Carnegie Mellon University
MELLON COLLEGE OF SCIENCE

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN THE FIELD OF PHYSICS

TITLE: "Large scale 3D mapping of the intergalactic medium with the Lyman-alpha forest using hydrodynamic simulations and SDSS–III BOSS DR12"

PRESENTED BY: Melih Ozbek

ACCEPTED BY THE DEPARTMENT OF PHYSICS

RUPERT CROFT 4/27/17
RUPERT CROFT, CHAIR PROFESSOR DATE

STEPHEN GAROFF 4/27/17
STEPHEN GAROFF, DEPT HEAD DATE

APPROVED BY THE COLLEGE COUNCIL

REBECCA DOERGE 4/27/17
REBECCA DOERGE, DEAN DATE

Large-scale 3D mapping of the intergalactic medium with the Lyman alpha forest using hydrodynamic simulations and SDSS–III BOSS DR12

by

Melih Ozbek

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
at
Carnegie Mellon University
Department of Physics
Pittsburgh, Pennsylvania

Advised by Professor Rupert Croft

April 16, 2017

Abstract

Maps of the large-scale structure of the Universe at redshifts $2 < z < 4$ can be made with the Lyα forest, which are complementary to low-redshift galaxy surveys. We apply the Wiener interpolation method of Caucci et al. to construct three-dimensional maps from sets of Lyman α forest spectra, evaluating its performance with cosmological hydrodynamic simulations and also applying these methods to BOSS DR12 data. We discuss local smoothing and also show that the simulated map resolutions are accessible to current observational surveys. We find from the simulation study that both the density field and the statistical properties of the IGM are recovered well enough that the resulting IGM maps can be meaningfully considered to represent large-scale maps of the universe, in agreement with Caucci et al., on larger scales and for sparser sightlines than had been tested previously. Quantitatively, for sightline parameters comparable to current and near future surveys, the correlation coefficient between true and reconstructed fields is $r > 0.9$ on scales $> 30\,h^{-1}$Mpc. The properties of the maps are relatively insensitive to the precise form of the covariance matrix used. The final BOSS quasar Lyman α forest sample allows us to make maps with a resolution of $\sim 60\,h^{-1}$Mpc over a volume of $\sim 15\,h^{-3}Gpc^3$ between redshifts 1.9 and 2.3. We make large maps of the BOSS DR12 IGM field and study global statistical properties of the maps with auto–correlation, slice plots, local peaks, probability density functions, point-by-point scatter, and percolation techniques to identify individual structures.

Acknowledgments

This thesis is dedicated to my dear parents. My mother, the most wonderful and loving person I have ever known, thank you for your constant encouragement and support. And my father, who was a cosmologist at heart, thank you for introducing me to the world of physics at a young age, you have been my inspiration my entire life.

I would like to thank my advisor Rupert Croft for his invaluable guidance during my PhD. Thanks to my office mates, colleagues, and other faculty at Carnegie Mellon for useful discussions.

Finally, I would like to thank my thesis committee, Hy Trac, Rachel Mandelbaum and Carles Badenes for their suggestions and helpful ideas.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Cosmology

Throughout the history of mankind, there have been two main stages to unfolding the mysteries of nature: Obtaining data and interpreting it correctly. Cosmology, the study of the origin and structural evolution of the Universe, is unique in that regard, as the data have always been readily available. Tycho Brahe proved, before telescopes were invented, that the Sun is much more distant than the Moon from the Earth, by measuring the position of stars and planets. Johannes Kepler, a student of Brahe, improved on his mentor's work by using elliptical planetary orbits instead of circular ones and formulating the laws of planetary motion. Isaac Newton expanded Galileo's ideas, describing the motion of celestial objects within a more rigorous scientific framework in late 17th century. In his cosmological model, space had Euclidean geometry and was infinite in all four dimensions, however, it was neither expanding nor contracting.

In early 20th century, the Newtonian static Universe was still the widely accepted model, which Einstein used for his work on general relativity. In his work between 1914–1917, he mentions various limitations of the Newtonian theory (e.g. the inconsistencies associated with the Newtonian boundary conditions of the Universe: "the field equations of gravitation which I have championed hitherto still need a slight modification, so that on the basis of the general theory of relativity those fundamental difficulties may be avoided ... as confronting the Newtonian theory." (zur allgemeinen Relativitätstheorie, 1917). Hence, although not initially present, Einstein also included a cosmological constant (Λ) in his field equations in order to balance out the attractive nature of gravity and hence allow for a static Universe:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu} \tag{1.1}$$

where $G_{\mu\nu}$ is the Einstein tensor, $g_{\mu\nu}$ is the metric tensor, $T_{\mu\nu}$ is the energy–momentum tensor, and $G = c = 1$. In principle, it describes the relation between the geometry of spacetime and the distribution and movement of energy and matter.

Edwin Hubble's discovery that the Universe is expanding (Hubble, 1929) suggested that the cosmological constant may be unnecessary: An expanding universe indicated by general relativity was seemingly correct. Although Einstein subsequently regretted introducing the idea of the cosmological constant and discarded it from his field equations, it was eventually discovered that Λ is small, but not zero: Examining Type Ia supernovae, two independent groups (Riess et al., 1998; Perlmutter et al., 1999) came to the conclusion that the expansion of the Universe is accelerating (for a review on the discovery of dark energy (DE) and the accelerating Universe, see (Frieman et al., 2008)). Since gravity alone could only slow down the expansion of the Universe, this marked the birth of new physics and subsequently dark energy, associated with negative pressure. The origins of dark energy still remain a mystery, however, independent observations from the cosmic microwave background (CMB), galaxies and the large–scale structure (LSS) support its existence. Furthermore, studying the motions of galaxies near the edge within the Coma cluster, the existence of another type of matter was proven (Zwicky, 1933), which only interacts gravitationally. Full–mission Planck observations (Ade et al., 2015, 2016) yield current relative density values as $\Omega_\Lambda = 0.68$ for dark energy, $\Omega_c = 0.27$ for dark matter (DM) and $\Omega_m = 0.05$ for baryonic matter (Figure 1.1).

The standard model of cosmology (Lambda Cold Dark Matter Model, or ΛCDM) provides a sufficiently good explanation for the accelerating expansion of the Universe, as well as the cosmic microwave background, big bang nucleosynthesis (light element abundances) and the large–scale structure. Temperature, the cosmic scale factor a or the redshift z are the three commonly used means of measuring time since the Big Bang:

$$a = \frac{1}{1+z} = \frac{\lambda_{emit}}{\lambda_{obs}} \tag{1.2}$$

where z is the redshift, also defined with the ratio of the emitted and elongated wavelengths of an electromagnetic wave as it propagates with the expanding Universe, e.g. a redshift of 2 means that electromagnetic waves had only a third of their wavelengths compared to what is observed today, or equivalently, the Universe was a third of its current size. The value of the dimensionless scale factor a is defined as unity for today.

This thesis is about studying the large–scale structure at very big scales, but first we need to understand how the different cosmic eras throughout the history of the Universe are interconnected before the LSS started evolving. While the ΛCDM cosmology can answer many aspects that have to do with the history of the Universe, there are still many mysteries that we cannot fully solve.

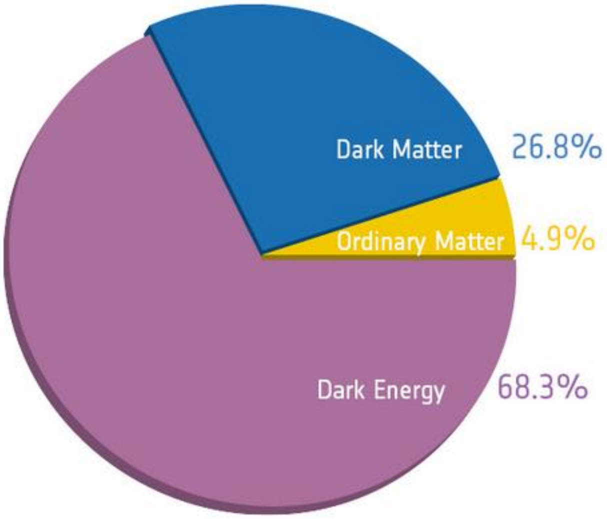


Figure 1.1: Relative densities of dark energy, dark matter and ordinary matter according to Planck observations (figure obtained from ESA / Planck).

1.1.1 Inflation

Inflation refers to the very short period at the beginning of the Big Bang, during which the Universe grew at a tremendous rate, to 10^{30} of its original size, within 10^{-32} seconds. The idea of inflation is that the Universe was dominated by vacuum energy with an equation of state $p < -\rho/3$. In view of the scaling of the radiation energy density ρ_r with the inverse fourth power of the scale factor, the condition required by general relativity, $\rho_\nu > \rho_r$, is trivially satisfied.

In principle, it is a viable method examine findings about the later Universe and go back in time to study earlier eras of the Universe, those that are closer in time to the Big Bang. This is how the inflation theory was introduced (Guth, 1981; Guth and Pi, 1982; Linde, 1982; Albrecht and Steinhardt, 1982; Hawking, 1982), in order to explain the fact that the anisotropies seen in the CMB are small (about one in 10^5) within regions that are not necessarily in causal contact, commonly known as the horizon problem (Figure 1.2). Attempts to enlarge the cone within an inflationary solution seem to suggest that inflation should be driven not by ordinary matter or radiation, but negative pressure, which means a form of dark energy (for a discussion, see Dodelson (2003)). According to Einstein's relativity, regions within the same light cone are connected causally, in other words, events forward in time (upwards in the light cone) can only be affected by events located below it, within the cone. The comoving horizon of the cone, η, is related to the cosmic scale factor a, which explains the proper distance between objects moving with the Hubble flow.

The flatness problem is yet another issue that the ΛCDM model fails to fully solve.

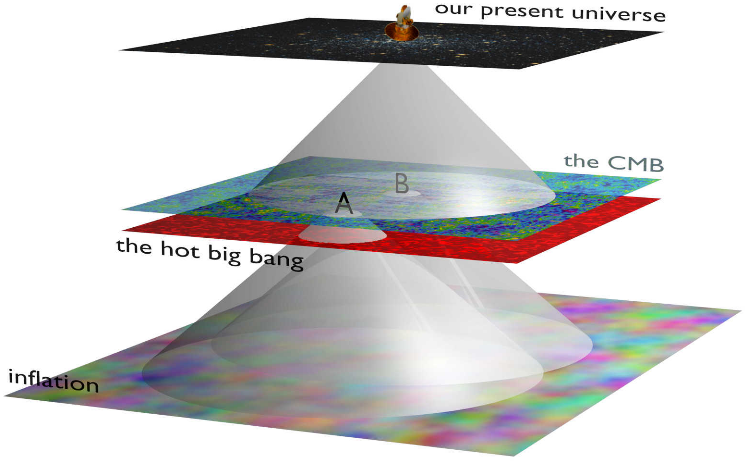


Figure 1.2: Horizon problem: causality connections between different epochs of the Universe (figure obtained from Yi (2014)).

After the introduction of the inflation theory, it was realized that it could also explain primordial density fluctuations. The spatial curvature of the Universe is very nearly zero, suggested by the CMB anisotropy and type Ia supernova observations. This is not necessarily preferred by Friedmann equations, and inflation tries to explain this phenomenon also (Berera et al., 1999). Inflation predicts a power law for the primordial power spectrum, describing the different length scale behavior:

$$P(s) \propto k^{n_s - 1} \tag{1.3}$$

The Planck survey gives the spectral index of curvature perturbations to be $n_S = 0.968 \pm 0.006$, meaning a nearly scale–invariant power spectrum. The reader is referred to (Yi, 2014) for a review of different inflation models.

1.1.2 Cosmic Microwave Background

Shortly after the Big Bang, the Universe was a hot and dense ionized opaque soup of baryons, coupled with photons. When the Universe cooled down sufficiently, Thomson scattering stopped and neutral hydrogen atoms were formed. The photons decoupled from baryons, their mean free path became effectively infinite, thus the Universe became transparent for the first time. These first free photons make up the cosmic microwave background (CMB). Tracing them back, the last time the CMB photons interacted with baryonic matter was the last scattering surface, during the recombination epoch. This was at a redshift of about $z = 1100$, when the scale factor a had a value of about 10^{-3}. Since a lot of information is lost due to scattering, it is not possible to use CMB photons directly to study the Universe prior to recombination.

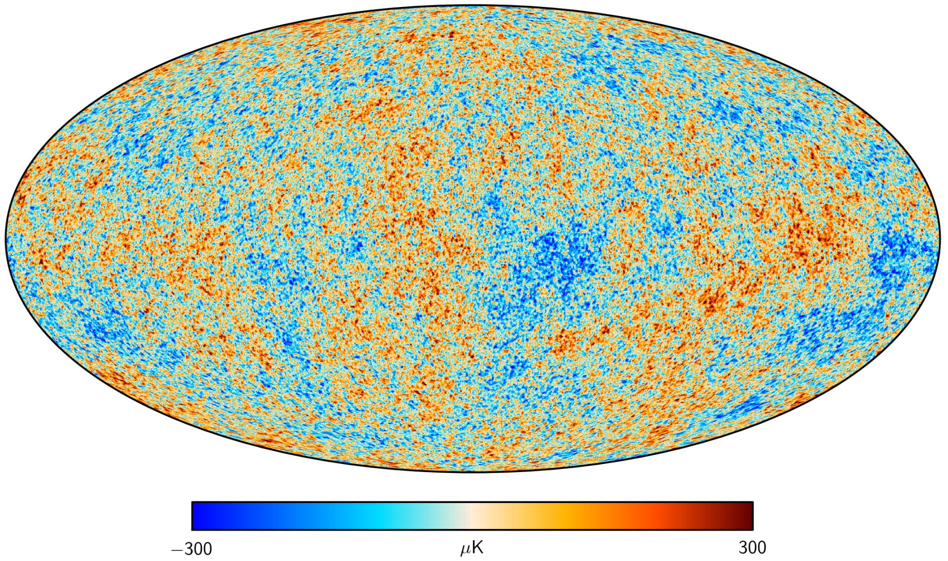


Figure 1.3: Planck CMB Temperature Map showing mild anisotropies of the order of 10^{-5} (figure obtained from ESA / Planck).

A snapshot of the earliest moments of recombination can be detected today in the mostly uniform infrared black-body radiation of the CMB. The latest and highest fidelity map was made with the Planck mission, using the SMICA semi–blind spectral–matching algorithm (Adam et al. (2016), see Figure 1.3). The temperature of the Universe was about 4000 K at $z = 1100$ at the last scattering surface, but due to cosmic expansion, the energy of the CMB photons decreased as they traveled for more than 13 billion years and are now detected at $T_0 = 2.725$ K. The first accurate detection of the black–body curve of the CMB, a successful prediction of the ΛCDM cosmology, was by the FIRAS instrument on the Cosmic Background Explorer (COBE) satellite in 1989 (Mather et al. (1990), see Figure 1.4). This is the most precise cosmological black body spectrum ever measured.

As a blackbody, the energy density of CMB photons is proportional to T^4, as can be shown by integrating the Planck formula, describing the energy density:

$$u(\nu)d\nu = \frac{8\pi h}{c^3}\frac{\nu^3}{e^{h\nu/kT}-1}d\nu \tag{1.4}$$

where h is the Planck constant, c is speed of light and ν is frequency. The T^4 dependence can be derived from the fact that the CMB energy density is a product of the number density and the energy per photon, which have a redshift dependence of $(1+z)^3$ and $(1+z)$, respectively, and keeping mind that the temperature also scales as $1+z$. A straightforward calculation shows that the energy of a typical CMB photon today is several 10^{-4} eV.

An interesting feature of the CMB map is that it is very uniform, as mentioned in §1.1.1. The mild temperature anisotropies (about $10^{-5}K$) are now a well defined

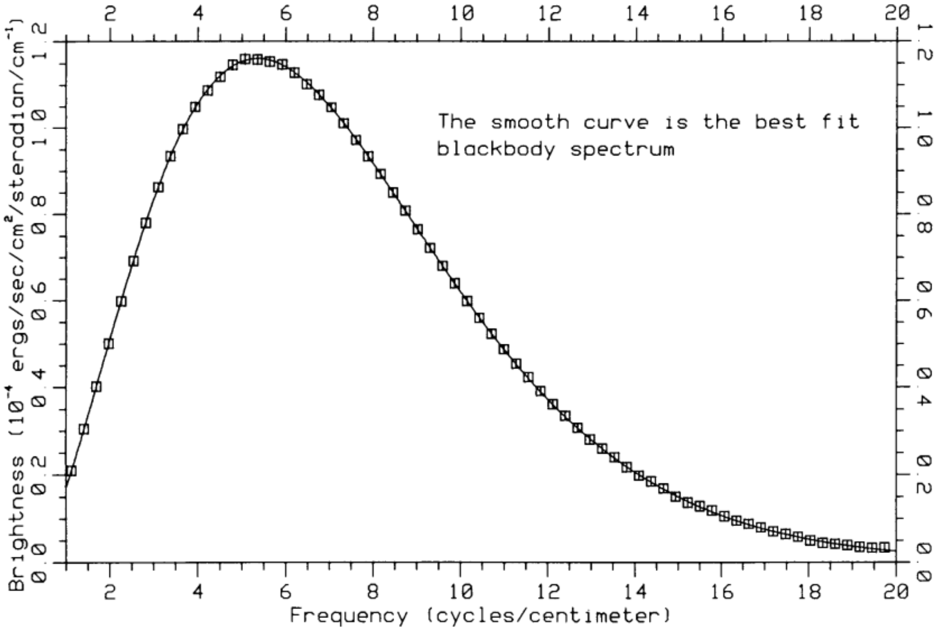


Figure 1.4: Cosmic Background Explorer (COBE) accurately detected the blackbody curve of the CMB (figure obtained from Mather et al. (1990)).

subject (White et al., 1994). The observations for these temperature fluctuations in the CMB, first accurately measured by COBE, improved in fidelity with every major experiment. Wilkinson Microwave Anisotropy Probe (WMAP) observations (Bennett et al., 2013; Hinshaw et al., 2013) made it possible to observe much finer features. Particularly, the noise levels of the CMB map from Planck are much lower than those of WMAP, thus adding valuable information.

1.2 Large Scale Structure

The CMB provides a window to the early Universe, where there are no galaxies, stars, or any form of large–scale structure. Gravity alone is the driving force that eventually evolved the initial Gaussian fluctuations into the tremendously complex filamentary cosmic web that we observe today. According to ΛCDM cosmology, cold dark matter haloes dominate the formation of structures. Since baryonic matter traces the total matter potential, studies on large–scale structure can constrain dark matter formation. Dark matter being cold refers to non–relativistic thermal velocities, first introduced in (Bond et al., 1984; Blumenthal et al., 1984). While the first N–body simulations had low resolutions (e.g. Davis et al. (1985)), the addition of the Press–Schlecter theory (Press and Schechter, 1974), which was modified to better estimate the number of dark matter halos and resolve galaxies (Mo and White, 1996) and the highly increased computing power since then have allowed recent successful DM simulations, having L_{box} sizes reaching hundreds of h^{-1}Mpc and following more than 2000^3 particles from $z = 127$ to the present (Springel, 2005; Springel et al., 2005; Boylan-Kolchin et al., 2009).

Hot DM comprises of weakly interacting particles like neutrinos, which were traveling at nearly the speed of light at high redshifts. By itself, hot DM is not enough to describe the density fluctuations in the distributions of galaxies (sizes of kpc scales and masses of about $10^{11} M_{\odot}$), the free streaming would have smoothed out the initial fluctuations. The Universe we observe today supports the existence of DE and is compatible with dark matter being mostly cold (Ade et al., 2016).

In order to identify the large–scale structure of the Universe, one can either observe structures directly with galaxy surveys, as we will examine in the following section, or indirectly, using absorption features. However, direct observation cannot continue indefinitely with increasing redshift, due to the scaling of galaxy surface brightness with redshift as $(1 + z)^{-4}$. Therefore, complementary methods are required for medium to high redshifts. Luminous red galaxies (LRGs) have a narrow interval of wavelength and luminosity and they can be observed to distances greater than those of regular galaxies, due to their greater luminosity (Postman and Lauer, 1995; Tegmark et al., 2006). As the redshift is increased, quasar (QSO) clustering and the absorption features of 21 cm and the Lyman alpha forest (Lyα) are the most common means of investigating the LSS, which will be explained in the following sections. A few examples of observational surveys that investigate the large scale structure at high redshifts are the Baryon Oscillation Spectroscopic Survey (BOSS), Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) (Adams et al., 2010) and Joint Dark Energy Mission (JDEM), currently replaced by Wide Field Infrared Survey Telescope (WFIRST) (Green et al., 2012).

1.2.1 Intergalactic Medium

The bulk of the mass of the Universe exists as a dilute medium in the void between galaxies, the intergalactic medium (IGM). Roughly half of the dark matter is thought to reside in the IGM volume, whereas for baryonic matter, the fraction located in the IGM is probably much higher. While there is no clear distinction for the boundaries of the IGM, the general understanding is that it is not bound to any galaxy as virialized matter. However, since the first galaxies started forming at redshifts $10 < z < 15$, this definition is limited to the epoch after the beginning of reionization. Earlier, in the absence of galaxies, the cosmic gas temperature was coupled to the CMB temperature, until $z \sim 147$, when the free electron ratio became too low to sustain the coupling any longer. The cosmic gas cooled adiabatically to ~ 2 K at $z = 10$. (Kulkarni et al., 2015). As the first galaxies in the Universe started to form, this affected the IGM with metal enrichment and their radiative backgrounds, ionizing almost all of it. This also heated the gas to several 10^4 K, which is the characteristic temperature for Lyα studies, smoothing the distribution of the IGM and affecting all subsequent galaxy formation (Ostriker and Ikeuchi, 1983). Therefore, there is close interplay between galaxy formation and the IGM. In general, the IGM is modeled well in the typical Lyα redshift range of $2 < z < 4$.

IGM studies yield many different insights into the history of the Universe, e.g., peak density analysis suggests that the first stars in the Universe formed most likely at $z \sim 65$ (Naoz et al., 2006). One can study the implications of galaxy formation on CMB anisotropies (Ostriker and Vishniac, 1986), test our models of structure formation and inflation (Seljak et al., 2005), impose limits on the mass of warm dark matter candidates (Viel et al., 2005), examine the filamentary topology (Sousbie et al., 2008; Caucci et al., 2008), and study the heights and the density of the peaks in the IGM to reveal information about the initial conditions of the Universe, specifically the linear matter power spectrum (Croft and Gaztañaga, 1998; Croft et al., 1999; De and Croft, 2007, 2010), the ionizing radiation and the large scale structure (Croft et al., 2002), and also the coldness of dark matter (Viel et al., 2013).

The optical depth, in general, is defined as the natural logarithm of the ratio of the incident to the transmitted intensity through a material. Flux is mapped to the optical depth of the IGM as $F = e^{-\tau}$, which is related to redshift as (McQuinn, 2015):

$$\tau_{Ly\alpha}(z) \approx 1.3\,\Delta_b \left(\frac{x_{HI}}{10^{-5}}\right) \left(\frac{1+z}{4}\right)^{3/2} \tag{1.5}$$

where the fluctuating Gunn–Peterson approximation is used: Lyα absorbers are assumed to trace the dark matter distribution and in a photoionization equilibrium with the UV background radiation, and peculiar velocities are ignored (Weinberg et al., 1997). Δ_b is the baryonic density in units of the cosmic mean ($\rho/\langle\rho\rangle$), often referred to as overdensity or density contrast. Typical overdensities are mild ($\delta \leq 10$) at high redshifts ($z > 2$), although in the Warm Hot Intergalactic Medium (WHIM) at $z \sim 0$, it is moderately higher ($\delta < 100$) because of the highly evolved structure, and the temperature is up to three orders of magnitude greater (Klar and Mücket, 2010). Evaluating Equation 1.5 at $z = 3$, the neutral hydrogen fraction is found to be $x_{HI} \sim 10^{-5}$, which translates to an HI number density of $n_{HI} \sim 10^{-10} cm^{-3}$. Again under the Gunn–Peterson approximation, temperature and density are related through the equation of state

$$T = T_0 \left(\frac{\rho}{\langle\rho\rangle}\right)^{\gamma-1} \tag{1.6}$$

where T_0 is the temperature at the mean density. If the IGM is reionized quickly compared to the cosmic expansion, then the index γ is close to unity, i.e., the gas is isothermal. During HeII reionization at $z \sim 3$, the gas was nearly isothermal (McQuinn et al., 2009). Away from reionization, due to the expansion of the Universe, the mean temperature T_0 decreases and the index γ increases from near unity immediately after reionization to ~ 1.5 at lower redshifts (Hui and Gnedin, 1997; Fang and

White, 2004), although a contradictory "inverted" temperature–density relation has also been proposed (Bolton et al., 2008; Garzilli et al., 2012).

At large scales (> 1 Mpc) and low densities, baryons in the IGM follow the total matter potential, therefore it is a tracer of dark matter. However, at characteristic scales smaller than ~ 100 kpc and high densities, pressure forces effectively result in smoothing of the gas, preventing it from tracing the collisionless dark matter. This exact scale for pressure smoothing, analogous to the classic Jeans length scale, is a function of the gas temperature and it depends on the entire thermal history (Binney and Tremaine, 1998). It is the scale under which the outward pressure is greater than the inward gravitational forces, regulating gas perturbations hydrodynamically. Furthermore, galaxy formation physics needs to be taken into account at these scales, as they affect the baryon distribution in the IGM. In this thesis, since we deal with smoothing scales as large as ~ 40 Mpc, we need not consider pressure smoothing in our analysis. However, it is important to be aware of small scale behavior, which may be important, for example, in future peak density studies, provided the spatial resolution is sufficiently high to observe sub–Mpc scales.

1.2.2 Galaxy Surveys

Dark matter halos form by the hierarchical aggregation or collapse of massive systems (Lacey and Cole, 1993, 1994; Springel et al., 2006), characterized by their large overdensities. The gravitational interaction within the DM halos retain the gas heated to temperatures $T_{virial} \gtrsim 10^4$ K, photoheated due to ionizing background ration (Mesinger and Dijkstra, 2008). As the gas cools and fragments, galaxies start forming as concentrated luminous clumps within them (White and Rees, 1978; Bromm and Yoshida, 2011). Dwarf galaxies eventually evolve into the galaxies we see today, forming towards the end of the reionization epoch, between the redshifts $6 < z < 15$ (for a study on the redshift constraints for reionization, see Pritchard et al. (2010)). According to recent observations, the star formation rate within galaxies seems to have peaked at $z \sim 1 - 2$, while the vast majority of the stars in our Universe has formed from $z = 3$ to present (Baugh et al., 1998).

On scales smaller than 10 h^{-1}Mpc, galaxy distribution can be modeled with a power law $\xi(r) = (r_0/r)^\gamma$, with the correlation length $r_0 \sim 5\,h^{-1}$Mpc. The angular galaxy correlation function can also be described by a power law: $w(\theta) = A_w\theta^{1-\gamma}$, where $\gamma \sim 1.7$ at angular scales less than about 10 degrees, after which it declines rapidly (Peebles (1980), for a recent study on angular correlations, see Connolly et al. (2002)).

Fundamentally, there is no reason to assume that the observed light from galaxy surveys shows the mass distribution of galaxies in the Universe exactly. Kaiser (1984) showed that galaxies should have a large bias due to being rare objects forming at the highest density peaks above a threshold. In general, the galaxy distribution is expected to be non–local and a tracer of the underlying dark–matter density in a stochastic manner. On large scales, the mean overdensity of galaxies can be related to the mean overdensity of mass with a linear bias:

$$b = \delta_g/\delta_m \tag{1.7}$$

where δ_g denotes the overdensity of galaxies, δ_m is the overdensity of the underlying total mass and b is the linear bias. The respective power spectra (or equivalently, the corresponding two–point correlation functions) can be related as $P_g(k) = b^2 P_m(k)$. This bias is shown to be scale dependent from simulations and observations, for example, 2dFGRS found a linear bias close to unity on large scales (Verde et al., 2002; Gaztañaga et al., 2005).

Recent galaxy surveys have been able to map out the galaxy distribution up to $z \sim 0.3$, although Sloan Sky Digital Survey (SDSS) has observed galaxies at redshifts as high as 0.8. Until now, the most important galaxy surveys have been The Las Campanas Redshift Survey (Shectman et al., 1996), The Center For Astrophysics Redshift Survey (Falco et al., 1999), The CNOC2 Field Galaxy Redshift Survey (Yee et al., 2000), The 2dF Galaxy Redshift Survey (Colless et al., 2003), AGES: The AGN And Galaxy Evolution Survey (Kochanek et al., 2012) and DEEP2 Redshift Survey (Davis et al., 2003; Newman et al., 2013). The power spectrum of galaxy clustering on scales up to 300 h^{-1}Mpc has been measured with the 2dFGRS survey (Percival et al., 2001) and BOSS (Anderson et al., 2014), which is found to be compatible with the filamentary structure predicted by a CDM Universe (Colberg et al., 2005).

As mentioned before, it is difficult to observe objects at higher redshifts due to rapidly decreasing brightness, especially beyond $z > 1$. Furthermore, galaxy spectra between $1.5 < z < 3$ do not have observable features which correspond to the most sensitive part of optical wavelengths of observational equipments. For example, the wavelengths accessible to the BOSS experiment of the SDSS are between 3600 and 10000 Å, which would correspond to original galaxy spectra between 1500 and 3000 Åbetween $1.5 < z < 3$, but there are no clear features to be observed in this range. It should be noted that spectra at shorter wavelengths can be examined with telescopes above the atmosphere. Hydrogen absorption at 912 Å makes it possible to observe distant galaxies at $z > 3$ with ground–based telescopes. However, due to the low surface brightness of these distant objects, it is necessary to rely on other methods such as 21 cm intensity mapping and the Lyα forest.

1.2.3 21 cm Intensity Mapping

21 cm intensity mapping is observed as an emission line, related to the splitting in the hyperfine structure of the ground state of neutral hydrogen. First thought to be noise due to the Sun, it was first detected using a microwave radiometer built specifically for this purpose (Ewen and Purcell, 1951). The term intensity mapping refers to translating the received redshifted frequency to comoving distance and the amplitude of the signal to mass density.

Since the 21 cm signal only originates from neutral hydrogen, a natural choice of redshift interval is before the end of reionization at $z \sim 6$, with the highest signal–to–noise ratio expected at $z = 9 - 10$ (Pritchard and Loeb, 2012). For example, James Webb Space Telescope (JWST) will potentially image some of the early galaxies at $10 < z < 15$, launching in 2018. Neutral gas shielded within spiral galaxies can also be studied with 21 cm methods (Scodeggio and Gavazzi, 1993). While most such structures are observed at $z < 0.1$ (Peterson and Suarez, 2012), some studies extend this range to intermediate redshifts (Gupta et al., 2009a,b). Furthermore, the neutral gas and ionization fractions can be studied at redshifts as high as $z \sim 10$ and thus obtain information about the reionization epoch (Santos et al., 2008).

In the Lyα forest redshift range of $2 < z < 4$, the low neutral hydrogen ratio of $x_H \sim 10^{-5}$ makes it a less attractive choice for 21 cm intensity mapping, in view of the 21 cm signal being much weaker a signal than galactic and extragalactic foregrounds (Curran and Whiting, 2012). From this perspective, 21 cm intensity mapping and the Lyα forest are complementary methods. Furthermore, studying the two methods together means looking at both HI emission and absorption, hence gaining information about the location and the amount. In terms of the strength of the signal, the Lyα forest is a much more sensitive method, since the absorptional cross section is about 10^5 larger for Lyα absorption compared to that of 21 cm. Combining the results of the two methods, better maps can be made by cross–correlating the two. The power spectrum calculated with simulations at $z = 2.4$ are already suggesting good agreement for modes with wavenumbers $k < 0.2\ hMpc^{-1}$ (Carucci et al., 2016). In particular, it is possible to enhance our understanding of the reionization era by constraining cosmological variables (Pritchard et al., 2010), keeping in mind the complementary nature of the redshift ranges available to the two methods mentioned above.

1.2.4 The Lyman Alpha Forest

The Lyman Alpha Forest (Lyα) refers to the imprints of the intervening HI clouds along lines of sight (LOS) as absorption lines in quasar spectra, observed in the ultraviolet (UV) and optical wavelength range. As discussed in §1.2.2, galaxy surveys are limited to low redshifts and the 21 cm intensity mapping method requires a relatively high neutral hydrogen fraction, therefore, other probes such as the Lyα forest are necessary to obtain quantitative information at high redshifts. The comoving space density of tracer objects in current large–scale maps of the Universe is $\sim 3.6 \times 10^{-4} h^3 \mathrm{Mpc}^{-3}$ at $z \sim 0.5$, which declines rapidly with increasing redshift. At high redshifts ($z > 2$), quasars are the only tracers that can be used to make such maps, where the space density of objects is about two orders of magnitude lower: $\sim 10^{-6} h^3 \mathrm{Mpc}^{-3}$. This highlights another feature of studying the large scale structure in the IGM: It probes mild overdensities, while galaxy surveys do not.

Quasars (quasi-stellar radio sources, or QSOs) are extremely luminous objects in the proximity of a black hole within a galaxy, falling into its accretion disk. QSOs are a type of active galactic nuclei (AGNs). Unlike the relatively narrower spectra of stellar populations, QSOs have mostly flat spectra, making them a useful cosmological tool. The first absorption lines in QSO spectra were seen in 1960s, independently by several groups (Scheuer 1965; Gunn and Peterson 1965; Shklovskii 1965). Gunn & Peterson also predicted a trough, which is when the absorption lines are showing complete saturation due to the amount of intervening HI being above a certain threshold. In 1970, Roger Lynds also observed prominent absorption lines blueward (shorter wavelength side) of the Lyα emission line in the furthest quasar observed to date, QSO 4C 05.34 at $z = 2.88$ (Lynds, 1971). He attributed the regularity of these lines to a preferred transition, happening at many different redshifts along the line of sight from the particular quasar, now known as the Lyα forest.

On their path from a QSO to the Earth, photons are absorbed by intervening neutral hydrogen atoms at ground state, thus providing the required energy for the transition to the first excited state ($1s \to 2p$), which happens at a rest wavelength of $\lambda_{Ly\alpha} = 1215.67$ Å. Due to the expansion of the Universe, the wavelength of photons are stretched by a factor of $1 + z$. Therefore, along the LOS from the QSO to the Earth, wavelengths of photons that are smaller than $\lambda_{Ly\alpha}$ will be stretched to it eventually, and some of these photons will be absorbed within neutral hydrogen clouds at redshifts lower than that of the particular quasar. The absorption line wavelength, as observed from the Earth, provides redshift (distance) information, while the strength of the dip provides the amount of matter (density) information at that location along the skewer. This is the Lyα forest (Figure 1.5, top panel), a collection of absorption lines blueward of the strong Lyα emission line of the quasar. The region redward of it are absorbed due to other chemical transitions (metal lines). The redshift information of the QSO can be read off immediately by simply dividing the QSO Lyα emission wavelength (~ 5600 Å) by the Lyα transition wavelength (1215.7 Å), which yields $1 + z \approx 4.6$. Therefore, the QSO is found to be at a redshift of $z \approx 3.6$. The bottom panel is a zoomed–in portion of the Lyα forest in order to see finer features in a selected narrow redshift range. The absorption in flux and the density is inversely correlated: The middle panel of Figure 1.5 shows how neutral hydrogen density information can be inferred from observed flux (Springel et al., 2006).

When one has saturation in the Lyman-α forest, the Lyman-β transition should provide information due to its lower cross section, which makes it a potentially better probe at high overdensities as high as 10 times the mean density at $z = 2 - 3$ (Shull et al., 2000). Since the Lyman-β absorption occurs at a lower rest wavelength of 1026 Å, the Lyman-α forest overlaps the Lyman-β forest, and therefore, statistical techniques are needed to make use of the Lyman-β information (Dijkstra et al., 2003; Iršič et al., 2013). In principle, higher order transitions could be used together with the Lyman-α forest in mapmaking also.

The Lyα forest consists of mild overdensities, typically between 0.1 and 10 times

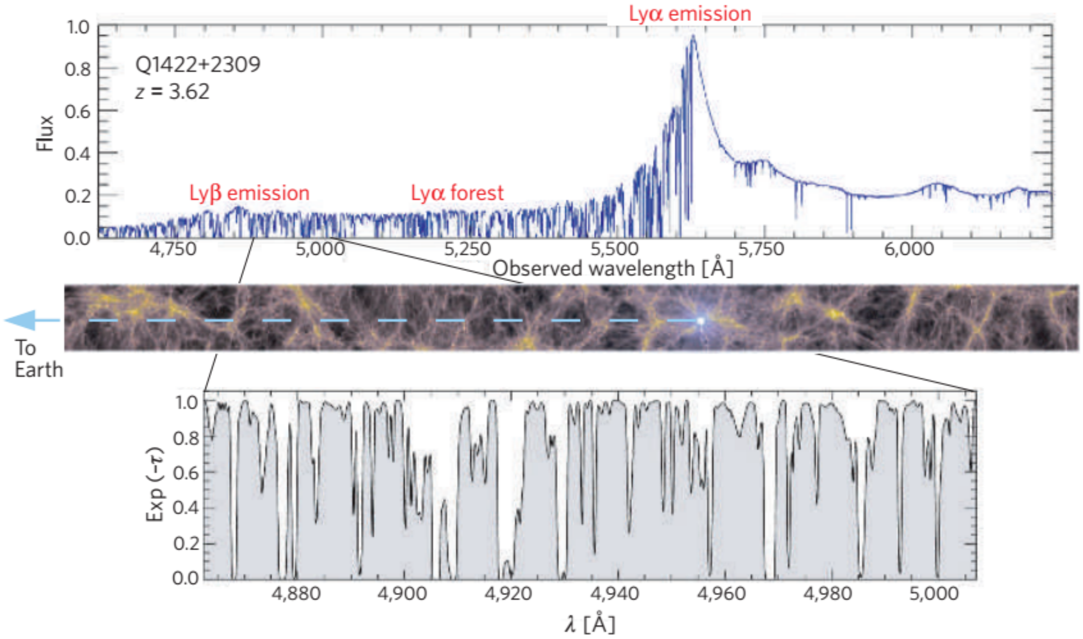


Figure 1.5: High resolution spectrum of a quasar at $z = 3.62$. The absorption lines blueward of the emission line make up the Lyα forest (top panel). The corresponding intervening large–scale structure, the neutral hydrogen clouds, is shown in the middle panel (figure obtained from Springel et al. (2006)).

the cosmic mean. Baryons and dark matter trace each other well. For most of the volume, it is possible to relate the previously defined optical depth, column density, temperature and baryon density via

$$\tau \propto n_{HI} \propto \rho_b^2 T^{-0.7} \Gamma^{-1} \tag{1.8}$$

where $T^{-0.7}$ accounts for the temperature dependence of the recombination rate and Γ describes the rate of ionization of HI clouds due to the UV background (Weinberg et al., 2003). With the assumption that the IGM is in ionization equilibrium, the Γ term can be ignored. To a good approximation, the baryonic density and the optical depth follow the relation $\tau \propto A(\rho/\bar{\rho})^{\beta}$, where A is a redshift dependent term and $\beta \sim 1.6$. This is commonly referred to as the fluctuating Gunn–Peterson approximation (Rauch, 1998). Although this theoretical picture neglects effects like peculiar velocities, collisional ionization and thermal broadening, simulations with all of these effects are shown to obey the relation between the optical depth and the underlying mass density as mentioned above (Croft et al., 1997).

The absorber density along Lyα skewers is mostly uniform, except in the vicinity of a QSO, they are seen to decrease. This can interpreted as QSOs ionizing nearby clouds due to their huge ultraviolet (UV) flux. A common statistic for absorbers is column density, the number of absorbers per unit length per unit area, often expressed in cm^{-2}. Absorption systems with a column density of $N_{HI} < 10^{17} cm^{-2}$ are available to the Lyα forest. Those with $10^{17} < N_{HI} < 10^{20} cm^{-2}$ are Lyman limit systems, which are optically thick to ionizing radiation. The general topology of the IGM seems to be sheet–like for $N_{HI} < 10^{14} cm^{-2}$, filamentary for $N_{HI} \sim 10^{15} cm^{-2}$ and clouds for $N_{HI} > 10^{16} cm^{-2}$. Finally, the densest absorption systems are DLAs (damped Lyα systems) with $N_{HI} > 10^{20} cm^{-2}$, self–shielded and mostly neutral, with damping wings due to Gaussian thermal and Lorenztian pressure broadening. Since DLAs contain most of the neutral gas mass in the Universe, it is suggested that these gas clumps are closely related to galaxy formation (Maller et al., 2001, 2003). As an example, in the bottom panel of Fig 1.5, it appears that at there are two DLAs at $\lambda = 4920$ Å and 4970 Å, easily recognized due to the deep trough and damping wings, corresponding to $z = 3.05$ and $z = 3.09$, respectively.

It is important to consider the number evolution of the hydrogen clouds and the relevant observational methods available. From the present to $z \sim 1$, the comoving number of clouds does not change in a noticeable manner. A power law $(1 + z)^{\gamma}$ with $2 < \gamma < 3$ describes the trend becoming sharper in $1 < z < 2$. This becomes even more steep as the redshift approaches $z \sim 4$. The Lyα forest becomes more transmissive due to cosmological expansion and the probable consumption of gas into stars as the large scale structure evolves. For example, the mean optical depth is $\tau_{eff} = 0.016(1+z)^{1.1}$ over $0 < z < 1.2$, much lower than it is in the typical Lyα range $2 < z < 4$, where $\tau_{eff} = 0.00211(1 + z)^{3.7}$ shows a much sharper dependence of the optical depth on the redshift (Meiksin, 2006). It is worth noting that the sparse Lyα forest at low redshifts does not contribute significantly to the fact that we cannot detect most of the baryons $z \sim 0$ (the missing baryon problem, Bregman (2007)). In any case, only limited wavelengths are available to ground surveys, therefore limiting the redshift ranges available to Lyα forest studies. For example, the lower limit of wavelength for the BOSS spectrograph is 3600 Å, therefore the structures at lower redshifts such as $z < 2$ are impossible to observe.

1.2.4.1 Observational Surveys

The Lyα forest has shed light on many cosmological unknowns within the last two decades. Prior to the late 80s, the two major limiting factors in Lyα studies were the low spectral resolution and signal–to–noise (S/N) ratio. In 1994, the 10 m Keck telescope with the High Resolution Spectrograph (HIRES) was a breakthrough, with a high spectral resolution of up to 67000 and signal–to–noise ratios of ~ 100 (Vogt et al., 1994). UVES on the Very Large Telescope, and later SDSS further improved on Keck, which will be discussed in §1.3.2.

As the observational sky surveys advanced, new analytical and numerical methods appeared. The evolution of the observed Lyα forest was shown to be consistent via numerical simulations with photo–ionization, and the two–point correlation agrees with that of the galaxy distribution (Mücket et al., 1995). The findings showed that the sharp neutral hydrogen limit imposed by Gunn & Peterson should be altered to the continuous absorption features seen in the Lyα forest due to the uniform medium at high redshifts (Reisenegger and Miralda-Escude, 1995). The HeII Gunn–Peterson absorption effect is predicted, however, with fiducial CDM models (Miralda-Escude et al., 1996; Croft et al., 1997). This is due to the relative strength of the HeII transition at 304 Å, which may be an even better probe than HI. Linear power spectrum of mass fluctuations was recovered using hydrodynamic simulations and the same method was also applied to a QSO spectrum from Keck HIRES (Croft et al., 1998). The baryon acoustic oscillations (BAO) signal, which results from the interplay between gravity and pressure in primordial anisotropies, and with the photonic pressure removed after decoupling, sets a standard cosmic comoving length of ~ 150 Mpc. The Lyα forest of BOSS quasars provide BOA information in the three dimensional correlation function of the transmitted flux for $2.1 < z < 3.5$ (Busca et al., 2013). These results all follow from absorption lines, but it is also possible to study the large scale structure with Lyα emission lines of QSOs by examining cross–correlations with galaxy spectra at high redshifts ($2 < z < 3.5$) and calculating the contribution to star formation rate, using SDSS and BOSS data (Croft et al., 2016).

Besides the many aspects of Lyα forest cosmology already discussed, it also acts as a probe of large scale structure at high redshifts. The quality of the density inference from QSO spectra depends heavily on the observed QSO density in observational sky surveys. Before recent large-scale structure surveys such as BOSS (Dawson et al., 2013), the sky density of known background quasars in this redshift range over most of the sky was of the order of 1 per square degree (e.g., from the 2dF quasar survey (Miller et al., 2002; Outram et al., 2003; Croom et al., 2004; Miller et al., 2004) and from SDSS I and II (Schneider et al., 2002, 2003, 2005; Richards et al., 2006). Except for some small areas with higher observed densities of objects (e.g., Rollinde et al. 2003) the Lyman-α forest was treated as a collection of discrete 1D individual quasar sightlines.

Recently, however, the increasing number of discovered quasars with suitable redshifts ($z > 2$) for ground based study has made it possible to correlate information over large scales in three dimensions. BOSS features a high QSO density of ~ 15 deg^{-2} ($\sim 180,000$ QSOs in the redshift range $2.15 < z < 4$ over 10,000 deg^2). Each QSO typically provides Lyman-α forest information along a skewer of length ~ 400 h^{-1}Mpc, and the typical mean separation for spectra in BOSS is ~ 20 comoving h^{-1}Mpc (Lee et al., 2013). This is what has enabled clustering statistics of the Lyman-α forest to be measured in three dimensions as mentioned above. In the future, one can expect yet higher densities of sightlines and more precise measurements, in view of the fact that even more quasars will be available for analysis (e.g., 45 deg^{-2} proposed for Mid–Scale Dark Energy Spectroscopic Instrument (MS–DESI; (Levi et al., 2013), see Table 2.1).

Despite the developments with the recent surveys, the QSO field is still sparse in the Lyα forest range $2 < z < 4$, and due to their sharply decreasing luminosity with redshift, QSOs are increasingly more difficult to observe at even higher redshifts. It took more than 30 years to observe a quasar far enough to prove the existence of the complete Gunn–Peterson trough by observing a spectrum at $z = 6.28$ from Sloan Digital Sky Survey data (Becker et al., 2001). This also suggests that epoch of reionization is ending at $z \sim 6$. Recent observations show that the observed QSO distribution reaches the maximum at $z \sim 2.25$, rapidly decreasing after this peak with very few QSO observations for redshifts greater than 4 (Pâris et al. (2016), see Figure 1.6). The highest redshift quasar observed is at $z = 6.440$.

Figure 1.7 shows the projected areal sky density of quasars in MS–DESI. The observational data set we will use in §4, which is taken from SDSS–III BOSS, was multiplied by a factor of 5.25 to extrapolate from the true count of observed BOSS QSOs (black curve). This also shows a peak at $z \sim 2.25$, similar to the one in Figure 1.6. For purposes of recovering the IGM, however, the red curve shows a more realistic density of probes: At each redshift bin, all quasars up to $\Delta z \sim 0.5$ further than that bin are taken into account, in view of the fact that a quasar illuminates about 400 h^{-1}Mpc along a skewer, starting from about 100 h^{-1}Mpc in front of it (in order to avoid having to model the Lyα emission line of the QSO). However, this must be treated as an upper bound, as quasars can be located far away from each other, not contributing to the red curve. For the redshift range of interest ($2 < z < 2.5$), the number of probing quasars for MS–DESI are as high as 70. This makes IGM topology with the Lyα forest a suitable choice when the MS–DESI data will be available.

1.3 Data

This thesis consists mainly of two sections: First we evaluate the performance of our map making with hydrodynamical simulations, followed by applying our methods to observational data for Sloan Digital Sky Survey in order to create real maps of the IGM at Gpc scales. In this section, we describe the details of the data used.

1.3.1 Smoothed Particle Hydrodynamics

The simulation used to test our IGM map making methods was run with the highly parallelized GADGET–2 code, a smoothed particle hydrodynamics (SPH) implementation with the N–body method (Springel, 2005; Di Matteo et al., 2012). It improves on the first version of GADGET (Springel et al., 2001) with more accuracy, speed of computation and better memory efficiency. Amongst the many possible uses of SPH applications, difficult problems posed in cosmology are solved successfully with the flexibility it allows.

SPH is a local interaction with an adaptive smoothing length. Naturally, the resolution depends on the local density. In the limit of infinitely many particles, it can

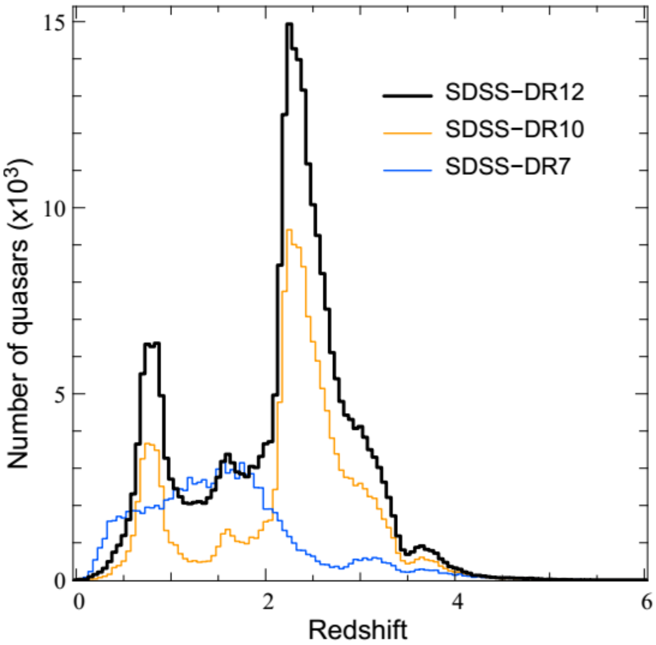


Figure 1.6: The evolution of the redshift distribution of QSOs from SDSS-III through several data releases. There is a sharp peak at $z \sim 2.25$, followed by a rapid decline (figure obtained from Pâris et al. (2016)).

be proven that the SPH method solves Euler's equations. The fluid dynamics equations reflect conservation laws as partial differential equations, which are transferred to integral equations through interpolation with a suitable kernel (Allahdadi et al., 1993). Energy and mass are explicitly conserved within these continuum equations. As it appeared in the original calculations (Gingold and Monaghan, 1977), it is best to assume a Gaussian kernel for the physical interpretation of the local smoothing, although many different kernels can be used with a necessity for normalization and a preference for compact support and vanishing first and second moments in order to make computations more tractable (Monaghan, 1992). The kernel typically encompasses at least $20-50$ particles. The SPH method has the advantage that the kernel can be included separately in a subroutine or a table, and it can be changed later trivially if the need arises.

In the SPH formulation, calculations are carried out based on particle separations only. The absence of a mesh (used in Eulerian methods) makes it possible to compute large deformations easily, hence making it a valuable computational tool. In principle, the Lagrangian calculation, as used in GADGET–2, comprises a moving frame, which yields more accurate results, compared to the Eulerian formulation. However, the Eulerian calculation generally has better applicability. Being a Lagrangian method, the SPH approach comprises full time derivatives in the equations explaining fluid properties of the particles such as density, the specific internal energy and the velocity

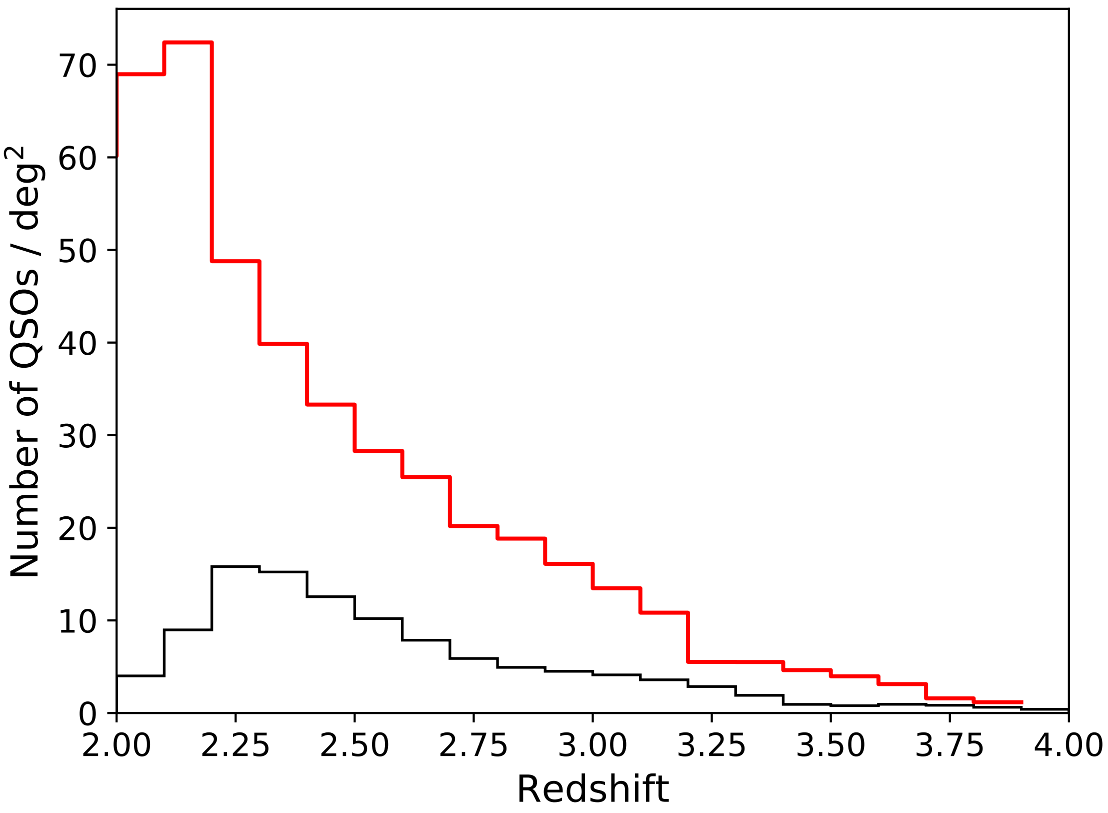


Figure 1.7: The projected quasar distribution per square degree as a function of redshift for the survey MS–DESI is shown by black color. The red curve shows the quasar density probing the flux field as a function of redshift, taking all quasars with $\Delta z \sim 0.5$ into account.

components, compared to partial derivatives in the Eulerian treatment.

SPH simulations make it possible for baryonic gas particles and dark matter to be tagged separately, as it was for the simulation used in this thesis. As an example, the Millenium simulation (Springel et al., 2005; Bett et al., 2007; Overzier et al., 2012) used 10 billion dark matter particles (and no baryonic matter), which were evolved in the ΛCDM cosmology in a cube over 2 billion light–years (500 h^{-1}Mpc) on a side, with high resolution. This cube size and the number of particles used in Millenium simulation are similar to those of the simulation we used in this thesis.

1.3.2 Sloan Digital Sky Survey

Prior to SDSS–III and SDSS-IV, the first two generations of SDSS (York et al., 2000) had been the largest redshift survey, providing redshifts of over one million galaxies in spectroscopic observations between 2000 and 2008 over a sky coverage of about 8000 square degrees. Located at the Apache Point Observatory (APO), Sunspot, New Mexico, it uses a dedicated 2.5 m telescope with a mosaic CCD camera operating in five optical bands. The continuous wavelength coverage is approximately 3800 Å to 9200 Å and the wavelength resolution ($\lambda/\Delta\lambda = c/\Delta v$) was 1800. Most of the sky was observed once or twice, with the exception of "Stripe 82" over 300 deg^2, observed between 70 and 90 times in order to look for high quality supernovae and improve photometric calibration (Frieman et al., 2007; Jiang et al., 2014).

Since the goal of obtaining the spectra of $\sim 10^6$ galaxies and $\sim 10^5$ QSOs was not entirely fulfilled in the first iteration of the survey, SDSS–II continued the observations, while also adding Sloan Extension for Galactic Understanding and Exploration (SEGUE) for a star survey (Abazajian et al., 2009). The third generation of the Sloan Digital Sky Survey (SDSS–III) used the same 2.5 m telescope (Eisenstein et al., 2011), taking data from 2008 to 2014 and catalogued more than 4 million unique spectra (Alam et al., 2015), with an extended sky coverage of over 10000 square degrees. It consisted of four separate surveys: Baryon Oscillation Spectroscopic Survey (BOSS) (see Dawson et al. (2013), BOSS will be explained in detail in §1.3.2.1), Sloan Extension for Galactic Understanding and Exploration (SEGUE–2), which surveyed 119000 unique stars (Yanny et al., 2009; Bond et al., 2010; Deason et al., 2011; Gómez et al., 2012), The APO Galactic Evolution Experiment (APOGEE), (Frinchaboy et al., 2013; Anders et al., 2014) and The Multi-object APO Radial Velocity Exoplanet Large-area Survey (MARVELS), which used a new technique to observe 60 stars at the same time to observe radial velocity variations. Until the original SDSS spectrographs were decommissioned in 2009, 860836 galaxies, 116003 quasars and 521990 stars were observed (Albareti et al., 2016).

All of the cumulative data that have been observed with the Sloan Digital Sky Survey are made public. In this thesis, we use Data Release 12 (DR12), which contains all the data that have been taken until 2014, and it also contains the first spectra of MARVELS. The sky coverage is almost completely within the Northern Hemisphere, within the right ascension (RA) and the declination (DEC) range $0^\circ < DEC < 30^\circ$ and $120^\circ < RA < 240^\circ$, although it also covered a relatively small area in the southern galactic cap. Figure 1.8 shows the sky coverage evolution of data releases 9 through 12.

SDSS–IV's latest and final release is Data Release 13 (DR13), which consists of three main programs: APOGEE–2, MaNGA and eBOSS. SDSS–IV observations began in July 2014 and they are planned to continue until 2020. For a technical explanation on what is included in DR13, the reader is referred to (Albareti et al., 2016).

The power spectrum of galaxy clustering on scales up to 300 h^{-1}Mpc has been

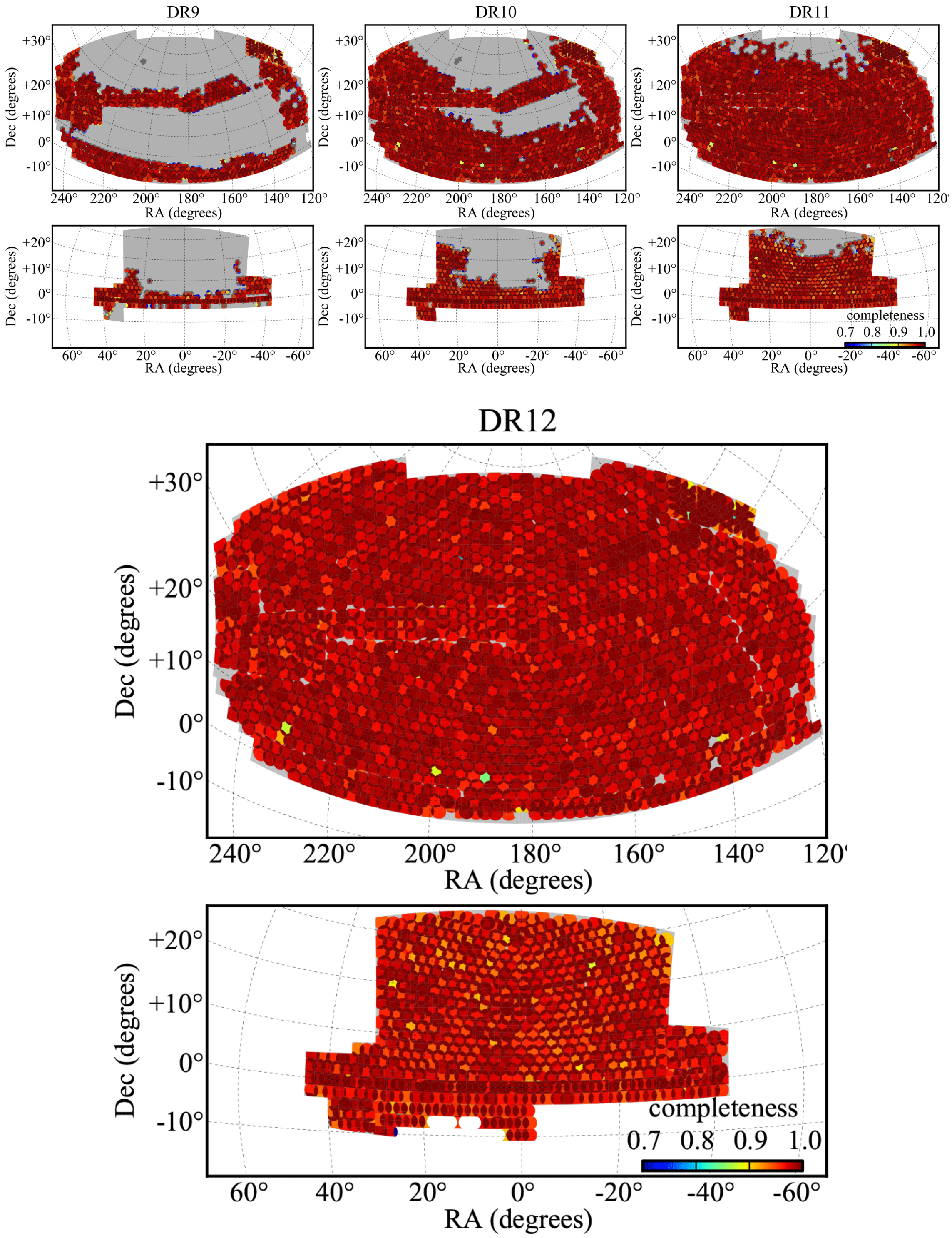


Figure 1.8: Footprints of BOSS Data Releases DR9 through DR12 (figure obtained from Alam et al. (2016)).

measured with the 2dFGRS survey (Percival et al., 2001) and BOSS (Anderson et al., 2014), found to be in accordance with the filamentary structure as predicted by a ΛCDM Universe (Colberg et al., 2005). The BAO signal shows up as a bump in the galaxy correlation function (Eisenstein et al., 2005), measured with a galaxy sample between $0.16 < z < 0.47$, and (Anderson et al., 2014), where the range was $0.43 < z < 0.57$). Since galaxies become fainter rapidly with increasing redshift, scaling as $(1 + z)^{-4}$, one has to rely on other methods to detect the BAO feature at higher redshifts, and to study the large–scale structure, which can be done with quasar spectra. The Lyman-α forest of BOSS quasars provide BOA information in the three dimensional correlation function of the transmitted flux for $2.1 < z < 3.5$ (Busca et al., 2013).

Since the Lyα absorption features of the IGM are imprinted in on QSO spectra, it is possible to map the large scale structure at high redshifts by interpolating the Lyα skewers. Naturally, the sky density of observed QSOs, illuminating 1–D skewers along the LOS for $\sim 400\ h^{-1}$Mpc in the typical Lyα forest redshift range, directly affect the fidelity and the resolution of the resulting inferred map, as will be explained in §2.1. While most QSOs observed to date fall in the Lyα forest redshift range (Lee et al., 2013), there also exist studies which examine quasar absorption spectra at higher redshifts (Perrotta et al., 2016).

In earlier studies, the correlation in the Lyα absorption skewers could only be measured along individual lines of sight due to the sparse nature of quasar density (e.g. Kaiser and Peacock (1991)). The only way of analysing the transverse correlation of the IGM was with the absorption spectra of high–redshift quasar pairs (Coppolani et al., 2006). While the QSO areal density was ~ 1 deg^{-2} in previous surveys such as 2df (Miller et al., 2002; Outram et al., 2003; Miller et al., 2004) and SDSS I and II (Schneider et al., 2002, 2003, 2005), the current SDSS–III has reached a significantly higher QSO density of ~ 15 deg^{-2}, with 175000 QSOs in the redshift range $2.1 < z < 3.5$, making it possible to correlate the Lyα absorption skewers in three dimensions and enabling the study of large–scale structure at high redshifts. The three–dimensional correlation function of the Lyα forest was first measured in (Slosar et al., 2011). Using COSMOS Lyman-Alpha Mapping And Tomography Observations (CLAMATO survey), it was possible to create density maps of the IGM at Mpc resolutions for the first time (Lee et al., 2014a). For a study about the exposure time necessary in order to reach certain map resolutions, the reader is referred to (Lee et al., 2014b).

1.3.2.1 BOSS and eBOSS

The successful detection of the baryon acoustic oscillations imprint in luminous red galaxies from SDSS data (Eisenstein et al., 2005) paved the way for an upgraded version of the original spectrographs in order to improve on the previous results. Shortly after the end of SDSS–II phase, in 2009, the capabilities of the SDSS equipment were enhanced with new CCDs (charged–coupled device detectors) having smaller pixels (15 μm), less noise, better quantum efficiency and 500 fibers feeding the two spectrographs. The wavelength coverage was also extended slightly to 3560 Å $< \lambda <$ 10400 Å. Until this change, the SEGUE–2 survey used the original spectrographs within the first year of SDSS–III.

The BOSS spectrographs have surveyed an additional 1,372,737 galaxies (Anderson et al., 2014; Sánchez et al., 2017), 294,512 quasars (Pâris et al., 2016) and 247,216 stars (Albareti et al., 2016). APOGEE has observed 156,393 high resolution IR spectra and MARVELS contributed with 3233 stars with radial velocity measurements. For purposes of large scale structure studies, which is the focus of this thesis, the highly increased BOSS sky quasar density enabled the three dimensional interpolation of Lyα skewers and made it possible to create the IGM maps we present here.

The fourth phase of the Sloan Digital Sky Survey (SDSS–IV) is made possible due to the success and significance of the previous phases (Blanton et al., in preparation). The fourth phase includes surveys spanning new redshift intervals, new galaxies, and the parts of the Milky Way and dwarf galaxies that are only observable from the Southern Hemisphere (Figure 1.9).

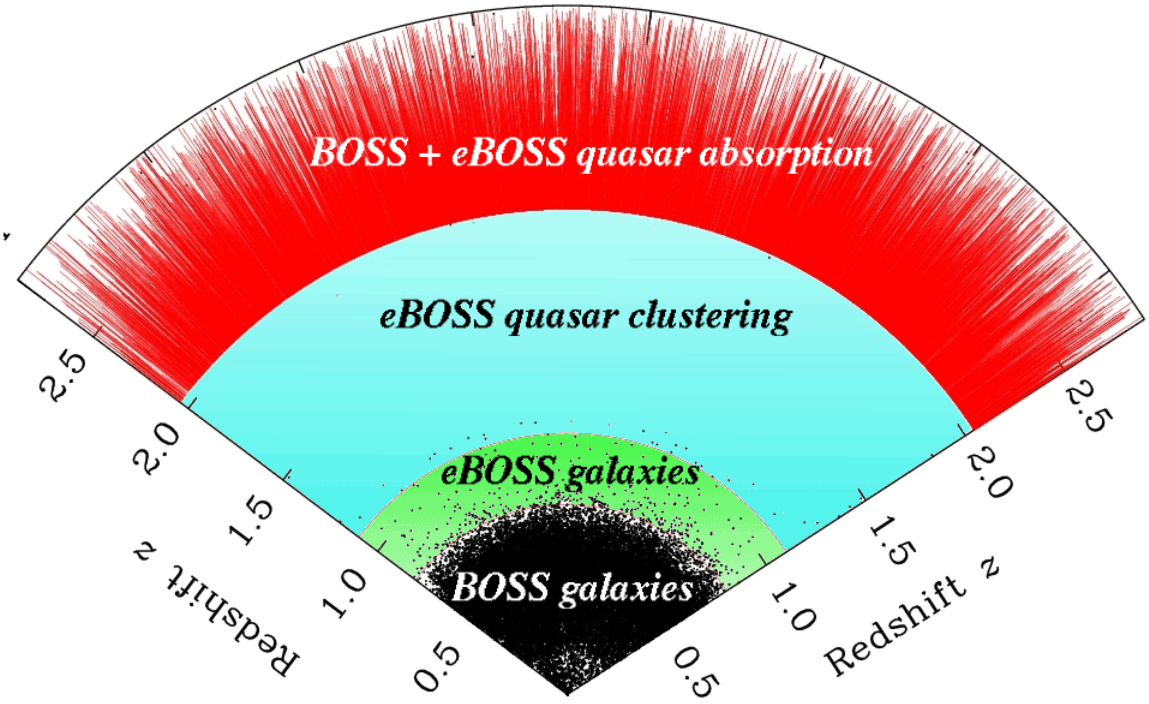


Figure 1.9: Redshift coverage of BOSS and eBOSS large scale structure with galaxies and the Lyα forest (figure obtained from the SDSS eBOSS web page: www.sdss.org/surveys/eboss).

The main purpose of Extended Baryon Oscillation Spectroscopic Survey (eBOSS) (Dawson et al., 2016), one of the main surveys of SDSS–IV, is to improve the BAO measurement obtained by BOSS in the clustering of matter within a relatively larger redshift range $0.6 < z < 2.2$. Spectroscopic redshifts of more than 400,000 LRGs and about 200,000 Emission–Line Galaxies (ELGs) are going to be targeted to provide two novel ways of BAO clustering measurements over the interval $0.6 < z < 1.1$.

In total, 500,000 quasars will be observed between $0.9 < z < 2.2$. As an addition to the 180,000 Lyα QSOs BOSS has already observed, eBOSS will contribute with 120,000 new QSO spectra. The new Lyα forest sample will allow more precise measurements, improving BOSS constraints by a factor of 1.4. Furthermore, new redshift space distortion measurements will be possible, as well as non–Gaussianity in the primordial field and the total mass of neutrino species will be examined. The future of quasar surveys is going to allow almost doubling of the number of quasars known to date: a quasar sky density of $\sim 45\,\mathrm{deg}^{-2}$ is proposed for the the Mid–Scale Dark Energy Spectroscopic Instrument (MS–DESI, see Levi et al. (2013)), launching in 2018 with the 4m Mayall telescope, covering 14000 square degrees of the sky.

1.4 Reconstructing the flux field

As discussed in the previous chapter, the increasing density of quasars makes it possible to interpolate the Lyα skewers in three dimensions. In principle, different filtering methods can be considered to make big maps of the Universe at high redshifts. In this thesis, we employ Wiener interpolation for this task, although other methods are also possible, such as local polynomial smoothing (Cisewski et al., 2014).

1.4.1 Wiener Interpolation

Norbert Wiener introduced the idea of the Wiener filter in the 1940s with his work on interpolation, extrapolation and smoothing of stationary time series (Wiener, 1949). The Wiener theory forms the foundation of interpolation methods that minimize the mean squared error between a filter output and the desired signal at the same location. The noise is assumed to be additive.

The Wiener filter offers great flexibility for different kinds of data sets. The data can be continuous or discrete and arbitrarily spaced. By adjusting the correlation function, the filter can be made local or as global as desired. In physics applications, weights of data points located further than a certain smoothing length scale are often negligible, e.g. for a Gaussian filter. The interpolation can also be oscillatory, negative at certain distances, and can be freely adjusted as desired. However, there are two main drawbacks of using this filter. The most important one is that the filtering process requires matrix inversions with dimensions having the same as the data. For large data sets, this can also be computationally expensive. As a remedy, the interpolation can be parallelized if the interpolation is sufficiently local. Also, smooth correlation functions can result in singular processes, making matrix inversion a different task. The other disadvantage of this method is the necessity to know, or correctly estimate the correlation function of the data.

The filter estimates the signal at a given location by averaging the values of nearby points, with the weighting given by the coefficient vector **w**. The output signal is computed as

$$\begin{aligned}\hat{x}(m) &= \sum_{k=0}^{P-1} w_k y(m-k) \\ &= \mathbf{w^T y} \end{aligned} \tag{1.9}$$

where m is the discrete index, $y^T = [y(m), y(m-1), ..., y(m-P-1)]$ is the filter input signal and the Wiener filter coefficient vector is $w^{\mathbf{T}} = [w_0, w_1, ..., w_{P-1}]$ (Vaseghi, 2000). The Wiener filter error signal $e(m)$ is the difference between the desired signal $x(m)$ and the filter output signal $\hat{x}(m)$:

$$\begin{aligned}e(m) &= x(m) - \hat{x}(m) \\ &= x(m) - \mathbf{w^T y}\end{aligned} \tag{1.10}$$

In vector form, Equation 1.10 can be expressed as

$$\begin{pmatrix} e_0 \\ e_1 \\ e_2 \\ ... \\ e_{N-1} \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ ... \\ x_{N-1} \end{pmatrix} - \begin{pmatrix} y_0 & y_{-1} & y_{-2} & ... & y_{1-P} \\ y_1 & y_0 & y_{-1} & ... & y_{2-P} \\ y_2 & y_1 & y_0 & ... & y_{3-P} \\ ... & ... & ... & ... & ... \\ y_{N-1} & y_{N-2} & y_{N-3} & ... & y_{N-P} \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ ... \\ w_{P-1} \end{pmatrix} \tag{1.11}$$

$$\mathbf{e} = \mathbf{x} - \mathbf{Yw} \tag{1.12}$$

where **e** is the error vector, **x** is the output signal, and $\mathbf{Yw} = \hat{\mathbf{x}}$ is the input. Here, the number of signal samples N must be greater than the filter size P for obtaining a unique solution. If not, the matrix equation in Equation 1.11 is said to be underdetermined, with an infinite number of solutions with zero estimation error. When $N > P$, the matrix equation is overdetermined with a unique solution, which usually yields non–zero error. Under this condition, Wiener filter coefficients are calculated in such a way that the mean squared error is minimized. From Equation 1.10, the mean squared error is given by

$$\begin{aligned}E[e^2(m)] &= E[(x(m) - \mathbf{w^T y})^2] \\ &= E[x^2(m)] - 2\mathbf{w^T} E[\mathbf{y}x(m)] + \mathbf{w^T} E[\mathbf{yy^T}]\mathbf{w} \\ &= r_{xx}(0) - 2\mathbf{w^T r_{yx}} + \mathbf{w^T R_{yy} w}\end{aligned} \tag{1.13}$$

where E is the expectation operator, $\mathbf{r_{xy}} = E[x(m)\mathbf{y}(m)]$ is the cross–correlation vector of the input and the desired signals and $\mathbf{R_{yy}} = E[\mathbf{y}(m)\mathbf{y^T}(m)]$ is the auto–correlation matrix of the input signal. Equation 1.13 is a function of the filter coefficient vector $\mathbf{w}$ and the only extremum is the minimum point. Taking the derivative of this equation yields

$$\begin{aligned}\frac{\partial}{\partial \mathbf{w}} E[e^2(m)] &= -2E[x(m)\mathbf{y}(m)] + 2\mathbf{w^T}E[\mathbf{y}(m)\mathbf{y^T(m)}] \\ &= -2\mathbf{r_{yx}} + 2\mathbf{w^T}\mathbf{R_{yy}} \end{aligned} \tag{1.14}$$

where the gradient vector is defined as

$$\frac{\partial}{\partial \mathbf{w}} = \left[\frac{\partial}{\partial w_0}, \frac{\partial}{\partial w_1}, \frac{\partial}{\partial w_2}, ..., \frac{\partial}{\partial w_{P-1}}\right]^T \tag{1.15}$$

Setting Equation 1.15 to zero gives the Wiener filter coefficient vectors, minimizing the mean squared error:

$$\mathbf{w} = \mathbf{R_{yy}^{-1}}\,\mathbf{r_{yx}} \tag{1.16}$$

Alternatively, to calculate $\mathbf{w}$, one can also use the fact that the squared error is minimized when the expectation of the product of the error and the data is zero, due to orthogonality. Combining Equations 1.9 and 1.16, the desired signal $\hat{x}(m)$ with the Wiener filter, in terms of the input y is then

$$\begin{aligned}\hat{x}(m) &= [\mathbf{R_{yy}^{-1}}\,\mathbf{r_{yx}}]^T y \\ &= \mathbf{r_{yx}^T}\,[\mathbf{R_{yy}^{-1}}]^T y \end{aligned} \tag{1.17}$$

If the input y can be expressed with a signal component and additive random noise, the Wiener filter can be trivially optimized for noise reduction. In view of the $\mathbf{R_{yy}}$ term in Equation 1.17, it is necessary to know, or estimate how the input data is correlated. The cross–correlation vector of the input and the output, $\mathbf{r_{xy}}$, also needs to be known. For the case where the noise in the input signal is uncorrelated ($\mathbf{R_{yy}} = \mathbf{R_{y'y'}} + \mathbf{R_{nn}}$), Equation 1.17 becomes

$$\hat{x}(m) = \mathbf{r_{yx}^T}\,[(\mathbf{R_{y'y'}} + \mathbf{R_{nn}})^{-1}]^T y \tag{1.18}$$

where $\mathbf{R_{nn}}$ is the auto–correlation matrix of the noise and the primed y variable denotes the input without noise.

1.4.2 Inverting the flux field

In this thesis, the optical depth (τ) data along lines of sight from SDSS DR12 (and the simulations) is the main observable. As we will describe in §2, we choose to work with flux ($e^{-\tau}$) directly to study the large scale structure. Therefore, it should be established that this is equivalent to making maps of the actual IGM density field.

The relation between the optical depth $\tau_\ell(w)$ along the LOS ℓ at position $\mathbf{x}_{\perp,\ell} = (y_\ell, z_\ell)$ and the neutral hydrogen density n_{HI} in velocity space w is given by

$$\tau_\ell(w) = \frac{c\sigma_0}{H(\bar{z})\sqrt{\pi}} \int\int \left(\int_{-\infty}^{\infty} \frac{n_{HI}(x, \mathbf{x}_\perp)}{b(x, \mathbf{x}_\perp)} exp - \left\{ - \frac{[w - x - v_p(x, \mathbf{x}_\perp)]^2}{b(x, \mathbf{x}_\perp)^2} \right\} dx \right) \delta_D(\mathbf{x}_\perp - \mathbf{x}_{\perp,\ell}) d^2\mathbf{x}_\perp \tag{1.19}$$

where σ_0 is the effective cross–section for resonant line scattering, $H(\bar{z})$ is the Hubble constant at mean redshift $\bar{z}$ and $v_p(x)$ is the projection of the peculiar velocity along the LOS (Pichon et al., 2001). The double integration over $\mathbf{x}_\perp$ is along the two directions perpendicular to the parallel lines of sight. δ_D is the two dimensional Dirac delta function. As τ is known, the task at hand is to invert Equation 1.19 to obtain n_{HI}.

We have already mentioned the relation between the neutral hydrogen density, the dark matter density and the temperature in Equation 1.8. Equations of state (Hui and Gnedin, 1997) are given by

$$T(\mathbf{x}) = \bar{T} \left(\frac{\rho_{DM}(\mathbf{x})}{\bar{\rho}_{DM}} \right)^{2\beta} \tag{1.20}$$

$$n_{HI}(\mathbf{x}) = \bar{n}_{HI} \left(\frac{\rho_{DM}(\mathbf{x})}{\bar{\rho}_{DM}} \right)^{\alpha} \tag{1.21}$$

$$b(\mathbf{x}) = 13\, kms^{-1} \sqrt{\frac{\bar{T}}{10^4 K}} \left(\frac{\rho_{DM}(\mathbf{x})}{\bar{\rho}_{DM}} \right)^{\beta} \tag{1.22}$$

where the parameter β is in the range $0 < \beta < 0.31$ (the upper bound comes from the asymptotic value at $z = 0$) and the dark matter scaling parameter is $\alpha = 2 - 1.4\beta$. Variables with a horizontal bar are mean values along the LOS. The Doppler parameter $b(\mathbf{x})$ is a function of the local temperature of the IGM at a given point and x is the real space coordinate. For the length scales of interest in this work, the thermal broadening due to Equation 1.22 can be ignored. Hence, Equation 1.19 becomes

$$\tau_\ell(w) = A(\bar{z}) \int\int \left(\frac{\rho_{DM}[w - v_p(x(w, \mathbf{x}_\perp), \mathbf{x}_\perp)]}{\bar{\rho}_{DM}} \right)^{\alpha} \delta_D(\mathbf{x}_\perp - \mathbf{x}_{\perp,\ell}) d^2\mathbf{x}_\perp, \tag{1.23}$$

$$A(\bar{z}) = n_{\bar{HI}} \frac{c\sigma_0}{H(\bar{z})} \tag{1.24}$$

The equations above are evaluated for all LOS independently. Since we are working in redshift space, $v_p = 0$. Redshift distortion at large scales do not change the topology and statistics of the field significantly, at mildly non–linear scales. Therefore, to constrain the 3D field in Equation 1.23, appropriate values are chosen for α and β, thus deciding the equations of state.

In the original work where this inversion method was introduced, the authors tested it with N–body simulations (Figure 1.10) and shortly after the original work, with a $z = 2.4$ quasar spectrum from Ultraviolet Visual Echelle Spectrograph (UVES) to constrain parameters and study the small–scale structure (Rollinde et al., 2001).

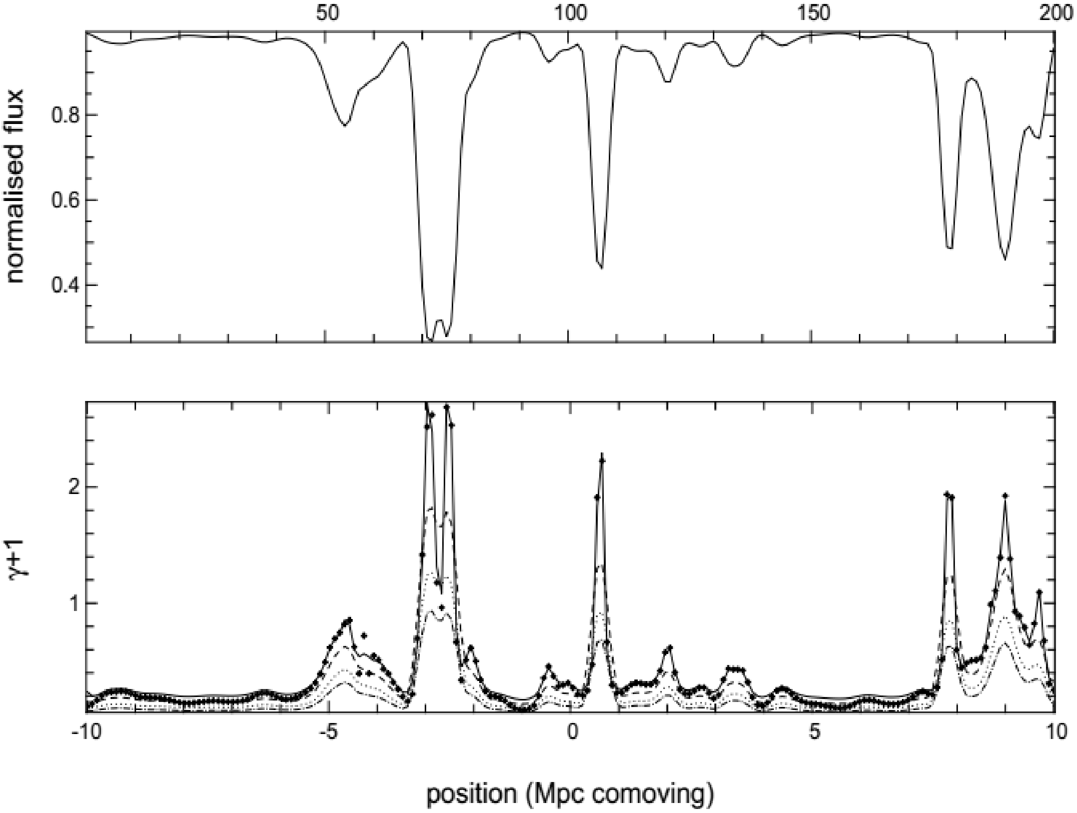


Figure 1.10: Different lines correspond to inversion of the Lyα forest with different equations of state. The top panel shows the flux along the simulation skewer. Black dots at the bottom panel correspond to the simulated density. Equation 1.20 is evaluated with $\bar{T} = 10^4 K$ and $\beta = 0.2$. Peculiar velocities are ignored. Other curves correspond to different $\bar{T}$ values at $\beta = 0.2$ (figure obtained from Pichon et al. (2001)).

1.5 Thesis Overview

1.5.1 Motivation

The Wiener interpolation method has proven to be a useful tool for reconstructing the large scale structure with sparse and noisy data (Zaroubi et al., 1994; Fisher et al., 1995; Pichon et al., 2001; Rollinde et al., 2001; Caucci et al., 2008).

The general idea for making 3D density maps in a big volume is to first carry out the inversion along LOSs to obtain the underlying density field, followed by interpolating between them with the Wiener filter. A usual scenario, showing the LOSs and the underlying density field for simulated fields is depicted in Figure 1.11, where although the methods are the same, the authors were able to introduce additional constraints from the Lyman-β forest (Petitjean et al., 2001). It is worth noting that for simulations, LOSs are parallel to each other, resulting in a uniform spatial resolution for the resulting maps. However, the fact that the quasar distribution in observational surveys is heavily dependent on redshift introduces the necessity for adjustments to the method. In principle, only volumes with a sufficient LOS density should be chosen for reconstructing the IGM field. This can also be seen from the condition $N > P$ in Equation 1.11, which if violated, can potentially lead to singular processes during matrix inversion. A natural choice for the redshift interval for making maps, given the quasar distribution (Figure 1.6), is $2 < z < 2.5$.

In this thesis, we improve on the previous work, first with a cosmological hydrodynamical simulation, using a larger box size and sparser sightline densities which mimic those of observational surveys, in order to evaluate the performance of our methods. Then, we apply the same procedure to SDSS–III BOSS DR12 data to make big observational maps of the IGM at $2 < z < 2.5$ and examine global and local statistical properties of the field. 3D maps of the IGM at such large scales at high redshifts have never been made before, making our work unprecedented.

1.5.2 Thesis Plan

In the next section, we describe our peer reviewed and published work with recovering the simulated fields and studying statistics of the field, which is based on (Ozbek et al., 2016). Observational requirements for obtaining certain map resolutions, as well as an alternative local polynomial interpolation technique are discussed in §3. In §4, we apply the Wiener interpolation method to observational data from SDSS-III BOSS QSO spectra to make big maps of the Universe at high redshifts. We focus on local statistics in §5, discussing percolation techniques to find superclusters in the map. Finally, in §6, we end our work with a summary and concluding remarks.

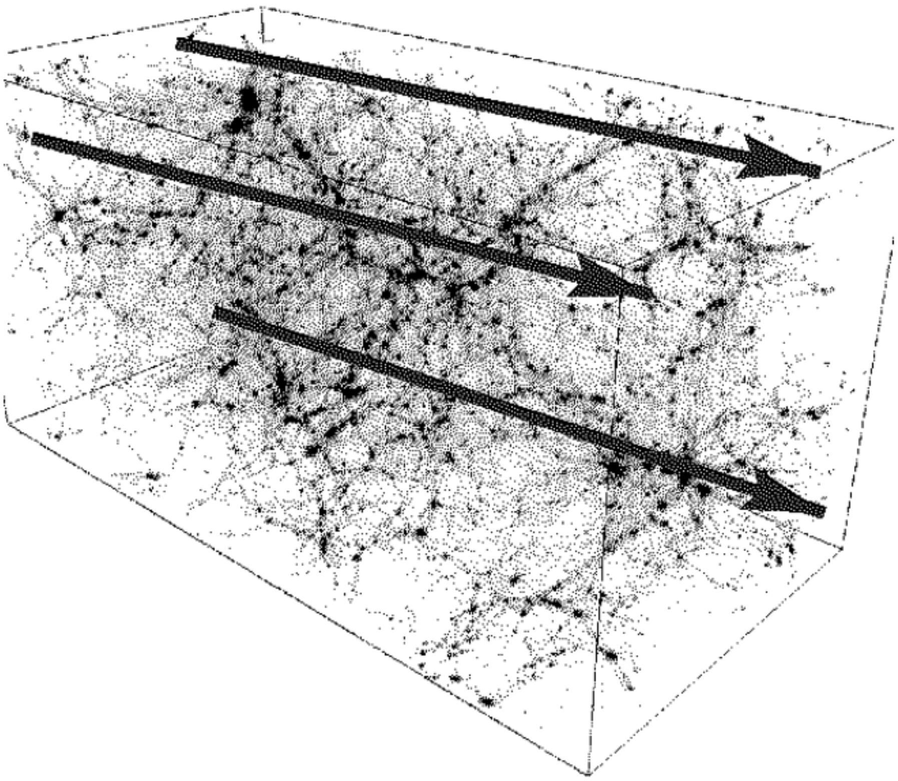


Figure 1.11: This figure shows a general picture of the LOS distribution and the underlying density field (figure obtained from Petitjean et al. (2001)). Darker colors show areas that are denser.

Chapter 2

3D Mapping of the Intergalactic Medium with Simulations

The work in this chapter has been published as Ozbek, Croft and Khandai (2016).

2.1 Introduction

In order to evaluate the expected performance of map-making reconstruction on Lyman-α forest data from BOSS and other observational surveys, we make use of a large hydrodynamic cosmological simulation of the ΛCDM model. We use the smoothed particle hydrodynamics code P–GADGET (see Springel 2005; Di Matteo et al. 2012) to evolve a distribution of $2 \times 4096^2 = 137$ billion particles in a cubical periodic volume of side length 400 h^{-1}Mpc. The simulation cosmological parameters were $h = 0.702$, $\Omega_\Lambda = 0.725$, $\Omega_m = 0.275$, $\Omega_b = 0.046$, $n_s = 0.968$ and $\sigma_8 = 0.82$. The mass per particle was 1.19×10^7 $h^{-1}M_\odot$ (gas) and 5.92×10^7 $h^{-1}M_\odot$ (dark matter). A gravitational force resolution of 3.25 kpc/h comoving was used. The power spectrum of the simulation initial conditions was taken from CAMB (Lewis et al., 2000). The simulation was run with an ultraviolet background radiation field consistent with Haardt and Madau (1995). Cooling and star formation were included. However the latter used a lower density threshold than usual (for example in Springel and Hernquist 2002) so that gas particles are rapidly converted to collisionless gas particles. This was done to speed up execution of the simulation. As a result the stellar properties of galaxies in the simulation are not predicted reliably but this has no significant effect on the diffuse IGM that gives rise to the Lyman-α forest. Black hole formation and feedback from stars were also switched off in the simulation.

2.1.1 Data

We use two simulation snapshots at redshifts of $z = 2$ and $z = 3$ to generate two sets of Lyman-α spectra using information from the particle distribution (Hernquist

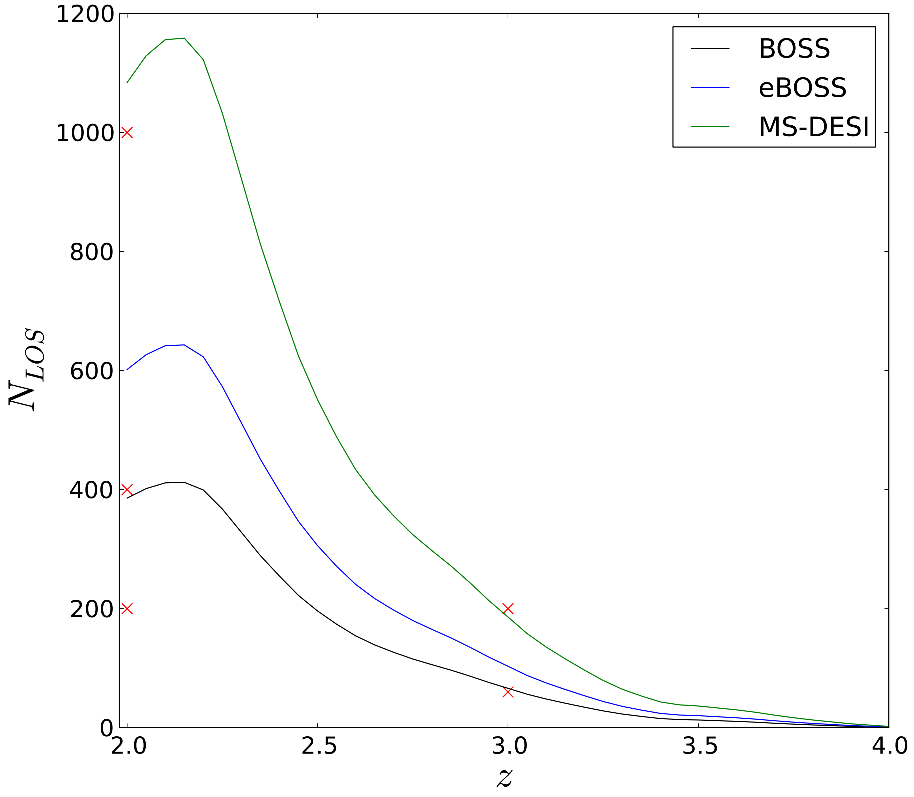


Figure 2.1: N_{LOS} passing through our simulation volume at different redshifts according to BOSS, eBOSS and MS–DESI (assuming the other two experiments have the same distribution of QSOs with respect to redshift as BOSS). The red markers show the fiducial choices for our work with simulation data. The BOSS quasar catalog indicates that the number of quasars peaks at z $\sim$ 2.25 and decreases rapidly at higher redshifts.

	BOSS	**eBOSS**	**MS–DESI**
$n_{\rm LOS}/\deg^2$	16	25	45
Sky coverage (deg^2)	10400	7500	14000
Lyα QSOs (thousands)	180	250	1000
Spectral Resolution	~ 2000	~ 2000	~ 3500

Table 2.1: LOS density, sky coverage, targeted Lyman-α QSOs and spectral resolution parameters for three sky surveys

et al., 1996). We make spectra set out on a grid with $176^2 = 30,976$ evenly spaced sightlines, resulting in 2.27 h^{-1}Mpc spacing. This can be compared in the line of sight direction with BOSS pixels of width $\Delta v = 69.02$ km s^{-1} (Lee et al., 2013), which is $\sim$ 0.6 h^{-1}Mpc at $z = 3$. Each simulation sightline was generated with high resolution, 10,560 pixels, in order to resolve the thermal broadening when computing the optical depth. The spectra were then downsampled (by averaging the transmitted flux over 60 pixels) to 176 pixels. The full set of simulation data sets therefore consist of 176^3 data values each.

In order to roughly approximate the noise which will be expected in observational data, we add random uncorrelated Gaussian pixel noise to the data sets with a signal to noise ratio S/N = 1 or 2 per unit simulation pixel in 176 pixels per 400 h^{-1}Mpc sightline. This is similar to BOSS which has S/N of order unity (Lee et al., 2015). If BOSS Lyman-α data were binned to 9.3 h^{-1}Mpc pixels, the mean S/N ratio would be 2.5. Similarly, the simulation data with S/N = 1, when binned to pixels of the same size results in an S/N ratio of 4. Therefore, we use the S/N = 1 case as a close match for the BOSS noise level, whereas the other noise level, S/N = 2, is given as an example with less noise. A random subset of the 176^2 sightlines was chosen, according to LOS area densities from the experiments BOSS (Alam et al., 2015), eBOSS (Raichoor et al., 2015) and MS–DESI (Levi et al., 2013) from our simulation box to carry out the reconstruction (see Table 2.1 and Figure 2.1). For example, at $z = 2$, $N_{\rm LOS} \sim 400$ LOS passing through the simulation volume mimics the LOS density for BOSS, while $N_{\rm LOS} \sim 1000$ mimics that of MS–DESI. Since most of our conclusions are inferred from data sets with $N_{\rm LOS} = 200$, one should expect observational results with BOSS data to be of even higher fidelity. We define $N_{\rm LOS}$ as the number of lines of sight chosen to reconstruct the entire volume in the simulation box, whereas $n_{\rm LOS}$ (see, e.g., Table 2.1) denotes the total number of sightlines along the entire redshift range from the observer to quasars.

Our data points derived from the simulation are optical depths, ($\tau = -\log_e F/F_0$), where F is the flux received at a certain location in space and F_0 is the unabsorbed flux. The data are convolved with the peculiar velocity, therefore we work in redshift space. In (Caucci et al., 2008) the authors worked with the density field directly. We will however work with the flux, and our maps will be reconstructions of the three

Sample	**Redshift**	$\mathbf{N}_{\mathrm{LOS}}$	**Noise**	$\langle \mathbf{d}_{\mathrm{LOS}} \rangle (h^{-1}\mathrm{Mpc})$
z2_N200	2	200	Noiseless	28.28
z2_N200_SN2	2	200	S/N=2	28.28
z2_N200_SN1	2	200	S/N=1	28.28
z2_N400	2	400	Noiseless	20.00
z2_N400_SN2	2	400	S/N=2	20.00
z2_N400_SN1	2	400	S/N=1	20.00
z2_N1000	2	1000	Noiseless	12.65
z2_N1000_SN2	2	1000	S/N=2	12.65
z2_N1000_SN1	2	1000	S/N=1	12.65
z3_N60	3	60	Noiseless	51.64
z3_N200	3	200	Noiseless	28.28
z3_N200_SN2	3	200	S/N=2	28.28

Table 2.2: Our choices of simulated data sets at redshifts 2 and 3 with different LOS density and noise levels.

dimensional flux field, in redshift space. The relation between the gas density and the optical depth is

$$\delta(x) = \frac{1}{\alpha} \log(\frac{\tau(x)}{A(\bar{z})}) \tag{2.1}$$

where $\delta(x) \approx \frac{\rho-\bar{\rho}}{\bar{\rho}}$ is the density contrast, and α and $A(\bar{z})$ are redshift dependent factors. We present our results in terms of flux contrast, $\delta_F = (F/\langle F \rangle) - 1$, where $\langle F \rangle$ is the mean transmitted flux computed from all spectra.

2.1.2 Simulated datasets

We have made 12 simulated data sets with different sightline densities and noise levels. These are summarised in Table 2.2. Sightline density choices were made to mimic those of current or future observational surveys, as shown in Figure 2.1 with red markers. Some data sets have a LOS density that is even lower than that of BOSS (e.g. the data set labelled "z2_N200") but still allow an accurate recovery of the flux field, as we will see in §2.2. Data sets with higher sightline densities (e.g. "z2_N1000", which is comparable to that of MS-DESI) result in an even better inference of the field. Therefore, as observational surveys find more quasars, even more accurate density maps will be available with the Lyα forest.

2.1.3 Wiener Interpolation

There are several methods which can be used to interpolate between the sparse absorption skewers in the Lyman-α forest. For example, recent work by Cisewski et al. (2014) used local polynomial smoothing for this purpose. The method we choose in this thesis is Wiener filtering, pioneered in this context by Pichon et al. (2001), and used by Caucci et al. (2008), and Lee et al. (2014b) to make maps from simulated data, and by Lee et al. (2014a) to make the first 3 dimensional maps from observations.

We consider the values of the flux contrast in the reconstructed field to be entries in a column vector $\mathbf{M}$, and the values of the flux contrast in the absorption skewer data to be entries in a column vector $\mathbf{D}$. In general the entries of $\mathbf{M}$ will represent values on a uniform grid of voxels as we are constructing a map which covers all space within the map boundary. In our simulation tests, they will be covering the cubical simulation volume uniformly. We choose not to make use of the simulation periodic boundary conditions, in order to mimic some aspects of real data. Using Wiener filtering, the reconstructed 3D field $\mathbf{M}$ can be inferred from the absorption skewer data $\mathbf{D}$ by computing

$$\mathbf{M} = \mathbf{C_{MD}} \cdot (\mathbf{C_{DD}} + \mathbf{N})^{-\mathbf{1}} \cdot \mathbf{D} \tag{2.2}$$

where $\mathbf{C_{MD}}$ and $\mathbf{C_{DD}}$ are the map–data and data–data covariance matrices and $\mathbf{N}$ is the diagonal noise matrix. In the present work, we assume the noise to be uncorrelated, so that the entries of $\mathbf{N}$ are inversely proportional to the square root of the number of pixels in each cell. The covariance matrices encode the expected correlation structure of the field. In most of our work we use the following simple form advocated by Pichon et al. (2001) and Caucci et al. (2008):

$$\begin{aligned} \mathbf{C}(x_1, x_2, \mathbf{x_{1\perp}}, \mathbf{x_{2\perp}}) = \sigma^2 \times \exp\Big(- \frac{(x_1 - x_2)^2}{L_{||}^2}\Big) \times \\ \exp\Big(- \frac{|\mathbf{x_{1\perp}} - \mathbf{x_{2\perp}}|^2}{L_\perp^2}\Big) \end{aligned} \tag{2.3}$$

where $(\boldsymbol{x_1} - \boldsymbol{x_2})$ and $|\mathbf{x_{1\perp}} - \mathbf{x_{2\perp}}|$ represent the distances between two pixels, parallel and perpendicular to the LOSs respectively, $\boldsymbol{L_{||}}$ and $\boldsymbol{L_\perp}$ are correlation lengths parallel and perpendicular to the LOSs, while the variance $\boldsymbol{\sigma^2}$ is calculated directly from the field. The $\mathbf{C_{DD}}$ covariance matrix contains correlation information between the initial data points only (the $\mathbf{D}$ array), whereas $\mathbf{C_{MD}}$ contains information of the 3D pixel locations of the map to be inferred and the $\mathbf{D}$ array. The similarity between Equations 1.18 and 2.2 are obvious. As previously mentioned in §1, the covariance of the input data $\mathbf{C_{DD}}$ needs to be known.

In order to test how well the reconstruction works as a function of line of sight density, we make several different data samples by choosing a subset of our lines of sight at random. The areal density of the sightlines are those that correspond to some current and planned experiments, e.g. BOSS (Dawson et al., 2013) and DESI (Levi et al., 2013) (see Table 2.1 and Figure 2.1).

Our numerical code to carry out the reconstruction splits the simulation volume into "subcubes". The interpolation is then carried out separately for each subcube in parallel with the others and in the final step the results are combined to form the whole reconstructed simulation cube. In order to make the calculations more computationally tractable, we decrease the resolution of the field from 176^3 to 44^3 pixels.

We introduce a buffer volume on the edges of the subcubes, allowing them to overlap, in order to avoid edge artifacts. In our fiducial reconstruction of the simulation we use 64 subcubes overall and a buffer of 40 h^{-1}Mpc on each side for each subcube. Each subcube therefore has a side length of 180 h^{-1}Mpc, including the buffer regions. We have tested and checked that adjusting the number of subcubes or changing the number of pixels does not significantly alter the results.

The code used does not take into account the periodic boundary conditions of the whole simulation box, in order to approximate the situation which will occur for observational data. This means that the reconstruction will be less accurate near the edges of the cubical simulation volume. For this reason, when choosing slice images to compare real and reconstructed fields, we choose the middle planes of the cube rather than the edges. We find by visual inspection that there is no significant difference in the quality of reconstructions when increasing the separation from the box edge by greater than 50 h^{-1}Mpc. We repeat some statistical calculations after truncating the cube by 50 h^{-1}Mpc from each edge in order to test the importance of edge artifacts.

The resolution of the maps is determined by the mean separation between quasar lines of sight:

$$\langle d_{\rm LOS} \rangle = \frac{L_{Box}}{\sqrt{N_{\rm LOS}}} \tag{2.4}$$

For $N_{\rm LOS} = 200$, $\langle d_{\rm LOS} \rangle$ is equal to 28.28 h^{-1}Mpc. For a study of the map resolution as a function of exposure time, the reader is referred to Lee et al. (2014b). Pichon et al. (2001) have shown that typical values for the correlation lengths $L_{||}$ and $L_{\perp}$ should be of the order of $\langle d_{\rm LOS} \rangle$ in order to avoid numerical instabilities in the matrix inversion leading to spurious structures. We smooth both true and reconstructed fields with an isotropic Gaussian filter with a standard deviation $\sigma_S = 1.4\langle d_{\rm LOS} \rangle$, in the latter case after carrying out the reconstruction.

2.2 Analysis

After the reconstruction of the field we analyze the relationship between the true and reconstructed fields, bearing in mind that both have been smoothed, as stated above.

2.2.1 Scatter Plots

We first show point by point scatter plots in Figure 2.2. We plot the reconstructed flux contrast, δ_{recon} against the true flux contrast, δ_{orig}. The results of the point to point comparison of the fields are summarised in Table 2.3.

Throughout the thesis, we use "original field" (or "true field") with the meaning that we keep all of the Lyman-α skewers in the cube, while "recovered field" or "reconstructed field" means the flux field inferred with the given LOS density with the quasars located at redshift 2 or 3. In the top left panel of Figure 2.2, we show the results for the the $N_{\rm LOS} = 200$ dataset with no noise. We can see the slope of the relation between δ_{recon} and δ_{orig} is biased (this was also found by Lee et al. 2014a), in the sense that the recovered field has more contrast than the original field. After fitting a linear regression we find that the slope is 1.73, whereas the y–intercept is consistent with zero. This bias depends on the interplay between the Wiener filter smoothing scales and fluctuations in the field that are missed in the sparse sampling. The bias is larger when the number density of sightlines is low (compare the top left panel of Figure 2.2 which has a fitted slope of 1.73 and a sightline density 5 times less than the bottom left panel, which has a fitted slope of 1.35). Any correction for this bias is likely to be empirical, and therefore in the rest of our analysis we apply the simplest correction, by renormalizing the δ_{recon} according to the slope of the regression.

In order to quantitatively test the quality of the reconstruction, we compute the error by calculating the ratio of the root mean square (RMS) of the pixel by pixel difference to the RMS of the true field (δ_{orig}) which only includes 95 per cent of the true pixels ($\pm 2\sigma$ from the mean, which is normalized to zero), therefore avoiding outlier points:

$$e_{\%} = 100 \frac{\sqrt{\sum (\delta_{\rm orig} - \delta_{\rm recon})^2}}{4\sqrt{\sum \delta_{\rm orig}^2}} \tag{2.5}$$

From Figure 2.2, we can see from the top row of panels that the addition of noise to the input field does affect the reconstruction. The RMS error (after bias correction) is 20.4 per cent for the noiseless case and 37.0 per cent and 57.3 per cent for the cases with S/N=2 and S/N=1 respectively in the pixels of size 9.30 h^{-1}Mpc used in our analysis. We remind the reader that the BOSS Lyα forest data when rebinned in this way has a mean S/N ratio of 2.5. Therefore, we estimate that the RMS error for the reconstructed BOSS data will be 36.7 per cent.

Sample	**RMS**($\delta_{\mathbf{orig}}$)	**RMS**($\delta_{\mathbf{recon}}$)	**RMS**($\delta_{\mathbf{diff}}$)	% **Error**	**Pearson Coefficient (r)**	$\sigma_{\mathbf{S}}(h^{-1}\mathrm{Mpc})$
z2_N200	0.00817	0.0104	0.00666	20.4	0.783	39.6
z2_N200_SN2	0.00809	0.0139	0.0120	37.0	0.578	39.6
z2_N200_SN1	0.00805	0.0195	0.0185	57.3	0.412	39.6
z2_N400	0.0122	0.0154	0.00957	19.6	0.790	28.0
z2_N400_SN2	0.0121	0.0228	0.0199	41.0	0.530	28.0
z2_N400_SN1	0.0121	0.0452	0.0417	86.1	0.421	28.0
z2_N1000	0.0193	0.0234	0.0133	17.2	0.824	17.8
z2_N1000_SN2	0.0192	0.0342	0.0284	36.9	0.563	17.8
z2_N1000_SN1	0.0195	0.0364	0.0307	39.3	0.537	17.8
z3_N60	0.00432	0.00695	0.00599	34.7	0.622	72.3
z3_N200	0.0128	0.0187	0.0139	27.1	0.686	39.6
z3_N200_SN2	0.0127	0.0379	0.0370	72.8	0.334	39.6

Table 2.3: RMS values, percentage error, Pearson coefficient and the standard deviation for the isotropic Gaussian filter size for all samples. The reader is referred to Table 2.2 for the sample definitions.

We also provide the Pearson coefficient (r) as a measure of the linear correlation between δ_{orig} and δ_{recon}. Total positive correlation between the original field and the reconstruction would correspond to $r = 1$, while no correlation would be $r = 0$ and total negative correlation would be $r = -1$. At redshift $z = 2$ with $N_{\rm LOS} = 200$, $r = 0.783$, and it is even higher when the number of sightlines is increased, as expected.

Increasing the number of sightlines, as shown in the bottom left panel of Figure 2.2 allows the smoothing scale to be reduced and the resolution of finer features in the flux contrast field. The RMS flux contrast fluctuations increase to 0.0193 for this sample (z2_N1000, see Table 2.3) and the percentage error on the reconstruction stays approximately the same as the lower resolution reconstruction in the top left panel.

Finally we show results for the higher redshift, $z = 3$ in the bottom right panel of Figure 2.2. For the sample z3_N200, we use the same number of sightlines, as the top left ($z = 2$) panel, but the RMS accuracy of the reconstruction is lower by a factor of 0.75. The quality of the reconstruction is low for the sample z3_N60, as can be seen in Figure 2.11. However, the percentage error in Table 2.3 is comparable to those from other samples due to the fact that the dynamic range in the true field is low because of the much higher σ_S value for that sample. The flux contrast fluctuations are larger at the higher redshift, because of the greater overall optical depth in the Lyα forest, but this does not translate into a better reconstruction.

2.2.2 Cross Correlation

In the previous section we have seen how the true and reconstructed fields compare on a point–by–point basis for one specific value of the smoothing scale, given by $\sigma = \sqrt{2}\langle d_{\rm LOS}\rangle$. We can also study how similar the fields are at different scales. Instead of using different filter scales, we make use of the correlation functions of the field, measuring the correlation between points separated within a distance r. This translates to a probability which is in excess of a random distribution.

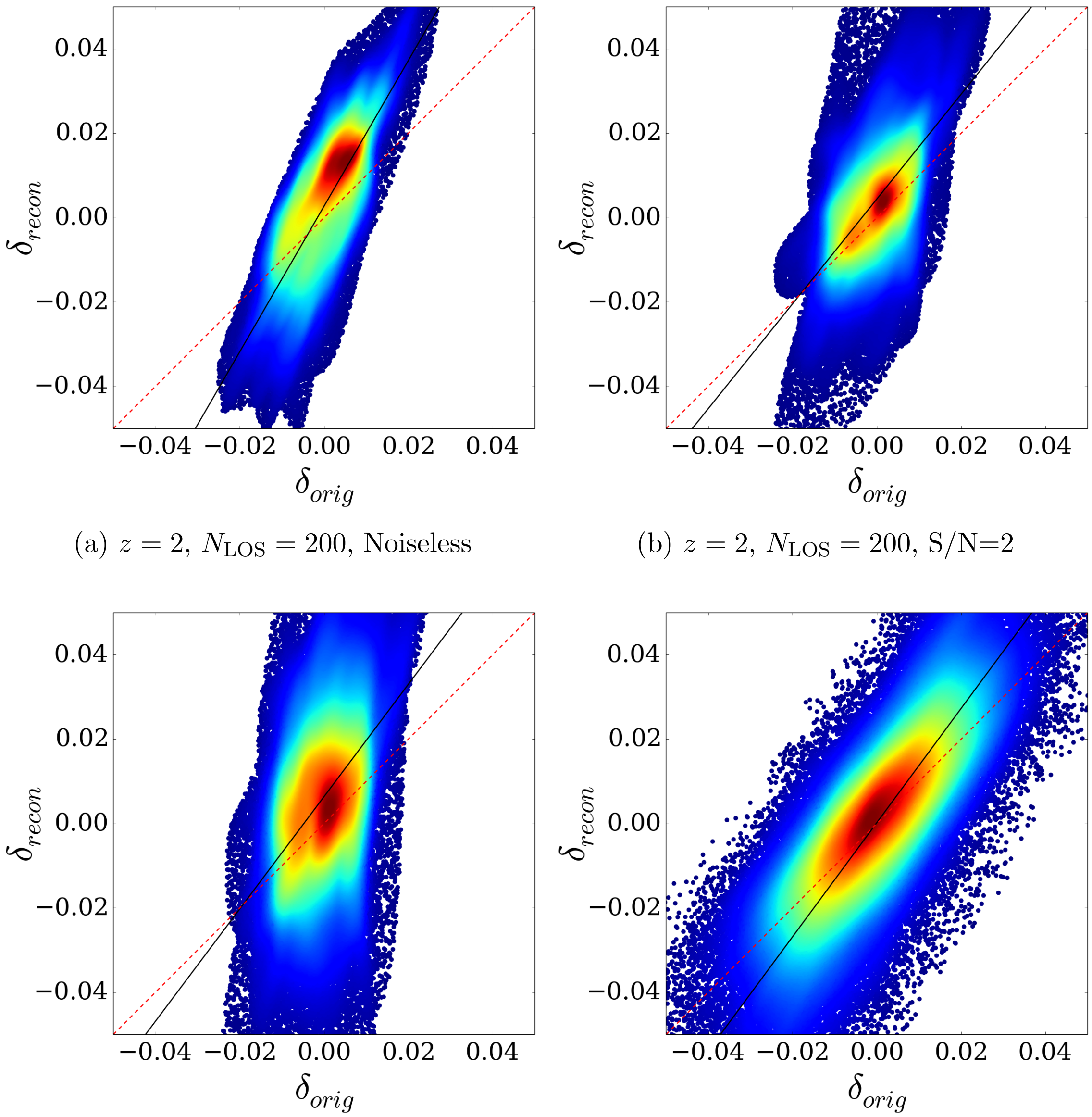
We compute the auto–correlation of the true field, the cross correlation and the standardized cross correlation between the true and the recovered fields. In general, the correlation function of two fields 1 and 2 is defined by:

$$\xi_{12}(r) = < \delta_1(x)\delta_2(x+r) > \tag{2.6}$$

This denotes the excess probability of finding pairs separated by a distance r. For the auto–correlation, 1 and 2 represent the same fields. Due to the periodic boundary conditions of the simulation cube, flux values are wrapped around over the edges when calculating correlations. In the top row of Figure 2.3, correlations tend to zero rapidly after the smoothing scale. The standardized cross–correlation is defined by:

$$C_{12}(r) = \xi_{12}(r)/\sqrt{\xi_{11}(r) \cdot \xi_{22}(r)}, \tag{2.7}$$

where ξ_{12} is the cross-correlation and ξ_{11} and ξ_{22} are the auto–correlations. C_{12} enables us to quantify the accuracy of the reconstruction as a function of scale, with

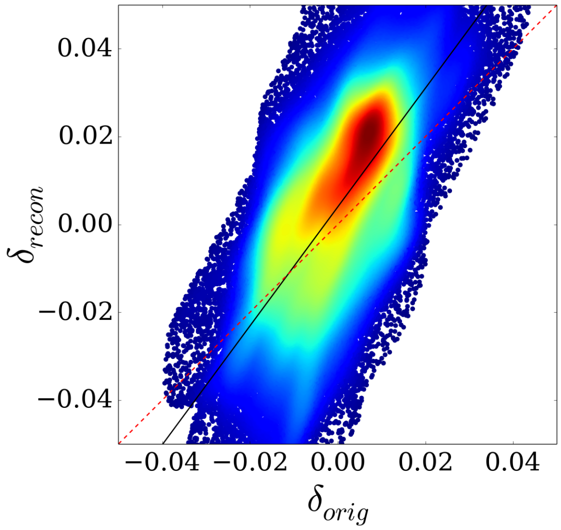


(a) $z = 2$, $N_{\text{LOS}} = 200$, Noiseless

(b) $z = 2$, $N_{\text{LOS}} = 200$, S/N=2

(c) $z = 2$, $N_{\text{LOS}} = 200$, S/N=1

(d) $z = 2$, $N_{\text{LOS}} = 1000$, Noiseless



(e) $z = 3$, $N_{\rm LOS} = 200$, Noiseless

Figure 2.2: Scatter plots of the true flux contrast (δ_{orig}) in the simulated maps compared to the reconstructed flux contrast (δ_{recon}). We show reconstructions at different redshifts, sightline densities and signal to noise ratios, as follows: The top row, which is from an analysis at $z = 2$ with $N_{\rm LOS} = 200$, demonstrates the effect of adding noise to our data and carrying out the reconstruction in order to mimic observational data. From left to right, the analyses for data which are noiseless, S/N=2 and S/N=1 are shown. The bottom left is from an analysis at $z = 2$ with $N_{\rm LOS} = 1000$. Bottom right plot shows our results at $z = 3$ with $N_{\rm LOS} = 200$. The black lines indicate linear regression fits and the red dashed lines show the $y = x$ line. colors show the density of the points, with red being the densest and blue denoting the most sparse. For each plot, the red area contains 68 per cent of the data points.

a value of unity indicating perfect fidelity. It should be noted that we do not expect good agreement for scales smaller than the fiducial smoothing length, the natural resolution of the map. According to Equation 2.4, the fields with $N_{\rm LOS} = 200$ are smoothed with $\sigma_S = 39.6\ h^{-1}$Mpc, while those with $N_{\rm LOS} = 1000$ are smoothed with $\sigma_S = 17.7\ h^{-1}$Mpc.

If we look at the case of low sampling density, $N_{\rm LOS} = 200$, we see that C_{12} can be as low as 0.45 on scales which are approximately 1.5–2 times the smoothing scale. Repeating the same analysis for unsmoothed fields with this LOS density does give better agreement at low scales – clearly, some structure is erased due to smoothing. For larger values of $N_{\rm LOS}$, on the other hand, we get good agreement despite the smoothing (Figure 2.3, bottom row). After smoothing both fields with the fiducial smoothing length, the recovery improves with distance until very large distances, where it decreases again. This decrease is due to edge effects, as on large scales (close to $\sim 200\ h^{-1}$Mpc), much of the volume is close to an edge. In panel (d) of Figure 2.3 we restrict the measurements of ξ_{12} to the volume of the simulation cube left after eliminating all regions within 50 h^{-1}Mpc of an edge, which alleviates the issue completely. Comparing with panel (c) of Figure 2.3, we can see that the agreement on small scales has also been improved, showing the positive effects of getting rid of the edge artifacts. In a large survey such as BOSS which spans a contiguous volume of several gigaparsecs, most of the volume will be much further from an edge than 50 h^{-1}Mpc, so that edge effects should be a small issue.

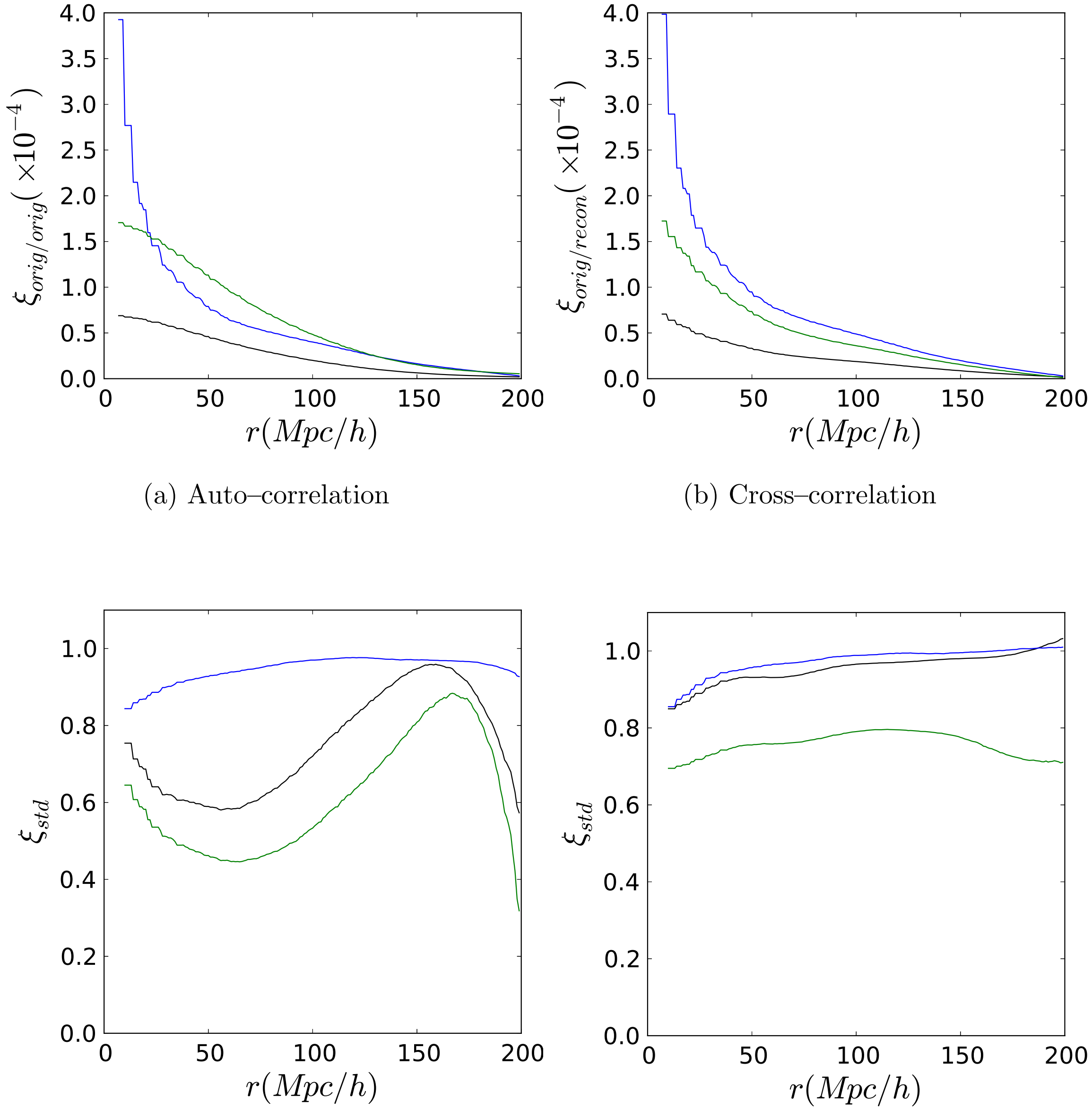
When the edge effects have been removed, we can see that the C_{12} measurement is close to 1 for all scales greater than the smoothing filter scale at $z = 2$, indicating essentially perfect statistical agreement. At redshift $z = 3$, C_{12} is never greater than 0.8, which may indicate that the less evolved structures at higher redshift make accurate reconstruction more difficult.

2.2.3 Non–Gaussianity

The density field probed by the Lyα forest is expected to be in the mildly non-linear regime. When smoothed on large scales, which we necessarily must do in order to construct our interpolated maps, we expect that the flux probability distribution should be quite close to Gaussian. Indeed this Gaussian assumption underlies the reconstruction carried out with the Wiener filter in Equation 2.3. It is therefore of interest to compare the reconstructed and true flux probability density functions with each other and with a normal distribution.

2.2.3.1 Probability Density Functions

Before we examine the distributions quantitatively with Kolmogorov–Smirnov tests, it is useful to study them visually (Figure 2.4) with probability density functions (PDFs). The solid grey line and the area under it shows the distribution of the original field, while the solid black line indicates the distribution of the recovered field. The red dotted line is a Gaussian fit, centered at zero and having the standard deviation which is equal to that of the original field. The areas under all curves are normalized to unity. In the top row, from left to right, the recovered field is recognizably more Gaussian as the LOS density is doubled. This behavior can also be seen in the Kolmogorov–Smirnov analysis (Table 2.4, rows 1 and 4). The reason for the horizontal spread being narrower in panel (a) relative to panel (b) is the difference in smoothing lengths: The size of the kernel for the data set with the lower LOS density ($N_{\rm LOS} = 200$) is 40 per cent greater than the other one (Table 2.2). Since the data is normalized to zero, more flux values are smoothed out with values closer to zero. On the other hand, for the data set at $z = 3$, not only is the recovered field markedly less Gaussian, but also, the spread in the flux contrast values is much greater than that of the original field. This suggests that the fidelity of the



(a) Auto–correlation

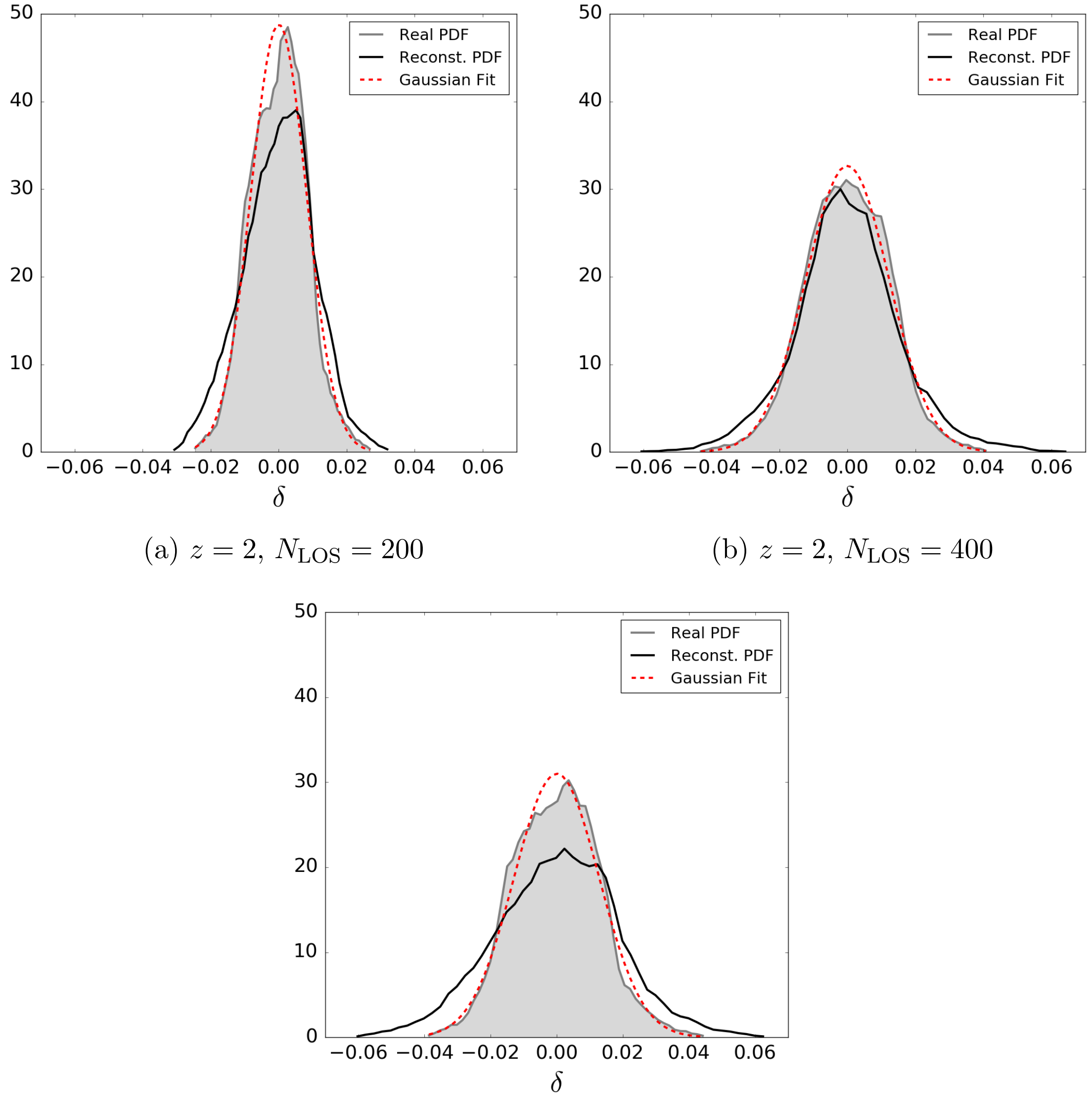
(b) Cross–correlation

(c) Standardized cross–correlation

(d) Standardized cross–correlation (truncated cube)

Figure 2.3: Correlation functions of the true and reconstructed fields as a function of scale. In each panel, the black color curves show the results for $z = 2$ with $N_{\rm LOS} = 200$, the blue color $z = 2$ with $N_{\rm LOS} = 1000$ and green $z = 3$ with $N_{\rm LOS} = 200$. Panel (a) shows the auto–correlation function of the true field and panel (b) the cross–correlation function of the true and reconstructed fields. Panel (c) shows the standardized cross–correlation function for the entire simulation volume computed using the equation $C_{12}/\sqrt{A_1 \cdot A_2}$. Panel (d) shows the standardized cross–correlation function computed only for the part of the simulation volume that is at least 50 h^{-1}Mpc from an edge.

reconstruction is not as high as in the $z = 2$ case. We have observed much better recovery at $z = 2$ than at $z = 3$, a trend that is recognizable in every plot.



(a) $z = 2$, $N_{\mathrm{LOS}} = 200$

(b) $z = 2$, $N_{\mathrm{LOS}} = 400$

(c) $z = 3$, $N_{\mathrm{LOS}} = 200$

Figure 2.4: Probability density distributions for flux contrast for the original (grey), reconstructed (black) field and a Gaussian centered at zero with the standard deviation matching that of the real field (red). The top row indicates PDFs for the data set at $z = 2$ for varying LOS densities similar to that BOSS. The bottom panel shows the PDF for the data set at $z = 3$.

2.2.3.2 Kolmogorov–Smirnov Tests

In order to look for deviations in the flux pdfs, we use the Kolmogorov–Smirnov (KS) test. We first compute the mean and the standard deviation σ of a Gaussian fit to the true field. We construct a cumulative probability distribution from this and compare it to the cumulative probability distributions of the true field and the reconstructions, for various $N_{\rm LOS}$ values, redshifts and levels of noise.

With the KS test, we compute quantitative measures of the similarity of the flux pdfs to normal distributions. The test statistic (D value) is the maximum of the difference in the cumulative distribution functions of the particular field being tested and the Gaussian. The closer this value is to 0, the more likely it is that the data sets have been drawn from the same distribution. Furthermore, the p value, which is computed from the test statistic, represents the significance level threshold below which the null hypothesis (that the data sets come from the same distribution) will be accepted.

When computing the flux pdfs, we would like the data points to be as independent as possible, and so our data points should at least be separated by distances greater than the smoothing scale, because smoothing would correlate the measurements. Because of this, we downsample each data set, picking only a 5^3 grid of values (i.e., data points separated by 80 h^{-1}Mpc in each direction in the 400 h^{-1}Mpc volume). Our KS test results are shown in Table 2.4, for different line of sight densities, redshifts and signal to noise ratios.

In all cases for the true field we find high p values and low D values, which means that the original field was approximately Gaussian to start with. The recovered field also shows the same property as well in Table 2.4, for recovery from data samples with no added noise. When noise was added, however, the reconstructed maps became significantly non–Gaussian, with the p value decreasing as the signal to noise ratio decreased. Furthermore, the fact that the D and p values between samples with different LOS density are significantly different can be attributed to the fact that the smoothing filter size depends on the LOS density itself. As we have mentioned above, the noisiest data, which has S/N=1, is significantly worse than the majority of BOSS data, for example, but the effect of noise in changing the pdf shape of the reconstructed field should still be borne in mind in an analysis of observed data.

2.2.4 Peaks In the Density Field

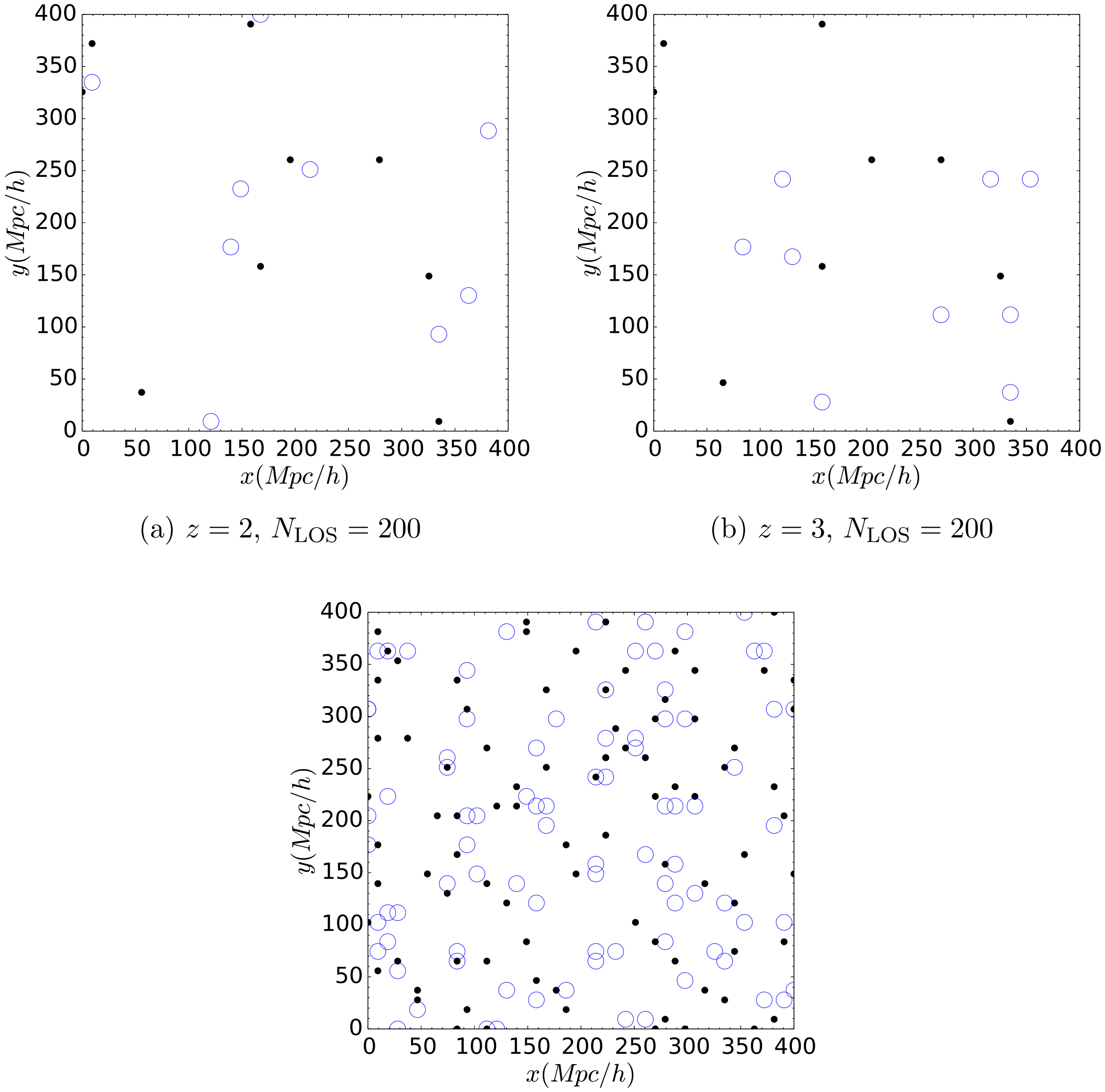
Searching for local maxima in the reconstructed flux density field offers one means of defining objects and finding them. Such peaks are likely to correspond to the locations of forming clusters or superclusters of galaxies. The properties of these density maxima can be used to constrain the cosmological model (Bardeen et al., 1986; Croft and Gaztañaga, 1998; De and Croft, 2007, 2010). It is therefore of interest to compare the peaks of the reconstructed flux density field with those in the true flux density field in the simulation.

DATA SET	**D value**	**p value**
$\mathbf{z = 2, N_{LOS} = 200}$, **Noiseless**		
Original	0.067	0.63
Reconstruction	0.060	0.77
$\mathbf{z = 2, N_{LOS} = 200}$, **S/N=2**		
Original	0.068	0.60
Reconstruction	0.14	0.011
$\mathbf{z = 2, N_{LOS} = 200}$, **S/N=1**		
Original	0.058	0.81
Reconstruction	0.23	4.4×10^{-6}
$\mathbf{z = 2, N_{LOS} = 400}$, **Noiseless**		
Original	0.044	0.97
Reconstruction	0.069	0.57
$\mathbf{z = 2, N_{LOS} = 1000}$, **Noiseless**		
Original	0.088	0.28
Reconstruction	0.057	0.83
$\mathbf{z = 3, N_{LOS} = 200}$, **Noiseless**		
Original	0.065	0.68
Reconstruction	0.098	0.17

Table 2.4: Kolmogorov–Smirnov test results giving the probability (p value) of the flux PDFs of the real and the reconstructed data being drawn from a Gaussian distribution. We show results for different numbers of quasar sightlines through our simulation volume, $N_{\rm LOS}$, redshifts and signal to noise ratios.

We search for density peaks in the three dimensional volume of the simulation. Density and flux are inversely related, therefore we identify a simulation 3D pixel as a local peak if its flux value is the smallest amongst the 26 neighboring 3D pixels surrounding it. As expected, we find that the number density of local peaks is strongly dependent on the smoothing filter size. We find that for a filter size of 39.6 h^{-1}Mpc, appropriate for $N_{\rm LOS} = 200$, we find 9 local peaks in the simulation volume at $z = 2$ (Figure 2.5a), and for a filter size of 17.8 h^{-1}Mpc, appropriate for $N_{\rm LOS} = 1000$, we find 87 peaks (Figure 2.5c), where both of these figures are for the true field. Furthermore, $z = 2$ and $z = 3$ data samples have very similar true peak locations for the same filter size. When noise is added to mimic observational data, we discover 8 local peaks for the real simulation flux field, while the reconstructed field contains 10 local peaks for the data set z2_N200_SN1.

In Figure 2.5 we show a comparison between the true field local peaks and those of the reconstructed field. The number of peaks in both cases match exactly for the three different combinations of $N_{\rm LOS}$ and redshift shown. The positions of the peaks are visually a reasonable match, with better agreement for $z = 2$ than $z = 3$. The structures traced out in the plot with $N_{\rm LOS} = 1000$ by the reconstructed peaks do seem visually to trace out those in the real peaks.



(a) $z = 2$, $N_{\rm LOS} = 200$

(b) $z = 3$, $N_{\rm LOS} = 200$

(c) $z = 2$, $N_{\rm LOS} = 1000$

Figure 2.5: Coordinates of potential superclusters (local peaks) from noiseless analysis at $z = 2$ and $z = 3$ with varying LOS densities. Black dots show the results from the true field, while blue empty circles show the results from the recovered field. Most black dots are enclosed by or neighboring a blue circle, indicating accurate statistics of the recovered field – when two supercluster candidates are on top of each other.

There is not a one to one correspondence however. We can quantify the level of agreement by counting the number of peaks in the true field which have a peak in the reconstructed field within one smoothing length. This is 33.3 per cent for $z = 2$, $N_{\rm LOS} = 200$. The expectation from randomly positioned peaks with the same number density (computed using 1000 Monte Carlo trials) is 11.5 per cent. This means that the reconstruction is a factor of 2.89 better than random. The equivalent measures for $z = 3$, $N_{\rm LOS} = 200$ and $z = 2$, $N_{\rm LOS} = 1000$ are 11.1 per cent and 32.1 per cent peaks within 1 smoothing length respectively and factors of 1.13 and 10.4 better than random.

This type of analysis could be potentially extended to look at the structures that are enclosed within isodensity contours. This would reveal the morphology of the IGM on large scales. At the scales we are probing here (smoothing scales > 10 h^{-1}Mpc), sheet and filament-like topologies are relatively difficult to see (as we shall see in our visualizations in the next section). On smaller scales, these characteristics are readily apparent in simulated maps (e.g. Pichon et al. 2001) and the first maps made from observational data with these techniques (Lee et al., 2014a). The most straightforward cosmological constraints will come from the peak density, and the reconstruction technique does very well: We get perfect agreement between the real and reconstructed fields for noiseless comparison, and for the noisy case (with z2_N200_SN1), the number of peaks agrees within 25 per cent.

2.2.5 Slice Images

We now turn to a visual comparison of the structures in the real and reconstructed maps. The three-dimensional datacubes have z axes oriented parallel to the line of sight, and x and y axes perpendicular to it. The sampling of pixels in a mock data set is therefore different depending on the plot orientation, and this could influence the recovery of structure. We therefore show two orientations for each plot, one in the $y - z$ plane (an "x" slice) and one in the $x - y$ plane (a "z" slice). In our plots we show the flux contrast in a slice of thickness one grid cell. As our volumes are 44 cells on a side, this corresponds to a thickness $400/43 = 9.3$ h^{-1}Mpc.

We have also seen in §2.2.1 that there is a bias in the reconstructed field which leads it to have higher contrast. Changes in $N_{\rm LOS}$ and adjusting the correlation lengths do not alter this and so we follow (Lee et al., 2014a) in applying a bias correction before visualizing the fields.

In Figures 2.6 through 2.12 we present image slices through the simulation volume. The images show the flux contrast $\delta_F = (F/\langle F\rangle) - 1$, which means that low values correspond to high values of the matter density. The red color shows these higher density regions and white those of lower density. The image slices are taken from the center of the cube in directions parallel and perpendicular to the LOSs. As explained in Section 3, the motivation for choosing to display slices through the center of the cube is because we are not using information about the periodic boundary conditions in the simulation when carrying out the reconstruction. The edges of each image slice will therefore give an idea of how well the reconstruction would succeed at the edges of a survey volume. We have checked other random slices and verified that the reconstruction recovers the general features of the original field, even when close to edges of the simulation volume.

In Figure 2.6 we can see the results for our lowest density of sightlines (we have $N_{\mathrm{LOS}} = 200$), at redshift $z = 2$. We can see that the general morphology of the field is recognizably similar in the true and reconstructed maps. In detail, the maps have some differences, but the maxima, minima and their gross shapes are fairly well reproduced, and one could therefore expect that observational data from the BOSS survey (which has approximately this number density of quasars) would yield visually quite accurate maps of the large scale structure, at least when smoothed on the relevant filter scale (a filter of 39.6 h^{-1}Mpc was used here).

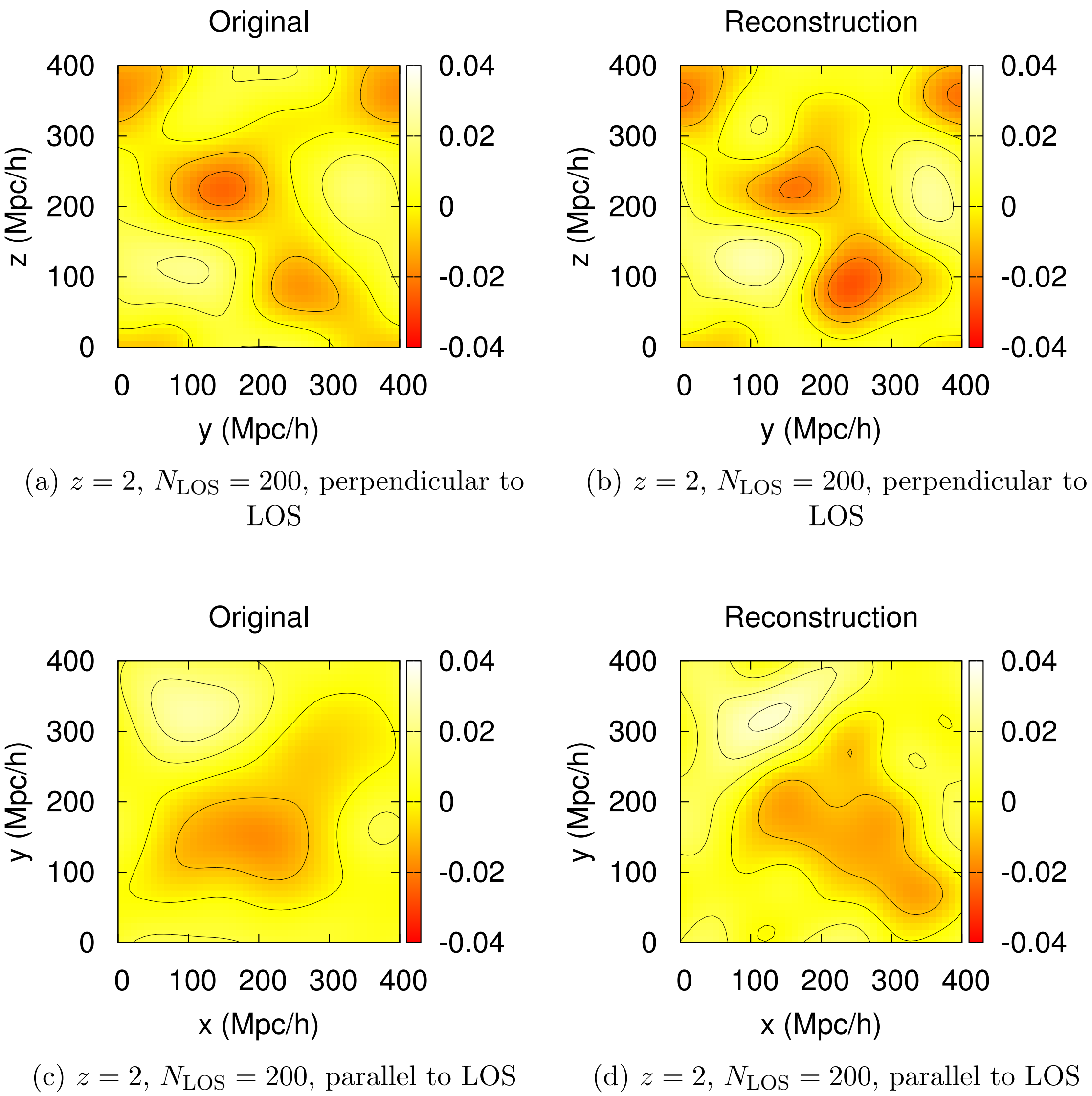
The top and bottom rows of Figure 2.6 show results for slices parallel and perpendicular to the line of sight. We see no obvious difference in the fidelity of reconstruction for each, and there is no obvious sign of the discrete sampling of the field by pixels and sightlines (which is different for the top and bottom rows). General features of the field are recovered well, especially for mildly dense regions.

Figures 2.7 and 2.8 demonstrate the effect of adding uncorrelated Gaussian noise to the flux field in order to better mimic observational data, and how the fidelity of the reconstruction changes when noise is introduced. Noise levels (S/N = 1 or 2) indicate the amount of noise for a simulation pixel (~ 0.76 h^{-1}Mpc) wide. Since our pixels are rebinned to ~ 9 h^{-1}Mpc, the added noise is reduced by a factor of 3.4. Hence, the difference between the true fields before and after adding noise is small. However, the reconstruction is sensitive to the amount of noise, therefore the fidelity of the noisy reconstruction is noticeably worse, especially for overdense and underdense regions. Due to the sensitivity of the reconstruction, we observe recognizably greater dynamic range in the reconstructed field when noise is added.

We increase the density of the sightlines in Figures 2.9 and 2.10. As a result, the quality of the reconstruction is visually better. Although the reconstruction code does not take into account the periodic boundary conditions of the simulation, the fields are comparable even at the edges. This is likely due to lower smoothing levels, as the smoothing level scales inversely with N_{LOS}.

Fluctuations in the flux field are greater at higher redshifts. As Figures 2.11 and 2.12 clearly show, this results in a decrease in the fidelity of the reconstruction. It is obvious that at redshift $z = 3$, the LOS density of $N_{\mathrm{LOS}} = 60$ is not good enough to yield a comparable reconstructed field. For observational data, we naturally expect a better map at $z = 2$ than at $z = 3$, as the LOS density is higher at $z = 2$. In this study, although the LOS density is set to be the same at both redshifts, we get a better recovery of the field at $z = 2$.

The recovery is accurate for scales larger than $\sim 1.4\langle d_{\mathrm{LOS}}\rangle$, as found in (Caucci et al., 2008), especially for mildly dense regions (standardized correlation plots). Due to the isotropic nature of the recovery and the smoothing, we do not notice any significant statistical difference between the directions parallel to and perpendicular to LOSs. If an anisotropic approach is found to be a significant improvement in future studies, they can be implemented with different correlation lengths in Wiener interpo-



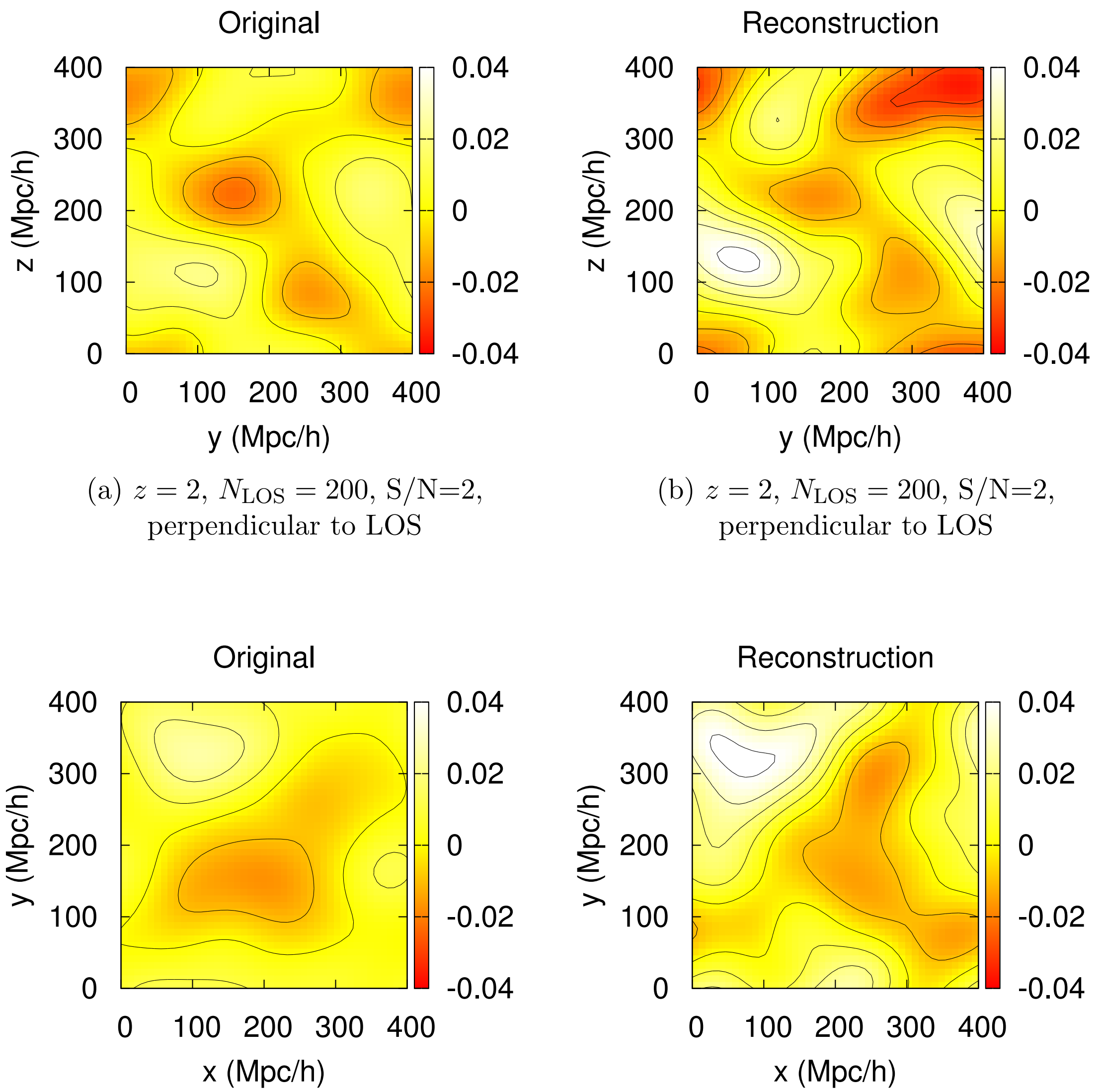
(a) $z = 2$, $N_{\mathrm{LOS}} = 200$, perpendicular to LOS

(b) $z = 2$, $N_{\mathrm{LOS}} = 200$, perpendicular to LOS

(c) $z = 2$, $N_{\mathrm{LOS}} = 200$, parallel to LOS

(d) $z = 2$, $N_{\mathrm{LOS}} = 200$, parallel to LOS

Figure 2.6: Slices extracted from the middle planes of the simulation cube are shown at $z = 2$ with $N_{\mathrm{LOS}} = 200$, without pixel noise. The color scale indicates flux contrast, δ_f. The top row shows slices perpendicular to LOSs, whereas the bottom row shows slices in the parallel direction. True field slices are given in (a) and (c), while (b) and (d) show reconstructed field slices. The smoothed reconstructed field recovers the general features of the simulation.



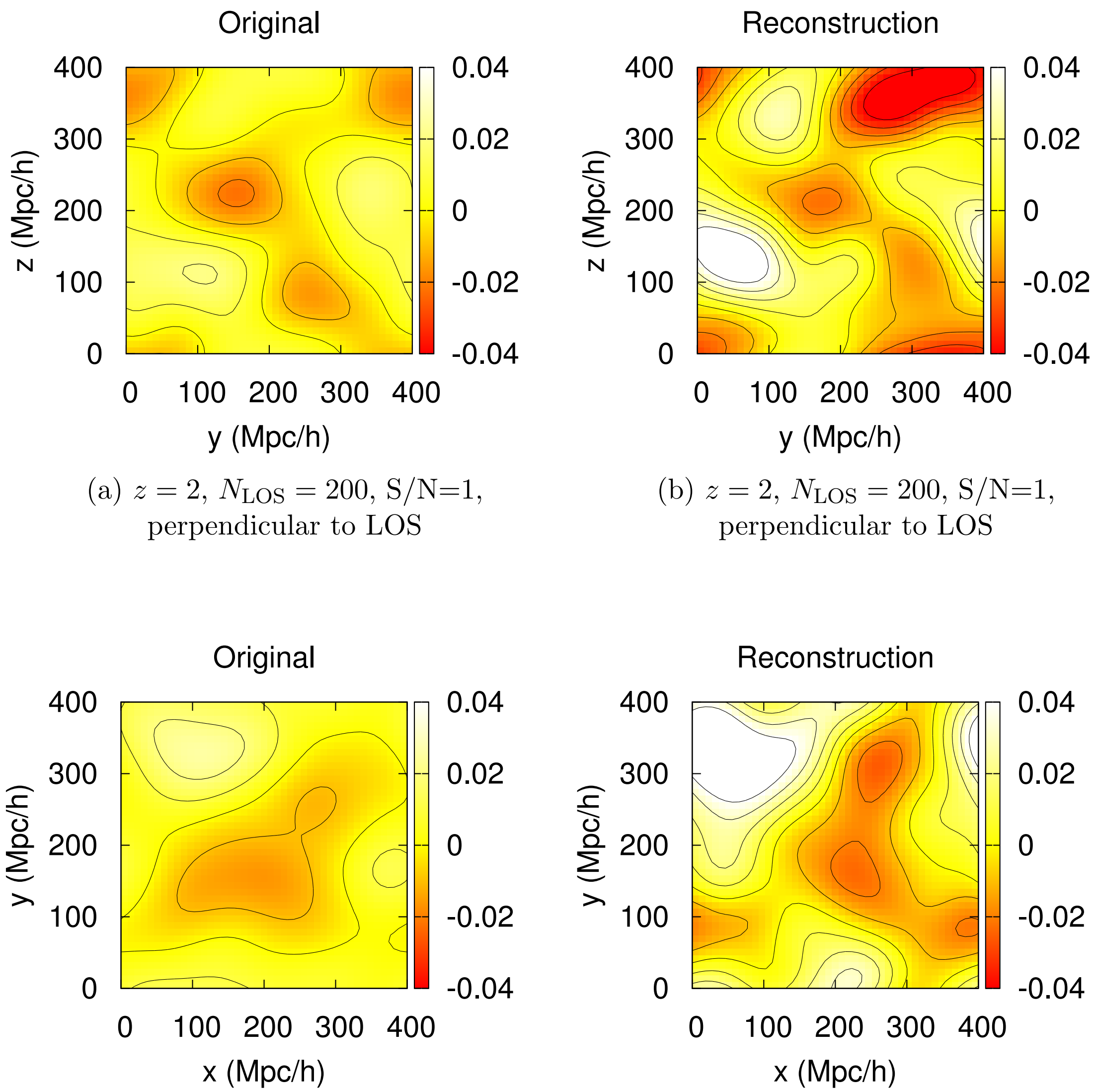
(a) $z = 2$, $N_{\mathrm{LOS}} = 200$, S/N=2, perpendicular to LOS

(b) $z = 2$, $N_{\mathrm{LOS}} = 200$, S/N=2, perpendicular to LOS

(c) $z = 2$, $N_{\mathrm{LOS}} = 200$, S/N=2, parallel to LOS

(d) $z = 2$, $N_{\mathrm{LOS}} = 200$, S/N=2, parallel to LOS

Figure 2.7: Slices extracted from the middle planes of the simulation cube are shown at $z = 2$ with $N_{\mathrm{LOS}} = 200$, with Gaussian pixel noise added (S/N = 2). The top row shows slices perpendicular to LOSs, whereas the bottom row shows slices in the parallel direction. True field slices are given in (a) and (c), while (b) and (d) show reconstructed field slices. The smoothed reconstructed field recovers the general features of the simulation.



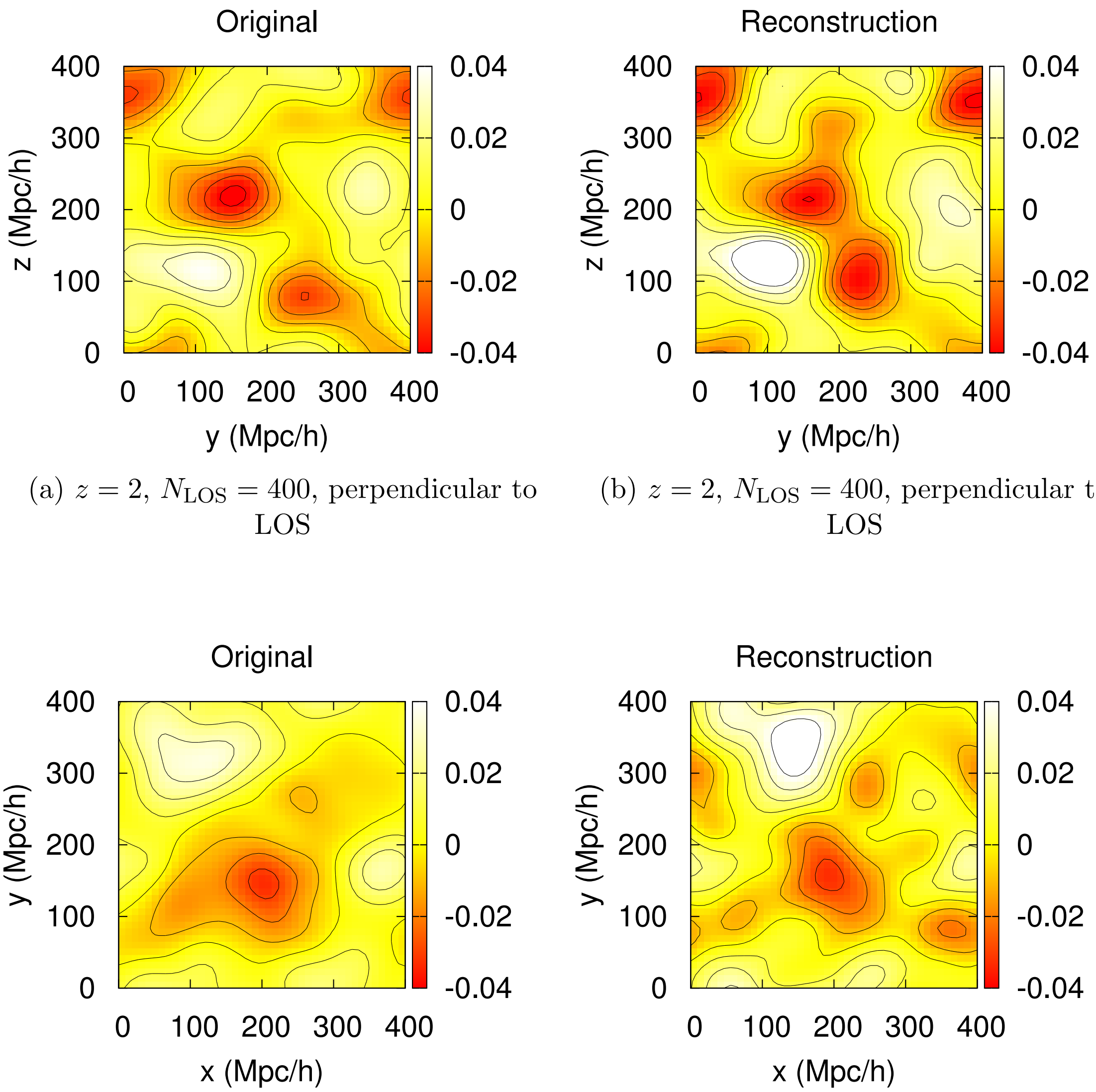
(a) $z = 2$, $N_{\text{LOS}} = 200$, S/N=1, perpendicular to LOS

(b) $z = 2$, $N_{\text{LOS}} = 200$, S/N=1, perpendicular to LOS

(c) $z = 2$, $N_{\text{LOS}} = 200$, S/N=1, parallel to LOS

(d) $z = 2$, $N_{\text{LOS}} = 200$, S/N=1, parallel to LOS

Figure 2.8: Slices extracted from the middle planes of the simulation cube are shown at $z = 2$ with $N_{\text{LOS}} = 200$, with Gaussian pixel noise added (S/N = 1). The top row shows slices perpendicular to LOSs, whereas the bottom row shows slices in the parallel direction. True field slices are given in (a) and (c), while (b) and (d) show reconstructed field slices. The smoothed reconstructed field recovers the general features of the simulation.



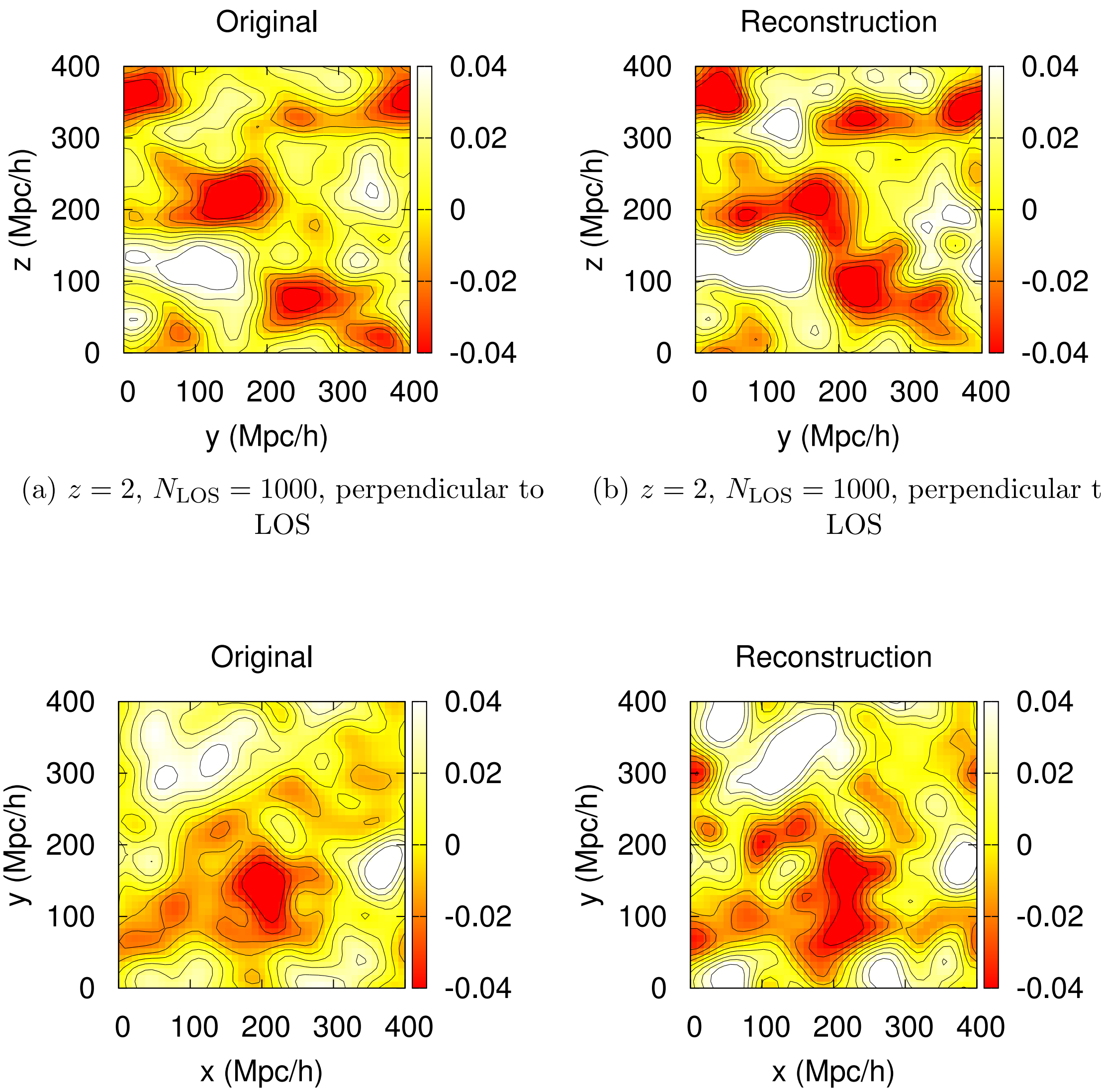
(a) $z = 2$, $N_{\mathrm{LOS}} = 400$, perpendicular to LOS

(b) $z = 2$, $N_{\mathrm{LOS}} = 400$, perpendicular to LOS

(c) $z = 2$, $N_{\mathrm{LOS}} = 400$, parallel to LOS

(d) $z = 2$, $N_{\mathrm{LOS}} = 400$, parallel to LOS

Figure 2.9: Slices extracted from the middle planes of the simulation cube are shown at $z = 2$ with $N_{\mathrm{LOS}} = 400$, without pixel noise. The top row shows slices perpendicular to LOSs, whereas the bottom row shows slices in the parallel direction. True field slices are given in (a) and (c), while (b) and (d) show reconstructed field slices. The smoothed reconstructed field recovers the general features of the simulation.



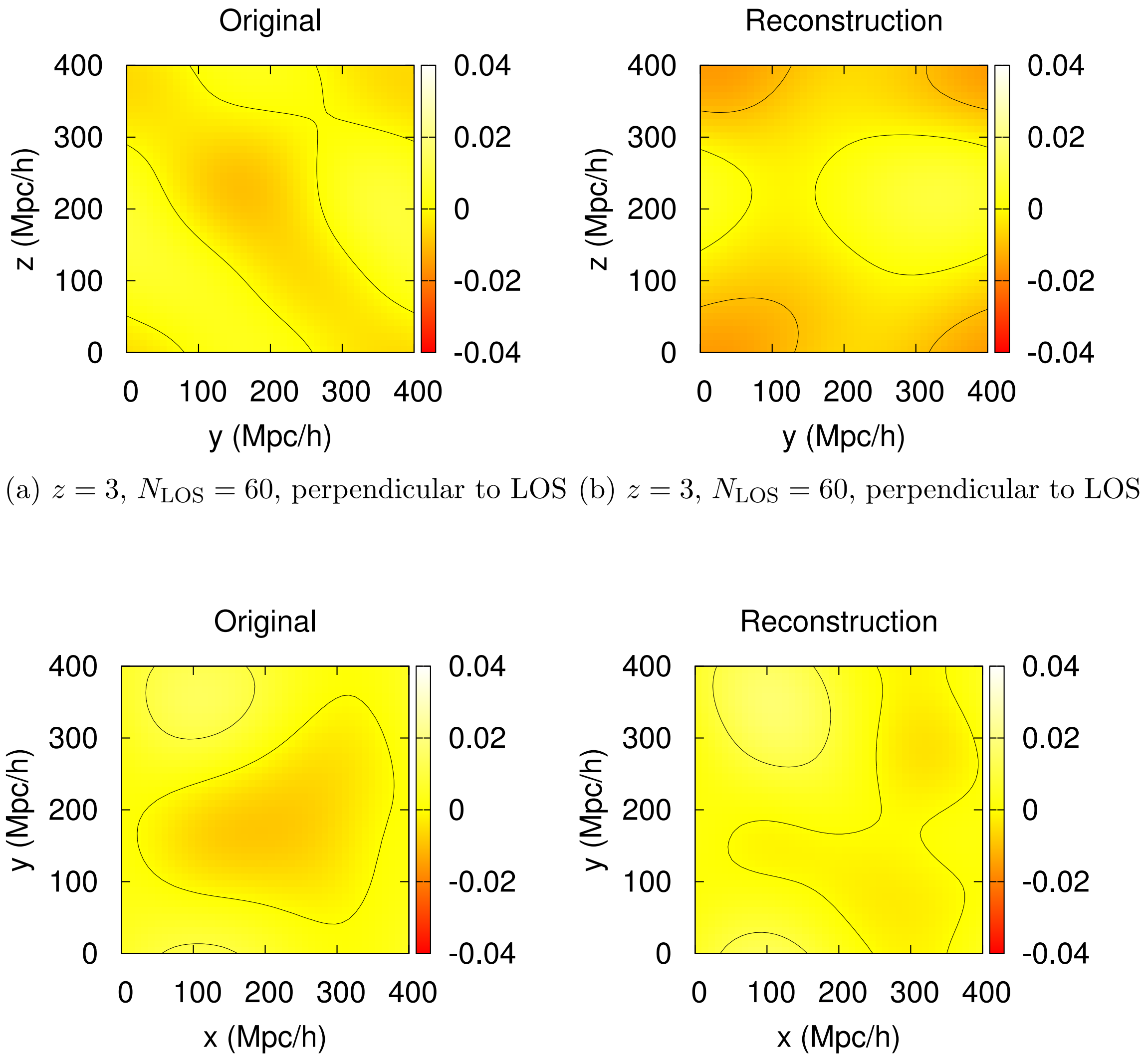
(a) $z = 2$, $N_{\text{LOS}} = 1000$, perpendicular to LOS

(b) $z = 2$, $N_{\text{LOS}} = 1000$, perpendicular to LOS

(c) $z = 2$, $N_{\text{LOS}} = 1000$, parallel to LOS

(d) $z = 2$, $N_{\text{LOS}} = 1000$, parallel to LOS

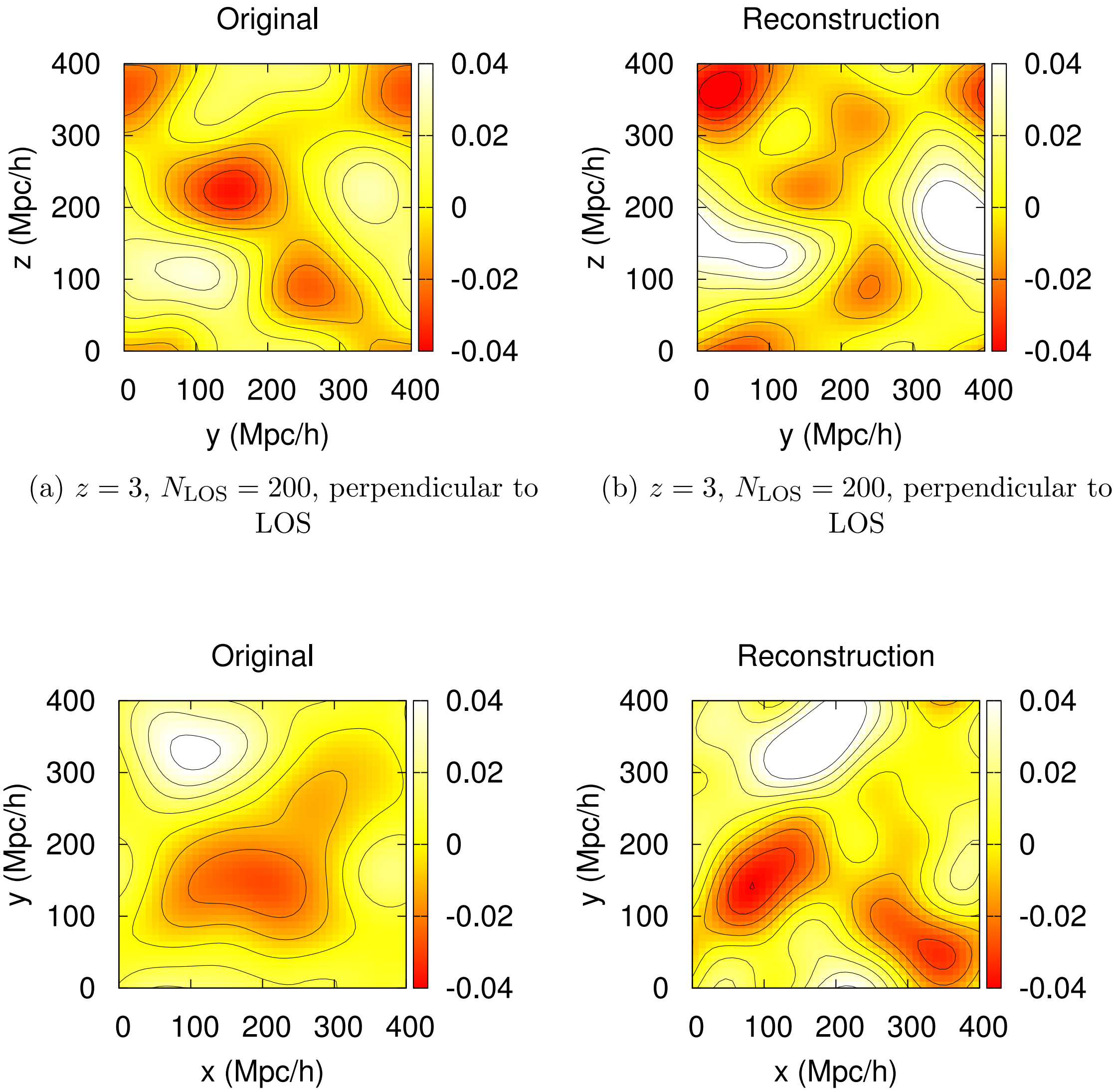
Figure 2.10: Slices extracted from the middle planes of the simulation cube are shown at $z = 2$ with $N_{\text{LOS}} = 1000$, without pixel noise. The top row shows slices perpendicular to LOSs, whereas the bottom row shows slices in the parallel direction. True field slices are given in (a) and (c), while (b) and (d) show reconstructed field slices. The smoothed reconstructed field recovers the general features of the simulation.



(a) $z = 3$, $N_{\text{LOS}} = 60$, perpendicular to LOS (b) $z = 3$, $N_{\text{LOS}} = 60$, perpendicular to LOS

(c) $z = 3$, $N_{\text{LOS}} = 60$, parallel to LOS (d) $z = 3$, $N_{\text{LOS}} = 60$, parallel to LOS

Figure 2.11: Slices extracted from the middle planes of the simulation cube are shown at $z = 3$ with $N_{\text{LOS}} = 60$, without pixel noise. The top row shows slices perpendicular to LOSs, whereas the bottom row shows slices in the parallel direction. True field slices are given in (a) and (c), while (b) and (d) show reconstructed field slices. The smoothed reconstructed field cannot recover the general features of the simulation well when the areal density of the absorption skewers is low.



(a) $z = 3$, $N_{\mathrm{LOS}} = 200$, perpendicular to LOS

(b) $z = 3$, $N_{\mathrm{LOS}} = 200$, perpendicular to LOS

(c) $z = 3$, $N_{\mathrm{LOS}} = 200$, parallel to LOS

(d) $z = 3$, $N_{\mathrm{LOS}} = 200$, parallel to LOS

Figure 2.12: Slices extracted from the middle planes of the simulation cube are shown at $z = 3$ with $N_{\mathrm{LOS}} = 200$, without pixel noise. The top row shows slices perpendicular to LOSs, whereas the bottom row shows slices in the parallel direction. True field slices are given in (a) and (c), while (b) and (d) show reconstructed field slices. The smoothed reconstructed field recovers the general features of the simulation, although the quality of the reconstruction is lower than that with the $z = 2$ data set.

lation and an anisotropic Gaussian filter. In our maps, as N_{LOS} increases, naturally, the recovery gets substantially better. This means that with future experiments like eBOSS and MS–DESI, which have higher areal density of LOSs, a very accurate large map of the IGM can be generated.

Adding noise to our data (pixel by pixel) and carrying out the recovery is an important step in order to better simulate real data from experiments. It is clearly seen from the figures that adding noise makes the recovery of overdense and underdense regions significantly worse. Furthermore, our results with the data set at $z = 3$ are significantly worse than the other data set at $z = 2$.

2.2.6 Observed Correlations in the Covariance Matrix

Along with Caucci et al. 2008, and Lee et al., 2014, we use a simple Gaussian form for the correlation function which appears in the Wiener interpolation covariance matrix (Equation 2.3). This is motivated by simplicity, and the fact that it is well behaved numerically at large separations. One might expect covariance matrices computed from the actual correlation functions of the field to give more accurate reconstruction results, however, and we now test this.

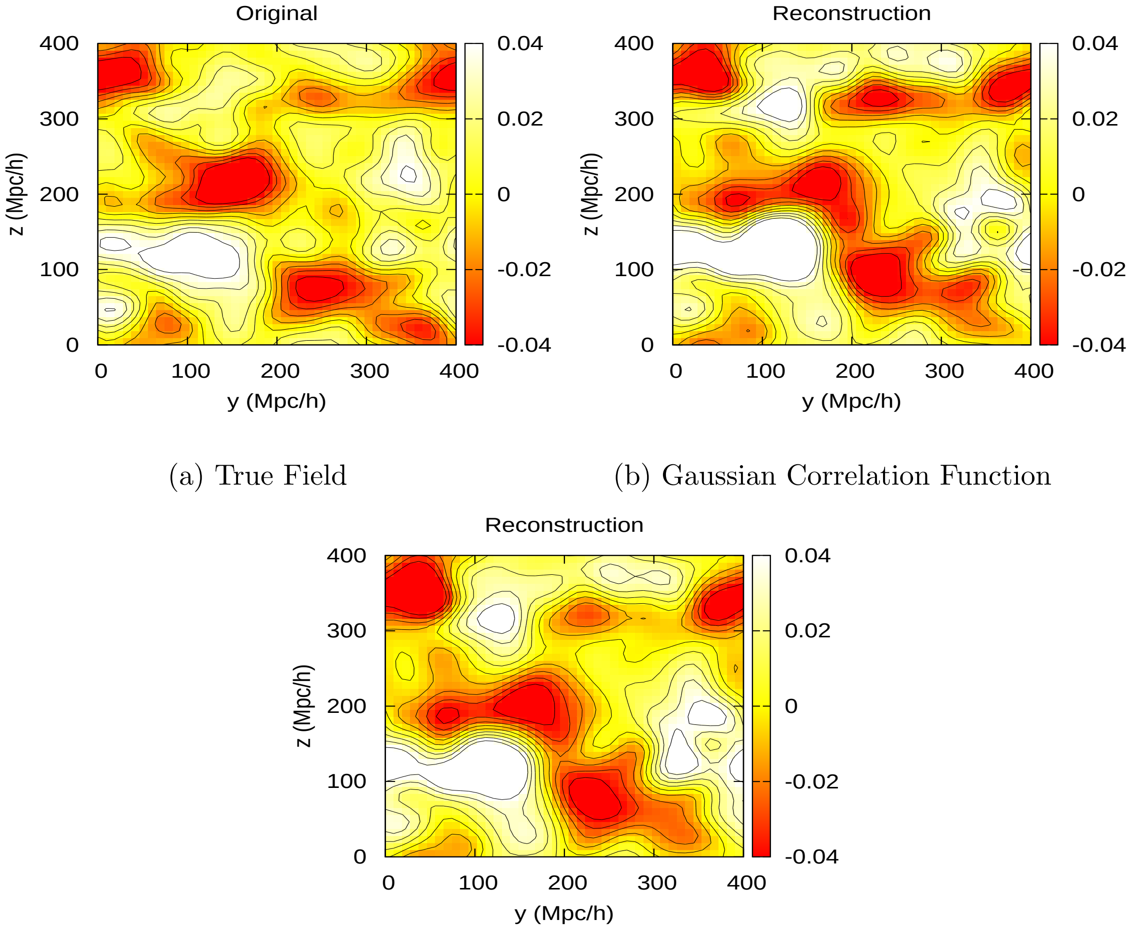
In (Slosar et al., 2011), the three-dimensional correlation function of the absorption in the Lyman-α forest was measured for the first time. The measurement was extended to greater than $100\,h^{-1}$Mpc scales by (Busca et al., 2013) and (Slosar et al., 2013). We use this measurement of correlation function to construct a correlation matrix instead of the Gaussian covariances we have used (Equation 2.3).

The correlation function measured from the observational Lyα forest data is anisotropic because of redshift distortions. We construct the correlation matrix not from the observational data results of Slosar et al. (2011), but from the linear theory CDM model consistent with the data. This redshift space model fit is given by Equations 4.5 – 4.13 of (Slosar et al., 2011). We use these equations, along with the linear theory correlation function from Section 2, and the following parameters: bias factor $b = 0.2$, and redshift distortion factor $\beta = 1.5$ to compute $\xi_F(r_\perp, r_\parallel)$, the flux correlation function for line of sight separation $r_\parallel$ and transverse separation $r_\perp$. The Wiener covariance (replacing Equation 2.3) is then given by

$$\mathbf{C}(x_1, x_2, \mathbf{x_{1\perp}}, \mathbf{x_{2\perp}}) = \xi_F(r_\perp, r_\parallel) \tag{2.8}$$

where $r_\parallel = (x_1 - x_2)$ and $r_\perp = |\mathbf{x_{1\perp}} - \mathbf{x_{2\perp}}|$.

After reconstructing the simulation field using the CDM fit to the (Slosar et al., 2011) results in the covariance matrix, we compare the results to our fiducial reconstruction technique (Figures 2.13, 2.14). We find that the recovery of the field with the fiducial (Gaussian) correlation functions yields slightly better results than with the CDM correlation function. Although the dynamic range with the Gaussian correlation seems to be slightly higher, general features of the original field are recovered



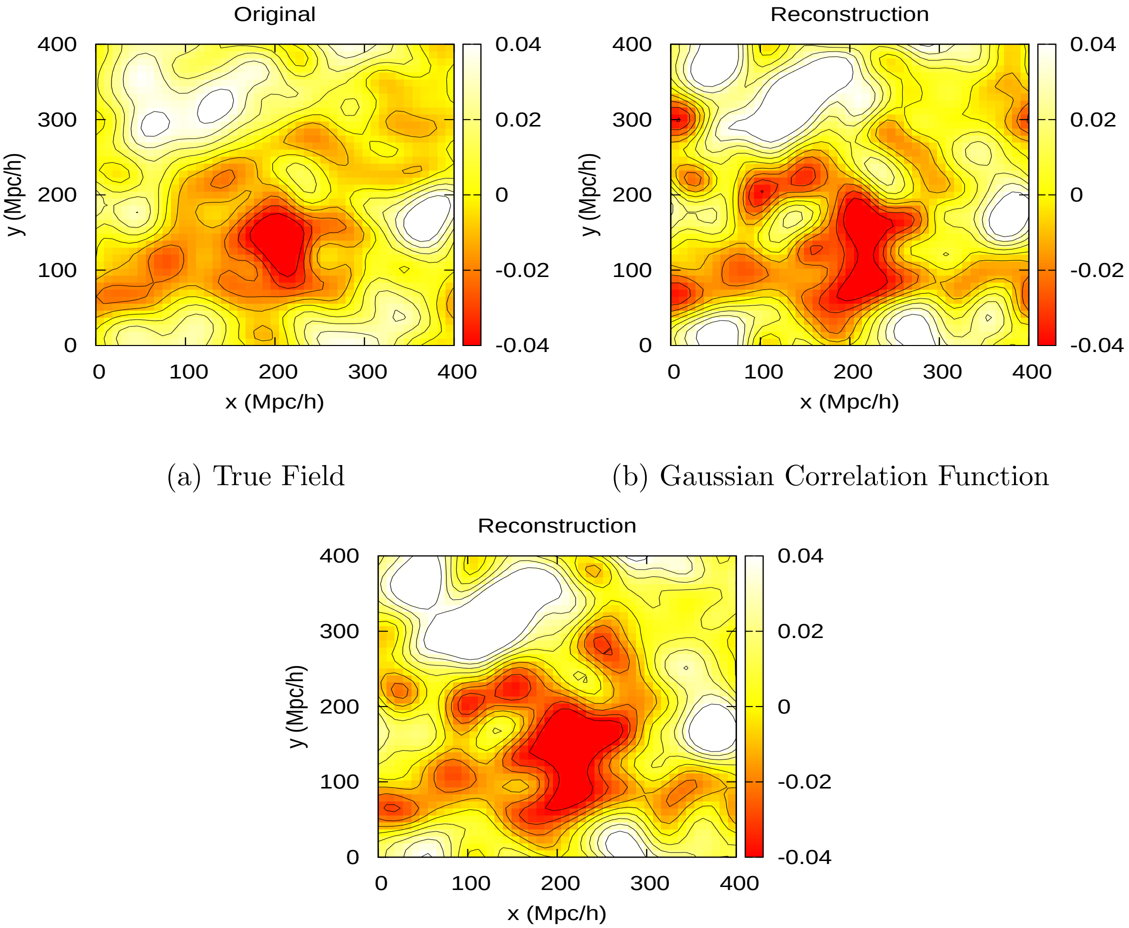
(a) True Field

(b) Gaussian Correlation Function

(c) CDM Correlation Function

Figure 2.13: Slice plots from the middle of the cube are shown above from a data sample at redshift $z = 2$ with $\mathrm{N_{LOS}} = 1000$. We use the original Gaussian correlation function as well as a correlation function measured from observations to recover the field. Slices are from the middle of the cube, perpendicular to LOSs. The recovery of the field using the Gaussian correlation function yields better results than using the CDM correlation function obtained from observations.

better. For example, for the data set z2_N1000, instead of our original result of the RMS percentage error 17.2, we find 20.3 with the CDM correlation function, which is significantly worse. It is worth noting that the actual correlation function in the simulations is probably not the same as it is estimated by Slosar et al. (2011).

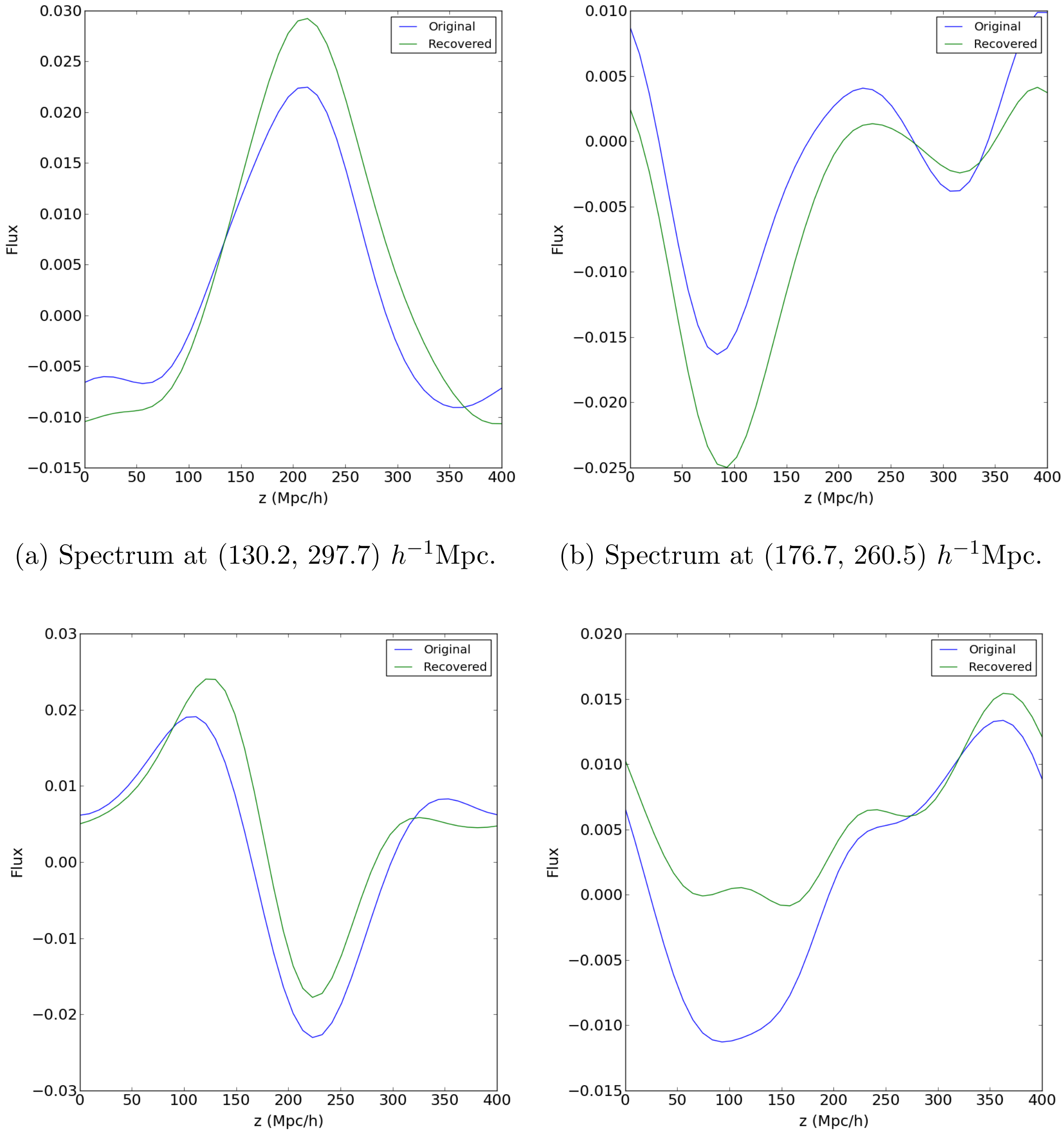


(a) True Field

(b) Gaussian Correlation Function

(c) CDM Correlation Function

Figure 2.14: Slice plots from the middle of the cube are shown above from a data sample at redshift $z = 2$ with $\mathrm{N}_{\mathrm{LOS}} = 1000$. We use the original Gaussian correlation function as well as a correlation function measured from observations to recover the field. Slices are from the middle of the cube, parallel to LOSs. The recovery of the field using the Gaussian correlation function yields better results than using the CDM correlation function obtained from observations.



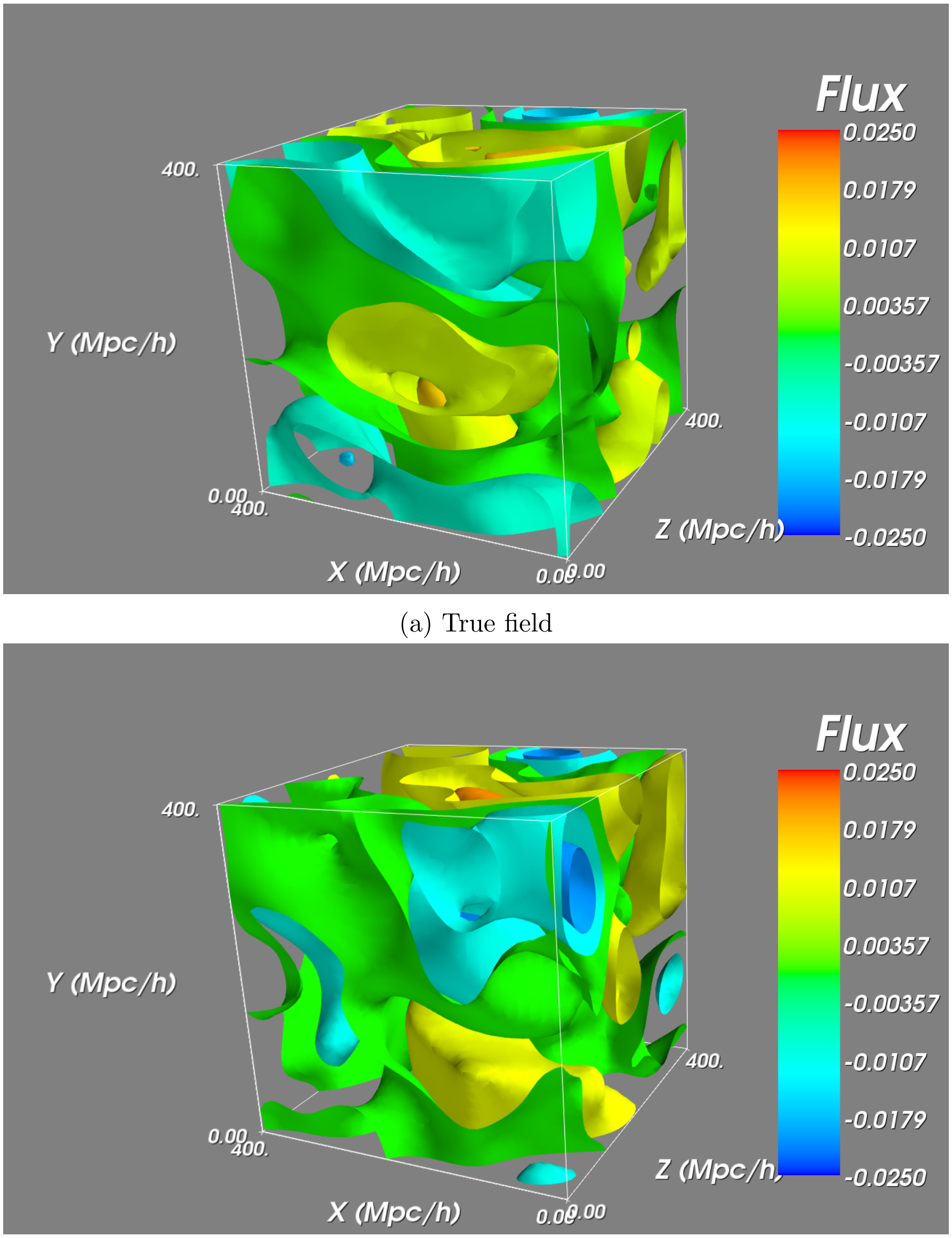
(a) Spectrum at (130.2, 297.7) h^{-1}Mpc.

(b) Spectrum at (176.7, 260.5) h^{-1}Mpc.

(c) Spectrum at (186.0, 130.2) h^{-1}Mpc.

(d) Spectrum at (316.3, 353.5) h^{-1}Mpc.

Figure 2.15: Random LOS comparisons are shown above from a data sample at redshift $z = 2$ with $N_{\mathrm{LOS}} = 200$. Pixel values are compared along a single LOS at the coordinates given in the captions. General features of individual spectra are captured by the recovered field.



(a) True field

(b) Reconstructed field

Figure 2.16: 3D visuals of the true and reconstructed fields show good agreement overall. Blue color shows denser regions. The fidelity of the reconstruction is especially high for mildly dense regions and away from the edges.

2.3 Conclusions

Using Wiener interpolation, we reconstruct the entire simulation box with a subset of the Lyman–alpha absorption skewers chosen randomly. This subset of skewers, $N_{\rm LOS}$, sets a natural resolution of our maps. The number of the skewers chosen at random is decided by matching it with the areal LOS density of current and future spectroscopic surveys such as BOSS and MS–DESI. Using the Lya forest with this method, one can make maps of the large scale structure at high redshifts ($2 < z < 3.5$).

The standardized cross correlation plot (Figure 2.3, panel(c)) indicates that the reconstruction is much better at $z = 2$ than at $z = 3$ using BOSS areal LOS densities. Naturally, the fidelity of the reconstruction is better as $N_{\rm LOS}$ is increased. Truncating the cube 50 h^{-1}Mpc from each edge, in order to remove the edge artifacts resulting from periodic boundary conditions of the simulation and to better mimic observational data, yields significantly better reconstruction (Figure 2.3, panel(d)).

We find that the data set at $z = 2$ yields clearly better results than the one at $z = 3$ for the simulation, even with the same $N_{\rm LOS}$. This is most easily understood in terms of the growth of structure through gravitational instability between $z = 3$ and $z = 2$. For observational surveys, in view of the fact that the areal LOS density is also much greater at $z = 2$ than at $z = 3$, one naturally expects that the large scale structure map will be significantly better at lower redshifts.

The overall bias seen in point to point flux values in real and reconstructed fields is an issue which does not have an easy explanation. Adjusting $N_{\rm LOS}$, the correlation lengths and the buffer length does not change the situation, but using an empirical bias correction allows the fields to be well-reconstructed.

In the high redshift range covered by the Lya forest, the IGM density field is expected to be in the mildly non-linear regime, therefore, we look for non-Gaussianity in the probability density functions of our reconstructed maps. While its behavior is nearly Gaussian for noiseless data samples, it becomes less Gaussian as the noise level is increased, as Table 2.4 indicates.

We provide more visualizations to study the general characteristics of the reconstruction: In Figure 2.15, we show one–dimensional visual comparisons along four lines of sight chosen at random using the data set $z2\_N200$, whose source density matches the areal LOS density of BOSS. We observe that the recovered skewers capture general features of the original ones, especially when the flux is changing slowly along the LOS. Figure 2.16 shows 3D visuals of the true field and the reconstructed field in the simulation volume.

Since the smoothing levels used in this study are greater than 10 h^{-1}Mpc, it is not possible to see the filamentary structure in the IGM topology. We remind the reader that there are no wide-field galaxy surveys that can detect the topology of the IGM at $z > 2$, as it is increasingly expensive to detect galaxies at higher redshifts to reach a high source density, even with 8–10 m telescopes (Le Fèvre et al., 2013). However, searching for local peaks allows us to discover the potential locations of the superclusters, which can be cross–correlated with galaxy surveys. Both the number of the local peaks and their locations are reproduced reasonably well by the interpolation.

Slice images allow a visual comparison between the original and the reconstructed fields. General features of the flux field are well reproduced by the interpolation, especially for mildly overdense regions.

Using a correlation matrix derived from a CDM–fit to observational data instead of the simple Gaussian correlation matrix used in our fiducial Wiener filtering leads to a slightly worse recovery of the field.

As an improvement, the isotropic smoothing of both fields can be altered, as one does not necessarily expect the same statistics parallel and perpendicular to the LOSs. Furthermore, as future surveys like eBOSS and MS–DESI discover more quasars, the fidelity of the large scale structure maps will improve. For example, in order to reach resolutions in the sub 10 h^{-1}Mpc regime at $z \sim 2$ to study the IGM filamentary structure, QSO densities of over 100 deg^{-2} will be necessary.

Having evaluated our map making methods with simulations, we move on to applying it on observational data to create IGM maps, using SDSS–III DR12, in §4.

Chapter 3

Local Polynomial Smoothing and Observational Requirements

This chapter includes my contributions to the published papers (Cisewski et al., 2014) and (Lee et al., 2014b).

We first follow (Cisewski et al., 2014), which uses non–parametric local polynomial smoothing to make maps of the IGM using the Lyα forest, and compare the results with §2, where Wiener interpolation was the method of choice for interpolation. By reconstructing an observational volume similar to the simulation, we also discuss the feasibility and the LOS density necessary for such maps. We extend observational requirements by studying exposure times and spatial resolutions necessary to reach desired Lyα map resolutions from (Lee et al., 2014b).

3.1 Non–parametric 3D Map of the IGM

3.1.1 Introduction

In the previous chapter, we showed that Wiener interpolation allows using a collection of the sparse Lyα skewers to make maps of the IGM at high redshifts ($z > 2$). However, there are other interpolation methods we can use, with different advantages and drawbacks. One such method is local polynomial smoothing (Cleveland et al., 1992; Wasserman, 2006).

One of the disadvantages of Wiener interpolation is the necessity of knowing the spatial distribution of the data prior to applying the filter (the $\mathbf{C_{DD}}$ term in Equation 2.2), for which we had used Equation 2.3, and also the form from observed correlations in Slosar et al. (2011). Unless the auto–correlation of the field is known well, errors will be introduced in the resulting map. Local polynomial regression mitigates this by using an adaptive function f, a polynomial of degree d:

$$y_i = f(x_i) + \epsilon_i \tag{3.1}$$

where y_i represents the measured optical depth from Lyα skewers, x_i are the locations of the voxels, ϵ_i are the independent random errors with expectation 0 and f is the unknown map of the density field in a given region. The index i denotes individual data points.

The estimation of the function f is local, meaning the global properties of the field need not be known, but only information about the neighborhood of each point i is needed. This is achieved by calculating a smoothing parameter α, which is a value between 0 and 1 that determines the fraction of the full data set to be used for estimating the function at a given point. Hence, a larger α value results in a smoother map. This also has the added benefit of altering the smoothing adaptively in the volume: Areas with a greater number of observations are assigned a lower α value, while sparse regions use a greater neighborhood distance through a higher α value. This can be particularly useful in making maps with QSO spectra, as the highly clustered nature of these objects cause variation in their sky density.

The local polynomial function f at some point x is calculated by finding the coefficients $(a_0, a_1, ..., a_d)$ that minimizes the squared error

$$\sum_{i=1}^{n} (y_i - p_x(x_i; a))^2 K\left(\frac{x_i - x}{h_\alpha}\right) \tag{3.2}$$

where $p_x(x_i; a) = a_0 + a_1(x_i - x)^2 + ... + \frac{a_d}{d!}(x_i - x)^d$ is the desired polynomial and h_α is the bandwidth available in the neighborhood of that location, describing the percentage of the full data set available for local smoothing. The choice for the kernel K does not significantly alter the results (Fan et al., 1997). The estimate is at the point x is simply given by $f(x) = p_x(x, a)$.

We estimate the performance of local polynomial smoothing by using the same simulated data set from §2, a cube of side length 400 h^{-1}Mpc with 176^3 data points, where the QSOs are located at $z = 2$. From this data set, with the same LOS densities used in §2, we are showing analyses with subsets of the data containing 100, 200 and 1000 LOSs chosen at random. These correspond to our data sets named z2_N200, z2_N1000 etc., which are self explanatory. After an appropriate α is chosen for each data set, the same smoothing level is applied to the full data set for comparing the fidelity of the reconstructed fields. Finally, we use this method on a sample of 234 quasars from BOSS DR9 within a volume similar to that of the simulation to reconstruct an observational volume at a redshift range of $2.2 \lesssim z \lesssim 2.3$.

In order to evaluate the performance of the resulting IGM maps, we present slice images taken from the simulation cube and the observational volume, and we provide standardized cross–correlation analysis, drawing comparisons between the results of §2 and §3.

3.1.2 Analysis

The first step is to decide on the optimal smoothing parameter α for all data sets. This is done by minimizing the predictive risk, using generalized cross–validation (Fan and Gijbels, 1996). This sets a smoothing length, and therefore a natural resolution for the reconstructed maps. Naturally, the map is smoother for the subsets containing less LOSs, where we cannot expect to resolve fine features. For each reconstruction, the corresponding ”true” maps are taken from the full data set, but smoothed with the same α used for that particular reconstruction. We visually compare the fields with slice images and discuss the performance of the interpolation as a function of scale with standardized correlation plots.

3.1.3 Slice Images

3.1.3.1 Simulation

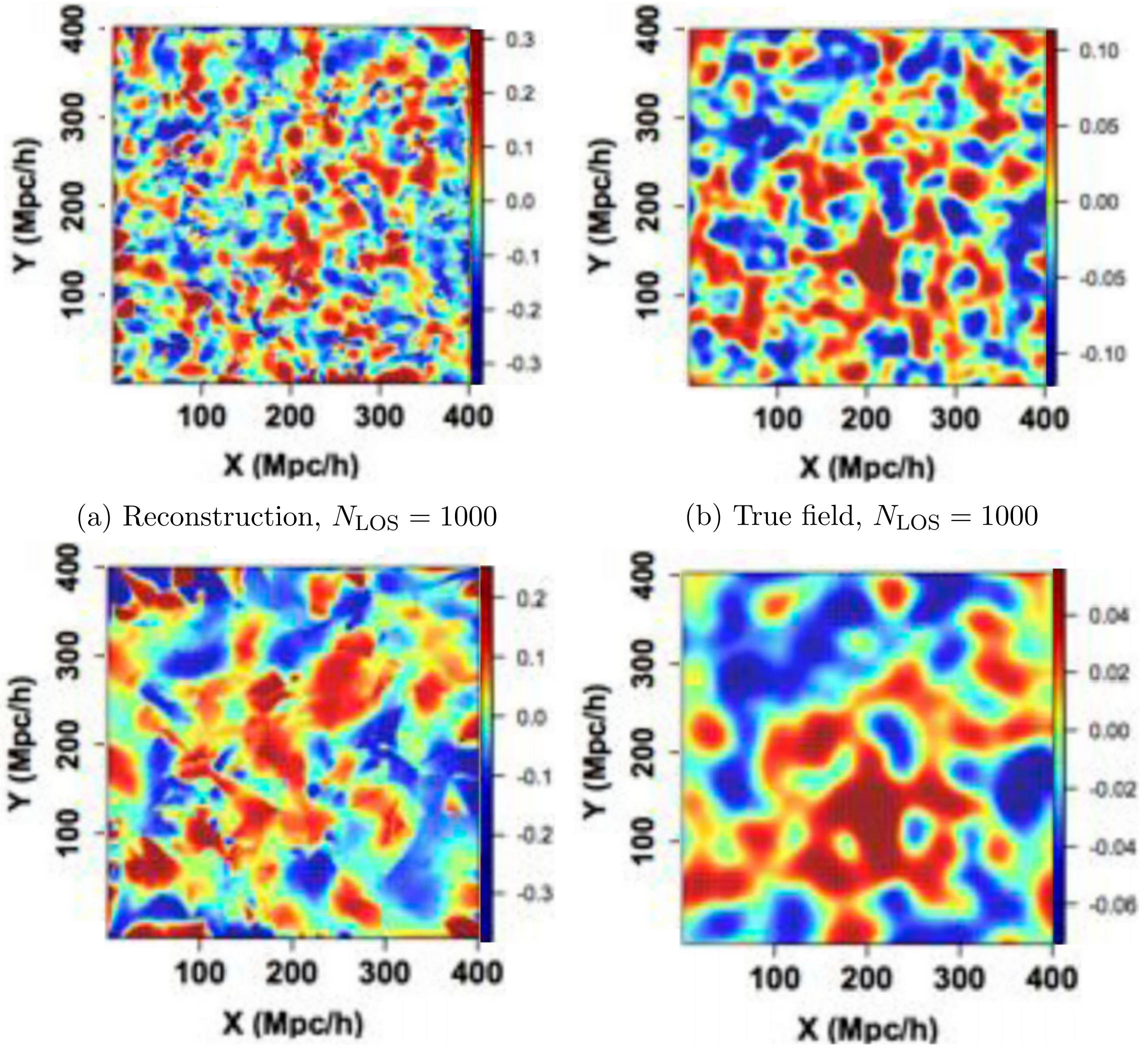
Slice images for the simulated volume are located at the same fixed z value, across a plane that is perpendicular to the LOSs. This makes it possible to directly compare not only the fidelity of the reconstruction, but also the effect of the smoothing levels between different LOS densities on the resulting map.

In Figure 3.1, we plot the negative of the delta flux, hence the red color corresponds to denser regions. Panels (a) and (c) are the reconstructions with LOS densities 1000 and 200, respectively. They can be directly compared with the true field given in panels (b) and (d), which have similar smoothing levels due to the same α value as the corresponding reconstruction. It is clear that the map with the higher line of sight density in panel (a) yields better fidelity than the one in panel (c). Finer features of the map are resolved in panel (a) and visually, and it compares well with the corresponding map in panel (b). The inferred field with $\mathrm{N}_{\mathrm{LOS}} = 200$ also captures general features of the original field successfully, especially for scales larger than $\sim 50\,h^{-1}$Mpc. The dynamic ranges in both instances of the reconstruction are higher than those of the original map, despite having the same α parameter. This may be due to the fact that the limited number of observations in reconstructed maps cause a greater variation in the resulting flux field.

3.1.3.2 BOSS DR9 QSO Sample

After evaluating the performance of local smoothing with simulations, we apply the same interpolation method to a subset of the Lyα forest data from BOSS SDSS–III DR9, within a volume that is approximately the same as the simulation. This analysis contains 234 QSOs with 24596 data points (with optical depth values) within a right ascension (RA) range of 205 and 211, and a declination (DEC) range between -3 and 3. The redshift interval is between 2.2 and 2.3.

A slice image of the reconstructed BOSS IGM field is provided in Figure 3.2. Since



(a) Reconstruction, $N_{\mathrm{LOS}} = 1000$

(b) True field, $N_{\mathrm{LOS}} = 1000$

(c) Reconstruction, $N_{\mathrm{LOS}} = 200$

(d) True field, $N_{\mathrm{LOS}} = 200$

Figure 3.1: Slices in the perpendicular direction with $\mathrm{N}_{\mathrm{LOS}}$ values of 200 and 1000 are shown. Higher density regions correspond to the red color, as the negative of delta flux $(-\delta)$ are the values in the visualization.

this observational LOS density is similar to that of the simulation with 200 LOS, it is expected that the corresponding slice images have similar features between this figure and Figure 3.1, panel (c). The general scale of structures are indeed comparable, as one can observe with the number of red blobs, which show high density regions. However, as the standardized cross–correlations will show in the next section, we need LOS densities that are higher than 234 to expect maps with high fidelity.

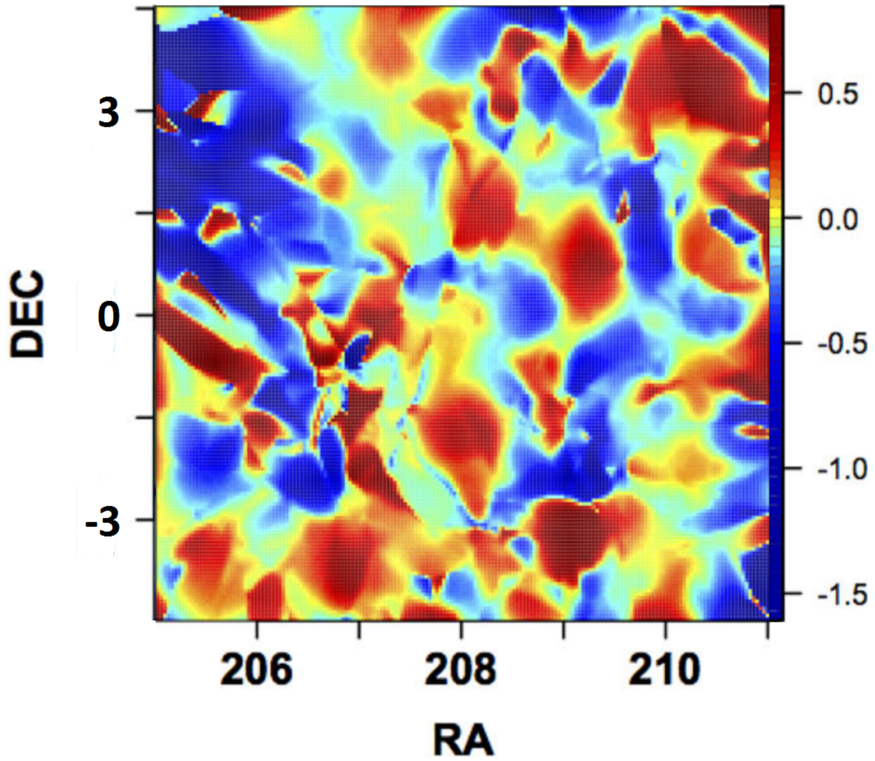


Figure 3.2: The IGM delta flux field in a volume similar to that of the simulation is inferred using a sample of 234 BOSS DR9 QSO spectra. Higher density regions correspond to the red color, as the negative of delta flux ($-\delta$) are the values in the visualization.

3.1.4 Standardized Cross–Correlation

The standard cross correlation was defined in §2.2.2 as $C_{12}(r) = \xi_{12}(r)/\sqrt{\xi_{11}(r) \cdot \xi_{22}(r)}$, where C_{12} is the cross correlation between the true field and the inferred field, while the other two terms are the auto–correlations of the individual fields. While slice images provide a qualitative means of estimating the performance of the local polynomial smoothing, standardized cross–correlations test the fidelity of the inference quantitatively as a function of scale: An asymptotic value of unity represents perfect reconstruction.

Amongst the three panels in Figure 3.3, the data set with the highest LOS density (1000 LOS, panel (a)) is the only one that yields acceptable results. The standardized cross–correlation value exceeds 0.7 at $\sim 40\,h^{-1}$Mpc, approaching values as high as ~ 0.85 at larger scales. The other two panels show that those LOS densities of 200 and 100 are not acceptable for interpolation with local polynomial smoothing. It is worth noting, however, that panels (b) and (c) represent data sets with LOS densities that are below that of BOSS. Nonetheless, they are useful for setting an approximate lower bound for the performance of this interpolation technique. We also provide auto–correlations of the data sets with the same LOS density levels in Figure 3.4. For comparison, the right column contains the true field, smoothed with the same bandwidth that is available to the corresponding reconstruction (hr corresponds to ”high resolution”, representing the true field).

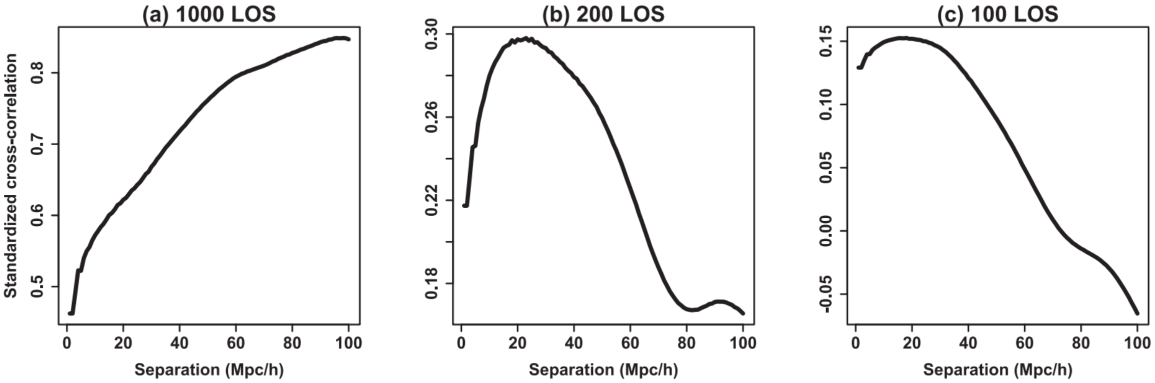


Figure 3.3: Standardized cross–correlations for varying LOS densities in the simulation volume. The data set with 1000 LOS performs the best, especially at scales larger than $40\,h^{-1}$Mpc.

3.2 Observational Requirements

In this section, we summarize some of the relevant ideas from (Lee et al., 2014b) that determine the spectral resolution power and areal LOS densities that are necessary for observational surveys to obtain certain observational IGM map resolutions ($\sim 1-5\,h^{-1}$Mpc) with the current generation of 8–10 m telescopes.

So far, we have shown that Wiener interpolation can be used to infer the underlying IGM field at high redshifts with simulations. The feasibility of using this technique with current observational surveys can be evaluated with the relation between the desired map resolution, areal LOS density and the observational spectral power:

$$\langle d \rangle \approx \sqrt{1/\mathrm{n_{LOS}}} \tag{3.3}$$

$$\langle d \rangle \approx \left(\frac{\mathrm{n_{LOS}}}{4200\,\mathrm{deg}^{-2}} \right)^{-1/2} \left(\frac{1+z}{3.25} \right)^{-3/2} \tag{3.4}$$

$$R > 1300 \left(\frac{1\,h^{-1}\mathrm{Mpc}}{\langle d \rangle} \right) \left(\frac{1+z}{3.25} \right)^{-1/2} \tag{3.5}$$

where we use the areal LOS density for $\mathrm{n_{LOS}}$. The ansatz that forms the basis of this technique is that the typical LOS separation $\langle d \rangle$ sets an approximate map resolution, also mentioned in (Caucci et al., 2008). Equation (3.4), expressed in h^{-1}Mpc,

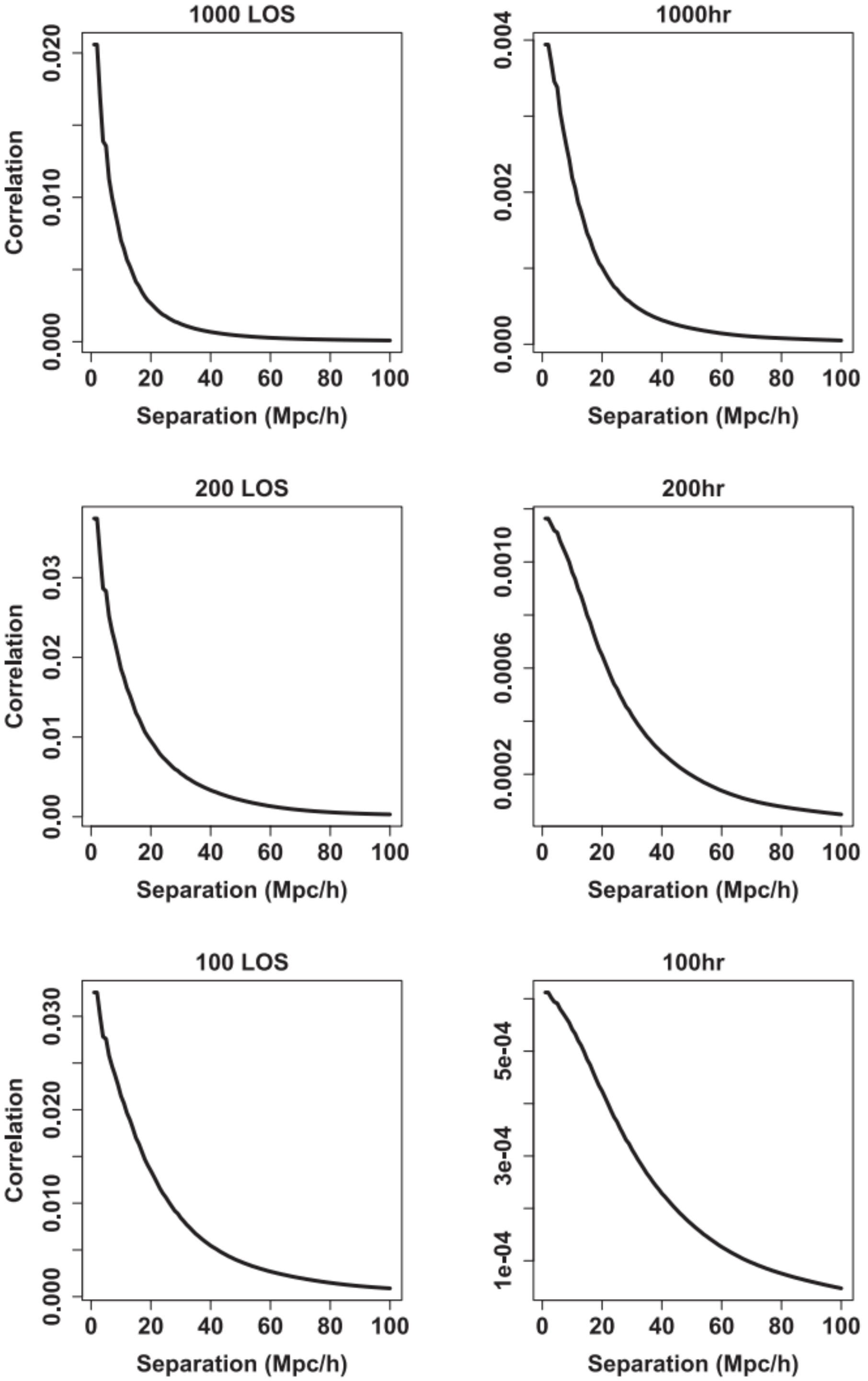


Figure 3.4: Auto–correlations are given for varying LOS densities in the simulation volume. Correlations of the reconstructed fields are on the left column. The figures on the right are prepared with the original field (hr corresponds to ”high resolution”, i.e. the original field), smoothed with the same bandwidth as the inferred field for that specific LOS density.

introduces a redshift dependence from the angular diameter distance. Since the areal LOS density for BOSS is $\sim$ 17 deg^{-2}, at the redshift where the quasar distribution is maximized ($z \sim 2.25$), Equation 3.4 yields $\langle d \rangle \approx 16\,h^{-1}$Mpc. This is very nearly our map resolution in §2, for the datasets with $\mathrm{N_{LOS}} = 1000$. Similarly, the targeted areal quasar density for MS–DESI ($\sim$ 45 deg^{-2}) yields $\langle d \rangle \approx 9.7\,h^{-1}$Mpc.

Finally, Equation 3.5 sets the minimum spectral resolution power ($R = \lambda/\Delta\lambda$) requirement as a function of the map resolution. For $\langle d \rangle = 17\,h^{-1}$Mpc and $z = 2.25$, Equation 3.5 gives $R > 76$, which is an order of magnitude lower than that of BOSS. For map resolutions of several h^{-1}Mpc, a spectral resolution power of at least $R \approx 1000$ is necessary, which is satisfied by current and future observational surveys such as BOSS, eBOSS and MS–DESI.

3.3 Conclusions

As an alternative to Wiener interpolation, we use local polynomial smoothing to interpolate both simulated Lyα skewers with the same simulation from §2, and we also apply this method to a sample of 234 QSOs from BOSS DR9. We observe a visual similarity between the BOSS reconstruction and the simulated inference with comparable LOS densities.

We evaluate the performance of local polynomial smoothing qualitatively by comparing slice images of the true field and the reconstructed field. General features of the field are captured well, especially for the reconstruction with 1000 LOS. However, comparing Figure 3.1 with the slice images from §2 (Figures 2.6 through 2.12), it is obvious that Wiener interpolation does a better job at mimicking the true field, both in terms of general fidelity and the comparable dynamic range. This observation is further supported by standardized cross–correlation analysis: For local smoothing, the only acceptable inference is with the data set that contains 1000 LOS, whereas LOS densities of 100 and 200 both yield completely unacceptable fidelity at all scales. In §2, we had found that Wiener interpolation gives perfectly acceptable results with $\mathrm{N_{LOS}} = 200$ and above, even when edge artifacts are still present (Figure 2.3, panels (c) and (d)). However, for the final data release of BOSS, the LOS count for an observational volume that has the same size as the simulation is as high as 400 (Figure 2.1, black curve), and even higher for surveys such as eBOSS and MS–DESI. Hence, we expect that local polynomial smoothing should be an acceptable means of inferring the observational IGM density field. The advantage of the adaptable smoothing level due to the nature of local polynomial smoothing can be especially useful, since QSOs are unevenly spaced. We also checked that adding random Gaussian noise does not change the performance of the map significantly.

In order to resolve IGM maps that are close to the Jeans scale, high areal LOS densities of the order of 1000 deg^{-2} will be necessary (e.g.Lee et al. (2014b), Figure 4). Currently, map resolutions of $\sim 20\,h^{-1}$Mpc are accessible to BOSS, due to the high QSO density and spectral resolution. However, it is possible to increase this resolution locally by limiting the reconstruction to regions where the QSO counts are relatively higher, as we will demonstrate in the next chapter.

Chapter 4

3D mapping of the Intergalactic Medium with the SDSS-III DR12 Lyman-α Forest

4.1 Reconstruction

We follow the Wiener interpolation method (Pichon et al., 2001; Caucci et al., 2008; Lee et al., 2014b,a; Ozbek et al., 2016) to carry out the reconstruction of the DR 12 flux field. As an alternative to this interpolation method, a nonparametric methodology can also be used, e.g. local polynomial smoothing (Cisewski et al., 2014).

Lyα skewers from DR12 are placed into a grid and a data column vector $\mathbf{D}$ is constructed from it, which contains the delta flux (δ_F) information. In order to negate the effects of peculiar velocities, we work in redshift space. The maps in this study are made with the delta flux field $\delta_F = (F/\langle F\rangle) - 1$ directly, where the flux F is defined as follows: $F/F_0 = e^{-\tau}$, where F is flux at a certain point in space, F_0 is unabsorbed flux, τ is the optical depth and $\langle F\rangle$ is calculated from all spectra. In (Caucci et al., 2008), the authors had opted to work with the density field of the simulation, however, it is physically equivalent to use the flux field instead, as these quantities are related:

$$\delta = \frac{1}{\alpha}\log(\frac{\tau}{A(\bar{z})}) \tag{4.1}$$

where $\delta \approx \frac{\rho-\bar{\rho}}{\bar{\rho}}$ is the density contrast, and α and $A(\bar{z})$ are factors that are redshift dependent.

We place the data vector $\mathbf{D}$ into a cube that is 8892 comoving h^{-1}Mpc long on a side and contains 500^3 regularly spaced voxels, which results in a cell spacing of 17.82 h^{-1}Mpc. The three–dimensional field $\mathbf{M}$, a collection of cell values which span the whole cube volume, is inferred from $\mathbf{D}$ using Wiener interpolation, and the data–data ($\mathbf{C_{DD}}$) and map–data ($\mathbf{C_{MD}}$) covariance matrices are as defined in §2.1.3:

$$\mathbf{M} = \mathbf{C_{MD}} \cdot (\mathbf{C_{DD}} + \mathbf{N})^{-1} \cdot \mathbf{D} \tag{4.2}$$

$$\mathbf{C}(x_1, x_2, \mathbf{x_{1\perp}}, \mathbf{x_{2\perp}}) = \sigma^2 \times \exp\Big(-\frac{(x_1 - x_2)^2}{L_{||}^2}\Big) \times \exp\Big(-\frac{|\mathbf{x_{1\perp}} - \mathbf{x_{2\perp}}|^2}{L_{\perp}^2}\Big) \tag{4.3}$$

where the noise matrix $\mathbf{N}$ is again assumed to be additive, diagonal and therefore uncorrelated, since we do not include the covariant terms between different cells. The non–zero entries of $\mathbf{N}$ are inversely proportional to the square root of the number of pixels in each cell. The distances between pixels are $(x_1 - x_2)$ along the direction parallel to LOSs, and $|\mathbf{x_{1\perp}} - \mathbf{x_{2\perp}}|$ denote the distances that are perpendicular. $L_{||}$ and $L_{\perp}$ are the correlation lengths parallel and perpendicular to the LOSs, and the variance σ^2 is calculated directly from the field.

After binning the cube volume of comoving $8892^3\ h^{-3}$Mpc^3 to 500^3 cells, we split it into smaller chunks that contain 5^3 cells to carry out the reconstruction in parallel with our Fortran 90 code, followed by stitching the cubes together into the original volume, which speeds up the computation tremendously. A buffer length of 80 h^{-1}Mpc exists between the chunks, which is slightly greater than the effective resolution of the map ($\sim 60\ h^{-1}$Mpc, explained in the next section), to diminish edge artefacts. We have checked that changing the number of cells (e.g. to 250^3, used for 3D visualizations) does not significantly alter the fidelity of the reconstruction.

In (Caucci et al., 2008), it was shown that for scales greater than $1.4\langle d_{\rm LOS}\rangle$, general features of the field can be recovered well. In order to avoid fictitious structures which are finer than that resolution, we smooth the field with an isotropic 3D Gaussian filter with a sigma of $\sigma_S = 1.4\langle d_{\rm LOS}\rangle$ after the reconstruction (see Equations 4.5 and 4.6). Evaluating these equations at $z = 2.5$ yields $\sigma_S \sim 60\,h^{-1}$Mpc, which is our fiducial smoothing level throughout this chapter.

4.2 Quasars in DR12 and the Flux Field

For producing QSO spectra from CCD images, we follow the usual procedure of subtracting the background noise and fitting the mean spectrum to a low order polynomial (Hui et al., 2001):

$$\bar{N_Q^\alpha} = \sum_a C^a p^{a\alpha} \tag{4.4}$$

where $\bar{N_Q^\alpha}$ is the quasar count, which is a sum of the raw count and the additive background noise, p_0, p_1, p_2 ... are the polynomials and the C^a coefficients are determined with using a linear estimator from the raw quasar counts.

We use the 12th iteration of the Data Release (DR12) from SDSS–III (Pâris et al., 2012; Alam et al., 2015). Quasars in DR12 sample the density field very sparsely. In

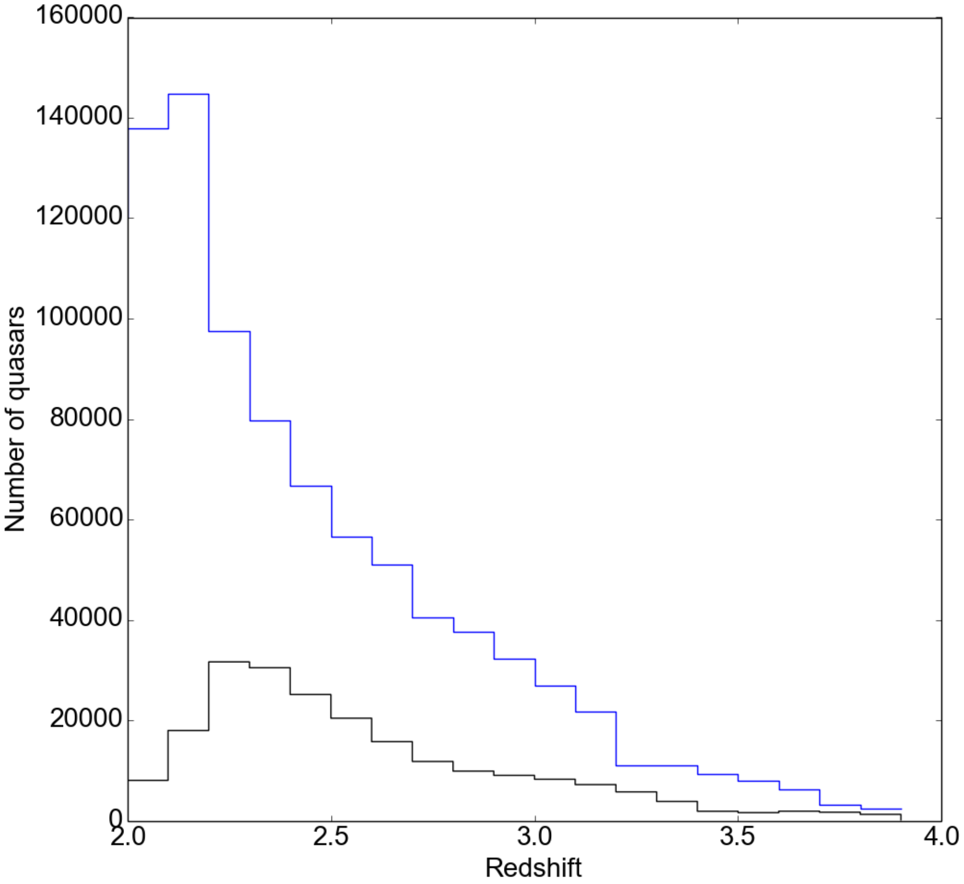


Figure 4.1: Quasar distribution in our data set as a function of redshift is shown by black color. The blue curve depicts the quasars probing the density field as a function of redshift.

the pre-reconstruction stage, we find 217161 quasars, with the peak at $z \sim 2.25$, shown with the black color in Figure 4.1.

Quasars probe the IGM along one–dimensional Lyα skewers typically for a redshift interval $\Delta z \sim 0.5$ in the range $2 \lesssim z \lesssim 3.5$ along lines of sight. For example, at $z = 3$, each individual spectrum contains one–dimensional density information for a ~ 400 h^{-1}Mpc skewer, starting about 100 h^{-1}Mpc in front of a quasar. Accordingly, the volume illuminated by the quasars can be calculated as a function of redshift. In addition to the black curve which shows the true count of quasars, the blue curve in Figure 4.1 shows the density field probed by quasars as a function of redshift for the entire volume. This is estimated by counting all the quasars within a redshift delta of $\Delta z = 0.5$ further from the observer. Using this information, the areal line of sight

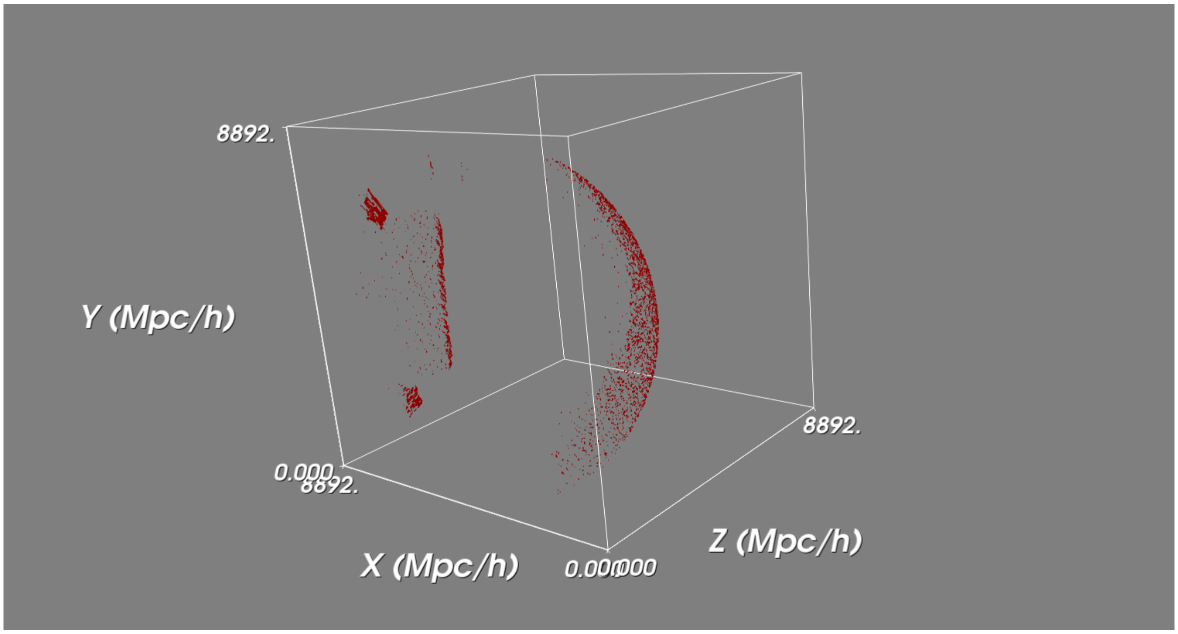


Figure 4.2: 3D Quasar distribution in the reconstruction cube.

density $\mathrm{N}_{\mathrm{areal}}$ and the mean line of sight (LOS) separation $\langle d_{\mathrm{LOS}} \rangle$ can be calculated:

$$\mathrm{N}_{\mathrm{areal}} = \frac{n_{QSO} \times (\mathrm{A}_{\mathrm{total}}/\mathrm{A}_{\mathrm{DR12}})}{4\pi r^2} \tag{4.5}$$

$$\langle d_{\mathrm{LOS}} \rangle = \frac{1}{\sqrt{\mathrm{N}_{\mathrm{areal}}}} \tag{4.6}$$

where $\mathrm{A}_{\mathrm{total}}$ is the total sky area (41253 square degrees), $\mathrm{A}_{\mathrm{DR12}}$ is the DR12 sky coverage (10400 deg^2) and r is the comoving radial distance, and $\langle d_{\mathrm{LOS}} \rangle$ is the mean line of sight separation, which sets a natural resolution for the resulting maps. The number of quasars (n_{QSO}) can be read off from Figure 4.1 as the blue curve for a specific redshift interval.

Quasars are highly clustered objects (Croom et al., 2004; Myers et al., 2006; Oogi et al., 2016), with similar clustering properties as galaxies at low redshifts: For $z < 2.8$, the correlation function can be fit with a power law $(r/r_0)^\gamma$, where $\gamma \approx 1.8$ and $r_0 \approx 5\,h^{-1}$Mpc. Hence, during the quasar search algorithm, it is inevitable that some pairs of quasars are binned into the same cell, thus decreasing the effective line of sight density. By labeling the furthest non–zero pixel along a spectrum as a quasar, we find 94876 such objects when the data were binned in a 500 cell cube. If one increases the cell size for binning to 250 cells along a side instead (as was done for 3D visualizations for memory considerations), this number decreases to 24137 (Figure 4.2).

The line of sight direction is along z, and the observer is located at $(x_{obs}, y_{obs}, z_{obs}) =$

$(4446.039, 4089.707, 843.561)\, h^{-1}$Mpc. Following the arguments in (Caucci et al., 2008) and using equations 4.5 and 4.6, we choose $L_{\perp} = L_{\parallel} = \langle d_{\mathrm{LOS}} \rangle$ and $\sigma_S \sim 1.4\langle d_{\mathrm{LOS}} \rangle$, i.e., $40\, h^{-1}$Mpc for the isotropic correlation length and $60\, h^{-1}$Mpc for the isotropic Gaussian smoothing kernel standard deviation size, respectively.

There are some minor differences between earlier work with cosmological hydrodynamical simulations (Caucci et al., 2008; Ozbek et al., 2016) and this study which uses observational data, however, we believe they are not significant enough so that the fundamental ideas are still valid. As in the previously mentioned articles, we assume that lines of sight towards quasars are parallel to each other, given that the quasars are sufficiently far away and the large redshift interval along the box that has a comoving length of 8892 h^{-1}Mpc on a side. This "distant observer" idea is shown to be a good approximation (White, 2014). It is also neglected that the quasars are located at different points along the lines of sight, trivially, unlike the simulations in (Caucci et al., 2008; Ozbek et al., 2016).

In §2, we had successfully created simulated large scale maps of the IGM using LOS densities as low as $\mathrm{N_{LOS}} = 200$ (see Figure 1 in (Ozbek et al., 2016)). Here the volume density $\mathrm{N_{LOS}}$ simply means the number of simulated Lyα skewers in a volume of $400^3\mathrm{Mpc}^3\mathrm{h}^{-3}$. For the observational reconstruction, although the skewers do not probe the entire LOS direction (z), we can still calculate an effective volume density $\mathrm{N_{LOS}}$ by simply dividing the number of non–zero pixels by the number of cells on a side. This is done in Section 4.3, for a subcube in the reconstruction box, which yields a LOS density comparable to that of BOSS, which is $\mathrm{N_{LOS}} = 184$. A typical BOSS pixel is about 1.5 h^{-1}Mpc wide, and assuming a signal to noise ratio of S/N=1, which becomes S/N $\sim$ 5 when rebinned to our pixel size in this chapter, and using Table 3 in (Ozbek et al., 2016), the RMS error for the present study can be estimated as $\sim$ 28 per cent.

When making preliminary IGM maps, the redshift dependence of the field is obvious (Figure 4.3). The red dashed curve shows the mean. The gradual decrease in flux is monotonic in the comoving radial range given in the figure, which corresponds to the redshift range $2 < z < 2.5$. We remind the reader that the evolution of the optical depth in the given range is $\tau_{eff} = A(1+z)^b$ (Meiksin, 2006), where A and b are constants. Since the flux and the optical depth are inversely correlated, this decreasing trend is expected.

Most flux contrast values in Figure 4.3 are in the immediate vicinity of zero, which appears as a dark grey horizontal strip. By visual inspection, more positive values are observed, compared to the negatives, in the outlier points. Also, positive outlier flux values close to $z \sim 2$ outnumber the ones close to $z \sim 2.5$, which may explain the decreasing mean. We bin the data into redshift bins to subtract out the redshift evolution of the field. By subtracting the mean flux at every redshift bin, we make sure that the flux contrast values average out to zero at every redshift, as intended, and continue all subsequent analysis after this normalization.

Instead of using Cartesian coordinates, as mentioned earlier in this section, a more

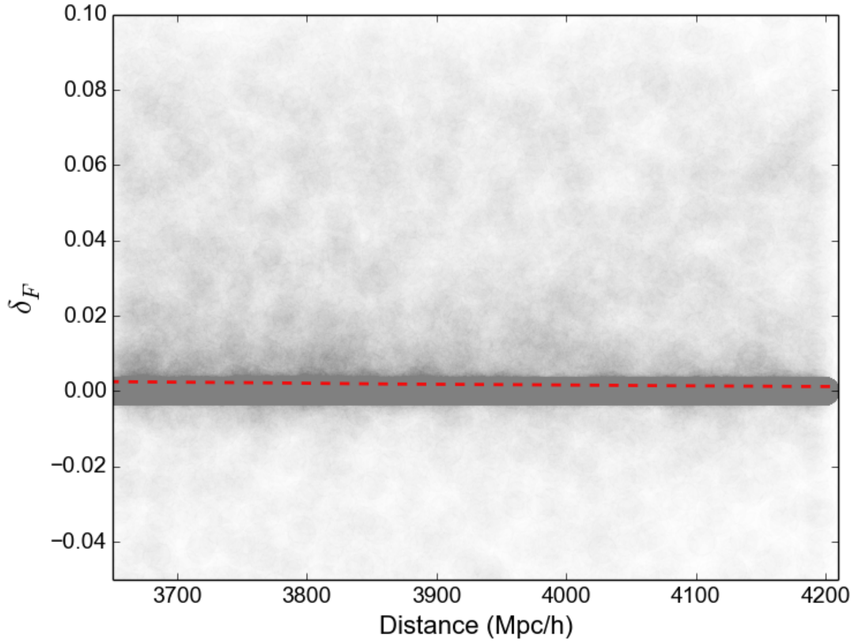


Figure 4.3: The mean flux at a given comoving distance from the observer is shown with the red curve. By subtracting out this evolution, it is ensured that the average flux is normalized to zero at every redshift bin.

natural means of studying the DR12 flux field is to analyse the redshift evolution of it, prior to smoothing, where we apply a redshift cut as $2 < z < 2.5$ while limiting the right ascension (RA) and declination (DEC) range. The redshift interval is approximately converted to a delta comoving radial distance, which varies from 0 h^{-1}Mpc at $z = 2.0$ and 464 h^{-1}Mpc at $z = 2.5$. The distance coverage of the RA range in the produced maps are also trivially calculated. Five such stripes across the North Galactic Cap (NGC) are shown in Figure 4.6, where we bin the field using RA, DEC and z intervals while considering only non–zero pixels: The right ascension range is $120° < \mathrm{RA} < 240°$ for all five panels, while the distance along the stripe in the circumferential direction varies at DEC = 10° along the x axis in Figure 4.6. Darker shades of red color correspond to underdense regions in the stripe graphs.

Comparing flux levels amongst varying declinations, it is observed that the mean flux is minimized when the declination is decreased to its lowest value of 10°, which may mean that the overall absorption in the Lyα forest is the greatest: The mean flux for DEC = 10° through DEC = 30°: 0.00582, 0.00665 and 0.00724, respectively. Comparing the recent BOSS footprints across new data releases from SDSS, one notices that these NGC stripes were mostly not covered in the given RA range in earlier data releases such as DR9 (Pâris et al., 2012), which had covered only 3275 deg^2. A visual inspection of the DR12 BOSS NGC stripes (Figure 1.8) shows clearly that the one at DEC = 10° is the most complete of all three, which may explain the trend in the overall mean. Furthermore, the fact that the standard deviation also shows an increasing trend from left to right (0.119, 0.129 and 0.132, respectively) suggests that there exists more contrast and hence more structure at higher declination, which can also be seen with visual inspection.

After applying the isotropic Gaussian smoothing and the linear bias to the reconstructed DR12 flux field, we can clearly see the large–scale structure (Figure 4.4. By visual inspection, it's readily noticable how closely the DR12 IGM field and the quasar distribution are related. Quantitatively, their cross–correlation was investigated with DR11 (Font-Ribera et al., 2014) and auto–correlation of the IGM in (Slosar et al., 2011). For the the Lyα auto–correlation graph (Figure 4.5), we only consider non–empty pixels in the cube, apply a redshift cut for the DR12 field as $2 < z < 2.5$ and use the ngtot = 500 data set, as the one with ngtot = 250 seems to give misleading correlation information for small scales due to binning, i.e. for scales less than 40 h^{-1}Mpc. From our work with simulations, we know that we can expect the field to be recovered well at large scales (Figure 2.3).

4.3 Subcube

Aside from reconstructing the delta flux field for the entire cube, we can also "zoom-in" on smaller sub–cubes in order to obtain a better resolution for the reconstruction and to focus on local statistics. The volume of this sub–cube is $\sim$ 10000 times smaller than that of the entire construction volume. For this sub–cube, we follow a similar computational procedure as we did for the entire cube, by putting the observational pixels into a grid and carrying out Wiener interpolation in order to infer the field. The entire cube, which is 8892 h^{-1}Mpc long on a side, contains 500 cells on a side which results in a resolution of 17.82 h^{-1}Mpc (or 250 cells on a side for some 3D visualizations to make it computationally tractable, with a cell spacing of 35.71 h^{-1}Mpc). By focusing on sub–cubes, we increase that resolution slightly to 13.62 h^{-1}Mpc and can study local statistics like the auto-correlation and local extrema points, and also draw comparisons with our previous work with hydrodynamical cosmological simulations (Ozbek et al., 2016).

The observational sub–cube, which is 422.3 h^{-1}Mpc long on a side, has the furthest corner from the observer at a comoving distance 3670.063 h^{-1}Mpc away from the observer, at a redshift of $z = 1.98$. The number of non–empty pixels along the cube is 5897, which results in an effective number of LOS density of 184 $\mathrm{N_{LOS}}$, as there are 32 cells on a side. For the 3D interpolation, we set the correlation lengths and the isotropic Gaussian smoothing standard deviation to be 40 h^{-1}Mpc. The mean line of sight separation $\langle d_{\mathrm{LOS}} \rangle$ sets a natural size for the correlation lengths and the smoothing length, which can be explained following the arguments in (Caucci et al., 2008): $L_\perp = L_\parallel = \sigma_S = 1.4\langle d_{\mathrm{LOS}} \rangle_{eff}^{subcube} = \frac{L_{box}^{subcube}}{\sqrt{\mathrm{N_{LOS}}}} \sim 40\, h^{-1}$Mpc. Since this resolution

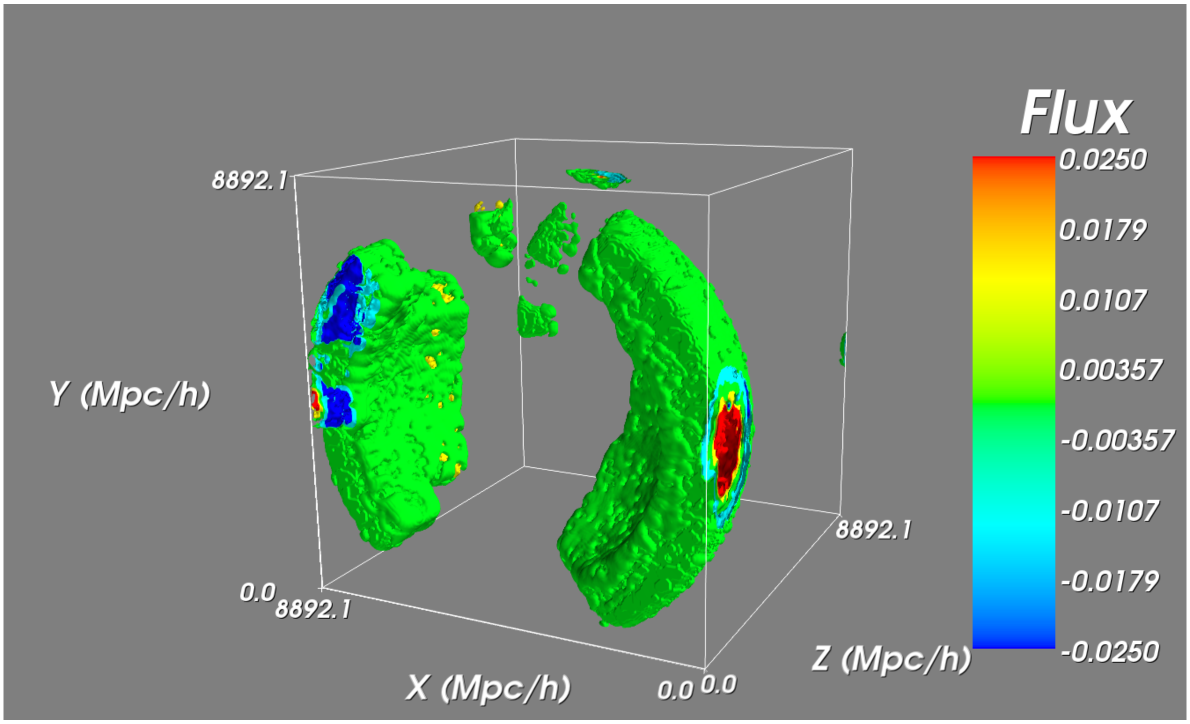


Figure 4.4: Delta flux 3D distribution in the reconstructed cube. Blue color shows overdense regions, while red denotes underdense regions in the IGM. We are showing five isocontour surfaces at the mean, $\pm\sigma$ and $\pm 2\sigma$.

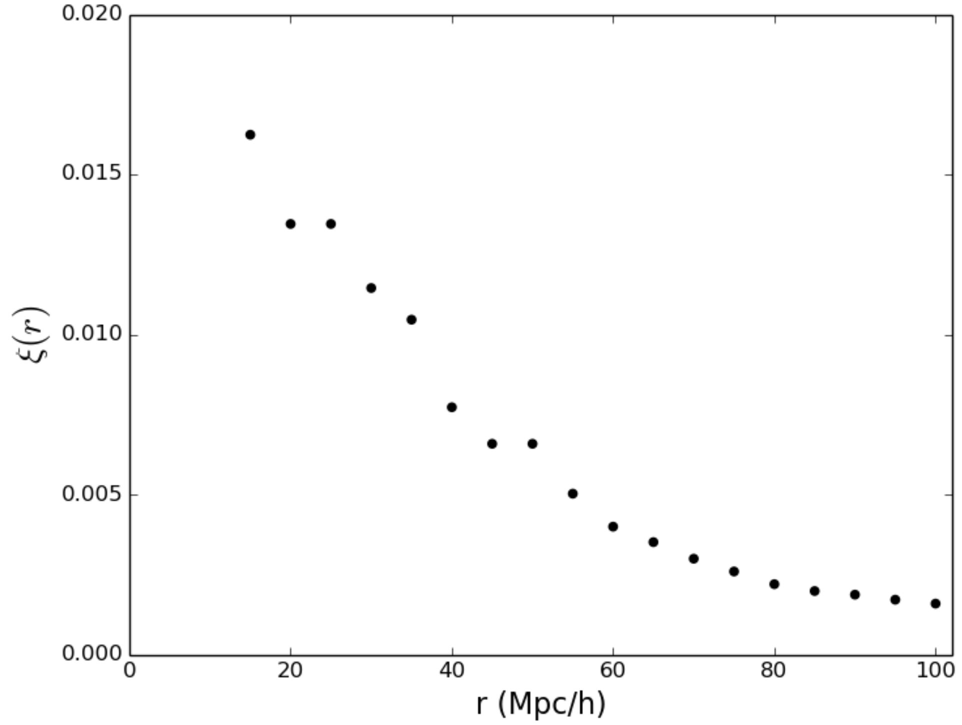
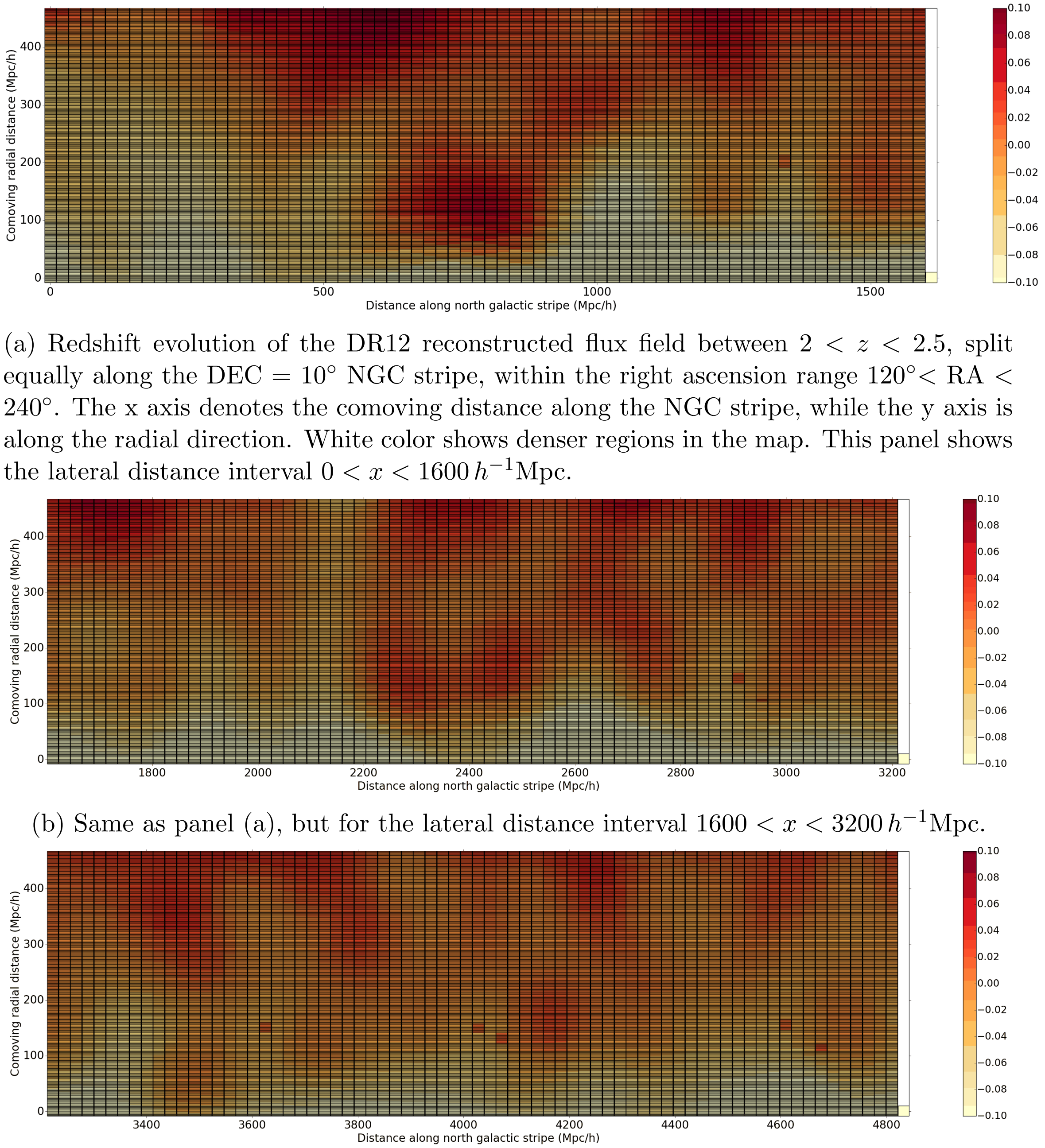


Figure 4.5: Auto–correlation of the DR12 flux field, where we limit the pixels in the cube between $2 < z < 2.5$.

is comparable to the one obtained with the BOSS space density of sightlines and also given the fact that the size of the simulation cube in Ozbek et al. 2016 and the redshift is also similar (400 h^{-1}Mpc, $z = 2$), it is reasonable to compare the statistical properties of those fields (see Figure 1 in our previously mentioned simulation paper (Ozbek et al., 2016) for a comparison of $\mathrm{N}_{\mathrm{LOS}}$ for recent surveys, and Figures 5 through 11 for maps with different LOS densities). In that paper, we had studied local extrema statistics of the field, similar to the analysis in §5.1. In the observational sub–cube, we find 9 local density maxima, which is in perfect agreement with the simulation cube statistics in (Ozbek et al., 2016). Furthermore, we find 7 local density minima, suggesting that the data distribution is slightly skewed, which is compatible with the overall trend that will be provided in the next chapter (Table 5.1).

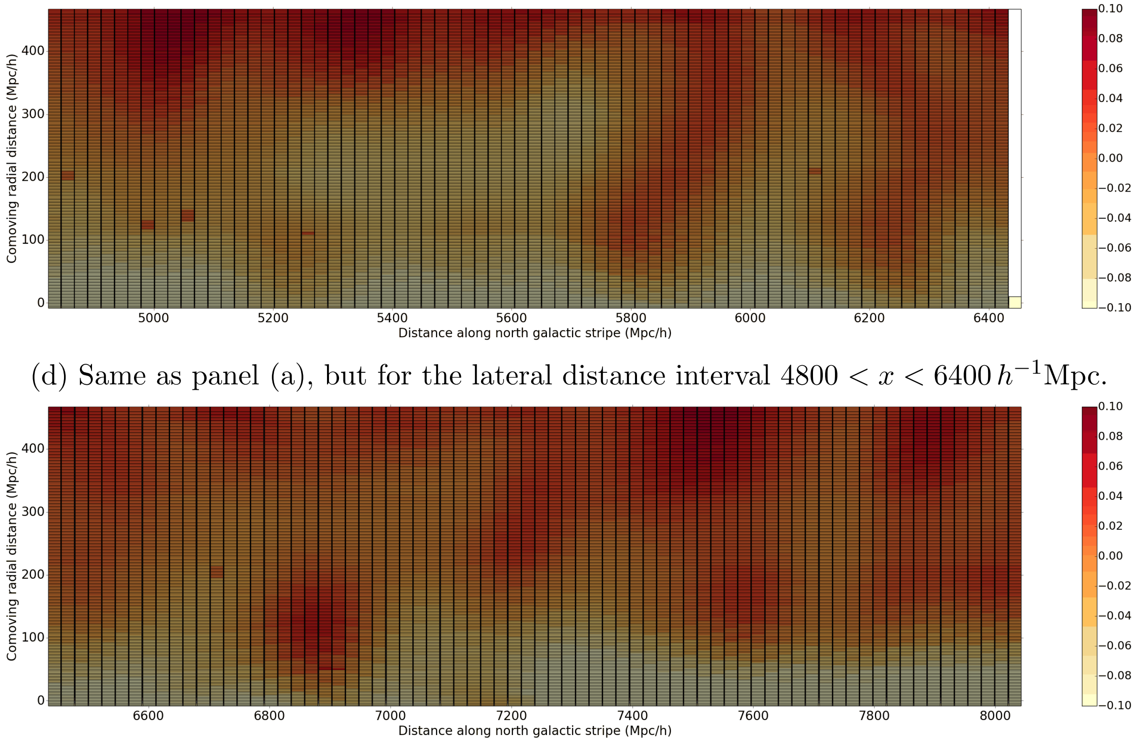
In Figure 2.16, we had shown 3D visualizations of the original simulation field and its reconstruction in panels a and b. This volume has a similar value to that of the DR12 sub–cube we analyzed, 400^3 and 422^3 h^{-3}Mpc^3, respectively. For the simulation, we had noticed that the linear bias for the simulated reconstructed field was deviating from unity slightly, and that it was approaching unity with increasing sightline density. By comparing the $\mathrm{N}_{\mathrm{LOS}}$ values between the simulation (Ozbek et al., 2016) and DR12, we decide on a proper linear bias and correct the inferred field with it. In this particular case the bias is 1.5. The analysis is carried out after smoothing



(a) Redshift evolution of the DR12 reconstructed flux field between $2 < z < 2.5$, split equally along the DEC = 10° NGC stripe, within the right ascension range 120°< RA < 240°. The x axis denotes the comoving distance along the NGC stripe, while the y axis is along the radial direction. White color shows denser regions in the map. This panel shows the lateral distance interval $0 < x < 1600\,h^{-1}$Mpc.

(b) Same as panel (a), but for the lateral distance interval $1600 < x < 3200\,h^{-1}$Mpc.

(c) Same as panel (a), but for the lateral distance interval $3200 < x < 4800\,h^{-1}$Mpc.

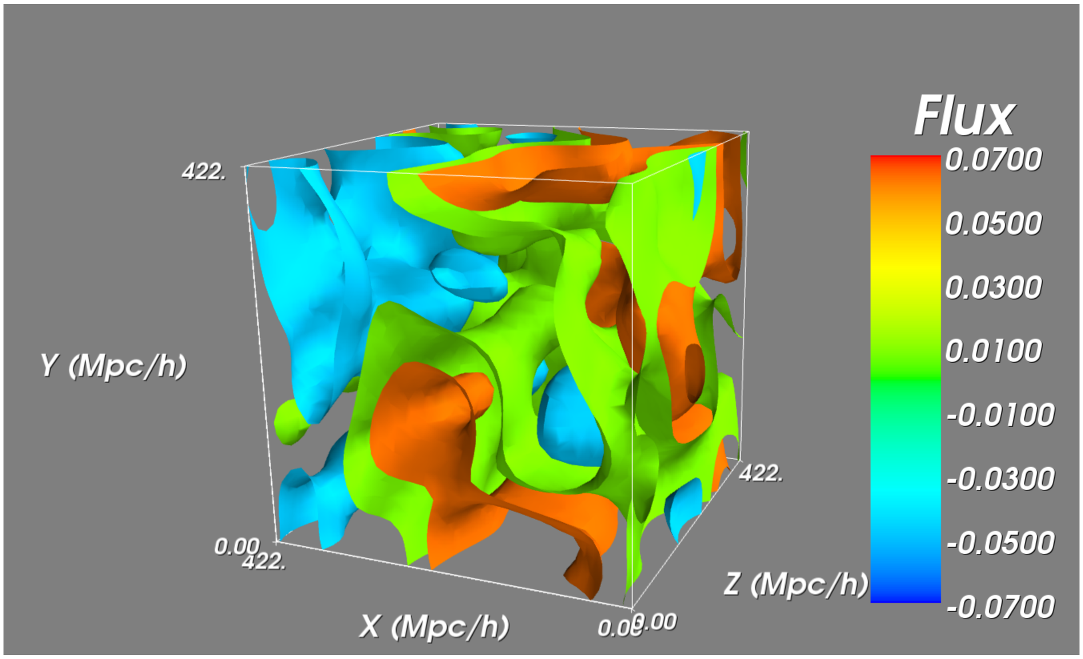


(d) Same as panel (a), but for the lateral distance interval $4800 < x < 6400\,h^{-1}$Mpc.

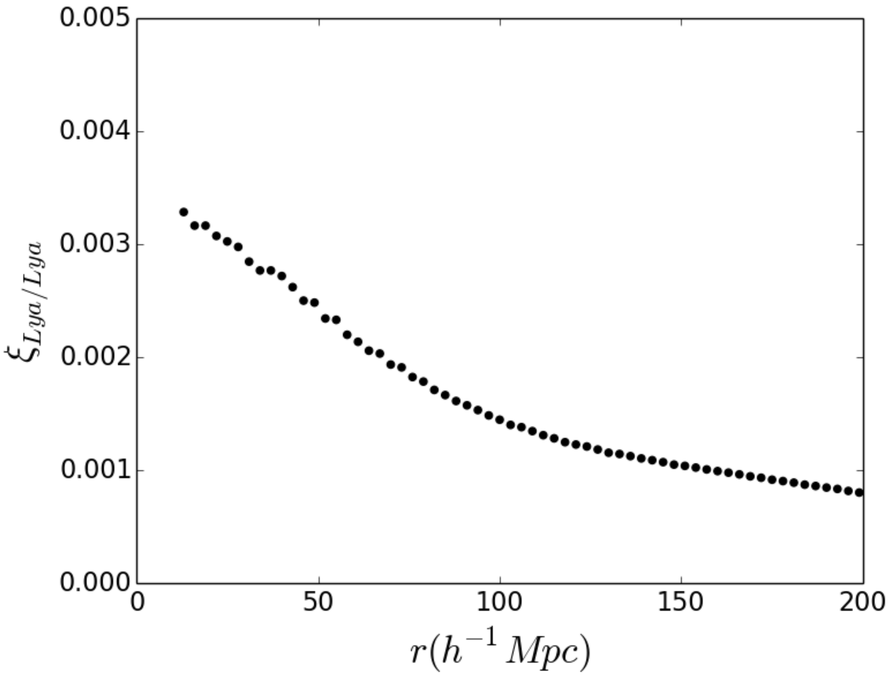
(e) Same as panel (a), but for the lateral distance interval $6400 < x < 8055\,h^{-1}$Mpc.

Figure 4.6: IGM field along the DEC=10° NGC stripe.

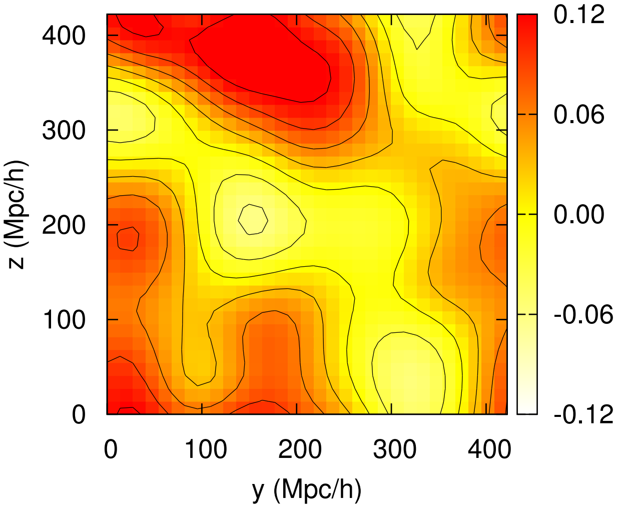
and applying the linear bias (Figure 4.7, panel (a)). The auto–correlation of the field quickly approaches zero (Figure 4.7, panel (b)). In general, it is not possible to observe the BAO signal in the correlation analyses in this thesis due to the large smoothing levels, which erases that feature. Finally, in Figures 4.7c and 4.7d, we show slices that split the sub–cube into two equal halves. The first slice (c) is an "x slice", along a plane that is perpendicular to the LOS direction. The other one is parallel to the LOS direction, a "z slice", given in panel (d). Due to the inverse relation between the flux and density, red color shows underdense regions in both the 3D visuals and the slice images.



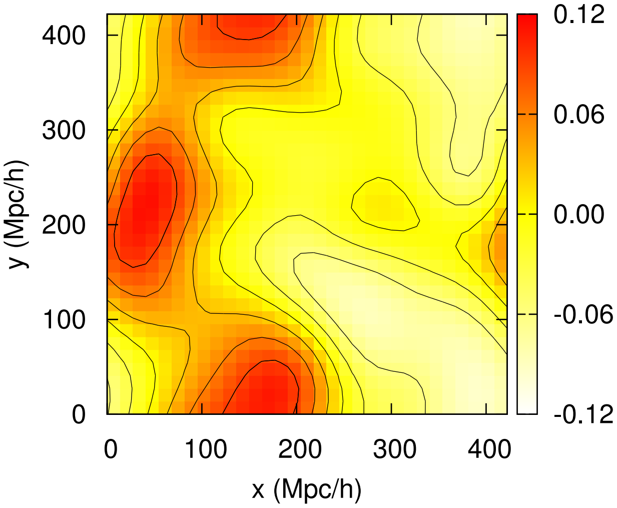
(a) Subcube isosurfaces extracted from the whole DR12 volume.



(b) Auto–correlation of the subcube.



(c) A slice image across the x direction.



(d) A slice image across the z direction.

Figure 4.7: A 422^3 $\text{Mpc}^3\text{h}^{-3}$ volume extracted from the whole DR12 reconstruction cube. 3D flux distribution and the auto–correlation are given in panels (a) and (b). We show slices taken from two perpendicular planes splitting the cube into two equal halves across the center in (c) and (d): An "x slice" and a "z slice", respectively. Red color denotes underdense regions in both the 3D visuals and the 2D slice images.

4.4 Summary and Conclusions

We have shown that using the Lyman alpha forest data from BOSS DR12, the large–scale structure of the Universe for the redshift interval $2 < z < 2.5$ can be mapped using Wiener interpolation. Comparing the observational LOS density with the ones used in our paper on simulated flux fields (Ozbek et al., 2016), we decide on an empirical linear bias of 1.6 and an isotropic smoothing level of 60 h^{-1}Mpc, which sets a natural spatial resolution. Equations 4 and 5 show the dependence of the map resolution on the number of quasars discovered in the survey and the survey sky coverage.

217161 QSO spectra are placed into a 8892^3 h^{-3}Mpc^3 comoving volume which is binned into 250^3 cells. After subtracting out the redshift evolution, we provide the auto–correlation of the δ_F field and its cross–correlation with the QSO field.

The IGM is expected to show mildly non–linear behavior in the redshift range covered in this thesis, hence we look for non–Gaussianity by comparing local extrema statistics between the observational flux field and a simulated Gaussian field, which agree within 28.0 per cent. In order to obtain finer resolutions and draw quantitative statistical comparisons with our paper (Ozbek et al., 2016), we also "zoom–in" on a subcube that is 422 h^{-1}Mpc long on a side, which mimics the geometrical properties of the simulated field. The local peak statistics of the subcube are observed to have perfect statistical agreement with that of the cosmological hydrodynamical simulations provided in our previously mentioned paper, and we also provide the auto–correlation of the flux field in the subcube.

Chapter 5

Superclusters

In order to identify overdense and underdense regions in the large–scale structure, we examine local peaks the in the flux field and carry out percolation analysis in the following sections. We also study the filamentarity of the field and study various statistics of the superclusters we find in the volume.

5.1 Introduction

It is predicted by inflation that the large–scale structure on scales larger than the non–linear regime have evolved from Gaussian primordial perturbations (Bardeen et al., 1983). The linear evolution of the cosmic baryonic matter on large scales and its extension to the Lyα forest at reveals that the IGM is mildly non–linear in the redshift range which is relevant for this study (Fang et al., 1993; Bi and Davidsen, 1997). Even at low redshifts, the galaxy distribution is shown to display topological properties of a mostly Gaussian random field, at least on scales larger than 8 h^{-1}Mpc (Colley, 1997).

Finding local extrema points provides a means of identifying objects, and it can be used to constrain cosmology (Bardeen et al., 1986; Croft and Gaztañaga, 1998; De and Croft, 2007, 2010). Locations of these extrema are candidates for voids and superclusters. We identify a local peak as a pixel that has the greatest absolute value of the flux amongst the 26 adjacent pixels that surround it. In general, the number of peaks can suggest how evolved a field is, as overdense regions tend to grow more dense, while underdense regions tend to grow more underdense over time.

5.2 Comparison with a Linear Density Field

In order to test the deviation of the reconstructed DR12 IGM field from initial Gaussian fields, we create a simulated linear density field with our Fortran 90 code. A random realization of the random field on a grid is created by picking Fourier modes

Data	**Smoothing**	**Maxima**	**Minima**
DR12	None	25360	25708
DR12	$\sigma_S = 40\,h^{-1}$Mpc	4888	5101
DR12	$\sigma_S = 60\,h^{-1}$Mpc	1295	1645
Linear Field	None	106858	106480
Linear Field	$\sigma_S = 40\,h^{-1}$Mpc	2800	2731
Linear Field	$\sigma_S = 60\,h^{-1}$Mpc	932	922

Table 5.1: Extrema points analysis for DR12 and Linear Fields flux fields. The optimal level for similar statistics occurs when $\sigma_S = 60\,h^{-1}$Mpc, the fiducial smoothing standard deviation level.

from the CDM power spectrum and then performing a fast Fourier transform. The cell spacing on the grid is regular and it is matched with that of the observational field. However, the box size is much smaller. The linear field realization contains a significantly less number of pixels than the reconstructed DR12 data: The DR12 flux field is binned to 500^3 (or 250^3 when computationally preferable) pixels after the reconstruction, while the linear field only has 128^3 pixels. After counting the number of non–empty pixels in the DR12 field that were used for local extrema analysis, we scale the numbers found from the linear density field accordingly. In Table 5.1, we give the number of minima and maxima for the DR12 flux field for different smoothing levels and we also provide the same information for linear density fields in order to compare these statistics. In order to avoid fictitious structures at small scales, it is important to smooth the recovered IGM field. Therefore, besides no smoothing, we also use two additional smoothing levels: $\sigma_S = 40$ and $\sigma_S = 60$ h^{-1}Mpc.

Clearly, the local extrema statistics of the two fields, i.e. the numbers of local minima and maxima, are comparable when smoothed. The best agreement is observed to be achieved when the standard deviation size for smoothing is set to 60 h^{-1}Mpc, especially for the number of maxima, which agrees within 30 per cent. This suggests that at large values of the smoothing radius, the DR12 flux field is showing statistical properties which are compatible with a Gaussian random field. The numbers we find for the DR12 flux field are only slightly higher, which suggests that structures have not evolved significantly, hence showing mildly non–linear field behavior. In §2, we had already shown that simulated fields inferred with Wiener interpolation were observed to preserve local extrema statistics.

In order to visually compare the DR12 flux field and the linear field, we also provide probability density functions (PDFs) in Figure 5.1. The σ values correspond to the standard deviation of the linear field. The distributions agree to a good extent, especially around mild overdensities (negative flux values).

The local peak analysis can be extended with percolation techniques (a friends of friends algorithm) to provide candidates for superclusters, which are analyzed in

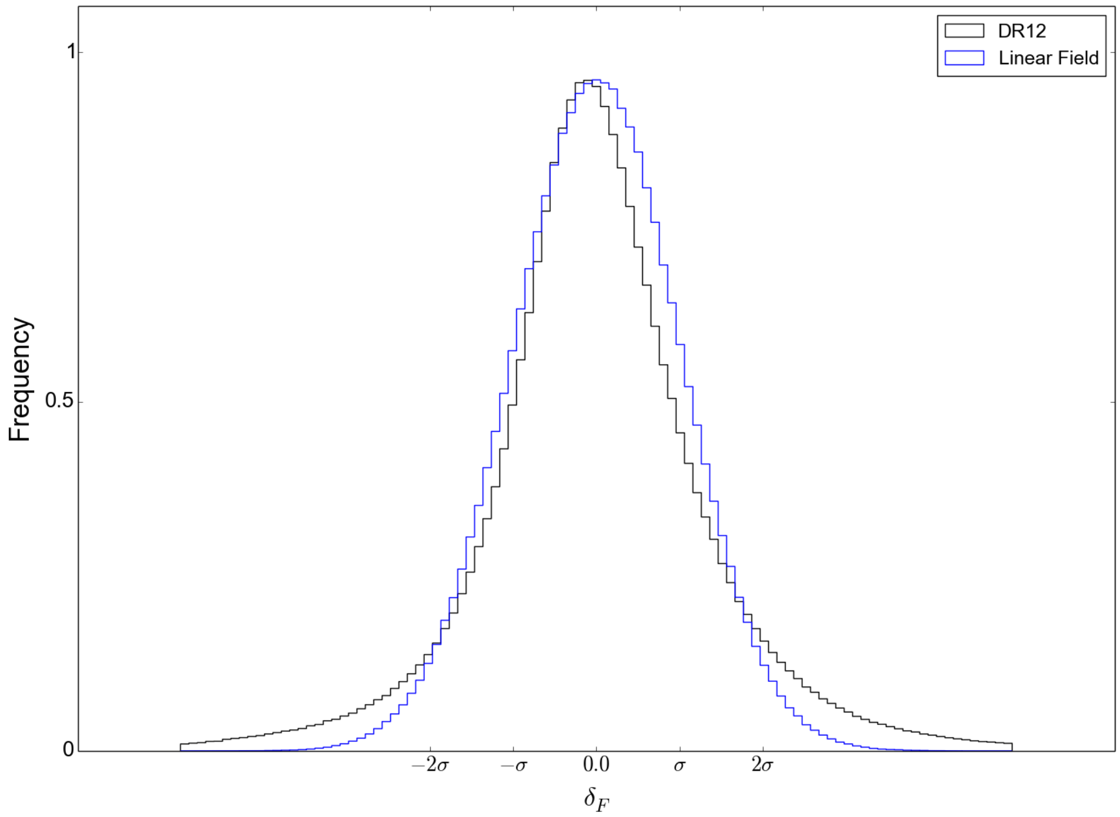


Figure 5.1: Probability Density Functions of the DR12 flux field and the linear field.

depth in the following section.

5.3 Percolation

5.3.1 Method

Percolation is a technique in statistical physics and mathematics which studies how clusters are connected in a random graph. Bond percolation, introduced by Broadbent and Hammersley (1957), is the classical problem in percolation theory, which describes the probability of whether an open path exists between two extreme points in a ”medium” for a ”fluid” to flow through. Here, fluid and medium can have different meanings, e. g., electrons migrating over an atomic lattice (Last and Thouless, 1971), molecules penetrating a porous solid, or superclusters and voids in the ΛCDM cosmology (Shandarin et al., 2004).

The percolation technique produces a friends of friends group catalogue for superclusters by identifying all cell pairs that are neighbors. After setting a cutoff value for the normalized flux field, all cells containing flux values below that threshold (i.e., cells that are denser) are tested for connectedness. We only search for superclusters, although the algorithm can also be used to examine voids. The minimum number of cells in a group is a free parameter, which is chosen to be 100 for all subsequent analyses. For computational tractability of the output, the data in the cube were rebinned to 250^3 pixels, resulting in a regular cell width of 35.71 h^{-1}Mpc. In view of this large cell spacing and the large number of minimum cells chosen, the size of the groups we can resolve with the percolation analysis is of the order of several hundred h^{-1}Mpc.

5.3.2 The Search for Superclusters

We use the previously mentioned percolation algorithm to identify dense regions in the observational volume. Aside from the minimum number of cells, the cutoff threshold is the other major parameter in the search for structures. We notice that by varying the threshold for flux δ_C, the number of superclusters changes rapidly. For greater values of δ_C, a large number of cells are eliminated due to not being dense enough, therefore discarding groups wherever they are connected, if the number of cells within that group is less than 100. Conversely, as δ_C is decreased, the number of groups are diminished for a different reason: Superclusters that are in proximity, but isolated for higher δ_C values, start to merge together into larger groups, an effect which is readily seen by visually inspecting the 3D figures of superclusters for varying δ_C values. We notice that the maximum number of groups is observed when the cutoff value is $\delta_C \sim 2.0\sigma$, where σ is the standard deviation of the field. Comparing this result with that of a Gaussian linear field yields similar behavior: the number of groups is maximized at $\delta_C = 1.5\sigma$.

We study the statistical properties of the superclusters we find quantitatively. In Figure 5.3, panels (a) and (b) show the RA and DEC locations of the superclusters. Most of them are located in the range $120° \lesssim \mathrm{RA} \lesssim 240°$, with a fairly random distribution in declination, which represents the region on the right hand side in the 3D volume, with the color range blue to light yellow, in Figure 5.2. We had already plotted some stripes of the data in this range in Figure 4.6. It is readily seen that there are no superclusters in the range $250° \lesssim \mathrm{RA} \lesssim 330°$, as this region at such low declinations is not included in the SDSS sky coverage. This also shows up as a big gap in Figure 4.4 between the two main blobs.

5.3.3 Morphology of Identified Superclusters

Since the theoretical prediction of pancake-like structures in (Zel'Dovich, 1970), planarity vs. filamentarity of structures remain an unresolved controversy. In the percolation analysis, most superclusters identified consist of several hundred cells. Visual inspection of 3D figures of superclusters suggests that the supercluster morphology has both filamentary and planar features (Figure 5.2). Numbers in the graph and the corresponding colors represent different supercluster groups, they do not contain any other physical information. It is worth noting that we find more superclusters with the smoothed data set than with the one with no smoothing (49 and 16, respectively), and they tend to be less massive, typically by an order or magnitude. All superclusters are detected at $z \sim 2.0$.

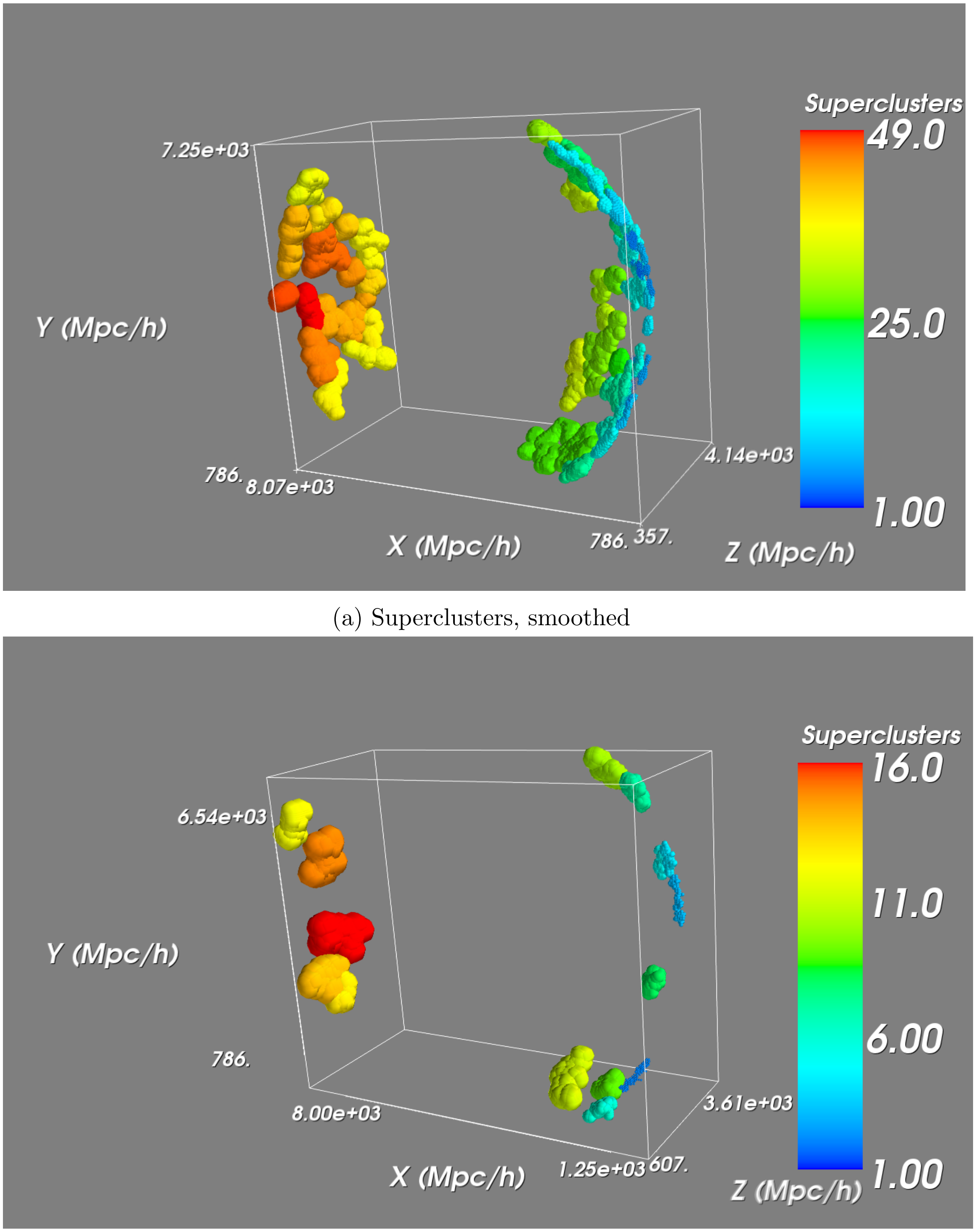
In order to inspect the morphology of superclusters quantitatively, one can study the correlation between the longest dimension of the supercluster and its mass (Figure 5.3, panels (c) and (d)). We identify the longest dimension in supercluster groups by computing the separation of the centers of the two most widely separated cells and adding one cell size to it. For the data set that has been smoothed with the fiducial Gaussian kernel size 60 h^{-1}Mpc, both a linear and a quadratic regression fit are acceptable for the scaling between mass versus longest dimension, but a linear fit yields a slightly larger overall relative error, by about 14.0 per cent. When the same analysis is repeated with the data set that has not been smoothed after reconstruction, however, a linear regression fit yields about 10.1 per cent less overall relative error, suggesting that the morphology is mostly linear. Since smoothing tends to erase filamentarity, this apparent discrepancy is actually expected. Performing a Pearson test between the mass and the longest dimension of the superclusters yields a Pearson correlation coefficient of 0.88 for the smoothed data set and 0.81 for the one with no smoothing. The fact that the correlation is more than 80 per cent in both cases supports the idea that these two quantities are mostly linearly correlated.

Naturally, the large smoothing kernel size decreases the patchiness of the field, making it more even. While smoothing definitely does not introduce new structures in a physical sense, it seems to erase filamentary features at small scales and it allows more supercluster groups to pass the test of connectedness for a minimum of 100 cells required for percolation. A more detailed filamentarity analysis can be done using shapefinders with Minkowski functionals, and it has been shown that most SDSS superclusters have filamentarities that are larger than their planarities in DR7 (Einasto et al., 2011).

5.3.4 Masses of Superclusters

Superclusters are the largest isolated overdense regions in the Universe, with characteristic dimensions from several h^{-1}Mpc up to ~ 100 h^{-1}Mpc, but in this study, they are typically hundreds of h^{-1}Mpc wide for the longest dimension (Figure 5.3, panels (c) and (d)). The reason for this relatively bigger scale is mainly due to the relatively large smoothing radius and cell spacing.

Having identified the superclusters in the DR12 IGM field, we compute their masses via



(a) Superclusters, smoothed

(b) Superclusters, no smoothing

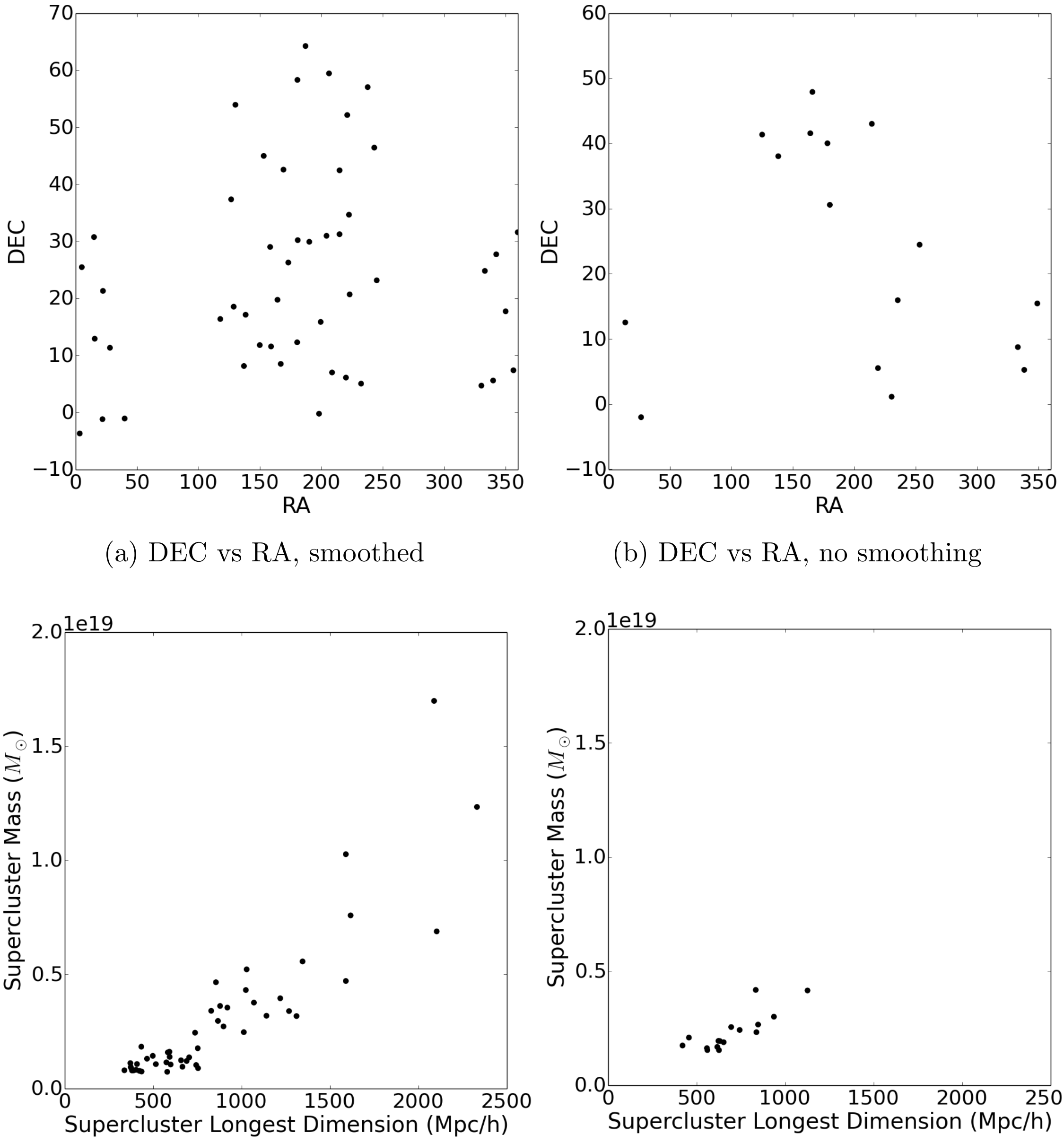
Figure 5.2: Panels (a) and (b), smoothed and with no smoothing, show locations of supercluster groups in the reconstruction cube with percolation analysis, for the smoothed case and no smoothing, respectively. Numbers show individual superclusters.

$$M_{SC} = M_C N_C \left(1 - \frac{\langle \delta_F \rangle}{b}\right) \tag{5.1}$$

$$M_C = m L_C^3 \tag{5.2}$$

where m is the solar mass per cubic h^{-1}Mpc ($= 7.6 \times 10^{10} M_\odot / 0.7$), L_C is the cell width, M_C is the mass per cell, N_C is the number of cells in a given supercluster, $\langle \delta_F \rangle$ is the mean delta flux in the supercluster and b is the linear bias of the forest ($= 0.2$, Slosar et al. (2011)).

In Figure 5.3, panels (c) through (f) summarize our findings. We had already discussed panels (c) and (d) in the morphology section. The relation between the mean delta flux in a supercluster and its mass is provided in panel (e). For most superclusters, the mean delta flux values are concentrated at $\delta_F \sim -0.1$ for the smoothed case, following a slightly negative correlation with supercluster mass. This is compatible with the analysis in similar studies. For example, (Mukae et al., 2016), Figure 2 shows a similar trend for the mean flux of a quasar sightline as a function of nearby galaxy overdensities.

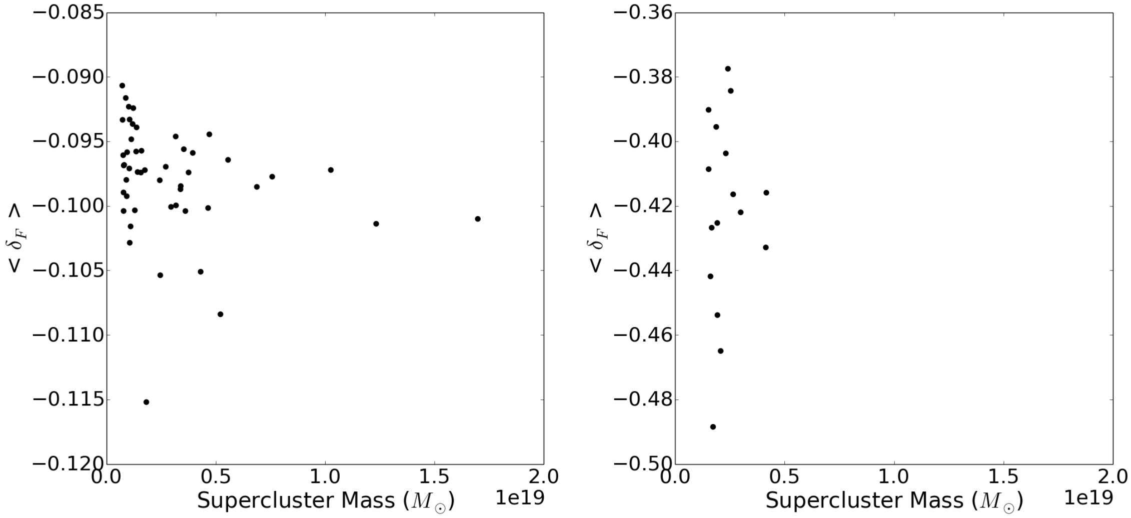


(a) DEC vs RA, smoothed

(b) DEC vs RA, no smoothing

(c) Mass vs longest dimension, smoothed

(d) Mass vs longest dimension, no smoothing



(e) Mean delta flux vs mass, smoothed

(f) Mean delta flux vs mass, no smoothing

Figure 5.3: Panels (a) and (b) show the right ascension and declination of the supercluster locations in the sky. Supercluster mass vs longest dimension is plotted in panels (c) and (d), smoothed and with no smoothing, respectively. The mean delta flux values tend to accumulate around $< \delta_F > \sim -0.10$ for the smoothed case (panel (e)).

5.4 Conclusions

Identifying structures in the IGM field can constrain physical parameters, as shown in previous literature. First we compare the local peak statistics of inferred DR12 IGM field with that of a Gaussian linear density field, and find that the local peak statistics agree at all smoothing levels. The best agreement is achieved at the fiducial smoothing of 60 h^{-1}Mpc.

We extend the local peak analysis using percolation techniques, and identify 49 supercluster candidates at $z \sim 2.0$. The large isotropic smoothing level tends to erase the filamentary properties of the topology of the overdense regions in the Universe. Also, the number of supercluster candidates is significantly reduced without smoothing. This suggests that the map resolution, and therefore the smoothing level, has a direct effect on the supercluster statistics.

The distribution of superclusters across the sky seem to be random, with no preference for any RA or DEC range. We find that typical supercluster masses are of the order of $10^{19}\,\mathrm{M}_\odot$, due to their large sizes (Figure 5.3, panel (e) shows sizes of up to three orders of magnitude greater than the typical protocluster length scale, e.g. see Mukae et al. 2016). We observe a slightly negative correlation between the mean delta flux level and supercluster mass.

A brief analysis of the morphology suggests that the observed structures have both filamentary and pancake–like properties, with a tendency towards a more filamentary topology, however, it is difficult to arrive at conclusive results with the current map resolution. When the quasar density increases in future observational surveys by a factor of ~ 2.5, this will allow the resolution to increase to levels finer than ~ 10 h^{-1}Mpc, enabling a more conclusive topological analysis of the intergalactic medium with this interpolation technique.

Chapter 6

Discussion and Conclusions

6.1 Summary

The Lyα forest has become the primary means of studying the large scale structure at high redshifts. We show with cosmological hydrodynamical simulations run with the P–GADGET code that for BOSS–like QSO densities, IGM maps can be made at redshifts greater than 2 by interpolating between the QSO skewers, using Wiener interpolation. The resulting maps show high fidelity, especially at regions with mild overdensities.

Local polynomial smoothing is an alternative to Wiener interpolation, however, we find that lower LOS densities (e.g. half of the LOS density for BOSS) are not acceptable for local smoothing, although this does not pose a problem for Wiener interpolation.

After evaluating the performance of simulations, we apply the same methods to the BOSS DR12 Lyα forest and create large maps of the IGM field. Since the IGM is in the mildly non–linear regime within the redshift range $2 < z < 4$, we test for non–Gaussianity in both simulations and observational maps via Kolmogorov–Smirnov tests with the cumulative distribution functions and local peak comparisons, finding consistent results. In the IGM maps, it is possible to identify structures using the percolation algorithm, which creates a friends of friends group catalog. We examine the distribution of the identified structures and study various statistics. For the morphology of the superclusters, we observe a mostly filamentary picture, and we also examine various properties like the longest distance, the mass and the mean flux of these individual structures.

6.2 Discussion

We have shown that Wiener interpolation is a feasible means of inferring the IGM field in the Lyα forest by studying various statistics of the reconstructed simulation

Dataset	**When**	**Area (deg^2)**	$\mathbf{N_{spectra}}$	**Mean Separation**
BOSS DR12	2016	10000	160,000	15 arcmin
eBOSS	2014–2018	7500	270,000	10 arcmin
CLAMATO	2014–2018	0.8	1000	1.7 arcmin
WEAVE	2018–2020	6000	400,000	7.5 arcmin
DESI	2018–2022	14000	840,000	7.5 arcmin
MSE	2025–	1000	1,000,000	1.9 arcmin

Table 6.1: Some relevant parameters for future Lyα forest observational datasets. Of these, BOSS (Dawson et al., 2013) has been completed, eBOSS (Dawson et al., 2016) and CLAMATO (Lee et al., 2014a) are ongoing, and WEAVE (Dalton et al., 2012) and DESI (Aghamousa et al., 2016) are about to start. The survey labelled MSE is a potential star forming galaxy with the proposed Mauna Kea Spectrosopic Explorer instrument[0].

field and applying our methods to the observational data set from BOSS DR12. In general, the areal tracer density requirements for making maps using the Lyα forest are easier to meet than the high volume requirements necessary for galaxy surveys.

There are some minor differences between the inference of the simulated field in §2 and the observations in §3. The former has all the Lyα skewers in parallel, which is still a good approximation for the observational case. However, the skewers do not span the entire reconstruction cube in the observational field, unlike the simulation, which may change the fidelity of the subsequent interpolation. We believe that this is a minor difference and our methods should still yield observational maps of the IGM with high fidelity, at least in regions where the LOS density is sufficiently high.

We note that both the matrix form of the auto–correlation of the sparse Lyα skewers ($\mathbf{C_{DD}}$ in Equation 2.2) and the smoothing algorithm leave room for improvement. Although we tested an alternative form of the auto–correlation using observed correlations of the field from Slosar et al. (2011), it did not improve our results. If the correlation of the Lyα skewers is known with better accuracy, the Wiener filter may be able to reduce the error in the resulting IGM fields, compared to our current choice of assuming a Gaussian auto–correlation. Also, since we do not necessarily expect the same statistics in the direction parallel and perpendicular to the LOSs, the isotropic nature of the Gaussian smoothing can be modified to better suit the needs of the IGM map. Furthermore, an adaptive smoothing length, depending on the local map resolution, should depict a more accurate picture for the reconstruction.

[0] http://mse.cfht.hawaii.edu/

6.3 Future Work

The findings of this thesis should be revisited, once higher LOS densities are available with surveys such as e–BOSS and MS–DESI. This will result in an increased map resolution and therefore an overall higher fidelity in all aspects. We are summarizing relevant features of some future surveys in Table 6.1.

One of the main goals of BOSS, eBOSS and also future Lyα forest surveys such as DESI is to observe the BAO signal. This has succeeded in BOSS (e.g., Busca et al. (2013); Bautista et al. (2017)), but there is much more information that can be extracted from this data. Map making can form part of this. For example, with higher sightline densities, we can make higher resolution maps, resolving not only superclusters with sizes of hundreds of Mpc as we have done with BOSS data, but protoclusters with size of order 10 Mpc or less (Stark et al., 2015). The topology of the intergalactic medium can be explored, for example looking at the genus (number of holes) per unit volume as a function of smoothing scale, which was proposed decades ago (Gott III et al., 1987) as a probe of primordial non–Gaussianity. So far, this has only been done with low redshift galaxy data (e.g., James et al. (2007)), which involves a lot of smoothing as the datasets are point-like, and also with the CMB (Gott et al., 2007). Topology studies with the IGM would mean looking at a new regime, with less gravitational non–linearity than with low redshift galaxy data, and with high LOS density the chance to look at smaller scales as less smoothing is necessary. It has even been suggested that the behavior of genus with smoothing scale could be used as a standard ruler to constrain dark energy (Zunckel et al., 2011).

Three dimensional maps have been very important in the history of cosmology. For example, slices plotted from the CfA redshift survey (De Lapparent et al., 1986), such as the "stickman", were for many people the first evidence of large–scale structure in the Universe and the first widely noted filamentary structures. With the 3D IGM maps, we can also display the morphology of structures at higher redshifts and check them against the expectations of theory. If surprises are out there, such as non–Gaussian voids or unexpected structures, the Lyα maps will be a good way to find them because the Universe has never been mapped at such high resolution at these redshifts.

If the small–scale structure of the IGM is resolved well, it will be possible to constrain the early dark matter power spectrum, since the baryon field is a tracer of the underlying dark matter field, for scales larger than the Jeans smoothing length. The space density of the local peaks in the observational IGM field can be used to constrain the linear matter power spectrum, regardless of the tracer (De and Croft, 2007), and the neutrino mass fraction can also be constrained (De and Croft, 2010). Furthermore, it is possible to cross–correlate the Lyα forest with other large scale structure tracers such as weak–lensing maps (Massey et al., 2007), the Lyman β forest (Iršič et al., 2013), quasars (Font-Ribera et al., 2014), and 21cm intensity mapping (Carucci et al., 2016).

Bibliography

Kevork N Abazajian, Jennifer K Adelman-McCarthy, Marcel A Agüeros, Sahar S Allam, Carlos Allende Prieto, Deokkeun An, Kurt SJ Anderson, Scott F Anderson, James Annis, Neta A Bahcall, et al. The seventh data release of the sloan digital sky survey. *The Astrophysical Journal Supplement Series*, 182(2):543, 2009.

R Adam, PAR Ade, N Aghanim, M Arnaud, M Ashdown, J Aumont, C Baccigalupi, AJ Banday, RB Barreiro, JG Bartlett, et al. Planck 2015 results-ix. diffuse component separation: Cmb maps. *Astronomy & Astrophysics*, 594:A9, 2016.

Joshua J Adams, Guillermo A Blanc, Gary J Hill, Karl Gebhardt, Niv Drory, Lei Hao, Ralf Bender, Joyce Byun, Robin Ciardullo, Mark E Cornell, et al. Hetdex pilot survey for emission-line galaxies-i. survey design, performance, and catalog. *arXiv preprint arXiv:1011.0426*, 2010.

PAR Ade, N Aghanim, M Arnaud, M Ashdown, J Aumont, C Baccigalupi, AJ Banday, RB Barreiro, N Bartolo, E Battaner, et al. Planck 2015 results. xiv. dark energy and modified gravity. *arXiv preprint arXiv:1502.01590*, 2015.

PAR Ade, N Aghanim, M Arnaud, M Ashdown, J Aumont, C Baccigalupi, AJ Banday, RB Barreiro, JG Bartlett, N Bartolo, et al. Planck 2015 results-xiii. cosmological parameters. *Astronomy & Astrophysics*, 594:A13, 2016.

Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, et al. The desi experiment part i: Science, targeting, and survey design. *arXiv preprint arXiv:1611.00036*, 2016.

Shadab Alam, Franco D Albareti, Carlos Allende Prieto, F Anders, Scott F Anderson, Brett H Andrews, Eric Armengaud, Éric Aubourg, Stephen Bailey, Julian E Bautista, et al. The eleventh and twelfth data releases of the sloan digital sky survey: Final data from sdss-iii. *arXiv preprint arXiv:1501.00963*, 2015.

Shadab Alam, Metin Ata, Stephen Bailey, Florian Beutler, Dmitry Bizyaev, Jonathan A Blazek, Adam S Bolton, Joel R Brownstein, Angela Burden, Chia-Hsun Chuang, et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: cosmological analysis of the dr12 galaxy sample. *arXiv preprint arXiv:1607.03155*, 2016.

Franco D Albareti, Carlos Allende Prieto, Andres Almeida, Friedrich Anders, Scott Anderson, Brett H Andrews, Alfonso Aragon-Salamanca, Maria Argudo-Fernandez, Eric Armengaud, Eric Aubourg, et al. The thirteenth data release of the sloan digital sky survey: First spectroscopic data from the sdss-iv survey mapping nearby galaxies at apache point observatory. *arXiv preprint arXiv:1608.02013*, 2016.

Andreas Albrecht and Paul J Steinhardt. Cosmology for grand unified theories with radiatively induced symmetry breaking. *Physical Review Letters*, 48(17):1220, 1982.

Firooz A Allahdadi, Theodore C Carney, Jim R Hipp, Larry D Libersky, and Albert G Petschek. High strain lagrangian hydrodynamics: a three dimensional sph code for dynamic material response. Technical report, DTIC Document, 1993.

F Anders, C Chiappini, BX Santiago, HJ Rocha-Pinto, L Girardi, LN da Costa, MAG Maia, M Steinmetz, I Minchev, M Schultheis, et al. Chemodynamics of the milky way-i. the first year of apogee data. *Astronomy & Astrophysics*, 564:A115, 2014.

Lauren Anderson, Éric Aubourg, Stephen Bailey, Florian Beutler, Vaishali Bhardwaj, Michael Blanton, Adam S Bolton, J Brinkmann, Joel R Brownstein, Angela Burden, et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: baryon acoustic oscillations in the data releases 10 and 11 galaxy samples. *Monthly Notices of the Royal Astronomical Society*, 441(1):24–62, 2014.

James M Bardeen, Paul J Steinhardt, and Michael S Turner. Spontaneous creation of almost scale-free density perturbations in an inflationary universe. *Physical Review D*, 28(4):679, 1983.

James M Bardeen, JR Bond, Nick Kaiser, and AS Szalay. The statistics of peaks of gaussian random fields. *The Astrophysical Journal*, 304:15–61, 1986.

CM Baugh, S Cole, CS Frenk, and Cedric G Lacey. The epoch of galaxy formation. *The Astrophysical Journal*, 498(2):504, 1998.

Julian E Bautista, Nicolás G Busca, Julien Guy, James Rich, Michael Blomqvist, Hélion du Mas des Bourboux, Matthew M Pieri, Andreu Font-Ribera, Stephen Bailey, Timothée Delubac, et al. Measurement of bao correlations at $z = 2.3$ with sdss dr12\lya-forests. *arXiv preprint arXiv:1702.00176*, 2017.

Robert H Becker, Xiaohui Fan, Richard L White, Michael A Strauss, Vijay K Narayanan, Robert H Lupton, James E Gunn, James Annis, Neta A Bahcall, J Brinkmann, et al. Evidence for reionization at z 6: Detection of a gunn-peterson trough in a z= 6.28 quasar. *The Astronomical Journal*, 122(6):2850, 2001.

CL Bennett, D Larson, JL Weiland, N Jarosik, G Hinshaw, N Odegard, KM Smith, RS Hill, B Gold, M Halpern, et al. Nine-year wilkinson microwave anisotropy probe (wmap) observations: final maps and results. *The Astrophysical Journal Supplement Series*, 208(2):20, 2013.

Arjun Berera, Marcelo Gleiser, and Rudnei O Ramos. A first principles warm inflation model that solves the cosmological horizon and flatness problems. *Physical Review Letters*, 83(2):264, 1999.

Philip Bett, Vincent Eke, Carlos S Frenk, Adrian Jenkins, John Helly, and Julio Navarro. The spin and shape of dark matter haloes in the millennium simulation of a λ cold dark matter universe. *Monthly Notices of the Royal Astronomical Society*, 376(1):215–232, 2007.

HongGuang Bi and Arthur F Davidsen. Evolution of structure in the intergalactic medium and the nature of the lyα forest. *The Astrophysical Journal*, 479(2):523, 1997.

James Binney and Scott Tremaine. Galactic dynamics. *Princeton Series in Astrophysics,(Princeton University Press, Princeton, NJ, 1987).[Google Books].(Cited on page 36.)*, 1998.

George R Blumenthal, SM Faber, Joel R Primack, and Martin J Rees. Formation of galaxies and large scale structure with cold dark matter. *Nature*, 1984.

JS Bolton, M Viel, T-S Kim, MG Haehnelt, and RF Carswell. Possible evidence for an inverted temperature–density relation in the intergalactic medium from the flux distribution of the lyα forest. *Monthly Notices of the Royal Astronomical Society*, 386(2):1131–1144, 2008.

J. R. Bond, J. Centrella, and J. R. Szalay, A. S.and Wilson. *Dark Matter Shocked Pancakes*, pages 87–99. Springer Netherlands, Dordrecht, 1984.

Nicholas A Bond, Željko Ivezić, Branimir Sesar, Mario Jurić, Jeffrey A Munn, Adam Kowalski, Sarah Loebman, Timothy C Beers, Julianne Dalcanton, Constance M Rockosi, et al. The milky way tomography with sdss. iii. stellar kinematics. *The Astrophysical Journal*, 716(1):1, 2010.

Michael Boylan-Kolchin, Volker Springel, Simon DM White, Adrian Jenkins, and Gerard Lemson. Resolving cosmic structure formation with the millennium-ii simulation. *Monthly Notices of the Royal Astronomical Society*, 398(3):1150–1164, 2009.

Joel N Bregman. The search for the missing baryons at low redshift. *arXiv preprint arXiv:0706.1787*, 2007.

Simon R Broadbent and John M Hammersley. Percolation processes. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 53, pages 629–641. Cambridge Univ Press, 1957.

Volker Bromm and Naoki Yoshida. The first galaxies. *arXiv preprint arXiv:1102.4638*, 2011.

Nicolás G Busca, Timothee Delubac, James Rich, Stephen Bailey, Andreu Font-Ribera, David Kirkby, J-M Le Goff, MM Pieri, Anze , É Aubourg, JE Bautista, et al. Baryon acoustic oscillations in the lyα forest of boss quasars. *Astronomy & Astrophysics*, 552:A96, 2013.

Isabella P Carucci, Francisco Villaescusa-Navarro, and Matteo Viel. The cross-correlation between 21cm intensity mapping maps and the lyman-alpha forest in the post-reionization era. *arXiv preprint arXiv:1611.07527*, 2016.

S Caucci, S Colombi, C Pichon, E Rollinde, P Petitjean, and T Sousbie. Recovering the topology of the igm at z˜ 2. *arXiv preprint arXiv:0801.4335*, 2008.

Jessi Cisewski, Rupert AC Croft, Peter E Freeman, Christopher R Genovese, Nishikanta Khandai, Melih Ozbek, and Larry Wasserman. Non-parametric 3d map of the intergalactic medium using the lyman-alpha forest. *Monthly Notices of the Royal Astronomical Society*, 440(3):2599–2609, 2014.

William S Cleveland, Eric Grosse, and William M Shyu. Local regression models. *Statistical models in S*, 2:309–376, 1992.

Jörg M Colberg, K Simon Krughoff, and Andrew J Connolly. Intercluster filaments in a λcdm universe. *Monthly Notices of the Royal Astronomical Society*, 359(1): 272–282, 2005.

Matthew Colless, Bruce A Peterson, Carole Jackson, John A Peacock, Shaun Cole, Peder Norberg, Ivan K Baldry, Carlton M Baugh, Joss Bland-Hawthorn, Terry Bridges, et al. The 2df galaxy redshift survey: final data release. *arXiv preprint astro-ph/0306581*, 2003.

Wesley N Colley. Two-dimensional topology of large-scale structure in the las campanas redshift survey. *The Astrophysical Journal*, 489(2):471, 1997.

Andrew J Connolly, Ryan Scranton, David Johnston, Scott Dodelson, Daniel J Eisenstein, Joshua A Frieman, James E Gunn, Lam Hui, Bhuvnesh Jain, Stephen Kent, et al. The angular correlation function of galaxies from early sloan digital sky survey data. *The Astrophysical Journal*, 579(1):42, 2002.

F Coppolani, P Petitjean, F Stoehr, E Rollinde, C Pichon, S Colombi, MG Haehnelt, B Carswell, and R Teyssier. Transverse and longitudinal correlation functions in the intergalactic medium from 32 close pairs of high-redshift quasars. *Monthly Notices of the Royal Astronomical Society*, 370(4):1804–1816, 2006.

Rupert AC Croft and Enrique Gaztañaga. The space density of galaxy peaks and the linear matter power spectrum. *The Astrophysical Journal*, 495(2):554, 1998.

Rupert AC Croft, David H Weinberg, Neal Katz, and Lars Hernquist. Intergalactic helium absorption in cold dark matter models. *The Astrophysical Journal*, 488(2): 532, 1997.

Rupert AC Croft, David H Weinberg, Neal Katz, and Lars Hernquist. Recovery of the power spectrum of mass fluctuations from observations of the lyα forest. *The Astrophysical Journal*, 495(1):44, 1998.

Rupert AC Croft, Wayne Hu, and Romeel Dave. Cosmological limits on the neutrino mass from the ly α forest. *Physical Review Letters*, 83(6):1092, 1999.

Rupert AC Croft, David H Weinberg, Mike Bolte, Scott Burles, Lars Hernquist, Neal Katz, David Kirkman, and David Tytler. Toward a precise measurement of matter clustering: Lyα forest data at redshifts 2-4. *The Astrophysical Journal*, 581(1):20, 2002.

Rupert AC Croft, Jordi Miralda-Escudé, Zheng Zheng, Adam Bolton, Kyle S Dawson, Jeffrey B Peterson, Donald G York, Daniel Eisenstein, Jon Brinkmann, Joel Brownstein, et al. Large-scale clustering of lyman α emission intensity from sdss/boss. *Monthly Notices of the Royal Astronomical Society*, 457(4):3541–3572, 2016.

Scott M Croom, RJ Smith, BJ Boyle, T Shanks, L Miller, PJ Outram, and NS Loaring. The 2df qso redshift survey–xii. the spectroscopic catalogue and luminosity function. *Monthly Notices of the Royal Astronomical Society*, 349(4):1397–1418, 2004.

SJ Curran and MT Whiting. Complete ionization of the neutral gas: Why there are so few detections of 21 cm hydrogen in high-redshift radio galaxies and quasars. *The Astrophysical Journal*, 759(2):117, 2012.

Gavin Dalton, Scott C Trager, Don Carlos Abrams, David Carter, Piercarlo Bonifacio, J Alfonso L Aguerri, Mike MacIntosh, Chris Evans, Ian Lewis, Ramon Navarro, et al. Weave: the next generation wide-field spectroscopy facility for the william herschel telescope. In *SPIE Astronomical Telescopes+ Instrumentation*, pages 84460P–84460P. International Society for Optics and Photonics, 2012.

Marc Davis, George Efstathiou, Carlos S Frenk, and Simon DM White. The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, 292:371–394, 1985.

Marc Davis, Sandra M Faber, Jeffrey Newman, Andrew C Phillips, Richard S Ellis, Charles C Steidel, C Conselice, Alison L Coil, DP Finkbeiner, David C Koo, et al. Science objectives and early results of the deep2 redshift survey. In *Astronomical Telescopes and Instrumentation*, pages 161–172. International Society for Optics and Photonics, 2003.

Kyle S Dawson, David J Schlegel, Christopher P Ahn, Scott F Anderson, Éric Aubourg, Stephen Bailey, Robert H Barkhouser, Julian E Bautista, Alessandra Beifiori, Andreas A Berlind, et al. The baryon oscillation spectroscopic survey of sdss-iii. *The Astronomical Journal*, 145(1):10, 2013.

Kyle S Dawson, Jean-Paul Kneib, Will J Percival, Shadab Alam, Franco D Albareti, Scott F Anderson, Eric Armengaud, Éric Aubourg, Stephen Bailey, Julian E Bautista, et al. The sdss-iv extended baryon oscillation spectroscopic survey: overview and early data. *The Astronomical Journal*, 151(2):44, 2016.

Soma De and Rupert AC Croft. Peaks in the cosmological density field: sensitivity to initial power spectrum, redshift distortions and galaxy halo occupation. *Monthly Notices of the Royal Astronomical Society*, 382(4):1591–1600, 2007.

Soma De and Rupert AC Croft. Peaks in the cosmological density field: parameter constraints from 2df galaxy redshift survey data. *Monthly Notices of the Royal Astronomical Society*, 401(3):1989–1998, 2010.

Valérie De Lapparent, Margaret J Geller, and John P Huchra. A slice of the universe. *The Astrophysical Journal*, 302:L1–L5, 1986.

AJ Deason, V Belokurov, and NW Evans. Rotation of halo populations in the milky way and m31. *Monthly Notices of the Royal Astronomical Society*, 411(3):1480–1494, 2011.

Tiziana Di Matteo, Nishikanta Khandai, Colin DeGraf, Yu Feng, RAC Croft, Julio Lopez, and Volker Springel. Cold flows and the first quasars. *The Astrophysical Journal Letters*, 745(2):L29, 2012.

Mark Dijkstra, Adam Lidz, and Lam Hui. Beyond lyman-α: Constraints and consistency tests from the lyman-beta forest. *arXiv preprint astro-ph/0305498*, 2003.

Scott Dodelson. *Modern cosmology*. Academic Press, San Diego, CA, 2003. URL `https://cds.cern.ch/record/1282338`.

Maret Einasto, LJ Liivamägi, E Tago, E Saar, E Tempel, J Einasto, VJ Martinez, and P Heinämäki. Sdss dr7 superclusters-morphology. *Astronomy & Astrophysics*, 532:A5, 2011.

Daniel J Eisenstein, Idit Zehavi, David W Hogg, Roman Scoccimarro, Michael R Blanton, Robert C Nichol, Ryan Scranton, Hee-Jong Seo, Max Tegmark, Zheng Zheng, et al. Detection of the baryon acoustic peak in the large-scale correlation function of sdss luminous red galaxies. *The Astrophysical Journal*, 633(2):560, 2005.

Daniel J Eisenstein, David H Weinberg, Eric Agol, Hiroaki Aihara, Carlos Allende Prieto, Scott F Anderson, James A Arns, Éric Aubourg, Stephen Bailey, Eduardo Balbinot, et al. Sdss-iii: Massive spectroscopic surveys of the distant universe, the milky way, and extra-solar planetary systems. *The Astronomical Journal*, 142(3): 72, 2011.

Harold I Ewen and Edward M Purcell. Observation of a line in the galactic radio spectrum. In *Classics in Radio Astronomy*, pages 328–330. Springer, 1951.

Emilio E Falco, Michael J Kurtz, Margaret J Geller, John P Huchra, James Peters, Perry Berlind, Douglas J Mink, Susan P Tokarz, and Barbara Elwell. The updated zwicky catalog (uzc) 1, 2, 3. *Publications of the Astronomical Society of the Pacific*, 111(758):438–452, 1999.

Jianqing Fan and Irene Gijbels. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press, 1996.

Jianqing Fan, Theo Gasser, Irène Gijbels, Michael Brockmann, and Joachim Engel. Local polynomial regression: optimal kernels and asymptotic minimax efficiency. *Annals of the Institute of Statistical Mathematics*, 49(1):79–99, 1997.

Li-Zhi Fang, Hongguang Bi, Shouping Xiang, and Gerhard Boerner. Linear evolution of cosmic baryonic medium on large scales. *The Astrophysical Journal*, 413:477–485, 1993.

Taotao Fang and Martin White. Probing the statistics of the temperature-density relation of the intergalactic medium. *The Astrophysical Journal Letters*, 606(1):L9, 2004.

KB Fisher, Ofer Lahav, Yehuda Hoffman, Donald Lynden-Bell, and Saleem Zaroubi. Wiener reconstruction of density, velocity and potential fields from all-sky galaxy redshift surveys. *Monthly Notices of the Royal Astronomical Society*, 272(4):885–908, 1995.

Andreu Font-Ribera, David Kirkby, Jordi Miralda-Escudé, Nicholas P Ross, Anže Slosar, James Rich, Éric Aubourg, Stephen Bailey, Vaishali Bhardwaj, Julian

Bautista, et al. Quasar-lyman α forest cross-correlation from boss dr11: Baryon acoustic oscillations. *Journal of Cosmology and Astroparticle Physics*, 2014(05): 027, 2014.

Joshua Frieman, Michael Turner, and Dragan Huterer. Dark energy and the accelerating universe. *arXiv preprint arXiv:0803.0982*, 2008.

Joshua A Frieman, Bruce Bassett, Andrew Becker, Changsu Choi, David Cinabro, Fritz DeJongh, Darren L Depoy, Ben Dilday, Mamoru Doi, Peter M Garnavich, et al. The sloan digital sky survey-ii supernova survey: Technical summary. *The Astronomical Journal*, 135(1):338, 2007.

Peter M Frinchaboy, Benjamin Thompson, Kelly M Jackson, Julia O'Connell, Brianne Meyer, Gail Zasowski, Steven R Majewski, S Drew Chojnowksi, Jennifer A Johnson, Carlos Allende Prieto, et al. The open cluster chemical analysis and mapping survey: local galactic metallicity gradient with apogee using sdss dr10. *The Astrophysical Journal Letters*, 777(1):L1, 2013.

Antonella Garzilli, JS Bolton, T-S Kim, S Leach, and Matteo Viel. The intergalactic medium thermal history at redshift z= 1.7–3.2 from the lyα forest: a comparison of measurements using wavelets and the flux distribution. *Monthly Notices of the Royal Astronomical Society*, 424(3):1723–1736, 2012.

E Gaztañaga, P Norberg, CM Baugh, and DJ Croton. Statistical analysis of galaxy surveysii. the three-point galaxy correlation function measured from the 2dfgrs. *Monthly Notices of the Royal Astronomical Society*, 364(2):620–634, 2005.

Robert A Gingold and Joseph J Monaghan. Smoothed particle hydrodynamics: theory and application to non-spherical stars. *Monthly notices of the royal astronomical society*, 181(3):375–389, 1977.

Facundo A Gómez, Ivan Minchev, Brian W OShea, Young Sun Lee, Timothy C Beers, Deokkeun An, James S Bullock, Chris W Purcell, and Álvaro Villalobos. Signatures of minor mergers in the milky way disc–i. the segue stellar sample. *Monthly Notices of the Royal Astronomical Society*, 423(4):3727–3739, 2012.

J Richard Gott, Wesley N Colley, Chan-Gyung Park, Changbom Park, and Charles Mugnolo. Genus topology of the cosmic microwave background from the wmap 3-year data. *Monthly Notices of the Royal Astronomical Society*, 377(4):1668–1678, 2007.

J Richard Gott III, David H Weinberg, and Adrian L Melott. A quantitative approach to the topology of large-scale structure. *The Astrophysical Journal*, 319:1–8, 1987.

J Green, P Schechter, C Baltay, R Bean, D Bennett, R Brown, C Conselice, M Donahue, X Fan, BS Gaudi, et al. Wide-field infrared survey telescope (wfirst) final report. *arXiv preprint arXiv:1208.4012*, 2012.

James E Gunn and Bruce A Peterson. On the density of neutral hydrogen in intergalactic space. *The Astrophysical Journal*, 142:1633–1641, 1965.

N Gupta, R Srianand, Patrick Petitjean, P Noterdaeme, and DJ Saikia. 21-cm absorbers at intermediate redshifts. *arXiv preprint arXiv:0902.3016*, 2009a.

N Gupta, R Srianand, Patrick Petitjean, P Noterdaeme, and DJ Saikia. A complete sample of 21-cm absorbers at z 1.3: Giant metrewave radio telescope survey using mg ii systems. *Monthly Notices of the Royal Astronomical Society*, 398(1):201–220, 2009b.

Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, Jan 1981. doi: 10.1103/PhysRevD.23.347. URL `http://link.aps.org/doi/10.1103/PhysRevD.23.347`.

Alan H Guth and So-Young Pi. Fluctuations in the new inflationary universe. *Physical Review Letters*, 49(15):1110, 1982.

Francesco Haardt and Piero Madau. Radiative transfer in a clumpy universe: Ii. the utraviolet extragalactic background. *arXiv preprint astro-ph/9509093*, 1995.

Stephen W Hawking. The development of irregularities in a single bubble inflationary universe. *Physics Letters B*, 115(4):295–297, 1982.

Lars Hernquist, Neal Katz, David H Weinberg, and Jordi Miralda-Escude. The lyman-alpha forest in the cold dark matter model. *The Astrophysical Journal Letters*, 457 (2):L51, 1996.

G Hinshaw, D Larson, E Komatsu, DN Spergel, CL Bennett, J Dunkley, MR Nolta, M Halpern, RS Hill, N Odegard, et al. Nine-year wilkinson microwave anisotropy probe (wmap) observations: cosmological parameter results. *The Astrophysical Journal Supplement Series*, 208(2):19, 2013.

Edwin Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929.

Lam Hui and Nickolay Y Gnedin. Equation of state of the photoionized intergalactic medium. *Monthly Notices of the Royal Astronomical Society*, 292(1):27–42, 1997.

Lam Hui, Scott Burles, Uroš Seljak, Robert E Rutledge, Eugene Magnier, and David Tytler. On estimating the qso transmission power spectrum. *The Astrophysical Journal*, 552(1):15, 2001.

Vid Iršič, Anže Slosar, Stephen Bailey, Daniel J Eisenstein, Andreu Font-Ribera, Jean-Marc Le Goff, Britt Lundgren, Patrick McDonald, Ross O'Connell, Nathalie Palanque-Delabrouille, et al. Detection of lyβ auto-correlations and lyα-lyβ cross-correlations in boss data release 9. *Journal of Cosmology and Astroparticle Physics*, 2013(09):016, 2013.

J Berian James, Geraint F Lewis, and Matthew Colless. Topology of large-scale structure in the 2df galaxy redshift survey. *Monthly Notices of the Royal Astronomical Society*, 375(1):128–136, 2007.

Linhua Jiang, Xiaohui Fan, Fuyan Bian, Ian D McGreer, Michael A Strauss, James Annis, Zoë Buck, Richard Green, Jacqueline A Hodge, Adam D Myers, et al. The sloan digital sky survey stripe 82 imaging data: Depth-optimized co-adds over 300 deg2 in five filters. *The Astrophysical Journal Supplement Series*, 213(1):12, 2014.

Nick Kaiser. On the spatial correlations of abell clusters. *The Astrophysical Journal*, 284:L9–L12, 1984.

Nick Kaiser and JA Peacock. Power-spectrum analysis of one-dimensional redshift surveys. *The Astrophysical Journal*, 379:482–506, 1991.

Jochen Sebastian Klar and JP Mücket. A detailed view of filaments and sheets in the warm-hot intergalactic medium-i. pancake formation. *Astronomy & Astrophysics*, 522:A114, 2010.

CS Kochanek, DJ Eisenstein, RJ Cool, Nelson Caldwell, RJ Assef, BT Jannuzi, C Jones, SS Murray, WR Forman, Arjun Dey, et al. Ages: the agn and galaxy evolution survey. *The Astrophysical Journal Supplement Series*, 200(1):8, 2012.

Girish Kulkarni, Joseph F Hennawi, Jose Oñorbe, Alberto Rorai, and Volker Springel. Characterizing the pressure smoothing scale of the intergalactic medium. *The Astrophysical Journal*, 812(1):30, 2015.

Cedric Lacey and Shanu Cole. Merger rates in hierarchical models of galaxy formation–ii. comparison with n-body simulations. *Monthly Notices of the Royal Astronomical Society*, 271(3):676–692, 1994.

Cedric Lacey and Shaun Cole. Merger rates in hierarchical models of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 262(3):627–649, 1993.

BJ Last and DJ Thouless. Percolation theory and electrical conductivity. *Physical Review Letters*, 27(25):1719, 1971.

O Le Fèvre, P Cassata, O Cucciati, B Garilli, O Ilbert, V Le Brun, D Maccagni, C Moreau, M Scodeggio, L Tresse, et al. The vimos vlt deep survey final data release: a spectroscopic sample of 35 016 galaxies and agn out to z˜ 6.7 selected with 17.5 iab 24.75. *Astronomy & Astrophysics*, 559:A14, 2013.

Khee-Gan Lee, Stephen Bailey, Leslie E Bartsch, William Carithers, Kyle S Dawson, David Kirkby, Britt Lundgren, Daniel Margala, Nathalie Palanque-Delabrouille, Matthew M Pieri, et al. The boss lyα forest sample from sdss data release 9. *The Astronomical Journal*, 145(3):69, 2013.

Khee-Gan Lee, Joseph F Hennawi, Casey Stark, J. Xavier Prochaska, and Martin White. Lya forest tomography from background galaxies: The first megaparsec-resolution large-scale structure map at z¿2. *The Astrophysical Journal*, 788(1):8, September 2014a.

Khee-Gan Lee, Joseph F Hennawi, Martin White, Rupert AC Croft, and Melih Ozbek. Observational requirements for lyα forest tomographic mapping of large-scale structure at z˜ 2. *The Astrophysical Journal*, 788(1):49, May 2014b.

Khee-Gan Lee, Joseph F Hennawi, David N Spergel, David H Weinberg, David W Hogg, Matteo Viel, James S Bolton, Stephen Bailey, Matthew M Pieri, William Carithers, et al. Igm constraints from the sdss-iii/boss dr9 lyα forest transmission probability distribution function. *The Astrophysical Journal*, 799(2):196, 2015.

M. Levi, C. Bebek, T. Beers, R. Blum, R. Cahn, D. Eisenstein, B. Flaugher, K. Honscheid, R. Kron, O. Lahav, P. McDonald, N. Roe, D. Schlegel, and representing the DESI collaboration. The DESI Experiment, a whitepaper for Snowmass 2013. *ArXiv e-prints*, August 2013.

Michael Levi, Chris Bebek, Timothy Beers, Robert Blum, Robert Cahn, Daniel Eisenstein, Brenna Flaugher, Klaus Honscheid, Richard Kron, Ofer Lahav, et al. The desi experiment, a whitepaper for snowmass 2013. *arXiv preprint arXiv:1308.0847*, 2013.

Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of cosmic microwave background anisotropies in closed friedmann-robertson-walker models. *The Astrophysical Journal*, 538(2):473, 2000.

A. D. Linde. A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, 108:389–393, February 1982. doi: 10.1016/0370-2693(82)91219-9.

Roger Lynds. The absorption-line spectrum of 4c 05.34. *The Astrophysical Journal*, 164:L73, 1971.

Ariyeh H Maller, Jason X Prochaska, Rachel S Somerville, and Joel R Primack. Damped lyman alpha systems and galaxy formation models-i. the radial distribution of cold gas at high z. *Monthly Notices of the Royal Astronomical Society*, 326 (4):1475–1488, 2001.

Ariyeh H Maller, Jason X Prochaska, Rachel S Somerville, and Joel R Primack. Damped lyman alpha systems and galaxy formation models–ii. high ions and lyman-limit systems. *Monthly Notices of the Royal Astronomical Society*, 343(1):268–278, 2003.

Richard Massey, Jason Rhodes, Richard Ellis, Nick Scoville, Alexie Leauthaud, Alexis Finoguenov, Peter Capak, David Bacon, Hervé Aussel, Jean-Paul Kneib, et al. Dark matter maps reveal cosmic scaffolding. *Nature*, 445(7125):286–290, 2007.

John C Mather, ES Cheng, RE Eplee Jr, RB Isaacman, SS Meyer, RA Shafer, R Weiss, EL Wright, CL Bennett, NW Boggess, et al. A preliminary measurement of the cosmic microwave background spectrum by the cosmic background explorer (cobe) satellite. *The Astrophysical Journal*, 354:L37–L40, 1990.

Matthew McQuinn. The evolution of the intergalactic medium. *arXiv preprint arXiv:1512.00086*, 2015.

Matthew McQuinn, Adam Lidz, Matias Zaldarriaga, Lars Hernquist, Philip F Hopkins, Suvendra Dutta, and Claude-André Faucher-Giguère. He ii reionization and its effect on the intergalactic medium. *The Astrophysical Journal*, 694(2):842, 2009.

Avery Meiksin. Colour corrections for high-redshift objects due to intergalactic attenuation. *Monthly Notices of the Royal Astronomical Society*, 365(3):807–812, 2006.

Andrei Mesinger and Mark Dijkstra. Ultraviolet radiative feedback during the advanced stages of reionization. *Monthly Notices of the Royal Astronomical Society*, 390(3):1071–1080, 2008.

Lance Miller, AM Lopes, RJ Smith, SM Croom, BJ Boyle, T Shanks, and P Outram. Possible arcminute-separation gravitational lensed qsos in the 2df qso survey. *arXiv preprint astro-ph/0210644*, 2002.

Lance Miller, SM Croom, BJ Boyle, NS Loaring, RJ Smith, T Shanks, and P Outram. 200-mpc-sized structure in the 2df qso redshift survey. *Monthly Notices of the Royal Astronomical Society*, 355(2):385–394, 2004.

Jordi Miralda-Escude, Renyue Cen, Jeremiah P Ostriker, and Michael Rauch. The lyα forest from gravitational collapse in the cold dark matter+ λ model. *The Astrophysical Journal*, 471(2):582, 1996.

HJ Mo and Simon DM White. An analytic model for the spatial clustering of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 282(2):347–361, 1996.

Joe J Monaghan. Smoothed particle hydrodynamics. *Annual review of astronomy and astrophysics*, 30(1):543–574, 1992.

Jan P Mücket, Patrick Petitjean, Ronald E Kates, and Rüdiger Riediger. Evolution of the lya forest: a consistent picture. *arXiv preprint astro-ph/9508129*, 1995.

Shiro Mukae, Masami Ouchi, Koki Kakiichi, Nao Suzuki, Yoshiaki Ono, Zheng Cai, Akio K Inoue, Yi-Kuan Chiang, Takatoshi Shibuya, and Yuichi Matsuda. Cosmic galaxy-igm hi relation at $z \sim 2-3$ probed in the cosmos/ultravista 1.6 deg^2 field. *arXiv preprint arXiv:1605.00379*, 2016.

Adam D Myers, Robert J Brunner, Gordon T Richards, Robert C Nichol, Donald P Schneider, Daniel E Vanden Berk, Ryan Scranton, Alexander G Gray, and Jon Brinkmann. First measurement of the clustering evolution of photometrically classified quasars. *The Astrophysical Journal*, 638(2):622, 2006.

Smadar Naoz, Shay Noter, and Rennan Barkana. The first stars in the universe. *Monthly Notices of the Royal Astronomical Society: Letters*, 373(1):L98–L102, 2006.

Jeffrey A Newman, Michael C Cooper, Marc Davis, SM Faber, Alison L Coil, Puragra Guhathakurta, David C Koo, Andrew C Phillips, Charlie Conroy, Aaron A Dutton, et al. The deep2 galaxy redshift survey: Design, observations, data reduction, and redshifts. *The Astrophysical Journal Supplement Series*, 208(1):5, 2013.

Taira Oogi, Motohiro Enoki, Tomoaki Ishiyama, Masakazu AR Kobayashi, Ryu Makiya, and Masahiro Nagashima. Quasar clustering in a galaxy and quasar formation model based on ultra high-resolution n-body simulations. *Monthly Notices of the Royal Astronomical Society: Letters*, 456(1):L30–L34, 2016.

JP Ostriker and S Ikeuchi. Physical properties of the intergalactic medium and the lyman-alpha absorbing clouds. *The Astrophysical Journal*, 268:L63–L68, 1983.

JP Ostriker and El T Vishniac. Generation of microwave background fluctuations from nonlinear perturbations at the era of galaxy formation. *The Astrophysical Journal*, 306:L51–L54, 1986.

PJ Outram, Fiona Hoyle, T Shanks, SM Croom, BJ Boyle, L Miller, RJ Smith, and AD Myers. The 2df qso redshift surveyxi. the qso power spectrum. *Monthly Notices of the Royal Astronomical Society*, 342(2):483–495, 2003.

R Overzier, G Lemson, RE Angulo, E Bertin, J Blaizot, BMB Henriques, G-D Marleau, and SDM White. The millennium run observatory: first light. *Monthly Notices of the Royal Astronomical Society*, 2012.

Melih Ozbek, Rupert AC Croft, and Nishikanta Khandai. Large-scale 3d mapping of the intergalactic medium using the lyman alpha forest. *Monthly Notices of the Royal Astronomical Society*, 456:3610–3623, 2016.

I. Pâris, P. Petitjean, N. P. Ross, A. D. Myers, É. Aubourg, A. Streblyanska, S. Bailey, É. Armengaud, N. Palanque-Delabrouille, C. Yèche, F. Hamann, M. A. Strauss, F. D. Albareti, J. Bovy, D. Bizyaev, W. N. Brandt, M. Brusa, J. Buchner, J. Comparat, R. A. C. Croft, T. Dwelly, X. Fan, A. Font-Ribera, J. Ge, A. Georgakakis, P. B. Hall, L. Jian, K. Kinemuchi, E. Malanushenko, V. Malanushenko, R. G. McMahon, M.-L. Menzel, A. Merloni, K. Nandra, P. Noterdaeme, D. Oravetz, K. Pan, M. M. Pieri, F. Prada, M. Salvato, D. J. Schlegel, D. P. Schneider, A. Simmons, M. Viel, D. H. Weinberg, and L. Zhu. The Sloan Digital Sky Survey Quasar Catalog: twelfth data release. *ArXiv e-prints*, August 2016.

Isabelle Pâris, Patrick Petitjean, É Aubourg, Stephen Bailey, Nicholas P Ross, Adam D Myers, Michael A Strauss, Scott F Anderson, Eduard Arnau, Julian Bautista, et al. The sloan digital sky survey quasar catalog: ninth data release. *Astronomy & Astrophysics*, 548:A66, 2012.

Phillip James Edwin Peebles. *The large-scale structure of the universe*. Princeton university press, 1980.

Will J Percival, Carlton M Baugh, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, Shaun Cole, Matthew Colless, Chris Collins, Warrick Couch, Gavin Dalton, et al. The 2df galaxy redshift survey: the power spectrum and the matter content of the universe. *Monthly Notices of the Royal Astronomical Society*, 327(4): 1297–1306, 2001.

Saul Perlmutter, G Aldering, G Goldhaber, RA Knop, P Nugent, PG Castro, S Deustua, S Fabbro, A Goobar, DE Groom, et al. Measurements of ω and λ from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2):565, 1999.

Serena Perrotta, Valentina D'Odorico, J Xavier Prochaska, Stefano Cristiani, Guido Cupani, Sara Ellison, Sebastian Lòpez, George D Becker, Trystyn AM Berg, Lise Christensen, et al. Nature and statistical properties of quasar associated absorption systems in the xq-100 legacy survey. *arXiv preprint arXiv:1605.04607*, 2016.

JB Peterson and Enrique Suarez. Intensity mapping with the 21-cm and lyman alpha lines. *arXiv preprint arXiv:1206.0143*, 2012.

Patrick Petitjean, Emmanuel Rollinde, Bastien Aracil, Christophe Pichon, and Stéphane Colombi. 3d spatial distribution of the intergalactic medium. *arXiv preprint astro-ph/0109082*, 2001.

C Pichon, JL Vergely, E Rollinde, S Colombi, and P Petitjean. Inversion of the lyman-α forest: 3d investigation of the intergalactic medium. *Arxiv preprint astro-ph/0105196*, 2001.

Marc Postman and Tod R Lauer. Brightest cluster galaxies as standard candles. *The Astrophysical Journal*, 440:28–47, 1995.

William H Press and Paul Schechter. Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *The Astrophysical Journal*, 187:425–438, 1974.

Jonathan R Pritchard and Abraham Loeb. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, 2012.

Jonathan R Pritchard, Abraham Loeb, and J Stuart B Wyithe. Constraining reionization using 21-cm observations in combination with cmb and lyα forest data. *Monthly Notices of the Royal Astronomical Society*, 408(1):57–70, 2010.

A Raichoor, J Comparat, T Delubac, J-P Kneib, C Yèche, H Zou, FB Abdalla, K Dawson, X Fan, Z Fan, et al. The sdss-iv extended baryonic oscillation spectroscopic survey: selecting emission line galaxies using the fisher discriminant. *arXiv preprint arXiv:1505.01797*, 2015.

Michael Rauch. The lyman alpha forest in the spectra of qsos. *arXiv preprint astro-ph/9806286*, 1998.

Andreas Reisenegger and Jordi Miralda-Escude. The gunn-peterson effect from underdense regions in a photoionized intergalactic medium. *arXiv preprint astro-ph/9502063*, 1995.

Gordon T Richards, Michael A Strauss, Xiaohui Fan, Patrick B Hall, Sebastian Jester, Donald P Schneider, Daniel E Vanden Berk, Chris Stoughton, Scott F Anderson, Robert J Brunner, et al. The sloan digital sky survey quasar survey: Quasar luminosity function from data release 3. *The Astronomical Journal*, 131(6):2766, 2006.

Adam G Riess, Alexei V Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M Garnavich, Ron L Gilliland, Craig J Hogan, Saurabh Jha, Robert P Kirshner, et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009, 1998.

E Rollinde, P Petitjean, C Pichon, S Colombi, B Aracil, V D'Odorico, and MG Haehnelt. The correlation of the lyman α forest in close pairs and groups of high-redshift quasars: clustering of matter on scales of 1–5 mpc. *Monthly Notices of the Royal Astronomical Society*, 341(4):1279–1289, 2003.

Emmanuel Rollinde, Patrick Petitjean, and Christophe Pichon. Physical properties and small-scale structure of the lyman-α forest: Inversion of the he 1122-1628 uves spectrum. *Astronomy & Astrophysics*, 376(1):28–42, 2001.

Ariel G Sánchez, Jan Niklas Grieb, Salvador Salazar-Albornoz, Shadab Alam, Florian Beutler, Ashley J Ross, Joel R Brownstein, Chia-Hsun Chuang, Antonio J Cuesta, Daniel J Eisenstein, et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: combining correlated gaussian posterior distributions. *Monthly Notices of the Royal Astronomical Society*, 464(2):1493–1501, 2017.

Mário G Santos, Alexandre Amblard, Jonathan Pritchard, Hy Trac, Renyue Cen, and Asantha Cooray. Cosmic reionization and the 21 cm signal: comparison between an analytical model and a simulation. *The Astrophysical Journal*, 689(1):1, 2008.

PAG Scheuer. A sensitive test for the presence of atomic hydrogen in intergalactic space. *Nature*, 207:963, 1965.

Donald P Schneider, Gordon T Richards, Xiaohui Fan, Patrick B Hall, Michael A Strauss, Daniel E Vanden Berk, James E Gunn, Heidi Jo Newberg, Timothy A Reichard, C Stoughton, et al. The sloan digital sky survey quasar catalog. i. early data release. *The Astronomical Journal*, 123(2):567, 2002.

Donald P Schneider, Xiaohui Fan, Patrick B Hall, Sebastian Jester, Gordon T Richards, Chris Stoughton, Michael A Strauss, Mark SubbaRao, Daniel E Vanden Berk, Scott F Anderson, et al. The sloan digital sky survey quasar catalog. ii. first data release. *The Astronomical Journal*, 126(6):2579, 2003.

Donald P Schneider, Patrick B Hall, Gordon T Richards, Daniel E Vanden Berk, Scott F Anderson, Xiaohui Fan, Sebastian Jester, Chris Stoughton, Michael A Strauss, Mark SubbaRao, et al. The sloan digital sky survey quasar catalog. iii. third data release. *The Astronomical Journal*, 130(2):367, 2005.

Marco Scodeggio and Giuseppe Gavazzi. 21 centimeter study of spiral galaxies in clusters. iii-neutral gas content, star formation, and radio continuum properties. *The Astrophysical Journal*, 409:110–125, 1993.

Uroš Seljak, Alexey Makarov, Patrick McDonald, Scott F Anderson, Neta A Bahcall, J Brinkmann, Scott Burles, Renyue Cen, Mamoru Doi, James E Gunn, et al. Cosmological parameter analysis including sdss ly α forest and galaxy bias: constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy. *Physical Review D*, 71(10):103515, 2005.

Sergei F Shandarin, Jatush V Sheth, and Varun Sahni. Morphology of the supercluster–void network in λcdm cosmology. *Monthly Notices of the Royal Astronomical Society*, 353(1):162–178, 2004.

Stephen A Shectman, Stephen D Landy, Augustus Oemler, Douglas L Tucker, Huan Lin, Robert P Kirshner, and Paul L Schechter. The las campanas redshift survey. *arXiv preprint astro-ph/9604167*, 1996.

IS Shklovskii. Physical conditions in the gaseous envelope of 3c-273. *Soviet Astronomy*, 8:638, 1965.

J Michael Shull, Mark L Giroux, Steven V Penton, Jason Tumlinson, John T Stocke, Edward B Jenkins, H Warren Moos, William R Oegerle, Blair D Savage, Kenneth R Sembach, et al. Fuse observations of the low-redshift lyman-beta forest. *arXiv preprint astro-ph/0005011*, 2000.

Anže Slosar, Andreu Font-Ribera, Matthew M Pieri, James Rich, Jean-Marc Le Goff, Éric Aubourg, Jon Brinkmann, Bill Carithers, Romain Charlassier, Marina Cortês, et al. The lyman-α forest in three dimensions: measurements of large scale flux correlations from boss 1st-year data. *Journal of Cosmology and Astroparticle Physics*, 2011(09):001, 2011.

Anže Slosar, Vid Iršič, David Kirkby, Stephen Bailey, Timothée Delubac, James Rich, Éric Aubourg, Julian E Bautista, Vaishali Bhardwaj, Michael Blomqvist, et al. Measurement of baryon acoustic oscillations in the lyman-α forest fluctuations in boss data release 9. *Journal of Cosmology and Astroparticle Physics*, 2013(04):026, 2013.

T Sousbie, C Pichon, S Colombi, D Novikov, and D Pogosyan. The 3d skeleton: tracing the filamentary structure of the universe. *Monthly Notices of the Royal Astronomical Society*, 383(4):1655–1670, 2008.

Volker Springel. The cosmological simulation code gadget-2. *Monthly Notices of the Royal Astronomical Society*, 364(4):1105–1134, 2005.

Volker Springel and Lars Hernquist. Cosmological sph simulations: A hybrid multiphase model for star formation. *arXiv preprint astro-ph/0206393*, 2002.

Volker Springel, Naoki Yoshida, and Simon DM White. Gadget: a code for collisionless and gasdynamical cosmological simulations. *New Astronomy*, 6(2):79–117, 2001.

Volker Springel, Simon DM White, Adrian Jenkins, Carlos S Frenk, Naoki Yoshida, Liang Gao, Julio Navarro, Robert Thacker, Darren Croton, John Helly, et al. Simulations of the formation, evolution and clustering of galaxies and quasars. *nature*, 435(7042):629–636, 2005.

Volker Springel, Carlos S Frenk, and Simon DM White. The large-scale structure of the universe. *Nature*, 440(7088):1137–1144, 2006.

Casey W Stark, Martin White, Khee-Gan Lee, and Joseph F Hennawi. Protocluster discovery in tomographic ly α forest flux maps. *Monthly Notices of the Royal Astronomical Society*, 453(1):311–327, 2015.

Max Tegmark, Daniel J Eisenstein, Michael A Strauss, David H Weinberg, Michael R Blanton, Joshua A Frieman, Masataka Fukugita, James E Gunn, Andrew JS Hamilton, Gillian R Knapp, et al. Cosmological constraints from the sdss luminous red galaxies. *Physical Review D*, 74(12):123507, 2006.

Saeed V Vaseghi. Wiener filters. *Advanced Digital Signal Processing and Noise Reduction, Saeed V. Vaseghi Copyright© 2000 John Wiley & Sons Ltd ISBNs: 0-471-62692-9 (Hardback): 0-470-84162-1 (Electronic)*, page 178, 2000.

Licia Verde, Alan F Heavens, Will J Percival, Sabino Matarrese, Carlton M Baugh, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, Shaun Cole, Matthew Colless, et al. The 2df galaxy redshift survey: the bias of galaxies and the density of the universe. *Monthly Notices of the Royal Astronomical Society*, 335(2):432–440, 2002.

Matteo Viel, Julien Lesgourgues, Martin G Haehnelt, Sabino Matarrese, and Antonio Riotto. Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with wmap and the lyman-α forest. *Physical Review D*, 71(6): 063534, 2005.

Matteo Viel, George D Becker, James S Bolton, and Martin G Haehnelt. Warm dark matter as a solution to the small scale crisis: New constraints from high redshift lyman-α forest data. *Physical Review D*, 88(4):043502, 2013.

Steven S Vogt, Steven L Allen, Bruce C Bigelow, L Bresee, William E Brown, T Cantrall, Albert Conrad, M Couture, C Delaney, Harland W Epps, et al. Hires: the high-resolution echelle spectrometer on the keck 10-m telescope. In *1994 Symposium on Astronomical Telescopes & Instrumentation for the 21st Century*, pages 362–375. International Society for Optics and Photonics, 1994.

L Wasserman. All of nonparametric statistics, ser, 2006.

David H Weinberg, Romeel Dav'e, Neal Katz, and Juna A Kollmeier. The lyman-alpha forest as a cosmological tool. *arXiv preprint astro-ph/0301186*, 2003.

DH Weinberg, L Hernsquit, N Katz, R Croft, and J Miralda-Escudé. Structure and evolution of the intergalactic medium from qso absorption line system. *Paris: France Publisher*, page 133, 1997.

Martin White. The zel'dovich approximation. *Monthly Notices of the Royal Astronomical Society*, page stu209, 2014.

Martin White, Douglas Scott, and Joseph Silk. Anisotropies in the cosmic microwave background. *Annual Review of Astronomy and Astrophysics*, 32:319–370, 1994.

Simon DM White and MJ Rees. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. *Monthly Notices of the Royal Astronomical Society*, 183(3):341–358, 1978.

Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*, volume 7. MIT press Cambridge, MA, 1949.

Brian Yanny, Constance Rockosi, Heidi Jo Newberg, Gillian R Knapp, Jennifer K Adelman-McCarthy, Bonnie Alcorn, Sahar Allam, Carlos Allende Prieto, Deokkeun An, Kurt SJ Anderson, et al. Segue: A spectroscopic survey of 240,000 stars with g= 14-20. *The Astronomical Journal*, 137(5):4377, 2009.

HKC Yee, SL Morris, H Lin, RG Carlberg, PB Hall, Marcin Sawicki, DR Patton, Gregory D Wirth, E Ellingson, and CW Shepherd. The cnoc2 field galaxy redshift survey. i. the survey and the catalog for the patch cnoc 0223+ 00. *The Astrophysical Journal Supplement Series*, 129(2):475, 2000.

Wang Yi. Inflation, cosmic perturbations and non-gaussianities. *Communications in Theoretical Physics*, 62(1):109, 2014.

Donald G York, J Adelman, John E Anderson Jr, Scott F Anderson, James Annis, Neta A Bahcall, JA Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, et al. The sloan digital sky survey: Technical summary. *The Astronomical Journal*, 120(3):1579, 2000.

S Zaroubi, Y Hoffman, KB Fisher, and O Lahav. Wiener reconstruction of the large scale structure. *arXiv preprint astro-ph/9410080*, 1994.

YA B Zel'Dovich. Gravitational instability: An approximate theory for large density perturbations. *Astronomy and astrophysics*, 5:84–89, 1970.

Caroline Zunckel, J Richard Gott, and Ragnhild Lunnan. Using the topology of large-scale structure to constrain dark energy. *Monthly Notices of the Royal Astronomical Society*, 412(2):1401–1408, 2011.

Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. sitzungsberichte der k. *Preussischen Akademie der Wissenschaften zu Berlin*, 1:142–152, 1917.

Fritz Zwicky. Die rotverschiebung von extragalaktischen nebeln. *Helvetica Physica Acta*, 6:110–127, 1933.