



**HAL**  
open science

# Baryon Acoustic Oscillations in the Large Scale Structures of the Universe as seen by the Sloan Digital Sky Survey

Julian Bautista

► **To cite this version:**

Julian Bautista. Baryon Acoustic Oscillations in the Large Scale Structures of the Universe as seen by the Sloan Digital Sky Survey. Cosmology and Extra-Galactic Astrophysics [astro-ph.CO]. Université Paris Diderot, 2014. English. NNT: . tel-01389967

**HAL Id: tel-01389967**

**<https://theses.hal.science/tel-01389967>**

Submitted on 4 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Ecole Doctorale Particules, Noyaux et Cosmos 517

UNIVERSITÉ PARIS DIDEROT SORBONNE PARIS CITE

Laboratoire Astroparticules Cosmologie

## THÈSE

pour obtenir le grade de

Docteur en Sciences

Spécialité : COSMOLOGIE

# Baryon Acoustic Oscillations in the Large Scale Structures of the Universe as seen by the Sloan Digital Sky Survey

par

Julián Ernesto BAUTISTA

Soutenue le 15 Septembre 2014.

### Jury :

<i>Président :</i>	M. Stavros KATZANEVAS	- Laboratoire AstroParticule et Cosmologie
<i>Rapporteurs :</i>	M. Alain BLANCHARD	- Institut de Recherche en Astrophysique et Planétologie
	M. Jean-Paul KNEIB	- Ecole Polytechnique Fédérale de Lausanne
<i>Examineurs :</i>	M. Julien GUY	- Lawrence Berkeley National Laboratory
	M. Jordi MIRALDA-ESCUDE	- Institut de Cincies del Cosmos Barcelona
	M. David WEINBERG	- Ohio State University
<i>Directeurs :</i>	M. Nicolás G. BUSCA	- Laboratoire Astroparticules et Cosmologie
	M. James RICH	- Centre Energie Atomique, Saclay









*“BAO is easy, it is just double for loop...”*

Jim Rich



# *Abstract*

This thesis exposes my contribution to the measurement of baryon acoustic oscillations (BAO) using Ly $\alpha$  forests and galaxies, with the cosmological interpretation of results.

The Baryon Oscillation Spectroscopic Survey (BOSS) - part of the Sloan Digital Sky Survey III - measured spectra of  $\sim 1.3$  million galaxies and  $\sim 150,000$   $z > 2$  quasars during four years of observations. For the galaxy BAO measurement, I contributed to the development of a new optimal estimator of the correlation function. Using Data Release 9 CMASS galaxy catalogs, we obtained errors on the BAO peak position estimate reduced by 20% compared to Landy-Szalay estimator.

My contribution to the BAO measurement in the Ly $\alpha$  forest covers from simulations, to data analysis and cosmological interpretation.

Concerning simulations, I developed a method to generate spatially correlated random fields dedicated to the production of mock Ly $\alpha$  forest absorption fields. This method has many advantages: computing time scaling linearly with the number of quasars, low memory requirements and no need to separate the survey into independent regions.

I created the *MockExpander* package, that converts an absorption field into a realistic set of quasar spectra. These spectra follow BOSS characteristics, including resolution, noise, sky residuals, mis-estimates of pixel errors. Astrophysical features such as metal absorption and high column density systems are also implemented. The BOSS collaboration currently uses the MockExpander to build mock catalogs.

These mock catalogs were extensively used to test our analysis chain. I quantified the possible sources of systematics, such as continuum fitting methods, sky residuals, errors on pixel noise, calibration features and metal absorption. No strong evidence was found for systematic errors on the BAO peak estimates using mock catalogs.

Using DR11 data, the measurement of the radial and transverse BAO peak position yield, respectively,  $c/r_d H(\bar{z}) = 9.18 \pm 0.28(1\sigma) \pm 0.6(2\sigma)$  and  $D_A(\bar{z})/r_d = 11.28 \pm 0.65(1\sigma)_{-1.2}^{+2.8}(2\sigma)$ , where  $r_d$  is the comoving sound horizon and  $\bar{z} = 2.34$  is the mean redshift of the measurement. Assuming  $r_d = 147.4$  Mpc, value derived from cosmic microwave background measurements, our results give  $H(\bar{z}) = 222 \pm 7(1\sigma) \text{ km s}^{-1} \text{ Mpc}^{-1}$  and the  $D_A(\bar{z}) = 1662 \pm 96(1\sigma) \text{ Mpc}$ . This pair of values is at  $1.8\sigma$  from the prediction given by a  $\Lambda$ CDM model with Planck best-fit parameters.



# *Acknowledgements*

It is time to thank some very important people. Without them, I would never have successfully accomplished this thesis. To better transmit my gratefulness, I have chosen to write these words in different languages.

First, I would like to thank Jean-Paul Kneib, Alain Blanchard and Julien Guy for reading my thesis, bringing some insightful comments. I thank David Weinberg for his presence in my jury. I thank Jordi, for being in my jury, for all the transmitted knowledge in this subject, and for trusting my capacities to work with his students.

En français, un merci à Stavros d'avoir présidé mon jury, en espérant que les années à venir en tant que directeur seront fructueuses. Merci à Pierre Binétruy pour les deux ans à la tête de l'APC.

Je tiens beaucoup à remercier JC, qui m'a permis de connaître l'APC, qui, avec son carisme, m'a toujours motivé à la recherche, qui m'a conduit pendant les premiers mois de thèse. Merci beaucoup Jim, d'avoir transmis toute ta passion par la cosmologie et ton savoir depuis les cours NPAC jusqu'à aujourd'hui. Merci d'être là quand Nicolas était loin. Je remercie Yannick, pour qui j'ai beaucoup de respect et qui a toujours trouvé du temps dans son agenda de directeur de groupe, pour nous aider et nous motiver. Merci à Cécile, Jean Kaplan pour le suivi, Ken pour la bonne humeur, Pierre pour les cours de Python (désolé, je suis plutôt Chrome que Firefox).

Si ces années à l'APC ont été si agréables et inoubliables, c'est grâce aux thésards. Merci aux conseils et à la sagesse des vieux: Mariana (o mejor, gracias), Seb pour le sourire et les débats foot, Gaël pour les ragots, George et Florian pour leur humour noir, Marie-Anne pour avoir mis un peu de grâce dans ce monde masculin, Loïc pour les débats cosmologiques sans fin, Alex, Giulio (campeur), Castex (ton royaume n'a pas duré), Josquin, Joseph et Guigui. La génération suivant la nôtre a apporté encore plus d'ambiance à l'APC: merci Alexis, pour la PHD, pour l'union des thésards et pour l'ambiance quoi (popopopo-popo-popo-popo-po), Jibril pour les discussions sur les filles qui nous font du bien et du mal, Maxime pour les galettes maison, Henri (casse-toi au FACE), à Aaahhh-lelek-Ileyk-Ileyk-Ileyk, à Julien Peloton pour être l'Obama des thésards, à Julien Treguer pour les débats Lyman- $\alpha$ . Merci et bons vents aux tout jeunes Camille, Léa, Mikhail et Davide.

Cependant notre génération était la meilleure, et mérite le plus grand merci. Je pourrais écrire des tonnes de mots et de souvenirs qui ont fait ces trois ans passer vite, mais restons succinct. Je commence par ceux qui sont gravés dans mon coeur depuis l'époque du Master, et pour toujours. Ben, merci d'avoir toujours été la meilleure compagnie dans le bureau, d'être le meilleur prof de culture française (parisienne) du monde. Mamatt, merci pour mettre l'ambiance dans les soirées Caféoztiques et de m'avoir accueilli en temps de galère dans ton appart'. Romain, l'exemple de copain idéal, qui nous a fait les meilleures crêpes flambées du monde. Julien, pour le style original et contagieux, toujours dispo pour partager une bonne binch. Flavious, une âme Lebowskienne de sourire indestructible, merci pour les meilleures blagues et d'avoir partagé tes savoirs en musique. Je remercie Maïca, pour nous avoir montré le professionnalisme nécessaire

pour publier en *Nature*. Merci aux autres NPACs, Pierre, Agnès, Guillaume, Asénath, Alice, entre autres pour les inombrables soirées de tout type.

Je remercie Marchou, Hippo, Doudou, Chichou, Grani et Lê, sans qui les matchs de foot n'auraient pas été les mêmes, ainsi que Guitou, pour la confiance.

É hora de agradecer pessoas em português! Adivinha quem? Nunca poderia deixar de agradecer minha família brasileira na França. Ao Papaizinho, que sempre estava lá pra me escutar e me aconselhar, pra jogar bola, pra dançar, ver jogo, melhor companhia pra tudo que eu fazia, obrigado por tudo. À Sofia, por toda a energia e amor que você transmite à todo mundo (em especial ao Papaizinho). Ao Mila, que *sempre* responde meus chamados pra tomar um café ou uma cerveja e conversar durante horas, sobre tudo. Ao Tikago, por ser um exemplo de força de vontade, que se transformou nesses três anos: de físico tímido à consultor e dançarino profissional. Ao Louzada, pelas várias frases *cult*. À Karina, por cuidar desse povo e de mim quando eu vinha! Obrigado, família.

En castellano (por favor!), quiero agradecer a mi jefe, Nicolás. Gracias por ser un jefe genial, por haberme enseñado tantas cosas, por mostrarme como se trabaja, por haberme ayudado siempre, de cerca o de lejos. Gracias por haberme dado la oportunidad de hacer una tesis con vos, de poder practicar el castellano. Sin vos, esta tesis no existiría, y yo no estaría ahora donde estoy. Abraço cara!

Gracias mami, por el amor infinito con el cual siempre me cuidaste, de cerca o de lejos. Gracias papi, por ser mi héroe incansable y por mostrarme como el conocimiento y el trabajo te pueden llevar lejos, a veces demasiado lejos. Gracias Germán, por ser el mejor hermano que lucha hasta hoy por conseguir su lugar. Robinho, por ser irmão e um ótimo papai. Los amo, los extraño, les dedico esta tesis, y a toda la familia.

Gracias Carolinda, por todo el amor y cariño, que me dio y me da energías para seguir adelante. Sin ti no hubiera terminado esta tesis. Je t'aime, à dead.

Dedico esta tesis especialmente a mi hermano Marcelo, mi tío Cacho, mi abuelo Justi, mi abuelo Baroncho, que se fueron muy temprano para verme doctor.

# Contents

<b>Abstract</b>	<b>vii</b>
<b>Acknowledgements</b>	<b>ix</b>
<b>1 Introduction to Cosmology</b>	<b>1</b>
1.1 A Little Word on General Relativity . . . . .	3
1.2 Homogeneous and Isotropic Cosmology . . . . .	4
1.3 The Contents of our Universe . . . . .	6
1.3.1 Photons . . . . .	6
1.3.2 Baryons . . . . .	7
1.3.3 Dark Matter . . . . .	7
1.3.4 Neutrinos . . . . .	8
1.4 Linear Perturbation Theory . . . . .	9
1.4.1 The Perturbed Metric . . . . .	10
1.4.2 Boltzmann Equation . . . . .	10
1.4.3 Perturbing Einstein's Equations . . . . .	12
1.4.4 Evolution of Perturbations: Matter Power Spectrum . . . . .	13
1.4.5 The Baryon-Photon Fluid : BAO . . . . .	15
1.4.6 The Cosmic Microwave Background . . . . .	16
1.5 Baryon Acoustic Oscillations as a Cosmological Probe . . . . .	18
<b>2 Tracing Dark Matter : The Lyman-<math>\alpha</math> Forest</b>	<b>23</b>
2.1 Introduction to the Lyman- $\alpha$ forest . . . . .	23
2.2 The Gunn-Peterson trough . . . . .	25
2.3 Reionization and the photo-ionization equilibrium . . . . .	27
2.4 Understanding the Ly $\alpha$ forest through simulations . . . . .	29
2.5 The Ly $\alpha$ forest and large scale structures . . . . .	30
2.6 First clustering measurements using the Lyman- $\alpha$ forest . . . . .	33
2.6.1 One-dimensional power spectrum . . . . .	33
2.6.2 Three-dimensional correlation function . . . . .	34
<b>3 The Baryon Oscillation Spectroscopic Survey</b>	<b>37</b>
3.1 The Sloan Digital Sky Survey . . . . .	37
3.2 BOSS: The main scientific goals . . . . .	38
3.3 The telescope . . . . .	39
3.4 The camera . . . . .	39
3.5 The spectrographs . . . . .	41
3.6 The imaging . . . . .	42
3.7 Target selection: galaxies . . . . .	42
3.8 Target selection: quasars . . . . .	43



3.9	Observing	44
3.10	Data reduction	45
3.11	Data releases	47
<b>4</b>	<b>BAO in Galaxy Clustering</b>	<b>49</b>
4.1	Measuring the galaxy clustering	49
4.2	BAO with BOSS galaxies	52
4.2.1	The galaxy samples	52
4.2.2	Masks	53
4.2.3	Weighting galaxies	53
4.2.4	The measurements	54
4.2.5	Fitting for isotropic BAO	55
4.2.6	Results	57
4.3	An Optimal Estimator	58
4.3.1	Commonly used estimators	59
4.3.2	An optimized estimator	60
4.3.3	Performance on lognormal mock catalogs	62
4.3.4	The iterative optimal estimator	63
4.3.5	Performance on PTHalos mock catalogs	64
4.3.6	Application to real data: DR9 CMASS	65
<b>5</b>	<b>Mock catalogs of Quasar Lyman-<math>\alpha</math> forests surveys</b>	<b>69</b>
5.1	Generating the absorption field	69
5.1.1	The lognormal model for the forests	70
5.1.2	Cholesky decomposition of correlation matrix	71
5.1.3	Random sampling of modes : theory	73
5.1.4	Random sampling of modes : tests	76
5.2	Producing BOSS mock data-sets	79
5.2.1	Raw transmission field, $F_\lambda$	80
5.2.1.1	High column density (HCD) systems	80
5.2.1.2	Metals	81
5.2.2	Quasar continua $C_\lambda$	82
5.2.3	The BOSS kernel	82
5.2.4	Flux normalization $M_\lambda$	82
5.2.5	The noise $N_\lambda$	83
5.2.6	Sky mis-subtraction $f_\lambda^{\text{sky}}$	85
5.2.7	The <i>MockExpander</i> package	86
<b>6</b>	<b>Measuring the correlation function of the Lyman-<math>\alpha</math> forest</b>	<b>87</b>
6.1	Continuum fitting	87
6.1.1	Generalities	87
6.1.2	The Gaussian method: C1	88
6.1.3	The Non-Gaussian method: C2	89
6.1.4	The PCA method: C3	90
6.2	The correlation function	91
6.2.1	Weights	91
6.2.2	The measurement	92
6.2.3	Details about implementation	93
6.3	The covariance matrix of $\hat{\xi}$	93
6.3.1	Sub-sampling	94
6.3.2	The Wick approximation	95
6.3.3	Smoothing the covariance matrix	96
6.4	Measuring the BAO peak position	96

---

6.4.1	The model . . . . .	96
<b>7</b>	<b>Analyzing Mock Catalogs</b>	<b>101</b>
7.1	Generalities . . . . .	101
7.2	Distortion of correlations . . . . .	102
7.3	Continuum fitting . . . . .	103
7.3.1	Residuals . . . . .	103
7.3.2	Convergence of iterative method for C2 . . . . .	104
7.3.3	Mean transmission . . . . .	105
7.4	Pixel variances . . . . .	106
7.5	Correlation function . . . . .	107
7.6	Covariance matrix . . . . .	109
7.7	Measuring BAO . . . . .	112
<b>8</b>	<b>Analysis of Real Data</b>	<b>117</b>
8.1	Our sample: Data Release 11 . . . . .	117
8.2	Mean transmission and weights . . . . .	118
8.3	The correlation function . . . . .	118
8.4	The covariance matrix . . . . .	120
8.5	BAO peak position measurement . . . . .	124
<b>9</b>	<b>Study of Systematic Effects</b>	<b>129</b>
9.1	Balmer features . . . . .	129
9.2	Metal contamination . . . . .	132
<b>10</b>	<b>Cosmology from BAO measurements</b>	<b>135</b>
10.1	Implications of BAO in the Ly $\alpha$ forest . . . . .	135
10.2	Implications from galaxy BAO measurements . . . . .	139
10.3	Combining Ly $\alpha$ forests and galaxies . . . . .	141
	<b>Conclusions</b>	<b>145</b>
	<b>Bibliography</b>	<b>147</b>



# Chapter 1

## Introduction to Cosmology

### Prelude

One century ago, Edwin Hubble observed for the first time the correlation between distance and recession velocity of galaxies millions of parsecs away. This close-to-linear relationship was, and still is, interpreted as the expansion of our Universe, which is observable only in scales much larger than the size of galaxies (of tens of kiloparsecs). Using Cepheid stars as standard candles to infer distances, he measured an expansion rate of  $H_0 \sim 500 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , showing for the first time evidence for the expansion of the Universe. Hubble's result is one order of magnitude larger than the accepted value today,  $H_0 \sim 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  (Riess et al., 2011, Herrnstein et al., 1999) for many reasons<sup>1</sup>, nevertheless the astonishing conclusion keeps unchanged since then: the Universe *is* expanding.

In 1996, the instrument FIRAS on board of the COBE satellite measured the cosmic microwave background radiation (CMB) spectrum showing that these photons follow a 2.7 Kelvin black-body distribution (Mather et al., 1994). This is a clear evidence that these photons were once thermalized with matter, meaning that the Universe was much denser and hotter than today, giving another strong evidence for the expansion.

Cepheids are not bright enough if we try to measure them at larger distances. Supernovae of type Ia (hereafter SNIa) are one the most energetic phenomena in the Universe, and can be seen from much larger distances. A possible explanation for these events is the explosion of a white dwarf when its mass reaches the Chandrasekhar limit of  $\sim 1.4$  solar masses ( $M_\odot = 2,0 \times 10^{30} \text{ kg}$ ) after accretion from its companion star, liberating as much radiation as the whole host galaxy. If you can observe galaxies, you are able to observe supernovae. These objects have standardizable light curves (Kowal, 1968) and can be used as standard candles. Even if we do not know the intrinsic luminosity of these objects, relative measurements of the observed flux can tell their relative distance. In 1998, measurements indicated for the first time that distant SNIa appear

---

<sup>1</sup>Hubble was wrong twice. First, he calibrated distance from M31 using Cepheids from a different kind of those in the Milky Way. Then, instead of measuring Cepheids in more distant galaxies, he observed bright HII regions.

dimmer than expected in a constant expansion Universe. This showed that the expansion is actually *accelerated* (Riess et al., 1998, Perlmutter et al., 1999).

From Newton to Einstein, our interpretation of gravity results in attraction between massive objects. How galaxies can be accelerated in the opposite way as gravity tell us? What is causing this anti-intuitive acceleration? Are we misinterpreting observations somehow? Or the behavior of gravity is more complex than described by Einstein’s General Relativity (hereafter, GR)?

Our current description of gravity, GR, is proven to be consistent with measurements up to the scales of our Solar System. Most of gravity-related measurements beyond the Solar System are “indirect”, meaning that they cannot contest the validity of GR, but they use GR to infer physical properties, e.g. the mass of exoplanets or the mass of gravitational lenses. The accelerated expansion of the Universe might be just another hint that GR is not a complete description of gravity. Improving measurements on large distances, and therefore in the past, is very important to access new physics.

A recent method to infer distances uses a standard ruler instead of standard candles as a probe of the cosmic expansion. What is this ruler? As said before, observations show that the Universe was so hot and dense in the past that nucleons, electrons and photons acted as a single fluid, a plasma. During this hot period, each plasma overdensity was growing because of gravity. On the other hand, this plasma was being pulled away of these potentials due to thermal pressure. This competition between gravity and pressure around these density fluctuations lead to the propagation of density waves outwards of potential wells. These sound waves propagate until the pressure falls to zero after recombination. When the protons and electrons stop feeling the pressure of photons, defined as *drag epoch*, the sound waves stop propagating and leave a slightly over-dense shell of baryonic matter around each original perturbation. This phenomenon is then called *baryon acoustic oscillations* (BAO). The radius of these shells is, since drag epoch, only affected by the expansion of the Universe, with negligible gravitational collapse into the central region of original perturbations. Therefore, measuring this radius at different times gives a direct measurement of the expansion history of the Universe.

In this manuscript I describe how physics of the expansion can be inferred using measurements of the baryon acoustic scale from the distribution of matter in the Universe. In this chapter, the basics of the current cosmological model and the physics of BAO are presented. In Chapter 2, we describe the Ly $\alpha$  forest and its relationship with the cosmic distribution of matter. In Chapter 3, I summarize the Baryon Oscillation Spectroscopic Survey, conceived to measure BAO in the distribution of galaxies and using the Ly $\alpha$  forest of quasars. In Chapter 4, I explain how to use galaxies to measure BAO, summarizing the last BOSS BAO measurements, and describing the new optimal estimator for this kind of measurement. The measurement of BAO using the Ly $\alpha$  forest is the main subject of this thesis and is sub-divided into many chapters. Chapter 5 describes the generation of mock catalogs dedicated to the measurement of BAO in the Ly $\alpha$  forest. The analysis chain used to perform these measurements is presented in Chapter 6. Numerous tests of our methods were performed using mock catalogs in Chapter 7. Once validated, our analysis methods were applied to real data in Chapter 8. In Chapter 9, I discuss some of the systematic effects related to this measurement, using both data and mock catalogs. Chapter 10 presents the cosmological implications of the measurements presented in previous chapters.

In models based on GR<sup>2</sup>, there is a simple model that can describe an Universe in accelerated expansion. This model is based in the cosmological principle, that states: “Our Universe is statistically *homogeneous and isotropic* on cosmological scales”. It means that, at scales much larger than the largest structures such as clusters of galaxies ( $\sim 1$  Mpc), matter seems to be distributed homogeneously in the Universe, and that there are no preferred directions. In this chapter, we briefly expose this model. However, it is obvious that the distribution of matter in the Universe is not uniform, otherwise galaxies, filaments, planets or us would not exist. These structures originate from density perturbations are difficult to describe analytically using GR, but they can be included as an extension of the homogeneous and isotropic model. This will be discussed in section 1.4.

## 1.1 A Little Word on General Relativity

Using the assumption that all free-falling frames are equivalent, Einstein concluded that gravity is not a force. Actually, space-time becomes curved in the presence of matter, and matter free-falls following this curved space-time over trajectories called *geodesics*. Massless particles such as photons are subject to the curvature of space-time and also have curved trajectories in presence of matter. Space-time curvature is usually described by a metric tensor  $g_{\mu\nu}$  characterizing the invariant line element  $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$  ( $\mu = 0$  corresponding to time and  $\mu = 1, 2, 3$  to space coordinates). This metric depends on the energy content filling space. This relationship is summarized by Einstein’s field equations. These are non-linear, second order differential equations for the metric  $g_{\mu\nu}$ . They are written as

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \quad (1.1)$$

where  $G_{\mu\nu} = R_{\mu\nu} - \frac{R}{2}g_{\mu\nu}$  is the difference of the Ricci tensor and the metric scaled by half of the scalar curvature  $R = R^\mu_\mu$ .  $T_{\mu\nu}$  is the energy-momentum tensor, representing energy and momentum densities, and momentum fluxes (equivalent to pressure).  $G$  is the Newton’s gravitational constant and  $c$  is the speed of light in vacuum. The 4x4 tensors in Eq. 1.1 are symmetric, yielding a system of 10 independent equations (6 of them are the same because of symmetry). The curved geometry described by the Ricci tensor is well defined if it satisfies Bianchi identities. These identities come from differential geometry considerations, adding 4 more constraints on Einstein’s equations. These are expressed as  $G^\mu_{\nu;\mu} = 0$ , where the “;” symbol denotes a covariant derivative (which is a well-defined derivative in curved space-time). Bianchi identities are related to the physical independence of Einstein’s equations on the choice of coordinate system. This freedom is called *gauge invariance*. To solve these equations, one needs to fix a given coordinate system. It is interesting to note that Bianchi identities, pure geometrical constraints, imply local energy conservation  $T^\mu_{\nu;\mu} = 0$ .

<sup>2</sup>No experiment to date has shown evidence against GR with direct measurements.

## 1.2 Homogeneous and Isotropic Cosmology

The cosmological principle restricts the range of possibilities when choosing a metric describing space-time. Isotropy allows us to define the spatial metric  $g_{ij}$ <sup>3</sup> through the expression of an infinitesimal 4-line element  $ds$  with no spatial dependence. The time-dependence is allowed by the cosmological principle. This dependence is set by an arbitrary function of time  $a(t)$ , that multiplies  $g_{ij}$  and works as a space dilator. For convenience, we write  $ds$  in spherical coordinates. Formally, the radial dependence of the metric is set by the homogeneity condition, implying the spatial scalar curvature  $R = R_i^i$  (contraction of the space-space Ricci tensor) to be also constant over space. As stated before, the gauge freedom imposes us to arbitrarily fix some terms of the metric. Usually the choice is a diagonal metric such that  $g_{00} = 1$  and  $g_{0i} = 0$ , which simplifies equations. The homogeneous and isotropic metric is thus written as :

$$ds^2 = c^2 dt^2 - a^2(t) \left[ \frac{dr^2}{1 - \kappa r^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right], \quad (1.2)$$

where  $\kappa = R/6$  is the curvature of the space. If  $\kappa = 0$ , space is flat and infinite (critical). If  $\kappa < 0$  space is hyperbolic and infinite (open), while  $\kappa > 0$  corresponds to a spherical finite space (closed). The function  $a(t)$ , commonly called *scale factor*, describes the evolution of the spatial metric, allowing the description of an expanding Universe. This metric was proposed by Friedmann, Roberston and Walker, therefore referred as FRW metric hereafter.

The FRW metric allows us to completely compute the left hand side of Eq. 1.1 as a function of the scale factor. The right-hand side of the same equation is also simplified under the assumptions of homogeneity and isotropy. Describing the energetic content of our Universe as perfect fluid in thermodynamic equilibrium, the energy-momentum tensor has the simplified form :

$$T_{\mu\nu} = \left( \rho + \frac{p}{c^2} \right) u_\mu u_\nu + p g_{\mu\nu}, \quad (1.3)$$

where  $\rho$  is the energy density,  $p$  the pressure,  $u_\mu$  is its 4-velocity. The cosmological principle forces  $u_0 = 1$  and  $u_i = 0$ , meaning that the fluid is locally at rest with respect to the frame chosen. Also, it sets  $\rho$  and  $p$  constant over space (they can depend on time).

Putting these ingredients on Eq. 1.1, we obtain two differential equations for the evolution of the scale factor. The first one comes from the time-time component. The second one comes from the trace of Eq. 1.1. They are written as,

$$\left( \frac{\dot{a}}{a} \right)^2 = H^2 = \frac{8\pi G}{3} \rho - \frac{\kappa c^2}{a^2}, \quad (1.4)$$

$$\left( \frac{\ddot{a}}{a} \right) = -\frac{4\pi G}{3} \left( \rho + \frac{3p}{c^2} \right). \quad (1.5)$$

The dots here represent derivatives with respect to time  $t$ , and  $H = \dot{a}/a$  is the Hubble expansion rate. Either combining equations above or using Bianchi identities ( $T^\mu_{\nu;\mu} = 0$ ), we obtain the energy conservation equation:

$$\dot{\rho} = -3H \left( \rho + \frac{p}{c^2} \right) \quad (1.6)$$

---

<sup>3</sup>Latin letters are used for the three spatial components while Greek letters for all the four, including time.

We have three unknowns,  $\rho(t)$ ,  $p(t)$  and  $a(t)$ , and only two independent equations. The last equation needed for solving this system of equations is given by the *equation of state* of the fluid, that relates density and pressure. The equation of state can be obtained from the quantum definitions of energy density and pressure. In the early hot Universe, reactions occurred at such rates that different species were in thermal equilibrium, sharing the same temperature  $T$ . The energy density and pressure are given by averages of energy  $E$  and momentum  $p$  weighted by the distribution of particles in phase-space  $f(\vec{x}, \vec{p})$ . These averages can be expressed as

$$\rho_i = g_i \int \frac{d^3p}{(2\pi\hbar)^3} f_i(\vec{x}, \vec{p}) E(p) \quad (1.7)$$

$$P_i = g_i \int \frac{d^3p}{(2\pi\hbar)^3} f_i(\vec{x}, \vec{p}) \frac{p^2}{3E(p)} \quad (1.8)$$

where the  $i$  subscript denotes different species,  $g_i$  is the species degeneracy (e.g. for photons  $g = 2$  because of the two spin-states) and  $E(p) = \sqrt{(pc)^2 + (mc^2)^2}$  is the 0th component of the 4-momentum. Depending if we consider bosons or fermions in equilibrium, the distribution  $f(p) = [e^{(E-\mu)/k_B T} \pm 1]^{-1}$  is the Bose-Einstein (taking the “−” sign in this formula) or Fermi-Dirac (taking the “+” sign), respectively. These distributions do not depend on the positions  $\vec{x}$  or the orientation of the momentum  $\vec{p}$  because of homogeneity and isotropy. To a good approximation, the chemical potential  $\mu$  is much smaller than the temperature over almost all times and all species, therefore allowing us to set  $\mu = 0$ . This also means that the number of particles and anti-particles is the same, which is not true only for baryons.

If the rest mass energy  $mc^2$  of a given species is zero or much smaller than its average kinetic energy  $\sim k_B T$  ( $mc^2 \ll k_B T$ , where  $k_B$  is the Boltzmann constant and  $T$  is the temperature of the system) these species need a relativistic description. We refer to these as *relativistic* species. In this approximation, the integrals in Eqs. 1.7 and 1.8 can be solved analytically, giving a simple relationship between energy density and pressure:  $P_R = \rho_R/3$  (the  $R$  subscript stands for “relativistic”).

The opposite case, for *non-relativistic* species, the rest mass energy is greater than its temperature ( $mc^2 \gg k_B T$ ), meaning that the momentum of these particles is negligible compared to their rest energy. Therefore,  $P_{NR} \approx 0$  (the  $NR$  subscript stands for “non-relativistic”). We often refer to non-relativistic particles as *pressure-less*.

Given these equations of state, solutions for the Friedmann equations can be found. It is possible to generalize the equation of state for different kinds of fluids by setting  $p = w\rho$ , where  $w$  is a constant ( $w = 1/3$  for relativistic fluids and  $w = 0$  for non-relativistic). The solutions for  $\rho(t)$  is

$$\rho(t) = \rho_0 [a(t)]^{-3(w+1)} , \quad (1.9)$$

which is just translating how species are diluted by the expansion. For  $a(t)$ , analytic solutions can be obtained by assuming one species at a time. Setting  $a(0) = a_0 = 1$  and also assuming flat space ( $\kappa = 0$ ), the solution for a given species is

$$\ln [a(t)] = \frac{2}{3(w+1)} \ln t \quad (1.10)$$



Note that there is an special case, where  $w = -1$  which means that the pressure is negative. In this case, the energy density  $\rho = \rho_0$  and the Hubble expansion rate  $H(t)$  are both constants over time. This implies an exponential growth of the scalar factor with time  $a(t) = e^{Ht}$ . This solution is important because it reproduces an accelerated expansion, as inferred by SNIa observations. This is the origin of the *cosmological constant*  $\Lambda$ , once added by Einstein in his field equations when trying to model a static universe with matter. The cosmological constant has two interpretations depending on which side of Eq. 1.1 it is written. Considered on the left-hand side of the equation, the cosmological constant is a modified theory of gravity, with a constant term  $\Lambda g_{\mu\nu}$  added to  $G_{\mu\nu}$  that cannot be derived from first principles. Added on the right-hand side, it describes a fluid of negative pressure filling all the space, not being affected by the expansion, commonly referred as *dark energy*. This energy has something similar to the vacuum energy in quantum electrodynamics. However, theoretical expectations from QED disagree with observations by  $\sim 120$  orders of magnitude. The accelerated expansion is therefore one of the most important unsolved problems in physics today.

### 1.3 The Contents of our Universe

The generally assumed species in our Universe are photons, baryons, cold dark matter and neutrinos. In these section, the influence of these species on the evolution of our Universe is described. As in the previous section, we still assume homogeneity and isotropy.

It is useful to define here the critical energy density  $\rho_{cr}$ , which is the density the Universe would have in a flat space,  $\kappa = 0$ . It is directly related to the expansion rate of the Universe today,  $H_0$  by one of the Friedmann equations (Eq. 1.4),

$$\rho_{cr} = \frac{3H_0^2}{8\pi G}. \quad (1.11)$$

It is common to quote the energy density of different species in the present relative to this critical density, using  $\Omega_i = \rho_i/\rho_{cr}$ . The curvature and the cosmological constant energy densities can also be defined as fractions of the critical density.

#### 1.3.1 Photons

Photons are mass-less particles responsible for the electromagnetic interactions of charged particles. Most of the photons filling out our Universe come from early epochs. They compose the Cosmic Microwave Background (CMB). These photons are thermalized and follow very closely a  $T_\gamma = 2.725$  Kelvin black-body distribution as measured by the COBE satellite (Mather et al., 1994). Therefore, the energy density of CMB photons can be written as

$$\rho_\gamma = 2 \int \frac{d^3p}{(2\pi\hbar)^3} \frac{1}{e^{pc/k_B T_\gamma} - 1} pc = \frac{\pi^2 k_B^4}{15c^3 \hbar^3} T_\gamma^4, \quad (1.12)$$

which depends only on the temperature  $T_\gamma$ . Since Eq. 1.9 for  $w = 1/3$  (photons are relativistic) gives  $\rho_\gamma \propto a^{-4}$ , we derive that the CMB photon temperature is inversely proportional to the scale factor,  $T_\gamma \propto a^{-1}$ . It is often useful and intuitive to use the temperature of photons, instead of the scale factor or time, to define different epochs of our Universe. However, this is not valid in very early epochs due to the change in the photon distribution when electrons and positrons annihilate. It is actually the density of entropy  $s \propto a^{-4}$  that can be used as a time indicator.

### 1.3.2 Baryons

Protons, neutrons and electrons are usually referred in cosmology as *baryonic* matter, even though electrons are leptons. The term baryonic is used because electrons are much lighter than protons and neutrons, not contributing significantly to the energy budget of these particles. Baryonic matter interacts electromagnetically (through photons), and is in general measurable with electromagnetic instruments.

The interaction of baryons with photons was very important when the Universe was hot and dense, giving origin to baryon acoustic oscillations, the main subject of this work.

Many methods were used to measure the baryon density in the Universe and they seem to agree in results (Fukugita et al., 1998). The two more precise techniques come from the measurement of baryon acoustic oscillations in the CMB, and from the measurement of the abundances of light elements. These measurements agree in a baryon density of about 4% of the total energy density of the Universe today.

Unlike photons, baryons cannot be described as a gas at a given temperature and zero chemical potential. Over most of the cosmic history, baryons are non-relativistic ( $w = 0$ ), meaning that their energy density evolves as  $\rho_b \propto a^{-3}$ .

### 1.3.3 Dark Matter

The measurements of baryon density uses information coming through photons. However, mass measurements can be performed using information extracted from the dynamics of gravitational potentials.

In this context, using Newtonian dynamics one can derive the total mass of gravitationally bound structures based on the measured velocities of its components. Measurements of galaxy rotation curves (Zwicky, 1937) and of the total mass of galaxy clusters based on the X-ray emission of the intra-cluster gas (Vikhlinin et al., 2006) point to the presence of non-baryonic mass which does not interact electromagnetically. This invisible matter is therefore called *dark matter*.

Gravitational potentials responsible for temperature fluctuations on the CMB also give strong evidence for dark matter. The observed amplitude of these fluctuations is much larger (factor of 100) than expected by models with only baryonic matter (see section 1.4 for more details). Also if there were only baryons, temperature fluctuations should be larger than observed by the same factor in order to produce the observed structures today.

Dark matter, as the cosmological constant, can be interpreted as a wrong description of gravity on large scales. There are many theories that try to explain dark matter as an extension or a modification of Newtonian gravity that would describe accelerations with more complex dependence with distances than  $r^{-2}$  (see [Famaey and McGaugh \(2012\)](#) for a review). Even if these theories are able to explain rotation curves and the observed mass-luminosity relationship, they fail to describe most cosmological observations such as CMB temperature fluctuations or the matter power spectrum.

Another possible explanation is that dark matter is composed by particles beyond the Standard Model. These particles would interact only through gravity and weak interactions, therefore referred as weak interacting massive particles (WIMPs). Many experiments try to measure a signal coming from the scatter of WIMPs on high sensitivity fluids. None of them found to date any convincing signal (see [Bergström \(2012\)](#) for a review). The larger density of dark matter in the center of halos would increase the rate of annihilations of WIMPs into gamma rays. In many models, it is expected that a gamma-ray peaked signal from the center of our Galaxy would be visible on Earth (see [Funk \(2013\)](#) for a review). Observations show no evidence for such emission to date ([The Fermi-LAT Collaboration et al., 2013](#)).

In any case, dark matter would act as a pressure-less fluid ( $w = 0$ ) whose energy density decays as  $\rho_{DM} \propto a^{-3}$ . The pressure-less nature of dark matter is essential to explain the observed amplitude of CMB fluctuations and the clustering of galaxies today (details on section 1.4).

### 1.3.4 Neutrinos

Neutrinos are relativistic fermions that interact weakly with matter. Neutrino dedicated experiments are converging towards the consensus that neutrinos actually are massive. The three observable families of neutrinos (electronic, muonic and taunic) would be a quantum linear combination of three in principle non-observable neutrino eigenstates. This implies that neutrinos oscillate between different families, a measurable effect ([Ahmad et al., 2002](#)) possibly explained by the neutrino mass. These experiments show that the difference between the square of two neutrino eigenstate masses is at least of  $\Delta m^2 = 0.0027\text{eV}^2$  meaning that at least one neutrino would have a mass of at least 0.04 eV ([Beringer et al., 2012](#)).

As photons, neutrino scattering was an efficient process at very early times and are also thermalized at a temperature  $T_\nu$ . The difference is that neutrinos stopped scattering efficiently at much earlier times than photons due to the nature of the weak interaction. Except for the fact that the photon temperature is slightly increased when positrons annihilates with electrons, neutrinos have approximately the same temperature as photons.

The expansion history of the Universe is affected by neutrinos because they act as a relativistic fluid, as photons. As described later in section 1.4, neutrinos also affect the evolution of perturbations. Therefore, it is possible to measure the effect of neutrinos on the CMB temperature fluctuations and on structure formation, and extract information about their mass and the number of neutrino families. This shows the beautiful link between the physics of extremely large objects and infinitesimally small fundamental particles.

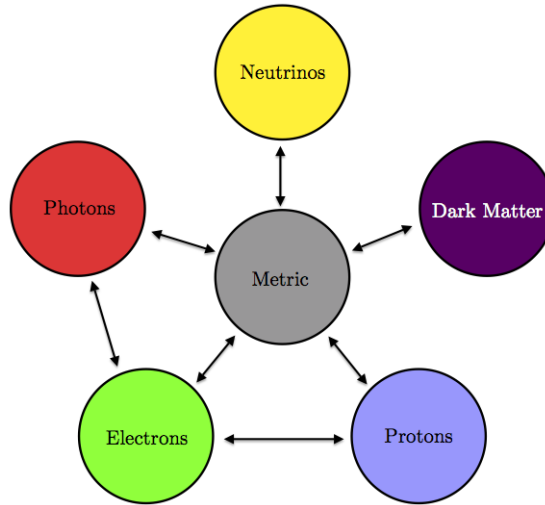


Figure 1.1: Scheme of interactions between different components of Einstein's field equations. Figure based on Fig. 4.1 of [Dodelson \(2003\)](#).

## 1.4 Linear Perturbation Theory

In section 1.2, a homogeneous and isotropic model for the Universe was introduced. Because of homogeneity, no galaxies or other structures can be formed in this model. Also, baryon acoustic oscillations (BAO), a phenomenon originated by density fluctuations, cannot exist. Because of the complexity of GR equations (Eq. 1.1), the evolution of density perturbations cannot be treated analytically. However, considering them *small* compared to the average density (which is a good approximation for early times) allows us to expand equations to first order. Differential equations become linear in this approximation.

The first step is to consider small fluctuations around the mean density,

$$\rho(\vec{x}) = \bar{\rho} [1 + \delta(\vec{x})] , \quad (1.13)$$

with  $\delta \ll 1$ . The evolution of the mean density  $\bar{\rho}$  is described by the homogeneous cosmology. The interest is turned to the description of the density perturbations,  $\delta(\vec{x})$ , responsible for structure formation. Due to gravitational attraction, all species have peculiar velocities  $\vec{v}$  which need to be taken into account.

The evolution of density perturbations results from an interplay between density fields of different species (baryons, dark matter, photons and neutrinos) and the space-time metric. The interaction between species and the metric is governed by Einstein equations and the Boltzmann equation (see § 1.4.2). Fig. 1.1 schematizes these interactions just described. Photons and electrons interact mainly through Thomson scattering and all species interact with the metric (gravity). Neutrinos interact with electrons through weak forces, but cross-sections are negligible with respect to Thompson scattering between photons and electrons.

### 1.4.1 The Perturbed Metric

To describe an homogeneous, isotropic and expanding Universe, we have chosen a coordinate system for which the metric is diagonal (Eq. 1.2). To treat perturbations, we need to consider a more general metric that allows space-dependence, eventually containing non-diagonal terms. Metric perturbations can be divided on three classes<sup>4</sup>: scalar, vector and tensor perturbations. Scalar perturbations dominate the evolution of the density field that describes structure formation and BAO. In most cosmological models, vector perturbations quickly decay and become negligible since early times. Tensor perturbations are source of primordial gravitational waves, which produce a special polarization pattern of CMB photons. Hereafter, only scalar perturbations are considered (for an introduction of non-scalar perturbations see [Dodelson \(2003\)](#), chapter 5).

The perturbed metric can be chosen to be described by two scalar functions (since we are dealing with scalar perturbations),  $\Phi$  and  $\Psi$ , both depending on space and time. This choice is commonly referred as a “gauge” choice, that fixes the employed coordinate system. The metric can be written as

$$g_{00}(\vec{x}, t) = -1 - 2\Psi(\vec{x}, t) , \quad (1.14)$$

$$g_{0i}(\vec{x}, t) = 0 , \quad (1.15)$$

$$g_{ij}(\vec{x}, t) = a^2(t)\delta_{ij} [1 + 2\Phi(\vec{x}, t)] . \quad (1.16)$$

In all our development, setting  $\Phi$  and  $\Psi$  to zero should give us the homogeneous zero-order cosmology discussed previously. Usually,  $\Psi$  corresponds to the Newtonian potential and  $\Phi$  to the spatial curvature.

In linear theory, these functions are treated as small quantities. All quadratic and higher order terms are discarded. Of course, this is an approximation that works astonishingly well for the statistical description of the CMB temperature fluctuations and the formation of large scale structures (larger than  $\sim 40 h^{-1}\text{Mpc}$ ), including BAO. The reason is that, in the standard model, potential perturbations are small ( $\Phi \sim \Psi \sim 10^{-5}$ ) since very early times and never grow (actually decaying at radiation epoch). For a better small scale theory, the higher order terms need to be considered.

### 1.4.2 Boltzmann Equation

One of the ingredients in the description of perturbations is the Boltzmann equation. It describes the evolution of particles distributions in phase-space (position and momentum).

Let  $f(\vec{x}, \vec{p})$  be the distribution in phase-space of particles of a given species. The Boltzmann equation is written as

$$\frac{df}{dt} = C[f] , \quad (1.17)$$

---

<sup>4</sup>This division in three classes of perturbations is a consequence of the decomposition theorem (see [Dodelson \(2003\)](#), section 5.4).

where  $C[f]$  is a functional that defines the collision, creation and annihilation terms. This term takes into account possible interactions with other species. The left hand term is a *total* derivative of the distribution with respect to time, that makes appear terms taking into account the expansion of the Universe and metric perturbations.

To translate the idea of this equation, we consider the case of photons, for which the distribution in phase-space is given by the Bose-Einstein distribution (with the chemical potential  $\mu$  set to zero),

$$f(\vec{x}, p, \hat{p}, t) = \left[ \exp \left\{ \frac{p}{T(t) [1 + \Theta(\vec{x}, \hat{p}, t)]} \right\} - 1 \right]^{-1}, \quad (1.18)$$

where we include temperature fluctuations through the  $\Theta(\vec{x}, \hat{p}, t)$  function. The zero-order temperature  $T$  depends on time (it is proportional to  $a^{-1}(t)$ , § 1.3). We assume that the temperature contrast  $\Theta$  depends only on the position (translating inhomogeneities) and on the direction of photon momentum  $\hat{p}$  (translating anisotropies). This assumption discards the dependence of  $\Theta$  on the momentum magnitude  $p$  because the change in this magnitude is negligible for Compton scattering processes considered next.

The zero-order Boltzmann equation (setting  $\Theta = 0$ ) gives us simply the dependence of temperature with the scale factor already derived previously ( $T \propto a^{-1}$ ). The first order Boltzmann equation describes the evolution of  $\Theta$ . The total derivative on the left side of Eq. 1.17 can be developed (considering only first order terms),

$$\left. \frac{df}{dt} \right|_{\text{first order}} \propto \frac{\partial \Theta}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Theta}{\partial x^i} + \frac{\partial \Phi}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Psi}{\partial x^i} \quad (1.19)$$

The first two terms are related to the free streaming of photons, and the last two account for the gravitational effects from the metric.

The collision term on the right-hand side of Eq. 1.17 takes into account Compton scattering of photons. The computation of this term is an usual calculation in classical physics (energies involved are not relativistic). The result depends on the Thompson scattering cross-section  $\sigma_T$ , the velocity of baryons  $\vec{v}_b$  and on the density of electrons  $n_e$ . It is written as

$$C[f(\vec{p})] \propto -n_e \sigma_T [\Theta_0 - \Theta(\hat{p}) + \hat{p} \cdot \vec{v}_b], \quad (1.20)$$

where  $\Theta_0(\vec{x}, t) = (4\pi)^{-1} \int d\Omega \Theta(\vec{p}, \vec{x}, t)$  is the monopole of temperature fluctuations (integral over the momentum direction).

Assembling Eqs. 1.19 and 1.20, writing them as function of conformal time  $\eta \equiv t/a(t)$  for clarity, and using dots for derivatives with respect to  $\eta$ , leads to

$$\dot{\Theta} + \hat{p}^i \frac{\partial \Theta}{\partial x^i} + \dot{\Phi} + \hat{p}^i \frac{\partial \Psi}{\partial x^i} = n_e \sigma_T a [\Theta_0 - \Theta(\hat{p}) + \hat{p} \cdot \vec{v}_b]. \quad (1.21)$$

If we Fourier transform all these variables, the resulting Fourier amplitudes obey ordinary differential equations, which are much simpler to solve. Derivatives with respect to position become scalar products with wavevector  $\vec{k}$ . This procedure is commonly employed in perturbation theory.

In Fourier space, the previous equation becomes

$$\dot{\Theta} + ik\mu\Theta + \dot{\Phi} + ik\mu\Psi = -\dot{\tau} [\Theta_0 - \Theta + \mu v_b] , \quad (1.22)$$

where we introduce the cosine of the angle between the wavevector and the photon direction,  $\mu = \hat{k} \cdot \hat{p}$ , and the optical depth  $\tau(\eta) = \int_{\eta}^{\eta_0} d\eta' n_e \sigma_T a$ . For clarity, symbols denoting Fourier amplitudes are the same as for configuration space if not otherwise stated.

For dark matter, the Boltzmann equation is much simpler due to the absence of collision terms. Final equations for the evolution of dark matter overdensities  $\delta = \rho/\bar{\rho} - 1$  and velocities  $v$  are given by

$$\dot{\delta} + ikv + 3\dot{\Phi} = 0 , \quad (1.23)$$

$$\dot{v} + \frac{\dot{a}}{a}v + ik\Psi = 0 . \quad (1.24)$$

For electrons and protons, Coulomb scattering ( $e + p \rightarrow e + p$ ) needs to be considered in the collision term of Boltzmann equation (Eq. 1.17). The Coulomb scattering rate is much larger than the expansion rate at all epochs of interest. This tight coupling forces electron and proton overdensities to a common value,  $\delta_e = \delta_p = \delta_b$  (where  $b$  stands for baryons). Similarly, velocities of the two species are the same. In Fourier space, the resulting equations for baryon density perturbations are given by

$$\dot{\delta}_b + ikv_b + 3\dot{\Phi} = 0 , \quad (1.25)$$

$$\dot{v}_b + \frac{\dot{a}}{a}v_b + ik\Psi = \dot{\tau} \frac{4\rho_\gamma}{3\rho_b} [3i\Theta_1 + v_b] . \quad (1.26)$$

where  $\Theta_1(\vec{x}, \eta) \equiv i \int_{-1}^1 \mu \Theta(\vec{p}, \vec{x}, \eta) d\mu/2$  is the local dipole of temperature fluctuations.

### 1.4.3 Perturbing Einstein's Equations

The Boltzmann equation describes the evolution of distributions of different species under action of the Newtonian potential  $\Phi$  and curvature  $\Psi$ . To solve this system of differential equations, two more equations are needed. In the following, we see that Einstein's field equations give us the evolution of  $\Phi$  and  $\Psi$  as functions of densities of all species.

Using the perturbed metric (Eqs. 1.14, 1.15 and 1.16) to compute the Einstein's equations (Eq. 1.1) is a long computation. Details are not included here since no interesting physical assumption is needed. As with Boltzmann equations, only terms linear in perturbations are considered. The zero-order part of these equations leads simply to the homogeneous equations of section 1.2. The first-order equations give the evolution of  $\Phi$  and  $\Psi$ .

The first of the two missing equations come from the time-time component of the Einstein's equation and the second from the *traceless* component of the space-space part. Final equations in Fourier space are given by

$$k^2\Phi + 3\frac{\dot{a}}{a} \left( \dot{\Phi} - \frac{\dot{a}}{a}\Psi \right) = 4\pi G a^2 (\rho_m \delta_m + 4\rho_r \Theta_{r,0}) , \quad (1.27)$$

$$k^2 (\Phi + \Psi) = -32\pi G a^2 \rho_r \Theta_{r,2} , \quad (1.28)$$

where  $\rho_m$  and  $\rho_r$  are the densities of total matter (dark + baryonic) and total radiation (photons + neutrinos), and  $\Theta_{r,2}$  is known as anisotropic stress and represents the quadrupole of the radiation temperature in a given point (in configuration space).

The first equation is just a generalized Poisson equation in an expanding Universe. It shows that the evolution of potentials is different if the Universe is dominated by radiation or matter, which sets the right-hand side of the equation. The second equation defines the difference between  $\Phi$  and  $\Psi$  as a function of the anisotropic stress. Usually,  $\Theta_{r,2} = 0$  is assumed, yielding an equivalence relation between gravitational potential and curvature,  $\Psi = -\Phi$ .

These two last equations complete the system of Boltzmann equations, allowing one to solve for all variables: density and velocity perturbations for each species and the metric potentials. The evolution of baryon and photon perturbations describes how BAO occurs in the primordial plasma before recombination. The cosmic microwave power spectrum is described by solutions for the photon temperature fluctuations  $\Theta$  and the gravitational potential  $\Phi$ , projected over the last scattering surface.

#### 1.4.4 Evolution of Perturbations: Matter Power Spectrum

Solutions for the differential equations described previously depend on the choice of initial conditions. The problem of the initial conditions is related to the problem of the origin of our Universe, which is one of the most intriguing questions in physics. The quasi-homogeneous CMB temperature and the flatness of space-time support the idea that, at very early times, the whole Universe was causally connected (becoming quasi-homogeneous). The Universe would have passed through a rapid exponential expansion called *inflation* (Tsujiikawa, 2003, Peebles, 1993). Inflation could in principle explain the seeds of structures (density fluctuations) by assuming they come from quantum fluctuations of a single field, that became macroscopic after the rapid expansion epoch. Enlarged quantum fluctuations would set initial conditions for the solutions of equations discussed in the previous section. Even though inflation seems to explain homogeneity and flatness, no direct observational proof of inflation exist to date.

The main property of quantum fluctuations is that they are Gaussian and uncorrelated among scales. If the potential driving inflation is very close to flat (a “slow-roll” potential), Fourier modes of density fluctuations have variances that are scale-invariant. The power spectrum  $P(\vec{k})$  of density Fourier modes  $\delta(\vec{k})$  is defined as

$$\langle \delta(\vec{k}) \delta(\vec{k}') \rangle = (2\pi)^3 \delta_D(\vec{k} - \vec{k}') P(\vec{k}) , \quad (1.29)$$

where  $\delta_D$  is a Dirac distribution. The scale invariant power spectrum is defined as  $P(k) = A k^{n_s}$ , with  $n_s = 1$ <sup>5</sup>. The statistical description of perturbations through its power spectrum is more

<sup>5</sup> (This footnote ends at the next page) The scale independent object is actually the dimensionless power  $\Delta^2(k) = k^3 P_\Phi(k) / 2\pi^2$ . Let  $d^3\mathcal{P}$  be power inside a volume  $d^3k$  of the gravitational potential fluctuations  $\Phi$ , given by

$$d^3\mathcal{P} = \frac{d^3k}{(2\pi)^3} P_\Phi(k) . \quad (1.30)$$



interesting than the evolution of each single mode. Linear perturbation theory predicts that large scale modes keep their Gaussianity until later times. Non-linear physics affect small scale modes, coupling different wavelengths, and creating non zero odd point correlation functions.

The power spectrum of dark matter is usually used as reference for the study of large scale structures because dark matter is practically not affected by other species or thermal pressure, simply falling into gravitational potential wells. Dark matter fluctuations are affected only indirectly by radiation during radiation dominated era. During matter dominated era, dark matter itself dominates potential wells. Baryons have a small effect in the overall shape of the power spectrum apart from small wiggles due to BAO.

Figure 1.2 shows the evolution of the dark matter power spectrum as a function of time, or redshift. Initial conditions are set with a scale-independent power spectrum  $P(k) \propto k$ . Once the wavelength  $\lambda_k = 2\pi k^{-1}$  of a given density mode is smaller than the Hubble radius  $D_H = c/H(z)$ , this mode is in causal contact and dark matter starts to collapse. Gravity makes matter density perturbations of all scales grow with time, except during radiation dominated era, where the rate of growth is much smaller because of the thermal pressure. Modes outside the horizon are not in causal contact and do not feel the pressure, and keep growing, causing the break in the power spectrum slope. This peaks moves to smaller  $k$  (larger scales) because horizon grows and increasingly larger modes become causally in contact. As matter starts to dominate at  $z < 2000$ , perturbations inside and outside the horizon grow at same rate and the break in the power spectrum no longer moves.

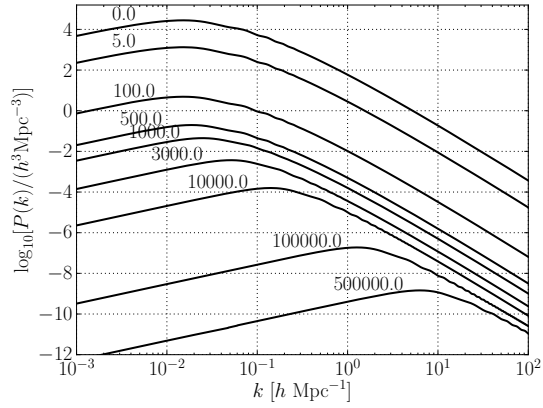


Figure 1.2: Evolution of the matter power spectrum  $P(k)$  starting at  $z = 500000$  (bottom line) to  $z = 0$  (top line).

(End of previous page footnote) Integrating over the solid angle leads to

$$d\mathcal{P} = \frac{k^2 dk}{2\pi^2} P_\Phi = \frac{k^3 P_\Phi(k)}{2\pi^2} d \ln k = \Delta^2(k) d \ln k . \quad (1.31)$$

The scale independence means that  $\Delta^2$  does not depend on  $k$ . Energy density fluctuations  $\delta$  are related to the potential fluctuations  $\Phi$  through the Poisson equation, which in Fourier space is written as

$$\delta = \frac{k^2}{4\pi G \bar{\rho}} \Phi , \quad (1.32)$$

yielding

$$P(k) = \langle |\delta|^2 \rangle \propto k^4 \langle |\Phi|^2 \rangle = k^4 P_\Phi(k) . \quad (1.33)$$

Therefore, a scale invariant power spectrum expressed in terms of energy density fluctuations is

$$\Delta^2(k) = C \Rightarrow P_\Phi = 2\pi^2 k^{-3} C \Rightarrow P(k) = (2\pi^2 C) k , \quad (1.34)$$

which has a dependence in  $k$  as expressed in the text.

### 1.4.5 The Baryon-Photon Fluid : BAO

The baryon effect on the dark matter power spectrum is small but essential for the purpose of this work and it is worth introducing it.

Mixing Boltzmann and Einstein equations in perturbation theory (Eqs. 1.22 to 1.28) results in a differential equation for temperature fluctuations in the baryon-photon fluid before recombination. This equation is similar to a forced damped harmonic oscillator. The damping comes from the expansion of the Universe, and the “force” comes from the gravitational potentials driving the growth of perturbations. Solutions for this equation are oscillatory functions of  $kr_s(\eta)$ , where  $r_s$  is the comoving distance<sup>6</sup> traveled by a sound wave at conformal time  $\eta$  and at sound speed  $c_s$ , i.e.,

$$r_s(\eta) \equiv \int_0^\eta d\eta' c_s(\eta') = \int_\infty^{z(\eta)} dz' \frac{c_s(z')}{H(z')} \quad (1.35)$$

The sound speed in the plasma,  $c_s$ , is given by

$$c_s(z) \equiv \sqrt{\frac{1}{3[1 + R(z)]}}, \quad (1.36)$$

where  $R(z) = 3\rho_b(z)/4\rho_\gamma(z)$  is the ratio of energy densities of baryons and photons (also in Eq. 1.26).

In configuration space, the oscillating solutions describe a sound wave propagation in the plasma around each initial density perturbation. Figure 1.3 illustrates the evolution of the correlation function in real space for four species: dark matter, photons, baryons and neutrinos. On the first three panels, baryons and photons are coupled as a single fluid, and the perturbation propagates as a sound wave with speed  $c_s$ . Since very early times (first panel) neutrinos are decoupled from other species and free-stream at speed  $c$ . Dark matter is pressure-less so perturbations grow with no propagation of sound waves. At  $z \sim 1100$ , the Universe cools down below energies  $E < 1$  eV and electrons start to recombine with protons. Photons have their mean free path increased due to the decreasing number density of electrons. At some point, photon mean scattering time becomes larger than Hubble time (age of the Universe at that time,  $H^{-1}(z)$ ), defining the decoupling of photons from baryons at redshift  $z_*$ . This decoupling marks the emission of the cosmic microwave background (CMB) photons that we observe today. Soon after, baryons stop feeling the radiation pressure and no longer propagate as a sound wave, leaving a overdense spherical shell around each initial perturbation. This defines the *drag epoch* at redshift  $z_{\text{drag}}$ . The radius  $r_d$  of this shell is given by the sound horizon at  $z_{\text{drag}}$ ,  $r_d = r_s(z_{\text{drag}})$ . Decoupling of photons happens slightly before drag epoch  $z_{\text{drag}} < z_*$  because photon number density is much larger than baryon number density by a factor of  $\sim 10^{10}$ . In simple words, photons are more numerous and stop “feeling” baryons before baryons stop “feeling” photons around them. In the current standard cosmological model,  $z_* \sim 1090$  and  $z_{\text{drag}} \sim 1060$ .

The last two panels in Fig 1.3 show that, after recombination, dark matter and baryons mix their gravitational potentials finishing with the same correlation function on large scales. The

<sup>6</sup>A comoving length  $\chi$  is defined by  $l/a(t)$ , where  $l$  is the physical size of the object today and  $a(t)$  is the scale factor. Comoving lengths are constant over time in an expanding Universe, while physical lengths expand accordingly (on cosmological scales).

BAO feature left by baryons at  $z_{\text{drag}}$  is also present in the dark matter correlation function at low redshift.

### 1.4.6 The Cosmic Microwave Background

The cosmic microwave background (CMB) is a projected picture in the sky of the Universe at  $z_*$ . Once the photons decouple from baryons, they scatter for the last time and free-stream. CMB photons arriving on Earth today contain information about temperature  $\Theta$  of the plasma and about gravitational potential fluctuations  $\Phi$  in the region where they last scattered (known as *last scattering surface*). The potential information comes from the gravitational redshift (resp. blueshift) experienced by the photon escaping an overdense (resp. underdense) region. An overdense region produces more bluer photons because of its temperature  $\Theta$ , but these photons are redshifted when escaping its potential well at the last scattering surface: these effects have opposite sign. It is actually the gravitational redshift that dominates over the local temperature (which is defined by local density) when defining the observed photon temperature in the CMB today. Against intuition, hot spots in the observed CMB correspond to underdense regions at  $z_*$ .

The statistics of the observed temperature fluctuations are related to the matter power spectrum projected in the sphere. Instead of Fourier space, temperature fluctuations are decomposed in a linear combination of spherical harmonics  $Y_{\ell m}$ , where  $\ell$  is the multipole degree, related to a separation angle  $\Delta\theta$  in the sky through  $\ell = \pi/\Delta\theta$ . The variance  $C_\ell = \langle |a_{\ell m}|^2 \rangle$  of the measured amplitudes  $a_{\ell m}$  for each spherical harmonic  $Y_{\ell m}$  is called temperature power spectrum. This angular power spectrum  $C_\ell$  is the spherical and discrete analog of the three dimensional power spectrum  $P(k)$  for a given frequency  $k$  in Fourier space.

Figure 1.4 shows the re-scaled temperature angular power spectrum  $\mathcal{D}_\ell = \ell(\ell+1)C_\ell/2\pi$  measured by four experiments: Planck satellite (Planck Collaboration et al., 2014), Wilkinson Microwave Anisotropy Probe (WMAP, Bennett et al. (2013), Hinshaw et al. (2013)), South Pole Telescope (SPT, Bleem et al. (2012), Hou et al. (2014)) and Atacama Cosmology Telescope (ACT, Das et al. (2011)). The first two experiments are satellites able to make full sky maps while the last two are ground based experiments aiming small angular scales (high  $\ell$  values). Up to nine peaks due to baryon acoustic oscillations are clearly seen in the power spectrum, allowing a precise measurement of the angular acoustic scale, given by  $\theta_* = 0.596724^\circ \pm 0.00038^\circ$ . The fitted model (dashed line) perfectly agrees with data, showing the success of the current cosmological model.

The acoustic scale, the overall amplitude of the power spectrum and the relative heights of the peaks are all related to the matter and baryon energy densities,  $\rho_m$  and  $\rho_b$ , because they are the most relevant quantities setting the evolution of acoustic oscillations prior to CMB emission. The temperature of the CMB,  $T_{\text{CMB}} = 2.7260 \pm 0.0013$  K (Fixsen, 2009), gives the energy density of the photons today  $\rho_\gamma$  (Eq. 1.12) that scales back in time as  $(1+z)^4$  because of the expansion. Assuming three mass-less neutrino species, its possible to compute  $\rho_R = \rho_\gamma + \rho_\nu$ . Using the measured values for radiation and matter densities in Eq. 1.35, we obtain an estimate for the sound horizon comoving length,  $r_d = 147.47 \pm 0.59$  Mpc.

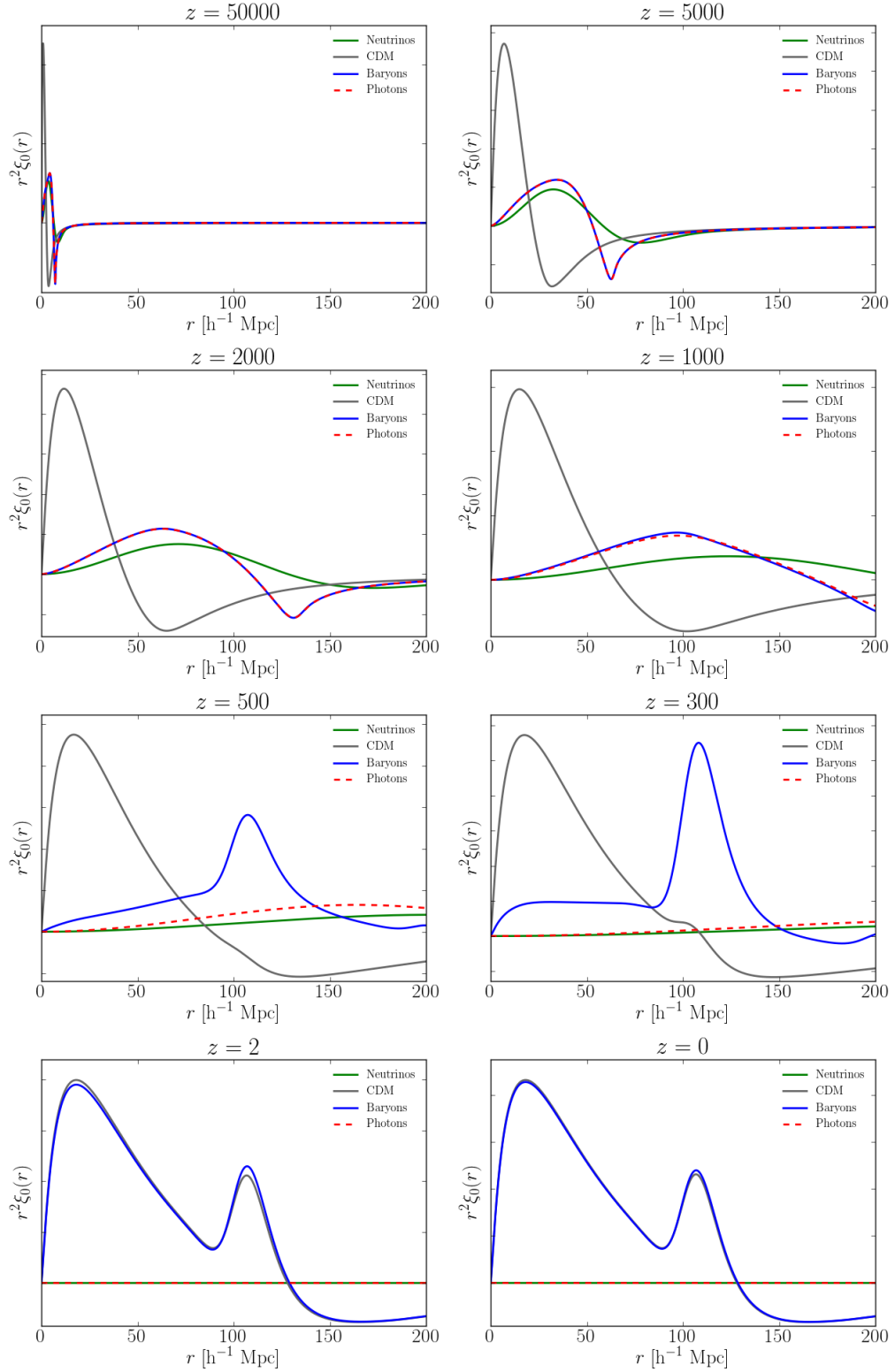


Figure 1.3: Evolution of the correlation function of dark matter (gray), photons (red dotted), baryons (blue solid) and neutrinos (green).

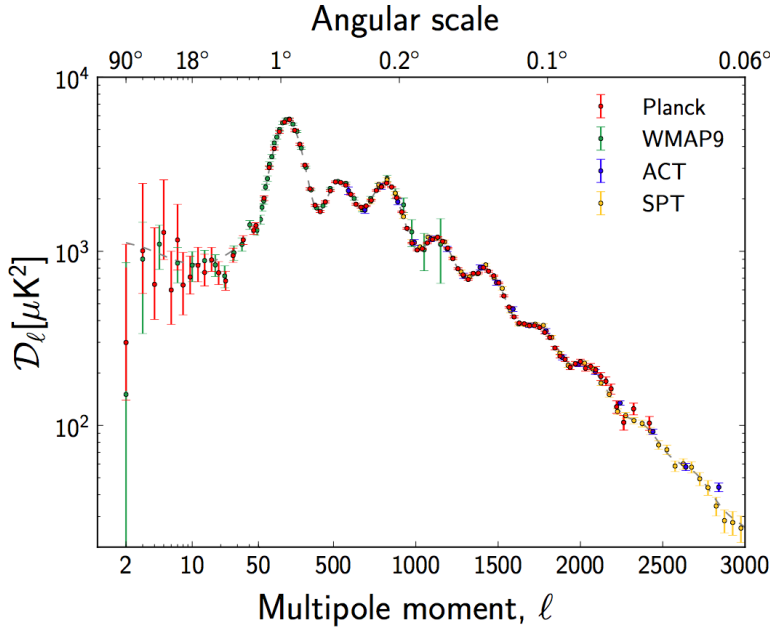


Figure 1.4: Measured temperature angular power spectra  $\mathcal{D}_\ell = \ell(\ell + 1)C_\ell/2\pi$  of Planck, WMAP9, ACT, and SPT. The model plotted is Planck’s best-fit model including Planck temperature, WMAP polarization, ACT, and SPT. Error bars include cosmic variance. The horizontal axis is logarithmic up to  $\ell = 50$ , and linear beyond.

## 1.5 Baryon Acoustic Oscillations as a Cosmological Probe

The most interesting feature of BAO in the correlation function is that the peak position in comoving coordinates practically does not change during the whole period from recombination to nowadays. In other words, the physical size of the sound horizon is *only* affected by the expansion of the Universe. The reason for that is the large size of the sound horizon, of about 150 Mpc (larger than any collapsed structure) that protects this feature from non-linear effects of structure formation. This means that the BAO scale  $r_s$  can be used as a standard ruler for cosmological measurements of distance or the expansion itself, similar to the concept of using supernovae of type Ia as standard candles.

One of the differences with respect to supernovae is that BAO is not a single object measurable in the sky. The signal is very weak, so statistical measurements over large volumes are needed. BAO is often referred to as a statistical standard ruler (SSR) for this reason.

The baryon acoustic feature is a three-dimensional ruler. Therefore, it is possible to measure it in many orientations in the sky with respect to us, e.g. in the radial direction or transverse direction. In practice, we measure angles in the sky and redshift differences as shown in Fig. 1.5. If we assume that BAO has a known comoving size, using the measured angular size of BAO we can infer the angular distance  $D_A(z)$  to the redshift  $z$  of this measurement. Analogously, the radial BAO scale expressed as a redshift separation (or difference of velocities) gives us a measurement of the local expansion rate  $H(z)$ . We explain now how this works.

Geometrical inferences described above are not as simple due to the expanding, and possibly curved, space-time. The infinitesimal comoving radial displacement of a photon is given by

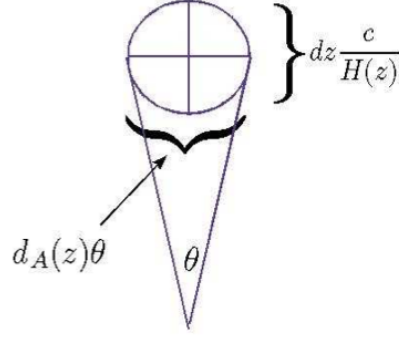


Figure 1.5: The radial length of an object is given by  $c dz/H(z)$  where  $dz$  is the difference in redshift between the front and back of the object while the transverse size of the object is  $D_A(z)\theta$  and  $\theta$  is its angular size. In the case of BAO, the diameter can be theoretically determined, allowing a measurement of  $D_A(z)$  and  $H(z)$  separately. Figure from [Bassett and Hlozek \(2009\)](#).

Eq. 1.2 with  $ds = d\theta = d\phi = 0$ , which can be written as

$$d\chi(z) = \frac{c}{a(t)} dt = -\frac{c}{H(z)} dz = -D_H(z) dz, \quad (1.37)$$

where

$$D_H(z) = \frac{c}{H(z)} \quad (1.38)$$

is the Hubble distance. The Hubble expansion rate,  $H(z)$ , can be written as a function of the energy density fractions of different species as

$$H(z) = H_0 \sqrt{\Omega_m(1+z)^3 + \Omega_{DE}f(z) + \Omega_k(1+z)^2 + \Omega_r(1+z)^4}. \quad (1.39)$$

The function  $f(z)$  defines the evolution of the dark energy density. A cosmological constant corresponds to  $f(z) = 1$ . The comoving distance at redshift  $z$  is given by the integral of the photon radial displacement from  $z$  to now ( $z = 0$ ),

$$\chi(z) = c \int_0^z \frac{dz'}{H(z')}. \quad (1.40)$$

The angular diameter distance  $D_A(z)$  is defined such that an object of size  $L$  at redshift  $z$  has an angular size in the sky of  $\Delta\theta = L/D_A(z)$ . It can be expressed as a function of the comoving distance  $\chi(z)$  as

$$D_A(z) = \frac{c}{H_0(1+z)\sqrt{\Omega_k}} \sin\left(\sqrt{\Omega_k}\chi(z)\frac{H_0}{c}\right). \quad (1.41)$$

where  $\Omega_k$  is the curvature “energy density”. If  $\Omega_k < 0$ , the sine function becomes a hyperbolic sine. In the limit of flat cosmology,  $\Omega_k = 0$ , then this expression is simply  $D_A(z) = \chi(z)/(1+z)$ .

Figure 1.6 shows how  $D_A$  and  $D_H$  vary as a function of redshift for three cosmological models, assuming flatness ( $\Omega_k = 0$ ),  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $\Omega_m = 0.27$  and a cosmological constant  $\Omega_\Lambda = 0.73$  for the fiducial choice. When increasing or decreasing  $\Omega_\Lambda$  by  $\pm 5\%$  the value of  $\Omega_m$  changes accordingly to conserve flatness ( $\Omega_m + \Omega_\Lambda = 1$ ).

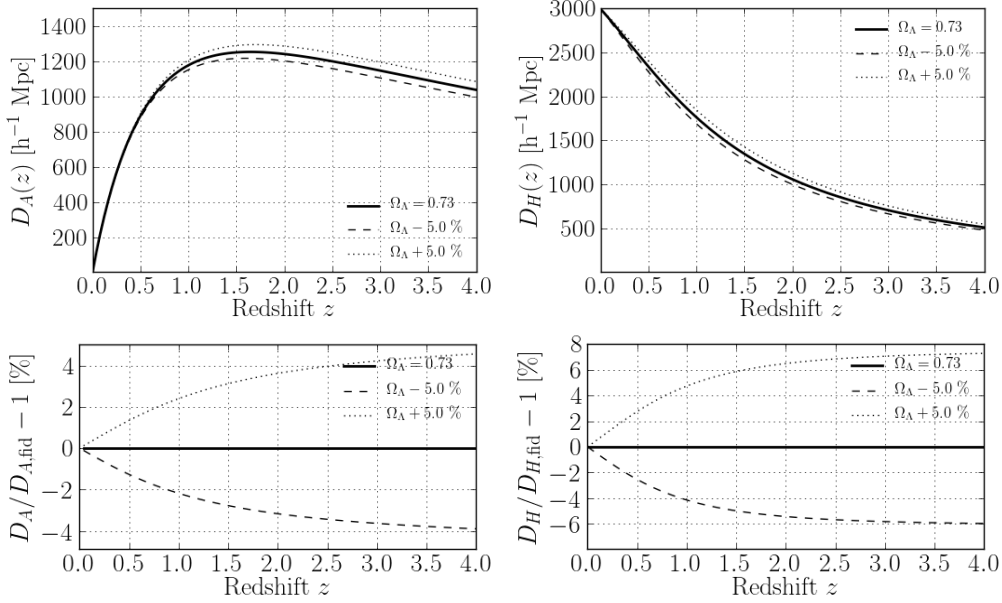


Figure 1.6: Evolution with redshift  $z$  of the angular diameter distance  $D_A$  and the Hubble distance  $D_H$  defined in Eqs. 1.41 and 1.38, respectively, for a flat  $\Lambda$ CDM cosmology with  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . Dotted (dashed) lines show the same curves with  $\Omega_\Lambda$  increased (decreased) by 5%. Bottom panels show the change of those distances in percents relative to the center cosmology.

Measurements of transverse and radial BAO gives the angular  $\Delta\theta$  and radial size expressed in redshift units  $\Delta z$  of the sound horizon  $r_d$ . This allows us to infer  $D_A(z)$  and  $D_H(z)$  through:

$$D_A(z) = \frac{1}{\Delta\theta} \frac{r_d}{1+z} \quad \text{and} \quad D_H(z) = \Delta z \frac{r_d}{1+z}; \quad (1.42)$$

There are two possibilities to proceed with observations:

- To consider that  $r_d$  is a standard ruler with unknown size. Therefore, similarly for supernovae studies for which the intrinsic luminosity of explosions is not known, we need to perform measurements of the sound horizon at many redshifts and work with ratios of distances,  $D_A(z_1)/D_A(z_2)$  and  $D_H(z_1)/D_H(z_2)$ , where  $z_1$  and  $z_2$  are the redshifts of two different measurements of BAO. This can be considered as the most conservative and model independent approach.
- Assume that  $r_d$  is known and it is calculable using Eq. 1.35 and a set of cosmological parameters derived from measurements using other probes such as the CMB. Latest measurement of BAO in the CMB [Planck Collaboration et al. \(2013\)](#) yields 0.1% precision for the angular size of the acoustic horizon :  $(1.04148 \pm 0.00066) \times 10^{-2}$ . These values constrain the matter density  $\Omega_m h^2$ , the Hubble constant  $h$  and the baryon density  $\Omega_b h^2$  to 0.3% precision assuming a flat  $\Lambda$ CDM model. This results in  $r_d = 147.4 \text{ Mpc}$  comoving, which can be used to derive distances. This approach is clearly more model dependent than the first one.

Either procedure give measurements of  $H(z)$ , which depend on the amount of different species in the Universe, in particular dark energy for which not much information is available to date.

Measurements of BAO at low and intermediate redshift combined with CMB measurements allow one to constrain models of dark energy with time-dependent equation of state ( $f(z)$  in Eq. 1.39) and other exotic models, e.g., decaying dark matter into relativistic particles.

In Chapter 10 we present cosmological constraints obtained with  $D_A$  and  $H$  measurements using galaxies at  $z = 0.32$  and  $0.57$ , and Ly $\alpha$  forests at  $z = 2.34$ , all observed with the Baryon Oscillation Spectroscopic Survey (BOSS).





## Chapter 2

# Tracing Dark Matter : The Lyman- $\alpha$ Forest

In this chapter is explained how the Ly $\alpha$  forest traces the underlying dark matter density field allowing their use to measure baryon acoustic oscillations.

The first part of the discussion is based on the review by [Rauch \(1998\)](#) and references therein. We start by introducing the Ly $\alpha$  forest as an astrophysical object in § 2.1. The theoretical interpretation is discussed in § 2.2 which is linked to the reionization process (§ 2.3). All this theory is put into simulations, described in § 2.4. The use of Ly $\alpha$  forest as a probe of dark matter fluctuations is introduced in § 2.4 using hydrodynamic cosmological simulations. First measurements of large scale correlations using real data are presented in § 2.6.

### 2.1 Introduction to the Lyman- $\alpha$ forest

The Lyman- $\alpha$  forest is the region of quasar spectrum containing large amount of thin absorption lines giving it the aspect of a “forest”. The forest extends from the Ly $\alpha$  emission peak (at rest-frame wavelength of  $\lambda_{\text{Ly}\alpha} = 121.6$  nm) of the quasar spectrum to bluer wavelengths. Figure 2.1 gives an example of a high-resolution quasar spectrum and its Ly $\alpha$  forest observed with the Keck telescope ([Womble et al., 1996](#)). Ly $\alpha$  forests are usually observed on spectra of high-redshift quasars ( $z > 2$ ) because at those redshifts the forests is observable with visible light spectrographs at wavelengths larger than 360 nm.

According to the standard model of quasars (or quasi-stellar objects) ([Greenstein and Schmidt, 1964](#)), these objects are galaxies with a very active central region, very luminous in radio, visible and UV light, historically thought to be stars. These objects are thought to be powered by a super-massive black hole accreting the gas surrounding it ([Kazanas et al., 2012](#)). This creates an intense continuum emission spectrum over a large region of the electromagnetic spectrum, with some characteristic emission lines from hydrogen transitions, such as Lyman- $\alpha$  and Lyman- $\beta$ , and also from ionized metals such as SiIV, CIII, CIV, MgII and NV (blended with Ly $\alpha$ ).

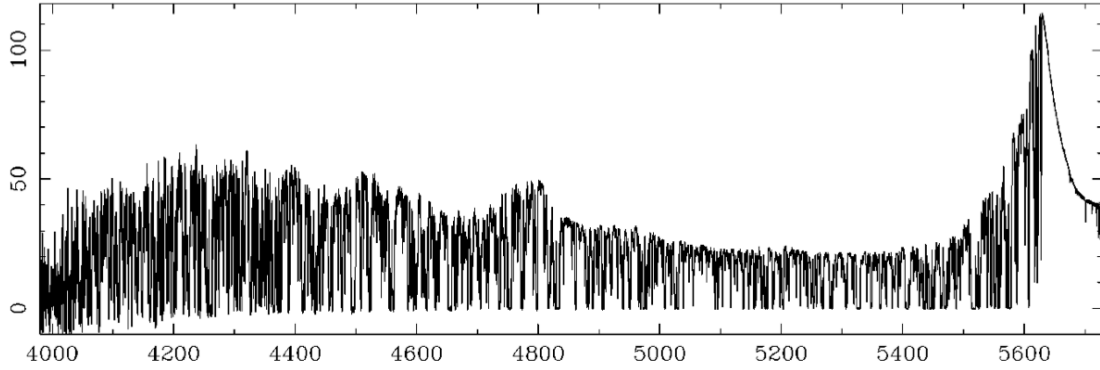


Figure 2.1: High resolution [full width at half maximum (FWHM)  $\approx 6.6\text{km s}^{-1}$ ] spectrum of the  $z_{em} = 3.62$  QSO1422+23 ( $V = 16.5$ ), taken with the Keck High Resolution Spectrograph (HIRES) (signal-to-noise ratio  $\sim 150$  per resolution element, exposure time 25,000 s). Data from Womble et al (1996). The right most peak is the Ly $\alpha$  emission line of the quasar and the peak at  $\sim 4800 \text{ \AA}$  is Ly $\beta$ .

The physical explanation for the nature of the Ly $\alpha$  forest was historically a subject of debate (see [Rauc \(1998\)](#) for a review). It was first suggested that these lines would come from ejected material from the quasar. However, the huge momentum requirements for ejection ([Goldreich and Scoville, 1976](#)), the detection of galaxies at same redshifts as some metal absorption systems ([Bergeron, 1986](#)), the detection of high metallicity gas in systems close to the quasar and low metallicities at more than 30,000 km/s away from it ([Petitjean and Bergeron, 1994](#)) are strong evidences showing that the absorbers are at cosmological distances from the quasar.

Giving the large distance of absorbers from the quasar and that the forest “stops” at the Ly $\alpha$  emission peak, one plausible interpretation for the Ly $\alpha$  forest is that absorbers are composed of neutral hydrogen, absorbing the redshifted light of the quasar at the Ly $\alpha$  resonance,  $\lambda_{\text{Ly}\alpha} = 121.6 \text{ nm}$ . Indeed, for wavelengths smaller than Ly $\beta$ , neutral hydrogen atoms are able to absorb at both wavelengths. Therefore, redwards of Ly $\beta$  peak of the quasar spectrum, Ly $\alpha$  and Ly $\beta$  absorption are mixed together. Similar mixing happens for the higher order resonances of the neutral hydrogen up to the Lyman limit,  $\lambda = 91.2 \text{ nm}$ . Any photon with energies higher than this limit are absorbed, creating a break in the spectrum. These observations confirm that the forest is absorption by neutral hydrogen in the IGM. The intensity of the absorption depends mainly in the local density of neutral hydrogen (see § 2.2).

A small fraction of the lines hidden in the forest are not caused by neutral hydrogen but belong to UV transitions from several common metal or heavy element ions (various ionization stages of C, O, Mg, Si, Fe, and Al are most frequently seen). These metal lines are often associated with strong Ly $\alpha$  lines. By stacking spectra in the rest frame of these Ly $\alpha$  absorbers, metal lines appear very clearly (see Fig. 9.5). These metals are produced by stars inside the gas clouds, enriching the medium through supernova explosions.

If the local density of HI is higher than for usual Ly $\alpha$  lines, the gas becomes optically thick to ionizing radiation, and a discontinuity at the Lyman limit (91.2 nm) is detectable. They are usually referred as Lyman limit systems (LLS). In systems with even higher densities, the ionizing background light cannot reach central regions rendering the gas predominantly neutral, a process commonly called self-shielding. The full absorption of the lines and the peculiar velocities of the

gas create a large trough in the spectrum. The damping wings of the Lorentzian component of the absorption profile begin to be detected, reaching their maximum in the so called Damped Lyman- $\alpha$  systems (DLAs).

When significant amounts of matter are being expelled from the quasar over its line of sight, broad absorption lines (BALs) are produced in its spectrum at wavelengths corresponding to Ly $\alpha$  and other metals. These lines are broad due to the velocity dispersion of the outflow gas. Since the outgoing matter comes from the quasar itself, BALs usually appear bluewards of the emission lines of the quasar. How close from the emission line depends on outwards velocity of the gas.

## 2.2 The Gunn-Peterson trough

Gunn and Peterson estimated for the first time, in 1965, the amount of neutral hydrogen in the intergalactic medium (IGM) needed to match observations of the Ly $\alpha$  forest at  $z \sim 2$  (Gunn and Peterson, 1965). The conclusion was that most of the hydrogen in the IGM is actually ionized at  $z \sim 2$ . This calculation is detailed in this section.

If we define  $F(\lambda)$  as being the ratio between the observed flux in the forest,  $f(\lambda)$ , and the original unabsorbed flux,  $C(\lambda)$ , at observed wavelength  $\lambda$ , we can express this fraction as a function of the optical depth,  $\tau(\lambda)$ , as

$$F(\lambda) = \frac{f(\lambda)}{C(\lambda)} = e^{-\tau(\lambda)}. \quad (2.1)$$

We refer to the transmitted flux fraction  $F$  as *transmission* hereafter. The optical depth,  $\tau(\lambda)$ , quantifies the probability of a photon, observed at  $\lambda$ , to be scattered during its travel between the quasar and us.

The scattering probability  $d\tau$  of a photon over a infinitesimal proper length interval  $dl = a(t) dr$  at time  $t$  (where  $a(t)$  is the scale factor as defined in § 1.2), is

$$d\tau = n_{\text{HI}}(t)\sigma(\nu_s) dl, \quad (2.2)$$

where  $n_{\text{HI}}(t)$  is the local number density of neutral hydrogen atoms at time  $t$  (assumed to be uniform over space) and  $\sigma(\nu)$  is the excitation cross-section for the neutral hydrogen, function of the local photon frequency  $\nu$ . The function  $\sigma(\nu)$  has the form

$$\sigma(\nu) = \frac{\pi q^2}{m_e c} g(\nu). \quad (2.3)$$

where  $q$  is the electron charge,  $m_e$  the electron mass and  $g(\nu)$  is the profile function such that  $\int_0^\infty g(\nu) d\nu = 1$ .

For neutral hydrogen, the profile function  $g(\nu)$  has many narrow resonant peaks in the form of Lorentzian profiles centered at each resonant frequency. Those frequencies correspond to transitions between quantum levels of energy. Transitions between the ground state and other

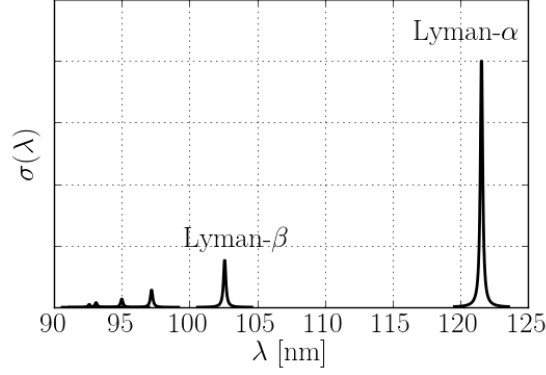


Figure 2.2: Scheme of the resonance peaks of the absorption cross-section of the neutral hydrogen as a function of wavelength  $\lambda$ . The widths of those peaks is only illustrative.

levels are characterized by the Lyman series. Figure 2.2 schematizes  $\sigma(\lambda)$  for the Lyman series. The most significant cross-section among the Lyman series is the Ly $\alpha$  resonance at  $\lambda = 121.6$  nm.

The number density of neutral hydrogen atoms  $n_{\text{HI}}(t)$  in Eq. 2.2 can be computed from the baryon energy density. The baryon energy density today,  $\rho_b$ , is given by a fraction  $\Omega_b$  of the critical density  $\rho_{cr} = 3H_0^2/8\pi G$  (Eq. 1.11, assuming a flat metric). In a past time, or redshift  $z$ , this density was larger by a factor of  $(1+z)^3$  due to the expansion of the Universe. Considering that most of baryons are hydrogen and helium atoms, and that the helium fraction is given by  $Y \sim 0.24$ , we have that the hydrogen energy density at redshift  $z$  is  $\rho_b(1-Y)(1+z)^3$ . Neglecting the contribution of electrons to the hydrogen energy density and assuming that a fraction  $X$  of those atoms is neutral, the number of neutral hydrogen atoms per unit volume is given by

$$n_{\text{HI}}(z) = X \frac{3H_0^2}{8\pi G} \frac{\Omega_b}{m_p} (1-Y)(1+z)^3 = 5 \times 10^{-6} \text{cm}^{-3} X \left( \frac{H_0}{70 \text{ km s}^{-1} \text{ Mpc}^{-1}} \right)^2 \frac{\Omega_b}{0.045} \left( \frac{1+z}{3} \right)^3, \quad (2.4)$$

where  $m_p$  is the mass of the proton.

Let  $\nu$  be the frequency of an observed photon in the Ly $\alpha$  forest of a quasar at redshift  $z_0$ . Due to the expansion, the photon frequency in the past goes as  $\nu_s = \nu(1+z)$ . The total optical depth is given by

$$\tau = \int_0^{z_0} d\tau(z) = \int_0^{z_0} n_{\text{HI}}(z) \sigma[\nu(1+z)] \frac{dl}{dz} dz. \quad (2.5)$$

The term  $dl/dz$  is equal to  $D_H(z)/(1+z)$  (Eq. 1.38). If the photon belongs to the Ly $\alpha$  forest, it means that between  $z = 0$  and  $z = z_0$  the only peak of the profile  $g(\nu)$  in the range of the integral is the one corresponding to the Ly $\alpha$  peak. The width of this profile depends on the IGM temperature, but even at high temperatures ( $\sim 10^6$  K) the velocity dispersion is  $\sim 10^{-4}c$  which is much smaller than the velocity corresponding to a redshift  $z_0 \sim 2$ . Thus, the profile function can be approximated as a Dirac delta centered at  $\nu = \nu_\alpha$ . The integral becomes

$$\tau(\nu) = \int_0^{z_0} n_{\text{HI}}(z) \frac{\pi q^2}{m_e c} f \delta^D[\nu(1+z) - \nu_\alpha] \frac{c}{H(z)(1+z)} dz, \quad (2.6)$$

where  $f = 0.416$  is the oscillator strength of the Ly $\alpha$  resonance. We can factor all functions out of the integral by evaluating them at  $z = \nu_\alpha/\nu - 1 = z_\alpha$ . At observed frequency  $\nu$ , it depends

only on the physics at the absorber redshift  $z_\alpha$ . Assuming  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_m = 0.27$  and  $\Omega_\Lambda = 0.73$ , we have, at  $H(z_\alpha \sim 3) \sim 94 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , approximately

$$\tau \sim 10^5 X . \quad (2.7)$$

Observations show that the optical depth is of order unity in average. Unless the assumptions are wrong by more than an order of magnitude, this result implies that the fraction of hydrogen in neutral form is of order  $X \sim 10^{-5}$ , meaning that the IGM is mainly ionized at  $z \sim 3$ . In other words, if the IGM was neutral with  $X \sim 1$ , the optical depth would be  $\tau \sim 10^5$  meaning that  $F = e^{-\tau} \sim 0$ , i.e., no light would be observed in the Ly $\alpha$  forest, clearly going against observations.

The optical depth can also be expressed as

$$\tau(\nu) = \Sigma(\nu) N_{\text{HI}}(z_\alpha) , \quad (2.8)$$

where  $\Sigma(\nu)$  is the cross-section integrated over the line of sight and  $N_{\text{HI}}(z_\alpha)$  is the neutral hydrogen column density, expressed in units of  $\text{cm}^{-2}$ . It is common to quantify the absorption of a given line in the forest by its corresponding column density. Most Ly $\alpha$  forest absorption lines have column densities between  $10^{13}$  and  $10^{16} \text{ cm}^{-2}$ . More dense systems can completely absorb the light from the quasar at Ly $\alpha$ , also showing a clear break at the Lyman limit. Those systems are then referred as Lyman limit systems. For systems denser than  $10^{20} \text{ cm}^{-2}$  a large trough in spectra can be observed even in low resolution spectra, therefore those are referred as Damped Lyman Alpha (DLA) systems (Lowrance et al., 1972, Beaver et al., 1972, Wright et al., 1979). These systems are often associated to dense regions of halos forming galaxies (Smith et al., 1979).

## 2.3 Reionization and the photo-ionization equilibrium

We discuss now how the Universe became ionized between  $z \sim 20$  and 6 after being neutral since recombination at  $z \sim 1000$ . We introduce the concept of photo-ionization equilibrium in the IGM that sets the properties of the Ly $\alpha$  forest.

Recombination is characterized by the evolution of the fraction  $X$  of the hydrogen in its neutral form. It can be computed from the Boltzmann equation (Eq. 1.17), where the collision term has three components,

$$\frac{dX}{dt} = \alpha_{\text{rec}} n_e (1 - X) - \gamma_c n_e X - \Gamma X . \quad (2.9)$$

The first term in the right hand side accounts for the recombination of electrons with free protons, producing a photon. The rate of radiative recombination is given by  $\alpha_{\text{rec}} = 4 \times 10^{-13} (T/10^4 \text{ K})^{-0.7} \text{ cm}^3/\text{s}$  (Peebles, 1968, Zeldovich et al., 1968). This expression does not consider recombination processes where the resulting photon has exactly  $E_\infty = 13.6 \text{ eV}$ , i.e., a photon able to ionize another hydrogen atom. The second term corresponds to collisional ionization processes. The rate of ionization by collisions is given by  $\gamma_c$  which is relatively important only if  $T > 10^6 \text{ K}$ , temperatures much higher than those considered here, of about  $10^4 \text{ K}$ . The

third term describes ionization by the UV photon background at rate  $\Gamma$ . The UV background can be composed by photons of cosmological (CMB) or astrophysical origin (stars, quasars).

At recombination at  $z \sim 1000$ , the cosmic photon temperature falls below 3000 K and the ionization process is no longer an efficient process; electrons start to recombine with protons. The neutral fraction  $X$  rapidly approaches unity. Translated in terms of Eq. 2.9, this means that the third term in the right hand side becomes significantly smaller than the first one. The recombination marks the beginning of the dark ages.

At redshifts  $z \sim 20$ , stars begin to form inside galaxies and to emit a significantly large amount of UV photons, increasing  $\Gamma$  in Eq. 2.9. Quasars also have an intense UV emission contributing to  $\Gamma$ . At the same time, the average recombination time becomes larger than Hubble time. Both factors lead to the ionization of the neutral hydrogen surrounding galaxies, forming spheres of ionized hydrogen. Those spheres grow and merge together, characterizing the ionization of the IGM with  $X \ll 1$ .

The photo-ionization rate is related to the UV background spectrum by

$$\Gamma = \int_{\nu_L}^{\infty} d\nu 4\pi \frac{J(\nu)}{h\nu} \sigma(\nu) \quad (2.10)$$

where  $J(\nu)$  is the background radiation intensity and  $\nu_L$  is the Lyman limit frequency. An extensive work was performed by Haardt & Madau (Haardt and Madau, 1996) to compute the spectrum of the ionizing photon background  $J(\nu)$  created by galaxies and quasars, also taking into account the contribution from helium atoms. Their conclusion is that  $\Gamma \sim 10^{-12} \text{ s}^{-1}$  at  $2 < z < 4$ . This value is relatively more important at those redshifts in Eq. 2.9 than the recombination rate.

The IGM becomes more and more ionized until the density of neutral atoms is such that ionization probability is negligible. The fraction  $X$  becomes constant around  $10^{-5}$ , reaching the so called *photo-ionization equilibrium*. In photo-ionization equilibrium,  $dX/dt = 0$ , we have from Eq. 2.9 (neglecting the collision term because of low temperatures)

$$\alpha_{\text{rec}} n_e (1 - X) \approx \Gamma X, \quad (2.11)$$

yielding (since  $X \ll 1$ ),

$$X \approx \frac{\alpha_{\text{rec}} n_e}{\Gamma}. \quad (2.12)$$

This equation is used when computing the Ly $\alpha$  forest on simulations, relating the optical depth based and the density of gas  $n_e$  through Eq. 2.7. Simulations are further described in § 2.4.

Another ingredient in those equations is the IGM mean temperature  $T$ , that sets the rate of recombination  $\alpha_{\text{rec}}$ . For an ionized hydrogen gas, the temperature is of the order of  $10^5$  K, at which a significant fraction of electrons have energies larger than the bounding energy of the hydrogen  $E_{\infty} = 13.6 \text{ eV}$ .

## 2.4 Understanding the Ly $\alpha$ forest through simulations

First hydrodynamic cosmological simulations tried to reproduce main properties of the IGM and the Ly $\alpha$  forest from a given structure formation model (Cen et al., 1994, Petitjean et al., 1995, Miralda-Escudé et al., 1996). They are based on Eulerian dark matter N-body simulations including hydrodynamics of the gaseous component, galaxy formation and a cosmological constant. They observed that the basic properties of the IGM depend weakly on the cosmological model. Since then, many numerical techniques were employed to simulate the IGM. The smoothed particle hydrodynamics (SPH) technique was first employed by Zhang et al. (1995) and Hernquist et al. (1996) within a CDM model, using Eulerian and Lagrangian frames, respectively. Eulerian techniques (fixed grid) are able to better resolve the low density regions producing the lowest column density Ly $\alpha$  forest. Lagrangian codes (particle frame) are better adapted inside denser regions (e.g. inside mini-halos or galaxies) where a larger dynamical range is needed. Lagrangian codes model better the distribution of high column density systems in the forests such as DLAs (Katz et al., 1996).

Besides small quantitative differences between different simulations of the Ly $\alpha$  forest, the qualitative conclusions are basically the same. Low column density systems ( $N_{\text{HI}} < 10^{14} \text{ cm}^{-2}$ ) are formed by sheet-like gaseous structures. Gas accretes through weak shocks and settles in a dense, central cooling layer, presumably to form stars. At the lowest column densities, gas remains unshocked and just bounces back because of the hydrostatic pressure. The gas is partly confined by gravity and partly by ram pressure. Higher column density clouds arise in more filamentary structures, with column density contours of  $N_{\text{HI}} \sim 10^{14} \text{ cm}^{-2}$  extending continuously and with relatively constant thickness ( $\sim 40\text{-}100 \text{ kpc}$  proper) over megaparsec distances. With increasing column density, the absorber geometry becomes rounder; column density contours at  $N_{\text{HI}} > 10^{16} \text{ cm}^{-2}$  invariably are spherical, entering the regime where the absorbers more closely correspond to minihalos; there the enclosed gas column is high enough to make the absorption system appear as a Lyman limit or damped Ly $\alpha$  system. Figure 2.3 shows the spatial appearance of the Ly $\alpha$  absorbers. The visual appearance of the low column density, sheet-like filamentary structure has been aptly described as a ‘‘Cosmic Web’’. Looking at the higher column density, optically thick gas on scales of several megaparsecs gives a somewhat different impression of chains of mini or larger halos, lining up like pearls on a string, quite similar to the structure seen in N-body simulations of the dark matter distribution. Confirming earlier analytical work, a large fraction of all baryons (80-90%) resides in the low column density Ly $\alpha$  forest, mostly in the column density range  $10^{14} < N_{\text{HI}} < 10^{15.5}$ .

Figure 2.4 shows the temperature-density relation  $T(n)$  for a random line of sight of a smoothed-particle hydrodynamical (SPH) simulation (Haehnelt et al., 1996). It reveals a significant departure from the photo-ionization equilibrium (Eq. 2.12) given by the solid line in the diagram, except for high density regions  $n > 10^{-3} \text{ cm}^{-3}$ . The temperature-density relation is generally steeper than the equilibrium curve because the lower density gas cools by expansion, while the gas in the density range  $10^{-5}\text{-}10^{-3} \text{ cm}^{-3}$  is heated by adiabatic compression or shock heating, non-linear processes that add the scatter observed in the diagram. For low density regions, the



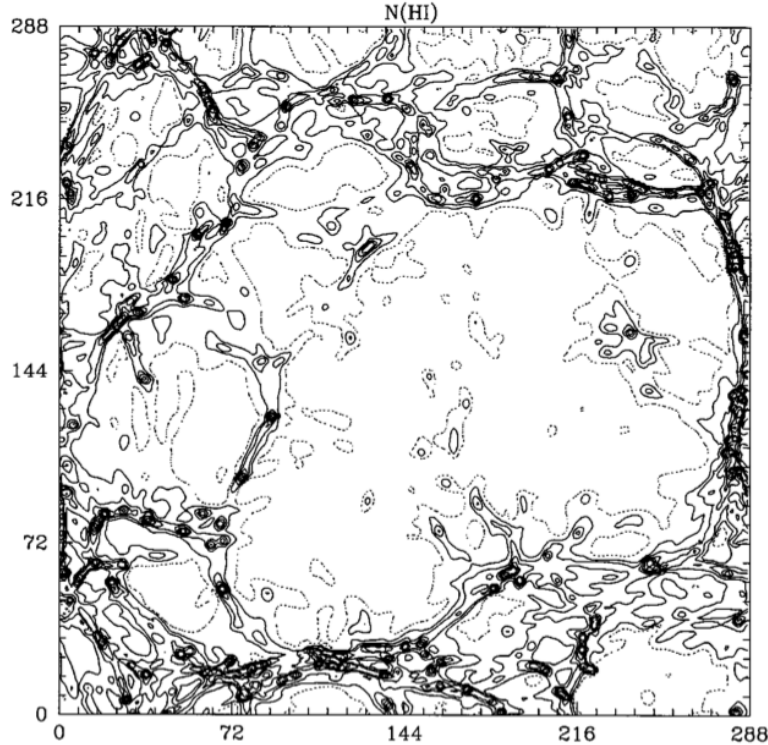


Figure 2.3: Slice of the  $10 h^{-1}\text{Mpc}$  hydrodynamic simulation of a  $\Lambda\text{CDM}$  model showing contours of neutral hydrogen column density, with contours  $10^{12+0.5i} \text{ cm}^{-2}$ ,  $i = 1$  shown dotted and  $i = 2, 3, \dots$  shown solid. Figure from [Miralda-Escudé et al. \(1996\)](#).

temperature density relation is approximated by a power-law,

$$T = T_0 \left( \frac{\rho_b}{\bar{\rho}_b} \right)^\gamma \quad (2.13)$$

where  $\gamma$  lies between 0.3 and 0.6 ([Hui and Gnedin, 1997](#)). This relation is often used in semi-analytical simulations where the gas temperature is not obtained from hydrodynamics and needs to be inferred from the local density.

The evolution of the forest with time at high redshift is mainly driven by the expansion of the universe and the increase of the mean ionization of the gas. The Hubble expansion dilutes the gas inside low density regions, reducing the amount of absorption. The increase in the UV photons (due to the increased amount of stars) reduces the neutral fraction of the hydrogen in those regions, also reducing the absorption.

## 2.5 The Ly $\alpha$ forest and large scale structures

In § 2.2, the Gunn-Peterson (GP) trough was computed assuming a constant baryon density over space. In this section we consider fluctuations (GP fluctuations hereafter) around this mean density, relating them to the underlying dark matter field on large scales. The model for the power spectrum of the Ly $\alpha$  forest transmission is described here.

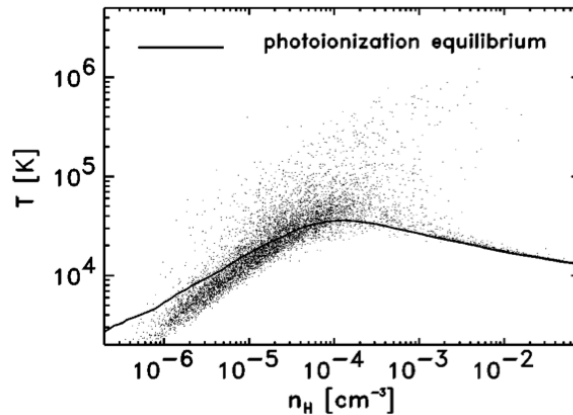


Figure 2.4: Density-temperature ( $n - T$ ) diagram of the Ly $\alpha$  forest at redshift 3.1 from a smoothed particle hydrodynamics (SPH) simulation of a CDM universe. Each dot represents the mean values of the total hydrogen density  $n$  and the gas temperature  $T$  along a random line of sight (LOS) through the simulated box. The solid curve gives the locus of thermal photo-ionization equilibrium. Departures from this curve are due to the dynamical nature of the universe. For all but the most dense regions, expansion cooling in voids (low density regions) and heating by compression and shocks during gravitational collapse steepen the  $T(n)$  relation compared to the equilibrium curve. Figure extracted from [Haehnelt et al. \(1996\)](#).

As discussed in § 2.4, most of the Ly $\alpha$  absorption comes from low density regions presumably associated to small dark matter fluctuations  $\delta = \rho/\bar{\rho} - 1 \lesssim 1$ . For being small, those fluctuations are not far inside the non-linear regime, and are well described by linear perturbation theory (§ 1.4). Also, we should consider peculiar velocities of the gas that introduce a distortion in the observed clustering of absorbers.

In redshift space, the observed position of a tracer is a sum of the radial distance (in velocity units) and the radial velocity. The velocity gradients give rise to additional perturbations called redshift-space distortions (RSD). One can view this as a mapping from the real space to the redshift space with the total number of tracers being conserved. In the large scale limit the RSD take a particularly simple form first derived by [Kaiser \(1987\)](#). The RSD in Fourier space enhances perturbations along the radial direction breaking the isotropy of the original modes  $\delta(\vec{k})$ . This can be written as

$$\delta_s(\vec{k}) = (1 + \beta\mu^2)\delta(\vec{k}), \quad (2.14)$$

where the  $\beta$  is the RSD parameter defined below, and  $\mu$  is the cosine of the angle between  $\vec{k}$  and the line of sight.

For the Ly $\alpha$  forest, the observable is the transmission  $F = e^{-\tau}$ . Let's consider transmission fluctuations  $\delta_F$  defined as

$$\delta_F = \frac{F}{\bar{F}} - 1, \quad (2.15)$$

related to dark matter fluctuations  $\delta$  through a bias function  $b_F$  as

$$b_F = \left. \frac{\partial \delta_F}{\partial \delta} \right|_{\eta}, \quad (2.16)$$

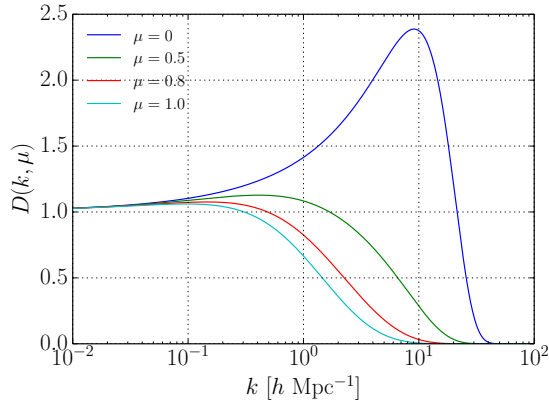


Figure 2.5: Damping kernel of the Ly $\alpha$  forest three-dimensional power spectrum (Eq. 2.19) for four different values of  $\mu = k_{\parallel}/k = 0, 0.5, 0.8$  and 1. Table 2.1 show the values used for the parameters.

where the derivative is taken at constant  $\eta = H^{-1} \partial v_p / \partial r$ , the velocity perturbation in the radial direction. Analogously we define a velocity bias  $b_{\eta}$  as

$$b_{\eta} = \left. \frac{\partial \delta_F}{\partial \eta} \right|_{\delta}. \quad (2.17)$$

On large scales, the fluctuations are still in the linear regime, and considering both biases as scale-independent functions is a good approximation (Kaiser, 1987).

Usually the large scale velocity bias is unity for tracers such as galaxies or gas because they follow the dark matter flow. However, the transformation between gas density (or  $\tau$ ) and the observable  $F = e^{-\tau}$  is non-linear, making the velocity bias become different from unity (Seljak, 2012).

In 2003, McDonald (2003) used hydro-particle-mesh simulations of Ly $\alpha$  forests to compute the three-dimensional flux power spectrum. From these measurements, he obtained values for the large scale bias  $b_F$  and the RSD parameter  $\beta$  (defined as the ratio between the velocity bias  $b_{\eta}$  and the density bias  $b_F$ ). He modeled the flux power spectrum  $P_F(\vec{k})$  as being related to the linear dark matter power spectrum  $P_{\text{lin}}(k)$  (described in § 1.4.4) through

$$P_F(\vec{k}) = b_F^2 (1 + \beta \mu^2)^2 P_{\text{lin}}(k) D(\vec{k}), \quad (2.18)$$

where

$$D(\vec{k}) = \exp \left\{ \left[ \frac{k}{k_{\text{NL}}} \right]^{\alpha_{\text{NL}}} - \left[ \frac{k}{k_P} \right]^{\alpha_P} - \left[ \frac{k_{\parallel}}{k_V(k)} \right]^{\alpha_V} \right\} \quad (2.19)$$

and  $k_V(k) = k_{V0} (1 + k/k'_V)^{\alpha'_V}$ . Figure 2.5 shows the shape of  $D(\vec{k})$  for the parameters in Table 2.1. The first term in the exponential allows for the isotropic increase in power due to non-linear growth, the second term for the isotropic suppression by pressure, and the third for the suppression by non-linear peculiar velocities and temperature along the line of sight.

The flux power spectrum  $P_F(\vec{k})$  was computed over simulations in different box sizes and resolution : 20, 40, 80  $h^{-1} \text{Mpc}$  per side containing  $512^3$ ,  $256^3$  and  $512^3$  particles, respectively. Those simulations were produced to test the effect of resolution on the small scale power, while larger

Table 2.1: Parameters defining the three-dimensional flux power spectrum (Eq. 2.18) obtained from HPM simulations of McDonald (2003).

Parameter	Value
$b_F$	$-0.131 \pm 0.017$
$\beta$	$1.580 \pm 0.022$
$k_{\text{NL}}$	$6.77 \text{ hMpc}^{-1}$
$\alpha_{\text{NL}}$	0.550
$k_P$	$15.9 \text{ hMpc}^{-1}$
$\alpha_P$	2.12
$k_{V0}$	$0.819 \text{ hMpc}^{-1}$
$\alpha_V$	1.50
$k'_V$	$0.917 \text{ hMpc}^{-1}$
$\alpha'_V$	0.528

box simulations probe the large scale power. The flux-density bias  $b_F$  and the velocity bias  $b_\eta$  were found by numerically computing derivatives in Eq. 2.16 and 2.17, averaging results from few realizations. The result is  $b_F = -0.131 \pm 0.017$  and  $\beta = b_\eta/b_F = 1.580 \pm 0.022$ . The density bias is negative because more density *reduces* the transmission  $F$ . In the next section we show that these values are in agreement with measurements using real data. It is important to note that those simulations are not sufficiently large to probe BAO scales. No simulation to date was able to simulate such large scales ( $\sim 150$  Mpc) while resolving correctly the Jeans scale of the forest ( $\sim 200$  kpc). Therefore, all statements about large-scale power are assumptions based on the intermediate scale power.

Using the obtained values of  $b_F$  and  $\beta$ , it was possible to fit the power spectrum model (Eq. 2.18) over simulation measurements in order to obtain values for the parameters defining the damping kernel (Eq. 2.19). Resulting parameters are summarized in Table 2.1. This is the model used to produce mock forests in Chapter 5. By Fourier transforming we obtain the model for the three dimensional flux correlation function  $\xi_F(\vec{r})$  which is the main observable of this thesis.

## 2.6 First clustering measurements using the Lyman- $\alpha$ forest

In this section we summarize some of the main results concerning the clustering in the Ly $\alpha$  forest, that are complementary to the work on simulations presented previously.

### 2.6.1 One-dimensional power spectrum

Before the advent of a dedicated large volume/high density quasar survey such as BOSS, the cosmological exploitation of Ly $\alpha$  forest data was achieved via the measurements of the clustering along individual lines of sight, through the one-dimensional correlation function  $\xi_{1D}$  or power

spectrum  $P_{1D}(k)$ . First evidences for clustering among different absorption lines were found by [Sargent et al. \(1980\)](#), [Webb and Malkan \(1986\)](#), [Muecket and Mueller \(1987\)](#).

Using less than 50 high resolution spectra, the first measurements of  $P_{1D}(k)$  were performed by [Croft et al. \(1999\)](#), [McDonald et al. \(2000\)](#), [Croft et al. \(2002\)](#), [Viel et al. \(2004\)](#), obtaining the first cosmological constraints on  $\sigma_8$  and  $n_s$  from the amplitude and slope of  $P_{1D}$ . Those measurements are sensitive to small scales,  $0.003 < k < 0.03 \text{ km s}^{-1}$  (equivalent to  $\sim 2\text{-}20 h^{-1}\text{Mpc}$  at  $z \sim 3$ ).

With the advent of SDSS, the number of quasars could increase by two orders of magnitude, even if this measurement was not the main goal of the survey. The 1D power spectrum was measured by [McDonald et al. \(2006\)](#) using  $\sim 3,000$  spectra, achieving sub-percent statistical errors in the measurement of the amplitude ( $\sim 0.6\%$ ) and slope ( $\sim 0.005$ ). An important effort was made to understand the systematic errors that can affect the final power spectrum, either from instrumental nature (sky residuals, pixel error mis-estimation, resolution effects) or from astrophysical origin (e.g. metal absorption).

The latest published measurement ([Palanque-Delabrouille et al., 2013](#)) used a sample of 13,821 quasar spectra from the Baryon Oscillation Spectroscopic Survey (BOSS, more on next chapter). This analysis followed closely methods from [McDonald et al. \(2006\)](#), yielding tighter constraints on two cosmological parameters: the variance of the power spectrum of the density field smoothed over spheres of radius  $8 h^{-1}\text{Mpc}$ ,  $\sigma_8 = 0.83 \pm 0.03$ , and slope of the primordial power spectrum,  $n_s = 0.97 \pm 0.02$ , when combined with  $H_0$  constraints from the cosmic ladder and the cosmic microwave background in a  $\Lambda\text{CDM}$  model. For comparison, the same parameters found by the Planck Collaboration are  $\sigma_8 = 0.834 \pm 0.027$  and  $n_s = 0.9616 \pm 0.0094$ , showing a consistency between results using totally different probes.

## 2.6.2 Three-dimensional correlation function

With BOSS, the quasar density in the sky is such that correlation measurements among different lines of sight becomes possible ([McDonald and Eisenstein, 2007](#), [McQuinn and White, 2011](#)).

The first measurement of the three dimensional flux correlation function  $\xi_F(\vec{r})$  ([Slosar et al., 2011](#)) using first-year data from BOSS contained  $\sim 14,000$  quasars and is shown in Fig. 2.6. This measurement put in evidence for the first time the long-range correlations among different lines of sight, reaching comoving separations up to  $60 h^{-1}\text{Mpc}$ . Also for the first time, the effect of redshift-space distortions in the Ly $\alpha$  correlation function was observed and quantified.

The first measurements of the density bias  $b_F$  and the RSD parameter  $\beta$  were performed assuming linear theory (Eq. 2.18). At 95% confidence, it was found  $-0.24 < b_F < -0.16$  and  $0.44 < \beta < 1.20$ , at  $z = 2.25$ , with a well constrained combination  $b(1 + \beta) = 0.336 \pm 0.012$ . The errors on  $\beta$  are asymmetric, with  $\beta = 0$  excluded at over  $5\sigma$  confidence level. The value of  $\beta$  is somewhat low compared to predictions from simulations discussed in § 2.4. One of the possible reasons underlined by the authors for the low value of  $\beta$  (relative to expectations for the Ly $\alpha$  forest alone) would be the presence of high column density systems (HCDs) and metal line absorption in the forests, that increase the density bias (and therefore reducing  $\beta$ ) as observed in mock

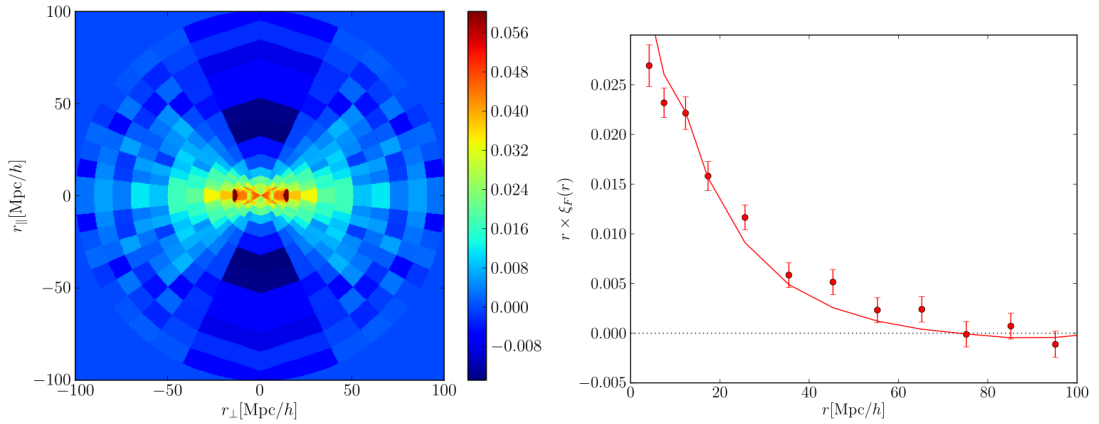


Figure 2.6: First measurement of the flux correlation function in three dimensions. The left panel shows the correlation as a function of parallel  $r_{\parallel}$  and transverse separations  $r_{\perp}$  with respect to the line of sight. The right panel shows the same correlations spherically averaged, with its best fit model.

catalogs. The effect of metals and HCDs is also part of this thesis and results are shown in Chapter 9.

The first measurement of BAO in the Ly $\alpha$  forest correlation function was reported in the subsequent BOSS publication using two years of data (Busca et al., 2013, Slosar et al., 2013, Kirkby et al., 2013), where correlations over separations of more than  $160 h^{-1}Mpc$  are statistically significant. This measurement and its updated version using the full BOSS data sample are the subject of this thesis. The full analysis and results are detailed in the next chapters.



## Chapter 3

# The Baryon Oscillation Spectroscopic Survey

In this chapter I introduce the Baryon Oscillation Spectroscopic Survey project, part of the third generation of the Sloan Digital Sky Survey. I present its main goals, the telescope and its instruments for photometry and spectroscopy, the observation routine and the data reduction pipeline. This Chapter is based on [Dawson et al. \(2013\)](#), [Eisenstein et al. \(2011\)](#) (and references therein).

### 3.1 The Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS) is one of the most successful astronomical surveys to date. It was built for deep digital optical imaging of one fourth of the sky, approximately  $10,000 \text{ deg}^2$ , and spectroscopy of a set of selected objects over  $5,700 \text{ deg}^2$ . It started in the year 2000 and is still operating in its third generation.

The two first versions of the survey SDSS-I and II measured, during eight years of operation, more than 930,000 galaxies and 120,000 quasars, allowing the construction of the largest three dimensional map of the Universe structure. The project has a huge scientific impact, with more than 5000 publications related to SDSS data.

One of the most impressive results is the first detection of baryon acoustic oscillations in the clustering of 46,000 luminous red galaxies at a mean redshift of 0.35 ([Eisenstein et al., 2005](#)).

The third generation of the SDSS ([Eisenstein et al., 2011](#)) started in autumn 2008 based on the same telescope of SDSS-II with improved instruments in order to cover three main themes: dark energy and cosmological parameters, the history and structure of the Milky Way, and the population of giant planets around other stars. Four different projects cover those themes, including BOSS (described later). They are listed below:



- SEGUE-2 : it is the continuation of the SDSS-II Sloan Extension for Galactic Understanding and Exploration (SEGUE), measuring medium-resolution ( $R = \lambda/\Delta\lambda \approx 1800$ ) optical spectra of 118,000 stars in a variety of target categories, probing chemical evolution, stellar kinematics and substructure, and the mass profile of the dark matter halo from the solar neighborhood to distances of 100 kpc.
- APOGEE : the Apache Point Observatory Galactic Evolution Experiment, obtained high-resolution ( $R \approx 30,000$ ), high signal-to-noise ratio ( $S/N \geq 100$  per resolution element),  $H$ -band ( $1.51 \mu\text{m} < \lambda < 1.70 \mu\text{m}$ ) spectra of  $10^5$  evolved, late-type stars, measuring separate abundances for  $\sim 15$  elements per star and creating the first high-precision spectroscopic survey of all Galactic stellar populations (bulge, bar, disks, halo) with a uniform set of stellar tracers and spectral diagnostics.
- MARVELS : The Multi-object APO Radial Velocity Exoplanet Large-area Survey monitored radial velocities of more than 8000 FGK stars with the sensitivity and cadence ( $10\text{--}40 \text{ m s}^{-1}$ ,  $\sim 24$  visits per star) needed to detect giant planets with periods up to two years, providing an unprecedented data set for understanding the formation and dynamical evolution of giant planet systems.

### 3.2 BOSS: The main scientific goals

Since the first measurements of the acoustic peak in the large-scale structures of the Universe in 2005 by the 2dF Galaxy Redshift Survey (Cole et al., 2005) and by SDSS-II (Eisenstein et al., 2005), the door was open for new and more accurate measurements of BAO using larger and deeper survey, which is the main goal of BOSS.

BOSS is divided into two main spectroscopic surveys, executed simultaneously over  $10,000 \text{ deg}^2$  of the sky. The first targets 1.5 million massive galaxies in the redshift range of  $0.15 < z < 0.7$ . The effective volume traced using BOSS galaxies is eight times larger than for SDSS-II. The second survey targets 150,000 quasars over  $2.1 < z < 3.5$  where the Ly $\alpha$  forest absorption in the SDSS spectral range can be used as a tracer of structures at high-redshift. The surface density of quasars in the sky is sufficiently high in order to able the first measurements of three-dimensional structure traced by the Ly $\alpha$  forests.

With this amount of data, forecasts can predict the precision in measurements of the angular diameter distance  $D_A(z)$  and the Hubble expansion rate  $H(z)$  at the completion of the surveys. Using galaxies, it is expected (Eisenstein et al., 2011) measurements of  $D_A$  and  $H$  with  $1\sigma$  precision of 1.0% and 1.8%, respectively, at  $z = 0.35$  (bin width  $0.2 < z < 0.5$ ), and with precision of 1.0% and 1.7%, respectively, at  $z = 0.6$  ( $0.5 < z < 0.7$ ). The errors at the two redshifts are essentially uncorrelated, while the errors on  $D_A(z)$  and  $H(z)$  at a given redshift are anti-correlated (Seo and Eisenstein, 2003). For the Ly $\alpha$  forest survey, forecasts (McDonald and Eisenstein, 2007) indicate final measurement of 4.5% and 2.6%, respectively, on  $D_A(z)$  and  $H(z)$  at an effective redshift  $z \sim 2.5$ , assuming  $10,000 \text{ deg}^2$  covered with a mean density of 15 quasars  $\text{deg}^{-2}$ . No systematic effects were taken into account in this prediction, that can come from metal absorption, damped Ly $\alpha$  systems or broad absorption lines.

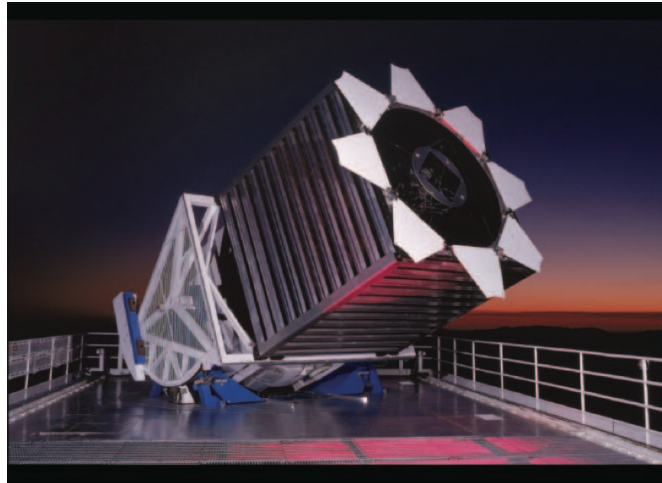


Figure 3.1: The SDSS 2.5m telescope at the Apache Point Observatory ready to start observations. Figure from [Gunn et al. \(2006\)](#).

### 3.3 The telescope

The Sloan Digital Sky Survey telescope is located at Apache Point Observatory, New Mexico, United States. The telescope is a modified two-corrector Ritchey-Chrétien design with a 2.5 m,  $f/2.25$  primary, a 1.08 m secondary, a Gascoigne astigmatism corrector, and one of a pair of interchangeable highly aspheric correctors near the focal plane, one for imaging and the other for spectroscopy. The final focal ratio is  $f/5$ . The telescope is instrumented by a wide-area, multiband CCD camera and a pair of fiber-fed double spectrographs. The telescope include the following: (1) A  $3^\circ$  diameter (0.65 m) focal plane that has excellent image quality and small geometric distortions over a wide wavelength range (300–1,060 nm) in the imaging mode, and good image quality combined with very small lateral and longitudinal color errors in the spectroscopic mode. The unusual requirement of very low distortion is set by the demands of time-delay-and-integrate (TDI) imaging. (2) Very high precision motion to support open-loop TDI observations. (3) A unique wind baffle/enclosure construction to maximize image quality and minimize construction costs. The telescope had first light in 1998 May and began regular survey operations in 2000. More details about the telescope design can be found in [Gunn et al. \(2006\)](#).

### 3.4 The camera

The camera used by BOSS for its imaging phase is the same used in previous versions of SDSS. It was optimized to work in TDI imaging and better cover the 3 deg field of view of the telescope. The full description of the SDSS camera is found in [Gunn et al. \(1998\)](#).

Figure 3.2 shows the camera design. Six columns of five charged coupled devices (CCDs), one for each photometric band, of  $2k \times 2k$  pixels each. Twenty-four smaller CCDs of  $2k \times 400$  pixels are placed around the main CCDs and are used for astrometry and calibration. The TDI drift scan direction is upwards in the Fig. 3.2.

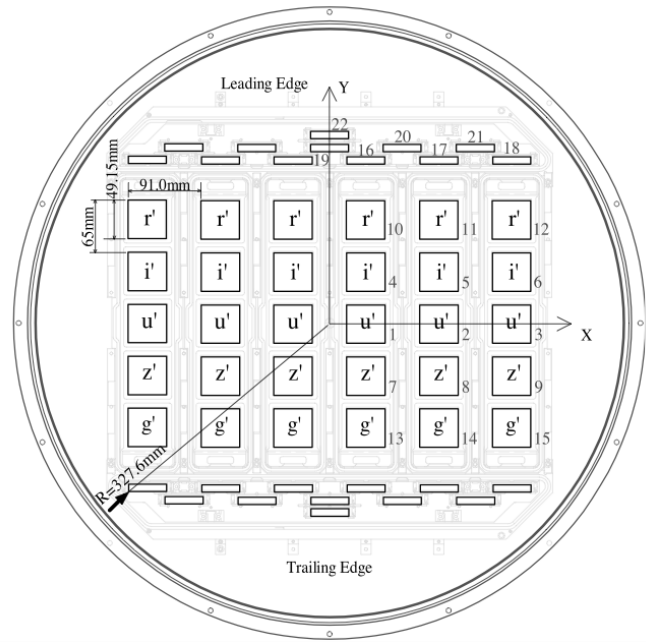


Figure 3.2: The SDSS photometric camera. Optical layout of the focal plane of the SDSS camera. Field 22 (top and bottom) are focus CCDs; Fields 16-21 are astrometric chips, and 1-15 are the photometric array. The time-delay and integrate drift scan direction is upward, so a star traverses the array from top to bottom. Figure from [Gunn et al. \(1998\)](#).

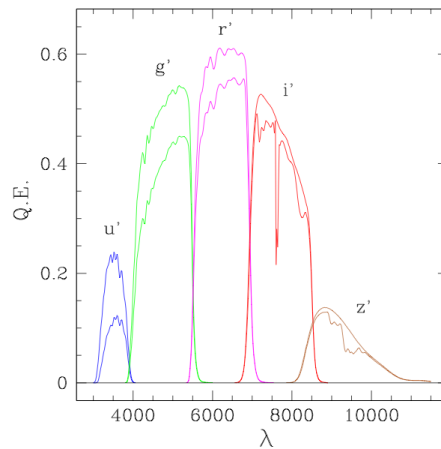


Figure 3.3: System quantum efficiency for each filter/detector system in the photometric array as a function of wavelength. The expected throughput of the optics is included; the lower of each pair of curves includes the expected atmospheric extinction. Figure from [Gunn et al. \(1998\)](#).

The main photometric CCDs are equipped with filters going from UV up to near infra-red (the limit of silicon detectors). Figure 3.3 shows the quantum efficiency of each filter already taking into account the optics, CCDs response and the atmospheric extinction. The low relative efficiency of the  $u$  and  $z$  filters justify their positioning near the center of the focal plane.

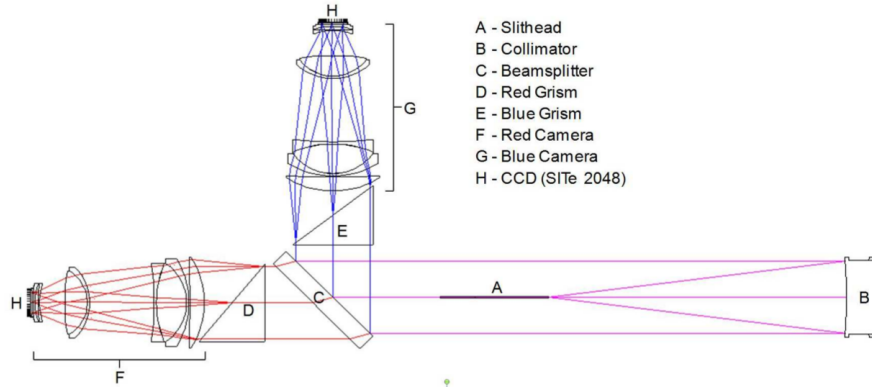


Figure 3.4: Optical layout of BOSS spectrographs. Light enters each spectrograph through the fibers, which terminate at a curved slit plate mounted inside the slithead (A). The slit plate positions the fiber ends on a radius concentric with the spherical collimating mirror (B). The 45 degree dichroic beamsplitter (C) reflects the blue portion of the bandpass ( $\lambda < 605$  nm) and transmits the red wavelengths ( $\lambda > 605$  nm). Immediately after the beamsplitter in each channel is a grism (D, E), dispersing the light, which enters the all-refractive eight-element cameras (F, G). Each camera contains a single  $4k \times 4k$  CCD (H, i) with  $15 \mu\text{m}$  pixels. Figure from [Smee et al. \(2013\)](#).

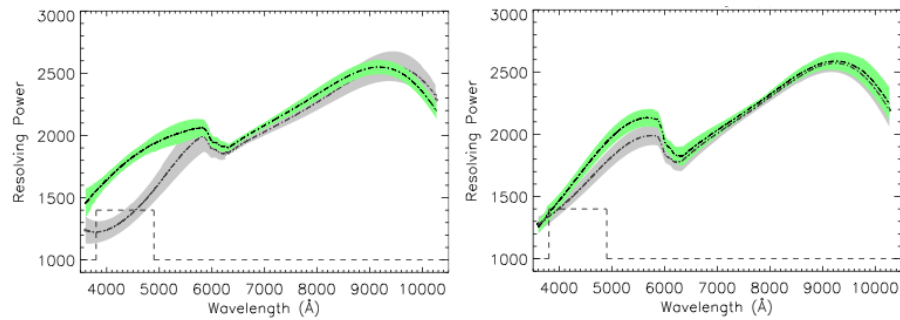


Figure 3.5: The requirements and the measured resolving power for the two BOSS spectrographs. The results for Spectrograph 1 are shown in the left panel and results for Spectrograph 2 are shown in the right panel. The dashed black curve is the requirement for the resolving power. The green curve is the 68% confidence limit about the mean of the resolving power for the central fiber and is representative of 80% of the fibers. The gray curve is the 68% confidence limit about the mean of the resolving power for a representative fiber near the edge of the spectrograph slit. Figure from [Smee et al. \(2013\)](#).

### 3.5 The spectrographs

BOSS is constituted of two identical spectrographs. The employed technology is similar to their SDSS-II versions. The multi-fiber optic system is the main characteristic of this system. The numbers of optical fibers increased from 600 in SDSS-II to 1000 for BOSS. Improvements on the spectrographs quality were needed to attain BOSS scientific requirements. Full technical details about BOSS spectrographs can be found in [Smee et al. \(2013\)](#). We discuss some main their characteristics in this section that are important for analysis described on the next chapters.

The spectral resolution is measured from calibration arc images taken before each set of science exposures. The one-dimensional arc image is first masked to include only pixels within 12 pixels of the center of each arc line on each fiber. A Gaussian of width  $\sigma_\lambda$  is fit to each spectral profile using the 25 unmasked pixels. A fourth order Legendre polynomial model is fit to the derived

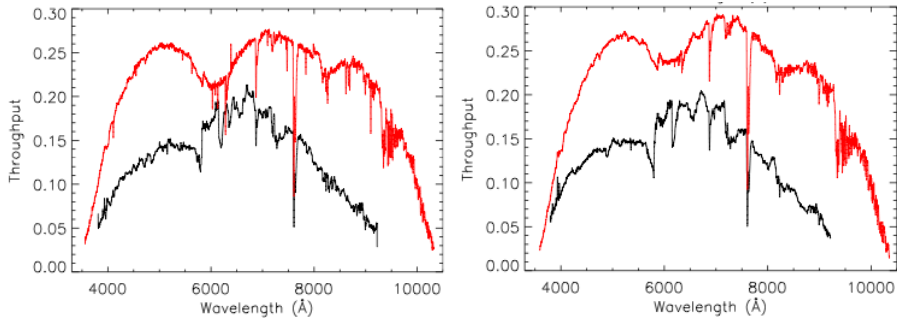


Figure 3.6: Throughput curves for SDSS (black) and BOSS (red). The results for Spectrograph 1 are displayed in the left panel. The results for Spectrograph 2 in the right panel. Figure from [Smee et al. \(2013\)](#).

$\sigma_\lambda$  as a function of wavelength to model the dispersion over the full wavelength range. The resolving power for SDSS and BOSS is then formally defined as  $R = \lambda/2.35\sigma_\lambda$ . The resolution is the FWHM of the Gaussian,  $2.35 \times \sigma_\lambda$ .

The throughput for SDSS was measured by averaging the throughput from 84 different standard stars observed under photometric conditions at an airmass of  $\sim 1.0$  and seeing  $\leq 1.15''$ . Because BOSS has smaller fibers, the measured throughput on the BOSS spectrographs is more susceptible to guiding errors. We attempt to mitigate this effect by averaging the throughput over the four stars that produce the highest throughput on each camera. The throughput for BOSS was measured by averaging the throughput of 24 stars over three plates, split evenly between the two spectrographs.

### 3.6 The imaging

The first stage of the survey, between the years 2000 and 2005, consisted on imaging the full  $11,600 \text{ deg}^2$  of the Northern Galactic Cap using five photometric filters in the  $u$ ,  $g$ ,  $r$ ,  $i$  and  $z$  bands. During the falls of the years 2008 and 2009,  $3,100 \text{ deg}^2$  of Southern Galactic Cap were also imaged. The final image covers  $14,555 \text{ deg}^2$ , more than one fourth of the sky, and contains almost a billion objects detected individually up to a magnitude of  $r \sim 22.5$ .

The next step consists in the spectroscopy of all interesting objects : galaxies and quasars for BOSS. Spectroscopy is much more time consuming process than imaging, therefore a careful selection of the targets for spectroscopy needs to be performed prior to observation. For BOSS, our main targets are galaxies and  $z > 2$  quasars.

The collaboration developed algorithms for selecting these kind of targets. The next two sections briefly describes the process, which is a bit more complex for quasar targets.

### 3.7 Target selection: galaxies

For galaxies, the main goal is to select galaxies composing a homogeneous distributed high density sample over the covered volume. Instead of selecting luminous red galaxies (LRG) as

in SDSS I and II, BOSS galaxies of a same mass and no intrinsic color restrictions are selected, resulting in a denser sample ( $\bar{n} \sim 3 \times 10^4 h^{-3} \text{Mpc}^3$  at  $z = 0.6$ ). For BOSS, two classes of galaxies were selected based on their redshift, called “LOWZ” and “CMASS” samples.

Target selection algorithms for BOSS are based on SDSS-II algorithms (Eisenstein et al., 2001) for LRGs, where the following combinations of colors,

$$c_{\parallel} = 0.7(g - r) + 1.2(r - i - 0.8) , \quad (3.1)$$

$$c_{\perp} = (r - i) - (g - r)/0.4 - 0.18 , \quad (3.2)$$

$$d_{\perp} = (r - i) - (g - r)/8.0 , \quad (3.3)$$

are appropriate to separate galaxies based on their evolution. Targets are selected by cuts made in these new color space. To a good approximation, the perpendicular colors track the location of the 400nm break in galaxy spectra, which provides an initial redshift selection.

The LOWZ algorithm is designed to select red galaxies at  $0.15 < z < 0.45$ . This sample includes galaxies with  $16 < r < 19.5$ ,  $r < 13.6 + c_{\parallel}/0.3$ , and  $|c_{\perp}| < 0.2$ . The CMASS sample is designed to be approximately stellar mass limited above  $z \sim 0.45$  and is built using the following cuts:  $17.5 < i < 19.9$ ,  $d_{\perp} > 0.55$ , and  $i < 19.86 + 1.6(d_{\perp} - 0.8)$ . Stellar mass measurements of BOSS galaxies using photometry show that the distribution of masses is narrow and centered at  $\log(M/M_{\odot}) \sim 11.3$  (Maraston et al., 2013), showing that the algorithms successfully select a constant mass sample of galaxies. Further details about galaxy target selection can be found in Parejko et al. (2013) for LOWZ and in Tojeiro et al. (2012) for CMASS.

### 3.8 Target selection: quasars

The main goal of the BOSS quasar sample is to provide a sufficiently dense sample of lines-of-sight in order to measure BAO through three dimensional Ly $\alpha$  forest correlations. Science forecasts for BOSS (McDonald and Eisenstein, 2007, McQuinn and White, 2011) show a linear dependence of the clustering S/N with surface density of quasars. The quasar luminosities should also be adequate for good spectroscopic measurements. At minimum, 15 quasars  $\text{deg}^{-2}$  at redshifts  $2.1 < z < 3.5$  with  $g_{\text{PSF}} < 22.0$  are required to make this measurement. Hopefully, since the forests trace the IGM gas density, there is no requirements on the homogeneity of the quasar sample as it is the case for galaxies clustering measurements.

To select quasar targets is a more challenging task compared to galaxies. The main reason is that photometric quasar colors are very similar to the colors of stars as shown in Figure 3.7. During the first year of BOSS, four different methods were developed to improve quasar selection using photometric data :

- a “Kernel Density Estimation” (Richards et al., 2009), which measures the densities of quasars and stars in color–color space from training sets and uses these to select high probability targets;

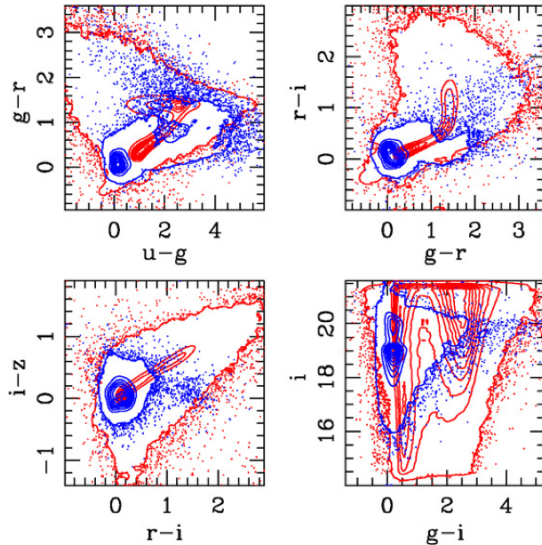


Figure 3.7: Color–color and color–magnitude distribution of objects in the training sets of the “Kernel Density Estimation” algorithm for quasar target selection. Quasars are given in blue (75,382 objects). “Stars” are given in red (429,908 objects). The (linear) contour levels are relative to the peak in each sample. Figure from [Richards et al. \(2009\)](#).

- a “Likelihood” approach ([Kirkpatrick et al., 2011](#)) which determines the likelihood that each object is a quasar, given its photometry and models for the stellar and quasar loci;
- an “Extreme Deconvolution” (XDQSO) ([Bovy et al., 2011](#)) selection, which performs a density estimation of stars and quasars by incorporating photometric uncertainties;
- an artificial neural network ([Yèche et al., 2010](#)), which takes as input the SDSS photometry and errors from a training set in order to run a classification scheme (star versus quasar) and generate a photometric redshift estimate.

Quasar targets fall into five distinct categories. The CORE sample includes targets selected by an uniform method, and since the second year of observations, by the XDQSO method. The BONUS sample includes additional targets selected non-uniformly using full photometric data in order to increase the quasar density. The other three categories include known quasars observed by SDSS (KNOWN\_MIDZ), lower redshift quasars useful for studies of metal absorption (KNOWN\_SUPPZ) and radio confirmed quasars (FIRST).

### 3.9 Observing

As in SDSS-I and II, the multi-object spectroscopy technology is used allowing the observation of hundreds of spectra simultaneously. The focal plane of the telescope is equipped with a 1.5 degree radius spectroscopic plate, accommodating 1000 optical fibers of 2” diameter. Figure 3.8

Science targets are assigned to fibers in plates in a process referred as “tiling” ([Blanton et al. 2003](#)). The tiling process is intended to maximize the fraction of targets that are assigned fibers (tiling completeness) while minimizing the number of tiles required to complete observations







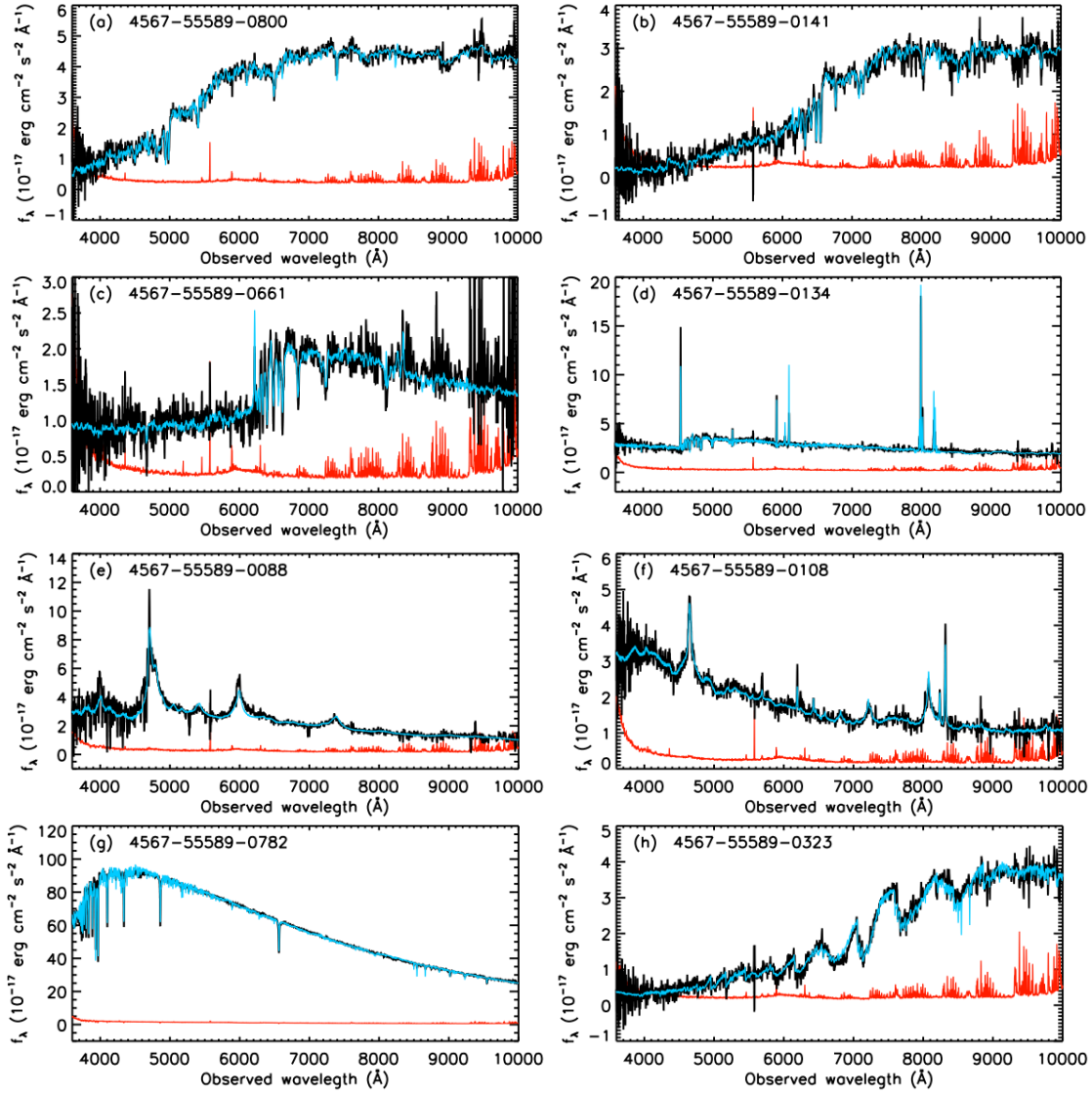


Figure 3.9: Examples of BOSS spectra, with a resolution of  $R \approx 2000$ . Black lines show data (smoothed over a 5 pixel window), cyan lines show best-fit redshift/classification model, and red lines show the extraction pipeline. Spectra are labeled by PLATE-MJD-FIBERID. Individual objects are (a) redshift  $z = 0.256$  LOWZ galaxy; (b) redshift  $z = 0.649$  CMASS galaxy; (c) redshift  $z = 0.669$  CMASS galaxy with post-starburst continuum; (d) redshift  $z = 0.217$  starburst galaxy (from QSO target sample); (e) redshift  $z = 2.873$  quasar; (f) redshift  $z = 0.661$  quasar; (g) spectrophotometric standard star; and (h) M star (from CMASS target sample). Figure extracted from (Bolton et al., 2012).

from two-dimensional images into one-dimensional spectra. The second step is to classify objects and estimate redshifts (Bolton et al., 2012).

Raw CCD frames are pre-processed by subtracting a bias model, their bias overscan, subtracting a dark current model, and dividing by a pixel flat-field model for each CCD. The read noise is measured from the bias overscan region for each CCD amplifier for each exposure. Per-pixel variance is estimated using the measured read noise and the observed photon counts in each pixel. The inverse variance is multiplied by a known CCD defect mask, and cosmic rays are identified to mask affected pixels.

Sky subtraction is performed using a model for background derived from the sky fibers that were assigned during plate design. The background varies with fiber position to account for smoothly varying differences across the focal plane. Similarly, the spectral response (i.e., the flux calibration) is determined over the focal plane using models fit to the spectra from the standard stars that were assigned in plate design.

Finally, the spectra from individual exposures are combined into a co-added frame for each fiber on a resampled grid that is linear in  $\log(\lambda)$ . Blue and red data are combined into one spectrum covering the full wavelength range of the instrument, 361-1014nm.

Pixel variances are propagated into the co-add spectrum pixels, discarding covariances among different exposures. The current pipeline performs this extraction as a single pass, resulting in a known bias due to estimating the variance from the data rather than iteratively updating the noise model with the statistics of the extracted spectra. In the limit of zero flux, the technique systematically assigns a slightly larger variance to pixels that fluctuate toward higher flux values, leading to a weighted mean that is suppressed below its true value. This problem can affect Ly $\alpha$  forest measurements that use this variance estimates and needs to be corrected. We further discuss this in Chapter 5 and 6.

Each object is identified through a fit of a star, galaxy, or quasar template into their combined spectrum. Figure 3.9 shows some typical spectra of different types, with its successfully fitted template. Due to some low signal-to-noise spectra the identification is sometimes inaccurate; those spectra are flagged with warning tags. The pipeline reports that 79% of quasar spectra have a confident identification and redshift estimate. Approximately 51.5% of the quasar targets are confirmed as quasars, the other part being mostly F-stars. However, only 33.6% of the quasar targets are  $2.1 < z < 3.5$  quasars.

In addition to an automatic classification of objects, a visual inspection of all quasar targets and some galaxy targets was performed in order to track errors made by the automatic pipeline. This impressive work is described in Pâris et al. (2012). The inspection could increase by 1.7% the sample of  $2.1 < z < 3.5$  quasars. Also, it found that 0.6% of pipeline-confirmed quasars have bad redshift estimates which occurs in the presence of strong broad absorption lines or with very low S/N spectra. Figure 3.10 shows some examples of classification errors committed by the pipeline. Also, the visual inspection accurately flags the presence of damped Ly $\alpha$  systems or broad absorption lines that affect the Ly $\alpha$  forest correlation measurements.

### 3.11 Data releases

Every year the BOSS collaboration makes publicly available the data in Data Releases<sup>1</sup>. Figure 3.11 shows the evolution of the sky coverage by BOSS data during three years, corresponding to Data Release 9, 10, and 11. The final release is DR12 and will be public together with DR11 in ends of 2014.

---

<sup>1</sup>Data Releases can be found in [sdss3.org](http://sdss3.org)

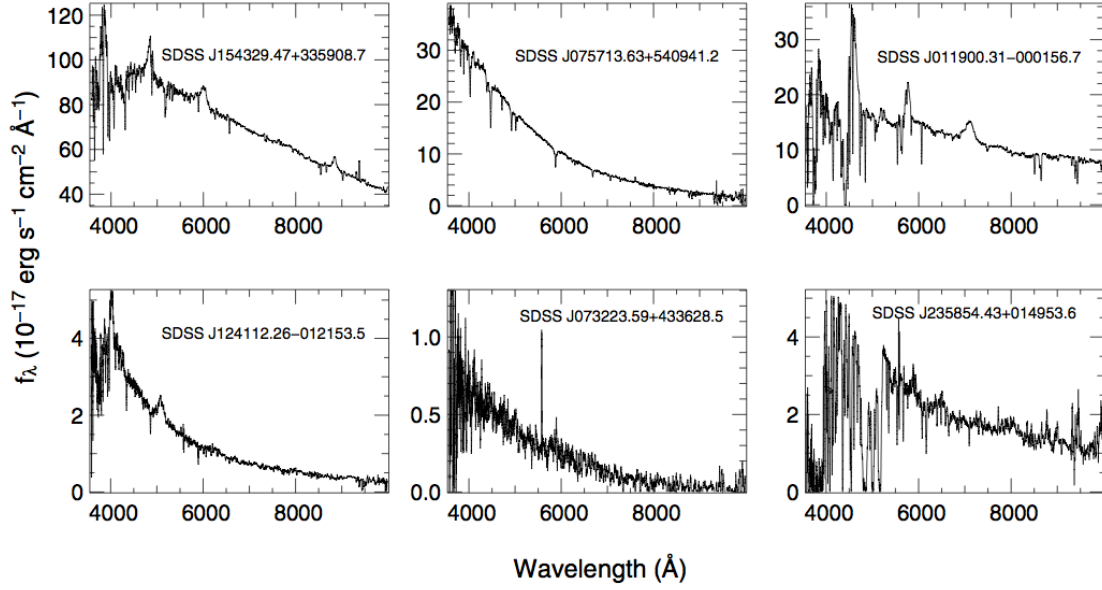


Figure 3.10: First column: examples of  $z > 2$  quasars classified as STAR by the BOSS pipeline. The overall shape of the spectrum is similar to the spectrum of F stars. Second column: examples of stars identified as QSO by the BOSS pipeline. Strong absorption lines or wiggles in the spectrum can mimic quasar features. Third column: examples of  $z > 2$  quasars for which the BOSS pipeline provides an inaccurate redshift estimate that must be corrected during the visual inspection. The pipeline is confused by the strong absorption lines. The spectra were boxcar median smoothed over 5 pixels. Figure extracted from [P aris et al. \(2012\)](#).

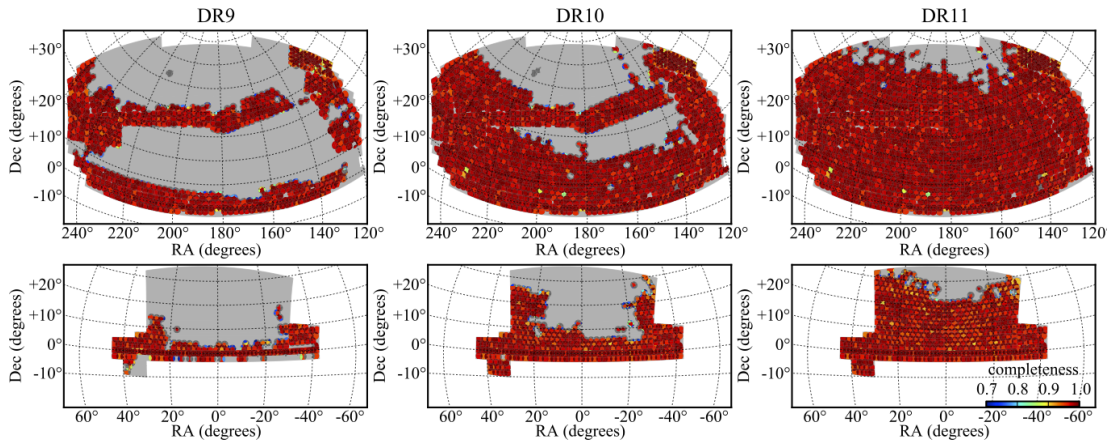


Figure 3.11: Evolution of the BOSS sky coverage from DR9 to DR11. Top panels show our observations in the North Galactic Cap (NGC) while lower panels show observations in the South Galactic Cap (SGC). Colors indicate the spectroscopic completeness within each sector as indicated in the key in the lower right panel. Gray areas indicate our expected footprint upon completion of the survey. The total sky coverage in DR9, DR10, and DR11 is 3,275 deg<sup>2</sup>, 6,161 deg<sup>2</sup>, and 8,377 deg<sup>2</sup>, respectively.

## Chapter 4

# BAO in Galaxy Clustering

In this chapter we discuss how to measure the BAO scale using galaxies as tracers of the matter density field. The basics of galaxy clustering methods are presented in § 4.1. The 1% BAO measurement using BOSS CMASS galaxies is summarized in § 4.2, followed by a description of a new optimal estimator of the correlation function that yields up to 20% improvement in the BAO scale measurement when applied on DR9 data. This new estimator is part of my original contribution of this chapter, published in [Vargas-Magaña et al. \(2013\)](#).

### 4.1 Measuring the galaxy clustering

As introduced in Chapter 1, the clustering of matter is characterized by its power spectrum,  $P(k)$ , or equivalently the correlation function,  $\xi(r)$ , from which the BAO scale can be extracted. Clustering measurements require estimates of the matter density in every point of space. However, dark matter is impossible to observe directly, and baryonic matter is in general too diffuse in space to be observable, except in galaxies or galaxy clusters (observable with X-rays or radio). The hydrogen in the intergalactic medium is observable through the Ly $\alpha$  forest, as discussed in the previous chapter, or through the 21cm emission line in radio, but those measurements are only possible at redshifts larger than  $z \sim 2$ , where observations are still time-consuming. At lower redshift, galaxies and galaxy clusters are still the best observable reservoirs of baryons. Galaxies form in the central regions of dark matter halos, which are themselves large dark matter overdensities ( $\delta\rho/\rho \gg 200$ ). Galaxies are thus the most convenient tracer of the matter distribution at low redshifts ( $z \lesssim 1$ ).

Building a three dimensional map of the galaxy distribution gives information about the underlying dark matter density field. Let  $n(\vec{x})$  be the galaxy number density at a given position  $\vec{x}$  in space. If the mean density of galaxies in the Universe is  $\bar{n}$ , the galaxy number density fluctuations at a given position in space  $\vec{x}$  can be defined in analogous manner as the dark matter overdensities as

$$\delta_n(\vec{x}) = \frac{n(\vec{x}) - \bar{n}}{\bar{n}} . \quad (4.1)$$

Statistical properties of the galaxy density field fluctuations, defined as

$$\xi_g(\vec{x}, \vec{x}') = \langle \delta_n(\vec{x}) \delta_n(\vec{x}') \rangle, \quad (4.2)$$

are a well suited observable to study galaxy clustering and BAO.

Since densities of dark matter and galaxies are related, we also expect a relationship between the galaxy correlation function  $\xi_g$  and the dark matter correlation function  $\xi$ . This relationship between galaxy and matter densities, also known as density bias, depends on the physics of galaxy formation inside halos of a given mass, and on the number of halos as a function of mass (known as mass function). Therefore, the bias is, in general, a complex, non-linear and non-local function. However, physical models (Bardeen et al., 1986, Cole and Kaiser, 1989) and numerical simulations predict a linear constant bias  $b$  on large scales, so that  $\xi_g = b^2 \xi$ . On smaller scales (below  $\sim 10 h^{-1} \text{Mpc}$ ), this linear approximation does not hold anymore; the clustering of galaxies is stronger than dark matter halos due to interactions between galaxies inside single collapsed structures. Many existing models take into account the clustering of galaxies on small scales by modeling dark matter halo profiles and the clustering of galaxies inside halos (Seljak, 2000, Peacock and Smith, 2000, Cooray and Sheth, 2002).

Galaxy surveys are therefore an important tool to measure clustering and had become popular in the last 20 years, because of its power in constraining cosmological models. First galaxy surveys predicted the need of a cosmological constant to explain the observed clustering (Efstathiou et al., 1990). A galaxy survey is simply a three dimensional map of a sample of galaxies. It is very important to have accurate locations of galaxies in order to precisely measure number densities, reducing errors in clustering estimates. In a galaxy survey, the angular position of galaxies is measured very precisely but their radial distance is not directly accessible. Instead of distances, we measure redshifts and these are converted into distances by assuming a fiducial cosmological model. It is crucial to measure redshifts accurately in order to correctly map the radial distribution of the matter. Spectroscopic surveys such as BOSS are able to use emission lines of spectra in order to measure redshifts with typical errors of  $\sim 0.001(1+z)$ . Photometric surveys also are able to measure redshifts but with larger errors. Future photometric surveys such as the Large Synoptic Survey Telescope (LSST) expect to measure redshifts with typical errors of order  $\sim 0.015(1+z)$  (Ivezic et al., 2008).

The use of redshift as radial distance estimates leads to physical effects on the clustering measurements. When transforming redshifts into distances, it is implicitly assumed that the measured recession velocity is only due to the Hubble flow. However, galaxies also have their own peculiar velocities because of their infall towards potential wells. These peculiar velocities are added to the recession velocity caused by the expansion. It is the combined velocity that gives the observed redshift of a given galaxy. When converting redshifts to distances, there is an associated error caused by this peculiar motion. Peculiar velocities thus modify the clustering in the radial direction while the angular clustering is left unchanged. This effect is commonly referred to as *redshift-space distortions* (RSD). The amount of distortion caused by peculiar velocities can be predicted for a given theory of gravity and structure formation. In the limit of large scales and linear theory, RSD can be simply modeled (Kaiser, 1987, Hamilton, 1992) as an enhancement of

Fourier modes of the density field in the radial direction, as discussed previously in Chapter 2 for the case of Ly $\alpha$  forests.

Once the position of the galaxies of the survey is mapped, the galaxy correlation function  $\xi_g(\vec{r})$  can be estimated as a function of the separation  $\vec{r}$  between galaxies. If the survey is sufficiently large, correlations at separations of  $100 h^{-1}\text{Mpc}$  become significant and the BAO peak position can be measured. The BAO peak in the galaxy clustering was measured for the first time in 2005 with  $\sim 50,000$  SDSS-II luminous red galaxies over a comoving volume of  $0.72 h^{-3}\text{Gpc}^3$  (Eisenstein et al., 2005). Concurrently the Two Degree Field Galaxy Redshift Survey (2dFGRS) also reported the detection of BAO in the galaxy power spectrum (Cole et al., 2005).

We describe now how to estimate the correlation function from a given galaxy survey. In principle it is necessary to know the average density of galaxies  $\bar{n}$  in order to compute galaxy density fluctuations  $\delta_n$  over the survey. A given survey observes a finite volume of the Universe set by the sky coverage and the galaxy selection criteria, and this needs to be correctly taken into account when computing the average density  $\bar{n}$ . Also, the completeness and purity of observations affect these estimates. The effective observed volume is often referred as the survey *mask*. In order to correct for boundary effects, uncorrelated mock galaxies are randomly distributed over the survey mask. Instead of explicitly computing  $\bar{n}$  over the survey, the correlations of real galaxies are measured by “comparison” with the correlations of this random catalog of galaxies (expected to be zero).

Correlations are measured by counting pairs of galaxies at a given comoving separation. Let  $DD(\vec{s})$  be the number of pairs of galaxies of the survey separated by a comoving distance  $\vec{s} = \vec{x} - \vec{x}'$ , normalized by the total number of unique pairs in the survey (if the survey contains  $N_d$  galaxies, then the normalization is  $N_d(N_d - 1)/2$ ). If galaxies are clustered, we expect to observe more pairs of galaxies at smaller separations than at larger separations compared to a uniform distribution of galaxies. The random galaxy catalog allows us to make this comparison. Let  $RR(\vec{s})$  be the normalized number of randomly distributed galaxy pairs as a function of separation. An estimator of the correlation function can be written as

$$\hat{\xi}(\vec{s}) = \frac{DD(\vec{s})}{RR(\vec{s})} - 1. \quad (4.3)$$

The expected value,

$$\langle 1 + \hat{\xi}(\vec{s}) \rangle = \frac{1 + \xi(\vec{s})}{1 + \xi_\Omega}, \quad (4.4)$$

shows that the estimator is biased (Landy and Szalay, 1993), where  $\xi_\Omega$  is the average of the correlation function over the mask, also known as integral constraint. This bias appears because the average number of galaxies  $\bar{n}$  is estimated from the survey itself, a procedure that might incorrectly take into account very large scale fluctuations. Landy and Szalay defined an unbiased estimator as

$$\hat{\xi}(\vec{s}) = \frac{DD - 2DR + RR}{RR}(\vec{s}), \quad (4.5)$$

where  $DR(\vec{s})$  is the number of “cross” pairs between random and real galaxies. This estimator also has smaller errors than those given by the estimator defined in Eq. 4.3. This is the most commonly used estimator for galaxy clustering measurements and BAO. The BAO measurement



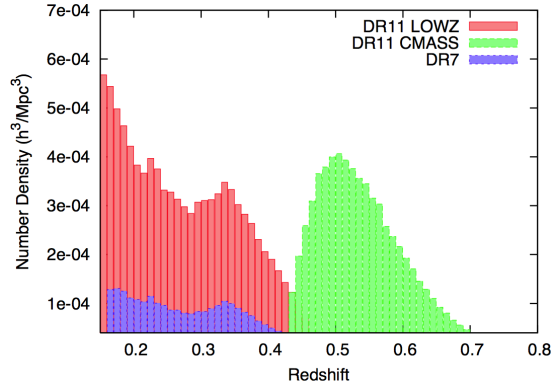


Figure 4.1: Histograms of the galaxy number density as a function of redshift for LOWZ (red) and CMASS (green) samples we analyse. We also display the number density of the SDSS-II DR7 LRG sample in order to illustrate the increase in sample size provided by BOSS LOWZ galaxies.

using galaxy clustering was performed by BOSS and is described in the next section (§ 4.2). In § 4.3, we show that an even more precise correlation function estimator can be constructed, yielding  $\sim 20\%$  improvement on BAO peak position measurements.

## 4.2 BAO with BOSS galaxies

### 4.2.1 The galaxy samples

One of the main goals of the BOSS project was to measure BAO to 1% using galaxy clustering. To achieve this goal, more than 70% observing time was dedicated to galaxies. As described in § 3.7, BOSS selected two samples of galaxies from photometric measurements in order to measure spectra and redshifts: the LOWZ sample ( $0.15 < z < 0.43$ ) and the CMASS sample ( $0.43 < z < 0.7$ ), totalizing more than 1.35 million galaxy targets. Figure 4.1 summarizes the redshift distribution of those samples used for BAO measurements.

After the end of spectroscopic observations, a small fraction of the targets was not observed and some spectrum were discarded. The reasons for a given target not being observed are:

- SDSS-II already obtained a good redshift for the object; these are denoted as *known*;
- a target of different type (e.g., a quasar) is within  $62''$  (the angular size of sky region measured by a single fiber); these are denoted as *missed*;
- another target of the same type is within  $62''$ ; these are denoted *cp* for “close-pair”;

while a spectrum is discarded when :

- The spectrum reveals that the object is a star (characterizing an error in the target selection procedure); these are denoted *star*;

- The automatic data reduction pipeline fails the redshift measurement of the galaxy spectrum; these are denoted *fail*.

### 4.2.2 Masks

Using the MANGLE software (Swanson et al., 2008) the sky areas covered by BOSS could be tracked for completeness estimations. This software divides the survey footprint into spherical polygonal sectors. The completeness in each sector is defined as

$$C_{BOSS} = \frac{N_{\text{obs}} + N_{cp}}{N_{\text{targ}} - N_{\text{known}}} , \quad (4.6)$$

where  $N$  is the number of objects in the sector, “obs” denotes observed and “targ” denotes target. For the analysis are discarded any sector where  $C_{BOSS} < 0.7$ . About 5% of the surveyed sky surface was masked after all cuts. Figure 3.11 shows the footprint of BOSS divided in sectors with their completeness.

### 4.2.3 Weighting galaxies

Each galaxy is weighted to take into account the problems listed previously. A galaxy is up-weighted if its nearest neighbor (of the same target class) had a redshift failure ( $w_{zf}$ ) or a redshift of that neighbor was not obtained because it was in a close-pair ( $w_{cp}$ ).

For CMASS, it was found that the galaxy number density is significantly anti-correlated with the stellar density and seeing (Ross et al., 2011, Ho et al., 2012). Basically the sky regions containing more stars hide possible galaxies behind them. Seeing also reduces the quality of photometric observations thus reducing the number of galaxy targets at high redshift. To take into account those effects additional weights,  $w_{star}$  and  $w_{see}$ , were added to each galaxy. In summary, each galaxy is counted as

$$w_{tot} = (w_{cp} + w_{zf} - 1)w_{star}w_{see} . \quad (4.7)$$

In addition to this systematic weight, Feldman, Kaiser and Peacock proposed to weight galaxies in order to optimize the signal to noise of the measurement based on the clustering amplitude (Feldman et al., 1994). The FKP weights are defined as

$$w_{FKP} = \frac{1}{1 + \bar{n}(z_i)P_0} , \quad (4.8)$$

where  $\bar{n}(z_i)$  is the mean density at redshift  $z_i$  and  $P_0 = 20,000 h^{-3}\text{Mpc}^3$  is an approximated value for the power spectrum that optimizes the BAO measurement. This ignores the scale dependence of the power spectrum, but this approximation does not affect final results and errors.

Using these weights, we define the effective volume of the survey by

$$V_{eff} = \sum_i \left( \frac{\bar{n}(z_i)P_0}{1 + \bar{n}(z_i)P_0} \right)^2 \Delta V(z_i) , \quad (4.9)$$



where  $\Delta V(z_i)$  is the effective volume of a redshift shell. The sum is performed over 200 redshift shells. For DR11, the effective volume of the LOWZ sample is  $2.4 \text{ Gpc}^3$  and for CMASS,  $6.6 \text{ Gpc}^3$ .

#### 4.2.4 The measurements

The spherically averaged correlation function  $\xi(s)$  (or monopole) and its analogous in Fourier space, the power spectrum  $P(k)$ , were estimated from the same galaxy catalog. We discuss here only isotropic measurements from [Anderson et al. \(2013\)](#). Details about anisotropic BAO measurements using galaxies can be found in [Vargas Magaña et al. \(2013\)](#), [Beutler et al. \(2013\)](#), [Samushia et al. \(2014\)](#), [Chuang et al. \(2013\)](#), [Sánchez et al. \(2014\)](#).

To compute distances from redshifts and for the model fitting described later, we used a fiducial cosmology following a flat  $\Lambda$ CDM model with  $\Omega_m = 0.274$ ,  $h = 0.7$ ,  $\Omega_b h^2 = 0.0224$ ,  $n_s = 0.95$  and  $\sigma_8 = 0.8$ .

The correlation function  $\xi(s)$  was computed using the Landy-Szalay estimator (Eq. 4.5), summing pair-counts into bins of width  $8 \text{ } h^{-1}\text{Mpc}$  spaced between  $29 < s < 200 \text{ } h^{-1}\text{Mpc}$  (22 bins). To calculate  $P(k)$ , we employed the FKP estimator ([Feldman et al., 1994](#)) in a Fourier grid of size  $2048^3$ ,  $4000 \text{ } h^{-1}\text{Mpc}$  along each side enclosing the survey, including both Northern and Southern Galactic Caps (NGC and SGC). The power spectrum was computed averaging mode amplitudes in bins of  $\Delta k = 8 \times 10^{-3} \text{ } h\text{Mpc}^{-1}$ , between  $0.02 < k < 0.3 \text{ } h\text{Mpc}^{-1}$  (35 bins).

Non-linear structure formation effects reduce the statistical sensitivity of the BAO measurement because their effect is to smooth the correlation function (and the BAO peak) ([Bassett and Hlozek, 2009](#)). Therefore, we used a procedure to *reconstruct* the linear density field ([Eisenstein et al., 2007](#)) and revert partially the effects of non-linear growth and large-scale peculiar velocities. The idea is to use the predicted relation between density and displacements in Lagrangian perturbation theory and, using the measured density field of our galaxy catalog, move each galaxy by an small amount in the opposite sense of the displacement field. The reconstruction technique improves the precision of the BAO scale measurements by a factor of 1.54 in average for BOSS (value estimated from mocks), as already shown in [Padmanabhan et al. \(2012\)](#), [Anderson et al. \(2012\)](#). This improvement depends on how well the density of galaxies can be estimated in the survey. A survey with a large shot-noise power component could not benefit from reconstruction.

Mock catalogs were created following the PTHalos methodology ([Manera et al., 2013](#)) for both the LOWZ and CMASS samples. These mock catalogs are based on second-order Lagrangian theory to create the matter density fields, using a empirical halo occupation function to populate halos with BOSS-like galaxies. The mocks reproduce the correlation function of the data catalog over the scales of interest and reproduce the redshift distribution and completeness of the data samples. A total of 1000 mocks were produced and analyzed for the LOWZ sample and 600 for CMASS. The cosmological model used to create these mock catalogs is the same as the one used to analyze data, presented previously.

As in previous galaxy clustering studies, covariance matrices for  $\xi(s)$  and  $P(k)$  are defined by the scatter of  $\xi(s)$  and  $P(k)$  around the mean, estimated from all mock catalogs. The full detailed

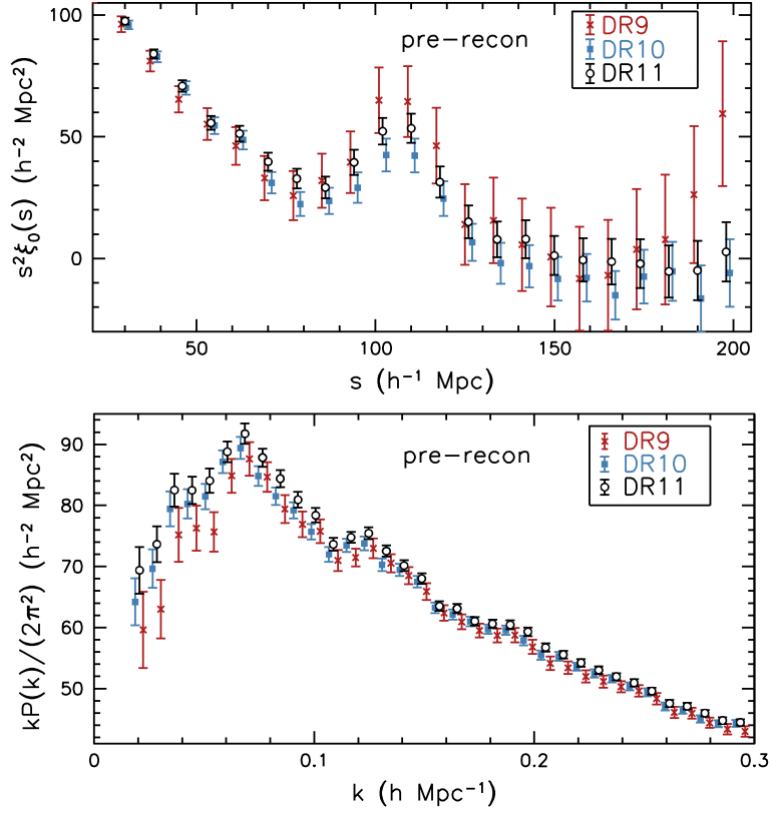


Figure 4.2: Top panel: The measured monopole of the CMASS galaxy correlation function, multiplied by the square of the scale,  $s$ , for each of the BOSS data releases. These figures are shown pre-reconstruction. For clarity, the DR10 data have been shifted horizontally by  $+1 h^{-1}\text{Mpc}$  and the DR9 data by  $-1 h^{-1}\text{Mpc}$ . Bottom panel: The measured spherically averaged CMASS galaxy power spectrum, multiplied by the frequency scale,  $k$ , for each of the BOSS data releases. For clarity, the DR9 data have been shifted by  $+0.002 h\text{Mpc}^{-1}$  and the DR10 data by  $-0.002 h\text{Mpc}^{-1}$ .

procedure is described in Percival et al. (2013). Also, the errors on the covariance matrix estimate from mocks were propagated into the covariance matrix itself.

Figure 4.2 shows the estimates of  $\xi(s)$  and  $P(k)$  described in this section for three samples in increasing order of number of galaxies : DR9, DR10 and DR11. The first thing to note is the clear presence of the BAO both in the correlation function, a peak near  $100 h^{-1}\text{Mpc}$ , and in the power spectrum, giving the oscillating pattern. The second remark is that error bars are significantly reduced between DR9 and DR11 samples by factor of  $\sim 2$ .

#### 4.2.5 Fitting for isotropic BAO

The BAO position in spherically averaged 2-point measurements is fixed by the comoving projection of the sound horizon at the drag epoch,  $r_d$ , and provides a measure of

$$D_V(z) \equiv [z(1+z)^2 D_A^2(z) D_H(z)]^{1/3}, \quad (4.10)$$

where  $D_A$  and  $D_H$  were defined in § 1.5. If we measure our 2-point functions using a fiducial cosmological model, then, to very good approximation, the observed BAO position depends on the scale dilation parameter

$$\alpha \equiv \frac{D_V(z)}{r_d} \left( \frac{r_d}{D_V(z)} \right)_{\text{fid}}, \quad (4.11)$$

which simply tells us about the relative position of the BAO peak with respect to the expected position in the fiducial cosmology. If  $\alpha > 1$ , the actual BAO peak is shifted towards scales smaller than expected, and vice-versa. This is the parameter that is fitted over the clustering measurements to extract the BAO position.

We describe now the models used to fit the measured correlation function and power spectrum. For both, we fit a template given by the fiducial cosmology re-scaled by the  $\alpha$  parameter, with an added broadband smooth function to take into account possible systematic effects and marginalize any information coming from the full shape. The idea is to fit only for the BAO peak position.

To compute the power spectrum template, we first compute the linear power spectrum  $P^{\text{lin}}$  produced by CAMB<sup>1</sup> (Lewis et al., 2000). We split it into two components, one oscillatory  $O^{\text{lin}}$  and the other smooth  $P^{\text{sm,lin}}$ , that return the original power spectrum when multiplied together. The full model fitted to the power spectrum is

$$P^{\text{fit}}(k) = \left[ B_P^2 P^{\text{sm,lin}}(k) + A_1 k + A_2 + \frac{A_3}{k} + \frac{A_4}{k^2} + \frac{A_5}{k^3} \right] \left[ 1 + (O^{\text{lin}}(k/\alpha) - 1) e^{-\frac{1}{2}k^2 \Sigma_{nl}^2} \right], \quad (4.12)$$

where the  $A_i$  and  $B_P$  are additive and multiplicative nuisance parameters to be marginalized over. The non-linear effects are modeled by the Gaussian damping kernel with characteristic scale  $\Sigma_{nl}$  that is also fitted using a Gaussian prior of width  $4 h^{-1}\text{Mpc}$  centered at  $8.3 h^{-1}\text{Mpc}$  (for pre-reconstruction) and  $4.6 h^{-1}\text{Mpc}$  (post-reconstruction). Those prior values were derived from fits over mocks.

The template for the fit of the correlation function  $\xi^{\text{mod}}(s)$  is obtained by the Fourier transform of the power spectrum model  $P^{\text{mod}}$  given by

$$P^{\text{mod}}(k) = P^{\text{nw}}(k) \left[ 1 + \left( \frac{P^{\text{lin}}(k)}{P^{\text{nw}}(k)} - 1 \right) e^{-\frac{1}{2}k^2 \Sigma_{nl}^2} \right], \quad (4.13)$$

where  $P^{\text{nw}}$  is a “de-wiggled” template power spectrum, where the BAO oscillations are subtracted out (Eisenstein and Hu, 1998). The correlation function model is

$$\xi^{\text{fit}}(s) = B_\xi^2 \xi^{\text{mod}}(\alpha s) + a_1 + \frac{a_2}{s} + \frac{a_3}{s^2}. \quad (4.14)$$

Once again, the  $a_i$  and the  $B_\xi$  are nuisance parameters to marginalize out the broadband shape of our measurement.

Each model was fitted over observations assuming that they were drawn from multi-variate Gaussian distributions, finding the combination of parameters that minimizes the  $\chi^2$  function.

<sup>1</sup><http://camb.info>

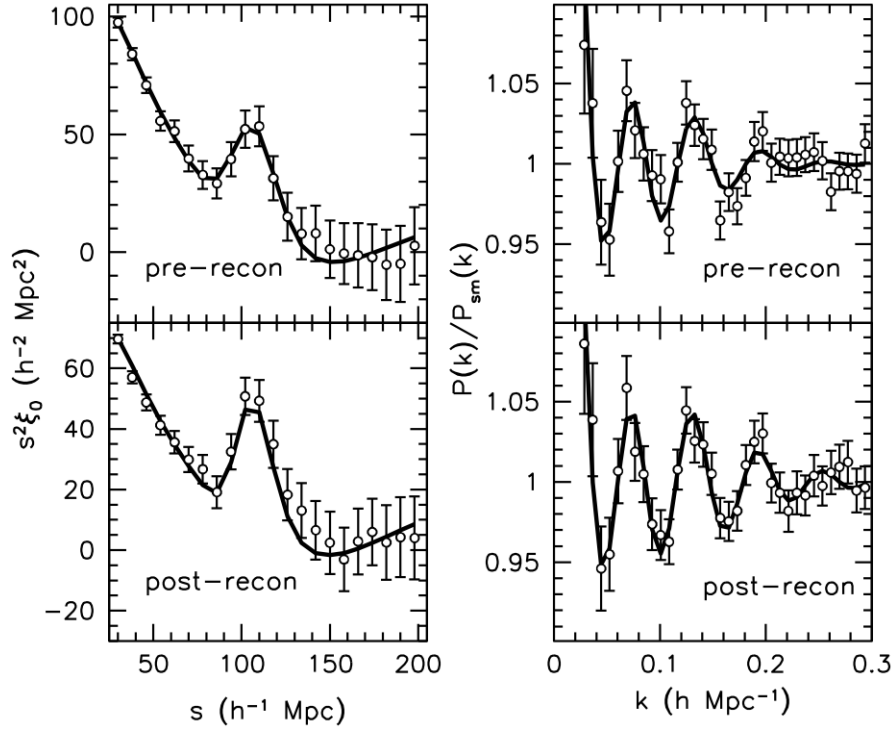


Figure 4.3: DR11 CMASS clustering measurements (black circles) with  $\xi(s)$  shown in the left panels and  $P(k)$  in the right panels. The top panels show the measurements prior to reconstruction and the bottom panels show the measurements after reconstruction. The solid lines show the best-fit BAO model in each case. One can see that reconstruction has sharpened the acoustic feature considerably for both  $\xi(s)$  and  $P(k)$ .

#### 4.2.6 Results

In order to test our modeling, mock catalogs were also fitted. The average  $\alpha$  is consistent with unity for both the correlation function and power spectrum, before and after reconstruction procedure is applied. The scatter around the mean  $\alpha$  of about 1.6% for the pre-reconstruction and 0.92% for the post-reconstruction case. This shows that we expect to obtain a measurement of  $\alpha$  at sub-percent precision on real data. Also, this scatter is consistent with the mean error estimated for each individual realization. This shows that our analysis chain does not lead to biased results, with correct estimation of errors, allowing us to apply it on real data.

Figure 4.3 shows the resulting fit over real data correlation function and power spectrum, for the pre and post-reconstruction cases on the CMASS sample. The sharpening of the BAO features is clearly seen in both estimators, justifying the observed increase in accuracy observed with the recovered  $\alpha$  for mock catalogs. Table 4.1 summarizes the isotropic measurements of DR11 for CMASS and LOWZ samples. For LOWZ we only show the result of the combination of  $\xi(s)$  and  $P(k)$ . We prefer to not detail here how the combination is performed, but we show the results to give an idea of the errors achieved. Indeed, we see that for the CMASS sample we reach sub-percent precision for the post-reconstruction case, for the correlation function and power spectrum.

The errors shown in Table 4.1 do not include possible systematic errors. In order to estimate the systematic contribution to errors, tests were performed by changing fitting methods, binning and

Table 4.1: Isotropic BAO scale measurements recovered from BOSS data. The errors quoted are purely statistical. The  $\chi^2$  values for the LOWZ sample are not shown because it refers to a combined value of  $\alpha$  from different estimators.

Estimator	$\alpha$	$\chi^2/\text{dof}$
DR11 CMASS ( $z = 0.57$ )		
post-recon $P(k)$	$1.0114 \pm 0.0093$	18/27
post-recon $\xi(s)$	$1.0209 \pm 0.0091$	16/17
pre-recon $P(k)$	$1.023 \pm 0.015$	33/27
pre-recon $\xi(s)$	$1.031 \pm 0.013$	14/17
DR11 LOWZ ( $z = 0.32$ )		
$P(k) + \xi(s)$	$1.018 \pm 0.020$	–

fitting to NGC and SGC separately. We found that the  $\alpha$  value does not change considerably in any of those tests. However, for some of them we see an increase in the errors by  $\sim 10\%$  in average. Therefore we decided to increase the final consensus error on  $\alpha$  by 10% to account for those possible systematics.

Even though the procedure is not described here, the BOSS collaboration also performed BAO measurements on the anisotropic correlation function and power spectrum, where effects of redshift space distortions are seen. Different analysis chains were developed inside the collaboration (Beutler et al., 2013, Samushia et al., 2014, Chuang et al., 2013, Sánchez et al., 2014) yielding consistent results for the radial and transverse dilation parameters  $\alpha_{\parallel}$  and  $\alpha_{\perp}$ . The consensus values given by Anderson et al. (2013) are

$$\alpha_{\parallel} = 0.968 \pm 0.033 \quad \text{and} \quad \alpha_{\perp} = 1.045 \pm 0.015 \quad (4.15)$$

The cosmological implications of those measurements are discussed in Chapter 10, after the presentation of BAO results using the Ly $\alpha$  forests of BOSS quasars.

### 4.3 An Optimal Estimator

The two point correlation function is one of the main statistical tools for the study of large scale structures. As explained previously, the idea is not to directly measure the galaxy number density within the survey volume but sample this density with galaxy locations. The observable is the average number of galaxy pairs at a given separation. However, galaxies near the edges of the survey have naturally less neighbors than they should; this observable becomes biased, which needs to be corrected for in an optimal way. This issue does not occur for instance when measuring the correlation function using absorption measurements on the Ly $\alpha$  forests because in principle they directly translate the amplitude of the underlying density field (see Chapter 2).

In this section, a novel estimator for the two-point correlation function of galaxies is introduced. Its performance can be optimized for a given galaxy survey geometry. First, I motivate this effort, showing that various well-known estimators for the two-point correlation function have

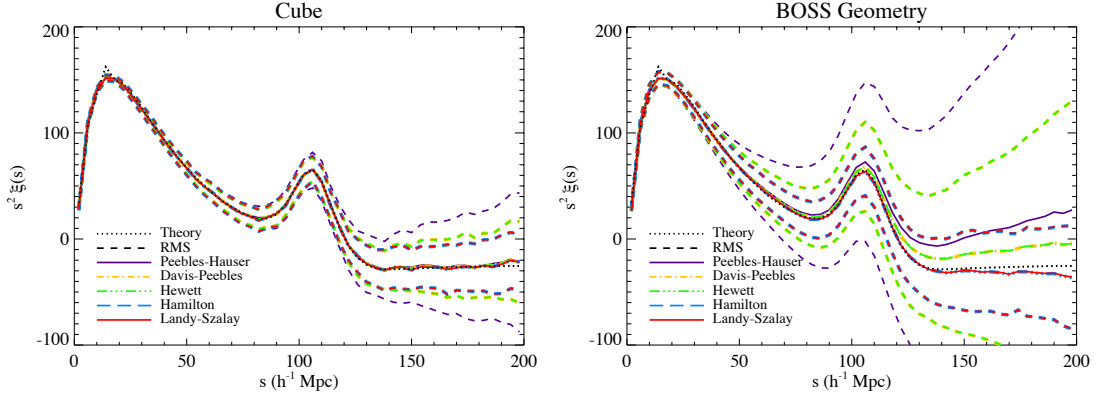


Figure 4.4: Input two-point correlation function (dotted black line) and reconstructed ones using the various estimators available in the literature (solid lines of various colors) for a cubic geometry (left) or a realistic (BOSS DR9) survey volume (right). The dashed lines of various colors represent the RMS of the corresponding estimators. Note that the Hamilton and Landy-Szalay lines are exactly superposed as well as the Davis-Peebles and Hewett lines.

bias and variance strongly dependent on the survey geometry. The commonly used Landy-Szalay estimator has been shown to be both unbiased and of minimal variance in the limit of a vanishing correlation function. We show that, in the cases of our interest (where the correlation function is not zero), the Landy-Szalay estimator does not reach the Poisson noise limit. The novel estimator reaches lower errors compared to Landy-Szalay.

For pedagogical reasons, we start with a simpler but biased version of our optimal estimator and then we develop a simple iterative procedure that allows the final estimator to be unbiased and model independent, while improving the accuracy by around 20-25% with respect to the Landy-Szalay estimator. Then, an application of our estimator on the DR9 CMASS sample is performed to show the improvement on the two-point correlation function measurement and cosmological parameters with respect to previous analyses.

### 4.3.1 Commonly used estimators

Estimators of the two-point correlation function  $\xi(s)$  ( $s$  being the comoving separation) have been studied by various authors (Landy and Szalay, 1993, Peebles and Hauser, 1974, Hewett, 1982, Davis and Peebles, 1983, Hamilton, 1993). Generically, pair counts in data are compared to pair counts in random samples that follow the geometry of the survey. Let's assume a catalog of  $n_d$  objects in the data sample and  $n_r$  in the random sample and calculate three sets of numbers of pairs as a function of the binned comoving separation  $s^2$ :

- within the data sample, leading to  $dd(s)$  that can be normalized to the total number of pairs as  $DD(s) = \frac{dd(s)}{n_d(n_d - 1)/2}$ .

<sup>2</sup>The number of pairs can be spherically averaged in the simplest approach. Its dependence on the angle with respect to the line of sight can be considered in a more elaborated analysis, in order to account for the sensitivity to angular distance in the transverse direction and  $H(z)$  in the radial one (see Cabre and Gaztanaga (2008) for details).

- within the random sample, leading to  $rr(s)$  normalized as  $RR(s) = \frac{rr(s)}{n_r(n_r - 1)/2}$ .
- among both samples (cross correlation) leading to  $dr(s)$  normalized as  $DR(s) = \frac{dr(s)}{n_r n_d}$ .

The most common estimators discussed in the literature are:

- $\hat{\xi}_{PH}(s) = \frac{DD}{RR} - 1$  (Peebles and Hauser, 1974)
- $\hat{\xi}_{Hew}(s) = \frac{DD - DR}{RR}$  (Hewett, 1982)
- $\hat{\xi}_{DP}(s) = \frac{DD}{DR} - 1$  (Davis and Peebles, 1983)
- $\hat{\xi}_H(s) = \frac{DD \times RR}{DR^2} - 1$  (Hamilton, 1993)
- $\hat{\xi}_{LS}(s) = \frac{DD - 2DR + RR}{RR}$  (Landy and Szalay, 1993)

The usual estimator (Landy and Szalay, 1993) is given by Eq. 4.5 is an unbiased and minimum variance estimator in the case of a *null* correlation function. Since galaxy surveys are increasingly large and dense yielding significant correlation amplitudes, this estimator is no longer a minimum variance estimator though still unbiased. We provide a way to empirically compute a biased minimum variance estimator, and an iterative method to estimate the bias and correct for it.

### 4.3.2 An optimized estimator

Given that common estimators of the correlation function are linear combinations of ratios of pair counts ( $DD$ ,  $DR$  and  $RR$ ), we start by writing a generalized estimator:

$$\hat{\xi} = \sum_i c_i \mathbf{R}_i, \quad (4.16)$$

where each  $c_i$  is a coefficient (non-necessarily integer) associated to the  $i$ -th ratio,

$$\mathbf{R}_i = DD^m DR^n RR^{-m-n},$$

which is a generalized ratio of pair counts. We can classify these ratios as being of  $k$ -th order if the integers  $m$  and  $n$  take values between  $-k$  and  $k$ . Consequently, there is a single possible 0-th order ratio ( $\mathbf{R}_0 = 1$ ), six 1st order ratios and twelve 2nd order ratios. Table 4.2 summarizes those ratios. We do not use 3rd or higher order ratios in this study. Our generalized estimator is thus a linear combination of 19 ratios ( $i \in [0, 18]$  in Eq. 4.16). As an illustration, the Landy-Szalay estimator is defined a linear combination of  $DD/RR$ ,  $DR/RR$  and unity, with coefficients given by 1,  $-2$  and 1, respectively. For clarity, the dependence of correlation functions on separation is taken into account by considering  $\hat{\xi}$ ,  $\mathbf{R}_i$ ,  $DD$ ,  $DR$  and  $RR$  as vectors, where each component corresponds to a separation bin.

Table 4.2: The nineteen ratios formed by data-data ( $DD$ ), data-random ( $DR$ ) and random-random ( $RR$ ) galaxy pair counts up to 2nd order.

0th order		
1		
1st order terms		
$\frac{DD}{RR}$	$\frac{DR}{RR}$	$\frac{DR}{DD}$
$\frac{RR}{DD}$	$\frac{RR}{DR}$	$\frac{DD}{DR}$
2nd order terms		
$\frac{DR \times RR}{DD^2}$	$\frac{RR^2}{DD^2}$	$\frac{DR \times DD}{RR^2}$
$\frac{DD^2}{RR^2}$	$\frac{DR^2}{RR^2}$	$\frac{DD^2}{DR^2}$
$\frac{RR^2}{DR^2}$	$\frac{DD \times RR}{DR^2}$	$\frac{RR^2}{DD \times DR}$
$\frac{DR^2}{DD \times RR}$	$\frac{DD^2}{DR \times RR}$	$\frac{DR^2}{DD^2}$

The next step consists in finding the coefficients  $c_i$  such that the estimator in Eq. 4.16 has minimal error. We use numerical simulations to perform find these coefficients. First, we produce hundreds of mock realizations of the galaxy survey following a given input correlation function  $\xi_{\text{th}}$ . Second, for each realization we compute the pair counts  $DD$ ,  $DR$ , and  $RR$ , and the nineteen ratios  $\mathbf{R}_i$ . Third, we search for the set of coefficients  $c_i$  that minimizes the scatter of the estimated correlation function (Eq. 4.16) around the input mock correlation function  $\xi_{\text{th}}$ . This last step is translated as a minimization of

$$\chi^2 = \sum_j \left[ \hat{\xi}_j - \xi_{\text{th}} \right]^T \cdot C^{-1} \cdot \left[ \hat{\xi}_j - \xi_{\text{th}} \right] , \quad (4.17)$$

with respect to the coefficients  $c_i$  (implicit in the definition of  $\hat{\xi}_j$ ). The  $j$  index stands for the  $j$ -th mock realization.  $C$  is the covariance matrix of our estimate, which needs to be estimated before the minimization. However, we do not have in advance our final estimate of  $\xi$ . Therefore, we estimated  $C$  using the scatter around the mean of measurements using the Landy-Szalay estimator.

The estimator  $\hat{\xi}$  defined by the optimal set of coefficients  $c_i$  might be biased, even though it is a minimal variance estimator by construction. It means that, using our estimator, the mean correlation function of the mock catalogs differs from the input model ( $\langle \hat{\xi} \rangle \neq \xi_{\text{th}}$ ). We refer to this bias as *residual bias*, distinguishing it from the physical density bias. If the amount of this mean residual bias is known, it is possible to correct each estimation for it, obtaining an unbiased minimal variance estimator. We show an example of the residual bias in the next section. In § 4.3.4 we propose an iterative method to circumvent the residual bias problem.



### 4.3.3 Performance on lognormal mock catalogs

To test the optimal estimator (Eq. 4.16), we used mock galaxy catalogs. Using the footprint and the redshift distribution of DR9 CMASS galaxies (Figures 3.11 and 4.1), we produced 120 mock catalogs where the density field is assumed to be lognormal and following a flat  $\Lambda$ CDM correlation function. The input cosmological parameters are  $h = 0.7$ ,  $\Omega_m = 0.27$ ,  $\Omega_b = 0.045$ ,  $\sigma_8 = 0.8$ , and  $n_s = 1.0$ . Galaxies are drawn following a Poisson statistics using the value of density field as the mean of the Poisson distribution. Catalogs with galaxies randomly distributed in the same survey geometry were also produced to compute  $RR$  and  $DR$  terms. Our random catalogs contain 30 times more galaxies than correlated galaxy catalogs.

We proceeded to the construction of the optimal estimator. Following the steps described in the previous section, we computed for each mock catalog the pair counts  $DD$ ,  $DR$  and  $RR$ , and the ratios  $\mathbf{R}_i$  as functions of comoving separation. We considered separations between 0 and 200  $h^{-1}$ Mpc, using 50 bins 4  $h^{-1}$ Mpc wide. The correlation function using Landy-Szalay estimator was computed for each realization. Using the 120 estimated correlation functions, we calculated the covariance matrix  $C_{LS}$  needed in Eq. 4.17. The set of coefficients  $c_i$  defining the optimal estimator were found by minimizing the  $\chi^2$ , using only bins over separations between 40 and 200  $h^{-1}$ Mpc (typical on BAO analyses).

We computed the correlation function using the optimal estimator over all mock realizations. The average correlation function over realizations,  $\langle \hat{\xi} \rangle$ , and its covariance matrix,  $C_{opt}$ , were also computed. Figure 4.5 shows the residual bias,  $B = \langle \hat{\xi} \rangle - \xi_{th}$ , and the root mean squared errors (RMS, defined as the square root of the diagonal terms of the covariance matrix), for both the Landy-Szalay and our optimal estimator. We observe that the scatter of correlations measured with the optimal estimator is  $\sim 30\%$  smaller over most of the separations range. However, the optimal estimator has a residual bias that is also non-smooth, with a BAO peak shaped feature at the exact same position as the BAO peak in the correlation function. Even though this residual bias is smaller than the scatter of the measurements, it is worrying because of the peak shape that might introduce systematic errors on the measurement of the BAO peak position in real data.

We checked for the dependence of the residual bias on the peak position. We produced 9 sets of 120 mock galaxy catalogs. For each set of mocks, the BAO peak position was slightly shifted in separation. We observed that, when computing the optimal estimator in each set separately, the BAO peak shape position in the residual bias is the same as in the input correlation function.

The mean residual bias of the optimal estimator is meant to be used as a correction for each individual estimation of the correlation function. However, if the “true” BAO peak position of the sample we are measuring differs from the input peak position of our simulations, correcting for the residual bias could move the peak and change final results. This is indeed observed in Fig. 4.6, where the squares and crosses show the systematic error in recovering the input value of the BAO peak position, characterized by the  $\alpha$  dilation parameter (the fitting procedure for  $\alpha$  is described in the next section). We see that the optimal estimator in this form cannot be used for measurements of the BAO peak. In the next section, we describe an iterative method to

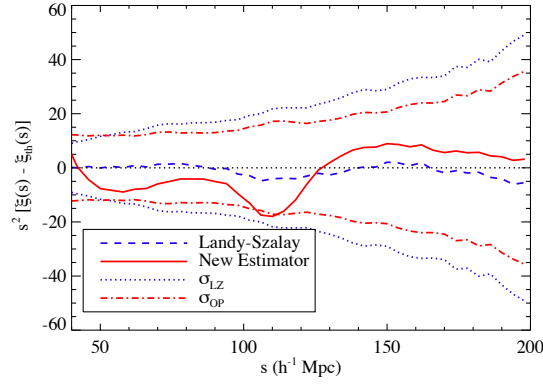


Figure 4.5: Residuals with respect to the input correlation function for the Landy-Szalay estimator (dashed-blue) and for the minimum variance estimator (solid-red). We also show the corresponding  $\pm 1\sigma$  (RMS from 120 lognormal simulations) for Landy-Szalay (dotted-blue) and for the minimum variance estimator (dot-dashed-red). Residuals and RMS are rescaled by  $s^2$ . The optimal estimator has been obtained by fitting for nineteen coefficients  $c_i$  of Eq. 4.16, limiting the  $\chi^2$  in Eq. 4.17 to the region  $[40, 200] h^{-1}\text{Mpc}$ . This range has been chosen accordingly to the typical range used in BAO analysis.

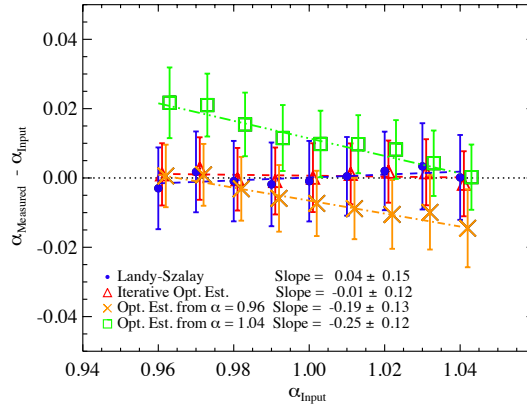


Figure 4.6: Bias on  $\alpha_{\text{Measured}}$  as a function of  $\alpha_{\text{Input}}$  (from lognormal simulations) for Landy-Szalay (blue points), iterative optimal (red triangles) and two non-iterative minimum-variance estimators with  $\alpha = 0.96$  (orange crosses) and  $\alpha = 1.04$  (green squares). The error bars indicate the RMS among the lognormal simulations. Simple linear functions were fitted to the points in each case with corresponding slopes shown in the legend. A strong bias can be seen for the non-iterative estimators while the iterative optimal and Landy-Szalay estimators are not biased.

circumvent this model dependence while keeping errors smaller than those of the Landy-Szalay estimator.

#### 4.3.4 The iterative optimal estimator

In order to correct for the model dependent residual bias of the optimal estimator, we developed an iterative method that avoid systematic errors on the estimation of the BAO peak position in the galaxy correlation function.

Basically the idea is to iterate between two steps : 1) constructing a new optimal estimator from mocks with a BAO peak at a given position, and 2) getting the new BAO peak position from the measured correlation function, while correcting for the residual bias. In other words, on each iteration the fitted BAO peak position is used to create a new set of mock catalogs from which a new optimal estimator is constructed. We iterate until the convergence of the measured value of the BAO peak position. This procedure avoids the use of a optimal estimator build on a set of mocks with a BAO peak shifted with respect to the peak from the data.

The fitting procedure of this iterative process is similar to the procedure presented in § 4.2. As before, we work only with the spherically averaged correlation function, also known as the monopole of the 3D correlation function. The model  $\xi_m$  fitted over the estimated correlation function is a defined as

$$\xi_m(r) = b^2 \xi_t(\alpha r) + a_0 + \frac{a_1}{r} + \frac{a_2}{r^2}, \quad (4.18)$$

which is a sum of two functions: a scaled cosmological template and a broadband smooth function. The template  $\xi_t$  is obtained from our fiducial cosmological model using CAMB. This template is scaled by the dilation parameter  $\alpha$  in the separations direction and has its amplitude set by a bias parameter  $b$ . Our broadband model is the same as in § 4.2. The scaling factor  $\alpha$  in the template is the parameter containing the information about the BAO peak position relative to the prediction of the cosmological model used to build the template. The model  $\xi_m$  contains five free parameters. We perform the fit over the measured correlation function using separations between 40 and 200  $h^{-1}\text{Mpc}$ , yielding 35 degrees of freedom.

Once the BAO peak position is measured through the  $\alpha$  parameter, the next step is to create a new set of mock catalogs using an input correlation function with the BAO peak placed at the last estimated position. However, instead of creating hundreds of mock catalogs at each iteration which would be computationally expensive, we create only 9 sets of mocks with their input BAO peak position determined by 9 different  $\alpha$  parameters regularly distributed between 0.96 and 1.04. For each set of mocks we built a new optimal estimator. At each iteration the current  $\alpha$  is not exactly one of the 9 values computed above. Therefore, we perform an interpolation between results of two estimators corresponding to the two nearest values of  $\alpha$  in the grid. Details of how the interpolation is performed and how the covariance matrix of this measurement is estimated are presented in section 4.1 of [Vargas-Magaña et al. \(2013\)](#).

The iterations stop when the difference between the two last fitted  $\alpha$  parameters is below 0.0001, which is achieved in general after less than ten iterations.

#### 4.3.5 Performance on PTHalos mock catalogs

We tested the performance of the iterative optimal estimator on a more realistic set of mock catalogs also used on the BAO measurements with BOSS: the PTHalos mocks described in § 4.2. Additional tests were performed in LasDamas suite of N-body simulations for SDSS-II and are described in [Vargas-Magaña et al. \(2013\)](#).

Since the cosmological model of PTHalos is slightly different than the one used for our mocks, we expect to measure a BAO peak placed such that  $\alpha = 1.002$  in our fiducial cosmology.

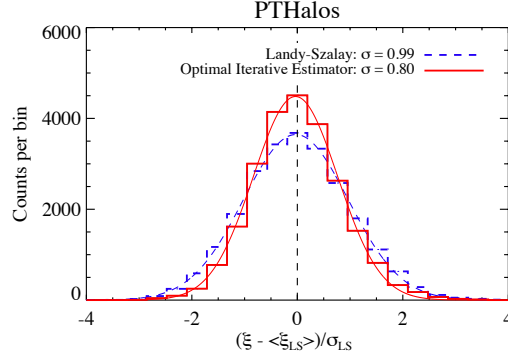


Figure 4.7: “Pull” distribution of correlation functions measured with PTHalos mock catalogs in the range  $40 < s < 200 h^{-1}\text{Mpc}$  with the Landy-Szalay (dashed-blue) and the iterative optimal estimator (solid red). The standard deviation of the Gaussian fit shows a smaller scatter for the latter estimator.

We used the iterative version of the optimal estimator to measure the correlation functions of all PTHalos realizations. Again, the covariance matrix was estimated through the scatter of different realizations around the mean correlation function. To make comparisons, we also computed the correlation functions using Landy-Szalay estimator.

Figure 4.7 shows the “pull” histogram of the correlation functions, i.e., the residuals of the correlation function relative to the average Landy-Szalay correlation function, normalized to the empirical RMS of the Landy-Szalay estimator,  $(\hat{\xi} - \langle \xi_{\text{LS}} \rangle) / \sigma_{\text{LS}}$ . By construction, the width of the pull distribution for the Landy-Szalay estimator is close to one while it is 0.8 for our estimator. This corresponds to a gain of 20% in errors, similar to what was obtained in the lognormal mock catalogs previously. We also observe that there is no residual bias in the estimation that would move the mean away from zero.

Fitting for BAO in each realization gives the distribution of  $\alpha$  parameters shown in Fig. 4.8. We observe that the expected value of  $\alpha$  is recovered by both estimators in average. However, the optimal estimator values show a reduced scatter around the mean of about 20% compared to the Landy-Szalay estimator, similar improvement as observed in the correlation function errors shown in Fig. 4.7.

We conclude that our method is robust and unbiased, allowing the application to real data.

#### 4.3.6 Application to real data: DR9 CMASS

We used the iterative optimal estimator to estimate the spherically averaged correlation function  $\xi(s)$  of the DR9 CMASS sample of galaxies. Results are shown in Fig. 4.9. The covariance matrix is obtained from the scatter around the mean of measurements on PTHalos mock catalogs, as performed in Anderson et al. (2012). The Landy-Szalay was also computed for comparison.

We fitted the correlation function as described in § 4.3.4. It resulted in a  $\chi^2 = 27.7$  for 35 degrees of freedom for the Landy-Szalay estimator while  $\chi^2 = 29.5$  for the iterative optimal estimator.

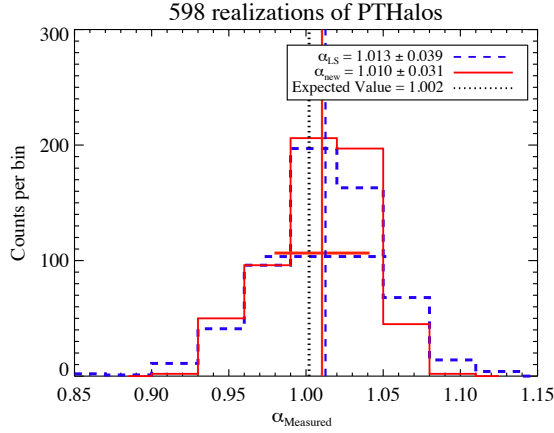


Figure 4.8: Histogram of  $\alpha_{\text{Measured}}$  for the 610 realizations of PTHalos mock galaxy catalogs using the Landy-Szalay (dashed blue) and the iterative optimal estimators (solid red). The average values over the realizations, shown in the legend, are represented as vertical lines with the same line styles, while the RMS of the histograms are shown as horizontal thick lines. The expected theoretical value is shown as a vertical black dotted line.

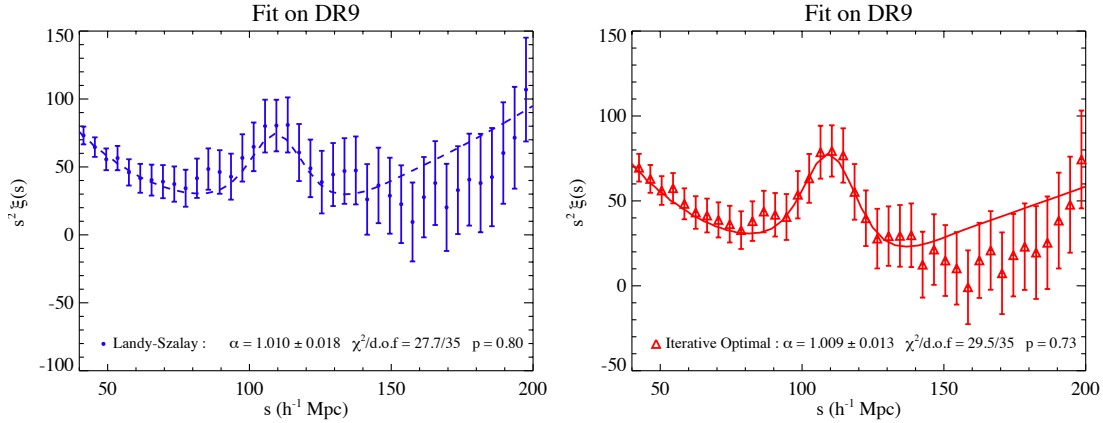


Figure 4.9: Correlation functions obtained for DR9 CMASS sample using the Landy-Szalay (left) and the iterative optimal estimator (right panel) with their best-fit models. The recovered  $\alpha_{\text{Measured}}$  are given in the legend, together with the  $\chi^2/\text{d.o.f.}$  and the probability of observing this value of  $\chi^2$ . Errors come from the diagonal terms of the covariance matrix and do not show the correlations among neighboring separation bins.

The  $\alpha$  values obtained are

$$\alpha_{\text{LS}} = 1.010 \pm 0.018, \quad \text{for the Landy - Szalay estimator and} \quad (4.19)$$

$$\alpha_{\text{opt}} = 1.009 \pm 0.013, \quad \text{for the Iterative Optimal Estimator.} \quad (4.20)$$

We see a 28% improvement in the accuracy of the fit of the BAO peak position when using our estimator. In Fig. 4.10 is shown that the  $\alpha$  parameter obtained for DR9 CMASS and its error are in agreement with measurements on PTHalos mock catalogs.

The values of  $\alpha$  obtained using our two estimators are in agreement between them and with the official measurement presented in Anderson et al. (2012) on the same DR9 CMASS sample using Landy-Szalay correlation function (without reconstruction):  $\alpha = 1.016 \pm 0.017$ . With reconstruction, the official measurement gains 6% in accuracy, while with our iterative optimal

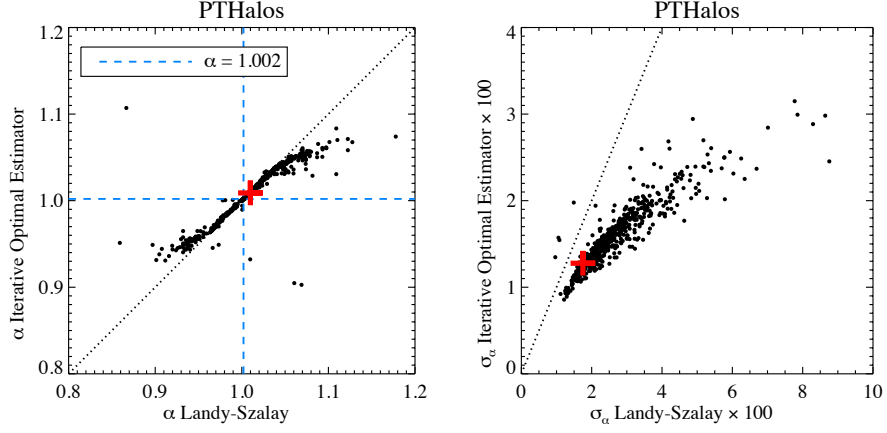


Figure 4.10: Comparison of the best-fit values of the isotropic  $\alpha$  dilation parameter (left) and their errors (right panel), obtained from correlation functions estimated using Landy-Szalay and the iterative optimal estimators on PTHalos mocks (black dots) and for the DR9 CMASS sample (red cross). Dashed blue lines show the expected value of  $\alpha$  for PTHalos mocks and black dotted lines show the identity relation in both panels.

estimation this improvement is even larger. A good test would be to use the optimal estimator on reconstructed catalogs and to quantify the total gain in accuracy.

There is plenty of room for improvement of the optimal estimator, allowing a better understanding of its performances. First, the estimator could be extended to treat anisotropic correlation functions. Second, a deeper analysis of the behavior of the coefficients  $c_i$  per iteration could tell us more about the relative importance of each ratio  $\mathbf{R}_i$ , and how they improve the accuracy of our estimates. Third, a more accurate model for the correlation function, used both in the mock generation and in the final fits for BAO. Fourth, the optimal estimator needs to be applied to larger samples, for instance DR12, and see whether the improvements are still significant.



## Chapter 5

# Mock catalogs of Quasar Lyman- $\alpha$ forests surveys

In this chapter we discuss the production of mock catalogs of BOSS Ly $\alpha$  forests including cosmological large scale correlations. These mocks were extensively used to test our BAO analysis (described in Chapters 6 and 7).

The first step, the generation of the correlated absorption field over quasar lines of sight is discussed in § 5.1. Two methods are presented: the official one (Font-Ribera et al., 2011) used on BOSS mocks, and a new method, less computationally demanding. The second step consists in creating realistic BOSS spectra, adding instrumental and astrophysical effects to absorption fields. All steps are detailed in § 5.2.

The original contribution of this Chapter is the development and testing of the new method to generate absorption fields (§ 5.1.3 and § 5.1.4). Also, I developed the *MockExpander* package that automatically transforms absorption fields into realistic BOSS spectra, allowing users to turn on/off different instrumental or astrophysical features. This package was extensively used for tests of the BAO analysis in Chapter 7; it will be publicly available at the Data Release 12 (December 2014).

### 5.1 Generating the absorption field

In this section we discuss how to generate the absorption field of Ly $\alpha$  forests of BOSS quasars. I introduce the method used for the official DR11 mock catalogs, then I describe an alternative for the first method that is significantly computationally cheaper and more flexible to changes in the mock properties.

For both methods, we use the lognormal approximation for the optical depth which allows for a sufficiently precise reproduction of the flux PDF of the Ly $\alpha$  forest, while also taking into account its evolution in redshift. The underlying Gaussian random field is correlated according to a power



spectrum. This implicitly means that the three-point function and other higher order odd-point functions vanish for this field. This is an approximation since non-linear growth effects couple different Fourier modes creating non-vanishing odd-point functions, but for the purposes of our mocks this approximation is sufficiently good.

The usual method to generate a correlated Gaussian random field following a given power spectrum is to draw random Fourier amplitudes  $\delta(\vec{k})$  for  $\vec{k}$ -modes in defined over a regular grid. These amplitudes are drawn from a centered Normal distribution with variance equal to the power spectrum. Using a Fast Fourier Transform (FFT) over this  $k$ -grid, we convert Fourier amplitudes into a configuration space density field  $\delta(\vec{x})$ , also defined over a regular grid with periodic boundary conditions. This method was used when creating galaxy mock catalogs for the study of the optimal estimator in § 4.3. Thanks to the available FFT algorithms, the FFT method is very fast.

For a Ly $\alpha$  forest survey such as BOSS, the large volume ( $\sim 20 h^{-3}\text{Gpc}^3$ ) and the resolution needed to simulate forests observed with BOSS spectrographs (resolving better than  $1 h^{-1}\text{Mpc}$  at  $z = 2.3$ ) makes the FFT method impractical due to hardware limitations<sup>1</sup>. Also, this method would compute the density field in a regular grid, over regions not probed by the quasar lines of sight, representing a needless computing effort.

Therefore, two methods were developed to circumvent those limitations and are described next. Both methods are adapted to compute a correlated density field *only* over quasar lines of sight.

### 5.1.1 The lognormal model for the forests

The goal is to obtain forests for which the transmissions  $F$  are correlated. The statistics of the transmittance fluctuations  $\delta_F = F/\bar{F} - 1$  should follow the input flux power spectrum given by the model described in § 2.5,

$$P_F(\vec{k}, z) = b^2(z)g^2(z)(1 + \beta\mu^2)^2 P_{\text{lin}}(k, z = z_0)D(\vec{k}), \quad (5.1)$$

where the input cosmological model used to compute  $P_{\text{lin}}(k)$  is the same fiducial model employed in the Ly $\alpha$  BAO analysis described in Chapter 8 :

$$\Omega_m = 0.27, \quad \Omega_b h^2 = 0.0227, \quad h = 0.7, \quad n_s = 0.96, \quad \sigma_8 = 0.8. \quad (5.2)$$

The linear bias of the Ly $\alpha$  forest is  $b(z) = 0.14[(1+z)/(1+z_0)]^{2.9}$ , where  $z_0 = 2.25$ , the growth rate is  $g(z) = 1/(1+z)$  in the matter-dominated era (to a good approximation) and the redshift-space distortion parameter is  $\beta = 1.4$ . The exponential damping term  $D(\vec{k})$  is a parametric function fitted to match hydro simulations (McDonald, 2003) basically describing modeling the small scale clustering of the Ly $\alpha$  forest.

The final correlated variables  $F$  are given by the exponential of the optical depth  $\tau$ , which is itself the exponential of a correlated Gaussian random field  $\delta_G$  in our lognormal approximation.

<sup>1</sup>The FFT algorithms store the full grid of amplitudes at a time in RAM memory, which might be impractical if the number of points in the grid is too large.

We can rewrite this transformation in a general form

$$F = \exp \left[ -a(z)e^{b(z)\delta_G} \right], \quad (5.3)$$

that ensures, for any positive values of  $a$  and  $b$ , that  $0 < F < 1$ . The functions  $a(z)$  and  $b(z)$  are computed to match the mean transmitted flux fraction  $\bar{F}(z)$  and its variance,  $\sigma_F^2(z)$ , given by the integral of the power spectrum of Eq. 5.1, both functions of redshift. Our models for  $\bar{F}(z)$  and  $\sigma_F^2(z)$  match observations of the 1D Ly $\alpha$  power spectrum (McDonald et al., 2006). They are given by

$$\ln \bar{F}(z) = \ln(0.8) \left[ \frac{1+z}{1+z_0} \right]^{3.2} \quad (5.4)$$

and

$$\sigma_F^2(z) = \sigma_F^2(z = z_0) \left[ \frac{1+z}{1+z_0} \right]^\gamma, \quad (5.5)$$

where  $\gamma = 3.8$ . At  $z_0 = 2.25$ , using the fiducial model (Eq 5.2) we obtain  $\sigma_F^2(z = z_0) = 0.110$ , which yields  $a(z_0) = 0.018$  and  $b(z_0) = 3.91$ .

In the next two sections, we describe how to generate a correlated Gaussian random field  $\delta_G$ . However, the transmission  $F$  is not Gaussian, it is a non-linear function of a Gaussian variable. Since both are random variables with probability densities satisfying  $p_F(F) dF = p_G(\delta_G) d\delta_G$  (the relation between  $\delta_G$  and  $F$  is bijective, where  $p_G$  is the normal distribution), it is possible to write a relationship between the transmission correlation function  $\xi_F = \langle F_1 F_2 \rangle$  and the correlation function of the Gaussian random variables  $\xi_G = \langle \delta_{G1} \delta_{G2} \rangle$ , given by

$$\xi_F(\xi_G) = \langle F(\delta_{G1}) F(\delta_{G2}) \rangle = \int_{-\infty}^{\infty} d\delta_{G1} \int_{-\infty}^{\infty} d\delta_{G2} \frac{\exp \left[ -\frac{\delta_{G1}^2 + \delta_{G2}^2 - 2\delta_{G1}\delta_{G2}\xi_G}{2(1-\xi_G^2)} \right]}{2\pi\sqrt{1-\xi_G^2}} F(\delta_{G1}) F(\delta_{G2}). \quad (5.6)$$

The former,  $\xi_F$ , is just Fourier transform of the transmission power spectrum (Eq. 5.1). The latter,  $\xi_G$ , is computed as a function of  $\xi_F$  after inverting Eq. 5.6. This function depends only on the relation between  $F$  and  $\delta_G$  (Eq. 5.3), which translates into a dependence on  $a(z)$  and  $b(z)$ . It can be tabulated for many values of  $\xi_G$  and inverted, so we are able to convert values of  $\xi_F$  into  $\xi_G$  (note that Eq. 5.6 does not depend on separation explicitly). The resulting  $\xi_G$  or its Fourier transform, the Gaussian power spectrum  $P_G$ , are used in the next sections to create correlated Gaussian variables. Since  $a(z)$  and  $b(z)$  are functions of redshift, the relations  $\xi_F \leftrightarrow \xi_G$  and  $P_F \leftrightarrow P_G$  also depend on redshift.

Next are described two different methods that use the Gaussian correlation function  $\xi_G(\vec{r})$  and the Gaussian power spectrum  $P_G(\vec{k})$  to generate a spatially correlated Gaussian random field,  $\delta_G(\vec{x})$ .

### 5.1.2 Cholesky decomposition of correlation matrix

This method uses matrix operations to rotate a vector of uncorrelated variables  $\eta_G$  into a correlated one  $\delta_G$ , such that

$$\langle \delta_{Gi} \delta_{Gj} \rangle = \xi_G(\vec{r}_i - \vec{r}_j). \quad (5.7)$$

Let  $\eta_G$  be an uncorrelated Gaussian random variable with zero mean and unity variance. First, for each Ly $\alpha$  forest pixel of the survey we associate a random value  $\eta_{Gi}$ , where the index  $i$  refers to that pixel. Second, we compute, for the same pixels, linear combinations of all  $\eta_{Gi}$  in order to make correlated Gaussian variables,  $\delta_{Gi}$ , written as

$$\delta_{Gi} = \sum_{j=1}^{N_p N_q} A_{ij} \eta_{Gj}, \quad \text{or in vectorial notation,} \quad \vec{\delta}_G = \mathbf{A} \vec{\eta}_G. \quad (5.8)$$

$N_p$  is the number of pixels per line of sight (we start by assuming it is the same for all quasars) and  $N_q$  is the number of quasars in the survey. The total number of pixels of the survey is  $N_p \times N_q$ . The coefficients  $A_{ij}$  of the matrix  $\mathbf{A}$  are to be determined. Reminding that we want to satisfy Eq. 5.7, this corresponds to

$$\langle \vec{\delta}_G^T \vec{\delta}_G \rangle = \mathbf{C} = \mathbf{A}^T \mathbf{A}, \quad (5.9)$$

where the elements of the matrix  $\mathbf{C}$  are given by  $C_{ij} = \xi_G(\vec{r}_i - \vec{r}_j)$ .

Basically that means that the wanted matrix  $\mathbf{A}$  is the ‘‘square-root’’ of  $\mathbf{C}$ . The coefficients of  $\mathbf{A}$  can be obtained by a Cholesky decomposition of  $\mathbf{C}$ . The Cholesky decomposition transforms  $\mathbf{C}$  into a product of a triangular matrix and its transpose.

The Cholesky decomposition is a simple procedure but limited by the fact that the matrix  $\mathbf{A}$  contains  $(N_p N_q)^2$  elements. For a BOSS-like survey,  $N_p \sim 500$  and  $N_q \sim 10^4$ , meaning that the Cholesky decomposition needs to store in memory a  $10^{12}$  element matrix simultaneously, not possible with the current available technology.

To overcome this issue, a physical argument can be used to divide the problem into smaller ones. First, we need to assume that lines of sight are parallel. Let  $\delta_G(r_{\parallel}, \vec{r}_{\perp})$  be the correlated Gaussian variable at position  $r_{\parallel}$  of the line of sight at coordinate  $\vec{r}_{\perp}$  ( $r^2 = r_{\parallel}^2 + r_{\perp}^2$ ). If we perform a one-dimensional Fourier transform of  $\delta_G$  in the direction of the line of sight only, we obtain  $\tilde{\delta}_G(k_{\parallel}, \vec{r}_{\perp})$ . Those one-dimensional Fourier modes have the following correlation:

$$\langle \tilde{\delta}_G(k_{\parallel}, \vec{r}_{\perp}) \tilde{\delta}_G(k'_{\parallel}, \vec{r}'_{\perp}) \rangle = \frac{1}{2\pi} \int d\vec{k}_{\perp} e^{i\vec{k}_{\perp} \cdot \vec{r}_{\perp}} \int d\vec{k}'_{\perp} e^{i\vec{k}'_{\perp} \cdot \vec{r}'_{\perp}} \delta^D(k_{\parallel} + k'_{\parallel}) \delta^D(\vec{k}_{\perp} + \vec{k}'_{\perp}) P_G(\vec{k}) \quad (5.10)$$

$$= 2\pi \delta^D(k_{\parallel} + k'_{\parallel}) P_{\times}(k_{\parallel}, |\vec{x}_{\perp} - \vec{x}'_{\perp}|), \quad (5.11)$$

where the symbol  $\delta^D$  stands for the Dirac delta distribution,  $P_G(\vec{k})$  is the Gaussian power spectrum of  $\delta_G$ , and

$$P_{\times}(k_{\parallel}, r_{\perp}) = \frac{1}{2\pi} \int_{k_{\parallel}}^{\infty} k_{\perp} dk_{\perp} \frac{\sin(k_{\perp} r_{\perp})}{k_{\perp} r_{\perp}} P_G(k_{\parallel}, k_{\perp}). \quad (5.12)$$

This shows that modes  $\tilde{\delta}_G$  on different lines of sight are independent except when  $k_{\parallel} = k'_{\parallel}$ . Therefore, the problem is now separated for each value of  $k_{\parallel}$ , reducing the original correlation matrix into  $N_p$  matrices of  $N_q^2$  elements each.

We proceed as follows : 1) we build a  $(k_{\parallel}, r_{\perp})$  grid, 2) for each value of  $k_{\parallel}$ , we compute the correlations between the  $\tilde{\delta}_G$  using Eq. 5.12 to build the  $\mathbf{C}$  matrix of  $N_q^2$  elements, 3) we perform

a Cholesky decomposition of each of the  $\mathbf{C}$  matrices to obtain the triangular matrices  $\mathbf{A}$ , 4) we generate a set of independent Gaussian variables,  $g_{ij}$  for each quasar and each value of  $k_{\parallel}$ , then rotating them through the  $\mathbf{A}$  matrices to obtain  $\tilde{\delta}_G = \mathbf{A}g$ , 5) finally, we perform a one-dimensional inverse Fourier Transform to convert the  $\tilde{\delta}_G$  back to the real-space field  $\delta_G$ , that are correlated as implied by the Gaussian power spectrum  $P_G(\vec{k})$ .

The one-dimensional Fourier transform in the direction of the line of sight simplifies the procedure but implies that all pixels of the a given line of sight follow the same power spectrum. Consequently, the redshift evolution of  $P_G(\vec{k})$  cannot be directly implemented. This problem is solved by creating the same survey many times with a different power spectrum but same seeds for the random number generator. Each survey is computed at a different redshift and the final survey is obtained by interpolation.

Even after decomposing the problem, matrices involved still have a large number of elements ( $N_q^2 \sim 10^8$  for BOSS). Therefore, we further simplify the problem by dividing the full survey into independent sub-regions, each with a smaller number of quasars. More details on this on § 5.2, where we present specific characteristics of BOSS mock catalogs.

This method, even though it is memory demanding, was extensively tested in [Font-Ribera et al. \(2011\)](#) and it was used to generate the absorption field for the official BOSS Ly $\alpha$  forest mock catalogs, described in § 5.2.

### 5.1.3 Random sampling of modes : theory

We describe now an alternative method to the previous one. The idea of this method is to Fourier transform a set of modes  $\delta(\vec{k})$  in order to obtain a correlated Gaussian random field  $\delta_G(\vec{r})$  but, instead of using the usual FFT in a regular grid of modes, we compute explicitly the integral using  $\vec{k}$ -modes randomly distributed over the Fourier space, as a Monte Carlo integration. As we demonstrate next, this method does not require large amounts of memory, neither the division of the survey in sub-regions. Computing time scales linearly with the number of lines of sight instead of quadratically. There is also no need to make the assumption of parallel lines of sight and the redshift evolution implementation is straightforward. In addition, thanks to the random sampling of the modes, the usual limits in frequency of FFT methods practically do not exist anymore.

In general, a correlated random variable  $\delta_G(\vec{r})$  can be written as a Fourier transform of  $\delta(\vec{k})$  as

$$\delta_G(\vec{r}) = S^{-1} \sum_j^{N_{\text{modes}}} \delta(\vec{k}_j) \exp(i\vec{k}_j \cdot \vec{r}) + \text{c.c.} , \quad (5.13)$$

where  $S$  is a normalization factor defined later and “c.c.” is the notation for the complex conjugate of term in the sum, needed to make  $\delta_G$  purely real. The mode amplitudes  $\delta(\vec{k}_j)$  are also complex quantities. This sum is the discrete version of the integral form of the Fourier Transform, which is a true equality in the limit of infinite number of modes, covering all possible real values of wavevectors  $-\infty < k < \infty$ . Of course in practice for numerical computations, only a finite number of wavevectors are used.

In a FFT, the wavevectors  $\vec{k}_j$  are sampled in a regular three-dimensional grid, that is related to the regular grid of positions  $\vec{r}_j$  where the field is sampled through  $\vec{k}_j = (L\pi/r_j)\hat{r}_j$ . There are as many modes as positions in the grid. This is the reason why the FFT method is not adapted to generate mock Ly $\alpha$  forest: in a 3D regular grid, most of the points would lay outside the quasar lines of sight, representing a waste of computing effort.

In our new method, the wavevectors  $\vec{k}_j$  are not sampled in a grid, but randomly. The sum in Eq. 5.13 is explicitly computed for all the real-space locations  $\vec{r}$  of interest. The values of the wavevectors  $\vec{k}_j$  are randomly drawn from a probability distribution function (PDF)  $p(\vec{k}_j)$  defined to be proportional to the Gaussian power spectrum  $P_G(\vec{k})$  as

$$p(\vec{k}) d^3k = \frac{1}{(2\pi)^3} \frac{P_G(\vec{k})}{\xi_G(0)} d^3k, \quad (5.14)$$

where  $\xi_G(0) = (2\pi)^{-3} \int d^3k P_G(\vec{k})$ , assumed to be finite. Since our Gaussian random field has variance unity, we set  $\xi_G(0) = 1$  hereafter.

Theoretically, there are no limits for the range covered by  $k_j = |\vec{k}_j|$ . However, the shape of the power spectrum itself sets a natural range. The  $k_j$  values where the power spectrum value is negligible are drawn with negligible probability. This automatically sets the  $k$  range, interesting for clustering measurements. In practice, a limited range is needed in order to numerically define the PDF from Eq. 5.14. We use the range defined by  $k_j \in [0, k_{\text{cut}}]$ , where the value of  $k_{\text{cut}}$  is chosen to be after the fall-off region of the power spectrum damping tail. This is another advantage of this method, that can take into account a much larger range of  $k_j$  compared to the usual FFT method.

The amplitudes  $\delta(\vec{k}_j)$  of each mode are chosen to be independent complex Gaussian random variables of zero mean,  $\langle \delta(\vec{k}_j) \rangle = \langle \delta^*(\vec{k}_j) \rangle = 0$ , and unitary variance,  $\langle \delta(\vec{k}_j) \delta^*(\vec{k}_j) \rangle = 1$ . At this point, one can imagine the result: drawing  $n$  modes with amplitudes  $\delta(\vec{k}_j)$  in a Fourier volume element  $d^3k$  centered at a given  $\vec{k}$  gives that the sum of these variables have a total variance equal to  $n$ . The fraction  $n/N_{\text{modes}}$  of modes inside this volume element is proportional to probability distribution defined by the power spectrum  $P(\vec{k})$ . The effective result is that those  $n$  modes have variance equal to  $P(\vec{k})$ , when choosing the appropriate value for the normalization  $S$  in Eq. 5.13. This gives a normalization factor  $S = \sqrt{2N_{\text{modes}}}$  (the “2” accounts for the complex conjugate term).

In the following we show that taking expected values over the distributions we recover the correct correlation function  $\xi_G$  for the  $\delta_G$ . Expressing the expected value of the product of two correlated variables  $\delta_G(\vec{r}_1)$  and  $\delta_G(\vec{r}_2)$  we have

$$\begin{aligned} \langle \delta_G(\vec{r}_1) \delta_G(\vec{r}_2) \rangle &= \frac{1}{2N_{\text{modes}}} \sum_{j,l}^{N_{\text{modes}}} \left\langle \delta(\vec{k}_j) \delta(\vec{k}_l) \exp\left(+i\vec{k}_j \cdot \vec{r}_1 + i\vec{k}_l \cdot \vec{r}_2\right) + \text{c.c.} \right. \\ &\quad \left. + \delta(\vec{k}_j) \delta^*(\vec{k}_l) \exp\left(+i\vec{k}_j \cdot \vec{r}_1 - i\vec{k}_l \cdot \vec{r}_2\right) + \text{c.c.} \right\rangle. \end{aligned} \quad (5.15)$$

This expression simplifies significantly since there is no correlation between the amplitudes  $\delta(\vec{k}_j)$  and the wavevectors  $\vec{k}_j$ , drawn from two different PDFs. Therefore, all crossed terms in the expression above vanish. Also, Fourier amplitudes of different wavevectors are independent so only diagonal terms with  $j = l$  do not vanish when taking ensemble averages. Therefore, we have

$$\left\langle \delta_G(\vec{r}_1) \delta_G(\vec{r}_2) \right\rangle = \frac{1}{2N_{\text{modes}}} \sum_j^{N_{\text{modes}}} \left\langle \exp \left[ i\vec{k}_j \cdot (\vec{r}_1 - \vec{r}_2) \right] + \text{c.c.} \right\rangle. \quad (5.16)$$

The expected value in the sum can be written as an integral the wavevector PDF,  $p(\vec{k}_j)$ . Since the PDF is proportional to the power spectrum, this integral yields the Gaussian correlation function  $\xi_G(\vec{r})$ . In other words,

$$\begin{aligned} \left\langle \exp \left[ i\vec{k} \cdot (\vec{r}_1 - \vec{r}_2) \right] \right\rangle &= \int d^3k \exp \left[ i\vec{k} \cdot (\vec{r}_1 - \vec{r}_2) \right] p(\vec{k}) \\ &= \frac{1}{(2\pi)^3} \int d^3k \exp \left[ i\vec{k} \cdot (\vec{r}_1 - \vec{r}_2) \right] P_G(\vec{k}), \end{aligned} \quad (5.17)$$

$$\left\langle \exp \left[ i\vec{k} \cdot (\vec{r}_1 - \vec{r}_2) \right] \right\rangle = \xi_G(\vec{r}_1 - \vec{r}_2). \quad (5.18)$$

Finally, Eq. 5.16 becomes,

$$\left\langle \delta_G(\vec{r}_1) \delta_G(\vec{r}_2) \right\rangle = \frac{1}{2N_{\text{modes}}} \sum_j^{N_{\text{modes}}} 2\xi_G(\vec{r}_1 - \vec{r}_2) = \xi_G(\vec{r}_1 - \vec{r}_2). \quad (5.19)$$

We have demonstrated that the random sampling of modes produces a correlated Gaussian random field following an input power spectrum in the limit of infinite number of modes  $N_{\text{modes}}$ . Having a finite number of modes introduces an intrinsic error on the equality of Eq. 5.18. This error states how far from  $P_G(\vec{k})$  can be the actual variance of the modes drawn in a given realization. The question is how many modes are needed such that this error is smaller compared to the other source of errors in the survey, e.g., the sample variance or noise. In § 5.1.4 we try to answer this question.

### Including redshift-space distortions

Include linear redshift-space distortions (RSD) in this method is straightforward because, in Fourier space, their effect is to amplify the isotropic modes  $\delta_{\text{iso}}(k)$  in radial direction, which can be described by the RSD operator (Kaiser, 1987)

$$\delta(\vec{k}) = (1 + \beta\mu_k^2)\delta_{\text{iso}}(k), \quad (5.20)$$

where  $\mu_k = \hat{k} \cdot \vec{r}$  is the cosine of the angle between the wavevector  $\vec{k}$  and the direction  $\hat{r}$  of the line of sight, and  $\beta$  is the RSD parameter of the Ly $\alpha$  forest.

In our method, a given fluctuation  $\delta_G(\vec{r})$  is computed explicitly with Eq. 5.13. Since the position  $\vec{r}$  is known, it is possible to compute  $\mu_k$  and apply the RSD operator for each individual mode while computing the sum. Therefore, the input power spectrum used in the wavevector PDF (Eq. 5.14) is the isotropic one  $P_G(k)$  and RSD are applied when computing  $\delta_G$ . An interesting feature of this procedure is that there is no need of the assumption of parallel lines of sight. The RSD operator is local and this operation works with the spherical geometry of a BOSS-like survey.

The value used for  $\beta$  in Eq. 5.20 is the same as in the flux power spectrum in Eq. 5.1. The redshift-space distortions are correctly conserved even after applying the non-linear transformation between  $\delta_G$  and  $F$ .

### Redshift evolution of the clustering

We model the redshift evolution of the Ly $\alpha$  forest clustering through an scaling of the power spectrum amplitude proportional to  $(1+z)^\gamma$  (Eq. 5.5).

As for the RSD, the redshift evolution could be applied by scaling the Fourier modes  $\delta(\vec{k})$  in the sum of Eq. 5.13 by  $[(1+z)/(1+z_0)]^\gamma$ , but this procedure does not work. The reason for this is the non-linear relation between  $\delta_G$  and  $F$  (Eq. 5.3), for which we assumed  $\langle \delta_G^2 \rangle = 1$  for all redshifts. Scaling the amplitudes of  $\delta_G$  would change their variance and the transformation to  $F$  would not work anymore.

If no scaling on  $\delta_G$  is applied, their variances would still be unity and we could expect that the functions  $a(z)$  and  $b(z)$  would automatically scale the flux power spectrum when converting  $\delta_G$  to  $F$ . They actually do a correct scaling of  $\sigma_F$  but the scaling of the correlation function  $\xi_F$  is incorrect.

The problem is that, when computing the Gaussian power spectrum  $P_G(k)$  from the flux power spectrum  $P_F(k)$  using Eq. 5.6, a constant scaling factor in the latter does not simply translate into a constant scaling factor in former; the scaling is actually a function of  $k$ .

Therefore, we compute the Gaussian power spectrum at different redshifts  $P_G(k, z)$ . When performing the sum in Eq. 5.13 for a pixel at redshift  $z$ , each mode  $\delta(k)$  is scaled by the function

$$R(k, z) = \sqrt{\frac{P_G(k, z)}{P_G(k, z = 2.25)}}, \quad (5.21)$$

which also conserves the unity of the variance of  $\delta_G$ . This procedure artificially modifies the power spectrum amplitude in a  $k$  and  $z$  dependent way.

#### 5.1.4 Random sampling of modes : tests

In this section, we test the random sampling method by comparing with the Cholesky decomposition method. The question we want to answer is how many modes  $N_{\text{modes}}$  are required for

constructing an absorption field for a BOSS-like survey having the correctly modeled correlation function.

### The idea

If the number of modes  $N_{\text{modes}}$  is not sufficiently high, the underlying power spectrum  $P_G^{N_{\text{modes}}}(k)$  can be significantly different from the desired power spectrum  $P_G(k)$ . This can be visualized in the limit case where just one  $k^*$  mode is drawn: the underlying power spectrum is just a Dirac delta distribution centered in  $k^*$ . A new realization would also have a Dirac power spectrum but centered in a different  $k$  value. Many independent realizations of  $\delta_G$  fields produced with a finite number of modes will have their underlying power spectrum fluctuating around the desired one. If the underlying power spectrum  $P_G^{N_{\text{modes}}}(k)$  of each realization could be measured perfectly, those measurements would scatter around  $P_G(k)$  due to the finite number of modes. This scatter should be smaller as we increase the number of modes used in each realization. We refer to this scatter as the “finite-mode variance”, hereafter.

The problem is that, for a given survey, a perfect measurement of the underlying power spectrum is not possible. For a finite survey, there is always an intrinsic error on the estimate of the power spectrum due to the finite number of pixels sampling a given scale, called sample variance. This is equivalent for the correlation function. Therefore, the scatter of measurements over many mock realizations have both contributions combined, sample and finite-mode variance.

The idea of our test is thus to compute the spread of correlation functions estimated over a set of realizations of the same survey, and see how this spread depends on the number of modes. We consider that  $N_{\text{modes}}$  is sufficiently large when this spread is dominated by the sample variance (which does not depend on  $N_{\text{modes}}$ ).

### Our test samples

We produced realizations of mock Ly $\alpha$  forest surveys using the random sampling method in order to analyze the effect of the number of drawn modes. Our forests contain only the absorption field, with no noise or continuum model added. The quasar positions were chosen to be the same as the DR11 quasars. For simplicity, we chose a sub-sample of the DR11 NGC quasars, shown in Fig. 5.1. Our larger sample contains  $\sim 60,000$  quasars.

Forests were computed in pixels of  $0.5 h^{-1}\text{Mpc}$  comoving over the wavelength range between  $\lambda_{\text{Ly}\beta}$  and  $\lambda_{\text{Ly}\alpha}$  in the rest-frame of each quasar. This comoving size for the bins is the same used for the official DR11 mock catalogs created with the Cholesky decomposition method. It is smaller than the typical BOSS spectrograph bin (of about  $1 \text{ \AA}$  that corresponds to  $0.7 h^{-1}\text{Mpc}$  at  $z = 2.3$ ).

Sets of 20 realizations of mock surveys were produced for five values of  $N_{\text{modes}}$ :  $3 \times 10^4$ ,  $10^5$ ,  $3 \times 10^5$ ,  $10^6$  and  $3 \times 10^6$  modes.



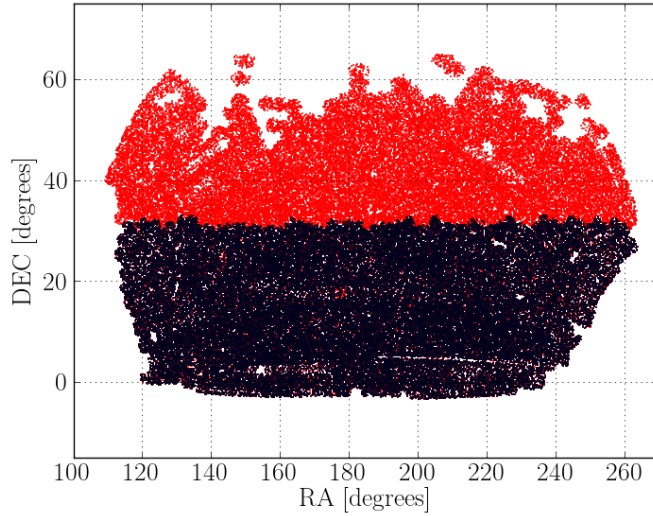


Figure 5.1: Angular distribution of the quasars used (black dots) to build mock catalogs for testing the random sampling method of generating the correlated Gaussian random field. These quasars are a sub sample of DR11 quasars (red dots) in the North Galactic Cap.

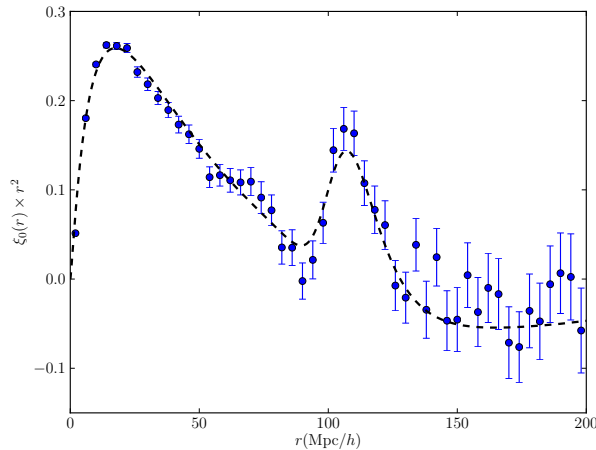


Figure 5.2: Monopole of the Ly $\alpha$  forest correlation function of using  $3 \times 10^6$  modes estimated from skewers without noise and exact continua. The dashed line is showing the input theory.

### Analysis of convergence

We used the standard analysis chain described in § 6 to estimate the correlation function  $\hat{\xi}(r_{\parallel}, r_{\perp})$  and its covariance matrix for each realization. We used  $50 \times 50$  bins of width  $4 h^{-1}$  Mpc covering separations between 0 and  $200 h^{-1}$  Mpc parallel and transverse to the line of sight.

Figure 5.2 shows an example of the estimated monopole of the correlation function (defined as the average of  $\xi(r, \mu)$  over  $0 < \mu < 1$  for fixed  $r$ ) for our largest sample of forests built using  $N_{\text{modes}} = 3 \times 10^6$ . We observe a very good agreement between the input theory and our measurement, showing that the random sampling of modes works correctly and reproduces the BAO peak.

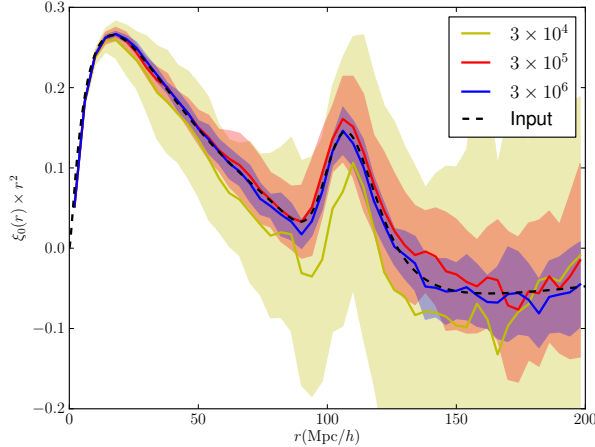


Figure 5.3: Average and RMS of monopoles of the Ly $\alpha$  forest correlation function from 20 realizations of mock catalogs using different number of modes. The dashed line is showing the input theory.

Figure 5.3 shows the mean and standard deviation of measured monopoles over each set of twenty realizations, for three values of  $N_{\text{modes}}$ :  $3 \times 10^4$ ,  $3 \times 10^5$  and  $3 \times 10^6$ . We clearly see that the scatter reduces when increasing the number of modes as expected.

For future work, it is to be quantified how many modes a given survey needs for convergence of the variance. Non-diagonal terms of the covariance matrix should also in principle converge thanks to the Wick theorem, that decomposes the 4-point function into sum of products of 2-point functions. If the 2-point function converges, the Wick theorem would imply the convergence of the 4-point function.

## 5.2 Producing BOSS mock data-sets

In the previous section we described the process of generating the correlated transmitted flux fraction  $F$  along quasar lines of sight. To generate realistic mocks of the data, these spectra need to be transformed to include astrophysical features such as a quasar continuum, high column density absorbers, metal absorption lines, and instrumental features such as noise, spectrophotometric calibration, sky subtraction residuals, etc. In this section we explain how we included all these features in DR11 mock catalogs.

These mock data are intended to mimic BOSS *coadded* spectra, that result from the coaddition of several successive 20-minute individual exposures of the same object, in a given plate and a single night<sup>2</sup>. Most of our discussion below refers to coadded spectra except for the section on noise properties where a distinction between the noise in each exposure and the noise in the coadds is necessary.

<sup>2</sup>The number of exposures is determined by the target overall signal-to-noise of the given plate.

The expansion procedure is summarized by the following *expansion* equation:

$$f_\lambda = [(F_\lambda \cdot C_\lambda) * \tilde{W}_\lambda(R_p, R_w) + N_\lambda] \cdot M_\lambda + f_\lambda^{\text{sky}}. \quad (5.22)$$

where  $f_\lambda$  is the simulated flux and the parameters on the r.h.s are as follows:  $F_\lambda$  is the raw transmission field, set to 1 outside the Ly $\alpha$  forest;  $C_\lambda$  is the PCA-generated quasar continuum;  $\tilde{W}(R_p, R_w)$  is the Fourier transform of the BOSS resolution and pixelization kernel (eq. 5.23) which is convolved with the product  $F_\lambda C_\lambda$ ;  $N_\lambda$  is the noise computed from our model (eq. 5.25);  $M_\lambda$  is a linear function of  $\log \lambda$  (eq. 5.24) used to ensure that each mock spectrum has the same mean flux and spectral index as the corresponding real spectrum; and  $f_{\text{sky}}$  is the added sky subtraction residuals (Fig. 5.6). The following subsections describe in detail each of the components of this expansion equation.

### 5.2.1 Raw transmission field, $F_\lambda$

The procedure for generating the correlated transmission field is described in section 5.1, and yields a field of mean absorption and variance given by equations 5.4 and 5.5. That transmission field models regions of optically thin neutral hydrogen absorption. However, some of the regions probed by the Ly $\alpha$  forest will inevitably contain higher density systems and non-Ly $\alpha$  absorbers. We describe the addition of these systems in the next two sub-sections.

#### 5.2.1.1 High column density (HCD) systems

Most of the Ly $\alpha$  absorption in a spectrum arises from regions that are highly ionized and optically thin to ionizing photons. However, most of the neutral hydrogen is in dense systems with high neutral hydrogen column density, which produce wavelength intervals of complete absorption surrounded by damped wings. These structures affect the measured Ly $\alpha$  transmission correlations in two ways. First, they affect the size of the Ly $\alpha$  forest fluctuations directly impacting the variance in the resulting long-range 3D correlations. Second, since these systems are themselves biased differently than the optically thin regions, they will also affect these correlations themselves.

Damped Ly $\alpha$  systems (DLAs) have strong damped wings that allow for their easy identification, but Lyman limit systems (LLS) of lower column density can also affect the correlations even if their damped wings are weak and individually not detectable.

We therefore insert systems of neutral hydrogen column density  $N_{\text{HI}} > 10^{17.2} \text{ cm}^{-2}$ , which we collectively designate as high column density systems (HCDs), following the procedure that is described in (Font-Ribera and Miralda-Escudé, 2012). In brief, HCDs are distributed only in pixels where the transmission  $F$  is lower than a certain threshold  $F_0$ , defined such that the probability to have an optical depth  $\tau$  larger than  $\tau_0 = -\ln(F_0)$  is 1%. The column density of the HCDs are randomly drawn from an analytical model (Zheng and Miralda-Escudé, 2002) calibrated to match observations (Noterdaeme et al., 2009) from SDSSII-DR7. Voigt profiles are included in these regions assuming a constant Doppler parameter  $b_D = 70 \text{ km s}^{-1}$ .

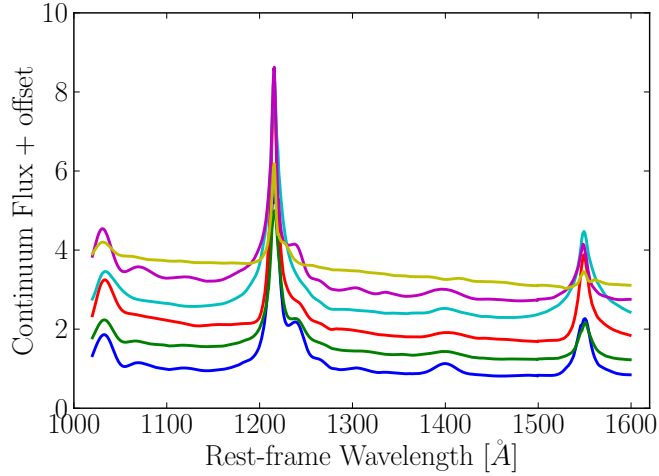


Figure 5.4: Example continua from 5 random mocks showing variations in line shapes and the continuum shape within the Ly $\alpha$  forest.

The effect of HCDs on the Ly $\alpha$  forest correlation function and the measurement errors was studied previously in (Font-Ribera and Miralda-Escudé, 2012).

### 5.2.1.2 Metals

In addition to absorption from hydrogen, metals present in the intergalactic medium can also absorb quasar light at discrete wavelengths inside the quasar forest. We added metal absorption with a procedure that assumes that all significant metal absorption is associated with significant Ly $\alpha$  absorption. For each mock spectra, we considered pixels for which the Ly $\alpha$  transmittance  $F(\lambda)$  was below 0.4. For these cases, we decrease the flux by  $\delta F_{\text{Met}}$  of their corresponding flux bin at wavelength  $\lambda + \Delta\lambda_{\text{Met}}$  for a given metal line.

Application of this method requires a list of significant metal lines and appropriate values for the absorption,  $\delta F_{\text{Met}}$ . This was done following the procedure of Slosar et al. (2011) where quasar spectra were stacked after shifting each spectra so that it is centered on a wavelength of high absorption. The stacked spectra then have, necessarily, a strong absorption at the center, but also absorption lines at wavelengths separated from the center by  $\lambda_i - \lambda_0$  where  $\lambda_0$  is the wavelength of absorption by the species responsible for the strong absorption and  $\lambda_i$  is the wavelength of absorption of any other absorber. With these stacks, we identified 21 correlated absorption features, 9 of them corresponding to Ly $\alpha$ -metal correlations and the others to metal-metal correlations (arising from the strong absorber being due to a metal line rather than to Ly $\alpha$ ). We designate these metal-metal correlations as “shadows” hereafter (Pieri et al., 2010).

Using the line catalog and the deduced flux decrements, we added metal lines (including shadow lines) to the mock spectra. No scatter is added to these flux decrements.

### 5.2.2 Quasar continua $C_\lambda$

The quasar continuum  $C$  for each line-of-sight is constructed using eight higher ranked eigenspectra of the Principal Component Analysis (PCA) of high resolution quasar spectra (Suzuki et al., 2005). The final continuum is a sum of a mean shape and a linear combination of these eigenspectra for which amplitudes were randomly sampled following a centered Gaussian distribution with the corresponding standard deviation. These eigenspectra cover restframe wavelengths 102–160 nm spanning from the Lyman- $\beta$  peak through C IV  $\lambda$ 1549. Wavelengths above and below this range are discarded in the mock spectra. Figure 5.4 shows 5 example mock continua with variations in line shapes and the shape of the continuum within the Ly $\alpha$  forest. The random sampling of the PCA eigenvalue amplitudes occasionally leads to negative continua at some wavelength bins (this happens on 0.5% of the continua), and in this case the continuum is discarded and a new set of random amplitudes is drawn.

### 5.2.3 The BOSS kernel

BOSS spectrographs cover the wavelength range 361 nm - 1014 nm with a resolving power  $\lambda/\Delta\lambda$  varying from 1300 in the blue end to 2600 in the red end. Each data spectrum has its own estimate of the wavelength dispersion per pixel. In BOSS coadded data, the pixels are logarithmic in wavelength with steps of  $\Delta \log_{10}(\lambda) = 10^{-4}$  corresponding to 69 km s $^{-1}$ .

As described in § 5.1.2, the raw absorption fields were produced over the lines-of-sight using a grid in comoving space with bins of 0.5  $h^{-1}$ Mpc. To match BOSS spectra binning and resolution, we first compute the mean wavelength dispersion (PSF)  $R_w$  and mean pixel width  $R_p$  over the Ly $\alpha$  forest region of the corresponding data spectrum. Each raw field was then convolved using the following kernel (in Fourier space):

$$W(k, R_p, R_w) = \exp\left(-\frac{k^2 R_w^2}{2}\right) \left[\frac{\sin(kR_p/2)}{kR_p/2}\right]. \quad (5.23)$$

We then match the binning by taking, for each data pixel, the absorption value of the closest pixel of the smoothed raw field.

### 5.2.4 Flux normalization $M_\lambda$

We wish to ensure that each mock quasar has a mean forest flux and spectral index equal to those of the corresponding real quasar. Specifically, we normalize the noise-free mock quasar flux  $f_{\text{mock}}$  to the data flux  $f_{\text{data}}$  by solving for  $M_0$  and  $M_1$  in

$$f_{\text{data}} = f_{\text{mock}}(M_0 + M_1 \log_{10} \lambda), \quad (5.24)$$

over the rest-frame wavelength ranges  $104.1 < \lambda < 118.5$  nm (inside the Ly $\alpha$  forest) and  $127.0 < \lambda < 150.0$  nm (between the Ly $\alpha$  and CIV emission peaks). Then,  $f_{\text{mock}}$  is multiplied by the factor  $M_\lambda = M_0 + M_1 \log_{10} \lambda$ . These fits are done using the inverse variance given by the pipeline

as fit weights and ignoring all masked pixels. For DR9 mock data-sets, the fit was performed without weighting leading a slightly larger number of bad fits, nearly 1% of the full sample.

The parameter  $M_1$  effectively corrects for quasar spectral distortions introduced by the SDSS optics that are currently not corrected by the pipeline. The Sloan 2.5-m telescope has a chromatic focal plane and lacks an atmospheric dispersion corrector. As a result, the optimal position for a spectrograph fiber is a wavelength and airmass dependent quantity. Galaxy targets and calibration stars are optimized for 540 nm, while quasar targets are offset both along and across the focal plane to optimize the signal-to-noise at 400 nm for Ly $\alpha$  forest studies. This offset means that the flux calibration vectors derived from the standard stars are not correct for the quasars and result in a flux mis-calibration which depends upon wavelength, airmass, seeing, guiding, and the location on the focal plane (Dawson et al., 2013).

### 5.2.5 The noise $N_\lambda$

The noise,  $N_\lambda$ , added to the fluxes of a given mock quasar is a random number taken from a Gaussian distribution of mean zero and with a variance determined by the noise model for the corresponding real quasar. The noise models are most naturally expressed using the total number of photo-electrons  $p_{\text{tot}}$  (signal plus sky) since in an ideal system the variance would be equal to  $p_{\text{tot}}$ . In practice the model gives the variance  $\sigma_{\text{phot}}^2$  as a linear function of  $p_{\text{tot}}$ :

$$\sigma_{\text{phot}}^2 = N_0 + N_1 p_{\text{tot}} \quad (5.25)$$

The coefficient  $N_0$  reflects the CCD readout noise and other systematic effects that are independent of the photon flux. The coefficient  $N_1$  would be unity for pure Poisson photon noise in the absence of systematics, but in practice  $N_1 \geq 1$  owing to sky subtraction and flux calibration errors.

For each mock quasar, the parameters  $N_0$  and  $N_1$  are found by fitting  $p_{\text{tot}}$  as a function of  $\sigma_{\text{phot}}^2$ , both obtained from its corresponding real quasar. This requires the use of the calibration vector  $c(\lambda)$  to transform fluxes to photo-electrons:

$$p_{\text{tot}}(\lambda) = [f_{\text{QSO}}(\lambda) + f_{\text{sky}}(\lambda)] / c(\lambda) \quad (5.26)$$

$$\sigma_{\text{phot}}^2(\lambda) = \sigma^2(\lambda) / c(\lambda)^2 \quad (5.27)$$

where  $\sigma^2(\lambda)$  is the estimated flux variance of the data. We fit a linear noise model for each spectrum, using pixels from the blue side of the spectrograph, Fig. 5.5 shows an example of a fit for  $N_0$  and  $N_1$  for one quasar spectrum in DR11.

Of course the accuracy of the noise model is only as good as the accuracy of  $\sigma(\lambda)$ . We do not use the pipeline-calculated  $\sigma(\lambda)$  because there are systematic deviations between it and the true  $\sigma(\lambda)$ . This was demonstrated by studying smooth regions redwards of the Ly $\alpha$  peak, between 142.0 and 151.0 nm where there are no strong emission lines. For each QSO spectrum we fit a 3rd order polynomial in this region to obtain an estimate of the continuum. The polynomial fit includes 5- $\sigma$  outlier rejection to be robust against metal absorption and data artifacts. We then

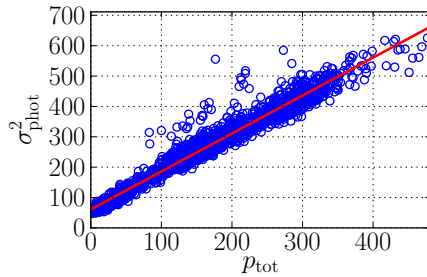


Figure 5.5: Example of a noise model fitting for spectrum 5493-56009-0504 (plate-MJD-fiber). The red line shows the result of the linear fit of the photon variance (Eq. 5.25) in the data points for a given spectrum.

organize the data in wavelength bins and measure, in each bin, the spread of the pull distribution by fitting a Gaussian to this distribution in the range  $[-2, 2]$ . Deviations from unity of the sigma parameter of this Gaussian directly measure any pipeline mis-estimation of the noise variance.

We find that for spectra in individual exposures, the pipeline noise estimates are accurate at the 1–2% level. However, coadded spectra have an observer-frame wavelength dependent error in the estimated noise. For  $\lambda < 600$  nm where only the blue spectrograph contributes, it is approximately proportional to the square root of the ratio of the coadd to individual spectrum spectral bin sizes. Below  $\sim 475.0$  nm, where the coadded bins are smaller than the original exposures, the pipeline underestimates the noise by 0–10%, while above  $\sim 470$  nm, where the coadded bins are larger than the original exposures, it overestimates by noise by 0–15%. Similar estimates were obtained in Palanque-Delabrouille et al. (2013). The “Concordance Ly $\alpha$  forest Sample” of Lee et al. (2013) uses a slightly different parametrization which has the same functional form in wavelength, but also includes a flux-dependent correction and a per-quasar overall normalization.

This parametrization of the pipeline noise correction is used when generating the noise for mock spectra and for purposefully mis-reporting the noise in those spectra to simulate the mis-estimated noise of the real data.

In addition to  $\sigma(\lambda)$ , we also need the calibration vector  $c(\lambda)$  to convert the mock and coadded data spectra into photo-electrons. The BOSS data individual exposures include the calibration vector  $c_i(\lambda)$  that converts between observed photo-electrons  $p_i(\lambda)$  and calibrated flux  $f_i(\lambda)$ . The coadd is performed as a weighted simultaneous spline fit to the individual exposures and the resulting effective calibration vector is not calculated. We re-derive the effective calibration vector using the approximation that the coadded photo-electrons are the unweighted sum of the individual exposure photo-electrons:  $p = \sum_i p_i(\lambda) = r(\lambda) \sum_i f_i(\lambda)/c_i(\lambda)$ , where  $r(\lambda)$  corrects for the wavelength dependent difference in bin-size between individual exposures and the coadded spectra.

Since the coadded flux is normalized to the same units as the individual exposures (ergs/s/cm<sup>2</sup>/Å), we may factor out  $f(\lambda) \simeq f_i(\lambda)$  such that  $p(\lambda) = f(\lambda)r(\lambda) \sum_i c_i^{-1}(\lambda)$ . Thus the effective calibration vector to convert between coadded photo-electrons and coadded flux is:

$$c(\lambda) = \left( r(\lambda) \sum_i c_i^{-1}(\lambda) \right)^{-1} \quad (5.28)$$

For this calculation we use only the blue exposures, therefore limiting the wavelength range of mock spectra from 360 to 633 nm.

With the parameters  $N_0$  and  $N_1$  for each quasar, it is simple to add realistic noise to the mock spectra. We compute the mock quasar flux from the product of the transmittance  $F$  and the generated PCA quasar continuum  $C$ , and add the same sky flux  $f_{\text{sky}}$  that is used in the data. We then compute the mock photons  $p$  and noise  $\sigma_p$  at each pixel,

$$p = (F \cdot C + f_{\text{sky}})/c \quad (5.29)$$

$$\sigma_p^2 = N_0 + N_1 p \quad (5.30)$$

We add noise using a Gaussian distribution with mean 0 and sigma  $\sigma_p$ :

$$\tilde{p} = p + \mathcal{N}(0, \sigma_p) \quad (5.31)$$

We convert back into quasar flux  $\tilde{f}_{\text{QSO}}$  :

$$\tilde{f}_{\text{QSO}} = \tilde{p}c - f_{\text{sky}} \quad (5.32)$$

$$\sigma_{\tilde{f}} = c\sigma_p \quad (5.33)$$

The final products are the noiseless mock spectrum, a noisy realization of that spectrum, the true inverse variance of that noise, and a noisy measurement of that inverse variance to mimic the fact that the real data measurement errors are themselves a noisy estimate of the true errors .

The noisy measurement of the inverse variance is generated as follows. As explained previously, the noise mis-calibration is defined by the ratio of wavelength bin sizes of individual exposures and coadded spectrum  $r(\lambda)$ . We also add Gaussian random fluctuations to this noise estimates with standard deviation proportional to the photon variance itself. The final mock photon noise estimate  $\tilde{\sigma}_{\tilde{f}}$  is given by  $\sigma_{\tilde{f}}(\lambda)r(\lambda) + \mathcal{N}(0, \sqrt{2}\sigma_f)$ .

The model presented here assumes that the noise in different pixels of the same spectrum is uncorrelated. This is likely not realistic because covariance among neighboring pixels is introduced by rebinning. However, other sources of small scale correlations in these mock catalogs are not correctly modeled by the input power spectrum, since the log-normal model was built to fit large-scale correlations. Therefore we do not consider any noise correlation between neighboring bins.

## 5.2.6 Sky mis-subtraction $f_{\lambda}^{\text{sky}}$

Figure 5.6 (left) shows the median residual of BOSS sky spectra after the sky model has been subtracted. While this is only a 1–2% bias in the sky subtraction, it can be large compared to the Ly $\alpha$  forest flux, which is typically faint compared to the sky. Fig. 5.6 (right) shows the relative flux between noise-free simulated Ly $\alpha$  forest flux and the residual sky. The median is



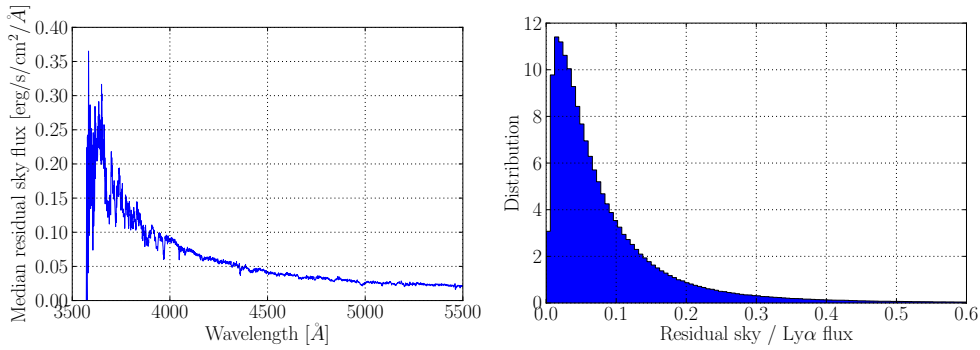


Figure 5.6: Median residual from BOSS sky-subtracted sky fibers (left panel), and the distribution of the ratio of that residual to simulated Ly $\alpha$  forest flux (right panel).

5.6%, the mean is 9.9%, and there is a tail reaching up to 1.0. Note that this is an additive component to mock spectra, unlike other mis-calibrations that are multiplicative.

Mock spectra add this median sky subtraction residual times a random constant scatter with mean 1.0 and RMS 0.1 so that each spectrum receives a slightly different sky subtraction residual bias.

### 5.2.7 The *MockExpander* package

These mocks are essential in the study of systematic effects, either physical or instrumental, that could affect our measurements. They are extensively used in this work in § 7 to test our BAO analysis chain.

The *MockExpander* package was developed to make the use of mocks easier. This package is a compilation of codes that transform raw absorption fields into realistic spectra, with all features discussed in this section. The user has the choice of turning on and off some systematics to test their effect on their analysis. Therefore, the user only needs to download the absorption fields and real data. This package is also easy to customize, allowing the user to modify our expansion recipe.

In December 2014, the *MockExpander* and the full set of mock catalogs will be publicly available for the community in the Data Release 12.

## Chapter 6

# Measuring the correlation function of the Lyman- $\alpha$ forest

In this chapter we describe all the steps of the measurement of the 3D correlation function of the Ly $\alpha$  forests, and the procedure to estimate of the BAO peak position.

Continuum fitting methods are discussed in § 6.1. Then, we compute the optimal weights for the correlation function estimate in § 6.2. The covariance matrix, estimated through two different methods described in § 6.3, is used on the measurement of the BAO peak position (§ 6.4).

Being just a descriptive chapter about the methods, all tests using mock catalogs showing the limitations of our analysis are explained in the next chapter, and the application on real data in the subsequent chapter.

This chapter is part of the original contribution of this thesis, also part of the following publications: [Busca et al. \(2013\)](#), [Delubac et al. \(2014\)](#). The BAO fitting procedure is published in [Kirkby et al. \(2013\)](#).

### 6.1 Continuum fitting

The first step on the Ly $\alpha$  BAO analysis is to compute the transmitted flux fluctuations  $\delta_F$  (defined in Eq. 2.15) of Ly $\alpha$  forests. In this section, we present three different automatic procedures to estimate flux fluctuations from BOSS quasar spectra.

#### 6.1.1 Generalities

The transmitted flux fraction  $F$  in a given pixel centered at wavelength  $\lambda$  is simply the ratio of the observed flux  $f(\lambda)$  and the unabsorbed flux coming from the quasar  $C(\lambda)$ , also referred as

continuum. Rewriting Eq. 2.15 in terms of  $f$  we have,

$$\delta_F(\lambda) = \frac{f(\lambda)}{C(\lambda)\bar{F}(\lambda)} - 1. \quad (6.1)$$

To compute  $\delta_F$ , we can choose between estimating either the product  $C\bar{F}$ , or  $C$  and  $\bar{F}$  independently. The continuum  $C$  is different for each quasar while the function  $\bar{F}$  is assumed to be the same for all quasars.

In high resolution quasar spectra the continuum level  $C$  is estimated manually since it is more visible (as in Fig. 2.1). Samples of high resolution spectra usually contain at most hundreds of quasars making the manual procedure feasible. However, BOSS quasar spectra have lower resolution and lower signal-to-noise, making individual continuum level estimates more complex. Furthermore, the BOSS sample contains hundreds of thousand spectra, making automatic continuum fitting algorithms preferred.

First, a model for the continuum is needed. Quasar continua can be basically described by a power-law and some emission lines which can have different relative intensities depending on the composition of the quasar environment. The continuum region of our interest, over the forest, is essentially smooth. Therefore, a basic model of the continuum shape over the forest can be build by simply stacking forests in their rest-frame. This is used by the continuum fitting methods C1 and C2 described in § 6.1.2 and § 6.1.3, respectively. The C2 method was chosen as the standard method for all the analyses presented in the subsequent chapters. C2 was the most tested method on mock catalogs, and uses a slightly more refined algorithm than C1.

Another continuum model can be built from a set of high resolution quasar spectra and a principal component analysis (PCA) (Suzuki et al., 2005, Suzuki, 2006). A hand-fitted continuum is estimated for each high-resolution quasar, then the mean continuum and its covariance matrix are computed. The eigenvectors of the covariance matrix are extracted, ranked by its eigenvalues in decreasing order. The principal components are defined by a subset of eigenvectors with larger eigenvalues. Usually, about ten principal components are considered. A given quasar spectrum can have its continuum modeled by a mean shape plus a linear combination of these components. This is the case of method C3 described in § 6.1.4. Two samples of PCA on quasar spectra are available (Suzuki et al., 2005, Pâris et al., 2011). The first sample was used to estimate BOSS quasar continua in Lee et al. (2013), and to build mock catalogs for the BAO analysis (see § 5). This method was not chosen as the standard one because of the lack of systematic testing on mock catalogs (that are constructed using PCA templates).

### 6.1.2 The Gaussian method: C1

This method assumes that the transmitted flux fraction  $F$  is a Gaussian random field. A fixed shape is fitted over each forest through a  $\chi^2$  minimization. This method is not adapted to estimate the continuum level of the quasar,  $C(\lambda)$ , and the mean transmission  $\bar{F}(z)$  separately, but to directly estimate the product of both,  $C(\lambda)\bar{F}(z)$ .

We describe now the steps of this method. First, all spectra are normalized by their respective observed mean flux in the interval 127.5 and 128.5 nm in the quasar rest-frame. Second, normalized forests are stacked in order to obtain a forest shape that will be used afterwards for fitting individual forests. Actually, ten different stacks are produced corresponding to ten independent quasar redshift ranges, between 2.1 and 3.5. Each stack is an estimate of the average product of the normalized continuum times the absorption,  $\langle C(\lambda)F(z) \rangle$ . The third and last step is to fit the stacked shape over the forest of each quasar  $q$  (using the stack associated to the redshift of the quasar). In order to take in to account quasar intrinsic diversity and also wavelength dependent spectro-photometric errors, a power-law is multiplied to the stacked shape,

$$C_q(\lambda)\bar{F}(z) = a_q\lambda^{b_q} \langle C(\lambda)\bar{F}(z) \rangle, \quad (6.2)$$

where the two parameters,  $a_q$  and  $b_q$ , are such that they minimize the weighted squared residuals of the normalized flux for each quasar.

### 6.1.3 The Non-Gaussian method: C2

This method might be considered as an improvement of the Gaussian method since it actually estimates the continuum level in the forest. The main improvement comes from the modeling of the probability density function (PDF) of the transmission  $F$ .

The used flux PDF model  $P(F, z)$  is the same used to generate mock catalogs. It assumes a redshift dependent lognormal distribution for the optical depth,  $\tau$  in Eq. 2.1 (more details in Chap. 5). In practice, this PDF is computed from the mock raw transmission fields. It is important to match the rebinning of pixels in mock and data. Rebinning significantly alters the flux PDF shape, and might introduce biases in the continuum estimation. In our analysis, we rebin forests into “analysis pixels” composed of three BOSS pixels.

The probability density  $P_i$  of observing flux  $f_i$  in each pixel  $i$  assuming a continuum  $C(\lambda_i)$  is the convolution of  $P(F, z_i)$  with the noise of that pixel  $\sigma_i$ , taken to be Gaussian, summarizing:

$$P_i(f_i, C_q(\lambda_i), z_i) \propto \int_0^1 dF P(F, z_i) \exp \left[ \frac{-(C_q F - f_i)^2}{2\sigma_i^2} \right]. \quad (6.3)$$

Similar to method C1, in this method the continuum for each quasar is a product of a stacked mean shape  $\bar{C}(\lambda_{\text{rf}})$  and a linear function in  $\log \lambda$ ,

$$C(\lambda) = (a_q + b_q \log \lambda) \bar{C}(\lambda_{\text{rf}}). \quad (6.4)$$

In this method, the mean shape  $\bar{C}$  is the same for all quasar redshifts. The parameters  $a_q$  and  $b_q$  are determined for each quasar by maximizing a likelihood defined by the product of probabilities for each pixel,

$$L(C_q) = \prod_i P_i[f_i, C_q(\lambda_i)]. \quad (6.5)$$

The last step for computing  $\delta_F$  is to estimate the mean transmission  $\bar{F}(\lambda)$ . The value of  $\bar{F}(z)$  could be derived analytically, if our flux PDF model  $P(F, z)$  was the true distribution of flux

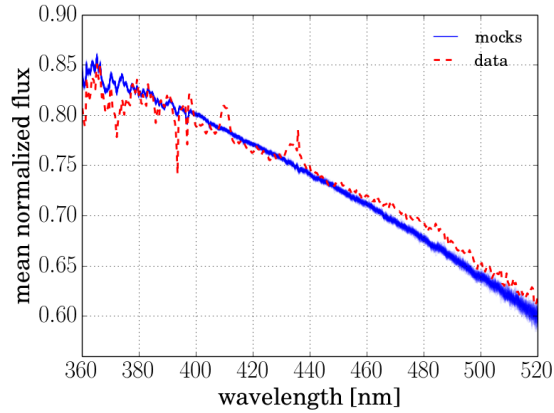


Figure 6.1: The measured mean transmitted flux fraction  $\bar{F}$  as a function of observed wavelength  $\lambda$  for data (red dashed) and for a set of ten mock catalogs (blue solid).

fractions. Therefore, we estimate  $\bar{F}$  from fitted forests by requiring resulting fluctuations  $\delta_F$  to have zero mean.

### The iterative procedure

In order to be less dependent on the mock flux PDF used in this method, we developed an iterative procedure that extract more information from the data itself when fitting continua.

First, we fit all continua using a constant mean shape  $\bar{C}(\lambda_{\text{rf}}) = 1$  (we assume that this shape is the same for all quasars at all redshifts). By stacking the resulting forests in the rest-frame of their quasars, we obtain the shape for the next iteration, where we fit again all continua. Furthermore, at each step we re-estimate the mean transmission  $\bar{F}(z)$  and the correction for pixel error mis-estimations, as discussed in § 6.2.1. Usually, up to six iterations are needed for convergence (more details in § 7.3).

#### 6.1.4 The PCA method: C3

This method uses a set of PCA samples described in § ?? to estimate quasar continua, not exclusively over the forest. We summarize here this method, but the full description and tests are published in Lee et al. (2013).

The basic idea of this method is to fit amplitudes of the principal components from Suzuki et al. (2005) between 102 and 160 nm in the quasar rest-frame. Inside the forest, measurements of the mean transmitted flux  $\bar{F}(\lambda)$  are used to correct the continuum level, a procedure called “mean flux regulation”.

Figure 6.2 gives an example of a randomly chosen spectrum with the continua estimated by the three methods presented here.

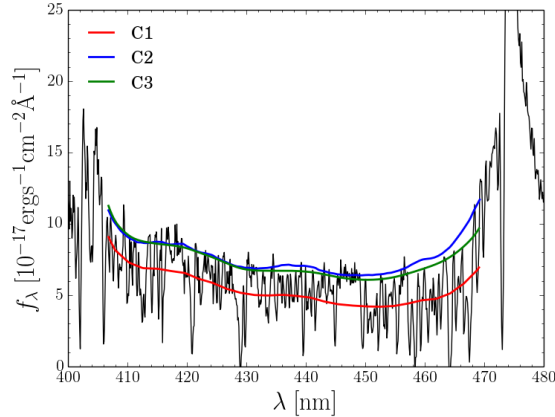


Figure 6.2: An example of a BOSS quasar spectrum of redshift 3.239. The red and blue lines cover the forest region used here,  $104.0 < \lambda_{\text{rf}} < 120.0$ . This region is sandwiched between the quasar’s  $\text{Ly}\beta$  and  $\text{Ly}\alpha$  emission lines respectively at 435 and 515 nm. The blue (green) line is the C2 (C3) model of the continuum,  $C_q(\lambda)$ , and the red line is the C1 model of the product of the continuum and the mean absorption,  $C_q(\lambda)\bar{F}(z)$ . (See text.)

## 6.2 The correlation function

Once the flux fluctuations  $\delta_F$  are estimated for all pixels in the survey, we compute the binned correlation function using the following estimator,

$$\hat{\xi}(A) = \frac{\sum_{i,j \in A} w_i w_j \delta_i \delta_j}{\sum_{i,j \in A} w_i w_j}, \quad (6.6)$$

where  $A$  is a bin of a separation grid. Separations can be computed using observable quantities such as redshifts and angles (as in Slosar et al. (2013)), or by converting them into comoving separation using a fiducial cosmology (as in Busca et al. (2013), Delubac et al. (2014)). This expression tells us that for a given bin  $A$ , we only average the product of fluctuations of pairs  $(i, j)$  of pixels whose separation falls inside  $A$ . The  $w_i$  are the weights of each pixel, defined next.

### 6.2.1 Weights

Before estimating the correlation function, it is important to correctly compute the weight of each pixel, taking into account the estimated error for each flux measurement and the redshift evolution of the  $\text{Ly}\alpha$  forest. This will affect not only the estimation of the correlation function  $\xi$  itself but also its covariance matrix (§ 6.3).

A discussion on the optimal weights for the  $\text{Ly}\alpha$  correlation function  $\xi$  is developed in McQuinn and White (2011). We use a simplified version, where the weight is the inverse of the total pixel variance, where the total pixel variance  $\sigma_i^2$  is a sum of pixel instrumental variance  $\sigma_{i,\text{pipeline}}^2$  (estimated by the automatic data-reduction pipeline) and intrinsic variance of the  $\text{Ly}\alpha$  forest  $\sigma_{\text{LSS}}^2(z)$  (redshift dependent), i.e.,

$$\sigma_i^2 = \frac{\sigma_{i,\text{pipeline}}^2}{\eta(z_i)} + \sigma_{\text{LSS}}^2(z_i). \quad (6.7)$$

The function  $\eta(z)$  is included to correct for an eventual systematic error in the estimation of the pixel variance by the pipeline. Indeed, many studies using BOSS spectra (McDonald et al., 2006, Palanque-Delabrouille et al., 2013, Lee et al., 2013) have shown that  $\eta(z) \neq 1$  in general. The instrumental pixel noise  $\sigma_{\text{pipeline}}$  is underestimated by 0-10% at  $\lambda < 450\text{nm}$  and overestimated for  $\lambda > 450\text{ nm}$  by 0-15%. The reason for these mis-estimates comes from the coaddition process of individual exposures by the automatic pipeline, that does not propagate errors precisely. Individual exposures also suffer from this problem, but with error mis-estimates smaller than 5%.

In practice, the functions  $\eta(z)$  and  $\sigma_{\text{LSS}}(z)$  are found by fitting Eq. 6.11 through the measurement of the true pixel variance  $\langle \delta_F^2 \rangle$  as a function of the furnished pipeline variance  $\sigma_{\text{pipeline}}^2$ . The relation  $\langle \delta_F^2 \rangle (\sigma_{\text{pipeline}}^2)$  is measured in ten pixel redshift bins. Tests presented in § 7.4 show that this model for the pixel variance is accurate to first order. The data measurement is shown in § 8.2.

The final weight is such that it maximizes the signal to noise ratio of the correlation function. The amplitude of correlations increases as  $(1+z)^\gamma$  with  $\gamma = 3.8$  (McDonald et al., 2006) but also does the intrinsic variance. In the approximation of independent bins of the correlation function, the variance of the estimator (Eq. 6.6) can be written as

$$\text{Var} [\hat{\xi}(A)] = \frac{\sum_{i,j \in A} w_i^2 w_j^2 \sigma_i^2 \sigma_j^2}{\left[ \sum_{i,j \in A} w_i w_j \right]^2}, \quad (6.8)$$

Since  $\xi_{ij}(z) = (1+z_i)^{\gamma/2} (1+z_j)^{\gamma/2} \xi_{ij}(z_0)$ , maximizing the signal to noise ratio given by

$$\left( \frac{S}{N} \right)^2 = \frac{\langle \hat{\xi}(A) \rangle^2}{\text{Var} [\hat{\xi}(A)]} \approx \frac{\left( \sum_{i,j \in A} w_i w_j \delta_i \delta_j \right)^2}{\sum_{i,j \in A} w_i^2 w_j^2 \sigma_i^2 \sigma_j^2}, \quad (6.9)$$

with respect to  $(w_i w_j)$  gives

$$w_i w_j \propto \frac{(1+z_i)^{\gamma/2} (1+z_j)^{\gamma/2}}{\sigma_i^2 \sigma_j^2}. \quad (6.10)$$

Therefore the final weight  $w_i$  for each pixel is given by

$$w_i = \frac{(1+z_i)^{\gamma/2}}{\left[ \frac{\sigma_{i,\text{pipeline}}^2}{\eta(z_i)} + \sigma_{\text{LSS}}^2(z_i) \right]}. \quad (6.11)$$

## 6.2.2 The measurement

For this work, we compute the correlation function (Eq. 6.6) as a function of comoving pixel separation in three dimensions. Therefore, using  $\Omega_m = 0.27$  in a flat  $\Lambda$ CDM cosmology we translate angles in the sky (RA and DEC) and  $z$  of pixels into comoving positions  $\vec{r}_i$  in units of  $h^{-1}\text{Mpc}$  using Eqs. 1.40 and 1.41.

The separation vector between two pixels  $\vec{r} = \vec{r}_i - \vec{r}_j$  can be decomposed in radial separations  $r_{\parallel}$  and separation transverse to the line of sight  $r_{\perp}$ , with  $r^2 = r_{\parallel}^2 + r_{\perp}^2$ . The binned correlation function  $\hat{\xi}_A$  is computed in a grid of bins  $4 h^{-1}\text{Mpc}$  wide, covering from 0 to  $200 h^{-1}\text{Mpc}$  in both radial and transverse separations, a total of 2500 bins. It is common to define the cosine of the angle between the separation vector  $\vec{r}$  and the line of sight direction  $\hat{z}$  as  $\mu = \hat{r} \cdot \hat{z} = r_{\parallel}/r$ .

When computing  $\hat{\xi}_A$ , we exclude pairs of pixels from the same quasar to avoid contamination coming from correlated differences between the true and the estimated continuum level. We also exclude pairs of pixels that have nearly the same observed wavelength, i.e.,  $r_{\parallel} < 4 h^{-1}\text{Mpc}$ , and that belong to the same plate. This is to avoid spurious correlations coming from the automatic data reduction, that uses constant quantities over a given plate (e.g. the sky subtraction algorithm).

### 6.2.3 Details about implementation

The computation of a correlation function is a straightforward procedure but it can be time demanding for a BOSS-like survey.

Our code was implemented to run using parallel threads. Each thread has a list of forests, and each forest has its own list of neighboring forests. The neighbors lists are obtained by first converting the  $200 h^{-1}\text{Mpc}$  transverse separation (maximal separation for correlation measurement) into an angle in the sky, which is  $3.5 \text{ deg}$  in our fiducial cosmology. Then, using the HEALPIX package we can find neighbors present in cells at less than  $3.5 \text{ deg}$  from the central quasar. For most quasars far from the survey boundaries, there are in average 400 neighboring forests. Those lists are not only useful for computing the correlation function, but also for the estimate of the covariance matrix by the Wick method (see § 6.3.2).

Once the forests are computed, they are exported into FITS format files with the wavelength, the comoving distance, the weight, and the continuum level of each pixel, and also the list of neighbors. Each forest has also the plate number, that is used to define the sub-regions for the sub-sampling method of computing the covariance matrix (§ 6.3.1).

During the correlation function computation, we associate a redshift for a pixel pair defined as the arithmetic mean of the pixel redshifts. This allows us to separate the correlation function in many disjoint redshift bins without losing pairs. We associate a pixel pair, instead of a single pixel, to a given bin.

## 6.3 The covariance matrix of $\hat{\xi}$

To measure the BAO peak position (§ 6.4), we fit a BAO model over the binned correlation function  $\hat{\xi}(A)$ . This fit requires an estimate of the covariance matrix of this measurement. Since our measurement represents a vector of 2500 separation bins, our covariance matrix is a  $2500 \times 2500$  element matrix. We estimate the covariance matrix  $C(A, B)$  using two methods, sub-sampling and a Wick expansion of the four-point function of the  $\delta_F$  field. Tests performed



in § 8.4 show that both methods yield results in agreement. We describe now each of these methods.

### 6.3.1 Sub-sampling

Our first method consists in dividing the survey in sub-samples of  $N_p$  independent regions and computing the scatter around the mean of estimated correlation functions over the sub-samples.

To define the regions, we used the focal-plane plates. If a pixel pair has its pixels in different plates, this pair accounts for correlations of the plate with lower right-ascension (this is of course an arbitrary choice).

We can rewrite Eq. 6.6 in terms of the sub-sample correlation function  $\hat{\xi}^p$  as

$$\hat{\xi}(A) = \frac{\sum_{p=1}^{N_p} w_A^p \hat{\xi}^p(A)}{\sum_{p=1}^{N_p} w_A^p}, \quad (6.12)$$

where  $w_A^p$  is the sum of weights in the sub-sample  $p$ . For the DR11 sample,  $N_p = 2044$ . We ensure that a given quasar accounts for a unique sub-sample.

The sub-sampling covariance matrix can be written as

$$C(A, B) = \langle \hat{\xi}(A) \hat{\xi}(B) \rangle - \langle \hat{\xi}(A) \rangle \langle \hat{\xi}(B) \rangle = S_{AB}^{-1} \sum_{p=1}^{N_{\text{plates}}} w_A^p w_B^p \left[ \langle \hat{\xi}^p(A) \hat{\xi}^p(B) \rangle - \langle \hat{\xi}(A) \rangle \langle \hat{\xi}(B) \rangle \right]. \quad (6.13)$$

In this expression we are neglecting correlations between  $\xi^p(A)$  and  $\xi^{p'}(A)$  for  $p \neq p'$ . In our case, this means that the correlation functions estimated in two different plates are uncorrelated. This approximation is only valid to first order because forests in different plates are all correlated among themselves. However, the contribution to the correlation function of forests near the plate boundary is small because of the larger number of pairs inside the same plate. The term  $S_{AB}$  is a normalization factor defined later (Eq. 6.16). We estimate the quantity inside the sum through

$$\hat{C}^p(A, B) = \hat{\xi}^p(A) \hat{\xi}^p(B) - \hat{\xi}(A) \hat{\xi}(B). \quad (6.14)$$

We remark that the  $N_p$  sub-samples give a very noisy measurement for the non-diagonal terms of the covariance matrix. Therefore, we developed a smoothing method in order to reduce this noise (see § 6.3.3).

Taylor et al. (2012) present a discussion about the precision of the covariance matrix estimates using sub-samples. In their work, it is argued that, in order to get 5% errors in the covariance matrix estimate, and if the number of data points of the measurement  $N_D$  (in our case is the number of correlation bins  $A$  effectively used in cosmological fits) is  $\gg 10^2$ , we need a number of sub-samples  $N_S > N_D$  and a fractional accuracy of  $< \sqrt{2/N_D}$  in the covariance. In our case, there are slightly more sub-samples than measurement bins ( $N_S \sim 2000$  and  $N_D \sim 1500$ ). The smoothing method presented in § 6.3.3 is able to make the covariance estimate achieve the required level of accuracy.

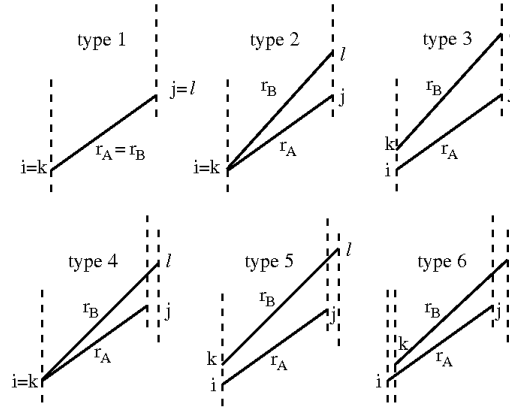


Figure 6.3: The six types of pairs of pairs used in the computation of the covariance matrix. The dashed lines refer to quasar lines-of-sight. The variances are dominated by types 1, 2 and 3. The  $(r_{\perp} - r'_{\perp} = 0)$  covariances are dominated by types 2 and 3.

### 6.3.2 The Wick approximation

The second method consists in using the approximation that  $\delta_F$  is a Gaussian field, writing the four-point function as a function of two-point functions (Wick theorem). The covariance between  $\xi$  for two separation bins  $A$  and  $B$  is therefore

$$C(A, B) = S_{AB}^{-1} \sum_{i,j \in A} \sum_{k,l \in B} w_i w_j w_k w_l [\xi_{ik} \xi_{jl} + \xi_{il} \xi_{jk}] . \quad (6.15)$$

where the pairs  $(i, j)$  and  $(k, l)$  refer to the pixels separated by the vector  $\vec{r}_A \in A$  and  $\vec{r}_B \in B$ . The normalization factor is basically the sum of the weights contributing for the sum in Eq. 6.15, written as

$$S_{AB} = \sum_{i,j \in A} (w_i w_j) \sum_{k,l \in B} (w_k w_l) . \quad (6.16)$$

In Figure 6.3 we illustrate the six possible configurations of pixel pairs naming them in decreasing order of importance to the final covariance.

Computing the full sum over pairs-of-pairs would require a large amount of computer time. Instead, we used a random sub-set of pairs-of-pairs. This also introduces noise in the covariance matrix as in the sub-sampling method, therefore needing a smoothing before the use in BAO measurements.

Remark that, for most diagrams, the correlation function between pixels in the same line of sight, i.e., the one-dimensional correlation function  $\xi_{1D}(\lambda, \Delta\lambda)$  is needed and therefore is also computed separately, in observable units (wavelength).

After computing separately the contribution of each diagram we observe that the diagonal elements of the covariance matrix are dominated ( $\sim 97\%$ ) by the two-quasar diagrams, T1 (contributing  $\sim 60\%$ ), T2 (25%) and T3 (15%) in Fig. 6.3. The dominant covariances have their larger contributions from T2 and T3. The structure of the covariance matrix and its dominant terms are analyzed in § 8.4.

### 6.3.3 Smoothing the covariance matrix

Given the low number of regions in the sub-sampling method or the use of only a fraction of pairs-of-pairs in the Wick method, the resulting covariance matrices  $C(A, B)$  are noisy. A smoothing procedure is needed to reduce this noise before we use these covariances in fits for the BAO peak position.

The smoothing procedure is inspired by its structure, analyzed in § 7.6. Basically we make the assumption that the correlation coefficient, defined as

$$c(A, B) = \frac{C(A, B)}{\sqrt{C(A, A)C(B, B)}}, \quad (6.17)$$

depends only of the difference of separations of bins  $A$  and  $B$ , i.e., on  $\Delta r_{\parallel} = r_{\parallel} - r_{\parallel}'$  and  $\Delta r_{\perp} = r_{\perp} - r_{\perp}'$ . Therefore, we average all correlation coefficients with same  $\Delta r_{\parallel}$  and  $\Delta r_{\perp}$ . Then, we assign this averaged value  $\bar{c}(\Delta r_{\parallel}, \Delta r_{\perp})$  to each element of  $c(A, B)$ .

We also avoid the noisy averages by setting to zero all elements for which the average correlation coefficient is compatible with zero. The error of a given average is defined by the standard deviation of the coefficients around this average.

The final smoothed covariance matrix is defined as

$$\tilde{C}(A, B) = \bar{c}(\Delta r_{\parallel}, \Delta r_{\perp}) \sqrt{C(A, A)C(B, B)}. \quad (6.18)$$

The assumptions around the smoothing of the covariance matrices are validated in § 7.6. In § 7.7, measurements of the BAO peak on mock catalogs show reasonable distribution of errors, which validate our estimates of the covariance matrix up to a few percent level.

## 6.4 Measuring the BAO peak position

After estimating the flux fluctuations and computed the correlation function and its covariance matrix, the BAO peak is clearly visible both in mock and in data (see § 7.5 and § 8.3). We proceed then to the measurement of the peak position, essential to extract cosmological information. In this section we expose the main points of the fitting procedure, also described in Kirkby et al. (2013)<sup>1</sup>.

### 6.4.1 The model

The idea is to fit the estimated binned correlation function  $\hat{\xi}(A)$  with a model composed of two functions. First, a template  $\xi_{\text{cosmo}}$  coming from a given cosmological model (see § 2.4). Second, a smooth function of separation without any peak-like feature, hereafter called “broadband”  $\xi_{bb}$ , used to marginalize out the full shape of the correlation function that might be not correctly

<sup>1</sup>The fitter, `BAOfit`, is publicly available on <http://github.com/baofit>

modeled by the template  $\xi_{\text{cosmo}}$  or to correct for any additive systematic correlations (function of separation), such as the distortion caused by the continuum fitting procedure (see § 7.1). Our model is written as,

$$\xi_{\text{model}}(\vec{r}, \alpha_{\parallel}, \alpha_{\perp}) = \xi_{\text{cosmo}}(\vec{r}, \alpha_{\parallel}, \alpha_{\perp}) + \xi_{bb}(\vec{r}), \quad (6.19)$$

where the  $\xi_{\text{cosmo}}$  is also decomposed into a smooth function and a peak function,

$$\xi_{\text{cosmo}}(\vec{r}, \alpha_{\parallel}, \alpha_{\perp}) = \xi_{\text{smooth}}(\vec{r}) + a_{\text{peak}} \cdot \xi_{\text{peak}}(\alpha_{\parallel} r_{\parallel}, \alpha_{\perp} r_{\perp}). \quad (6.20)$$

The parameter  $a_{\text{peak}}$  determines the amplitude of the BAO peak. The dilation parameters  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  “move the peak”, defining the position of the BAO peak relative to the position predicted by the fiducial cosmological model of the template  $\xi_{\text{cosmo}}$ , respectively in the radial and transverse direction. The dilation parameters contain information that will be translated into cosmological constraints in Chapter 10.

The angular size of the sound horizon at redshift  $\bar{z}$  is, to a good approximation, given by  $D_A(\bar{z})/r_d$  and its radial length in redshift space is  $D_H(\bar{z})/r_d$ , where  $D_A(\bar{z})$  is the angular diameter distance,  $D_H(\bar{z}) = c/H(\bar{z})$  is the Hubble distance and  $r_d$  is the comoving sound horizon at drag epoch. All those quantities were defined in § 1.5. The  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  parameters can be written as ratios of the angular size and radial size of the BAO peak position with respect to the fiducial cosmology values (denoted with subscript “fid”) as

$$\alpha_{\parallel} = \frac{D_H(\bar{z})/r_d}{[D_H(\bar{z})/r_d]_{\text{fid}}} \quad \text{and} \quad \alpha_{\perp} = \frac{D_A(\bar{z})/r_d}{[D_A(\bar{z})/r_d]_{\text{fid}}}. \quad (6.21)$$

The function  $\xi_{\text{cosmo}}$  is calculated from the power spectrum using the following procedure. We model the Ly $\alpha$  forest power spectrum including redshift-space distortions and nonlinear effects as

$$P(\vec{k}) = b_F^2 (1 + \beta \mu_k^2)^2 [P_{\text{peak}}(k) \exp(-k^2 \Sigma^2(\mu_k)/2) + P_{\text{smooth}}(k)], \quad (6.22)$$

where  $\mu_k \equiv \hat{z} \cdot \hat{k}$ ,  $b$  is the Ly $\alpha$  forest bias parameter and  $\beta$  is the redshift-space distortion parameter. Here, we have defined  $P_{\text{peak}}(k) = P_{\text{lin}}(k) - P_{\text{smooth}}(k)$ , where  $P_{\text{lin}}$  is the linear-theory matter power spectrum computed with CAMB (Lewis et al., 2000) using the fiducial cosmological parameters and  $P_{\text{smooth}}$  is the same CAMB power spectrum with the BAO feature erased by fitting a third order polynomial over the wiggled region. The exponential function in Eq. 6.22 models the anisotropic nonlinear broadening from structure growth with  $\Sigma^2(\mu_k) = \mu_k^2 \Sigma_{\parallel}^2 + (1 - \mu_k^2) \Sigma_{\perp}^2$  and is only applied to the BAO feature. The default values we have adopted are  $\Sigma_{\parallel} = 6.41 h^{-1} \text{Mpc}$  and  $\Sigma_{\perp} = 3.26 h^{-1} \text{Mpc}$ .

Instead of computing  $\xi_{\text{cosmo}}(\vec{r})$  with a direct 2D Fourier transform of  $P(\vec{k})$ , we first decompose  $P(\vec{k})$  in multipoles using the base of Legendre polynomials  $L_{\ell}(\mu_k)$ . In the distant observer approximation (Kaiser, 1987), the three first non-zero multipoles are  $\ell = 0, 2$  and 4. The Legendre

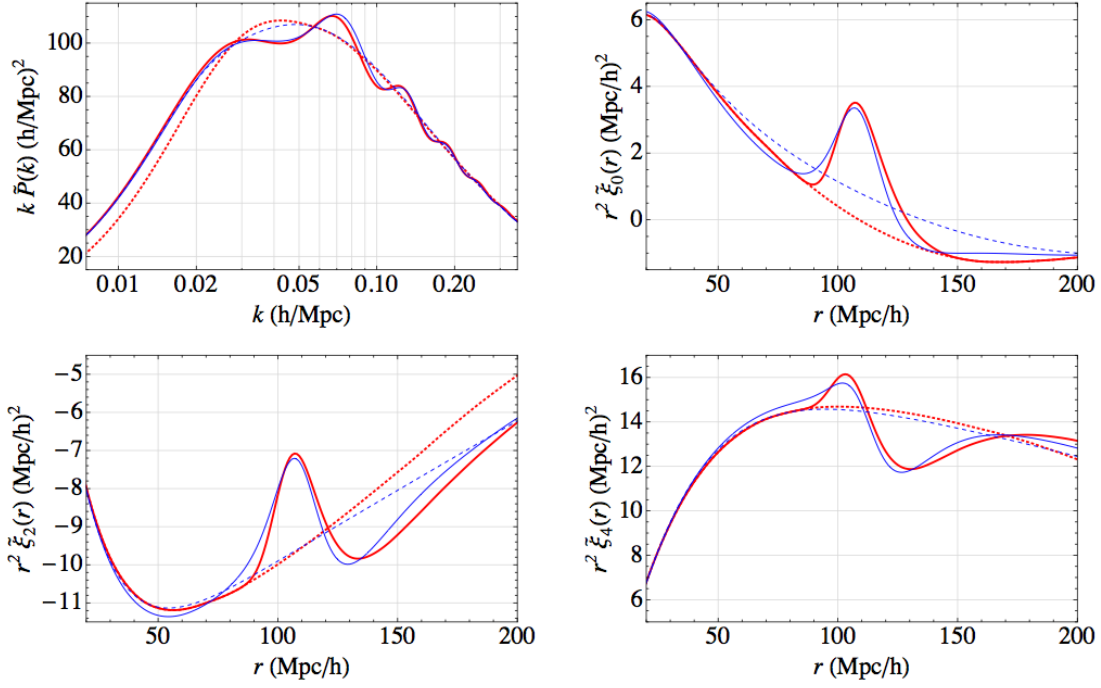


Figure 6.4: Cosmological linear models calculated for  $z = 2.25$  and assuming a flat universe with  $\Omega_\Lambda = 0.73$ ,  $h = 0.7$ ,  $\Omega_b h^2 = 0.0227$  and  $n_s = 0.97$ . Panels show the ( $k$ -weighted) power spectrum (top-left) and the ( $r^2$ -weighted) correlation function monopole (top-right), quadrupole (bottom-left), and hexadecapole (bottom-right). Curves are calculated with CAMB (thick, red) and using Eisenstein and Hu (1998) (light, blue) with solid curves showing the full cosmological model and dotted (dashed) curves showing the corresponding CAMB (no-wiggles of Eisenstein and Hu (1998)) smooth model. Figure extracted from Kirkby et al. (2013).

polynomials associated to these multipoles are given by

$$L_0(\mu) = 1 \quad (6.23)$$

$$L_2(\mu) = \frac{1}{2}(3\mu^2 - 1) \quad (6.24)$$

$$L_4(\mu) = \frac{1}{8}(35\mu^4 - 30\mu^2 + 3), \quad (6.25)$$

which give the following formulas for the power spectrum multipoles,

$$P_\ell(k) = \frac{2\ell + 1}{2} \int_{-1}^{+1} P(k, \mu_k) L_\ell(\mu_k) d\mu_k. \quad (6.26)$$

Due to the non-linear broadening of the peak, higher order multipoles of the power spectrum are also non zero but they are negligible when compared to the first ones. We compute the multipoles of the correlation function  $\xi_{\ell, \text{cosmo}}(r)$  by a Bessel transform of  $P_\ell(k)$ , written as

$$\xi_{\ell, \text{cosmo}}(r) = \frac{i^\ell}{2\pi^2} \int_0^\infty k^2 j_\ell(kr) P_\ell(k) dk, \quad (6.27)$$

where  $j_\ell$  is the spherical Bessel function. Figure 6.4 shows the input linear isotropic power spectrum  $P_{\text{lin}}(k)$  and the multipoles,  $\xi_{0, \text{cosmo}}$ ,  $\xi_{2, \text{cosmo}}$  and  $\xi_{4, \text{cosmo}}$  of the correlation function with their respective sideband functions (without BAO features).

Table 6.1: Summary of fitted parameters of the correlation function model for BAO measurements (Eq. 6.19).

Parameter	Description
$\alpha_{\parallel}$	Dilation parameter over the line of sight
$\alpha_{\perp}$	Dilation parameter in the transverse direction
$b_F$	Density bias
$\beta$	Redshift-space distortion parameter
$a_{\text{peak}}$	Amplitude of BAO peak ( $a_{\text{peak}} = 0$ means no peak)
$\Sigma_{\parallel}$	Non-linear damping scale in the parallel direction
$\Sigma_{\perp}$	Non-linear damping scale in the transverse direction
$a_{ij}$	Broadband amplitudes

Finally, the template correlation function  $\xi_{\text{cosmo}}$  is just the combination of these multipoles,

$$\xi_{\text{cosmo}}(\vec{r}) = \sum_{\ell=0,2,4} L_{\ell}(\mu) \xi_{\ell,\text{cosmo}}(r). \quad (6.28)$$

The broadband function  $\xi_{bb}$  is also written as a combination of Legendre polynomials times decreasing functions of separation  $r$ ,

$$\xi_{bb}(\vec{r}) = \sum_{j=0}^{j_{\text{max}}} \sum_{i=0}^{i_{\text{max}}} a_{ij} \frac{L_{2j}(\mu)}{r^i}, \quad (6.29)$$

where the  $L_{2j}$  is the Legendre polynomial of order  $2j$ . The coefficients  $a_{ij}$  are free parameters in the fit. Our standard model uses  $i_{\text{max}} = j_{\text{max}} = 2$  but we also performed the fit with other limits.

A summary of the fitted parameters is given in Table 6.1. There are 7 parameters defining  $\xi_{\text{cosmo}}$  and  $i_{\text{max}} \times j_{\text{max}}$  parameters defining the broadband function  $\xi_{bb}$ . In § 8.5 we analyze changes in results when letting some of these parameters fixed at some fiducial value or free (with or without a prior). The standard fit uses fiducial values for  $\Sigma_{\perp}$  and  $\Sigma_{\parallel}$ , and sets  $a_{\text{peak}} = 1$ . It thus has four physical free parameters,  $b, \beta, \alpha_{\perp}, \alpha_{\parallel}$ , and nine free broadband parameters  $a_{ij}$ . The standard fit uses a smaller domain in separations of the correlation function,  $40 < r < 180 h^{-1}\text{Mpc}$ , letting remain a total of 1515 bins (from the 2500 originally), resulting in 1502 degrees of freedom. The fitted parameters are such that they minimize

$$\chi^2 = \sum_{A,B} \left[ \hat{\xi}(A) - \xi_{\text{model}}(A, \alpha_{\parallel}, \alpha_{\perp}) \right] C^{-1}(A, B) \left[ \hat{\xi}(B) - \xi_{\text{model}}(B, \alpha_{\parallel}, \alpha_{\perp}) \right]. \quad (6.30)$$

The 68 and 95.5% confidence intervals (1 and  $2\sigma$ ) defining errors on  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  are computed through the marginalized  $\chi^2$  surfaces, and finding the  $\alpha$  values where  $\chi^2(\alpha) - \chi_{\text{min}}^2 = 1$  and 4, respectively.

In this Chapter was presented our full analysis chain for BAO measurements in the Ly $\alpha$  forest flux correlation function. This chain was developed and tested using mock and real data, as described in the following two chapters.



# Chapter 7

## Analyzing Mock Catalogs

In this chapter we validate the analysis chain described in Chapter 6 by using measurements on mock catalogs (Chapter 5). We perform tests on most of the analysis steps showing satisfying results and also their limitations. In § 7.3 we perform some tests related to the continuum fitting procedure, quantifying errors in our estimates. In § 7.4 we estimate the optimal weights on mock catalogs, comparing with the input values. The correlation function measurement is discussed in § 7.5 and its covariance matrix in § 7.6. Finally, in § 7.7 is shown that our analysis chain is unbiased when applied to our mock catalogs.

All subjects treated in this chapter are part of my original contribution.

### 7.1 Generalities

$\text{Ly}\alpha$  forest mock catalogs are created based on a two step process: first, production of absorption fields with correlations given by a cosmological model (see § 5.1); second, production of realistic spectra by adding quasar continua, metals, high column density systems, noise, sky subtraction errors and noise mis-calibration (see § 5.2). Correlation function measurements on mock catalogs are affected by all features added in the expansion. In this Chapter, the effect of each feature on results is studied. We use the following samples of mock spectra:

- noiseless forests: the pure absorption field;
- with continuum model and noise: we test the impact of the continuum fitting;
- mocks with sky subtraction residuals;
- mocks with errors in the pixel noise estimate;
- “expanded mocks”, that include all features above.



## 7.2 Distortion of correlations

We start by a theoretical motivation of the tests. The measured correlation function can be distorted from the true cosmological correlation due to spurious correlations coming from continuum fitting, sky residuals and noise. This section is based on Kirkby et al. (2013).

The correlation function between two pixels indexed  $i$  and  $j$  in a given separation bin  $A$  can be written as

$$\xi_{ij \in A} = \langle \delta_i \delta_j \rangle - \langle \delta_i \rangle \langle \delta_j \rangle = \frac{\langle F_i F_j \rangle - \langle F_i \rangle \langle F_j \rangle}{\bar{F}(\lambda_i) \bar{F}(\lambda_j)}, \quad (7.1)$$

where

$$F_i = \frac{f_i - f_{i,\text{sky}}}{f_{i,\text{cont}}} \quad (7.2)$$

is the measured transmitted flux fraction, which is the total flux  $f_i$  subtracted by the estimated sky flux  $f_{i,\text{sky}}$  and normalized by an estimate of the continuum level  $f_{i,\text{cont}}$ , and  $\bar{F}(\lambda_i)$  is the estimated mean transmission. Since these quantities are estimates, we can express them as function of the true quantities (denoted with tildes) as

$$f_i = \tilde{f}_i + \epsilon_i \quad (7.3)$$

$$f_{i,\text{sky}} = \tilde{f}_{i,\text{sky}} + s(\lambda_i) + \epsilon_{i,\text{sky}} \quad (7.4)$$

$$f_{i,\text{cont}} = \tilde{f}_{i,\text{cont}} + c_i \quad (7.5)$$

where  $\epsilon_i$  and  $\epsilon_{i,\text{sky}}$  represent the pixel noise,  $s(\lambda_i)$  accounts for any wavelength-dependent residual sky-subtraction bias (e.g. Fig. 5.6), and  $c_i$  describes the continuum fitting error. Putting these expressions into Eq. 7.2, we find

$$F_i = \frac{\tilde{f}_i - \tilde{f}_{i,\text{sky}} - s(\lambda_i) + \epsilon_i - \epsilon_{i,\text{sky}}}{\tilde{f}_{i,\text{cont}} + c_i} \approx \tilde{F}_i [1 - S_i - C_i + E_i] \quad (7.6)$$

where we defined the dimensionless quantities

$$\tilde{F}_i \equiv \frac{\tilde{f}_i - \tilde{f}_{i,\text{sky}}}{\tilde{f}_{i,\text{cont}}}, \quad S_i \equiv \frac{s(\lambda_i)}{\tilde{f}_i - \tilde{f}_{i,\text{sky}}}, \quad C_i \equiv \frac{c_i}{\tilde{f}_{i,\text{cont}}}, \quad E_i \equiv \frac{\epsilon_i - \epsilon_{i,\text{sky}}}{\tilde{f}_i - \tilde{f}_{i,\text{sky}}}, \quad (7.7)$$

and assuming  $C_i \ll 1$ . Taking ensemble averages, we obtain

$$\xi_{ij} \approx r_{ij} \left[ \tilde{\xi}_{ij} (1 + A_{ij}) + B_{ij} \right] \quad (7.8)$$

where  $\tilde{\xi}_{ij} \equiv \langle \tilde{F}_i \tilde{F}_j \rangle / \langle \tilde{F}_i \rangle \langle \tilde{F}_j \rangle - 1$ ,  $r_{ij} \sim 1$  describes the discrepancy between the assumed and true mean transmission.  $A_{ij}$  is a multiplicative distortion of the true correlation function

$$A_{ij} \equiv \langle S_i S_j \rangle + \langle C_i C_j \rangle + \langle E_i E_j \rangle + \langle S_i \rangle \langle C_j \rangle + \langle S_j \rangle \langle C_i \rangle - \langle S_i \rangle - \langle S_j \rangle - \langle C_i \rangle - \langle C_j \rangle, \quad (7.9)$$

and  $B_{ij}$  is an additive distortion

$$B_{ij} \equiv \langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle + \langle C_i C_j \rangle - \langle C_i \rangle \langle C_j \rangle + \langle E_i E_j \rangle. \quad (7.10)$$

We assumed that the  $S$ ,  $C$  and  $E$  are mutually uncorrelated.

Equation 7.8 shows that the estimated correlation function is distorted by correlations in errors on continuum estimation, sky residuals, and noise. Any multiplicative distortion comes mainly from a mismatch between the assumed and the true mean transmission, whereas additive distortion is due to correlated continuum fit errors.

The distortion is observed in mock and real data, as we will see in the two following chapters. In mocks, the distortion can be quantified since we know the true correlation function. In the following, we show that sky residuals also distort the correlation as predicted by Eq. 7.8. This distortion is the main motivation for including an additive broadband function when fitting for BAO, as explained in § 6.4.

## 7.3 Continuum fitting

### 7.3.1 Residuals

The first natural question addressed to continuum fitting methods is how well they estimate the original continua. On mock catalogs the true continuum level  $C_{\text{true}}(\lambda)$  is known, therefore it can be used to study errors in the continuum estimate. In this section the analysis concerns results from method C2 only because it is the method with which I have worked. The C1 and C3 methods are furnished by other collaborators.

Figure 7.1 shows differences between the true and the estimated continuum of a random mock forest. The estimated continuum does not reproduce exactly the original continuum. PCA-generated continua from mock spectra have an intrinsic spectral shape diversity not taken into account by method C2, that uses the same spectral shape to fit all forests. Therefore, our method adds errors to transmission estimates. However, these differences can bias the correlation function only if they are correlated among different lines-of-sight, as pointed out in § 7.2. Since generated mock continua are independent<sup>1</sup> those differences are in principle not correlated.

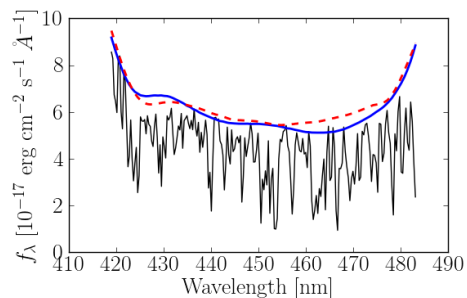


Figure 7.1: Fitted (blue solid) and true (red dashed) continuum level over a randomly selected mock forest.

For a consistency check, we computed the mean and scatter of residuals  $\hat{C}(\lambda) - C_{\text{true}}(\lambda)$  as a function of the rest-frame wavelength and forest mean signal to noise ratio (S/N). The S/N of each forest is computed as the mean S/N of forest pixels ( $104 < \lambda < 120$  nm).

<sup>1</sup>In principle, the continuum shape of a quasar depends on the quasar local environment, therefore making those shapes uncorrelated among quasars.

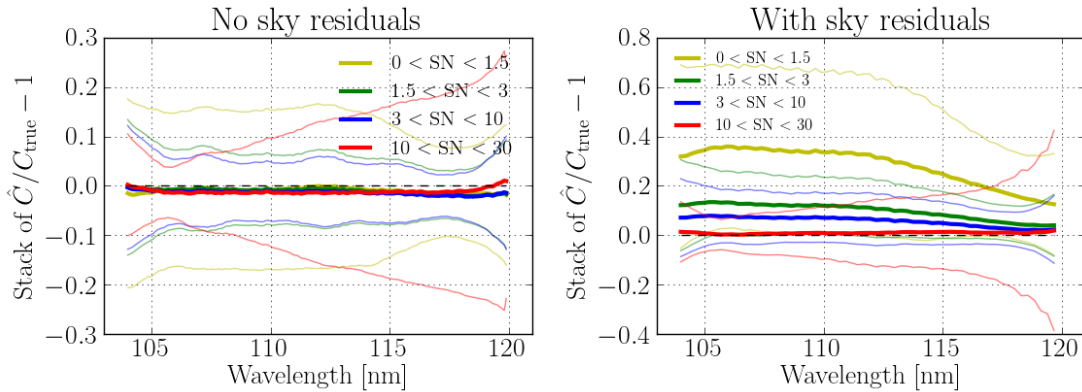


Figure 7.2: Mean relative deviation of the estimated continuum with respect to the true mock continuum as a function of rest-frame wavelength and signal to noise. Left (right) panel shows results on mocks without (with) sky residuals. The thin lines show the mean  $\pm 1\sigma$  values.

Figure 7.2 show results with two types of mocks : with and without sky residuals added, both without noise mis-calibration (see next). We see that mocks with no sky residuals have almost no bias in the continuum estimation, while those containing residuals have an average bias that increases for lower S/N. For the second case, we observe a clear dependence of the mean residual error with S/N. Forests with  $S/N > 3$ , corresponding to  $\sim 50\%$  of the forests, have relative mean residuals of at most 7% with scatter around the mean consistent with zero bias. Forest with  $S/N < 1.5$  have larger biases in their continua because of the high amplitude of sky residuals relative to the actual quasar flux (Fig. 5.6). However, forests within this S/N range represent less than 10% of the forests and contribute even less to the correlation function due to their low weight.

We computed continuum residuals for mocks with mis-calibrated pixel noise. Our analysis estimates a correction for this mis-calibration ( $\eta(z)$  in Eq. 6.11). The iterative procedure (§ 6.1.3) that computes  $\eta(z)$  correctly takes into account this systematic and do not introduce biases in the continuum estimation.

### 7.3.2 Convergence of iterative method for C2

The iterative procedure for the method C2 (§ 6.1.3) was implemented to make measurements less dependent on the mock input flux PDF.

Figure 7.3 shows the convergence of estimates of the mean continuum shape  $\bar{C}(\lambda_{\text{rf}})$ , the mean transmission  $\bar{F}(z)$ , the noise correction  $\eta(z)$  and the forest intrinsic variance  $\sigma_{\text{LSS}}^2(z)$  on mocks containing sky residuals and mis-calibrated noise. All estimates converge near the third iteration. Results from first iteration differ considerably from those of the last one, showing the importance of the iterative procedure in the estimate of flux fluctuations. However, as we will show in § 7.4, the final measurements of  $\sigma_{\text{LSS}}(z)$  and  $\eta(z)$  do not fully agree with the input values. The reasons for this disagreement are discussed in § 7.4.

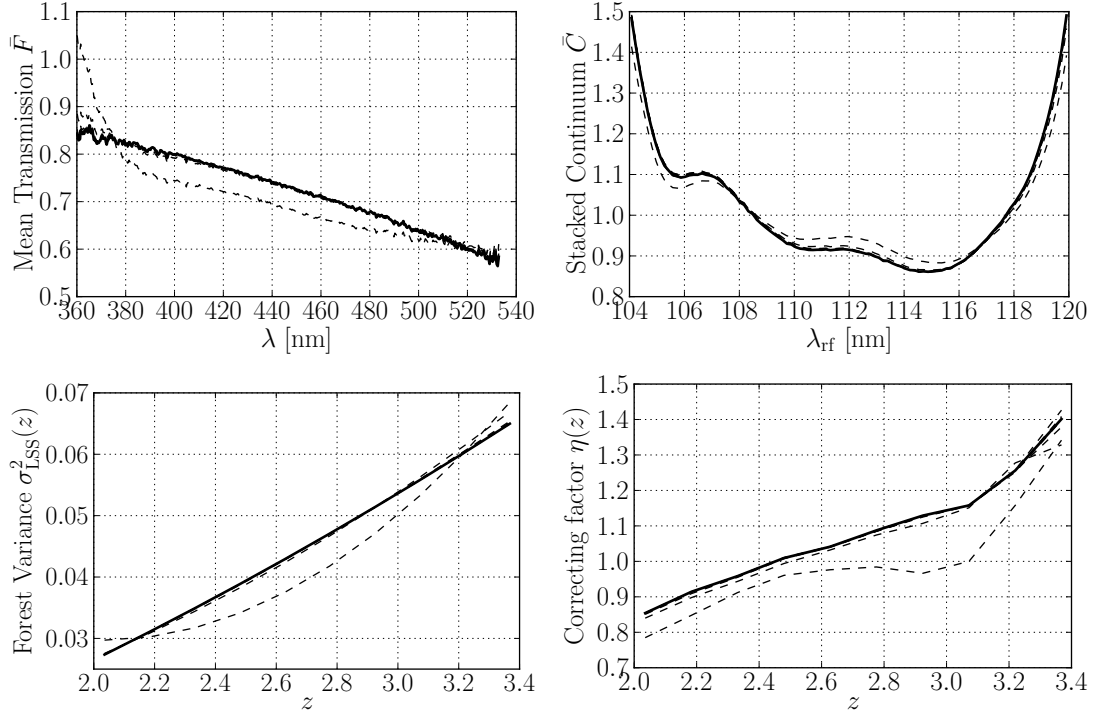


Figure 7.3: Convergence of different observables with the iterative procedure for continuum fitting method C2. Dashed lines show first five iterations and solid line the last one. Due to the fast convergence at the third iteration, not all curves are visible. Top left: mean transmission as a function of redshift. Top right: Mean stacked continuum shape as function of rest-frame wavelength. Bottom left: intrinsic variance of the forest function of redshift. Bottom right: correction for the pixel noise estimates.

### 7.3.3 Mean transmission

The estimation of the mean transmission level  $\bar{F}$  of forests results from the requirement of a zero mean  $\delta_F$  field. We remind that only the C2 method is able to measure  $\bar{F}$  since C1 directly measures the product  $C(\lambda)\bar{F}$  and C3 fits continua by matching  $\bar{F}$  to a given model. Note, however, mean transmission measurements might contain any other systematic effect present in average on spectra, such as sky residuals. Therefore,  $\bar{F}$  is not a measurement of the true mean transmission of the absorption field if the systematic effects are not suppressed.

Figure 7.4 shows the estimated mean transmission relative to the input model for mocks with increasing number of added systematic effects. In order to see the effect of our continuum fitting method, we also estimated  $\bar{F}$  on forests built with the true continuum level  $C_{\text{true}}$ . Noise miscalibration does not affect our results, as also observed with the continuum residuals (§ 7.3.1), the main systematic effect comes from sky residuals. Noiseless mocks have their mean transmission correctly estimated (magenta). When using noisy spectra but with the true continuum and no sky residuals, a small bias of less than one percent in the  $\bar{F}$  estimate is seen (red line). The effect of the sky residuals is seen by the green and blue lines at low redshift where these residuals introduce large fluctuations (the same shape as in left panel of Fig. 5.6). When fitting the continuum level, the  $\bar{F}$  estimate is biased differently, smoothly increasing with redshift (blue and yellow). In principle, these errors distort the correlation function multiplicatively, but, as we can see in Fig. 7.4, the multiplicative factor is very close to unity (deviations up to 2%).

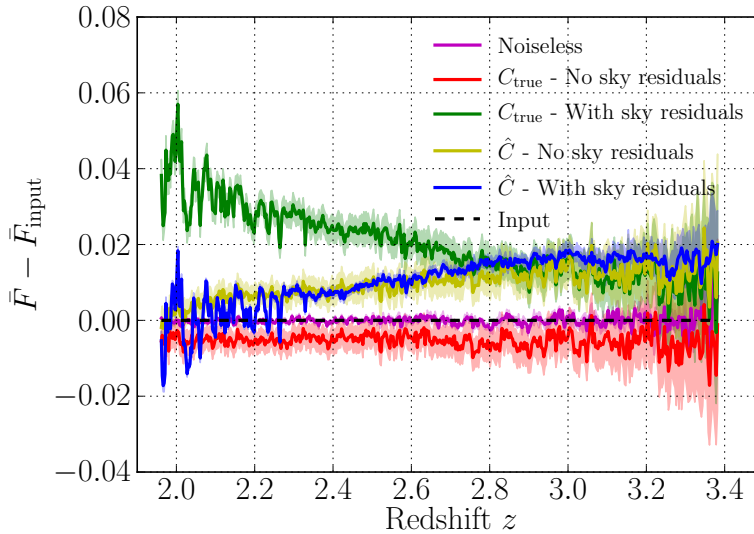


Figure 7.4: Effect of continuum fitting and sky subtraction residuals in the estimated mean transmission (with respect to the input values). Mocks surveys with different systematics are considered: noiseless and true continuum (magenta); true continuum, noise and no sky residuals (red); true continuum, noise and sky residuals (green); estimated continuum, noise, no sky residuals (yellow); estimated continuum, noise and sky residuals (blue).

Furthermore, the BAO peak position is not affected by discrepancies in the  $\bar{F}$  estimate because the consequent distortion is smooth.

## 7.4 Pixel variances

The procedure used to compute optimal pixel weights  $w_i$  for the correlation function, described in § 6.2.1, assumes that the pixel variance  $\sigma^2$  has two contributions. The first is intrinsic to the forest and is related to the fluctuations of the underlying density field,  $\sigma_{\text{LSS}}^2(z)$ . The second contribution comes from the instrumental pixel noise estimated by the pipeline,  $\sigma_{\text{pipeline}}$ , which is known to be biased, needing a wavelength (or redshift) dependent correction,  $\eta(z)$ . In this section, we use mocks to test whether the input values are recovered, and if not, what is the error on those estimates. As in the previous sections, we analyze the effect of different systematics on the pixel variance estimates using mocks defined in § 7.1.

Figure 7.5 compares the intrinsic variance (left) and the noise correction estimates (right panel) with input values for different mock samples:

- For noiseless mocks, the input values of  $\sigma_{\text{LSS}}^2$  are correctly recovered. When using either the fitted continuum or the true continuum, the intrinsic variance is slightly overestimated: up to 10% at  $z < 3.0$  for the fitted continuum and up to 8% for the true continuum. At higher redshift, the disagreement is slightly larger but it concerns a small fraction of the pixels (less than 10%). The noise mis-calibration does not affect significantly these conclusions.
- The noise correction  $\eta(z)$  for noiseless mocks is unity by definition (magenta). For mocks without noise mis-calibration, a flat function of  $z$  is recovered for fitted (yellow) and true

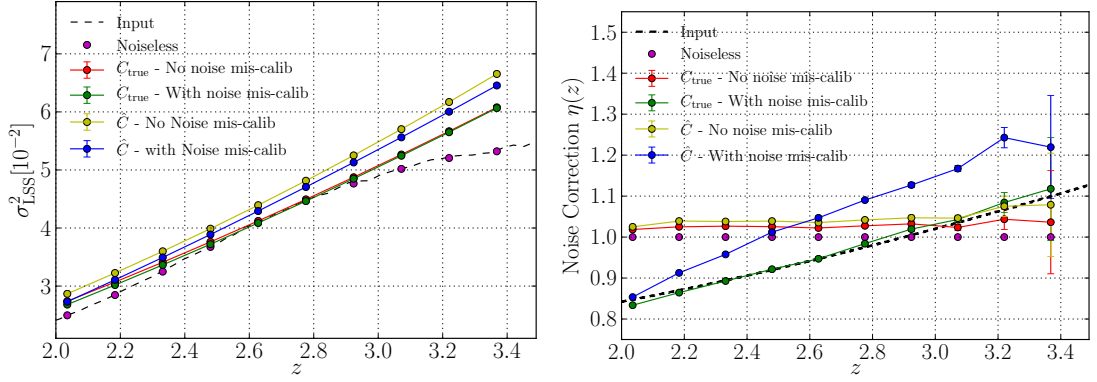


Figure 7.5: Effects of noise mis-calibration and continuum fitting on the estimate of the intrinsic forest variance  $\sigma_{\text{LSS}}^2$  (left panel) and the noise correction term  $\eta$  (right panel). The dashed line shows the input values.

continuum (red), but with values  $\sim 3$  to 5% higher than unity. Purposefully including a biased estimate of pixel errors on mocks, we see that fitted (blue) and true continuum (green) cases does not give the same result. While using the true continuum  $C_{\text{true}}$  yields the correct mis-calibration, using the fitted continuum  $\hat{C}$  gives a biased estimate of  $\eta(z)$ . This would indicate that continuum fitting adds variance to the pixels. To understand this bias, we analyzed some fits of the relation between  $\sigma^2$  and  $\sigma_{\text{pipeline}}^2$ , discussed next. Sky residuals do not affect these results, since they modify  $\bar{F}$  and not the variances.

Figure 7.6 shows examples of the fitted pixel variance model over measurements on mocks with fitted continua, for two redshift bins. In practice for a given bin in redshift, the noise correction  $\eta(z)$  is the inverse of the slope of the fitted model, and the intrinsic variance  $\sigma_{\text{LSS}}^2(z)$  is the intercept of the model with the y-axis (i.e., the variance when  $\sigma_{\text{pipeline}}^2 = 0$ ). Looking at residuals of the fit we see that the relation between  $\langle \delta_F^2 \rangle$  and  $\sigma_{\text{pipeline}}^2$  is not exactly linear. Also, we see that the intrinsic variance is underestimated because points at low  $\sigma_{\text{pipeline}}^2$  are above the model. However, in the left panel of Fig. 7.5 we saw that the intrinsic variance is actually overestimated with respect to input values. These observations are similar for other mock sets, e.g., with true continuum, or with noise mis-calibration or sky residuals. An additional source of variance is being added to the intrinsic variance not coming from continuum fitting or sky residuals. The intercepts and slopes are thus less representative of the true behavior of the variances, that would need a model with more free parameters to be correctly described. Nevertheless, residuals in Fig. 7.6 are less than one standard deviation from the fitted model for most of the range, so we consider our linear approximation valid at first order. In any case, the variances are not being underestimated and we can consider this slight overestimation as a conservative marge on errors.

## 7.5 Correlation function

In this section we study the measured correlation function obtained from the estimator described in § 6.2 applied to mock catalogs, studying the impact of continuum fitting and systematic effects.

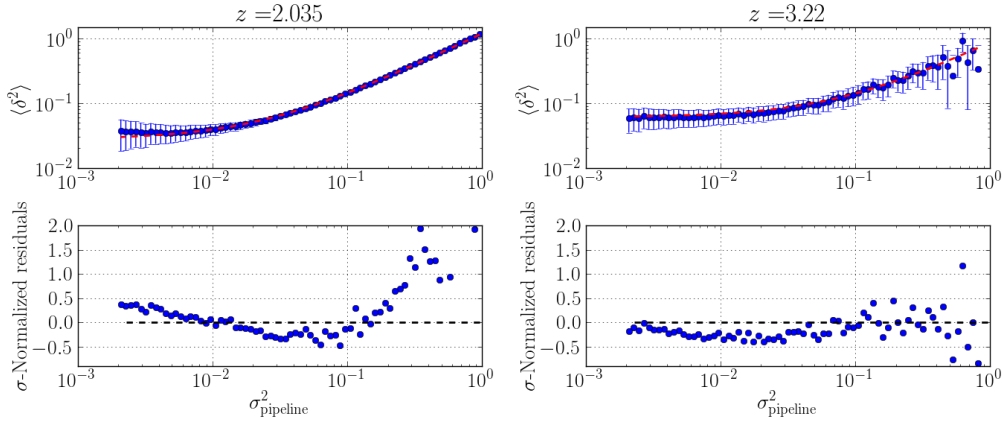


Figure 7.6: Fits of pixel variance model on a mock realization for two redshift bins (shown in the title). The bottom plots show the residuals of the fit normalized by the error bars. We show only results using forests computed using the estimated continua. Similar results are obtained on mock sets with true continua or without systematic effects.

Figure 7.7 presents all 100 correlation functions computed with the standard analysis on “expanded mocks”. The correlation functions are represented by averages in  $\mu = r_{\parallel} / \sqrt{r_{\parallel}^2 + r_{\perp}^2}$  ranges. Respectively, three weighted averages of  $\hat{\xi}$  are plotted:  $\xi_{\perp}$  ( $0 < \mu < 0.5$ ),  $\xi_{int}$  ( $0.5 < \mu < 0.8$ ) and  $\xi_{\parallel}$  ( $0.8 < \mu < 1$ ), as function of the comoving separation  $r$ . Blue solid lines show the mean stacked measurement, while dashed lines show the one standard deviation region.

The BAO peak shows itself very clearly, in particular on  $\xi_{\parallel}$ . As explained in Chapter 2, the enhancement of the peak in the radial direction is due to the strong redshift-space distortions (RSD) of the Ly $\alpha$  forest.

Since the goal of this study is the measurement of the BAO peak, we are not interested in the clustering on scales smaller than  $40 h^{-1}\text{Mpc}$ , where the theoretical modeling is difficult, or larger than  $180 h^{-1}\text{Mpc}$ , where we do not expect significant signal. Therefore, our standard range for model fitting is defined by separations between  $40 < r < 180 h^{-1}\text{Mpc}$  (see § 7.7).

A direct comparison of results in Fig. 7.7 with the input theory is not possible because measurements are affected by the distortion, as discussed in § 7.2. Figure 7.8 compares the correlation function of the stack of ten realizations computed using the true continuum level (red and green) with the stack of 100 realizations with fitted continuum (blue). The input theory (dashed black) and the correlation function of noiseless forests (magenta) are also presented. We see that, even using the actual continuum level with no sky residuals added (red), the mean correlation function does not agree with the input theory, as opposed to the correlation function computed with noiseless forests. This happens because of the difference in the effective redshift  $z_{\text{eff}}$  of the two measurements. The noise in forest pixels depend on redshift, changing the relative weight of low and high redshift pixels. The effective redshift for the true continuum case is  $z_{\text{eff}} = 2.33$  and for the noiseless case is  $z_{\text{eff}} = 2.18$ , which yields an effective amplitude difference of 20%, assuming a  $(1+z)^{3.8}$  evolution of the correlation function amplitude, in agreement with Fig. 7.8. For the fitted continuum case, the effective redshift is the same as for the true continuum case, confirming that the differences indeed come from distortion effects.

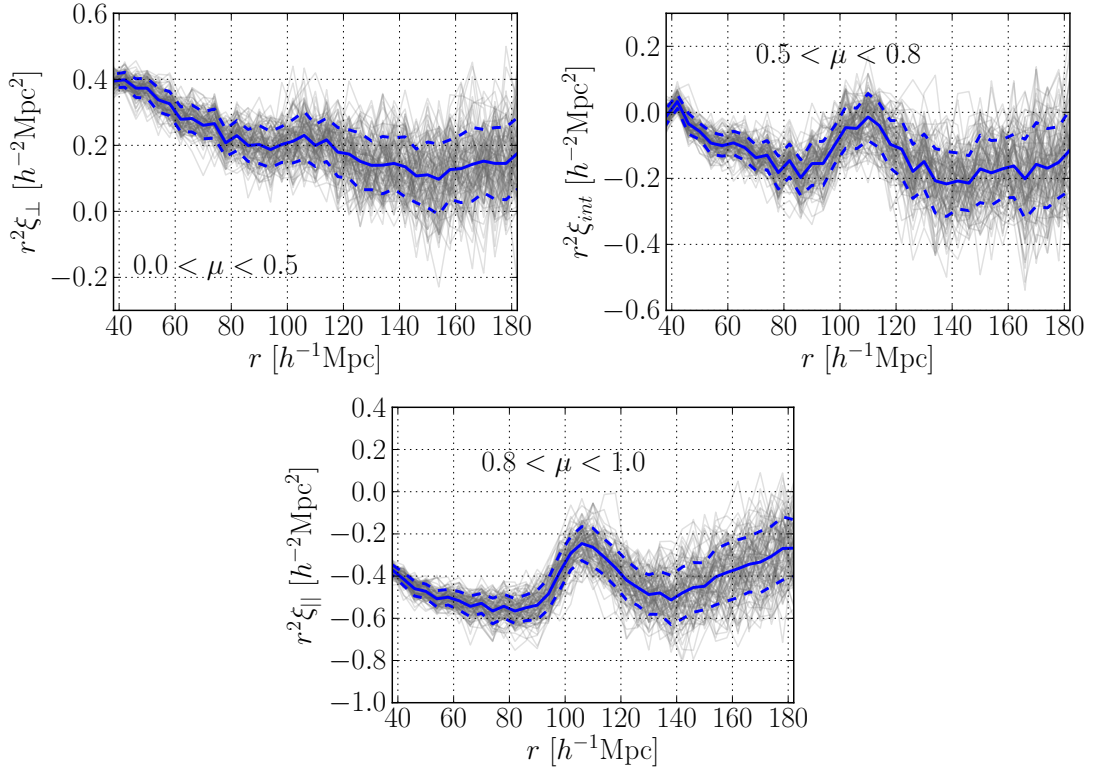


Figure 7.7: Correlation function from mock DR11 catalogs represented in averages over  $\mu$  ranges,  $\xi_{\perp}$  ( $0 < \mu < 0.5$ , top left),  $\xi_{int}$  ( $0.5 < \mu < 0.8$ , top right) and  $\xi_{\parallel}$  ( $0.8 < \mu < 1$ , bottom). The average and scatter around the mean ( $\pm 1$  standard deviation) are shown in blue solid and dashed lines respectively.

In § 7.7, we analyse the dependence of the inferred values for the BAO peak position and its errors on the continuum used (true or estimated).

## 7.6 Covariance matrix

The covariance matrix  $C(r_{\parallel}, r_{\perp}, r_{\parallel}', r_{\perp}')$  for the 100 expanded mock catalogs were computed individually using the sub-sampling method (§ 6.3.1) with the smoothing procedure that assumes a dependence of the correlation coefficients  $c(r_{\parallel}, r_{\perp}, r_{\parallel}', r_{\perp}')$ , defined in Eq. 6.17, only on  $\Delta r_{\parallel} = r_{\parallel} - r_{\parallel}'$  and  $\Delta r_{\perp} = r_{\perp} - r_{\perp}'$ . The sub-sampling method is computationally much faster than the Wick approximation calculation (§ 6.3.2). Tests using data (see § 8.4) show that both methods agree at surprisingly high levels. Therefore, all following tests are performed using smoothed sub-sampling covariance matrices.

The diagonal elements of  $C$  (the variance of  $\xi$ ,  $\text{Var}(\xi)$ ) have simple structure. The quantity  $\text{Var}(\xi) \times n_{\text{pairs}}$  is basically scale dependent. Figure 7.9 shows histograms of this quantity for four sets of mock catalogs (the same as previously). The noise mis-calibration does not affect these variances, while sky residuals *decrease* the variance of the correlation function by 10%. This is probable because the mean transmission  $\bar{F}$  is over estimated for mocks containing sky residuals,



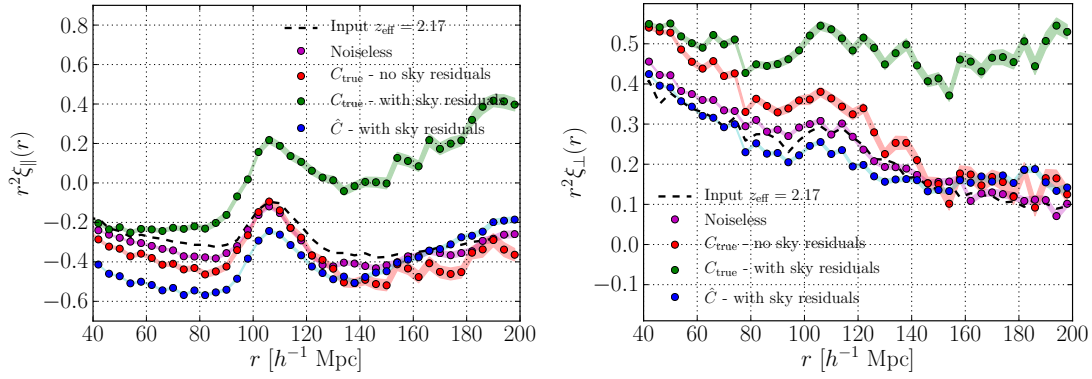


Figure 7.8: Radial (left) and transverse (right) correlation functions of the stack of mock realizations using: raw forests without noise (magenta points, 10 realizations), true continuum without sky residuals (red points, 10 realizations), true continuum containing sky residuals (green points, 10 realizations), and estimated continuum with sky residuals (blue points, 100 realizations). The dashed line shows the input theory of mock forests computed at the effective mean redshift of noiseless forests.

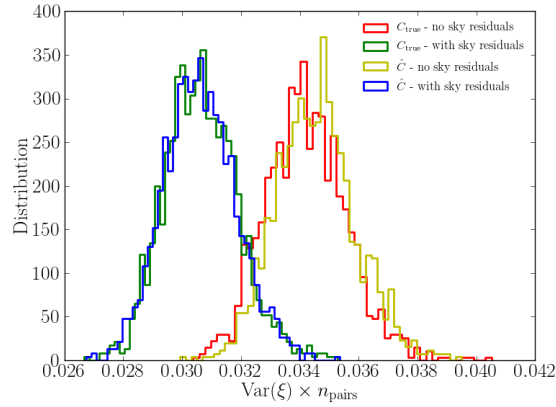


Figure 7.9: Histograms of the diagonal elements of the covariance matrix times the number of pairs contributing to each bin, for four sets of mock catalogs.

reducing the variance of  $\delta_F = F/\bar{F} - 1$ . Using the true continuum instead of the fitted one does not change variances.

The smoothing procedure of the covariance matrix should work if the assumption that all correlation coefficients corresponding to the same  $\Delta r_{\parallel}$  and  $\Delta r_{\perp}$  are drawn from the same distribution, and that this distribution does not depend on  $r_{\parallel}$  or  $r_{\perp}$  for given values of  $\Delta r_{\parallel}$  and  $\Delta r_{\perp}$ . Figure 7.10 presents distributions of  $c$  as functions of  $r_{\parallel}$  and  $r_{\perp}$ , given  $(\Delta r_{\parallel}, \Delta r_{\perp})$ . We see that these distributions, or their means, do not change significantly with either  $r_{\parallel}$  or  $r_{\perp}$ , validating our assumptions.

Figure 7.11 shows the distributions of the correlation coefficients for the six smallest combinations of  $(\Delta r_{\parallel}, \Delta r_{\perp})$ . From each distribution it is possible to estimate their mean value  $\bar{c}$  and the error  $\sigma_{\bar{c}}$  of this mean for the pair  $(\Delta r_{\parallel}, \Delta r_{\perp})$ . If the averaged coefficient is less than  $2\sigma_{\bar{c}}$  distant from zero, we set it to zero to avoid noisy coefficients.

Figure 7.12 shows the smoothed correlation coefficients for three values of  $\Delta r_{\perp} = 0, 4$  and  $8 h^{-1}\text{Mpc}$  as functions  $\Delta r_{\parallel}$ . We observe that the coefficients fall very quickly to zero as  $\Delta r_{\perp}$

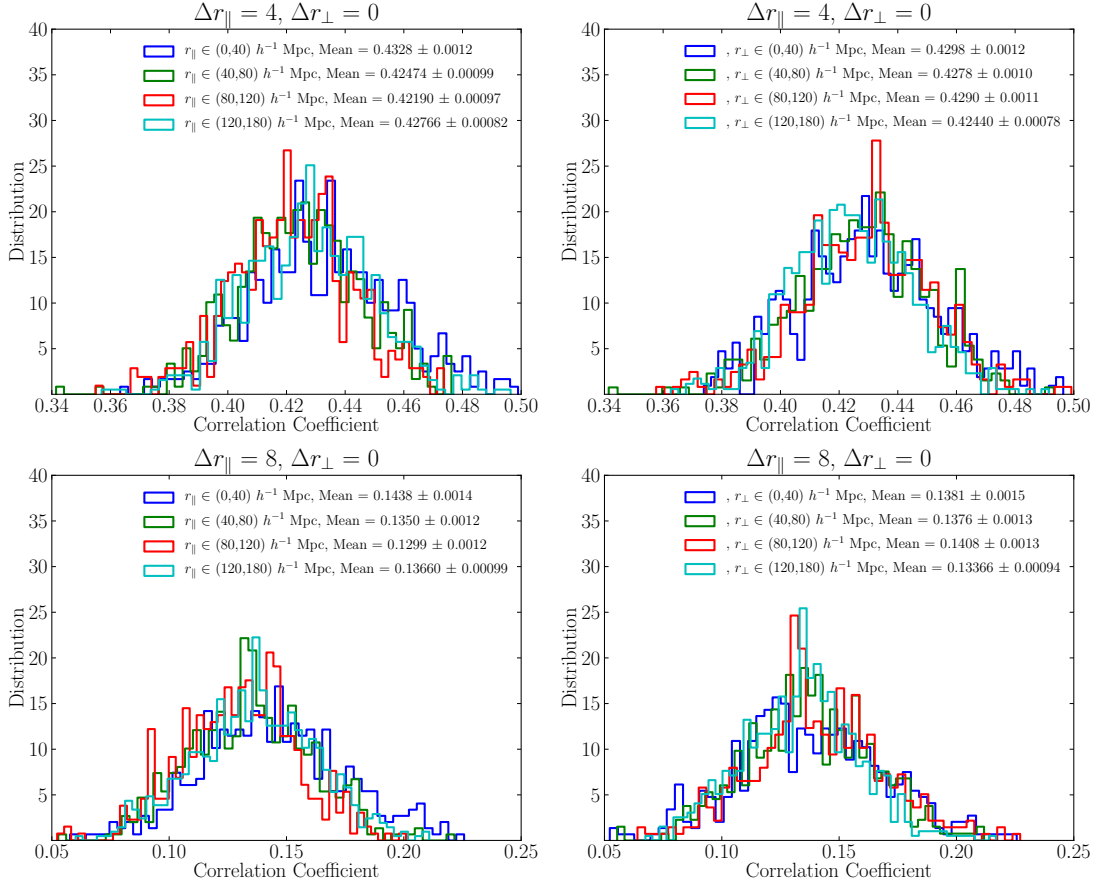


Figure 7.10: Histograms of correlation coefficients  $c(r_{\parallel}, r_{\perp}, r'_{\parallel}, r'_{\perp})$  of four ranges of  $r_{\parallel}$  (left panels) and  $r_{\perp}$  (right panels), for  $\Delta r_{\perp} = 0$  and  $\Delta r_{\parallel} = 4$  (top panels) and  $8 h^{-1}$  Mpc (bottom panels).

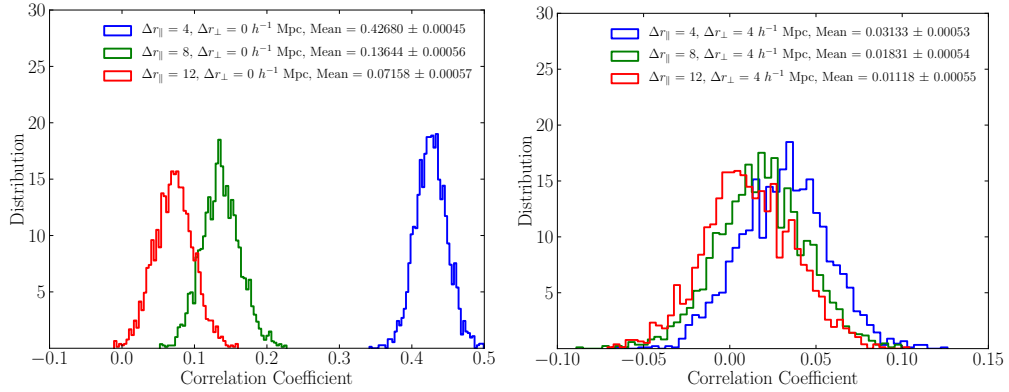


Figure 7.11: Histograms of the correlation coefficient elements for six pairs of  $(\Delta r_{\parallel}, \Delta r_{\perp})$ .

increases, becoming smaller than 1% for  $\Delta r_{\parallel} > 10 h^{-1}$  Mpc when  $\Delta r_{\perp} > 4 h^{-1}$  Mpc. In galaxy clustering measurements, the covariance matrix is estimated from measurements on multiple realizations of mock catalogs. The measurement described in § 4.2 uses 600 of them. For Ly $\alpha$  the 100 realizations are not sufficient for a precise estimate of the covariance matrix. Furthermore, those mock catalogs do not reproduce all identified systematic effects and have disagreements with respect to real data (see § 8.3) that affect the covariance matrix. Nevertheless, ignoring those factors we computed the covariance matrix using the 100 correlation measurements, then

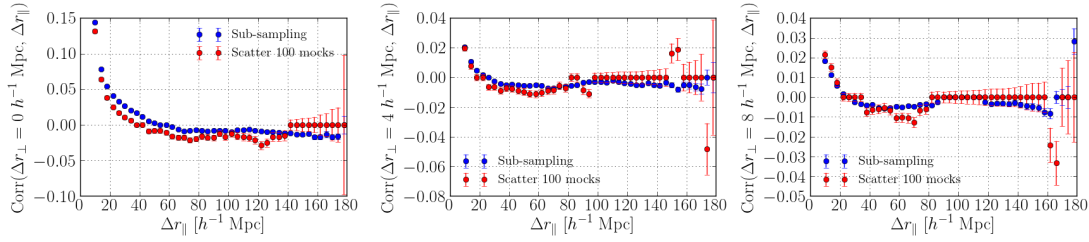


Figure 7.12: Averaged correlation coefficients as a function of  $(\Delta r_{\parallel}, \Delta r_{\perp})$  for the three first  $\Delta r_{\perp}$  bins. We compare coefficients from a sub-sampling covariance matrix from a single realization (blue points) with the covariance of the 100 mock realizations (red points).

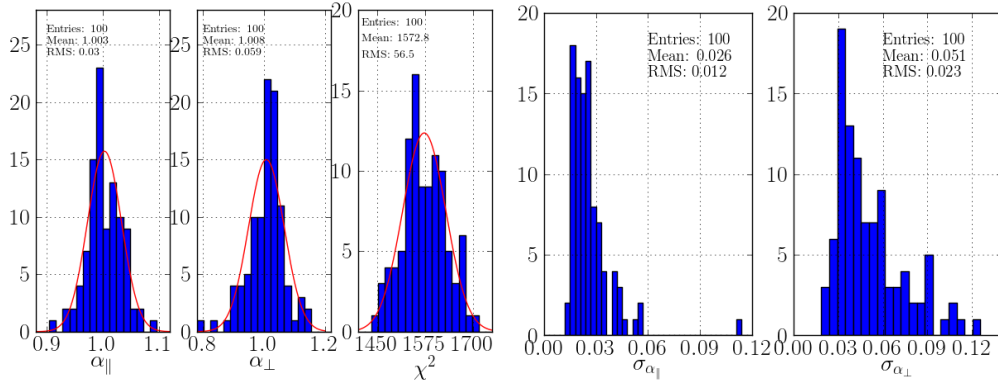


Figure 7.13: Summary of the results of fits for  $(\alpha_{\parallel}, \alpha_{\perp})$  for the 100 mock catalogs. The histograms show the best-fit values, the minimum  $\chi^2$  values and the  $1\sigma$  uncertainties.

smoothing it as for the sub-sampling method. The averaged correlation coefficients  $c(\Delta r_{\parallel}, \Delta r_{\perp})$  are shown in Fig. 7.12 and compared to those obtained by sub-sampling of one realization. There is a significant disagreement between both methods for almost all values of  $(\Delta r_{\parallel}, \Delta r_{\perp})$  where the “scatter” correlations are not consistent with zero. The reason for this disagreement is still an open question.

## 7.7 Measuring BAO

Once correlation functions and covariance matrices are estimated for each mock realization, we performed fits of the BAO template in order to recover the BAO peak position. Since the answer is known ( $\alpha_{\parallel} = 1, \alpha_{\perp} = 1$ ), this test can tell us about possible biases in the standard analysis.

The first two panels of Figure 7.13 show the distribution of recovered values for  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  and their estimated  $1\sigma$  errors. The mean recovered  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  are consistent with unity showing no significant bias in our methodology for extracting the BAO peak position. The number of measurements of  $\alpha_{\parallel}$  within 1 and  $2\sigma$  is, respectively, 61 and 93, consistent with the expected numbers, 68 and 95.5. For  $\alpha_{\perp}$  these numbers are 68 and 95. For the combined measurement of  $(\alpha_{\parallel}, \alpha_{\perp})$ , 70 are within  $1\sigma$  and 93 within the  $2\sigma$  contours.

In the third panel of Fig. 7.13, we see that the  $\chi^2$  distribution has mean  $\sim 1573$  and r.m.s. of 56, which is more than  $1\sigma$  greater than the expected value of 1502 corresponding to the number of

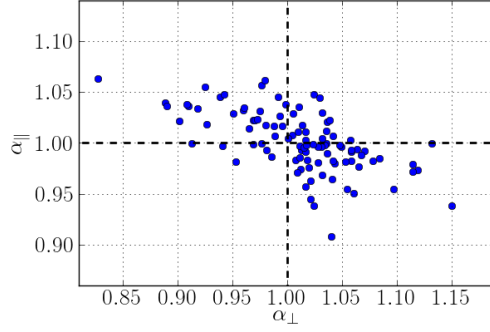


Figure 7.14: Measured  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  for the 100 mock catalogs.

degrees of freedom of the standard fit (1515 separation bins and 13 parameters of Table 6.1). This would mean that either our model is not accurate enough or that our covariance matrix is slightly underestimated. We remind that our template  $\xi_{\text{cosmo}}$  and our broadband  $\xi_{bb}$  are both computed with a limited number of multipoles (see § 6.4). Increasing the number of multipoles in the broadband adds free parameters and the  $\chi^2$  is then reduced (see § 8.5). On the other hand, our covariance matrix for mocks is estimated using the sub-sampling method with smoothing, that might have small inaccuracies. Nevertheless, if the underestimation of covariance matrix is the reason for this distribution of  $\chi^2$ , the error would be of less than 2% which does not change significantly our final errors on  $(\alpha_{\parallel}, \alpha_{\perp})$ .

The distributions of estimated errors  $\sigma_{\alpha_{\parallel}}$  and  $\sigma_{\alpha_{\perp}}$  on the best-fit dilation parameters (two right panels in Fig. 7.13) show that the typical error on  $\alpha_{\parallel}$  is 2.6% and on  $\alpha_{\perp}$ , 5.1%, both with a relative scatter of  $\sim 50\%$ . In § 8.5 it will be shown that the errors obtained with real data are consistent with these distributions.

Figure 7.14 shows the anti-correlation between the measurements of  $\alpha_{\parallel}$  and  $\alpha_{\perp}$ . The correlation coefficient is about  $-0.6$ .

In order to quantify the effect of continuum fitting on BAO estimates, we fitted for  $(\alpha_{\parallel}, \alpha_{\perp})$  on mock correlation functions obtained with the true continuum level. Figure 7.15 compares values of retrieved  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  parameters between ten mocks with true continuum and the same mocks with fitted continuum. In the radial direction, the BAO peak is measured with a mean uncertainty of 2.1% when using the true continua while for fitted continua we have a mean 2.6% error in  $\alpha_{\parallel}$ , corresponding to a 20% difference. In the transverse direction the “improvement” on errors is larger, of 30%. The mean  $\chi^2$  of the same ten realizations reduces from 1545 to 1503 when using the true continuum instead of the fitted one. This gives an idea of how continuum modeling increases errors on recovered BAO parameters, even given the low number of realizations studied.

A check for our BAO fitting procedure consists in comparing input and recovered values for the Ly $\alpha$  forest bias parameter,  $b$ , and the redshift-space distortion parameter,  $\beta$ , which have input values of  $b = -0.14$  and  $\beta = 1.4$  in mocks. In practice, the fitted parameters are  $\beta$  and  $b(1 + \beta)$ , which set, respectively, the relative and absolute amplitude of transverse and radial correlation functions. Correct values of those parameters are difficult to recover since the smooth component of the correlation function is affected by distortions caused by the continuum fitting and sky residuals. Figure 7.16 shows the distribution of the best-fit  $\beta$  and  $b(1 + \beta)$  from 100

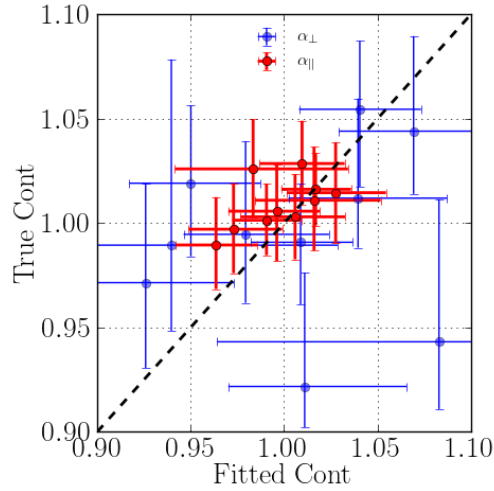


Figure 7.15: Comparison of the  $\alpha_{\parallel}, \alpha_{\perp}$  values obtained from 10 mock surveys with forests built with a fitted continuum or the true continuum.

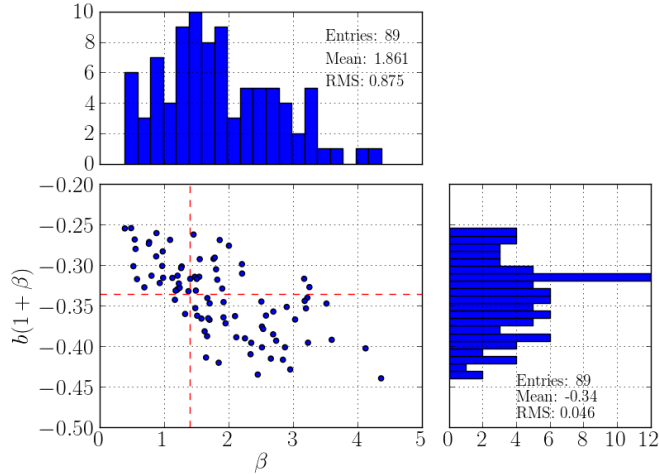


Figure 7.16: Distribution of fitted  $\beta$  and  $b(1 + \beta)$  for the 100 mock catalogs. There are 11 outliers outside the plot ranges, giving unphysical values for both parameters. The dashed red lines show the input values.

mock catalogs containing sky residuals and fitted continua. Excluding 11 outliers which have unphysical values, we obtain that the mean  $\beta$  is  $1.86 \pm 0.09$  which is in disagreement with the input value of 1.4. The mean  $b(1 + \beta)$  parameter,  $-0.340 \pm 0.005$ , is in  $1\sigma$  agreement with the input value of  $-0.336$ . The bias value could be derived from the combination of  $\beta$  and  $b(1 + \beta)$ , but since the fitted  $\beta$  values are typically not correct, the derived bias values would also have systematic errors included. In order to improve estimates of  $\beta$  and  $b$ , a better understanding of distortions is needed.

We conclude, based on the tests performed in our mock catalogs that the method for estimating the  $\text{Ly}\alpha$  forest correlation function and extracting the BAO peak position leads to unbiased results. We are able to correctly estimate the covariance matrix of the correlation function as shown by the distributions of the dilation parameters,  $\alpha_{\parallel}$  and  $\alpha_{\perp}$ , and their respective errors. These estimates were obtained from 100 mock realizations analyzed with our standard methods.

---

Of course, this conclusion is considering the systematic effects added in our mock catalogs discussed up to now. Real data contains other systematic effects, such as absorption by high column density systems, metal absorption, and instrumental effects on spectra such as the presence of calibration residuals. In the following two chapters we characterize the effect of these systematic effects on the correlation function, using data and improved mock catalogs.



## Chapter 8

# Analysis of Real Data

In this chapter, we finally present the Ly $\alpha$  BAO measurement using real data. Results of each step of the analysis are discussed and compared with mock measurements. The data sample is presented in § 8.1, the mean transmission and variances estimates in § 8.2; the correlation function is shown in § 8.3 and a discussion about the covariance matrix in § 8.4. In § 8.5, measurements of the BAO peak position using a variety of data sets and different analysis assumptions are exposed. We conclude that our measurement is robust against those changes in the analysis. The standard analysis gives  $\alpha_{\parallel} = 1.054^{+0.032}_{-0.031}$  and  $\alpha_{\perp} = 0.973^{+0.056}_{-0.051}$ , a result consistent with the fiducial model within  $2\sigma$ .

### 8.1 Our sample: Data Release 11

The Data Release 11 of BOSS was used for the analysis described in this chapter. As described in § 3.10, the visual inspection of DR11 spectra associated to quasar targets confirmed 158,401 quasars with redshifts  $z$  in the useful range for Ly $\alpha$  forest observations:  $2.1 < z < 3.5$ .

The visual inspection could identify quasars containing broad absorption lines (BALs) that were not used in our analysis, leaving 140,579 quasars, even though it is possible to use a fraction of their forest as in Slosar et al. (2013). An additional cut requiring a minimum number of good quality forest pixels (50 “analysis pixels” as explained in § 6) yields 137,562 quasars. Those numbers are summarized in Table 8.1.

Damped Ly $\alpha$  systems (DLAs) were also flagged by the visual inspection, but they were also identified and characterized by an automatic procedure detailed in Noterdaeme et al. (2012). Forests containing DLAs were not excluded from the analysis. Instead, using their measured redshift and column density we fitted forest continua adding Voigt profiles over DLAs. Forest pixels where the absorption from the DLA is less than 20% were preserved. Since DLAs are strong absorption systems, metal lines associated to these DLAs produce strong absorption lines in the forests. Those lines were masked over widths of 0.2 or 0.3nm depending on the strength of the line and 0.4 for Ly $\beta$ .



Table 8.1: Summary of the DR11 Ly $\alpha$  forest sample.

Set	Number
All QSOs	239,148
$2.1 < z < 3.5$	158,401
Flagged with BALs	17,822
Flagged with DLAs	17,518
Flagged with both	1,046
Low number of analysis pixels	3,017
Final Ly $\alpha$ forest sample	137,562

Forests were rebinned into analysis pixels corresponding to 3 BOSS pixels, yielding pixels of  $\Delta \log \lambda = 3 \times 10^{-4}$  which corresponds to  $\Delta v = c\Delta\lambda/\lambda = 207 \text{ km s}^{-1} \sim 0.2 \text{ nm}$  or  $\sim 2 h^{-1} \text{ Mpc}$  at  $z = 2.3$ . The total sample of 137,562 forests provides  $2.4 \times 10^7$  transmission measurements over a volume of  $\sim 50 h^{-3} \text{ Gpc}^3$ , the largest volume ever used for clustering studies.

All forest had their continua fitted using the three methods presented in § 6.1. As in the previous chapter, the discussion is focused on method C2, the method for which I contributed in this thesis. Final results with other continuum fitting methods are compared when fitting for the BAO peak in § 8.5 as a systematic check.

## 8.2 Mean transmission and weights

Figure 8.1 shows the obtained  $\bar{F}(\lambda)$ , compared with the same measurement performed on mock catalogs. We observe the expected evolution on redshift ( $z = \lambda/\lambda_{\text{Ly}\alpha} - 1$ ). We also observe some data reduction features, such as the galactic calcium absorption doublet at 393 and 397 nm, and the Balmer residuals due to flux calibration errors near 398, 410 and 435 nm, corresponding to Balmer transitions of hydrogen. These Balmer bumps are present in each forests and could in principle introduce systematic errors in the correlation measurements. The effect of these Balmer lines on the correlation function is further studied in § 9.1. At wavelengths  $\lambda < 400 \text{ nm}$ , fluctuations due sky subtraction residuals are seen both in data and mocks.

Once continua are fitted and the mean transmission estimated, we estimated the optimal pixel weights for the correlation function measurements (see § 6.2.1). Figure 8.2 shows the estimated intrinsic variance,  $\sigma_{\text{LSS}}^2$ , and the noise correction term,  $\eta(z)$ , for DR11 forests from C2 method. Intrinsic variances of real data have a redshift evolution similar to mocks. The noise correction of real spectra is 20% smaller than the correction for the mocks. However, as shown in Fig. 8.3 and previously discussed in § 7.4, the fit of our pixel variance model is not very accurate, but correct to first order. Fit residuals for both mock and data show the same dependence on  $\sigma_{\text{pipeline}}^2$ .

## 8.3 The correlation function

Following the same procedure as for mock catalogs in § 7.5 we computed the correlation function for DR11 Ly $\alpha$  forests. The result is shown in Fig. 8.4, represented by averages over the same

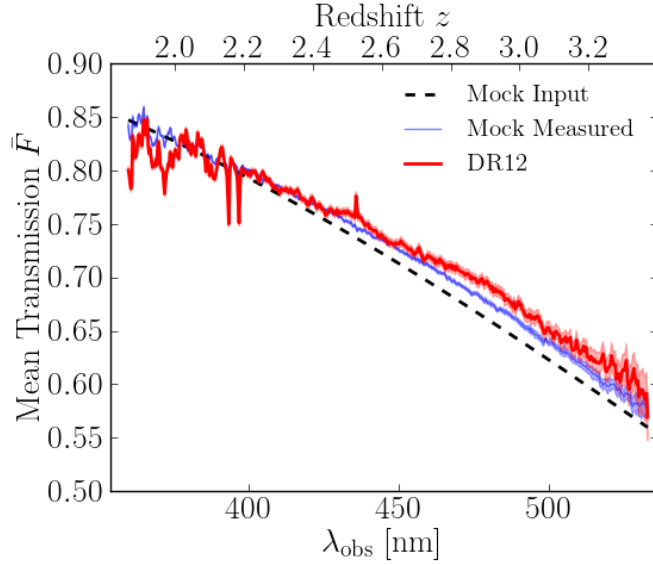


Figure 8.1: Mean transmission  $\bar{F}$  as function of observed wavelength estimated using continuum fitting method C2 over DR11 forests (red). As a reference, the mock measurement (blue) is also shown with its input values. The bumps observed at 398, 410 and 435 nm (Balmer transitions of hydrogen) are spectro-calibration errors in data.

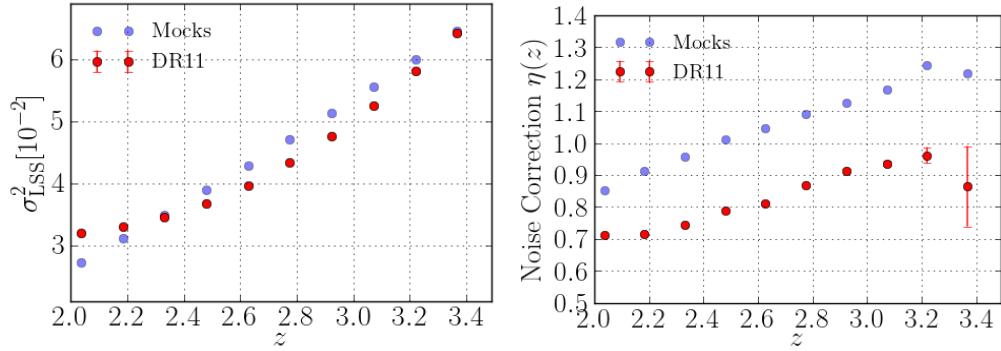


Figure 8.2: The estimated intrinsic variance of the forests and the noise correction term as function of redshift for DR11 forests (red points) compared to mock estimates (blue points).

three  $\mu$  intervals as for mock catalogs. The error bars come from the diagonal of the covariance matrix, summed over the same  $\mu$  ranges. The BAO peak is seen in the three panels, especially in  $\xi_{\parallel}(r)$ . Comparing with DR9 measurements of Busca et al. (2013) where the peak was only visible in the radial direction, the error bars were reduced by a factor  $\sim 3$ , which agrees with expectations given that DR11 has three times more data.

The superimposed curves in Fig 8.4 are two fitted models. The first obtained when letting free the peak position dilation parameters,  $\alpha_{\parallel}$  and  $\alpha_{\perp}$ , and the second is the same but fixing  $\alpha_{\parallel} = \alpha_{\perp} = 1$ . Fit results are discussed in § 8.5, after the discussion about the covariance matrix estimate.

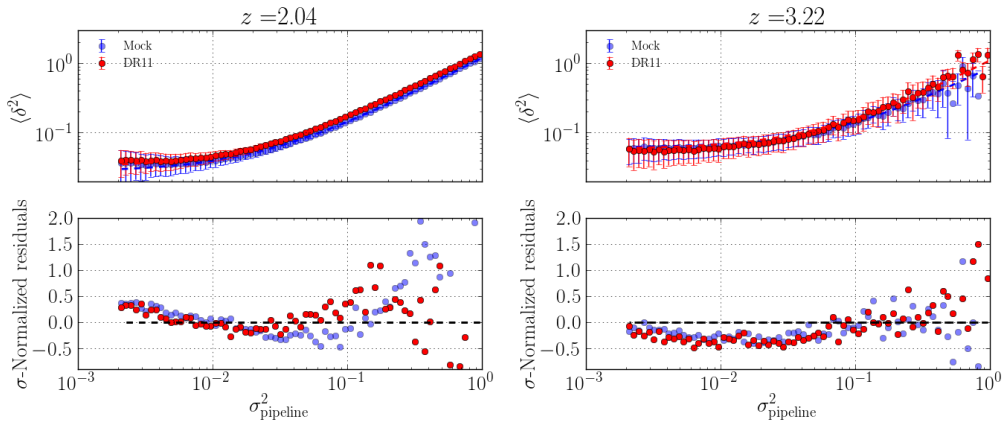


Figure 8.3: Detail of the pixel variance model fit for two redshift bins and their relative residuals for DR11 forests (red points) and for a mock realization (blue points).

## 8.4 The covariance matrix

The covariance matrix  $C(A, B)$  of the correlation function  $\xi(A)$  was computed using the two methods presented in § 6.3: sub-sampling and Wick approximation. We smoothed these covariance matrices by assuming that correlation coefficients (Eq. 6.17) depend only on  $\Delta r_{\parallel} = r_{\parallel} - r_{\parallel}'$  and  $\Delta r_{\perp} = r_{\perp} - r_{\perp}'$ . We have shown in § 7.6 that this assumption is valid. For the sub-sampling method, 2044 plates were used as sub-samples of the full survey.

The full  $2500 \times 2500$  element covariance matrix has a relatively simple structure. As for mocks, the diagonal terms are, in a good approximation, inversely proportional to the number of pairs and independent of separation, as shown in Fig. 8.5,

$$C(A, A) \sim \frac{0.041}{N_{\text{pairs}}(A)}. \quad (8.1)$$

This value is twice the value that would have been obtained if considering all pixels independent. It is larger because of the correlation between neighboring pixels of the same quasar. It means that the measurement of  $\xi(r_{\parallel}, r_{\perp})$  using a pair of pixels in two different quasars is not independent of the measurement of  $\xi(r_{\parallel}, r_{\perp})$  using a different pair of pixels in the same two quasars.

Figure 8.5 also shows that diagonal terms of the mock covariance matrix are 30% smaller than for data. The reason is that mock fluctuations  $\delta_F$  have lower variance than data fluctuations as shown in Fig. 8.6. Taking the square of these variances, which is approximately the T1 terms of the covariance matrix, gives roughly the same 30% factor between data and mocks. This difference have two possible origins: the presence of undetected high column density systems in data, or a bad correction of the pixel noise estimate.

Figure 8.7 shows averaged correlation coefficients of the non-diagonal terms of the covariance matrix, as functions of  $\Delta r_{\parallel}$  for the first  $\Delta r_{\perp}$  bins, computed for both sub-sampling and Wick expansion. The first remarkable observation is that both estimates agree for almost the full range of separations. This validates the use of the sub-sampling method for computing the covariance matrices for cosmological fits. Indeed, fitting using both covariances yields no significant difference in final results.

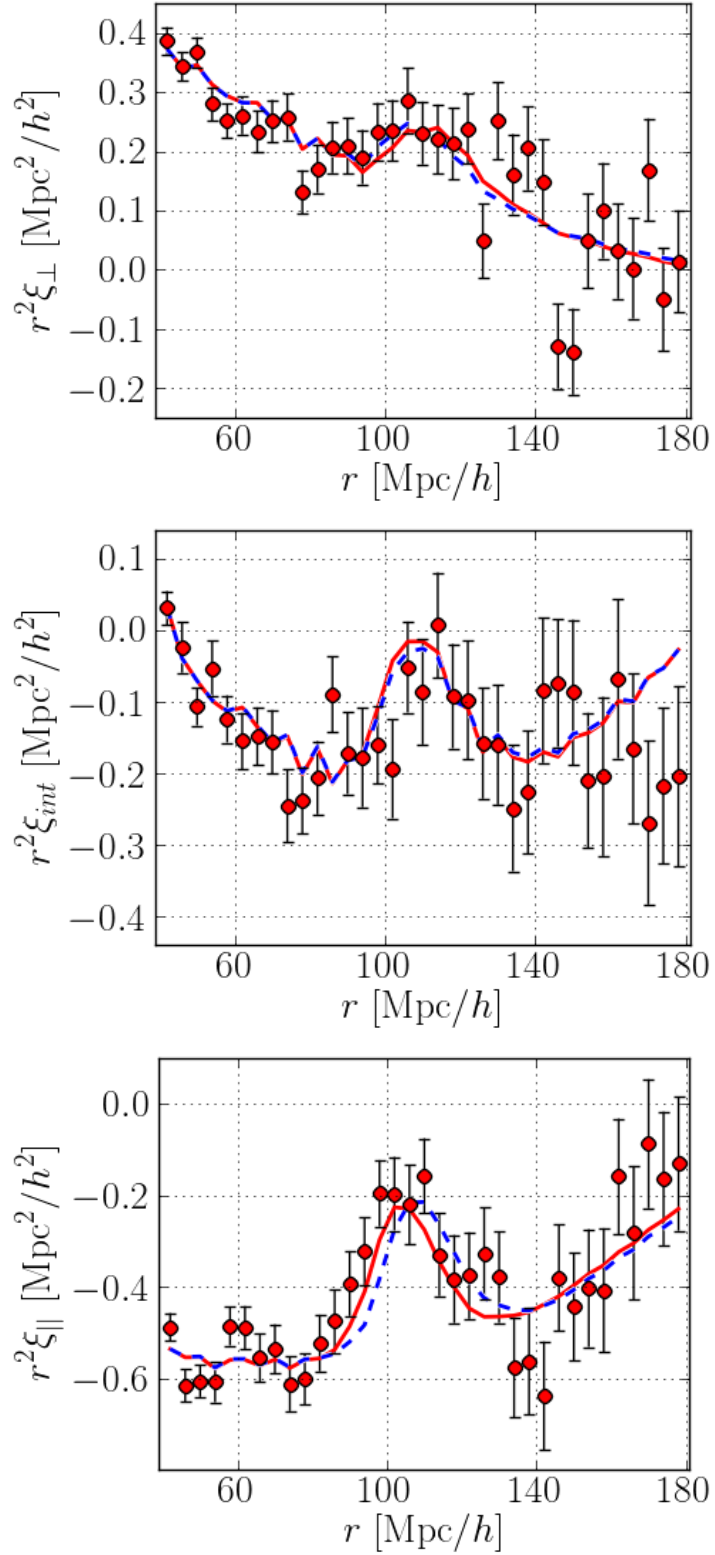


Figure 8.4: The measured Ly $\alpha$  forest correlation functions (continuum C2) in three angular regions:  $\mu < 0.5$  (top),  $0.8 > \mu > 0.5$  (middle), and  $\mu > 0.8$  (bottom), where  $\mu$  is the central value of  $r_{\parallel}/\sqrt{r_{\parallel}^2 + r_{\perp}^2}$  in each  $(r_{\parallel}, r_{\perp})$  bin. The curves show the results of fits as described in § 6.4. The full curve is best fit and the dashed curve is best fit when the parameters  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  are both set to unity. The irregularities in the fits are due to the use of  $(r_{\parallel}, r_{\perp})$  bins rather than  $(r, \mu)$  bins.

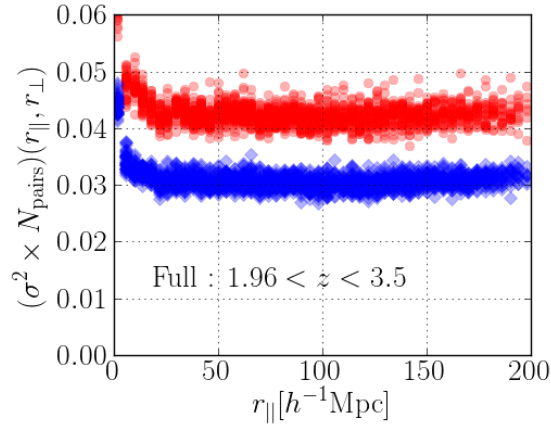


Figure 8.5: Variance of the correlation function times the number of pairs in each bin for data (red points) and for a mock realization (blue points) as functions of  $r_{\parallel}$ . This quantity is almost scale-invariant, except for small values of  $r_{\parallel}$ .

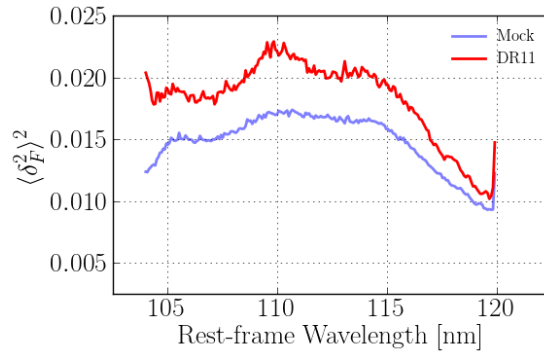


Figure 8.6: Variance squared of the transmission fluctuations  $\delta_F$  as a function of rest-frame wavelength, computed for data forests and for a mock realization.

The structure of the covariance is simplified by the fact that we are close to the situation where lines of sight are parallel, so correlations of type T2 and T3 dominate the covariance (the T1 term contributes only to the diagonal terms). The reasons for this structure is made clear by the Wick expansion in § 6.3 relating the covariance to correlations within pairs-of-pairs of pixels. To the extent that two neighboring lines of sight are parallel, these terms contribute only to the elements of the covariance matrix where  $r_{\perp} = r'_{\perp}$  (or  $\Delta r_{\perp} = 0$ ). These elements are shown as a function of  $\Delta r_{\parallel}$  in Fig. 8.7 (top left). They follow closely the line of sight correlation function found within individual forests. The Ly $\alpha$ -SiII correlation peak is clearly seen at  $\Delta r_{\parallel} \sim 20 h^{-1} \text{Mpc}$ .

For the terms of the covariance matrix where  $r_{\perp} \neq r'_{\perp}$  (or  $\Delta r_{\perp} > 0$ ), only pairs-of-pairs of pixels of three or four quasars can contribute (T4, T5 and T6 in Fig. 6.3). As explained previously, these have sub-dominant contributions for the covariance. As shown in Fig. 8.7 (bottom panels) these terms fall to zero very quickly when increasing  $\Delta r_{\perp}$  or  $\Delta r_{\parallel}$ .

The statistical accuracy of the covariance matrix estimate with the sub-sampling method is  $\sim 0.02$ , number related to the number of both sub-samples and bins over which correlation coefficients (Eq. 6.17) are averaged over in the smoothing process. As described in § 6.3.3, all

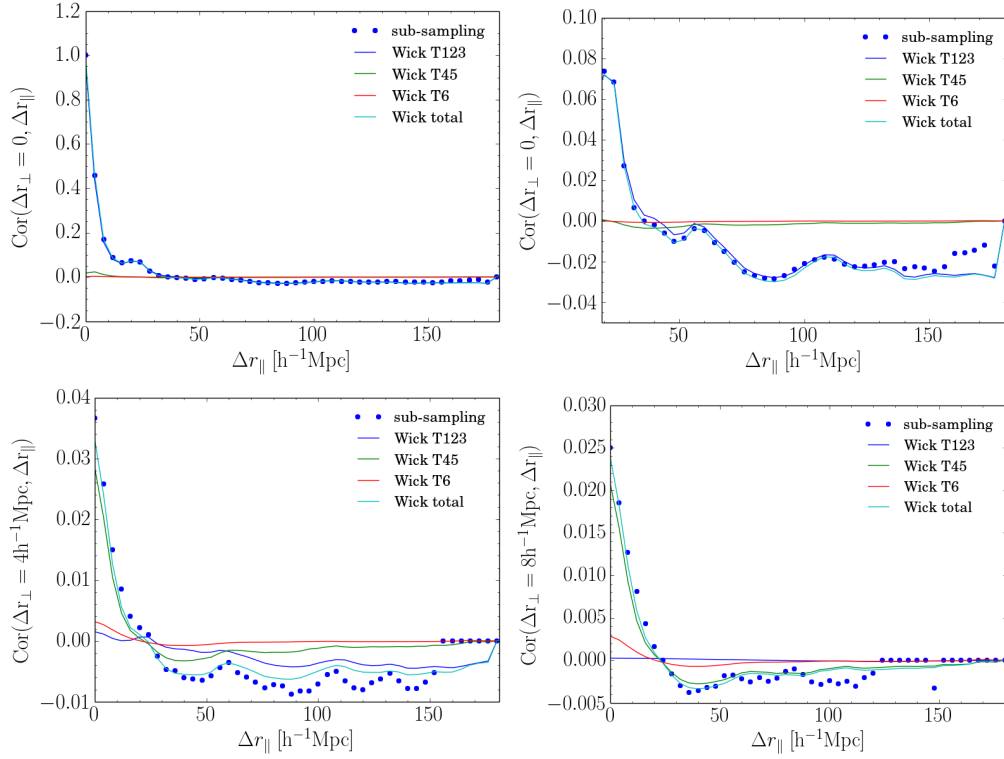


Figure 8.7: The correlation  $C(r_{\perp}, r_{\parallel}, r_{\perp}', r_{\parallel}') / [\text{Var}(r_{\perp}, r_{\parallel}) \text{Var}(r_{\perp}', r_{\parallel}')]^{1/2}$  as a function of  $r_{\parallel} - r_{\parallel}'$  (averaged over  $(r_{\perp}', r_{\parallel}')$ ). The top figures are for  $r_{\perp} - r_{\perp}' = 0$  over the full range of  $\Delta r_{\parallel}$  (left) and for  $\Delta r_{\parallel} > 20 h^{-1} \text{Mpc}$  (right). The bottom two figures are for  $r_{\perp} - r_{\perp}' = 4 h^{-1} \text{Mpc}$  (left) and for  $r_{\perp} - r_{\perp}' = 8 h^{-1} \text{Mpc}$  (right). Shown are the correlations determined by sub-sampling and by a Wick expansion. The latter correlations are decomposed into the pair-of-pair types, T1-T6, from Fig. 6.3.

elements for which the averaged correlation coefficient is consistent with no correlation are set to zero.

In Fig. 8.8 we compare the averaged correlation coefficients from data and from a mock realization, both computed using the sub-sampling method. Since the peak of Ly $\alpha$ -SiII is clearly seen in the covariance, for a better comparison we also compute the covariance matrix for a mock realization containing metal absorption in the forests. We see in Fig. 8.8 that the Ly $\alpha$ -SiII peak in the covariance is reasonably well reproduced in mocks. Correlations look similar but show slightly different dependence with  $\Delta r_{\parallel}$  and  $\Delta r_{\perp}$ . For example, at  $\Delta r_{\perp} = 0$ , the data correlations are larger than for mocks by at most 20% in the  $0 < \Delta r_{\parallel} < 30 h^{-1} \text{Mpc}$  range, whereas for the next transverse separation bin ( $\Delta r_{\perp} = 4 h^{-1} \text{Mpc}$ ) the mock correlations are larger by 30% in the same range.

The differences in covariances between mock and data can be understood as coming from differences in the small scale clustering. Following the arguments of the Wick expansion of the covariance matrix, small scale correlations are dominant for the total covariance (T2 and T3 terms). However, the input model for the mock correlation function was built for the large scale measurement of the BAO peak, while the small scales were not precisely modeled (e.g., using hydro-simulations).

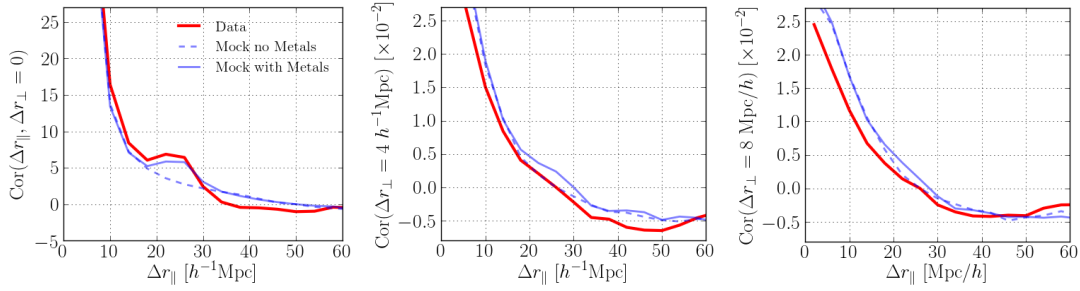


Figure 8.8: Correlation coefficient  $C(r_{\perp}, r'_{\perp}, r_{\parallel}, r'_{\parallel}) / \sqrt{\text{Var}(r_{\perp}, r_{\parallel})\text{Var}(r'_{\perp}, r'_{\parallel})}$  of the 3D correlation function as a function of parallel separation difference  $\Delta r_{\parallel} = r_{\parallel} - r'_{\parallel}$ . The values shown are the average of all correlation matrix elements with the same  $\Delta r_{\parallel}$  and  $\Delta r_{\perp}$ , for  $\Delta r_{\perp} = 0$  (left),  $4 h^{-1}\text{Mpc}$  (center) and  $8 h^{-1}\text{Mpc}$  (right panel). Data is shown in red solid lines, mock measurements in blue, with (solid) and without metals (dashed).

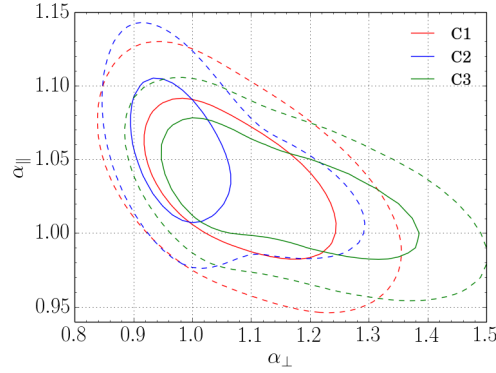


Figure 8.9: Constraints on  $(\alpha_{\parallel}, \alpha_{\perp})$  using the three continuum estimators, C1 (red), C2 (blue) and C3 (green). The solid and dashed contours correspond to  $1\sigma$  and  $2\sigma$  ( $\Delta\chi^2 = 2.3, 6.2$ ).

## 8.5 BAO peak position measurement

Once the correlation function and its covariance matrix were estimated, and since the BAO peak was clearly visible near the expected position, we could estimate its position by fitting the correlation function model over the measurement, as described in § 6.4. The fitting procedure was already tested in § 7.7 over mock catalogs, where we showed that we are able to recover unbiased results with correct estimation of errors.

Table 8.2 shows the results of fits over data for a variety of data sets and analysis assumptions. The first line lists our standard analysis using C2 continua,

$$\alpha_{\parallel} = 1.054^{+0.032}_{-0.031} \quad \text{and} \quad \alpha_{\perp} = 0.973^{+0.056}_{-0.051}, \quad (8.2)$$

for which the fitted model is shown by the red line in Fig. 8.4. The precision on  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  inferred from our fitting procedure are in agreement with the distribution of errors on mock catalogs shown in Fig. 7.13.

The  $1\sigma$  and  $2\sigma$   $\chi^2$  contours in the  $(\alpha_{\parallel}, \alpha_{\perp})$  plane are shown in Fig. 8.9 for the three continuum

Table 8.2: Results for the standard fit and modified fits. The standard fit uses the C2 continuum, a broadband defined by  $(i_{\max}, j_{\max}) = (2, 2)$ , a forest defined by  $104.0 < \lambda_{\text{rf}} < 120.0$  nm, and  $a_{\text{peak}} = 1$ .

analysis	$\alpha_{\parallel}$	$\alpha_{\perp}$	$\beta$	$b(1 + \beta)$	$\chi_{\min}^2/DOF$
standard(C2)	$1.054^{+0.032}_{-0.031}(1\sigma)^{+0.069}_{-0.063}(2\sigma)$	$0.973^{+0.056}_{-0.051}(1\sigma)^{+0.199}_{-0.103}(2\sigma)$	$1.47 \pm 0.87$	$-0.39 \pm 0.05$	1499.0/(1515-13)
C1	$1.038^{+0.035}_{-0.037}(1\sigma)^{+0.073}_{-0.074}(2\sigma)$	$1.054^{+0.132}_{-0.093}(1\sigma)^{+0.246}_{-0.176}(2\sigma)$	$3.47 \pm 2.78$	$-0.42 \pm 0.05$	1571.5/(1515-13)
C3	$1.038^{+0.026}_{-0.039}(1\sigma)^{+0.054}_{-0.071}(2\sigma)$	$1.041^{+0.259}_{-0.063}(1\sigma)^{+0.259}_{-0.126}(2\sigma)$	$2.28 \pm 1.24$	$-0.46 \pm 0.05$	1603.7/(1515-13)
$\beta$ -prior ( $1.4 \pm 0.4$ )	$1.055^{+0.032}_{-0.031}(1\sigma)^{+0.068}_{-0.063}(2\sigma)$	$0.972^{+0.053}_{-0.051}(1\sigma)^{+0.117}_{-0.102}(2\sigma)$	$1.41 \pm 0.36$	$-0.39 \pm 0.04$	1499.1/(1515-13)
$a_{\text{peak}}$ free	$1.054^{+0.035}_{-0.031}(1\sigma)^{+0.078}_{-0.063}(2\sigma)$	$0.973^{+0.057}_{-0.052}(1\sigma)^{+0.232}_{-0.104}(2\sigma)$	$1.49 \pm 1.08$	$-0.38 \pm 0.24$	1499.0/(1515-14)
$\Sigma_{\perp} = \Sigma_{\parallel} = 0$	$1.053^{+0.029}_{-0.028}(1\sigma)^{+0.062}_{-0.059}(2\sigma)$	$0.961^{+0.055}_{-0.052}(1\sigma)^{+0.254}_{-0.103}(2\sigma)$	$1.31 \pm 0.87$	$-0.34 \pm 0.05$	1501.2/(1515-13)
$\Sigma_{\perp}, \Sigma_{\parallel}$ free	$1.063^{+0.041}_{-0.036}(1\sigma)^{+0.101}_{-0.073}(2\sigma)$	$0.976^{+0.053}_{-0.05}(1\sigma)^{+0.124}_{-0.102}(2\sigma)$	$1.56 \pm 0.80$	$-0.46 \pm 0.07$	1497.2/(1515-15)
no special DLA treatment	$1.049^{+0.038}_{-0.034}(1\sigma)^{+0.089}_{-0.068}(2\sigma)$	$0.954^{+0.053}_{-0.049}(1\sigma)^{+0.132}_{-0.096}(2\sigma)$	$0.36 \pm 0.46$	$-0.33 \pm 0.06$	1489.7/(1515-13)
$104.5 < \lambda_{\text{rf}} < 118.0$ nm	$1.052^{+0.041}_{-0.041}(1\sigma)^{+0.145}_{-0.094}(2\sigma)$	unconstrained	$2.37 \pm 2.81$	$-0.34 \pm 0.07$	1448.2/(1515-13)
No spectra with DLAs	$1.031^{+0.035}_{-0.035}(1\sigma)^{+0.074}_{-0.074}(2\sigma)$	$1.073^{+0.117}_{-0.082}(1\sigma)^{+0.228}_{-0.171}(2\sigma)$	$2.38 \pm 1.91$	$-0.43 \pm 0.06$	1506.5/(1515-13)
$z < 2.295$	$0.996^{+0.052}_{-0.054}(1\sigma)^{+0.113}_{-0.134}(2\sigma)$	$0.89^{+0.064}_{-0.053}(1\sigma)^{+0.148}_{-0.108}(2\sigma)$	$1.10 \pm 0.92$	$-0.31 \pm 0.06$	1523.0/(1515-13)
$z > 2.295$	$1.096^{+0.037}_{-0.036}(1\sigma)^{+0.079}_{-0.073}(2\sigma)$	$0.994^{+0.057}_{-0.049}(1\sigma)^{+0.155}_{-0.1}(2\sigma)$	$1.61 \pm 1.05$	$-0.48 \pm 0.06$	1479.1/(1515-13)
$a_{\text{peak}} = 0$	-	-	-	-	1526.2/(1515-11)

fitting methods. They show that errors on those parameters are non-Gaussian and the anti-correlation between them, consistent with the observed anti-correlation for mocks in Fig. 7.14. The errors on  $\alpha_{\perp}$  extend asymmetrically to large values, as expected from the visual impression in Fig. 8.4.

The next seven lines of Table 8.2 present the results using the same data, but with different choices of analysis procedures : use of non-standard continuum fitting C1 and C3; adding a Gaussian prior to the redshift distortion parameter  $\beta$  around the nominal value of 1.4 with a width of 0.4; freeing the peak amplitude  $a_{\text{peak}}$ ; fitting the non-linearity parameters,  $\Sigma_{\parallel}$  and  $\Sigma_{\perp}$ , or setting them to zero (and thus not correcting for non-linear effects); using spectra containing DLAs but using their pixels as if they were not a systematic (no fit of damping wings). Since all those fits use the same data, significant differences in results would indicate a systematic effect caused by our analysis chain or fitting procedure. All those seven sets give  $(\alpha_{\parallel}, \alpha_{\perp})$  values with differences smaller than  $1\sigma$ . However the only noticeable effect is seen when using non-standard continuum fitting methods C1 and C3, for which errors in  $\alpha_{\perp}$  are larger. This is somehow expected because of the low significance of the transverse BAO peak in our measurements. The errors on  $\alpha_{\parallel}$  are robust with respect to changes in the continuum fitting method.

The next two lines in Table 8.2 show results obtained with the method C2 on forests with reduced wavelength range,  $104.5 < \lambda_{\text{rf}} < 118.0$  nm (the same range used in the DR9 analysis (Busca et al., 2013)); and discarding forests containing detected DLAs. Both results are consistent with the standard ones within  $1\sigma$ . The error bars are increased by  $\sim 50$  and 13% since those data sets represent smaller amounts of data.

The next two lines present results obtained when dividing the full sample into two redshift bins. The redshift of a pixel pair is defined by the average of the redshift of both pixels. We have chosen to split the sample at  $z = 2.295$  that represents the median of the pixel pair redshift distribution. Both samples, low and high redshift, are fairly independent and results agree at the  $2\sigma$  level. The difference between the low and high  $z$  values for  $(\alpha_{\parallel}, \alpha_{\perp})$  is consistent



Table 8.3: Fit results with the C2 continuum with a modified fitting range [standard:  $40 < r < 180 \text{ h}^{-1}\text{Mpc}$ ] and number of terms in the broadband (Eq. 6.29) [standard:  $(i_{\max}, j_{\max}) = (2, 2)$ ].

	$\alpha_{\parallel}$	$\alpha_{\perp}$	$\chi^2_{\min}/DOF$
standard (C2)	$1.054 \pm 0.032$	$0.973 \pm 0.055$	1499.0/(1515-13)
range ( $h^{-1}\text{Mpc}$ )			
$60 < r < 180$	$1.045 \pm 0.032$	$0.986 \pm 0.063$	1391.8/(1415-13)
$40 < r < 160$	$1.052 \pm 0.033$	$0.974 \pm 0.053$	1139.2/(1177-13)
$(i_{\max}, j_{\max})$			
(2,3)	$1.057 \pm 0.032$	$0.970 \pm 0.050$	1484.2/(1515-16)
(3,2)	$1.050 \pm 0.033$	$0.987 \pm 0.067$	1497.8/(1515-16)
(3,3)	$1.051 \pm 0.034$	$0.986 \pm 0.068$	1479.2/(1515-20)

with the differences observed with mock catalogs, shown in the left panel of Fig. 8.10. Those measurements have naturally larger error bars due to the lower statistics of those samples. The right panel of Fig. 8.10 shows the correlations between low and high  $z$  measurements of  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  using mock catalogs. The data point is also shown for comparison. The values of  $\alpha_{\perp}$  are only 2% correlated while the values of  $\alpha_{\parallel}$  have a correlation of 16.6%. This higher correlation for  $\alpha_{\parallel}$  is certainly due to the larger number of pixels near  $z = 2.295$  contributing to correlations in both redshift bins.

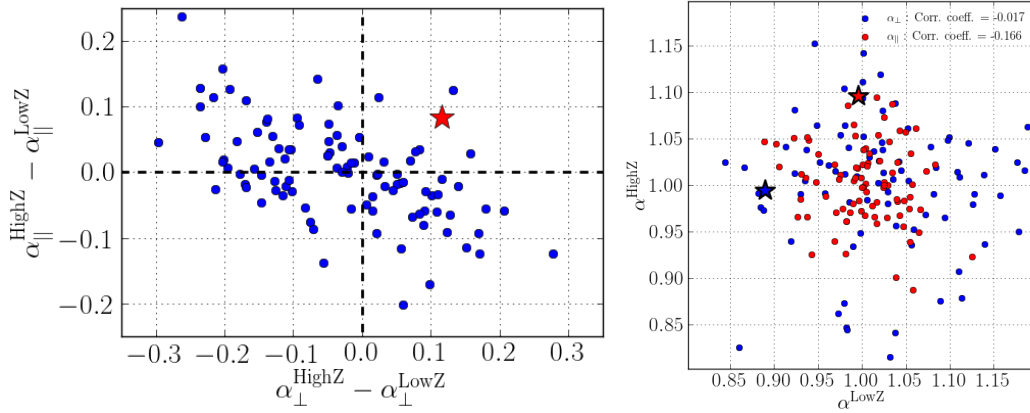


Figure 8.10: Left: Difference in best-fit  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  values between high redshift ( $z > 2.295$ ) and low redshift ( $z < 2.295$ ) subsets of the 100 mock realizations and the result for data (red star). Compared to Fig 7.14, the plot shows the degraded precision resulting from the division of the data into two redshift bins. Right: comparison between low and high redshift values for the same parameters showing correlations between both measurements in both redshift bins.

The last line of Table 8.2 shows the fit result when assuming that there is no BAO peak in the correlation function, fixing the amplitude of the peak to zero. The  $\chi^2$  difference with respect to the first line reveals  $\Delta\chi^2 = 27.2$  for two extra degrees of freedom, corresponding to a  $5\sigma$  detection.

Table 8.3 shows results when changing the separation range and the broadband model (Eq. 6.29) in the fits of the correlation function. The broadband model is included to take into account systematic effects affecting the broadband of the correlation function, e.g., the distortion from

continuum fitting or smooth correlated calibration errors. Results show no significant change on the derived  $(\alpha_{\parallel}, \alpha_{\perp})$  parameters, indicating that the broadband is performing as required. Of particular significance, adding greater freedom to the broadband function has only a  $\sim 10\%$  impact on the size of the  $\alpha_{\parallel}$  error, although it has a larger impact of about 20-30% on the  $\alpha_{\perp}$  error.

The uncertainties on  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  reported in this section are derived from  $\chi^2$  surfaces and are purely statistical. Physical and instrumental effects could add systematic errors in our measurements of the BAO peak position. In the next chapter we describe and characterize the impact of those potential systematics. No strong evidence for a systematic was found in the tests we performed to this point.

We conclude that our measurement of the DR11 Ly $\alpha$  forest correlation function is robust given our methodology and the possible choices of cuts on data. The BAO peak is clearly seen in the correlation function shown in Fig. 8.4. The covariance matrix of our measurement was computed with two different methods yielding same results. Our standard analysis gives a BAO peak position consistent with the fiducial model within  $2\sigma$  on the radial direction and within  $1\sigma$  in the transverse direction. The cosmological interpretation of those results is discussed in Chapter 10.



## Chapter 9

# Study of Systematic Effects

In this chapter we explore the possible systematic effects that could introduce a change in the estimated position of the BAO peak, by the introduction of spurious non-smooth correlations in our estimation of  $\xi(r_{\parallel}, r_{\perp})$  described in the previous chapters. In § 9.1 we study the effect of calibration errors mistakenly introduced in all BOSS spectra by pipeline at wavelengths corresponding to Balmer transitions of the hydrogen. In § 9.2 we characterize the effect of metal absorption that is naturally correlated with Ly $\alpha$  absorption. We conclude that all those systematics indeed include spurious correlations but too small to be significant for BAO measurements.

### 9.1 Balmer features

Going back to our measurement of the mean transmitted flux fraction  $\bar{F}(z)$  (Fig. 8.1), some bumps, not present in mocks, are clearly visible. The wavelengths of these bumps correspond to the Balmer transitions of the hydrogen atom. The most visible ones correspond to H $\delta$  (410 nm) and H $\gamma$  (434 nm).

The most likely explanation for those features is problems on estimation of the flux calibration vectors used by the pipeline described in § 3.10 to convert CCD electron counts into flux for each spectrum. When performing observations, for each spectrograph some fibers of the plate are pointed to standard F-stars (bottom left panel in Fig. 3.9) with known spectral shape, that are then used to compute these flux calibration vectors. A spline function is fitted over each standard spectrum (in electron units) of which the flux is known. Calibration vectors are defined as the ratio between the known flux and the fitted number of electrons. For a given plate exposure, two calibration vectors are computed for the blue side and two for the red side.

The left panel of Figure 9.1 shows one example of such calibration vector for the blue spectrograph. The broadband shape translates the throughput of the spectrograph (Fig. 3.6). We smooth this broadband shape using a running average over 200 wavelength pixels, building a smoothed calibration vector. Dividing the original vector by its smoothed version allows one to see small fluctuations. These fluctuations are shown in the right panel of Fig. 9.1 where at

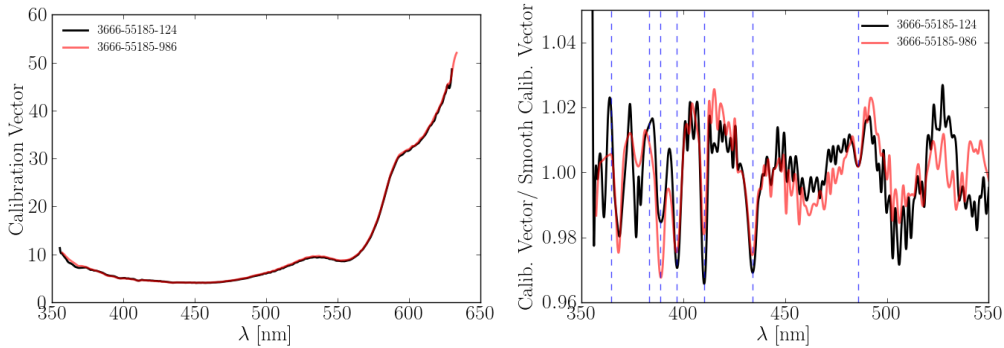


Figure 9.1: Left: examples of flux calibration vectors for one blue exposure and two fibers on distinct half-plates. All calibration vectors of a given half-plate are the same. The half-plates are defined by the numbers of fibers inside the ranges: 1-500 and 501-1000. Right: same calibration vectors normalized by their smoothed versions. Blue dashed lines show the Balmer transition wavelengths.

least five Balmer features appear clearly at 389, 397, 410, 434 and 486 nm. These features come from bad modeling of the fitted spline over the Balmer emission present in the standard spectra, leaving these residuals in the final calibration. Due to this problem, all BOSS spectra contain Balmer features included in their spectrum, including quasars and galaxies.

Those calibration errors at Balmer wavelengths might introduce additional correlations between flux fluctuations that would create non-smooth structure in the  $\text{Ly}\alpha$  correlation function. Those structures could bias our BAO peak position measurement. Therefore, we first checked the significance of these lines through the calibration vectors themselves. Second, we computed the correlation function of calibration vectors to see if they have important non-smooth correlations.

For a given half-plate (500 fibers), the calibration vector  $C(\lambda)$  is the same. Figure 9.2 shows the mean and the scatter around the mean of fluctuations  $\Delta C(\lambda)/\bar{C}(\lambda)$  of the calibration vectors around their broadband shape  $\bar{C}(\lambda)$  from a sub-sample of 896 half-plates. We see that the Balmer lines are significant with relative amplitudes up to  $0.02 \pm 0.004$ . If not corrected, these calibration errors would lead to additional  $\delta_F \sim 0.02$  over Balmer wavelengths. The subtraction of the mean  $\bar{F}(z)$  when estimating  $\delta_F$  removes this effect in average but does not correct each spectrum individually. It is thus important to check whether those fluctuations create spurious spatial correlations.

We computed the correlation function of calibration vector fluctuations  $\delta_C = C(\lambda)/\bar{C} - 1$ . We placed these fluctuations in the same position as the  $\text{Ly}\alpha$  forests. Figure 9.3 shows results. We do observe additional correlations at  $r_{\parallel} \sim 64, 100$  and  $160 h^{-1}\text{Mpc}$  corresponding to the correlations between, respectively, the pairs (389, 397)nm, (397,410)nm and the mix of (389,410) and (410,434)nm. However, these additional correlations of about  $10^{-7}$  are small when compared to the noise level of  $\hat{\xi}(A)$  at those separations, of about  $10^{-5}$ . This implies no significant change in the  $\text{Ly}\alpha$  correlation function. Furthermore, removing from the analysis pixel pairs near (397,410)nm, near BAO peak separations, does not generate any measurable change in the BAO peak position.

We conclude that the Balmer residuals do not affect BAO measurements at our level of precision. Nevertheless, this work was important for the improvement of the automatic data reduction

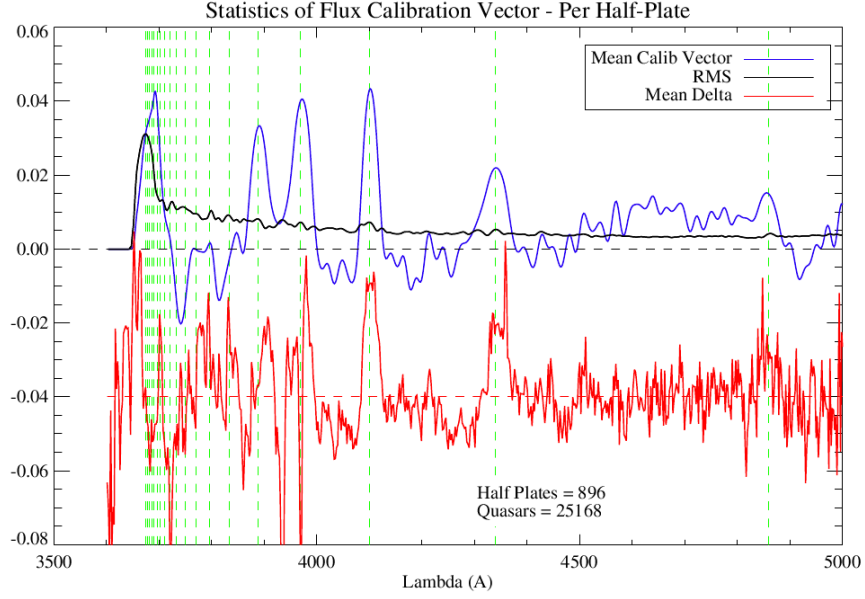


Figure 9.2: Average and RMS of calibration vectors with their broadband shape removed. Green dashed lines are the positions of the hydrogen Balmer transitions. Red line is the stacking of  $\delta_F$  from data.

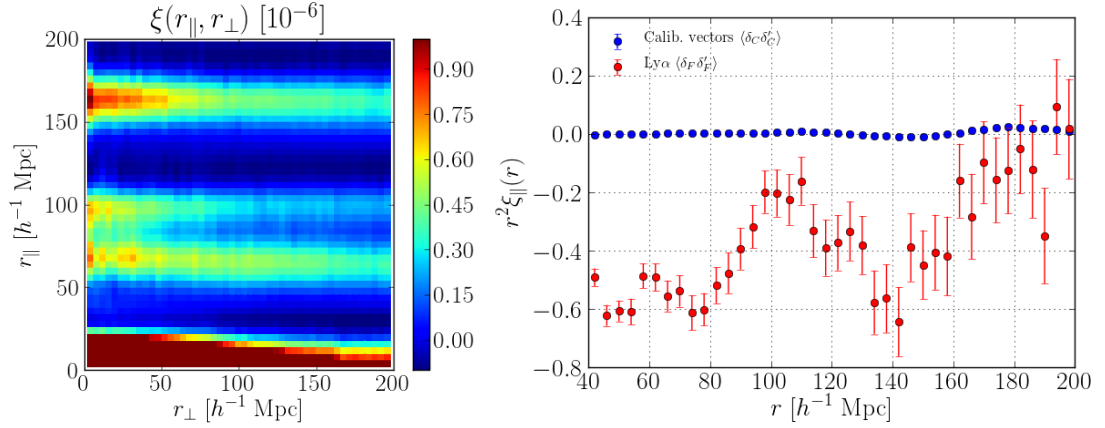


Figure 9.3: Left : Correlation function of calibration vector fluctuations as function of separation ( $r_{\parallel}, r_{\perp}$ ). Right: The same correlation averaged over  $0.8 < \mu < 1.0$ , compared with the DR11 Ly $\alpha$  correlation function.

pipeline, that took into account a correction for these errors in the new version of the calibration vectors.

The correction we proposed is simple: for each of the five significant Balmer lines, we make a linear interpolation of the between the values of the vector taken at  $\lambda_i \pm \Delta\lambda_i$ , where  $\Delta_i$  varies from 2.5 to 3.5 nm. We were able to fix the fluxes of individual spectra by multiplying them by the ratio of uncorrected and corrected calibration vectors. Since the calibration vectors are associated to individual exposures, for each quasar we performed the correction on individual exposures and we made our own coadded spectra using a inverse variance weighted sum. Figure 9.4 compares mean transmissions  $\bar{F}$  obtained from our coadded and corrected spectra with respect to the uncorrected ones. A clear improvement is seen on at least three of the five lines for which the

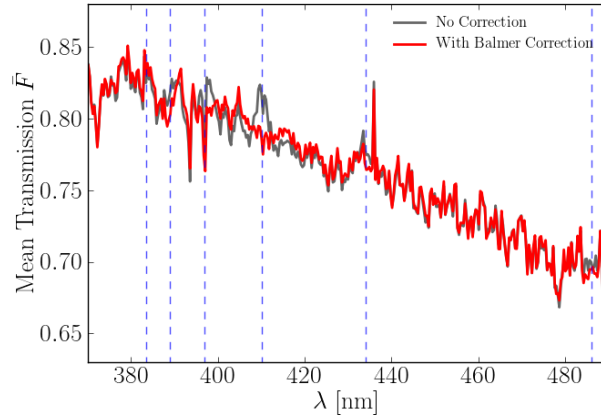


Figure 9.4: Mean Transmission of Ly $\alpha$  forest computed with a sub-sample of  $\sim 20,000$  quasars, for which we applied the correction for the Balmer artifacts.

correction was applied. This improvement was already implemented on the pipeline associated to Data Release 12.

## 9.2 Metal contamination

Another possible source of spurious correlation is metal absorption lines in the forest. As explained in § 2.1, metals in the IGM can absorb quasar light in the forest. Metal absorption lines are hard to identify on individual forests, but they are visible when stacking forests at the rest-frame of each absorption line having a transmission  $F$  lower than some threshold  $F_{\text{cut}}$ . Figure 9.5 shows a stacked spectrum of absorption lines with  $-0.05 < F < 0.05$  from Pieri et al. (2013), revealing different metal components of the IGM such as Si, N, O, C.

Since the metal and Ly $\alpha$  absorption are correlated, correlations as a functions of separation might appear in the Ly $\alpha$  correlation function at separations corresponding to the differences between wavelengths of Ly $\alpha$  and other metals. For example, the correlation between Ly $\alpha$  (121.6nm) and SiII (126nm) absorption correspond radial separations at  $z \sim 2.3$  of about  $r_{\parallel} = 110 h^{-1}\text{Mpc}$ . This excess peak-shaped correlation might introduce systematic errors in the BAO peak shape and position.

The effect of metals on the BAO measurements was characterized through mock catalogs. We included metals in ten DR11 mock catalogs following procedure described in § 5 and Pieri et al. (2013). Each realization was analyzed with the usual procedure, from the continuum fitting to the computation of the correlation function and the fit for the BAO peak position. The left panel of Fig. 9.6 shows correlation function differences between mocks with and without metals. We see that most of the effect of metals is concentrated on separations near to the line of sight ( $r_{\perp} < 30 h^{-1}\text{Mpc}$ ), showing smoothed peaks due to redshift space distortions. In the right panel, we show correlations averaged over  $\mu > 0.8$ ,  $\xi_{\parallel, \text{Met}}(r)$ , and the mean difference  $\Delta\xi_{\parallel}(r) = \xi_{\parallel, \text{Met}} - \xi_{\parallel}$  between results from the same realization with and without metals. We see that, except for the peaks at  $r \sim 20$  and  $175 h^{-1}\text{Mpc}$ , no peak-like feature appears in  $\xi_{\parallel}(r)$  that would change the BAO peak shape or position. At this precision level, metals act only as

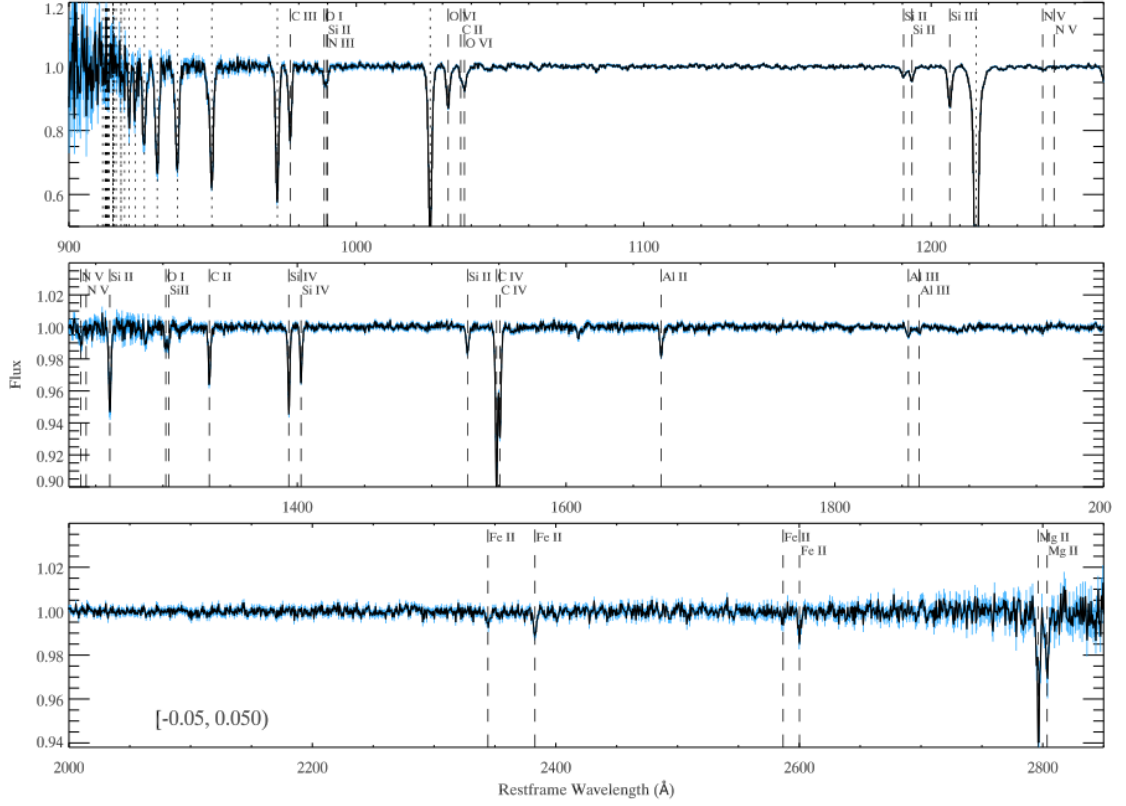


Figure 9.5: Composite spectrum of Ly $\alpha$  absorbers selected with transmission between  $-0.05 < F < 0.05$ . Error bars are shown in blue. Vertical dashed lines indicate metal lines identified and dotted vertical lines denote the locations of the Lyman series. A sample of 242,150 forests from DR9 were used. Figure extracted from [Pieri et al. \(2013\)](#).

a new source of noise in our estimates. Fitting for the BAO peak position parameters,  $\alpha_{\parallel}$  and  $\alpha_{\perp}$ , yields a mean difference of  $\Delta\alpha_{\parallel} = 0.002 \pm 0.003$  and  $\Delta\alpha_{\perp} = 0.003 \pm 0.009$ , indicating no significant shifts in the BAO peak position.

This result was obtained assuming that our metal implementation on mocks is realistic. However our model could be limited by different factors: the experimental uncertainty on the true amplitude of individual metal lines derived from the stacks, or the leak of scatter in the amount of metal absorption added in mock spectra. However, it is difficult to improve the characterization of the metal absorption distribution from data because metal lines are often blended with Ly $\alpha$  absorption inside the forest.

Given those uncertainties in our metal model, we also performed a model independent test using only data to see the importance of the metal correlations for BAO. This test consists in fitting the correlation function but excluding bins corresponding to transverse separations below a given  $r_{\perp, \min}$ , therefore neglecting the region where metal correlations are expected to be seen. Different values for  $r_{\perp, \min}$  were used and results are shown in Fig. 9.7. We see in this plot that no significant change is observed in  $(\alpha_{\parallel}, \alpha_{\perp})$  when cutting data bins, except for the natural increase in error bars due to the reduced number of degrees of freedom in the fits.

Given these tests, we conclude that at this precision, metals do not affect our measurements of the BAO peak position in a significant way.



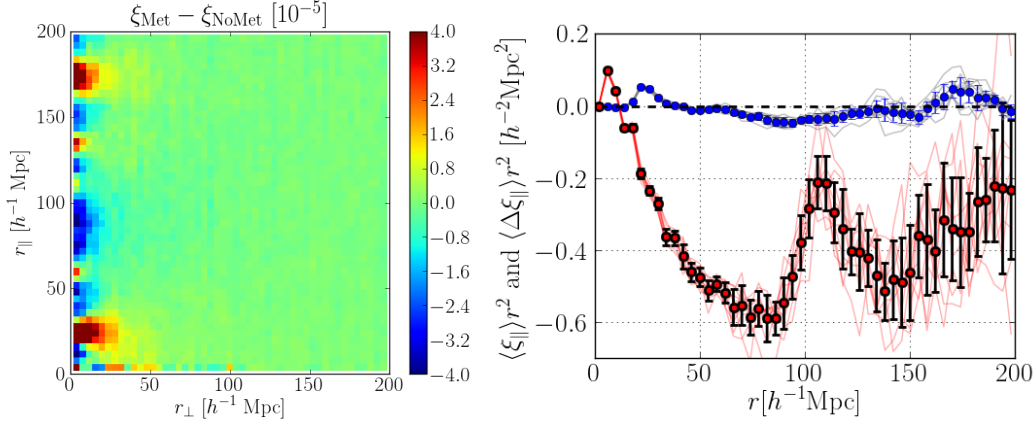


Figure 9.6: The effect of metals on the stacked measurement correlation function of 10 mock sets. Left: difference between correlations measured of mock with and without metal absorption included. Right: the same difference averaged over  $0.8 < \mu < 1.0$  (blue points) compared with the measurement itself (red points). The light red and blue lines show the results for individual mock sets. Error bars are the standard deviation of the 10 estimates.

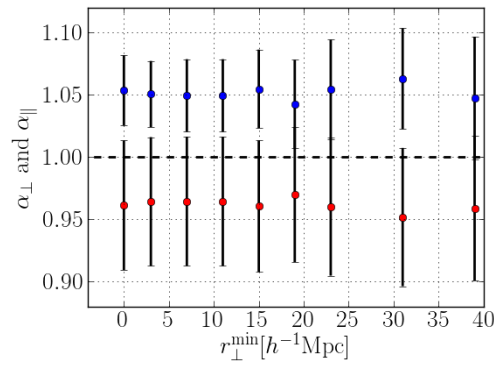


Figure 9.7: Values  $\alpha_{\parallel}$  (blue dots) and  $\alpha_{\perp}$  (red dots) recovered from the DR11 data for different choices of the minimum transverse separation,  $r_{\perp}^{\text{min}}$ , used in the fit. The measured values do not change significantly when eliminating the small  $r_{\perp}$  bins that may be contaminated by correlations due to absorption by metals.

## Chapter 10

# Cosmology from BAO measurements

This chapter is dedicated to the cosmological implications of the BAO measurement presented previously. In § 10.1 we discuss the consequences of estimates using Ly $\alpha$  forests presented in § 8.5). The same kind of discussion is made with the galaxy measurements of § 4 in § 10.2. Finally, in § 10.3 we combine galaxy and Ly $\alpha$  BAO and we discuss the derived cosmological parameters.

### 10.1 Implications of BAO in the Ly $\alpha$ forest

In this section, implications of BAO measurements on cosmological parameters are analyzed. When combined with CMB measurements, Ly $\alpha$  BAO at  $z = 2.34$  has the power to constrain dark energy density.

We compare now our results with predictions given by the  $\Lambda$ CDM model using CMB measurements. Two different sets of CMB constraints are used and summarized in Tab. 10.1. The first one is obtained from a combination of Planck temperature (Planck Collaboration et al., 2013) and WMAP polarization measurements<sup>1</sup>, denoted Planck+WP hereafter. The second set uses WMAP nine-year data (polarization included, Bennett et al. (2013)) combined with data from high resolution experiments ACT and SPT<sup>2</sup> (Calabrese et al., 2013), denoted WMAP9+ACT+SPT hereafter. These two CMB data sets have slightly different constraints on cosmological parameters. In particular, the matter density parameter of Planck+WP is 6% larger than the value obtained with WMAP9+ACT+SPT, which results in a disagreement of  $1.5\sigma$  on the Hubble constant  $h$  between these two sets.

<sup>1</sup>The WMAP polarization breaks the degeneracy between the power spectrum amplitude  $A_s$  and the optical depth to the last scattering surface  $\tau(z_*)$ , increasing the accuracy in the measured matter  $\Omega_m h^2$  and baryon densities  $\Omega_b h^2$  by about 12%.

<sup>2</sup>The high resolution experiments are more sensitive to foreground emissions and point sources in the CMB. Full-sky experiments use this data to fit more accurately foreground models. The more accurate foreground extraction and the larger angular range probed reduces error bars in cosmological parameters, in particular the angular sound horizon at recombination (due to the larger number of acoustic peaks).

Table 10.1: Parameters of the fiducial flat  $\Lambda$ CDM cosmological model used for this analysis, the flat  $\Lambda$ CDM model derived from Planck and low- $\ell$  WMAP polarisation data, ‘Planck + WP’ (Planck Collaboration et al., 2013), and a flat  $\Lambda$ CDM model derived from the WMAP, ACT and SPT data (Calabrese et al., 2013). The models are defined by the cold dark matter, baryon and massive neutrinos densities, the Hubble constant and the number of light neutrino species. The sound horizon at the drag epoch,  $r_d$  is calculated using CAMB (which can be approximated with equation (55) of Anderson et al. (2013) to a precision of 0.1%).

	fiducial	Planck + WP	WMAP9 +ACT+SPT
$\Omega_m h^2$	0.1323	0.14305	0.1347
$= \Omega_c h^2$	0.1090	0.12038	0.1122
$+ \Omega_b h^2$	0.0227	0.022032	0.02252
$+ \Omega_\nu h^2$	0.0006	0.0006	0
$h$	0.7	0.6704	0.714
$N_\nu$	3	3	3
$\Omega_m$	0.27	0.3183	0.265
$r_d$ (Mpc)	149.7 (104.80 $h^{-1}$ )	147.4 (98.79 $h^{-1}$ )	149.1 (106.4 $h^{-1}$ )
$D_A(2.34)/r_d$	11.59	11.76	11.47
$D_H(2.34)/r_d$	8.708	8.570	8.648

The standard fit values for  $(\alpha_{\parallel}, \alpha_{\perp})$  from Table 8.2,

$$\alpha_{\parallel} = 1.054_{-0.031}^{+0.032} \quad \text{and} \quad \alpha_{\perp} = 0.973_{-0.051}^{+0.056},$$

combined with the fiducial values from Table 10.1 yield the following results

$$\frac{D_H(2.34)}{r_d} = 9.18 \pm 0.28(1\sigma) \pm 0.6(2\sigma) \quad (10.1)$$

$$\frac{D_A(2.34)}{r_d} = 11.28 \pm 0.65(1\sigma) {}_{-1.2}^{+2.8}(2\sigma). \quad (10.2)$$

The blue shading in Figure 10.1 shows 68.3% and 95.5% likelihood contours for these parameters, showing the anti-correlation between them. These constraints can be expressed equivalently as

$$\begin{aligned} H(z = 2.34) &= (222 \pm 7 \text{ km s}^{-1} \text{ Mpc}^{-1}) \times \frac{147.4 \text{ Mpc}}{r_d} \\ D_A(z = 2.34) &= (1662 \pm 96 \text{ Mpc}) \times \frac{r_d}{147.4 \text{ Mpc}}, \end{aligned} \quad (10.3)$$

where we have scaled by the value  $r_d = 147.4$  Mpc from the Planck+WP model in Table 10.1.

Our measurements of  $(D_A/r_d, D_H/r_d)$  can be compared to expected values obtained from CMB measurements. The Planck+WP values in a  $\Lambda$ CDM model from Tab. 10.1, (8.570, 11.76) are at  $1.8\sigma$  from our measurements, while the WMAP9+ACT+SPT values, (8.648, 11.47), are at  $1.6\sigma$ . These differences are mainly driven by the slightly higher value of our  $D_H/r_d$  estimate. On the other hand, our estimate of  $D_A/r_d$  has large errors and is in agreement with both CMB predictions in Tab. 10.1. These results are independent of the choice for the fiducial cosmology used in the analysis. We re-analyzed our data using a different fiducial cosmology based on Planck+WP parameters, which yields same results.

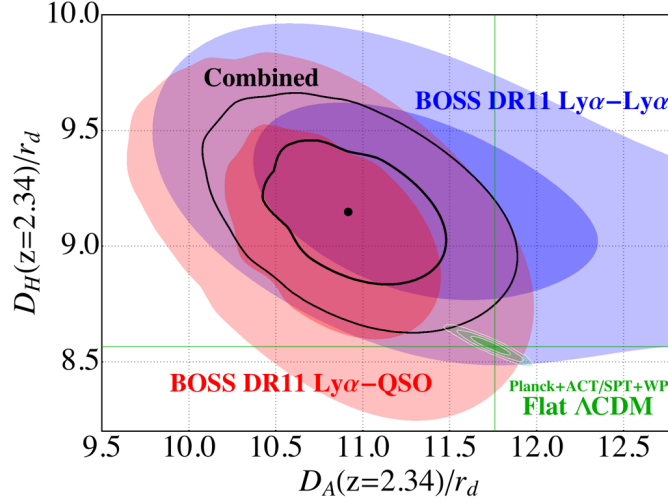


Figure 10.1: Constraints on  $(D_A/r_d, D_H/r_d)$ . Contours show 68.3% ( $\Delta\chi^2 = 2.3$ ) and 95.5% ( $\Delta\chi^2 = 6.2$ ) contours from the Ly $\alpha$  forest auto-correlation (this work, blue), the quasar Ly $\alpha$  forest cross-correlation (Font-Ribera et al., 2013) (red), and the combined constraints (black). The green contours are CMB constraints calculated using the Planck+WP+SPT+ACT chains (Planck Collaboration et al., 2013) assuming a flat  $\Lambda$ CDM cosmology.

In order to see what kind of models are in agreement with our measured values of  $(D_A/r_d, D_H/r_d)$ , we computed the probability distribution for parameters of an open  $\Lambda$ CDM model (which allows non zero curvature). Results are shown in Fig. 10.2, where contours give the 68 and 95.5% confidence regions for  $\Omega_m$  and  $\Omega_\Lambda$ . In addition to our measurements, priors on the value of the Hubble constant  $h$  and on the baryon density  $\Omega_b h^2$  were used when computing these contours. The prior on  $h$  is given by a Gaussian centered at 0.706 with 0.032 as standard deviation, covering measurements from both the distance ladder (Riess et al., 2011) and the CMB assuming a  $\Lambda$ CDM model (Planck Collaboration et al., 2013). For  $\Omega_b h^2$ , we used a Gaussian prior based on the Planck+WP measurement given by  $\Omega_b h^2 = 0.02205 \pm 0.00033$ . Figure 10.2 shows that our BAO measurement at  $z = 2.34$  prefers lower values for  $\Omega_m$  in a *flat* cosmology than the Planck prediction (represented by the star). In an open  $\Lambda$ CDM, the Ly $\alpha$  measurement has a slight preference for a closed universe. However,  $\Omega_k < 1$  is not detected significantly due to the small constraining power of a single BAO measurement. In § 10.3, we combine this result with BAO measurements using galaxies at lower redshift.

Using the same forests, a BAO measurement was performed using the cross-correlation between quasars and the Ly $\alpha$  forest (Font-Ribera et al., 2013). Results are shown by the red contours in Fig. 10.1. Due to the higher density bias of quasars (around 4), the cross-correlation BAO signal is significant even if the quasar density is low. Also, for the same reason the redshift space distortion effect is reduced ( $\beta$  is the ratio between the velocity and density biases), yielding similar errors in both radial and transverse directions. The error in  $D_A/r_d$  is even smaller than the same value for the Ly $\alpha$  auto-correlation. The  $(D_A/r_d, D_H/r_d)$  measurement for the cross-correlation alone also shows a  $1.5\sigma$  difference with respect to  $\Lambda$ CDM predictions from CMB measurements.

The statistical errors in the cross-correlation BAO measurement are dominated by the quasar shot noise while the auto-correlation BAO measurement errors are dominated by instrumental

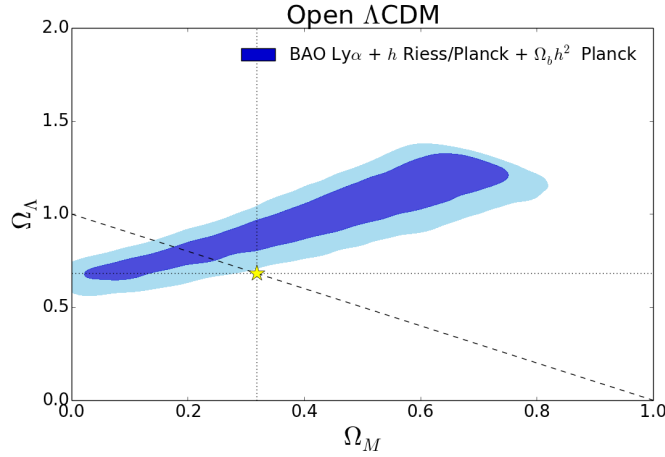


Figure 10.2: Constraints on the o $\Lambda$ CDM parameters  $(\Omega_\Lambda, \Omega_m)$  based on the auto-correlation contours of figure 10.1. The contours show 68.3% and 95.5% confidence levels. The Planck value of  $\Omega_b h^2$  is assumed together with a Gaussian prior for  $H_0 = 70.6 \pm 3.2 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . The yellow star is the Planck  $\Lambda$ CDM measurement and the dashed line corresponds to a flat universe.

errors and small-scale clustering (for instance the two quasar diagram in the Wick theorem computation, see § 6.3.2). For this reason, the statistical errors in the two BAO measurements are almost completely uncorrelated. We have therefore combined the two likelihood surfaces as independent to produce the joint likelihood contours shown by the solid lines in Fig. 10.1. Marginalized 1-d constraints from the combined likelihood are:

$$\frac{D_H(2.34)}{r_d} = 9.15_{-0.21}^{+0.20} (1\sigma) \text{ }_{-0.42}^{+0.40} (2\sigma) \quad (10.4)$$

$$\frac{D_A(2.34)}{r_d} = 10.93_{-0.34}^{+0.35} (1\sigma) \text{ }_{-0.65}^{+0.75} (2\sigma) . \quad (10.5)$$

The combined measurement of  $(D_A/r_d, D_H/r_d)$  differs from CMB predictions by  $2.5\sigma$ , as shown in Fig. 10.1. Central values differ by 7% from the green contours from Planck+WP+ACT+SPT (which are very similar to Planck+WP in Tab. 10.1). The “tension” between the CMB-constrained flat  $\Lambda$ CDM model and the auto-correlation measurement of  $D_H$  is evident in the top panel of Figure 8.4, where the peak in the data is visually to the left of the peak in the fiducial model (and would be even more to the left of the Planck+WP model).

This tension could be interpreted as coming from systematic errors in our auto-correlation measurement. However, during Chapters 7 and 9 we performed tests on all steps of our analysis chain using both mock catalogs and real data. The continuum fitting methods, noise properties, sky-residuals, metals, high column density systems, model fitting methods: all those possible systematic sources were investigated, and no systematic errors could be found at this precision level.

The same kind of systematic tests could not be performed for the cross-correlation BAO measurement. This is because our mock catalogs do not simulate the quasar- $\text{Ly}\alpha$  cross-correlation. The absorption fields of these mocks are build completely ignoring correlations with quasars. In principle, mocks including those correlations can be produced with our method discussed in

§ 5.1.3, where the same correlated field could be used to place quasars and their Ly $\alpha$  forest. However, we leave this topic as subject for future work.

While it is premature to conclude that a major modification of  $\Lambda$ CDM is needed, it is nevertheless interesting to note what sort of changes are indicated by the data. The most widely discussed extensions to flat  $\Lambda$ CDM, allowing non-zero space curvature or a dark energy equation-of-state with  $w \neq -1$ , do not readily resolve the tension seen in Fig. 10.1 without running afoul of other constraints. This is because of the necessity of decreasing  $D_A$ (2.34) while increasing  $D_H$ (2.34) (which would solve the tension), something not easily done since the former is related to the integral of the latter. In addition, as we will see in the next section, models also need to match galaxy BAO measurements at lower redshift, which are more consistent with the CMB.

## 10.2 Implications from galaxy BAO measurements

In Chapter 4 we presented the BAO measurement using two samples of BOSS galaxies : LOWZ and CMASS, yielding measurements of  $D_V/r_d$  (isotropic BAO),  $D_A/r_d$  (transverse BAO) and  $D_H(z)/r_d$  (radial BAO) at effective redshifts of 0.32 and 0.57, respectively. These measurements are obtained from the fits of  $\alpha$ ,  $\alpha_\perp$  and  $\alpha_\parallel$ , respectively, on the galaxy correlation function. The values from Table 4.1 and Eq. 4.15 yields, using the fiducial cosmology from Anderson et al. (2013), for the isotropic case,

$$D_V(0.32) = (1264 \pm 25 \text{ Mpc}) \left( \frac{r_d}{r_{d,\text{fid}}} \right), \quad (10.6)$$

$$D_V(0.57) = (2056 \pm 20 \text{ Mpc}) \left( \frac{r_d}{r_{d,\text{fid}}} \right), \quad (10.7)$$

and for the anisotropic case,

$$H(0.32) = (81.7^{+4.0}_{-4.4} \text{ km s}^{-1} \text{ Mpc}^{-1}) \left( \frac{r_{d,\text{fid}}}{r_d} \right), \quad (10.8)$$

$$D_A(0.32) = (965 \pm 42 \text{ Mpc}) \left( \frac{r_d}{r_{d,\text{fid}}} \right), \quad (10.9)$$

$$H(0.57) = (96.8 \pm 3.4 \text{ km s}^{-1} \text{ Mpc}^{-1}) \left( \frac{r_{d,\text{fid}}}{r_d} \right), \quad (10.10)$$

$$D_A(0.57) = (1421 \pm 20 \text{ Mpc}) \left( \frac{r_d}{r_{d,\text{fid}}} \right), \quad (10.11)$$

These constraints are also shown in Fig. 10.3 and Fig. 10.4 compared with predictions obtained from Planck and WMAP+SPT/ACT Bennett et al. (2013), Hou et al. (2014), Sievers et al. (2013) measurements.

In Fig. 10.3, one can see the 5% tension between the two sets of CMB results, coming from differences on the values of  $\Omega_m h^2$ . The current galaxy BAO data fall in between the two predictions and are clearly consistent with both. CMASS measurements are in agreement with WiggleZ results (Parkinson et al., 2012) at  $z = 0.6$  while LOWZ is compatible with SDSS-II measurements at  $z = 0.275$  (Percival et al., 2010) and at  $z = 0.35$  (Padmanabhan et al., 2012). The 6dFGRS

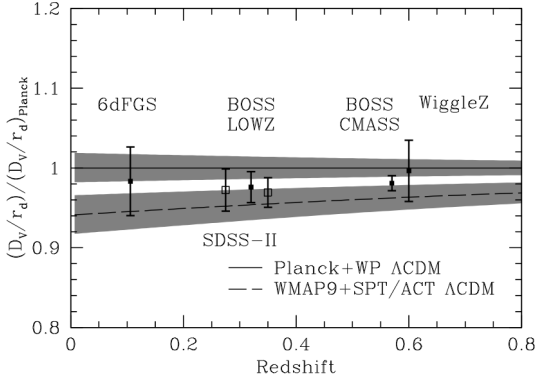


Figure 10.3: The  $D_V(z)/r_d$  measured from galaxy surveys, divided by the best-fit flat  $\Lambda$ CDM prediction from the Planck data. All error bars are  $1\sigma$ . The Planck prediction is a horizontal line at unity, by construction. The dashed line shows the best-fit flat  $\Lambda$ CDM prediction from the WMAP+SPT/ACT results (Bennett et al., 2013). The grey regions shows the  $1\sigma$  variation in the predictions for  $D_V(z)$ , which are dominated by uncertainties in  $\Omega_m h^2$ . SDSS-II results are plotted as open symbols, from Percival et al. (2010) at  $z = 0.275$  and from Padmanabhan et al. (2012) at  $z = 0.35$ . Figure from Anderson et al. (2013).

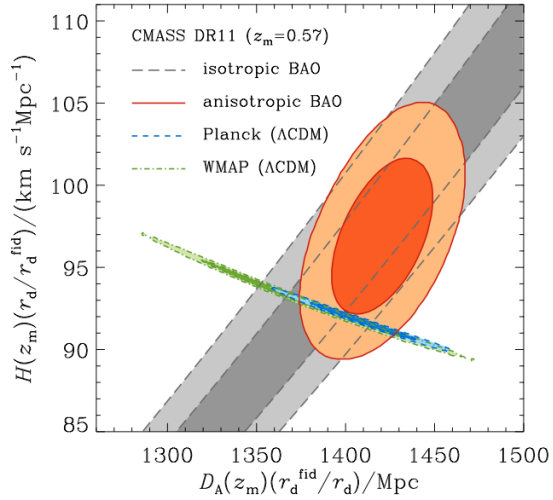


Figure 10.4: Comparison of the 68 and 95% constraints in the  $D_A(0.57)(r_{d,\text{fid}}/r_d) - H(0.57)(r_{d,\text{fid}}/r_d)$  plane from CMASS consensus anisotropic (orange) and isotropic (grey) BAO constraints. The Planck contours correspond to Planck+WMAP polarization (WP) and no lensing. The green contours show the constraints from WMAP9. Figure from Anderson et al. (2013).

(Beutler et al., 2011) result at  $z = 0.1$  is also shown. In § 10.3, we use the 6dFGRS, LOWZ and CMASS values of  $D_V/r_d$  to create a combined set of galaxy BAO measurements for constraining cosmological parameters.

Figure 10.4 shows the constraints of the anisotropic BAO measurement of CMASS galaxies in the  $D_A - H$  plane. The 68 and 95% confidence levels define the contours. The contours from the isotropic measurement are also plotted, creating the gray band seen in this plot (since  $D_V \propto D_A^2/H$ ). The expected values in a  $\Lambda$ CDM model using WMAP and Planck measurements yields narrow ellipses due to the high precision in determining the sound horizon at  $z_* = 1089$ . As discussed in § 1.4.6, the amplitude and the relative height of the acoustic peaks in the CMB temperature power spectrum give values for  $\Omega_m h^2$  and  $\Omega_b h^2$ . Combining with the measurement of the sound horizon breaks the degeneracy between  $\Omega_m$  and  $h$  in a  $\Lambda$ CDM model. The main uncertainty in CMB measurements is the value of the cold dark matter density  $\Omega_c h^2$ . Planck constraints are tighter since they observe 9 acoustic peaks instead of 3, increasing the accuracy on the measurement of the acoustic scale. The anisotropic BAO measurement with CMASS galaxies is in  $1\sigma$  agreement with CMB predictions, as also observed with the isotropic estimates.

Both isotropic and anisotropic measurements provide no indication that additional parameters are needed to describe the expansion history beyond those in flat  $\Lambda$ CDM. However, if one considers more general models with curvature  $\Omega_k$  or a different equation of state for dark energy  $w_{\text{DE}}$  (assumed to be constant over time), the lower redshift BAO measurements are complementary to CMB. This is summarized in Figure 10.5 where we show constraints on  $\Omega_k$  and  $w$  separately,

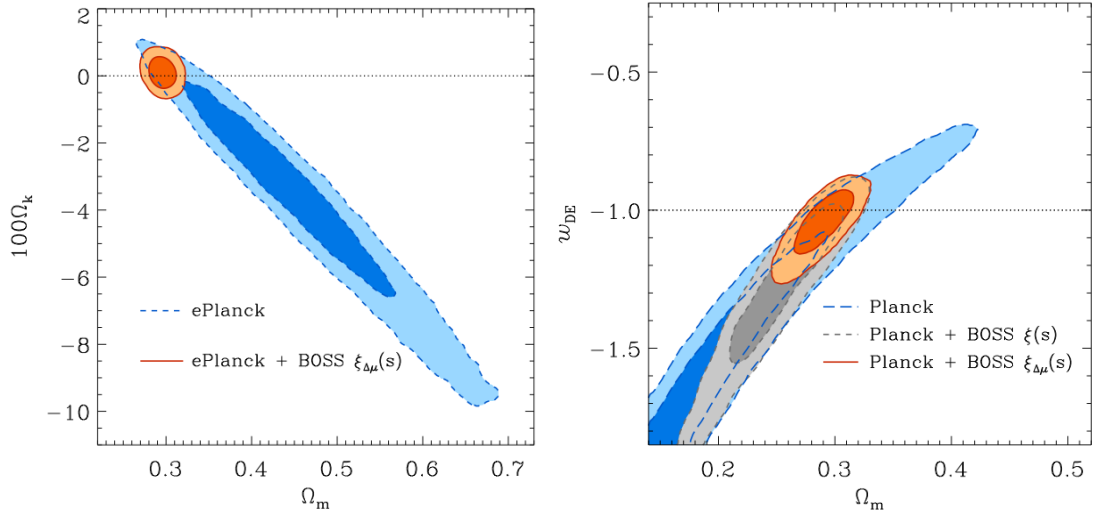


Figure 10.5: Constraints on the curvature  $\Omega_k$  (left panel), the equation of state of dark energy  $w_{\text{DE}}$  (right panel), and matter density fraction  $\Omega_m$ , for ePlanck (Planck+HighL) and CMASS+LOWZ isotropic ( $\xi(s)$ ) and anisotropic ( $\xi_{\Delta\mu}(s)$ ) BAO measurements. Dotted lines show the  $\Lambda$ CDM values:  $\Omega_k = 0$  and  $w_{\text{DE}} = -1$ . Figures from [Sánchez et al. \(2014\)](#).

when combining CMASS and LOWZ anisotropic measurements with CMB measurements, denoted ePlanck (Planck+highL).

In the left panel of Fig. 10.5, BAO measurements at low redshift break the so called geometrical degeneracy ([Efstathiou and Bond, 1999](#)) of CMB measurements, relating models with same prediction for the angular scale of the acoustic peaks. This explains the large contours of ePlanck (Planck+highL) alone in the  $\Omega_k - \Omega_m$  plane. This case gives  $100\Omega_k = -4.2^{+2.7}_{-1.7}$ . When combining with BOSS anisotropic measurements, this degeneracy is broken yielding  $100\Omega_k = 0.10 \pm 0.29$ , in good agreement with a flat model.

In the right panel of Fig. 10.5, a similar degeneracy is observed for the dark matter equation of state in the  $w_{\text{DE}} - \Omega_m$  plane when using only Planck measurements. The isotropic BAO partially breaks this degeneracy, giving  $w_{\text{DE}} = -1.28^{+0.24}_{-0.16}$ . Using the anisotropic measurements reduces even more this degeneracy, resulting in  $w_{\text{DE}} = -1.049 \pm 0.078$ . These results show that the combination CMB and LSS data constrain the dark energy equation of state to 8%, with results in agreement with a cosmological constant ( $w_{\text{DE}} = -1$ ). The same conclusions are obtained by using WMAP data instead of Planck, with only a small shift in the mean value of  $w_{\text{DE}}$  giving  $-0.964 \pm 0.077$ .

Letting both curvature and dark energy equation of state free, the degeneracy is larger. Anisotropic BAO measurements again reduce considerably this degeneracy, yielding  $100\Omega_k = 0.02 \pm 0.43$  and  $w_{\text{DE}} = -1.05 \pm 0.11$ , values consistent with flat  $\Lambda$ CDM.

### 10.3 Combining Ly $\alpha$ forests and galaxies

In this section we discuss the combination of BAO measurements using galaxies at lower redshift and using Ly $\alpha$  forests at  $z = 2.3$ . In the previous section we discussed whether each of these



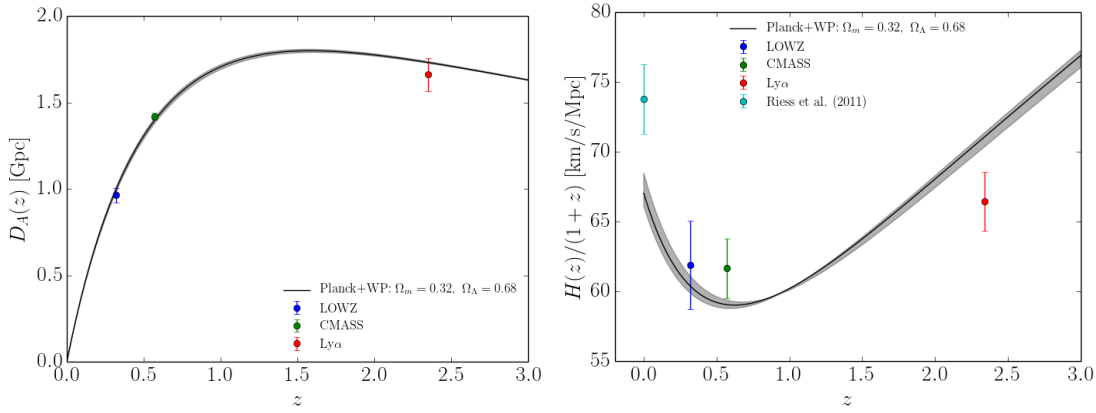


Figure 10.6: Measurements of  $D_A(z)$  (left) and  $H(z)/(1+z)$  (right) for three BOSS data-sets: LOWZ, CMASS and Ly $\alpha$  forest. The gray area shows the Planck prediction for a  $\Lambda$ CDM model.

measurements agrees with predictions from the CMB. Here, we analyze the combination of all those estimates.

Figure 10.6 summarizes the measurements of  $D_A(z)$  and  $H(z)$  discussed previously using BOSS data: LOWZ, CMASS and Ly $\alpha$  forest (auto-correlation only). The  $1\sigma$  expectations from Planck measurements in a  $\Lambda$ CDM model are also shown for comparison. The left panel shows a very good agreement within  $1\sigma$  of all  $D_A(z)$  estimates with the Planck prediction. The right panel of Fig. 10.6 shows  $H(z)/(1+z)$  estimates. This quantity is simply a rewriting of derivative of the scale factor  $\dot{a}(t) = a(t)H(t)$ , that shows more clearly the expected deceleration and subsequent acceleration of the Universe in a  $\Lambda$ CDM cosmology with a non-zero cosmological constant. We see that the galaxy measurements are in agreement within  $1\sigma$  with the Planck expectations, with a CMASS value somewhat higher also seen in Fig. 10.3 which is related to a high value of  $\Omega_m$  measured by Planck. The Ly $\alpha$  measurement of  $H(z)$  is also below Planck predictions by  $1.8\sigma$ , as discussed in § 10.1. If combined with the quasar-Ly $\alpha$  cross-correlation measurement at the same  $z$ , this tension increases to  $\sim 2.5\sigma$ . For this reason the combination of CMB, galaxies and Ly $\alpha$  BAO measurements is not straightforward in the context of a standard  $\Lambda$ CDM model. A more general model is needed to fit all measurements at a time.

For the combination of our measurements, we considered a more general cosmological model with curvature  $\Omega_k$  and an equation of state for dark-energy  $w$ , called  $ow$ CDM. We used Monte Carlo chains to compute the combined likelihood surfaces and determine these parameters together with  $\Omega_m$ ,  $\Omega_\Lambda$ . Instead of using all Planck  $\Lambda$ CDM estimates, as in § 10.1, we have chosen to be more conservative and only use a prior on the baryon density  $\Omega_b h^2 = 0.02207 \pm 0.00033$  from Planck and a larger prior on  $h = 0.706 \pm 0.032$  allowing Planck and direct measurement values for  $H_0$  to be equally likely.

Figure 10.7 shows three marginalized likelihoods for  $\Omega_m$ ,  $\Omega_\Lambda$  and  $w$ , and 68 and 95% likelihood contours in three parameter spaces,  $(\Omega_\Lambda, \Omega_m)$ ,  $(w, \Omega_m)$  and  $(w, \Omega_\Lambda)$  for three data sets: galaxies alone (green), Ly $\alpha$  forest alone, and both combined. The Planck  $\Lambda$ CDM central values are shown with stars. In this generalized model, the Ly $\alpha$  measurement alone gives very loose constraints in these parameters due to a geometrical degeneracy. The same happens for galaxies alone, even

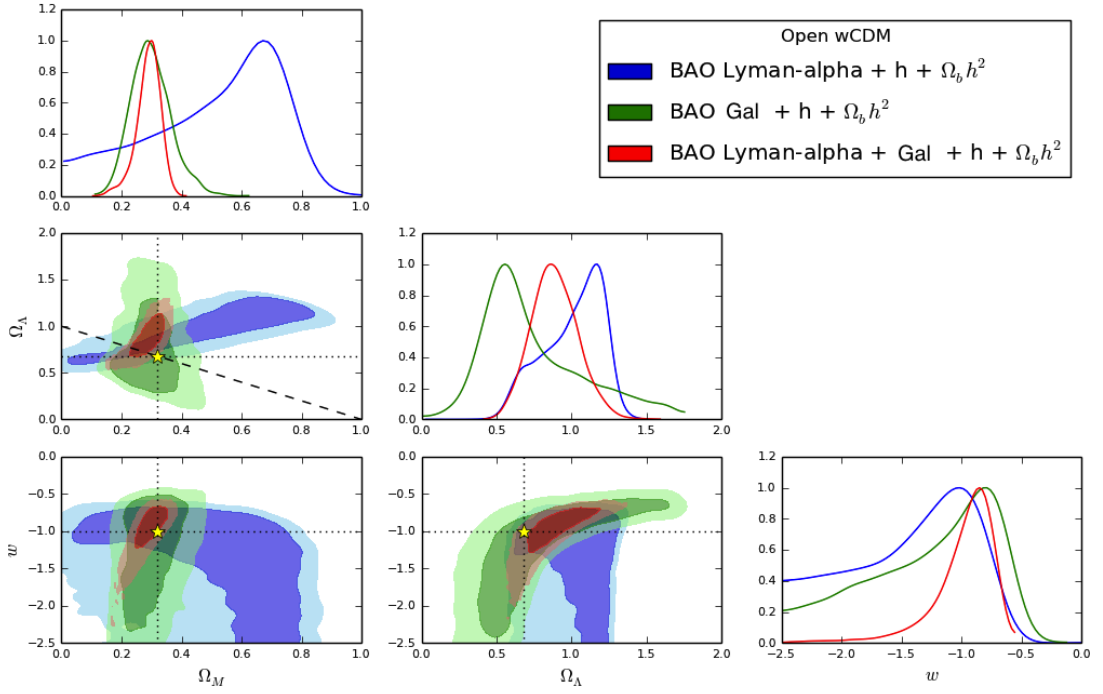


Figure 10.7: Open  $w$ CDM constraints for combined BAO measurements.

though the constraints on  $\Omega_m$  have reasonable values in agreement with CMB measurements with not so large errors ( $\sim 0.1$ ). When combining galaxy and Ly $\alpha$  BAO measurements, constraints are much tighter due to the orthogonality between likelihood surfaces arriving from the redshift difference of the two measurements. With the combined likelihood, even in this general model, we are able to recover values in agreement with  $\Lambda$ CDM at  $1\sigma$  level. This shows that BAO measurements are able to measure, with a more conservative prior on the baryon density and  $h$ :

- a non-zero dark energy component with a equation of state compatible with a cosmological constant;
- reasonable values for the matter density;
- and values for the curvature compatible with a flat model.

These results demonstrate the power of BAO as a probe for cosmological measurements over a very large range of redshifts. Also, they show the impressive agreement of the  $\Lambda$ CDM model with all measurements discussed here.



# Conclusions

In this thesis, I presented baryon acoustic oscillations measured in the large scales structures of the Universe seen by the BOSS survey. My work was mainly dedicated to the measurement at  $z = 2.35$  using  $\text{Ly}\alpha$  forests, but I also contributed to the  $z = 0.57$  measurement using galaxies.

BOSS is the largest spectroscopic survey to date, that measured more than 1.3 million luminous galaxies over  $0.1 < z < 0.7$ , and 150,000 quasar spectra at  $2.0 < z < 3.5$ , covering the largest volume for large scale structure studies.

Measurements of galaxy clustering lead to a sub per cent level estimate on the isotropic BAO peak position at  $z = 0.57$  with the DR11 CMASS sample. This estimate is compatible with expectations of the concordance  $\Lambda$ CDM model with cosmological parameters derived from the cosmic microwave background temperature power spectrum. Still using galaxies, I developed an iterative optimal estimator for isotropic correlation functions. Applied to Data Release 9 CMASS galaxies, the new estimator gives an improvement of 20% in errors of the BAO peak position with respect to the usual estimator Landy-Szalay.

The measurement of BAO using the  $\text{Ly}\alpha$  forests is the main topic of this thesis, where my contributions were important over all steps of the analysis, from observations, data reduction, simulations and tests for systematic effects in our analysis, to the final measurement and its interpretation.

Concerning mock  $\text{Ly}\alpha$  forest catalogs, I developed a new method to generate spatially correlated Gaussian random fields useful in the creation of mock  $\text{Ly}\alpha$  absorption fields. My new method leads to same results as the usual method but with several improvements: computing time scaling linearly with the number of quasars, virtually no memory requirements and no need to divide the survey into independent regions. For the second step of mock creation, which consists in creating realistic BOSS quasar spectra from absorption fields, I developed the *MockExpander* package that performs this transformation, also allowing the user to turn on and off different systematic effects. This package was distributed to the collaboration and will be publicly available as part of DR12.

These mocks were used to perform exhaustive tests on the full BAO analysis chain. I have studied the effect of many possible sources of systematic effects on the final measurement: noise mis-estimates, continuum fitting, sky residuals, calibration errors, metal and high column density absorption. No strong evidence for systematic errors or biases were found at this statistical precision level. Mock BAO measurements are consistent with the input model, with reasonable

error estimates, validating our methods to compute the covariance matrix and the model fitting procedure.

The BAO peak in the DR11 Ly $\alpha$  correlation function was detected at high significance, leading to a 3% measurement of its position in the radial direction and 5.4% in the transverse direction. Changing slightly the data set and analysis assumptions results in sub- $\sigma$  shifts in our measurement. The derived Hubble expansion rate at  $z = 2.35$  is at  $1.8\sigma$  from  $\Lambda$ CDM expectations coming from Planck measurements of the CMB. This difference might just be a statistical fluctuation or an indication for new physics happening between  $z_* \sim 1000$  and  $z = 2.35$ . All tests performed with data and mock catalogs indicate that the systematic contribution is negligible.

When combined with the DR11 quasar-Ly $\alpha$  cross-correlation, both the angular diameter distance and the Hubble expansion rate are in tension with Planck measurements at  $2$  and  $2.5\sigma$ , respectively. Translating this into cosmological parameters, the matter density fraction needed to fit  $z = 2.35$  BAO measurements should be lowered from  $0.3$  to  $0.2$ , while increasing the dark energy fraction in a flat universe. Also, if allowed to vary, the dark energy density at  $z = 2.35$  is at  $2.5\sigma$  below the expectation from a cosmological constant.

BAO in the large scale structures at low redshift have the power to break the geometrical degeneracy of CMB measurements. If not just a statistical unlucky result, our measurements seem to indicate that new physics would need to be included in our current cosmological model. Our  $2.5\sigma$  discrepancy with predictions is not statistically significant but it demands a thorough consideration. The next generation of BOSS, called “extended BOSS” aims the measurement of BAO using galaxies and quasar clustering in the redshift range not probed in this work ( $0.7 < z < 2.0$ ). The eBOSS survey uses the same telescope as BOSS, but targets different objects. The Dark Energy Spectroscopic Survey (DESI) is the ambitious improved version of BOSS, using a 4-meter class telescope and a focal plane observing 5,000 objects simultaneously. DESI aims the measurement of BAO with all probes of large scale structure, galaxies, quasars and Ly $\alpha$  forest, from  $0.1 < z < 4$ . DESI will not only be able to improve the measurements discussed in this thesis, but it will perform a full measurement of  $H(z)$  over the probed redshift range. These measurements will help us in understanding more about about the origin of the expansion, opening new doors for the exploration of new physics.

# Bibliography

- Adam G. Riess, Lucas Macri, Stefano Casertano, Hubert Lampeitl, Henry C. Ferguson, Alexei V. Filippenko, Saurabh W. Jha, Weidong Li, and Ryan Chornock. A 3% solution: Determination of the hubble constant with the hubble space telescope and wide field camera 3. *The Astrophysical Journal*, 730:119, April 2011. ISSN 0004-637X. doi: 10.1088/0004-637X/730/2/119. URL <http://adsabs.harvard.edu/abs/2011ApJ...730..119R>.
- J. R. Herrnstein, J. M. Moran, L. J. Greenhill, P. J. Diamond, M. Inoue, N. Nakai, M. Miyoshi, C. Henkel, and A. Riess. A 4% geometric distance to the galaxy NGC4258 from orbital motions in a nuclear gas disk. *Nature*, 400(6744):539–541, August 1999. ISSN 00280836. doi: 10.1038/22972. URL <http://arxiv.org/abs/astro-ph/9907013>. arXiv:astro-ph/9907013.
- J. C. Mather, E. S. Cheng, D. A. Cottingham, R. E. Eplee, Jr., D. J. Fixsen, T. Hewagama, R. B. Isaacman, K. A. Jensen, S. S. Meyer, P. D. Noerdlinger, S. M. Read, L. P. Rosen, R. A. Shafer, E. L. Wright, C. L. Bennett, N. W. Boggess, M. G. Hauser, T. Kelsall, S. H. Moseley, Jr., R. F. Silverberg, G. F. Smoot, R. Weiss, and D. T. Wilkinson. Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument. *The Astrophysical Journal*, 420:439–444, January 1994. ISSN 0004-637X. doi: 10.1086/173574. URL <http://adsabs.harvard.edu/abs/1994ApJ...420..439M>.
- C. T. Kowal. Absolute magnitudes of supernovae. *The Astronomical Journal*, 73:1021–1024, December 1968. ISSN 0004-6256. doi: 10.1086/110763. URL <http://adsabs.harvard.edu/abs/1968AJ.....73.1021K>.
- Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, B. Leibundgut, M. M. Phillips, David Reiss, Brian P. Schmidt, Robert A. Schommer, R. Chris Smith, J. Spyromilio, Christopher Stubbs, Nicholas B. Suntzeff, and John Tonry. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116:1009–1038, September 1998. ISSN 0004-6256. doi: 10.1086/300499. URL <http://adsabs.harvard.edu/abs/1998AJ....116.1009R>.
- S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and Supernova Cosmology Project. Measurements of omega and lambda from 42 high-redshift supernovae. *The Astrophysical*

- Journal*, 517:565–586, June 1999. ISSN 0004-637X. doi: 10.1086/307221. URL <http://adsabs.harvard.edu/abs/1999ApJ...517..565P>.
- M. Fukugita, C. J. Hogan, and P. J. E. Peebles. The cosmic baryon budget. *The Astrophysical Journal*, 503:518, August 1998. ISSN 0004-637X. doi: 10.1086/306025. URL <http://adsabs.harvard.edu/abs/1998ApJ...503..518F>.
- F. Zwicky. On the masses of nebulae and of clusters of nebulae. *The Astrophysical Journal*, 86:217, October 1937. ISSN 0004-637X. doi: 10.1086/143864. URL <http://adsabs.harvard.edu/abs/1937ApJ...86..217Z>.
- A. Vikhlinin, A. Kravtsov, W. Forman, C. Jones, M. Markevitch, S. S. Murray, and L. Van Speybroeck. Chandra sample of nearby relaxed galaxy clusters: Mass, gas fraction, and mass-temperature relation. *The Astrophysical Journal*, 640:691–709, April 2006. ISSN 0004-637X. doi: 10.1086/500288. URL <http://adsabs.harvard.edu/abs/2006ApJ...640..691V>.
- Benoît Famaey and Stacy S. McGaugh. Modified newtonian dynamics (MOND): Observational phenomenology and relativistic extensions. *Living Reviews in Relativity*, 15, 2012. ISSN 1433-8351. doi: 10.12942/lrr-2012-10. URL <http://relativity.livingreviews.org/Articles/lrr-2012-10/>.
- L. Bergström. Dark matter evidence, particle physics candidates and detection methods. *Annalen der Physik*, 524:479–496, October 2012. ISSN 0003-3804. doi: 10.1002/andp.201200116. URL <http://adsabs.harvard.edu/abs/2012AnP...524..479B>.
- Stefan Funk. Indirect detection of dark matter with gamma rays. *ArXiv e-prints*, 1310:2695, October 2013. URL <http://adsabs.harvard.edu/abs/2013arXiv1310.2695F>.
- The Fermi-LAT Collaboration, :, M. Ackermann, A. Albert, B. Anderson, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, E. Bissaldi, E. D. Bloom, E. Bonamente, A. Bouvier, T. J. Brandt, J. Bregeon, M. Brigida, P. Bruel, R. Buehler, S. Buson, G. A. Caliandro, R. A. Cameron, M. Caragiulo, P. A. Caraveo, C. Cecchi, E. Charles, A. Chekhtman, J. Chiang, S. Ciprini, R. Claus, J. Cohen-Tanugi, J. Conrad, F. D’Ammando, A. de Angelis, C. D. Dermer, S. W. Digel, E. do Couto e Silva, P. S. Drell, A. Drlica-Wagner, R. Essig, C. Favuzzi, E. C. Ferrara, A. Franckowiak, Y. Fukazawa, S. Funk, P. Fusco, F. Gargano, D. Gasparrini, N. Giglietto, M. Giroletti, G. Godfrey, G. A. Gomez-Vargas, I. A. Grenier, S. Guiriec, M. Gustafsson, M. Hayashida, E. Hays, J. Hewitt, R. E. Hughes, T. Jogler, T. Kamae, J. Knödseder, D. Kocevski, M. Kuss, . Larsson, L. Latronico, M. Llana Garde, F. Longo, F. Loparco, M. N. Lovellette, P. Lubrano, G. Martinez, M. Mayer, M. N. Mazziotta, P. F. Michelson, W. Mitthumsiri, T. Mizuno, A. A. Moiseev, M. E. Monzani, A. Morselli, I. V. Moskalenko, S. Murgia, R. Nemmen, E. Nuss, T. Ohsugi, E. Orlando, J. F. Ormes, J. S. Perkins, F. Piron, G. Pivato, T. A. Porter, S. Rainò, R. Rando, M. Razzano, S. Razzaque, A. Reimer, O. Reimer, S. Ritz, M. Sánchez-Conde, N. Sehgal, C. Sgrò, E. J. Siskind, P. Spinelli, L. Strigari, D. J. Suson, H. Tajima, H. Takahashi, J. B. Thayer, L. Tibaldo, M. Tinivella, D. F. Torres, Y. Uchiyama, T. L. Usher, J. Vandenbroucke, G. Vianello, V. Vitale, M. Werner, B. L. Winer, K. S. Wood, M. Wood, G. Zaharijas, and S. Zimmer. Dark matter constraints from observations of 25 milky way satellite galaxies with the fermi large area telescope. *ArXiv e-prints*, 1310:828, October 2013. URL <http://adsabs.harvard.edu/abs/2013arXiv1310.0828T>.

- Q. R. Ahmad, R. C. Allen, T. C. Andersen, J. D. Anglin, J. C. Barton, E. W. Beier, M. Bercovitch, J. Bigu, S. D. Biller, R. A. Black, I. Blevis, R. J. Boardman, J. Boger, E. Bonvin, M. G. Boulay, M. G. Bowler, T. J. Bowles, S. J. Brice, M. C. Browne, T. V. Bullard, G. Bühler, J. Cameron, Y. D. Chan, H. H. Chen, M. Chen, X. Chen, B. T. Cleveland, E. T. H. Clifford, J. H. M. Cowan, D. F. Cowen, G. A. Cox, X. Dai, F. Dalnoki-Veress, W. F. Davidson, P. J. Doe, G. Doucas, M. R. Dragowsky, C. A. Duba, F. A. Duncan, M. Dunford, J. A. Dunmore, E. D. Earle, S. R. Elliott, H. C. Evans, G. T. Ewan, J. Farine, H. Fergani, A. P. Ferraris, R. J. Ford, J. A. Formaggio, M. M. Fowler, K. Frame, E. D. Frank, W. Frati, N. Gagnon, J. V. Germani, S. Gil, K. Graham, D. R. Grant, R. L. Hahn, A. L. Hallin, E. D. Hallman, A. S. Hamer, A. A. Hamian, W. B. Handler, R. U. Haq, C. K. Hargrove, P. J. Harvey, R. Hazama, K. M. Heeger, W. J. Heintzelman, J. Heise, R. L. Helmer, J. D. Hepburn, H. Heron, J. Hewett, A. Hime, M. Howe, J. G. Hykawy, M. C. P. Isaac, P. Jagam, N. A. Jelley, C. Jillings, G. Jonkmans, K. Kazkaz, P. T. Keener, J. R. Klein, A. B. Knox, R. J. Komar, R. Kouzes, T. Kutter, C. C. M. Kyba, J. Law, I. T. Lawson, M. Lay, H. W. Lee, K. T. Lesko, J. R. Leslie, I. Levine, W. Locke, S. Luoma, J. Lyon, S. Majerus, H. B. Mak, J. Maneira, J. Manor, A. D. Marino, N. McCauley, A. B. McDonald, D. S. McDonald, K. McFarlane, G. McGregor, R. Meijer Drees, C. Mifflin, G. G. Miller, G. Milton, B. A. Moffat, M. Moorhead, C. W. Nally, M. S. Neubauer, F. M. Newcomer, H. S. Ng, A. J. Noble, E. B. Norman, V. M. Novikov, M. O'Neill, C. E. Okada, R. W. Ollerhead, M. Omori, J. L. Orrell, S. M. Oser, A. W. P. Poon, T. J. Radcliffe, A. Roberge, B. C. Robertson, R. G. H. Robertson, S. S. E. Rosendahl, J. K. Rowley, V. L. Rusu, E. Saettler, K. K. Schaffer, M. H. Schwendener, A. Schülke, H. Seifert, M. Shatkay, J. J. Simpson, C. J. Sims, D. Sinclair, P. Skensved, A. R. Smith, M. W. E. Smith, T. Spreitzer, N. Starinsky, T. D. Steiger, R. G. Stokstad, L. C. Stonehill, R. S. Storey, B. Sur, R. Tafirout, N. Tagg, N. W. Tanner, R. K. Taplin, M. Thorman, P. M. Thornewell, P. T. Trent, Y. I. Tserkovnyak, R. Van Berg, R. G. Van de Water, C. J. Virtue, C. E. Waltham, J.-X. Wang, D. L. Wark, N. West, J. B. Wilhelm, J. F. Wilkerson, J. R. Wilson, P. Wittich, J. M. Wouters, and M. Yeh. Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory. *Phys. Rev. Lett.*, 89(1):011301, June 2002. doi: 10.1103/PhysRevLett.89.011301. URL <http://link.aps.org/doi/10.1103/PhysRevLett.89.011301>.
- J. Beringer, J. F. Arguin, R. M. Barnett, K. Copic, O. Dahl, D. E. Groom, C. J. Lin, J. Lys, H. Murayama, C. G. Wohl, W. M. Yao, P. A. Zyla, C. AMSler, M. Antonelli, D. M. Asner, H. Baer, H. R. Band, T. Basaglia, C. W. Bauer, J. J. Beatty, V. I. Belousov, E. Bergren, G. Bernardi, W. Bertl, S. Bethke, H. Bichsel, O. Biebel, E. Blucher, S. Blusk, G. Brooijmans, O. Buchmueller, R. N. Cahn, M. Carena, A. Ceccucci, D. Chakraborty, M. C. Chen, R. S. Chivukula, G. Cowan, G. D'Ambrosio, T. Damour, D. de Florian, A. de Gouvêa, T. DeGrand, P. de Jong, G. Dissertori, B. Dobrescu, M. Doser, M. Drees, D. A. Edwards, S. Eidelman, J. Erler, V. V. Ezhela, W. Fetscher, B. D. Fields, B. Foster, T. K. Gaisser, L. Garren, H. J. Gerber, G. Gerbier, T. Gherghetta, S. Golwala, M. Goodman, C. Grab, A. V. Gribsan, J. F. Grivaz, M. Grünwald, A. Gurtu, T. Gutsche, H. E. Haber, K. Hagiwara, C. Hagmann, C. Hanhart, S. Hashimoto, K. G. Hayes, M. Heffner, B. Heltsley, J. J. Hernández-Rey, K. Hikasa, A. Höcker, J. Holder, A. Holtkamp, J. Huston, J. D. Jackson, K. F. Johnson, T. Junk, D. Karlen, D. Kirkby, S. R. Klein, E. Klempt, R. V. Kowalewski, F. Krauss,



- M. Kreps, B. Krusche, Yu. V. Kuyanov, Y. Kwon, O. Lahav, J. Laiho, P. Langacker, A. Liddle, Z. Ligeti, T. M. Liss, L. Littenberg, K. S. Lugovsky, S. B. Lugovsky, T. Mannel, A. V. Manohar, W. J. Marciano, A. D. Martin, A. Masoni, J. Matthews, D. Milstead, R. Miquel, K. Mönig, F. Moortgat, K. Nakamura, M. Narain, P. Nason, S. Navas, M. Neubert, P. Nevski, Y. Nir, K. A. Olive, L. Pape, J. Parsons, C. Patrignani, J. A. Peacock, S. T. Petcov, A. Piepke, A. Pomarol, G. Punzi, A. Quadt, S. Raby, G. Raffelt, B. N. Ratcliff, P. Richardson, S. Roesler, S. Rolli, A. Romaniouk, L. J. Rosenberg, J. L. Rosner, C. T. Sachrajda, Y. Sakai, G. P. Salam, S. Sarkar, F. Sauli, O. Schneider, K. Scholberg, D. Scott, W. G. Seligman, M. H. Shaevitz, S. R. Sharpe, M. Silari, T. Sjöstrand, P. Skands, J. G. Smith, G. F. Smoot, S. Spanier, H. Spieler, A. Stahl, T. Stanev, S. L. Stone, T. Sumiyoshi, M. J. Syphers, F. Takahashi, M. Tanabashi, J. Terning, M. Titov, N. P. Tkachenko, N. A. Törnqvist, D. Tovey, G. Valencia, K. van Bibber, G. Venanzoni, M. G. Vincter, P. Vogel, A. Vogt, W. Walkowiak, C. W. Walter, D. R. Ward, T. Watari, G. Weiglein, E. J. Weinberg, L. R. Wiencke, L. Wolfenstein, J. Womersley, C. L. Woody, R. L. Workman, A. Yamamoto, G. P. Zeller, O. V. Zenin, J. Zhang, R. Y. Zhu, G. Harper, V. S. Lugovsky, P. Schaffner, and Particle Data Group. Review of particle physics. *Phys. Rev. D*, 86(1):010001, July 2012. doi: 10.1103/PhysRevD.86.010001. URL <http://link.aps.org/doi/10.1103/PhysRevD.86.010001>.
- Scott Dodelson. *Modern cosmology*. 2003. ISBN 0-12-219141-2. URL <http://adsabs.harvard.edu/abs/2003moco.book.....D>.
- Shinji Tsujikawa. Introductory review of cosmic inflation. *arXiv:hep-ph/0304257*, April 2003. URL <http://arxiv.org/abs/hep-ph/0304257>. arXiv: hep-ph/0304257.
- P. J. E. Peebles. *Principles of Physical Cosmology*. 1993. URL <http://adsabs.harvard.edu/abs/1993ppc..book.....P>.
- Planck Collaboration, P. A. R. Ade, N. Aghanim, M. I. R. Alves, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, H. Aussel, C. Baccigalupi, A. J. Banday, R. B. Barreiro, R. Barrena, M. Bartelmann, J. G. Bartlett, N. Bartolo, S. Basak, E. Battaner, R. Battye, K. Benabed, A. Benoit, A. Benoit-Lévy, J.-P. Bernard, M. Bersanelli, B. Bertin-court, M. Bethermin, P. Bielewicz, I. Bikmaev, A. Blanchard, J. Bobin, J. J. Bock, H. Böhringer, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, H. Bourdin, J. W. Bowyer, M. Bridges, M. L. Brown, M. Bucher, R. Burenin, C. Burigana, R. C. Butler, E. Calabrese, B. Cappellini, J.-F. Cardoso, R. Carr, P. Carvalho, M. Casale, G. Castex, A. Catalano, A. Challinor, A. Chamballu, R.-R. Chary, X. Chen, H. C. Chiang, L.-Y. Chiang, G. Chon, P. R. Christensen, E. Churazov, S. Church, M. Clemens, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, B. Comis, F. Couchot, A. Coulais, B. P. Crill, M. Cruz, A. Curto, F. Cuttaia, A. Da Silva, H. Dahle, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, T. Déchelette, J. Delabrouille, J.-M. Delouis, J. Démoclès, F.-X. Désert, J. Dick, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, O. Fabre, E. Falgarone, M. C. Falvella, Y. Fantaye, J. Fergusson, C. Filliard, F. Finelli, I. Flores-Cacho, S. Foley, O. Forni, P. Fosalba, M. Frailis, A. A. Fraisse, E. Franceschi, M. Freschi, S. Fromenteau, M. Frommert, T. C. Gaier, S. Galeotta, J. Gallegos, S. Galli, B. Gandolfo, K. Ganga, C. Gauthier, R. T. Génova-Santos, T. Ghosh, M. Giard,

- G. Giardino, M. Gilfanov, D. Girard, Y. Giraud-Héraud, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Haissinski, J. Hamann, F. K. Hansen, M. Hansen, D. Hanson, D. L. Harrison, A. Heavens, G. Helou, A. Hempel, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, S. Ho, M. Hobson, W. A. Holmes, A. Hornstrup, Z. Hou, W. Hovest, G. Huey, K. M. Huffenberger, G. Hurier, S. Ilić, A. H. Jaffe, T. R. Jaffe, J. Jasche, J. Jewell, W. C. Jones, M. Juvela, P. Kalberla, P. Kangaslahti, E. Keihänen, J. Kerp, R. Kesitalo, I. Khamitov, K. Kiiveri, J. Kim, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, F. Lacasa, G. Lagache, A. Lähteenmäki, J.-M. Lamarre, M. Langer, A. Lasenby, M. Lattanzi, R. J. Laureijs, A. Lavabre, C. R. Lawrence, M. Le Jeune, S. Leach, J. P. Leahy, R. Leonardi, J. León-Tavares, C. Leroy, J. Lesgourgues, A. Lewis, C. Li, A. Liddle, M. Liguori, P. B. Lilje, M. Linden-Vørnle, V. Lindholm, M. López-Caniego, S. Lowe, P. M. Lubin, J. F. Macías-Pérez, C. J. MacTavish, B. Maffei, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, D. Marinucci, M. Maris, F. Marleau, D. J. Marshall, P. G. Martin, E. Martínez-González, S. Masi, M. Massardi, S. Matarrese, T. Matsumura, F. Matthai, L. Maurin, P. Mazzotta, A. McDonald, J. D. McEwen, P. McGehee, S. Mei, P. R. Meinhold, A. Melchiorri, J.-B. Melin, L. Mendes, E. Menegoni, A. Mennella, M. Migliaccio, K. Mikkelsen, M. Millea, R. Miniscalco, S. Mitra, M.-A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, N. Morisset, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, M. Negrello, N. P. H. Nesvadba, C. B. Netterfield, H. U. Nørgaard-Nielsen, C. North, F. Noviello, D. Novikov, I. Novikov, I. J. O'Dwyer, F. Orieux, S. Osborne, C. O'Sullivan, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Paladini, S. Pandolfi, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, P. Paykari, D. Pearson, T. J. Pearson, M. Peel, H. V. Peiris, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, P. Platania, D. Pogosyan, E. Pointecouteau, G. Polenta, N. Ponthieu, L. Popa, T. Poutanen, G. W. Pratt, G. Prézeau, S. Prunet, J.-L. Puget, A. R. Pullen, J. P. Rachen, B. Racine, A. Rahlin, C. Räth, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, A. Riazuelo, S. Ricciardi, T. Riller, C. Ringeval, I. Ristorcelli, G. Robbers, G. Rocha, M. Roman, C. Rosset, M. Rossetti, G. Roudier, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Ruiz-Granados, B. Rusholme, E. Salerno, M. Sandri, L. Sanselme, D. Santos, M. Savelainen, G. Savini, B. M. Schaefer, F. Schiavon, D. Scott, M. D. Seiffert, P. Serra, E. P. S. Shellard, K. Smith, G. F. Smoot, T. Souradeep, L. D. Spencer, J.-L. Starck, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, F. Sureau, P. Sutter, D. Sutton, A.-S. Suur-Uski, J.-F. Sygnet, J. A. Tauber, D. Tavagnacco, D. Taylor, L. Terenzi, D. Texier, L. Toffolatti, M. Tomasi, J.-P. Torre, M. Tristram, M. Tucci, J. Tuovinen, M. Türler, M. Tuttlebee, G. Umama, L. Valenziano, J. Valiviita, B. Van Tent, J. Varis, L. Vibert, M. Viel, P. Vielva, F. Villa, N. Vittorio, L. A. Wade, B. D. Wandelt, C. Watson, R. Watson, I. K. Wehus, N. Welikala, J. Weller, M. White, S. D. M. White, A. Wilkinson, B. Winkel, J.-Q. Xia, D. Yvon, A. Zacchei, J. P. Zibin, and A. Zonca. Planck 2013 results. i. overview of products and scientific results. *Astronomy & Astrophysics*, May 2014. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/201321529. URL <http://arxiv.org/abs/1303.5062>. arXiv: 1303.5062.
- C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, E. Komatsu, M. R. Nolte, L. Page, D. N. Spergel, E. Wollack,

- J. Dunkley, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Nine-year wilkinson microwave anisotropy probe (WMAP) observations: Final maps and results. *The Astrophysical Journal Supplement Series*, 208:20, October 2013. ISSN 0067-0049. doi: 10.1088/0067-0049/208/2/20. URL <http://adsabs.harvard.edu/abs/2013ApJS...208...20B>.
- G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R.olta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. Nine-year wilkinson microwave anisotropy probe (WMAP) observations: Cosmological parameter results. *The Astrophysical Journal Supplement Series*, 208:19, October 2013. ISSN 0067-0049. doi: 10.1088/0067-0049/208/2/19. URL <http://adsabs.harvard.edu/abs/2013ApJS...208...19H>.
- L. Bleem, P. Ade, K. Aird, J. Austermann, J. Beall, D. Becker, B. Benson, J. Britton, J. Carlstrom, C. L. Chang, H. Cho, T. de Haan, T. Crawford, A. Crites, A. Datesman, M. Dobbs, W. Everett, A. Ewall-Wice, E. George, N. Halverson, N. Harrington, J. Henning, G. Hilton, W. Holzappel, S. Hoover, J. Hubmayr, K. Irwin, R. Keisler, J. Kennedy, A. Lee, E. Leitch, D. Li, M. Lueker, D. P. Marrone, J. McMahon, J. Mehl, S. Meyer, J. Montgomery, T. Montroy, T. Natoli, J. Nibarger, M. Niemack, V. Novosad, S. Padin, C. Pryke, C. Reichardt, J. Ruhl, B. Saliwanchik, J. Sayre, K. Schafer, E. Shirokoff, K. Story, K. Vanderlinde, J. Vieira, G. Wang, R. Williamson, V. Yefremenko, K. W. Yoon, and E. Young. An overview of the SPTpol experiment. *J Low Temp Phys*, 167(5-6):859–864, June 2012. ISSN 0022-2291, 1573-7357. doi: 10.1007/s10909-012-0505-y. URL <http://link.springer.com/article/10.1007/s10909-012-0505-y>.
- Z. Hou, C. L. Reichardt, K. T. Story, B. Follin, R. Keisler, K. A. Aird, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H.-M. Cho, T. M. Crawford, A. T. Crites, T. de Haan, R. de Putter, M. A. Dobbs, S. Dodelson, J. Dudley, E. M. George, N. W. Halverson, G. P. Holder, W. L. Holzappel, S. Hoover, J. D. Hrubes, M. Joy, L. Knox, A. T. Lee, E. M. Leitch, M. Lueker, D. Luong-Van, J. J. McMahon, J. Mehl, S. S. Meyer, M. Millea, J. J. Mohr, T. E. Montroy, S. Padin, T. Plagge, C. Pryke, J. E. Ruhl, J. T. Sayre, K. K. Schaffer, L. Shaw, E. Shirokoff, H. G. Spieler, Z. Staniszewski, A. A. Stark, A. van Engelen, K. Vanderlinde, J. D. Vieira, R. Williamson, and O. Zahn. Constraints on cosmology from the cosmic microwave background power spectrum of the 2500 deg<sup>2</sup> SPT-SZ survey. *The Astrophysical Journal*, 782:74, February 2014. ISSN 0004-637X. doi: 10.1088/0004-637X/782/2/74. URL <http://adsabs.harvard.edu/abs/2014ApJ...782...74H>.
- Sudeep Das, Tobias A. Marriage, Peter A. R. Ade, Paula Aguirre, Mandana Amiri, John W. Appel, L. Felipe Barrientos, Elia S. Battistelli, John R. Bond, Ben Brown, Bryce Burger, Jay Chervenak, Mark J. Devlin, Simon R. Dicker, W. Bertrand Doriese, Joanna Dunkley, Rolando Dünner, Thomas Essinger-Hileman, Ryan P. Fisher, Joseph W. Fowler, Amir Hajian, Mark Halpern, Matthew Hasselfield, Carlos Hernández-Monteagudo, Gene C. Hilton, Matt Hilton, Adam D. Hincks, Renée Hlozek, Kevin M. Huffenberger, David H. Hughes, John P. Hughes, Leopoldo Infante, Kent D. Irwin, Jean Baptiste Juin, Madhuri Kaul, Jeff

- Klein, Arthur Kosowsky, Judy M. Lau, Michele Limon, Yen-Ting Lin, Robert H. Lupton, Danica Marsden, Krista Martocci, Phil Mausekopf, Felipe Menanteau, Kavilan Moodley, Harvey Moseley, Calvin B. Netterfield, Michael D. Niemack, Michael R. Nolte, Lyman A. Page, Lucas Parker, Bruce Partridge, Beth Reid, Neelima Sehgal, Blake D. Sherwin, Jon Sievers, David N. Spergel, Suzanne T. Staggs, Daniel S. Swetz, Eric R. Switzer, Robert Thornton, Hy Trac, Carole Tucker, Ryan Warne, Ed Wollack, and Yue Zhao. The atacama cosmology telescope: A measurement of the cosmic microwave background power spectrum at 148 and 218 GHz from the 2008 southern survey. *The Astrophysical Journal*, 729:62, March 2011. ISSN 0004-637X. doi: 10.1088/0004-637X/729/1/62. URL <http://adsabs.harvard.edu/abs/2011ApJ...729...62D>.
- D. J. Fixsen. The temperature of the cosmic microwave background. *The Astrophysical Journal*, 707:916–920, December 2009. ISSN 0004-637X. doi: 10.1088/0004-637X/707/2/916. URL <http://adsabs.harvard.edu/abs/2009ApJ...707..916F>.
- Bruce A Bassett and Renée Hlozek. Baryon acoustic oscillations. *arXiv:0910.5224*, October 2009. URL <http://arxiv.org/abs/0910.5224>. Dark Energy, Ed. P. Ruiz-Lapuente (2010, ISBN-13: 9780521518888).
- Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, E. Battaner, K. Benabed, A. Benoît, A. Benoit-Lévy, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. Bobin, J. J. Bock, A. Bonaldi, J. R. Bond, J. Borrill, F. R. Bouchet, M. Bridges, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, B. Cappellini, J.-F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R.-R. Chary, X. Chen, H. C. Chiang, L.-Y. Chiang, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, J.-M. Delouis, F.-X. Désert, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, T. C. Gaier, S. Galeotta, S. Galli, K. Ganga, M. Giard, G. Giardino, Y. Giraud-Héraud, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Haissinski, J. Hamann, F. K. Hansen, D. Hanson, D. Harrison, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, Z. Hou, W. Hovest, K. M. Huffenberger, A. H. Jaffe, T. R. Jaffe, J. Jewell, W. C. Jones, M. Juvela, E. Keihänen, R. Kesitalo, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J.-M. Lamarre, A. Lasenby, M. Lattanzi, R. J. Laureijs, C. R. Lawrence, S. Leach, J. P. Leahy, R. Leonardi, J. León-Tavares, J. Lesgourgues, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Cañiego, P. M. Lubin, J. F. Macías-Pérez, B. Maffei, D. Maino, N. Mandolesi, M. Maris, D. J. Marshall, P. G. Martin, E. Martínez-González, S. Masi, M. Massardi, S. Matarrese, F. Matthai, P. Mazzotta, P. R. Meinhold, A. Melchiorri, J.-B. Melin, L. Mendes, E. Menegoni, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M.-A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen,

- F. Noviello, D. Novikov, I. Novikov, I. J. O'Dwyer, S. Osborne, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, D. Pearson, T. J. Pearson, H. V. Peiris, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, P. Platania, E. Pointecouteau, G. Polenta, N. Ponthieu, L. Popa, T. Poutanen, G. W. Pratt, G. Prézeau, S. Prunet, J.-L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, S. Ricciardi, T. Riller, I. Ristorcelli, G. Rocha, C. Rosset, G. Roudier, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, E. P. S. Shellard, L. D. Spencer, J.-L. Starck, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, F. Sureau, D. Sutton, A.-S. Suur-Uski, J.-F. Sygnet, J. A. Tauber, D. Tavagnacco, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, M. Tucci, J. Tuovinen, M. Türler, G. Umama, L. Valenziano, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, N. Vittorio, L. A. Wade, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Wilkinson, D. Yvon, A. Zacchei, and A. Zonca. Planck 2013 results. XVI. cosmological parameters. *ArXiv e-prints*, 1303:5076, March 2013. URL <http://adsabs.harvard.edu/abs/2013arXiv1303.5076P>.
- Michael Rauch. The Lyman alpha forest in the spectra of quasistellar objects. *Annual Review of Astronomy and Astrophysics*, 36(1):267–316, 1998. doi: 10.1146/annurev.astro.36.1.267. URL <http://www.annualreviews.org/doi/abs/10.1146/annurev.astro.36.1.267>.
- D. S. Womble, W. L. W. Sargent, and R. S. Lyons. Heavy elements in the Lyman- $\alpha$  forest: Abundances and clustering at  $z = 3$ . volume 206, page 249, eprint: arXiv:astro-ph/9511035, 1996. URL <http://adsabs.harvard.edu/abs/1996ASSL..206..249W>.
- Jesse L. Greenstein and Maarten Schmidt. The quasi-stellar radio sources 3c 48 and 3c 273. *The Astrophysical Journal*, 140:1, July 1964. ISSN 0004-637X. doi: 10.1086/147889. URL <http://adsabs.harvard.edu/abs/1964ApJ...140....1G>.
- Demosthenes Kazanas, Keigo Fukumura, Ehud Behar, Ioannis Contopoulos, and Chris Shrader. Toward a unified AGN structure. *The Astronomical Review*, 7:92–123, July 2012. URL <http://adsabs.harvard.edu/abs/2012AstRv...7c..92K>.
- P. Goldreich and N. Scoville. OH-IR stars. I - physical properties of circumstellar envelopes. *The Astrophysical Journal*, 205:144–154, April 1976. ISSN 0004-637X. doi: 10.1086/154257. URL <http://adsabs.harvard.edu/abs/1976ApJ...205..144G>.
- J. Bergeron. The MG II absorption system in the QSO PKS 2128-12 - a galaxy disc/halo with a radius of 65 KPC. *Astronomy and Astrophysics*, 155:L8–L11, January 1986. ISSN 0004-6361. URL <http://adsabs.harvard.edu/abs/1986A%26A...155L...8B>.
- P. Petitjean and J. Bergeron. CIV QSO absorption systems and properties of galactic haloes at high redshift. *Astronomy and Astrophysics*, 283:759, March 1994. ISSN 0004-6361. URL <http://adsabs.harvard.edu/abs/1994A%26A...283..759P>.
- James E. Gunn and Bruce A. Peterson. On the density of neutral hydrogen in intergalactic space. *The Astrophysical Journal*, 142:1633–1641, November 1965. ISSN 0004-637X. doi: 10.1086/148444. URL <http://adsabs.harvard.edu/abs/1965ApJ...142.1633G>.

- John L. Lowrance, Donald C. Morton, Paul Zucchino, J. B. Oke, and Maarten Schmidt. The spectrum of the quasi-stellar object PHL 957. *The Astrophysical Journal*, 171:233, February 1972. ISSN 0004-637X. doi: 10.1086/151275. URL <http://adsabs.harvard.edu/abs/1972ApJ...171..233L>.
- E. A. Beaver, E. M. Burbidge, C. E. McIlwain, H. W. Epps, and P. A. Strittmatter. Digicon spectrophotometry of the quasi-stellar object PHL 957. *The Astrophysical Journal*, 178:95–104, November 1972. ISSN 0004-637X. doi: 10.1086/151768. URL <http://adsabs.harvard.edu/abs/1972ApJ...178...95B>.
- A. E. Wright, D. C. Morton, B. A. Peterson, and D. L. Jauncey. Spectroscopy of PKS 1157+014 - a QSO with no lyman alpha emission. *Monthly Notices of the Royal Astronomical Society*, 189:611–620, December 1979. ISSN 0035-8711. URL <http://adsabs.harvard.edu/abs/1979MNRAS.189..611W>.
- H. E. Smith, B. Margon, and M. Jura. On the abundances in the  $z = 2.81$  absorbing material toward the quasi-stellar object PKS 0528-250. *The Astrophysical Journal*, 228:369–374, March 1979. ISSN 0004-637X. doi: 10.1086/156854. URL <http://adsabs.harvard.edu/abs/1979ApJ...228..369S>.
- P. J. E. Peebles. Recombination of the primeval plasma. *The Astrophysical Journal*, 153:1, July 1968. ISSN 0004-637X. doi: 10.1086/149628. URL <http://adsabs.harvard.edu/abs/1968ApJ...153...1P>.
- Y. B. Zeldovich, V. G. Kurt, and R. A. Syunyaev. Recombination of hydrogen in the hot model of the universe. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, 55:278–286, July 1968. ISSN 0044-4510. URL <http://adsabs.harvard.edu/abs/1968ZhETF...55..278Z>.
- Francesco Haardt and Piero Madau. Radiative transfer in a clumpy universe. II. the ultraviolet extragalactic background. *The Astrophysical Journal*, 461:20, April 1996. ISSN 0004-637X. doi: 10.1086/177035. URL <http://adsabs.harvard.edu/abs/1996ApJ...461...20H>.
- Renyue Cen, Jordi Miralda-Escudé, Jeremiah P. Ostriker, and Michael Rauch. Gravitational collapse of small-scale structure as the origin of the lyman-alpha forest. *The Astrophysical Journal Letters*, 437:L9–L12, December 1994. ISSN 0004-637X. doi: 10.1086/187670;. URL <http://adsabs.harvard.edu/abs/1994ApJ...437L...9C>.
- Patrick Petitjean, J. P. Mueket, and Ronald E. Kates. The LY alpha forest at low redshift: Tracing the dark matter filaments. *Astronomy and Astrophysics*, 295:L9–L12, March 1995. ISSN 0004-6361. URL <http://adsabs.harvard.edu/abs/1995A%26A...295L...9P>.
- Jordi Miralda-Escudé, Renyue Cen, Jeremiah P. Ostriker, and Michael Rauch. The ly alpha forest from gravitational collapse in the cold dark matter + lambda model. *The Astrophysical Journal*, 471:582, November 1996. ISSN 0004-637X. doi: 10.1086/177992;. URL <http://adsabs.harvard.edu/abs/1996ApJ...471..582M>.
- Yu Zhang, Peter Anninos, and Michael L. Norman. A multispecies model for hydrogen and helium absorbers in lyman-alpha forest clouds. *The Astrophysical Journal Letters*, 453:L57, November 1995. ISSN 0004-637X. doi: 10.1086/309752;. URL <http://adsabs.harvard.edu/abs/1995ApJ...453L...57Z>.



- Lars Hernquist, Neal Katz, David H. Weinberg, and Jordi Miralda-Escudé. The Lyman- $\alpha$  forest in the cold dark matter model. *The Astrophysical Journal Letters*, 457:L51, February 1996. ISSN 0004-637X. doi: 10.1086/309899;. URL <http://adsabs.harvard.edu/abs/1996ApJ...457L..51H>.
- Neal Katz, David H. Weinberg, Lars Hernquist, and Jordi Miralda-Escudé. Damped Lyman- $\alpha$  and Lyman limit absorbers in the cold dark matter model. *The Astrophysical Journal Letters*, 457:L57, February 1996. ISSN 0004-637X. doi: 10.1086/309900. URL <http://adsabs.harvard.edu/abs/1996ApJ...457L..57K>.
- Martin G. Haehnelt, Michael Rauch, and Matthias Steinmetz. Non-equilibrium effects on line-of-sight size estimates of QSO absorption systems. *Monthly Notices of the Royal Astronomical Society*, 283:1055–1060, December 1996. ISSN 0035-8711. URL <http://adsabs.harvard.edu/abs/1996MNRAS.283.1055H>.
- Lam Hui and Nickolay Y. Gnedin. Equation of state of the photoionized intergalactic medium. *Monthly Notices of the Royal Astronomical Society*, 292:27, November 1997. ISSN 0035-8711. URL <http://adsabs.harvard.edu/abs/1997MNRAS.292...27H>.
- Nick Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227:1–21, July 1987. URL <http://adsabs.harvard.edu/abs/1987MNRAS.227...1K>.
- Uroš Seljak. Bias, redshift space distortions and primordial nongaussianity of nonlinear transformations: application to Ly- $\alpha$  forest. *Journal of Cosmology and Astro-Particle Physics*, 03:004, March 2012. ISSN 1475-7516. doi: 10.1088/1475-7516/2012/03/004. URL <http://adsabs.harvard.edu/abs/2012JCAP...03..004S>.
- Patrick McDonald. 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. *The Astrophysical Journal*, 585:34–51, March 2003. doi: 10.1086/345945;. URL <http://adsabs.harvard.edu/abs/2003ApJ...585...34M>.
- W. L. W. Sargent, P. J. Young, A. Boksenberg, and D. Tytler. The distribution of Lyman- $\alpha$  absorption lines in the spectra of six QSOs - evidence for an intergalactic origin. *The Astrophysical Journal Supplement Series*, 42:41–81, January 1980. ISSN 0067-0049. doi: 10.1086/190644. URL <http://adsabs.harvard.edu/abs/1980ApJS...42...41S>.
- W. Webb and M. Malkan. Spectrum and polarization of the continuum from accretion disks in active galaxies and quasars. volume 266, page 15, 1986. ISBN 0075-8450. doi: 10.1007/3-540-17195-9\_2. URL <http://adsabs.harvard.edu/abs/1986LNP...266...15W>.
- J. P. Muecket and V. Mueller. The clustering of Ly- $\alpha$  absorption lines in the QSO spectra. *Astrophysics and Space Science*, 139:163–174, December 1987. ISSN 0004-640X. doi: 10.1007/BF00643822. URL <http://adsabs.harvard.edu/abs/1987Ap%26SS.139..163M>.
- Rupert A. C. Croft, David H. Weinberg, Max Pettini, Lars Hernquist, and Neal Katz. The power spectrum of mass fluctuations measured from the Ly $\alpha$  forest at redshift  $z=2.5$ . *The Astrophysical Journal*, 520:1–23, July 1999. ISSN 0004-637X. doi: 10.1086/307438;. URL <http://adsabs.harvard.edu/abs/1999ApJ...520...1C>.

- Patrick McDonald, Jordi Miralda-Escudé, Michael Rauch, Wallace L. W. Sargent, Tom A. Barlow, Renyue Cen, and Jeremiah P. Ostriker. The observed probability distribution function, power spectrum, and correlation function of the transmitted flux in the Ly $\alpha$  forest. *The Astrophysical Journal*, 543:1–23, November 2000. ISSN 0004-637X. doi: 10.1086/317079;. URL <http://adsabs.harvard.edu/abs/2000ApJ...543....1M>.
- Rupert A. C. Croft, David H. Weinberg, Mike Bolte, Scott Burles, Lars Hernquist, Neal Katz, David Kirkman, and David Tytler. Toward a precise measurement of matter clustering: Ly $\alpha$  forest data at redshifts 2–4. *The Astrophysical Journal*, 581:20–52, December 2002. ISSN 0004-637X. doi: 10.1086/344099;. URL <http://adsabs.harvard.edu/abs/2002ApJ...581...20C>.
- Matteo Viel, Martin G. Haehnelt, and Volker Springel. Inferring the dark matter power spectrum from the Lyman  $\alpha$  forest in high-resolution QSO absorption spectra. *Monthly Notices of the Royal Astronomical Society*, 354:684–694, November 2004. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2004.08224.x;. URL <http://adsabs.harvard.edu/abs/2004MNRAS.354..684V>.
- Patrick McDonald, Uroš Seljak, Scott Burles, David J. Schlegel, David H. Weinberg, Renyue Cen, David Shih, Joop Schaye, Donald P. Schneider, Neta A. Bahcall, John W. Briggs, J. Brinkmann, Robert J. Brunner, Masataka Fukugita, James E. Gunn, Željko Ivezić, Stephen Kent, Robert H. Lupton, and Daniel E. Vanden Berk. The Ly $\alpha$  forest power spectrum from the Sloan Digital Sky Survey. *The Astrophysical Journal Supplement Series*, 163:80–109, March 2006. doi: 10.1086/444361;. URL <http://adsabs.harvard.edu/abs/2006ApJS...163...80M>.
- Nathalie Palanque-Delabrouille, Christophe Yèche, Arnaud Borde, Jean-Marc Le Goff, Graziano Rossi, Matteo Viel, Éric Aubourg, Stephen Bailey, Julian Bautista, Michael Blomqvist, Adam Bolton, James S. Bolton, Nicolás G. Busca, Bill Carithers, Rupert A. C. Croft, Kyle S. Dawson, Timothée Delubac, Andreu Font-Ribera, Shirley Ho, David Kirkby, Khee-Gan Lee, Daniel Margala, Jordi Miralda-Escudé, Demitri Muna, Adam D. Myers, Pasquier Noterdaeme, Isabelle Pâris, Patrick Petitjean, Matthew M. Pieri, James Rich, Emmanuel Rollinde, Nicholas P. Ross, David J. Schlegel, Donald P. Schneider, Anže Slosar, and David H. Weinberg. The one-dimensional Ly $\alpha$  forest power spectrum from BOSS. arXiv e-print 1306.5896, June 2013. URL <http://arxiv.org/abs/1306.5896>.
- Patrick McDonald and Daniel J. Eisenstein. Dark energy and curvature from a future baryonic acoustic oscillation survey using the Lyman- $\alpha$  forest. *Physical Review D*, 76:63009, September 2007. ISSN 1550-7998. doi: 10.1103/PhysRevD.76.063009. URL <http://adsabs.harvard.edu/abs/2007PhRvD...76f3009M>.
- Matthew McQuinn and Martin White. On estimating Ly $\alpha$  forest correlations between multiple sightlines. *Monthly Notices of the Royal Astronomical Society*, 415:2257–2269, August 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.18855.x;. URL <http://adsabs.harvard.edu/abs/2011MNRAS.415.2257M>.
- Anže Slosar, Andreu Font-Ribera, Matthew M Pieri, James Rich, Jean-Marc Le Goff, Éric Aubourg, Jon Brinkmann, Nicolas Busca, Bill Carithers, Romain Charlassier, Marina Cortés, Rupert Croft, Kyle S Dawson, Daniel Eisenstein, Jean-Christophe Hamilton, Shirley Ho, Khee-Gan Lee, Robert Lupton, Patrick McDonald, Bumbarija Medolin, Jordi Miralda-Escudé, Demitri Muna, Adam D Myers, Robert C Nichol, Nathalie Palanque-Delabrouille,



- Isabelle Pâris, Patrick Petitjean, Yodovina Piškur, Emmanuel Rollinde, Nicholas P Ross, David J Schlegel, Donald P Schneider, Erin Sheldon, Benjamin A Weaver, David H Weinberg, Christophe Yèche, and Donald G York. The Lyman-alpha forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data. *arXiv:1104.5244*, April 2011. URL <http://arxiv.org/abs/1104.5244>.
- N. G. Busca, T. Delubac, J. Rich, S. Bailey, A. Font-Ribera, D. Kirkby, J.-M. Le Goff, M. M. Pieri, A. Slosar, É. Aubourg, J. E. Bautista, D. Bizyaev, M. Blomqvist, A. S. Bolton, J. Bovy, H. Brewington, A. Borde, J. Brinkmann, B. Carithers, R. A. C. Croft, K. S. Dawson, G. Ebelke, D. J. Eisenstein, J.-C. Hamilton, S. Ho, D. W. Hogg, K. Honscheid, K.-G. Lee, B. Lundgren, E. Malanushenko, V. Malanushenko, D. Margala, C. Maraston, K. Mehta, J. Miralda-Escudé, A. D. Myers, R. C. Nichol, P. Noterdaeme, M. D. Olmstead, D. Oravetz, N. Palanque-Delabrouille, K. Pan, I. Pâris, W. J. Percival, P. Petitjean, N. A. Roe, E. Rollinde, N. P. Ross, G. Rossi, D. J. Schlegel, D. P. Schneider, A. Sheldon, E. S. Sheldon, A. Simmons, S. Snedden, J. L. Tinker, M. Viel, B. A. Weaver, D. H. Weinberg, M. White, C. Yèche, and D. G. York. Baryon acoustic oscillations in the Lyman-alpha forest of BOSS quasars. *Astronomy and Astrophysics*, 552:96, April 2013. ISSN 0004-6361. doi: 10.1051/0004-6361/201220724;. URL <http://adsabs.harvard.edu/abs/2013A%26A...552A..96B>.
- Anže Slosar, Vid Iršič, David Kirkby, Stephen Bailey, Nicolás G. Busca, Timothée Delubac, James Rich, Éric Aubourg, Julian E. Bautista, Vaishali Bhardwaj, Michael Blomqvist, Adam S. Bolton, Jo Bovy, Joel Brownstein, Bill Carithers, Rupert A. C. Croft, Kyle S. Dawson, Andreu Font-Ribera, J.-M. Le Goff, Shirley Ho, Klaus Honscheid, Khee-Gan Lee, Daniel Margala, Patrick McDonald, Bumbarija Medolin, Jordi Miralda-Escudé, Adam D. Myers, Robert C. Nichol, Pasquier Noterdaeme, Nathalie Palanque-Delabrouille, Isabelle Pâris, Patrick Petitjean, Matthew M. Pieri, Yodovina Piškur, Natalie A. Roe, Nicholas P. Ross, Graziano Rossi, David J. Schlegel, Donald P. Schneider, Nao Suzuki, Erin S. Sheldon, Uroš Seljak, Matteo Viel, David H. Weinberg, and Christophe Yèche. Measurement of baryon acoustic oscillations in the Lyman-alpha forest fluctuations in BOSS data release 9. *Journal of Cosmology and Astro-Particle Physics*, 04:026, April 2013. ISSN 1475-7516. doi: 10.1088/1475-7516/2013/04/026;. URL <http://adsabs.harvard.edu/abs/2013JCAP...04..026S>.
- David Kirkby, Daniel Margala, Anže Slosar, Stephen Bailey, Nicolás G. Busca, Timothée Delubac, James Rich, Julian E. Bautista, Michael Blomqvist, Joel R. Brownstein, Bill Carithers, Rupert A. C. Croft, Kyle S. Dawson, Andreu Font-Ribera, Jordi Miralda-Escudé, Adam D. Myers, Robert C. Nichol, Nathalie Palanque-Delabrouille, Isabelle Pâris, Patrick Petitjean, Graziano Rossi, David J. Schlegel, Donald P. Schneider, Matteo Viel, David H. Weinberg, and Christophe Yèche. Fitting methods for baryon acoustic oscillations in the Lyman-alpha forest fluctuations in BOSS data release 9. *Journal of Cosmology and Astro-Particle Physics*, 03:024, March 2013. ISSN 1475-7516. doi: 10.1088/1475-7516/2013/03/024;. URL <http://adsabs.harvard.edu/abs/2013JCAP...03..024K>.
- Kyle S. Dawson, David J. Schlegel, Christopher P. Ahn, Scott F. Anderson, Éric Aubourg, Stephen Bailey, Robert H. Barkhouser, Julian E. Bautista, Alessandra Beifiori, Andreas A. Berlind, Vaishali Bhardwaj, Dmitry Bizyaev, Cullen H. Blake, Michael R. Blanton, Michael Blomqvist, Adam S. Bolton, Arnaud Borde, Jo Bovy, W. N. Brandt, Howard Brewington,

- Jon Brinkmann, Peter J. Brown, Joel R. Brownstein, Kevin Bundy, N. G. Busca, William Carithers, Aurelio R. Carnero, Michael A. Carr, Yanmei Chen, Johan Comparat, Natalia Connolly, Frances Cope, Rupert A. C. Croft, Antonio J. Cuesta, Luiz N. da Costa, James R. A. Davenport, Timothée Delubac, Roland de Putter, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Daniel J. Eisenstein, S. Escoffier, Xiaohui Fan, N. Filiz Ak, Hayley Finley, Andreu Font-Ribera, R. Génova-Santos, James E. Gunn, Hong Guo, Daryl Haggard, Patrick B. Hall, Jean-Christophe Hamilton, Ben Harris, David W. Harris, Shirley Ho, David W. Hogg, Diana Holder, Klaus Honscheid, Joe Huehnerhoff, Beatrice Jordan, Wendell P. Jordan, Guinevere Kauffmann, Eyal A. Kazin, David Kirkby, Mark A. Klaene, Jean-Paul Kneib, Jean-Marc Le Goff, Khee-Gan Lee, Daniel C. Long, Craig P. Loomis, Britt Lundgren, Robert H. Lupton, Marcio A. G. Maia, Martin Makler, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Marc Manera, Claudia Maraston, Daniel Margala, Karen L. Masters, Cameron K. McBride, Patrick McDonald, Ian D. McGreer, Richard G. McMahon, Olga Mena, Jordi Miralda-Escudé, Antonio D. Montero-Dorta, Francesco Montesano, Demitri Muna, Adam D. Myers, Tracy Naugle, Robert C. Nichol, Pasquier Noterdaeme, Sebastián E. Nuza, Matthew D. Olmstead, Audrey Oravetz, Daniel J. Oravetz, Russell Owen, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, John K. Parejko, Isabelle Pâris, Will J. Percival, Ismael Pérez-Fournon, Ignasi Pérez-Ràfols, Patrick Petitjean, Robert Pfaffenberger, Janine Pforr, Matthew M. Pieri, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Rafael Rebolo, James Rich, Gordon T. Richards, Constance M. Rockosi, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, J. A. Rubiño-Martin, Lado Samushia, Ariel G. Sánchez, Conor Sayres, Sarah J. Schmidt, Donald P. Schneider, C. G. Scóccola, Hee-Jong Seo, Alaina Shelden, Erin Sheldon, Yue Shen, Yiping Shu, Anže Slosar, Stephen A. Smee, Stephanie A. Snedden, Fritz Stauffer, Oliver Steele, Michael A. Strauss, Alina Streblyanska, Nao Suzuki, Molly E. C. Swanson, Tomer Tal, Masayuki Tanaka, Daniel Thomas, Jeremy L. Tinker, Rita Tojeiro, Christy A. Tremonti, M. Vargas Magaña, Licia Verde, Matteo Viel, David A. Wake, Mike Watson, Benjamin A. Weaver, David H. Weinberg, Benjamin J. Weiner, Andrew A. West, Martin White, W. M. Wood-Vasey, Christophe Yèche, Idit Zehavi, Gong-Bo Zhao, and Zheng Zheng. The baryon oscillation spectroscopic survey of SDSS-III. *The Astronomical Journal*, 145:10, January 2013. ISSN 0004-6256. doi: 10.1088/0004-6256/145/1/10. URL <http://adsabs.harvard.edu/abs/2013AJ....145...10D>.
- Daniel J. Eisenstein, David H. Weinberg, Eric Agol, Hiroaki Aihara, Carlos Allende Prieto, Scott F. Anderson, James A. Arns, Éric Aubourg, Stephen Bailey, Eduardo Balbinot, Robert Barkhouser, Timothy C. Beers, Andreas A. Berlind, Steven J. Bickerton, Dmitry Bizyaev, Michael R. Blanton, John J. Bochanski, Adam S. Bolton, Casey T. Bosman, Jo Bovy, W. N. Brandt, Ben Breslauer, Howard J. Brewington, J. Brinkmann, Peter J. Brown, Joel R. Brownstein, Dan Burger, Nicolas G. Busca, Heather Campbell, Phillip A. Cargile, William C. Carithers, Joleen K. Carlberg, Michael A. Carr, Liang Chang, Yanmei Chen, Cristina Chiappini, Johan Comparat, Natalia Connolly, Marina Cortes, Rupert A. C. Croft, Katia Cunha, Luiz N. da Costa, James R. A. Davenport, Kyle Dawson, Nathan De Lee, Gustavo F. Porto de Mello, Fernando de Simoni, Janice Dean, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Edward M. Edmondson, Jacob M. Eiting, Stephanie Escoffier, Massimiliano Esposito, Michael L. Evans, Xiaohui Fan, Bruno Femenía Castellá, Leticia Dutra Ferreira, Greg Fitzgerald, Scott W.

- Fleming, Andreu Font-Ribera, Eric B. Ford, Peter M. Frinchaboy, Ana Elia García Pérez, B. Scott Gaudi, Jian Ge, Luan Ghezzi, Bruce A. Gillespie, G. Gilmore, Léo Girardi, J. Richard Gott, Andrew Gould, Eva K. Grebel, James E. Gunn, Jean-Christophe Hamilton, Paul Harding, David W. Harris, Suzanne L. Hawley, Frederick R. Hearty, Joseph F. Hennawi, Jonay I. González Hernández, Shirley Ho, David W. Hogg, Jon A. Holtzman, Klaus Honscheid, Naohisa Inada, Inese I. Ivans, Linhua Jiang, Peng Jiang, Jennifer A. Johnson, Cathy Jordan, Wendell P. Jordan, Guinevere Kauffmann, Eyal Kazin, David Kirkby, Mark A. Klaene, G. R. Knapp, Jean-Paul Kneib, C. S. Kochanek, Lars Koesterke, Juna A. Kollmeier, Richard G. Kron, Hubert Lampeitl, Dustin Lang, James E. Lawler, Jean-Marc Le Goff, Brian L. Lee, Young Sun Lee, Jarron M. Leisenring, Yen-Ting Lin, Jian Liu, Daniel C. Long, Craig P. Loomis, Sara Lucatello, Britt Lundgren, Robert H. Lupton, Bo Ma, Zhibo Ma, Nicholas MacDonald, Claude Mack, Suvrath Mahadevan, Marcio A. G. Maia, Steven R. Majewski, Martin Makler, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Claudia Maraston, Daniel Margala, Paul Maseman, Karen L. Masters, Cameron K. McBride, Patrick McDonald, Ian D. McGreer, Richard G. McMahon, Olga Mena Requejo, Brice Ménard, Jordi Miralda-Escudé, Heather L. Morrison, Fergal Mullally, Demitri Muna, Hitoshi Murayama, Adam D. Myers, Tracy Naugle, Angelo Fausti Neto, Duy Cuong Nguyen, Robert C. Nichol, David L. Nidever, Robert W. O'Connell, Ricardo L. C. Ogando, Matthew D. Olmstead, Daniel J. Oravetz, Nikhil Padmanabhan, Martin Paegert, Nathalie Palanque-Delabrouille, Kaike Pan, Parul Pandey, John K. Parejko, Isabelle Pâris, Paulo Pellegrini, Joshua Pepper, Will J. Percival, Patrick Petitjean, Robert Pfaffenberger, Janine Pforr, Stefanie Phleps, Christophe Pichon, Matthew M. Pieri, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Beatriz H. F. Ramos, I. Neill Reid, Celine Reyle, James Rich, Gordon T. Richards, George H. Rieke, Marcia J. Rieke, Hans-Walter Rix, Annie C. Robin, Helio J. Rocha-Pinto, Constance M. Rockosi, Natalie A. Roe, Emmanuel Rollinde, Ashley J. Ross, Nicholas P. Ross, Bruno Rossetto, Ariel G. Sánchez, Basilio Santiago, Conor Sayres, Ricardo Schiavon, David J. Schlegel, Katharine J. Schlesinger, Sarah J. Schmidt, Donald P. Schneider, Kris Sellgren, Alaina Shelden, Erin Sheldon, Matthew Shetrone, Yiping Shu, John D. Silverman, Jennifer Simmerer, Audrey E. Simmons, Thirupathi Sivarani, M. F. Skrutskie, Anže Slosar, Stephen Smee, Verne V. Smith, Stephanie A. Snedden, Keivan G. Stassun, Oliver Steele, Matthias Steinmetz, Mark H. Stockett, Todd Stollberg, Michael A. Strauss, Alexander S. Szalay, Masayuki Tanaka, Aniruddha R. Thakar, Daniel Thomas, Jeremy L. Tinker, Benjamin M. Tofflemire, Rita Tojeiro, Christy A. Tremonti, Mariana Vargas Magaña, Licia Verde, Nicole P. Vogt, David A. Wake, Xiaoke Wan, Ji Wang, Benjamin A. Weaver, Martin White, Simon D. M. White, John C. Wilson, John P. Wisniewski, W. Michael Wood-Vasey, Brian Yanny, Naoki Yasuda, Christophe Yèche, Donald G. York, Erick Young, Gail Zasowski, Idit Zehavi, and Bo Zhao. SDSS-III: Massive spectroscopic surveys of the distant universe, the milky way, and extra-solar planetary systems. *The Astronomical Journal*, 142:72, September 2011. ISSN 0004-6256. doi: 10.1088/0004-6256/142/3/72;. URL <http://adsabs.harvard.edu/abs/2011AJ....142...72E>.
- D. J Eisenstein, I. Zehavi, D. W Hogg, R. Scoccimarro, M. R Blanton, R. C Nichol, R. Scranton, H. Seo, M. Tegmark, Z. Zheng, S. Anderson, J. Annis, N. Bahcall, J. Brinkmann, S. Burles, F. J Castander, A. Connolly, I. Csabai, M. Doi, M. Fukugita, J. A Frieman, K. Glazebrook, J. E Gunn, J. S Hendry, G. Hennessy, Z. Ivezić, S. Kent, G. R Knapp, H. Lin, Y. Loh, R. H

- Lupton, B. Margon, T. McKay, A. Meiksin, J. A. Munn, A. Pope, M. Richmond, D. Schlegel, D. Schneider, K. Shimasaku, C. Stoughton, M. Strauss, M. SubbaRao, A. S Szalay, I. Szapudi, D. Tucker, B. Yanny, and D. York. Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies. *arXiv:astro-ph/0501171*, January 2005. doi: 10.1086/466512. URL <http://arxiv.org/abs/astro-ph/0501171>. *Astrophys.J.*633:560-574,2005.
- Shaun Cole, Will J. Percival, John A. Peacock, Peder Norberg, Carlton M. Baugh, Carlos S. Frenk, Ivan Baldry, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, Matthew Colless, Chris Collins, Warrick Couch, Nicholas J. G. Cross, Gavin Dalton, Vincent R. Eke, Roberto De Propriis, Simon P. Driver, George Efstathiou, Richard S. Ellis, Karl Glazebrook, Carole Jackson, Adrian Jenkins, Ofer Lahav, Ian Lewis, Stuart Lumsden, Steve Maddox, Darren Madgwick, Bruce A. Peterson, Will Sutherland, and Keith Taylor. The 2df galaxy redshift survey: power-spectrum analysis of the final data set and cosmological implications. *Monthly Notices of the Royal Astronomical Society*, 362:505–534, September 2005. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2005.09318.x. URL <http://adsabs.harvard.edu/abs/2005MNRAS.362..505C>.
- Hee-Jong Seo and Daniel J. Eisenstein. Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys. *The Astrophysical Journal*, 598:720–740, December 2003. ISSN 0004-637X. doi: 10.1086/379122. URL <http://adsabs.harvard.edu/abs/2003ApJ...598..720S>.
- James E. Gunn, Walter A. Siegmund, Edward J. Mannery, Russell E. Owen, Charles L. Hull, R. French Leger, Larry N. Carey, Gillian R. Knapp, Donald G. York, William N. Boroski, Stephen M. Kent, Robert H. Lupton, Constance M. Rockosi, Michael L. Evans, Patrick Waddell, John E. Anderson, James Annis, John C. Barentine, Larry M. Bartoszek, Steven Bastian, Stephen B. Bracker, Howard J. Brewington, Charles I. Briegel, Jon Brinkmann, Yorke J. Brown, Michael A. Carr, Paul C. Czarapata, Craig C. Drennan, Thomas Dombeck, Glenn R. Federwitz, Bruce A. Gillespie, Carlos Gonzales, Sten U. Hansen, Michael Harvanek, Jeffrey Hayes, Wendell Jordan, Ellyne Kinney, Mark Klaene, S. J. Kleinman, Richard G. Kron, Jurek Kresinski, Glenn Lee, Siriluk Limmongkol, Carl W. Lindenmeyer, Daniel C. Long, Craig L. Loomis, Peregrine M. McGehee, Paul M. Mantsch, Eric H. Neilsen, Jr., Richard M. Neswold, Peter R. Newman, Atsuko Nitta, John Peoples, Jr., Jeffrey R. Pier, Peter S. Prieto, Angela Prosapio, Claudio Rivetta, Donald P. Schneider, Stephanie Snedden, and Shu-i. Wang. The 2.5 m telescope of the sloan digital sky survey. *The Astronomical Journal*, 131:2332–2359, April 2006. ISSN 0004-6256. doi: 10.1086/500975. URL <http://adsabs.harvard.edu/abs/2006AJ...131.2332G>.
- J. E. Gunn, M. Carr, C. Rockosi, M. Sekiguchi, K. Berry, B. Elms, E. de Haas, Ž . Ivezić, G. Knapp, R. Lupton, G. Pauls, R. Simcoe, R. Hirsch, D. Sanford, S. Wang, D. York, F. Harris, J. Annis, L. Bartoszek, W. Boroski, J. Bakken, M. Haldeman, S. Kent, S. Holm, D. Holmgren, D. Petravick, A. Prosapio, R. Rechenmacher, M. Doi, M. Fukugita, K. Shimasaku, N. Okada, C. Hull, W. Siegmund, E. Mannery, M. Blouke, D. Heidtman, D. Schneider, R. Lucinio, and J. Brinkman. The sloan digital sky survey photometric camera. *The Astronomical Journal*,

116:3040–3081, December 1998. ISSN 0004-6256. doi: 10.1086/300645. URL <http://adsabs.harvard.edu/abs/1998AJ....116.3040G>.

Stephen A. Smee, James E. Gunn, Alan Uomoto, Natalie Roe, David Schlegel, Constance M. Rockosi, Michael A. Carr, French Leger, Kyle S. Dawson, Matthew D. Olmstead, Jon Brinkmann, Russell Owen, Robert H. Barkhouser, Klaus Honscheid, Paul Harding, Dan Long, Robert H. Lupton, Craig Loomis, Lauren Anderson, James Annis, Mariangela Bernardi, Vaishali Bhardwaj, Dmitry Bizyaev, Adam S. Bolton, Howard Brewington, John W. Briggs, Scott Burles, James G. Burns, Francisco Javier Castander, Andrew Connolly, James R. A. Davenport, Garrett Ebelke, Harland Epps, Paul D. Feldman, Scott D. Friedman, Joshua Frieman, Timothy Heckman, Charles L. Hull, Gillian R. Knapp, David M. Lawrence, Jon Loveday, Edward J. Mannery, Elena Malanushenko, Viktor Malanushenko, Aronne James Merrelli, Demitri Muna, Peter R. Newman, Robert C. Nichol, Daniel Oravetz, Kaike Pan, Adrian C. Pope, Paul G. Ricketts, Alaina Shelden, Dale Sandford, Walter Siegmund, Audrey Simmons, D. Shane Smith, Stephanie Snedden, Donald P. Schneider, Mark SubbaRao, Christy Tremonti, Patrick Waddell, and Donald G. York. The multi-object, fiber-fed spectrographs for the sloan digital sky survey and the baryon oscillation spectroscopic survey. *The Astrophysical Journal*, 146:32, August 2013. ISSN 0004-6256. doi: 10.1088/0004-6256/146/2/32. URL <http://adsabs.harvard.edu/abs/2013AJ....146...32S>.

Daniel J. Eisenstein, James Annis, James E. Gunn, Alexander S. Szalay, Andrew J. Connolly, R. C. Nichol, Neta A. Bahcall, Mariangela Bernardi, Scott Burles, Francisco J. Castander, Masataka Fukugita, David W. Hogg, Željko Ivezić, G. R. Knapp, Robert H. Lupton, Vijay Narayanan, Marc Postman, Daniel E. Reichart, Michael Richmond, Donald P. Schneider, David J. Schlegel, Michael A. Strauss, Mark SubbaRao, Douglas L. Tucker, Daniel Vanden Berk, Michael S. Vogeley, David H. Weinberg, and Brian Yanny. Spectroscopic target selection for the sloan digital sky survey: The luminous red galaxy sample. *The Astronomical Journal*, 122:2267–2280, November 2001. ISSN 0004-6256. doi: 10.1086/323717. URL <http://adsabs.harvard.edu/abs/2001AJ....122.2267E>.

Claudia Maraston, Janine Pforr, Bruno M. Henriques, Daniel Thomas, David Wake, Joel R. Brownstein, Diego Capozzi, Jeremy Tinker, Kevin Bundy, Ramin A. Skibba, Alessandra Beifiori, Robert C. Nichol, Edd Edmondson, Donald P. Schneider, Yanmei Chen, Karen L. Masters, Oliver Steele, Adam S. Bolton, Donald G. York, Benjamin A. Weaver, Tim Higgs, Dmitry Bizyaev, Howard Brewington, Elena Malanushenko, Viktor Malanushenko, Stephanie Snedden, Daniel Oravetz, Kaike Pan, Alaina Shelden, and Audrey Simmons. Stellar masses of SDSS-III/BOSS galaxies at  $z \sim 0.5$  and constraints to galaxy formation models. *Monthly Notices of the Royal Astronomical Society*, 435:2764–2792, November 2013. ISSN 0035-8711. doi: 10.1093/mnras/stt1424. URL <http://adsabs.harvard.edu/abs/2013MNRAS.435.2764M>.

John K. Parejko, Tomomi Sunayama, Nikhil Padmanabhan, David A. Wake, Andreas A. Berlind, Dmitry Bizyaev, Michael Blanton, Adam S. Bolton, Frank van den Bosch, Jon Brinkmann, Joel R. Brownstein, Luiz Alberto Nicolaci da Costa, Daniel J. Eisenstein, Hong Guo, Eyal Kazin, Marcio Maia, Elena Malanushenko, Claudia Maraston, Cameron K. McBride, Robert C. Nichol, Daniel J. Oravetz, Kaike Pan, Will J. Percival, Francisco Prada, Ashley J. Ross, Nicholas P. Ross, David J. Schlegel, Don Schneider, Audrey E. Simmons, Ramin Skibba,

- Jeremy Tinker, Rita Tojeiro, Benjamin A. Weaver, Andrew Wetzel, Martin White, David H. Weinberg, Daniel Thomas, Idit Zehavi, and Zheng Zheng. The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: the low-redshift sample. *Monthly Notices of the Royal Astronomical Society*, 429:98–112, February 2013. ISSN 0035-8711. doi: 10.1093/mnras/sts314. URL <http://adsabs.harvard.edu/abs/2013MNRAS.429...98P>.
- Rita Tojeiro, W. J Percival, J. Brinkmann, J. R Brownstein, D. Eisenstein, M. Manera, C. Maraston, C. K McBride, D. Duna, B. Reid, A. J Ross, N. P Ross, L. Samushia, N. Padmanabhan, D. P Schneider, R. Skibba, A. G Sanchez, M. E. C Swanson, D. Thomas, J. L Tinker, L. Verde, D. A Wake, B. A Weaver, and G. Zhao. The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: measuring structure growth using passive galaxies. *arXiv:1203.6565*, March 2012. URL <http://arxiv.org/abs/1203.6565>.
- Gordon T. Richards, Adam D. Myers, Alexander G. Gray, Ryan N. Riegel, Robert C. Nichol, Robert J. Brunner, Alexander S. Szalay, Donald P. Schneider, and Scott F. Anderson. Efficient photometric selection of quasars from the sloan digital sky survey. II.  $\sim 1,000,000$  quasars from data release 6. *The Astrophysical Journal Supplement Series*, 180:67–83, January 2009. ISSN 0067-0049. doi: 10.1088/0067-0049/180/1/67. URL <http://adsabs.harvard.edu/abs/2009ApJS...180...67R>.
- Jessica A. Kirkpatrick, David J. Schlegel, Nicholas P. Ross, Adam D. Myers, Joseph F. Hennawi, Erin S. Sheldon, Donald P. Schneider, and Benjamin A. Weaver. A simple likelihood method for quasar target selection. *The Astrophysical Journal*, 743:125, December 2011. ISSN 0004-637X. doi: 10.1088/0004-637X/743/2/125. URL <http://adsabs.harvard.edu/abs/2011ApJ...743..125K>.
- Jo Bovy, Joseph F. Hennawi, David W. Hogg, Adam D. Myers, Jessica A. Kirkpatrick, David J. Schlegel, Nicholas P. Ross, Erin S. Sheldon, Ian D. McGreer, Donald P. Schneider, and Benjamin A. Weaver. Think outside the color box: Probabilistic target selection and the SDSS-XDQSO quasar targeting catalog. *The Astrophysical Journal*, 729:141, March 2011. ISSN 0004-637X. doi: 10.1088/0004-637X/729/2/141. URL <http://adsabs.harvard.edu/abs/2011ApJ...729..141B>.
- Ch. Yèche, P. Petitjean, J. Rich, E. Aubourg, N. Busca, J.-Ch. Hamilton, J.-M. Le Goff, I. Paris, S. Peirani, Ch. Pichon, E. Rollinde, and M. Vargas-Magaña. Artificial neural networks for quasar selection and photometric redshift determination. *Astronomy and Astrophysics*, 523:14, November 2010. ISSN 0004-6361. doi: 10.1051/0004-6361/200913508. URL <http://adsabs.harvard.edu/abs/2010A%26A...523A..14Y>.
- Adam S. Bolton, David J. Schlegel, Éric Aubourg, Stephen Bailey, Vaishali Bhardwaj, Joel R. Brownstein, Scott Burles, Yan-Mei Chen, Kyle Dawson, Daniel J. Eisenstein, James E. Gunn, G. R. Knapp, Craig P. Loomis, Robert H. Lupton, Claudia Maraston, Demitri Muna,



- Adam D. Myers, Matthew D. Olmstead, Nikhil Padmanabhan, Isabelle Pâris, Will J. Percival, Patrick Petitjean, Constance M. Rockosi, Nicholas P. Ross, Donald P. Schneider, Yiping Shu, Michael A. Strauss, Daniel Thomas, Christy A. Tremonti, David A. Wake, Benjamin A. Weaver, and W. Michael Wood-Vasey. Spectral classification and redshift measurement for the SDSS-III baryon oscillation spectroscopic survey. *The Astronomical Journal*, 144:144, November 2012. ISSN 0004-6256. doi: 10.1088/0004-6256/144/5/144. URL <http://adsabs.harvard.edu/abs/2012AJ....144..144B>.
- I. Pâris, P. Petitjean, É. Aubourg, S. Bailey, N. P. Ross, A. D. Myers, M. A. Strauss, S. F. Anderson, E. Arnau, J. Bautista, D. Bizyaev, A. S. Bolton, J. Bovy, W. N. Brandt, H. Brewington, J. R. Browstein, N. Busca, D. Capellupo, W. Carithers, R. A. C. Croft, K. Dawson, T. Delubac, G. Ebelke, D. J. Eisenstein, P. Engelke, X. Fan, N. Filiz Ak, H. Finley, A. Font-Ribera, J. Ge, R. R. Gibson, P. B. Hall, F. Hamann, J. F. Hennawi, S. Ho, D. W. Hogg, Ž. Ivezić, L. Jiang, A. E. Kimball, D. Kirkby, J. A. Kirkpatrick, K.-G. Lee, J.-M. Le Goff, B. Lundgren, C. L. MacLeod, E. Malanushenko, V. Malanushenko, C. Maraston, I. D. McGreer, R. G. McMahon, J. Miralda-Escudé, D. Muna, P. Noterdaeme, D. Oravetz, N. Palanque-Delabrouille, K. Pan, I. Perez-Fournon, M. M. Pieri, G. T. Richards, E. Rollinde, E. S. Sheldon, D. J. Schlegel, D. P. Schneider, A. Slosar, A. Shelden, Y. Shen, A. Simmons, S. Snedden, N. Suzuki, J. Tinker, M. Viel, B. A. Weaver, D. H. Weinberg, M. White, W. M. Wood-Vasey, and C. Yèche. The sloan digital sky survey quasar catalog: ninth data release. *Astronomy and Astrophysics*, 548:66, December 2012. ISSN 0004-6361. doi: 10.1051/0004-6361/201220142;. URL <http://adsabs.harvard.edu/abs/2012A%26A...548A..66P>.
- M. Vargas-Magaña, J. E. Bautista, J.-Ch. Hamilton, N. G. Busca, É. Aubourg, A. Labatie, J.-M. Le Goff, S. Escoffier, M. Manera, C. K. McBride, D. P. Schneider, and Ch. N. A. Willmer. An optimized correlation function estimator for galaxy surveys. *Astronomy and Astrophysics*, 554:131, June 2013. ISSN 0004-6361. doi: 10.1051/0004-6361/201220790;. URL <http://adsabs.harvard.edu/abs/2013A%26A...554A.131V>.
- J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The statistics of peaks of gaussian random fields. *The Astrophysical Journal*, 304:15–61, May 1986. ISSN 0004-637X. doi: 10.1086/164143. URL <http://adsabs.harvard.edu/abs/1986ApJ...304...15B>.
- Shaun Cole and Nick Kaiser. Biased clustering in the cold dark matter cosmogony. *Monthly Notices of the Royal Astronomical Society*, 237:1127–1146, April 1989. ISSN 0035-8711. URL <http://adsabs.harvard.edu/abs/1989MNRAS.237.1127C>.
- Uroš Seljak. Analytic model for galaxy and dark matter clustering. *Monthly Notices of the Royal Astronomical Society*, 318:203–213, October 2000. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2000.03715.x. URL <http://adsabs.harvard.edu/abs/2000MNRAS.318..203S>.
- J. A. Peacock and R. E. Smith. Halo occupation numbers and galaxy bias. *Monthly Notices of the Royal Astronomical Society*, 318:1144–1156, November 2000. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2000.03779.x. URL <http://adsabs.harvard.edu/abs/2000MNRAS.318.1144P>.
- Asantha Cooray and Ravi Sheth. Halo models of large scale structure. *Physics Reports*, 372:1–129, December 2002. ISSN 0370-1573. doi: 10.1016/S0370-1573(02)00276-4. URL <http://adsabs.harvard.edu/abs/2002PhR...372...1C>.

- G. Efstathiou, W. J. Sutherland, and S. J. Maddox. The cosmological constant and cold dark matter. *Nature*, 348:705–707, December 1990. ISSN 0028-0836. doi: 10.1038/348705a0. URL <http://adsabs.harvard.edu/abs/1990Natur.348..705E>.
- Z. Ivezić, J. A. Tyson, E. Acosta, R. Allsman, S. F. Anderson, J. Andrew, R. Angel, T. Axelrod, J. D. Barr, A. C. Becker, J. Becla, C. Beldica, R. D. Blandford, J. S. Bloom, K. Borne, W. N. Brandt, M. E. Brown, J. S. Bullock, D. L. Burke, S. Chandrasekharan, S. Chesley, C. F. Claver, A. Connolly, K. H. Cook, A. Cooray, K. R. Covey, C. Cribbs, R. Cutri, G. Daues, F. Delgado, H. Ferguson, E. Gawiser, J. C. Geary, P. Gee, M. Geha, R. R. Gibson, D. K. Gilmore, W. J. Gressler, C. Hogan, M. E. Huffer, S. H. Jacoby, B. Jain, J. G. Jernigan, R. L. Jones, M. Juric, S. M. Kahn, J. S. Kalirai, J. P. Kantor, R. Kessler, D. Kirkby, L. Knox, V. L. Krabbenndam, S. Krughoff, S. Kulkarni, R. Lambert, D. Levine, M. Liang, K-T. Lim, R. H. Lupton, P. Marshall, S. Marshall, M. May, M. Miller, D. J. Mills, D. G. Monet, D. R. Neill, M. Nordby, P. O'Connor, J. Oliver, S. S. Olivier, K. Olsen, R. E. Owen, J. R. Peterson, C. E. Petry, F. Pierfederici, S. Pietrowicz, R. Pike, P. A. Pinto, R. Plante, V. Radeka, A. Rasmussen, S. T. Ridgway, W. Rosing, A. Saha, T. L. Schalk, R. H. Schindler, D. P. Schneider, G. Schumacher, J. Sebag, L. G. Seppala, I. Shipsey, N. Silvestri, J. A. Smith, R. C. Smith, M. A. Strauss, C. W. Stubbs, D. Sweeney, A. Szalay, J. J. Thaler, D. Vanden Berk, L. Walkowicz, M. Warner, B. Willman, D. Wittman, S. C. Wolff, W. M. Wood-Vasey, P. Yoachim, H. Zhan, and for the LSST Collaboration. LSST: from science drivers to reference design and anticipated data products. *ArXiv e-prints*, 0805:2366, May 2008. URL <http://adsabs.harvard.edu/abs/2008arXiv0805.2366I>.
- A. J. S. Hamilton. Measuring  $\omega$  and the real correlation function from the redshift correlation function. *The Astrophysical Journal Letters*, 385:L5–L8, January 1992. doi: 10.1086/186264. URL <http://adsabs.harvard.edu/abs/1992ApJ...385L...5H>.
- Stephen D. Landy and Alexander S. Szalay. Bias and variance of angular correlation functions. *The Astrophysical Journal*, 412:64, July 1993. ISSN 0004-637X, 1538-4357. doi: 10.1086/172900. URL <http://adsabs.harvard.edu/doi/10.1086/172900>.
- Molly E. C. Swanson, Max Tegmark, Michael Blanton, and Idit Zehavi. SDSS galaxy clustering: luminosity and colour dependence and stochasticity. *Monthly Notices of the Royal Astronomical Society*, 385:1635–1655, April 2008. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2008.12948.x. URL <http://adsabs.harvard.edu/abs/2008MNRAS.385.1635S>.
- Ashley J. Ross, Shirley Ho, Antonio J. Cuesta, Rita Tojeiro, Will J. Percival, David Wake, Karen L. Masters, Robert C. Nichol, Adam D. Myers, Fernando de Simoni, Hee Jong Seo, Carlos Hernández-Monteagudo, Robert Crittenden, Michael Blanton, J. Brinkmann, Luiz A. N. da Costa, Hong Guo, Eyal Kazin, Marcio A. G. Maia, Claudia Maraston, Nikhil Padmanabhan, Francisco Prada, Beatriz Ramos, Ariel Sanchez, Edward F. Schlafly, David J. Schlegel, Donald P. Schneider, Ramin Skibba, Daniel Thomas, Benjamin A. Weaver, Martin White, and Idit Zehavi. Ameliorating systematic uncertainties in the angular clustering of galaxies: a study using the SDSS-III. *Monthly Notices of the Royal Astronomical Society*, 417:1350–1373, October 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.19351.x. URL <http://adsabs.harvard.edu/abs/2011MNRAS.417.1350R>.



Shirley Ho, Antonio Cuesta, Hee-Jong Seo, Roland de Putter, Ashley J. Ross, Martin White, Nikhil Padmanabhan, Shun Saito, David J. Schlegel, Eddie Schlafly, Uros Seljak, Carlos Hernández-Monteagudo, Ariel G. Sánchez, Will J. Percival, Michael Blanton, Ramin Skibba, Don Schneider, Beth Reid, Olga Mena, Matteo Viel, Daniel J. Eisenstein, Francisco Prada, Benjamin A. Weaver, Neta Bahcall, Dimitry Bizyaev, Howard Brewinton, Jon Brinkman, Luiz Nicolaci da Costa, John R. Gott, Elena Malanushenko, Viktor Malanushenko, Bob Nichol, Daniel Oravetz, Kaike Pan, Nathalie Palanque-Delabrouille, Nicholas P. Ross, Audrey Simmons, Fernando de Simoni, Stephanie Snedden, and Christophe Yèche. Clustering of sloan digital sky survey III photometric luminous galaxies: The measurement, systematics, and cosmological implications. *The Astrophysical Journal*, 761:14, December 2012. ISSN 0004-637X. doi: 10.1088/0004-637X/761/1/14. URL <http://adsabs.harvard.edu/abs/2012ApJ...761...14H>.

Hume A. Feldman, Nick Kaiser, and John A. Peacock. Power-spectrum analysis of three-dimensional redshift surveys. *The Astrophysical Journal*, 426:23–37, May 1994. ISSN 0004-637X. doi: 10.1086/174036. URL <http://adsabs.harvard.edu/abs/1994ApJ...426...23F>.

Lauren Anderson, Eric Aubourg, Stephen Bailey, Florian Beutler, Vaishali Bhardwaj, Michael Blanton, Adam S. Bolton, J. Brinkmann, Joel R. Brownstein, Angela Burden, Chia-Hsun Chuang, Antonio J. Cuesta, Kyle S. Dawson, Daniel J. Eisenstein, Stephanie Escoffier, James E. Gunn, Hong Guo, Shirley Ho, Klaus Honscheid, Cullan Howlett, David Kirkby, Robert H. Lupton, Marc Manera, Claudia Maraston, Cameron K. McBride, Olga Mena, Francesco Montesano, Robert C. Nichol, Sebastian E. Nuza, Matthew D. Olmstead, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, John Parejko, Will J. Percival, Patrick Petitjean, Francisco Prada, Adrian M. Price-Whelan, Beth Reid, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Cristiano G. Sabiu, Shun Saito, Lado Samushia, Ariel G. Sanchez, David J. Schlegel, Donald P. Schneider, Claudia G. Scoccola, Hee-Jong Seo, Ramin A. Skibba, Michael A. Strauss, Molly E. C. Swanson, Daniel Thomas, Jeremy L. Tinker, Rita Tojeiro, Mariana Vargas Magana, Licia Verde, David A. Wake, Benjamin A. Weaver, David H. Weinberg, Martin White, Xiaoying Xu, Christophe Yèche, Idit Zehavi, and Gong-Bo Zhao. The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Baryon acoustic oscillations in the data release 10 and 11 galaxy samples. *arXiv:1312.4877 [astro-ph]*, December 2013. URL <http://arxiv.org/abs/1312.4877>.

Mariana Vargas Magaña, Shirley Ho, Xiaoying Xu, Ariel G. Sánchez, Ross O’Connell, Daniel J. Eisenstein, Antonio J. Cuesta, Will J. Percival, Ashley J. Ross, Eric Aubourg, Stéphanie Escoffier, David Kirkby, Marc Manera, Donald P. Schneider, Jeremy L. Tinker, and Benjamin A. Weaver. SDSS-III baryon oscillation spectroscopic survey: Analysis of potential systematics in fitting of baryon acoustic feature. *ArXiv e-prints*, 1312:4996, December 2013. URL <http://adsabs.harvard.edu/abs/2013arXiv1312.4996V>.

Florian Beutler, Shun Saito, Hee-Jong Seo, Jon Brinkmann, Kyle S. Dawson, Daniel J. Eisenstein, Andreu Font-Ribera, Shirley Ho, Cameron K. McBride, Francesco Montesano, Will J. Percival, Ashley J. Ross, Nicholas P. Ross, Lado Samushia, David J. Schlegel, Ariel G. Sánchez, Jeremy L. Tinker, and Benjamin A. Weaver. The clustering of galaxies in the

- SDSS-III baryon oscillation spectroscopic survey: Testing gravity with redshift-space distortions using the power spectrum multipoles. *arXiv:1312.4611 [astro-ph]*, December 2013. URL <http://arxiv.org/abs/1312.4611>.
- Lado Samushia, Beth A. Reid, Martin White, Will J. Percival, Antonio J. Cuesta, Gong-Bo Zhao, Ashley J. Ross, Marc Manera, Éric Aubourg, Florian Beutler, Jon Brinkmann, Joel R. Brownstein, Kyle S. Dawson, Daniel J. Eisenstein, Shirley Ho, Klaus Honscheid, Claudia Maraston, Francesco Montesano, Robert C. Nichol, Natalie A. Roe, Nicholas P. Ross, Ariel G. Sánchez, David J. Schlegel, Donald P. Schneider, Alina Streblyanska, Daniel Thomas, Jeremy L. Tinker, David A. Wake, Benjamin A. Weaver, and Idit Zehavi. The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: measuring growth rate and geometry with anisotropic clustering. *Monthly Notices of the Royal Astronomical Society*, 439:3504–3519, April 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu197. URL <http://adsabs.harvard.edu/abs/2014MNRAS.439.3504S>.
- Chia-Hsun Chuang, Francisco Prada, Florian Beutler, Daniel J. Eisenstein, Stephanie Escoffier, Shirley Ho, Jean-Paul Kneib, Marc Manera, Sebastian E. Nuza, David J. Schlegel, Donald P. Schneider, Benjamin A. Weaver, Joel R. Brownstein, Kyle S. Dawson, Claudia Maraston, and Daniel Thomas. The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: single-probe measurements from CMASS and LOWZ anisotropic galaxy clustering. *arXiv:1312.4889 [astro-ph]*, December 2013. URL <http://arxiv.org/abs/1312.4889>.
- Ariel G. Sánchez, Francesco Montesano, Eyal A. Kazin, Eric Aubourg, Florian Beutler, Jon Brinkmann, Joel R. Brownstein, Antonio J. Cuesta, Kyle S. Dawson, Daniel J. Eisenstein, Shirley Ho, Klaus Honscheid, Marc Manera, Claudia Maraston, Cameron K. McBride, Will J. Percival, Ashley J. Ross, Lado Samushia, David J. Schlegel, Donald P. Schneider, Ramin Skibba, Daniel Thomas, Jeremy L. Tinker, Rita Tojeiro, David A. Wake, Benjamin A. Weaver, Martin White, and Idit Zehavi. The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: cosmological implications of the full shape of the clustering wedges in the data release 10 and 11 galaxy samples. *Monthly Notices of the Royal Astronomical Society*, 440:2692–2713, May 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu342. URL <http://adsabs.harvard.edu/abs/2014MNRAS.440.2692S>.
- Daniel J. Eisenstein, Hee-Jong Seo, Edwin Sirko, and David N. Spergel. Improving cosmological distance measurements by reconstruction of the baryon acoustic peak. *The Astrophysical Journal*, 664:675–679, August 2007. ISSN 0004-637X. doi: 10.1086/518712. URL <http://adsabs.harvard.edu/abs/2007ApJ...664..675E>.
- Nikhil Padmanabhan, Xiaoying Xu, Daniel J Eisenstein, Richard Scalzo, Antonio J Cuesta, Kushal T Mehta, and Eyal Kazin. A 2% distance to  $z=0.35$  by reconstructing baryon acoustic oscillations - i : Methods and application to the sloan digital sky survey. *arXiv:1202.0090*, January 2012. URL <http://arxiv.org/abs/1202.0090>.
- Lauren Anderson, Eric Aubourg, Stephen Bailey, Dmitry Bizyaev, Michael Blanton, Adam S Bolton, J. Brinkmann, Joel R Brownstein, Angela Burden, Antonio J Cuesta, Luiz N. A da Costa, Kyle S Dawson, Roland de Putter, Daniel J Eisenstein, James E Gunn, Hong Guo, Jean-Christophe Hamilton, Paul Harding, Shirley Ho, Klaus Honscheid, Eyal Kazin, D. Kirkby,

- Jean-Paul Kneib, Antione Labatie, Craig Loomis, Robert H Lupton, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Marc Manera, Claudia Maraston, Cameron K McBride, Kushal T Mehta, Olga Mena, Francesco Montesano, Demetri Muna, Robert C Nichol, Sebastian E Nuza, Matthew D Olmstead, Daniel Oravetz, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, John Parejko, Isabelle Paris, Will J Percival, Patrick Petitjean, Francisco Prada, Beth Reid, Natalie A Roe, Ashley J Ross, Nicholas P Ross, Lado Samushia, Ariel G Sanchez, David J. Schlegel Donald P Schneider, Claudia G Scoccola, Hee-Jong Seo, Erin S Sheldon, Audrey Simmons, Ramin A Skibba, Michael A Strauss, Molly E. C Swanson, Daniel Thomas, Jeremy L Tinker, Rita Tojeiro, Mariana Vargas Magana, Licia Verde, Christian Wagner, David A Wake, Benjamin A Weaver, David H Weinberg, Martin White, Xiaoying Xu, Christophe Yèche, Idit Zehavi, and Gong-Bo Zhao. The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Baryon acoustic oscillations in the data release 9 spectroscopic galaxy sample. *arXiv:1203.6594*, March 2012. URL <http://arxiv.org/abs/1203.6594>.
- Marc Manera, Roman Scoccimarro, Will J. Percival, Lado Samushia, Cameron K. McBride, Ashley J. Ross, Ravi K. Sheth, Martin White, Beth A. Reid, Ariel G. Sánchez, Roland de Putter, Xiaoying Xu, Andreas A. Berlind, Jonathan Brinkmann, Claudia Maraston, Bob Nichol, Francesco Montesano, Nikhil Padmanabhan, Ramin A. Skibba, Rita Tojeiro, and Benjamin A. Weaver. The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: a large sample of mock galaxy catalogues. *Monthly Notices of the Royal Astronomical Society*, 428:1036–1054, January 2013. ISSN 0035-8711. doi: 10.1093/mnras/sts084. URL <http://adsabs.harvard.edu/abs/2013MNRAS.428.1036M>.
- Will J. Percival, Ashley J. Ross, Ariel G. Sanchez, Lado Samushia, Angela Burden, Robert Crittenden, Antonio J. Cuesta, Mariana Vargas Magana, Marc Manera, Florian Beutler, Chia-Hsun Chuang, Daniel J. Eisenstein, Shirley Ho, Cameron K. McBride, Francesco Montesano, Nikhil Padmanabhan, Beth Reid, Shun Saito, Donald P. Schneider, Hee-Jong Seo, Rita Tojeiro, and Benjamin A. Weaver. The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Including covariance matrix errors. *arXiv:1312.4841 [astro-ph]*, December 2013. URL <http://arxiv.org/abs/1312.4841>.
- Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of cosmic microwave background anisotropies in closed friedmann-robertson-walker models. *The Astrophysical Journal*, 538:473–476, August 2000. ISSN 0004-637X. doi: 10.1086/309179. URL <http://adsabs.harvard.edu/abs/2000ApJ...538..473L>.
- Daniel J. Eisenstein and Wayne Hu. Baryonic features in the matter transfer function. *The Astrophysical Journal*, 496:605–614, March 1998. ISSN 0004-637X. doi: 10.1086/305424. URL <http://adsabs.harvard.edu/abs/1998ApJ...496..605E>.
- P. J. E. Peebles and M. G. Hauser. Statistical analysis of catalogs of extragalactic objects. III. the shane-wirtanen and zwicky catalogs. *The Astrophysical Journal Supplement Series*, 28:19, November 1974. ISSN 0067-0049. doi: 10.1086/190308. URL <http://adsabs.harvard.edu/abs/1974ApJS...28...19P>.

- P. C. Hewett. The estimation of galaxy angular correlation functions. *Monthly Notices of the Royal Astronomical Society*, 201:867–883, December 1982. ISSN 0035-8711. URL <http://adsabs.harvard.edu/abs/1982MNRAS.201..867H>.
- M. Davis and P. J. E. Peebles. A survey of galaxy redshifts. v - the two-point position and velocity correlations. *The Astrophysical Journal*, 267:465–482, April 1983. ISSN 0004-637X. doi: 10.1086/160884. URL <http://adsabs.harvard.edu/abs/1983ApJ...267..465D>.
- A. J. S. Hamilton. Toward better ways to measure the galaxy correlation function. *The Astrophysical Journal*, 417:19, November 1993. doi: 10.1086/173288. URL <http://adsabs.harvard.edu/abs/1993ApJ...417...19H>.
- Anna Cabre and Enrique Gaztanaga. Clustering of luminous red galaxies i: large scale redshift space distortions. *arXiv:0807.2460*, July 2008. doi: 10.1111/j.1365-2966.2008.14281.x. URL <http://arxiv.org/abs/0807.2460>. Mon.Not.Roy.Astron.Soc.393:1183-1208,2009.
- Andreu Font-Ribera, Patrick McDonald, and Jordi Miralda-Escudé. Generating mock data sets for large-scale lyman- $\alpha$  forest correlation measurements. *arXiv:1108.5606*, August 2011. URL <http://arxiv.org/abs/1108.5606>.
- Andreu Font-Ribera and Jordi Miralda-Escudé. The effect of high column density systems on the measurement of the lyman  $\alpha$  forest correlation function. *arXiv:1205.2018*, May 2012. URL <http://arxiv.org/abs/1205.2018>.
- Zheng Zheng and Jordi Miralda-Escudé. Self-shielding effects on the column density distribution of damped ly $\alpha$  systems. *The Astrophysical Journal Letters*, 568:L71–L74, April 2002. ISSN 0004-637X. doi: 10.1086/340330. URL <http://adsabs.harvard.edu/abs/2002ApJ...568L..71Z>.
- P. Noterdaeme, P. Petitjean, C. Ledoux, and R. Srianand. Evolution of the cosmological mass density of neutral gas from sloan digital sky survey II - data release 7. *Astronomy and Astrophysics*, 505:1087–1098, October 2009. ISSN 0004-6361. doi: 10.1051/0004-6361/200912768. URL <http://adsabs.harvard.edu/abs/2009A%26A...505.1087N>.
- Matthew M. Pieri, Stephan Frank, David H. Weinberg, Smita Mathur, and Donald G. York. The composite spectrum of strong ly $\alpha$  forest absorbers. *The Astrophysical Journal Letters*, 724:L69–L73, November 2010. ISSN 0004-637X. doi: 10.1088/2041-8205/724/1/L69;. URL <http://adsabs.harvard.edu/abs/2010ApJ...724L..69P>.
- Nao Suzuki, David Tytler, David Kirkman, John M. O’Meara, and Dan Lubin. Predicting QSO continua in the ly $\alpha$  forest. *The Astrophysical Journal*, 618:592–600, January 2005. doi: DOI:10.1086/426062;eprintid:arXiv:astro-ph/0306577. URL <http://adsabs.harvard.edu/abs/2005ApJ...618..592S>.
- Khee-Gan Lee, Stephen Bailey, Leslie E. Bartsch, William Carithers, Kyle S. Dawson, David Kirkby, Britt Lundgren, Daniel Margala, Nathalie Palanque-Delabrouille, Matthew M. Pieri, David J. Schlegel, David H. Weinberg, Christophe Yèche, Éric Aubourg, Julian Bautista, Dmitry Bizyaev, Michael Blomqvist, Adam S. Bolton, Arnaud Borde, Howard Brewington,

- Nicolás G. Busca, Rupert A. C. Croft, Timothée Delubac, Garrett Ebelke, Daniel J. Eisenstein, Andreu Font-Ribera, Jian Ge, Jean-Christophe Hamilton, Joseph F. Hennawi, Shirley Ho, Klaus Honscheid, Jean-Marc Le Goff, Elena Malanushenko, Viktor Malanushenko, Jordi Miralda-Escudé, Adam D. Myers, Pasquier Noterdaeme, Daniel Oravetz, Kaike Pan, Isabelle Pâris, Patrick Petitjean, James Rich, Emmanuel Rollinde, Nicholas P. Ross, Graziano Rossi, Donald P. Schneider, Audrey Simmons, Stephanie Snedden, Anže Slosar, David N. Spergel, Nao Suzuki, Matteo Viel, and Benjamin A. Weaver. The BOSS *ly $\alpha$*  forest sample from SDSS data release 9. *The Astronomical Journal*, 145:69, March 2013. ISSN 0004-6256. doi: 10.1088/0004-6256/145/3/69;. URL <http://adsabs.harvard.edu/abs/2013AJ...145...69L>.
- Timothée Delubac, Julian E. Bautista, Nicolás G. Busca, James Rich, David Kirkby, Stephen Bailey, Andreu Font-Ribera, Anže Slosar, Khee-Gan Lee, Matthew M. Pieri, Jean-Christophe Hamilton, Éric Aubourg, Michael Blomqvist, Jo Bovy, J. Brinkmann, William Carithers, Kyle S. Dawson, Daniel J. Eisenstein, Jean-Paul Kneib, J.-M. Le Goff, Daniel Margala, Jordi Miralda-Escudé, Adam D. Myers, Robert C. Nichol, Pasquier Noterdaeme, Ross O’Connell, Matthew D. Olmstead, Nathalie Palanque-Delabrouille, Isabelle Pâris, Patrick Petitjean, Nicholas P. Ross, Graziano Rossi, David J. Schlegel, Donald P. Schneider, David H. Weinberg, Christophe Yèche, and Donald G. York. Baryon acoustic oscillations in the  $ly\{\alpha\}$  forest of BOSS DR11 quasars. *arXiv:1404.1801 [astro-ph]*, April 2014. URL <http://arxiv.org/abs/1404.1801>.
- Nao Suzuki. Quasar spectrum classification with principal component analysis (PCA): Emission lines in the *ly $\alpha$*  forest. *The Astrophysical Journal Supplement Series*, 163:110–121, March 2006. ISSN 0067-0049. doi: 10.1086/499272. URL <http://adsabs.harvard.edu/abs/2006ApJS...163..110S>.
- I. Pâris, P. Petitjean, E. Rollinde, E. Aubourg, N. Busca, R. Charlassier, T. Delubac, J.-Ch Hamilton, J.-M. Le Goff, N. Palanque-Delabrouille, S. Peirani, Ch Pichon, J. Rich, M. Vargas-Magaña, and Ch Yèche. A principal component analysis of quasar UV spectra at  $z\sim 3$ . *arXiv:1104.2024*, April 2011. doi: 10.1051/0004-6361/201016233. URL <http://arxiv.org/abs/1104.2024>.
- Andy Taylor, Benjamin Joachimi, and Thomas Kitching. Putting the precision in precision cosmology: How accurate should your data covariance matrix be? *arXiv e-print 1212.4359*, December 2012. URL <http://arxiv.org/abs/1212.4359>.
- P. Noterdaeme, P. Petitjean, W. C. Carithers, I. Pâris, A. Font-Ribera, S. Bailey, E. Aubourg, D. Bizyaev, G. Ebelke, H. Finley, J. Ge, E. Malanushenko, V. Malanushenko, J. Miralda-Escudé, A. D. Myers, D. Oravetz, K. Pan, M. M. Pieri, N. P. Ross, D. P. Schneider, A. Simmons, and D. G. York. Column density distribution and cosmological mass density of neutral gas: Sloan digital sky survey-III data release 9. *Astronomy and Astrophysics*, 547:L1, November 2012. ISSN 0004-6361. doi: 10.1051/0004-6361/201220259. URL <http://adsabs.harvard.edu/abs/2012A%26A...547L...1N>.
- Matthew M. Pieri, Michael J. Mortonson, Stephan Frank, Neil Crighton, David H. Weinberg, Khee-Gan Lee, Pasquier Noterdaeme, Stephen J. Bailey, Nicolas Busca, Jian Ge, David Kirkby, Britt Lundgren, Smita Mathur, Isabelle Paris, Nathalie Palanque-Delabrouille,

- Patrick Petitjean, James Rich, Nicholas P. Ross, Donald P. Schneider, and Donald G. York. Probing the circumgalactic medium at high-redshift using composite BOSS spectra of strong Lyman-alpha forest absorbers. *arXiv:1309.6768 [astro-ph]*, September 2013. URL <http://arxiv.org/abs/1309.6768>. arXiv: 1309.6768.
- Erminia Calabrese, Renée A. Hlozek, Nick Battaglia, Elia S. Battistelli, J. Richard Bond, Jens Chluba, Devin Crichton, Sudeep Das, Mark J. Devlin, Joanna Dunkley, Rolando Dünner, Marzieh Farhang, Megan B. Gralla, Amir Hajian, Mark Halpern, Matthew Hasselfield, Adam D. Hincks, Kent D. Irwin, Arthur Kosowsky, Thibaut Louis, Tobias A. Marriage, Kavilan Moodley, Laura Newburgh, Michael D. Niemack, Michael R. Nolta, Lyman A. Page, Neelima Sehgal, Blake D. Sherwin, Jonathan L. Sievers, Cristóbal Sifón, David N. Spergel, Suzanne T. Staggs, Eric R. Switzer, and Edward J. Wollack. Cosmological parameters from pre-Planck cosmic microwave background measurements. *Physical Review D*, 87:103012, May 2013. ISSN 1550-7998. doi: 10.1103/PhysRevD.87.103012. URL <http://adsabs.harvard.edu/abs/2013PhRvD..87j3012C>.
- Andreu Font-Ribera, David Kirkby, Nicolas Busca, Jordi Miralda-Escudé, Nicholas P. Ross, Anže Slosar, Éric Aubourg, Stephen Bailey, Vaishali Bhardwaj, Julian Bautista, Florian Beutler, Dmitry Bizyaev, Michael Blomqvist, Howard Brewington, Jon Brinkmann, Joel R. Brownstein, Bill Carithers, Kyle S. Dawson, Timothée Delubac, Garrett Ebelke, Daniel J. Eisenstein, Jian Ge, Karen Kinemuchi, Khee-Gan Lee, Viktor Malanushenko, Elena Malanushenko, Moses Marchante, Daniel Margala, Demitri Muna, Adam D. Myers, Pasquier Noterdaeme, Daniel Oravetz, Nathalie Palanque-Delabrouille, Isabelle Pâris, Patrick Petitjean, Matthew M. Pieri, Graziano Rossi, Donald P. Schneider, Audrey Simmons, Matteo Viel, Christophe Yèche, and Donald G. York. Quasar-Lyman  $\alpha$  forest cross-correlation from BOSS DR11 : Baryon acoustic oscillations. arXiv e-print 1311.1767, November 2013. URL <http://arxiv.org/abs/1311.1767>.
- Jonathan L. Sievers, Renée A. Hlozek, Michael R. Nolta, Viviana Acquaviva, Graeme E. Addison, Peter A. R. Ade, Paula Aguirre, Mandana Amiri, John William Appel, L. Felipe Barrientos, Elia S. Battistelli, Nick Battaglia, J. Richard Bond, Ben Brown, Bryce Burger, Erminia Calabrese, Jay Chervenak, Devin Crichton, Sudeep Das, Mark J. Devlin, Simon R. Dicker, W. Bertrand Dorise, Joanna Dunkley, Rolando Dünner, Thomas Essinger-Hileman, David Faber, Ryan P. Fisher, Joseph W. Fowler, Patricio Gallardo, Michael S. Gordon, Megan B. Gralla, Amir Hajian, Mark Halpern, Matthew Hasselfield, Carlos Hernández-Monteagudo, J. Colin Hill, Gene C. Hilton, Matt Hilton, Adam D. Hincks, Dave Holtz, Kevin M. Huffenberger, David H. Hughes, John P. Hughes, Leopoldo Infante, Kent D. Irwin, David R. Jacobson, Brittany Johnstone, Jean Baptiste Juin, Madhuri Kaul, Jeff Klein, Arthur Kosowsky, Judy M. Lau, Michele Limon, Yen-Ting Lin, Thibaut Louis, Robert H. Lupton, Tobias A. Marriage, Danica Marsden, Krista Martocci, Phil Mauskopf, Michael McLaren, Felipe Menanteau, Kavilan Moodley, Harvey Moseley, Calvin B. Netterfield, Michael D. Niemack, Lyman A. Page, William A. Page, Lucas Parker, Bruce Partridge, Reed Plimpton, Hernan Quintana, Erik D. Reese, Beth Reid, Felipe Rojas, Neelima Sehgal, Blake D. Sherwin, Benjamin L. Schmitt, David N. Spergel, Suzanne T. Staggs, Omelan Stryzak, Daniel S. Swetz, Eric R. Switzer, Robert Thornton, Hy Trac, Carole Tucker, Masao Uehara, Katerina Visnjic, Ryan Warne,



- Grant Wilson, Ed Wollack, Yue Zhao, and Caroline Zuncke. The atacama cosmology telescope: Cosmological parameters from three seasons of data. *arXiv:1301.0824*, January 2013. URL <http://arxiv.org/abs/1301.0824>.
- Will J. Percival, Beth A. Reid, Daniel J. Eisenstein, Neta A. Bahcall, Tamas Budavari, Joshua A. Frieman, Masataka Fukugita, James E. Gunn, Željko Ivezić, Gillian R. Knapp, Richard G. Kron, Jon Loveday, Robert H. Lupton, Timothy A. McKay, Avery Meiksin, Robert C. Nichol, Adrian C. Pope, David J. Schlegel, Donald P. Schneider, David N. Spergel, Chris Stoughton, Michael A. Strauss, Alexander S. Szalay, Max Tegmark, Michael S. Vogeley, David H. Weinberg, Donald G. York, and Idit Zehavi. Baryon acoustic oscillations in the sloan digital sky survey data release 7 galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 401: 2148–2168, February 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2009.15812.x. URL <http://adsabs.harvard.edu/abs/2010MNRAS.401.2148P>.
- David Parkinson, Signe Riemer-Sørensen, Chris Blake, Gregory B. Poole, Tamara M. Davis, Sarah Brough, Matthew Colless, Carlos Contreras, Warrick Couch, Scott Croom, Darren Croton, Michael J. Drinkwater, Karl Forster, David Gilbank, Mike Gladders, Karl Glazebrook, Ben Jelliffe, Russell J. Jurek, I.-hui Li, Barry Madore, D. Christopher Martin, Kevin Pimblett, Michael Pracy, Rob Sharp, Emily Wisnioski, David Woods, Ted K. Wyder, and H. K. C. Yee. The WiggleZ dark energy survey: Final data release and cosmological results. *arXiv:1210.2130*, October 2012. URL <http://arxiv.org/abs/1210.2130>.
- Florian Beutler, Chris Blake, Matthew Colless, D. Heath Jones, Lister Staveley-Smith, Lachlan Campbell, Quentin Parker, Will Saunders, and Fred Watson. The 6df galaxy survey: baryon acoustic oscillations and the local hubble constant. *Monthly Notices of the Royal Astronomical Society*, 416:3017–3032, October 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.19250.x. URL <http://adsabs.harvard.edu/abs/2011MNRAS.416.3017B>.
- G. Efstathiou and J. R. Bond. Cosmic confusion: degeneracies among cosmological parameters derived from measurements of microwave background anisotropies. *Monthly Notices of the Royal Astronomical Society*, 304:75–97, March 1999. ISSN 0035-8711. doi: 10.1046/j.1365-8711.1999.02274.x. URL <http://adsabs.harvard.edu/abs/1999MNRAS.304...75E>.