

..... NNT/NL : 2020AIXM0001/001ED000

Habilitation à Diriger des Recherches

Présentée à Aix-Marseille Université

le XX janvier 2022 par

Julián Ernesto BAUTISTA

Dark-energy with Spectroscopic Observations of the Universe

Discipline

Physique et Sciences de la Matière

Spécialité

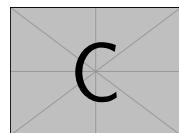
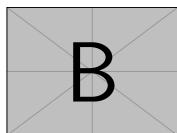
Astrophysique et Cosmologie

Laboratoire

Centre de Physique des Particules de Marseille

Composition du jury

•	Prénom NOM	Rapporteur·e
•	Affiliation	
•	Prénom NOM	Rapporteur·e
•	Affiliation	
•	Prénom NOM	Examinateur·rice
•	Affiliation	
•	Prénom NOM	Président·e du jury
•	Affiliation	
	Prénom NOM	Tuteur
	Affiliation	



Je soussigné, Julián Ernesto Bautista, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille, le [date]



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International](#).

Résumé

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Mots clés: géométrie algorithmique, complexe planaire et rectangulaire, géodésique, courbure globale non-positive

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Keywords: computational geometry, planar and rectangular complex, geodesic, global nonpositive curvature

Acknowledgements

Le modèle de thèse AMU n'existerait pas sans la contribution des doctorants. Nous souhaitons remercier tout particulièrement [Mickaël Bojados](#), [Flora Cordoleani](#) et [Florian Caullery](#) pour leur aide précieuse et la qualité de leurs fichiers sources LaTeX. La mise à jour effectuée en 2018 doit beaucoup à l'excellent travail de [Dorian Depriester](#).

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placeat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Contents

Résumé	3
Abstract	4
Acknowledgements	5
Contents	6
1 Introduction	8
1.1 The history of our Universe and its big open questions	8
1.2 Evidences for the accelerated expansion of the Universe	9
1.3 Model of an expanding Universe	10
1.3.1 Assumptions and ingredients	10
1.3.2 The expansion rate	10
1.3.3 Distances in cosmology	11
1.3.4 Dark-energy models	12
1.4 Model of the large-scale structures	12
1.4.1 Statistical description of perturbations	13
1.4.2 Configuration and Fourier space	14
1.4.3 Cosmological dependency of the power spectrum	14
1.4.4 The amplitude of the power spectrum	15
1.5 Cosmological probes of expansion	16
1.5.1 Direct measurements of H_0	16
1.5.2 Type-Ia supernovae	17
1.5.3 Big Bang nucleosynthesis	18
1.5.4 Baryon acoustic oscillations	19
1.5.5 The cosmic microwave background	20
1.5.6 Redshift-space distortions	21
1.5.7 Weak gravitational lensing	23
2 Observing the Universe with spectroscopy	24
2.1 Selecting the objects to observe	24
2.2 Pointing fibres to the sky	25
2.3 From electrons to spectra	26
2.4 From spectra to redshifts	28
2.5 From redshifts to cosmology	28
3 The high-redshift Universe and its forests	29
3.1 Forests as a tracer of the Universe's structures	29
3.2 Baryon acoustic oscillations in the forests	31
3.2.1 Transmission field	32
3.2.2 Two-point correlation functions	33

3.2.3	Correction matrices	34
3.2.4	The model	35
3.2.5	BAO constraints	36
3.3	Impact of redshift errors	38
3.4	Weak-lensing of forests	41
3.5	Voids in tomographic maps	44
4	The mid-redshift Universe and its galaxies	45
4.1	From redshifts to clustering catalogues	45
4.2	Reconstruction of linear density field	50
4.3	Mock catalogues	51
4.4	From catalogues to clustering estimates	52
4.4.1	Correlation function	52
4.4.2	Power spectrum	53
4.4.3	The clustering of eBOSS DR16 LRGs	55
4.5	Baryon acoustic oscillations	57
4.6	Redshift-space distortions	59
4.7	Joint clustering analysis in Fourier and Configuration space	63
4.8	Cross-correlation with radio surveys	66
5	The low-redshift Universe and its velocities	69
5.1	Measuring peculiar velocities	69
5.1.1	The Tully-Fisher relation of spiral galaxies	70
5.1.2	The Fundamental Plane of elliptical galaxies	71
5.1.3	Type-Ia supernovae	72
5.2	Biases affecting velocities	74
5.3	Growth-rate measurements with velocities	75
5.3.1	Maximum likelihood method	76
5.3.2	Compressed two-point statistics	76
5.3.3	Density-velocity comparison	78
5.3.4	Forward-model of density and velocity field	78
5.4	Current measurements	79
5.5	DESI and ZTF	79
Conclusion		82
Bibliography		83

1 Introduction

Contents

1.1	The history of our Universe and its big open questions	8
1.2	Evidences for the accelerated expansion of the Universe	9
1.3	Model of an expanding Universe	10
1.3.1	Assumptions and ingredients	10
1.3.2	The expansion rate	10
1.3.3	Distances in cosmology	11
1.3.4	Dark-energy models	12
1.4	Model of the large-scale structures	12
1.4.1	Statistical description of perturbations	13
1.4.2	Configuration and Fourier space	14
1.4.3	Cosmological dependency of the power spectrum	14
1.4.4	The amplitude of the power spectrum	15
1.5	Cosmological probes of expansion	16
1.5.1	Direct measurements of H_0	16
1.5.2	Type-Ia supernovae	17
1.5.3	Big Bang nucleosynthesis	18
1.5.4	Baryon acoustic oscillations	19
1.5.5	The cosmic microwave background	20
1.5.6	Redshift-space distortions	21
1.5.7	Weak gravitational lensing	23

1.1 The history of our Universe and its big open questions

If we gather all knowledge in physics humanity could learn to this day, we can build an interesting story for how our Universe evolved since *very* early times. This great story that can explain most of what we observe in the sky, while being consistent with experimental results here on Earth. However, this story has several weak spots. Either because where we do not have a satisfactory physical explanation for what we observe, or because we simply cannot observe what happens by then (and potentially we never will).

The story in a nutshell goes as follows: the “beginning” of the Universe¹, also known as the *Big Bang* or *inflation*, was an epoch of exponential expansion of space that happened around 13.8 billion years ago. Inflation transformed quantum, microscopic fluctuations of space into macroscopic density fluctuations of the Universe’s constituents: quarks, photons,

¹The actual beginning of the Universe can be seen as a complex philosophical question or simply as an ill defined physical concept. This discussion is clearly beyond the scope of this work and/or my capabilities. Here, I decide to name as beginning the first epoch of the Universe for which we have a widely known physical theory to describe it.

neutrinos and dark-matter. After inflation, the Universe continues to expand but much slowly. The average temperature of the Universe decreases adiabatically. Quarks start to gather to form protons, neutrons and relics of heavier atoms, such as deuterium, tritium, helium, lithium and so on. All this during the first few minutes of the Universe's history, in the so called *Big Bang nucleosynthesis*. After 380 thousand years, electrons rebind with protons to form neutral hydrogen, and photons scatter for their last time with them as the Universe becomes more diffuse. The photons from this epoch are observable today and are known as the *cosmic microwave background*. The Universe goes through its *dark ages*, where hydrogen is mostly neutral and stars have not yet formed. After several million years, gravity clumps hydrogen into dense clouds that reach high enough temperatures to start thermonuclear reactions in their core. The first stars are formed. Then the first galaxies. These galaxies merge into larger galaxies, galaxies form small virialized groups, large groups, clusters, super clusters. Clusters, filamentary structures and voids compose the so called cosmic web. At around half of the Universe's age, the expansion starts to speed up again, accelerating, as if gravity became repulsive on very large scales. Then, on one of the many billions of galaxies, the *Milky Way*, planet Earth formed around a somewhat isolated star, the Sun, and here we are today, observing the sky, trying to explain the whole Universe.

This story is based on the assumption that general relativity (GR) is the theory of gravity, that the Universe can be modelled as statistically homogeneous and isotropic on large enough scales, and that the Universe is composed of the standard model particles and interactions, with the addition of two exotic components: dark-energy and dark-matter.

The physical origin of dark-energy and dark-matter is a mystery and one of the weak spots of the Universe's story. Another weak spot is the physics of inflation, but observables related to that time are limited. [Any others ?] Solutions for these problems would represent major breakthroughs in physics and large teams of scientists are dedicated to them, either theoretically or experimentally.

My past work has been focused on the observations related to dark-energy. The rest of this chapter will be dedicated to explaining the problem of accelerated expansion, the dark-energy model and other possible solutions, and how do we observe the expansion.

1.2 Evidences for the accelerated expansion of the Universe

Not long after developing his theory of gravity, Einstein wanted to write a model for a static Universe filled with matter. Since gravity is attractive, his solution was to include a new constant term, Λ , in the equations. This will be later known as the *cosmological constant*, one of the simplest models of dark-energy in agreement with a wide variety of cosmological observations.

The first indirect evidences for dark-energy come from measurements of galaxy clustering in the 80's (Maddox et al. 1990; Efstathiou et al. 1990). At the end of the 90's, two independent teams measured the expansion of the Universe using type-Ia supernovae (Riess et al. 1998; Perlmutter et al. 1999). Their observations could only be explained by an Universe containing around 30 per cent of matter and 70 per cent of dark energy, in the form of a cosmological constant. Few years later, first measurements of the temperature fluctuations in the cosmic microwave background were consistent with an Universe with a flat space (Balbi et al. 2000; de Bernardis et al. 2000), which in combination with measurements of clustering, supernovae and the local expansion rate (Mould et al. 2000), definitely confirmed that the Universe's expansion is accelerating. Age estimates of globular stellar clusters, which are supposed to be among the

oldest astrophysical objects, were also indicating that the Universe had to be older than the age predicted by models of an Universe only filled with matter (Chaboyer 1998). The following decade was enriched by the first measurements of the baryon acoustic oscillations in the distribution of galaxies (Eisenstein et al. 2005; Cole et al. 2005), re-confirming the accelerated nature of the expansion.

Since the 2010's, we live in the so-called *precision cosmology* era, where several observatories were constructed with the main goal of improving the precision of all these measurements just discussed. Even with the most up-to-date results, with errors of the order of a few percent, there is yet no evidence for a departure from the model of Universe governed by GR, composed by 70 per cent of this mysterious cosmological constant and by 25 per cent by this mysterious dark-matter.

1.3 Model of an expanding Universe

1.3.1 Assumptions and ingredients

In order to define some important cosmological parameters, I will quickly review the basics of the current most accepted model for our Universe in the limit of very large scales, where it can be considered homogenous and isotropic. This is also known as the *background* model, since the formation of structures (see next section) can be modeled as small perturbations around it. It is also assumed that gravity is described by general relativity, which interlinks the fabric of space-time and energy densities of the Universe constituents. For these constituents, we assume the presence of *baryons* (protons, electrons, atoms), photons, neutrinos, cold dark-matter and dark-energy.

The expansion of space is usually parametrised by the scale factor $a(t)$, a function of time t that dictates how physical and comoving distances relate to each other. We assume that $a(t_0) = 1$ today ($t_0 = 13.8$ Gyr) and that $a \rightarrow 0$ as $t \rightarrow 0$. The speed of the expansion and its acceleration are simply derivatives with respect to time, $\dot{a}(t)$ and $\ddot{a}(t)$, respectively. The expansion rate of the Universe is defined as $H(t) = \dot{a}(t)/a(t)$ and takes the Hubble constant value today, $H(t_0) = H_0 = 100h$ km/s/Mpc ~ 70 km/s/Mpc. Since the scale factor is a monotonically increasing function of time (in our simple model), a is commonly used to describe a given cosmic epoch.

One of the consequences of the expansion is that relativistic species, such as photons and massless neutrinos, loose energy or are *redshifted* as they propagate. In astronomy, the redshift z is defined as the relative difference in wavelength between the one emitted by source and the one observed. Thus, in an expanding Universe the scale factor is related to the redshift as $a = 1/(1+z)$. Redshift is also commonly used to describe cosmic epoch or distances. Today corresponds to $z = 0$ while $z \rightarrow \infty$ as $t \rightarrow 0$.

1.3.2 The expansion rate

The expansion rate can be derived from GR equations and be written as a function of redshift and the densities of each constituent

$$H^2(z) = H_0^2 [\Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_{de}(z) + \Omega_k(1+z)^2] \quad (1.1)$$

where $\Omega_x = \rho_x/\rho_{\text{crit}}$ is the ratio between the density constituent x today and the critical energy density $\rho_{\text{crit}} = 3H_0^2/8\pi G$, which is the density required for a flat space geometry. The

subscripts m , r , de and k stand for non-relativistic species (baryons, cold dark-matter and non-relativistic neutrinos), relativistic species (photons and relativistic neutrinos), dark-energy and curvature, respectively. The curvature “density” parameter is defined as a function of the others, $\Omega_k = (1 - \sum_x \Omega_x)$, and it is zero in a flat space geometry. Non-relativistic species dilute in an expanding universe so their energy density is proportional to $a^{-3} = (1 + z)^3$. Relativistic species dilute as well but have an extra factor of a^{-1} due to redshifting. Dark-energy is written as a general function of redshift. In the simplest case of a cosmological constant, $\Omega_{de}(z) = \Omega_\Lambda$, i.e. a constant. The dependency of curvature with a^{-2} simply comes from the field equations.

1.3.3 Distances in cosmology

Distances are non-trivially defined in an expanding, potentially non-flat, Universe. Therefore, different types of distances are more appropriate for certain observables. The *Hubble distance* is related to the expansion rate as

$$D_H(z) = \frac{c}{H(z)}, \quad (1.2)$$

and it is commonly used in measurements of baryon acoustic oscillations in the line-of-sight direction (see Chapters 3 and 4) or to represent the size of a causally connected region of the Universe.

The *comoving distance* to an object at redshift z is written as

$$D_C(z) = \int_0^z dz' D_H(z'). \quad (1.3)$$

This expression yields the distance travelled by a photon from a source towards us but factoring out the scale factor, effectively “removing” the effect of expansion.

The *comoving angular diameter distance* $D_M(z)$ is useful when distances are inferred from angles, which are affected by the curvature of space. An object with comoving size l at a redshift z would be seen with an angular size θ . This allows us to define $D_M(z) = l/\theta(z)$ as

$$D_M(z) = D_C(z) \text{sinc}\left(\sqrt{-\Omega_k} \frac{D_C(z)}{D_H(z=0)}\right), \quad (1.4)$$

where $\text{sinc}(x) = \sin(x)/x$. This distance is used in observations of gravitational lensing and in measurements of the baryon acoustic oscillations in the transverse direction to the line-of-sight.

Analogously, the *luminosity distance* $D_L(z)$ relates the flux f received by an object at redshift z with intrinsic luminosity L . We define then $D_L(z) = \sqrt{L/4\pi f(z)}$ as

$$D_L(z) = D_M(z)(1+z). \quad (1.5)$$

This distance is used in supernova cosmology, where fluxes are used to infer dark-energy properties. Since supernovae fluxes vary by orders of magnitude, it is handy to define the *distance modulus* as the logarithm of the luminosity distance:

$$\mu(z) = 5 \log_{10} \frac{D_L(z)}{10 \text{ pc}}. \quad (1.6)$$

This definition is such that it agrees with the definition of magnitude of an object.

In all the distances defined above, the dependency to the cosmological energy densities Ω_x

happens only through $H(z)$ (Eq. 1.1). This is how we can constrain these parameters through distance measurements. However, all cosmological distance measurements are *relative*, which means they are defined to an arbitrary normalisation². It is only by observing the *evolution* of these relative distance measurements as a function of redshift that we can constrain cosmological parameters. Note that in all the distances defined above, the dependency with H_0 only impacts this arbitrary normalisation, so it cannot be measured directly. It is useful to factor out this dependency with H_0 , and write distances in units of $h^{-1}\text{Mpc}$ (numerically equivalent to set $H_0 = 100 \text{ km/s/Mpc}$). These are the units used in all works studying the large-scale structures and it is also used throughout this manuscript.

1.3.4 Dark-energy models

In the Eq. 1.1 for the expansion rate of the Universe, the dark-energy density term is written as a general function of redshift. As mentioned before, the simplest model is to consider a cosmological constant $\Omega_{de}(z) = \Omega_\Lambda$. The cosmological constant can be thought as equivalent to a fluid with relativistic pressure $p_{de} = -\rho_{de}$. This model can be extended by considering a different equation of state $p_{de} = p_{de}(\rho_{de}) = w(z)\rho_{de}$, where $w(z)$ can be a constant or a more general function of redshift, under the condition that $w(z) < -1/3$, needed to obtain an accelerated expansion in a dark-energy dominated Universe. One widely known parametrisation is given by $w(a) = w_0 + w_a(1 - a)$ (Chevallier and Polarski 2001; Linder 2003).

The litterature contains a huge variety of models that attempt to be more physically motivated than those just presented (see Weinberg et al. 2013, section 2.2 for a review). Some of them suggest that maybe general relativity breaks down on large enough scales, causing the expansion to accelerate. They suggest extensions or modifications to Einstein's theory of general relativity, and are commonly known as *modified gravity* models.

1.4 Model of the large-scale structures

The Universe is clearly not homogenous and isotropic. Matter clustered under the influence of gravity, creating the cosmic web, composed of clusters, filaments, empty regions. The model described in the previous section only describes the Universe as a whole, without any inhomogeneities. If we want to describe the evolution of the structures we need to consider an extension to the background model.

Observations of the large-scale structures and the cosmic microwave background (CMB) indicate that the matter density field today is the evolution of tiny density perturbations from the early Universe. Temperature fluctuations in the CMB are of the order of 10^{-5} , while densities today can reach values several orders of magnitude larger than the average density. The most widely accepted theory is that the CMB fluctuations, and today's structures, are originated from the inflation-grown quantum fluctuations, that evolved under the influence of gravity and pressure in the hot dense plasma-like epoch of our Universe.

It is been possible to model the physics of the evolution of these tiny fluctuations using *linear perturbation theory*. The predictions of this theory are an excellent match to observations of the CMB and to the late-time distribution of matter on large scales, i.e., larger than few tens of Mpc. On smaller scales however, gravity becomes highly non-linear and more advanced calculations

²Two exceptions are parallax distances which are an absolute measurements based on the known distance between the Earth and the Sun and gravitational wave distances, which amplitudes depend on the distance and on the masses of the progenitors, though the masses affect the wave form as well, breaking the degeneracy.

are required. There are two main approaches to model non-linearities: theoretical calculations going beyond linear terms in perturbation theory or numerical n-body simulations. [Develop?]

The main idea of perturbation theory is to consider that the density of a given species x is given by:

$$\rho_x(\vec{x}, t) = \bar{\rho}_x(t) [1 + \delta_x(\vec{x}, t)], \quad (1.7)$$

where the bar indicates average over the whole Universe and $\delta_x \ll 1$ is a small perturbation. The average density $\bar{\rho}_x$ follows the background evolution described in section 1.3. A new set of equations can be derived for the evolution of δ_x for each species depending if x denotes relativistic species (photons, hot baryons, hot neutrinos) or non-relativistic species (cold baryons, cold neutrinos or dark-matter). Only some families of models also consider fluctuations in the dark-energy density.

The evolution of δ_x for a given species is dictated by both Boltzmann and Einstein's GR equations. Boltzmann equations describe the evolution of phase-space distributions of each constituent, considering collisions (pressure) and particle creation and annihilation. GR equations describe how each species influence gravitational potentials which in turn make perturbations grow.

Another important ingredient in these equations is the *peculiar velocities* $\vec{v}_x(\vec{x}, t)$ of each species, particularly the non-relativistic ones since relativistic species are assumed to have $v \sim c$. They influence the evolution of perturbations through Boltzmann, Euler and continuity equations. These velocities are observable and are an important probe of the cosmological model and on the strength of gravity itself. They will be a key actor in Chapters 4 and 5.

1.4.1 Statistical description of perturbations

The value of the matter density field at a given position \vec{x} is hardly observable and therefore does not contain much cosmological information, since it is just the evolution of a single random realisation of the initial conditions. In cosmology we are mostly interested by the statistical properties of the density field, or of its perturbations. Particularly, most cosmological information is contained in the first moments of the density field, starting from the two-point function (the one-point function is zero by definition), which is defined as

$$\xi(\vec{x}, \vec{y}) = \langle \delta(\vec{x}) \delta(\vec{y}) \rangle \quad (1.8)$$

where $\langle \cdot \rangle$ denotes an ensemble average. Given that we only have one single realisation of the Universe, we assume that averages over space (in an infinite space) are equivalent to ensemble averages.

The assumptions of homogeneity and isotropy of the background can be applied in a statistical sense to the perturbations. Statistical homogeneity makes ξ not to depend on the specific locations \vec{x} and \vec{y} but only on their separation $\vec{r} = \vec{y} - \vec{x}$. Statistical isotropic makes ξ no longer depend on the orientation of \vec{r} , only on its absolute value $r = |\vec{r}|$. Therefore, the two-point correlation function simplifies to a function of r :

$$\xi(r) = \langle \delta(\vec{x}) \delta(\vec{x} + \vec{r}) \rangle. \quad (1.9)$$

The correlation function $\xi(r)$ is a powerful observable which can be compared to predictions from a given cosmological model.

1.4.2 Configuration and Fourier space

The time evolution of linear density perturbations $\delta(\vec{x})$ is dictated by a set of second-order differential equations, containing derivatives with respect to time and space. It is thus convenient to apply Fourier transforms to these equations, where derivatives with respect to space become simple products. Separations \vec{r} in the so-called *Configuration space* become wavevectors \vec{k} in *Fourier space*. The density perturbations in Fourier space are defined by

$$\tilde{\delta}(\vec{k}) = \int d^3x e^{-i\vec{k}\cdot\vec{x}} \delta(\vec{x}), \quad (1.10)$$

and its inverse is

$$\delta(\vec{x}) = \frac{1}{(2\pi)^3} \int d^3k e^{+i\vec{k}\cdot\vec{x}} \tilde{\delta}(\vec{k}). \quad (1.11)$$

As previously, we are interested in the two (or more) point functions of the field. In Fourier space, the two-point correlation function is known as the *power spectrum* and is defined as

$$\langle \tilde{\delta}(\vec{k}) \tilde{\delta}(\vec{k}') \rangle = (2\pi)^3 P(k) \delta_{\text{Dirac}}^3(\vec{k} - \vec{k}'), \quad (1.12)$$

where $\delta_{\text{Dirac}}^3(\vec{x})$ is a three-dimensional Dirac distribution. The assumption of spatial homogeneity is guaranteed by the properties of Fourier space, while the assumption of isotropy makes the power spectrum to be only a function of the absolute value of the wavevector $k = |\vec{k}|$.

When considering the evolution of perturbations in Fourier space to the linear level, each mode $\tilde{\delta}(\vec{k})$ evolves independently of the others. We often say that there is no *mode coupling* in linear theory.

The power spectrum $P(k)$ is an observable, as powerful as the correlation function, to constrain cosmological models. While in theory these two contain exactly the same compressed statistical information, in practice the analyses of real data use different finite ranges of scales which are affected differently by systematic effects. Therefore, analyses of real data in configuration and Fourier space complement each other. Such analyses are presented in Chapter 4.

1.4.3 Cosmological dependency of the power spectrum

The power spectrum $P(k)$, or the correlation function $\xi(r)$, of density perturbations are an excellent probe of the cosmological model. It encodes information accumulated of the whole Universe's history, since the Big Bang until today.

At the end of inflation, we think that perturbations were roughly Gaussian (as are quantum fluctuations of the vacuum), having a nearly scale independent power spectrum defined by

$$P_0(k) = A_s k^{n_s - 1}, \quad (1.13)$$

where $A_s \sim 2 \times 10^{-9}$ is the amplitude of scalar perturbations and $n_s \sim 0.96$ is the scalar spectral index. The term scalar refers to standard density (or gravitational potential) perturbations, while tensorial perturbations (of the space-time metric) refer to primordial gravitational waves. Vectorial perturbations decay and rapidly become negligible in most common cosmological models.

After inflation, the majority of the energetic budget of Universe was held by photons and neutrinos. Since their energy density decays as a^{-4} , their contribution quickly drops. This radiation-dominated era lasted until $z \sim 5000$, or $t \sim 20$ kyr, when the energy density of matter

became the dominant source. Matter density decays at a slower rate, proportional to a^{-3} . Dark-energy became dominant at $z \sim 0.5$, or $t \sim 9$ Gyr, since its energy density is constant (or close to constant). The actual duration of each era depends on relative values of the energy densities, parametrised by Ω_x (see 1.3.2).

Add figure Matter perturbations grew at different rates during these different eras. In radiation-dominated era, large-scale modes (large values of r or small values of k) grew while small-scale ones oscillated, due to battle between gravity and radiative pressure. These are known as *baryon acoustic oscillations*. These oscillations are imprinted in the power spectrum of the photon-baryon fluid, at the time of the CMB, and in the matter power spectrum at later times. The radiative pressure prohibited the growth of small-scale perturbations, so the matter power spectrum is damped, which results in this hill-like shape. In the matter-dominated era, the radiative pressure becomes negligible and all modes grow equally (neglecting non-linear growth). When dark-energy is dominant, the expansion accelerates and growth of structures slows down slightly.

Well after recombination, matter (baryons + dark) density perturbations obey the following differential equation, if we assume GR and linear theory:

$$\ddot{\delta}(\vec{x}, t) + 2H(z)\dot{\delta}(\vec{x}, t) - \frac{3}{2}\Omega_m H_0^2(1+z)^3\delta(\vec{x}, t) = 0. \quad (1.14)$$

If we assume that $\delta(\vec{x}, t) = \delta(\vec{x}, t_0)G(t)/G(t_0)$, we can factor out the spatial dependency and only solve for the time-dependent term $G(t)$, known as the *growth factor*. Its logarithmic derivative with respect to the scale factor $a(t)$ is called *growth rate* of structures:

$$f(a) \equiv \frac{d \ln G(a)}{d \ln a}, \quad (1.15)$$

which is a key observable for which the predicted value depends on the cosmological parameters and on the assumed theory of gravity (e.g., GR in Eq. 1.14).

In linear theory, peculiar velocities are simply related to density perturbations via the continuity equation

$$\nabla \cdot \vec{v}(\vec{x}, a) \equiv \theta(\vec{x}, a) = -a^2 H(a) \frac{d\delta(\vec{x}, a)}{da} = -aH(a)f(a)\delta(\vec{x}, a) \quad (1.16)$$

where θ is the velocity divergence, a convenient scalar field describing velocities. In Fourier space, the velocity field is simply proportional to the density. The velocity divergence power spectrum $P_{\theta\theta}$ is an important ingredient for models of redshift-space distortions, as we discuss in the following section.

1.4.4 The amplitude of the power spectrum

The matter power spectrum $P(k)$ represents the variance of density perturbations at a given wavenumber k . The total variance of the density field is therefore the sum of all contributions over all available three dimensional modes. This is equivalent to the correlation function at zero separation:

$$\sigma^2 = \xi(r=0) = \frac{1}{(2\pi)^3} \int d^3 k P(k) = \frac{1}{2\pi^2} \int_0^\infty dk k^2 P(k), \quad (1.17)$$

which diverges if we use the linear prediction for the matter power spectrum. This is because the linear power spectrum is the evolution of a nearly scale invariant power spectrum, equivalent to white noise, so the variance simply keeps increasing as we consider larger ranges of scales. Therefore, it is convenient to smooth the linear density field using a three-dimensional top-hat filter with radius R . The variance of the smoothed field is then

$$\sigma_R^2 = \frac{1}{2\pi^2} \int_0^\infty dk k^2 P(k) W_R^2(k), \quad (1.18)$$

where $W_R(k) = 3[\sin(kR) - kR \cos(kR)]/(kR)^3$ is the Fourier transform of the top-hat filter.

In studies of large-scales structures, it is common to use σ_R as a parameter defining the amplitude of the linear power spectrum. As discussed before, this amplitude depends on the values of A_s (Eq. 1.13) and h (which determines the age of the Universe and how long structures could have grown). Therefore σ_R is degenerate with A_s and h though is more representative of the amplitude of $P(k)$ at later times. Other energy density parameters Ω_x also affect this amplitude, but these also change the overall shape and are not degenerate with σ_R .

The variance σ_R is a decreasing function of R . Historically, the chosen values for R give variances near unity, which correspond to the regime where linear theory should break. Recently, the value of $R = 8h^{-1}\text{Mpc}$ has been used in several analysis, though it has been argued that choosing $R = 12 \text{ Mpc}$ (without the h dependency) is a better choice to break degeneracies (Sánchez 2020).

In measurements of redshift-space distortions and peculiar velocities, there is a degeneracy between the value of f and the amplitude of the power spectrum, σ_8 . Therefore, measurements can only constrain the combination $f\sigma_8$.

1.5 Cosmological probes of expansion

In this section, I present the basics of the cosmological observables that allow us to learn about dark-energy. Some of these are mostly sensitive to the background evolution presented in section 1.3.2, while others depend on the matter perturbations and their statistical properties as discussed in section 1.4.

1.5.1 Direct measurements of H_0

The Hubble constant H_0 is probably one of the first cosmological parameters to be estimated by the one giving its name (Hubble 1929). Today, there are a few techniques to estimate H_0 which yield roughly independent results. See Riess 2020 for a quick review of the latest results.

The most traditional method is known as the *distance ladder*. The idea is to measure the distance-redshift relationship of objects well in the Hubble flow. The distance to the closest objects in the Solar neighbourhood can be estimated with the parallax method. The Gaia satellite has the current largest catalogue of parallax measurements, of more than a billion stars. Parallax is one of the few direct distance estimating methods. They can be used to measure the intrinsic luminosity of some objects thought to be standard candles, such as Cepheid stars, RR Lyrae, and the largest red giant stars. Distances to farther objects can be estimated using these candles. They can also be used to calibrate other brighter standard candles, such as type-Ia supernovae, which can be observed at the largest distances. Riess et al. 2021 contains the latest measurements using Gaia parallaxes, Cepheids and type-Ia supernovae to determine

$H_0 = 73.0 \pm 1.4$ km/s/Mpc. Freedman et al. 2019 is latest measurement of distances using the tip of the red giant branch.

The distance to the galaxy NGC 4258 could be determined thanks to the presence of a water maser orbiting the center of this galaxy. The proper motions of several clouds orbiting close to the central massive black hole could be measured both in radial and angular directions, strongly constraining the dynamics of the system (Herrnstein et al. 1999). The most recent measurement yields $D = 7.576 \pm 0.082$ (stat.) ± 0.076 (sys.) Mpc (M. J. Reid et al. 2019), which can be used as an alternative to Cepheids to anchor the distance ladder and provide an estimate to H_0 .

Another alternative method is based on the time-delay of the signals emitted from a quasar behind a strong gravitational lens. The lens creates multiple images of the same background quasar, but the light paths have slightly different lengths. Since quasars are variable objects, the same variability is observed with a delay of several days between the different images. By modelling the distribution of matter between the quasar and us, it is possible to convert these time delays into an estimate of H_0 . The “ H_0 Lenses in COSMOGRAIL’s Wellspring”, or H0LiCOW, collaboration produced the latest comprehensive measurement of H_0 using strongly-lensed quasars (Wong et al. 2019). An alternative measurement has been performed using strong lenses from the Sloan (SLACS). They are compared to H0LiCOW in Birrer et al. 2020.

More recently, gravitational waves from a merger of two black holes were observed for the first time by the LIGO collaboration (T. L. S. Collaboration and the Virgo Collaboration 2016). Not long after, a merger of two neutron stars was observed by both LIGO and Virgo, but this time an electromagnetic counterpart was also detected (T. L. S. Collaboration and T. V. Collaboration 2017), pointing to the galaxy where the event occurred. Given the well predicted shape of the wave form and its dependency on the masses of the neutron stars, it was possible to estimate the distance to the host galaxy. Combining with a measurement of its redshift, a single event could provide an estimate of H_0 (Abbott et al. 2017). This opened the field of cosmology using standard sirens, i.e., gravitational wave sources.

The cosmic microwave background can yield an estimate of H_0 but it is strongly degenerate with other unknown parameters, such as curvature or dark-energy densities. Some of these degeneracies are reduced when considering the effect of gravitational lensing of the CMB or when combining with other probes of late times. Therefore, precise measurements of H_0 from the CMB alone are only possible when considering more restrictive models, such as a flat space with a cosmological constant.

1.5.2 Type-Ia supernovae

As mentioned earlier, type-Ia supernovae (SNIa) can be used as standard candles for absolute distance measurements, if they are anchored by another method. The SNIa can also be used alone to produce relative distance measurements and constrain dark-energy. In this case, no anchor is required and the only assumption is that the SNIa intrinsic brightness does not evolve with redshift.

The main observable are the fluxes of SNIa as a function of time in different photometric bands, known as light-curves. Spectroscopic follow up observations of the explosion can confirm the type of the supernova based on the features present in their spectra. Another key ingredient is the redshift of the host galaxy of the SNIa, which is commonly measured with spectroscopy as well.

Spectro-photometric models of SNIa are used to fit the observed light-curves, yielding their apparent magnitudes at peak luminosity in a given photometric band. These models account for correlations between color and duration of light curves and the peak magnitude, reducing the intrinsic scatter in these magnitudes from 40% to roughly 15%. It is essential to obtain accurate and precise measurements of SNIa fluxes in order to obtain the best cosmological constraints. Large sets of realistic simulations of the data are required in order to correct for selection effects. The final product of the analysis is a set of distance moduli (Eq. 1.6) and their host-galaxy redshift, which can then be compared to models of expansion of the Universe.

Distance moduli depend on dark-energy through the integral of $H^{-1}(z)$. In order to obtain good constraints on dark-energy properties, it is important to have a large redshift coverage, at least covering the transition between matter-dominated to dark-energy-dominated eras, so $0 < z < 1$. The advantage of SNIa for dark-energy studies is that they can span these redshifts with a high sampling rate, which helps in the study of the expansion-rate. One of the inconveniences is that SNIa are complex and poorly understood astrophysical events, the intrinsic scatter in luminosity cannot be reduced to better than 12%, potentially limiting the gains constraining power from future experiments.

The latest comprehensive study of SNIa combines data from more than a dozen projects into a single sample, the Pantheon sample, from which dark-energy constraints were derived (D. M. Scolnic et al. 2018). The Dark Energy Survey also measured a more recent sample of few hundreds of SNe (Brout et al. 2019b; Brout et al. 2019a; Kessler et al. 2019), but not all of them have spectroscopic confirmation that they are of type Ia. Their constraints on dark-energy are not yet competitive compared to the Pantheon sample. The Zwicky Transient Facility (ZTF, Graham et al. 2019) is currently observing the Northern sky on a search for transient events and is expected to discover around 5000 spectroscopically confirmed SNIa at low redshifts ($0 < z < 0.12$) until 2023. After 2023, the Rubin Observatory Legacy Survey of Space and Time (LSST) will take over in the Southern hemisphere and will discover more than 300 thousand SNIa up to $z < 0.5$.

While ZTF and Rubin's samples of SNIa will not decrease significantly the errors on dark-energy parameters, they cover a redshift range where other powerful probes, such as BAO or weak-lensing, lack of statistical power. Furthermore, at lower redshifts when dark-energy is dominant, SNIa are complementary to redshift-space distortions (RSD) when it comes to testing the validity of General Relativity or constraining alternate models of gravity, as solutions for dark-energy.

Not only SNIa are a great probe of the expansion history, they can also provide peculiar velocities of their host galaxies via their inferred distances. These peculiar velocities and their statistical properties can complement RSD analysis when measuring the growth-rate of structures (Eq. 1.15). Estimates of $H(z)$ and $f(z)$ can help break degeneracies between simple dark-energy models and more involved models of gravity (Kim and Linder 2020; Graziani et al. 2020). Chapter 5 is dedicated to this topic, to which I plan to dedicate the future years of my research.

1.5.3 Big Bang nucleosynthesis

In the post-inflation Universe, when temperatures are below the equivalent of 100 MeV, perturbations in the matter and radiation fields are quite small and most of the physics is dictated by the interactions between protons, electrons, neutrons, and photons, as described by the Boltzmann equations. As the Universe cools down and rarefies, protons and neutrons start forming atoms of deuterium, tritium, helium and heavier elements. This process is known as

the Big Bang nucleosynthesis (BBN).

The relative amount of each of the formed elements depends on the expansion rate of the Universe at that time as well as the physical density of baryons $\omega_b = \Omega_b h^2$ and radiation $\omega_r = \Omega_r h^2$. Given that the energies involved are within reach of particle accelerators, these reactions can be studied with great detail on Earth, allowing us to build accurate models of the BBN (see Pitrou et al. 2021 for the latest calculations and references therein).

Observations of the primordial abundances can be compared to the predictions by BBN models. Abundances of deuterium, helium, and others can be estimated from spectroscopic observations of HII regions in metal-poor galaxies or from absorption lines of the intergalactic medium in quasar spectra. The most up-to-date measurements of the primordial helium-4 abundance yields $Y_p = 0.2453 \pm 0.0034$ (Aver et al. 2021), while the deuterium one is $D/H = (2.527 \pm 0.030) \times 10^{-5}$ (Pitrou et al. 2021). While observations are consistent with BBN models for deuterium and helium, the abundance of lithium-7 exhibits a factor 3 discrepancy, which is a huge problem in BBN but quite often neglected.

The temperature fluctuations in the cosmic microwave background (CMB) are also very sensitive to the physical baryon density ω_b . Historically, the values obtained from CMB have been in good agreement with BBN measurements, showing that baryons make up to around 16% of the total matter content of the Universe, or $\omega_b = (2.195 \pm 0.022) \times 10^{-2}$. The agreement between two quite independent probes is one of the great successes of the current cosmological model, though they also enforce the need for a dark-matter component.

1.5.4 Baryon acoustic oscillations

Baryon acoustic oscillations (BAO) is the name given to the propagation of sound waves in the primordial plasma (baryons and photons), prior to recombination. Because of the high pressure on small scales at those times, each initial density perturbation had a spherical density wave around them propagating outwards at the speed of sound in that medium. The speed of sound in the plasma is given by

$$c_s(z) \equiv \sqrt{\frac{1}{3[1 + R(z)]}} \quad (1.19)$$

where $R(z) = 3\rho_b/4\rho_\gamma$ is the baryon-to-photon ratio. The propagation of sound waves occurred until the temperatures and densities dropped to values such that baryons no longer felt the pressure from photons, known as the *drag epoch*, which is close in time to the recombination (but technically not the same epoch). This process left a slight overdense shell around each initial perturbation with a radius given by

$$r_{\text{drag}} \equiv r_s(z_{\text{drag}}) = \int_{\infty}^{z_{\text{drag}}} dz' \frac{c_s(z')}{H(z')} \quad (1.20)$$

which is known as the *sound horizon at drag epoch*, or the BAO scale. Today, r_d has a physical size of about 147 Mpc, much larger than any collapsed structure in the Universe. As one can see from Eqs. 1.19 and 1.20, r_{drag} mainly depends on ω_b and ω_c assuming the CMB gives a precise estimate of ω_γ . The dependency of r_{drag} is mainly through ω_b , which is useful when constraining dark energy models for instance.

After recombination, the sound horizon scale only increases in size due to the expansion, or equivalently, its comoving size remains unchanged. Therefore, the BAO scale is a great *standard ruler* to study the expansion rate of the Universe. In practice, the BAO scale is observed

statistically in the two-point function of the matter density field, so it is often classified as an statistical standard ruler. In configuration space, the correlation function $\xi(r)$ presents a small peak at separations corresponding to the BAO scale r_d , while in Fourier space the power spectrum $P(k)$ contains an oscillatory pattern as a function of scale with a period proportional to the BAO scale. In Chapters 3 and 4, I present my past work in the measurement of the BAO scale using Ly α forests and galaxies, respectively, as tracers of the matter density field.

Given that a galaxy survey is made of angular positions and redshifts, which are true observables, the BAO peak is effectively measured as an angle $\Delta\theta_{\text{BAO}}$ or as a difference in redshift Δz_{BAO} . These can be modelled as ratios of distances to the BAO scale r_d as

$$\Delta\theta_{\text{BAO}}(z_{\text{eff}}) = \frac{D_M(z_{\text{eff}})}{r_d} \quad (1.21)$$

$$\Delta z_{\text{BAO}}(z_{\text{eff}}) = \frac{D_H(z_{\text{eff}})}{r_d} \quad (1.22)$$

where z_{eff} is the effective redshift of the galaxy survey, D_M is the comoving angular diameter distance (Eq. 1.4), and D_H is the Hubble distance (Eq. 1.2).

There are two ways BAO can be used to constrain cosmological models, depending if we assume that r_{drag} is known or not. If r_{drag} is known, i.e. given by Eq. 1.20 using some value for ω_b (given by the CMB or BBN for instance), then BAO measurements are converted to absolute distance measurements which depend only on $H(z)$. Therefore, BAO can constrain H_0 , curvature and dark energy. If r_{drag} is supposed to be unknown but still a standard ruler, then BAO constraints the ratio of $E(z) = H(z)/H_0$ to the combination $H_0 r_{\text{drag}}$ which is now degenerate. This is similar to type Ia supernovae, where their absolute magnitude is degenerate with H_0 . By combining several BAO measurements at different effective redshifts, BAO is a powerful probe of dark energy and curvature.

Current BAO measurements span effective redshifts from 0.1 to 2.3. The Sloan Digital Sky Survey (SDSS, Eisenstein et al. 2011; Blanton et al. 2017) has measured more than 2 million redshifts spectroscopically in the past twenty years, producing the largest maps to date of the distribution of matter in the Universe. In addition to SDSS, surveys such as FastSound (Okumura et al. 2016), Vipers (Pezzotta et al. 2017), 6 degree field galaxy survey (6dFGS, Beutler et al. 2012) and WiggleZ (Parkinson et al. 2012) also produced BAO measurements using galaxies, though the volumes probed and the number of galaxies is inferior to SDSS. The latest cosmological constraints from BAO are described in Alam et al. 2021.

1.5.5 The cosmic microwave background

The cosmic microwave background (CMB) is one of the richest cosmological probes of all. The photons we receive today last scattered on baryons at a redshift of about 1100, corresponding to roughly 380 000 years after the Big Bang. It is the oldest information that we can measure from the Universe today.

The average temperature $T_{\text{CMB}} = 2.72548 \pm 0.00057$ K (Mather et al. 1994; Fixsen 2009) of these photons, measured the last time by the COBE satellite, tell us on how much energy density from radiation there is in the Universe. From the black body Bose-Einstein distribution of photons, we have

$$\rho_\gamma = \frac{\pi^2 k^4}{15 \hbar^3 c^3} T_{\text{CMB}}^4 \quad (1.23)$$

The fluctuations around this average temperature, ΔT , and the polarisation of these ph-

tons trace the structures back at the recombination epoch. Given its early-times nature, the temperature and polarisation fields are extremely well described by Gaussian statistics. Most of the information is therefore contained in the two point functions of these fields. Given that these fields are two dimensional, functions of the position angle in the sky (θ, ϕ), it is convenient to decompose these fields into a basis of spherical harmonic functions $Y_{\ell m}(\theta, \phi)$, which amplitudes are denoted $a_{\ell m}$. The angular power spectrum is simply the variance of these amplitudes: $C_{\ell} = \langle a_{\ell m} a_{\ell m}^* \rangle$, which in linear theory is simply a function of ℓ and independent of m . The temperature and both polarisation modes (E and B) can be decomposed into spherical harmonics, so we can estimate all cross power spectra, e.g., $C_{\ell}^{TE} = \langle a_{\ell m}^T a_{\ell m}^{E*} \rangle$ is the cross temperature and E-mode polarisation power spectrum.

The temperature and polarisation auto and cross power spectra are exquisitely well modelled by linear perturbation theory. In the most basic Λ CDM model, the power spectrum is a function of only six parameters: the angular scale of baryon acoustic oscillations θ_* , the physical density of baryons ω_b and dark-matter ω_c , the primordial power spectrum amplitude A_s and slope n_s , and the optical depth τ . The fact that such a model with so few free parameters can describe so well the observations is one of the greatest achievements in modern physics. A great description of the CMB physics from first principles can be found in Dodelson and Schmidt 2020. Several codes are available to compute models for the CMB such as CAMB³ (Lewis et al. 2000) and CLASS⁴ (Lesgourgues 2011).

Explain how the power spectrum depends on the parameters

Latest measurements: Planck satellite (Planck Collaboration et al. 2020), Atacama Cosmology Telescope (ACT, Aiola et al. 2020), South Pole Telescope (SPT, Aylor et al. 2017; Balkenhol et al. 2021; Dutcher et al. 2021), BICEP/Keck (BICEP/Keck et al. 2021) for B-modes.

1.5.6 Redshift-space distortions

Matter is not static in the Universe. Due to gravity, matter flows from underdense regions towards overdense ones. These velocities are commonly called *peculiar velocities*. The radial component of peculiar velocities alters the observed redshift of an object, due to the Doppler effect. Therefore, the total observed redshift z_{obs} is a combination of its cosmological redshift z_{cos} , due to the expansion of the Universe, and the peculiar redshift z_{pec} , due to the Doppler effect:

$$1 + z_{\text{obs}} = (1 + z_{\text{cos}})(1 + z_{\text{pec}}) \quad (1.24)$$

where

$$z_{\text{pec}} = \sqrt{\frac{1 + \frac{\vec{v} \cdot \hat{n}}{c}}{1 - \frac{\vec{v} \cdot \hat{n}}{c}}} - 1 \approx \frac{\vec{v} \cdot \hat{n}}{c} \quad (1.25)$$

The first equality is the definition of the relativistic Doppler effect, \vec{v} is the velocity and \hat{n} is the radial unitary vector. The right-hand side is the non-relativistic approximation, which is accurate for typical velocities of matter flows in our Universe.

When converting observed redshifts into comoving distances, peculiar velocities introduce a small error such that

$$s(z_{\text{obs}}) \approx r(z_{\text{cos}}) + \frac{(1 + z_{\text{cos}})}{H(z_{\text{cos}})} \vec{v} \cdot \hat{n}, \quad (1.26)$$

where s is the *redshift-space* distance and r is the *real-space* distance. The errors caused by

³<https://camb.info/>

⁴https://lesgourg.github.io/class_public/class.html

peculiar velocities distort the observed distribution of matter/galaxies since we cannot in general decouple the velocities and the Hubble flow. This effect is observable and named *redshift-space distortions* (RSD). RSD are a powerful probe of the dynamics of the matter field, i.e., on how velocities are related to the densities, and how they contribute to the growth of structures over time. More generally, observations of RSD can be used to test the validity of general relativity since it is a probe of the strength of gravity.

The density contrast of matter is modified when observed in redshift-space relative to real-space, given the redshift-space distortions. Mass conservation implies that

$$[1 + \delta_s(\vec{s})] d^3 s = [1 + \delta(\vec{x})] d^3 x, \quad (1.27)$$

such that in Fourier space we obtain, in the linear regime (Kaiser 1987)

$$\delta_s(\vec{k}) = (1 + f\mu^2)\delta(\vec{k}) \quad (1.28)$$

where f is the growth-rate of structures defined in Eq. 1.15 and $\mu = \vec{k} \cdot \hat{n}/k$ or the cosine of the angle between the wavevector and the line of sight. The power spectrum in redshift-space assuming linear perturbations is simply

$$P_s(k, \mu) = (1 + f\mu^2)^2 P_m^{\text{lin}}(k) \quad (1.29)$$

This means that the redshift-space matter power spectrum has radial modes enhanced by a factor of $(1 + f)^2$ (where $f \sim 1$) relative to modes transverse to the line of sight. This enhancement is observable with galaxy surveys. If the normalisation of the linear matter power spectrum is parametrised by σ_8 (Eq. 1.18), then the actual measured quantity is the product $f\sigma_8$. Transposed to a given effective redshift z_{eff} of a galaxy survey, we need to scale σ_8 , which is usually defined at $z = 0$, using the growth factor $D(z)$, such that the measured quantity is $f(z)\sigma_8 D(z)$.

The observable $f(z)\sigma_8 D(z)$ is mostly sensitive to the amount of dark matter ω_m , which drives the growth of structures, and $H(z)$, which damps the growth (see Eq. 1.14). Therefore, it can be used to constrain dark energy models. More generally, measurements of $f(z)\sigma_8 D(z)$ can be used to test the the validity of general relativity on large scales, which is the commonly assumed theory of gravity.

The latest most relevant measurements were performed using data from the Sloan Digital Sky Survey (SDSS), including

- the SDSS Main Galaxy Sample Howlett et al. 2015,
- the Baryon Oscillation Spectroscopic Survey (BOSS, Alam et al. 2017),
- the extended BOSS luminous red galaxy sample (eBOSS LRG, Bautista et al. 2021; Gil-Marín et al. 2020),
- the eBOSS emission line galaxy sample (eBOSS ELG, Tamone et al. 2020; de Mattia et al. 2021),
- the eBOSS quasar sample (eBOSS QSO, Hou et al. 2021; Neveux et al. 2020),

In chapter 4 I present my contributions to the measurement of the growth-rate of structures using the eBOSS LRG sample (Bautista et al. 2021). The cosmological implications of SDSS growth-rate measurements are described in Alam et al. 2021.

1.5.7 Weak gravitational lensing

Photons follow space-time geodesics. Space-time is distorted in the presence of a source of gravitational potential, which is typically in the form of a mass concentration. Therefore, the matter distribution in the Universe bends photon trajectories from distant sources. This phenomenon receives the name of *gravitational lensing*, since the theory describing light propagation on a gravitational field is analogous to classical optics. Gravitational lensing is a rich cosmological probe since the distortion of photon trajectories depends on both the baryonic and dark matter. From lensing measurements we can learn about the total matter distribution in an expanding Universe.

In the case of an extended source of photons, e.g., a galaxy, light from different angular positions within the source take slightly different paths towards the observer. If this extended source is placed in the background of a mass concentration - the lens - each path will suffer a slightly different bending, which depends on the impact parameter of each photon relative to the center of this lens. Also, due to conservation of surface brightness (is this the correct name?), the total flux is also increased. The final result is an image shifted, distorted and brighter.

Lensing measurements use shifts, distortions and increase in flux to determine properties of the lenses. These measurements are hard since we do not have access to the original unlensed position, shape and flux of a given galaxy. One remarkable exception is lensing in its strong regime. When the ratio between impact parameter of the source and mass of the lens is below some threshold - the Einstein radius - multiple images of the same source are created. Strong lensing allows us to estimate the mass and density profile of a given lens. Moreover, if the source is variable in time, this variability is slightly delayed between each of the multiple images, which allows us to constrain the expansion of the Universe. As mentioned in section 1.5.1, strongly lensed quasars (variable sources) are used to constrain H_0 , under the assumption that we can properly model the lens density profile.

However, statistical measurements of lensing, such as correlations with overdensities, can be performed by making assumptions on the statistical properties of the unlensed sample of galaxies.

weak regime we measure shapes and correlate themselves (shear-shear) or with the galaxy field (galaxy-shear)

Sensitive to: the amplitude of the shear power spectrum, $S_8 = \sigma_8 \sqrt{\Omega_m / 0.3}$.

Latest measurements: Kilo Degree Survey (KiDS, Heymans et al. 2021), Dark Energy Survey Year 3 (DES-Y3, D. Collaboration et al. 2021), Hyper Supreme Cam first year (HSC, Hikage et al. 2019)

2 Observing the Universe with spectroscopy

Contents

2.1 Selecting the objects to observe	24
2.2 Pointing fibres to the sky	25
2.3 From electrons to spectra	26
2.4 From spectra to redshifts	28
2.5 From redshifts to cosmology	28

In the last two decades, spectroscopy became one of the most powerful techniques to survey the galaxies of our Universe. Particularly thanks to its capability to obtain precise galaxy redshifts, spectroscopy allows us to build precise maps of the distribution of matter in three dimensions.

This chapter is an overview on how to observe galaxies with spectroscopy and how the data is treated from the target selection all the way to the redshifts. I expose my work on improving the spectroscopic data reduction pipeline for the extended Baryon Oscillation Spectroscopic Survey (eBOSS), for which I was the *Lead Data Scientist* for 3 years.

Naturally, this chapter will focus on the spectroscopic observations with the Sloan Digital Sky Survey (SDSS), but the majority of the concepts introduced here also apply to the Dark Energy Spectroscopic Instrument (DESI).

2.1 Selecting the objects to observe

The first step to build a spectroscopic survey is to pre-select objects to be observed. This step, known as *target selection*, is required since you need to know where to point the optical fibers that take the light from an object to the spectrograph. Therefore, we cannot simply observe all objects in a given field, we need to choose which targets to observe.

For the target selection, a prior *photometric or imaging* survey is required. In the first years of SDSS, a photometric survey was carried out, covering more than XXX deg² of the sky York et al. 2000. The focal plane was equipped with six rows of five charge-coupled devices (CCD), each one covered with one of the SDSS filters: *u, g, r, i* or *z*. J. E. Gunn et al. 1998; Doi et al. 2010. A technique named drift-scanning was used to continuously observe “stripes” of constant declination during the night. SDSS was the first of its kind to produce a systematic survey of the Universe in the optical domain.

Images were reduced using the SDSS photometric pipeline Lupton et al. 2001; Padmanabhan et al. 2008. Fluxes/magnitudes and their uncertainties were computed for each detected object in five colour bands. Based on their fluxes and angular sizes relative to the point-spread function (PSF), each object received a photometric classification as star or galaxy.

The final list of objects with their respective fluxes and angular positions is the input for targeting algorithms. These algorithms aim to select a given type of object for spectroscopic follow-up, based solely on their fluxes and colors. For galaxy surveys, it is vital to be able to

Table 2.1: Surveys and their target selection algorithms

Sample	Redshift range	Reference
SDSS Main Galaxy Sample	$0.0 < z < 0.2$	Strauss et al. 2002
BOSS LOWZ galaxies	$0.2 < z < 0.4$	B. Reid et al. 2016
BOSS CMASS galaxies	$0.4 < z < 0.7$	B. Reid et al. 2016
BOSS Ly α forest quasars	$2.0 < z < 3.5$	N. P. Ross et al. 2012
eBOSS LRGs	$0.6 < z < 1.0$	Prakash et al. 2016
eBOSS ELGs	$0.7 < z < 1.1$	A. Raichoor et al. 2017
eBOSS quasars as tracers	$0.8 < z < 2.2$	Myers et al. 2015
eBOSS Ly α quasars	$2.0 < z < 3.5$	Myers et al. 2015 N. Palanque-Delabrouille et al. 2016

distinguish between galaxies - the objects of our interest - and stars - which belong to our own galaxy and have a distinct scientific purpose. Additional colour cuts also help selecting a given redshift range for particular types of galaxies.

Since we are interested in the clustering of galaxies, it is essential to obtain a relatively homogeneous angular density of targets so to avoid spurious correlations. Target selection algorithms enforce a requirement of about 15 per cent angular fluctuations on the number density of targets. Residual fluctuations have to be corrected before any clustering measurements, I will discuss about this in section 2.5.

Table 2.1 provides a summary of target selection algorithms for several types of galaxy types and redshift ranges. (see https://www.sdss.org/science/technical_publications)

2.2 Pointing fibres to the sky

Once the targets are chosen, we need to define the observing strategy for spectroscopy. This strategy is defined based on several constraints, such as

- the focal plane dimensions, which is a one meter diameter plate holding 1000 optical fibers;
- the field of view of the telescope, which is about 5 deg^2 for the plates;
- the number of fibers. There are a total of 1000 fibers available of which 80 are used for sky observations and 20 for standard stars;
- the size of the footprint, which is roughly $10\,000 \text{ deg}^2$;
- the fiber completeness, i.e., the fraction of targets receiving an optical fiber. The completeness has to be usually above a certain threshold over all the footprint;
- exposure times and total observing time available. Exposure times are dependent on the average signal-to-noise ratio of the observed targets, which should reach a certain threshold. The total observing time of the program is roughly four to five years.
- visibility window of a given patch of the sky at a given time of the year;
- priority for fiber assignment. Some types of targets have higher priority than others, which affects the fiber completeness of the low-priority samples.

The process of dividing the sky into overlapping projections of the focal plane is called *tiling*. A detailed description of the tiling algorithm can be found in Blanton et al. 2003. Once the tiling and fiber assignment are set, this information is sent to the plate production and drilling of holes that will hold the optical fibers. Focal plane plates are drilled a few months before observing and are unique for a particular patch of the sky and observing time. Not observing the plate at the designed times causes loss of flux due to different refraction caused by the atmosphere.

The drilled plates are sent to the Apache Point Observatory (APO) in New Mexico where the Sloan 2.5-meter Telescope is based. On the mountain, observers attach the plates into cartridges that will fit at the focal plane of the telescope. There are about 15 cartridges, each equipped with 1000 optical fibers. The fibers are plugged by hand, by one or two observers during the afternoon preceding the observation night. Plates are unplugged from their cartridge once a large enough number of exposures has been taken. A minimum of 3 exposures are taken per plate.

The light of the objects is transported by the optical fibers through the Sloan spectrographs. There are two spectrographs attached at the focal plane of the telescope. Each spectrograph receives the light from 500 fibers and passes it through a beamsplitter, dividing it into a red and blue channels. Each channel has its own grism that spreads the light over wavelength before hitting the CCDs. The blue camera observes roughly from 3500 to 6000 Å and the red camera from 6000 to 10500 Å. The resolution $R \equiv \lambda/\Delta\lambda$ increases with wavelength from 1500 to 2000 on the blue camera and from 2000 to 2500 on the red camera.

In addition to the science exposures (observing galaxies), a small set of calibration exposures is also taken, typically at the beginning and at the end of the observing run. Flat exposures are taken with lamps that emit over all wavelengths. The light is passed through the spectrographs, so we refer to these as fiber-flats, as opposed to the exposures taken without the spectrographs, named super-flats. With another type of lamp, which emit narrow lines at some specific wavelengths, arc exposures are obtained, that are used to derive the wavelength-pixel relation.

2.3 From electrons to spectra

This section describes the data reduction pipeline of spectroscopic observations by the SDSS telescope, for which I contributed as the eBOSS Lead Data Scientist. This automated pipeline transforms the counts stored in CCDs into calibrated spectra, for which redshifts are estimated. The software, named `idlspc2d`, was written in Interactive Data Language (IDL) and can be found online¹. The latest version used in Data Release 16 of eBOSS data is v5_13_0 (Ahumada et al. 2020).

The dispersed light of each object falls onto CCD detectors containing 2048x2048 square 24 μ m pixels. There are 500 traces per CCD, except when fibers are broken or unplugged by accident. The traces of each spectra are parallel and slightly curved towards the edges. They are separated by about 7 pixels.

The first step of the data reduction is to remove bias and dark, mask cosmic rays and other known bad pixels, convert counts into electrons using estimated gain values, and correct for the super-flat image (flat taken without the spectrograph).

The next step is to extract the total number of counts per wavelength and per object. One of the axis of the CCD is aligned with the wavelengths, but we do not know the wavelength

¹https://svn.sdss.org/public/repo/eboss/idlspc2d/tags/v5_13_0/

solution at this point. The other axis is aligned with fiber number. The extraction of the fluxes is performed by bundles of 20 fibers. A set of 20 Gaussian profiles plus a third order polynomial term are fit simultaneously over the counts. From the best-fit parameters of the Gaussian, one can compute the total flux at each wavelength for each object. This fit is performed regardless of whether the fibers contain flux from sky, stars or galaxies. The extraction is performed similarly to science, flat and arc exposures.

One important aspect of extraction is: what do we use as weights in the fit? For SDSS-II and III, the extraction used the total estimated variance of each pixel, assumed to be Poisson with mean equal to the number of observed hits in the pixel. However, in SDSS-IV eBOSS we pushed the limits of the instrument by observing fainter objects. In this regime, we started to observe biases due to this weighting scheme in the extraction. The ideal extraction would use the true variance (Horne 1986), not the estimated one, as a weight. Using the estimated one yields a bias in the final fluxes. We modified the extraction algorithm such that it would use a flux-independent weight for fit, yielding unbiased fluxes. The price to pay is that this extraction is less optimal, yielding slightly larger flux uncertainties. Biased fluxes affected particularly the analysis of Ly α forests, as described in Chapter 3 or in the appendix of Bautista et al. 2017.

The fiber-flat images are used to calculate the traces positions and widths (more precisely than in science images) and to correct for throughput variations across fibers. The arc images are used to calculate the wavelength solution and the dispersion in the wavelength direction based on the line widths. Sky lines in science exposures are eventually used to do small adjustments to the arc-image solution.

The flux in the sky fibers are used to fit a sky model in units of counts. A polynomial dependency over fiber index is used to account to variations over the focal plane. This sky model is then subtracted from all science spectra, including the sky spectra themselves and calibration stars. The sky-subtracted sky fibers are a good metric to evaluate the quality of sky subtraction algorithm.

The next step of the reduction is the *flux calibration*, which converts the observed counts into flux in physical units. The spectra of standard stars are the main ingredient of the flux calibration. The stars chosen for calibration are of type F, with small variations in temperature and metallicity. Physical models of their spectral emission can be obtained through complex stellar synthesis calculations, including all absorption features. The goal is to fit absorption lines to the data in order to determine the exact model for each observed star. We start by isolating the absorption lines in the observed spectrum (in units of counts) by fitting a smooth function over its shape, then dividing the whole spectrum by this model. The same is done for the physical stellar models, so the fit only uses the absorption lines information. Once the best parameters of the star are found, a calibration vector is constructed by simply taking the ratio of the observed counts to the full model including its smooth shape. A set of ten stars are fit independently and a single calibration vector is obtained from them. This final calibration vector is then applied to all other galaxy spectra in order to convert their flux into physical units. During the last years of eBOSS, I updated the set of physical stellar models from the Kurucz model to those used in DESI (ref? Allende-Prieto?) which have a larger diversity in stellar parameters and more precise absorption features. This update contributed to a significant reduction of flux calibration residuals (add Figure??) computed using stacks of spectral regions of quasars without emission or absorption lines.

Once individual science exposures are converted into physical flux units, we proceed to the coaddition of these exposures into a single set of spectra. For a given object, a B-spline is fitted over all observed spectra, using a new wavelength grid with constant steps in $\log_{10} \lambda$

of 10^{-4} . The best-fit spline is the final coadded spectrum for this object. The coaddition is made independently for each object. At this stage, we also combine spectra from blue and red cameras into one single spectrum covering the full wavelength range of 3500 to 10500 Å. The last step of this process is the calculation of potential broadband distortions in flux caused by atmospheric differential refraction (ADR). Using information from the plate design, the position of the fibers in the plate and the actual observations (airmass, hour-angle), one can compute a correction vector that is also applied to all spectra.

During the whole reduction, pixels or whole fibers can be masked due to any issues, or when the robustness of flux calculations is compromised. Each pixel has a flux and an uncertainty estimates, the latter is expressed as an inverse variance, which can be used directly as a weight in scientific analyses.

This concludes the so-called two-dimensional reductions, that convert 2D CCD images into a set of calibrated coadded sky-subtracted 1D spectra.

2.4 From spectra to redshifts

The next and last step of the automated data reduction pipeline is the spectral classification and redshift measurement. A set of physical templates of stars, galaxies and quasars is fit to each spectrum by a simple χ^2 minimization. For each template, we scan over several values of redshifts to obtain the minimum χ^2 . The redshift ranges probed depend on the type of template: stars are at $z = 0$ with allowed peculiar velocities, galaxies are in $0 < z < 2$ and quasars in $0 < z < 7$. The five pairs of template-redshift producing the smaller χ^2 values are stored. Redshift uncertainties are estimated using the χ^2 profiles around the minima. If the difference between the first and second best-fit χ^2 values is smaller than a certain threshold (corresponding to roughly 5σ for one parameter), a warning flag is set, meaning that the classification is not to be trusted.

A program of visual inspection of reduced spectra, with their automated classification and redshift estimates, is carried out following the first observations of a new survey. Visual inspection is an important step in order to verify the quality of the data reduction and the classification process. A truth table containing the visually confirmed redshifts and classifications for thousands of spectra is one of the results of the visual inspection. This truth table is used to compute the actual density of tracers that are obtained for a given target selection, as a function of redshift. Spectral features caused by problems in the reduction were then reported to the pipeline team.

During the first months of eBOSS, I coordinated a program of visual inspection where about 15 members analysed about a thousand spectra each (with some overlap for cross-checking results). This provided a truth table for the Luminous Red Galaxies from eBOSS. Similar programs were carried out for the eBOSS Emission Line Galaxies and quasars. The inspections of eBOSS quasars continued through the whole program, where a small fraction of the spectra without confident automated classifications were inspected. In DESI, a larger visual inspection program was put in place for the same goals.

2.5 From redshifts to cosmology

Description of catalog generation. Shall this go on the clustering part ?

3 The high-redshift Universe and its forests

Contents

3.1	Forests as a tracer of the Universe's structures	29
3.2	Baryon acoustic oscillations in the forests	31
3.2.1	Transmission field	32
3.2.2	Two-point correlation functions	33
3.2.3	Correction matrices	34
3.2.4	The model	35
3.2.5	BAO constraints	36
3.3	Impact of redshift errors	38
3.4	Weak-lensing of forests	41
3.5	Voids in tomographic maps	44

Until today, the Ly α forest in quasar spectra has been the only tracer of large-scale structures producing measurements of baryon acoustic oscillations (BAO) at redshifts above 2. Given the higher redshifts and the smaller scales probed, the clustering of the forest also yielded competitive constraints on the sum of the mass of the neutrino species, when combining it with measurements of the cosmic microwave background (CMB) anisotropies. Current and future spectroscopic surveys plan to observe denser samples of Ly α forests to improve our understanding of the $z > 2$ Universe.

In this chapter, I expose my contributions to the study of dark energy through Ly α forest observations. Section 3.1 introduce the main concepts used in this chapter. Section 3.2 focus on the BAO measurements from BOSS and eBOSS surveys for which I provided key contributions. My thesis¹ was dedicated to this topic. Section 3.3 and 3.4 present two projects carried out by my former PhD student Samantha Youles at the University of Portsmouth, UK, who graduated on the 18th of March 2022.

3.1 Forests as a tracer of the Universe's structures

The Ly α forest is the name for a series of absorption lines observed in quasar spectra caused by the presence of neutral hydrogen along the line of sight between us and the quasars. Figure XX shows an example of a optical quasar spectrum with its Ly α forest. These lines are seen bluewards of the Ly α emission peak of the quasar, which lies at the quasar rest-frame wavelength $\lambda_{\text{rest}} = 1216 \text{ \AA}$. The absorption is not limited to the first transition; bluewards of the Ly β peak ($\lambda_{\text{rest}} = 1025 \text{ \AA}$) we also observe Ly β absorption lines on top of the Ly α ones, and similarly for all Lyman series until the Lyman break at 912 \AA .

The advantage of the forests is that it provides a tomographic view of the matter distribution along the line of sight of the quasar. This is because, as the light propagates outwards of the quasar, the Universe expands, causing the whole spectrum to be redshifted. When the

¹<https://tel.archives-ouvertes.fr/tel-01389967>

redshifted spectrum hits a hydrogen atom, the light being absorbed at the Ly α transition in the *atom* rest-frame is no longer at the Ly α wavelength in the *quasar* rest-frame: it a bluer wavelength. If a given quasar sits at a redshift z_q and a given intervening atom sits at a redshift $z_a < z_q$, the Ly α photon being absorbed by the atom corresponds to a photon at $\lambda_{\text{rest}} = 1216(1+z_a)/(1+z_q)$ Å in the quasar rest-frame. In our frame, the observed wavelength of the atom absorption, or equivalently of its Ly α line, is $\lambda_{\text{obs}} = 1216(1+z_a)$ Å. Note the observed wavelength does not depend on the quasar redshift. We only need the quasar redshift to identify where the Ly α forest is in the observed spectrum, such that we can assign each observed wavelength to a quasar rest-frame wavelength. The result is that a single forest of lines maps the neutral hydrogen density across a large range in redshifts. One single quasar spectrum can yield a map of the large-scale distribution of matter along its line of sight over up to several hundreds of megaparsecs (Mpc).

How much neutral hydrogen is needed to create a forest of Ly α absorption lines? Not a lot as it turns out. Given the high cross-section of the Ly α resonance, very low densities of neutral hydrogen are sufficient to create a line. At redshifts $2 < z < 4$ where Ly α forests are observed in the optical, typical densities found in the intergalactic medium (IGM), $n_{\text{HI}} \sim 10^{XXX}$ cm $^{-3}$, are enough to absorb a significant fraction of the photons. Denser regions completely absorb the light and create saturated lines (zero flux). This is the case of the so-called *high-column density systems* (HCDs), which are often associated with galaxies or proto-galaxies. The extreme cases can be observed with high-resolution spectroscopy, where we can also observe a non-saturated deuterium absorption from which we derive constraints on big-bang nucleosynthesis (BBN, see section 1.5.3).

The Ly α forest, and therefore the amounts of neutral hydrogen in the IGM, is tightly connected with the process of reionisation of the Universe. As more stars and quasars form, more ultraviolet light is produced, progressively ionising the neutral hydrogen. On top of that, the Universe continuously expands, diluting it. Thus, the average absorption of forests decreases with time, or increases with redshift. At redshifts above 6, most of the light bluewards of the Ly α emission is absorbed, while at redshifts below 1, the Ly α absorption is minimal. Forests from $z_q > 6$ quasars are used to put constraints on reionisation (ref?).

The Ly α forest and its connection with the large-scale structures have been studied theoretically since works by James E. Gunn and B. A. Peterson 1965. Most of recent advances are thanks to numerical simulations. In order to simulate forests, a full hydrodynamic n-body simulations are required, including all the complex baryonic physics. Add more refs. On large scales, it has been shown that the forest can be considered as linear tracer of the underlying matter field McDonald et al. 2000; McDonald 2003; McDonald et al. 2005. On smaller scales, the gas physics is more complex to model given the effects of pressure and thermal velocities. Feedback from supernovae explosions, AGN or star forming galaxies also play an important role to model the small-scale clustering (Chabanier et al. 2020). The small scales are interesting due to their potential to constrain neutrino masses for instance.

The first measurements of clustering using the Ly α forests were based solely on the two-point statistics along the same line of sight (Croft et al. 1999; McDonald et al. 2006). Given the good wavelength sampling of a forest, limited by the resolution of our spectrographs, the small scales are easily accessible in the radial/wavelength direction. This type of measurement is still performed today and allow us to obtain tight upper limits on the total mass of neutrino species (Nathalie Palanque-Delabrouille et al. 2020), when combined with measurements of the cosmic microwave background.

It is only with the Baryon Oscillation Spectroscopic Survey (BOSS) from SDSS-III that we

could study the clustering of Ly α forests in three dimensions. Thanks to the density of quasars observed by BOSS, of about 15 deg^{-2} , it was possible for the first time to estimate correlations using absorption from neighboring lines of sight. Also thanks to the large area of sky covered by BOSS, the first measurement of baryon acoustic oscillations (BAO) using forests was achieved Busca et al. 2013; Slosar et al. 2013; Kirkby et al. 2013.

Quasars also trace the matter field. While their density is not homogenous enough across the sky to measure clustering using these quasars alone, they can be cross-correlated with the forests. The cross-correlation between quasars and Ly α forests is interesting because it is mostly independent of the Ly α forest auto-correlation. This is mainly because the quasar sample has a low density (shot-noise dominated). The first measurement of BAO in the quasar-forest cross-correlation was also possible with BOSS (Font-Ribera et al. 2014). The BAO constraints from the auto and cross correlations could be combined assuming that these two are independent.

Since the first measurements of BAO using forests, the BOSS and eBOSS collaborations published BAO constraints with increasingly larger samples and improved analysis. They are associated with the official SDSS Data Releases (DR):

- DR9: First measurement of large-scale Ly α correlations without BAO (Slosar et al. 2011);
- DR9: First detection of BAO in the Ly α auto-correlation (Busca et al. 2013; Slosar et al. 2013; Kirkby et al. 2013);
- DR11: First detection of BAO in the quasar- Ly α cross-correlation (Font-Ribera et al. 2014), updated auto-correlation measurement (Delubac et al. 2015);
- DR12: Final BOSS auto-correlation (Bautista et al. 2017) and cross-correlation measurements (du Mas des Bourboux et al. 2017);
- DR14: Updated measurements with eBOSS data (de Sainte Agathe et al. 2019; Blomqvist et al. 2019);
- DR16: Final Ly α BAO measurements of SDSS (du Mas des Bourboux et al. 2020).

3.2 Baryon acoustic oscillations in the forests

In this section I detail the methodology I used in my past work to measure BAO with a set of Ly α forests, highlighting my personal contributions. I consider the case of SDSS forests, which are observed in the optical domain with low-resolution spectroscopy. Forest have a rather low signal-to-noise ratio on average, so the methods are fit to this type of data. Similar methods are employed in the measurement of the line-of-sight (or one-dimensionnal) power-spectrum, but we focus on BAO here.

The main steps of the BAO analysis with Ly α forests are:

- estimate of the transmission field and their associated weights;
- estimate of the two-point correlation functions, including the cross-correlation with quasars;
- estimate of correction matrices due to distortions of continuum fitting and metals;
- fit of the BAO model over measured correlations.

3.2.1 Transmission field

As described in section 3.1, the amount of absorbed flux at a given wavelength (redshift) is related to the density of neutral hydrogen and therefore is a tracer of structures.

The main observable used to compute correlations is the so-called *transmission* F , defined as

$$F(\hat{n}, \lambda) = f(\hat{n}, \lambda) / C(\hat{n}, \lambda) = \exp\{-\tau(\hat{n}, \lambda)\} \quad (3.1)$$

where $f(\hat{n}, \lambda)$ is the observed flux and $C(\hat{n}, \lambda)$ is the unabsorbed/original flux level, for a quasar line-of-sight at angular position \hat{n} and observer-frame wavelength λ , which can be linked to the absorber redshift $z = \lambda/\lambda_\alpha - 1$ if the absorption is due to Ly α . We can also express the transmission as a function of the optical depth τ as in the right-hand side of Eq. 3.1.

The challenge is to estimate transmissions from the observed fluxes of a set of quasar spectra, particularly given that our data is low-resolution and relatively noisy. This makes it hard to “see” the unabsorbed flux level C , also known as the *continuum* level, requiring automated methods. Several past attempts to achieve this used either principal-component analysis techniques (Lee et al. 2012) where the templates were built from high-resolution and high signal-to-noise data; or maximum likelihood methods accounting for the non-Gaussian nature of the probability density function of F (Busca et al. 2013). During my PhD, I particularly tried to merge these last two methods into a single one, without success. Those methods, while more sophisticated and flexible in their modelling of C , they suffer from the additional noise added to the estimated transmission due to noisier estimates of C . The simplest method will prevail in the latest analyses, which consists in averaging forests in their rest-frame to obtain a single average shape for C , or *mean continuum* \bar{C} . This shape is then fitted to each individual forest with a linear tilt in $\log \lambda$, i.e., $C(\hat{n}, \lambda) \equiv \bar{C}(\lambda)[a_0(\hat{n}) + a_1(\hat{n}) \log \lambda]$, where a_0 and a_1 are fitted coefficients per quasar.

Once the continuum level C is estimated for each forest, one can compute the transmission F and their *fluctuations* δ_F around the mean, defined as

$$\delta_F(\hat{n}, \lambda) = \frac{F(\hat{n}, \lambda)}{\langle F \rangle} - 1 \quad (3.2)$$

where $\langle F \rangle$ is the ensemble-averaged transmission of the Universe.

The $\langle F \rangle$ is actually an evolving function of time (or redshift or observed-frame wavelength) and it is important to take this evolution into account when computing δ_F . Several measurements of this quantity exist (Faucher-Giguère et al. 2008; Pâris et al. 2011; Becker et al. 2013; Kamble et al. 2020), thought the latest BAO measurements do not use them directly to compute δ_F . One could in principle use the forests themselves to estimate $\langle F \rangle$, by stacking forests in their observer-frame (by contrast with the mean continuum that is a stack in the quasar rest-frame), though this requires a good estimate of C . If one express δ_F as a function of the observed flux f , the continuum C and the average transmission $\langle F \rangle$, one sees that there are degeneracies. The latest BAO analyses therefore fits a linear function that models the product $C\langle F \rangle$ simultaneously.

All methods to estimate the transmission fluctuations also suffer from the fact that they use information the forests themselves to estimate C or $C\langle F \rangle$. This creates spurious correlations between a given pixel in a given forest and a neighboring pixel in the same forest. As I will discuss in section 3.2.3, spurious correlations are also introduced between pixels in different lines-of-sight, distorting the correlation function. We name this effect the *distortion by continuum fitting*; it needs to be correctly modelled when fitting the correlation function.

Typical ranges chosen to define the Ly α forest and extract the transmission field are between 1040 and 1200 Å in the quasar rest-frame. Recent analyses also consider the Ly β forest region, between 920 and 1020 Å, where both Ly α and Ly β absorption are present. The absorption in this region can be assigned a redshift that depends on the choice of λ_{rest} , that can be either Ly α = 1216 Å or Ly β = 1025 Å. These rest-frame wavelengths are separated enough such that both can be used in clustering measurements without much contamination at separations below $200 h^{-1} \text{Mpc}$.

Uncertainties of the transmission fluctuations are also estimated from the data themselves. Typically two major components are taken into account when defining uncertainties: instrumental noise and the intrinsic variance of absorbers. The former is essentially the output of the data reduction pipeline, described in Chapter 2, corrected with some normalisation factor. The latter is estimated from the variance of δ_F observed in the data. The intrinsic variance of δ_F is an increasing function of redshift, while the instrumental noise is typically decreasing with observer-frame wavelength, since forest mostly lie at the blue end of the spectrographs. The inverse of the final pixel uncertainty squared is used as a weight in the estimates of the correlation function.

Word on DLAs ?

Word on calibration residuals and my work to find/fix the problem ?

3.2.2 Two-point correlation functions

Once the transmission fluctuations δ_F and their weights w are estimated for all pixels of all forests, we compute their auto-correlation function and their cross-correlation with quasars (or any other point tracer) in three dimensions, as a function of comoving radial and transverse separations.

The first step is to convert redshifts to comoving distances using a fiducial cosmology. With angular positions and distances, we can obtain the separation $\vec{r} = (r_\perp, r_\parallel)$ between two tracers, where r_\perp is the component orthogonal to the line-of-sight and r_\parallel is the component along the line-of-sight.

The correlations are estimated in bins of separation. We consider usually bins of $4 h^{-1} \text{Mpc}$ from 0 to $200 h^{-1} \text{Mpc}$ in both radial and transverse directions. For the cross-correlation with quasars, the radial separation can be negative, meaning that the absorber is closer than the quasar from us. Let A be the index of a separation bin, the auto-correlation is defined as

$$\xi_A^{\text{auto}} = \frac{\sum_{i,j \text{ if } \vec{r}_{ij} \in A} w_i w_j \delta_{F,i} \delta_{F,j}}{\sum_{i,j \text{ if } \vec{r}_{ij} \in A} w_i w_j} \quad (3.3)$$

where the indexes i, j denote a given pixel in the forest. The sum is over all pairs of absorbers for which the separation $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$ falls inside bin A .

The cross-correlation of absorbers with quasars is estimated with the following estimator:

$$\xi_A^{\text{cross}} = \frac{\sum_{i,j \text{ if } \vec{r}_{ij} \in A} w_i w_j \delta_{F,i} \delta_{Q,j}}{\sum_{i,j \text{ if } \vec{r}_{ij} \in A} w_i w_j}, \quad (3.4)$$

which is very similar to the auto-correlation one in Eq. 3.3. The index i runs over absorbers while j runs over quasars, but there is no $\delta_{F,j}$ term. This estimator is valid under a few assumptions: sparsity of quasars, weak cross-correlation and weak auto-correlation. These assumptions are discussed in detail in the appendix B of Font-Ribera et al. 2012.

The covariance matrix is estimated using sub-samples of the full survey. Under the assumption that each sub-sample s is independent, the estimator of the covariance between ξ_A and ξ_B is

$$C_{AB} = \frac{1}{W_A W_B} \sum_s W_A^s W_B^s [\xi_A^s \xi_B^s - \bar{\xi}_A \bar{\xi}_B], \quad (3.5)$$

where W_A^s is the total weight in bin A of sub-sample s and $W_A \equiv \sum_s W_A^s$. This estimator has been tested with 100 realisations of synthetic datasets (mocks) and it shows to be robust at the current precision level.

The data vector contains around 50x50 bins, making the covariance too large (2500x2500) for the usual number of available sub-samples (~ 1000). To avoid the matrix to be singular, we apply a smoothing to it. The smoothing procedure considers that the correlation coefficients, defined as $\rho_{AB} = C_{AB}/\sqrt{C_{AA}C_{BB}}$, are only a function of $\Delta r_\perp \equiv r_{\perp B} - r_{\perp A}$ and $\Delta r_\parallel \equiv r_{\parallel B} - r_{\parallel A}$. By averaging all correlation coefficients with the same $(\Delta r_\perp, \Delta r_\parallel)$, we obtain a 50x50 matrix which is now positive definite. The new covariance matrix is constructed by taking these averaged coefficients and multiplying them by the variances. This method is used to estimate covariance matrices for both auto and cross correlation functions, but also for the cross-covariance between ξ_A^{auto} and ξ_B^{cross} , used in joint fits (see section 3.2.4).

3.2.3 Correction matrices

There are two important effects to be taken into account when modelling the Ly α correlation functions: the distortion by the continuum fitting, discussed in section 3.2.1, and the contamination by metal absorption. The most recent analyses use matrices to convolve a binned cosmological model and obtain the final template to be compared to the measured binned correlations. We describe both corrections below.

The distortion matrix is built under the assumption that the “mistake” we make when fitting the continuum C is also a linear function of $\log \lambda$. We therefore remove completely the components of $\delta_F(\lambda)$ that are proportional to a constant plus a slope. This procedure removes not only the potential mistake we made, but also the cosmological signal proportional to a constant plus a slope, though it is ok since these components should not contain much BAO information. This removal of a linear component of δ_F is referred to as *projection*². Thanks to the linearity of the problem, the model for the correlation function of the projected δ_F field can be written as a linear function of the true cosmological correlation function. For the case of a binned correlation function, this relation can be written with the help of a matrix, the *distortion matrix*. The distortion matrix was first introduced in Bautista et al. 2017 for the BAO analysis of DR12 forests, where we showed that the distortion matrix was correctly describing the clustering of mock forests after continuum fitting.

The contamination due to *metal*³ absorption is mainly due to four transitions of silicon atoms for which their rest-frame wavelength is close to the Ly α transition. When computing the auto-correlation of forests, metal absorption creates confusion because it is interpreted as a distinct Ly α absorber when in fact it is the same. For example, the SIII absorption happens at $\lambda_{\text{rest}} = 1207 \text{ \AA}$, close to Ly α for which $\lambda_{\text{rest}} = 1216 \text{ \AA}$. In comoving coordinates, assuming a standard fiducial cosmology, this corresponds to $\Delta r_\parallel \approx 21 h^{-1} \text{Mpc}$ at $z \sim 2.3$. At these radial separations and when $r_\perp \sim 0$, we see a spike in the correlation function. This peak is not due

²This is equivalent to the removal of modes contaminated by photometric systematics in galaxy clustering, as studied in Paviot et al. 2022.

³The term *metal* is used in astronomy for elements heavier than helium.

to Ly α absorption separated by $21 h^{-1}\text{Mpc}$ but it is the cross-correlation between Ly α and SIII absorption at near *zero* separation. It is the same region in the Universe absorbing both in Ly α and in SIII, though the latter has a smaller amplitude. This happens for all metals with rest-frame wavelenghts close to Ly α . For instance, one of the SiII transitions creates a spike at $111 h^{-1}\text{Mpc}$, right at the BAO peak. Since Delubac et al. 2015, I worked on carefully adding metal absorption to mock forests in order to quantify their impact on BAO constraints. In Bautista et al. 2017, we introduced a model for the Ly α -metal cross-correlation to be fitted altogether with the BAO model. Similarly to the distortion modelling, the Ly α -metal correlations can be described as a matrix times the Ly α -Ly α correlation function. This matrix is referred to as the *metal matrix*. The amplitude of Ly α -metal correlations is a free parameter in the BAO fits. We found that marginalising over these amplitudes makes BAO constraints slightly larger, though more conservative and robust.

3.2.4 The model

The model used to fit for the BAO scale on the measured correlation functions is based on a linear matter power spectrum with some empirical terms that account for redshift-space distortions, non-linear clustering, high column density systems and binning effects. The model is constructed in Fourier space and it is Fourier transformed to obtain $\xi(r_\perp, r_\parallel)$. The latest version of the model is described in du Mas des Bourboux et al. 2020 and I shortly present it in this section.

The model can be written, in Fourier space, as

$$P^{\text{model}}(\vec{k}) = b_i b_j (1 + \beta_i \mu_k^2) (1 + \beta_j \mu_k^2) P_{\text{QL}}(\vec{k}) F_{\text{NL}}(\vec{k}) G_{\text{bin}}(\vec{k}), \quad (3.6)$$

where \vec{k} is the wavevector, with modulus k and $\mu_k = k_\parallel/k$; b_i and β_i are the linear bias and redshift-space distortions parameters, respectively, and i is a index referring to absorbers or quasars; G_{bin} accounts for the binning of the correlation function, F_{NL} is a empirical term that accounts for the non-linear effects on small scales, and P_{QL} is the linear matter power spectrum with a empirical anisotropic damping applied to the BAO peak component. The linear matter power spectrum is computed from a Boltzmann solver code, such as CAMB (Lewis et al. 2000).

The non-linear term F_{NL} is only included for the quasar-Ly α cross-correlation and accounts for the effect of Fingers-of-God on small-scales as well as an eventual contribution from uncertainties of redshift estimates, which can be relatively large for quasars ($\sim 500 \text{ km/s}$). This term can be modelled as a Gaussian or as a Lorentzian, as a function of a velocity dispersion parameter σ_v (in units of km/s). The impact of redshift errors was studied in detail in Youles et al. 2022 and it is also described in section 3.3.

The effect of high-column density systems, if identified and modelled when fitting the continuum, is negligible. For the HCDs that are not identified, the extended absorption (called “wings”) causes k_\parallel -dependent broadening of the clustering (Font-Ribera and Miralda-Escudé 2012; Rogers et al. 2018). This effect is modelled through effective biases b_i^{eff} and RSD parameters β_i^{eff} that depend on the biases of the Ly α forest and those from HCDs, plus a function that depends on the distribution of column densities of the unidentified HCDs.

The metal correlations are accounted as additive terms to the Ly α correlations

$$\xi_A^{\text{model}} = \xi_A^{\text{Ly}\alpha \times \text{Ly}\alpha} + \sum_m \xi_A^{\text{Ly}\alpha \times m} + \sum_{m_1, m_2} \xi_A^{m_1 \times m_2}, \quad (3.7)$$

where each $\xi^{m_1 \times m_2}$ can be written as a metal matrix $M_{A,B}$ times the $\xi_B^{\text{Ly}\alpha \times \text{Ly}\alpha}$ as mentioned in section 3.2.3. The metal matrix is defined as

$$M_{A,B} = \frac{1}{W_A} \sum_{(m,n) \in A, (m,n) \in B} w_m w_n, \quad (3.8)$$

where $(m, n) \in A$ refers to pixel separations computed assuming Ly α absorption, while $(m, n) \in B$ refers to pixels separations computed assuming metal absorption. This formalism was introduced in Blomqvist et al. 2018.

Quasars emit UV photons which ionise their surroundings slightly more than the average intergalactic medium. This is known as the *proximity effect* and it changes the average properties of Ly α absorption and therefore its clustering. When considering forests around a given quasar but not to its own forest, we refer to it as the transverse proximity effect. This effect is accounted for in the quasar- Ly α cross-correlation with the following parametric form:

$$\xi^{TP} = \xi_0^{TP} \left(\frac{1 h^{-1} \text{Mpc}}{r} \right)^2 \exp(-r/r_{UV}) \quad (3.9)$$

where $r^2 = r_{\parallel}^2 + r_{\perp}^2$, $r_{UV} = 300 h^{-1} \text{Mpc}$ is fixed and ξ_0^{TP} is a fitted amplitude.

The spectroscopic data reduction pipeline described in section 2.3 also introduces spurious correlations between spectra. For instance, sky-subtraction and flux calibration are two processes that induce correlation between two pixels at the same observed wavelength but in two distinct lines-of-sight, provided they are observed with the same spectrograph (set of 500 fibers). These correlations have been studied using quasar spectral regions deprovided of Ly α absorption, such as the MgII spectral region ($\lambda_{\text{rf}} \in [2600, 2760] \text{\AA}$). An empirical model for this effect is included: a Gaussian, function of r_{\perp} , centered in $r_{\perp} = 0$ but only when $r_{\parallel} = 0$. Due to the effect of distortion caused by continuum fitting, these spurious correlations due to data reduction also propagate to pairs with $r_{\parallel} \neq 0$. The amplitude and characteristic length of this effect are two free parameters in the fit of the Ly α auto-correlation function.

The correlation function model accounts for the distortion matrix, as discussed in section 3.2.3. The distorted correlations are written as

$$\xi_A^{\text{dist}} = \sum_{A'} D_{AA'} \xi_{A'}^{\text{model}}. \quad (3.10)$$

It is also possible to add arbitrary smooth functions of separation in order to account for any unexplained correlations and improve the fit to the data. These functions are used to marginalise over any potential information on BAO coming from the full-shape of the correlation functions, which is avoided. This is commonly done in galaxy clustering analysis as I will show in Chapter 4.

3.2.5 BAO constraints

The last essential ingredient to be added to the model of correlations are the Alcock-Paczynski (AP) parameters. Thanks to the observed baryon acoustic scale, which we know should be isotropic and have a comoving size of $r_d \sim 105 h^{-1} \text{Mpc}$, the AP parameters can be well constrained with the data. These AP parameters account for small differences between the assumed fiducial cosmology, used to convert redshifts into distances, and the measured cosmology extracted from the observed BAO peak position.

We commonly use two AP parameters, encoding radial and transverse dilation effects. They

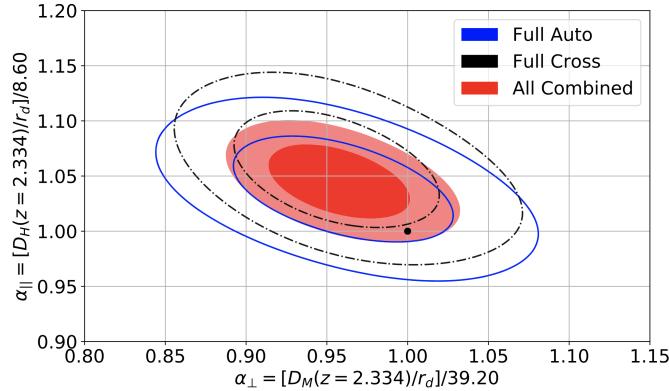


Figure 3.1: BAO constraints from the SDSS DR16 sample of Ly α forests. The contours show the 68 and 95 per cent confidence levels for the Alcock-Paczynski parameters α_{\perp} and α_{\parallel} from the fits to the Ly α -Ly α auto-correlation (blue solid), the quasar-Ly α cross-correlation (black, dash-dotted), and the joint fit of auto+cross (red shades). Given the high level of shot-noise of the quasar sample, the covariance between constraints from the auto and cross correlation are found to be negligible. Extracted Figure 12 from du Mas des Bourboux et al. 2020.

can be written as ratios of the observed Δz_{BAO} and $\Delta \theta_{\text{BAO}}$ at a given effective redshift z_{eff} from equations 1.22 and 1.21 of section 1.5.4, yielding

$$\alpha_{\parallel}(z_{\text{eff}}) = \frac{D_H(z_{\text{eff}})}{D_H^{\text{fid}}(z_{\text{eff}})} \frac{r_d^{\text{fid}}}{r_d} \quad (3.11)$$

$$\alpha_{\perp}(z_{\text{eff}}) = \frac{D_M(z_{\text{eff}})}{D_M^{\text{fid}}(z_{\text{eff}})} \frac{r_d^{\text{fid}}}{r_d} \quad (3.12)$$

where D_H and D_M are respectively the Hubble and comoving angular diameter distances (Eqs. 1.2 and 1.4).

The AP parameters α_{\parallel} and α_{\perp} are used to scale separations when computing the correlation function model as

$$\xi(r'_{\perp}, r'_{\parallel}) = \xi(\alpha_{\perp} r_{\perp}, \alpha_{\parallel} r_{\parallel}), \quad (3.13)$$

and they are let free when fitting correlations. The BAO constraints are defined by the posteriors of α_{\perp} and α_{\parallel} , after marginalising over all other free parameters of the model.

Figure 3.1 shows the current BAO constraints from the SDSS DR16 sample of Ly α forests (containing both Ly α and Ly β absorption) and quasars (du Mas des Bourboux et al. 2020). The marginalised constraints from the joint fit of auto and cross-correlations are

$$D_H(z = 2.334)/r_d = 8.99^{+0.20}_{-0.19} \quad (3.14)$$

$$D_M(z = 2.334)/r_d = 37.5^{+1.2}_{-1.1} \quad (3.15)$$

$$\rho(D_H/r_d, D_M/r_d) = -0.45 \quad (3.16)$$

which represent a 2.2 per cent BAO measurement in the radial direction and a 3.2 per cent in the transverse direction.

3.3 Impact of redshift errors

This section concerns work led by Samantha Youles, PhD candidate under my supervision at the University of Portsmouth between 2018 and 2022. The full details are presented in Youles et al. 2022.

A key element in the BAO analysis using forests is the quasar redshifts, which can be challenging to measure precisely when compared to galaxy redshifts. Quasar spectra are basically composed of thermal continuous emission from their accretion disk around their supermassive black holes in the centre of their host galaxies plus some broad emission lines. Typical quasar spectra from SDSS or DESI do not present detected narrow absorption or emission lines from the host galaxy itself, which would considerably improve the uncertainties on their redshifts. Previous studies using quasars as tracers for clustering (Zarrouk et al. 2018; Lyke et al. 2020) estimated that quasar redshift uncertainties range from 100 km/s at $z \sim 1$ to 400 km/s at $z \sim 2$, while galaxies (LRGs or ELGs) have redshift uncertainties below ~ 50 km/s. Of course, these uncertainties depend on the spectrograph and on which broad emission lines are visible for a given quasar.

In addition to statistical uncertainties in the determination of redshift, Shen et al. 2016 has shown that redshifts derived from broad emission lines of quasars have intrinsic scatter and systematic shifts of the order of hundreds of km/s with respect to the host galaxy redshift. Current redshift fitting algorithms based on templates do not account for these shifts. While Youles et al. 2022 studies the impact of scatter, we did not study in detail the impact of systematic shifts.

The uncertainties in the redshift measurements smear the two-point statistics of the quasars over separations along the line-of-sight. The smearing directly impacts the BAO measurement, particularly reducing the signal-to-noise ratio of estimates of α_{\parallel} (see section 3.2.4).

Non-linearities induce large peculiar velocities on scales below a few Mpc and are commonly modelled as a random scatter in the radial direction. These non-linearities are known as *fingers-of-God* (FoG) and they smear the two-point statistics in a similar manner than redshift errors, though I will show the differences.

In this section, these two types of extra scatter in quasar redshifts are parametrised by

- $\sigma_{v,z}$ as the scatter induced by uncertainties in redshift determination, or redshift errors.
- $\sigma_{v,\text{FoG}}$ as the scatter induced by non-linear motions on small scales, or Fingers-of-God.

Using DESI mock catalogues of the 5-year sample of Ly α forests and quasars, we studied the impact of each of these types of extra scatter on the clustering and on BAO constraints. We discovered a new feature arising from redshift errors that modifies the shape of the two-point statistics in a non-trivial manner, though it does not impact BAO constraints expected with DESI.

Figure 3.2 presents correlation functions, both for the quasar- Ly α cross-correlation and for the Ly α auto-correlation, in averages over $\mu_r \equiv r_{\parallel}/r$. There are three sets of mocks

- 1) fiducial ones where $\sigma_{v,\text{FoG}} = 150$ km/s and $\sigma_{v,z} = 0$,
- 2) mocks with $\sigma_{v,\text{FoG}} = 500$ km/s and $\sigma_{v,z} = 0$ and
- 3) mocks with $\sigma_{v,\text{FoG}} = 150$ km/s and $\sigma_{v,z} = 500$ km/s.

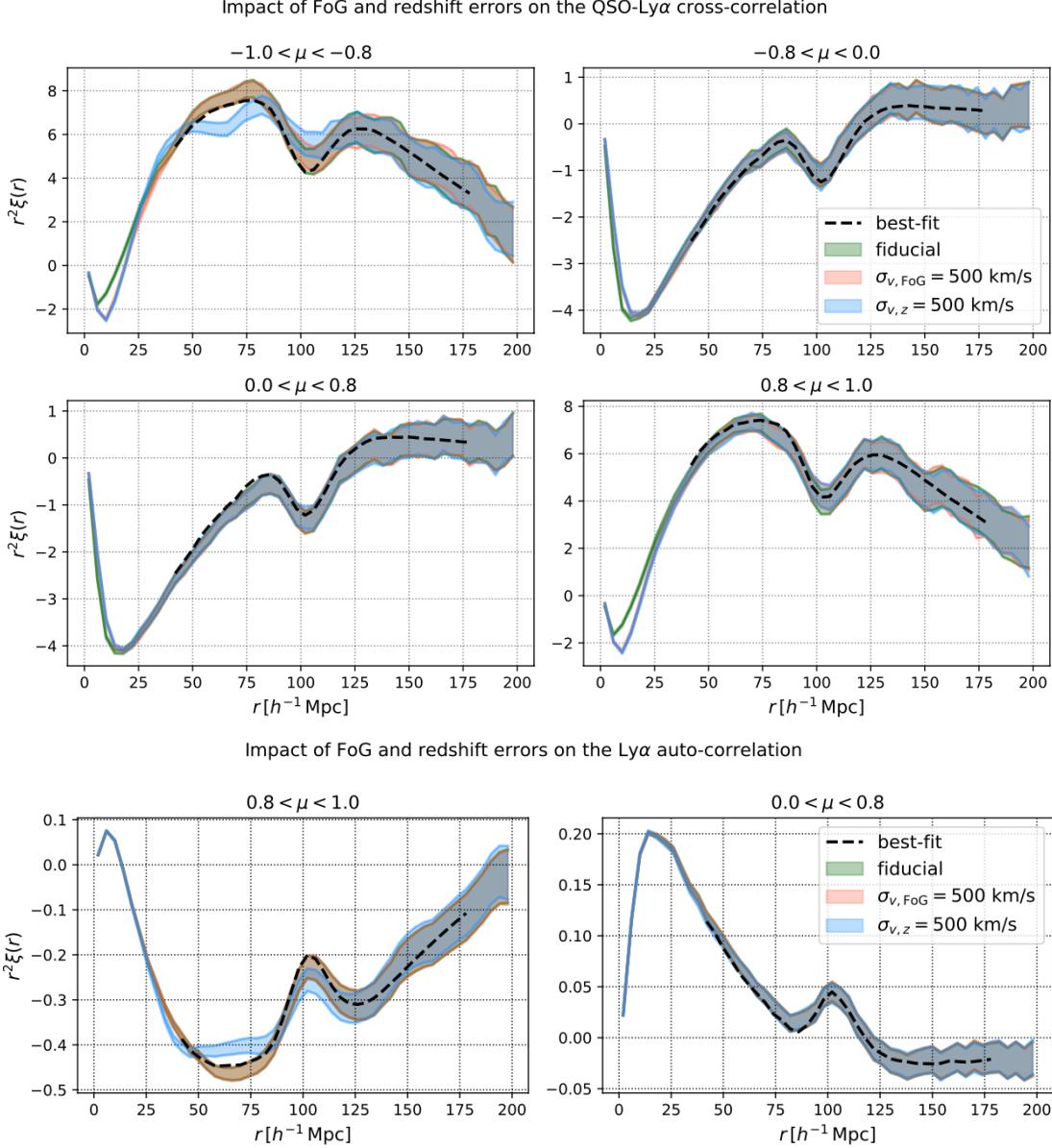


Figure 3.2: Wedges of the correlation functions versus comoving separation. Top four panels show the QSOx Ly α cross-correlation (top four panels) while the bottom two panels show the Ly α auto-correlation of forests. These are estimated from DESI mock catalogues reproducing the full 5-year sample. The green bands show the measurement from the standard mocks, with a small value of $\sigma_{v,\text{FoG}} = 150 \text{ km/s}$ and no redshift errors; the pink bands (indistinguishable from the green in the cross-correlation) correspond to the mocks generated with an extreme value of $\sigma_{v,\text{FoG}} = 500 \text{ km/s}$; the blue bands show the results for mocks with large redshift errors of $\sigma_{v,z} = 500 \text{ km/s}$. These measurements are computed from the average of 10 realisations of the complete 5-year DESI survey, and the width of the bands corresponds to the scatter between realisations. Note that $\mu > 0$ ($\mu < 0$) corresponds to configurations where the Ly α pixels are behind (in front of) the neighbouring quasar.

Each set of mocks contains ten realisations, results are averaged for each set. The colour bands show the scatter amongst realisations and are equivalent to the noise expected for the 5-year sample. We can see that FoG (orange bands) do not have a significant impact except on the cross-correlation on small scales. However, redshift errors do create a new oscillatory feature both in the auto and cross-correlations at separations of $20\text{--}50 h^{-1}\text{Mpc}$ but only those along the line-of-sight. Using other mock catalogues with increasing values of $\sigma_{v,z}$ (not shown here) makes these features stronger. This work puts in evidence this effect for the first time.

In Youles et al. 2022, we claim that the oscillatory features appearing when adding redshift errors to quasars is due to two main factors: the smoothing of emission lines in the continua and the clustering of the quasars. These two factors *combined* create this new effect. The smoothing of the mean continuum is illustrated in Figure 3.3 for increasingly large values of $\sigma_{v,z}$. These oscillatory features defined by the difference between true and estimated continua are imprinted on the clustering thanks to the fact that quasars themselves are clustered.

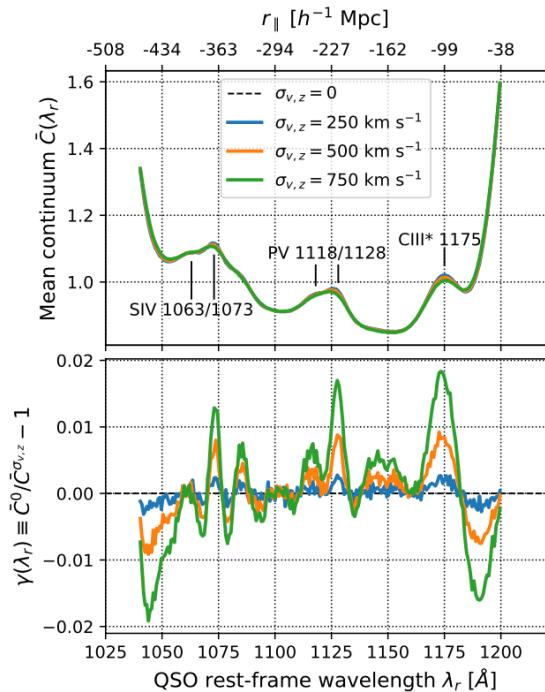


Figure 3.3: Effect of redshift errors on the mean continuum. The top panel shows the mean continua for no-forest mocks without errors, and with increasingly large redshift errors: $\sigma_{v,z} = 0, 250, 500$ and 750 km/s . The positions of some of the strongest emission lines are annotated. The lower panel shows the ratio between the continua with and without redshift errors. The top axis on the top panel shows the radial separations between pixels and the quasar, assuming that the quasar is at redshift $z_q = 2.3$.

We tested our hypothesis by creating two special sets of mock catalogues, one where quasars are deprovided of forest absorption, and one where quasars are not clustered. With the first set of mocks, features are still present with the same amplitudes as the standard mocks. This points to the fact that the Ly α forest and their correlations do not play a role in this effect. For the second set of mocks, the features disappear in the correlation functions, demonstrating

that the quasar clustering is a key ingredient to explain the effect.

While this model is satisfactory to explain the features in the cross-correlation, the auto-correlation seems to require an additional ingredient that we could not identify. We leave this investigation for future work.

We performed BAO fits on all sets of mocks for different values of $\sigma_{v,z}$ and $\sigma_{v,\text{FoG}}$. We employ the exact same model described in section 3.2.4, though we did not include metals or high-column density systems. We do not attempt to model the effect of redshift errors when fitting for BAO. We see an increased value for the χ^2 of the fits when increasing the value of $\sigma_{v,z}$, simply because these features are not modelled properly. The inferred BAO parameters α_{\parallel} and α_{\perp} do not see significant shifts or trends versus $\sigma_{v,z}$ other than a slight increase in their uncertainties.

We believe that for the final analysis of the 5-year DESI sample the impact of redshift errors has to be correctly modelled in order to reduce the uncertainties in BAO parameters or to extract cosmological information from the full-shape of the correlation function Cuceu et al. 2021.

3.4 Weak-lensing of forests

This section also concerns work led by Samantha Youles, PhD candidate under my supervision at the University of Portsmouth between 2018 and 2022.

Weak-gravitational lensing is a powerful cosmological probe since it is sensitive to the total (baryons and dark) matter density field, as discussed in section 1.5.7. The quasar and Ly α forest samples used in the measurement of BAO lie at $2.0 < z < 4.0$ and are subject to the effect of lensing from the matter distribution between $0 < z < 2.0$. Current BAO analysis such as the one described in section 3.2 do not take any lensing effect into account. In this section we describe how one could potentially extract weak-lensing information from a realistic set of Ly α forests.

Until now, no measurement of weak gravitational lensing with sources at $z \sim 2$ has been produced, though it has been studied theoretically and with simplified simulations. Croft et al. 2018, Metcalf et al. 2018, and Metcalf et al. 2020 looked into a regular grid of mock quasars with arcminute separations, populated with Gaussian random forests. The close, regular spacing of the forests created a near-continuous field enabling the use of a quadratic estimator adapted from 21cm intensity mapping analyses Poutsidou and Metcalf 2014; Poutsidou and Metcalf 2015. Even though these techniques are able to reconstruct the lensing potential, they are not yet readily applicable to real survey data given its non-regular sampling, noise, volumes and densities probed.

Figure 3.4 illustrates the impact of lensing over a pair of Ly α forests caused by an intervening overdensity. The lens creates an image of the pair at a slightly larger angle with respect to the observer, virtually increasing the transverse separation r_{\perp} between the two forests to a larger value r_{\perp}^L . Of course, the expected correlation function between two given pixels is unchanged and given by $\xi(r_{\perp}, r_{\parallel}) = \langle \delta_1 \delta_2 \rangle$ where the brackets indicate an ensemble average. When observing a set of lensed quasars, the measured correlations are wrongly assigned to the separation given by $(r_{\perp}^L, r_{\parallel}) \neq (r_{\perp}, r_{\parallel})$. Here we assume that the radial separations changes are negligible, $r_{\parallel}^L \sim r_{\parallel}$. If one could measure the correlation function uniquely using pairs of forests that are background of overdense lenses, the results would be a dilated correlation function in the transverse direction. Alternatively, considering only underdense lenses, the effect would be an opposite dilation in the transverse direction.

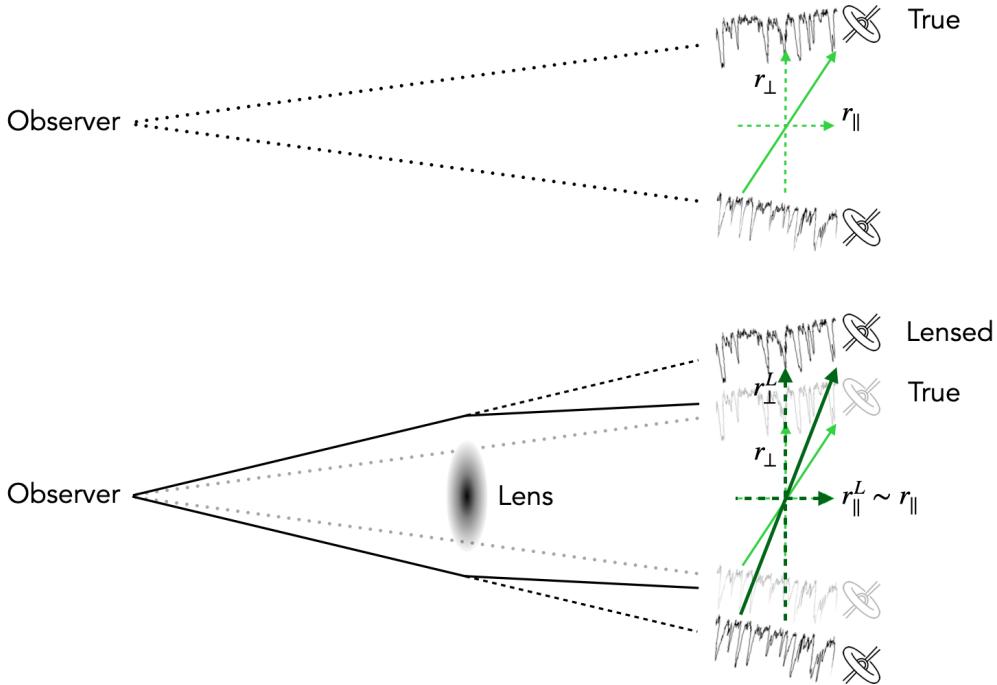


Figure 3.4: Illustration of the effect of gravitational lensing over a pair of Ly α forests. The top figure shows the unlensed case, the arrows define the parallel and transverse separations, r_{\parallel} and r_{\perp} respectively, of a pair of Ly α forest pixels (or a pixel-quasar pair). The bottom panel shows the lensed case, where an overdense region acts as a “thin” lens, creating the illusion that the background pairs are actually farther away from each other in the transverse direction. The measured correlations should be the same in both cases, though in the lensed case this correlation is assigned to a wrong transverse separation. The radial effect is relatively small.

I describe how we attempt to reconstruct the lensing signal from a simpler estimator. One should first note that when averaging correlations over all available pairs in a survey, the lensing effect from over and underdense lenses averages to a negligible net impact on the correlation function, though it increases its noise. However, we can attempt to detect a difference between the average correlation function $\bar{\xi}(r_{\parallel}, r_{\perp})$ and the local correlation function in a given patch of the sky $\xi(\hat{n}, r_{\parallel}, r_{\perp})$. We can write a first-order approximate relation between the local difference $\Delta\xi \equiv \xi(\hat{n}, r_{\parallel}, r_{\perp}) - \bar{\xi}(r_{\parallel}, r_{\perp})$ and the local convergence field $\kappa(\hat{n})$. In practice, each pair of forest given us an estimate of $\xi(\hat{n}, r_{\parallel}, r_{\perp})$, though very noisy. Our estimates of the convergence $\hat{\kappa}(\hat{n})$ are assigned into a map by choosing \hat{n} to be the mid-point between each pair of lines of sight. Convergence maps are constructed using both the auto-correlation function of forests and the quasar- Ly α cross-correlation function, denoted $\hat{\kappa}_{\alpha\alpha}$ and $\hat{\kappa}_{q\alpha}$ respectively. Given the fact that auto- and cross-correlations are close to be independent (see section 3.2.2), the convergence maps from auto and cross can be combined into a single map, denoted $\hat{\kappa}_{q\alpha+\alpha\alpha}$.

Angular power spectra are computed from the estimated maps as a function of angular multipole number ℓ . Since our convergence maps are quite noisy, computing the auto power spectra of these maps, e.g. $\langle \hat{\kappa}\hat{\kappa} \rangle_{\ell}$, is not convenient since these would also be dominated by noise. Instead, we use high signal-to-noise ratio convergence maps to compute cross angular

power spectra. In real cases, one can use a lensing convergence or potential maps derived from the cosmic microwave background or one can construct a density map of foreground sources from galaxy surveys.

We tested our methodology using simulations. We used the 5-year DESI mock catalogues of Ly α forests also used in the study of redshift errors (section 3.3). Gaussian lensing convergence maps were created from a fiducial cosmology and the linear matter power spectrum computed with *camb*. We displaced the quasars following these “input” convergence maps and applied our methodology to build convergence maps from the auto- and cross-correlation functions. The analysis was performed 100 times with different realisations of the convergence maps (the forests were kept the same).

Figure 3.5 shows the resulting power spectra compared to the input convergence maps as a function of multipole ℓ . There are two effects to be taken into account when comparing to theory. First, the survey mask changes the amplitude of the power spectra by a factor equal to the fraction of sky covered (to first order). The resulting power is shown as a blue line. Second, the power spectra depend on the choice of maximum transverse separation r_{\perp}^{\max} between quasars used to build the convergence maps. We refer to this non-trivial function of ℓ describing the variation of power with r_{\perp}^{\max} as “damping”. Both effects combined are shown as a purple line.

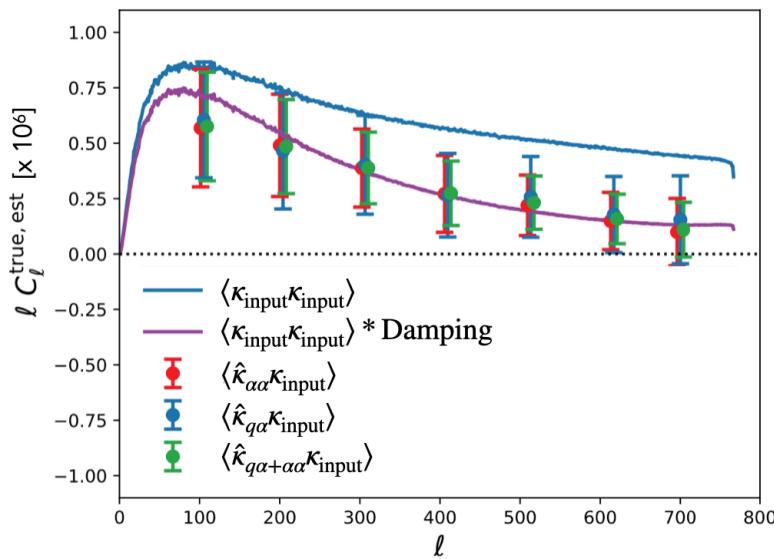


Figure 3.5: Angular power spectra of convergence maps from simulations of the 5-year DESI sample of Ly α forests. The blue line indicates the auto power spectra of the input convergence when accounting for effect of survey mask. The purple line also includes the effect of the choice of maximum transverse separation between lines of sight used when building the convergence maps. The points show the estimated cross power from maps computed from the auto-correlation of forests (red), the cross-correlation between quasars and forests (blue), and the combination of both (green).

There is an excellent agreement, though noisy, between estimated cross-power spectra and their input values. Of course, the “input” map is not available in a real setting so one should consider the noise of this second map (e.g. CMB lensing map) when producing realistic

forecasts.

The results presented in this section are still under development and have not yet been published. We refer the reader to S. Youles thesis for more details on methods and tests.

3.5 Voids in tomographic maps

Ravoux et al. 2022

4 The mid-redshift Universe and its galaxies

Contents

4.1	From redshifts to clustering catalogues	45
4.2	Reconstruction of linear density field	50
4.3	Mock catalogues	51
4.4	From catalogues to clustering estimates	52
4.4.1	Correlation function	52
4.4.2	Power spectrum	53
4.4.3	The clustering of eBOSS DR16 LRGs	55
4.5	Baryon acoustic oscillations	57
4.6	Redshift-space distortions	59
4.7	Joint clustering analysis in Fourier and Configuration space	63
4.8	Cross-correlation with radio surveys	66

At redshifts between 0 and 2, galaxies can be used as tracers of the matter distribution. With the statistics of the galaxy distribution we can measure the characteristic scale of the baryon acoustic oscillations (BAO) as well as extract information about the growth-rate of structures from the anisotropies caused by redshift-space distortions (RSD). As presented in Chapter 1, BAO and RSD are powerful probes of dark-energy and theories of gravity.

In this chapter, I overview my contributions for the study of dark-energy with galaxy clustering. Section 4.1 describes how to create a galaxy clustering catalogue from the spectroscopic observations, including corrections for all known systematic effects caused by photometry or spectroscopy. Section 4.2 describes reconstruction techniques used to increase the precision of BAO measurements. Section 4.3 presents the mock catalogues that are used to test our methodology and section 4.4 I overview the two-point statistics estimates, either in Fourier space or in configuration space. In section 4.5 I present the BAO measurements I performed with the SDSS sample of luminous red galaxies, while section 4.6 focus on the RSD constraints. Section 4.7 overviews recent work carried out by my PhD students Tyann Dumerchat and Vincenzo Aronica on the joint analysis of galaxy clustering in Fourier and configuration space.

The work described in this chapter is published in the following articles: Bautista et al. 2018; Bautista et al. 2021; Gil-Marín et al. 2020; A. J. Ross et al. 2020; Zhao et al. 2021.

4.1 From redshifts to clustering catalogues

An essential step in the cosmological analysis of galaxy survey data is to convert convert a list of galaxy redshifts (see Chapter 2) into a catalogue from which we can define overdensities $\delta_n(\vec{x}) = n(\vec{x})/\bar{n} - 1$, where $n(\vec{x})$ is the number density of galaxies in a volume element located at position \vec{x} . The quantity \bar{n} is the average galaxy number density over the probed *volume*. Therefore, it is important to 1) define precisely what is the volume observed, or in jargon terms,

the survey window function; and 2) to make sure that the number density of galaxies is tracing the actual cosmological fluctuations and not any spurious contamination.

The simplest form of survey window function would be a function of position \vec{x} that yields 1 if \vec{x} is inside the survey volume and 0 else. One convenient way to define this function is using a large set of randomly and uniformly distributed positions encompassing the survey volume. Each random position would have a weight of either 1 or 0, depending if it is located inside or outside the observed volume. This list of positions and their weights is referred to as the *random catalogue* or the *randoms*. Typically, randoms are sampled at some arbitrary higher number density than the galaxy average number density (typically between 20 or 50 times higher), to reduce the noise in the description of the window function. As we will see next, the random catalogue is slightly more complex: the weights are used to account for observational completeness and systematic effects. In SDSS analyses, the random catalogue is a product of an angular and a radial selection function.

I describe now the procedure to define the angular footprint. The starting footprint is the one used in the process of target selection (section 2.1). Figure 4.1 displays the eBOSS footprint for three tracers (LRGs, ELGs and QSOs) compared to the BOSS footprint and the stellar density map from Gaia. Regions with bad photometric or around bright stars are masked out by removing randoms belonging to these regions (or assigning them a null weight). After tiling, fibre assignment and spectroscopic observations, the footprint can be divided into an unique set of sectors. Each sector corresponds to regions observed by one or more plates. The *fibre completeness* of each sector is defined as the ratio between the number of spectroscopically observed targets and the number of available targets in the sector. Randoms are sub-sampled or de-weighted following the fibre completeness to correct for its effect.

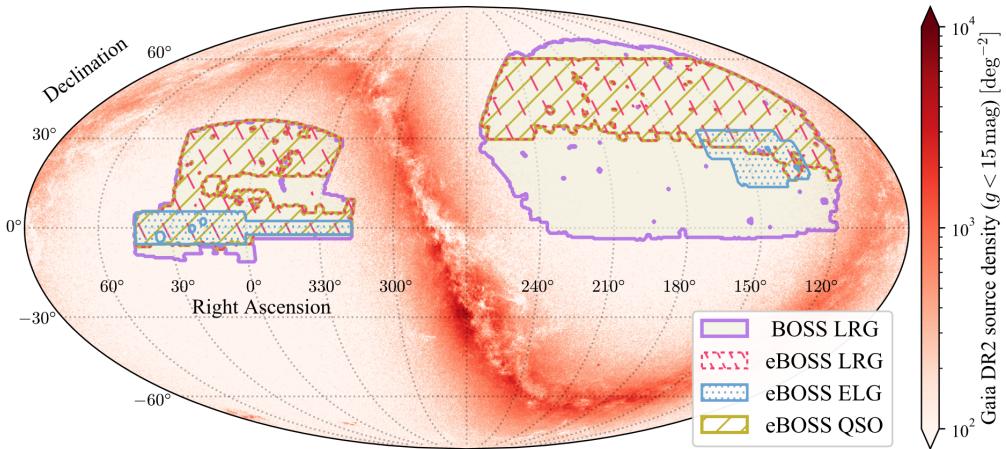


Figure 4.1: The DR16 footprint for each eBOSS tracer: Luminous Red Galaxies (LRG), Emission Line Galaxies (ELG) and quasars (QSO). The BOSS DR12 LRGs area is also shown, as well as the density map of Gaia DR2 sources with $g < 15$ mag. Figure extracted from Zhao et al. 2021.

Some targets cannot be observed due to its proximity to another observed target and the finite size of fibres, corresponding to 62 arcseconds in the sky. We refer to these events as *fibre collisions*. Collisions might happen not only to pairs of galaxies, but to any group of targets with linking lengths smaller than 62 arcsec. Depending on the number of plates observing a giving sector, some collisions might be solved but the missing ones impact the measured

clustering on small scales if not corrected. While there are several methods to solve collisions (Guo et al. 2012; Bianchi and Percival 2017), we used the simplest up-weighting technique, where the nearest observed galaxy is up-weighted by the number of non-observed targets within the collision group. This assumes that non-observed targets are also galaxies of the same target type and that they are physically close (angularly and radially) to the observed ones. On scales above a few $h^{-1}\text{Mpc}$, this simple correction is a good approximation.

Two additional sets of weights are defined to correct for spurious density fluctuations contaminating cosmological fluctuations in which we are interested. These weights could be applied to randoms but we chose to apply these to the galaxy themselves.

The first set of weights corrects for angular fluctuations caused by correlations between galaxy number density and photometry-related quantities, such as Galactic extinction, stellar density, imaging depth or sky flux. They are commonly referred to as *photometric weights*. Figure 4.2 displays these correlations for the LRG sample both before and after applying correction weights. For the analysis of eBOSS DR16 LRGs, I have implemented¹ a multi-dimension linear regression that assumes these correlations are linear and accounts for the fact that some photometric quantities are correlated between themselves (e.g., stellar density and Galactic extinction). This multi-linear method is based on work described in Prakash et al. 2016. Of course, more advanced methods have been developed since then, based on machine learning algorithms (Rezaie et al. 2020; Chaussidon et al. 2021). It is important to evaluate the performance of these algorithms with mock catalogues where we add artificial contamination and quantify how well we recover the initial raw power spectrum after correcting for them. There is a risk of overfitting and removing some large-scale cosmological modes which can be harmful for clustering analysis, particularly those aiming at measuring primordial non-Gaussianity (Rezaie et al. 2021; Mueller et al. 2021).

¹https://github.com/julianbautista/eboss_clustering/blob/master/python/systematic_fitter.py

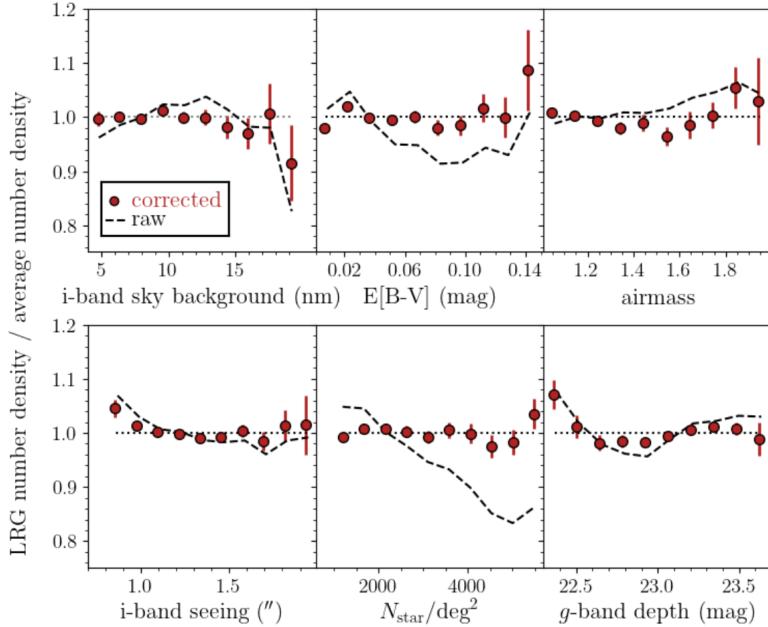


Figure 4.2: Fluctuations in the angular LRG density as a function of various imaging properties and Galactic foregrounds. The dashed curves show these fluctuations before any correction while red points show the result of applying a linear correction for stellar density and Galactic extinction. Figure extracted from A. J. Ross et al. 2020.

The second set of weights corrects for the so-called *redshift failures*, which are spectra with low signal-to-noise ratio (S/N) not yielding a statistically significant redshift measurement. If redshift failures were randomly distributed across the sky, they would not introduce any particular bias to the clustering. However, given the particular configuration and throughput of the SDSS spectrographs, these failures imprint a pattern in the sky which contaminates clustering. The left panel of Figure 4.3 displays the average failure rate of eBOSS LRGs as a function of position in the focal plane. This pattern is due to the fact that fibres at the side edges of the focal plane transport light to the edges of the CCDs which have a lower-than-average throughput. The right panel of Figure 4.3 shows the dependency of the average redshift failure rate as a function of the signal-to-noise ratio of the observation². When modelling these dependencies of failure rates with S/N (or fibre number or focal plane position), one can correct these spurious fluctuations by assigning a weight inversely proportional to the modelled rate. These *redshift-failure weights* are assigned to the galaxies with confident redshift measurements, to compensate for the eventual non-confident ones at the same location in the focal plane.

²The signal-to-noise ratio of an observation was defined internally as a function of the signal-to-noise ratio of all observed spectra and their magnitudes.

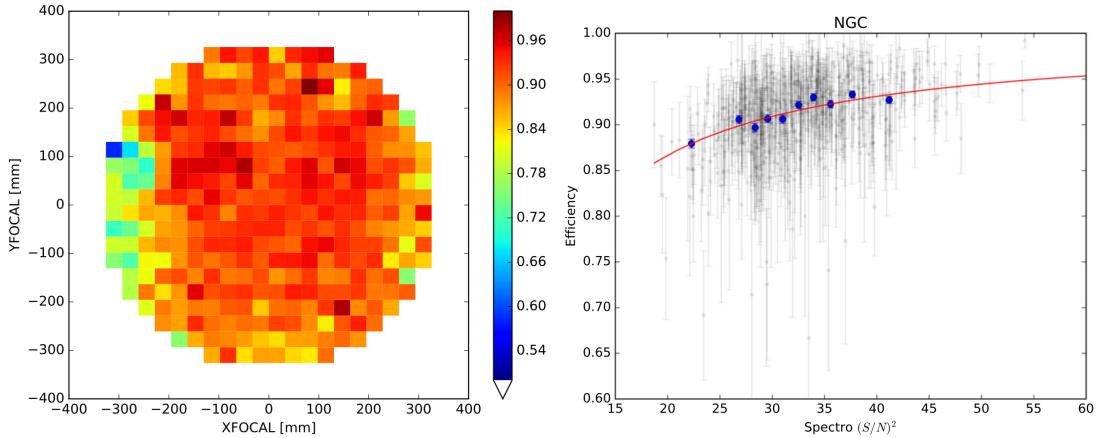


Figure 4.3: Average redshift efficiency as a function of physical position of the optical fiber in the focal plane (left panel) and as a function of the observation signal-to-noise squared (right panel). The left panel shows how redshift-failures imprint a pattern on the sky and therefore must be corrected. The right panel shows the underlying reason for the failures. The model (red line) is used to compute correction weights. Figures extracted from Bautista et al. 2018.

Once all galaxies have their correction weights (fibre collision, photometric and redshift-failure) and the randoms are corrected by fibre completeness, it is time to compute the radial selection function. The idea is to assign redshifts to randoms such that they follow the same redshift distribution as the galaxies. The redshift distribution is often quantified by the comoving galaxy (weighted) number density $\bar{n}(z)$. Figure 4.4 shows the comoving density of eBOSS tracers as a function of redshift. For randoms to match this distribution, we randomly assign galaxy redshifts to each random, with repetition. Another possibility would be to fit some smooth function over the observed $\bar{n}(z)$ and draw random redshifts from the resulting model. All these methods introduce the so called *radial integral constraint*, i.e., the fact that the observed $\bar{n}(z)$ is not the true ensemble averaged function of the Universe, but derived from the volume-limited realisation that we observe. The impact on clustering of the radial integral constraint is inversely proportional to the area of the footprint (de Mattia and Ruhlmann-Kleider 2019) and it is significant for surveys such as the eBOSS ELG (see Figure 4.1 and Tamone et al. 2020; de Mattia et al. 2021).

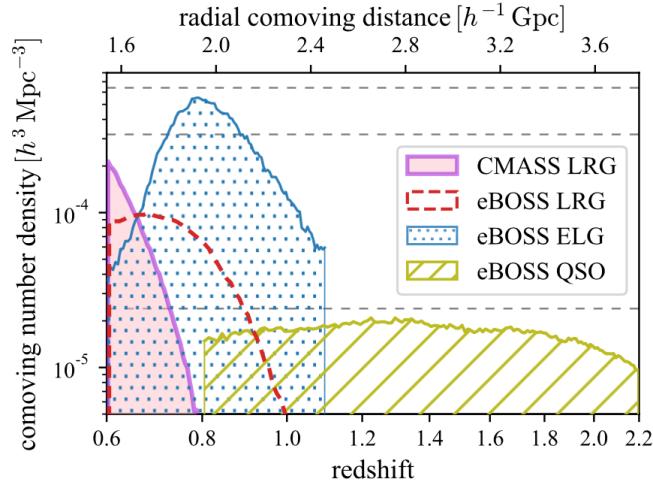


Figure 4.4: The weighted comoving number densities $\bar{n}(z)$ of eBOSS DR16 tracers and BOSS DR12 CMASS LRGs, with all the photometric and spectroscopic systematic weights included. The comoving distances and volumes are computed assuming a flat- Λ CDM cosmology with $\Omega_m = 0.31$. Figure extracted from Zhao et al. 2021.

The final step of the catalogue production consists in adding an additional set of weights to galaxies that optimise the signal-to-noise ratio of clustering measurements when the redshift distribution is not uniform. These weights were first introduced by Feldman et al. 1994 and have been applied to all clustering analysis ever since. They are commonly referred to as *FKP weights* and can be written for a galaxy i as

$$w_{i, \text{FKP}} = \frac{1}{1 + \bar{n}(z_i) P_0} \quad (4.1)$$

where P_0 is a rough value of the power spectrum of the tracers at some particular scale of choice, usually BAO scales for large-scale clustering measurements. For eBOSS analyses, we chose a scale of $k \sim 0.1 h \text{Mpc}^{-1}$ for which $P_0 = 10000 h^{-3} \text{Mpc}^3$ for LRGs, $P_0 = 4000 h^{-3} \text{Mpc}^3$ for ELGs and $P_0 = 6000 h^{-3} \text{Mpc}^3$ for QSOs.

In summary, the random catalogue describes the angular and radial selection functions of the galaxy survey. Each random has a weight $w_r = w_{\text{comp}} w_{\text{FKP}}$, where w_{comp} accounts the spectroscopic completeness of targets and w_{FKP} are optimal weights for clustering. Each galaxy has a weight given by $w_g = w_{\text{col}} w_{\text{photo}} w_{\text{fail}} w_{\text{FKP}}$, where w_{col} accounts for fibre collisions, w_{photo} corrects fluctuations caused by photometry and w_{fail} corrects redshift failures. These weights are employed when computing correlation functions or power spectra (see 4.4).

4.2 Reconstruction of linear density field

Bulk motions of galaxies on scales of tens of Mpc impact their two-point statistics, acting as a smoothing of the clustering (Eisenstein et al. 2007b). In particular, these motions smooth/broaden the BAO peak, reducing the precision on our measurement of its position. In the past decades, several methods have been proposed to reconstruct the linear density field and remove these bulk motions from a galaxy survey, therefore sharpening the BAO peak. These methods are referred to as *reconstruction* methods. Reconstruction is now an essential

part of standard BAO measurements from galaxy surveys, since they improve significantly their precision.

The simplest form of reconstruction is based on a theoretical relationship between the galaxy density field on large scales (our observable) and the displacement field connecting the observed field (Eulerian frame) to its past version (Lagrangian frame). Lagrangian perturbation theory (LPT) can predict these displacements (see Bernardeau et al. 2002 for a review). To first order, the relation between displacements and density is linear (Zel'dovich 1970). From the observed density field in our survey, we can compute the corresponding linear displacements and move galaxies back. Since we want to “move” the density field, we also move randoms by the same displacements. This procedure successfully removes most of the broadening of the BAO peak (Nusser and Dekel 1992; Eisenstein et al. 2007a). The first order LPT reconstruction is often called *Zeldovich reconstruction*. Extending to second order LPT does not improve BAO reconstruction significantly (Seo et al. 2010) but it helps on recovering the velocity field (Kitaura et al. 2012). Zeldovich reconstruction has been used in all BAO measurements from SDSS (BOSS and eBOSS). I implemented a python version³ of the Zeldovich reconstruction that uses Fast Fourier Transforms while accounting for the large angle effects (Burden et al. 2014; Burden et al. 2015). Alternatively, displacements can be computed in configuration space using finite difference approximations (Padmanabhan et al. 2012) though these are slower and less precise than iterative FFTs.

Since we do not observe the actual density of matter, a bias relation between galaxy number density and matter density has to be assumed in reconstruction. Usually we assume a linear bias relation, where the bias value is estimated from the clustering of the standard galaxy catalogue. This is a reasonable approximation since reconstruction uses fluctuations on scales larger than $\sim 15 h^{-1}\text{Mpc}$. It is also possible to remove linear redshift-space distortions, by assuming a value for the growth-rate of structures f (see section 1.5.6) used to convert the displacements into velocities. The radial components of the estimated velocities are then removed from each galaxy. Therefore, the clustering of reconstructed catalogues depend on the fiducial values of bias and f , but less on the distance-redshift relation (Carter et al. 2020).

There are other backward reconstruction techniques such as those based on the least action principle (LAP). First proposed by Peebles 1989, it was extended to cosmological applications by Nusser and Branchini 2000; Sarpa et al. 2019. Sarpa et al. 2021 applied it for the BAO measurement of the BOSS DR12 galaxy sample.

Reconstruction is essential to improve the precision on BAO measurements. However, due to imperfections in recovering the true initial conditions, reconstruction has not yet been used in attempts to derive cosmological constraints from the full-shape of the reconstructed two-point statistics. Therefore, analysis of redshift-space distortions are based on the non reconstructed clustering (see section 4.6).

4.3 Mock catalogues

An important limitation in cosmology is that we only have a single realisation of our Universe. Therefore simulations are an essential tool to overcome this limitation.

We can recognise two types of simulations mainly used in cosmology. First, n-body simulations (eventually including gas) solve numerically for the formation of structures and are computationally expensive. They focus on the fully non-linear aspects and require a higher

³https://github.com/julianbautista/eboss_clustering/

resolution setting, preventing them to simulate very large volumes or producing numerous realisations. N-body simulations are still the baseline for tests of perturbation theory models. See Angulo and Hahn 2022 for a review on cosmological n-body simulations. The second type of simulations are the so-called *mock catalogues* or simply *mocks*. They are lower resolution and do not fully describe the non-linear clustering but they are faster to produce both in large volumes and in large number of realisations. Their are key for the understanding of biases in best-fitted quantities and on their uncertainties in clustering analyses. Given the large number of realisations, mocks have been the most reliable method to estimate covariance matrices of clustering measurements. Also, mocks are used to study the impact of observational systematic effects on our measurements. All previous cosmological measurements from galaxy surveys have associated sets of n-body simulations and mock catalogues. I will describe the example I was involved in, the case of EZMOCKS produced for the eBOSS DR16 clustering measurements.

A set of 1000 EZMOCK realisations of the eBOSS survey were produced for three tracers: LRGs, ELGs and QSOs. They are fully described in Zhao et al. 2021. These mocks employ an effective Zeldovich approximation to evolve the density field in longer time steps (Chuang et al. 2015). By construction, these mocks match the large-scale clustering of the target data sample, including 2 and 3-point statistics. They have some extra free parameters to simulate a biased galaxy sample and non-linear motions on small-scales. The eBOSS EZMOCKS also include light-cone effects produced by interpolating boxes at different redshifts.

The last step in the production of EZMOCKS was the addition of systematic effects. I was responsible of implementing all known observational effects from the data onto the mocks. A simple script⁴ gathers the information from the real catalogues and simulates fibre collisions, spurious fluctuations caused by photometry and spectroscopy. These contaminated mocks are then corrected using the same algorithms as for the real data, as described in section 4.1. The clustering before and after adding those effects is extensively compared to real data (Zhao et al. 2021) and the agreement is found to be excellent on the scales of interest for BAO and RSD measurements.

4.4 From catalogues to clustering estimates

Once the catalogue of galaxies and the associated random catalogue are ready, any statistical quantity can be computed, such as n -point statistics. As discussed in section 1.4, these statistics can be computed in either configuration space or Fourier space. This section describes how to obtain 2-point function estimates, in both spaces, which are the statistics I worked with.

4.4.1 Correlation function

The estimate of the correlation function $\xi(\vec{r})$ is based on counting pairs of galaxies in bins of separation \vec{r} and comparing it to same numbers obtained from the random catalogue. For the estimator we use, there is no need to calculate galaxy number densities $n(\vec{x})$ or overdensities $\delta_n(\vec{x})$ in some mesh. To optimise the variance of the estimator while keeping it unbiased, Landy and Szalay 1993 proposed to include the cross-pairs between galaxies and randoms in

⁴https://github.com/julianbautista/eboss_clustering/blob/master/bin/ezmicks_add_systematics.py

the estimator, yielding:

$$\xi(\vec{r}_i) = \frac{DD(\vec{r}_i) - 2DR(\vec{r}_i)}{RR(\vec{r}_i)} + 1, \quad (4.2)$$

where $DD(\vec{r}_i)$ (or $RR(\vec{r}_i)$) are the normalised paircounts between galaxies (or randoms) at bin i corresponding to separation \vec{r}_i , $DR(\vec{r}_i)$ is the normalised cross-paircounts between galaxies and randoms. The normalisation of the paircounts is given by the total number of unique pairs in the catalogue, i.e., $N(N-1)/2$ if the catalogue contains N galaxies or $N_g N_r$ for the cross pairs. When using weights as described in section 4.1, nothing changes except that each pair of galaxies (or randoms) are weighted as $w_{ij} = w_i w_j$, where w_i, w_j are the weights of each element of the pair.

More recently, new ways to account for fibre collisions or incompleteness have been proposed (Bianchi and Percival 2017; Percival and Bianchi 2017) where each pair of galaxies would have a weight $w_{ij} \neq w_i w_j$, i.e., not a product of individual weights. This makes the calculation slightly slower since these weights have to be defined for each pair. These new estimators were first used with the eBOSS tracers (Mohammad et al. 2020) and are currently being employed within DESI.

The estimator in Eq. 4.2 changes slightly when dealing with catalogues resulting from reconstruction algorithms, where both galaxies and randoms are shifted in order to reduce the impact of bulk motions (see section 4.2). The Landy-Szalay estimator is now written as

$$\xi^{\text{recon}}(\vec{r}_i) = \frac{DD(\vec{r}_i) - 2DS(\vec{r}_i) + SS(\vec{r}_i)}{RR(\vec{r}_i)}, \quad (4.3)$$

where S represents the shifted random catalogue and the R is the usual random catalogue as before. See Padmanabhan et al. 2009; Padmanabhan et al. 2012 for a detailed derivation of this estimator.

Most analyses of the two-point correlation function do not use the full $\xi(\vec{r}_i)$, where \vec{r}_i are bins in 2D separations such as bins in $(r_{\parallel}, r_{\perp})$ or (r, μ_r) . Most analysis further compress the estimated correlation function into a new basis with fewer degrees of freedom. Some analyses uses the correlation function *wedges* $\xi_{\mu_{r1}, \mu_r}(r)$, which are averages of ξ over $\mu_{r1} < \mu < \mu_{r2}$, as already discussed in section 3.2.2. Other analyses, such as the one described in this chapter, decompose ξ into a basis of Legendre polynomials $L_\ell(\mu_r)$ such that $\xi(r, \mu_r) = \sum_\ell \xi_\ell(r) L_\ell(\mu_r)$, where $\xi_\ell(r)$ are the correlation function *multipoles*. Usually two or three wedges or multipoles are considered when comparing to clustering models.

One of the most recent codes to compute correlation function accounting for all features described above is PYCORR⁵ which is being currently developed for DESI.

4.4.2 Power spectrum

The power spectrum is the analogous of the correlation function in Fourier space. It requires a Fourier Transform of the galaxy overdensity field, which introduces some effects that have to be correctly taken into account when fitting for clustering models. The power spectrum estimation is based on work by Feldman et al. 1994; Bianchi et al. 2015; Hand et al. 2017 and it is shortly described here.

The first step is to assign galaxies and randoms into a three-dimensional mesh, defining an

⁵<https://github.com/cosmodesi/pycorr>

“overdensity” function $F(\vec{x}_i)$ as

$$F(\vec{x}_i) = n(\vec{x}_i) - \alpha n_{\text{rand}}(\vec{x}_i), \quad (4.4)$$

where n_{rand} are the *weighted* number density of galaxies and randoms in voxel centered at \vec{x}_i , α is the ratio of the total weighted number of galaxies to the total weighted number of randoms. The field $F(\vec{x})$ is also known as the FKP field.

The field $F(\vec{x})$ can be Fourier transformed numerically using Fast Fourier Transform (FFT) algorithms. The power spectrum would be defined as the absolute value squared of this field. Power spectrum *multipoles* are commonly used, as it was the case in configuration space. They can be expressed as:

$$P_\ell(k) = \frac{1}{I_{22}} \left[(2\ell + 1) \left\langle F_0(\vec{k}) F_\ell(-\vec{k}) \right\rangle_{V_k} - (1 + \alpha) I_{12} \delta_{0\ell} \right] \quad (4.5)$$

where the brackets $\langle \cdot \rangle$ indicate average over spherical shells in k-space with radius $|\vec{k}|$, $F_\ell(\vec{k})$ are the overdensity multipoles in k-space given by

$$F_\ell(\vec{k}) = \int d^3x F(\vec{x}) e^{i\vec{k}\cdot\vec{r}} L_\ell(\hat{k}\cdot\hat{x}), \quad (4.6)$$

and the normalisation factors I_{ab} are defined as

$$I_{ab} \equiv \int d^3x [n(\vec{x})]^a [w_{\text{FKP}}(\vec{x})]^b. \quad (4.7)$$

The first term in Eq. 4.5 is the usual term defining the power spectrum, i.e., the square of the absolute value of Fourier fluctuations, while the second term is the so called *shot-noise* and only affects the power spectrum monopole.

Since we use Fourier transforms which rely on periodic boundary conditions, the mesh used to compute $F(\vec{x})$ has to be large enough to encompass the survey volume with some padding around it. Also, because of the assumption of periodicity, the estimated power spectra will be convolved by the window/selection function of the survey. This window has to be correctly computed using the randoms and it is a key ingredient to be accounted for when modelling the clustering.

Since we use numerical FFTs, the resolution of the mesh defines the scales being probed. If the mesh has a total comoving size of L on a side and has N voxels on a side, the fundamental mode is given by $k_F = \pi/L$ while the highest (or Nyquist) frequency is $k_{\text{Ny}} = \pi N/L$. Typically meshes with 512 or 1024 voxels on a side are sufficient to cover the scales of interest for BAO or RSD analyses in current survey volumes, $0.01 < k < 0.30 (\text{Mpc}/h)^{-1}$.

The estimated power spectra are also affected by the discreteness nature of the mesh on scales close the Nyquist frequency. The assignment of galaxies to the mesh can be optimised to reduce those effects. Also, interlacing techniques can also further reduce mesh effects. These effects and proposed solutions are described in detail in Sefusatti et al. 2016 and are implemented in the package NBODYKIT⁶ or PYPower⁷.

⁶<https://nbodykit.readthedocs.io/>

⁷<https://github.com/cosmodesi/pypower>

4.4.3 The clustering of eBOSS DR16 LRGs

Figure 4.5 and 4.6 show the power spectrum $P_\ell(k)$ and correlation function $\xi_\ell(r)$ multipoles for $\ell = 0, 2, 4$ for the combined sample of BOSS DR12 and eBOSS DR16 luminous red galaxies and their respective EZMOCK catalogues. Figure 4.5 shows results from standard catalogues while Figure 4.6 displays results from reconstructed catalogues.

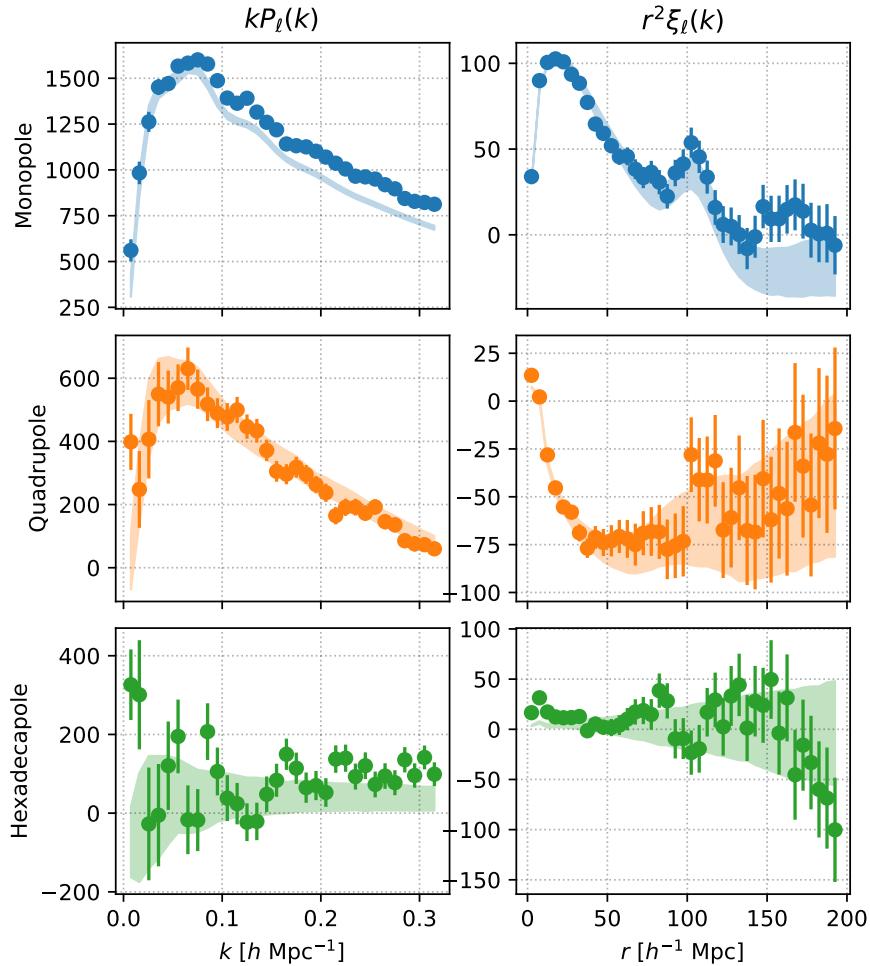


Figure 4.5: Pre-reconstruction multipoles of the two-point statistics for the combination of BOSS DR12 and eBOSS DR16 luminous red galaxies. Left panels contain the power spectrum multipoles (scaled by k) and right panels the correlation function multipoles (scaled by r^2). Points with errorbars are real data while shaded regions show the mean and standard deviation of 1000 realisations of EZMOCKS.

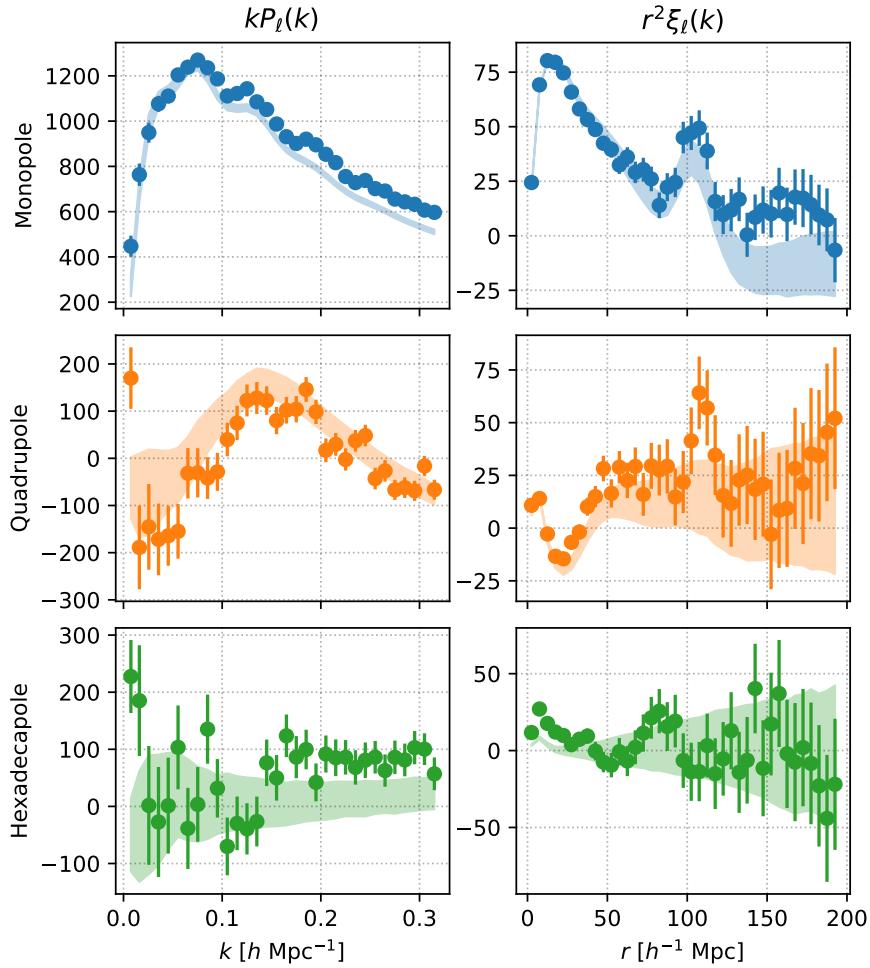


Figure 4.6: Same as Figure 4.5 but for reconstructed catalogues. We used the Zeldovich reconstruction method with removal of redshift-space distortions. The BAO features are more prominent and the quadrupole is significantly reduced compared to results with the standard catalogues.

The uncertainties on our measurements of P_ℓ and ξ_ℓ were derived by measuring the same statistics on the full sample of 1000 realisations of EZMOCKS. By reproducing with high fidelity the observational features and systematic effects, the covariance matrix obtained with mocks should account for the extra scatter induced by them. However on small scales the clustering of the mocks is not accurate compared to real data or to more realistic n-body simulations. Figure 4.7 shows the resulting covariance matrices normalised by the diagonal elements (correlation coefficients). We see how power spectra multipoles are generally less correlated than correlation function multipoles. Reconstruction also helps reducing these correlations. Also shown are the cross-covariance between Fourier space and configuration space measurements, which will be discussed in section 4.7. As expected, these two spaces are highly correlated since they originate from the same sample of galaxies and randoms.

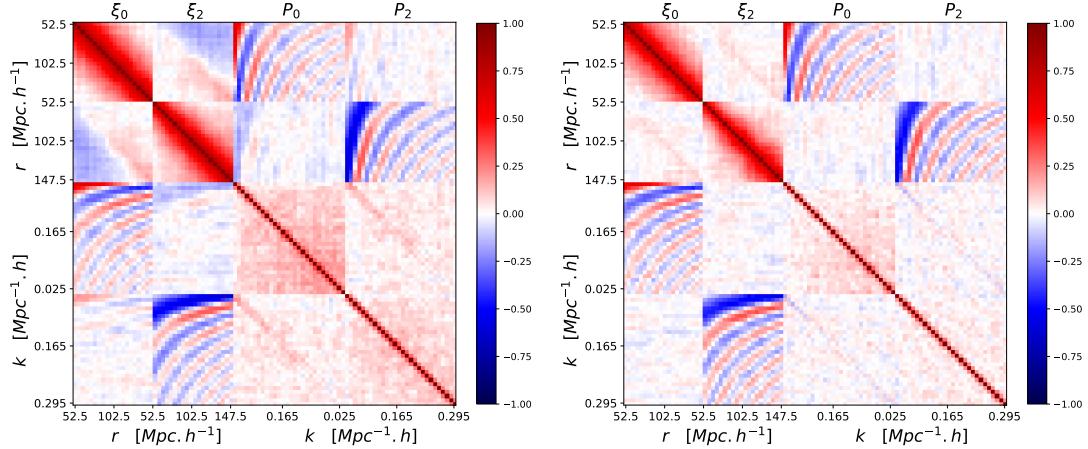


Figure 4.7: Correlation coefficients from covariance matrices obtained from 1000 measurements of power spectrum and correlation function multipoles, P_ℓ and ξ_ℓ respectively, from EZMOCKS reproducing the combined BOSS DR12 and eBOSS DR16 LRG sample. Left panel shows results from the standard catalogues while the right panel shows the results from Zeldovich reconstructed catalogues.

4.5 Baryon acoustic oscillations

The feature from baryon acoustic oscillations (BAO) is clearly visible as oscillations in the power spectrum multipoles and as a sharp peak in the correlation function multipoles (Figures 4.5 and 4.6). In this section I describe how we measure the BAO scale, which is used as a standard ruler in the study of the expansion of the Universe. The work is detailed further in Bautista et al. 2021; Gil-Marín et al. 2020.

The model used to fit BAO is based on the linear matter power spectrum with a simplified treatment of bias, redshift-space distortions and non-linearities. It is basically the same model as used in the BAO measurements from Ly α forests described in section 3.2.4 but without the treatment of metal absorption, high-column density systems and distortions which do not apply here. A linear bias and linear redshift-space distortions is assumed. The non-linearities that broaden the BAO peak are modelled by two Gaussian smoothing terms, one for radial separations and one for transverse separations. We also decompose the peak part from the smooth part of the model, scaling only the peak part by the dilation factors α_{\parallel} and α_{\perp} , defined in Eqs. 3.11 and 3.12. The fit is performed in the space of multipoles instead of fitting for the full two-point statistics. Smooth power-laws of separation are added with free amplitudes in order to marginalise any potential information coming from the smooth part of the correlation function. To account for the removal of redshift-space distortions after reconstruction, an empirical damping term is added to the model, such that the linear RSD term $(1 + \beta\mu^2)$ becomes $(1 + S(k)\beta\mu^2)$ where $S(k) \equiv 1 - e^{-k^2\Sigma_{\text{rec}}^2/2}$ (Seo et al. 2016).

The methodology was tested with the sample of 1000 EZMOCKS and also an additional set of 84 NSERIES n-body simulations (section 4.3). In principle, BAO is a feature on large enough scales such that testing the method on approximate mocks is sufficient, although we also looked at results on more realistic n-body simulations. In order to understand potential biases

in best-fit BAO parameters, we fit for BAO on the average clustering of all available realisations for EZMOCKS and NSERIES mocks, effectively increasing the volume of the survey by a factor of 1000 and 84 respectively. Thanks to the large number of EZMOCKS, we were able to study the performance of our fitter statistically, by looking at the ensemble of 1000 individual BAO fits.

Figure 4.8 displays the results of fits to the mean clustering of mocks for catalogues pre- and post-reconstruction, also as a function of the fiducial choice for Ω_m^{fid} used in converting distances to redshifts. First we notice how biases are much smaller for reconstructed catalogues than standard ones, likely thanks to a better modelling of non-linearities affecting the BAO peak which are reduced in the reconstructed case. Second, uncertainties of reconstructed results are reduced relative to pre-reconstruction catalogues, showing the good performance of the method to remove bulk motions. Third, there is a very weak dependency of our results with the choice of Ω_m^{fid} , which is reassuring.

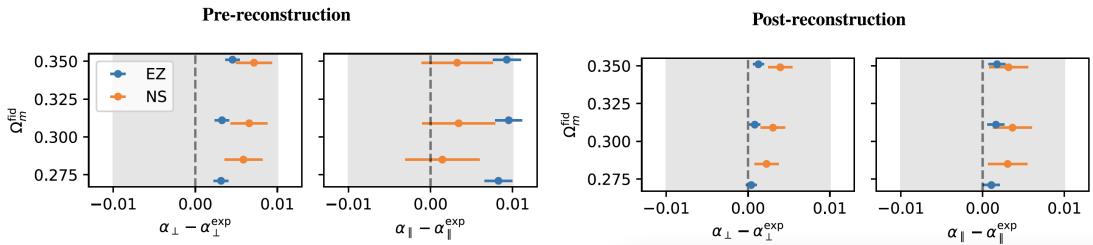


Figure 4.8: Impact of choice of fiducial cosmology in the recovered values of α_{\parallel} and α_{\perp} from BAO fits to the stacks of 1000 multipoles the EZMOCKS (blue) and 84 NSERIES mocks (orange), for pre- (left panels) and post-reconstruction (right panels). Associated error bars correspond to the uncertainty estimated from the mean of the mocks. The grey shaded areas correspond to 1 per cent shifts. For comparison, the uncertainty on real data is 1.9 per cent for α_{\perp} and 2.6 per cent for α_{\parallel} in the post-reconstruction case.

Another important test of our methodology consists in checking uncertainty estimation. With the sample of 1000 BAO measurements from mocks, we can study them statistically. Our uncertainties are derived by scanning the likelihood profile and finding the intervals where the parameters α_{\perp} and α_{\parallel} increase the χ^2 by unity⁸ relative to its minimum. First, we compare the standard deviation of all best-fit values to the average uncertainty found in all 1000 mocks. If uncertainties are correctly estimated then these two numbers match. Second, we could look at the pull distribution of each parameter x_i , defined as $Z_i \equiv (x_i - \bar{x}_i)/\sigma_{x_i}$. If the uncertainties are well behaved and Gaussian, then we should obtain $\bar{Z}_i = 0$ and $\text{Var}(Z_i) = 1$. Table 6 in Bautista et al. 2021 shows these statistics for BAO measurements. The results validate our methodology and we applied it to data.

Figure 4.9 shows the best-fit correlation function model and the BAO constraints for the eBOSS DR16 LRG sample. Converting the constraints in $(\alpha_{\perp}, \alpha_{\parallel})$ into ratios of distances, we obtained:

$$\mathbf{D}_{\text{BAO}, \xi_\ell} = \begin{pmatrix} D_M/r_d \\ D_H/r_d \end{pmatrix} = \begin{pmatrix} 17.86 \pm 0.33 \\ 19.34 \pm 0.54 \end{pmatrix}, \quad (4.8)$$

which correspond to a BAO measurement at 1.9 per cent in the transverse direction and 2.8 per cent in the radial direction. Radial and transverse BAO measurements are correlated by

⁸In the case marginalised errors in one parameter. For two parameters, the 68 and 95% confidence intervals are defined by $\Delta\chi^2 = 2.3, 5.9$ respectively.

-33% for this dataset. These are the best constraints obtained from galaxies at redshifts above 0.6.

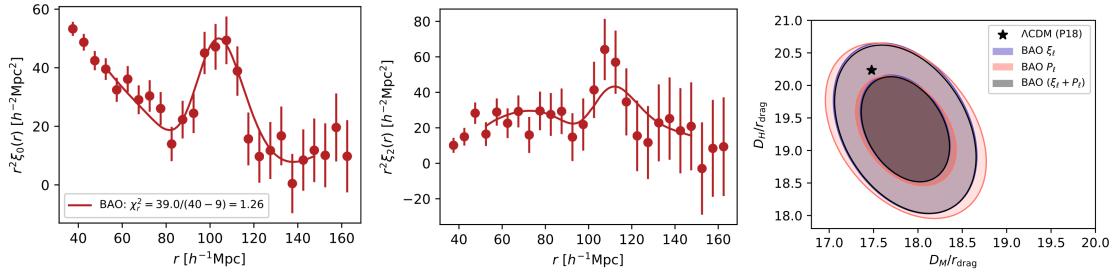


Figure 4.9: Two left panels show the best-fit model for the correlation function monopole and quadrupole for the reconstructed eBOSS LRG sample. The right panel shows the two-dimensional constraints on the distance to sound horizon ratios, for the configuration and Fourier space BAO analyses. The consensus BAO result is shown in grey. The star points to the prediction of a flat Λ CDM model with Planck 2018 best-fit parameters.

4.6 Redshift-space distortions

The analysis of redshift-space distortions (RSD) aims to measure the growth-rate of structures f , as introduced in section 1.5.6, by quantifying the anisotropies in the clustering of galaxies. As we seen before, assuming an incorrect fiducial cosmology when converting redshifts into distances also introduces anisotropies (the AP effect). Therefore in practice, an RSD analysis yields simultaneously the combination $f(z)\sigma_8(z)$ and the distance ratios $D_H(z)/r_{\text{drag}}, D_M(z)/r_{\text{drag}}$.

The main differences with respect to the BAO analysis (see previous section) is that the information about anisotropies come from the full shape of the two-point statistics, not only the peak. This complicates the RSD analysis in that we need to have accurate models for the full shape of the clustering down to some scale, and we cannot add empirical terms to be marginalised over as we did in the BAO analysis. Due to non-linearities of matter clustering and galaxy bias, it is increasingly hard to model scales below those well described by linear theory. For these mildly non-linear scales, we need more advanced perturbation theory models. A huge variety of theoretical models have been developed in the past years (see Bernardeau et al. 2002 for a review).

In the analysis of eBOSS data, we used two types of models: the first one is based on regularised perturbation theory with second order bias and RSD correction terms (Taruya et al. 2010; Taruya et al. 2012). We commonly refer to it as the *TNS model*. The second one is based on convolution Lagrangian perturbation theory with Gaussian streaming (Carlson et al. 2013; L. Wang et al. 2014). We refer to it as the *CLPT-GS model*. I implemented these models in python language in the package BAOPY⁹.

The TNS model is expressed as:

$$P_g(k, \mu_k) = D_{\text{FoG}}(k, \mu_k) [P_{g,\delta\delta}(k) + 2f\mu_k^2 P_{g,\delta\theta}(k) + f^2\mu_k^4 P_{\theta\theta}(k) + b_1^3 A(k, \mu_k, \beta) + b_1^4 B(k, \mu_k, \beta)], \quad (4.9)$$

⁹<https://baopy.readthedocs.io/>

where b_1 is the first order linear galaxy bias, $\beta \equiv f/b_1$, $(P_{g,\delta\delta}, P_{g,\delta\theta}, P_{\theta\theta})$ are the non-linear isotropic galaxy auto power spectrum, galaxy-velocity divergence cross power spectrum and the velocity divergence auto power spectrum, respectively, which are themselves functions of first and second order bias parameters. These functions are provided by the RegPT scheme, implemented in PYREGPT¹⁰. The functions A and B are the RSD correction terms. The function D_{FoG} can be either a Gaussian or a Lorentzian of $k_{\parallel} = k\mu_k$ with characteristic scale σ_{FoG} corresponding to the velocity dispersion on small scales. This model is computed in Fourier space, but it can be Fourier transformed to obtain a model in configuration space.

The CLPT-GS model is uniquely defined in configuration space, and writes as:

$$1 + \xi_{g,s}(r_{\perp}, r_{\parallel}) = \int \frac{1}{\sqrt{2\pi [\sigma_{12}^2(r) + \sigma_{\text{FoG}}^2]}} [1 + \xi_g(r)] \exp \left\{ -\frac{[r_{\parallel} - y - \mu v_{12}(r)]^2}{2[\sigma_{12}^2(r) + \sigma_{\text{FoG}}^2]} \right\} dy \quad (4.10)$$

where $\xi_g(r)$, $v_{12}(r)$, and $\sigma_{12}(r)$ are obtained from CLPT and contains two bias terms. The dependency with f is via a scaling of the functions v_{12}, σ_{12} that are proportional to velocity. The integral corresponds to the convolution by velocities in the radial direction, characteristic of the Gaussian streaming formalism.

Both the TNS and CLPT-GS models had to be tested against n-body simulations. We were particularly interested in the scales of validity of these models for the case of the eBOSS LRG sample. Given the number of available realisations, we used the n-body suite of 84 realisations of NSERIES simulations produced for the BOSS CMASS sample. We fitted the average correlation function of these simulations and measured the parameters $f\sigma_8, \alpha_{\perp}, \alpha_{\parallel}$. Figure 4.10 presents the results compared to expected values for both TNS and CLPT-GS models. The shaded green area shows the fiducial choice to be used in the data.

¹⁰<https://github.com/adematti/pyregpt>

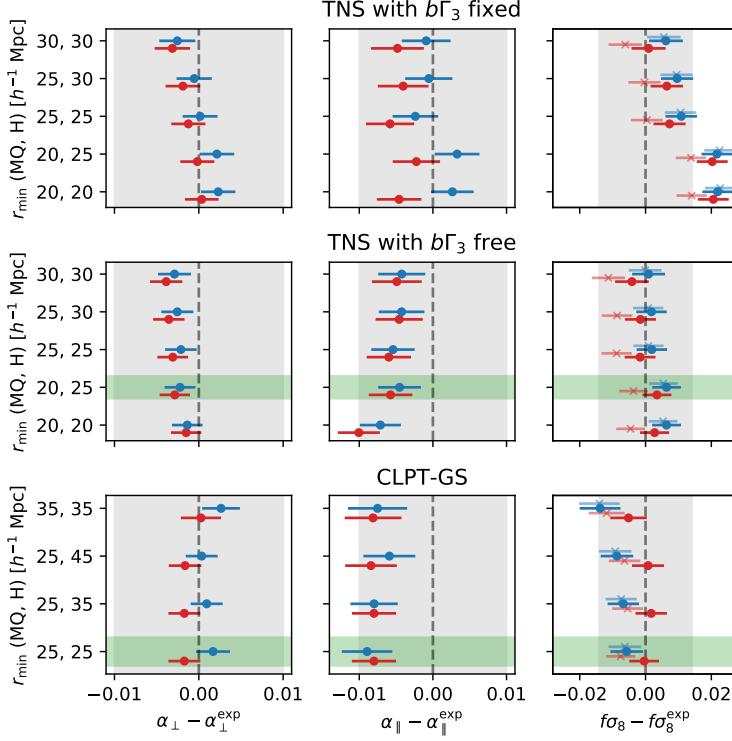


Figure 4.10: Biases in the measurement of $f\sigma_8, \alpha_{\parallel}, \alpha_{\perp}$ obtained from full-shape fits to the average of 84 multipoles from the NSERIES mocks as a function of the separation range used. The y-axis displays the value of the minimal separation r_{\min} used in fits of the monopole, quadrupole (MQ) and hexadecapole (H). Top and mid rows display results for the TNS model when fixing or letting free the bias parameter $b\Gamma_3$ respectively. Bottom row presents results for the CLPT-GS model. The blue circles correspond to the analysis using $\Omega_m^{\text{fid}} = 0.286$ (the true value of simulations) while the red circles correspond to $\Omega_m^{\text{fid}} = 0.31$. The crosses in the $f\sigma_8$ panels correspond to their values before the scaling of σ_8 discussed in Figure 4.11. The gray shaded areas correspond to 1 per cent errors in $\alpha_{\perp}, \alpha_{\parallel}$ and to 3 per cent in $f\sigma_8$. The green shaded area shows our choice for baseline analysis for TNS and CLPT-GS models.

Using the same simulations, we tested the dependency of our constraints to the choice of fiducial cosmology. Note that the fiducial cosmology enters twice in the modelling, first when converting redshifts to distances; second, when computing the linear matter power spectrum template $P_m^{\text{lin}}(k)$, used as main ingredient for both models. Figure 4.11 presents the results for three choices of fiducial cosmology. While for $\alpha_{\perp}, \alpha_{\parallel}$ the biases or trends are negligible, $f\sigma_8$ shows a strong dependency with Ω_m^{fid} , shown as crosses in the figure. This dependency is mainly caused by the assumed σ_8 value, that corresponds to the one derived from the fiducial $P_m^{\text{lin}}(k)$ via Eq. 1.18. The resulting f of the fit is scaled by this σ_8 . We can reduce this dependency by recomputing σ_8 using $R = 8\alpha h^{-1}\text{Mpc}$, where $\alpha \equiv \alpha_{\perp}^{(2/3)}\alpha_{\parallel}^{(1/3)}$ is the isotropic dilation factor obtained from the best-fit $\alpha_{\perp}, \alpha_{\parallel}$. In effect, this keeps the scale at which σ_8 is fitted fixed relative to the data in units of $h^{-1}\text{Mpc}$, which only depends on Ω_m^{fid} . This is an alternative approach to the recently proposed σ_{12} parametrisation (Sánchez

2020), where the radius of the top-hat function is set to $R = 12$ Mpc (in units of Mpc) instead of $R = 8 h^{-1}$ Mpc (in units of h^{-1} Mpc). The circles in the right panel of Figure 4.11 shows the results with the recomputed σ_8 for each case, showing that this procedure indeed reduces the trend of $f\sigma_8$ with Ω_m^{fid} . This topic is also discussed in Appendix A of Alam et al. 2021.

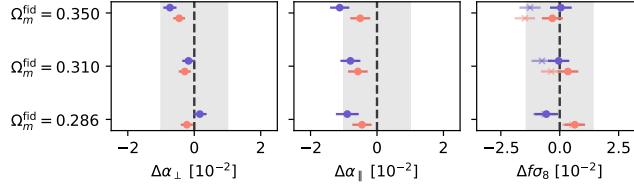


Figure 4.11: Biases in best-fit parameters for both CLPT-GS (blue) and TNS (red) models from fits to the average multipoles of 84 NSERIES mocks. Shaded grey areas show the equivalent of 1 per cent error for α_{\perp} , α_{\parallel} and 3 per cent for $f\sigma_8$. In the right panel, crosses indicate $f\sigma_8$ values when σ_8 is not recomputed as described in the text. The true cosmology of the mocks is $\Omega_m = 0.286$. For reference, the errors on our data sample are ~ 2 , 3 and 10 per cent for α_{\perp} , α_{\parallel} , $f\sigma_8$ respectively.

After validating the model, we tested the impact of different halo occupation distribution (HOD) models and observational effects, both introduced in simulations. We found that none of these modifications introduces significant biases our constraints of $(f\sigma_8, \alpha_{\perp}, \alpha_{\parallel})$, though the shifts we observe were added as systematic uncertainties in quadrature to the statistical ones. Table 10 of Bautista et al. 2021 summarises the systematic errors which are assumed to be diagonal (no covariance between the three parameters) for simplicity. For our RSD analysis, systematic errors are about half of the statistical ones for $f\sigma_8$ and slightly less for $\alpha_{\perp}, \alpha_{\parallel}$. In quadrature, final uncertainties increase by up to 20% due to systematic errors.

We also attempted to analyse our biases and uncertainties statistically with the set of 1000 EZMOCKS. However, an important caveat is that these mocks only solve approximatively the clustering, and therefore are not the better suited to test RSD fitting. We proceeded to fit the 1000 realisations and we compared the average uncertainty to the standard deviation of best-fit parameters among all mocks. These numbers agree sufficiently well, particularly for $f\sigma_8$. The pull distributions of each parameters do not point to any under/over estimation of uncertainties.

Figure 4.12 presents the best-fit TNS and CLPT-GS models to the correlation function multipoles of the eBOSS DR16 LRG sample. Both models are basically indistinguishable even though their are derived differently. The constraints are also similar between the two models, so we combined them into a single results by averaging their best-fit parameters and their covariances. Figure 4.13 show the final constraints on $f\sigma_8$ and the distance ratios, compared to the same analysis in Fourier space. We also show the consensus result computed using the method by Sánchez et al. 2017, which assumes Gaussianity of each posterior distribution (see the next section for a further discussion). The final RSD constraints from the correlation function fits are:

$$\mathbf{D}_{\text{RSD}, \xi} = \begin{pmatrix} D_M/r_d \\ D_H/r_d \\ f\sigma_8 \end{pmatrix} = \begin{pmatrix} 17.42 \pm 0.40 \\ 20.46 \pm 0.70 \\ 0.460 \pm 0.050 \end{pmatrix} \quad (4.11)$$

which correspond to a 11% measurement of $f\sigma_8$ at $z_{\text{eff}} = 0.7$. This uncertainty is slightly

reduced when combining with the Fourier space results and with BAO results from the previous section. The combination between Fourier and configuration space results will be discussed in the following section.

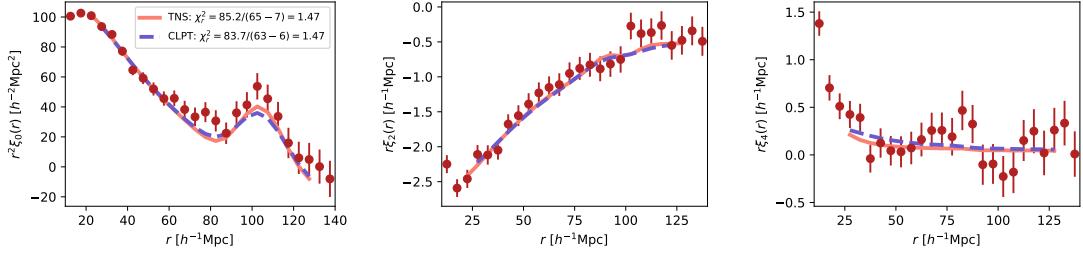


Figure 4.12: Best-fits models of redshift-space distortions (RSD) to the eBOSS DR16 LRG multipoles. Left, mid and right panel display mono, quad and hexadecapole, respectively. Note that only the monopole is scaled by r^2 while the others are scaled by r . The CLPT-GS model is shown by the blue dashed line while the TNS model is shown by the red solid line.

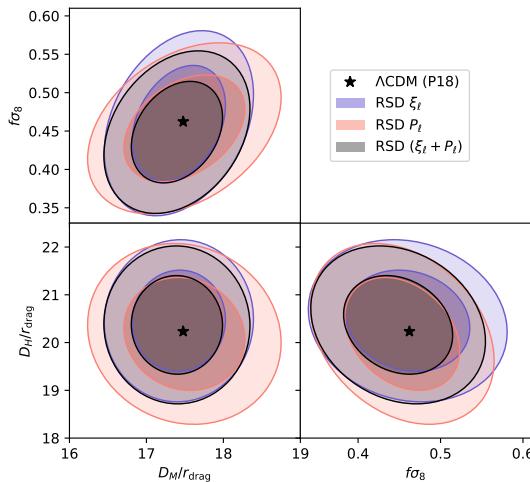


Figure 4.13: Constraints on D_M/r_d , D_H/r_d and $f\sigma_8$ at $z_{\text{eff}} = 0.7$ from the full-shape RSD analysis of the completed eBOSS LRG sample. Contours show 68 and 95 per cent confidence regions for the analyses in configuration space (blue), Fourier space (red) and the combined (grey). The expected values in a flat Λ CDM model with best-fit parameters from Planck 2018 results is indicated as a black star.

4.7 Joint clustering analysis in Fourier and Configuration space

From the same catalogue of angular positions and redshifts, we can perform two different clustering analyses: one in Fourier space (FS) and the other in Configuration space (CS). How to define a consensus result from both analysis which might yield slightly different results? In this section I expose the work lead by my PhD students Tyann Dumerchat and Vincenzo Aronica on how to combine BAO and RSD constraints from both spaces into a single consensus

result.

Sánchez et al. 2017 proposed a method to combine two correlated Gaussian posteriors into a single one. I refer to it as the Gaussian approximation (GA) method. The GA method was used in BOSS (Alam et al. 2017) and it was slightly improved in eBOSS (Bautista et al. 2021). The main idea is that we want to combine a set of vectors containing the best-fit parameters of different analysis (and their respective error matrix) into a single vector of parameters and a single error matrix. For example in BOSS and eBOSS, we aimed at combining the vectors $\vec{p} \equiv f\sigma_8, \alpha_{\perp}, \alpha_{\parallel}$ obtained from FS and CS analyses. The GA method uses an estimate the correlations between \vec{p}_{FS} and \vec{p}_{CS} , and built a large covariance matrix used to combine the vectors and error matrices. Currently the best method to obtain these correlations between FS and CS results is using mock catalogues. This provides with an estimate of a sort of “ensemble averaged” correlations (though limited by the number of realisations). However, the real data is a single realisation of the ensemble, and the correlations between FS and CS are also susceptible to fluctuate, just as the individual parameters and their uncertainties. In Bautista et al. 2021 we developed a method to account for this fact, where we adjust the correlations for each realisation based on the vectors \vec{p} and their error matrices of that particular realisation.

In **dumerchatBaryonAcousticOscillations2022**, we developed an analysis where we fit jointly the correlation function and power spectrum while accounting for their covariance. I refer to it as the joint space (JS) method. The covariance is built at the multipole level, using a set of mock catalogues. Figure 4.7 shows the correlation coefficients of CS and FS multipoles, including the cross-correlation terms. The wiggles observed in the cross-correlation terms can be simply explained by the fact that CS and FS are related via Hankel transforms (see Appendix A of **dumerchatBaryonAcousticOscillations2022** for an explicit derivation). We tested and validated this framework on the BAO analysis of the eBOSS LRG sample (section 4.5) where we fit for a single α_{\perp} and α_{\parallel} , though amplitudes of power-laws are independent for each space. The main advantage of the JS method is that we do not assume Gaussian posteriors at the parameter level, only at the multipole level when fitting our model (which is quite common assumption). The posteriors on the resulting $(\alpha_{\perp}, \alpha_{\parallel})$ can also be non-Gaussian. Figure 4.14 displays few examples from BAO fits to mock realisations where the CS and FS posteriors are Gaussian (left) or non-Gaussian (right panels), and the resulting combination from GA and JS methods. For the non-Gaussian cases, the GA method fails to correctly describe the posterior, underestimating uncertainties.

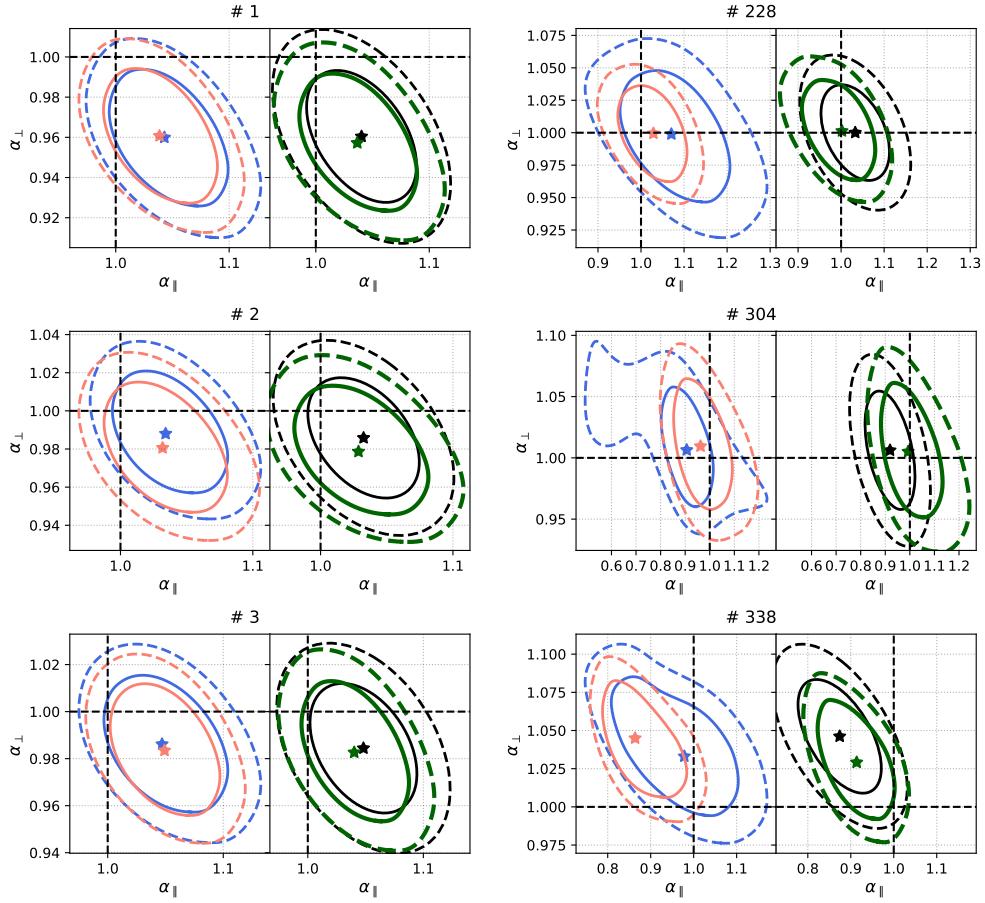


Figure 4.14: Posteriors on the BAO fits of some EZMOCKS realisations. We display cases where contours found in configuration or Fourier space are Gaussian (left) and non Gaussian(right). Red contours are for Fourier space (FS), blue for configuration space (CS), black for the Gaussian approximation (GA) and green for the joint space (JS) fits. The expected value is the intersection of the dotted black lines, and the best fit values are described by a star for the JS and GA methods. We can see how the JS method yields better combined results that are not necessarily Gaussian. Extracted from [dumerchatBaryonAcousticOscillations2022](#).

Vincenzo Aronica is currently working on the application of joint space analysis for the case of redshift-space distortions (RSD). He implemented the TNS model (section 4.4.3) in Fourier and configuration space in the BAOPY package. The tests on the suite of 84 realisations of NSERIES n-body simulations are promising, though there are some subtleties to solve related to the relative normalisation (or linear bias) amplitudes between FS and CS. Currently we use the same values of biases for both space, in addition to $\alpha_{\perp}, \alpha_{\parallel}, f\sigma_8$. Vincenzo will recompute power spectra for a given set of mock catalogues using a consistent normalisation and window function calculations provided by PYPOWER.

The work on BAO and RSD joint fits in Fourier and configuration space will be discussed and hopefully applied to the analysis of DESI data.

4.8 Cross-correlation with radio surveys

The study of large-scale structures will soon be improved thanks to observations of the distribution of neutral hydrogen (HI) via the emission of photons from the 21cm hyperfine transition of the electron. Future 21cm surveys such as the Square Kilometer Array (SKA) will be able to observe the Universe with intensity mapping up to redshifts of 3, opening an alternate window to that provided by spectroscopic surveys. The term *intensity mapping* refers to the fact that we measure the integrated flux from unresolved sources at a given location in space.

During my postdoctoral stay at the University of Portsmouth, I contributed to the study of a currently available 21cm survey produced by the Green Bank Telescope (GBT). I was particularly interested in the study of cross-correlations between the 21cm observed by GBT and the eBOSS galaxies observed in the same volume. With the cross-correlation function we could infer the product of the average neutral hydrogen fraction Ω_{HI} at redshift $z_{\text{eff}} = 0.8$ with the bias of HI b_{HI} . These parameters are important ingredients for models of clustering of the 21cm emission. I overview the main findings of this work in this section but further details are described in Wolz et al. 2022.

The use of 21cm intensity mapping to measure the large-scale structures has been extensively studied theoretically (Bharadwaj et al. 2001; Battye et al. 2004; McQuinn et al. 2006; Chang et al. 2008) and with simulations (Mesinger et al. 2011; Villaescusa-Navarro et al. 2018 and references therein). On large scales, 21cm flux fluctuations trace those of dark matter density, while on smaller scales the physics of the gas and its connection to halos complexifies the modelling. Including redshift-space distortions, a linear model for the 21cm power spectrum is essentially the same as the linear model for galaxy clustering. In Villaescusa-Navarro et al. 2018, an extensive study using the IllustrisTNG n-body hydro simulation quantified several ingredients commonly used in more complex models of the 21cm power spectra, such as the HI content of halos, the HI probability density function, the HI bias with respect to matter and others.

The data sets we used in our study were the eBOSS luminous red galaxy (LRG) and emission line galaxy (ELG) samples in one hand, and the Green Bank Telescope (GBT) 21cm maps in the other hand. The eBOSS LRG sample is already described in section 4.1 while the ELG sample is described in Anand Raichoor et al. 2020. The GBT maps are observations of a 100 deg^2 patch of sky at frequencies from 700 to 900 MHz divided into 256 channels, corresponding to HI emission between $0.6 < z < 1.0$, perfectly overlapping with our galaxy samples in area and redshift. The angular resolution of HI observations varies from 0.31 deg at 700 MHz to 0.25 deg at 900 MHz. In order to reduce the impact of varying PSF, all channels are convolved by an artificial beam of 0.44 deg . The maps we used are an extension to those produced by Masui et al. 2013, including extra scans which increase the area and depth. The raw data is contaminated by radio frequency interference and some channels are therefore masked. Foregrounds are removed with a technique named *FastICA*, which stands for fast independent component analysis. I was provided several maps with different number of foreground components removed. Once the final area of GBT maps was fixed, I cut the eBOSS LRG and ELG samples to the same area. Figure 4.15 shows a redshift slice (or one channel) of the foreground-cleaned GBT HI map and the LRG and ELG angular distribution in the same patch of sky.

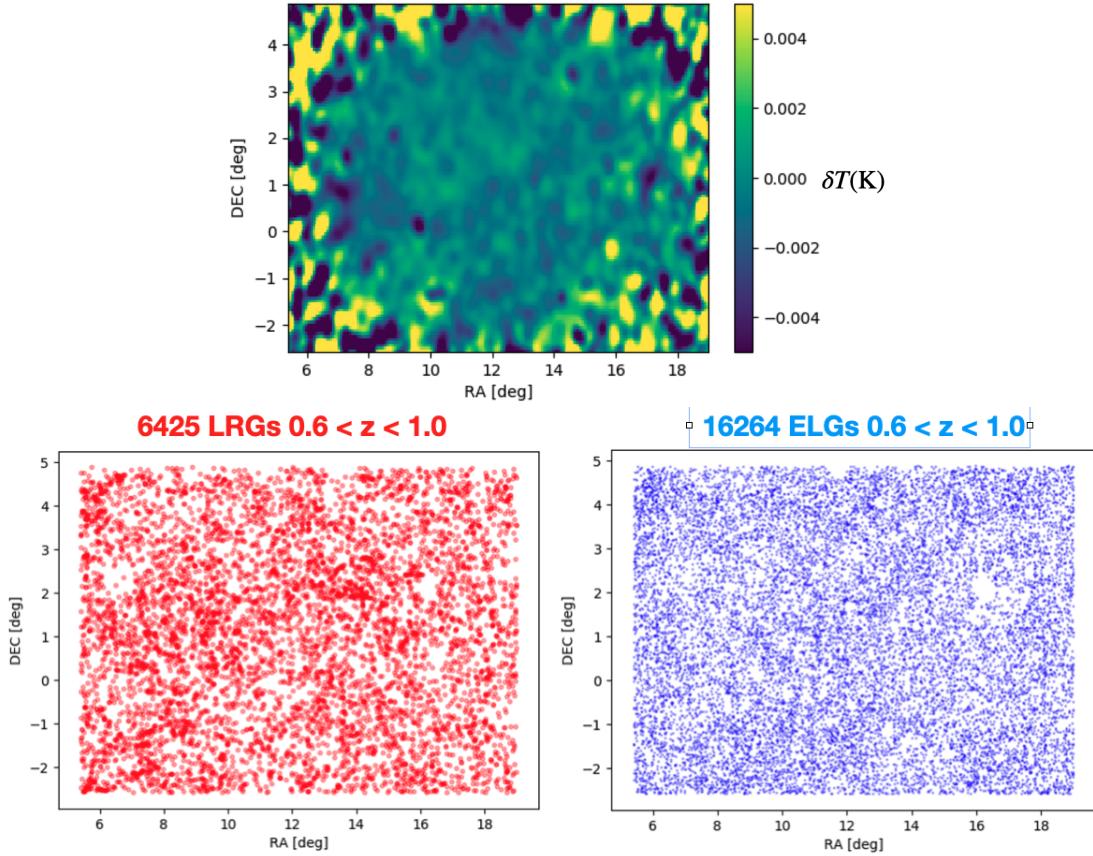


Figure 4.15: Top panel: a redshift slice of the 21cm HI map from the Green Bank Telescope after foreground removal. Bottom panels: the angular distribution of all eBOSS LRGs and ELGs falling in the same sky area.

The cross-correlations between eBOSS galaxies and 21cm temperature maps was computed in Fourier space by Wolz et al. 2022 and in configuration space by me. The cross-correlation estimator I used is the same used for the quasar- $\text{Ly}\alpha$ forest cross-correlation described in section 3.2.2, which assumes that galaxies (in this case) are shot-noise dominated and that voxels of the 21cm map are independent. These assumptions are not quite valid for our sample but I left further developments for future work. For instance, one could perform the study of cross-correlations with a maximum likelihood method (see e.g. section ??) or in Fourier space but correctly accounting for the convolution by the window function of these surveys, which was not explicitly done in Wolz et al. 2022.

Figure 4.16 displays the monopoles and quadrupoles of the cross-correlation function between 21cm temperature fluctuations and eBOSS galaxies. Both ELG and LRG show non-zero correlations with the temperature maps that extend to cosmological scales. In particular, we clearly see a non-zero quadrupole, which was measured for the first time in this analysis. The covariance matrix of the multipoles was computed with the jackknife technique, using 96 independent subsets of the data. Due to the estimator I used, neighbouring separation bins in the monopole are strongly correlated.

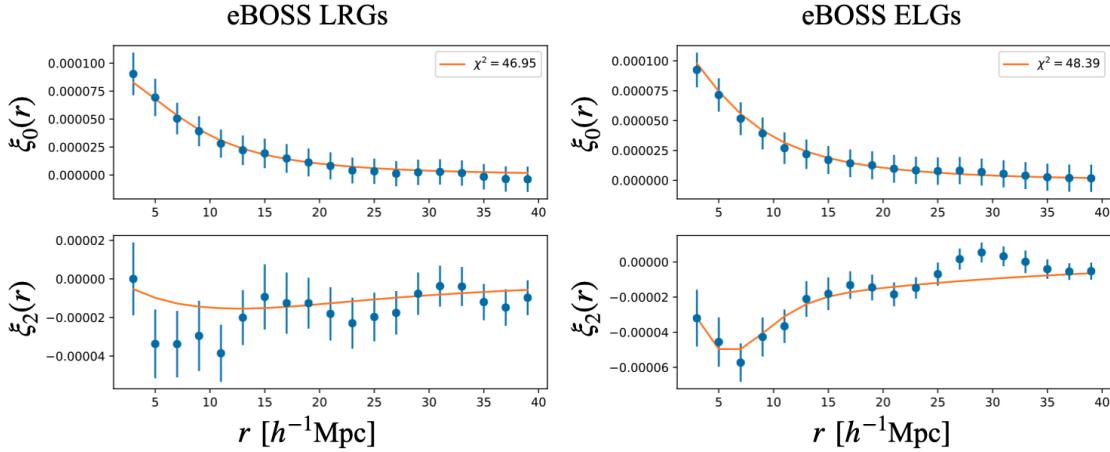


Figure 4.16: Cross-correlation function multipoles between 21cm temperature fluctuation maps from the Green Bank Telescope and the eBOSS samples of Luminous Red Galaxies (left) and Emission Line Galaxies (right). Uncertainties are computed with jackknife resampling and solid lines correspond to the best-fit models.

I attempted to fit the temperature-galaxy cross-correlation with the following model:

$$P(k, \mu) = \frac{b_{\text{gal}} b_{\text{HI}} \bar{T}_{\text{HI}}}{(1 + k^2 \mu^2 \Sigma_{\text{FoG}}^2)^2} (1 + \beta_{\text{gal}} \mu^2) (1 + \beta_{\text{HI}} \mu^2) P_{\text{lin}}(k) e^{-k^2 R_{\text{beam}}^2 (1 - \mu^2)/2} \quad (4.12)$$

which includes a term for Fingers-of-God with dispersion Σ_{FoG} , and a damping in the transverse (angular) direction to account for the effect of the beam of the GBT instrument, with characteristic scale R_{beam} . The solid lines in Figure 4.16 show the best-fit models, which have in total 6 free parameters. The χ^2 values are also shown in the figure, which correspond to 30 data points (24 degrees of freedom). The fits are generally well behaved. I do not show values for the best-fit parameters and their uncertainties because of the caveats I mentioned before related to the estimator and the covariance matrices. Also at the time of my analysis, we did not have mock catalogues to test my methodology.

Wolz et al. 2022 proceeded with the analysis of the cross power spectrum monopole and fitted for the amplitude of the signal, after testing our methods on mock catalogues. Using different scale ranges, we obtained constraints on the combination $\Omega_{\text{HI}} b_{\text{HI}} r$ at $z_{\text{eff}} = 0.8$, where r is an empirical cross-correlation term that depends on the galaxy sample. Table 1 of Wolz et al. 2022 summarises results. For instance, for conservative choices for the analysis, we find $\Omega_{\text{HI}} b_{\text{HI}} r = (0.48 \pm 0.12) \times 10^{-3}$ for GBT x ELGs and $(0.38 \pm 0.12) \times 10^{-3}$ for GBT x LRGs, corresponding to 5 and 4.2σ detections respectively.

This work is on the pathway for precision measurements with future 21cm surveys such as SKA (Square Kilometre Array Cosmology Science Working Group et al. 2020).

5 The low-redshift Universe and its velocities

Contents

5.1	Measuring peculiar velocities	69
5.1.1	The Tully-Fisher relation of spiral galaxies	70
5.1.2	The Fundamental Plane of elliptical galaxies	71
5.1.3	Type-Ia supernovae	72
5.2	Biases affecting velocities	74
5.3	Growth-rate measurements with velocities	75
5.3.1	Maximum likelihood method	76
5.3.2	Compressed two-point statistics	76
5.3.3	Density-velocity comparison	78
5.3.4	Forward-model of density and velocity field	78
5.4	Current measurements	79
5.5	DESI and ZTF	79

At low redshifts, peculiar velocities of galaxies can be directly measured and be used to learn about gravity on large scales. As presented in chapters 1 and 4, velocities are generally not observable other than by their indirect impact on the galaxy clustering through redshift-space distortions (RSD). At $z < 0.2$, the volume that can be probed by galaxy surveys is insufficient to obtain precise measurements of the growth-rate of structures f solely with RSD. The additional information from peculiar velocities significantly improves these constraints, to the point where one can perform consistency tests of the standard cosmological model or of our current model for gravity: general relativity.

In this chapter I will expose the recent work we developed within the cosmology group at the Centre de Physique des Particules de Marseille in trying to combine galaxy surveys and peculiar velocities, particularly those obtained from type-Ia supernovae.

5.1 Measuring peculiar velocities

As discussed in section 1.5.6, redshifts of galaxies can be seen as a mixture of two distinct types of redshift. The first one is due to the expansion of the Universe and we refer to it as *cosmological redshift* z_{cos} . The second one is due to Doppler effect and the peculiar velocities of galaxies with respect to a comoving frame, we refer to it as *peculiar redshift* z_{pec} . The total observed redshift z_{obs} can be written as:

$$1 + z_{\text{obs}} = (1 + z_{\text{cos}})(1 + z_{\text{pec}}) \quad (5.1)$$

Measuring peculiar velocities or redshifts z_{pec} means breaking the degeneracy between z_{cos} and z_{obs} . The observed redshift of a given galaxy can be precisely measured via spectroscopy but its cosmological redshift requires a measurement of its distance or age, of which we can

derive z_{\cos} through a cosmological model. Therefore, peculiar velocity surveys can be seen simply as distance surveys.

Measuring distance of individual objects on cosmological scales is quite challenging since it commonly involves standard or standardisable candles. For cosmological distances, three types of standardisable candles have been used by peculiar velocity surveys: spiral galaxies, elliptical galaxies and type-Ia supernovae. Many decades of studies of galaxies have pointed to interesting correlations between their luminosity or absolute magnitudes, which are not directly observable, to properties or physical quantities that can be directly observed and do not depend on the galaxy distance. Two famous correlations are the Tully-Fisher (TF) relation for spiral galaxies and the Fundamental Plane (FP) for ellipticals, which we will describe shortly after. Needless to say, type-Ia supernovae (SNIa) have been proven to be excellent standardisable candles when accounting for correlations between their luminosity and the duration and colour of their light-curves. As discussed in section 1.5.2, SNIa are by themselves a key cosmological probe of the Universe's expansion.

The precision of distance measurements is highly dependent on the modelling of correlations between luminosity, from which we derive luminosity distances, and their observed properties. Even after standardisation, each method remains with some intrinsic scatter in distances. For instance, TF and FP distances have generally a scatter of about 20% while SNIa can reduce it to about 8%. For cosmological applications, one SNIa distance is worth 4 to 5 TF or FP distances to obtain the same precision.

5.1.1 The Tully-Fisher relation of spiral galaxies

The Tully-Fisher relation is an empirical link between the mass of spiral galaxies and their rotational velocity at the regime where it does not vary much with distance to the center. Tully and Fisher 1977 where the first to find a correlation between a distance-independent observable, the asymptotic rotational velocity, and mass or luminosity, and to use it as a method to infer galaxy distances. The physical origin of the relation is the virial theorem of a collapsed object.

Later works found that the infrared luminosity correlates better with rotational velocity, i.e., it reduces the remaining scatter. This is because infrared wavelengths are less sensitive to dust extinction than optical wavelengths. McGaugh et al. 2000 have later shown that the standard Tully-Fisher relation breaks down for galaxies with small rotation velocities. However, those galaxies contain relatively more gas than stars. If the total baryonic mass is used instead of only stellar mass, the relation holds even for these smaller galaxies. This is known today as the baryonic Tully-Fisher (BTF) relation.

One of the largest samples of BTF distance measurements to date is the one provided by the fourth version of the Cosmicflows project (Kourkchi et al. 2022), containing about 10 000 distances. Figure 5.1 and 5.2 show the angular and redshift distribution of this catalogue. The BTF relation of the sample is calibrated using a set of 600 galaxies in 20 well known clusters. They used information from HI fluxes from several radio telescopes. The HI fluxes are better tracer of the gas content of the galaxies and the HI line width is related to the rotation velocity of these galaxies. For stellar masses, they used photometry from SDSS in optical and the WISE satellite in the infrared. The inferred scatter in their TF relation is of about 45% in the distance modulus (or absolute magnitudes) which corresponds to a scatter of $\ln(10)/5 \times 45\% = 18\%$ in distances. Their sample extends up to $z \sim 0.05$. Hoffman et al. 2021 describes how to compute peculiar velocities from this sample of TF distances, including a correct treatment of uncertainties.

In section 5.3 I will present some of the growth-rate measurements using this or a previous version of the CosmicFlows catalogue of distances.

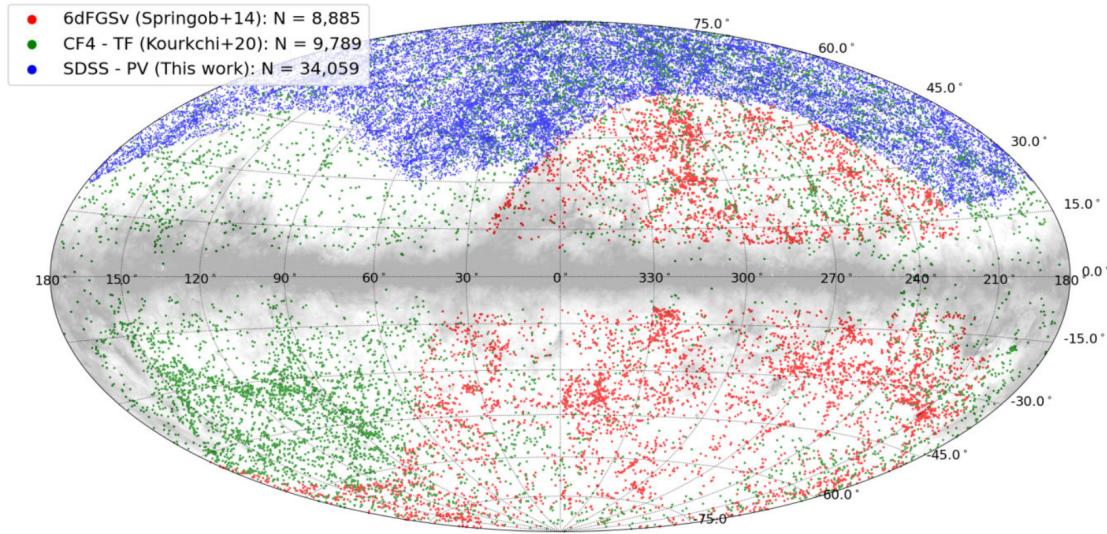


Figure 5.1: Angular distribution of peculiar velocity samples in Galactic coordinates: SDSS (blue) and 6dFGSv (red) from Fundamental Plane and Cosmicflows-4 (green) from Tully-Fisher relation. Figure extracted from Howlett et al. 2022.

5.1.2 The Fundamental Plane of elliptical galaxies

The equivalent of the Tully-Fisher relation for elliptical galaxies is named the Fundamental Plane (FP, Djorgovski and Davis 1987). It is an extension in two-dimensions of a previously thought-to-be one-dimensional relation between the luminosity of the elliptical galaxy and the velocity dispersion of their stars near their centres. The one-dimensional version is known as the Faber-Jackson relation (Faber and Jackson 1976). Later, this relation was extended to two-dimensions to relate the effective radius of the galaxy (in kpc) to the velocity dispersion and the average surface brightness within the effective angular radius. The last two quantities are distance-independent and can be observed with spectroscopy and photometry respectively. From these, we can derive the effective radius which is then converted to a distance measurement.

The largest sample of FP distances to date is the one produced from the SDSS project (Howlett et al. 2022) containing about 34 000 distances. This sample recently surpassed the previous larger sample from the 6-degree Field Galaxy Survey peculiar velocity sample (6dFGSv, Springob et al. 2014) which contained nearly 9 000 distances. Figure 5.1 and 5.2 show the angular and redshift distribution of these two catalogues, respectively. The final estimated scatter of the SDSS sample is about 50% in distance moduli which corresponds to about 23% in distances, which is similar to the scatter of TF distances.

In section 5.3 I will present some of the growth-rate measurements using the 6dFGSv sample of distances. Currently there are no published results using the larger SDSS sample.

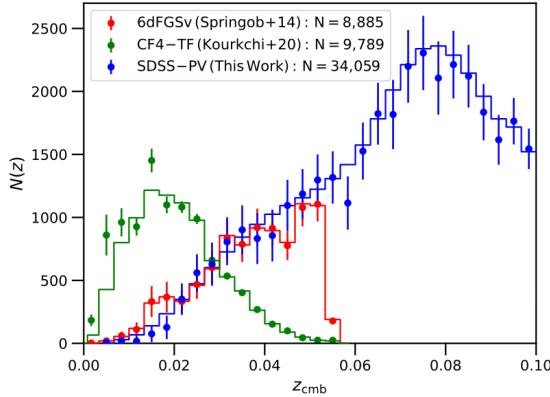


Figure 5.2: The redshift distribution of the distance samples: SDSS (blue) and 6dFGSv (red) from Fundamental Plane and Cosmicflows-4 (green) from Tully-Fisher relation. Points with error bars show the number of galaxies per bin of width 1000 km/s, where error bars include cosmic variance computed from mock catalogues. Solid lines are the mean redshift distributions from these mocks. Figure extracted from Howlett et al. 2022.

5.1.3 Type-Ia supernovae

As mentioned earlier, type-Ia supernovae (SNIa) have been used as distance indicators to measure the Hubble constant and the equation of state of dark energy. The state of the art results based on SNIa cover up to redshifts of 2.

Since SNIa provide distances, one can also estimate peculiar velocities by looking at the residuals with respect to the best-fit Hubble diagram. Peculiar velocities modify mostly the observed redshifts than the observed magnitude. Magnitudes are only affected by the relativistic beaming, which scales luminosity distances as $(1 + z_{\text{pec}})^2$, so it is a second-order effect. Figure 5.3 sketches how peculiar velocities would impact an ideal case of a SNIa Hubble Diagram without any kind of scatter. We can see how displacements are majoritarily along the redshift direction.

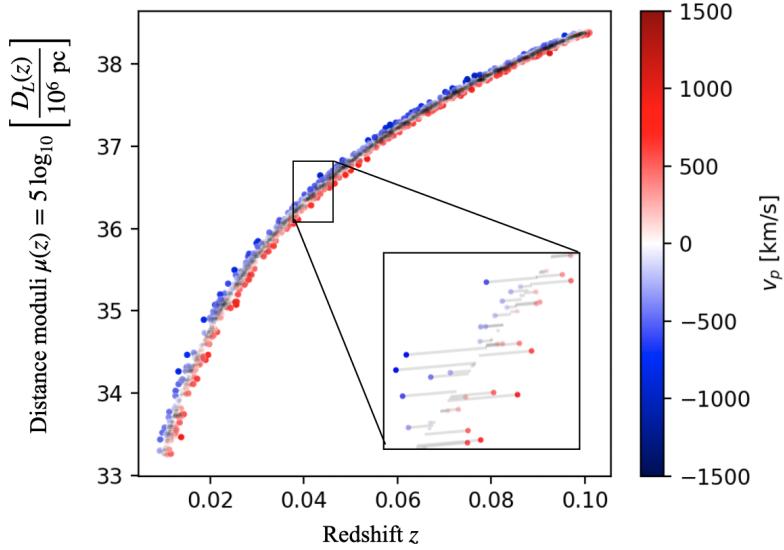


Figure 5.3: Effect of peculiar velocities on an ideal case of a Hubble Diagram of type-Ia supernovae without any intrinsic scatter or observational uncertainties. The displacements along the redshift direction are proportional to $(1 + z_{\text{pec}})$ while those along the distance moduli direction are proportional to $\log_{10}(1 + z_{\text{pec}})$. The colourbar indicates the simulated peculiar velocity of each SNIa.

Until now, peculiar velocities were mostly treated as nuisance when constructing the Hubble Diagram. E. R. Peterson et al. 2021 attempted to correct for peculiar velocities of some SNIa of the Pantheon+ sample (Dan Scolnic et al. 2021) by using reconstruction techniques on galaxy survey data and by identifying groups of galaxies. Peculiar velocities from SNIa have not been widely used to measure growth-rate for two main reasons. First, the small number of detected SNIa at low redshifts, where uncertainties in velocities are not prohibitive. Second, the inhomogeneity in the available samples which originate from several telescopes, cameras, each with its own observing strategy or prior selection of targets. As we will discuss later, selection effects are important when constructing velocity fields with any type of distance tracer.

The Zwicky Transient Facility (ZTF, Graham et al. 2019) is currently observing the whole available sky from the Northern hemisphere and will produce a large sample of about 5 000 SNIa at $z < 0.1$ without any prior selection. This will be a homogeneous set ideal to perform statistical measurements with their peculiar velocities. Figure 5.4 compares the angular distribution of SNIa of the Pantheon+ sample to the ZTF Year 1 sample. Figure 5.5 compares the redshift distribution of the same samples. In section 5.5, I will present the plan of the project of measuring the clustering of ZTF peculiar velocities in combination with galaxies from DESI.

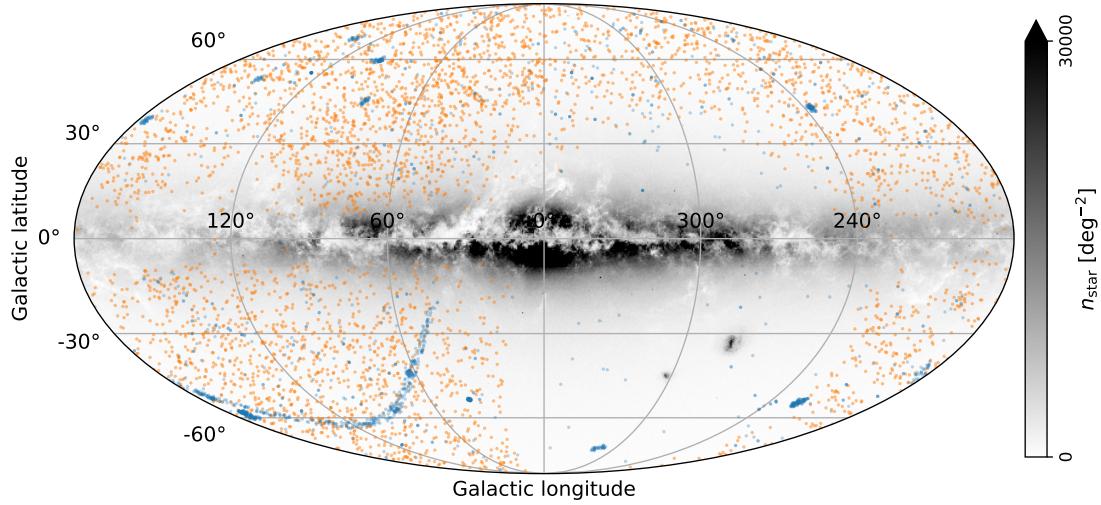


Figure 5.4: Angular distribution of type-Ia supernovae in Galactic coordinates for the Pantheon+ sample (blue) and for ZTF Year 2 sample (orange). On the background we see the stellar density distribution from Gaia.

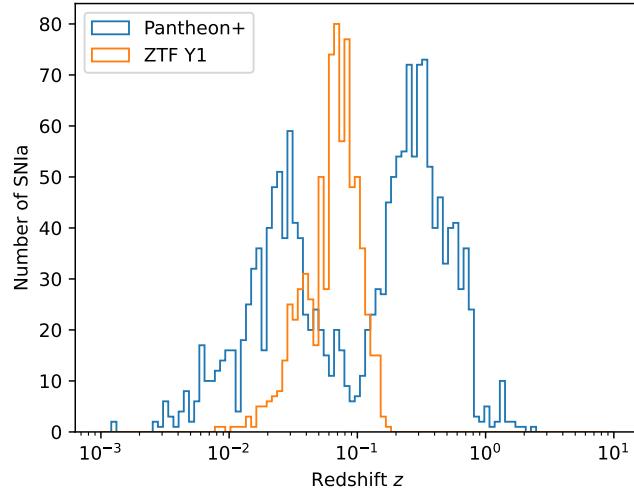


Figure 5.5: Redshift distribution of host galaxies of type-Ia supernovae for the Pantheon+ sample (blue) and for ZTF Year 1 sample (orange).

5.2 Biases affecting velocities

One of the main difficulties with distance measurements based on standardisable candles and their statistical properties (see next section) is the impact of certain types of biases. The review by Strauss and Willick 1995 discuss all these biases and how they impact clustering analyses. They can be classified in three types: selection bias, homogeneous and inhomogeneous Malmquist biases. The term Malmquist is often used to refer to selection bias, but I will follow the convention of Strauss and Willick 1995.

The selection bias, as the name indicates, is caused by some selection or detection criteria on the sample of objects. For instance, it is pretty common to impose some cuts on signal-to-noise ratio of fluxes or some magnitude limit. For objects close to the boundary, only the brighter objects pass the cuts. If one is interested in, say, the average distance of these objects, we will observe a bias relative to the true average distance. The intrinsic scatter in luminosity of all standardisable candles worsens this type of bias and it complicates any attempts to correct for these biases. Typical analyses of SNIa often use realistic simulations of the data, including the physical properties of SNIa that create the scatter and instrumental noise, to model these selection effects. The analysis of the Pantheon SNIa sample used a technique named Bayesian Estimation Applied to Multiple Species (BEAMS, Kunz et al. 2007) with Bias Corrections (BBC, R. Kessler and D. Scolnic 2017). This technique also accounts for mis-classified supernovae.

The homogeneous Malmquist bias manifests itself due to the fact that the number of objects observed as a function of distance to us increases with distance squared. When averaging objects in a given distance or redshift range, they are not uniformly distributed inside this range, so the average distance is higher than the true distance, often assigned to the center of the range. Is this true ? Check!

The inhomogeneous Malmquist bias is a similar effect but generalised to the fluctuations in density (due to structures) along the line of sight. This bias depends on the precise shape of $\delta(\hat{n}, d)$ where \hat{n} is a given direction in the sky and d is the distance to us.

It is possible to correctly account for all these biases in a fully Bayesian treatment of probability density functions of the distances (Graziani et al. 2019; Boruah et al. 2021).

5.3 Growth-rate measurements with velocities

With a set of distance or peculiar velocity measurements, we can study them statistically and compare to predictions from cosmological models, just as we did with the overdensities of fluxes in chapter 3 or galaxies in chapter 4.

One can see this additional set of observables as a scalar field, the radial velocity field $v_r(\vec{x}_i)$, sampled at the positions of the galaxies from which they were measured \vec{x}_i , where $i = 1, \dots, N_{\text{gal}}$. Instead of radial velocities, it is quite common to use the log of the ratio between the distance inferred from the redshift and the distance inferred from the luminosity, called log-distance ratio. Log-distance ratios have uncertainties that can be approximated by a Gaussian, whereas in velocity space the uncertainties become log-normal are harder to handle.

The first statistic that one can think of measuring is the correlation function of the radial velocities $\xi_{v_r v_r} \equiv \langle v_r(\vec{x}_i) v_r(\vec{x}_j) \rangle$ (or the equivalent for log-distance ratios). If one has, in addition to the velocities, a galaxy survey over the same volume, one can use the observed overdensity field $\delta(\vec{x})$ to cross-correlate with the observed velocity field, i.e., $\xi_{\delta_g v_r} \equiv \langle \delta(\vec{x}) v_r(\vec{y}) \rangle$. The auto-correlation of the overdensity field $\xi_{\delta_g \delta_g} \langle \delta(\vec{x}) \delta(\vec{y}) \rangle$ is the equivalent to the one used in BAO and RSD measurements. These three statistics can be analysed jointly when comparing to models, increasing the constraining power.

I will shortly describe a few methods available in the litterature that allows us to connect models to the obseved densities and velocities, at the field level or at the two-point function level.

5.3.1 Maximum likelihood method

The idea is to maximise the likelihood, assumed to be a multivariate Gaussian on the velocities,

$$\mathcal{L}(\vec{p}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det[C(\vec{p})]^{1/2}} \exp \left[-\frac{1}{2} \nu_r^T C(\vec{p})^{-1} \nu_r \right] \quad (5.2)$$

with respect to a set of parameters \vec{p} . Here ν_r is a vector containing each measured peculiar velocity and the covariance matrix $C(\vec{p})$ is the model for the two-point statistics of ν_r . Each element of the covariance is given by $C_{ij} = \xi_{\nu_r \nu_r}^{\text{model}}(\vec{x}_i, \vec{x}_j, \vec{p}) + \delta_{ij}^K \sigma_i \sigma_j$, where σ_i is a diagonal noise term, often accounting for instrumental uncertainties as well as intrinsic scatter of distance measurements. The first term in the covariance is the cosmic variance and it is connected to a cosmological model.

The cosmological part of the covariance C_{ij} is often written as a Fourier transform of the two-point statistics in Fourier space, often described by the velocity or the velocity-divergence power spectrum. Most of previous works have used linear perturbation theory with some empirical terms that account for non-linearities. In linear theory, the velocity field $\vec{v}(\vec{x})$ is irrotational, meaning it is completely defined as the gradient of a scalar field $\theta(\vec{x})$. This velocity divergence is equal to the overdensity of matter $\delta(\vec{x})$ for linear perturbations. Therefore the two-point statistics of the velocity is proportional to the linear matter power spectrum (and thus to σ_8^2) and to the linear growth-rate of structures f .

The maximum likelihood method has been used on several peculiar velocity samples such as the 6dFGSv (Campbell et al. 2014; Springob et al. 2014) with velocities alone (Johnson et al. 2014), in combination with density field (Adams and Blake 2017; Adams and Blake 2020), or adding a few SNIa (Huterer et al. 2017).

At CPPM, Bastien Carreres is currently employing this technique to analyse the ZTF sample of type-Ia supernovae. I have implemented the basic algorithm and Bastien continued its development, adapting it to work with log-distance ratio instead of velocities, several mesh assignement schemes and models of clustering. He also implemented a whole pipeline to produce realistic simulations of ZTF SNIa light-curves, which can be painted on n-body simulations, and to analyse them as real data. He is currently studying the impact of selection biases on the estimates of $f\sigma_8$ using the maximum likelihood method. First results on simulated data show that an unbiased measurement is possible if we limit our analysis to SNIa at $z < 0.05$. We will apply our methodology to real data once the final ZTF SNIa is set.

5.3.2 Compressed two-point statistics

The compression of two-point statistics means that we combine measurements into separation or wavenumber bins, depending if we are in configuration or Fourier space. Estimating compressed statistics is the most common method for standard clustering analysis, as we saw in earlier chapters, where we aim to estimate the correlation function or the power spectrum. The modelling happens at the compressed statistics level, so there are generally less degrees of freedom than for the maximum likelihood method.

In configuration space, the velocity auto correlation cannot simply be written as a unique function of separation \vec{r} between two galaxies, since velocities are a vector field. K. Gorski 1988 introduced the correlation tensor defined as

$$\Psi_{i,j}(\vec{r}_A, \vec{r}_B) \equiv \langle \nu_i(\vec{r}_A) \nu_j(\vec{r}_B) \rangle \quad (5.3)$$

where i, j denote each component of the peculiar velocity field \vec{v} . If the field is irrotational, homogeneous and isotropic, the correlation tensor can be written as

$$\Psi_{i,j}(\vec{r}) = [\Psi_{\parallel}(r) - \Psi_{\perp}(r)] \hat{r}_{Ai} \hat{r}_{Bj} + \Psi_{\perp}(r) \delta^K_{ij} \quad (5.4)$$

where $\Psi_{\parallel}(r)$ and $\Psi_{\perp}(r)$ are functions describing the correlations the components of the peculiar velocities parallel and perpendicular to the separation vector \vec{r} . Both functions can be written as transforms of the linear matter power spectrum.

The correlations between the radial velocities of two galaxies relative to the observer, as depicted in Figure 5.6, can be written as

$$\langle u_A(\vec{x}) u_B(\vec{x} + \vec{r}) \rangle = \Psi(r) \cos \theta_{AB} + [\Psi_{\parallel}(r) - \Psi_{\perp}(r)] \cos \theta_A \cos \theta_B. \quad (5.5)$$

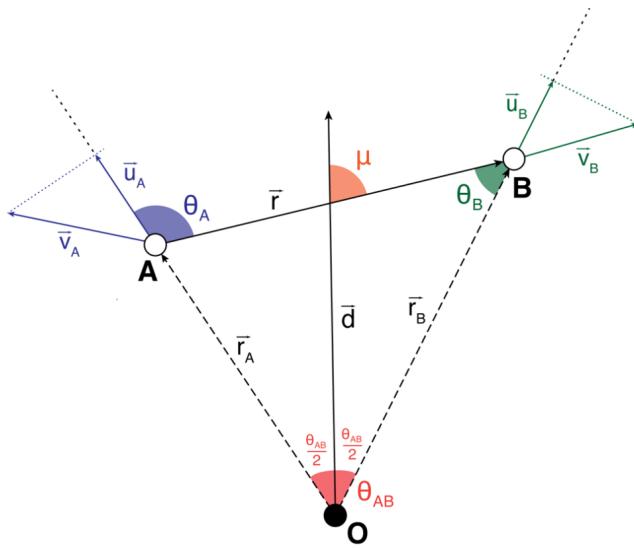


Figure 5.6: Non-parallel lines-of-sight of a pair of galaxies A and B as seen by an observer O. \vec{r}_A and \vec{r}_B are their position vectors, respectively, and \vec{r} is the separation vector between them. The peculiar velocities of galaxies A and B are represented by \vec{v}_A and \vec{v}_B , and the radial component of these velocities are represented by \vec{u}_A and \vec{u}_B . The pair line-of-sight \vec{d} is defined at the mid-angle and $\mu \equiv \hat{d} \cdot \hat{r}$. Figure extracted from Turner et al. 2022.

K. M. Gorski et al. 1989 developed estimators that we can relate to a model that depends on Ψ_{\parallel} and Ψ_{\perp} . This type of correlation functions were used in Y. Wang et al. 2018; Dupuy et al. 2019; Turner et al. 2021; Turner et al. 2022.

In Fourier space, given the sparsity of the data, it is hard to use FFTs on the observed velocity field. This is because voxels without data cannot be assigned zero velocity: the field would be ill-defined there. For this reason, Howlett 2019 suggests to use the momentum power spectrum, a quantity commonly used in the analysis of velocities in n-body simulations. The momentum is simply the mass-weighted velocity field, written as $\vec{p}(\vec{x}) = [1 + \delta(\vec{x})] \vec{v}(\vec{x})$. In the absence of galaxies or velocity measurements, the momentum is well defined and equal to zero.

The first application of the momentum power spectrum method was performed by Qin et al.

2019 on the combined sample of 6dFGSv and the 2MASS Tully-Fisher survey (2MTF).

The analysis in compressed statistics is generally faster and more intuitive than other methods. The data versus model comparison is commonly easier with compressed statistics. However, one has to be careful when accounting for survey properties, wide-angle effects or redshift evolution. These effects are generally easier to deal with in other methods.

5.3.3 Density-velocity comparison

The link between density and velocity fields is straightforward on quite large scales, where linear theory is valid. Another method to test for the consistency of the standard cosmological model is to compare the observed (radial) velocity field with predictions based on the observed density field. This method is often referred to as density-velocity comparison.

Reconstruction techniques discussed in section 4.2 aim to compute past trajectories of galaxies and therefore their velocities based on the observed density field. Reconstruction yields a prediction for the radial velocities at the locations where we measured them from the distance surveys. The inputs are often some value for the large scale bias b of the galaxy sample as well as the growth-rate f . Commonly, an empirical amplitude A is fitted to the relation $v_r^{\text{obs}} = A v_r^{\text{model}}$. Any deviations from $A = 1$ would indicate an inconsistency in the standard cosmological model. From this amplitude, one can quote a value for $f\sigma_8$ if a bias value and a σ_8 value are assumed.

Several applications of this method can be found in the literature. Davis et al. 2011 used the Two Micron All-Sky Redshift Survey (2MRS or 2M++ for its extended version) galaxy catalogue as the input for reconstruction and fitted the inverse Tully-Fisher relation to spiral galaxies from the Spiral Field I-band ++ survey (SFI++). Carrick et al. 2015 improved the analysis and reported results using the same samples. Boruah et al. 2020 added peculiar velocities from 465 SNIa to the SFI++ and the 2MTF samples. Said et al. 2020 used a larger dataset of nearly 16k peculiar velocities by combining 6dFGSv and SDSS samples, and compared to the density field from the 2M++.

At CPPM, Elena Sarpa has a more advanced reconstruction technique called extended Fast Action Minimization (eFAM, Sarpa et al. 2021) which can be readily used to implement the density-velocity comparison. The eFAM technique yields higher-order trajectories than Zeldovich reconstruction, which might lead to more realistic predictions for the peculiar velocities. This technique will be potentially applied to the DESI and ZTF data, after a detailed comparison with the maximum likelihood method.

5.3.4 Forward-model of density and velocity field

Instead of estimating summary statistics with the density and velocity fields, one can attempt to fit directly the observed fields with some model that depend on cosmological parameters. Typically, the model is composed of some cosmological parameters and a set of initial density and velocity fields (one amplitude parameter per mode). These initial conditions are then evolved linearly or using more advanced techniques and then compared to observations. The final product is a posterior on cosmological parameters and on the amplitudes of the initial fields. This method is commonly referred to as *forward-modelling*.

In forward modelling, complications due to selection functions are slightly easier to deal with in principle. One has to simulate observations on the evolved fields just as one does when creating mock catalogues. Also, by modelling the observed field entirely, we capture more information from higher-order statistics than just the two-point function. The inconvenience of

Table 5.1: Measurements of the growth-rate of structures from peculiar velocity and galaxy survey data.

Reference	Dataset	Method	$f\sigma_8$
Beutler et al. 2012	6dFGSz	RSD	0.423 ± 0.055
Blake et al. 2018	6dFGSz	RSD	0.380 ± 0.120
Johnson et al. 2014	6dFGSv	Max-likelihood	$0.428^{+0.079}_{-0.068}$
Huterer et al. 2017	6dFGSv	Max-likelihood	$0.428^{+0.048}_{-0.045}$
Adams and Blake 2020	6dFGSz,6dFGSv	Max-likelihood	$0.384 \pm 0.052^{\text{stat}} \pm 0.061^{\text{syst}}$
Qin et al. 2019	6dFGSv,2MTF	Compressed 2pt	$0.404^{+0.082}_{-0.081}$
Howlett et al. 2017b	2MTF	Compressed 2pt	$0.510^{+0.090}_{-0.080}$
Nusser 2017	CF3,2MRS	Compressed 2pt	0.400 ± 0.080
Dupuy et al. 2019	CF3	Compressed 2pt	$0.430 \pm 0.030^{\text{obs}} \pm 0.110^{\text{cosmic}}$
Davis et al. 2011	2MRS,SFI++	$\delta - v$ comparison	0.310 ± 0.040
Carrick et al. 2015	2MRS,SFI++	$\delta - v$ comparison	0.401 ± 0.024
Boruah et al. 2020	2M++,2MTE,SFI++,A2 SNIa	$\delta - v$ comparison	0.400 ± 0.017
Said et al. 2020	2M++,6dFGSv,SDSS-FP	$\delta - v$ comparison	0.311 ± 0.027

this method is the computing time required to perform the inference, given the large number of free parameters and the time to evolve each sample from initial conditions.

Lavaux 2016 implemented the forward-modelling applied to both density and velocity fields. A similar approach was used on CosmicFlows3 data by Graziani et al. 2019. In Boruah et al. 2021 they discuss the treatment of the inhomogenous Malmquist bias in the Bayesian formalism (see section 5.2). In Prideaux-Ghee et al. 2022, they used a Bayesian hierarchical modelling approach to fit both density and velocity fields using a Hamiltonian Monte Carlo sampling of the likelihood.

5.4 Current measurements

Several measurements were made in the past decade using a variety of distance survey data and methods. Table 5.1 attempts to summarise all measurements and their methodologies.

5.5 DESI and ZTF

Both the Dark Energy Spectroscopic Instrument (DESI) and the Zwicky Transient Facility (ZTF) are currently observing similar area of the sky and building one of the best samples for a joint study of densities and peculiar velocities for cosmology. One of the main advantages is the great overlap in volume of both surveys. The 14k deg^2 footprint of DESI will be completely covered by ZTF observations, which cover up to 18k deg^2 of area.

The main science goal of DESI is to produce precise measurements of the Universe's expansion and the growth history of structures, in order to test models of dark energy and gravity. DESI will observe more than 20 million galaxies and quasars from up to redshifts of 4.5. The bright time of the survey, i.e., when the moon is up, is dedicated to building a flux limited sample of galaxies at low redshift: the Bright Galaxy Survey (BGS). The BGS will collect roughly 10 million galaxies with r -band magnitude $r < 19.5$, which lie at $z < 0.6$. A fainter $19.5 < r < 20.175$ sample will extend and increase the completeness in stellar mass of the BGS

sample. This sample is not only ideal for BAO and standard RSD measurements, but also for higher order statistical measurements, multi-tracer analyses, given its high completeness and simple selection function. Most of the ZTF SNIa will have their host-galaxy redshifts measured by the DESI BGS. It is expected that the BGS observations will complete by 2024 the latest.

The ZTF survey main goal is to observe the transient sky in three optical bands. Most of the sky area is covered every two nights with a given filter, which is excellent cadence for discovery and study of SNIa. At its current configuration, ZTF is expected to discover around 15k supernovae, out of which about 30% will have spectroscopic classifications. The host galaxy redshifts are obtained from archival data, some of which will be observed by DESI in the near future. We expect a final sample of 5k cosmology-level spectroscopic SNIa at $z < 0.1$.

Figure 5.7 compares the density of tracers between DESI BGS and ZTF SNIa, expected at the end of their 5-year programs. The density of ZTF SNIa falls quickly after $z \sim 0.06$ where selection effects start to kick in. Rough estimates show that the ZTF SNIa sample is *complete*, i.e., all potentially observable SNIa were observed, at $z < 0.05$. While the density of ZTF SNIa is several orders of magnitudes lower than those of BGS galaxies, the fact that each SNIa yields a measurement of the velocity field makes them a powerful tracer nevertheless.

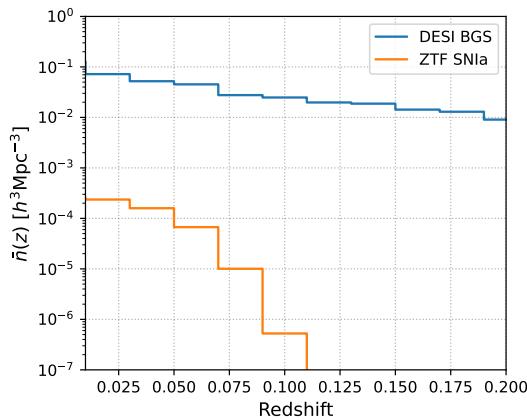


Figure 5.7: Comoving density of tracers as a function of redshift for the galaxies of the DESI Bright Galaxy Survey (blue) and for type-Ia supernovae from ZTF (orange). Densities were computed assuming a flat ΛCDM model with Planck 2018 best-fit parameters, based on simulated data and scaled to match numbers expected at the end of both programs. Both samples overlap in redshift and in sky coverage, which is ideal for joint studies of densities and peculiar velocities.

I studied the constraining power of DESI BGS and ZTF SNIa using the Fisher formalism. I computed the uncertainty expected on $f\sigma_8$ from a RSD analysis with DESI BGS galaxies by themselves or from a joint RSD + PV analysis of DESI BGS and ZTF SNIa. I employed the same Fisher forecast code¹ described in Howlett et al. 2017a. The inputs of the calculation are the expected comoving density of tracers for galaxies and peculiar velocities at the end of both surveys, the intrinsic scatter of the distance indicators parametrised by $\alpha \equiv \sigma(D_L)/D_L \sim 10$ or 20%, and the scale range (in Fourier space) assumed for the measurement parametrised by $k_{\max} \sim 0.1$ or $0.2 h\text{Mpc}^{-1}$. The model for the galaxy clustering is a simple linear RSD model with galaxy bias b and a Gaussian Fingers-of-God term function of a galaxy velocity

¹https://github.com/CullanHowlett/PV_fisher

Table 5.2: Fisher forecast on measurements of the growth-rate of structures times the normalisation of the power spectrum $f\sigma_8(z_{\text{eff}}) = 0.45$ where $z_{\text{eff}} = 0.08$ for the complete samples of DESI BGS and ZTF type-Ia supernovae. Results are shown for different analysis choices: k_{max} is the maximum wavenumber used in the analysis, $\sigma(D_L)/D_L$ is the fractional error on luminosity distances due to intrinsic scatter. The assumed number densities of tracers is the one shown in Figure 5.7. The total assumed sky area with overlapping coverage from both experiments is 14k deg^2 .

Dataset	k_{max}	$\sigma(D_L)/D_L$	$\sigma(f\sigma_8(z_{\text{eff}}))/f\sigma_8(z_{\text{eff}})$
DESI BGS	0.1	-	0.58
DESI BGS	0.2	-	0.21
ZTF SNIa	0.1	0.05	0.22
ZTF SNIa	0.1	0.10	0.35
ZTF SNIa	0.2	0.05	0.19
ZTF SNIa	0.2	0.10	0.32
DESI BGS + ZTF SNIa	0.1	0.05	0.12
DESI BGS + ZTF SNIa	0.1	0.10	0.20
DESI BGS + ZTF SNIa	0.2	0.05	0.09
DESI BGS + ZTF SNIa	0.2	0.10	0.13

dispersion σ_v . The model for the velocity correlations is similar, scaling with f , but with a different empirical correction for the fact we observe velocities in redshift-space (Koda et al. 2014). This empirical function depends on an extra dispersion parameter σ_u . A total of four free parameters define the model. I focused on the constraints on $f\sigma_8$ by using tracers over $0 < z < 0.1$, which corresponds to a measurement at an effective redshift $z_{\text{eff}} = 0.08$.

Results of the Fisher forecast are shown in Table 5.2 for different samples and analysis choices. First, we can see how the uncertainties on $f\sigma_8$ are significantly reduced when including ZTF SNIa, particularly when the intrinsic scatter is small and when using a larger range of scales. We see an improvement of a factor of 2.3 at most. Second, we see how increasing the scale range k_{max} also improves constraints. This is why it is important to be able to model the clustering of galaxies and peculiar velocities on non-linear scales. Third, we see how improving the standardisation of SNIa also helps reducing uncertainties on $f\sigma_8$. In conclusion, a simple Fisher forecast predicts a 9% measurement of $f\sigma_8$ in the optimistic scenario and a 20% in the pessimistic scenario.

Conclusion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Bibliography

- Abbott, B. P. et al. (2017). “A Gravitational-Wave Standard Siren Measurement of the Hubble Constant”. In: *Nature* 551.7678 (7678), pp. 85–88. DOI: [10.1038/nature24471](https://doi.org/10.1038/nature24471) (cit. on p. 17).
- Adams, Caitlin and Chris Blake (11, 2017). “Improving Constraints on the Growth Rate of Structure by Modelling the Density–Velocity Cross-Correlation in the 6dF Galaxy Survey”. In: *Monthly Notices of the Royal Astronomical Society* 471.1, pp. 839–856. DOI: [10.1093/mnras/stx1529](https://doi.org/10.1093/mnras/stx1529) (cit. on p. 76).
- (21, 2020). “Joint Growth-Rate Measurements from Redshift-Space Distortions and Peculiar Velocities in the 6dF Galaxy Survey”. In: *Monthly Notices of the Royal Astronomical Society* 494.3, pp. 3275–3293. DOI: [10.1093/mnras/staa845](https://doi.org/10.1093/mnras/staa845) (cit. on pp. 76, 79).
- Ahumada, Romina et al. (1, 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”. In: *The Astrophysical Journal Supplement Series* 249, p. 3. DOI: [10.3847/1538-4365/ab929e](https://doi.org/10.3847/1538-4365/ab929e) (cit. on p. 26).
- Aiola, Simone et al. (30, 2020). “The Atacama Cosmology Telescope: DR4 Maps and Cosmological Parameters”. In: *Journal of Cosmology and Astroparticle Physics* 2020.12, pp. 047–047. DOI: [10.1088/1475-7516/2020/12/047](https://doi.org/10.1088/1475-7516/2020/12/047) (cit. on p. 21).
- Alam, Shadab et al. (1, 2017). “The Clustering of Galaxies in the Completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological Analysis of the DR12 Galaxy Sample”. In: *Monthly Notices of the Royal Astronomical Society* 470, pp. 2617–2652. DOI: [10.1093/mnras/stx721](https://doi.org/10.1093/mnras/stx721) (cit. on pp. 22, 64).
- Alam, Shadab et al. (28, 2021). “Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from Two Decades of Spectroscopic Surveys at the Apache Point Observatory”. In: *Physical Review D* 103.8, p. 083533. DOI: [10.1103/PhysRevD.103.083533](https://doi.org/10.1103/PhysRevD.103.083533) (cit. on pp. 20, 22, 62).
- Angulo, Raul E. and Oliver Hahn (1, 2022). “Large-Scale Dark Matter Simulations”. In: *Living Reviews in Computational Astrophysics* 8, p. 1. DOI: [10.1007/s41115-021-00013-z](https://doi.org/10.1007/s41115-021-00013-z) (cit. on p. 52).
- Aver, Erik, Danielle A. Berg, Keith A. Olive, Richard W. Pogge, John J. Salzer, and Evan D. Skillman (1, 2021). “Improving Helium Abundance Determinations with Leo P as a Case Study”. In: *Journal of Cosmology and Astroparticle Physics* 2021.03, p. 027. DOI: [10.1088/1475-7516/2021/03/027](https://doi.org/10.1088/1475-7516/2021/03/027) (cit. on p. 19).
- Aylor, K. et al. (21, 2017). “A Comparison of Cosmological Parameters Determined from CMB Temperature Power Spectra from the South Pole Telescope and the Planck Satellite”. In: *The Astrophysical Journal* 850.1, p. 101. DOI: [10.3847/1538-4357/aa947b](https://doi.org/10.3847/1538-4357/aa947b) (cit. on p. 21).
- Balbi, A. et al. (1, 2000). “Constraints on Cosmological Parameters from MAXIMA-1”. In: *The Astrophysical Journal* 545, pp. L1–L4. DOI: [10.1086/317323](https://doi.org/10.1086/317323) (cit. on p. 9).
- Balkenhol, L. et al. (4, 2021). “Constraints on \$\Lambda\$CDM Extensions from the SPT-3G 2018 \$EE\$ and \$TE\$ Power Spectra”. In: *Physical Review D* 104.8, p. 083509. DOI: [10.1103/PhysRevD.104.083509](https://doi.org/10.1103/PhysRevD.104.083509) (cit. on p. 21).
- Battye, Richard A., Rod D. Davies, and Jochen Weller (1, 2004). “Neutral Hydrogen Surveys for High-Redshift Galaxy Clusters and Protoclusters”. In: *Monthly Notices of the Royal Astro-*

Bibliography

- nonical Society* 355, pp. 1339–1347. DOI: [10.1111/j.1365-2966.2004.08416.x](https://doi.org/10.1111/j.1365-2966.2004.08416.x) (cit. on p. 66).
- Bautista, Julian E. et al. (1, 2017). “Measurement of Baryon Acoustic Oscillation Correlations at $z = 2.3$ with SDSS DR12 Ly α -Forests”. In: *Astronomy and Astrophysics* 603, A12. DOI: [10.1051/0004-6361/201730533](https://doi.org/10.1051/0004-6361/201730533) (cit. on pp. 27, 31, 34, 35).
- Bautista, Julian E. 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”. In: *The Astrophysical Journal* 863.1, p. 110. DOI: [10.3847/1538-4357/aacea5](https://doi.org/10.3847/1538-4357/aacea5) (cit. on pp. 45, 49).
- Bautista, Julian E. et al. (1, 2021). “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”. In: *Monthly Notices of the Royal Astronomical Society* 500, pp. 736–762. DOI: [10.1093/mnras/staa2800](https://doi.org/10.1093/mnras/staa2800) (cit. on pp. 22, 45, 57, 58, 62, 64).
- Becker, George D., Paul C. Hewett, Gábor Worseck, and J. Xavier Prochaska (1, 2013). “A Refined Measurement of the Mean Transmitted Flux in the Ly-alpha Forest over $2 < z < 5$ Using Composite Quasar Spectra”. In: *Monthly Notices of the Royal Astronomical Society* 430, pp. 2067–2081. DOI: [10.1093/mnras/stt031](https://doi.org/10.1093/mnras/stt031) (cit. on p. 32).
- Bernardeau, F., S. Colombi, E. Gaztanaga, and R. Scoccimarro (2002). “Large-Scale Structure of the Universe and Cosmological Perturbation Theory”. In: *Physics Reports* 367.1-3, pp. 1–248. DOI: [10.1016/S0370-1573\(02\)00135-7](https://doi.org/10.1016/S0370-1573(02)00135-7) (cit. on pp. 51, 59).
- Beutler, Florian et al. (11, 2012). “The 6dF Galaxy Survey: $z \approx 0$ Measurement of the Growth Rate and Sigma_8”. In: *Monthly Notices of the Royal Astronomical Society* 423.4, pp. 3430–3444. DOI: [10.1111/j.1365-2966.2012.21136.x](https://doi.org/10.1111/j.1365-2966.2012.21136.x) (cit. on pp. 20, 79).
- Bharadwaj, Somnath, Biman B. Nath, and Shiv K. Sethi (1, 2001). “Using HI to Probe Large Scale Structures at $z \sim 3$ ”. In: *Journal of Astrophysics and Astronomy* 22, pp. 21–34. DOI: [10.1007/BF02933588](https://doi.org/10.1007/BF02933588) (cit. on p. 66).
- Bianchi, Davide, Héctor Gil-Marín, Rossana Ruggeri, and Will J. Percival (1, 2015). “Measuring Line-of-Sight-Dependent Fourier-space Clustering Using FFTs”. In: *Monthly Notices of the Royal Astronomical Society* 453, pp. L11–L15. DOI: [10.1093/mnrasl/slv090](https://doi.org/10.1093/mnrasl/slv090) (cit. on p. 53).
- Bianchi, Davide and Will J. Percival (2017). “Unbiased Clustering Estimation in the Presence of Missing Observations”. In: *Monthly Notices of the Royal Astronomical Society* 472.1, p. 1106. DOI: [10.1093/mnras/stx2053](https://doi.org/10.1093/mnras/stx2053) (cit. on pp. 47, 53).
- BICEP/Keck et al. (26, 2021). “A Demonstration of Improved Constraints on Primordial Gravitational Waves with Delensing”. In: *Physical Review D* 103.2, p. 022004. DOI: [10.1103/PhysRevD.103.022004](https://doi.org/10.1103/PhysRevD.103.022004) (cit. on p. 21).
- Birrer, S. et al. (1, 2020). “TDCOSMO - IV. Hierarchical Time-Delay Cosmography – Joint Inference of the Hubble Constant and Galaxy Density Profiles”. In: *Astronomy & Astrophysics* 643, A165. DOI: [10.1051/0004-6361/202038861](https://doi.org/10.1051/0004-6361/202038861) (cit. on p. 17).
- Blake, Chris, Paul Carter, and Jun Koda (1, 2018). “Power Spectrum Multipoles on the Curved Sky: An Application to the 6-Degree Field Galaxy Survey”. In: *Monthly Notices of the Royal Astronomical Society* 479, pp. 5168–5183. DOI: [10.1093/mnras/sty1814](https://doi.org/10.1093/mnras/sty1814) (cit. on p. 79).
- Blanton, Michael R. et al. (1, 2003). “An Efficient Targeting Strategy for Multiobject Spectrograph Surveys: The Sloan Digital Sky Survey “Tiling” Algorithm”. In: *The Astronomical Journal* 125, pp. 2276–2286. DOI: [10.1086/344761](https://doi.org/10.1086/344761) (cit. on p. 26).

Bibliography

- Blanton, Michael R. et al. (2017). "Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe". In: *The Astronomical Journal* 154.1, p. 28. DOI: [10.3847/1538-3881/aa7567](https://doi.org/10.3847/1538-3881/aa7567) (cit. on p. 20).
- Blomqvist, Michael et al. (2018). "The Triply-Ionized Carbon Forest from eBOSS: Cosmological Correlations with Quasars in SDSS-IV DR14". In: *Journal of Cosmology and Astroparticle Physics* 2018.5, p. 029. DOI: [10.1088/1475-7516/2018/05/029](https://doi.org/10.1088/1475-7516/2018/05/029) (cit. on p. 36).
- Blomqvist, Michael et al. (2019). "Baryon Acoustic Oscillations from the Cross-Correlation of Ly α Absorption and Quasars in eBOSS DR14". In: *Astronomy and Astrophysics* 629, A86. DOI: [10.1051/0004-6361/201935641](https://doi.org/10.1051/0004-6361/201935641) (cit. on p. 31).
- Boruah, Supranta S., Michael J. Hudson, and Guilhem Lavaux (1, 2020). "Cosmic Flows in the Nearby Universe: New Peculiar Velocities from SNe and Cosmological Constraints". In: *Monthly Notices of the Royal Astronomical Society* 498, pp. 2703–2718. DOI: [10.1093/mnras/staa2485](https://doi.org/10.1093/mnras/staa2485) (cit. on pp. 78, 79).
- Boruah, Supranta S., Guilhem Lavaux, and Michael J. Hudson (30, 2021). "Reconstructing Dark Matter Distribution with Peculiar Velocities: Bayesian Forward Modelling with Corrections for Inhomogeneous Malmquist Bias" (cit. on pp. 75, 79).
- Brout, D. et al. (2019a). "First Cosmology Results Using SNe Ia from the Dark Energy Survey: Analysis, Systematic Uncertainties, and Validation". In: *The Astrophysical Journal* 874.2, p. 150. DOI: [10.3847/1538-4357/ab08a0](https://doi.org/10.3847/1538-4357/ab08a0) (cit. on p. 18).
- Brout, D. et al. (1, 2019b). "First Cosmology Results Using Type Ia Supernovae from the Dark Energy Survey: Photometric Pipeline and Light-curve Data Release". In: *The Astrophysical Journal* 874, p. 106. DOI: [10.3847/1538-4357/ab06c1](https://doi.org/10.3847/1538-4357/ab06c1) (cit. on p. 18).
- Burden, A., W. J. Percival, and C. Howlett (1, 2015). "Reconstruction in Fourier Space". In: *Monthly Notices of the Royal Astronomical Society* 453, pp. 456–468. DOI: [10.1093/mnras/stv1581](https://doi.org/10.1093/mnras/stv1581) (cit. on p. 51).
- Burden, A., W. J. Percival, M. Manera, Antonio J. Cuesta, Mariana Vargas Magana, and Shirley Ho (1, 2014). "Efficient Reconstruction of Linear Baryon Acoustic Oscillations in Galaxy Surveys". In: *Monthly Notices of the Royal Astronomical Society* 445, pp. 3152–3168. DOI: [10.1093/mnras/stu1965](https://doi.org/10.1093/mnras/stu1965) (cit. on p. 51).
- Busca, N. G. et al. (1, 2013). "Baryon Acoustic Oscillations in the Ly α Forest of BOSS Quasars". In: *Astronomy and Astrophysics* 552, A96. DOI: [10.1051/0004-6361/201220724](https://doi.org/10.1051/0004-6361/201220724) (cit. on pp. 31, 32).
- Campbell, Lachlan A. et al. (1, 2014). "The 6dF Galaxy Survey: Fundamental Plane Data". In: *Monthly Notices of the Royal Astronomical Society* 443, pp. 1231–1251. DOI: [10.1093/mnras/stu1198](https://doi.org/10.1093/mnras/stu1198) (cit. on p. 76).
- Carlson, Jordan, Beth Reid, and Martin White (1, 2013). "Convolution Lagrangian Perturbation Theory for Biased Tracers". In: *Monthly Notices of the Royal Astronomical Society* 429, pp. 1674–1685. DOI: [10.1093/mnras/sts457](https://doi.org/10.1093/mnras/sts457) (cit. on p. 59).
- Carrick, Jonathan, Stephen J. Turnbull, Guilhem Lavaux, and Michael J. Hudson (11, 2015). "Cosmological Parameters from the Comparison of Peculiar Velocities with Predictions from the 2M++ Density Field". In: *Monthly Notices of the Royal Astronomical Society* 450.1, pp. 317–332. DOI: [10.1093/mnras/stv547](https://doi.org/10.1093/mnras/stv547) (cit. on pp. 78, 79).
- Carter, Paul, Florian Beutler, Will J. Percival, Joseph DeRose, Risa H. Wechsler, and Cheng Zhao (1, 2020). "The Impact of the Fiducial Cosmology Assumption on BAO Distance Scale Measurements". In: *Monthly Notices of the Royal Astronomical Society* 494, pp. 2076–2089. DOI: [10.1093/mnras/staa761](https://doi.org/10.1093/mnras/staa761) (cit. on p. 51).

Bibliography

- Chabanier, Solène et al. (1, 2020). “The Impact of AGN Feedback on the 1D Power Spectra from the Ly α Forest Using the Horizon-AGN Suite of Simulations”. In: *Monthly Notices of the Royal Astronomical Society* 495, pp. 1825–1840. DOI: [10.1093/mnras/staa1242](https://doi.org/10.1093/mnras/staa1242) (cit. on p. 30).
- Chaboyer, B. (1, 1998). “The Age of the Universe.” In: *Physics Reports* 307, pp. 23–30. DOI: [10.1016/S0370-1573\(98\)00054-4](https://doi.org/10.1016/S0370-1573(98)00054-4) (cit. on p. 10).
- Chang, Tzu-Ching, Ue-Li Pen, Jeffrey B. Peterson, and Patrick McDonald (1, 2008). “Baryon Acoustic Oscillation Intensity Mapping of Dark Energy”. In: *Physical Review Letters* 100, p. 091303. DOI: [10.1103/PhysRevLett.100.091303](https://doi.org/10.1103/PhysRevLett.100.091303) (cit. on p. 66).
- Chaussidon, Edmond et al. (2, 2021). “Angular Clustering Properties of the DESI QSO Target Selection Using DR9 Legacy Imaging Surveys”. In: *Monthly Notices of the Royal Astronomical Society* 509.3, pp. 3904–3923. DOI: [10.1093/mnras/stab3252](https://doi.org/10.1093/mnras/stab3252) (cit. on p. 47).
- Chevallier, Michel and David Polarski (1, 2001). “Accelerating Universes with Scaling Dark Matter”. In: *International Journal of Modern Physics D* 10, pp. 213–223. DOI: [10.1142/S0218271801000822](https://doi.org/10.1142/S0218271801000822) (cit. on p. 12).
- Chuang, Chia-Hsun, Francisco-Shu Kitaura, Francisco Prada, Cheng Zhao, and Gustavo Yepes (2015). “EZmocks: Extending the Zel'dovich Approximation to Generate Mock Galaxy Catalogues with Accurate Clustering Statistics”. In: *Monthly Notices of the Royal Astronomical Society* 446.3, p. 2621. DOI: [10.1093/mnras/stu2301](https://doi.org/10.1093/mnras/stu2301) (cit. on p. 52).
- Cole, Shaun et al. (1, 2005). “The 2dF Galaxy Redshift Survey: Power-Spectrum Analysis of the Final Data Set and Cosmological Implications”. In: *Monthly Notices of the Royal Astronomical Society* 362, pp. 505–534. DOI: [10.1111/j.1365-2966.2005.09318.x](https://doi.org/10.1111/j.1365-2966.2005.09318.x) (cit. on p. 10).
- Collaboration, DES et al. (27, 2021). “Dark Energy Survey Year 3 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing” (cit. on p. 23).
- Collaboration, The LIGO Scientific and The Virgo Collaboration (16, 2017). “GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral”. In: *Physical Review Letters* 119.16, p. 161101. DOI: [10.1103/PhysRevLett.119.161101](https://doi.org/10.1103/PhysRevLett.119.161101) (cit. on p. 17).
- Collaboration, The LIGO Scientific and the Virgo Collaboration (11, 2016). “Observation of Gravitational Waves from a Binary Black Hole Merger”. In: *Physical Review Letters* 116.6, p. 061102. DOI: [10.1103/PhysRevLett.116.061102](https://doi.org/10.1103/PhysRevLett.116.061102) (cit. on p. 17).
- Croft, Rupert A. C., Alessandro Romeo, and R. Benton Metcalf (1, 2018). “Weak Lensing of the Lyman α Forest”. In: *Monthly Notices of the Royal Astronomical Society* 477, pp. 1814–1821. DOI: [10.1093/mnras/sty650](https://doi.org/10.1093/mnras/sty650) (cit. on p. 41).
- Croft, Rupert A. C., David H. Weinberg, Max Pettini, Lars Hernquist, and Neal Katz (1, 1999). “The Power Spectrum of Mass Fluctuations Measured from the LYalpha Forest at Redshift Z=2.5”. In: *The Astrophysical Journal* 520, pp. 1–23. DOI: [10.1086/307438](https://doi.org/10.1086/307438); (cit. on p. 30).
- Cuceu, Andrei, Andreu Font-Ribera, Benjamin Joachimi, and Seshadri Nadathur (25, 2021). “Cosmology beyond BAO from the 3D Distribution of the Lyman- α Forest” (cit. on p. 41).
- Davis, Marc, Adi Nusser, Karen L. Masters, Christopher Springob, John P. Huchra, and Gerard Lemson (1, 2011). “Local Gravity versus Local Velocity: Solutions for β and Non-Linear Bias”. In: *Monthly Notices of the Royal Astronomical Society* 413, pp. 2906–2922. DOI: [10.1111/j.1365-2966.2011.18362.x](https://doi.org/10.1111/j.1365-2966.2011.18362.x) (cit. on pp. 78, 79).
- De Bernardis, P. et al. (1, 2000). “A Flat Universe from High-Resolution Maps of the Cosmic Microwave Background Radiation”. In: *Nature* 404, pp. 955–959. DOI: [10.1038/35010035](https://doi.org/10.1038/35010035) (cit. on p. 9).

Bibliography

- Delubac, Timothée et al. (1, 2015). “Baryon Acoustic Oscillations in the Ly α Forest of BOSS DR11 Quasars”. In: *Astronomy and Astrophysics* 574, A59. DOI: [10.1051/0004-6361/201423969](https://doi.org/10.1051/0004-6361/201423969) (cit. on pp. 31, 35).
- De Mattia, Arnaud and Vanina Ruhlmann-Kleider (1, 2019). “Integral Constraints in Spectroscopic Surveys”. In: *Journal of Cosmology and Astroparticle Physics* 08, p. 036. DOI: [10.1088/1475-7516/2019/08/036](https://doi.org/10.1088/1475-7516/2019/08/036) (cit. on p. 49).
- De Mattia, Arnaud et al. (1, 2021). “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”. In: *Monthly Notices of the Royal Astronomical Society* 501, pp. 5616–5645. DOI: [10.1093/mnras/staa3891](https://doi.org/10.1093/mnras/staa3891) (cit. on pp. 22, 49).
- De Sainte Agathe, Victoria et al. (2019). “Baryon Acoustic Oscillations at $z = 2.34$ from the Correlations of Ly α Absorption in eBOSS DR14”. In: *Astronomy and Astrophysics* 629, A85. DOI: [10.1051/0004-6361/201935638](https://doi.org/10.1051/0004-6361/201935638) (cit. on p. 31).
- Djorgovski, S. and Marc Davis (1, 1987). “Fundamental Properties of Elliptical Galaxies”. In: *The Astrophysical Journal* 313, p. 59. DOI: [10.1086/164948](https://doi.org/10.1086/164948) (cit. on p. 71).
- Dodelson, Scott and Fabian Schmidt (18, 2020). *Modern Cosmology - 2nd Edition*. 2nd ed. Elsevier (cit. on p. 21).
- Doi, Mamoru et al. (1, 2010). “Photometric Response Functions of the Sloan Digital Sky Survey Imager”. In: *The Astronomical Journal* 139, pp. 1628–1648. DOI: [10.1088/0004-6256/139/4/1628](https://doi.org/10.1088/0004-6256/139/4/1628) (cit. on p. 24).
- Du Mas des Bourboux Helion, Hélion et al. (2020). “The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Ly α Forests”. In: *The Astrophysical Journal* 901.2, p. 153. DOI: [10.3847/1538-4357/abb085](https://doi.org/10.3847/1538-4357/abb085) (cit. on pp. 31, 35, 37).
- Du Mas des Bourboux, Hélion et al. (1, 2017). “Baryon Acoustic Oscillations from the Complete SDSS-III Ly α -Quasar Cross-Correlation Function at $z = 2.4$ ”. In: *Astronomy and Astrophysics* 608, A130. DOI: [10.1051/0004-6361/201731731](https://doi.org/10.1051/0004-6361/201731731) (cit. on p. 31).
- Dupuy, Alexandra, Helene M Courtois, and Bogna Kubik (11, 2019). “An Estimation of the Local Growth Rate from Cosmicflows Peculiar Velocities”. In: *Monthly Notices of the Royal Astronomical Society* 486.1, pp. 440–448. DOI: [10.1093/mnras/stz901](https://doi.org/10.1093/mnras/stz901) (cit. on pp. 77, 79).
- Dutcher, D. et al. (13, 2021). “Measurements of the E-Mode Polarization and Temperature-E-Mode Correlation of the CMB from SPT-3G 2018 Data”. In: *Physical Review D* 104.2, p. 022003. DOI: [10.1103/PhysRevD.104.022003](https://doi.org/10.1103/PhysRevD.104.022003) (cit. on p. 21).
- Efstathiou, G., W. J. Sutherland, and S. J. Maddox (1, 1990). “The Cosmological Constant and Cold Dark Matter”. In: *Nature* 348, pp. 705–707. DOI: [10.1038/348705a0](https://doi.org/10.1038/348705a0) (cit. on p. 9).
- Eisenstein, Daniel J., Hee-Jong Seo, Edwin Sirk, and David N. Spergel (1, 2007a). “Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak”. In: *The Astrophysical Journal* 664, pp. 675–679. DOI: [10.1086/518712](https://doi.org/10.1086/518712) (cit. on p. 51).
- Eisenstein, Daniel J., Hee-Jong Seo, and Martin White (1, 2007b). “On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter”. In: *The Astrophysical Journal* 664, pp. 660–674. DOI: [10.1086/518755](https://doi.org/10.1086/518755) (cit. on p. 50).
- Eisenstein, Daniel J. et al. (2005). “Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies”. In: *The Astrophysical Journal* 633.2, p. 560. DOI: [10.1086/466512](https://doi.org/10.1086/466512) (cit. on p. 10).

Bibliography

- Eisenstein, Daniel J. et al. (1, 2011). "SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems". In: *The Astronomical Journal* 142, p. 72. DOI: [10.1088/0004-6256/142/3/72](https://doi.org/10.1088/0004-6256/142/3/72); (cit. on p. 20).
- Faber, S. M. and R. E. Jackson (1, 1976). "Velocity Dispersions and Mass-to-Light Ratios for Elliptical Galaxies." In: *The Astrophysical Journal* 204, pp. 668–683. DOI: [10.1086/154215](https://doi.org/10.1086/154215) (cit. on p. 71).
- Faucher-Giguère, Claude-André, Jason X. Prochaska, Adam Lidz, Lars Hernquist, and Matias Zaldarriaga (1, 2008). "A Direct Precision Measurement of the Intergalactic Ly α Opacity at $2 \leq z \leq 4.2$ ". In: *The Astrophysical Journal* 681, pp. 831–855. DOI: [10.1086/588648](https://doi.org/10.1086/588648) (cit. on p. 32).
- Feldman, Hume A., Nick Kaiser, and John A. Peacock (1994). "Power-Spectrum Analysis of Three-dimensional Redshift Surveys". In: *The Astrophysical Journal* 426, p. 23. DOI: [10.1086/174036](https://doi.org/10.1086/174036) (cit. on pp. 50, 53).
- Fixsen, D. J. (1, 2009). "The Temperature of the Cosmic Microwave Background". In: *The Astrophysical Journal* 707, pp. 916–920. DOI: [10.1088/0004-637X/707/2/916](https://doi.org/10.1088/0004-637X/707/2/916) (cit. on p. 20).
- Font-Ribera, Andreu and Jordi Miralda-Escudé (2012). "The Effect of High Column Density Systems on the Measurement of the Lyman- α Forest Correlation Function". In: *Journal of Cosmology and Astro-Particle Physics* 2012.7, p. 028. DOI: [10.1088/1475-7516/2012/07/028](https://doi.org/10.1088/1475-7516/2012/07/028) (cit. on p. 35).
- Font-Ribera, Andreu et al. (28, 2012). "The Large-Scale Cross-Correlation of Damped Lyman Alpha Systems with the Lyman Alpha Forest: First Measurements from BOSS". In: *Journal of Cosmology and Astroparticle Physics* 2012.11, pp. 059–059. DOI: [10.1088/1475-7516/2012/11/059](https://doi.org/10.1088/1475-7516/2012/11/059) (cit. on p. 33).
- Font-Ribera, Andreu et al. (1, 2014). "Quasar-Lyman-\$\alpha\$ Forest Cross-Correlation from BOSS DR11: Baryon Acoustic Oscillations". In: *Journal of Cosmology and Astro-Particle Physics* 05, p. 027. DOI: [10.1088/1475-7516/2014/05/027](https://doi.org/10.1088/1475-7516/2014/05/027) (cit. on p. 31).
- Freedman, Wendy L. et al. (29, 2019). "The Carnegie-Chicago Hubble Program. VIII. An Independent Determination of the Hubble Constant Based on the Tip of the Red Giant Branch". In: *The Astrophysical Journal* 882.1, p. 34. DOI: [10.3847/1538-4357/ab2f73](https://doi.org/10.3847/1538-4357/ab2f73) (cit. on p. 17).
- Gil-Marín, Héctor et al. (1, 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". In: *Monthly Notices of the Royal Astronomical Society* 498, pp. 2492–2531. DOI: [10.1093/mnras/staa2455](https://doi.org/10.1093/mnras/staa2455) (cit. on pp. 22, 45, 57).
- Gorski, Krzysztof (1, 1988). "On the Pattern of Perturbations of the Hubble Flow". In: *The Astrophysical Journal Letters* 332, pp. L7–L11. DOI: [10.1086/185255](https://doi.org/10.1086/185255) (cit. on p. 76).
- Gorski, Krzysztof M., Marc Davis, Michael A. Strauss, Simon D. M. White, and Amos Yahil (1, 1989). "Cosmological Velocity Correlations - Observations and Model Predictions". In: *The Astrophysical Journal* 344, pp. 1–19. DOI: [10.1086/167771](https://doi.org/10.1086/167771) (cit. on p. 77).
- Graham, Matthew J. et al. (1, 2019). "The Zwicky Transient Facility: Science Objectives". In: *Publications of the Astronomical Society of the Pacific* 131, p. 078001. DOI: [10.1088/1538-3873/ab006c](https://doi.org/10.1088/1538-3873/ab006c) (cit. on pp. 18, 73).
- Graziani, R. et al. (1, 2019). "The Peculiar Velocity Field up to $z \sim 0.05$ by Forward-Modelling Cosmicflows-3 Data". In: *Monthly Notices of the Royal Astronomical Society* 488, pp. 5438–5451. DOI: [10.1093/mnras/stz078](https://doi.org/10.1093/mnras/stz078) (cit. on pp. 75, 79).

Bibliography

- Graziani, R. et al. (24, 2020). “Peculiar Velocity Cosmology with Type Ia Supernovae” (cit. on p. 18).
- Gunn, J. E. et al. (1, 1998). “The Sloan Digital Sky Survey Photometric Camera”. In: *The Astronomical Journal* 116, pp. 3040–3081. DOI: [10.1086/300645](https://doi.org/10.1086/300645) (cit. on p. 24).
- Gunn, James E. and Bruce A. Peterson (1, 1965). “On the Density of Neutral Hydrogen in Intergalactic Space.” In: *The Astrophysical Journal* 142, pp. 1633–1641. DOI: [10.1086/148444](https://doi.org/10.1086/148444) (cit. on p. 30).
- Guo, Hong, Idit Zehavi, and Zheng Zheng (2012). “A New Method to Correct for Fiber Collisions in Galaxy Two-point Statistics”. In: *The Astrophysical Journal* 756.2, p. 127. DOI: [10.1088/0004-637X/756/2/127](https://doi.org/10.1088/0004-637X/756/2/127) (cit. on p. 47).
- Hand, Nick, Yin Li, Zachary Slepian, and Uros Seljak (3, 2017). “An Optimal FFT-based Anisotropic Power Spectrum Estimator”. In: *Journal of Cosmology and Astroparticle Physics* 2017.07, pp. 002–002. DOI: [10.1088/1475-7516/2017/07/002](https://doi.org/10.1088/1475-7516/2017/07/002) (cit. on p. 53).
- Herrnstein, J. R. et al. (1999). “A Geometric Distance to the Galaxy NGC4258 from Orbital Motions in a Nuclear Gas Disk”. In: *Nature* 400.6744 (6744), pp. 539–541. DOI: [10.1038/22972](https://doi.org/10.1038/22972) (cit. on p. 17).
- Heymans, Catherine et al. (2021). “KiDS-1000 Cosmology: Multi-probe Weak Gravitational Lensing and Spectroscopic Galaxy Clustering Constraints”. In: *Astronomy & Astrophysics* 646, A140. DOI: [10.1051/0004-6361/202039063](https://doi.org/10.1051/0004-6361/202039063) (cit. on p. 23).
- Hikage, Chiaki et al. (1, 2019). “Cosmology from Cosmic Shear Power Spectra with Subaru Hyper Suprime-Cam First-Year Data”. In: *Publications of the Astronomical Society of Japan* 71, p. 43. DOI: [10.1093/pasj/psz010](https://doi.org/10.1093/pasj/psz010) (cit. on p. 23).
- Hoffman, Yehuda, Adi Nusser, Aurélien Valade, Noam I. Libeskind, and R. Brent Tully (1, 2021). “From Cosmicflows Distance Moduli to Unbiased Distances and Peculiar Velocities”. In: *Monthly Notices of the Royal Astronomical Society* 505, pp. 3380–3392. DOI: [10.1093/mnras/stab1457](https://doi.org/10.1093/mnras/stab1457) (cit. on p. 70).
- Horne, K. (1, 1986). “An Optimal Extraction Algorithm for CCD Spectroscopy.” In: *Publications of the Astronomical Society of the Pacific* 98, pp. 609–617. DOI: [10.1086/131801](https://doi.org/10.1086/131801) (cit. on p. 27).
- Hou, Jiamin et al. (1, 2021). “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”. In: *Monthly Notices of the Royal Astronomical Society* 500, pp. 1201–1221. DOI: [10.1093/mnras/staa3234](https://doi.org/10.1093/mnras/staa3234) (cit. on p. 22).
- Howlett, Cullan (1, 2019). “The Redshift-Space Momentum Power Spectrum - I. Optimal Estimation from Peculiar Velocity Surveys”. In: *Monthly Notices of the Royal Astronomical Society* 487, pp. 5209–5234. DOI: [10.1093/mnras/stz1403](https://doi.org/10.1093/mnras/stz1403) (cit. on p. 77).
- Howlett, Cullan, Aaron S. G. Robotham, Claudia D. P. Lagos, and Alex G. Kim (1, 2017a). “Measuring the Growth Rate of Structure with Type IA Supernovae from LSST”. In: *The Astrophysical Journal* 847, p. 128. DOI: [10.3847/1538-4357/aa88c8](https://doi.org/10.3847/1538-4357/aa88c8) (cit. on p. 80).
- Howlett, Cullan, Ashley J. Ross, Lado Samushia, Will J. Percival, and Marc Manera (1, 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 $z = 0.15$ ”. In: *Monthly Notices of the Royal Astronomical Society* 449, pp. 848–866. DOI: [10.1093/mnras/stu2693](https://doi.org/10.1093/mnras/stu2693) (cit. on p. 22).

- Howlett, Cullan et al. (1, 2017b). “2MTF – VI. Measuring the Velocity Power Spectrum”. In: *Monthly Notices of the Royal Astronomical Society* 471.3, pp. 3135–3151. DOI: [10.1093/mnras/stx1521](https://doi.org/10.1093/mnras/stx1521) (cit. on p. 79).
- Howlett, Cullan et al. (1, 2022). “The Sloan Digital Sky Survey Peculiar Velocity Catalogue”. In: *Monthly Notices of the Royal Astronomical Society* 515, pp. 953–976. DOI: [10.1093/mnras/stac1681](https://doi.org/10.1093/mnras/stac1681) (cit. on pp. 71, 72).
- Hubble, Edwin (15, 1929). “A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae”. In: *Proceedings of the National Academy of Sciences* 15.3, pp. 168–173. DOI: [10.1073/pnas.15.3.168](https://doi.org/10.1073/pnas.15.3.168) (cit. on p. 16).
- Huterer, Dragan, Daniel L. Shafer, Daniel M. Scolnic, and Fabian Schmidt (1, 2017). “Testing LCDM at the Lowest Redshifts with SN Ia and Galaxy Velocities”. In: *Journal of Cosmology and Astroparticle Physics* 05, p. 015. DOI: [10.1088/1475-7516/2017/05/015](https://doi.org/10.1088/1475-7516/2017/05/015) (cit. on pp. 76, 79).
- Johnson, Andrew et al. (11, 2014). “The 6dF Galaxy Survey: Cosmological Constraints from the Velocity Power Spectrum”. In: *Monthly Notices of the Royal Astronomical Society* 444.4, pp. 3926–3947. DOI: [10.1093/mnras/stu1615](https://doi.org/10.1093/mnras/stu1615) (cit. on pp. 76, 79).
- Kaiser, Nick (1, 1987). “Clustering in Real Space and in Redshift Space”. In: *Monthly Notices of the Royal Astronomical Society* 227, pp. 1–21. DOI: [10.1093/mnras/227.1.1](https://doi.org/10.1093/mnras/227.1.1) (cit. on p. 22).
- Kamble, Vikrant, Kyle Dawson, Hélon du Mas des Bourboux, Julian Bautista, and Donald P. Scheinder (1, 2020). “Measurements of Effective Optical Depth in the Ly α Forest from the BOSS DR12 Quasar Sample”. In: *The Astrophysical Journal* 892, p. 70. DOI: [10.3847/1538-4357/ab76bd](https://doi.org/10.3847/1538-4357/ab76bd) (cit. on p. 32).
- Kessler, R et al. (1, 2019). “First Cosmology Results Using Type Ia Supernova from the Dark Energy Survey: Simulations to Correct Supernova Distance Biases”. In: *Monthly Notices of the Royal Astronomical Society* 485.1, pp. 1171–1187. DOI: [10.1093/mnras/stz463](https://doi.org/10.1093/mnras/stz463) (cit. on p. 18).
- Kessler, R. and D. Scolnic (1, 2017). “Correcting Type Ia Supernova Distances for Selection Biases and Contamination in Photometrically Identified Samples”. In: *The Astrophysical Journal* 836, p. 56. DOI: [10.3847/1538-4357/836/1/56](https://doi.org/10.3847/1538-4357/836/1/56) (cit. on p. 75).
- Kim, Alex G. and Eric V. Linder (1, 2020). “Complementarity of Peculiar Velocity Surveys and Redshift Space Distortions for Testing Gravity”. In: *Physical Review D* 101, p. 023516. DOI: [10.1103/PhysRevD.101.023516](https://doi.org/10.1103/PhysRevD.101.023516) (cit. on p. 18).
- Kirkby, David et al. (2013). “Fitting Methods for Baryon Acoustic Oscillations in the Lyman-\$\alpha\$ Forest Fluctuations in BOSS Data Release 9”. In: *Journal of Cosmology and Astroparticle Physics* 2013.3, p. 024. DOI: [10.1088/1475-7516/2013/03/024](https://doi.org/10.1088/1475-7516/2013/03/024) (cit. on p. 31).
- Kitaura, Francisco-Shu, Raul E. Angulo, Yehuda Hoffman, and Stefan Gottlöber (1, 2012). “Estimating Cosmic Velocity Fields from Density Fields and Tidal Tensors”. In: *Monthly Notices of the Royal Astronomical Society* 425, pp. 2422–2435. DOI: [10.1111/j.1365-2966.2012.21589.x](https://doi.org/10.1111/j.1365-2966.2012.21589.x) (cit. on p. 51).
- Koda, Jun et al. (1, 2014). “Are Peculiar Velocity Surveys Competitive as a Cosmological Probe?” In: *Monthly Notices of the Royal Astronomical Society* 445, pp. 4267–4286. DOI: [10.1093/mnras/stu1610](https://doi.org/10.1093/mnras/stu1610) (cit. on p. 81).
- Kourkchi, Ehsan, R. Brent Tully, Helene M. Courtois, Alexandra Dupuy, and Daniel Guinet (31, 2022). “Cosmicflows-4: The Baryonic Tully-Fisher Relation Providing ~10,000 Distances” (cit. on p. 70).

Bibliography

- Kunz, Martin, Bruce A. Bassett, and Renée A. Hlozek (18, 2007). “Bayesian Estimation Applied to Multiple Species”. In: *Physical Review D* 75.10, p. 103508. DOI: [10.1103/PhysRevD.75.103508](https://doi.org/10.1103/PhysRevD.75.103508) (cit. on p. 75).
- Landy, Stephen D. and Alexander S. Szalay (1993). “Bias and Variance of Angular Correlation Functions”. In: *The Astrophysical Journal* 412, p. 64. DOI: [10.1086/172900](https://doi.org/10.1086/172900) (cit. on p. 52).
- Lavaux, Guilhem (1, 2016). “Bayesian 3D Velocity Field Reconstruction with VIRBIUS”. In: *Monthly Notices of the Royal Astronomical Society* 457, pp. 172–197. DOI: [10.1093/mnras/stv2915](https://doi.org/10.1093/mnras/stv2915) (cit. on p. 79).
- Lee, Khee-Gan, Nao Suzuki, and David N. Spergel (1, 2012). “Mean-Flux-Regulated Principal Component Analysis Continuum Fitting of Sloan Digital Sky Survey Ly α Forest Spectra”. In: *The Astronomical Journal* 143, p. 51. DOI: [10.1088/0004-6256/143/2/51](https://doi.org/10.1088/0004-6256/143/2/51) (cit. on p. 32).
- Lesgourgues, Julien (14, 2011). “The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview”. In: (cit. on p. 21).
- Lewis, Antony, Anthony Challinor, and Anthony Lasenby (1, 2000). “Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models”. In: *The Astrophysical Journal* 538, pp. 473–476. DOI: [10.1086/309179](https://doi.org/10.1086/309179) (cit. on pp. 21, 35).
- Linder, Eric V. (1, 2003). “Exploring the Expansion History of the Universe”. In: *Physical Review Letters* 90, p. 091301. DOI: [10.1103/PhysRevLett.90.091301](https://doi.org/10.1103/PhysRevLett.90.091301) (cit. on p. 12).
- Lupton, R., J. E. Gunn, Z. Ivezić, G. R. Knapp, and S. Kent (1, 2001). “The SDSS Imaging Pipelines”. In: 238, p. 269 (cit. on p. 24).
- Lyke, Brad W. et al. (1, 2020). “The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release”. In: *The Astrophysical Journal Supplement Series* 250, p. 8. DOI: [10.3847/1538-4365/aba623](https://doi.org/10.3847/1538-4365/aba623) (cit. on p. 38).
- Maddox, S. J., G. Efstathiou, W. J. Sutherland, and J. Loveday (1, 1990). “Galaxy Correlations on Large Scales.” In: *Monthly Notices of the Royal Astronomical Society* 242, p. 43. DOI: [10.1093/mnras/242.1.43P](https://doi.org/10.1093/mnras/242.1.43P) (cit. on p. 9).
- Masui, K. W. et al. (2013). “Measurement of 21 Cm Brightness Fluctuations at z 0.8 in Cross-correlation”. In: *The Astrophysical Journal Letters* 763.1, p. L20. DOI: [10.1088/2041-8205/763/1/L20](https://doi.org/10.1088/2041-8205/763/1/L20) (cit. on p. 66).
- Mather, J. C. et al. (1, 1994). “Measurement of the Cosmic Microwave Background Spectrum by the COBE FIRAS Instrument”. In: *The Astrophysical Journal* 420, pp. 439–444. DOI: [10.1086/173574](https://doi.org/10.1086/173574) (cit. on p. 20).
- McDonald, Patrick (1, 2003). “Toward a Measurement of the Cosmological Geometry at \$z \sim 2\$: Predicting Ly\$\alpha\$ Forest Correlation in Three Dimensions and the Potential of Future Data Sets”. In: *The Astrophysical Journal* 585, pp. 34–51. DOI: [10.1086/345945](https://doi.org/10.1086/345945) (cit. on p. 30).
- McDonald, Patrick et al. (1, 2000). “The Observed Probability Distribution Function, Power Spectrum, and Correlation Function of the Transmitted Flux in the Ly\$\alpha\$ Forest”. In: *The Astrophysical Journal* 543, pp. 1–23. DOI: [10.1086/317079](https://doi.org/10.1086/317079) (cit. on p. 30).
- McDonald, Patrick et al. (1, 2005). “The Linear Theory Power Spectrum from the Ly\$\alpha\$ Forest in the Sloan Digital Sky Survey”. In: *The Astrophysical Journal* 635, pp. 761–783. DOI: [10.1086/497563](https://doi.org/10.1086/497563) (cit. on p. 30).
- McDonald, Patrick et al. (1, 2006). “The Ly\$\alpha\$ Forest Power Spectrum from the Sloan Digital Sky Survey”. In: *The Astrophysical Journal Supplement Series* 163, pp. 80–109. DOI: [10.1086/444361](https://doi.org/10.1086/444361); (cit. on p. 30).

Bibliography

- McGaugh, S. S., J. M. Schombert, G. D. Bothun, and W. J. G. de Blok (1, 2000). “The Baryonic Tully-Fisher Relation”. In: *The Astrophysical Journal* 533, pp. L99–L102. DOI: [10.1086/312628](https://doi.org/10.1086/312628) (cit. on p. 70).
- McQuinn, Matthew, Oliver Zahn, Matias Zaldarriaga, Lars Hernquist, and Steven R. Furlanetto (1, 2006). “Cosmological Parameter Estimation Using 21 Cm Radiation from the Epoch of Reionization”. In: *The Astrophysical Journal* 653, pp. 815–834. DOI: [10.1086/505167](https://doi.org/10.1086/505167) (cit. on p. 66).
- Mesinger, Andrei, Steven Furlanetto, and Renyue Cen (2011). “21CMFAST: A Fast, Seminumerical Simulation of the High-Redshift 21-Cm Signal”. In: *Monthly Notices of the Royal Astronomical Society* 411.2, p. 955. DOI: [10.1111/j.1365-2966.2010.17731.x](https://doi.org/10.1111/j.1365-2966.2010.17731.x) (cit. on p. 66).
- Metcalf, R. Benton, Rupert A. C. Croft, and Alessandro Romeo (2018). “Noise Estimates for Measurements of Weak Lensing from the Ly α Forest”. In: *Monthly Notices of the Royal Astronomical Society* 477.2, p. 2841. DOI: [10.1093/mnras/sty806](https://doi.org/10.1093/mnras/sty806) (cit. on p. 41).
- Metcalf, R. Benton, Nicolas Tessore, and Rupert A. C. Croft (2020). “Reconstructing the Gravitational Lensing Potential from the Lyman- α Forest”. In: *Astronomy and Astrophysics* 642, A122. DOI: [10.1051/0004-6361/202038056](https://doi.org/10.1051/0004-6361/202038056) (cit. on p. 41).
- Mohammad, Faizan G. et al. (1, 2020). “The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Pairwise-Inverse Probability and Angular Correction for Fibre Collisions in Clustering Measurements”. In: *Monthly Notices of the Royal Astronomical Society* 498, pp. 128–143. DOI: [10.1093/mnras/staa2344](https://doi.org/10.1093/mnras/staa2344) (cit. on p. 53).
- Mould, Jeremy R. et al. (1, 2000). “The Hubble Space Telescope Key Project on the Extragalactic Distance Scale. XXVIII. Combining the Constraints on the Hubble Constant”. In: *The Astrophysical Journal* 529, pp. 786–794. DOI: [10.1086/308304](https://doi.org/10.1086/308304) (cit. on p. 9).
- Mueller, Eva-Maria et al. (25, 2021). “The Clustering of Galaxies in the Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Primordial Non-Gaussianity in Fourier Space” (cit. on p. 47).
- Myers, Adam D. et al. (2, 2015). “The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Quasar Target Selection”. In: *The Astrophysical Journal Supplement Series* 221.2, p. 27. DOI: [10.1088/0067-0049/221/2/27](https://doi.org/10.1088/0067-0049/221/2/27) (cit. on p. 25).
- Neveux, Richard et al. (1, 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”. In: *Monthly Notices of the Royal Astronomical Society* 499, pp. 210–229. DOI: [10.1093/mnras/staa2780](https://doi.org/10.1093/mnras/staa2780) (cit. on p. 22).
- Nusser, Adi (1, 2017). “Velocity-Density Correlations from the Cosmicflows-3 Distance Catalogue and the 2MASS Redshift Survey”. In: *Monthly Notices of the Royal Astronomical Society* 470, pp. 445–454. DOI: [10.1093/mnras/stx1225](https://doi.org/10.1093/mnras/stx1225) (cit. on p. 79).
- Nusser, Adi and Enzo Branchini (1, 2000). “On the Least Action Principle in Cosmology”. In: *Monthly Notices of the Royal Astronomical Society* 313, pp. 587–595. DOI: [10.1046/j.1365-8711.2000.03261.x](https://doi.org/10.1046/j.1365-8711.2000.03261.x) (cit. on p. 51).
- Nusser, Adi and Avishai Dekel (1, 1992). “Tracing Large-Scale Fluctuations Back in Time”. In: *The Astrophysical Journal* 391, p. 443. DOI: [10.1086/171360](https://doi.org/10.1086/171360) (cit. on p. 51).
- Okumura, Teppei et al. (1, 2016). “The Subaru FMOS Galaxy Redshift Survey (FastSound). IV. New Constraint on Gravity Theory from Redshift Space Distortions at $z \sim 1.4$ ”. In: *Publications of the Astronomical Society of Japan* 68.3. DOI: [10.1093/pasj/psw029](https://doi.org/10.1093/pasj/psw029) (cit. on p. 20).

Bibliography

- Padmanabhan, Nikhil, Martin White, and J. D. Cohn (1, 2009). “Reconstructing Baryon Oscillations: A Lagrangian Theory Perspective”. In: *Physical Review D* 79, p. 063523. DOI: [10.1103/PhysRevD.79.063523](https://doi.org/10.1103/PhysRevD.79.063523) (cit. on p. 53).
- Padmanabhan, Nikhil et al. (1, 2008). “An Improved Photometric Calibration of the Sloan Digital Sky Survey Imaging Data”. In: *The Astrophysical Journal* 674, pp. 1217–1233. DOI: [10.1086/524677](https://doi.org/10.1086/524677) (cit. on p. 24).
- Padmanabhan, Nikhil et al. (2012). “A 2 per Cent Distance to $z = 0.35$ by Reconstructing Baryon Acoustic Oscillations - I. Methods and Application to the Sloan Digital Sky Survey”. In: *Monthly Notices of the Royal Astronomical Society* 427.3, p. 2132. DOI: [10.1111/j.1365-2966.2012.21888.x](https://doi.org/10.1111/j.1365-2966.2012.21888.x) (cit. on pp. 51, 53).
- Palanque-Delabrouille, N. et al. (1, 2016). “The Extended Baryon Oscillation Spectroscopic Survey: Variability Selection and Quasar Luminosity Function”. In: *Astronomy and Astrophysics* 587, A41. DOI: [10.1051/0004-6361/201527392](https://doi.org/10.1051/0004-6361/201527392) (cit. on p. 25).
- Palanque-Delabrouille, Nathalie et al. (2020). “Hints, Neutrino Bounds, and WDM Constraints from SDSS DR14 Lyman-\$\alpha\$ and Planck Full-Survey Data”. In: *Journal of Cosmology and Astroparticle Physics* 2020.04, pp. 038–038. DOI: [10.1088/1475-7516/2020/04/038](https://doi.org/10.1088/1475-7516/2020/04/038) (cit. on p. 30).
- Pâris, I. et al. (2011). “A Principal Component Analysis of Quasar UV Spectra at $z \sim 3$ ”. In: *Astronomy and Astrophysics* 530, A50. DOI: [10.1051/0004-6361/201016233](https://doi.org/10.1051/0004-6361/201016233) (cit. on p. 32).
- Parkinson, David et al. (2012). “The WiggleZ Dark Energy Survey: Final Data Release and Cosmological Results”. In: *Physical Review D* 86.10, p. 103518. DOI: [10.1103/PhysRevD.86.103518](https://doi.org/10.1103/PhysRevD.86.103518) (cit. on p. 20).
- Paviot, Romain et al. (1, 2022). “Angular Systematics-Free Cosmological Analysis of Galaxy Clustering in Configuration Space”. In: *Monthly Notices of the Royal Astronomical Society* 512, pp. 1341–1356. DOI: [10.1093/mnras/stac560](https://doi.org/10.1093/mnras/stac560) (cit. on p. 34).
- Peebles, P. J. E. (1, 1989). “Tracing Galaxy Orbits Back in Time”. In: *The Astrophysical Journal* 344, p. L53. DOI: [10.1086/185529](https://doi.org/10.1086/185529) (cit. on p. 51).
- Percival, Will J. and Davide Bianchi (2017). “Using Angular Pair Upweighting to Improve 3D Clustering Measurements”. In: *Monthly Notices of the Royal Astronomical Society: Letters* 472.1, pp. L40–L44. DOI: [10.1093/mnrasl/slx135](https://doi.org/10.1093/mnrasl/slx135) (cit. on p. 53).
- Perlmutter, S. et al. (1, 1999). “Measurements of Omega and Lambda from 42 High-Redshift Supernovae”. In: *The Astrophysical Journal* 517, pp. 565–586. DOI: [10.1086/307221](https://doi.org/10.1086/307221) (cit. on p. 9).
- Peterson, Erik R. et al. (7, 2021). “The Pantheon+ Analysis: Evaluating Peculiar Velocity Corrections in Cosmological Analyses with Nearby Type Ia Supernovae” (cit. on p. 73).
- Pezzotta, A. et al. (2017). “The VIMOS Public Extragalactic Redshift Survey (VIPERS). The Growth of Structure at $0.5 < z < 1.2$ from Redshift-Space Distortions in the Clustering of the PDR-2 Final Sample”. In: *Astronomy and Astrophysics* 604, A33. DOI: [10.1051/0004-6361/201630295](https://doi.org/10.1051/0004-6361/201630295) (cit. on p. 20).
- Pitrou, Cyril, Alain Coc, Jean-Philippe Uzan, and Elisabeth Vangioni (10, 2021). “A New Tension in the Cosmological Model from Primordial Deuterium?” In: *Monthly Notices of the Royal Astronomical Society* 502.2, pp. 2474–2481. DOI: [10.1093/mnras/stab135](https://doi.org/10.1093/mnras/stab135) (cit. on p. 19).
- Planck Collaboration et al. (1, 2020). “Planck 2018 Results. VI. Cosmological Parameters”. In: *Astronomy and Astrophysics* 641, A6. DOI: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910) (cit. on p. 21).
- Pourtsidou, A. and R. Benton Metcalf (1, 2014). “Weak Lensing with 21 Cm Intensity Mapping at $z \sim 2$ –3”. In: *Monthly Notices of the Royal Astronomical Society* 439, pp. L36–L40. DOI: [10.1093/mnrasl/slt175](https://doi.org/10.1093/mnrasl/slt175) (cit. on p. 41).

Bibliography

- Pourtsidou, A. and R. Benton Metcalf (1, 2015). “Gravitational Lensing of Cosmological 21 Cm Emission”. In: *Monthly Notices of the Royal Astronomical Society* 448, pp. 2368–2383. DOI: [10.1093/mnras/stv102](https://doi.org/10.1093/mnras/stv102) (cit. on p. 41).
- Prakash, Abhishek et al. (1, 2016). “The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Luminous Red Galaxy Target Selection”. In: *The Astrophysical Journal Supplement Series* 224, p. 34. DOI: [10.3847/0067-0049/224/2/34](https://doi.org/10.3847/0067-0049/224/2/34) (cit. on pp. 25, 47).
- Prideaux-Ghee, James, Florent Leclercq, Guilhem Lavaux, Alan Heavens, and Jens Jache (31, 2022). “Field-Based Physical Inference From Peculiar Velocity Tracers” (cit. on p. 79).
- Qin, Fei, Cullan Howlett, and Lister Staveley-Smith (1, 2019). “The Redshift-Space Momentum Power Spectrum - II. Measuring the Growth Rate from the Combined 2MTF and 6dFGSV Surveys”. In: *Monthly Notices of the Royal Astronomical Society* 487, pp. 5235–5247. DOI: [10.1093/mnras/stz1576](https://doi.org/10.1093/mnras/stz1576) (cit. on pp. 77, 79).
- Raichoor, A. et al. (1, 2017). “The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Final Emission Line Galaxy Target Selection”. In: *Monthly Notices of the Royal Astronomical Society* 471, pp. 3955–3973. DOI: [10.1093/mnras/stx1790](https://doi.org/10.1093/mnras/stx1790) (cit. on p. 25).
- Raichoor, Anand et al. (11, 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”. In: *Monthly Notices of the Royal Astronomical Society* 500.3, pp. 3254–3274. DOI: [10.1093/mnras/staa3336](https://doi.org/10.1093/mnras/staa3336) (cit. on p. 66).
- Ravoux, Corentin et al. (21, 2022). “First Measurement of the Correlation between Cosmic Voids and the Lyman-\$\alpha\$ Forest” (cit. on p. 44).
- Reid, Beth et al. (1, 2016). “SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: Galaxy Target Selection and Large-Scale Structure Catalogues”. In: *Monthly Notices of the Royal Astronomical Society* 455, pp. 1553–1573. DOI: [10.1093/mnras/stv2382](https://doi.org/10.1093/mnras/stv2382) (cit. on p. 25).
- Reid, M. J., D. W. Pesce, and A. G. Riess (2019). “An Improved Distance to NGC 4258 and Its Implications for the Hubble Constant”. In: *The Astrophysical Journal* 886.2, p. L27. DOI: [10.3847/2041-8213/ab552d](https://doi.org/10.3847/2041-8213/ab552d) (cit. on p. 17).
- Rezaie, Mehdi, Hee-Jong Seo, Ashley J. Ross, and Razvan C. Bunescu (1, 2020). “Improving Galaxy Clustering Measurements with Deep Learning: Analysis of the DECaLS DR7 Data”. In: *Monthly Notices of the Royal Astronomical Society* 495, pp. 1613–1640. DOI: [10.1093/mnras/staa1231](https://doi.org/10.1093/mnras/staa1231) (cit. on p. 47).
- Rezaie, Mehdi et al. (28, 2021). “Primordial Non-Gaussianity from the Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey I: Catalogue Preparation and Systematic Mitigation”. In: *Monthly Notices of the Royal Astronomical Society* 506.3, pp. 3439–3454. DOI: [10.1093/mnras/stab1730](https://doi.org/10.1093/mnras/stab1730) (cit. on p. 47).
- Riess, Adam G. (2020). “The Expansion of the Universe Is Faster than Expected”. In: *Nature Reviews Physics* 2.1, pp. 10–12. DOI: [10.1038/s42254-019-0137-0](https://doi.org/10.1038/s42254-019-0137-0) (cit. on p. 16).
- Riess, Adam G. et al. (1, 1998). “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant”. In: *The Astronomical Journal* 116, pp. 1009–1038. DOI: [10.1086/300499](https://doi.org/10.1086/300499) (cit. on p. 9).
- Riess, Adam G. et al. (2021). “Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with \$\Lambda\$”. In: 908.1, p. L6. DOI: [10.3847/2041-8213/abdbaf](https://doi.org/10.3847/2041-8213/abdbaf) (cit. on p. 16).

Bibliography

- Rogers, Keir K., Simeon Bird, Hiranya V. Peiris, Andrew Pontzen, Andreu Font-Ribera, and Boris Leistedt (2018). “Correlations in the Three-Dimensional Lyman-alpha Forest Contaminated by High Column Density Absorbers”. In: *Monthly Notices of the Royal Astronomical Society* 476.3, p. 3716. DOI: [10.1093/mnras/sty603](https://doi.org/10.1093/mnras/sty603) (cit. on p. 35).
- Ross, Ashley J. et al. (1, 2020). “The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Large-scale Structure Catalogues for Cosmological Analysis”. In: *Monthly Notices of the Royal Astronomical Society* 498, pp. 2354–2371. DOI: [10.1093/mnras/staa2416](https://doi.org/10.1093/mnras/staa2416) (cit. on pp. 45, 48).
- Ross, Nicholas P. et al. (2012). “The SDSS-III Baryon Oscillation Spectroscopic Survey: Quasar Target Selection for Data Release Nine”. In: *The Astrophysical Journal Supplement Series* 199.1, p. 3. DOI: [10.1088/0067-0049/199/1/3](https://doi.org/10.1088/0067-0049/199/1/3) (cit. on p. 25).
- Said, Khaled, Matthew Colless, Christina Magoulas, John R. Lucey, and Michael J. Hudson (1, 2020). “Joint Analysis of 6dFGS and SDSS Peculiar Velocities for the Growth Rate of Cosmic Structure and Tests of Gravity”. In: *Monthly Notices of the Royal Astronomical Society* 497, pp. 1275–1293. DOI: [10.1093/mnras/staa2032](https://doi.org/10.1093/mnras/staa2032) (cit. on pp. 78, 79).
- Sánchez, Ariel G. (2, 2020). “Arguments against Using h^{-1}Mpc Units in Observational Cosmology”. In: *Physical Review D* 102.12, p. 123511. DOI: [10.1103/PhysRevD.102.123511](https://doi.org/10.1103/PhysRevD.102.123511) (cit. on pp. 16, 61).
- Sánchez, Ariel G. et al. (2017). “The Clustering of Galaxies in the Completed SDSS-III Baryon Oscillation Spectroscopic Survey: Combining Correlated Gaussian Posterior Distributions”. In: *Monthly Notices of the Royal Astronomical Society* 464.2, p. 1493. DOI: [10.1093/mnras/stw2495](https://doi.org/10.1093/mnras/stw2495) (cit. on pp. 62, 64).
- Sarpa, E., C. Schimd, E. Branchini, and S. Matarrese (1, 2019). “BAO Reconstruction: A Swift Numerical Action Method for Massive Spectroscopic Surveys”. In: *Monthly Notices of the Royal Astronomical Society* 484, pp. 3818–3830. DOI: [10.1093/mnras/stz278](https://doi.org/10.1093/mnras/stz278) (cit. on p. 51).
- Sarpa, E., A. Veropalumbo, C. Schimd, E. Branchini, and S. Matarrese (1, 2021). “Extended Fast Action Minimization Method: Application to SDSS-DR12 Combined Sample”. In: *Monthly Notices of the Royal Astronomical Society* 503, pp. 540–556. DOI: [10.1093/mnras/stab378](https://doi.org/10.1093/mnras/stab378) (cit. on pp. 51, 78).
- Scolnic, D. M. et al. (1, 2018). “The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample”. In: *The Astrophysical Journal* 859, p. 101. DOI: [10.3847/1538-4357/aab9bb](https://doi.org/10.3847/1538-4357/aab9bb) (cit. on p. 18).
- Scolnic, Dan et al. (1, 2021). “The Pantheon+ Analysis: The Full Dataset and Light-Curve Release”. In: *arXiv e-prints* (cit. on p. 73).
- Sefusatti, E., M. Crocce, R. Scoccimarro, and H. M. P. Couchman (1, 2016). “Accurate Estimators of Correlation Functions in Fourier Space”. In: *Monthly Notices of the Royal Astronomical Society* 460, pp. 3624–3636. DOI: [10.1093/mnras/stw1229](https://doi.org/10.1093/mnras/stw1229) (cit. on p. 54).
- Seo, Hee-Jong, Florian Beutler, Ashley J. Ross, and Shun Saito (11, 2016). “Modeling the Reconstructed BAO in Fourier Space”. In: *Monthly Notices of the Royal Astronomical Society* 460.3, pp. 2453–2471. DOI: [10.1093/mnras/stw1138](https://doi.org/10.1093/mnras/stw1138) (cit. on p. 57).
- Seo, Hee-Jong et al. (1, 2010). “High-Precision Predictions for the Acoustic Scale in the Non-linear Regime”. In: *The Astrophysical Journal* 720, pp. 1650–1667. DOI: [10.1088/0004-637X/720/2/1650](https://doi.org/10.1088/0004-637X/720/2/1650) (cit. on p. 51).
- Shen, Yue et al. (2016). “The Sloan Digital Sky Survey Reverberation Mapping Project: Velocity Shifts of Quasar Emission Lines”. In: *The Astrophysical Journal* 831.1, p. 7. DOI: [10.3847/0004-637X/831/1/7](https://doi.org/10.3847/0004-637X/831/1/7) (cit. on p. 38).

Bibliography

- Slosar, Anže et al. (2011). “The Lyman-\$\alpha\$ Forest in Three Dimensions: Measurements of Large Scale Flux Correlations from BOSS 1st-Year Data”. In: *Journal of Cosmology and Astroparticle Physics* 2011.9, p. 001. DOI: [10.1088/1475-7516/2011/09/001](https://doi.org/10.1088/1475-7516/2011/09/001) (cit. on p. 31).
- Slosar, Anže et al. (1, 2013). “Measurement of Baryon Acoustic Oscillations in the Lyman-\$\alpha\$ Forest Fluctuations in BOSS Data Release 9”. In: *Journal of Cosmology and Astroparticle Physics* 04, p. 026. DOI: [10.1088/1475-7516/2013/04/026](https://doi.org/10.1088/1475-7516/2013/04/026); (cit. on p. 31).
- Springob, Christopher M. et al. (11, 2014). “The 6dF Galaxy Survey: Peculiar Velocity Field and Cosmography”. In: *Monthly Notices of the Royal Astronomical Society* 445.3, pp. 2677–2697. DOI: [10.1093/mnras/stu1743](https://doi.org/10.1093/mnras/stu1743) (cit. on pp. 71, 76).
- Square Kilometre Array Cosmology Science Working Group et al. (1, 2020). “Cosmology with Phase 1 of the Square Kilometre Array Red Book 2018: Technical Specifications and Performance Forecasts”. In: *Publications of the Astronomical Society of Australia* 37, e007. DOI: [10.1017/pasa.2019.51](https://doi.org/10.1017/pasa.2019.51) (cit. on p. 68).
- Strauss, Michael A. and Jeffrey A. Willick (1, 1995). “The Density and Peculiar Velocity Fields of Nearby Galaxies”. In: *Physics Reports* 261.5, pp. 271–431. DOI: [10.1016/0370-1573\(95\)00013-7](https://doi.org/10.1016/0370-1573(95)00013-7) (cit. on p. 74).
- Strauss, Michael A. et al. (1, 2002). “Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample”. In: *The Astronomical Journal* 124, pp. 1810–1824. DOI: [10.1086/342343](https://doi.org/10.1086/342343) (cit. on p. 25).
- Tamone, Amélie et al. (1, 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”. In: *Monthly Notices of the Royal Astronomical Society*. DOI: [10.1093/mnras/staa3050](https://doi.org/10.1093/mnras/staa3050) (cit. on pp. 22, 49).
- Taruya, Atsushi, Francis Bernardeau, Takahiro Nishimichi, and Sandrine Codis (1, 2012). “Direct and Fast Calculation of Regularized Cosmological Power Spectrum at Two-Loop Order”. In: *Physical Review D* 86, p. 103528. DOI: [10.1103/PhysRevD.86.103528](https://doi.org/10.1103/PhysRevD.86.103528) (cit. on p. 59).
- Taruya, Atsushi, Takahiro Nishimichi, and Shun Saito (1, 2010). “Baryon Acoustic Oscillations in 2D: Modeling Redshift-Space Power Spectrum from Perturbation Theory”. In: *Physical Review D* 82, p. 063522. DOI: [10.1103/PhysRevD.82.063522](https://doi.org/10.1103/PhysRevD.82.063522) (cit. on p. 59).
- Tully, R. B. and J. R. Fisher (1977). “A New Method of Determining Distances to Galaxies.” In: *Astronomy and Astrophysics* 54, pp. 661–673 (cit. on p. 70).
- Turner, Ryan J., Chris Blake, and Rossana Ruggeri (22, 2021). “Improving Estimates of the Growth Rate Using Galaxy-Velocity Correlations: A Simulation Study” (cit. on p. 77).
- (8, 2022). *A Local Measurement of the Growth Rate from Peculiar Velocities and Galaxy Clustering Correlations in the 6dF Galaxy Survey*. DOI: [10.48550/arXiv.2207.03707](https://doi.org/10.48550/arXiv.2207.03707) (cit. on p. 77).
- Villaescusa-Navarro, Francisco et al. (2018). “Ingredients for 21 Cm Intensity Mapping”. In: *The Astrophysical Journal* 866.2, p. 135. DOI: [10.3847/1538-4357/aadba0](https://doi.org/10.3847/1538-4357/aadba0) (cit. on p. 66).
- Wang, Lile, Beth Reid, and Martin White (1, 2014). “An Analytic Model for Redshift-Space Distortions”. In: *Monthly Notices of the Royal Astronomical Society* 437, pp. 588–599. DOI: [10.1093/mnras/stt1916](https://doi.org/10.1093/mnras/stt1916) (cit. on p. 59).
- Wang, Yuyu, Christopher Rooney, Hume A Feldman, and Richard Watkins (11, 2018). “The Peculiar Velocity Correlation Function”. In: *Monthly Notices of the Royal Astronomical Society* 480.4, pp. 5332–5341. DOI: [10.1093/mnras/sty2224](https://doi.org/10.1093/mnras/sty2224) (cit. on p. 77).

Bibliography

- Weinberg, David H., Michael J. Mortonson, Daniel J. Eisenstein, Christopher Hirata, Adam G. Riess, and Eduardo Rozo (10, 2013). “Observational Probes of Cosmic Acceleration”. In: *Physics Reports* 530.2, pp. 87–255. DOI: [10.1016/j.physrep.2013.05.001](https://doi.org/10.1016/j.physrep.2013.05.001) (cit. on p. 12).
- Wolz, Laura et al. (1, 2022). “H I Constraints from the Cross-Correlation of eBOSS Galaxies and Green Bank Telescope Intensity Maps”. In: *Monthly Notices of the Royal Astronomical Society* 510, pp. 3495–3511. DOI: [10.1093/mnras/stab3621](https://doi.org/10.1093/mnras/stab3621) (cit. on pp. 66–68).
- Wong, Kenneth C. et al. (1, 2019). “H0LiCOW - XIII. A 2.4 per Cent Measurement of H0 from Lensed Quasars: 5.3sigma Tension between Early- and Late-Universe Probes”. In: *Monthly Notices of the Royal Astronomical Society* 498, pp. 1420–1439. DOI: [10.1093/mnras/stz3094](https://doi.org/10.1093/mnras/stz3094) (cit. on p. 17).
- York, Donald G. et al. (1, 2000). “The Sloan Digital Sky Survey: Technical Summary”. In: *The Astronomical Journal* 120, pp. 1579–1587. DOI: [10.1086/301513](https://doi.org/10.1086/301513) (cit. on p. 24).
- Youles, Samantha et al. (13, 2022). “The Effect of Quasar Redshift Errors on Lyman-\$\alpha\$ Forest Correlation Functions”. DOI: [10.48550/arXiv.2205.06648](https://arxiv.org/abs/2205.06648) (cit. on pp. 35, 38, 40).
- Zarrouk, Pauline et al. (1, 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”. In: *Monthly Notices of the Royal Astronomical Society* 477, pp. 1639–1663. DOI: [10.1093/mnras/sty506](https://doi.org/10.1093/mnras/sty506) (cit. on p. 38).
- Zel'dovich, Ya B. (1970). “Gravitational Instability: An Approximate Theory for Large Density Perturbations.” In: *Astronomy and Astrophysics* 5, pp. 84–89 (cit. on p. 51).
- Zhao, Cheng et al. (1, 2021). “The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: 1000 Multi-Tracer Mock Catalogues with Redshift Evolution and Systematics for Galaxies and Quasars of the Final Data Release”. In: *Monthly Notices of the Royal Astronomical Society* 503, pp. 1149–1173. DOI: [10.1093/mnras/stab510](https://doi.org/10.1093/mnras/stab510) (cit. on pp. 45, 46, 50, 52).