Université Paris Sud - XI
École doctorale « Particules, Noyaux et Cosmos »(ED 517)
Institut de Recherche sur les lois fondamentales de l'Univers, CEA Saclay

Thèse de doctorat

Spécialité physique

Soutenue le 27 juin 2014

par

Arnaud Borde

One-Dimensional Power Spectrum and Neutrino Mass in the Spectra of BOSS

Directrice de thèse : Nathalie Palanque-Delabrouille

Rapporteurs : Romain Teyssier
David H. Weinberg

Composition du jury :
Président du jury : Réza Ansari
Examinateurs : David Elbaz
Julien Lesgourgues
Nathalie Palanque-Delabrouille
Romain Teyssier

Contents

Introduction

When my friends or my family ask me what I am doing, I answer "I'm preparing a Ph.D. in cosmology". Then the next question usually is "But what is cosmology?" and this one is much more complicated. However, in a few words, it can be defined as the science of the origin, development and structure of the Universe. I also often compare it to astrophysics, which studies how individual categories of objects or media like supernovae, black holes or the interstellar medium work whereas cosmology studies the Universe as a whole.

Cosmology as we know it today takes its theoretical foundations in the theory of general relativity developed by Albert Einstein in 1916. The main concepts of these almost centenary foundations are presented in chapter 1. First, I separately present the geometry and the content of the Universe before linking them through Einstein's famous equation. I also highlight two fundamental yet still mysterious concepts of the currently simplest cosmological model that explains the observations: dark matter and dark energy.

The evolution of cosmology has also led to the development of new probes to understand the Universe. Using several fields of physics like particle physics or plasma physics, these probes allow us to study the Universe at places where and epochs when it contains no bright objects. The best known is the cosmological microwave background, a relic of the first light of the Universe, emitted only 380 000 years after the Big Bang. It was discovered by mere chance in 1964 and the study of its anisotropies, often seen as the seeds of current structures such as galaxies or clusters, is surely one of the most important branches of cosmology today. Chapter 2 of this work presents the evolution from the primordial fluctuations to the matter power spectrum that reflects the universe structure. In this chapter, I also present the main probe used in this thesis: the Lyman-α forest and the quasi stellar objects (or quasars) that are used as distant lighthouses.

The end of my thesis is almost simultaneous with the end of the Sloan Digital Sky Survey III. This unique survey is the third generation of a project started in the eighties that aims at studying the large scale structures of the Universe. The Baryon Oscillation Spectroscopic Survey (BOSS) is part of it and is a spectroscopic survey whose main goal is to detect the baryon acoustic oscillations using the Lyman-α forest to probe the intergalactic medium. The project and the associated instrument is presented in chapter 3.

The baryon acoustic oscillations are not the only features that can be measured using quasar spectra and the one-dimensional Lyman-α forest power spectrum is another. This measurement forms the first part of my work, especially with the development of a new technique based on a likelihood function. It required careful selection, preparation and calibration of the data, along with a precise estimation of the systematic uncertainties to match the excellent level of statistical uncertainties provided by the unprecedented number of spectra at our disposal. This study is presented in detail in chapter 4.

The measurement of the one-dimensional power spectrum finds all its interest when combined with numerical cosmological simulations. Used together, they can constrain the value of parameters of cosmological models. But evolving a large number of particles, even when considering only gravity, is not an easy task and requires specific techniques. Chapter 5 briefly presents the most common algorithms used in cosmological simulations without going into detail but aiming at giving the reader the necessary basis to understand the next step of my work.

Finally, chapter 6 presents the grid of cosmological simulations that I developed in order to quantify the effect of several cosmological and astrophysical parameters on the one-dimensional power spectrum. For this work, I also implemented and tested a splicing technique that was developed a few years ago but was never used. This chapter ends with the preliminary constraints of cosmological parameters that result from the combination of these simulations and our measurement of the one-dimensional Lyman-α forest power spectrum.

Chapter 1

Cosmology

"The eternally incomprehensible thing about the world is its comprehensibility."
—Albert Einstein, *Physics and Reality* (1936)

THIS chapter introduces some general concepts of modern cosmology that form the background of this thesis. It is strongly inspired from lectures I had during my studies: "General Relativity" by Philipe Jetzer (U. of Zürich), "Theoretical cosmology" by Uros Seljack (U. of Zürich) and "Astrophysics II" by Simon Lily (ETH Zürich). Some parts are also inspired by the book "Fundamentals of Cosmology" by James Rich (CEA Saclay).

1.1 General Relativity: Space

1.1.1 We All Live in a Manifold

According to the theory of general relativity time and space are represented as a manifold, i.e. a topological space that locally looks like the "ordinary" Euclidian $\mathbb{R}^n$ space (for example one can map a 2-sphere on the usual $\mathbb{R}^2$ plane using the two coordinates (θ, ϕ)). This manifold has 4 dimensions: one for time and three for space. To describe such a manifold one requires a coordinate system and a metric tensor. The coordinate system $\mathbf{x} = (x^0, x^1, x^2, x^3)$ will help us define where (and when since we included time) an event occurs. The metric tensor $g_{\mu\nu}(\mathbf{x})$ will tell us how far from one another are two events. The distance ds between the points $\mathbf{x}$ and $\mathbf{x} + \mathrm{d}\mathbf{x}$ is given by:

$$\mathrm{d}s^2 = g_{\mu\nu}(\mathbf{x})\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu. \tag{1.1}$$

By construction the metric tensor is symetric ($g_{\mu\nu} = g_{\nu\mu}$) and may depend on the position where it is considered. This dependence means that space can be curved. In spherical coordinates a general metric where time is measured by comoving clock can be written

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)^2 \left(g^2(\chi, \theta, \phi)\mathrm{d}\chi^2 + f^2(\chi, \theta, \phi) \left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2 \right) \right) \tag{1.2}$$

where

$$f(\chi) = \begin{cases} \sinh\chi & \text{3-hyperbola, negative curvature} \\ \chi & \text{plane, zero curvature} \\ \sin\chi & \text{3-sphere, positive curvature} \end{cases} \tag{1.3}$$

and $a(t)$ is a scale factor. One should note that a 3-hyperbola universe is an open universe (its volume is infinite) whereas a 3-sphere universe is a closed universe (finite volume).

1.1.2 Principles

One of the bases of theoretical cosmology is the "Cosmological principle" which can be stated as:

> At any epoch, the universe appears the same to all observers, regardless of their individual locations.

As a consequence, any observer at rest in a comoving frame should see an isotropic and homogeneous universe. Because of the existence of structures and objects like stars and galaxies, the universe is often said to be only statistically homogeneous and isotropic at large scale. It is originally a philosophical statement which emphasizes the fact that we are not at the center of the Universe. However, observations of the Cosmic Microwave Background (CMB) such as in Bennett et al. (2013) and Planck Collaboration et al. (2013) allow us to test isotropy from our terrestrial point of view. Studies of other effects like galaxy clustering or the Sunyaev-Zel'dovich effect allow tests from other points of view (Planck Collaboration et al., 2013; Scrimgeour et al., 2012). There is currently no observation at odds with the hypothesis of large scale isotropy and homogeneity.

Another principle is at the center of general relativity: the equivalence principle as stated by Albert Einstein (Einstein, 1916):

1. inertial and gravitational mass are equal,
2. gravitational forces are equivalent to inertial forces,
3. in a local inertial frame we have the known laws of special relativity without gravitation.

I illustrate this principle and its consequence for simple cases in figure 1.1.

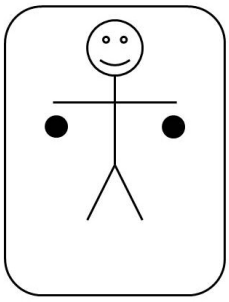
1.1.3 The Friedmann-Lemaître-Robertson-Walker Metric

Applying the cosmological principle to a general metric, one can derive the Friedmann-Lemaître-Robertson-Walker metric. Its expression in the spherical coordinate system, which is well adapted to describe a homogeneous and spherically symmetric space, is

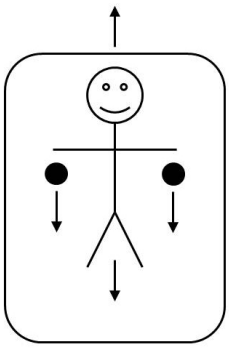
$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a(t)^2\left\{\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right\} \tag{1.4}$$

where $a(t)$ is the cosmic scale factor and k a parameter that can have three different values corresponding to different curvatures scenarios:

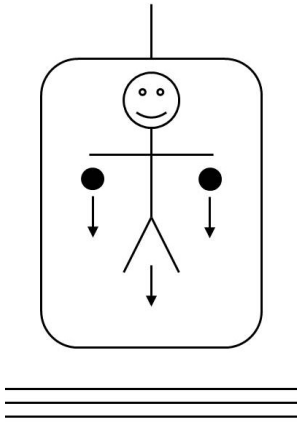
$$k = \begin{cases} -1 & \text{hyperbola, negative curvature} \\ 0 & \text{plane, zero curvature} \\ +1 & \text{sphere, positive curvature} \end{cases} . \tag{1.5}$$



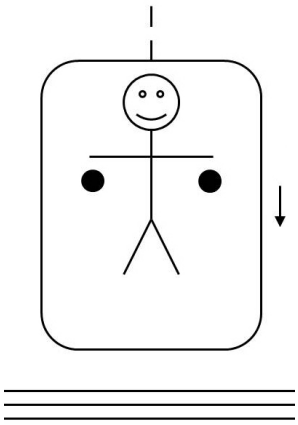
(a) An experimenter and his two stones freely floating somewhere in outer space, i.e. in the absence of forces.



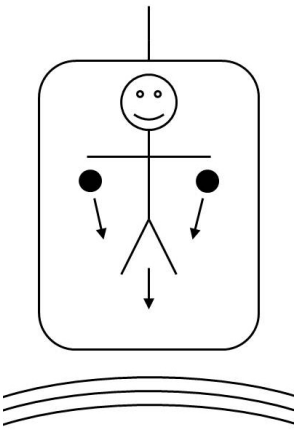
(b) An upwards constant acceleration mimics the effect of a gravitational field: experimenter and stones drop to the bottom of the box.



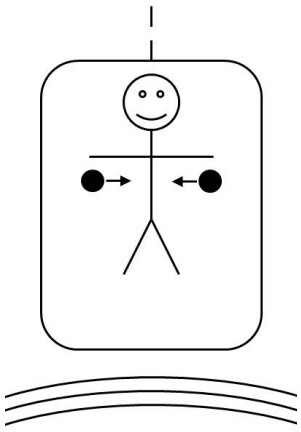
(c) The effect of a constant gravitational field: indistinguishable for our experimenter from that of a constant acceleration as in the previous figure.



(d) Free fall in a gravitational field has the same effect as no gravitational field: experimenter and stones float.



(e) The experimenter and his stones in a non-uniform gravitational field: the stones will approach each other slightly as they fall to the bottom of the elevator.



(f) The experimenter and his stones are freely falling in a non-uniform gravitational field: his body floats, so do the stones, but they move closer together, indicating the presence of some external forces.

Figure 1.1 – Illustration of the equivalence principle using an elevator, an observer, two stones and gravitational fields

In the case of a non-zero curvature, $a(t)$ is the radius of curvature of the considered space. For the flat case $k = 0$, $a(t)$ depends on the choice of coordinates and has thus no physical meaning. However its variation rate, called the Hubble parameter, does:

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \tag{1.6}$$

Current time is often noted t_0 and more generally, the 0 subscript denotes quantities taken at time t_0. A common convention is to take $a_0 = 1$, which is always possible through a good choice of coordinates when $k = 0$.

1.1.4 An Expanding Universe

Depending on the sign of $\dot{a}$, the Universe can be contracting ($\dot{a} < 0$), static ($\dot{a} = 0$) or expanding ($\dot{a} > 0$). These are three really different homogeneous universes with different past and future. One of the major breakthroughs of the past century is the discovery that our universe is an expanding universe. This is the basis of the *Big Bang* theory and is supported by three strong observational facts:

— The Hubble law: in 1929, Edwin Hubble combined his measurements of galaxy distances with Vesto Slipher and Milton Humason's measurements of the redshifts associated with the galaxies and found that the recession velocity v (found as if the redshift was due to a Doppler effect) of galaxies is approximately proportional to their distance to Earth d. The law is expressed by the equation $v = H_0 d$ where the proportionality constant H_0 is named the Hubble constant and corresponds to today's value of the Hubble parameter. The original measurement of Hubble (1929) is given in figure 1.2.

— The Cosmic Microwave Background is a black body thermal radiation, a relic of the decoupling between photons and matter. Its low temperature ($\sim 2.3\,\mathrm{K}$) is a consequence of the expansion of the Universe.

— The Primordial Nucleosynthesis: the abundance of light elements (hydrogen, helium and lithium) can be explained with an expanding universe.

1.2 Components

Up to now we only talked about the space and time we live in, but our universe is not empty. In this section we will focus on matter (or energy as they are equivalent thanks to the famous $E = mc^2$ equation) in an expanding universe.

1.2.1 Energy Conservation

In a dynamic universe of total energy density ρ with pressure p, one can apply the first law of thermodynamics, giving the variation of energy $\mathrm{d}U$ of a system for an adiabatic change of volume $\mathrm{d}V$:

$$\mathrm{d}U = -p\mathrm{d}V. \tag{1.7}$$

In the considered universe this becomes:

$$\frac{\mathrm{d}\rho a^3}{\mathrm{d}t} = -p\frac{\mathrm{d}a^3}{\mathrm{d}t}, \tag{1.8}$$

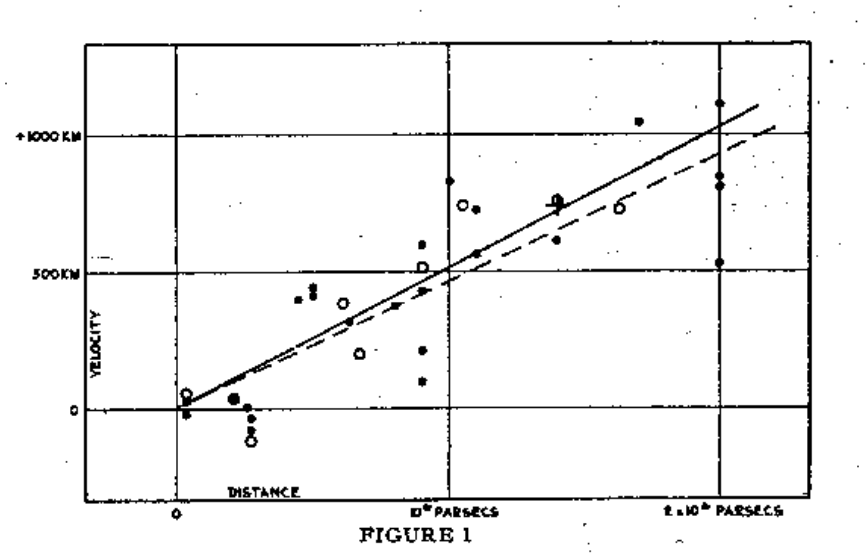


Figure 1.2 – The original Hubble diagram from Hubble (1929). The distances in this diagram are underestimated because of a bad distance calibration and it led Hubble to a significantly overestimated value of $500\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ for H_0. Here is the original caption: *"Radial velocities, corrected for solar motion, are plotted against distances estimated from involved stars and mean luminosities of nebulae in a cluster. The black discs and full line represent the solution for solar motion using the nebulae individually; the circles and broken line represent the solution combining the nebulae into groups; the cross represents the mean velocity corresponding to the mean distance of 22 nebulae whose distances could not be estimated individually."*

which, after expansion of the first derivative, gives us the equation of energy conservation in an expanding universe:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -3H(\rho + p). \tag{1.9}$$

this equation illustrates the conservation of the total energy density. However, the Universe is made of several components and if one wants to know about a specific one, it is necessary to write down the Boltzmann equation for this component. Dealing with the interactions between the different particles and the evolution of the number of particles of each species, the Boltzmann equation is usually quite complicated. However, there are two cases where it can be heavily simplified:

— when interaction times are large compared to the expansion characteristic time H^{-1}. Then the comoving number density of particles is conserved:

$$\frac{\mathrm{d}na^3}{\mathrm{d}t} = 0. \tag{1.10}$$

— When the different species are in thermal equilibrium. In this particular case, the entropy is conserved for an adiabatic expansion.

We will now examine what this means for matter depending on whether or not it is relativistic. Relativistic matter (particles with a velocity v close the speed of light c) is also called radiation.

1.2.2 Non-Relativistic Matter

Non-relativistic matter, often called *dust* and denoted by the subscript m, has no pressure. Thus the energy conservation equation can be rewritten:

$$\frac{\mathrm{d}\rho_m}{\mathrm{d}t} = -3H\rho_m. \tag{1.11}$$

Once integrated, it gives the proportionality relation

$$\rho_m \propto a^{-3}. \tag{1.12}$$

This is what one can expect for ordinary matter: the density is inversely proportional to the volume.

1.2.3 Relativistic Matter

Radiation is driven by the equation of state $\rho_r = 3p_r$ which we can inject in the energy conservation equation to get

$$\frac{\mathrm{d}\rho_r}{\mathrm{d}t} = -4H\rho_r, \tag{1.13}$$

leading to the relation

$$\rho_r \propto a^{-4}. \tag{1.14}$$

This means that in addition to the dilution of the energy density due to the expansion of the Universe, radiations undergo an additional loss of energy, proportional to a^{-1}. Since for a radiation the energy is proportional to its wavelength this means that this diminution of energy corresponds to a decrease of wavelength. This is the cosmological redshift z, defined by

$$1 + z = \frac{\lambda_{\text{emission}}}{\lambda_{\text{reception}}} \tag{1.15}$$

where $\lambda_{\text{emission}}$ is the wavelength of the radiation when it is emitted and $\lambda_{\text{reception}}$ is the wavelength seen by the observer. Since the loss of energy is proportional to a^{-1} we get

$$1 + z = \frac{a_{\text{reception}}}{a_{\text{emission}}} \tag{1.16}$$

where a_{emission} and $a_{\text{reception}}$ are the values of the scale factor respectively at emission and reception time. Since a is coordinate dependent, we can always choose a to be 1 at the reception time on Earth. Thus (suppressing the subscript since there is no ambiguity):

$$1 + z = \frac{1}{a}. \tag{1.17}$$

In other words, a measure of the redshift of a radiation is a measure of the cosmic scale factor at the time of emission.

1.3 Linking Container and Content

1.3.1 Einstein's Equation

Einstein's field equation gives us the link between the geometry of the universe, represented by the metric, and the content of the universe represented by the energy-momentum tensor $T^{\mu\nu}$:

$$R_{\mu\nu} - \frac{\mathcal{R}}{2} g_{\mu\nu} + \Lambda g_{\mu\nu} = -\frac{8\pi G}{c^4} T_{\mu\nu} \tag{1.18}$$

where:

— $R_{\mu\nu}$ is the Ricci tensor (contraction of the curvature tensor: $R_{\mu\nu} = R^{\lambda}_{\mu\lambda\nu}$)
— $\mathcal{R}$ is the scalar curvature (contraction of the Ricci tensor: $\mathcal{R} = g^{\mu\nu}R_{\nu\mu} = R^{\mu}_{\mu}$)
— Λ is the cosmological constant, which was first introduced by Einstein to allow for a static universe, but was removed afterwards. Nowadays, it is usually replaced by a component (dark energy) with constant energy density. More general formulations of dark energy allow this component to have a redshift-dependent energy density.

For an ideal fluid the energy-momentum tensor is given by

$$T_{\mu\nu} = \left(\rho + \frac{p}{c^2}\right) u^{\mu}u^{\nu} - g^{\mu\nu}p \tag{1.19}$$

with

— u^{μ}, the four-velocity,
— ρ, the proper energy density (including mass energy through $E = mc^2$),
— p, the pressure of the fluid.

1.3.2 Friedmann's Equations

The Einstein's equation in the case of an isotropic and homogeneous universe can be reduced to two scalar equations (instead of ten since we are dealing with 4×4 symetric tensors). If one considers the time component and the trace, it respectively gives the first and second Friedmann equations.

1.3.2.1 The First Friedmann's Equation

Taking the time component ($\mu = \nu = 0$) of equation 1.18 one gets

$$H(t)^2 = \frac{8\pi G}{3}\rho(t) - \frac{k}{a(t)^2} + \frac{\Lambda}{3} \tag{1.20}$$

where $\rho = \rho_{\text{matter}} + \rho_{\text{radiation}}$ is the sum of the non-relativistic and relativistic matter energy densities and $k = -1, 0, +1$ depends on the curvature of the Universe. This equation links the evolution of the scale factor (through the Hubble parameter) to the energetic content of the Universe. If the Universe is flat, the equation can be rewritten in the form (omitting time dependence)

$$H^2 = \frac{8\pi G}{3}\left(\rho + \frac{\Lambda}{8\pi G}\right). \tag{1.21}$$

It is then useful to see the cosmological constant term as an energy density $\rho_{\Lambda} = \frac{\Lambda}{8\pi G}$. Such an energy density has very particular properties which we describe in 1.4.2. One can then obtain the energy density for a flat universe:

$$\rho_{\text{flat}} = \frac{3H^2}{8\pi G}. \tag{1.22}$$

We can now define the critical energy density as the current energy density of a flat universe:

$$\rho_c \equiv \frac{3H_0^2}{8\pi G} \tag{1.23}$$

which allows us to rewrite equation 1.20 as

$$\left(\frac{H}{H_0}\right)^2 = \frac{\rho_{tot}}{\rho_c} - \frac{k}{H_0^2 a^2} \tag{1.24}$$

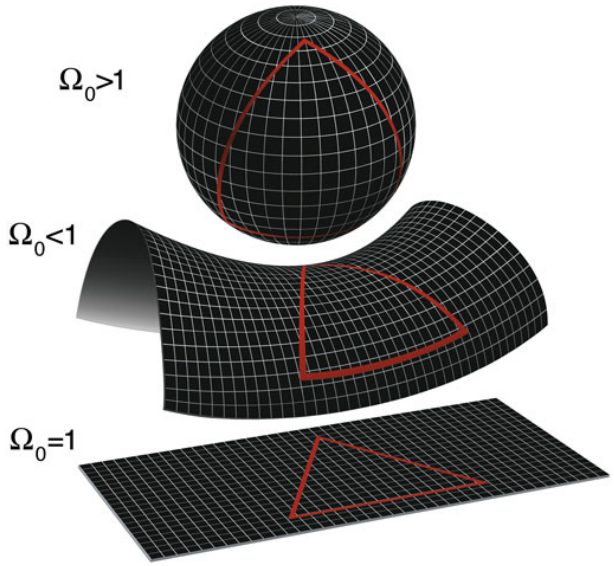


Figure 1.3 – The three different possible topologies for a universe: closed, open, flat (from top to bottom). Credits: NASA/WMAP Science Team

where $\rho_{tot} = \rho_{\text{matter}} + \rho_{\text{radiation}} + \rho_\Lambda$. If we write this equation at $t = t_0$ we get

$$\frac{\rho_{tot}(t_0)}{\rho_c} = \frac{H_0^2 + k}{H_0^2}. \tag{1.25}$$

Thus the position of $\Omega_0 \equiv \frac{\rho_{tot}(t_0)}{\rho_c}$ with respect to 1 determines the topology of the universe:

$$\begin{cases} \Omega_0 < 1 \Leftrightarrow k = -1 & \text{open universe} \\ \Omega_0 = 1 \Leftrightarrow k = 0 & \text{flat universe} \\ \Omega_0 > 1 \Leftrightarrow k = +1 & \text{closed universe} \end{cases}. \tag{1.26}$$

The three different cases are illustrated in figure 1.3.

1.3.2.2 The Second Friedmann's Equation

We now take the trace of Einstein's equation:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \tag{1.27}$$

with $\rho = \rho_{\text{matter}} + \rho_{\text{radiation}}$ and $p = p_{\text{matter}} + p_{\text{radiation}}$. As for the first Friedmann's equation we introduce $\rho_\Lambda = \frac{\Lambda}{8\pi G}$, and get

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho_{tot} - 3\rho_\Lambda + 3p). \tag{1.28}$$

Thus if we want to describe the cosmological constant as another component like matter or radiation we need to introduce the corresponding pressure term $p_\Lambda = -\rho_\Lambda$. Then the previous equation becomes

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho_{tot} + 3p_{tot}). \tag{1.29}$$

In a general case, one can define the equation of state of a component as $\rho = wp$. From what we have seen before, we have $w = 0$ for non-relativistic matter and $w = 1/3$ for radiation. The

energy conservation equation then becomes

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -3H\rho(1+w) \tag{1.30}$$

And thus the general solution is

$$\rho \propto a^{-3(1+w)}. \tag{1.31}$$

This formalism allows us to easily derive solutions to the Friedmann equation in the cases where only one component is considered (or dominating over the others). For example, in a flat universe ($k = 0$), one gets

$$a(t) \propto t^{\frac{2}{3(1+w)}}. \tag{1.32}$$

This means that the time dependence of the expansion of the universe will allow to distinguish the dominating component at a given time. One can also include the cosmological constant term in this formalism by defining $w' = p'/\rho'$ where

$$\rho' = \rho + \frac{\Lambda}{8\pi G} \quad \text{and} \quad p' = p - \frac{\Lambda}{8\pi G}. \tag{1.33}$$

Of course, if the different components are non-interacting one can use a linear combination of the individual solutions to get the solution for a more realistic multi-component universe.

1.3.3 Ω's

For each component, we can define the normalized density parameter "Ω_i":

$$\Omega_i = \frac{\rho_{i,0}}{\rho_c}, \qquad i = m, r, \Lambda. \tag{1.34}$$

We can similarly define an "Ω" for the curvature term:

$$\Omega_k = -\frac{k}{H_0^2}. \tag{1.35}$$

Using these parameters, we can derive the first Friedmann's equation classic form:

$$\left(\frac{H}{H_0}\right)^2 = \Omega_k a^{-2} + \Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_\Lambda. \tag{1.36}$$

This equation has five parameters (H_0 and four Ω_i). At $t = t_0$, it becomes

$$\Omega_m + \Omega_r + \Omega_\Lambda = 1 - \Omega_k. \tag{1.37}$$

And thus there are only three out of four Ωs that are independent. We also find again that for a flat universe, the total energy density is equal to the critical energy density:

$$\Omega_m + \Omega_r + \Omega_\Lambda = 1 \quad \text{or} \quad \rho_{m,0} + \rho_{r,0} + \rho_{\Lambda,0} = \rho_c. \tag{1.38}$$

1.3.4 Cosmological Distances

We have seen that spacetime is closely related to its energetic content. Cosmology uses this strong link to constrain either the container or the content. However, our only tools are telescopes. That means that the only thing we are able to measure from space is a flux of photons, its direction and sometimes its time evolution and its spectral composition. We will see how we can get information on the Universe from these measurements.

1.3.4.1 Comoving Distance

We consider a photon emitted at spherical coordinates $(t_e, \chi_e, \theta_e, \phi_e)$ that we observe at $(t_0, \chi = 0, \theta_e, \phi_e)$ since the progation occurs at constant θ and ϕ. Moreover, photons follow $\mathrm{d}s^2 = 0$, so according to equation 1.1 we have

$$\frac{\mathrm{d}t}{a(t)} = \mathrm{d}\chi \tag{1.39}$$

which we can integrate to obtain

$$\chi_e = \int_{t_e}^{t_0} \frac{\mathrm{d}t}{a(t)}. \tag{1.40}$$

χ_e is called the comoving distance but in fact it is a coordinate (which is time independent) and it is $a_0\chi_e$ which is homogeneous to a distance. It is usual to perform the change of variable from t to z which leads to

$$\chi(z_e) = \int_{z_e}^{0} \frac{\mathrm{d}z}{H(z)}. \tag{1.41}$$

Thus if one knows both the spectral redshift and the comoving distance of a galaxy, constraints on the integral of $H(z)$ can be derived. As $H(z)$ depends on H_0 and the Ω's, one also obtains constraints on these parameters. With several measurements at different redshifts, it is possible to break the degeneracies. However, while measuring a redshift is quite easy, using emission lines for example, measuring the comoving distance is not.

1.3.4.2 Luminosity Distance

Fortunately, one can get information on distances using the intensity and the wavelength of the photon flux. Let us consider a source at comoving distance χ emitting N photons of average frequency ν_e at time t_e during the infinitesimal duration $\mathrm{d}t$. The intrinsic luminosity L_e of the source is thus

$$L_e = \frac{N\nu_e}{\mathrm{d}t}. \tag{1.42}$$

The surface $S(\chi)$ of the sphere centered on the source is given by

$$S(\chi) = 4\pi f^2(\chi) \tag{1.43}$$

where $f(\chi)$ is defined in 1.2. In the case of a simple static universe, the received flux F_S would simply be

$$F_S = \frac{L_e}{S(\chi)}. \tag{1.44}$$

However, in an expanding universe the photon's frequency is proportional to the scale factor $\nu_0 \propto a = \frac{1}{1+z}$ and the arrival times are spread out: the signal is received in the time interval $\mathrm{d}t' = (1+z)\,\mathrm{d}t$. The luminosity after propagation is then

$$L_0 = \frac{N\nu_0}{\mathrm{d}t'} = \frac{L_e}{(1+z)^2} \tag{1.45}$$

and the flux we receive is

$$F_0 = \frac{L_0}{4\pi f^2(\chi)} = \frac{F_S}{(1+z)^2}. \tag{1.46}$$

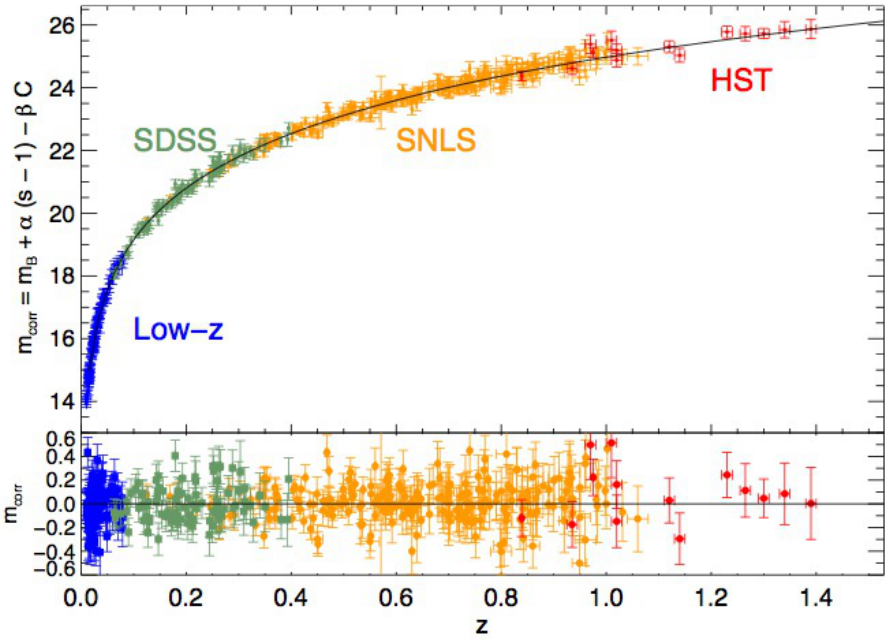


Figure 1.4 – Corrected absolute brightness ($m_{\rm corr} \propto log(D_L)$ as a function of redshift for 472 supernovae, including measurements from the SNLS, HST, the Sloan Digital Sky Survey, and other facilities for the low-z sample. The detailed shape of the diagram especially at large redshift is evidence for the accelerating expansion of the universe. The lower panel shows the residual from the best fit of dark energy model with w as a free parameter. The results is $w = -0.90^{+0.17}_{-0.24}$ which is compatible with a cosmological constant model. Credits: Conley et al. (2011)

The luminosity distance is then defined as

$$D_L = (1+z) f(\chi) \tag{1.47}$$

so that the received flux can be written such that D_L would be the source's distance in a flat and static universe:

$$F_0 = \frac{L_e}{4\pi D_L^2}. \tag{1.48}$$

This equation is interesting because if one knows the intrinsic luminosity L_e and the flux F_0 received from a given source, one can derive its luminosity distance. In addition, knowing the redshift allows one to derive the comoving distance in a flat universe.

An astrophysical object for which the intrinsic luminosity is known is called a "standard candle". There are currently two main types of standard candles used in cosmology: Cepheids (variable stars whose luminosity is linked to their periodicity) and type Ia supernovae. Type Ia supernovae are expected to arise from stellar binaries in which at least one is a white dwarf. This white dwarf accretes matter from its companion until it reaches the Chandrasekhar mass, point at which it explodes into a supernova. Because the explosion always occurs at the characteristic mass, these supernovae have very similar intrinsic luminosities. Compared to the Cepheids, supernovae are very bright and thus can be seen up to very high redshifts. Hundreds of such supernovae have been observed by dedicated experiments like the SuperNova Legacy Survey (Astier et al., 2006), the Sloan Digital Sky Survey Supernova Survey (Holtzman et al., 2009; Kessler et al., 2009) or with the Hubble Space Telescope (Riess et al., 2007). The latest Hubble diagram combining those measurements is shown in figure 1.4.

1.3.4.3 Angular Distance

Let us now consider an object of physical size l, tranverse to the line of sight, at the comoving distance χ_e, from which a photon is emitted at each side at time t_e, as shown in figure 1.5. In a static universe, assuming $\chi_e \gg l$, these photons are seen separated by an angle

$$\theta_S = \frac{l}{f(\chi_e)}. \tag{1.49}$$

However, this is not the case in a homogeneous expanding universe where photons follows trajectories at constant angular coordinates (θ, ϕ). Since the comoving size of the object is $\frac{l}{a}$, we are in fact seeing an angle

$$\theta_0 = \frac{l}{af(\chi_e)} = (1+z)\theta_S. \tag{1.50}$$

Thus we can define the angular distance D_A:

$$D_A = \frac{f(\chi_e)}{1+z} \tag{1.51}$$

so that the angle can be rewritten in the same form as in a flat static universe:

$$\theta_0 = \frac{l}{D_A}. \tag{1.52}$$

The principle to get constraints on the geometry or the content of the Universe is the same as with the luminosity distance but one needs to know the physical size of an object instead of its intrinsic luminosity. Such objects are called standard rulers and one (the baryon acoustic oscillations) will be presented in section 2.1.1.1.

1.4 ΛCDM

ΛCDM is a cosmological model of a flat, homogeneous and isotropic universe in which gravity is governed by the theory of general relativity as seen in the previous sections. It is currently the simplest model that provides predictions of "basic" observations such as

— the cosmic microwave background,
— the large scale structures in the distribution of galaxies,
— the abundance of light elements (hydrogen, helium and lithium),
— and the accelerating expansion of the Universe.

The energetic content of the Universe in this model is described with three main components given with their respective contribution at t_0 according to the recent results from the Planck team (Planck Collaboration et al., 2013):

— ordinary matter (4%),
— cold (non-relativistic) dark matter (25%),
— a cosmological constant or dark energy (70%),
— neutrinos,
— photons.

The contribution of each component is illustrated in figure 1.6 along with the repartition at $z \sim 1100$ when the CMB was emitted. Interpreting recent cosmological data with a ΛCDM model implies that our Universe is dominated by two components of unknown nature, dark matter and dark energy, which are briefly described below.

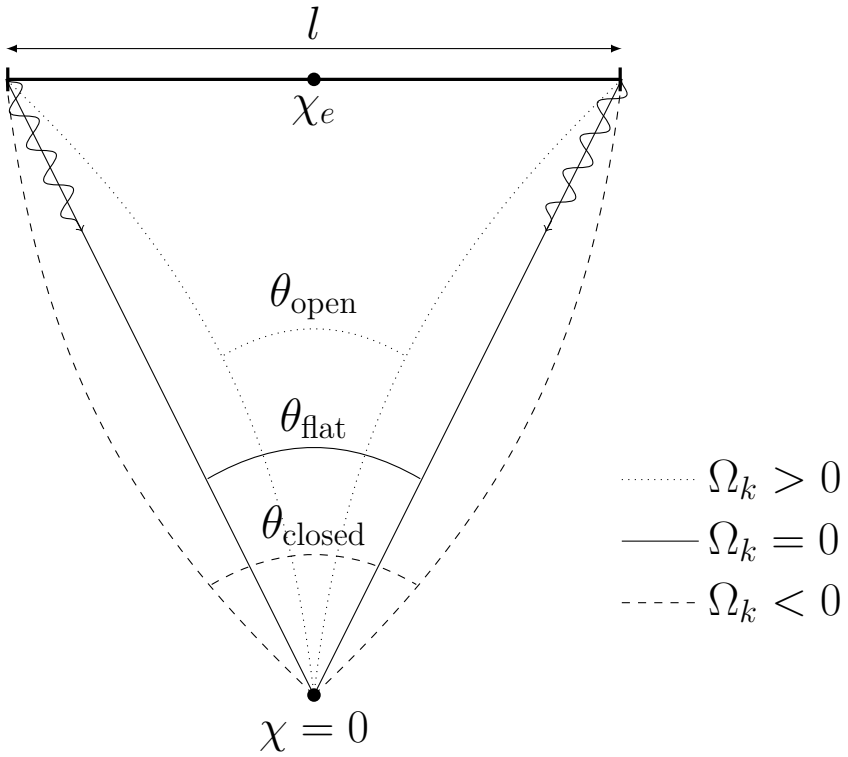


Figure 1.5 – Object of size l standing at comoving distance χ_e tranverse to the line of sight. The photons emitted at each end follow a path dependent of the curvature of the universe. In an open universe ($\Omega_k > 0$) the angle θ_{open} separating them when they arrive at the observer in $\chi = 0$ is inferior to the angle θ_{flat} for a flat universe ($\Omega_k = 0$), which itself is inferior to the angle θ_{closed} for a closed universe ($\Omega_k < 0$).

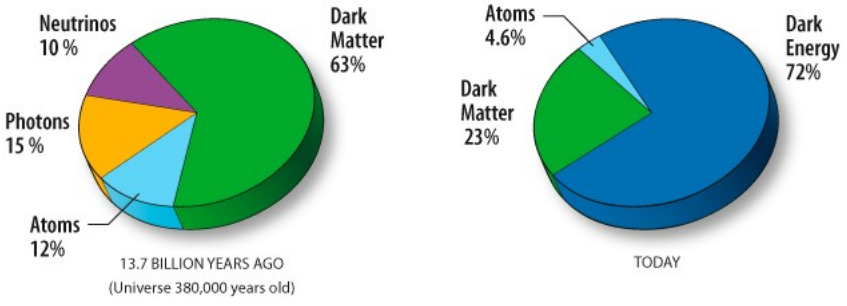


Figure 1.6 – Content of the Universe: at the time of decoupling between atoms and photons (left) and today (right). Credits: NASA/WMAP Science Team

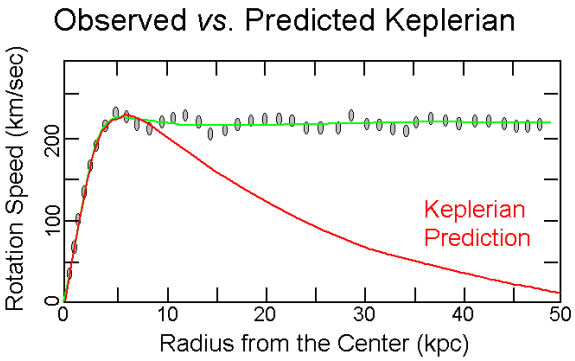


Figure 1.7 – Comparison between the predicted Keplerian rotation curve (red), accounting only for visible matter (stars) compared to the actual observed rotation curve (green). Credits: Prof. Richard Pogge, Ohio State University

1.4.1 Dark Matter

Dark matter is a hypothetic (no direct detection has been made yet) type of matter which does not (or very faintly) interact with photons and consequently cannot be seen. However, its properties can be inferred from its gravitational effects. It was first proposed in 1932 by Jan Ort to account for the orbital velocities of stars in the Milky Way and in 1933 by Fritz Zwicky to account for the peculiar velocities of galaxies in clusters. The first effect is presented in figure 1.7. Since then, other observational evidence has been found, including gravitational lensing and the shape of the angular fluctuations in the cosmic microwave background spectrum.

1.4.2 Dark Energy

Whereas the first Friedmann's equation (1.20) drives the first derivative of the scale factor a, the second one (1.29) drives the evolution of the second derivative $\ddot{a}$. In particular we have $\ddot{a} \propto (\rho_{tot} + 3p_{tot})$. Ordinary matter and radiation both have $\rho > 0$ and $p > 0$. Thus because there is some ordinary matter around us, the Universe should be either in a decelerating expansion or in a contracting state.

Because he did not like the idea of a dynamic universe, Einstein introduced the cosmological constant in his equation to allow for static solutions. We have seen that one can treat this constant as an energetic component, the so-called *dark energy*. We have also seen that the energy density and the pressure of this component have an opposite sign: $\rho_\Lambda = -p_\Lambda$; i.e. $w = -1$. Thanks to this, one can now obtain a positive or null second derivative of a. One should note that a redshift (or time) dependent equation of state $\rho_\Lambda = w(z)p_\Lambda$ is also possible and many experiments are trying to measure w at various epochs to distinguish between the models for dark energy. Distance measurements such as the Hubble diagram shown in figure 1.4 are powerful tools to provide constraints on dark energy models (Sullivan et al., 2011). Figure 1.8 shows the constraints one can get on w (and Ω_m) using only supernovae assuming a flat universe with a constant dark energy equation of state.

Einstein dropped his cosmological constant when the Universe was discovered to be expanding. But it was at that time thought that this expansion could only be decelerating. It was reintroduced recently when it was discovered through the observation of type Ia supernovae that this expansion was in fact accelerating (Riess et al., 1998; Perlmutter et al., 1999). However,

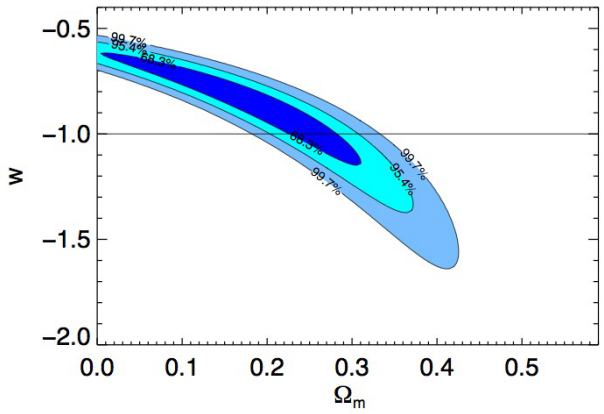


Figure 1.8 – Statistical supernovae only constraints on w, Ω_m from SNLS assuming a flat universe with a constant dark energy equation of state. Credits: Conley et al. (2011)

the fact that this constant is the result of the action of a mysterious dark energy is not obvious and other interpretations exist (like modified gravity models).

Chapter 2

From Quantum Fluctuations to Lyman-α Forest

"It has been said that something as small as the flutter of a butterfly's wing can ultimately cause a typhoon halfway around the world."
—Chaos Theory

After introducing the basics of modern cosmology, we take a deeper look at the perturbations of the matter density field and their consequences. We also look at the main tool used in this work: the one-dimensional matter power spectrum and the Lyman-α forest, its tracer. Most generalities about quasars and the Lyman-α forest come from Osmer (2006) and Rauch (2006).

In the previous chapter, results were obtained in the context of general relativity, respecting the cosmological principles. The metric tensor $g_{\mu\nu}$ and the energy-momentum tensor $T_{\mu\nu}$ were considered homogeneous and isotropic. However, it is obvious from basic observations that the Universe is not perfectly homogeneous and isotropic. The next step is thus to introduce perturbations in these tensors and to study how these perturbations evolve. The resulting calculations are long, complicated and are not relevant for this work and therefore are not detailed here. They can be found, for example, in "Cosmology" by Steven Weinberg or in "Fundamentals of Cosmology" by Jim Rich for a more intuitive approach. I will here stay on a qualitative level. After some elements on these perturbations, I will present the quasars, some of the most luminous objects in the universe that can hence be used for studies at high redshift. In a last section, I will introduce the main tool used for this work: the Lyman-α forest, a remarkable feature in the quasars spectra that may be used to probe the matter distribution in the intergalactic medium.

2.1 From Primordial Fluctuations to Matter Power Spectrum

2.1.1 Primordial Fluctuations

Primordial fluctuations are perturbations of the density, velocity, ... fields which are at the origin of the formation of structures (clusters, galaxies, ...) that we observe nowadays. In inflation models, these fluctuations are described as quantum fluctuations in the plasma that formed the Universe at its beginning. These fluctuations are then brought to a macroscopic scale during a very intense expansion phase called inflation. Afterwards, these fluctuations evolve during the eras of radiation or matter domination.

In the context of general relativity, three different types of perturbations are possible: scalar, vector and tensor perturbations. For simplicity, we will here only consider scalar perturbations. These scalar perturbations can be either adiabatic or of isocurvature. The first one corresponds to fluctuations in the total energy density ρ whereas the second type comes from fluctuations in the composition such as the ratio of baryon to photon number densities η:

$$\begin{array}{ll} \text{adiabatic:} & \Delta\rho/\rho \neq 0 \quad \Delta\eta/\eta = 0 \\ \text{isocurvature:} & \Delta\rho/\rho = 0 \quad \Delta\eta/\eta \neq 0 \end{array}$$

The fluctuations observed in the cosmological microwave background (see section 2.1.2) are in agreement with adiabatic primordial fluctuations. Such fluctuations are also produced by single field inflation models. We will now see how these fluctuations lead to the baryon acoustic oscillations that can serve as a standard ruler as seen in the previous chapter.

2.1.1.1 Baryon Acoustic Oscillations

The theory of structure formation relies mainly on gravitational instability, that is the idea that in overdense regions of the universe, gravitational collapse through self-gravity is stronger than expansion. Thus overdensities tend to grow over time. On the other hand in underdense regions expansion is stronger and these regions become more underdense over time.

Recombination represents the epoch when the Universe is cool enough for electrons and protons to form hydrogen atoms. At this time, there are no longer enough free electrons for photons to significantly scatter on them: the Universe becomes transparent. Before recombination, Thomson scattering (between photons and electrons) is predominant and the free-streaming scale of photons is much smaller than the size of the horizon $H^{-1}(z)$. Photons and electrons are thus strongly coupled. In addition, protons and electrons interact through the Coulomb force. These three types of particles are coupled and form a unique fluid called the baryon-photon plasma in which density perturbations evolve like sound waves.

To see what is a baryon acoustic oscillation, let us consider a point-like, adiabatic perturbation. Originally, all the components are affected by this perturbation: cold (non-relativistic) dark matter, baryons, photons and ultrarelativistic neutrinos). Neutrinos only interact weakly and are too fast and too light to be affected by gravitation, so they just stream away from the initial perturbation. Cold dark matter is only affected by gravity and thus only stands growing at the original position. Because the baryon-photon plasma is very hot and dominated by photons at this time, it has a strong pressure compared to its density. The initial overdensity is thus also an initial overpressure. As pressure tries to equalize itself with the surroundings, this results in an expanding spherical sound wave. It is exactly like the common soundwave, with the sound speed in the plasma c_s approximately equal to $c/\sqrt{3}$. The baryon and photon

perturbation is carried outwards and its density drops as the energy is spread over the expanding spherical wave as shown in figure 2.1a.

As time goes on, the wave continues to propagate, the neutrinos spread out and dark matter accumulates in the overall density perturbation. Not only is the peak in dark matter growing, but the width of the perturbation is also increasing since background material is attracted. This can be seen in figure 2.1b.

At the time of recombination, electrons and nuclei start to combine to form neutral atoms, and photons and baryons become decoupled. The photon perturbation begins to spread out like the neutrinos earlier as shown on figure 2.1c. As decoupling happens, the photon pressure decreases (in addition, the photons are getting cooler) and thus the wave starts to slow down.

The process continues until the photons have completely leaked out of the perturbation, and the sound wave has almost stopped propagating. We are left with a dark matter perturbation around the origin and a gas perturbation in a shell of about 150 Mpc (comoving). This is what is seen in figure 2.1d.

However, baryons and dark matter attract each other through gravitation and this causes the two perturbations to act on each other. More precisely, the two overdensities are increasing in response to the combined gravitational forces of both components (see figure 2.1e). Eventually, the two perturbations look very similar and the spherical shell of gas has imprinted itself in the dark matter: this is the acoustic peak (figure 2.1f).

Only one perturbation was described here but since they are very small, they can be linearly summed. Galaxy formation occurs in the overdense regions and although most of it happens at the position of the original fluctuations, there is a tiny excess in the regions that are 150 Mpc away from these initial perturbations. This length scale, originally related to a density excess shortly after recombination, is expected to be also present in the distribution of matter in the later universe. It can therefore be used as a standard ruler to probe our cosmological model. This is detected as a single *acoustic peak* in the correlation function of galaxies or a series of *acoustic oscillations* in the corresponding power spectrum (Eisenstein et al., 2005; Anderson et al., 2013).

2.1.2 Cosmological Microwave Background

The CMB is the most important source for cosmological constraints. I will here briefly present its main characteristics.

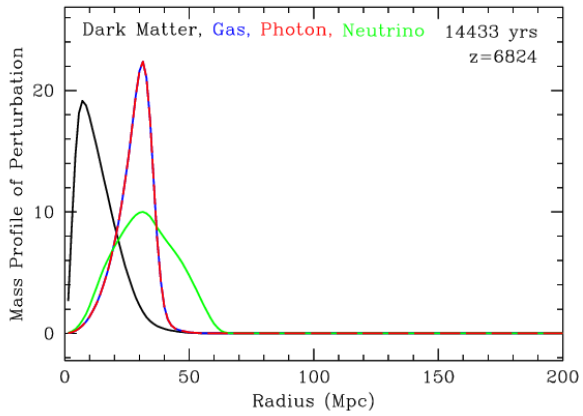
Before recombination, the Thomson scattering rate is high and thus the Universe is opaque to light. After recombination, there are almost no more free electrons and thus the Thomson scattering rate is almost zero: the Universe is transparent. The photons that were in the initial plasma are freed and propagate on a straight line until they reach us nowadays [1]. This electromagnetic radiation is called the cosmological microwave background (CMB). It is still observable today.

The CMB presents a quasi-perfect blackbody temperature spectrum. The intensity I at a given wavelength ν is therefore given by Planck's law:

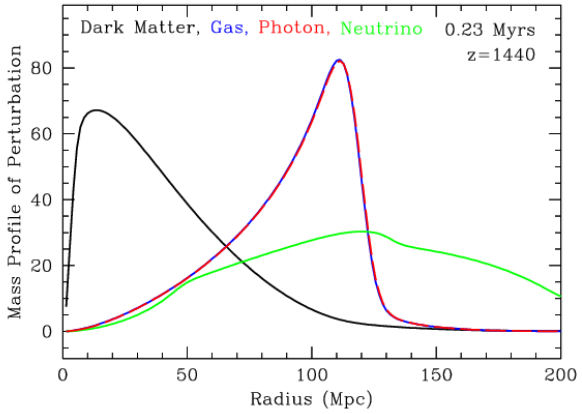
$$I(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1} \tag{2.1}$$

where T is the temperature of the blackbody. This temperature presents some small fluctuations (or anisotropies) ($\Delta T/T \sim 10^{-5}$) as a function of position of the measurement on the sky. These

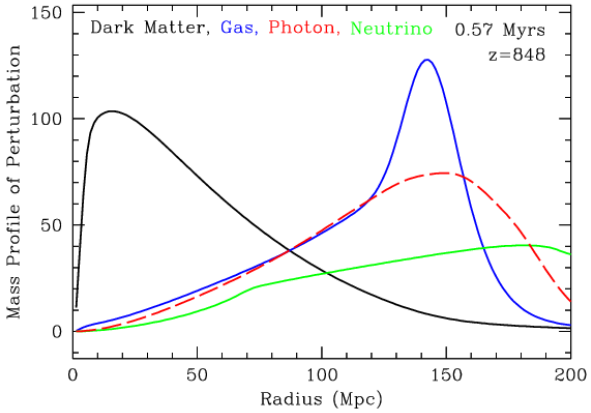
1. We do not take reionisation into account at this point.



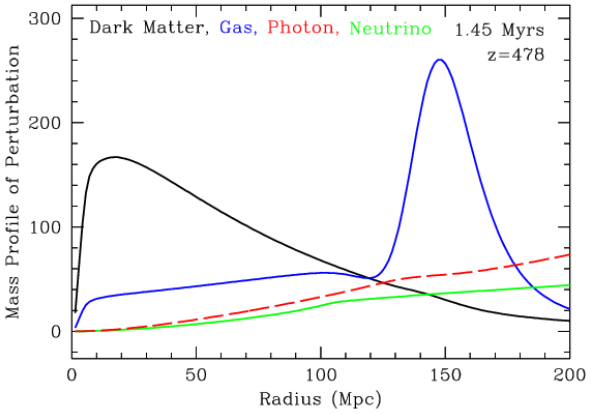
(a) Near the initial time, the photons and baryons travel outward as a pulse.



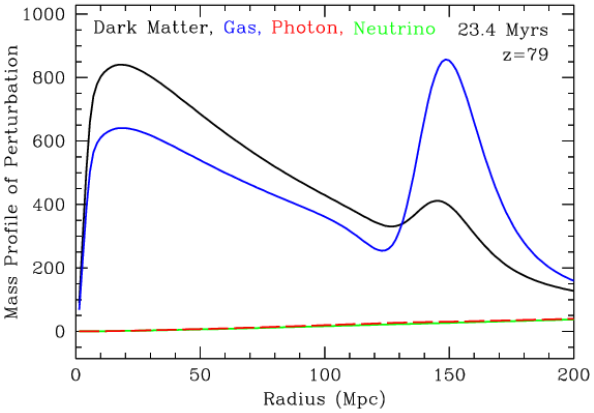
(b) Approaching recombination, one can see the wake in the cold dark matter raised by the outward-going pulse of baryons and relativistic species.



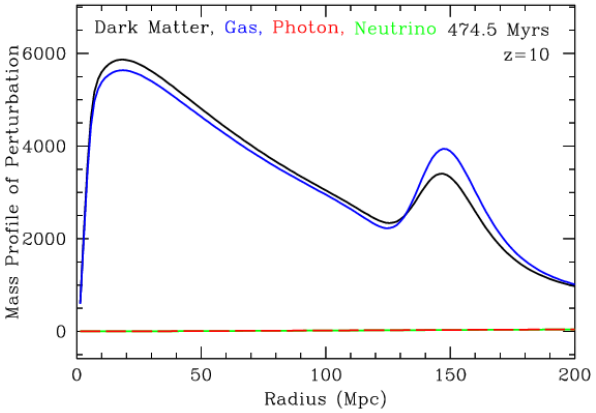
(c) At recombination, the photons leak away from the baryonic perturbation.



(d) With recombination complete, we are left with a dark matter perturbation toward the center and a baryonic perturbation in a shell.



(e) Gravitational instability now takes over, and new baryons and dark matter are attracted to the overdensities.



(f) The gas and dark matter peaks now look alike. The acoustic peak has decreased in contrast because dark matter, which has no peak initially, outweighs the gas.

Figure 2.1 – Evolution of the radial mass (fractional) profile versus comoving radius of an initially pointlike overdensity located at the origin. The units of the mass profile are arbitrary but are correctly scaled between the panels. These figures were made by suitable transforms of the transfer functions created by `CMBFAST` (Seljak and Zaldarriaga, 1996; Zaldarriaga and Seljak, 2000). Credits: Eisenstein et al. (2007).

fluctuations are usually studied using a decomposition of the temperature measurements into spherical harmonics:

$$T(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\theta,\phi) \tag{2.2}$$

where (θ,ϕ) are the equatorial angular coordinates on the sky and $Y_{lm}(\theta,\phi)$ the spherical harmonic functions with the corresponding amplitude a_{lm}. Y_{lm} quantifies the fluctuations on the scale $\Delta\theta \sim \pi/l$. Therefore, the 00 harmonic represents the average on the sky and a_{00} is thus the mean temperature of the CMB:

$$a_{00} = 2.725\,48 \pm 0.000\,57\,\mathrm{K} \tag{2.3}$$

as computed in Fixsen (2009) from core measurements. The next harmonic ($l = 1$) is dominated by the velocity of the solar system with respect to the CMB. This has been measured in Lineweaver et al. (1996) to be $a_{10} = 3.358 \pm 0.023\,\mathrm{mK}$ which corresponds to a speed of $369.0 \pm 2.5\,\mathrm{km\,s^{-1}}$.

To study the relative importance of the various harmonics, one can look at the variance of the temperature anisotropies, that can be written (since spherical harmonics are orthogonal to each other):

$$\left\langle \Delta T^2 \right\rangle = \frac{1}{4\pi} \sum_{l=2}^{\infty} \sum_{m=-l}^{l} |a_{lm}|^2 . \tag{2.4}$$

At this point, we introduce the coefficients C_l defined by

$$C_l = \left\langle |a_{lm}|^2 \right\rangle_m \tag{2.5}$$

where the average is computed on the $2l+1$ values of a_{lm}. The variance of the anisotropies can thus be rewritten

$$\left\langle \Delta T^2 \right\rangle = \frac{1}{4\pi} \sum_{l=2}^{\infty} (2l+1) C_l. \tag{2.6}$$

We define the power spectrum of the temperature fluctuations:

$$\Delta_T^2 = \frac{l(l+1)}{2\pi} C_l. \tag{2.7}$$

The map of the CMB anisotropies and the corresponding power spectrum obtained by the Planck collaboration is presented in figure 2.2 (Planck Collaboration et al., 2013). These anisotropies mostly come from the state of the Universe when photons decoupled from baryons. Three contributions must be used to describe the primary CMB anisotropies:

— density fluctuations,
— gravitational field perturbations,
— velocity anisotropies.

Before recombination and decoupling of the photons, the temperature of the photon-baryon plasma is directly linked to the density: overdense regions are intrinsically hotter than the average and underdense regions are cooler. When decoupling occurred, the blackbody radiation emitted by overdense regions had a slightly hotter temperature than the average, whereas underdense regions emitted a slightly cooler radiation.

The second contribution is linked to the perturbation of the gravity field induced by the density perturbation: to escape an overdense region a photon will undergo a gravitational redshift, whereas a photon escaping an underdense one will be blueshifted. This effect is known as

the Sachs-Wolfe effect, from the first two cosmologists to calculate it (Sachs and Wolfe, 1967). It breaks the direct link between the plasma density at the time of decoupling and the observed temperature of the CMB. This gravitational effect dominates over the intrinsic fluctuations, so that overdense regions, although intrinsically hotter, end up being seen colder than average. The Sachs-Wolfe temperature relation described above holds for the adiabatic perturbations we are considering here, but is not necessarily valid with more complex models such as isocurvature perturbations or topological defects.

The last contribution is a Doppler effect: because of the propagation of the acoustic waves in the plasma, there is a non-uniform velocity field. Hence, the velocity in the direction of the line of sight induces a redshift (if the perturbation is moving away) or a blueshift (if the perturbation is coming closer) of the emitted photon.

In addition to these effects that come from the state of the plasma when recombination occurs, photons of the CMB can be perturbed while travelling towards us, causing the so-called secondary anisotropies. For example, the formation of the first stars and galaxies around $z = 11$ reionised almost all of the neutral hydrogen of the intergalactic medium, causing some Thomson scattering. About 10% of CMB photons undergo a Thomson scatter before reaching us (Planck Collaboration et al., 2013). The thermal motions of large amounts of hot ionised gas, such as those found around clusters of galaxies, also affect the primary perturbations: this is the Sunyaev-Zel'dovich thermal effect.

It is worth noting that several softwares exist to calculate the power spectrum from a given cosmological model such as the Code for Anisotropies in the Microwave Background (`CAMB`) [2] (Lewis et al., 2000) that was built from `CMBFast` (Seljak and Zaldarriaga, 1996).

2.1.3 Matter Power Spectrum

2.1.3.1 Power Spectrum

As long as the fluctuations defined earlier are small ($\delta = \rho/\bar{\rho} - 1 \ll 1$), one can treat their evolution in Fourier space because each mode evolves independently from the others. We here consider the comoving coordinate $\mathbf{r}$. The transformation from Fourier space to real space is given by (dropping time dependence)

$$\hat{\delta}(\mathbf{k}) = \int \delta(\mathbf{r}) e^{-i\mathbf{r}.\mathbf{k}} \mathrm{d}\mathbf{r} \tag{2.8}$$

and the inverse transformation is

$$\delta(\mathbf{r}) = \frac{1}{(2\pi)^3} \int \hat{\delta}(\mathbf{k}) e^{i\mathbf{k}.\mathbf{r}} \mathrm{d}\mathbf{k}. \tag{2.9}$$

Since we are using comoving coordinates, the wave-vector $\mathbf{k}$ is also comoving, and in particular the physical length of its wavelength $\lambda_k(t)$ depends on the scale factor $a(t)$:

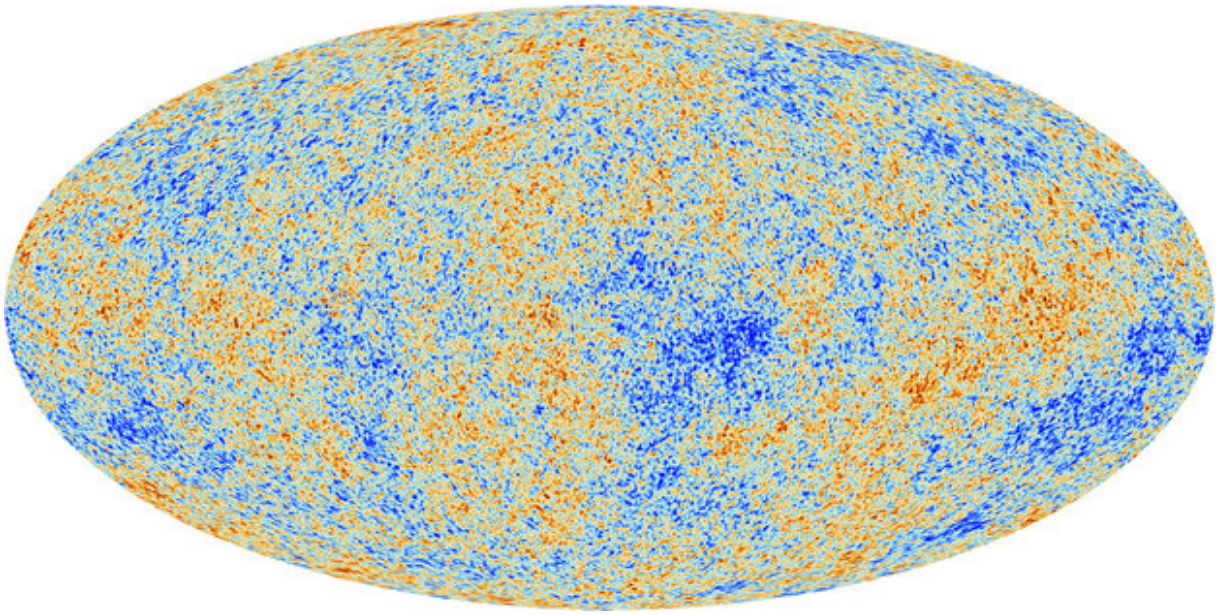
$$\lambda_k(t) = \frac{2\pi}{k} a(t) \tag{2.10}$$

where $k = |\mathbf{k}|$.

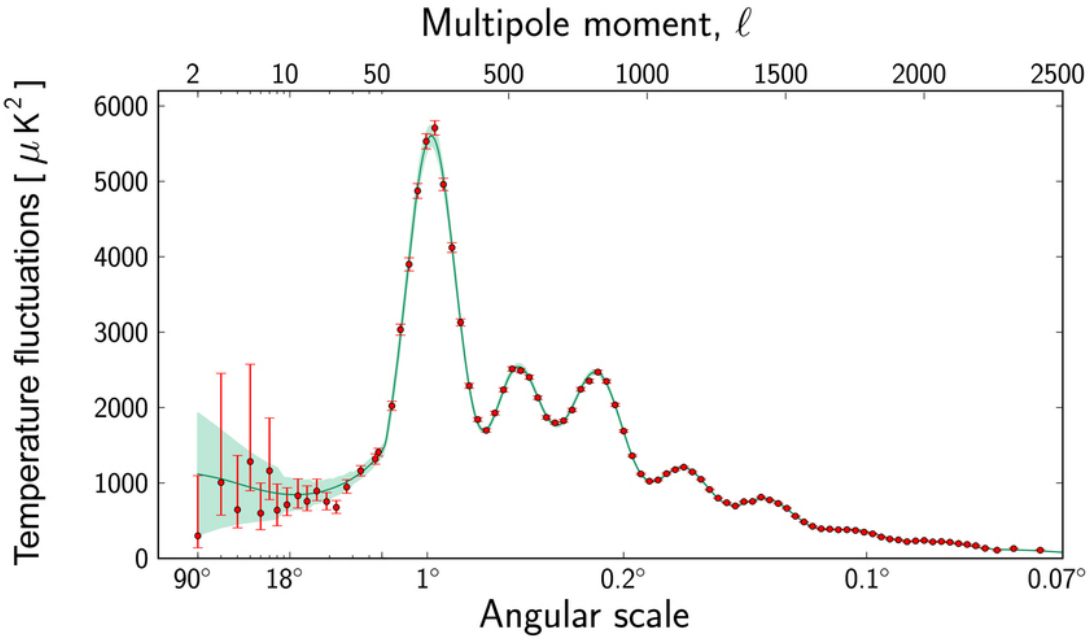
The 3D power spectrum of the fluctuations is then defined as

$$P_{3D}(\mathbf{k}) = \left| \hat{\delta}(\mathbf{k}) \right|^2 \tag{2.11}$$

2. `http://camb.info`



(a) The anisotropies of the Cosmic microwave background (CMB) as observed by Planck. Red indicates hotter regions, whereas blue indicate cooler regions. The center of the map, where the CMB is masked by the Milky Way is a reconstruction that does not affect the measured statistical properties of the anisotropies.



(b) Power spectrum of the temperature fluctuations detected by Planck at different angular scales on the sky, starting at ninety degrees on the left side of the graph, through to the smallest scales on the right hand side. The red dots are measurements made with Planck; shown with error bars. The green curve represents the best fit of the ΛCDM model to the Planck data. The pale green area around the curve shows the predictions of the variations of the ΛCDM model that best agree with the data.

Figure 2.2 – The Cosmic Microwave Background map and power spectrum by Planck. Credits: ESA and the Planck collaboration

which in the case of a statistically isotropic and homogeneous field is often reduced to

$$P_{3D}(k) = \langle P_{3D}(\mathbf{k}) \rangle \tag{2.12}$$

where the average is taken on all the modes $\mathbf{k}$ so that $|\mathbf{k}| = k$. One can then link the variance of fluctuations in real space to this power spectrum through

$$\left\langle \delta^2(\mathbf{r}) \right\rangle = \frac{1}{(2\pi)^3} \int \left| \hat{\delta}(\mathbf{k}) \right|^2 \mathrm{d}\mathbf{k} = \frac{1}{2\pi^2} \int_0^\infty k^2 P_{3D}(k) \mathrm{d}k \tag{2.13}$$

The power spectrum $P_{3D}(k)$ has the dimension of a volume and to work with dimensionless quantities, the 3D spectrum of relative fluctuations $\Delta_{3D}(k)$ is often defined with

$$\Delta^2_{3D}(k) = \frac{k^3 P_{3D}(k)}{2\pi^2}. \tag{2.14}$$

It is related to the variance of δ by

$$\left\langle \delta^2(\mathbf{r}) \right\rangle = \int_0^\infty \frac{\Delta^2_{3D}(k)}{k} \mathrm{d}k. \tag{2.15}$$

When the field $\delta(t, \mathbf{r})$ can only be measured in one direction x, usually the line of sight direction as with the Lyman-α forests (see section 2.3), we can define the 1D power spectrum $P_{1D}(k_\parallel)$:

$$P_{1D}(k_\parallel) = \frac{1}{2\pi} \int_0^\infty \delta(x)^2 e^{-ik_\parallel x} \mathrm{d}x \tag{2.16}$$

where $k_\parallel$ is the radial component of the wavevector (and $\mathbf{k}_\perp$ its tranverse component). The 1D and 3D power spectra are related by

$$P_{1D}(k_\parallel) = \frac{1}{(2\pi)^2} \int P_{3D}(k_\parallel, \mathbf{k}_\perp) \mathrm{d}\mathbf{k}_\perp. \tag{2.17}$$

or

$$P_{1D}(k_\parallel) = \frac{1}{2\pi} \int_{k_\parallel}^\infty k P_{3D}(k) \mathrm{d}k \tag{2.18}$$

It is also worth noting that the power spectrum is related to the two point correlation function $\xi(r)$ that describes the excess probability, compared with a random distribution, of finding two objects or overdensities at distance r:

$$\xi(r) = \langle \delta(\mathbf{r}_1) \delta(\mathbf{r}_2) \rangle_r \tag{2.19}$$

where $r = |\mathbf{r}|$ and the average is done on all the couples of points $\mathbf{r}_1$, $\mathbf{r}_2$ so that $|\mathbf{r}_2 - \mathbf{r}_1| = r$. The two functions are related by

$$\xi(r) = \frac{1}{2\pi^2} \int_0^\infty k^2 \frac{\sin(kr)}{kr} P_{3D}(k) \mathrm{d}k \tag{2.20}$$

or inversely

$$P_{3D}(k) = \frac{1}{2\pi^2} \int_0^\infty r^2 \frac{\sin(kr)}{kr} \xi(r) \mathrm{d}r. \tag{2.21}$$

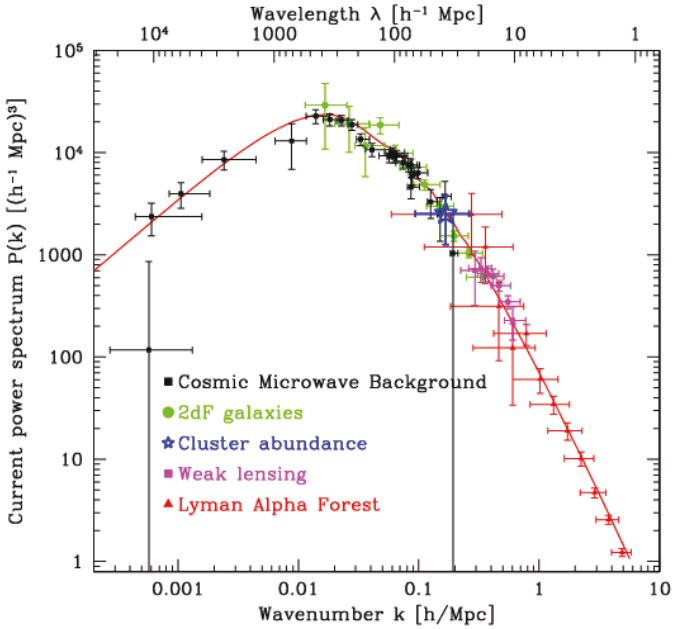


Figure 2.3 – Linear matter power spectrum $P(k)$ versus wavenumber, from various measurements of cosmological structures. The solid line is the best fit from a ΛCDM model. Credits: Tegmark and Zaldarriaga (2002)

This expression explains why the acoustic peak of the correlation function is seen as oscillations in the power spectrum.

The power spectrum of matter density fluctuations has now been measured with great accuracy. Figure 2.3 (Tegmark and Zaldarriaga, 2002) shows some measurements using CMB anisotropies, galaxy large scale structure from 2dF, weak lensing of galaxy shapes, the Lyman-α forest, and a single point for galaxy clusters at $8\,\mathrm{Mpc}\,h^{-1}$.

2.1.3.2 Bias and Redshift Distorsion

Measurement of the matter power spectrum is an alternative to the CMB to put constraints on cosmological parameters. However, it is quite difficult to precisely measure the density of matter as most of it is in the form of dark matter that we cannot see directly. What is currently done is to measure the distribution of baryonic tracers which is assumed to be the same as the distribution of matter. In addition, to measure the baryon distribution one must use tracers such as galaxies, quasars or Lyman-α forests. However, the distribution of these observables does not follow exactly the baryon distribution and the difference between them is modelled by a bias b such that

$$\frac{\Delta n_{\mathrm{observable}}}{n_{\mathrm{observable}}} = b\frac{\Delta\rho}{\rho} \tag{2.22}$$

where $n_{\mathrm{observable}}$ is the density of the tracer under study and ρ the matter density. This bias could be a function of many parameters like the environment, the type of galaxy, the distance... but it is so difficult to estimate that it is usually taken as a fixed scalar. This implies that the systematic errors when measuring the matter power spectrum are often dominated by the estimation of the bias.

When measuring the power spectrum, one must also take into account another effect called redshift distortion. It comes from the use of redshift to determine the radial position of an object. The redshift of an object is the sum of a cosmological redshift due to the Universe expansion and a Doppler spectral shift (blueshift or redshift) due to the peculiar velocity of the object. It induces an anisotropy of the power spectrum that is characterized by a parameter β introduced in Kaiser (1987):

$$P(\mathbf{k}_\perp, k_\parallel) = (1 + \beta\mu_k^2)^2 P(k) \tag{2.23}$$

where μ_k is the cosine of the angle between the line of sight and the wavevector $\mathbf{k}$.

2.2 Quasi Stellar Objects

Before introducing the Lyman-α forest, which will be our matter tracer in this work, it is necessary to present the objects that will serve as a background light: the quasars. I will go through their main characteristics without going into deep details as it is not relevant for this work.

Quasi stellar objects, or quasars, or QSOs, are believed to be powered by the accretion of matter onto super massive black holes at the center of galaxies, a process that can emit more energy than thermonuclear reactions. Today, quasars are considered to be the most luminous members of the general class of objects called active galactic nuclei, or AGNs. They are also the most luminous objects in the universe.

The main observed properties of quasars are the following:

— nuclei that appear starlike in optical images. Extended emission can now often be detected around the nucleus, and jets extending away from the nucleus are occasionally seen;

— spectra showing broad emission lines with widths greater than observed in galaxies;

— radiation of similar intensity at energies ranging from γ-rays to the far-infrared (100 µm) region of the spectrum, with a decrease in intensity at radio frequencies;

— redshifts ranging from 0.1 to 7;

— variability in brightness at different wavelengths on timescales as short as days or weeks;

— luminosities as high as 10^{14} solar luminosities.

2.2.1 Nature

The prodigious luminosities of quasars in combination with their small size led theorists to consider gravity as their energy source immediately upon their discovery. However, the required gravitational potential energy can only be extracted if quasars contain compact objects with masses of a hundred million Suns and sizes about that of the solar system. In addition, the gravitational field of such objects is so strong that the effects of general relativity are dominant in their vicinity. Since their discovery, gravity has remained as the consensus source of energy for quasars, although the concepts of the nature of the central source have changed considerably. For example, “supermassive stars”, were originally postulated as the central objects in quasars, but they were quickly abandoned for black holes.

Today, the working hypothesis for quasars and AGNs is that they are powered by massive (10^6–$10^9\,\mathrm{M}_\odot$) black holes. The radiated energy we detect from them comes from matter being accreted onto the black hole. The radii of black holes in AGNs and quasars range from a few solar radii to 20 astronomical units.

Figure 2.4 – Artist's view of a quasar. Credits: ESO/M. Kornmesser

It is most likely that much of the matter surrounding the black hole is in the shape of a disk. The matter in the disk orbits the black hole and moves inward as it loses angular momentum from some source of viscosity in the disk. The gas in the inner regions of the disk is expected to be hot enough, about 10^4–10^6 K, to account for the thermal component of the continuum radiation at ultraviolet wavelengths.

2.2.2 Redshifts

The large redshifts of quasars are one of their distinctive features. The value of the greatest known redshift is $z = 7.085$ (Mortlock et al., 2011) and the Baryon Oscillations Sky Survey (see chapter 3) counts 45 QSOs with a redshift greater than 5 (Pâris et al., 2014). However, most known quasars have redshifts less than 2.5. This occurs for two reasons: such quasars are easier to find because they are brighter, and there are intrinsically more quasars with $z < 2.5$ (Palanque-Delabrouille et al., 2013). The lowest redshift limit for quasars is normally set at 0.1 because objects at lower redshift are usually resolvable as galaxies and are catalogued as active galactic nuclei (AGN).

In addition to providing our first view of the universe at $z > 5$, quasars also proved to be valuable cosmological probes because they are luminous enough to yield clues on the properties of matter along the line of sight (see sections 2.2.5 and 2.3).

2.2.3 Variability

Most quasars are believed to show variations in light at the 10–40% level over timescales of days to years. Furthermore, variations are observed in the strength and shape of emission lines and in continuum emission at x-ray, ultraviolet and radio wavelengths.

This variability of the continuum emission of the quasars has been known since their discovery (Matthews and Sandage, 1963). It has been used to both discover some quasars (van den Bergh et al., 1973; Rengstorf et al., 2006) or select them for surveys (Geha et al., 2003; Dobrzycki et al., 2003; Palanque-Delabrouille et al., 2011).

Observations of the variability of quasars carried out at different wavelengths have provided some of the most direct information on the nature of the inner regions in quasars and AGNs (Kawaguchi et al., 1998; Trevese et al., 2001). These programmes have shown that the emission line regions are considerably smaller and closer to the central source than was originally thought.

2.2.4 Continuum Emission and Emission Lines

Quasars emit radiations from γ-rays and x-rays to the far infrared (100 µm). For most of them, the amount of energy emitted in each band is remarkably similar, in contrast to thermal radiation from stars, which is much more peaked and restricted in wavelength. Most quasars at $z \leq 2.5$ are bright at ultraviolet wavelengths, a property that is very helpful to distinguish them in sky surveys from the more numerous stars which are usually faint at these wavelengths.

Although quasars were discovered through their radio emission, about 90% of quasars do not emit strongly at radio wavelengths and are classified as radio-quiet. Radio-loud quasars are typically 100 times brighter at radio wavelengths than radio-quiet quasars.

The continuum emission in quasars appears to arise from a combination of thermal and non-thermal processes. In any event, the continuum radiation from quasars demonstrates that some very energetic processes are involved. Furthermore, the continuum radiation at the highest energies tends to show the most variability and the shortest timescales, which is another indication of the extreme conditions that exist near quasars.

The spectra of quasars are characterized by the large widths of the emission lines and wide range of ionization. The widths are produced by motions of gas in the emitting region, where a mixture of infall, rotation and ejection probably occurs. The widths are consistent with the emission region being at a distance of light months to a few light-years from a central black hole.

The strongest emission lines in quasar spectra come from hydrogen, carbon and magnesium, with lines of nitrogen, oxygen, iron and other elements also being visible. The observed levels of ionization range from neutral for hydrogen and oxygen to five times ionized oxygen and even more highly ionized iron. An example of a QSO spectrum is given in figure 2.5.

2.2.5 Absorption Lines

Absorption lines were detected in quasars within a few years of their discovery and were initially believed to be rare. Today, the study of quasar absorption lines is a major topic.

Absorption lines may be placed in three categories according to the distance of the absorbing gas from the central emission source of the quasar:

— intrinsic systems,
— associated systems,
— intervening systems.

Intrinsic systems arise from the quasar itself, mainly its accretion disk or jets. The broad absorption line (BAL) quasars, whose spectra show broad absorption lines on the shorter-wavelength side of the emission lines, are an example of such absorbing systems. The width of the absorption features indicates outflow velocities that can exceed $30\,000\,\mathrm{km\,s^{-1}}$. An example is given in figure 2.6a.

Associated systems show absorption features from elements such as hydrogen, carbon and magnesium. Such systems have redshifts close to that of the emission lines in quasars but they produce absorption lines that are narrower. They are believed to arise in gas associated with the host galaxy of the quasar or of the environment in which the host galaxy resides.

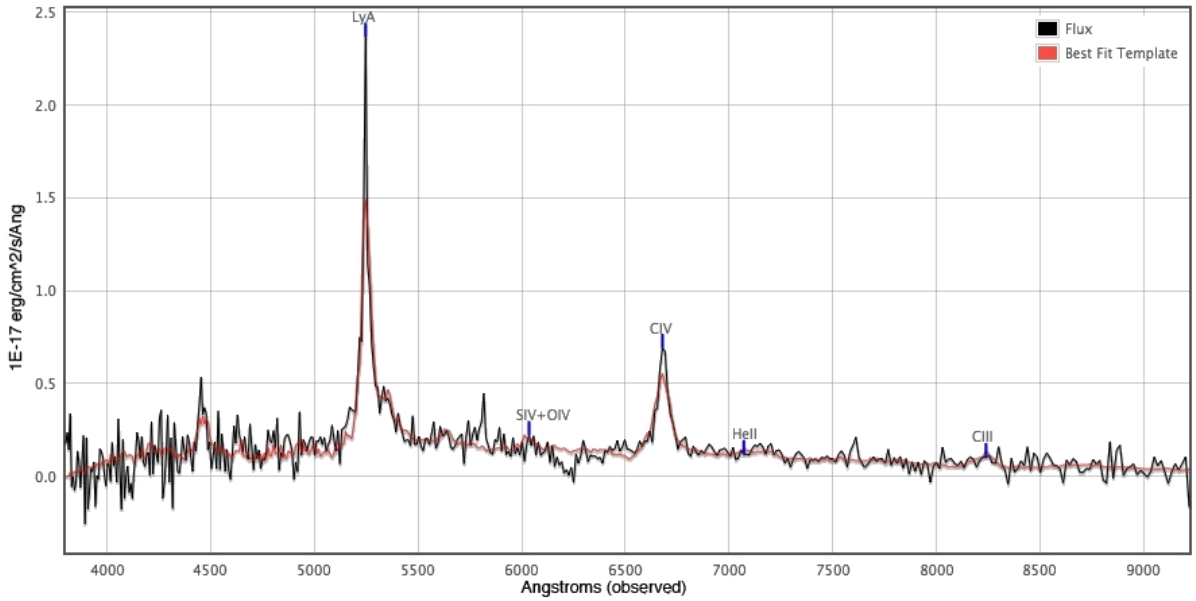


Figure 2.5 – A BOSS quasar spectrum (J034424.41+003724.2, $z = 3.317$) with the corresponding emission lines.

Intervening systems have redshifts smaller than the emission lines in quasars and arise in clouds of gas unrelated to the quasar that lie along the line of sight. In this case, the quasar serves as a background beacon that enables the study of the gas. Studies of absorption lines in quasars have provided crucial information about gas in the early universe.

Intervening systems may themselves be classified in three groups according to the column density of the absorbing gas:

1. damped Lyman-α systems (DLA),

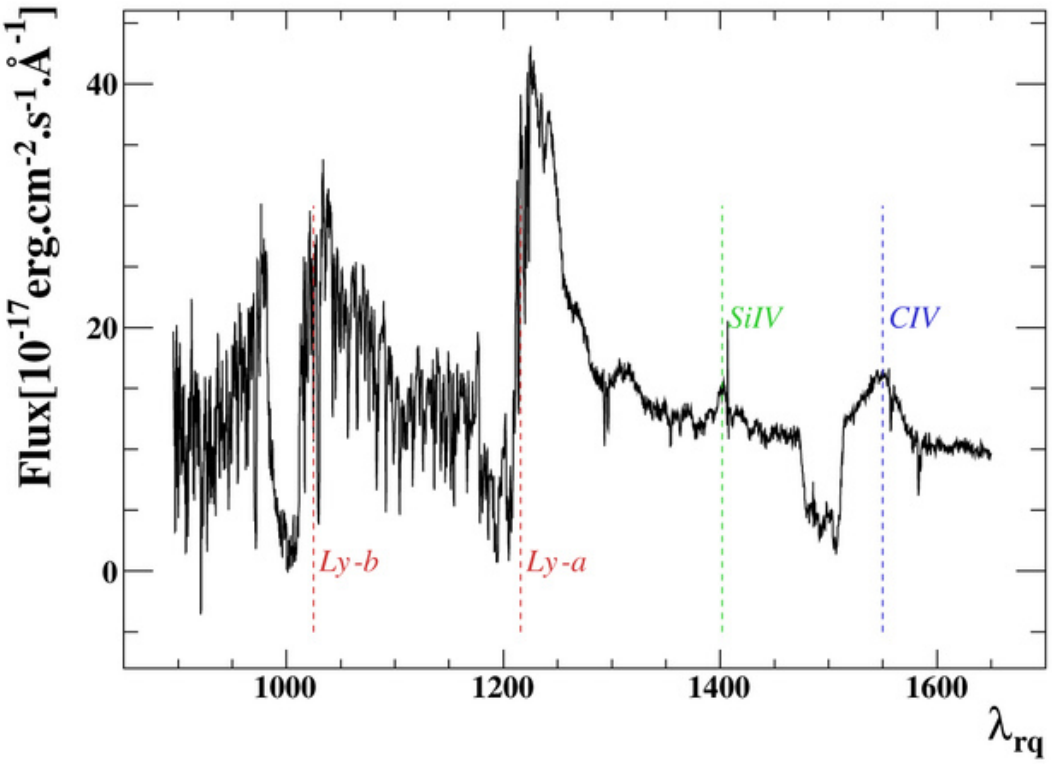
2. intermediate systems,

3. the Lyman-α forest.

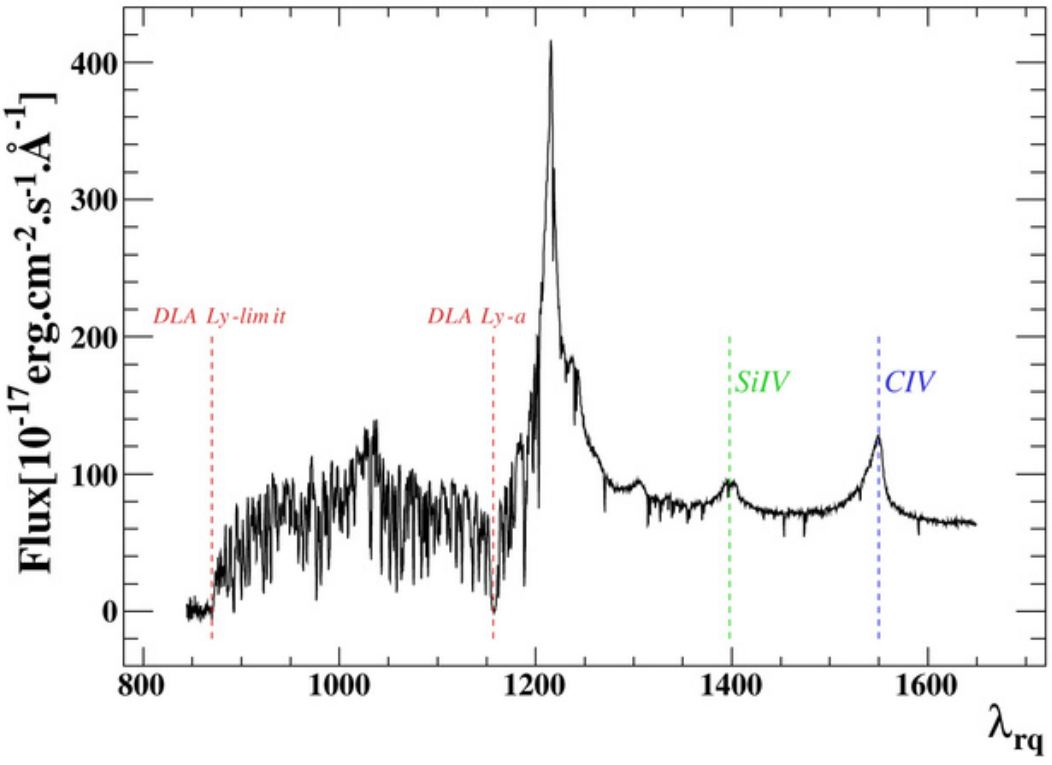
Damped Lyman-α systems represent the highest column densities, typically in excess of 10^{20} neutral hydrogen atoms per cm^2, and are so called because a wide saturation in the absorption is observed. Such systems occur for instance when the line of sight passes directly through a galaxy. An example is shown is figure 2.6b.

Intermediate systems, with a characteristic column density of 10^{17} neutral hydrogen atoms per cm^2, also show absorption lines from elements such as carbon, magnesium and other metals. They are commonly attributed to the halos of galaxies.

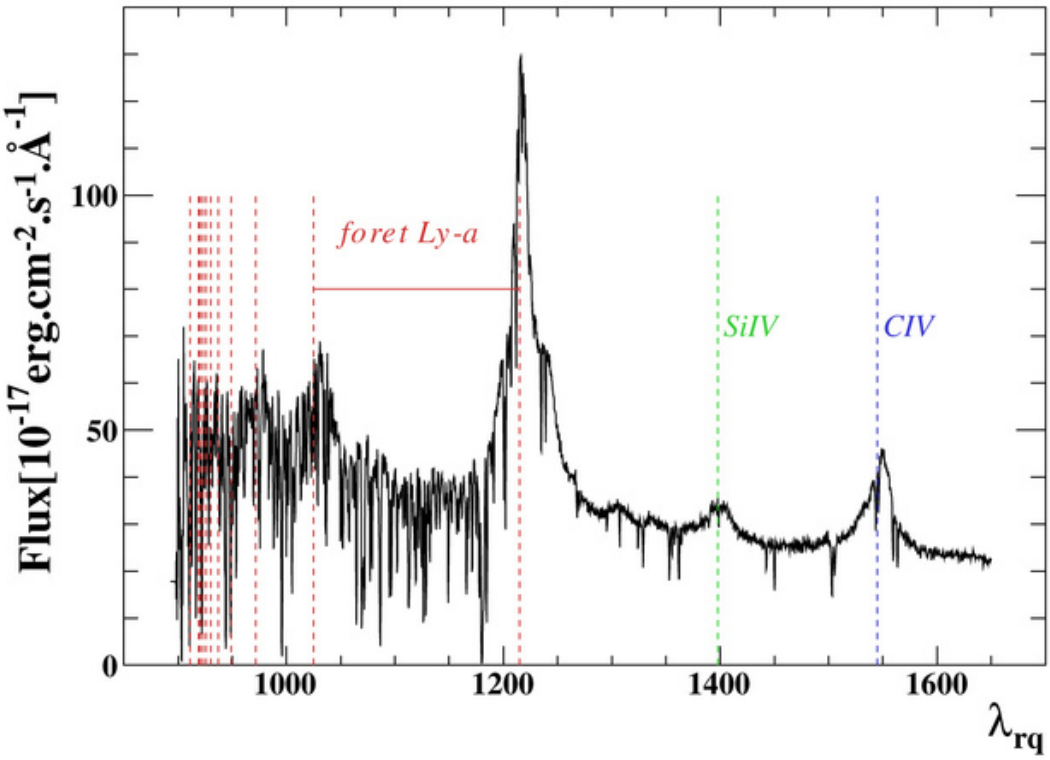
The Lyman-α forest lines are the most numerous and ubiquitous absorption features and are attributed to clouds of gas in the intergalactic medium. They are seen in a range of column densities beginning at the limit of detectability around 10^{12-13} neutral hydrogen atoms per cm^2 and extending up into the range of the intermediate systems. They are treated in more details in the next section. An example is shown in figure 2.6c.



(a) Broad absorption line system (BAL) spectrum. Characteristic absorptions are clearly visible on the left of C IV, Lyman-α and Lyman-β emission lines.



(b) Damped Lyman-α system (DLA) spectrum



(c) "Regular" Lyman-α forest spectrum. The dotted red lines represent the Lyman series from Lyman-α at 1215.67 Å to the Lyman continuum at 911.3 Å.

Figure 2.6 – BOSS quasar spectra of a broad absorption line (BAL) system, a damped Lyman-α (DLA) system and a "regular" Lyman-α forest. Spectra are represented using the wavelength in the quasar rest frame $\lambda_{\rm rq}$ and they all present high average signal to noise ratio with respect to other BOSS spectra.

2.3 The Lyman-α forest

2.3.1 The Gunn-Peterson Effect

The Lyman-α forest is an absorption phenomenon seen in the spectra of high redshift QSOs. It is the only direct observational evidence we have of the existence and properties of the general intergalactic medium.

On its way to us, the light of a bright, distant QSO passes through intervening intergalactic gas. Absorption by the gas modifies the spectra of the background QSO and imprints a record of the gas clouds physical and chemical states on the observed spectrum. It is like a giant cosmic slide projector, where a QSO plays the role of the light bulb, and the intervening gas clouds are the slides, changing the colors of the light source by absorbing parts of the (white) spectrum. Lyman-α forest absorption in a QSO spectrum was predicted and first detected by Gunn and Peterson (1965). An illustration of the "building" of the Lyman-α forest is shown in figure 2.7.

The name "Lyman-α forest" refers to the appearance of the optical QSO spectra, which show hundreds of sharp absorption lines, mostly from the neutral hydrogen Lyman-α line at $\lambda_{\rm Ly\alpha} = 1215.67$ Å, on the otherwise rather smooth QSO continuum. It starts below the Lyman-α emission peak and most studies use only the region between the Lyman-α and the Lyman-β to avoid the presence of multiple Lyman absorptions. The absorption systems spread out

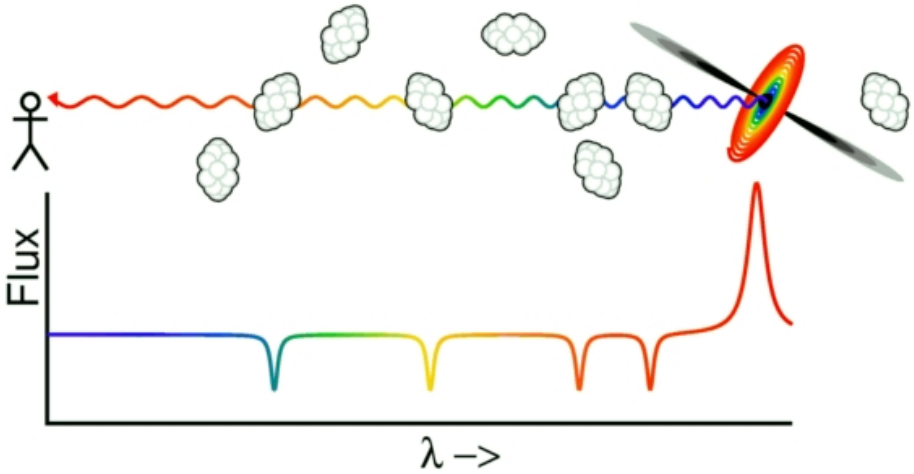


Figure 2.7 – Illustration of the making of the Lyman-α forest along a line of sight. Credits: Edward L. Wright

into a "forest" of lines because each line is redshifted by a different amount in proportion to the absorbing cloud's distance from us. Let us consider a quasar at redshift $z_q = 3$ and a cloud of intergalactic hydrogen on the line of sight at $z_c = 2.8$. The Lyman-α emission peak is observed at $\lambda_e = \lambda_{\mathrm{Ly}\alpha}(1 + z_q) = 4860\,\text{Å}$ but the Lyman-α absorption is observed at $\lambda_a = \lambda_{\mathrm{Ly}\alpha}(1 + z_c) = 4617\,\text{Å}$, which is on the left of the emission peak. An example of a high-resolution spectrum is shown in figure 2.8.

Clouds with H I column densities larger than $10^{17}\,\mathrm{cm}^{-2}$ show a discontinuity due to continuous absorption at a rest frame wavelength of 912 Å, beyond the limit of the Lyman series. These "Lyman limit systems" occupy a column density regime where a gas cloud starts shielding itself against ionizing radiation from the outside. Clouds with even higher column densities ($N > 10^{19}\,\mathrm{cm}^{-2}$) are the "damped Lyman-$\alpha$" systems where the gas is almost completely self-shielded and mostly neutral.

Lyman-α forest absorption systems have now been observed from redshift zero (with UV satellites) up to the highest redshifts at which background light sources can still be found (currently $z \sim 5 - 7$).

One can also define other Lyman forests like the Lyman-β forest which starts below the Lyman-β emission peaks and corresponds to absorption at 1025 Å. One should note that of course Lyman-α absorption from nearer absorbers is also present in the Lyman-β forest due to more distant absorbers.

2.3.2 Cosmology with the Lyman-α Forest

The Lyman-α forest presents a clear interest for cosmology because it provides information on the content of the intergalactic medium in neutral hydrogen. Hydrogen, neutral and ionised, is assumed to be the major constituent of the intergalactic medium. The observed flux f_{obs} in the Lyman-α forest is the fraction the flux f_e emitted by the quasar that is not absorbed by neutral hydrogen of the intergalactic medium. Therefore we can define the transmitted flux fraction F as

$$F = \frac{f_{obs}}{f_e}. \tag{2.24}$$

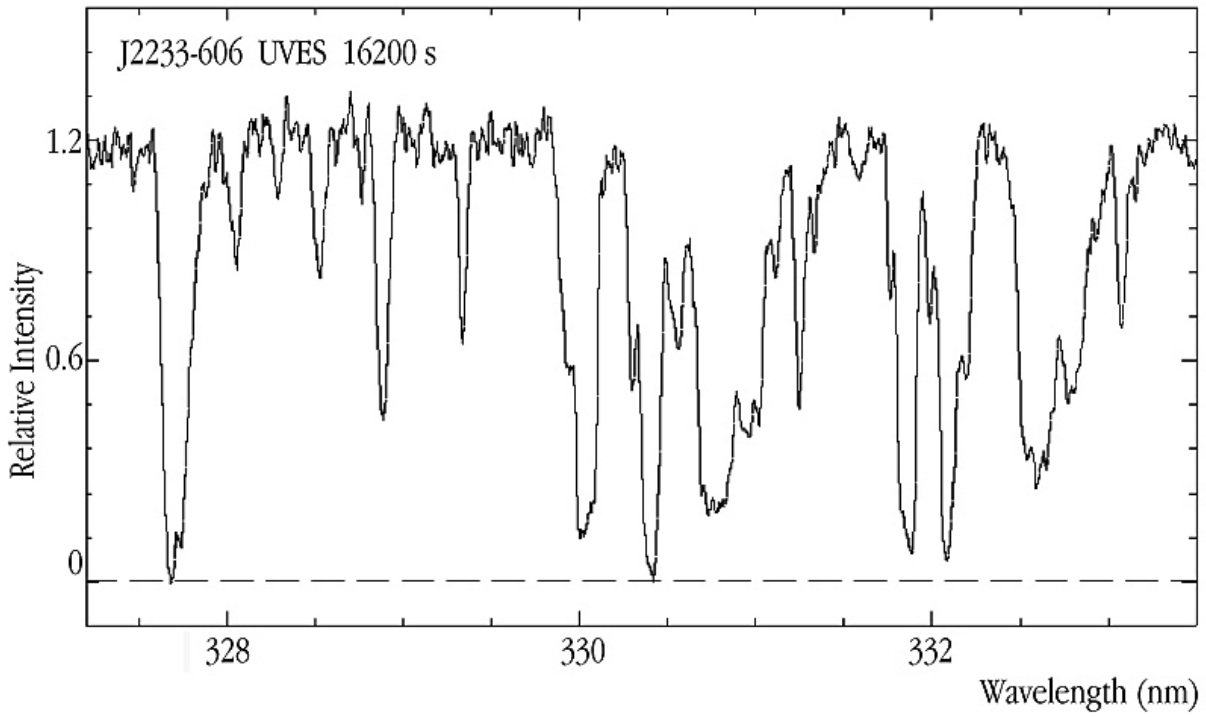


Figure 2.8 – Section of the high-resolution UV spectrum of the quasar J2233-606 in the region of the Lyman-α forest at redshift $z \sim 1.7$, obtained by the Ultraviolet and Visual Echelle Spectrograph (UVES) on the Very Large Telescope (VLT). Credits: ESO

By definition it is also given by

$$F = e^{-\tau} \tag{2.25}$$

where τ is the optical depth. The development of hydrodynamic simulations (Cen et al., 1994; Zhang et al., 1995; Hernquist et al., 1996; Croft et al., 1997) and analytical models (Bi, 1993; Miralda-Escude and Rees, 1993; Hui and Gnedin, 1997) allowed to show that this optical depth is proportional to the density of neutral hydrogen ($\tau \propto n_{\mathrm{H\,I}}$). The density is itself a tracer of the baryon density ρ_b through the relation $n_{\mathrm{H\,I}} \propto \rho_b^{\beta}$ with $\beta \sim 1.5 - 2.0$. The previous equation thus becomes

$$F = e^{-k\rho_b^{\beta}} \tag{2.26}$$

where k is the proportionality constant between the optical depth τ and the baryon density ρ_b. However, because this coupling is non-linear and the fact that the constant k is not well-known, it is usually impossible to evaluate the baryon density.

The cross section of the Lyman-α transition is very high and the fact that the absorption is not completely saturated comes from the high ionisation of the intergalactic medium (Gunn and Peterson, 1965). This ionisation comes from the UV photon flux generated by star forming galaxies and active galactic nuclei and thus depends on the redshift. Moreover, the Universe expansion induces a diminution of the hydrogen mean density with time, proportional to $(1+z)^3$. These two dependencies make the Lyman-α forest bias dependent on the redshift.

Using this link, Croft et al. (1998) have suggested a technique for recovering the initial power spectrum of density fluctuations directly from the fluctuations of the optical depth measured from Lyman-α forest spectra. One can use the Lyman-α forest to constrain cosmological parameters, for example the Ωs or $H(z)$ seen in chapter 1. On the one hand, one must use QSO

spectra to obtain the transmitted flux power spectra of Lyman-α forests and recover the power spectrum of density fluctuation. On the other hand, numerical simulations with different input parameters must be run to extract the power spectrum of mass density fluctuations. Then one can compare the measured power spectrum with the simulated ones, inferring the best set of input parameters to reproduce the observations. This work is based on this idea and chapters 4 and 6 will present in detail the recovery of the density fluctuation power spectra from the BOSS QSO spectra and the grid of hydrodynamical simulations used for cosmological constraints.

Chapter 3

The Baryon Oscillation Sky Survey

"Make no small plans. Dream no small dreams."
—George Ellery Hale, 1868–1938

THE spectroscopic survey BOSS is originally dedicated to the measurement of the baryon acoustic oscillation in both the spatial distribution of galaxies and the Lyman-α forests. However, thanks to the fabulous amount of data and their quality, a lot of other projects have been achieved and will be achieved with it. This chapter details the main characteristics of the survey: the strategy, the instruments and the data reduction. In addition to technical publications this can be found on the SDSS[1] and SDSS-III[2] websites.

3.1 The Sloan Digital Sky Survey (SDSS)

3.1.1 SDSS and SDSS-II

With the original goal to get a better unerstanding of the large scale structure of the Universe, the Sloan Digital Sky Survey was imagined in the eighties. This large and deep sky survey has been made possible thanks to the quick development of CCD sensors along with the fast growing data processing capabilities. This project led to the construction of a dedicated 2.5m Ritchey-Chrétien telescope at the Apache Point Observatory (New Mexico, USA). Sitting at an elevation of 2788 metres, the telescope started to observe in 2000. More details about the telescope can be found in section 3.3.1. The first two generations of SDSS (SDSS and SDSS-II) were completed respectively from 2000 to 2005 and 2005 to 2008. The results were a photometric survey of 11 663 $\deg^2$ in five wavelength bands along with a spectroscopic survey of 1 640 960 spectra. This includes more than 930 000 galaxies, 120 000 quasars and 460 000 stars. The combination of the astrometric data of the photometric survey and the redshifts measured with the spectroscopic data, made possible the realization of a three dimensional map of a region of the Universe. All the data were made available to the scientific community with the seventh data release (DR7) and can be found on the SDSS website [3]. The study of this tremendous amount of data had

1. `http://www.sdss.org`
2. `http://www.sdss3.org`
3. `http://www.sdss.org/dr7`

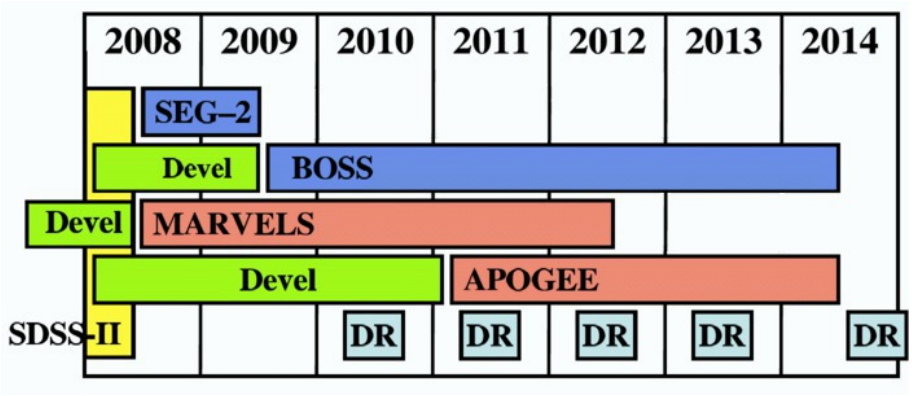


Figure 3.1 – High-level SDSS-III schedule. Dark-time observing programmes are marked in blue and bright-time observing programmes in red. Hardware development activities are marked in green, and light blue squares represent the five planned public SDSS-III data releases.

a big impact on our knowledge in both astrophysics and cosmology. At this time, more than 5000 publications mention SDSS in their title (including SDSS-III) and they share more than 200 000 citations. One of the most famous results from SDSS is the first ever detection of baryon acoustic oscillations by Eisenstein et al. (2005) using a sample of around 46 000 luminous red galaxies (LRG). This allowed a measurement of absolute distance at $z = 0.35$ with a precision of 5%.

3.1.2 SDSS-III

The third generation of SDSS started in Autumn 2008, directly following SDSS-II. Using the same telescope, it is built on four different surveys: The Apache Point Observatory Galactic Evolution Experiment (APOGEE), the Multi-object APO Radial Velocity Exoplanet Large-area Survey (MARVELS), the Sloan Extension for Galactic Understanding and Exploration 2 (SEGUE-2) and the Baryon Oscillation Spetroscopic Survey (BOSS). Figure 3.1 presents the high-level schedule of the four SDSS-III surveys and their development activities. Whereas BOSS will be presented in greater details in the next section, I will here briefly describe the other three surveys.

3.1.2.1 SEGUE-2

The Sloan Extension for Galactic Understanding and Exploration 2 produced spectra for almost 118 000 stars inside the galactic halo of our galaxy. With apparent magnitude below 19, these stars belong to different stellar populations but they are all between 10 kpc and 60 kpc of the galactic centre. Combined with the already taken 230 000 spectra from SEGUE-1, the data of SEGUE-2 highlighted the complexity of the kinematic and chemical substructures of the stellar halo of the Milky Way, giving us insight on the formation and metal enrichment of our galaxy. SEGUE-2 used the telescope dark time between Autumn 2008 and Spring 2009, ending before the start of BOSS in Autumn 2009. Spectra taken with the SEGUE-2 spectrograph cover the range 385−920 nm, with a mean resolution of 2000 and an average signal-to-noise ratio of 25.

3.1.2.2 APOGEE

The APO Galactic Evolution Experiment is a survey that aims at observing around 100 000 red giant stars with apparent magnitude below 12.5 and located in different regions of the Milky Way (bulb, disk, bar and halo). Measuring with high precision the peculiar velocity and the chemical composition of these stars, APOGEE studies both the dynamic and the chemical history of our galaxy. The observations started in Spring 2011 and ended in Spring 2014. It is using the "bright" time –when the Moon is more than 60% illuminated– that BOSS cannot use. Because it observes much brighter stars, the APOGEE spectra are much better than those of SEGUE-2: their average signal-to-noise ratio is around 100 and the typical resolution is 20 000.

3.1.2.3 MARVELS

The Multi-object APO Radial Velocity Exoplanet Large-area Survey was a spectroscopic survey designed to observe 11 000 bright stars of our galaxy. Each star was to be observed between 25 and 35 times on an eighteen months period, looking at their radial velocity to look for giant gaseous exoplanets. Started in Autumn 2008, and observing during "bright" time, the goal of the survey was to constrain theoritical models of formation and evolution of giant planets systems. However, the required resolution of the spectra was never reached and the project was stopped in 2012.

3.2 BOSS overview and strategy

3.2.1 Overview

The first goal of BOSS is the observation of the BAO, with enough sensitivity to measure it in both the radial and transverse direction, thus allowing independent measurement of the Hubble parameter $H(z)$ and the angular distance $D_A(z)$. To achieve this, BOSS carries on two spectroscopic surveys on more than $10\,000\,\mathrm{deg}^2$ of the sky.

The first survey consists in the observation of 1.5 million galaxies with redshift ranging from 0.15 to 0.7 and a mean density of $150\,\mathrm{deg}^{-2}$. This survey can itself be divided in two samples: one called LOW-Z for galaxies with $0.15 \leq z \leq 0.43$ which aims at completing the original SDSS and SDSS-II surveys; and a second one called CMASS at higher redshift ($0.43 \leq z \leq 0.7$) with a mean density of $120\,\mathrm{deg}^{-2}$. For the first time ever, the precision obtained with CMASS will be limited by cosmic variance (the fact that we have only one universe to observe) and not shot noise.

The second survey aims at observing at least 150 000 quasars at high redshift ($2.15 \leq z \leq 3.5$) with a surface density of $17\,\mathrm{deg}^{-2}$. This should allow us to probe the intergalatic medium through the study of the Lyman-α forests along the quasars lines-of-sight with enough precision to compute the correlation function of neutral hydrogen with a 3% precision.

In addition to these two main surveys, BOSS allocates 5% of its observable objects to minor scientific projects called ancillary programmes. They are proposed by SDSS-III members and they have dealt with very different subjects such as high energy blazars, very low mass stars or quasar selection using their variability.

3.2.2 Observations

Thanks to very good weather conditions, BOSS is ahead of schedule and is finishing its last observations at this time (February 2014). This period of spectra acquisition was preceded by a photometric phase to complete the original SDSS and SDSS-II data. The resulting image of the sky was the base for the selection of the target for the spectroscopic stage. The main characteristics of these two phases are presented here.

3.2.2.1 Photometry

From 1998 to 2005, SDSS took images of $11\,600\,\mathrm{deg}^2$ of the northern hemisphere sky, including a contiguous portion of $7600\,\mathrm{deg}^2$ in the northern galactic cap (NGC). This was done in five wavelength bands (called u', g', r', i', z') with a dedicated 54 CCDs camera described in section 3.3.2. During Autumns 2008 and 2009, BOSS reused the same camera to cover around $3100\,\mathrm{deg}^2$ in the southern galactic cap (SGC). The resulting image, presented in figure 3.2 covers a total of $14\,555\,\mathrm{deg}^2$ and includes more than 932 million individually detected objects. All the photometric data were made public with the eighth data release (DR8) and can be accessed via the corresponding website [4].

All the photometric data were taken under "photometric conditions" (no clouds and an atmospheric extinction depending only on the airmass), with the Moon under the horizon and a seeing better than $2''$ in the r' band. A technique called time-delay integration or drift scanning was used. In this observing technique, the telescope stands still pointing in a given direction while the sky passes under the effect of the rotation of the Earth. The camera thus integrates the light of an object from the moment it enters the field of view to the moment it leaves it, the CCDs being read out at the sideral rate as the sky drifts by. This allows to use more than 90% of the available observing time for actual observations. The exposure time for each object is 55 s in each band. Because of the very large field of view of the telescope (3°) it is necessary to operate the drift scanning along great circles to avoid transit-time differences across the imaging CCD array.

3.2.2.2 Spectroscopy

Because taking a spectrum requires a substantially longer exposure time, it is not possible to have a spectrum for every object that had been identified in the photometric survey. Thus the first step of the spectroscopic survey is to select the targets for which one wants spectra. The second step is the tiling, which aims at optimizing the observations of the selected targets in order to minimize the total observation time and thus the overall duration of the project. Then, it is possible to start the observations. This section presents these three steps.

The BOSS quasars selection was done in a limited portion of the SDSS photometric survey, representing $7578\,\mathrm{deg}^2$ in the NGC and $2663\,\mathrm{deg}^2$ in the SGC. The main reason for this reduction is to avoid the galactic plane, where identifying quasars is more diffcult. However, thanks to very good weather conditions during the survey, the observations are ahead of planning and ended in February 2014. The leftover time is going to be used for additional ancillary programmes.

To obtain the desired precision on the BAO measurement with Lyman-α forest an average density of $15\,\mathrm{deg}^{-2}$ is required. This corresponds to a limit magnitude of 22 in the g band (McDonald and Eisenstein, 2007; McQuinn and White, 2011). Unlike galaxies that are approaching

4. `http://www.sdss3/dr8/`

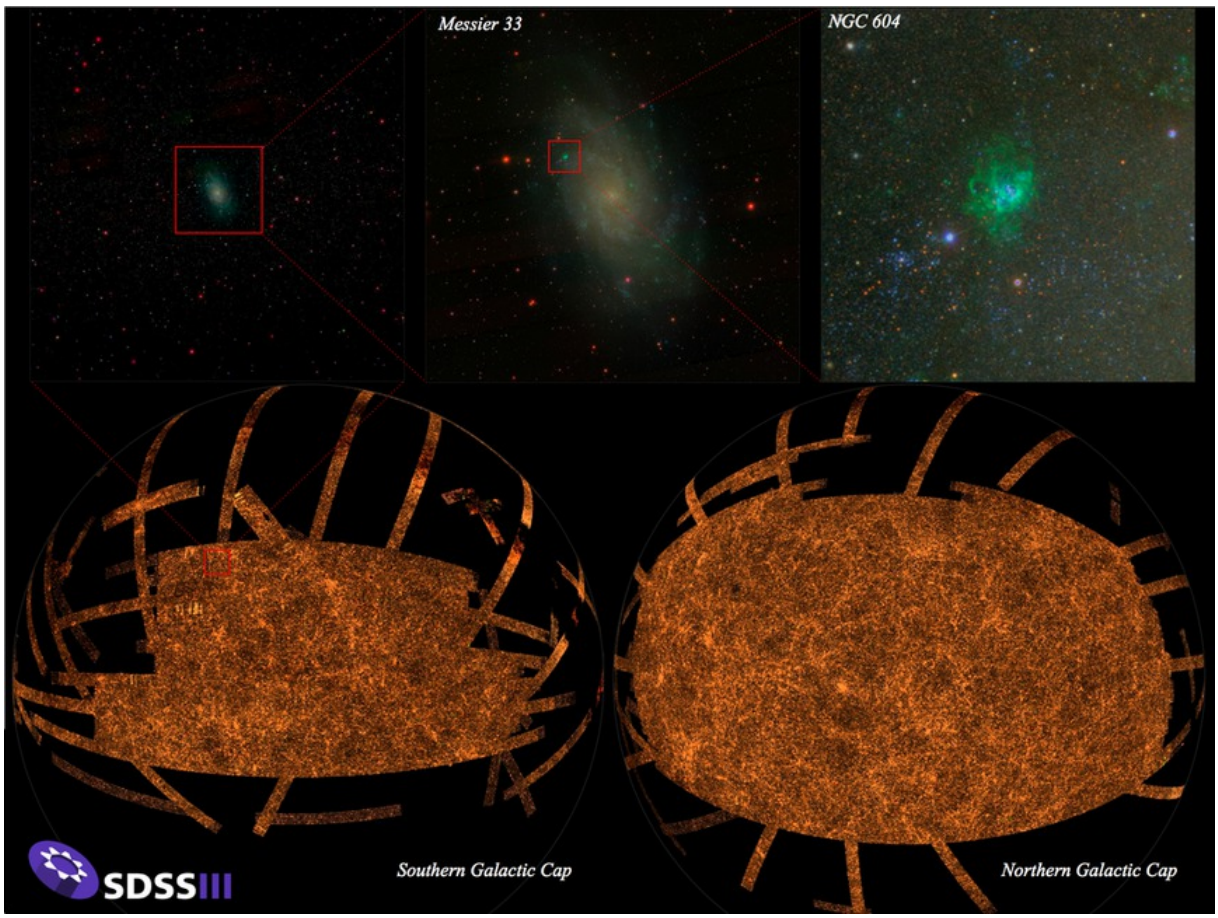


Figure 3.2 – The top left picture shows the SDSS-III view of a small part of the sky, centered on the galaxy Messier 33 (M33), at different zoom levels. The figure at the bottom is a map of the whole sky derived from the SDSS-III image. Both the northern and southern galactic caps are shown. The bands coming from the observation strategy are partially visible. Credits: M. Blanton and the SDSS-III collaboration

the limitation from cosmic variance, shot noise is by far the current limitation of the measurement of the BAO feature in quasar spectra. Thus one wants to maximize the density of observed quasars. Several methods have been developed to select quasars from photometric surveys, most of them using the positions of objects in different colour-colour planes. A detailed study of the quasar selection in BOSS can be found in Ross et al. (2012). Moreover a selection method using quasar variability is presented in Palanque-Delabrouille et al. (2011). Using these complementary multiple techniques, BOSS observed between $15\,\mathrm{deg}^{-2}$ and $18\,\mathrm{deg}^{-2}$ high-redshift quasars ($z > 2.15$), depending on the star density, and the availibility of additional data.

The tiling of the sky is done with 1000 optical fibres inserted in an aluminium plate that eventually lies at the telescope focal plane (see section 3.3), covering $7\,\mathrm{deg}^2$ (3° diameter). The process of tiling of the $10\,000\,\mathrm{deg}^2$ of the survey is described in Blanton et al. (2003). The goal is to minimize the number of required plates while maximizing the number of objects that will be assigned to a fibre, taking into account that there is a required minimum separation of $3''$ between two fibres thus requiring two overlapping plates to observe close objects. However, to allow for some flexibility, like further optimization at later times in the selection algorithm, the tiling is not done on the all $10\,000\,\mathrm{deg}^2$ at once but on 31 chunks called BOSS1 to BOSS31.

While the projected surface of a plate on the sky is about $7\,\mathrm{deg}^2$, the plate overlap reduces the average surface per plate in BOSS to $5{-}5.5\,\mathrm{deg}^2$.

Out of the 1000 fibres of a plate, 80 are positioned where there is no detected object, taking spectra of the "black" sky that will be used to subtract sky background, like light pollution from human activity or atmosphere emission, from science fibres. Another 20 fibres are used for calibration purpose, pointing at F-type stars. On the remaining 900 fibres, 5 are assigned to objects observed with at least another plate, to control the reproducibility of the observations. Therefore 895 fibres are available, upon which 160 to 200 are for quasars targets, 560 to 630 for galaxies targets and 20 to 90 for ancillary projects.

The plates preparation is done several months in advance at the University of Washington. To determine the position of each hole, the best sideral time for the observation is estimated. The exact position of each fibre is then determined taking into account the atmospheric refraction. Because this refraction depends on the wavelength, the position of the fibre will not be the same if one wants to optimize the blue or red part of the spectrum. For quasars, the holes are centered on the light at 4000 Å to maximize the signal to noise ratio in the Lyman-α forest, whereas they are centered on 5400 Å for the galaxies. In the focal plane, this represents $0{-}300\,\mu\mathrm{m}$ depending on the distance to the optical axis (at the centre of the plate). In addition to the 1000 holes for the targets, 16 holes are drilled for pointing and monitoring the observation. Holes are also drilled at the emplacement of bright stars to avoid parasitic reflexions on the plate.

Once the plates are drilled, they are sent to the observatory where they are prepared by the on-site staff before each observation night. For this, plates that will be observed during the night are placed into a cartridge during the day. The thousand fibres are plugged onto it by hand, one at a time. Plugging a cartridge takes around forty-five minutes for two persons. Up to nine cartridges can thus be done per day, which also corresponds to the maximum number of cartridges that are needed in a given night since each plate is observed for typically one hour and astronomical nights are at most nine hours long.

Observations start when the sun is 12° under the horizon. The first cartridge is then placed at the focal plane of the telescope. The observation starts with measurements of arc lamps, to calibrate the spectra, then go on with fifteen minutes of data taking. After fifteen minutes, a simple software calibrates and extracts the one dimensionnal spectrum. This simplified extraction aims at estimating the signal to noise ratio of the spectra. This is done because a cartridge is said observed when every galaxy spectrum has a squared signal to noise ratio $(S/N)^2 \geq 20$ for the red camera and $(S/N)^2 \geq 10$ for the blue camera. This criterion is used because $(S/N)^2$ depends linearly on the exposure time. If the criterion is not met, the plate is reobserved again for fifteen minutes. Once a plate is observed, the next cartridge is loaded. At the end of the night, if the last plate is not finished, it is left on the telescope and will be reobserved at the beginning of the next time. The criterion only deals with galaxy spectra because the gain on BAO measurement coming from the improvement of signal to noise in quasar spectra is said to be less than the gain from a larger survey surface (McDonald and Eisenstein, 2007; Font-Ribera et al., 2012). However, this assertion is challenged for next surveys like eBOSS and DESI.

3.3 The Instruments

As described in Gunn et al. (2006), the realization of a multi-band photometric survey, covering a large fraction of the observable sky with an associated ambitious spectroscopic survey requires several technical innovations:

— a telescope with a wide field of view, a very low distortion focal plane, a very precise pointing and the capacity to switch between photometry and spectroscopy easily. The main characteristics of the telescope are described in section 3.3.1.
— A camera able to cover the field of view of the telescope and allowing simultaneous observations in different photometric bands. It also needs a good precision for astrometric calibration. The SDSS camera is described in section 3.3.2.
— Spectrographs able to take hundreds of spectra simultaneously on a wide wavelength range. The BOSS spectrographs are presented in section 3.3.3.
— An acquisition system allowing both data storage and real time control quality.
— A fast data pipeline, to efficiently calibrate the data and identify the objects both in photometry and spectroscopy. The data reduction pipeline is detailed in section 3.4.

3.3.1 The Telescope

To fulfil the goals of SDSS, the telescope must have a wide field of view, a focal plane with very low distortion, a very precise pointing and the ability to switch between photometric and spectroscopic observations. With all these requirements in mind, a 2.5 m telescope with a 3° field of view has been designed and then built at the Apache Point Observatory (APO). This observatory is located in New-Mexico at 2800 metres above sea level. This section presents the main characteristics of the telescope that can be seen in figure 3.3b. A detailed presentation of the optics and mechanics can be found in Gunn et al. (2006).

The SDSS telescope is a Ritchey-Chrétien: its two mirrors (primary and secondary) are hyperbolic. In addition to the classical Ritchey-Chrétien design, it has two optical correctors. The first one is a Gascoigne corrector which aims at reducing the optical system astigmatism and the second one is a pair that are switched depending on the observation mode (photometric or spectroscopic). Figure 3.3a shows a schema of the telescope.

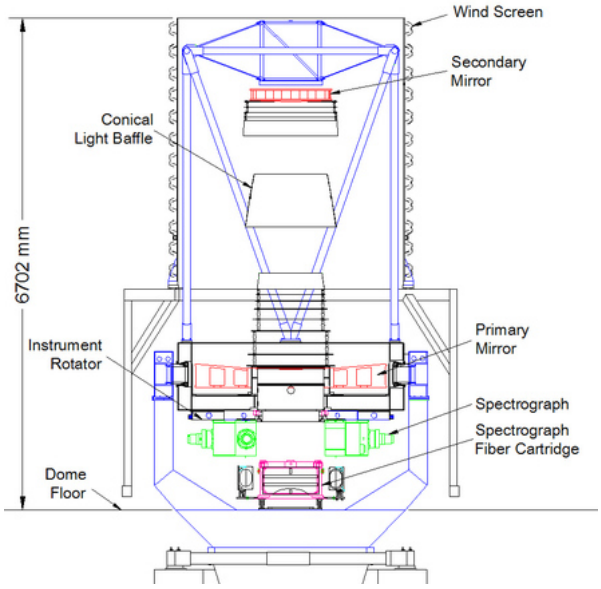
The primary mirror has a diameter of 2.5 m and a focal to diameter ratio of $f/2.25$. It has a hole of 1.17 m in its center. There is a conical light baffle above it to prevent parasitic light to reach the focal plane.

The secondary mirror has a diameter of 1.08 m. It also has a light baffle, that in addtion to the secondary miror, blocks 27% of the light reaching the telescope.

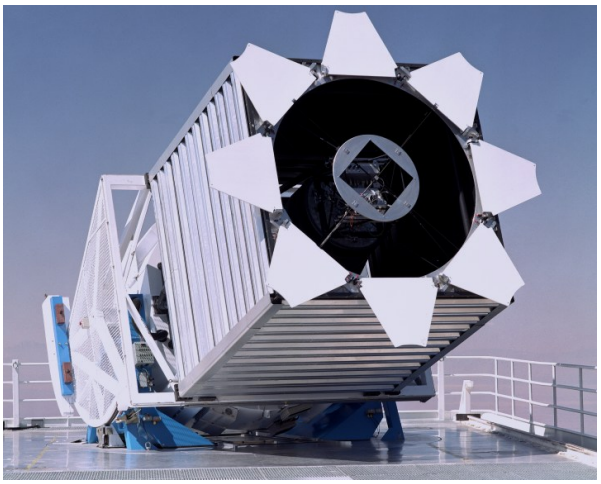
The two mirrors form an optical system with a focal to diameter ratio of $f/5.0$ and a 3° field of view. The two mirrors are separated by only 3.6 m, so that the focal is located 0.76 m behind the primary mirror, allowing an easy access to the instruments.

The first corrector is of Gascoigne type and it is the last optical piece common to the photometric and spectrometric configuration. That is why it is sometimes called the “common” corrector. Its goal is to drastically reduce the astigmatism which is one of the major drawbacks of Ritchey-Chrétien telescopes.

The second corrector or “final” corrector is not the same depending on the configuration of the telescope. These configurations are optimized for each case and are briefly presented in section 3.3.2 for the photometric corrector and in section 3.3.3 for the BOSS spectroscopic corrector.



(a) SDSS telescope schema: the mirrors are in red, the spectrographs in green and a cartridge in pink. Credits: Smee et al. (2012)



(b) A photograph of the 2.5 m SDSS telescope.

Figure 3.3 – The SDSS telescope

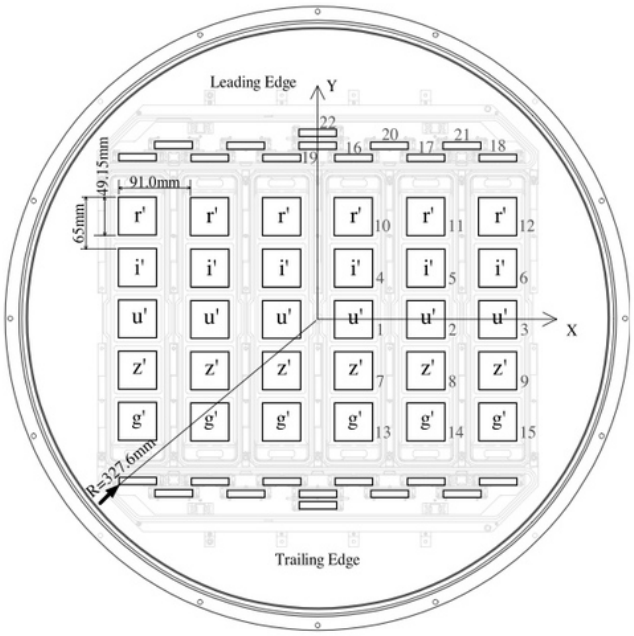
3.3.2 The Camera

The camera used for the BOSS photometric phase was the same as for SDSS and SDSS-II because it was already optimized for the operating mode and the 3° field of view of the telescope. A full study of the camera can be found in Gunn et al. (1998).

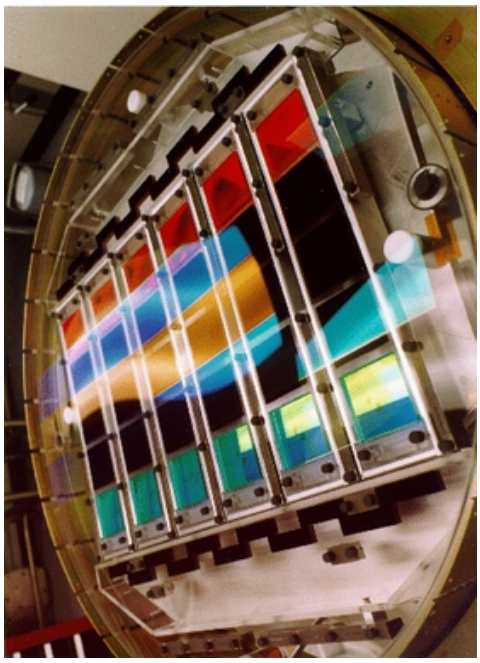
The first piece of the camera is an optical corrector that aims at correcting the distortion at the focal plane induced by the telescope optical system. In addition to being an optical corrector this lens is also a structural piece of the camera on which the detectors are mounted. It plays a very important role in the preservation of the mechanical properties of the camera thus having a big impact on the image quality and the astrometry.

The next piece of the camera is a matrix of 6 columns of 5 CCDs each, one for each photometric band (u', g', r', i' and z'). The size of one CCD is 2048 × 2048 pixels. The remaining space on the focal plane is used for 24 smaller CCDs (2048 × 400 pixels), 22 being used for the astrometry and 2 for the focus. The repartition of these smaller CCDs around the bigger ones is visible on figure 3.4. The drift scanning is performed from top to bottom on schema 3.4a, therefore a star entering the focal plane successively encounters an astrometric CCD, the photometric CCDs in the order r', i', u', z', g' and eventually a second astrometry dedicated CCD.

All the CCDs dedicated to the photometry are identical but they have filters to select the wavelength band of observation. The CCDs can observe from the UV atmospheric cut off around 3000 Å to the limit of silicon based detectors close to 11 000 Å. The wavelength range and quantum efficiency (defined as the ratio between the numbers of collected electrons to the number of incident photons) of each filter is shown in figure 3.5. The very low efficiency of the u' band is the reason why it is located at the centre of the focal plane. The magnitude detection limits are defined for a signal to noise ratio of 5. They are approximately $u'_{\text{limit}} = 22.1\,\text{mag}$, $g'_{\text{limit}} = 23.2\,\text{mag}$, $r'_{\text{limit}} = 23.1\,\text{mag}$, $i'_{\text{limit}} = 22.5\,\text{mag}$ and $z'_{\text{limit}} = 20.8\,\text{mag}$.



(a) The focal plane organization for the camera used for SDSS, SDSS-II and BOSS. The matrix of photometric CCDs is at the centre, whereas the astrometric and focus CCDs are at the top and bottom. The drift scanning direction is from top (leading edge) to bottom (trailing edge). Credits: Gunn et al. (1998)



(b) A photograph of the SDSS camera.

Figure 3.4 – The SDSS camera

3.3.3 The Spectrographs

BOSS has two identical spectrographs that have inherited some of the technologies developed for the first two generations of SDSS. The main one is that they are fed by optical fibres plugged at the focal plane. However, because of BOSS requirements several upgrades were done, like an increase in the number of fibres or the optical efficiency. This section summarizes the most important characterictics of the spectrographs for which a full study has been conducted in Smee et al. (2012).

3.3.3.1 Cartridges

To get the spectra, an aluminium plate is put at the focal plane of the telescope. This plate is 3.2 mm thick, has a diameter of 81.3 cm and weighs 4.3 kg. The telescope transforms the position of an object on a sky in declination and right ascension coordinates with respect to the optical axis in a position on the focal plane in cartesian coordinates. Therefore, the plate is drilled, in advance, where optical rays from an object are going to converge. Optical fibres are then plugged in these holes to redirect the light to the spectrograph grism, which is a combination of a prism and a grating. The plates and the optical fibres are held by an aluminium structure, the whole forming a “cartridge” that can be seen in figure 3.6. A cartridge can be easily mounted on the telescope by one person and several cartridges (9 are available) can be plugged in advance, making this system easy to operate and flexible.

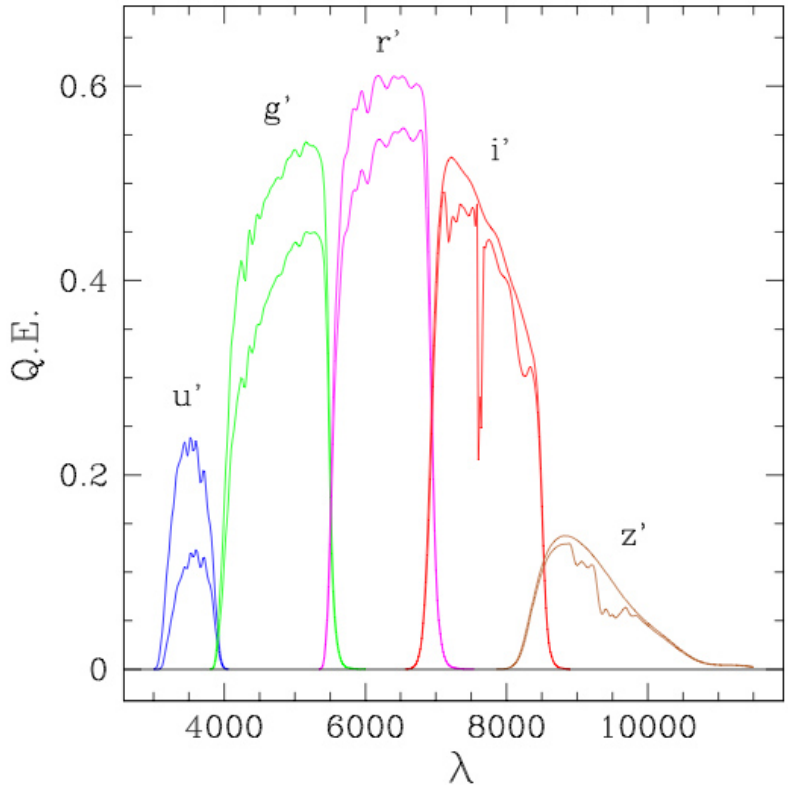
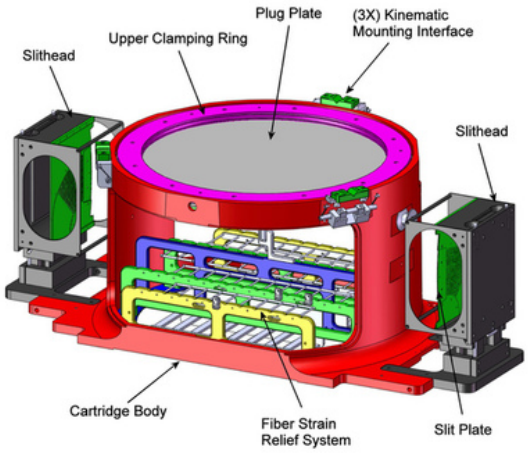
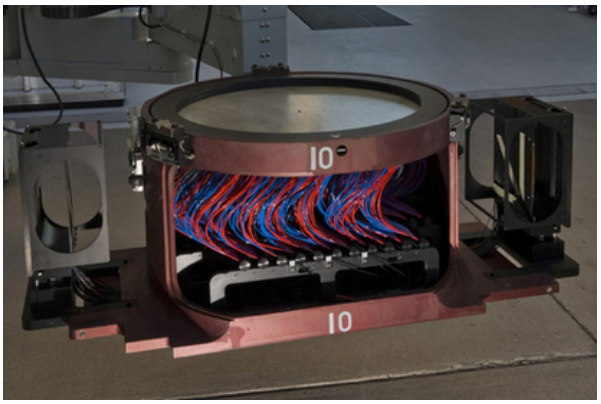


Figure 3.5 – Quantum efficiency of the optical system in each band. The upper curves take into account the transmission of the filters, the quantum efficiency of the CCDs and the losses in the optical system. The lower curves add the atmospheric extinction. Credits: Gunn et al. (1998)



(a) BOSS cartridge schema



(b) A BOSS cartridge photograph, ready to be mounted on the telescope. The optical fibres are visible in red and blue. The two slits on each side are designed to go inside the spectrographs.

Figure 3.6 – A BOSS cartridge. Credits: Smee et al. (2012)

3.3.3.2 Optical Fibres

In order to increase the capacity of the survey, the number of fibres per plate went from 640 to 1000 for SDSS-III/BOSS. To do so without changing most of the telescope optics, the size of the fibres were decreased from 180 µm (3″ on the sky) to 120 µm (2″ on the sky). It also contributed to a better signal to noise ratio for high redshift objects by decreasing the contribution of the sky background. Each fibre collects the light at the focal plane in a cone of 0.1 numerical aperture [5] and gives back the light in a slightly wider cone of 0.125 numerical aperture. This deterioration is due to the propagation of light inside the optical fibre. At the end of the fibre, the collimator is built to collect light from a cone of 0.125 numerical aperture, any ray outside this cone is lost. Thus there is a requirement to control the beam width at the output of the fibre to maximize the yield of the system. To do so, the spectrographs are mounted directly on the telescope (as seen on figure 3.3a), limiting both the displacement between the two ends of the fibres and the mechanical constraints. It also allows a minimization of the fibre length, each of them measuring 1830 ± 25 mm.

The 1000 fibres are divided in two sets of 500, grouped in bundles that are directed to two slits, one at each end of the cartridge. Each slit goes into a spectrograph when the cartridge is mounted on the telescope.

3.3.3.3 Optical System and CCDs

The optical system of the BOSS spectrographs is almost identical to the original system of SDSS. However, the wavelength range has been increased from $3900\,\text{Å} \lesssim \lambda \lesssim 9000\,\text{Å}$ to $3600\,\text{Å} \lesssim \lambda \lesssim 10\,000\,\text{Å}$. The extension on the blue side allows to increase the detection of the Lyman-α forest at low redshift: considering one can use the forest up to 10 Å of the Lyman-α emission line, i.e. 1206 Å, the minimum redshift is thus $z_{\min} \approx 3600/1206 - 1 \sim 2.0$. The increase on the red side was motivated by the study of galaxy spectra discontinuity. Both ameliorations were made possible thanks to technological progress on CCDs, grisms and optical transmission of optical systems in general.

A schema of the optical system of a spectrograph is presented in figure 3.7. Light arrives in the spectrograph through a slit located at the end of a cartridge. This slit holds 500 fibres. First, the light goes on a collimator which reflects it to a beamsplitter in the form of a parallel beam of 160 mm in diameter. At the beamsplitter, wavelengths shorter than 6400 Å are reflected in direction of the blue channel whereas longer wavelengths are transmitted to the red channel. In each channel, the beam is diffracted by a grism whose density is 400 lines per millimetre for the red channel and 520 lines per millimetre for the blue channel. Cameras are located after the grisms and are made of a succession of lenses : a singlet, a triplet, a doublet and eventually two field flatteners just in front the CCD detectors. These CCD detectors are made of 4096×4096 pixels whose sides are 15 µm long. The path of light rays inside the cameras is shown in figure 3.8.

3.3.3.4 Optical Performances

The efficiency of the optical system has been predicted (Smee et al., 2012) as a function of wavelength, taking into account every component from the atmospheric absorption to the

5. The numerical aperture NA of a system is defined as $\text{NA} = n.\sin\theta$ where n is the index of refraction of the medium in which the system is working, and θ the maximum possible angle (with respect to the optical axis) of a light ray entering the system.

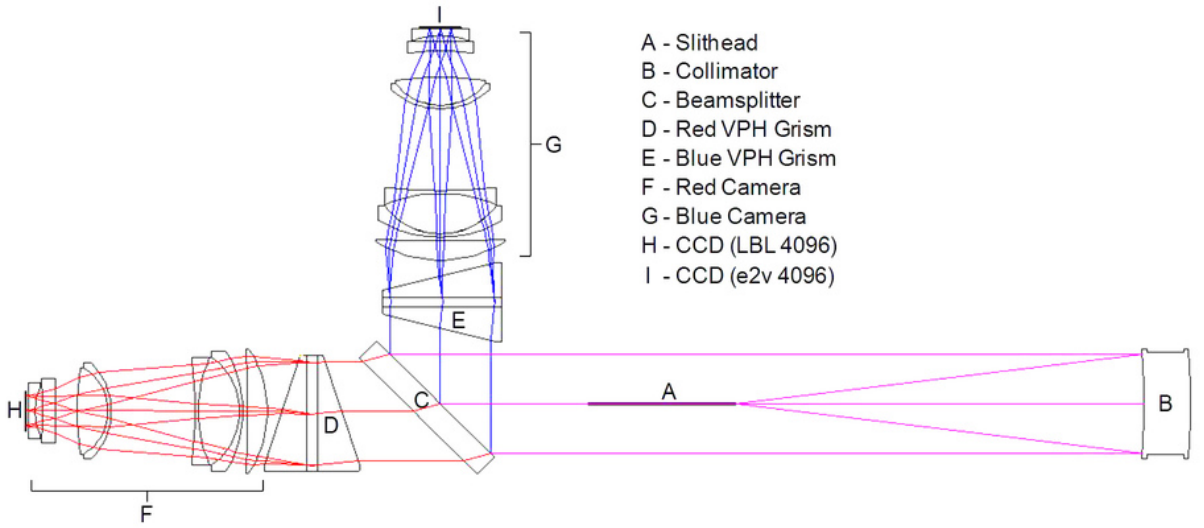


Figure 3.7 – Optical schema of a BOSS spectrograph. Light comes from the slithead (A) which holds the 500 fibres bundle. It is directed to the collimator (B) which returns it in the opposite direction as a parallel beam. The beamsplitter (C) splits the beam into a red and a blue components. Each component is diffracted by a grism (D and E) and then focused by the camera (F and G) on the CCDs (H and I). Credits: Smee et al. (2012)

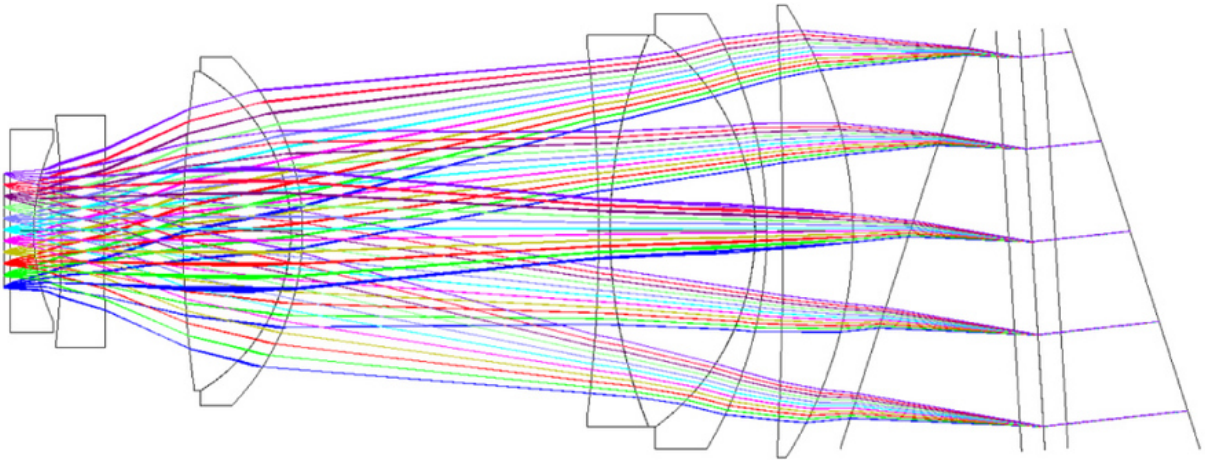


Figure 3.8 – Path of the light rays diffracted by the grism (on the right) and focused on the CCD (on the left) by the optical system of the camera (in the middle). Credits: Smee et al. (2012)

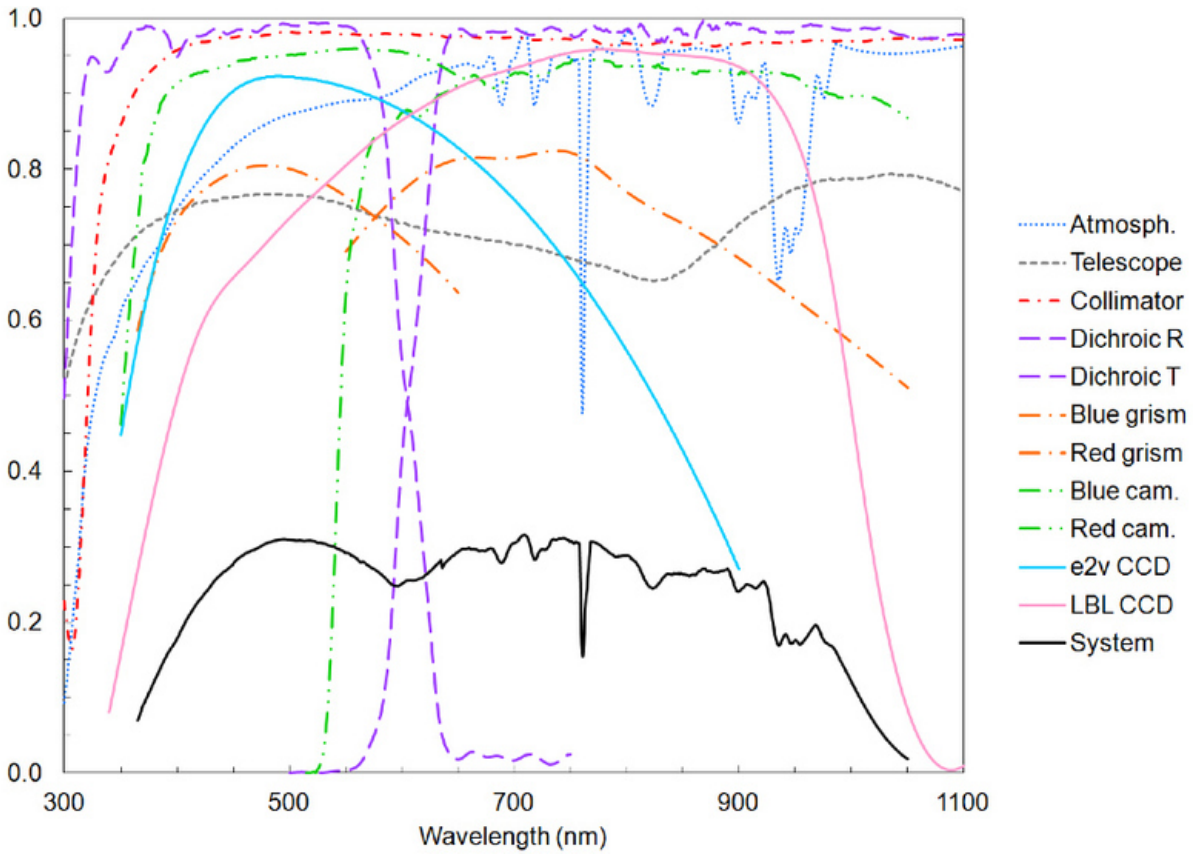
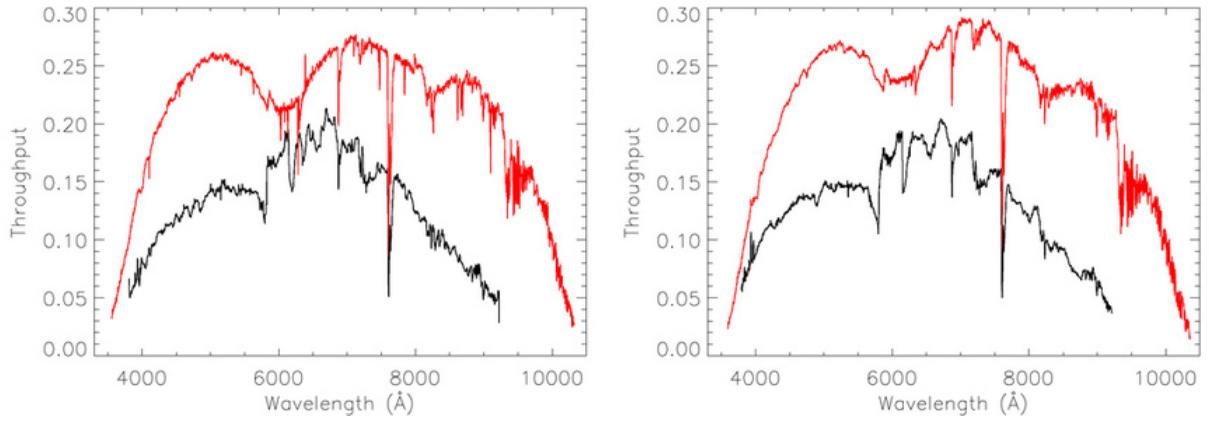


Figure 3.9 – The predicted optical efficiency of each BOSS component and the total efficiency of the whole system (black curve). Credits: Smee et al. (2012)

quantum efficiency of the CCDs. The results are presented in figure 3.9. The maximum efficiency is expected to be around 30% in the red channel as well as in the blue one. This prediction is to be compared with the efficiency measured by observing 84 standard stars with the BOSS spectrographs as shown in figure 3.10. For each star, the efficiency is defined as the ratio between the measured flux and the incoming flux outside the atmosphere. The average yield of each spectrograph is then computed. The two spectrographs have very similar efficiency. However, the measured efficiency is slightly lower than the predicted one, with 26% in the blue channel and 28% in the red one. Nevertheless, they represent a great increase compared to the SDSS spectrographs whose efficiency is also shown in figure 3.10.

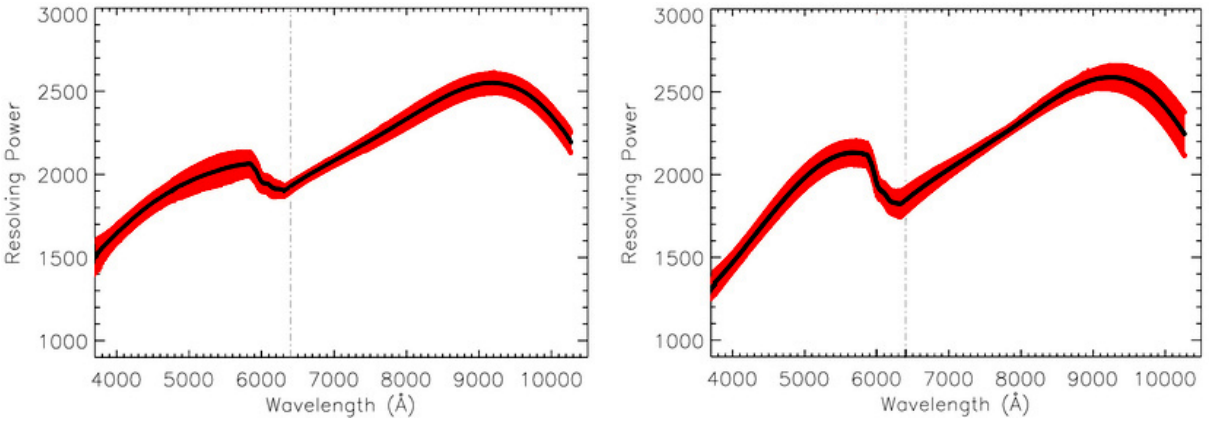
The spectral resolution is measured before each plate observation, using arc lamps dedicated to the calibration. First, the arc lamps are observed by the spectrographs. Then, each emission line in the arc lamp spectra are modelled by a Gaussian of width σ_i. The computed values of σ_i are fitted by a fourth order polynomial to get σ_λ, the width of a line as a function of its wavelength on the full wavelength range of the spectrographs. The resolving power is then defined as $R(\lambda) = \frac{\lambda}{2.35\sigma_\lambda}$ where $2.35\sigma_\lambda$ is the full width at half-maximum of a line at wavelength λ. The resolving power averaged on a hundred plates is presented on figure 3.11 for the two spectrographs. The step around 6200 Å is due to the change of CCDs (blue channel to red channel).

The resolving power of the spectrographs does not only depend on the wavelength but also on the position in the focal plane of the CCD, i.e. on the fibre number. This dependence is shown in figure 3.12. For all CCDs, the optical quality is lower on the sides where the resolving



(a) Spectrograph 1 (b) Spectrograph 2

Figure 3.10 – Measured optical efficiency for each BOSS spectrograph (red), compared to the efficiency of the original SDSS spectrographs (black). credits: Smee et al. (2012)



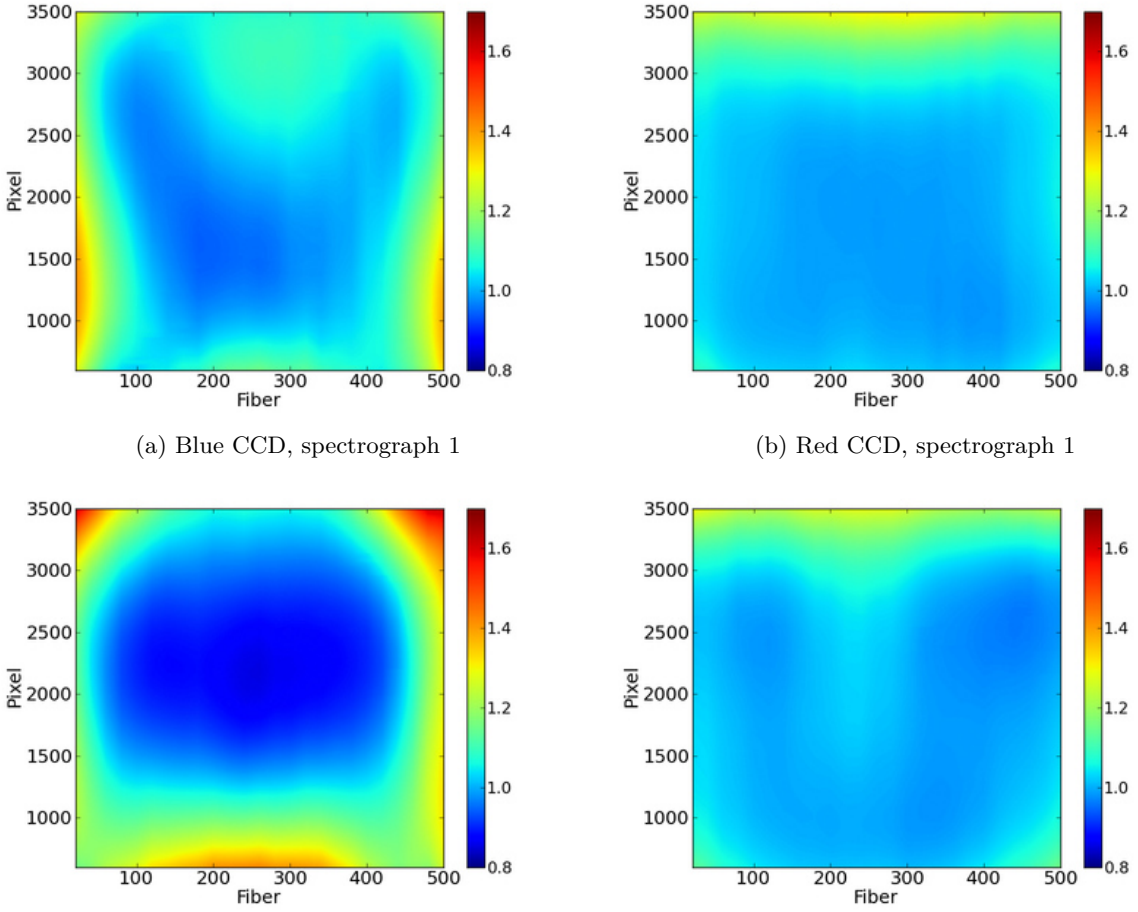
(a) Spectrograph 1 (b) Spectrograph 2

Figure 3.11 – The resolving power for the BOSS spectrographs. The red regions correspond to the regions that contain 68% of the plates in the measurement. The dashed line corresponds to the transition between the blue CCD and the red CCD. Adapted from Smee et al. (2012)

power can be up to two times less than the resolving power at the centre. However, the red CCDs seem much more homogeneous than the blue ones with a constant resolution up to the pixel ~ 3000 which corresponds to $\lambda \approx 9500\,\text{Å}$. The determination of the resolution is here an important matter as it has a strong influence on the measurement of the one dimensional power spectrum (see 4).

3.4 Data Reduction

The data reduction software works plate by plate. It has been designed to extract, calibrate, coadd, classify and estimate the redshifts of the thousand objects of a plate using all the available expositions. This software (Bolton et al., 2012), which is briefly described in 3.4.1, is able to distinguish with a good efficiency quasars, galaxies and stars. However, it was complemented by a visual inspection of all the quasar spectra as described in section 3.4.2. The software suffers from some caveats, e.g. in the noise and the resolution estimation, that will be discussed in section 4.2.



(a) Blue CCD, spectrograph 1

(b) Red CCD, spectrograph 1

(c) Blue CCD, spectrograph 2

(d) Red CCD, spectrograph 2

Figure 3.12 – Width of an emission line in pixels as a function of its position in the focal plane, for each of the four CCDs of the spectrographs. The blue CCDs are on the left, the red ones on the right. Credits: Smee et al. (2012)

3.4.1 Pipeline Software

The first step deals with the individual exposure and transforms the two dimensional images of the CCDs in one dimensional spectra. Those spectra are then calibrated using the arc lamps data of the plate. The per pixel variance is estimated from the readout noise and the number of photons recorded in each pixel. The inverse variance is then multiplied by a factor that takes the different known flaws of the CCDs into account. The pixels that were hit by a cosmic ray are identified and masked. Finally, the flux is calibrated using the spectra of standard stars that were observed on the plate for this purpose.

The second step consists in adding the different individual exposures to obtain a coadded spectrum for each fibre. Each spectrum has the data of both the blue and red CCDs, covering the full range of wavelength of the instruments 3650−10 400 Å. The spectra are then rescaled to be linear in $\log(\lambda)$. The variance is estimated using the variance of each exposure, the covariance being neglected. For objects that have been observed multiple times, the different spectra are compared and the spectrum with the best signal to noise ratio is kept as the “primary” spectrum.

The last step is to fit the spectra with models of star, galaxy and quasar spectra. This procedure is described in Aihara et al. (2011) and aims at determining the type of the object and its redshift.

For the tenth data release of SDSS (DR10), the BOSS collaboration has made public 1 880 584 galaxy spectra, 312 309 quasar spectra and 269 615 star spectra along with all the other data like individual exposure, pixel variance, calibration vector, ... Data are available on the corresponding SDSS website [6].

3.4.2 Visual Inspection

As described in section 2.2, quasars have a broad variety of spectra characteristics like DLA and BAL, that make their automatic classification difficult. To overcome this problem, members from the French Participation Group (FPG) of SDSS have visually inspected every quasar spectrum observed by BOSS. This visual inspection allows a reliable identification as well as a precise determination of the redshift. A full description of this inspection is given in Pâris et al. (2012). For the tenth data release, 321 579 spectra have been inspected leading to a quasar catalogue of 166 583 spectra (Pâris et al., 2014). The inspected spectra are all the quasar targets, from the main survey and the ancillary programmes, along with some galaxy spectra that were classified as quasars by the pipeline software.

The visual inspection allows an estimation of the efficiency of the automatic classification of the spectra. On the one hand, it shows that the star sample is very pure with less than 0.1% of the objects classified as stars being quasars in reality. On the other hand, the high redshift ($z > 2.15$) securely-identified quasar sample has more contamination with 0.5% of the objects classified as quasars being either stars or lower-redshift quasars. This contamination is much higher for the low-redshift ($z < 2.15$) quasar sample where it reaches 5%. In addition some of the quasars identified by the pipeline are flagged as unsecure detections that cannot be used later on in the analyses wihout a visual inspection stage to confirm or correct these identifications. This is the case for 8% of the $z > 2.2$ quasars and 25% of the $z < 2.2$ quasars. Almost half of the latter are indeed visually identified as stars and not low redshift quasars. While this is no concern for BOSS since only high redshift quasars are used in the analyses, this high contamination requires a revision of the pipeline for application to eBOSS where both low and high redshift quasars will be targeted.

Thanks to the visual inspection, it is also possible to check the efficiency of the automatic redshift determination of quasars. It shows that only 0.3% of the quasars have an error greater than 0.1 ($\Delta z > 0.1$). Such errors mainly occur for quasars with a redshift $z < 2.0$ with a non visible Lyman-α emission line.

Eventually, the visual inspection also allows the detection of particular spectra features likes DLA and BAL. However, it should be noted that an automatic procedure has now been developed to determine the presence of a DLA (by measuring the column density of the associated system) and to give a "balnicity index" to spectra with a BAL feature (Noterdaeme et al., 2009). The presence of DLAs or of a BAL in quasar spectra are characteristics that are used in the cosmological analyses based on BOSS quasar data.

6. `http://www.sdss3.org/dr10`

Chapter 4

The One-Dimensional Lyman-α Forest Power Spectrum from BOSS

"Failure is not an option"
—Apollo 13, 1995

RECOVERING the one-dimensional Lyman-α forest power spectrum using the quasar spectra of BOSS (data release 9), is the first part of the work presented in this thesis. Due to the unprecedented statistical power of the data, a detailed study of all the uncertainties associated with the reduction pipeline and the reconstruction methods had to be conducted. This study led to the publication of a paper, Palanque-Delabrouille et al. (2013), on which this chapter is based and the results were presented in Borde et al. (2013), as well as in several status report presentations during BOSS and SDSS collaboration meetings.

This work uses the background presented in the previous chapter and is outlined as follows: first I present the motivations for this work and the previous studies in section 4.1. Then, in sections 4.2 and 4.3, I present the BOSS data selection, preparation and calibration with an emphasis on the determination of the level of noise in the spectra and the spectrograph resolution. In section 4.4, I describe the two complementary methods we have developed to analyse the data. In section 4.5 our estimates of the systematic uncertainties associated with each method or due to our imperfect knowledge of the instrument performances are detailed. The final results are given in section 4.6. Conclusions and perspectives are presented in section 4.7 and will lead to the next part of my work: simulating universes.

4.1 Motivations

As we have seen in chapter 2 and section 2.3 , the information embedded in the Lyman-α forest can be used to probe the amplitude and shape of the power spectrum of mass fluctuations (Croft et al., 1998; Gnedin, 1998; Hui et al., 1999; Gaztanaga and Croft, 1999; Nusser and Haehnelt, 1999; Feng and Fang, 2000; McDonald et al., 2000; Hui et al., 2001), and to constrain cosmology through the study of redshift-space distortions and the Alcock-Paczynski test (Alcock

and Paczyński, 1979; Hui et al., 1999; McDonald and Miralda-Escude, 1999; Croft et al., 2002), the mass of neutrinos (Seljak et al., 2005; Viel et al., 2010), or the BAO peak position (McDonald and Eisenstein, 2007). Initially, the Lyman-α forest power spectrum was studied exclusively along the line of sight, by measuring the correlation separately in each quasar spectrum, starting with the use of small numbers of high-resolution spectra: 1 Keck HIRES spectrum (Croft et al., 1998), 19 spectra from the Hershel telescope on La Palma or the AAT (Croft et al., 1999), 8 Keck HIRES spectra (McDonald et al., 2000), a set of 30 Keck HIRES and 23 Keck LRIS spectra (Croft et al., 2002), or a set of 27 high resolution UVES/VLT QSO spectra at redshifts ~ 2 to 3 (Kim et al., 2004b,a; Viel et al., 2004).

A substantial breakthrough was achieved with the measurement of the Lyman-α forest power spectrum based on the much larger sample of 3035 medium-resolution ($R = \Delta\lambda/\lambda \approx 2000$) quasar spectra from SDSS (York et al., 2000) by McDonald et al. (2006). The large number of observed quasars allowed detailed measurements with well characterized errors of the power spectrum up to larger scales, probing the linear regime and providing cosmological constraints (McDonald et al., 2005; Seljak et al., 2005).

We have seen in chapter 3 that SDSS III (Eisenstein et al., 2011) and more precisely BOSS (Dawson et al., 2013) was especially designed to target quasars at redshift $z > 2$, which are useful for the Lyman-α forest analysis, and to obtain spectra of a much larger number of them than in the previous phases of SDSS. This large number of quasars allowed for a detailed measurement of the Lyman-α power spectrum in three-dimensional redshift space (as a function of the transverse and parallel directions) in Slosar et al. (2011), using the first 14000 quasars of BOSS. For the first time, the redshift distortions predicted in linear theory of large-scale structure by gravitational evolution (Kaiser, 1987) were detected in the Lyman-α forest. With the quasars in the Data Release 9 (Ahn et al., 2012), containing more than 60000 quasars with observed Lyman-α forest absorption (Pâris et al., 2012; Lee et al., 2013), the measurement of the redshift space power spectrum has been extended up to the scales of the baryon acoustic oscillations, yielding the highest redshift measurement of the acoustic peak position and providing new constraints on the history of the expansion of the universe (Busca et al., 2013; Slosar et al., 2013; Kirkby et al., 2013).

The measurement of the three-dimensional power spectrum uses only information from the flux correlation of pixel pairs in different quasar spectra that are relatively close in the sky. However, the correlation of pixel pairs on the same quasar spectrum provides complementary, useful information on the Lyman-α correlation along the line of sight, which is also important to constrain the physical parameters of the Lyman-α forest. The one-dimensional power spectrum P_{1D} and the three-dimensional power spectrum P_{3D} are related through equation 2.18:

$$P_{1D}(k_\parallel) = \frac{1}{2\pi} \int_{k_\parallel}^{\infty} k P_{3D}(k) \mathrm{d}k \tag{4.1}$$

If all the relevant scales could be treated in the limit of linear theory, the three-dimensional power spectrum should be simply related to the mass power spectrum according to $P_{3D}(k_\parallel, \mathbf{k}_\perp) = b_\delta^2 P(k)(1 + \beta k_\parallel^2/k^2)^2$, where $k = |\mathbf{k}| = \sqrt{k_\parallel^2 + |\mathbf{k}_\perp|^2}$, and b_δ and β are the density bias and redshift distortion parameters of the Lyman-α forest (McDonald, 2003; Slosar et al., 2011). However, linear theory is valid only on large scales, and even though the linear expression is valid for P_{3D} when k is small, the one-dimensional P_{1D} is affected by the non-linearities of small scales even for very small values of $k_\parallel$. The theoretical interpretation of measurements of P_{1D}

is therefore always dependent on the non-linear physics of the intergalactic medium on small scales.

In this chapter, I present our measurement of the 1D transmission power spectrum of the Lyman-α forest from a sample of 13 821 quasar spectra, which are selected as the highest quality spectra among the set of 61 931 quasars at $z > 2.15$ from the DR9 quasar catalogue of Pâris et al. (2012).

Historically, two approaches have been used to measure the 1D power spectrum of the fluctuations in the transmitted flux fraction F. The first is done directly in Fourier space, computing the Fourier transform of $\delta = F/\langle F\rangle - 1$ for each quasar spectrum and obtaining the power spectrum from these Fourier modes (Croft et al., 1998, 2002; Viel et al., 2004). The second approach uses a likelihood method to compute in real space the covariance matrix of δ (or line of sight correlation function) as a function of the pixel pair separation in the spectra (McDonald et al., 2006). The 1D power spectrum is the Fourier transform of the Lyman-α correlation function obtained in this way. The two methods have their own advantages and drawbacks in terms of, for example, robustness, processing speed, accounting of instrumental effects, precision, ... To benefit from their complementarity, we have developed two independent analysis pipelines based on either technique. The implementation and development of the likelihood approach was my main contribution to this work.

4.2 Data Calibration

4.2.1 BOSS Reduction Pipeline Limits

At low redshift ($z < 2.5$), the 1D power spectrum has a significant contribution from photon noise, so it is quite sensitive to the precision with which the noise level in the data is known. The spectrograph wavelength resolution is also a major issue on small scales (i.e. large k-modes) where it abruptly reduces the power spectrum, by a factor of ~ 2 at $k = 0.01\,(\mathrm{km/s})^{-1}$ and by a factor of 5−10 at $k = 0.02\,(\mathrm{km/s})^{-1}$. The accuracy with which noise and spectrograph resolution are determined in the automated pipeline is insufficient for the purpose of this analysis. We have therefore developed techniques, described in the following sections, to derive corrections and achieve the required accuracy. These refinements were not necessary for the measurement of the large-scale 3D Lyman-α correlation function (Busca et al., 2013; Slosar et al., 2013) since the BAO feature occurs at much larger scales than the size of the resolution element, and the noise in the data only affects the amplitude of the power spectrum and not the correlation function where the BAO peak is seen. Instead, we here aim at measuring the absolute level of the power spectrum, which is directly affected by the level of noise, down to scales of the order of a few Mpc i.e., of a few pixels, where an accurate knowledge of the spectrograph resolution is crucial.

4.2.2 Calibration of Pixel Noise

The noise provided by the SDSS-III pipeline is known to suffer from systematic underestimates (McDonald et al., 2006; Desjacques et al., 2007). To investigate the extent of this issue, we examine the pixel variance in spectral regions that are intrinsically smooth and flat. We use two 50 Å regions of quasar spectra (hereafter "side-bands"), redwards of the Lyman-α peak: $1330\,\text{Å} \leq \lambda_{\mathrm{RF}} \leq 1380\,\text{Å}$ and $1450\,\text{Å} \leq \lambda_{\mathrm{RF}} \leq 1500\,\text{Å}$. These bands are not affected by the Lyman-α absorption and have a quasar unabsorbed emission flux that is relatively wavelength

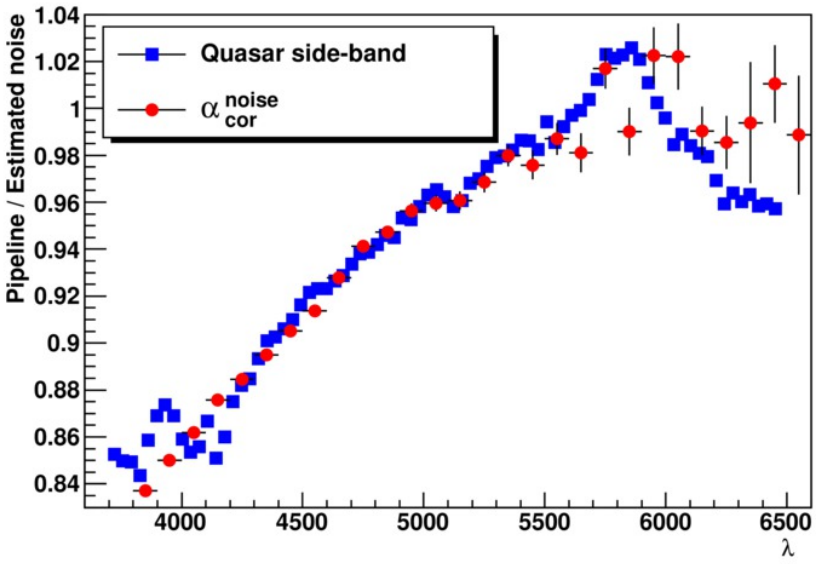


Figure 4.1 – Ratio of the pipeline noise estimate to the actual flux dispersion in the spectra. The blue squares denote this ratio as estimated from the quasar $1330\,\text{Å} \leq \lambda_{\text{RF}} \leq 1380\,\text{Å}$ and $1450\,\text{Å} \leq \lambda_{\text{RF}} \leq 1500\,\text{Å}$ side bands. The red points indicate the correction from our procedure (equation 4.2) as a function of mean forest wavelength.

independent. For each individual quasar, we compute the ratio of the mean pipeline error estimate in the band, $\langle\sigma_p\rangle$, to the root-mean-square (rms) of the pixel-to-pixel flux dispersion within the same band. This quantity is averaged over all DR9 quasars, giving us a wavelength dependent measure of the accuracy of the pipeline noise estimate because of the distribution of quasar redshifts (blue points in figure 4.1). For a perfect noise estimation, the plotted quantity would be unity at all wavelengths; on the other hand, under (over) estimates will produce values below (above) unity. The flux dispersion in the blue part of the spectra ($\lambda \leq 4000\,\text{Å}$) is seen to be about 15% larger than expected from the noise level given by the pipeline. The discrepancy decreases with increasing wavelength, and the two estimates are in agreement at $\lambda \simeq 5700\,\text{Å}$.

This test clearly indicates a wavelength dependent miscalibration of the noise. However, since some of the flux dispersion in the quasar side bands can arise from intervening metals along the sightline (see correction of the metal contribution to the power spectrum in section 4.6.1), this procedure could overestimate the true noise. Lee et al. (2013) provided a per-quasar correction to the pipeline noise that was sufficient for BAO studies, but still not accurate enough for this power-spectrum analysis. Here, we recalibrate the pixel noise for each quasar as described below. This new correction deviates from the one described in Lee et al. (2013) at most by a few percent.

We make use of the typically 4 to 7 individual exposures that contribute to a given quasar coadded spectrum and split them into two interleaved sets: one containing the odd and the other the even exposures. For each set, we compute the weighed average spectrum with weights equal to the pixel inverse variance σ_p^{-2} given by the BOSS pipeline, binned into pixels of width $\Delta \log_{10}(\lambda) = 10^{-4}$ as for the final coadded spectrum. We then compute a "difference spectrum" $\Delta\phi$ by subtracting the spectrum obtained for one set from that for the other set. In this process, we mask all pixels affected by sky emission lines (see section 4.5.1) by setting to 0 the value of the corresponding pixels in the difference spectrum. The difference spectrum should have all physical signals removed and should contain only signal fluctuations. It can therefore be used to

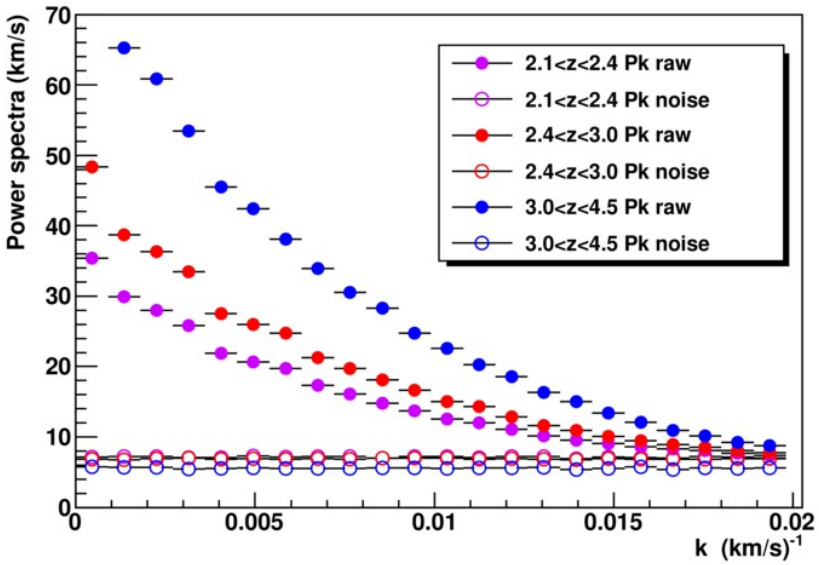


Figure 4.2 – Average power spectra of the raw (filled dots) and of the difference (open circles) signal for three ranges in Lyman-α redshifts.

directly determine the level of noise in the data, irrespective of any miscalibration of the pixel noise in the reduction pipeline. This procedure also has the advantage of evaluating the noise level for each individual spectrum and not on a statistical basis.

We compute the power spectrum of the difference spectrum $P_{\rm diff}^{noise} = |\mathcal{F}(\Delta\phi)|^2$, where $\mathcal{F}$ represents the Fourier transform. In figure 4.2, we plot the average of $P_{\rm diff}^{\rm noise}$ computed over the Lyman-α forest of quasars, for three ranges in Lyman-α redshifts (or equivalently three ranges in observed wavelength). The noise is expected to be white, and $P_{\rm diff}^{\rm noise}$ is indeed seen to be scale independent to an accuracy sufficient for our purposes. For comparison, we also show in the figure the power spectrum of coadded spectra where both signal and noise are present. The noise power spectrum becomes of the same order of magnitude as the raw power spectrum at small scales ($k \sim 0.02\,({\rm km/s})^{-1}$) and small redshifts ($z < 2.4$). This is therefore the region where it is most important to accurately determine its contribution.

We then derive the "pipeline noise power spectrum" $P_{\rm pipe}^{\rm noise}$ from the error σ_p given by the pipeline in each pixel. $P_{\rm pipe}^{\rm noise}$ would be the true noise power spectrum if the pipeline error estimates were correct. For each individual quasar, we thus derive a correction coefficient of the pixel flux error as

$$\alpha_{\rm cor}^{noise} = \sqrt{\frac{\left\langle P_{\rm pipe}^{\rm noise} \right\rangle_k}{\left\langle P_{\rm diff}^{\rm noise} \right\rangle_k}}, \tag{4.2}$$

where the power spectra are computed in both cases over the pixels in the quasar forest and averaged over k. In figure 4.1, the value of the correction term is shown (red dots), averaged over all the DR9 quasars; as before, the distribution of quasar redshifts provides a wavelength dependent measurement. We observe, on average, an excellent agreement between $\alpha_{\rm cor}^{\rm noise}$ and the noise miscalibration estimated in quasar side-bands. In the latter case, however, the estimate is derived from lower-redshift quasars whose side-band covers the same wavelength region as the Lyman-α forest of higher-redshift quasars. The method based on spectrum differences, in contrast, uses the forest data directly and is thus a better estimate of the noise in each quasar

spectrum. For each quasar, the corrected pixel error σ is derived from the pipeline pixel error σ_p by $\sigma(\lambda) = \sigma_p(\lambda)/\alpha_{\rm cor}^{\rm noise}$.

4.2.3 Calibration of Spectrograph Resolution

For each coadded spectrum, the spectral resolution is provided by the BOSS reduction pipeline (Bolton et al., 2012). As the measurement of the 1D power spectrum at small scales is extremely sensitive to the spectrograph resolution, we first investigated the resolution given by the pipeline and we determined a correction table.

4.2.3.1 Spectrograph Resolution in the BOSS Pipeline

As said in chapter 3, in BOSS, spectral lamps are used to provide the wavelength calibration, as described in Smee et al. (2012). In the present work, we are mostly interested in the calibration of the blue CCD. It is obtained from the illumination with a mercury-cadmium arc lamp which has seven principal emission lines in the blue and the green parts of the spectrum.

The spectral resolution is measured from the calibration arc lamp images taken before each set of science exposures. The pipeline procedure fits a Gaussian distribution around the position of each mercury and cadmium line. The mean m_λ and the width σ_λ of the Gaussian determine, respectively, the absolute wavelength on the CCD and the resolution of the spectrograph at the line wavelength. A fourth-order Legendre polynomial is fit to the different σ_λ to model the dispersion over the full wavelength range.

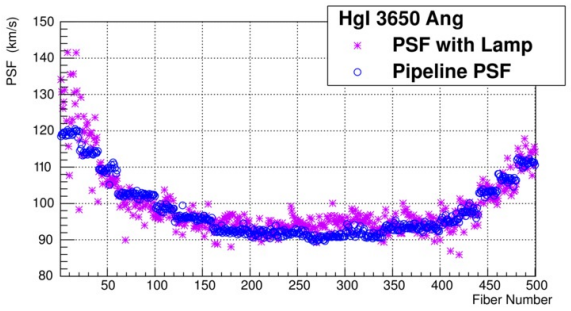
4.2.3.2 Precision of Pipeline Resolution

The BOSS reduction pipeline provides the spectrograph resolution $\sigma(\lambda, i)$ for each pixel i of each spectrum. On a set of plates, we have performed our own Gaussian fits on the mercury and cadmium lines, and we compare our measurement to the resolution given by the BOSS pipeline. We observe systematic shifts that depend on two parameters: wavelength (given by the emission wavelength of the line), and position of the spectrum on the CCD (given by the fibre number). Each CCD has 500 fibres, with numbers 1 and 500 corresponding to CCD edges while numbers near 250 correspond to the central region of the CCD. This comparison is illustrated as a function of fibre number in figures 4.3a, 4.3b and 4.3c corresponding to three lines of mercury and cadmium. The disagreement is at most of the order of a few percent. It is larger in the central region of the CCD and smaller on the edges. The disagreement increases with wavelength and reaches 10% on the blue CCD, near $\lambda = 6000\,\text{Å}$.

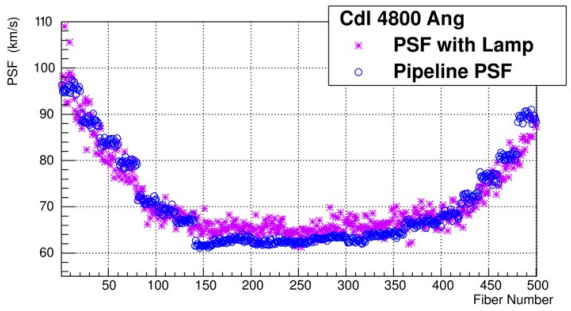
We also check the wavelength calibration using the brightest sky line observed on the blue CCD: the O I line at 5577 Å. The comparison between the BOSS pipeline and our computation of the resolution (see figure 4.3d) shows a similar discrepancy similar to the one observed directly with the mercury arc lamp for similar wavelengths.

4.2.3.3 Correction of Pipeline Resolution

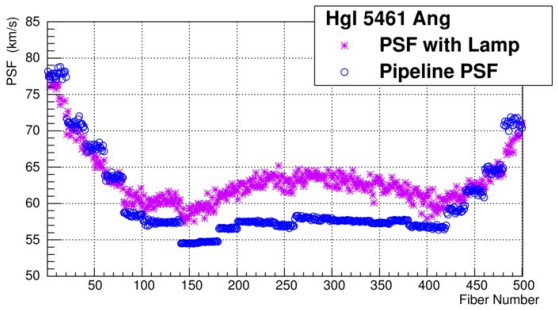
In our analysis, we start from the resolution given by the BOSS reduction pipeline, to which we apply a correction to take into account the discrepancy we observe between the pipeline resolution and our estimate, whether with the arc lamp or a sky line. Figure 4.4a shows the correction as a function of wavelength for spectra in the central region of the CCD. The amplitude of this correction is small, of the order of 10% in the worst case (central spectra and large



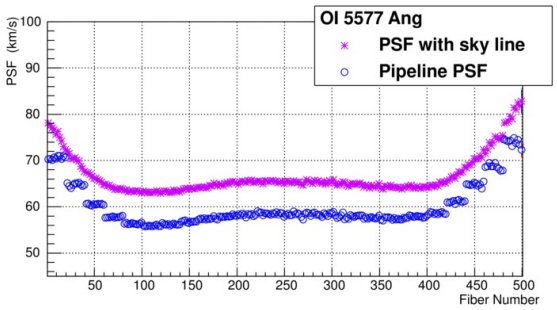
(a) Comparison with the mercury line at 3650 Å from the arc lamp



(b) Comparison with cadmium line at 4800 Å from the arc lamp



(c) Comparison with the mercury line at 5461 Å from the arc lamp



(d) Comparison with the O I sky line at 5577 Å

Figure 4.3 – Comparison of the resolution given by the pipeline (blue circles) and our computation (purple crosses) for arc lamp or sky line as a function of fibre number (ie. position of spectrum on CCD).

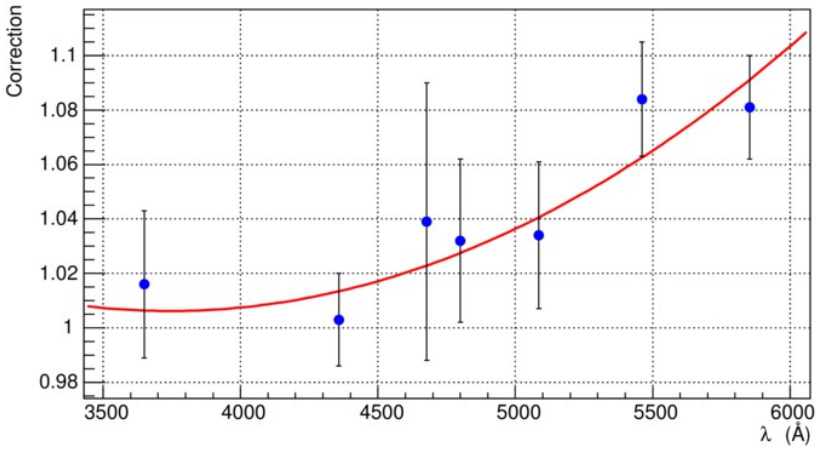
wavelength for the blue CCD). Figure 4.4b shows the 2D correction to the resolution that we apply in our analysis, as a function of wavelength (second-order polynomial) and fibre number (bounded first-order polynomial).

4.3 Quasar Selection and Data Preparation

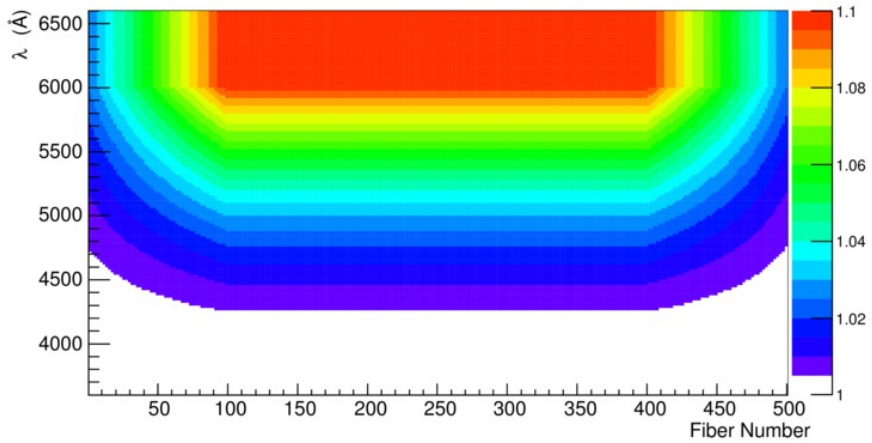
4.3.1 Data Selection

We define the Lyman-α forest by the range $\lambda_{\rm RF} = 1050-1180$ Å, thus at least 7000 $(\mathrm{km/s})^{-1}$ away from the quasar Lyman-β and Lyman-α emission peaks. We limit ourselves to wavelengths above the detector cutoff, i.e., to $\lambda > 3650$ Å, corresponding to an absorber redshift of $z = 2.0$.

The Lyman-α forest spans a redshift range $\Delta z \sim 0.4$ for a quasar at a redshift $z_{\rm qso} = 2.5$, and $\Delta z \sim 0.6$ for a quasar redshift $z_{\rm qso} = 5.0$. In order to improve our redshift resolution to $\Delta z \leq 0.2$ without affecting the k-resolution too much, and at the same time to reduce the computation time for the likelihood approach (see details in section 4.4), we split the Lyman-α forest into one, two or three (depending on the length of the Lyman-α forest) consecutive and non-overlapping sub-regions of equal length, hereafter called "z-sectors". A non-truncated Lyman-α forest contains 507 pixels and is divided into three z-sectors of 169 pixels each. At low redshift ($z_{\rm qso} < 2.5$) the forest extension is limited by the CCD UV cutoff. In practice, the forest is divided into three z-sectors down to a forest length of 180 pixels, into two z-sectors for a forest length comprised between 90 and 180 pixels and not subdivided otherwise. This procedure ensures that the redshift range spanned by a z-sector is at most 0.2.



(a) Correction of the pipeline resolution for spectra in the middle of the CCD (fibre numbers ~ 250). The curve is the best second-order polynomial fit to the measurements at the arc-lamp wavelengths.



(b) Two dimensional correction of the pipeline resolution as a function of fibre number (ie. position of spectrum on CCD) and wavelength.

Figure 4.4 – Correction to the pipeline resolution

With a pixel size $\Delta v = c\Delta\lambda/\lambda = 69\,(\mathrm{km/s})^{-1}$, the smallest k-mode is therefore between $k_{\mathrm{min}} = 5 \times 10^{-4}\,(\mathrm{km/s})^{-1}$ and $k_{\mathrm{min}} = 1 \times 10^{-3}\,(\mathrm{km/s})^{-1}$ depending on the actual z-sector length. Our largest possible mode is determined by the Nyquist-Shannon limit at $k_{\mathrm{Nyquist}} = \pi/\Delta v = 4.5 \times 10^{-2}\,(\mathrm{km/s})^{-1}$, but we limit our analysis to $k_{\mathrm{max}} = 2 \times 10^{-2}\,(\mathrm{km/s})^{-1}$ because of the large window function correction (mostly due to the spectrograph resolution, see figure 4.10) for modes of larger k.

We use the quasars from the DR9 quasar catalogue of BOSS (Pâris et al., 2012). The full catalog contains 61 931 quasars, of which we selected the best 13 821 on the basis of their mean signal-to-noise ratio (SNR) in the Lyman-α forest, spectrograph resolution ($\bar{R}$), and quality flags on the pixels. Flags are also set during the visual scanning of the spectra; we reject all quasars that have broad absorption line features (BAL), damped Lyman-α (DLA) or detectable Lyman limit (LLS) systems in their forest.

The total noise per pixel decreases on average with wavelength, by about a factor of 2 between 3650 Å and 4000 Å and by another factor of 2 between 4000 Å and 6000 Å. We reject quasars with SNR<2, where the SNR is averaged over the Lyman-α forest. This criterion mostly

Criteria	Incremental rejection
Mean forest redshift > 2.15	46%
$SNR > 2.0$	36%
$\bar{R} < 85\,\mathrm{km\,s^{-1}}$	40%
Not BAL	12%
Not DLA	19%

Table 4.1 – Summary of main quasar selection cuts and fraction of quasars passing previous cuts rejected at each step.

removes low-redshift quasars, since they have their Lyman-α forest in the blue, hence noisiest, part of the spectrograph. The spectrograph resolution R varies slightly with wavelength, from typically $82\,\mathrm{km\,s^{-1}}$ (at 1σ) at 3650 Å to $61\,\mathrm{km\,s^{-1}}$ at 6000 Å. It also varies with the position of the spectrum on the CCD (see figure 4.3), with a resolution in the central part that is about $7\,\mathrm{km\,s^{-1}}$ smaller than in the outer regions. We reject quasars with a resolution, averaged over the Lyman-α forest, over $85\,\mathrm{km\,s^{-1}}$ to limit the effect of the velocity resolution in the derived power spectrum. We also remove quasars with pixels in their Lyman-α forest that are masked by the pipeline ($< 2\%$ of the sample). The purpose of these restrictions is to ensure that the systematic uncertainty coming from the precision with which the spectrograph noise and resolution can be calibrated remains smaller than the statistical uncertainty of the estimated power spectra. These uncertainties will be detailed in section 4.5.2. Because both the noise and the resolution are worse in the blue part of the spectrograph, these cuts affect the low-redshift more than the high-redshift quasars. Since the former are also much more numerous, we can thus improve the quality of our sample in a region where the systematic uncertainties would otherwise dominate over the statistical ones.

Table 4.1 summarizes the impact of our cuts on the quasar sample and figure 4.5 shows the distributions of the quasar redshifts and of the z-sector mean redshifts for the quasars and z-sectors that pass these criteria.

In figure 4.6, we show the average quasar spectra obtained by averaging the spectra of all the DR9 BOSS quasars passing the above cuts, split into 5 redshift bins from $z = 2.3$ to 4.3. Broad quasar emission lines are clearly visible, such as Lyman-β at $\lambda_{\mathrm{RF}} \sim 1026$ Å, Lyman-α at $\lambda_{\mathrm{RF}} \sim 1216$ Å, N V at $\lambda_{\mathrm{RF}} \sim 1240$ Å, Si IV at $\lambda_{\mathrm{RF}} \sim 1400$ Å and C IV at $\lambda_{\mathrm{RF}} \sim 1549$ Å. The absorption by Lyman-α absorbers along the quasar lines-of-sight appears blueward of the quasar Lyman-α emission peak, with more absorption (and thus less transmitted flux) at high redshift.

We calculate the 1D power spectra in twelve redshift bins of width Δz=0.2 and centred on $z_c = 2.2$ to $z_c = 4.4$. The mean redshift of the Lyman-α absorbers of a given z-sector determines the redshift bin to which it contributes. While the Lyman-α forest of a quasar spectrum may cover several redshift bins, a given z-sector only contributes to a single bin, thus avoiding correlations between redshift bins. The redshift span of a z-sector, at most 0.2, is well adapted to the size of our redshift bins.

4.3.2 Sky Lines Masking

Sky lines affect the data quality by increasing significantly the pixel noise. The procedure used to identify them is detailed in Lee et al. (2013). It is briefly summarized here.

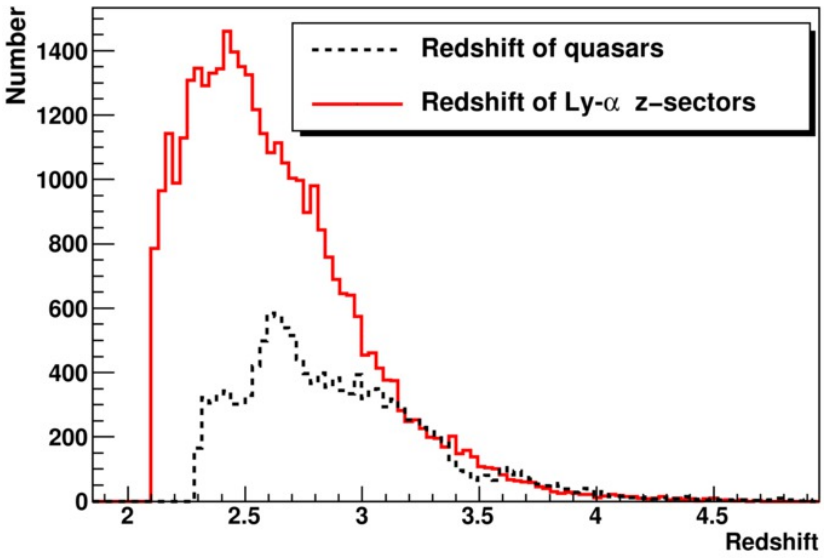


Figure 4.5 – Redshift distribution of the 13 821 quasars selected in the analysis, and mean redshift distribution of each z-sector of their Lyman-α forest.

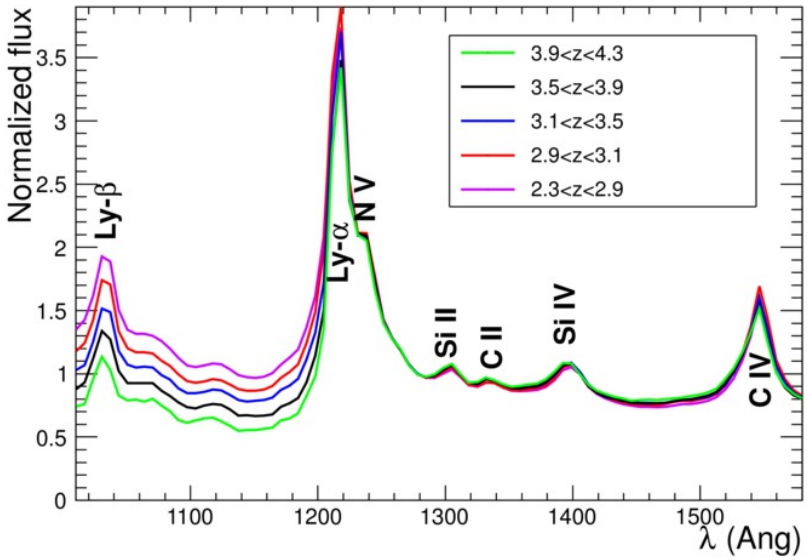


Figure 4.6 – Average quasar spectra in five redshift bins. All spectra are normalized at $\lambda = 1280\,\text{Å}$.

We use the sky calibration fibers, and compute the mean and the root mean square of the residuals measured on the sky-subtracted spectrum obtained with the standard BOSS pipeline. We define a "sky continuum" as the running average of the residual rms fluctuation centered around a ±25 pixel window, and generate a list of sky lines from all the wavelengths that are 25% above the sky continuum. The continuum, measured with the unmasked pixels, and the sky line list are iterated until they converge. To this list, we add the calcium H and K galactic absorption lines near $\lambda = 3934\,\text{Å}$ and $\lambda = 3966\,\text{Å}$. We then mask all pixels that are within 1.5 Å of the listed wavelengths.

We apply the mask differently in the Fourier transform and the likelihood methods. For the Fourier transform, we replace the flux of each masked pixel by the average value of the flux over the unmasked forest. This procedure introduces a k-dependent bias in the resulting power spectrum that reaches at most 15% at small k for the $3.5 < z \leq 3.7$ redshift bin, which contains 5577 Å, the strongest sky emission line (O I). We correct this bias a posteriori, as explained in section 4.5.1. For the likelihood method, the masked pixels are simply omitted from the data vector. We have checked (see details in section 4.5.1) that in this case we observe no bias on the resulting power spectrum.

4.3.3 Quasar Continuum

The normalized transmitted flux fraction $\delta(\lambda)$ is estimated from the pixel flux $f(\lambda)$ by:

$$\delta(\lambda) = \frac{f(\lambda)}{f_{\text{qso}}^{1280} \times C_{\text{qso}}(\lambda, z_{\text{qso}}) \times \bar{F}(z_{\text{Ly}\alpha})} - 1, \tag{4.3}$$

where f_{qso}^{1280} is a normalization equal to the mean flux in a 20 Å window centered on $\lambda_{RF} = 1280\,\text{Å}$, $C_{\text{qso}}(\lambda, z_{\text{qso}})$ is the normalized unabsorbed flux (the mean quasar "continuum") and $\bar{F}(z_{\text{Ly}\alpha})$ is the mean transmitted flux fraction at the H I absorber redshift. Pixels affected by sky line emission are not included when computing the normalization. Since the mean quasar continuum is flat in the normalization region, the rejection of a few pixels does not bias the mean pixel value. The product $C_{\text{qso}}(\lambda, z_{\text{qso}}) \times \bar{F}(z_{\text{Ly}\alpha})$ is assumed to be universal for all quasars at redshift z_{qso} and is computed by stacking appropriately normalized quasar spectra f/f_{qso}^{1280}, thus averaging out the fluctuating Lyman-α absorption. The product $f_{\text{qso}}^{1280} \times C_{\text{qso}}(\lambda, z_{\text{qso}}) \times \bar{F}(z_{\text{Ly}\alpha})$ represents the mean expected flux, and the transmitted flux fraction is given by $F(\lambda) = f(\lambda)/(f_{\text{qso}}^{1280} \times C_{\text{qso}}(\lambda, z_{\text{qso}}))$. For a pixel at rest-frame wavelength λ_{RF} of a quasar at redshift z_{qso}, the corresponding H I absorber redshift $z_{\text{Ly}\alpha}$ can be inferred from $1 + z_{\text{Ly}\alpha} = \lambda_{\text{RF}}/\lambda_{\text{Ly}\alpha} \times (1 + z_{\text{qso}})$, where $\lambda_{\text{Ly}\alpha} \simeq 1216\,\text{Å}$.

Figure 4.7 shows the product $C_{\text{qso}}(\lambda, z_{\text{qso}})\bar{F}(z_{\text{Ly}\alpha})$ of the quasar continuum with the mean transmitted flux fraction as a function of rest-frame wavelength and Lyman-α redshift. Figure 4.8 shows the projections of the 2D distribution of figure 4.7 onto the redshift and the wavelength axis. The former shows $\left\langle f(\lambda)/f_{\text{qso}}^{1280}) \right\rangle_\lambda$ averaged over wavelength and is proportional to the mean transmitted flux fraction, and the latter shows the mean unabsorbed quasar spectrum $C_{\text{qso}}(\lambda)$ normalized to f_{qso}^{1280}. The mean transmitted flux fraction is well fitted by a function of the form $\exp\left[-\alpha(1+z)^\beta\right]$, with $\alpha \sim 0.0046$ and $\beta \sim 3.3$, in agreement with previous measurements of the effective optical depth τ_{eff} where $\bar{F} \propto \exp(-\tau_{\text{eff}})$ (see e.g. Meiksin (2009) for a review and Becker et al. (2011) for a more recent measurement).

The values in the 2D table, $C_{\text{qso}}(\lambda, z_{\text{qso}}) \times \bar{F}(z_{\text{Ly}\alpha})$, differ from those of the product $C_{\text{qso}}(\lambda) \times \bar{F}(z_{\text{Ly}\alpha})$ by up to 5%, possibly due to variations in the mean quasar continuum with redshift.

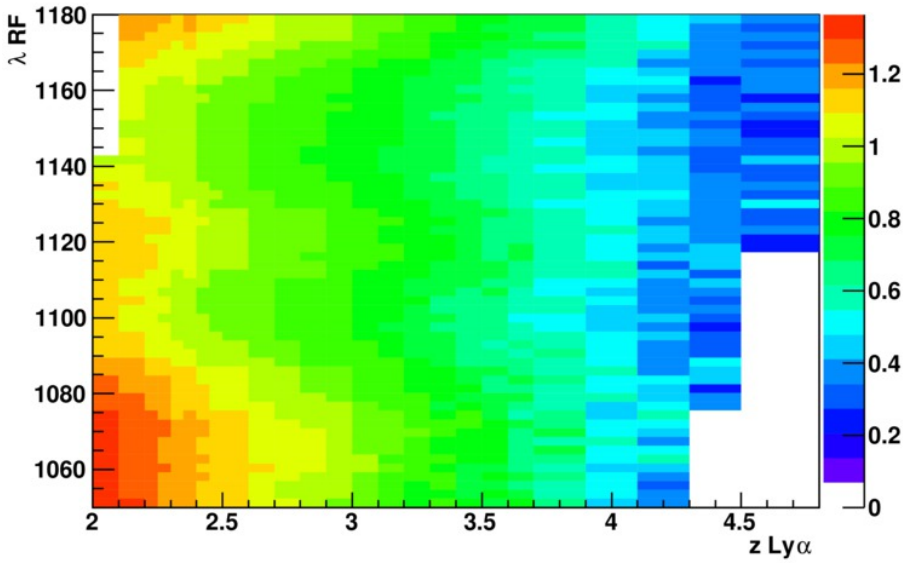


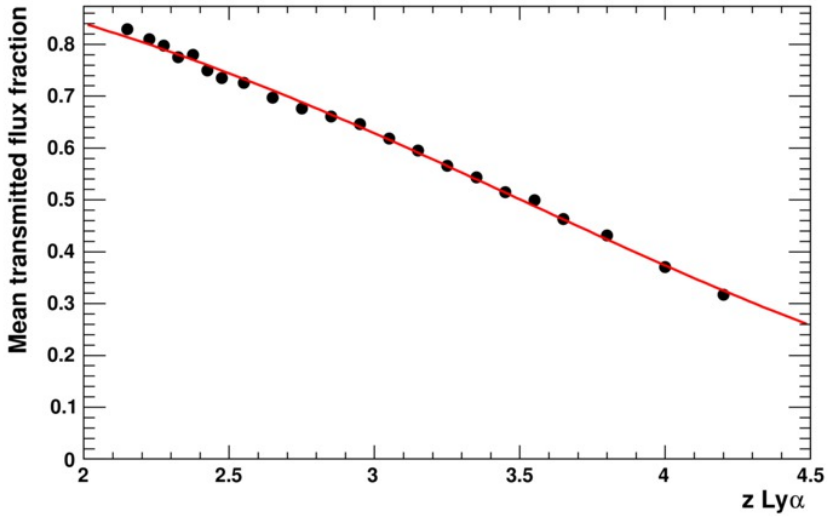
Figure 4.7 – Product of the quasar continuum $C_{\rm qso}(\lambda, z_{\rm qso})$ by the mean transmitted flux fraction $\bar{F}(z)$ as a function of rest-frame wavelength and Lyman-α redshift. This 2D table is used to compute the normalized flux transmission fraction $\delta(\lambda)$.

Despite its lower statistical precision for a given wavelength and redshift, we therefore use the 2D table.

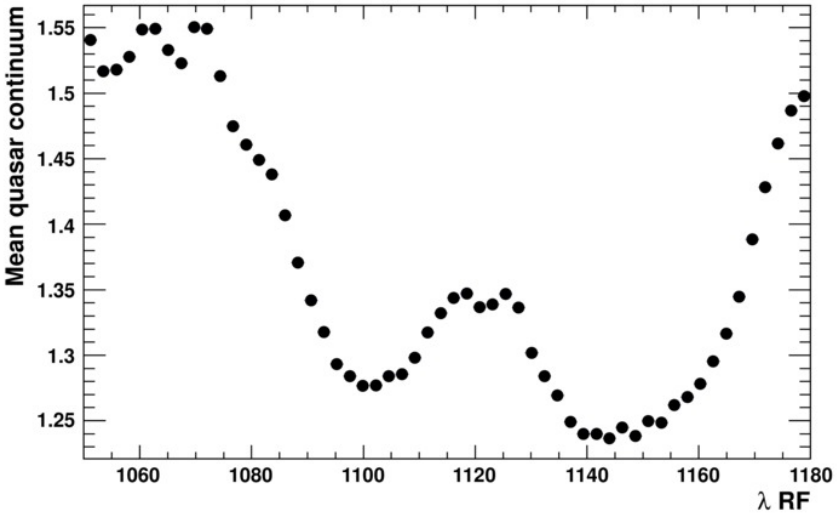
Figure 4.9 shows the resulting mean $\delta(\lambda)$ as a function of the observed wavelength. The mean fluctuates around zero at the 2% level with correlated features that are due to the imperfect spectrograph calibration and absorption. These features include the calcium H-K doublet (3934 Å, 3968 Å) from Milky Way absorption, and Balmer lines Hγ, Hδ and Hϵ (4341 Å, 4102 Å, 3970 Å) that are residuals from the use of F-stars as spectrocalibration standards. Busca et al. (2013) have studied these features in detail and concluded that they had quasar-to-quasar variations of less than 20% of the mean Balmer artifact deviations. To remove their contribution to the Lyman-α power spectrum, we subtract the mean residual of figure 4.9 from $\delta(\lambda)$.

4.4 Methods for Determining $P(k)$

We apply two methods to compute the one-dimensional power spectrum. The first one is based on a Fourier transform. It is fast and robust, thus allowing many tests leading to a better understanding of the impact of the different ingredients entering the analysis. We use it to test the impact of, for instance, different selections of quasars on the precision of the resulting power spectra, or various algorithms to mask sky emission lines. The second method relies upon a maximum likelihood estimator in real space. It can take into account variations in the noise or in the spectrograph resolution at the pixel level instead of through global factors, and is therefore expected to be more precise than a Fourier transform. It also offers a natural way to mask pixels affected by sky emission lines, as explained in section 4.4.2. However, it is more sensitive than the Fourier transform to details in the implementation of the method, it is susceptible to convergence problems in the presence of noisy spectra and is more time-consuming. It is therefore not as flexible for algorithm testing. The power spectra obtained



(a) Mean transmitted flux fraction $\bar{F}(z_{\mathrm{Ly}\alpha})$ as a function of Lyman-α absorber redshift. The overlaid curve is $\exp\left[-0.0046 \times (1+z)^{3.3}\right]$.



(b) Mean quasar continuum $C_{\mathrm{qso}}(\lambda)$ as a function of rest-frame wavelength, averaged over all selected quasars.

Figure 4.8 – Projection of the 2D distribution of figure 4.7 onto the redshift and the wavelength axis

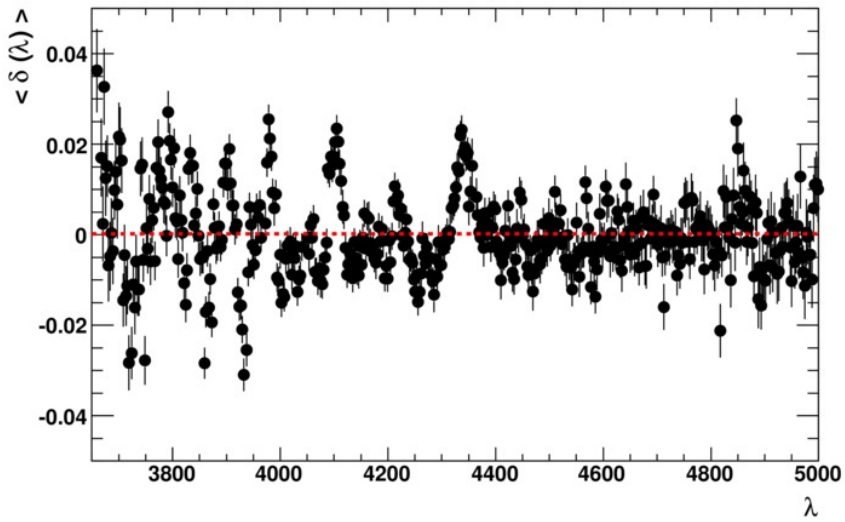


Figure 4.9 – Mean of $\delta(\lambda)$ as a function of wavelength in Å. Systematic offsets from zero are seen at the 2% level due to imperfections in the spectrograph calibration.

with the two approaches are in good agreement. Their comparison provides an estimate of the systematic uncertainty on our measurement (see section 4.6.3).

4.4.1 Fourier Transform Approach

4.4.1.1 Measurement of the Power Spectrum with a Fourier Transform

To measure the one-dimensional power spectrum $P_{1D}(k)$ we decompose each absorption spectrum $\delta(\lambda)$ into Fourier modes and estimate their variance as a function of wave number k. In practice, we do this by computing the discrete Fourier transform of the flux transmission fraction $\delta = F/\langle F\rangle - 1$ as described in Croft et al. (1998), using a fast Fourier transform (FFT) algorithm. The use of a FFT requires that the pixels be equally spaced. This condition is satisfied with the quasar coadded spectra provided by the SDSS pipeline (Bolton et al., 2012): the spectra are computed with a constant pixel width $\Delta\left[\log(\lambda)\right] = 10^{-4}$, and the velocity difference between pixels, i.e., the relative velocity of absorption systems at wavelengths $\lambda + \Delta\lambda/2$ and $\lambda - \Delta\lambda/2$, is

$$\Delta v = c\frac{\Delta\lambda}{\lambda} = c\Delta\left[\ln(\lambda)\right] = c\ln(10)\Delta\left[\log(\lambda)\right]. \tag{4.4}$$

The coadded spectra thus have pixels equally spaced in Δv. Throughout this paper we therefore use velocity instead of observed wavelength. Similarly, the wave vector $k \equiv 2\pi/\Delta v$ is measured in $(\mathrm{km/s})^{-1}$.

In the absence of instrumental effects (noise and resolution of the spectrograph), the one-dimensional power spectrum can be simply written as the ensemble average over quasar spectra of $P^{\mathrm{raw}}(k) \equiv \left|\mathcal{F}\left(\delta(\Delta v)\right)\right|^2$, where $\mathcal{F}\left(\delta(\Delta v)\right)$ is the Fourier transform of the normalized flux transmission fraction $\delta(\Delta v)$ in the quasar Lyman-α forest binned in pixels of width Δv.

When taking into account the noise in the data and the impact of the spectral resolution of the spectrograph, δ can be expressed as $\delta = s + n$, with s the signal and n the noise, and the

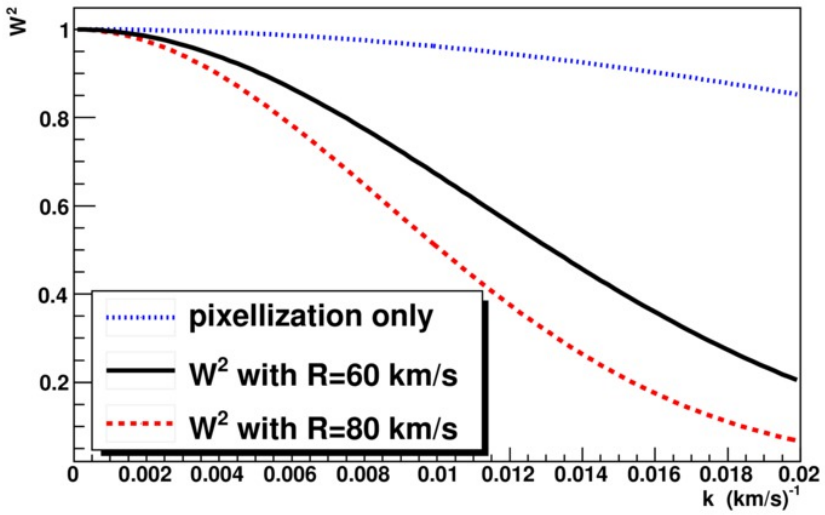


Figure 4.10 – Window function $W^2(k, R, \Delta v)$, with $\Delta v = 69\,\mathrm{km\,s^{-1}}$, reproducing the spectrum binning and the impact of the spectrograph resolution, for a resolution $R = 60\,\mathrm{km\,s^{-1}}$ typical at $\lambda > 5000\,\text{Å}$ and $R = 80\,\mathrm{km\,s^{-1}}$ typical at $\lambda < 4300\,\text{Å}$. For comparison, we also show the contribution from the pixellization only (equivalent to $R = 0$).

estimator of the one-dimensional power spectrum is

$$P_{1D}(k) = \left\langle \frac{P^{\mathrm{raw}}(k) - P^{\mathrm{noise}}(k)}{W^2(k, R, \Delta v)} \right\rangle, \tag{4.5}$$

where $\langle\rangle$ denotes the ensemble average over quasar spectra and where

$$P^{\mathrm{noise}}(k) \equiv |\mathcal{F}(n(\Delta v))|^2 . \tag{4.6}$$

The window function corresponding to the spectral response of the spectrograph is defined by

$$W(k, R, \Delta v) = \exp\left[-\frac{(kR)^2}{2}\right] \times \frac{\sin(k\Delta v/2)}{(k\Delta v/2)}, \tag{4.7}$$

where Δv and R are respectively the pixel width and the spectrograph resolution. Both quantities are in $\mathrm{km\,s^{-1}}$, and R should not be confused with the dimensionless resolving power of the spectrograph. We illustrate in figure 4.10 the effect of the spectrograph resolution on the window function $W^2(k, R, \Delta v)$ for different values of R.

The power spectrum obtained with this simple analysis is given in figure 4.11. In this power spectrum two main effects can be seen: the upturn on small scales, which can mostly be explained by an underestimation of the noise by the pipeline (especially visible on the lower redshifts), and the excess of power mostly present in the $z \sim 3.6$ bin due to the presence of the strong O I sky emission line at 5577 Å. Other sky emission lines are affecting the power spectrum at all redshifts. The effect of uncertainties in the modelling of the spectrograph resolution is here dominated by the other two. This justifies the corrections presented in section 4.2.

4.4.1.2 Computation of $P_{1D}(k)$ with a FFT

We compute the Fourier transform using the efficient FFTW package from Frigo and Johnson (1998). Compared to the likelihood approach described in the next section, the Fourier transform

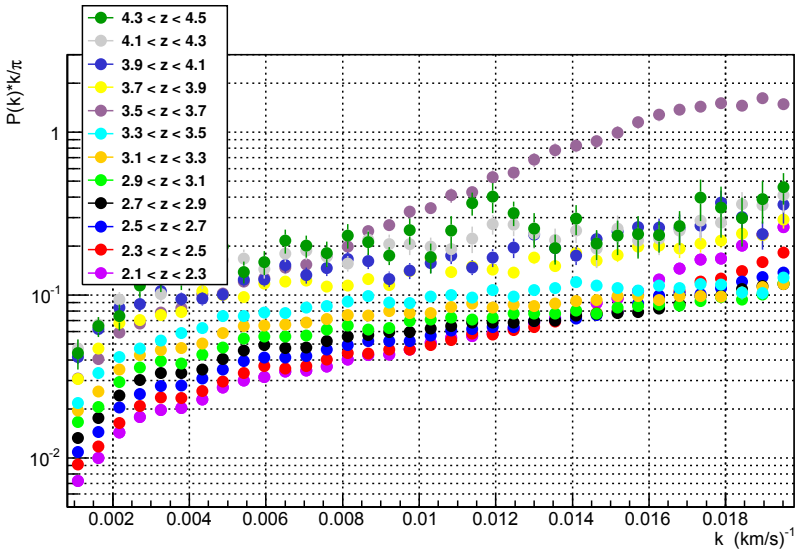


Figure 4.11 – One-dimensional power spectrum with the noise and resolution taken directly from the pipeline. The sky lines are not corrected.

is much faster, but it requires some simplifying hypotheses in the treatment of the noise and of the spectrograph resolution. We explain these simplifications below. Sky emission lines are also treated in a simplified way as described in section 4.3.2. The results provided by this simple method are complementary to the likelihood approach.

Although the redshift of the absorbing neutral hydrogen increases with wavelength along the spectrum of a given quasar, the power spectrum is considered to be computed at its average redshift. As explained in section 4.3.1, we improve the redshift resolution of the measured power spectra by splitting the Lyman-α forest of each quasar into redshift z-sectors (see section 4.3.1). The computation is done separately on each z-sector instead of on the entire Lyman-α forest. The mean redshift of the Lyman-α absorbers in the z-sector determines the redshift bin to which the z-sector contributes.

The noise power spectrum $P^{\rm noise}(k,z)$ is taken as the power spectrum $P^{\rm noise}_{\rm diff}$ on the z-sector, computed as explained in section 4.2.2. Since $P^{\rm noise}_{\rm diff}$ is flat with k, we improve the statistical precision on our determination of the level of the noise power spectrum by taking the average of $P^{\rm noise}_{\rm diff}(k)$ for $k < 0.02\,({\rm km/s})^{-1}$.

Finally, we apply the correction of the spectrograph resolution and pixellization by dividing by $W^2(k,\bar{R},\Delta v)$, where $\bar{R}$ is the mean value of the spectrograph resolution R averaged over the z-sector. The value of R is given by the pipeline and corrected by following the prescription described in section 4.2.3. For a given spectrum, R varies by less than 10% over the Lyman-α forest (less than 3% over a z-sector), and the impact of this simplification is negligible.

We rebin the final power spectrum onto a predefined grid in k-space, giving equal weight to the different Fourier modes that enter each bin. The final 1D power spectrum is obtained by averaging the corrected power spectra of all the contributing z-sectors of all selected quasars, as expressed in equation 4.5.

4.4.2 Likelihood Approach

We also estimate $P_{1D}(k)$ using a maximum likelihood estimator derived from methods developed for studies of the cosmic microwave background anisotropy (Bond et al., 1998; Seljak, 1998a,b). This method guarantees optimal performance for Gaussian or nearly Gaussian distributions, and can be applied here ensuring minimal variance, although the power spectrum estimates are not Gaussian distributed. Our approach involves a direct maximization of the likelihood function and is not based on the quadratic maximum estimation as in McDonald et al. (2006). It is slower but provides the values of $P_{1D}(k)$ with their covariance matrix at the maximum of the likelihood.

4.4.2.1 The Likelihood Function

We model the normalized flux transmission fraction $\delta_l = F_l/\langle F\rangle - 1$ measured in pixel l as the sum of contributions from signal and noise: $\delta_l = s_l + n_l$. We assume that signal and noise are independent, with zero mean and that there is no correlation between the noise of different pixels. Thus the corresponding covariance matrices are given by

$$C_{lm}^{\text{signal}} = \langle s_l s_m\rangle \quad \text{and} \quad C_{lm}^{\text{noise}} = \langle n_l n_m\rangle = \sigma_l\sigma_m\delta_{lm}, \tag{4.8}$$

where δ_{lm} is the Kronecker symbol and $\sigma_l = \sigma_l^{\text{pipeline}}/\alpha_{\text{cor}}^{\text{noise}}$ (pipeline estimate and its correction). The total covariance matrix can therefore be written as:

$$C_{lm} = \langle \delta_l\delta_m\rangle = C_{lm}^{\text{signal}} + C_{lm}^{\text{noise}}. \tag{4.9}$$

The signal covariance matrix can be derived from the 1D power spectrum by

$$C_{lm}^{\text{signal}} = \int_{-\infty}^{+\infty} P_{1D}(k)\times\exp\left[-ik\Delta v\times(l-m)\right]\mathrm{d}k \tag{4.10}$$

$$= \int_{0}^{+\infty} P_{1D}(k)\times 2\cos\left[k\Delta v\times(l-m)\right]\mathrm{d}k. \tag{4.11}$$

$$\tag{4.12}$$

We can approximate $P_{1D}(k)$ by $\mathbf{P} = (P_1,\ldots,P_i,\ldots,P_N)$, a discrete set of N values of $P_i \equiv P_{1D}\left(\frac{k_{i-1}+k_i}{2}\right)$ for the mode sample $(k_0,\ldots,k_i,\ldots,k_N)$. The previous integral can then be approximated by:

$$C_{lm}^{\text{signal}}(\mathbf{P}) = \sum_{i=1}^{N} P_i \int_{k_{i-1}}^{k_i} 2\cos\left[k\Delta v\times(l-m)\right]\mathrm{d}k. \tag{4.13}$$

Taking the spectrograph resolution into account and using the definition of the window function given in equation 4.7, the covariance matrix becomes:

$$C_{lm}^{\text{signal}}(\mathbf{P}) = \sum_{i=1}^{N} P_i \int_{k_{i-1}}^{k_i} 2\cos\left[k\Delta v\times(l-m)\right]\times W(k,R_l,\Delta v)\times W(k,R_m,\Delta v)\mathrm{d}k \tag{4.14}$$

where R_i and Δv are respectively the spectrograph resolution for pixel i and the pixel width (which is the same for all pixels here but could have been pixel dependent).

For spectrum i containing M_i pixels, we can define the likelihood function $\mathcal{L}_i$ as:

$$\mathcal{L}_i(\mathbf{P}) = \frac{1}{(2\pi)^{M_i/2}\sqrt{\det(C)}} \exp\left(-\frac{\delta^T C^{-1}\delta}{2}\right) \tag{4.15}$$

For stability reasons, we do not fit a single spectrum at a time but instead combine N_{sp} spectra corresponding to the same redshift bin into a common likelihood. The likelihood is the product:

$$\mathcal{L}(\mathbf{P}(z)) = \prod_{i=1}^{N_{sp}} \mathcal{L}_i(\mathbf{P}(z)). \tag{4.16}$$

We can then search for the vector $\mathbf{P}(z)$ (i.e. the parameters $P_i(z)$) that maximizes this likelihood.

4.4.2.2 Extraction of $P_{1D}(k)$

We extract the $P_{1D}(k)$ power spectrum from the likelihood $\mathcal{L}(\mathbf{P}(z))$ that combines several spectra in the same redshift bin. We use the `MINUIT` (James and Roos, 1975) library to minimize the term $-2\ln(\mathcal{L})$. This minimization provides the value and the error of each $P_i(z)$ and the covariance matrix between the different $P_i(z)$. The method implemented in `MINUIT` may be slow but it is robust; it rarely falls into secondary minima.

As the minimization can take a few hundred iterations, I have optimized our fitting procedure. The computation time of the likelihood is limited by the inversion of the covariance matrix C. Therefore, to reduce the size of the matrix C (number of pixels), we do the computation on the z-sectors defined in section 4.3.1, instead of on the entire Lyman-α forest.

As said above, the noise covariance matrix is assumed diagonal, i.e., without correlation terms. Each diagonal element is equal to the square of the pixel error estimated by the pipeline, $\sigma_i^{\text{pipeline}}$, multiplied by the square of the correction factor $\alpha_{\text{cor}}^{\text{noise}}$ defined in equation 4.2.

I use the Cholesky factorization [1] to increase the speed of the matrix inversion of the positive-definite matrix C. The Cholesky decomposition is roughly twice as efficient as the usual LU decomposition and it is numerically more precise.

Finally, in the product of the individual likelihoods of equation 4.16, we take $N_{sp} = 100$ where in practice N_{sp} is the number of z-sectors and not the number of quasar spectra (the z-sectors of a given spectrum can contribute to different redshift bins). While a large N_{sp} improves the fit convergence by making the fit more stable, we nevertheless restrict the number of z-sectors to be fitted simultaneously in order to limit the minimization to a reasonable CPU time. I also tested that convergence has been reached at better than the percent-level at $N_{sp} = 100$ and that there is no need to further increase N_{sp}, as shown on figure 4.12. We determine the final $\mathbf{P}(z)$ by averaging over the N_b bunches of N_{sp} z-sectors (with $N_b \times N_{sp}$ being the total number of z-sectors that enter a given redshift bin). The total covariance matrix M_{cov}^{tot} is computed as $(M_{\text{tot}}^{\text{cov}})^{-1} = \sum_{i=1}^{N_b} (M_i^{\text{cov}})^{-1}$.

The typical CPU time for the minimization of one bunch of 100 z-sectors is about 10 to 15 minutes, performing between five and six hundred iterations before convergence. For this analysis, we ran on a farm of wenty-four computers, which allowed us to compute the independent power spectra for different redshift bins in parallel. The total wall-clock time for the full analysis is approximately twelve hours.

1. Cholesky factorization is a decomposition of a Hermitian, positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose: $A = LL^\star$ or $A = LL^T$ if A is real.

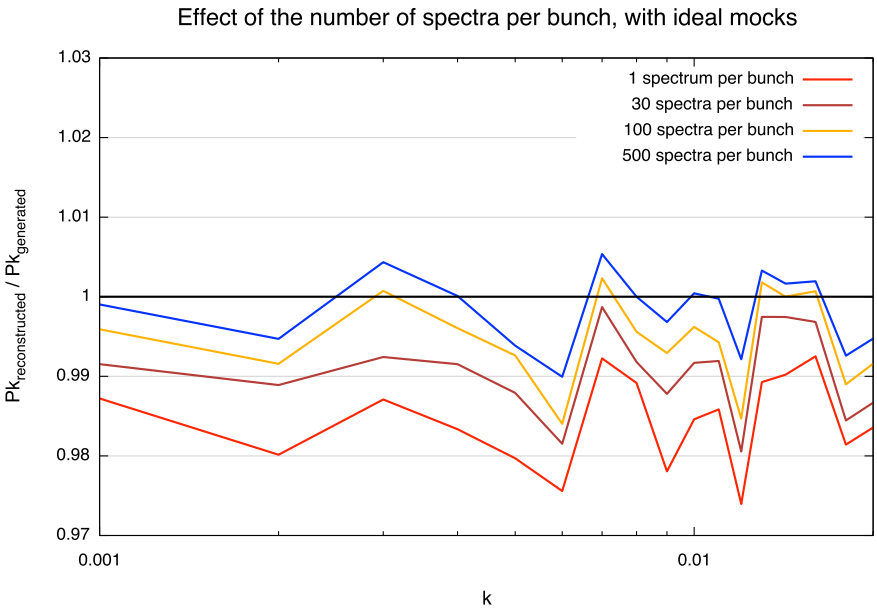


Figure 4.12 – Ratio of the reconstructed and generated power spectra for different number of spectra per bunch. We use 2000 ideal mock spectra with infinite resolution and no noise.

4.5 Systematic Uncertainties

In this section, we study the biases and systematic uncertainties that affect our analysis. We correct our results for the identified biases, and we estimate systematic uncertainties that we summarize in section 4.6 [2].

The biases and uncertainties arise from two different origins. In section 4.5.1, the biases related to the analysis methods, either Fourier transform or likelihood, are presented assuming that the instrumental noise and resolution are perfectly known. Then in section 4.5.2, the systematics due to our imperfect knowledge of the instrument characteristics are described and quantified using the data themselves.

4.5.1 Biases in the Analyses and Related Systematics

We study here the biases and systematic uncertainties introduced at each step of the data analysis. We estimate their impact using mock spectra. We compute the "bias" of the method as the ratio of the measured flux power spectrum to the flux power spectrum that was generated in the mock spectra.

We generate mock spectra with the following procedure. First, a redshift and a magnitude in band g are chosen at random from the real BOSS spectra. Second, an unabsorbed flux spectrum is drawn for each quasar from a random selection of principal component analysis amplitudes following the procedure of Pâris et al. (2011), and flux normalized to the selected magnitude. Third, the Lyman-α forest absorption is generated following a procedure adapted from Font-Ribera et al. (2012) in which the authors provide an algorithm for generating any spectrum of the transmitted flux fraction $F(\lambda)$ from a Gaussian random field $g(\lambda)$. Specifically, they present a recipe for choosing the parameters a and b and the power spectrum $P_g(k)$ such that

2. Full results tables are available at `http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/559/A85`

the transformation $F(\lambda) = \exp\left[-a\exp\left(bg(\lambda)\right)\right]$ yields the desired power spectrum and mean value of $F(\lambda)$. In practice we generate a suite of transmitted flux fraction spectra for twelve redshifts that reproduce the observed power. For each wavelength pixel, $F(\lambda)$ is obtained by interpolation between redshifts according to the actual Lyman-α absorption redshift of the pixel. The unabsorbed flux is multiplied by $F(\lambda)$ and convolved with the spectrograph resolution. In practice, the spectra are generated with a pixel width that is one third of a SDSS pixel, and about one third of the spectral resolution. We checked that this size was small enough to properly take into account the spectral resolution. Finally, noise is added according to BOSS throughput and sky noise measurements as was done in Le Goff et al. (2011), and the spectrum is rebinned to the SDSS pixel size.

The determination of the transmitted flux fraction requires an estimate of the quasar unabsorbed flux obtained as explained in section 4.3.3 and equation 4.3. As a starting point, we have checked that using the generated values for the quasar continuum $C_{\rm qso}(\lambda)$ and for the mean transmitted flux $\bar{F}(z)$ allows recovery of exactly the input power spectra in the absence of noise. Using instead our estimated value of $C_{\rm qso}(\lambda, z_{\rm qso}) \times \bar{F}(z)$ produces an overestimate of the power spectrum of the order of 2%, and is k-independent over the range of interest. To have a better estimate of the continuum on a quasar-by-quasar basis and allow for tilts in the flux calibration, we considered a more sophisticated method consisting of multiplying the average shape by a factor $A + B\lambda_{\rm RF}$ where A and B were fitted for each quasar. This method was not retained, however, because it generated a larger overestimate ($\sim 6\%$).

We studied the impact of our correction for the spectrograph spectral resolution $W(k, R, \Delta v)$ by using mocks where W was either similar to that of BOSS (including both pixellization and spectrograph resolution), or was reduced to the contribution of pixellization only (see figure 4.10). We found a negligible bias (less than 0.1%) in both the Fourier transform method and the likelihood method.

The removal of the noise contribution to the Lyman-α power spectrum introduces a bias in both methods. For mock spectra, the noise power spectrum is white and we determine its level directly from the pixel errors. For the Fourier transform approach, the removal of the noise power spectrum on mock spectra analysed with the true quasar continuum produces a small (2%) underestimate.

The likelihood method is much more sensitive than the Fourier transform approach to the level of noise and to the relative levels of noise and signal power spectra. It results in biases that can reach $\sim 13\%$ at low redshift ($z < 2.3$) and small scales ($k > 0.015$), where noise is high and signal is low (see figure 4.2). The cause of this bias has not been identified. To correct it, we produced mock spectra covering the range in $P^{\rm noise}$ and $P^{\rm raw}$ observed in the data. While the noise is white, the k-dependence of $P^{\rm raw}$ provides, with each power spectrum, a wide range of relative values of $P^{\rm noise}$ and $P^{\rm raw}$. We measured a systematic overestimate of the power spectrum (see figure 4.13), which we modelled by $c_0 + c_1 \times P^{\rm noise}/P^{\rm raw} + c_2 \times P^{\rm noise}$. We found $c_0 = 0.999$, $c_1 = 0.082$ and $c_2 = 0.007$. This bias is determined from a full analysis (determination of the quasar continuum, correction for spectrograph resolution and for noise); it thus takes into account the systematic biases from all the above steps. We assign a systematic uncertainty on the resulting power spectrum equal to 30% of the correction (see figure 4.14).

The masking of the sky emission lines is implemented in different ways in the two analysis methods. In the likelihood approach, where the relevant pixels are simply omitted, the masking procedure results in no measurable bias. For the Fourier transform approach, we estimate the impact of the masking procedure by applying it on mock spectra that do not include emission

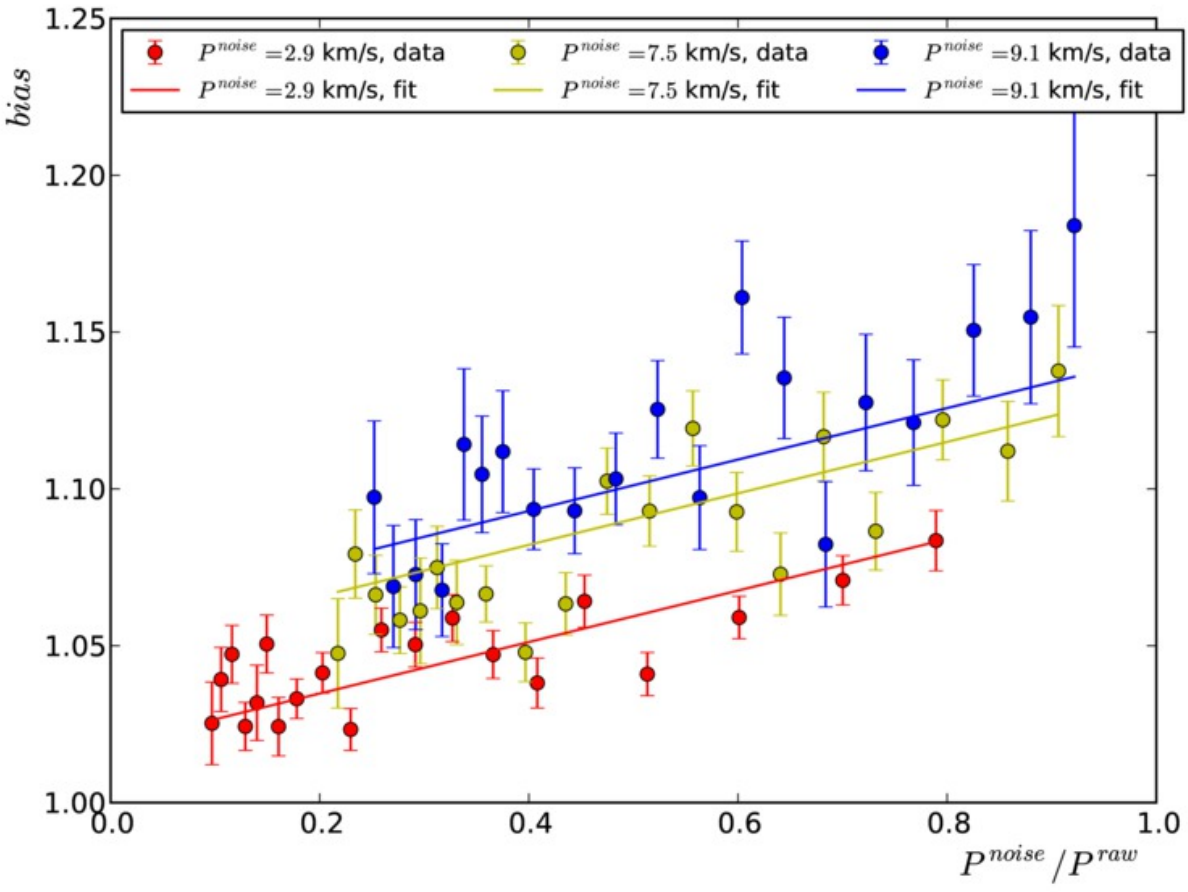


Figure 4.13 – Overestimation of the mock power spectrum determined from the likelihood method as a function of $P^{\rm noise}/P^{\rm raw}$, for different values of $P^{\rm noise}$. The curves illustrate the best fit model.

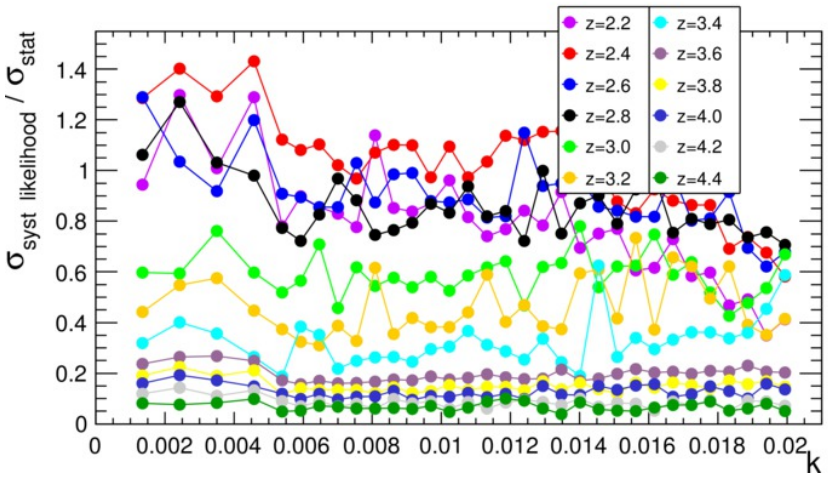


Figure 4.14 – Systematic uncertainty related to the correction of the noise-related bias in the likelihood method, relative to the statistical uncertainty, for each redshift bin.

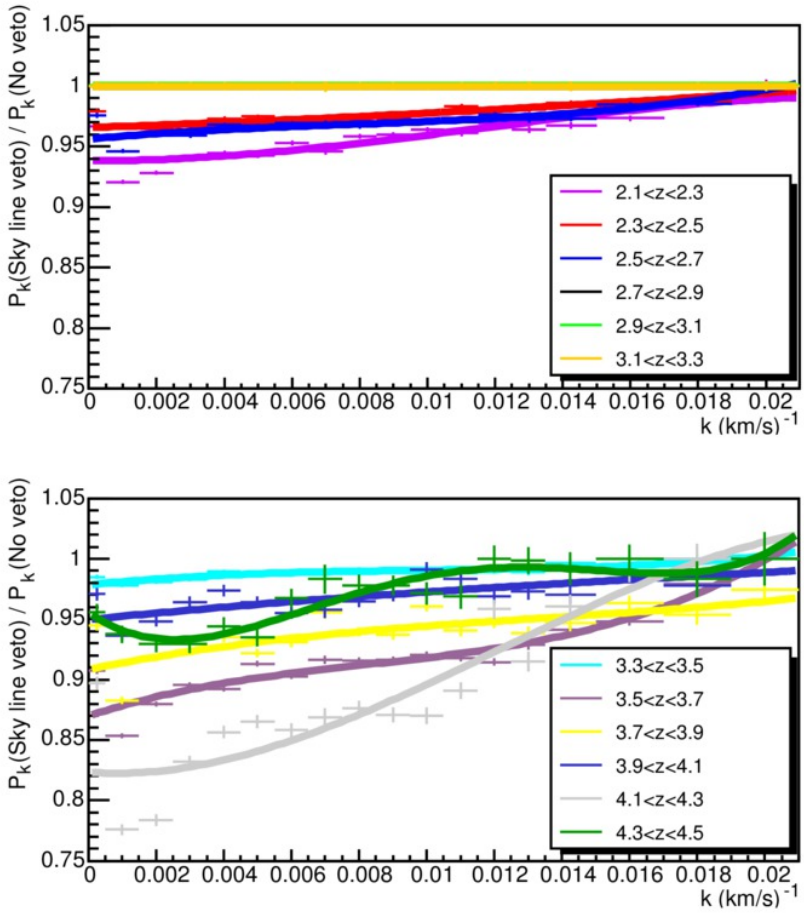


Figure 4.15 – Underestimation of the power spectrum due to the masking of the sky emission lines for the Fourier transform approach. The curves are polynomial fits to the measured k-dependent bias for each redshift bin. No strong sky lines enter the forest in $2.7 < z < 3.3$, implying no systematic uncertainty in this redshift range.

from sky lines. The result is illustrated in figure 4.15. No strong sky line enters the forest for the redshift range $2.7 \leq z \leq 3.3$, which explains why no bias is observed in the corresponding redshift bins. The largest bias occurs for large-scale modes where most of the effect is related to the relative number of masked pixels in the forest. The effect on small scales is more sensitive to the distribution and size of the masked regions. The bias tends to decrease with increasing k, to become negligible near $k = 0.02\,(\mathrm{km/s})^{-1}$. It is modeled by a third-degree polynomial (except for the $4.3 < z < 4.5$ redshift bin where a fourth-degree polynomial is used) that is used to correct the measured power spectrum. Again, we assign a systematic uncertainty on the resulting power spectrum equal to the 30% of the correction. As illustrated in figure 4.16, the systematic uncertainty is larger at small k, but it remains sub-dominant compared to the statistical uncertainty for all modes.

Table 4.2 summarizes the sources of bias identified in both analysis methods. The final power spectra are corrected for these under or over estimations. As explained above, we infer k and z systematic uncertainty dependences associated with these corrections. Their values are given along with the power spectrum measurements [3].

3. `http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/559/A85`

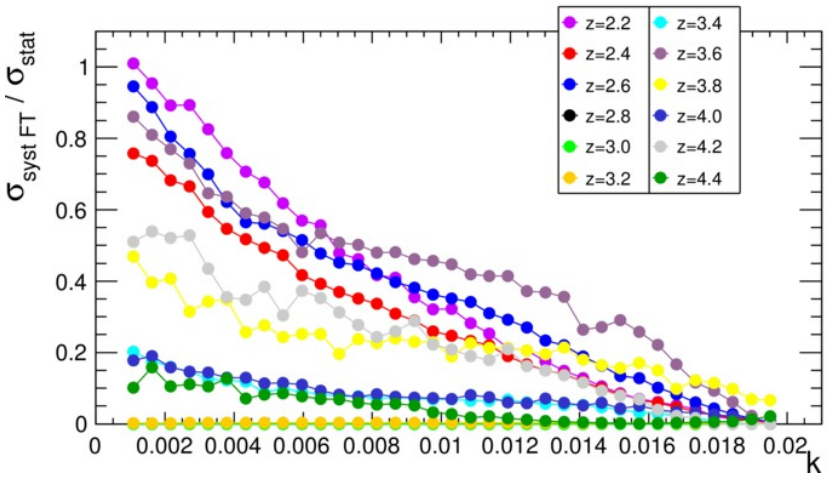


Figure 4.16 – Systematic uncertainty related to the masking of the sky lines in the Fourier transform approach, relative to the statistical uncertainty, for each redshift bin.

	Fourier transform	Likelihood
QSO continuum	1.02	1.02
Spectrograph resolution	–	–
Noise in the data[a].	0.98	1.00 to 1.13
Masking of sky lines . . .	0.82 to 1.00	1.00

[a] The noise-related bias was measured in the Fourier transform using the true continuum and is to be added to the other biases; for the likelihood, it includes systematic effects from all steps.

Table 4.2 – Bias introduced at different steps of the analyses.

4.5.2 Instrumental Uncertainties and Associated Systematics

The two main sources of instrumental uncertainties are related to the estimate of the noise and the resolution. The techniques to correct these two effects are respectively described in section 4.2.2 and 4.2.3. Here we present the associated systematics.

The power spectrum of the noise is obtained by computing the Fourier transform of a "difference spectrum" between the individual exposures of a single quasar. In a similar way to what was done for figure 4.1, we compare the side-band measurement of the noise (estimated as the flux rms in the $1330\,\text{Å} < \lambda_{\text{RF}} < 1380\,\text{Å}$ side-band of a quasar) either to the pipeline noise or to our determination of the noise from the difference power spectrum. Using the distribution of quasar redshifts, we show these distributions as a function of wavelength on the left plot of figure 4.17: the red curve shows the ratio of the average pipeline noise over the side-band noise; the green curve is the ratio of our estimate of the pixel noise, using the correction factor given in equation 4.2, over the side-band noise. After correction, the distribution is flat in wavelength and centered on 1.0, as expected. The right plot of figure 4.17 shows the distribution of the green points shown in the left plot of the figure. Its spread provides an estimate of the remaining uncertainty on the noise. From a Gaussian fit as shown on right plot of figure 4.17, we assign a conservative $\sim 1.5\%$ systematic error on the noise estimate.

We applied a similar method to derive the systematic error related to the resolution. In this case, we plotted the ratio of our resolution measurement to the resolution given by the pipeline

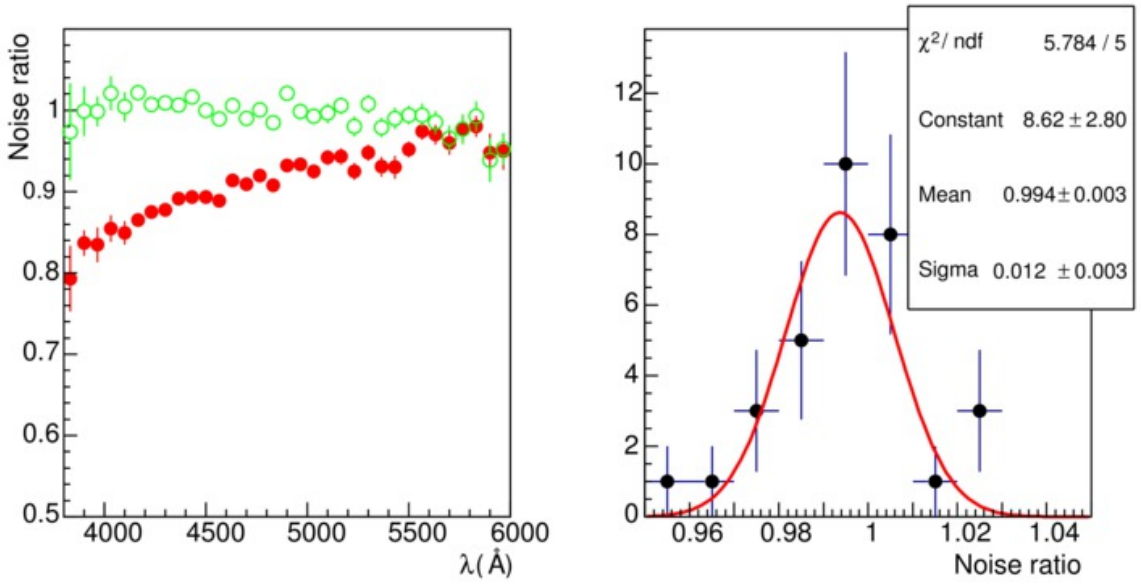


Figure 4.17 – Left: ratio of pipeline to side-band estimates of the noise as a function of the wavelength without correction (red dots) and when including the correction of equation 4.2 (green circles). Right: distribution of the residuals of the noise ratio with including noise correction; the rms of the green distribution, $\sim 1.5\%$, gives an estimate of the uncertainty on the noise correction.

(red) as a function of the fibre number (see the left plot of figure 4.18). In green, the pipeline resolution is corrected by our model of section 4.2.3. The rms of the residual distribution (in green) with respect to 1.0 yields a value of about 3% for the systematic error on the spectrograph resolution.

We determine the final impact of each of these two systematic effects using the data. We increase, for instance, our estimate of the noise for all the quasar spectra selected for the data analysis by the observed dispersion of 1.5%. We then apply the full procedure to measure the 1D power spectrum $P(k, z)$ with this new estimate of the noise. Finally, for each bin, we compare the new power spectrum, $P^{\rm new}(k, z)$, to the nominal power spectrum $P^{\rm init}(k, z)$. We define the systematic error to be $\sigma_P^{\rm sys}(k, z) \equiv 30\% \times |P^{\rm new}(k, z) - P^{\rm init}(k, z)|$. This is a conservative approach since we here consider a systematic effect acting in the same direction for all the quasars. The impact on the power spectrum of these systematic uncertainties are illustrated in figure 4.19. The largest systematic on the noise estimate is for low-redshift bins. However, with the cut we have applied on the spectrum signal-to-noise ratio, its contribution is at most of 70% $\sigma_{\rm stat}$. The systematic on the resolution estimate becomes dominant, over all other sources of uncertainties, at small scales for $z < 3.0$. The stringent cut we have applied on the mean resolution in the forest ($\bar{R} < 85\,{\rm km\,s^{-1}}$), however, has reduced it by almost a factor 5 compared to its value in the absence of such a cut.

4.6 Power Spectrum Measurement

We apply the methods presented previously to the BOSS data and measure the flux power spectrum in the Lyman-α forest region. It contains two components: the signal, arising from H I absorption, and the background due to absorption by all species other than atomic hydrogen

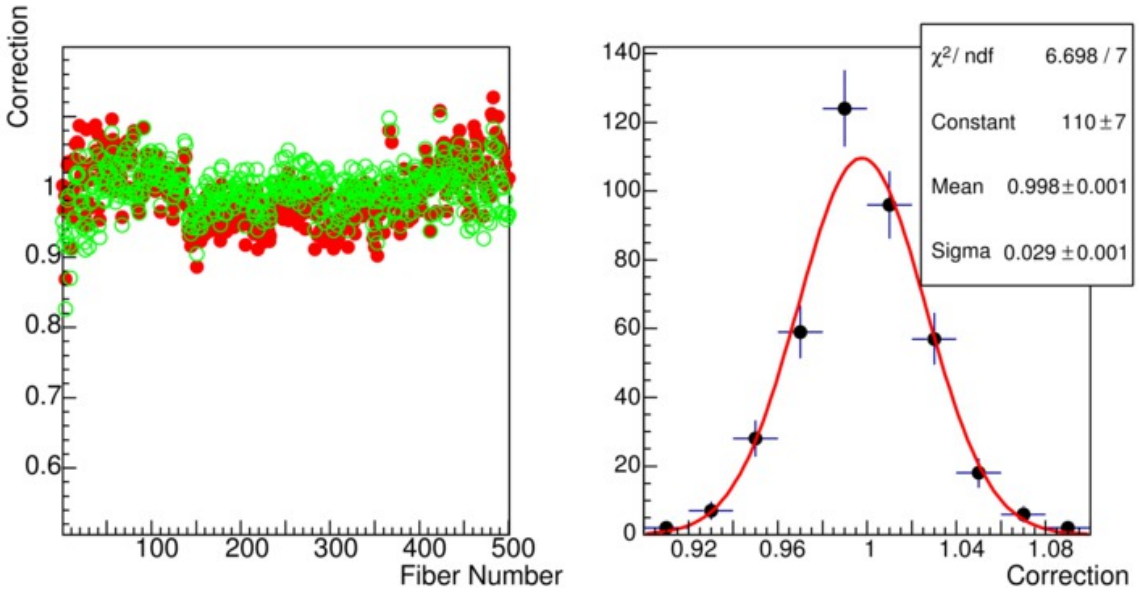


Figure 4.18 – Left: discrepancy between pipeline and arc lamp resolution as a function of the fibre number before correction (red dots) and after correction of section 4.2.3 (green circles). This plot is obtained for the Cd line at ~ 4800 Å. Right: distribution of the residuals of the resolution correction; the rms of the green distribution, $\sim 3.0\%$, provides an estimate of the uncertainty on the resolution correction.

(hereafter "metals"). In this section, we explain how we separate each contribution and we conclude by summarizing the obtained results.

Absorption at an observed wavelength λ receives contributions from any atomic species, i, absorbing at wavelength λ_i, if the absorption redshift $1 + z_i = \lambda/\lambda_i$ satisfies $z_i < z_{\rm qso}$. We want to subtract the background from metals. To do this, we use two methods that work respectively for species with $\lambda_i > \lambda_{{\rm Ly}\alpha}$ and $\lambda_i \sim \lambda_{{\rm Ly}\alpha}$.

For the first case, the wavelength of the metal line is far from Lyman-α. If its absorption falls in the Lyman-α forest of a quasar, then the Lyman-α absorption from the same redshift absorber is outside (further blue) the forest wavelength range. It therefore presents no correlation with the Lyman-α absorption. The summed absorption at λ due to all such species can be determined by studying absorption at λ in quasars with $1 + z_{\rm qso} < \lambda/\lambda_{{\rm Ly}\alpha}$ for which Lyman-α absorption makes no contribution. The subtraction of the background for this first case is described in section 4.6.1.

For the second case, atomic hydrogen and the metal species produce correlated absorption within the Lyman-α forest (Pieri et al., 2010). The 1D correlation function will have a peak at wavelength separations corresponding to hydrogen and metallic absorption at the same redshift: $\Delta\lambda/\lambda = 1 - \lambda_i/\lambda_{{\rm Ly}\alpha}$. The main contribution in this second case comes from Si III. The strategy adopted to subtract this second category of background is described in section 4.6.2.

In section 4.6.3, we present the final results in such a way that the reader can access directly the signal power spectrum and the different contaminating components.

4.6.1 Uncorrelated Background Subtraction

The uncorrelated background due to metal absorption in the Lyman-α forest cannot be estimated directly from the power spectrum measured in this region. We address this issue

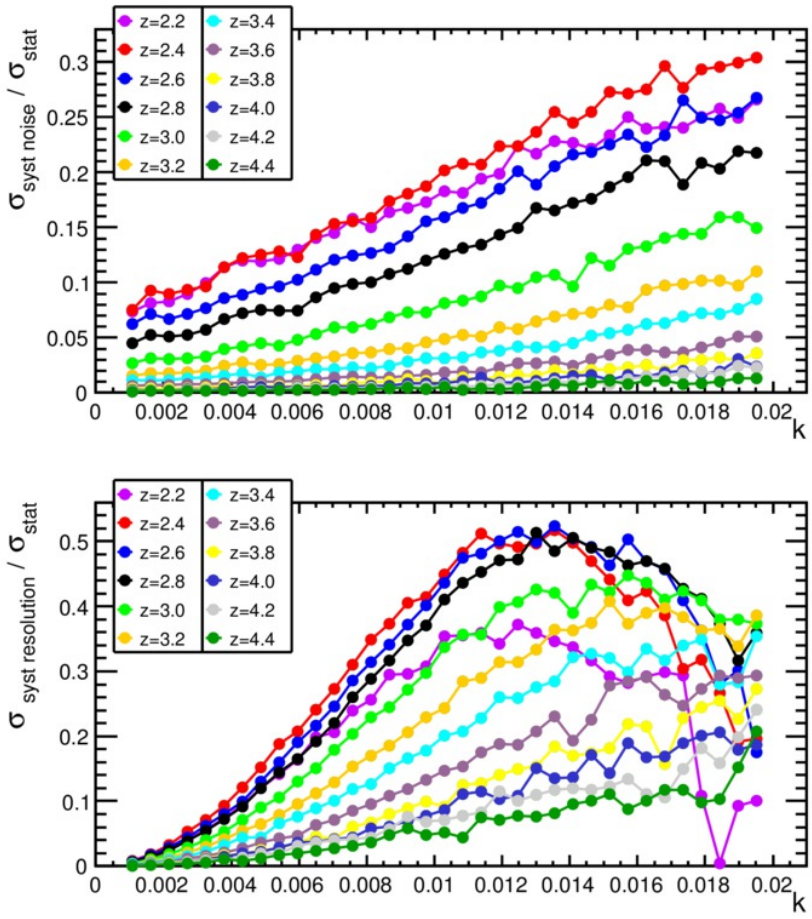


Figure 4.19 – Systematic uncertainties related to the estimate of the noise (upper plot) or of the spectrograph resolution (bottom plot), relative to the statistical uncertainty, for each redshift bin.

by estimating the background components in side bands located at longer wavelength than the Lyman-α forest region. We measure the power spectrum in these side bands and subtract it from the Lyman-α power spectrum measured in the same gas redshift range. This method is purely statistical; we use different quasars to compute the Lyman-α forest and the metal power spectra for a given redshift bin. This approach is inspired by the method described in McDonald et al. (2006). However, our approach is simpler and more robust because it relies only on control samples and does not require any modelling.

In practice, we define two side bands that correspond, in the quasar rest frame, to the wavelength ranges $1270\,\text{Å} < \lambda_{\text{RF}} < 1380\,\text{Å}$ and $1410\,\text{Å} < \lambda_{\text{RF}} < 1520\,\text{Å}$. The power spectrum measured in the first side band includes the contribution from all metals with $\lambda_{\text{RF}} > 1380\,\text{Å}$, including absorption from Si IV and C IV. The second side band also includes C IV but excludes the Si IV absorption. For our analysis, we use the first side band ($1270\,\text{Å} < \lambda_{\text{RF}} < 1380\,\text{Å}$) to subtract the metal contribution in the power spectrum, and measurement in the second side band constitutes an important consistency check.

We determine the metal power spectrum in the same observed wavelength range as the Lyman-α forest power spectrum from which it is being subtracted. For instance, for the first redshift bin, $2.1 < z < 2.3$, we measure the power spectrum in the first side band, corresponding to $3650\,\text{Å} < \lambda < 4011\,\text{Å}$, i.e., using quasars with a redshift $z \sim 1.9$. Quasars in a given redshift

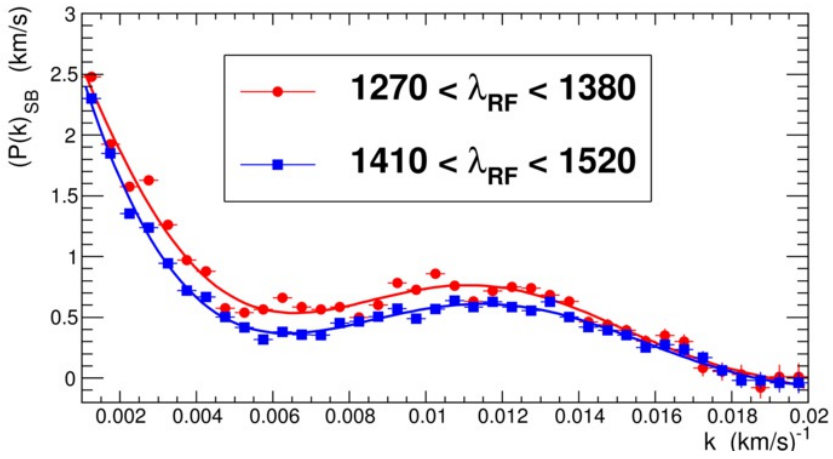


Figure 4.20 – Power spectrum $P_{\rm SB}(k)$ computed for side band regions above the Lyman-α forest region. The red dots and the blue squares are respectively for the two side bands defined in the rest frame by $1270\,\text{Å} < \lambda_{\rm RF} < 1380\,\text{Å}$ and $1410\,\text{Å} < \lambda_{\rm RF} < 1520\,\text{Å}$. Each power spectrum is fitted with a sixth-degree polynomial.

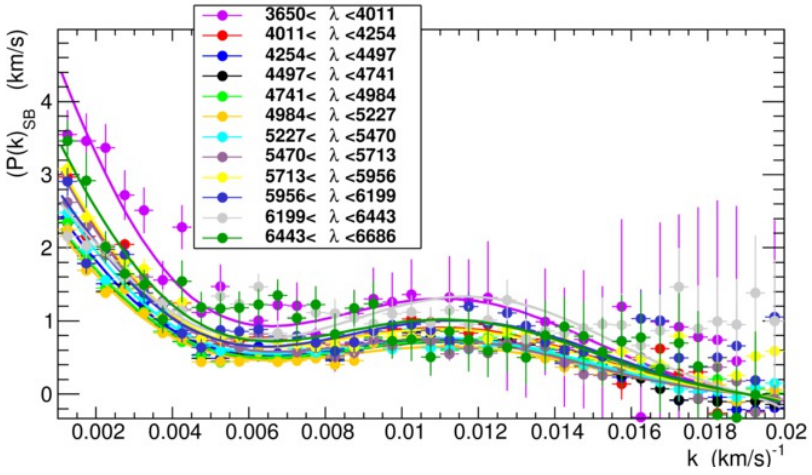
window have their two side-bands corresponding to fixed observed wavelength windows, which in turn match a specific redshift window of Lyman-α forest.

The power spectra $P_{\rm SB}(k)$ shown in figure 4.20 are obtained with $\sim 40,000$ quasars with redshifts in the range $1.7 < z < 4.0$, passing similar quality cuts as the quasars for the Lyman-α forest analysis. The shapes of $P_{\rm SB}(k)$ are similar for the two side bands. As expected, for the second side band, corresponding to $1410\,\text{Å} < \lambda_{\rm RF} < 1520\,\text{Å}$, which excludes Si IV, the amplitude of $P_{\rm SB}(k)$ is smaller. We fit the distribution $P_{\rm SB}(k)$ with a sixth-degree polynomial. We will use this fitted function as a template to parametrise the $P_{\rm SB}(k)$ measured for each wavelength window (see figure 4.21).

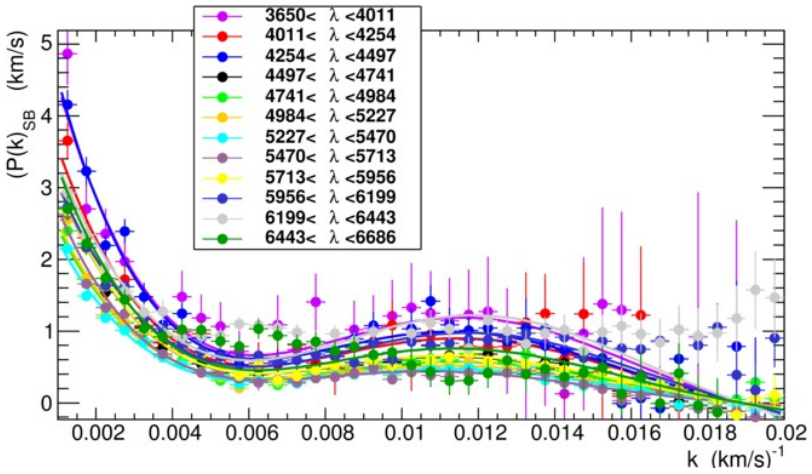
As the shape and the magnitude of the power spectrum vary from one wavelength window to another, we have parameterised this as the product of the fixed shape obtained in figure 4.20, with a variable first-degree polynomial, with two free parameters that are different for each wavelength window. This adequately fits the measured power in all the wavelength windows (see figure 4.21). From these two parametric functions, we extract the value of the power spectrum $P_{\rm SB}(k, z)$ for each k and for each Lyman-α redshift window.

The statistical uncertainty on $P_{\rm SB}$ is largest where we have the smallest number of quasars to measure the metal contribution. This occurs in the $z \sim 2.2$ redshift bin for which we only have about four hundred quasars (at $z_{\rm qso} \sim 1.7$) instead of about 4000 on average for the other bins. For $z \sim 2.2$, the uncertainty on the metal correction, derived from the statistical precision on the first-degree polynomial fit, is of the order of 10%.

An uncertainty on our metal correction will have the largest impact relative to the measured $P(k)$ in the Lyman-α region when the absolute $P(k)$ has the smallest value. This again occurs for $z \sim 2.2$, which therefore constitutes the worst case both in terms of statistical uncertainty and relative level of the correction. Even in this worst case, the metal power spectrum is less than ten percent of the Lyman-α power spectrum. The uncertainty of the metal correction is therefore less than one percent of the Lyman-α $P(k)$ across our whole sample.



(a) Power spectrum in the side band 1270 Å $< \lambda_{\rm RF} <$ 1380 Å



(b) Power spectrum in the side band 1410 Å $< \lambda_{\rm RF} <$ 1520 Å

Figure 4.21 – Power spectrum $P_{\rm SB}(k)$ computed for side band regions above the Lyman-α forest region for different λ windows. Each λ region corresponds to one redshift bin. Each power spectrum is fitted by the product of the sixth-degree polynomial obtained in figure 4.20 and a first-degree polynomial in which the two parameters are free.

4.6.2 Si III Cross-Correlation

The correlated background due to absorption by Lyman-α and Si III from the same gas cloud along the quasar line of sight can be estimated directly in the power spectrum. Since Si III absorbs at $\lambda = 1206.50$ Å, it appears in the data auto-correlation function $\xi_{\rm tot}(v) = \langle \delta(x)\delta(x+v) \rangle$ as a bump at $\Delta v = 2271\,{\rm km\,s^{-1}}$, and in the power spectrum as wiggles with peak separations of $\Delta k = 2\pi/\Delta v = 0.0028\,({\rm km/s})^{-1}$. Following the approach suggested by McDonald et al. (2006), we model the Si III structure as being equal to that of the Lyman-α forest up to an overall normalization: $\delta_{\rm tot} = \delta(v) + a \times \delta(v + \Delta v)$ where $\delta(v)$ is for Lyman-α only. The corresponding correlation function is

$$\xi_{\rm tot}(v) = (1 + a^2)\xi(v) + a\xi(v + \Delta v) + a\xi(v - \Delta v) \tag{4.17}$$

and the corresponding power spectrum

$$P_{\rm tot} = (1 + a^2)\, P(k) + 2\, a \cos(\Delta v\, k) P(k)\ , \tag{4.18}$$

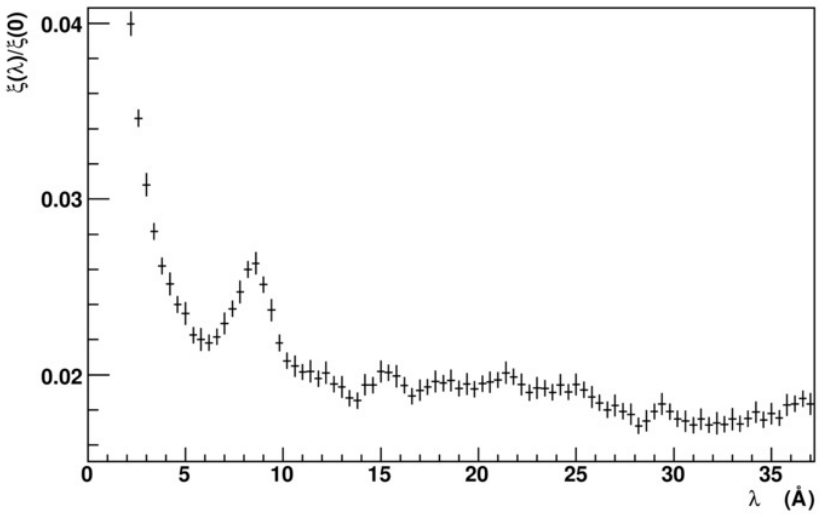


Figure 4.22 – Normalized correlation function $\xi_{\rm tot}(\lambda)/\xi_{\rm tot}(0)$ in the Lyman-α forest region. A clear peak is visible at $\Delta\lambda = 9.2\,\text{Å}$ corresponding to the Lyman-α–Si III cross correlation.

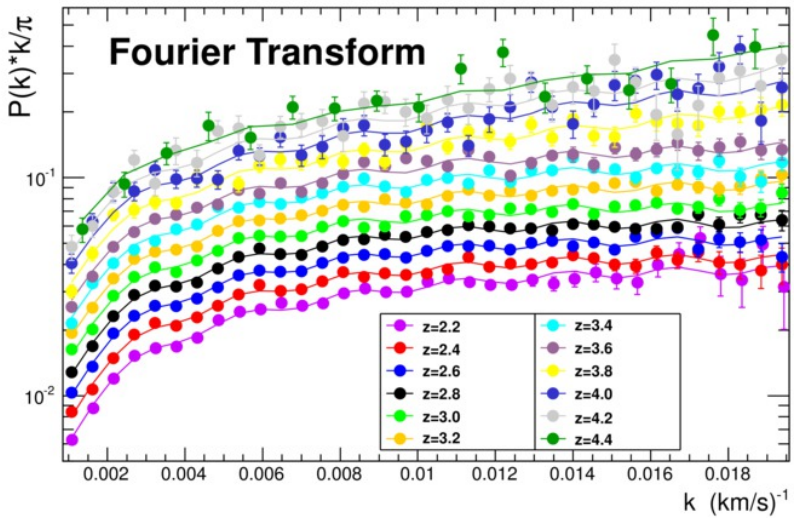
where $\xi(v)$ and $P(k)$ are for Lyman-α–Lyman-α correlations. We clearly detect, in the power spectrum, the oscillatory pattern due to the Si III–Lyman-α cross correlation (see figure 4.23), or equivalently a peak near $\Delta\lambda = 9.2\,\text{Å}$ in the correlation function. The correlation function is shown in figure 4.22. We do not observe any other significant metal features seen in Pieri et al. (2010), such as Si II lines (at 22.4 Å and 25.3 Å) or N V lines (at 23.2 Å and 27.1 Å). However, some weak contribution may be present from metals where they do not produce signals distinct from each other or from the greater Lyman-α signal.

The measured normalization evolves with redshift roughly as $a(z) = f_{\rm Si\,III}/(1-\bar{F}(z))$, where $\bar{F}(z)$ is the mean transmitted flux fraction defined in section 4.3.3. With a simple fit, we find a normalization factor $f_{\rm Si\,III} = 0.008 \pm 0.001$, similar to the value $f \sim 0.011$ measured by McDonald et al. (2006) on a sample of three thousand SDSS quasars.

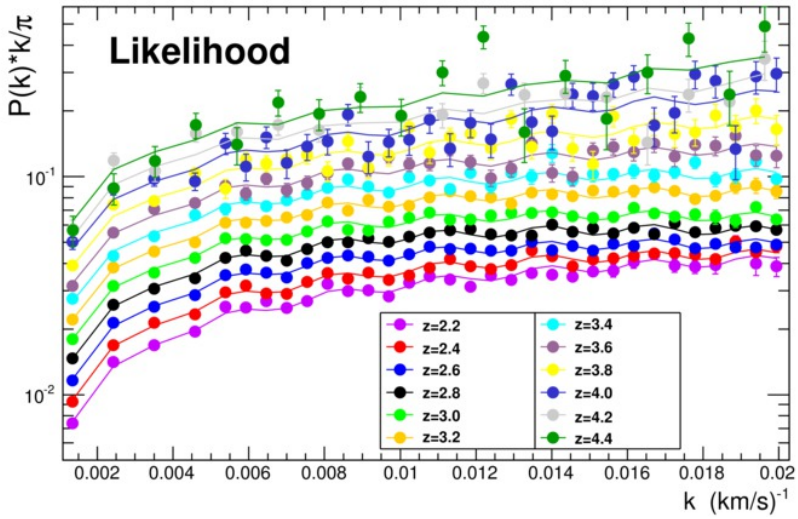
4.6.3 Summary of Experimental Results

Figure 4.23 shows the one-dimensional Lyman-α forest power spectrum obtained with the Fourier transform and the likelihood method. Figure 4.24 demonstrates a good agreement between the methods, although they are quite different in the treatment of the sky line masking, the noise subtraction and the resolution correction. Moreover, the agreement with the previous SDSS measurements (see McDonald et al. (2006)) is also remarkable. The only significant discrepancy between SDSS and BOSS is observed for the low z and high k region where the noise subtraction is difficult since the noise level is similar to the signal level. The uncertainty in this region is covered by the use of the systematic errors given in section 4.5 or by using nuisance parameters.

In order to compare the measured power spectrum for SDSS and BOSS, and also to compare the results of the Fourier transform and the likelihood methods in a quantitative way, we define an empirical function $P^{\rm emp}$ with which we fit each power spectrum distribution. This function, written in equation 4.19, has five free physical parameters: an amplitude A_F corresponding to the amplitude of the power spectrum at the pivot mode k_0 and pivot redshift z_0, a slope



(a) Lyman-α forest 1D power spectrum obtained with the Fourier transfrom method.



(b) Lyman-α forest 1D power spectrum obtained with the likelihood method.

Figure 4.23 – One-dimensional Lyman-α forest power spectrum obtained with the Fourier transform method and the likelihood method. The metal contribution estimated in section 4.6.1, is subtracted. The power spectrum is fitted with the empirical function of equation 4.19.

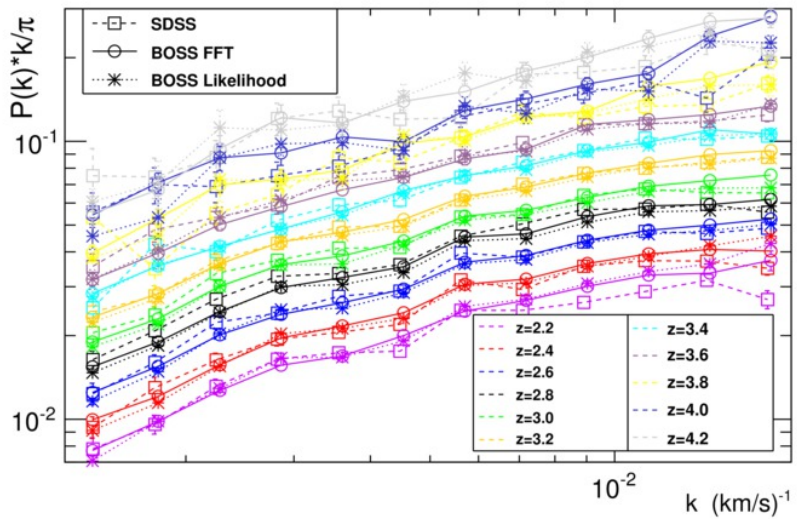


Figure 4.24 – Comparison of the 1D Lyman-α forest power spectrum obtained in BOSS and in SDSS (McDonald et al., 2006) over the redshift range, $z = 2.1{-}4.3$. For BOSS, we show the results for the two methods, Fourier transform and likelihood, and we use the same k bins as in McDonald et al. (2006).

Parameter	SDSS	BOSS Fourier transform	BOSS likelihood
A_F	0.062 ± 0.002	0.067 ± 0.001	0.064 ± 0.001
n_F	-2.64 ± 0.04	-2.50 ± 0.02	-2.55 ± 0.02
α_F	-0.13 ± 0.02	-0.08 ± 0.01	-0.10 ± 0.01
B_F	3.30 ± 0.14	3.36 ± 0.06	3.55 ± 0.07
β_F	-0.28 ± 0.09	-0.29 ± 0.04	-0.28 ± 0.05

Table 4.3 – Results of the fit by the empirical function $P^{\rm emp}(k,z)$ (see definition in equation 4.19) of the SDSS and BOSS datasets over the redshift range, $z = 2.1{-}4.3$. These five parameters should not be used for any quantitative science as the χ^2 remain ~ 1.4 even after adding nuisance parameters in the fit.

$n_F = \mathrm{d}\ln P/\mathrm{d}\ln k|_{(k_0,z_0)}$, a curvature $\alpha_F = \mathrm{d}\ln n_F/\mathrm{d}\ln k|_{(k_0,z_0)}$, and two parameters, B_F and β_F, that model the redshift evolution of the power spectrum. In addition, we introduce nuisance parameters to take into account the correlation between H I and Si III (parameter a), and the imperfection of our resolution and noise models. We choose a pivot point in the middle of our measurements, $k_0 = 0.009\,(\mathrm{km/s})^{-1}$ and $z_0 = 3.0$. The results of the fits are summarized in table 4.3. The agreement between the different methods and datasets is good. All five parameters are within 1 or 2 σ of one another.

$$\begin{aligned}\frac{k}{\pi} \times P^{\rm emp}(k,z) = A_F \times &\left(\frac{k}{k_0}\right)^{3+n_F+\alpha_F \ln\left(\frac{k}{k_0}\right)+\beta_F \ln\left(\frac{1+z}{1+z_0}\right)} \\ &\times \left(\frac{1+z}{1+z_0}\right)^{B_F} \\ &\times \Big(1 + a^2 + 2a\cos(\Delta v\, k)\Big)\end{aligned} \tag{4.19}$$

The full results are available at the *Centre de Données astronomiques de Strasbourg (CDS)*[4]. Two tables summarize, for each redshift bin, the results for the one-dimensional Lyman-α forest power spectrum. The different components (P_{1D}, $P^{\rm noise}$ and $P^{\rm metals}$) are given in these tables. $P^{\rm metals}$ only takes into account the uncorrelated background computed in section 4.6.1. Two other columns represent the statistical and systematical uncertainty on P_{1D}. We added in quadrature all the systematic uncertainties studied in section 4.5.The correlation matrices are illustrated in figure 4.25 for the first eight redshift bins. The maximum correlation is about twenty percent for neighbouring $k-$modes, and the correlation rapidly drops under ten percent.

4.7 Conclusions

We have developed two independent methods to measure the one-dimensional power spectrum of the transmitted flux in the Lyman-α forest. The first method is based on a Fourier transform, and I have developed a second approach which relies upon a maximum likelihood estimator. The two methods are independent and present different systematic uncertainties due to the techniques used to mask pixels contaminated by sky emission lines, or to take into account the spectrograph resolution and the noise contribution to the Lyman-α power spectrum, which differ in the two approaches. The determination of the noise level in the data spectra was subject to a novel treatment, because of its significant impact on the derived power spectrum.

We applied these two methods to 13 821 quasar spectra from SDSS-III/BOSS selected from a larger sample of almost 90 000 DR9 BOSS spectra on the basis of their high quality, large signal to noise ratio, and small value of the spectral resolution. The power spectra measured using either method are in good agreement over all twelve redshift bins from $\langle z\rangle = 2.2$ to $\langle z\rangle = 4.4$. We determined the systematic uncertainties on our measurements coming both from the analysis method and from our knowledge of the instrument characteristics.

The logical next step is to use these results to extract cosmological constraints, with a study similar to Viel et al. (2006). The improvement in precision with respect to previous studies on SDSS data (McDonald et al., 2006) should allow constraints tighter by a factor 2 to 3. This is done thanks to a new set of hydrodynamical simulations for the Lyman-α forest that we ran with an upgrade in both resolution and box size to match the sensitivity of our measurement and which, in addition, includes massive neutrinos. These simulations are presented in chapter 6 after a short introduction on cosmological simulations in chapter 5.

4. `http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/559/A85`

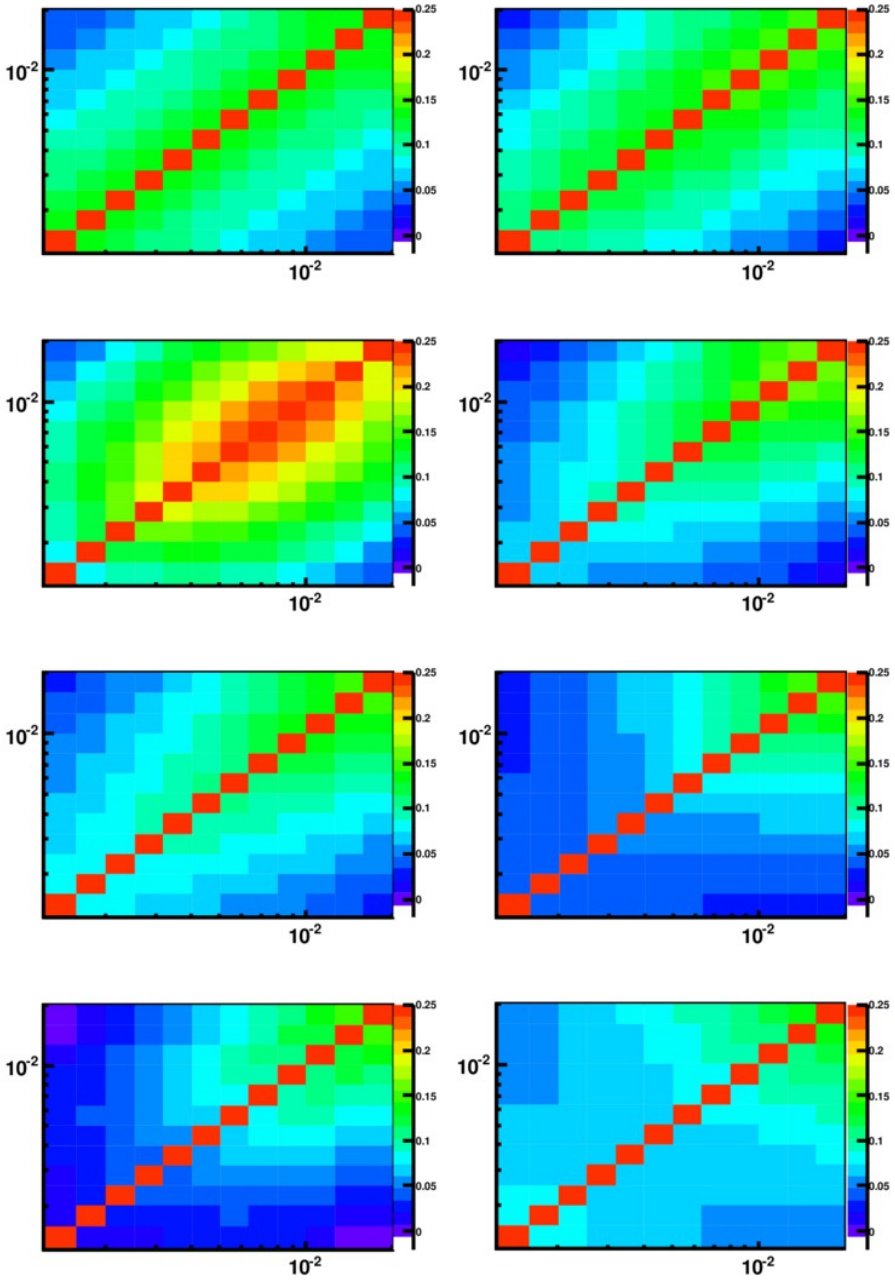


Figure 4.25 – Correlation matrices between the different k-modes for the first 8 redshift bins, with $z = 2.1-4.3$, smoothed by 2D second degree polynomials. The color range is identical in all 8 plots, with red for all values above 0.25.

Chapter 5

Simulating the universe

"Never send a human to do a machine's job."
—Matrix, 1999

COMPUTATIONAL cosmology is the field of science that studies the modelling of the Universe and structure formation using numerical simulations. Such simulations can be seen as the only "experiment" –as opposed to observation– to check theories about the evolution of the Universe or its composition. Thanks to computer hardware and simulation code improvement in the last decades, it is now possible to model in detail volumes as large as the observable universe in the *Dark Energy Universe Simulation Full Universe Run* (DEUS FUR) (Alimi et al., 2012). Section 5.2 is summarized from Booth (2007). Other sections take inspiration mainly from Booth (2007) and Dolag et al. (2008).

The goal of this chapter is to give the reader an insight into what are the main techniques to simulate an evolving universe. After a brief history of cosmological simulations, I will present the main algorithms used to make the gravity calculations. I will then present the basics of the smoothed particle hydrodynamics methods for evolving a gas. This method was used in the simulation we did to evolve baryonic gas (see chapter 6). Finally, I will say a few words about pre-initial particles distributions and initial conditions generations. I will not present the details of the calculations and algorithms as they are not necessary to understand the next chapters.

5.1 History of Simulations

Structure formation is believed to be mainly (but not only) driven by gravitational collapse of the primordial fluctuations (see chapter 2). Up to some point, the evolution of these fluctuations can be studied using a linear perturbation theory (Zeldovich, 1970). However, when the overdensity $\delta = \rho(\mathbf{r})/\langle\rho\rangle - 1$ becomes close to unity, these calculations are no longer accurate. As today's structures correspond to δ values ranging from $\delta \simeq -1$ in voids to $\delta \simeq 10^6$ in the galaxies' central regions, linear calculus cannot explain the current observational data. Computing the evolution of perturbations in a non-linear regime is very difficult and cannot be done analytically by hand. Thus the only accurate method is to use numerical simulations.

Nowadays, these simulations are used in every field of cosmology and astrophysics and they are so sophisticated that they have a research branch of their own: *computational cosmology*.

To follow the evolution of primordial fluctuations up to actual structures like stars, galaxies, clusters, ... there are in fact two steps. The first one is to generate initial conditions according to the chosen structure formation model (see section 5.4), and the second is to evolve these initial conditions forward in time using a numerical integrator for the equations of interest (both gravity and hydrodynamics if required). The initial density field is sampled by particles or mesh and therefore the codes that evolve such fields are referred to as N-body codes.

The current generation of cosmological simulations has ancestors that date back to several decades. The first gravitational N-body simulation of interacting galaxies was performed using an analogue optical computer (Holmberg, 1941): gravity was represented by the flux from 37 lightbulbs, with photocells and galvanometers used to measure and display the inverse square law force. The first astronomical N-body computations using digital computers were made in the early 1960s (von Hoerner, 1960, 1963; Aarseth, 1963). However, these early simulations were limited to at most about a hundred particles. The first truly cosmological simulations of structure formation were the N-body integrations of Press and Schechter (1974) to study hierarchical clustering.

Since then algorithms have advanced considerably and a lot of different solvers for the Poisson equation were developed:

— direct summation of the pairwise forces (Aarseth et al., 1979; Frenk et al., 1983);
— fast Fourier transforms on grids: PM (particle mesh) methods (Efstathiou and Eastwood, 1981; Klypin and Shandarin, 1983; White et al., 1983);
— hierarchical trees (Barnes and Hut, 1986; Jernigan and Porter, 1989);
— mesh relaxations methods (Brandt, 1977);
— hybrid methods like P^3M (Efstathiou and Eastwood, 1981; Couchman et al., 1995; Wadsley and Bond, 1996) or Tree-PM (Wadsley et al., 2004; Springel, 2005).

At the same time algorithms to evaluate hydrodynamics forces were also developed like the smoothed particle hydrodynamics (SPH) method (Gingold and Monaghan, 1977; Monaghan, 1992) or the piecewise parabolic method (PPM) (Colella and Woodward, 1984; Woodward and Colella, 1984).

The last decade has seen an impressive growth not only in the size of cosmological simulations: simulations of Gpc^3 with billions of particles are now almost common, but also in the sophistication of the physical process involved.

5.2 Gravity Calculations

For a density field $\rho(\mathbf{r})$, the associated gravitational potential $\Phi(\mathbf{r})$ is given by the Poisson equation:

$$\nabla^2\Phi(\mathbf{r}) = 4\pi G\rho(r), \tag{5.1}$$

where G is the gravitational constant. In a cosmological comoving frame this equation is written (Peebles, 1980):

$$\nabla^2\Phi(\mathbf{r}) = 4\pi G a^2 \left[\rho(r) - \langle\rho\rangle\right]. \tag{5.2}$$

The obtained potential can then be used to compute the gravitational acceleration at any point. However, continuous fields cannot be represented in a computer and one has in addition to discretize the phase-space distribution of matter (position and velocity) with enough discrete

points to get the important features. I will here briefly present four different methods to calculate the gravitational force: direct summations, Fourier transforms on a regular grid, hierarchical trees and iterations on a (potentially irregular) grid. Eventually, I will compare these techniques and introduce hybrid methods that are used nowadays.

5.2.1 Particle Particle Methods

The simplest way to calculate the gravitational force on a point mass due to other point particles is to use the direct summation or particle particle (PP) method. It explicitly sums the contribution of each particle to the gravitational potential:

$$\Phi(\mathbf{r}) = -G \sum_i \frac{m_i}{\sqrt{|\mathbf{r} - \mathbf{r}_i|^2 + \epsilon^2}}, \tag{5.3}$$

where ϵ is a smoothing term which avoids divergence of the gravitational force at small radii ($r < \epsilon$). It is introduced to avoid the appearance of arbitrarily large velocity when two particles are very close. In addition one must not forget that the simulation particles do not represent actual particles or stars but rather a group of particles. PP codes are exact up to the numerical precision of the computer and they were used in the very first simulations (Aarseth, 1963; Hénon, 1964). However, such codes are very computationally intensive as they scale as $\mathcal{O}(N_p^2)$ where N_p is the number of considered particles. Even with recent computers they cannot be used for large simulations.

5.2.2 Particle Mesh Methods

To lower the number of required operations, one can evaluate the gravitational potential on a regular grid and interpolate the forces on the particles. The process is made of three different steps:

1. mapping the discrete particles onto a uniform mesh,
2. solving the Poisson equation to obtain the gravitational potential on the mesh,
3. differentiating the potential at the location of each particle to obtain the gravitational force.

We here consider a cubic volume of side length L, with N_p particles of masses m_i. It is thus possible to write the mass distribution at any point $\mathbf{r}$ as

$$m(\mathbf{r}) = \sum_{i=0}^{N-1} m_i \delta(\mathbf{r} - \mathbf{r}_i) \tag{5.4}$$

Then using a cubic grid of spacing Δ, it is possible to compute the mass density at any point $\mathbf{r}_g$ on the grid through

$$\rho(\mathbf{r}_g) = \frac{1}{\Delta^3} \sum_{i=0}^{N-1} m_i W\left(|\mathbf{r}_g - \mathbf{r}_i|\right) \tag{5.5}$$

where W is a function representing the assignment scheme of the discrete mass distribution on the grid. Several schemes can be used such as the nearest grid point (NGP), where all the particle mass is assigned to the closest grid cell; the cloud in cell (CIC) scheme, where the mass is distributed over the nearest two cells in each direction or the even more complicated

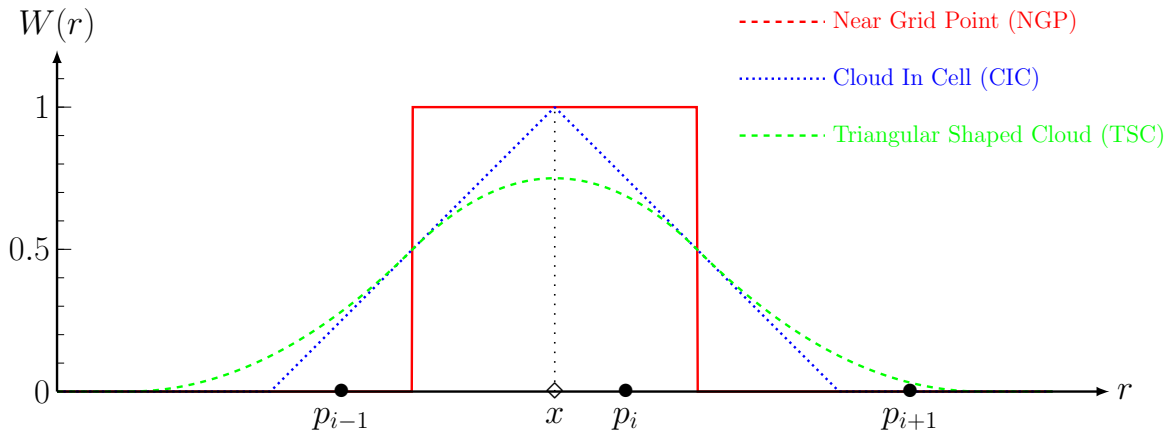


Figure 5.1 – Three mass assignement schemes for a particle located at point x with closest grid points p_{i-1}, p_i and p_{i+1}. The NGP method assigns all the mass to p_i, the CIC method assigns the mass to the nearest points p_i and p_{i-1} and the TSC method distributes the mass over the three points.

triangular shaped cloud scheme (TSC). A one-dimensional illustration of these schemes is given in figure 5.1 and the corresponding functions are

$$W_{\mathrm{NGP}}(\mathbf{r}) = \begin{cases} 1 & \text{if } |\mathbf{r}| < \frac{\Delta}{2} \\ 0 & \text{otherwise} \end{cases}; \tag{5.6}$$

$$W_{\mathrm{CIC}}(\mathbf{r}) = \begin{cases} 1 - \frac{|\mathbf{r}|}{\Delta} & \text{if } |\mathbf{r}| < \Delta \\ 0 & \text{otherwise} \end{cases}; \tag{5.7}$$

$$W_{\mathrm{TSC}}(\mathbf{r}) = \begin{cases} \frac{3}{4} - \left(\frac{|\mathbf{r}|}{\Delta}\right)^2 & \text{if } |\mathbf{r}| \leq \frac{\Delta}{2} \\ \frac{1}{2}\left(\frac{3}{2} - \frac{|\mathbf{r}|}{\Delta}\right)^2 & \text{if } \frac{\Delta}{2} < |\mathbf{r}| \leq \frac{3\Delta}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{5.8}$$

Choosing one of these schemes depends upon several factors: higher order schemes are more computationally expensive but they reduce the amount of noise introduced in the resulting solution. Of course, a lower order scheme can be partially compensated with a tighter grid.

Once the density field is computed, the Poisson equation is solved on the grid in Fourier space. It is the most computationally efficient method thanks to the existing methods to perform fast Fourier transform and the linearity of the equation. Details of the calculation of the potential using a Green's function can be found for example in Dolag et al. (2008).

After the calculation of the potential, the force field can be computed on the grid points by differentiating the potential since $\mathbf{F}(\mathbf{r}) = -\nabla\Phi(\mathbf{r})$. This is usually done by approximating the partial derivative with a finite difference scheme, for example with a central difference:

$$\left.\frac{\partial\Phi}{\partial x}\right|_{i,j,k} = \frac{\Phi_{i+1,j,k} - \Phi_{i-1,j,k}}{2\Delta} + \mathcal{O}(\Delta^2) \tag{5.9}$$

The forces are then assigned back to the particles using the same scheme employed for the density field construction to avoid annoying self-forces.

It is worth noting that although this method is quite efficient as it scales as $\mathcal{O}(N_p + N_c \log(N_c))$ where N_p is the number of particles and N_c the number of grid cells, its force resolution is limited

by the size of the mesh, especially when several particles fall into the same cell. To increase the resolution, additional methods are required.

5.2.3 Tree Methods

Instead of arranging space in a regular grid, one can also arrange particles into a hierachy of groups, i.e. a tree. With respect to particle particle methods, the computation of the forces is more efficient as distant particles in the tree are treated as groups. There exists a multitude of different ways to organize an ensemble of particles in a volume into a tree, such as the Barnes and Hut octree (Barnes and Hut, 1986) or trees based on nearest neighbour pairing (Jernigan and Porter, 1989).

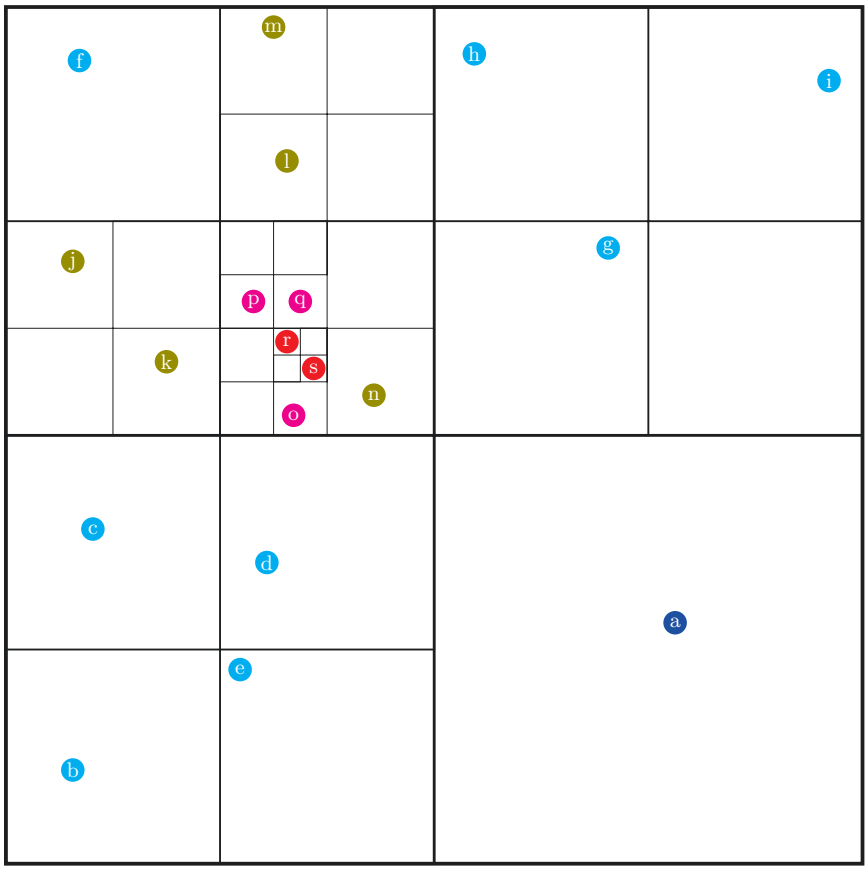
A Jernigan-Porter tree is constructed as follows: the two closest particles are joined together to form a node to which is associated the centre of mass of the two particles. Using this node as a new particle, the process is iterared until only one centre of mass remains. The construction of the Barnes-Hut tree is illustrated in two dimensions in figure 5.2. In three dimensions, the volume is recursively divided into eight equal cubes (thus the name octree): if a cube contains more than one particle, the process is repeated; if it contains exactly one particle, it becomes a leaf of the tree; if it has no particle, it is dropped.

It is worth noting that the choice of a type of tree only affects storage requirements and the efficiency of the gravity forces calculations, the procedure used to calculate the force is unchanged. To compute the force that applies to a single particle, we walk down the tree starting from the root using an opening criterion like $r > l/\epsilon$ where r is the distance between the considered particle and the current node, l is the spatial size of the node and ϵ is an accuracy parameter. If the criterion is satisfied, it means that the node is far enough from the particle so that the force coming from this node is added to the total force acting on the particle. If the criterion is unsatisfied, the node is opened and the criterion is applied to its children. The process is repeated until every particle is contributing to the force, directly or through a node.

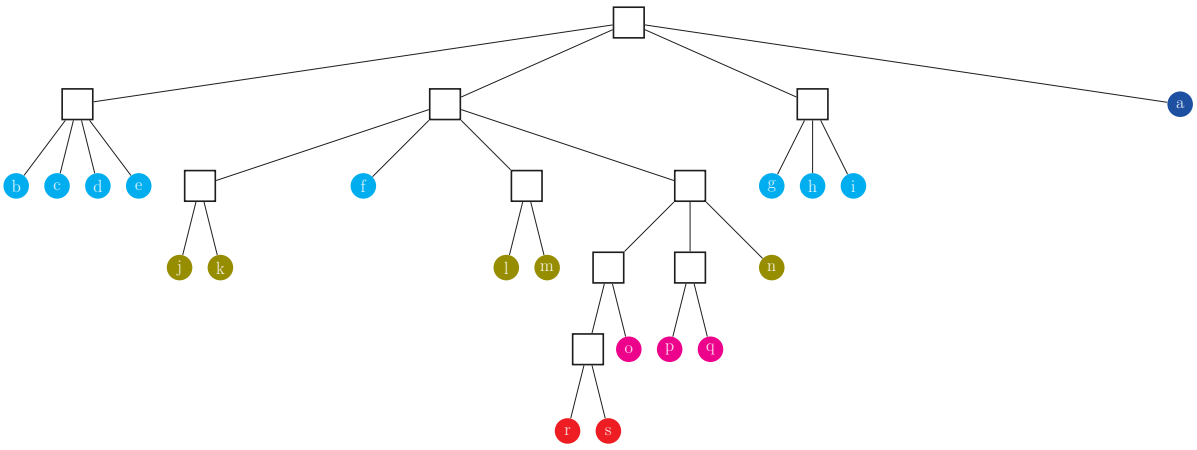
The force generated by a node can be evaluated at first order using the centre of mass or at higher order using multipole expansion. This of course produces only an approximation of the true force coming from the particles included in the node. However, the error can be made arbitrarily small by tuning the accuracy parameter ϵ, which controls how deep we go down the tree when doing the calculation and by increasing the order to which the forces are calculated for each node. To achieve maximum efficiency two approaches are used: low order expansion of the force but a deep walk in the tree (Springel, 2005) or a high order expansion but a shallow walk in the tree (Wadsley et al., 2004).

5.2.4 Mesh Relaxation Methods

Instead of using Fourier transform to solve the Poisson equation on a grid, one can also use iterative methods. This is especially interesting when using a non-uniform grid, as with adaptative mesh refinement (AMR) as computing the Fourier transform on such a grid is much slower than on a uniform grid. Solving the Poisson equation starts and ends with the same mapping and reverse mapping between the particles and the grid as in the particle mesh methods. The difference comes from the method used to solve the Poisson equation on a grid.



(a) Recursive division of space to construct the quadtree: starting with the whole space representing the root node of the tree, space is recursively divided in four equal pieces if it contains more than one particle.



(b) The tree resulting from the above division of space: each square containing more than one particle is a node and each square with exactly one particle is a leaf.

Figure 5.2 – Two-dimension version of the Barnes-Hut octree, the quadtree. Each circle represents a single particle, the colours represent the different levels of the tree.

On a non-uniform grid, it is possible to approximate a second derivative of the gravitational field with

$$\frac{\partial^2\Phi}{\partial x^2} \approx \frac{2\Phi_{i-1,j}}{\Delta x_i(\Delta x_i + \Delta x_{i+1})} - \frac{2\Phi_{i,j}}{\Delta x_i \Delta x_{i+1}} + \frac{2\Phi_{i+1,j}}{\Delta x_{i+1}(\Delta x_i + \Delta x_{i+1})} \tag{5.10}$$

where $\Delta x_i = |x_i - x_{i-1}|$. This can be used to approximate the Laplacian (here in two dimensions):

$$\begin{aligned}
\nabla^2\Phi_{i,j} &= \left.\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2}\right|_{i,j} \\
&\approx \frac{2\Phi_{i-1,j}}{\Delta x_i(\Delta x_i + \Delta x_{i+1})} - \frac{2\Phi_{i,j}}{\Delta x_i \Delta x_{i+1}} + \frac{2\Phi_{i+1,j}}{\Delta x_{i+1}(\Delta x_i + \Delta x_{i+1})} \\
&\qquad + \frac{2\Phi_{i,j-1}}{\Delta y_i(\Delta y_i + \Delta y_{i+1})} - \frac{2\Phi_{i,j}}{\Delta y_i \Delta y_{i+1}} + \frac{2\Phi_{i,j+1}}{\Delta y_{i+1}(\Delta y_i + \Delta y_{i+1})} \\
&= \rho_{i,j}.
\end{aligned} \tag{5.11}$$

Then, assuming a uniform grid with spacing h for simplicity, the Poisson equation can be rewritten

$$\Phi_{i,j} \approx \frac{1}{4}\left(\Phi_{i-1,j} + \Phi_{i+1,j} + \Phi_{i,j-1} + \Phi_{i,j+1} + h^2\rho_{i,j}\right). \tag{5.12}$$

This equation can be solved iteratively using the Jacobi method: starting from a guess of the potential, usually the potential from the precedent timestep, the equation is used to compute an improved estimate of the potential until convergence. The Jacobi method can be improved using the Gauss-Seidel iteration where the already updated values are used in the equation, it was shown to greatly improve the speed of the convergence (Hockney and Eastwood, 1988), thus reducing the number of iterations and saving computational time. However, the Gauss-Seidl scheme makes the iterations difficult to parallelize as each calculation depends on the previous ones. The solution comes from the "red-black" ordering where all the red (respectively black) cells can be computed independently allowing them to be distributed on different processors. This method is illustrated in figure 5.3.

The main disadvantage of the relaxation technique is that since each point interacts only with its closest neighbours, the propagation of a perturbation is slow and requires many iterations to be echoed on the whole grid. This problem can however be solved by using hybrid techniques as we will see in the next section.

5.2.5 Comparison of Methods and Hybrid Methods

After describing the methods in the previous section, I will now focus on their advantages and disadvantages as well as introduce some hybrid techniques, which combine the best features of two or more methods.

Particle Particle (PP) Methods: The main advantages of the particle-particle methods are that they are the easiest to understand and they are accurate up to the machine precision. However, they have a big drawback: the number of calculations scales as $\mathcal{O}(N_p^2)$ where N_p is the number of particles involved. Even with specialized hardware like GRAPE (Sugimoto et al., 1990), codes based on the direct summation of the forces are limited to small simulations (on a desktop computer, only a few thousand particles can be reasonably evolved: it takes around

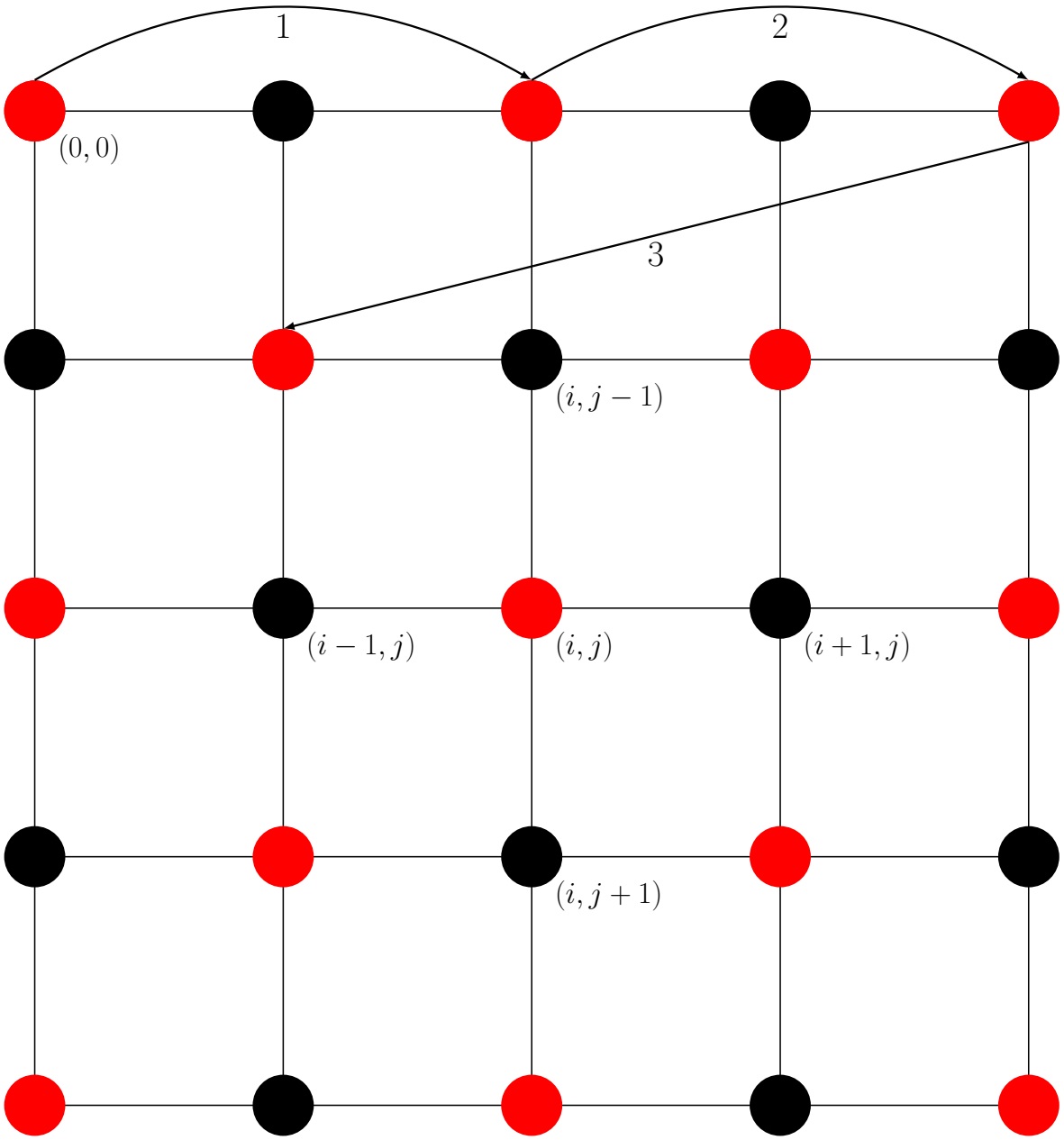


Figure 5.3 – Red-black ordering scheme. Each circle represents a point on the mesh identified here by two indices i and j. When iterating to estimate the gravitational potential, a first loop is done over all the red points (following the arrows labelled 1, 2 and 3) updating the potential on the way. Then a second loop is done, on the black points this time, using the updated values of the red points to obtain the new estimate.

two hours to evolve 64^3 particles with a basic non-parallel C code, but it scales very well when using several CPU cores).

Particle Mesh (PM) Methods: In these methods, the errors only come from the interpolations between the particles and the grid as the Poisson equation is solved exactly on the grid points. As a consequence, the gravitational forces cannot be accurately calculated below a few times the grid spacing. In addition, the grid spacing is usually limited by the available memory as dividing the grid spacing by two implies eight times more grid cells. However, the particle mesh methods are very computationally efficient as they scale as $\mathcal{O}(N_p + N_c \log N_c) = \mathcal{O}(N_c \log N_c)$ (Hockney and Eastwood, 1988) where N_c is the number of cells in the grid.

Tree Methods: Tree-based methods solve the Poisson equation to any desired accuracy by tuning the opening criterion and the order of the force expansion. These methods scale as $\mathcal{O}(N_p \log N_P)$. Thus they are not as efficient as the particle mesh methods but they do not have a resolution limit due to a finite grid size.

Mesh Relaxation Methods: Like tree methods, mesh relaxation methods can solve the Poisson equation up to any accuracy, here by tuning the convergence criterion. However, this technique requires an adaptative meshing to be really efficient as a large number of iterations is required to achieve convergence on many gridpoints as perturbations travel from neighbour to neighbour only.

Hybrid Methods: In order to benefit from the advantages of several techniques, it is possible to combine them. Usually a method is used to compute the short-range forces while another computes the long-range ones. For example, it is possible to improve the PM methods by explicitly computing the forces between close (basically, the particles that fit into a single cell) particles as in the PP technique: this is called a P^3M code (Hockney and Eastwood, 1988; Efstathiou and Eastwood, 1981). Such codes can solve the Poisson equation up to any required accuracy and scale as $\mathcal{O}(N_n \log N_p)$ where N_n is the average number of particles used for the PP part (Springel, 2005). A good balance must be done as one does not want the PP calculations to become dominant. To this already hybrid technique, it is possible to add adaptative meshing to reduce the need for direct computations (Couchman et al., 1995), but this usually implies strong memory requirement to store the grid.

Another hybrid method that is commonly employed, mixes a tree-based method for the short range forces and a PM method for the long range forces. This Tree-PM (Xu, 1995) method allows to compute the short range forces at an arbitrary accurary and benefits from the efficiency of the PM method for the long range ones.

5.3 Smoothed Particle Hydrodynamics

To describe a gas one needs several equations: the equations of motion and the equation of state. In cosmology, most gases can be described with a polytropic equation of state:

$$p = A\rho^\gamma \tag{5.13}$$

where γ is the adiabatic index and A a constant at least in space (and often in time). Three Navier-Stokes equations describe the motion of a gas in a cosmological context:

— the continuity equation (conservation of mass):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \tag{5.14}$$

— the conservation of momentum:

$$\frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot (\rho \mathbf{v}^2) + \nabla p = \rho \mathbf{g}, \tag{5.15}$$

— the conservation of energy:

$$\frac{\partial \rho E}{\partial t} + \nabla \left[(\rho E + p) \mathbf{v} \right] = \rho \mathbf{v} \cdot \mathbf{g}. \tag{5.16}$$

In these equations, p is the pressure of the gas, ρ its density, $\mathbf{v}$ its velocity, $\mathbf{g}$ the gravitational acceleration and E is the internal energy per unit mass. Of course, the gas is also subject to gravity and obeys the Poisson equation.

Smooth particle hydrodynamics (SPH) is a Lagrangian method for solving the equations of hydrodynamics, or in a simpler form it is the answer to the question: "How does one compute the density field at any point from an arbitrary distribution of point mass particles?". One solution is the one described in section 5.2.2 using a mass weighing scheme, but it has limitations on accuracy, speed and resolution. Another solution, illustrated in figure 5.4, is to use the SPH estimator, developed by Gingold and Monaghan (1977) and Lucy (1977):

$$\rho(\mathbf{r}) = \sum_i m_i W(\mathbf{r} - \mathbf{r}_i, h) \tag{5.17}$$

where W is a weighing function (or smoothing kernel) and h is a scale parameter governing the fall of W with respect to the distance between particles. This scale parameter is usually taken so that the "mass inside the smoothing sphere" is kept constant (Springel and Hernquist, 2002).

The smoothing kernel must have several properties including:

— normalization (because of mass conservation $\int \rho \, \mathrm{d}V = \sum_i m_i$):

$$\int_V W(\mathbf{r}' - \mathbf{r}_i, h) \, \mathrm{d}V' = 1, \tag{5.18}$$

— it must be positive and decrease smoothly,

— symmetric with respect to $\mathbf{r} - \mathbf{r}'$: $W(\mathbf{r}' - \mathbf{r}, h) = W(|\mathbf{r} - \mathbf{r}'|, h)$,

— flat central portion, so that the density is not strongly changed by a small displacement of a close neighbour.

The first natural choice was a Gaussian kernel:

$$W(\mathbf{r} - \mathbf{r}', h) = \frac{1}{h^3 \pi \sqrt{\pi}} \exp\left[-\frac{|\mathbf{r} - \mathbf{r}'|^2}{h^2} \right] \tag{5.19}$$

which is infinitely smooth. However, it implies interaction with all the particles and thus is computationally expensive. It is nowadays replaced by kernels with compact support like the cubic spline (Monaghan, 1985):

$$W(\mathbf{r} - \mathbf{r}, h) = \frac{1}{\pi} \begin{cases} \frac{1}{4} \left(2 - \frac{r}{h}\right)^3 - \left(1 - \frac{r}{h}\right)^3 & 0 \le r < h \\ \frac{1}{4} \left(2 - \frac{r}{h}\right)^3 & h \le r < 2h \\ 0 & \text{otherwise} \end{cases}. \tag{5.20}$$

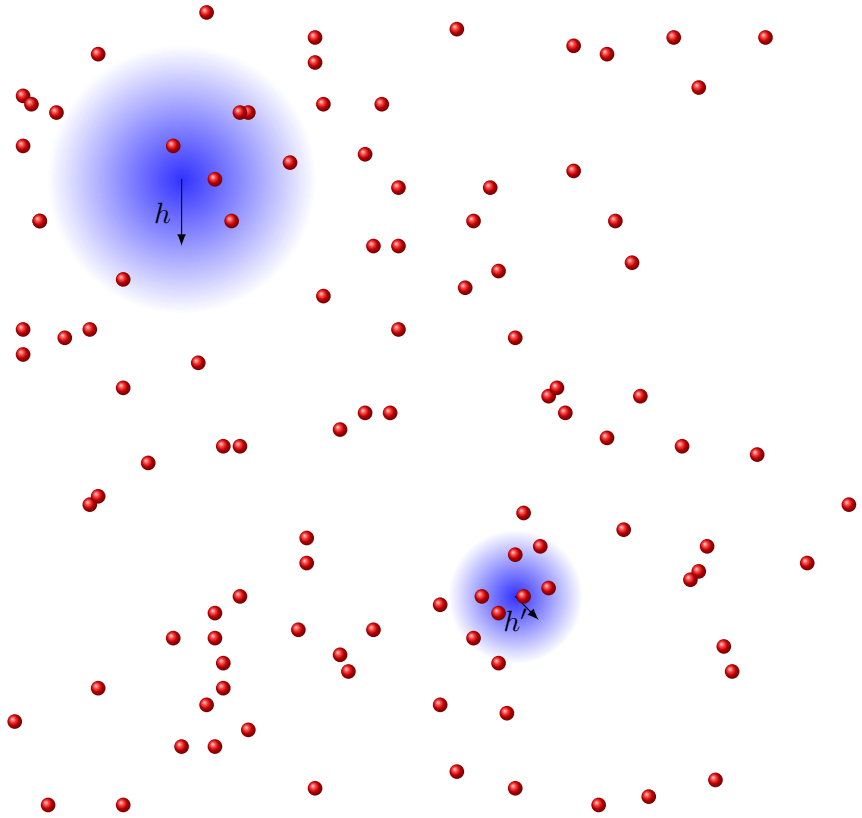


Figure 5.4 – Density field computation using SPH: the density is calculated at an arbitrary point with a weighed sum over the neighbouring particles. The weight decrease with the particle distance is described by a scale factor h.

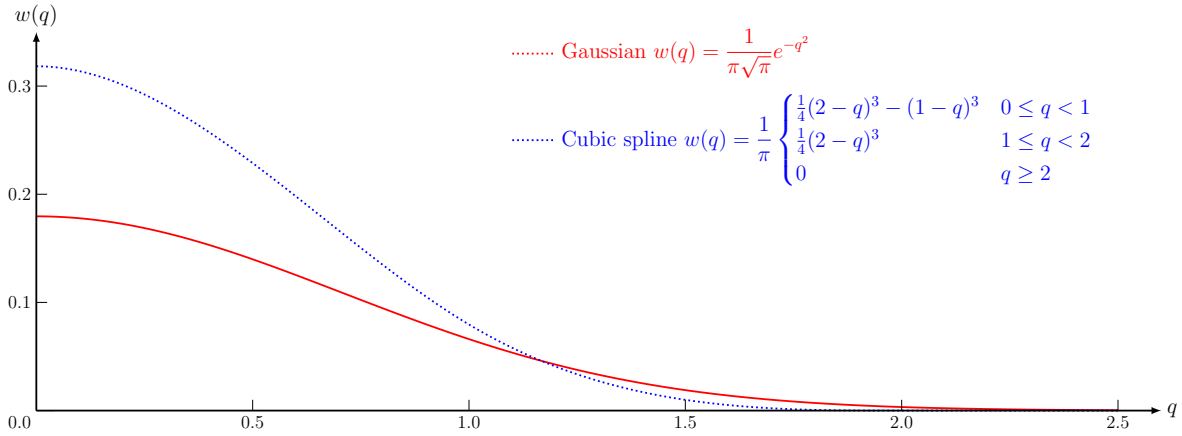


Figure 5.5 – Illustration of two kernel functions for SPH. I here plot $w(q)$ where $q = |\mathbf{r} - \mathbf{r}'|/h$ and $W(|\mathbf{r} - \mathbf{r}'|, h) = \frac{1}{h^3} w(q)$

Both kernel functions are represented in figure 5.5. Higher order splines can be used for additional smoothness but at the cost of a more expensive evaluation.

A very interesting thing about SPH is that one can derive all the hydrodynamics equations from the density estimator using for example the discrete Lagrangian

$$L = \sum_i m_i \left(\frac{1}{2} v_i^2 - u_i(\rho_i, s_i) \right), \tag{5.21}$$

were m_i, v_i, u_i, ρ_i, s_i are respectively the mass, the velocity, the internal energy, the density and the entropy of particle i. Then assuming that this Lagrangian is differentiable, that the evolution of the gas is adiabatic (no change in entropy) and neglecting the discreteness in the time integral, we get the full set of SPH equations for evaluating ρ, $\mathbf{v}$ and u:

$$\rho = \sum_i m_i W(\mathbf{r} - \mathbf{r}_i, h); \qquad h = h(\rho), \tag{5.22}$$

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\sum_i m_i \left[\frac{P}{\Omega \rho^2} \nabla W(\mathbf{r} - \mathbf{r}_i, h) + \frac{P}{\Omega \rho^2} \nabla W(\mathbf{r} - \mathbf{r}_i, h_i) \right], \tag{5.23}$$

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{P}{\Omega \rho^2} \sum_i m_i (\mathbf{v} - \mathbf{v}_i) \cdot \nabla W(\mathbf{r} - \mathbf{r}_i, h), \tag{5.24}$$

where

$$\Omega = 1 - \frac{\partial h}{\partial \rho} \sum_i m_i \frac{\partial W(\mathbf{r} - \mathbf{r}_i, h)}{\partial h} \tag{5.25}$$

and P is the pressure. Quantities without subscript are evaluated at the point of interest $\mathbf{r}$. The details of the calculations can be found in Price (2012a). More complete reviews and discussions of SPH can be found in Monaghan (1992), Monaghan (2005), Rosswog (2009), Springel (2010) and Price (2012b).

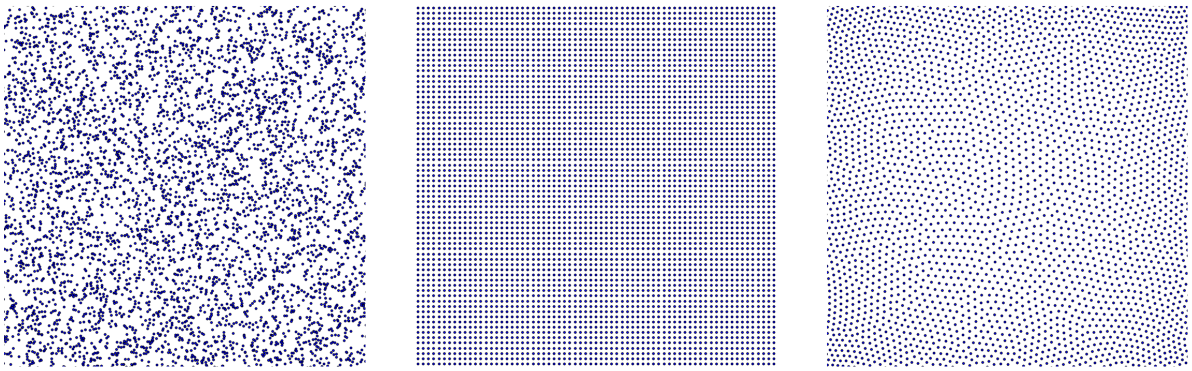
Grid based codes like `RAMSES` (Teyssier, 2002) or `ENZO` (The Enzo Collaboration et al., 2013) which are using adaptive mesh refinement (AMR) solve the hydrodynamical equations using Godunov scheme (Godunov, 1959) extension in which data are no longer approximated by piecewise constant functions but rather piecewise linear function (van Leer, 1979) or piecewise parabolic function (Colella and Woodward, 1984; Woodward and Colella, 1984).

5.4 Initial Conditions

Generating good initial conditions for cosmological simulation is important as errors can be amplified by the simulation. The goal is to represent a power spectrum using discrete particles. I will first briefly discuss the pre-initial condition, i.e. the distribution of particles on which the power spectrum is applied. Then we will see how to generate proper initial conditions.

5.4.1 Pre-initial Conditions

Before the generation of actual initial conditions that can be used as an input to a N-body code, it is important to have "pre-initial" conditions; a distribution of particles that represents an unperturbed system. I only summarize here three types of distributions: randomly placed particles, particles on a grid and glass distributions, which are represented in figure 5.6. An initial study of pre-initial conditions can be found in White (1994) and a more recent one in L'Huillier et al. (2014) .



(a) Random uniform distribution (b) Grid distribution (c) Glass distribution

Figure 5.6 – Three different types of pre-inital particle distribution: randomly (uniform) placed particles, particles placed on a regular grid and glass distribution.

The first distribution, which uses randomly placed particles following a uniform law, should not be used as it strongly suffers from Poisson noise and it has an initial power spectrum corresponding to a white noise. Evolving such a distribution will lead to the formation of non linear objects, which should not be the case as no perturbations have been set yet.

A better approach is to use a distribution of particles placed on a regular grid. However, the grid structure implies preferred directions along the axis that may develop into artefacts, especially at small scale.

The current preferred solution is to use a glass distribution (White, 1994). To construct such a distribution, one starts from a random distribution which is then evolved using an inversed (repulsive) gravitational force. The particle distribution is then smoothed, removing large-scale power as in the random distribution and suppressing preferred direction as in the grid distribution.

5.4.2 Generation of Initial Conditions

As seen in chapter 2, the original quantum density fluctuations form a random Gaussian field in a homogeneous and isotropic space. It is a (discretized) field $\phi(\mathbf{r})$ with a Gaussian N-point distribution function, i.e. the probability to have $\phi(\mathbf{r}_1) = x_1$, ..., $\phi(\mathbf{r}_N) = x_N$ is given by

$$P(\mathbf{X})\mathrm{d}\mathbf{X} = \frac{1}{2\pi\sqrt{\det C}} \exp\left[-\frac{\mathbf{X}C^{-1}\mathbf{X}^T}{2}\right] \mathrm{d}\mathbf{X} \tag{5.26}$$

where $\mathbf{X} = (x_1, \ldots, x_N)$ and C is the covariance matrix which for $N = 2$ is given by:

$$C = \begin{bmatrix} \sigma^2 & \xi(x_1, x_2) \\ \xi(x_2, x_1) & \sigma^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & \xi(r) \\ \xi(r) & \sigma^2 \end{bmatrix} \tag{5.27}$$

where $r = |\mathbf{r}_2 - \mathbf{r}_1|$.

Equivalently, in k-space, the field can be written:

$$\hat{\phi}(\mathbf{k}) = |\hat{\phi}(\mathbf{k})|e^{i\theta} = |\hat{\phi}(k)|e^{i\theta} \tag{5.28}$$

where the phase θ is random (with uniform distribution) and the amplitude $|\hat{\phi}(k)|$ follows a Rayleigh probability distribution function:

$$P(x)\mathrm{d}x = \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right] \mathrm{d}x. \tag{5.29}$$

whose variance is related to the power spectrum of the field through

$$\sigma^2 = \frac{k^3}{(2\pi)^3} P(k) \tag{5.30}$$

Thus the power spectrum of a Gaussian random field is sufficient to describe it.

The initial power spectrum (i.e. just after inflation) is usually written as

$$P(k) = A \left(\frac{k}{k_0} \right)^{n_s} \tag{5.31}$$

where k_0 is called the pivot mode, A and n_s are respectively the amplitude and the spectral index of the primordial power spectrum. Because it would be too expensive to evolve a system starting at the epoch of inflation, initial conditions are generated at a much lower redshift $z \sim 10 - 100$ with the assumption that the evolution up to the starting point is linear. Therefore, the initial power spectrum has to be corrected by a transfer function T:

$$P(k, z) = A \left(\frac{k}{k_0} \right)^{n_s} |T(k, z)|^2 \tag{5.32}$$

Such transfer functions are nowadays computed by numerical codes like `CMBFAST` (Seljak and Zaldarriaga, 1996) or its successor `CAMB` (Lewis et al., 2000).

To generate an initial overdensity field in a cubic volume of side length L, one should first generate the amplitude of the Fourier transform of the overdensity field $\delta(\mathbf{r})$ at the different modes available in the box:

$$k_l = \frac{2l\pi}{L} \tag{5.33}$$

This can be done with

$$\hat{\delta}(k) = \sigma \sqrt{-2 \ln U} \tag{5.34}$$

where U is a random variable drawn from the uniform distribution on the interval $]0, 1]$. In conjunction with a random phase, it gives the Fourier transform of the over density field which can then be recovered in real space using the inverse Fourier transform.

From a pre-initial distribution of particles and a density field $\rho = \langle \rho \rangle (\delta + 1)$, it is now possible to compute the initial particle distribution in position and velocity. This is done using the Zel'dovich approximation (Zeldovich, 1970; Shandarin and Zeldovich, 1989):

$$\mathbf{r}_i = \mathbf{x}_i - D(t) \nabla \Phi(\mathbf{x}_i) \tag{5.35}$$

where $\mathbf{r}_i$ is the new position of the particle i sitting at $\mathbf{x}_i$ in the pre-initial distribution, $D(t)$ is the growth rate of linear density fluctuations and Φ is the gravitational potential generated by the density field. The form of $D(t)$ depends on the chosen cosmology. The corresponding velocities are given by

$$\mathbf{v}_i = -\frac{\mathrm{d}D}{\mathrm{d}t}(t) \nabla \Phi(\mathbf{x}_i) \tag{5.36}$$

The Zel'dovich approximation corresponds to the first order of the Lagrangian perturbation theory and can be pushed to second order for more accurate initial conditions, especially low starting redshifts (Scoccimarro, 1998; Crocce et al., 2006; L'Huillier et al., 2014).

We now have the full procedure for generating initial conditions from a given power spectrum $P(k)$ (which depends on several cosmological parameters, see chapter 2). It can be summarized in 4 points:

— generate a random field in k-space with variance proportional to $P(k)$ at each k,
— take the inverse Fourier transform of this field, to obtain an overdensity field with the required power spectrum,
— generate a pre-initial distribution of particles,
— displace the particles and set their initial velocities using either the Zel'dovich approximation or the second order Lagrangian perturbation theory.

I have presented the main techniques to simulate an ensemble of particles that interact through gravitation. We have also seen the basics of smoothed particle hydrodynamics and of initial conditions generation. All this theoretical foundations will be useful to understand the next chapter in which I will present the grid of simulations that have been run to study the effect of various parameters on the one-dimensional matter power spectrum.

Chapter 6

Lyman-α Forest Power Spectrum with Hydrodynamical Simulations

"All we have to decide is what to do with the time that is given to us."
—John Ronald Reuel Tolkien,
The Fellowship of the Ring (1954)

SIMULATING universes with different combinations of cosmological parameters is the last required step before being able to extract cosmological constraints from the measured Lyman-α forest power spectrum (see chapter 4). Due to the quality of this measurement I had to go further than what has already been done in the past, in order to match the precisions of the measurements and of the simulations. This work led to the publication of a paper, Borde et al. (2014), and was presented in several collaboration meetings. The neutrino implementation is described and studied in detail in Rossi et al. (2014).

In this chapter, I present a suite of cosmological N-body simulations made with `GADGET-3` with cold dark matter and baryons, specifically aiming at modelling the low-density regions of the intergalactic medium as probed by the Lyman-α forests at high redshift that were used in the study of chapter 4. The outline of the chapter is as follows: I first present the motivations for this work and the previous studies in section 6.1. In section 6.2, I describe our grid and the values chosen for the different parameters we varied. Then, in section 6.3, I present the simulation pipeline, along with the solutions I chose to solve issues such as the generation of the initial conditions or the radiative cooling and heating processes that occur in the intergalactic medium (IGM). Tests that were made to determine the required characteristics of our simulations in the light of our goals are presented in section 6.4. In section 6.5, I describe the splicing technique that is applied in order to obtain simulations with the desired resolution and box size. The validity of our grid approach is discussed in section 6.6. In section 6.7, I eventually present preliminary constraints on the probed cosmological parameters. Conclusions and perspectives are given in section 6.8. A recapitulation of all the simulations performed for this study is given in appendix A.

6.1 Motivations

In chapter 2, we have seen how the Lyman-α forest absorption can be used as a tracer of the varying density of intergalactic gas expected from the growth of structure from primordial fluctuations in the Universe (Croft et al., 1998). We use it to compute the flux power spectrum of the Lyman-α forest. It can later be used to constrain cosmology (Alcock and Paczyński, 1979; Hui et al., 1999; McDonald and Miralda-Escude, 1999; Croft et al., 2002; McDonald et al., 2005; Seljak et al., 2005), the baryonic acoustic oscillation peak position (McDonald and Eisenstein, 2007) or the sum of the masses of neutrinos (Seljak et al., 2005; Viel et al., 2010).

Whereas the measurement of the three-dimensional power spectrum uses only information from the flux correlation of pixel pairs in different quasar spectra and thus provides information on rather large scales, the one-dimensional power spectrum P_{1D} is linked to the three-dimensional power spectrum P_{3D} through the equation 2.18:

$$P_{1D}(k_{\parallel}) = \frac{1}{2\pi} \int_{k_{\parallel}}^{\infty} k P_{3D}(k) \mathrm{d}k \tag{6.1}$$

The one-dimensional power spectrum uses the correlation of pixel pairs on the same quasar spectrum and thus provides a complementary, useful information on smaller scales that are fundamental to constrain the physical parameters of the Lyman-α forest. The one-dimensional P_{1D} is probing scales at the transition from linear to non-linear regime. Therefore, cosmological simulations are required to provide insight on the non-linear physics of the intergalactic medium on the small scales probed by P_{1D}. Such simulations are then used to constrain various cosmological and astrophysical parameters that have an effect on the power spectrum (Viel et al., 2004; McDonald et al., 2005; Viel and Haehnelt, 2006; Bolton et al., 2008; Viel et al., 2010; Bird et al., 2011, 2012).

An intergalactic medium heated exclusively by photo-ionization can be modelled with hydrodynamic simulations (Cen, 1992; Zhang et al., 1995; Hernquist et al., 1996; Hui and Gnedin, 1997; Hui et al., 1997; Viel et al., 2004) and the physics at play in this model is well understood. However, mechanisms such as radiative transfer effects during hydrogen and helium reionization (Abel and Haehnelt, 1999) or the mechanical effects of galactic winds and quasar outflows can quickly complicate this simple picture, making simulations even more useful.

Here, I present a set of 36 cosmological smoothed particle hydrodynamics (SPH) and N-body simulations that reproduce the impact on the one-dimensional matter power spectrum of the values taken by the most relevant cosmological and astrophysical parameters, including the sum of neutrino masses. Only the baryonic particles undergo a SPH treatment (see section 5.3), i.e. they receive an additional hydrodynamic acceleration, and their internal entropy per unit mass is evolved as an independent thermodynamic variable. The requirements in terms of box size, resolution and redshift coverage of our simulations were derived from the Data Release 9 quasar catalogue (Pâris et al., 2012; Ahn et al., 2012) of the Baryon Oscillation Spectroscopic Survey (Dawson et al., 2013). I extrapolate these requirements so that this suite of simulations may also be used for future spectroscopic surveys such as eBOSS [1] (planned for 2014-2018) and even for preliminary studies for DESI [2] (Schlegel et al., 2011) (2018-2023).

In order to realize this study, I have ran some preliminary simulations, including the convergence test presented in section 6.4, using a total of around 250 thousand hours of CPU time on

1. `http://www.sdss3.org/future/eboss.php` and `http://www.sdss3.org/future/sdss4.pdf`
2. `http://desi.lbl.gov`

various systems at the National Energy Research Scientific Computing Center [3] and at the *Très Grand Centre de Calcul*[4]. Using these simulations as a basis for time consumption estimations, I have written two proposals (one at the French level and one at the European level), resulting in the allocation of respectively seven million and five hundred thousand hours of CPU of the machine Curie thin node [5]. This machine is made of 5040 computing nodes, each of them having two eight-core Intel processors at 2.7 GHz and 64 GB of memory. An additional one hundred thousand hours were used to finalize some studies.

6.2 Simulation Grid

Ideally, in order to derive confidence intervals on each parameter of a cosmological model with eight to ten free parameters, one would like to compute theoretical predictions for thousands of models, exploring most of the parameter space. Statistical frameworks have been studied to optimize the precision of the model for a reduced number of simulations, such as Latin hypercube sampling (Tang, 1993). While this method is superior to a random sampling of the parameters, for instance as regards the attained precision (McKay et al., 2000), it still requires a large number of simulations. Latin hypercube sampling has only been tested so far to predict the power spectrum on large scales, using low resolution simulations, of order 128^3 particles for a 450 Mpc.h^{-1} box (Heitmann et al., 2009; Schneider et al., 2011).

When dealing with Lyman-α data, running large numbers of simulations is not possible due to the high execution time of each hydrodynamical simulation. Hence, people have developed various approximate methods in which a restricted number of simulations is used either to calibrate a flux-to-matter power-spectrum bias function or to Taylor expand the flux power-spectrum with respect to cosmological parameters in the vicinity of a best-fit model. For cosmological predictions of the power spectrum in the Lyman-α regime where hydrodynamical simulations are required, the grid approach as presented in (Viel et al., 2006) is generally adopted (cf. (Wang et al., 2013) for instance for a recent application). This is the method we have selected for this work.

The second-order Taylor expansion about our best-guess model is given by

$$\begin{aligned} f(\mathbf{x}+\mathbf{\Delta x}) =& f(\mathbf{x}) \\ &+ \sum_i \frac{\partial f}{\partial x_i}(\mathbf{x})\Delta x_i \\ &+ \frac{1}{2}\sum_i\sum_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})\Delta x_i \Delta x_j. \end{aligned} \tag{6.2}$$

With n parameters, the total number of simulations required to get the Taylor expansion coefficients is $1+2n+n(n-1)/2$, where the terms account for, respectively, the central (or best-guess) model, two other values of each parameter to derive the first and second-order derivatives, and the simultaneous variation of each pair of parameters to compute the cross derivatives (cf. figure 6.1). With this lattice, all derivatives are approximated to second order except the cross derivatives which are approximated to first order. This approximation is justified by the fact that the parameters are reasonably decoupled, and it allowed us to reduce the CPU time consumption since second-order cross derivatives would require additional $n(n-1)/2$ simulations.

3. `http://www.nersc.gov/`
4. `http://www-hpc.cea.fr/fr/complexe/tgcc.htm`
5. `http://www-hpc.cea.fr/fr/complexe/tgcc-curie.htm`

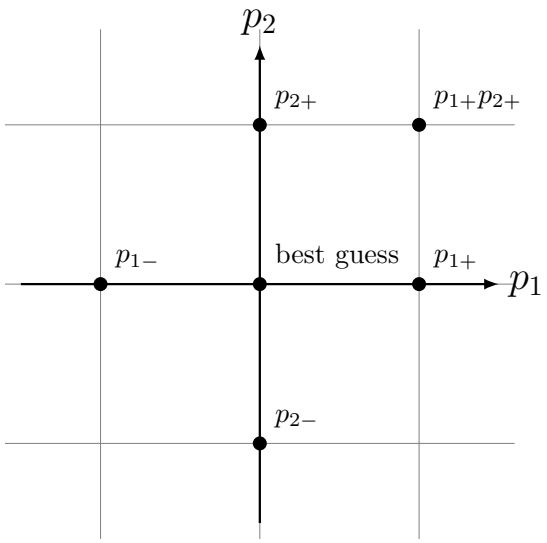


Figure 6.1 – Illustration of the required grid for a second-order Taylor expansion in a two-dimensional parameter space.

Parameter	Best fit	68% limits
n_s	0.9624	0.9616 ± 0.0094
σ_8	0.8344	0.834 ± 0.027
Ω_m	0.3175	0.314 ± 0.020
H_0	67.11	67.4 ± 1.4

Table 6.1 – Cosmological parameters values from Planck temperature power spectrum alone. I give best fit and 68% confidence limits.

Our central model assumes massless neutrinos. As we do not simulate neutrinos with negative mass, we therefore use a forward finite difference to compute the derivatives with respect to neutrino mass.

6.2.1 Simulation Variable Parameters

To model the physics of the Universe, we introduced two categories of parameters that are varied in the simulations: cosmological parameters that describe the cosmological model in the simplest case of ΛCDM assuming a flat Universe with massless or massive neutrinos, and astrophysical parameters that model the astrophysics within the IGM and the relation between temperature and density of the gas. A summary of all simulations performed to compute the coefficients of the Taylor expansion is given in appendix A.

6.2.1.1 Cosmological Parameters

This first category contains five parameters: the amplitude of the matter power spectrum σ_8, the spectral index of primordial density fluctuations n_s, the matter density Ω_m, the Hubble constant H_0 and the sum of neutrino masses $\sum m_\nu$.

The values for our central model (with massless neutrinos) are in agreement with the latest best-fit values from Planck (Planck Collaboration et al., 2013), which I recall in table 6.1.

I chose the range of variation for these parameters so as to include other recent constraints from the Wilkinson Microwave Anisotropy Probe seven year data (Komatsu et al., 2011), the South Pole Telescope data (Hou et al., 2014) and the SuperNova Legacy Survey three year data (Conley et al., 2011; Sullivan et al., 2011), thus taking into account the fact that results from Planck for H_0 (respectively Ω_m) are low (respectively high) compared to other measurements. Central value and range for each of the cosmological parameters are given in table 6.2.

6.2.1.2 Astrophysical Parameters

This category includes two redshift-dependent parameters that describe the temperature-density relation of the IGM for $\rho/\langle\rho\rangle \leq 10$:

$$T(\rho, z) = T_0(z) \times \left(\frac{\rho}{\langle\rho\rangle}\right)^{\gamma(z)-1}, \tag{6.3}$$

where ρ is the baryonic density, $T_0(z)$ is a normalization temperature and $\gamma(z)$ a logarithmic slope. In the post processing step, I scale the effective optical depth $\tau_{\text{eff}} = -\ln(\langle F\rangle) = -\ln(\langle e^{-\tau}\rangle)$, where F is the flux and τ is the optical depth, so that it follow a power law $\tau_{eff}(z) = \tau_A \times (1+z)^{\tau_S}$. I can thus test different mean flux normalizations and evolutions with redshift by varying the parameters τ_A and τ_S.

In the absence of a clear consensus on the heating history of the IGM, I took the $T(\rho)$ measurements from Becker et al. (2011) assuming $\gamma = 1.3$ as our central model, and I chose a wide variation around these values so that other recent measurements (Garzilli et al., 2012; Lidz et al., 2010; Schaye et al., 2000) fall into the explored range. The evolution with redshift of $\gamma(z)$ and $T_0(z)$ in our simulations is therefore designed to reproduce the $T(\rho)$ measurements presented by Becker et al. (2011) through an adaptation of the cooling routines in the simulation code. Thus I only need to fix those two parameters at a given redshift, in our case $z = 3.0$, which corresponds to the central redshift of our study. In practice, I do not set $T_0(z=3)$ and $\gamma(z=3)$ but instead use two internal code parameters, `AMPL` and `GRAD`, that alter the amplitude and density dependence of the photo-ionization heating rates, such that $\epsilon_f = \texttt{AMPL} \times \delta^{\texttt{GRAD}} \times \epsilon_i$ where ϵ's are the heating rates and δ is the over-density. T_0 and γ are evaluated after the simulations have run, as explained at the end of section 6.3. Given the one-to-one correspondence between (T_0, γ) and (`AMPL`, `GRAD`), I prefer to keep on quoting T_0 and γ since these parameters have a physical meaning and can be compared to other studies.

6.2.1.3 Grid Values

The central values and variation ranges of the parameters of our study are summarized in table 6.2. With six varying parameters, this represents a total of 36 cosmological simulations in our grid. In addition I also ran simulations with $\sum_\nu m_\nu = 0.1\,\text{eV}, 0.2\,\text{eV}$ to test different ranges for the neutrino derivatives.

6.2.2 Requirements

I base our minimal requirements for the resolution and box size of our simulations on the largest currently-available spectroscopic survey: SDSS-III/BOSS (Dawson et al., 2013). Those requirements are in part determined by the resolution of the spectrograph and the extension of the Lyman-α forest data (see chapter 3). In addition, because the 1D power spectrum results

Parameter	Central value	Range
n_s	0.96	± 0.05
σ_8	0.83	± 0.05
Ω_m	0.31	± 0.05
H_0	67.5	± 5
$T_0(z=3)$	14000	± 7000
$\gamma(z=3)$.	1.3	± 0.3
$\sum m_\nu$	0.0	$+0.4$, $+0.8$

Table 6.2 – Central values and variation ranges of the parameters for our simulation grid.

from an integral over the 3D power spectrum up to $k = \infty$ (cf. Eq. 2.18), the resolution of the simulation has to be of the size of the smallest structures in the transverse direction. For structures in local hydrostatic equilibrium, this would be the Jeans scale, of order a few 100 kpc at $z = 3$. In an SPH approach, over-dense regions are sampled with much higher spatial resolution than average, so the difficulty does reside there. Under-dense regions, on the other hand, might not necessarily be in local hydrostatic equilibrium. The decisive solution to ensure that the simulations do resolve the relevant structures is therefore to perform a convergence test.This is the first step that was done to fix the characteristics of our simulations. The details are given in section 6.4. I describe below a first estimate of the simulation requirements driven by instrumental issues and experimental data.

The quasar coadded spectra provided by the SDSS pipeline (Bolton et al., 2012) are computed with a constant pixel width of $\Delta v = 69\,\mathrm{km\,s^{-1}}$. The largest possible mode is determined by the Nyquist-Shannon limit at $k_{\mathrm{Nyquist}} = \pi/\Delta v = 4.5 \times 10^{-2}\,(\mathrm{km/s})^{-1}$. For the other bound, the smallest k-mode is driven by the extension of the Lyman-α forest which lies between the Lyman-α and Lyman-β emissions respectively at 1216 Å and 1026 Å. The minimum k-mode one can obtain must take into account that the exploitable Lyman-α forest is smaller than the separation of the two emission peaks due to their respective widths. In addition, instrumental constraints often make the first and last theoretical modes very difficult to obtain with reasonable precision from data. McDonald et al. (2006) used the region $1041\,\text{Å} < \lambda_{\mathrm{rest}} < 1185\,\text{Å}$ which corresponds to $\Delta v \simeq 4 \times 10^4\,\mathrm{km\,s^{-1}}$ and $k_{\mathrm{min}} = 1.6 \times 10^{-4}\,(\mathrm{km/s})^{-1}$. They restricted their analysis however to $k_{\mathrm{min}} = 1.4 \times 10^{-3}\,(\mathrm{km/s})^{-1}$ and $k_{\mathrm{max}} = 1.8 \times 10^{-2}\,(\mathrm{km/s})^{-1}$. Using more recent BOSS data, we computed the 1D power spectrum from forest lengths corresponding to a third of the total available range in order to restrain the redshift span to $\Delta z = 0.2$ at most (see chapter 4). This led to $k_{\mathrm{min}} \sim 1.0 \times 10^{-3}\,(\mathrm{km/s})^{-1}$. I therefore consider simulations that should cover the minimal range $1 \times 10^{-3}\,(\mathrm{km/s})^{-1} < k < 2 \times 10^{-2}\,(\mathrm{km/s})^{-1}$, which corresponds approximately to $0.1\,(\mathrm{Mpc}/h)^{-1} < k < 2\,(\mathrm{Mpc}/h)^{-1}$ at $z \sim 3$.

In numerical simulations, the two important parameters are the size of the box L that determines the smallest k-mode, and the ratio $N^{1/3}/L$, where N is the number of particles, that drives the largest k-mode. One may note that due to the computational algorithms used nowadays in simulations, such as smoothed particle hydrodynamics (SPH) or adaptive mesh refinement (AMR) in which "resolution follows density", particle spacing in regions of interest will be significantly smaller than $L/N^{1/3}$. However, I will use this approximation as it reflects a worst-case scenario and is thus fairly conservative.

For a simulation evolving N particles in a box of size L, the smallest k-mode is given by $k_{\min} = 2\pi/L$, and the Nyquist limit by $k_{\text{Nyquist}} = (2\pi N^{1/3})/(2L)$. Therefore, the theoretical minimum requirement for a simulation to reproduce the BOSS DR9 data precision is a box of about $100\,\text{Mpc}\,h^{-1}$ with $\sim 100^3$ particles of each type. These theoretical requirements will be refined with the convergence tests provided in section 6.4. In particular, we will see that many more particles are needed to achieve the required resolution.

6.3 Pipeline

All the components of our simulation workflow are represented on figure 6.2. The first part of the pipeline is the production of the initial condition snapshot. This is done in the linear approximation with perturbations treated up to second order. The simulations are then performed using both N-body and hydrodynamic (SPH) treatments for the baryon component. The post-processing stage takes the result of the simulations and computes the power spectra that will be compared to data through the Taylor expansion described earlier. In addition to baryon and dark matter particles, simulations with massive neutrinos include them as a third type of particles.

The products of our suite of simulations are obtained at thirteen predefined redshifts, equally spaced every $\Delta z = 0.2$ from $z = 2.2$ to $z = 4.6$. Our selection of redshifts reflects the possibilities of current and foreseen large-scale spectroscopic surveys. The lower bound results from the UV cut-off of CCDs at $\lambda \sim 350$ nm that prevents the observation of Lyman-α below $z \sim 2.2$. The upper bound results from the quasar luminosity function that peaks near $z \sim 2$ and drops significantly for $z > 3$. The density of QSOs at $z > 4$ is of the order of $2\,\text{deg}^{-2}$ for a limiting magnitude $g < 23$ as expected for the future DESI survey. This is less than an order of magnitude smaller than at $z \sim 2$ (Palanque-Delabrouille et al., 2013).

6.3.1 `CAMB`

The Code for Anisotropies in the Microwave Background (`CAMB`)[6] (Lewis et al., 2000) is a numerical Boltzmann code written in Fortran 90. It is a parallelized line-of-sight integration code developed from `CMBFAST` (Seljak and Zaldarriaga, 1996) and `COSMICS` (Bertschinger, 1995), which is widely used (and thus tested) to calculate not only the lensed cosmic microwave background temperature and polarization spectra but also linear matter power spectra for different species of particles (in our case baryons, dark matter and sometimes neutrinos).

`CAMB` is here used to compute the transfer functions and linear power spectra that will be used in the next step to compute the initial displacement of particles (see section 5.4).

6.3.2 `2LPTIC`

All our simulations are tuned to obtained a given σ_8 at $z = 0$. This is done with the `spnorm` Python script that rescales the total matter power spectrum $P_S^{CAMB}(z_i)$ issued from `CAMB` before generating the initial conditions such that

$$P_S(z_i) = P_S^{\texttt{CAMB}}(z_i) \times \left[\frac{\sigma_8(z_i)}{\sigma_8^{\texttt{CAMB}}(z_i)}\right]^2, \tag{6.4}$$

6. `http://camb.info`

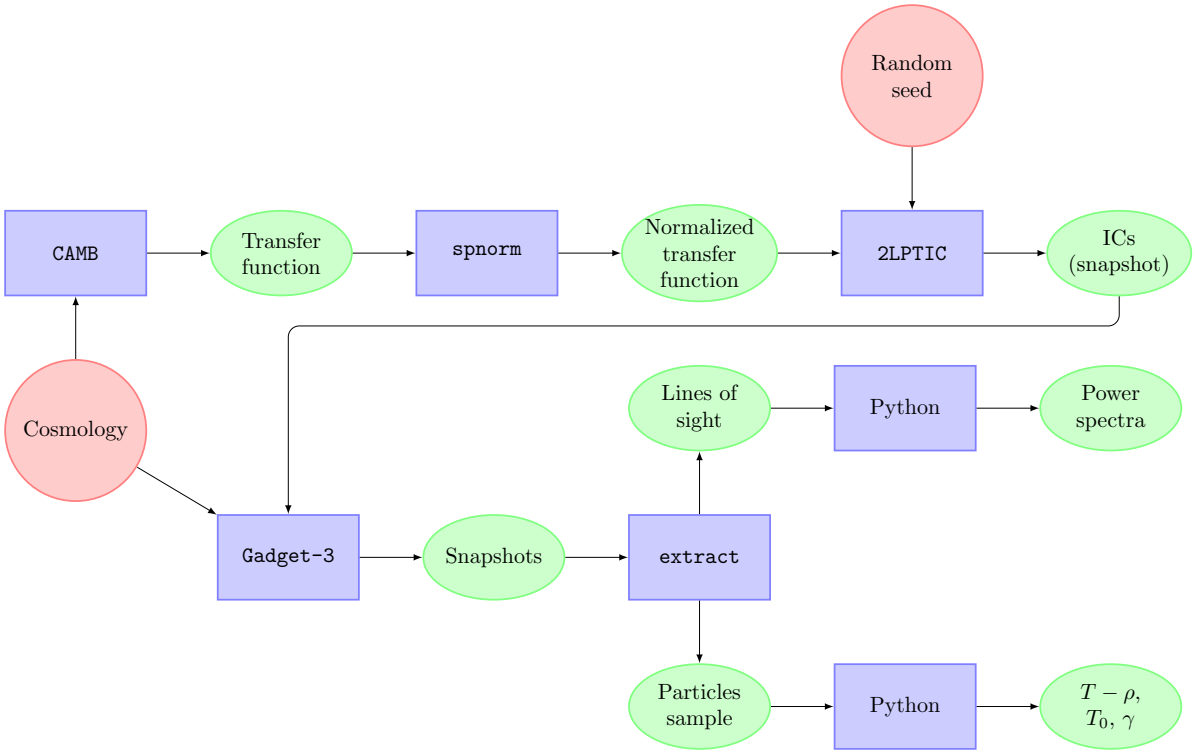


Figure 6.2 – Our simulation pipeline: red circles represent input from the user, blue rectangles are softwares and scripts and green ellipses stand for outputs from these softwares.

where z_i is the redshift at which the initial conditions are run, σ_8^{CAMB} is the value of σ_8 obtained with `CAMB` for a chosen cosmological model, and

$$\sigma_8(z_i) = \frac{\sigma_8(z=0)}{\sigma_8^{\texttt{CAMB}}(z=0)} \times \sigma_8^{\texttt{CAMB}}(z_i)\,. \tag{6.5}$$

This allows us to avoid rerunning `CAMB` with the corresponding initial spectral amplitude A_s. While `CAMB` includes radiation this is not the case for `GADGET-3`. However, if the radiation contribution is certainly present, its impact is very small in terms of matter power spectrum at the scales relevant for the present work and will be negligible in terms of flux power spectrum. So low-redshift evolution of the simulation is reproducing the wanted matter power spectrum in a regime in which radiation contribution can safely be ignored. The rescaled power spectra are then used as input to the `2LPTIC`[7] code that provides initial conditions based on second order Lagrangian Perturbation Theory (2LPT), rather than first-order (Zel'dovich approximation). The rescaling explicitly forces the simulation to achieve the desired value of σ_8 at $z = 0$. The choice of second order precision initial conditions is motivated by the discussion in Crocce et al. (2006) and the fact that I run cosmological simulations including neutrinos as a new particle type (Rossi et al., 2014). Indeed, because of their high velocity, neutrinos require initial conditions taken at rather low redshift in order to reduce Poisson noise (Ali-Haimoud and Bird, 2012; Bird et al., 2012). Initial conditions for all the grid simulations are run with the same seed. Comparison of power spectra obtained with different initial conditions are presented in section 6.4.

7. `http://cosmo.nyu.edu/roman/2LPT/`

6.3.3 Gadget-3

GADGET-3 (GAlaxies with Dark matter and Gas intEracT) is a massively parallel tree-PM code for cosmological simulations, originally developed by Volker Springel and collaborators (Springel et al., 2001; Springel, 2005). It is written in ANSI C, and uses the standardized message passing interface (MPI) along with several open-source libraries (GSL [8], FFTW [9]). Long range gravitational interactions are computed with a particle-mesh method (see section 5.2.2) and short range ones using a tree (see section 5.2.3). Gas-dynamics is followed with smoothed particle hydrodynamics (SPH, see section 5.3); collisionless dark matter, baryonic gas and neutrinos when present are represented by particles.

Since its original version (GADGET-1), the code has undergone a series of improvements and optimizations over several years (GADGET-2 and 3), to maximize the work-load balance and the efficiency in memory consumption and communication bandwidth. In what follows, I briefly describe the key features of the code.

GADGET-3 follows a collisionless fluid with standard N-body methods, and an ideal gas with smoothed particle hydrodynamics (SPH). The code solves simultaneously the dynamics of the collisionless component and of the ideal gas, both subject to and coupled by gravity in an expanding background space. The N-body implementation only differs from other cosmological codes by the accuracy of the gravitational field computation. A number of further physical processes have also been implemented in GADGET-3, from radiative cooling/heating physics to non-standard dark matter dynamics, star formation (I use a simple model to save computational time where dense and cool particles are directly converted to stars) and feedback. In figure 6.3, I present the evolution of a filament with redshift and in figure 6.4 I show the image of the four components (for a simulation with massive neutrinos, three otherwise) of a snapshot made with splotch [10]. Such realizations can be used for visual confirmation before quantitative analysis as well as for public outreach and education.

Several optimization strategies have been added in GADGET-3. These include a Peano-Hilbert space decomposition, a massively parallel version of the Fast Fourier Transform library, the possibility of splitting the simulation data across several files (to facilitate and speed-up the input/output process), and the fact that the code can be run efficiently on a large number of processors (several thousand at least). In its current version, GADGET-3 is highly efficient in memory consumption (it allocates up to 80 bytes per particle) and communication bandwidth, is versatile and flexible, accurate and fast. Another important aspect is the scalability of the code, i.e. its performance when the number of processors is increased, which has currently been tested up to 16 000 cores.

I started all our simulations at $z_i = 30$ with initial conditions based on second-order Lagrangian perturbation theory (Crocce et al., 2006), and adopted the same gravitational softening for the different species considered (i.e. baryonic gas, dark matter, stars and neutrinos), which however varies the length of the box and the size of the mesh chosen. Speciffically, I set the gravitational softening length (expressed in GADGET-3 internal units, and roughly equivalent to the Plummer softening length) to 0.81 for the simulations having a $25\,\mathrm{Mpc}\,h^{-1}$ boxsize and 768^3 particles per specie, while the softening is 3.25 for the other two runs, i.e., the $25\,\mathrm{Mpc}\,h^{-1}$ boxsize, 192^3 particles and the $100\,\mathrm{Mpc}\,h^{-1}$ boxsize, 768^3 particles. I use the QUICKLYA option in GADGET-3 to simulate the Lyman-α forest, assuming a helium mass fraction of $Y = 0.24$ in

8. http://www.gnu.org/software/gsl/
9. http://www.fftw.org/
10. http://www.mpa-garching.mpg.de/~kdolag/Splotch

the primordial gas. We neglect metals and the evolution of elementary abundances, as well as feedback processes and galactic winds. Along the lines of Viel et al. (2010), we adopted a simplified criterion for star formation: all gas particles whose overdensity with respect to the mean is above 1000 and whose temperature is below 1×10^5 K are turned into stars particles immediatly.

6.3.4 `extract`

The `GADGET-3` snapshots contain various fields among which the position **p** and velocity **v** for dark matter, gas particles and neutrinos if present. It also contains fields that are specific to the SPH treatment of gas particles: internal energy U, density ρ, electron fraction N_e, hydrogen fraction $N_{\rm H}$ and smoothing length h. I use these fields to extract two samples:

— **a particle sample:** I extract a subsample of particles to study the temperature-density relation. For each particle the temperature is derived with the formula

$$k_B T = U \times (\gamma - 1) \times \mu M_{\rm H}, \tag{6.6}$$

where $\mu = 1/(X_{\rm H}(0.75 + N_e) + 0.25)$. γ is the adiabatic index (5/3 for monoatomic gas), $M_{\rm H}$ is the mass of an hydrogen atom, k_B is the Boltzmann constant and $X_{\rm H}$ is the hydrogen fraction by mass. Figure 6.15 illustrates typical temperature-density diagrams obtained from this particle sample.

— **a line of sight sample:** following the traditional procedure in one-dimensional flux power studies (Croft et al., 2002; Gnedin and Hamilton, 2002), I extract lines of sight (LOS) from the simulation cube choosing random origin and axis. For each pixel of each LOS, I derive density ρ, temperature T, peculiar velocity v and optical depth τ, all for H I only using the SPH equation:

$$A(\mathbf{r}) = \sum_j m_j \frac{A_j}{\rho_j} W\left(|\mathbf{r} - \mathbf{r}_j|, h_j\right) \tag{6.7}$$

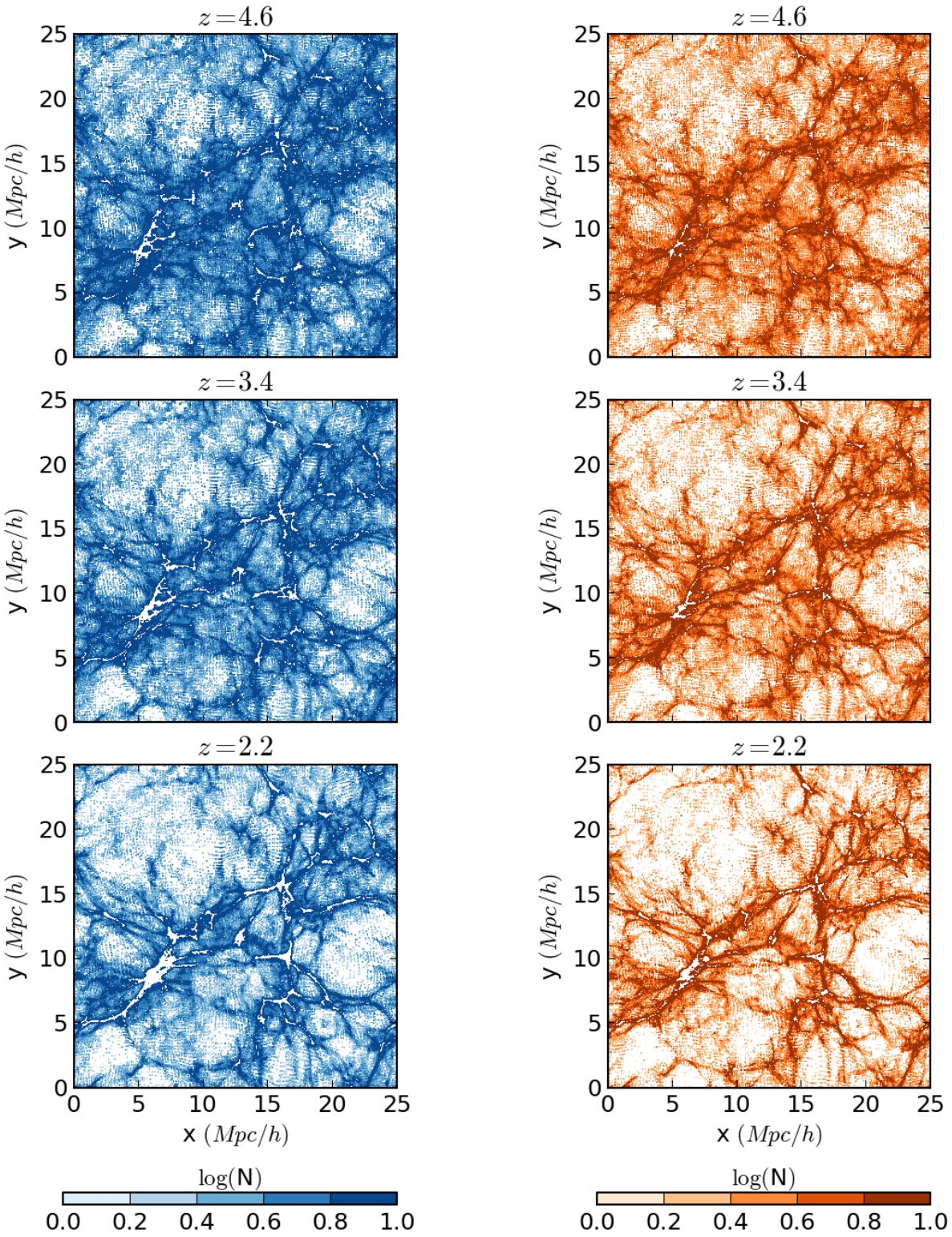
where A is a scalar quantity, $\mathbf{r}$ a position in the cube, h the smoothing length, and W a kernel function. The index j runs on all particles. The 3D cubic spline kernel described in section 5.3 is used:

$$W(q_j) = \frac{1}{\pi} \begin{cases} 1 + q_j^2(-\frac{3}{2} + \frac{3}{4}q_j) & |q_j| \leq 1 \\ \frac{1}{4}(2 - q_j)^3 & 1 < |q_j| \leq 2 \\ 0 & |q_j| \geq 2 \end{cases} \tag{6.8}$$

where $q_j = |\mathbf{r} - \mathbf{r_j}| / h_j$. These LOS are not mock spectra, in the sense that they do not match any properties (such as noise, resolution, metals absorption, ...) of observational data. The quantity I am particularly interested in is the optical depth for H I, from which I then compute the mean transmitted flux for each pixel.

6.3.5 Post-Processing

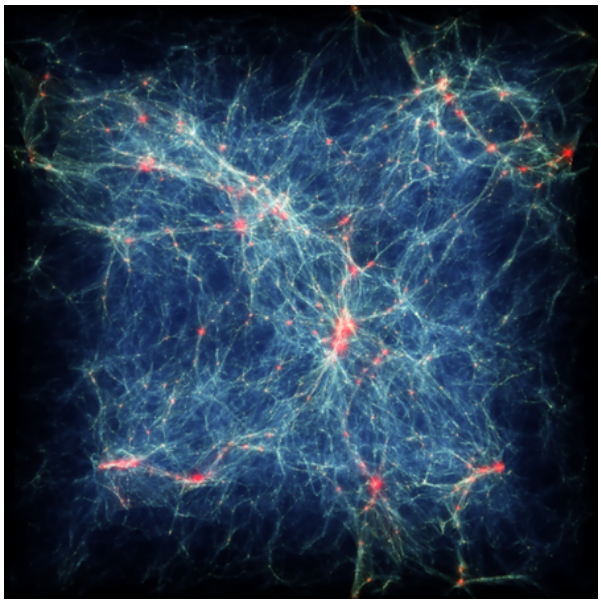
The post-processing stage allows us to extract two categories of results from our simulations. The first one is the large-box high resolution power spectrum that is derived by an appropriate combination of the power spectra from three lower-resolution or smaller box simulations, using the splicing technique described in section 6.5. The second one is a particle sample that is used



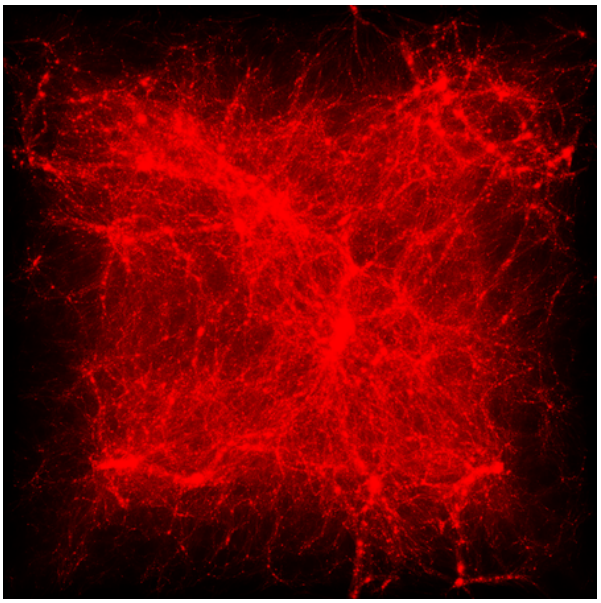
(a) Baryonic gas

(b) Dark matter

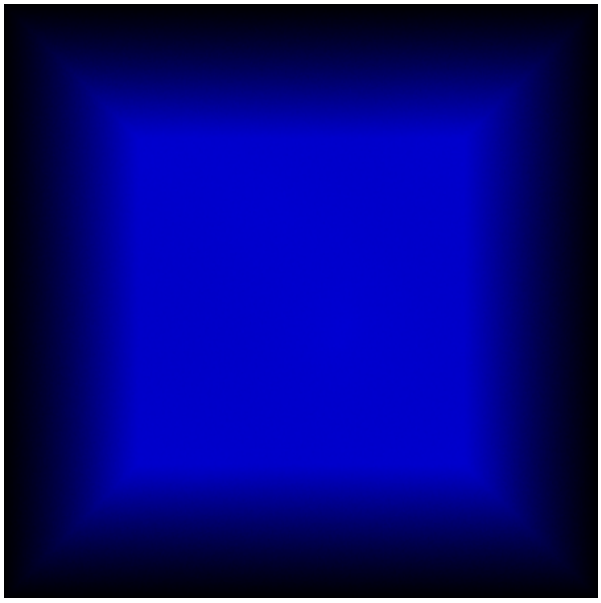
Figure 6.3 – Slices of baryon and dark matter snapshots ($2.5\,\mathrm{Mpc}\,h^{-1}$ depth), at three different redshifts, extracted from a simulation with 192^3 particles per type in a $(25\,\mathrm{Mpc}\,h^{-1})^3$ box. As expected, there are very few differences between the distributions for the two types of particles but dark matter is more clumped especially at low redshift. Colour represents particle number density. Out-of-scale densities (whether underflow or overflow) are white.



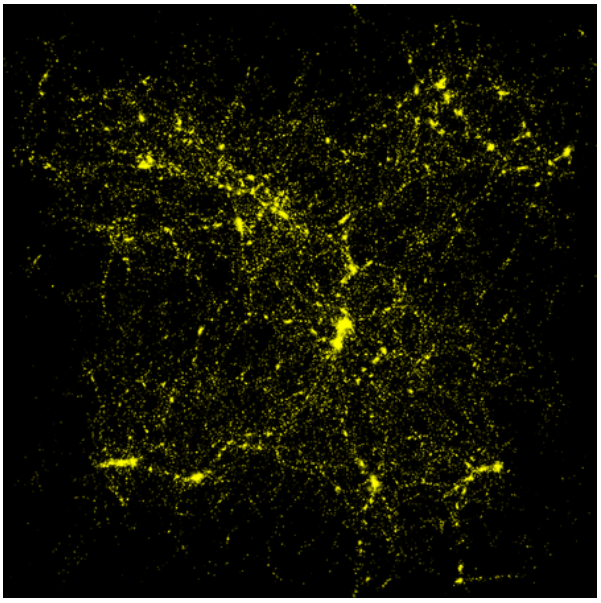
(a) Baryonic gas



(b) Dark matter



(c) Neutrinos



(d) Stars

Figure 6.4 – Visualisation using `splotch` of the four components that are evolved in a simulation with massive neutrinos. These pictures represent a box of $25\,(\mathrm{Mpc}/h)^3$ with 768^3 particles of each initial species at $z = 2.2$. It was run using our best-guess cosmology plus massive neutrinos with $\sum_\nu m_\nu = 0.8\,\mathrm{eV}$. For the baryonic gas, which is the only species that undergoes a SPH treatment, colour represents gas temperature (from blue to red) and density is mapped to intensity. Dark matter follows the same distribution as the gas, as expected. Even with our very simple model of star formation, stars are well located along the overdensities of gas. Since they are very light, neutrinos are not virialised and their clustering is not visible here (but it is present).

to derive the parameters $T_0(z)$ and $\gamma(z)$ in the IGM. This is performed by estimating the location of the most populated region of the diagram using the mode of the 2D distribution, for particles lying in the region defined by $\log(\delta) \in [-0.5, 0.0]$ and $\log(T/1\,\mathrm{K}) < 5.0$, with $\delta = \rho/\langle\rho\rangle$. Given the large tail of particles toward the high temperature regions where clusters of galaxies reside, in particular at low redshift, the mode was preferred to the mean since it is not affected by the precise choice of the (δ, T) bounds used to define the IGM. I estimate the mode by taking bins of 1000 K and computing the barycenter of the five highest bins. A linear fit is then performed using these points.

At this stage, I fix the photo-ionization rate (or equivalently the UV flux) by requiring the effective optical depth at each redshift to follow the empiracal power law

$$\tau_{eff}(z) = \tau_A(1+z)^{\tau_S} \tag{6.9}$$

where $\tau_A = 0.0025$ and $\tau_S = 3.7$ in agreement with observations (Meiksin, 2009). The rescaling coefficients, determined independently for each redshift using all the line of sight pixels, are typically between -20% and $+20\%$. We perform this normalization a posteriori since it is computationally much cheaper than finding and fixing the appropriate photo-ionization rate a priori for each of the simulations. as explained in Theuns (2005), however, this is justified by the fact that when the gas is highly ionized and in photo-ionization equilibrium, as is the case for the Lyman-α forest, the total heating rate per unit volume is independent of the amplitude of the UV flux. Gas dynamics can thus be considered not to be affected by the UV flux. The power spectrum is then computed from the scaled flux, and averaged over all line of sight.

6.4 Tests

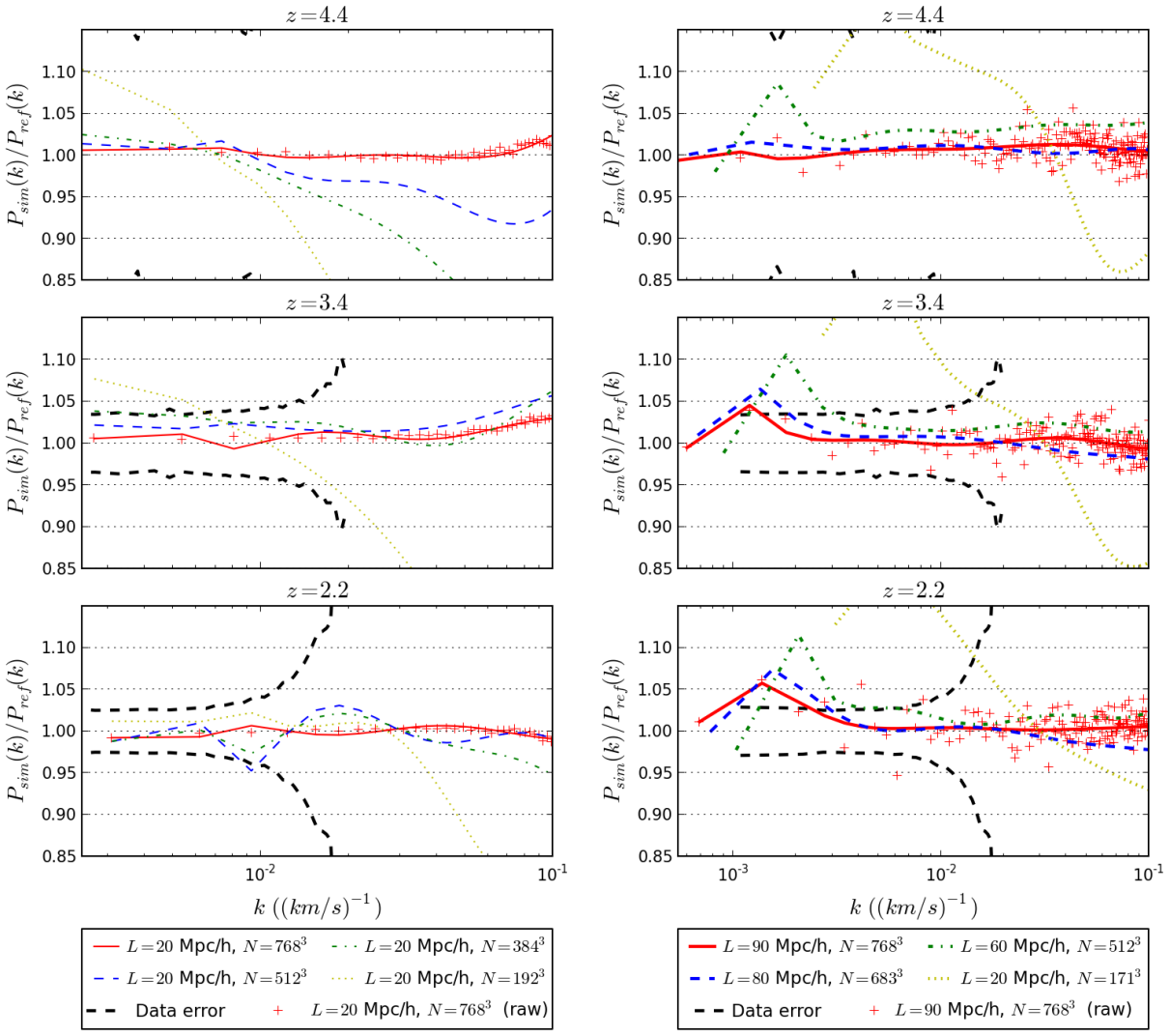
I describe below the convergence tests I performed to determine the mass resolution (or particle number) and box size needed to meet our requirements, that are based on the measurement errors on the power spectrum (see chapter 4), which set the convergence levels to be achieved in the simulations. I will also show the short study that was conducted on initial conditions. As for the grid, the initial conditions for all simulations are obtained with the same seed. The cosmology used in these tests has $(\Omega_m, \Omega_b h^2, h, \sigma_8, n_s) = (0.31, 0.021, 0.675, 0.83, 0.96)$. I ran two sets of simulations: the first set with simulations having the same box size L of $\sim 20\,\mathrm{Mpc}\,h^{-1}$ but changing the particle loading N^3 and therefore the mass resolution, the second with simulations having the same mass resolution but varying volumes, keeping L/N fixed at ~ 0.12. These two sets are listed in table 6.3 and the results at three different redshifts are presented in figures 6.5a and 6.5b. Hereafter, I will use the notation (L, N) to represent a simulation with N^3 particles for each species (dark matter and baryons) in a box of size $L\,\mathrm{Mpc}\,h^{-1}$ on a side.

These convergence tests are more stringent than what has been done before, justified by our aim to use our simulation suite for comparison to data of higher quality. For instance, to probe the effect of the box size, Viel et al. (2004) compared their simulations to a reference simulation with $(L, N) = (120, 200)$ and thus $L/N \sim 0.60$, and Bolton and Becker (2009) to a reference simulation $(80, 400)$ i.e., $L/N \sim 0.20$. This is to be compared to our L/N of 0.12. As regards the convergence on the mass-resolution, I explored a similar range of mass-resolutions as Bolton and Becker (2009) (in contrast, Viel et al. (2004) restricted to a minimum mass per particle three times larger), but using a $20\,\mathrm{Mpc}\,h^{-1}$ box instead of $10\,\mathrm{Mpc}\,h^{-1}$.

The most difficult redshifts at which to achieve convergence are those at $z > 3$, since the mean flux level becomes very small at such epochs and even under-dense regions, which are less

Mass-resolution test	Box-size test
(20,1024)	**(120,1024)**
(20,768)	(90,768)
(20,512)	(80,683)
(20,384)	(60,512)
(20,192)	(20,171)

Table 6.3 – The two sets of simulations used for convergence tests, with the reference simulation indicated in bold. (L, N) refers to a simulation with N^3 particles per species in a box of size $L\,\mathrm{Mpc}\,h^{-1}$ on a side.



(a) Mass-resolution tests, the reference simulation has $L = 20\,\mathrm{Mpc}\,h^{-1}$ and 2×1024^3 particles.

(b) Box-size tests, the reference simulation has $L = 120\,\mathrm{Mpc}\,h^{-1}$ and 2×1024^3 particles.

Figure 6.5 – Convergence tests for mass resolution and box size at three different redshifts. The curves are fifth order polynomial functions fitted to the data for better visibility. All values of the power spectrum ratio are shown for the (20,768) (left) and the (90,768) (right) cases. The dashed black curves illustrate $1\sigma_{\mathrm{stat}}$ uncertainties measured in BOSS data. Data uncertainties exceed plot boundaries at $z = 4.2$.

well sampled than average in an SPH framework, are producing absorptions. High redshift bins, however, are very important since gravitational collapse tends to suppress the differences in the linear-theory power spectra.These bins therefore highlight primordial differences between the matter power spectra resulting from different contributions of the various cosmological constituents. Some convergence problems can also arise at $z \sim 2$ due to the fact that strong systems, which are very non-linear, might be simulated inadequately due to cosmic variance or lack of resolution. Low-redshift bins are also those where the measurements from QSO spectra have the smallest statistical error bars, making the convergence criteria tighter.

6.4.1 Mass Resolution

I computed the ratio of the power spectra of each of the simulations listed in the first column of table 6.3 to the power spectrum of the (120,1024) simulation. The results presented in figure 6.5a show that an excess of power on large scales (small k) and a lack of power on small scales (large k) appear with decreasing resolution. As expected, this effect is stronger at higher redshift where the Lyman-α forest probes low density regions, which are less well resolved in the SPH treatment since it is the mass (and not the spatial) resolution that is kept fixed. Further details about this effect can be found in Bolton and Becker (2009).

The dashed curves in figure 6.5a illustrate the level of the $1\sigma_{\rm stat}$ statistical uncertainties observed in the BOSS analysis (Palanque-Delabrouille et al., 2013) at each redshift. At $z = 4.2$, the experimental uncertainties are larger than the maximum $\pm 15\%$ departure allowed on the plot and no longer appear.

Simulations with a mass resolution at least as good as for the $(20, 512)$ simulation all deviate by less than 2.5% from the highest mass-resolution power spectrum over our minimal k-range. This corresponds to a mean mass per particle of $M = 2.2 \times 10^5\,{\rm M}_\odot\,h^{-1}$. Extending to $k_{\rm max} = 0.1\,({\rm km/s})^{-1}$, the (20,512) simulation deviates by $\sim 10\%$ at the largest redshift.

6.4.2 Box Size

The results of figure 6.5b show that the box size has an effect on all scales, and not only on the large scales that approach the Nyquist limit. This is due to the non-linear coupling of modes during gravitational evolution, and to the fact that even on scales close to the box size, mass-fluctuations are not fully linear.

As before, the dashed curves in figure 6.5b illustrate the level of the $1\sigma_{\rm stat}$ data uncertainties at each redshift. To reach $k_{\rm min} = 1 \times 10^{-3}\,({\rm km/s})^{-1}$, we see that we need a box size of at least $90\,{\rm Mpc}\,h^{-1}$. The most significant constraint comes from the largest scales that cannot be probed (or not with adequate precision) otherwise.

6.4.3 Summary of Convergence Requirements

In conclusion, the ideal simulation for our study should use a $\sim\ 100\,{\rm Mpc}\,h^{-1}$ box and a mass resolution roughly equivalent to a $(20, 614)$ simulation, which translates into 3072^3 particles of each species. The mean mass of a gas particle is then $M = 1.2 \times 10^5\,{\rm M}_\odot\,h^{-1}$.

Although convergence tests are specific to each problem and each statistical property for which convergence is sought, we can briefly compare our conclusions to the results obtained by other studies. To infer the dark matter power spectrum from the Lyman-α forest in high resolution QSO absorption spectra covering $0.003 < k < 0.03\,({\rm km/s})^{-1}$, Viel et al. (2004) chose

Starting redshift	Approximation
30	Zel'dovich
30	2LPT
99	Zel'dovich
99	2LPT

Table 6.4 – The four simulations used for initial conditions tests, all with $L = 100\,\mathrm{Mpc}\,h^{-1}$ and $N = 768^3$. The Zel'dovich approximation is the first order approximation corresponding to the second order Lagrangian perturbation theory (2LPT) approximation.

a $(60, 400)$ simulation, i.e. a mass per gas particle of $\sim 4 \times 10^7\,\mathrm{M}_\odot\,h^{-1}$. To resolve the high redshift Lyman-α forest in smoothed particle hydrodynamics simulations, a problem similar to our own, Bolton and Becker (2009) found that a box size of at least $40\,\mathrm{Mpc}\,h^{-1}$ is preferable at all redshifts. They also found that while a mean gas particle mass $M_{\mathrm{gas}} \leq 1.6 \times 10^6\,\mathrm{M}_\odot\,h^{-1}$ is required at $z = 2$, a mass resolution at least eight times better is needed at $z = 5$, i.e. $M_{\mathrm{gas}} \leq 2 \times 10^5\,\mathrm{M}_\odot\,h^{-1}$. Our requirements are thus more stringent than those selected in the past, both in terms of box size and mass resolution.

Several tens of such simulations, as needed to compute our grid of cosmological simulations, would require several tens of millions of hours to be run, which is not an acceptable computational time. I address and solve this issue with the splicing technique presented in section 6.5.

6.4.4 Initial Conditions

To compare the effect of different initial conditions in terms of starting redshift and order of the approximation, I ran four simulations with $L = 100\,\mathrm{Mpc}\,h^{-1}$ and $N = 768^3$ using our best guess cosmology. These simulations are listed in table 6.4. Initial conditions using a first order approximation (Zel'dovich) are produced with `NGENIC` and second order ones are produced using `2LPTIC`.

Using these simulations, four comparisons are interesting to look at.

— Comparison of the power spectra obtained with initial conditions from both `NGENIC` and `2LPTIC` produced at $z = 99$ is given in figure 6.6. There is almost no difference between the two approximations at this redshift. However, the inclusion of neutrinos as a new type of particle forces us to start at the lowest redshift possible (Rossi et al., 2014).

— Comparison of the power spectra obtained with initial conditions from both `NGENIC` and `2LPTIC` produced at $z = 30$ is given in figure 6.7. We see that first order initial conditions lead to an excess of power of the order of a few percent, justifying the use of second order initial conditions. The effect is stronger at higher redshift since non-linear evolution tends to suppress differences in the linear-theory spectra.

— Comparison of the power spectra obtained with initial conditions from `2LPTIC` at $z = 99$ and $z = 30$ is given in figure 6.8. This last comparison shows that the error coming from low starting redshift is not trivial nor negligible and must be taken into account when using the power spectra.

— Comparison of the power spectra obtained with initial conditions from both `NGENIC` at $z = 99$ and $z = 30$ is given in figure 6.9. The effect of the low starting redshift is stronger than in figure 6.8, thus justifying the use of second-order approximation.

Although the differences between the power spectra derived from `NGENIC` or `2LPTIC` initial conditions may look big, they do not exceed 5% in the domain restricted to the k-range of our

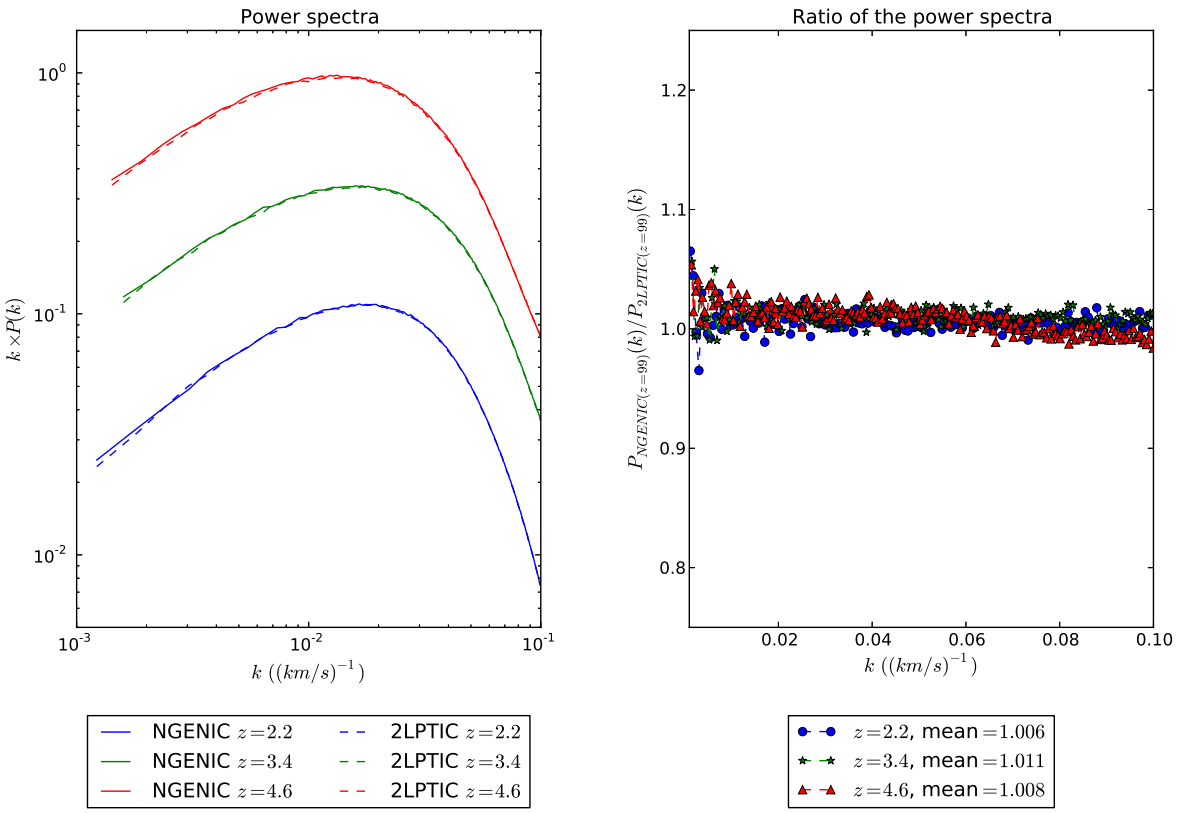


Figure 6.6 – Comparison of power spectra from initial conditions generated with `NGENIC` and `2LPTIC` at $z = 99$

study ($k = 0.001 - 0.02\,(\mathrm{km/s})^{-1}$). In addition, the error is strongly diminishing with redshift. Extrapolating to $z = 0$ yields a difference in good agreement with what L'Huillier et al. (2014) have obtained for a similar comparison at $z = 0$ (with starting redshifts of 100, 50 and 23).

6.5 Splicing

In the previous section I have estimated that simulating a flux power spectrum covering the range $k = 1 \times 10^{-3}\,(\mathrm{km/s})^{-1}$ to $k = 2 \times 10^{-2}\,(\mathrm{km/s})^{-1}$) with a unique simulation at sufficient precision for every redshift in the range $2.2 < z < 4.6$ requires $N = 3072^3$ particles of each species in a box of size $L = 100\,\mathrm{Mpc}\,h^{-1}$. To obtain power spectra of equivalent resolution and box size in a reasonable computational time, I use the technique described in McDonald (2003). In this method, competing demands of large box size and high resolution are solved by splicing together the power spectra from pairs of large and small box simulations, using $L = 100\,\mathrm{Mpc}\,h^{-1}$ for the large-scale power, and $L = 25\,\mathrm{Mpc}\,h^{-1}$ for the small-scale power, both with $N = 768^3$. One must then correct the large box size simulation for the lack of resolution, and the small box size for the lack of non-linear coupling between the highest and the lowest k-modes. The corrections are computed using a transition $(25, 192)$ simulation that has the same resolution as a $(100, 768)$ and the same box size as a $(25, 768)$.

One needs to distinguish three regimes when computing the full power spectra:

— $\mathbf{k < k_{min,25}}$, where $k_{min,25} = 2\pi/25\,\mathrm{Mpc}\,h^{-1}$ is the minimum k present in a box of $25\,\mathrm{Mpc}\,h^{-1}$ side. The spliced flux power P_F is the flux power $P_{F,100,768}$ of the $(100, 768)$ simulation here taken as our reference, corrected for its low-resolution by a k-independent

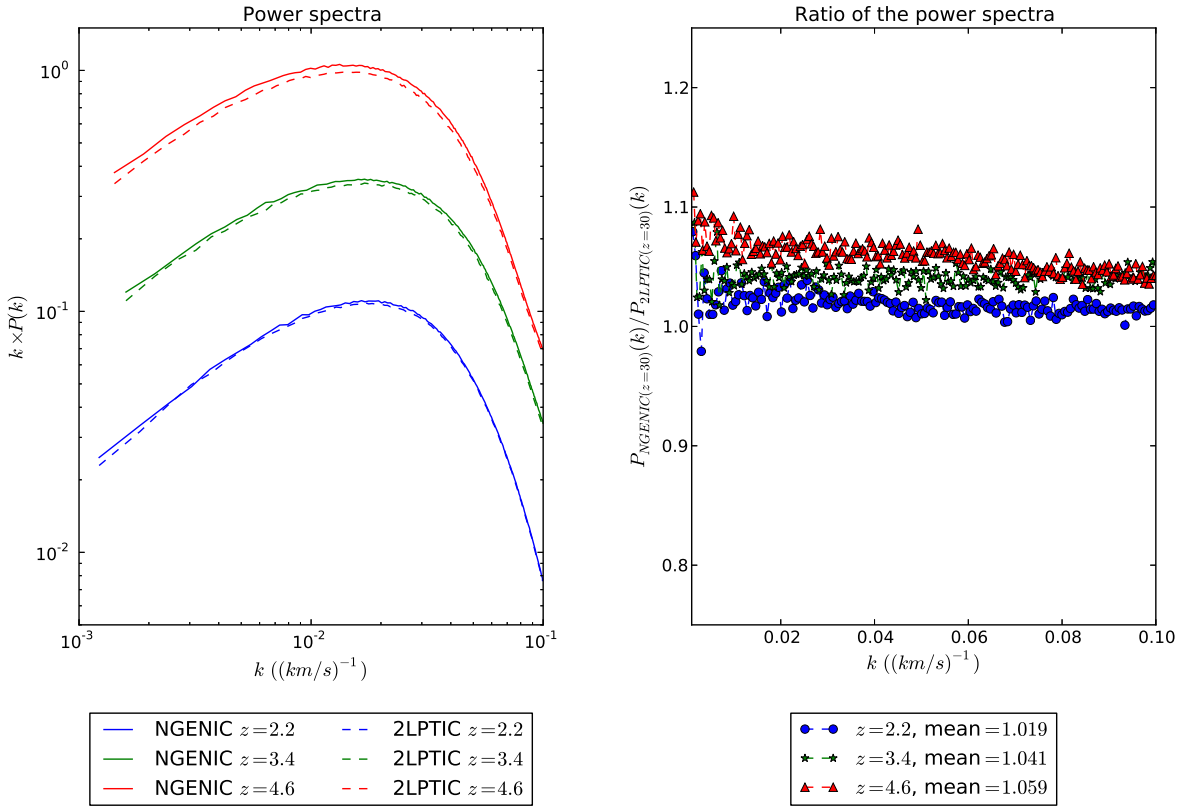


Figure 6.7 – Comparison of power spectra from initial conditions generated with `NGENIC` and `2LPTIC` at $z = 30$

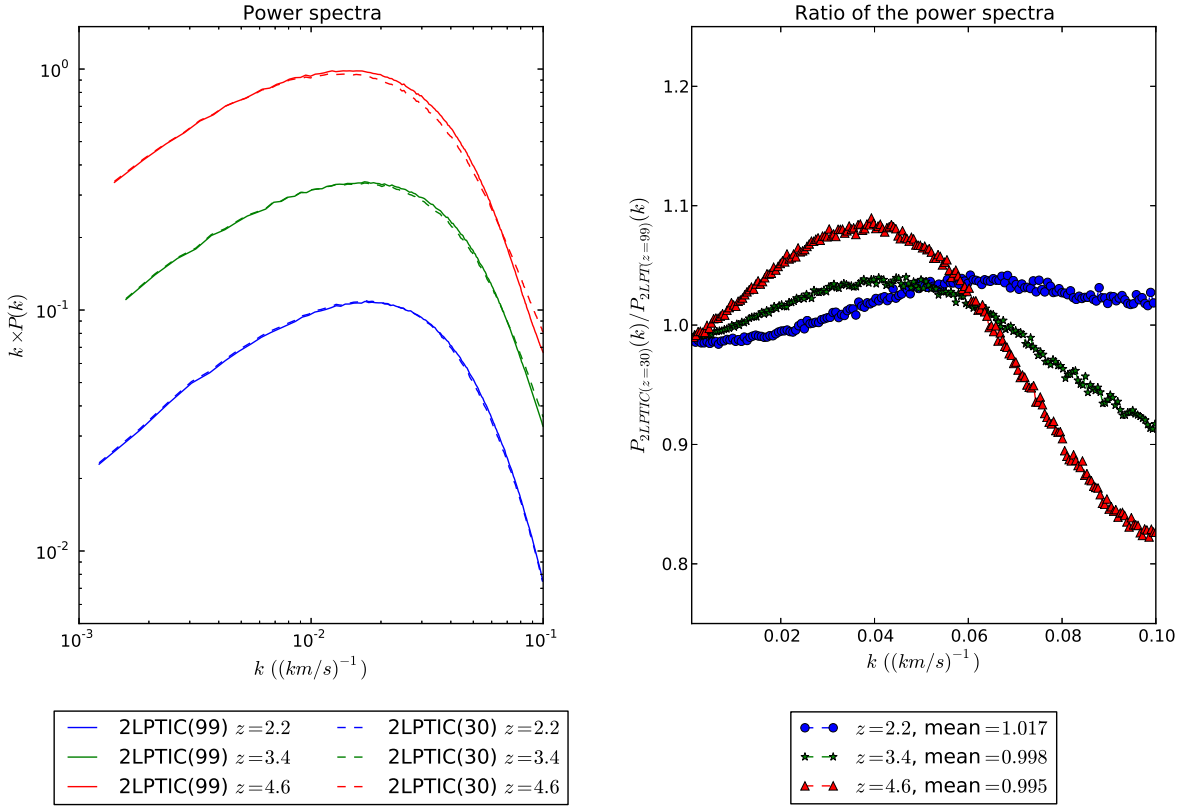


Figure 6.8 – Comparison of power spectra from initial conditions generated with `2LPTIC` at $z = 30$ and $z = 99$

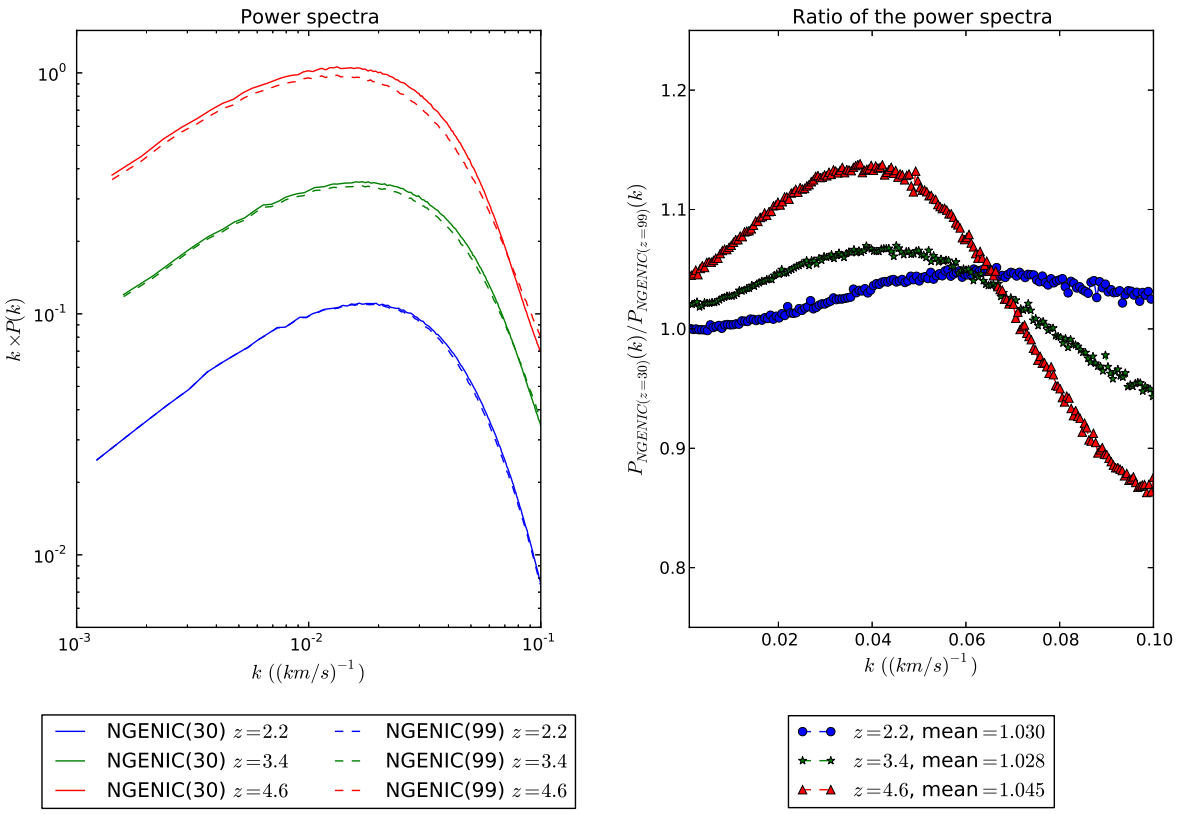


Figure 6.9 – Comparison of power spectra from initial conditions generated with `NGENIC` at $z = 30$ and $z = 99$

factor evaluated at $k_{min,25}$:

$$P_F(k) = P_{F,100,768}(k) \times \frac{P_{F,25,768}(k_{min,25})}{P_{F,25,192}(k_{min,25})}. \tag{6.10}$$

The possibility of using a constant factor for the largest k-modes has been tested in McDonald (2003).

— $\mathbf{k_{min,25} < k < k_{Nyq,100}/4}$, where $k_{Nyq,100} = 768\pi/100\,\mathrm{Mpc}\,h^{-1}$ is the Nyquist wave number of the large box. In this regime I use a similar correcting ratio, but taken at the wave number k at which the flux power is calculated:

$$P_F(k) = P_{F,100,768}(k) \times \frac{P_{F,25,768}(k)}{P_{F,25,192}(k)}. \tag{6.11}$$

This is mathematically equivalent to considering the high-resolution simulation $(25, 768)$ as our reference, and correcting it for its small box size.

— $\mathbf{k > k_{Nyq,100}/4}$. At these large k-modes, the resolution correction is no longer a small factor. I thus take the $(25, 768)$ simulation as our reference, and correct its limited box size by a k-independent factor evaluated at the fixed splicing point $k = k_{Nyq,100}/4$:

$$P_F(k) = P_{F,25,768}(k) \times \frac{P_{F,100,768}(k_{Nyq,100}/4)}{P_{F,25,192}(k_{Nyq,100}/4)}\,. \tag{6.12}$$

The splicing technique is applied for each redshift at which I compute the power spectrum. I illustrate the method and its accuracy at two different redshifts on figure 6.10, using a set of

smaller-box size simulations to enhance the contrast between the different power spectra, as well as to allow the comparison to a full resolution run (labelled "exact" on the figure) with 1024^3 particles of each species in a $(100\,\mathrm{Mpc}\,h^{-1})^3$ box. For this illustration, the large box-size, the large resolution and the transition simulations are $(100, 256)$, $(25, 256)$ and $(25, 64)$ simulations respectively. The spliced power spectra obtained at $z = 2.2$ and $z = 4.6$ are presented on the left of figure 6.10, along with the exact power spectrum and the individual runs entering the splicing estimate. The correction coefficient with respect to the reference power spectrum in each regime of k-modes is shown on the middle plots of figure 6.10. In the intermediate regime, the resolution correction increases towards smaller scales, reaching almost 50% for the set of simulations illustrated here. It shows less scale-dependence when taken as a box size correction to the large-resolution power spectrum. The error resulting from the splicing, i.e. the ratio between the spliced power spectrum and the "exact" power spectrum, is shown on the right plots of figure 6.10. Although the correction factors show discontinuities and large values at the boundary where the simulation chosen as reference changes, the spliced power spectrum is continuous by construction and the splicing errors stay well contained. In the large mode regime, at $k > k_{Nyq,100}/4$, it is unclear wether a constant box size correction or even any correction at all is indeed the optimal combination, since both the correction factor and the residuals are at the same level of about 0.95. this regime, however, is only probed by the medium resolution SDSS-III/BOSS data in the highest redshift bins where measurements uncertainties significantly exceed the splicing errors. Its optimization is thus beyond scope of this work.

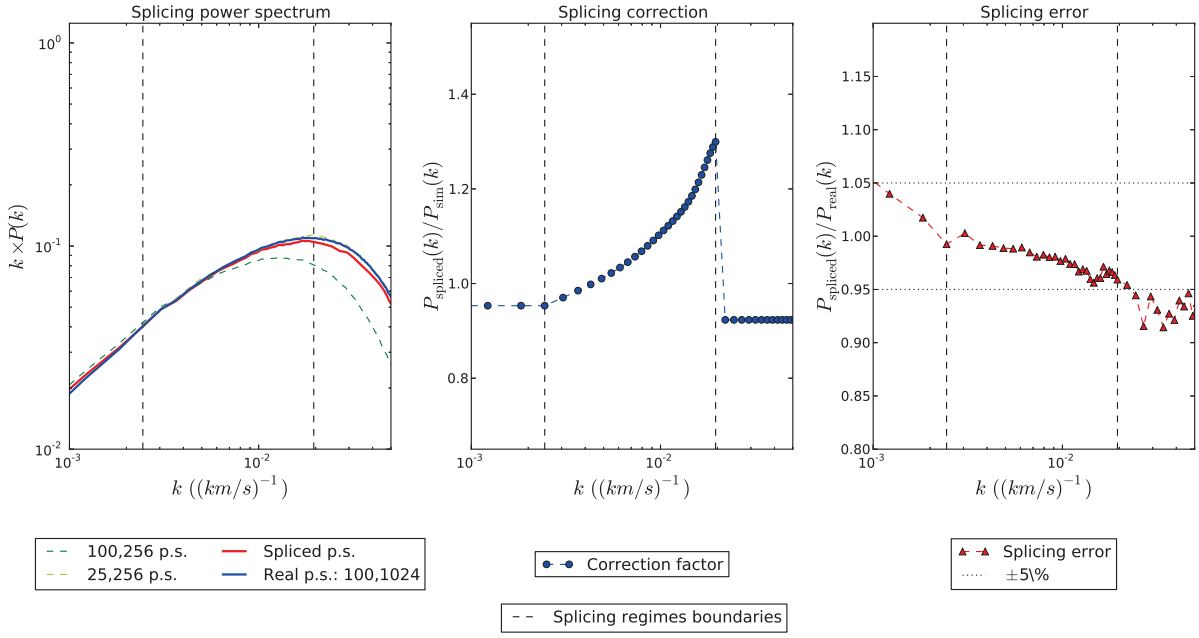
For the box size and resolution chosen for our simulation suite, the last regime begins at $k = 5.3 \times 10^{-2}\,(\mathrm{km/s})^{-1}$ for $z = 3.0$, which is beyond the maximum mode that can be reached with BOSS or eBOSS data. The maximum correction factor, obtained for $k = 2 \times 10^{-2}\,(\mathrm{km/s})^{-1}$, is thus smaller than in the previous illustration. It ranges from 5% at $z = 2.2$ to 22% at $z = 4.6$.

I estimate the accuracy of the technique from the splicing residuals, defined as the ratio of the spliced to the exact power spectrum. The splicing residuals show no clear dependence on redshift. In figure 6.11, they are plotted for $z = 2.2$ and $z = 4.2$, along with the statistical uncertainty at the same redshifts obtained in the most recent BOSS analysis (Palanque-Delabrouille et al., 2013). Over the k-range of interest for BOSS data, the residuals have an average of -0.98 with a rms of 0.01. The largest excess is seen near $k = 1 \times 10^{-3}\,(\mathrm{km/s})^{-1}$. A simulation with a larger box size would be needed to reduce the splicing residuals further. For the purpose of this study, the splicing technique is accurate at the 2% level over the entire k-range of interest.

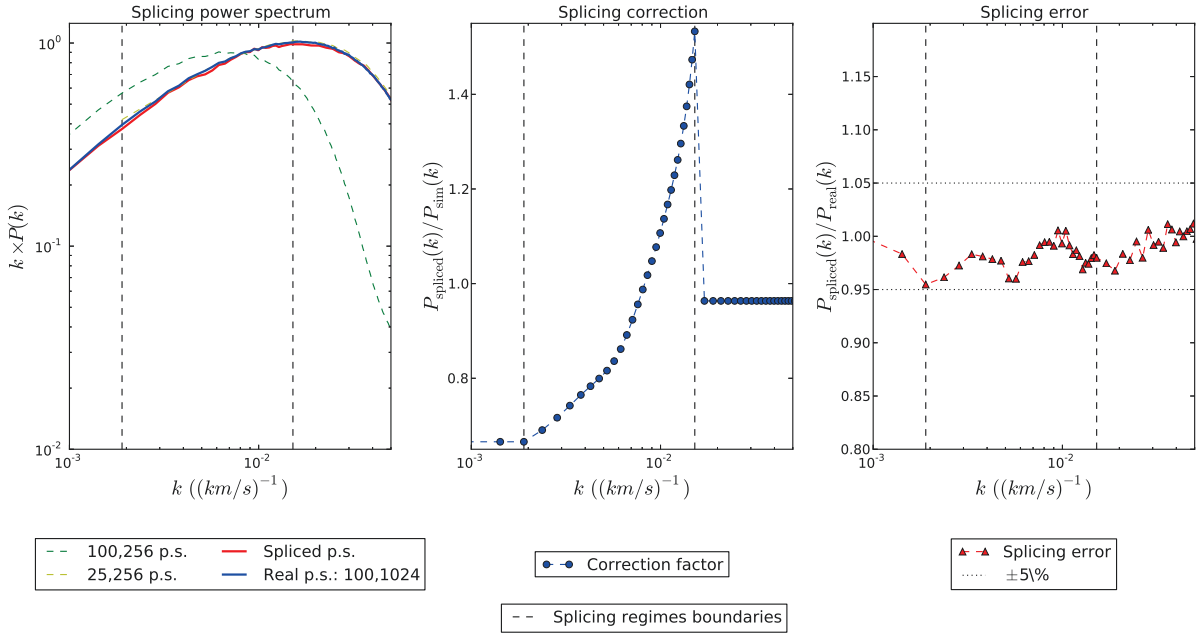
6.6 Simulation Checks

Several checks were performed to validate our simulations. I first verified that the power spectrum of independent simulations obtained either with different cosmological parameters or different seeds is consistent with the power spectrum derived from the Taylor expansion of section 6. Finally, I discuss some characteristics of our simulations. In particular, I describe the effect on the flux power spectrum of some of the parameters we have varied, and I show the $T - \rho$ diagrams from which I derive the two parameters $T_0(z)$ and $\gamma(z)$ that describe the IGM.

I also test that the simulation uncertainties were artficially limited by an oversampling of the simulated volume: this was done by considering different numbers n of lines of sight (from 5000 to 100 000). For each redshift and mode, the power spectrum value is taken as the mean over the n lines of sight and the uncertainty on the mean as the rms of the distribution divided



(a) Splicing at $z = 2.2$



(b) Splicing at $z = 4.6$

Figure 6.10 – Illustration of the splicing technique at the two extreme redshifts with simulations of either 64^3 or 256^3 particles of each species in a volume of $(25\,\mathrm{Mpc}\,h^{-1})^3$ or $(100\,\mathrm{Mpc}\,h^{-1})^3$. Dashed vertical lines illustrate the regime boundaries. Left: power spectra of the two basis simulations $(100, 256)$ and $(25, 256)$, the spliced power spectrum and the "exact" $(100, 1024)$ power spectrum. Middle: correction factors computed with respect to the power spectrum chosen as reference in each regime. Right: splicing error (residuals). Although the correction presents a discontinuity, the spliced power spectrum is continuous by construction.

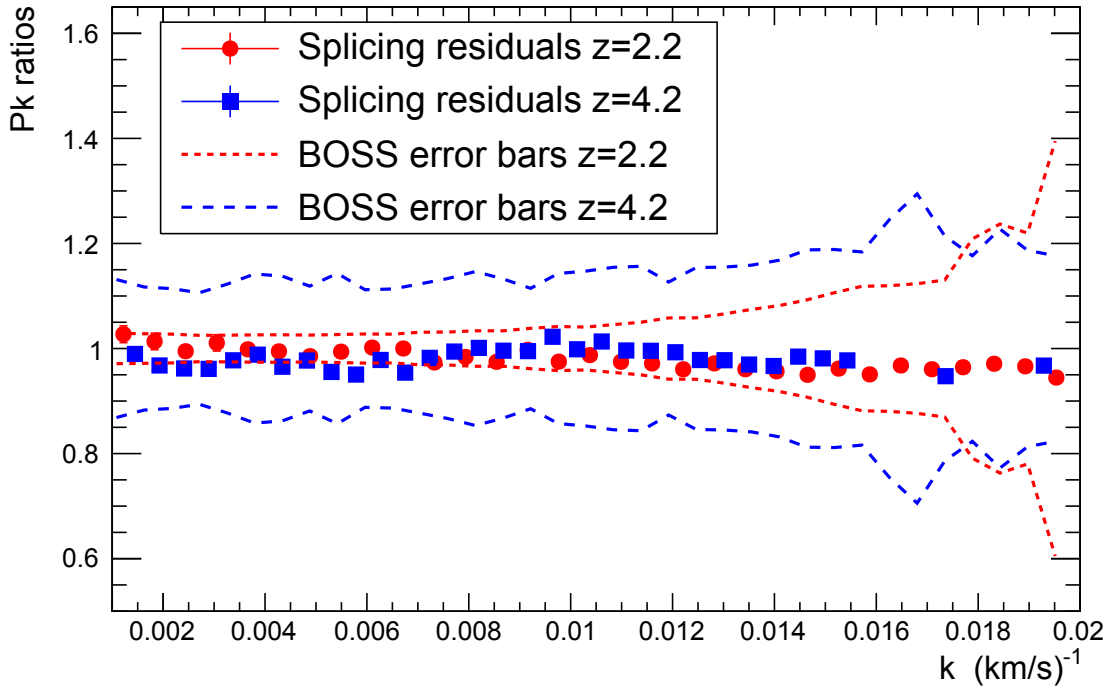


Figure 6.11 – Residuals of the spliced to the exact power spectrum for $z = 2.2$ and $z = 4.2$. The dashed curves illustrate the level of statistical uncertainties in current data ($1\sigma_{\rm stat}$).

by $\sqrt{n}$. I checked that the uncertainty on each point of the power spectrum did scale as the square root of the number of lines of sight used to compute it, at better than the percent level.

6.6.1 Checks with Independent Simulations

I validate our second-order Taylor expansion by comparing its prediction to the simulated power spectrum for simulations other than those used in the grid. For each simulation, I extract ten sub-samples of 10 000 lines of sight each with random origin and axis. For each redshift and mode, the power spectrum value is taken as the mean over the 100 000 lines of sight and the uncertainty on the mean as the *rms* of the distribution divided by $\sqrt{10}$.

I produced a new simulation with the same parameters as our central simulation but using a different random seed to compute the initial conditions. Snapshots of the resulting gas distribution in the two cases are shown in figure 6.12. The derived power spectra for the two seeds are in excellent agreement on low scales. On the largest scales, the two power spectra can differ by up to 2 to 3σ at all redshifts, indicating a cosmic variance contribution to the uncertainty on the simulated power spectrum since the simulation box has the size of the largest modes measured. On smaller scales however, the power spectra from either seed are in excellent agreement. The cosmological fit on the second seed simulation using a Taylor expansion built with the first seed is shown in table 6.5. The values do not differ by more than 1σ between the two seeds.

I also ran additional simulations conserving the same seed but with different cosmological and astrophysical parameters. For each sub-sample, I compute the power spectra corresponding to the twelve redshifts in the range $z = [2.2 - 4.6]$. We then perform a simple fit of the six parameters (n_s, σ_8, Ω_m, H_0, `AMPL`, `GRAD`) using our second-order Taylor expansion as model. The methodology of the fit is described in section 6.7.

Parameter	Input value	Fitted value
n_s	0.96	0.969 ± 0.004
σ_8	0.83	0.839 ± 0.005
Ω_m	0.31	0.28 ± 0.01
H_0	67.5	65.0 ± 1.0
$\gamma(z = 3)$.	1.3	1.38 ± 0.03
$T_0(z = 3)$	14000	13750 ± 1000

Table 6.5 – Comparison of the input parameters and the fitted parameters for a different seed in the simulation.

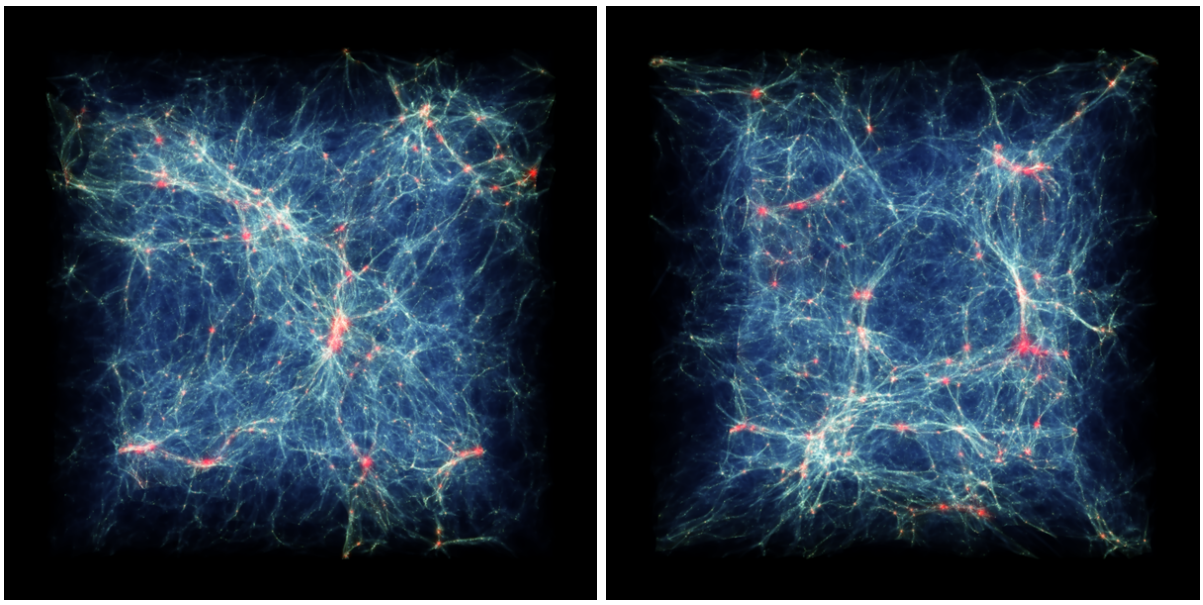


Figure 6.12 – Visualisation using `splotch` of the baryonic gas from a `GADGET-3` snapshot taken at $z = 2.2$ for two simulations run with identical parameters but different random seeds to compute the initial conditions. Both simulations are using 768^3 particles per species in a $(25\,\mathrm{Mpc}\,h^{-1})^3$ box. Colour represents gas temperature (from blue to red) and density is mapped to intensity. Left for the random seed used for the grid, right for a different random seed.

The results are summarized in table 6.6. The last column shows the average of the fitted values over the ten sub-samples. The uncertainty on the fitted value is estimated as the *rms* of the distribution over the ten sub-samples divided by $\sqrt{10}$. These four configurations probe different regions of the parameter phase space. They give results in good agreement with the input cosmological and astrophysical parameters. The level of accuracy achieved with these validation tests is three to five times better than the errors we expect on these parameters from a fit to data given the uncertainties of Palanque-Delabrouille et al. (2013). These validations demonstrate the robustness of our second-order Taylor expansion over the entire parameter space covered by my grid simulations.

Finally, we also test our neutrino implementation as a new type of particle by running a simulation $\Sigma_\nu m_\nu = 0.01\,\mathrm{eV}$. We found it fully compatible with our central model where there are no "neutrino particles", but which account for massless neutrinos. More details about this test can be found in Rossi et al. (2014).

Parameter	Input value	Fitted value
n_s	0.93	0.933 ± 0.002
σ_8	0.85	0.839 ± 0.008
Ω_m	0.31	0.311 ± 0.003
H_0	67.5	66.0 ± 1.0
$\gamma(z=3)$.	1.3	1.30 ± 0.03
$T_0(z=3)$	14000	14800 ± 600
n_s	0.96	0.962 ± 0.002
σ_8	0.83	0.827 ± 0.009
Ω_m	0.31	0.313 ± 0.003
H_0	67.5	66.5 ± 1.0
$\gamma(z=3)$.	1.47	1.48 ± 0.03
$T_0(z=3)$	10000	10250 ± 400
n_s	0.93	0.930 ± 0.002
σ_8	0.86	0.845 ± 0.007
Ω_m	0.30	0.300 ± 0.003
H_0	66.0	66.5 ± 1.0
$\gamma(z=3)$.	1.16	1.18 ± 0.04
$T_0(z=3)$	10000	11000 ± 600

Table 6.6 – Comparison of the input parameters and the fitted parameters for three different sets of parameters conserving the same seed in the simulations.

6.6.2 Power Spectrum

Figure 6.13 illustrates the impact on the power spectrum of our four cosmological parameters. I compare the power spectrum computed from our best-guess model to the one obtained when varying each parameter, one at a time. I note that the dependence on the value of the four parameters is as expected according to their physical meaning. I briefly explain the different behaviours below.

The spectral index n_s represents the evolution of the primordial density fluctuations with respect to k through $\mathcal{P}(k) \propto k^{n_s-1}$. A larger n_s therefore increases the power at large k, as seen in the top left panel of figure 6.13.

The parameter σ_8 measures the *rms* amplitude of the linear matter density fluctuations today in spheres of size $8\,\mathrm{Mpc}\,h^{-1}$, and thus determines the normalization of the matter power spectrum. In a first approximation, increasing the value of σ_8 therefore increases the power spectrum on all scales, as shown in the top right panel of figure 6.13. A slightly larger effect, however, is seen on large scales, since an excess in the amplitude of the fluctuations will favour the merging of small scale fluctuations, thus enhancing the power on larger scales. This slight trend is purely non-linear and not expected in the evolution with σ_8 of the linear power-spectrum.

The present-day Hubble constant H_0 (in units of velocity/distance) allows the conversion from distance-space to k-space (units of inverse velocity). Therefore, if H_0 is increased, a given distance will correspond to a higher k, thus leading to an increase of power since the power spectrum, which is a decreasing function of k, is shifted to the right. This is indeed what is observed in the lower left panel of figure 6.13.

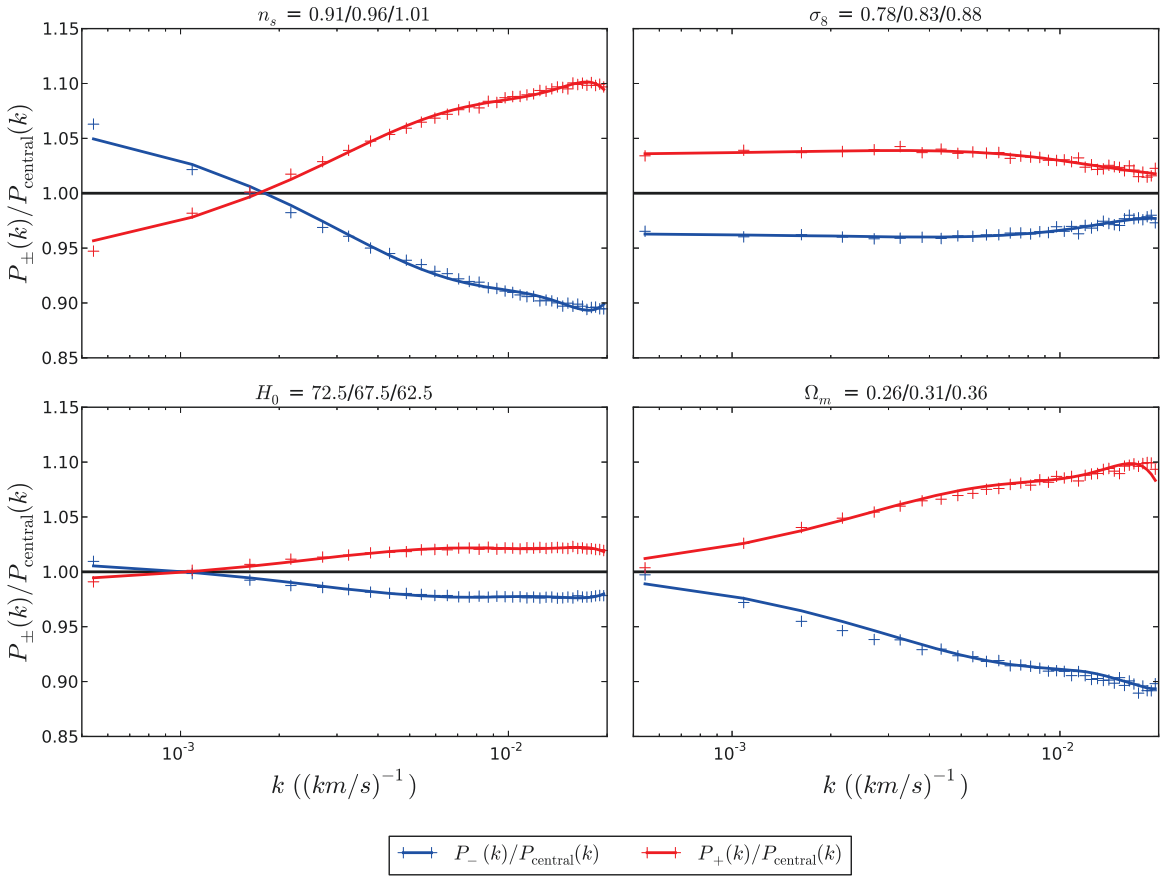


Figure 6.13 – Effect of the parameters n_s, σ_8, H_0 and Ω_m on the power spectrum (limited to the k-range of our study) at $z = 3.2$. $P_+(k)$ and $P_-(k)$ refer to the power spectra extracted from the simulations using the upper and lower limit on each considered parameter respectively. The fit to the points is a fifth order polynomial function and is here only for better visibility.

The parameter Ω_m quantifies the fraction of matter density in a flat Universe. Because Ω_m and the dark energy density Ω_Λ vary in opposite directions, a higher Ω_m delays the onset of dark energy domination, thus increasing the time available for structure formation. In addition, in a larger Ω_m universe, more structures (in particular small ones that would not collapse otherwise) will be formed, leading to an increase of the power spectrum, especially at high k. This is in agreement with the plots in the lower right panel of figure 6.13.

Finally, figure 6.14 shows the effect of different masses of neutrinos on the one-dimensional power spectra. I use simulations normalized to the same mean flux and with the same σ_8 at $z = 0$, which means that the absolute level of the ratios has no meaning here since σ_8 and $\Sigma_\nu m_\nu$ are strongly correlated. As expected, the larger the mass, the bigger the effect on the power spectrum. Because having massive neutrinos basically consists in adding additional matter, the overall effect is an increase of the power spectrum. The effect is stronger on large scales since on small scales neutrinos, which are almost relativistic and interact only through gravity, tend to suppress fluctuations. In other words, if the power spectra for the different neutrino masses were normalized to the same power on large scales (as done for instance by constraining its value to a large-scale measurement such as done with CMB data), then the impact of higher neutrino mass is a higher suppression of the power spectrum on small scales.

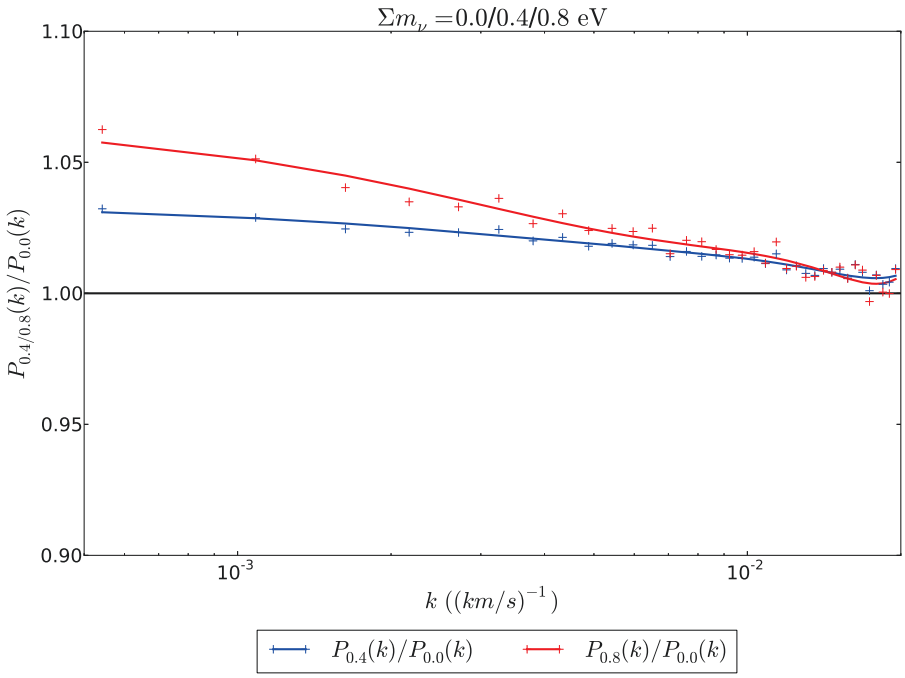


Figure 6.14 – Effect of the parameter $\Sigma_\nu m_\nu$ on the power spectrum (limited to the k-range of our study) at $z = 3.2$. $P_{0.4}(k)$ and $P_{0.8}(k)$ refer to the power spectra extracted from the simulations using neutrinos with a total mass of 0.4 eV and 0.8 eV respectively. $P_{0.0}(k)$ is the power spectrum of our central model, which takes into account massless neutrinos but does not include them as a third species of particle. The fit to the points is a fifth order polynomial function and is here only for better visibility.

6.6.3 Density-Temperature Relation

In figure 6.15, I present the $T - \rho$ diagrams obtained from our central simulation at each of the snapshot redshifts. It is possible to distinguish three different populations – the IGM, the stars, and the clusters – with a clear evolution with redshift for each of them.

The IGM is described by the low density and low temperature particles. This is the regime that dominates at high redshift. At later times, however, fewer and fewer particles reside in this part of the $T - \rho$ diagram, since they are captured by collapsing over-densities. I use this region to extract the $T_0(z)$ and $\gamma(z)$ parameters, displayed in figure 6.16, where they are compared to the measurements of Becker et al. (2011).

The particles with higher temperature correspond to clusters and galactic gas. As expected, their density increases as structures are formed in the simulation box. They therefore become more prominent at lower redshifts.

In our simulations dedicated to the study of the IGM through the Lyman-α forest measurements, star formation undergoes a simplified treatment, which reflects as the sharp cut-off at $\log(\delta) \simeq 3$. Any particle sufficiently dense and cool is transformed into a star particle. The latter is used for gravity force calculation, but does not undergo SPH treatment like baryonic gas does.

6.7 Cosmological Application

I present here a preliminary cosmological interpretation of our results. The goal is to show what can be excepted from our method. The current results do not take into account any bias

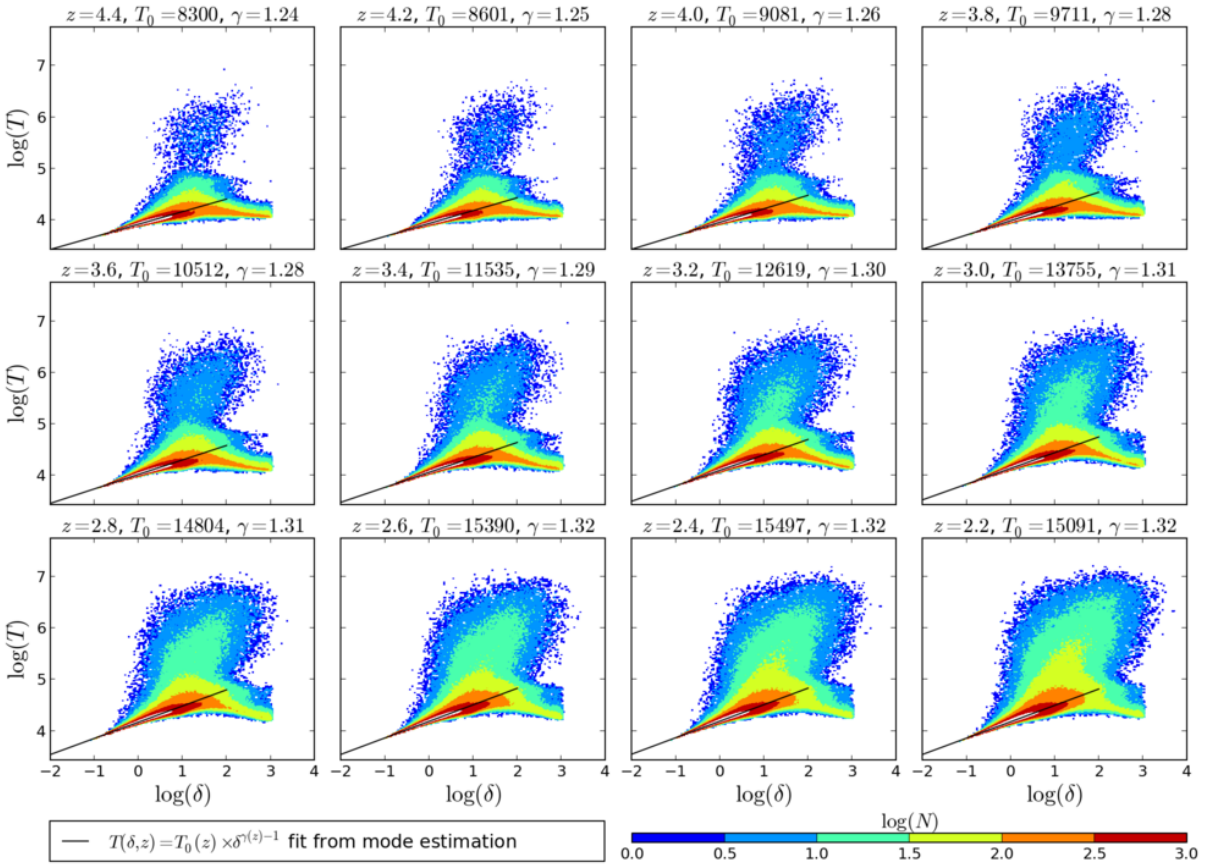


Figure 6.15 – Temperature-density diagrams at various redshifts. Colour represents the particle density in logarithmic scale. The black line represents the fitted $T - \rho$ relation from several mode-estimated points. δ is the normalized density $\rho / \langle\rho\rangle$.

like those induced by the splicing or the low starting redshift. I here use only twelve redshift bins out of the thirteen produced by the hydrodynamical simulations because the current data do not allow a measurement in the highest bin $\langle z \rangle = 4.6$. We also compare and combine our results with the cosmological results of Planck (Planck Collaboration et al., 2013).

6.7.1 Direct Comparison to SDSS-III/BOSS DR9 Data

In chapter 4, the one-dimensional Lyman-α forest power spectrum is measured with 13 821 quasar spectra from SDSS-III/BOSS DR9 selected on the basis of their high quality, large signal-to-noise ratio, and good spectral resolution. The power spectra are measured over twelve redshift bins from $\langle z \rangle = 2.2$ to $\langle z \rangle = 4.4$, and scales from $0.001\,(\mathrm{km/s})^{-1}$ to $0.02\,(\mathrm{km/s})^{-1}$ (see figure 6.17).

In order to compare the measured power spectrum to the one obtained for our central model simulation, I normalized the simulation power spectrum at each redshift by constraining the effective optical depth to follow a power law evolution, $\tau_{eff}(z) = \tau_A \times (1+z)^{\tau_S}$, where $\tau_A = 0.0025$ and $\tau_S = 3.7$. To account for the effect of the correlated Si III absorption, we correct the simulated power spectrum by a multiplicative term, $1 + a^2 + 2a\cos(vk)$ with $a = f_{\mathrm{Si\,III}}/(1 - \langle F \rangle(z))$ following the suggestion of McDonald et al. (2006). The parameter $f_{\mathrm{Si\,III}}$ is adjusted and v is fixed at $2271\,\mathrm{km\,s^{-1}}$. We model the imperfection of the resolution of BOSS spectra through a

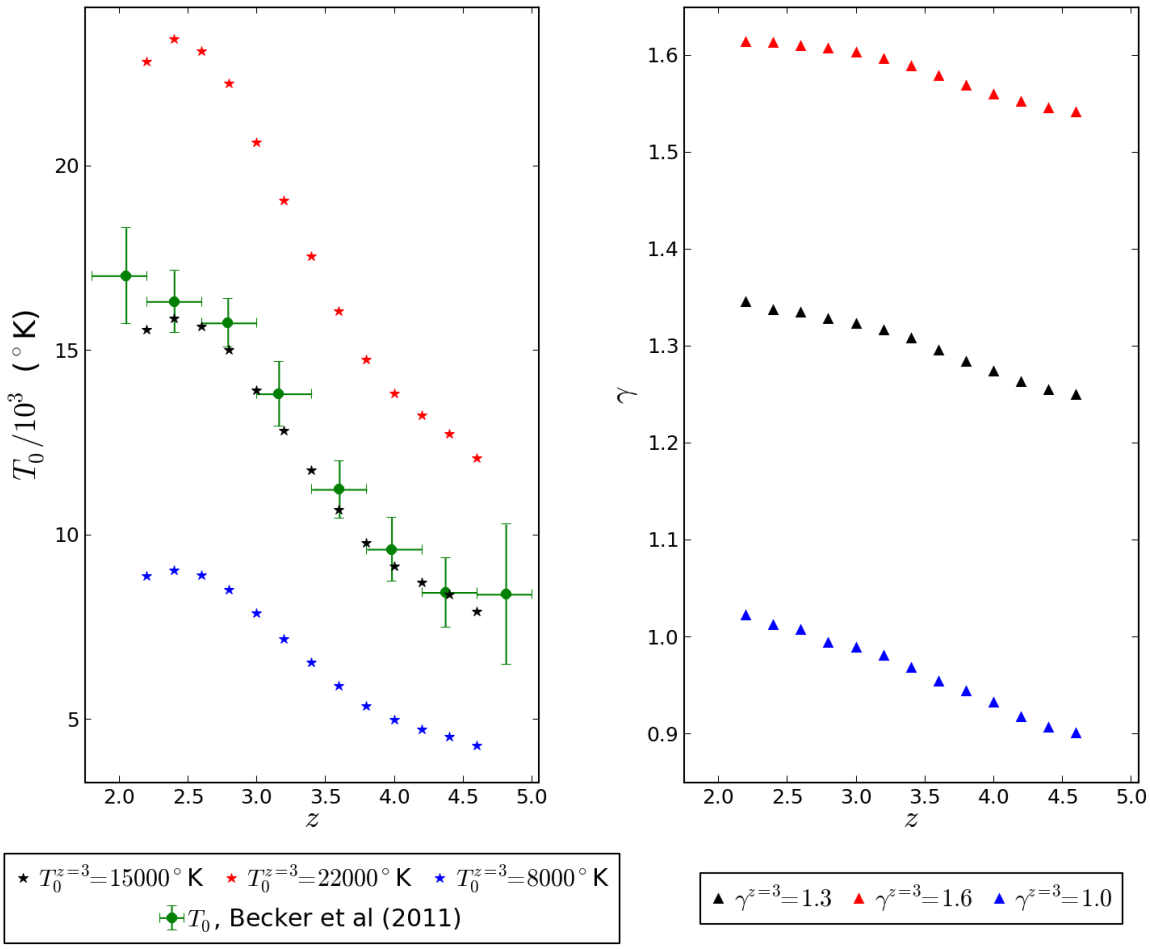


Figure 6.16 – Evolution with redshift of T_0 and γ for the different values of these two parameters used in the grid. The plotted values are extracted from a sample of particles using a mode estimation as explained in the text. I overlaid the measurements of Becker et al. (2011) (for $\gamma = 1.3$) for comparison.

multiplicative term. Finally, we allow for imperfection in the noise estimate of the BOSS spectra with eight additive terms (one for each redshift bin).

Figure 6.17 illustrates the good agreement between the data and our central simulation. Without any adjustment of the cosmological and astrophysical parameters, the χ^2 per number of degrees of freedom is already better than 1.2. The good agreement between data and simulation covers the whole redshift range, $z = [2.1 - 4.5]$. This simple comparison demonstrates the improvement obtained with these simulations over the previous generation of simulations (Viel et al., 2010) which use smaller range in k ($0.001\,41 - 0.017\,78\,(\mathrm{km/s})^{-1}$) and in z (eight redshift bins from $z = 2.2$ to $z = 3.6$). These simulations also have lower resolution than ours since a typical simulation has $L/N = 0.15$ and they all share the same box size and resolution corrections.

6.7.2 Fit Methodology

For a given cosmological model defined by the n cosmological, astrophysical and nuisance parameters $\Theta = (\theta_1, \ldots, \theta_n)$, and for a data set of power spectra $P(k_i, z_j)$ measured with Gaus-

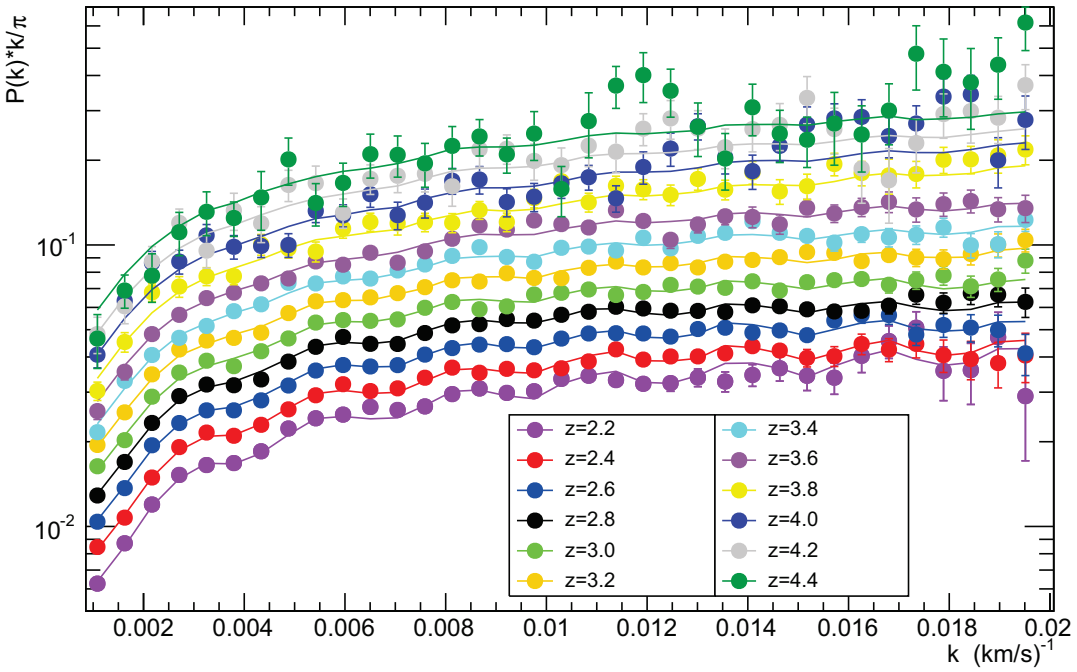


Figure 6.17 – One-dimensional Lyman-α forest power spectrum obtained with BOSS spectra. The dots are the measured power spectrum of chapter 4. The solid line represents the power spectrum for our central model after adjustment of nuisance parameters to account for imperfect modelling of the instrumental parameters in the 1D power spectrum measurement.

sian experimental errors $\sigma_{i,j}$, the likelihood function can be written as:

$$\mathcal{L}\left(P,\sigma,\Theta\right)=\prod_{i,j}\frac{1}{\sqrt{2\pi}\sigma_{i,j}}\times\exp\left(-\frac{\left[P(k_i,z_j)-P^{\mathrm{th}}(k_i,z_j)\right]^2}{2\sigma_{i,j}^2}\right) \tag{6.13}$$

where $P^{\mathrm{th}}(k_i,z_j)$ is the predicted value of the power spectrum for the bin k_i and redshift z_j. A χ^2 notation is adopted, which means that the following quantity is minimized:

$$\chi^2(P,\sigma,\Theta)=-2\ln\left(\mathcal{L}\left(P,\sigma;\Theta\right)\right)+\chi^2_{\mathrm{ext}}. \tag{6.14}$$

The second term of equation 6.14 represents the external constraints on the cosmological model. In our study, it is used to combine our results with those of Planck using the correlation matrix R^{Planck} and the error $\sigma_i^{\mathrm{Planck}}$ given by the Planck collaboration:

$$\chi^2_{\mathrm{Planck}}(\Theta)=\Delta^T C^{-1}\Delta \tag{6.15}$$

where C is the covariance matrix given by $C_{i,j}=R^{\mathrm{Planck}}_{i,j}\sigma_i^{\mathrm{Planck}}\sigma_j^{\mathrm{Planck}}$ and $\Delta_i=\theta_i^{\mathrm{Planck}}-\theta_i$ with $\theta_i^{\mathrm{Planck}}$ being the value of parameter i from the Planck results. In this study, the minimization of $\chi^2(P,\sigma;\Theta)$ is performed with the `MINUIT` package (James and Roos, 1975).

Most recent Lyman-α analyses use Markov Chain Monte Carlo simulations (Viel et al., 2010) with Bayesian inference. The debate between the Bayesian and the frequentist statistical approaches is beyond the scope of this work. The philosophical difference between the two methods should not generally lead, in the end, to major differences in the determination of

physical parameters and their confidence intervals when the parameters stay in a physical region (Yèche et al., 2006). It is currently tested in our group and preliminary results show no important differences between Bayesian and frequentist approaches.

Our work is based on the "frequentist" (or "classical") confidence level method originally defined by Neyman (1937). This avoids any potential bias due to the choice of priors. In addition, we have also found ways to improve the calculation speed, which gives our program some advantages over Bayesian programs.

We first determine the minimum χ^2_0 of $\chi^2(P, \sigma; \Theta)$ with all the cosmological parameters as free parameters. Then, to set a confidence level CL on any individual cosmological parameter θ_i, we scan the variable θ_i: for each fixed value of θ_i, we minimize again $\chi^2(P, \sigma; \Theta)$ but with $n-1$ free parameters. The χ^2 difference, $\Delta\chi^2(\theta_i)$, between the new minimum and χ^2_0, allows us to compute the confidence level on the variable, assuming that the experimental errors are Gaussian,

$$\mathrm{CL}(\theta_i) = 1 - \int_{\Delta\chi^2(\theta_i)}^{\infty} f_{\chi^2}(t; N_{dof})\,\mathrm{d}t, \tag{6.16}$$

with

$$f_{\chi^2}(t; N_{dof}) = \frac{e^{-t/2} t^{N_{dof}/2-1}}{\sqrt{2^{N_{dof}}}\,\Gamma(N_{dof}/2)} \tag{6.17}$$

where Γ is the Gamma function and the number of degrees of freedom N_{dof} is equal to 1. This method can be easily extended to two variables. In this case, the minimizations are performed for $n-2$ free parameters and the confidence level $\mathrm{CL}(\theta_i, \theta_j)$ is derived from equation 6.16 with $N_{dof} = 2$.

By definition, this frequentist approach does not require any marginalization to determine the sensitivity on a single individual cosmological parameter. Moreover, in contrast with Bayesian treatment, no prior on the cosmological parameters is needed. With this approach, the correlations between the variables are naturally taken into account and the minimization fit can explore the whole phase space of the cosmological, astrophysics and nuisance parameters.

6.7.3 Results

We performed two fits: one with massless neutrinos and another with possibly massive neutrinos. Because Lyman-α data alone poorly constrain T_0 and γ, we include a restriction on the range of T_0 ($14\,000 \pm 3000\,\mathrm{K}$ at 1σ) to keep it in physically reasonable region. Removing this constraint does not change the results of the fit on the other parameters. We also impose a constraint on H_0 because of the very strong H_0-n_s degeneracy with a gaussian likelihood with mean value and standard deviation based on Planck 2013 results: $H_0 = 67.4 \pm 1.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. The results of the fits are given in tables 6.7 and 6.8 and contours of confidence level in the σ_8-n_s and σ_8-$\sum_\nu m_\nu$ planes are given in figure 6.18.

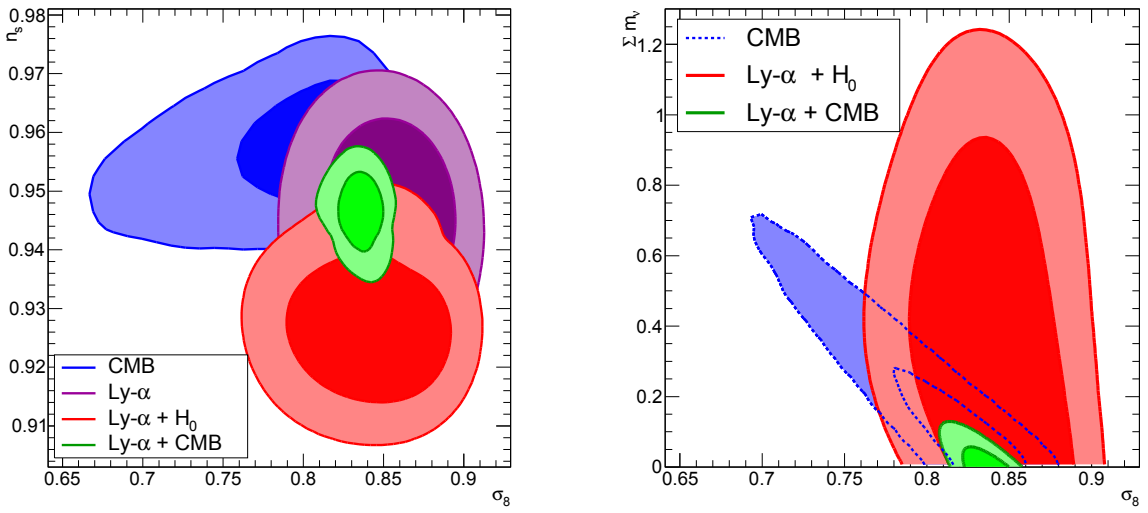
In the two cases the ratio of χ^2 by the number of degree of freedom is very good at 1.12. It is also worth noting that introducing massive neutrinos does not change significantly the result of the other parameters, except for σ_8 that shifts by 2σ. This is explained by the strong correlation between the two parameters (more than 50%), which simultaneously explains why the error on σ_8 increases by a factor 2 in the fit with massive (instead of massless) neutrinos. The strong constraints on $\Sigma_\nu m_\nu$ do not come from Lyman-α alone (which only gives $\Sigma_\nu m_\nu = 0.45 \pm 0.47\,\mathrm{eV}$), nor from tension between Lyman-α and Planck on other parameters as they agree remarkably well. As seen in the left plot of figure 6.18, for instance, Lyman-α and Planck

Parameter	Value
σ_8	0.86 ± 0.02
n_s	0.929 ± 0.007
Ω_m	0.291 ± 0.013
H_0	66.8 ± 1.4
T_0	11600^{+2600}_{-2800}
γ	1.1 ± 0.2
τ_A	0.0026 ± 0.0001
τ_S	3.67 ± 0.02
$f_{\mathrm{Si\,III}}$	0.0061 ± 0.0004

Table 6.7 – Result of the fit (frequentist approach) of the power spectrum measured with BOSS in the range $z = [2.1 - 4.5]$, with a fixed $\sum_\nu m_\nu = 0\,\mathrm{eV}$ and constraints on H_0 and T_0.

Parameter	Value
σ_8	0.85 ± 0.04
n_s	0.929 ± 0.007
Ω_m	0.293 ± 0.014
H_0	66.8 ± 1.4
T_0	11600^{+800}_{-800}
γ	1.2 ± 0.2
τ_A	0.0026 ± 0.0001
τ_S	3.67 ± 0.02
$f_{\mathrm{Si\,III}}$	0.0061 ± 0.0004
$\sum_\nu m_\nu$ ...	< 1.1 (95% C.L.)

Table 6.8 – Result of the fit (frequentist approach) of the power spectrum measured with BOSS in the range $z = [2.1 - 4.5]$ with $\sum_\nu m_\nu$ as an additional free parameter and constraints on H_0 and T_0.



(a) Contours in the σ_8-n_s plane

(b) Contours in the σ_8-$\sum_\nu m_\nu$ plane

Figure 6.18 – 2D confidence level contours in two different phase space planes obtained with a frequentist approach. The Lyman-α results, the Planck CMB results and their combination are given. H_0 represent the gaussian constraint that is imposed on this parameter.

2D contours in n_s-σ_8 are in agreement within 1σ. They rather come from the degeneracies between the two sets of parameters that are different in the two data sets, as can be seen on the right panel of figure 6.18. This is also the case for the parameters that mainly affect the overall levels of the one-dimensional power spectrum (τ_A, `AMPL` and noise correction) which are strongly correlated (up to 80%) when using Lyman-α alone. Combinations with other probes and extended Planck results (e.g. polarization results) are planned and should allow even tighter constraints. However, it is already worth noting that our combined constraints, with a 95% confidence level of $\Sigma_\nu m_\nu < 0.1\,\mathrm{eV}$, has already reached the sensitive region where one can discriminate between normal and inverted neutrino mass hierarchy (Lesgourgues and Pastor, 2006).

6.8 Conclusion

I have designed and produced a grid of cosmological simulations that is useful to extract constraints on cosmological parameters from Lyman-α surveys, whether current like SDSS-III/BOSS or future like SDSS-IV/eBOSS. These simulations cover the redshift range $2.2 - 4.6$. They explore the cosmological parameters n_s, σ_8, H_0, Ω_m and $\sum_\nu m_\nu$ over a large range centered on Planck measurements, as well as the astrophysical parameters T_0 and γ in a range covering recent results. The details about the simulations with massive neutrinos can be found in Rossi et al. (2014) which is published as a companion paper to Borde et al. (2014).

Using the splicing technique of McDonald (2003), I computed 1D power spectra from simulations equivalent to a $100\,(\mathrm{Mpc}/h)^3$ box filled with 3072^2 particles of each species (here dark matter, baryonic gas and sometimes neutrinos), noted $(100, 3072)$, from lower-resolution $(100, 768)$ and smaller box-size $(25, 768)$ simulations, combined using a $(25, 192)$ transition simulation. I show that the splicing technique allows us to approximate the exact full-resolution large-box simulation with an accuracy at a 2% level.

While one full-size high-resolution $(100, 3072)$ simulation would have required several million hours, one equivalent set of three simulations consumes an average of 50 000 hours of CPU time (100 000 hours for simulations including neutrinos), most of it being spent by `GADGET`. The data volume produced by each set is 1.6 terabytes (2 terabytes for simulations including neutrinos). Therefore, the whole grid represents about 4 million hours of CPU time and a volume of 70 terabytes of data.

From the 1D power spectra that we computed at each point of the grid, we derived a second-order Taylor expansion around our best-guess model. It describes the evolution of the 1D power spectrum with changes in either the cosmological or the astrophysical parameters that we studied. I have performed several check runs to ensure the quality and validity of the simulation grid, using either different seeds, or off-the-grid values of the cosmological and astrophysical parameters. These checks were all consistent with the power spectrum predicted using our second-order Taylor expansion, thus validating it.

We compared our central simulation to published data from BOSS and showed that they were already in good agreement without any adjustment of any of the simulation parameters. I also presented preliminary cosmological constraints from the combination of the experimental results of chapter 4 and the simulations described in this chapter. These results were obtained assuming a flat ΛCDM model with massless or massive neutrinos. These results currently do not take any bias into account but some have already been identified like those associated to the splicing and the low starting redshift.

There are still some potential bias and systematic errors that need to be studied, and a new proposal was made to ask for the computational time to do so. Additional parameters can also be included in the same context. However, due to the presence of the cross terms that are necessary for an accurate modelling of the likelihood function that illustrates the variation of the power spectrum in all directions of this growing parameter-space, adding new parameters will become more and more expensive in terms of calculation time.

Conclusion

Doing my thesis in the SDSS collaboration has been a privilege: it currently comprises a few of the most experienced cosmologists and physicists and the largest catalogue of astrophysical objects. The first two generations had a very strong impact on science in general, with regular public releases of the data: even six years after, the DR7 release (the last of SDSS-II) still leads to scientific results. There is no doubt that the third generation will have the same influence.

For this work, I performed two tasks that aim at the same goal: providing constraints on cosmological parameters in the ΛCDM model. First, we measured the one-dimensional Lyman-α power spectrum. I developed a method based on a likelihood function that allows a pixel by pixel correction of noise and resolution. It also provides an efficient way to treat the sky lines that are found to be inside the Lyman-α forest. This method, and a more classical one based on a Fourier transform, were applied to 13 821 quasar spectra selected from about 60 000 DR9 BOSS spectra. This leads to the most precise measurement of the one-dimensional power spectrum of our time. Such precision (a few percent) was made possible by a careful study of the systematic uncertainties linked either to instrumental effects or to the methods themselves so as to bring them to the level of the statistical uncertainties. Because of this, repeating the measurement on the next releases will require a lot of work to keep the two sources of uncertainties at the same level.

Secondly, I developed a grid of hydrodynamical simulations to explore the effect of several cosmological and astrophysical parameters on the one-dimensional matter power spectrum. These parameters include the sum of neutrino masses in addition to the main parameters of the ΛCDM model (σ_8, n_s, Ω_m, H_0). This grid was built to perform a second-order Taylor expansion (including cross-terms) around a best-guess model corresponding mostly to the recent Planck results. To obtain the necessary resolution for our simulations, we used with success a splicing technique developed a few years ago. I also performed many checks to ensure the quality and validity of our grid. We also studied several sources of biases and uncertainties like the initial conditions redshift, the splicing technique or the random seed used to generate the initial conditions. Although several sources still need to be explored, we performed fits using the measurement obtained in the first part to extract constraints on cosmology. These first results are very encouraging, especially when combined with another probe like the cosmological microwave background anisotropies recently measured by Planck. More precise combinations are being performed and a proposal for the required computational time has just been submitted to explore additional sources of potential biases. In the end, this study should lead to some of the tightest cosmological constraints, especially on the sum of the neutrino masses.

Appendix A

List of simulations

We summarize in table A.1 all the simulations mentioned in the paper. For box size and number of particles, (L, N) refers to a simulation with N^3 particles per species (gas, dark matter or neutrinos, thus $2 \times N^3$ or $3 \times N^3$ total particles) in a box of size $L\,\mathrm{Mpc}\,h^{-1}$ on a side. Standard parameters are $(n_s, \sigma_8, \Omega_m, H_0, T_0, \gamma) = (0.83, 0.96, 0.31, 67.5, 14.000, 1.3)$. Unless parameter names are explicitly listed, values are given for all parameters in the order just mentioned. All parameters are assumed to have their standard value unless specified otherwise. Except for the simulations performed for the convergence tests or to compute the exact power spectrum in the splicing test, all simulations are using the splicing technique to combine each set of three simulations into a single one of equivalent size to the largest box and equivalent mass-resolution to the best mass resolution.

Type	Box size, particles (L, N)	Simulation parameters
Convergence tests simulations		
Convergence: resolution	(20,1024)	Standard
Convergence: resolution	(20,768)	Standard
Convergence: resolution	(20,512)	Standard
Convergence: resolution	(20,384)	Standard
Convergence: resolution	(20,192)	Standard
Convergence: box size	(120,1024)	Standard
Convergence: box size	(90,768)	Standard
Convergence: box size	(80,683)	Standard
Convergence: box size	(60,512)	Standard
Convergence: box size	(20,171)	Standard
Grid simulations		
Central model	(25,768)+(100,768)+(25,192)	Standard
Derivatives n_s	(25,768)+(100,768)+(25,192)	$n_s = 0.91$
Derivatives n_s	(25,768)+(100,768)+(25,192)	$n_s = 1.01$
Derivatives σ_8	(25,768)+(100,768)+(25,192)	$\sigma_8 = 0.83$
Derivatives σ_8	(25,768)+(100,768)+(25,192)	$\sigma_8 = 0.93$
Derivatives Ω_m	(25,768)+(100,768)+(25,192)	$\Omega_m = 0.26$
Derivatives Ω_m	(25,768)+(100,768)+(25,192)	$\Omega_m = 0.36$
Derivatives H_0	(25,768)+(100,768)+(25,192)	$H_0 = 62.5$
Derivatives H_0	(25,768)+(100,768)+(25,192)	$H_0 = 72.5$
Derivatives T_0	(25,768)+(100,768)+(25,192)	$T_0 = 7000\,\mathrm{K}$

Derivatives T_0	(25,768)+(100,768)+(25,192)	$T_0 = 21\,000\,\mathrm{K}$
Derivatives γ	(25,768)+(100,768)+(25,192)	$\gamma = 1.0$
Derivatives γ	(25,768)+(100,768)+(25,192)	$\gamma = 1.6$
Derivatives $\sum_\nu m_\nu$	(25,768)+(100,768)+(25,192)	$\sum_\nu m_\nu = 0.4\,\mathrm{eV}$
Derivatives $\sum_\nu m_\nu$	(25,768)+(100,768)+(25,192)	$\sum_\nu m_\nu = 0.8\,\mathrm{eV}$
Cross-term n_s-σ_8	(25,768)+(100,768)+(25,192)	$n_s = 1.01$, $\sigma_8 = 0.93$
Cross-term n_s-Ω_m	(25,768)+(100,768)+(25,192)	$n_s = 1.01$, $\Omega_m = 0.36$
Cross-term n_s-H_0	(25,768)+(100,768)+(25,192)	$n_s = 1.01$, $H_0 = 72.5$
Cross-term n_s-T_0	(25,768)+(100,768)+(25,192)	$n_s = 1.01$, $T_0 = 21.000$
Cross-term n_s-γ	(25,768)+(100,768)+(25,192)	$n_s = 1.01$, $\gamma = 1.6$
Cross-term σ_8-Ω_m	(25,768)+(100,768)+(25,192)	$\sigma_8 = 0.93$, $\Omega_m = 0.36$
Cross-term σ_8-H_0	(25,768)+(100,768)+(25,192)	$\sigma_8 = 0.93$, $H_0 = 72.5$
Cross-term σ_8-T_0	(25,768)+(100,768)+(25,192)	$\sigma_8 = 0.93$, $T_0 = 21.000$
Cross-term σ_8-γ	(25,768)+(100,768)+(25,192)	$\sigma_8 = 0.93$, $\gamma = 1.6$
Cross-term Ω_m-H_0	(25,768)+(100,768)+(25,192)	$\Omega_m = 0.36$, $H_0 = 72.5$
Cross-term Ω_m-T_0	(25,768)+(100,768)+(25,192)	$\Omega_m = 0.36$, $T_0 = 21.000$
Cross-term Ω_m-γ	(25,768)+(100,768)+(25,192)	$\Omega_m = 0.36$, $\gamma = 1.6$
Cross-term H_0-T_0	(25,768)+(100,768)+(25,192)	$H_0 = 72.5$, $T_0 = 21.000$
Cross-term H_0-γ	(25,768)+(100,768)+(25,192)	$H_0 = 72.5$, $\gamma = 1.6$
Cross-term T_0-γ	(25,768)+(100,768)+(25,192)	$T_0 = 21.000$, $\gamma = 1.6$
Cross-term $\sum_\nu m_\nu$-n_s	(25,768)+(100,768)+(25,192)	$\sum_\nu m_\nu = 0.8\,\mathrm{eV}$, $n_s = 1.01$
Cross-term $\sum_\nu m_\nu$-σ_8	(25,768)+(100,768)+(25,192)	$\sum_\nu m_\nu = 0.8\,\mathrm{eV}$, $\sigma_8 = 0.93$
Cross-term $\sum_\nu m_\nu$-Ω_m	(25,768)+(100,768)+(25,192)	$\sum_\nu m_\nu = 0.8\,\mathrm{eV}$, $\Omega_m = 0.36$
Cross-term $\sum_\nu m_\nu$-H_0	(25,768)+(100,768)+(25,192)	$\sum_\nu m_\nu = 0.8\,\mathrm{eV}$, $H_0 = 72.5$
Cross-term $\sum_\nu m_\nu$-T_0	(25,768)+(100,768)+(25,192)	$\sum_\nu m_\nu = 0.8\,\mathrm{eV}$, $T_0 = 21\,000\,\mathrm{K}$
Cross-term $\sum_\nu m_\nu$-γ	(25,768)+(100,768)+(25,192)	$\sum_\nu m_\nu = 0.8\,\mathrm{eV}$, $\gamma = 1.6$
Additional simulations		
Alt. derivatives $\sum_\nu m_\nu$	(25,768)+(100,768)+(25,192)	$\sum_\nu m_\nu = 0.1\,\mathrm{eV}$
Alt. derivatives $\sum_\nu m_\nu$	(25,768)+(100,768)+(25,192)	$\sum_\nu m_\nu = 0.2\,\mathrm{eV}$
Splicing test: grid-like	(25,256)+(100,256)+(25,64)	Standard
Splicing test: exact	(100,1024)	Standard
Initial condition test	(100,768)	$z_{\mathrm{start}} = 99$, 2^{nd} order
Initial condition test	(100,768)	$z_{\mathrm{start}} = 30$, 2^{nd} order
Initial condition test	(100,768)	$z_{\mathrm{start}} = 99$, 1^{st} order
Initial condition test	(100,768)	$z_{\mathrm{start}} = 30$, 1^{st} order
Check: random seed	(25,768)+(100,768)+(25,192)	Standard
Check: off-grid 1	(25,768)+(100,768)+(25,192)	$(0.93, 0.85, 0.31, 67.5, 14000, 1.3)$
Check: off-grid 2	(25,768)+(100,768)+(25,192)	$(0.96, 0.83, 0.31, 67.5, 10000, 1.47)$
Check: off-grid 3	(25,768)+(100,768)+(25,192)	$(0.93, 0.86, 0.30, 66, 10000, 1.16)$

Table A.1 – Simulations used in this work.

List of Tables

List of Figures

Bibliography

Aarseth, S. J. (1963). Dynamical evolution of clusters of galaxies, I. *Monthly Notices of the Royal Astronomical Society 126*, 223–255.

Aarseth, S. J., E. L. Turner, and I. Gott, J. R. (1979). N-body simulations of galaxy clustering – I - Initial conditions and galaxy collapse times. *The Astrophysical Journal 234*, 13–26.

Abel, T. and M. G. Haehnelt (1999). Radiative Transfer Effects during Photoheating of the Intergalactic Medium. *The Astrophysical Journal 520*(1), L13–L16. `arXiv:astro-ph/9903102`.

Ahn, C. P., R. Alexandroff, C. Allende Prieto, S. F. Anderson, T. Anderton, et al. (2012). The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey. *The Astrophysical Journal Supplement Series 203*(2), 21. `arXiv:1207.7137`.

Aihara, H., C. Allende Prieto, D. An, S. F. Anderson, E. Aubourg, et al. (2011). The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III. *The Astrophysical Journal Supplement Series 193*(2), 29. `arXiv:1101.1559`.

Alcock, C. and B. Paczyński (1979). An evolution free test for non-zero cosmological constant. *Nature 281*(5730), 358–359.

Ali-Haimoud, Y. and S. Bird (2012). An efficient implementation of massive neutrinos in non-linear structure formation simulations. *Monthly Notices of the Royal Astronomical Society 428*(4), 3375–3389. `arXiv:1209.0461`.

Alimi, J.-m., V. Bouillot, Y. Rasera, V. Reverdy, P.-S. Corasaniti, et al. (2012). DEUS Full Observable ΛCDM Universe Simulation: the numerical challenge. `arXiv:1206.2838`.

Anderson, L., E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton, et al. (2013). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample. *Monthly Notices of the Royal Astronomical Society 427*(4), 3435–3467. `arXiv:1203.6594`.

Astier, P., J. Guy, N. Regnault, R. Pain, E. Aubourg, et al. (2006). The Supernova Legacy Survey: measurement of Ω_m, Ω_Λ and w from the first year data set. *Astronomy & Astrophysics 447*(1), 31–48. `arXiv:astro-ph/0510447v1`.

Barnes, J. and P. Hut (1986). A hierarchical $\mathcal{O}(N \log N)$ force-calculation algorithm. *Nature 324*(6096), 446–449.

Becker, G. D., J. S. Bolton, M. G. Haehnelt, and W. L. W. Sargent (2011). Detection of extended He II reionization in the temperature evolution of the intergalactic medium. *Monthly Notices of the Royal Astronomical Society 410*(2), 1096–1112. arXiv:1008.2622.

Bennett, C. L., D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *The Astrophysical Journal Supplement Series 208*(2), 20. arXiv:1212.5225.

Bertschinger, E. (1995). COSMICS: Cosmological Initial Conditions and Microwave Anisotropy Codes. arXiv:astro-ph/9506070.

Bi, H. (1993). Lyman-alpha absorption spectrum of the primordial intergalactic medium. *The Astrophysical Journal 405*(2), 479.

Bird, S., H. V. Peiris, M. Viel, and L. Verde (2011). Minimally parametric power spectrum reconstruction from the Lyman-α forest. *Monthly Notices of the Royal Astronomical Society 413*(3), 1717–1728. arXiv:1010.1519.

Bird, S., M. Viel, and M. G. Haehnelt (2012). Massive neutrinos and the non-linear matter power spectrum. *Monthly Notices of the Royal Astronomical Society 420*(3), 2551–2561. arXiv:1109.4416.

Blanton, M. R., H. Lin, R. H. Lupton, F. M. Maley, N. Young, et al. (2003). An Efficient Targeting Strategy for Multiobject Spectrograph Surveys: the Sloan Digital Sky Survey "Tiling" Algorithm. *The Astronomical Journal 125*(4), 2276–2286. arXiv:astro-ph/0105535.

Bolton, A. S., D. J. Schlegel, E. Aubourg, S. Bailey, V. Bhardwaj, et al. (2012). Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey. *The Astronomical Journal 144*(5), 144. arXiv:1207.7326.

Bolton, J. S. and G. D. Becker (2009). Resolving the high redshift Lyα forest in smoothed particle hydrodynamics simulations. *Monthly Notices of the Royal Astronomical Society: Letters 398*(1), L26–L30. arXiv:0906.2861.

Bolton, J. S., M. Viel, T.-S. Kim, M. G. Haehnelt, and R. F. Carswell (2008). Possible evidence for an inverted temperature–density relation in the intergalactic medium from the flux distribution of the Lyα forest. *Monthly Notices of the Royal Astronomical Society 386*(2), 1131–1144. arXiv:0711.2064.

Bond, J., A. Jaffe, and L. Knox (1998). Estimating the power spectrum of the cosmic microwave background. *Physical Review D 57*(4), 2117–2137. arXiv:astro-ph/9708203.

Booth, C. M. (2007). *Simulating Molecular Cloud Regulated Star Formation in Galaxies*. Ph. D. thesis, University of Durham.

Borde, A., N. Palanque-Delabrouille, G. Rossi, M. Viel, J. Bolton, et al. (2014). New approach for precise computation of Lyman-alpha forest power spectrum with hydrodynamical simulations. arXiv:1401.6472.

Borde, A., C. Yeche, N. Palanque-Delabrouille, G. Rossi, R. Croft, et al. (2013). Measurement of the 1D Lyman-α power spectrum with the DR9 BOSS quasar data. In AAS (Ed.), *AAS 221st meeting*, Long Beach, CA.

Brandt, A. (1977). Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computation 31*(138), 333–333.

Busca, N. G., T. Delubac, J. Rich, S. Bailey, A. Font-Ribera, et al. (2013). Baryon acoustic oscillations in the Lyα forest of BOSS quasars. *Astronomy & Astrophysics 552*, A96. arXiv:1211.2616.

Cen, R. (1992). A hydrodynamic approach to cosmology - Methodology. *The Astrophysical Journal Supplement Series 78*, 341–364.

Cen, R., J. Miralda-Escude, J. P. Ostriker, and M. Rauch (1994). Gravitational collapse of small-scale structure as the origin of the Lyman-alpha forest. *The Astrophysical Journal 437*, L9–L12.

Colella, P. and P. R. Woodward (1984). The Piecewise Parabolic Method (PPM) for gas-dynamical simulations. *Journal of Computational Physics 54*(1), 174–201.

Conley, A., J. Guy, M. Sullivan, N. Regnault, P. Astier, et al. (2011). Supernova Constraints and Systematic Uncertainties from the First 3 Years of the Supernova Legacy Survey. *The Astrophysical Journal Supplement Series 192*(1), 29. arXiv:1104.1443.

Couchman, H. M. P., P. A. Thomas, and F. R. Pearce (1995). Hydra: an Adaptive-Mesh Implementation of P^3M-SPH. *The Astrophysical Journal 452*, 797. arXiv:astro-ph/9409058.

Crocce, M., S. Pueblas, and R. Scoccimarro (2006). Transients from initial conditions in cosmological simulations. *Monthly Notices of the Royal Astronomical Society 373*(1), 369–381. arXiv:astro-ph/0606505.

Croft, R. A. C., D. H. Weinberg, M. Bolte, S. Burles, L. Hernquist, et al. (2002). Toward a Precise Measurement of Matter Clustering: Lyman-alpha Forest Data at Redshifts 2–4. *The Astrophysical Journal 581*(1), 20–52. arXiv:astro-ph/0012324.

Croft, R. A. C., D. H. Weinberg, N. Katz, and L. Hernquist (1997). Intergalactic Helium Absorption in Cold Dark Matter Models. *The Astrophysical Journal 488*(2), 532–549. arXiv:astro-ph/9611053.

Croft, R. A. C., D. H. Weinberg, N. Katz, and L. Hernquist (1998). Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Lyman-alpha Forest. *The Astrophysical Journal 495*(1), 44–62. arXiv:astro-ph/9708018.

Croft, R. A. C., D. H. Weinberg, M. Pettini, L. Hernquist, and N. Katz (1999). The Power Spectrum of Mass Fluctuations Measured from the Lyman-alpha Forest at Redshift z = 2.5. *The Astrophysical Journal 520*(1), 1–23. arXiv:astro-ph/9809401.

Dawson, K. S., D. J. Schlegel, C. P. Ahn, S. F. Anderson, E. Aubourg, et al. (2013). The Baryon Oscillation Spectroscopic Survey of SDSS-III. *The Astronomical Journal 145*(1), 10. arXiv:1208.0022.

Desjacques, V., A. Nusser, and R. K. Sheth (2007). The probability distribution function of the Lyman-α transmitted flux from a sample of Sloan Digital Sky Survey quasars. *Monthly Notices of the Royal Astronomical Society 374*(1), 206–219. arXiv:astro-ph/0608167.

Dobrzycki, A., L. M. Macri, K. Z. Stanek, and P. J. Groot (2003). Variability-selected Quasars behind the Small Magellanic Cloud. *The Astronomical Journal 125*(3), 1330–1335. arXiv:astro-ph/0210601.

Dolag, K., S. Borgani, S. Schindler, A. Diaferio, and A. M. Bykov (2008). Simulation Techniques for Cosmological Simulations. *Space Science Reviews 134*(1-4), 229–268. arXiv:0801.1023.

Efstathiou, G. and J. Eastwood (1981). On the clustering of particles in an expanding Universe. *Monthly Notices of the Royal Astronomical Society 194*, 503–525.

Einstein, A. (1916). Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik 354*(7), 769–822.

Eisenstein, D., I. Zehavi, D. Hogg, R. Scoccimarro, M. Blanton, et al. (2005). Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies. *The Astrophysical Journal 633*, 560. arXiv:astro-ph/0501171.

Eisenstein, D. J., H.-j. Seo, E. Sirko, and D. N. Spergel (2007). Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak. *The Astrophysical Journal 664*(2), 675–679. arXiv:astro-ph/0604362.

Eisenstein, D. J., D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, et al. (2011). SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way and Extra-Solar Planetary Systems. *The Astronomical Journal 142*(3), 72. arXiv:1101.1529.

Feng, L.-l. and L.-z. Fang (2000). Non-Gaussianity and the Recovery of the Mass Power Spectrum from the Lyα Forest. *The Astrophysical Journal 535*(2), 519–529. arXiv:astro-ph/0001348.

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

Font-Ribera, A., P. McDonald, and J. Miralda-Escudé (2012). Generating mock data sets for large-scale Lyman-α forest correlation measurements. *Journal of Cosmology and Astroparticle Physics 2012*(01), 17. arXiv:1108.5606.

Frenk, C. S., S. D. M. White, and M. Davis (1983). Nonlinear evolution of large-scale structure in the universe. *The Astrophysical Journal 271*, 417.

Frigo, M. and S. Johnson (1998). FFTW: an adaptive software architecture for the FFT. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, Volume 3, pp. 1381–1384. IEEE.

Garzilli, A., J. S. Bolton, T.-S. Kim, S. Leach, and M. Viel (2012). The intergalactic medium thermal history at redshift z = 1.7-3.2 from the Lyα forest: a comparison of measurements using wavelets and the flux distribution. *Monthly Notices of the Royal Astronomical Society 424*(3), 1723–1736. arXiv:1202.3577.

Gaztanaga, E. and R. A. C. Croft (1999). Predictions for the clustering properties of the Lyman-alpha forest – I. One-point statistics. *Monthly Notices of the Royal Astronomical Society 309*(4), 885–904. arXiv:astro-ph/9811480.

Geha, M., C. Alcock, R. A. Allsman, D. R. Alves, T. S. Axelrod, et al. (2003). Variability-selected Quasars in MACHO Project Magellanic Cloud Fields. *The Astronomical Journal 125*(1), 1–12. arXiv:astro-ph/0209513.

Gingold, R. A. and J. J. Monaghan (1977). Smoothed particle hydrodynamics - Theory and application to non-spherical starstle. *Monthly Notices of the Royal Astronomical Society 181*, 375–389.

Gnedin, N. Y. (1998). Probing the Universe with the Lyman-alpha forest – II. The column density distribution. *Monthly Notices of the Royal Astronomical Society 299*(2), 392–402. arXiv:astro-ph/9706286.

Gnedin, N. Y. and A. J. S. Hamilton (2002). Matter power spectrum from the Lyman-alpha forest: myth or reality? *Monthly Notices of the Royal Astronomical Society 334*(1), 107–116. arXiv:astro-ph/0111194.

Godunov, S. K. (1959). A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Matematicheskii Sbornik 47*(89), 271–306.

Gunn, J. E., M. Carr, C. Rockosi, M. Sekiguchi, K. Berry, et al. (1998). The Sloan Digital Sky Survey Photometric Camera. *The Astronomical Journal 116*(6), 3040–3081. arXiv:astro-ph/9809085.

Gunn, J. E. and B. A. Peterson (1965). On the Density of Neutral Hydrogen in Intergalactic Space. *The Astrophysical Journal 142*, 1633.

Gunn, J. E., W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, et al. (2006). The 2.5 m Telescope of the Sloan Digital Sky Survey. *The Astronomical Journal 131*(4), 2332–2359. arXiv:astro-ph/0602326.

Heitmann, K., D. Higdon, M. White, S. Habib, B. J. Williams, et al. (2009). The Coyote Universe – II. Cosmological Models and Precision Emulation of the Nonlinear Matter Power Spectrum. *The Astrophysical Journal 705*(1), 156–174. arXiv:0902.0429.

Hénon, M. (1964). L'évolution initiale d'un amas sphérique. *Annales d'Astrophysique 27*, 83.

Hernquist, L., N. Katz, D. H. Weinberg, and J. Miralda-Escudé (1996). The Lyman-Alpha Forest in the Cold Dark Matter Model. *The Astrophysical Journal 457*(2), 1–11. arXiv:astro-ph/9509105.

Hockney, R. W. and J. W. Eastwood (1988). *Computer simulation using particles*. Bristol, PA, USA: Taylor & Francis, Inc.

Holmberg, E. (1941). On the Clustering Tendencies among the Nebulae – II. A Study of Encounters Between Laboratory Models of Stellar Systems by a New Integration Procedure. *The Astrophysical Journal 94*, 385.

Holtzman, J. a., J. Marriner, R. Kessler, M. Sako, B. Dilday, et al. (2009). The Sloan Digital Sky Survey-II: Photometry and Supernova Ia Light Curves from the 2005 data. *The Astronomical Journal 136*(6), 2306–2320. arXiv:0908.4277.

Hou, Z., C. L. Reichardt, K. T. Story, B. Follin, R. Keisler, et al. (2014). Constraints on Cosmology from the Cosmic Microwave Background Power Spectrum of the 2500 deg^2 SPT-SZ Survey. *The Astrophysical Journal 782*(2), 74. arXiv:1212.6267.

Hubble, E. (1929). A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences 15*(3), 168–173.

Hui, L., S. Burles, U. Seljak, R. E. Rutledge, E. Magnier, et al. (2001). On Estimating the QSO Transmission Power Spectrum. *The Astrophysical Journal 552*(1), 15–35. arXiv:astro-ph/0005049.

Hui, L. and N. Y. Gnedin (1997). Equation of state of the photoionized intergalactic medium. *Monthly Notices of the Royal Astronomical Society 292*(1), 27–42. arXiv:astro-ph/9612232.

Hui, L., N. Y. Gnedin, and Y. Zhang (1997). The Statistics of Density Peaks and the Column Density Distribution of the Ly-alpha Forest. *The Astrophysical Journal 486*(2), 599–622. arXiv:astro-ph/9608157.

Hui, L., A. Stebbins, and S. Burles (1999). A Geometrical Test of the Cosmological Energy Contents Using the Lyman-alpha Forest. *The Astrophysical Journal 511*(1), L5–L8. arXiv:astro-ph/9807190.

James, F. and M. Roos (1975). Minuit - a system for function minimization and analysis of the parameter errors and correlations. *Computer Physics Communications 10*(6), 343–367.

Jernigan, J. G. and D. H. Porter (1989). A tree code with logarithmic reduction of force terms, hierarchical regularization of all variables, and explicit accuracy controls. *The Astrophysical Journal Supplement Series 71*, 871.

Kaiser, N. (1987). Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society 227*(1), 1–21.

Kawaguchi, T., S. Mineshige, M. Umemura, and E. L. Turner (1998). Optical Variability in Active Galactic Nuclei: Starbursts or Disk Instabilities? *The Astrophysical Journal 504*(2), 671–679. arXiv:astro-ph/9712006.

Kessler, R., A. C. Becker, D. Cinabro, J. Vanderplas, J. a. Frieman, et al. (2009). First-Year Sloan Digital Sky Survey-Ii Supernova Results: Hubble Diagram and Cosmological Parameters. *The Astrophysical Journal Supplement Series 185*(1), 32–84. arXiv:0908.4274.

Kim, T.-S., M. Viel, M. G. Haehnelt, B. Carswell, and S. Cristiani (2004a). Erratum: The power spectrum of the flux distribution in the Lyman-alpha forest of a large sample of UVES QSO Absorption Spectra (LUQAS). *Monthly Notices of the Royal Astronomical Society 351*(4), 1471–1472.

Kim, T. S., M. Viel, M. G. Haehnelt, R. F. Carswell, and S. Cristiani (2004b). The power spectrum of the flux distribution in the Lyman-α forest of a large sample of UVES QSO absorption spectra (LUQAS). *Monthly Notices of the Royal Astronomical Society 347*(2), 355–366. arXiv:astro-ph/0308103.

Kirkby, D., D. Margala, A. Slosar, S. Bailey, N. G. Busca, et al. (2013). Fitting methods for baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9. *Journal of Cosmology and Astroparticle Physics 2013*(03), 38. arXiv:1301.3456.

Klypin, A. and S. Shandarin (1983). Three-dimensional numerical model of the formation of large-scale structure in the Universe. *Monthly Notices of the Royal Astronomical Society 204*, 891–907.

Komatsu, E., K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, et al. (2011). Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *The Astrophysical Journal Supplement Series 192*(2), 18. arXiv:1001.4538.

Le Goff, J. M., C. Magneville, E. Rollinde, S. Peirani, P. Petitjean, et al. (2011). Simulations of BAO reconstruction with a quasar Ly-α survey. *Astronomy & Astrophysics 534*, A135. arXiv:1107.4233.

Lee, K.-g., S. Bailey, L. E. Bartsch, W. Carithers, K. S. Dawson, et al. (2013). The BOSS Lyman-α forest sample from SDSS data release 9. *The Astronomical Journal 145*(3), 69. arXiv:1211.5146.

Lesgourgues, J. and S. Pastor (2006). Massive neutrinos and cosmology. *Physics Reports 429*(6), 307–379. arXiv:astro-ph/0603494.

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

L'Huillier, B., C. Park, and J. Kim (2014). Effects of the initial conditions on cosmological N-body simulations. arXiv:1401.6180.

Lidz, A., C.-A. Faucher-Giguère, A. Dall'Aglio, M. McQuinn, C. Fechner, et al. (2010). A Measurement of Small-Scale Structure in the $2.2 \leq z \leq 4.2$ Lyman-alpha Forest. *The Astrophysical Journal 718*(1), 199–230. arXiv:0909.5210.

Lineweaver, C. H., L. Tenorio, G. F. Smoot, P. Keegstra, A. J. Banday, et al. (1996). The Dipole Observed in the COBE DMR 4 Year Data. *The Astrophysical Journal 470*, 38. arXiv:astro-ph/9601151.

Lucy, L. B. (1977). A numerical approach to the testing of the fission hypothesis. *The Astronomical Journal 82*, 1013.

Matthews, T. A. and A. R. Sandage (1963). Optical Identification of 3c 48, 3c 196, and 3c 286 with Stellar Objects. *The Astrophysical Journal 138*, 30.

McDonald, P. (2003). Toward a Measurement of the Cosmological Geometry at $z \sim 2$: Predicting Lyα Forest Correlation in Three Dimensions and the Potential of Future Data Sets. *The Astrophysical Journal 585*(1), 34–51. arXiv:astro-ph/0108064.

McDonald, P. and D. Eisenstein (2007). Dark energy and curvature from a future baryonic acoustic oscillation survey using the Lyman-alpha forest. *Physical Review D 76*(6), 18. arXiv:astro-ph/0607122.

McDonald, P. and J. Miralda-Escude (1999). Measuring the Cosmological Geometry from the Lyman Alpha Forest along Parallel Lines of Sight. *The Astrophysical Journal 518*(1), 24–31. arXiv:astro-ph/9807137.

McDonald, P., J. Miralda-Escude, M. Rauch, W. L. W. Sargent, T. A. Barlow, et al. (2000). The Observed Probability Distribution Function, Power Spectrum, and Correlation Function of the Transmitted Flux in the Lyα Forest. *The Astrophysical Journal 543*(1), 1–23. arXiv:astro-ph/9911196.

McDonald, P., U. Seljak, S. Burles, D. J. Schlegel, D. H. Weinberg, et al. (2006). The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey. *The Astrophysical Journal Supplement Series 163*(1), 80–109. arXiv:astro-ph/0405013.

McDonald, P., U. Seljak, R. Cen, D. Shih, D. H. Weinberg, et al. (2005). The Linear Theory Power Spectrum from the Lyα Forest in the Sloan Digital Sky Survey. *The Astrophysical Journal 635*(2), 761–783. arXiv:astro-ph/0407377.

McKay, M. D., R. J. Beckman, and W. J. Conover (2000). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code. *Technometrics 42*(1), 55–61.

McQuinn, M. and M. White (2011). On estimating Lyα forest correlations between multiple sightlines. *Monthly Notices of the Royal Astronomical Society 415*(3), 2257–2269. arXiv:1102.1752.

Meiksin, A. (2009). The physics of the intergalactic medium. *Reviews of Modern Physics 81*(4), 1405–1469. arXiv:0711.3358.

Miralda-Escude, J. and M. J. Rees (1993). Tests for the minihalo model of the Lyman alpha forest. *Monthly Notices of the Royal Astronomical Society 260*(3), 617–624.

Monaghan, J. (1985). Extrapolating B splines for interpolation. *Journal of Computational Physics 60*(2), 253–262.

Monaghan, J. J. (1992). Smoothed Particle Hydrodynamics. *Annual Review of Astronomy and Astrophysics 30*(1), 543–574.

Monaghan, J. J. (2005). Smoothed particle hydrodynamics. *Reports on Progress in Physics 68*(8), 1703–1759.

Mortlock, D. J., S. J. Warren, B. P. Venemans, M. Patel, P. C. Hewett, et al. (2011). A luminous quasar at a redshift of $z = 7.085$. *Nature 474*(7353), 616–9. arXiv:1106.6088.

Neyman, J. (1937). Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 236*(767), 333–380.

Noterdaeme, P., P. Petitjean, C. Ledoux, and R. Srianand (2009). Evolution of the cosmological mass density of neutral gas from Sloan Digital Sky Survey II – Data Release 7. *Astronomy & Astrophysics 505*(3), 1087–1098. arXiv:0908.1574.

Nusser, A. and M. Haehnelt (1999). A first step towards a direct inversion of the Lyman forest in QSO spectra. *Monthly Notices of the Royal Astronomical Society 303*(1), 179–187. arXiv:astro-ph/9806109.

Osmer, P. (2006). Quasistellar Objects : Overview.

Palanque-Delabrouille, N., C. Magneville, C. Yèche, S. Eftekharzadeh, A. D. Myers, et al. (2013). Luminosity function from dedicated SDSS-III and MMT data of quasars in $0.7 < z < 4.0$ selected with a new approach. *Astronomy & Astrophysics 551*, A29. arXiv:1209.3968.

Palanque-Delabrouille, N., C. Yèche, A. Borde, J.-M. Le Goff, G. Rossi, et al. (2013). The one-dimensional Lyα forest power spectrum from BOSS. *Astronomy & Astrophysics 559*, A85. arXiv:1306.5896.

Palanque-Delabrouille, N., C. Yeche, A. D. Myers, P. Petitjean, N. P. Ross, et al. (2011). Variability selected high-redshift quasars on SDSS Stripe 82. *Astronomy & Astrophysics 530*, A122. arXiv:1012.2391.

Pâris, I., P. Petitjean, E. Aubourg, S. Bailey, N. P. Ross, et al. (2012). The Sloan Digital Sky Survey quasar catalog: ninth data release. *Astronomy & Astrophysics 548*, A66. arXiv:1210.5166.

Pâris, I., P. Petitjean, E. Aubourg, N. P. Ross, A. D. Myers, et al. (2014). The Sloan Digital Sky Survey quasar catalog: tenth data release. *Astronomy & Astrophysics 563*, A54. arXiv:1311.4870.

Pâris, I., P. Petitjean, E. Rollinde, E. Aubourg, N. Busca, et al. (2011). A principal component analysis of quasar UV spectra at $z \sim 3$. *Astronomy & Astrophysics 530*, A50. arXiv:1104.2024.

Peebles, P. J. E. (1980). *The Large-Scale Structure of the Universe*. Princeton, N.J.: Princeton University Press.

Perlmutter, S., G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, et al. (1999). Measurements of Ω and Λ from 42 High-Redshift Supernovae. *The Astrophysical Journal 517*(2), 565–586. arXiv:astro-ph/9812133v1.

Pieri, M. M., S. Frank, D. H. Weinberg, S. Mathur, and D. G. York (2010). The Composite Spectrum of Strong Lyman Alpha Forest Absorbers. *The Astrophysical Journal 724*(1), L69–L73. arXiv:1001.5282.

Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, et al. (2013). Planck 2013 results. I. Overview of products and scientific results. arXiv:1303.5062.

Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, et al. (2013). Planck 2013 results. XVI. Cosmological parameters. arXiv:1303.5076.

Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, et al. (2013). Planck intermediate results. XIII. Constraints on peculiar velocities. arXiv:1303.5090.

Press, W. H. and P. Schechter (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *The Astrophysical Journal 187*, 425.

Price, D. J. (2012a). Smoothed particle hydrodynamics and magnetohydrodynamics. *Journal of Computational Physics 231*(3), 759–794. arXiv:1012.1885.

Price, D. J. (2012b). Smoothed Particle Hydrodynamics: Things I wish my mother taught me. In R. Capuzzo-Dolcetta, M. Limongi, and A. Tornambè (Eds.), *Advances in computational astrophysics: methods, tools, and outcome*, Cefalù, Italy, pp. 249.

Rauch, M. (2006). Lyman Alpha Forest.

Rengstorf, A. W., R. J. Brunner, and B. C. Wilhite (2006). A Synoptic Multiwavelength Analysis of a Large Quasar Sample. *The Astronomical Journal 131*(4), 1923–1933. arXiv:astro-ph/0512476.

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

Riess, A. G., L.-g. Strolger, S. Casertano, H. C. Ferguson, B. Mobasher, et al. (2007). New Hubble Space Telescope Discoveries of Type Ia Supernovae at $z \geq 1$: Narrowing Constraints on the Early Behavior of Dark Energy. *The Astrophysical Journal 659*(1), 98–121. arXiv:astro-ph/0611572.

Ross, N. P., A. D. Myers, E. S. Sheldon, C. Yèche, M. a. Strauss, et al. (2012). The SDSS-III Baryon Oscillation Spectroscopic Survey: Quasar Target Selection for Data Release Nine. *The Astrophysical Journal Supplement Series 199*(1), 3. arXiv:1105.0606.

Rossi, G., N. Palanque-Delabrouille, A. Borde, M. Viel, C. Yeche, et al. (2014). A Suite of Hydrodynamical Simulations for the Lyman-Alpha Forest with Massive Neutrinos. arXiv:1401.6464.

Rosswog, S. (2009). Astrophysical smooth particle hydrodynamics. *New Astronomy Reviews 53*(4-6), 78–104. arXiv:0903.5075.

Sachs, R. K. and A. M. Wolfe (1967). Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. *The Astrophysical Journal 147*, 73.

Schaye, J., T. Theuns, M. Rauch, G. Efstathiou, and W. L. Sargent (2000). The thermal history of the intergalactic medium. *Monthly Notices of the Royal Astronomical Society 318*(3), 817–826. arXiv:astro-ph/9912432.

Schlegel, D., F. Abdalla, T. Abraham, C. Ahn, C. A. Prieto, et al. (2011). The BigBOSS Experiment. arXiv:1106.1706.

Schneider, M. D., S. Cole, C. S. Frenk, and I. Szapudi (2011). Fast Generation of Ensembles of Cosmological N-Body Simulations via Mode Resampling. *The Astrophysical Journal 737*(1), 11. arXiv:1103.2767.

Scoccimarro, R. (1998). Transients from initial conditions: a perturbative analysis. *Monthly Notices of the Royal Astronomical Society 299*(4), 1097–1118. arXiv:astro-ph/9711187.

Scrimgeour, M. I., T. Davis, C. Blake, J. B. James, G. B. Poole, et al. (2012). The WiggleZ Dark Energy Survey: the transition to large-scale cosmic homogeneity. *Monthly Notices of the Royal Astronomical Society 425*(1), 116–134. arXiv:1205.6812.

Seljak, U. (1998a). Cosmography and Power Spectrum Estimation: A Unified Approach. *The Astrophysical Journal 503*(2), 492–501. arXiv:astro-ph/9710269.

Seljak, U. (1998b). Weak Lensing Reconstruction and Power Spectrum Estimation: Minimum Variance Methods. *The Astrophysical Journal 506*(1), 64–79. arXiv:astro-ph/9711124.

Seljak, U., A. Makarov, P. McDonald, S. Anderson, N. Bahcall, et al. (2005). Cosmological parameter analysis including SDSS Lyα forest and galaxy bias: Constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy. *Physical Review D 71*(10), 103515. arXiv:astro-ph/0407372.

Seljak, U. and M. Zaldarriaga (1996). A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *The Astrophysical Journal 469*, 437. arXiv:astro-ph/9603033.

Shandarin, S. F. and Y. B. Zeldovich (1989). The large-scale structure of the universe: Turbulence, intermittency, structures in a self-gravitating medium. *Reviews of Modern Physics 61*(2), 185–220.

Slosar, A., A. Font-Ribera, M. M. Pieri, J. Rich, J.-m. L. Goff, et al. (2011). The Lyman-α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data. *Journal of Cosmology and Astroparticle Physics 2011*(09), 001–001. arXiv:1104.5244.

Slosar, A., V. Iršič, D. Kirkby, S. Bailey, N. G. Busca, et al. (2013). Measurement of baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9. *Journal of Cosmology and Astroparticle Physics 2013*(04), 58. arXiv:1301.3459.

Smee, S., J. E. Gunn, A. Uomoto, N. Roe, D. Schlegel, et al. (2012). The Multi-Object, Fiber-Fed Spectrographs for SDSS and the Baryon Oscillation Spectroscopic Survey. *The Astronomical Journal 146*(2), 43. arXiv:1208.2233.

Springel, V. (2005). The cosmological simulation code GADGET-2. *Monthly Notices of the Royal Astronomical Society 364*(4), 1105–1134. arXiv:astro-ph/0505010.

Springel, V. (2010). Smoothed Particle Hydrodynamics in Astrophysics. *Annual Review of Astronomy and Astrophysics 48*(1), 391–430. arXiv:1109.2219.

Springel, V. and L. Hernquist (2002). Cosmological smoothed particle hydrodynamics simulations: the entropy equation. *Monthly Notices of the Royal Astronomical Society 333*(3), 649–664. arXiv:astro-ph/0111016.

Springel, V., N. Yoshida, and S. D. White (2001). GADGET: a code for collisionless and gasdynamical cosmological simulations. *New Astronomy 6*(2), 79–117. arXiv:astro-ph/0003162.

Sugimoto, D., Y. Chikada, J. Makino, T. Ito, T. Ebisuzaki, et al. (1990). A special-purpose computer for gravitational many-body problems. *Nature 345*(6270), 33–35.

Sullivan, M., J. Guy, A. Conley, N. Regnault, P. Astier, et al. (2011). SNLS3: Constraints on Dark Energy Combining the Supernova Legacy Survey Three-Year Data with Other Probes. *The Astrophysical Journal 737*(2), 102. arXiv:1104.1444.

Tang, B. (1993). Orthogonal Array-Based Latin Hypercubes. *Journal of the American Statistical Association 88*(424), 1392–1397.

Tegmark, M. and M. Zaldarriaga (2002). Separating the early universe from the late universe: Cosmological parameter estimation beyond the black box. *Physical Review D 66*(10), 103508. arXiv:astro-ph/0207047.

Teyssier, R. (2002). Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. *Astronomy & Astrophysics 385*(1), 337–364. arXiv:astro-ph/0111367.

The Enzo Collaboration, G. L. Bryan, M. L. Norman, B. W. O'Shea, T. Abel, et al. (2013). ENZO: An Adaptive Mesh Refinement Code for Astrophysics. *The Astrophysical Journal Supplement Series 211*(2), 19. arXiv:1307.2265.

Theuns, T. (2005). Numerical simulations of quasar absorbers. *Proceedings of the International Astronomical Union 1*(C199), 185. arXiv:astro-ph/0507570.

Trevese, D., R. G. Kron, and A. Bunone (2001). Continuum Variability of Active Galactic Nuclei in the Optical–Ultraviolet Range. *The Astrophysical Journal 551*(1), 103–110. arXiv:astro-ph/0012408.

van den Bergh, S., E. Herbst, and C. Pritchet (1973). A search for faint variable objects. *The Astronomical Journal 78*, 375.

van Leer, B. (1979). Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method. *Journal of Computational Physics 32*(1), 101–136.

Viel, M. and M. G. Haehnelt (2006). Cosmological and astrophysical parameters from the Sloan Digital Sky Survey flux power spectrum and hydrodynamical simulations of the Lyman-α forest. *Monthly Notices of the Royal Astronomical Society 365*(1), 231–244. arXiv:astro-ph/0508177.

Viel, M., M. G. Haehnelt, and A. Lewis (2006). The Lyman alpha forest and WMAP year three. *Monthly Notices of the Royal Astronomical Society: Letters 370*(1), L51–L55. arXiv:astro-ph/0604310.

Viel, M., M. G. Haehnelt, and V. Springel (2004). Inferring the dark matter power spectrum from the Lyman-α forest in high-resolution QSO absorption spectra. *Monthly Notices of the Royal Astronomical Society 354*(3), 684–694. arXiv:astro-ph/0404600.

Viel, M., M. G. Haehnelt, and V. Springel (2010). The effect of neutrinos on the matter distribution as probed by the intergalactic medium. *Journal of Cosmology and Astroparticle Physics 2010*(06), 015–015. arXiv:1003.2422.

von Hoerner, S. (1960). Die numerische Integration des n-Körper-Problemes für Sternhaufen. I. *Zeitschrift für Astrophysik 50*, 184–214.

von Hoerner, S. (1963). Die numerische Integration des n-Körper-Problems für Sternhaufen, II. *Zeitschrift für Astrophysik 57*, 47–82.

Wadsley, J. and J. R. Bond (1996). SPH P-cubed MG Simulations of the Lyman Alpha Forest. In D. A. Clarke and M. J. West (Eds.), *Computational Astrophysics; 12th Kingston Meeting on Theoretical Astrophysics*, Halifax, Nova Scotia, Canada, pp. 332.

Wadsley, J., J. Stadel, and T. Quinn (2004). Gasoline: a flexible, parallel implementation of TreeSPH. *New Astronomy 9*(2), 137–158. arXiv:astro-ph/0303521.

Wang, M.-y., R. Croft, A. Peter, A. Zentner, and C. Purcell (2013). Lyman-α forest constraints on decaying dark matter. *Physical Review D 88*(12), 123515. arXiv:1309.7354.

White, S. D. M. (1994). Formation and Evolution of Galaxies: Les Houches Lectures. arXiv:astro-ph/9410043.

White, S. D. M., C. S. Frenk, and M. Davis (1983). Clustering in a neutrino-dominated universe. *The Astrophysical Journal 274*, L1.

Woodward, P. and P. Colella (1984). The numerical simulation of two-dimensional fluid flow with strong shocks. *Journal of Computational Physics 54*(1), 115–173.

Xu, G. (1995). A New Parallel N-Body Gravity Solver: TPM. *The Astrophysical Journal Supplement Series 98*, 355. arXiv:astro-ph/9409021.

Yèche, C., A. Ealet, A. Réfrégier, C. Tao, A. Tilquin, et al. (2006). Prospects for dark energy evolution: a frequentist multi-probe approach. *Astronomy & Astrophysics 448*(3), 831–842. arXiv:astro-ph/0507170.

York, D. G., J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, et al. (2000). The Sloan Digital Sky Survey: Technical Summary. *The Astronomical Journal 120*(3), 1579–1587. arXiv:astro-ph/0006396.

Zaldarriaga, M. and U. Seljak (2000). CMBFAST for Spatially Closed Universes. *The Astrophysical Journal Supplement Series 129*(2), 431–434. arXiv:astro-ph/9911219.

Zeldovich, Y. B. (1970). Gravitational instability: An approximate theory for large density perturbations. *Astronomy & Astrophysics 5*, 84–89.

Zhang, Y., P. Anninos, and M. L. Norman (1995). A Multispecies Model for Hydrogen and Helium Absorbers in Lyman-Alpha Forest Clouds. *The Astrophysical Journal 453*(2), 14. arXiv:astro-ph/9508133.

Acknowledgments

J'en arrive maintenant à la partie la plus plaisante de la rédaction de ce manuscrit, qui n'aurait pu voir le jour sans le soutien d'un certain nombre de personnes.

Je tiens d'abord à remercier Nathalie Palanque-Delabrouille, pour m'avoir encadré tout au long de cette thèse et de m'avoir permis de réaliser la partie simulation cosmologique. Son enthousiasme et sa bonne humeur permanente ont assurément déteint sur moi et mon permis de franchir les obstacles avec sérennité. Au final, comme je l'ai souligné lors de ma soutenance, j'avais plus l'impression de travailler avec une collègue que d'avoir une directrice de thèse.

Au-delà de Nathalie, je souhaite également remercier l'ensemble du groupe cosmologie du SPP pour son accueil lors de ces trois années. L'ensemble de nos discussions, sur la physique ou non, m'ont sans aucun doute énormément enrichi. Merci donc à Christophe Y. pour m'avoir accueilli au sein du groupe, à Jean-Marc pour son calme à toute épreuve, à Étienne que j'espère un jour retrouver sur un course, à Jim pour m'avoir fait partager son immense culture, à Vanina qui fut une formidable marraine, à Christophe M. pour les longues discussions geeks que nous avons pu avoir dans son bureau, à Jean-Baptiste notre spécialiste CMB et enfin à Didier malheureusement arrivé trop tard pour que nous puissions travailler ensemble. Je n'oublie pas non plus nos compagnons de repas : Roy, Céline et Dominique.

Je souhaite aussi chaleureusement remercier Romain Teyssier et David Weinberg pour avoir accepté d'être les rapporteurs de ce manuscrit, ainsi que Julien Lesgourgue, Réza Ansari et David Elbaz pour avoir fait partie de mon jury.

Cette thèse ne se serait pas aussi bien déroulé sans l'appui de la collaboration BOSS et du groupe Lyman-α en particulier. L'appui fournit aux jeunes thésards ou post-docs dans cette collaboration m'a l'air sans pareil et contribue à mon avis fortement à son immense succès. Je tiens à remercier en particulier (en espérant n'oublier personne) Matteo Viel, Patrick McDonald et James Bolton pour leur aide et leurs explications.

Merci à mes parents pour leur soutien sans faille pendant ces onze années d'études. À ma mère pour avoir relu avec courage l'ensemble de mon manuscrit et à mon père pour tout ses bricolages et pour ce magnifique tonneau-bar.

S'il y en a qui m'ont réellement supporté pendant ces trois ans, il s'agit bien de l'ensemble de la famille Habib : Nadine, Daniel, Myriam, Jonathan et Yaëlle, je vous dois beaucoup (et même plus) et ma porte vous sera toujours ouverte, sachez-le.

Dans mon esprit, une thèse ne peut pas se faire sans être bien entouré. Un grand merci donc à tous mes amis pour l'ensemble des moments passés ensemble. Merci aux thésards du SPP pour leur bonne humeur et pour avoir supporté mes références militaire, merci Timothée, Léa, Nicolas, Pierre et Thibaut. Merci aux amis que j'ai pu revoir aux US : Frédéric/Copernic, Adrianne, Antony, Sonia, Alexandre. Merci à tout les participants des soirées jeux : Raphaël/Xelnor, Yann/Drynn, François-Nicolas/Fneuh, Hélène, Marie-Alix/Max, Maïlys, Cosme, Édouard/HP, Virginie/Vinie, Charles/Sheik...Merci aux horribles pour les soirées

BCG (binouzes, catch et gras) : Damien/ZoinX, Sébastien/V4, François-Baptiste/Françbat, Nicolas/Cortot, Adrien/Mekki. Merci à l'ensemble des membres de polytechnique.org : Thomas/Fulanor, Anne/Kithyane, Nicolas/Iooss, Bernado...Merci aux cavaliers de l'écurie al Andalus pour tout les bons moments passé à cheval. Merci aussi à tous les TOS que j'ai pu croisé au local Faërix et qui ont acceuilli comme il se doit le VC que je suis. J'ai aussi une pensée pour les jeunes mariés Manon et Guillaume.

Enfin, je tiens à remercier la Direction Générale de l'Armement pour m'avoir permis de faire cette thèse. J'espère mettre à leur profit la maturité et l'autonomie acquises pendant ces trois années.