is that the redshift distribution does not depend on the angular position on the sky in each chunk of the survey. Thus, both radial and angular integral constraints can be imposed at the same time since radial and angular selection functions are independent.

Figure 4.13 (right column) shows the effects of the combined *binned* and *pixelated* schemes on N-series mocks. They add up in the hexadecapole, and partially cancel in the quadrupole.

	Baseline global IC	Pixelated angular IC	Systematics, pixelated angular IC	Expected
$\alpha_\parallel$	0.988 ± 0.031	0.991 ± 0.032	0.992 ± 0.032	0.989
$\alpha_\perp$	0.980 ± 0.018	0.979 ± 0.020	0.979 ± 0.020	0.979
$f\sigma_8$	0.472 ± 0.037	0.471 ± 0.045	0.463 ± 0.044	0.470
χ^2	36.6 ± 9.7	36.9 ± 9.3	37.5 ± 9.7	ndof = 35

Table 4.2 – Mean and standard deviation of the cosmological parameters fitted on the 84 N-series mocks, corresponding to Figure 4.15. Error bars should be divided by $\sqrt{84} \sim 10$ to obtain errors on the mean of the mocks.

This is expected as $\mathcal{L}_2$ is negative around $\mu = 0$ (where the angular integral constraint removes signal) and positive around $\mu = 1$ (where the radial integral constraint plays up). As with the *pixelated* scheme alone (Figure 4.13, left column), angular systematics are mitigated in the quadrupole and hexadecapole, but a multiplicative effect remains in the monopole.

Let us model the radial and angular integral constraints in a row:

$$\delta^{\rm cic}(\mathbf{r}) = W(\mathbf{r}) \left\{ \delta(\mathbf{r}) - \int d^3x W_{\rm rad}(\mathbf{x}) \epsilon_{\rm rad}(\mathbf{r},\mathbf{x}) \delta(\mathbf{x}) - \int d^3x W_{\rm ang}(\mathbf{x}) \epsilon_{\rm ang}(\mathbf{r},\mathbf{x}) \delta(\mathbf{x}) \right. \\ \left. + \int d^3x W_{\rm ang}(\mathbf{x}) \epsilon_{\rm ang}(\mathbf{r},\mathbf{x}) \int d^3y W_{\rm rad}(\mathbf{y}) \epsilon_{\rm rad}(\mathbf{x},\mathbf{y}) \delta(\mathbf{y}) \right\}. \tag{4.194}$$

Since the radial and angular parts of $W(\mathbf{r})$ are independent, the last term is just the integral of the density contrast over the whole (chunk) footprint, i.e. the global integral constraint. Thus, the two integral constraints commute. Then, building up the correlation function multipoles of the observed density fluctuations (4.194), we find 16 terms:

$$\xi_\ell^{\rm cic}(s) = \xi_\ell^{\rm c}(s) - IC_\ell^{\delta,\rm rad}(s) - IC_\ell^{\rm rad,\delta}(s) + IC_\ell^{\rm rad,rad}(s) \tag{4.195}$$
$$- IC_\ell^{\delta,\rm ang}(s) - IC_\ell^{\rm ang,\delta}(s) + IC_\ell^{\rm ang,ang}(s) \tag{4.196}$$
$$+ IC_\ell^{\delta,\rm glo}(s) + IC_\ell^{\rm glo,\delta}(s) + IC_\ell^{\rm glo,glo}(s) \tag{4.197}$$
$$- IC_\ell^{\rm glo,rad}(s) - IC_\ell^{\rm rad,glo}(s) - IC_\ell^{\rm glo,ang}(s) - IC_\ell^{\rm ang,glo}(s)$$
$$+ IC_\ell^{\rm rad,ang}(s) + IC_\ell^{\rm ang,rad}(s). \tag{4.198}$$

Terms (4.195), (4.196) and (4.197) correspond to the radial, angular and global integral constraints, while terms (4.198) are the cross-integral constraints, given by a formula similar to Eq. (4.207), with:

$$\mathcal{W}_{\ell p}^{i,j,(n)}(s,\Delta) = \frac{(2\ell+1)(2p+1)}{(4\pi)^2 A} \int d\Omega_s \int d\Omega_\Delta \int d^3y W(\mathbf{y}) W(\mathbf{y}-\mathbf{s}) \\ \int d^3x x^{-n} W_{\rm ic}(\mathbf{x}) W_{\rm ic}(\mathbf{x}-\mathbf{\Delta}) \mathcal{L}_\ell(\hat{\mathbf{y}}\cdot\hat{\mathbf{s}}) \mathcal{L}_p(\hat{\mathbf{x}}\cdot\hat{\mathbf{\Delta}}) \epsilon_i(\mathbf{y},\mathbf{x}) \epsilon_j(\mathbf{y}-\mathbf{s},\mathbf{x}-\mathbf{\Delta}). \tag{4.199}$$

We recall that to speed up practical calculations, since $IC_\ell^{i,j}(s)$ are linear in $\xi_p^{{\rm ep},(n)}$, all $\mathcal{W}_{\ell p}^{i,j,(n)}$ can be directly summed over to yield a single $IC_\ell(s)$ term.

As in Section 4.3.6.2, window function calculations are based on the survey selection function radially-tuned on one realisation of mock data using the *binned* scheme. As in Section 4.3.7.2, the same window functions are used to calculate the angular IC correction for both uncontaminated and contaminated *pixelated* mocks.

	Baseline global IC	Binned, pixelated radial x angular IC	Systematics, binned, pixelated radial x angular IC	Expected
$\alpha_\parallel$	0.988 ± 0.031	0.988 ± 0.032	0.990 ± 0.032	0.989
$\alpha_\perp$	0.980 ± 0.018	0.981 ± 0.020	0.980 ± 0.020	0.979
$f\sigma_8$	0.472 ± 0.037	0.475 ± 0.044	0.467 ± 0.043	0.470
χ^2	36.6 ± 9.7	36.5 ± 9.3	37.0 ± 9.6	ndof $= 35$

Table 4.3 – Mean and standard deviation of the cosmological parameters fitted on the 84 N-series mocks, corresponding to Figure 4.16. Error bars should be divided by $\sqrt{84} \sim 10$ to obtain errors on the mean of the mocks.

As shown in Figure 4.16 and Table 4.3, the combined (radial x angular) integral constraint accounts well for the *binned* and *pixelated* schemes in the mocks: cosmological parameters are recovered within the uncertainty on the mean of the mocks (column 2 in Table 4.3). Cosmological fits of contaminated mocks are not further degraded by adding the radial integral constraint on top of the angular one. As in Section 4.3.7.2, a bias of 12% to 20% of the error on a single realisation can be seen on $f\sigma_8$ (comparing column 3 to columns 2 and 1).

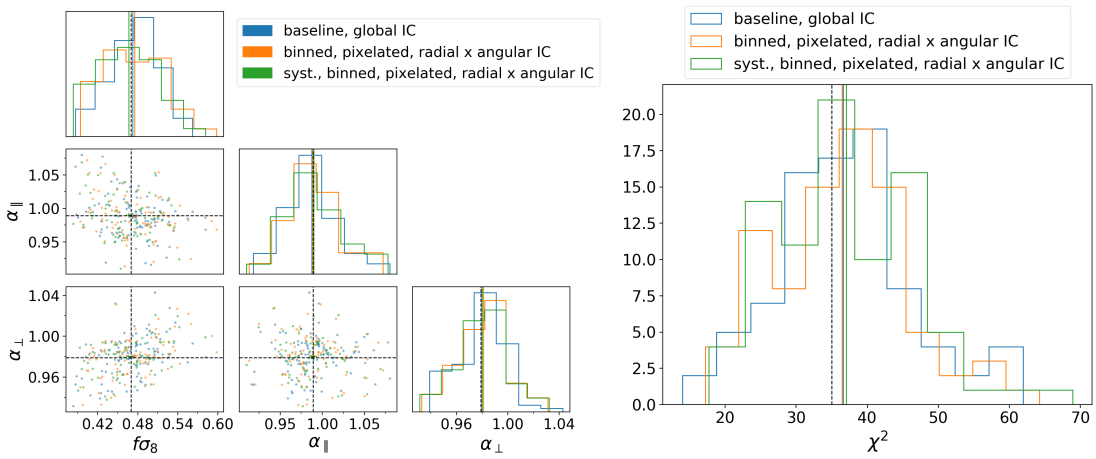

Figure 4.16 – Same as Figure 4.15, with the radial and angular integral constraints combined.

4.3.7.4 Caveat: multiplicative systematics

We emphasise that the angular (pixel) integral constraint can only mitigate the additive part of angular systematics. Let us call $c(\mathbf{r})$ the contamination signal, assumed constant over a pixel. First, let us suppose c to be purely additive. Applying the *pixelated* scheme, we would measure the power spectrum of the density fluctuations:

$$\delta^{\rm cic}(\mathbf{r}) = W(\mathbf{r})\left\{1+\delta(\mathbf{r})+c(\mathbf{r})\right\} - W(\mathbf{r})\frac{\int d^3x W(\mathbf{x})\left\{1+\delta(\mathbf{x})+c(\mathbf{x})\right\}\epsilon_{\rm ang}(\mathbf{r},\mathbf{x})}{\int d^3x W(\mathbf{x})\epsilon_{\rm ang}(\mathbf{r},\mathbf{x})}. \tag{4.200}$$

Then, as $c(\mathbf{r})$ is constant over a pixel, $c(\mathbf{x})\epsilon_{\rm ang}(\mathbf{r},\mathbf{x}) = c(\mathbf{r})\epsilon_{\rm ang}(\mathbf{r},\mathbf{x})$ and:

$$\delta^{\rm cic}(\mathbf{r}) = W(\mathbf{r})\left\{\delta(\mathbf{r}) - \int d^3x W_{\rm ang}(\mathbf{x})\delta(\mathbf{x})\epsilon_{\rm ang}(\mathbf{r},\mathbf{x})\right\}. \tag{4.201}$$

As expected, $c(\mathbf{r})$ disappears from the analysis.

Now, let us consider $c(\mathbf{r})$ to be multiplicative (as implemented in our contamination model (4.193)):

$$\delta^{\rm cic}(\mathbf{r}) = W(\mathbf{r})c(\mathbf{r})\left\{1+\delta(\mathbf{r})\right\} - W(\mathbf{r})\frac{\int d^3x W(\mathbf{x})c(\mathbf{x})\left\{1+\delta(\mathbf{x})\right\}\epsilon_{\rm ang}(\mathbf{r},\mathbf{x})}{\int d^3x W(\mathbf{x})\epsilon_{\rm ang}(\mathbf{r},\mathbf{x})} \tag{4.202}$$

$$= W(\mathbf{r})c(\mathbf{r})\left\{\delta(\mathbf{r}) - \int d^3x W_{\rm ang}(\mathbf{x})\delta(\mathbf{x})\epsilon_{\rm ang}(\mathbf{r},\mathbf{x})\right\}, \tag{4.203}$$

i.e. $c(\mathbf{r})$ multiplies the selection function $W(\mathbf{r})$. However, by definition, $c(\mathbf{r})$ is unknow and therefore cannot be taken into account in $W(\mathbf{r})$. The resulting multiplicative systematics can explain the observed bias on $f\sigma_8$ (compensating for a higher $b_1\sigma_8$). A possible way to alleviate this effect may consist in estimating the angular survey selection function (including $c(\hat{\mathbf{r}})$) with the *pixelated* scheme. This would induce a bias which can be estimated using the method presented in Section 4.3.5.1. We report the reader to (e.g. Shafer and Huterer, 2015) for a fully coherent treatment of unknown multiplicative systematics in the case of the angular power spectrum.

We showed that one could mitigate potential, unknown, systematics at the price of a small increase of statistical errors by applying an angular integral constraint in the data. Such a technique may be helpful to reduce the impact of poorly understood angular target density variations with photometric conditions, or can be considered as a regular consistency test to check for remaining angular systematics.

4.3.8 FKP weights: a potential source of systematics

In Section 4.3.6.1 we mentioned a potential bias due to FKP weights, which we apply to both data and randoms. These weights are given by:

$$w_{\rm FKP} = \frac{1}{1+n(z)P_0}, \tag{4.204}$$

using $P_0 = 20\,000\,{\rm Mpc}^3\,h^{-3}$. FKP weights hence require an estimation of the true redshift density $n(z)$ in absence of clustering. If $n(z)$ is computed from the data itself, using narrow bins in z, FKP weights may bias clustering measurements since they overweight (downweight) underdensities (overdensities).

As mentioned in Section 4.3.6.1, the redshift density $n(z)$ is computed in bins $\Delta z = 0.005$. Figure 4.17 shows power spectrum measurements obtained with the *binned* scheme, with different $n(z)$ estimations. Measuring $n(z)$ from the full CMASS sample induces a negligible bias with respect to taking the true $n(z)$. However, measuring $n(z)$ separately in each of the 6 chunks dividing the CMASS footprint, as would be natural to do, leads to an additional bias on all scales of all multipoles. Though the density $n(z)$ also enters the power spectrum normalisation (see Eq. (4.68)), the main effect comes from the smoothing of the clustering along the line of sight, as can be seen from the loss of power in the quadrupole and hexadecapole. This effect is multiplicative and would be difficult to take into account.

A simple way to prevent FKP weights from biasing clustering measurements would be to fit a simple spline to the redshift distribution in wide redshift bins to reduce the correlation between $w_{\rm FKP}$ and the density field. As stated in Section 4.3.6.1, we choose for simplicity to estimate $n(z)$ from each data realisation, using the full CMASS sample (instead of the 6 different chunks), making the bias from $w_{\rm FKP}$ almost invisible.

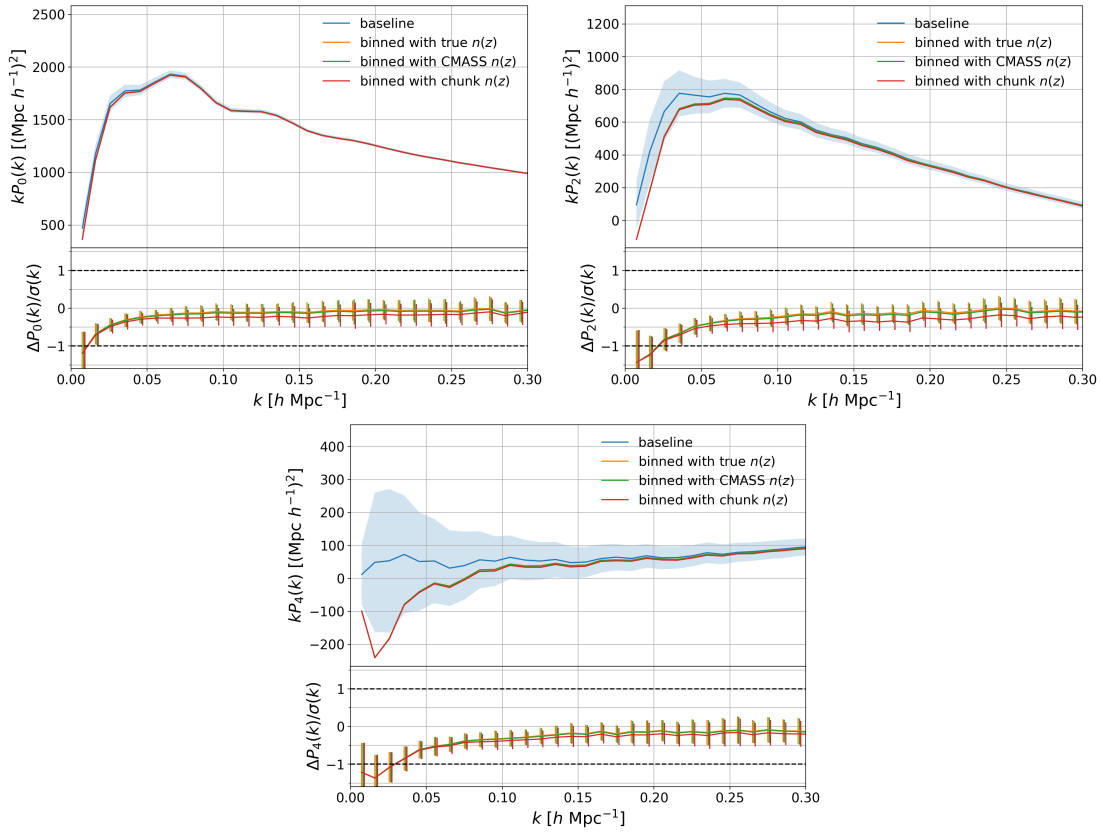

Figure 4.17 – Top panels: power spectrum multipoles (upper left: monopole, upper right: quadrupole, bottom: hexadecapole) obtained using the *binned* scheme and different $n(z)$ estimations. In orange, we use the true density $n(z)$ from the radial selection function (rescaled for each realisation of the mocks, as explained at the end of Section 4.3.6.1). In green, $n(z)$ is estimated from each full CMASS mock realisation. In red, $n(z)$ is measured from each mock realisation, in the 6 chunks separately. We recall the *baseline* in blue, obtained with the true selection function. The blue shaded area represents the standard deviation of the mocks. Bottom panels: difference of the *binned* scheme with the different (true, full CMASS, chunk) $n(z)$ estimations to the *baseline*, with the standard deviation of the difference given by the error bars, normalised by the standard deviation of the mocks.

4.3.9 Integral constraint for the Landy-Szalay estimator

In Section 4.3.3 we derived the integral constraint effect on the Yamamoto power spectrum estimator. However, integral constraints also impact configuration space measurements: Eq. (4.179) expresses the impact of integral constraints on the *window-convolved* correlation function.

A commonly used estimator to compute the 2-point correlation function is the Landy-Szalay estimator (Landy and Szalay, 1993):

$$\hat{\xi}(s,\mu) = \frac{DD(s,\mu) - DR(s,\mu) - RD(s,\mu) + RR(s,\mu)}{RR(s,\mu)} \tag{4.205}$$

where $DD(s,\mu)$, $DR(s,\mu)$ and $RR(s,\mu)$ are the data - data, data - random and random - random pairs counts. We take them to be normalised by the weighted number of pairs, i.e. $\left(\sum w_{g,i}\right)^2 - \sum w_{g,i}^2$, $\left(\sum w_{g,i}\right)\left(\sum w_{s,j}\right)$ and $\left(\sum w_{s,i}\right)^2 - \sum w_{s,i}^2$, respectively.

In fact, the multipoles of the expectation value of the numerator of Eq. (4.205) are simply given by Eq. (4.179). The denominator of Eq. (4.205) aims at removing the window function ef-

fect from the numerator, i.e. its multipoles simply correspond to the window function multipoles $\mathcal{W}_\ell^{\delta,\delta,(0)}(s)$ (4.138) (up to some normalisation). Therefore:

$$\hat{\xi}_\ell^{\rm ic}(s) = \xi_\ell(s) - IC_{\ell,\rm now}^{\delta,\rm ic}(s) - IC_{\ell,\rm now}^{\rm ic,\delta}(s) + IC_{\ell,\rm now}^{\rm ic,ic}(s) \tag{4.206}$$

with:

$$IC_{\ell,\rm now}^{i,j}(s) = \int \Delta^2 d\Delta \sum_{p,n} \frac{4\pi}{2p+1} \Delta^n \xi_p^{\mathbf{d},(n)}(\Delta) \mathcal{W}_{\ell p,\rm now}^{i,j,(n)}(s,\Delta) \tag{4.207}$$

and where:

$$\mathcal{W}_{\ell p,\rm now}^{i,j,(n)}(s,\Delta) = \frac{2\ell+1}{2} \int_{-1}^{1} d\mu \frac{\sum_q \mathcal{W}_{qp}^{i,j,(n)}(s,\Delta)\mathcal{L}_q(\mu)}{\sum_q \mathcal{W}_q^{\delta,\delta,(0)}(s)\mathcal{L}_q(\mu)} \mathcal{L}_\ell(\mu). \tag{4.208}$$

Note that the same line of sight as in the Landy-Szalay estimator (usually the midpoint line of sight) should be used for the cosine angle with respect to the separation $\mathbf{s}$ in $\mathcal{W}_{qp}^{i,j,(n)}(s,\Delta)$. The same line of sight definition is a natural choice for the cosine angle with respect to the separation $\mathbf{\Delta}$ and hence for $\xi_p^{\mathbf{d},(n)}(\Delta)$ as explained in the end of Section 4.3.3. Similarly, we obtain the shot noise contribution from Eq. (4.188) following:

$$SN_{\ell,\rm now}^{i,j}(s) = \frac{2\ell+1}{2} \int_{-1}^{1} d\mu \frac{\sum_q SN_q^{i,j}(s)\mathcal{L}_q(\mu)}{\sum_q \mathcal{W}_q^{\delta,\delta,(0)}(s)\mathcal{L}_q(\mu)} \mathcal{L}_\ell(\mu). \tag{4.209}$$

The integral constraint correction specified above was successfully used for the clustering analysis of the eBOSS ELG sample in configuration space (Tamone et al., 2020; Hou et al., 2020).

4.3.10 Conclusions

We revisited the notion of integral constraints which we showed to be useful in clustering analyses to account for biases related to calibrating the survey selection function partly on data.

We presented a general formalism to account for integral constraints that completes the existing models for the global integral constraint (Peacock and Nicholson, 1991; Beutler et al., 2014; Wilson et al., 2017). We discussed the window function normalisation and the shot noise contribution to the integral constraints and included wide-angle effects following (Beutler et al., 2019). Our formalism was used to model the radial integral constraint effect induced by inferring the radial selection function from actual data.

We indeed noticed that the common practice of drawing random redshifts based on data may significantly bias large scale clustering measurements. This effect can be particularly large compared to the window function effect if the survey is composed of several patches whose radial selection functions must be treated separately. In particular, its impact is likely to be significant in analyses focusing on large scales, e.g. dedicated to primordial non-Gaussianity (Ross et al., 2013; Castorina et al., 2019).

Applying our modelling of the radial integral constraint to a RSD analysis, the bias induced on cosmological parameters was shown to be successfully reduced. Though more statistics would be required to push our validation further, the potentially remaining bias is already below the expected statistical uncertainty of future galaxy surveys.

As a further application, we showed that we could similarly apply an angular integral constraint to help mitigating unknown angular systematics. This angular integral constraint can be combined to the radial one, as required if the radial selection function is also estimated from the data. Though ony additive systematics can be fully accounted for, our scheme performs well enough with multiplicative systematics on a BOSS CMASS-like survey.

We noted that a potential bias can emerge when the selection function used in the window function calculations for integral constraint corrections is estimated from the data itself. We suggested a way to estimate this bias, which remained very subdominant in our analysis. A workaround for future analyses may consist in predicting the radial selection function from first principles (e.g. from the luminosity function), without relying on the observed data, while marginalising on possible unknowns in the cosmological fits. Another idea would be to directly project out radial modes from the estimation of the power spectrum by assigning them infinite variance (e.g. Elsner et al., 2016).

Our modelling of the radial integral constraint was successfully utilised in ELG (de Mattia et al., 2020; Tamone et al., 2020) and QSO (Neveux et al., 2020; Hou et al., 2020) Fourier and configuration space clustering analyses. The next chapter is devoted to the clustering analysis of the eBOSS ELG sample.

Bibliography

Abramo, L. R., L. F. Secco, and A. Loureiro (2016). Fourier analysis of multitracer cosmological surveys. MNRAS *455*(4), 3871–3889. arXiv:1505.04106.

Alam, S., M. Ata, S. Bailey, F. Beutler, D. Bizyaev, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. MNRAS *470*, 2617–2652. arXiv:1607.03155.

Alcock, C. and B. Paczynski (1979). An evolution free test for non-zero cosmological constant. Nature *281*, 358.

Ata, M., F. Baumgarten, J. Bautista, F. Beutler, D. Bizyaev, et al. (2018). The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2. MNRAS *473*, 4773–4794. arXiv:1705.06373.

Ballinger, W. E., J. A. Peacock, and A. F. Heavens (1996). Measuring the cosmological constant with redshift surveys. MNRAS *282*, 877. arXiv:astro-ph/9605017.

Beutler, F., E. Castorina, and P. Zhang (2019). Interpreting measurements of the anisotropic galaxy power spectrum. J. Cosmology Astropart. Phys. *2019*(3), 040. arXiv:1810.05051.

Beutler, F., S. Saito, H.-J. Seo, J. Brinkmann, K. S. Dawson, et al. (2014). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: testing gravity with redshift space distortions using the power spectrum multipoles. MNRAS *443*(2), 1065–1089. arXiv:1312.4611.

Beutler, F., H.-J. Seo, S. Saito, C.-H. Chuang, A. J. Cuesta, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: anisotropic galaxy clustering in Fourier space. MNRAS *466*, 2242–2260. arXiv:1607.03150.

Bianchi, D., H. Gil-Marín, R. Ruggeri, and W. J. Percival (2015). Measuring line-of-sight-dependent Fourier-space clustering using FFTs. MNRAS *453*(1), L11–L15. arXiv:1505.05341.

Blake, C., S. Brough, M. Colless, W. Couch, S. Croom, et al. (2010). The WiggleZ Dark Energy Survey: the selection function and z = 0.6 galaxy power spectrum. MNRAS *406*, 803–821. arXiv:1003.5721.

Burden, A., N. Padmanabhan, R. N. Cahn, M. J. White, and L. Samushia (2017). Mitigating the impact of the DESI fiber assignment on galaxy clustering. *Journal of Cosmology and Astro-Particle Physics 2017*, 001. arXiv:1611.04635.

Castorina, E., N. Hand, U. Seljak, F. Beutler, C.-H. Chuang, et al. (2019). Redshift-weighted constraints on primordial non-Gaussianity from the clustering of the eBOSS DR14 quasars in Fourier space. J. Cosmology Astropart. Phys. *2019*(9), 010. arXiv:1904.08859.

Castorina, E. and M. White (2018). Beyond the plane-parallel approximation for redshift surveys. MNRAS *476*(4), 4403–4417. arXiv:1709.09730.

Colombi, S., A. Jaffe, D. Novikov, and C. Pichon (2009). Accurate estimators of power spectra in N-body simulations. MNRAS *393*(2), 511–526. arXiv:astro-ph/0811.0313.

d'Amico, G., J. Gleyzes, N. Kokron, K. Markovic, L. Senatore, et al. (2020). The cosmological analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure. J. Cosmology Astropart. Phys. *2020*(5), 005. arXiv:1909.05271.

de Mattia, A. and V. Ruhlmann-Kleider (2019). Integral constraints in spectroscopic surveys. J. Cosmology Astropart. Phys. *2019*(8), 036. arXiv:1904.08851.

de Mattia, A., V. Ruhlmann-Kleider, A. Raichoor, A. J. Ross, A. Tamone, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the emission line galaxy sample from the anisotropic power spectrum between redshift 0.6 and 1.1. *arXiv e-prints*, arXiv:2007.09008. arXiv:2007.09008.

Elsner, F., B. Leistedt, and H. V. Peiris (2016). Unbiased methods for removing systematics from galaxy clustering measurements. MNRAS *456*(2), 2095–2104. arXiv:1509.08933.

Feldman, H. A., N. Kaiser, and J. A. Peacock (1994). Power-Spectrum Analysis of Three-dimensional Redshift Surveys. ApJ *426*, 23. arXiv:astro-ph/9304022.

Fisher, R. (1935). *The design of experiments. 1935.* Edinburgh: Oliver and Boyd.

Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman (2013). emcee: The mcmc hammer. *PASP 125*, 306–312. 1202.3665.

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

Gil-Marín, H., W. J. Percival, J. R. Brownstein, C.-H. Chuang, J. N. Grieb, et al. (2016). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the LOS-dependent power spectrum of DR12 BOSS galaxies. MNRAS *460*, 4188–4209. arXiv:1509.06386.

Górski, K. M., E. Hivon, A. J. Banday, B. D. Wand elt, F. K. Hansen, et al. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ *622*(2), 759–771. arXiv:astro-ph/0409513.

Hamilton, A. J. S. (2000). Uncorrelated modes of the non-linear power spectrum. MNRAS *312*(2), 257–284. arXiv:astro-ph/9905191.

Hand, N., Y. Feng, F. Beutler, Y. Li, C. Modi, et al. (2018). nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure. AJ *156*, 160. arXiv:1712.05834.

Hand, N., Y. Li, Z. Slepian, and U. Seljak (2017). An optimal FFT-based anisotropic power spectrum estimator. J. Cosmology Astropart. Phys. *2017*(7), 002. arXiv:1704.02357.

Hou, J., A. G. Sánchez, A. J. Ross, A. Smith, R. Neveux, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from anisotropic clustering analysis of the Quasar Sample in configuration space between redshift 0.8 and 2.2. *arXiv e-prints*, arXiv:2007.08998. arXiv:2007.08998.

iminuit team. iminuit – a python interface to minuit. https://github.com/iminuit/iminuit. Accessed: 2018-03-05.

Ivanov, M. M., M. Simonović, and M. Zaldarriaga (2020). Cosmological parameters from the BOSS galaxy power spectrum. J. Cosmology Astropart. Phys. *2020*(5), 042. arXiv:1909.05277.

James, F. and M. Roos (1975). Minuit – a system for function minimization and analysis of the parameter errors and correlations. *Computer Physics Communications 10*, 343–367.

Jing, Y. P. (2005). Correcting for the Alias Effect When Measuring the Power Spectrum Using a Fast Fourier Transform. ApJ *620*(2), 559–563. arXiv:astro-ph/0409240.

Kitaura, F.-S., S. Rodríguez-Torres, C.-H. Chuang, C. Zhao, F. Prada, et al. (2016). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release. MNRAS *456*, 4156–4173. arXiv:1509.06400.

Klypin, A., G. Yepes, S. Gottlöber, F. Prada, and S. Heß (2016). MultiDark simulations: the story of dark matter halo concentrations and density profiles. MNRAS *457*, 4340–4359. arXiv:1411.4001.

Landy, S. D. and A. S. Szalay (1993). Bias and Variance of Angular Correlation Functions. ApJ *412*, 64.

Laureijs, R. (2009). Euclid Assessment Study Report for the ESA Cosmic Visions. *arXiv e-prints*, arXiv:0912.0914. arXiv:0912.0914.

Neveux, R., E. Burtin, A. de Mattia, A. Smith, A. J. Ross, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from the anisotropic power spectrum of the Quasar sample between redshift 0.8 and 2.2. *arXiv e-prints*, arXiv:2007.08999. arXiv:2007.08999.

Peacock, J. A. and D. Nicholson (1991). The large-scale clustering of radio galaxies. MNRAS *253*, 307–319.

Prakash, A., T. C. Licquia, J. A. Newman, A. J. Ross, A. D. Myers, et al. (2016). The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Luminous Red Galaxy Target Selection. *The Astrophysical Journal Supplement Series 224*(2), 34. arXiv:1508.04478.

Raichoor, A., J. Comparat, T. Delubac, J. P. Kneib, C. Yèche, et al. (2017). The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: final emission line galaxy target selection. MNRAS *471*(4), 3955–3973. arXiv:1704.00338.

Reid, B., S. Ho, N. Padmanabhan, W. J. Percival, J. Tinker, et al. (2016). SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues. MNRAS *455*, 1553–1573. arXiv:1509.06529.

Reimberg, P., F. Bernardeau, and C. Pitrou (2016). Redshift-space distortions with wide angular separations. J. Cosmology Astropart. Phys. *2016*(1), 048. arXiv:1506.06596.

Ross, A. J., W. J. Percival, A. Carnero, G.-b. Zhao, M. Manera, et al. (2013). The clustering of galaxies in the SDSS-III DR9 Baryon Oscillation Spectroscopic Survey: constraints on primordial non-Gaussianity. MNRAS *428*, 1116–1127. arXiv:1208.1491.

Ross, A. J., W. J. Percival, A. G. Sánchez, L. Samushia, S. Ho, et al. (2012). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: analysis of potential systematics. MNRAS *424*, 564–590. arXiv:1203.6499.

Ruggeri, R., W. J. Percival, E.-M. Mueller, H. Gil-Marín, F. Zhu, et al. (2019). The extended Baryon Oscillation Spectroscopic Survey: testing a new approach to measure the evolution of the structure growth. MNRAS *484*(3), 4100–4112. arXiv:1712.03997.

Samushia, L., W. J. Percival, and A. Raccanelli (2012). Interpreting large-scale redshift-space distortion measurements. MNRAS *420*(3), 2102–2119. arXiv:1102.1014.

Scoccimarro, R. (2015). Fast estimators for redshift-space clustering. Phys. Rev. D *92*(8), 083532. arXiv:1506.02729.

Sefusatti, E., M. Crocce, R. Scoccimarro, and H. M. P. Couchman (2016). Accurate estimators of correlation functions in Fourier space. MNRAS *460*(4), 3624–3636. arXiv:1512.07295.

Shafer, D. L. and D. Huterer (2015). Multiplicative errors in the galaxy power spectrum: self-calibration of unknown photometric systematics for precision cosmology. MNRAS *447*, 2961–2969. arXiv:1410.0035.

Slepian, Z. and D. J. Eisenstein (2018). A practical computational method for the anisotropic redshift-space three-point correlation function. MNRAS *478*, 1468–1483. arXiv:1709.10150.

Springel, V. (2005). The cosmological simulation code GADGET-2. MNRAS *364*, 1105–1134. arXiv:astro-ph/0505010.

Szalay, A. S., T. Matsubara, and S. D. Landy (1998). Redshift-Space Distortions of the Correlation Function in Wide-Angle Galaxy Surveys. ApJ *498*(1), L1–L4. arXiv:astro-ph/9712007.

Tamone, A., A. Raichoor, C. Zhao, A. de Mattia, C. Gorgoni, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Growth rate of structure measurement from anisotropic clustering analysis in configuration space between redshift 0.6 and 1.1 for the Emission Line Galaxy sample. *arXiv e-prints*, arXiv:2007.09009. arXiv:2007.09009.

Tegmark, M., A. J. S. Hamilton, M. A. Strauss, M. S. Vogeley, and A. S. Szalay (1998). Measuring the Galaxy Power Spectrum with Future Redshift Surveys. ApJ *499*(2), 555–576. arXiv:astro-ph/9708020.

Tinker, J. (2016). in prep.

Vogeley, M. S. and A. S. Szalay (1996). Eigenmode Analysis of Galaxy Redshift Surveys. I. Theory and Methods. ApJ *465*, 34. arXiv:astro-ph/9601185.

White, M., J. L. Tinker, and C. K. McBride (2014). Mock galaxy catalogues using the quick particle mesh method. MNRAS *437*(3), 2594–2606. arXiv:1309.5532.

Wilson, M. J., J. A. Peacock, A. N. Taylor, and S. de la Torre (2017). Rapid modelling of the redshift-space power spectrum multipoles for a masked density field. MNRAS *464*(3), 3121–3130. arXiv:1511.07799.

Yamamoto, K., M. Nakamichi, A. Kamino, B. A. Bassett, and H. Nishioka (2006). A Measurement of the Quadrupole Power Spectrum in the Clustering of the 2dF QSO Survey. PASJ *58*, 93–102. arXiv:astro-ph/0505115.

Zarrouk, P., E. Burtin, H. Gil-Marín, A. J. Ross, R. Tojeiro, et al. (2018). The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: measurement of the growth rate of structure from the anisotropic correlation function between redshift 0.8 and 2.2. MNRAS *477*, 1639–1663. arXiv:1801.03062.

Clustering analysis of the eBOSS ELG sample

The first four chapters of this manuscript told the story of density fluctuations, in the primordial Universe until the drag epoch (Chapter 1), how they grew and became non-linear (Chapter 2) and how to measure them (Chapter 4) in galaxy surveys (Chapter 3). In this chapter we propose to apply the previous developments to the clustering analysis of the eBOSS ELG sample, which we publish in de Mattia et al. (2020).

Specifically, we perform a RSD analysis of this spectroscopic sample, using the model built in Chapter 2, describing the dark matter density field, redshift space distortions and galaxy bias, up to quasi non-linear scales. However, a clear signature can be robustly measured in the galaxy power spectrum without resorting to such a complex machinery: baryon acoustic oscillations, explained in Chapter 1. Indeed, their typical scale of 150 Mpc makes them quite robust to non-linear evolution of the density field, and their typical pattern to observational systematics. We thus also perform a BAO analysis, which we finally combine with the RSD analysis, to measure $f(z_{\rm eff})\sigma_8(z_{\rm eff})$, and ratios $D_{\rm H}(z_{\rm eff})/r_{\rm drag}$ and $D_{\rm M}(z_{\rm eff})/r_{\rm drag}$, at the effective redshift of the sample $z_{\rm eff} = 0.85$.

We start by presenting the analysis methodology in Section 5.1, then assess our RSD and BAO theoretical models in Section 5.2. The generation of the eBOSS DR16 ELG sample data clustering catalogues, including correction schemes for known systematic effects is described in Section 5.3, as well as approximate mocks used to test the robustness of our analysis pipeline with respect to observational systematics, as discussed in Section 5.4. Cosmological fits and their implications are presented in Section 5.5. These measurements are combined with configuration space results of Tamone et al. (2020) in Section 5.5.3. We conclude in Section 5.5.4.

5.1 Analysis methodology

We provide the practical details of the estimation of power spectrum multipoles in Section 5.1.1. The geometric corrections described in Section 4.2 and Section 4.3 entering the power spectrum RSD model are recapped in Section 5.1.2. We describe the BAO template used to model these oscillations in Section 5.1.3, and briefly review the technique of *reconstruction* to enhance the BAO feature in clustering measurements in Section 5.1.4. We detail the model parameters in Section 5.1.5, and how the model is compared to power spectrum measurements in Section 5.1.6. We finally detail the choice of effective redshift in Section 5.1.7.

5.1.1 Power spectrum estimation

In this section we detail our measurements of the power spectrum multipoles of the galaxy density field in a periodic box (used in Section 5.2) and within a real, sky-cut geometry (used in Section 5.3 and beyond).

5.1.1.1 Periodic box

We first calculate the density contrast:

$$\delta_g(\mathbf{r}) = \frac{n_g(\mathbf{r})}{\bar{n}_g} - 1 \tag{5.1}$$

where $n_g(\mathbf{r})$ is the galaxy density at comoving position $\mathbf{r}$, and $\bar{n}_g$ its average over the whole box of volume V. Taking the Fourier transform $\delta_g(\mathbf{k})$ of this field, power spectrum multipoles are

calculated as:

$$P_\ell(k_\mu) = \frac{2\ell+1}{VV_{k_\mu}} \int_{V_{k_\mu}} d^3k \delta_g(\mathbf{k})\delta_g(-\mathbf{k})\mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\eta}) - P_\ell^{\rm noise}(k) \tag{5.2}$$

$\mathcal{L}_\ell$ being the Legendre polynomial of order ℓ and $\hat{\eta}$ the global line of sight, which we choose to be one axis of the box. The shot noise term is non-zero for the monopole only:

$$P_0^{\rm noise} = \frac{1}{\bar{n}_g}. \tag{5.3}$$

We use the implementation of the periodic box power spectrum estimator in the Python toolkit `nbodykit` (Hand et al., 2018). The density contrast field $\delta_g(\mathbf{r})$ in Eq. (5.1) is interpolated on a 512^3 mesh following the triangular shaped cloud scheme (TSC, see Section 4.1.7). In the following (see Section 5.2), the box size is $3000\,\mathrm{Mpc}\,h^{-1}$ and thus the Nyquist frequency is $k_N \simeq 0.5\,h\,\mathrm{Mpc}^{-1}$, more than twice larger than the maximum wavenumber used in the RSD analysis ($k = 0.2\,h\,\mathrm{Mpc}^{-1}$). We checked that using a 700^3 mesh ($k_N \simeq 0.7\,h\,\mathrm{Mpc}^{-1}$) does not change our measurement in a detectable way. Then, the term $\delta_g(\mathbf{k})$ in Eq. (5.2) is calculated with a FFT of the interpolated density contrast and the interlacing technique described in Section 4.1.7 is used to mitigate aliasing effects.

The integral $\int_{V_{k_\mu}} d^3k$ in Eq. (5.2) is performed in spherical shells of $\Delta k = 0.01\,h\,\mathrm{Mpc}^{-1}$, starting from $k = 0\,h\,\mathrm{Mpc}^{-1}$. The discrete $\mathbf{k}$-space mesh makes the angular mode distribution irregular at large scales, an effect which we account for in the model following Section 4.2.4 (with a fixed line of sight $\hat{\eta}$).

5.1.1.2 Real survey geometry

As mentioned in Section 4.1.6, we use the implementation of the Yamamoto estimator in the Python toolkit `nbodykit` (Hand et al., 2018) to measure power spectra for real survey geometries. The density field is interpolated on a 512^3 mesh with the TSC scheme. The interlacing technique is again employed to mitigate aliasing effects. Here we use a box size of $4000\,\mathrm{Mpc}\,h^{-1}$, so the Nyquist frequency is $k_N \simeq 0.4\,h\,\mathrm{Mpc}^{-1}$. We also checked that using a 700^3 mesh ($k_N \simeq 0.55\,h\,\mathrm{Mpc}^{-1}$) does not change our measurement significantly.

The integral $\int_{V_{k_\mu}} d^3k$ in Eq. (4.71) is also performed in spherical shells of $\Delta k = 0.01\,h\,\mathrm{Mpc}^{-1}$, starting from $k = 0\,h\,\mathrm{Mpc}^{-1}$, unless otherwise stated. We include the impact of the irregular mode distribution on large scales in the model following Section 4.2.4 (with a varying line of sight).

5.1.1.3 Fiducial cosmology

To obtain the FKP field as a function of comoving position $\mathbf{r}$ we turn galaxy redshifts into distances assuming a fiducial cosmology. This fiducial cosmology will be also used (unless otherwise stated) to compute the linear matter power spectrum for the RSD and BAO analyses in Section 5.1.2 and Section 5.1.3. For both purposes, we utilised the same fiducial cosmology as in Section 4.3.6.2 [1]:

$$\begin{gathered} h = 0.676,\ \Omega_{m,0} = 0.31,\ \Omega_{\Lambda,0} = 0.69,\ \omega_{b,0} = 0.022, \\ \sigma_{8,0} = 0.80,\ n_s = 0.97,\ \sum m_\nu = 0.06\,\mathrm{eV}. \end{gathered} \tag{5.4}$$

1. Note that only $\Omega_{m,0}$ (and $\Omega_{\Lambda,0}$) matter for the redshift to comoving distance (in $\mathrm{Mpc}\,h^{-1}$) conversion.

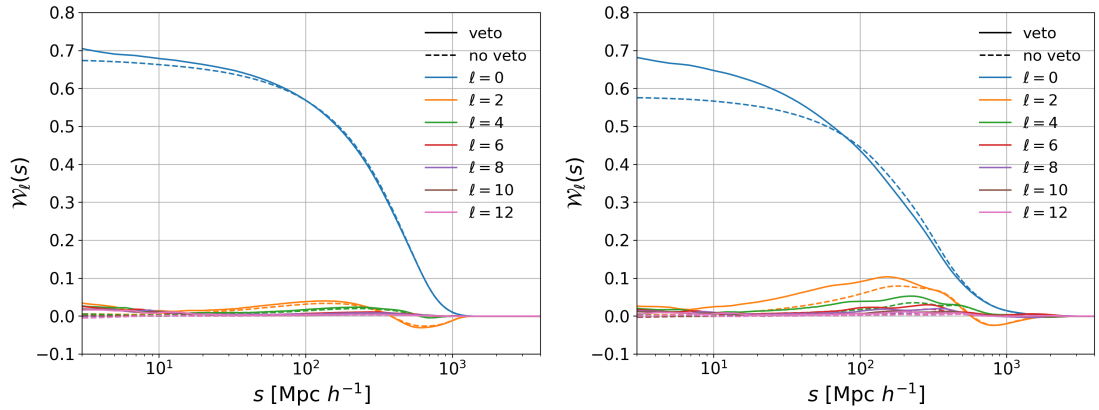

Figure 5.1 – Window function multipoles (left: NGC, right: SGC) of the EZ mocks (reproducing the eBOSS ELG sample), before (dashed lines) and after (continuous lines) application of the veto masks. The height difference between the window function monopoles is explained by the area covered by the veto masks (see text for details).

Within this fiducial cosmology, that will be used throughout this chapter (unless otherwise stated), the comoving sound horizon at the redshift at which the baryon-drag optical depth equals unity (see Section 1.3.5) is $r_{\rm drag}^{\rm fid} = 147.77\,{\rm Mpc}\,h^{-1}$.

5.1.2 RSD model

We use the RSD TNS model as detailed in Section 4.2.5. Let us simply recall that the linear matter power spectrum $P_m^{\rm lin}(k)$ for perturbation theory calculation is kept fixed in the cosmological fits, as usual in BOSS and eBOSS clustering analyses (e.g. Beutler et al., 2017; Gil-Marín et al., 2016, 2020). In the following we detail the geometry effects related to the eBOSS ELG sample.

5.1.2.1 Survey geometry

As detailed in Section 4.2.2, the observed galaxy density is modulated by the survey selection function. Figure 5.1 shows the window function multipoles of the EZ mocks (reproducing the eBOSS ELG sample, see Section 5.3.3): the monopole has a non-zero slope even below $\lesssim 5\,{\rm Mpc}\,h^{-1}$ due to the fine-grained veto masks. We recall that contrary to previous clustering analyses imposing window functions to converge to 1 on small scales, we properly normalise these window functions by the same term as the power spectrum estimation. For comparison purposes, we also plot the window function without veto masks applied; in this case, the monopole stabilises faster on small scales. The area entering the estimation of A (through the density n_s in Eq. (4.68)) used to normalise the unmasked window function (see Eq. (4.138)) has been kept fixed to that in the masked case; since veto masks remove more area in the SGC than in NGC (see Section 5.3.1.1), the unmasked SGC window function is relatively lower than the masked case compared to NGC. In Section 5.4 we further check that veto masks do not bias cosmological measurements with our treatment of the window function.

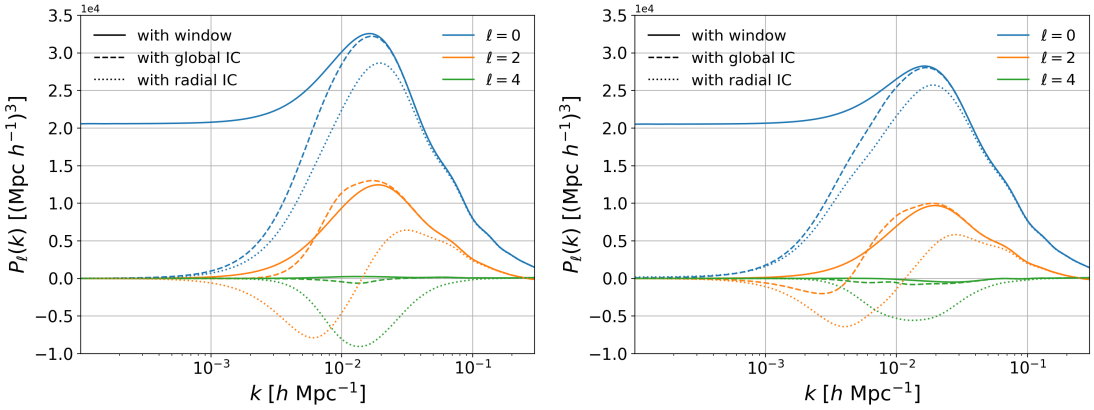

Figure 5.2 – Power spectrum multipoles (left: NGC, right: SGC; blue: monopole, red: quadrupole, green: hexadecapole) of the RSD model. The window function effect only is taken into account in continous lines, and the additional impact of the global and radial integral constraints (IC) are shown in dashed and dotted lines, respectively. For this figure we choose $f = 0.8$, $b_1 = 1.4$, $b_2 = 1$, $\sigma_v = 4\,\mathrm{Mpc}\,h^{-1}$.

5.1.2.2 Integral constraints

As explained in Section 4.3.1, in the Yamamoto estimator the normalisation of the random catalogue is tuned on the data, leading to a global integral constraint (GIC), which we model following Section 4.3.3. Moreover, as discussed in Section 5.3.1.6 redshifts of the random catalogue sampling the selection function are randomly drawn from the data (*shuffled* scheme), leading to a radial integral constraint (RIC), as explained in Section 4.3.6. We therefore model this effect following Section 4.3.3. The impact of the global and radial integral constraints on the power spectrum multipoles is shown in Figure 5.2. Note that the integral constraint formalism will also be used to account for our mitigating angular observational systematics, as suggested in Section 4.3.7.2 and detailed in Section 5.4.

5.1.3 BAO template

BAO analyses rely exclusively on the acoustic scale in the power spectrum or correlation function to measure the Alcock-Paczynski effect (see Section 4.2.1) due to the difference between the fiducial cosmology used in the analysis and the true, observed cosmology. One can measure the scaling parameters parallel ($\alpha_\parallel$) and transverse ($\alpha_\perp$) to the line of sight, using a model for the anisotropic power spectrum (see e.g. Beutler et al., 2017; Gil-Marín et al., 2016). Since the eBOSS ELG sample has a rather low signal-to-noise, we rather focus on an isotropic BAO fit (e.g. Ata et al., 2018).

We checked that the amplitude of the power spectrum measured at $k \simeq 0.1\,h\,\mathrm{Mpc}^{-1}$ on post-reconstruction mock catalogues (see Section 5.3.3) is roughly constant over μ, suggesting that the BAO information is isotropically distributed. Thus, the monopole is optimal for single-parameter BAO-scale measurement, which can be used to constrain the following combination (Eisenstein et al., 2005; Ross et al., 2015):

$$\alpha = \alpha_\parallel^{1/3} \alpha_\perp^{2/3}. \tag{5.5}$$

α can be translated into a measurement of the volume-averaged distance $D_{\rm V}(z) = \left(D_{\rm M}^2(z) D_{\rm H}(z) z\right)^{1/3}$ at the effective redshift of the sample $z_{\rm eff}$ through a relation similar to Eq. (4.113):

$$\alpha = \frac{D_{\rm V}(z_{\rm eff}) r_{\rm drag}^{\rm fid}}{D_{\rm V}^{\rm fid}(z_{\rm eff}) r_{\rm drag}}. \tag{5.6}$$

To fit the isotropic BAO feature, we use the same power spectrum template (dubbed *wiggle* template) as in previous analyses of BOSS and eBOSS (Beutler et al., 2017; Gil-Marín et al., 2016; Ata et al., 2018):

$$P(k, \alpha) = P_{\rm sm}(k) \mathcal{O}_{\rm damp}(k/\alpha) \tag{5.7}$$

where:

$$\mathcal{O}_{\rm damp}(k) = 1 + \left[\mathcal{O}(k) - 1\right] e^{-\frac{1}{2}\Sigma_{\rm nl}^2 k^2}. \tag{5.8}$$

$\mathcal{O}(k)$ is obtained by taking the ratio of the linear matter power spectrum $P_m^{\rm lin}(k)$ to the no-wiggle power spectrum of Eisenstein and Hu (1998), augmented by a five order polynomial term, fitted such that $\mathcal{O}(k)$ oscillates around 1. We take:

$$P_{\rm sm}(k) = B_{\rm nw}^2 P_{\rm nw}(k) + \sum_{i=-2}^{i=2} A_i k^i, \tag{5.9}$$

where $P_{\rm nw}(k) = P_m^{\rm lin}(k)/\mathcal{O}(k)$. Again, the linear matter power spectrum $P_m^{\rm lin}(k)$ is computed with the fiducial cosmology (5.4) (except otherwise stated) and kept fixed in the cosmological fits. The number of broadband parameters A_i is found such that the BAO template Eq. (5.7) can reproduce the mean of the EZ mocks (see Section 5.3.3) within 10% of the uncertainty on the data power spectrum measurement. To specify the BAO detection and for plotting purposes in Section 5.5.1, we will use the no-wiggle template obtained by removing the oscillation pattern in Eq. (5.7) (i.e. keeping only the $P_{\rm sm}(k)$ factor).

The effect of the survey window function is accounted for according to Section 4.2.2 through the (dominant) monopole term only, since the power spectrum template is isotropic. This is legitimate since broadband terms typically absorb these smooth distortions of the power spectrum. We checked that totally ignoring the window function effect leads to a negligible change of $\simeq 10^{-3}$ in the α measurement with the eBOSS ELG data. We also neglect the integral constraints, as their impact will be shown to be negligible in Section 5.4. Finally, we cannot include a correction for the irregular μ sampling as the power spectrum template in Eq. (5.7) is isotropic; this is not an issue since the correction seen in the case of the RSD model (a relative change of $\simeq 10^{-3}$ in the cosmological parameters) is very small.

5.1.4 Reconstruction

Non-linear evolution of the density field and redshift space distortions tend to smooth out baryon acoustic oscillations (see Section 2.4, Figure 2.17). Hence, a method, called reconstruction, was developed to enhance the BAO feature in the galaxy 2-point correlation function by (partially) undoing the aforementioned non-linear features. Reconstruction proceeds by displacing galaxies (and randoms, accounting for the selection function) by the reverse motion due to gravity evolution. This displacement is usually computed assuming the Zel'dovich approximation (see Section 2.1.4).

Assuming linear perturbation theory, the linear displacement field Ψ can be obtained by solving in real space (index i runs over the three spatial coordinates):

$$\partial_i \Psi_i = -\delta_m \tag{5.10}$$

with δ_m the matter density contrast. As a function of the galaxy density contrast in redshift space, δ_g, one must solve the following equation at any position $\mathbf{r}$:

$$\partial_i \Psi_i + f \partial_i \left(\mathbf{\Psi} \cdot \hat{\mathbf{r}} \right) \hat{r}_i = -\frac{\delta_g}{b} \tag{5.11}$$

with b the Eulerian linear bias. The shift $\mathbf{\Psi}$ can be easily computed using FFTs, in the flat sky approximation (constant line of sight) (Burden et al., 2015). This is however not possible with the varying line of sight of spectroscopic surveys due to (the curl of) the $\left(\mathbf{\Psi} \cdot \hat{\mathbf{r}} \right) \hat{\mathbf{r}}$ term. Hence, an iterative procedure to solve for $\mathbf{\Psi}$ with FFTs was developed in Burden et al. (2015); Bautista et al. (2018). Next, RSDs can be removed with:

$$\mathbf{\Psi}_{\mathrm{RSD}} = -f \left(\mathbf{\Psi} \cdot \hat{\mathbf{r}} \right) \hat{\mathbf{r}}. \tag{5.12}$$

In practice, we perform three reconstruction iterations, assuming the growth rate parameter $f = 0.82$ and the linear bias $b = 1.4$, as expected within the ΛCDM model for ELGs at the effective redshift of the sample. The density contrast field is smoothed by a Gaussian kernel of width $15\,\mathrm{Mpc}\,h^{-1}$ to estimate $\mathbf{\Psi}$. The choice of these reconstruction conditions and the assumed fiducial cosmology were shown to have very small impact on the BAO measurements in Vargas-Magaña et al. (2018) and Carter et al. (2020).

In this paper, isotropic BAO fits are performed on both pre- and post-reconstruction monopole power spectra, while the RSD analysis makes use of the monopole, quadrupole and hexadecapole of the pre-reconstruction power spectrum. As we will see in Section 5.5, the posterior of the RSD only measurement is significantly non-Gaussian, making it hard to combine with the isotropic BAO posteriors. We thus also use jointly the pre-reconstruction multipoles with the post-reconstruction monopole (taking into account their cross-covariance) to perform a combined RSD and isotropic BAO fit.

5.1.5 Parameter estimation

In the RSD analysis, fitted cosmological parameters are the growth rate of structure f and the scaling parameters $\alpha_\parallel$ and $\alpha_\perp$. We recall that $\alpha_\parallel$ and $\alpha_\perp$ can be translated into $D_{\mathrm{H}}/r_{\mathrm{drag}}$ and $D_{\mathrm{M}}/r_{\mathrm{drag}}$ at the effective redshift of the sample z_{eff} using Eq. (4.113). Since f is very degenerate with the power spectrum normalisation σ_8, we quote the combination $f\sigma_8$. As discussed in Gil-Marín et al. (2020); Bautista et al. (2020), we take σ_8 as the normalisation of the power spectrum at $8 \times \alpha\mathrm{Mpc}\,h^{-1}$ (instead of $8\,\mathrm{Mpc}\,h^{-1}$), with $\alpha = \alpha_\parallel^{1/3}\alpha_\perp^{2/3}$ as measured from the fit. We emphasise that the quoted $f\sigma_8$ measurement can be straightforwardly compared to any $f\sigma_8$ prediction, as usual. The sensitivity of our RSD measurements on the assumed fiducial cosmology is discussed in Section 5.2.4. We consider 4 nuisance parameters for the RSD fit: the linear and second order biases b_1 and b_2, the velocity dispersion σ_v and $A_g = N_g/P_0^{\mathrm{noise}}$, with N_g the constant galaxy stochastic term (see Eq. (2.150)) and P_0^{noise} the measured Poisson shot noise (see Eq. (4.67)). Again, b_1 and b_2 are almost completely degenerate with σ_8, so we quote $b_1\sigma_8$ and $b_2\sigma_8$.

For the isotropic BAO fit, the fitted cosmological parameter is α, which can be translated into $D_{\mathrm{V}}/r_{\mathrm{drag}}$ at z_{eff} using Eq. (5.6). Nuisance parameters are B_{nw} and the broadband terms $(A_i)_{i\in[-2,2]}$ in Eq. (5.9). These last terms are fixed by solving the least square problem for each value of α, B_{nw}. The non-linear damping scale Σ_{nl} is fixed using N-body simulations in Section 5.2.5.

For the combined RSD and post-reconstruction isotropic BAO fit, we use parameters from both analyses. We rely on Eq. (5.5) to relate α from the isotropic BAO fit to the $\alpha_{\parallel}$ and $\alpha_{\perp}$ scaling parameters of the RSD fit. We fix $B_{\rm nw}$ to b_1, as this choice introduced no detectable bias in the fits of the EZ mocks (see Section 5.4). The varied parameters are reported in Table 5.1.

The fitted k-range of the RSD measurement is $0.03 - 0.2\,h\,\mathrm{Mpc}^{-1}$ for the monopole and quadrupole and $0.03 - 0.15\,h\,\mathrm{Mpc}^{-1}$ for the hexadecapole. We choose such a minimum k to avoid large scale systematics and non-Gaussianity. For the isotropic BAO fit we use the monopole between 0.03 and $0.3\,h\,\mathrm{Mpc}^{-1}$.

5.1.6 Likelihood

As is in some other eBOSS analyses (e.g. Raichoor et al., 2020; Neveux et al., 2020; Bautista et al., 2020), we use a frequentist approach to estimate the scaling parameter α for the isotropic BAO analysis. Bayesian inference is used to obtain posteriors for the eBOSS ELG RSD (and RSD + BAO) measurements. For the sake of computing time, we use a frequentist estimate of the cosmological parameters from the N-body based and approximate mocks and to perform data robustness tests.

In the frequentist approach, we perform a χ^2 minimisation using the `Minuit` (James and Roos, 1975) algorithm, as explained in Section 4.2.5. Errors are determined by likelihood profiling: the error on parameter p_i is obtained by scanning the $p_i \to \min_{p_{j\neq i}} \chi^2(\mathbf{p})$ profile until the χ^2 difference to the best fit reaches $\Delta\chi^2 = 1$ (while minimising over other parameters p_j).

In the case of N-body mocks with periodic boundary conditions (see Section 5.2.2), we compute an analytical covariance matrix following Grieb et al. (2016). In the case of data (see Section 5.3.1) or sky-cut mocks (from N-body simulations in Section 5.2 or approximate mocks in Section 5.3.3 and 5.3.4), the power spectrum covariance matrix is estimated from approximate mocks (lognormal, EZ or GLAM-QPM mocks). We thus apply the Hartlap correction factor (Hartlap et al., 2007) to the inverse of the covariance matrix C measured from the mocks:

$$\mathbf{\Psi} = (1 - D)\,\mathbf{C}^{-1}, \quad D = \frac{n_b + 1}{n_m - 1} \tag{5.13}$$

with n_b the number of bins and n_m the number of mocks. In addition, in order to propagate the uncertainty on the estimation of the covariance matrix we rescale the parameter covariance (Dodelson and Schneider, 2013; Percival et al., 2014) by:

$$m_1 = \frac{1 + B\,(n_b - n_p)}{1 + A + B\,(n_p + 1)}, \tag{5.14}$$

with n_p the number of varied parameters and:

$$A = \frac{2}{(n_m - n_b - 1)\,(n_m - n_b - 4)}, \tag{5.15}$$

$$B = \frac{n_m - n_b - 2}{(n_m - n_b - 1)\,(n_m - n_b - 4)}. \tag{5.16}$$

When cosmological fits are performed on the same mocks used to estimate the covariance matrix, the covariance of the obtained best fits is rescaled by:

$$m_2 = (1 - D)^{-1}\,m_1 \tag{5.17}$$

to be compared with the covariance derived from the likelihood, including corrections (5.13) and (5.14). We discuss further corrections to a mock-based covariance matrix in Section 6.2. In

	RSD	BAO	RSD + BAO
varied parameters	f, $\alpha_\parallel$, $\alpha_\perp$, b_1, b_2, σ_v, A_g	α, $B_{\rm nw}$, $(A_i)_i$	f, $\alpha_\parallel$, $\alpha_\perp$, b_1, b_2, σ_v, A_g, $(A_i)_i$
NGC/SGC specific	b_1, b_2, σ_v, A_g	$B_{\rm nw}$, $(A_i)_i$	b_1, b_2, σ_v, A_g, $(A_i)_i$
priors (MCMC)	$f \geq 0$, $b_1 \geq 0$, $\sigma_v \geq 0$	-	$f \geq 0$, $b_1 \geq 0$, $\sigma_v \geq 0$

Table 5.1 – Varied parameters, and their priors in the case of Bayesian inference (MCMC). Priors are all flat, with infinite bounds, except for those mentioned in the bottom row. No MCMC is run for the BAO only analysis. In all cases (including MCMC), parameters $(A_i)_{i\in[-2,2]}$ are solved analytically (see text).

particular, we propose a new version of this formula, accounting for a combined measurement of several likelihoods, which we use when fitting both the NGC and SGC (though we will refer to the formulae above in this chapter). The magnitude of the rescaling (5.14) is of order 5.5% at most (for the combined RSD + BAO measurements).

In the Bayesian approach, which we use to produce the posterior of the eBOSS ELG RSD and combined RSD + BAO measurements, the uncertainty on the covariance matrix estimation is marginalised over following Sellentin and Heavens (2016):

$$\mathcal{L}(\mathbf{x}^d|\mathbf{p}) = \frac{c_b\,|\mathbf{C}|^{-\frac{1}{2}}}{\left\{1+\frac{1}{n_m-1}\left[\mathbf{x}^d-\mathbf{x}^t(\mathbf{p})\right]^T\mathbf{C}^{-1}\left[\mathbf{x}^d-\mathbf{x}^t(\mathbf{p})\right]\right\}^{\frac{n_m}{2}}}, \qquad c_b = \frac{\Gamma\left(\frac{n_m}{2}\right)}{\left[\pi\left(n_m-1\right)\right]^{\frac{n_b}{2}}\Gamma\left(\frac{n_m-n_b}{2}\right)}, \tag{5.18}$$

where we note the power spectrum measurements (data) $\mathbf{x}^d$ and the model (theory) $\mathbf{x}^t$ as a function of parameters $\mathbf{p}$. In the limit $n_m \to +\infty$ one recovers a standard Gaussian likelihood. Eq. (5.18) will be further discussed in Section 6.2.3. The combined NGC and SGC likelihood is trivially the product of NGC and SGC likelihoods. Our posterior is the product of Eq. (5.18) with flat priors on all parameters, infinite for all of them, except for f, b_1 and σ_v, which are lower-bounded by 0 (see Table 5.1).

As mentioned in Section 4.2.5, to sample the posterior distribution we run Monte Carlo Markov Chains (MCMC) with the package `emcee` (Foreman-Mackey et al., 2013). We run 8 chains in parallel and check their convergence using the Gelman-Rubin criterion $R-1<0.02$ (Gelman and Rubin, 1992).

5.1.7 Effective redshift

In order to match the definition used for other eBOSS tracers and analyses, the effective redshift $z_{\rm eff}$ of the ELG sample between $0.6<z<1.1$ is calculated as the average over pairs of galaxies:

$$z_{\rm eff} = \frac{\sum_{i,j} w_{{\rm tot},i} w_{{\rm tot},j} (z_{g,i}+z_{g,j})/2}{\sum_{i,j} w_{{\rm tot},i} w_{{\rm tot},j}}, \tag{5.19}$$

where $w_{{\rm tot},i}$ is the total galaxy weight (completeness weight and FKP weight, see Section 5.3.1) and the sum is performed over all galaxy pairs between $25\,{\rm Mpc}\,h^{-1}$ and $120\,{\rm Mpc}\,h^{-1}$. We measure $z_{\rm eff}=0.845$ for the combined NGC and SGC (NGC alone: 0.849, SGC alone: 0.841). We checked that this result varies by less than 0.4% when including pairs between $0\,{\rm Mpc}\,h^{-1}$ and $200\,{\rm Mpc}\,h^{-1}$. We also compute the effective redshift corresponding to the cuts $0.7<z<1.1$, which will be used in Section 5.5: $z_{\rm eff}=0.857$ for the combined NGC and SGC (NGC alone: 0.860, SGC alone: 0.853). The typical variation (using Eq. (5.4)) corresponding to the $\simeq 0.8\%$ difference between the effective redshift of NGC and SGC is 0.2% on $f\sigma_8$, 0.4% on $D_{\rm H}/r_{\rm drag}$ and 0.6% on $D_{\rm M}/r_{\rm drag}$, small compared to the statistical uncertainty (see Section 5.5).

Definition (5.19) is not very satisfactory as (i) the separation range is not well specified and (ii) we expect the effective redshift to actually depend on the full compression of galaxy position into cosmological parameters, hence on the power spectrum estimator and the scale of ranges used in the fit. Let us derive another definition of the effective redshift, more specific to the Fourier-space analysis. With the Yamamoto estimator (4.46), in the local plane parallel approximation, assuming the (perfectly known) survey selection function varies slowly compared to the correlation function, and with infinitely thin V_{k_μ} sampling we measure (in average):

$$\left\langle \hat{P}_\ell(k) \right\rangle = \frac{\int d^3r \bar{n}^2(r) P_\ell(k, z(r))}{\int d^3r \bar{n}^2(r)} \tag{5.20}$$

where $\bar{n}(r)$ is the expected mean density of weighted galaxies in the absence of clustering and $P_\ell(k, z)$ is the true galaxy power spectrum at redshift z. Taylor expanding the power spectrum about the effective redshift $z_{\rm eff}$:

$$P_\ell(k, z) = P_\ell(k, z_{\rm eff}) + P'_\ell(k, z_{\rm eff})(z - z_{\rm eff}) + \cdots \tag{5.21}$$

and substituting this expression into Eq. (5.20), we find that the expected value of the Yamamoto estimator can be approximated at first order by the quantity that we actually model:

$$\tilde{P}_\ell(k) = \frac{\int d^3r \bar{n}^2(r) P_\ell(k, z_{\rm eff})}{\int d^3r \bar{n}^2(r)} \tag{5.22}$$

if we use the effective redshift:

$$z_{\rm eff} = \frac{\int d^3r \bar{n}^2(r) z(r)}{\int d^3r \bar{n}^2(r)} \simeq \frac{\sum_{i=1}^{N_g} w_{\rm tot,i} n_{g,i} z_{g,i}}{\sum_{i=1}^{N_g} w_{\rm tot,i} n_{g,i}} \tag{5.23}$$

where $n_{g,i} = n(z_i)$ the estimated radial density and $w_{\rm FKP,i}$ the FKP weights (see Section 5.3.1). We checked that the value obtained with this definition of the effective redshift agrees with Eq. (5.19) to the 0.5% level.

5.2 Mock challenge

In this section we validate our implementation of the RSD TNS model and isotropic BAO template (presented in Section 5.1.2 and 5.1.3) against mocks based on N-body simulations, which are expected to more faithfully reproduce the small scale, non-linear galaxy clustering.

Though some work explored the relation between star-forming ELGs and halos (see Section 2.2.3), using either semi-analytical models of galaxy formation (e.g. Gonzalez-Perez et al., 2018) or observational data (e.g. Favole et al., 2016; Guo et al., 2019), there is still significant uncertainty on the way star-forming ELGs populate halos, from the mean HOD shape to the position and velocities of ELGs. Therefore, we aim at testing the robustness of the RSD model measurements over a wide range of galaxy to halo relations in Section 5.2.1 and Section 5.2.2. We refer the interested reader to Alam et al. (2020) and Avila et al. (2020) for a complete description of this mock challenge. In addition, we assess the impact of assuming a fixed template cosmology for the calculation of the theoretical model in Section 5.2.4, and perform similar tests for the BAO analysis in Section 5.2.4.

5.2.1 MultiDark mocks

A first set of mocks is based on the MultiDark simulation (see Section 2.2.1.5).

Dark matter halos were populated with galaxies following two HOD models: a standard HOD (*SHOD*) and a HOD quenched at high mass. The standard HOD is composed of a smooth step function for central galaxies and is given by (for lisibility we change some conventions with respect to Alam et al. (2020) to match those of Avila et al. (2020)):

SHOD

$$\langle N_{\rm cen}(M)\rangle = \frac{1}{2}A_c\left[1+{\rm erf}\left(\frac{\log M-\mu}{\sigma}\right)\right] \tag{5.24}$$

μ sets the typical minimum mass scale for a halo to host a central galaxy, and hence is tightly related to the galaxy bias (more massive halos are more biased).

The second mean central HOD considered is called high mass quenched (*HMQ*, Alam et al., 2020), to account for the decrease of the number of central galaxies in high mass halos noted in Section 2.2.3:

HMQ

$$\langle N_{\rm cen}(M)\rangle = 2A\phi(M)\Phi(\gamma M)+\frac{1}{2Q}\left[1+{\rm erf}\left(\frac{\log M-\log M_c}{0.01}\right)\right], \tag{5.25}$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma_M}e^{-\frac{(x-\log M_c)^2}{2\sigma_M^2}} \tag{5.26}$$

$$\Phi(x) = \int_{-\infty}^{x}\phi(t)dt = \frac{1}{2}\left[1+{\rm erf}\left(\frac{x-\log M_c}{\sqrt{2}\sigma_M}\right)\right] \tag{5.27}$$

$$A = \frac{p_{\rm max}-1/Q}{\max(2\phi(x)\Phi(\gamma x))}. \tag{5.28}$$

$\max(2\phi(x)\Phi(\gamma x))$ is the maximum of $2\phi(x)\Phi(\gamma x)$. The role of M_c is similar to that of μ in Eq. (5.24) i.e. it is the typical minimum halo mass. $p_{\rm max}$ increases the height of the low-mass Gaussian part relative to the high-mass plateau, whose level is tuned by Q.

In both *SHOD* and *HMQ* cases, for satellite galaxies, $\langle N_{\rm sat}(M)\rangle$ is a power-law above a certain halo mass M_0:

$$\langle N_{\rm sat}(M)\rangle = \begin{cases} 0 & M\leq M_0 \\ A_s\left(\frac{M-M_0}{M_1}\right)^{\alpha} & M>M_0 \end{cases} \tag{5.29}$$

which is kept for all the HOD models considered in this mock challenge. The number of central galaxies is sampled from Bernouilli's probability law (of mean $\langle N_{\rm cen}(M)\rangle \leq 1$). Those are placed at the halo centre, with the velocity of their host halo. A Poisson law (of parameter $\langle N_{\rm sat}(M)\rangle$) is used to sample the number of satellites in each halo. Satellite galaxy positions are sampled from a Navarro et al. (1996) profile (see Eq. (2.95)) and their velocities from a Gaussian of width the dispersion of particle velocities within their halo. Then, different deviations around these baseline cases were considered, as summarised in Table 5.2. For type $0.1R_{200}$ central galaxies are displaced following a Gaussian of zero mean and width $0.1R_{200}$, where R_{200} is the radius within which the mean density is 200 the matter density at the simulation redshift (see Section 2.2.2.2). Then, velocities of satellite galaxies are decreased (type $\alpha_v = 0.5$) or increased (type $\alpha_v = 1.5$) by 50%. The concentration (for satellite galaxies) was varied by -50% (type $K = 0.5$) and $+50\%$ (type $K = 1.5$). As discussed in Section 2.2.2.6, the way galaxies populate dark matter

short form	**description**	**simulations**
baseline	*SHOD* (5.24) and *HMQ* (5.25), positions from NFW and velocities from DM particles	MD, OR
$0.1R_{200}$	central galaxies are off-centred with a Gaussian distribution of width $0.1R_{200}$	MD, OR
$\alpha_v = 0.5$	satellite galaxies have 50% lower velocity dispersion than DM particles	MD, OR
$\alpha_v = 1.5$	satellite galaxies have 50% higher velocity dispersion than DM particles	MD, OR
$K = 0.5$	satellite galaxies have 50% lower concentration than dark matter	MD
$K = 1.5$	satellite galaxies have 50% higher concentration than dark matter	MD
$(A_{\rm cen}, A_{\rm sat}) = (0.3, 0.0)$	assembly bias; central galaxies occupation is correlated with halo concentration ($A_{\rm cen} = 0.3$)	MD
$(A_{\rm cen}, A_{\rm sat}) = (0.0, 0.3)$	assembly bias; satellite galaxies occupation is correlated with halo concentration ($A_{\rm sat} = 0.3$)	MD
$(A_{\rm cen}, A_{\rm sat}) = (0.3, 0.3)$	assembly bias; central and Satellite galaxies occupation is correlated with halo concentration ($A_{\rm cen} = A_{\rm sat} = 0.3$)	MD
$v \times 0.8$	peculiar velocities of galaxies are scaled lower by 20%	MD, OR
$v \times 1.2$	peculiar velocities of galaxies are scaled higher by 20%	MD, OR

Table 5.2 – Some of HOD variations around *SHOD* (5.24) and *HMQ* (5.25) mean HODs considered for the mock challenge, applied to MultiDark (MD) or OuterRim (OR) simulations. Adapted from Alam et al. (2020).

halos can depend on halo secondary properties; the so-called assembly bias. Hence, Alam et al. (2020) introduce assembly bias using the *decorated* (Hearin et al., 2016) HOD of Zentner et al. (2019). The central ($X = \text{cen}$) and satellite ($X = \text{sat}$) mean decorated HOD are given by:

$$\left\langle N_X^{\rm dec}(M, c) \right\rangle = \langle N_X(M) \rangle + \begin{cases} \delta N_X(M) & c > c_{\rm med} \\ -\delta N_X(M) & c \leq c_{\rm med} \end{cases} \tag{5.30}$$

where $c_{\rm med}$ is the median concentration, and:

$$\delta N_{\rm cen}(M) = A_{\rm cen} \min\left[\langle N_{\rm cen}(M) \rangle, 1 - \langle N_{\rm cen}(M) \rangle\right] \tag{5.31}$$

$$\delta N_{\rm sat}(M) = A_{\rm sat} \langle N_{\rm sat}(M) \rangle \tag{5.32}$$

Eq. (5.30) varies the number of galaxies depending on their concentration. The strength of assembly bias is controlled by $A_{\rm cen} \in [-1, 1]$ and $A_{\rm sat} \in [-1, 1]$. Alam et al. (2020) tested different combinations: $(A_{\rm cen}, A_{\rm sat}) = (0.3, 0)$, $(A_{\rm cen}, A_{\rm sat}) = (0, 0.3)$, $(A_{\rm cen}, A_{\rm sat}) = (0.3, 0.3)$. Finally, all galaxy peculiar velocities were upscaled (downscaled) by 20% for type $v \times 0.8$ ($v \times 1.2$). In these cases, the measured growth of structure f is expected to equally vary by 20% (−20%). All these cases are recapped in Table 5.2, and the corresponding power spectrum measurements are shown in Figure 5.3 (top). A high velocity dispersion ($\alpha_v = 1.5$) is analogous to an increased Finger-of-God effect and thus significantly damps the quadrupole (while conserving the same large scale bias).

The galaxy density reaches $3 \times 10^{-3}\, h^3\,{\rm Mpc}^{-3}$, about 10 times the mean eBOSS ELG density, such that the shot noise is very low.

We derived a covariance matrix from a set of 500 lognormal mocks produced with `nbodykit`, in the MDPL2 cosmology of Eq. (2.88), assuming a bias of 1.4 and with the same density of $3 \times 10^{-3}\,h^3\,\mathrm{Mpc}^{-3}$. We checked that the agreement between N-body based and lognormal mocks was satisfactory on the whole k-range of the cosmological fit.

Both N-body based and lognormal mocks were analysed with the fiducial cosmology of Eq. (5.4), as for the eBOSS ELG data. We thus accounted for the appropriate window function and global IC effect in the model, and we included the correction for the irregular μ distribution at large scales, as presented in Section 5.1.2. As shown in Figure 5.3 (bottom), the fitted cosmological parameters are found to be within 1σ of the expected values, even for mocks with rescaled galaxy velocities, where the offset in the fitted $f\sigma_8$ values corresponds to the 20% offset in velocity. However, the obtained uncertainties were only half of those expected with the eBOSS ELG sample, which was not sufficient to derive an accurate modelling systematic budget. We thus focused on larger mocks.

5.2.2 OuterRim mocks

Two other sets of mocks were based on the $z_{\mathrm{snap}} = 0.865$ snapshot of the OuterRim (Heitmann et al., 2019) simulation (see Section 2.2.1.5). We analysed these mocks with the OuterRim cosmology of Eq. (2.89), imposing periodic boundary conditions. Therefore, there is no window effect and we only included the correction for the irregular μ distribution (see Section 4.2.4) at large scales. A Gaussian covariance matrix was calculated following Grieb et al. (2016) for each of these mocks, taking their measured power spectrum as input.

A first set of mocks using the *SHOD* and *HMQ* HODs was produced, with 8 types, corresponding to types *baseline*, $0.1R_{200}$, $\alpha_v = 0.5$, $\alpha_v = 1.5$, $K = 0.5$, $K = 1.5$, $v \times 0.8$, $v \times 1.2$ of the MultiDark-based mocks above (see Table 5.2). Indeed, contrary to the MultiDark case, the concentration of each halo was not provided as part of the OuterRim halo catalogues, such that no test relying on the concentration could be implemented. The galaxy density ranges between $3 \times 10^{-3}\,h^3\,\mathrm{Mpc}^{-3}$ (*SHOD*) and $4 \times 10^{-3}\,h^3\,\mathrm{Mpc}^{-3}$ (*HMQ*). Power spectrum measurements are shown in Figure 5.4 (left) and are similar (with less noise) to Figure 5.3. Also in Figure 5.4 (right) are shown best fit cosmological parameters, which are in good agreement with the expected values.

As mentioned at the end of Section 2.3.2, following the assumption of initial bias being local in Lagrangian space, as a baseline bias coefficients b_{s2} and $b_{3\mathrm{nl}}$ are fixed according to:

$$b_{s2} = -\frac{4}{7}\,(b_1 - 1)\,, \tag{5.33}$$

$$b_{3\mathrm{nl}} = \frac{32}{315}\,(b_1 - 1)\,. \tag{5.34}$$

The impact of such an assumption is shown in Figure 5.5; no bias can be seen on cosmological parameters compared to the case were b_{s2} and $b_{3\mathrm{nl}}$ are left free.

A second set of OuterRim mocks was produced based on results from models of galaxy formation and evolution (Avila et al., 2020).

Three mean HOD models were explored. The first (*HOD-1*) is the same as *SHOD* (Eq. (5.24)) which rises from 0 to 1 with a typical transition scale at $\log(M) = \mu$. The shape of the mean HOD typically corresponds to galaxies selected by their magnitude or stellar mass. The completeness A_c is the ratio of observed to the total number of galaxies. The HOD for satellite galaxies is always given by Eq. (5.29).

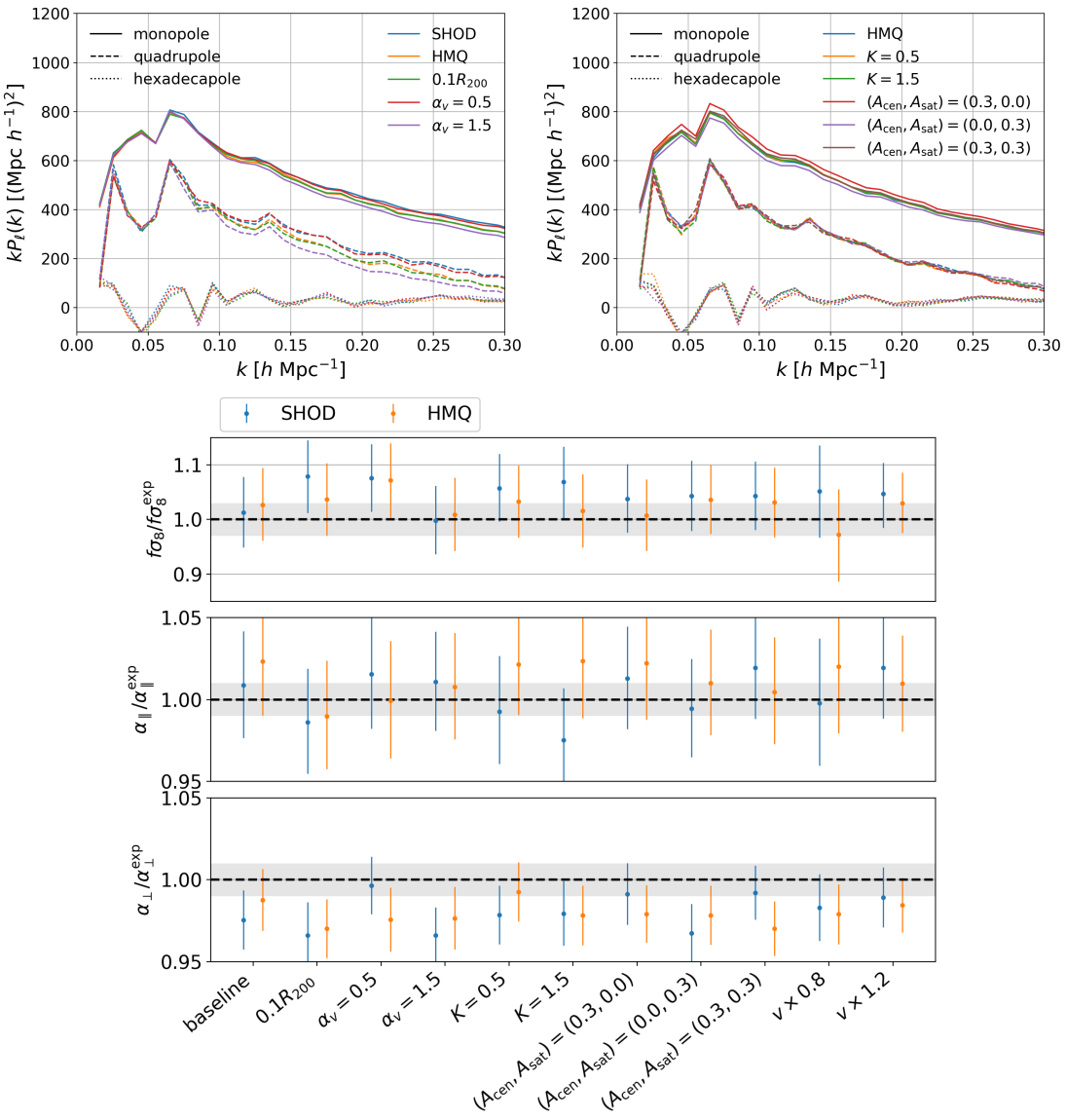

Figure 5.3 – Top: power spectrum multipoles measured on the MultiDark-based mocks as described in Table 5.2. The mean HOD *HMQ* is considered by default (for $0.1R_{200}$, $\alpha_v = 0.5$ and $\alpha_v = 1.5$ curves). Bottom: best fit cosmological parameters. The gray shaded area represents an error of 3% on $f\sigma_8$ and 1% on the scaling parameters on both sides of the reference values in the OuterRim cosmology.

Since the population of star-forming ELGs tends to be quenched at high halo mass, the step function of Eq. (5.24) is replaced by a Gaussian for the second HOD model:

HOD-2

$$\langle N_{\rm cen}(M)\rangle = \frac{A_c}{\sqrt{2\pi}\sigma} e^{-\frac{(\log M - \mu)^2}{2\sigma^2}} \tag{5.35}$$

The third mean HOD model includes an asymmetry of the mean central HOD at $\log M > \mu$, modelled as a decaying power-law:

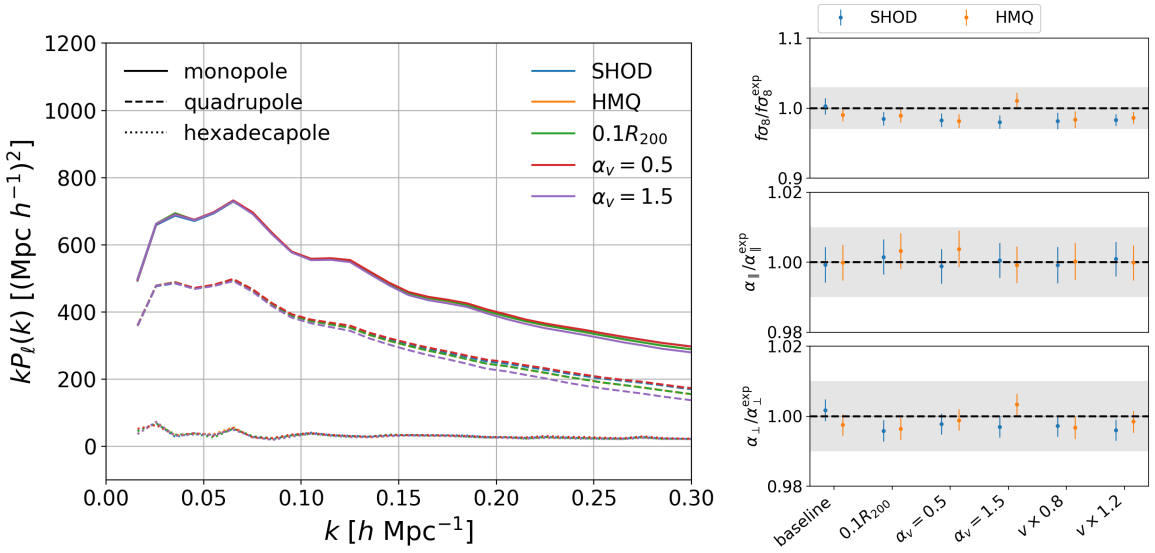

Figure 5.4 – Left: power spectrum multipoles (averaged on x, y and z lines of sight) measured on the first series of OuterRim-based mocks as described in Table 5.2. The mean HOD *HMQ* is considered by default (for $0.1R_{200}$, $\alpha_v = 0.5$ and $\alpha_v = 1.5$ curves). Right: ratio of the parameter best fits to their expected values (for the x line of sight).

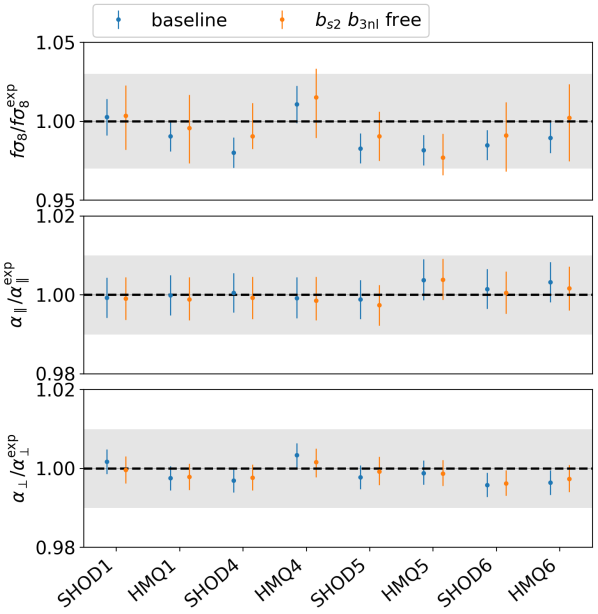

Figure 5.5 – Left: ratio of the best fit parameters to their expected values obtained for the first series of OuterRim-based mocks as described in Table 5.2 (for the x line of sight). In blue (*baseline*) we fix the bias parameters b_{s2} and b_{3nl} with Eq. (5.34) and we leave them free in orange.

HOD-3 (baseline)

$$\langle N_{\rm cen}(M)\rangle = \begin{cases} \frac{A_c}{\sqrt{2\pi}\sigma} e^{-\frac{(\log M-\mu)^2}{2\sigma^2}} & \log M \leq \mu \\ \frac{A_c}{\sqrt{2\pi}\sigma}\left(\frac{M}{10^\mu}\right)^\gamma & \log M > \mu \end{cases} \tag{5.36}$$

which is the model taken as a baseline. The parameters σ, α, γ as well as $\log M_0 - \mu$, $\log M_1 - \mu$ are determined by a fit to the HOD form found by Gonzalez-Perez et al. (2018), such that only

μ, A_c and A_s are left free. A_c and A_s are related to the density $\bar{n}$:

$$\bar{n} = \int dM \frac{dn}{dM}(M_h) \left[\langle N_{\rm cen}(M)\rangle + \langle N_{\rm sat}(M)\rangle\right] \tag{5.37}$$

and the satellite fraction $f_{\rm sat}$ is:

$$f_{\rm sat} = \frac{1}{\bar{n}} \int dM \frac{dn}{dM}(M_h) \langle N_{\rm sat}(M)\rangle \tag{5.38}$$

while μ determines the galaxy bias b_g:

$$b_g = \frac{1}{\bar{n}} \int dM b(M) \frac{dn}{dM}(M_h) \left[\langle N_{\rm cen}(M)\rangle + \langle N_{\rm sat}(M)\rangle\right] \tag{5.39}$$

where $\frac{dn}{dM}(M_h)$ is the halo mass function and $b(M)$ is the bias to halo mass relation measured on the OuterRim simulation. Then, to compare these different HODs, Avila et al. (2020) fix $\bar{n}$ and b_g to the value measured onto the small scale eBOSS ELG clustering. Hence, only $f_{\rm sat}$ is varied. Actually, Avila et al. (2020) found that $f_{\rm sat}$ had a larger impact on the small scale clustering than the HOD type itself. This is confirmed by Figure 5.6, showing (left) the power spectrum measurements and (right) the best fit cosmological parameters, varying the HOD type and satellite fraction. Note that there is some scatter depending on the axis of the box used a line of sight; an effect which we study further in Section 6.1.

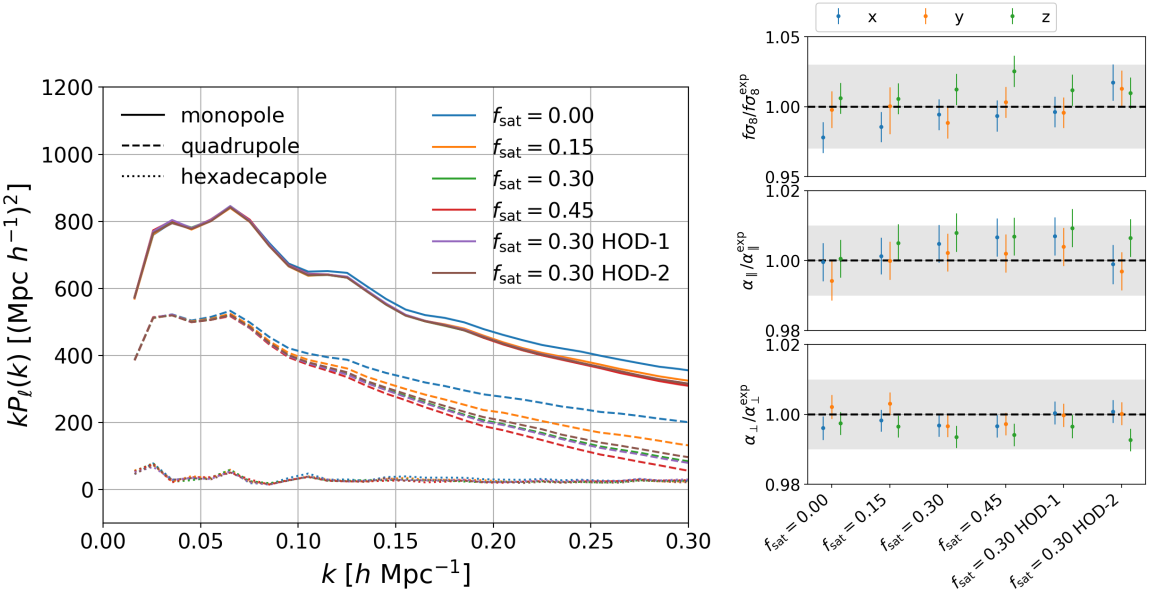

Figure 5.6 – Left: power spectrum multipoles (averaged on x, y and z lines of sight) measured on the second series of OuterRim-based mocks (see text). *HOD-3* (5.36) is considered by default. Right: ratio of the best fit parameters to their expected values for the three different lines of sight.

As in Section 5.2.1, central galaxies are placed at the halo centre, with the velocity of their host halo. As a baseline, a Poisson law (of parameter $\lambda = \langle N_{\rm sat}(M)\rangle$) is used to sample the number of satellites in each halo. The scatter in the number of satellites is thus $\sqrt{\lambda}$. However, as mentioned in Section 2.2.3, Jiménez et al. (2019) found that the HOD scatter has a significant impact on the one halo term, and that a negative binomial distribution (*neg. bin.*) should be preferred for star-forming galaxies (based on semi-analytical models of galaxy formation):

$$P(N|r,p) = \frac{\Gamma(N+r)}{\Gamma(r)\Gamma(N+1)} p^r (1-p)^N \qquad p = \frac{1}{(1+\beta)^2},\ r = \frac{\lambda}{\beta(2+\beta)} \tag{5.40}$$

This distribution, which has mean λ and an increased scatter of $(1+\beta)\sqrt{\lambda}$ ($\beta > 0$) is considered in addition to the standard Poisson distribution. Finally, Avila et al. (2020) consider a next integer (*next int.*) distribution:

$$P(N|\lambda) = \begin{cases} 1 - (\lambda - \text{int}(\lambda)) & N = \text{int}(\lambda) \\ \lambda - \text{int}(\lambda) & N = \text{int}(\lambda) + 1 \\ 0 & \text{else} \end{cases} \tag{5.41}$$

which has mean λ and a reduced scatter of $\sqrt{[1 - (\lambda - \text{int}(\lambda))]\,[\lambda - \text{int}(\lambda)]}$, with $\text{int}(\lambda)$ the whole part of λ. Avila et al. (2020) find that the HOD scatter mostly increases the projected correlation function as small scales. In practice, the type of distribution has a small impact on the large scale clustering measurements, as displayed in Figure 5.7.

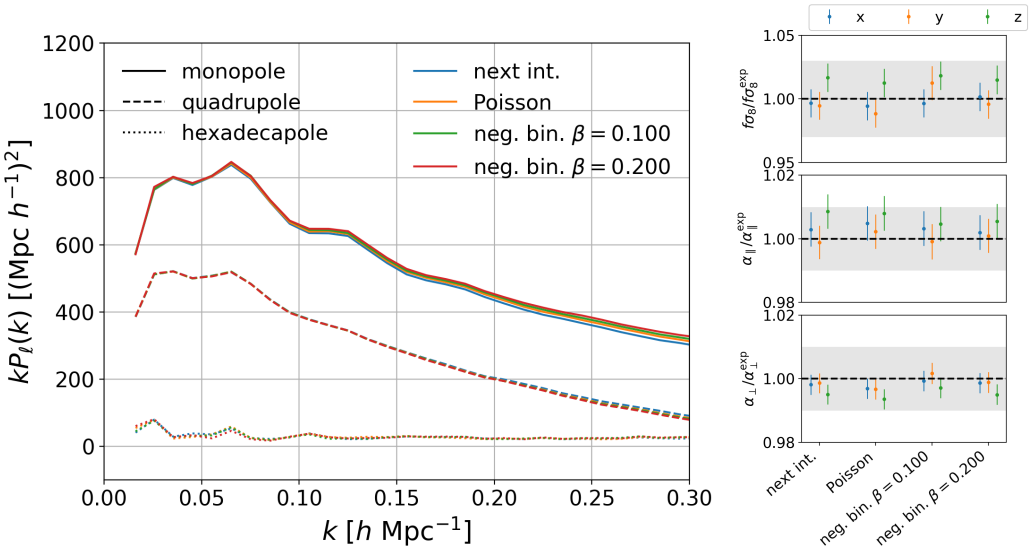

Figure 5.7 – Left: power spectrum multipoles (averaged on x, y and z lines of sight) measured on the second series of OuterRim-based mocks (see text). *HOD-3* (5.36) is assumed. Right: ratio of the best fit parameters to their expected values (for the three lines of sight).

As a baseline, satellite positions are sampled from a NFW profile (*NFW*). The concentration (see Eq. (2.96)) is by default set to the prescription of Klypin et al. (2016). As mentioned in Section 2.2.3, star-forming galaxies are preferentially found in the outskirts of halos (Orsi and Angulo, 2018). A parameter K, scaling the concentration, is thus introduced. Velocities are assigned from the virial theorem (Bryan and Norman, 1998; Avila et al., 2018), i.e. following a Gaussian with mean the halo velocity and standard deviation:

$$\sigma_v = 476 \times 0.9 \left[\Delta_{\text{vir}} E^2(z)\right]^{1/6} \left(\frac{M}{10^{15} M_\odot h^{-1}}\right)^{1/3} \text{km s}^{-1}, \qquad E(z) = \frac{H(z)}{H_0}. \tag{5.42}$$

Another explored possibility is to randomly select (with its velocity) a dark matter particle of the halo as a galaxy (*part.*). We note that Avila et al. (2020) (and references therein) find that the halo profile obtained in such a way is quite different from NFW, as (i) approximately half of dark matter halos are not relaxed at the redshift of interest (when NFW profile could only apply to relaxed halos) and (ii) the profile of (relaxed) dark matter halos appears to be better described by an Einasto (Einasto, 1965) profile (Child et al., 2018). To take into account the

halo concentration, distances of galaxies to the halo centre are rescaled by $1/K$. The change in concentration is found to have a small impact on clustering measurements, as confirmed by Figure 5.8.

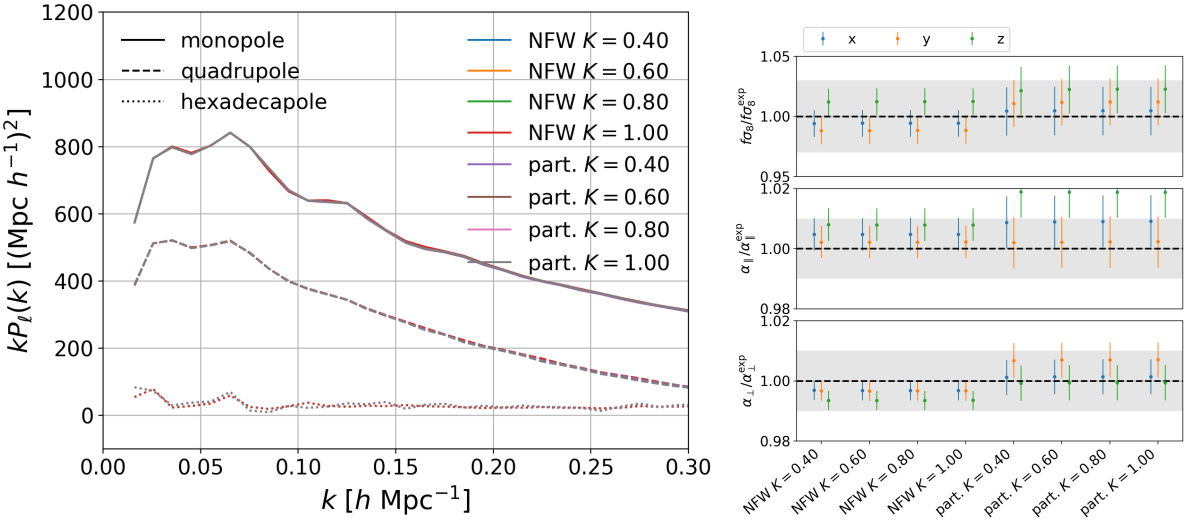

Figure 5.8 – Left: power spectrum multipoles (averaged on x, y and z lines of sight) measured on the second series of OuterRim-based mocks (see text). *HOD-3* (5.36) is assumed. Right: ratio of the best fit parameters to their expected values (for the three lines of sight).

It is not obvious that galaxy velocities should follow those of dark matter inside halos. Hence, Avila et al. (2020) further introduce a velocity bias α_v. In the case of the *part.* scheme, α_v rescales the satellite galaxy relative velocity to the halo velocity. When the *NFW* scheme is considered, α_v scales the velocity dispersion σ_v as given by Eq. (5.42). Furthermore, as mentioned in Section 2.2.3, Orsi and Angulo (2018) found that star-forming galaxies tend to have a high radial velocity dispersion and a net infalling velocity; therefore Avila et al. (2020) try giving an infall component following a Gaussian of mean $v_t = 500\,\mathrm{km\,s^{-1}}$ and dispersion $200\,\mathrm{km\,s^{-1}}$. Main differences due to both velocity bias and infall velocities are found in the quadrupole. Both act indeed as a Finger-of-God term and decrease the amplitude of the quadrupole, as can be seen in Figure 5.9.

Finally, Avila et al. (2020) fit the small-scale eBOSS ELG data power spectrum monopole, quadrupole and hexadecapole. Overall, they find a preference for low concentration ($K < 1$), and velocity bias ($\alpha_v > 1$) — though α_v gets more compatible with 1 once K is left free, and a mild preference for the *part.* scheme. Different probability laws for sampling satellites are preferred depending on the parameters which are let free in the fit.

The number density of the 24 mocks we analysed ranges between the approximate mean eBOSS ELG density $2 \times 10^{-4}\,h^3\,\mathrm{Mpc}^{-3}$ (*part.* and some *NFW* mocks) and $2 \times 10^{-3}\,h^3\,\mathrm{Mpc}^{-3}$ (*NFW* mocks).

No evidence for an overall systematic bias of the model was found when analysing these mocks, as reported in Alam et al. (2020).

5.2.3 Blind OuterRim mocks

We participated in a blind mock challenge dedicated to the ELG sample (see Alam et al. (2020), Section 8). For simplicity, as for $v \times 0.8$ and $v \times 1.2$ cases of Table 5.2, only galaxy

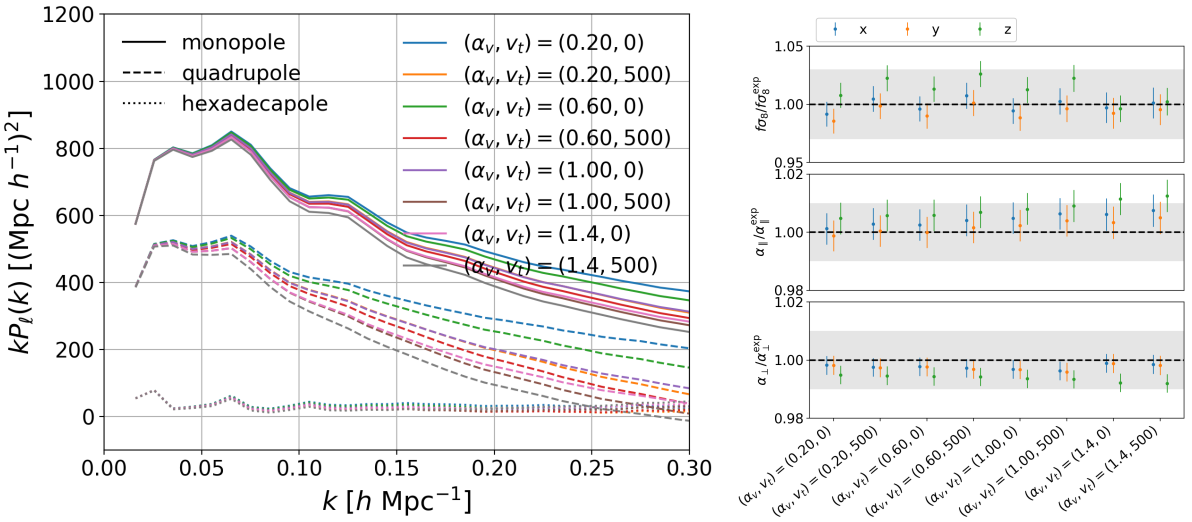

Figure 5.9 – Left: power spectrum multipoles (averaged on x, y and z lines of sight) measured on the second series of OuterRim-based mocks (see text). *HOD-3* (5.36) is assumed. Right: ratio of the best fit parameters to their expected values (for the three lines of sight).

peculiar velocities were scaled, by an initially unknown value -50% (type 1), -25% (type 2) and 0% (type 3), thus changing the expected value of the growth rate by the same amounts. The background cosmology was kept fixed, hence the scaling parameters were expected to agree with 1. 30 to 40 realisations for each type of mocks (3 for *SHOD* and *HMQ*) were produced with a density of the order of the eBOSS ELG mean density ($\simeq 2\times10^{-4}\,h^3\,\mathrm{Mpc}^{-3}$). The OuterRim boxes were analysed the same way as in Section 5.2.2, and the velocity rescaling was then disclosed. Though velocities were scaled by as much as 50%, no significant systematic shift in $f\sigma_8$ can be seen in Figure 5.10, confirming the robustness of our RSD model.

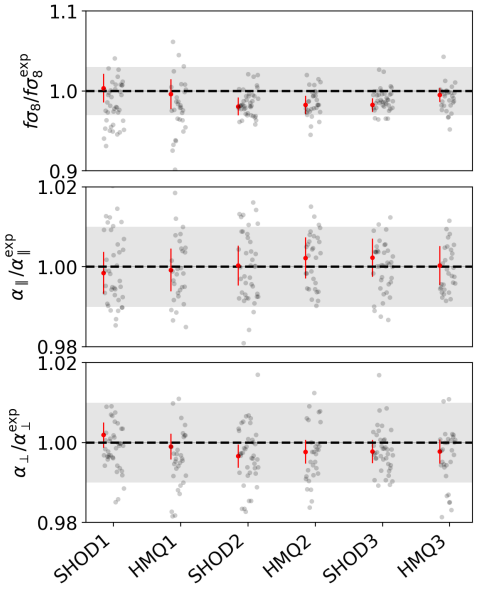

Figure 5.10 – Ratio of the best fit parameters to their expected values for the blind OuterRim-based mocks (for the x line of sight). The red data points and error bars are the measurements obtained on concatenated mocks.

Since no obvious systematic shift was noticed with previous unblind mocks (Section 5.2.1 and Section 5.2.2), the systematic uncertainties were derived from this blind mock challenge in Alam et al. (2020) (Section 9): 1.6% on $f\sigma_8$, 0.8% on $\alpha_\parallel$ and 0.7% on $\alpha_\perp$ [2]. This also reflects the typical scatter due to the varying HOD prescriptions in Section 5.2.2. We do not scale these errors by a factor of 2 as in Alam et al. (2020), since we further take into account the effect of the fiducial cosmology in Section 5.2.4 in a conservative way.

5.2.4 Fiducial cosmology

As stated in Section 5.1.2, we keep the so-called *template cosmology* used to compute the linear power spectrum for the RSD model ($P_m^{\rm lin}(k)$ in Section 5.1.2) fixed, thereby implicitly assuming that the scaling parameters and the growth rate are an unbiased compression of the cosmological dependence of the galaxy power spectrum.

Hence, as Hou et al. (2020); Neveux et al. (2020); Gil-Marín et al. (2020); Bautista et al. (2020) we test the dependence of the measurement of cosmological parameters with respect to the template cosmology. To this end, we reanalyse the first set of OuterRim mocks of Table 5.2 using different template cosmologies. We consider first the fiducial cosmology of the data analysis (5.4) and also scale each of the cosmological parameters (h, $\omega_{c,0}$, $\omega_{b,0}$ and n_s) of Eq. (2.89) by $\pm 10\%$ to $\pm 20\%$ (typically 30σ variations of Planck Collaboration et al. (2018) CMB (TT, TE, EE, lowE, lensing) and BAO constraints).

Note that for simplicity we do not change the fiducial cosmology (2.89) used in the analysis (power spectrum estimation and Gaussian covariance matrix) and thus rescale the fitted $\alpha_\parallel$ and $\alpha_\perp$ accordingly to determine σ_8 as in Section 5.1.5.

Results are shown in Figure 5.11. Scaling parameters are well recovered. The best fit $f\sigma_8$ value is primarily sensitive to the template h and n_s values. Taking the root mean square of the difference (averaged over all types, HODs, and lines of sight) to the expected values gives the following systematic uncertainties: 2.6% on $f\sigma_8$ and 0.4% on scaling parameters. Note that without the σ_8 rescaling described in Section 5.1.5 the systematic uncertainties related to the choice of template cosmology would have been twice larger for $f\sigma_8$.

We add the above uncertainties in quadrature to those derived in Section 5.2.3 to obtain the final RSD modelling systematics: 3.0% on $f\sigma_8$, 0.9% on $\alpha_\parallel$ and 0.8% on $\alpha_\perp$.

5.2.5 Isotropic BAO

We also test the robustness of the isotropic BAO analysis with respect to variations in the HOD and template cosmology. Again, we consider the first set of OuterRim mocks (types 1, 4, 5, 6, with *SHOD* and *HMQ* HODs) presented in Section 5.2.2, apply reconstruction (with the parameters set in Section 5.1.4), and measure their power spectrum using the OuterRim cosmology (2.89) as fiducial cosmology.

Using the isotropic BAO model described in Section 5.1.3, we first find (with the OuterRim cosmology as template cosmology) a damping parameter $\Sigma_{\rm nl}$ value of $8\,{\rm Mpc}\,h^{-1}$ ($4\,{\rm Mpc}\,h^{-1}$) to fit the pre-reconstruction (post-reconstruction) power spectrum. We use these values in the rest of the chapter, unless stated otherwise.

We then perform the post-reconstruction isotropic BAO fits with the different template cosmologies introduced in Section 5.2.4. Results are shown in Figure 5.12. The measured α

2. This systematic uncertainty budget was consistently updated using our prescription for σ_8 discussed in Section 5.1.5 — leading to a minor relative decrease of 4% on the systematic error for $f\sigma_8$.

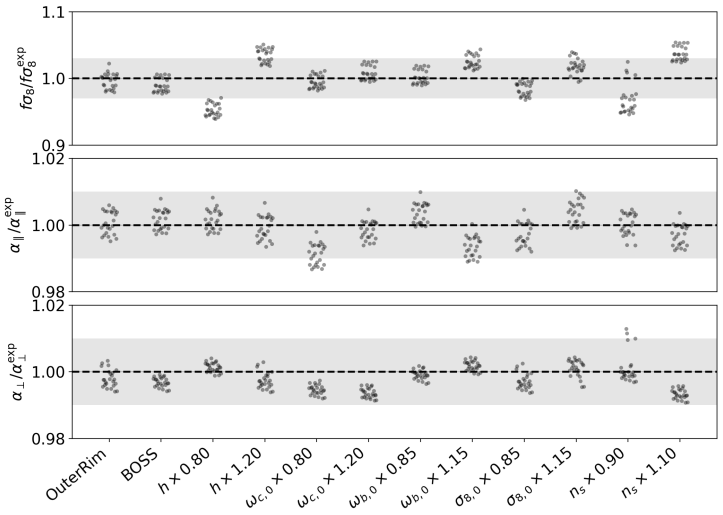

Figure 5.11 – Ratio of the best fit parameters to their expected values for the OuterRim-based mocks of Table 5.2, using x, y, z lines of sight and different template cosmologies. The gray shaded area represents an error of 3% on $f\sigma_8$ and 1% on the scaling parameters on both sides of the reference values in the OuterRim cosmology.

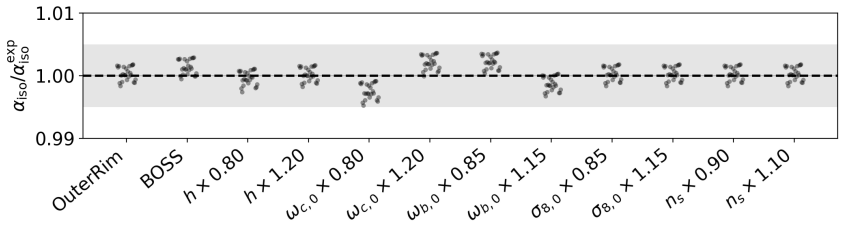

Figure 5.12 – Ratio of the isotropic BAO best fit parameters to their expected values for the OuterRim-based mocks of Table 5.2, using x, y, z lines of sight and different template cosmologies. The gray shaded area represents an error of 0.5% on α on either side of the reference value in the OuterRim cosmology.

value shows very small dependence with the template cosmology, as also found in e.g. Carter et al. (2020). The same is true for the HOD model. Taking the root mean square of differences between best fit and expected values gives a systematic uncertainty of 0.2% on α, which we take as BAO modelling systematics.

In addition, to quantify how typical the data BAO measurements are (see Section 5.5.1), we generate accurate mocks designed to reproduce the ELG sample. An OuterRim box (satellite fraction of 0.14, no velocity bias) is trimmed to the eBOSS ELG footprint, including veto masks and radial selection function, in order to take into account their impact on reconstruction. We then cut 6 nearly independent mocks for NGC and SGC with 3 different orientations. The original box was replicated by 20% to enclose the total SGC footprint. As the ELG density in the OuterRim box is much larger than the observed ELG density, we draw 4 disjoint random subsamples for each sky-cut mock. The number of galaxies in the mock samples match that of the data to better than 1%. Next, in order to have sufficient statistics to characterise the data BAO detection, we randomly generate 1000 fake mock power spectra following a multivariate Gaussian. The Gaussian mean is taken from the pre- and post-reconstruction power spectrum measurements obtained from the above sky-cut OuterRim mocks, and its covariance matrix is

given by the *baseline* EZ mocks. These fake post-reconstruction power spectra will be used in Section 5.5.1 to quantify the probability of the BAO measurements in the data.

We have thus demonstrated the validity of our RSD and BAO theoretical models and derived a related (small) systematic budget. Another part of the systematic budget will stem from observational effects. To this end, we explain the generation of data clustering catalogues in the next section.

5.3 Data and mock catalogues

We first detail in Section 5.3.1 the generation of eBOSS ELG clustering catalogues. We review observational systematics and how they were corrected for at the catalogue level. Based on these results, we detail how we implemented systematics into raw mock catalogues, in order to test the correction schemes and the analysis pipeline as a whole. Such tests will be discussed in the next section.

5.3.1 Generation of eBOSS ELG clustering catalogues

As explained in Section 4.1.6, clustering analyses rely on density fluctuations: the observed data density is compared to the survey selection function, which is sampled with a random synthetic catalogue (as in any BOSS and eBOSS analysis). To do so, we have to produce both data and random clustering catalogues, such that the expected data density is given by the density of the random catalogue. Therefore, observational systematics, i.e. non-cosmological fluctuations in the galaxy density, can be either corrected for in the data or the random catalogues.

5.3.1.1 Veto masks

Ignoring any masks or systematics, one would expect the galaxy density to be isotropic in average. Therefore, the random catalogue is first generated with a uniform random sampling of the sphere, with a density of $1 \times 10^4\,\mathrm{deg}^{-2}$ (Raichoor et al., 2020). Then, the veto masks introduced in Section 3.2.3 and Section 3.3.1.4, which we recap in Table 5.3 for completeness, are applied to both the data and random catalogues. We recall that bits 1 to 5 were applied to the catalogue of targets before fibre assignment (a few targets are removed due to a slight change in the implementation). Bits 1 to 7 are implemented at the brick level using the `brickmask` code [3], bit 8 with a `HEALPix` (Górski et al., 2005) map and bit 9 to 11 with a `mangle` (Hamilton and Tegmark, 2004) mask.

In addition, we remove two tiles in `eboss22`, where the spectroscopic quality is anomalously low.

These masks are applied prior to any other operation on data and random catalogues.

5.3.1.2 Fibre assignment

As discussed in Section 3.3.1, all targets may not receive a fibre in regions of high density, due to the finite number of fibres per plate. In addition, two targets closer than a fibre collision radius ($62''$ on the sky) cannot not be both allocated fibres from a single plate. Hence, some targets will miss from the final clustering sample, in a way depending on the target density and plate overlaps — which must be accounted for.

3. https://github.com/cheng-zhao/brickmask/releases/tag/v1.0

bit	mask	removed area [deg^2]	removed targets
1	not *grz*	67.2	27
2	*xy* bug	49.7	0
3	recovered `decam_anymask`	210.1	142
4	`tycho2inblob`	4.7	0
5	bright objects	57.6	7
6	Gaia stars	54.0	17456
7	Mira star	12.5	3555
8	imprecise bit 3	0.1	15
9	centerpost	0.6	166
10	TDSS_FES targets	1.3	308
11	bad photometric calibration	72.7	16325
-	`eboss22` low-quality plates	13.9	3123
-	total	436.5	41124

Table 5.3 – Angular veto masks applied to the eBOSS ELG catalogues. Masks 1 to 5 were applied prior to the fibre assignment, when all other masks were applied after spectroscopic observations. Extracted from Raichoor et al. (2020).

The fibre assignment procedure outputs collision groups, i.e. groups of targets which are collided with one another. Due to these collisions, some fraction of these targets could not be assigned a fibre. We thus weight each fibre-assigned ELG by the collision pair weight $w_{\rm cp,i}$ given by the number of targets over the number of valid fibres within each collision group. Invalid fibres are those flagged by `ZWARNING`, as described in Section 3.3.3.2; i.e. they suffered from hardware issues.

Collided targets or targets with invalid fibres are declared *resolved* when they lie in the same collision group as a target with a valid fibre. We then define the tiling completeness as the ratio of the number of resolved targets to the number of targets in each sector, where a sector consists in the area covered by intersecting tiles. As shown in Figure 5.13, the tiling completeness may not reach 1 (i.e. some collision groups do not contain valid fibres) due to invalid fibres or an insufficient density of fibres compared to ELG targets. Hence, the tiling completeness is included in the random systematic weight $w_{\rm sys,i}$. By default, targets with collided or invalid fibres are assigned $w_{\rm cp,i} = 0$.

5.3.1.3 Redshift failures

The redshift determination pipeline detailed in Section 3.3.3 does not provide a reliable redshift (as defined in Section 3.3.3.3) for all spectra. We will see that the probability of redshift failures (i.e. redshifts which are not reliable) depends on the target angular position, resulting in spurious angular contamination signal which has to be removed.

The spectroscopic success rate (SSR) is defined as the ratio of the number of valid redshifts over the total number of spectra but those of spectroscopically-confirmed stars (whose distribution contains no cosmological signal). From now on, we therefore consider all objects, but spectroscopically-confirmed stars, with a valid fibre.

The SSR varies with observing conditions; we therefore express it as a function of the median of the median signal-to-noise ratio across each spectrum measured at each pointing pSN, in each

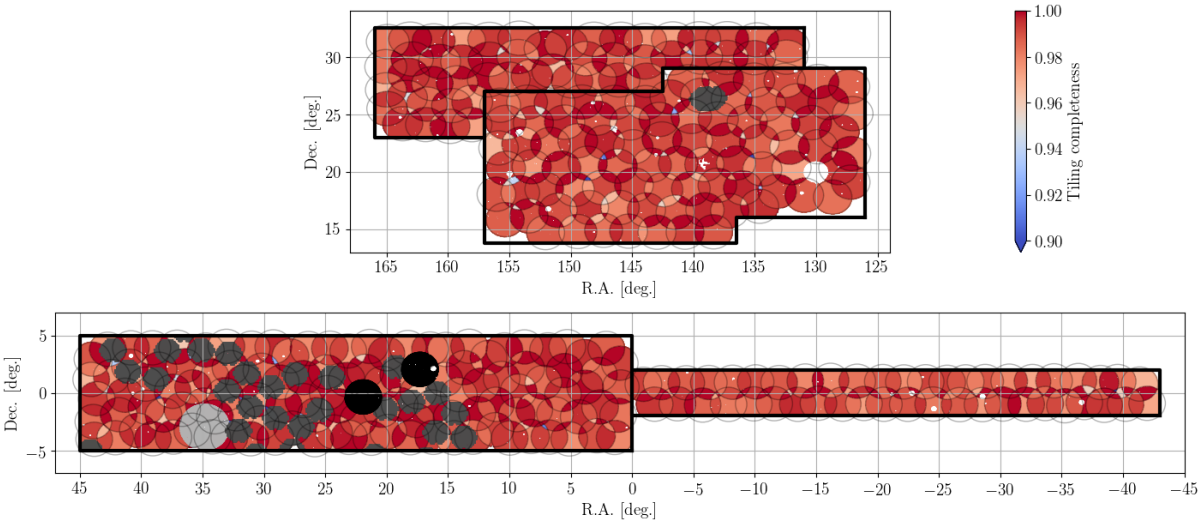

Figure 5.13 – The eBOSS ELG tiling footprint, weighted by the tiling completeness. The light grey disk in the corner at (R.A., Dec.) = (34.84°, −2.98°) corresponds to the Mira star mask (bit 7 in Table 5.3). The dark grey disks mask DECam pointings with bad photometric calibration (bit 11 in Table 5.3), while black disks correspond to the two plates in `eboss22` of low spectroscopic quality. Taken from Raichoor et al. (2020).

chunk (`eboss21`, `eboss22`, `eboss23`, `eboss25`) and half-spectrograph ($1 \leq$ FIBERID ≤ 250, $251 \leq$ FIBERID ≤ 500, $501 \leq$ FIBERID ≤ 750, $751 \leq$ FIBERID ≤ 1000). Indeed, based on Section 3.1.2.3 (especially Figure 3.6), spectrograph 2 ($501 \leq$ FIBERID ≤ 1000) has a slightly higher throughput than spectrograph 1 ($1 \leq$ FIBERID ≤ 500). In addition, half spectrographs are found to have different response: the median signal-to-noise is 0.91 and 0.87 in the first and second halves of spectrograph 1 and 0.94 and 0.88 in the first and second halves of spectrograph 2, respectively. This effect is not well understood. We fit, through a χ^2 minimisation, the following function:

$$f_{\rm noz}^{\rm pSN}({\rm pSN}) = c_0 - c_1 \left|{\rm pSN} - c_2\right|^{c_3} \tag{5.43}$$

to the measured SSR in each subsample.

Variations of the SSR are shown as a function of the median signal-to-noise ratio in Figure 5.14 (grey points). Trends are corrected once galaxies are weighted by the inverse of the fitted Eq. (5.43) relation (in each chunk and half-spectrograph).

Also, we noticed in Section 3.1.2.3 the variation of the resolving power and throughput as a function of the fibre position in the spectrographs. Since fibres are grouped by bundles, which are in average allocated to the same region of the plate, this translates into variations of the SSR as a function of the position in the focal plane, encoded in the variables XFOCAL, YFOCAL. We describe the SSR variations, for each chunk and half-spectrograph, with the following function:

$$f_{\rm noz}^{\rm XYFOCAL}(x, y) = c_0 - c_1 \left|x - c_2\right|^{c_3} - c_4 \left|y - c_5\right|^{c_6} \tag{5.44}$$

where x and y are the centre coordinates of each of the (non-empty) 20×20 XYFOCAL bins. The c_i coefficients are measured separately in each chunk and half-spectrograph. To avoid double counting redshift failures, we weight each galaxy by the inverse of the fitted Eq. (5.43) relation to perform the fit of Eq. (5.44). As shown in Figure 5.15, the SSR decreases neatly near the boundaries of the focal plane, a trend which is well modelled by Eq. (5.44).

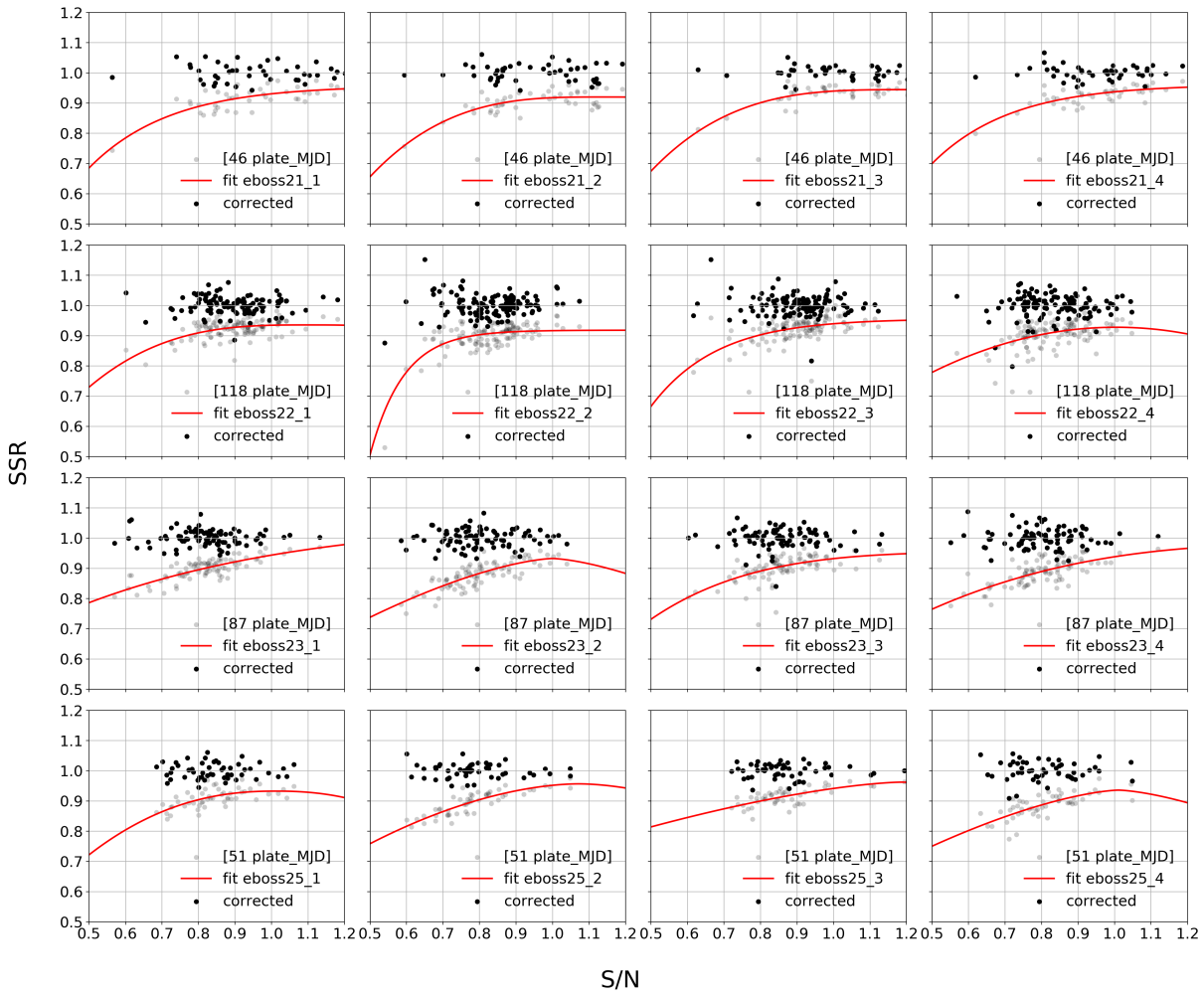

Figure 5.14 – Variations of the SSR as a function of the median signal-to-noise ratio in each chunk (row 1: `eboss21`, row 2: `eboss22`, row 3: `eboss23`, row 4: `eboss25`) and half-spectrograph (column 1: spectrograph 1, first half, column 2: spectrograph 1, second half, column 3: spectrograph 2, first half, column 4: spectrograph 2, second half). Grey points are before correction, black points after weighting by the inverse of the fitted relation of Eq. (5.43).

The final redshift failure weight, applied on the data, is given for each object by:

$$w_{\rm noz,i} = \frac{1}{f_{\rm noz}^{\rm pSN}({\rm pSN}_i) f_{\rm noz}^{\rm XYFOCAL}({\rm XFOCAL}_i, {\rm YFOCAL}_i)}, \tag{5.45}$$

where pSN_i is the median signal-to-noise in the chunk and half-spectrograph, and $\mathrm{XFOCAL}_i, \mathrm{YFOCAL}_i$ the object position in the focal plane. $f_{\rm noz}^{\rm XYFOCAL}({\rm XFOCAL}_i, {\rm YFOCAL}_i)$ is clipped between $[0.5, 1.5]$ to avoid outliers (in practice, this clipping does not impact data weights). Objects that have an unreliable redshift or stars are assigned $w_{\rm noz,i} = 0$.

To obtain the final clustering sample (i.e. the sample to be used for clustering analyses), we apply the following quality cuts:

$$\text{spectroscopic success rate} \geq 0 \tag{5.46}$$

$$\text{tiling completeness} \geq 0.5 \tag{5.47}$$

$$0.6 < z < 1.1 \tag{5.48}$$

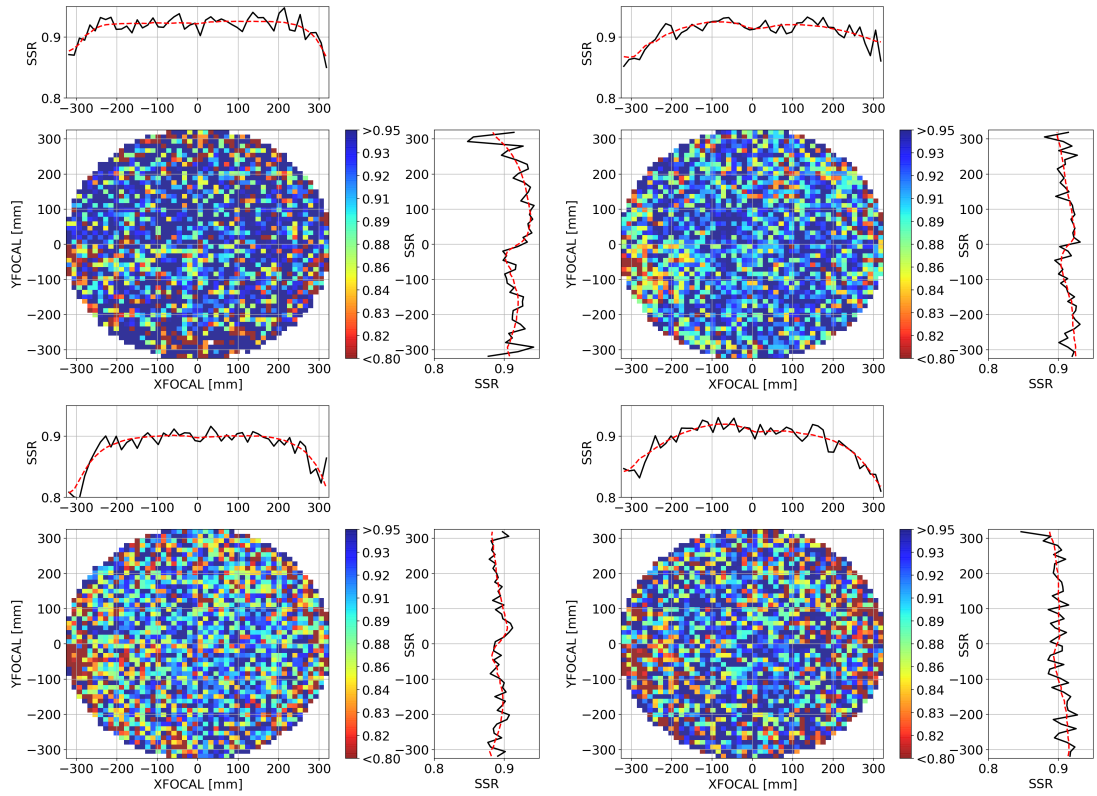

Figure 5.15 – Variations of the SSR (black lines) as a function of the position on the focal plane (top left: `eboss21`, top right: `eboss22`, bottom left: `eboss23`, bottom right: `eboss25`). The fitted model of Eq. (5.44) is shown in red lines.

The two first criteria are applied at the sector level. The first cut requires that each sector contains at least one galaxy with a reliable redshift (which effectively removes one galaxy with a bad redshift in the whole ELG sample). The second cut has no practical effect on the eBOSS ELG data, since all sectors have a tiling completeness greater than 0.5. The same cuts are applied to the random catalogue.

5.3.1.4 Angular photometric systematics

After correcting for fibre incompleteness and redshift failures, remaining non-cosmological variations of the angular density of galaxies in the final clustering sample are caused by fluctuations in the target density due to varying photometric conditions.

We bin photometric parameters in a `HEALPix` map with nside $= 256$ (pixel area of $\simeq 188\,\mathrm{arcmin}^2$). We consider galactic extinction and dust temperature (Schlegel et al., 1998), HI column density (Lenz et al., 2017), Gaia stellar density (stars from DR2 (Gaia Collaboration et al., 2018) with magnitude $12 < g < 17$), DECaLS [4] 5σ depth (galdepth_g,r,z) and seeing (psfsize_g,r,z) in the g, r, z bands.

Random objects and galaxies from the final clustering sample (see Eq. (5.48)), including weights (tiling completeness for the randoms, $w_{\mathrm{cp,i}}$ and $w_{\mathrm{noz,i}}$ for the data) are binned in this `HEALPix` map; we call n_{dat} and n_{ran} the resulting per-pixel quantities. n_{ran} is normalised so that its sum matches the sum of n_{dat} in each chunk (`eboss21`, `eboss22`, `eboss23`, `eboss25`)

4. http://legacysurvey.org/dr3/description/

and represents the expected density of targets in the absence of photometric systematics (and clustering). Then, the density $n_{\rm dat}$ is modelled as the multilinear model:

$$f_{\rm sys}(i_{pix}) = n_{{\rm ran},i_{pix}} \left(\epsilon + \sum_{t \in \mathcal{T}} c_t t_{i_{pix}} \right) \tag{5.49}$$

which is fitted in each chunk to the per-pixel $n_{\rm dat}$ through a χ^2 minimisation. We consider a diagonal covariance matrix with Poisson errors $\sigma_{i_{pix}} = \sqrt{n_{{\rm ran},i_{pix}}}$. We checked that minimising a log Poisson likelihood instead does not change our results, and verified with mocks that a Gaussian likelihood with diagonal covariance matrix is overall a good approximation. The set of photometric templates $\mathcal{T}$ is: galactic extinction and logarithm of the dust temperature, log of HI column density, stellar density, seeing and depth flux. The photometric weight is then given by:

$$w_{\rm sys,i} = f_{\rm sys}(i_{pix}) \tag{5.50}$$

calculated on the `HEALPix` map. As an example, Figure 5.16 shows the variations of the ELG NGC angular density with respect to the stellar density and seeing in g-band, with and without applying the angular weight (5.50). As stellar density increases, the probability for a galaxy to be blended with a star increases. Hence the ELG density diminishes with stellar density, as shown in Figure 5.16 (left). When photometric conditions are poorer, we also expect the ELG density to decrease, explaining the trends with the seeing on Figure 5.16 (right). We see that the weight $w_{\rm sys,i}$ correctly removes these trends.

The mean of photometric weights $w_{\rm sys,i}$ over all ELG targets is then normalised to 1 in each chunk.

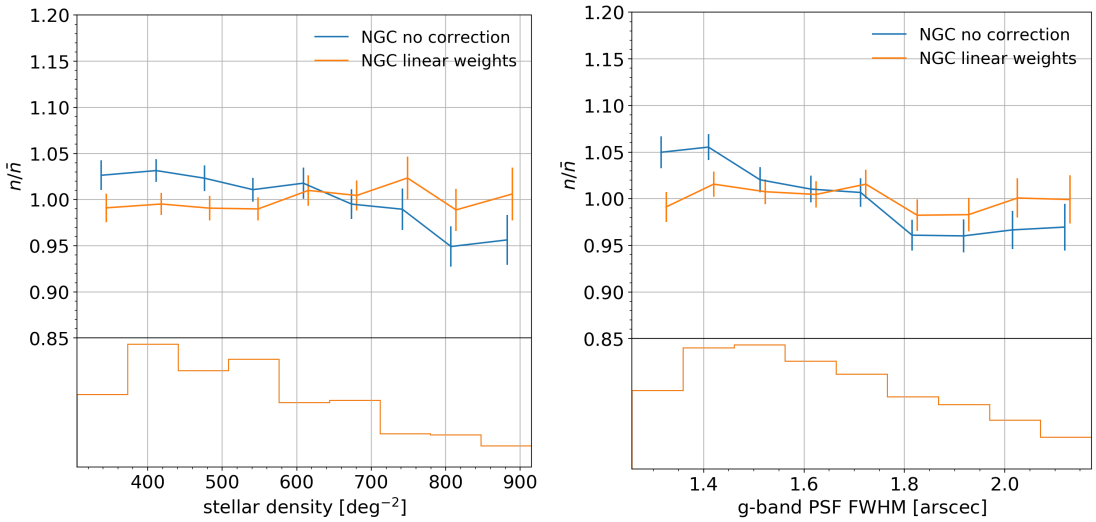

Figure 5.16 – Variations of the (weighted) ELG NGC angular density, calculated as the number of weighted data over weighted randoms (in the final clustering sample, see Eq. (5.48)) in bins of the template p, with p the stellar density (left) and the seeing in g-band (right). Variations are shown before (blue) and after (orange) the additional weights (5.50) are applied. Error bars are given by the dispersion, scaled by the square root of the sum of FKP weights in each bin to take into account cosmic variance, as pointed out by Ross et al. (2017). Bottom panels display the distribution of ELG as a function of stellar density (left) and seeing in g-band (right).

Note that by definition such multilinear regression nulls the correlation between (weighted) ELG density and the considered set of photometric templates. In other words, this method im-

poses angular integral constraints, in a very similar manner as discussed in Section 4.3, and hence biases clustering measurements on large scales. This effect has been checked to be subdominant for the eBOSS ELG RSD and BAO analyses.

Let us mention two other attempts to model photometric systematics. Rezaie et al. (2020) aimed at modelling photometric systematics using neural nets. They constructed their target catalogue applying the eBOSS ELG NGC target selection to a later data release of DECaLS (DR7). Their method is more powerful in that it can account for complex, non-linear relations between the angular galaxy density field and photometric templates. Also, overfitting (leading to the angular integral constraints discussed above) is mitigated by performing cross-validation (neural nets are trained and assessed on different data sets) and elimination of useless templates. Their method yields quite similar results to the simple multilinear regression discussed above for the specific case of the eBOSS ELG footprint, using specific DECaLS DR3 and DR5 data. Another approach, called `Obiwan` and detailed in Kong et al. (2020) consists in injecting fake galaxies into DECam images and process them through `legacypipe` (see Section 3.2.2.5) and the eBOSS ELG target selection, in order to assess the response of the whole imaging pipeline, including the software implementation. This method is free of the angular constraints discussed above. Also, it is very useful in general to check for possible biases in the imaging pipeline. `Obiwan` was run on the eBOSS ELG `eboss23` footprint and provided similar results as the multilinear regression presented above.

5.3.1.5 Completeness weights

The data completeness weights, including corrections for photometric systematics ($w_{\rm sys,i}$), fibre collisions ($w_{\rm cp,i}$) and redshift failures ($w_{\rm noz,i}$) are given by:

$$w_{\rm comp,i} = w_{\rm sys,i} w_{\rm cp,i} w_{\rm noz,i} \tag{5.51}$$

while random weights are simply $w_{\rm comp,i} = w_{\rm sys,i}$ ($w_{\rm cp,i}$ and $w_{\rm noz,i}$ being set to 1) with $w_{\rm sys,i}$ of the random catalogue accounting for the tiling completeness (see Section 5.3.1.2) — unlike those of the data catalogue. By convention, data completeness weights are normalised such that their sum reflects the number of redshifts one would have obtained without inefficiencies due to fibre collisions and redshift failures (tiling incompleteness being accounted for in the randoms). Hence, $w_{\rm noz,i}$ is rescaled such that the mean of the data completeness weights $w_{\rm comp,i}$ of ELGs with a reliable redshift or stars (for them and for this calculation we take $w_{\rm noz,i} = 1$) is equal to the mean of $w_{\rm sys,i}$ over all resolved targets. This normalisation convention is useful to specify the initial redshift density required in mocks such that the final number of mock objects is not lower than the number of data galaxies after the aforementioned inefficiencies are implemented.

5.3.1.6 Depth-dependent redshift density

The random angular density is designed to reproduce the expected angular ELG density after corrections for fibre collisions, redshift failures and photometric systematics are included. However, the random catalogue still misses a redshift information. As already discussed in Section 4.3.2, the radial selection function is hard to model. Hence, as in any SDSS clustering analysis, the redshift distribution is directly transferred from the data to the random catalogue, using the *shuffled* scheme (Samushia et al., 2012; Ross et al., 2012; Reid et al., 2016): randoms redshifts are randomly picked from the data redshift distribution. If the radial selection function

was expected to be isotropic, the latter operation could be performed on the whole footprint. However, as we discuss in the following, the eBOSS ELG radial selection function varies on the sky.

The eBOSS ELG programme targeted quite faint targets, even for DECaLS imaging, which results in a significant uncertainty in the magnitude estimation. Hence, targets are expected to scatter in and out of the selection box. Especially, as described in Section 3.2.3, we can expect some targets from low redshifts to scatter in the selection box, such that the density at low redshift will be higher in regions of shallower imaging. We can also anticipate this effect to be larger in the NGC where the depth is 0.5 mag lower, despite the more conservative, redder, cut of the selection box presented in Section 3.2.3.

Hence, the ELG redshift distribution is expected to change with imaging depth. This is shown in Figure 5.17, displaying the redshift density of `eboss23` binned into different intervals in depth. Since imaging depth is not constant on the sky (as could be expected from the variations of number of photometric observations presented in Figure 3.11), the radial selection function varies with the angular position. This effect was first noticed in the eBOSS ELG sample, since other SDSS programmes targeted brighter targets, with a more isotropic imaging, as permitted by the drift-scan strategy of the SDSS imaging survey (see Section 3.1.1.1).

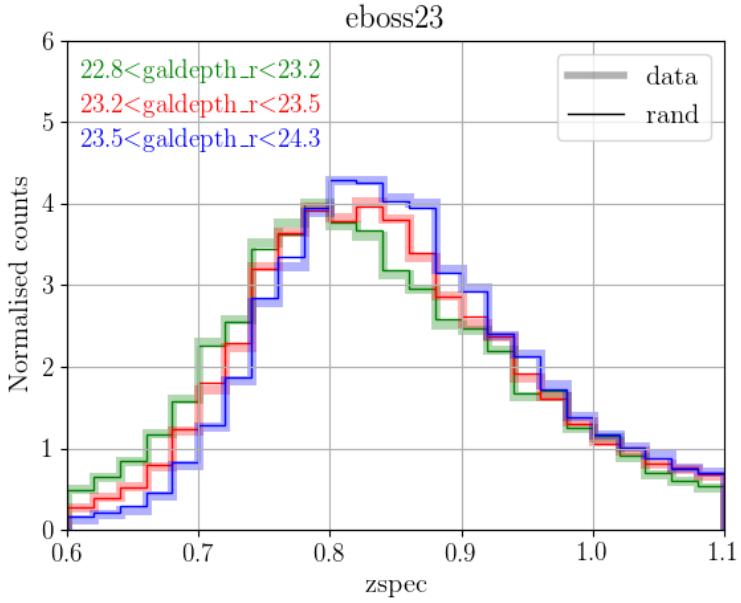

Figure 5.17 – Redshift density of `eboss23`, binned in three intervals of depth in r-band. Data are shown in thick lines, randoms in thin lines. Taken from Raichoor et al. (2020).

We describe the depth variations with a single variable f_{grz}, which is a linear combination of the depth flux f_b in the g, r and z bands:

$$f_{grz} = \epsilon + \sum_{b \in \{g,r,z\}} c_b f_b. \tag{5.52}$$

The coefficients ϵ and c_b of this linear combination are fitted in each chunk (`eboss21`, `eboss22`, `eboss23`, `eboss25`) through a χ^2 minimisation (taking into account data completeness weights) comparing each ELG redshift to f_{grz} at its angular position. We note that the redshift density was found have smaller dependence with respect to other photometric templates, the largest

one being the seeing in g and r bands in `eboss23`, which is already accounted for by the f_{grz} dependence.

Random objects of each chunk are split into 3 quantiles of f_{grz}. The same intervals are used to bin ELG data. Then, data redshifts in the final clustering sample are randomly assigned to random objects in each of these intervals (dubbed `chunk_z`), with a probability proportional to their completeness weight $w_{\rm comp,i}$, following the *shuffled* scheme. Note that such a chunk-splitting increases the radial integral constraint discussed in Section 4.3.6, since the redshift distribution of the random catalogue is tuned to match that of the data in smaller chunks.

Finally, random weights $w_{\rm sys,i}$ are normalised to ensure that the sum of weighted data over the sum of weighted randoms is the same in each `chunk_z`.

5.3.1.7 Calculation of redshift density and FKP weights

Data redshift density is calculated in each chunk, by binning weighted data from the final clustering sample (without redshift cut) into redshift slices of size $\Delta z = 0.005$, starting at $z = 0$. The comoving volume within the shell delimited by redshifts $z_1 < z_2$:

$$V(z_1, z_2) = \frac{\mathcal{A}}{3}\left[\chi(z_2)D_{\rm M}(z_2)^2 - \chi(z_1)D_{\rm M}(z_1)^2\right] \tag{5.53}$$

is calculated assuming the fiducial cosmology (5.4).

The effective area $\mathcal{A}$ used for the calculation is given by the number of randoms in the final clustering sample, weighted by the tiling completeness. As mentioned in Section 5.3.1.4, such a definition ensures that the redshift density $n(z)$ matches the one we would have obtained without spectroscopic (fibre assignment and redshift determination) incompleteness, which is convenient for the mock generation process described below. However, it does not necessarily provide a correct normalisation for the Yamamoto estimator, which must be accounted for in the model, as discussed in Section 4.2.2.3.

The redshift density $n(z)$ is evaluated at each data and random redshift z by linear interpolation between the mean (or centre, if the bin is empty) of each redshift bin. FKP weights (see Eq. (4.39)):

$$w_{\rm FKP,i} = \frac{1}{1 + n(z_i)P_0} \tag{5.54}$$

are assigned to both data and randoms, using $P_0 = 4000\,{\rm Mpc}^3\,h^{-3}$, close to the measured power spectrum monopole at $k \simeq 0.1\,h\,{\rm Mpc}^{-1}$. As mentioned in Section 4.3.8, these FKP weights may bias clustering measurements since they overweight (downweight) underdensities (overdensities), but this systematic bias has been checked to be small enough for our analysis.

The total weights, to be applied on both data and random objects are given by:

$$\boxed{w_{\rm tot,i} = w_{\rm FKP,i}w_{\rm comp,i}} \tag{5.55}$$

with $w_{\rm comp,i}$ as given by Eq. (5.51).

We shall next check the robustness of the correction schemes detailed above and the analysis pipeline as a whole to the aforementioned systematics, using mock catalogues, which are also used to estimate the power spectrum covariance matrix.

5.3.2 Approximate mocks

Performing a data analysis requires estimating the likelihood of the observed data (compressed statistics). In particular, the likelihood of the observed power spectrum or correlation function is generally assumed to be Gaussian, such that it is fully specified by its mean (the theoretical model) and a covariance matrix.

A first solution to compute these covariance matrices is to rely on analytic methods. Simplest covariance matrices assume Gaussian statistics (e.g. Grieb et al., 2016), some include perturbation theory calculation (e.g. Wadekar and Scoccimarro, 2019) or partial tuning on jacknife estimates (e.g. Philcox et al., 2020). Though these methods are promising in providing reliable covariance matrices, it appears more difficult to include observational systematic uncertainties. In general, mocks remain useful to test for potential bias due to observational systematics.

Since N-body mocks are computationally costly to produce (see Section 2.2.1), one has to resort to faster methods (see Chuang et al. (2015) for a comparison of different methods). They include Effective Zel'dovich approximation mocks (EZ, Chuang et al., 2015), Fast Particle Mesh (FastPM, Feng et al., 2016), GaLAxy Mocks (GLAM, Klypin and Prada, 2018), PerturbAtion Theory Catalog generator of Halo and galaxY distributions (PATCHY, Kitaura et al., 2014), and Quick Particle Mesh (QPM, White et al., 2014a). FastPM, GLAM, and QPM are fast particle mesh solvers. Instead, EZ and PATCHY evolve density fields generated by perturbation theories and tracers are sampled using an effective bias model.

In this analysis, we consider two sets of approximate mocks, EZ and GLAM-QPM. The generation of raw catalogues, i.e. without observational systematics, is described in the following.

5.3.2.1 EZ mocks

Details about the generation of the 1000 systematic-free EZ mocks can be found in Zhao et al. (2020). We briefly recap the generation steps hereafter for completeness. The matter density field of EZ mocks is evolved through the Zel'dovich approximation (see Section 2.1.4), which is easily solved in Fourier space using FFTs. This field is then populated with tracers based on a parameterised bias model. This effective bias prescription includes linear, non-linear, deterministic, and stochastic effects, which can be calibrated with the 2- and 3-point clustering statistics from observations or N-body simulations. Chuang et al. (2015) showed that the redshift space power spectra of EZ mocks are accurate to the 5% level for $k \lesssim 0.3\,h\,\mathrm{Mpc}^{-1}$, for a low computational cost.

For the eBOSS analyses, Zhao et al. (2020) generated boxes of side length $5\,\mathrm{Gpc}\,h^{-1}$ and grid size 1024^3, with the same number of dark matter particles. These particles are evolved through the Zel'dovich approximation; we note the resulting dark matter density field ρ_m. The fiducial cosmology of these mocks is that of the MultiDark simulation [5] (except for σ_8), i.e. flat ΛCDM with [6]:

$$\begin{aligned} h = 0.6777,\ \Omega_{m,0} = 0.307115, \Omega_{b,0} = 0.048206, \\ \sigma_{8,0} = 0.8225,\ n_s = 0.9611. \end{aligned} \tag{5.56}$$

5. https://www.cosmosim.org/cms/simulations/mdpl2/

6. Note that n_s is rounded to 0.96 on the MultiDark webpage; private email from Cheng Zhao on April 2, 2019.

The final ELG number density is $6.4 \times 10^{-4}\, h^3\,\mathrm{Mpc}^{-3}$. The galaxy density field follows the relation:

$$\rho_g = \theta(\rho_m - \rho_c)\rho_{\mathrm{sat}} \left[1 - \exp\left(-\rho_m/\rho_{\mathrm{exp}}\right)\right] B_s. \tag{5.57}$$

θ is a step function, with $\theta(x \geq 0) = 1$ and $\theta(x < 0) = 0$, such that halos (galaxies) only form above some critical density ρ_c (see Section 2.3.1). The exponential term ensures saturation of the galaxy density with the dark matter density ρ_m. B_s is a stochastic bias, following a Gaussian distribution above 1 and a lognormal distribution below, of mean 0 and standard deviation 10. Due to its degeneracy with B_s, ρ_{sat} is fixed to an arbitrary value.

The probability for each grid cell to contain n_g galaxies is set to a power law $\propto b^{n_g}$ $(b < 0 < 1)$. Given the target total number of galaxy one obtains the number of cells $n_{\mathrm{cell}}(n_g)$ to be populated with n_g galaxies. Then, the maximum number of galaxies in one cell $n_{g,\mathrm{max}}$ is assigned to the $n_{\mathrm{cell}}(n_{g,\mathrm{max}})$ cells where ρ_g of Eq. (5.57) is highest. The process is repeated for decreasing n_g values. Galaxies are placed at the location of dark matter particles if there are any, else are given uniform random locations inside the cell. Galaxies are assigned velocities as predicted by linear theory, with an additional isotropic Gaussian random variable of mean 0 and standard deviation v.

NGC and SGC are spatially far away and are thus generated from independent initial conditions. Cubic boxes are remapped (an operation which preserves the periodicity of the galaxy density field, see Carlson and White 2010) into a parallelepipoid of sides $5/\sqrt{2}\mathrm{Gpc}\,h^{-1}$ and $5\sqrt{2}\mathrm{Gpc}\,h^{-1}$, such that all eBOSS tracers are contained inside it (for cross-correlation studies). Approximate light cones are built by concatenating different snapshots with redshift ranges $[0.6, 0.7]$, $[0.7, 0.75]$, $[0.75, 0.8]$, $[0.8, 0.85]$, $[0.85, 0.9]$, $[0.9, 1.0]$, and $[1.0, 1.1]$.

The `make_survey` software (White et al., 2014b) is used to rotate and trim the cubic boxes to the ELG footprint. The observed redshift, including redshift space distortions is given by:

$$z_{\mathrm{conf}} = z_{\mathrm{real}} + (1 + z_{\mathrm{real}})\,\frac{v}{c}, \tag{5.58}$$

with v the peculiar (proper) galaxy velocity along the line-of-sight, c the speed of light, z_{real} the real redshift (corresponding to the real space comoving distance).

The mocks are first trimmed following the ELG tiling geometry [7] using the `mangle` software.

Bias parameters (ρ_c, ρ_{exp}, b, v) of these light-cone mocks are calibrated on the observed data 2- and 3-point statistics in 4 overlapping redshift bins and their value is extrapolated to any effective redshift using a second order Taylor expansion in redshift.

5.3.2.2 GLAM-QPM mocks

The generation of the 2003 raw (systematic-free) GLAM-QPM mocks is discussed in Lin et al. (2020); we briefly recap the main steps of the mock creation in the following. The matter density field is generated with GaLAxy Mock (GLAM) simulations and is sampled by halos using the Quick Particle-Mesh (QPM) method. Halos are then populated with galaxies following a HOD approach.

GLAM simulations evolve 1500^3 particles in a box of side length $3\,\mathrm{Gpc}\,h^{-1}$, hence with a mass resolution of $6.8 \times 10^{11}\,\mathrm{M}_\odot\,h^{-1}$, starting at $z = 100$, with 94 time steps, increasing with

7. https://data.sdss.org/sas//ebosswork/eboss/ebosstilelist/trunk/outputs/eboss21/geometry-eboss21.ply

redshift. The following cosmology is assumed:

$$h = 0.678,\ \Omega_m = 0.307, \omega_b = 0.022,$$
$$\sigma_{8,0} = 0.828,\ n_s = 0.96. \tag{5.59}$$

Then, dark matter particles are binned into 8 equally spaced bins of $\mu = \ln(1+\delta)$, and the bias is measured in each bin. Bias b is related to halo mass M_h following Tinker et al. (2008), allowing a mapping between μ and halo mass M_h to be determined. Then, particles with a local density of μ_0 are turned into halos of mass M_h, with a Gaussian PDF of mean $\mu(M_h)$ and standard deviation 0.1, with the constraint that the target halo mass function of Tinker et al. (2010) should be reproduced.

Then, these halos are populated with galaxies following a HOD. The radial distribution of satellites is assumed to follow a NFW profile, with concentration-mass relation as given in Macciò et al. (2007). The mean HOD shape is the same as Eq. (5.35) for centrals and Eq. (5.29) for satellites. Parameters were fitted to the data using a coarse grid search (fixing the completeness and mass scale σ). The satellite fraction is found to be $\simeq 17.4\%$, in agreement with e.g. Favole et al. (2016). The obtained galaxy number density of $6.7 \times 10^{-4}\, h^3\,\mathrm{Mpc}^{-3}$. The best fit HOD is shown in Figure 5.18 (top left) and is compared to other HODs found in the literature (Gonzalez-Perez et al., 2018; Guo et al., 2019) (see Section 2.2.3). The same figure shows the projected correlation function (top right), convenient to ignore RSD effects (see Davis and Peebles, 1983), and the correlation function monopole and quadrupole (bottom). Compared to ELG measurements, GLAM-QPM mocks lack clustering around $s \simeq 1\,\mathrm{Mpc}\,h^{-1}$, at the transition between the one halo and two halo terms.

Mocks are trimmed to the tiling geometry; then, veto masks and the radial selection function measured in chunks are applied.

5.3.3 Generation of EZ mock catalogues

We first implement observational systematics in the raw EZ mock catalogues described above.

5.3.3.1 Depth-dependent redshift density

Raw mocks have a constant density. We thus have to imprint the radial selection function of the observed data, which, as seen in Section 5.3.1.6, depends on imaging depth.

First, the reference `chunk_z` is assigned to any object in both mock data and randoms by taking the `chunk_z` value of the nearest object in the same chunk of the eBOSS ELG random catalogue. Both mock data and randoms are downsampled to match the redshift density of the final eBOSS ELG clustering sample (rescaled to the mock fiducial cosmology) in each `chunk_z`. As can be seen in Figure 5.19, the data redshift distribution is well reproduced in each `chunk_z`, except the density peak in `eboss22` (top right) which is truncated. This is due to the mock density being slightly too low to incorporate all observational systematics, a major contribution coming from angular photometric systematics.

However, despite this slight mismatch in redshift distribution we still want to keep the same average final number of objects in the mocks as in the data clustering catalogues. Hence, the redshift density is slightly rescaled so that the resulting average (over the 1000 mocks) number of mock galaxies matches the weighted number of objects in the eBOSS clustering sample in each chunk — plus missed targets contributing to the tiling incompleteness. Table 5.4 and Table 5.5

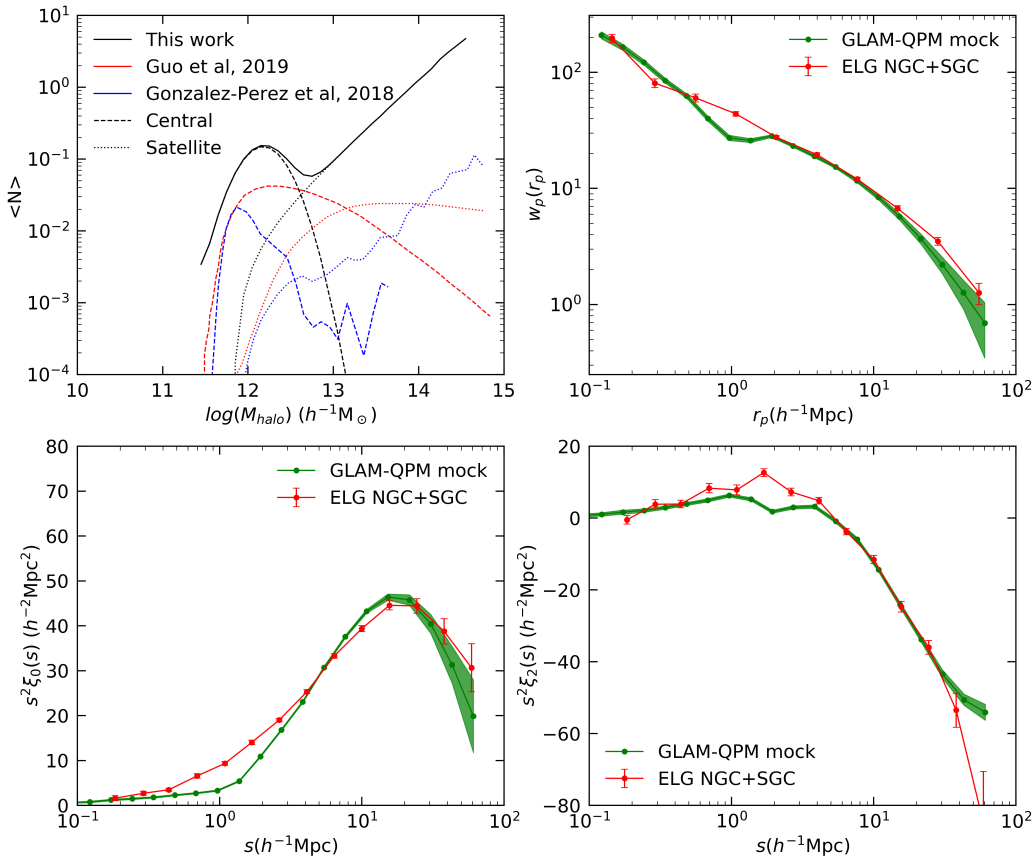

Figure 5.18 – Top left: HOD model for the GLAM-QPM mocks compared to other models in the redshift range $0.8 < z < 0.9$ (Gonzalez-Perez et al., 2018; Guo et al., 2019). Top right: projected correlation function of 100 GLAM-QPM mocks (green) and data (red, averaged over NGC and SGC). Bottom: correlation function monopole (left) and quadrupole (right). Data error bars are jacknife estimates, while the width of the green regions is given by the standard deviation of the mocks.

report the number of objects in the data and mock catalogues. They match very well, at the sub-percent level.

After this first downsampling to reproduce the data redshift density in each `chunk_z`, all objects are kept in mock catalogues, making it possible to compare clustering measurements with and without a specific observational systematics using the same files. For convenience, a veto flag includes all photometric veto masks and cuts to obtain the clustering sample and we adopt the convention that a weight of 0 means that the mock galaxy should not be considered if the corresponding systematics is applied.

5.3.3.2 Veto masks

Photometric veto masks are applied on mock data and randoms the same way as to eBOSS data (see Section 5.3.1.1). To match the correct target density for the fibre assignment (see Section 5.3.3.4), objects from the true ELG data are injected into the mocks. For this we first include all spectroscopically-confirmed stars. This is not sufficient to reach the required target

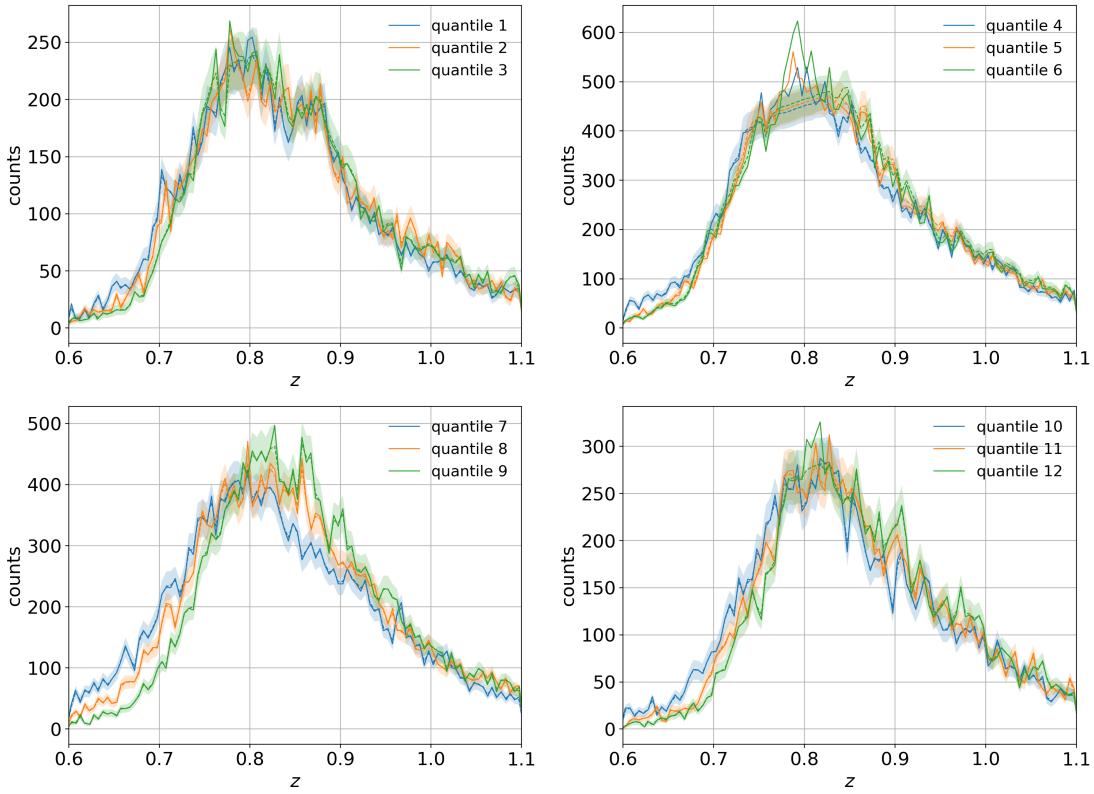

Figure 5.19 – Redshift histograms for each `chunk_z` of the different chunks (top left: `eboss21`, top right: `eboss22`, bottom left: `eboss23`, bottom right: `eboss25`): eBOSS data (continous lines) and mean of EZ mocks (dashed lines), with their dispersion (shaded area). The EZ mocks redshift density is slightly truncated (e.g. the peak in the top right panel) and rescaled to obtain the correct final number of objects.

density, since in the data a number of low or high redshift objects (outside the redshift range $[0.6, 1.1]$) also contribute to the ELG target density.

Hence, a random subsample (changing from mock to mock) of non-star fibre-assigned ELG objects, not rejected by bits 1, 2, 3, 4 and 5 of the veto mask 5.3 — since those were applied before the fibre assignment — is added to the mock sample. Redshifts of the latter objects are randomly reassigned using the sample of non-star eBOSS ELGs with a redshift outside $[0.6, 1.1]$. Note that we do not use directly objects outside the redshift range $[0.6, 1.1]$, as their SSR is biased towards lower values and would thus alter the implementation of redshift failures described in Section 5.3.3.5.

Note that all these objects from the ELG target sample will be discarded by cuts on redshift quality (for stars) or reshift value (for the other eBOSS objects) and hence will not be part of the final mock clustering sample. An additional flag is added to keep track of these real eBOSS objects.

5.3.3.3 Photometric systematics

As noted in Section 5.3.1.4, the eBOSS ELG sample suffers from angular photometric systematics, which we consider as partly unknown. We thus follow the most agnostic approach to implement these systematics.

Mock data objects are downsampled following an `HEALPix` map (nside = 256) built upon the observed angular target density, smoothed with a Gaussian beam of radius 1°. This map

is displayed on Figure 5.20. All real eBOSS objects are kept unmasked, since they were by definition selected as part of the ELG targets.

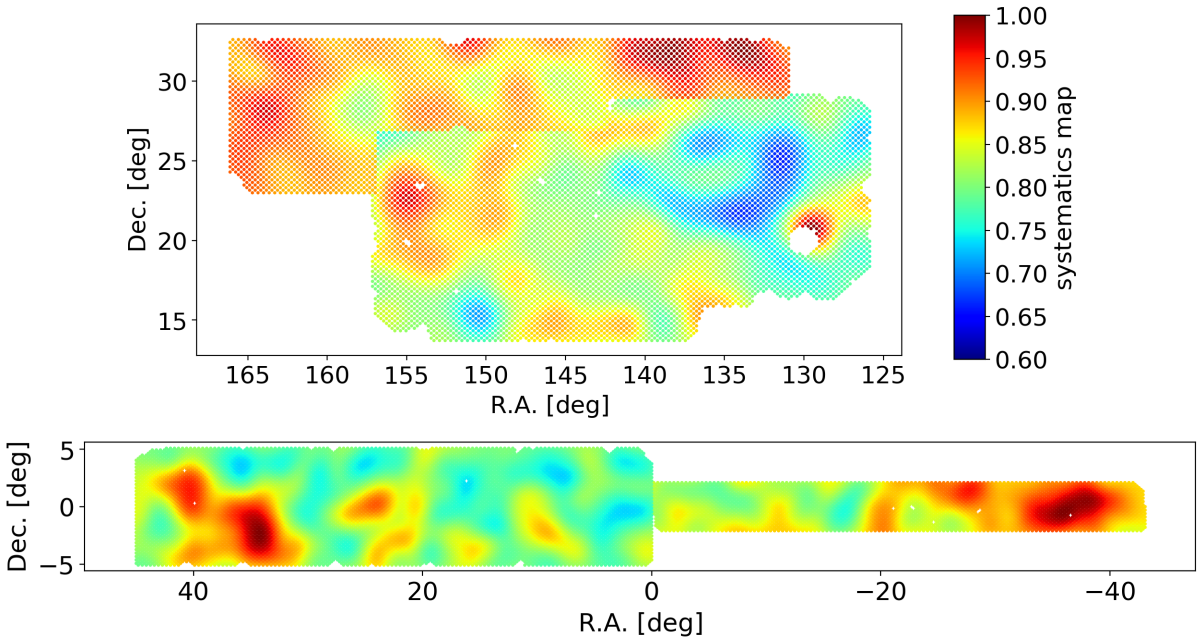

Figure 5.20 – Map of angular photometric systematics used to downsample the mocks (top: NGC, bottom: SGC). This map is obtained by smoothing the data target density with a Gaussian beam of size 1° in each chunk.

5.3.3.4 Fibre collisions

The positions of ELG tiles are transferred from observed to mock data using the tiling files [8]. We do not intend to reproduce the variations of the tile positions with target density resulting from the optimisation described in Section 3.3.1. We indeed expect the potential coupling between data density (and hence clustering) and tiling completeness to be very low for the eBOSS ELG sample, as the tiling density was high.

The implementation of fibre collisions in the mocks is essentially based on the algorithm of Guo et al. (2012) implemented into `nbodykit` (Hand et al., 2018) [9]. In essence, this algorithm forms collision groups with a Friends-of-Friends algorithm of linking length the fibre collision radius, 62″. Then, targets are split between a decollided set (the largest set of targets which are not in collision with one another) and a collided set (targets which are in collision with one of the decollided set). To this end, for each collision group:

— If the multiplicity is 2, one of the two targets is randomly added to the collided set

— Else (if the multiplicity is greater than 2) targets to be added to the collided set are preferentially those with:

1. collisions with the highest number of other targets;
2. collisions with targets that have the lowest number of collisions with other targets;

8. https://data.sdss.org/sas/ebosswork/eboss/ebosstilelist/trunk/outputs/eboss21/
9. https://nbodykit.readthedocs.io/en/latest/api/_autosummary/nbodykit.algorithms.fibrecollisions.html#nbodykit.algorithms.fibrecollisions.fibreCollisions

If several targets are in the exact same conditions, one of them is randomly added to the collided set. Then, the above selections are repeated until all collisions are removed.

We modify this algorithm to deal with different priority targets; namely, targets of lower priority are always added to the collided set first. We call the resulting algorithm `FCP` (fibre collisions with priority).

Just as for the fibre assignment to real ELG targets, we process each chunk separately. We consider all mock objects, except those flagged by bits 1, 2, 3, 4 and 5 of Table 5.3 — as those were applied to the data target catalogues prior to fibre assignement. As noted in Section 3.3.1.4, some targets of the Time Domain Spectroscopic Survey (TDSS, Ruan et al., 2016) ($\simeq 50$ per tile) were tiled at the same time as eBOSS ELGs. A subset of them, TDSS FES targets (206, 451, 469, 280 in chunks `eboss21`, `eboss22`, `eboss23` and `eboss25`, respectively) had the same priority (0) as ELGs, while other RQS targets were targeted with lower priorities (1, 2 and 3). Actually, the tiling was run in 3 successive runs, with the condition that the fibres assigned in previous runs are left untouched. eBOSS ELGs (priority 0) were allocated fibres in the first run, with only TDSS FES (0) and RQS1 (1) targets. Since later runs did not impact the fibre assignment of this first run, we only have to consider the latter for implementation in the mocks.

Hence TDSS FES (RQS1) targets are added to the ELG targets, with their priority 0 (1), to obtain the total sample for fibre assignment. We further take into account the tiling geometry of the eBOSS ELG sample: some fibre collisions can indeed be resolved in regions where multiple tiles overlap (sectors). For this we loop over the minimal number of overlapping tiles, $N_{\rm tiles}$. We start at $N_{\rm tiles} = 1$, with all targets in the collided set (and none in the decollided one). Then, the `FCP` algorithm described above is run on the set of collided targets for which the number of overlapping tiles is $\geq N_{\rm tiles}$. If there are none, the algorithm terminates. Else, `FCP` outputs a set of decollided targets, which is merged with the existing one, and a set of collided targets, which we take as the new one. Collision group identifiers, used to compute $w_{\rm cp,i}$ weights, are set in the first run.

Then, the added TDSS (FES and RQS1) targets are removed from the mock sample (since they are not part of the eBOSS ELG sample). As noted in Section 5.3.1.2, some fibres could not be used to measure spectra due to hardware issues. We thus match each mock ELG with its nearest neighbour in the eBOSS targets (rejecting only bits 1, 2, 3, 4 and 5 of Table 5.3); if the latter received an invalid fibre, it is propagated to the mock object.

Just as for the eBOSS ELG data, by the end of this procedure we obtain mock targets which have been allocated a fibre, some targets that were collided with these and others that received an invalid fibre. In addition, collision groups were determined.

5.3.3.5 Redshift failures

The final source of incompleteness is redshift failures.

We adopt an agnostic scheme to implement redshift failures into the baseline mocks. Each mock ELG (with a valid fibre) receives the spectroscopic properties (redshift reliability, median signal-to-noise ratio, plate and observation date and XFOCAL, YFOCAL) of the nearest ELG in the eBOSS data sample of targets with valid fibres and not spectroscopically classified as stars — as those were already added to the mock sample in Section 5.3.3.2.

To test the impact of such a procedure (which statistically overestimates the impact of redshift failures), we produced a second set of mocks. In this case, mock objects are declared as redshift failures according to the probability $f_{\rm noz}^{\rm pSN} \times f_{\rm noz}^{\rm XYFOCAL}$ (see Eq. (5.43) and (5.44))

of their nearest neighbour in eBOSS data. This procedure is thus stochastic, as opposed to the previous one, which is deterministic (once fibre angular positions are provided).

The clustering measurements on these sets of mocks (statistically differing only through their implementation of redshift failures) will be compared in Section 5.4.

5.3.3.6 Systematic weights and corrections

Once observational systematics have been implemented to the mocks, we can apply to them the correction schemes detailed in Section 5.3.1. For this we use the exact same code as for the eBOSS ELG sample, ensuring a perfectly similar treatment between data and mocks.

Veto masks The same veto flags as in the eBOSS data (see Section 5.3.1.1) are applied before calculation of any correction scheme.

Fibre collisions The `FCP` algorithm outputs collision group identifiers, which are used to compute $w_{\mathrm{cp,i}}$ the same way as in Section 5.3.1.2. The tiling completeness is also calculated and applied to the randoms. Figure 5.21 shows the tiling success rate (TSR), defined as the number of valid fibres over the total number of fibres in each sector, for eBOSS data and one realisation of the EZ mocks. Similarly, Figure 5.22 displays the tiling completeness, which we recall is the ratio of the number of resolved targets to the number of targets in each sector. On both figures, mocks reproduce very well the patterns seen in the data, especially as a function of the number of tile overlaps.

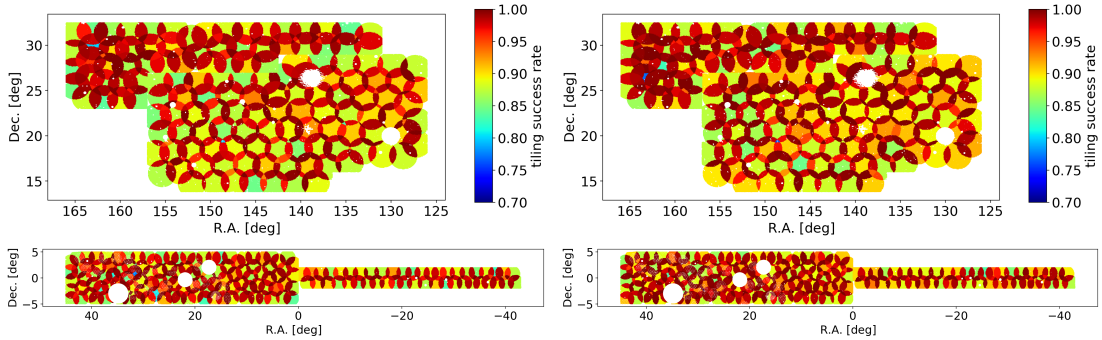

Figure 5.21 – TSR in the eBOSS data (left) and in one realisation of the EZ mocks (right), in NGC (top) and SGC (bottom). The TSR is fully predicted by the `FCP` algorithm.

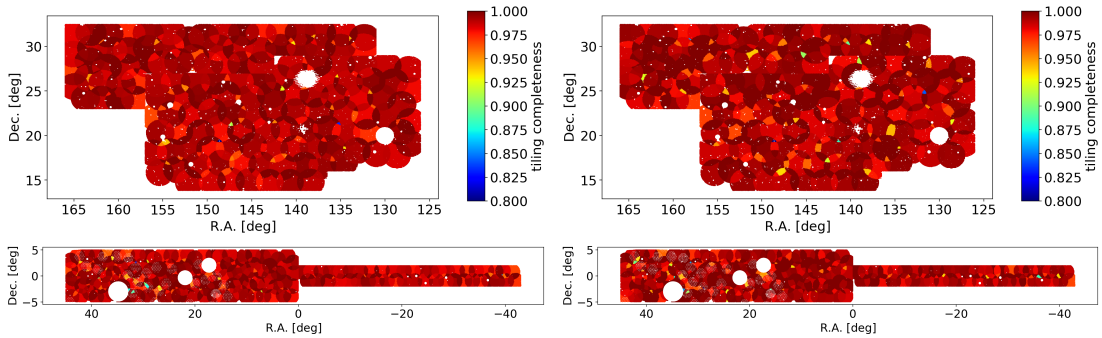

Figure 5.22 – BOSS tiling completeness in the eBOSS data (left) and one realisation of the EZ mocks (right), in NGC (top) and SGC (bottom).

Redshift failures The correction scheme described in Section 5.3.1.3 is applied to mock data using the assigned redshift reliability, median signal-to-noise ratio, plate and observation date and XFOCAL, YFOCAL. The equivalent of Figure 5.14 and Figure 5.15, obtained with the observed data, are Figure 5.23 and Figure 5.24, obtained with one realisation of the EZ mocks. Data and mock variations of the spectroscopic success rate with median signal-to-noise and position on the focal plane look very much alike (as could be expected by construction). In addition, Figure 5.25 displays the variation of the spectroscopic success rate with the fibre identifier — which is correlated to the position on the focal plane. Again, both data and mocks show similar trends, such as a drop in SSR close to 1, 500 and 1000, corresponding to the spectrograph boundaries. Those trends are mostly removed when the weight $w_{\rm noz,i}$ is applied.

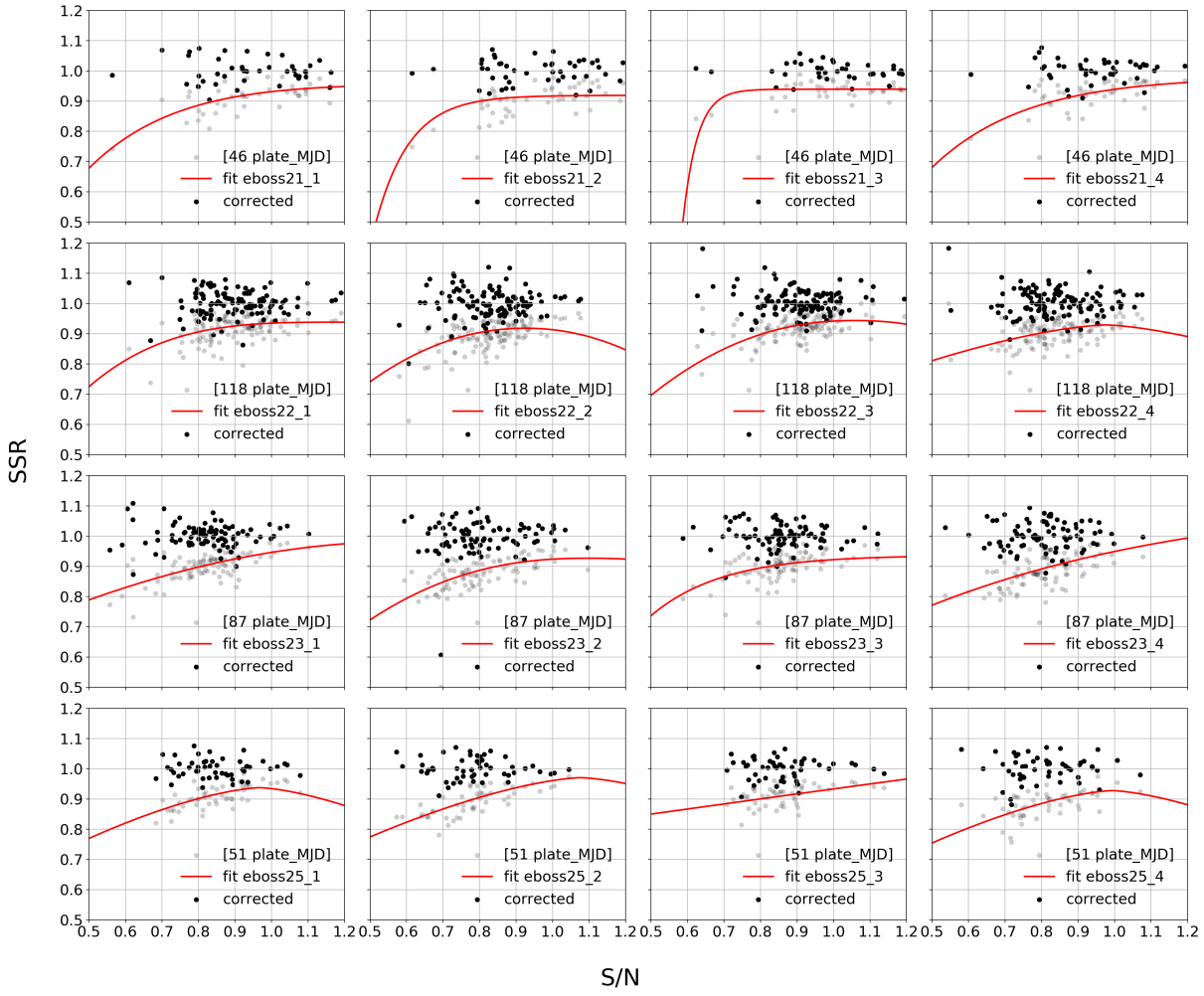

Figure 5.23 – Same as Figure 5.14, for one realisation of the EZ mocks.

Angular photometric systematic The data photometric map is used to bin weighted mock data and randoms. The same correction scheme as in Section 5.3.1.4 is applied to the mocks. Figure 5.26 displays the map of systematic weights for data and one realisation of the EZ mocks. Both data and the mock show similar trends, as a consequence of the data-based systematic map injected into the mocks shown in Figure 5.20. Indeed, regions of low (high) density in Figure 5.20 are attributed high (low) $w_{\rm sys,i}$.

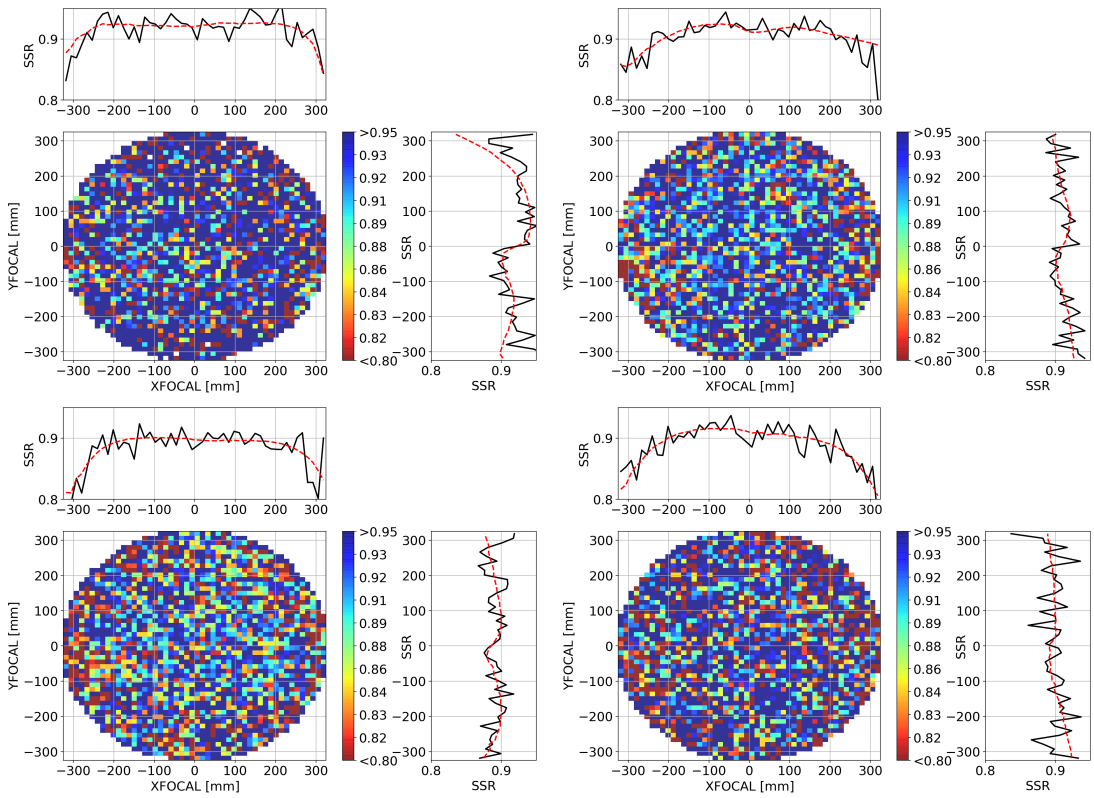

Figure 5.24 – Same as Figure 5.15, for one realisation of the EZ mocks.

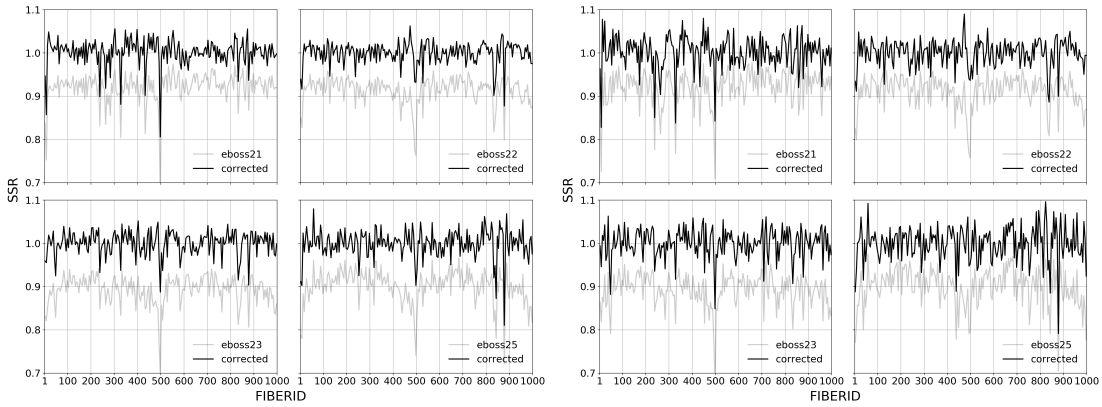

Figure 5.25 – Variations of SSR as a function of fibre ID (shaded: before redshift failures correction, plain: after $w_{\mathrm{noz,i}}$ correction) in the eBOSS data (left) and in one realisation of the EZ mocks (right). We use bins of 5 fibres.

Quality cuts Eq. (5.48) are applied to mock catalogues and weights are normalised the same way as in Section 5.3.1.5.

Depth-dependent redshift density By construction (see Section 5.3.3.1), the radial selection function of mock catalogues varies with imaging depth. Mock randoms and data are first assigned the depth value of their nearest neighbour in the eBOSS randoms. The same correction scheme as in Section 5.3.1.6 is then applied to the mocks. Thus, due to variations in the mock object angular positions and redshifts, the calculated `chunk_z` is not exactly the same as the reference input from the eBOSS ELG sample.

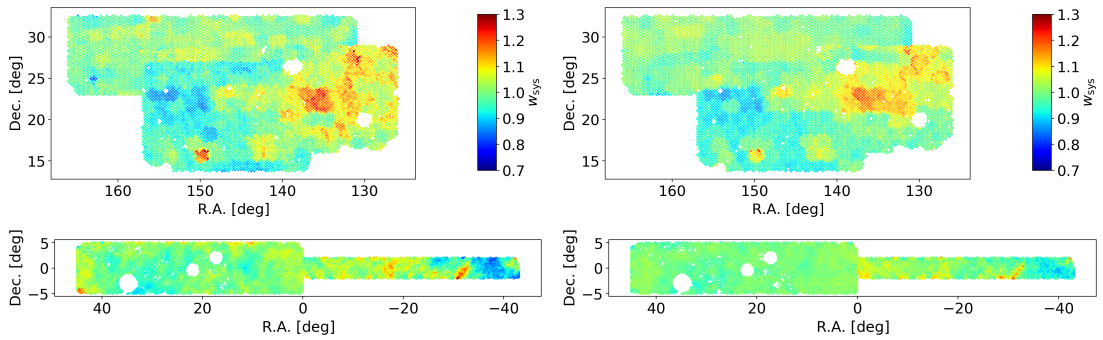

Figure 5.26 – Photometric systematic weights in the eBOSS data (left) and in one realisation of the EZ mocks (right), in NGC (top) and SGC (bottom).

	eboss21	eboss22	eboss23	eboss25	NGC	SGC	ALL
$N_{\rm targ}$	36,314	79,880	70,935	42,565	113,500	116,194	229,694
$N_{\rm obs}$	34,217	76,097	66,221	40,456	106,677	110,314	216,991
$N_{\rm missed}$	408	675	660	358	1,018	1,083	2,101
$N_{\rm cp}$	1,689	3,108	4,054	1,751	5,805	4,797	10,602
$N_{\rm gal}$	31,200	69,071	58,648	36,166	94,814	100,271	195,085
$N_{\rm star}$	333	512	544	315	859	845	1,704
$N_{\rm fail}$	2,684	6,514	7,029	3,975	11,004	9,198	20,202
$N_{\rm used}$	28,029	61,938	51,432	32,337	83,769	89,967	173,736
Mean TSR	0.9412	0.9521	0.9332	0.9496	0.9390	0.9485	0.9437
Mean comp.	0.9887	0.9915	0.9906	0.9915	0.9909	0.9906	0.9908
Mean SSR	0.9207	0.9135	0.8932	0.9016	0.8962	0.9159	0.9059

Table 5.4 – Number statistics of the eBOSS ELG sample. $N_{\rm targ}$ is the number of targets (after all masks 5.3 are applied), $N_{\rm obs}$ the number of targets with valid fibres, $N_{\rm missed}$ the number of targets which were not resolved, $N_{\rm cp}$ the number of collided targets, $N_{\rm gal}$ the number of reliable redshifts, $N_{\rm star}$ the number of spectroscopically-confirmed stars, $N_{\rm fail}$ the number of redshift failures, $N_{\rm used}$ the size of the clustering sample (after quality cuts Eq. (5.48)). *Mean TSR* is the mean tiling success rate (as computed in the random catalogue), *mean comp.* the mean tiling completeness and *mean SSR* the mean spectroscopic success rate.

Calculation of redshift density and FKP weights Redshift density $n(z)$ is calculated the same way as in Section 5.3.1.7, except that the effective area is taken to be constant, equal to that of the eBOSS ELG sample. Indeed, this area is a purely arbitrary choice (it is fully compensated for with an equivalent normalisation of the window functions in the model, see Section 4.2.2.3) and thus should not enter the estimation of the clustering measurement covariance.

Number statistics Various number statistics are provided for the observed ELG data in Table 5.4. We provide the similar quantities for EZ mocks in Table 5.5. Agreement between data and EZ mock number statistics is of the order of 0.5%, thereby confirming observational systematics have been correctly implemented.

Figure 5.27 shows the different systematics and corrections applied successively to the EZ mocks. One can already see that angular photometric systematics (*photo*) are the dominant ones. Another important effect is due to the *shuffled* scheme used to assign redshifts to randoms

	eboss21	eboss22	eboss23	eboss25	NGC	SGC	ALL
$N_{\rm targ}$	36,308	79,883	70,913	42,556	113,469	116,191	229,660
$N_{\rm obs}$	34,437	76,399	66,666	40,621	107,287	110,836	218,123
$N_{\rm missed}$	342	649	599	340	939	991	1,930
$N_{\rm cp}$	1,529	2,835	3,648	1,595	5,243	4,365	9,608
$N_{\rm gal}$	31,432	69,279	59,035	36,267	95,302	100,711	196,013
$N_{\rm star}$	322	499	523	306	828	820	1,649
$N_{\rm fail}$	2,683	6,621	7,109	4,048	11,157	9,304	20,461
$N_{\rm used}$	28,170	61,992	51,612	32,347	83,960	90,161	174,121
Mean TSR	0.9485	0.9564	0.9401	0.9545	0.9455	0.9539	0.9498
Mean comp.	0.9906	0.9919	0.9916	0.9920	0.9917	0.9915	0.9916
Mean SSR	0.9221	0.9133	0.8934	0.9003	0.8960	0.9161	0.9062

Table 5.5 – Mean number statistics of the 1000 EZ mocks. See Table 5.4 for the definition of the various quantities and a comparison to the statistics of the observed data.

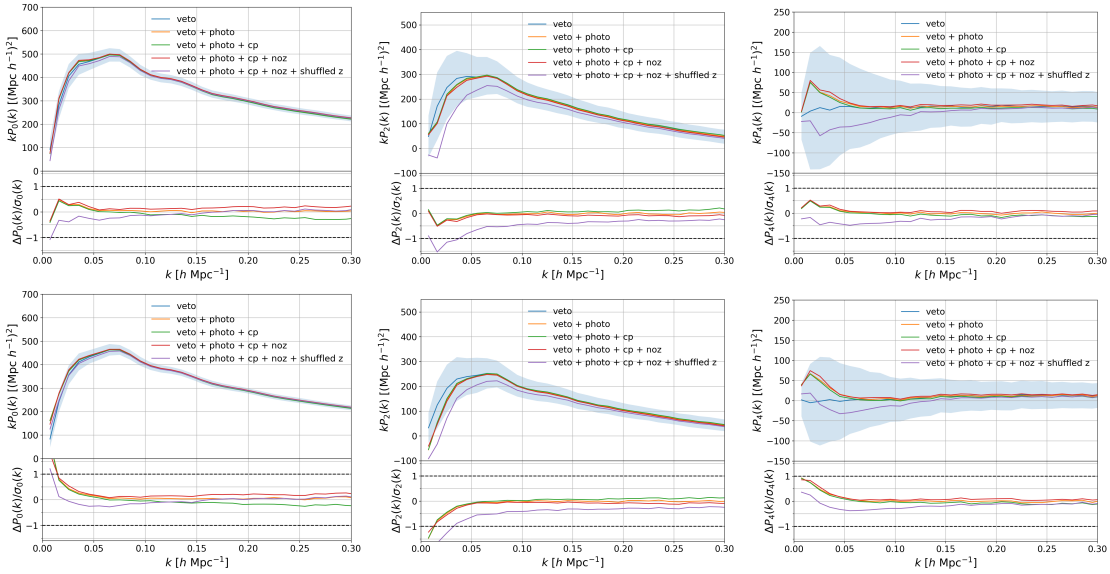

Figure 5.27 – Power spectrum measurements (left: monopole, middle: quadrupole, right: hexadecapole, top: NGC, bottom: SGC) of the EZ mocks, with different systematics and corrections applied successively. The blue shaded region represents the standard deviation of the mocks with veto masks only. Bottom panels: difference of the various schemes to the reference (with veto flag only), normalised by the standard deviation of the mocks. Note that redshift failures, as implemented in the EZ mocks, partially cancel the effect of fibre collisions.

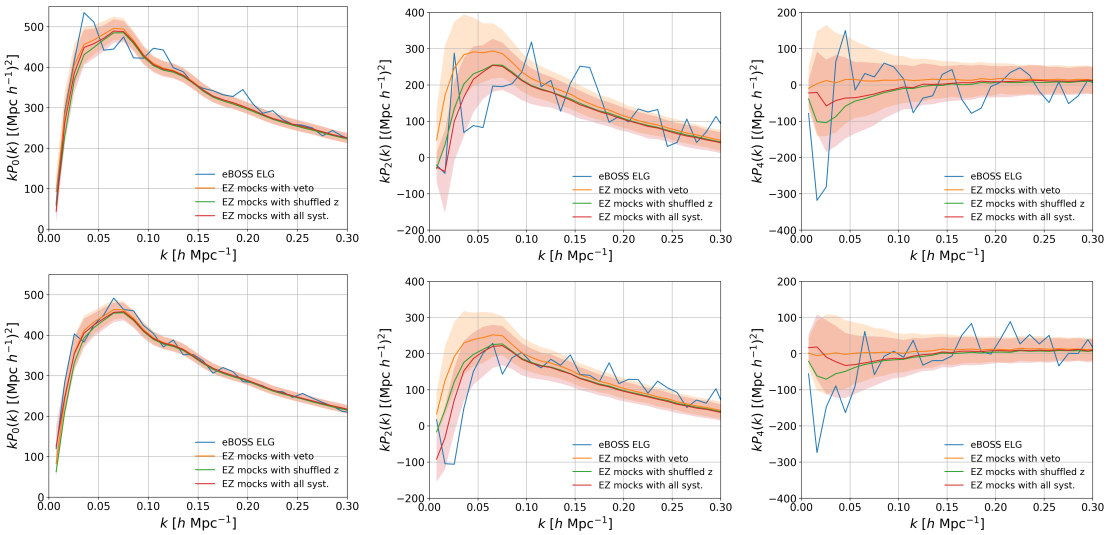

Figure 5.28 – Power spectrum measurements (left: monopole, middle: quadrupole, right: hexadecapole, top: NGC, bottom: SGC) of the eBOSS data (blue) and the mean and standard deviation (shaded region) of the EZ mocks without (orange) and with (red) all systematics. EZ mocks with the *shuffled* scheme only (green) do not include observational systematics.

from the mock data redshift distribution, which leads to the aforementioned radial integral constraint, clearly visible in the quadrupole and hexadecapole at large scale.

Figure 5.28 displays the power spectrum measurement of the eBOSS ELG sample (blue curve), together with the mean of the EZ mocks with veto masks only applied (*baseline*, orange). Accounting for the *shuffled* scheme in the mocks (green) resolves part of the difference between data and mocks in the quadrupole and hexadecapole on large scales. Including all systematics and corrections (red), the agreement with observed data is improved in the quadrupole.

5.3.4 Generation of GLAM-QPM mock catalogues

We also added observational effects to GLAM-QPM mocks. Contrary to EZ mocks, we do not implement variations of the redshift distribution with imaging depth, nor imaging systematics. However, all other systematics (fibre collisions and redshift failures) are treated the same way as for EZ mocks. Again, all systematic corrections are applied the exact same way to the mocks as to the eBOSS ELG catalogues.

The number statistics of GLAM-QPM mocks are provided in Table 5.6. Those are also in very good agreement with data (see Table 5.4).

Figure 5.29 displays the power spectrum measurement of the eBOSS ELG sample (blue curve). The mean of the GLAM-QPM mocks with veto masks only applied (*baseline*) is shown in orange. Including all implemented systematics and corrections (green), the agreement with observed data is improved.

We have been in charge of the implementation of observational systematics in mock catalogues, and their correction in both data and mocks, for the eBOSS collaboration. Main improvements compared to previous analyses consist in (i) implementing realistic observational systematics (e.g. fibre collisions) and (ii) applying the exact same corrections (including practical implementation) to both data and mocks in order to be best reproduce the statistical properties of the eBOSS ELG sample.

	eboss21	eboss22	eboss23	eboss25	NGC	SGC	ALL
$N_{\rm targ}$	36,308	79,907	70,941	42,556	113,497	116,215	229,711
$N_{\rm obs}$	34,329	76,202	66,511	40,532	107,043	110,531	217,574
$N_{\rm missed}$	339	651	602	340	943	990	1,932
$N_{\rm cp}$	1,640	3,055	3,827	1,684	5,511	4,695	10,206
$N_{\rm gal}$	31,340	69,098	58,877	36,189	95,066	100,437	195,504
$N_{\rm star}$	322	499	523	306	828	821	1,649
$N_{\rm fail}$	2,668	6,605	7,111	4,037	11,148	9,273	20,420
$N_{\rm used}$	28,078	61,814	51,453	32,273	83,726	89,892	173,618
Mean TSR	0.9455	0.9536	0.9376	0.9524	0.9431	0.9511	0.9472
Mean comp.	0.9907	0.9919	0.9915	0.9920	0.9917	0.9915	0.9916
Mean SSR	0.9223	0.9133	0.8931	0.9004	0.8959	0.9161	0.9061

Table 5.6 – Mean statistics of the 2003 GLAM-QPM mocks. See Table 5.4 for the definition of the various quantities and a comparison to the statistics of the observed data.

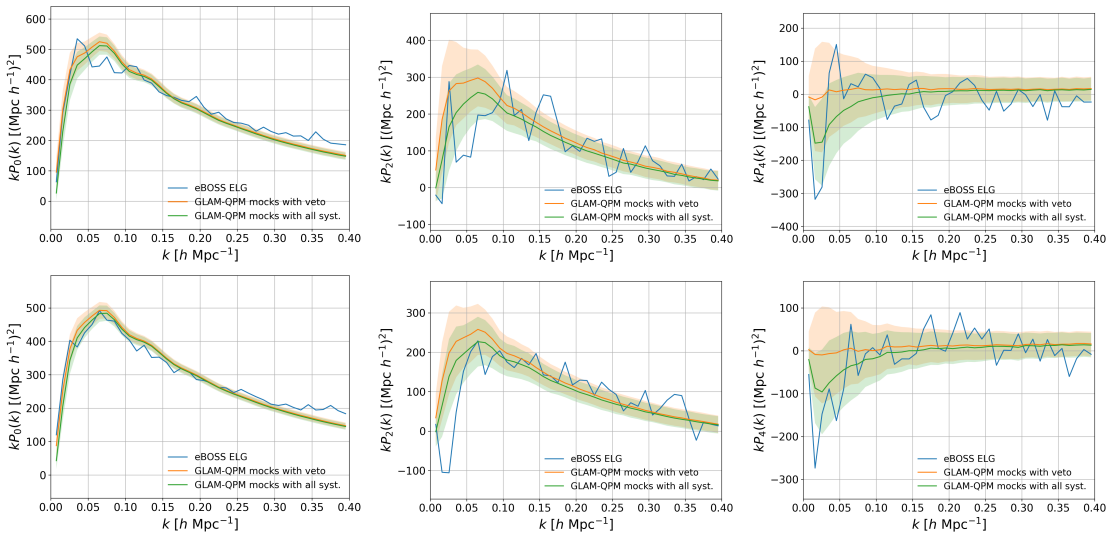

Figure 5.29 – Power spectrum measurements (left: monopole, middle: quadrupole, right: hexadecapole, top: NGC, bottom: SGC) of the eBOSS data (blue) and the mean of the GLAM-QPM mocks without (orange) and with (green) all systematics.

The mock catalogues we have just built, including observational systematics, will be used to test the analysis pipeline as described in the next section.

5.4 Testing the analysis pipeline using mock catalogues

In this section we first check our analysis pipeline and review how the observational systematics introduced in the approximate mocks impact BAO and RSD measurements. Although some systematic effects are difficult to model accurately in mocks, these can still be used to derive reliable estimates for part of the systematic uncertainties.The other systematic uncertainties will be estimated from the data itself in Section 5.5.

In all tests, to fit each type of mocks, we use the covariance matrix built from the same mocks, unless otherwise stated.

	$f\sigma_8$	$D_\mathrm{H}/r_\mathrm{drag}$	$D_\mathrm{M}/r_\mathrm{drag}$
EZ boxes at $z_\mathrm{snap} = 0.876$	$0.43088^{+0.00017}_{-0.00017}$	$18.2186^{+0.0034}_{-0.0034}$	$20.7175^{+0.0027}_{-0.0027}$
EZ boxes at $z_\mathrm{eff} = 0.845$	$0.43391^{+0.00017}_{-0.00017}$	$18.5590^{+0.0034}_{-0.0034}$	$20.1526^{+0.0026}_{-0.0026}$
Mean of EZ mocks no veto ($z_\mathrm{eff} = 0.845$)	$0.4346^{+0.0017}_{-0.0017}$	$18.477^{+0.033}_{-0.033}$	$20.103^{+0.028}_{-0.028}$
Mean of EZ mocks baseline ($z_\mathrm{eff} = 0.845$)	$0.4341^{+0.0017}_{-0.0017}$	$18.472^{+0.033}_{-0.034}$	$20.127^{+0.031}_{-0.030}$

Table 5.7 – Comparison of the RSD measurements on EZ boxes at redshift $z_\mathrm{snap} = 0.876$ and extrapolated at $z_\mathrm{eff} = 0.845$ (given their cosmology), with those from the sky-cut EZ mocks, with and without veto masks. For the EZ boxes we quote the mean and standard deviation of the best fit measurements, divided by the square root of the number of realisations (300). For the sky-cut mocks, error bars are given by the $\Delta\chi^2 = 1$ level on the mean of the mocks.

For reasons that will be justified in Section 5.5, we will use NGC and SGC (NGC + SGC) or SGC only power spectrum measurements and vary the redshift range. The baseline result will use NGC + SGC, and the redshift ranges $0.7 < z < 1.1$ and $0.6 < z < 1.1$ for the RSD and BAO fits, respectively.

5.4.1 Survey geometry effects

The RSD model presented in Section 5.1.2 neglects the evolution of the cosmological background within the redshift range of the eBOSS ELG sample. To test the impact of this assumption on clustering measurements, we first fit 300 EZ periodic boxes at redshift $z_\mathrm{snap} = 0.876$ (see Section 5.3.2.1), using a Gaussian covariance matrix, as in Section 5.2.2. We compare these measurements to those obtained on the mean of the *no veto* EZ mocks, that is including the (approximate) light-cone and global (tiling) footprint. In this case, we apply the corresponding window function treatment (Section 5.1.2.1) and the global integral constraint (Section 5.1.2.2) in the model. To ease the comparison, which we present in Table 5.7, we extrapolate the best fits to the EZ boxes at redshift $z_\mathrm{snap} = 0.876$ to the effective redshift $z_\mathrm{eff} = 0.845$ of the EZ mocks, using their input cosmology of Eq. (5.56). The difference between the extrapolated mean of the best fits to the EZ boxes and the best fit to the mean of the EZ mocks is 0.2% on $f\sigma_8$, 0.4% on $D_\mathrm{H}(z)/r_\mathrm{drag}$ and 0.2% on $D_\mathrm{M}(z)/r_\mathrm{drag}$, fully negligible compared to the dispersion of the mocks (12%, 6% and 5% respectively, see Table 5.9), validating our modelling approximation of the eBOSS ELG survey as a single snapshot at redshift $z_\mathrm{eff} = 0.845$. We finally apply veto masks to EZ mocks and in the window function calculation. In this case, again, the change in best fit parameters is $\simeq 0.1\%$, compatible with the error bars (*baseline* versus *no veto*). We also checked that increasing the sampling of the window function in the small separation $s \to 0$ limit has virtually no impact (0.01%) on the cosmological measurement. These tests validate our treatment of the window function with the fine-grained eBOSS ELG veto masks.

The total shifts between the *baseline* sky-cut mocks and the EZ boxes are 0.1%, 0.5% and 0.1% for $f\sigma_8$, $\alpha_\parallel$ and $\alpha_\perp$. We take them as systematic shifts (under the generic denomination *survey geometry*).

5.4.2 Fibre collisions

Fibre collisions are shown to be the dominant observational systematics in the eBOSS QSO sample (Neveux et al., 2020; Hou et al., 2020). The impact of fibre collisions can be seen on EZ mocks by comparing the green to the orange curves in Figure 5.27. Here we test their

	NGC	SGC
tile overlap	0.44	0.35
collision fraction	0.47	0.39
simulated collision fraction	0.46	0.38
EZ simulated collision fraction	0.46 ± 0.005	0.38 ± 0.004
GLAM-QPM simulated collision fraction	0.46 ± 0.005	0.39 ± 0.005

Table 5.8 – Different estimates of the fibre collisions fraction f_s. See text for details.

effect on cosmological fits to GLAM-QPM mocks, as these mocks are not further impacted by photometric systematics.

We report in Table 5.9 the best fits to 2003 GLAM-QPM *baseline* mocks with only geometry and veto masks applied (*baseline*) and to the mocks where fibre collisions are simulated (*fibre collisions*). We find a systematic shift of 2.5% on $f\sigma_8$ (22% of the dispersion of the mocks), 0.6% on $\alpha_\parallel$ (9%) and 0.5% on $\alpha_\perp$ (10%).

The impact of fibre collisions can be mitigated following Hahn et al. (2017), if the fraction of collided pairs f_s and the fibre collision angular scale D_{fc} are known. In the Hahn et al. (2017) correction, $f_s = 1$ corresponds to all galaxy pairs closer than the fibre collision angular scale being unobserved. Because of tile overlaps, this fraction is reduced. The fraction of collided pairs f_s can then be estimated in several ways:

— *tile overlap*: the fraction of the survey area without tile overlap, estimated using the synthetic catalogue. This assumes all collisions are resolved in tile overlaps;

— *collision fraction*: the number of targets which were collided with another one (including the relevant TDSS targets), divided by the number of targets that would be assigned a fibre without tile overlap. This number is simulated with the same algorithm as that used for the EZ and GLAM-QPM mocks to implement fibre collisions, except the effect of tile overlaps (see Section 5.3.3 and 5.3.4);

— *simulated collision fraction*: same as *collision fraction*, but also simulating the number of data targets which were collided with another one (including the relevant TDSS targets), taking into account tile overlaps;

— *EZ simulated collision fraction*: same as *simulated collision fraction*, in the EZ mocks;

— *GLAM-QPM simulated collision fraction*: same as *simulated collision fraction*, in the GLAM-QPM mocks.

All these estimates are calculated with veto masks applied and are reported in Table 5.8, using 50 mocks (for *EZ* and *GLAM-QPM simulated collision fraction*). They all agree within 2%. The modelling of fibre collisions in Hahn et al. (2017) is actually based on their impact on the projected correlation function. Figure 5.30 displays the ratio of the projected correlation function of the GLAM-QPM mocks with fibre collisions corrected by $w_{\rm cp,i}$ to the true one (without fibre collisions): f_s, given by the height of the step function (see Hahn et al. 2017), is in very good agreement with the above estimates provided in Table 5.8. We therefore choose the corresponding values $f_s = 0.46$ for NGC and $f_s = 0.38$ for SGC.

For the fibre collision angular scale D_{fc}, we take the comoving distance corresponding to the fibre collision radius $62''$ at the effective redshift of the eBOSS ELG sample $z_{\rm eff} = 0.845$. The

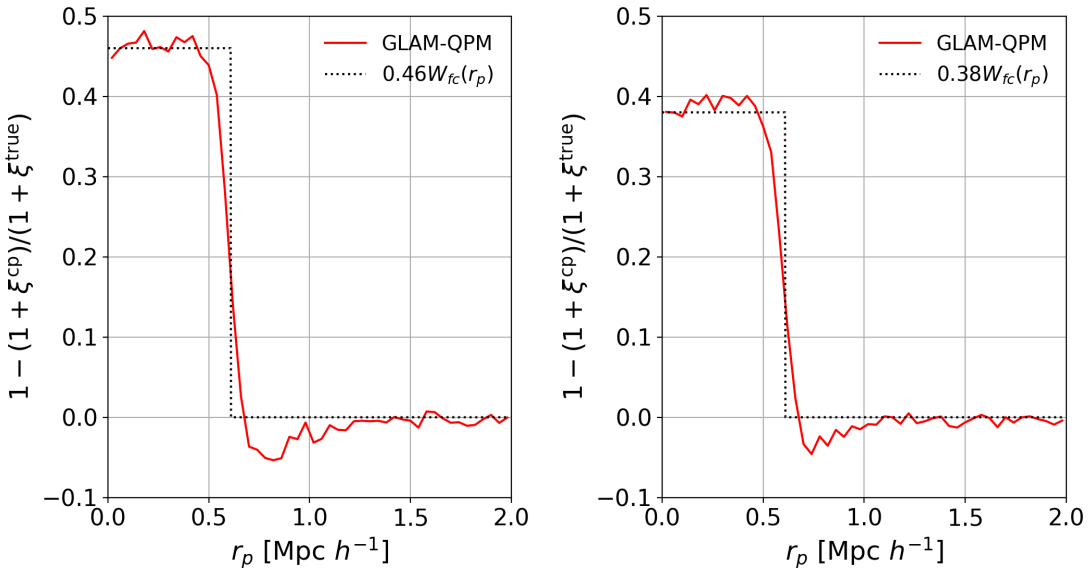

Figure 5.30 – Ratio of the $w_{cp,i}$-corrected projected correlation function to the true projected correlation function, presented in the form $1-(1+\xi^{cp})/(1+\xi^{true})$, as obtained in 379 GLAM-QPM mocks and in the model of Hahn et al. (2017) (left: NGC, right: SGC). See text for details, and Figure 8 of Hahn et al. (2017) for comparison.

obtained value, 0.61 Mpc h^{-1}, provides good modelling of the effect as can be seen in Figure 5.30. For the redshift cut $0.7 < z < 1.1$, a similar calculation provides $D_{fc} = 0.62\,\mathrm{Mpc}\,h^{-1}$.

The parameters f_s and D_{fc} being determined, the Hahn et al. (2017) correction can be included in the RSD model. Best fits to the GLAM-QPM mocks with fibre collisions are in very good agreement with the *baseline* mocks once the correction is included: the potential remaining systematic bias is 0.3% on $f\sigma_8$, 0.1% on $\alpha_\parallel$ and 0.0% on $\alpha_\parallel$ — 3%, 2% and 0% of the dispersion of the mocks, respectively (*fibre collisions + Hahn et al.* versus *baseline* in Table 5.9). We therefore include this correction as a baseline in the following.

Note that Bianchi and Percival (2017); Percival and Bianchi (2017) developed a method to correct for such missing observations in the n-point (configuration space) correlation function using n-tuple upweighting; for an application to the eBOSS samples (including ELG), we refer the reader to Mohammad et al. (2020). This method has been very recently extended to the Fourier space analysis by Bianchi and Verde (2019). We do not apply this technique to the eBOSS ELG sample, since most of this analysis was completed before this publication and because the effect of fibre collisions appears subdominant, especially after the Hahn et al. (2017) correction.

5.4.3 Radial integral constraint

As mentioned in Section 5.3.3, the *shuffled* scheme, used to assign data redshifts to randoms is responsible for a major shift of the power spectrum multipoles (purple versus red curves in Figure 5.27) due to the radial integral constraint (see Section 5.1.2.2). We report in Table 5.9 the cosmological measurements from RSD fits without (*baseline*, GIC) and with the *shuffled* scheme (*shuffled*, GIC), while keeping the global integral constraint (GIC) in the model: the induced systematic shift is 0.4% on $f\sigma_8$ (4% of the dispersion of the mocks), 4.4% on $\alpha_\parallel$ (66%) and 3.9% on $\alpha_\perp$ (78%). Modelling the radial integral constraint (RIC) removes most of this

bias: the remaining shift is 0.2% on $f\sigma_8$ (2% of the dispersion of the mocks), 0.2% on $\alpha_\parallel$ (3%) and 0.3% on $\alpha_\perp$ (5%).

5.4.4 Remaining angular systematics

In Section 5.3.3.3 we mentioned the large angular photometric systematics of the eBOSS ELG sample, which we attempted to introduce in the EZ mocks (orange versus blue curves in Figure 5.27). These systematics bias cosmological measurements from RSD fits, as can be seen in Table 5.9: comparing the fits on contaminated mocks, including the fibre collision correction of Section 5.4.2 (*all syst., fc*) to uncontaminated mocks (*baseline*, GIC), one notices a bias of 8.8% on $f\sigma_8$, 2.4% on $\alpha_\parallel$ and 1.6% on $\alpha_\perp$, corresponding to a significant shift of respectively 75%, 35% and 32% of the dispersion of the best fits to the mocks.

We propose to mitigate these residual systematics by applying the *pixelated* scheme as described in Section 4.3.7.2. In Table 5.9 we report the RSD measurements without (*baseline*, GIC) and with the full angular and radial integral constraints (ARIC) modelled, applying the *shuffled* and *pixelated* schemes to the uncontaminated mock data, for two pixel sizes: nside = 64 ($\simeq 0.84\,\mathrm{deg}^2$) and nside = 128 ($\simeq 0.21\,\mathrm{deg}^2$). The combined radial and angular integral constraint is correctly modelled, generating only a small potential bias of 1.1% on $f\sigma_8$, 0.5% and 0.4% on scaling parameters (which amounts to 10%, 8% and 7% of the dispersion of the mocks, respectively) for nside = 64. A similar shift is seen with nside = 128. The *pixelated* scheme increases statistical uncertainties by a reasonable fraction of $\simeq 10\%$.

Finally, Figure 5.31 shows the best fits to the *baseline* (blue) and contaminated (red) EZ mocks. Measurements obtained when applying the *pixelated* scheme (nside = 64) to the contaminated mocks and modelling the ARIC are shown in blue. The systematic bias quoted at the beginning of the section is clearly reduced and becomes 2.5% on $f\sigma_8$ (21% of the dispersion of the mocks), 0.4% on $\alpha_\parallel$ (6%) and 0.5% on $\alpha_\perp$ (12%) with nside = 64, slightly less with nside = 128 (see Table 5.9, *all syst & pix64, fc* with respect to *baseline, GIC*).

5.4.5 Likelihood Gaussianity

In Section 5.1.6 we assumed that we could use a Gaussian likelihood to compare data and model. While this may be accurate enough by virtue of the central limit theorem when the number of modes is high enough, it may break down on large scales where statistics is lower and mode coupling due to the survey geometry, and, in our specific case, the radial and angular integral constraints, occurs (see e.g. Hahn et al. 2019).

Comparing the median of the fits to each individual *baseline* EZ mocks to the fit to the mean of the mocks (see first two rows of second series of results in Table 5.9, *baseline, GIC* versus *mean of mocks baseline, GIC*), we observe shifts of 2.2% on $f\sigma_8$ (19% of the dispersion of the mocks), 1.3% on $\alpha_\parallel$ (21%) and 0.9% on $\alpha_\perp$ (18%). This bias could be due to either non-Gaussianity of the power spectrum likelihood or model non-linearity.

To test a potential bias coming from the breakdown of such a Gaussian assumption, we produce 1000 fake power spectra following a Gaussian distribution around the mean of the EZ mocks, with the covariance of the mocks, and fit them with our model (using the same covariance matrix). Results are reported in Table 5.9 (*fake all syst. & pix64, fc*). Shifts with respect to the true mocks (*all syst. & pix64, fc*) are 0.2% on $f\sigma_8$, 0.1% on $\alpha_\parallel$ and 0.2% on $\alpha_\perp$.

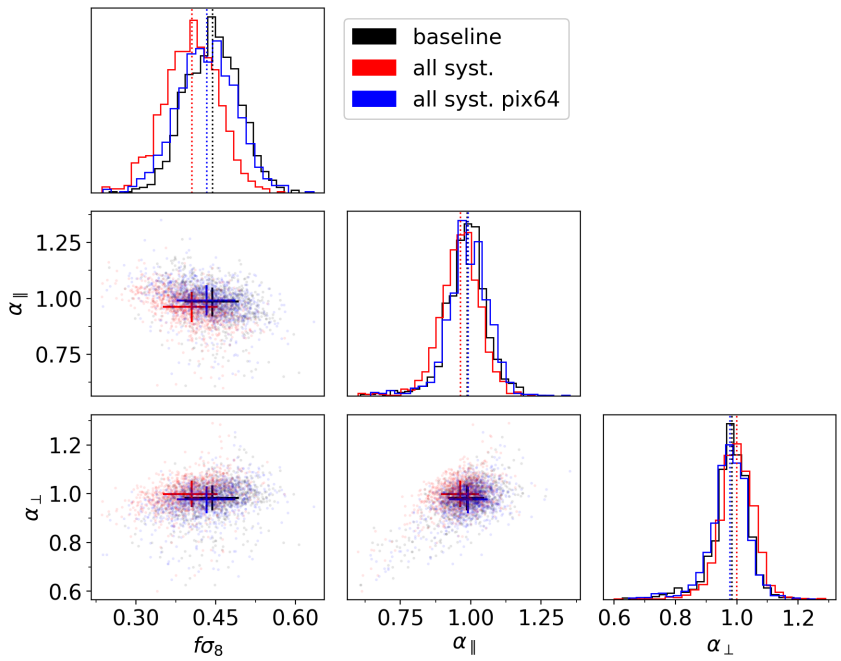

Figure 5.31 – Best fits to the baseline (black) and fully contaminated (red) EZ mocks. In blue, the *pixelated* scheme is applied on contaminated mocks to mitigate angular systematics. Dotted vertical lines in the histograms and crosses in the scatter plots point to the median of the best fit values, while the size of the crosses is given by the 16% and 84% percentiles.

We therefore conclude that one can safely use the Gaussian likelihood to compare data and model power spectra. We also attribute the shifts between the fit to the mean of the EZ mocks and the median of the fits to each mock to model non-linearity.

5.4.6 Isotropic BAO measurements

In Table 5.10 and hereafter, as in Ata et al. (2018); Raichoor et al. (2020), we qualify BAO detections as α measurements for which the best fit value and its error bar (determined by the $\Delta\chi^2 = 1$ level) are within the range $[0.8\alpha^{\rm exp}, 1.2\alpha^{\rm exp}]$ ($\alpha^{\rm exp}$ being the expected α value, given the fiducial and mock cosmologies). Statistics are provided for the $N_{\rm det}$ mocks with BAO detections. As we include covariance matrix corrections (Hartlap factor, see Eq. (5.13)) and correction to the parameter covariance matrix (m_1 factor, see Eq. (5.14)) in the α measurement on each mock, we follow Percival et al. (2014) and provide the standard deviation S of the α measurement corrected by $\sqrt{m_2}$, with m_2 given by Eq. (5.17).

As stated in Section 5.2.5, we fix $\Sigma_{\rm nl}$ to $8\,{\rm Mpc}\,h^{-1}$ (resp. $4\,{\rm Mpc}\,h^{-1}$) when fitting pre-reconstruction (resp. post-reconstruction) power spectra. Pre-reconstruction α measurements on both EZ and GLAM-QPM mocks are biased slightly high, as can be seen from Table 5.10 (*baseline pre-reconstruction* versus *baseline*). This is in line with the expected shift of the BAO peak caused by the non-linearity of structure formation (Padmanabhan et al., 2009; Ding et al., 2018). On the contrary, post-reconstruction α measurements do not show any bias, at the $0.43/\sqrt{1000} \simeq 0.1\%$ level.

The radial integral constraint effect was noticed to have a significant impact on RSD cosmological measurements (Section 5.4.3). We find its impact to be negligible on the post-reconstruction

	$f\sigma_8$	$\alpha_\parallel$	$\alpha_\perp$
RSD only	GLAM-QPM mocks		
baseline	$0.442^{+0.050}_{-0.050}$	$0.997^{+0.064}_{-0.066}$	$0.990^{+0.048}_{-0.054}$
fibre collisions	$0.453^{+0.050}_{-0.049}$	$0.990^{+0.063}_{-0.066}$	$0.995^{+0.048}_{-0.054}$
fibre collisions + Hahn et al.	$0.443^{+0.049}_{-0.050}$	$0.998^{+0.063}_{-0.066}$	$0.990^{+0.048}_{-0.055}$
RSD only	EZ mocks	tests of IC	
mean of mocks baseline, GIC	$0.4341^{+0.0017}_{-0.0017}$	$0.9997^{+0.0018}_{-0.0019}$	$0.9926^{+0.0015}_{-0.0015}$
baseline, GIC	$0.444^{+0.050}_{-0.052}$	$0.987^{+0.061}_{-0.066}$	$0.984^{+0.050}_{-0.053}$
shuffled, GIC	$0.446^{+0.049}_{-0.054}$	$0.943^{+0.060}_{-0.069}$	$1.022^{+0.053}_{-0.052}$
shuffled, RIC	$0.443^{+0.051}_{-0.054}$	$0.985^{+0.061}_{-0.064}$	$0.986^{+0.049}_{-0.052}$
shuffled & pix64, ARIC	$0.449^{+0.053}_{-0.057}$	$0.982^{+0.063}_{-0.064}$	$0.987^{+0.055}_{-0.053}$
shuffled & pix128, ARIC	$0.450^{+0.054}_{-0.061}$	$0.983^{+0.064}_{-0.066}$	$0.987^{+0.055}_{-0.053}$
RSD only	EZ mocks	mitigation	
all syst. fc	$0.405^{+0.049}_{-0.054}$	$0.964^{+0.065}_{-0.069}$	$0.999^{+0.055}_{-0.054}$
all syst. & pix64 fc	$0.433^{+0.054}_{-0.056}$	$0.990^{+0.069}_{-0.067}$	$0.978^{+0.052}_{-0.059}$
all syst. & pix128 fc	$0.438^{+0.056}_{-0.057}$	$0.987^{+0.070}_{-0.068}$	$0.979^{+0.054}_{-0.058}$
fake all syst. & pix64 fc	$0.434^{+0.054}_{-0.059}$	$0.991^{+0.060}_{-0.058}$	$0.979^{+0.053}_{-0.055}$
RSD + BAO	EZ mocks		
mean of mocks baseline, GIC	$0.4384^{+0.0016}_{-0.0017}$	$1.0031^{+0.0017}_{-0.0019}$	$0.9979^{+0.0012}_{-0.0016}$
baseline, GIC	$0.445^{+0.048}_{-0.048}$	$0.994^{+0.057}_{-0.051}$	$0.994^{+0.037}_{-0.040}$
shuffled & pix64, ARIC	$0.449^{+0.050}_{-0.051}$	$0.987^{+0.057}_{-0.055}$	$0.995^{+0.041}_{-0.044}$
all syst., fc	$0.404^{+0.049}_{-0.049}$	$0.973^{+0.055}_{-0.054}$	$1.010^{+0.043}_{-0.048}$
all syst. & pix64, fc	$0.434^{+0.053}_{-0.051}$	$1.000^{+0.058}_{-0.056}$	$0.989^{+0.046}_{-0.048}$
all syst. & pix64, fc, $B_{\rm nw}$ free	$0.434^{+0.053}_{-0.052}$	$1.002^{+0.057}_{-0.056}$	$0.989^{+0.045}_{-0.048}$
all syst. & pix128, fc	$0.439^{+0.054}_{-0.051}$	$0.998^{+0.059}_{-0.057}$	$0.990^{+0.046}_{-0.047}$
RSD + BAO	EZ mocks	$0.7 < z < 1.1$	
all syst. & pix64, fc	$0.436^{+0.054}_{-0.058}$	$0.998^{+0.061}_{-0.055}$	$0.991^{+0.045}_{-0.051}$
photo syst. & pix64	$0.451^{+0.055}_{-0.054}$	$0.987^{+0.061}_{-0.057}$	$0.996^{+0.043}_{-0.046}$
photo + cp syst. & pix64, no fc	$0.457^{+0.051}_{-0.055}$	$0.986^{+0.056}_{-0.064}$	$1.001^{+0.042}_{-0.047}$
photo + cp syst. & pix64, fc	$0.446^{+0.051}_{-0.055}$	$0.993^{+0.059}_{-0.063}$	$0.996^{+0.042}_{-0.046}$
all syst. & pix64, fc, no $w_{\rm noz,i}$	$0.435^{+0.056}_{-0.057}$	$1.002^{+0.063}_{-0.056}$	$0.992^{+0.043}_{-0.050}$
all syst. randnoz & pix64, fc	$0.446^{+0.054}_{-0.051}$	$0.989^{+0.058}_{-0.061}$	$0.994^{+0.046}_{-0.051}$
all syst. randnoz & pix64, fc, no $w_{\rm noz,i}$	$0.445^{+0.056}_{-0.056}$	$0.992^{+0.059}_{-0.060}$	$0.995^{+0.045}_{-0.050}$
all syst. & pix64, fc, GLAM-QPM cov	$0.435^{+0.058}_{-0.061}$	$1.000^{+0.065}_{-0.061}$	$0.988^{+0.045}_{-0.053}$
all syst. & pix64, fc, no syst. cov	$0.435^{+0.056}_{-0.059}$	$1.002^{+0.061}_{-0.058}$	$0.992^{+0.044}_{-0.050}$
all syst. & pix64, $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$	$0.435^{+0.056}_{-0.060}$	$0.997^{+0.062}_{-0.058}$	$0.987^{+0.046}_{-0.049}$
all syst. & pix64, fc +1/2 k-bin	$0.439^{+0.055}_{-0.058}$	$0.996^{+0.064}_{-0.053}$	$0.993^{+0.046}_{-0.052}$
fake all syst. & pix64, fc	$0.437^{+0.058}_{-0.066}$	$0.999^{+0.061}_{-0.062}$	$0.992^{+0.048}_{-0.059}$

Table 5.9 – Impact of systematics on RSD and RSD + BAO measurements on the GLAM-QPM and EZ mocks. We quote the median and the 16% and 84% percentiles as a metric of the centre and dispersion of the measurements.

	$\langle\alpha\rangle$	$\langle\sigma\rangle$	S (uncorrected)	$N_{\rm det}/N_{\rm tot}$	$\langle\chi^2\rangle/dof$
EZ mocks	1.0003				
baseline pre-reconstruction	1.004	0.047	0.051 (0.049)	942/1000	$40.7/(54-13) = 0.992$
mean of mocks baseline	1.0017	0.0011	–	–	$671/(54-13) = 16.4$
baseline	1.000	0.033	0.043 (0.042)	981/1000	$41.4/(54-13) = 1.01$
shuffled	1.000	0.033	0.043 (0.042)	979/1000	$41.4/(54-13) = 1.01$
all syst.	1.002	0.033	0.042 (0.042)	979/1000	$41.5/(54-13) = 1.01$
photo syst.	1.000	0.034	0.043 (0.042)	978/1000	$41.4/(54-13) = 1.01$
photo + cp syst.	1.001	0.034	0.043 (0.042)	983/1000	$41.6/(54-13) = 1.01$
all syst., no $w_{\rm noz,i}$	1.001	0.034	0.043 (0.042)	985/1000	$42.3/(54-13) = 1.03$
all syst. rand noz	0.999	0.034	0.043 (0.042)	982/1000	$41.5/(54-13) = 1.01$
all syst. rand noz, no $w_{\rm noz,i}$	1.000	0.034	0.044 (0.043)	986/1000	$42.1/(54-13) = 1.03$
all syst., GLAM-QPM cov	1.000	0.036	0.044 (0.043)	981/1000	$39.8/(54-13) = 0.970$
all syst., no syst. cov	1.000	0.034	0.043 (0.042)	986/1000	$42.0/(54-13) = 1.03$
all syst. + 1/2 k-bin	1.002	0.034	0.044 (0.043)	977/1000	$41.4/(54-13) = 1.01$
all syst. $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$	1.002	0.038	0.042 (0.041)	974/1000	$41.5/(54-13) = 1.01$
fake all syst.	1.001	0.034	0.044 (0.043)	982/1000	$41.8/(54-13) = 1.02$
all syst. SGC	1.003	0.045	0.057 (0.056)	926/1000	$19.9/(27-7) = 0.995$
GLAM-QPM mocks	0.9992				
baseline pre-reconstruction	1.002	0.046	0.047 (0.047)	1907/2003	$41.2/(54-13) = 1.01$
baseline	0.998	0.031	0.040 (0.040)	1969/2003	$42.5/(54-13) = 1.04$
all syst.	0.997	0.032	0.043 (0.042)	1973/2003	$42.4/(54-13) = 1.03$

Table 5.10 – Isotropic BAO measurements on EZ and GLAM-QPM OuterRim mocks in different conditions. We quote statistics for the $N_{\rm det}$ mocks with BAO *detection*, i.e. mocks for which $\alpha - \sigma_{\rm low} > 0.8\alpha^{\rm exp}$ and $\alpha + \sigma_{\rm up} < 1.2\alpha^{\rm exp}$. $\langle\alpha\rangle$ is the mean α, $\langle\sigma\rangle$ the mean σ $(= (\sigma_{\rm low} + \sigma_{\rm up})/2)$. S is the standard deviation of α, rescaled by $\sqrt{m_2}$, with m_2 given by Eq. (5.17) (the uncorrected value is provided in brackets). Expected values $\alpha^{\rm exp}$ are given at the top of each sub-table (in the $\langle\alpha\rangle$ column).

isotropic BAO measurements (*shuffled* versus *baseline*). We thus do not model any RIC correction for the isotropic BAO fits, as it would have required an increased computation time.

Adding all observational systematics and their correction scheme (*all syst.*), the isotropic BAO fits to EZ mocks shift by a negligible 0.1%, while no change is seen with GLAM-QPM mocks (which do not include angular photometric systematics).

To support data robustness tests presented in Section 5.5, we apply systematics successively to the EZ mocks. Fibre collisions lead to a negligible α shift of 0.1% (*photo + cp syst.* versus *photo syst.*).

Redshift failures do not impact the α measurement (*all syst.* versus *photo + cp syst.*). Ignoring the correction weight $w_{\rm noz,i}$ and removing redshift failures from the mocks used to build the covariance matrix is equally harmless (*all syst. no* $w_{\rm noz,i}$ versus *all syst.*). A negligible shift is seen as well when the correction weight $w_{\rm noz,i}$ is not used, and redshift failures are removed from the mocks used to build the covariance matrix (*all syst. rand noz & pix64, fc, no* $w_{\rm noz,i}$). In the above, redshift failures are implemented in the EZ mocks following a deterministic process: a mock object is declared a redshift failure if the redshift of its nearest neighbour in the data could not be reliably measured, which in average overestimates the angular impact of redshift failures. As explained in Section 5.3.3.5, we therefore consider a second set of mocks,

where redshift failures are applied to the EZ mocks with a probability following the model fitted on the data catalogue. In this case (*rand noz*), shifts in the measured α due to redshift failures are equally small.

The recovered α does not change when k-bin centres are shifted by half a bin ($0.005\,h\,\mathrm{Mpc}^{-1}$, *all syst. + 1/2 k-bin*). Fitting the EZ mocks with the covariance matrix estimated from GLAM-QPM mocks results in a small 0.1% shift of α measurements. The same behaviour is seen when using the covariance matrix from EZ mocks without systematics (with the *shuffled* scheme only, *no syst. cov*).

As in Section 5.4.5 we again generate and fit (*fake all syst.*) 1000 fake power spectra following a Gaussian distribution with mean and covariance matrix inferred from the contaminated EZ mocks. A negligible shift of 0.1% of α is seen with respect to the true mocks (*all syst.*), showing that one can safely use a Gaussian likelihood to compare data and model power spectra.

A small shift of 0.1% is seen between the fit to the mean of the mocks and the mean of the fits to each individual mock, which we label as model non-linearity in the following.

Based on the previous tests, we determine two systematic effects to be included in the final systematic budget: the model non-linearity, since it can only be measured on mocks, and fibre collisions, as we believe our modelling of the effect in the EZ mocks (see Section 5.3.3) to be quite representative of the actual data fibre collisions. We directly take the shift in α attributed to model non-linearity as a systematic bias. The 0.1% α shift attributed to fibre collisions is below twice the mock-to-mock dispersion divided by the square root of the number of mocks, i.e. $2 \times 3.8\%/\sqrt{962} = 0.2\%$, a value which we take as a systematic uncertainty, following the same procedure as in Neveux et al. (2020); Gil-Marín et al. (2020).

One would notice that the mean error on α measurements on EZ and GLAM-QPM mocks (defined by the $\Delta\chi^2 = 1$ level, see Section 5.1.6) is systematically and significantly lower than the dispersion of the best fit values. The value of the damping parameter $\Sigma_{\rm nl} = 4\,\mathrm{Mpc}\,h^{-1}$ is chosen to match the BAO amplitude seen in the reconstructed OuterRim mocks (see Section 5.2.5). However, the BAO amplitude is significantly less pronounced in the EZ mocks (see e.g. Raichoor et al. 2020), hence favouring a larger $\Sigma_{\rm nl} = 6\,\mathrm{Mpc}\,h^{-1}$. When this value is used, the distribution of the residuals $(\alpha - \langle\alpha\rangle)/\sigma$ of mocks with BAO detection is consistent with a standard normal distribution, as shown by the Kolmogorov-Smirnov test [10] of Figure 5.32. Using a lower $\Sigma_{\rm nl}$ artificially decreases the error on the BAO fits to EZ or GLAM-QPM mocks. Since we determined $\Sigma_{\rm nl}$ on the more accurate OuterRim-based mocks, we conclude that statistical errors quoted on the data measurement are fairly estimated.

5.4.7 Combined RSD and BAO measurements

As already mentioned in Section 5.1.5, we combine the RSD and BAO likelihoods, taking into account the cross-covariance between pre- and post-reconstruction power spectrum measurements.

In Table 5.9, one can notice a small shift between the RSD and the RSD + BAO fit to the mean of the *baseline* EZ mocks: 1.0% on $f\sigma_8$, 0.3% on $\alpha_\parallel$ and 0.5% on $\alpha_\perp$, which we quote as systematic error related to the technique of combining RSD and BAO likelihoods. These shifts may come from residual systematic differences between BAO template and mocks which contaminate the RSD part of the likelihood through its cross-covariance with the BAO part.

10. Non-parametric statistical test to determine the consistency between a sample and a probability law or another sample, based on the supremum of the difference of their cumulative distribution function.

We do not investigate this effect further since these biases remain small ($< 10\%$) compared to the dispersion of the mocks (and thus to the data measurement errors).

Again, RSD + BAO measurements on contaminated (*all syst, fc*) mocks are strongly biased: 9.2% on $f\sigma_8$ (86% of the dispersion of the mocks), 2.1% on $\alpha_\parallel$ (41%) and 1.6% on $\alpha_\perp$ (42%). When applying the *pixelated* scheme (nside = 64), one recovers reasonable systematic shifts with respect to (*baseline, GIC*) of 2.5% on $f\sigma_8$, 0.6% on $\alpha_\parallel$ and 0.5% on $\alpha_\perp$. These shifts reduce further when using nside = 128, but we choose the *pixelated* scheme with nside = 64 as it induces a bias which we estimate small enough for our analysis since it represents 24% of the dispersion of the mocks on $f\sigma_8$, 11% on $\alpha_\parallel$ and 13% on $\alpha_\perp$. In addition, the *pixelated* scheme (which involves integrating over all scales of the model correlation function) has only been tested up to nside = 64 with N-body based mocks in Section 4.3.6.3. Moreover, the data clustering measurement is also plagued by the complex dependence of $n(z)$ with imaging quality, which we only partly removed through the `chunk_z` splitting of the radial selection function in Section 5.3.1.6. This will require estimating the potential residual systematics from the data itself (see Section 5.5.2), which will prove to be large so that the previously mentioned shifts become subdominant.

We note the potential systematic bias induced by applying the radial and angular integral constraints (*shuffled & pix64, ARIC* versus *baseline, GIC*): 0.8% on $f\sigma_8$, 0.6% on $\alpha_\parallel$ and 0.1% on $\alpha_\perp$. These shifts are more than twice the mock-to-mock dispersion, divided by the square root of the number of mocks (0.4% on $f\sigma_8$, 0.1% on $\alpha_\parallel$ and $\alpha_\perp$). We therefore account for the ARIC modelling in our systematic budget by taking an error of 0.8% on $f\sigma_8$, 0.6% on $\alpha_\parallel$ and 0.1% on $\alpha_\perp$.

The isotropic BAO template of Eq. (5.7) contains a bias term $B_{\rm nw}$, which we so far forced to be equal to the linear bias b_1 of the RSD model. We try to let it free ($B_{\rm nw}$ *free*) and see no shift on cosmological parameters. We thus keep $B_{\rm nw} = b_1$ in the following.

As in Section 5.4.5 we generate and fit (*fake all syst.*) 1000 fake power spectra following a Gaussian distribution with mean and covariance matrix inferred from the contaminated EZ mocks. Negligible shifts of 0.3% on $f\sigma_8$, 0.1% on $\alpha_\parallel$ and 0.1% on $\alpha_\perp$ are seen with respect to the true mocks (*all syst. & pix64, fc*), showing that using a Gaussian likelihood to compare data and model power spectra is accurate enough. However, we find small systematic shifts of 1.6% on $f\sigma_8$ (15% of the dispersion of the mocks), 1.0% on $\alpha_\parallel$ (17%) and 0.4% on $\alpha_\perp$ (11%) between the median of the best fits to each individual mock and the fit to the mean of the mocks (*baseline, GIC* versus *mean of mocks baseline, GIC*). As in Section 5.4.5, we attribute this bias to the model non-linearity, which one would note is slightly reduced compared to the RSD only analysis.

We check that the error bars measured on each individual mock (defined by the $\Delta\chi^2 = 1$ level, see Section 5.1.6) are correctly estimated by performing a similar test as done on the post-reconstruction isotropic BAO fits in Section 5.4.6. For this Kolmogorov-Smirnov test, we use $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$, and keep only mocks for which the best fit $\alpha_\parallel$ and $\alpha_\perp$ and their error bars (divided by the $\alpha_\parallel$ and $\alpha_\perp$ expected values) are within the range $[0.8, 1.2]$. The residuals are in correct agreement with a standard normal distribution, as expected. We checked that the $f\sigma_8$ residuals remain very compatible with a standard normal distribution when considering all mocks (i.e. without cut on $\alpha_\parallel$ and $\alpha_\perp$ best fits and error bars).

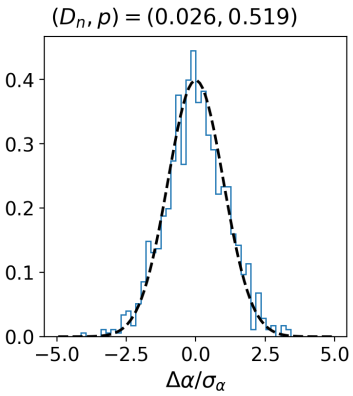

Figure 5.32 – Kolmogorov-Smirnov test on the residuals of EZ mocks, with $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$, for the BAO fits.

5.4.8 Further tests

In Section 5.5, we will justify our choice to fit the data with the redshift cut $0.7 < z < 1.1$. The expected shift on $f\sigma_8$ due to the change in effective redshift is 0.3%. As can be seen in Table 5.9 (*all syst. pix64, fc*), the effect of this redshift cut on the EZ mocks is negligible.

As in Section 5.4.6, to support the data robustness tests presented in Section 5.5, we apply systematics successively to the EZ mocks. Fibre collisions shift $f\sigma_8$, $\alpha_\parallel$ and $\alpha_\perp$ by 1.1%, 0.6% and 0.0%, respectively (*photo + cp syst. & pix64, fc* versus *photo syst. & pix64*). The $\alpha_\perp$ shift lies below twice the mock-to-mock dispersion divided by the square root of the number of mocks, 0.2%, which we therefore take as a systematic uncertainty for this parameter. This is not the case for $f\sigma_8$ (0.5%) and $\alpha_\parallel$ (0.2%), for which we take the measured shifts 1.1%, 0.6% as systematic uncertainty, following the same procedure as in e.g. Neveux et al. (2020). In Section 5.4.2 our several estimates of the f_s parameter required for the Hahn et al. (2017) fibre collision correction differed by 2% at most. To assess the impact of this additional uncertainty, we compare best fits with and without the Hahn et al. (2017) fibre collision correction: we find shifts on $f\sigma_8$, $\alpha_\parallel$ and $\alpha_\perp$ of 1.3%, 0.1% and 0.5%, respectively. Multiplying these variations by the uncertainty of 2% leads to a very small additional uncertainty, which we thus neglect.

Despite the weights $w_{\rm noz,i}$ and the *pixelated* scheme, redshift failures (red versus green curves in Figure 5.27) produce shifts of 2.3%, 0.5% and 0.5% on $f\sigma_8$, $\alpha_\parallel$ and $\alpha_\perp$ (*all syst. & pix64, fc* versus *photo + cp syst. & pix64, fc*). These shifts become 2.6%, 0.9% and 0.4% on $f\sigma_8$, $\alpha_\parallel$ and $\alpha_\perp$ when the correction weight $w_{\rm noz,i}$ is not used, and redshift failures are removed from the mocks used to build the covariance matrix (*all syst. & pix64, fc, no* $w_{\rm noz,i}$). A much smaller systematic shift is seen with respect to angular photometric systematics and fibre collisions only (*photo + cp syst. & pix64, fc*) for EZ mocks with the stochastic implementation of redshift failures (*all syst. rand noz & pix64, fc*): 0.1% on $f\sigma_8$, 0.4% on $\alpha_\parallel$ and 0.1% on $\alpha_\perp$.

We finally test the robustness of our analysis when using a covariance matrix measured from the GLAM-QPM mocks (without angular photometric systematics), from EZ mocks without systematics (*no syst. cov*), and when shifting the k-bin centres by half a bin ($0.005\,h\,{\rm Mpc}^{-1}$, *all syst. + 1/2 k-bin*). In all these cases, best fits to the EZ mocks remain stable.

We therefore conclude that our analysis pipeline is robust enough to perform the BAO and RSD + BAO measurements on the eBOSS ELG data. Based on the previous tests, four

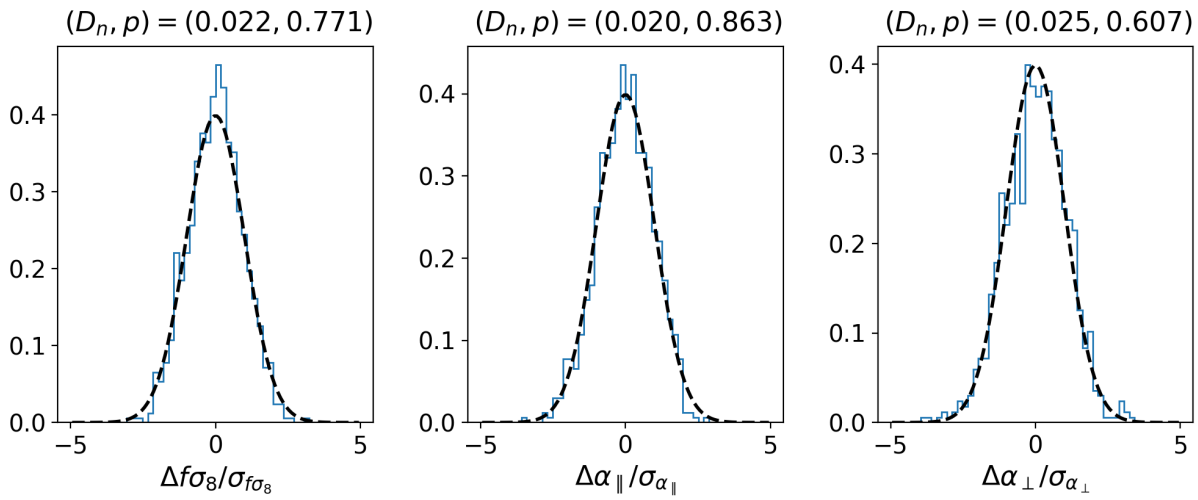

Figure 5.33 – Kolmogorov-Smirnov test on the residuals of EZ mocks, with $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$, for the RSD + BAO fits.

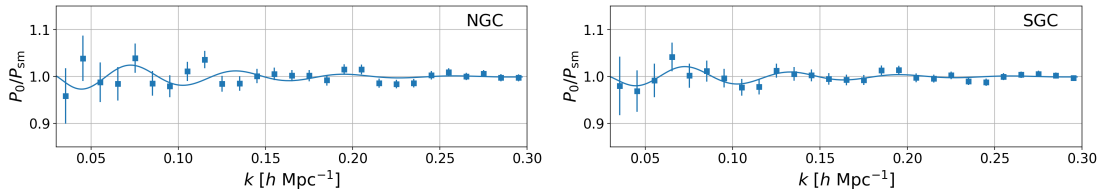

Figure 5.34 – Isotropic BAO fit (left: NGC, right: SGC), in the baseline case: NGC + SGC, $0.6 < z < 1.1$, $\Sigma_{\rm nl} = 4\,{\rm Mpc}\,h^{-1}$. Both data (points with error bars from EZ mocks) and model (continuous line) are divided by the *no-wiggle* power spectrum.

systematic effects estimated on mocks (survey geometry, model non-linearity, ARIC modelling, fibre collisions) will be included in the final systematic budgets presented in the next section.

5.5 Results

In this section we present isotropic BAO, RSD, and combined RSD + BAO measurements on the eBOSS DR16 ELG data, discuss robustness tests of those results and provide the final error budgets, including statistical and systematic contributions. In particular, systematic uncertainties are estimated from data itself where we consider mocks cannot give a reliable estimate.

5.5.1 Isotropic BAO measurements

As decided in Section 5.2.5, we take $\Sigma_{\rm nl} = 4\,{\rm Mpc}\,h^{-1}$ as fiducial value for the BAO damping and BAO templates are computed within the fiducial cosmology (5.4), except otherwise stated. Figure 5.34 shows the BAO oscillation pattern fitted to the observed ELG NGC + SGC data. One would note that NGC does not show a clear BAO feature, contrary to SGC. The best fit α and its 1σ error are provided for both SGC and NGC + SGC fits in Table 5.11. For both

	α	χ^2/dof
SGC only	z cuts	
$0.6 < z < 1.1$	$0.997^{+0.032}_{-0.035}$	$12.4/(27-7) = 0.619$
$0.65 < z < 1.1$	$0.989^{+0.034}_{-0.034}$	$13.1/(27-7) = 0.655$
$0.7 < z < 1.1$	$0.995^{+0.039}_{-0.038}$	$13.1/(27-7) = 0.654$
$0.75 < z < 1.1$	$0.993^{+0.040}_{-0.038}$	$20.1/(27-7) = 1.01$
SGC only	$0.6 < z < 1.1$	
no `chunk_z`	$0.991^{+0.035}_{-0.036}$	$13.8/(27-7) = 0.692$
no `chunk_z` G-Q cov	$0.988^{+0.034}_{-0.036}$	$15.2/(27-7) = 0.762$
$\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$	$0.993^{+0.040}_{-0.040}$	$13.1/(27-7) = 0.653$
no $w_{\rm noz,i}$	$0.997^{+0.032}_{-0.033}$	$11.7/(27-7) = 0.584$
GLAM-QPM cov	$0.997^{+0.034}_{-0.036}$	$12.3/(27-7) = 0.613$
no syst. cov	$0.997^{+0.033}_{-0.033}$	$13.1/(27-7) = 0.654$
500 mocks in cov	$1.000^{+0.032}_{-0.033}$	$12.6/(27-7) = 0.632$
+ 1/2 k-bin	$1.003^{+0.030}_{-0.031}$	$16.2/(27-7) = 0.810$
OR cosmo (rescaled)	$0.999^{+0.025}_{-0.026}$	$13.3/(27-7) = 0.664$
NGC + SGC	z cuts	
$0.6 < z < 1.1$ (baseline)	$0.986^{+0.025}_{-0.028}$	$42.8/(54-13) = 1.04$
$0.65 < z < 1.1$	$0.984^{+0.026}_{-0.027}$	$42.5/(54-13) = 1.04$
$0.7 < z < 1.1$	$0.982^{+0.028}_{-0.032}$	$44.8/(54-13) = 1.09$
$0.75 < z < 1.1$	$0.961^{+0.035}_{-0.041}$	$49.0/(54-13) = 1.19$
NGC + SGC	$0.6 < z < 1.1$	
no `chunk_z`	$0.973^{+0.031}_{-0.036}$	$42.1/(54-13) = 1.03$
no `chunk_z` G-Q cov	$0.970^{+0.029}_{-0.031}$	$40.2/(54-13) = 0.980$
$\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$	$0.979^{+0.033}_{-0.038}$	$43.4/(54-13) = 1.06$
no $w_{\rm noz,i}$	$0.984^{+0.025}_{-0.026}$	$46.1/(54-13) = 1.12$
GLAM-QPM cov	$0.988^{+0.026}_{-0.027}$	$38.8/(54-13) = 0.946$
no syst. cov	$0.984^{+0.026}_{-0.026}$	$40.3/(54-13) = 0.984$
500 mocks in cov	$0.991^{+0.025}_{-0.027}$	$38.8/(54-13) = 0.946$
+ 1/2 k-bin	$0.991^{+0.025}_{-0.027}$	$52.1/(54-13) = 1.27$
OR cosmo (rescaled)	$0.992^{+0.022}_{-0.023}$	$40.2/(54-13) = 0.980$
OR cosmo (rescaled), $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$	$0.988^{+0.026}_{-0.029}$	$40.4/(54-13) = 0.986$

Table 5.11 – Isotropic BAO best fits on the eBOSS DR16 ELG sample. Error bars are defined by the $\Delta\chi^2 = 1$ level.

fits, $\alpha \pm \sigma$ lies well in $[0.8, 1.2]$ [11], the criterion used in Section 5.4.6 to qualify *detections* in the mocks. However, in the NGC alone, the best fit α value is 0.79, such that the aforementioned criterion is not met. The same test, using the same BAO template in fiducial cosmology (5.4), is applied to EZ mocks (with $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$, to match their BAO signal amplitude) and to sky-cut OuterRim mocks of Section 5.2.5 (with $\Sigma_{\rm nl} = 2.4\,{\rm Mpc}\,h^{-1}$, to match their BAO signal amplitude), as reported in Table 5.12 ($\alpha \pm \sigma \notin [0.8, 1.2]$ line). Sky-cut OuterRim mocks (hereafter OR mocks), based on N-body simulations, provide the expected BAO detections in the absence of non-Gaussian contributions due to systematics [12]. In contrast, EZ mocks include known data systematics, but their BAO amplitude is lower than expected given their cosmology. Altogether, we expect the correct BAO *detection* rate to lie between values derived from EZ mocks and OR mocks. One notices that 9% of the EZ mocks and 2.5% of the OR mocks fail to meet the $\alpha \pm \sigma \in [0.8, 1.2]$ criterion in both the NGC and SGC. Therefore, the probability that α does not lie in $[0.8, 1.2]$ within errors, for either the NGC or the SGC, ranges from 5% (OR mocks) to 17% (EZ mocks), such that, with this criterion, the behaviour of the data is not very unexpected. This is in line with conclusions drawn in the configuration space BAO analysis (Raichoor et al., 2020).

To further quantify the BAO signal we compute the χ^2 difference between the best fits obtained with the *wiggle* and *no-wiggle* power spectrum templates (see Section 5.1.3). The χ^2 profiles using the *wiggle* and *no-wiggle* power spectrum templates are shown in Figure 5.35. Combining NGC and SGC we find $\Delta\chi^2 = -1.95$ (1.4σ) at a best fit value denoted $\alpha_{\rm NSGC}$ in the following. Note however that the best fit α value may not be relevant to compute the $\Delta\chi^2$ criterion when too far from the true one if the data (or mock) vector is too noisy. Therefore, for data or mock fits performed on each cap (NGC and SGC) separately, we quote in Table 5.12 the $\Delta\chi^2$ value evaluated at the corresponding NGC + SGC best fit value rather than at the respective NGC or SGC best fits, which are more subject to noise. We also provide the $\Delta\chi^2$ taken at the expected α value, given our fiducial cosmology, $\alpha^{\rm exp}$ ($\alpha^{\rm exp} = 1$ for data). We find that for NGC + SGC, the mean $\Delta\chi^2(\alpha = \alpha_{\rm NSGC})$ is lower in the mocks, meaning a better BAO *detection*. However, 18% EZ mocks and 7% OR mocks have larger $\Delta\chi^2$ values, i.e. worse BAO *detection*, than the data (see $N\,(> \Delta\chi^2(\alpha = \alpha_{\rm NSGC}))$ line in Table 5.12). So according to this criterion, the behaviour of the data is not very unexpected. A similar conclusion holds when taking $\Delta\chi^2$ at $\alpha = \alpha^{\rm exp}$ (see $N\,(> \Delta\chi^2(\alpha = \alpha^{\rm exp}))$ in Table 5.12). Focusing on the SGC, $\Delta\chi^2(\alpha = \alpha_{\rm NSGC})$ is smaller (better BAO *detection*) in data than in 85% of the EZ mocks and 60% of the OR mocks. However, only 1.8% of EZ mocks and 0.6% of OR mocks show a larger $\Delta\chi^2(\alpha = \alpha_{\rm NSGC})$ (worse BAO *detection*) than the NGC data. Therefore, the probability for such a poor BAO *detection* to happen in either NGC or SGC is approximately twice higher, of the order of a few percents. Again similar conclusions hold when taking $\Delta\chi^2$ at $\alpha = \alpha^{\rm exp}$. We emphasise however that the above figures are tied to the statistics used to qualify the BAO *detection*.

We have seen that contrary to SGC the poor BAO detection in the NGC is statistically unlikely (even given the known observational systematics implemented in the EZ mocks). Let us now discuss whether one can combine the two caps. We note that NGC photometry is shallower than SGC and thus more prone to (potentially unknown) photometric systematics (Raichoor

11. Here we assume that fiducial cosmology (5.4) agrees with the true one such that the BAO peak positions differ by much less than 20%.

12. We note however that in the OuterRim cosmology (2.89) the BAO amplitude (and hence signal-to-noise of BAO fits) is slightly larger than the Planck Collaboration et al. (2018) best fit model.

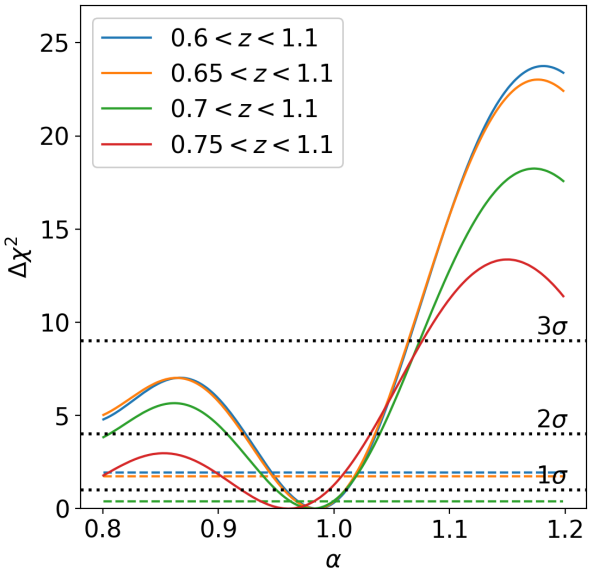

Figure 5.35 – χ^2 profiles of the isotropic BAO fits for different lower redshift cuts, fitting NGC + SGC, relative to the minimum value obtained with the *wiggle* power spectrum template. Continuous (resp. dashed) lines show the χ^2 profile using the *wiggle* (resp. *no-wiggle*) power spectrum template. Systematic uncertainties of Table 5.13 are not included.

	NGC	SGC	NGC + SGC
data			
$\Delta\chi^2(\alpha = \alpha_{\rm NSGC})$	3.58	−5.53	−1.95
$\Delta\chi^2(\alpha = \alpha^{\rm exp})$	3.97	−5.61	−1.63
all syst. EZ mocks			
$N\,(\alpha \pm \sigma \notin [0.8, 1.2])$	91	90	26
$\Delta\chi^2(\alpha = \alpha_{\rm NSGC})$	−3.22	−2.43	−5.65
$N\,(> \Delta\chi^2(\alpha = \alpha_{\rm NSGC}))$	18	846	184
$\Delta\chi^2(\alpha = \alpha^{\rm exp})$	−2.62	−1.92	−4.54
$N\,(> \Delta\chi^2(\alpha = \alpha^{\rm exp}))$	21	881	269
sky-cut OuterRim mocks			
$N(\alpha \pm \sigma \notin [0.8, 1.2])$	25	23	6
$\Delta\chi^2(\alpha = \alpha_{\rm NSGC})$	−5.38	−4.48	−9.85
$N\,(> \Delta\chi^2(\alpha = \alpha_{\rm NSGC}))$	6	601	68
$\Delta\chi^2(\alpha = \alpha^{\rm exp})$	−4.72	−4.07	−8.79
$N\,(> \Delta\chi^2(\alpha = \alpha^{\rm exp}))$	17	654	102

Table 5.12 – Data versus mock BAO *detection*, according to different criteria. The BAO signal is noticeable in eBOSS SGC post-reconstruction data, not in NGC data, where α and its error bar are not within $[0.8, 1.2]$ ($\alpha \pm \sigma \notin [0.8, 1.2]$). $N\,(\alpha \pm \sigma \notin [0.8, 1.2])$ is the number of mocks not satisfying this criterion, for EZ mocks (using $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$) with all systematics implemented and for OuterRim mocks (using $\Sigma_{\rm nl} = 2.4\,{\rm Mpc}\,h^{-1}$). The χ^2 difference $\Delta\chi^2$ between the *wiggle* and *no-wiggle* template best fits for data and mocks is provided at the α value measured in the NGC + SGC combination ($\Delta\chi^2(\alpha = \alpha_{\rm NSGC})$) and at the fiducial α value ($\Delta\chi^2(\alpha = \alpha^{\rm exp})$). $N\left(> \Delta\chi^2(\alpha = \alpha_{\rm NSGC})\right)$ and $N\left(> \Delta\chi^2(\alpha = \alpha^{\rm exp})\right)$ are the number of mocks which show a larger $\Delta\chi^2$ than the data (and hence weaker BAO *detection*) at $\alpha = \alpha_{\rm NSGC}$ and $\alpha = \alpha^{\rm exp}$, respectively.

et al., 2020). SGC only and combined NGC + SGC fits are similar, to the 1.1% level ($0.6 < z < 1.1$ in Table 5.11), which is statistically expected, as seen in 755/1000 EZ mocks (considering both tails). Hence, there is no hint of a strong, unexpected systematic shift in the combined fit, due the addition of potentially contaminated NGC data. We also note that this 1.1% shift is smaller than 1.3%, the uncertainty related to photometric systematics included in our systematic budget (see Table 5.13), and hence is already accounted for if the NGC was the major source of photometric systematics. As a conclusion, we do not see any reason to reject NGC data in the fit. Moreover, combining NGC + SGC turns out to be more optimal than considering SGC alone, even given the poor BAO detection in NGC. To check this, we select EZ mocks with $\Delta\chi^2(\alpha = \alpha_{\rm NSGC}) > 0$ in the NGC but $\Delta\chi^2(\alpha = \alpha_{\rm NSGC}) < 0$ in the SGC among mocks for which the NGC + SGC combination fulfills $\alpha \pm \sigma \in [0.8, 1.2]$. The dispersion of α measurements in the obtained sample of 130 mocks is 0.048 for NGC + SGC, less than 0.058 for SGC alone. In addition, a Kolmogorov-Smirnov test (similar to that of Figure 5.32) does not show any misestimation of error bars (p-value of 0.533 for the residuals to be consistent with a standard normal distribution). A similar reduction of error bars is seen in data in Table 5.11 ($^{+0.025}_{-0.028}$ versus $^{+0.032}_{-0.035}$). Hence, we find it legitimate to combine NGC and SGC measurements.

To further check that error bars are still correctly estimated in the low signal-to-noise regime, we select NGC + SGC EZ mocks for which $\Delta\chi^2(\alpha = \alpha_{\rm NSGC})$ is larger than that observed in the data (while still restricting to $\alpha \pm \sigma \in [0.8, 1.2]$). Again, in this sample of 165 EZ mocks, a Kolmogorov-Smirnov test shows no hint for a misestimation of error bars (the p-value for the residuals to be consistent with a standard normal distribution is 0.459). Therefore, the method to estimate statistical uncertainties in data appears to be correct. We note however that data statistical error bars ($^{+0.025}_{-0.028}$, see Table 5.11) are smaller than those seen in mocks on average ($\sigma \simeq 0.033$, see Table 5.10); 246/1000 EZ mocks have smaller statistical error bars than data (using $\Sigma_{\rm nl} = 4\,{\rm Mpc}\,h^{-1}$ for both data and mocks). In a sample of EZ mocks for which $\Delta\chi^2(\alpha = \alpha_{\rm NSGC})$ is within ± 1 of the data $\Delta\chi^2(\alpha = \alpha_{\rm NSGC})$, which may be considered as representative of the data affinity for BAO, we find 6/157 mocks to have smaller statistical error bars than the data [13]. Due to the low average BAO amplitude in the EZ mocks and the additional residual systematics in the data, this fraction slightly underestimates the probability to obtain smaller error bars than in data. Altogether, though small, data statistical error bars are not too unlikely.

We now turn to stability tests performed on data. Beforehand, we emphasise that despite the low significance of the BAO signal, a robust measurement of the BAO position is possible because the relative amplitude of oscillations is imposed as a prior in the BAO model (Section 5.1.3). In other words, though a model without BAO is not disfavoured by the data, a model with BAO far from the maximum of the likelihood is a significantly worse fit to the data.

In Raichoor et al. (2020) (Figure 10), variations of the redshift density with photometric depth were noted to be relatively higher in the low redshift end, $0.6 < z \lesssim 0.7$. We therefore test the robustness of our measurement with the lower z-limit. Best fits do not move significantly with the lower z cut (given the change in the sample statistics), as can also be seen in Figure 5.35. For example, between $0.6 < z < 1.1$ and $0.7 < z < 1.1$ the best fit α moves by 0.4%. This is not significant as a larger shifts happen in 378/1000 (considering both tails: 741/1000) EZ mocks. We thus use the full redshift range $0.6 < z < 1.1$ for our isotropic BAO measurement.

13. This fraction increases to 39/157 when comparing data total error bars (including systematics, see Table 5.13) to mock statistical error bars.

The α measurement remains very stable with the assumed $\Sigma_{\rm nl}$ value in SGC only. For NGC + SGC, some 0.8% shift (27% of the statistical uncertainty) is obtained between $\Sigma_{\rm nl} = 4\,{\rm Mpc}\,h^{-1}$ and $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$. A larger shift happens for 97/1000 (considering both tails: 214/1000) EZ mocks. To account for the uncertainty in the expected amplitude of the BAO signal, we include this shift as an additional uncertainty.

We estimate the potential residual systematics due to the imperfect modelling of the variations of the survey selection function with imaging quality (see Section 5.3.1.4 and Section 5.3.1.6) by comparing the measurement obtained with the baseline correction to that without any mitigation. Namely, random redshifts are taken separately in each `chunk` (instead of `chunk_z`) to measure the data power spectrum. The covariance matrix is built from mocks where we do not introduce photometric systematics nor variations of the redshift density with `chunk_z`. We take care to change the model window functions (see Section 5.1.2.1) accordingly. A shift in α of 1.3%, which we take as systematic uncertainty, is noticed between these two configurations. One would notice that this shift is mainly driven by the NGC, since a variation of 0.6% only is seen in the SGC, as can be expected since photometry is deeper in the SGC than in the NGC.

The residual systematics remaining after the redshift failure correction (see Section 5.3.1.3) are similarly estimated by comparing the measurement obtained with the baseline correction to that without any mitigation (*no* $w_{\rm noz,i}$). Namely, no $w_{\rm noz,i}$ are applied to measure the data power spectrum, while we do not introduce redshift failures in the mocks to construct the covariance matrix. This leads to a shift in α of 0.2%; a larger variation is seen in 438/1000 (considering both tails: 851/1000) EZ mocks, making it quite likely. Note however that data and EZ mock shifts are expected to match since a mock object is deterministically declared a redshift failure if the redshift of its nearest neighbour in the data could not be reliably measured. As mentioned in Section 5.4.8, a second set of mocks was produced with redshift failures being randomly drawn from the model fitted to the data; in this case, we find a larger variation in 378/1000 (considering both tails: 826/1000) EZ mocks. Hence, the shift seen in the data is well explained by the mocks. We conservatively include it in the systematic budget to account for the uncertainty in the redshift failure correction.

Changing the covariance matrix for that based on GLAM-QPM mocks, which does not include angular nor radial photometric systematics leads to a small α shift of 0.4% in the *no* ***`chunk_z`*** case (*no* ***`chunk_z`*** *G-Q cov* versus *no* ***`chunk_z`***). Similarly, in the baseline case, using a covariance matrix built from GLAM-QPM or EZ mocks without systematics (only the *shuffled* scheme) induces small shifts of 0.2% (*GLAM-QPM cov* versus *baseline*). A larger shift is observed in the former case for 283/1000 (considering both tails: 732/1000) EZ mocks, in the latter case for 377/1000 (considering both tails: 713/1000) EZ mocks. The parameter covariance matrix correction (see Eq. (5.14)) already leads to an increase of 0.7% of the statistical error, which is higher than obtained by summing the shifts above in quadrature (0.3%). We therefore do not include any systematic uncertainty related to the choice of the covariance matrix. Dividing the number of mocks used to build the covariance matrix by 2, we find a shift in α of 0.5% (*500 mocks in cov* versus *baseline*). This would lead to a 1.9% increase of the error if added in quadrature, while the increase of error bars required to account for the change in covariance matrix between these two configurations is 1.5%. Hence, the shift seen in the fit is compatible with a statistical fluctuation and we conclude that the estimation of the covariance matrix is robust enough for this measurement.

Moving the centre of the k-bin by half a bin ($0.005\,h\,{\rm Mpc}^{-1}$, *+ 1/2 k-bin*) leads to a α shift of 0.5%, which is compatible with a statistical fluctuation since a larger shift is observed

source	α
linear	
model non-linearity (from EZ mocks)	0.1%
quadrature	
modelling systematics (from mock challenge)	0.2%
damping term $\Sigma_{\rm nl}$	0.8%
photometric systematics	1.3%
fibre collisions (from EZ mocks)	0.2%
redshift failures	0.3%
statistics	$^{+0.025}_{-0.028}$
systematics	$^{+0.006}_{-0.005}$
total	$^{+0.031}_{-0.033}$

Table 5.13 – Error budget for post-reconstruction isotropic BAO measurements on the eBOSS DR16 ELG sample. Percentages are provided with respect to the α value. The last three lines (statistics, systematics and total) recap the absolute statistical error bar, the systematic contribution (total minus statistical), and the total error bar, respectively.

for 307/1000 (considering both tails: 595/1000) EZ mocks. We therefore do not include any additional uncertainty related to the choice of k-bins in our systematic budget.

Finally, we change the fiducial cosmology (5.4) for the OuterRim cosmology (2.89), and for comparison purposes we report the α value rescaled to our fiducial cosmology. Some shift can be seen (*OR cosmo* versus *baseline*), which we relate to the change of damping term. Indeed, we checked that analysing a power spectrum in our fiducial cosmology with an OuterRim template leads to a preferred $\Sigma_{\rm nl} \simeq 5.4\,{\rm Mpc}\,h^{-1}$. The measurement obtained with $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$ and OuterRim cosmology (*OR cosmo* $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$) is indeed close (within 0.2%) to the measurement using $\Sigma_{\rm nl} = 4\,{\rm Mpc}\,h^{-1}$ and our fiducial cosmology ($0.6 < z < 1.1$ *(fiducial)*). This shift can be seen in 428/1000 (considering both tails: 840/1000) EZ mocks and is within the 0.2% modelling uncertainty derived in Section 5.2.5. We therefore do not include any additional uncertainty due to the assumed fiducial cosmology.

Overall, despite the mild preference for BAO in the eBOSS ELG sample, the BAO measurement appears relatively robust.

The final error budget is reported in Table 5.13. Since the 0.2% α shift attributed to model non-linearity is closer to a bias than a systematic uncertainty, we decided to be conservative and to add it linearly to the statistical uncertainty. All other contributions to the systematic budget are uncertainties due to the model accuracy (modelling systematics) or our limited knowledge of the survey selection function (photometric systematics, fibre collisions, redshift failures) or are analysis choices (damping term). We thus add them in quadrature to the statistical uncertainty.

Our final post-reconstruction isotropic BAO measurement is $\alpha = 0.986^{+0.031}_{-0.033}$, including statistical and systematic uncertainties. In terms of the volume-averaged distance $D_{\rm V}(z)$, we find:

$$\boxed{D_{\rm V}(z_{\rm eff} = 0.845)/r_{\rm drag} = 18.33^{+0.57}_{-0.62}} \tag{5.60}$$

independently of the assumed fiducial cosmology.

In order to generate the BAO likelihood profile for further cosmological inference, we first rescale $\Delta\chi^2(\alpha) = \chi^2(\alpha) - \chi^2(\alpha_0)$ (with α_0 the best fit value) by the inverse of the parameter

covariance rescaling (5.14). To include the systematic error budget we further rescale $\Delta\chi^2(\alpha)$ by the ratio $(\sigma_{\alpha,\rm stat}/\sigma_{\alpha,\rm tot})^2$, with $\sigma_{\alpha,\rm tot}$ and $\sigma_{\alpha,\rm stat}$ the total and statistical upper (lower) error bars when $\alpha > \alpha_0$ ($\alpha < \alpha_0$). We finally provide the BAO likelihood $e^{-\Delta\chi^2(D_{\rm V}/r_{\rm drag})/2}$, with $D_{\rm V} = \alpha D_{\rm V}^{\rm fid}$.

5.5.2 Combined RSD and BAO measurements

RSD and combined RSD + BAO measurements are reported under different fitting conditions in Table 5.14. As in Section 5.5.1, we take $\Sigma_{\rm nl} = 4\,{\rm Mpc}\,h^{-1}$ as fiducial value for the BAO damping and RSD and BAO templates are computed within the fiducial cosmology (5.4), except otherwise stated. Again, we quote results for the combined caps NGC + SGC and SGC alone.

As already noted in Section 5.5.1, variations of the redshift density with photometric depth were observed to be relatively higher in the low redshift end. We therefore test the robustness of our measurement with respect to the lower redshift cut. For the RSD + BAO measurement, shifts of 6.4%, 4.1% and 0.4% are seen for $f\sigma_8$, $\alpha_\parallel$ and $\alpha_\perp$, respectively, when changing the redshift range $0.6 < z < 1.1$ to $0.7 < z < 1.1$ for the pre-reconstruction power spectrum (RSD part of the likelihood). Larger shifts are observed for 159/1000 (considering both tails: 296/1000) EZ mocks for $f\sigma_8$, 19/1000 (32/1000) for $\alpha_\parallel$ and 374/1000 (732/1000) for $\alpha_\perp$. The somewhat low probability of the change in $\alpha_\parallel$ may point towards some unaccounted systematics in the lower redshift end. These potential systematics were without effect on the BAO measurement, as noted in Section 5.5.1.

For our final measurement we therefore conservatively choose the redshift range $0.7 < z < 1.1$ for the RSD part of the likelihood, and keep $0.6 < z < 1.1$ for the BAO part, as decided in Section 5.5.1. Following Eq. (5.19), the effective redshift for the redshift range $0.6 < z < 1.1$ is 0.845 while it is 0.857 for $0.7 < z < 1.1$. We choose the effective redshift as 0.85 for the combined RSD + BAO measurement, and check that the expected variations of $f\sigma_8$, $D_{\rm H}/r_{\rm drag}$ and $D_{\rm M}/r_{\rm drag}$ over the redshift range $0.845-0.857$ within our fiducial cosmology are small (0.3%, 0.7% and 1.1%, respectively) compared to the statistical uncertainty on data (24%, 9% and 4%, respectively).

Best fit models of ELG power spectra in the NGC and SGC are compared with data in Figure 5.37, while Figure 5.36 shows the posteriors of the RSD and RSD + BAO measurements. One would note that the RSD + BAO combination helps reducing the posterior tails while not changing the central values. In particular, the combination of RSD with BAO removes secondary local minima in the contours. An illustration of this ill-shaped RSD only posteriors is the large change for the RSD only best fit to the NGC + SGC data in the redshift range $0.7 < z < 1.1$ with respect to $0.6 < z < 1.1$, as reported in Table 5.14. In particular, the bump seen on the left side of the peak in the $\alpha_\parallel$ marginal RSD only posterior in Figure 5.36 corresponds to the position of the best fit value $\alpha_\parallel = 0.871^{+0.109}_{-0.061}$. We checked that the overall shift in the RSD only posterior contours when changing the redshift range is much smaller than that of the RSD only best fits and consistent with that of the RSD + BAO case. Note that in the following the systematic budget will be derived for the RSD + BAO combination, for which the shape of the posterior is more Gaussian.

As in Section 5.5.1, we test the robustness of our result with respect to the choice of the $\Sigma_{\rm nl}$ value: we find a 1.9% shift on $f\sigma_8$, 0.1% on $\alpha_\parallel$ and 0.8% on $\alpha_\perp$ between $\Sigma_{\rm nl} = 4\,{\rm Mpc}\,h^{-1}$ and $\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$, which we add to our uncertainty budget. Larger shifts are seen for 104/1000

	$f\sigma_8$	$\alpha_\parallel$	$\alpha_\perp$	χ^2/dof
RSD only	SGC only	z cuts		
$0.6 < z < 1.1$	$0.348^{+0.078}_{-0.079}$	$0.974^{+0.100}_{-0.084}$	$0.952^{+0.071}_{-0.062}$	$31.7/(46-7) = 0.812$
$0.65 < z < 1.1$	$0.367^{+0.086}_{-0.085}$	$0.982^{+0.097}_{-0.093}$	$0.968^{+0.090}_{-0.073}$	$29.2/(46-7) = 0.749$
$0.7 < z < 1.1$	$0.38^{+0.10}_{-0.11}$	$1.013^{+0.087}_{-0.098}$	$1.04^{+0.13}_{-0.13}$	$29.8/(46-7) = 0.763$
$0.75 < z < 1.1$	$0.33^{+0.23}_{-0.10}$	$0.98^{+0.11}_{-0.16}$	$0.936^{+0.328}_{-0.082}$	$31.3/(46-7) = 0.802$
RSD only	SGC only	$0.7 < z < 1.1$		
GLAM-QPM cov	$0.429^{+0.081}_{-0.082}$	$1.022^{+0.082}_{-0.079}$	$1.120^{+0.098}_{-0.100}$	$28.6/(46-7) = 0.733$
no syst. cov	$0.409^{+0.077}_{-0.082}$	$1.027^{+0.076}_{-0.074}$	$1.110^{+0.087}_{-0.098}$	$29.3/(46-7) = 0.752$
+ 1/2 k-bin	$0.374^{+0.093}_{-0.099}$	$1.023^{+0.083}_{-0.079}$	$1.052^{+0.104}_{-0.098}$	$25.0/(46-7) = 0.640$
OR cosmo (rescaled)	$0.371^{+0.090}_{-0.106}$	$1.029^{+0.075}_{-0.082}$	$1.07^{+0.10}_{-0.12}$	$24.9/(46-7) = 0.638$
RSD + BAO	SGC only	z cuts		
$0.6 < z < 1.1$	$0.327^{+0.084}_{-0.105}$	$1.017^{+0.133}_{-0.087}$	$0.967^{+0.059}_{-0.061}$	$51.0/(73-12) = 0.837$
$0.65 < z < 1.1$	$0.348^{+0.090}_{-0.107}$	$1.013^{+0.120}_{-0.087}$	$0.976^{+0.064}_{-0.066}$	$46.1/(73-12) = 0.756$
$0.7 < z < 1.1$	$0.335^{+0.099}_{-0.124}$	$1.004^{+0.122}_{-0.097}$	$0.986^{+0.071}_{-0.083}$	$50.0/(73-12) = 0.819$
$0.75 < z < 1.1$	$0.33^{+0.12}_{-0.16}$	$1.01^{+0.16}_{-0.12}$	$0.964^{+0.087}_{-0.096}$	$52.4/(73-12) = 0.859$
RSD + BA0	SGC only	$0.7 < z < 1.1$		
GLAM-QPM cov	$0.425^{+0.088}_{-0.092}$	$0.950^{+0.095}_{-0.103}$	$1.033^{+0.058}_{-0.066}$	$53.5/(73-12) = 0.877$
no syst. cov	$0.386^{+0.079}_{-0.085}$	$0.979^{+0.079}_{-0.073}$	$1.028^{+0.057}_{-0.062}$	$47.6/(73-12) = 0.780$
+ 1/2 k-bin	$0.31^{+0.11}_{-0.12}$	$1.04^{+0.12}_{-0.10}$	$0.973^{+0.074}_{-0.078}$	$47.6/(73-12) = 0.780$
OR cosmo (rescaled)	$0.31^{+0.10}_{-0.11}$	$1.01^{+0.11}_{-0.11}$	$0.983^{+0.069}_{-0.070}$	$45.4/(73-12) = 0.743$
RSD only	NGC + SGC	z cuts		
$0.6 < z < 1.1$	$0.250^{+0.124}_{-0.067}$	$1.15^{+0.11}_{-0.28}$	$0.919^{+0.038}_{-0.039}$	$87.6/(92-11) = 1.08$
$0.65 < z < 1.1$	$0.259^{+0.116}_{-0.068}$	$1.15^{+0.10}_{-0.27}$	$0.922^{+0.040}_{-0.041}$	$83.9/(92-11) = 1.04$
$0.7 < z < 1.1$	$0.382^{+0.053}_{-0.056}$	$0.871^{+0.109}_{-0.061}$	$0.901^{+0.043}_{-0.050}$	$81.3/(92-11) = 1.00$
$0.75 < z < 1.1$	$0.365^{+0.062}_{-0.073}$	$0.905^{+0.146}_{-0.080}$	$0.887^{+0.045}_{-0.053}$	$66.6/(92-11) = 0.822$
RSD only	NGC + SGC	$0.7 < z < 1.1$		
GLAM-QPM cov	$0.308^{+0.104}_{-0.074}$	$1.08^{+0.11}_{-0.27}$	$0.942^{+0.047}_{-0.083}$	$80.3/(92-11) = 0.991$
no syst. cov	$0.388^{+0.052}_{-0.054}$	$0.865^{+0.106}_{-0.061}$	$0.910^{+0.046}_{-0.052}$	$83.2/(92-11) = 1.03$
+ 1/2 k-bin	$0.376^{+0.055}_{-0.134}$	$0.887^{+0.272}_{-0.071}$	$0.910^{+0.047}_{-0.048}$	$68.5/(92-11) = 0.846$
OR cosmo (rescaled)	$0.289^{+0.128}_{-0.075}$	$1.08^{+0.11}_{-0.27}$	$0.933^{+0.043}_{-0.075}$	$73.4/(92-11) = 0.906$
RSD + BAO	NGC + SGC	z cuts		
$0.6 < z < 1.1$	$0.271^{+0.059}_{-0.057}$	$1.129^{+0.078}_{-0.090}$	$0.938^{+0.030}_{-0.030}$	$140./(146-21) = 1.12$
$0.65 < z < 1.1$	$0.281^{+0.059}_{-0.057}$	$1.122^{+0.076}_{-0.085}$	$0.938^{+0.031}_{-0.031}$	$137/(146-21) = 1.10$
$0.7 < z < 1.1$ (baseline)	$0.289^{+0.068}_{-0.066}$	$1.085^{+0.087}_{-0.107}$	$0.941^{+0.035}_{-0.034}$	$141/(146-21) = 1.13$
$0.75 < z < 1.1$	$0.319^{+0.068}_{-0.069}$	$1.062^{+0.088}_{-0.092}$	$0.937^{+0.037}_{-0.037}$	$123/(146-21) = 0.984$
RSD + BAO	NGC + SGC	$0.7 < z < 1.1$		
no `chunk_z`	$0.261^{+0.059}_{-0.058}$	$1.109^{+0.070}_{-0.074}$	$0.928^{+0.036}_{-0.036}$	$153/(146-21) = 1.23$
no `chunk_z` GLAM-QPM cov	$0.262^{+0.061}_{-0.061}$	$1.137^{+0.077}_{-0.078}$	$0.926^{+0.035}_{-0.034}$	$136/(146-21) = 1.09$
$\Sigma_{\rm nl} = 6\,{\rm Mpc}\,h^{-1}$	$0.283^{+0.101}_{-0.069}$	$1.086^{+0.096}_{-0.284}$	$0.934^{+0.036}_{-0.036}$	$141/(146-21) = 1.13$
no $w_{\rm noz,i}$	$0.306^{+0.071}_{-0.068}$	$1.088^{+0.089}_{-0.112}$	$0.939^{+0.034}_{-0.034}$	$144/(146-21) = 1.15$
GLAM-QPM cov	$0.305^{+0.066}_{-0.065}$	$1.101^{+0.078}_{-0.081}$	$0.956^{+0.037}_{-0.035}$	$132/(146-21) = 1.06$
no syst. cov	$0.326^{+0.072}_{-0.078}$	$1.03^{+0.10}_{-0.11}$	$0.952^{+0.037}_{-0.038}$	$125/(146-21) = 1.00$
500 mocks in cov	$0.301^{+0.071}_{-0.070}$	$1.084^{+0.082}_{-0.092}$	$0.941^{+0.037}_{-0.036}$	$127/(146-21) = 1.02$
+ 1/2 k-bin	$0.287^{+0.067}_{-0.064}$	$1.082^{+0.081}_{-0.087}$	$0.943^{+0.035}_{-0.033}$	$140./(146-21) = 1.12$
OR cosmo (rescaled)	$0.271^{+0.066}_{-0.063}$	$1.098^{+0.075}_{-0.086}$	$0.938^{+0.034}_{-0.032}$	$129/(146-21) = 1.03$
Bayesian	$0.289^{+0.060}_{-0.075}$	$1.085^{+0.104}_{-0.090}$	$0.941^{+0.036}_{-0.037}$	$141/(146-21) = 1.13$

Table 5.14 – RSD and RSD + BAO best fits to the eBOSS DR16 ELG sample. Error bars are defined by the $\Delta\chi^2 = 1$ level, except in the *Bayesian* case, where we consider the minimum interval which contains 68% of the MCMC samples.

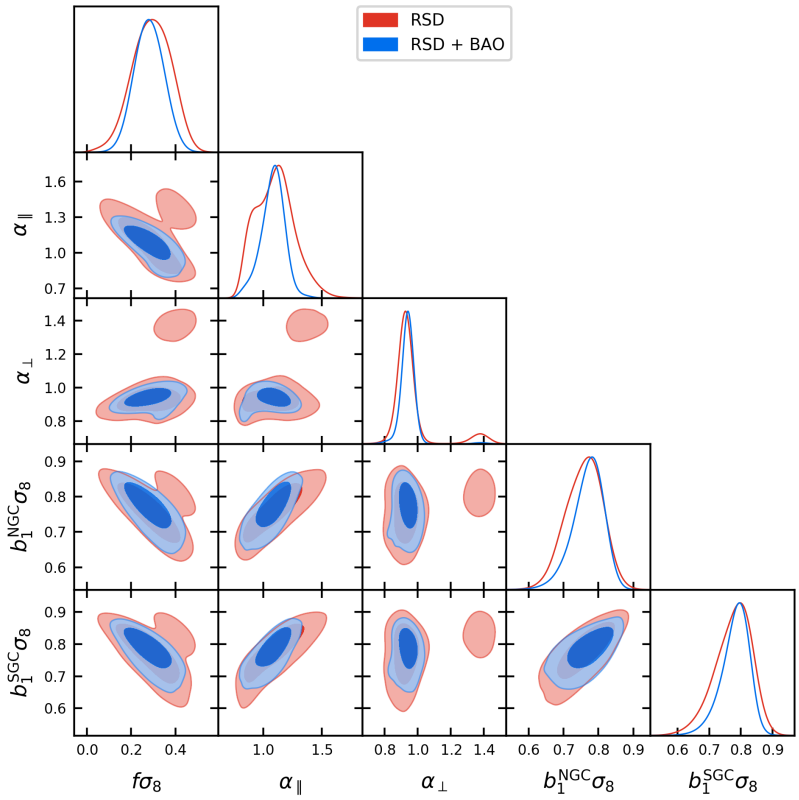

Figure 5.36 – Posteriors of the RSD and RSD + BAO measurements. Contours contain 68% and 95% of the probability. Systematic uncertainties of Table 5.15 are not included.

(both tails: 142/1000), 402/1000 (861/1000) and 211/1000 (345/1000) EZ mocks for $f\sigma_8$, $\alpha_\parallel$ and $\alpha_\perp$.

The uncertainty in the modelling of the selection function variations with photometry is estimated similarly to Section 5.5.1: we compare the baseline correction, using the *pixelated* scheme, to the best fit measurement (*no* ***chunk_z***) obtained without any angular mitigation scheme nor modelling of the variations of the radial selection function with `chunk_z`. The covariance matrix is also built from EZ mocks without angular nor radial photometric systematics. We measure shifts of 9.4% on $f\sigma_8$, 2.2% on $\alpha_\parallel$ and 1.4% on $\alpha_\perp$, which we take as additional systematic uncertainty.

Similarly, the impact of redshift failures is estimated by comparing the baseline measurement with the best fit obtained on the data without including the $w_{\rm noz,i}$ correction weight (*no* $w_{\rm noz,i}$). We measure shifts of 6.1% on $f\sigma_8$, 0.3% on $\alpha_\parallel$ and 0.3% on $\alpha_\perp$, which we take as systematic uncertainty. Larger variations happen for 176/1000 (considering both tails: 368/1000), 479/1000 (856/1000) and 409/1000 (882/1000) EZ mocks, respectively. If we rather consider EZ mocks where redshift failures are stochastic, based on the model fitted to the observed data, we find larger shifts for 146/1000 (considering both tails: 318/1000), 499/1000 (880/1000) and 403/1000 (862/1000) EZ mocks for $f\sigma_8$, $\alpha_\parallel$ and $\alpha_\perp$, respectively.

Dividing the number of mocks used to build the covariance matrix by 2, we find a shift of 4.3% on $f\sigma_8$, 0.1% on $\alpha_\parallel$ and 0.0% on $\alpha_\perp$ (*500 mocks in cov* versus *baseline*). This would lead to a 1.7%, 0.0% and 0.0% increase of the error if added in quadrature, less than 5.8%, the typical increase of error bars required to account for the change in covariance matrix between these two

configurations. We thus conclude that the RSD + BAO covariance matrix is stable with respect to the number of mocks.

Changing the covariance matrix built from contaminated EZ mocks for the one obtained from GLAM-QPM mocks, we find a 5.6% shift on $f\sigma_8$, 1.5% on $\alpha_\parallel$ and 1.6% on $\alpha_\perp$. Larger variations happen for 171/1000, 218/1000, 107/1000 EZ mocks, respectively (considering both tails: 363/1000, 402/1000 and 290/1000). Even larger shifts are seen when using a covariance matrix based on EZ mocks without systematics (only the *shuffled* scheme, *no syst. cov*): 12.9% shift on $f\sigma_8$, 4.7% on $\alpha_\parallel$ and 1.2% on $\alpha_\perp$. These changes happen for 12/1000, 18/1000, 252/1000 EZ mocks, respectively (considering both tails: 27/1000, 50/1000 and 428/1000).

One would also notice a large change (26.7% on $f\sigma_8$, 5.4% on $\alpha_\parallel$ and 4.7% on $\alpha_\perp$) for the SGC only measurements with the GLAM-QPM covariance matrix (*GLAM-QPM cov*). A larger shift in $f\sigma_8$ happens for 4/1000 EZ mocks (considering both tails: 18/1000). One may be concerned by some possible coupling between the data and the covariance matrix from contaminated EZ mocks, as the map of angular systematics injected in these mocks was inferred directly from the observed data, after smoothing (see Section 5.3.3.3). To check the importance of this potential effect, we generate a new systematic map based on the angular target density measured on one contaminated EZ *reference mock*, and contaminate the other 999 EZ mocks with this new map. Best fits to the *reference mock* using the original covariance matrix and using the new one are compared. Performing RSD fits to the SGC, we find small shifts of 3.5% for $f\sigma_8$, 0.2% for $\alpha_\parallel$ and 0.3% for $\alpha_\perp$, which are not significant since larger shift are seen in 116/999, 471/999 and 440/999 EZ mocks (considering both tails: 219/999, 901/999 and 783/999).

The variations in the best fit parameters with the covariance matrix being very untypical, for conservativeness we include in our systematic budget the largest shifts seen in the NGC + SGC fit, namely those obtained with a covariance matrix based on EZ mocks without systematics (only the *shuffled* scheme, *no syst. cov*), i.e. 12.9% on $f\sigma_8$, 4.7% on $\alpha_\parallel$ and 1.2% on $\alpha_\perp$.

Moving the centre of the k-bin by half a bin ($0.005\,h\,\mathrm{Mpc}^{-1}$) leads to a shift of 0.6% for $f\sigma_8$ and 0.2% for $\alpha_\parallel$ and $\alpha_\perp$. A larger $f\sigma_8$ shift is observed for 393/1000 (considering both tails: 931/1000) EZ mocks, 463/1000 (884/1000) for $\alpha_\parallel$ and 539/1000 (922/1000) for $\alpha_\perp$. As in Section 5.5.1, since these shifts are compatible with mocks, we do not include them in the systematic budget.

Changing the fiducial cosmology of Eq. (5.4) for the OuterRim cosmology of Eq. (2.89), we find moderate shifts of 6.0% on $f\sigma_8$ and 1.2% on $\alpha_\parallel$ and 0.3% on $\alpha_\perp$ (which are rescaled to the fiducial cosmology Eq. (5.4) for comparison purposes). Larger shifts are observed for 263/1000 (considering both tails: 349/1000) EZ mocks for $f\sigma_8$, 283/1000 (474/1000) for $\alpha_\parallel$ and 312/1000 (831/1000) for $\alpha_\perp$ and hence are fully compatible with a statistical fluctuation. Since we accounted for the change of fiducial cosmology in the systematic modelling budget, we do not quote any other related systematic uncertainty.

The final error budget is reported in Table 5.15. The systematic bias related to the analysis methodology, namely the survey geometry, the RSD and BAO combination, the model non-linearity and the modelling of the angular and radial integral constraints (ARIC) are summed together in quadrature and added linearly to the statistical error bars. Other contributions are uncertainties due to our limited understanding of the ELG small scale clustering (modelling systematics) or the survey selection function (photometric systematics, fibre collisions, redshift failures) or consist in analysis choices (damping term, covariance matrix, fiducial cosmology). These other terms are added in quadrature to the statistical error bars.

source	$f\sigma_8$	$\alpha_\parallel$	$\alpha_\perp$
linear			
survey geometry (from EZ mocks)	0.1%	0.5%	0.1%
RSD + BAO combination (from EZ mocks)	1.0%	0.3%	0.5%
model non-linearity (from EZ mocks)	1.6%	1.0%	0.4%
ARIC modelling (from EZ mocks)	0.8%	0.6%	0.1%
quadrature			
modelling systematics (from mock challenge)	3.0%	0.9%	0.8%
damping term $\Sigma_{\rm nl}$	1.9%	0.1%	0.8%
photometric systematics	9.4%	2.2%	1.4%
fibre collisions (from EZ mocks)	1.1%	0.6%	0.2%
redshift failures	6.1%	0.3%	0.3%
covariance matrix	12.9%	4.7%	1.2%
statistics	$^{+0.060}_{-0.075}$	$^{+0.104}_{-0.090}$	$^{+0.036}_{-0.037}$
systematics	$^{+0.024}_{-0.021}$	$^{+0.029}_{-0.031}$	$^{+0.012}_{-0.011}$
total	$^{+0.085}_{-0.096}$	$^{+0.13}_{-0.12}$	$^{+0.048}_{-0.049}$

Table 5.15 – Error budget for RSD + BAO measurements on the eBOSS DR16 ELG sample. Percentages are provided with respect to the parameter value. The last three lines (statistics, systematics and total) recap the absolute statistical error bar, the systematic contribution (total minus statistical), and the total error bar, respectively.

Including both statistical and systematic uncertainties, our final combined RSD + BAO measurement is (in the *Bayesian* case): $f\sigma_8 = 0.289^{+0.085}_{-0.096}$, $\alpha_\parallel = 1.08^{+0.13}_{-0.12}$ and $\alpha_\perp = 0.941 \pm 0.049$.

In terms of angular and Hubble distances, we find:

$$\begin{aligned} f\sigma_8(z_{\rm eff} = 0.85) &= 0.289^{+0.085}_{-0.096} \\ D_{\rm H}(z_{\rm eff} = 0.85)/r_{\rm drag} &= 20.0^{+2.4}_{-2.2} \\ D_{\rm M}(z_{\rm eff} = 0.85)/r_{\rm drag} &= 19.17 \pm 0.99 \end{aligned} \tag{5.61}$$

As can be seen in Figure 5.36, the linear bias combination $b_1\sigma_8$ is correlated with $f\sigma_8$. Fixing σ_8 to the fiducial cosmology in Eq. (5.4) we find $b_1^{\rm NGC} = 1.49 \pm 0.10$ and $b_1^{\rm SGC} = 1.52^{+0.10}_{-0.11}$, in agreement with previous studies (e.g. Comparat et al., 2013). Best fit values and errors for all parameters are given in Table 5.16. No discrepancy can be seen between NGC and SGC nuisance parameters ($b_1\sigma_8$, $b_2\sigma_8$, σ_v and A_g). $A_g^{\rm NGC}$ and $A_g^{\rm SGC}$ are compatible with zero, as expected with nearly Poisson shot noise.

To include the systematic error budget in the above RSD + BAO posteriors, we rescale the distance of each MCMC sample $(f\sigma_8, \alpha_\parallel, \alpha_\perp)$ to the median values by the ratio of total to statistical only errors. For example, the new $f\sigma_8$ position is:

$$\frac{\sigma_{f\sigma_8,{\rm tot}}}{\sigma_{f\sigma_8,{\rm stat}}}\left[f\sigma_8 - {\rm median}(f\sigma_8)\right] + {\rm median}(f\sigma_8). \tag{5.62}$$

$\sigma_{f\sigma_8,{\rm tot}}$ and $\sigma_{f\sigma_8,{\rm stat}}$ are the upper (lower) total and statistical error bars with respect to the median when $f\sigma_8 > {\rm median}(f\sigma_8)$ ($f\sigma_8 < {\rm median}(f\sigma_8)$). We proceed similarly for $\alpha_\parallel$ and $\alpha_\perp$.

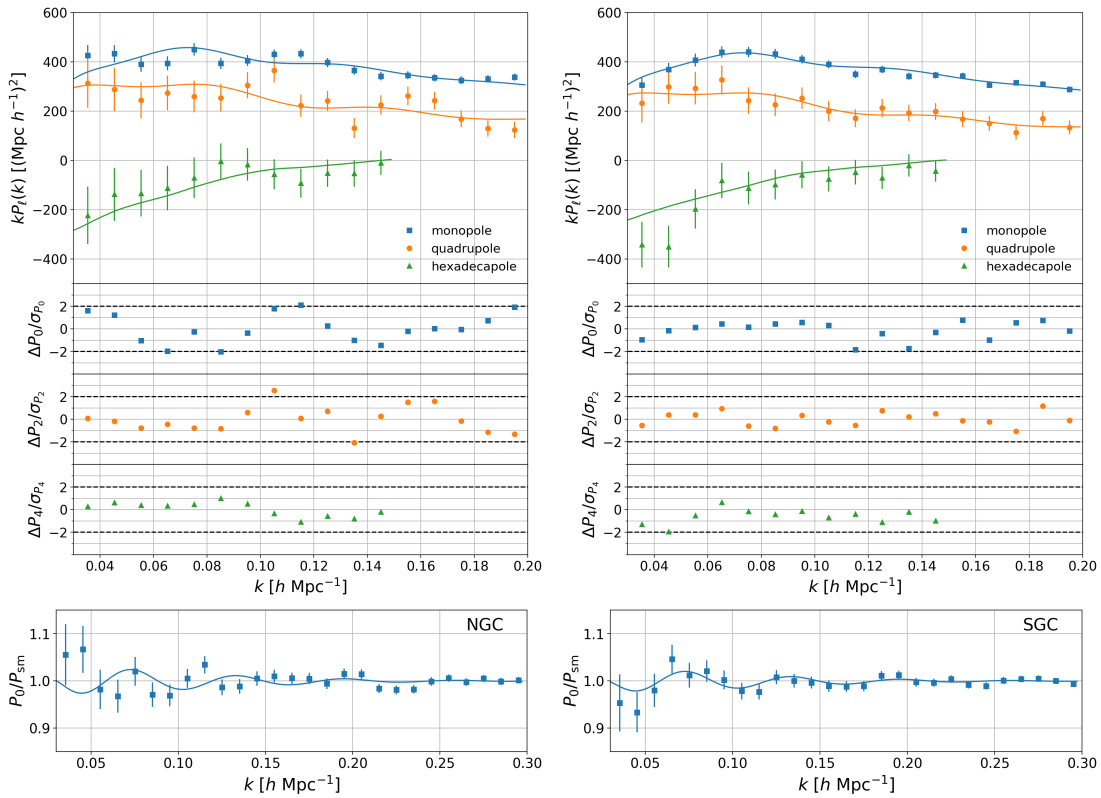

Figure 5.37 – Combined RSD + BAO fit (left: NGC, right: SGC): data points with error bars from the EZ mocks and best fit model as continuous line for the power spectrum multipoles (top), normalised residuals for every power spectrum multipole (middle) and the BAO oscillation pattern (bottom).

	best fit	mean	median	$\Delta\chi^2 = 1$	68% interval	std deviation	$[16\%, 84\%]$
$f\sigma_8$	0.289	0.282	0.281	$^{+0.068}_{-0.066}$	$0.289^{+0.060}_{-0.075}$	0.067	$0.289^{+0.061}_{-0.074}$
$\alpha_\parallel$	1.085	1.082	1.089	$^{+0.087}_{-0.107}$	$1.085^{+0.104}_{-0.090}$	0.103	$1.085^{+0.094}_{-0.102}$
$\alpha_\perp$	0.941	0.942	0.940	$^{+0.035}_{-0.034}$	$0.941^{+0.036}_{-0.037}$	0.061	$0.941^{+0.035}_{-0.039}$
$b_1^{\mathrm{NGC}}\sigma_8$	0.779	0.770	0.774	$^{+0.048}_{-0.066}$	$0.779^{+0.042}_{-0.044}$	0.046	$0.779^{+0.035}_{-0.053}$
$b_1^{\mathrm{SGC}}\sigma_8$	0.795	0.782	0.787	$^{+0.046}_{-0.079}$	$0.795^{+0.038}_{-0.046}$	0.045	$0.795^{+0.029}_{-0.057}$
$b_2^{\mathrm{NGC}}\sigma_8$	-0.23	-0.14	-0.18	$^{+0.97}_{-0.75}$	$-0.23^{+0.77}_{-0.76}$	0.72	$-0.23^{+0.87}_{-0.66}$
$b_2^{\mathrm{SGC}}\sigma_8$	-0.1	-0.01	-0.02	$^{+1.2}_{-1.0}$	$-0.06^{+0.89}_{-0.85}$	0.79	$-0.06^{+0.93}_{-0.81}$
A_g^{NGC}	0.02	0.04	0.01	$^{+0.24}_{-0.15}$	$0.02^{+0.12}_{-0.17}$	0.17	$0.02^{+0.20}_{-0.13}$
A_g^{SGC}	-0.04	0.02	-0.04	$^{+0.35}_{-0.15}$	$-0.04^{+0.14}_{-0.16}$	0.20	$-0.04^{+0.26}_{-0.11}$
σ_v^{NGC}	2.53	2.14	2.23	$^{+0.73}_{-0.93}$	$2.53^{+0.72}_{-1.18}$	0.94	$2.53^{+0.54}_{-1.42}$
σ_v^{SGC}	3.05	2.73	2.81	$^{+0.73}_{-0.82}$	$3.05^{+0.66}_{-0.98}$	0.88	$3.05^{+0.51}_{-1.16}$

Table 5.16 – Parameters for the eBOSS ELG combined RSD + BAO likelihood. We provide the best fit, as well as the mean and the median of the MCMC samples. Error bars are given around the best fit. The $\Delta\chi^2 = 1$ errors in parameter x are given by the x values for which χ^2 is increased by 1 compared to the best fit, when minimising with respect to all other parameters. The 68% interval for parameter x corresponds to the smallest interval of x which contains 68% of the MCMC samples. We also quote the standard deviation of the MCMC samples. The $[16\%, 84\%]$ interval is given by the 16% and 84% percentiles of the MCMC samples. Systematic uncertainties of Table 5.15 are not included.

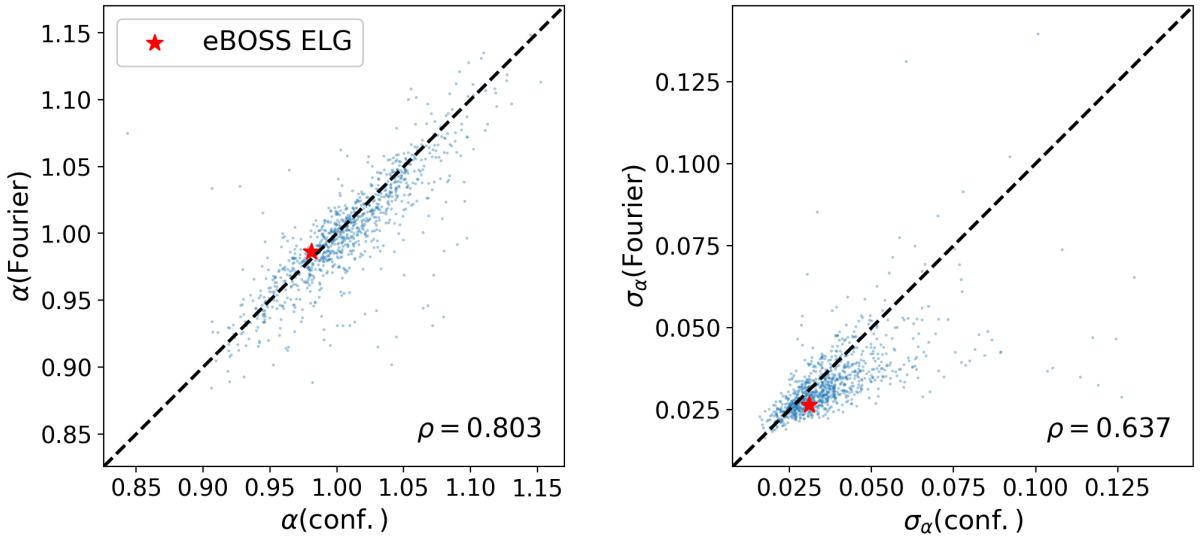

Figure 5.38 – Post-reconstruction isotropic BAO measurements on the EZ mocks including all systematics, in Fourier and configuration space (left: best fits, right: errors, taken to be the mean of the upper and lower error bars). The red cross shows the eBOSS ELG data.

5.5.3 Consensus

BAO and RSD analyses of the eBOSS ELG sample are also performed in configuration space, as detailed in Raichoor et al. (2020); Tamone et al. (2020).

Post-reconstruction isotropic BAO measurements in Fourier and configuration space are compared for each mock in Figure 5.38. Both measurements are well correlated ($\rho = 0.8$). Raichoor et al. (2020) measurement, $D_V(z_{\rm eff} = 0.845)/r_{\rm drag} = 18.23^{+0.58}_{-0.58}$ (statistical only) is $< 0.2\sigma$ away from our Fourier space measurement, $D_V(z_{\rm eff} = 0.845)/r_{\rm drag} = 18.33^{+0.46}_{-0.52}$ (statistical only). A larger difference in the best fits occurs for 223/956 EZ mocks including all systematics (considering both tails: 684/1000), while we find 387/956 EZ mocks (considering both tails: 473/1000) with a larger difference in the mean of the lower and upper error bars. The good agreement between configuration and Fourier space measurements is shown by the data cross in Figure 5.38, lying close to the diagonal. We choose the Fourier space measurement as consensus as it has a lower statistical uncertainty.

This measurement, $D_V(z_{\rm eff} = 0.845)/r_{\rm drag} = 18.33^{+0.57}_{-0.62}$, is 0.6σ below the Planck Collaboration et al. (2018) CMB-based (TT, TE, EE, lowE, lensing) prediction.

Combining RSD and BAO measurements in configuration space and including systematic uncertainties, Tamone et al. (2020) find:

$$\begin{aligned} f\sigma_8(z_{\rm eff} = 0.85) &= 0.35 \pm 0.10 \\ D_H(z_{\rm eff} = 0.85)/r_{\rm drag} &= 19.1^{+1.9}_{-2.0} \\ D_M(z_{\rm eff} = 0.85)/r_{\rm drag} &= 19.9 \pm 1.0 \end{aligned} \tag{5.63}$$

These values are 0.7σ, 0.5σ and 0.7σ away from our Fourier space median $f\sigma_8$, $D_H/r_{\rm drag}$ and $D_M/r_{\rm drag}$ values, respectively. Comparing the best fits instead, differences are 0.4σ, 0.4σ and 0.7σ, respectively. Again comparing best fits, larger differences occur in 386/1000 (considering both tails: 543/1000), 193/1000 (358/1000) and 246/1000 (311/1000) EZ mocks including

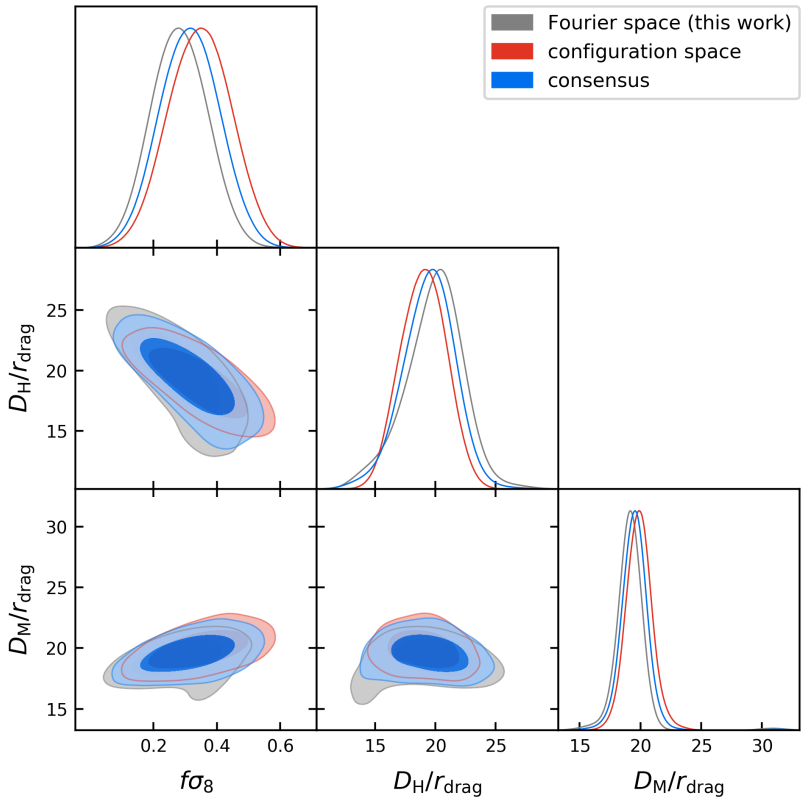

Figure 5.39 – Fourier and configuration space, and combined posteriors of the final eBOSS ELG RSD + BAO measurements (including systematic uncertainties). Contours contain 68% and 95% of the probability.

all systematics. Considering posterior medians instead, larger differences occur in 158/1000 (considering both tails: 185/1000), 119/1000 (234/1000) and 256/1000 (324/1000) best fits to EZ mocks including all systematics. To combine these two measurements, since posteriors are not Gaussian but show comparable error bars, we translate them such that their two medians are located at the mean median and take the mean of the two posteriors as consensus. This method leads to an unbiased measurement if both measurements are unbiased, and is conservative about the final error bars. We show the two posteriors and their combination in Figure 5.39. The consensus RSD + BAO eBOSS ELG measurement is thus:

$$\begin{aligned} f\sigma_8(z_{\rm eff}=0.85) &= 0.315 \pm 0.095 \\ D_{\rm H}(z_{\rm eff}=0.85)/r_{\rm drag} &= 19.6^{+2.2}_{-2.1} \\ D_{\rm M}(z_{\rm eff}=0.85)/r_{\rm drag} &= 19.5 \pm 1.0 \end{aligned} \tag{5.64}$$

These measurements are 1.4σ, 0.5σ and 0.9σ from the Planck Collaboration et al. (2018) CMB-based (TT, TE, EE, lowE, lensing) predictions for $f\sigma_8$, $D_{\rm H}/r_{\rm drag}$ and $D_{\rm M}/r_{\rm drag}$, respectively.

We interpolate the consensus MCMC posterior (including systematic error bars) on a grid of $f\sigma_8$, $D_{\rm H}/r_{\rm drag}$ and $D_{\rm M}/r_{\rm drag}$ which is used, together with the BAO likelihood profile (including systematic error bars), for the combined eBOSS cosmological constraints presented in eBOSS Collaboration et al. (2020).

5.5.4 Conclusions

The eBOSS ELG sample is the largest and highest redshift emission line galaxy spectroscopic sample ever assembled, the third one to used for cosmology (Blake et al., 2011; Contreras et al., 2013; Okumura et al., 2016) and the first one targeted in SDSS. Together with the configuration space analysis of Tamone et al. (2020), we provided measurements of the growth rate of structure and cosmological distances at a redshift of 0.85. This analysis was complicated by various observational and analysis artefacts, which fostered reconsideration of previous practices in clustering analyses, as we detailed in the previous chapter. First, the fine-grained veto masks applied to the eBOSS ELG data led us to revise the way window functions are normalised in the model, as detailed in Section 4.2.2.3. We also noticed that measuring the expected redshift distribution from observed data itself (using the *shuffled* scheme) led to a significant bias of cosmological parameters, which we corrected by modelling the induced radial integral constraint as explained in Section 4.3.2. Finally, residual angular systematics due to the inhomogeneous photometry, which could not be treated following the standard technique of template regression, were mitigated using the *pixelated* scheme as discussed in Section 4.3.7.

Some observational systematics may remain, due to our incomplete understanding of the relation between the ELG target density and the imaging properties in the early DECaLS release used in the eBOSS ELG target selection.

Let us note however that this work helped improve other eBOSS clustering analyses (using QSOs and LRGs as tracers of the matter density field), in particular through the modelling of radial integral constraints and the implementation of realistic systematics into the mocks. The present analysis should also prove helpful for the next generation of spectroscopic surveys, DESI and Euclid (Laureijs et al., 2011), which will both measure 20 million ELG spectra.

5.6 SDSS cosmological results

eBOSS is the last survey of the SDSS designed for cosmological clustering analyses. SDSS redshift surveys belong to Stage-III experiments, as defined by the Dark Energy Task Force (DEFT, Albrecht et al., 2006). The DEFT Figure-of-Merit (FoM) is taken as the inverse square root of the determinant of the measured w_0, w_a covariance matrix in the CPL parametrisation (1.54) of the dark energy equation of state, letting curvature density $\Omega_{K,0}$ free. This FoM was 11 for the Stage-II experiments presented in Sullivan et al. (2011). Stage-III experiments increase this FoM by a typical factor of 3, and Stage-IV experiments by a factor of 10. DESI (DESI Collaboration et al., 2016) will be the first Stage-IV experiment as it will reach a FoM of 704, including Planck data, in the best scenario (see Section 6.3.2). Its first cosmological constraints are expected in 2023.

Hence, cosmology with galaxy redshift surveys has reached a turning point in terms of constraining power, calling for a timely review of the role of large scale structure in cosmological measurements, as presented in eBOSS Collaboration et al. (2020).

5.6.1 Data sets

For this review, highlighting SDSS results, 8 SDSS samples are considered. For all galaxy and quasar clustering samples, BAO and large scale RSD measurements are used, yielding constraints on $f\sigma_8$, $D_{\rm M}/r_{\rm drag}$ and $D_{\rm H}/r_{\rm drag}$ (or their combination $D_{\rm V}/r_{\rm drag}$). For Lyman-α samples, only

BAO measurements ($D_\mathrm{H}/r_\mathrm{drag}$, $D_\mathrm{M}/r_\mathrm{drag}$) are considered. These measurements are shown in Figure 5.40 and briefly recapped hereafter:

1. the Main Galaxy Sample (MGS, $0.07 < z < 0.2$), collected in SDSS-I and -II, contains 63 163 galaxy redshifts. BAO ($D_\mathrm{V}/r_\mathrm{drag}$) and RSD measurements are provided in Ross et al. (2015); Howlett et al. (2015).

2. BOSS DR12 galaxies ($0.2 < z < 0.6$), 1.2 M redshifts collected by BOSS during SDSS-III. Originally, three overlapping redshift bins were considered: $0.2 < z < 0.5$, $0.4 < z < 0.6$ and $0.5 < z < 0.75$. The high redshift objects $z > 0.6$ are combined with the following eBOSS LRG sample, so that the last redshift bin from BOSS $0.5 < z < 0.75$ is dropped. Bins $0.2 < z < 0.5$ and $0.4 < z < 0.6$ contain 604 001 and 686 370 galaxies, respectively. BAO ($D_\mathrm{H}/r_\mathrm{drag}$, $D_\mathrm{M}/r_\mathrm{drag}$) and RSD measurements are provided in Alam et al. (2017).

3. eBOSS LRGs ($0.6 < z < 1.0$), targeted by eBOSS, which obtained 174 816 redshifts. These are supplemented by the high redshift $z > 0.6$ BOSS galaxies, increasing the sample size to 377 458 galaxies. BAO ($D_\mathrm{H}/r_\mathrm{drag}$, $D_\mathrm{M}/r_\mathrm{drag}$) and RSD measurements are provided in Gil-Marín et al. (2020); Bautista et al. (2020).

4. eBOSS ELGs ($0.6 < z < 1.1$), fully targeted by eBOSS, consisting of 173 736 galaxies. BAO ($D_\mathrm{V}/r_\mathrm{drag}$) and RSD ($D_\mathrm{H}/r_\mathrm{drag}$, $D_\mathrm{M}/r_\mathrm{drag}$, $f\sigma_8$) measurements are provided in de Mattia et al. (2020); Tamone et al. (2020).

5. eBOSS QSOs ($0.8 < z < 2.2$), mainly targeted by eBOSS, but containing 18% of objects from SDSS-I, -II and -III, leading to a sample size of 343 708 quasars. BAO ($D_\mathrm{H}/r_\mathrm{drag}$, $D_\mathrm{M}/r_\mathrm{drag}$) and RSD measurements are provided in Neveux et al. (2020); Hou et al. (2020).

6. Lyman-α forests ($1.8 < z < 3.5$) found in quasar spectra, probing variations of the absorption due to neutral hydrogen. The BOSS sample of 157 845 spectra ($2 < z < 3.5$) without broad absorption lines has been complemented and partly reobserved by eBOSS, yielding a total sample of 210 005 spectra ($2.1 < z < 3.5$). Both an auto-correlation and a cross-correlation with 341 468 quasars ($z > 1.77$) were performed. BAO ($D_\mathrm{H}/r_\mathrm{drag}$, $D_\mathrm{M}/r_\mathrm{drag}$) measurements are provided in du Mas des Bourboux et al. (2020).

External data sets are CMB measurements (TT, TE, EE, lowE and lensing) (Planck Collaboration et al., 2019a, 2018b), Pantheon supernovae (SN) Ia (Scolnic et al., 2018) and DES Year 1 3×2pt (cosmic shear, galaxy clustering and galaxy - galaxy lensing) correlation function measurements (Abbott et al., 2018).

We will see in the next sections that the probes above are complementary: BAO and SN data are powerful to constrain respectively curvature and the dark energy equation of state. Growth rate measurements can be used to test deviations to general relativity. We will thus highlight extensions to the baseline flat ΛCDM model. Specifically, will be considered hereafter an open ΛCDM model (oΛCDM), i.e. with free curvature, a CDM model with free dark energy (DE) equation of state (EoS) (wCDM), an open CDM model with free DE EoS (owCDM), a CDM model with DE EoS given by the CPL parametrisation (w_0w_aCDM) and an open CDM model with DE EoS given by the CPL parametrisation (ow_0w_aCDM).

For all figures in this section, σ_8 and all density parameters are taken at $z = 0$.

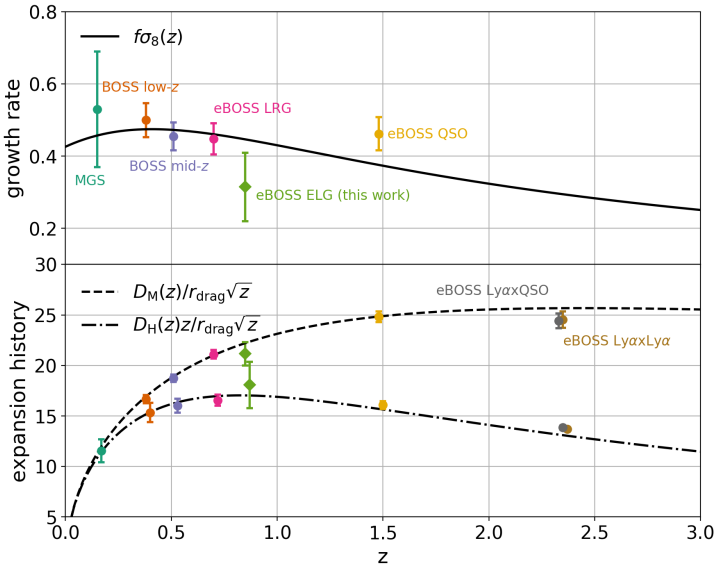

Figure 5.40 – SDSS growth rate and BAO measurements used in the final cosmological inference. For the MGS sample ($z_{\rm eff} = 0.15$) the $D_{\rm V}/r_{\rm drag}$ measurement is plotted as a $D_{\rm M}/r_{\rm drag}$ one. Continuous lines are ΛCDM predictions with Planck Collaboration et al. (2018) CMB (TT, TE, EE, lowE, lensing) parameters.

5.6.2 BAO constraints

In this section, we restrict CMB data to temperature and polarisation (CMB T&P, i.e. TT, TE, EE, lowE of Planck Collaboration et al. 2018), dropping lensing measurements, which probe the growth of structure.

Within the oΛCDM, BAO measurements alone lead to $\Omega_{\Lambda,0} = 0.637^{+0.084}_{-0.074}$, a 8σ confidence detection of dark energy. Such a constraint is achieved through the combination of measurements at different redshifts, hence breaking degeneracies between $\Omega_{m,0}$ and $\Omega_{K,0}$. The corresponding contours are shown in Figure 5.41 (top left). SN data alone are not as powerful and provide $\Omega_{\Lambda,0} = 0.73 \pm 0.11$. Both BAO and SN data are compatible with a flat Universe. Combining these data sets with Planck CMB T&P (which favours a closed Universe) pulls $\Omega_{K,0}$ towards 0, as shown in Figure 5.41 (top right). Especially, CMB T&P and BAO alone yield $\Omega_{K,0} = -0.0001 \pm 0.0018$.

Allowing for a free DE EoS (wCDM model), BAO and SN data favour $w = -1$, i.e a cosmological constant, as can be seen in Figure 5.41 (bottom panel). In particular, BAO data alone yield $w = 0.69 \pm 0.15$, compatible with a cosmological constant ($w = -1$). Also, the SN $\Omega_{m,0} - w_0$ contours are orthogonal to that of the CMB T&P (which favours $w < -1$), yielding a 3.3% constraint on w, compatible with -1.

In addition, BAO and SN data can be used to set constraints on H_0 through the *inverse distance ladder*. Specifically, given a value of $r_{\rm drag}$, BAO provide a measurement of cosmological distances than can be used to calibrate SN luminosity distances and yield a present-day value H_0, in a given cosmological model. Under the assumptions of standard pre-recombination physics and a well-measured mean temperature of the CMB, $r_{\rm drag}$ only depends on $\omega_{c,0}$ and $\omega_{b,0}$. Taking for these two parameters priors from CMB T&P, BAO and SN data yield percent constraints on H_0. As can be seen in Figure 5.42 (left), H_0 measurements are stable with respect to the model expansion history: ΛCDM, owCDM or ow_0w_aCDM, in the latter case providing a precise

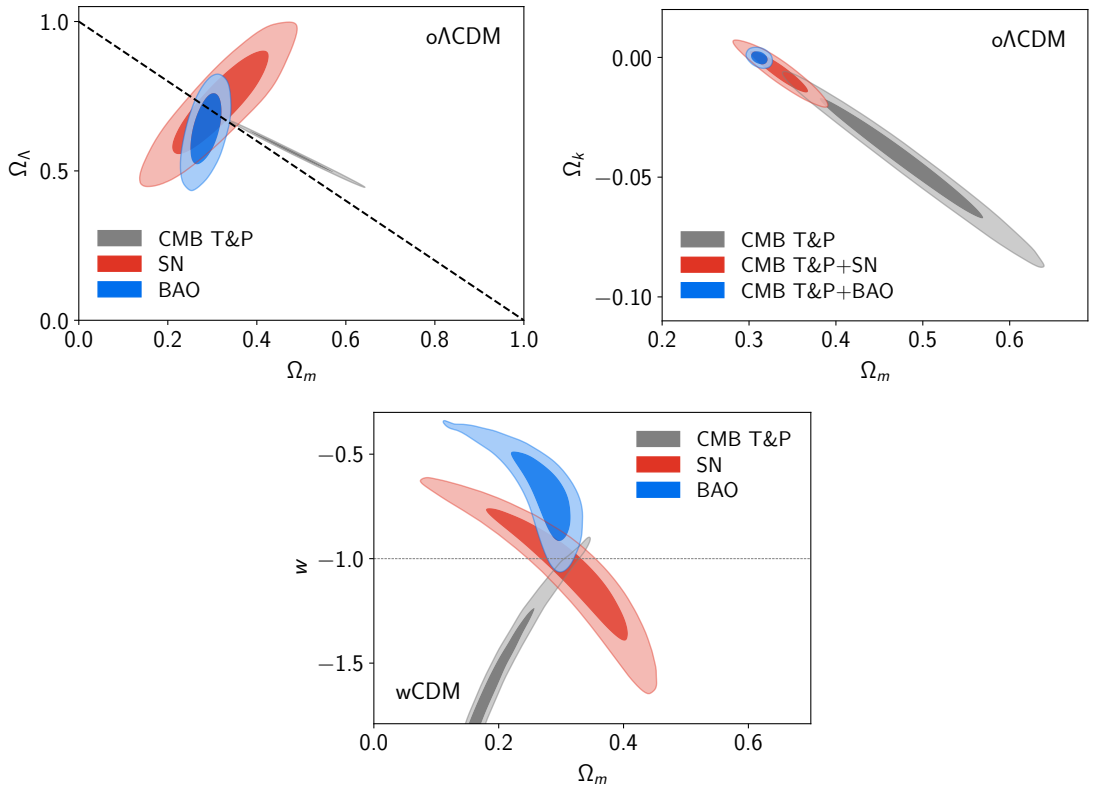

Figure 5.41 – Top left: $\Omega_{m,0} - \Omega_{\Lambda,0}$ 68% and 95% contours, in the oΛCDM model. Top right: same, in the $\Omega_{m,0} - \Omega_{K,0}$ plane. Bottom: $\Omega_{m,0} - w$ 68% and 95% contours, within the wCDM model. Dashed lines indicate the parameters values within the ΛCDM model. Taken from eBOSS Collaboration et al. (2020).

measurement of $H_0 = 67.87 \pm 0.86\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, which would not be possible with CMB data alone.

To make sure the measured value of H_0 is insensitive to potential systematics in CMB anisotropies, $\omega_{b,0}$ entering r_{drag} can be constrained with BBN measurements. The radiation density is set by COBE/FIRAS (Fixsen, 2009), and $\omega_{c,0}$ is constrained by BAO alone. Figure 5.42 (right) shows that the constraining power of BAO is due to high redshift measurements (QSO and Lyman-α) having a different degeneracy in the $\Omega_{m,0} - H_0$ plane than low redshift ones. Using the BBN prior and in the ΛCDM model, the Hubble parameter is found to be $H_0 = 67.35 \pm 0.97\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, in excellent agreement with the value quoted above.

This value of H_0 is in tension with late-time measurements, making no assumption on the physics in the early Universe. Using measurements from SH0ES (Riess et al., 2018) (calibrating SN luminosity distance with Cepheids) and H0LiCOW measurements (based on time delays of lensed quasars), Wong et al. (2019) find $H_0 = 73.8 \pm 1.1\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (as mentioned in Section 1.1.4), in $\simeq 5\sigma$ tension with the above measurements.

5.6.3 RSD constraints

RSD measurements are particularly powerful to constrain the dark energy equation of state. Indeed, in a wCDM model, a low value of w increasingly reduces $f\sigma_8$ towards lower redshifts. Hence, RSD measurements reduce uncertainty on w by more than a factor of two over CMB data, as can be seen in Figure 5.43, yielding $w = -1.09 \pm 0.11$.

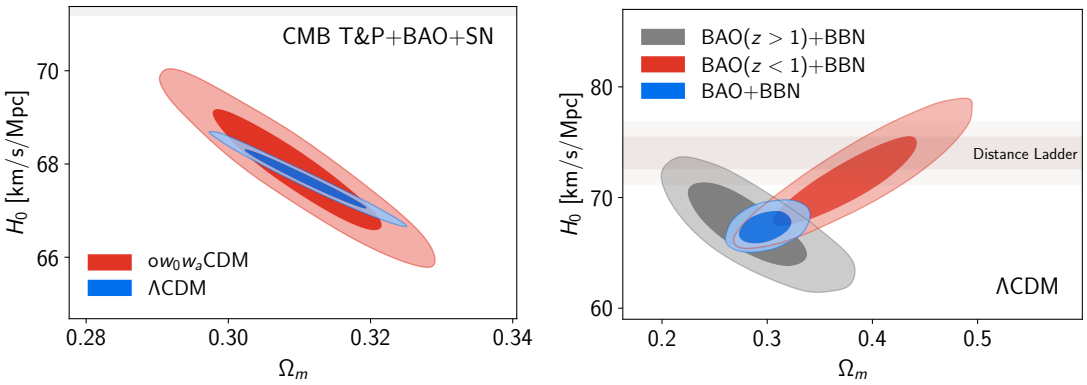

Figure 5.42 – Left: $\Omega_{m,0} - H_0$ 68% and 95% contours, in the ΛCDM and ow_0w_aCDM models, BAO and SN constraints using Planck priors for $\omega_{b,0}$ and $\omega_{c,0}$. Right: same parameter space, within the ΛCDM model, low and high-z BAO constraints using a BBN prior for $\omega_{b,0}$. The red shaded bands correspond to the 68% and 95% intervals of the Riess et al. (2019) measurement $H_0 = 74.03 \pm 1.42\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Taken from eBOSS Collaboration et al. (2020).

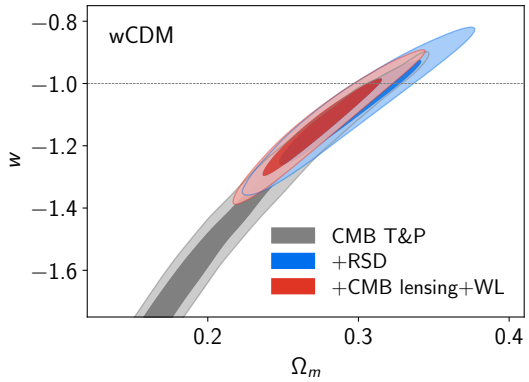

Figure 5.43 – Left: $\Omega_{m,0} - w$ 68% and 95% contours, within the wCDM model. The dashed line indicates the value of w within the ΛCDM model. Taken from eBOSS Collaboration et al. (2020).

However, the current RSD measurements do not help constrain the oCDM model, since the effect of curvature on the growth of structure is mainly visible at low z (covered by the MGS sample), where $f\sigma_8$ error bars are large. DES data instead help reduce the uncertainty on $\Omega_{K,0}$ by a factor of 3.4.

RSD and weak lensing measurements alone provide constraints in the $\sigma_{8,0} - \Omega_{m,0}$ plane, as shown in Figure 5.44 (left). The late time structure growth (probed by RSD and galaxy and CMB lensing) is consistent (albeit with large contours) with that inferred from the early Universe (CMB T&P) assuming a ΛCDM model, providing a strong consistency test of this cosmological model.

Finally, growth measurements can be used to put constraints on modification to general relativity (GR). Let us introduce the parameters $\Sigma(a), \mu(a)$, modifying the Poisson equation (2.5) through (Abbott et al., 2019):

$$k^2\Phi = -4\pi G_\mathrm{N} a^2 (1+\mu)\,\bar{\rho}\delta \tag{5.65}$$

$$k^2 (\Phi + \Psi) = -8\pi G_\mathrm{N} a^2 (1+\Sigma)\,\bar{\rho}\delta \tag{5.66}$$

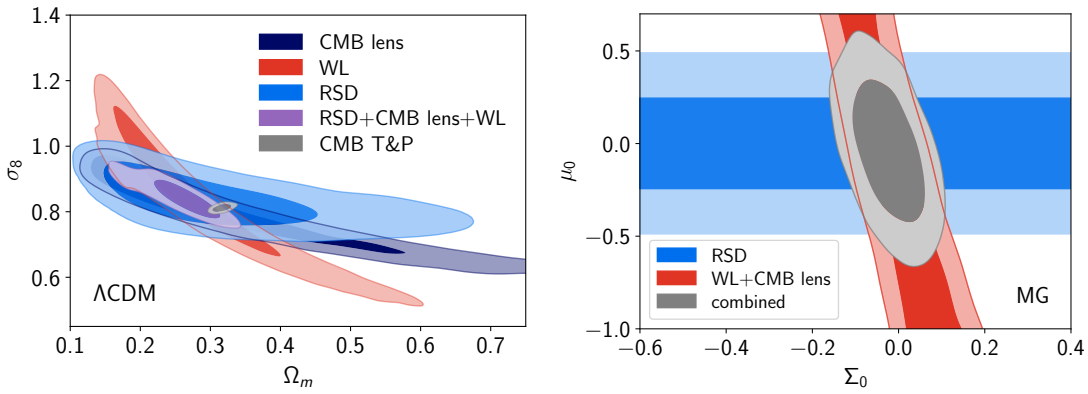

Figure 5.44 – Left: $\Omega_{m,0} - \sigma_{8,0}$ 68% and 95% contours, within the ΛCDM model. Right: 68% and 95% contours in the $\Sigma_0 - \mu_0$ extension to the ΛCDM model. Taken from eBOSS Collaboration et al. (2020).

where we recall Φ and Ψ are the Bardeen potentials related to the time and spatial part of the metric (see Eq. (1.109)). Following Abbott et al. (2019), the time dependence of $\mu(a)$ and $\Sigma(a)$ is parametrised through:

$$\mu(a) = \mu_0 \frac{\Omega_\Lambda(a)}{\Omega_{\Lambda,0}} \qquad \Sigma(a) = \Sigma_0 \frac{\Omega_\Lambda(a)}{\Omega_{\Lambda,0}}. \tag{5.67}$$

In GR, $\Psi = \Phi$ and hence $\mu_0 = \Sigma_0 = 0$. RSD (in the non-relativistic limit) is only sensitive to Φ and thus can only probe μ_0. Weak lensing measurements are also sensitive to the propagation of light dictated by $\Phi + \Psi$ and hence to Σ_0. As can be see from Figure 5.44 (right), combining RSD and weak lensing measurements (DES and Planck lensing) yields $\mu_0 = -0.04 \pm 0.25$ and $\Sigma_0 = -0.024 \pm 0.054$, consistent with the GR prediction of $\mu_0 = \Sigma_0 = 0$.

5.6.4 Full fits

Combining all probes with SDSS BAO and RSD measurements, a fit is performed assuming a (flat) ΛCDM model. All SDSS data are found to be consistent with the best fit model. The overall compatibility of SDSS, DES, SN and Planck data can be checked in Figure 5.45. For both SDSS and DES contours, priors on $\omega_{b,0}$ and n_s are those of Planck Collaboration et al. (2018b), namely $n_s = 0.96 \pm 0.02$ and $\omega_{b,0} = 0.0222 \pm 0.0005$, motivated by BBN. This latter prior has little effect on DES contours but helps constrain $r_{\rm drag}$ for SDSS results. DES data prefer lower $\Omega_{m,0}$ values than Planck Collaboration et al. (2018), translating in some differences in $\sigma_{8,0}$ and H_0. Abbott et al. (2018) argue there is no evidence for tension with CMB data. SDSS and SN data are in good agreement with either DES or Planck measurements.

Results are provided in Table 5.17 for the various cosmological models considered. In the ΛCDM case, $\Omega_{\Lambda,0}$ is measured at the 0.7% level, with a gain of 1.78 over CMB data (excluding lensing) alone, mainly driven by BAO and SN.

Cosmological parameters remain stable, consistent with the ΛCDM case, whatever model is assumed. Error bars on dark energy density and the Hubble parameter only increase by a factor of 1.5 and 2.3, respectively, when allowing the most flexible model for expansion history ow_0w_aCDM. This is made possible by the complementarity of different probes, as shown in Figure 5.46. CMB + SN and CMB + BAO + RSD $\Omega_{K,0} - w_0$ contours are orthogonal; as seen in Section 5.6.2, while SN are powerful to constrain the DE EoS, BAO data provide stringent constraints on curvature. In the $w_0 - w_a$ plane, CMB + SN achieve slightly better constraints

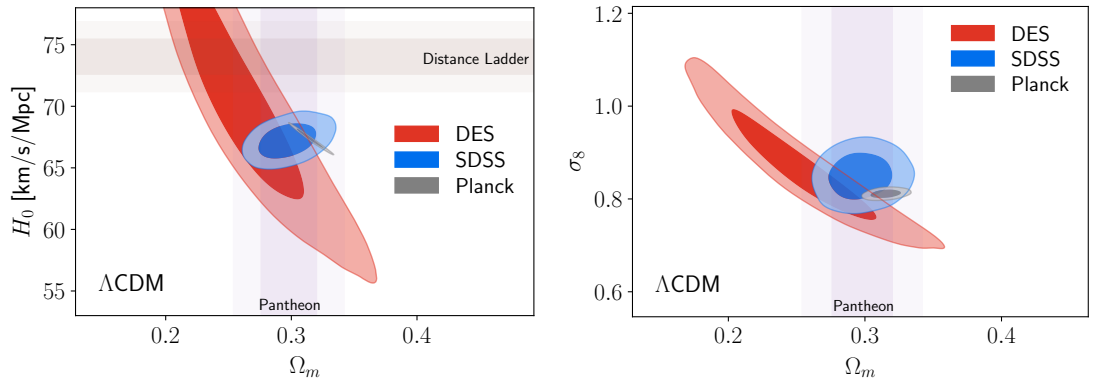

Figure 5.45 – Left: $\Omega_{m,0} - H_0$ 68% and 95% contours, within the ΛCDM model. Right: same, in the $\Omega_{m,0} - \sigma_{8,0}$ plane. Purple and red shaded bands correspond to the 68% and 95% intervals of the Pantheon sample measurement ($\Omega_{m,0} = 0.298 \pm 0.022$) and Riess et al. (2019) measurement $H_0 = 74.03 \pm 1.42\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, respectively. Taken from eBOSS Collaboration et al. (2020).

model	$\Omega_{\mathrm{DE},0}$	H_0	$\sigma_{8,0}$	$\Omega_{K,0}$	w_0	w_a
ΛCDM	0.6959(47)	68.19(36)	0.8073(56)	–	–	–
oΛCDM	0.6958(48)	68.21(55)	0.8076(65)	0.0001(17)	–	–
wCDM	0.6992(66)	68.64(73)	0.8128(92)	–	−1.020(27)	–
owCDM	0.6997(69)	68.59(73)	0.8127(91)	0.0004(19)	−1.023(30)	–
w_0w_aCDM	0.6971(69)	68.47(74)	0.8139(93)	–	−0.939(73)	$-0.31^{+0.28}_{-0.24}$
ow_0w_aCDM	0.6988(72)	68.20(81)	0.8140(93)	−0.0023(22)	−0.912(81)	$-0.48^{+0.36}_{-0.30}$

Table 5.17 – Marginalised values and in parentheses the 68% confidence limits for cosmological models using Planck, Pantheon SN, SDSS, and DES data. $\Omega_{\mathrm{DE},0}$ is the dark energy density today and corresponds to $\Omega_{\Lambda,0}$ in the case of ΛCDM-based models. Extracted from eBOSS Collaboration et al. (2020).

than CMB + BAO + RSD, but precision on w_0 and w_a increases by roughly a factor of 2 when all data sets are combined, yielding measurements consistent with a flat ΛCDM model: w_0, w_a and $\Omega_{K,0}$ are 1.1σ, 1.3σ and 1.0σ from -1, 0 and 0, respectively.

Shifting the pivot scale factor $a = 1$ of the CPL parametrisation (1.54) to a_p, we can write $w(a) = w_p + w_a(a_p - a)$. Given the full data set, in the w_0w_aCDM extension, the $w_p - w_a$ covariance matrix can be made diagonal at the pivot redshift $z_p = 1/a_p - 1 = 0.34$, yielding $w_p = -1.018 \pm 0.028$ and $w_a = 0.31^{+0.28}_{-0.24}$. Errors on w_p are only 4% larger than on w in the wCDM model: the cost of allowing another degree of freedom in the DE EoS is minimal.

Combining Planck and SDSS data, the DEFT FoM in the $w_0 - w_a$ plane, assuming no curvature, is 38.4. Adding DES data and the Pantheon sample increases the FoM by another factor of 3.5, yielding a final FoM of 134. This value only degrades to 92 in the ow_0w_aCDM model.

Cosmological data sets can also be used to measure neutrino masses. The sum of neutrino masses (previously fixed to $\sum m_\nu = 0.06\,\mathrm{eV}$) is left free on top of the flat ΛCDM model (νΛCDM) or wCDM model ($m_\nu w$CDM). Constraints on the sum of neutrino masses are provided in Table 5.18. In all cases, one massive and two massless neutrinos are assumed — a good approximation for masses of interest — and a prior on the sum of neutrino masses $\sum m_\nu \geq 0$ is imposed, thus ignoring the non-zero mass bounds evidenced by neutrino oscillations (see Section 1.2.3.2). Combining all data sets, at the 95% confidence level, $\sum m_\nu < 0.111\,\mathrm{eV}$ in the νΛCDM model, which degrades to $\sum m_\nu < 0.161\,\mathrm{eV}$ when allowing for a free DE EoS ($m_\nu w$CDM model). As

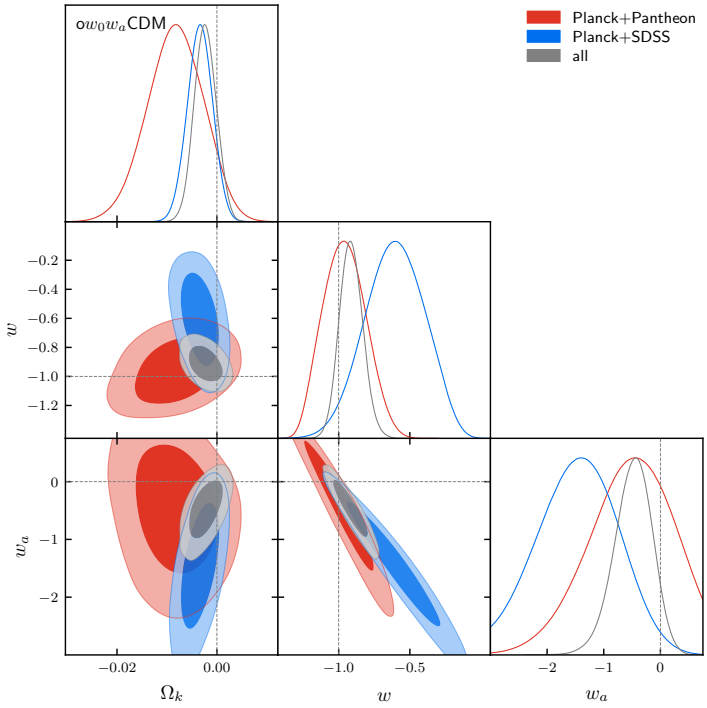

Figure 5.46 – Posteriors in the $\Omega_{K,0}$, w_0 and w_a space, assuming the ow_0w_aCDM model. Dashed lines indicate the parameter values within the ΛCDM model. Contours contain 68% and 95% of the probability. Taken from eBOSS Collaboration et al. (2020).

can be seen in Figure 5.47, the preferred value of $\sum m_\nu$ is negative. Hence, Gaussian fits are performed to provide insight into the constraining power of the different probes, independently of the preferred $\sum m_\nu$ value. Improvement on top of Planck constraints is mainly driven by BAO data, through the impact of neutrino masses onto the expansion rate: a factor 2 gain is achieved. RSD yield another 20% reduction in uncertainty, through the damping effect of neutrino free-streaming on the growth of structure. The constraining power of RSD is expected to overcome that of BAO with Stage-IV experiments, such as DESI. Adding DES data slightly degrades the constraint on $\sum m_\nu$ from 0.099 eV to 0.111 eV (in the $\nu\Lambda$CDM model), due to the preferred lower $\Omega_{m,0}$ value. Also provided in Table 5.18 is the probability $P_{\rm unphy}$ that the sum neutrino masses lies in the unphysical region $0 < \sum m_\nu < 0.0588\,\text{eV}$, where 0.0588 eV is the minimum total neutrino mass (in the normal hierarchy) according to the observation of neutrino oscillations. The obtained $P_{\rm unphy}$ values show no strong evidence for an unphysical sum of neutrino masses. Also reported is the ratio $P_{\rm inv}/P_{\rm norm}$, with:

$$P_{\rm norm} = \int_{0.0588\,\text{eV}}^{\infty} p(m_\nu) dm_\nu \tag{5.68}$$

$$P_{\rm inv} = \int_{0.0995\,\text{eV}}^{\infty} p(m_\nu) dm_\nu, \tag{5.69}$$

0.0995 eV being the minimum sum of neutrino masses in the inverted hierarchy. A high ratio $P_{\rm inv}/P_{\rm norm}$ would indicate a strong preference for the inverted over the normal hierarchy, which is not the case with current data sets.

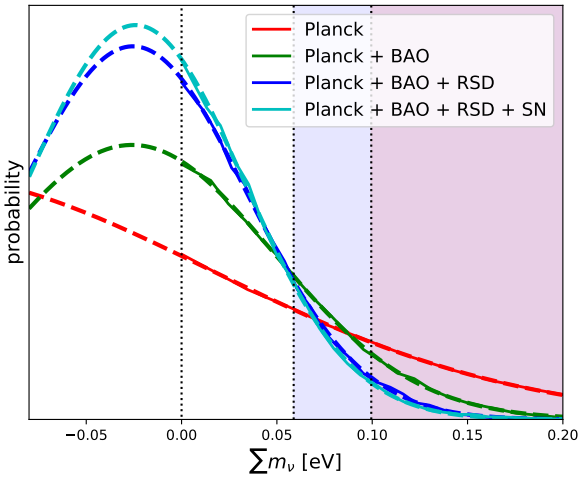

Figure 5.47 – Posteriors of the sum of neutrino masses, within the $\nu\Lambda$CDM model. Posteriors are normalised such that their integral for $\sum m_\nu > 0\,\mathrm{eV}$ is unity. Dashed lines are Gaussian fits. Purple and red regions are low limits on $\sum m_\nu$ in the normal and inverted hierachies, respectively. Taken from eBOSS Collaboration et al. (2020).

model	95% upper limit [eV]	$P_{\mathrm{inv}}/P_{\mathrm{norm}}$	P_{unphy}	Gaussian fit to posterior [eV]
Planck	0.252	0.64	0.43	
Planck + BAO	0.129	0.36	0.64	-0.026 ± 0.074
Planck + BAO + RSD	0.102	0.24	0.76	-0.026 ± 0.060
Planck + SN	0.170	0.49	0.56	-0.076 ± 0.106
Planck + BAO + RSD + SN	0.099	0.22	0.78	-0.024 ± 0.057
Planck + BAO + RSD + SN + DES	0.111	0.27	0.71	-0.014 ± 0.061
Planck + BAO + RSD + SN ($m_\nu w$CDM)	0.139	0.40	0.61	-0.033 ± 0.082
Planck + BAO + RSD + SN + DES ($m_\nu w$CDM)	0.161	0.48	0.56	-0.048 ± 0.097

Table 5.18 – Constraints on neutrino masses and relative probabilities of neutrino models. The $\nu\Lambda$CDM model is assumed as a baseline. Extracted from eBOSS Collaboration et al. (2020).

5.6.5 Conclusions

Of interest is to compare the currently obtained Stage-III cosmological constraints to those prior to the BOSS results, called Stage-II, as illustrated in Figure 5.48 with an open $m_\nu w$CDM model ($m_\nu o w$CDM). For Stage-II, the CMB data set consists in WMAP (Hinshaw et al., 2013; Bennett et al., 2013), SN come from the JLA sample (Betoule et al., 2014) and BAO measurements from SDSS DR7 (Percival et al., 2010) and the 2-degree Field Galaxy Redshift Survey (Colless et al., 2001). For Stage-II + SDSS, the aforementioned BAO measurements are replaced by the final SDSS BAO + RSD data. Conversely, the Stage-III without SDSS case is obtained by replacing final SDSS BAO + RSD data by the previous BAO measurements.

Overall, Stage-III measurements are fully compatible with Stage-II results, the largest deviation being for $\sigma_{8,0}$. Stage-III measurements confirm consistency with a flat ΛCDM model, as can be seen with $\Omega_{K,0}$ and w_0 moving towards 0 and -1, respectively. Also, bounds on the sum of neutrino masses have become significantly tighter, without sign for a non-minimal sum of neutrino masses.

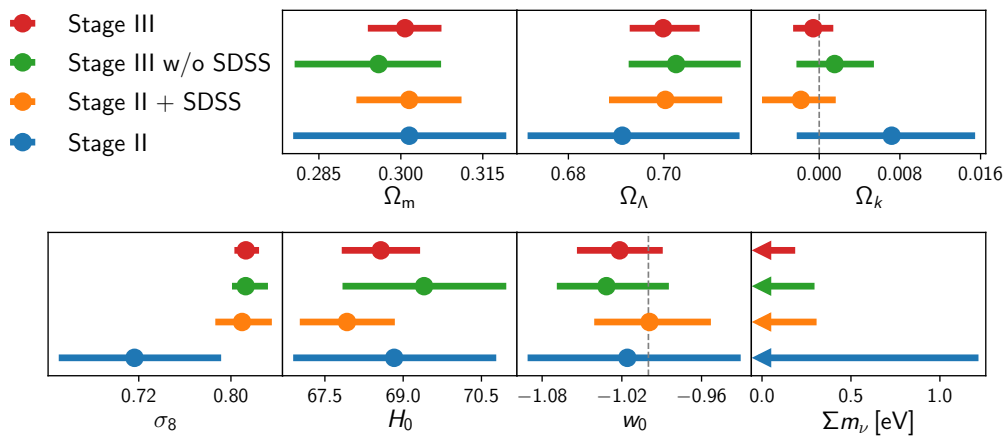

Figure 5.48 – Uncertainties in parameters of the $m_\nu o w$CDM model with Stage-II and Stage-III experiments. Limits on the sum of neutrino masses are given at the 95% confidence level. See text for details on the different data sets involved. Taken from eBOSS Collaboration et al. (2020).

The precision on all shown parameters has increased by at least a factor of 2.5, from Stage-II to Stage-III. The largest reduction in uncertainties are for $\Omega_{K,0}$ (by 4.5), $\sigma_{8,0}$ (7.0), and $\sum m_\nu$ (7.1). The FoM[14] in the w, $\Omega_{K,0}$, $\sum m_\nu$, H_0, and $\sigma_{8,0}$ space has increased from 11 to 44 between Stage-II and Stage-III. Adding only final SDSS to Stage-II leads already to a remarkable improvement by a factor of 2.1, demonstrating the significant role of BAO and RSD measurements in constraining the cosmological model, especially its parameters $\Omega_{K,0}$, H_0, and $\sum m_\nu$.

The role played by cosmological measurements with large scale structure will become increasingly important with next generation, Stage-IV galaxy surveys, such as DESI.

14. defined as the determinant of the covariance matrix of the $N = 5$ parameters to the power $-1/(2N)$

Bibliography

Abbott, T. M. C., F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, et al. (2018). Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. Phys. Rev. D *98*(4), 043526. arXiv:1708.01530.

Abbott, T. M. C., F. B. Abdalla, S. Avila, M. Banerji, E. Baxter, et al. (2019). Dark Energy Survey year 1 results: Constraints on extended cosmological models from galaxy clustering and weak lensing. Phys. Rev. D *99*(12), 123505. arXiv:1810.02499.

Alam, S., M. Ata, S. Bailey, F. Beutler, D. Bizyaev, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. MNRAS *470*(3), 2617–2652. arXiv:1607.03155.

Alam, S., A. de Mattia, A. Tamone, S. Ávila, J. A. Peacock, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: N-body Mock Challenge for the eBOSS Emission Line Galaxy Sample. *arXiv e-prints*, arXiv:2007.09004. arXiv:2007.09004.

Alam, S., J. A. Peacock, K. Kraljic, A. J. Ross, and J. Comparat (2020). Multitracer extension of the halo model: probing quenching and conformity in eBOSS. MNRAS *497*(1), 581–595. arXiv:1910.05095.

Albrecht, A., G. Bernstein, R. Cahn, W. L. Freedman, J. Hewitt, et al. (2006). Report of the Dark Energy Task Force. *arXiv e-prints*, astro–ph/0609591. arXiv:astro-ph/0609591.

Ata, M., F. Baumgarten, J. Bautista, F. Beutler, D. Bizyaev, et al. (2018). The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2. MNRAS *473*(4), 4773–4794. arXiv:1705.06373.

Avila, S., M. Crocce, A. J. Ross, J. García-Bellido, W. J. Percival, et al. (2018). Dark Energy Survey Year-1 results: galaxy mock catalogues for BAO. MNRAS *479*(1), 94–110. arXiv:1712.06232.

Avila, S., V. Gonzalez-Perez, F. G. Mohammad, A. de Mattia, C. Zhao, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: exploring the Halo Occupation Distribution model of Emission Line Galaxies. *arXiv e-prints*, arXiv:2007.09012. arXiv:2007.09012.

Bautista, J. E., R. Paviot, M. Vargas Magaña, S. de la Torre, S. Fromenteau, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1. *arXiv e-prints*, arXiv:2007.08993. arXiv:2007.08993.

Bautista, J. E., M. Vargas-Magaña, K. S. Dawson, W. J. Percival, J. Brinkmann, et al. (2018). The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations at Redshift of 0.72 with the DR14 Luminous Red Galaxy Sample. ApJ *863*(1), 110. arXiv:1712.08064.

Bennett, C. L., D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. ApJS *208*(2), 20. arXiv:1212.5225.

Betoule, M., R. Kessler, J. Guy, J. Mosher, D. Hardin, et al. (2014). Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. A&A *568*, A22. arXiv:1401.4064.

Beutler, F., H.-J. Seo, A. J. Ross, P. McDonald, S. Saito, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Fourier space. MNRAS *464*(3), 3409–3430. arXiv:1607.03149.

Beutler, F., H.-J. Seo, S. Saito, C.-H. Chuang, A. J. Cuesta, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: anisotropic galaxy clustering in Fourier space. MNRAS *466*, 2242–2260. arXiv:1607.03150.

Bianchi, D. and W. J. Percival (2017). Unbiased clustering estimation in the presence of missing observations. MNRAS *472*(1), 1106–1118. arXiv:1703.02070.

Bianchi, D. and L. Verde (2019). Confronting missing observations with probability weights: Fourier space and generalised formalism. *arXiv e-prints*, arXiv:1912.08803. arXiv:1912.08803.

Blake, C., S. Brough, M. Colless, C. Contreras, W. Couch, et al. (2011). The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift z=0.9. MNRAS *415*(3), 2876–2891. arXiv:1104.2948.

Bryan, G. L. and M. L. Norman (1998). Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons. ApJ *495*(1), 80–99. arXiv:astro-ph/9710107.

Burden, A., W. J. Percival, and C. Howlett (2015). Reconstruction in Fourier space. MNRAS *453*(1), 456–468. arXiv:1504.02591.

Carlson, J. and M. White (2010). Embedding Realistic Surveys in Simulations Through Volume Remapping. ApJS *190*(2), 311–314. arXiv:1003.3178.

Carter, P., F. Beutler, W. J. Percival, J. DeRose, R. H. Wechsler, et al. (2020). The Impact of the Fiducial Cosmology Assumption on BAO Distance Scale Measurements. MNRAS. arXiv:1906.03035.

Child, H. L., S. Habib, K. Heitmann, N. Frontiere, H. Finkel, et al. (2018). Halo Profiles and the Concentration-Mass Relation for a ΛCDM Universe. ApJ *859*(1), 55. arXiv:1804.10199.

Chuang, C.-H., F.-S. Kitaura, F. Prada, C. Zhao, and G. Yepes (2015). EZmocks: extending the Zel'dovich approximation to generate mock galaxy catalogues with accurate clustering statistics. MNRAS *446*(3), 2621–2628. arXiv:1409.1124.

Chuang, C.-H., C. Zhao, F. Prada, E. Munari, S. Avila, et al. (2015). nIFTy cosmology: Galaxy/halo mock catalogue comparison project on clustering statistics. MNRAS *452*(1), 686–700. arXiv:1412.7729.

Colless, M., G. Dalton, S. Maddox, W. Sutherland , P. Norberg, et al. (2001). The 2dF Galaxy Redshift Survey: spectra and redshifts. MNRAS *328*(4), 1039–1063. arXiv:astro-ph/0106498.

Comparat, J., E. Jullo, J.-P. Kneib, C. Schimd, H. Shan, et al. (2013). Stochastic bias of colour-selected BAO tracers by joint clustering-weak lensing analysis. MNRAS *433*(2), 1146–1160. arXiv:1302.5655.

Contreras, C., C. Blake, G. B. Poole, F. Marin, S. Brough, et al. (2013). The WiggleZ Dark Energy Survey: measuring the cosmic growth rate with the two-point galaxy correlation function. MNRAS *430*(2), 924–933. arXiv:1302.5178.

Davis, M. and P. J. E. Peebles (1983). A survey of galaxy redshifts. V. The two-point position and velocity correlations. ApJ *267*, 465–482.

de Mattia, A., V. Ruhlmann-Kleider, A. Raichoor, A. J. Ross, A. Tamone, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the emission line galaxy sample from the anisotropic power spectrum between redshift 0.6 and 1.1. *arXiv e-prints*, arXiv:2007.09008. arXiv:2007.09008.

DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, et al. (2016). The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, arXiv:1611.00036. arXiv:1611.00036.

Ding, Z., H.-J. Seo, Z. Vlah, Y. Feng, M. Schmittfull, et al. (2018). Theoretical systematics of Future Baryon Acoustic Oscillation Surveys. MNRAS *479*(1), 1021–1054. arXiv:1708.01297.

Dodelson, S. and M. D. Schneider (2013). The effect of covariance estimator error on cosmological parameter constraints. Phys. Rev. D *88*(6), 063537. arXiv:1304.2593.

du Mas des Bourboux, H., J. Rich, A. Font-Ribera, V. de Sainte Agathe, J. Farr, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Baryon acoustic oscillations with Lyman-α forests. *arXiv e-prints*, arXiv:2007.08995. arXiv:2007.08995.

eBOSS Collaboration, S. Alam, M. Aubert, S. Avila, C. Balland, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory. *arXiv e-prints*, arXiv:2007.08991. arXiv:2007.08991.

Einasto, J. (1965). On the Construction of a Composite Model for the Galaxy and on the Determination of the System of Galactic Parameters. *Trudy Astrofizicheskogo Instituta Alma-Ata 5*, 87–100.

Eisenstein, D. J. and W. Hu (1998). Baryonic Features in the Matter Transfer Function. ApJ *496*(2), 605–614. arXiv:astro-ph/9709112.

Eisenstein, D. J., I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, et al. (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. ApJ *633*(2), 560–574. arXiv:astro-ph/0501171.

Favole, G., J. Comparat, F. Prada, G. Yepes, E. Jullo, et al. (2016). Clustering properties of g-selected galaxies at z $\sim$ 0.8. MNRAS *461*(4), 3421–3431. arXiv:1507.04356.

Feng, Y., M.-Y. Chu, U. Seljak, and P. McDonald (2016). FASTPM: a new scheme for fast simulations of dark matter and haloes. MNRAS *463*(3), 2273–2286. arXiv:1603.00476.

Fixsen, D. J. (2009). The Temperature of the Cosmic Microwave Background. ApJ *707*(2), 916–920. arXiv:0911.1955.

Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman (2013). emcee: The mcmc hammer. *PASP 125*, 306–312. 1202.3665.

Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, et al. (2018). Gaia Data Release 2. Summary of the contents and survey properties. A&A *616*, A1. arXiv:1804.09365.

Gelman, A. and D. B. Rubin (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statist. Sci. 7*, 457–472.

Gil-Marín, H., J. E. Bautista, R. Paviot, M. Vargas-Magaña, S. de la Torre, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0. *arXiv e-prints*, arXiv:2007.08994. arXiv:2007.08994.

Gil-Marín, H., W. J. Percival, A. J. Cuesta, J. R. Brownstein, C.-H. Chuang, et al. (2016). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: BAO measurement from the LOS-dependent power spectrum of DR12 BOSS galaxies. MNRAS *460*(4), 4210–4219. arXiv:1509.06373.

Gonzalez-Perez, V., J. Comparat, P. Norberg, C. M. Baugh, S. Contreras, et al. (2018). The host dark matter haloes of [O II] emitters at 0.5 < z < 1.5. MNRAS *474*(3), 4024–4038. arXiv:1708.07628.

Górski, K. M., E. Hivon, A. J. Banday, B. D. Wand elt, F. K. Hansen, et al. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ *622*(2), 759–771. arXiv:astro-ph/0409513.

Grieb, J. N., A. G. Sánchez, S. Salazar-Albornoz, and C. Dalla Vecchia (2016). Gaussian covariance matrices for anisotropic galaxy clustering measurements. MNRAS *457*(2), 1577–1592. arXiv:1509.04293.

Guo, H., X. Yang, A. Raichoor, Z. Zheng, J. Comparat, et al. (2019). Evolution of Star-forming Galaxies from z = 0.7 to 1.2 with eBOSS Emission-line Galaxies. ApJ *871*(2), 147. arXiv:1810.05318.

Guo, H., I. Zehavi, and Z. Zheng (2012). A New Method to Correct for Fiber Collisions in Galaxy Two-point Statistics. ApJ *756*(2), 127. arXiv:1111.6598.

Hahn, C., F. Beutler, M. Sinha, A. Berlind, S. Ho, et al. (2019). Likelihood non-Gaussianity in large-scale structure analyses. MNRAS *485*(2), 2956–2969. arXiv:1803.06348.

Hahn, C., R. Scoccimarro, M. R. Blanton, J. L. Tinker, and S. A. Rodríguez-Torres (2017). The Effect of Fiber Collisions on the Galaxy Power Spectrum Multipoles. MNRAS *467*(2), 1940–1956. arXiv:1609.01714.

Hamilton, A. J. S. and M. Tegmark (2004). A scheme to deal accurately and efficiently with complex angular masks in galaxy surveys. MNRAS *349*(1), 115–128. arXiv:astro-ph/0306324.

Hand, N., Y. Feng, F. Beutler, Y. Li, C. Modi, et al. (2018). nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure. AJ *156*, 160. arXiv:1712.05834.

Hartlap, J., P. Simon, and P. Schneider (2007). Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. A&A *464*(1), 399–404. arXiv:astro-ph/0608064.

Hearin, A. P., A. R. Zentner, F. C. van den Bosch, D. Campbell, and E. Tollerud (2016). Introducing decorated HODs: modelling assembly bias in the galaxy-halo connection. MNRAS *460*(3), 2552–2570. arXiv:1512.03050.

Heitmann, K., H. Finkel, A. Pope, V. Morozov, N. Frontiere, et al. (2019). The Outer Rim Simulation: A Path to Many-Core Supercomputers. *arXiv e-prints*, arXiv:1904.11970. arXiv:1904.11970.

Hinshaw, G., D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. ApJS *208*(2), 19. arXiv:1212.5226.

Hou, J., A. G. Sánchez, A. J. Ross, A. Smith, R. Neveux, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from anisotropic clustering analysis of the Quasar Sample in configuration space between redshift 0.8 and 2.2. *arXiv e-prints*, arXiv:2007.08998. arXiv:2007.08998.

Howlett, C., A. J. Ross, L. Samushia, W. J. Percival, and M. Manera (2015). The clustering of the SDSS main galaxy sample - II. Mock galaxy catalogues and a measurement of the

growth of structure from redshift space distortions at z = 0.15. MNRAS *449*(1), 848–866. arXiv:1409.3238.

James, F. and M. Roos (1975). Minuit – a system for function minimization and analysis of the parameter errors and correlations. *Computer Physics Communications 10*, 343–367.

Jiménez, E., S. Contreras, N. Padilla, I. Zehavi, C. M. Baugh, et al. (2019). Extensions to the halo occupation distribution model for more accurate clustering predictions. MNRAS *490*(3), 3532–3544. arXiv:1906.04298.

Kitaura, F. S., G. Yepes, and F. Prada (2014). Modelling baryon acoustic oscillations with perturbation theory and stochastic halo biasing. MNRAS *439*, L21–L25. arXiv:1307.3285.

Klypin, A. and F. Prada (2018). Dark matter statistics for large galaxy catalogues: power spectra and covariance matrices. MNRAS *478*(4), 4602–4621. arXiv:1701.05690.

Klypin, A., G. Yepes, S. Gottlöber, F. Prada, and S. Heß (2016). MultiDark simulations: the story of dark matter halo concentrations and density profiles. MNRAS *457*(4), 4340–4359. arXiv:1411.4001.

Kong, H., K. J. Burleigh, A. Ross, J. Moustakas, C.-H. Chuang, et al. (2020). Removing Imaging Systematics from Galaxy Clustering Measurements with Obiwan : Application to the SDSS-IV extended Baryon Oscillation Spectroscopic Survey Emission Line Galaxy Sample. *arXiv e-prints*, arXiv:2007.08992. arXiv:2007.08992.

Laureijs, R., J. Amiaux, S. Arduini, J. L. Augueres, J. Brinchmann, et al. (2011). Euclid Definition Study Report. *arXiv e-prints*, arXiv:1110.3193. arXiv:1110.3193.

Lenz, D., B. S. Hensley, and O. Doré (2017). A New, Large-scale Map of Interstellar Reddening Derived from H I Emission. ApJ *846*(1), 38. arXiv:1706.00011.

Lin, S., J. L. Tinker, A. Klypin, F. Prada, M. R. Blanton, et al. (2020). The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: GLAM-QPM mock galaxy catalogs for the Emission Line Galaxy Sample. *arXiv e-prints*, arXiv:2007.08996. arXiv:2007.08996.

Macciò, A. V., A. A. Dutton, F. C. van den Bosch, B. Moore, D. Potter, et al. (2007). Concentration, spin and shape of dark matter haloes: scatter and the dependence on mass and environment. MNRAS *378*(1), 55–71. arXiv:astro-ph/0608157.

Mohammad, F. G., W. J. Percival, H.-J. Seo, M. J. Chapman, D. Bianchi, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Pairwise-Inverse-Probability and Angular Correction for Fibre Collisions in Clustering Measurements. *arXiv e-prints*, arXiv:2007.09005. arXiv:2007.09005.

Navarro, J. F., C. S. Frenk, and S. D. M. White (1996). The Structure of Cold Dark Matter Halos. ApJ *462*, 563. arXiv:astro-ph/9508025.

Neveux, R., E. Burtin, A. de Mattia, A. Smith, A. J. Ross, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from the anisotropic power spectrum of the Quasar sample between redshift 0.8 and 2.2. *arXiv e-prints*, arXiv:2007.08999. arXiv:2007.08999.

Okumura, T., C. Hikage, T. Totani, M. Tonegawa, H. Okada, et al. (2016). The Subaru FMOS galaxy redshift survey (FastSound). IV. New constraint on gravity theory from redshift space distortions at z $\sim$ 1.4. PASJ *68*(3), 38. arXiv:1511.08083.

Orsi, Á. A. and R. E. Angulo (2018). The impact of galaxy formation on satellite kinematics and redshift-space distortions. MNRAS *475*(2), 2530–2544. arXiv:1708.00956.

Padmanabhan, N., M. White, and J. D. Cohn (2009). Reconstructing baryon oscillations: A Lagrangian theory perspective. Phys. Rev. D *79*(6), 063523. arXiv:astro-ph/0812.2905.

Percival, W. J. and D. Bianchi (2017). Using angular pair upweighting to improve 3D clustering measurements. MNRAS *472*(1), L40–L44. arXiv:1703.02071.

Percival, W. J., B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, et al. (2010). Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. MNRAS *401*(4), 2148–2168. arXiv:0907.1660.

Percival, W. J., A. J. Ross, A. G. Sánchez, L. Samushia, A. Burden, et al. (2014). The clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: including covariance matrix errors. MNRAS *439*(3), 2531–2541. arXiv:1312.4841.

Philcox, O. H. E., D. J. Eisenstein, R. O'Connell, and A. Wiegand (2020). RASCALC: a jackknife approach to estimating single- and multitracer galaxy covariance matrices. MNRAS *491*(3), 3290–3317. arXiv:1904.11070.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2018). Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, arXiv:1807.06209. arXiv:1807.06209.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2019a). Planck 2018 results. V. CMB power spectra and likelihoods. *arXiv e-prints*, arXiv:1907.12875. arXiv:1907.12875.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2018b). Planck 2018 results. VIII. Gravitational lensing. *arXiv e-prints*, arXiv:1807.06210. arXiv:1807.06210.

Raichoor, A., A. de Mattia, A. J. Ross, C. Zhao, S. Alam, et al. (2020). The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Large-scale Structure Catalogues and Measurement of the isotropic BAO between redshift 0.6 and 1.1 for the Emission Line Galaxy Sample. *arXiv e-prints*, arXiv:2007.09007. arXiv:2007.09007.

Reid, B., S. Ho, N. Padmanabhan, W. J. Percival, J. Tinker, et al. (2016). SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues. MNRAS *455*, 1553–1573. arXiv:1509.06529.

Rezaie, M., H.-J. Seo, A. J. Ross, and R. C. Bunescu (2020). Improving Galaxy Clustering Measurements with Deep Learning: analysis of the DECaLS DR7 data. MNRAS. arXiv:1907.11355.

Riess, A. G., S. Casertano, W. Yuan, L. Macri, B. Bucciarelli, et al. (2018). Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant. ApJ *861*(2), 126. arXiv:1804.10655.

Riess, A. G., S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic (2019). Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM. ApJ *876*(1), 85. arXiv:1903.07603.

Ross, A. J., F. Beutler, C.-H. Chuang, M. Pellejero-Ibanez, H.-J. Seo, 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. MNRAS *464*(1), 1168–1191. arXiv:1607.03145.

Ross, A. J., W. J. Percival, and M. Manera (2015). The information content of anisotropic Baryon Acoustic Oscillation scale measurements. MNRAS *451*(2), 1331–1340. arXiv:1501.05571.

Ross, A. J., W. J. Percival, A. G. Sánchez, L. Samushia, S. Ho, et al. (2012). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: analysis of potential systematics. MNRAS *424*, 564–590. arXiv:1203.6499.

Ross, A. J., L. Samushia, C. Howlett, W. J. Percival, A. Burden, et al. (2015). The clustering of the SDSS DR7 main Galaxy sample - I. A 4 per cent distance measure at z = 0.15. MNRAS *449*(1), 835–847. arXiv:1409.3242.

Ruan, J. J., S. F. Anderson, P. J. Green, E. Morganson, M. Eracleous, et al. (2016). The Time-Domain Spectroscopic Survey: Understanding the Optically Variable Sky with SEQUELS in SDSS-III. ApJ *825*(2), 137. arXiv:1602.02752.

Samushia, L., W. J. Percival, and A. Raccanelli (2012). Interpreting large-scale redshift-space distortion measurements. MNRAS *420*(3), 2102–2119. arXiv:1102.1014.

Schlegel, D. J., D. P. Finkbeiner, and M. Davis (1998). Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. ApJ *500*(2), 525–553. arXiv:astro-ph/9710327.

Scolnic, D. M., D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, et al. (2018). The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. ApJ *859*(2), 101. arXiv:1710.00845.

Sellentin, E. and A. F. Heavens (2016). Parameter inference with estimated covariance matrices. MNRAS *456*(1), L132–L136. arXiv:1511.05969.

Sullivan, M., J. Guy, A. Conley, N. Regnault, P. Astier, et al. (2011). SNLS3: Constraints on Dark Energy Combining the Supernova Legacy Survey Three-year Data with Other Probes. ApJ *737*(2), 102. arXiv:1104.1444.

Tamone, A., A. Raichoor, C. Zhao, A. de Mattia, C. Gorgoni, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Growth rate of structure measurement from anisotropic clustering analysis in configuration space between redshift 0.6 and 1.1 for the Emission Line Galaxy sample. *arXiv e-prints*, arXiv:2007.09009. arXiv:2007.09009.

Tinker, J., A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, et al. (2008). Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality. ApJ *688*(2), 709–728. arXiv:astro-ph/0803.2706.

Tinker, J. L., B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, et al. (2010). The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests. ApJ *724*(2), 878–886. arXiv:1001.3162.

Vargas-Magaña, M., S. Ho, A. J. Cuesta, R. O'Connell, A. J. Ross, et al. (2018). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: theoretical systematics and Baryon Acoustic Oscillations in the galaxy correlation function. MNRAS *477*(1), 1153–1188. arXiv:1610.03506.

Wadekar, D. and R. Scoccimarro (2019). The Galaxy Power Spectrum Multipoles Covariance in Perturbation Theory. *arXiv e-prints*, arXiv:1910.02914. arXiv:1910.02914.

White, M., J. L. Tinker, and C. K. McBride (2014a). Mock galaxy catalogues using the quick particle mesh method. MNRAS *437*(3), 2594–2606. arXiv:1309.5532.

White, M., J. L. Tinker, and C. K. McBride (2014b). Mock galaxy catalogues using the quick particle mesh method. MNRAS *437*(3), 2594–2606. arXiv:1309.5532.

Wong, K. C., S. H. Suyu, G. C. F. Chen, C. E. Rusu, M. Millon, et al. (2019). H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes. *arXiv e-prints*, arXiv:1907.04869. arXiv:1907.04869.

Zentner, A. R., A. Hearin, F. C. van den Bosch, J. U. Lange, and A. Villarreal (2019). Constraints on assembly bias from galaxy clustering. MNRAS *485*(1), 1196–1209. arXiv:1606.07817.

Zhao, C., C.-H. Chuang, J. Bautista, A. de Mattia, A. Raichoor, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: one thousand multi-tracer mock catalogues with redshift evolution and systematics for galaxies and quasars of the final data release. *arXiv e-prints*, arXiv:2007.08997. arXiv:2007.08997.

Towards DESI

In this chapter we develop on projects complementing the large scale structure analysis presented in the previous chapter. We first investigate in Section 6.1 the large variations of $f\sigma_8$ seen when fitting mock power spectra measured with different lines of sight, as noted in Section 5.2.2, and notice that the error can be significantly reduced when averaging over multiple lines of sight. In Chapter 5, the power spectrum covariance matrix for cosmological inference was derived from a finite set of simulations. We study the consequence of the uncertainty in the covariance matrix on posterior measurements in Section 6.2. Finally, as a further opening to DESI, its forecast cosmological constraints are presented in Section 6.3. We emphasise that estimating and marginalising over observational systematics is key to guaranteeing accurate cosmological measurements with the forecast precision. We thus briefly present an alternative to the n-point correlation function analyses, known as forward modelling of the galaxy density field, which allows to properly marginalise, and even estimate, observational systematics, together with the cosmological inference.

6.1 Anisotropic measurements from mocks with different lines of sight

In Section 5.2.2, e.g. Figure 5.6, we noted some $\simeq 1\sigma$ scatter in the best fit cosmological parameters to the OuterRim-based mocks when using different lines of sight (see also Smith et al., 2020), as a consequence of the scatter in the power spectrum measurements. We first compute the expected cross-correlation between power spectrum multipoles measured using different lines of sight (Section 6.1.1), and emphasise that a significant reduction in uncertainties can be achieved by averaging power spectrum measurements over multiple lines of sight (Section 6.1.2). Finally, as a result, we show that $f\sigma_8$ measurements with two orthogonal lines of sight are significantly anti-correlated, explaining the variations seen in Section 5.2.2, and that averaging over three orthogonal lines of sight thus yields a relative reduction of uncertainty of better than $1/\sqrt{3}$ (Section 6.1.3).

These results are presented in Smith et al. (2020).

6.1.1 Cross-correlation of power spectrum measurements

Let us first consider power spectrum multipoles $\hat{P}^u_\ell$ and $\hat{P}^v_\ell$, measured with different (global) lines of sight $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$. Their cross-covariance matrix can be written as:

$$C^{uv}_{\ell\ell' ij} = \left\langle \left[\hat{P}^u_\ell(k_i) - \left\langle \hat{P}^u_\ell(k_i)\right\rangle\right]\left[\hat{P}^v_{\ell'}(k_j) - \left\langle \hat{P}^v_{\ell'}(k_j)\right\rangle\right]\right\rangle \tag{6.1}$$

$$= C_{\ell\ell' ij}\rho^{uv}_{\ell\ell' ij}, \tag{6.2}$$

where $C_{\ell\ell' ij}$ is the covariance matrix of power spectrum measurements with the same line of sight, and $\rho^{uv}_{\ell\ell' ij}$ is the cross-correlation, which we aim to calculate. By rearranging Eq. (6.2), we can write $\rho^{uv}_{\ell\ell' ij}$ as:

$$\rho^{uv}_{\ell\ell' ij} = \frac{\left\langle \left[\hat{P}^u_\ell(k_i) - \left\langle \hat{P}^u_\ell(k_i)\right\rangle\right]\left[\hat{P}^v_{\ell'}(k_j) - \left\langle \hat{P}^v_{\ell'}(k_j)\right\rangle\right]\right\rangle}{C_{\ell\ell' ij}} \tag{6.3}$$

$$= \frac{\left\langle \hat{P}^u_\ell(k_i)\hat{P}^v_{\ell'}(k_j)\right\rangle - \left\langle \hat{P}^u_\ell(k_i)\right\rangle\left\langle \hat{P}^v_{\ell'}(k_j)\right\rangle}{\left\langle \hat{P}^u_\ell(k_i)\hat{P}^u_{\ell'}(k_j)\right\rangle - \left\langle \hat{P}^u_\ell(k_i)\right\rangle\left\langle \hat{P}^u_{\ell'}(k_j)\right\rangle}, \tag{6.4}$$

where we arbitrarily took $\hat{\mathbf{u}}$ as line of sight in the denominator. The first term in the numerator of Eq. (6.4) yields (see Eq. (5.2)):

$$\left\langle \hat{P}^u_\ell(k)\hat{P}^v_{\ell'}(k')\right\rangle \propto \int d\Omega_u d\Omega_v \left\langle \delta^u_g(\mathbf{k})\delta^u_g(-\mathbf{k})\delta^v_g(\mathbf{k}')\delta^v_g(-\mathbf{k}')\right\rangle \mathcal{L}_\ell(\cos\theta_u)\mathcal{L}_{\ell'}(\cos\theta_v), \tag{6.5}$$

where the prefactors can be dropped, since they are the same in both the numerator and denominator of Eq. (6.4), and will cancel out. Assuming Gaussianity of δ_g, Wick theorem (2.54) gives:

$$\begin{aligned}\left\langle \delta^u_g(\mathbf{k})\delta^u_g(-\mathbf{k})\delta^v_g(\mathbf{k}')\delta^v_g(-\mathbf{k}')\right\rangle &= \left\langle \delta^u_g(\mathbf{k})\delta^u_g(-\mathbf{k})\right\rangle \left\langle \delta^v_g(\mathbf{k}')\delta^v_g(-\mathbf{k}')\right\rangle \\ &+ \left\langle \delta^u_g(\mathbf{k})\delta^v_g(\mathbf{k}')\right\rangle \left\langle \delta^u_g(-\mathbf{k})\delta^v_g(-\mathbf{k}')\right\rangle \\ &+ \left\langle \delta^u_g(-\mathbf{k})\delta^v_g(\mathbf{k}')\right\rangle \left\langle \delta^u_g(\mathbf{k})\delta^v_g(-\mathbf{k}')\right\rangle .\end{aligned} \tag{6.6}$$

The first of these terms is $\left\langle \hat{P}^u_\ell(k_i)\right\rangle \left\langle \hat{P}^v_{\ell'}(k_j)\right\rangle$, which can be dropped, since it is being subtracted out in the numerator of Eq. (6.4).

Let us now compute the cross-correlation $\left\langle \delta^u_g(\mathbf{k})\delta^v_g(\mathbf{k}')\right\rangle$. Let us write ψ_u ($\psi_z = -fv_z$ in Eq. (2.159)) the RSD displacement field along line of sight $\hat{\mathbf{u}}$, such that the redshift space position $\mathbf{s}_u$ is related to the real space position $\mathbf{r}$ through $\mathbf{s}_u = \mathbf{r} + \psi_u\hat{\mathbf{u}}$. Eq. (2.165) can be straightforwardly extended to the cross power spectrum of the redshift space density contrast seen with two lines of sight $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$:

$$P^{uv}_s(\mathbf{k}) = \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \left\langle e^{i(k_u\psi_u(\mathbf{x})-k_v\psi_v(\mathbf{r}+\mathbf{x}))}\left[\delta^r_g(\mathbf{r})+\partial_u\psi_u(\mathbf{r})\right]\left[\delta^r_g(\mathbf{r}+\mathbf{x})+\partial_v\psi_v(\mathbf{r}+\mathbf{x})\right]\right\rangle , \tag{6.7}$$

where we noted the real space galaxy density field δ^r_g. We also used $k_u = \mathbf{k}\cdot\hat{\mathbf{u}} = k\cos\theta_u$ (and a similar expression for k_v), with $\cos\theta_u$ the cosine of the angle between the wave vector $\mathbf{k}$ and line of sight $\hat{\mathbf{u}}$ (and similarly for $\cos\theta_v$). As in Section 2.4.1, in the limit $|k\psi_u|, |k\psi_v| \ll 1$ one can derive the equivalent of the Kaiser formula (2.163):

$$P^{uv}_s(\mathbf{k}) = b^2\left(1+\beta\cos^2\theta_u\right)\left(1+\beta\cos^2\theta_v\right)P^{\rm lin}_m(k) \tag{6.8}$$

Here we assumed a linear galaxy bias b ($\delta^r_g = b\delta_m$ with δ_m the real space matter density contrast), used $\beta = f/b$ and $P^{\rm lin}_m(k)$ the linear matter power spectrum.

Let us now specify the cosine angle $\cos\theta_v$. Taking the line of sight $\hat{\mathbf{u}}$ as reference, we can write the unit wavevector as:

$$\hat{\mathbf{k}} = (\cos\theta_u, \sin\theta_u\cos\phi_u, \sin\theta_u\sin\phi_u) \tag{6.9}$$

where $\cos\theta_u = \hat{\mathbf{k}}\cdot\hat{\mathbf{u}}$, and ϕ_u is the azimuthal angle around $\hat{\mathbf{u}}$. In the same coordinate system we can also write the second line of sight as:

$$\hat{\mathbf{v}} = (\cos\theta_{uv}, \sin\theta_{uv}\cos\phi_{uv}, \sin\theta_{uv}\sin\phi_{uv}) , \tag{6.10}$$

where θ_{uv} is the angle between $\hat{\mathbf{v}}$ and $\hat{\mathbf{u}}$, and ϕ_{uv} is the azimuthal angle of $\hat{\mathbf{v}}$ around $\hat{\mathbf{u}}$. Hence:

$$\begin{aligned}\cos\theta_v = \hat{\mathbf{k}}\cdot\hat{\mathbf{v}} &= \cos\theta_u\cos\theta_{uv} + \sin\theta_u\cos\phi_u\sin\theta_{uv}\cos\phi_{uv} + \sin\theta_u\cos\phi_u\sin\theta_{uv}\sin\phi_{uv} \\ &= \cos\theta_u\cos\theta_{uv} + \sin\theta_u\sin\theta_{uv}\cos\phi_v,\end{aligned} \tag{6.11}$$

where, in the last equality, we used $\phi_v = \phi_{uv} - \phi_u$.

We eventually have:

$$\left\langle \delta_g^u(\mathbf{k})\delta_g^v(\mathbf{k}')\right\rangle = (2\pi)^3\,\delta_D^{(3)}(\mathbf{k}+\mathbf{k}')\left[P_s^{uv}(\mathbf{k}) + P^{\rm noise}(\mathbf{k})\right], \tag{6.12}$$

where we introduced the shot noise term $P^{\rm noise}(\mathbf{k})$, which comes from the discrete nature of the galaxy density field. As in Section 4.1.2 one can define a regular grid of cells with galaxy occupation number $n_i \in \{0,1\}$, positions $\mathbf{r}_i$ and volume δV. Hence, we have:

$$\left\langle e^{i(k_u\psi_u(\mathbf{r}_i)-k_v\psi_v(\mathbf{r}_i))}n_i^2\right\rangle = \left\langle e^{i(k_u\psi_u(\mathbf{r}_i)-k_v\psi_v(\mathbf{r}_i))}n_i\right\rangle \tag{6.13}$$

$$= \left\langle e^{i(k_u\psi_u(\mathbf{r}_i)-k_v\psi_v(\mathbf{r}_i))}\left(1+\delta_g^r(\mathbf{r})\right)\right\rangle \bar{n}\delta V \tag{6.14}$$

where $\bar{n}$ is the mean density. The shot noise contribution to $\left\langle e^{i(k_u\psi_u(\mathbf{r})-k_v\psi_v(\mathbf{r}+\mathbf{x}))}\delta_g^r(\mathbf{r})\delta_g^r(\mathbf{r}+\mathbf{x})\right\rangle$ of Eq. (6.7) thus reads $\left\langle e^{i(k_u\psi_u(\mathbf{r})-k_v\psi_v(\mathbf{r}+\mathbf{x}))}\left(1+\delta_g^r(\mathbf{r})\right)\right\rangle \delta_D^{(3)}(\mathbf{x})/\bar{n}$. After integration over $\mathbf{x}$ in Eq. (6.7):

$$P^{\rm noise}(\mathbf{k}) = \left\langle \frac{1}{\bar{n}} e^{i(k_u\psi_u-k_v\psi_v)}\left(1+\delta_g^r\right)\right\rangle \tag{6.15}$$

where δ_g^r, ψ_u and ψ_v are taken at the same real position. Expanding $e^{i(k_u\psi_u-k_v\psi_v)}$ into series and applying Wick theorem, the above formula yields products involving $\left\langle \psi_u\delta_g^r\right\rangle$ (and $\left\langle \psi_v\delta_g^r\right\rangle$), which, as we will see in the following, vanishes at linear order. Hence, $\left\langle e^{i(k_u\psi_u-k_v\psi_v)}\delta_g^r\right\rangle = 0$ i.e. the δ_g^r term can be dropped in Eq. (6.15). Next, the phase shift term $\left\langle e^{i(k_u\psi_u-k_v\psi_v)}\right\rangle$ is the characteristic function of the Gaussian random variable, $k_u\psi_u - k_v\psi_v$, which has a mean of zero, and variance $\sigma^2_{\psi,uv}$. Hence the shot noise contribution reads:

$$P^{\rm noise}(\mathbf{k}) = \frac{e^{-\frac{1}{2}\sigma^2_{\psi,uv}(k)}}{\bar{n}} \tag{6.16}$$

where $\sigma^2_{\psi,uv}(k)$ is given by:

$$\sigma^2_{\psi,uv}(k) = \left\langle (k_u\psi_u - k_v\psi_v)^2\right\rangle \tag{6.17}$$

$$= k_u^2\left\langle\psi_u^2\right\rangle + k_v^2\left\langle\psi_v^2\right\rangle - 2k_uk_v\left\langle\psi_u\psi_v\right\rangle. \tag{6.18}$$

In linear perturbation theory, the displacement ψ_u due to RSD is given in Fourier space by (assuming no vorticity, see Eq. (2.52) and Eq. (2.159)):

$$\psi_u(\mathbf{k}) = if\frac{k_u}{k^2}\delta_m(\mathbf{k}). \tag{6.19}$$

Therefore, taking the Fourier transform, we find:

$$\begin{aligned}\langle\psi_u\psi_v\rangle &= f^2\int\frac{d^3k}{(2\pi)^3}\frac{k_uk_v}{k^4}P_m^{\rm lin}(k)\\ &= \frac{f^2}{(2\pi)^3}\int dkP_m^{\rm lin}(k)\int_0^{2\pi}d\phi_v\int_0^{\pi}\sin\theta_u d\theta_u\cos\theta_u\left(\cos\theta_u\cos\theta_{uv}+\sin\theta_u\sin\theta_{uv}\cos\phi_v\right)\\ &= f^2\sigma_d^2\cos\theta_{uv}\end{aligned} \tag{6.20}$$

with the one-dimensional dispersion of the displacement field (see Eq. (2.166)):

$$\sigma_d^2 = \frac{1}{6\pi^2}\int dkP_m^{\rm lin}(k). \tag{6.21}$$

Of course, we have $\langle\psi_u^2\rangle = \langle\psi_v^2\rangle = f^2\sigma_d^2$ since $\cos\theta_{uu} = \cos\theta_{vv} = 1$. That the cross-covariance $\langle\psi_u\psi_v\rangle / \langle\psi_u^2\rangle$ of the RSD displacement fields differs from ±1 for two lines of sight $\hat{\mathbf{u}} \neq \pm\hat{\mathbf{v}}$ is precisely the reason why $\hat{P}_\ell^u$ and $\hat{P}_\ell^v$ do not correlate perfectly, as we will see in the following. Especially, if the two lines of sight are orthogonal ($\cos\theta_{uv} = 0$), ψ_u and ψ_v do not correlate. Note also that $\left\langle\psi_u\delta_g^r\right\rangle$ yields an integral over $d\theta_u\cos\theta_u$ equal to zero, which justifies dropping δ_g^r in Eq. (6.15) as discussed previously. Eventually:

$$P^{\rm noise}(\mathbf{k}) = \frac{e^{-\frac{1}{2}\left(k_u^2+k_v^2-2k_uk_v\cos\theta_{uv}\right)f^2\sigma_d^2}}{\bar{n}}. \tag{6.22}$$

The shot noise contribution is largest ($1/\bar{n}$) when $f = 0$, as expected, and vanishes as $f \to +\infty$.

Finally, putting everything together, and using $d\Omega_u = d\phi_v d\theta_u \sin\theta_u$, the cross-correlation $\rho^{uv}_{\ell\ell' ij}$ can be written as

$$\rho^{uv}_{\ell\ell' ij} = \delta_{ij}\frac{\kappa^{uv}_{\ell\ell'}(k_i)}{\kappa_{\ell\ell'}(k_i)}, \tag{6.23}$$

with:

$$\begin{aligned}\kappa^{uv}_{\ell\ell'}(k) = \int_0^{2\pi} d\phi_v \int_0^{\pi} d\theta_u \sin\theta_u \Big[b^2\left(1+\beta\cos^2\theta_u\right)\left(1+\beta\cos^2\theta_v\right)P_m^{\rm lin}(k) \\ + \frac{1}{\bar{n}}e^{-\frac{1}{2}k^2\left(\cos^2\theta_u+\cos^2\theta_v-2\cos\theta_u\cos\theta_v\cos\theta_{uv}\right)f^2\sigma_d^2}\Big]^2 \mathcal{L}_\ell(\cos\theta_u)\mathcal{L}_{\ell'}(\cos\theta_v),\end{aligned} \tag{6.24}$$

$$\kappa_{\ell\ell'}(k) = 2\pi\int_0^{\pi} d\theta_u \sin\theta_u \left[b^2\left(1+\beta\cos^2\theta_u\right)^2 P_m^{\rm lin}(k) + \frac{1}{\bar{n}}\right]^2 \mathcal{L}_\ell(\cos\theta_u)\mathcal{L}_{\ell'}(\cos\theta_u). \tag{6.25}$$

Note that the above formula is valid at first order in $1/(P_m^{\rm lin}(k)\bar{n})$. Shot noise yields other terms, even at tree order in perturbation theory, see e.g. Meiksin and White (1999); Howlett and Percival (2017). Neglecting shot noise, the k and b dependence disappear; then, for two orthogonal lines of sight x and y, the limits for the different multipoles as $\beta \to 0$ and $\beta \to +\infty$ are:

$$\begin{aligned}\lim_{\beta\to0}\rho^{xy}_{00ii}(\beta) &= 1 & \lim_{\beta\to+\infty}\rho^{xy}_{00ii}(\beta) &= \frac{3}{35} \\ \lim_{\beta\to0}\rho^{xy}_{22ii}(\beta) &= -\frac{1}{2} & \lim_{\beta\to+\infty}\rho^{xy}_{22ii}(\beta) &= -\frac{33}{5810} \\ \lim_{\beta\to0}\rho^{xy}_{44ii}(\beta) &= \frac{3}{8} & \lim_{\beta\to+\infty}\rho^{xy}_{44ii}(\beta) &= \frac{5239}{199080}.\end{aligned}$$

Figure 6.1 displays the β dependence of the cross-correlation coefficients for two orthogonal lines of sight and neglecting shot noise. The power spectrum quadrupoles are always anti-correlated. Note also that the correlation of the off-diagonal terms (monopole - quadrupole, monople - hexadecapole and quadrupole - hexadecapole) between two orthogonal lines of sight is small, suggesting that there is no much redundant information in the two measurements using orthogonal lines of sight. In the next section, we take advantage of these findings to reduce the uncertainty of the measured power spectrum multipoles.

6.1.2 Reduction in power spectrum measurement uncertainties

Let us define a new estimator of the power spectrum multipoles, averaging over three orthogonal lines of sight:

$$\hat{P}_\ell^{3-\rm los}(k) = \frac{\hat{P}_\ell^x(k) + \hat{P}_\ell^y(k) + \hat{P}_\ell^z(k)}{3}. \tag{6.26}$$

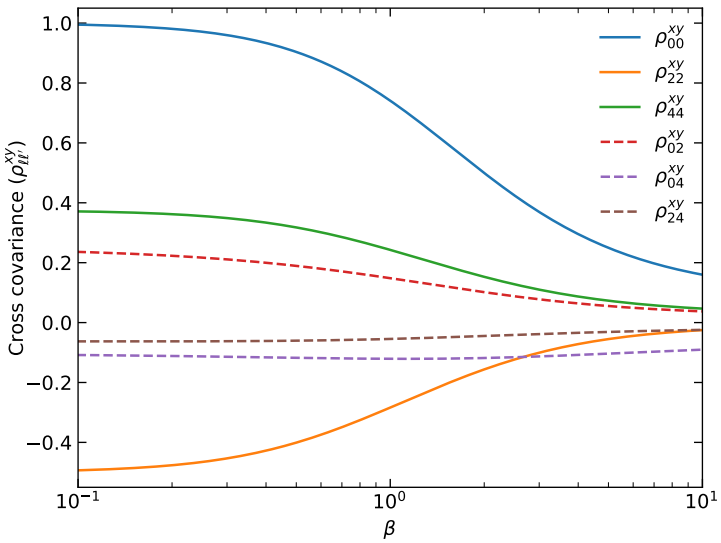

Figure 6.1 – Cross-correlation between power spectrum multipoles using two orthogonal lines of sight, in the zero shot noise limit. Taken from Smith et al. (2020).

The reduction in variance, when averaging over three orthogonal lines of sight, is (using Eq. (6.4)):

$$\rho_{\ell\ell' ij}^{3-\mathrm{los}} = \frac{1 + 2\rho_{\ell\ell' ij}^{xy}}{3}, \tag{6.27}$$

where we arbitrarily took $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ as lines of sight. For the case of no shot noise, the limits are:

$$\begin{aligned}
\lim_{\beta\to 0} \rho_{00ii}^{3-\mathrm{los}}(\beta) &= 1 & \lim_{\beta\to+\infty} \rho_{00ii}^{3-\mathrm{los}}(\beta) &= \frac{41}{105} \\
\lim_{\beta\to 0} \rho_{22ii}^{3-\mathrm{los}}(\beta) &= 0 & \lim_{\beta\to+\infty} \rho_{22ii}^{3-\mathrm{los}}(\beta) &= \frac{2872}{8715} \\
\lim_{\beta\to 0} \rho_{44ii}^{3-\mathrm{los}}(\beta) &= \frac{7}{12} & \lim_{\beta\to+\infty} \rho_{44ii}^{3-\mathrm{los}}(\beta) &= \frac{104779}{298620}.
\end{aligned} \tag{6.28}$$

Noticeably, in the $\beta \to 0$ limit, the variance of the quadrupole is reduced to zero. The predicted reduction of errors is shown with continuous lines in Figure 6.2. The error reduction in the quadrupole is much lower than $1/\sqrt{3}$, the expected value in absence of correlation. To check our theoretical prediction, we generated 100 Gaussian random fields, applied the Kaiser formula for different values of β, and computed the reduction in error bars when averaging the measured power spectrum multipoles over 3 orthogonal lines of sight. The results are shown with the data points, in perfect agreement with the theoretical predictions.

One can go further and average over all possible lines of sight ω of the solid angle Ω,

$$\hat{P}_\ell^{\mathrm{all-los}}(k) = \frac{1}{4\pi} \int d\Omega \hat{P}_\ell^{\omega}(k). \tag{6.29}$$

Hence:

$$\left\langle \hat{P}_\ell^{\mathrm{all-los}}(k) \hat{P}_{\ell'}^{\mathrm{all-los}}(k') \right\rangle = \frac{1}{(4\pi)^2} \int d\Omega \int d\Omega' \left\langle \hat{P}_\ell^{\omega}(k) \hat{P}_{\ell'}^{\omega'}(k') \right\rangle. \tag{6.30}$$

Following from statistical isotropy, a reference line of sight can be chosen, e.g. $\hat{\mathbf{u}}$, such that

$$\left\langle \hat{P}_\ell^{\mathrm{all-los}}(k) \hat{P}_{\ell'}^{\mathrm{all-los}}(k') \right\rangle = \frac{1}{4\pi} \int d\Omega \left\langle \hat{P}_\ell^{\omega}(k) \hat{P}_{\ell'}^{u}(k') \right\rangle \tag{6.31}$$

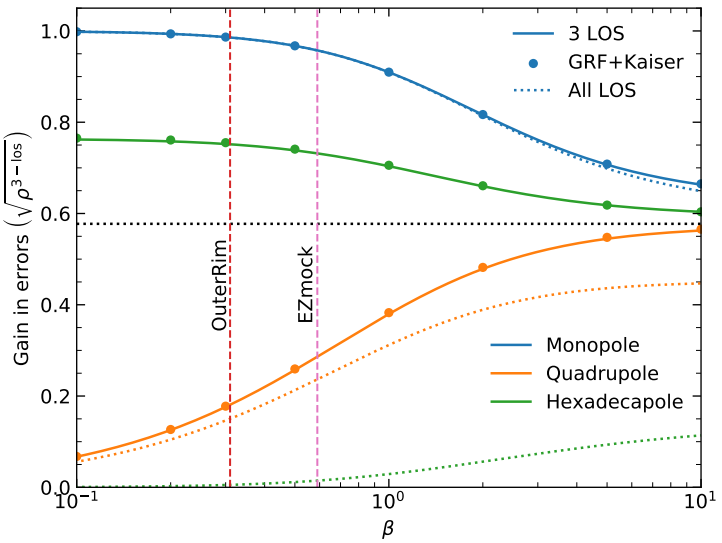

Figure 6.2 – Reduction in uncertainties when averaging power spectrum measurements over 3 orthogonal lines of sight. The prediction (6.27) is shown in continuous lines, and data points are measurements from Gaussian random fields implementing the Kaiser formula, hence without shot noise. Coloured dotted lines show the error reduction, averaging over all lines of sight, predicted by (6.32). The horizontal black dotted line indicates a reduction of errors by $1/\sqrt{3}$. Taken from Smith et al. (2020).

and therefore, the variance reduction is:

$$\rho^{\mathrm{all-los}}_{\ell\ell' ij} = \frac{1}{2}\int_0^\pi \sin\theta_{uv}\rho^{uv}_{\ell\ell' ij}. \tag{6.32}$$

In the case of no shot noise, when averaging over all lines of sight, the limits are:

$$\begin{aligned}
\lim_{\beta\to 0}\rho^{\mathrm{all-los}}_{00ii}(\beta) &= 1 & \lim_{\beta\to+\infty}\rho^{\mathrm{all-los}}_{00ii}(\beta) &= \frac{9}{25}\\
\lim_{\beta\to 0}\rho^{\mathrm{all-los}}_{22ii}(\beta) &= 0 & \lim_{\beta\to+\infty}\rho^{\mathrm{all-los}}_{22ii}(\beta) &= \frac{20592}{101675}\\
\lim_{\beta\to 0}\rho^{\mathrm{all-los}}_{44ii}(\beta) &= 0 & \lim_{\beta\to+\infty}\rho^{\mathrm{all-los}}_{44ii}(\beta) &= \frac{155584}{7838775}.
\end{aligned} \tag{6.33}$$

Already, with 3 orthogonal lines of sight, the relative variance of the quadrupole is pinned down to 0 in the limit $\beta \to 0$. Averaging over all lines of sight also sets the relative variance of the hexadecapole to 0 as $\beta \to 0$, while there is also a reduction in the limits as $\beta \to +\infty$ for all multipoles. The full β dependence can be seen in Figure 6.2 (dotted lines). Note however that such an averaging saturates the $\ell \neq \ell'$ cross-terms of the covariance matrix of $\hat{P}^{\mathrm{all-los}}_\ell(k)$, $\rho^{\mathrm{all-los}}_{\ell\ell' ij}C_{\ell\ell' ij}$, which is hence singular and cannot be used for cosmological parameter inference.

We also tested our formula Eq. (6.23) further, including the shot noise term, using the same 300 EZ mock boxes as used in the eBOSS ELG analysis in Section 5.4. Results are shown in Figure 6.3. To compute the theoretical prediction from Eq. (6.23), $f\sigma_d$ is taken as the standard deviation of the RSD displacements in the EZ mocks. β and b are the mean values obtained by fitting the measured power spectrum monopoles, quadrupoles and hexadecapoles with a linear Kaiser model damped by a Lorentzian Finger-of-God term over the k-range $0.03 < k < 0.12\,h\,\mathrm{Mpc}^{-1}$ (see Section 6.1.3 below). Both theoretical predictions and measurements are in good agreement, up to scales $0.15\,h\,\mathrm{Mpc}^{-1}$. Some divergence arises above that scale in the quadrupole, which we attribute to either non-linearity or the approximate nature of EZ mocks.

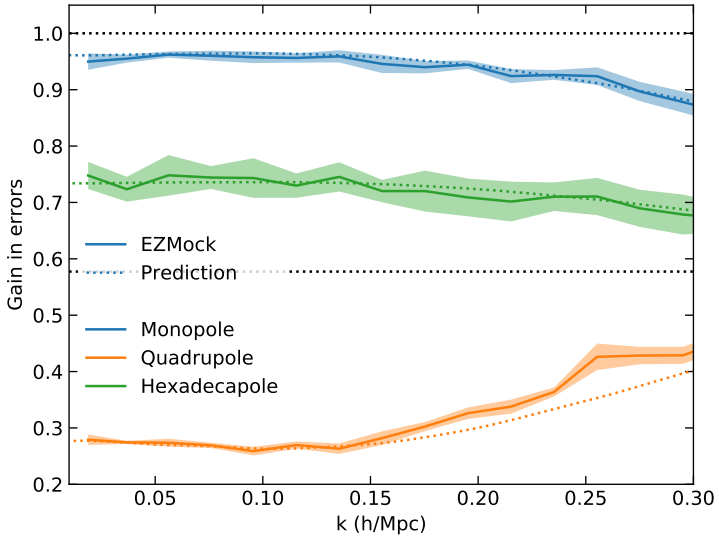

Figure 6.3 – Ratio of the error on the power spectrum multipoles measured on 300 EZ mocks and averaged over 3 orthogonal lines of sight to the average of the error for each line of sight. Shaded area represent the uncertainties estimated by splitting the 300 mocks into 6 sets of 50 mocks. The prediction (6.27), including shot noise, is shown in coloured dotted lines. The middle horizontal black dotted line indicates a reduction of errors by $1/\sqrt{3}$. Taken from Smith et al. (2020).

6.1.3 Reduction in RSD measurement uncertainties

In this section, we propagate the cross-covariance between power spectrum measurements determined previously up to the measurements of the cosmological parameters. We proceed with a Fisher forecast.

Cosmological parameters are estimated through a χ^2 minimisation,

$$\chi^2(\mathbf{p}) = \left(x_i^d - x_i^t(\mathbf{p})\right) \Psi_{ij} \left(x_j^d - x_j^t(\mathbf{p})\right), \tag{6.34}$$

where $\mathbf{x}^d$ is the measured power spectrum, $\mathbf{\Psi}$ is the inverse covariance matrix and $\mathbf{x}^t(\mathbf{p})$ is the model, which depends upon the parameters, $\mathbf{p}$. Taking the derivative of $\chi^2(\mathbf{p})$ with respect to p_α yields

$$0 = \frac{\partial x_i^t}{\partial p_\alpha} \Psi_{ij} \left(x_j^d - x_j^t\right). \tag{6.35}$$

At first order around the true parameter value $\mathbf{p}^t$ we have $x_i^t(\mathbf{p}) = x_i^t(\mathbf{p}^t) + \frac{\partial x_i^t}{\partial p_\alpha}(p_\alpha - p_\alpha^t)$, with all derivatives with respect to p_α taken at p_α^t. Without loss of generality, we set $p_\alpha^t = 0$ and $x_i^t(\mathbf{p}^t) = 0$. Eq. (6.35) can therefore be rearranged to give:

$$\frac{\partial x_i^t}{\partial p_\alpha} \Psi_{ij} \frac{\partial x_j^t}{\partial p_\beta} \hat{p}_\beta = \frac{\partial x_i^t}{\partial p_\alpha} \Psi_{ij} x_j^d. \tag{6.36}$$

Writing the Fisher matrix as:

$$F_{\alpha\beta} = \frac{\partial x_i^t}{\partial p_\alpha} \Psi_{ij} \frac{\partial x_j^t}{\partial p_\beta} \tag{6.37}$$

we obtain:

$$\hat{p}_\alpha = F_{\alpha\beta}^{-1} \frac{\partial x_i^t}{\partial p_\beta} \Psi_{ij} x_j^d. \tag{6.38}$$

The covariance of the measured parameters, when the power spectra measurements are calculated with respect to the same line of sight, is simply:

$$\langle \hat{p}_\alpha \hat{p}_\beta \rangle = F^{-1}_{\alpha\alpha'} \frac{\partial x^t_i}{\partial p_{\alpha'}} \Psi_{ij} \left\langle x^d_j x^d_{j'} \right\rangle \Psi_{j'i'} \frac{\partial x^t_{i'}}{\partial p_{\beta'}} F^{-1}_{\beta'\beta} \tag{6.39}$$

$$= F^{-1}_{\alpha\beta}. \tag{6.40}$$

For two different lines of sight $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$, the covariance is:

$$\left\langle \hat{p}^u_\alpha \hat{p}^v_\beta \right\rangle = F^{-1}_{\alpha\alpha'} \frac{\partial x^t_i}{\partial p_{\alpha'}} \Psi_{ij} \left\langle x^{d,u}_j x^{d,v}_{j'} \right\rangle \Psi_{j'i'} \frac{\partial x^t_{i'}}{\partial p_{\beta'}} F^{-1}_{\beta'\beta} \tag{6.41}$$

$$= F^{-1}_{\alpha\alpha'} \frac{\partial x^t_i}{\partial p_{\alpha'}} \Psi_{ij} C^{uv}_{jj'} \Psi_{j'i'} \frac{\partial x^t_{i'}}{\partial p_{\beta'}} F^{-1}_{\beta'\beta} \tag{6.42}$$

where $\mathbf{C}^{uv}$ is the cross-covariance between measurements of the power spectrum from the two different lines of sight $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$, which is estimated in Section 6.1.1. The same calculations derived in Section 6.1.2 can be used to calculate the reduction on uncertainties of cosmological parameters, when averaging over multiple lines of sight.

Let us compare the obtained prediction to RSD fits performed on the 300 EZ mocks. We consider two fitting ranges: (i) $0.03\,h\,\mathrm{Mpc}^{-1} < k < 0.12\,h\,\mathrm{Mpc}^{-1}$ and (ii) $0.03\,h\,\mathrm{Mpc}^{-1} < k < 0.20\,h\,\mathrm{Mpc}^{-1}$, using the monopole, quadrupole and hexadecapole. In case (i), we use the Kaiser model (2.164) damped by a Lorentzian Finger-of-God term (2.197b). In case (ii), we consider the TNS model as described in Section 2.4.3. We fit mock power spectra measured with 3 orthogonal lines separately, i.e. we do not perform the power spectrum averaging (6.27).

Fisher forecasts are computed following Eq. (6.42). For $\mathbf{C}$ we take the covariance matrix used in the fits. $\mathbf{C}^{uv}$ is determined from Eq. (6.23), using the mean f, β values obtained in the fits and $f\sigma_d$ as previously measured in the mocks.

Results are reported in Table 6.1. $f\sigma_8$ measurements with two orthogonal lines of sight are found to be anti-correlated. Since $f\sigma_8$ is related to the quadrupole-to-monopole ratio, this is in line with the anti-correlation of the power spectrum quadrupoles noted in Section 6.1.1. This also explains why the measured $f\sigma_8$ can vary significantly with the line of sight as noted in Section 5.2.2. Hence, averaging the measured $f\sigma_8$ over the three lines of sight yields a reduction in uncertainties larger than $1/\sqrt{3}$. For the wide fitting range (ii), including quasi non-linear scales, the reduction of uncertainties in the $f\sigma_8$ measurement is somewhat lower, but still better than $1/\sqrt{3}$. There is also a non-negligible error reduction in the measurement of $\alpha_\parallel$, especially for the narrow fitting range (i). For both the narrow (i) and wide (ii) fitting ranges, theoretical predictions are in relatively good agreement with the measurements from the mocks.

In conclusion, averaging anisotropic power spectrum measurements from simulations with different lines of sight is an easy way to increase the precision of such mock challenges, for (almost) no computational cost. This may prove useful for future mock challenges, as organised by DESI. Before moving on to DESI prospects, let us first present the impact of estimating covariance matrices from simulations onto parameter inference.

6.2 Mock-based covariance matrices

In this section we recap and develop ideas about the estimation of the covariance matrix of a Gaussian likelihood measured from simulations (mocks). We first review the state-of-the-art frequentist correction (Section 6.2.1) and extend it to the case of the combination of likelihoods

(i) $0.03 - 0.12\,h\,\mathrm{Mpc}^{-1}$	$f\sigma_8$	$\alpha_\parallel$	$\alpha_\perp$
ρ^{xy}	-0.244 ± 0.016	0.420 ± 0.037	0.723 ± 0.024
prediction	-0.274	0.405	0.726
$\sqrt{\rho^{3-\mathrm{los}}}$	0.414 ± 0.013	0.783 ± 0.016	0.903 ± 0.009
prediction	0.388	0.777	0.904
(ii) $0.03 - 0.2\,h\,\mathrm{Mpc}^{-1}$	$f\sigma_8$	$\alpha_\parallel$	$\alpha_\perp$
ρ^{xy}	-0.076 ± 0.041	0.480 ± 0.038	0.756 ± 0.022
prediction	-0.035	0.507	0.754
$\sqrt{\rho^{3-\mathrm{los}}}$	0.532 ± 0.025	0.808 ± 0.017	0.915 ± 0.008
prediction	0.557	0.819	0.914

Table 6.1 – Measured values of the cross-correlation, ρ^{xy}, and the reduction of uncertainties, $\sqrt{\rho^{3-\mathrm{los}}}$, for the parameters $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$ fitted on the EZ mocks. Error bars are estimated by splitting the 300 mocks into 6 sets of 50 mocks. We also report the prediction from the Fisher analysis.

which are known to be independent (Section 6.2.2). The obtained formulae were utilised in the eBOSS ELG analysis (see Section 5.1.6). We then present a Bayesian approach (Section 6.2.3), which we show to fail in the case of a high-dimensional data vector.

6.2.1 Frequentist correction

Let us first recap the correction of (Hartlap et al., 2007; Dodelson and Schneider, 2013; Percival et al., 2014). We consider a number n_m of independent Gaussian observations $\mathbf{m}_i$ (the mocks) of size n_b, with true mean $\mathbf{m}^t$ and covariance matrix $\mathbf{C}^t$. Then, $\mathbf{M} = \sum_{i=1}^{n_m} (\mathbf{m}_i - \overline{\mathbf{m}}) (\mathbf{m}_i - \overline{\mathbf{m}})^T$ with $\overline{\mathbf{m}} = \sum_{i=1}^{n_m} \mathbf{m}_i$ follows a Wishart distribution (Wishart, 1928; Anderson, 1958):

$$\mathcal{W}(\mathbf{M}|\mathbf{C}^t) = \frac{1}{2^{\frac{(n_m-1)n_b}{2}} |\mathbf{C}^t|^{\frac{n_m-1}{2}} \Gamma_{n_b}\left(\frac{n_m-1}{2}\right)} |\mathbf{M}|^{\frac{n_m-n_b-2}{2}} e^{-\frac{1}{2}\mathrm{Tr}\left(\mathbf{C}^{t^{-1}}\mathbf{M}\right)}, \tag{6.43}$$

which is the joint probability of the $n_b(n_b-1)/2$ free elements of $\mathbf{M}$.

$\mathbf{M}^{-1}$ follows an inverse Wishart distribution:

$$\mathcal{W}^{-1}(\mathbf{M}^{-1}|\mathbf{C}^t) = \frac{|\mathbf{C}^t|^{\frac{n_m-1}{2}}}{2^{\frac{(n_m-1)n_b}{2}} \Gamma_{n_b}\left(\frac{n_m-1}{2}\right)} |\mathbf{M}|^{-\frac{n_m+n_b}{2}} e^{-\frac{1}{2}\mathrm{Tr}\left(\mathbf{C}^t\mathbf{M}^{-1}\right)}. \tag{6.44}$$

Integrating $\mathbf{M}^{-1}\mathcal{W}^{-1}(\mathbf{M}^{-1}|\mathbf{C}^t)$ over the ensemble of positive-definite matrices yields:

$$\left\langle \mathbf{M}^{-1} \right\rangle = \frac{1}{n_m - n_b - 2} \mathbf{C}^{t^{-1}}. \tag{6.45}$$

Hence, an unbiased estimate of $\mathbf{\Psi}^t = \mathbf{C}^{t^{-1}}$ is given by (see Eq. (5.13), Hartlap et al. 2007):

$$\mathbf{\Psi} = \frac{n_m - n_b - 2}{n_m - 1} \left(\frac{\mathbf{M}}{n_m - 1}\right)^{-1} \tag{6.46}$$

$$= (1 - D)\,\mathbf{C}^{-1} \qquad \text{with} \qquad \mathbf{C} = \frac{\mathbf{M}}{n_m - 1}, \qquad D = \frac{n_b + 1}{n_m - 1}. \tag{6.47}$$

Note however that the estimated precision matrix $\mathbf{\Psi} = \mathbf{\Psi}^t + \Delta\mathbf{\Psi}$ has an intrinsic scatter (Dodelson and Schneider, 2013):

$$\langle \Delta\Psi_{ij} \Delta\Psi_{i'j'} \rangle = A \Psi^t_{ij} \Psi^t_{i'j'} + B \left(\Psi^t_{ii'} \Psi^t_{jj'} + \Psi^t_{ij'} \Psi^t_{ji'} \right) \tag{6.48}$$

with (Taylor et al., 2013):

$$A = \frac{2}{(n_m - n_b - 1)(n_m - n_b - 4)}, \tag{6.49}$$

$$B = \frac{n_m - n_b - 2}{(n_m - n_b - 1)(n_m - n_b - 4)}. \tag{6.50}$$

Let us consider a Gaussian likelihood for the measurements $\mathbf{x}^d$:

$$L(\mathbf{x}^d|\mathbf{\Psi}) \propto e^{-\frac{1}{2}\left(\mathbf{x}^d - \mathbf{x}^t\right)^T \mathbf{\Psi}\left(\mathbf{x}^d - \mathbf{x}^t\right)}. \tag{6.51}$$

At first order the measured parameters (assuming their true value to be zero, without loss of generality) are given by Eq. (6.38):

$$\hat{p}_\alpha = [F + \Delta F]^{-1}_{\alpha\alpha'} \frac{\partial x^t_i}{\partial p_{\alpha'}} \Psi_{ij} \left(x^d_j - x^t_j\right) \tag{6.52}$$

where (see Eq. (6.37)):

$$F_{\alpha\beta} = \frac{\partial x^t_i}{\partial p_\alpha} \Psi^t_{ij} \frac{\partial x^t_j}{\partial p_\beta} \tag{6.53}$$

$$\Delta F_{\alpha\beta} = \frac{\partial x^t_i}{\partial p_\alpha} \Delta\Psi_{ij} \frac{\partial x^t_j}{\partial p_\beta}. \tag{6.54}$$

As can be seen from Eq. (6.52), the $\hat{p}_\alpha$ measurement can be biased if (i) the covariance matrix depends on the model (which we have ignored so far) and (ii) if the estimated covariance matrix correlates with the data, $\left\langle \Psi_{ij} x^d_j \right\rangle \neq 0$. We will ignore these possibilities in the following, though e.g. (ii) should be kept in my mind when covariance matrices are estimated from the data itself, e.g. with jacknife methods.

At first order, $[F + \Delta F]^{-1}_{\alpha\beta} = F^{-1}_{\alpha\beta} + \left[F^{-1}\Delta F F^{-1}\right]_{\alpha\beta}$. Then, the leading order parameter covariance is $\langle \hat{p}_\alpha \hat{p}_\beta \rangle \ni F^{-1}_{\alpha\beta}$ and the next-to-leading (second order) contribution is (Dodelson and Schneider, 2013):

$$\langle \hat{p}_\alpha \hat{p}_\beta \rangle \ni F^{-1}_{\alpha\alpha'} \left[\frac{\partial x^t_i}{\partial p_{\alpha'}} \frac{\partial x^t_{i'}}{\partial p_{\beta'}} C^t_{jj'} (\Delta\Psi)_{ij} (\Delta\Psi)_{i'j'} \right] F^{-1}_{\beta\beta'} \tag{6.55}$$

$$- \left[F^{-1}\Delta F F^{-1}\Delta F F^{-1}\right]_{\alpha\beta} \tag{6.56}$$

where we recall that $C^t_{ij} = \left\langle \left(x^d_i - x^t_i\right)\left(x^d_j - x^t_j\right) \right\rangle$. The A term from Eq. (6.48) gives $AF_{\alpha\beta}$ for Eq. (6.55) and $-AF_{\alpha\beta}$ for Eq. (6.56). The B term from Eq. (6.48) gives $B\,(n_b + 1)\,F_{\alpha\beta}$ for Eq. (6.55) and $-B\,(n_p + 1)\,F_{\alpha\beta}$ for Eq. (6.56), with n_p the number of parameters (dimension of the Fisher matrix F). Hence, the total, up to second order parameter covariance reads (Dodelson and Schneider, 2013):

$$V_{\alpha\beta} = \langle \hat{p}_\alpha \hat{p}_\beta \rangle = [1 + B\,(n_b - n_p)]\,F_{\alpha\beta}. \tag{6.57}$$

Note however that we do not have access to $F_{\alpha\beta}$ (6.54), since we do not know the true inverse covariance $\mathbf{\Psi}^t$. Instead, in the Gaussian case, the parameter covariance matrix derived from the likelihood (given by the inverse of its second log-derivative or likelihood profiling) is (Percival et al., 2014):

$$\sigma^2_{\alpha\beta} = [F + \Delta F]^{-1}_{\alpha\beta}\,. \tag{6.58}$$

The first order average contributions of ΔF are zero, but the second order ones read:

$$\sigma^2_{\alpha\beta} \ni \left[F^{-1}\Delta F F^{-1}\Delta F F^{-1}\right]_{\alpha\beta} \tag{6.59}$$

i.e. Eq. (6.56). Following from the previous calculation we have, at second order:

$$\sigma^2_{\alpha\beta} = \left[1 + A + B\left(n_p + 1\right)\right] F_{\alpha\beta}. \tag{6.60}$$

Therefore, on top of the Hartlap factor (6.47), the full correction to be applied to the data parameter covariance estimated from the likelihood is:

$$m_1 = \frac{V_{\alpha\beta}}{\sigma^2_{\alpha\beta}} = \frac{1 + B\left(n_b - n_p\right)}{1 + A + B\left(n_p + 1\right)}. \tag{6.61}$$

Let us consider the case where we fit the mocks used to estimate the inverse covariance matrix $\mathbf{\Psi}$. The covariance of the mocks is $\left\langle \left(x^d_i - x^t_i\right)\left(x^d_j - x^t_j\right)\right\rangle = C_{ij} = (1-D)\left[\Psi^t + \Delta\Psi\right]^{-1}_{ij}$, hence with Eq. (6.52) the covariance of the best fit parameters obtained on the mocks is (Percival et al., 2014):

$$\langle \hat{p}_\alpha \hat{p}_\beta \rangle = (1-D)^{-1}\left[F + \Delta F\right]^{-1}_{\alpha\beta}. \tag{6.62}$$

Hence, to be compared with the parameter covariance derived from the likelihood in which both the Harlap factor of Eq. (6.47) and m_1 of Eq. (6.61) are included, the covariance of best fit parameter values obtained from the mocks should be rescaled by (Percival et al., 2014):

$$m_2 = \frac{V_{\alpha\beta}}{\langle \hat{p}_\alpha \hat{p}_\beta \rangle} = \frac{V_{\alpha\beta}}{\sigma^2_{\alpha\beta}} \frac{\sigma^2_{\alpha\beta}}{\langle \hat{p}_\alpha \hat{p}_\beta \rangle} \tag{6.63}$$

$$= m_1 (1-D)^{-1}. \tag{6.64}$$

6.2.2 Combination of independent likelihoods

As outlined in Section 5.1.6, the prescription detailed above to rescale the parameter covariance matrix is formally not correct when combining two (e.g. NGC and SGC) likelihoods with mock-based covariance matrices which are further assumed to be independent (i.e. their cross-covariance is zero).

Indeed, an intuitive use of Eq. (6.61) would take n_b as the number of bins in either NGC or SGC and n_p the total number of parameters. Then, however, the constraint $n_p \leq n_b$ inherent to this formula appears artificial. Indeed, there is no issue with having $n_p > n_b$ as long as n_p is less than the total number of bins. Another approach would be to take n_b as the total number of bins in NGC and SGC; though this would be correct if we estimated the combined NGC and SGC covariance from mocks, this does not apply to our case where we impose the cross-covariance between NGC and SGC to be zero. Therefore, we have to revise Eq. (6.61) in the context of a block diagonal covariance matrix.

As in the previous section, the Hartlap correction is applied to the inverse NGC covariance matrix to obtain the precision matrix (see Eq. (6.47)):

$$\mathbf{\Psi}^{\mathrm{NGC}} = \left(1 - D^{\mathrm{NGC}}\right)\left(\mathbf{C}^{\mathrm{NGC}}\right)^{-1}, \quad D^{\mathrm{NGC}} = \frac{n_b^{\mathrm{NGC}} + 1}{n_m^{\mathrm{NGC}} - 1} \tag{6.65}$$

with n_b^{NGC} the number of bins and n_m^{NGC} the number of mocks in NGC; similarly for SGC. The full precision matrix $\mathbf{\Psi}$ is a block-diagonal matrix built from $\mathbf{\Psi}^{\mathrm{NGC}}$ and $\mathbf{\Psi}^{\mathrm{SGC}}$. The corrections

to the parameter covariance are a bit less trivial, as the errors on each inverse covariance matrix must be propagated through the Fisher information.

The covariance of the precision matrix fluctuations of Eq. (6.48) becomes:

$$\langle \Delta\Psi_{ij}\Delta\Psi_{i'j'}\rangle = \left(A^{\rm NGC}_{iji'j'} + A^{\rm SGC}_{iji'j'}\right)\Psi^t_{ij}\Psi^t_{i'j'} + \left(B^{\rm NGC}_{iji'j'} + B^{\rm SGC}_{iji'j'}\right)\left(\Psi^t_{ii'}\Psi^t_{jj'} + \Psi^t_{ij'}\Psi^t_{ji'}\right) \tag{6.66}$$

where $A^{\rm NGC}_{iji'j'}$ is constant (equal to $A^{\rm NGC}$) if indices $iji'j'$ all lie in the same block NGC, zero elsewhere (i.e. cross-covariance terms between NGC and SGC are zero), and similarly for SGC and B terms. Let us consider the contribution from the A terms first. From Eq. (6.55) one gets the contribution:

$$\langle \hat{p}_\alpha \hat{p}_\beta\rangle \ni F^{-1}_{\alpha\alpha'}\left[A^{\rm NGC}F^{\rm NGC}_{\alpha'\beta'} + A^{\rm SGC}F^{\rm SGC}_{\alpha'\beta'}\right]F^{-1}_{\beta\beta'} \tag{6.67}$$

where we split the total Fisher information $\mathbf{F} = \mathbf{F}^{\rm NGC} + \mathbf{F}^{\rm SGC}$ (since NGC and SGC are independent). Eq. (6.56) gives:

$$\langle \hat{p}_\alpha \hat{p}_\beta\rangle \ni -F^{-1}_{\alpha\alpha'}F^{-1}_{\beta'\alpha''}F^{-1}_{\beta''\beta}\left(A^{\rm NGC}F^{\rm NGC}_{\alpha'\beta'}F^{\rm NGC}_{\alpha''\beta''} + A^{\rm SGC}F^{\rm SGC}_{\alpha'\beta'}F^{\rm SGC}_{\alpha''\beta''}\right). \tag{6.68}$$

Let us move to the B terms. From Eq. (6.55) one gets the contribution:

$$\langle \hat{p}_\alpha \hat{p}_\beta\rangle \ni F^{-1}_{\alpha\alpha'}\left[B^{\rm NGC}\left(n^{\rm NGC}_b + 1\right)F^{\rm NGC}_{\alpha'\beta'} + B^{\rm SGC}\left(n^{\rm SGC}_b + 1\right)F^{\rm SGC}_{\alpha'\beta'}\right]F^{-1}_{\beta\beta'}, \tag{6.69}$$

and from Eq. (6.56):

$$\begin{aligned}\langle \hat{p}_\alpha \hat{p}_\beta\rangle \ni -F^{-1}_{\alpha\alpha'}F^{-1}_{\beta'\alpha''}F^{-1}_{\beta''\beta}\Big(&B^{\rm NGC}F^{\rm NGC}_{\alpha'\alpha''}F^{\rm NGC}_{\beta'\beta''} \\ &+B^{\rm SGC}F^{\rm SGC}_{\alpha'\alpha''}F^{\rm SGC}_{\beta'\beta''} + B^{\rm NGC}F^{\rm NGC}_{\alpha'\beta''}F^{\rm NGC}_{\beta'\alpha''} + B^{\rm SGC}F^{\rm SGC}_{\alpha'\beta''}F^{\rm SGC}_{\beta'\alpha''}\Big).\end{aligned} \tag{6.70}$$

Formulae above could be evaluated numerically, but we will consider a simpler case in the following. We assume that $F^{\rm NGC}_{\alpha\beta}$ and $F^{\rm SGC}_{\alpha\beta}$ are block-diagonal, with specific parameters for NGC and SGC which are uncorrelated, and a set of common parameters $\alpha\beta$ for which the Fisher information content can be written $F^{\rm NGC}_{\alpha\beta} = f^{\rm NGC}F_{\alpha\beta}$ (resp. $F^{\rm SGC}_{\alpha\beta} = f^{\rm SGC}F_{\alpha\beta}$), with $f^{\rm NGC} + f^{\rm SGC} = 1$. We also assume the common parameters to be uncorrelated to the NGC and SGC specific parameters [1]. Then, the contribution from the A and B terms to the covariance of the common parameters is simply the sum of terms $\mathcal{C}_1$ (from Eq. (6.55)) and $\mathcal{C}_2$ (from Eq. (6.56)):

$$\langle \hat{p}_\alpha \hat{p}_\beta\rangle \ni \left[\mathcal{C}^{\rm NGC}_1 + \mathcal{C}^{\rm SGC}_1 - \mathcal{C}^{\rm NGC}_2 - \mathcal{C}^{\rm SGC}_2\right]F^{-1}_{\alpha\beta} \tag{6.71}$$

where:

$$\mathcal{C}^{\rm NGC}_1 = A^{\rm NGC}f^{\rm NGC} + B^{\rm NGC}f^{\rm NGC}\left(n^{\rm NGC}_b + 1\right) \tag{6.72}$$

and

$$\mathcal{C}^{\rm NGC}_2 = A^{\rm NGC}\left(f^{\rm NGC}\right)^2 + B^{\rm NGC}f^{\rm NGC}\left(n^{\rm NGC}_{\rm eff} + f^{\rm NGC}\right) \tag{6.73}$$

where we use the effective number of parameters $n^{\rm NGC}_{\rm eff} = n^{\rm NGC}_{\rm sp} + f^{\rm NGC}n_{\rm co}$, with $n^{\rm NGC}_{\rm sp}$ the number of parameters specific to NGC (similarly for SGC) and $n_{\rm co}$ the number of parameters in common.

The contribution from the A and B terms to the covariance of e.g. the NGC specific parameters is obtained by forcing $f^{\rm NGC} = 1$ (then $f^{\rm SGC} = 0$) in Eq. (6.72) and Eq. (6.73) — while

1. This is of course not the case in practice.

keeping $n_{\rm eff}^{\rm NGC}$ fixed. In this case, only errors from the NGC precision matrix contribute to the parameter covariance. In the simplified case where $n_{\rm co} = 0$, we recover Eq. (6.57):

$$\langle \hat{p}_\alpha \hat{p}_\beta \rangle \ni B^{\rm NGC} \left[n_b^{\rm NGC} - n_{\rm sp}^{\rm NGC} \right] F^{-1}_{\alpha\beta}. \tag{6.74}$$

Adding up A and B contributions, the parameter covariance is:

$$V_{\alpha\beta} = \left[1 + \mathcal{C}_1^{\rm NGC} + \mathcal{C}_1^{\rm SGC} - \mathcal{C}_2^{\rm NGC} - \mathcal{C}_2^{\rm SGC} \right] F^{-1}_{\alpha\beta}. \tag{6.75}$$

Again, the parameter covariance estimated from the likelihood has for second order contribution Eq. (6.59), i.e. Eq. (6.56), which has just been calculated. Then:

$$\sigma^2_{\alpha\beta} = \left[1 + \mathcal{C}_2^{\rm NGC} + \mathcal{C}_2^{\rm SGC} \right] F^{-1}_{\alpha\beta}. \tag{6.76}$$

Therefore, the full correction to apply to the parameter covariance estimated from the likelihood is:

$$m_1 = \frac{V_{\alpha\beta}}{\sigma^2_{\alpha\beta}} = \frac{1 + \mathcal{C}_1^{\rm NGC} + \mathcal{C}_1^{\rm SGC} - \mathcal{C}_2^{\rm NGC} - \mathcal{C}_2^{\rm SGC}}{1 + \mathcal{C}_2^{\rm NGC} + \mathcal{C}_2^{\rm SGC}}. \tag{6.77}$$

In case we fit the mocks used to produce the covariance matrix, the covariance of the measurements x_i is just C_{ij}:

$$\left\langle \left(x_i^d - x_i^t \right) \left(x_j^d - x_j^t \right) \right\rangle = C_{ij} \tag{6.78}$$

$$= \left(1 - D_{ij}^{\rm NGC} \right) \Psi_{ij}^{-1} + \left(1 - D_{ij}^{\rm SGC} \right) \Psi_{ij}^{-1} \tag{6.79}$$

where $D_{ij}^{\rm NGC} = D^{\rm NGC}$ if ij lie in the NGC block, zero otherwise. Then:

$$\langle \hat{p}_\alpha \hat{p}_\beta \rangle = [F + \Delta F]^{-1}_{\alpha\alpha'} [F + \Delta F]^{-1}_{\beta\beta'} \frac{\partial x_i^t}{\partial p_{\alpha'}} \frac{\partial x_{i'}^t}{\partial p_{\beta'}} \Psi_{ii'} \Psi_{jj'} \left\langle \left(x_i^d - x_i^t \right) \left(x_j^d - x_j^t \right) \right\rangle \tag{6.80}$$

$$= [F + \Delta F]^{-1}_{\alpha\alpha'} [F + \Delta F]^{-1}_{\beta\beta'} \left\{ \left(1 - D^{\rm NGC} \right) [F + \Delta F]^{\rm NGC}_{\alpha'\beta'} + \left(1 - D^{\rm SGC} \right) [F + \Delta F]^{\rm SGC}_{\alpha'\beta'} \right\} \tag{6.81}$$

$$= [F + \Delta F]^{-1}_{\alpha\beta} \left[\left(1 - D^{\rm NGC} \right) f^{\rm NGC} + \left(1 - D^{\rm SGC} \right) f^{\rm SGC} \right]. \tag{6.82}$$

To be compared with the parameter covariance derived from the likelihood in which both the Harlap factor of Eq. (6.65) and m_1 of Eq. (6.77) are included, the covariance of the distribution of best fit parameter values obtained from the mocks should be rescaled by:

$$m_2 = \frac{V_{\alpha\beta}}{\langle \hat{p}_\alpha \hat{p}_\beta \rangle} = \frac{V_{\alpha\beta}}{\sigma^2_{\alpha\beta}} \frac{\sigma^2_{\alpha\beta}}{\langle \hat{p}_\alpha \hat{p}_\beta \rangle} \tag{6.83}$$

$$= m_1 \left[\left(1 - D^{\rm NGC} \right) f^{\rm NGC} + \left(1 - D^{\rm SGC} \right) f^{\rm SGC} \right]^{-1}. \tag{6.84}$$

The extension to a higher number of combined independent likelihoods is straightforward.

The m_1 factor described above yields a correction of 5% only on the parameter covariance matrix of the eBOSS ELG analysis. Hence, above calculations may appear as an academic exercise. However, to extract most information from future data sets, there is a growing interest in using data representation of higher dimensionality, such as the bispectrum, or the power spectrum in small wavebands. Typically, to keep the same correction (6.47), the increase of the number of mocks n_m should be roughly proportional to the increase of the number of bins n_b. As the computational cost of mocks is still relatively high, uncertainties in the mock-based covariance matrix may be significant for these future analyses. There is thus interest for a more accurate approach to the propagation of such uncertainties in the cosmological inference, beyond the Gaussian approximation made so far.

6.2.3 A Bayesian approach

The prescription described in the previous sections is fundamentally frequentist. A Bayesian approach was developed in Sellentin and Heavens (2016). The Wishart distribution Eq. (6.43) is the probability distribution of the mock-based covariance matrix $\mathbf{C} = \mathbf{M}/(n_m - 1)$, given the true covariance matrix $\mathbf{C}^t$. To obtain the PDF of the true covariance matrix given $\mathbf{C}$, Sellentin and Heavens (2016) consider a Jeffreys prior:

$$\pi(\mathbf{C}^t) = \left|\mathbf{C}^t\right|^{-\alpha} \qquad \text{with} \qquad \alpha = \frac{n_b + 1}{2}. \tag{6.85}$$

We introduce the parameter α as it will prove useful in the following. The PDF of the true covariance matrix is (up to some normalisation) the product of Eq. (6.43) with the prior $\pi(\mathbf{C}^t)$:

$$\begin{aligned} p(\mathbf{C}^t|\mathbf{M}) &\propto \mathcal{W}(\mathbf{M}|\mathbf{C}^t)\pi(\mathbf{C}^t) && (6.86)\\ &\propto \left|\mathbf{C}^t\right|^{-\frac{n_m-1}{2}-\alpha} e^{-\frac{1}{2}\mathrm{Tr}\left(\mathbf{C}^{t^{-1}}\mathbf{M}\right)} && (6.87)\\ &= \frac{|\mathbf{M}|^{\frac{n_m-n_b-2+2\alpha}{2}}}{2^{\frac{(n_m-n_b-2+2\alpha)n_b}{2}}\Gamma_{n_b}\left(\frac{n_m-n_b-2+2\alpha}{2}\right)} \left|\mathbf{C}^t\right|^{-\frac{n_m-1}{2}-\alpha} e^{-\frac{1}{2}\mathrm{Tr}\left(\mathbf{C}^{t^{-1}}\mathbf{M}\right)}. && (6.88) \end{aligned}$$

Normalisation factors were reintroduced in the last equality, noticing that the first line is simply an unormalised inverse Wishart distribution of $n_m - n_b - 2 + 2\alpha$ degrees of freedom. Writing $\delta = \mathbf{x}^d - \mathbf{x}^t$, the data likelihood, given $\mathbf{M}$ can be obtained by marginalising over $\mathbf{C}^t$:

$$\begin{aligned} L(\mathbf{x}^d|\mathbf{M}) &\propto \int d\mathbf{C}^t \left|\mathbf{C}^t\right|^{-\frac{1}{2}} e^{-\frac{1}{2}\delta^T\mathbf{C}^{t^{-1}}\delta} p(\mathbf{C}^t|\mathbf{M}) && (6.89)\\ &\propto \int d\mathbf{C}^t \left|\mathbf{C}^t\right|^{-\frac{1}{2}} e^{-\frac{1}{2}\delta^T\mathbf{C}^{t^{-1}}\delta} \left|\mathbf{C}^t\right|^{-\frac{n_m-1}{2}-\alpha} e^{-\frac{1}{2}\mathrm{Tr}\left(\mathbf{C}^{t^{-1}}\mathbf{M}\right)} && (6.90)\\ &\propto \int d\mathbf{C}^t \left|\mathbf{C}^t\right|^{-\frac{n_m}{2}-\alpha} e^{-\frac{1}{2}\mathrm{Tr}\left(\mathbf{C}^{t^{-1}}\left(\mathbf{M}+\delta\delta^T\right)\right)} && (6.91)\\ &\propto \left|\mathbf{M} + \delta\delta^T\right|^{-\frac{n_m-n_b-1+2\alpha}{2}} && (6.92) \end{aligned}$$

where the last line follows from noticing the previous line boils down to integrating an inverse Wishart distribution of $n_m - n_b - 1 + 2\alpha$ degrees of freedom. Applying the matrix determinant lemma, $\left|\mathbf{M} + \delta\delta^T\right| = |\mathbf{M}|\left(1 + \delta^T\mathbf{M}^{-1}\delta\right)$, we finally obtain a t-student distribution of $n_m - 2n_b - 1 + 2\alpha$ degrees of freedom:

$$L(\mathbf{x}^d|\mathbf{M}) = \frac{c_b\,|\mathbf{M}|^{-\frac{1}{2}}}{(1 + \delta^T\mathbf{M}^{-1}\delta)^{\frac{n_m-n_b-1+2\alpha}{2}}}, c_b = \frac{\Gamma\left(\frac{n_m-n_b-1+2\alpha}{2}\right)}{\pi^{\frac{n_b}{2}}\Gamma\left(\frac{n_m-2n_b-1+2\alpha}{2}\right)} \tag{6.93}$$

where we reintroduced normalisation factors. This Bayesian formulation is appealing because it does not require correcting the posterior with the numerical factor m_1 of Eq. (6.61), inferred using a Gaussian approximation of the true posterior. Also, the combination of independent likelihoods is straightforward, through a simple product, without the need to specify where the Fisher information comes from, as in Section 6.2.2.

We have found this approach to work correctly (according to the criterion described below) for $n_b = 1$, which is the test case of Sellentin and Heavens (2016). In higher dimensions $n_b > 2$, their approach appear to fail, as can be seen in Figure 6.4. For this figure, we repeat 5000

times the experiment which consists in (i) generating $n_m = 200$ mocks following a multivariate Gaussian of dimension $n_b = 128$, (ii) calculating the corresponding mock covariance matrix $\mathbf{C}$, (iii) fitting a one-parameter linear model to another Gaussian mock, using the previous covariance matrix, (iv) calculating the posterior cumulative distribution function $F(t)$, i.e. the integral of the posterior from $-\infty$ to the true value of the model parameter $t = 0$. Then, in Figure 6.4, we report, for each value $cdf \in [0, 1]$, the difference with respect to cdf of the rate of $F(t) < cdf$, with $F(t)$ determined for each of the 5000 realisations. Our criterion (the so-called *frequentist matching*) for the posterior (and hence the measurement error bar) to be well predicted is that the obtained curve should be flat, equal to zero. By definition, taking the true covariance matrix $\mathbf{C}^t$ (instead of $\mathbf{C}$ estimated from the 200 mocks) for the fits yields well predicted posteriors (black). Taking $\mathbf{C}$ instead (red), $F(t)$ is underpredicted (compared to the rate of $F(t) < cdf$) when $cdf < 0.5$, i.e. error bars are underestimated. Applying the Hartlap factor Eq. (6.47) yields better results (orange), but still not quite correct. The Percival et al. (2014) correction (green) provides very good posteriors. Instead, the Sellentin and Heavens (2016) correction (deep blue) fails, at the level of the Harlap correction alone (orange).

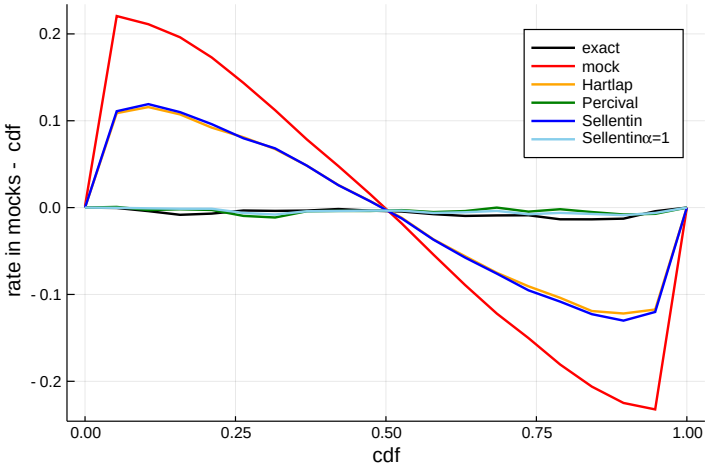

Figure 6.4 – For each value $cdf \in [0, 1]$, difference with respect to cdf of the rate of $F(t) < cdf$, with $F(t)$ the value of the posterior cumulative distribution function at the truth t ($= 0$ in our test case) for each of the 5000 realisations of the experiment (see text), with $n_b = 128$ and $n_m = 200$. We show the resulting curve when considering the exact covariance matrix $\mathbf{C}^t$, known from theory (black), using directly the covariance matrix $\mathbf{C}$ estimated from mocks (red), applying the Hartlap factor (Eq. (6.47), orange), the Percival et al. (2014) correction m_1 (Eq. (6.61), green), the Sellentin and Heavens (2016) prescription (Eq. (6.93) with $\alpha = (n_b + 1)/2$, deep blue), and Eq. (6.93) with $\alpha = 1$ (light blue).

We attribute the failure of the Sellentin and Heavens (2016) prescription to the somewhat unmotivated Jeffreys prior of Eq. (6.85). We therefore tried keeping $\alpha = 1$ as is the case when $n_b = 1$. In this case, the posteriors are very well predicted, as can be seen in Figure 6.4 (light blue).

Note that Eq. (6.88) is normalisable only if the number of degrees of freedom is $n_m - n_b - 2 + 2\alpha > n_b - 1$ i.e. $\alpha > (2n_b - n_m + 1)/2$. This is necessarily the case if $\alpha = (n_b + 1)/2$ as in Sellentin and Heavens (2016). However, setting $\alpha = 1$ yields the constraint $n_m > 2n_b - 1$. Actually, Eq. (6.88) can be considered as an improper prior for Eq. (6.91), and hence does not really have to be normalisable, if Eq. (6.91) indeed converges. Here, the condition is slightly relaxed: $n_m - n_b - 1 + 2\alpha > n_b - 1$; for $\alpha = 1$, $n_m > 2n_b - 2$. In the case of a linear model

samples	redshift range	targets [deg^{-2}]	good redshifts
LRG	$0.4 - 1.0$	350	4.0 M
ELG	$0.6 - 1.6$	2400	17.1 M
QSO (tracers)	< 2.1	170	1.7 M
QSO (Lyman-α)	> 2.1	90	0.7 M
total in dark/grey time		3010	23.6 M
bright galaxies	$0.05 - 0.4$	700	9.8 M

Table 6.2 – Summary statistics of the DESI samples. Extracted from DESI Collaboration et al. (2016).

$\mathbf{x}^t(\mathbf{p})$, Eq. (6.93), together with flat priors on the model parameters $\mathbf{p}$, would give a t-student distribution in the model parameters, of $n_m - n_b - n_p - 1 + 2\alpha$ degrees of freedom, i.e. the condition becomes, for $\alpha = 1$, $n_m > n_b + n_p - 2$. It is still not clear whether these conditions are meaningful. Another route to choose a prior for $\mathbf{C}^t$ that ensures frequentist matching is right Haar priors (Berger et al., 2015), but they do not allow analytical integration over the covariance matrix as in Eq. (6.91).

Of course, to mitigate the impact of the limited number of mocks on the final error bars (and their accurate prediction), a simple idea is to decrease n_b, i.e. compress the data vector into a shorter form. This is however not lossless if the model has a non-linear dependence in the parameters. To avoid this drawback, another idea is to keep the data vector the same size, but to reduce the scatter in the covariance matrix, e.g. by resorting to a parametrised model covariance, to be constrained with mocks. We have started to explore further that possibility (McDonald et al., 2020), which is beyond the scope of this manuscript.

6.3 DESI prospects

The 20-year history of SDSS spectroscopic galaxy surveys has just terminated with the eBOSS final data release, in this day of July 20, 2020. DESI will start spectroscopic operations in a few months. After presenting this new survey (Section 6.3.1), which will be 10 times larger than previously gathered SDSS samples, we emphasise on its constraining power (Section 6.3.2) and the subsequent progress to be made on controlling both theoretical and observational systematics (Section 6.3.3). To tackle the latter challenge we finally present a method which consists in forward modelling the galaxy density field (Section 6.3.4).

6.3.1 Presentation

DESI (DESI Collaboration et al., 2016) will measure spectra during approximately five years, using the 4 m Mayall telescope located at Kitt Peak, Arizona, with an instrumented field of view of 7.5 deg^2. The footprint area is 14 000 deg^2, split between NGC and SGC. Given DESI field of view, the footprint can be covered with 2000 tiles, which we call a layer.

DESI will target four classes of objects: bright galaxies, LRGs, ELGs and QSOs. QSO spectra will be used as tracers of the matter density field or of the neutral hydrogen density through the Lyman-α forest. The summary statistics for the four samples is provided in Table 6.2.

LRGs, ELGs and QSOs will be observed in dark (and grey) time. Five tiling layers will be achieved during DESI operations (1940 h per year), with an average exposure time of 1800 s (DESI Collaboration et al., 2016). The moon will be below the horizon during 80% of spectroscopic

data	FoM	a_p	$\sigma(w_p)$	$\sigma(\Omega_{K,0})$
galaxy BAO	133	0.69	0.023	0.0013
galaxy and Lyman-α forest BAO	169	0.71	0.022	0.0011
BAO + gal. broadband to $k < 0.1\,h\,\mathrm{Mpc}^{-1}$	332	0.74	0.015	0.0009
BAO + gal. broadband to $k < 0.2\,h\,\mathrm{Mpc}^{-1}$	704	0.73	0.011	0.0007

Table 6.3 – DESI forecasts for the DEFT (Albrecht et al., 2006) Figure-of-Merit (FoM), the error $\sigma(w_p)$ on $w_p = w(a_p)$ at the pivot scale factor a_p and the error $\sigma(\Omega_{K,0})$ on the curvature energy density $\Omega_{K,0}$ (fiducial values: $w_0 = -1$, $w_a = 0$ and $\Omega_{K,0} = 0$), at the 68% confidence level. All lines contain BGS, BOSS in $0.45 < z < 0.6$, DESI LRGs, ELGs and QSOs and Planck Collaboration et al. (2014) data. In the two last lines, BAO constraints include Lyman-α forest. Extracted from DESI Collaboration et al. (2016).

observations. In this dark time, LRG and QSO targets will be targeted in priority. Conversely, grey time will be mainly dedicated to measuring ELG spectra, which are less sensitive to variations in the sky subtraction due to the strong emission lines. The DESI LRG sample will contain 4 M galaxy redshifts, 3 times more than the combined BOSS and eBOSS LRG samples, of total size 1.4 M redshifts (see Section 5.6.1). The DESI QSO tracer sample will contain 1.7 M redshifts, 5 times more than the SDSS equivalent sample, and 0.7 M Lyman-α QSOs, 3 times more than SDSS. However, the largest DESI sample is by far ELGs, with 17.1 M redshifts: this is 100 times the size of the eBOSS ELG sample.

Sky will be too bright during 250 h per year to observe ELGs. This observing time will be dedicated in priority to the Bright Galaxy Survey (BGS). Remaining fibres will be allocated to old stars in our galaxies, as part of the Milky Way Survey (MWS). Three layers of the dark time tiling will be dedicated to these two surveys. As a result, BGS will provide redshifts for 10 M galaxies, in the range $0.05 < z < 0.4$. This sample will be 10 times larger than the SDSS Main Galaxy Sample (see Section 5.6.1).

6.3.2 Forecasts

In Figure 6.5 we add DESI Fisher forecast error bars, provided in DESI Collaboration et al. (2016), on top of the SDSS RSD and BAO clustering measurements shown in Figure 5.40. In the optimistic scenario, using a maximum wavenumber $k_{\mathrm{max}} = 0.2\,h\,\mathrm{Mpc}^{-1}$ in the fits, the forecast relative uncertainty in $f\sigma_8$, $D_{\mathrm{H}}/r_{\mathrm{drag}}$ and $D_{\mathrm{M}}/r_{\mathrm{drag}}$ is $\simeq 2\%$ in 18 $\Delta z = 0.1$ redshift slices in the redshift range $0.05 < z < 1.85$. Low redshift measurements are provided by the BGS sample, measurements above $z > 0.6$ by the LRGs, ELGs and QSOs, and high redshift BAO measurements by Lyman-α forests (auto- and cross-correlation with QSO tracers).

DESI cosmological forecasts are extensively presented in DESI Collaboration et al. (2016). In Table 6.3 we report the expected Figure-of-Merit of the DEFT (Albrecht et al., 2006). The error on the dark energy equation of state is quoted at the pivot scale factor a_p where the $w_0 - w_a$ covariance matrix is diagonal (see Section 5.6.4). Even with BAO measurements alone, the FoM is 133, more than the 110 threshold value adopted to define a Stage-IV experiment. Noticeably, using the full shape (broadband) of the power spectrum yields a great improvement in the FoM, up to 704 if the analysis is pushed to scales up to $0.2\,h\,\mathrm{Mpc}^{-1}$.

Apart from constraints on the dark energy EoS, DESI will also help probe the inflation. First, in combination with Planck, DESI will exquisitely measure n_s and the running α_s with a precision of 0.19% (marginalising over the sum of neutrino masses), an improvement of 1.7 and

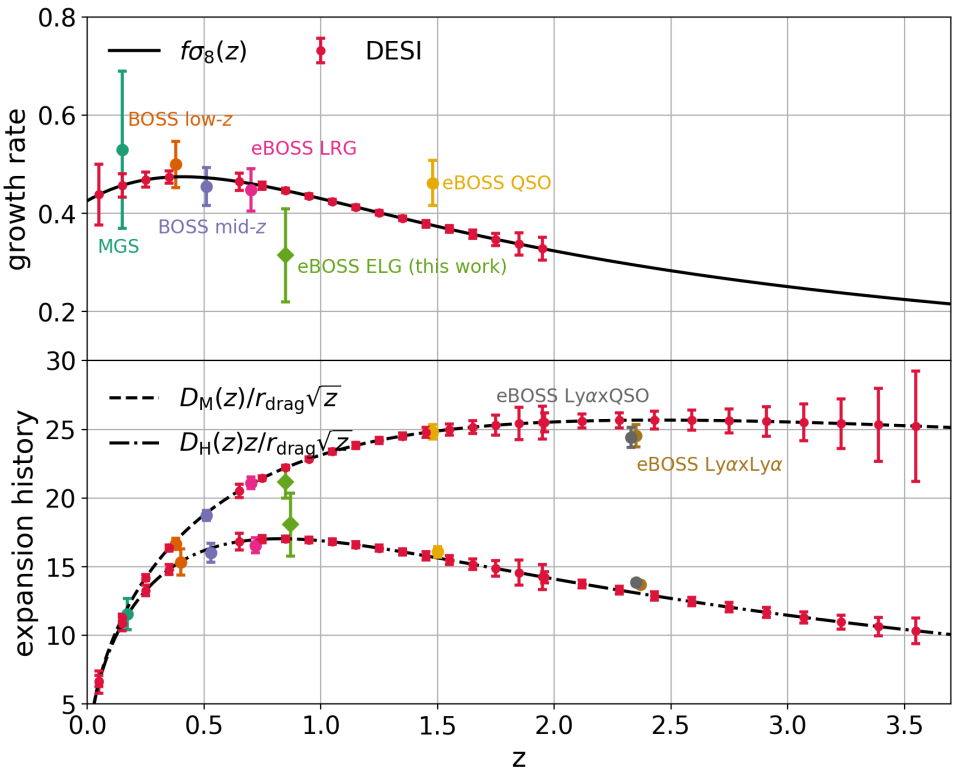

Figure 6.5 – SDSS growth rate and BAO measurements of Figure 5.40, with DESI forecast error bars, provided in DESI Collaboration et al. (2016) (with a survey area of 14 000 deg^2 and $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$).

2.7 over Planck Collaboration et al. (2014) alone. Constraints come both from the broadband power spectrum up to $k = 0.2\,h\,{\rm Mpc}^{-1}$ and Lyman-α forests. Good constraints on n_s help shrink contours in the $n_s - r$ plane, as displayed in Figure 1.8, thereby providing constraints for the amount of gravitational waves generated by inflation. As explained in Section 1.2.5, n_s is also related to the number of e-folds of expansion, and measuring a large value ($> 10^{-3}$) of α_s would exclude the standard single-field slow-roll models of inflation. Also, DESI alone will put 1σ errors on $f_{\rm NL}^{\rm local}$ of $\sigma(f_{\rm NL}^{\rm local}) = 5$, similar to Planck Collaboration et al. (2014) alone (see Section 1.2.5), which will be divided by 2 when combining both data sets. Finally, DESI will help put constraints on the sum of neutrino masses, as summarised in Table 6.4. Including Planck and DESI data sets, the uncertainty on the sum of neutrino masses may reach 0.020 eV. In the case neutrino masses are minimal in the normal hierarchy, the inverted hierarchy could be excluded at the 3σ level. Noticeably, most constraints are expected to come from the galaxy broadband power spectrum.

6.3.3 Challenges

As seen in the previous forecasts, the galaxy broadband power spectrum will play a key role in all DESI cosmological constraints. Namely, the measurement of the spectral index and its running, or the study of neutrino masses requires a fine understanding of the shape of the power spectrum at $k \geq 0.1\,h\,{\rm Mpc}^{-1}$, hence a well-controlled galaxy bias model or galaxy - dark matter halo connection (see Section 2.3), including galaxy velocities. The state-of-the-art eBOSS modelling systematic budget, for example, is already at the order of 2% for $f\sigma_8$ (see Section 5.2, and Smith et al. 2020; Gil-Marín et al. 2020; Bautista et al. 2020), which is

data	$\sigma(\sum m_\nu)$ [eV]
Planck	0.56
Planck + BAO	0.087
Planck + BAO + gal. broadband to $k < 0.1\,h\,\mathrm{Mpc}^{-1}$	0.030
Planck + BAO + gal. broadband to $k < 0.2\,h\,\mathrm{Mpc}^{-1}$	0.021
Planck + BAO + Lyman-α forest	0.041
Planck + BAO + gal. broadband to $k < 0.2\,h\,\mathrm{Mpc}^{-1}$ + Lyman-α forest	0.020

Table 6.4 – DESI forecast 68% uncertainty for the sum of neutrino masses (fiducial value: $\sum m_\nu = 0.06\,\mathrm{eV}$). BAO measurements include DESI galaxies, quasars and Lyman-α forests. CMB constraints are from Planck Collaboration et al. (2014). Extracted from DESI Collaboration et al. (2016).

equal to the typical DESI expected statistical accuracy for each redshift slice of Figure 6.5. Note that these modelling errors could be relatively easily divided by roughly a factor 2 by varying the template cosmology as part of the cosmological inference (e.g. Chudaykin et al., 2020). Conversely, studies of primordial non-Gaussianity require an accurate modelling of scales $k \leq 0.1\,h\,\mathrm{Mpc}^{-1}$, which includes wide-angle and integral constraint corrections (see Section 4.2.3 and Section 4.3), as well as relativistic effects (e.g. Wang et al., 2020).

Of course, the impact of observational systematics should be well understood too. On small scales, fibre collisions will play an important role. On large scales, the clustering signal will be impacted by variations in the redshift efficiency or quality of the imaging (see Section 5.3.1). Purely angular systematics can be mitigated (see e.g. Section 4.3.7, Tamone et al. 2020 and Shafer and Huterer 2015), but three-dimensional variations of the survey selection function, as seen in Section 5.3.1.6, are much more complex to account for. Again, eBOSS typical systematic uncertainties were of the order of 2% (see Neveux et al., 2020; Hou et al., 2020; Gil-Marín et al., 2020; Bautista et al., 2020) to 9% (see Section 5.5.2 and Tamone et al. 2020), which is too large for the DESI expected accuracy.

One would have also noted in Chapter 5 that the propagation of modelling and observational systematics down to cosmological parameters is a bit cumbersome, requiring many tests with mock catalogues, and inaccurate, neglecting cross-correlations between the different systematic and statistical errors. In the next section we present an analysis method that allows proper marginalisation over these observational systematics.

6.3.4 Forward modelling

The forward modelling approach consists in reconstructing the observed galaxy density from the matter density field initial conditions, as depicted in Figure 6.6. Namely, an initial random matter density field is generated in a cubic box. This field is then evolved forward in time through a gravity solver. The galaxy density field is inferred from the matter density field through a bias model, taking into account redshift space distortions, and is then submitted to the survey selection function. The likelihood of the observed galaxy density field given the simulated one is computed, and its value is used to iterate over, or to sample, initial conditions of the matter field.

In the following we more specifically describe the approach of Lavaux et al. (2019), called BORG (Bayesian Origin Reconstruction from Galaxies), which was applied to the SDSS BOSS galaxy sample.

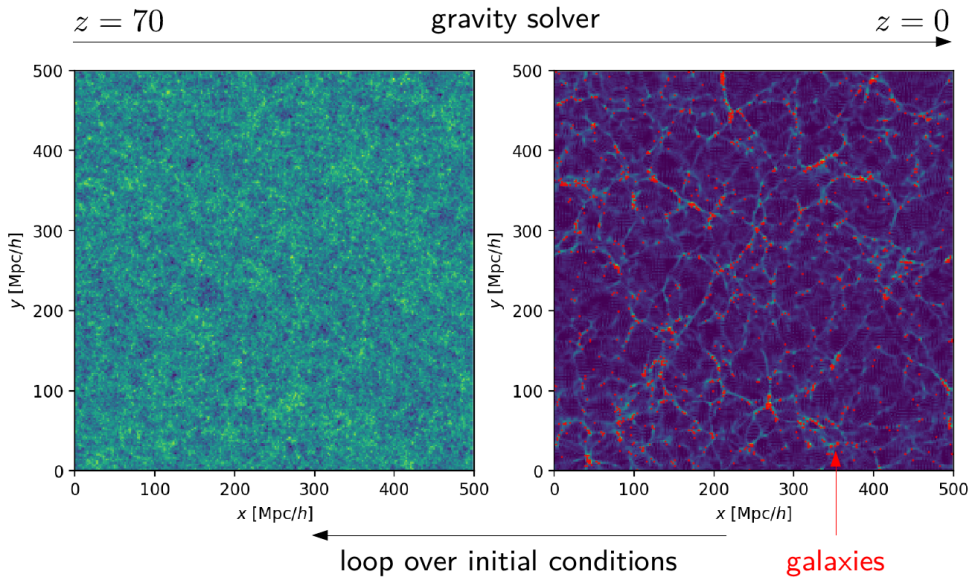

Figure 6.6 – Illustration of the forward modelling approach for reconstructing the galaxy density field (obtained with simulated data).

In Lavaux et al. (2019), the density field is evolved through Lagrangian perturbation theory, as given by Eq. (2.48). The time evolution of structure formation is accounted for by taking the linear growth factor D_+ at the scale factor corresponding to the comoving distance q. This is claimed to be accurate enough at the scales of interest ($16\,\mathrm{Mpc}\,h^{-1}$). However, Jasche and Lavaux (2019) also implemented second order Lagrangian perturbation theory and a particle-mesh solver (see Section 2.2.1.3). In Lavaux et al. (2019), to account for galaxy bias, the model number of galaxies in cell i is taken to be the quadratic form:

$$N_g^{(i)} = \boldsymbol{\Delta}_i^T \mathbf{Q} \boldsymbol{\Delta}_i \tag{6.94}$$

with $\mathbf{Q}$ a positive definite matrix, sampled as part of the Bayesian inference, and $\boldsymbol{\Delta}_i$ the vector containing the density contrast averaged over multiple grid cells (up to 3^3) centred around i at different powers (up to 2). This formula is in effect an expansion over the pure linear bias, taking into account short range non-locality of galaxy bias. Note that more complex bias models have been designed, e.g. using neural networks (Charnock et al., 2020). Then, Lavaux et al. (2019) write the model density of galaxies, including survey selection effects, as:

$$N_{g,\mathrm{sel}}^{(i)} = W_i \prod_{t \in \mathcal{T}} (1 + c_t t)\, N_g^{(i)} \tag{6.95}$$

where W_i is the value of the model survey selection function at cell i, and $\mathcal{T}$ is a set of photometric templates (E(B-V), sky flux, airmass, seeing, stellar density). Coefficients c_t are sampled as part of the Bayesian inference. The set of templates $\mathcal{T}$ may however not be complete, hence Lavaux et al. (2019) further marginalise over the value A_p scaling the survey selection function W in HEALPix pixels (nside = 256), with a Jeffreys prior $\pi(A_p) \propto 1/A_p$, as proposed by Porqueres et al. (2019). Bias, template and pixel A_p parameters are specific to the 4 comoving distance bins used in the Bayesian inference for each of the CMASS and LOWZ NGC and SGC samples.

Initial conditions of the matter field, as well as bias, photometric template and pixel parameters are left free. Their posterior is sampled with the Hamiltonian Markov Chain Monte Carlo algorithm, which requires computation of the adjoint likelihood gradient, as detailed in Jasche

and Lavaux (2019). The result of the Bayesian reconstruction process is thus a sampling of the matter density field and velocities at any point in our past light-cone.

The posterior reconstructed final density field obtained by Lavaux et al. (2019) is shown in Figure 6.7 (left). One can clearly see the separation between the LOWZ and CMASS samples at distance of around 900 Mpc h^{-1}. Outside the BOSS footprint, the mean density field is zero in average, as expected. The filamentary nature of the large scale structure is clearly visible. Also in Figure 6.7 (right) is shown the standard deviation of the galaxy density field. Variance is high in regions of low or zero observed galaxy density, since there is no information to constrain the reconstruction of the density field.

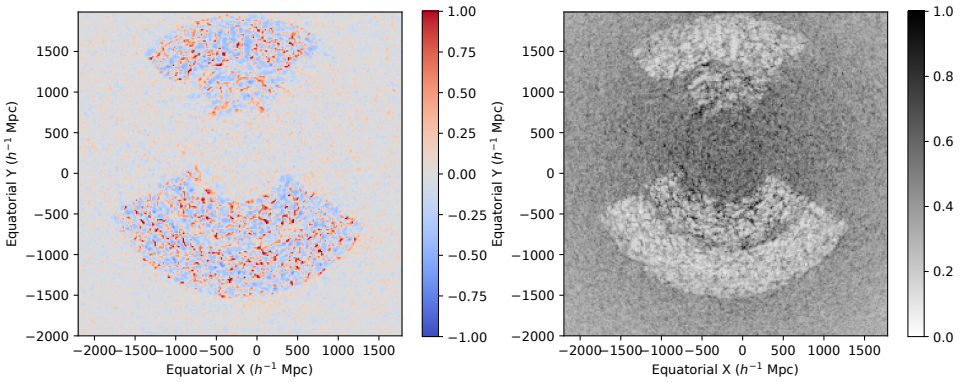

Figure 6.7 – Mean (left) and standard deviation (right) of the reconstructed final matter density field. Taken from Lavaux et al. (2019).

Lavaux et al. (2019) note that BOSS data favours a non-zero contribution from the density field averaged over multiple cells in the bias model, and hence evidence for non-local bias. Concerning observational systematics, they find residual, redshift dependent, variations of the observed galaxy density with seeing (r-band) and skyflux (u-band), as shown in Figure 6.8 (left). In addition to the dependence of galaxy density with these photometric templates, Lavaux et al. (2019) observe roughly iso-Dec. stripe patterns in the pixel maps of posterior A_p coefficients, with relative variations reaching up to 30%, as shown in Figure 6.8 (right).

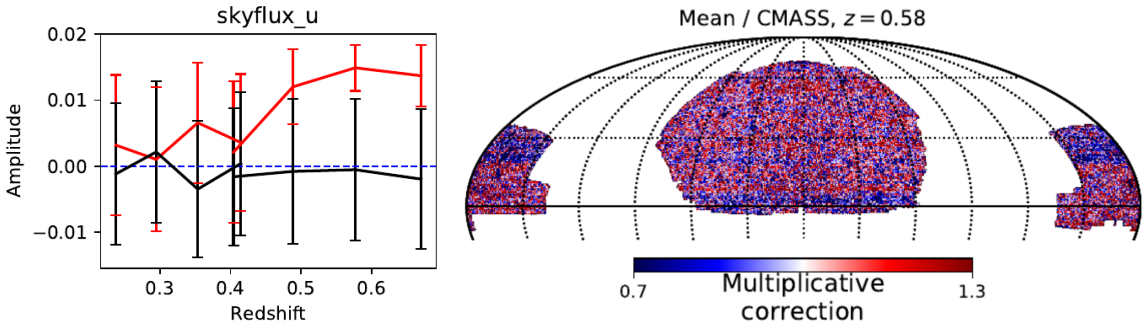

Figure 6.8 – Left: mean and error bar of the coefficient scaling the template of sky flux in u band, in 4 redshift bins for LOWZ and CMASS samples (red: NGC, black: SGC). Right: mean pixel systematic map A_p in the CMASS redshift bin $z = 0.58$. Extracted from Lavaux et al. (2019).

The origin of these systematics remains unclear. In any case, this study shows the large potential of the Bayesian forward modelling approach to unveiling observational systematics in galaxy surveys — with the ultimate purpose of deriving robust cosmological constraints. For

example, Ramanah et al. (2019) ran an Alcock-Paczynski test on simulated data, and found significantly tighter constraints than standard BAO techniques (see Section 5.1.3). The long-term goal is to further extend this framework for complete inference of cosmological parameters.

In addition to the analysis of galaxy redshift surveys, this forward modelling approach may have other interesting cosmological applications. Jasche and Lavaux (2019) showed that they could recover masses of galaxy clusters, in agreement with weak lensing, X-ray or velocity measurements (see Section 1.2.4.2). Lavaux et al. (2019) computed the convergence map of CMB backlight passing through the reconstructed gravitational potential, and found clear correlation with the CMB-infered convergence maps (Planck Collaboration et al., 2018), providing evidence for the potential of galaxy clustering and CMB lensing cross-correlation with the Bayesian forward modelling framework. Similar cross-correlations may be performed with galaxy weak lensing. We also note that the forward modelling approach was applied to the density field reconstruction with (simulated) Lyman-α forests by Porqueres et al. (2019). To summarise, this framework appears extremely powerful to analyse and combine different probes of the large scale structure in the Universe.

Bibliography

Albrecht, A., G. Bernstein, R. Cahn, W. L. Freedman, J. Hewitt, et al. (2006). Report of the Dark Energy Task Force. *arXiv e-prints*, astro–ph/0609591. arXiv:astro-ph/0609591.

Anderson, T. (1958). *An introduction to multivariate statistical analysis*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley.

Bautista, J. E., R. Paviot, M. Vargas Magaña, S. de la Torre, S. Fromenteau, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1. *arXiv e-prints*, arXiv:2007.08993. arXiv:2007.08993.

Berger, J. O., J. M. Bernardo, and D. Sun (2015). Overall Objective Priors. *arXiv e-prints*, arXiv:1504.02689. arXiv:math.ST/1504.02689.

Charnock, T., G. Lavaux, B. D. Wandelt, S. Sarma Boruah, J. Jasche, et al. (2020). Neural physical engines for inferring the halo mass distribution function. MNRAS *494*(1), 50–61. arXiv:1909.06379.

Chudaykin, A., M. M. Ivanov, and M. Simonović (2020). CLASS-PT: non-linear perturbation theory extension of the Boltzmann code CLASS. *arXiv e-prints*, arXiv:2004.10607. arXiv:2004.10607.

DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, et al. (2016). The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, arXiv:1611.00036. arXiv:1611.00036.

Dodelson, S. and M. D. Schneider (2013). The effect of covariance estimator error on cosmological parameter constraints. Phys. Rev. D *88*(6), 063537. arXiv:1304.2593.

Gil-Marín, H., J. E. Bautista, R. Paviot, M. Vargas-Magaña, S. de la Torre, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement

of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0. *arXiv e-prints*, arXiv:2007.08994. arXiv:2007.08994.

Hartlap, J., P. Simon, and P. Schneider (2007). Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. A&A *464*(1), 399–404. arXiv:astro-ph/0608064.

Hou, J., A. G. Sánchez, A. J. Ross, A. Smith, R. Neveux, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from anisotropic clustering analysis of the Quasar Sample in configuration space between redshift 0.8 and 2.2. *arXiv e-prints*, arXiv:2007.08998. arXiv:2007.08998.

Howlett, C. and W. J. Percival (2017). Galaxy two-point covariance matrix estimation for next generation surveys. MNRAS *472*(4), 4935–4952. arXiv:1709.03057.

Jasche, J. and G. Lavaux (2019). Physical Bayesian modelling of the non-linear matter distribution: New insights into the nearby universe. A&A *625*, A64. arXiv:1806.11117.

Lavaux, G., J. Jasche, and F. Leclercq (2019). Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data. *arXiv e-prints*, arXiv:1909.06396. arXiv:1909.06396.

McDonald, P. et al. (2020). Another approach to likelihood analysis with repeated mocks to constrain covariance. *in prep*. arXiv:2012.xxxxx.

Meiksin, A. and M. White (1999). The growth of correlations in the matter power spectrum. MNRAS *308*(4), 1179–1184. arXiv:astro-ph/9812129.

Neveux, R., E. Burtin, A. de Mattia, A. Smith, A. J. Ross, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from the anisotropic power spectrum of the Quasar sample between redshift 0.8 and 2.2. *arXiv e-prints*, arXiv:2007.08999. arXiv:2007.08999.

Percival, W. J., A. J. Ross, A. G. Sánchez, L. Samushia, A. Burden, et al. (2014). The clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: including covariance matrix errors. MNRAS *439*(3), 2531–2541. arXiv:1312.4841.

Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, et al. (2014). Planck 2013 results. XVI. Cosmological parameters. A&A *571*, A16. arXiv:1303.5076.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2018). Planck 2018 results. VIII. Gravitational lensing. *arXiv e-prints*, arXiv:1807.06210. arXiv:1807.06210.

Porqueres, N., J. Jasche, G. Lavaux, and T. Enßlin (2019). Inferring high-redshift large-scale structure dynamics from the Lyman-α forest. A&A *630*, A151. arXiv:1907.02973.

Porqueres, N., D. Kodi Ramanah, J. Jasche, and G. Lavaux (2019). Explicit Bayesian treatment of unknown foreground contaminations in galaxy surveys. A&A *624*, A115. arXiv:1812.05113.

Ramanah, D. K., G. Lavaux, J. Jasche, and B. D. Wand elt (2019). Cosmological inference from Bayesian forward modelling of deep galaxy redshift surveys. A&A *621*, A69. arXiv:1808.07496.

Sellentin, E. and A. F. Heavens (2016). Parameter inference with estimated covariance matrices. MNRAS *456*(1), L132–L136. arXiv:1511.05969.

Shafer, D. L. and D. Huterer (2015). Multiplicative errors in the galaxy power spectrum: self-calibration of unknown photometric systematics for precision cosmology. MNRAS *447*(3), 2961–2969. arXiv:1410.0035.

Smith, A., E. Burtin, J. Hou, R. Neveux, A. J. Ross, et al. (2020). The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: N-body Mock Challenge for the Quasar Sample. *arXiv e-prints*, arXiv:2007.09003. arXiv:2007.09003.

Smith, A., A. de Mattia, E. Burtin, C.-H. Chuang, and C. Zhao (2020). Reducing the Variance of Redshift Space Distortion Measurements from Mock Galaxy Catalogues with Different Lines of Sight. *arXiv e-prints*, arXiv:2007.11417. arXiv:2007.11417.

Tamone, A., A. Raichoor, C. Zhao, A. de Mattia, C. Gorgoni, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Growth rate of structure measurement from anisotropic clustering analysis in configuration space between redshift 0.6 and 1.1 for the Emission Line Galaxy sample. *arXiv e-prints*, arXiv:2007.09009. arXiv:2007.09009.

Taylor, A., B. Joachimi, and T. Kitching (2013). Putting the precision in precision cosmology: How accurate should your data covariance matrix be? MNRAS *432*(3), 1928–1946. arXiv:1212.4359.

Wang, M. S., F. Beutler, and D. Bacon (2020). Impact of Relativistic Effects on the Primordial Non-Gaussianity Signature in the Large-Scale Clustering of Quasars. *arXiv e-prints*, arXiv:2007.01802. arXiv:2007.01802.

Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika 20A*(1-2), 32–52. https://academic.oup.com/biomet/article-pdf/20A/1-2/32/530655/20A-1-2-32.pdf.

Conclusion

On this day of July 20, 2020, the SDSS releases its final clustering measurements. Though the primary focus of this manuscript was on eBOSS ELGs, other samples were analysed following a similar methodology. The theoretical prediction (correlation function or power spectrum) was tested against N-body simulations and an associated systematic budget was derived. Data clustering catalogues were constructed, after selection of reliable data, including weighting schemes to correct for observational systematics. These were implemented in fast approximate simulations, called mocks, to estimate their residual impact on the whole analysis pipeline after correction procedures. These mocks were also used to estimate the covariance matrix of the measured statistics (correlation function or power spectrum) in order to perform the final cosmological inference with the observed data.

We emphasised however that the analysis of the eBOSS ELG sample was rendered peculiarly difficult due to the complex photometric systematics and the relatively low signal over noise. Noting that the radial selection function varied significantly with imaging depth, we had no other choice but estimating these variations from the data itself. We observed that such a procedure significantly biases clustering measurements on large scales, which we interpreted as a radial integral constraint that is common to any SDSS clustering analysis. The modelling and numerical implementation of this effect is an original work of this thesis. Within the same framework, we developed a pixelation scheme to mitigate purely angular photometric systematics, which we also applied to the eBOSS ELG sample. As the posterior of the pure RSD measurement was significantly non-Gaussian, we combined RSD and BAO analyses at the likelihood level. After comparison with mocks, most observational systematics were estimated on the data itself for conservativeness.

During this work, we actively took part to the eBOSS collaboration. In particular, we extensively contributed to the creation of ELG data clustering catalogues. Our implementation of the theoretical RSD model was used for the analysis of other eBOSS samples. We also clarified some details of Fourier space studies. The effect of the radial integral constraint was tested in all analyses, and we provided the tools for practical implementation for the quasar sample. In addition, we significantly improved the implementation of observational systematics in mocks

with procedures which were extended to other samples. This work was presented in numerous conferences.

Together with the analyses themselves, the eBOSS collaboration released the cosmological implications of the SDSS clustering samples. Cosmological distances and growth rate measurements were provided in seven redshift bins, spanning the range between 0.15 and 2.33, which allowed constraints to be put on curvature, the dark energy equation of state and the sum of neutrino masses. In addition, measurements of growth rate are best combined with those of weak lensing to constrain deviations to general relativity. Altogether, SDSS has been a major step forward in understanding the large scale structure and testing the consistency of the standard cosmological model.

Future spectroscopic surveys, such as DESI or the satellite Euclid, will allow the precision of clustering measurements to be increased by an order of magnitude. Apart from the dark energy equation of state, tight constraints are expected for the sum of neutrino masses, the power spectrum scalar index, and primordial non-Gaussianity. These exciting tests of the cosmological model poses stimulating challenges. Focusing on small scale clustering, theoretical predictions will have to be tested and possibly improved further. Impacting all scales, observational systematics will have to be well controlled and thoroughly propagated to the final measurements. Finally, a huge effort has to be put to set up a robust, user-friendly, collaborative analysis pipeline.

As galaxy surveys enter their golden era, many analysis techniques are being developed. Forward modelling the galaxy density field appears as a promising approach to consistently include and constrain observational systematics as part of the cosmological inference. Huge potential is also foreseen in the cross-correlation of galaxy clustering with other probes of the matter density field, such as galaxy or CMB weak lensing. Large scale structure studies have definitely a bright future ahead.

Résumé substantiel

Au cours de cette thèse nous avons effectué l'analyse en spectre de puissance de l'échantillon des galaxies à raie d'émission (ELG) du programme eBOSS de la collaboration SDSS. Plus spécifiquement, dans la distribution spatiale de ces galaxies situées à un décalage spectral entre 0.6 et 1.1, nous avons mesuré la taille des oscillations acoustiques de baryons et le taux de croissance des structures. Une telle étude, se consacrant aux grandes structures de l'univers, s'appelle une analyse de *clustering*. Nous commençons par des éléments de cosmologie dans la Section 4. La modélisation de la variation de densité des galaxies, et plus particulièrement sa statistique à deux points, le spectre de puissance, est discutée dans la Section 5. Nous présentons le relevé photométrique et spectroscopique dans la Section 6. A l'issue de ce relevé, les positions angulaires et les décalages spectraux des galaxies sont compressés sous la forme du spectre de puissance, comme expliqué dans la Section 7. Dans cette section, nous insistons en particulier sur un biais appelé contrainte intégrale, qui a donné lieu à une publication premier auteur. Forts de ces enseignements, nous présentons l'analyse de l'échantillon des ELG d'eBOSS, publiée dans un second article premier auteur, ainsi que les implications cosmologiques de SDSS dans la Section 8. Nous terminons par une présentation du nouveau relevé DESI, et une ouverture sur une nouvelle technique d'analyse dans la Section 9.

4 Le modèle standard de la cosmologie

Dans cette section nous passons rapidement en revue les éléments clés de la cosmologie moderne, qui se révéleront utiles dans l'analyse de l'échantillon des ELG d'eBOSS : les mesures de distance et la dynamique de l'univers à grande échelle (Section 4.1) et les perturbations dans l'univers primordial (Section 4.2).

4.1 L'univers homogène

Puisque l'on s'intéresse à la distribution spatiale des galaxies, il convient de spécifier comment déterminer leur distance. La géométrie de l'espace-temps est décrite par une métrique. Dans un univers homogène et isotrope, comme le postule le principe cosmologique, une longueur infinitésimale ds s'écrit :

$$ds^2 = -dt^2 + a^2(t)\gamma_{ij}dx^i dx^j = a^2(\eta)\left(-d\eta^2 + \gamma_{ij}dx^i dx^j\right) \tag{96}$$

où

$$\gamma_{ij}dx^i dx^j = d\chi^2 + f_K^2(\chi)d\Omega^2. \tag{97}$$

Cette métrique est appelée métrique de Friedmann-Lemaître-Robertson-Walker. Dans ces formules, a est le facteur d'échelle de l'univers, t et η les temps cosmiques et conformes, $x_{i\in[1,3]}$ sont les coordonnées comobiles d'espace, et γ_{ij} la partie spatiale de la métrique. f_K est une fonction de χ et de la courbure K. Dans un univers plat, $K = 0$ et $f_K(\chi) = \chi$, ce que l'on suppose dorénavant. En effet, comme on le verra dans la Section 8.5 les observations actuelles du fond diffus cosmologique (incluant son lentillage par les grandes structures), des distances de supernovae Ia lointaines et des oscillations acoustiques de baryons favorisent un courbure extrêmement faible.

Les spectres des galaxies que l'on observe sont décalés vers le rouge. La longueur d'onde λ_e d'un photon émis à un temps t_e et reçu à un temps t_0 est dilatée par un facteur :

$$1 + z(t_e) = \frac{\lambda_0}{\lambda_e} = \frac{a(t_0)}{a(t_e)} \tag{98}$$

où z_e est appelé décalage spectral. Dans toute la suite, on pose $a(t_0) = 1$. Mesurer le décalage spectral d'une galaxie donne donc accès au facteur d'échelle $a(t_e)$ à l'instant où sa lumière a été émise. Si l'on veut spécifier la distance de la galaxie, il faut s'appuyer sur la métrique (96).

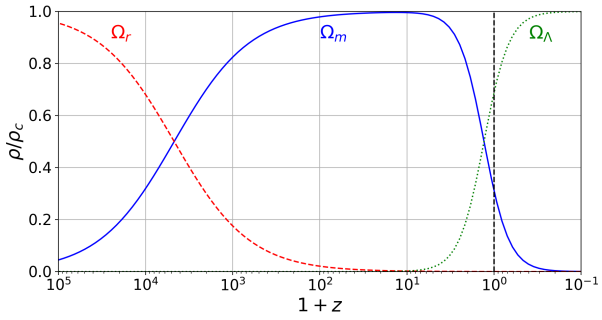

Figure 7.1 – Les paramètres de densité en fonction du décalage spectral: $\Omega_m(z) = \Omega_c(z) + \Omega_b(z)$ et $\Omega_r(z) = \Omega_\gamma(z) + \Omega_\nu(z)$ (les neutrinos sont ici considérés sans masse).

La distance radiale comobile d'un objet émettant à un décalage spectral z_e est l'intégrale selon une géodésique de genre lumière ($ds^2 = 0$) :

$$\chi(z_e) = \int_0^{z_e} \frac{dz}{H(z)} \tag{99}$$

où $H = \dot{a}/a$ est le paramètre de Hubble, $\dot{}$ symbolise une dérivée par rapport au temps cosmique t. On définit aussi la distance angulaire comobile $D_M(z_e)$ comme la grandeur qui relie la surface comobile d'un objet dS_e à un décalage spectral z_e à l'angle solide $d\Omega_0$ vu par l'observateur, $dS_e = D_M(z_e)^2 d\Omega_0$. Partant de l'Eq. (96), il est clair que :

$$D_{\rm M}(z) = f_K(\chi(z)) \tag{100}$$

Pour calculer ces distances, il convient de spécifier $H(z)$. Pour cela, dans le cadre de la relativité générale, les équations d'Einstein s'expriment, en supposant la métrique ci-dessus, sous la forme des équations de Friedmann-Lemaître. La première donne :

$$H^2(z) = H_0^2 \sum_X \Omega_{X,0}(1+z)^{3(1+w_X)} \tag{101}$$

Dans cette équation, $\Omega_{X,0}$ est le rapport de densité d'énergie du composant X sur l'énergie critique, et w_X son équation d'état ($P = w\rho$ avec P la pression et ρ la densité). $H_0 = H(z=0)$ est la constante de Hubble ; on notera aussi $H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

Dans le modèle ΛCDM, le modèle plus simple qui rende compte de la majorité des observations cosmologiques actuelles et que nous utiliserons dans la suite, les composants de l'univers aujourd'hui sont la matière noire froide (CDM, $\Omega_{c,0} \simeq 26\%$ et $w_c = 0$), la matière baryonique ou matière ordinaire ($\Omega_{b,0} \simeq 5\%$ et $w_b \simeq 0$), la radiation (photons et neutrinos si relativistes, $\Omega_r \simeq 0.007\%$ et $w_r = 1/3$) et l'énergie noire ($\Omega_{\Lambda,0} \simeq 69\%$ et $w_\Lambda = -1$). Par ailleurs, on négligera la courbure ($\Omega_{K,0} = 0$ et $w_K = -1/3$). Dans ce modèle, l'évolution des paramètres de densité est montrée en fonction du décalage spectral sur la Figure 7.1.

4.2 Perturbations primordiales

Pour le moment, nous avons décrit un univers homogène et isotrope, ce qui semble une bonne description à des échelles $\gtrsim 200\,\mathrm{Mpc}$, mais n'est évidemment pas vrai à plus petite échelle, puisque l'on observe des grandes structures, e.g. des murs cosmiques, des filaments, des galaxies, des étoiles, et nous-mêmes. On définit le contraste de densité :

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}} \tag{102}$$

avec ρ le champ de density et $\bar{\rho}$ sa valeur moyenne dans l'univers. Dans la suite on suppose le principe ergodique, ce qui nous permet d'identifier $\bar{\rho}$ avec $\langle\rho\rangle$, la moyenne d'ensemble du champ ρ. Par définition, le moment d'ordre 1 du champ (sa moyenne) est nulle, on s'intéressera donc aux statistiques d'ordre supérieur.

Les fluctuations du champ de densité ont été générées dans l'univers primordial, probablement au cours d'une phase d'expansion accélérée appelée inflation, laquelle fournit par ailleurs une explication à la platitude observée de l'univers. A l'issue de cette phase, les fluctuations suivent une statistique quasiment gaussienne. Elles sont donc entièrement caractérisées par leur moment d'ordre 2, la fonction de corrélation à deux points $\xi(\mathbf{s})$:

$$\xi(\mathbf{s}) = \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{s})\rangle \tag{103}$$

ou le spectre de puissance $P(\mathbf{k})$:

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle = (2\pi)^3\,\delta_D^{(3)}(\mathbf{k}+\mathbf{k}')P(\mathbf{k}) \tag{104}$$

où $\mathbf{k}$ est le vecteur d'onde. A l'issue de l'inflation, les fluctuations de densité ont un spectre de puissance $\propto k^{n_s}$, avec $n_s = 0.9665 \pm 0.0038$ l'indice spectral. Se succèdent ensuite des périodes où l'univers est dominé par la radiation, jusqu'à $z_{\rm eq} \simeq 3400$, puis par la matière (jusqu'à $z \simeq 0.3$) (voir Figure 7.1). Les fluctuations de densité de la matière croissent comme le facteur d'échelle $a \propto \mathcal{H}^{-2}$ pour des échelles super-Hubble ($k \ll \mathcal{H} = aH$) ou lorsque l'univers est dominé par la matière, mais restent (quasiment) constantes pour des modes sub-Hubble ($k \gg \mathcal{H}$) dans l'ère de radiation. En conséquence, les modes $k \gg k_{\rm eq} = \mathcal{H}(z_{\rm eq})$, qui sont restés sub-Hubble tout du long de l'ère de la radiation, sont amortis d'un facteur $(k_{\rm eq}/k)^2$. Il en découle que le spectre de puissance de la matière présente la forme générale suivante :

$$P_m^{\rm lin}(k) \propto \begin{cases} k^{n_s} & k \ll k_{\rm eq} \\ k^{n_s-4} & k \gg k_{\rm eq} \end{cases} \tag{105}$$

Pour l'instant nous avons ignoré le couplage de la matière et de la radiation par la diffusion Thomson. En prenant ce terme en compte, on montre qu'à des échelles sub-Hubble mais plus grandes que le parcours moyen des photons, le contraste de densité des photons obéit à une équation d'oscillateur harmonique. Les oscillations produites sont appelées oscillations acoustiques de baryons (BAO). Lorsque la température de l'univers chute, les électrons se recombinent avec les noyaux pour former (essentiellement) de l'hydrogène neutre. La densité électronique et donc le taux de diffusion Thomson chutent rapidement. Les photons sont libérés autour d'un décalage spectral $z_\star$, où la profondeur optique est unité. Ils forment le fond diffus cosmologique (CMB), dont les fluctuations relatives de température sont de l'ordre de 10^{-5} et portent la marque des BAO. Le décalage spectral où se découplent les baryons est appelé $z_{\rm drag}$. Les distributions de baryons et matière noire se rejoignent sous l'effet de la gravité. Les BAO sont alors visibles dans le spectre de puissance des galaxies, qui tracent le champ de matière, comme on le verra dans la Section 5.2. L'échelle BAO correspond à l'horizon sonique au découplage des baryons, $r_{\rm drag}$. Comparer cette échelle standard dans la distribution observée de galaxies à sa valeur attendue permet de contraindre les distances parallèles et transverses à la ligne de visée, comme on le décrira dans la Section 7.2.

5 Le modèle du spectre de puissance

Jusqu'après le découplage, les perturbations de densité sont bien décrites dans un régime linéaire. Néanmoins, celles-ci s'amplifient sous l'effet de la gravité ; il conviendra donc de décrire les perturbations dans un régime (semi) non-linéaire (Section 5.1), en tenant compte du biais des galaxies (Section 5.2) et des distortions dans l'espace des décalages spectraux (Section 5.3).

5.1 Théorie des perturbations

Nous décrirons la matière noire froide et les baryons comme un seul fluide de matière de contraste de densité ρ et de vitesse particulière $\mathbf{u}$, sans collision, en évolution dans un potentiel gravitationnel Φ, dans un univers en expansion. Les deux premiers moments de l'équation de Vlasov donnent les équations de continuité et d'Euler, auxquelles s'ajoute l'équation de Poisson pour la gravité pour former le système suivant :

$$\delta' + \partial_i \left[(1+\delta)\, u_i\right] = 0 \tag{106}$$

$$u_i' + \mathcal{H} u_i + u_j \partial_j u_i = -\partial_i \Phi - \frac{1}{\rho}\partial_j \left(\rho \sigma_{ij}\right) \tag{107}$$

$$\Delta \Phi = \frac{3}{2}\mathcal{H}^2 \Omega_m \delta \tag{108}$$

Dans ces équations, $'$ symbolise une dérivée par rapport au temps conforme η ($d\eta = dt/a$), $\mathcal{H} = aH$, et δ est le contraste de densité de l'Eq. (102). σ_{ij} est le tenseur anisotrope, que l'on peut supposer nul dans le régime semi non-linéaire, avant que les structures ne s'effondrent et virialisent.

En linéarisant les équations précédentes ($\delta, u_i \ll 1$), on obtient :

$$\delta'' + \mathcal{H}\delta' - \frac{3}{2}\mathcal{H}^2 \Omega_m \delta = 0, \tag{109}$$

dont les solutions s'écrivent comme la somme d'un mode croissant $D_+(a)$ et d'un mode décroissant $D_-(a)$: $\delta(\mathbf{k}, a) = C_+(\mathbf{k}) D_+(a) + C_-(\mathbf{k}) D_-(a)$. Notons que dans le cas spécifique d'un univers dominé par la matière $D_+(a) \propto a$ et $D_-(a) \propto a^{-3/2}$. Le taux de croissance des structures est alors défini comme la dérivée logarithmique :

$$f = \frac{d \ln D_+}{d \ln a} \tag{110}$$

Dans le régime linéaire, la vorticité du champ de vitesses est diluée dans l'expansion ; on la suppose donc nulle dans la suite, de sorte que $\mathbf{u}$ est entièrement décrit par sa divergence θ. Toujours dans le régime linéaire, θ est directement relié à δ :

$$\theta = \nabla \cdot \mathbf{u} = -f\mathcal{H}\delta \tag{111}$$

La description précédente est satisfaisante lorsque $\delta, u_i \ll 1$, mais beaucoup d'information se trouve dans des échelles semi non-linéaires, que nous allons explorer à présent.

Pour cela, nous définissons le vecteur $\mathbf{\Psi} = \left(\delta, -\frac{\theta}{f\mathcal{H}}\right)$. Le système d'équations (108) se réécrit :

$$\Psi_a(\mathbf{k}, u) = g_{ab}(u)\phi_b(\mathbf{k}) + \int_0^u dv g_{ab}(u-v) \int \frac{d^3k_1 d^3k_2}{(2\pi)^3} \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_{12}) \gamma_{bcd}(\mathbf{k}_1, \mathbf{k}_2) \Psi_c(\mathbf{k}_1, v) \Psi_d(\mathbf{k}_2, v) \tag{112}$$

avec $u = \ln D_+$. $g_{ab}(u)$ est un propagateur linéaire, faisant évoluer linéairement le champ initial $\phi_b(\mathbf{k})$ dans le temps. $\gamma_{bcd}(\mathbf{k}_1, \mathbf{k}_2)$ est le vertex de l'interaction entre les champs $\Psi_c(\mathbf{k}_1, v)$ et $\Psi_d(\mathbf{k}_2, v)$, pour laquelle $\delta_D^{(3)}(\mathbf{k} - \mathbf{k}_{12})$ (avec $\mathbf{k}_{12} = \mathbf{k}_1 + \mathbf{k}_2$) assure la conservation du vecteur d'onde (imposée par l'homogénéité statistique). $g_{ab}(u-v)$ fait évoluer le champ obtenu dans le temps, depuis son interaction à $v \leq u$ jusqu'au temps présent u. La solution est construite avec un développement perturbatif :

$$\Psi_a(\mathbf{k}, u) = \sum_{n=1}^{+\infty} \Psi_a^{(n)}(\mathbf{k}, u). \tag{113}$$

A partir de ces solutions, qui peuvent être dérivées à l'aide de diagrammes, le spectre de puissance s'exprime selon :

$$P_{ab}(\mathbf{k},u) = \sum_{l=0}^{+\infty} P_{ab}^{(l)}(\mathbf{k},u) \tag{114}$$

où le terme à l boucles $P_{ab}^{(l)}(\mathbf{k},u)$ est donné par :

$$(2\pi)^3\,\delta_D^{(3)}(\mathbf{k}+\mathbf{k}')P_{ab}^{(l)}(\mathbf{k},u) = \sum_{m=1}^{2l+1}\left\langle \Psi_a^{(m)}(\mathbf{k},u)\Psi_b^{(2l+2-m)}(\mathbf{k}',u)\right\rangle . \tag{115}$$

Celui-ci corrèle $2l+2$ champs initiaux — les corrélations d'un nombre impair de champs initiaux étant nulles comme ces derniers sont supposés suivre une statistique gaussienne.

Cependant, cette théorie des perturbations standard n'a pas de bonnes propriétés de convergence : les contributions d'ordre croissant se compensent partiellement et décroissent lentement. Des techniques de resommation on été développées. Elles s'appuient sur une généralisation du propagateur linéaire g_{ab} au cas non-linéaire (Crocce and Scoccimarro, 2006) Le propagateur non-linéaire à plusieurs points $\Gamma_{ac_1\cdots c_p}$ est défini par Bernardeau et al. (2008):

$$\frac{1}{p!}\left\langle \frac{\delta^p\Psi_a(\mathbf{k},u_f)}{\delta\phi_{c_1}(\mathbf{k}_1)\cdots\delta\phi_{c_p}(\mathbf{k}_p)}\right\rangle = \frac{\delta_D^{(3)}(\mathbf{k}-\mathbf{k}_1-\cdots\mathbf{k}_p)}{(2\pi)^{3(p-1)}}\Gamma_{ac_1\cdots c_p}(\mathbf{k}_1\cdots\mathbf{k}_p,u_f,u_i). \tag{116}$$

$\Gamma_{ac_1\cdots c_p}(\mathbf{k}_1\cdots\mathbf{k}_p,u_f,u_i)$ relie le champ non-linéaire $\Psi_a(\mathbf{k},u_f)$ aux conditions initiales $\phi_{c_1}(\mathbf{k}_1)\cdots\phi_{c_p}(\mathbf{k}_p)$ à u_i. La condition $\delta_D^{(3)}(\mathbf{k}-\mathbf{k}_1-\cdots\mathbf{k}_p)$ découle de l'homogénéité statistique. Le spectre de puissance non-linéaire s'exprime alors simplement :

$$P_{ab}(k,u_f) = \sum_{p=1}^{+\infty} p!\int \frac{d^3q_1\cdots d^3q_p}{(2\pi)^{3(p-1)}}\delta_D^{(3)}(\mathbf{k}-\mathbf{q}_1-\cdots\mathbf{q}_p)\Gamma_{a,(p)}(\mathbf{q}_1\cdots\mathbf{q}_p,u_f,u_i) \\ \Gamma_{b,(p)}(\mathbf{q}_1\cdots\mathbf{q}_p,u_f,u_i)P_0(q_1)\cdots P_0(q_p) \tag{117}$$

$P_0(q_p)$ est le spectre de puissance des conditions initiales des champs, $\langle\phi_a(\mathbf{k})\phi_b(\mathbf{k}')\rangle = (2\pi)^3\,\delta_D^{(3)}(\mathbf{k}+\mathbf{k}')\varepsilon_a\varepsilon_b P_0(\mathbf{k})$, et $\Gamma_{a,(p)} = \Gamma_{ac_1\cdots c_p}\varepsilon_{c_1}\cdots\varepsilon_{c_p}$. Les propagateurs non-linéaires sont calculés à bas k (grandes échelles) par un développement perturbatif. A grand k (petites échelles), les diagrammes dominants à chaque ordre peuvent être resommés de manière exacte, de sorte que le propagateur est exponentiellement amorti (Crocce and Scoccimarro, 2006; Bernardeau et al., 2008). La théorie des perturbations régularisées (RegPT, Taruya et al. 2012) est une technique d'interpolation entre ces deux limites. Elle permet de prédire le spectre de puissance de la matière avec une précision de 1% jusqu'à $k = 0.23\,h\,\mathrm{Mpc}^{-1}$ à $z=1$, alors que pour la même précision le régime linéaire s'arrête à $k = 0.1\,h\,\mathrm{Mpc}^{-1}$.

Bien qu'un programme en `Fortran` fût disponible [1], nous avons recodé la prescription RegPT en `C`, enrobé de `Python`, permettant la parallélisation des calculs et un ajustement des limites d'intégration, dans un module appelé `pyregpt` [2]. Nous avons implémenté dans le même module le calcul des termes de bias et de RSD, que nous allons décrire à présent.

5.2 Biais des galaxies

Notons que l'on n'observe pas directement la matière, mais des objets émettant ou absorbant de la lumière : galaxies, quasars, gaz, etc. Nous faisons l'hypothèse clé que le contraste de densité δ_g de ces objets peut être relié à celui de la matière δ. Nous adoptons une approche perturbative : la relation entre δ_g et δ est reliée par des coefficients de biais. Pour cela, nous suivons le développement de McDonald and Roy (2009). δ_g peut bien sûr dépendre de δ et θ. En sus, une dépendance dans le chemin parcouru par la

1. http://www2.yukawa.kyoto-u.ac.jp/~atsushi.taruya/regpt_code.html
2. https://github.com/ademattia/pyregpt

matière est introduite avec $\mathbf{u}$ et Φ. Un changement global de la vitesse, du potentiel gravitationnel et de sa dérivée (le champ) ne devant pas être observable, on introduit donc une dépendance de δ_g avec $\partial_i u_j$ et $\partial_i \partial_j \Phi$. δ_g est alors développé à l'ordre 3 en δ en fonction des variables ci-dessus, et en incluant un terme stochastique ϵ. Les termes du spectre de puissance de δ_g qui montrent une dépendance similaire en k sont regroupés. A l'issue de cette procédure, le spectre de puissance des galaxies s'exprime selon :

$$P_{gg}(k) = b_1^2 P_{\delta\delta}(k) + 2b_1 b_2 P_{b2,\delta} + 2b_1 b_{s2} P_{bs2,\delta} + 2b_1 b_{3\mathrm{nl}} \sigma_3^2(k) P_m^{\mathrm{lin}}(k) + b_2^2 P_{b22}(k) + 2b_2 b_{s2} P_{b2s2}(k) + b_{s2}^2 P_{bs22}(k) + N_g \tag{118}$$

où $P_{\delta\delta}(k)$ est le spectre de puissance de la matière, $P_{b2,\delta}$, $P_{bs2,\delta}$, $\sigma_3^2(k)P_m^{\mathrm{lin}}(k)$, $P_{b22}(k)$, $P_{b2s2}(k)$, $P_{bs22}(k)$ sont des termes de biais dépendant du spectre linéaire $P_m^{\mathrm{lin}}(k)$. b_1, b_2, b_{s2}, $b_{3\mathrm{nl}}$ et N_g sont des coefficients de biais. De la même manière, en supposant que les galaxies n'ont pas de biais de vitesse par rapport à la matière, on obtient le spectre de puissance croisé entre les galaxies et θ :

$$P_{g\theta}(k) = b_1 P_{\delta\theta}(k) + b_2 P_{b2,\theta} + b_{s2} P_{bs2,\theta} + b_{3\mathrm{nl}} \sigma_3^2(k) P_m^{\mathrm{lin}}(k) \tag{119}$$

où $P_{\delta\theta}(k)$ est le spectre de puissance croisé de la matière et θ, $P_{b2,\theta}$, $P_{bs2,\theta}$ et $\sigma_3^2(k)P_m^{\mathrm{lin}}(k)$ sont des termes de biais dépendant du spectre linéaire $P_m^{\mathrm{lin}}(k)$. En pratique, et c'est ce que nous faisons dans cette thèse, les coefficients b_{s2} et $b_{3\mathrm{nl}}$ peuvent être fixés avec b_1 en supposant que les galaxies sont un fluide couplé à la matière par la gravité, et que le biais est initialement local dans l'espace lagrangien (Chan et al., 2012; Baldauf et al., 2012; Saito et al., 2014).

5.3 Distorsions dans l'espace des décalages spectraux

Nous avons à présent une description du spectre de puissance des galaxies dans l'espace réel, c'est-à-dire de leur position réelle dans l'espace. En réalité, le décalage spectral mesuré des galaxies a (principalement) deux composantes : l'expansion de l'univers, et leur vitesse particulière. En transformant le décalage spectral des galaxies directement en distance (avec l'Eq. (99)), leur distribution des galaxies apparaît déformée dans l'espace dit des décalages spectraux (RSD).

La position dans l'espace des décalages spectraux est reliée à celle dans l'espace réel par :

$$\mathbf{s} = \mathbf{r} - f v_z \hat{\mathbf{z}} \tag{120}$$

avec $v_z = -\mathbf{u} \cdot \hat{\mathbf{z}} / (f\mathcal{H})$, où $\mathbf{u}$ est la vitesse particulière, f le taux de croissance des structures donné par l'Eq. (110) et $\hat{\mathbf{z}}$ est la ligne de visée. Soit δ_s le contraste de densité dans l'espace des décalages spectraux, la conservation de la matière implique que $[1 + \delta_s(\mathbf{s})]\, d^3 s = [1 + \delta(\mathbf{r})]\, d^3 r$. Ainsi, le spectre de puissance dans l'espace des décalages spectraux s'exprime selon :

$$P_s(\mathbf{k}) = \int d^3 x e^{-i\mathbf{k}\cdot\mathbf{x}} \left\langle e^{-ifk\mu\Delta v_z} \left[\delta(\mathbf{r}) + f\partial_z v_z(\mathbf{r})\right] \left[\delta(\mathbf{r}+\mathbf{x}) + f\partial_z v_z(\mathbf{r}+\mathbf{x})\right] \right\rangle \tag{121}$$

avec $\Delta v_z = v_z(\mathbf{r}) - v_z(\mathbf{r}+\mathbf{x})$. Cette expression révèle deux effets opposés :

1. l'effet Kaiser (Kaiser, 1987) : à l'ordre linéaire, où les vitesses sont cohérentes, le terme entre crochets corrèle $\left(1 + \mu^2 f\right)\delta$; ainsi, à grande échelle le spectre de puissance est plus grand que celui dans l'espace réel. Cet accroissement dépend de $f \simeq \Omega_m^{\gamma \simeq 0.55}$ (Linder, 2005), de sorte que les RSD permettent de sonder le contenu énergétique de l'univers (avec Ω_m) ou tester la relativité générale (avec γ).
2. l'effet Finger-of-God (Jackson, 1972): à petite échelle, due à la virialisation, les vitesses se décorrèlent du champ de densité ; le terme entre accolades $e^{-ifk\mu\Delta v_z}$ amortit le spectre de puissance. Cela se passe typiquement lorsque $k \gtrsim 1/\sigma_v$, avec $\sigma_v^2 = \frac{f^2}{3} \int \frac{d^3 q}{(2\pi)^3} \frac{P_m^{\mathrm{lin}}(q)}{q^2}$.

Le modèle de RSD est finalement donné par la formule :

$$P_s(k,\mu) = D_{\rm FoG}(k,\mu,\sigma_v)\left[P_{\delta\delta}(k) + 2f\mu^2 P_{\delta\theta}(k) + f^2\mu^4 P_{\theta\theta}(k) + A(k,\mu,f) + B(k,\mu,f)\right], \quad (122)$$

où le terme d'amortissement $D_{\rm FoG}(k,\mu,\sigma_v)$ vient de l'effet Finger-of-God, $P_{\delta\delta}(k)+2f\mu^2P_{\delta\theta}(k)+f^2\mu^4P_{\theta\theta}(k)$ correspond à l'effet Kaiser, et les terms $A(k,\mu,f)$ et $B(k,\mu,f)$ proviennent du couplage entre ces deux effets. Le biais des galaxies est enfin pris en compte en remplaçant $P_{\delta\delta}(k)$ et $P_{\delta\theta}(k)$ dans l'Eq. (122) par l'Eq. (118) et l'Eq. (119), respectivement, et en le factorisant des termes A et B en remplaçant f par $\beta = f/b_1$.

En résumé, les spectres de puissance $P_{\delta\delta}$ et $P_{\delta\theta}$ dans l'Eq. (118) et l'Eq. (119) sont calculés à 2 boucles avec la prescription RegPT, ainsi que les termes de RSD A et B. Nous adoptons une forme lorentzienne pour le terme de Finger-of-God (Cole et al., 1995), $D_{\rm FoG}(k,\mu,\sigma_v) = \left[1+(k\mu\sigma_v)^2/2\right]^{-2}$. La dispersion de vitesse σ_v ainsi que les coefficients de biais b_1, b_2 et N_g sont laissés libres lors de l'ajustement du modèle aux données. Notons qu'à l'ordre linéaire b_1 multiplie le spectre de puissance, et est donc complètement dégénéré avec sa normalisation σ_8. La même remarque s'applique à f, et, dans une moindre mesure, à b_2. En conséquence, nous renseignerons les combinaisons $f\sigma_8$, $b_1\sigma_8$ et $b_2\sigma_8$.

Rappelons enfin que les calculs de théorie des perturbations (spectre de puissance et termes A et B à 2 boucles, termes de biais à 1 boucle) sont réalisés avec notre module `pyregpt` [3]. Le spectre de puissance linéaire nécessaire à ces calculs est obtenu avec le code de Boltzmann CLASS (Blas et al., 2011).

Ce modèle théorique du spectre de puissance ne peut être comparé directement au spectre observé. Des effets de géométrie seront à prendre en compte, comme nous le verrons en Section 7.2. Cherchons auparavant à comprendre en quoi consiste un relevé spectroscopique de galaxies.

6 Relevé de galaxies

Les relevés de galaxies du Sloan Digital Sky Survey (SDSS, York et al., 2000), dont fait partie eBOSS, procèdent en deux étapes : un relevé photométrique (Section 6.2), où des images du ciel sont prises afin de sélectionner des galaxies cibles, pour en mesurer ensuite le spectre au cours du relevé spectroscopique (Section 6.3). Commençons par présenter le dispositif spectroscopique du SDSS.

6.1 L'instrument SDSS et BOSS

Le SDSS a été conçu au milieu des années 1980 comme un relevé photométrique multi-bande et un relevé spectroscopiques, dans le but de comprendre les grandes structures de l'univers en mesurant les décalages spectraux de 10^6 galaxies et 10^5 quasars sur un quart de la sphère céleste ($10\,000\,\deg^2$). La mise en service commença en 1998, et le relevé proprement dit débuta en mai 2000. Les deux étapes de photométrie et de spectroscopie sont réalisées avec le télescope de 2.5 m du SDSS situé au Nouveau Mexique (voir Figure 7.2), doté d'un champ de vue de 3°. Pour le relevé spectroscopique, la caméra du relevé photométrique est remplacée dans le plan focal du télescope par une plaque d'aluminium (800 mm de diamètre) où sont emmanchées les 640 fibres optiques. Celles-ci guident la lumière des cibles photométriques jusqu'à deux spectrographes, montés à l'arrière du télescope. Ces derniers comportent une voie bleue et une voie rouge, séparées par un dichroïque à 6000 Å (voir Figure 7.2). Les spectres sont projetés sur des CCD 2048×2048 (avec des pixels de 24 µm), en gardant un espace de 3 pixels entre chaque spectre (de largeur $\simeq 3$ pixels). Huit dispositifs supportant les fibres optiques ont été fabriqués ; ils sont préparés pendant la journée (les fibres sont montées sur la plaque d'aluminium) et peuvent être changés en quelques minutes par un opérateur pour chaque exposition.

Le système spectroscopique a été mis à jour pour le relevé BOSS (Dawson et al., 2013) qui commença en 2008 dans le but de mesurer la position des BAO à une précision du pourcent grâce à la mesure de décalages spectraux de 1.5 millions de galaxies à $z < 0.7$ et de spectre de quasars entre $2.15 < z < 2.35$ contenant des forêts Lyman-α [4]. Le nombre de fibres optiques a été augmenté à 1000 et leur

3. https://github.com/adematti/pyregpt
4. Les forêts Lyman-α sont constituées par l'absorption due à l'hydrogène neutre le long de la ligne de visée.

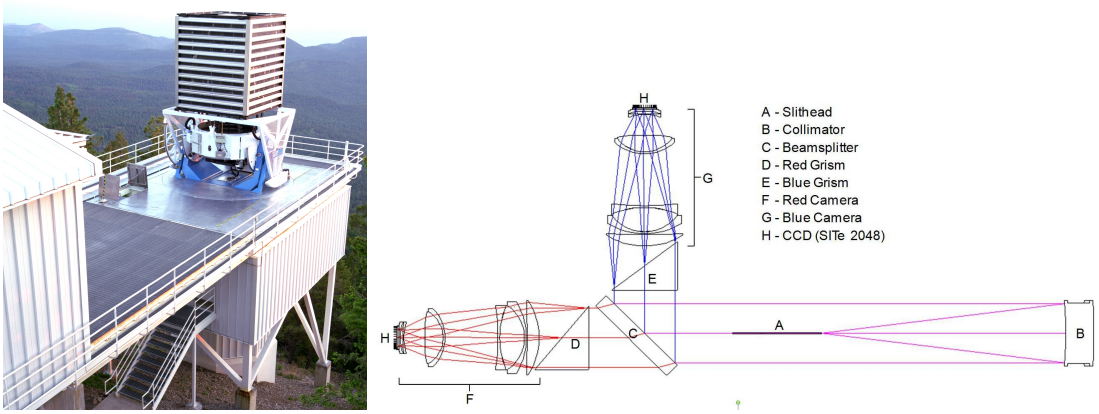

Figure 7.2 – Gauche : téléscope du SDSS, à l'Apache Point Oservatory, au Nouveau Mexique. L'enceinte du téléscope est visible à gauche de l'image, et le dispositif coupe-vent est monté sur le téléscope à droite. Droite : schéma optique des deux spectrographes du SDSS. Les fibres optiques sont fixées en A (slithead). Le faisceau est réfléchi par le collimateur (B), scindé entre les voies rouge et bleue par le dichroïque (C), et dispersé par les grismes rouge (D) ou bleu (E) avant d'entrer dans les caméras (F et G). Source : Smee et al. (2013).

diamètre réduit, pour cibler plus de galaxies, plus lointaines. Le grisme a été changé pour une meilleure transmission, et les CCD des spectrographes ont été remplacés par des CCD 4096×4096, avec des pixels de 15 µm. La résolution (rapport de la longueur d'onde sur la largeur à mi-hauteur de l'image d'une fibre sur le CCD) augmente de $\simeq 1400$ à l'extrême bleu (3560 Å) à $\simeq 2000$ à l'extrême rouge (10 400 Å). Cependant, la résolution et la transmission optiques diminuent sur les bords du plan focal, où la qualité optique est moindre.

Les spectrographes de BOSS ont été réutilisés pour eBOSS (Dawson et al., 2016), l'extension de cinq ans de BOSS. eBOSS a été conçu pour des mesures de *clustering* à plus haut décalage spectral, en ciblant 280 000 LRG (galaxies lumineuses rouges, avec $0.6 < z < 1.0$), 170 000 ELG ($0.6 < z < 1.1$), 330 000 QSO (quasars, avec $0.8 < z < 2.2$) comme traceurs et 60 000 QSO pour les forêts Lyman-α. La dernière publication de données (DR) d'eBOSS, dont l'échantillon d'ELG étudié dans cette thèse fait partie, se nomme DR16.

6.2 La photométrie

Contrairement aux autres relevés d'eBOSS utilisant les données photométriques optiques de SDSS-I-II-III (et proche infrarouge de WISE, Wright et al. 2010), l'échantillon d'ELG a été sélectionné avec le nouveau relevé photométrique DECaLS (Dey et al., 2019), qui est utilisé pour le futur relevé de galaxies DESI (Dark Energy Spectroscopic Instrument DESI Collaboration et al. 2016). Un tel choix tient à la profondeur de DECaLS, au moins une magnitude plus élevée que SDSS dans les trois bandes photométriques utilisées, g, r, z. Les images sont prises avec la caméra DECam (Flaugher et al., 2015), montée sur le télescope Blanco du Cerro Tololo Inter-American Observatory au Chili. Opérationnelle depuis 2012, 62 CCD (59 fonctionnels) pour la photométrie sont installés dans son champ de vue de $3.18\,\mathrm{deg}^2$. Le code `legacypipe` assure le traitement final des données : calibration photométrique et astrométrique, extraction des sources (réalisée par `The Tractor`, une approche dite de *forward modelling*) et génération des catalogues.

Les cibles ELG ont été sélectionnées dans le Nord (NGC, $554\,\mathrm{deg}^2$) et le Sud galactiques (SGC, $616\,\mathrm{deg}^2$) dans les relevés DECaLS DR3 et DR5 (pour la région `eboss25` dans le NGC). Comme le montre la Figure 7.3, le SGC bénéficie d'un nombre d'observations bien plus élevé. La profondeur photométrique y est une demi magnitude plus grande que dans le NGC. Comme les relevés DR3 et DR5 n'étaient pas publics pour le début du programme spectroscopique, les images ont été analysées de

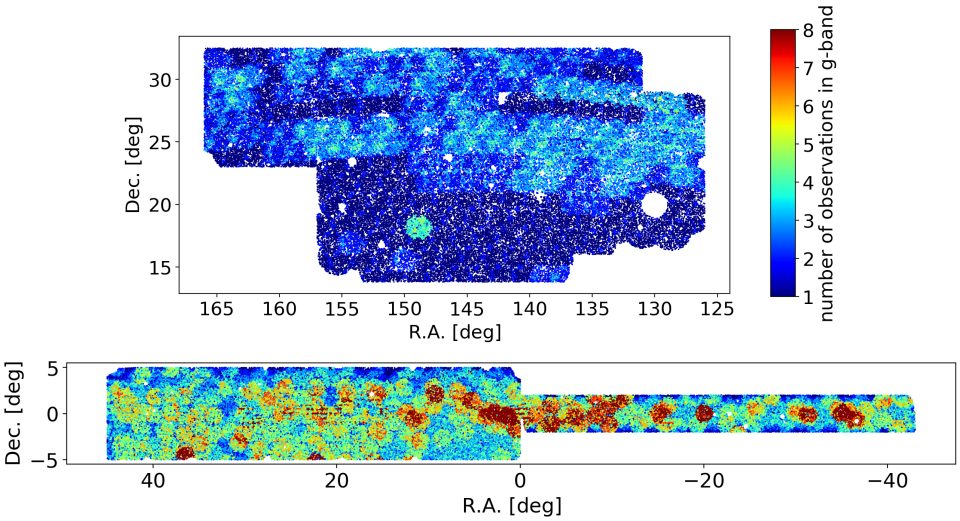

Figure 7.3 – Nombre d'observations dans la bande g pour les cibles ELG d'eBOSS (haut: NGC, bas: SGC). Le nombre d'observations dépasse 8 (le maximum affiché ici) pour 4% des cibles.

manière indépendante par l'équipe d'eBOSS. Une coupure en magnitude g, corrélée avec le flux [OII], et une coupure en couleur $g - r$, $-r - z$ pour l'intervalle en $0.6 < z < 1.1$ ont fourni une liste de 269 718 cibles. En sus, des masques ont été appliqués pour rejeter les régions de photométrie incorrecte : zones sans données, avec des pixels défectueux, proches des étoiles ou avec une calibration photométrique douteuse (Raichoor et al., 2020).

6.3 La spectroscopie

Comme nous l'avons décrit précédemment, les observations spectroscopiques se divisent en un certain nombre de pointés, chacun avec une disposition de 1000 fibres associée à une plaque d'aluminium disposée dans le plan focal. Les fibres ont un diamètre physique, 62″ sur le ciel, qui les empêchent d'être trop proches les unes des autres. L'objectif est de maximiser le taux d'allocation des fibres aux cibles (TSR pour *tiling success rate*), pour un nombre de pointés donné. La première étape consiste à allouer les fibres de manière optimale à chaque plaque. Tout d'abord, un algorithme dit de *Friends-of-Friends* permet de déterminer le plus grand ensemble de cibles (en tenant compte de possibles priorités) qui ne sont en collision avec une autre. Ces cibles sont ensuite affectées à des plaques par un algorithme dit de *network flow* (Blanton et al., 2003). Un deuxième algorithme de ce type permet de répartir les fibres aux cibles plus proches que 62″, en mettant à profit le recouvrement des plaques sur le ciel. Enfin, la position des plaques est ajustée dans des zones indépendantes appelées *chunk*, de manière récursive pour bénéficier des recouvrements dans les zones de grande densité de cibles (avec une pénalité pour les écarts à la distribution uniforme).

Les ELG d'eBOSS se sont vu attribuer 305 plaques de $\simeq$ 850 fibres. Le relevé est séparé en 4 *chunks* : `eboss23` et `eboss25` dans le NGC, `eboss21` et `eboss22` dans le SGC. Certaines cibles (TDSS, Ruan et al., 2016) ($\simeq$ 50 par plaque) ont été observées avec les mêmes plaques que les ELG d'eBOSS, une partie (FES) avec la même priorité, les autres ayant des priorités moins élevées. Pour tenir compte des possibles éliminations de fibres ELG au profit des cibles TDSS FES, un masque de rayon 62″ est appliqué autour de ces dernières. Par ailleurs, les fibres d'une plaque donnée ne peuvent être placées dans un rayon de 92″ autour du centre de la plaque, par lequel elle est maintenue dans le plan focal du télescope. Un masque supplémentaire en tient compte.

Une fois les fibres allouées aux cibles photométriques, le spectre de ces dernières peut être mesuré par les spectrographes décrits dans la Section 6.1. Un spectre obtenu par concaténation des données de SEQUELS (Sloan Extended Quasar, ELG and LRG Survey, Alam et al., 2015) et des observations pilotes des ELG d'eBOSS est montré sur la Figure 7.4. Comme leur nom le suggère, les spectres des ELG

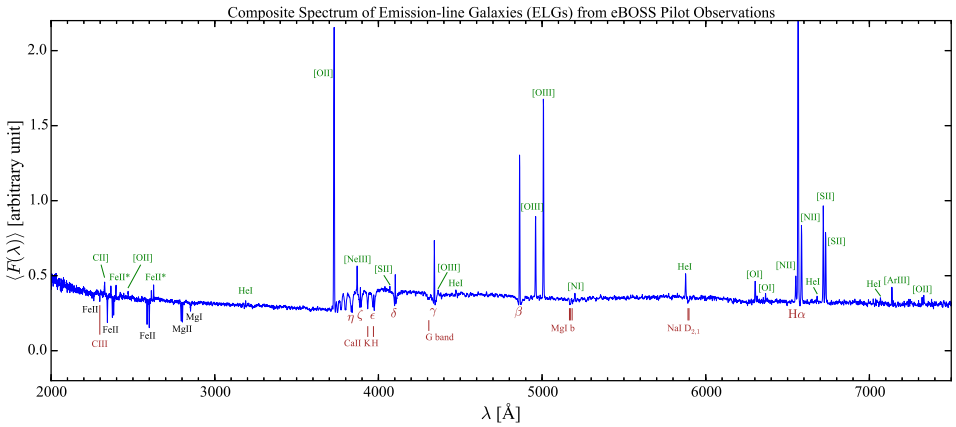

Figure 7.4 – Spectre concaténé de 12 000 ELG des observations SEQUELS et eBOSS. Les raies d'émission sont notées en vert, les raies d'absorption stellaire en rouge et les raies d'absorption du millieu interstellaire / circumgalactique en noir. Source : Zhu et al. (2015).

possèdent des raies d'émission, dont la plupart sont des raies nébulaires : le doublet [OII]($\lambda\lambda 3727, 3730$), les raies [OIII]($\lambda 5008$), [OI]($\lambda 6300$), [NII]($\lambda 6584$). Celles-ci sont émises par le gaz distribué dans la galaxie, photoexcité par la lumière ultraviolette produite par des étoiles chaudes à courte durée de vie, responsables du continuum à ~ 2000 Å. Ce dernier atténue la discontinuité de Balmer à 3646 Å qui se prolonge jusqu'à 6563 Å par la série de Balmer (η, ζ, ϵ, δ, γ, β, Hα). Les ELG à formation d'étoiles constituent un traceur intéressant pour les analyses de *clustering*. En effet, leur distribution suit le taux de formation d'étoiles, qui s'accroît jusqu'à $z \simeq 2$, où les galaxies rouges se font plus rares. De plus, les fortes raies d'émission facilitent la mesure du décalage spectral. En particulier, le doublet [OII] peut être résolu par un spectrographe possédant une résolution supérieure à 4000 — ce qui est le cas de DESI, rendant la mesure du décalage spectral extrêmement robuste. Voyons justement comment les décalages spectraux sont mesurés.

Comme mentionné dans la Section 6.1, les spectres s'étalent sur 3 colonnes de pixels. Le signal obtenu est débiaisé, converti en nombre d'électrons, et aplani, la lumière diffusée soustraite. Il est ensuite compressé en spectres unidimensionnels, en réalisant un ajustement pour chaque ligne de CCD. Le signal obtenu est enfin corrigé par le champ plat de chaque fibre, le modèle de ciel, les distorsions entre plusieurs observations et au sein du CCD, et calibré avec des spectres d'étoiles standard. Un nouvel algorithme de détermination du décalage spectral, `REDROCK`, a été utilisé pour eBOSS, afin de répondre aux objectifs d'eBOSS (Dawson et al., 2016) : 90% de mesures correctes du décalage spectral, avec moins de 1% de mesures catastrophiques, c'est-à-dire qui sont fausses de plus de 1000 km s^{-1} pour les ELG. Dans une première étape, les spectres mesurés sont ajustés par des patrons issus d'une décomposition PCA de spectres d'étoiles, galaxies et quasars. Les ajustements non physiques sont supprimés dans une deuxième étape, au cours de laquelle des modèles de spectres (appelés *archétypes*, 110 galaxies, 64 quasars et 40 étoiles) sont ajustés avec des polynômes de Legendre au spectre mesuré, dans le voisinage des 3 meilleurs ajustements obtenus à l'issue de la première étape. La mesure (décalage spectral, erreur, et classe d'objets) avec le meilleur χ^2 est retenue, et le différence de χ^2 (appelée $\Delta\chi^2$ dans la suite) au second meilleur ajustement est retenu comme critère de fiabilité de la mesure.

Le seuil $\Delta\chi^2 \geq 9$ est déterminé avec les galaxies rouges d'eBOSS (Ross et al., 2020), auquel s'ajoute la positivité du coefficient de l'archétype pour obtenir la sélection ZWARNING = 0. Pour les ELG, des coupures supplémentaires sont appliquées ; elles imposent un signal sur bruit médian minimum dans les bandes i et z du spectre, où se situe le doublet [OII] des galaxies à $z \simeq 1$, et des critères de détection de raies et du continuum (Comparat et al., 2016; Raichoor et al., 2020). Les mesures de décalages spectraux satisfaisant ces critères sont qualifiés de fiables. Tandis que l'algorithme de BOSS permettait d'obtenir des décalages spectraux fiables pour 74.0% des objets, `REDROCK` a augmenté cette efficacité (appelée SSR pour *spectroscopic success rate*) à 80.7%. La fraction de mesures catastrophiques a été estimée à 0.3%

avec $\sim 13\,000$ observations répétées. En sus, des inspections visuelles ont été réalisées sur un pointé du relevé eBOSS ELG, et 2 pointés du programme pilote (Comparat et al., 2016). Elles ont confirmé que les mesures fiables étaient correctes dans 99% des cas, avec une fraction de mesures catastrophiques de 1%, répondant ainsi aux exigences d'eBOSS. Des comparaisons avec des relevés externes ont permis de confirmer ces résultats (Raichoor et al., 2020).

7 Mesure du spectre de puissance des galaxies

Nous avons vu qu'un relevé spectroscopique de galaxies permet d'obtenir un catalogue de positions angulaires et décalages spectraux de galaxies. Dans cette section nous décrivons comment compresser ces données sous la forme d'un spectre de puissance (Section 7.1), ce qui induit différents effets géométriques qu'il convient alors de modéliser (Section 7.2). Certains de ces effets sont appelés contraintes intégrales (Section 7.3), lesquelles ont fait l'objet d'une publication premier auteur.

7.1 L'estimateur du spectre de puissance

Construisons le champ FKP $F(\mathbf{x})$ donné par (Feldman et al., 1994):

$$F(\mathbf{x}) = n_g(\mathbf{x}) - \alpha n_s(\mathbf{x}), \tag{123}$$

où $n_g(\mathbf{x})$ est la densité de galaxies (incluant de possibles poids) et $n_s(\mathbf{x})$ la densité d'un catalogue dit *synthétique* qui représente la fonction de sélection du relevé $W(\mathbf{r}) = \langle n_g(\mathbf{r}) \rangle$, la moyenne d'ensemble de la densité pondérée de galaxies. α est un facteur de normalisation, en théorie fixé à :

$$\alpha = \frac{\int d^3x W(\mathbf{x})}{\int d^3x n_s(\mathbf{x})}. \tag{124}$$

A partir de ce champ, on construit l'estimateur de Yamamoto (Yamamoto et al., 2006) :

$$\hat{P}_\ell(k_\mu) = \frac{2\ell+1}{AV_{k_\mu}} \int_{V_{k_\mu}} d^3k F_0(\mathbf{k}) F_\ell(-\mathbf{k}) - P_\ell^{\rm noise}(k_\mu) \tag{125}$$

où l'intégrale se fait sur une coquille en k_μ et :

$$F_\ell(\mathbf{k}) = \int d^3x e^{i\mathbf{k}\cdot\mathbf{x}} F(\mathbf{x}) \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{x}}), \tag{126}$$

où l'on a noté $\mathcal{L}_\ell$ le polynôme de Legendre d'ordre ℓ. La normalisation A est donnée par :

$$A = \alpha^2 \int d^3r n_s^2(\mathbf{r}). \tag{127}$$

Pour minimiser la variance de l'estimateur (125), les galaxies sont pondérées par le poids FKP (Feldman et al., 1994) :

$$w_{\rm FKP}(\mathbf{x}) = \frac{1}{1 + \bar{n}(\mathbf{x}) P_0} \tag{128}$$

où P_0 est la valeur typique du spectre de puissance aux échelles d'intérêt. Le bruit de Poisson dans l'Eq. (125) s'écrit :

$$P_\ell^{\rm noise}(k_\mu) = \frac{\delta_{\ell 0}}{A} \int d^3x \left[W(\mathbf{x}) w_g(\mathbf{x}) + \alpha^2 n_s(\mathbf{x}) w_s(\mathbf{x}) \right]. \tag{129}$$

Pour minimiser le bruit de Poisson dû au catalogue synthétique, on aura typiquement $n_s/n_g \propto \alpha^{-1} \sim 20 - 50$.

L'estimateur de Yamamoto (125) peut être obtenu avec le formalisme des estimateurs optimaux quadratiques, après un certain nombre d'approximations, dont la covariance du champ de fluctuations diagonale, l'approximation de ciel localement plat, la non correction de l'effet de fonction fenêtre — sujet sur lequel on reviendra dans la prochaine section. En pratique, l'Eq. (126) peut être mise sous la forme de transformées de Fourier, par exemple en développant le polynôme de Legendre $\mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}})$ en harmoniques sphériques (Hand et al., 2017). Nous utilisons l'implémentation de l'estimateur de Yamamoto fournie dans le module `nbodykit` (Hand et al., 2018), qui interpole le champ $F(\mathbf{x})$ sur une grille et calcule l'Eq. (126) avec des transformées de Fourier rapides, en atténuant les effets de repliement spectral avec la technique dite d'*interlacing* (Sefusatti et al., 2016).

7.2 Effets de géométrie

Les relevés spectroscopiques de galaxies permettent d'obtenir un catalogue de positions angulaires R.A., Dec. et de décalages spectraux z. Ceux-ci sont convertis en positions cartésiennes en supposant une cosmologie dite *fiducielle* fid, laquelle ne correspond pas nécessairement à la vraie cosmologie qui sous-tend les données. Ainsi, des distorsions sont introduites, connues sous le nom d'effet Alcock-Paczynski (Alcock and Paczynski, 1979). Les vecteurs d'onde parallèles et transverses à la ligne de visée dans la vraie cosmologie doivent être multipliés par :

$$q_\parallel = \frac{D_\mathrm{H}(z_\mathrm{eff})}{D_\mathrm{H}^\mathrm{fid}(z_\mathrm{eff})} \qquad \text{et} \qquad q_\perp = \frac{D_\mathrm{M}(z_\mathrm{eff})}{D_\mathrm{M}^\mathrm{fid}(z_\mathrm{eff})} \tag{130}$$

pour obtenir leur valeur dans la cosmologie fiducielle. Par ailleurs, pour des raisons de temps de calcul, le spectre initial de la matière est fixé à la cosmologie fiducielle dans le modèle du spectre de puissance. En contrepartie, $q_\parallel$ et $q_\perp$ sont laissés libres dans les ajustements cosmologiques. De manière effective, ceux-ci sont principalement fixés par l'échelle BAO r_drag, et sont donc sensibles aux combinaisons :

$$\alpha_\parallel = \frac{D_\mathrm{H}(z_\mathrm{eff})r_\mathrm{drag}^\mathrm{fid}}{D_\mathrm{H}^\mathrm{fid}(z_\mathrm{eff})r_\mathrm{drag}} \qquad \alpha_\perp = \frac{D_\mathrm{M}(z_\mathrm{eff})r_\mathrm{drag}^\mathrm{fid}}{D_\mathrm{M}^\mathrm{fid}(z_\mathrm{eff})r_\mathrm{drag}} \tag{131}$$

Par ailleurs, l'estimateur (125) mesure en fait le spectre de puissance des galaxies convolué par la fonction fenêtre du relevé. Cet effet est pris en compte dans le modèle du spectre de puissance. Pour cela, on travaille en espace de configuration. Les multipoles de la fonction de corrélation théorique sont obtenus à partir des multipoles du spectre de puissance par une transformée de Hankel (inverse). Ensuite, la fonction de corrélation $\xi_p(s)$ est multipliée par la fonction fenêtre $\mathcal{W}_q^{\delta,\delta}(s)$:

$$\xi_\ell^c(s) = \sum_{p,q} A_{\ell p}^q \frac{2\ell+1}{2q+1}\xi_p(s)\mathcal{W}_q^{\delta,\delta}(s), \tag{132}$$

avec $A_{\ell p}^q$ tel que $\mathcal{L}_\ell(\mu)\mathcal{L}_p(\mu) = \sum_{q=0}^{\ell+p} A_{\ell p}^q \mathcal{L}_q(\mu)$. Les multipôles de la fonction fenêtre sont donnés par :

$$\mathcal{W}_\ell^{\delta,\delta}(s) = \frac{2\ell+1}{4\pi A}\int d\Omega_s \int d^3x W(\mathbf{x})W(\mathbf{x}-\mathbf{s})\mathcal{L}_\ell(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}}) \tag{133}$$

où A est la normalisation (124). En pratique, cette fonction fenêtre est calculée par simple comptage de paires du catalogue synthétique. Jusqu'à présent, une seule ligne de visée a été considérée pour chaque paire de galaxies. Cette approximation de ciel localement plat ne tient plus lorsque la séparation s entre les galaxies est non négligeable par rapport à leur distance d à l'observateur. Ces effets de grand angle sont modélisés avec un développement de la fonction de corrélation en s/d (Szalay et al., 1998; Castorina and White, 2018). Des multipoles d'ordre impairs sont générés, lesquels se couplent au multipoles d'ordre impair de la fonction fenêtre (Beutler et al., 2019) et génèrent des corrections en s/d aux multipoles pairs

de la fonction de corrélation (132). Néanmoins, ces effets sont localisés à grande échelle et restent donc négligeables pour les analyses de *clustering* actuelles consacrées aux RSD et BAO.

Notons enfin que le spectre de puissance étant estimé sur une grille en $\mathbf{k}$ (Section 7.1), celui-ci est peu échantillonné au petit k, ce que l'on corrige dans le modèle du spectre de puissance (Beutler et al., 2017).

7.3 Contraintes intégrales

Jusqu'à présent, nous avons supposé la fonction de sélection $W(\mathbf{r}) = \langle n_g(\mathbf{r})\rangle$ entièrement connue. Ce n'est pas le cas en pratique. Par exemple, le nombre de galaxies attendues dans le relevé, $\int d^3x W(\mathbf{x})$, est très compliqué à prédire. En conséquence, on le remplacera par le nombre de galaxies réellement observées $\int d^3x n_g(\mathbf{x})$ dans l'Eq. (124), ce qui revient à imposer pour les fluctuations de densité $\delta^{\rm cic} = n_g - \alpha n_s$: $\int d^3r \delta^{\rm cic}(\mathbf{r}) = 0$. Cet effet, dite de contrainte intégrale globale, était déjà connu. Néanmoins, la fonction de sélection radiale du relevé, c'est-à-dire sa densité en décalge spectral, est aussi difficile à prévoir. En conséquence, dans les analyses de SDSS, celle-ci est directement mesurée à partir des données observées, par exemple avec la technique dite *shuffled* (Samushia et al., 2012; Ross et al., 2012; Reid et al., 2016) : les décalages spectraux du catalogue synthétique sont pris aléatoirement dans la distribution des données. Ainsi, l'intégrale de $\delta^{\rm cic}(\mathbf{r})$ est nulle dans toute tranche en distance radiale : il s'agit de l'effet de contrainte intégrale radiale. Nous avons formalisé l'effet des contraintes intégrales par la formule :

$$\delta^{\rm cic}(\mathbf{r}) = W(\mathbf{r})\left\{\delta(\mathbf{r}) - \int d^3x W_{\rm ic}(\mathbf{x})\delta(\mathbf{x})\epsilon_{\rm ic}(\mathbf{r},\mathbf{x})\right\} \qquad W_{\rm ic}(\mathbf{r}) = \frac{W(\mathbf{r})}{\int d^3x W(\mathbf{x})\epsilon_{\rm ic}(\mathbf{r},\mathbf{x})} \tag{134}$$

avec $\epsilon_{\rm ic}(\mathbf{r},\mathbf{x})$ une fonction caractérisant la contrainte intégrale : $\epsilon_{\rm glo}(\mathbf{r},\mathbf{x}) = 1$ pour la contrainte intégrale globale (GIC), $\epsilon_{\rm rad}(r,x)$ pour la contrainte intégrale radiale (RIC). De là, la fonction de corrélation à deux points de $\delta^{\rm cic}(\mathbf{r})$ s'écrit :

$$\boxed{\xi_\ell^{\rm cic}(s) = \xi_\ell^{\rm c}(s) - IC_\ell^{\delta,{\rm ic}}(s) - IC_\ell^{{\rm ic},\delta}(s) + IC_\ell^{{\rm ic},{\rm ic}}(s)} \tag{135}$$

avec $\xi_\ell^{\rm c}(s)$ donnée par l'Eq. (132) et :

$$IC_\ell^{i,j}(s) = \int \Delta^2 d\Delta \sum_p \frac{4\pi}{2p+1}\xi_p(\Delta)\mathcal{W}_{\ell p}^{i,j}(s,\Delta) \qquad (i,j)\in\{(\delta,{\rm ic}),({\rm ic},\delta),({\rm ic},{\rm ic})\} \tag{136}$$

Dans cette formule, $\mathcal{W}_{\ell p}^{\delta,{\rm ic}}(s,\Delta)$, $\mathcal{W}_{\ell p}^{{\rm ic},\delta}(s,\Delta)$ sont des fonctions de corrélation à 3 points de la fonction de sélection W, p. ex. :

$$\begin{aligned}\mathcal{W}_{\ell p}^{\delta,{\rm ic}}(s,\Delta) = \frac{(2\ell+1)(2p+1)}{(4\pi)^2 A}\int d\Omega_s \int d\Omega_\Delta \\ \int d^3x W(\mathbf{x})W(\mathbf{x}-\mathbf{s})W_{\rm ic}(\mathbf{x}-\mathbf{\Delta})\mathcal{L}_\ell(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}})\mathcal{L}_p(\hat{\mathbf{x}}\cdot\hat{\mathbf{\Delta}})\epsilon_{\rm ic}(\mathbf{x}-\mathbf{s},\mathbf{x}-\mathbf{\Delta}).\end{aligned} \tag{137}$$

Ces fonctions de corrélation à 3 points sont calculées à partir du catalogue synthétique, après avoir implémenté (et adapté) l'algorithme de Slepian and Eisenstein (2018). $\mathcal{W}_{\ell p}^{{\rm ic},{\rm ic}}(s,\Delta)$ est une fonction de corrélation à 4 points de W:

$$\begin{aligned}\mathcal{W}_{\ell p}^{{\rm ic},{\rm ic}}(s,\Delta) = \frac{(2\ell+1)(2p+1)}{(4\pi)^2 A}\int d\Omega_s \int d\Omega_\Delta \int d^3y W(\mathbf{y})W(\mathbf{y}-\mathbf{s}) \\ \int d^3x W_{\rm ic}(\mathbf{x})W_{\rm ic}(\mathbf{x}-\mathbf{\Delta})\mathcal{L}_\ell(\hat{\mathbf{y}}\cdot\hat{\mathbf{s}})\mathcal{L}_p(\hat{\mathbf{x}}\cdot\hat{\mathbf{\Delta}})\epsilon_{\rm ic}(\mathbf{y},\mathbf{x})\epsilon_{\rm ic}(\mathbf{y}-\mathbf{s},\mathbf{x}-\mathbf{\Delta})\end{aligned} \tag{138}$$

qui est séparée en un produit de fonctions de corrélation à 2 points $\mathbf{x},\mathbf{y}$ dans chaque ensemble où $\epsilon_{\rm ic}(\mathbf{x},\mathbf{y})$ est constant (et non nul). Notons que les contraintes intégrales ont une contribution venant du bruit de Poisson, que nous avons calculée. L'effet de la technique *shuffled* a été évalué sur des simulations. Sans

correction, un biais[5] était visible dans les paramètres $f\sigma_8$, $\alpha_\parallel$ et $\alpha_\perp$ ajustés aux simulations. Une fois la correction de contrainte intégrale radiale incluse dans le modèle avec les équations précédentes, aucun biais n'a pu @ être détecté.

Par ailleurs, nous avons proposé de supprimer les fluctuations de densité dans des pixels du ciel, afin d'atténuer des systématiques angulaires (une technique appelée pixélisation). Ce faisant, l'on introduit une contrainte intégrale angulaire qui peut être traitée de façon similaire. Nous avons par ailleurs montré que cette contrainte intégrale angulaire peut être combinée avec la contrainte intégrale radiale, dans le cas où cette dernière doit être modélisée. Enfin, les corrections de grand angle évoquées précédemment ont été incorporées dans le formalisme des contraintes intégrales. Ces résultats ont été présentés dans de Mattia and Ruhlmann-Kleider (2019).

En résumé, le modèle théorique du spectre de puissance est construit à partir de la théorie des perturbations (Section 5), en incluant l'effet d'Alcock-Paczynski et (si besoin) la convolution par la fonction fenêtre du relevé (Section 7.2) et les contraintes intégrales (Section 7.3).

8 Analyse

Nous avons vu comment mesurer le spectre de puissance d'un relevé de galaxies, et le modéliser. Dans cette section, nous appliquons ces développements à l'analyse de *clustering* de l'échantillon des ELG d'eBOSS, qui a fait l'objet de la publication de Mattia et al. (2020). Nous commençons par tester notre modèle théorique avec des simulations (Section 8.1) puis expliquons la construction des catalogues de données (Section 8.2) et de leurs simulations (Section 8.3) avant de présenter les résultats finaux de l'analyse (Section 8.4) et de la collaboration SDSS (Section 8.5).

8.1 Mock challenge

Dans un premier temps, il convient d'évaluer la précision du modèle théorique du spectre de puissance. Pour cela, nous prendrons comme référence des simulations dites à N corps. Celles-ci résolvent l'équation de Vlasov (dont les deux premiers moments sont l'équation de continuité et d'Euler) et l'équation de Poisson, en échantillonnant la fonction de distribution de la matière noire avec des particules (sans collisions), typiquement au nombre de 10^{10-12} dans les simulations cosmologiques actuelles. Nous avons principalement utilisé la simulation OuterRim (Heitmann et al., 2019), faisant évoluer $10\,240^3$ particules de matière noire de masse $1.85 \times 10^9\,\mathrm{M}_\odot\,h^{-1}$ dans un volume de $27\,\mathrm{Gpc}^3\,h^{-3}$. Sous l'action de la gravité, les particules forment des structures gravitationellement liées appelées halos de matière noire. Un algorithme de *Friends-of-Friends* permet de les identifier dans les simulations. Ensuite, une prescription appelée HOD (pour *halo occupation distribution*) donne la probabilité $P(N|M)$ de trouver N galaxies dans un halo de masse M. $P(N|M)$ est souvent séparé pour les galaxies centrales (situées au centre du halo) et les galaxies satellites. Cette prescription est généralement ajustée sur les prédictions de modèles semi-analytiques de formation des galaxies contraints par des observations. Pour des galaxies sélectionnées par masse stellaire, le nombre moyen de galaxies centrales $\langle N_{\mathrm{cent}}(M)\rangle$ est bien décrit par une fonction créneau, atteignant 1 à haute masse de halo M. En revanche, pour des galaxies sélectionnées par taux de formation d'étoiles (SFR pour *star formation rate*), le nombre de galaxies centrales diminue au-delà de $\sim 10^{12}\,\mathrm{M}_\odot\,h^{-1}$. Dans les deux cas, le nombre moyen de satellites $\langle N_{\mathrm{sat}}(M)\rangle$ est une fonction croissante de la masse. Néanmoins, les prescriptions trouvées dans la littérature pour les galaxies sélectionnées par SFR diffèrent significativement, selon les différents mécanismes de suppression du taux de formation d'étoiles considérés (Contreras et al., 2013).

L'objectif du *mock challenge* est de vérifier la robustesse du modèle de spectre de puissance des galaxies selon la relation entre galaxies et halos, en variant ses paramètres. Différentes formes de HOD (incluant une fonction créneau et une gaussienne en $\log M$ pour $\langle N_{\mathrm{cent}}(M)\rangle$) ont été testées (Alam et al., 2020). Les galaxies centrales ont été déplacées d'une fraction (1/10) du rayon du halo. Par ailleurs, le nombre de galaxies satellites a été modifié et leur dispersion de vitesse variée, de ±50%. Orsi and

5. qui dépend de la configuration du relevé, dans notre cas 2.8%, 0.9% et 0.6% sur $f\sigma_8$, $\alpha_\parallel$ et $\alpha_\perp$

Angulo (2018) suggèrent que les ELG se trouvent en majorité en périphérie des halos, avec une vitesse radiale centripète. Ainsi, les positions des galaxies satellites ont été modifiées (via la concentration du halo), et une vitesse radiale moyenne de $500\,\mathrm{km\,s^{-1}}$ ajoutée. Enfin, Jiménez et al. (2019) montre que la loi de probabilité a un impact significatif sur la corrélation intra-halo, et qu'une loi binomiale négative correspond le mieux aux galaxies à haut SFR — ce qui a été également testé (Avila et al., 2020). Enfin, la connexion entre halos et galaxies peut, en plus de la masse, dépendre de propriétés secondaires du halos — un effet appelé *assembly bias*, exploré en introduisant une dépendance dans la concentration selon Hearin et al. (2016). En sus, une analyse en aveugle a été effectuée : les vitesses des galaxies ont été multipliées par un facteur qui ne fut rélevé qu'après avoir fourni la mesure des paramètres cosmologiques — laquelle s'est révélée en accord avec leur valeur attendue. Cette batterie de tests a permis d'établir un budget d'incertitude systématique du modèle de 1.6% pour $f\sigma_8$, 0.8% et 0.7% pour $\alpha_\parallel$ et $\alpha_\perp$.

Comme mentionné en Section 7, le spectre de puissance initial utilisé dans les calculs de théorie des perturbations est gardé fixe dans la cosmologie fiducielle. L'impact de cette simplification a été quantifié avec les simulations OuterRim en modifiant les paramètres de la cosmologie utilisée pour calculer ce spectre de puissance dans un intervalle de largeur 30 fois les incertitudes de Planck Collaboration et al. (2018) (CMB et BAO). Au total, le budget d'incertitude systématique du modèle s'élève à 3.0% pour $f\sigma_8$, 0.9% pour $\alpha_\parallel$ et 0.8% pour $\alpha_\perp$.

Par ailleurs, nous avons réalisé une mesure du signal BAO isotropique dans les ELG d'eBOSS. La technique dite de reconstruction a été employée afin d'enlever (partiellement) l'évolution non-linéaire du champ de galaxies due à la gravité et aux RSD, afin de renforcer les BAO dans le spectre de puissance des données. Un modèle très simple, constitué des oscillations BAO multipliées par un patron doté d'une polynôme dont les coefficients sont laissés libres a été ajusté à ce spectre de puissance reconstruit. Toujours avec la simulation OuterRim, et en testant l'impact de la cosmologie utilisée pour le patron du modèle, le budget d'incertitude systématique pour l'analyse BAO seule a pu être fixé à 0.2% pour le paramètre de dilatation $\alpha = \alpha_\parallel^{1/3}\alpha_\perp^{2/3}$.

Cette étape dite de *mock challenge* nous a permis de noter que les mesures des paramètres cosmologiques, notamment $f\sigma_8$, varient fortement (typiquement 1σ) selon la ligne de visée utilisée pour l'implémentation des RSD dans les simulations. Nous avons montré dans Smith et al. (2020) que les quadrupoles du spectre de puissance selon deux lignes de visée orthogonales sont fortement anticorrélés, ce qui se traduit par une anticorrélation des valeurs mesurées de $f\sigma_8$. Nous avons alors proposé de moyenner les mesures selon plusieurs lignes de visées, afin d'accroître la précision statistique du *mock challenge* pour un coût marginal quasiment nul (estimation et ajustement de spectres de puissance).

8.2 Données

Nous avons activement participé à la création des catalogues de données, incluant des pondérations pour corriger d'effets systématiques liés à la photométrie et à la spectroscopie, ainsi que de catalogues simulés, appelés *mocks*. Comme nous l'avons expliqué dans la Section 7.1, la fonction de sélection est échantillonnée par un catalogue synthétique, que nous devons construire de telle sorte que la moyenne d'ensemble de la densité de données soit égale à celle du catalogue synthétique. Les masques liés à la photométrie et à l'attribution des fibres aux cibles présentés en Section 6 sont appliqués au catalogue de données et au catalogue synthétique. Des observations avec un faible signal sur bruit dans `eboss22` sont aussi masquées (3123 cibles). Pour tenir compte des collisions de fibres, les cibles ayant reçu une fibre sont pondérées par le rapport du nombre de cibles sur le nombre de fibres dans chaque groupe de collisions. La perte d'efficacité due aux fibres non fonctionnelles, ou à une insuffisante densité de fibres est prise en compte par pondération du catalogue synthétique.

L'algorithme de détermination du décalage spectral ne fournit pas une mesure fiable (selon les critères définis en Section 6.3) pour tous les spectres mesurés. La fraction de mesures fiables, le SSR, dépend des conditions d'observation. Une relation entre SSR et le signal sur bruit pour chaque pointé et dans chaque *chunk* et demi-spectrographe est ajustée, et les objets pondérés par l'inverse de cette relation. Par ailleurs, comme mentionné dans la Section 6, la transmission diminue sur les bords du plan focal.

Ainsi, la relation entre SSR et la position dans le plan focal est ajustée (après correction précédente) et les objets pondérés par l'inverse de cette relation.

Après les corrections liées à l'attribution des fibres aux cibles et la mesure du décalage spectral, la dernière source de variation angulaire de la densité de cibles provient de la photométrie. La densité de galaxies (pondérée par les corrections précédentes) est calculée dans des pixels `HEALPix` (Górski et al., 2005) avec nside = 256 (surface de pixels de $\simeq 188\,\mathrm{arcmin}^2$). Un modèle linéaire dans des variables photométriques (densité d'étoiles, extinction galactique, température de poussières, densité colonne de HI, profondeur photométrique et qualité d'image dans les bandes g, r, z) est ajustée à cette densité de galaxies, lesquelles sont aussi pondérées par l'inverse de la prédiction.

Enfin, des décalages spectraux sont sélectionnés aléatoirement dans les données et assignés au catalogue synthétique avec la technique *shuffled* (voir Section 7), dans l'intervalle $0.6 < z < 1.1$. Cependant, il a été noté que la fonction de sélection radiale varie en fonction de la profondeur de l'imagerie, donc de la position angulaire. Cette effet, dû à la faible luminosité des cibles par rapport à la profondeur du relevé, a été noté pour la première fois dans l'échantillon des ELG d'eBOSS, car les autres échantillons du SDSS sont composés de cibles plus lumineuses et disposant d'une profondeur d'imagerie plus isotrope. Pour tenir compte de ces variations angulaires de la fonction de sélection radiale, cette dernière a été mesurée dans des sous-échantillons du relevé ayant approximativement la même profondeur photométrique. Pour cela, une relation entre décalage spectral et une combinaison de la profondeur dans les bandes g, r, z a été ajustée sur les données. Cette nouvelle variable a permis de définir trois intervalles par *chunk* où les décalages spectraux des galaxies ont été assignés au catalogue synthétique. Cette technique augmente l'impact de la contrainte intégrale radiale, ce que nous avons pu modéliser selon la Section 7.3.

Enfin, la densité de galaxies est mesurée, et le poids FKP (voir l'Eq. (128), avec $P_0 = 4000\,\mathrm{Mpc}^3\,h^{-3}$) est assigné aux données et au catalogue synthétique.

8.3 Mocks

Afin de tester ces procédures de correction des données, des catalogues simulés appelés *mocks* ont été réalisés : 1000 *mocks* EZ et 2003 *mocks* GLAM-QPM. Ces *mocks* sont issus de simulations rapides du champ de densité des galaxies, soit avec la théorie des perturbations et un modèle de biais (*mocks* EZ), soit en résolvant le système de Vlasov-Poisson sur grille et en utilisant un HOD (*mocks* GLAM-QPM). Les masques angulaires des données ont été appliqués aux *mocks*. Des variations de la densité angulaire de cibles ont été introduites dans les EZ *mocks* à partir d'une carte de densité des données filtrée avec une gaussienne d'écart-type 1°. Ensuite, la fonction de sélection radiale a été appliquée dans les sous-échantillons de profondeur homogène décrits ci-dessus. Les collisions de fibres ont été implémentées avec une heuristique similaire à l'algorithme d'attribution des fibres présenté en Section 6 : les groupes de collisions sont trouvés avec un algorithme *Friends-of-Friends*, les fibres assignées aux cibles de manière à maximiser le TSR dans chaque groupe de collisions, en tenant des comptes des priorités avec les (vraies) cibles TDSS, et des recouvrements de plaques. Enfin, un décalage spectral dans un *mock* est déclaré fiable si c'est le cas de son plus proche voisin dans les données, ou (dans un second jeu de *mocks*) avec une probabilité prédite par le modèle ajusté aux données. Enfin, un catalogue synthétique est généré pour chaque *mock*, où toutes les pondérations et le tirage des décalages spectraux décrits précédemment pour les données sont appliquées (avec le même code). Une méthodologie similaire est utilisée pour les *mocks* GLAM-QPM, à la différence près qu'aucune systématique photométrique ni variation de la fonction de sélection radiale avec la profondeur photométrique n'est implémentée. L'accord entre les différentes efficacités (e.g. TSR, SSR) des données et des *mocks* est excellent, de l'ordre de 0.5%.

La chaine d'analyse a pu être testée avec ces *mocks*, notamment : l'effet de la géométrie du relevé et des masques angulaires, de la contrainte intégrale radiale, des collisions de fibres (lesquelles ont été modélisées avec la technique de Hahn et al. (2017)) et des systématiques angulaires résiduelles. Par ailleurs, les postérieurs des données pour les mesures RSD et BAO étant significativement non gaussiens, nous avons combiné les multipoles du spectre de puissance pré et post reconstruction, de manière à effectuer l'ajustement conjoint des modèles de RSD et de BAO reconstruit. Cette méthodologie a pu être vérifiée grâce aux *mocks*. Enfin, les *mocks* nous ont permis de remarquer que le quadrupole des

source	α
linéaire	
non-linéarité du modèle (mocks EZ)	0.1%
quadrature	
systématiques du modèle (mock challenge)	0.2%
terme d'amortissement $\Sigma_{\rm nl}$	0.8%
systématiques photométriques	1.3%
collisions de fibres (mocks EZ)	0.2%
échecs de mesure des décalage spectraux	0.3%
statistique	$^{+0.025}_{-0.028}$
systématique	$^{+0.006}_{-0.005}$
total	$^{+0.031}_{-0.033}$

Table 7.1 – Budget d'erreur pour la mesure du BAO isotropique post-reconstruction. Les pourcentages sont relatifs à la valeur de α. Les trois dernières lignes (statistique, systématique et total) donnent la barre d'erreur statistique, la barre d'erreur systématique (totale moins statistique), et la barre d'erreur totale, respectivement.

données était significativement plus bas qu'attendu à grande échelle, ce que nous avons pu reproduire en partie avec l'injection de systématiques angulaires dans les *mocks*, comme mentionné précédemment. En appliquant la technique de pixélisation introduite en Section 7 (avec une surface de pixel de $\simeq 0.84\,\mathrm{deg}^2$), les biais sur $f\sigma_8$, $\alpha_\parallel$ et $\alpha_\perp$ dus à ces systématiques ont pu être réduits de $\simeq 70\%$. La pixélisation a donc été appliquée aux données pour atténuer de potentielles systématiques angulaires résiduelles.

Ces *mocks* nous ont aussi permis de calculer les matrices de covariance des fonctions de vraisemblance pour les ajustements cosmologiques. Dû au nombre limité de *mocks*, le bruit dans l'estimation de la matrice de covariance doit être propagé à la mesure finale. Nous avons utilisé la fonction de vraisemblance proposée par Sellentin and Heavens (2016) pour l'analyse bayésienne (mesure RSD finale). Pour l'analyse fréquentiste, nous avons utilisé la méthode de Dodelson and Schneider (2013); Percival et al. (2014) qui consiste à corriger les erreurs de la mesure cosmologique par un facteur numérique et que nous avons étendue à la combinaison de deux fonctions de vraisemblance indépendantes (NGC et SGC).

8.4 Résultats

Même après reconstruction, le spectre de puissance montre une faible détection des BAO (1.4σ), car le signal BAO est inexistant dans le NGC. Néanmoins, nous avons montré qu'une mesure robuste de la position du signal BAO reste possible, dès lors que son amplitude est fixée a priori. Le budget d'erreurs systématiques pour l'analyse BAO est présenté dans la Table 7.1. Une partie des incertitudes systématiques provient du *mock challenge*, et des *mocks* (non-linéarité du modèle, collisions de fibres). Compte-tenu des possibles systématiques résiduelles dans l'échantillon ELG, le reste des incertitudes systématiques est estimé à partir des données elles-mêmes : systématiques photométriques et échecs de mesure des décalages spectraux. Pour cela, on prend la différence de l'ajustement cosmologique de référence avec celui obtenu sans les corrections idoines dans le catalogue de données (variations de la fonction de sélection radiale en fonction de la profondeur et pixélisation pour les systématiques photométriques, poids pour les échecs de mesure des décalages spectraux) et sans les effets systématiques idoines dans la matrice de covariance. Il est à noter qu'aucun signal BAO n'était détectable dans la moitié Nord (NGC) du relevé. Néanmoins, la position du BAO a pu être mesurée de manière relativement robuste en imposant l'amplitude des oscillations BAO dans le modèle du spectre de puissance (terme d'amortissement $\Sigma_{\rm nl}$ fixé à $4\,\mathrm{Mpc}\,h^{-1}$). Nous incluons une incertitude supplémentaire liée à ce choix dans le budget d'incertitude systématique. La mesure finale, en terme de distance comobile volumique $D_{\rm V}(z) = \left(D_{\rm M}^2(z) D_{\rm H}(z) z\right)^{1/3}$ et en incluant les incertitudes systématiques est :

$$D_{\rm V}(z_{\rm eff} = 0.845)/r_{\rm drag} = 18.33^{+0.57}_{-0.62} \tag{139}$$

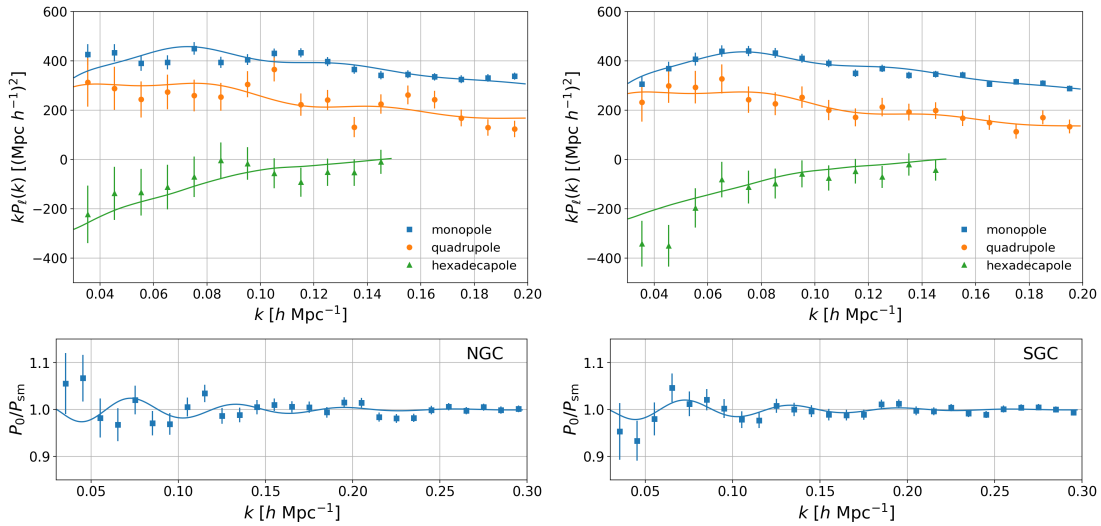

Figure 7.5 – Ajustement combiné RSD + BAO (gauche : NGC, droite : SGC) : points de mesure avec erreur venant des *mocks* EZ et meilleur ajustement du modèle (lignes continues) pour les multipoles (haut) et les oscillations BAO (bas).

source	$f\sigma_8$	$\alpha_\parallel$	$\alpha_\perp$
linéaire			
géométrie du relevé (EZ *mocks*)	0.1%	0.5%	0.1%
combinaison RSD + BAO (EZ *mocks*)	1.0%	0.3%	0.5%
non-linéarité du modèle (EZ *mocks*)	1.6%	1.0%	0.4%
modélisation des contraintes intégrales (EZ *mocks*)	0.8%	0.6%	0.1%
quadrature			
systématiques du modèle (mock challenge)	3.0%	0.9%	0.8%
terme d'amortissement $\Sigma_{\rm nl}$	1.9%	0.1%	0.8%
systématiques photométriques	9.4%	2.2%	1.4%
collisions de fibres (EZ *mocks*)	1.1%	0.6%	0.2%
échecs de mesure de décalages spectraux	6.1%	0.3%	0.3%
matrice de covariance	12.9%	4.7%	1.2%
statistique	$^{+0.060}_{-0.075}$	$^{+0.104}_{-0.090}$	$^{+0.036}_{-0.037}$
systématique	$^{+0.024}_{-0.021}$	$^{+0.029}_{-0.031}$	$^{+0.012}_{-0.011}$
total	$^{+0.085}_{-0.096}$	$^{+0.13}_{-0.12}$	$^{+0.048}_{-0.049}$

Table 7.2 – Budget d'erreur pour la mesure combinée RSD + BAO, similaire à la Table 7.1.

L'ajustement combiné RSD + BAO (avec pixelisation) est présenté sur la Figure 7.5. Le budget d'erreurs systématiques pour l'analyse combinée RSD + BAO est présenté dans la Table 7.2. Là encore, certaines incertitudes systématiques sont évaluées à partir du *mock challenge* et des *mocks* (géométrie du relevé, combinaison RSD + BAO, non-linéarité du modèle, modélisation des contraintes intégrales radiales et angulaires, collisions de fibres). Les autres sont estimées à partir des données. On notera des contributions significatives des systématiques photométriques et des échecs de mesure des décalages spectraux, estimées de la même façon que pour l'analyse BAO. Par ailleurs, l'ajustement cosmologique s'est révélé relativement instable selon la matrice de covariance utilisée, ce qui a été inclus dans le budget d'erreur systématique. La mesure finale, en terme de distances $D_{\rm H}$ et $D_{\rm M}$ et en incluant les incertitudes systématiques est $f\sigma_8(z_{\rm eff}) = 0.289^{+0.085}_{-0.096}$, $D_{\rm H}(z_{\rm eff})/r_{\rm drag} = 20.0^{+2.4}_{-2.2}$ et $D_{\rm M}(z_{\rm eff})/r_{\rm drag} = 19.17 \pm 0.99$ à $z_{\rm eff} = 0.85$.

Cette mesure a été combinée avec celle réalisée en espace de configuration (en utilisant la fonction de corrélation), comme le montre la Figure 7.6 (gauche) : les deux postérieurs ont été déplacés jusqu'à la moyenne de leur médiane, et moyennés. En incluant les incertitudes systématiques, la mesure finale

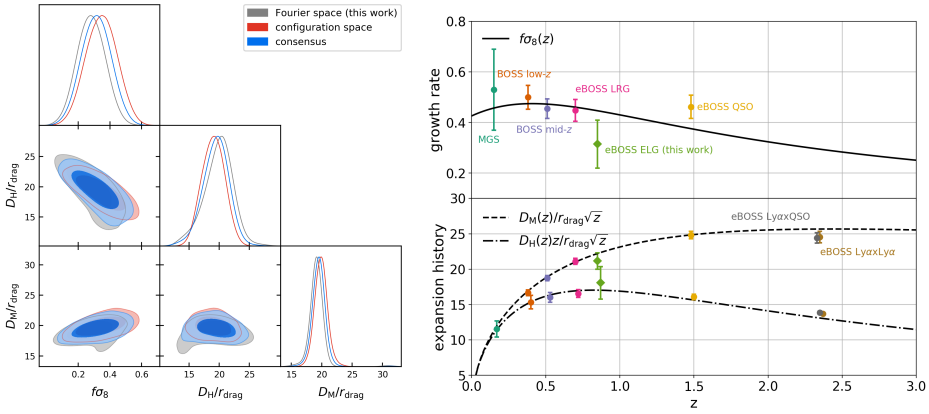

Figure 7.6 – Gauche : postérieurs des analyses en espace de Fourier et de configuration, et leur combinaison, pour l'analyse RSD + BAO (en incluant les incertitudes systématiques) des ELG d'eBOSS. Les contours contiennent 68% et 95% de la probabilité. Droite : mesures de taux de croissance des structures et du BAO par SDSS, utilisées dans l'analyse finale. Pour l'échantillon MGS ($z_{\rm eff} = 0.15$), la mesure $D_{\rm V}/r_{\rm drag}$ est montrée comme celles de $D_{\rm M}/r_{\rm drag}$. Les lignes continues sont les prédictions du modèle ΛCDM avec les paramètres de Planck Collaboration et al. (2018) issus du CMB (TT, TE, EE, lowE, lensing).

après combinaison est :

$$\begin{aligned} f\sigma_8(z_{\rm eff} = 0.85) &= 0.315 \pm 0.095 \\ D_{\rm H}(z_{\rm eff} = 0.85)/r_{\rm drag} &= 19.6^{+2.2}_{-2.1} \\ D_{\rm M}(z_{\rm eff} = 0.85)/r_{\rm drag} &= 19.5 \pm 1.0 \end{aligned} \tag{140}$$

Comme la mesure du BAO isotropique seul est en excellent accord ($< 0.2\sigma$) entre espace de Fourier et espace de configuration, et que la mesure en espace de Fourier a une barre d'erreur statistique légèrement plus faible, cette dernière a été choisie comme mesure de consensus pour l'échantillon des ELG d'eBOSS.

8.5 Résultats cosmologiques de SDSS

Les mesures des analyses BAO et RSD + BAO ont été utilisées dans l'analyse finale de SDSS, présentée dans eBOSS Collaboration et al. (2020). Cette dernière utilise les points de données montrés sur la Figure 7.6 (droite) : MGS (Ross et al., 2015; Howlett et al., 2015), BOSS low-z, BOSS mid-z (Alam et al., 2017), eBOSS LRG (Gil-Marín et al., 2020; Bautista et al., 2020), eBOSS ELG (de Mattia et al., 2020; Tamone et al., 2020), eBOSS QSO (Neveux et al., 2020; Hou et al., 2020), auto-corrélation eBOSS Lyman-α et cross-corrélation Lyman-α et QSO (du Mas des Bourboux et al., 2020).

Dans un modèle ΛCDM avec courbure libre, les mesures BAO à différents décalages spectraux permettent de briser la dégénérescence entre $\Omega_{m,0}$ et $\Omega_{K,0}$ et ainsi de mettre en évidence une constante cosmologique non nulle à 8σ. Par ailleurs, une fois combinées avec les données du CMB T&P (température et polarisation [6] de Planck Collaboration et al. 2018), les mesures BAO permettent de diviser par 10 l'incertitude sur la densité de courbure $\Omega_{K,0}$, laquelle est compatible avec 0 à une précision de 1.8×10^{-3}. Par ailleurs, dans un univers plat, les mesures BAO seules permettent de contraindre l'équation d'état de l'énergie noire w à -0.69 ± 0.15, compatible avec une constante cosmologique ($w = -1$). Enfin, des contraintes peuvent être posées sur H_0 en combinant BAO et supernovae Ia (SN, Scolnic et al. 2018) grâce à *l'échelle de distance inverse* : étant donnée une valeur de $r_{\rm drag}$ [7], les BAO fournissent une mesure des distances cosmologiques qui peut être utilisée pour calibrer la distance de luminosité des SN et contraindre H_0, en supposant un modèle cosmologique. Grâce aux points de données à différents décalages

6. c'est-à-dire TT, TE, EE et lowE
7. qui dépend de $\Omega_{c,0}h^2$ and $\Omega_{b,0}h^2$, contraints avec CMB T&P ou la nucléosynthèse primordiale

spectraux, les mesures de H_0 obtenues avec cette technique restent relativement peu sensibles au modèle de l'expansion de l'univers ; dans un modèle flexible avec courbure libre et équation d'état de l'énergie noire à deux paramètres [8], $H_0 = 67.87 \pm 0.86\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Cette mesure est à $\simeq 5\sigma$ des mesures locales (e.g. $H_0 = 73.8 \pm 1.1\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, Wong et al. 2019), ne faisant pas d'hypothèses sur la physique baryonique dans l'univers primordial.

Les mesures RSD, combinées avec CMB T&P, permettent de diviser par deux l'incertitude sur l'équation d'état de l'énergie noire, $w = -1.09 \pm 0.11$. Les RSD seules posent des contraintes dans le plan $\sigma_{8,0} - \Omega_{m,0}$, au même titre que le lentillage du CMB (Planck Collaboration et al., 2018) ou le lentillage faible des galaxies (DES, Abbott et al. 2018). Ces trois dernières sondes permettent de contraindre des déviations à la relativité générale, par exemple avec deux paramètres permettant des variations autour de l'équation de Poisson Abbott et al. (2019) : $\mu_0 = -0.04 \pm 0.25$ et $\Sigma_0 = -0.024 \pm 0.054$, compatibles avec la prédiction de la relativité générale ($\mu_0 = \Sigma_0 = 0$).

Pour conclure, il est intéressant d'analyser la réduction des incertitudes sur les paramètres cosmologiques permise par les récentes observations. Dans un modèle avec courbure, masse des neutrinos $\sum m_\nu$ et équation d'état de l'énergie noire libres, les contraintes expérimentales datant d'une dizaine d'années [9] fournissent une figure de mérite [10] dans l'espace w, $\Omega_{K,0}$, $\sum m_\nu$, H_0, et $\sigma_{8,0}$ égale à 11. En ajoutant les données du SDSS, la figure de mérite passe à 23 et en ajoutant DES et Planck, à 44. Ceci montre l'apport capital joué par les grands relevés de galaxies du SDSS dans les contraintes cosmologiques actuelles. Cette tendance ne pourra que se confirmer avec les relevés futurs, comme DESI.

9 Perspectives

L'instrument DESI, monté sur le télescope Mayall de 4 m situé à Kitt Peak, en Arizona, mesurera les décalages spectraux de 10 fois plus de galaxies et quasars que l'ensemble de SDSS, sur une surface de $14\,000\,\mathrm{deg}^2$. En particulier, les analyses RSD permettront de mesurer $f\sigma_8$, $D_\mathrm{H}/r_\mathrm{drag}$ and $D_\mathrm{M}/r_\mathrm{drag}$ à une précision relative de $\simeq 2\%$ dans 18 intervalles en décalage spectral entre $0.05 < z < 1.85$. Les mesures à bas z sont effectuées avec l'échantillon de galaxies brillantes (BGS), au-delà de $z > 0.6$ par les LRG, ELG et QSO, et les mesures du BAO à haut z par les forêts Lyman-α (auto- et cross-corrélation avec les QSO). DESI permettra de contraindre significativement l'équation d'état de l'énergie noire, l'indice spectral (relié à l'inflation), la masse des neutrinos et le paramètre $f_\mathrm{NL}^\mathrm{local}$ des non-gaussianités primordiales. Pour chacun de ces paramètres, il s'agira d'ajuster précisément la forme du spectre de puissance. Pour cela, il est nécessaire comprendre les systématiques liées à l'observation et à l'analyse à toutes les échelles.

En particulier, il sera capital d'élucider les variations tridimensionnelles de la fonction de sélection du relevé, à l'instar de celles constatées dans le relevé des ELG d'eBOSS (Section 8.2). Afin de propager correctement les incertitudes sur cette fonction de sélection lors de l'inférence cosmologique, une approche nouvellement développée, appelée *forward modelling*, apparaît prometteuse. Elle consiste à faire évoluer un champ de densité initial tridimensionnel sous l'effet de la gravité, et de comparer le champ de galaxies ainsi simulé à celui réellement observé avec une fonction de vraisemblance, laquelle est utilisée pour itérer ou échantillonner les conditions du champ de densité initial. Cette technique a été mise en oeuvre sur le relevé BOSS par Lavaux et al. (2019), qui ont mis en évidence des systématiques photométriques qui n'avaient pas été trouvées avec les analyses de BOSS. Cette technique permet en outre de produire des champs de densité et de vitesse, avec leur dispersion et de combiner différentes sondes (telles le lentillage gravitationnel). En conclusion, les relevés spectroscopiques de galaxies s'imposent peu à peu comme une sonde incontournable de la cosmologie observationnelle et leur analyse met en jeu des techniques de plus en plus sophistiquées.

8. $w = w_0 + (1-a)\,w_a$ (Chevallier and Polarski, 2001; Linder, 2003)

9. WMAP (Hinshaw et al., 2013; Bennett et al., 2013) pour le CMB, JLA (Betoule et al., 2014) pour les supernovae Ia et SDSS DR7 (Percival et al., 2010) et 2dFGRS (Colless et al., 2001) pour les mesures BAO

10. définie comme le déterminant de la matrice de covariance des $N = 5$ paramètres à la puissance $-1/(2N)$

Bibliography

Abbott, T. M. C., F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, et al. (2018). Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. Phys. Rev. D *98*(4), 043526. arXiv:1708.01530.

Abbott, T. M. C., F. B. Abdalla, S. Avila, M. Banerji, E. Baxter, et al. (2019). Dark Energy Survey year 1 results: Constraints on extended cosmological models from galaxy clustering and weak lensing. Phys. Rev. D *99*(12), 123505. arXiv:1810.02499.

Alam, S., F. D. Albareti, C. Allende Prieto, F. Anders, S. F. Anderson, et al. (2015). The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III. ApJS *219*(1), 12. arXiv:1501.00963.

Alam, S., M. Ata, S. Bailey, F. Beutler, D. Bizyaev, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. MNRAS *470*(3), 2617–2652. arXiv:1607.03155.

Alam, S., A. de Mattia, A. Tamone, S. Ávila, J. A. Peacock, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: N-body Mock Challenge for the eBOSS Emission Line Galaxy Sample. *arXiv e-prints*, arXiv:2007.09004. arXiv:2007.09004.

Alcock, C. and B. Paczynski (1979). An evolution free test for non-zero cosmological constant. Nature *281*, 358.

Avila, S., V. Gonzalez-Perez, F. G. Mohammad, A. de Mattia, C. Zhao, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: exploring the Halo Occupation Distribution model of Emission Line Galaxies. *arXiv e-prints*, arXiv:2007.09012. arXiv:2007.09012.

Baldauf, T., U. Seljak, V. Desjacques, and P. McDonald (2012). Evidence for quadratic tidal tensor bias from the halo bispectrum. Phys. Rev. D *86*(8), 083540. arXiv:1201.4827.

Bautista, J. E., R. Paviot, M. Vargas Magaña, S. de la Torre, S. Fromenteau, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1. *arXiv e-prints*, arXiv:2007.08993. arXiv:2007.08993.

Bennett, C. L., D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. ApJS *208*(2), 20. arXiv:1212.5225.

Bernardeau, F., M. Crocce, and R. Scoccimarro (2008). Multipoint propagators in cosmological gravitational instability. Phys. Rev. D *78*(10), 103521. arXiv:astro-ph/0806.2334.

Betoule, M., R. Kessler, J. Guy, J. Mosher, D. Hardin, et al. (2014). Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. A&A *568*, A22. arXiv:1401.4064.

Beutler, F., E. Castorina, and P. Zhang (2019). Interpreting measurements of the anisotropic galaxy power spectrum. J. Cosmology Astropart. Phys. *2019*(3), 040. arXiv:1810.05051.

Beutler, F., H.-J. Seo, S. Saito, C.-H. Chuang, A. J. Cuesta, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: anisotropic galaxy clustering in Fourier space. MNRAS *466*, 2242–2260. arXiv:1607.03150.

Blanton, M. R., H. Lin, R. H. Lupton, F. M. Maley, N. Young, et al. (2003). An Efficient Targeting Strategy for Multiobject Spectrograph Surveys: the Sloan Digital Sky Survey "Tiling" Algorithm. AJ *125*(4), 2276–2286. arXiv:astro-ph/0105535.

Blas, D., J. Lesgourgues, and T. Tram (2011). The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. J. Cosmology Astropart. Phys. *2011*(7), 034. arXiv:1104.2933.

Castorina, E. and M. White (2018). Beyond the plane-parallel approximation for redshift surveys. MNRAS *476*(4), 4403–4417. arXiv:1709.09730.

Chan, K. C., R. Scoccimarro, and R. K. Sheth (2012). Gravity and large-scale nonlocal bias. Phys. Rev. D *85*(8), 083509. arXiv:1201.3614.

Chevallier, M. and D. Polarski (2001). Accelerating Universes with Scaling Dark Matter. *International Journal of Modern Physics D 10*(2), 213–223. arXiv:gr-qc/0009008.

Cole, S., K. B. Fisher, and D. H. Weinberg (1995). Constraints on Omega from the IRAS redshift surveys. MNRAS *275*(2), 515–526. arXiv:astro-ph/9412062.

Colless, M., G. Dalton, S. Maddox, W. Sutherland , P. Norberg, et al. (2001). The 2dF Galaxy Redshift Survey: spectra and redshifts. MNRAS *328*(4), 1039–1063. arXiv:astro-ph/0106498.

Comparat, J., T. Delubac, S. Jouvel, A. Raichoor, J. P. Kneib, et al. (2016). SDSS-IV eBOSS emission-line galaxy pilot survey. A&A *592*, A121. arXiv:1509.05045.

Contreras, S., C. M. Baugh, P. Norberg, and N. Padilla (2013). How robust are predictions of galaxy clustering? MNRAS *432*(4), 2717–2730. arXiv:1301.3497.

Crocce, M. and R. Scoccimarro (2006). Memory of initial conditions in gravitational clustering. Phys. Rev. D *73*(6), 063520. arXiv:astro-ph/0509419.

Dawson, K. S., J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, et al. (2016). The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data. AJ *151*(2), 44. arXiv:1508.04473.

Dawson, K. S., D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, et al. (2013). The Baryon Oscillation Spectroscopic Survey of SDSS-III. AJ *145*(1), 10. arXiv:1208.0022.

de Mattia, A. and V. Ruhlmann-Kleider (2019). Integral constraints in spectroscopic surveys. J. Cosmology Astropart. Phys. *2019*(8), 036. arXiv:1904.08851.

de Mattia, A., V. Ruhlmann-Kleider, A. Raichoor, A. J. Ross, A. Tamone, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the emission line galaxy sample from the anisotropic power spectrum between redshift 0.6 and 1.1. *arXiv e-prints*, arXiv:2007.09008. arXiv:2007.09008.

DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, et al. (2016). The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, arXiv:1611.00036. arXiv:1611.00036.

Dey, A., D. J. Schlegel, D. Lang, R. Blum, K. Burleigh, et al. (2019). Overview of the DESI Legacy Imaging Surveys. AJ *157*(5), 168. arXiv:1804.08657.

Dodelson, S. and M. D. Schneider (2013). The effect of covariance estimator error on cosmological parameter constraints. Phys. Rev. D *88*(6), 063537. arXiv:1304.2593.

du Mas des Bourboux, H., J. Rich, A. Font-Ribera, V. de Sainte Agathe, J. Farr, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Baryon acoustic oscillations with Lyman-α forests. *arXiv e-prints*, arXiv:2007.08995. arXiv:2007.08995.

eBOSS Collaboration, S. Alam, M. Aubert, S. Avila, C. Balland, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory. *arXiv e-prints*, arXiv:2007.08991. arXiv:2007.08991.

Feldman, H. A., N. Kaiser, and J. A. Peacock (1994). Power-Spectrum Analysis of Three-dimensional Redshift Surveys. ApJ *426*, 23. arXiv:astro-ph/9304022.

Flaugher, B., H. T. Diehl, K. Honscheid, T. M. C. Abbott, O. Alvarez, et al. (2015). The Dark Energy Camera. AJ *150*(5), 150. arXiv:1504.02900.

Gil-Marín, H., J. E. Bautista, R. Paviot, M. Vargas-Magaña, S. de la Torre, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0. *arXiv e-prints*, arXiv:2007.08994. arXiv:2007.08994.

Górski, K. M., E. Hivon, A. J. Banday, B. D. Wand elt, F. K. Hansen, et al. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ *622*(2), 759–771. arXiv:astro-ph/0409513.

Hahn, C., R. Scoccimarro, M. R. Blanton, J. L. Tinker, and S. A. Rodríguez-Torres (2017). The Effect of Fiber Collisions on the Galaxy Power Spectrum Multipoles. MNRAS *467*(2), 1940–1956. arXiv:1609.01714.

Hand, N., Y. Feng, F. Beutler, Y. Li, C. Modi, et al. (2018). nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure. AJ *156*, 160. arXiv:1712.05834.

Hand, N., Y. Li, Z. Slepian, and U. Seljak (2017). An optimal FFT-based anisotropic power spectrum estimator. J. Cosmology Astropart. Phys. *2017*(7), 002. arXiv:1704.02357.

Hearin, A. P., A. R. Zentner, F. C. van den Bosch, D. Campbell, and E. Tollerud (2016). Introducing decorated HODs: modelling assembly bias in the galaxy-halo connection. MNRAS *460*(3), 2552–2570. arXiv:1512.03050.

Heitmann, K., H. Finkel, A. Pope, V. Morozov, N. Frontiere, et al. (2019). The Outer Rim Simulation: A Path to Many-Core Supercomputers. *arXiv e-prints*, arXiv:1904.11970. arXiv:1904.11970.

Hinshaw, G., D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. ApJS *208*(2), 19. arXiv:1212.5226.

Hou, J., A. G. Sánchez, A. J. Ross, A. Smith, R. Neveux, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from anisotropic clustering analysis of the Quasar Sample in configuration space between redshift 0.8 and 2.2. *arXiv e-prints*, arXiv:2007.08998. arXiv:2007.08998.

Howlett, C., A. J. Ross, L. Samushia, W. J. Percival, and M. Manera (2015). The clustering of the SDSS main galaxy sample - II. Mock galaxy catalogues and a measurement of the growth of structure from redshift space distortions at z = 0.15. MNRAS *449*(1), 848–866. arXiv:1409.3238.

Jackson, J. C. (1972). A critique of Rees's theory of primordial gravitational radiation. MNRAS *156*, 1P. arXiv:astro-ph/0810.3908.

Jiménez, E., S. Contreras, N. Padilla, I. Zehavi, C. M. Baugh, et al. (2019). Extensions to the halo occupation distribution model for more accurate clustering predictions. MNRAS *490*(3), 3532–3544. arXiv:1906.04298.

Kaiser, N. (1987). Clustering in real space and in redshift space. MNRAS *227*, 1–21.

Lavaux, G., J. Jasche, and F. Leclercq (2019). Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data. *arXiv e-prints*, arXiv:1909.06396. arXiv:1909.06396.

Linder, E. V. (2003). Exploring the Expansion History of the Universe. Phys. Rev. Lett. *90*(9), 091301. arXiv:astro-ph/0208512.

Linder, E. V. (2005). Cosmic growth history and expansion history. Phys. Rev. D *72*(4), 043529. arXiv:astro-ph/0507263.

McDonald, P. and A. Roy (2009). Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS. J. Cosmology Astropart. Phys. *2009*(8), 020. arXiv:0902.0991.

Neveux, R., E. Burtin, A. de Mattia, A. Smith, A. J. Ross, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from the anisotropic power spectrum of the Quasar sample between redshift 0.8 and 2.2. *arXiv e-prints*, arXiv:2007.08999. arXiv:2007.08999.

Orsi, Á. A. and R. E. Angulo (2018). The impact of galaxy formation on satellite kinematics and redshift-space distortions. MNRAS *475*(2), 2530–2544. arXiv:1708.00956.

Percival, W. J., B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, et al. (2010). Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. MNRAS *401*(4), 2148–2168. arXiv:0907.1660.

Percival, W. J., A. J. Ross, A. G. Sánchez, L. Samushia, A. Burden, et al. (2014). The clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: including covariance matrix errors. MNRAS *439*(3), 2531–2541. arXiv:1312.4841.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2018). Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, arXiv:1807.06209. arXiv:1807.06209.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2018). Planck 2018 results. VIII. Gravitational lensing. *arXiv e-prints*, arXiv:1807.06210. arXiv:1807.06210.

Raichoor, A., A. de Mattia, A. J. Ross, C. Zhao, S. Alam, et al. (2020). The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Large-scale Structure Catalogues and Measurement of the isotropic BAO between redshift 0.6 and 1.1 for the Emission Line Galaxy Sample. *arXiv e-prints*, arXiv:2007.09007. arXiv:2007.09007.

Reid, B., S. Ho, N. Padmanabhan, W. J. Percival, J. Tinker, et al. (2016). SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues. MNRAS *455*, 1553–1573. arXiv:1509.06529.

Ross, A. J., J. Bautista, R. Tojeiro, S. Alam, S. Bailey, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Large-scale structure catalogues for cosmological analysis. MNRAS *498*(2), 2354–2371. arXiv:2007.09000.

Ross, A. J., W. J. Percival, A. G. Sánchez, L. Samushia, S. Ho, et al. (2012). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: analysis of potential systematics. MNRAS *424*, 564–590. arXiv:1203.6499.

Ross, A. J., L. Samushia, C. Howlett, W. J. Percival, A. Burden, et al. (2015). The clustering of the SDSS DR7 main Galaxy sample - I. A 4 per cent distance measure at z = 0.15. MNRAS *449*(1), 835–847. arXiv:1409.3242.

Ruan, J. J., S. F. Anderson, P. J. Green, E. Morganson, M. Eracleous, et al. (2016). The Time-Domain Spectroscopic Survey: Understanding the Optically Variable Sky with SEQUELS in SDSS-III. ApJ *825*(2), 137. arXiv:1602.02752.

Saito, S., T. Baldauf, Z. Vlah, U. Seljak, T. Okumura, et al. (2014). Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum. Phys. Rev. D *90*(12), 123522. arXiv:1405.1447.

Samushia, L., W. J. Percival, and A. Raccanelli (2012). Interpreting large-scale redshift-space distortion measurements. MNRAS *420*(3), 2102–2119. arXiv:1102.1014.

Scolnic, D. M., D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, et al. (2018). The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. ApJ *859*(2), 101. arXiv:1710.00845.

Sefusatti, E., M. Crocce, R. Scoccimarro, and H. M. P. Couchman (2016). Accurate estimators of correlation functions in Fourier space. MNRAS *460*(4), 3624–3636. arXiv:1512.07295.

Sellentin, E. and A. F. Heavens (2016). Parameter inference with estimated covariance matrices. MNRAS *456*(1), L132–L136. arXiv:1511.05969.

Slepian, Z. and D. J. Eisenstein (2018). A practical computational method for the anisotropic redshift-space three-point correlation function. MNRAS *478*, 1468–1483. arXiv:1709.10150.

Smee, S. A., J. E. Gunn, A. Uomoto, N. Roe, D. Schlegel, et al. (2013). The Multi-object, Fiber-fed Spectrographs for the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey. AJ *146*(2), 32. arXiv:1208.2233.

Smith, A., A. de Mattia, E. Burtin, C.-H. Chuang, and C. Zhao (2020). Reducing the Variance of Redshift Space Distortion Measurements from Mock Galaxy Catalogues with Different Lines of Sight. *arXiv e-prints*, arXiv:2007.11417. arXiv:2007.11417.

Szalay, A. S., T. Matsubara, and S. D. Landy (1998). Redshift-Space Distortions of the Correlation Function in Wide-Angle Galaxy Surveys. ApJ *498*(1), L1–L4. arXiv:astro-ph/9712007.

Tamone, A., A. Raichoor, C. Zhao, A. de Mattia, C. Gorgoni, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Growth rate of structure measurement from anisotropic clustering analysis in configuration space between redshift 0.6 and 1.1 for the Emission Line Galaxy sample. *arXiv e-prints*, arXiv:2007.09009. arXiv:2007.09009.

Taruya, A., F. Bernardeau, T. Nishimichi, and S. Codis (2012). Direct and fast calculation of regularized cosmological power spectrum at two-loop order. Phys. Rev. D *86*(10), 103528. arXiv:1208.1191.

Wong, K. C., S. H. Suyu, G. C. F. Chen, C. E. Rusu, M. Millon, et al. (2019). H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes. *arXiv e-prints*, arXiv:1907.04869. arXiv:1907.04869.

Wright, E. L., P. R. M. Eisenhardt, A. K. Mainzer, M. E. Ressler, R. M. Cutri, et al. (2010). The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. AJ *140*(6), 1868–1881. arXiv:1008.0031.

Yamamoto, K., M. Nakamichi, A. Kamino, B. A. Bassett, and H. Nishioka (2006). A Measurement of the Quadrupole Power Spectrum in the Clustering of the 2dF QSO Survey. PASJ *58*, 93–102. arXiv:astro-ph/0505115.

York, D. G., J. Adelman, J. Anderson, John E., S. F. Anderson, J. Annis, et al. (2000). The Sloan Digital Sky Survey: Technical Summary. AJ *120*(3), 1579–1587. arXiv:astro-ph/0006396.

Zhu, G. B., J. Comparat, J.-P. Kneib, T. Delubac, A. Raichoor, et al. (2015). Near-ultraviolet Spectroscopy of Star-forming Galaxies from eBOSS: Signatures of Ubiquitous Galactic-scale Outflows. ApJ *815*(1), 48. arXiv:1507.07979.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Particules, hadrons, énergie et noyau : instrumentation, imagerie, cosmos et simulation (PHENIICS)

Titre: Des relevés spectroscopiques de galaxies à raies d'émission aux contraintes cosmologiques : d'eBOSS à DESI

Mots clés: cosmologie, relevé spectroscopique de galaxies, croissance des structures, distorsions dans l'espace des décalages spectraux, oscillations acoustiques de baryons, collaboration eBOSS

Résumé: Les relevés spectroscopiques de galaxies sont riches d'information cosmologique. Les ondes acoustiques qui se sont propagées dans l'Univers primordial ont laissé une signature dans la distribution de matière, appelée oscillations acoustiques de baryons (BAO), à une échelle caractéristique de 150 mégaparsecs. Mesurer cette échelle dans la distribution des galaxies permet de sonder le taux d'expansion de l'Univers au cours de son histoire. Par ailleurs, les mesures des décalages spectraux des galaxies sont sensibles à leur vitesse particulière (RSD), permettant par là même de mesurer le taux de croissance des structures et de tester la relativité générale à grande échelle. Cette thèse est consacrée à la mesure par spectre de puissance des BAO et des RSD dans l'échantillon de 173 736 galaxies à raies d'émission (ELG) de la collaboration eBOSS, à un décalage spectral effectif de 0.85. Un soin particulier a été apporté à la validation de l'implémentation du modèle théorique du spectre de puissance des galaxies, et à l'estimation et correction des systématiques observationnelles, grâce à des simulations réalistes. Les principales systématiques observationnelles de ce relevé proviennent des fluctuations de la densité de cibles avec la qualité de l'échantillon photométrique utilisé pour eBOSS, dont une version plus récente est exploitée pour le relevé spectroscopique de nouvelle génération DESI. En particulier, en plus de notables fluctuations dans la densité angulaire de cibles, de fortes variations de densité en décalage spectral ont été notées en fonction de la profondeur de l'imagerie, un effet mineur dans les autres relevés d'eBOSS. Les variations résiduelles de densité angulaire ont été atténuées en supprimant le contraste de densité au-delà d'une certaine échelle, tandis que les fluctuations dans la fonction de sélection radiale ont été prises en compte en divisant le relevé en sous-ensembles de profondeur photométrique équivalente. Ce faisant, le spectre de puissance mesuré est amorti à grande échelle, ce qui nécessite de corriger la prédiction théorique par des termes dits de contrainte intégrale, un projet original de cette thèse, qui a aussi permis d'améliorer certains procédés des analyses réalisées jusqu'alors. Les mesures RSD et BAO obtenues avec les ELG d'eBOSS sont combinées avec celles des autres relevés de grandes structures de la collaboration SDSS, et les implications cosmologiques sont exposées. Cette thèse se termine par un bilan des différentes systématiques observationnelles et d'analyse entachant la mesure cosmologique, et de pistes pour contrôler le budget systématique des futurs relevés spectroscopiques, comme DESI.

Title: From emission line galaxy spectroscopic surveys to cosmological constraints: from eBOSS to DESI

Keywords: cosmology, spectroscopic galaxy survey, growth of structure, redshift space distortions, baryon acoustic oscillations, eBOSS collaboration

Abstract: Spectroscopic galaxy surveys contain a wealth of cosmological information. Acoustic waves that propagated in the primordial Universe left a signature on the matter distribution, called baryon acoustic oscillations (BAO), at a characteristic scale of 150 megaparsecs. Measuring this standard ruler in the distribution of galaxies provides access to the expansion history of the Universe. In addition, redshift measurements of galaxies are sensitive to their peculiar velocities (RSD), allowing to probe the growth rate of structure and test general relativity on large scales. This thesis is dedicated to the power spectrum analysis of BAO and RSD in the sample of 173,736 emission line galaxies (ELGs) of the eBOSS collaboration, at an effective redshift of 0.85. Special care was devoted to validate the implementation of the galaxy power spectrum theoretical model, and to estimate and correct observational systematics, with the help of realistic simulations. The main observational systematics of this sample stem from fluctuations of the galaxy target density with the quality of the photometric sample used for eBOSS, which is an early version of that utilised in the next generation spectroscopic survey DESI. In particular, besides the noticeable fluctuations of the angular density of targets, strong variations in the redshift density with imaging depth were noted - an effect which is usually assumed to be minor. Residual variations of the angular target density were mitigated by suppressing the density contrast above a certain scale, while fluctuations in the radial survey selection function were accounted for by dividing the data set in subsamples of equivalent photometric depth. Doing so, the measured galaxy power spectrum is damped at large scale, which requires to correct the theoretical prediction for the so-called integral constraints, an original work of this thesis that also allowed to improve some techniques of clustering analyses. The RSD and BAO measurements obtained with eBOSS ELGs are combined with the results from the other clustering samples of the SDSS collaboration, and cosmological implications are presented. This manuscript ends with a recap of different observational and analysis systematics hindering the cosmological measurements, and gives ideas to control the systematic budget of future spectroscopic surveys, like DESI.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France