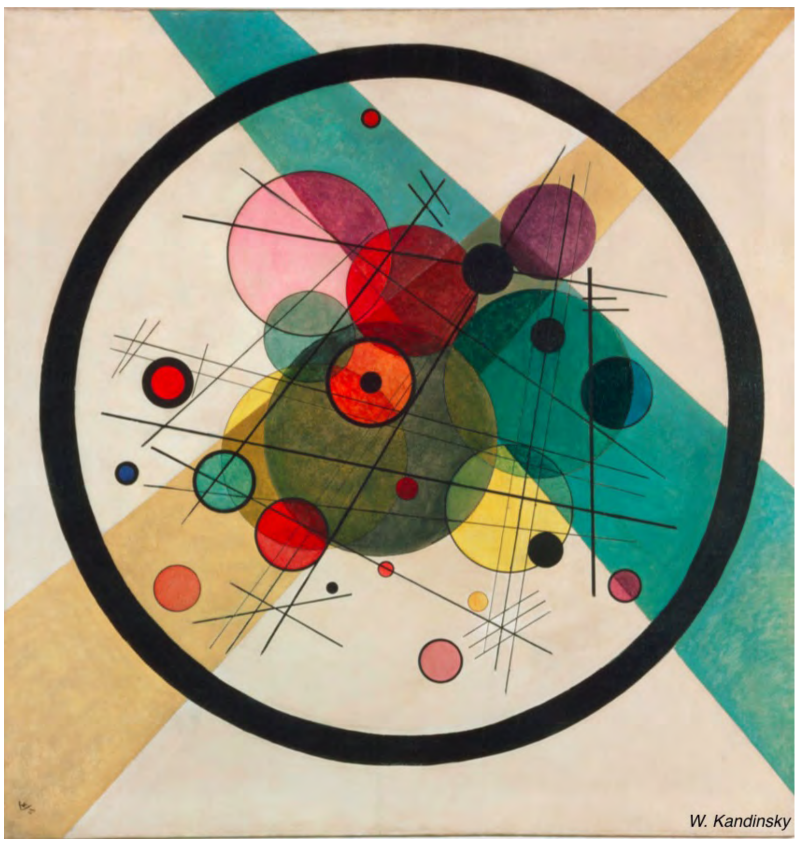
Counting Holes in the Universe

Master's Thesis by Mikkel Stockmann
University of Copenhagen, Dark Cosmology Centre

September 15, 2014



Supervised by
Signe Riemer-Sørensen, University of Oslo (UiO)
Tamara M. Davis, University of Queensland (UQ)
Steen H. Hansen, University of Copenhagen, Dark Cosmology Centre

Abstract

The matter in the Universe is distributed in web-like filaments with large regions of empty spaces in between, called voids. Recent N-body simulations by Villaescusa-Navarro et al. (2011) show a relation between the number of voids and the currently unknown mass of the neutrino, the lightest standard model particle. We were inspired by the results to investigate whether this effect is measurable in observational data. Since voids per definition contain very little matter, they are difficult to measure and can instead be classified from the empty regions between the patterns of the matter distribution. The matter distribution can be established from hydrogen clouds shadowing quasars that are among the most luminous light sources in the Universe.

We investigate the possibility of counting the number of voids, in UVES (VLT) high resolution quasar spectra ($2 \leq z \leq 5$), interpreted as the space between the Lyman-α absorption lines, in order to perform the first void count on real quasar data. From artificial mock spectra, I develop an automatic void counting method that enables us to make a statistical void count. To avoid subjective bias we determine our coding selection criteria from a subsample of 50 quasar spectra before running a blind analysis on the remaining quasar sample. We produce the to date first statistical void count comparison between simulated and observed quasar spectra along line of sight. We find an observed void count offset from the number of voids in simulation but to conclude any further we must increase our number statistics using the HIRES (Keck) data sample.

Acknowledgements

I thank the groups led by Micheal Murphy and John Webb that have provided us with the high resolution quasar data. Vincent Dumont categorised the Keck and VLT spectra in the online database CODDAM, making them easy accessible, thanks. Also a special thanks to Graziano Rossi and Arnaud Borde that provided us with the simulated spectra. When developing the void counting code, I found help in Martin Sparre's expertise in numerical techniques which was greatly appreciated. I thank Amanda and Nina for correction my drafts, showing me better ways to express the points. I owe great thanks to my supervisors Associate Professor Tamara Davis and especially Dr. Signe Riemer-Sørensen for spending many hours discussing interesting physics, and guiding me through the process of this research project. I would also like to thank Dark Cosmology Centre for providing me with the connection to Tamara Davis and Signe Riemer-Sørensen that have made this project possible.

Contents

1 Introdution

The distribution of galaxies from redshift surveys reveal a variety of large scale structures in the Universe. Dense clusters (up to few Mpc in size) is connected by filaments (stretching up to tens of Mpc's) that is sewn around the large under-dense voids (spanning from tens to hundreds of Mpc's).
These enormous structures described in Bond et al. (1996) are pictured in Figure 1, where the galaxies (kpc size) are represented as yellow dots, which in large collections and strings represent the clusters and filaments with the dark empty regions being the voids. This cosmic web is believed to reflect an evolved state of the primordial density patterns of the early Universe.

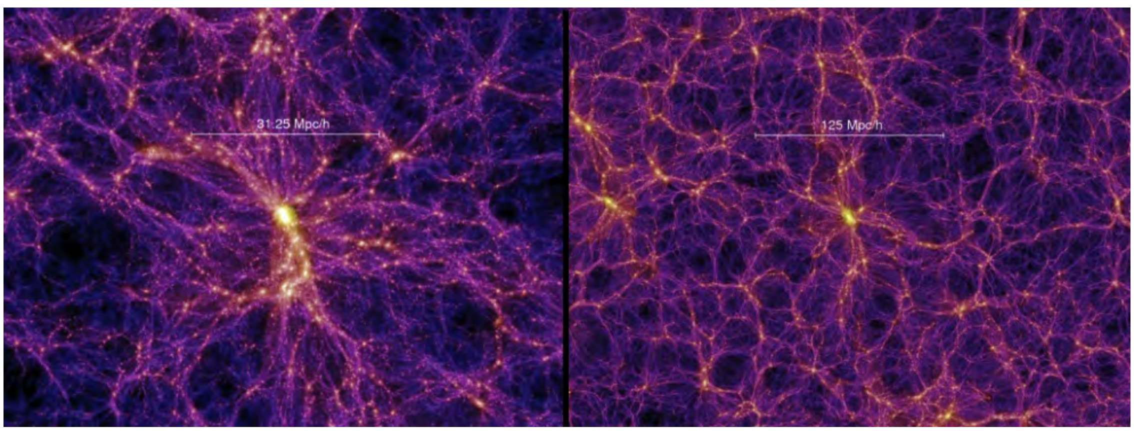


Figure 1: *The Large Scale Structure (LSS) of the Universe from a dark matter simulation. The left plot is a zoom of the right. The large empty regions are voids that barely contain any galaxies. Each yellow dot is a galaxy, with a diameter of 100.000 light years. The collections of yellow dots are groups of galaxies, and the big crossroad of filaments in the middle is a super cluster. The simulation was performed in Garching, Germany by Springel et al. Boylan-Kolchin et al. (2009).*

The traditional focus has been the easy observable structures, but in this project I focus on the voids. The voids occupy the majority of the volume in the Universe and strongly influence the growth of the large structures. The interior void mass affects how quickly the matter is evacuated into the denser areas, and at which pace the large structures of the Universe is formed. The statistics and dynamics of voids therefore provides possible ways of testing todays model of structure formation.

Quasars, one of the most luminous objects observed, are visible out to large redshifts, providing us with information from the early Universe. They are believed to be galaxies hosting super massive black holes ($M \sim 10^8\ M_\odot$) that send out jets of very energetic light (Urry and Padovani, 1995). On its way towards us the light passes through neutral hydrogen clouds (galaxies and proto-galaxies), which leave absorption signatures in the spectrum (see Figure 2). Counting the number of signatures in the Lyman-α forest directly determine the number of voids along line of sight (LOS).

We were motivated to count voids in observed spectra from recent simulations that illuminated the relationship between the sum of neutrino masses and the number of voids along LOS (Villaescusa-Navarro et al., 2011). The aim of the project was to develop an automatic code to count the number of voids along LOS in observed high resolution quasar spectra ($2 \leq z \leq 5$) from the UVES spectrograph (King et al., 2012). I constructed the code from artificial mock spectra to avoid subjective bias. A blind analysis was then performed on the resulting quasar sample, to determine if the observed void count distribution match the number of voids in spectra from N-body simulations by Rossi et al. (2014).

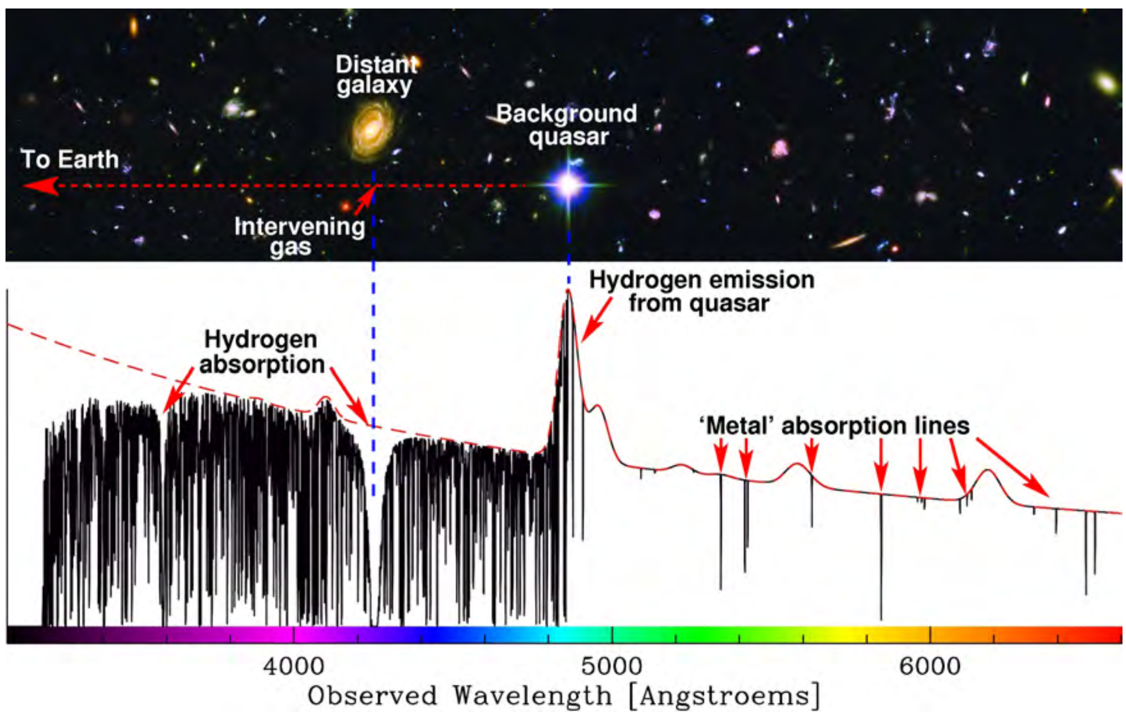


Figure 2: *A quasar spectrum revealing a forest of absorption lines. The absorption occurs when the emitted light from the quasar interacts with the neutral hydrogen distributed on cosmological scales on its way to the observer. Credit: Michael Murphy and John Webb.*

The reader less familiar with cosmology can skip to section A for a historical and conceptual overview of astrophysics. Section 2 describes the cosmological background and the effect neutrinos have on voids as the Universe evolves. The concept of the project can be found in 3, and section 4 introduces the different data types and how they are prepared for the analysis. Section 5 explains the practical challenges and solutions occurring when observing voids. A robust void count is developed in section 6, whereafter the uncertainty from the quasar spectrum normalization is estimated (section 7).

We use the final quasar sample from section 4.3 to perform the first robust void count on observed spectra in section 8. The statistical methods used throughout the analysis are covered in Appendix B. We summarize the results in section 9, followed by ideas on further work (section 9.1) before we briefly conclude in section 10. Appendix C, includes a sample of extra figures also referred to throughout the thesis.

2 Cosmological Background

In this chapter I will cover the cosmological background that provides the basis of the thesis as well as subsequent discussions. The first subsection will introduce the concordance model, using the expanding Universe, the cosmological principle, and Friedmann's equation for accelerated expansion. Section 2.2 will address the formation of structure from Big Bang up until the galaxies and clusters we observe today. I will especially focus on the evolution and dynamics of voids and their interplay with cosmological neutrinos.

2.1 The Expansion of the Universe

When we look out in the Universe we observe enormous structures like galaxies (kpc size) and galaxy clusters (Mpc size). They are linked into a cosmic web of filaments sewn around the large empty regions called voids, as illustrated in Figure 1. We observe all extragalactic objects to increase their distance to us (as well as to each other), and we interpret this as an expanding Universe (Hubble, 1929a). An expansion of the Universe implies that at earlier times it was smaller, hotter, and denser. This fact lays the foundation of the Big Bang paradigm (Lemaître, 1927).
The time of the Big Bang is defined to be, when the distance between all structures were zero, about 13.7 billion years ago (Planck Collaboration, 2013). The Universe started in the Big Bang, expanded, cooled, created atoms, and molecules which later condensed into the planets, stars, galaxies, and clusters we see today.
Edwin Hubble was the first astronomer to observe the expansion of the Universe, by measuring how the recession velocity of galaxies are related to the distance. This we call Hubble's law (Hubble, 1929b):

$$v = cz = H_0 D. \tag{1}$$

Hubble found a linear relation between the recession velocity and the distance, described by Hubble's constant. The recession velocity, v of a galaxy can be found be measuring its redshift,

$$z = \frac{\lambda_{obs}}{\lambda_{em}} - 1 \tag{2}$$

where λ is respectively the observed (obs) and emitted (em) wavelength. Hubble's constant describes the rate at which the Universe expands, and is usually presented like $H_0 = 100\ h\ km\ s^{-1}\ Mpc^{-1}$ with the current value h being 0.6780 ± 0.0077 (Planck Collaboration, 2013). Unless otherwise stated we adopt this value throughout the thesis.

2.1.1 The Cosmological Principle

In light of the Copernican principle we assume that our observation point in the Universe is not privileged so we live in a so-called isotropic Universe. When we further observe that the matter distribution, on scales larger than 100 Mpc/h, is roughly homogenous, we can exploit these symmetries to write down a metric that describes the relation between time (dt) and space ($d\chi, d\phi$) in an expanding Universe (Friedmann, 1922a).

$$ds^2 = -c^2dt^2 + R(t)^2\Big[d\chi^2 + S_k^2(\chi)d\psi^2\Big] \tag{3}$$

This metric is known as the Friedman-Lemaître-Robertson-Walker metric but often referred to as FLRW. The spatial part is subjected to a scaling factor, $R(t)$, that ensures isotropic treatment of the radial and angular expansion of the Universe. The $S_k(\chi)$ describes how the angular part changes with different geometrical curvature. $S_k(\chi) = sin(\chi), \chi, sinh(\chi)$ for $k = 1,\ 0,\ -1$ corresponding to a closed, flat and open Universe respectively. The size of the acoustic scales in CMB map suggest that we exist in a flat Universe with $k = 0$. The time, dt, in the FLRW metric is cosmic time that is sliced so the density is uniform for all observers at each time, t.

Einstein introduced, through his theory of general relativity, an equation describing the relation between how space time curves (R) and the stress-energy in the space-time (T) resulting from gravitation (G).

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G T_{\mu\nu}, \tag{4}$$

When we apply the symmetries of the FLRW metric and solve the Einstein equation we obtain two equations describing the Universe as a fluid. The Friedmann equation and the cosmic fluid equation.

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{5}$$

$$H^2 + \dot{H} = \left(\frac{\ddot{a}}{a}\right) = -\frac{4\pi G}{3}(\rho + 3P) \tag{6}$$

These equations govern the relation between density (ρ), pressure (P), curvature (k) and size of the Universe in terms of scale factor (a). Consequently we can describe the evolution of the Universe purely from Einsteins theory of relativity and the assumptions of homogeneity and isotropy. This will be addressed in more detail in the next section.

2.1.2 The Friedmann Equation for a Dark Universe

Recent observations of supernova Ia suggest that the expansion of the Universe is accelerating (Perlmutter et al., 1999; Riess et al., 1998), and slowly increasing the rate at which the intergalactic distances are growing. When this effect, referred to as dark energy or Λ, is added to the Friedmann equation we can write it as follows

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}. \tag{7}$$

Hence the dark energy density is defined as $\rho_\Lambda = \Lambda/8\pi G$. We can write the total density as a sum of the matter, radiation and dark energy:

$$\rho_{tot} = \rho_M + \rho_{rad} + \rho_\Lambda. \tag{8}$$

Introducing the energy density relative to the critical density we have

$$\Omega_i = \frac{\rho_i}{\rho_{crit}}, \qquad \rho_{crit} = \frac{3H^2}{8\pi G}. \tag{9}$$

The critical density is the energy density of a flat Universe (k = 0) at present time. We can rewrite the Friedmann equation in terms of the relative energy density.

$$\Omega(a) - 1 = \frac{k}{H^2 a^2}, \qquad \Omega(a) = \Omega_M + \Omega_{rad} + \Omega_\Lambda \tag{10}$$

This version of the Friedmann equation easily shows the relation between Ω and curvature so $k = (1,\ 0,\ -1)$ gives us respectively $\Omega(a) > 1$ (closed), $\Omega(a) = 1$ (flat), $\Omega(a) < 1$ (open). We can define the energy density of curvature, $\Omega_k = k/H^2a^2$ and write the Friedmann equation in a intuitive form.

$$H(a)^2 = \left(\frac{\dot{a}}{a}\right)^2 = H_0^2\Big[\Omega_{rad,0}a^{-4} + \Omega_{M,0}a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda\Big] \tag{11}$$

Here we introduced $\Omega_x = \Omega_{x,0}a^u$, where u determines the rate at which the density of x drops with expansion. $\Omega_{x,0}$ is the relative energy density today, when we use the definition of $a(t_{today}) = 1$. Today the Planck Collaboration (2013) measure the curvature and radiation density to be very close to zero ($\Omega_{rad,0} \sim \Omega_k \sim 0$) so often this equation is written without these terms. In section 5.3 we show how to use the equation (rewritten) to estimate void sizes in the Universe.

The ΛCDM model best fits the Universe with energy components, $\Omega_{M,0} = 26.8$ % and $\Omega_\Lambda = 68.3$ % (Planck Collaboration, 2013). It is the model that describes an accelerated expansion, and have a dark matter to baryon (visible matter) ratio of, $\rho_{DM}/\rho_b \sim 5$. If we sum up the total energy densities, it means that dark matter and dark energy form ~ 95 % of the density, leaving all the matter we can see, and describe with the Standard Model of Particle Physics (Beringer, 2012), to constitute only ~ 5 % of the total energy density of the Universe.

2.2 Formation of Structures

The growth of structure through the age of the Universe has become what we observe as galaxies, clusters and voids today. The primordial density perturbations in the early Universe left an imprint that evolved and became what we today refer to as the Large Scale Structure (LSS) of clusters, filaments, and voids.

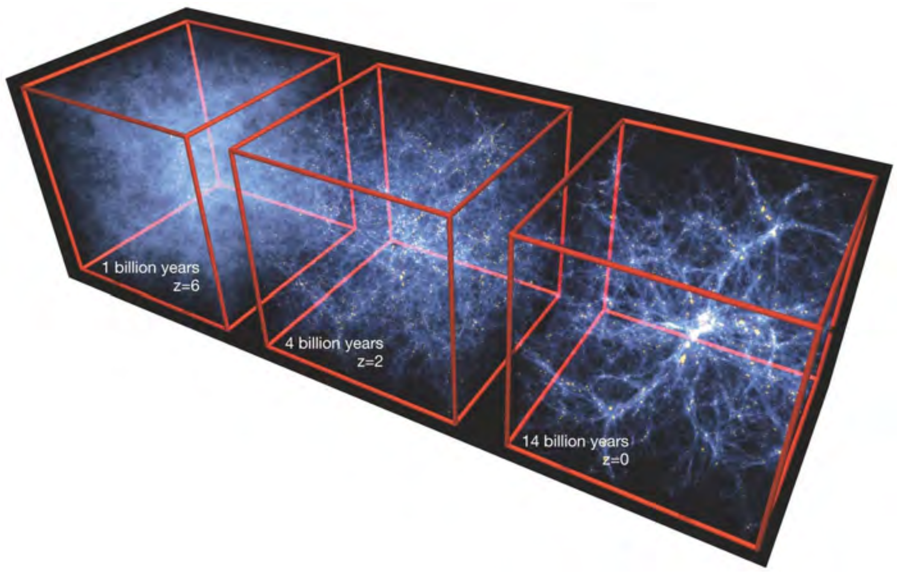


Figure 3: *The evolution of matter densities and voids shown in a N-body simulation box from $z = 6 \rightarrow 0$. Credit: Pearson Education 2008.*

At $10^{-36}s$ after Big Bang the Universe entered the inflationary state that made it expand at a rate faster than the speed of light, c, until $t = 10^{-32}s$. During this phase the tiny quantum fluctuations were enlarged and seeded the primordial density perturbations that gravitationally evolved into the LSS we see today.

While the Universe expanded, it cooled continually until the energy density of matter was similar to that of radiation at $t = 75.000$ yr (radiation-matter equality) . From here on the non-interacting dark matter started collapsing into the initial perturbations from inflation, while a veil of photons and baryons, on top, was locked into a plasma (due to a high interaction rate).

Not until the Universe had expanded and cooled sufficiently, did the electrons and photons decouple ($\sim 378,000$ years after Big Bang) and were free to move into the already created dark matter potentials. It was in these potentials the formation of stars and galaxies started (after $\sim 4,000,000$ years). Over- and under-densities became respectively the large clusters and empty voids we observe today. The shape and size of the LSS is strongly related to the evolution of the initial density perturbations, as quickly sketched above, through a complicated network of gravitational interactions and mergers. The formation of structures is too complicated to determine analytically and is usually studied via N-body simulations (Springel et al., 2005; Villaescusa-Navarro et al., 2011; Rossi et al., 2014; Vogelsberger et al., 2014).

2.2.1 The interplay between Neutrinos and Voids

In the Standard Model of particle physics (Beringer, 2012) the neutrino species forms 3 out of the 12 fundamental particles that makes up $\sim 5\%$ of the energy in the Universe (see Figure 4).

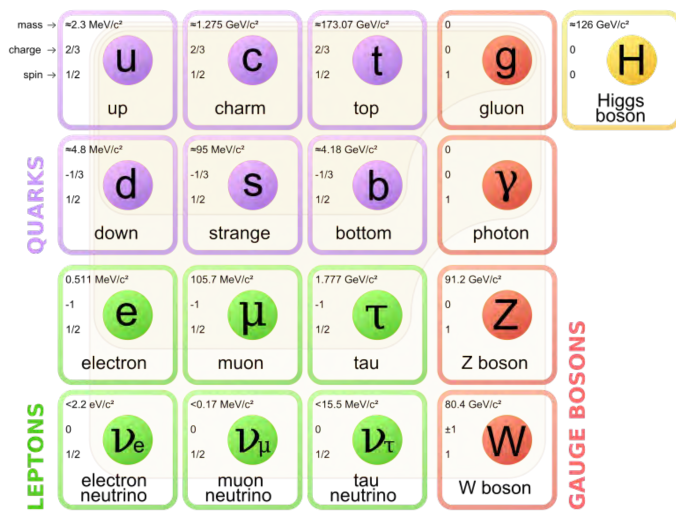


Figure 4: *A sketch of the 12 fundamental particles (green + purple), the 4 force carriers (red) and the Higgs particle (yellow). The fundamental blocks consist of 6 quarks and 6 leptons that make up all the luminous matter. Credit: Wikipedia.*

The family of neutrinos are fermions (half integer spin) and consists of three different kinds, the electron (ν_e), muon (ν_μ) and tau (ν_τ) neutrino (bottom three green particles in Figure 4). Unlike the electron they are neutral and only interact via the electroweak and gravitational force making them very difficult to detect (see e.g. IceCube Collaboration et al. (2013)). Both forces are weak at the subatomic scale, gravity being the weakest of the two, and they mostly pass through normal matter unimpeded. The neutrino oscillate between the different species which requires it to have mass (Daya Bay Collaboration,

2007). The cosmological neutrinos affect the evolution of the LSS. How is it possible for the smallest particle (10^{-24} m) of the Standard Model to affect voids (10^{15} m) the largest structures in the Universe?
To answer the question we first need to understand the formation scenarios of cold dark matter (CDM) and warm dark matter (WDM). Here cold and warm refers to the clustering properties of the dark matter determined by their temperature when decoupling from the initial plasma.

In the previous section I mentioned that the large scale structures of the Universe formed from initial over- and under-concentrations in the density field, leading to respectively galaxy clusters and voids at present day. We believe this formation scenario occurred from a bottom-up CDM scenario (Frenk and White, 2012), where small structures (pc - kpc size) form first, and then larger (Mpc size) in due course. Dark matter must have been dynamically cold to force a collapse that can create structures on the sizes we observe today. Neutrinos on the other hand are warm (high velocities) and follow a top-down formation scenario, where larger structures (Mpc size) form first and then later fragment into smaller (kpc - pc size). They are a WDM candidate due to their complicated gravitational effect in the non-linear regime.

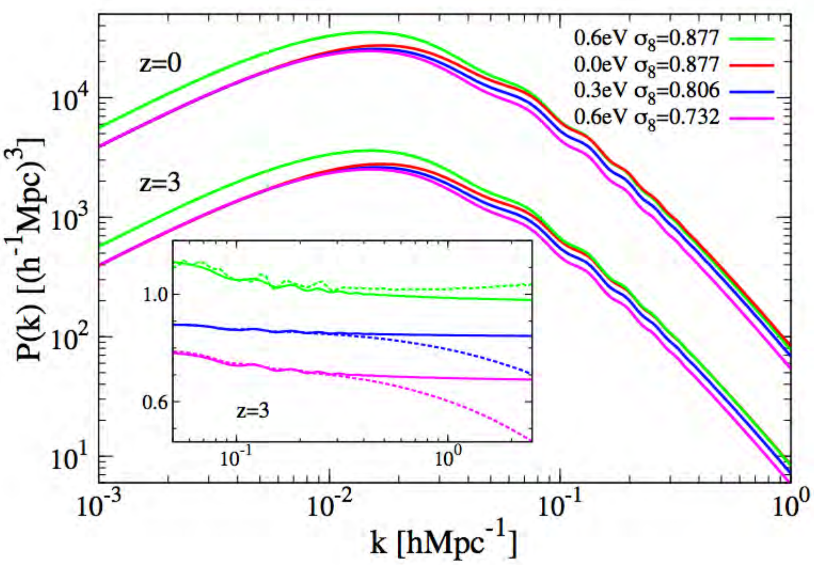


Figure 5: *The linear matter power spectrum from Villaescusa-Navarro et al. (2011). Two curves are shown for respectively $z = 0$ and 3 (upper, lower). Each simulation was run with increasing neutrino masses ($\sum_i m_{\nu_i}$) resulting in a damped power spectrum at small cosmological scales.*

The cosmological neutrinos, from the early Universe, are relativistic and redistribute their mass on small scales. This results in a delay of the gravitational collapse of DM structures, an effect that can be measured in the matter power spectrum that describes the amount of clustering on different scales. A cosmological power spectrum describes the amplitude of the density fluctuations (y-axis) that is present at each scale size (x-axis). In Figure 5, I have included the matter power spectrum from a simulation by Villaescusa-Navarro et al. (2011). The red, blue and purple lines describe respectively $\sum_i m_{\nu,i} = 0.0, 0.3, 0.6\ eV$, whereas the green line is an example of the red line with similar initial conditions ($\sigma_8 = 0.877$)[1] but $\sum_i m_{\nu,i} = 0.6\ eV$. The peak in the power spectrum at $k = (1-2) \cdot 10^{-2}$ represents the largest density amplitude in the Universe. In this figure k describes the inverse scale, where small values correspond to large physical sizes, e.g. $k = 10^{-2}\ h\text{Mpc}^{-1}$ represents a size of 100 Mpc/h. The cosmological neutrino suppresses the power spectrum on small physical scales < Mpc size (large k) proportional to their mass. This means they smooth out structures below their free streaming scale ($v < c$) trying to equilibrate in the medium.

In voids the neutrinos contribute to the interior mass, and have contradictory to heavy DM halos an important effect. The extra neutrino mass present, delay the rate at which CDM is abandoning the under-dense regions and thereby increasing the void size. In Villaescusa-Navarro et al. (2011) they solved the dynamical equations for an isolated spherical top-hat under-dense perturbation, and found that their simulated neutrinos modify the evolution of voids, making them smaller and denser. They find that this effect can be seen in the relation between $\sum_i m_{\nu_i}$ and number of voids along LOS.

[1]The σ_8 parameter describes the mass fluctuation amplitude (amount of gravitational clumping) in a sphere of 8 Mpc/h.

2.3 Quasars as Cosmological Probes

The Universe has on the largest scales been probed using the light from galaxies and galaxy clusters. The intervening clouds leaves imprints in the spectra via the Lyman absorption and can be used as tracers of structure, (see Seljak et al., 2006; Busca et al., 2013, and references therein). Voids are almost empty regions incorporated into the matter distribution and it seems fair to assume that places without Lyman absorption are potential voids.

In Viel et al. (2008) they find the 1D empty flux regions (Lyman-α voids) to be related to the 3D dark matter void distribution in simulations (see section 8.1), and we can safely attempt a realistic void count in quasar spectra.

To gather information over the course of structure formation we are interested in extending the void search to high redshifts ($2 \leq z \leq 5$), but also find sources with a light source luminous enough to reach us from these early times. The very luminous quasars are one of the most energetic light sources in the Universe and extend out to high redshifts, making them qualified candidates to detect voids.

Figure 6: *M. Kornmesser's (ESO) interpretation of a quasar. We see the distinct jet coming from the center of the accretion disc. This is the emission we use as background light to study intervening gas and voids.*

Quasars are an Active Galaxy Nuclei subclass, and was up until the 1980's a controversial subject but are now believed to be super massive black holes consuming gas in the centre of galaxies (Urry and Padovani, 1995). By doing so they release a large amount of energy into thin jets perpendicular to the in-falling gas disk (see Figure 6). The powerful jets light up clouds of gas along the LOS that can be used as tracers of structures (see both Figure 7 and 18). The Universe's visible component mainly consists of neutral hydrogen that produces, for each structure in the LOS, a Lyman signature in the spectra (Ryden and Peterson, 2009).

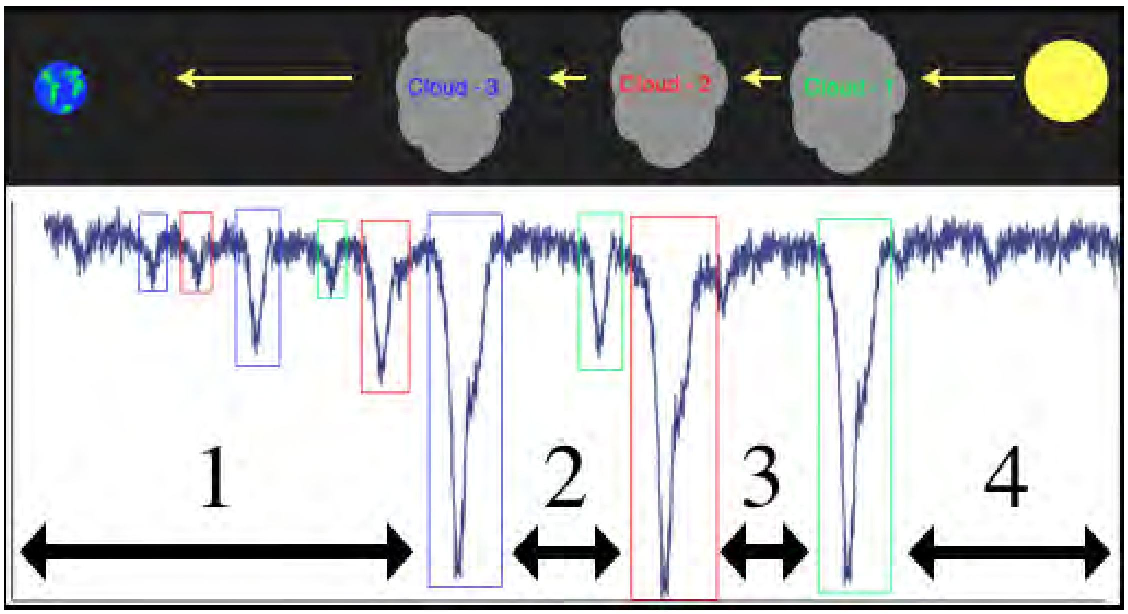


Figure 7: *I have made a simplified version of our Universe, with one source (yellow) e.g. a quasar, and three gas clouds (e.g. galaxies) along the LOS to Earth. Each cloud leaves a Lyman α signature in the spectrum, making it possible to determine the number of clouds in the LOS, and thereby the number of voids, in this case four.*

The signature occurs due to excitation jumps produced by the background quasar photons, and is known as the Lyman series when the electrons are excited from the ground state. Each jump to the ground state releases a photon of discrete energy, starting from the lowest energy photon; Lyman-α, -β, -γ,... ending at the ionization level of hydrogen (13.6 eV or 911 Å).

Counting each signature in a spectrum gives us the number of clouds along the LOS, which is easy convertible to the number of voids (the empty regions between the clouds). The light from galaxies and clouds in the Universe gets redshifted which can be seen in the spectra when the absorption signatures are shifted towards higher wavelength, making it possible to distinguish individual clouds along the LOS. The redshift is defined as

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} \tag{12}$$

where λ_{obs}, λ_{em} are the observed, emitted wavelengths respectively. The redshift increases with distance/look back time, and it is often used as a distance measure. In Figure 7 I have made a simple Universe with three clouds along the LOS, and restricted the signature to only exhibit α, β, γ Lyman absorption lines. The before mentioned redshift effect, can be seen as a shift of the lines towards the blue (left) end of the spectrum with respect to the quasar. In this example four voids are present in the spectrum. The real quasar spectrum in Figure 8 are unfortunately not as easy to analyze as the simplified example just mentioned. This challenge will be addressed in section 5.2.

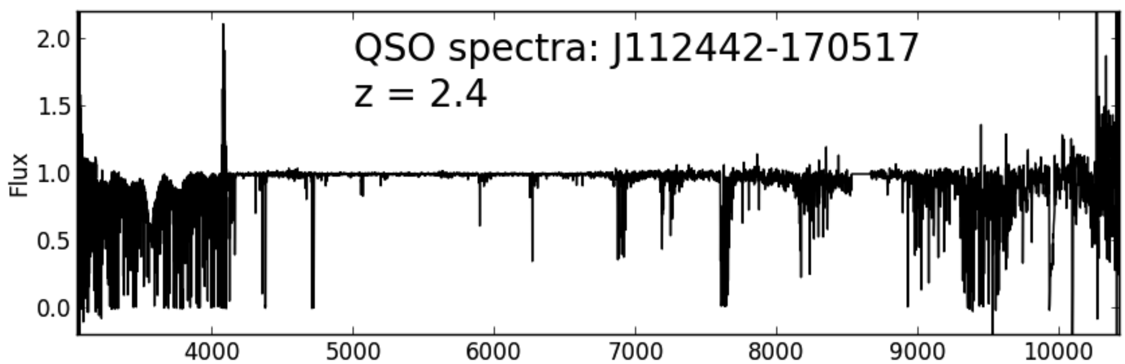


Figure 8: *The spectrum of the quasar with the complicated name, J112442-170517 located at* $z = 2.4$. *The x-axis shows the observed wavelength in Ångstrøms and the y-axis shows the observed flux normalized to the quasar emission flux so one can focus on the absorption features (standard procedure). Obtained from the sample of quasar data by King et al. (2012).*

3 Project Outline

In this thesis I will explore the possibility of counting voids in high resolution quasar spectra. Quasar spectra exhibit Lyman signatures, from each cloud of neutral hydrogen, encountered along the LOS (see e.g. Figure 2). The number of signatures in the Lyman-α forest is directly related to the number of voids.

I have developed an automated method that counts the number of voids following a strict set of selection criteria. The number of voids in each spectrum depends on how we define a void which again is affected by the quality of the data, flux density threshold, smoothing scale, and the effect of gravitational clumping through line blending. We use artificial mock spectra to determine the criteria and obtain a void count as independent of the chosen parameter values as possible.

I run our method on a subsample of 50 quasar spectra to test if our mock selection criteria are robust on real quasar spectra before running the analysis blindly on the remaining sample. We currently have access to 413 high resolution quasar spectra observed with the UVES[2] spectrograph on the world's largest telescope VLT[3] (King et al., 2012).

We make a prime quasar sample selected from good S/N and a low number of data-less pixels, to test possible systematic uncertainties. The aim is to produce a statistical void count at different redshifts, and use the simple void counting method to test if it matches the number of voids from numerical simulations. In Figure 9, I have included a flow chart showing the steps of my void counting code for the quasar and simulated spectra sample leading up to the analysis. The red boxes show the different code steps for the quasar spectra. Each specific step can be found in section 5.1 (Lyman-α forest), 6.1.1 (data-less pixel removal), and 6.1.2 (Fourier smoothing). The simulated spectra are assumed to be a perfect no noise spectra with only one coding step, the gaussian smoothing (see section 6.3.1).
With the processed spectra we perform a comparison between the simulated and observed void counts in section 8. We also reproduce the plots in Villaescusa-Navarro et al. (2011) from observational data.

[2] the Ultraviolet and Visual Echelle Spectrograph
[3] Very Large Telescope

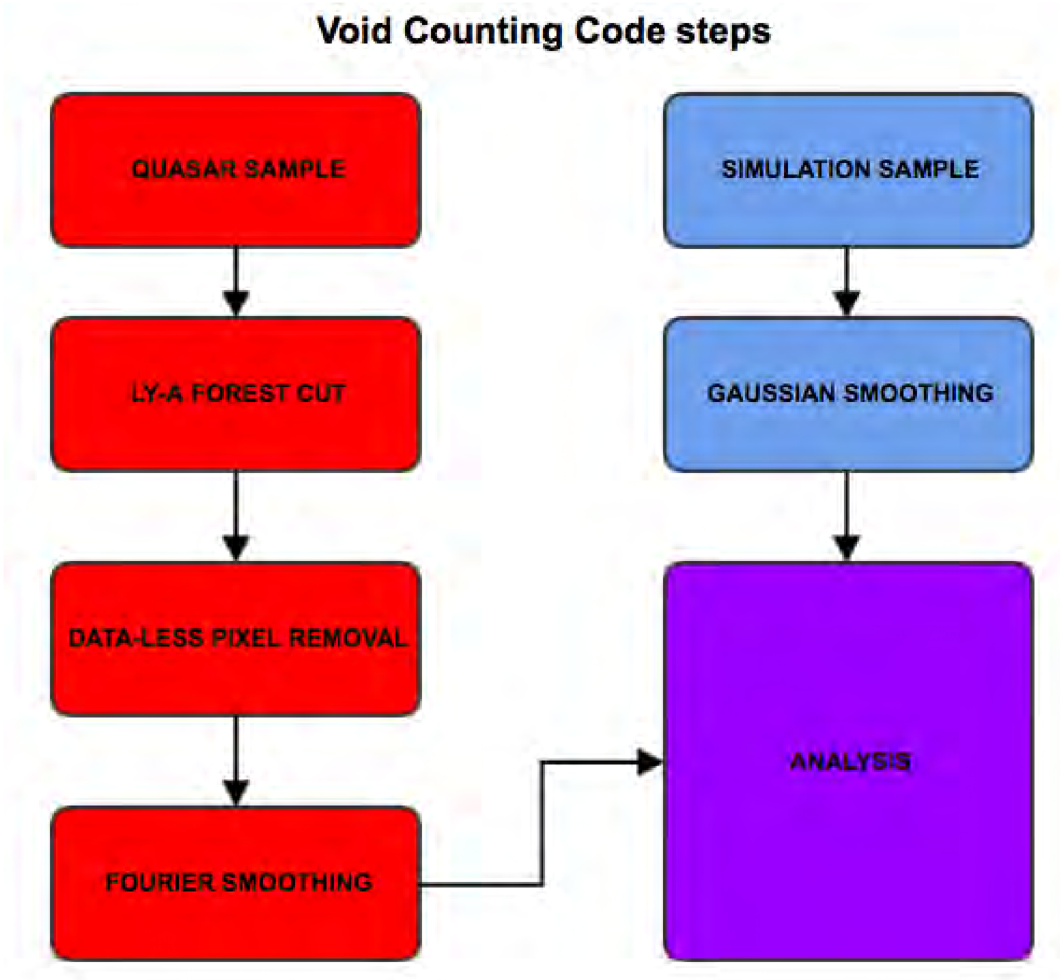


Figure 9: *A flow chart showing the important steps in the void counting code. The red colours indicate the quasar code steps, the blue the simulations, and the purple is the final comparison of the observed and simulated void count.*

4 Data Sample

In this chapter we introduce the artificial mock spectra, the simulated spectra, and the quasar spectra including a set of steps preparing the sample for the analysis.

4.1 Artificial Mock Spectra

Mock spectra are synthesized spectra that looks like real data, where we can artificially change different properties such as noise, redshift and line distribution. The mock spectra are constructed using the mock generator code from Liske et al. (2008).
We have investigated two types of mock spectra. The "perfect" spectra does not contain any noise similar to observational noise and as such all fluctuations origin in physical properties (void or matter). We compare the perfect spectra to identical spectra with observational noise imposed.

4.2 Simulated spectra

Graziano Rossi and his group at CEA - Saclay have provided us with the gas density distribution from their numerical simulations (Rossi et al., 2014). The best guess simulation is run with ΛCDM parameters from the Planck Collaboration (2013). They have used the standard Boltzmann code CAMB[4] to estimate the initial mass fluctuation amplitude, σ_8 (see footnote in section 2.2.1), at $z = 49$ where they kick-start their simulation.
The simulated spectra are produced by measuring optical depth of hydrogen (τ_{HI}) in the Lyman-α forest along Random Line Of Sights (RLOS), which translate into the flux in the following way,

$$F = e^{-\tau_{HI}}. \tag{13}$$

We have received 10.000 RLOS (arbitrarily positioned) in a 100 Mpc/h box with 768^3 particles at respectively $z = 2.2$ and 4.0. In Villaescusa-Navarro et al. (2011) they use $z = 2.2$ and 4.0 and we naturally adopted these choices, to construct a test of our results. If our preliminary void count comparison allow for constraining the neutrino mass within the errors, we have a grid of simulations with different cosmological parameters such as a varying neutrino mass ($\sum_i m_{\nu_i} = 0.1, 0.2, 0.3, 0.4$ and 0.8 eV) on which we can perform an additional analysis.

4.3 Quasar sample

We have access to 413 high resolution quasar spectra from UVES (VLT) situated in Chile. The raw quasar data have been reduced by the groups led by John Webb (University of New South Wales) and Michael Murphy (Swinburne University of Technology) and are accessible as normalized quasar emission spectra (King et al., 2012).

[4] http://camb.info/

It is to date one of the largest sample of normalized high resolution ($R \sim 80.000$[5]) optical spectrum. The UVES detector spans a wavelength width of $3000 - 11000$ Å [6]. The high resolution spectra have S/N noise values from, $2 - 128$ with a average values of 23 (see Appendix B.1). We have access to additional ~ 150 HIRES (Keck) quasar spectra, which apart from formatting issues can be directly implemented in a future analysis.

Below I will explain how we selected our quasar sample from the UVES (VLT), using the ranges of the Lyman-α forest. Artificial and unnatural spectra are blacklisted from the Lyman-α quasar sample (section 4.3.2) and from the resulting selection a prime sample is chosen using the signal to noise (S/N) and data-less pixel removal (section 4.3.3).

4.3.1 Lyman-α Forest selection

The position of the Lyman-α forest in the spectra depends on the redshift at which the quasar light was emitted. We select which of the spectra with their Lyman-α forest in the UVES detector wavelength range. We calculate the minimum and maximum redshift that a spectra can take before its Lyman-α forest is excluded by the detector range.

$$z_{lower} = \frac{\lambda_{obs,cut}}{\lambda_{\beta,em}} - 1 = \frac{3000\ \mathring{A}}{1025.72\ \mathring{A}} - 1 = 1.93 \tag{14}$$

$$z_{upper} = \frac{\lambda_{obs,cut}}{\lambda_{\alpha,em}} - 1 = \frac{11000\ \mathring{A}}{1215.67\ \mathring{A}} - 1 = 8.05 \tag{15}$$

The UVES quasar sample have redshifts in the range between $z = 0.014 - 5.25$. We can unfortunately only use the spectra between $1.93 \leq z \leq 8.05$ and must exclude a large part of our sample because their Lyman-α regions fall outside the wavelength range of the detector.

4.3.2 Unnatural spectra

A range of quasar spectra with $S/N < 3$ exhibited very artificial Fourier transformed spectra while the flux spectra looked normal except for few very bad pixel values. I have included an example in Figure 10 that clearly shows a supposedly normal spectrum with a corresponding artificial Fourier frequency spectrum. These "unnatural spectra“, may be arising from incorrect normalization, and will be removed from the quasar sample. These spectra need a manual re-reduction and continuum fitting before entering the void counting sample, but that is out of the scope of this project.

[5] http://www.eso.org/sci/facilities/paranal/instruments/uves/inst.html

[6] http://www.eso.org/sci/facilities/paranal/instruments/uves/doc/VLT-MAN-ESO-13200-1825_v94.1.pdf (section 3.1 Capabilities of the Instrument)

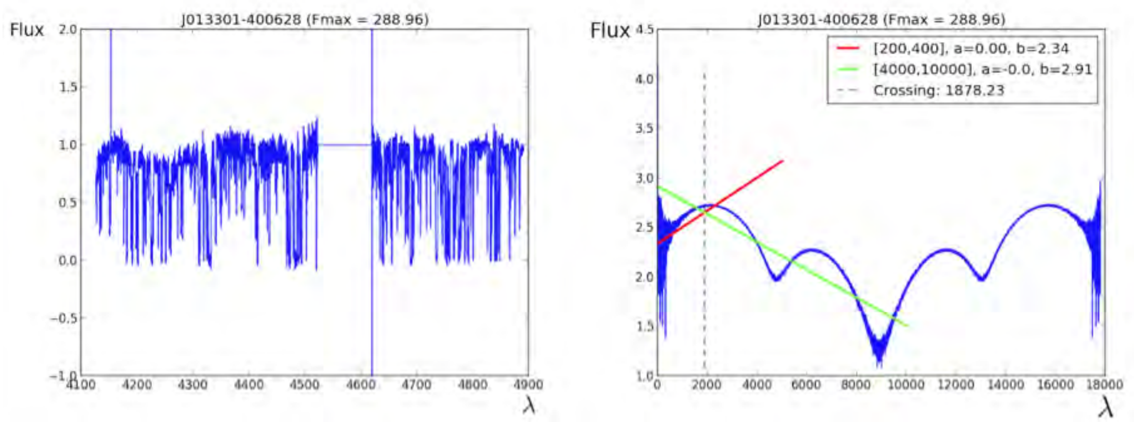


Figure 10: *An example of a unnatural spectrum (left) and the Fourier frequency spectrum (right). The spectra have a few bad pixels but nothing that gives away its very artificial frequencies that can be seen in the right plot. The functionality of the green and red line, in the right figure, will become clear in section 6.1.3.*

The spectra were classified using the flux difference, $\Delta F_{max} = F_{max} - F_{min}$, which for a perfectly normalized spectrum is maximum 1. Based on this, spectra with $\Delta F_{max} > 20$ have been blacklisted from the total quasar sample. The ΔF_{max} cut-off is determined from a set of spectra with artificial frequencies (this sample is also used in section 6.1.3). We have reduced our total quasar sample, and will perform our blinded analysis on the resulting (good) spectra. In Figure 11, I have plotted the ΔF_{max} against S/N. We see that most spectra with very low S/N values also have high ΔF_{max} values.

4.3.3 Prime sample

We construct a prime sample, of the total quasar sample, from spectra with good S/N and a low number of data-less pixels (D-LP), see section 6.1.1. It is done in order to test if the bad S/N spectra and the data-less pixel removal introduce artificial voids that adds a systematic error in the result. We choose the 50% best spectra with, $\overline{S/N} < S/N$ and $D - LP < \overline{D - LP}$. The prime sample will be used in the analysis of section 8.

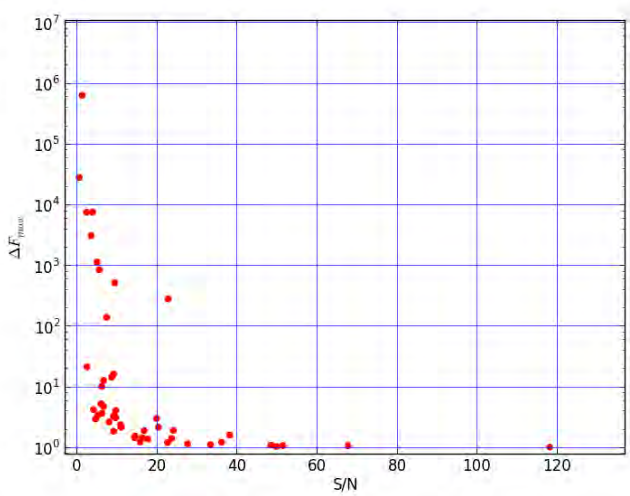


Figure 11: $\Delta F_{max} = F_{max} - F_{min}$ *plotted versus the S/N values. Spectra with high* ΔF_{max} *values clearly have unnatural pixel values and seem to be correlated with their S/N values. The* ΔF_{max} *cut-off is determined looking at the Fourier transformed spectra.*

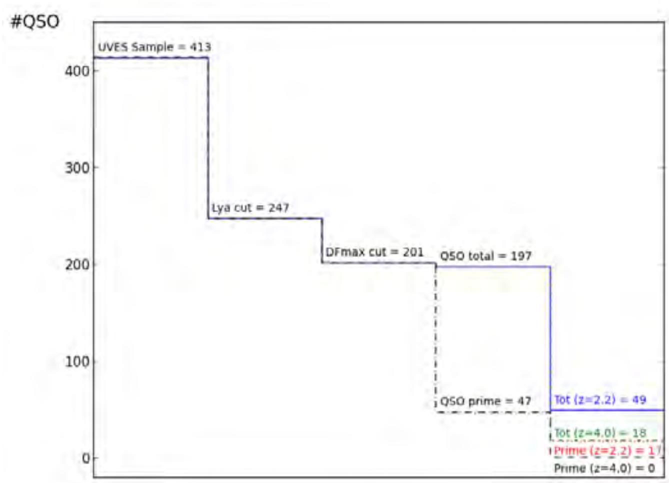


Figure 12: *The numbers of quasar spectra throughout the selections, such as redshift cut,* ΔF_{max}*, the total and prime and their consecutive redshift bins.*

4.3.4 Final sample

In Figure 12, I show an overview of the number of UVES quasars through the process of each selection. Here the "Lya cut" refers to the redshift cut from the detector resolution (section 4.3.1), the "DFmax" (ΔF_{max}) to the removal of unnatural spectra (section 4.3.2), followed by the numbers of the total and prime sample that in section 8 is splitted into redshift bins. The Lyman-α forest represent only a part of the total quasar spectrum and in order to make a statistical void count from our final sample we must ensure that we have a continuous redshift sample with enough Lyman-α forest overlaps.

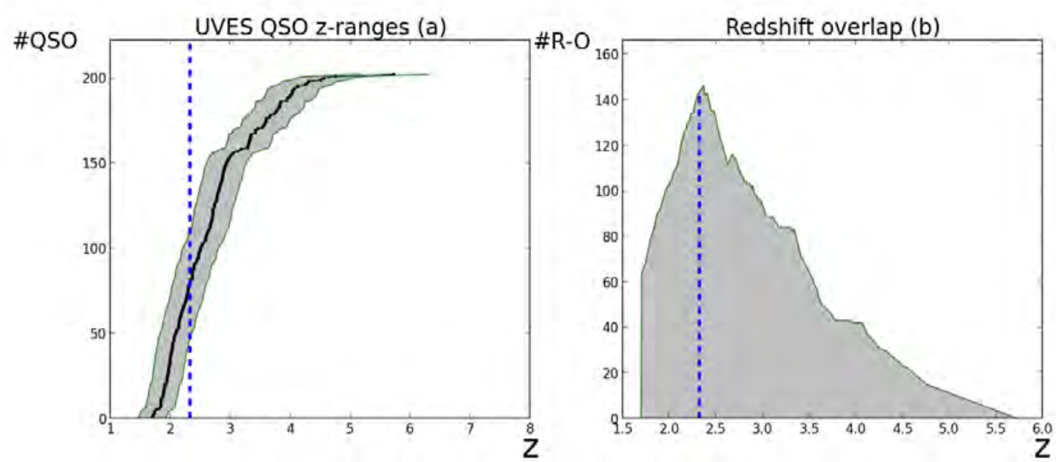


Figure 13: *The left figure (a) shows each of the UVES quasar spectra (y-axis) with their corresponding Lyman-α forest (grey section parallel to x-axis) shown in redshift space (x-axis). At the vertical blue dashed line at $z = 2.4$ I have counted the number of spectra with overlapping Lyman-α forest. The number of overlaps ($\#R - O$) as a function of redshift is shown in the right figure (b). E.g. at redshift at $z = 2.4$ we have around 140 quasar spectra with overlapping Lyman-α forest. The cut-off at $z = 1.75$ is due to the detector sensitivity range.*

I have in Figure 13 (a) plotted each of the UVES quasars Lyman-α range (y-axis) as a function of redshift (x-axis). We are interested in knowing how many Lyman-α regions overlap, and thus how many spectra can contribute with a void count at each redshift. The number of quasars crossing the blue dashed line (the number of overlaps) in figure (a) is shown in (b) as a function of redshift. This is a simplified test estimating that over 80 quasar spectra in the redshift range between [2 - 3.5] can participate in the statistical void count. For the region above $z = 3.5$, the number of overlaps decrease drastically and we will have to be more careful when using this range. This shows that we have enough quasar data to perform a good statistical void count around $z = 2 - 3.5$.

5 Observing Voids in Quasar spectra

When counting voids in quasar spectra, one immediately encounters problems such as the resolution of the signature near the Lyman forest and choice of flux density threshold. This section contains a practical explanation of how to produce a robust and simple void count using the Lyman-α forest (section 5.1). In section 5.2, I will address the importance of choosing a proper flux density threshold. In the last section I use equation (3) to make a back-of-the-envelope calculation of the relation between void sizes in spectra (Å) and comoving real void sizes (Mpc/h).

5.1 Lyman-α forest

Electrons absorb photons at different energies from neutral to ionized Hydrogen. The transition towards the ionization of Hydrogen creates a range of absorption lines known as the Lyman signatures (see Figure 14). They become increasingly close and indistinguishable when approaching the ionization level, and appear as a forest of lines, and so forth the name "Lyman forest". The forest complicates the void counting process and, in the first robust attempt, we avoid this region by only counting voids in the Lyman-α forest.

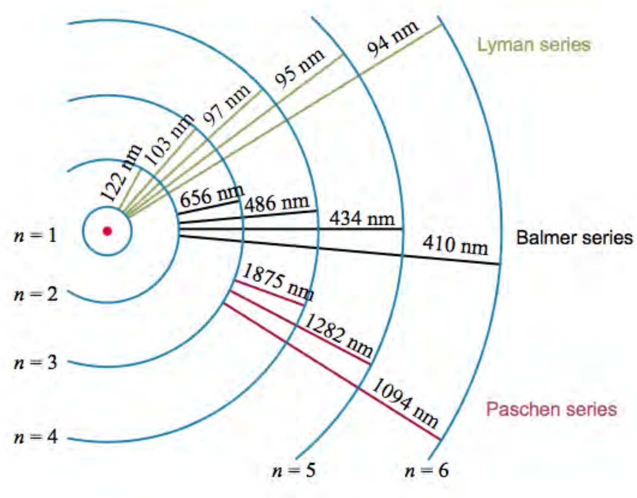


Figure 14: *A sketch of the Hydrogen quantum transitions. The Lyman series are the top green jump from or to the ground state (n=1).*

In wavelength, the Lyman-α forest ranges from the quasar's first Lyman-α (equal to the quasar's Lyman-α emission) to its Lyman-β absorption lines. Figure 15 shows this cut in the top panel (resized in the bottom panel). In this range, there will only be Lyman-α lines each originating from the individual gas clouds along the line of sight. Additional

Lyman-α lines can exist beyond the Lyman-β line, but focusing on the Lyman-α forest simplifies the void counting process significantly. In section 9.1, I sketch how we can use the Lyman-β forest to increase the void count statistics, but that is outside the scope of this project.

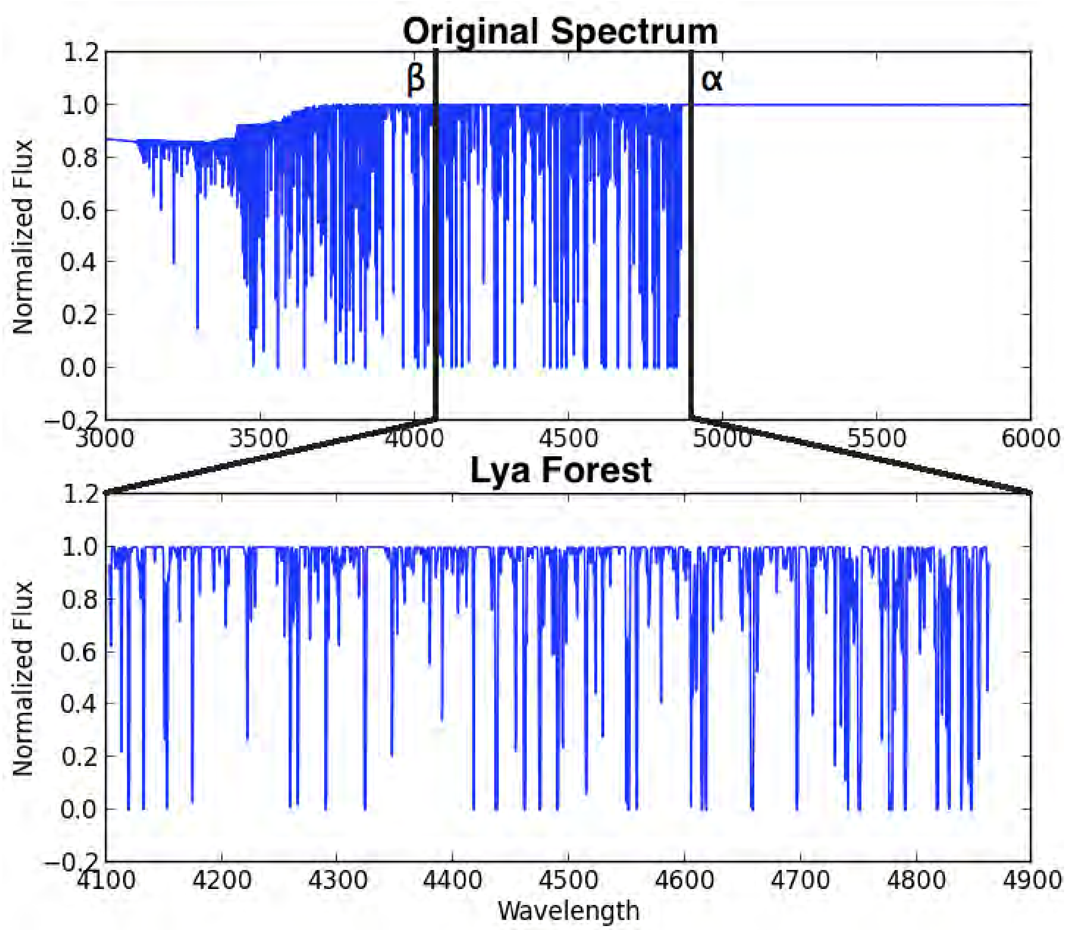


Figure 15: *The top panel is a normalized quasar spectrum with the black lines indicating the range of the Lyman-α forest resized in the bottom panel. The x-axis is in units of Ångstrøm.*

5.2 Flux density threshold

The flux density describes the amount of light that is absorbed at a given wavelength. For a normalized spectrum, the flux value $F = 0$ corresponds to maximum absorption contra $F = 1$ where no absorption occurs. Structures in the LOS have different column densities and show a variety of deep to shallow absorption lines. As a result, it is a matter of choice when deciding to count the number of voids for a certain flux density.

To clarify this further, I have in the upper panel of Figure 16, illustrated how different flux density thresholds, γ, can exhibit very different void counts. Low and high flux density threshold corresponds to respectively looking at large and small densities of structures/absorption lines. Setting the flux density threshold at low values (large densities) show few big voids and oppositely high threshold values (small densities) display many small voids.

The free choice of γ means that no true void count exists and that the void count is tightly connected to a specific density threshold. It resembles the fact that the circumference of an object can increase when the resolution is enlarged to include all the small scale cracks and peaks. It is not trivial to relate the real space density to the flux density, and often simulations are needed (e.g. Viel et al., 2008). A thorough coverage of quasar absorption line systems can be found in Petitjean (1998).

The void density is normally described like, $\delta = \rho \,/ < \rho > - \, 1$. It ranges from $\delta(\rho \to 0) = -1$ all the way to $\delta(\rho \to \infty) = \infty$. In Figure 17, the void density profile from Hamaus et al. (2014) is shown. The range of colours represents how voids, with different sizes, transitions from low to high density. Alas large voids automatically become more empty as their matter evacuates with a quicker pace due to pull from the heavy surrounding structures. We will return to this in section 8.1.

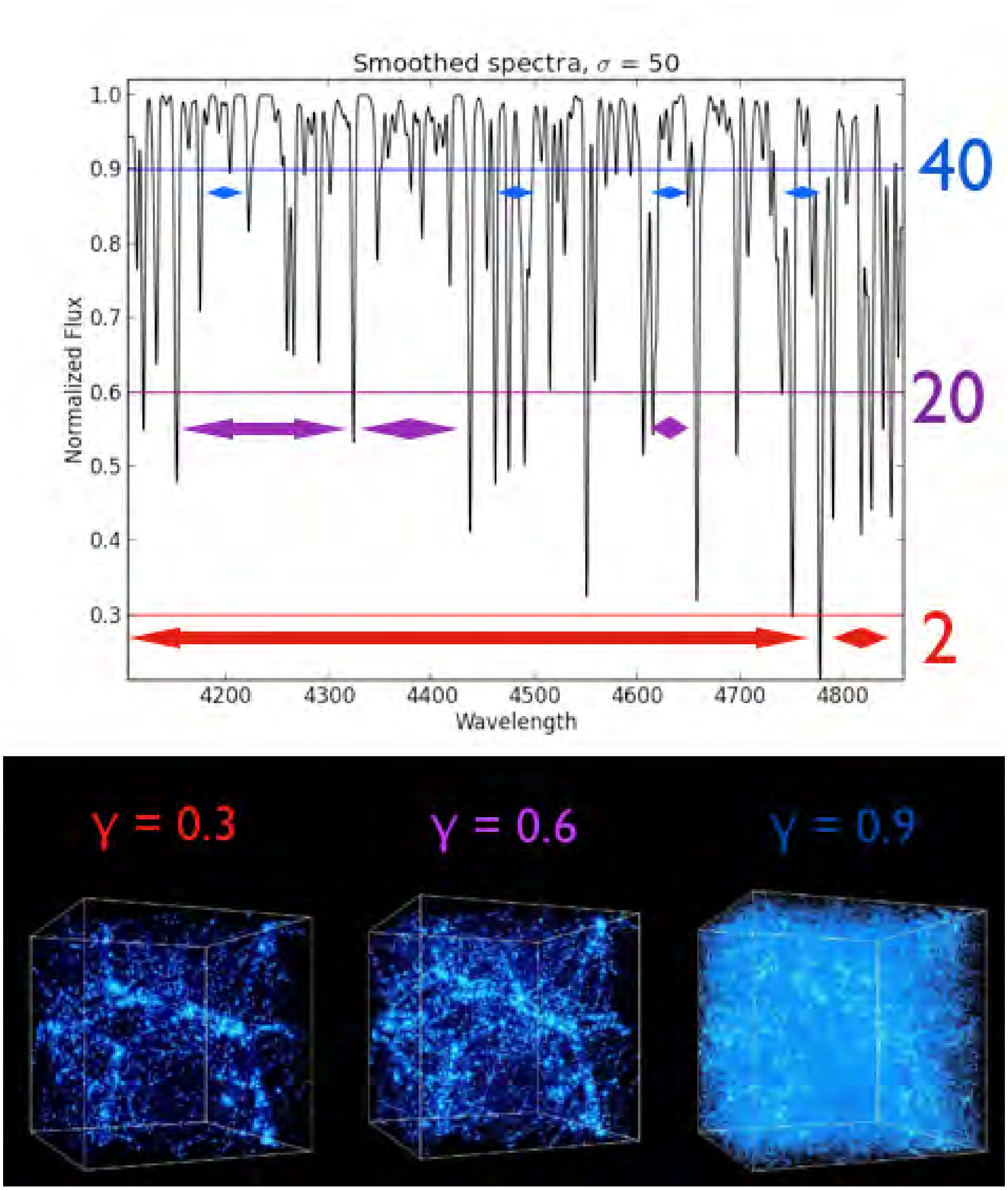


Figure 16: *The upper panel is a simplified spectrum with different flux density thresholds (horizontal lines). The interpretation of each threshold is shown in the lower panel (matching colors). Changing the threshold will affect the definition of a void and thereby the number of voids, as shown on the right side of the spectrum.*

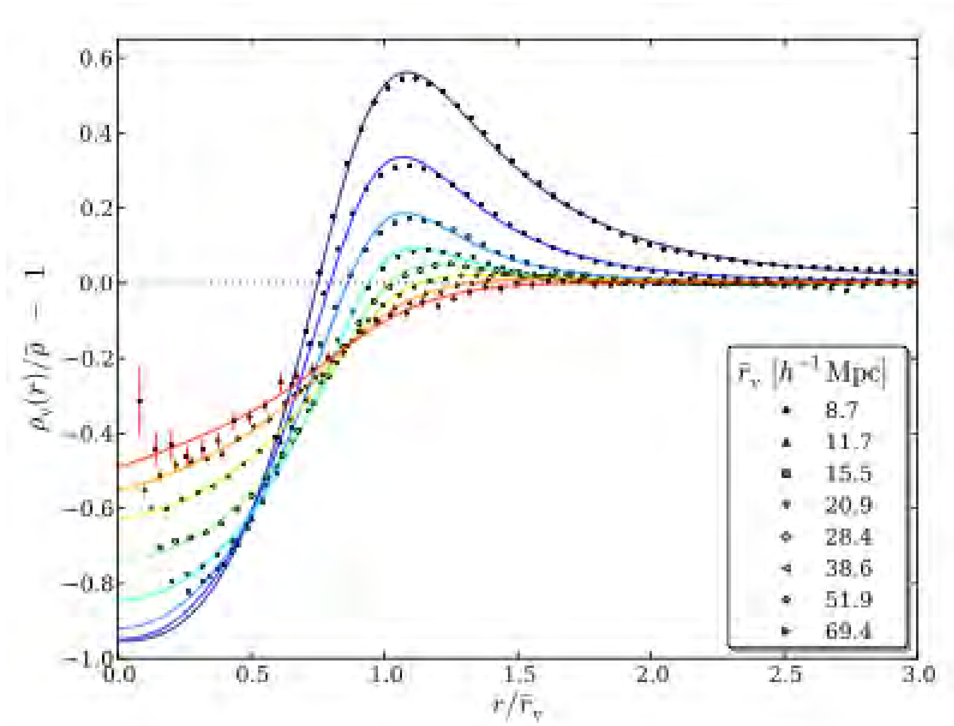


Figure 17: *The void density profile for a range of void sizes (Hamaus et al., 2014).*

5.3 Back-of-the-envelope void size calculation

We have, in the previous sections, introduced how to detect Lyman-α voids in quasar spectra using a specific flux density threshold. We will, in this section, link the void sizes, found in the spectra, with the real space comoving void sizes through a back-of-the-envelope calculation. These equations will be used in later sections to test if we smooth the spectra on the correct scales.

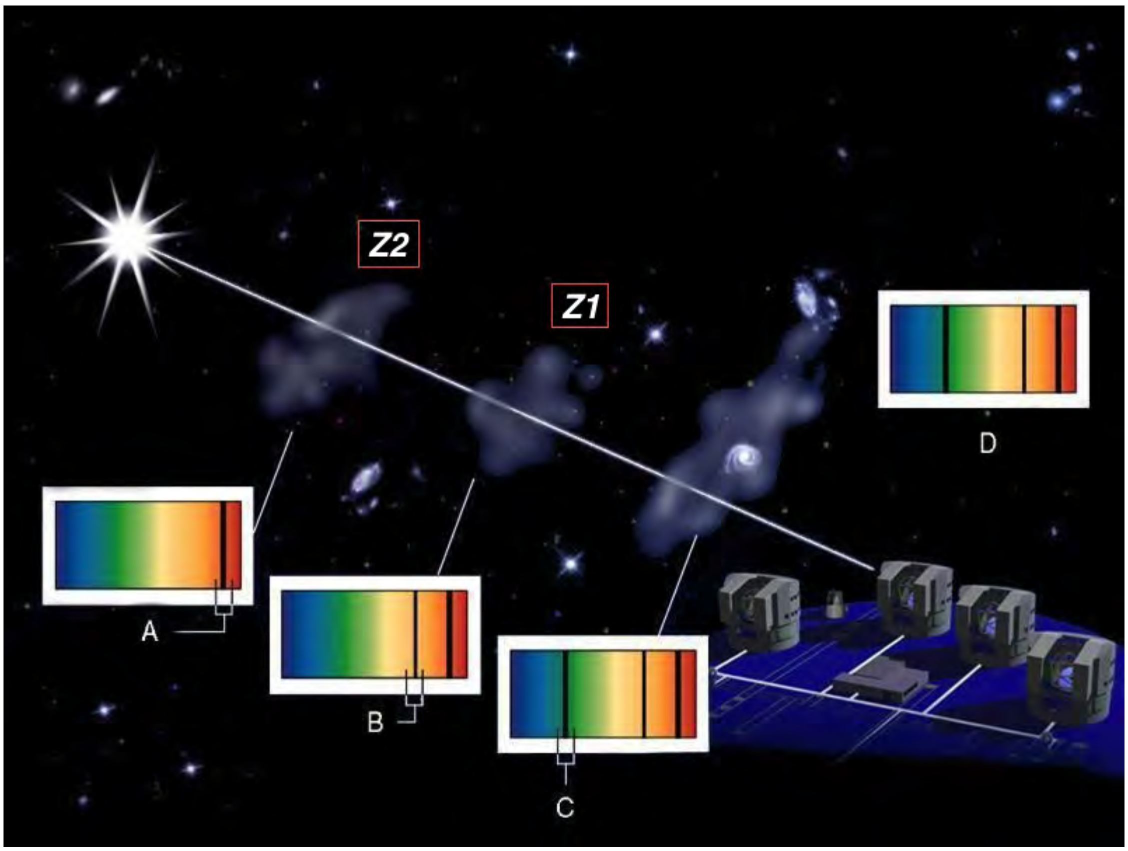


Figure 18: *The setup used to calculate the distance between the galaxies/clouds which are positioned around a void with redshift z_1 and z_2. Credit: Ed Janssen, ESO*

The analytical void size can be found by equating the difference in distance between two galaxies positioned at opposite sides of a void. I have sketched the setup in Figure 18. We use the Robertson-Walker metric to calculate the comoving void size

$$D_\chi = R_0\chi(z) = c\int_{z_1}^{z_2} \frac{dz}{H(z)}, \tag{16}$$

where $H(z)$ is the Hubble parameter. The comoving distance is obtained from eq. (3) following geodesic lines ($ds = 0$) only using χ due to symmetry ($d\psi = 0$). We insert $a = 1/(1 + z)$ into equation 11 which gives us the Hubble parameter as a function of redshift,

$$H(z) = H_0(1 + z)\left[1 + \Omega_M z + \Omega_\Lambda(1/(1 + z)^2 - 1)\right]^{1/2}. \tag{17}$$

We will use the equation in the redshift range $z = 2 - 7$ (see Figure 13) where $\Omega_{rad} \sim \Omega_k \sim 0$. We can now substitute $H(z)$ into equation 16 and construct the comoving void size,

$$R_{void} = D_2 - D_1 = c\left[\int_0^{z_2} \frac{dz}{H(z)} - \int_0^{z_1} \frac{dz}{H(z)}\right] = c\int_{z_1}^{z_2} \frac{dz}{H(z)}. \tag{18}$$

The integral is non-analytic and we solve it numerically leading to the black dashed line in figure 19. Different orders of polynomials (blue, red) are fitted on top of this comoving

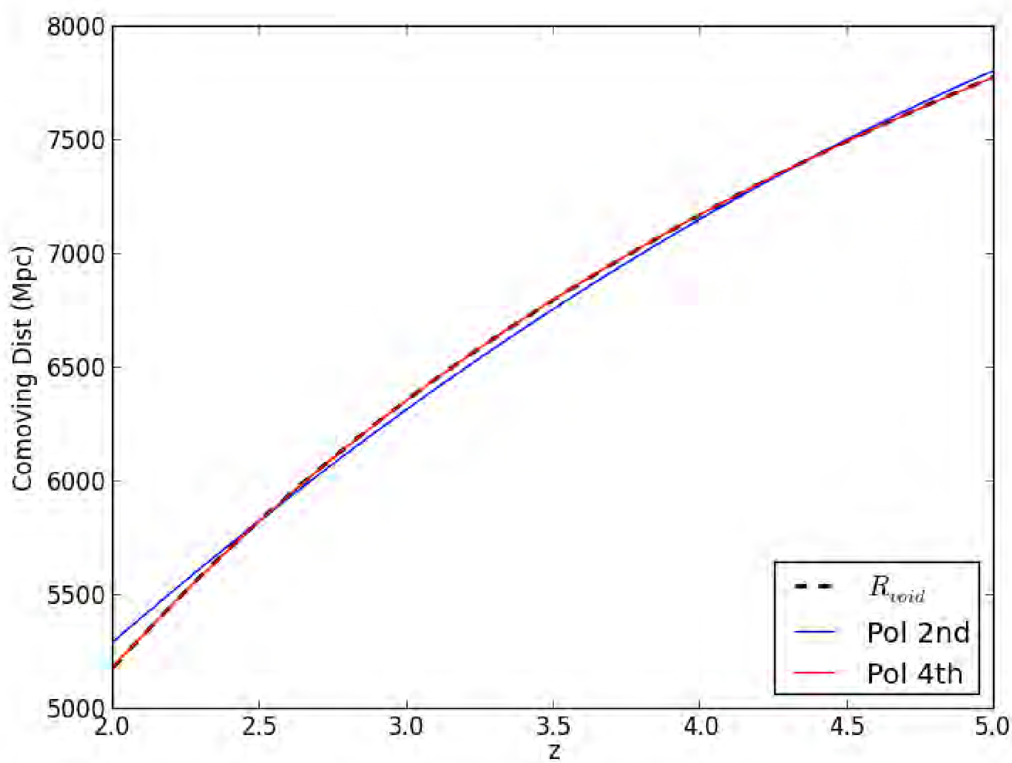


Figure 19: *The normalized comoving void size as a function of redshift (black line). A 2nd (blue) and 4th (red) order polynomial are fitted to the black dashed line and showed with respect to that.*

void size, R_{void} (black). These polynomials are used to create semi-analytic expressions describing R_{void}. I have used $[c] = km\ s^{-1}$ and $[H] = km\ Mpc^{-1}\ s^{-1}$ which give us $[R_{void}] = Mpc$.

Here, I list the best fit polynomials for the 2nd and 4th order:

$$Pol_2(z) = -92z^2 + 1482z + 2702 \tag{19}$$
$$Pol_4(z) = -2z^4 + 55z^3 - 546z^2 + 3007z + 958 \tag{20}$$

In Figure 19, we clearly see that both the 2nd and 4th order polynomial describe the comoving void size well in the range from $z = [2 - 7]$ equivalent to the quasar spectra range.

5.3.1 A Comoving Void Size

I solve the 2nd order equation (19) to obtain an approximative relation between Δz and $\Delta\lambda$. The second order equation is solved in the range covered by our quasar sample, $z = [2 - 7]$. Solving the 4th order equation, which describes the comoving void size, proved to be time consuming and the 2nd order solution was preferred.

The equation is now on the general form $R_{void} = az^2 + bz + c$, where a,b, and c are given in eq. (19)

$$R_{void} = \Big[az^2 + bz + c\Big]_{z_1}^{z_2} = a(z_2^2 - z_1^2) + b(z_2 - z_1). \tag{21}$$

We are interested in relating $\Delta z = z_2 - z_1$ to $\Delta\lambda$ and we start by substituting $z_2 = \Delta z + z_1$ into eq. (21)

$$R_{void} = a((\Delta z + z_1)^2 - z_1^2) + b(\Delta z + z_1 - z_1) \tag{22}$$
$$= a(\Delta z^2 + 2\Delta z z_1) + b\Delta z = \Delta z^2 a + \Delta z(2az_1 + b). \tag{23}$$

We assume that R_{void} is roughly constant over the range $z = [2 - 7]$, and call it C_{void}. We get the equation,

$$0 = \Delta z^2 a + \Delta z(2az_1 + b) - C_{void}. \tag{24}$$

We can solve Δz as a function of z_1. This places the redshift of the void at z_1, contra using the middle of the void, but we neglect this effect. We can write the solution to equation (24) like,

$$\boxed{\Delta z = \frac{-(2az_1 + b) \pm \sqrt{(2az_1 + b)^2 + 4aC_{void}}}{2a}}\ . \tag{25}$$

This equation relates the void size, C_v, in redshift size, Δz, as a function of redshift, z_1, see Figure 20. In order to get a feel for the real space (Mpc/h) void sizes we find in

Pan et al. (2012), that they range from $1-100$ Mpc/h. We continue the analysis with $C_v = 1, 10$ and 100 Mpc/h. The upper limit on 100 Mpc/h is chosen in accordance with Einasto et al. (1989), but we will focus on the choice of 1 Mpc/h in section 6.4.

The solution for both the positive and negative Δz is plotted in the subplot of Figure 20. We expect a rise in Δz for increasing z_1 value, and to that extend the red ($\Delta z(+)$) slope is clearly unphysical showing the opposite behavior. I have, in the same figure, made a zoom on the blue ($\Delta z(-)$) slope (large frame) that shows a shallow increase as a function of redshift.

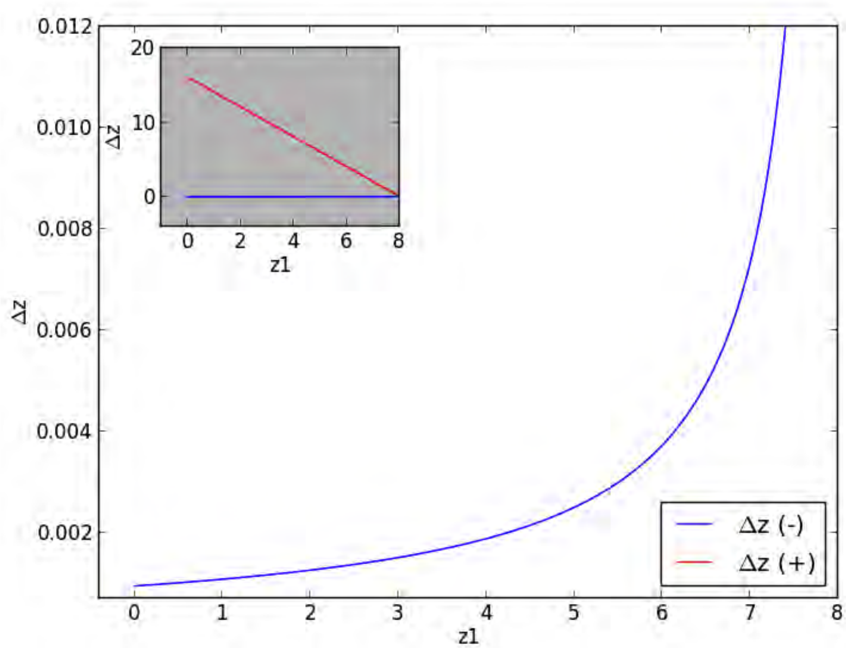


Figure 20: *The solutions to the 2nd order polynomial is represented by the positive (red) and negative (blue) solutions (grey box). The large frame is a zoom of the positive slope (blue, negative solution) that is our preferred solution showing a very shallow increase with redshift.*

5.3.2 Relating Δz and $\Delta\lambda$

We now have a relation between Δz and z_1 (the blue line in Figure 20) that we wish to relate to the wavelength width, $\Delta\lambda$ in our quasar spectra. The redshift, $z = \lambda_{obs}/\lambda_{em} - 1$ is substituted in the equation below to find the void width.

$$\begin{aligned}
\Delta z &= z_2 - z_1 = \left(\frac{\lambda_{obs,2}}{\lambda_{em,2}} - 1\right) - \left(\frac{\lambda_{obs,1}}{\lambda_{em,1}} - 1\right) && (26)\\
&= \frac{\lambda_{obs,2}}{\lambda_{em,2}} - \frac{\lambda_{obs,1}}{\lambda_{em,1}} && (27)\\
&= \lambda_{Ly\alpha}^{-1}(\lambda_{obs,2} - \lambda_{obs,1}) && (28)\\
\Rightarrow \Delta z &= \lambda_{Ly\alpha}^{-1}\Delta\lambda. && (29)
\end{aligned}$$

In eq. (28), I used the fact that we look at Lyman-α absorption lines to substitute $\lambda_{em,2} = \lambda_{em,1} = \lambda_{Ly\alpha}$. This provides us with a final equation based on the 2nd order polynomial

$$\boxed{\Delta\lambda(z_1) = \lambda_{Ly\alpha}\left[\frac{-(2az_1 + b) - \sqrt{(2az_1 + b)^2 + 4aC_{void}}}{2a}\right]}. \qquad (30)$$

Here $a = -92$, $b = 1482$, and $C_{void} = 1\ Mpc\ h^{-1} = 1.4\ Mpc$. In Figure 21, I have plotted the wavelength range, $\Delta\lambda$, and redshift range, Δz, as a function of redshift, z. The time between two emitted photons increase as a function of redshift resulting in both higher Δz and $\Delta\lambda$ with z.

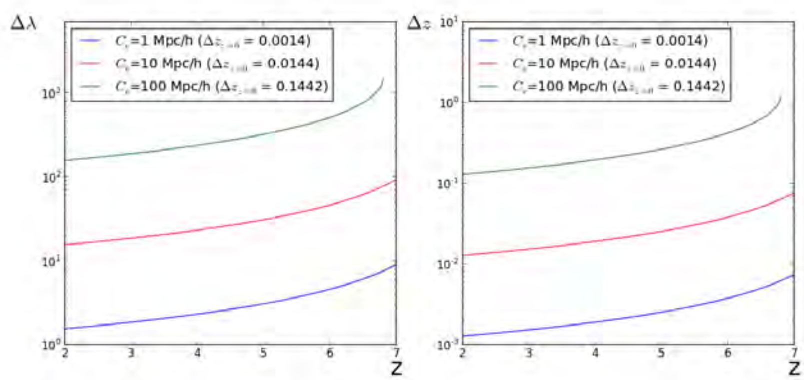


Figure 21: *The wavelength width in quasar spectra (Å), $\Delta\lambda$, as a function of redshift, z_1 (left) for a minimum void size of respectively 1 (blue), 10 (red) and 100 $Mpch^{-1}$ (green). We see that increasing the void size also increases $\Delta\lambda$ as expected. The redshift values in the legend refer to the redshift size of each of the chosen void sizes at $z = 0$. The right plot is similar but is shown with Δz instead.*

6 Improving the Void Count

Real quasar spectra are subjected to uncertainties in the form of noise, emission shape normalization and gravitational clumping of gas clouds leading to absorption line blending. In the following sections, we test the physical effects in order to quantify how much the uncertainties affect the void count.

6.1 Quasar spectra

Normalized quasar spectra are produced through an automated pipeline reduction. Here, instrumental noise (e.g. broken CCD pixels) can be reduced to artificially high flux values. Artificial lines give rise to a systematic error in the void count and, below (section 6.1.1), I will address how we treat these data-less pixels. When the artificial flux values are removed, the spectra still exhibit gaussian noise from the natural uncertainty in the detector. The noise in the spectra will add an unreal number of voids, and we prefer to remove the noise using a Fourier smoothing technique (see section. 6.1.2 and 6.1.3). We create a small sample of 50 real quasar spectra and, from these, define a set of selection criteria based on the void count that is the least sensitive to the above effects. The final void counting code will be fully automatic thus the criteria must be designed to work without human intervention.

6.1.1 Data-less pixel removal

Pipeline reduction problems result in extremely large flux (F) and error (σ_{Flux}) pixel values of $\sim 10^{30}$. In order to remove the artificial pixel values, I perform two cutting techniques using respectively the error and flux values. The quasar spectra are usually normalized to have $F = [0;1]$, and the cutting is based on this range. The first technique removes pixel values using the errors: $(0 - 5 \cdot \sigma_{flux}) < F < (1 + 5 \cdot \sigma_{flux})$. This removes lines with artificially large errors. On top of this, I do a flux cut: $-5 < Flux < 5$, which removes the large artificial flux values. In Figure 22, I have plotted an example of the error (left column) and flux (right column) spectra before (top) and after (bottom) a data-less pixel removal. In the right column of the figure is an example of how the data removal slopes the flat regions. This token will be flattened out and the artificial extra void count removed when we in the following section smooth the spectra.

When removing the data-less pixels, we construct artificial empty regions that resemble voids. We take this effect into account when using the prime sample (section 4.3.3) in the analysis of section 8.

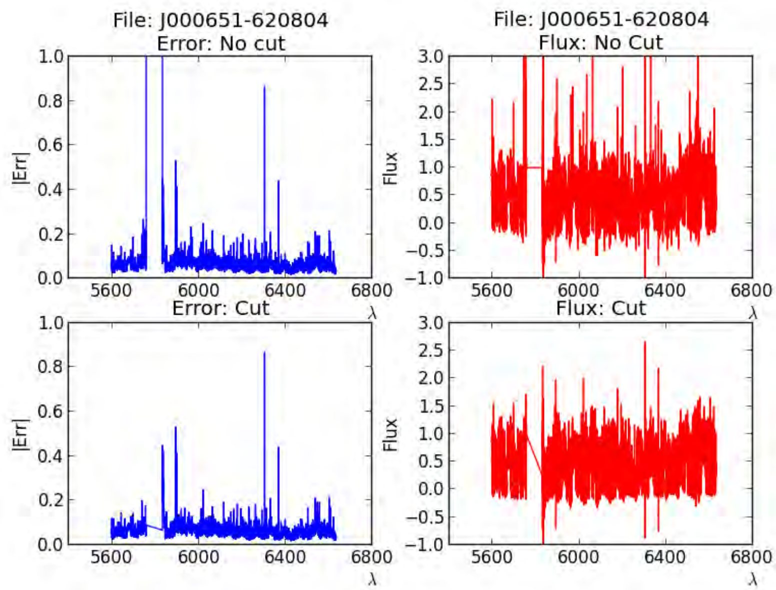


Figure 22: *An example of the normalized flux errors and flux spectrum (left and right) shown before and after (top and bottom) using the cutting method to remove the unrealistic errors (left side). Quasar spectra: J000651-620803.*

6.1.2 Fourier Smoothing

The level of noise (amount of artificial lines) affects the voids count in the spectra. Smoothing the spectra to remove the noise and only counting the voids at cosmological scales are therefore crucial. We choose to use a Fourier method that smoothes the spectra according to the amount of noise it exhibits. In section 6.4 it is shown that the noise-lines are narrower than real voids (> 1 Mpc$/h$) and is smoothed by the Fourier technique. A frequency power spectrum can be constructed from Fourier transforming a normalized flux spectrum. High frequencies resemble the small scale fluctuations and can be removed. The initial spectrum, $f(x)$ is Fourier transformed using the equation

$$F(p) = \int_{-\infty}^{\infty} f(x) e^{-ipx} dx. \tag{31}$$

The low frequency signals can be inverse Fourier transformed with

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(p) e^{ipx} dp \tag{32}$$

returning a smoothed spectrum. This process is demonstrated in Figure 23, that shows each of the three steps (left to right) for three different degrees of smoothing (top to bottom) of a mock spectrum. The first column is the un-smoothed spectra, the second the frequency space and the last represents the smoothed spectra. The blue color in column two shows the part of the frequency space that is reverse Fourier transformed into the smoothed version in column three. Low cut values result in a high amount of smoothing. The top row, with a cut value of 1750, is the strongest smoothed spectrum.

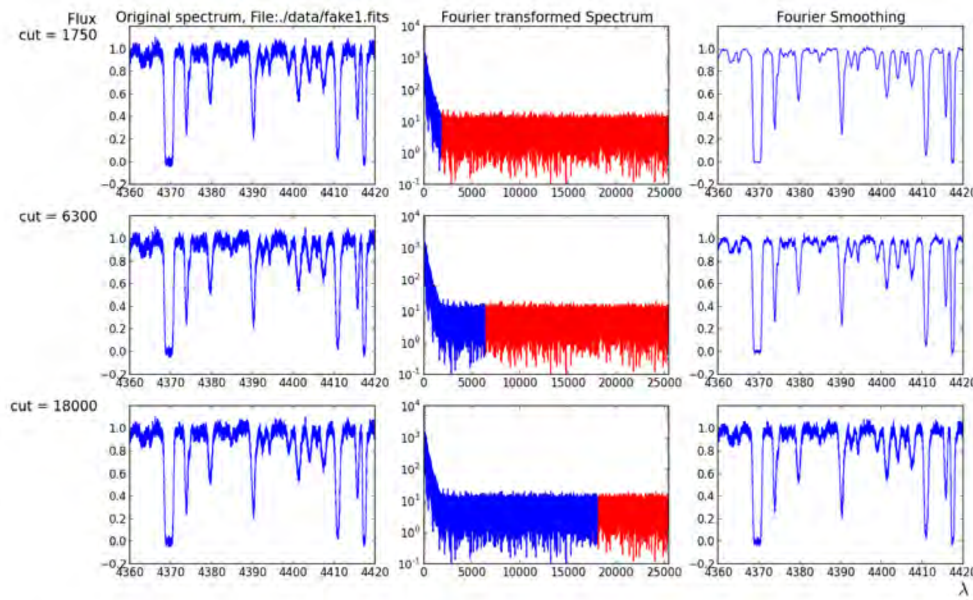


Figure 23: *The plot shows three degrees of smoothing (rows). The first column is the initial spectra before smoothing, then the frequency space and last the smoothed spectra. The smoothing is determined by the frequency cutting value, and low cuts represent high amount of smoothing.*

Smoothing sensitivity

I test the void count sensitivity to Fourier smoothing by counting the number of voids in a mock spectrum as a function of threshold, γ, for a range of smoothing scales. Figure 24 shows the void count as a function of threshold for fake1 (artificial mock spectrum) at a range of Fourier smoothings, Cut(Hz) $= [0, 1750, 6300, 18000]$. On top of this is added the void count of perfect1 (a no-noise version of fake1) showing the optimal smoothing. Comparing perfect1 with the non-smoothed fake1 (Cut $= 0$) clearly shows that smoothing the spectrum is crucial to remove the noise that otherwise would be counted as voids. The line density increases when $\gamma \to 1$ and, as a result, the void count also increases. The rise in void count, for small threshold values in Figure 24, appears due to the noise at the bottom of saturated lines. The Fourier method smoothes lines below a certain size and is in accordance with Figure 24 showing the number of voids dropping proportional to the smoothing scale. The void count changes significantly at low and high flux density threshold, implying that these regions should be avoided when selecting our threshold criteria.

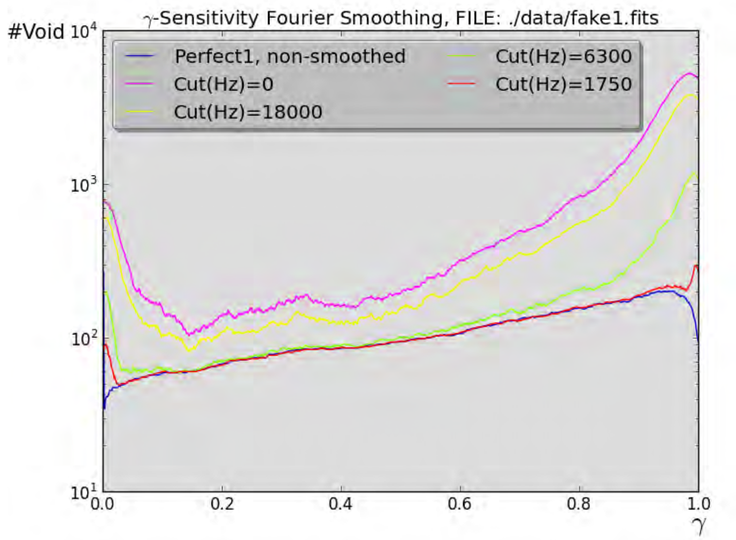


Figure 24: *The void count as a function of threshold for the perfect1 and fake1 mock spectra. The fake1 spectrum has been subjected to a range of Fourier smoothing scales.*

Automatizing the Fourier Smoothing

In Figure 24 the void count of fake1 resembles perfect1 when a smoothing of $\sim 1750\ Hz$ is used. This cut occurs at the bend in the Fourier frequency spectrum and suggests a noise dependent smoothing scale (see Figure 23). Based on this trend, we develop an automatic smoothing technique that we, in section 6.1.3, will test on real quasar spectra.

We suspect that the steep (low frequency) and shallow (high frequency) slopes arise from two kinds of frequencies: noise and cloud absorption lines. If this is correct, cutting the shallow slope (high frequencies) away should smooth the spectra by only removing the noise. To choose this cutting value consistently, we fitted lines to each of the slopes and chose the crossing point to be our frequency cutting value. The method is shown in Figure 25 where the crossing between the magenta and green lines sets the limit of what is reverse Fourier transformed into the smoothed spectrum.

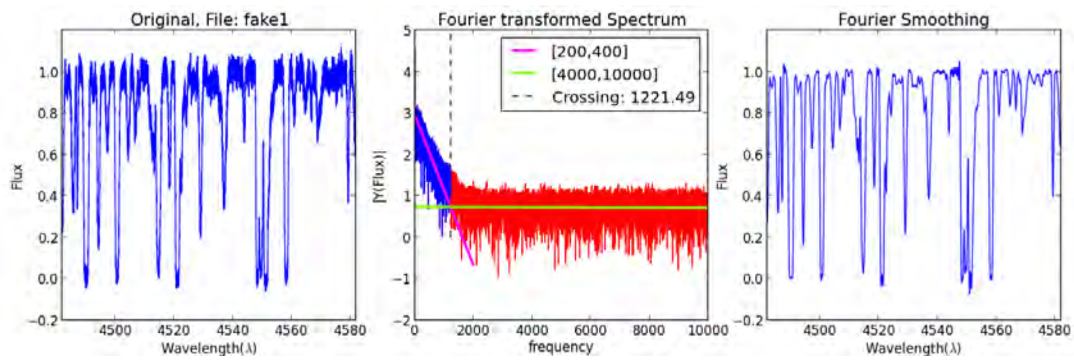


Figure 25: *The automatic Fourier smoothing applied to the mock spectra fake1. The frequency space shows a varying slope trend fitted by a magenta (slope) and green (flat) line to determine the smoothing cut value. In this case the ranges was found using a pixel dependent function to be respectively* $[200, 400]$ *and* $[4000, 10000]$*. In Figure 49 in Appendix C we show how we calibrated the fitting ranges to estimate the best fitting function.*

6.1.3 50 QSO subsample

The noise in the mock spectra is purely Gaussian, which is not necessarily the case for real observational noise that may have several sources. Consequently, testing if the frequency bend is present in real quasar spectra, we have decided to create a 50 QSO subsample. From this sample, we will determine our robust void count selection criteria including a flux density threshold.

Fourier method

To produce a consistent smoothing scale, we construct a pixel dependent function that fits the slope (magenta) and flat (green) part of the frequency spectrum (see Figure 25). We use the trends from three randomly chosen quasar spectra ($J193957-100241, J000651-620803, J120523-074232$) in the 50 subsample to fix the inner values to $[200, 400]$. The outer values ($[\mathrm{Npix}/5, \mathrm{Npix}/2.5]$) are determined from number of pixels (Npix) as the quantity of removed data-less pixels changes the frequency range of the spectrum. In Figure 49 in Appendix C are examples of how the ranges was altered.

One aim of the Fourier method was to develop a noise dependent smoothing scale. This is tested in Figure 26 where we clearly see that the more noisy quasar $J000651-620803$ is smoothed with a lower cut value (larger smoothing).

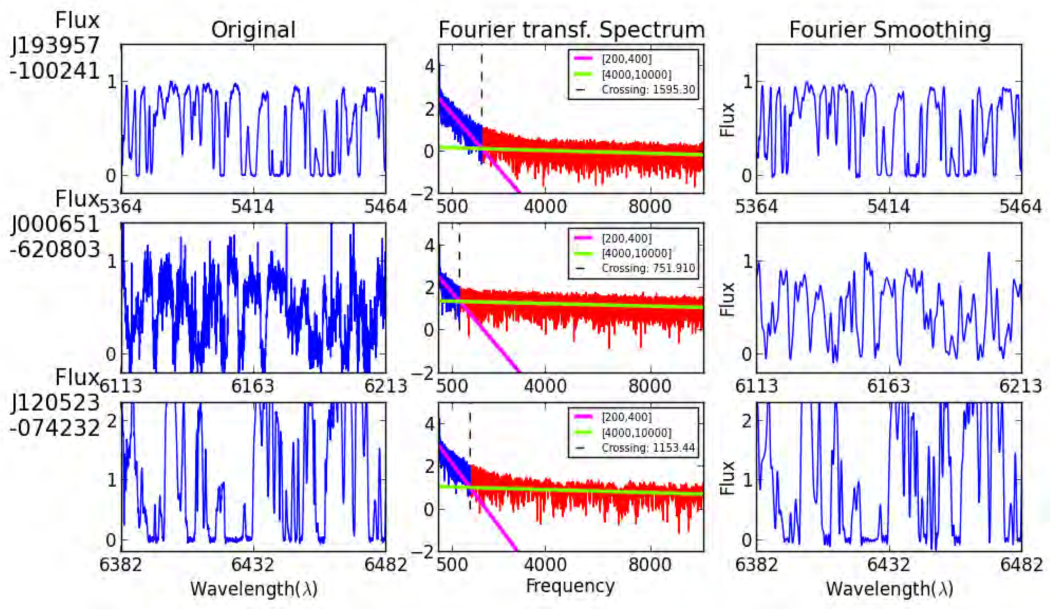


Figure 26: *The automatic Fourier smoothing method applied to the three noisy quasar spectra; J193957-100241, J000651-620803, J120523-074232. The first column is the un-smoothed spectra, the second the frequency space and last the smoothed spectra. The middle row is far more noisy than the top and bottom rows, but is smoothed accordingly.*

We have constructed a method that appears to smooth, according to the noise level, and we will now test its robustness on our new 50 quasar sample. The 50 sample reduces to 39 when removing spectra with artificial frequency spectra (see section 4.3.1 + 4.3.2 for quasar sample handling). I reproduce the frequency plots of each of the 39 quasar spectra (Figure 50 in Appendix C) and checked visually that the frequency spectra, fitting ranges, and smoothing scales look reasonable. Based on this, we conclude that our Fourier method smoothes according to the noise level and also proves to be robust on larger samples.

A stable flux density threshold

We discussed in section 6.1.2, under the paragraph **Smoothing Sensitivity**, that the void count changes significantly at low and high values of the flux density threshold (see Figure 24). Naturally, we avoid these regions but also wish to maximize our void count to improve the statistics. In Figure 27, I have included a single plot from Figure 51, in Appendix C, that shows the void count sensitivity and void count as a function of threshold, $\gamma = [0.5 - 0.7]$. The original figure, from Appendix C, includes the larger range $\gamma = [0.4 - 0.6], [0.5 - 0.7], [0.6 - 0.8]$. The left plot estimates the relative void count change between the minimum and maximum count in the respective threshold range. The right plot resembles Figure 24 showing the void count against threshold (limited range). Both the left and right plot are color coded with respect to the S/N value (see section 4.3.3 for definition). Our best threshold estimate is determined by the competing effects of a stable void count and the wish of improving the statistics through maximizing the void count to be $\gamma = 0.6$ with 0.5 and 0.7 as secondary candidates.

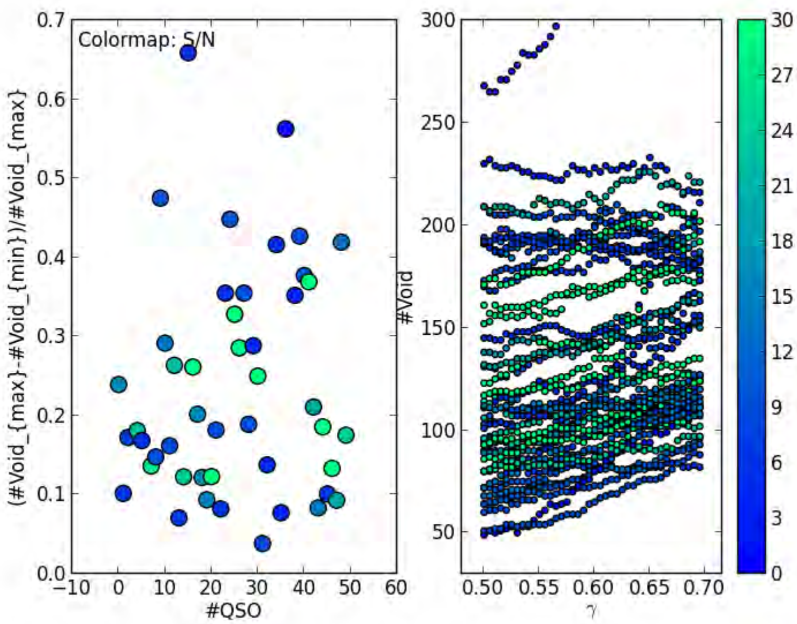


Figure 27: *The void count sensitivity (left) and the respective void count as a function of γ of the resulting 39 quasar (from the 50 sample) is plotted for $\gamma = [0.5 - 0.7]$. The void count sensitivity is constructed by measuring the relative difference of the void count, $(max(\#void) - min(\#void)) \cdot max(\#void)^{-1}$.*

6.2 Mock spectra

The matter distribution in the Universe is governed by gravity and, consequently, the matter tends to clump together resulting in a non-uniform distribution of clouds along the line of sight. The absorption lines, in the spectra, will follow the same trend and we are interested in how this affects both the Fourier method as well as the void count. To quantify both effects, we have made a range of clumpy mock spectra from the same mock generation code as used in Liske et al. (2008). We will focus the tests on the Fourier cut and the void count's sensitivity to line blending from gravitational clumping (clumpy1-5).

6.2.1 Gravitational Clumping

The effect of gravitational clumping can be seen in Figure 28, where the number of absorption lines from $z = 4.0 \rightarrow 2.2$ drops drastically. In Figure 29 we test if the gravitational clumping have an effect on the Fourier method. The 5 artificial and a real quasar spectra (top to bottom) are shown with each the original (non-smoothed) frequency and Fourier smoothed spectra (left to right).
Clumpy5 is the spectrum with the highest clumping factor, and clumpy1 the lowest. The quasar spectrum shown is J193957-100241, which we also used in Figure 26. The smoothing scale appears the same when the clumping factor is increased between the individual mock spectra as well as the included quasar spectrum. One thing to notice is that the Fourier method still appears to be smoothing accordingly to the amount of noise in the spectra, which is also why the smoothing scale vary slightly.

The gravitational clumping affects the void count through collapse of smaller structures (many lines) into bigger (fewer lines). The question is how quickly does the void count's change under the influence of gravitational clumping? In order to test this, we count the number of voids as a function of threshold for the same range of clumpy spectra (1-5). The result can be seen in Figure 30, similar to Figure 24, where the void count against flux density threshold is plotted for each of the clumpy mock spectra.The void count drops drastically when a clumping equal to clumpy 5 is present in the spectrum, whereas clumpy 1-4 exhibit similar void count's behaviour. I have chosen two quasar spectra with respectively $z = 2.2, 4.0$ and plotted their flux spectrum, in Figure 28. Here, $z = 2.2$ shows the largest clumping and resembles the clumpy 2 or 3, suggesting that 4 or 5 might be an unrealistic amount of clumping for our quasar range.

The consequence of line blending is a natural effect of gravitational clumping that leads to the void count drop for the clumpy5 spectrum. We have compared the number of voids in the fake1 mock file (known exactly) with the number of voids counted using our code, to show the blending effect visually. A random part of the spectrum (blue) is plotted in Figure 31 where we have over-plotted the lines from the mock file (black) and the lines

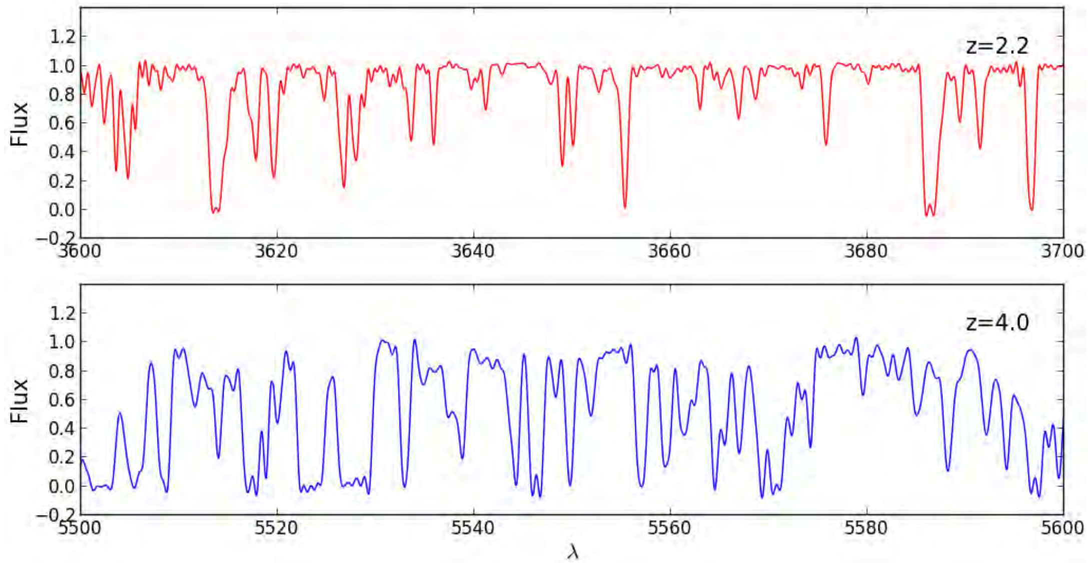


Figure 28: *Quasar spectra J000344-232354 (top) and J024401-013403 (bottom) with respectively $z = 2.2, 4.0$. The fewer lines in J000344-232354 ($z = 2.2$) compared to J024401-013403 ($z = 4.0$) clearly shows the effect of gravitational clumping of matter in the Universe.*

counted by the code (red). The effect of blending lines is clear at e.g. $\lambda = 4140$, where the three black lines blend and are counted as one red line. This results in fewer number of red lines compared to the known number of black lines (mock). This happens as lines that are close blend due to their intrinsic width and appear as one line lowering the count of lines in the spectra. The number of red lines in the spectra is also dependent on the void density threshold which is arbitrary set to $\gamma = 0.94$, and represented by the green dashed line.

The void count of clumpy 5 was significant lower than clumpy 1-4 (Figure 30) and we conclude that not only was the clumpy 5 an unrealistic amount of clumping but also that the void count is not sensitive to the realistic effect of line blending. When we later compare quasar and simulated spectra, we will count the voids with the exact same method, so a robust counting method based on our selection criteria is preferred to obtain a precise void count. The test shows that we understand the relation between threshold, blending of lines and the counted number of absorption lines in the spectra.

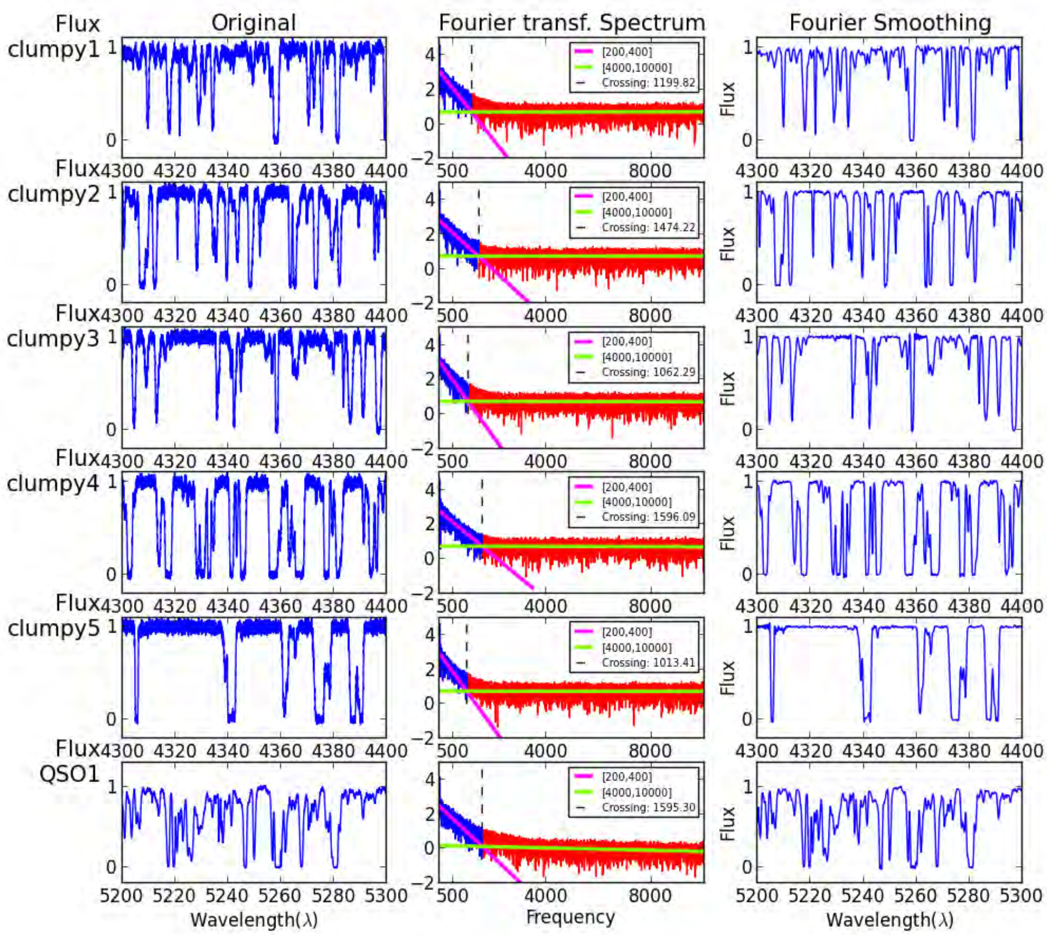


Figure 29: *The Fourier smoothing method is shown for the five clumpy spectra as well as a quasar spectrum (bottom). The first column shows the original un-smoothed spectra, then the frequency space and last the smoothed spectra. The top five rows are clumpy 1 to 5, and in the bottom we have included a real quasar spectra for comparison.*

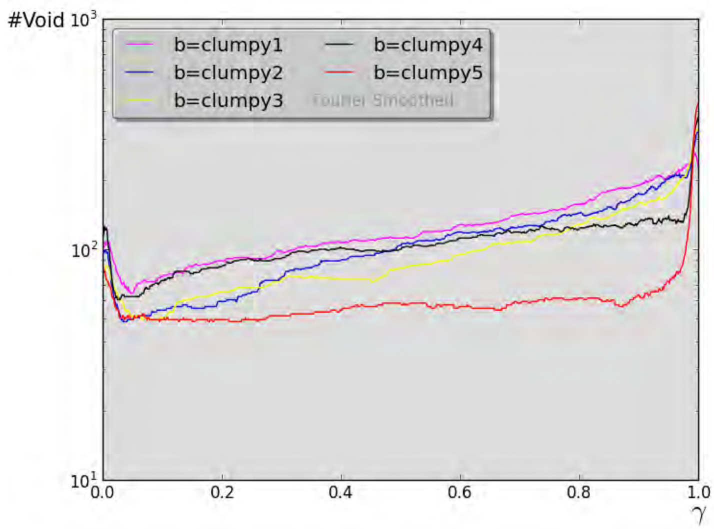


Figure 30: *The void count as a function of threshold, γ, for a range of clumpy spectra. The clumpyness ranges from clumpy1 (low) to clumpy5 (high). The spectra have been automatically Fourier smoothed as described in section 6.1.2.*

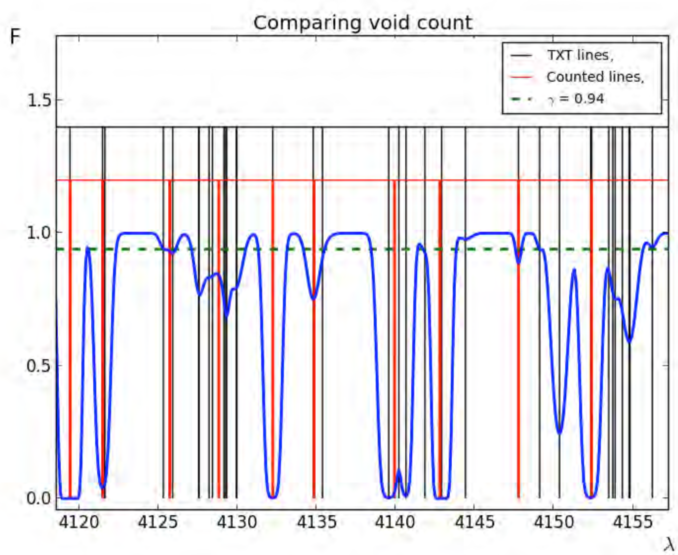


Figure 31: *An example of a mock spectrum (blue) where the underlying absorption lines are marked with black vertical lines. The green dashed line is an arbitrary choice of threshold for which the red vertical lines mark the counted lines. It is clear that where the black lines lie close/blend we only count one (red) which leads to a lower count.*

6.3 Simulated spectra

In the simulated spectra we expect no detector noise, bad pixels or pipeline reduction errors. This means that methods like data-less pixel removal and Fourier smoothing, normally used to improve the quasar spectra, will not be applied to the simulated spectra. We smooth the simulated flux spectra on a scale of 1 Mpc/h, in accordance with Villaescusa-Navarro et al. (2011), to remove substructure below the Jeans length like numerical resolution and astrophysical processes (e.g. feedback from galactic winds). I developed a gaussian smoothing technique (see section 6.3.1) that allows us to set the smoothing scale in $[\sigma] =$ Å. Figure 21 (left), in section 5.3.2, shows the result of equation 30, that is a relation between the wavelength width (Å) and redshift (z) for a comoving void size of $C_v = 1$ Mpc/h. We use this equation to determine the wavelength width of 1 Mpc/h that for different redshifts becomes $\sigma(z = [2.2,\ 4.0]) = 1.58,\ 2.28$ Å.

When running simulations it is customary to take snap shots of your simulation box at specific redshift intervals. In our case we have a box at $z = 2.2$ and 4.0 with RLOS extracting the column density from which the simulated flux spectra can be constructed. When you look into the Universe the spectra show absorption lines at different redshifts, whereas the absorption lines in a simulated spectra all occurs at the same redshift. The comparison between simulated (redshift space) and quasar spectra (real space) does not affect the number count of voids when you compare spectra slices of 100 Mpc/h. Although, if looking at the void sizes one should treat this bias with care. A good example of how to circumvent the redshift space bias and produce more realistic simulated spectra is shown in Viel et al. (2013) section *IV. The Mock QSO Sample.*

6.3.1 Gaussian Smoothing

The Gaussian smoothing method convolves the spectrum with a gaussian weighted function, which can be expressed as,

$$f(x_0) = \frac{\sum_i f(x_i) g(x_i)}{\sum_i g(x_i)}, \tag{33}$$

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}. \tag{34}$$

Here $g(x)$ is the gaussian function with mean μ and width σ, $f(x)$ is the function we wish to smooth (in our case the simulated flux spectrum), and $f(x_0)$ being the smoothed function value at x_0. The σ-parameter in g(x) determines the smoothing scale. In Figure 32, I have plotted the two simulated spectra before (full line) and after (dashed line) a Gauss smoothing of $\sigma = 1.58,\ 2.28$ Å for respectively $z = 2.2$ (upper) and 4.0 (lower). The different smoothing scales arise from the redshift stretching effect that that occurs as a result of the expanding space as well as the inertial frame transformation. The figure shows different amounts of gravitational clumping, where the evolution of the dense ($F \sim 0$) areas merge together over time, from $z = 4.0$ (lower blue) $\rightarrow$ 2.2 (upper red), creating more empty areas, or under-dense voids ($F \sim 1$) over time. The smoothing scale do not remove physical structures in no noise simulated spectra, demonstrating that our Fourier technique is calibrated correctly (more about this in section 6.4)

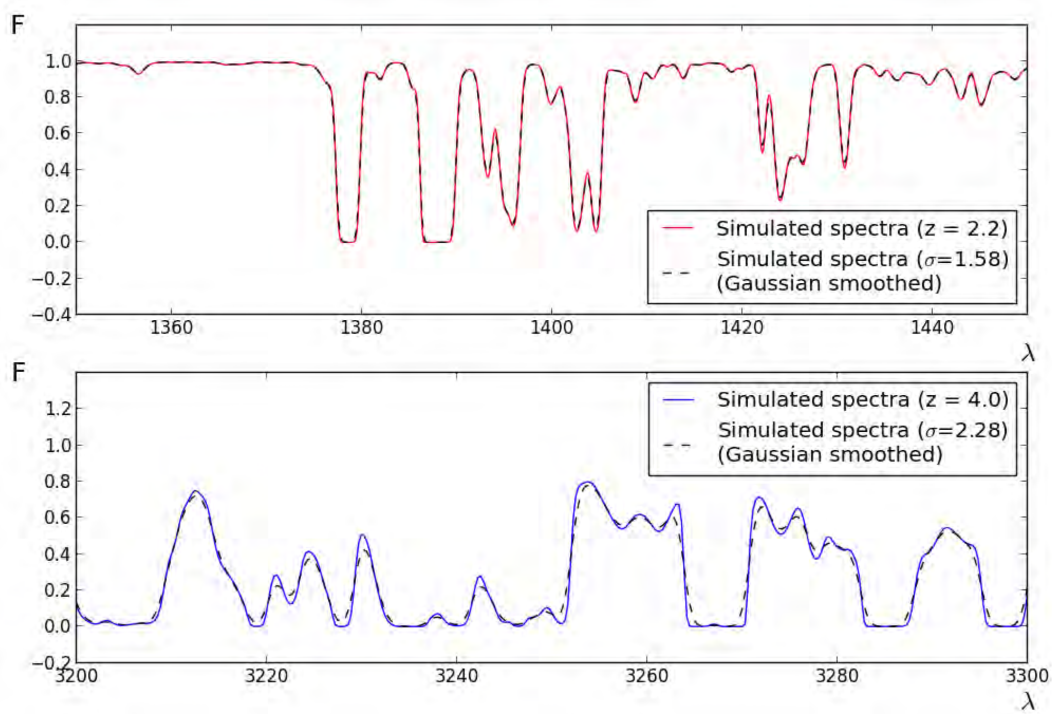


Figure 32: *Two simulated spectra plotted before (full line) and after (dashed line) the Gaussian smoothing of* 1 *Mpc/h for* $z = 2.2$ *(upper) and* 4.0 *(lower). The effect of gravitational clumping can be seen evolving from* $z = 4.0 \rightarrow 2.2$ *showing less but denser structures at lower redshift.*

6.4 Comparison of smoothing scales

It is possible to test if our Fourier method smoothes the spectra on the correct scales and do not accidentally remove or add real voids. I have, in Figure 33 plotted the void count between a perfect mock spectrum (no noise) and a fake mock spectrum (noise added) which is Gaussian smoothed with respectively $\sigma = 0, 3$ as well as a Fourier smoothing. We see that the fake mock spectra need to be smoothed with $\sigma = 3$ or the Fourier method in order to remove enough noise to resemble the perfect spectrum. This suggests that the Fourier method is calibrated correctly. To support the above claim, I have in Figure

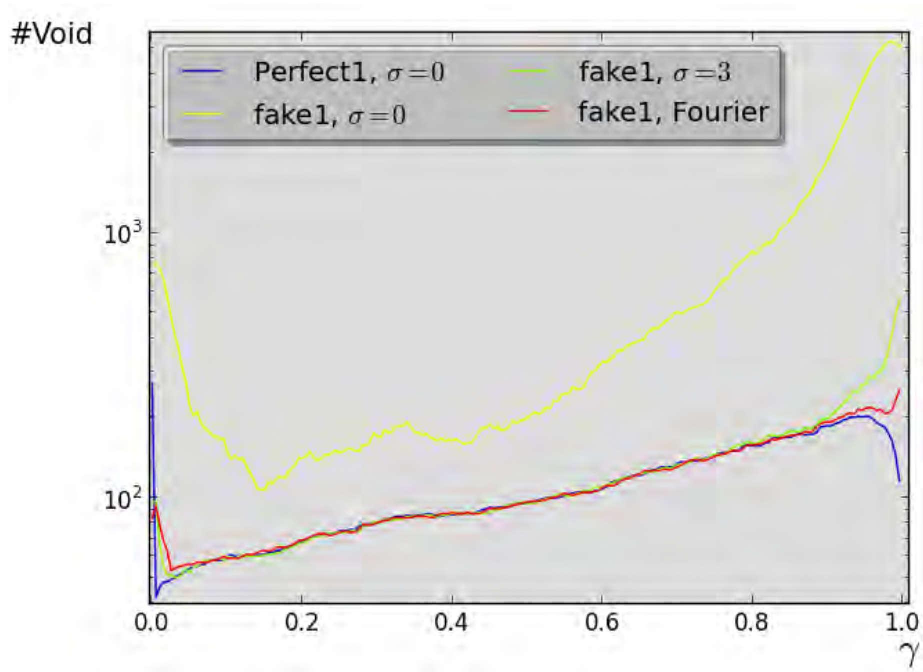


Figure 33: *The void count of perfect (no noise) mock spectrum (blue) compared to the fake (noise added) mock spectra smoothed by respectively a Gaussian smoothing with $\sigma = 0$ (yellow), 3 (green) and the automatic Fourier method. We clearly see that the void count of the yellow line ($\sigma = 0$) is very affected by the noise in the spectra. The green line ($\sigma = 3$) and the red line (Fourier smoothing) have, on the contrary, been smoothed accordingly and resembles the no noise blue line.*

34 plotted the void count as a function of threshold (γ) for each of the spectra in the 50 quasar subsample. The red lines are Gaussian smoothed with $\sigma = 1$ (upper) and 3 (lower) and the blue line represents the automatic Fourier smoothing for each of the 50 quasar subsample mentioned in section 6.1.3. I have marked the spectra exhibiting $\Delta F_{max} > 20$ (see section 4.3.2) with black circles.

I have chosen the Gaussian smoothing scale of 1 Mpc/h from Figure 21 (left) in the quasar redshift range $z = 2 - 4$ (see Figure 13) that results in a scale of $\sigma = 1 - 3$ (see also 6.3 for comparison with the simulated smoothing scale). The figures without circles all have a Fourier smoothing scale in the range of 1 Mpc/h (grey area), comparable to the simulated smoothing in section 6.3 where no structures/voids are affected by similar smoothing scale. We still prefer counting voids using the Fourier smoothing technique that takes the noise level into account. Here, it is important to notice that we trust that a blind void count between quasar and simulated spectra, using the same selection criteria, is the key to produce a robust result.

Metal lines are expected to be present in quasar spectra, but they will appear narrower than hydrogen lines due to larger atomic masses that lower the velocity distribution. In section 4. *Systematic Errors* of Villaescusa-Navarro et al. (2011), the typical size of metal lines is mention to be ~ 0.05 Mpc/h, which is well below our smoothing scale used for both quasar and simulated spectra. As a result, we conclude that metal lines of sizes ~ 0.05 Mpc/h will not affect the void count.

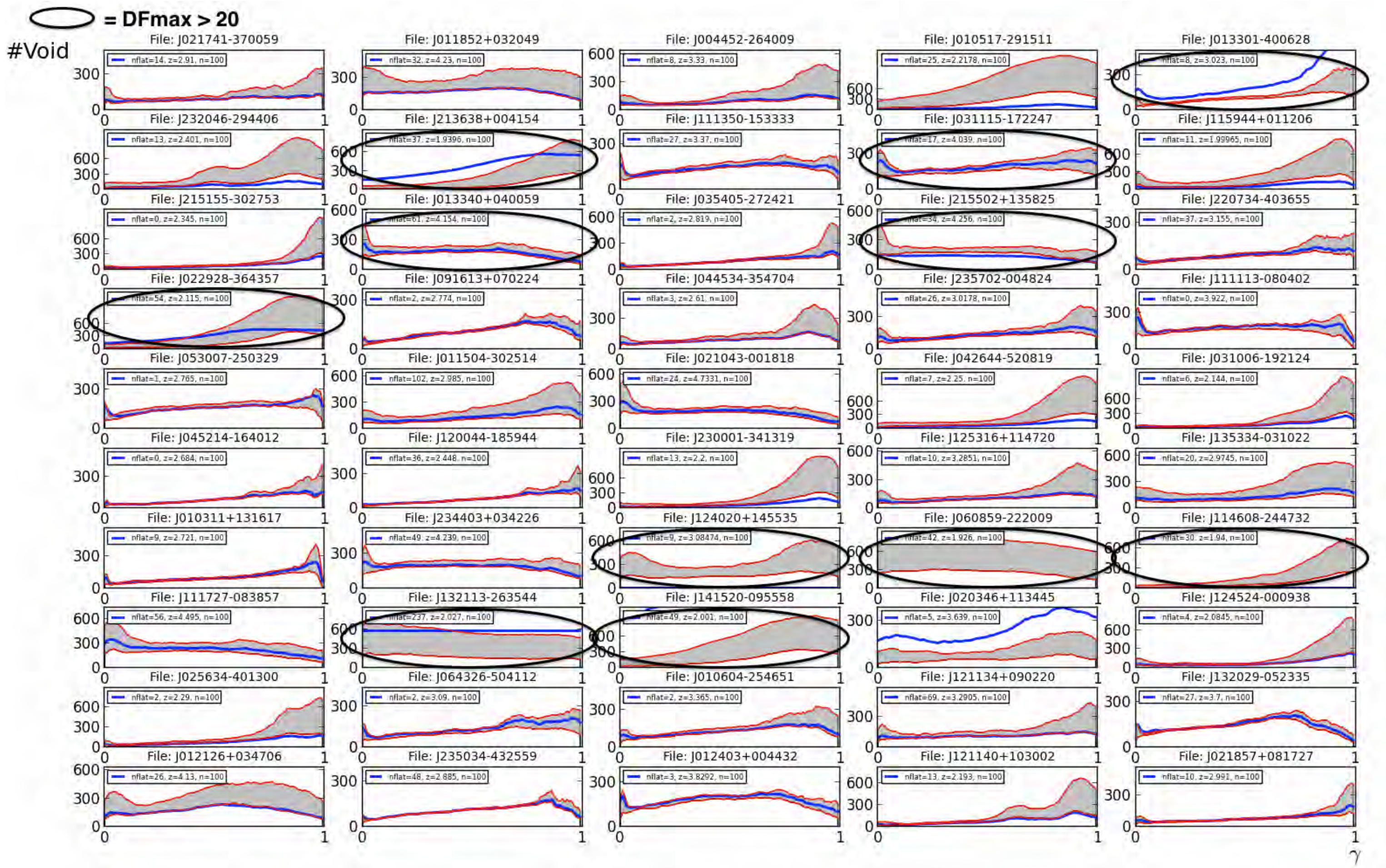


Figure 34: *The void count as a function of flux density threshold for our 50 quasar subsample. The red lines are Gaussian smoothed with $\sigma = 1$ (upper) and 3 (lower) and the blue line is the automatic Fourier smoothing. The black circles are spectra with $\Delta F_{max} > 20$ (see section 4.3.2) resulting in incorrect Fourier cut values.*

7 Quasar Normalization - Estimating the Void Count Uncertainty

The light from a quasar is absorbed on its way to the Earth, and the spectrum's background emission shape is difficult to determine with high precision. As a part of the pipeline processing, the spectra are normalized by a rigid polynomial fitting (e.g. a Chebyshev or a Lengendre polynomial)[7]. The normalization will mainly introduce an uncertainty in the left (low wavelength) part of the spectra where the absorption lines of the Lyman forest reside (see dip in the left side of top plot in Figure 15).

We wish to test how sensitive the void count is to the emission shape normalization by simulating the pipeline process. The uncertainty that arises can be simulated by multiplying the spectra with different variational functions and measuring how the void count changes. To mimic the pipeline reduction in King et al. (2012) and Murphy et al. (2003) we use a rigid Legendre polynomial to perturb the spectra. The aim is to perturb the continuum of the spectra, and to do so we create an initial Legendre continuum fit using the 3σ cutting method (see section 7.1).

The continuum fit is randomly perturbed and multiplied onto the original spectra to create a set of Random Continuum Perturbed (RCP) spectra (section 7.2). Each RCP spectra is now compared with the original to find the relative void count difference, $(\#\mathrm{V}_{original} - \#\mathrm{V}_{pert})/\#\mathrm{V}_{original}$ as well as the relative change between the spectra, the Kolmogorov-Smirnov (KS) D-value (see Appendix B.2). The D-value is a measure of the relative difference between two spectra and will, plotted against the void count difference, tell us how sensitive the void count is to the uncertainty of the normalization (section 7.3).

[7] see http://mathworld.wolfram.com/LegendrePolynomial.html

7.1 3σ Fitting

To construct the continuum I start by fitting an N'th order Legendre polynomial to the original spectrum. The points which are 3σ below the polynomial are removed and a new fit is made. Repeating this, results in a fit that converges towards the continuum. In Figure 35 I have fitted a Legendre polynomial (blue) to a spectrum (red). The green line shows the optimal fit after using the 3σ cutting method. The subplot is an optimization plot showing the convergence of the fit. The figure clearly shows that we can create a fit

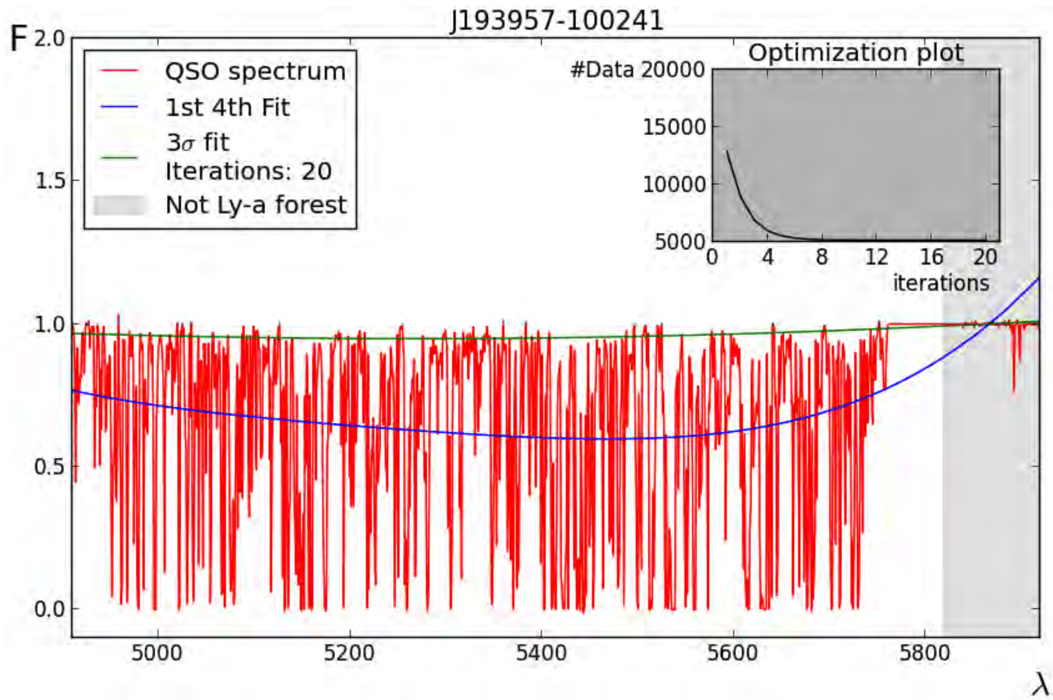


Figure 35: *The spectrum of QSO J193957-100241 (red), with the initial Legendre fit (blue). After 20 iterations of 3σ fitting we obtain the green line. The grey region in the right side is added due to its strong continuum around 1 that helps stabilize the continuum fit. An optimization curve is shown in the subplot with the amount of data as a function of number of iterations.*

that resembles the continuum of the spectrum using the 3σ method. To produce a better continuum fit I added an extra 100 Å on the right side of the Lyman-α forest (see grey region). Here the continuum is particularly good and stabilizes the 3σ fit (green line). Figure 35 is reproduced for 10 random quasar spectra from the 50 subsample and shows that the combination between rigidness and goodness of fit suggests using a $4th$ order Legendre polynomial. The continuum is well fitted within 10 iterations of the 3σ method.

7.2 Random Perturbation of the spectra

The 3σ constructed continuum fit is now used to construct the RCP spectra by multiplying each (randomly) perturbed continuum with the original spectrum. The continuum perturbations are made from adding random numbers to each of the Legendre polynomial constants according to their size. As a result the relative perturbation is dependent on the Legendre continuum fit that is different for each spectrum.

In the end, we use the KS D-value to measure the amount of perturbation in each spectra, so the exact values of the constant (initial perturbation) are less important. In Figure 36 I show a comparison between the original (red) and 15 RCP spectra (blue). Here the relative factor (denoted by Rel-pert - see legend) is set to 0.5.

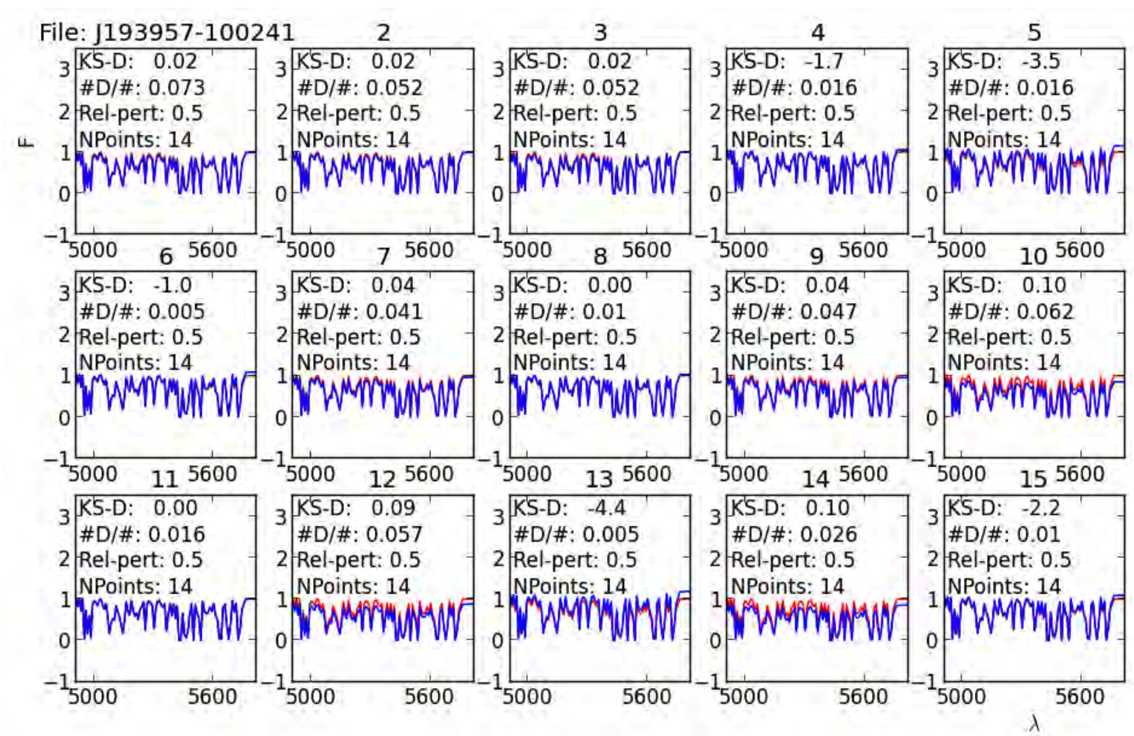


Figure 36: *Here I have plotted 15 randomly perturbed spectra (blue) compared to the original quasar spectra (red). The relative factor that determines the amount of perturbation is set to 0.5. Each of these RCP spectra will be used to test the normalization uncertainty.*

7.3 The Void Counts Sensitivity

The uncertainty of the normalization is affecting the void count in the spectra. I will count the relative number of voids between each RCP spectra and its original spectrum and compare this with their respective flux differences. The flux difference between two spectra is defined by calculating the D-value that represents the relative difference between the spectra.
In Figure 37, the relative void count difference, $(\#\mathrm{V}_{original} - \#\mathrm{V}_{pert})/\#\mathrm{V}_{original}$ is plotted with the flux difference between the spectra which is described respectively by the $\Delta\chi^2$ (left) and the D-value (right). The $\Delta\chi^2$ cannot determine whether the perturbed spectrum is above or below the original spectrum so the left figure fluctuates unnecessarily but serves to support the trend of the right figure. The KS test is on the contrary a cumulative sum that takes into account whether the RCP spectrum is raised or lowered, and appear more smooth. We will base our systematic error estimation of the normalization on the D-values.

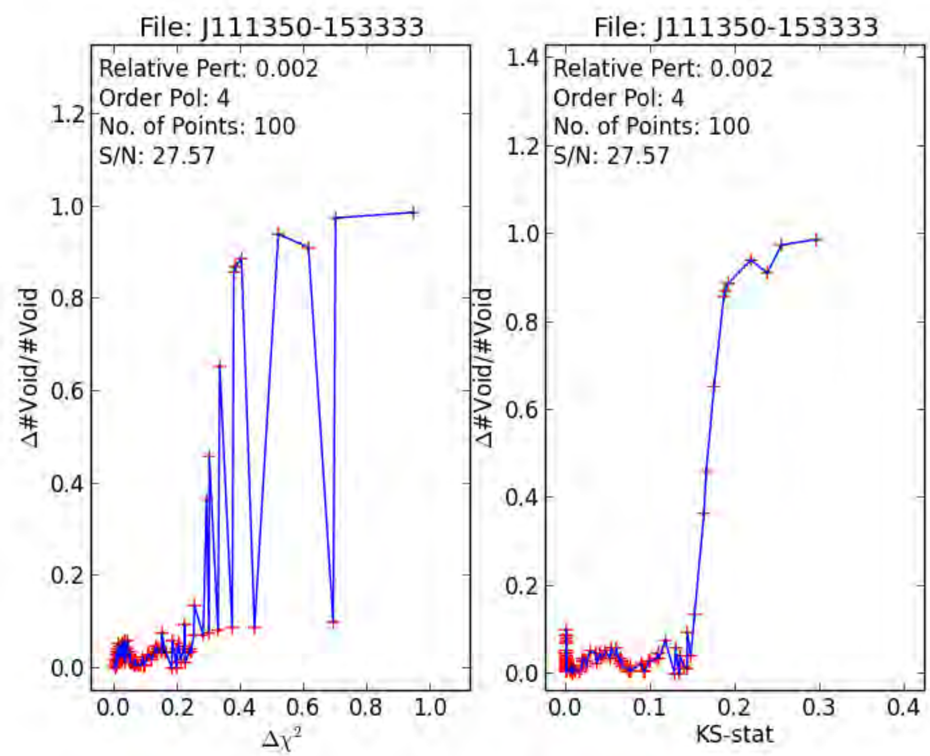


Figure 37: *The left plot shows the relative void count difference as a function of the $\Delta\chi^2$ and the right a function of the KS statistic's D-value. In the right plot a steady void count change of maximum* 5 % *is seen here when the normalization uncertainty is allowed to vary up to* 10 %.

The quasar spectrum in Figure 37 with a $S/N = 28$ shows in the right plot a maximum of 5 % void count change when the spectrum is perturbed up around 10 %. I have made this analysis on 10 random quasar spectra from the 50 subsample. In Table 1 I have listed the quasar spectra with their maximum void count uncertainty for variation of 10 % in the normalization (D-value) as well as the S/N value. The average value in the table describes the void count uncertainty as 17.5 %, when the normalization varies of the order of 10 % (see Murphy, 2013). There is not a strong correlation between the void count change and the S/N values when the D-value is fixed at 10 %, which suggest that factors other than the S/N affects the systematic normalisation uncertainty.

	Δ#Void/#Void (%)	D-value (%)	S/N
J011852+032049	0.2	0.1	6
J012403+004432	0.2	0.1	20
J025634-401300	0.4	0.1	11
J053007-250329	0.1	0.1	67
J091613+070224	0.1	0.1	51
J111350-153333	0.05	0.1	28
J121140+103002	0.1	0.1	14
J125316+114720	0.2	0.1	10
J132029-052335	0.2	0.1	36
J235702-004824	0.2	0.1	18
Average	**0.175**	**0.1**	25

Table 1: *The relative void change (left column), the relative flux spectra change (middle column) and the S/N values (right). In the bottom I have calculated the mean to estimate the statistical significant systematic error.*

8 The Void Count Analysis

We have introduced the cosmological background, the project outline and how to observe voids in quasar spectra followed by a range of improvements, testing, as well as a selection of good criteria that makes the void count insensitive to small parameter variations. We estimated the systematic uncertainty from the flux normalization and will below perform the analysis and comparison of the void count from observed and simulated Universes.

8.1 The Flux Density of Voids

The spatial extend of voids depends on how the density transition between voids and structures is defined. When observing voids in quasar spectra this transition is set by the flux threshold (γ). A high density transition level (similar to low flux density threshold) corresponds to large voids, as only the densest parts are defined as cosmological structures.

In order to compare our void counts to other published papers we adopt the preferred threshold convention from Viel et al. (2008). Here, they relate the voids found in 1D flux spectra (LOS) with the actual 3D dark matter distribution in simulations and recover the relation,

$$\delta + 1 = \rho \ / < \rho >= -0.9 \ F \ / < F > +1.9 \ . \tag{35}$$

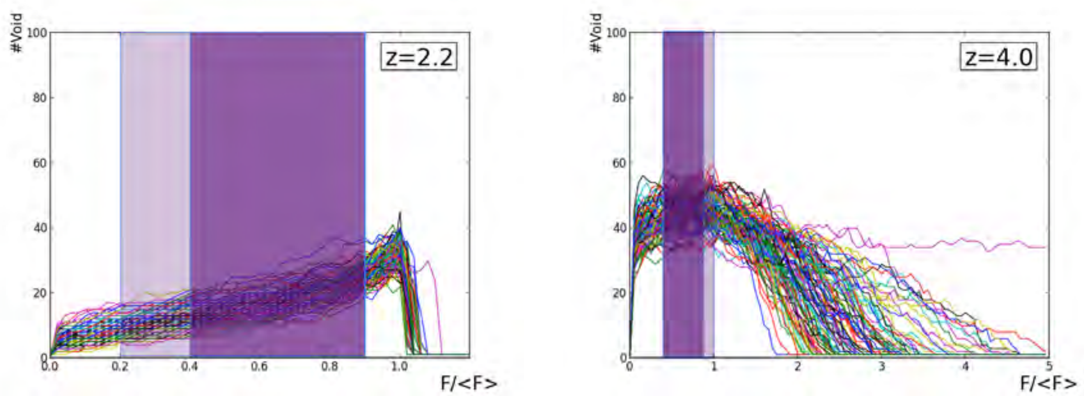
The equation relates the normalized density, $\rho \ / < \rho >$ (3D matter distribution), and the averaged flux, $F \ / < F >$ (1D flux distribution) at $z = 2.2$. We redefine our flux density threshold, γ to $F \ / < F >$ that is related to the real matter distribution. In the same article they also mention that the correlation between F and δ is strongest if only selecting voids larger than $7 - 10$ Mpc$/h$.

8.1.1 A stable flux density threshold

We determine a new preferred averaged flux density threshold ($F \ / < F >$) for which the void counts are robust and stable[8]. We use the simulated spectra to find a good threshold, in redshift bins $z = 2.2$ and 4.0, as the 50 quasar subsample lack the required statistics. We compare the void count from each redshift bin, and a common averaged flux threshold is chosen.

[8] We use stable to address a void count change as little as possible for a small variation in threshold.

SIMULATED SPECTRA



(a) **(b)**

Figure 38: *The void count as a function of the averaged threshold, $F \ / < F >$ for $z = 4.0$ (b) and 2.2 (a). The peak of the void count shifts towards higher threshold values because of the continually ongoing collapse of structures. The void density are stable in the light purple shaded areas, with $F \ / < F >= 0.2 - 0.9$ (a) and $0.4 - 1.0$ (b), where (a) and (b) overlap in dark purple area of $F \ / < F >= 0.4 - 0.9$.*

In Figure 38 I show the void count of simulated spectra as a function of $F \ / < F >$ for $z = 2.2$ (left) and 4.0 (right) to find a threshold with a stable void count. We select a threshold that can be used at both redshifts. The $z = 2.2$ and 4.0 plot in Figure 38 are most stable in the respective range from $F \ / < F > = 0.2 - 0.9$ and $0.4 - 1.0$ (see light purple shaded areas). Given the overlap between $0.4 - 0.9$ (dark purple shaded area) the clever choice of a robust threshold, one with a high void count that avoids the unsure high and low ends, is selected to be

$$F \ / < F > = 0.65 \text{ (primary)}. \tag{36}$$

Here $F \ / < F > = 0.55$ and 0.75 are chosen as secondary candidates, within the good threshold limit. It should be noted that the newly selected threshold is restricted by choosing a stable value covering void counts at both $z = 2.2$ and 4.0.

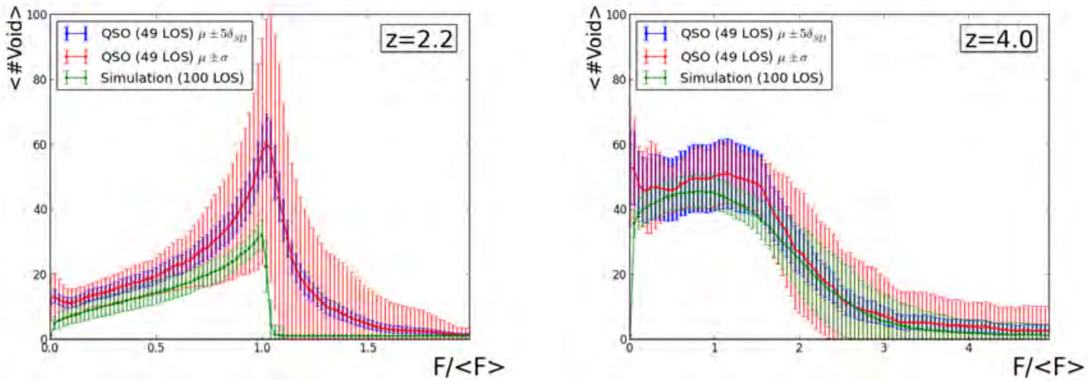
8.2 Mean number of voids

The number of voids decrease as the Universe grows older, joining each other as a result of matter falling into over-dense areas. The interior mass of the voids plays a part in how quickly the matter evacuates into the denser areas, relating the number of voids to the void mass. The void mass in simulations depends on the initial parameters, so by comparing the average void count between simulated and real quasar spectra we should be able (with good quasar statistics) to test if the interior void mass from simulations match the real Universe.

We perform our first statistical void count comparison between simulated and observed quasar spectra, using the best guess simulation from Rossi et al. (2014) (see Table 5 in Appendix C) and the large UVES quasar sample. The LOS from the simulated spectra of $z = 2.2$ and 4.0 have a spatial extend of 100 Mpc/h. In order to compare the void count between simulated and real spectra, we bin the quasar spectra on the same scale. The wavelength width similar to a 100 Mpc/h spatial slice is calculated using equation (30) setting the constant, $C_{void} = 100$ Mpc/h. The function then returns the wavelength width as a function of redshift (green line in left plot of Figure 21). The resulting number of quasar spectra after the binning, from the total and prime sample, can be seen in Figure 12.

In Figure 39 I have plotted the mean void count as a function of threshold for simulated (green) and quasar (red) spectra LOS at $z = 2.2$ (a) and $z = 4.0$ (b). The red and green error bars illustrate the spread, σ, from a gaussian fit whereas the blue error bars are the Standard Deviation On the Mean (SDOM), δ, determined from the normalisation uncertainty of 17.5 % (the error calculation is covered in Appendix B.3). I have multiplied the δ-error bars with 5 for it to be visual in the figure. The quasar void count error, σ (red), is larger than the simulated (green) because of the observational uncertainty and low number of spectra.

The average number of voids drop from $z = 4.0 \rightarrow 2.2$ as the Universe grows older. This effect of the void count drop is largest for the low flux density threshold, $(F \, / < F >) \leq 0.6$ that corresponds to only observing the most empty voids. In order for a void to become very under-dense it needs time and heavy structures surrounding it, pulling out the matter. Because of this, the emptiest voids are usually also the biggest and oldest. When heavy structures merge, their absorption lines blend together and naturally as time passes the void count ($\sim$ number of absorption lines) will also drop.



(a) **(b)**

Figure 39: *The mean void count and the resulting spread of all quasar void counts at each threshold value. The quasar and simulation data is represented by the red and green data points respectively. The green and red error bars illustrate the spread, σ, whereas the blue are the SDOM, δ.*

I will now test how well the void count of the best guess ΛCDM simulated spectra match the real number of voids in the Universe, using the two types of observational errors, σ and δ. In order to quantify this I calculated the χ^2 value (see Appendix B.4) between the simulated and quasar void counts, using both types of errors. The χ^2 and χ^2_{red} (χ^2/dof) are calculated using σ (red) and δ (blue), for both redshift bins and listed in table 2.

When comparing the simulated and quasar void count using the SDOM, δ, as the observed error value, we get a very poor χ^2_{red} value for both redshifts (see table 2). This suggests that either the SDOM is underestimated (or we miss contributions from other sources of errors) or that the quasar void count is not consistent with the number of voids from simulations and their common distributions can be rejected with a high significance level. Another explanation could be that the choice of cosmological parameters from the simulation introduces a systematic shift in the void count (see section 9.1.4). On the contrary if we calculate the difference using σ as the significant error, we get good χ^2_{red} values ($\chi^2_{\text{red}} \sim 1$), see table 2. The quasar void count slightly fits the number of voids from simulated spectra too well, which can be due to overestimated errors, that makes the two data sets appear very identical. The two datasets are consistent, given we use σ with a significance of 99 % implying that the number of voids from simulation fits the void count from quasar spectra in the real Universe.

The two error choices create two distinctive fits to the simulated spectra. The spread, σ, takes all errors into account and is most likely overestimated. On the other side the normalization uncertainty, δ, is estimated from a reliable perturbation method, which is probably not the only source of error. I have combined the two errors, σ and δ into ϵ using equation (43) from Appendix B.3. The χ^2 results are listed in table 2 (purple) which provide a more realistic guess of the similarity between the datasets. The χ^2_{red} mildly suggests that it is unlikely that the two datasets are consistent ($\chi^2_{\mathrm{red},\epsilon} \geq 1$). In order to determine if the simulated void counts are comparable with observed ones at a higher level of significance, we need to increase the amount of data.

z	χ^2_σ	$\chi^2_{\mathrm{red},\sigma}$	χ^2_δ	$\chi^2_{\mathrm{red},\delta}$	χ^2_ϵ	$\chi^2_{\mathrm{red},\epsilon}$
2.2	60	0.59	22103	223	1476	14.91
4.0	27	0.26	1884	19	563	5.68

Table 2: *The χ^2 and χ^2_{red} values colour coded to fit the red and blue errors in Figure 39 for both redshifts. The blue colour, representing the δ error shows a poor $\chi^2_{\mathrm{red}} > 1$. The data is described by the σ error having $\chi^2_{\mathrm{red}} < 1$, is on the contrary a good value. The purple shaded regions is the errors combined into the best error estimate possible, suggesting a poor fit, $\chi^2_{\mathrm{red},\epsilon} \geq 1$ between the data and the model.*

8.3 Void slices at uniform flux density

Voids are strange empty objects that slither and exist in-between the dense walls, filaments, and clusters of matter. As we just experienced, their size and the number along LOS change with the definition of the matter density transition. In our case, when observing the 1D spectra, this choice is set by the averaged flux density threshold. In section 8.1.1 we found a range of stable thresholds for which we now look at the distribution of voids, again comparing the quasar and simulated spectra.

In Figure 40 I have shown the void count probability distribution function (PDF) for the stable thresholds, $0.55, 0.65, 0.75$ (horizontal) at $z = 2.2$ (upper) and 4.0 (lower). The red, green and blue curves are respectively the simulated, prime and total spectra samples.

The effect of gravitational contraction, mentioned in section 8.2, results in a void number decrease as the Universe evolves (decreasing z). This is also visible comparing the upper and lower plots of Figure 40. The average void count drops, for both the simulated and quasar spectra, as the Universe grows from $z = 4.0 \rightarrow 2.2$.

The plots in Figure 40 resemble threshold slices (parallel to the y-axis) of Figure 39 where what before was the y-axis, have now become the x-axis. Similar the errors, σ and δ, represent the uncertainty of the void count based on the spread and normalization again. In Figure 40 it is clear that when including the spread, σ, as the uncertainty, the void count from the simulated (red) and total quasar sample (blue) easily overlaps and could be consistent datasets. Oppositely, when the error is determined by the normalization, δ, it appears unlikely that the void counts are from the same Universe.

I have included the prime sample from section 4.3.3 to test how the data-less pixel removal and the S/N values affect the void count. Unfortunately, when the best spectra were selected from the total sample, there were not enough spectra left at $z = 4.0$. The void count from the prime sample overlaps the total number of voids within both, σ and δ. This suggests that the data-less pixel removal and the S/N values do not affect the final void count, and cannot be a part of any systematic error.

The fact that the quasar void count distribution has a higher average value compared to the simulation (within the δ errors), appears to be quite interesting when listing the different systematic uncertainties, see table 3. In table 3 is listed the possible systematic uncertainties, their effect on the void count and the reasoning behind. Most of the cases result in a decrease of void count and cannot explain the higher observed average void count in observational spectra. Two cases can result in a increased void count: under-smoothing the spectra with the Fourier method and a downwards shift of the flux spectra (continuum) from incorrect normalization.

Possible systematic errors	Effect	Explanation
Large flat regions in the spectra	Decrease #voids	The absorption lines normally present would be missing in flat regions and therefore producing a lower void count.
Over smoothing (Fourier)	Decrease #voids	Over smoothing the spectra, removes additional absorption lines lowering the void count.
Under smoothing (Fourier)	Increasing #voids	Under smoothing the spectra, leaves noise in the spectra that will be counted as artificial extra voids, increasing the void count.
Normalization (sloped continuum)	Decreasing #voids	The threshold will cross less absorption lines in a sloped continuum, and the result would be a lower void count
Normalization (shift of continuum)	Averages out (increases #voids)	Shifting the continuum up or down, can lower or increase the void count. We expect with a statistical count that such an effect would average out. Although if the normalization is shifted down a higher number of voids would be counted.
higher gravitational clumping	Decreases #voids	In Figure 31 it is clear that only a higher amount of smoothing would affect the void count, although making it lower.

Table 3: *A list of possible systematic uncertainties, their effect on the void count and an explanation.*

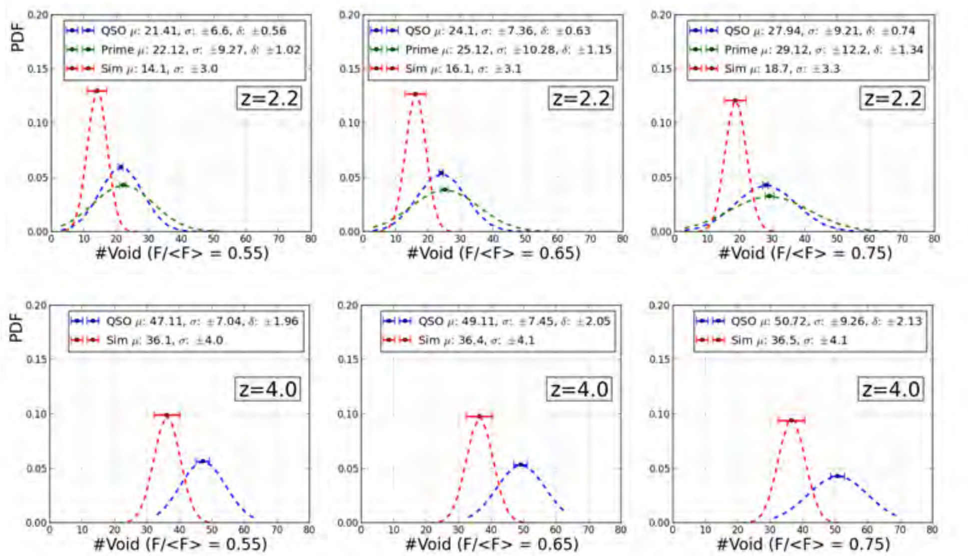


Figure 40: *The void count PDF as a function of the averaged threshold,* $F\ /<F>= 0.55, 0.65, 0.75$ *for* $z = 2.2$ *(upper) and* 4.0 *(lower).*

From Figure 33 and 34 in section 6.4 it should be clear that the Fourier smoothing scale is correctly calibrated, and that a under-smoothing is unlikely. The larger observed void count is unlikely to be from the downwards shift of the flux spectra. The effect is included in the normalization uncertainty, δ, that is to small to describe a systematic void count shift of that size.

The combination of the errors in section 8.2 supports the fact that more data is still needed before drawing any further conclusions. So far it is difficult to explain this discrepancy with the systematic errors listed in table 3. Another possibility is that the systematic error comes from the choice of cosmological parameters and/or conversion between density and flux spectra arising from the N-body simulation.

8.4 Neutrino Signatures on the High Transmission Regions of the Lyman-α Forest

The idea to count voids along LOS in quasar spectra was inspired by Villaescusa-Navarro et al. (2011). They found a relation between the sum of the neutrino masses, $\sum_i m_{\nu,i}$ and the number of Lyman-α voids along LOS. The main result in their paper is shown in Figure 41. Below I explain briefly what is seen in the plots.

Figure 3 (a)
In Figure 41 (a) they show the distribution of the most probable void count (PDF) at $z = 2.2$ (upper) and 4.0 (lower). Increasing the neutrino (red, blue, purple) mass clearly shifts the curve towards a lower void count peak. The green ($\sum_i m_{\mu_i} = 0.6\ eV, \sigma_8 = 0.877$) curve is a remake of the red ($\sum_i m_{\mu_i} = 0.0\ eV, \sigma_8 = 0.877$) with the same σ_8 parameter but different sum of neutrino masses. This results in the peak moving towards a higher void count.

Figure 4 (b)
Figure 41 (b) shows number of voids as a function of the averaged threshold ($F\ / < F >$) again for $z = 2.2$ (upper) and 4.0 (lower). The subfigure in the upper plot shows the curves divided by the red curve with zero neutrino mass. The figure sums up that the effect of neutrinos mentioned in section 2.2.1 have a clear effect on the number of voids along LOS.

We have reproduced Figure 41 (b) and (a), with observational quasar spectra in Figure 39 and 40 respectively. The trend of the decreasing void count as a function of $F\ / < F >$ in Figure 41 (b), resembles the void count decrease we also see in our Figure 39 in the ranges $F\ / < F >= 1.06 - 1.15$ ($z = 2.2$) and $F\ / < F >= 1.2 - 1.9$ ($z = 4.0$). Although we have been able to reproduce the figures, it is not possible for us to make any direct conclusions as they have used a different set of selection criteria. We defined our selection criteria based on newer simulations (Rossi et al., 2014) for which a grid would be available if the method was found to work. If we in the future obtain good enough statistics, and attempt to constrain the neutrino mass the key ideas behind this paper is important. They have optimized the observables to be most sensitive to neutrino mass difference.

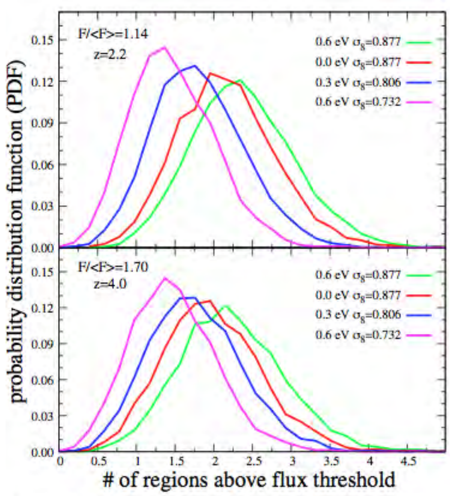


Figure 3. Probability distribution function (PDF) for the number of regions per path length of 100 h^{-1}Mpc above a threshold of $F/\langle F\rangle = 1.14$ (top), 1.70 (bottom) as a function of $\Sigma_i m_{\nu_i}$ and σ_8 at $z = 2.2$ (top) and $z = 4$ (bottom). The PDFs have long tails with a very low probability that extend up to 10-12. The $\sigma_8 - \Omega_\nu$ degeneracy is not perfect and can be broken by studying the spectra at different redshifts.

(a)

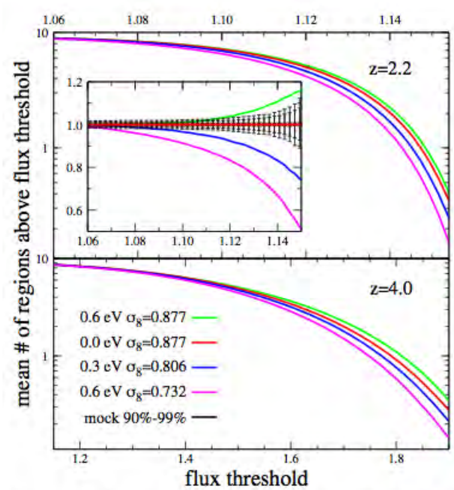


Figure 4. Average number of regions per path length of 100 h^{-1}Mpc as a function of flux threshold at redshift $z = 2.2$ (top) and $z = 4$ (bottom) for different neutrino masses and σ_8. The subplot in the upper panel shows the ratio between models with $\Sigma_i m_{\nu_i} \neq 0.0$ and the model with $\Sigma_i m_{\nu_i} = 0.0$. The black error bars indicate the 90% (interior tick marks) and 99% (exterior tick marks) confident intervals for a mock catalog consisting of 200 RLOS taken from the simulation with ($\Sigma_i m_{\nu_i} = 0.0$ eV, $\sigma_8 = 0.877$). Models with $\Sigma_i m_{\nu_i} = 0.3, 0.6$ and $\sigma_8 = 0.806, 0.732$ respectively can be ruled out with a high significance by using a catalog of 200 QSO spectra.

(b)

Figure 41: *Figure 3 (a) and 4 (b) from (Villaescusa-Navarro et al., 2011).*

8.5 Measuring the Effect of Gravity

The complex pattern of the Large Scale Structure emerges when the force of gravity pulls matter together as it evolves from an almost homogenous early Universe. Consequently a gravitational pattern also emerges from the under dense regions, that at different redshift resemble different stages of evolution. With time the voids "merge" and become bigger, leading us to expect fewer bigger voids at lower redshifts.

With my automatic void counting code it is possible to test if and how the number of voids decrease as the Universe grows older. We can do this by selecting quasar spectra at different redshifts and stacking their void count to produce an average void count over redshift. To follow our convention we choose to count voids in a 100 Mpc/h slice of the spectra centered at each of the given redshifts.

In Figure 42, I have plotted the average void count for the total (upper) and prime (lower) quasar sample as a function of redshift, z. The colours green (prime), red and blue (secondary) represent the averaged threshold value of respectively $F \;/ < F >= 0.65, 0.55, 0.75$ (see section 8.1.1).

From $z = 1.8 - 4$ ($1.8 - 3.5$ being the most significant region, see Figure 13) the number of voids increase with redshift, showing the same trend for both total and prime samples, and for each of the threshold choices $(0.55, 0.65, 0.75)$. The higher number of voids for increasing redshift is consistent with the expected (simple) picture that many small voids occupy the early Universe, and as the Universe grows older evolves to fewer large voids.

In section 7 table 1 we estimated the normalization uncertainty on the void count to be 17.5 %. To get the average void counts and their errors in Figure 42 we calculated the mean and the SDOM (see Appendix B.3). One thing to notice is that the error bars increase with redshift as a result of decreasing number statistics (see Figure 13). The turquoise crosses from simulated spectra at $z = 2.2$ and 4.0 both have lower average void count than what we found in the observed data, supporting the results from section 8.2 and 8.3.

We can compare the void count-redshift distribution to previous results from Petitjean (1998) who introduces a relation between the number of Lyman-α lines and redshift,

$$N(z) = N_0(1 + z)^k \,. \tag{37}$$

The number of lines is related to the number of voids like, $(N_{voids} = N_{lines} + 1)$. We calculate the normal and χ^2_{red} to estimate how well the model (equation 37) fits the observed data (see table 4). Focusing on the χ^2_{red} value it is clear that the prime sample produces the best fit. It appears that both samples fit the data well up to $z = 3.6$. For the total sample a peculiar rise in number of voids occur at $z = 4.4$, which worsen the

χ^2_{red} values. The outlying points are all more than 3σ (standard deviations) away from the model, which is very unlikely. Given that the data follows a normal distribution such an event only occurs in 1/370 making the resulting probability for such and outlier in a 16 point sample $16/370 = 0.04\,\%$. This is not good enough statistics to make a precise conclusion regarding the outlier whereas increasing the amount of quasar spectra may yield this possibility in the near future.

Sample	$F < F >$	N_0	k	χ^2	χ^2/dof
QSO Total	0.55	3.48	1.63	44.13	3.39
	0.65	5.36	1.36	46.49	3.58
	0.75	8.83	1.05	46.91	3.61
QSO Prime	0.55	4.33	1.47	12.16	0.94
	0.65	5.87	1.3	6.22	0.48
	0.75	8.93	1.05	7.88	0.61

Table 4: *The table list the fit values of Figure 42 for the total (upper) and prime (lower) quasar sample. I have colour coded each of the rows to fit the colour used in the figure for each sample, and list the threshold ($F\ /<F>$), fit parameters (N_0, k) and the χ^2 and χ^2/dof (χ^2_{red}).*

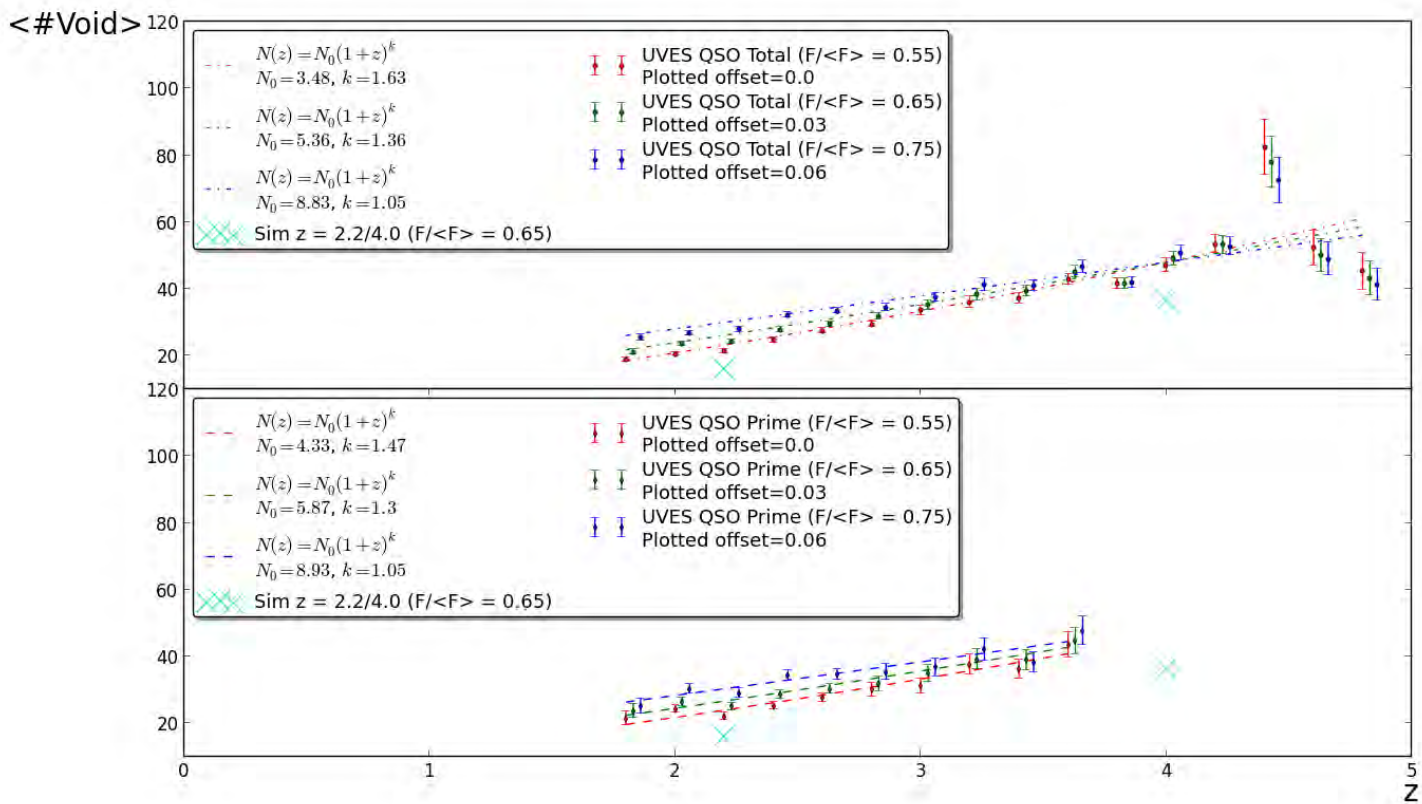


Figure 42: *The redshift evolution of the void count for respectively the total (upper) and the prime sample (lower) at different averaged flux density thresholds, $F\ /<F>= 0.55$ (red), 0.65 (green), and 0.75 (blue). I have added a small offset to the green (0.03) and blue (0.06) points in the x direction to make the error bars visible. To each of the threshold data sets is fitted the function, $N(z) = N_0(1+z)^k$ that describes the number of Lyman-α lines ($N_{voids} = N_{lines} + 1$) as a function of redshift. For $\chi^2_{\rm red}$ values see table 4. (Petitjean, 1998).*

9 Summary

I have developed a code, based on artificially created mock spectra, that can count Lyman-α voids in the 1D quasar flux distribution. The number of voids can be stacked to produce a statistical distribution as a function of redshift. To improve the robustness of the void count we run a series of tests on mock spectra and a randomly chosen subsample of 50 quasar spectra (section 6).

We construct a set of spectral selection criteria that optimizes the stability of the void count, which is affected by noise of different origin, e.g. atmospherical lines, cosmic rays or bad pixels. We remove the pipeline reduction errors which introduce data-less pixels (section 6.1.1) and construct an automatic Fourier method that smoothes the spectra according to the noise level, see section 6.1.2 and 6.4. The gravitational clumping of clouds along the line of sight blend the absorption lines in the spectra and lowers the void count. For realistic line blending the void counts on mock spectra are not affected, as shown in section 6.2.1. In accordance with Villaescusa-Navarro et al. (2011) we smooth the no-noise simulated spectra on a scale of ~ 1 Mpc/h (section 6.3), that remove astrophysical effects below the jeans scale but do not remove actual voids (> 1 Mpc).

A systematic error unfolds in the normalization of the quasar spectra and we quantify this void count uncertainty to be of the order of 17.5 % in section 7. The void counting method is also subjected to uncertainties, which are included in the spread of the proposed statistical approach ($\sim 30 - 50\%$). A final UVES quasar sample, on which the analysis is performed, is constructed from the usable spectra (properly reduced) as well as the range of the detector sensitivity resulting in a sample of 201 quasar spectra (see Figure 12 in section 4.3.4).

The stable flux density threshold was estimated to be $\gamma = 0.6$. We later discovered that the real-space dark matter voids were conventionally related to the 1D flux voids counted using a flux averaged threshold, $F\ /<F>$. The redefined prime values for $z = 2.2$ and 4.0 were found to be $F\ /<F> = 0.65$. With the selection criteria in place, the systematic error estimated and the final quasar sample ready, we redshift bin our spectra in chunks of 100 Mpc/h, resulting in a lower number of quasar spectra in each bin (see Figure 12) but better overall exploitation of the data set.

In section 8 I present our preliminary results from the, as far as we know, first statistical void count comparison between observed quasar and simulated spectra. The void count in Figure 39 and 40 from observed quasars appears to be shifted towards a higher average value compared to the number of voids in simulated spectra. When analyzing the void count distribution with the stable threshold ($F\ /<F> = 0.65$) the before mentioned shift is also present. The significance of the discrepancy depends on the favoured

error bar that either takes the form of the spread, σ, between the different LOS or the normalization uncertainty, δ. When the large σ error is adopted, the discrepancy of the void count (simulated vs. observed) becomes insignificant. Since the normalization uncertainty is much smaller than the spread, better statistics will increase the precision and will be able to tell whether the difference is real or not.
We attempt in section 8.2 to combine the errors, σ and δ (see table 2) to improve the significance of the χ^2 analysis. We get better χ^2_{red} values but still need to increase the amount of data to determine the origin of the difference.

In table 3 we list the possible systematic uncertainties and only the Fourier undersmoothing seems to be a realistic, although still unlikely, way to increase the observed void count. In section 8.4 we reproduce the two important figures from Villaescusa-Navarro et al. (2011) from observational data, to show that when very large data samples are available we can make a proper attempt at constraining the neutrino mass, ($\sum_i m_{\nu,i}$) using the grid of simulations provided by Rossi et al. (2014).

The effect of gravitational clumping that decreases the void count over time, can be seen in all the important Figures, 39, 40, and 42, from section 8. The fact that the observations behave as we expect, show that we trace real physical quantities. The void count as a function of redshift in Figure 42 display one peculiar point (near $z = 4.4$) that we should attend to when the statistics is increased. Currently the statistical uncertainty appears to be the dominating limitation at the moment, but can be improved with existing data from the Keck HIRES sample and possibly in the future with BOSS DR11 (see section 9.1.6).

9.1 Further Work

Throughout the thesis I have come across ideas to improve the coding and void count statistics, as well as applying new selection criteria for the smoothing scale, the redshift binning size, and the data sample. Below I cover these approaches in more detail.

9.1.1 Smoothing with redshift

When the light is emitted from the source to the observer, the redshift stretches the spectrum. A structure at $z = 4$ appears larger (wider in wavelength) than a similar structure at $z = 2$. The current smoothing uses the same scale for all wavelength where a better, or more precise, method would smooth the spectra with a scale that increases with redshift. The wavelength change, between $z = 2$ and 5 is approximately $\Delta\lambda = 5$ Å, see Figure 21.

9.1.2 Size of redshift bins

I have constructed the redshift bins such that the width (scale) can change. This allows us to increase our number statistics, e.g. using smaller ranges like 50 Mpc/h for both quasar and simulated spectra. Such a scale is in accordance with 99.9 % of the void sizes being $< 40 - 50$ Mpc/h around $z \sim 2$ Viel et al. (2008). We expect a similar trends at $z = 4.0$, where the void sizes naturally are smaller. We can try a range of redshift bin sizes to test how the number statistics of the void count changes and whether this technique is usable.

9.1.3 Lyman-β forest

Our void count statistics are limited by the void count per quasar spectrum. The challenge is that beyond the Lyman-α forest range each signature is either a new lower redshift structure or a higher order signature (Lyman β, δ, γ) from a high redshift structure. It is possible to determine the approximate location of the β-lines corresponding to the already counted α-lines using the formula describing the Lyman series and remove them to count the remaining α-lines in the Lyman-β forest. In principle this should apply to the entire range of the Lyman series, but realistically it will probably only work for Lyman β because of the increased probability of blending and the limited wavelength range of the detector.

9.1.4 Void count of a different cosmology

We can test the void count sensitivity to the simulation input by counting voids with the same criteria but for a slightly different cosmology, e.g. a higher matter component. If the void count is stable it implies that we can also compare our observed void count with void counts from a grid of simulations with e.g. different neutrino masses in order to constrain the $\sum_i m_{\nu_i}$.
If a systematic error arises from the specific choice of cosmological parameters, this should

also be clear when comparing void counts made from different cosmologies with our current quasar sample. In this case the void count may not be precise enough to provide information about $\sum_i m_{\nu_i}$ but can possibly help determine the standard cosmological parameters.

9.1.5 Counting Larger Voids

In Viel et al. (2008) they mention that the 1D flux follow the 3D dark matter distribution, especially, when counting voids $> 7 - 10$ Mpc/h. This opens a new and very easy way to count voids in both quasar and simulated spectra. No noise and only large scale voids will be present in the spectra after using a simple Gaussian smoothing, of ~ 10 Mpc/h. This also means that we do not need a Fourier smoothing that is slower. Such a void count can be compared to 3D galaxy void distributions, see e.g. Pan et al. (2012). When this is pursued we should keep in mind to check if the neutrino distinction still occurs at this void scale.

9.1.6 BOSS DR11 Sample

A recent study by Delubac et al. (2014) uses 137,562 low resolution quasars spectra in the redshift range $2.1 \leq z \leq 3.5$ to probe the baryon acoustic oscillation using the Lyman-α forest. Such a large sample of quasar spectra is especially interesting for a statistical void count. When they are reduced and normalized they would be a obvious choice to increase our void count significance.

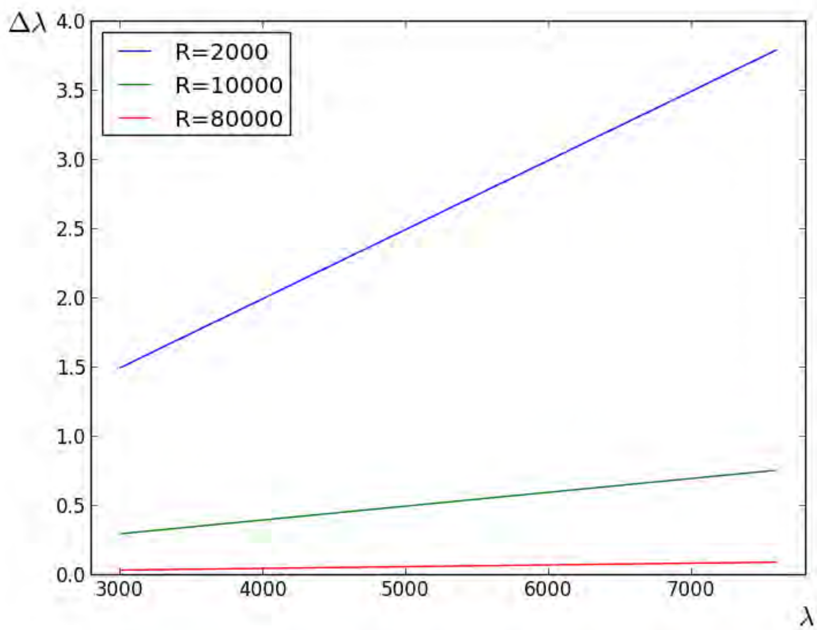


Figure 43: *A plot of $\Delta\lambda = \lambda/R$ (Å) for three different R-values,* $2000, 10.000, 80.000$. *Large R (high resolution) corresponds to small values of $\Delta\lambda$ the resolution width - and $R \sim 2000, 80.000$ respectively show a low and high resolution.*

To test, if the low resolution BOSS sample is sensitive to the voids, we compare the high and low resolutions with the corresponding wavelength scale. In Figure 43 I have plotted the resolution scale, $\Delta\lambda$, for three different resolutions as a function of the wavelength, λ. The typical UVES quasar spectra and BOSS sample have respectively a resolution of $R \sim 80.000$ (red) and $R \sim 2000$[9] (blue). The blue line changes from $\Delta\lambda = 1.5$ to 3.5 which roughly corresponds to 1 Mpc/h from $z = 2$ to 5 (equation 30). Thus it is possible for us to detect voids in BOSS low resolution spectra, especially using the idea from the previous section of counting voids $> 7-10$ Mpc/h - drastically reducing the void counting time.

[9] https://www.sdss3.org/instruments/boss_spectrograph.php

10 Concluding Remarks

The thesis approaches the first attempt to count the number of voids along LOS in high resolution quasar spectra, using the Lyman-α forest up to $z = 5$. I developed the code on mock spectra, where certain parameters could be varied and tested resulting in a robust void count. Using a 50 quasar subsample we made a range of improvements to produce a robust void count, with the void count spread and flux normalization determining the uncertainty. Determining the selection criteria from the 50 quasar sample allowed us to make a blinded void count on the remaining sample. We compared the void count distribution between observed quasar and simulated spectra. Although, in order to make further conclusions on their compatibility more data is needed.

We show that we can reproduce the figures from Villaescusa-Navarro et al. (2011) as a test of the void counting method. On top of this we show the first void count as a function of redshift where especially the formation of voids is clearly visible while inconclusive at the moment, the method carries some potential with an increase of data. The analysis was performed only on the UVES quasar sample of 201 spectra (after unusable spectra was removed), and we will within the nearest future add ~ 150 HIRES (Keck) spectra to the analysis. The code is constructed to run automatically, enabling future analysis of the void count distribution in very large datasets.

A Astronomy through the ages

Introductions to astronomy usually begins with the development of modern science by the Greek and Babylonian astronomers. Today we will take yet another leap backwards and begin by seeking answers to the questions of our existence in the first place.

The human curiosity has given us the means to learn and adapt in favour of survival (Darwin, 1860). Exploiting the seasonal variation of crops or the hunting patterns of predators gives us advantages in the natural selection. The world of today grants us the luxury of a high survival rate increasing the possibility to channelize our curiosity towards a deeper understanding of the world around us. One way of questioning our nature is to look beyond the boundaries of the Earth. We call this study astronomy, and it describes the motion of the celestial objects and their properties.

A.1 The First Astronomers

The first astronomers were priests that predicted the weather and the seasons that was of great importance to the agriculture. The first recorded signs of modern astronomy were found in Babylonian star catalogues from 1200 B.C., though it has been speculated whether the Pyramids of Giza (∼2500 B.C.) had a celestial significance. The Ancient Greeks developed astronomy using mathematics and deduction, which allowed them to perform more advanced studies of the heavens. They were the first to describe the Cosmos in a three dimensional geo-centric picture, with Plato (428-348 BC) and Aristotle (384-322 B.C.) introducing the symmetry of circles and spheres to explain the motion of the celestial objects.

The first cosmological model was born. This Ptolemaic system was a geo-centric model, that prevailed until the beginning of the 16th century, with only few extensions such as the usage of apparent magnitude (distant objects appear fainter). Aristarchus of Samos (∼ 310-230 BC) was the first to propose the idea of a helio-centric model in the 3rd century B.C., but the idea failed to gain interest, maybe affected by Ptolemy of Alexandria's (90-168) publications at the same time favoring a geo-centric picture. The geo-centric world view was not only present in the Hellenistic culture as the Mayan, Islamic, Indian and Chinese cultures developed similar models (Aaboe, 2001; Krupp, 2003).

The new paradigm in astronomy came with Nicolaus Copernicus (1473-1543) who introduced the helio-centric model with the Sun in the center. After Copernicus, the development of the model was influenced by men such as Tycho Brahe, Galileo Galilei and Johannes Kepler. Tycho Brahe (1546-1601) applied a comprehensive and accurate number of observations to prove that the planets move in epi-cycles (circles on top of circles) to confirm the geo-centric model, which he believed to be correct. Among the first men to point a telescope to the sky and perform more precise observations, was Galileo Galilei (1564-

1642), supporting the Copernican model. Johannes Kepler (1571-1630), the assistant to Tycho Brahe, used his supervisor's extensive observations to calculate the eccentricity of the planetary orbits and postulate the three *laws of planetary motion* describing the planets' motion around the Sun.

Figure 44: *A picture showing the discovery of the helio-centric model, by looking out into the Universe through the atmosphere of Earth. The picture is a digital remake of Camille Flammarion's (French astronomer) old wood engraving. Camille Flammarion (1888).*

The time for the apple to hit Isaac Newton's (1642-1727) head was up. This resulted in the publication of his book Mathematical Principles of Natural Philosophy, known as *Newton's Principia* (Newton, 1687). The book contained what today is referred to as *Newton's laws of motion*, which includes a description of the gravitational force that is the foundation of modern physics. Kepler's *laws of planetary motions* could be derived from *Newton's laws*, and the last evidence towards a geo-centric model was rejected in favor of the helio-centric Copernican picture (see Figure 44).

Here ends the first era of astronomy that was initiated by the Greek and Babylonian astronomers, who were ahead of their time with deductive mathematical methods. The geo-centric Ptolemaic model was followed and replaced by the Copernican helio-centric idea, which supported by Newton's laws became the concordant model. History shows that humans have a basic instinct of curiosity which have shaped the contemporary knowledge of the world around us. It is up to us to shape the knowledge of future generations.

A.2 Modern Cosmology

The size of the known Universe slowly grew as astronomers started to use better telescopes and discovered that the Sun was only one star among 10s of billions stars in our galaxy – The Milky Way. In 1924 the first stars were found residing outside the Milky Way in new galaxies (see Figure 45).

Observations revealed that the galaxies recedes from us, which was the first evidence to support a Big Bang model, initially proposed by Georges Lemaître (1894-1966) (Van den Bergh, 2011). After Hubble's discovery, the view of the Milky Way being the entire Universe was replaced by the Milky Way being one amongst many galaxies in an expanding Universe. In 1915 a big leap in the understanding of space, time, and mat-

Figure 45: *Our neighbor galaxy, Andromeda (M31). The picture is taken by the Hubble Space Telescope.*

ter occurred as Albert Einstein (1879-1955) published his Theory of General Relativity (Einstein, 1916). A description of a space-time that warps around massive object, and a clarification of the similarity between mass and energy. Einstein believed that the Universe was neither expanding nor contracting i.e. static and unchanging. This was contradictory to the prediction of relativity, thus he introduced a cosmological constant that made the Universe eternally static. When Edwin Hubble (1889-1953) published his findings in 1924 Einstein was forced to abandon his idea of a static Universe. The first person to apply general relativity to describe cosmology was the russian physicist and mathematician Alexander Friedmann (1888-1925) (Friedmann, 1922b). He derived an "expanding Universe" solution to the field equation of general relativity. This solution is now known as the Friedmann equations (see also section 2.1.1). Later he contributed to derive the Friedmann-Lemaître-Robertson-Walker (FLRW) metric that describes an isotropic and homogenous Universe with spatial curvature.

In 1927, Lemaître proposed that an expanding Universe must have been hotter and denser in the past, an event later named the Big Bang (Lemaître, 1927). Around 1930 *the missing matter* problem occurred while Fritz Zwicky (1898-1974) was investigating the motion of galaxies in the Coma Cluster Zwicky (1937). Today gravitational lensing effects among others imply a dark matter component (Frenk and White, 2012). Either Newton's laws are wrong, or there exists an invisible matter component, dubbed dark matter.

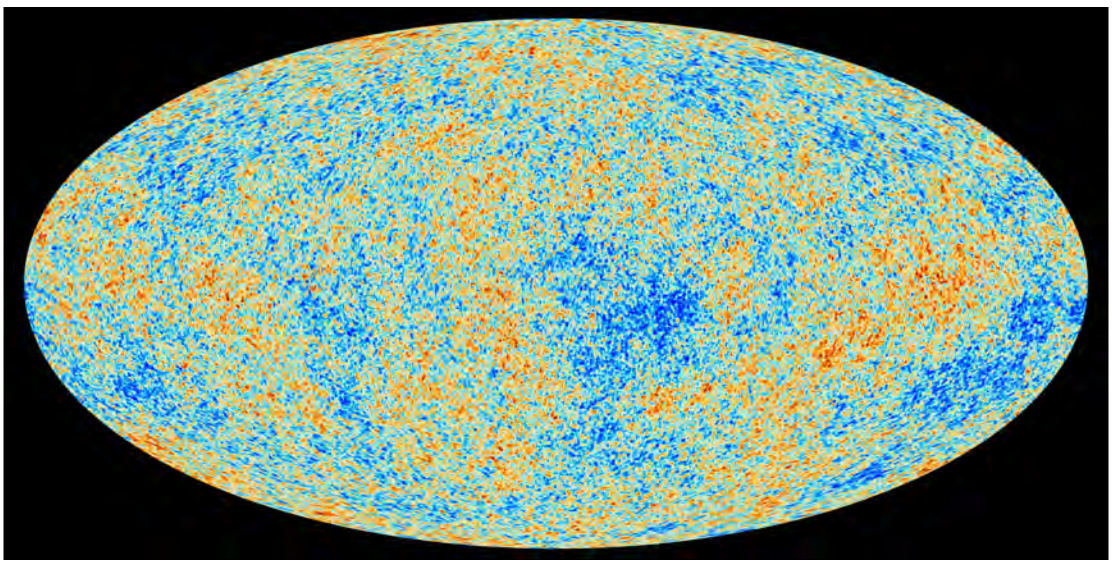


Figure 46: *The temperature fluctuations in Cosmic Microwave Background, last observed by the Planck surveyor. (Planck Collaboration, 2013).*

In 1965, one year before Lemaître's death, the Cosmic Microwave Background (CMB) radiation, a leftover of the hot glow, from the Big Bang was discovered (Penzias and Wilson, 1965) (see Figure 46). This confirmed the Hot Big Bang model that George Gamow (1904-1968) had developed and became the established cosmological model. To explain the paradox of an infinitely dense Universe at $t = 0$, the general view was that the Universe expanded and contracted in a oscillatory state, and therefore never reached infinite density. Stephen Hawking (1942-) among others (MacCallum, 2006) disproved the idea of an oscillatory Universe, in the 1960's, by showing that singularities exist in the solutions to Einstein's field equation. The Big Bang model applied to observations of the CMB gives a finite age of the Universe of 13.8 billion years (Planck Collaboration, 2013).

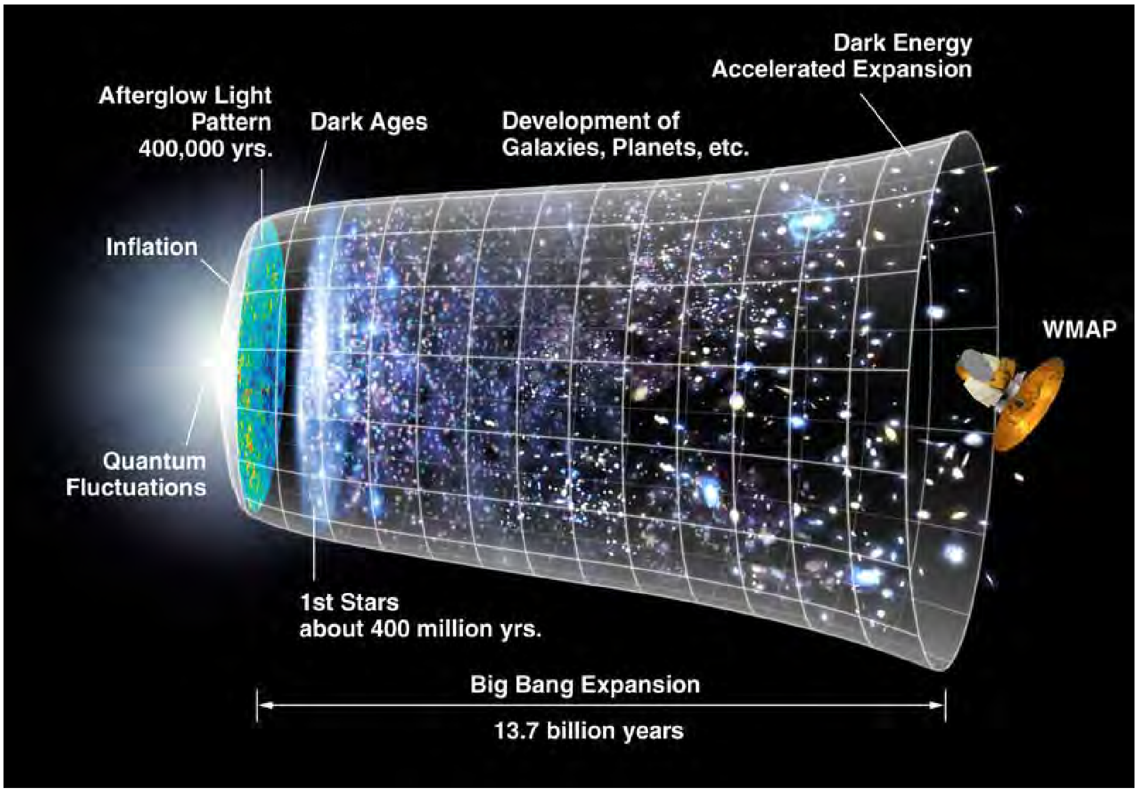


Figure 47: *The history and evolution of the Universe from Big Bang until what we see today, NASA/WMAP Science team.*

Yakov B. Zel'dovich (1914-1987) noticed that the Universe was simply too big to be homogenous and unlikely to be flat. Alan Guth (1947-) proposed a solution ten years later, in 1980, called the theory of inflation (Guth, 1981). This describes a blow up in the early universe that explains the size and flatness of the Universe as well as the monopole problem.

This was the demise of the last opposing theories, and the Hot Big Bang model was settled (Figure 47 sketches the Big Bang evolution). In 1990 the first measurements from the COsmic Background Explorer (COBE) found the CMB radiation to be an almost perfect black-body with a temperature of 2.725 K. This is supported by the Planck satellite, which made the latest CMB measurements (Figure 46). Small anisotropies in the CMB spectra were consistent with the expected gravitational effect of dark matter in the early Universe (Komatsu *et al.*, 2011). 8 years later in 1998 an accelerated expansion was suggested by Perlmutter et al. (1999); Riess et al. (1998) from a sample of fainter type Ia supernovae which resulted in a larger distance than earlier expected. The increase in distance requires a slower expansion in the past, and thereby mimics an acceleration of the expanding Universe, an effect we today refer to as Dark Energy, Λ. Hence the resulting model is called Λ Cold Dark Matter (ΛCDM) cosmology. Recently the BICEP2 collaboration published measurements that, despite of their trustworthiness, suggests the discovery of gravitational waves supporting the inflationary theory mentioned above (BICEP2 Collaboration et al., 2014).

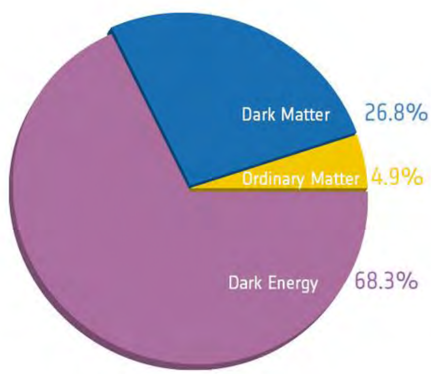


Figure 48: *The distribution of energy densities in the Universe today. (Planck Collaboration, 2013).*

We describe the Universe with energy density parameters $\Omega_\Lambda, \Omega_M, \Omega_b$ that respectively describes dark energy, dark matter, and baryonic matter. The distribution of energy densities can be seen in Figure 48. In an expanding Universe the galaxies get redshifted, which is an effect of space stretching the light (Davis and Lineweaver, 2004; Davis, 2005)[10]. The redshift can be used as a distance measure because light from distant objects gets redshifted more than light from nearby objects (see section 2.1).

We have now followed a century of great thinkers and innovators, which led us from Einstein's formulation of general relativity to Hubble's discovery of the expansion, to Friedmann's cosmology, even onwards to the discovery of dark matter, the CMB radiation, the inflationary model and last but not least the accelerated expansion.

[10]See also the introduction to redshift and energy conservation for non-experts here: `http://www.dark-cosmology.dk/~tamarad/papers/SciAm_BigBang.pdf`

B Statistical methods

I here introduce some of the statistical methods used throughout the thesis.

B.1 Signal to Noise

We calculate the signal to noise (S/N) value to estimate the amount of noise in the spectra. It can be done using different methods, but we choose to divide each flux value with its associated error to create a S/N array. To obtain a single number we take the median of the S/N array,

$$S/N = \overline{S/N_{array}} = \overline{\left(\frac{Flux_{array}}{Error_{array}}\right)}. \tag{38}$$

This method is slightly uncertain as a spectra might exhibit both a good and a bad part. With this in mind we still categorize the spectra using the S/N value shown above.

B.2 The Kolmogorov-Smirnov D-value

The KS test quantifies the difference between two datasets, using Cumulative Distribution Functions (CDF). A CDF is a cumulative sum and is defined as

$$CDF(x) = \frac{1}{N}\sum_{i=1}^{N} P(i < x). \tag{39}$$

To calculate the D-value we find the largest absolute distance between the two CDF

$$D = max\ |CDF_2(x) - CDF_1(x)|. \tag{40}$$

For a thorough introduction I refer to Barlow (1999).

B.3 Mean & Standard Deviation

In a dataset consisting of $\{x_1, x_2, x_3, ..., x_N\}$ the mean of x is

$$\bar{x} = \frac{\sum_i^N x_i}{N}. \tag{41}$$

The resulting error on $\bar{x}$, also known as the Standard Deviation Of Mean (SDOM) can be expressed,

$$\bar{\sigma}_x = \frac{\sqrt{\sum_i^N \sigma_{err,i}^2}}{N}. \tag{42}$$

Here, $\sigma_{err,i}$ describes the individual errors on x_i. When adding individual errors the formula below can be used,

$$\delta_{tot} = \frac{\sqrt{sum_i^N \delta_i^2}}{N}\ . \tag{43}$$

For examples I refer to Taylor (1997) book on error analysis.

B.4 χ^2 Calculation

When comparing two data sets the χ^2 method can be used to quantify the difference. The χ^2 is the sum of the normalized difference between the data and a chosen model,

$$\chi^2 = \sum_{i}^{N} \frac{(x_i - f(x_i))^2}{\sigma_{x,i}} \ . \tag{44}$$

Here, x_i, $\sigma_{x,i}$, and $f(x_i)$ represents the i'th data point, error, and model. If we normalize with the degrees of freedom, $\text{dof} = N - 1 - N_{var}$ we get a value quantifying the similarity, called the reduced χ^2,

$$\chi^2_{\text{red}} = \chi^2/\text{dof} = \chi^2/(N - 1 - N_{var}) \ . \tag{45}$$

N is number of data points and N_{var} is the number of variables used in the model. As a rule of thumb $\chi^2_{\text{red}} \gg 1$ imply a poor model, whereas $\chi^2_{\text{red}} < 1$ and $\chi^2_{\text{red}} > 1$ indicates over-fitting of the data (too large errors) and under-fitting the data (too small errors) respectively. A $\chi^2_{\text{red}} \sim 1$ suggest that the model within the error is a good fit to the data.

C Graphlist

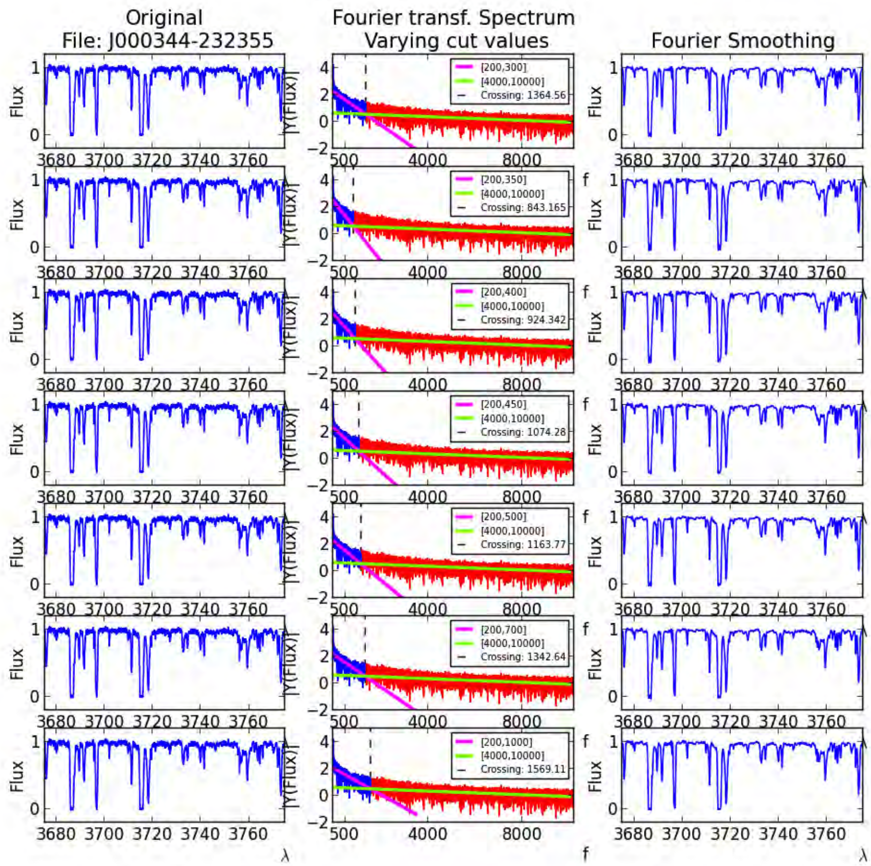


Figure 49: *An example of how we varied the fitting ranges using the pixel dependent function. Here is shown the quasar spectrum, J000344-232355 that shows the same trends as the artificial mock spectra. The least sensitive fitting values became:* $[200, 400]$ *and* $[Npix/5, Npix/2.5]$.

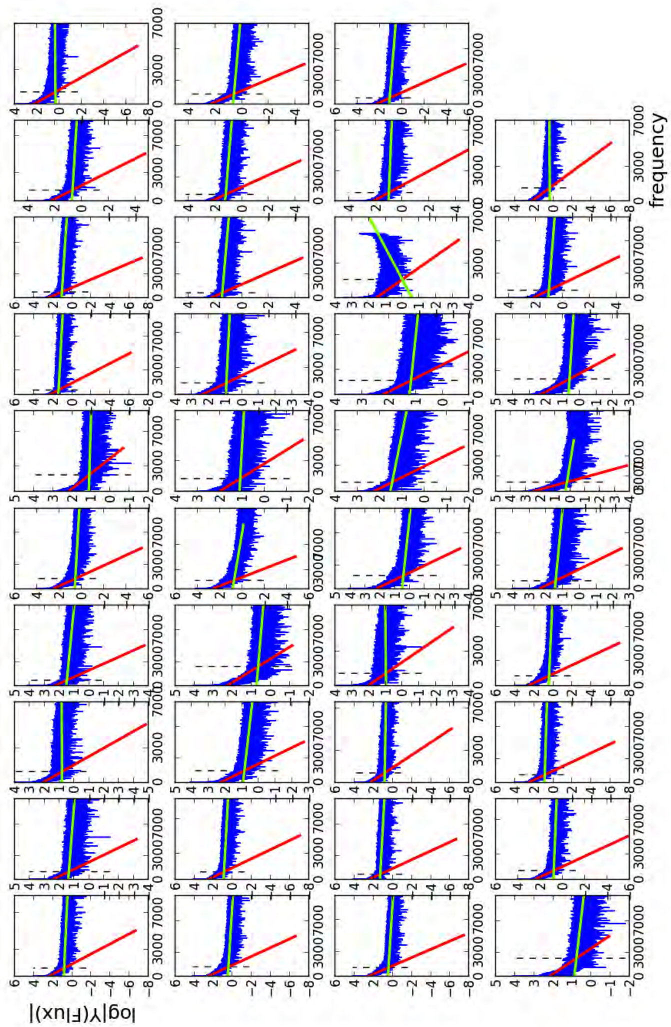
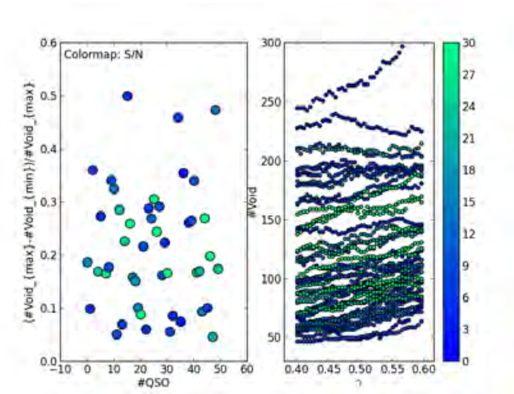
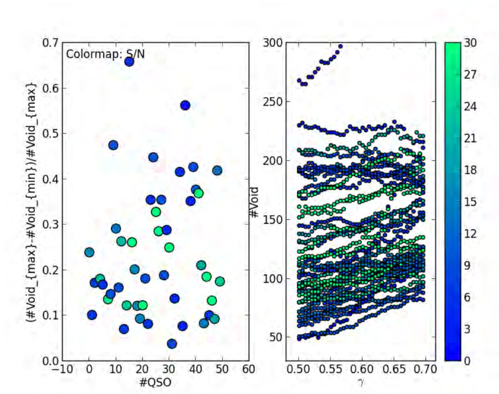


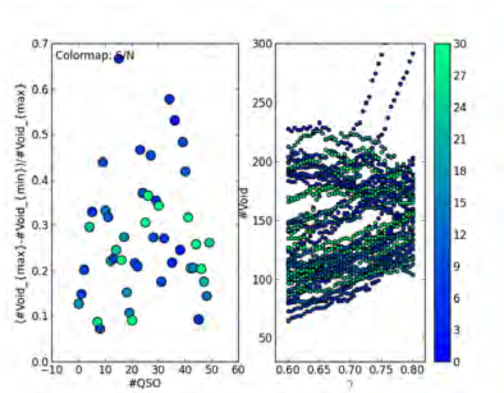
Figure 50: *The flux power spectrum of the 39 quasar (from the 50 subsample). The green and red fit to each of the Fourier frequency spectra show that we have produced a robust and automatic smoothing method.*



(a)



(b)



(c)

Figure 51: *The void count sensitivity (left) and the respective void count as a function of* γ *of the resulting 39 quasar (from the 50 sample) is plotted in each of the three figures a,b,c for respectively* $\gamma : [0.4 - 0.6], [0.5 - 0.7], [0.6 - 0.8]$. *The void count sensitivity is constructed by measuring the relative difference of the void count,* $(max(\#void) - min(\#void)) \cdot max(\#void)^{-1}$ *in the respective range of a,b, and c.*

Parameter	Value
$\sigma_8(z=0)$	0.83
n_s	0.96
H_0 [km s^{-1}Mpc^{-1}]	67.5
Ω_m	0.31
Ω_b	0.044
Ω_Λ	0.69
$T_0(z=3)[K]$	15000
$\gamma(z=3)$	1.3
Starting redshift	30

Table 5: *The best guess cosmological parameters from Rossi et al. (2014).*

References

Aaboe, A. (2001). *Episodes from the Early History of Astronomy.* Springer.

Barlow, R. (1999). Statistics - a guide to the use of statistical methods in the physical science. *Wiley*, pages 155–156.

Beringer, J. (2012). The review of Particle Physics. *http://pdg.lbl.gov/*.

BICEP2 Collaboration, Ade, P. A. R., Aikin, R. W., Barkats, D., Benton, S. J., Bischoff, C. A., Bock, J. J., Brevik, J. A., Buder, I., Bullock, E., Dowell, C. D., Duband, L., Filippini, J. P., Fliescher, S., Golwala, S. R., Halpern, M., Hasselfield, M., Hildebrandt, S. R., Hilton, G. C., Hristov, V. V., Irwin, K. D., Karkare, K. S., Kaufman, J. P., Keating, B. G., Kernasovskiy, S. A., Kovac, J. M., Kuo, C. L., Leitch, E. M., Lueker, M., Mason, P., Netterfield, C. B., Nguyen, H. T., O'Brient, R., Ogburn, R. W., Orlando, A., Pryke, C., Reintsema, C. D., Richter, S., Schwarz, R., Sheehy, C. D., Staniszewski, Z. K., Sudiwala, R. V., Teply, G. P., Tolan, J. E., Turner, A. D., Vieregg, A. G., Wong, C. L., Yoon, K. W., and Bicep2 Collaboration (2014). Detection of B-Mode Polarization at Degree Angular Scales by BICEP2. *Physical Review Letters*, 112(24):241101.

Bond, J. R., Kofman, L., and Pogosyan, D. (1996). How filaments of galaxies are woven into the cosmic web. *nat*, 380:603–606.

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., and Lemson, G. (2009). Resolving cosmic structure formation with the Millennium-II Simulation. *mnras*, 398:1150–1164.

Busca, N. G., Delubac, T., Rich, J., Bailey, S., Font-Ribera, A., Kirkby, D., Le Goff, J.-M., Pieri, M. M., Slosar, A., Aubourg, É., Bautista, J. E., Bizyaev, D., Blomqvist, M., Bolton, A. S., Bovy, J., Brewington, H., Borde, A., Brinkmann, J., Carithers, B., Croft, R. A. C., Dawson, K. S., Ebelke, G., Eisenstein, D. J., Hamilton, J.-C., Ho, S., Hogg, D. W., Honscheid, K., Lee, K.-G., Lundgren, B., Malanushenko, E., Malanushenko, V., Margala, D., Maraston, C., Mehta, K., Miralda-Escudé, J., Myers, A. D., Nichol, R. C., Noterdaeme, P., Olmstead, M. D., Oravetz, D., Palanque-Delabrouille, N., Pan, K., Pâris, I., Percival, W. J., Petitjean, P., Roe, N. A., Rollinde, E., Ross, N. P., Rossi, G., Schlegel, D. J., Schneider, D. P., Shelden, A., Sheldon, E. S., Simmons, A., Snedden, S., Tinker, J. L., Viel, M., Weaver, B. A., Weinberg, D. H., White, M., Yèche, C., and York, D. G. (2013). Baryon acoustic oscillations in the Lyα forest of BOSS quasars. *aap*, 552:A96.

Camille Flammarion (1888). L'atmosphere: Météorologie populaire. *Private publication.*

Darwin, C. R. (1860). *On the origin of species by means of natural selection, or the preservation of favored races in the struggle for life.* London: John Murray.

Davis, T. M. (2005). *Fundamental aspects of the expansion of the universe and cosmic horizons*. PhD thesis, Ph.D dissertation, 2005. Australia: University of New South Wales (Australia); 2005. Publication Number: AAT 0807562. DAI-B 66/01, Jul 2005.

Davis, T. M. and Lineweaver, C. H. (2004). Expanding Confusion: Common Misconceptions of Cosmological Horizons and the Superluminal Expansion of the Universe. *pasa*, 21:97–109.

Daya Bay Collaboration (2007). A Precision Measurement of the Neutrino Mixing Angle theta_13 using Reactor Antineutrinos at Daya Bay. *ArXiv High Energy Physics - Experiment e-prints*.

Delubac, T., Bautista, J. E., Busca, N. G., Rich, J., Kirkby, D., Bailey, S., Font-Ribera, A., Slosar, A., Lee, K.-G., Pieri, M. M., Hamilton, J.-C., Aubourg, É., Blomqvist, M., Bovy, J., Brinkmann, J., Carithers, W., Dawson, K. S., Eisenstein, D. J., Kneib, J.-P., Le Goff, J.-M., Margala, D., Miralda-Escudé, J., Myers, A. D., Nichol, R. C., Noterdaeme, P., O'Connell, R., Olmstead, M. D., Palanque-Delabrouille, N., Pâris, I., Petitjean, P., Ross, N. P., Rossi, G., Schlegel, D. J., Schneider, D. P., Weinberg, D. H., Yèche, C., and York, D. G. (2014). Baryon Acoustic Oscillations in the Ly-α forest of BOSS DR11 quasars. *ArXiv e-prints*.

Einasto, J., Einasto, M., and Gramann, M. (1989). Structure and formation of superclusters. IX - Self-similarity of voids. *mnras*, 238:155–177.

Einstein, A. (1916). Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik*, 354:769–822.

Frenk, C. S. and White, S. D. M. (2012). Dark matter and cosmic structure. *Annalen der Physik*, 524:507–534.

Friedmann, A. (1922a). Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386.

Friedmann, A. (1922b). Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386.

Guth, A. H. (1981). The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. *Phys.Rev.*, D23:347–356.

Hamaus, N., Sutter, P. M., and Wandelt, B. D. (2014). A Universal Density Profile for Cosmic Voids. *ArXiv e-prints*.

Hubble, E. (1929a). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173.

Hubble, E. (1929b). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173.

IceCube Collaboration, Abbasi, R., Abdou, Y., Ackermann, M., Adams, J., Aguilar, J. A., Ahlers, M., Altmann, D., Andeen, K., Auffenberg, J., and et al. (2013). Searches for High-energy Neutrino Emission in the Galaxy with the Combined ICECUBE-AMANDA Detector. *apj*, 763:33.

King, J. A., Webb, J. K., Murphy, M. T., Flambaum, V. V., Carswell, R. F., Bainbridge, M. B., Wilczynska, M. R., and Koch, F. E. (2012). Spatial variation in the fine-structure constant - new results from VLT/UVES. *mnras*, 422:3370–3414.

Komatsu *et al.*, E. (2011). Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *Astrophys. J. Suppl. 192*.

Krupp, E. C. (2003). *Echoes of the Ancient Skies: The Astronomy of Lost Civilizations*. Courier Dover Publications.

Lemaître, G. (1927). A homogeneous Universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *Ann. Soc. Sci. Brux.*

Liske, J., Grazian, A., Vanzella, E., Dessauges, M., Viel, M., Pasquini, L., Haehnelt, M., Cristiani, S., Pepe, F., Avila, G., Bonifacio, P., Bouchy, F., Dekker, H., Delabre, B., D'Odorico, S., D'Odorico, V., Levshakov, S., Lovis, C., Mayor, M., Molaro, P., Moscardini, L., Murphy, M. T., Queloz, D., Shaver, P., Udry, S., Wiklind, T., and Zucker, S. (2008). Cosmic dynamics in the era of Extremely Large Telescopes. *mnras*, 386:1192–1218.

MacCallum, M. A. H. (2006). Finding and using exact solutions of the Einstein equations. In Mornas, L. and Diaz Alonso, J., editors, *A Century of Relativity Physics: ERE 2005*, volume 841 of *American Institute of Physics Conference Series*, pages 129–143.

Murphy, M. T. (2013). Private communication. *Mail Correspondance*.

Murphy, M. T., Webb, J. K., and Flambaum, V. V. (2003). Further evidence for a variable fine-structure constant from Keck/HIRES QSO absorption spectra. *mnras*, 345:609–638.

Newton, I. (1687). *Newton's Philosophiae Naturalis Principia Mathematica*. Isaac Newton.

Pan, D. C., Vogeley, M. S., Hoyle, F., Choi, Y.-Y., and Park, C. (2012). Cosmic voids in Sloan Digital Sky Survey Data Release 7. *mnras*, 421:926–934.

Penzias, A. A. and Wilson, R. W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *apj*, 142:419–421.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., and Supernova Cosmology Project (1999). Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *apj*, 517:565–586.

Petitjean, P. (1998). QSO Absorption Line Systems. *ArXiv Astrophysics e-prints*.

Planck Collaboration (2013). Planck 2013 results. XVI. Cosmological parameters. *ArXiv:1303.5076*.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., and Tonry, J. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *aj*, 116:1009–1038.

Rossi, G., Palanque-Delabrouille, N., Borde, A., Viel, M., Yeche, C., Bolton, J. S., Rich, J., and Le Goff, J.-M. (2014). A Suite of Hydrodynamical Simulations for the Lyman-Alpha Forest with Massive Neutrinos. *ArXiv e-prints*.

Ryden, B. and Peterson, B. M. (2009). *Foundation of Astrophysics*. Pearson.

Seljak, U., Slosar, A., and McDonald, P. (2006). Cosmological parameters from combining the Lyman-α forest with CMB, galaxy clustering and SN constraints. *jcap*, 10:14.

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. *nat*, 435:629–636.

Taylor, J. R. (1997). An introduction to error analysis - the study of uncertainties in physical measurements. *University Science Books*.

Urry, C. M. and Padovani, P. (1995). Unified Schemes for Radio-Loud Active Galactic Nuclei. *pasp*, 107:803.

Van den Bergh, S. (2011). The Curious Case of Lemaître's Equation No. 24. *jrasc*, 105:151.

Viel, M., Becker, G. D., Bolton, J. S., and Haehnelt, M. G. (2013). Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-α forest data. *prd*, 88(4):043502.

Viel, M., Colberg, J. M., and Kim, T.-S. (2008). On the importance of high-redshift intergalactic voids. *mnras*, 386:1285–1293.

Villaescusa-Navarro, F., Vogelsberger, M., Viel, M., and Loeb, A. (2011). Neutrino Signatures on the High Transmission Regions of the Lyman-alpha Forest. *ArXiv:1106.2543*.

Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G. F., Nelson, D., and Hernquist, L. (2014). Introducing the Illustris Project: Simulating the coevolution of dark and visible matter in the Universe. *ArXiv e-prints*.

Zwicky, F. (1937). On the Masses of Nebulae and of Clusters of Nebulae. *apj*, 86:217.