Physics of the Intergalactic Medium: a Study of the Power Spectrum ot the Lyman Alpha Forest and the Metal Content of Damped Lyman Alpha Systems

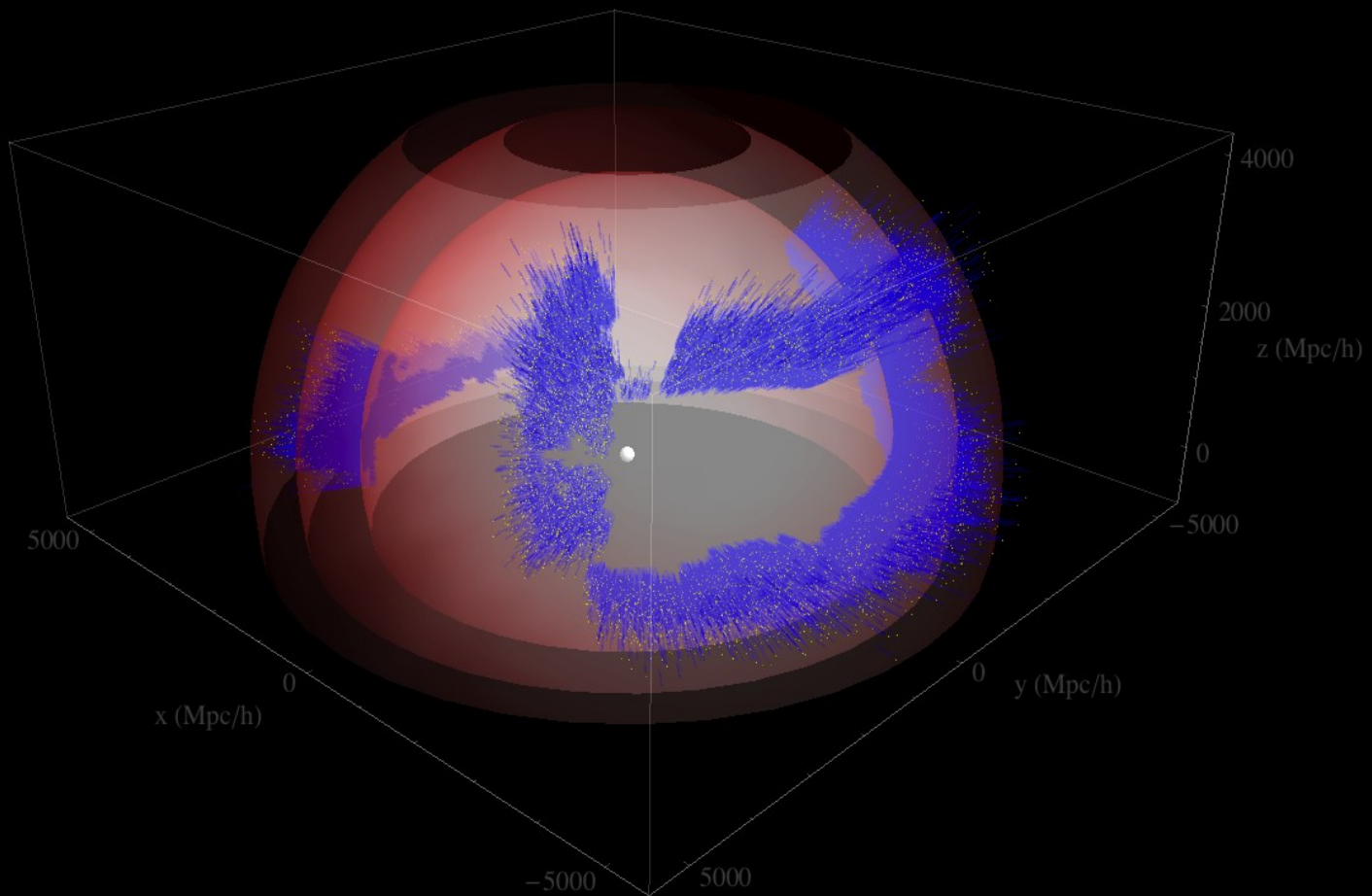Andreu Ariño i Prats

Physics of the Intergalactic Medium:

A study of the Power Spectrum of the Lyman alpha forest

and

The metal content of Damped Lyman Alpha Systems

Andreu Ariño i Prats

Director: Jordi Miralda Escudé

Universitat de Barcelona

DAM

IEEC

ICCUB

Autor: Andreu Ariño i Prats
Director: Jordi Miralda Escudé
Tutor: Alberto Manrique

Institut d'Estudis Espacials de Catalunya (IEEC)
Institut de Ciències del Cosmos (ICC-UB)
Departament d'Astromnomia i Meteorologia (DAM)
Universitat de Barcelona (UB)

Facultat de Fisica, Martí i Franques
08028, Barrcelona

Setembre 2014

Resum de la tesi.

En aquesta tesi presente treball relacionat majoritàriament en aprofundir al nostra comprensió del medi intergalàctic (IGM, Itergalactic Medium en anglès). L'IGM a grans trets correspon a tot allò que es troba entre les galàxies, com bé indica el nom. L'estudi d'aquesta tesi tracta de l'estudi del senyal que el gas neutre que es troba a l'IGM deixa en absorbir certes freqüències de l'espectre d'objectes més llunyans, com ara els Quàsars. Aquesta senyal, quan és produïda per hidrogen neutre, s'anomena "Bosc Lyman Alfa" (Ly\alpha forest en anglès), relacionada amb la transició Lyman Alfa de l'hidrogen. Segons la distribució del gas que es troba entre el quàsar i nosaltres, aquest deixarà una senyal característica a l'espectre del quàsar que ens permet un estudi detallat de els propietats del IGM.

Per a el detallat estudi d'aquest calen dues coses:

Primerament un extens sondeig observacional de centenars de milers d'espectres que ens permeten un gran mapeig del cel. En el nostre cas aquest estudi ha estat realitzat a la 3a fase del Sloan Digital Sky Survey dins del programa Barionic Oscilation Spectroscopic Surbey (SDSSIII-BOSS) en que s'extrau l'espectre d'aquests centenars de milers de quàsars.

En segon lloc és necessari un detallat anàlisi de tots els processos físics que puguen tindre algun efecte en el senyal que aquest gas neutre deixa en els espectres. Per a açò és necessari una bona comprensió de la teoria que explica les pertorbacions primordials i com aquestes acaben donant lloc al IGM. Per altra part simulacions que ens permeten tractar de modelitzar tots els complexos processos que ocorren en aquest medi i que donat a la seua natura altament no lineal i interrelació de moltes facetes, fan necessari de l'ús de càlculs numèrics només accessibles gràcies a les simulacions per ordinador.

En aquesta tesi doncs es tracten aquests 2 punts. Els quals composen la estructura de la tesi que es pot resumir a continuació:

Estudi de de l'espectre de potencies del Lyα forest en 3 dimensions per a escales no lineals amb l' ús de múltiples simulacions Hidrodinàmiques.

L'espectre de potencies descriu la probabilitat de trobar estructures de determinada mida, aquesta probabilitat bé directament determinada pel tipus de cosmologia que habitem, per tant entenent bé l'espectre de potències del IGM podrem extraure directament informació de la cosmologia.

L'IGM és després de el CMB (comsmological microwave background, en anglès) l'únic observable cosmològic capaç de mesurar directament l'espectre de potències ja que normalment jau en estructures que encara segueixen l'evolució lineal de la matèria. Però allò que finalment deixa traça en la senyal d'absorció depèn de física no lineal i d'escales més evolucionades. Es per això que per complementar l'estudi del IGM és necessari l'ús de simulacions Hidrodinàmiques. Les simulacions hidrodinàmiques són aquelles que tracten de simular la evolució del univers i on s'inclou no només la interacció gravitatòria de les partícules, si no també els efectes que els barions, partícules que interaccionen electromagnèticament, i tenen uns efectes molt rellevants en la formació d'estructura. Açò és transcendental per estudiar el IGM ja que els efectes dels barions, que formen a la fi els àtoms que deixaran la senyal del Lyα forest, són vitals per caracteritzar aquest medi.

Aleshores amb la col·laboració amb el Dr. Jordi Miralda Escudè i el Dr. Matteo Viel, i amb les aportacions de les simulacions realitzades pels Drs. Matteo Viel i Renyue Cen he procedit a

extraure i detalladament caracteritzar l'espectre de potencies del camp d'absorció deixat pel Lyα forest anant fins a escales no lineals. Un cop extret i desenvolupat el mètode per a fer-ho sistemàticament i de forma equivalent per a múltiples simulacions, hem produint un model per tal de descriure aquest espectre de potencies, que permet comparar característiques i prediccions per a cada una de les simulacions.

Estudi i evolució de la "Força dels Metalls" associats als sistemes Lyα Esmorteits.

Els sistemes Lyα Esmorteits (Damped Lyα Systems, DLAs en anglès) són sistemes d'alta densitat de columna presents al bosc Lyα els quals estan totalment saturats i tenen tan alt contingut d'àtoms neutres que absorbeixen longituds d'ona pròximes a la del redshift que els correspondria deixant senyals característiques que permeten calcular la densitat de columna d'aquestos sistemes. Aquestos formen part del IGM i permeten traçar aquest i les l'estructura de la que formen part. A més juguen un paper molt important el la formació de galàxies.

Coneguda la posició del DLA ens permet estudiar on trobar absorcions de línies espectrogràfiques corresponents a especies metàl·liques (altres àtoms al gas diferents de hidrogen i heli), cosa que ens porcionsa importants característiques dels DLAs.

Aleshores utilitzant l'ampli catàleg de BOSS amb he pogut accedir a més de trenta-mil d'aquestos objectes, i per a cada un d'ells mesurar la "força" o amplada equivalent (equivalent width en anglès) per a una serie de línies metàl·liques gràcies a l'ajut del Dr. George Becker. Un cop fet un nou catàleg d'aquestes línies metàl·liques hem desenvolupat una nova parametrització d'una propietat dels DLAs que hem anomenat "Força Metal·lica" (Metal Strength en anglès). Aquesta no és més que una suma pesada de les amplades equivalents mesurades per als metalls de cada DLA.

Aquest nou paràmetre em permet ara classificar el gran nombre de DLAs, aquesta classificació l'he fet servir per estudiar per una part agrupacions de DLAs, feta amb col·laboració amb en Luís Mas. Organitzant els sistemes depenent de la seua "Força Metàl·lica" permet l'estudi de propietats globals dels DLAs, com per exemple com evoluciona la metal·licitat i la dispersió de velocitats d'aquestos objectes, a més permet observar metalls mai abans vists en els DLAs. També puc vore com evoluciona la "Força Metàl·lica" amb redshift o edat del univers, sent aquest un nou i valuós observable que es pot utilitzar per comparar amb models i simulacions del univers que tracten d'estudiar els DLAs.

Finalment aquesta classificació es pot utilitzar per vore com esbiaixats, com de diferents respecte l'estructura de la matèria, es poden trobar aquestos sistemes depenent de la nova parametrització. Podent obtenir informació cosmològica d'ací.

Physics of the Intergalactic Medium:

A study of the Power Spectrum of the Lyman Alpha Forest

and

the Metal Content of Damped Lyman Alpha Systems.

Doctoral Thesis
Director: Jordi Miralda Escudé
Tutor: Alberto Manrique

Committee:
Jose Maria Solanes Majua,
Matteo Viel,
Nathalie Palanque Delabrouille.

Andreu Ariño i Prats

Contents

Acknowledgements

First of all I need to thank to this mysterious and full of diversity universe for existing. We still don't understand why there exists nothing at all, we begin to have some some hints in why the cosmos is the way it is but still many other universes and set of natural laws could be possible. Nevertheless one of the main properties of the world is that it has time, another aspect of the world that we don't fully grasp, and that the time has a preferred direction. Then in this universe we understand that there were some primordial perturbations that when they grow with the time and give rise to all the structure that now we can see around us. Specifically, from the remains of another star this wonderful spherical world in which we live was formed. On it, thanks to a process that we don't yet understanding, complex molecules that could reproduce themselves produced all the living diversity that exist in this little world. This was achieved following a simple set of logical rules that gave rise to evolution.

Among this diversity of living things I need to thank the the level of ingenuity and curiosity that is a defining part part of the humans beings. This trough time allowed to develop and discover a necessary tool and framework to decode the world: Mathematics. Mathematics are essential to describe, organize and understand our world, from the universe to the humans. Even more, math can create self contained worlds and structures on itself, that we still have to understand their links to our universe. Humans not only created math but also thanks to the rise of culture and society we have been able to develop many complex and diverse levels that allow for the sustained exploration of our world. Technology is also a intrinsic player of that endeavour, it is completely interlinked with the evolution of society and culture give rise to the Scientific Method. The scientific method is an amazing framework optimized for understanding the world, that has achieved all the breakthroughs in our understanding of the world in the past four centuries and that is the base of all modern society.

The same society, culture and technology evolved through time has allowed for a sustained and growing investment in doing scientific research, this has allowed me to follow all my studies and later research to pursue my aim of understanding the world, since all my studies have been paid by the rest of the society. Trough the way of getting to this point in my understanding I crossed my path with several people that helped me to large lengths on that direction. In the context of this thesis firstly I need to mention my supervisor Jordi, his extensive knowledge and experience, and continuous curiosity has always surprised me and offered me much to learn, specially from the resolution of physical problems. He made us, by the means of our basic knowledge and ideas, be able to find a first other solution of a given question of the nature. That is undoubtedly a great training for the mind and for a young learning researcher.

Then my scholarship, paid by the whole Spanish society, offered me the great opportunity to visit other centres of research, work with other people and know other researchers. My fist experience was in Trieste, where I worked with Matteo, he welcomed me in a magnificent fashion and showed me another way of working and living the research, his unique energy and passion where great to see in a fellow researcher. Also all the people that I met at Trieste soon accepted me as part of their group and let me to experience a rich scientific and social environment that made of my time there a wonderful experience. I later enjoyed a stay in Cambridge, where George welcomed me and showed me his support and guided me trough my first experience with observational data. Cambridge also offered me a wonderful scenery of history and rich field of many scientific research and topics at the tip of your hands, or bike. Finally I enjoyed a great stay in Heidelberg thanks to Khee-Gan that introduced me to

modelling of spectra. There I got the chance to be within the Enigma team, which was a very young and active group, working in many fields, sharing knowledge and always questioning any research done.

Then the time working in this thesis wouldn't have been the same if I didn't get to share my office with an amazing set of hobbits. Firstly Ignasi, we had in common the curiosity and energy that allowed us to organize many things in the department, and we increased our knowledge in informatics and physics asking endlessly each other any sort of question. Victor, that we captured in our office and he captured us in his climbing, he always has enlighten us with his great sentences, his passion for life and random usually enjoyable music. Lluís is an oddball that appeared in our office, a really experienced person in life that opened our world to the exterior of the research bubble with each of his magnificent stories, also his neverending energy and finally it was great to share one of my research projects with him. The final incorporation Satya completed the office and gave new dynamics and ideas to our little group the hobbits, that shared passion for cookies, and with our memorable and interminable discussions about food.

This group was a small sample of our terrace loving people. Only one students office enjoyed access to a terrace, but the terrace inhabitants allowed the rest of us (even the sad windowless like the hobbits) to share that magnificent place to enjoy the time, food and experiences. This provided the perfect environment to try to organize many activities and share valuable information of our PhD life.

Then the department is nothing without its people, some researchers joined the neverending curiosity of the cosmology fringe group and allowed us to use them to share their research and knowledge in a weekly basis. Part of that department is our IT guy, Gabi, that always has some great T-shirt to show or some weird trivia or odd feature to enlighten your day. And one has to always thank the invaluable effort of your administration, especially JR from the DAM that seems that has magical powers to solve any problem, Ariadna from the ICCUB always ready to help, and Anna in the IEEC never failing you with the hellish paperwork.

1. Introduction: Our place in the cosmos and tools to study it

1.1. THE HISTORICAL DEVELOPMENT OF COSMOLOGY

1.1.1 Early developments on cosmology

Humanity has long been wondering about our place in the universe. This has led us to always been asking about the nature of what we observe, and therefore about the physical laws governing the material world and the nature of the universe itself. Through the development of the Scientific Method we have been able to reliably answer many questions. However, the scientific method as understood today, where alternative hypotheses and theories are proposed which can make quantifiable predictions, which are then carefully tested against observations, was born in a decisive way only a few centuries ago. Previously, some advances on our understanding of the shape of the Earth and its place in the cosmos had occurred at a very slow pace: the size of the Earth was measured by Eratosthenes from the distance covered on the ground that corresponded to a certain change in the angular position of the Sun or a star, and the size and distance to the Moon was determined from the size of Earth's shadow during a lunar eclipse, and the diurnal parallax measurements of Hypparchus during the second and first centuries BC in classical Greece.

To reach a step further and to understand the nature of Earth as a planet of the Solar System, substantial progress did not occur until the Middle Ages had passed, when the heliocentric model was carefully studied by Copernicus in the late 16th century and defended by Galileo thanks in part to many observations revealed by the use of telescope. This event was the one of the main drivers of the Scientific Revolution. This revolution was further strengthen thanks to accurate observations of the motion of the planets by Tycho Brahe that gave rise to the discovery of the laws of motion of planets in the Solar System by Kepler in the early 17th century. All this was solidly grounded into scientific demonstration when Newton proved Kepler's laws from his Universal Law of Gravitation at the end of the 17th century. Newton's theoretical tour-de-force was able to explain such a large number of new phenomena (planet perturbations on their motions which result in corrections to Kepler's laws, the precise

motion of the Moon, Earth's ellipsoidal shape and its precession rate...) that no one could seriously doubt any longer about the structure of the Solar System, even though an actual observational proof of Earth's motion around the Sun arrived only in in the fist half of the 18 century with the discovery of the aberration of the light of stars by Bradley.

Our understanding of the cosmos that we inhabit went through its next revolution with the comprehension of the nature of next larger structure that we are part of, the Milky Way. Observations from Wright pointed to the Milkyway being a rotating disc of stars sustained by gravitation in the mid 18 century. Its shape was later confirmed by Hershel at the end of that century to be a rotating disc of which the Solar System is part.

Finally about one century latter we reached the position in which we were able to study nature of what lied outside the Milky Way galaxy. This was possible thanks to the steady development of both observations and theory that gave rise to the field of Cosmology. This is the topic that I am involved, the study of the universe, which has become a proper branch of the science tree since little less than a century ago. With it came the beginning of a clear picture in our long journey of understanding the nature and place in the Universe.

The theory that allowed for first time to describe the universe as a whole is one that accounts for the behaviour of the dominant force on large scales: gravity. Thanks to General Relativity understanding gravity as the relation between mass and the shape of space time allowed the precise description of the gravitational phenomena, this was contained in Einstein's Field equations developed in the 10s,

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} - \Lambda g_{\mu\nu}, \tag{1.1}$$

which explained briefly basically relates the geometry of the universe $G_{\mu\nu}$ and $g_{\mu\nu}$ (how space-time is bent) with its matter and energy distribution $T_{\mu\nu}$.

Soon the GR framework allowed to develop equations that provided a description of the evolution of space-time and its content as a whole, therefore it gave the basis to describe the universe and allowed to develop the first predictions of how a universe could evolve, assuming it is homogeneous and isotropic. Friedman, Lemetre, Roverson and Walker (FLRW) in the early 20s and 30s developed differential equations and metric that account for the scale factor of the universe (a is the scale factor of the universe normalized to 1 at the present day), its energy density ρ and pressure p:

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2} - \Lambda c^2 = -\frac{8\pi G}{c^2} p, \tag{1.2}$$

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2} - \frac{\Lambda c^2}{3} = \frac{8\pi G}{3}\rho. \tag{1.3}$$

These equations allowed various evolutionary paths depending on the balance of the components such as p, ρ, curvature (k), cosmological and constant (Λ) that are intersect characteristics of the universe; $\dot{a}$ is its temporal evolution, i.e. how fast is increasing or deceasing in size and $\ddot{a}$ its acceleration; c is the speed of light; and G is the constant of gravity. The universe could be ever expanding, contracting, it was demonstrated early on that it could not be static i.e., not evolving in time, as that solution is unstable if the universe has any minimal

inhomogeneity, as it is the case. Observations should then follow to discern between these various possibilities.

Astronomical observations came from much earlier than just century ago. They have always been linked one way or the other to our wondering about the cosmos. They have been continuously awakening our craving for understanding the world, and constitute the main source for the scientific revolution at the end of the medieval age. Ever since then they have been revolutionizing the understanding of the cosmos we live in, to higher or lower degree. Technology has also played a key role here, the development of optics, photography, spectrography many technologies that have come later allowed us to peer further and further in the universe. To be able to discern this enormous and expanding universe not only better telescopes where needed, but also a way to measure distances. With the help of photography applied to astronomical observations, Cepheids, stars with a constant period of oscillation of their brightness, could be used to measure distances once its periodicity and brightness was measured and it was inferred that they were tightly related (Leavitt, 1912). Using this and powerful telescopes such as the Mont Palomar, Cepheids stars could be observed in Andromeda, measuring then its distance and placing it very far from the Milky Way, Andromeda was found to be a galaxy on its own (Hubble, 1924) closing definitely the *Great Debate* on whenever the Milky Way was the Whole universe or just one galaxy out of many. To observational cosmology this was its main inception, for first time we could investigate the oceans of space that lied outside our galaxy.

Later on, not only distances but also velocities were measured for distant galaxies through the Doppler effect. It was discovered that the galaxies were moving away from us in all directions, and the further they where the faster they were moving from us (Hubble, 1929). This is how a law that related velocity and distance of distant galaxies was discovered. Hubble's law

$$v = H \cdot d, \tag{1.4}$$

where v is the velocity measured, H is the rate at which the universe is expanding, and d is the distance to the galaxy, allowed to compute distances to distant galaxies by simply measuring their velocity. Since in an expanding universe the galaxies are receding from us, the wavelength of the light that they emit is shifted to the red due to Doppler effect, how much their wavelengths were shifted to the red gives their redshift which is directly related to their velocity with respect us.

This placed us in an ever expanding universe in all directions. All the matter was moving away from each other if the *Copernican Principle* (that we are not in a special place of the universe) was followed. In the universe that meant that space itself was expanding. The evolution of the expansion rate with time($H(t)$) allows to rewrite equation 1.3 as the dependence of the expansion rate on the different density of the ingredients that define the universe,

$$H(t)^2 = \left(\frac{\dot{a}}{a}\right)^2 = H_0^2(\Omega_R a^{-4} + \Omega_M a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda), \tag{1.5}$$

where H_0 is the expansion time at present time, Ω_R is the radiation density , Ω_M is the matter density, Ω_k is the curvature curvature of the universe, and $\Omega_\Lambda = \Lambda$ from equation 1.1

FIGURE 1

Velocity-Distance Relation among Extra-Galactic Nebulae.
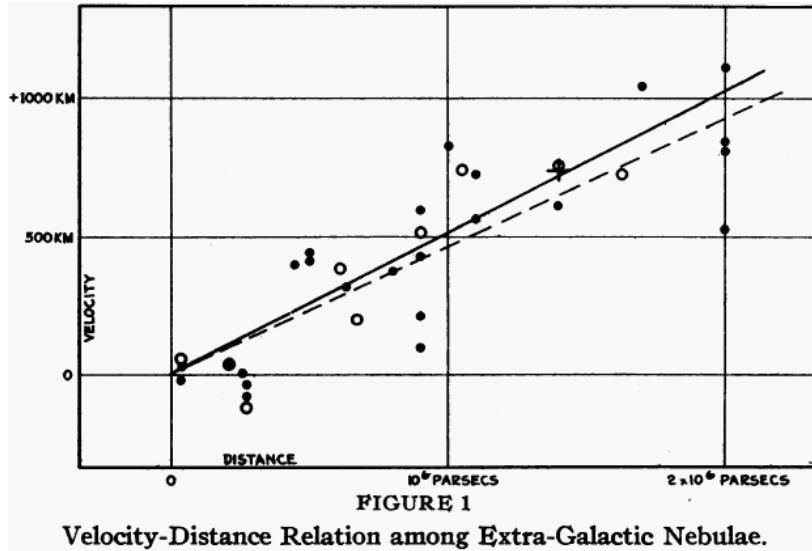
Figure 1.1: Original Hubble diagram where redshift and distance of galaxies is fitted by the simple equation 1.4.

is the cosmological constant, all the densities are values at current cosmological time. These densities are referred to the percentage of critical density necessary for the geometry of the universe to be flat.

$$\Omega \equiv \frac{8\pi G\rho}{3H^2}. \tag{1.6}$$

This provided vital clues for the theoretical models and the advent of theories like the Big Bang which modeled the universe as the outcome of an explosion that expanded universe from a really compact and really hot and dense plasma to the universe that we inhabit today, cooling as it expanded, forming all the structure that we see in the universe in the process. From that theory, the moment of this expansion started is taken as the moment the universe was born. The rate at which the universe is expanding is linked to the age of the universe, the faster it is expanding the shorter the time where all the universe was in hot dense plasma phase is. Therefore measuring the expansion rate combined with our theories of the universe allows to compute the age of the universe,

$$t_0 \sim H_0^{-1} \tag{1.7}$$

this will be important to test our theories and understanding and will impulse further developments on cosmology.

1.1.2 Main features of our cosmological model: Nucleosynthesis, Dark Matter and the primordial Power Spectrum

Early on, the collaboration of theory observations and technology gave rise to an essential part of the Big Bang theory. A combination of high energy physics and the expansion of the universe allowed to account for the ratios of light elements (Hydrogen isotopes, Helium isotopes, Lithium) measured in the local universe with the spectrographs of the epoch. This

allowed Primordial nucleosynthesis to be developed (R. Alpher 1948). Matter and antimatter should be in equilibrium, but for the universe to have matter in it, matter should be more abundant than antimatter. Computing how much of one there is over the other, or the ratio of radiation to matter, allows to make very accurate predictions of the abundances of the light elements. When compared with the observations of the light elements measured in gas and stars around us the prediction matched them. This was later refined and measured at various epochs, which combined with fine details of other theories allows the observations to match the predictions really nicely.

The next crucial ingredient on building the SCM was discovered in a paradigmatic example of how observations push theories forward. This is the fact that non baryonic Cold Dark Matter (CDM) constitutes an important fraction of the budget of the present universe. Curiously enough observations about the need for non luminous matter to account for velocities dispersions on stars in the Milky Way exist as early as the 20s (J.Jeans and J. Kapteyn 1922), and also in the 30s to explain the large velocity dispersion on galaxies in the viralized coma cluster in the 30s (F. Zwicky 1933). This apparent problem went unnoticed on the community for a long time (it has to be said that the non barionic nature of this matter was not introduced at that time). Until when in the 70s the rotation curves of many galaxies showed a non Keplerian motion with distance (Rubin and Ford 1970), those were consistent with the presence of matter not being observed. The addition of observations led to realize that the mass to light ratios increased as you observed larger structures. (Einasto, Kaasik, and Saar, Ostriker; Peebles, and Yahil 1974). A direct measure of Ω_m and Ω_b , those being very different, further indicated the presence of mass that was not accounted for, and that this mass could not be baryonic on nature. Many other observations afterwards pointed in the same direction, like X rays in clusters where emission depends on the mass, lensing where mass bends the light in the background, and also from the inhomogeneities on the CMB. These inhomogeneities should be smaller if decoupled DM was not present, since the presence of it modifies the epoch of reionization and therefore the growth of structures to that point. All of this soon led most of the community to accept that a big part of the content of the universe is non baryonic matter. The nature and properties of this matter will be, and still are, further explored, but nevertheless CDM constitutes on of the main ingredients of the CSM. The fact that it is estimated to be *Cold* (with small velocity dispersion) will take a long process in itself, and further developments on the cosmology field that continue until present day.

Then another paradigmatic example essential to the SCM is how a model predated observations. This is the case of the Cosmic Micorwave Bakground (CMB) and its properties. As introduced, from the Big Bang Nucleosynthesis a radiation remanent was already predicted arising from a time of phase transition, where matter and radiation came from being in equilibrium to be decoupled. This left a residual background of a black body at a characteristic temperature. It was eventually measured as a uniform microwave background with the spectrum of a black body that had cooled to a temperature of 2.7k (Penzias and Wilson, 1965). This large uniformity was one of the motivations of another theory that has become an important part of the CSM.

It is difficult to explain why the universe looks so similar in all directions, not only on the CMB but also large galaxy surveys. That uniformity would meant that as some point all parts of the universe where in causal contact, that would explain why they share the same characteristics of temperature and density. However this cna only be produced if the universe
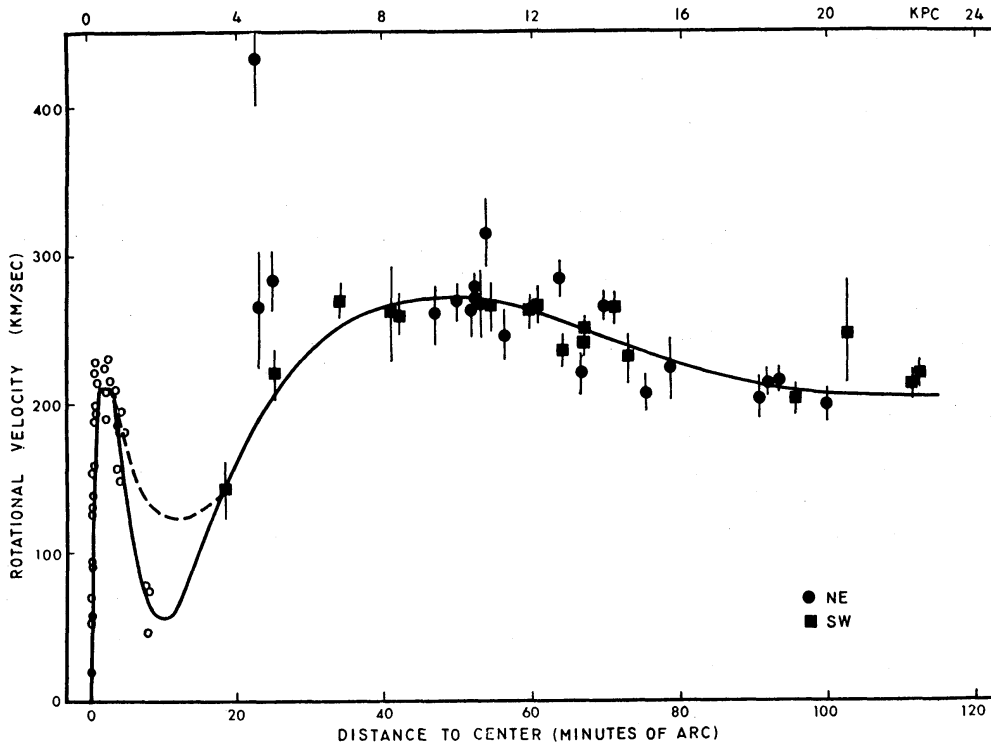
Figure 1.2: Original rotation curve of andromeda as measured by Rubin and Ford, 1970, where the velocity with radius does not follow a keplerian low 1.4.

was perfectly uniform since the beginning or that a ray of light had time to cross the entire size of the universe several times to exchange energy from one point to the other and thermalize it. Both options are disfavoured by theory and computations. Since its unlikely that only one state is chosen of all the possible states that each part of the universe can be. And from computations it was measured that a photon could not travel faster than the expansion of the universe, that means that horizon of the universe (how far can a photon travel since the origin of the universe) has always been smaller than the universe itself , therefore being unable to thermalize it at any point on its evolution. This could imply an extremely fast expansion at the beginning of the universe, where a really small part of the universe that was inside a common horizon expanded exponentially, stretching this uniform part of the universe so fast that big inhomogeneities could not be formed. (Zeldovich, 1978, Starobinsky , 1979; Guth 1980, Linde 1981, 1983). This is the inflation theory that can explain the horizon problem as well as others (like the flatness of the universe, where the density it is very precisely at its critical value, or originally, why we do not observe magnetic monopoles if they were created in the Big Bang). Therefore the theory of Inflation advanced a series of solutions and predictions that are now in the verge to be tested thanks to among other things, the polarization of the CMB (Plank collaboration 2014, BICEP2 collaboration 2014). From the CMB came the measure of the inhomogeneity of the CMB, its expected size and shape. These have been in accordance with predictions from the origin of the perturbations and its growth in the primordial universe (Bardeen 1980, Kodama and Sasaki 1984; Thorne 1980, Ellis, Matravers and Treciokas 1983). These inhomogeneities where first measured, when the COBE satellite was launched. Out of it, the cosmological model came in really strong shape. Those measurements have been

refined with exquisite precision on later missions of WMAP and Plank satellites.

1.1.3 The detection of Dark Energy and its addition in the cosmological model

The field of cosmology can not be understood without the aggregate of many contributions. An essential part of this aggregate comes from numerical and computational power, the advent of current cosmology is definitely linked to the advent of more and more advanced computers. This comes in the form of simulations, crunching of big data, analytical models and methods to analyse and describe the universe. This high amount of computational power is needed due to the nature of the problem studied, the complex dynamics and high amounts of data needed. All of this and the computer to process that are now available. After all, a modern cosmologist is like a cook that tries to reproduce a dish that can only be seen when served. You do not know how it is made but you need to guess the set of ingredients and steps that are necessary to cook it from the final complex result with all the information and tools that you have. All previously stated is just the discovery of the ingredients and the invention of the tools required to be able to cook our universe again. In this frame simulations are just the tools to actually cook the universe and taste it. They allow to see if the result matches real characteristics that are being observed or we are missing something. In that way computers have been used since the very beginning to try to work out the complex problems at hand, from simple dynamical evolution of dust like particles bounded by gravity to the ambition of fully simulating a universe with most of its main features evolving at the same time. This for example allowed to pin down the thermal characteristic of DM, if it was Hot, Warm or Cold; Cold being preferred by the results of simulations matching observations, although many other details had to be work out.

When analysing the ingredients necessary to reproduce the universe one comes to the fact that Cosmology is the curious case of a science in which the ingredients fundamental to it are not know, only its most basic characteristics are used. This is the case for the "dark" sectors that form the basic frame of cosmology. Non baryonic CDM was already introduced, the nature of which is highly hypothesized but so far (as the writing of this thesis) is not understood. Only its gravitational effects on ordinary matter and light are known. In the same way the last main ingredient for the CSM is another dark entity. At the end of the 80s a predicted slow down on the expansion of the universe due to gravity in the most evolved universe in the accepted Einstein de-Sitter model was not being observed. That would have repercussions on the age of the universe, and the growth of the structures within it. The discrepancy between model and observation where pointing that the characteristics of the late universe could not be explained by the simple model of big bang expansion and gravitational slowing down, but yet the models where not trustful enough to accept that conclusion. Finally just before the end of the millennium the final and definitive push to the shift of paradigm came thanks again to a technique to accurately measure distances further away than before. This allowed to characterise the expansion of the late universe. This method used supernovae as standard candles to measure distances. A specific kind of supernova was studied and understood well enough to relate its luminosity to its distance (Phillips, 1993) thanks to an intense supernova survey (Caln/Tololo 1989-1995). Once an accurate use of these standard candles was ready, two independent groups used them to measure the distances to many galaxies. From that they discovered that the universe was not only not slowing down but has to be expanding faster as it evolves. Thanks to this, an accelerated phase in the evolution of the universe could be directly measured (Riess et al. 1998 and Perlmutter et al. 1999).
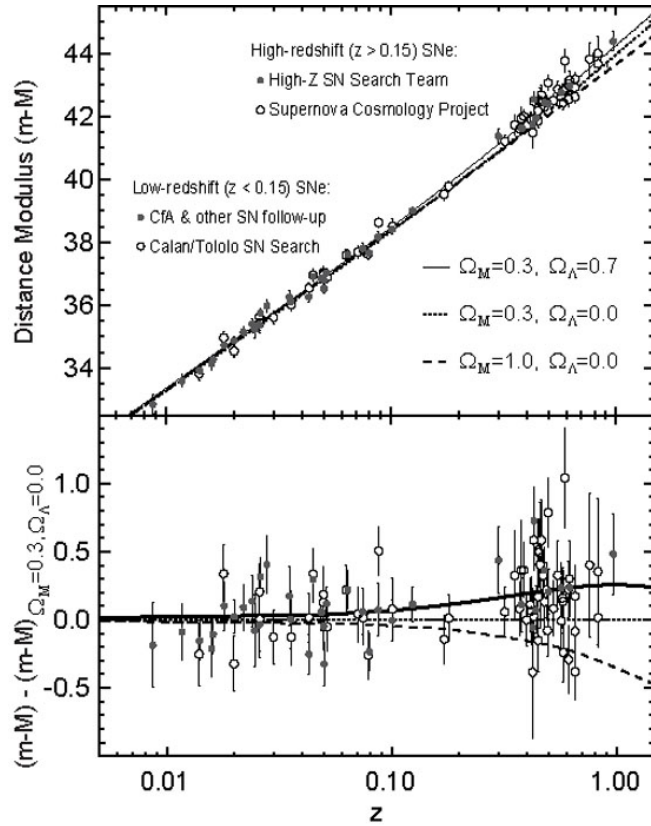
Figure 1.3: Original diagram similar to Hubble one where the best fit is a cosmology with cosmological constant, as measured by Riess et al, 1998.

The universe has to be expanding faster as it grows older to match the observations. That expansion could be understood as a cosmological constant arising from an integration constant of the field equations of general relativity. Cosmology was back at the beginning, closing an uncertainty that existed since the inception of cosmology, the value of the Cosmological Constant. Most importantly, if this was 0 or different from 0. A positive value could be understood as the pressure of the empty space, repealing all the parts of the universe from each other and therefore acting opposed to gravity. This really answer arose many interesting questions, but most importantly for this discussion a new and essential ingredient to build our universe was firmly discovered, this is the era dominated by Dark Energy (DE) both in the history of the universe and the history of cosmology.

Since the addition of DE, constituting most of the budget of the present universe, the ΛCDM model has dominated the scene as the Standard Cosmological paradigm. It has been really successful in explaining many of the elements of our universe and provides . But more importantly, it has put forward predictions that can be tested, and have been successfully so in the last years. Is during this time that cosmology has seen the biggest wealth of data until now and it continues to grow at extraordinary rates. As said, two more satellites to study the CMB where launched, WMAP and Plank tested finer and finer details this snapshot of the early universe. This holds an enormous amount of information of the epoch of reionization, also mixed into it there is information from before and after reionization, putting tight constrains on many parameters of the model. Ground observations of the CMB at smaller angular scales
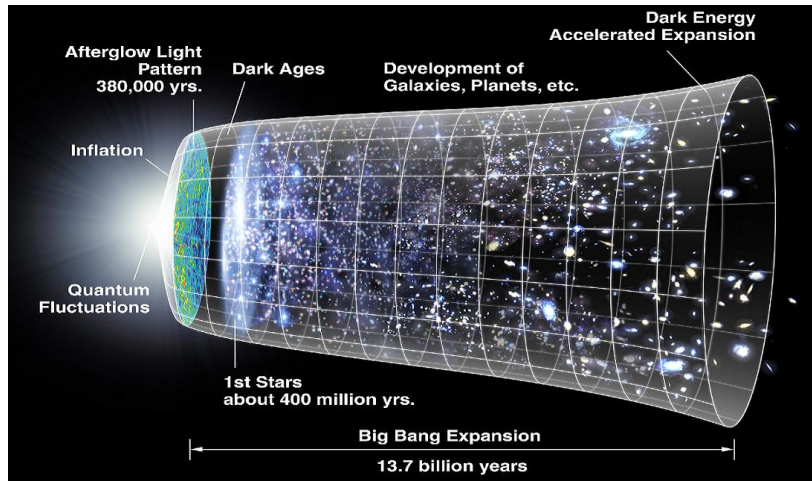
Figure 1.4: Schematic of the Cosmological standard model after the addition of dark energy.

also can test other characteristics of the universe. Other instruments like the Hubble Space telescope continue to push the frontiers of what can be seen on the early universe to validate or eliminate models of that epoch. And finally the most important new observational technique that has seen this decade is, in the same way that SN are used as standard candles, the use of a standard ruler that allows to characterize the universe at various epochs. This ruler is the Baryonic Acoustic Oscillation (BAO), an imprint on the structure of the universe itself. Then as long as you can observe large spans of the universe you can measure the BAO. Therefore since its size (but also its shape in more complex ways) depends on the characteristics of the cosmos, a wealth of data can be extracted from it.

Finally the ΛCDM framework allows to place all the main elements that shape our universe. Thanks to the observations and tools and understanding gained in the last few years we are in the process of grasping down many of the details of the universe that we live in.

1.2. THE INTERGALACTIC MEDIUM

Going through the history of cosmology we have seen that our knowledge in cosmology has been achieved by to the combination of many various disciplines and methods and observations. In this thesis I focus on one of the research fields that are most related to our understanding of the universe, the characterization of the intergalactic medium (here after, IGM).

Of the palette of tools that cosmology are available to study the universe, the IGM is one that is convenient and versatile. The study of the IGM provides with a view of the large structure of space at different time, ranging from the local universe to the epoch of recombination. This allows to study from very small objects that leave a clear imprint in our observations to large linear scales that can be traced by correlations between different parts of the IGM. Since the IGM allows to measure the structure of the universe, its careful analysis allows the measure of cosmological characteristics at various and interesting epoch of the universe, obtaining valuable cosmological information at different times. However this structure of the universe is more prominent in our observations the denser and more compact

it is, meaning it is in a very evolved state, no longer following the linear behaviour that best traces the cosmology that defines the universe. Also the obsrvations depend on non linear physics like thermal and pressure processes. Therefore it is of high importance to understand all the phenomena and characteristics of this structures that although might be small compare to cosmological scales are basic to extract information about the nature of the universe. And not only that, the IGM plays a major role on the galaxy formation, these dense parts of the structure of the IGM are the nurseries of the galaxies, and provide the material for galaxy formation. This way the IGM goes from small amplitude cosmological primordial perturbations of big size that evolve in to big amplitude perturbations but of small size that end up forming stars galaxies. Then to reconstruct the small perturbations on large scales on the primordial universe we make use of the small dense evolved structures that constitute the most prominent observation from the IGM . Summarizing to have a complete view of the IGM we need to understand the small to comprehend the big.

1.2.1 Our understanding of the Intergalactic Medium from the absorption in spectra

In the same way that cosmology has had its own development an history the IGM also has been shaped trough time until we achieved our current knowledge. In the 50s from fist principles it could already be predicted that there should exist a gaseous environment around galaxies (Spitzer, 1956). This mixture of mainly hydrogen and some helium should be ionized due to the radiation emitted by the stars residing in the galaxies. However this ionized plasma should be in thermodynamic equilibrium, meaning that some fraction of it should capture and electron and pass to a neutral atomic gaseous state again. This predicts a certain amount of neutral gas $n_I = x \cdot n_H$ in this ionized medium, were n_I is the number density of neutral hydrogen (I comes from an electron being in the 1st atomic level), n_H is the number density of the total hydrogen, and x is the neutral fraction, then

$$n_I \Gamma_i + n_e \cdot n_I \cdot \gamma_c(T) = \alpha_A(T) \cdot n_e(1-x)n_H, \tag{1.8}$$

where the first term accounts for the photoionization, being Γ_i the photo-ionization flux due to the astrophysical background of light; the 2nd term is the collision ionization, n_e is number density of electrons and $\gamma_c(T)$ the collision parameter, dependent on temperature. The rightmost term accounts for the recombination of the ionized hydrogen, being $\alpha_A(T)$ the photo-ionization factor, also depending on the temperature. From this, knowing the gas temperature, Γ_i and the n_e at each place and time, an estimation of the neutral fraction can be found. Once it is known that there is a certain amount of neutral hydrogen in the IGM (independently predicted by Gunn and Peterson, 1965; Scheuler, 1965; Schklovoski, 1965), that means that there are observational implications. Unlike ionized plasma, atomic gas strongly interacts with light, therefore it has to have an observational signal that can be detected. Again the technology helped here, the spectrographs and telescopes where evolved enough to be able to observe objects far enough that would have redshifted to the visible range the ultraviolet light absorbed by the Hydrogen Lyα transition of Hydrogen (1215Åat laboratory at redshift 3 would be at 4860Å, $\lambda = \lambda_0(1+z)$, well within the visible range of 3900 to 7000Å). Then it was expected that gas at high enough redshift between a luminous source and us would leave a characteristic absorption feature in the spectrum of that source bluewards of its Lyα emission. When a photon of the source is emitted it gets redshifted as it travels
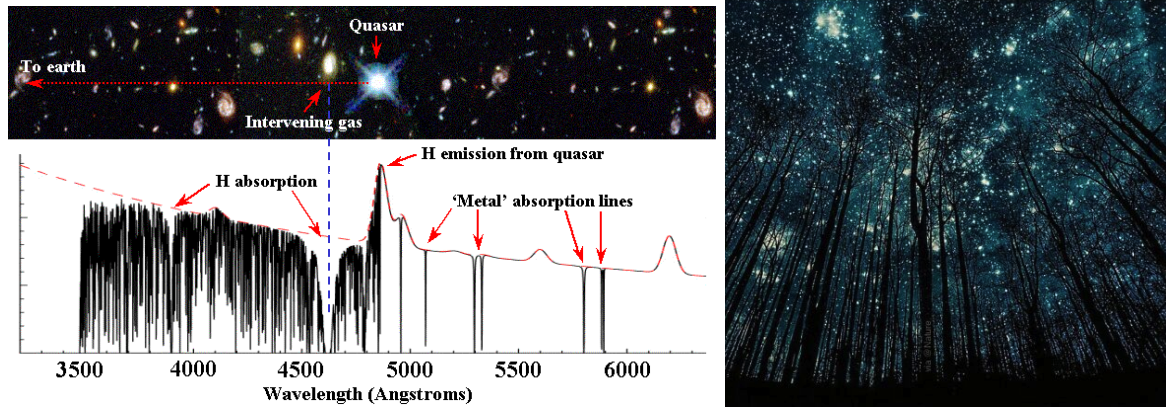
Figure 1.5: A didactic example of the looks of a quasar spectrum, with the absorption in the form of the Lyα forest and a Damped Lyα system (DLA) and its metals. This is compared with the absorption of a real forest, seen on the right.

the universe, as tehy get redsiftet they will have the frequwency of the Lyα transition of the neutral hydrogen gas that they cross and will be absorbed at that redshift, if al the IGM had a big enough fraction of neutral gas thet would produce he complete abortion all the photons in a section of the spectum, this feature was called *Gunn-Peterson Trough*. Then observations of the spectra of Quasars (Active Galaxy Nucleus that are the most bright objects in the universe) that found that the spectral region that should interact with the neutral Hydrogen was not completely absorbed, this allowed to put upper limits to the amount of intergalactic neutral Hydrogen (Gunn and Peterson, 1965).

Soon after it was also predicted that there should exist clumps of gas around galaxies that would produce narrow absorption features corresponding to narrow redshift space (Bahcall and Salpeter 1965). The theoretical predictions impulsed to observe even more objects with higher resolution and soon discrete absorbing lines where discovered (Lynds and Stockton, 1966; Burbidge, 1966; Kinman, 1966), and when their higher order Lyman series where detected it was proven that the absorption was indeed due to *Neutral Hydrogen* Lyα (Baldwin, 1974).

Once higher resolution spectrograph came at hand it was seen that there was hundreds or thousands of fine absorption lines in almost al the spectra analysed, small features only ranging few Angstroms. This was called the *Lyman α Forest* thanks to the narrow lines being similar to a densely wooden forest (Weymann 1981). The nature of this forest was open to speculation. Various models arose on the true nature of the medium that produced these fine absorbing features, but in principle non could be proven or disproven. At first only statistical observations such as relative abundance, frequency and distribution on the few spectra at hand was done in the 70s and 80s thanks to new telescopes and instruments. This allowed to set the basis for the nature of this absorbers as representing a new class of astrophysical object: intergalactic clouds with much larger rate of incidence and weak clustering than galaxies and metal absorbers (Sargent, 1980).

Finally thanks to some models put forward a clear picture of the nature of the medium that gave rise to the *Forest* started to come into shape. This is the Intergalactic Medium, eventually it came to be seen as a continuum medium with more or less linear perturbations over a mean. This perturbations could reliably explain the sharp and narrow features of absorption of the forest. This picture falls well inside the framework developed by the DM model, when the

first analytical models and simulations of dust like particles where done it was seen that they formed flattered structures and voids, but also knots and filaments. These simulations even though were not done for baryons, produced the Lyα structures as a byproduct of the CDM structure formation. The neutral gas that was present in the ionized medium could simply follow those potential wells of the CDM particles, deep enough to contain the heated photoionized plasma creating the overdensities that produced the absorption features, this gave a common origin for all the features of the forest. This picture was clearly established thanks to the advent of hydrodynamical simulations in which the baryons could be simulated alongside the CDM particles, allowing to produce synthetic Lyα forest that could match that of observations . Therefore being able to produce a representation of the evolution of the gas and the traces that it leaves, the Lyα forest (Cen, 1994). This picture has been improves since then thanks to the measurement the power spectrum of the forest (how probable is to find structures of a given size in the forest) and matching it with the one produced by hydrodinamical simulations (Croft, 1998; McDonald, 2000) observation, clearly representing the cosmological origin of the Lyα Forest.

1.2.2 The Intergalactic Medium as a Cosmological tool

Over the years trough the use of semianalytical models and hydrodynamical simulations a clear view was developed showing that the gas in the IGM is tracer of the gravitational density fluctuations of the matter in the universe. This allowed to develop a framework in which the IGM could be well tested and understood (Croft 1998, McDonald 2000). The IGM is aside from CMB the only environment from which the observable properties can be calculated from a simple set of cosmological initial conditions since the large reservoirs of the baryonic gas can be found in structures that are still within the linear regime. These reservoirs of neutral gas are the ones that leave the leave the Lyα forest as a imprint in the spectra of more distant objects.

That meant that the *Lyα Forest* was a perfect tool to study the structure formation in the universe. Since this structure formation would be sensitive to the cosmological model in which they grow, a measure of the forest would directly put constrains on the cosmology under which the structure was formed. This was firstly studied with the use of few thousands high redshift spectra which allowed to compute the IGM transmission 1D power spectrum (how probable is to find structures of a given size in line of sight from the source to us). This allowed to put strong constrains in cosmology, particularly important in neutrino masses (McDonald, 2006).

Since the Lyα forest traces the large structure of the universe, and as mentioned the BAO is a standard ruler that can be used to extract cosmological information at different redshifts, the BAO can also be measured from the forest giving a unique window to study the universe at that epoch (roughly 10 billion years ago) (Slosar, 2011).

All of this pointed to the importance of doing large surveys that covered big parts of the sky with thousands of objects at high redshift from which to obtain the spectrum and measure the Lyα forest and extract features like the BAO from it (McDonald and Eiseinstein, 2007). With that as one the main aims, the Sloan Digital Sky Survey with the Baryon Oscillation Spectroscopic Survey (SDSSIII-BOSS) program was developed to measure, among others, hundreds of thousands quasars spectra that allowed such measurement from the Lyα forest (Slosar, 2013).
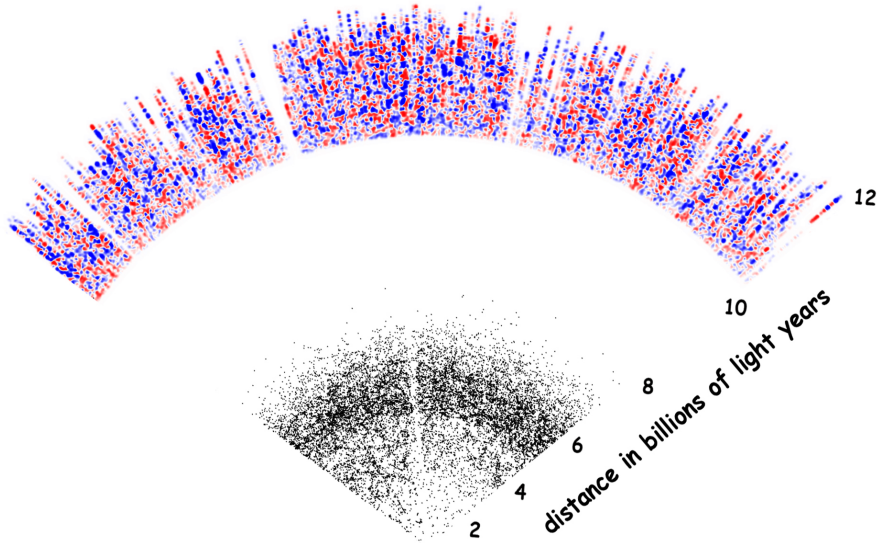
Figure 1.6: Visualization of the mapping of the SDSS-BOSS survey, dots are galaxies and colors represent the density field extracted from the Lyα forest.

However local non linear effects and features of the simulations like convergence of the simulations, non linearity, non uniform photoionization, winds, shocks... might leave important signal on the spectra that can upset some of the cosmological conclusions. For that we need to understand these non linear effects. That can be made thanks to large sets of hydrodynamical simulations. Once they became good enough to make reliable predictions of the physical properties of the physics of the IGM they could be used to make predictions of the effects of these non linear features on the forest. We have just reached that point in which we can start to have growing sets of hydrodynamical simulations to study some of the mentioned effects, as is demonstrated in this Thesis.

Finally it is important to mention that predictions from the IGM and its forest are not only useful for the big picture of the universe at large scales but also offer important answers on other fields, like galaxy formation, growth of structure, metal enrichment of the universe and the helium reionization that came latter than the Hydrogen one.

1.2.3 Understanding the small non linear scales to comprehend the large linear structure

As mentioned earlier to have a clear picture and understanding of the multiple facets of the IGM and be able to extract useful information from it, it is important to have a clear image and modeling of what are the basic physics that drive what we observe from the Lyα forest and the IGM, what structures give rise to the features that we observe.Tthat means understanding the phenomena that happens at structures of few parsecs to some megaparsecs, that way being able to extract information of cosmological scales and about the universe itself. That is the basic motivation to proceed in this thesis with the detailed study of the non linear evolution of the small scales of the IGM.

First as mentioned one has to understand the photoionization of the medium. Thanks to the firsts stars and other astrophysical sources like super novae and Quasars, simple EM interaction, coupled with heating, the originally neutral medium gets reionized into plasma. A fraction of this plasma still has some neutral gas in it, as described in 1.8. Therefore knowing the ionization history would teach us about the state of the IGM, or the other way around, a clear picture of the state of the IGM can teach us the ionization history.

Then what are the major repositories of this neutral gas are systems that produce the absorption signals. These absorbers are high density perturbations of the underling density field, that arises from non-linear gravitational evolution of the structure, however even having a non linear origin they ultimately trace the linear evolution of the matter in the universe with a bias that can be predicted and latter measured.

This linear evolution traces the large scales of the universe. But the correspondence between of the Lyα absorbers is not direct, there is the Bias, how much more significant is one observable with respect another, and Redshifts Distortions, we do not observe in the real space but in the redshift one, where velocities and positions get mixed by the way the light form the objects is redshifted, this distorts the mapping that we do of the universe. This bias and redshift distortions ultimately can be used to relate any observer the linear evolution of the matter linear distribution of the universe.

Finally the linear evolution depends directly on the of cosmology of the universe, therefore specific predictions can be made using a specific cosmological model and its physical parameters, producing an outcome that is what it would be expected if the model was the real representation of the universe. Then again to go all the way from the model to match observations we need to understand all the intermediate steps to reproduce observations. Then this can be matched to real observations to see if the model is able to reproduce them. Also from observations we can build all the way up and try to deduce the model that give rise to them.

Since this is a complex process it is difficult or not always possible or desirable to go from models to mock observations or from observations to specific models, therefore a way to compare them in the intermediate steps can be used. We can build statistics from observations that should be reproduced by the theories, models and simulations. Obtaining these statistical tools and their results allows a common baseline to compare all shorts of models and observations from different origins. Thus having this common ground allows for a more or less direct comparison to obtain our information about the world. Some of these methods that are used in this work are to obtain Lyα forest statistics from simulations, its power spectrum, abundance and metallicities of Damped Lyα systems from large datasets of observations. These allow to present predictions and expectations from simulations an observations that can be latter cheeked on other studies that want to determine main properties of the universe and its evolution like its cosmology, reionization, galaxy formation etc. Thus probing the value of the IGM as a formidable tool to understand our world.

1.2.4 IGM in this thesis

The IGM, by means the Lyα absorption in quasar spectra, is a unique tool allowing the study of the large scales of the universe, which in this way constrains several cosmological, the growth factor of the structure. Moreover the IGM also plays primordial role in galaxy formation. Understanding the physics of the IGM and the way non-linear processes affect the observed Lyα forest, hydrodynamical simulations are necessary because the non linear evolution of the physics cannot be treated in any analytic framework. At the same time, a large number of observations, which have been obtained from the Barionic Oscilation Spectrographic survey (BOSS) to be used in this thesis allows studding in unprecedented large number of Lyα forest from quasar spectra, allowing to achieve major breakthroughs.

We have seen the importance of the IGM in the boarder picture of cosmology, structure and galaxy formation. Furthermore It is clear that due to the nature of the IGM we need a good understanding not only of the IGM at large linear scales but also it is necessary to understand what happens at small and non linear scales, as these leave characteristics imprints in our observations that would affect the way we extract valuable information from these obervations. In this thesis I have worked with that aim, to understand the physics of the small and non linear scales of the IGM. In first place, as proposed by Jordi Mriralda Escudè I had the opportunity to develop a detailed study of the power spectrum in 3 dimensions of the transmission field of the IGM from hydrodynamical simulations. Then thanks to the valuable collaboration with Matteo Viel and also having at my disposition simulations from Renyue Cen I was able to amass a large set of simulations that supposed the basis of the work. From that large sample of simulations with different simulation and physical properties I am able to model the said non linear power spectrum.

In the 2nd chapter of this thesis I describe the methodology that I developed to extract the power spectrum from the transmission flux of the Hydrodynamical simulations. Using a set of simulations provided by various collaborators we study the effects of various simulation characteristics, such as resolution, box size, particles, and cell division, to see how they might affect the extracted power spectrum. Then a model was developed to fit the power spectrum of each of the simulations and in this way see the effects of different simulation and physical properties on the predictions from the power spectrum. Once the model fit has been obtained, the non linear power spectrum of the transmission can be related on the matter power spectrum, which on the form of its biased relation with the matter power spectrum contains the cosmological information. This relation is measured by the bias and redshift distortions that can be directly measured from simulations, making predictions for these values that could be later tested with the observations from current and future surveys like BOSS, eBOSS and DESI.

After the main results of the 3D power spectrum were achieved I was sent to collaborate with George Becker on studding another piece of the IGM, the Damped Lyman Alpha Systems (DLAs), sections of completely absorbed flux in the spectra of some of the high redshift objects, the quasars, observed by the SDSSIII-BOSS survey.

In the 3rd chapter we use the data from the large catalogue of Damped Lyα systems (DLAs) provided by BOSS. These DLAs, broad regions of completely absorbed flux in the spectra ob about 10% of the high redshift objects, the quasars. The DLAs are a large repository of neutral hydrogen, that is why they absorb all the flux that corresponds to their redshift. Since there is so much Neutral Hydrogen the absorption is damped, meaning that

wavelength close to the one of the DLA also get absorbed, forming characteristic Lorenz wings that are sensitive to the column density of Hydrogen of the DLA. DLAs are interesting objects that are expected to be very relevant for galaxy formation, and also because they trace the large structure, therefore also being relevant for cosmology. They are, similarly to IGM of which they form part, a bridge between the large cosmological framework and the galaxies that populate the universe. The BOSS survey has allowed us for the first time have a set of thousands of DLAs. Given this amount of data and the interest of the DLAs, we developed a new method and a new parameterization to study the DLAs in the framework of large surveys such as BOSS. First we constructed a method to measure the equivalent width of several metal absorption lines associated with each DLA, even if individually they are not detected because they can not be observed due to the noise and resolution of the spectrum. The equivalent width are combined in a newly defined parameter called Metal Strength, which is a an optimal combination of the equivalent width of the metals measured for a DLA. This classification can then have many applications to measure the mean stack, mean bias and mean redshift evolution with respect this new parameter. The Metal Strength is therefore used to analyse the evolution and characteristics of the various populations of DLAs.

2. The 3D power spectrum of the Lyα forest from simulations

> "The model made testable predictions, a rare but risky undertaking for astrophysical theories, that eventually led to its demise."
>
> — Michael Rauch.

2.1. INTRODUCTION

The Lyα forest is one of the main observational probes we have for studying the structure and evolution of the intergalactic medium (hereafter, IGM). The fraction of the flux in the spectrum of a source (typically a quasar), $F(\lambda)$, that is transmitted through the hydrogen gas in the region between the Lyα and $Ly\beta$ lines in the rest frame of the source provides us with a one-dimensional map of absorption along the line of sight. The way in which $F(\lambda)$ is related to the density, temperature and peculiar velocity gradient of the gas on small scales is non-linear, and can only be modelled in detail from hydrodynamic cosmological simulations (Cen et al., 1994; Zhang et al., 1995; Hernquist et al., 1996; Miralda-Escudé et al., 1996; Theuns et al., 1998) of the evolution of random realizations of the initial density field. However, in the limit of large scales, the transmission fraction averaged over a large region depends linearly on the mean overdensity and peculiar velocity gradient in the region, and the power spectrum of the transmission is simply proportional to the power spectrum of mass fluctuations, with the standard redshift distortions that were predicted initially for galaxy surveys (Kaiser, 1987; Hamilton, 1992).

This simple linear treatment of the Lyα forest applicable on large scales has led to its use as a cosmological tool to measure the power spectrum, first from single quasar spectra (where the projected one-dimensional power as a function of the parallel Fourier component only is measured; see Croft et al., 1998; McDonald et al., 2000; Croft et al., 2002; McDonald et al., 2006) and then in full redshift space, where the correlation in the transmission among parallel lines of sight is used (Slosar et al., 2011). As proposed in (McDonald & Eisenstein, 2007) and implemented in the Baryon Oscillation Spectroscopic Survey (BOSS; see Dawson et al., 2013a) of the Sloan Digital Sky Survey-III (SDSS-III, see Eisenstein et al., 2011a), the Lyα power spectrum is proving to be a powerful tool to measure the general large-scale mass power spectrum at high redshift, and in particular to measure the baryon acoustic oscillation scale

and the constraints it provides on the expansion rate and the angular diameter distance as a function of redshift (Busca et al., 2013; Slosar et al., 2013; Font-Ribera et al., 2014; Delubac et al., 2014).

In contrast to the large scales, the Lyα power spectrum at small scales is affected by a variety of non-linear physical processes that govern the evolution of the IGM. These physical processes are highly complex, and they may include several phenomena related to the formation of stars and quasars in galaxies that can perturb the IGM: the reionization and the inhomogeneous heating it causes, and the hydrodynamic effects from galactic winds and quasar jets. There is, however, a more simple assumption that can be made for the evolution of the IGM: that the ionization of the IGM is caused only by a nearly uniform radiation background, causing also a nearly uniform heating, and that shock waves arise only from the gravitational collapse of structure, and not from the ejection of any gas from galaxies due to supernovae-driven winds or quasars. Even though it is known that quasar jets and galaxy winds are present in the universe and they have some impact on the IGM, the volume they affect may in practice be very small, and it is useful to test first the most simple assumption for the evolution of the IGM against the observations. This simple model should be mostly described by only five parameters, which determine the statistical properties of the Lyα forest:

- The mean transmission $\bar{F}(z)$, which depends on the intensity of the cosmic ionizing background and is directly measured in the observations.

- The density-temperature relation, usually parameterized with the two parameters T_0 and γ in the power-law relation $T = T_0(1 + \delta)^\gamma$, where $\delta = \rho/\bar{\rho} - 1$ is the overdensity of the gas. When the IGM is heated in photoionization equilibrium and cools adiabatically due to Hubble expansion, one expects this power-law relation to hold with $\gamma \simeq 0.6$, but the relation may be altered by the heating due to HeII reionization (Hui & Gnedin, 1997). We describe this further below.

- The mass power spectrum of primordial perturbations near the characteristic Jeans scale of the IGM, $\lambda_J = 2\pi/k_J$, which we can parameterize also with two parameters as a power-law with free amplitude and index, $P(k) = A_\alpha(k/k_J)^{n_\alpha}$. The Jeans scale is related to the IGM temperature, although in detail it depends also on the entire thermal history (Gnedin & Hui, 1998), and therefore may be considered as a sixth parameter.

Even though the large-scale properties of the Lyα forest are simply understood from linear theory, there is a strong interest in understanding the small-scale, non-linear properties as well. There are several reasons for this: first, we need to test if our understanding of the IGM in terms of a simple uniform photoionization as mentioned above is essentially correct, or if there are important modifications due to a strong impact of galactic winds and jets (Kollmeier et al., 2006) or large inhomogeneities due to HeII reionization (McQuinn et al., 2009; Compostella et al., 2013). Second, the Lyα forest linear power spectrum depends on two bias factors, with values that can be measured and can be predicted from an understanding of the small-scale physics. Finally, the detailed comparison of the observed Lyα power spectrum, determined from a combination of large-scale surveys like BOSS and quasar pairs at small angular separations to better probe the small-scale transverse correlations, with predictions from detailed numerical simulations of the fully non-linear power spectrum, may offer us new clues to essential questions in cosmology, such as the impact of neutrino masses on the growth of structure, or limits on models of warm dark matter or other possible variations on the

nature of the dark matter. There is therefore a need to obtain reliable theoretical predictions for the non-linear power spectrum of the Lyα transmission fraction as a function of redshift from numerical simulations of a large array of cosmological models, in terms of the most important Lyα forest parameters mentioned above.

The goal for the theory of the non-linear Lyα forest is comparable to the that of numerical simulations of the hot, X-ray emitting gas in clusters of galaxies. Detailed determinations of the gas density and temperature distributions from X-ray observations and the Sunyaev-Zeldovich effect, together with the mass distribution from gravitational lensing and the kinematic distribution of galaxies, have spurred advances in the theoretical modelling of clusters, the comparison of numerical codes, and tests of the convergence of the results. At present, the abundance of clusters of galaxies can be used to infer the normalization of the mass power spectrum, but this determination depends on the uncertain relation between the observable properties from X-rays and the Sunyaev-Zeldovich temperature decrement to the cluster mass. This relation needs to be predicted from numerical simulations, and the theoretical modelling affects the comparison with the power spectrum normalization derived from CMB observations (e.g., Hasselfield et al., 2013). Similarly, the Lyα forest is sensitive to the amplitude of the power spectrum and several other cosmological parameters and physical properties of the IGM, but constraints on these quantities can only be inferred once we have a reliable understanding and modelling of non-linear effects on the observed properties of the Lyα forest.

The aim of this paper is to study several cosmological simulations of the Lyα forest, to study the non-linear spectrum that they predict, and to test for the conditions that the simulations must satisfy, in terms of resolution and the simulation volume, to reach a satisfying convergence of the results. This problem was first addressed in the pioneering paper by McDonald (2003), and here we attempt to continue this study by examining a large number of hydrodynamic simulations, characterizing the power spectrum with a new, more simple fitting formula with several non-linear parameters, and studying the dependence of the linear bias factors on the IGM properties. The simulations are described in §2, our technique for measuring and fitting the power spetrum is explained in §3, and the results for the power spectrum fits are presented in §4, with a discussion of the results and conclusion in §5.

2.1.1 The bias factors of the Lyman Alpha Forest

on several of a prediction for the value of the mean transmission in a region where the initial cosmological perturbation has average values of the overdensity $\delta = \rho/\bar{\rho} - 1$ and peculiar velocity gradient $\eta = (1 + z)/H(z) \, dv/dx$ (where v and x are the line-of-sight components of the peculiar velocity and comoving coordinate) is simply given in the linear approximation by $F = \bar{F}(1 + b_\delta \delta + b_\eta \eta)$, where b_δ and b_η are two bias factors for the Lyα forest (McDonald (2003) McDonald 2003). Therefore, when measuring large-scale properties where linear theory can be used, all of the complexities of non-linear collapse and the impact of galactic winds should be absorbed in the values of the two bias factors, while the shape of the Lyα power spectrum in redshift space should reflect the power spectrum of the initial cosmological perturbations with linear redshift distortions (Kaiser (1987) Kaiser 1987, Hamilton (1992) Hamilton 1992). Note, however, that this assumes that the impact of fluctuations in the intensity of the ionizing background and the heating rate due to photoionization can be neglected compared to cosmological density perturbations in the limit of large scales.

2.2. METHOD OF ANALYSIS OF THE SIMULATIONS

Our goal in this paper is to use cosmological hydrodynamic simulations of the intergalactic medium to predict the three-dimensional power spectrum of the Lyman alpha forest transmission fraction in redshift space, $P(k, \mu; z)$, where k and μ are the modulus and the cosine of the angle from the line of sight of the Fourier mode vector, and z is the redshift denoting the cosmological epoch. The simulations we use are described in §2.2.1. The method of analysis is inspired in that of McDonald (2003) and proceeds with the following steps: (1) starting from a grid of cells containing the hydrodynamic quantities of gas density, ionized fraction, temperature and velocity at a certain redshift output, the corresponding spectra of Lyα transmission are computed for the entire grid, using one of the simulation axes as the assumed line of sight direction, and the 3D Fast Fourier Transform of this transmission field is obtained (§2.2.2); (2) the mean value of $P(k, \mu)$ is computed in bins of these two variables, and errorbars are assigned which take into account the variance due to the finite simulation volume (§2.2.3); (3) a parameterized fitting function for $P(k, \mu)$ is chosen to obtain best-fit values of the parameters for several simulations §2.2.4.

2.2.1 Simulation characteristics

Two types of hydrodynamic simulations will be used in this paper. Most of our simulations use the particle-based GADGET code Springel (2005). We also use one simulation that is based on a fixed-grid Eulerian code, described in Cen (2010), to allow for a first comparison of the results for the two types of codes.

Table2.2.1 shows a list of all the simulations that will be used in this paper, including variations in the spatial grid size and spectral pixel size for the analysis of the Lyα forest. The first two columns give the comoving box size, L, and the number of dark matter particles in for the SPH simulations. Note that the simulation labeled Euler does not use particles, and the cells used to run the simulation are the same as those used to compute the cells and spectra. . The third gives the number of cells, N_c^3, in the uniform grid that is constructed to compute the density, temperature and velocity in real space (this is the same number of grid of cells for the Eulerian simulation) . The Lyα spectra are computed for each of the three axes of the simulation playing the role of the line of sight, with the number of pixels in each spectrum from each row of N_c cells of length L given in the fourth column; generally there are used as many pixels in the spectra as cells in the spatial grid, except in the analysis labelled P1024 where the number of pixels is doubled. The last columns give physical parameters of the simulations: the variable σ_8 parameterizing the present amplitude of linear perturbations, the mean temperature at the mean density T_0, and the power-law index that fits the density-temperature relation at low densities, which is described below in more detail. In general, models have variations of different parameters around the values of the fiducial model in the first row of Table2.2.1 , and they are labelled with names that refer to the parameter that is being varied.

Gadget-2

All simulations in Table2.2.1 except for the one designated as Euler were run using the publicly available Tree-Particle Mesh Smoothed Particle Hydrodynamics (SPH) GADGET-2

Name	Box size	Particles	Cells	Pixels	σ_8	γ	$\log(T_0)$
Fiducial	60 Mpc/h	512^3	512^3	512	0.88	1.6	4.3
P1024	60 Mpc/h	512^3	512^3	1024	0.88	1.6	4.3
C256	60 Mpc/h	256^3	256^3	256	0.88	1.6	4.3
R384C256	60 Mpc/h	384^3	256^3	256	0.88	1.6	4.3
R384	60 Mpc/h	384^3	512^3	512	0.88	1.6	4.3
R640	60 Mpc/h	640^3	512^3	512	0.88	1.6	4.3
L80	80 Mpc/h	512^3	512^3	512	0.88	1.6	4.3
L120	120 Mpc/h	768^3	512^3	512	0.88	1.6	4.3
Euler	50 Mpc/h	—	2048^3	2048	0.82	1.9	4.1
Lagrange	50 Mpc/h	512^3	512^3	512	0.82	1.9	4.1
G1.3	60 Mpc/h	512^3	512^3	512	0.88	1.3	4.3
G1.0	60 Mpc/h	512^3	512^3	512	0.88	1.0	4.3
G1T4	60 Mpc/h	512^3	512^3	512	0.88	1.0	4.0
S0.76	60 Mpc/h	512^3	512^3	512	0.76	1.6	4.3

Table 2.1: List of simulations and analysis variations used in this paper. The first column lists the name, the second the simulation box size, and the third indicates the number of particles used in the simulation (both for dark matter and gas), except for the simulation named Euler, which uses a fixed Eulerian grid instead of particles. The fourth column gives the number of cells used to represent the hydrodynamic variables in the spatial grid that is computed to obtain the Lyα forest spectra, and the fifth column is the number of pixels on the line of sight direction used to compute the Lyα spectra.

code Springel (2005).

The fiducial simulation uses a box of 60 comoving h^{-1} Mpc and 2×512^3 particles (for the total of gas and dark matter). Other simulations are run with larger boxes of 80 and 120 h^{-1} Mpc (L80 and L120) to test the effect of the missing large-scale power, or with different resolution to check the convergence as the particle masses are reduced. The cosmological model is flat ΛCDM with the following parameters: $\Omega_{0m} = 0.3$, $\Omega_{0\Lambda} = 0.7$, $\Omega_{0b} = 0.05$, $h = 0.7$, $n_s = 1$ and $\sigma_8 = 0.878$. The initial conditions are generated using the software CAMB [1] and the Zel'dovich approximation at the initial redshift of $z = 49$. The particle mesh grid used to calculate the long range forces is chosen to be equal to the number of gas particles 512^3, while the gravitational softening is $4\,\mathrm{kpc}/h$ in comoving units for the $60\,h^{-1}$ Mpc box and scales proportionally to the initial particle separation for the other simulations The hydrodynamics is followed according to Katz et al. (1996). Star formation is also included in the model with a simplified prescription that allows to convert into a star particle any gas particle of overdensity larger than 1000 and temperature colder than 10^5 K this has been demonstrated to have negligible impact on flux power statistics Viel et al. (2004).

We explore the impact of different thermal histories on the Lyα forest by modifying the Ultra Violet (UV) background photo-heating rate in the simulations, as in Bolton et al. (2008). A power-law temperature-density relation, $T = T_0(1 + \delta)^{\gamma-1}$, arises in the low density IGM $(1 + \delta < 10)$ as a natural consequence of the interplay between photo-heating and adiabatic cooling Hui & Gnedin (1997). We consider two different values for the temperature at mean density, T_0, and three different values for the power-law index of the temperature-density

---
[1] http://camb.info/readme.html

relation, γ, which bracket the most recent observational constraints.

Eulerian

In addition to the smoothed particle hydrodynamics simulations based on lagrangian approach, in this work it has also been analysed a different paradigm for hydrodynamical simulations, a rigid grid one with eulerian approach. This simulation provided by Renyue Cen is based on his series of simulations; (see Cen et al., 1990; Cen, 1992; Cen & Ostriker, 1993; Cen, 2002) . In this work the simulation used features a 50 Mpc/h box consisting on 2048^3 regularly distributed cells. Its cosmology is $\Omega_m = 0.28$, $\Omega_b = 0.04$, $\Omega_r = 0$, $Lambda = 0.72$, $w_0 = -1$, $w_a = 0.0$, $H_0 = 70$, $n_s = 0.96$, $\sigma_8 = 0.82$.

2.2.2 Extracting the Lyα power spectrum from the simulations

For the SPH simulations, we follow the SPH-formalism of Theuns et al. (1998) (their Appendix A4) to compute the hydrodynamic variables of gas density, temperature and velocity on a cartesian grid, and then extract mock Lyα spectra with a period equal to the simulation box size. For the Euler simulation, the cartesian grid on which the simulation is run is used directly to obtain the Lyα spectra, as in Miralda-Escudé et al. (1996). Each cartesian grid in real space from a simulation at a certain redshift results in three simulated boxes of Lyα spectra, taking each of the three axes as the line of sight. For each of the three axes, the spectra for the entire simulated box are computed, resulting in N_c^2 Lyα spectra, where N_c is the number of cells per periodic row of the spatial grid.

Apart from the parameters of each simulation, an additional parameter is necessary to compute the Lyα spectra: the intensity of the ionizing background, which can be altered to adjust the mean transmission $\bar{F}(z)$ to a certain value. We fix the mean transmission to the values given by the expression:

$$\bar{F}(z) = exp(-0.0023(1+z)^{3.65}), \tag{2.1}$$

which adequately fits the observational determinations in Kim et al. (2007). The computed Lyα spectra are modified to adjust this value of the mean transmission by using the approximation that the optical depth varies at each pixel as the inverse of the intensity of the ionizing background, and that the gas temperature is not affected by this background intensity. This assumes that collisional ionization can be neglected and that the atomic fraction is much smaller than unity, which is generally an excellent approximation (except in high density regions where the optical depth is very large in any case, and therefore does not affect the computed Lyα spectra). The use of this approximation avoids having to recompute the Lyα spectra every time that the mean transmission is adjusted to the required value in the expression above.

In the majority of cases, the Lyα spectra have a number of pixels equal to the number of cells in the spatial grid (see Table2.2.1). Only for the case P1024, the number of spectral pixels is doubled to test the impact on the power spectrum calculation. For this case, the transmission is averaged over each two neighbouring pixels after having computed the optical depths and readjusted the mean transmitted fraction.

A 3D Fast Fourier Transform is then applied to the entire box of Lyα spectra, for each of the three cases taking each axis as the line of sight. The routine fftn from the scipy pakage in Python is used for this computation. This results in $N_c^3/2$ independent Fourier modes, each with a modulus and a phase. The moduli are used to obtain the estimate of the Lyα power spectrum $P(k, \mu; z)$.

2.2.3 Fourier space binning and errorbars

The Fourier modes obtained from the three boxes of Lyα spectra from a given simulation analysis (using each of the three axes as the assumed line of sight) are then used to obtain the average of their squared amplitudes at all the possible values of k, μ that arise from the cubic grid. The total number of modes contributing to each value of k, μ is also stored as a weight, $w(k, \mu)$. As an example, it is useful to mention the number of independent modes that are obtained for the few smallest values of k allowed by the size of the box, L. The smallest Fourier modes have a modulus $k_1 \equiv 2\pi/L$. For the direction parallel to the line of sight (with $k_x = k_y = 0$ and $k_z = k_1$, if the z-axis is chosen as the line of sight), with $\mu = 1$, only one independent Fourier mode is obtained (the mode with $k_z = -k_1$ is not independent because of the condition that the Lyα transmission field is a real function). Two independent modes perpendicular to the line of sight are available (for $\mu = 0$): $k_x = k_1$, $k_y = 0$, and $k_x = 0$, $k_y = k_1$. The next smallest modes have $k = \sqrt{2}k_1$, with two independent modes for $\mu = 0$ ($k_x = k_1$, $k_y = k_1$, and $k_x = k_1$, $k_y = -k_1$), and four independent modes for $\mu = 1/\sqrt{2}$ (with either k_x or k_y being equal to +/- k_z). In general, eight independent modes are available for any values of k_x, k_y and k_z when they are all different from zero and $k_x \neq k_y$ (owing to the symmetry under two independent sign changes and under the exchange of k_x for k_y), which are used to estimate the Lyα power for the values $k = (k_x^2 + k_y^2 + k_z^2)^{1/2}$, $\mu = k_z/k$. In this way, estimates of the power at each of the set of discrete values of k, μ obtained from all the modes in the box are collected by averaging the square amplitudes of all the modes, and storing also the number of modes contributing, $w(k, \mu)$.

For large values of k, keeping the estimates of the power at all of the discrete sets of values of k, μ that are obtained from all the positions in a Cartesian cubic grid becomes impractical. At the same time, when obtaining fits to the form of $P(k, \mu)$, the number of values to be fitted needs to be reduced to ensure that the required computational time for performing the χ^2 minimization is reasonable. Therefore choose a threshold value k_t is chosen, above which the exact values of k, μ for each Fourier mode are no longer stored, and fixed bins of k, μ are instead chosen to average the square amplitudes of all the modes within each bin, this scale is chosen to be $k_t = 1[h/Mpc]$ for the $60[h/Mpc]$. The process to select this value is further explained in §A.1.6 Notice that in order to keep the same number discrete modes k_t depends on box size i.e. for the box of 120[Mpc/h] (twice the side length) $k_t = 0.5[h/Mpc]$.

To proceed with the binning also the largest $k = k_l$ mode must be selected, the fft returns modes up to the Nyquist length, which corresponds to the number of flux cells in each axis, this translates to scales much smaller than the ones to be considered on this work. The smallest scale to be used corresponds to the mode 100 for the Reference simulation of 60[Mpc/h], that is $k_l = 10.47[h/Mpc]$ This upper limit is kept teh same for all the different box sizes, unlike the case of k_t. Then the binning is done between k_t and k_l, the values used for the Fiducial simulation are 16×16 bins in k, μ for values in k, this leaves to 256 bins and 296 discrete modes, leaving a total of 552 fitting points for all simulations. The reason for this binning

values is described in A.1.4.

In addition to compute the fit using a MCMC method, each fitting point has a weight or "error" associated, for all data points exists a lower limit error know as ϵ which avoids the low k modes dominating the fit, this ϵ represents the percentage of the total value of the fitting point, $\epsilon = 0.05$ representing a 5% is used through the work. The reason for this values is explained in detail in A.1.3.

The errorbar for each of the power spectrum values is computed as:

$$\sigma(k, \mu) = P(k, mu) \left(1/\sqrt{w(k, \mu)} + \epsilon \right) \ . \tag{2.2}$$

When preparing the vector of the power spectrum $P_{3D}(k_{x,y,z})$, $k_{x,y,z}$ being the components along the 3 axis of the box, one is only interested in certain range of scales. It is desirable to cut the higher k modes, corresponding to lower scales. In this work, that scale corresponds to 10.47[h/Mpc] , that translates to different k mode depending on the simulation box size. The highest mode of each axis is chosen to be $k_{cut} = 2\pi \cdot (k_{lowscale} \cdot size)^{-1}$

After the fft the output has to be converted from the 3 axis realization to a $k_{\|,per}$ space. This straightforward, the $k_\|$ direction is the one the has been computed flowing the Hubble flux (the line of sight) $k_\| = k_x$, while $k_{per} = \sqrt{k_y^2 + k_z^2}$. Different values of $P(k_{x,y,z})$ contribute to the the same $P_{3D}(k_{\|,per})$. Therefore here it is important to keep the weights of each new $P(k_{\|,per})$, that is, the number of old modes that fall in the same new configuration of the parameters space.

The fft output is a 3D vector, the fft output has 2^3 times the elements from the initial 3D flux matrix. As seen in the documentation [2] it returns the modes from 0 to the number of elements in each axis (nyquist number), and then it comes back all the way to 0. The values are symmetric over one axis but not in *each* axis, therefore each combination of modes has $2^{n_{dim}-1}$ different contributions. This means that there are 4 different modes which contribute to the same element of the power spectrum , those four must be accounted for: each power spectrum value has at least 4 different contributions. This is needed to correctly sum all the modes.

Once all the previous is done, computing $P(k_{\|,per})$ is automatic. Just sum the number of elements that have the same pair of $k_{\|,per}$ and divide the number of elements that contribute to that value. To know the weight will be important since our way to compute the errors depends on it, the weight corresponds to the total number of elements that contribute to one final value in the new parameters space. The errors will be explained in A.1.3

Next is to transform to the data to the desired format in which the fitting procedure will be made. In the literature the chosen parametrization for the power spectrum is $P(k, \mu)$ because it has a better representation of the physical observed space. With modulus $k = \sqrt{k_\|^2 + k_{per}^2}$ and the projection along the line of sight $\mu = k_\|/k$.

Since to fit the data from the simulation, which is done using the ξ^2 test, the bigger the number of points, the longer it takes to compute it. The number of elements if a direct transformation is done is more than 10^5, moreover most of them representing large values of k, which correspond to low scales, deep in the nonlinear regime. Those very small scales

---

[2]http://docs.scipy.org/doc/numpy/reference/routines.fft.html

are not considered since resolution limits make them not trustworthy. Therefore the solution is to aggregate the values in a defined number of intervals of k, μ. On the other hand the number of elements that fall in intervals corresponding at low k, large scales, is small. This regions are the most interesting since map the transition from non linear to linear behaviour. Therefore is important to have the specific exact value of each of those modes, not putting them in intervals that only have few elements. Those two different treatments effectively imply dividing the spectrum of the data in 2 regions: for low k were each value is computed individually, and for bigger k all the elements are aggregated in intervals. The scale at which that happens is called transition scale or k_t

The way to chose this transition from one region to the other is related with the total number of data points used for the fitting procedure. It is desired that the the number of intervals in one region to be of the order of the number of low modes of the other. That must be explored until reaching an optimal configuration of enough data points to trace the power spectrum , but not too much making the fitting procedure too slow. That can vary depending on several factors, however it must be adequate for all the simulations analysed, and it is preferable to use always the same procedure to allow an easy comparison. This exploration is described in the appendix A.1.6

Finally the intervals are made both by considering the central point of the interval, or the mean value of all the elements that fall in that k, μ interval. Since even for the smallest bins there are large amounts of modes being considered, both methods return mostly the same outcome.

2.2.4 Parameterized fitting function for the Lyα power spectrum

When all the data is prepared, a theory to compare with is needed. The basic description of the evolution of the power spectrum is that of the linear evolution of it, that is plus corrections for the non linear evolution at small scales :

$$P_{3D}(k,\mu) = b_\delta^2(1 + \beta\mu^2)^2 \ P_{lin}(k)D(k,\mu), \tag{2.3}$$

where b^2 and β are the bias parameters that characterize how the actual values for the power spectrum depart from the initial power spectrum of the perturbations $P(k)$. The linear power evolves with redshift by the growth factor . At the scales that we are analysing the non linearities are parametrized by $D(k,\mu)$. There are several ways to express that, but basically some factors that trace how non linearities affect the growth of the power spectrum must be introduced. Tree non linear effects can be accounted for: The Power spectrum is enlarged at low scales by the effect of gravitational bounding and falling of the hallos one to another. Another is the suppression by the effect of pressure that keeps structure from collapsing. Finally the increasing of the velocity dispersion which smooths the inhomogeneity.

Therefore the non linearities are accounted by those factors:

$D(k,\mu) = e^{(nl-p-v)}$,

where nl, p, v are the functions that trace for the variation of power due to the non linear physical effects. Those can have different forms, and in this work several options have been checked. One will be chosen as the standard, from there all the different simulations will be fitted using the same expression for $D(k,\mu)$.

Different options were tried in this work for nl, p, v, here are present the most tested ones:

$$nl_0 = \left(\frac{k}{k_{nl}}\right)^{a_{nl}},$$

this form only tries to account for effects on the scales lower than k_{nl} and a_{nl} controls the shape with K.

$$nl_1 = k^3 P_{lin}(k) \cdot k_{nl},$$

this is chosen to always converge to linearity at large scales, as n can still be large in the low end of k shifting the effect of non linearity to larger scales than expected. The factor $k^3 P_{lin}(k)$ is adimensional.

$$p_0 = \left(\frac{k}{k_p}\right)^{a_p} \text{ tries to model the effect of the pressure at small scales, it stops the growth of}$$
structures when the pressure dominates.

$$p_1 = \left(\frac{k}{k_p}\right)^2,$$

here a_p is fixed to 2 because it is the expected evolution with the pressure and agrees with the values found when it is let be free on the fitting.

$$v_0 = \left(\frac{k \cdot \mu}{k_{v0}(1+k/k_{v1})^{a_{v1}}}\right)^{a_{v0}},$$

this form comes form the McDonald 2003 McDonald (2003) where it tries to model the peculiar velocity dispersion dependence both in scale k and angle from the line of sight μ.

$$v_1 = k^3 P_{lin}(k) \cdot \left(\frac{k}{k_v}\right)^{a_v} \mu^{b_v},$$

this is the most simple dependence of the power spectrum, to model the effect of peculiar velocities and thermal broadening that has been found. It uses one less parameter than the v_0 form.

Therefore with these factors 2 main different formulas have been used to model the non linear evolution of the power spectrum . The combination of nl_1, p_1 and v_1 (D_1) and nl_0, p_0 and v_0 (D_0), with D_0 having 10 parameters, 3 more than D_1. D_0 is the same as used in McD2003 and is useful for tests and comparison, but the default one will be D_1, as it behaves better and has less free parameters, as described in the appendix. A.1.1

That leaves the final expressions as:

$$D_0(k,\mu) = exp\left(P_{lin}(k)\left(\frac{k}{k_{nl}}\right)^{a_{nl}} - \left(\frac{k}{k_p}\right)^{a_p} - \left(\frac{k \cdot \mu}{k_{v0}(1+k/k_{v1})^{a_{v1}}}\right)^{a_{v0}}\right). \qquad (2.4)$$

$$D_1(k,\mu) = exp\left(k^3 P_{lin}(k)\left(k_{nl} - \left(\frac{k}{k_v}\right)^{a_v}\mu^{b_v}\right) - \left(\frac{k}{k_p}\right)^2\right). \qquad (2.5)$$
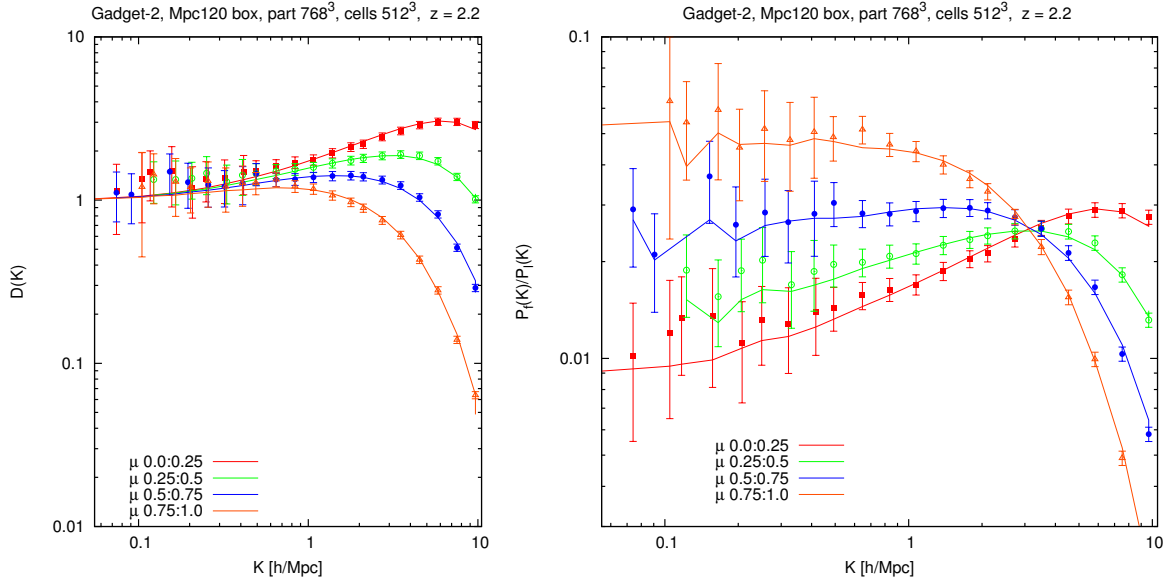
Figure 2.1:   power spectrum for the reference model with $D_1(k,\mu)$ eq. at redshift 2.2 and simulation L120, in function of k and in 4 bins of μ. Left are the therms due to the non linear behaviour, as can be seen those tend to 1 for large scales, one of the motivations to explore different $D(k,\mu)$ Right is the power spectrum from the simulations normalized by the linear power spectrum form the simulation. The lines represent the fit with the $D_1(k,\mu)$ eq. while the points represent the data from the simulation and the error bars represent the number of modes that fall in each interval of k,μ, with an added 5% error to all values to avoid large modes to have very small associated errors.

2.2.5 Results: Fiducial model

Once the equations to proceed with the fit are decided, next is to select the arrangement of the data to be fitted, and pass different tests to check which eq and configuration is better to proceed systematically with for the rest of the work. Since the number of modes for the 3D power spectrum grows as n^3 where n is the number of modes in one dimension, once one goes to sufficiently large n, which in its turn corresponds to small scales, the number of data points corresponding to those modes grows to an intractable amount. The points at high modes, small scales, must be aggregated in bins of k,μ as described previously, On the other hand, there are few modes to account for the large scales, those of the same size of the box and fractions of it, that is better to resolve them exactly. Therefore for some large scales (small k) the data points are directly computed from the modes, while for large k the large amount of data points is binned in intervals.

How to select how many bins are used and where to make the transition from small to large scales, and the effect that varying those things has is discussed through the appendix. 2.5 Another thing to be considered for the fitting is which error to associate to each data point, to proceed to do the fit. For the fitting a χ^2 test is used which uses the error for each data point to proceed with the minimization. It has been decided to assign an error depending on the amount of modes that contribute to a determinate k,μ values and bins. The problem with that is that the amount of modes that correspond to small (and therefore highly non

linear) k is huge compared with the linear counterpart.

Therefore similar as in McDonald (2003), to avoid the non linear terms dominating the fit, an additional error is added to all data points. How the error is computed and the effects that it has over the fitting is discussed in §A.1.3 . Next the behaviour of the two equations should be tested. It has been decided to use a different expression that the one in McDonald (2003) because it failed to pass some tests, basically it had a stronger than desired effect on the linear scales, therefore affecting greatly to the values of the bias parameters. That can be seen by how the value of $D(k, \mu)$ tends to 1 for the fitted parameters at large scales. Another test is considering the fit up to smaller or grater scales, if the values are similar that means that the influence of the small scales over the linear ones is not dominant. Those tests are described in A 2.5

Since $D_1(k, \mu)$ provides a good fit even with less parameters, behaves better at large scales, and passes more tests than other expressions do, it will be used as default fitting formula. Therefore:

$$P_{3D}(k, \mu) = b_\delta^2 (1 + \beta\mu^2)^2 \ P_{lin}(k) \ D_1(k, \mu). \tag{2.6}$$

2.2.6 Growth factor and linear power spectrum

Other detail to consider for the fit is the linear power spectrum that is used. It comes from the outcome of the CAMB [3] code using the cosmology of each of the simulations, as mentioned in §2.2.1 and it plays an important role on the fit, as it is present both in the linear an non linear terms. The growth factor is also to be considered as is the only place in which the cosmology affects the fitting. The linear power spectrum is computed for redshift 0, therefore the linear growth of perturbations must be scaled back to the redshift at which the fit is done. This growth of perturbations depends on the cosmology used, therefore it has an influence, although minor, on the fitting. Since it is minor it is not sensible to small changes in the cosmological parameters therefore the power spectrum can not be used to constrain the cosmology in this way.

2.2.7 Fitting procedure: MCMC

With a theoretical model and the data prepared, only to fit the one to another remains. That also can be done in several ways, the one chosen here, as pointed earlier, is the minimization of the χ^2 function:

$$\chi^2(x_{1,...,p}) = \sum_i^N \frac{[d_i - f_i(x_{1,...,p})]^2}{\sigma_i^2}, \tag{2.7}$$

d_i the data points, σ_i the error given to each data point, $f_i(x_{1,...,p})$ the function to be fitted to the data depending on $x_{1,...,p}$ being the parameters that will vary in order to minimize χ^2, $\frac{\partial \chi^2}{\partial x_{1,...,p=0}}$.

As explained in A.1.3 the values for $c\sigma_i$ are artificially increased for high k modes. Since the error affects to the value of the best fit for the χ^2 . This implies that now χ^2/N is no

---

[3]http://camb.inf

longer a reliable test of how good the fit is N being the number of data points minus the number of degrees of freedom, number of fitting parameters. For this case case χ^2/N is going to be quite small since $\epsilon = 0.05$, meaning that the error is over estimated, as it is the case. Through this work a simply minimization of the χ^2 will be used for the fitting.

For the chosen function with n DoF, a good method to explore all the possible values for the parameters and find the best combination is necessary. Also a method that returns the errors associated with each of the parameters, since it is very important to know to which degree the the best fitted values can be trusted. If they are very degenerate or if they give some valuable information should be known.

In this work a simple routine based on the Monte-Carlo Machovff Chain (MCMC) using the Metropolis algorithm has been developed to deal with the minimization process, finding the errors of each parameter, and how the degenerations and correlations between them. See (Bhanot, 1988) for a review on the Metropolis algorithm.

To improve the performance of the method, is desired to start the chain from values that are not far form the ones expected. It will take longer if they are not. Also it is recommendable to start the chain and the fitting from different points to check if it finds always the same minimum, if that were not the case that means it is falling each time in different local minimums, or has not completely explored all the minimum for very degenerated parameters. In those cases is simply run longer and bigger number of chains.
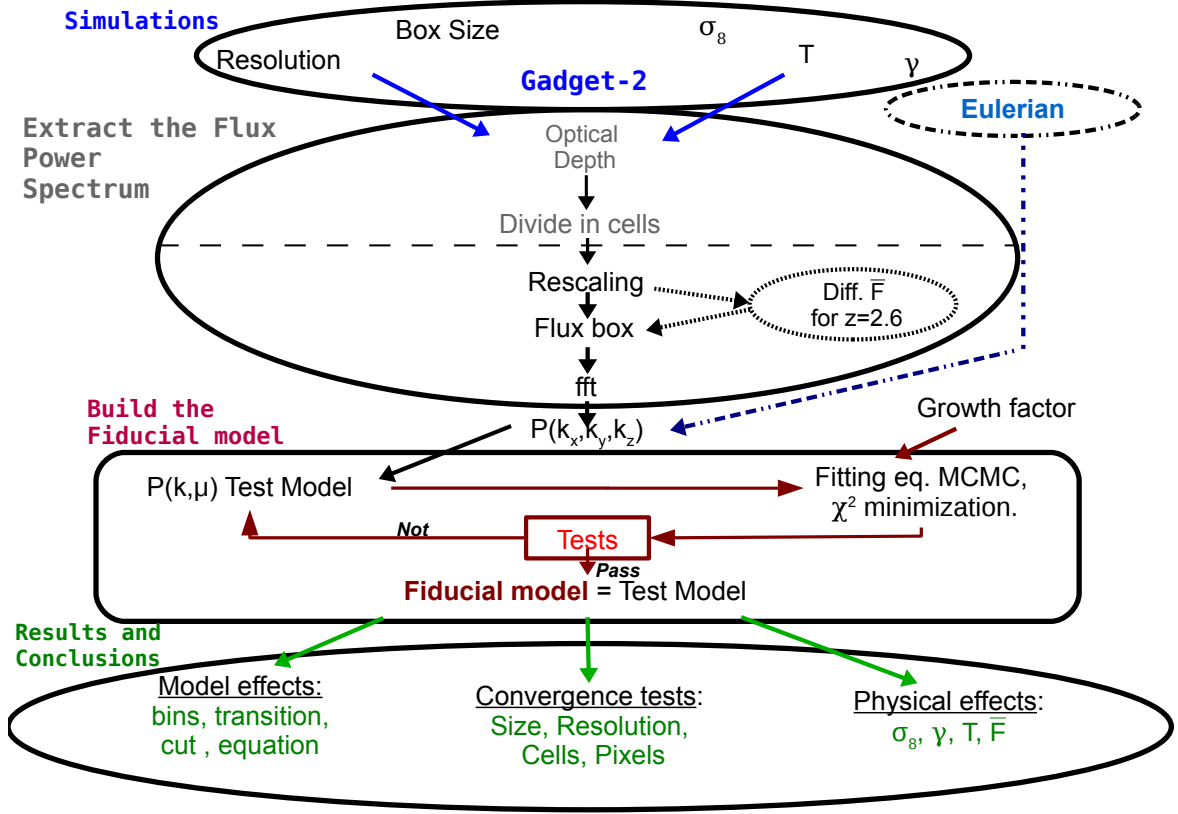
Figure 2.2: In this diagram is shown a comprehensive summary of the flow of the work. First we have a set of simulations, Gadget-2 with different resolutions an box sizes, and various modifications of the physical properties: amplitude of the perturbations σ_8, mean temperature T and temperature density relation γ. Then the outcome of this simulations for various redshifts is processed, by dividing by cells, extracting the neutral fraction and computing the optical depth. Then this has to be renormalized at a certain mean flux, and from that the fast fourier transform (fft) provides the power spectrum in the 3 axes $P(k_x, k_y, k_z)$. The eulerian simulation provides directly the power spectrum. The power spectrum is then transformed to k, μ coordinates. This is then used to create the standard method for all the simulations, a series of tests are used to select the equation to describe the non linear part of the spectrum, which will make use of the linear matter power spectrum. Through this method are selected the binning, scales (cut, transition) that are considered, and minimum error (ϵ). This determines the fiducial model to be used. From this are extracted the predictions, convergence tests and effects of the method used. This are the outcomes of the set of initial simulations and the methodology created to extract and fit the power spectrum for reach of the simulations.

2.3. CONVERGENCE TESTS

To be able to obtain strong results from this study, it is necessary to check that all the results hold. All the data analysed comes from cosmological simulations, therefore it is important to check if the results are or not critically dependent on the simulation used, and its characteristics. Also it has been checked in §2.5 the effect that the data reduction for the fitting on the values of the parameters.

the next step to proceed is to do the same analysis with simulations, modifying each of the characteristics that can have an effect on the results. For clarity we will divide those characteristics in two parts:

First one would be *Prostprocesing*, which is all the modifications that can be done to the power spectrum and the fitting procedure; the binning of the data intervals, the transition point in k space in which the power spectrum data points pass from being computed exactly to be computed in bins in k and μ, the added error to avoid the small scales dominating the fit, the effect of doing the fitting up to larger or smaller scales, renormalization of the flux to have the same mean flux in all boxes and finally how the results depend on the fitting formula used. Since those are technical proceedings and have been tested for the production of the fiducial model, they are discussed in the appendix 2.5.

Second one would be *Simulation characteristics*, which correspond to the different data that I use to compute the power spectrum of the flux. That is: different dimension of the simulation (box size), numerical resolution of the simulation, number of flux cells in the grid, number of pixels on the line of sight and finally using totally different simulation.

2.3.1 Box Size

As discussed in §2.2.1 the scales in which the power spectrum is characterised are of the order 0.1 to 100Mpc, therefore a simulation must be able to explore inside this range of scales with enough resolution. Simulations within this given box size will be used. However must be also checked that the results obtained are not simply due to the physical size of the simulation, i.e. physical process that occur only at those scales and might affect the power spectrum . Or simply check if the simulated physics scales well enough for bigger sections of the universe, and if not, how that could affect the results.

In order to accomplish that it is only needed to run with bigger boxes given the same initial conditions. For this propose the Gadget simulation has been run in 80Mpc/h with 512^3 particles and a 120Mpc/h with 768^3 particles, an "equivalent" resolution (i.e number of particles per volume) as the Reference simulation with 384^3 particles. Also to have an equivalent physical resolution the L120 with 512^3 cells vs. R384 with 256^3.

Then simply compare the results of the fittings for each redshift output for those simulations. In fig.2.3 is compared the power spectrum for L60 and L120 with equivalent spacial resolution (768^3 particles and 512^3 cells for 120Mpc/h; 256^3 cells and 384^3 for 60Mpc/h at z=2.2). It can be seen how the fit for the small scales is very similar, while for large scales the fit shows increasing difference for decreasing μ. Although this difference is minor and decreases when more simulations are stuck together for the small box, decreasing cosmic variance, the difference still remains. Therefore it would be important to understand the cause

of it, if is a simulation glitch or data processing. The fact that it depends on μ and the differences are more important along the LoS probably points to an effect of the large scales peculiar velocity playing a different role in each simulation. Having less power on L60 along the LoS hints that for it the peculiar velocities are more important, while for L120 maybe larger structures probably present in L120 but not on L60 on large scales limit the influence of the velocity smaller scales.

Visually the difference is more noticeable for large scales where the power is bigger for L120, but that does not translate to large fitting bias values fig.2.4 were it can be seen that the values are similar for the different boxes. But surprisingly the gradient of the values of β is reversed, something that one would not expect even if by small differences exist. L80 is in one extreme and L120 or L60 in the middle while one would expect the opposite. Notice that L80 has different physical resolution than L60 and L120, which have the same. Also the fiducial model has been chosen for it to give the same β as L120 at redshift 2.2, as explained in the appendix fig.A.1.4 . Such setting has not been done for the rest of the redshift outputs or other simulations. Yet the agreement, even if not total is quite good. And curiously for b_{delta}, and the non linear parameters seen in fig.2.5 now L80 does behave as expected for most of the parameters and redshifts the values falling in between L60 and L120. This behaviour might have to do with the method used to select the fiducial model, or the different physical resolutions used when comparing L80 with the others, nevertheless the effect minor but shows how complex this study is.
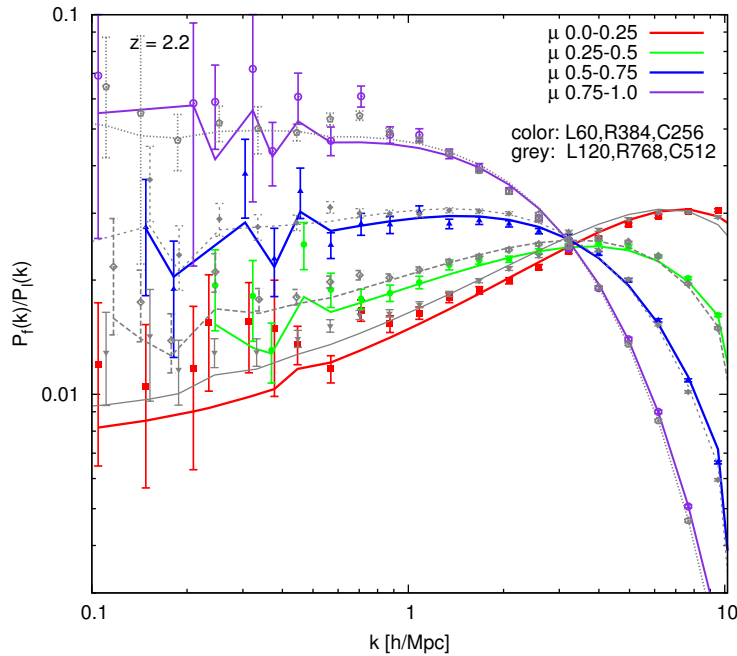

Figure 2.3:   Comparing the 120Mpc/h box with the 60Mpc/h with equivalent resolution, 768^3 vs. 384^3 particles and equivalent physical cell resolution 512 vs. 256. At z = 2.2 The agreement is evident for small scales but not much for large ones, yet the bias from the fit result in similar values.
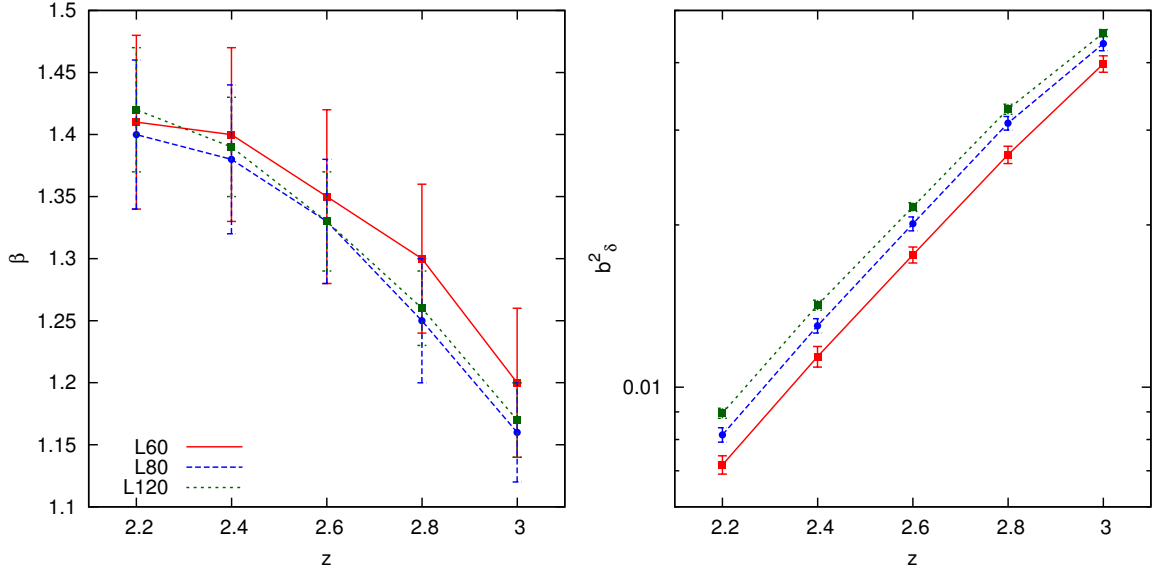
Figure 2.4: Comparing the bias parameters for different box sizes. Notice that while L60 and L120 have equivalent physical resolution, i.e. L60 has 384^3 particles and 256^3 cells, L120 has 768^3 par and 512^3 cells, L80 has different physical resolution since it has 512^3 particles and cells.

2.3.2 Resolution

To test if the results are sensitive to the resolution of the simulation, the same setup is used with different number of particles for the G-2 simulation: 382^3, 512^3 & 680^3. The same fiducial model has been used for all of them. Is expected that the major discrepancies will be important for small scales where the resolution plays a major role. This can be seen in fig.2.6 where the discrepancies are stronger for small scales, and mostly for angles perpendicular to the LoS (high k low μ). This agrees with what would be expected when having more resolution and its effect on the small scales. Curiously, similar to the box case, can be seen a slight but systematic increase on power at all scales, also greater for smaller μ values.

Nevertheless the difference is rather small if one only looks to the data points. This is reflected directly in the values for b_δ^2 remain almost the same, as seen in fig.2.7, here can be easily observed that the resolution has an effect on the parameter β and dependence in redshift, being the effects more clear at high z. Lower resolution decreasing systematically the values β wile higher resolution increases it. R384 and R640 agree surprisingly well for both bias parameters at low redshift, but R640 has bigger β for high redshift even though converges a low redshift. That is probably due to the ability of the simulation to resolve non linear scales depending on the resolution, not being enough particles at the non linear scales and high redshift, and therefore not being resolved the same way for higher and lower resolutions at those redshifts.

Interestingly when comparing the fits for the non linear parameters fig.2.8 there are no significant differences, nor strong discrepancies depending on redshift. Overall the parameters for R640 are more different but in a constant way. Except k_{nl} and k_p in which R640 has different trend. For k_p, if the interpretation given to it as to account for the scales in which
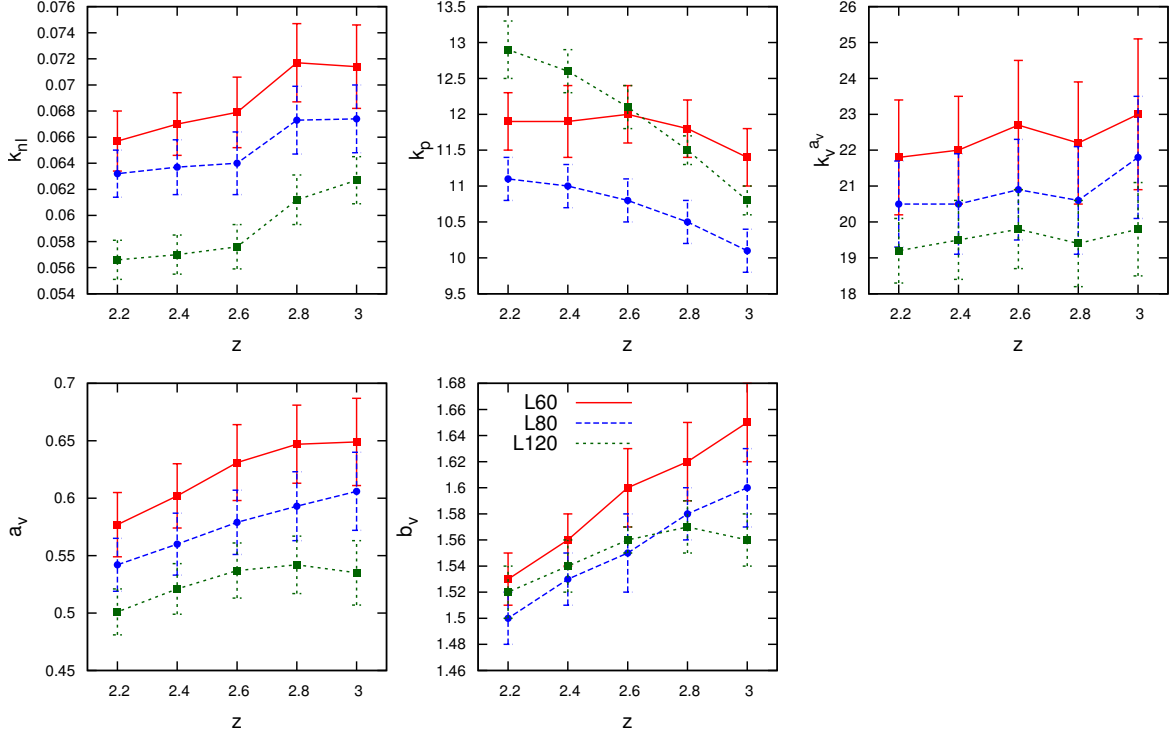
Figure 2.5: Comparing the nonlinear parameters for different box sizes. Notice that while L60 and L120 have equivalent physical resolution, i.e. L60 has 384^3 particles and 256^3 cells, L120 has 768^3 par and 512^3 cells, L80 has different physical resolution since it has 512^3 particles and cells.

the pressure plays a role, then it would reflect that this pressure is solved differently for the higher resolution and higher redshift, as it better analyses the non linear behaviour for those redshifts. But again that is only an interpretation, other tests and analysis should be developed to check those interpretations. Nevertheless the effects are small and are only over non linear parameters with strong degeneracies between themselves, therefore is difficult to isolate those effects.

Overall higher resolution reduces the trend on the redshift evolution of β, being now compatible with no evolution with redshift if the errors are considered. On the other hand this non evolution trend is unlikely for the lower resolution. Finally what that shows is that even effects at small scales like the resolution, have a noticeable effect on computing the bias parameters when this method is used. In this case making what was a clear evolution on β not so clear for higher once higher resolution is used.
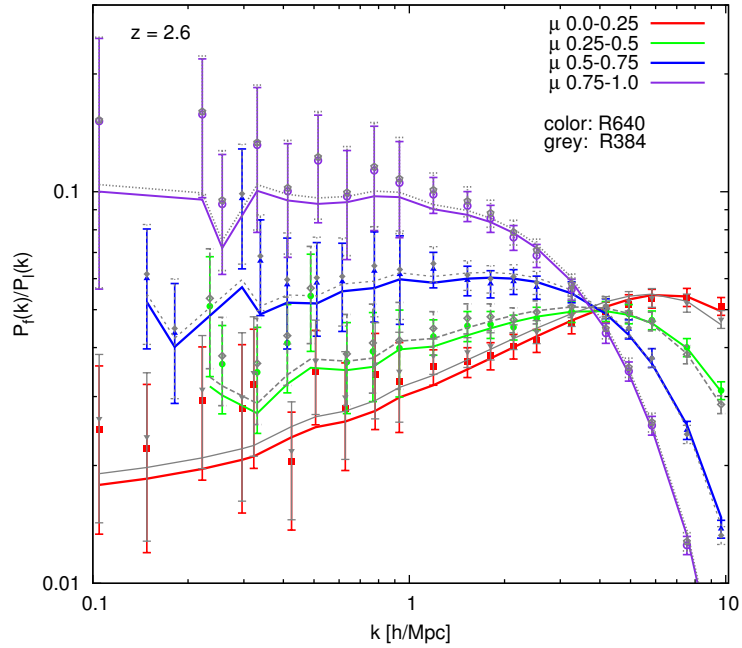
Figure 2.6: Comparing the power spectrum for the lower and higher resolution simulations (colors 640^3 vs grey 384^3 particles) for the 60Mpc/h box at redshift 2.6.
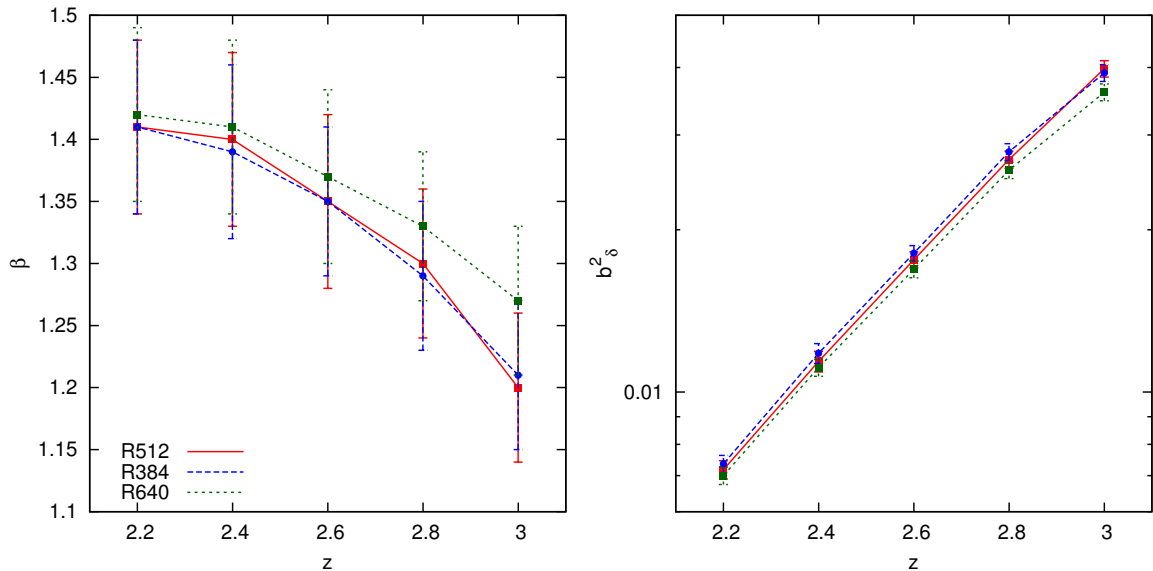

Figure 2.7: Different resolution for the 60Mpc/h box for the same initial conditions. Here the highest and lowest resolution simulations with 640^3 and 384^3 particles are compared with the reference 512^3 particles for 512^3 cells.
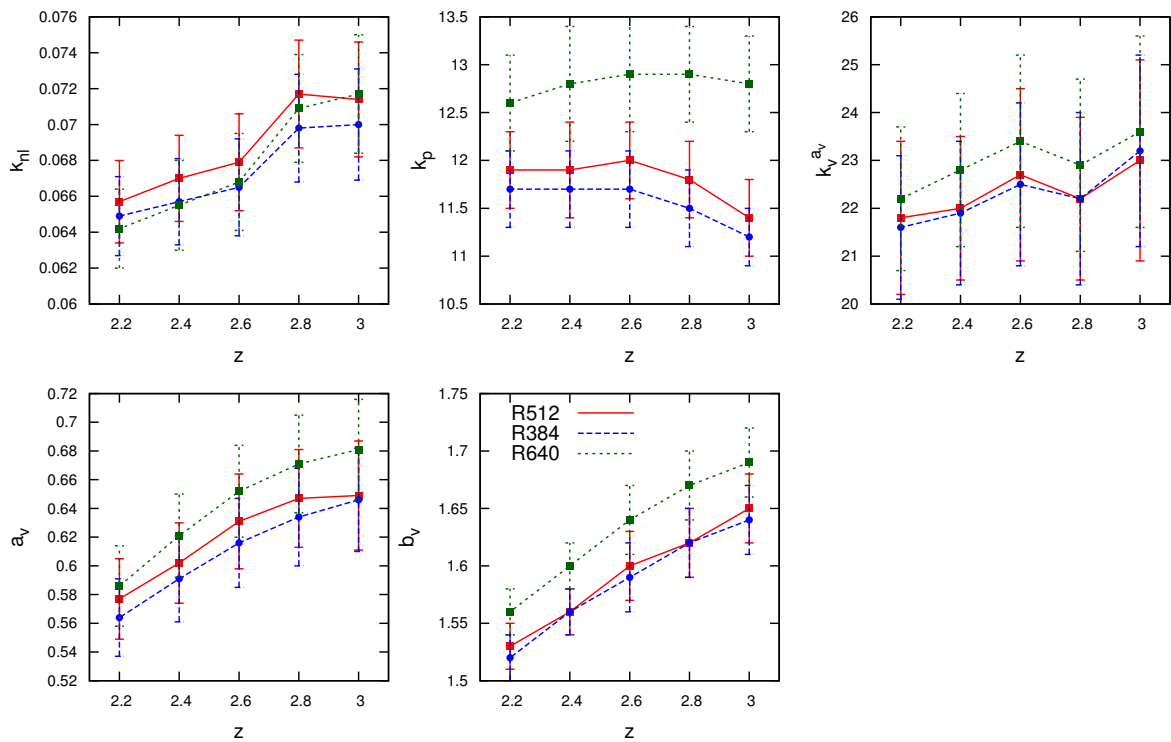
Figure 2.8: The non linear parameters for different resolutions, they do not show any significant trend or divergence other than the k_p one, which probably links with the differences in non linear scales for different resolutions.

2.3.3 Number of flux cells

Another effect related to the resolution in number of particles is the cell resolution for the flux box. Converting particles of the simulation to a box with regular grid. This can be done for different number of of total cells, the method to obtain those cells is explained in appendix A4 of Viel et al. (2004). From those cells that contain peculiar velocity, hydrogen density, and temperature, is easy to extract the flux and compute the power spectrum , the amplitude for each mode of the box. The more cells there are, the better the large modes are solved, however this implies that more memory is used to handle the three dimensional Fourier transform, and since is a tree dimensional box, increasing the cells increases the memory consumption as n^3_{cells}. Therefore selecting the minimum amount of cells to resolve all scales is needed. To obtain scales of 0.6Mpc/h for L60 means 100 cells per side, and for L120 should be 200. However when L120 is divided in 256 cells the power spectrum is not well solved for small scales, has to be considered that in other to produce those cells the neighbour particles are play a role,like in all the SPH systems, therefore cells too big can disturb that, and when computing the power spectrum more physical resolution is needed. All SPH simulations in this work will be divided in 512^3 cells as standard, to keep the same cells for comparison, except L60,R364,C256 in which is the physical resolution is the same as L120, described in §2.3.1. The problem is that modifying the total number of cells has an effect on the power spectrum , even for L60 the modes mainly at small scales, but also at large scales, is affected by selecting a box with different number of cells. This is probably due to the way the cells are build, but in any case this effect must be also considered and explored. In fig.2.9 is shown the effect of modifying the number of cells for the highest resolution simulation: R640, where the effect is stronger since there are more particles and the neighbours affected differ more depending on the number of cells. It can be seen how for small and intermediate scales the discrepancies are noticeable, and moreover for large scales in the C256 simulation the power of the data points is slightly lower for increasing μ. As there is less spacial resolution probably lower modes that would be taken into account by the neighbouring system, now are not resolved and contribute to increase the power at higher scales.

From fig.2.11 the different tendency for β and b_δ at z=3, the only major difference for this redshift, as seen in fig.2.10 is that the non linear terms perpendicular to the LoS are have a bigger difference, and probably that forces the fit to give smaller β. From the fitting eq is obvious that when $\mu \to 0$ the non linear effects have more importance over the bias parameters since the pareameters depending on μ have little or non effect now. This then motives once more to compute the power spectrum for the IGM, or at least the power spectrum perpendicular to the LoS since it constrains more strongly the bias, which describe the growth of the perturbations and the departure from the linear evolution. The fact that this discrepancy is stronger at z=3 is probably also due to having smaller structures which involve less particles to and a coarse cell can not resolve them good enough.

Finally the non linear parameters, one hand they differ significantly as seen in fig.2.12, probably reflecting the differences in the non linear scales, but on the other they do not show a great difference in trends for z=3 as is the case with β, except for the case of b_v and maybe k_p, showing this change in trends for z=3.
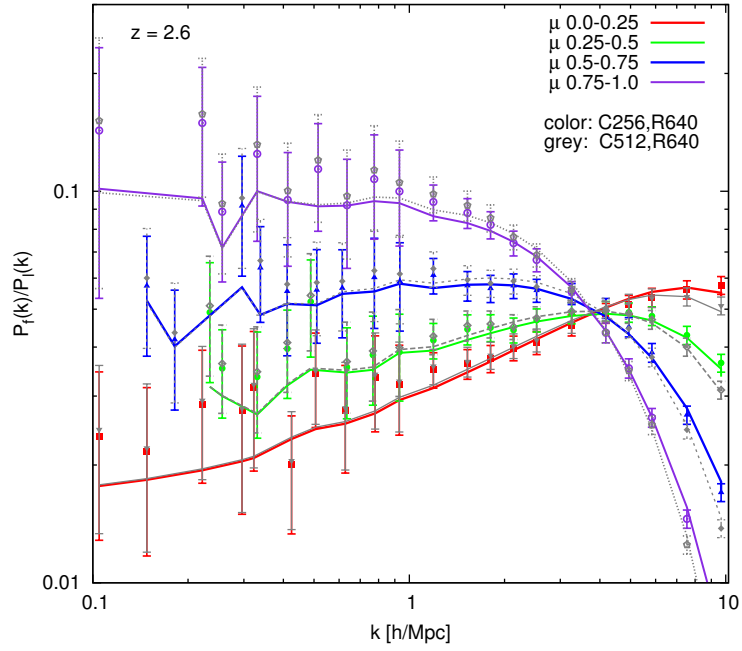
Figure 2.9: For the simulation with 640^3 particles, comparing for 512^3 (grey) and 256^3 cells (colors) at redshift 2.6. The number of cells chosen to compute the power spectrum form the flux has an appreciable influence, it increases the power for the small scales of the simulation.
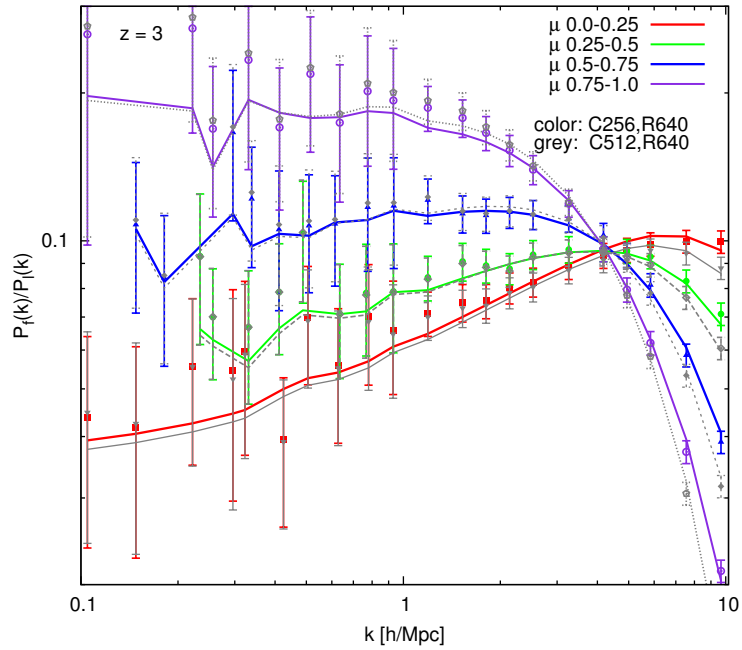

Figure 2.10: For the simulation with 640^3 particles, comparing for 512^3 (grey) and 256^3 cells (colors) at redshift 3. The the discrepancies for small scales are greater than at lower redshifts.
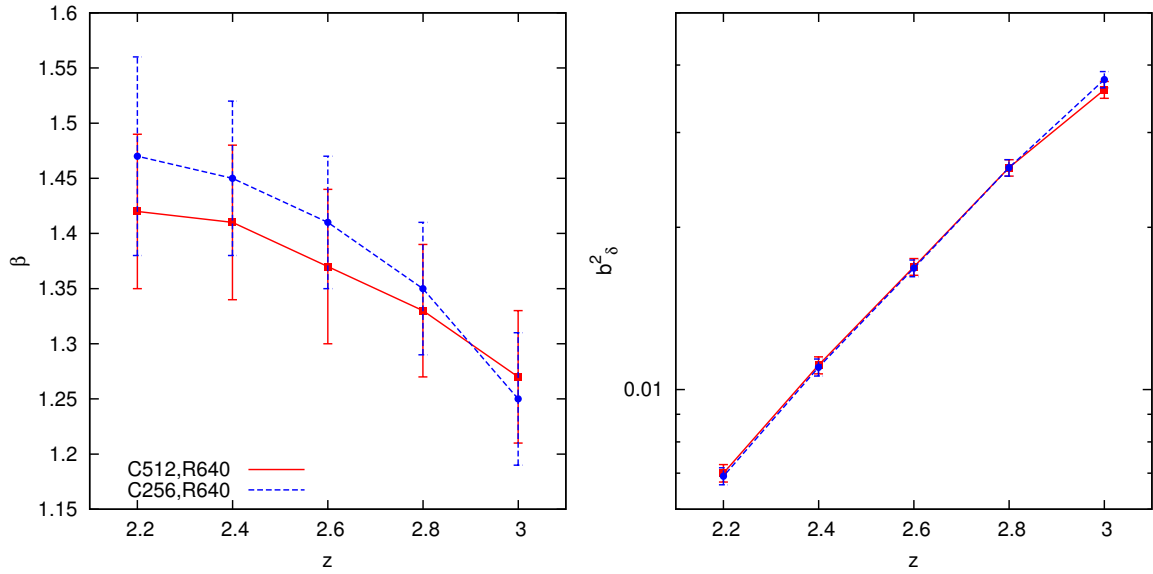
Figure 2.11: Effect of the flux number of cells on the power spectrum for the higher resolution 640^3 where the effect is stronger. Higher β values for C256 are mainly driven form small scales effects, as seen in fig.2.9, except for z=3 where the trend is reversed.
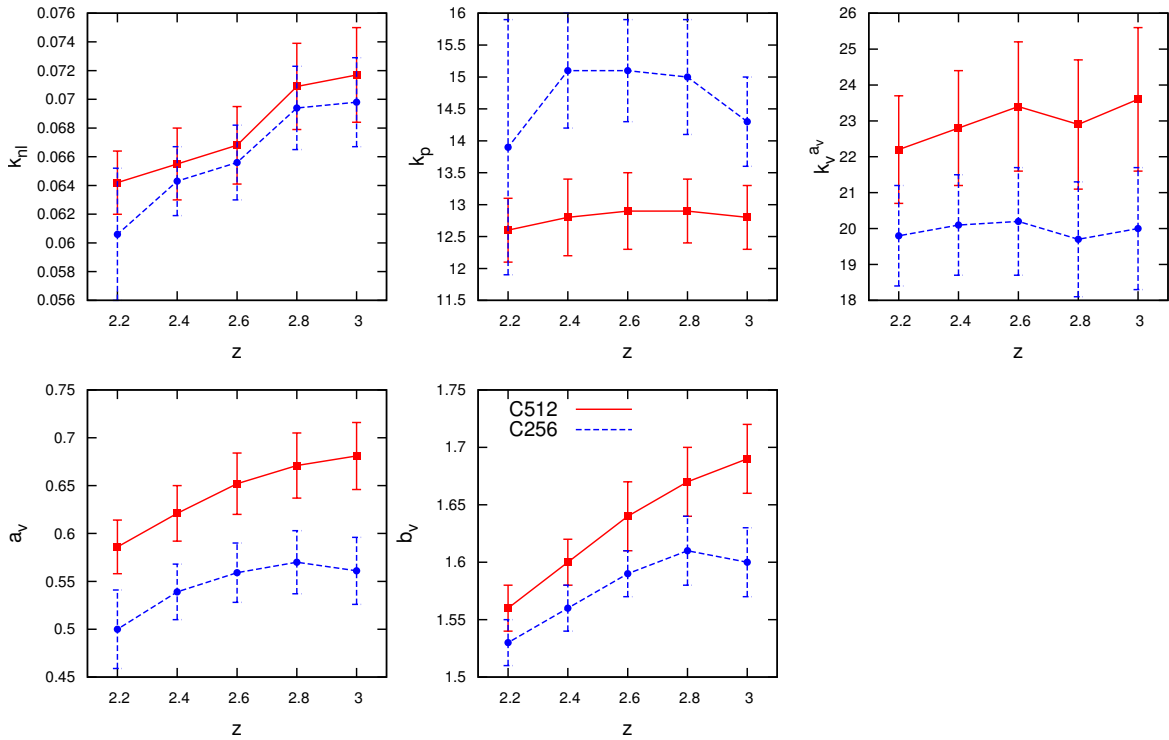

Figure 2.12: 256 versus 512 cells for R640 for the non linear parameters. The values are quite different, but the only breaking of the trend for z=3 is seen b_v, and to a minor degree in a_v and k_p.

2.3.4 Number of pixels in the line of sight direction

Finally another technical detail to check is the effect of increasing the number of pixels along the line of sight, that way having more resolution along this axis. Then the position space is stretched when converting to to the velocity space due to the Hubble flow, therefore more pixels along this direction could in principle modify the power spectrum . Increasing the pixels is done by dividing the optical depth (τ) along the line of sight in 1024, while in the perpendicular direction only 512. Latter when the flux is computed, 2 pixels of the optical depth from the LoS are combed to form a flux cell, having finally a cubic 512^3 cells box for the flux, but with the flux computed from more τ pixels along the LoS.

The curious thing is that increasing the pixels along the LoS has apparently stronger effect on the for the modes prependicular to the LoS, as seen in fig.2.13. The fact that those are affected is not surprising though, that bin contains some modes along the LoS, and moreover when building the cells the neighbouring particles in all directions are considered, as explained in §.2.3.3, therefore increasing the number of pixels in one direction affects all of them, and finally since the renormalization is affected by the total power, a decreasing or increasing along one direction is going to be redistributed for all, as the flux is modified up or down to have the same mean. The fact that the perpendicular seems to be the most affected is simply due to the fit procedure, also as explained in §.2.3.3 the the bias and the fitting is more sensitive when $\mu \to 0$, therefore small changes on this will affect mainly β and have an effect on the fit at all scales, which is the case seen in fig.2.13.

This modification, is seen to have similar but much weaker effect than increasing the resolution. Affecting the values of β for high redshifts, seen in fig.2.14, the similarities with the R640 are clear, but the increase in β is lower than for R640, pointing to some other effect which is present in one of the simulations but not the other.

For z=3 the slight decreasing in power reflected by smaller b_δ as well as β points again to some effect of the resolution with redshift, being the number of cells or pixels ore sensitive to smaller number of particles to describe non linear perturbations that have not grow much at high redsifts, therefore harbouring smaller number of particles, affecting that to the optical depth pixels build. Overall the effect is quite small, also in the fitting on the non linear parameters, not shown here, but again it shows the sensitivity of the fitting procedure to small differences on the non linear scales, especially if they are present perpendicular to the LoS.
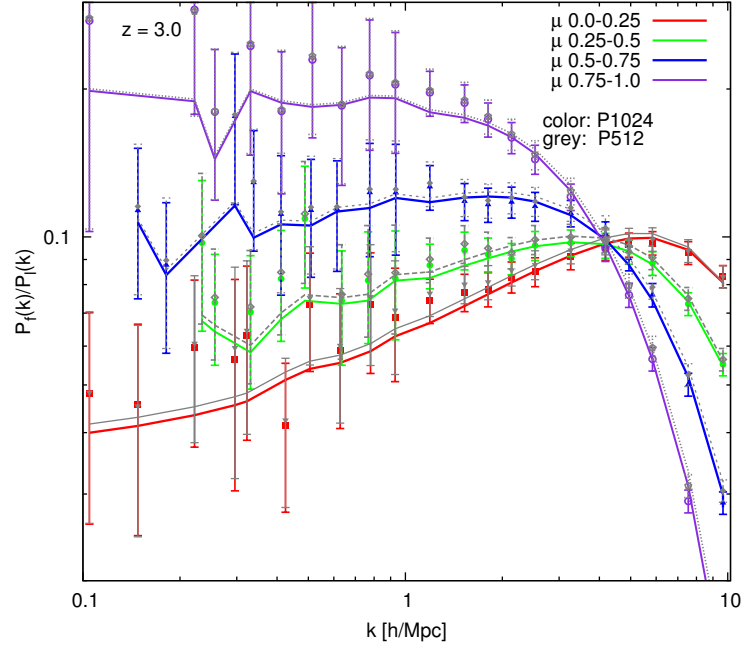
Figure 2.13: Comparing double number of pixels in the line of sight direction 512 to 1024 pixels at z = 3. The effect is small but affects the μ bin perpendicular to the LoS, which explains why β is affected greater for this redshift.
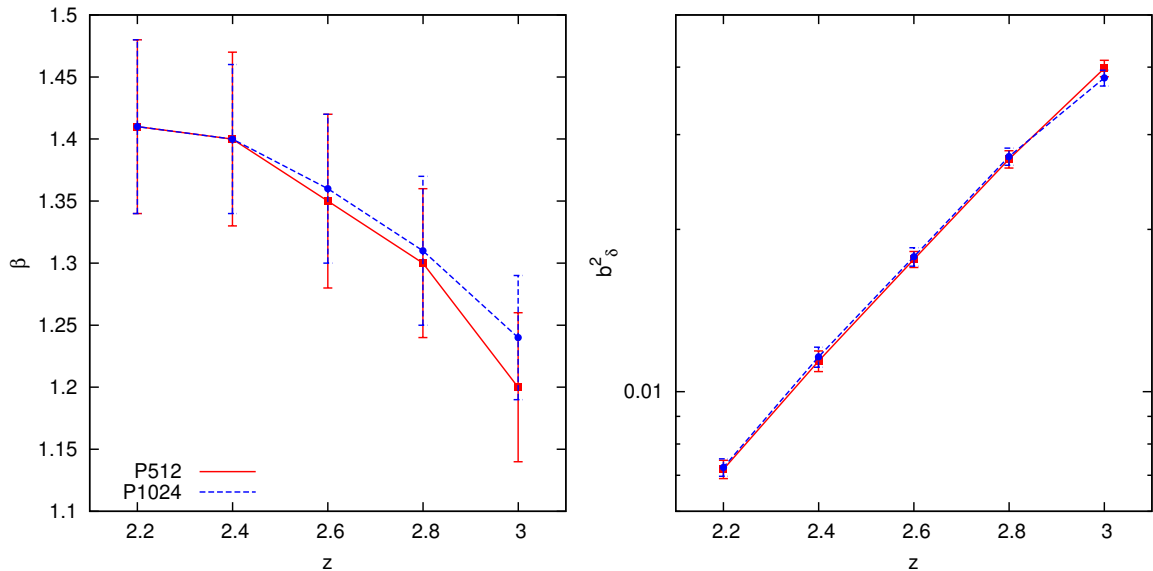

Figure 2.14: Bias parameters. Comparing the relevance in the fitting of increasing the spacial resolution along the LoS increasing the number of pixels in the line of sight whit the reference model. The effects go in the same direction as increasing the resolution, as seen in fig.2.7 .

2.3.5 Comparing Lagrangian and Eulerian simulations

Last thing to be tested is if the results are consistent despite the simulation that is being analysed, which means that they might be more general than the specific case studied.

For this work an Eulerian simulation could be also used, described in §2.2.1. Comparing it with the reference setting of the Gadget-2 simulation, the results for the bias parameters some differences, mostly in b_δ fig.2.16. The differences probably arise form many different characteristics of each simulation (see Table2.2.1). This translates in quite different power spectrum , seen in fig.2.15. When comparing the two simulations, all the postprocessing for both is the same, adjusting for different box sizes as done previously. In the event that the fiducial model can be safely applied to any given simulation or data sample, the discrepancies on the bias parameters could be used to select which kind of simulation better recovers the power spectrum of the IGM. Unfortunately since it is such a complex matter and depends on so many factors it will require a great effort to reach the point in which we are able to spot such differences. For the non linear parameters, the disagreement is even larger than the bias for most of the parameters fig.2.17. Most likely reflecting the combined effects of all the different physics and characteristics from both simulations.

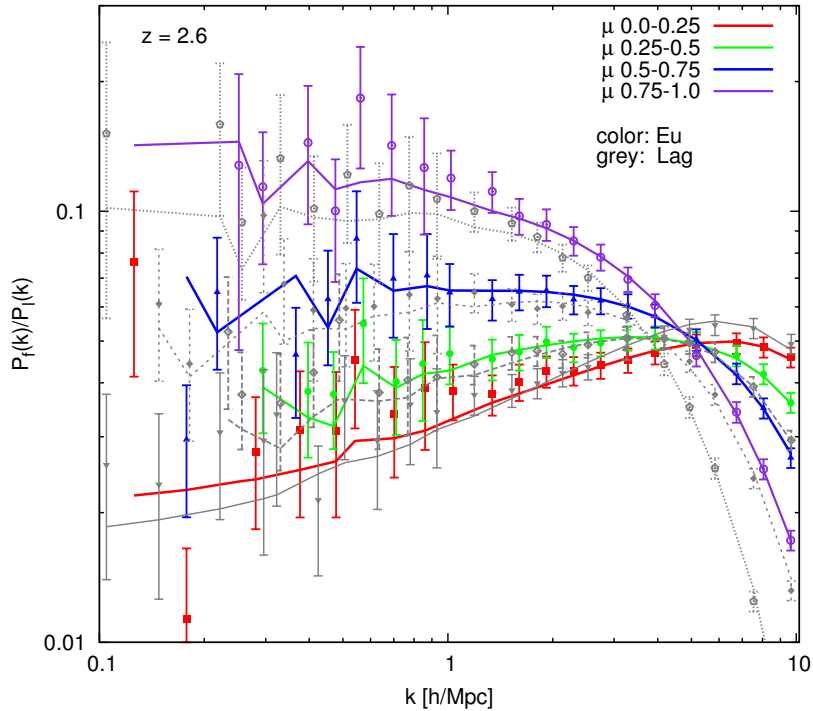**Figure 2.15**: Comparing the reference disposition for Gadget-2 in grey versus the Eulerian simulation at redshift 2.6, can be seen that the dispersion at low modes is much grater, affected also by the smaller size of the box and greater resolution, among many other differences. Visually can be seen as the fitted curves differ quite, which will translate to different values for the bias parameters.
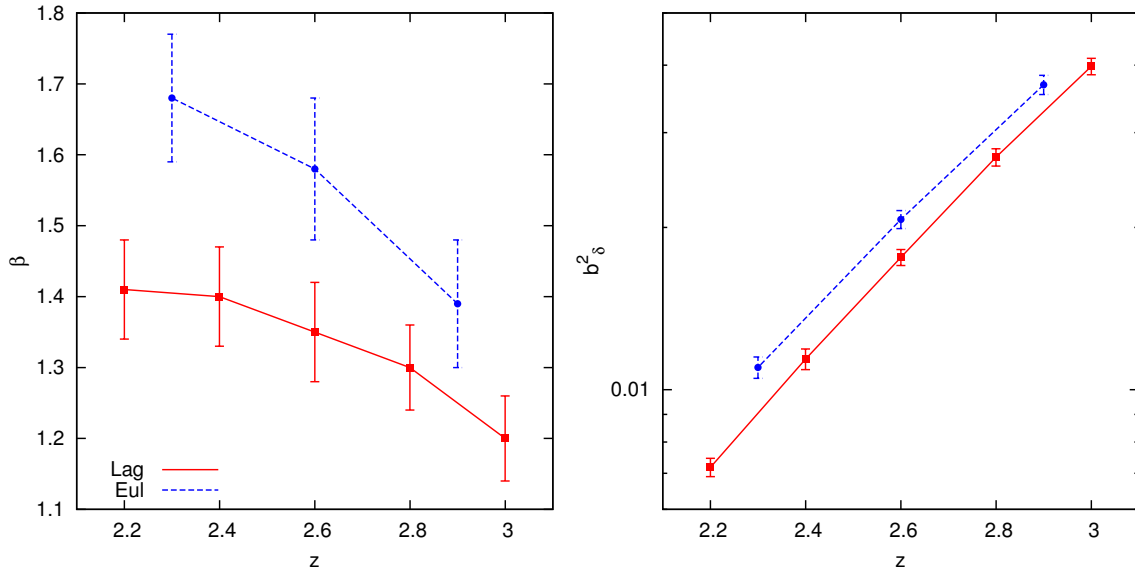
Figure 2.16: Comparing the 50Mpc/h Eulerian simulation with the 60Mpc/h reference for Gadget-2 . The discrepancies are most likely due to many major differences. With enough sensitivity using this method it could be used to check which simulation better reproduces observations.
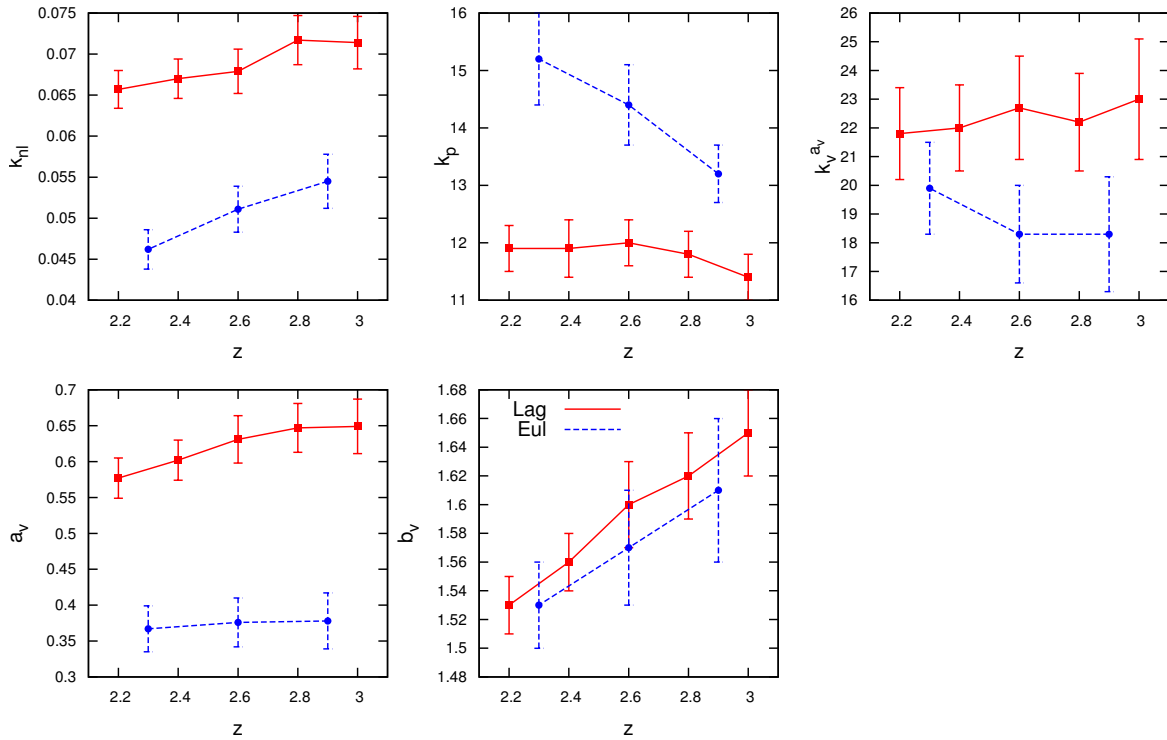

Figure 2.17: The non linear the parameters probably simply show expected differences in solving the non linear regions for Eulerian vs. Lagrangian simulations.

2.4. EFFECT OF DIFFERENT PHYSICAL PROPERTIES ON THE POWER SPECTRUM

All this work revolves around the bias parameters in the Lyα power spectrum . Therefore it is interesting to see how those are affected by the physics of the IGM, analysing those effects will help to better understand the physics of the IGM. If the bias are well known that will constrain the physics, or the other way around, i.e knowing the physics from the simulations and form them constrains can be extract for the bias, as done in this work. To understand how the physics affects the bias, and how that limits the possible characteristics of our universe, the best way to proceed is to play with the physics of the simulated universe, see what effect it has and which constrains can be found for the IGM. Four major ingredients that affect the IGM where chosen to play with, those are:

1. Mean temperature of the IGM.

2. Pressure-density relation (γ).

3. Amplitude of the perturbations (σ_8).

4. Mean flux of the IGM.

But firstly it will be analysed the the simple evolution with time, trough the different cosmological outputs from the simulations (redshift from 3 to 2). This is, the dependence of the bias parameters on the redshift, and how it is related whit the effect of the rest of the physics studied here.

2.4.1 Redshift

One simple yet important thing to observe is the evolution of the power spectrum with redshift, and with it, the bias and other parameters. When doing that, it is convenient to analyse the results from both Eulerian and Lagrangian simulations, despite the differences. That way more general results can be extracted.

To see the evolution on redshift for G-2, even if is not the standard $\gamma = 1.6$ but $\gamma = 1$, fig.2.27, fig.2.28 can be used. Here it can be seen how the linear scales are more important for high redshift, and how the amplitude of the power spectrum (highest to smallest power depending on μ) is smaller for high z. Also the point where the different cuts on μ intersect is more to the left (higher scales) for lower redshift, while the place where the power peaks for $\mu \to 0$ is more to the right for smaller z. This is general for all the cases and simulations explored, and simply reflects the growth of the non linear effects for a more evolved universe, as it is expected after all.

The evolution of the bias for the Eulerian and Lagrangian (G-2) simulations can be seen in fig.2.16. Both b_δ and β show a strong evolution with redshift. b_δ evolution is expected since the older the universe is, the more ionized it becomes. This erases power from the flux spectrum decreasing b_δ. The fact that the evolution follows a power law can be explained by the bias tracing the evolution of the mean flux that follows itself a power law as seen in eq.2.1.

For β no clear evolution can be expected since different factors play different roles, as will be seen in the next sections.   An increase in ionization results less absorption and more flux, which implies increasing β fig.2.19; but increasing temperature results in a decrease on β fig.2.19; finally increasing the perturbations results in bigger β fig.2.22. The evolution with redshift of the temperature-density relation is weak in the simulations, but it increases slightly with redshift, then an increase in γ means bigger β fig.2.25 but again this is a small effect.   Then there are opposite contributions, bigger perturbations and flux increase β, while higher mean temperature decreases it. As a summary $\beta \propto \sigma + \bar{F} - \bar{T}$. This accounts for the evolution on redshift and the change in trend, this change pointing to $\bar{T}$ gaining in relative importance over the others. Again this simplistic interpretation should be taken lightly, since other effects that evolve with redshift are not considered: from technical details with the simulations, to other physical aspects that can be of importance, like shocks or a major role of the temperature density relation, or different ways to proceed with the ionization. Also it is too simplistic since the ionization is related with the mean flux and also the mean temperature, which have opposite effects on β; those have been studied in different ways therefore more direct interdependence can not be ruled out.

For the non linear parameters, from fig.2.17 can be seen how the trends for all the parameters (except maybe $k_v^{a_v}$) are the same, despite the large differences in values. This points to some common evolution in redshift for the the effects that drive the non linear evolution of the power spectrum . Also can be noticed the quite neat linear evolution of b_v with redshift, and how despite all the other differences, both Eulerian and Lagrangian simulations agree quite well in both the values and slope of this parameter.

2.4.2 Mean Flux

One piece of physics for the IGM not well described yet is the ionization radiation background. Therefore it is interesting to see how different amounts of ionization, which translates in different mean flux, affects the data and the value for the fitting parameters. To achieve that a simple rescaling of the flux is done, in the same way of the renormalization of §2.4.2. It is done for one specific redshift of z = 2.6. The values to which the mean flux is rescaled are the ones corresponding to the mean flux for each of the other redshift outputs, that way they can be compared to the corresponding mean flux of each redshift. Direct comparison of the power spectrum can not be made since the power increases or decreases with redshift, however comparing the non linear power $D(k)$ is possible since are normalized to 1. In fig.2.18 can be seen how similar this terms are for 2 different redshifts, once they are rescaled to the same $\bar{F}$. At the same time can be seen how the differences are mainly dependent on μ, being almost the same perpendicular to the LoS but quite different along it. This shows how the non linear evolution can be well described by a change in the neutral content of the universe if the velocity distribution is not considered, but when it is directly involved on the resulting spectra, like in the LoS, other physics depending on redshift play a major part. This allows to disentangle the redshift evolution from the mean flux evolution.

All of that is reflected on the values of the bias. In fig.2.19 two effects can be observed: how the bias parameters change for different rescaled flux at the same redshift (blue line) and how they change with the normal redshift evolution (red line). The effect of the flux on b_δ is clear, being the main characteristic that dominates its evolution. However the slope still is different, hinting that other process have some relevance. For β it is seen how it also has a

strong effect on the evolution of the parameter, but accounts for a smaller slope of β, therefore having weaker effect on the β evolution with redshift. This evolution is expected from the μ dependence in redshift, mainly traced by β and other non linear parameters. This implies that this evolution is driven by other effects more dominant than the flux.

In fig.2.20 can be seen the dependence of the non linear parameters on the flux. All the parameters but b_v have inverted slopes. It is expected since the time evolution of the perturbations tends to increase velocity dispersion and hampers perturbation growth. This is enhanced along the LoS by the redshift distortions which tend to expand the modes, removing power along the LoS for smaller scales and increasing the power in larger scales. While a decreasing in flux tends to increase the power, since there is more neutral gas which traces more structures, especially the low density ones which better trace the linear evolution. This means that for lower flux the non linear terms are smaller. For the same $\bar{F}$ that translates in lowering the $D(k)$ for smaller redshifts, as seen in fig.2.18. It is curious that k_p has much stronger evolution only when modifying the flux, while is not that evident for the redshift evolution; translating the effect of increasing the mean flux as a pressure therm that hampers the growth of the perturbations.
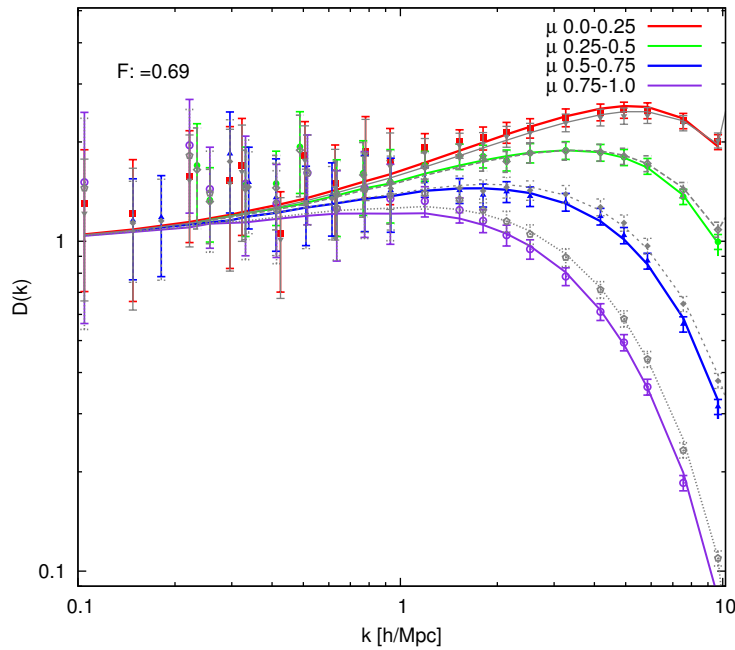

Figure 2.18: Non linear Part of the power spectrum fthe same mean flux but at different redshifts: Grey is the reference simulation at z=3; Colors are the rescaled power spectrum from z=2.6 to the mean flux of redshift 3. It can be seen how the differences are mainly dependent on μ, pointing to other effects dependent on redshift shifting it.
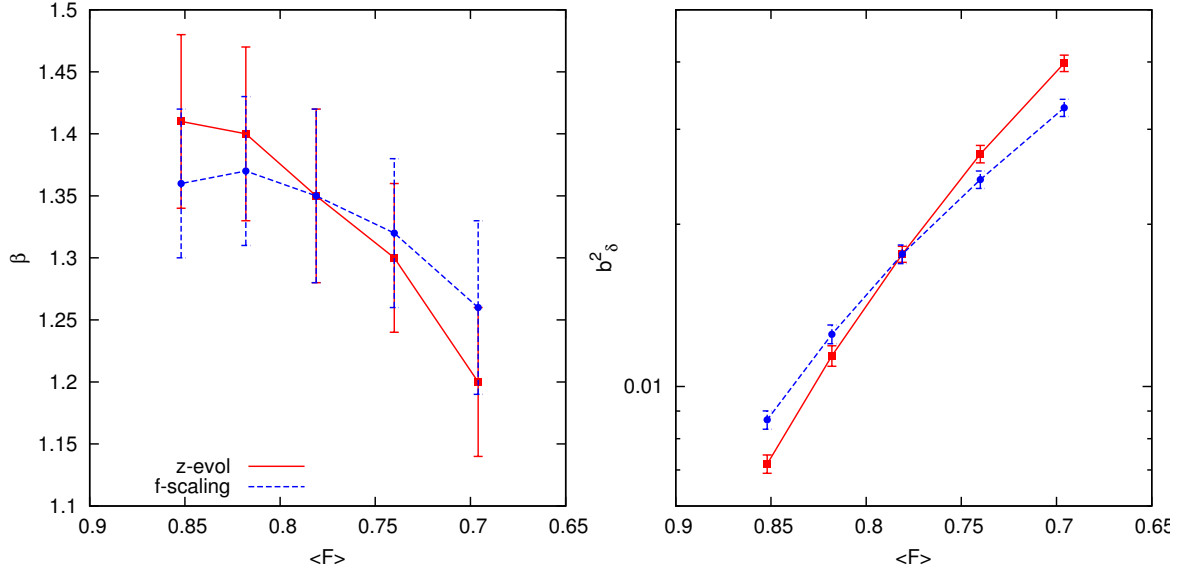
Figure 2.19: Values for the bias parameters. Rescaled fluxes compared with the equivalent flux for the analysed redshifts, being the same at the central value corresponding to z = 2.6.
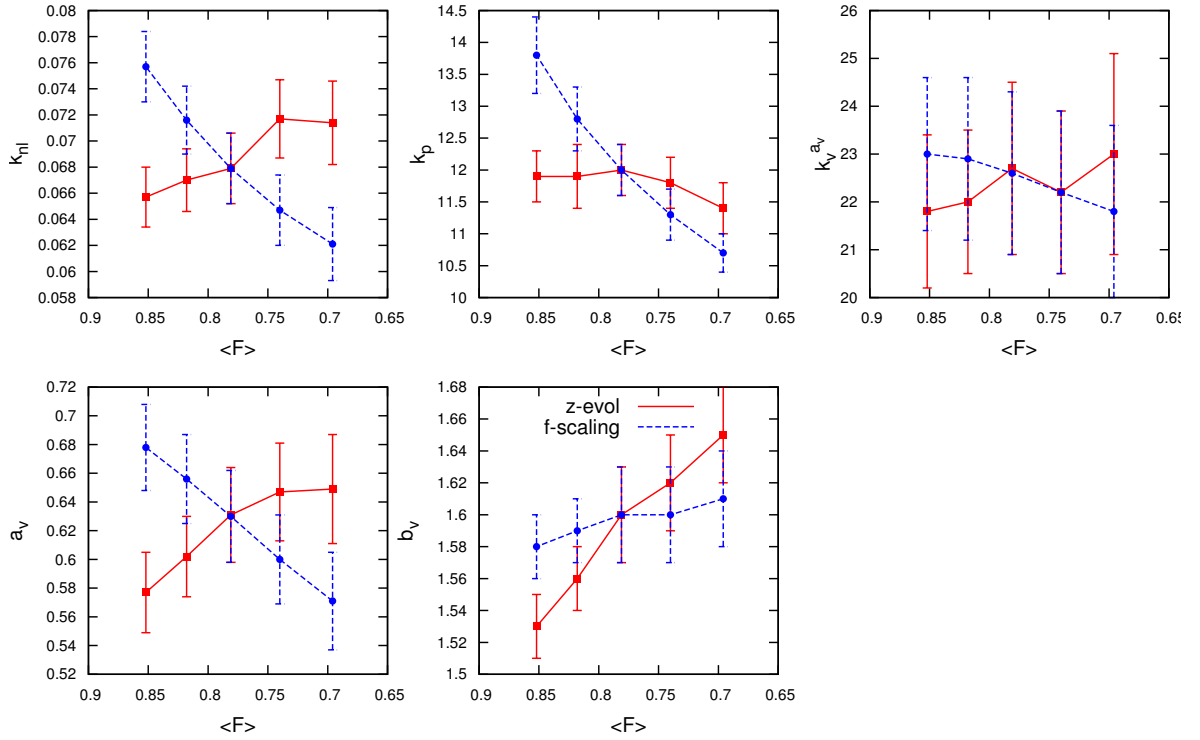

Figure 2.20: Evolution for the non linear parameters, rescaling the mean flux for redshift 2.6 compared with the values for the vias corresponding to the normal redshift evolution.

2.4.3 Overdensity amplitude

The amplitude of the perturbations σ_8, characterized by the amplitude of the linear power spectrum at 8Mpc/h, is not yet well constrain to de level of other cosmological parameters. It is seen that σ_8 has a strong effect on the structure collapsing since it will affect the the transition from linear to non linear scales, depending on the densities, having expected strong effect on the the redshift distortions and therefore on β.

This strong effect in β is clearly seen in fig.2.21 where there is a strong dependency on μ, this is reflected in fig.2.22. What is surprising is the small effect on b_δ despite the great differences on the power spectrum , being it much lower for large scales and larger σ_8. This points to a still a large effect of the non linear parts on the value of the bias parameters. There is also an strong evolution depending on k, being the effect stronger for k low and high μ or k high and low μ. All this makes one of the stronger effect seen, both in β and in the non linear parameters seen in fig.2.23, However it must be noticed that despite all those differences b_v remains the same even if it should represent the variation on μ for high k.

Although the effect on β is probably the most clear from the whole work, and has no strong redshift dependence, the values are still compatible within the given error bars, which means that the method should be greatly refined before strong conclusions can be derived from this results. This points to an effect due to σ_8, but is still far to give a quantitative dependence.
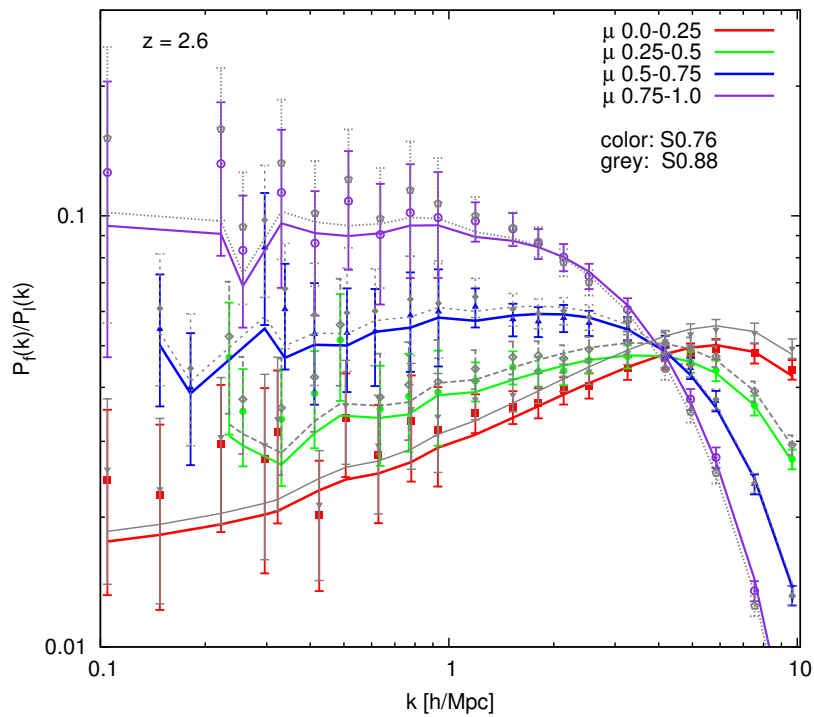
Figure 2.21: Comparing different power spectrum spectrums from simulations with reference $\sigma_8 = 0.88$ (grey) vs. $\sigma_8 = 0.72$ (colors). Can be clearly seen that it will have a strong effect on the redshift distortion parameter β, because of the strong dependency on μ. Also how the effect is stronger for low k and high μ or high k and low μ, but not low k and low μ or the opposite, pointing also to some dependence in k.
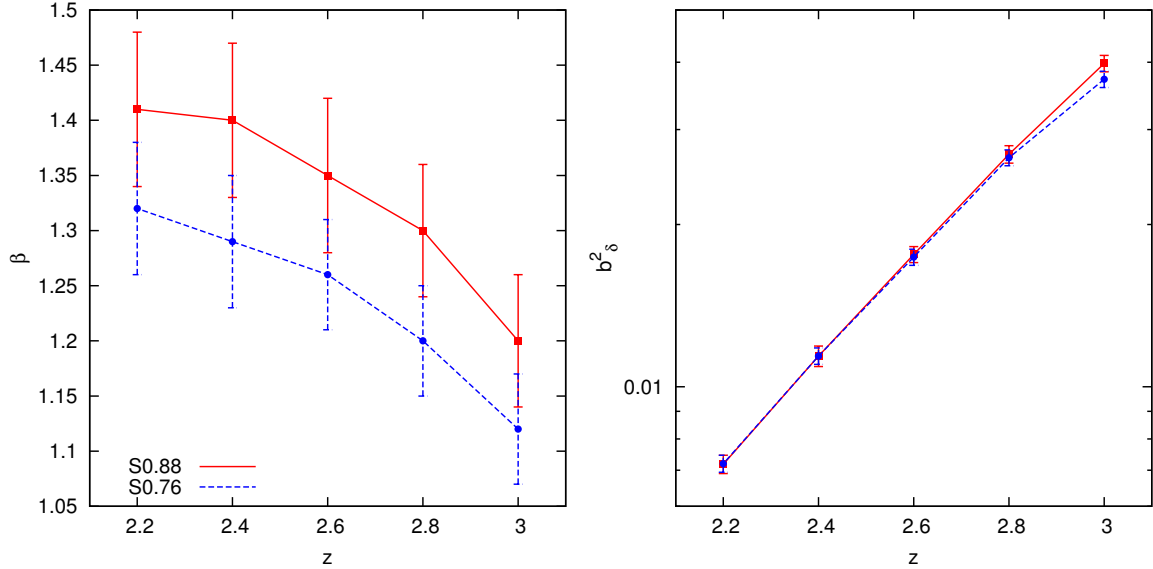
Figure 2.22: Comparing bias for the reference model with $\sigma_8 = 0.88$ and a model with $\sigma_8 = 0.76$ The overall effect of reducing the amplitude off the perturbations is a diminution of the redshift distortion parameter β as hinted by fig 2.21. There is not a different slope on redshift, which means that σ_8 has an overall effect with no redshift dependence.
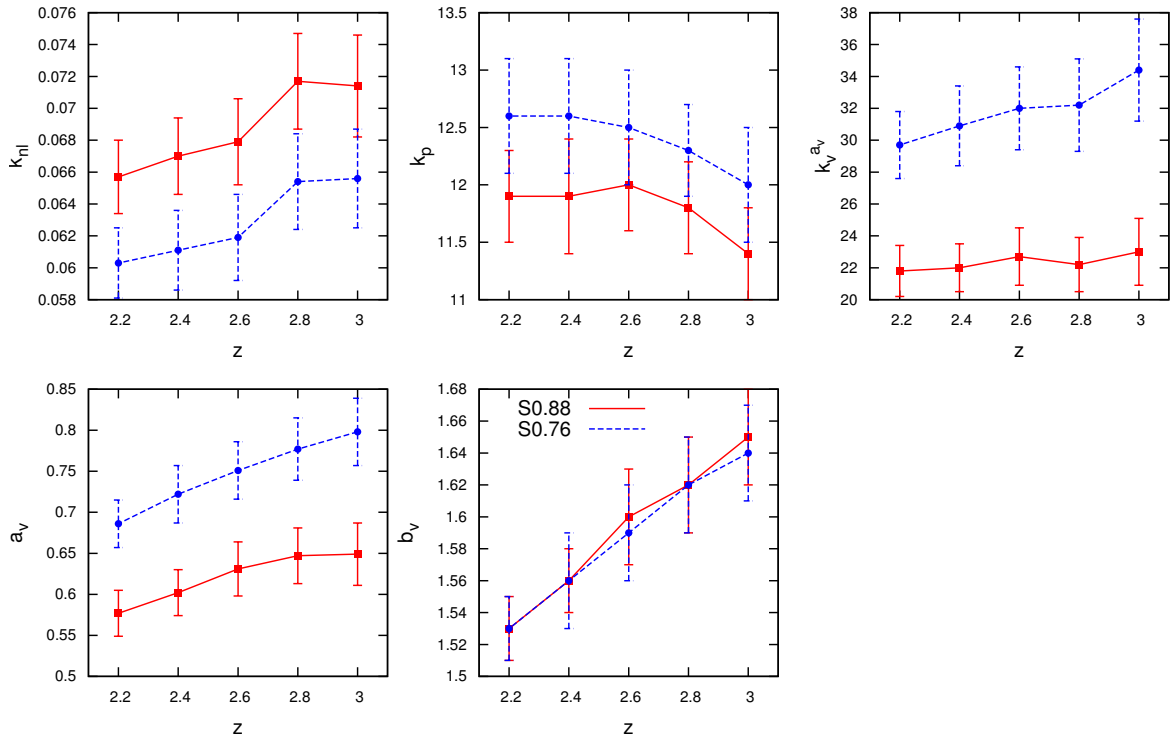
Figure 2.23: Effect of lowering σ_8 on the non linear parameters compared with the reference simulation (red line), it is interesting to notice that the effects although great for almost all the parameters is almost non-existent for b_v which should represent the power on μ for the non linear scales.

2.4.4 Temperature-Density relation

The temperature density relation (γ) models how hot the barionic gas is expected to be depending on its density, therefore $T \propto \rho^{\gamma}$ whit γ determining the relation between both. This is still not perfectly characterised and affects the evolution and growth of the structures. This effect can be better understood by introducing variations in the simulation and see how it alters the resulting power spectrum . In fig.2.24 can be seen an overall growth of the power spectrum for smaller γ, being more pronounced for small scales, which is expected as denser but not so hot halos produce more growth and therefore higher power. Also a slightly bigger increase in power for $\mu \to 0$ is present; this is even more pronounced for bigger redshifts, inducing to a greater reduction of β for higher redshifts as seen in fig.2.25. From fig.2.25, 2.26 can be also clearly deduced that the effect on the parameters when increasing γ is not linear; producing a strong modification in all parameters, going from $\gamma = 1.6$ to 1.3 but not much form 1.3 to 1.0. Meaning that for low enough $T - \rho$ relation it becomes a second order effect. Also from fig.2.26 is curious to notice how there is a strong redshift dependence on k_v^{av}, and it has a small inversion of the values respect to decreasing γ, probably both effect are related with the degeneracy between k_v^{av} and β which is the only other parameter that has some evolution on redshift.
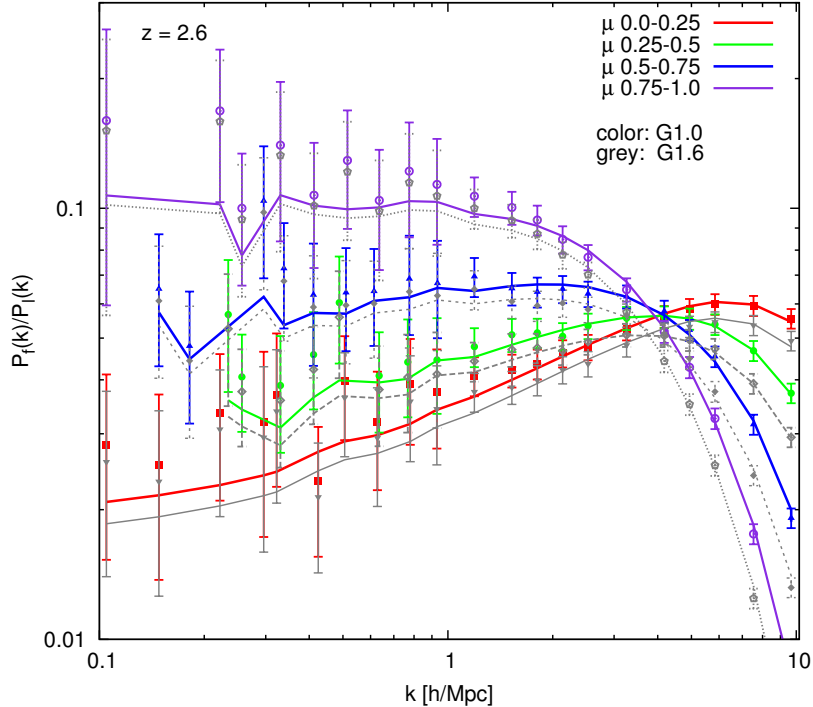

Figure 2.24:   Comparison of the reference simulation with $\gamma = 1.6$ (grey) versus one with $\gamma = 1.0$ (colors). The power is overall increased, being more significant for lower scales and $\mu \to 0$.
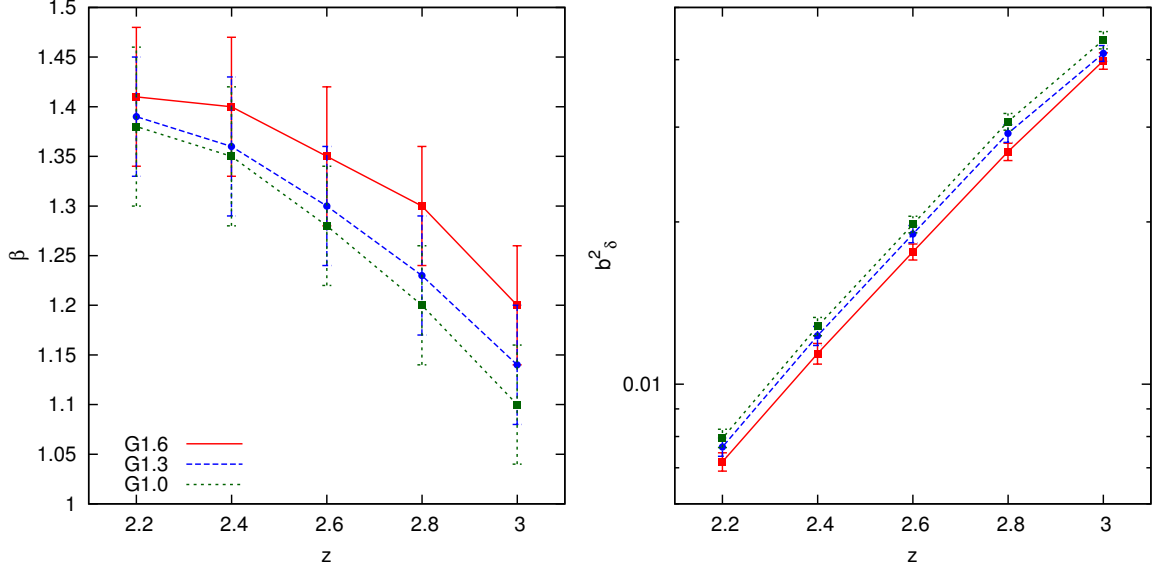
Figure 2.25:   Comparing how different values of γ affect the bias parameters for the reference simulation (L60, R512). It can be seen how it has a strong and noticeable effect on β, also minor in b_δ which represents the increase in power seen in fig.2.24.

2.4.5 Lower temperature for lower γ

Modifying the mean temperature at which the simulation develops is one major effect over the Lya, it can represent mechanisms as He II reionization, shell shocks, etc. Therefore is also interesting to study which is the effect that it has on the power spectrum and the corresponding bias, because so far the temperature of the IGM is not well characterized . Here has been chosen to lower this temperature for a $\gamma = 1$, where it has been seen in §2.4.4 that for low γ its effect is small. That way the sole effect from the temperature is more relevant. The way to alter the overall temperature is simply by modifying the ionization background until the desired temperature is achieved.

Overall lower temperature means lower power, except for the small scales, where it is bigger for all my values but $0.0 < \mu < 0.25$. When the temperature is decreased one expects more neutral gas, therefore more absorption and higher amplitude for the power, however with the renormalization where the optical depth is modified to have the same mean flux this effect is cancelled, this can be seen for the values Γ ; the other effect is that for lower temperatures the broadening of the absorption lines becomes smaller, therefore the transition between low density and high density gas becomes more sharp, and produces the overall effect of reducing the power, since there can be more saturated lines now which before, due to thermal broadening that expanded the absorption to different frequencies and contributed to different modes, resulting in more power for each mode. However this effect of increasing the power is overtaken by other effect at small scales and modes along the LoS. Here the decreasing in power due to the sharper absorption lines is overtaken by the redshift distorted structures $\mu \to 1$. Now those modes get enhanced due to sharper abortion contributing to the power of those scales.

For variations of the power, when $\mu \to 0$ the effect is better fitted by different values of
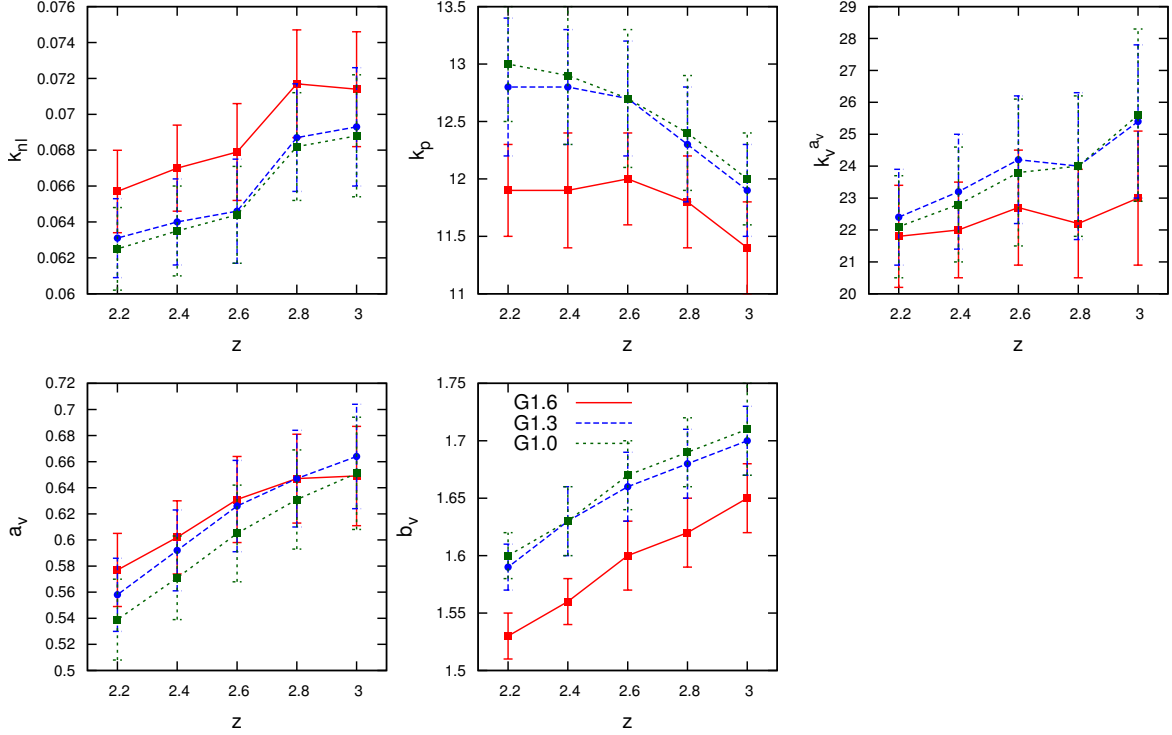
Figure 2.26:   Comparing how different values of γ affect the non linear parameters for the reference simulation (L60, R512). Here the differences are clear for all parameters for a transition from γ 1.6 to 1.3 but not so large for 1.3 to 1.0.

β, while for $\mu \to 1$ it is by b_v. This is clearly shown in fig.2.27 and fig.2.28, where visually the differences for the power spectrum are grater at z=3 for $\mu \to 1$ than $\mu \to 0$, while for z=2.2 is the opposite, resulting in the trends of β with redshift seen in fig.2.29, and in fig.2.30 for b_v. The opposite trend of b_v on redshift compared to β means that b_v scans better for differences when $\mu \to 1$ while β does it better for $\mu \to 0$. One would expect for redshift 3 to have lower β (if only the lower modes are considered) and higher β for redshift 2.2, but the effect is much stronger than expected, pointing again to strong relevance of the small scales to the bias. Nevertheless always consider that the errors for the fitting are large, and the fitted parameters have themselves large error bars form the χ^2 reduction method Finally k_p also has a trend on redshift, this time in the same way as β, here it can be clearly seen when comparing fig.2.27 and fig.2.28 that the differences at low scales are bigger for z = 2.2.
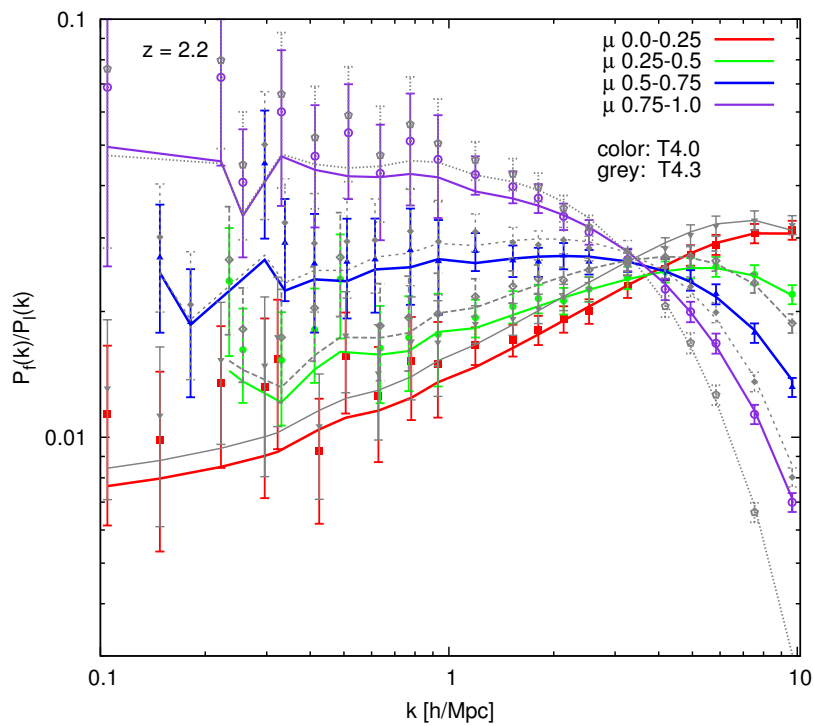
Figure 2.27: Different mean temperatures for the standard L60,R512 but with γ =1.0 at redshift 2.2. The effect is mostly an over all decrease of the power spectrum for lower mean T (color) although the power increases for high k and all μ bins except the lowest. Also there is bigger redshift distortion β for large scales and $\mu \to 0$, which translates to a higher fitted value for β.
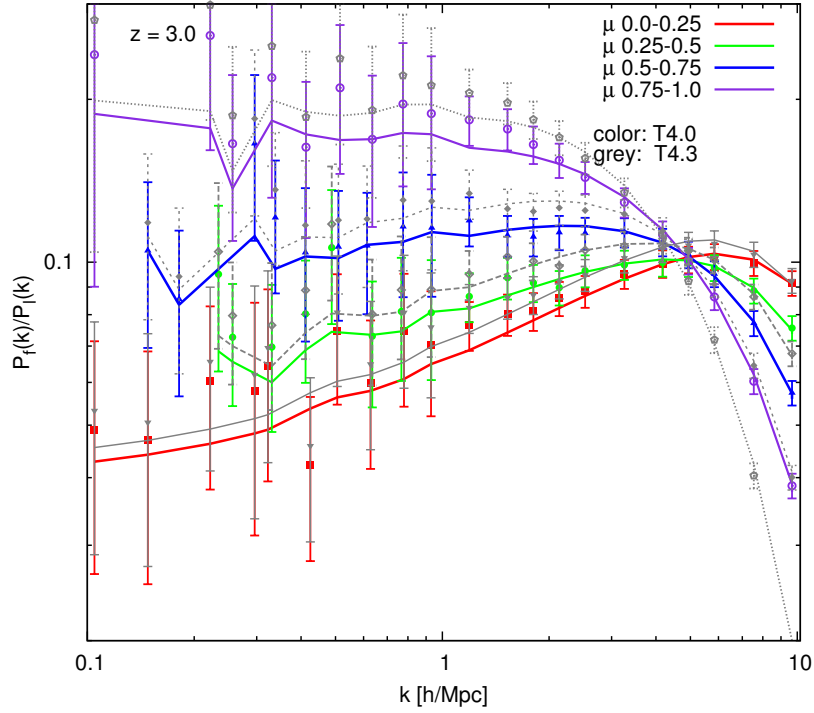
Figure 2.28: Different mean temperatures for the standard L640,R512 but $\gamma =1.0$ at redshift 3.0. Now with stronger redshift distortion β for large scales and $\mu \to 1$ than seen in the other redshift fig.2.27 , which translates to a lower b_v.
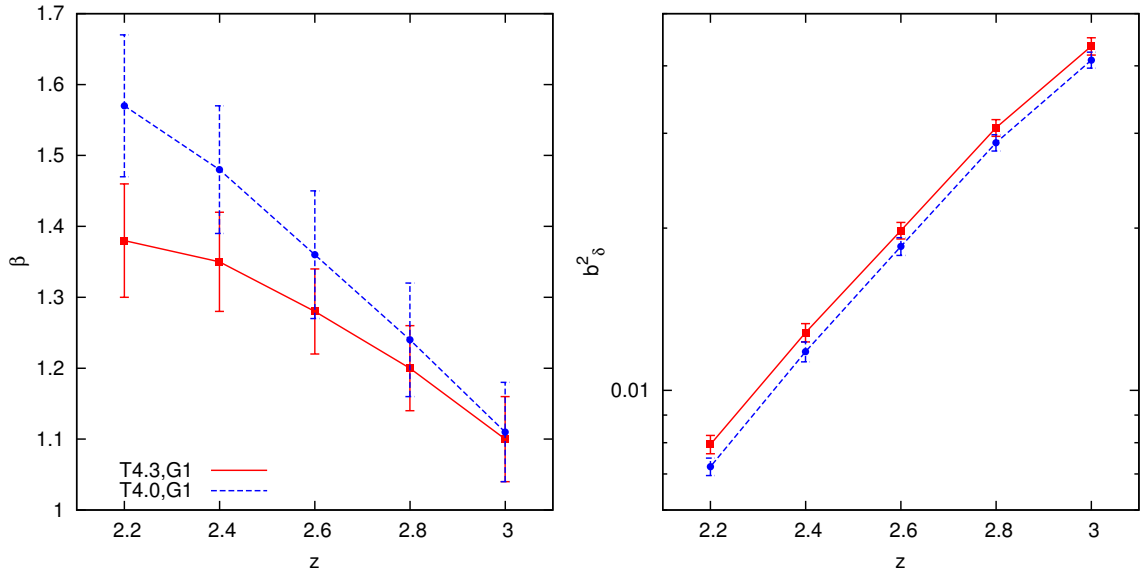
Figure 2.29: Modification of mean T for $\gamma = 1$ for the bias parameters. The effect over β is strong when lowering redshift, this results form the fitting eq being more sensible to changes when $\mu \to 0$, being the effect stronger for z=2.2 as seen in fig.2.27. For b_δ a small systematic decreasing for all redshift is what would be expected from fig.2.27, fig.2.28.
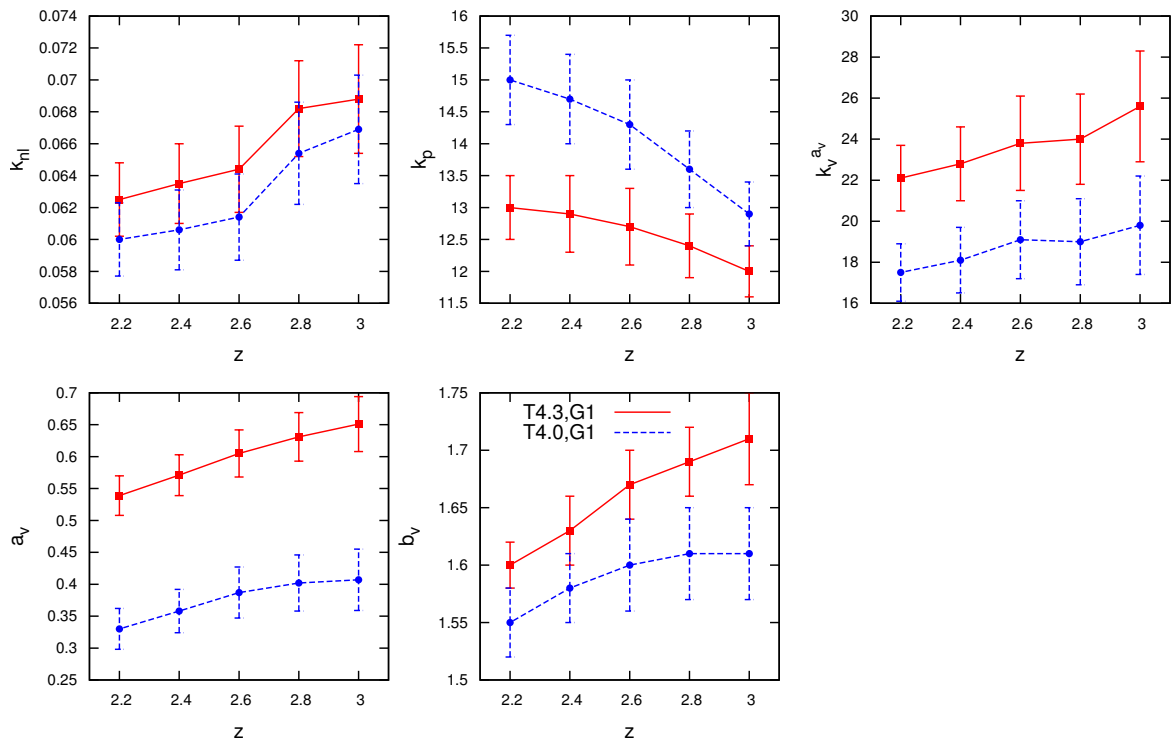
Figure 2.30: Non linear parameters for the different mean Temperature. It seems to have a clear effect on all the non linear parameters.

2.5. CONCLUSIONS

In this work a method has been created to extract the 3D power spectrum of from the transmission flux of a series very different hydrodynamical simulations, ranging from 0.1 h/Mpc to 10h/Mpc. This method has to be common for all simulations, allowing comparison between them and common predictions from the various simulations. To make sure that this is true a series of tests have been made that analyse how modifying some characteristics of the method affect the outcome of our analysis, with that a fiducial method has been established and used thorough the work. This specific methodology should be the one followed in future works to allow easy and direct comparisons.

Once the methodology is set, it is necessary to understand the effect of technical details like resolution or box since, these even though not being completely convergent show that with the current simulations specific predictions can be done with sufficient certainty. The main aim of this work can be pursued, the aim is to model the linear and non linear power spectrum in the form of a simple physically motivated equation with few parameters. Then the outcome from the simulations is fit to that modelisation. This has been done for the various simulations with different technical aspects and then with a fiducial simulation where the cosmology has been modified around a central, with that the effect of each variation on the model and fitted parameters can be seen and then build a small grid of these effects.

Crucially from this simulations and the model specific predictions and its uncertainty can be made for each simulation, and physical characteristics and cosmology, specifically estimation of the bias and redshift distortions of the Lyα have been made. This allows to test with observations these models and simulations done for redshift 3 to 2. This is aimed ultimately to be used to put constrains on these simulations, cosmological models and physics when comparing with Lyα surveys, whether current like SDSSIII-BOSS or future like SDSS-IV/eBOSS or DESI.

It is important to notice the relevance of works like the one done here, the transmission field of the Lyα forest is directly determined by the non linear physics that are only modelled by hydrodynamical simulations. As the number and density of surveys increases the perpendicular to the line of sight non linear scales are further and further mapped with better resolution, therefore the necessity of a good modelling of these that is already indispensable, will be even more necessary to understand and extract reliable cosmology information from these future studies.

This work is the baseline that lies the methodology and modelling, this can be easily extended in future work as long the same guidelines are followed to analyse many more simulations. This would allow ideally to generate a grid that assuming the simulations reproduce with fidelity the underlying physics of the universe, can be extensively used you constrain and understand the cosmology and structure formation.

A.1. BUILDING THE FIDUCIAL MODEL

Through this abstract a specific fiducial model is created. A fiducial model and a good characterisation of it is needed because an important goal of this work is to build a reliable method to measure the bias parameters form the power spectrum . In order to achieve that,

different simulations should be used to compare the results, finally providing a framework in which the comparisons are direct and independent on the steps of the method.

To build the fiducial model one has to correctly study the effect of varying each procedure. Each variation on the method produces slight that must be constrained to assert the robustness of the results. This is complex to proceed with since many factors play a role and have some minor or major effect that is coupled with others. This means that a systematic and foolproof method is difficult to be found, however finding a method that suits for the analysis and is reliable can be done, and it is done here.

Two main steps that determine the fiducial handling of the information extracted form the simulation are, on one hand *Processing* (how to prepare the data from the simulations to fit it with a model) and on the other *Fitting* (which model to fit and which fitting numerical method to use). Those are not fixed by any well determined mechanism, therefore an optimal configuration has to be found by educated guess. This forces to characterise how modifying them affects the results.

This appendix is organized as follows: In §A.1.1 teh different equations that can be used to fit the power spectrum are tested. In §A.1.2 the various equations are tested against the effect of using different ranges of scales. Next is to select the minimum error that a bin should have, this is done in §A.1.3. The same is done in §A.1.4 for selecting the binning to be used. Then is seen that the minimum error and the binning are related in §A.1.5. Similarly it is check in §A.1.6 the effect of selecting a different transition scale between binning and computing specific modes. Finally it is studied in §A.1.7 the effect of adding more simulations with same technical characteristics and physics but different initial conditions.

It should be reminded that this is a fiducial model build using L120 simulation as reference. It should be tested if the same or similar method will still hold for different simulations, or eventually if observational data that is eligible to be treated in a similar way. It is important to know that the final results are only given by this specific fitting and data processing method which will constitute the fiducial model. For the smallest boxes any modification on the configuration of those will alter (even if by an small amount) the result.

A.1.1 Different fitting equations

As mentioned in §2.2, depending on the fitting expression used, the results wil vary. Trough all the work eq. D_1 2.5 has been used to show all the results, however this expression is just physically motivation and its value is not fixed by any theory or empirical observation. This allows for other options to be considered, this are studied here in the same way as eq. D_1. For brevity only eq. $D_0$2.4 explored there. Several other fitting equations had been checked, those are not mentioned here, they simply where modifications of the ones shown, and which did not provide any improvement on the fitting. Here Eqs. D_0 2.4 and D_1 2.5 are compared, they are used to fit the non linear behaviour of the power spectrum , as described in §2.2.4.

The former formula D_0, extracted from McDonald (2003) had to be dropped.An important problem specific of D_0 is that it has not a perfect linear behaviour for the larger scales with the fitting points provided (it does not converge to 1 for the smallest modes), this is seen in the left panel of fig.32. Notice that in the this case, the errors attributed for the modes corresponding to large scales are much bigger than in McDonald (2003), here the total number

of simulations is much smaller, therefore the variance larger and that is represented as larger errors, this is described in §A.1.3.

A prevailing problem in the small 60Mpc/h boxes is that for both D_0 and D_1, when different limit scales k_l are considered as explained in §A.1.2, the fitted values returned for β change by a $10-30\%$. However when this is computed for the 120Mpc/h box, the fitted values are stable for D_1 but not D_0, implying that the fitting done with D_1 is more reliable.

A different formula has a strong effect to compute the bias parameters for the reference simulation (60Mpc/h, 512^3 particles), despite how similar curves it returns for the fit of the power spectrum fig.34. The effect on b_δ is due to the described inability to for the formula to achieve linear behaviour for large scales and high redshifts, that in its turn is caused by the degeneracy of b_δ with some of the non linear parameters, mostly k_{nl} and α_{nl}.

Finally the degeneracies found using D_1 eq.2.5 are shown. In fig.33, where the 1σ contours for the minimization of χ^2 is plotted.
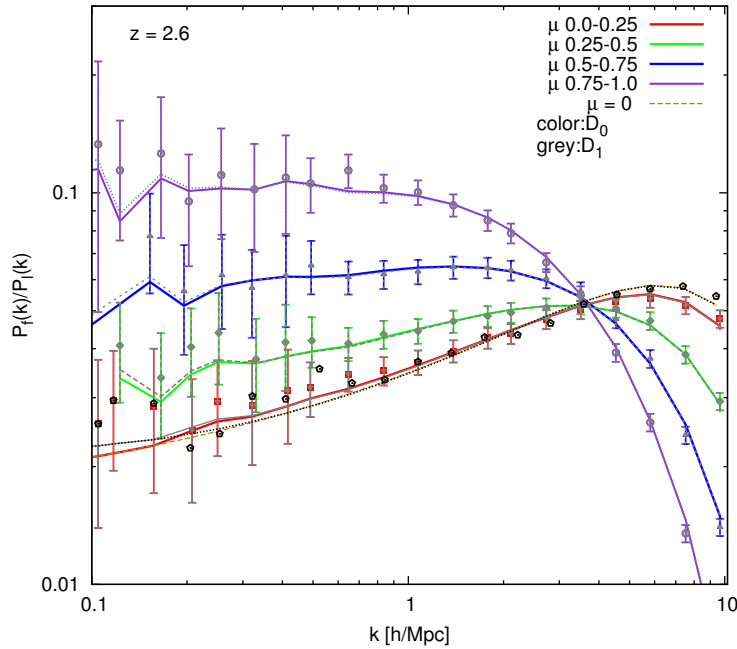
Figure 31: Comparison of the two fitting eq. D_1 (grey) and D_0 (colors), at z=2.2 for the L120 simulation. It can be seen that the differences are not strong visually, but that translates to noticeable differences in the value for the fitted parameters, mostly for the bias parameters.

A.1.2 Scale cut test

One easy test to check the behaviour of the above options is to cut part of the non linear scales, set a lower limit scale k_l, if they recover similar bias parameters it means that those are not strongly affected by the analysis done for the non linear scales. The cuts used are the default 10.47h/Mpc, half of that 5.23h/Mpc and also 3h/Mpc, which is not shown in fig.34 since has similar behaviour as 5.23h/Mpc cut. In fig.34 can bee seen how D_1 recovers similar bias values when scales up to different k_l are used, this tests plays a major role for selecting D_1 to be the fitting expression for the fiducial model, as is discussed in §A.1.1.
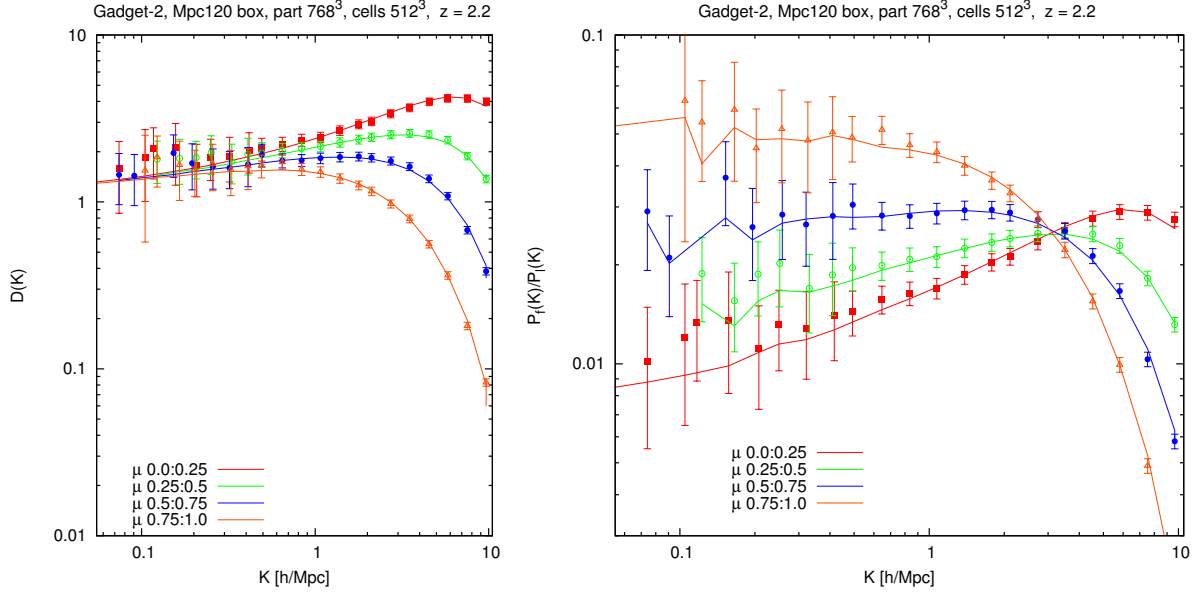
Figure 32: power spectrum data and Fitted curves with D_0 for L120, right panel. In the left panel the non linear terms are ploted and it is clearly seen how the non linear terms from $D(k, \mu)$ does not tend to 1 for the larger scales, meaning that they still have relevance on the linear terms, an effect that is not desired for a well behaved eq. To be compared with fig.2.1 where D_1 is used.

In order to correctly cut the smaller scales while retaining the valance between non linear and linear contributions, the weight and number of bins of the non linear part has to be modified. That is delicate to do because it has to be decided which equilibrium to preserve and how to do it. The method used is to move the binned interval to higher scales ($k_t = 0.8$) in order to preserve the same number of binned data points at the same h/Mpc intervals, while sacrificing modes computed for the specific values of k and μ. Other options consist in increasing the density of bins or reducing the error associated to the remaining bins, conserving that way the relative weight. Those different methods have also been also used to check the effect, returning similar results as the one shown.

A.1.3 Selecting the Error

In order to apply a fitting method, the points crafted from the simulation must have some variance, error or weights assigned. An optimal way to compute the variance associated to the simulations would be to obtain the dispersion of the modes, using several simulations with the same configuration but different initial perturbation field. However that is computationally time consuming.

The way chosen to do this is by using the total number of modes that contributes to the final value or fall in the k, μ interval. The bigger the number of elements the bigger is its weight and the lower is its associated variance or error. The way in which the final weight of points used for the fitting has strong effect on the results from the fitting procedure. Since the number of modes grows as k^3, the weight of the large modes which have a fixed binning
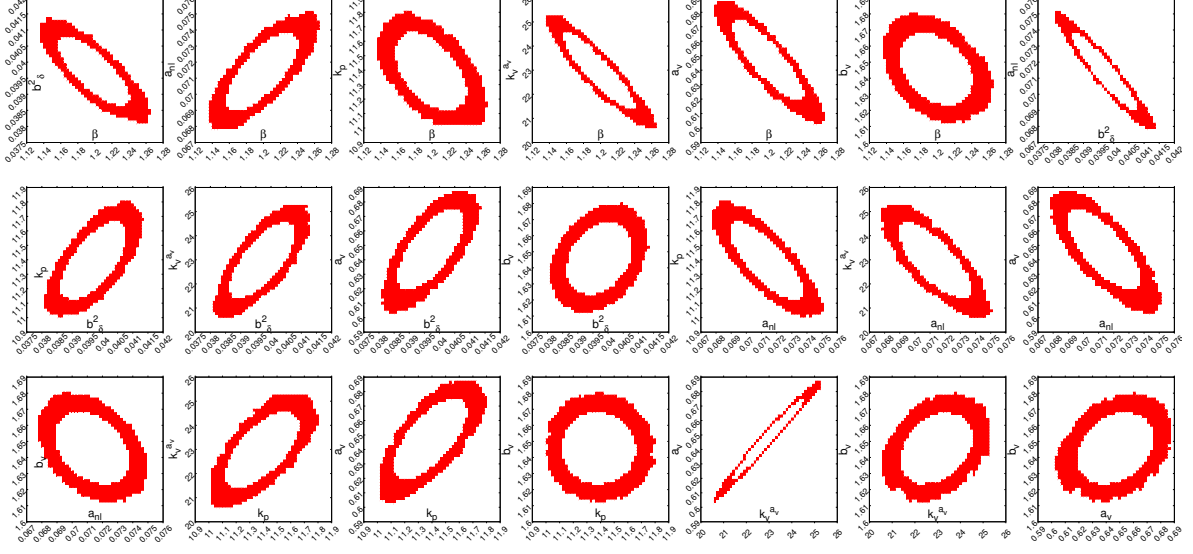
Figure 33:   Contours showing the degeneracies between the different fitting parameters for the reference simulation at redshift 2.6. Those contours show the 1σ region for 2 parameters obtained from the minimization of χ^2 using a MCMC. There are some degeneracies but all of them seem quite linear, showing elliptic shapes.

rapidly dominates the fitting. To avoid that a minimum "error" is fixed for all the fitting points, that way the smaller scales are no longer dominant in the fit.

The errors associated to the fitting points in this work are defined as:

$$\sigma(k,\mu) = P(k,\mu)/(\sqrt{w(k,\mu)}) + P(k,\mu) \cdot \epsilon; \qquad (8)$$

where $P_{3D}(k,\mu)$ is the power spectrum , $w(k,\mu)$, the weights and $P_{3D}(k,\mu) \cdot \epsilon$ is the factor added to avoid the higher k dominating the fitting. For large k, $1/(\sqrt{w(k,\mu)}) << \epsilon$ then the error simply approaches ϵ, for low k it only adds a small artificial uncertainty since the variance dominates those modes. The value ϵ accounts for how much this artificial increase of the error is.

To compute the weights simply keep track of the number of elements that fall in each value or interval of the new parameters space, r with just one more consideration, each element of the amplitude power spectrum has contribution of $2^{n_{dim}-1} = 4$, if one of the modes along one axis is 0 (only 2 dimensions, the perpendicular plane) the number of contributions is now $2^{2-1} = 2$. That must considered for the weighting, although they appear 4 times in the procedure, only 2 values are different, only 2 modes have information. The weigh associated is half the one given to the rest.

In fig.35 is shown the retuned fitted value of β at redshift 2.2 for different values of ϵ, maintaining all other characteristics of the method the same. z =2.2 is chosen because during the preparation of the work it has shown the most sensitivity to all the tests, due to the grater relevance of the non linear evolution, growing to larger scales. β is chosen as reference because also it is more variable than b_δ.It is seen how the fitting is affected by the value of ϵ, however the returned values for the fittings are stable as long as it is kept around few percentage points of the P_{3D} values. The values for two different box sizes (60 and 120Mpc/h) are plotted, can
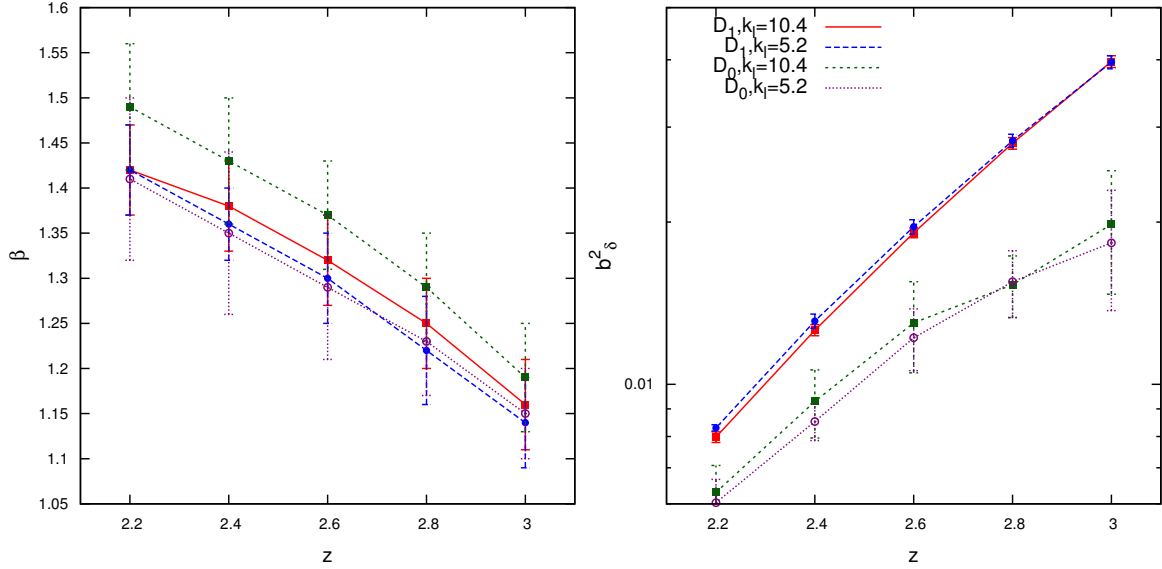
Figure 34: Comparison of the fitted bias parameters for the two eqs. when using up to different k_l, in this case D_1 to sclaes of k=[0.1:10.4] (red), and k=[0.1:5.2] (blue); D_0 scales k=[0.1:10.4] (green), and k=[0.1:5.2] (purple). Can be seen how D_1 behaves similar for both cuts while D_0 eq.2.4 changes despite having more parameters, that is due to the non linear terms having strong effect on the linear part. The differences in b_δ^2 between the 2 eq. show the effect of the non linear terms on the linear scales as those do not decrease to 1 for large scales, as seen in fig.32 .

be seen how for L120 the fitted β is stable for orders of magnitude in ϵ, that is not the case for L60 where it keeps evolving for big enough differences in ϵ.

To choose the fiducial value for ϵ, L120 will be used as the reliable source. For L120Mpc/h both b_δ and β are stable for different ϵ, but not in the case of 60Mpc/h. Then the value of ϵ is chosen in order to recover similar L120 fit. This fixes $\epsilon = 0.05$ that is, a 5% of the value of the P_{3D}. It is also practical since it is the one chosen in McDonald (2003). This value will be used as the fiducial one for all the work, it will not be modified unless otherwise specified.

Finally it has to be mentioned that although for z 2.2 the changes are notable, for z=3 similar tests have been done. In those cases the values returned for the fitting are really stable for variations on ϵ and all the other modifications, therefore the fiducial model is mostly constrained by the redshift 2.2, probably as hinted due to major relevance of the non linear behaviour reaching larger scales for lower redshifts as the perturbations have evolved and grown more. The same test have also been run for different number of particles and cells, all of them retuning similar results.

A.1.4 Effect of the binning of the power spectrum

As discussed in §2.2.3 the data for low scales must be binned and added together in intervals of $k\,\mu$, modes and angle from the line of sight for the perturbations. The effect of this binning on the outcome must be accounted for. Also must be considered that although it would be ideal to have as many bins as possible to better describe the data, this excessive binning has
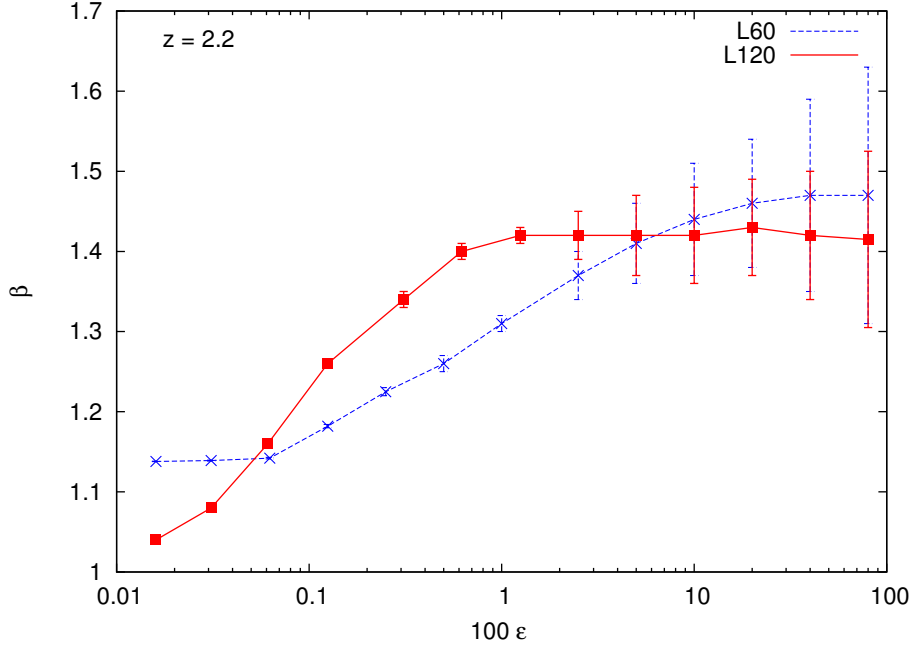
Figure 35: Returned values of β for the fitting with different ϵ (in this case the percentage is showed) in log scale at redshift 2.2 for L120 (red) and 60Mpc/h (blue). It is clear how L120 is sable for a great range of ϵ values while L60 changes greatly in the interval of interest, only stabilizing at very small or big ϵ. The crossing point at about 0.05 is chosen for the fiducial model. The errors increase or decrease simply because for greater ϵ a grater range of values can give similar χ^2, since only ϵ is modified.

also negative effects, i.e. considering too many data points or small binning adds more data on the non linear parts, out weighting it versus the large scales in which the number of modes is limited, also a large number of bins lows down the MCMC.

Choosing a convenient number of bins in k, μ space has to be done in similar way as choosing ϵ in §A.1.3, that means computing β for L120 and Fiducial 60Mpc/h at z=2.2 for different number of bins (as with ϵ other redshifts are also analysed, but z=2.2 has proven to be the most sensitive). The number of divisions in μ and k can be treated differently, but no appreciable effect has been seen when modifying one for a fixed value of the other with respect of varying the total number of bins by the same amount. The results are only sensitive to the total number of bins chosen, not the way the division is done in k, μ, prompting to chose the same number of divisions for both, the number of bins being the square of a integrer number. In fig.36 is shown variation of the total number of bins, it is clear how L120 is sable for the useful range of number of bins (not too large to slow down the MCMC, nor too mall to lose information) while 60 changes greatly in the interval of interest. They cross at little more than 200 bins, since number of bins is the square of an integrer then 225 and 256 are good options, but we settle at 256 since it means 16x16 divisions in k, μ and multiple of 2, for easy splitting. Notice that the total number of bins for L120 and L60 are not exactly the same, for comparison reasons when constructing the bins they are forced to have the same k interval when the boxes are different. For the case of 120Mpc/h that means adding few more bins, that explains why in fig.36 the number of bins is not exactly the same for 120 and 60.

As with the case of modifying ϵ only the total number of bins is modified, maintaining the rest of characteristics fixed. That translates in an increase of the errors returned from the MCMC fitting when less bins are considered, or a decrease when more bins are.
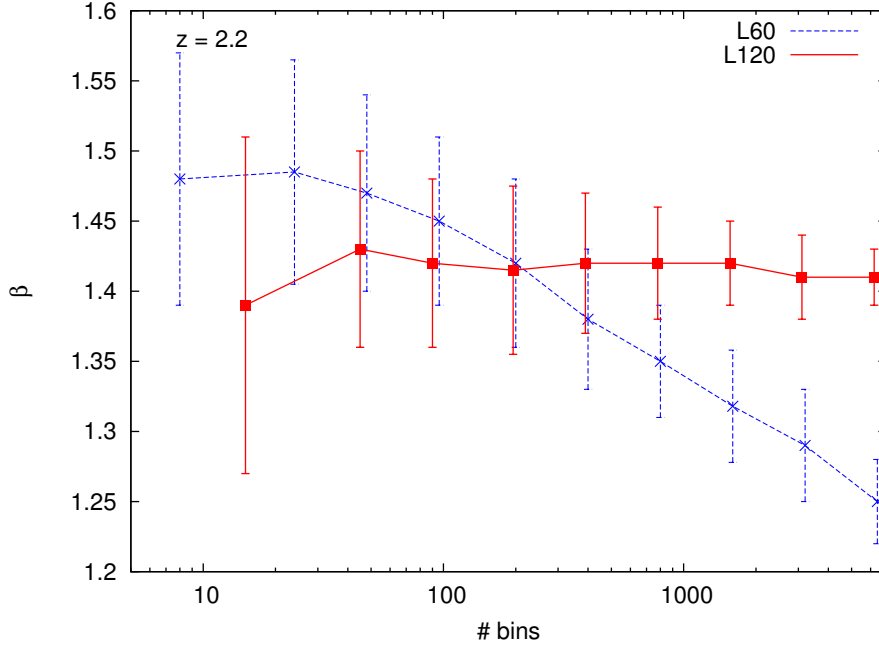
Figure 36: Returned values of β for the fitting with different number of bins in log scale at redshift 2.2 for L120, 640^3 particles (red); and 60Mpc/h, 512^3 particles (blue). The crossing point gives around 200 bins, that is translated in 14 k and 14 μ intervals. The errors increase or decrease simply because if there are less bins, less data points to compute χ^2, the fitting is less constrained.

A.1.5 Relation of error and binning effects

As can be deduced from fig.36 and fig.35 the influence of the increasing the binning is inverse to increasing ϵ. That means that those variations have opposite effect while have the same origins: giving more or less relative weight to small scales versus large ones, if the small non linear scales start to weight more (having smaller error or having more data describing those ranges), the bias is finally affected, going down to values that better describe the non linear scales rather than the linear ones. This also means that different combinations of ϵ and number of bins can give similar results for the bias, this is connected by the relation:

$$\epsilon \sim n^{3/2}; \tag{9}$$

where n stands for the number of total fitting points. The exponent $3/2$ is given by the cubic dependence of the number of modes with k, which reduces the error, and its square-root is simply the divergence. The values of $\epsilon = 0.05$ and number of bins $= 256$ have been chosen to build the fiducial model. An ϵ too small implies smaller error which does not reflect well the uncertainty, bigger ϵ implies bigger errors that might better reflect the uncertainty but also implies bigger number of data points, which will slow down the analysis. Also when

the errors become too large it is more difficult to extract any meaningful information from the data. Since computing directly all the modes for the power spectrum is computationally extremely time consuming, the procedure to reduce the data in bins, and then add an error is an artificial management of the raw data. Understanding those effects and selecting which best describes the original data is the objective of those tests. Therefore although the final values are chosen because of their convenience and the ability to return similar values to L120, other combinations are possible considering the degeneracy of n and ϵ. In fact has been tested that the values for the parameters remain the same if ϵ is modified in accordance of eq.9. The final values will be use for the large amount of simulations to be analysed, having a more or less constrain errors allow comparisons between the different simulations. With enough time it would be interesting to check which is the ideal combination and how it probably depends on the resdshift, resolution, and others. If in a future those results want to be compared with other computations is advised to use at least the same fiducial model if possible when comparisons are done. It should be easily doable and not time consuming.

A.1.6 Transition scale effect k_t

The effect of the of the scale interval for which the data points are computed exactly as opposite to the binned interval on k is also studied. This is needed for the fit, the fitting can be obtained directly form the value of the power spectrum at each k, μ, however there are too many modes in small intervals, it is computationally desirable to reduce the number of data points by staking them in bins. For the small k modes, on the other hand, is better if the specific modes are computed, since correspond the linear scales which are desired to be computed exactly. Since the number of modes grows as k^3, at certain k it is necessary to start the binning, that value will be the transition scale k_t. As with the other tests the transition scale is modified for the same simulation, for all redshifts and latter only for z = 2.2 both for L120 and Fiducial 60Mpc/h. This change differs with the others in that the results from the fitting are not sensible to it. For large differences in k_t the value for the different parameters remains almost constant both for L60 and L120. Only if a large k_t is used has the contra intuitive effect of slightly decreasing β, that is similar to giving more relative weight to the non linear scales. This can be explained because although there are more points in the linear regime, each one has bigger associated error and therefore its relative weight is smaller. Also those points have now more dispersion as they are not averaged over, therefore the fitting procedure would not be as sensitive to those points as compared to the less dispersed ones from bins. Those effects can be checked by increasing ϵ, which corrects for that variation.

Given that the selection of k_t has no big relevance on the results, and that a large value for ti is not desired as the number of modes increases rapidly with k, slowing down the computation, a value of $k_t(60) = 1.0h/Mpc$ is used as a transition scale, it is arbitrary scale, but is well inside the linear regime at those redshifts. It corresponds to 296 modes in k, μ for the 60Mpc/h box, which is a convenient number as is of the order of the number of bins chosen for the low k modes, i.e. 256 as seen in §.A.1.4 .

Note that if the same number of modes bellow k_t are to be maintained for different box size then k_t will have to change. If the box is doubled k_t must be halved. This is only necessary if this specific quantity wants to be maintained for comparison. Maintaining this number of modes constant translates to $k_t(120) = 0.5h/Mpc$ for 120Mpc/h.

A.1.7 Number of seeds

Another test is simply to check how using the same simulation with different initial conditions, a different seed, might affect the result. Also should be check if combining axis produced from different seeds is or not equivalent to combine the 3 axes from the same seed.

No appreciable difference has been observed when conniving 3 different axes from simulations with different seeds in the same way as done combining the 3 axes (x,y.z) of a simulation box fig.A.1.7, confirming the validity of the method of using the 3 orthogonal axes of a simulation to increase the statistical weight obtained from one given simulation.
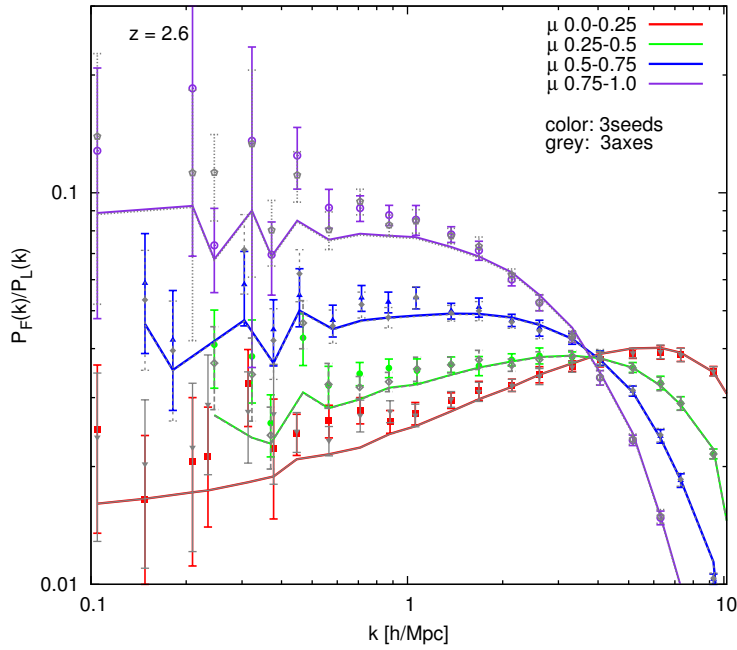

Figure 37:   Power for 1 box combining the Fourier transform along its 3 axes (grey) and combining 1 axe of 3 different simulations (colors). It can bee seen how the differences are minimum (only appreciable for the small modes) and the fit is virtually the same.

3. Metal content of the Damped Lyα Systems: the Metal Strength

> "Quantity has a quality all its own."
> — Joseph Stalin.

3.1. INTRODUCTION

For fist Time in history we are able to access to the high redshift universe in astonishing large numbers, the Baryon Oscillation Spectroscopic Survey Dawson et al. (2013b) inside the Sloan Digital Sky Survey-III (SDSS-BOSS)Eisenstein et al. (2011b) has obtained spectra for more than $160\,000$ quasars at $z > 2$. These spectra provide access to not only to the the source of them, the quasars, but also to the high redshifts gas that lies in between the emitting source and us. This in its turn is at a redshift high enough that we can observe from the ground the originally ultraviolet absorption features produced by neutral hydrogen and metal species that form this gas in the spectra of the background sources. The main absorbing features are objects that are present in a fraction of all the observed spectra, these are Damped Lyman alpha systems (DLAs from now on) , they are generally defined to have column densities $N_{HI} > 2 \times 10^{20}\,\mathrm{cm}^{-2}$. The SDSS-BOSS survey provides an unprecedentedly large sample of DLAs. These are systems with high column density, this systems have two useful characteristics: firstly the damped profile of their hydrogen Lyα line can be measured, providing the column density can from the absorption profile, secondly the high column density provides self-shielding of the external cosmic ionizing background, this implies that the hydrogen in these systems is mostly in atomic form (see Wolfe et al. (2005) for a review). then using the DLAs we have access to a sample of the atomic gas repository at that time. Their column density distribution yields directly the mean density of baryons that are contained in these systems, this account for a fraction of the critical density $\Omega_{DLA} \simeq 10^{-3}$ at redshifts $2 < z < 3.5$, or $\sim 2\%$ of all the baryons in the universe Noterdaeme et al. (2012).

DLAs have been known for a long time, bun only now in such large numbers. Previously studies of high resolution spectra have revealed fine characteristics like a diversity of *velocity structure* of the absorbers which is characterized by multiple components, ranging form single component with the typical velocity width of photoionized gas clouds of $\sim 10\,\mathrm{km\,s}^{-1}$ up to $\sim 100\,\mathrm{km\,s}^{-1}$ Prochaska & Wolfe (1997). The derived *metallicities* are generally low,

distributed over a broad range of 10^{-3} to $10^{-1} Z_\odot$, and on average declining slowly with redshift (see Rafelski et al. (2012) and see also Kulkarni et al. (2005)). The complex velocity profiles suggest a highly turbulent environment, and models of gaseous galactic halos in which cool clouds move in random orbits can generally explain the observations (e.g., Haehnelt et al. (1998), McDonald & Miralda-Escudé (1999)).

The nature of the DLAs within the paradigm of the Cold Dark Matter model of structure formation should follow the model for structure of hierarchical growth, starting from a mass power spectrum that predicts a number density of halos as a function of their mass at each time of the universe.

The need to better understand characteristics of the DLAs like the velocity structure and metallicities has impulsed us to the creation of new and creative method to use the data from large surveys like BOSS and be able to extract the most juicy results that can light our understanding of the DLAs and the universe in which they reside.

Despite the relatively low resolution and signal-to-noise of the BOSS spectra, the large number of observed DLAs allows one use high number statistics to study global properties of the systems.

We aim to develop a method to pinpoint two of the most important characteristics of the DLAs, as mentioned earlier: the velocity structure and metallicity, that are traced by the equivalent width of the metal lines of the DLAs. Being able to study them in large numbers even for low resolution spectra promises to deliver new and interesting results. For that end we developed a new method to compute equivalent width for 17 of the strongest low ionization lines associated with the DLAs, this methods is able to measure an equivalent width regardless of how weak or noisy a signal is, this has the drawback of having measurements that are not very reliable but on the other hand allows us to have large numbers of them, and several metal lines per DLA.

Then we aim to have the strongest way to use this wealth of newly measured equivalent widths. Therefore we crated a parametrisation that combines all the lines measured for a DLA, weighted by the expected mean value of the global population and corrected by the natural evolution of the equivalent width of the metals with the column density of the host DLA. We have named this parameter *Metal Strength* and we infer that it is a way to be able to account for the mixture of different gas phases, velocity dispersions and the metallicity. By the nature of the Metal Strength we can compute it for almost all the DLAs from BOSS. With that we are free to group the DLAs as it best suits our interests to study properties of the DLAs themselves and the universe and structures in which they reside.

This approach has the advantage of providing directly properties of population of DLAs, and allows to group them by their mixture of different gas phases and metal content, which is individually complex to model because of their large intrinsic random variation, but we can try to model it from these populations with large samples and from which we can extract average values.

Once we have group a set of DLAs we can extract average characteristics of the group by staking them, this allows to average out the intrinsic fluctuations of individual DLAs and see fine and weak features that would be otherwise lost due to the noise and intrinsic variability.

Another interesting property of the DLA population that has been extensively studied is the redshift evolution of their metallicity. Although the Metal Strength computes a combination of velocity dispersion and metal content, thanks to the large number of objects at

our hands we can proceed to study the mean redshift evolution of the Metal Strength with high precision. This provides the ground to be compared with other studies and models and simulations that try to mimic the metallicity and velocity evolution of structures like the DLAs.

Finally on the grounds of cosmology recent observations of the large-scale cross-correlation amplitude of DLAs with the Lyα forest absorption have determined their mean bias factor $b_{DLA} \simeq 2$ Font-Ribera et al. (2012), which indicates that the DLAs are distributed over a broad range of halo masses $10^9 M_\odot \lesssim M_h \lesssim 10^{13} M_\odot$. We can now compute the bias not just for the global average population but also for subpopulations according to this new parametrization, grouping the DLAs depending on their Metal Strength.

Therefore in this work we describe our method to compute these equivalent width of 17 metal lines for the DLAs in §3.3, then we present the way to compute the Metal Strength of the DLA and its correction by the column density in §3.4 and then we proceed to present a newly compiled catalogue containing the equivalent with for these lines and the value of the Metal strength for each DLA §3.5.1. We can then divide the total sample in three populations depending on their Metal Strength and stack them to see the evolution of the mean of each population with Metal strength §3.5.3, and see the evolution of the Metal Strength with redshift §3.3.2. In §3.6 we discuss the results.

3.2. DATA SAMPLING

For our study we make use of a DLA catalogue containing 34 366 DLA candidates (but only 19 698 have $\log N_{HI} > 20.3$, the column density used in the literature to catalogue an absorbing system as DLA) at a redshift between 1.9 and 5.7, this catalogue will be referred from now on as DR12DLA. DR12DLA is made from the SDSS-III BOSS Data Release 12 Quasar Catalogue (DR12Q). This is an update of the catalogue produced from Data Realese 9 see Noterdaeme et al. (2009) in which the detection of the DLAs in quasar spectra is made through a fully-automatic procedure based on profile recognition using Spearman correlation analysis based on Voigt-profile recognition from Noterdaeme et al. (2009). The data used in the present work is obtained through the application of this technique to the Pâris et al. (2014) to updated to DR12Q, it is basically an update of the DR9 DLA Catalogue Noterdaeme et al. (2012)).

Our calculations are performed over almost the whole catalogue in order to obtain big statistical samples for the overall DLA population. However, it is important to notice that this DLAs catalogue was designed to be as large as possible and with minimal cuts on data quality and robustness of DLA detection. Although by definition DLAs are those systems with hydrogen column densities greater than 2 x $10^{20} cm^{-2}$ we here use all absorption systems with $N_{HI} \geq 10^{20} cm^{-2}$. Due to this, some systems can be false positive DLA detections or can have significant uncertainties for the redshift and N_{HI}.

For each DLA we use the quasar continuum and noise for their measured wavelength from BOSS, which is measured in units of $log_{10}(\lambda)$ This corresponds to a fixed velocity pixel of 69km/s which i. All the operations will be made in velocity space because of the fixed width of the pixel. Just the equivalent width will be computed in Angstroms at the end.

The estimated redshift of the Quasar and DLA, its continuum to noise, and column density are extracted from Notredam's work. Also the fact that they might be flagged as a Lyβ form another DLA on the same quasar spectra, the objects flagged can be not used to avoid contamination with obvious false positives.

A table with a total of 872 skylines detected in the SDSS-BOSS survey is also used to avoid these parts of the spectra contaminating the measure of the equivalent width. A list of the the sky lines and description of the sky mask can be found in Blomqvist (2014). Also we make use of the strongest emission features of the red part of the quasar continuum seen in Pâris et al. (2012) to avoid using them to normalize that region of the spectrum.

3.3. Determining Metal lines equivalent width (W)

In this section we will proceed to compute the equivalent width of a set of lines that we consider can best give a estimation of the level of metallicity of the DLAs. We proceed to describe which are these lines to be used, how they are selected for each DLA and how their W is measured. Finally this allows us to make an update to DR12DLA catalogue, we add more W computed metals and the value measured with our method for the existing metals, which we have realized traces better the equivalent width.

3.3.1 Metal lines used for the analysis

For this work a set of metal lines will be used as a proxy of the metal content of the DLA. This lines are *17 low ionization* transitions that correspond to the lines with a mean $\bar{W} > 0.5$ from BOSS, as measured from Mas-Ribas et al. (2014), in the region between $1260\mathring{A}$ to $3000\mathring{A}$ at restframe, that means that they are redwards from the Lyα line, but within the spectrograph range. With the exception of and $MnII\, 2576$ since it has $\bar{W}$ just slightly > 0.5 and also since it has a long wavelength that usually falls in the reddest and nosiest part of the spectrum, making its signal really noisy.

In Table 3.3.1 those 17 lines are shown, it can be seen that they are low ionization lines in the restframe range mentioned. The equivalent with computed from staked DR10 spectra Mas-Ribas et al. (2014) is shown, this will be used latter in §3.4, it is compared with the mean value that we measure from our method to compute W, the description of this method . Then to have a vision of the relative importance of each line, we show the mean contribution of each line to the total weight of a computed $MetalStrength$ that we will use latter in §3.4. This basically shows how much on average that line contributes to the strength when that line is used.

$$C_i = \frac{1}{N_i} \sum_j \frac{\left(\bar{W}_i/\epsilon_{i,j}\right)^2}{\sum_i \left(\bar{W}_i/\epsilon_{i,j}\right)^2}, \tag{3.1}$$

where W_i is the equivalent width for each i metal extracted from the stack as shown in the table, $\epsilon_{i,j}$ is the error of the i metal line measured for each j DLA, and N_i is the number of times that each i metal has been measured.

The last column shows fraction of times that each line is used (F), $F_i = N_i/N_T$ where $N_T = 33\,939$ the total number of DLAs where at least one line is measured. It can be seen how non of the lines can be used for all the DLAs, specially the lines that are close to the limits of the wavelengths used.

Table 3.1: Metal lines used and their principal characteristics used for this work. The name of the metal line appears with its rounded metal transition at vacuum and at restframe. Then the W are shown for the stacked spectrum of the DLAs in DR10 Mas-Ribas et al. (2014), this is compared with the $\bar{W}$ computed in this work, the errors shown are the variance divided by the square-root of the number of lines measured. The mean contribution of Each line is shown as C, and the fraction of the total $33\,939$ DLAs in which a line is present is shown as F.

Name	$\lambda\,\AA$	Staked W DR10	$\bar{W}$	C_i	F_i
$SiII - Fe\,1260$	1260.42	0.53 ± 0.03	0.79 ± 0.02	0.37	0.02
$OI - SiII\,1303$	1303.20	0.68 ± 0.02	0.79 ± 0.011	0.34	0.20
$CII\,1334$	1334.53	0.50 ± 0.01	0.53 ± 0.006	0.46	0.35
$SiII\,1526$	1526.71	0.38 ± 0.03	0.38 ± 0.004	0.34	0.68
$FeII\,1608$	1608.45	0.19 ± 0.01	0.21 ± 0.004	0.11	0.67
$AlII\,1670$	1670.79	0.39 ± 0.01	0.39 ± 0.004	0.35	0.69
$SiII\,1808$	1808.01	0.05 ± 0.01	0.08 ± 0.006	0.02	0.67
$AlIII\,1854$	1854.72	0.10 ± 0.02	0.15 ± 0.006	0.04	0.61
$AlIII\,1862$	1862.79	0.06 ± 0.01	0.12 ± 0.008	0.01	0.61
$FeII\,2344$	2344.21	0.44 ± 0.02	0.42 ± 0.02	0.17	0.22
$FeII\,2374$	2374.46	0.21 ± 0.03	0.31 ± 0.03	0.04	0.19
$FeII\,2383$	2382.76	0.57 ± 0.03	0.61 ± 0.03	0.23	0.18
$FeII\,2587$	2586.65	0.42 ± 0.04	0.45 ± 0.02	0.10	0.14
$FeII\,2600$	2600.17	0.63 ± 0.07	0.61 ± 0.06	0.17	0.15
$MgII\,2796$	2796.35	1.06 ± 0.12	1.08 ± 0.04	0.28	0.08
$MgII\,2803$	2803.53	0.85 ± 0.14	0.91 ± 0.05	0.27	0.09
$MgI\,2853$	2852.96	0.23 ± 0.04	0.25 ± 0.04	0.02	0.07

To visualize this in Figure 3.1 it can be seen the mean contribution C_i of each line to the DLAs in which that line is present (dark grey) and this value weighted by the fraction of the total DLAs F_i in which this line is used. It can be seen how for example *SiII-FeII1260* which has a high mean contribution, this is not an important line in the global of the population since it only can be used in a small fraction of all the DLAs (618 out of 33 939)
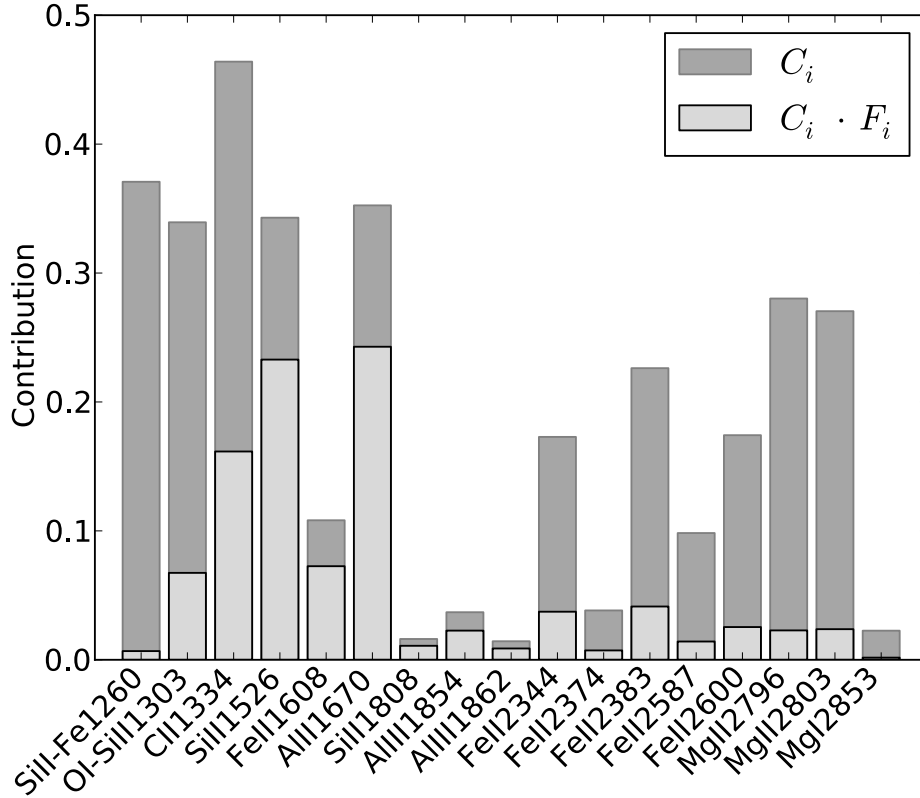

Figure 3.1: Dark grey, mean contribution of each line. On average to compute the metal strength S that line will represent the fraction shown here. Light grey, the contribution is weighted by the fraction of DLAs in which it appears, it is C_i multiplied by F_i, its sum is 1. This shows how a metal specie is more o less important in the global population.

It has to be noticed that case of OI-SiII1303 doublet is special, OI1302 and SiII1304 lie close enough for them to bend on the extremes, especially for high metallicity systems. Therefore the lines do not reach the continuum between them and therefore overlap. Then to avoid uncertainty when trying to split the 2 of them, it is better to measure both as a combined absorption feature. This will make this one of the strongest features when measured, since both OI1302 and SiII1304 are quite strong themselves and are more so combined. Unfortunately since these lie close to the Lyα line, most of the time these fall in the Lyα emitting region from the quasar, or in the Lyα forest if the DLA is far from the Quasar.

3.3.2 Redshifts

For each DLA we need an estimate of its redshift. It can be corrected using the metal lines associated with the DLA, but we have seen that the redshift estimated in DR12DLA is accurate for DR12, this means that the correction does not change the redshift by more than 1%, therefore DR12DLA redshift for the DLAs is used unless it is otherwise mentioned. During the work the metal lines in the Quasar spectrum are identified thanks to this redshift and the their laboratory wavelength. After that for each line the area around the expected position of the metal line should lie is put in restframe. All the rest of the operations will be made on this window on restframe, the size of this windows will be discussed latter §3.3.3. The redshift of the Quasar is also needed to determine the region where the analysis has to be done, only redwards from the Lyα emission. But it is not needed to be a precise estimation, as the area around the Lyα peak would not be used.

3.3.3 Windows for continuum fitting and W measurement

As mentioned in §3.3.2, a section of the spectra around each line is shifted to restframe. But how to select this section? For each line we will need a set of 2 windows centred around the expected position of the metal line.

A smaller *Measurement Window* around the line will be the one that should contain the whole absorbing feature, that means the pixels that have been affected by the absorption of that metal specie due to the width of the line. This window will be fixed for all the lines except for OI-SiII1303 where the two lines are measured together in the same window.   For this work we are assuming that all the lines for all the DLAs have similar width since their widths, or Doppler parameter value, is mainly due to the spectrograph resolution. therefore we will be using the same fixed window for all the lines of all the DLAs. We have check this by computing the Doppler parameter of a subsample containing the best resolved lines of the total population and seeing that it is indeed centred around the spectrograph resolution and the distribution is not wide.   The size of the *measurement window* is taken to be 10 pixels of 69km/s or 690km/s. With this size we have seen that the absorption feature sits well within the window and not many continuum pixels sit inside.

A bigger *Continuum Window* is set around the smaller measurement window, but the pixels of the this window are not used in the continuum one. This window will be used to set the value of the continuum around the line, as it is explained in §3.3.5. This is used to normalize the flux of the Quasar on the integration window. This window is also fixed for all the lines except when 2 lines would fall within the continuum window of eachother in order to avoid contamination on the continuum. In this case the continuum window is not centred in one line but around of both of the integration windows, having a common continuum window for both. This is the case for the pairs: $AlIII$ 1855-$AlIII$ 1863; $FeII$ 2374-$FeII$ 2383; $FeII$ 12587-$FeII$ 2600; $MgII$ 2796-$MgII$ 12804. The size is nevertheless fixed, only the where it is around and the distance to the line varies. The size of the *continuum window* is 50 pixels of 69km/s, 28 at each side of the integration window or 1 725km/s, with 1 pixel in between that is not used for neither.
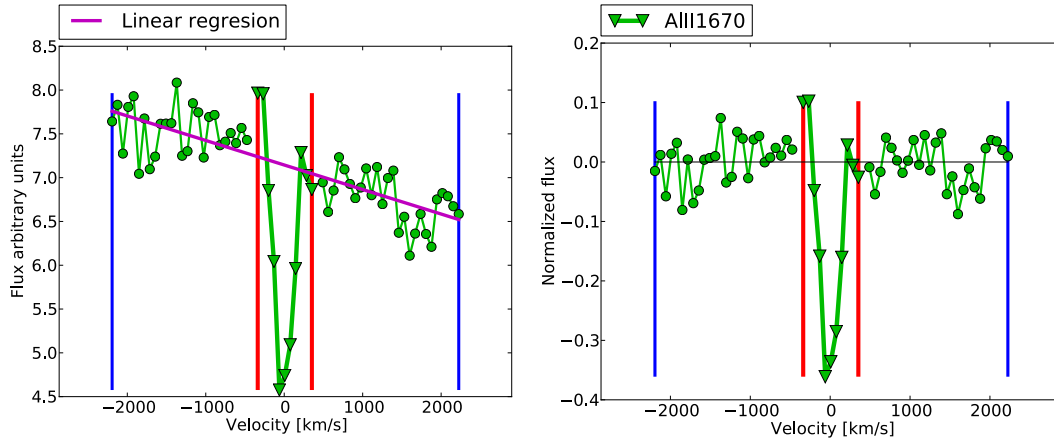
Figure 3.2: An example of the windows for a restframe flux in velocity, which is non normalized (left panel) and already normalized (right panel) around the absorbing metal line *AlII* 1670 for the DLA in Quasar spectrum 55182-3587-0100 from BOSS . The *Measurement Window* is what lies inside the red vertical lines, what lies outside corresponds and in between the blue vertical lines is used for to the *Continuum Window*. These pixels of the *Continuum Window* are fitted by a linear regression, this is shown in magenta in the right panel.

3.3.4 Selecting the Metal lines

Not all the 17 lines will be used for each DLA, only few of them at a time will have the conditions to be analysed . For example it is very unlikely that a line close to the Lyα from the DLA will fall on the red part of the continuum of the Quasar. That would mean that the DLA is almost on top of the Quasar. This happens for the SiII-Fe1260 line, this line is so close to the Lyα line that only 618 of them have been used 3.3.1. The criteria to select the lines for each DLA are that they have to fulfil the flowing set of conditions:

- The red end of the lines' continuum window has to lie redwards of the Lyα peak from the quasar. Given the Quasar redshift, the line must be at least $150 \cdot 69$km/s redwards of the Lyα peak as shown in Table 3.3.4. This is to avoid the line being in the Lyα and *NIV* 1240 emission, or directly in the forest, where the continuum is difficult to estimate.

- The red end of the lines' continuum window has to fall bluewards of the last pixel of the spectrograph. Given the highest wavelength measured the line must be at least $30 \cdot 69$km/s bluewards of the end of the spectra. This is to avoid the right continuum being cut when it is close to the end of the spectrum.

- The lines' integration window can not have any sky line in it. This is to avoid a sky line contaminating the equivalent width.This set of 872 skylines can be bound in Blomqvist (2014)

- The lines' continuum window can not have an emission line from the quasar within it or close to it. This is to avoid the emission feature upsetting the continuum estimation. The emission features to be avoided are those shown in Table 3.3.4, wich represent the strongest emission features of the red part of the spectra, as seen in Pâris et al. (2012)

and the separation s between absorbing and emitting line must be bigger than the one in the table, except for the Lyα case that is disused above.

Table 3.2: Strongest emission lines to be avoided in the red part of the continuum of the Quasar. The metal line ha to be at a separation bigger than s or at more than the number of pixels p from the emission line for the continuum to avoid being affected by the emission.

Name	λ	s	p
$Ly\alpha$-NV	1215.0 Å	10 350 km/s	150
$SiIIV$-OIV	1400.0 Å	4 029 km/s	61
CIV	1549.2 Å	4 029 km/s	61
$CIII$	1908.7 Å	4 029 km/s	61
$MgII$	2798.7 Å	4 029 km/s	61

Once the lines that meet these criterion are selected, these will be used through the rest of the analysis for that DLA. The average number of lines used for each DLA is 6 Figure 3.3, but not the 6 of them have to be strong one. For the ones that contribute the most (we have chosen the 8 that have $C > 0.2$) the most common is to have only a couple of these contributing, and in 2 818 cases non of the strong lines contribute. A few DLAs (427) do not have any lines computed.
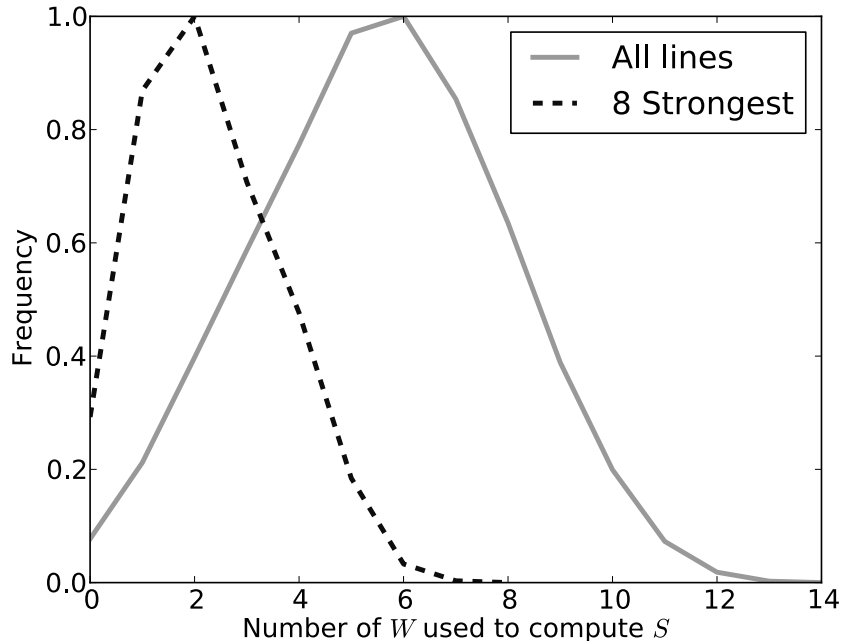

Figure 3.3: Relative amount of lines in which the equivalent width was computed for each DLA for the all 17 metal lines and for the 8 with contribution $C > 0.2$.

3.3.5 Continuum estimation

Once the lines and regions around them are selected and in restframe we need to normalize the region to compute the equivalent width of the line. In order to do that an effective and simple method is to compute the continuum around the line and normalize by it. The simplest estimator of this continuum is to assume that the continuum is flat in window used, therefore a linear regression can be fitted to the pixels on both sides of the *Continuum Window.* This simple example is shown in Figure 3.3.3 where the area around one absorbing metal line is shown before and after the normalization.

This nevertheless arises some problems that have to be tracked. A bad estimation of the continuum will result in big differences on the final equivalent width computed. Once the emission features of the quasar continuum are avoided as described in §3.3.3, the main problem to tackle is to avoid bad pixels or big outliers that will affect the estimation of the continuum. In order to deal with this first a linear fit weighted by the noise of each pixel is done. Then this is used to see if any pixel lies at more than 4σ from the first linear estimation. If that is the case this pixel is removed and the linear fit is done again. This is repeated until there are no more pixels at more than 4σ from the estimation, or when half of the pixels have been removed, in that case the equivalent width of that line is not computed due to bad continuum.

Once the linear fit is estimated, the flux pixels inside the integration window are normalized by it, and also the noise of these pixels. This can now be used to compute the equivalent width of each line and its uncertainty.

3.3.6 Measurement and error estimation of the W of the metal lines

There are different ways of computing the equivalent width. In this subsection we have decided for a method suited for the data we are dealing with, a low resolution and noisy one.

By simply summing all the pixels within the integration window and transforming them to Angstroms, $dx = 69km/s\lambda_M/C$ we have a raw estimate of the equivalent width. The associated error to this equivalent width is the sum of the noise of the pixels. This error underestimates the real uncertainty since the fitting of the continuum would add uncertainty here, but nevertheless this is a minimum estimation and it is sensitive to how noisy the region used is.

This method is unbiased as it is independent of any characteristics of line or the process used, but it gives the unrealistic solutions of negative equivalent width, this is due to some lines having effectively no absorbing signal, just noise. This noise can be either positive or negative therefore the sum over it can be either positive or negative. Actually the negative values with their errors should be and are compatible with 0 at 3σ. This means that positive values within 3σ of 0 are also likely to be non detection.

To test the method we have added mock lines in empty parts of the spectrum and we have successfully recovered their initial equivalent widths within the uncertainty estimated for spectra for different levels of noise. This can be further checked by computing the equivalent with in areas of the spectrum that we do not expect to be any absorbing line. If the result of the calculus is compatible with 0 then there is no bias on the measure.

3.4. METAL STRENGTH INDEX

Once we have a measurement for multiple metal lines for thousands of DLAs, we have produced a simple index of the matallicity to classify the DLAs according to their metal content. This *Metal Strength Index* uses the equivalent widths of different lines, allowing for better statistical power and larder number of DLAs with metal strength measured them if only one metal transition is used since not one line can be used for all DLAs. This combination also allows for reduced uncertainty or systematics, a combination of lines has less relative uncertainty that the measure of only one line.

The way to measure this metal strength S for each DLA is by computing:

$$S = \frac{\sum_i \left(\bar{W}_i/\epsilon_i\right)^2 \cdot \left(W_i/\bar{W}_i\right)}{\sum_i \left(\bar{W}_i/\epsilon_i\right)^2}, \tag{3.2}$$

where w_i is the equivalent width of each of the lines; ϵ_i is the error for the equivalent width of each line ; and $\bar{w}_i$ is the mean equivalent width computed for the whole population of DLAs from the BOSS catalogue, here both the staking value or the mean value seen in table 3.4.1 can be used. The staking from Mas-Ribas et al. (2014) is slightly more reliable since it has more statistics (uses all the DLAs) and its uncertainty is smaller. Nevertheless the 2 methods are compatible for all but the $Fe\,2374$ value. This metal strength combines the expected W from the global population, how each measured line is stronger or weaker with respect to that, and its contribution to the metal strength is weighted by this expected W and the uncertainty that each individual equivalent width for a metal. This way the lines that have stronger expected W and less relative noise contribute more to the Metal Strength parameter.

A representation of this *Metal Strength S* is shown in the blue curve of Figure 3.4.

And the Uncertainty for the metal strength is:

$$\epsilon_s = \left(\sum_i \left(\frac{\bar{W}_i}{\epsilon_i}\right)^2\right)^{-1/2}. \tag{3.3}$$

With this metal strength index we can build different groups of DLAs according to the metal content of the DLAs.

3.4.1 Correction for the dependence of W on column density of neutral Hydrogen

We have seen in Mas-Ribas et al. (2014) that there is a clear linear dependence of the equivalent width of the different species of metals with the logarithm of the neutral Hydrogen column density of the DLA to which they belong ($log(N_{HI})$). This can be seen in table A3 in Mas-Ribas et al. (2014) and also in our own analysis, but with less precision since the number of DLAs used for each subgroup is smaller. With this dependence in column density the *Metal Strength Index* can be further refined.

$$\bar{W}_i(N_{HI}) = a(log(N_{HI}) - 20) + b, \tag{3.4}$$

where a is the slope computed for the fit and b is the value of $W_i(\,log(N_{HI})=20)$. Now the equivalent width of each line is not just "weighted" by the mean of the population, but by what should be expected to be the mean of a given the logarithm of the column density. Therefore $\bar{W}_i \rightarrow \bar{W}_i(\,log(N_{HI}))$ and this linear dependence on the logarithm of the neutral Hydrogen column density is extracted from as mentioned from table A3 of Mas-Ribas et al. (2014).

In Table 3.4.1 the fitted parameters of the linear regression are shown.

Table 3.3: Values for the linear regression fitted to the evolution of the staked W for different intervals of N_{HI}, from 3.4.

Name	a	b
$SiII$-Fe 1260	0.41	0.34
OI-$SiII$ 1303	0.67	0.35
CII 1334	0.40	0.31
$SiII$ 1526	0.41	0.18
$FeII$ 1608	0.38	0.21
$AlII$ 1670	0.12	0.01
$SiII$ 1808	0.11	0.05
$AlIII$ 1854	0.05	0.03
$AlIII$ 1862	0.31	0.04
$FeII$ 2344	0.51	0.19
$FeII$ 2374	0.35	0.05
$FeII$ 2383	0.57	0.28
$FeII$ 2587	0.59	0.13
$FeII$ 2600	0.63	0.30
$MgII$ 2796	0.20	0.10
$MgII$ 2803	0.83	0.52
MgI 2853	0.88	0.50

In 3.4 it is shown that the effect o the correction on the total distribution is not noticeable. This is because the corrections at bigger and lower H_{NI} average out since the effects are contrary as seen in Figure 3.4.1 .
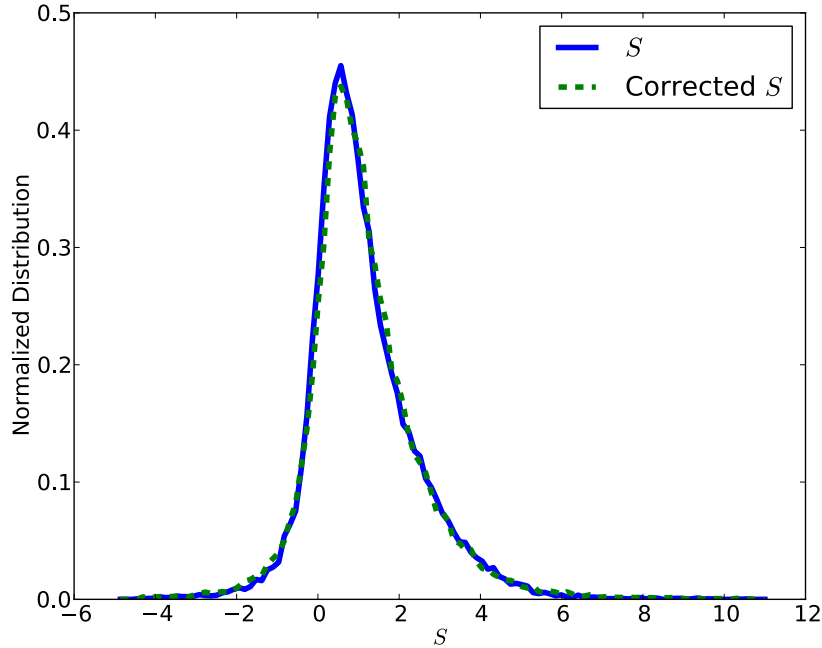

Figure 3.4: Dark grey, mean contribution of each line. On average to compute the Strength S that line will represent the fraction shown here. Light grey, the contribution is weighted by the fraction of DLAs in which it appears, basically it is C multiplied by F, its sum is 1. This shows how a metal specie is more o less important on the global population.

In 3.4.1 the sample is split depending on its $log(N_{HI})$ for the lower range ($log(N_{HI}) < 20.25$) shown in the upper panel it can clearly be seen how the correction affects deeply the population, the same is true for the lower panel. Using this the populations can be better defined. If $log(N_{HI})$ for a DLA is bigger than the mean then that translates in a smaller S since it is now "weighted" by a bigger $\bar{w}$ and the dispersion on S is smaller than when not considering the correction. The opposite happens when DLAs with smaller column density than the mean are considered.
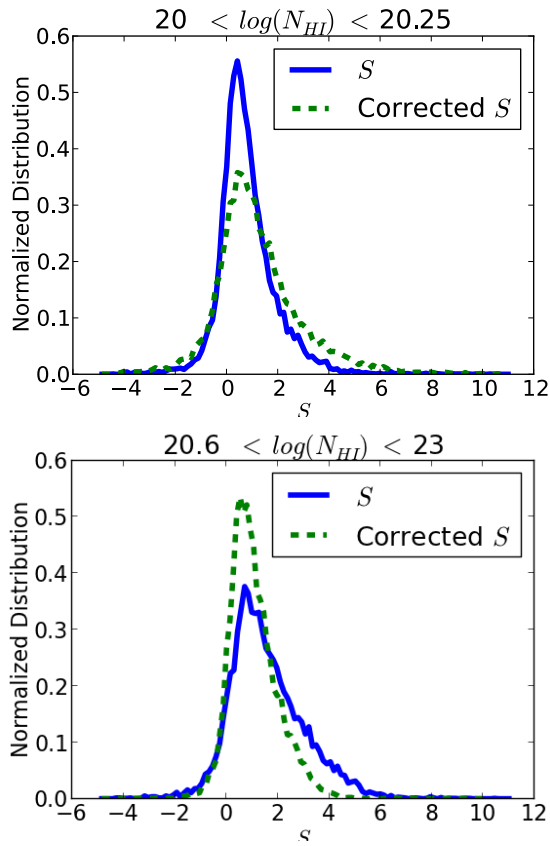
Figure 3.5: Comparison of the distribution of metal strength S when corrected by the dependence on the DLA column density $log(N_{HI})$ **Upper panel** shows the difference in distribution for the low density end $log(N_{HI}) < 20.25$ and **lower panel** shows the same for the high density end $log(N_{HI} > 20.6$

3.4.2 Splitting the DLAs into Metal Strength Populations

Once we have this *Metal Strength Index S* we can classify our population of DLAs according to it. A good way to start is to divide the population in 3 groups. To avoid mixing of objects belonging to different groups, only DLAs with a $err_s < 0.5$ are used, out of the 33 939, 22 849 DLAs still follow that requisite. With that 3 groups are made with roughly 6500 to 9000 objects in each. The 3 groups are:

- $S \geq 0.5$, Low metallicity group, with 7308 objects.

- $0.5 < S \geq 1.5$, Medium metallicity group, with 9010 objects.

- $1.5 < S$, High metallicity group, with 6531 objects.

3.5. RESULTS

Here we present the main results of the work, the elavoration of a a cataloge with the new way of measuring the equivalent widths and the metal strength for each DLA §3.5.1, an analysis of the redshift evolution of the metal strength parameter and a staking of the DLAs acording to different populations selected depending on their metal strength.

3.5.1 Catalogue of the metal lines

Finally with the results we have produced a list containing all the equivalent widths measured in the raw way. Each DLA contains the values for the column density, redshifts of DLA and Quasar, signal to noise, name of the Quasar spectra in Julian Fiver Plate units. And then the equivalent widths for all the 17 lines, with their error and a flag detecting strange features. When the value of the equivalent width is 0 that means it has not been measured. In addition to that the 2 final columns are the Metal Strength parameter measured as described in 3.4.1 and its uncertainty for each DLA.

This list differs from DR12DLA one in that the way of measuring the equivalent width is different and that more lines are considered. In DR12DLA the line is fitted by a Gaussian or Void profile and the equivalent width is measured from that. We have checked that this method overestimates the equivalent widths since it only integrates over the values that are below the continuum, therefore it is biased because the noise that tends to decrease the continuum flux is considered for the fit, but the one that increases it, if it is putted above the estimated continuum, is not used. This translates in an overestimation of the equivalent with by about $15 - 20\%$ on average.

Table 3.4: Sample of the catalogue, with some of the DLAs information: *Redshift of the Quasar, Julian-Fiver-Plate from BOSS, DLA redshift and Column density* as measured by Noterdaeme et al. (2009), then the measured W and its uncertainty for the 16 of the 17 lines (MgI 2853 lacks due to the format of the table.). Other spectrum characteristics are included in the public catalogue but are not shown here for brevity, these are: *thingID, flag Right Ascension, Declination, Continuum to Noise Ratio, and Balnicity index.*

Qz	JFP	DLAz	$log(N_{HI})$	SiII1260	eSiII1260
2.72	56604-7167-0290	2.49	21.09	0	0
2.48	56265-6151-0936	2.37	20.02	0	0
2.78	56190-6182-0338	2.34	21.22	0	0

OI-SiII1303	eOI-SiIIi1303	CII1334	eCII1334	SiII1526	eSiII1526
0	0	0	0	0.66	0.5
0.17	0.15	0.11	0.14	-0.04	0.14
0	0	0	0	1.35	0.1

FeII1608	eFeII1608	AlII1670	eAlII1670	SiII1808	SiII1808
-0.01	0.56	-0.26	0.47	0	0
0	0	0.04	0.15	-0.06	0.19
0.39	0.12	0	0	0.32	0.16

AlIII1854	eAlIII1854	AlIII1862	eAlIII1862	FeII2344	eFeII2344
-0.44	0.88	0	0	-0.74	1.7
0.29	0.2	-0.01	0.21	0	0
0.45	0.15	0	0	0	0

FeII2374	eFeII2374	FeII2382	eFeII2382	FeII2587	eFeII2587
0	0	0	0	0	0
0	0	0	0	0.15	0.24
0	0	0	0	0	0

FeII2600	eFeII2600	MgII2796	eMgII2796	MgII2803	eMgII2803
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

3.5.2 Redshift evolution of the Metal Strength

Once we have developed the metal strength as a measurement of the intrinsic properties of a DLA, driven by its velocity dispersion and metal content, then we can easily use this new parameter to study its evolution under other properties from the DLAs, thanks again to the large number of objects that we have at our disposition. One evolution that is very direct and pursued in the literature is the study of the evolution of the metallicity with redshift Rafelski et al. (2012), Kulkarni et al. (2005). Although in these studies the metallicity is directly measured, we can do a similar exercise with the evolution of our metal strength S with the redshift, we remind here that the metal strength connives a measurement of the velocity dispersion and metal content of the DLA. In Figure 3.5.2 we can see how the distribution of S clearly tends to be predominantly around lower values of S, this is more evident for redshift higher than 3. In Figure **??**, lower panel, we can see this trend with redshift more easily, here the mean S and its standard deviation is plotted for each redshift interval, we clearly see how for high redsift, 5 to 3.5, the Metal Strength increases sharply and then flattens for redshift 3 to 2, therefore this is not a linear monotonic evolution but it depends on the epoch of the universe having a seemingly sharp rise in redshift 4 to 3, probably tracing some phases of major structure and star formation. It has to be noted that a mean vale of $S = 1$ is what should be expected for the whole population, therefore mean values above 1 imply stronger than average and below 1 smaller than average metal strength.
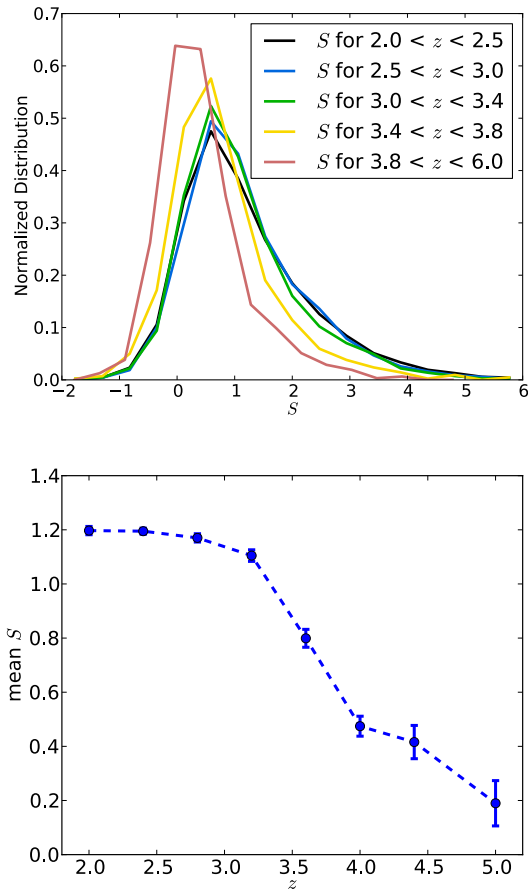
Figure 3.6: Evolution of the Metal Strength S with redshift. **Upper panel** shows the distribution of the metal strength for various cuts in redshift space. **Lower panel** shows the mean S and its mean its standar deviation for several redshift intervals. At higher redshift the errors grow as there are fewer DLAs in these bins that at lower redshift, the last bin is twice as big as the others and ranges from 5.4 to 4.6.

3.5.3 Stacked spectra for different Metallicity Estimation Populations

Thanks that we have a large number of observed DLAs, this allows one to examine the average metal-line absorption by averaging over many systems, and studying the dependence of the equivalent widths of any line with the Metal Strength parameter, presented in §3.2. This has the advantage of providing directly properties of the average population of DLAs, rather than individual systems which have a large intrinsic random variation and are more affected by the individual noise, are actually probing a mixture of different gas phases. Moreover, absorption lines that are located in the Lyα forest region can also be accurately measured after averaging over a large number of absorption systems while it is not possible to extract any useful information since the metal absorption lines are mixed with the Lyα lines from the forest.

With 3 different populations of DLAs selected by their metal strength we can do a different analysis. One thing that can be done is to see the evolution of the characteristics of each population with this metal strength. Since each DLA has low resolution and high signal to noise, a good solution to estimate the mean characteristics of each population is to build a synthetic model representative of the population. Traditionally this model has been constructed by stacking many different spectra that share similar features, this way these features (the DLA is our case) are highlighted while all the noise is suppressed. Here we can do the same for these 3 populations of objects, build a synthetic mean spectrum of each population made by stacking all of the objects belonging to each group. With that we can latter study the evolution of the main features with respect to the Metal Strength.

From the stack seen in Figure 3.5.3 we can extract many interesting points. Firstly and most obvious is the clear evolution of the equivalent width of all the absorption lines with increasing Metal Strength, black is the interval corresponding to the smallest S range (S up to 0.5), here many weak lines can not be discern, and even the strong ones have really small equivalent widths. The intermediate range in blue ($0.5 < S <, 1.5$) and the upper one in green (S bigger than 1.5) show a clear growth of the equivalent width of all the lines, weak and strong, ad also in the ones that usually lie in the Lyα forest and that are not used in this study. Also interestingly not only for the low ionization lines but also for the high ionization lines like CIV or NV we can see the same trend. All these trends in addition give strength to the method and the usefulness of the Metal Strength as a prove and important characteristic of the DLAs, since selecting only few equivalent width of each DLA that would be saturated in mots of the cases, we clearly see how all the rest of characteristics that this parameter aims to represent follow the same trend even though were not used a priory. This was in any case expected, as many studies show the correlation between all the metal features of the spectrum, nevertheless the fact that the correlation is seen here is non minor feat.

Then with the stacking we can also do more detailed studied. For example we can study the evolution for different kinds of lines, these that are strong and therefore are saturated for most of the cases, then they are only sensitive to the velocity dispersion (e.g. $SiII$ 1806, $AlII$ 1670, $FeII$ 1606), but not an increase in metallicity. On the other hand weak lines like $NiII$ 1455 usually wont be saturated and would not increase their equivalent width if the velocity dispersion increases, but only if their metallicity does. Then we can see using these tow polulations of lines how each of the characteristics that the Metal Strength traces evolves with it. Finally as mentioned we can do the same study as done with low ionization metals but for high ones. This lead to the position in which with a refined and thoughtfully

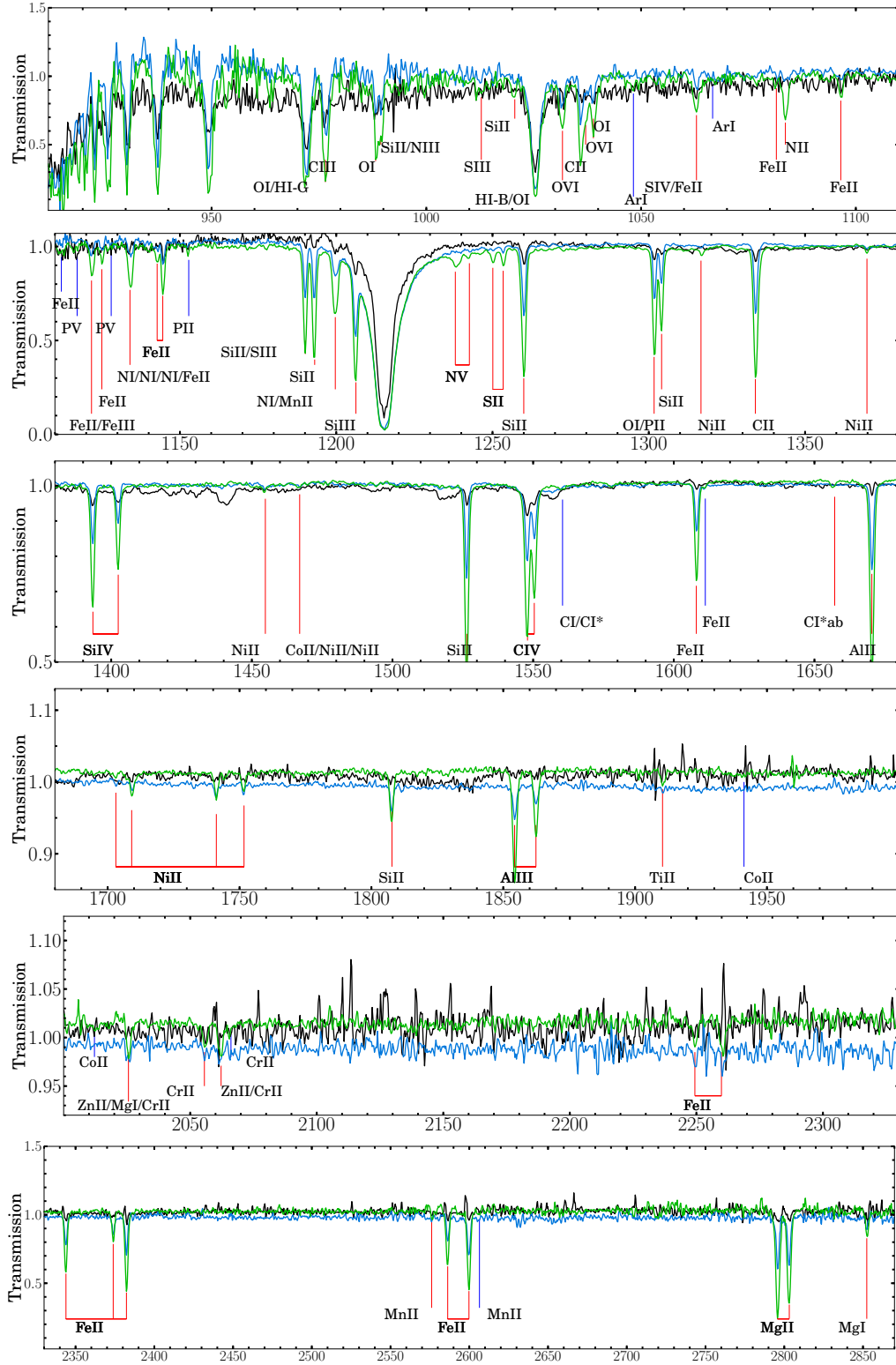**Figure 3.7**: Stack for 3 different populations, black is for $S < 0.5$, blue for $0.5 < S < 1.5$, and green for $1.5 < S$. The 11 lines marked in blue could be detected in even higher $S > 2.5$ as seen in 8.

tested method we can modelise this trends in metallicity and velocity dispersion with Metal Strength.

In the same way, if we just center in the higher order Lyman series (β, γ, δ, ϵ, ...), they are obviously saturated and therefore we just study the velocity dispersion. As we can clearly see in the upper panel of 3.5.3 the evolution with S is evident, therefore by selecting a higher or smaller Metal Strength we are effectively selecting different populations of DLAs with much larger or weaker velocity dispersion independently of column density of the Lyα because we already corrected by that.

And then we can let us enjoy the method and go to extreme cases, like selecting the DLAs with most Metal Strength to see what hidden features we can observe, thanks to that we have been able to pinpoint metal lines never observed before in spectra like BOSS, a total of 11 new transitions can be seen in Figure 8 when selecting a metal strength above 2.5 ($S > 2.5$). These lines mark with blue in 3.5.3 are new for this work, and some like PV have never been realibly measured before because they lie in the Lyα forest of the quasar. This points to a wealth of results can be achieved with this method and very interesting discoveries await as it is refined and the amount of available data increases over time.

3.6. Conclusions

During this work we developed a new and simple and easily exportable method to measure the equivalent width W of the 17 strongest low ionization metal transitions for the DLAs absorbing spectrum between $1260\mathring{A}$ and $3000\mathring{A}$ at restframe. With this long sentence we want to transmit that for the first time we have the tools that allow to compute better than ever an equivalent width for tens of thousands of individual lines belonging to thousands of DLAs form the database in DR12DLA from the BOSS survey DR12, large numbers of measurements is pursued as a main result of the work. We demonstrated a new and simpler method to estimate the equivalent width for low resolution spectra, this allowed to achieve a more complete and representative sample of a bigger than ever number of metal transitions. With this we have produced a new large catalogue with more W measured for more low ionization metal transitions than ever for this kind of data, and we achieved measurements for up to $33\,939$ DLAs.

Nevertheless we notice that the reliability of this measurements is not sufficient in itself to be a major breakthrough since the uncertainty in most of the newly computed equivalent width is intrinsically is high.

In other to compensate for this, we can take advantage of the large catalogue of equivalent widths and the fact that for each DLA we have measured several metal transitions (an average of 6, with 2 being strong ones), combined with the mean expected equivalent width for different column densities of the DLA ($\bar{W}(log(N_{H_I}))$). We created a new parameter, the *Metal Strength* which supposes way of classifying the DLAs by the metallicity, saturation and velocity dispersion of the metals associated to each DLA. This parameter is better than a direct measurement of the metallicity from low ionzation because it is better suited to be applied to this kind of data, a large dataset with thousands of low resolution and noisy spectra. This allows for unprecedented high numbers of objects to be analysed depending on this new parameter, the Metal Strength.

We have proceeded with several direct studies once we have the measurement of the equivalent widths and a way to group them depending on this new parameter. First we have build 3 populations of *low*, *medium* and *high* Metal Strength. To highlight the faint features and present in each population we have stacked the spectra corresponding these 3 different populations, this allows to reveal really faint meta lines, some never seen before. Also the metal evolution is very clear for all the lines, low and high ionization on the stack, this does not only shows how the mean absorptions features change with Metal strength, but also validates the method, as they evolved in the way that was expected. This also allows to compute the elution of the strongest lines which are likely to be saturated and allow to estimate the increase of velocity dispersion with Metal Strength, and the weak lines, which should not be saturated and therefore directly trace the growth in metallicity.

Then we studied the evolution of the Metal Strength with redshift, although this is not a direct evolution of metallicity with redshift as has been pursued in many studies, we have seen a clear and strong evolution of our parameter, which is also non monolithic, it flattens at redshift 3 to 2, has a strong growth at redshifts 4 to 3, and seems tat flattens again at redshift higher than 5, but here we are prudent on that since the statistics are low. This has direct implications on star, galaxy and structure formation trough the time, and could be used in the future to model and test simulations that aim to measure the metallicity and velocity dispersion history of the universe.

Finally we expect an evolution of the bias of the DLAs with respect the matter spectrum, depending on the Metal Strength, since the dispersion of velocities and high metallicity is associated with bigger halos, therefore the bigger the Metal Strength, the bigger the bias should be. Adding another tool to prove the cosmology.

In conclusion this work provides with a new, easy and reliable method to extract properties of a common and characteristic object of the universe, the DLAs. It constitutes a preliminary study of this new method and it can be further refined and polished to extract more useful data and more and precise conclusions, but we have shown that it delivers many interesting results and it promises to be a tool that can be used for many and diverse applications. The strength of this obviously will grow as the many surveys that are in the pipeline increase the number of objects that can be used to increase our population and be able to have more subdivisions and more fine testing.

A.1. High Metal Strength subsample stacking

Here we present the taking of the high Metal Strength subsample ($S > 2.5$) where 11 new weak lines are tantalizingly detected, those marked in blue as seen in Figure 8. Lines in the Lyα forest have never been detected before in individual spectra as they can be confused with the forest. The lines marked in red where detected in Mas-Ribas et al. (2014) metallic sample.
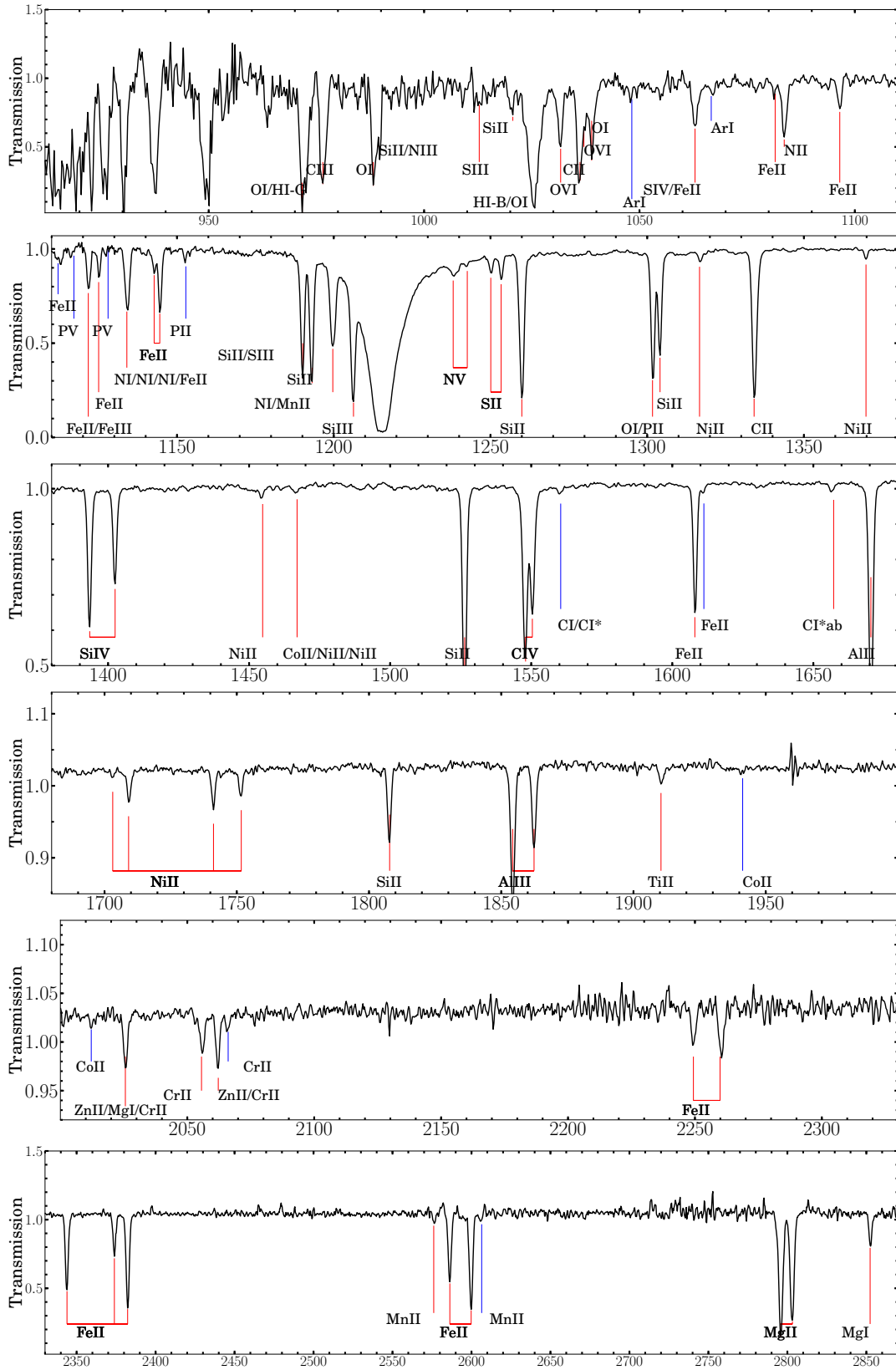
Figure 8: Stack for the high Metal Strength subsample ($S > 2.5$). The lines marked in red where detected in Mas-Ribas et al. (2014) metallic sample, the 11 lines marked in blue are detected only here.

Bibliography

Bhanot, G. 1988, Reports on Progress in Physics, 51, 429

Blomqvist, M. 2014, Sky Mask, SDSSIII-DR12, `https://trac.sdss3.org/wiki/BOSS/LyaForestsurvey/SkyMask`, accessed:2014-07-02

Bolton, J. S., Viel, M., Kim, T.-S., Haehnelt, M. G., & Carswell, R. F. 2008, MNRAS, 386, 1131

Busca, N. G., Delubac, T., Rich, J., et al. 2013, A&A, 552, A96

Cen, R. 1992, ApJS, 78, 341

—. 2002, ApJS, 141, 211

—. 2010, Formation and Evolution of Galaxies and the Intergalactic Medium, nASA Proposal #10-ATP10-73

Cen, R., Miralda-Escudé, J., Ostriker, J. P., & Rauch, M. 1994, ApJL, 437, L9

Cen, R., & Ostriker, J. P. 1993, ApJ, 417, 404

Cen, R. Y., Ostriker, J. P., Jameson, A., & Liu, F. 1990, ApJL, 362, L41

Compostella, M., Cantalupo, S., & Porciani, C. 2013, MNRAS, 435, 3169

Croft, R. A. C., Weinberg, D. H., Bolte, M., et al. 2002, ApJ, 581, 20

Croft, R. A. C., Weinberg, D. H., Katz, N., & Hernquist, L. 1998, ApJ, 495, 44

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013a, AJ, 145, 10

—. 2013b, AJ, 145, 10

Delubac, T., Bautista, J. E., Busca, N. G., et al. 2014, ArXiv e-prints, arXiv:1404.1801

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011a, AJ, 142, 72

—. 2011b, AJ, 142, 72

Font-Ribera, A., Miralda-Escudé, J., Arnau, E., et al. 2012, JCAP, 11, 59

Font-Ribera, A., Kirkby, D., Busca, N., et al. 2014, JCAP, 5, 27

Gnedin, N. Y., & Hui, L. 1998, MNRAS, 296, 44

Haehnelt, M. G., Steinmetz, M., & Rauch, M. 1998, ApJ, 495, 647

Hamilton, T. 1992, in Bulletin of the American Astronomical Society, Vol. 24, American Astronomical Society Meeting Abstracts, 1248

Hasselfield, M., Hilton, M., Marriage, T. A., et al. 2013, JCAP, 7, 8

Hernquist, L., Katz, N., Weinberg, D. H., & Miralda-Escudé, J. 1996, ApJL, 457, L51

Hui, L., & Gnedin, N. Y. 1997, MNRAS, 292, 27

Kaiser, N. 1987, MNRAS, 227, 1

Katz, N., Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19

Kim, T.-S., Bolton, J. S., Viel, M., Haehnelt, M. G., & Carswell, R. F. 2007, MNRAS, 382, 1657

Kollmeier, J. A., Miralda-Escudé, J., Cen, R., & Ostriker, J. P. 2006, ApJ, 638, 52

Kulkarni, V. P., Fall, S. M., Lauroesch, J. T., et al. 2005, ApJ, 618, 68

Mas-Ribas, L., Miralda-Escudé, J., Noterdaeme, P., et al. 2014, A&A, 000, 0000

McDonald, P. 2003, ApJ, 585, 34

McDonald, P., & Eisenstein, D. J. 2007, PRD, 76, 063009

McDonald, P., & Miralda-Escudé, J. 1999, ApJ, 519, 486

McDonald, P., Miralda-Escudé, J., Rauch, M., et al. 2000, ApJ, 543, 1

McDonald, P., Seljak, U., Burles, S., et al. 2006, ApJS, 163, 80

McQuinn, M., Lidz, A., Zaldarriaga, M., et al. 2009, ApJ, 694, 842

Miralda-Escudé, J., Cen, R., Ostriker, J. P., & Rauch, M. 1996, ApJ, 471, 582

Noterdaeme, P., Petitjean, P., Ledoux, C., & Srianand, R. 2009, A&A, 505, 1087

Noterdaeme, P., Petitjean, P., Carithers, W. C., et al. 2012, A&A, 547, L1

Pâris, I., Petitjean, P., Aubourg, É., et al. 2012, A&A, 548, A66

—. 2014, A&A, 563, A54

Prochaska, J. X., & Wolfe, A. M. 1997, ApJ, 487, 73

Rafelski, M., Wolfe, A. M., Prochaska, J. X., Neeleman, M., & Mendez, A. J. 2012, ApJ, 755, 89

Slosar, A., Font-Ribera, A., Pieri, M. M., et al. 2011, JCAP, 9, 1

Slosar, A., Iršič, V., Kirkby, D., et al. 2013, JCAP, 4, 26

Springel, V. 2005, MNRAS, 364, 1105

Theuns, T., Leonard, A., Efstathiou, G., Pearce, F. R., & Thomas, P. A. 1998, MNRAS, 301, 478

Viel, M., Haehnelt, M. G., & Springel, V. 2004, MNRAS, 354, 684

Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, ARAA, 43, 861

Zhang, Y., Anninos, P., & Norman, M. L. 1995, in Bulletin of the American Astronomical Society, Vol. 27, American Astronomical Society Meeting Abstracts, 1412