The Lyman Alpha Forest 1D Power Spectrum from the Baryon Oscillation Spectroscopic Survey

Vaishali Bhardwaj

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Matthew McQuinn, Chair

Scott Anderson

Patrick McDonald

Program Authorized to Offer Degree:
Department of Astronomy

University of Washington

Abstract

The Lyman Alpha Forest 1D Power Spectrum from the Baryon Oscillation Spectroscopic Survey

Vaishali Bhardwaj

Chair of the Supervisory Committee:
Professor Matthew McQuinn
Department of Astronomy

We measure the one dimensional power spectrum of the transmitted flux in the Lyα forest using 16,331 quasar spectra from the twelfth Data Release (DR12) of the Baryon Oscillation Spectroscopic Survey (BOSS). An analysis of this caliber requires a thorough understanding of the data, and the majority of this thesis is aimed at accurately determining the contribution of quasar continuum, sky, and throughput of the quasar spectra. We develop realistic mock spectra to test our code and our power spectrum measurement. A trustworthy measurement of the one dimensional power spectrum also requires removing the contamination from metal absorption in the forest. We account for this and subtract it from the total power measured. We compare our final results to previous results and check for consistency in our measurement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to start off by thanking Patrick McDonald for his guidance throughout this thesis. Although we were not in the same city for much of the time, he was always available to contact through email or by phone. His grasp of Cosmology is awe-inspiring and I feel fortunate to have some of that knowledge passed on to me.

As I have chosen an unusual path with an advisor at a different university than the University of Washington (UW), this could have proven to be a problem to the Department of Astronomy. I am grateful that not only did the department happily accept an unconventional situation, but Scott Anderson stepped up to provide me with the support and guidance that I needed as a younger graduate student.

Since then, the fortune has not subsided as Matt McQuinn recently came to the UW Department of Astronomy and volunteered to advise me even though I was well into my thesis work. Matt's ability to explain complex concepts of Cosmology proved extremely helpful. I enjoyed our discussions about Cosmology, the IGM, and sometimes food in Berkeley!

Since I began graduate school, Andreu Font-Ribera has gone from a graduate student himself, to a postdoc. As he is not related to my thesis in any official capacity, it is surprising how much he has invested in helping me get to this point today. Out of his own volition, he has been available to discuss all aspects of my thesis, whether it was explaining a Cosmological concept, reading my thesis draft at various stages, or Skype-ing with me with the camera pointed at the white board to work through a complex calculation. It is rare to find such benevolent people and I thoroughly appreciate Andreu's guidance and friendship.

I would like to thank the BOSS collaboration for not only providing me with the data for my thesis, but for creating a fun atmosphere in which to work. In particular, Anze Slosar has been there along the way, and does not cease to make me laugh with his humorous work

emails.

Thanking everyone individually will be difficult so I would like to thank the UW Astronomy Faculty for creating a healthy environment in the Astronomy department, where even an undergraduate can feel comfortable interacting with a faculty member. It sounds like it should be a given, but it is not at a lot of places. The graduate students have made this experience fun along the way, and I will miss them dearly.

One person in particular has been an integral part of my career in Astronomy. David Schlegel was my first advisor at UC Berkeley, and has since then been a mentor figure. Even after leaving Berkeley, he still made sure to not only keep in touch, but gave me multiple opportunities to work with the BOSS collaboration. I have never felt more capable of accomplishing great things than in his presence, as his faith in my abilities is unfaltering.

Last, I will thank my family. It is impossible to put into words what I am thanking them for, as the list is endless. It is obvious that they have been supportive, but in particular, they provided me with the confidence that in the worst of situations, I always have their love and care. The roughest of times are manageable when you know that, at the end of the day, you have the support of your family.

DEDICATION

To my parents, who have shown me the pleasure that comes from curiosity and individual thought.

Chapter 1

INTRODUCTION

It has always been the utmost desire of humanity to understand our origins. Whether it was the ancient Mayans' attempt to create an accurate calendar or Charles Darwin with his theory of evolution, curiosity about how our world has come to be has been the common link throughout history. Curiosity is only half the battle. Zora Neal Hurston once said, "Research is formalized curiosity. It is poking and prying with a purpose." It is through research that I aim to satiate a part of this endless curiosity regarding the Universe. The scientific method provides robust means to make incremental progress towards the collective knowledge of mankind and I apply this method to the study of Cosmology.

The field of cosmology aims to understand various aspects of the Universe, such as its origin, constituents, and evolution. As astronomers, we are tasked to study the foundations of our Universe by observing astronomical objects and their motions, to push the limits of our understanding of the cosmos by whatever light reaches us, without the advantage of conducting experiments in laboratories. We are fortunate that the nature of the Universe allows us to look back in time, giving us snapshots of the Universe at different stages in its evolution.

Considering the overall picture, one could conclude that our current understanding of the Universe is limited. It is now known that the Universe consists of approximately 68% dark energy, 27% dark matter and 5% ordinary matter, but we are still unaware of what exactly dark energy and dark matter are. Even of the 5% ordinary matter, much is yet to be understood. But, the more that is unknown, the bigger the opportunity for discovery. It may seem disappointing that we do not understand the majority of constituents of the Universe, but in this case, the numbers do not do justice to the current situation. It is truly

remarkable how much has been uncovered in the field of Cosmology through the resources available, such as indirect measurements of unseen phenomena. To fully appreciate how far the field has come and ultimately, the relevance of this thesis, it is necessary to familiarize oneself with the current paradigm of cosmology.

1.1 A Brief History of Cosmology

Our current view of the Universe consists of a hot Big Bang, with a brief period of exponential growth known as inflation, followed by the creation and evolution of large scale structure (LSS) in an expanding Universe. We can trace the origin of cosmology as a subject of study to the early 1900s with Albert Einstein's General Relativity and Alexander Friedmann's equations. Employing the assumption that the Universe is homogeneous and isotropic, or the Cosmological principle, these equations predict the expansion of space. Yet, the discoveries made by Edwin Hubble in 1929 mark the start of observational cosmology as we know it. Edwin Hubble not only showed that controversial observations of spiral nebulae were, in fact, galaxies other than our own, but by measuring radial velocities of galaxies, he revealed a correlation between recessional velocity and distance to the galaxy, the Hubble Law [Hubble (1929)]. The cosmological implication of this result is an expanding universe. An outcome of this recessional velocity is that the wavelength of light from distant objects is Doppler shifted to longer, redder wavelengths. This redshift, as it is called, gives us a measurement of distance through the Hubble Law and due to the constant speed of light, also provides us with a measurement of time.

To further this idea, Georges Lemaître proposed that projecting an expanding universe back in time would lead to all the mass of the Universe confined to one single point, the first allusion to the idea of a "Big Bang" [Lemaître (1927)]. The Big Bang model states that all the matter in the Universe was once in a hot and dense state and has been expanding ever since. Since the development of the theory, it has come to be the accepted theory of origin by corroboration from many observations.

In addition to this evidence of the expanding Universe, the Big Bang model is able to

explain the abundance of elementary elements in the Universe through primordial Big Bang Nucleosynthesis (BBN) [Gamow (1948)]. The abundances of light elements (namely, helium, deuterium, lithium) were thought to have been created in the interiors of stars. This idea was unable to predict the high abundance of helium, approximately 25%, that we observe. BBN predicts that the creation of these elements must occur in a hot, dense environment, only possible at a time very soon after the Big Bang.

The detection of cosmic microwave background (CMB) radiation in 1964 by Arno Penzias and Robert Wilson [Penzias & Wilson (1965)] solidified the Big Bang Theory into universal acceptance. In the early Universe, photons and baryons were coupled in a plasma. As the Universe expands and cools, electrons are captured by ions to create atoms. The Universe becomes transparent and the photons can free stream. The photons from this time have been propogating ever since, thus reducing their energy due to the expansion of space. We detect this radiation as a thermal blackbody with a temperature of 2.726 K in the microwave region of the electromagnetic spectrum.

A more complete picture of the origin and makeup of the Universe has been in constant development since the start of the field of cosmology. In 1933, Fritz Zwicky noticed that the orbital velocities of galaxies in the Coma Cluster corresponded to a total mass that was much higher than that predicted by its luminosity. He postulated that there was unseen matter, *dark matter*, providing this additional mass [Zwicky (1937)]. Dark matter was accepted as one of the main constituents in our Universe in the 1970's when Vera Rubin measured the velocities of stars in galaxies and found a similar result [Rubin & Ford (1970)]. The stars' velocities were much higher than predicted by the mass of the luminous matter. Therefore, the only explanation was the existence of dark matter, which interacts only gravitationally and extremely weakly with electromagnetic radiation. Big Bang models began including this component, called Cold Dark Matter (CDM), where "cold" refers to the low speeds of dark matter. In the CDM model, large scale structure formed from gravitational collapse of small fluctuations in the early Universe.

In the 1980's, Alan Guth and Alexei Starobinsky independently proposed the idea of

a short-lived phase of exponential expansion of the Universe, an inflationary epoch [Guth (1981); Starobinsky (1982)]. Quantum fluctuations grew rapidly in this inflationary phase becoming the small fluctuations that formed large scale structure we see today. This sudden increase of many orders of magnitude also provides an understanding of why the Universe is homogeneous and isotropic.

Although the CMB had been detected in 1964, CMB measurements from COBE in 1992 showed that, although nearly uniform, the CMB contains tiny anisotropies corresponding to inhomogeneities in the Universe. It is from these anisotropies, one part in 100,000 imprinted by inflation, that large scale structure forms [Smoot et al. (1992)]. This is consistent with the predictions of inflation, and these results are considered to be paramount to the acceptance of the inflation theory. Fig.1.1 shows the most recent results of the CMB anisotropies from *Planck* [Planck Collaboration et al. (2014)].

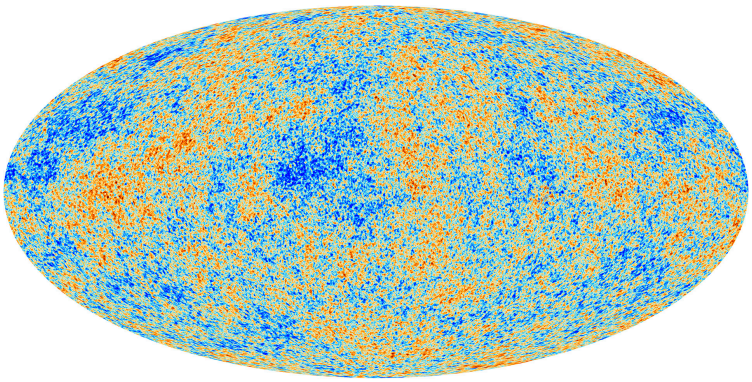

Figure 1.1 Cosmic Microwave Background from *Planck* showing anisotropies in temperature [Planck Collaboration et al. (2014)].

Cosmology has been an ever advancing field, as we are continually making exciting new

discoveries. One such monumental discovery, less than twenty years ago, came from observations of Type 1a supernovae which revealed that not only was the Universe expanding, the expansion rate was accelerating, as depicted in Fig.1.2 [Perlmutter et al. (1999); Riess et al. (1998)]. The existing model of the Universe could not account for this acceleration, and a non-zero cosmological constant was added to Einstein's equation. Current potential theories include a modified understanding of General Relativity or a negative pressure component, *dark energy*.

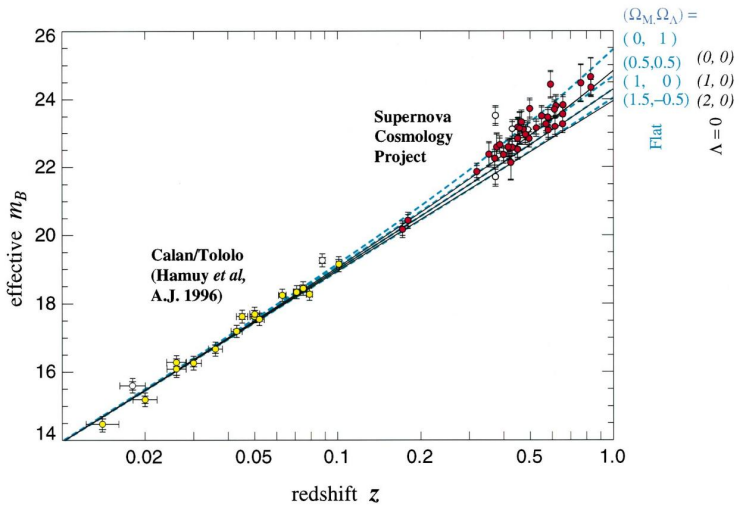

Figure 1.2 Results from Perlmutter et al. (1999) showing the accelerated expansion of the Universe as seen by the decrease in brightness at high redshift, measured by high redshift Type 1A supernovae.

1.2 The Intergalactic Medium

One tends to think that the most of the matter in the Universe lies in galaxies, which contain stars, planets, gas, and dust. It turns out that the majority of matter in the Universe lies in the voids between galaxies, known as the intergalactic medium (IGM). One can imagine that

looking far back in the history of the Universe, the IGM is the reservoir of gas from which galaxies formed. Understanding the makeup of this region is paramount to completing the picture of the Universe's history and structure formation.

We can follow the evolution of the IGM by understanding its ionization history. As the Universe cooled after the Big Bang, it became cool enough for electrons and protons to combine to form neutral atoms, the majority of which are Hydrogen. It was only after the first stars and galaxies were formed that this region was ionized during the *reionization epoch* of the Universe. Since then, quasars have played a role in also providing a ultraviolet (UV) background to ionize the neutral material in the region. At the present day, with more and more astrophysical processes occurring, much of the IGM is ionized.

Although this is the current understanding of the IGM, studying the density, temperature, and makeup of the IGM is not trivial. Fortunately, optical data of the the Lyman Alpha Forest in quasar spectra provides us with means to probe the IGM in the redshift range $2 < z < 4$.

1.3 The Lyα Forest in Quasar Spectra

We are now aware that the Universe is made up of ordinary matter, dark matter, and dark energy, but much is still not well understood. Studying the large scale structure through different astrophysical tracers is one of the ways to shed light on these subjects. Historically, galaxies have been used to map where the mass of the Universe lies. This is extremely useful in understanding where dense dark matter halos reside and their distribution, but there are limits to the scale at which these galaxies probe the Universe. Recent work with the Lyman Alpha Forest has allowed probing much smaller scales of the Universe, and a lower density region, the IGM, which contains the majority of the mass of the Universe.

The Lyman Alpha Forest refers to the absorption of the Lyman Alpha transition from the n = 1 to n = 2 orbital state of neutral hydrogen observed in high redshift quasar spectra. Although it was unknown for some time what exactly a quasar is, we now know that a quasar is a galaxy with a very high luminosity due to an accreting supermassive blackhole.

Due to the expansion of the universe, light from the quasar is redshifted on its path to the observer. Light that is redshifted to the specific wavelength of the Lyman alpha transition will be absorbed by neutral hydrogen in the intergalactic medium (IGM). In 1965, James Gunn and Bruce Peterson predicted that in an expanding universe homogeneously filled with gas, an absorption trough blueward of the Lyα emission line would be present due to the absorption of the redshifted Lyα line [Gunn & Peterson (1965)]. Around the same time, John Bahcall and Edwin Salpeter predicted that discrete lines in the spectra of quasars would be indicative of absorption from intervening clouds of neutral hydrogen in the line of sight [Bahcall & Salpeter (1965)]. These absorption features, named the *Lyman Alpha Forest* (Lyα forest), were discovered in the spectra of a single quasar in 1971 by Roger Lynds [Lynds (1971)].

Since the discovery of the Lyα forest, studies have shown that the prevailing picture for this region is a smooth, low density environment with absorption due to neutral hydrogen as opposed to discrete clouds. This fluctuating, photoionized gas in the IGM implies that the ordinary matter traces the underlying dark matter distribution [Cen et al. (1994); Zhang et al. (1995); Hernquist et al. (1996); Theuns et al. (1998)]. Fig. 1.3 shows a very high resolution quasar spectra at low redshift with no Lyα forest absorption visible, and a second spectra at high redshift with considerable Lyα forest contribution. Some of the absorption in this region is by UV transitions of metals as opposed to neutral hydrogen. The origin of these metals is enrichment from galactic winds. Regions of column densities higher than $\sim 10^{19}$, which may be the outer regions of galaxies, exhibit an absorption feature that is dominated by Lorentzian wings, and are named *Damped Lyman Alpha* (DLA) systems.

1.4 Observational Constraints on Cosmological Parameters

The aim of cosmology is to paint a picture of the origin and evolution of the Universe. As cosmologists, we observe the Universe at snapshots in its history and provide a model to describe the evolution from one stage to the next. Our model should be consistent through all epochs and ideally, a few independent cosmological parameters should determine the

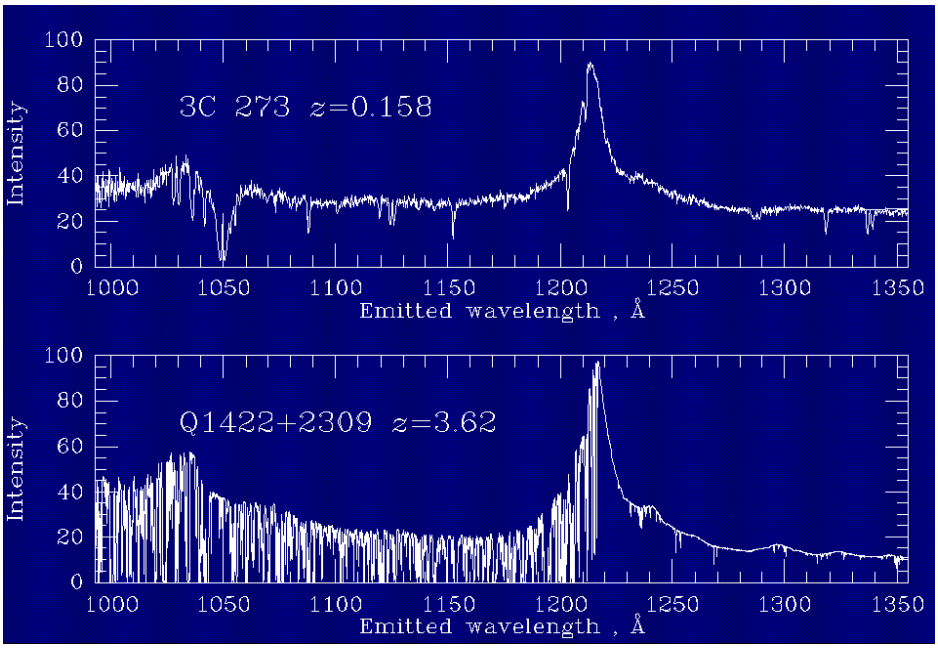

Figure 1.3 Two high resolution spectra showing low and high redshift quasars, respectively, with their Lyα absorption. The low redshift quasar shows much less absorption as there is not much neutral hydrogen that the light from the quasar passes through, as compared with the high redshift quasar.

overall evolution of the Universe.

The best current model for the Universe is the ΛCDM model, where Λ refers to the dark energy component at late time, and CDM is Cold Dark Matter, referring to the non relativistic nature of dark matter. This model consists of six independent parameters and corresponds to a spatially flat Universe. The primordial seeds of structure are Gaussian with a near scale-invariant spectrum of adiabatic fluctuations. As mentioned above, a model of the Universe should be powerful enough to describe all epochs of its origin and evolution. To test whether our model works, we need to observe the Universe at various stages in its history.

Different tracers probe different regions in redshift and on varying length scales. Combining observations from a variety of tracers has been extremely powerful in constraining cosmological parameters. The ΛCDM model is the current accepted model because it is able to describe observations including CMB, galaxy clustering, Baryon Acoustic Oscillations (BAO), and Type 1a supernovae in varying redshift ranges and scales.

One of the most powerful probes of cosmology, corresponding to the early epoch of the Universe, is the CMB. The CMB provides us with observations 377,000 years after the Big Bang. Detecting anisotropies in the temperature profile of the CMB is used to constrain the primordial power spectrum of density fluctuations which, in turn, grow to the structures present today. Therefore, using other tracers to understand the density distribution of the Universe at different stages can be related back to the primordial power spectrum. There are limits to this, as small scales undergo strong nonlinear evolution which erases the information of the primordial power spectrum. This is the case when using galaxy clustering or weak lensing as tracers. But on larger scales, these tracers provide a powerful method to break degeneracies that exist when using the CMB alone to constrain parameters. Although some constraints can be made on cosmology using observations of one tracer, the power of observational cosmology lies in breaking degeneracies by combining results.

One example of such a degeneracy, and the power of combining multiple tracers, is seen in current results from combining *Planck* CMB results with BAO results from the Baryon

Oscillation Spectroscopic Survey (BOSS). BAO provide a comoving standard candle from which we can constrain the expansion history of the Universe. The early Universe was extremely hot and dense such that photons, baryons, and electrons were tightly coupled in a plasma. Opposing forces of gravity and outward pressure from baryons created sound waves in the plasma. As the Universe expanded and cooled, baryons combined with electrons to create atoms, in what is called the recombination epoch. Photons were now able to free stream, leaving an imprint of baryons at the characteristic scale of the sound horizon at the moment of decoupling. This imprint is seen in the power spectrum of CMB anisotropies at high redshift but is also seen in galaxies at later time. The first measurement of BAO was done by Eisenstein et al. (2005) using SDSS galaxies and has since been measured to higher precision in Anderson et al. (2014) and Anderson et al. (2012) using BOSS galaxies. Figure 1.4 exemplifies how combining tracers can tighten constraints on cosmological parameters, in this case, the Hubble expansion rate, and the angular diameter distance.

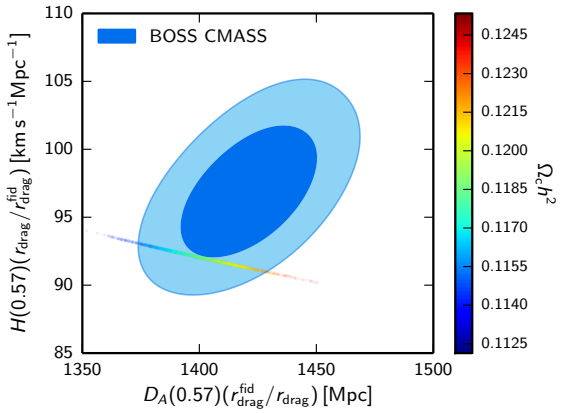

Figure 1.4 Combining results from *Planck* and BOSS BAO provide tight constraints on the Hubble expansion rate and the angular diameter distance at $z = 0.57$. The contours shown are 68% and 95% contours from Anderson et al. (2012), while *Planck* results are shown colored by their value of $\Omega_c h^2$ [Planck Collaboration et al. (2014)].

Combining tracers proves powerful to constrain cosmology, but there are limits to the

scales and redshift range that certain probes can provide. At the redshift range where galaxies are observable, nonlinear evolution is dominant rendering small scale information difficult to relate to primordial information. Using the Lyα forest can be particularly useful here, as the high redshift range probes an epoch where nonlinear evolution is not significant. Although early work with the Lyα forest used few high resolution spectra, the advent of spectroscopic surveys such as the Sloan Digital Sky Survey (SDSS) have made it possible to study large scale structure in the Universe using the Lyman Alpha Forest. BAO from galaxies can only be done at $z < 1$, but recent work by Slosar et al. (2013) and Busca et al. (2013) have measured BAO from the Lyα forest three-dimensional correlation function, extending the measurement of expansion out to much higher redshift, $z \sim 2.4$. An exciting outcome of this measurement is that this probed the region of the Universe during its deceleration phase as opposed to other tracers that can only detect regions where the dark energy component dominates as in Fig. 1.5.

Measuring the BAO uses the Lyα forest to understand large scales of the Universe, but another appeal of the Lyα forest is its ability to probe the smallest scales. Measurements of the one dimensional power spectrum take advantage of this property of the Lyα forest to constrain the smallest of scales in Cosmology. Early work by McDonald et al. (2006) used the SDSS-I spectroscopic data of $\sim$ 3000 quasars to measure the one-dimensional power spectrum of the Lyα forest. Recent work by Palanque-Delabrouille et al. (2013) measured the power spectrum using $\sim 13,000$ BOSS spectra. Combining our knowledge of large scales all the way out to the small scales as probed from the Lyα forest can constrain a variety of cosmological parameters. A unique constraint that the Lyα forest can provide is an upper limit on massive neutrinos. At the smallest of scales, neutrinos suppress power in dark matter clustering by free streaming, slowing the growth of CDM structure.

1.5 Thesis Project

The goal of this thesis is to make progress toward answering some basic questions regarding the Universe. Observing large-scale density fluctuations in the Universe is one of the best

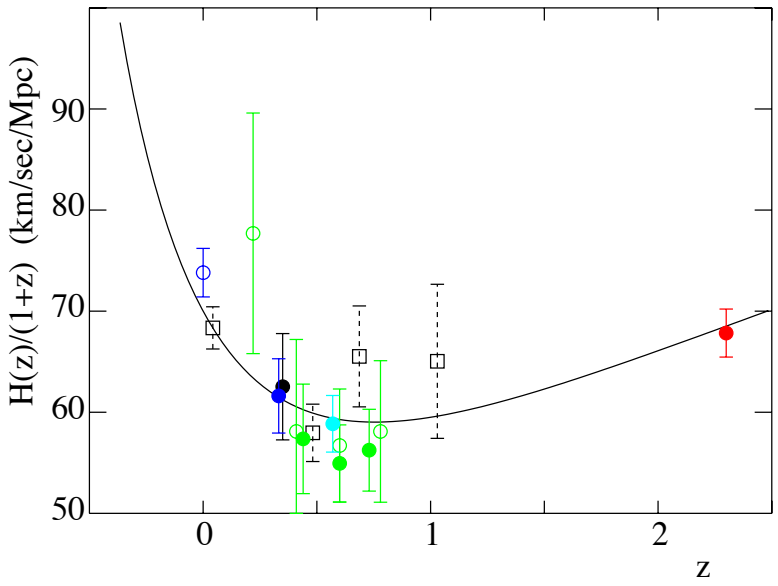

Figure 1.5 Results from Lyα forest BAO studies from Busca et al. (2013) compute the rate of expansion at a time before dark energy began to dominate, depicting a decreasing expansion rate of the Universe. The red data point is derived from BAO measurements of the Lyα forest using BOSS quasar spectra.

ways to approach these questions. Statistics of the observable density fluctuations can be used to infer statistics of the small initial perturbations from which they grew, and in turn to understand the physics of the very early Universe. Observing the evolution of large-scale structure (LSS) over time tells us about the present matter content of the Universe and the dynamical rules its evolution follows.

There are only a handful of good tracers of LSS, including the cosmic microwave background, galaxies, galaxy clusters, gravitational lensing, and the Lyα forest. This thesis project is specifically related to the Lyα forest as described above. The Lyα forest provides an approximate spatial map of the density field along the line of sight to the quasar. The bulk of the statistical power of the Lyα forest lies in the redshift range $2 \lesssim z \lesssim 4$, and it probes spatial scales down to a few tens of kiloparsecs.

The unique capability of the Lyα forest is that it probes an intermediate redshift range between the CMB at high redshift and other traces at lower redshift, such as galaxies. Relative to other tracers, it also probes relatively small scales while they are still linear, both because it is at higher z where structure is more linear and because strongly non-linear structures produce saturated absorption and therefore, Lyα forest statistics have little sensitivity to them. In contrast, other methods tend to be dominated by the highly non-linear peaks. The near linearity is an important aspect because the field is still tightly connected to the background evolution and initial conditions laid down in the early universe, and because it allows first-principle theoretical calculations.

Our target statistic to measure will be the one dimensional power spectrum of absoprtion fluctuations along single lines of sight. This is effectively the Fourier transform of the correlation function of a density field, $\delta(\lambda)$,

$$\delta(\lambda) = \frac{e^{-\tau(\lambda)}}{\langle e^{-\tau} \rangle} - 1 \tag{1.1}$$

where τ is the optical depth to Lyα absorption.

The first measurement of the power spectrum was done by Croft et al. (1998) with one quasar spectra. Since then, McDonald et al. (2006) measured this from $\sim$ 3000 spectra,

as shown in Fig.1.6. Recent work done by Palanque-Delabrouille et al. (2013) measured the power spectrum for a subset of the full BOSS data set, $\sim 13,000$ quasars, improving constraints on cosmological parameters by a factor of 2-3. Since the work of Palanque-Delabrouille et al. (2013), BOSS has observed $> 160,000$ spectra, which will provide us with an improvement in statistical errors on our measurement. Our analysis has been independent from that of others within the collaboration and as such differs in methodology. Since this analysis is also performed on BOSS spectra, we use the parameterization of $P(z,k)$ by Palanque-Delabrouille et al. (2013) to build our mock spectra as will be detailed below.

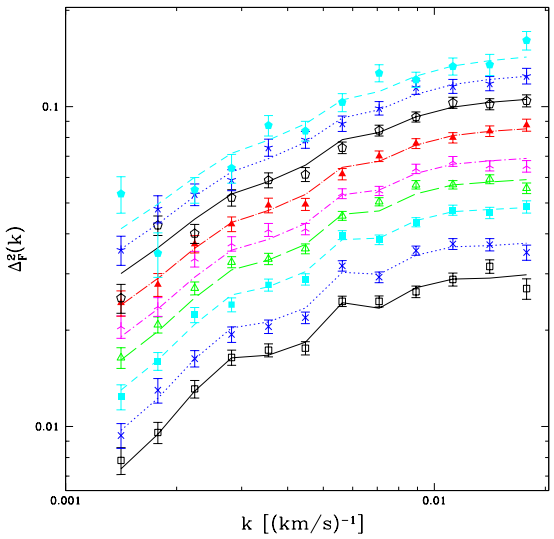

Figure 1.6 Results from McDonald et al. (2006) show the 1D Power Spectrum as measured from ~ 3000 quasar spectra from SDSS-I. Redshift bins increase in increments of $\Delta z = 0.02$, from the bottom at $z = 2.2$ to the top at $z = 3.8$.

Pushing the limits of our measurement to the smallest of scales requires a very clear understanding of the systematics in our data. The BOSS data reduction pipeline provides a decent measurement of various properties such as noise, sky, and throughput, but we will need to assess the accuracy of these values.

Our measurement is robust for the range of Fourier modes between $k = 0.0004 -$

0.02 s km^{-1}, where our measurement is in units of km s^{-1}, and we have used velocity units as opposed to observed wavelength with the relationship $\Delta v = c\Delta \ln(\lambda)$. This is an acceptable approximation since we do not measure power on scales large enough where the discrepancy will be considerable. We remove the contribution of metal absorption, which is measurable for transitions with $\lambda < 1300$ Å outside of the Lyα forest. For transitions $\lambda < \lambda_\alpha$, though, we cannot distinguish the difference between Hydrogen Lyα absorption and absorption due to metals, therefore we leave them as is to be modeled by those using the measurement for cosmological purposes.

Although this is outside the scope of this thesis, the Lyα forest can be combined with other probes to constrain cosmological parameters. Combining with the CMB results, we can constrain the scale dependence of the spectral index, α_s. Combining with simulations, we can constrain the rms amplitude of perturbations in 8 h^{-1}Mpc radius spheres, σ_8. The small scale power measurement combined with large scale measurements can be used to provide a bound on the sum of neutrino masses, Σm_ν. Over the redshift range $2 \lesssim z \lesssim 6$, we can measure the strength of the ionizing background radiation field, produced by stars and quasars. Finally, we can measure a temperature-density relation of the IGM over the redshift range $2 \lesssim z \lesssim 4$. The statistical improvement from the larger data set will improve linear power spectrum measurements and all constraints derived from them. Note that this thesis project presents the observational measurement of the Lyα forest flux power spectum, devoid of any cosmological interpretation of the result. It does not aim to constrain cosmological parameters but provides a measurement that can be used to do so. Fig. 1.7 shows the result of combining the information from various tracers and its ability to constrain the linear matter power spectrum through a range of scales.

We discuss the BOSS quasar spectra that makes up our data sample in Chapter 2.2. In Chapter 3, we describe our data model,. In Chapter 4, we outline the core elements of our analysis and detail the statistic we will be measuring, namely the one dimensional power spectrum. We describe our methods for building mock data sets in 5. In Chapter 6, we show the results of the power spectrum measurement and tests conducted. We conclude with some

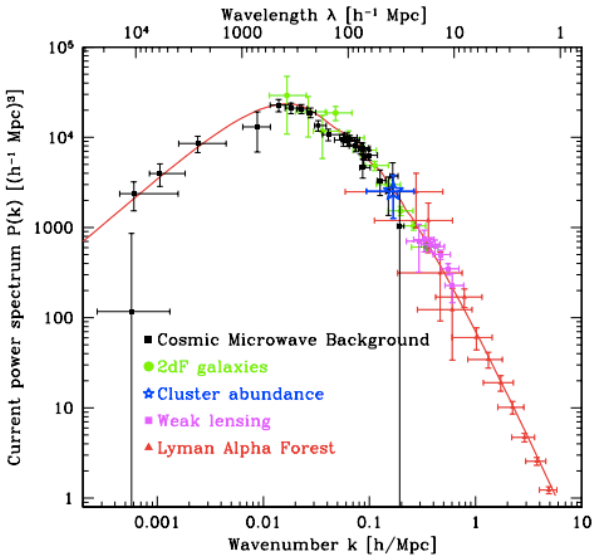

Figure 1.7 We see the power of combining the measurement of the power spectrum from various tracers. This figure by Max Tegmark shows the linear matter power spectrum using the CMB, galaxies, weak lensing, and the Lyman Alpha forest, from large scales to small scales, respectively.

remarks in Chapter 7.

Chapter 2

DATA SELECTION AND PREPARATION

For this thesis project, we construct a data set using quasar spectra from the Baryon Oscillation Spectroscopic Survey (BOSS) Data Release 12 [Alam et al. (2015)]. BOSS targets quasars specifically to be used for Lyα forest studies, and we will describe data from BOSS and the quasar catalog in §2.1. In §2.2, we discuss the cuts we apply to create our data sample.

2.1 BOSS Observation and Quasar Sample

The Baryon Oscillation Spectroscopic Survey [Dawson et al. (2013)] is the largest of four surveys in the Sloan Digital Sky Survey III [SDSS-III, Eisenstein et al. (2011)] using the 2.5-m Sloan telescope [Gunn et al. (2006)]. Spanning from 2009-2014, the main science goal of BOSS is to measure cosmic distances through baryon acoustic oscillations (BAO), the observed clustering of matter due to the imprint of sound waves from the early Universe. Measurement of the BAO scale, namely the comoving size of the imprint of sound waves at different redshifts, can be used to characterize the expansion history of the Universe and dark energy. The first measurements of the BAO scale were from spectroscopic data of galaxies from the Sloan Digital Sky Survey [Eisenstein et al. (2005), Cole et al. (2005)]. A measurement such as this requires a high volume of data, since the distances in question are so large. In addition, increasing the statistical precision on constraints of cosmological parameters requires even more data. To meet these goals, BOSS obtained the redshift and spectra of over 1.35 million galaxies, 300,000 quasars, of which $\sim 160,000$ are in the redshift range $2.2 < z < 3.5$ to probe the Lyman Alpha Forest. In addition to the imaging from SDSS-II [Abazajian et al. (2009)], SDSS-III imaged an additional 3000 square degrees, using a drift

scanning imaging camera [Gunn et al. (1998)] in five photometric bandpasses [Fukugita et al. (1996)]. The total photometric data, released in Data Release 8 [Aihara et al. (2011)], covers over 14,000 square degrees of the sky.

The survey uses two double spectrographs with a wavelength range of 3600 - 10000 Å fed by 1000 optical fibers, each subtending 2" on the sky to obtain the spectroscopic data of its targets. Objects are observed in multiple 900 second exposures. The resulting spectra have a resolving power of 1500-2600 [Smee et al. (2013)]. As compared to SDSS-I, which was used for previous Lyα forest work, the BOSS spectrograph has a longer wavelength range and higher sensitivity in the blue end of the spectrograph. This was designed specifically to target as much of the Lyα forest as possible. Pushing out to shorter wavelengths ensures that we can probe the Lyα forest to the lowest redshifts possible in the optical range. Quasar candidates are selected by a variety of algorithms as described in Ross et al. (2012).

The BOSS data reduction pipeline performs wavelength calibration, sky subtraction, flux calibration, and uses a local spline interpolation to coadd the four to seven 900-s exposures onto a fixed grid of pixel width $\Delta\log_{10}(\lambda) = 10^{-4}$ [Guy (2015, in prep)]. This does not guarantee that the noise estimate provided by the pipeline is uncorrelated between pixels. This is an issue for our analysis, as the pipeline does not provide a covariance matrix to describe the correlation between the noise, such that we would be misinterpreting the noise of the spectra in our analysis. In Chapter 3, we will discuss our methodology to tackle this problem.

Fig. 2.1 shows a schematic of the telescope. Note that the extensive design of this telescope, although necessary to achieve the desired result, shows how complex the effects of the telescope will be on the light coming from the quasar. For example, light is split into two, at 6000 $\mathring{A}$, and goes through separate grisms before hitting either a red or a blue CCD, which both exhibit different throughput properties. The coadded spectra then, in addition to combining multiple exposures, also combines the red and blue spectra at 6000 $\mathring{A}$. One can imagine that truly understanding the data is not a trivial task, and much of this thesis is aimed at quantifying the effects of throughput on the quasar spectra.

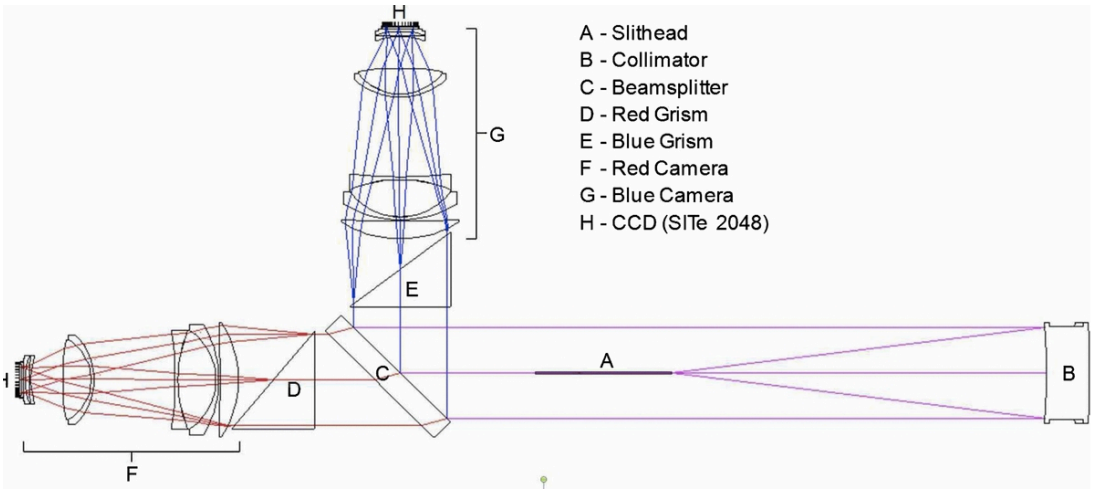

Figure 2.1 Schematic of the SDSS telescope with BOSS upgrade spectrographs from Smee et al. (2013). Many aspects of this analysis aim to understand particular features of the data which are a result of changes made to the hardware when compared to SDSS-I. Namely, there is a new spectrograph, plates with 1000 fibers of smaller radius than the 600 fibers of SDSS-I, and new CCDs.

Spectral classification and redshifts are provided by a fitting algorithm from the data reduction pipeline [Bolton et al. (2012)], but all quasar targets are visually inspected and confirmed thereafter in the DR12 Quasar Catalog provided by the French Participation Group [FPG, Pâris et al. (2014)]. An additional catalog for DR12, specifically an extension of the DR9 catalog [Noterdaeme et al. (2012)], which identifies Damped Lyman Alpha (DLA) systems in quasar spectra, is provided by Pasquier Noterdaeme to the internal SDSS-III collaboration. As will be discussed further in §2.2, we will need to remove quasars with DLA systems for a clean data sample.

Each BOSS plate has a 7 deg^2 field of view and allots 80 fibers to observe sky. The data reduction pipeline creates a Principal Component Analysis (PCA) sky template from the sky spectra on each plate to describe the contribution of sky, $S(\lambda)$, for a given target. The template is interpolated depending on the fiber's position on the plate. Resolution of the pixels is calculated during the wavelength calibration step. Lamps mounted around the telescope whose spectra is known are used to calibrate the wavelength of each exposure, at which time a measurement of resolution of each pixel is also recorded. Note that the resolution is calibrated exposure by exposure. Therefore, in the coadded spectra, the pipeline computes an average per pixel for the resolution. Spectral resolution plays a large role in the power spectrum measurement as the Gaussian resolution element dampens power, which is more noticeable at smaller scales. Our measurement will be dependent on whether this resolution is well understood.

As the raw data is recorded in photon counts, spectrophotometric standard stars are used to relate photon counts to flux units. 20 standard stars are observed on each plate and each exposure is flux calibrated individually. Unfortunately, the spectrophotometric calibration for BOSS spectra, which corresponds to the total integrated flux as compared to photometric magnitudes, is fraught with issues. The spectrophotometric errors were introduced in BOSS as a result of changes in the fiber size and plate design. In order to improve S/N of the Lyα region of high redshift quasars, fiber holes for quasar targets are offset as compared to the fibers of spectrophotometric standard stars which are used for calibration. In conjunction,

the smaller, 2 arcsecond fiber size (as compared to the SDSS fiber size of 3 arcseconds) allows the atmospheric differential refraction (ADR) to be more pronounced. The ADR is wavelength dependent and thus, more prominent at the blue end of the spectra, and depends on the observing angle. In addition, if the guiding of the telescope is imperfect, it is possible that all the light from the quasar is not entering the fiber [Margala et al. (2015)]. Figure 2.2 shows the quasar spectra of the same object observed at different times. As quasars exhibit variability in their overall brightness, we cannot use photometric data to normalize the spectra. We discuss this issue further in Chapter 3.

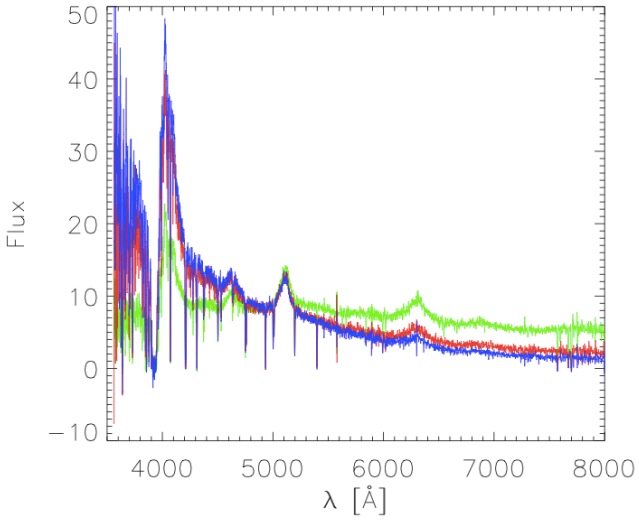

Figure 2.2 Three different observations of the same object, displaying the problems with spectrophotometry in the BOSS quasar spectra. Many aspects contribute to these errors including fiber positioning offsets between quasars and standard stars, smaller fiber size allowing atmospheric differential refraction to disperse light outside of the fiber, and misguiding, allowing the centroid of the target's point spread function (PSF) to drift across the fiber hole.

BOSS targets quasars in the redshift range, $2.2 < z < 3.5$ [Ross et al. (2012)]. specifi-

cally to obtain quasars with a considerable Lyα forest contribution. The low redshift limit corresponds to the limit of our spectrograph on the blue end, namely 3600 Å. The high redshift limit is chosen due to the decrease in quasar number density at $z > 3$ [Osmer (1982), Schmidt et al. (1995), Richards et al. (2006)] Fig.2.3 sheds light on the difficulty with quasar target selection by showing the overlap of the stellar locus and quasar targets. As we note in this plot in color-color space, spectroscopically confirmed stars and quasars with $z > 2.2$ show much overlap. Even after using many different techniques, the success rate of target selection is $\sim 50\%$.

2.2 Data Sample

Our power spectrum analysis will compute the power spectrum at a redshift range of $2 < z < 4$, but we will use BOSS quasars in the full redshift range for our measurement of the intrinsic quasar continuum. To construct our data sample, we first remove quasars in plates that were considered to be of bad quality according to the BOSS data reduction pipeline. We will also mask pixels based on results from the pipeline, which may be due to bad pixels on the CCD, or cosmic rays, etc. In Fig.2.4, we show the distribution of redshifts for quasars in the full BOSS quasar sample which will be used to compute the intrinsic quasar continuum. Our *gold sample*, as we will call it, will be a high signal-to-noise (S/N) quasar sample with S/N > 7. Our Lyα forest power spectrum measurements will be done on quasars with $z > 2$, but we will also need to assess the contamination from metals (namely, SiIV and CIV) which will require quasars with $z > 1.75$. In Fig.2.5, we show the S/N distribution of quasars with $z > 2$ in DR12, where we highlight the region that we use for our *gold sample*, namely S/N > 7.

The French Participation Group (FPG) has provided a quasar catalog in which spectra has been visually inspected. Upon this inspection, flags for Broad Absorption Line (BAL) quasars and those containing Damped Lyman Alpha (DLA) systems have been set, which we subsequently remove in our data samples. Previous work from McDonald et al. (2006) found that the addition of DLA may bias the result, therefore, we choose to remove them in

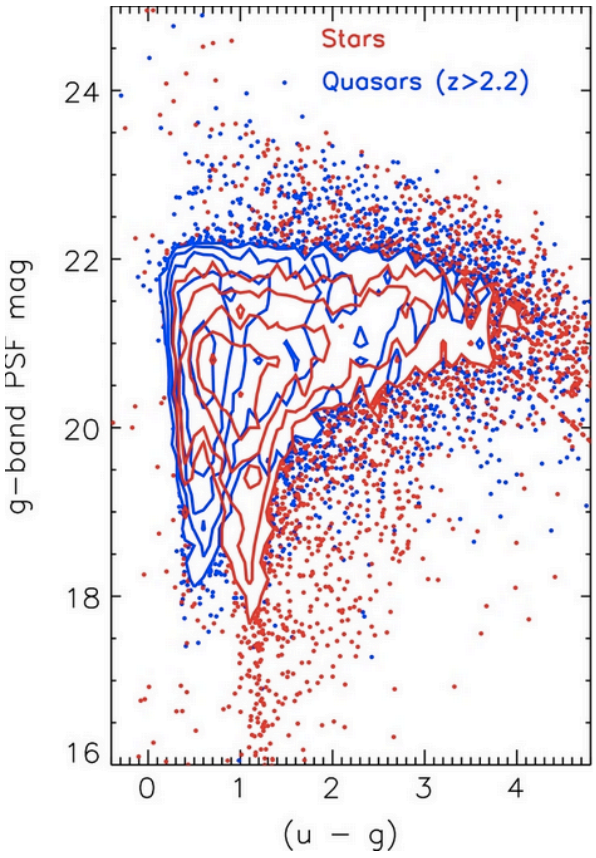

Figure 2.3 Quasars that are in the ideal redshift range for Lyα forest studies are also extremely difficult to target as they overlap with stars in color-color space. We plot u-g vs. g for spectroscopically confirmed stars and quasars with $z > 2.2$ which shows much overlap. Target selection requires sophisticated techniques to distinguish stars from quasars. Even after complex algorithms, the success rate of target selection is $\sim 50\%$.

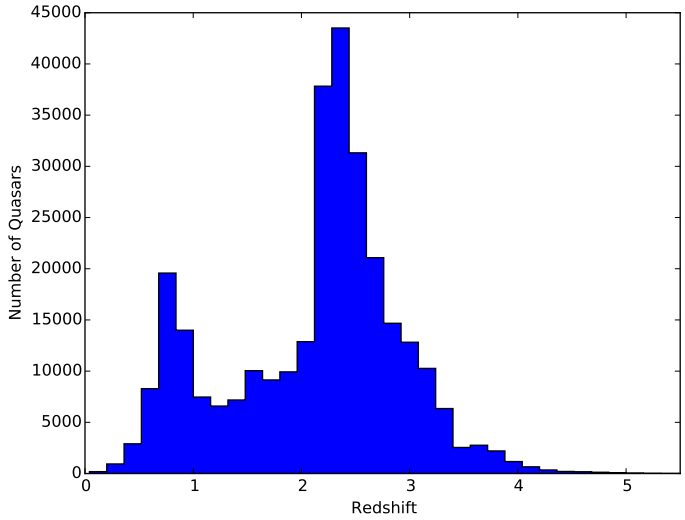

Figure 2.4 Distribution of quasar redshifts in BOSS DR12. The full redshift range will be used for continuum fitting, while only the region with a considerable Lyα forest contribution, namely $z > 2$, will be used for the power spectrum fit.

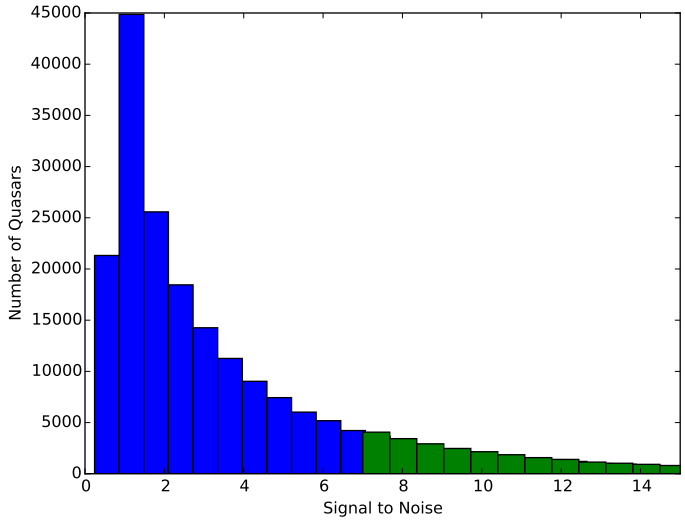

Figure 2.5 Distribution of Signal-to-Noise Ratio of quasars with $z > 2$ in BOSS DR12, highlighting the region in green used in the *gold sample* where S/N > 7.

our sample. In addition to the visual inspection, a more rigorous study on DLA systems has been conducted by Pasquier Noterdaeme which contains many more candidates for quasars with DLA systems. We will assess the effects that DLAs have on our power spectrum in Chapter 6.

Our measurement of the power in the Lyα forest is dependent on our understanding of the underlying quasar continuum. We therefore cut quasars where our parameterization of the quasar continuum does not fit a given quasar spectra to high precision. We quantify this goodness of fit by computing the probability of the χ^2 of the continuum fit. We remove quasars with $P(\chi^2) < 0.01$. There are assumptions that do not allow this probability distribution to be formally accurate such that more than 1% of the data is removed when making the above cut. Since our continuum model does not contain BAL or DLA systems, this statistic is also a method to identify BAL and DLA quasars missed by visual inspection, or those with other unusual properties. As an example, Fig.2.6 shows a quasar which passed the visual inspection but exhibits BAL features. Table 2.1 lists the number of quasars in the sample after each cut. The redshifts of pixels used in our Lyα forest analysis are shown in Fig. 2.7.

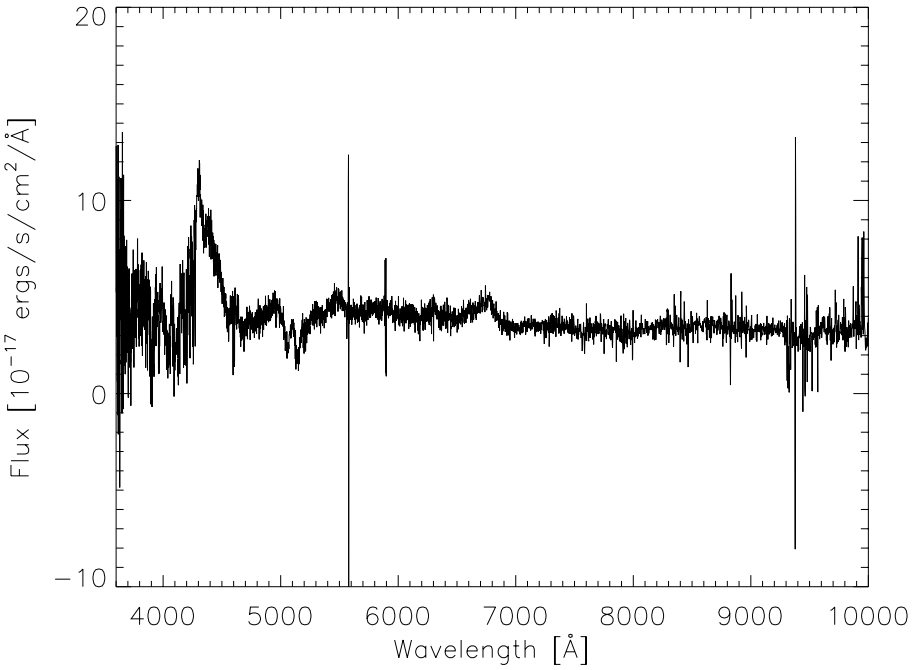

Figure 2.6 Example of an unusual quasar that has been cut due to a bad continuum fit as characterized by $P(\chi^2) < 0.01$. Cutting based on the probability of the χ^2 value of a continuum fit allows to remove quasars which may have passed a quick, visual inspection, but still contains features which make it difficult to characterize the intrinsic quasar continuum.

Total Quasars	297,301
$z > 1.75$	213,710
SNR > 7	35,329
No BAL	29,766
No DLA (visual inspection)	27,875
No DLA	26,506
$P(\chi^2) > 0.01$	22,037
Total $z > 1.75$ Quasars	22,037
Total $z > 2$ Quasars	16,331

Table 2.1 We provide information about our *gold sample* and the number of quasars remaining in the sample after applying cuts. Quasars with $z > 1.75$ are used to measure the power spectrum from the background, while the $z > 2$ sample is used for Lyα forest power spectrum measurement. This is called the *gold sample* throughout the thesis.

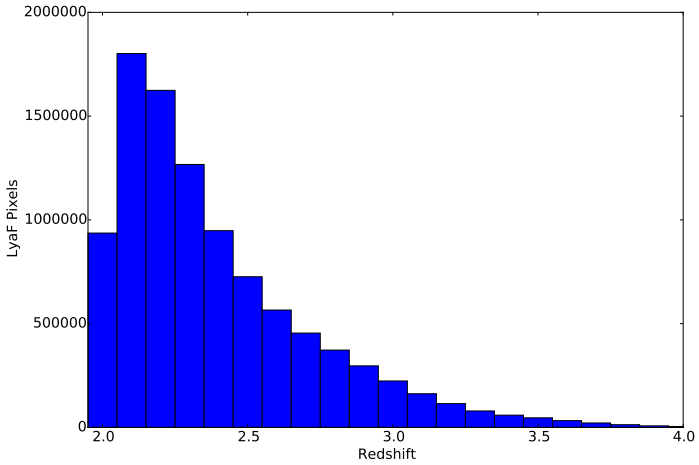

Figure 2.7 Histogram of redshifts for Lyα forest pixels for our *gold sample*. Due to target selection and limits of the spectrograph, quasars with sizeable contribution in the Lyα forest are those with $z > 2$.

Chapter 3

DATA MODEL

The goal of this thesis is to measure the power spectrum of the density fluctuations in the Lyα forest. A major task for this project is to decompose quasar spectra into its constituents, of which one will be the field of density fluctuations in the Lyα forest on which we can perform the power spectrum analysis. We must accurately model the full suite of components that contribute to the spectra, such as the underlying quasar continuum, added flux from the sky, throughput from the telescope and spectrographs, and noise. We could, in principle, use the values for sky, throughput, and noise from the BOSS reduction pipeline, but previous work [McDonald et al. (2006)] showed that the reduction pipeline is not rigorous enough for our needs. Even if we were to trust this, we still must completely understand the contribution of the quasar continuum and the mean transmission of the forest to ultimately return the relevant quantity of interest, the density of fluctuations about the mean transmission. The quantity of interest is the Lyα forest field of density fluctuations,

$$\delta_F(\lambda) = \frac{e^{-\tau(\lambda)}}{\langle e^{-\tau} \rangle} - 1, \tag{3.1}$$

where τ is the Lyα absorption optical depth and we are measuring the density fluctuations about a mean value, which we can write as

$$\bar{F}(z) = \left\langle e^{-\tau} \right\rangle, \tag{3.2}$$

where

$$z = \frac{\lambda}{\lambda_{Ly\alpha}} - 1 \tag{3.3}$$

is the redshift of the gas that would produce Lyα absorption at that wavelength, not the redshift of the quasar. $\lambda_{Ly\alpha}$ is the wavelength of the Lyα absorption line of the $n = 1$ to $n = 2$

energy transition corresponding to 1215.67 Å. The transmitted flux fraction, $F(\lambda) = e^{-\tau}$, is then

$$F(\lambda) = \bar{F}(z)[1 + \delta_F(\lambda)]. \tag{3.4}$$

Without including the effects of the telescope and sky, quasar spectra consists of the flux transmission of the Lyα forest and the intrinsic quasar continuum (where we refer to continuum as the total quasar contribution including emission and absorption lines),

$$Q(\lambda) = C(\lambda)\bar{F}(\lambda)[1 + \delta_F(\lambda)]. \tag{3.5}$$

If we could fully trust the BOSS reduction pipeline, we could model the quasar spectra using Eq. 3.5 and assume the throughput effects and sky subtraction was performed perfectly. We cannot assume this, and need to model corrections to the values predicted by the BOSS reduction pipeline. Therefore, our model to describe the flux at a pixel, i, includes correction factors to both throughput and sky,

$$f_i = C(\lambda_i)T(\lambda_i)[1 + \delta_T(\lambda_i)]\bar{F}(z_i)\,[1 + \delta_{F_i}] + S(\lambda_i)[1 + \delta_{TS}(\lambda)] + N(\lambda_i). \tag{3.6}$$

The term $C(\lambda_i)$ is the quasar continuum, $T(\lambda_i)$ is the throughput of the telescope and spectrograph with a correction factor, $\delta_T(\lambda)$. The term $\bar{F}(z_i)$ is the mean transmitted fraction of the forest pixels, as a function of redshift (where $z_i = \lambda_i/\lambda_{Ly\alpha} - 1$) and $\delta_F(\lambda_i)$ is the fluctuations in the Lyα forest. $S(\lambda_i)$ is the sky flux, with a correction factor, δ_{TS}, where we denote this with the subscript TS to remind the reader that there is a throughput component to this sky correction factor. $N(\lambda_i)$ is the noise. In this approximation, $\langle\delta_F\rangle = 0$ and $\langle N\rangle = 0$, whereas $\langle\delta_T\rangle$ and $\langle\delta_{TS}\rangle$ are not necessarily zero.

This relates to the power spectrum of fluctuations in the following way. In Eq. 3.6, our theoretical value is given by $m_i = C(\lambda_i)T(\lambda_i)\bar{F}(z_i)$ and is defined such that $< f_i >= m_i$. We can then compute the expectation value of two points, f_i and f_j.

$$\langle f_i f_j\rangle = \Big\langle (m_i[1 + \delta_{f_i}] + s_i + n_i)(m_j[1 + \delta_{f_j}] + s_j + n_j)\Big\rangle \tag{3.7}$$

$$= m_i m_j[1 + \xi_{ij}] + \langle s_i s_j\rangle + \langle n_i n_j\rangle \tag{3.8}$$

where

$$\xi_{ij}(r,i) =< \delta_i \delta_j > \tag{3.9}$$

is the flux correlation function, which is the Fourier transform of the power spectrum, $P(k)$. $< n_i n_j >= \sigma^2_{N_i} \delta^k_{ij}$ since the noise in both pixels is uncorrelated.

A large portion of this project is to fully understand the various contributors to the spectra. We are aware that our data model is not perfect, such that the probability distribution of χ^2 of the fit will not be formally good. Thus, when we cut quasars with $P(\chi^2) < 0.01$, as discussed in Chapter 2.2, we are actually removing more than 1% of the data. Some quasars exhibit Broad Absorption Line (BAL) features which are caused by high velocity outflows near the quasar. Damped Lyman Alpha (DLA) quasars are those in which the light from the quasar passes a high density region, most likely the outer edges of a galaxy, causing damping wings in the absorption at that region in the Lyα forest. Our continuum estimate does not properly model Broad Absorption Line (BAL) quasars or Damped Lyman Alpha (DLA) systems, so a $P(\chi^2) < 0.01$ cut will also remove outliers that contain BALs and DLAs which may have not been removed using the quasar catalog flags.

We compute mean quasar continuum, mean transmission of the forest, and correction factors to the BOSS pipeline measurements for throughput, sky, and noise as described in §3.1. In addition, we measure the variance of these values which are useful to employ to cull pixels that cannot be well modeled, as discussed in §3.2. We end this discussion of our data model by presenting the method of creating our own coadds with an accompanying noise correction factor in §3.3

3.1 Determination of Mean Parameters

The parameterization of the quasar flux in Eq. 3.6 implies that all quasars can be described by the same "global" parameters. There are certain single quasar parameters which have been absorbed into the global parameters, such as an amplitude for each quasar that is absorbed into the quasar continuum, $C(\lambda)$, that we also must measure. We perform an iterative method to measure global parameters, that are the same for all quasars, such

as a mean quasar continuum, mean transmission, throughput, sky, and noise, in addition to various single quasar parameters, which we discuss below. To fit the parameters, we employ the maximum likelihood method, where the likelihood function is modeled as a Gaussian. We compute the best fit value for a given parameter, one by one, holding the other parameters constant. For example, we first fit an amplitude to each quasar, holding all other parameters constant, then measure the global quasar continuum, holding all other parameters constant, then the mean absorption in the forest as a function of redshift, continuing through the parameters. To find the values that maximize the likelihood, we employ two different techniques: Newton-Raphson, or minimization routines from the GNU Scientific Library [GSL, Gough (2009)]. The Newton-Raphson method iterates by creating a step size for each parameter defined by:

$$\delta P = \mathcal{L}^{-1}_{,\alpha\beta}\ \mathcal{L}_{\beta} \tag{3.10}$$

where $\mathcal{L}_{,\beta}$ and $\mathcal{L}_{,\alpha\beta}$ are the first and second derivatives of the log likelihood with respect to a parameter, respectively. GSL minimization is a package we use that contains routines to find minima of multidimensional functions, which also uses the technique of creating iterative steps. Both methods are used to iterate to convergence.

We restrict our rest-frame wavelength range for the Lyα forest between 1041 Å $\lambda_{\rm rest} <$ 1185 Å to remove the tails of the Lyα and Lyman-β emission lines and absorption in the host galaxy. We show later in this section how this exact range is chosen. Our model for the intrinsic quasar continuum as a function of rest wavelength is:

$$C(\lambda_{\rm rest}) = A_q \bar{C}(\lambda_{\rm rest}) \left(\frac{\lambda_{\rm rest}}{\lambda_\star}\right)^{\alpha_q} \begin{cases} B_q & 1041\text{Å} < \lambda_{\rm rest} < 1185\text{Å}. \\ 1 & \text{otherwise}, \end{cases} \tag{3.11}$$

where A_q, α_q, and B_q are single quasar parameters, for quasar q. A_q is a quasar amplitude, α_q is a power law index to allow tilting of the continuum, with $\lambda_* = 1324$ Å. B_q is a constant introduced to ensure that the Lyα forest region of the spectrum is well modeled at least on average, and to allow marginalization over the normalization in this region as part of the Lyα forest power spectrum analysis. This is required as the continuum estimate is not perfect in

the Lyα forest because of extrapolation from outside the forest.

The parameterization of quasar continuum in Eq. 3.11 is following the work of McDonald et al. (2006), where they found that adding Principal Component Analysis (PCA) terms did not improve the result and this simple parameterization was sufficient for the Lyα forest analysis. Note that we assume that pixels are uncorrelated when computing the quasar continuum, although this is obviously not true in the Lyα forest region. This is another factor that will affect the χ^2 of the fit, resulting in a probability distribution of χ^2 that is not formally good.

In Figure 3.1, we plot $C(\lambda_{rest})$ which is computed by evaluating the maximum likelihood method. Plotted in Figure 3.2 is an example of a Quasar spectra with the quasar continuum overplotted.

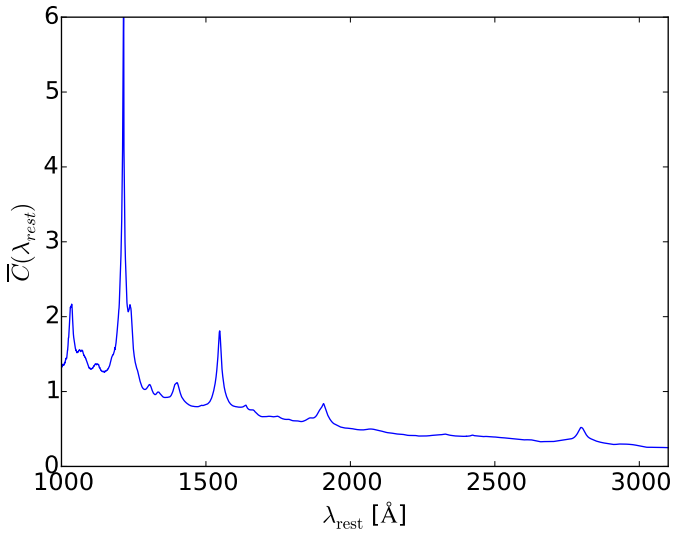

Figure 3.1 Mean quasar continuum as a function of rest wavelength measured from DR12 standard pipeline coadded spectra.

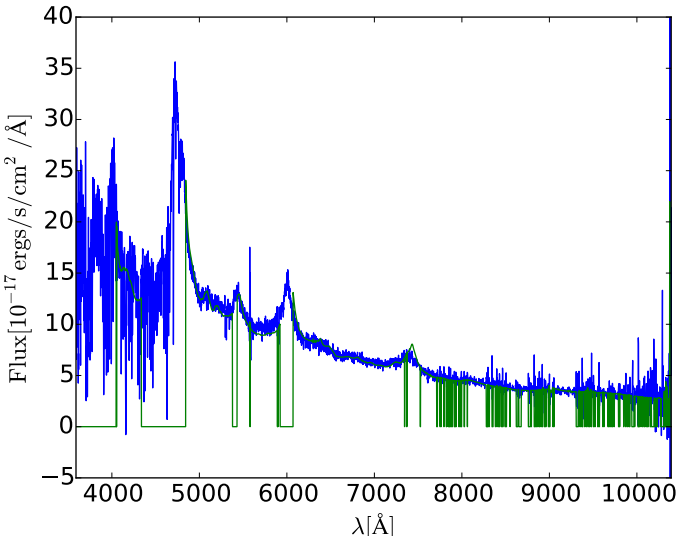

Figure 3.2 An example quasar spectra (blue) with mean continuum overplotted (green). Regions near the Lyα emission line are not fit due to the variable width of the Lyα emission line between quasars, and are therefore set to 0. Other regions are not fit if the BOSS pipeline has set a mask at that pixel, as shown where the continuum is set to 0. These regions are mostly those near bright sky lines.

We model the IGM absorption as

$$F_{\rm tot}(\lambda) = \prod_i \begin{cases} \bar{F}_i(\lambda)\,(1+\delta_F^i(\lambda)) & \lambda < \lambda_i(1+z_q). \\ 1 & \text{otherwise,} \end{cases} \tag{3.12}$$

where F_{tot} is the overall transmission fraction and $i = 1...N_a$ denotes a number of absorbing species with absorption at rest wavelength λ_i and corresponding mean absorption $\bar{F}_i$ with fluctuations, δ_F^i, around it. By construction, $\langle \delta_F^i \rangle = 0$ where the mean is over cosmological realizations. When the absorber is "behind" the quasar, i.e. $\lambda/(1+z_q) > \lambda_i$, we obviously cannot see the absorber's fluctuations.

We are making a number of simplifying assumptions. First, we limit our analysis to Lyα absorptions and therefore, from now on, $\bar{F}$ and δ_F will refer to mean absorptions and fluctuations in the Lyα forest. We will add additional absorption due to the effect of low redshift metals, as discussed in Chapter 6, but we do not model higher Lyman-series absorptions. Second, we ignore the region near the Lyα emission line of the quasar by avoiding pixels with velocity separations smaller than $\Delta v = 7650$ km s^{-1} from the quasar, since the width of the emission line varies quasar to quasar.

Our measurement of mean flux, $\bar{F}(z)$, as parameterized by Eq. 3.12, is shown in Fig.3.3. Note that this measurement is not calibrated to an absolute scale and thus, does not have theoretical significance. Our measurement is not necessarily consistent with more direct measurements from observations as is obvious at high redshifts, yet, as we are not interpreting these measurements in a scientific context, and divide our spectra by the product of mean flux, mean continuum, and throughput correction, this is a reasonable approximation.

Our measurements of throughput, sky, and noise differ from the above measurements in that we have a pipeline measurement and want to model the errors in that measurement. Therefore, we introduce a multiplicative factor $[1+\delta_T(\lambda)]$ to model the error in throughput as shown in Fig.3.4. Fig.3.5 shows some features that were not properly removed in the calibration, such as Balmer features and some interstellar absorption.

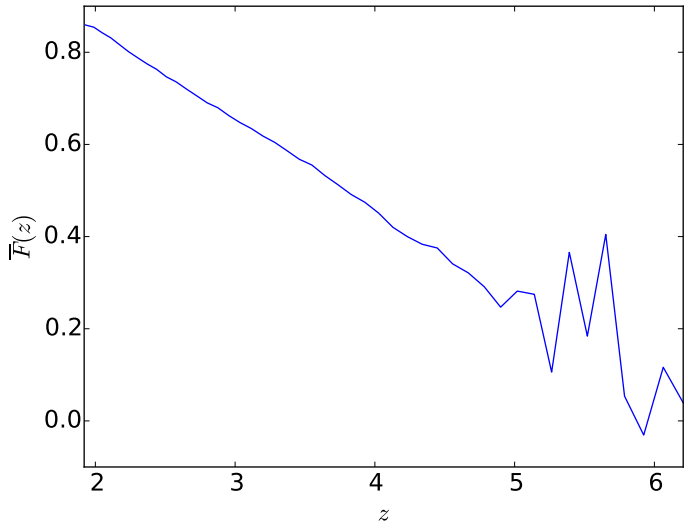

Figure 3.3 Mean IGM transmission fraction from DR12 standard pipeline coadds. The continuous function, $\bar{F}(z)$ is parameterized by linear interpolation in log spaced bins.

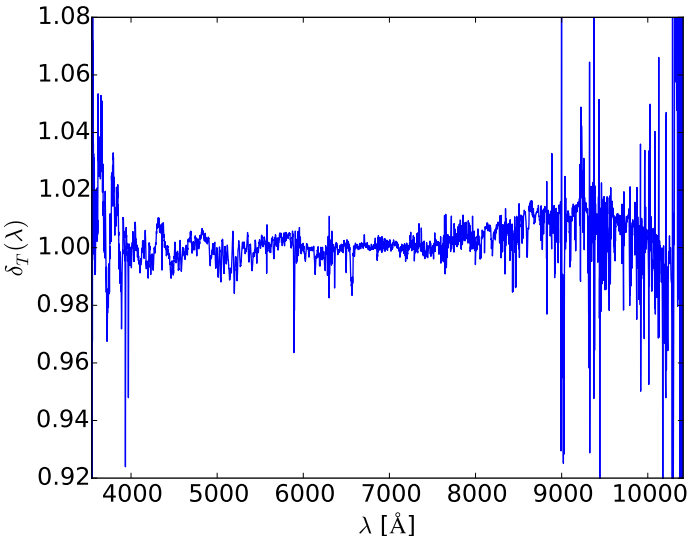

Figure 3.4 Mean throughput correction for DR12 standard pipeline coadded spectra. The mean throughput correction is parameterized in log space bins, using NGP interpolation.

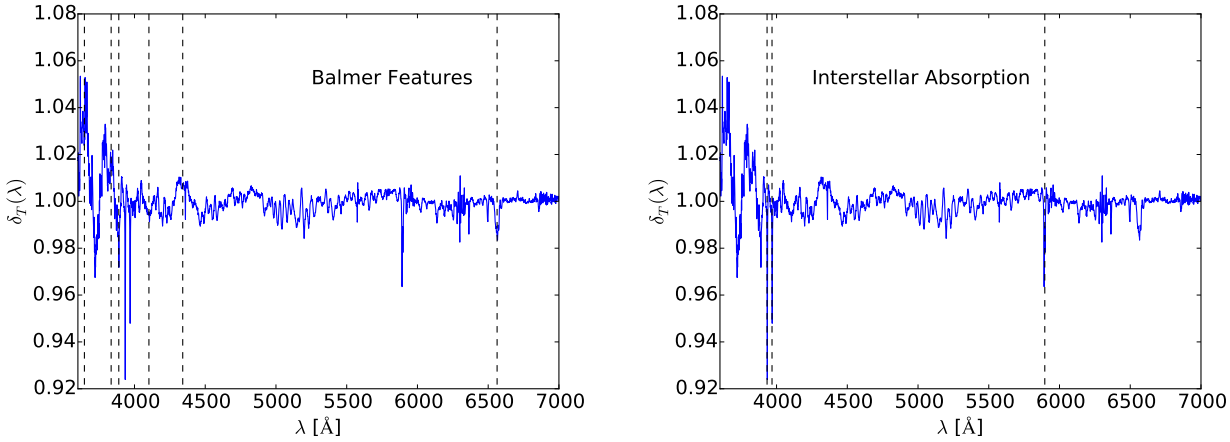

Figure 3.5 Mean throughput correction of quasars as a function of wavelength showing miscalibration of certain absorption features, namely Balmer, Calcium H & K, and NaID.

Fig.3.6 shows the measurement of sky correction, which we model as an additive term,

$$\Delta S(\lambda) = S(\lambda)\delta_{TS}(\lambda), \tag{3.13}$$

where $S(\lambda)$ is the pipeline estimate for the sky, and $\delta_{TS}(\lambda)$ is some product of errors in sky and throughput.

Table 3.1 lists the full suite of parameters used to describe the quasar spectra.

3.1.1 Breaking Degeneracies in the Data Model

The keen reader will notice that due to the multiplicative nature of continuum, mean transmission, and throughput, we suffer from degeneracies between the measurements. An example of such a degeneracy would be between the amplitude of the quasar and the quasar continuum. Namely, the model is degenerate under the transformation between the quasar amplitude and the quasar continuum, $A_q \to A_q c, \bar{C}(\lambda_{\rm rest}) \to \bar{C}(\lambda_{\rm rest})/c$ for any constant, c. We break this degeneracy by requiring

$$\frac{1}{\lambda_2 - \lambda_1}\int_{\lambda_1}^{\lambda_2} \bar{C}(\lambda_{\rm rest})d\lambda_{\rm rest} = 1, \tag{3.14}$$

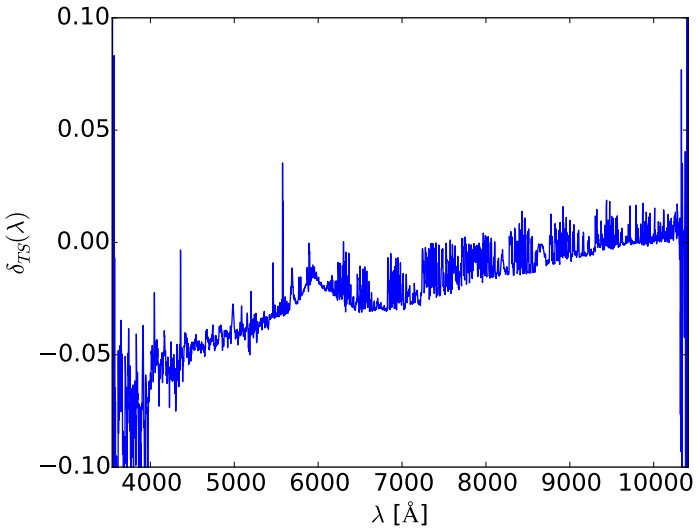

Figure 3.6 Mean sky correction measured from DR12 standard pipeline coadded spectra, parameterized in log space bins using NGP interpolation.

Parameter	type	number of parameters	meaning
$A_q(\lambda)$	per quasar	1	Normalization of quasar amplitude
$\alpha_q(\lambda)$	per quasar	1	Normalization of quasar amplitude in Lyα forest
$B_q(\lambda)$	per quasar	1	Power law slope multiplying quasar continuum
$\bar{C}(\lambda_r)$	quasar global	260	Mean quasar continuum
$\bar{F}(\lambda)$	quasar global	51	Mean transmission in the forest
$\delta_T(\lambda)$	instrument	4707	Mean Instrument Throughput
$\delta_{TS}(\lambda)$	instrument	4707	Mean Sky

Table 3.1 List of the parameters of our data model. Parameters are grouped into three sections: per quasar parameters, global quasar parameters and parameters describing instrument response.

for a side band in the red side of the spectrum with $\lambda_1 = 1268$Å and $\lambda_2 = 1380$Å. For the region in the forest, we suffer a similar degeneracy under the transformation $B_q \rightarrow B_q c, \bar{C}(\lambda_{rest})/c$. We break this degeneracy by applying a Gaussian prior on B_q with unit mean and 0.2 root-mean-square (r.m.s.), that effectively maintains the mean value of B_q very close to one. The overall slope of the mean continuum, $\bar{C}$, is also degenerate with the individual quasar slopes, α_q. Again, we break this degeneracy by applying a Gaussian prior on α_q with zero mean and 0.2 r.m.s. The mean continuum in the forest is also degenerate with the mean transmission, $\bar{F}$. One could multiply the mean flux by a power law and it would be reabsorbed in the mean continuum within the forest. We break this degeneracy by requiring $\bar{F}(z = 2.25) = 0.8$ and $\bar{F}(z = 3.25) = 0.59$. Degeneracy between mean throughput, $\delta_T(\lambda)$ and mean continuum can be broken by fixing $\delta_T = 0$ at $\lambda = 3540$ Å and $\lambda = 10450$Å.

3.2 Culling Pixels based on Variance Measurements

A unique addition to our measurement of mean values for the elements in our data model is the measurement of variances of the parameters. So, in addition to measuring the quasar continuum, we measure the continuum variance between quasar spectra, where the mean continuum is $\bar{C}$ and its variance is $\langle C^2 \rangle - \bar{C}^2 = \bar{C}^2 \sigma_C^2$. In Figure 3.7, we plot the continuum variance measurement. Large variance is measured near the Lyα emission line, and some metal lines such as CIV and MgII at 1549 Å and 2798 Å, respectively. This is consistent with the fact that these features are not exactly the same in all quasars. This plot is especially interesting in the Lyα forest region as shown in panel 2 of Figure 3.7 where it becomes apparent that we choose 1041 Å - 1185 Å specifically because the continuum variance in that region is very small, much smaller than the Lyα forest variance.

In Figs. 3.8 and 3.9, we show the measurement of throughput variance and sky variance, respectively. Here, throughput variance is defined as $\langle T^2 \rangle - \bar{T}^2 = \sigma_T^2$ and sky variance is $\langle S^2 \rangle - \bar{S}^2 = \sigma_S^2$. In particular, we can see the large variance caused by the bright sky line at 5577 Å in Fig. 3.9. N is the instrumental noise with zero mean and variance reported by the pipeline $\langle N^2 \rangle = \sigma_N^2$ to which we measure a noise correction factor as a function of

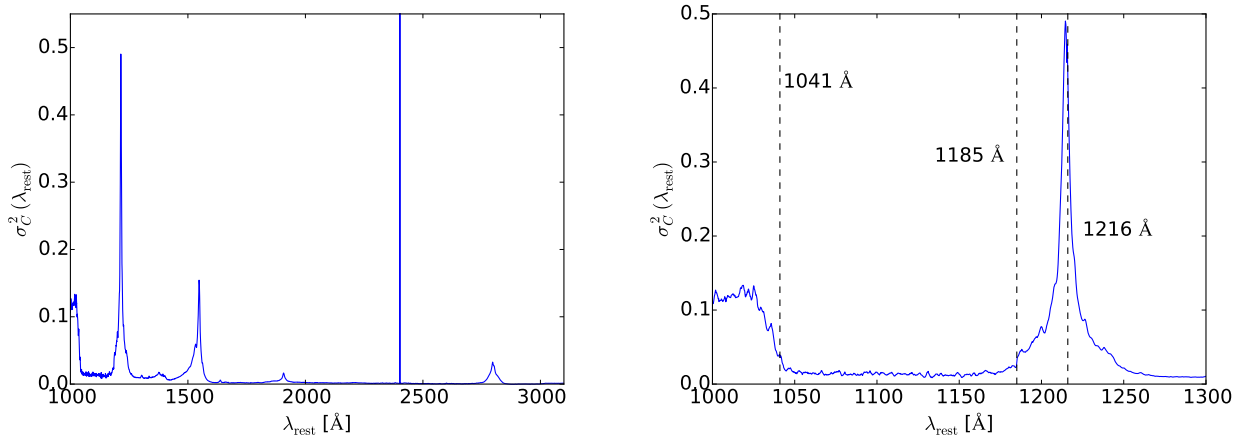

Figure 3.7 Continuum variance as a function of rest wavelength from DR12 standard pipeline coadds. Panel 2 shows a zoomed in view near the Lyα forest. The small variance between 1041Å$-$1185Å is proof that the region we use for the Lyα forest in our analysis is a reasonable choice, as the continuum is well modeled in this range.

wavelength, as shown in Fig. 3.10.

The measurement of variance is particularly useful in choosing regions to mask. For example, if sky variance is extremely high near a sky emission line, this will be apparent in our measurement of the variance and we can use this to mask the relevant region. The mean of our pixel is,

$$\langle f \rangle = \bar{C}\ \bar{F}\ \bar{T} + \bar{S} \tag{3.15}$$

and its variance,

$$\sigma_f^2 = \left\langle f^2 \right\rangle - \langle f \rangle^2 = \langle f \rangle^2 \left(\sigma_C^2 + \sigma_F^2 + \frac{\sigma_T^2}{\bar{T}^2} \right) + \sigma_S^2 + \sigma_N^2 \tag{3.16}$$

We can set a variance cut such that the pixel will be masked if the total variance of a pixel, σ_f^2, is above some threshold.

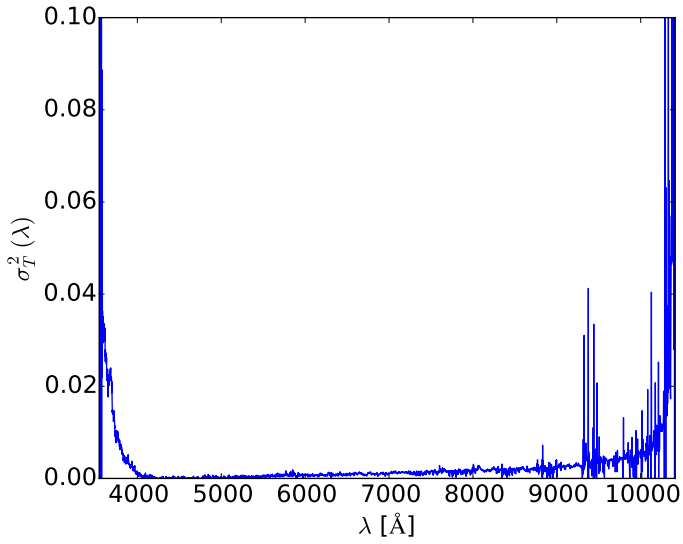

Figure 3.8 Throughput variance as a function of wavelength measured from DR12 standard pipeline coadds.

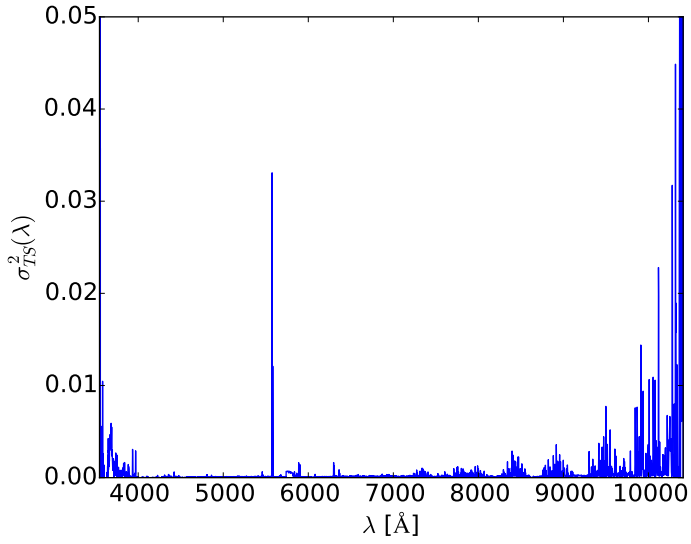

Figure 3.9 Sky correction variance as a function of wavelength measured from DR12 standard pipeline coadds.

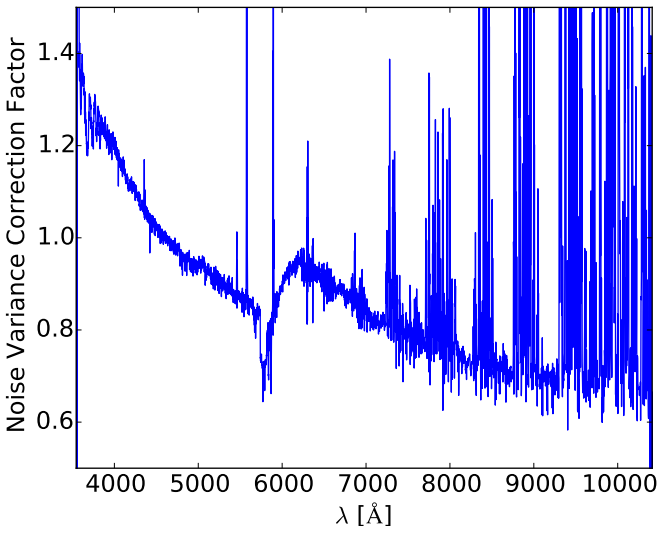

Figure 3.10 Noise variance multiplicative correction factor as a function of wavelength measured from DR12 standard pipeline coadds.

3.3 Coadded Spectra from Single Exposures

One of the elements of this analysis is that we use our own coadded spectra as opposed to those made by the BOSS data reduction pipeline. We create these coadded spectra from single exposures from the BOSS spectroscopic data. Although the BOSS project provides its own coadded spectra, this coadd is not up the standards that we require for our 1D power spectrum analysis. The following are problems that would affect our result that we need to fix. First, the BOSS project coadds are created from a spline interpolation which can cause mixing of power and does not ensure an uncorrelated noise estimate. In addition, noise is not estimated properly for BOSS coadded spectra, with no known reason as of yet for this problem. Therefore, deriving the noise estimate from single exposures is a more accurate measurement of true noise. Finally, spectrophotometric errors cause the throughput between exposures to be different. The same calibration vector is used on all exposures of one observation in the pipeline leading to differences in the throughput [Margala et al. (2015)].

Our methodology for creating the coadds is as follows. We create a fixed grid defined by the pixel width of the BOSS coadd spectra, namely $\Delta \log_{10}(\lambda) = 0.0001$. For each coadd, we compute the flux, f_{coadd}, inverse variance, $1/\sigma^2_{coadd}$, and resolution, r_{coadd}. Note that some quasars have been observed multiple times. This means that the observations could have occurred over mutliple nights resulting in different observing conditions and observing angles. Thus, the throughput could be quite different for multiple observations. We create two sets of coadds for each quasar, what we will call the *Primary* coadds and the *Combined* coadds. The *Primary* coadds consists of a coadd created from all the exposures of *one* observation, while the *Combined* coadds are created from all the exposures of *all* observations of that particular quasar. The majority of our analysis will use the *Primary* coadds, as the variance of the throughput between exposures is less.

We add the contribution from all N_{exp} single exposures using the nearest grid point and compute the weighted *intrinsic* flux and inverse variance at a grid point, x.

$$f_{coadd,x} = \frac{\sum_i^{N_{exp}} w_{int,i} f_{int,i}}{\sum_i^{N_{exp}} w_{int,i}} \tag{3.17}$$

$$\frac{1}{\sigma^2_{coadd,x}} = \sum_i^{N_{exp}} \frac{1}{\sigma^2_{f_{int,i}}} \tag{3.18}$$

where $w_{int,i} = \sigma^{-2}_{f_{int,i}}$ and we compute the weighted average of the resolution in the same way. This coadd is meant to sum the contribution of the *intrinsic* flux from the exposures, as opposed to the *observed* flux. For a pixel in an exposure, we define $f_{int,i}$ in Eq. 3.17 and $\frac{1}{\sigma^2_{f_{int,i}}}$ in Eq. 3.18 as:

$$f_{int,i} = \left(\frac{f_{obs,i} - s_i}{T_i} \right) \tag{3.19}$$

$$\frac{1}{\sigma^2_{f_{int,i}}} = \frac{1/\sigma^2_{f_{obs,i}}}{T_i^2} \tag{3.20}$$

where f_{int} is the *intrinsic* flux at a given pixel, f_{obs} is the *observed* flux as given by the BOSS pipeline, T is the throughput correction and s is the additive sky residual, all corresponding to single exposure i.

One of the advantages of using the individual exposures is to measure a single exposure throughput correction, exposure by exposure. Therefore, we can modify our throughput correction to include single exposure parameters such that

$$\delta_T(\lambda) = \delta_T^{global}(\lambda) + \sum_{n=0}^{N} a_n x^n(\lambda) \tag{3.21}$$

with $x = \log\frac{\lambda}{\lambda_0}$, where $\lambda_0 = 5500\mathring{A}$ and N is the maximum order considered. At the current state of the analysis, we have not implemented this single throughput exposure.

3.3.1 χ^2 Noise Correction Factor

In addition to the improved inverse variance computed for each coadd, McDonald et al. (2006) found that a χ^2 fit comparing the coadd flux to the contribution from each exposure was a good predictor of noise misestimation. For a pixel in a single exposure, we compare the intrinsic flux, $f_{int,i}$, to the corresponding coadd pixel, $f_{coadd,x}$,

$$\chi^2 = \sum_{i}^{N_{exp}} \frac{(f_{int,i} - f_{coadd,x})^2}{\sigma^2_{f_{int,i}}}. \tag{3.22}$$

where the degrees of freedom is $\nu = N_{exp} - 1$. We use our definitions of $f_{int,i}$ and $1/\sigma^2_{f_{int,i}}$ from Eqs. 3.19 - 3.20, and our coadded flux from Eq. 3.17. We define our noise correction factor such that $\sum_p \chi^2 = \sum_p \nu$, where the sum is taken over pixels in the region,

$$B_q \equiv \frac{\sum_p \chi^2_p}{\sum_p \nu} \tag{3.23}$$

where B_q is a multiplicative factor which is applied to the variance of the coadd. Fig. 3.11 shows a histogram of noise correction values for our *gold sample*.

3.4 Tests with Mock Spectra

We test the reliability of our method by creating mock spectra with input theories with the following properties. We use a single quasar parameters chosen from the distribution of those measured from real data. We set the continuum equal to one, execpt 5 narrow regions where

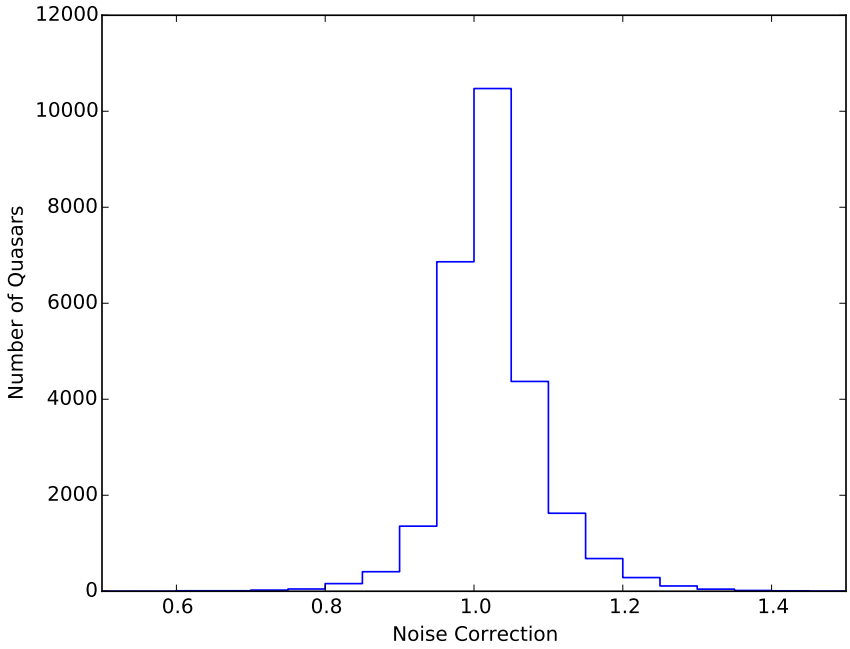

Figure 3.11 Histogram of noise correction values computed from χ^2/ν of single exposure contribution to our own coadds.

it is increased or decreased by 5%. The mean flux is set to a power law in optical depth. We set a sinusoidal throughput with 5% fluctuations. Sky is constant and equal to 0.01, and the noise variance correction is constant, equal to 1.1. Fig. 3.12 shows the results of these tests, showing reasonable convergence to the input parameters.

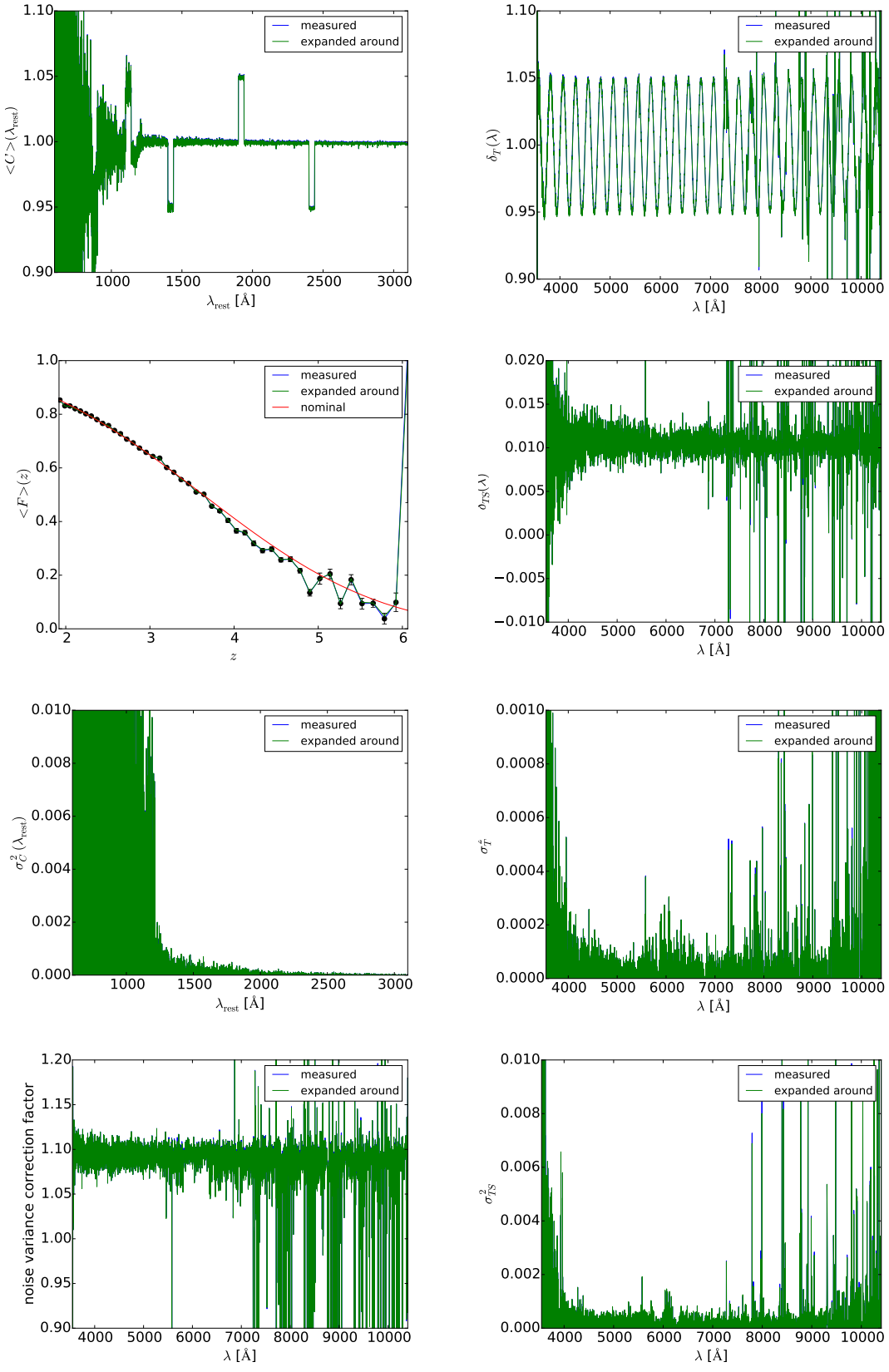

Figure 3.12 Testing convergence of data model parameters with 10,000 mock spectra.

Chapter 4

CORE ANALYSIS METHODS

This chapter is dedicated to the analysis techniques we implement throughout this project. Specifically, our analysis consists of fitting parameters to data with the ultimate goal of estimating band power amplitudes for the 1D power spectrum. We use the quadratic estimator as our basic statistic in fitting parameters throughout our analysis, whether it is band power amplitudes for the power spectrum, or parameters describing the intrinsic quasar continuum. We discuss the likelihood method and a simplification in implementation, the Optimal Quadratic Estimator, in §4.1. In §4.2, we describe our main deliverable, the power spectrum, and our method of applying the Quadratic Estimator to fit band power amplitudes.

4.1 Optimal Quadratic Method

For all elements of our analysis, we will be fitting parameters to the data by maximizing the Likelihood function. Some approximations and tricks will allow us to optimize this calculation. The likelihood, L, for a parameter, $\mathbf{p}$, is

$$L(\delta|\mathbf{p}) \propto \det(\mathbf{C})^{-1/2} \exp\left[-\frac{1}{2}\delta^t \mathbf{C}^{-1}\delta\right] = \exp\mathcal{L}, \tag{4.1}$$

where $\delta = o - m$, the observable minus the mean, and $\mathbf{C}$ is the covariance of δ. We start by approximating the Log Likelihood function, $\mathcal{L}$, by a Taylor expansion in the model parameters around a starting guess, $\mathbf{p}_0$,

$$\mathcal{L}(\mathbf{p}) \simeq \mathcal{L}(\mathbf{p}_0) + \frac{d\mathcal{L}}{dp_\alpha}\delta p_\alpha + \frac{1}{2}\frac{d^2\mathcal{L}}{dp_\alpha dp_\beta}\delta p_\alpha \delta p_\beta + ... \equiv \mathcal{L}(\mathbf{p}_0) + \mathcal{L}_{,\alpha}\delta p_\alpha + \frac{1}{2}\mathcal{L}_{,\alpha\beta}\delta p_\alpha \delta p_\beta + ... \tag{4.2}$$

where $\delta\mathbf{p} = \mathbf{p} - \mathbf{p}_0$. The function has a maximum at:

$$\mathcal{L}_{,\alpha} + \mathcal{L}_{,\alpha\beta}\delta p_\beta^{max} = 0, \tag{4.3}$$

or, solving for the parameter:

$$p_\alpha^{max} = p_\alpha^0 - \mathcal{L}^{-1}_{,\alpha\beta}\mathcal{L}_{,\beta}. \tag{4.4}$$

Our ultimate goal is to solve Eq.4.4. We can then iterate to convergence. We can write generic definitions of $\mathcal{L}_{,\alpha}$ and $\mathcal{L}_{,\alpha\beta}$ or we can simply define these for the case where $\mathbf{C}$ is the covariance matrix of the data where $\mathbf{C} = \mathbf{N} + \mathbf{S}_{,\alpha}\mathbf{p}_\alpha$, as is the case for band power amplitudes, $\mathbf{p}_\alpha$. Here, $\mathbf{S}$ is the signal covariance matrix that depends on the parameters, and $\mathbf{N}$ is the noise covariance matrix, which is independent of the band power amplitudes. The relevant equations then become

$$\mathcal{L}(\delta|\mathbf{p}) = -\frac{1}{2}\ln\det(\mathbf{C}) - \frac{1}{2}\delta^t\mathbf{C}^{-1}\delta, \tag{4.5}$$

where, for the case of the power spectrum, δ is the density fluctuation field. The first and second derivatives simplify to:

$$\mathcal{L}_{,\alpha}(\delta|\mathbf{p}) = \frac{1}{2}\delta^t\mathbf{C}^{-1}\mathbf{C}_{,\alpha}\mathbf{C}^{-1}\delta - \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,\alpha}\right]; \tag{4.6}$$

$$\mathcal{L}_{,\alpha\beta}(\delta|\mathbf{p}) = \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,\alpha}\mathbf{C}^{-1}\mathbf{C}_{,\beta}\right] - \delta^t\mathbf{C}^{-1}\mathbf{C}_{,\alpha}\mathbf{C}^{-1}\mathbf{C}_{,\beta}\mathbf{C}^{-1}\delta. \tag{4.7}$$

The standard quadratic estimator makes the additional assumption that the quadratic term can be replaced by its expectation value [Tegmark (1997); Seljak (1998)]. The expectation value of $\mathcal{L}_{,\alpha\beta}$ is the Fisher matrix:

$$F_{\alpha\beta} \equiv -\left\langle\mathcal{L}_{,\alpha\beta}\right\rangle = \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,\alpha}\mathbf{C}^{-1}\mathbf{C}_{,\beta}\right]. \tag{4.8}$$

By replacing $\mathcal{L}_{,\alpha\beta}$ in Eq.4.4 with its average value, $-F_{\alpha\beta}$, the iterated estimate $\hat{\mathbf{p}}$ is:

$$\begin{aligned}
\hat{\mathbf{p}}_\alpha &= F^{-1}_{\alpha\beta}(F_{\beta\gamma}\mathbf{p}^0_\gamma + \mathcal{L}_{,\beta}) = \frac{1}{2}\,\mathrm{Tr}[\mathbf{C}^{-1}\mathbf{S}_{,\beta}\mathbf{C}^{-1}\mathbf{S} + \delta^t\mathbf{C}^{-1}\mathbf{S}_{,\beta}\mathbf{C}^{-1}\delta - \mathbf{C}^{-1}\mathbf{S}_{,\beta}] && (4.9)\\
&= \frac{1}{2}F^{-1}_{\alpha\beta}\mathrm{Tr}[\delta^t\mathbf{C}^{-1}\mathbf{S}_{,\beta}\mathbf{C}^{-1}\delta - \mathbf{C}^{-1}\mathbf{S}_{,\beta}\mathbf{C}^{-1}\mathbf{N}] && (4.10)\\
&= \frac{1}{2}F^{-1}_{\alpha\beta}\mathrm{Tr}[\mathbf{C}^{-1}\mathbf{S}_{,\beta}\mathbf{C}^{-1}(\delta\delta^t - \mathbf{N})]. && (4.11)
\end{aligned}$$

We use this optimal quadratic estimator as the core statistic when fitting various parameters in our data model, e.g. continuum parameters, sky variance parameters, power spectrum band power amplitudes.

4.1.1 Techniques for Matrix Inversion

As shown in §4.1, the likelihood function and its derivatives require inverting the covariance matrix of the data. For continuum fitting, this encompasses the full spectra, which requires sophisticated techniques to minimize the time and memory required to perform this matrix inversion. We use the Cholesky decomposition to simplify the positive definite covariance matrix. Namely,

$$\mathbf{C} = \mathbf{L}\mathbf{L}^{\mathbf{T}}, \tag{4.12}$$

where $\mathbf{L}$ is a lower triangular matrix with real and positive diagonal entries and $\mathbf{L}^{\mathbf{T}}$ is the transpose, the upper triangular matrix.

4.2 Power Spectrum Estimation

The statistic we aim to measure is the power spectrum of fluctuations in the Lyα flux. This is essentially the Fourier transform of the velocity space correlation function. We describe how the power spectrum band power amplitudes can be fit using the optimal quadratic estimator as detailed in §4.1.

Before jumping into a discussion about how to measure the power spectrum, it is worth discussing why our chosen statistic is the power spectrum, as opposed to the correlation function. These two differ in that the power spectrum is the Fourier transform of the correlation function,

$$P(k) = \int_{-\infty}^{\infty} \xi_{ij}\ e^{-ikr} dr. \tag{4.13}$$

Therefore, there must be some advantage to using one over the other. For our analysis, band power measurements are more indepedent than measuring the correlation function. This is because Fourier modes are orthogonal, although not perfectly orthogonal as the length of the forest in our measurement is finite and this causes a correlation between Fourier modes. Resolution limits the ability to probe the smallest scales, as we will discuss later in this section. Yet, for the power spectrum, this only affects high-k modes, where $k = 2\pi/l$, and l is the length of the Fourier mode. For the correlation function, it is more difficult to assess at

what distances the resolution affects. As Baryon Acoustic Oscillations measurements deals with larger scales, the correlation function is a reasonable statistic. Since our aim is to probe the smallest of scales, the power spectrum is a more appropriate choice.

The reader may, at this point, then ask why we are using a likelihood method as opposed to performing a Fourier transform on the delta field to compute the power spectrum. There are two main advantages to using the likelihood method as opposed to the Fast Fourier Transform (FFT) method. First, if there is a region to mask within the forest for any reason, such as a bright sky line or a bad pixel on the CCD, if we mask pixels and use the FFT method, we will be missing the power contributed by this masked region at a certain Fourier mode. For the likelihood method, since we are fitting a power spectrum to the data, masking pixels does not cause such a bias. The second advantage is that the FFT method requires one resolution value for the entire Lyα forest range. This reduces its ability to accurately measure the power at high-k, or small scales, where the power spectrum is very sensitive to resolution. Another disadvantage of the FFT method is that redshift information becomes averaged when Fourier transformed, therefore the redshift information in the resultant power spectrum is not exact. The likelihood method is most easily applied for Gaussian probability distributions. The power spectrum of Lyα forest fluctuations is not exactly Gaussian but previous results from McDonald et al. 2006 showed that deviations from Gaussian are small. Although there are disadvantages to the FFT method, for completeness, we do measure our power spectrum using the FFT method as well. This proves to be a much faster method for code testing purposes and provides a second check as to whether our measurement is trustworthy, at least at large scales.

To estimate a given band power amplitude, $\hat{\mathbf{p}}_\alpha$, we use Eq. 4.11,

$$\hat{\mathbf{p}}_\alpha = \frac{1}{2} F^{-1}_{\alpha\beta} \mathrm{Tr}[\mathbf{C}^{-1}\mathbf{S}_{,\beta}\mathbf{C}^{-1}(\delta\delta^t - \mathbf{N})]. \tag{4.14}$$

The most difficult aspect of this calculation is to invert the covariance matrix and properly compute its derivative. We now discuss how to compute $\mathbf{C}$ and $\mathbf{C}_{,\alpha}$. We start with δ_F of Lyα fluctuations. For an overdensity, $\delta_F(r)$, in real space, we can write the correlation function

of two points, i and j, as

$$\xi_{ij}(r) \equiv \left\langle \delta_i^F \delta_j^F \right\rangle . \tag{4.15}$$

Transforming to Fourier space, the overdensity is:

$$\tilde{\delta}(k) = \int_{-\infty}^{\infty} \delta_F(r) e^{-ikr}\ dr. \tag{4.16}$$

Therefore, the power spectrum can be defined as a Fourier transform of the correlation function,

$$P(k) \equiv \left\langle \tilde{\delta}_i \tilde{\delta}_j \right\rangle = \int_{-\infty}^{\infty} \xi_{ij}\ e^{-ikr}\ dr. \tag{4.17}$$

Ultimately, we want to define the power spectrum as band power amplitudes to fit with the optimal quadratic estimator. This method requires fitting the power spectrum parameters to the density field as opposed to Fourier transforming the density field in order to obtain the resulting power spectrum.

We define our band power parameters as follows. We characterize the power spectrum, $P(z,k)$ with N_k bins in k and N_z bins in redshift. We start off with a fiducial function describing $P(z,k)$ or, for purposes of building up our code, various simple functions describing $P(z,k)$ (see Chapter 5). We smooth our power spectrum due to pixel width and resolution. This gives us $P(z,k,w,R)$ where w is the pixel width and R is the resolution, and the smoothing is as follows:

$$P(z,k,w,R) = P(z,k)\frac{\sin^2(kw/2)}{(kw/2)^2} e^{-(kR)^2} = P(z,k)\ W(k,w,R)\ , \tag{4.18}$$

where the smoothing term,$W(k,w,R)$, is the convolution of a top hat function for pixel width and a Gaussian kernel for resolution in Fourier space. For further discussion on smoothing, refer to Appendix A.

We compute the Fourier transform:

$$\xi(r,z,w,R) = \frac{1}{2\pi}\int_{-\infty}^{\infty} P(z,k,w,R)\ e^{ikr}\ dk, \tag{4.19}$$

where the correlation function is equivalent to the covariance matrix, $\mathbf{C}$, as defined in the optimal quadratic estimator in Eq.4.11,

$$\mathbf{C}_{ij} = \xi(z_{ij}, r_{ij}, w_{ij}, R_{ij}). \tag{4.20}$$

Here, i and j are the indices of a pixel pair and, as mentioned above, we are able to use the discrete values of resolution as opposed to one per spectra. The derivative, $\mathbf{C}_\alpha$, of the covariance matrix with respect to a parameter, α, is:

$$\begin{aligned}
\mathbf{C}_{ij,\alpha} &= \frac{\partial \mathbf{C}_{ij}}{\partial P_\alpha} \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\partial P(z,k,w,R)}{\partial P_\alpha}\ e^{ikr}\ dk \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} Z^\alpha(z_{ij})\ K^\alpha(k)\ W_{ij}(k)\ e^{ikr}\ dk \\
&= Z^\alpha(z_{ij})\frac{1}{2\pi}\int_{k_\alpha - dk/2}^{k_\alpha + dk/2} W_{ij}(k)\ e^{ikr_{ij}}\ dk
\end{aligned} \tag{4.21}$$

where dk is the width of the k bin, $W_{ij}(k) = W(k, w_{ij}, R_{ij})$, and we have decomposed the derivative into a function of redshfit $Z^\alpha(z)$ and a function of wave number $K^\alpha(k)$. This factorization allows us to describe the derivatives as product of a 1D function $Z^\alpha(z)$ times a 3D interpolated function that depends on (r,w,R). Since we are using nearest grid point (NGP) interpolation, both these functions will be one if $z \in z_\alpha$ and $k \in k_\alpha$ respectively, and zero otherwise. Plugging Eqs. 4.20 and 4.21 into Eq. 4.11 gives us the measurement for the band power amplitude. We iterate to reach convergence.

We define our k bins based on the specifics of the BOSS quasar spectra, where $k = 2\pi/l$, where l is the wavelength of the Fourier mode, in units of km s^{-1}. The pixel width of a coadd pixel is defined as 69 km s^{-1}, where $\Delta v = c\Delta \ln(\lambda)$ relates the wavelength to units of velocity. Our Lyα forest spans approximately 500 pixels, or 34,500 km s^{-1}. This corresponds to a fundamental mode of $k = 0.00018$ s km^{-1}. The Nyquist limit is $k = \pi/l$, and for our BOSS spectra, $k = 0.046$ s km^{-1}. Although this is the technical limit at which we can define our power spectrum, due to resolution variation, we will only be measuring our power spectrum in the range $k = 0.0002 - 0.02$ s km^{-1}. We add a large constant to the covariance matrix to remove sensitivity to the mean value of the forest.

An additional discussion regarding measurement error on the power spectrum is included in Appendix B.

Chapter 5

MOCK SPECTRA

We generate mock data sets as a means to test our power spectrum fitting code. There are many systematics that can affect our measurement, such as redshift evolution, variations in noise and methods used to measure the power specrtrum (either FFT or optimal quadratic estimator). We create a controlled environment in which to test these through mock quasar spectra.

It is important to note the difference between what we refer to as mock data sets and simulations. When constraining cosmological parameters from the power spectrum, by relating the 1D power spectrum to the linear matter power spectrum, one would need cosmological simulations built from first principles. This is an entirely different problem and to assess the consistency of the measurement of the 1D flux power spectrum, we do not need cosmological simulations. This drastically simplifies the problem, as we do not need to input the basics of cosmology to generate mock spectra. Since the data we use are essentially one dimensional skewers probing the Universe, we are able to create this mock data set without incorporating 3D clustering [Font-Ribera et al. (2012)], which makes this much more computationally manageable. Assuming we have a functional form of $P(z, k)$, we can use that as an input power to create mock spectra which mimics the most realistic delta field of fluctuations in the Lyα forest. We use a functional form of $P(z, k)$ as parameterized in the work of Palanque-Delabrouille et al. (2013).

As a major part of this thesis involves creating seamless code architecture that can be easily adopted to use data from different telescopes, part of our methodology for creating mock spectra is to accurately emulate how the density field is obtained, in addition to constructing a realistic portrayal of the density field. For example, we could bypass making

spectra altogether and just create a density field of fluctuations in the Lyα forest. Yet, to use this mock data set to its full capacity, we want to test whether our code can start from an observed spectrum and take the appropriate steps to recover the delta field, by removing quasar continuum, dividing by mean flux, etc.

To ensure that our power spectrum fitting code is robust, we do not immediately generate mocks mimicking the BOSS quasar data sample. In §5.1, we describe the method for creating the most basic mock spectra using white noise power with corresponding fit to the $P(z, k)$. In §5.2, we generate more complex mocks using the power spectrum derived from previous measurements [Palanque-Delabrouille et al. (2013)] and the BOSS quasar redshift distribution.

5.1 White Noise Mocks

The first and most basic test would be to define a Gaussian random field with a constant power, P, that we expect to return when running the fitting code. We start by creating a long, finely spaced line of sight data vector to which we will apply pixel and resolution smoothing, and cull the relevant area pertaining to the Lyα forest. The data vector will be defined as follows:

- To ensure the region of the Lyα forest for quasars with varied redshifts will be sampled, we need to define a data vector with a long enough comoving length: 4096 Mpc h^{-1}.

- We will be sampling our mock spectra to mimic BOSS spectra, namely a pixel width of 69 km s^{-1}. The bins of the data vector should be spaced finer than this, namely, $\Delta x = 10$ km s^{-1}.

- We define the central value of the data vector as $\lambda_c = 6000$ Å.

- Each pixel will be defined as a Gaussian random number with variance, σ_F^2:

$$\sigma_F^2(i) = \frac{P}{\Delta x} \tag{5.1}$$

- Define the Lyα forest region by using redshifts of quasars from BOSS DR12.

Our mock spectra is then defined as the part of the data vector which is in the restframe wavelength range of the Lyα forest and sampled at a typical pixel width of the BOSS spectrograph, 69 km s^{-1}. The spectra is then smoothed due to resolution and pixel width,

$$W(k, w, R) = \frac{\sin^2(kw/2)}{(kw/2)^2} e^{-(kR)^2} \tag{5.2}$$

which describes the convolution of a top hat function and a Gaussian kernel in Fourier space. We can then add arbitrarily small variance per pixel, $\sigma^2_{n,i}$, either a constant value for all pixels or variance of actual BOSS pixels for the corresponding quasar spectra in the catalog.

Moving forward, we will quantify the goodness of fit for a power spectrum measurement by a χ^2 comparing the band power amplitudes to a baseline. For the case of white noise mocks, we choose a constant value of P to which we can apply redshift evolution, in the form of a power law in z, such that

$$P(z, k) = C \left(\frac{1+z}{1+z_0} \right)^n \tag{5.3}$$

where $P = 27$ km s^{-1}, $z_0 = 3$ and n is a power law index of our choosing.

Tests with these mocks will consist of varying three parameters:

- Redshift evolution: $n = 0$, or $n = 2$
- N_z, number of redshift bins for $P(z, k)$: 1 or 10
- Variance, σ^2_N, of spectra: constant for all pixels or varied

In Fig. 5.1, we show the simplest case of constant power with no redshift evolution applied, and all power measured in one redshift bin with constant noise and resolution for all pixels, namely $\sigma^2_N = 0.1$ and $R = 69$ km s^{-1}. Table 5.1 shows the results associated with Fig.5.1 - 5.4, where we have not plotted the simple cases of one redshift bin. We find that the measurements exhibit reasonable χ^2 values.

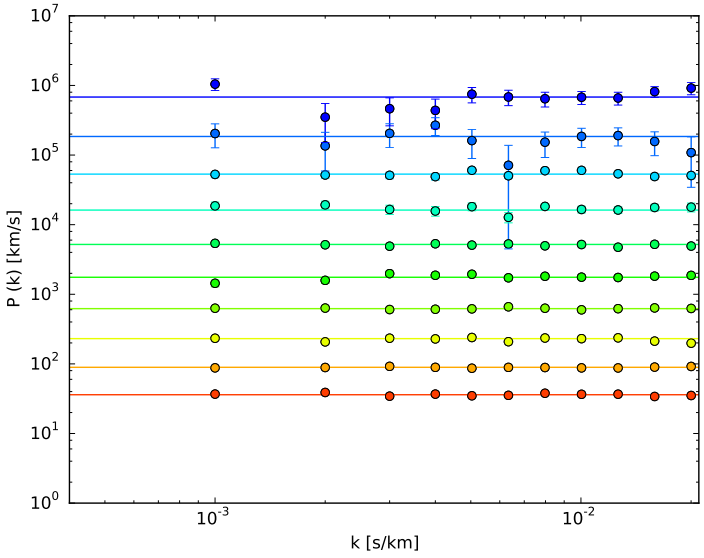

Figure 5.1 $P(z,k)$ fit on 20,000 White Noise mock spectra. Input power has redshift evolution with $n = 2$ power law index and constant $\sigma_N^2 = 0.1$ on all pixels. We measuring using $N_z = 10$ redshift bins. Refer to Table 5.1 for χ^2 information. Redshift bins, from bottom to top, correspond to $z = 2.15, z = 2.3, z = 2.48, z = 2.66, z = 2.85, z = 3.04, z = 3.25, z = 3.47, z = 3.7, z = 3.94$.

z-Evolution	N_z	σ_N^2	χ^2	N_{dof}	$P(\chi^2)$
None	1	0.1	5.0	11	0.93
z^2	1	0.1	2.8	11	0.99
z^2	10	0.1	107.59	110	0.55
None	10	0.1	107.15	110	0.56
None	10	varied	116.5	110	0.32
z^2	10	varied	117.18	110	0.30
z^2	10	1	96.8	110	0.81

Table 5.1 We show the results of tests done with white noise mock spectra. The parameters we vary are redshift evolution applied to the mock, number of z-bins, N_z, and constant or varied value of the noise variance in pixels, σ_N^2. We record the χ^2 and degrees of a freedom, N_{dof}, in addition to the probability of the χ^2.

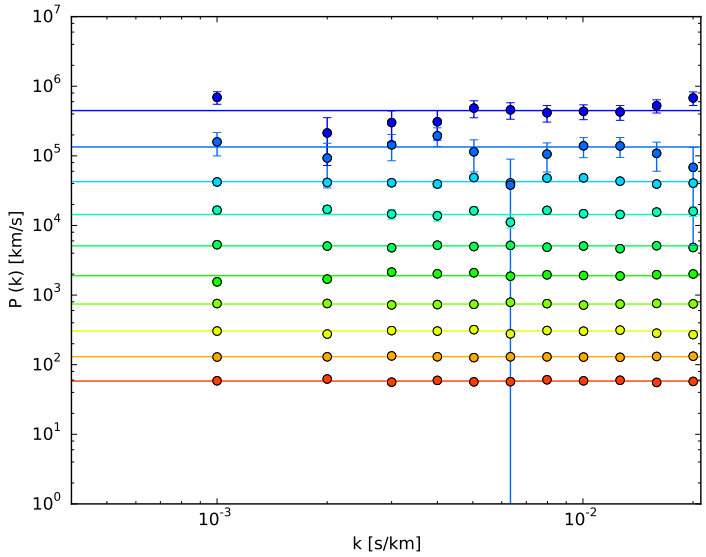

Figure 5.2 $P(z,k)$ fit on 20,000 White Noise mock spectra. Input power has no redshift evolution and constant $\sigma_N^2 = 0.1$ on all pixels. We measure using $N_z = 10$ redshift bins. Refer to Table 5.1 for χ^2 information.

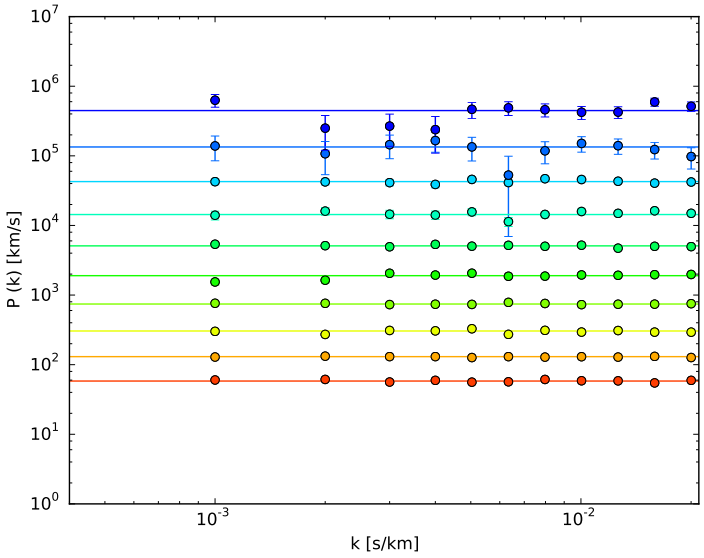

Figure 5.3 $P(z, k)$ fit on 20,000 White Noise mock spectra. Input power has no redshift evolution and varied noise on pixels using values from BOSS quasar spectra. We measuring using $N_z = 10$ redshift bins. Refer to Table 5.1 for χ^2 information.

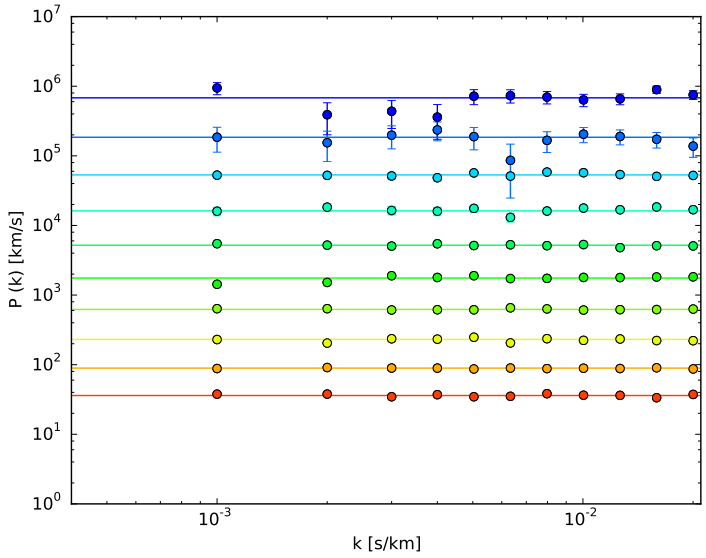

Figure 5.4 $P(z, k)$ fit on 20,000 White Noise mock spectra. Input power has redshift evolution with $n = 2$ power law index and varied noise on pixels using values from BOSS quasar spectra. We measure using $N_z = 10$ redshift bins. Refer to Table 5.1 for χ^2 information.

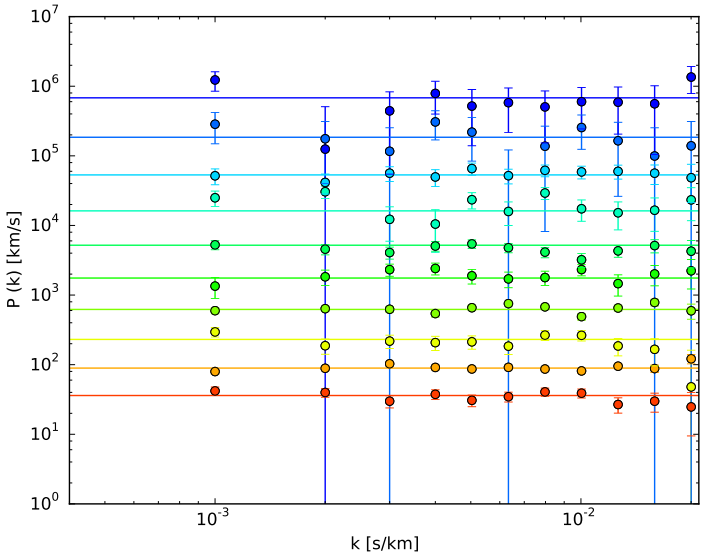

Figure 5.5 $P(z,k)$ fit on 20,000 White Noise mock spectra. Input power has rdshift evolution with $n = 2$ power law index and constant $\sigma_N^2 = 1$ on all pixels. We measure using $N_z = 10$ redshift bins. Refer to Table 5.1 for χ^2 information.

To ensure that our results on our mock tests are not reasonable by chance, we performed multiple power spectrum fits on many sets of mocks, and computed the probability distribution of the χ^2 associated with the mock data sets. In Fig. 5.6, we show the χ^2 probability distribution for 100 sets of mock data sets, each with 100 mock spectra with $N_{dof} = 10$ with the expected distribution overplotted.

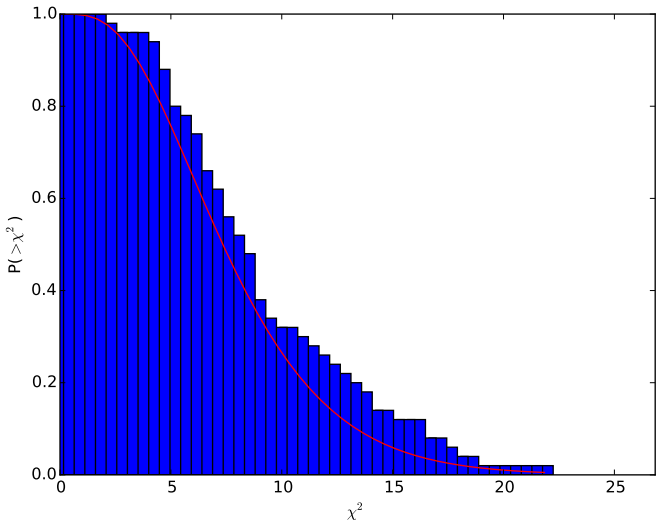

Figure 5.6 Probablity distribution for 100 mock data sets, each with 100 mock spectra. The chi^2 corresponds to $N_{dof} = 10$, with the expected distribution overplotted in red.

5.2 Mocks using $P(z,k)$ from Previous Measurements

Moving forward from the simple white noise mocks, we create mock spectra starting with the functional form of $P(z,k)$ as derived by Palanque-Delabrouille et al. (2013), which I will call the French Participation Group (FPG) mocks. For simplification of computing the mocks, we will factorize the true form of the function to make k and z separable. The following equation is the separable form of the FPG $P(z,k)$ formula:

$$P(z,k) = P(k) * P(z) = \frac{\pi A_F}{k_0}(\frac{k}{k_0})^{3+n_F+\alpha_F * \ln(k/k_0)} \ (\frac{1+z}{1+z_0})^{B_F}, \tag{5.4}$$

Parameter	Value
A_F	0.064
n_F	-2.55
α_F	-0.1
B_F	3.55

Table 5.2 Parameter values for the parameterized functional form of $P(z, k)$ from Palanque-Delabrouille et al. (2013).

where we list the values used for the various parameters in Table 5.2 and $k_0 = 0.009$ s km^{-1}.

We create a delta field using the Fourier modes defined by $P(k)$, at first ignoring the dependence on redshift. We do so by creating a long, fine skewer of Gaussian flux with the prescribed power. Although the actual density distribution of the Lyα forest is a log normal distribution, since our mocks are meant to test our code's ability to recover an input power, a Gaussian distribution is sufficient for our needs. To simulate the actual BOSS DR 12 data set, we start with the quasar spectra from our data sample. Using the redshift of the quasar, we define the Lyα forest region. Since we will be Fourier transforming the power to return our delta field, we must define one resolution and one pixel width for the forest pixels. We define the resolution as the mean of the resolution of pixels in the forest from the real spectra. Our mock spectra is mimicking coadded spectra, therefore, the pixel width is equal for all pixels in the spectra. We smooth the skewer due to resolution and pixel width using Eq.5.2, and depending on the redshift of each pixel in the forest, multiply the flux by the square root of the redshift dependence of the power.

Then, for the region outside the forest, our flux is:

$$f = C(\lambda) * T(\lambda) + S(\lambda) + N(\lambda) \tag{5.5}$$

whereas the region in the forest is:

$$f = C(\lambda)T(\lambda)F(z)[1 + \delta_F(\lambda)] + S(\lambda) + N(\lambda) \tag{5.6}$$

where $F(z)$ is our measurement for the mean transmission as a function of redshift, $C(\lambda)$ is the quasar continuum, $T(\lambda)$ is the throughput, $S(\lambda)$ is the sky, and $N(\lambda)$ is the noise, all described in Chapter 3.

Note that our mock spectra do not include absorption other than the Lyα forest absorption, although we will measure this in the BOSS spectra. With our FPG mocks, it becomes apparent why we limit our measurement of $P(z,k)$ to $k = 0.02$ s km^{-1}. Eq. 5.2 shows a strong dependence on resolution at high k, namely,

$$W(k,R) = e^{-(kR)^2}. \tag{5.7}$$

For the typical BOSS resolution of 69 km s^{-1}, there is an 85% suppression of power at $k = 0.02$ s km^{-1}. This implies that if our measurement of resolution is inaccurate by even 1 km s^{-1}, this would affect our $P(k)$ measurement by 5% at $k = 0.02$ s km^{-1}. Figure 5.7 shows the effect of resolution for a power spectrum with constant power of 27 km s^{-1}.

We can see the difficulty in recovering the power spectrum at high-k in Fig. 5.8, where we plot the power measured out to the Nyquist frequency, $k = 0.046$ s km^{-1}. The measurement starts to deviate from the baseline. Therefore, we limit our studies to $k < 0.02$ s km^{-1}. We use the FPG mocks to test the consistency between the Fourier transform method of measuring the power spectrum versus the Likelihood method. Figs. 5.9 and 5.10 show the power spectrum measurement using the FFT method and the optimal quadratic estimator, respectively. In Fig. 5.11, we show the residuals between the two, to determine whether there is a systematic difference between the two methods. We conclude that there is no systematic difference between the two methods, but will revisit this test for our power spectrum measurement on data.

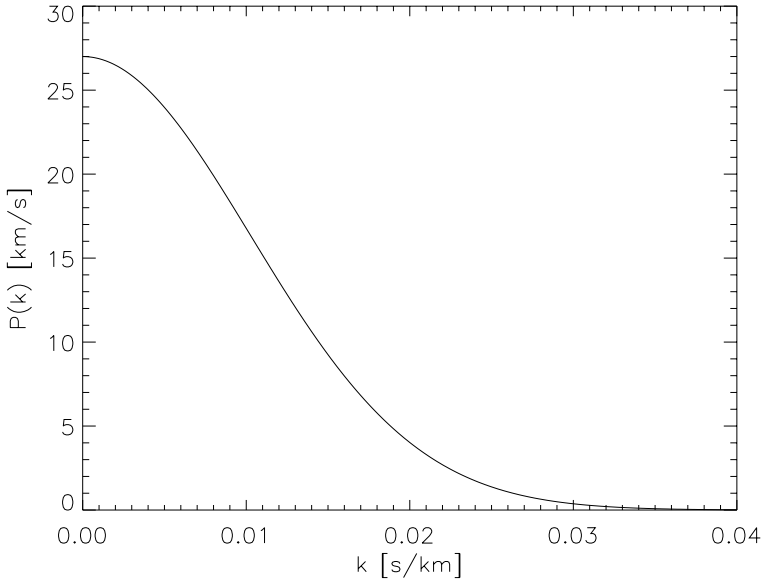

Figure 5.7 Gaussian smoothing due to pixel resolution causes damping of power at high k. For the typical resolution of the BOSS spectrograph, this translates to 85% suppression of power at $k = 0.02$ s km^{-1}.

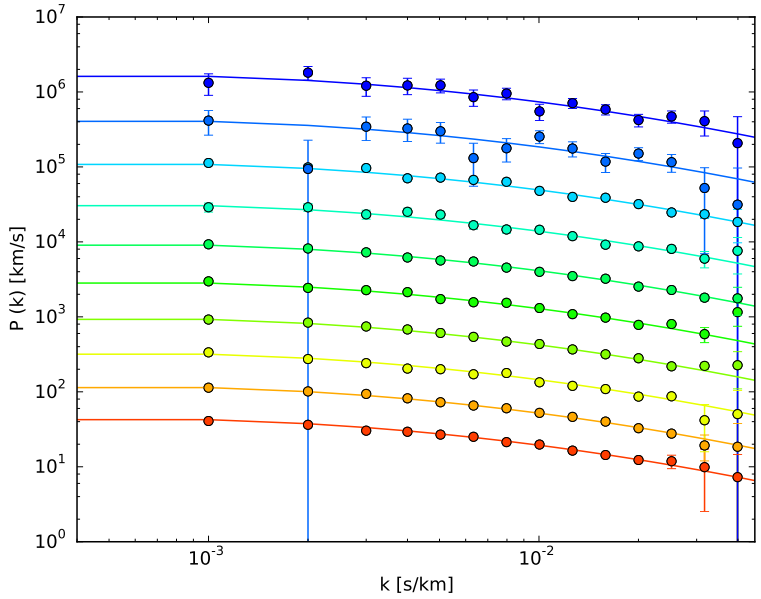

Figure 5.8 20,000 FPG Mocks made with noise from real quasar spectra. The curves are the input power spectrum used to make the mocks with the power spectrum measurement as circle points. We notice the deviation from the baseline at high-k, therefore we cut our measurements on real data to $k < 0.02$ s km^{-1}. Redshift bins as listed in caption of Fig. 5.1.

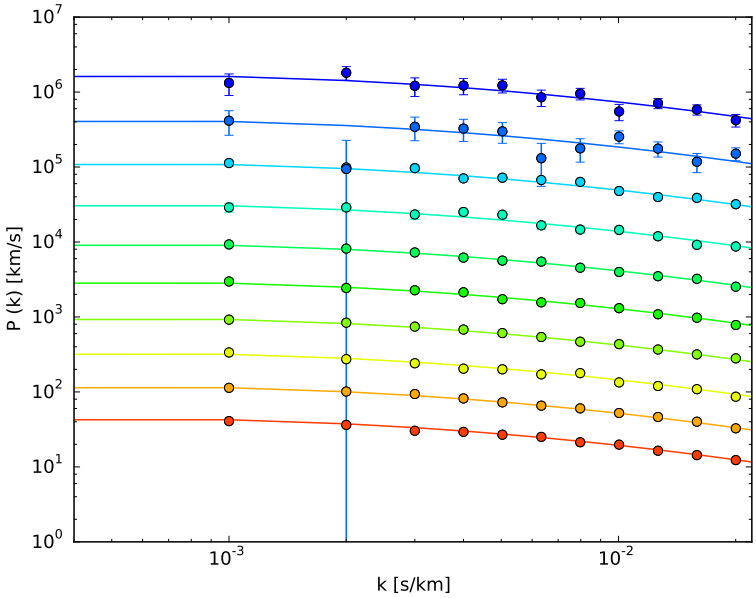

Figure 5.9 20,000 FPG Mocks made with noise from real quasar spectra using FFT method. The curves are the input power spectrum used to make the mocks with the power spectrum measurement as circle points. $\chi^2 = 108.1$ for $N_{dof} = 110$, corresponding to a probability of $P(\chi^2) = 0.53$. Redshift bins as listed in caption of Fig. 5.1.

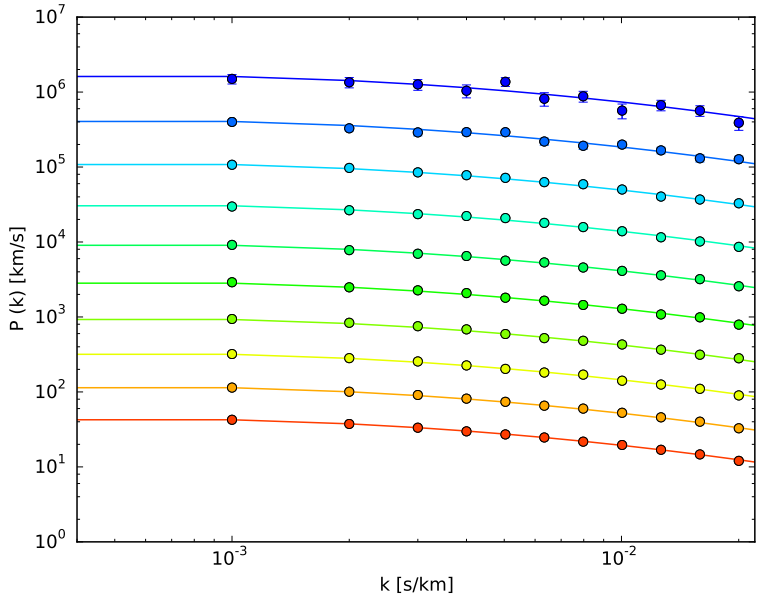

Figure 5.10 20,000 FPG Mocks made with noise from real quasar spectra using optimal quadratic estimator. The curves are the input power spectrum used to make the mocks with the power spectrum measurement as circle points. $\chi^2 = 141$ for $N_{dof} = 110$, corresponding to a probability of $P(\chi^2) = 0.02$. Redshift bins as listed in caption of Fig. 5.1.

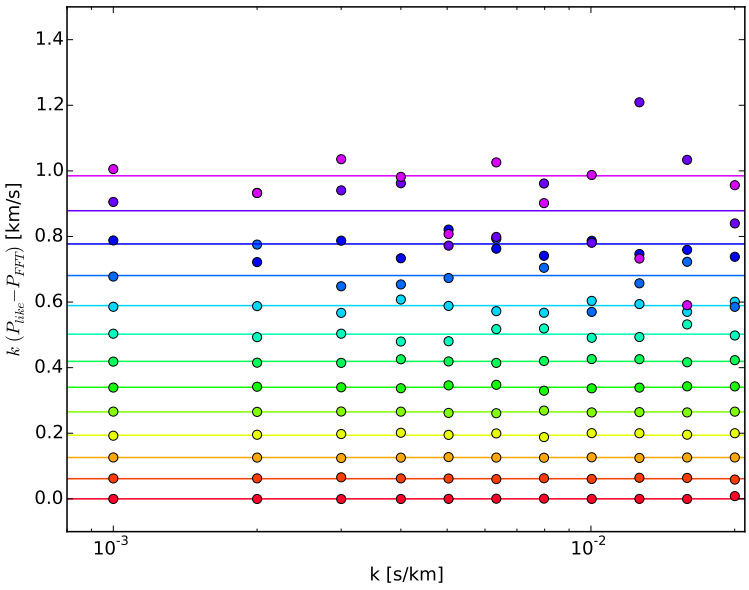

Figure 5.11 Residuals between optimal quadratic estimator measurement of $P(z,k)$ and FFT measurement. There does not appear to be any systematic difference between the methods.Redshift bins as listed in caption of Fig. 5.1.

Chapter 6

RESULTS: $P(Z, K)$ ON DATA

The final result of this thesis is a measurement of the power spectrum of the absorption of neutral hydrogen in the Lyα forest. Theoretical models predict a tight relationship between the Lyα absorption and the underlying matter density [Bi (1993); Bi et al. (1995); Reisenegger & Miralda-Escude (1995); Bi & Davidsen (1997); Croft et al. (1997); Hui et al. (1997)]. The ultimate goal for cosmologists, although out of scope of this thesis project, is to recover the matter power spectrum from the measurement of the Lyα absorption, and constrain cosmological parameters therein. Early work measured the power spectrum of Lyα absorption on a few dozen spectra [Croft et al. (1998, 1999); McDonald et al. (2000); Croft et al. (2002); Kim et al. (2004); Viel et al. (2004)]. Only in the past ten years have studies been conducted that increased the number of spectra by orders of magnitude, in order to minimize the error on cosmological parameters derived from the measurement. These studies include McDonald et al. (2006) and Palanque-Delabrouille et al. (2013).

Our analysis is done using the same survey as Palanque-Delabrouille et al. (2013), but includes data from the most recent data release. Ultimately, we use a similar number of spectra as Palanque-Delabrouille et al. (2013), but due to the rigorous cuts we apply, overall, our data set is of higher quality. Our power spectrum is measured on the *gold sample* of our own coadds which has the following cuts as described in §2.2:

- Signal-to-Noise Ratio > 7
- $P(\chi^2) > 0.01$, where this is the probability of χ^2 of the continuum fits
- Remove Broad Absorption Line Quasars

- Remove quasars with DLA systems

We use the notation $P_{\lambda_1,\lambda_2}(k,z)$ for the raw power measured in the interval $\lambda_1 < \lambda_{rest} < \lambda_2$. As discussed in Chapter 3, the Lyα forest range is defined as 1041 Å $< \lambda_{rest} <$ 1185 Å to ensure that we are measuring the power in a region where we can properly model our quasar spectra while keeping a safe distance away from the fluctuations in the Lyα emission line. We will refer to the final, background-subtracted power spectrum as $P_F(z,k)$. The background to be subtracted is the power in the wavelength range 1268 Å $< \lambda_{rest} <$ 1380 Å, which contains contamination from metals such as SiIV and CIV and removes calibration errors that we may have not accounted for properly. Therefore, the final, background-subtracted power spectrum is

$$P_F(z,k) = P_{1041,1185}(z,k) - P_{1268,1380}(z,k), \tag{6.1}$$

where $z = \lambda/\lambda_\alpha - 1$, the redshift of the Ly$\alpha$ absorption feature.

Before we can properly measure our final result, we need to assess the contamination of the Lyα forest due to metal absorption. §6.1 will delve into measurements of contaminating background power. Results following will use the covariance from this background power when measuring $P(z,k)$ of the Lyα forest, effectively subtracting this power as shown in Eq. 6.1. This chapter culminates with the final measurements of $P(z,k)$ of density fluctuations in the Lyα forest with background subtraction. Before delivering the final results, though, we assess the validity of cuts that we have made to our data sample in §6.2. To ensure consistency of our results and assess the systematics of the measurement, we conduct tests on data sets split into two on a particular parameter. We discuss these consistency tests in §6.3. We use the optimal quadratic estimator to compute the power spectrum in the majority of the tests we conduct, unless otherwise stated.

6.1 Metal Absorption

We now turn our attention to an additional absorption component from transitions of metals, where metals refers to elements with higher atomic number than Helium. In addition to the

Lyα absorption, metals in the IGM can contribute to absorption. (Bahcall et al. (1968); Sargent et al. (1980); Meyer & York (1987); Cowie et al. (1995)) At a given redshift, there are three contributing factors of absorption – the Lyα transition, metals with transitions at $\lambda < \lambda_\alpha$ and metal absorption for transitions at wavelengths greater than the Lyα transition. It is difficult to identify the absorption from metal absorption that occurs in the Lyα forest region, such as SiIII at 1206.5 Å but absorption outside the observed wavelength range of the forest can be modeled precisely. Specifically, we are aware of absorption of SiIV at 1393.75 and 1402.77 Å and CIV at 1548.2 and 1550.78 Å that contributes to the power in the forest, which would be misinterpreted as Lyα absorption were it not removed.

The current understanding of how these metals entered the IGM is that they have most likely been transported through galactic winds. (Cowie et al. (1995); Davé et al. (1998); Aguirre et al. (2001); Schaye et al. (2003); Oppenheimer & Davé (2006)). Although the necessity to understand the contribution of metals for our project is as a contaminant for the final measurement, the metal abundance in the IGM can be used to understand early star formation and feedback mechanisms in galaxies. Analysis of the metal abundances in the IGM has shown that 60-70% of systems with column densities in the range of $N_{HI} \geq 10^{13.6}$ cm^{-2} exhibit metal enrichment (Simcoe et al. (2004)). The implications are that not only does the IGM provide us with the reservoir of primordial material, but a sizable fraction is enriched. Better understanding of the redshift evolution of the metal absorption would drastically improve models of galaxy formation and the interaction with the IGM.

We measure power redward of the Lyα emission line, where there is no Lyα forest absorption. Here, one can measure most of the possible sources of contaminating background power, e.g, metals, sky subtraction errors, spectral calibration errors, misestimation of the detector noise, etc. Power measured in a given region accounts for the contamination from longer wavelength metal transitions. Note that there is no need to extrapolate from the longer wavelength metal absorption in the same quasar spectrum, which is due to gas at a different redshift. The correct background for the Lyα forest at a given redshift can be measured by looking at the same observed wavelength range in lower redshift quasars. Thus,

as mentioned above, our data sample includes low redshift quasars to study the background power.

We define mean transmission for a given metal line as:

$$F_S = \bar{F}_S(1 + \delta_S) \tag{6.2}$$

Given the flux of a pixel, f:

$$f = \bar{C}\bar{F}_F(1 + \delta_F)\bar{F}_S(1 + \delta_S) \tag{6.3}$$

$$< f > = \bar{C}\bar{F}_F\bar{F}_S \tag{6.4}$$

$$< f_i f_j > = \bar{C}^2 \bar{F}_F{}^2 \bar{F}_S{}^2 (1 + < \delta_{F_i}\delta_{F_j} > + < \delta_{S_i}\delta_{S_j} > + ...) \tag{6.5}$$

where, in addition to the absorption due to a metal, we have the quasar continuum and mean absorption of the Lyα forest. Although we know there may be absorption due to specific metal lines, there may be effects from absorption and throughput that we should account for by measuring redward of the Lyα emission line. By choosing the region $1268 < \lambda_{rest} < 1380$ we ensure that we account for the contribution of absorption from both SiIV and CIV, in addition to throughput effects.

In Figure 6.1, we plot the power, $P_{1268,1380}(z, k)$, for our *gold sample*. We define the redshift, $z = \lambda/\lambda_\alpha - 1$, so that the metal power occurs at the same observed wavelength as the Lyα absorption. Note that by computing power in this rest wavelength region, although the contribution for a given redshift for the Lyα absorption and the metal absorption does not come from the same quasar, we are measuring the exact amount of absorption due to metals at a given redshift and not an approximation.

The results in Figure 6.1 show a non-negligible measurement of power from this region. Two main features of this measurement are the bumps at $k = 0.003$ s/km and $k = 0.013$ s/km. We can attribute the first to the SiIV doublet at a separation of 1933 km/s and the second to CIV at a separation of 499 km/s, respectively. The evolution with redshift is not trivial to decouple. Although metal enrichment in the IGM increases with decreasing redshift, the ionizing background is also increasing, which can cancel each other out. The

goal of this thesis is not to analyze the metal absorption, but to remove it as a contaminant for the final result. Yet, with the extensive work accomplished by this thesis to understand the quasar spectra, as explained in previous chapters, future work could benefit greatly to attribute redshift evolution of metal absorption to underlying physics and astrophysical processes.

The reader should note that we have not accounted for correlation of metal transitions blueward of the Lyα transition with the Lyα line. There is no direct method to measure this and this is a correlation that should be modeled when using the power spectrum result to constrain cosmological parameters. We anticipate a cross correlation between Lyα and SiIII at 1206.50 Å, which would show up as a peak in the correlation function, and thus wiggles in the power spectrum. We will not fit for this in this thesis, but it would be necessary to model this before computing cosmological parameters from the 1D power spectrum. As this thesis is not concerned with computing cosmological parameters, it is out of scope of this thesis.

6.2 Tests on Data

Before delivering the final result of our power spectrum measurement, we must assess the validity of our methods and the cuts we make. There are certain choices we make on the methods we employ, e.g. whether its measuring the power using the optimal quadratic estimator or FFT, and it is necessary to make sure that these do not affect our results dramatically. For each test, we will vary only one parameter and note its effect on the power spectrum measurement. For all comparisons, we use the SiIV covariance as computed from the power measured in Fig. 6.1.

Our first measurement shows the results of computing the power spectrum on our own coadds using the optimal quadratic estimator as in Figs. 6.2 and 6.3. Although the three plots are showing the same result, we plot it in different ways to elucidate certain areas. For example, the top plot in Fig. 6.2 is at the correct overall scale with the power spectrum

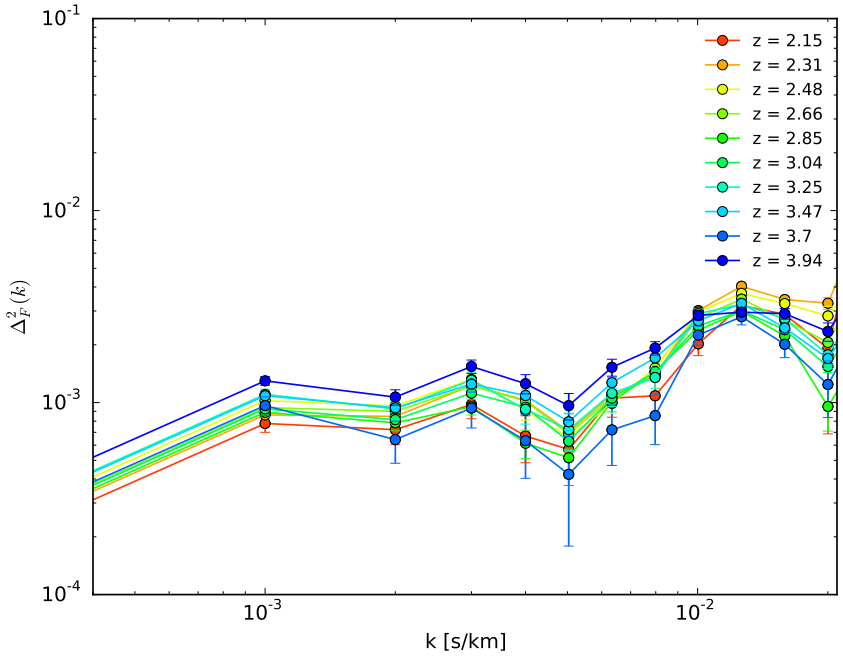

Figure 6.1 Background Power as measured in region between 1268 $\mathring{A} < \lambda_{\text{rest}} < 1380$ $\mathring{A}$ for our *gold sample*. This includes SiIV and CIV absorption, in addition to residual power from throughput. The bump at $k = 0.013$ s km^{-1} is probably due to CIV at a separation of 499 km/s. The bump at $k = 0.003$ s km^{-1} is probably due to the SiIV doublet at separation 1933 km/s. Redshift evolution is not monotonic and requires a better understanding of the physics in the IGM to disentangle.

measurement from Palanque-Delabrouille et al. (2013) overplotted. Yet, as it is hard to differentiate the measurements between redshift bins, the bottom plot in Fig. 6.2 has a constant offset between bins. In Fig. 6.3, we remove the power spectrum measurement from Palanque-Delabrouille et al. (2013) to clearly show what the overall result looks like. We plot $\Delta_F^2(k)$ vs. k, where we use the dimensionless quantity,

$$\Delta_F^2(k) = \frac{kP(k)}{\pi}. \tag{6.6}$$

The first aspect to notice in Fig. 6.2 is that, as compared to previous results from Palanque-Delabrouille et al. (2013), the lowest redshift bin, $z = 2.15$, deviates most from the baseline. This is expected, as the lowest redshift bin uses data points from the bluest end of the spectrograph, which is the region where noise is the least understood. Fig. 6.2 shows that, indeed, the most difficult area to probe is the highest-k region. As stated previously, power is suppressed by the resolution, and inaccuracies in the measurement of resolution will certainly be visible at the smallest of scales.

Our next set of figures is a measurement of the power spectrum using the FFT method. In Figs. 6.4 and 6.5, we plot the FFT result, but again, use the same plotting conventions as in Figs. 6.2 and 6.3 in order to properly analyze details in the result. We expect that the result of FFT will not be as accurate as the optimal quadratic estimator. Particularly, this should be noticeable in the high-k end of the power spectrum. As the FFT requires one resolution value for the whole Lyα forest region, we can imagine that the resolution will not be as accurate as in the optimal quadratic estimator, where we can use the values of resolution per pixel. Fig. 6.4 shows that, indeed, the high-k end of the power spectrum moves away from the baseline. A better test, though, would be to do a side-by-side comparison of the power spectrum results using the optimal quadratic estimator and the FFT method.

Fig. 6.6 plots both results using the optimal quadratic estimator, in circle points, and the FFT method, in 'x' points. Here, it is much more clear to see the difference between the two results. Although when comparing to the Palanque-Delabrouille et al. (2013), both

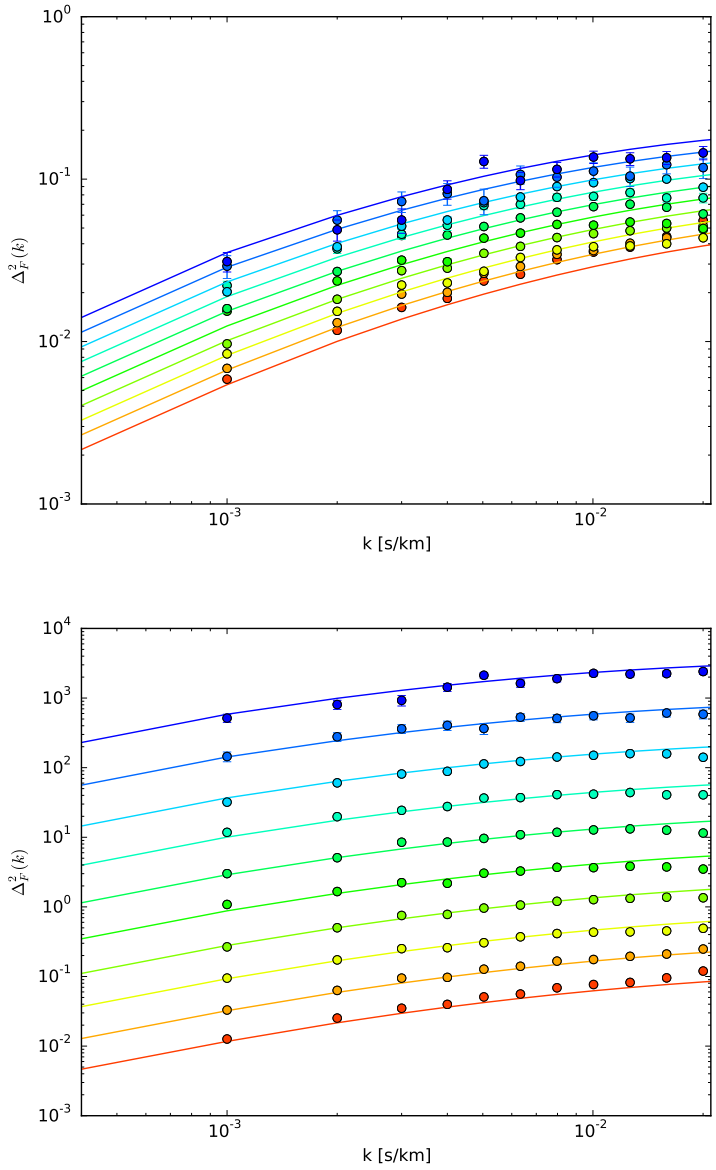

Figure 6.2 Power Spectrum measurement using the optimal quadratic estimator on our own coadds. Solid lines are power spectrum results from Palanque-Delabrouille et al. (2013) and solid circle points are our measurements of the power spectrum with error bars. Bottom plot is the same as the top but with a constant offset between redshift bins applied for clarity. Redshift bins are, from bottom to top, $z = 2.15, z = 2.3, z = 2.48, z = 2.66, z = 2.85, z = 3.04, z = 3.25, z = 3.47, z = 3.7, z = 3.94$. The most difficult redshift bins to measure are the lowest and highest redshift bins, $z = 2.15$ and $z = 3.94$, respectively.

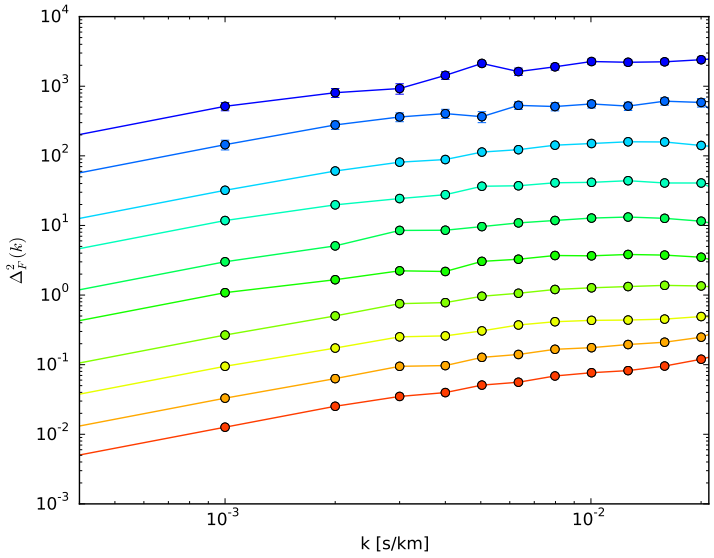

Figure 6.3 Same as bottom plot of Fig. 6.2 but without the power spectrum from Palanque-Delabrouille et al. (2013) overplotted. Redshift bins are, from bottom to top, $z = 2.15, z = 2.3, z = 2.48, z = 2.66, z = 2.85, z = 3.04, z = 3.25, z = 3.47, z = 3.7, z = 3.94$.

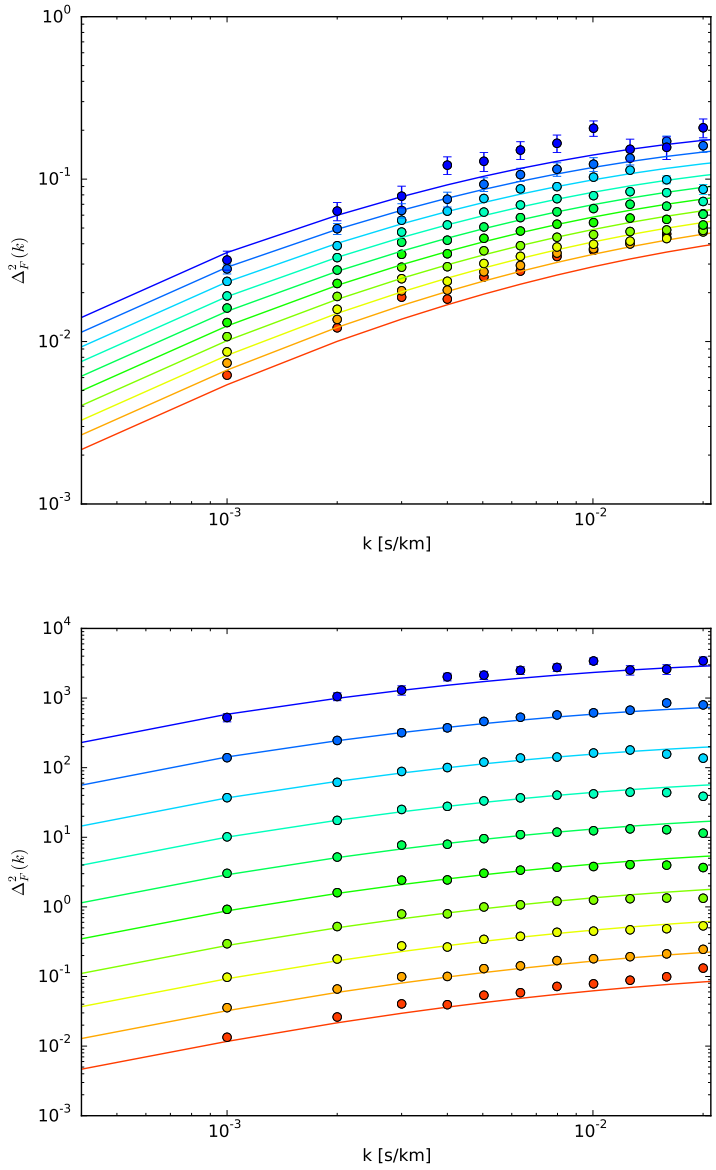

Figure 6.4 Power Spectrum measurement using the FFT method performed on our own coadds. Power spectrum measurement from Palanque-Delabrouille et al. (2013) is overplotted in solid lines. Bottom plot is the same as the top except a constant offset between redshift bins is added for clarity. Redshift bins are, from bottom to top, $z = 2.15, z = 2.3, z = 2.48, z = 2.66, z = 2.85, z = 3.04, z = 3.25, z = 3.47, z = 3.7, z = 3.94$. FFT method seems to inaccurately measure the highest redshift bin. More discussion regarding the difficulties of an FFT measurement are in Chapter 4.

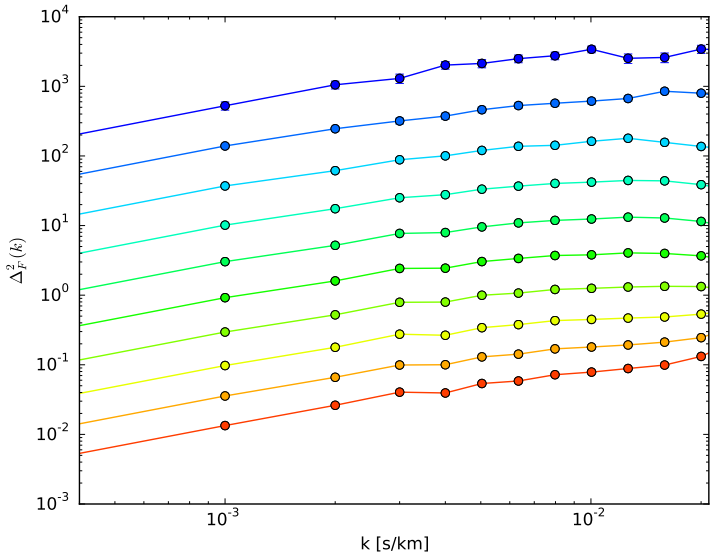

Figure 6.5 Power Spectrum measurement using the FFT method performed on our own coadds. We do not overplot the power spectrum measurement from Palanque-Delabrouille et al. (2013). Redshift bins are, from bottom to top, $z = 2.15, z = 2.3, z = 2.48, z = 2.66, z = 2.85, z = 3.04, z = 3.25, z = 3.47, z = 3.7, z = 3.94$.

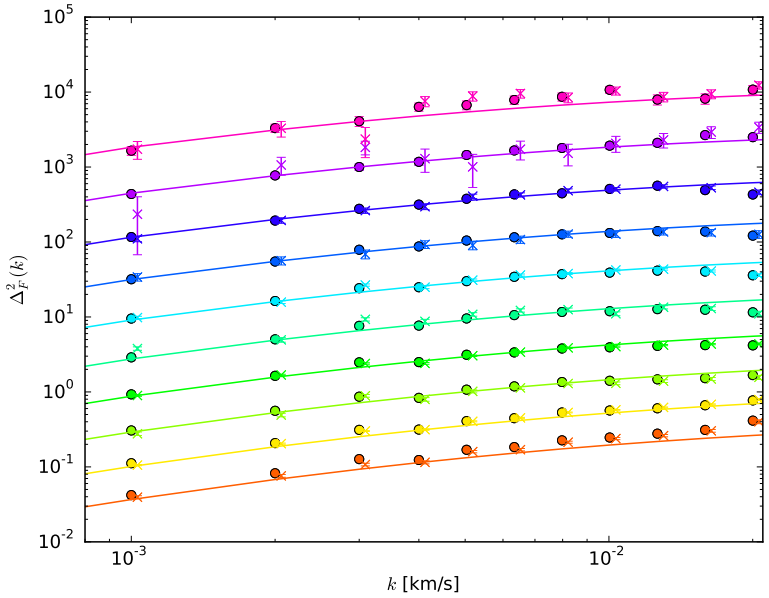

Figure 6.6 We plot two power spectrum measurements next to each other. The 'x' points are those computed using the FFT method and the circle points are those using the optimal quadratic estimator. Optimal quadratic estimator shows more consistency with previous results of Palanque-Delabrouille et al. (2013).

methods deviate from the baseline at high-k, this discrepancy is more pronounced for the FFT method. Another aspect that is noticeable is that the FFT seems to deviate much more greatly at high redshift bins. We conclude that, as expected, the optimal quadratic estimator is a better method to measure the power spectrum, and we will continue the rest of our tests using this method.

The next comparison we conduct is a comparison between our own coadds and the official BOSS coadds to understand whether there is a systematic offset between the two measurements. We use the same method as that which was used in the measurement of Fig. 6.2 - 6.3, but instead, use official coadds with their noise correction as a function of wavelength. We plot both results in Fig. 6.7. The first plot in Fig. 6.7 plots both results next to each other, while the second plot subtracts one measurement from the other to obtain the residuals between the two measurements. We find that the two results are very similar, except at the highest-k bin. In this bin, our own coadds seem to underestimate power at the high redshift bins and overestimate power at the low redshift bins. Since the only major distinction between the official coadds and our own coadds are different noise correction factors being applied, we should look into the noise correction factor for our own coadds to understand what could be happening at the smallest of scales.

Let us compare the results of power spectrum fits between our own coadds with a noise correction applied versus one without a noise correction. The noise correction we apply to our own coadds is a constant scale factor multiplied to the noise as computed from a χ^2 of the contribution of flux in a single exposure as compared to the coadd. The first plot in Fig. 6.8 shows both power spectrum measurements and the second shows the residuals between the two. We notice that applying this scale factor noise correction actually causes a deviation at the highest k bin. This begs the question of whether we should also apply a wavelength dependent noise correction factor.

We have been using the noise correction as measured from the χ^2 of the contribution of flux in a single exposure as compared to the coadd, but we should look into applying also a correction as a function of wavelength that has been applied to the official coadds. For the

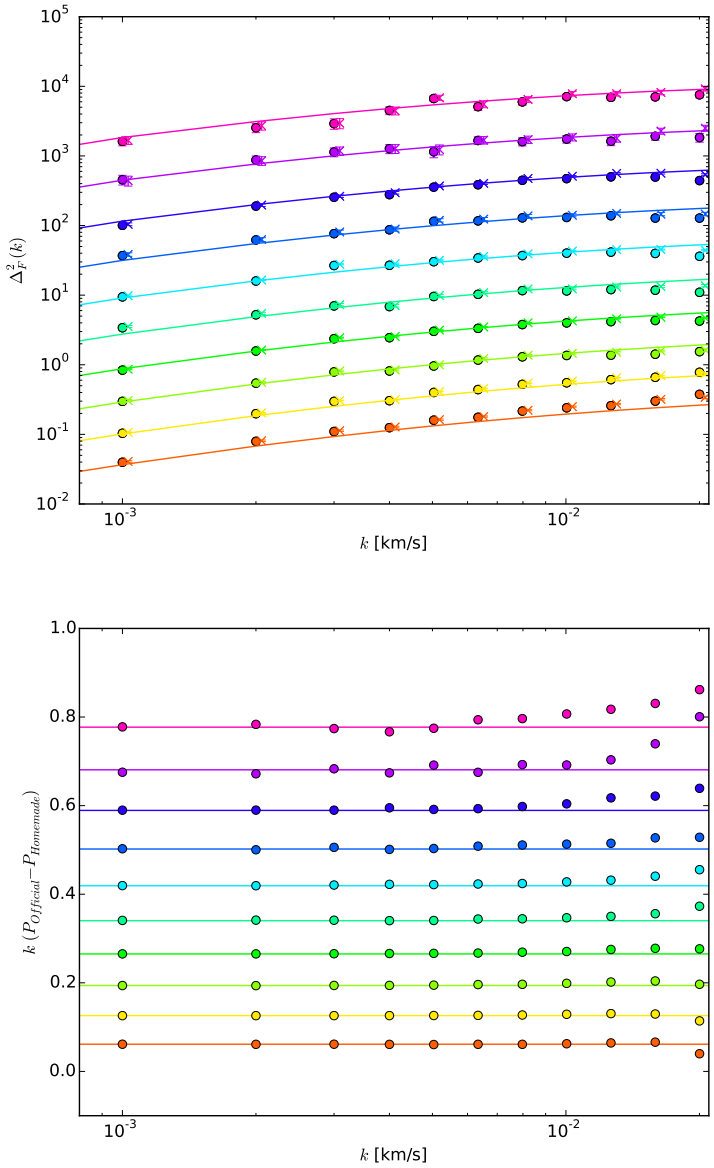

Figure 6.7 In the top plot, we show both power spectrum measurements done on our own coadds (circle points) with scale factor noise correction applied, and offical coadds ('x' points), with a wavelength dependent noise correction factor applied. In the bottom plot, we show the residuals between the $P_{official} - P_{homemade}$. We find a discrepancy at high-k, where it seems that the official coadds measure the high-k bin with more consistency to previous results from Palanque-Delabrouille et al. (2013). Redshift bins are the same as those in Fig. 6.2.

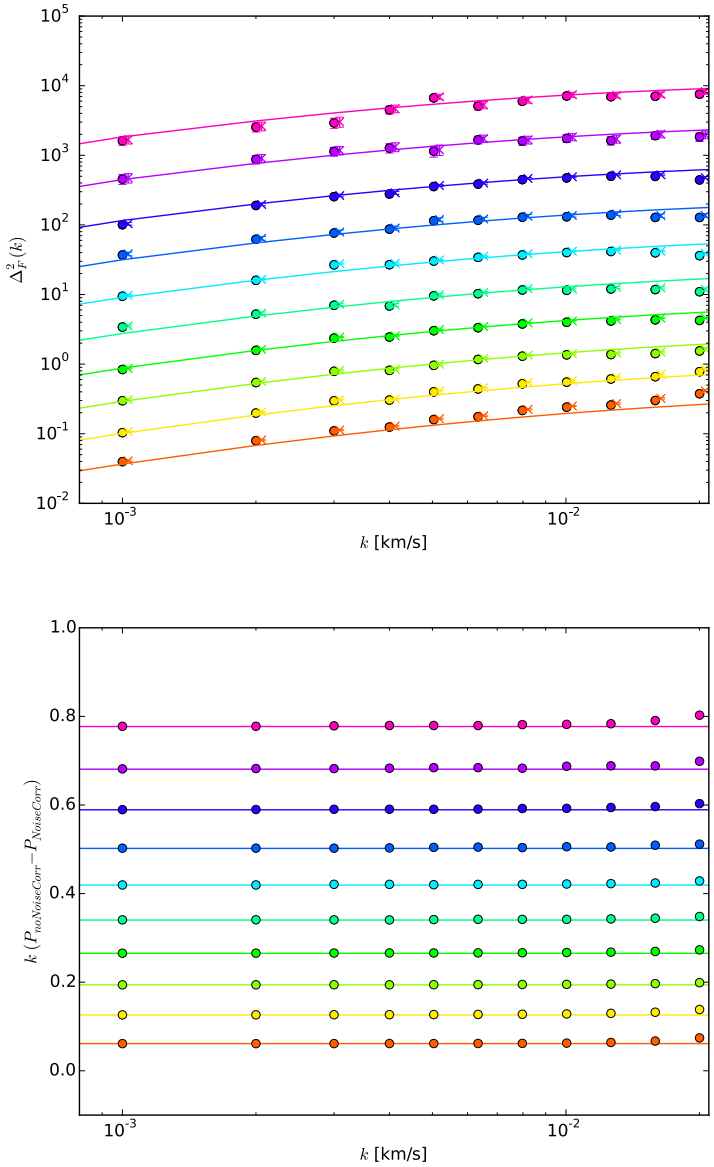

Figure 6.8 In the top plot, we show the fits of power spectrum measured on our own coadds without a noise correction ('x' points) and those with a scale factor noise correction (circle points). In the bottom plot, we show the residuals between that without a noise correction and that with a noise correction applied. The scale factor noise correction seems to still have a problem being consistent with previous results at the high-k bin. Redshift bins are the same as those in Fig. 6.2.

next test, therefore, we will be applying two noise corrections, one a constant multiplicative factor, and one as a function of wavelength. Fig. 6.9 shows the comparison between using both correction factors and only using one constant scale factor. Looking closely at the high-k bin of the power spectrum in Fig. 6.9 shows that, indeed, two correction factors, one scale factor and one wavelength dependent, does help to alleviate the problem at the high-k end of the power spectrum.

When creating an ideal data sample, we remove BAL and DLA quasars. This is performed because we know that these quasars can affect the power spectrum measurement, but it is necessary to make sure that we undertand and can explain the effect on the power spectrum. In the top plot of Fig. 6.10, we plot the power spectrum of only BAL quasars and quasars with no BAL contaminants. The bottom plot shows the residuals of $P_{BAL} - P_{NoBAL}$. We can see that the measurement on BAL quasars returns a systematically lower power. As a broad absorption line is reducing the overall flux in the forest region, one could assume that we would be measuring *more* power as there would seem to be more Lyα absorption, but that is not what we observe. It is important to remember, then, that we are first fitting the continuum and the mean flux to the quasar, and dividing to return the delta field of Lyα forest fluctutuations. Therefore, it makes sense that broad absorption lines would cause an overall lower continuum measurement to be divided out, leading to a lower power spectrum measurement for BAL quasars.

Having measured lower power for BAL quasars, the question is whether contamination from BAL quasars will drastically affect our power spectrum measurement. Although we have removed quasars that have been flagged as BAL from the quasar catalog, and our continuum $P(\chi^2)$ cut is able to cut further BAL candidates, we may still have some contamination. McDonald et al. (2006) found that including BAL quasars did not affect their measurement. For our measurement, the level of contamination would be low and should not, therefore, affect the measurement. The next test to conduct would be a comparison between quasars with and without BAL features versus quasars without BAL features to test how much the BAL contamination would affect the power spectrum measurement.

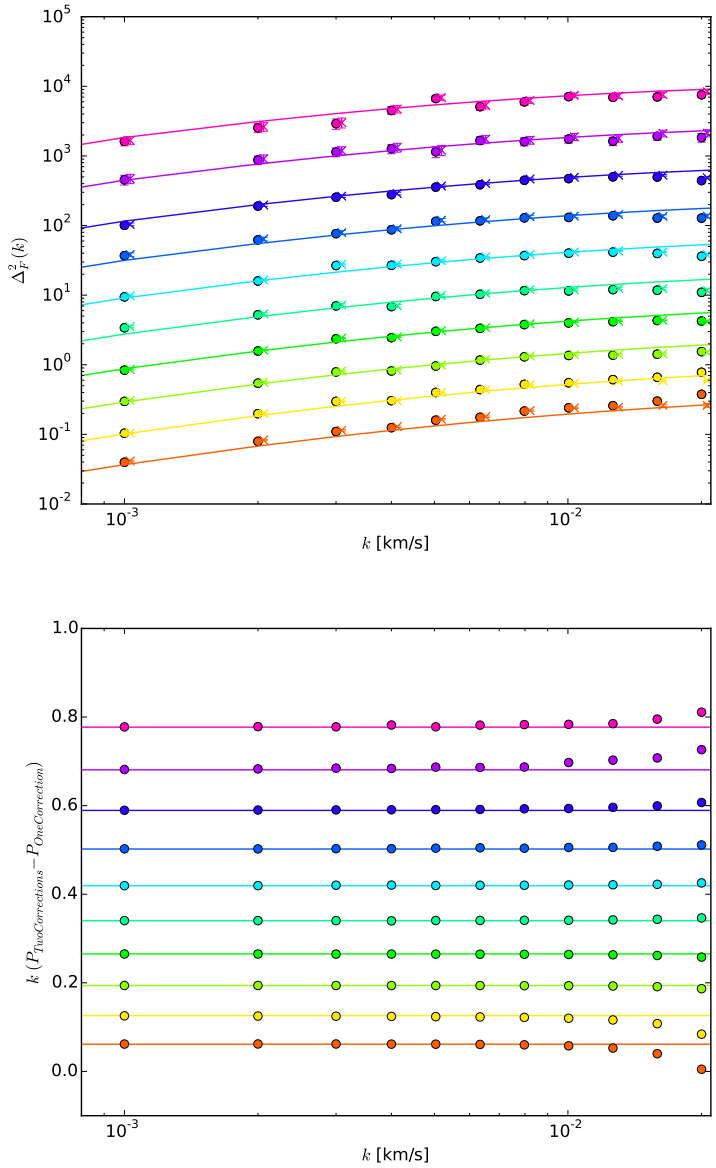

Figure 6.9 In the top plot, we show the fits on our own coadds using both corrections – scale factor and a noise correction as a function of wavelength ('x' points) as compared to that with just one noise correction, a scale factor. This figure shows that two noise correction factors help in removing the disparity between our results and previous results in solid lines from Palanque-Delabrouille et al. (2013). In the bottom plot, we show the residuals between one fit using two noise correction factors compared to that with one noise correction factor. Since noise will affect the smallest of scales, this plot is consistent in showing that the effect will be most prominent at the highest-k bin. Redshift bins are the same as those in Fig. 6.2.

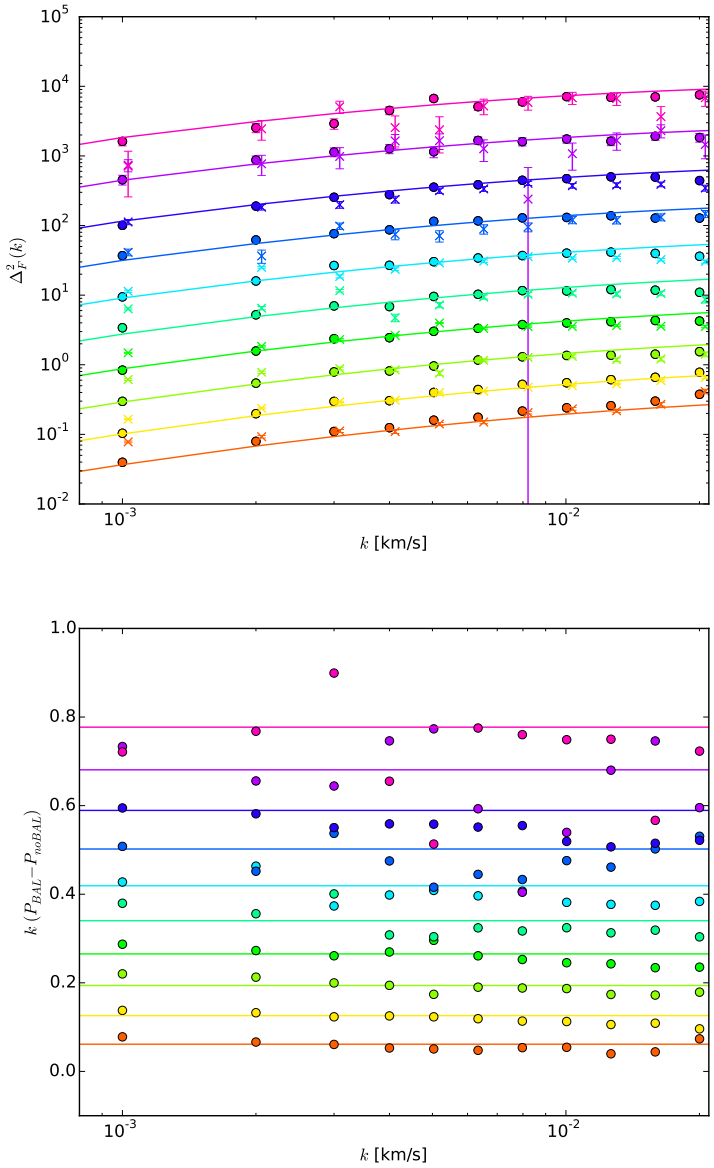

Figure 6.10 In the top plot, we show the power spectrum measurement on quasars with broad absorption lines ('x' points) and those without BALs (circle points) as compared to previous results from Palanque-Delabrouille et al. (2013) in solid lines. Error bars for the measurement on BAL quasars are larger because we are measuring the power spectrum on a smaller data set, approximately 10% the size of the measurement on quasars without BALs. In the bottom plot, we show power spectrum residuals, $P_{BAL} - P_{noBAL}$ comparing only quasars with BAL features and quasars without BAL features. It is evident that those with BAL features tend to underestimate the power, which is a result of an inaccurate measurement of the quasar continuum do to the BALs. Redshift bins are the same as those in Fig. 6.2.

We conduct a similar test, where we show the measurement between quasars with DLA systems and those with no DLA contaminants in Fig. 6.11. Again, one would think that Damped Lyman Alpha systems exhibit more absorption in the forest and would recover a stronger measurement of power for DLA quasars. We see that this is not true for the same reason as that for the BAL quasars. The continuum measurement step computes a lower measurement for the continuum, dividing out too much power in this region, leaving us with a lower power spectrum measurement for DLA quasars. As the redshift increases, the effect of the DLAs is more pronounced. This is probably due to a higher number of DLAs at higher redshift [Sánchez-Ramírez et al. (2015)].

As a last comparison, let us look at the difference between our own coadds made from *Primary* observations and *Combined* observations of single exposures. To reiterate, when we discuss *Primary* observations, we use all the exposures of *one* observation to create our coadds. For the *Combined* observations, we combine all exposures from *all* observations of the quasar. Not all quasars have multiple observations but it is possible that some quasars were observed multiple times, over different nights. Different observing conditions could result in different throughput of the spectra and, with future work, our individual throughput correction to each exposure should fix this. For our final product, that will be the case, but as of now, where we have not applied the single throughput correction, we may encounter differences in the power spectrum estimation. One such problem could be at the high-k, where a large variance between the resolution of exposures could mean that the weighted average resolution that we compute for our coadds is not an accurate enough representation. The first plot in Fig. 6.12 plots the two fits for *Primary* observations and *Combined* observations, and the bottom plot shows the residuals between the two. As expected, there is an effect on the high-k such that the *Combined* observations underestimate the power at the high-k end. Although we see this effect, it is difficult to make a decision about whether to only use the *Primary* observations. Of course, this is not ideal as we would be wasting multiple observations of the same quasar that reduces the noise in our spectra. To make a more informed decision, we should conduct another test when we apply a throughput correction

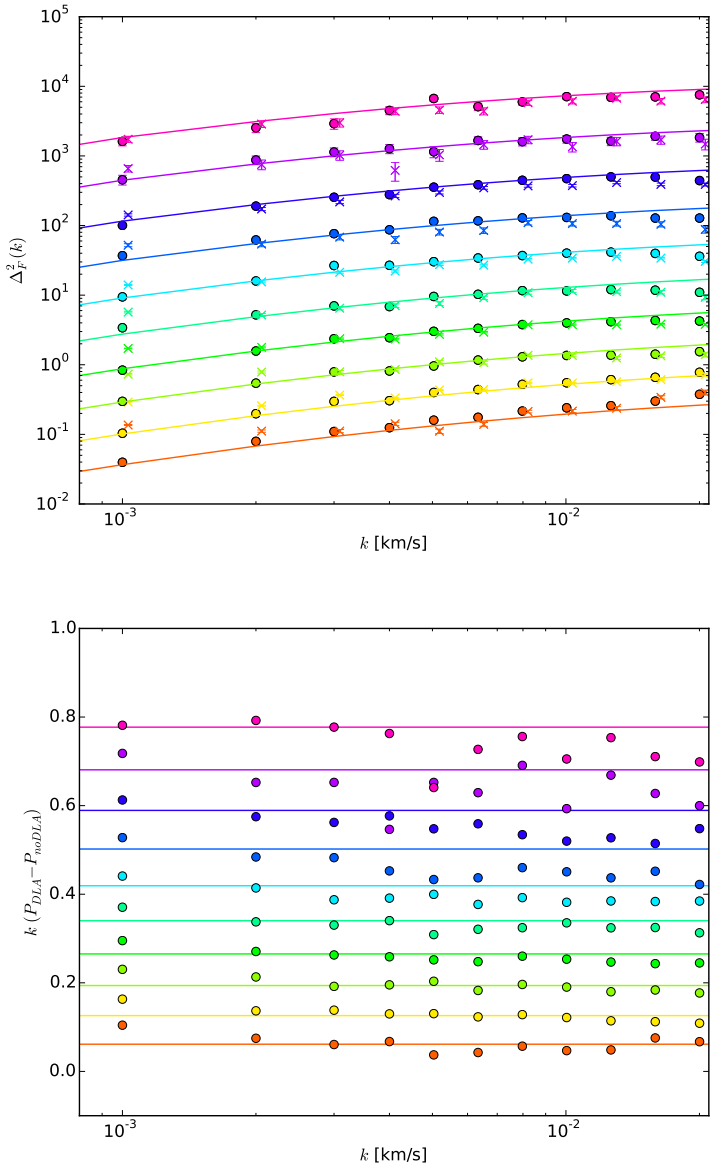

Figure 6.11 The top plot shows the power spectrum measurement of quasars with Damped Lyman Alpha systems ('x' points) and those without DLAs (circle points) with previous results from Palanque-Delabrouille et al. (2013) in solid lines. Error bars for those with DLAs are larger because we are using a smaller data set, approximately 10% of that without DLAs. The bottom plot shows residuals between power with DLAs and those without DLAs, $P_{DLA} - P_{noDLA}$ with previous results from Palanque-Delabrouille et al. (2013) in solid lines. Those with DLAs underestimate the power, which is a result of the pipeline measuring an inaccurate continuum fit due to the DLAs. Redshift bins are the same as those in Fig. 6.2.

exposure by exposure. We anticipate that the differences between the *Primary* observation power spectrum measurement and the *Combined* observation power spectrum measurement would be reduced.

Let us reiterate the tests we just conducted,

1. FFT vs Optimal Quadratic Estimator

2. Official Coadds vs. "Homemade" Coadds (with their respective noise correction factors)

3. Noise Correction for Homemade Coadds vs. No Noise Correction

4. Two Noise Correction factors for Homemade Coadds vs. One Noise Correction

5. BAL Quasars vs. Non-BAL Quasars

6. DLA Quasars vs. Non-DLA Quasars

7. Coadds made from *all* observations of a quasar vs. *Primary* observations

The tests conducted here all show that, indeed, there are noticeable effects when we change the methodology of our measurement. The next tests we should perform should ask how much of an effect this has on the power spectrum. Specifically, for the BAL and DLA quasars, ask how much contamination would have a noticeable effect on the power spectrum measurement. We can quantify this by comparing the χ^2 of the two fits. If the $\Delta\chi^2$ is over a certain threshold of tolerance, i.e. a level that would change the probability of the fit drastically, this means that we must be extremely careful to remove all the BAL or DLAs from our sample.

Our final, best fit result is plotted in Fig. 6.13-6.14, where we include two noise correction factors.

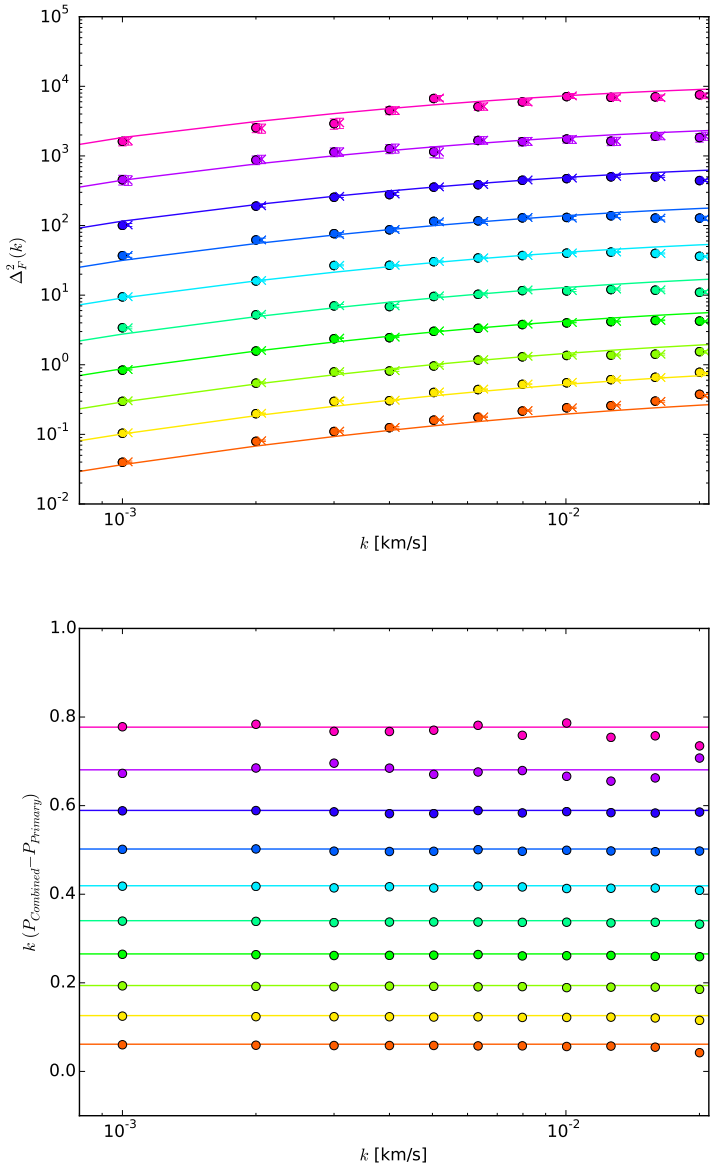

Figure 6.12 The top plot shows the power spectrum measurements for our own coadds made using all observations ('x' points) and those made using only the primary observation (circle points) with previous results from Palanque-Delabrouille et al. (2013) in solid lines. Creating coadds from all observations can include observations made over many nights, which would affect throughput and resolution measurements. The bottom plot shows residuals between power spectrum measurements, $P_{Combined} - P_{Primary}$, of coadds made from all observations and those made from primary observations. Effects from resolution are apparent at the high-k bin. Redshift bins are the same as those in Fig. 6.2.

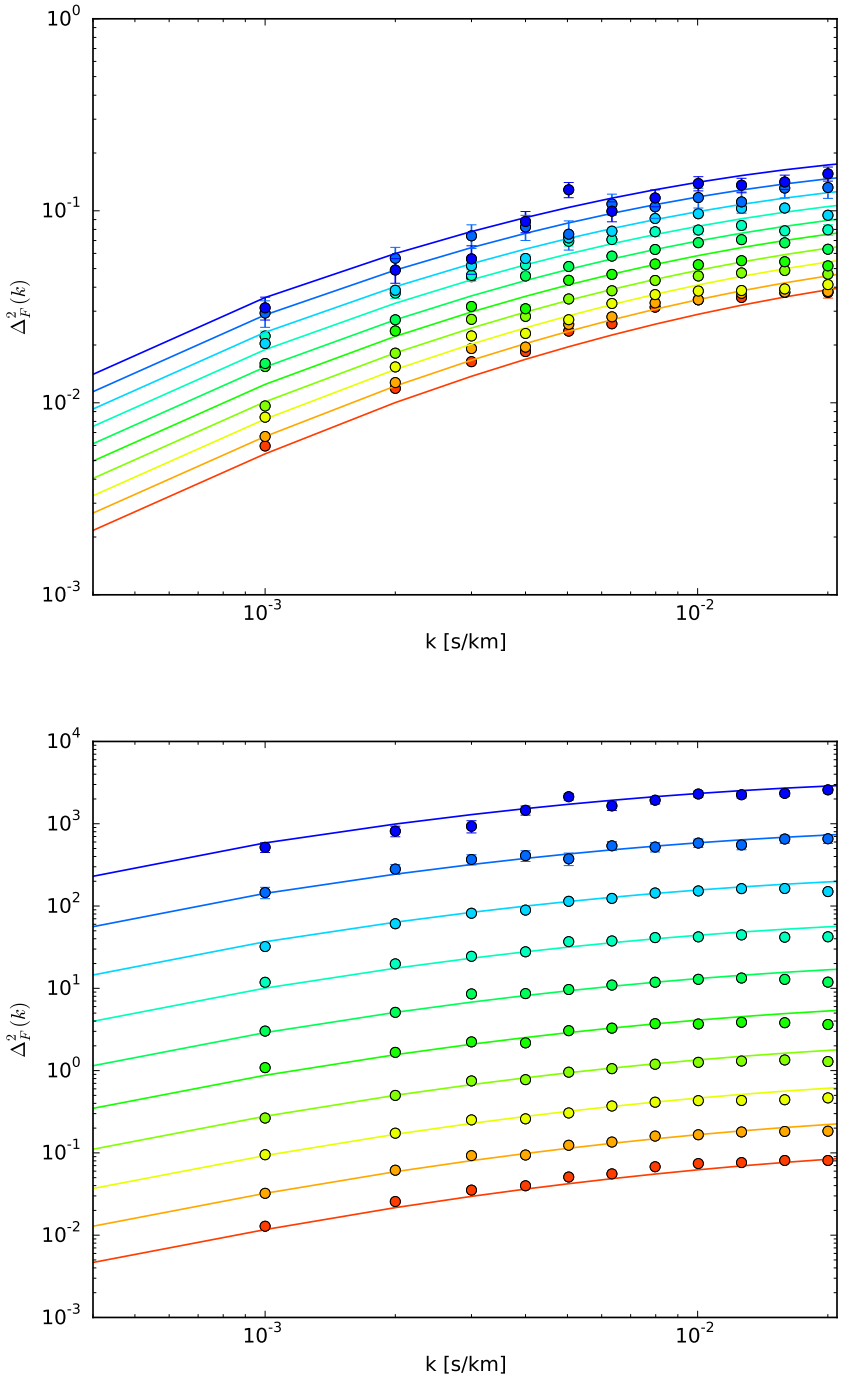

Figure 6.13 Our best fit result, using two noise correction factors applied to our own coadds. The bottom plot is the same as the top but with an offset to make it easier to distinguish between curves. Redshift bins are, from bottom to top, $z = 2.15, z = 2.3, z = 2.48, z = 2.66, z = 2.85, z = 3.04, z = 3.25, z = 3.47, z = 3.7, z = 3.94$.

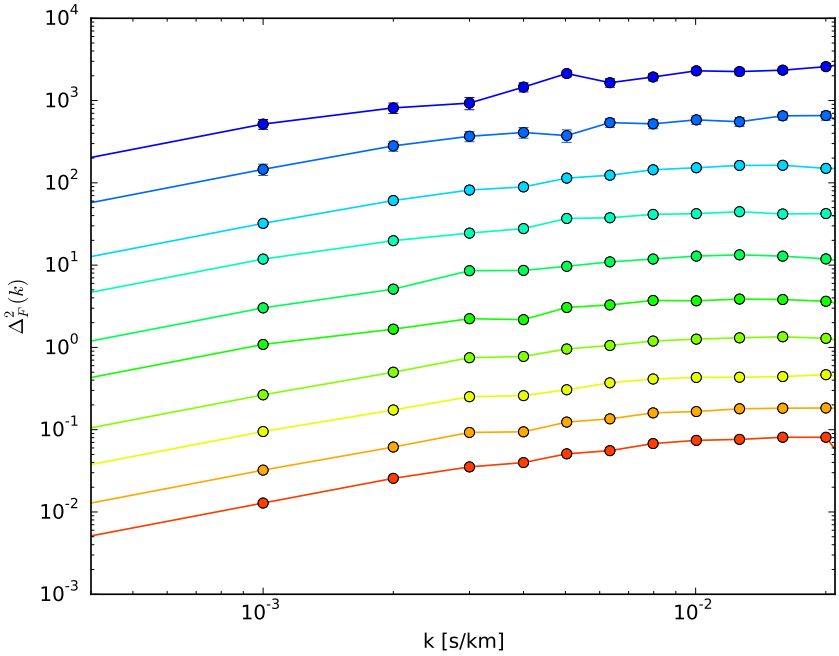

Figure 6.14 Same as the bottom plot of Fig. 6.13 but without the results from Palanque-Delabrouille et al. (2013) overplotted. Redshift bins are, from bottom to top, $z = 2.15, z = 2.3, z = 2.48, z = 2.66, z = 2.85, z = 3.04, z = 3.25, z = 3.47, z = 3.7, z = 3.94$.

Parameter	Sample 1	Sample 2	χ^2	d.o.f.
Signal to Noise Ratio	$7 < S/N < 14$	$S/N > 14$	198	110
Noise Correction Factor	$\chi^2/\nu < 1.1$	$\chi^2/\nu > 1.1$	217	110
Continuum $P(\chi^2)$	$0.01 < P(\chi^2) < 0.5$	$0.5 \leq P(\chi^2) < 0.99$	551	110

Table 6.1 We split the data into two sets to test for consistency between the power spectrum measurement. We return the χ^2 and degrees of freedom between the two measurements.

6.3 Split Data Tests

There are various aspects of our analysis that can affect our $P(z, k)$ measurement. We can identify systematic errors and discrepancies by splitting up the data set in two samples based on a given parameter and measure the χ^2 to quantify the degree of consistency. In Table 6.1, we list the various split data samples and the χ^2 and degrees of freedom for the measurement as computed by,

$$\chi^2 = (\mathbf{D_1} - \mathbf{D_2})^T (\mathbf{C}_1 + \mathbf{C}_2)(\mathbf{D_1} - \mathbf{D_2}) \tag{6.7}$$

where $\mathbf{D}$ and $\mathbf{C}$ refer to the power spectrum measurement and its covariance matrix, respectively.

Our first test is done by splitting the data based on S/N to test whether there is a dependence on noise for our fits. One of our samples has $7 < S/N < 14$ and the other has $S/N > 14$. We show this result in Fig. 6.15. This is a good test of whether we account for our noise properly. We can see that there is no large systematic offset between the two measurements, and with $\chi^2 = 198$ for 110 degrees of freedom, we exhibit decent consistency between the two data sets. Next, we compare the fits when we split on a noise correction factor of $\chi^2/\nu < 1.1$ and $\chi^2/\nu > 1.1$. This will test the validity of our noise correction factor and whether it is actually helps in quantifying the noise of our spectra. We plot the results in Fig. 6.16. There is no tendency for one data set to measure a systematically higher measurement than the other, yet since $\chi^2 = 217$ for 110 degrees of freedom, we may need to

conduct more split data tests to fully understand what is causing the discrepancy between the two measurements. For our last data split, we split based on the probability of χ^2 of

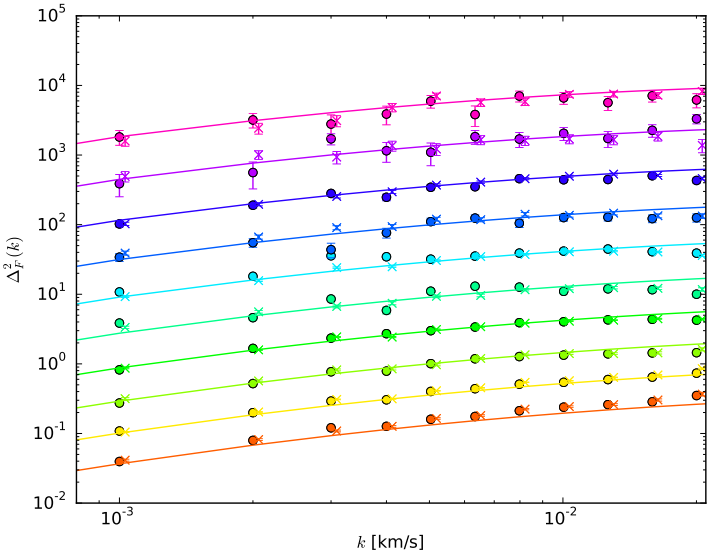

Figure 6.15 Tests done splitting up the data in two based on the signal to noise. Circle points refer to a data set with $7 < S/N < 14$, 'x' points refer to a data set with $S/N > 14$. The test has a $\chi^2 = 198$ for 110 degrees of freedom. Redshifts are defined in the caption of Fig. 6.4.

the continuum fit. Since we have cut our quasars with $P(\chi^2) < 0.01$, our two data samples consist of $0.01 < P(\chi^2) < 0.5$ and $0.5 < P(\chi^2) < 0.99$. The two fits are shown in Fig. 6.17. Note that specifically, this split data test does not display consistency between the two data sets. It will be necessary to delve into this issue further.

Splitting the data tests whether we are self consistent for our power spectrum measurement. It can be used to pinpoint areas of our analysis that are causing problems in our measurement. We have conducted three such split data tests, but there are others we should perform as well. A couple tests that would be useful are splitting based on resolution to quantify how the resolution really affects the high-k end of the power spectrum. Also, power

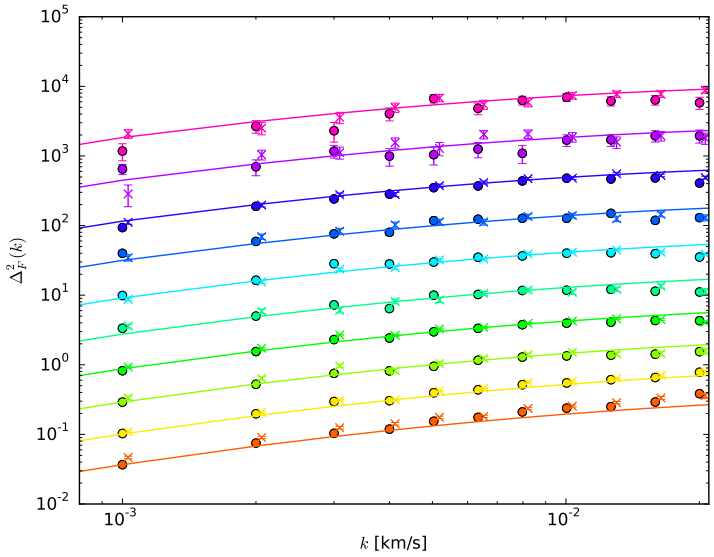

Figure 6.16 Tests done splitting up the data in two based on the noise correction factor for our own coadds. Circle points refer to a data set with $\chi^2/\nu < 1.1$ while 'x' points refer to the data set with $\chi^2/\nu > 1.1$. The test has a $\chi^2 = 217$ for 110 degrees of freedom. Redshifts are defined in the caption of Fig. 6.4.

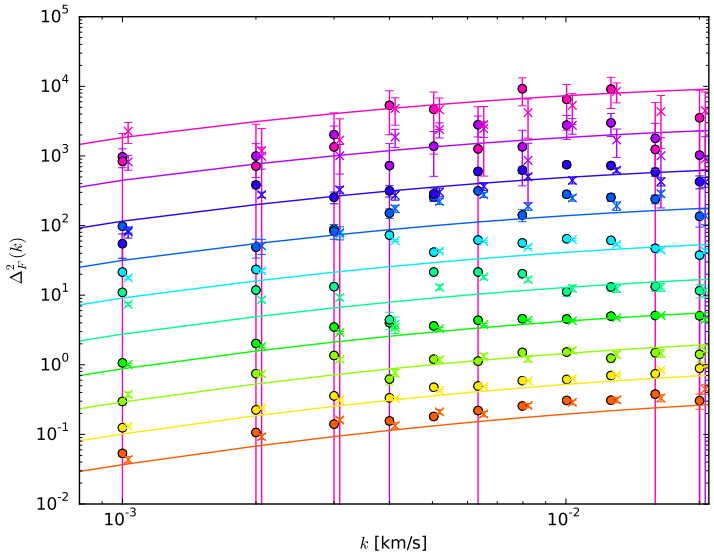

Figure 6.17 Tests done splitting up the data in two based on the probability of continuum fits. Circle points refer to data with $0.01 < P(\chi^2) < 0.5$. 'x' points refer to data with $0.5 \leq P(\chi^2) < 0.99$. $\chi^2 = 551$ for 110 degrees of freedom. Redshifts are defined in the caption of Fig. 6.4.

should be independent of the rest frame region of the quasar spectra. Splitting the data between the rest frame wavelengths of $1048 - 1113$ Å and $1113 - 1185$ Å should return similar results.

After performing these tests though, the next step is to test whether the discrepancy in measurements, if any, affects the cosmological parameters derived from the power spectrum. Therefore, future work should be conducted that tests the effect on cosmological parameters.

Chapter 7

CONCLUSIONS

We have presented the measurement of the 1D Power spectrum of the Lyα forest from a high quality quasar spectra sample from the Baryon Oscillation Spectroscopic Survey (BOSS). Specifically, our measurement includes the Fourier modes ranging from $0.0004 < k < 0.02$ s km^{-1} in the redshift range $2 < z < 4$. We restrict our sample to quasars with an average signal-to-noise (S/N) ratio > 7. In addition, we cut out quasars exhibiting Broad Absorption Lines (BAL) and those containing Damped Lyman Alpha (DLA) systems. We remove quasars where fits to the quasar continuum are not likely, in a probabalistic sense.

As BOSS pushes the limit in redshift range and S/N, we developed sophisticated techniques to remove contamination from the telescope, spectrograph, sky, and intrinsic quasar continuum to ultimately return the delta field of Lyα forest density fluctuations. We used the optimal quadratic estimator to parameterize the quasar continuum and mean flux of Lyα absorption as a function of redshift. Although the BOSS reduction pipeline computes the calibration vector to describe the effects of throughput of the telescope and spectrograph, we measured a throughput correction factor. BOSS also provides a measurement of the sky contribution to the quasar spectra, but we computed an additive correction factor to ensure a better representation of these effecs.

Our parameterization of quasar continuum consists of an individual quasar amplitude, a global continuum measurement as a function of rest wavelength, and a power law tilt per quasar. The region in the forest has an additional constant factor for normalization. The mean flux is measured as a function of redshift. Quasar spectra is divided by the product of quasar continuum, mean flux, and throughput, to return the density field of Lyα forest fluctuations.

We made mock spectra to test our code, which we built up from the simplest of mocks, to those accurately depicting the distribution of power observed in the Lyα forest. To mimic the Lyα forest, we used a parameterization of the power spectrum as developed by Palanque-Delabrouille et al. (2013), creating skewers of line of sight spectra from a Gaussian random field. Mocks were then used to not only ensure that we recover the power spectrum we used to create mocks, but also to test whether the code is seamless and can properly handle basic tasks such as reading the data from a quasar catalog, to culling the proper quasars and pixels, to finally returning the 1D power spectrum.

Mock spectra are limited in their ability to mimic observations perfectly. We know that in addition to Lyα absorption, contamination of our measurement can occur due to the absorption due to metals, such as SiIV and CIV. We measured this background contamination in the region $1268\mathring{A} < \lambda_{\rm rest} < 1380\mathring{A}$ which can also include residual throughput effects, and used the covariance from this power measurement when computing the power spectrum of the Lyα forest.

We created our own coadded spectra by combining exposures using the nearest grid point, ensuring that noise is not correlated between pixels. We computed a correction factor for the noise, determined by the sum of χ^2 and degrees of freedom of pixels between the coadded pixel and the contribution from exposures. Future work would include applying a per-exposure throughput correction when making the coadds.

We measured the power spectrum of Lyα fluctuations in the region $1041\mathring{A} < \lambda_{\rm rest} < 1185\mathring{A}$. We chose this region based on our measurements of variance for the theoretical predictions of our data model. In particular, we ensured that the continuum variance in this region is less than the variance of the Lyα forest. This continuum variance, and our measurements of variance of throughput correction and sky correction were also used to mask regions of the Lyα forest in which our prediction of theoretical values were not trustworthy.

We tested the effects of changing certain parameters in our power spectrum measurement, to assure ourselves that we understand the effects of these changes. Additional tests were conducted that split the data tests based on various parameters.

Our final measurement includes two correction factors, one constant scale factor applied to the data, and another noise correction as a function of wavelength. We recorded the errors on this measurement with the assumption that the likelihood function is Gaussian.

7.1 Completion of the Project

It is necessary to perform further tests to finalize this power spectrum measurement. Specifically, we still do not show completely consistent results during our split data tests. The main test, though, would be to use the power spectrum measurement to constrain cosmological parameters and test whether our differences in results for split data tests affect the derived cosmological parameters.

Our coadds are a better representation of a coadd made from many exposures than that from the BOSS reduction pipeline, but additional single exposure parameters, specifically those relating to throughput, would create even more accurate coadds. This would be the next step to take. Finally, we have not completely assessed whether the errors on our measurement are realistic and it would be helpful to use bootstrapping techniques to assess the real error bars.

This thesis has made strides in developing data analysis techniques necessary for studies of the Lyα forest that are not limited to the one dimensional power spectrum. Any analysis of the Lyα forest requires a thorough understanding of the quasar spectra, and techniques developed will be used in the future for a variety of projects. For example, this research group aims to measure the three dimensional power spectrum, which will use all techniques developed in this thesis.

In addition, this core software is developed such that adding data from future surveys will be relatively simple, as opposed to reinventing the wheel every time new data is obtained. We are looking forward to an additional 500,000 quasars from SDSS-IV, using the same BOSS spectrograph, and even further into the future with the Dark Energy Spectroscopic Instrument (DESI) which will obtain 3 million quasar spectra.

7.2 Lyman Alpha Forest with Future Surveys

The future of the Lyα forest is promising. Many new surveys are in development that will increase the ability to constrain cosmology using the Lyα forest. In the near future, two types of surveys will be integral in moving the Lyα forest studies forward. Combining information from these surveys, with CMB information from *Planck*, will prove to be extremely powerful.

The first of these surveys are spectroscopic redshift surveys, such as the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) and the Dark Energy Spectroscopic Instrument (DESI), which will provide many more quasar spectra with Lyα forest coverage, as mentioned above. In addition to the information directly from the Lyα forest, galaxy and quasar clustering provides us with point tracers of the underlying cosmic structure. Therefore, any increase in knowledge, whether directly related to Lyα forest or not, can be used in conjunction with Lyα forest studies to provide additional constraints. The second type of surveys are photometric surveys, such as the Dark Energy Survey (DES) and the Large Synoptic Survey Telescope (LSST). These will provide lensing, clustering, and cross-correlation information.

There has been considerable work done to predict the outcome of combining such surveys. Font-Ribera et al. (2014) performs Fisher matrix projections for the constraint of cosmological parameters using the above surveys, in addition to others. Font-Ribera et al. (2014) concludes that the sum of neutrino masses will be measured to $\sim 0.1 - 0.2$ eV, and the curvature parameter will be constrained to ± 0.001. Precise BAO measurements will be possible at redshifts greater than 2 and the Lyα forest power spectrum will be used to probe the running of the inflationary spectral index.

The future of the Lyα forest is promising. The work presented in this thesis is providing just the tip of the iceberg in data analysis techniques needed to fully use the Lyα forest to its full potential. As new data becomes available from telescopes and surveys, the future of cosmology lies in developing code that can be easily modified with the advent of more data and knowledge. My thesis has been just one step towards that ultimate goal.

Figures are reproduced by permission of the AAS.

BIBLIOGRAPHY

Abazajian, K. N., et al. 2009, ApJS, 182, 543, 0812.0649

Aguirre, A., Hernquist, L., Schaye, J., Weinberg, D. H., Katz, N., & Gardner, J. 2001, ApJ, 560, 599, astro-ph/0006345

Aihara, H., et al. 2011, ApJS, 193, 29, 1101.1559

Alam, S., et al. 2015, ApJS, 219, 12, 1501.00963

Anderson, L., et al. 2014, MNRAS, 441, 24, 1312.4877

Anderson, L., et al. 2012, MNRAS, 427, 3435, 1203.6594

Bahcall, J. N., Greenstein, J. L., & Sargent, W. L. W. 1968, ApJ, 153, 689

Bahcall, J. N., & Salpeter, E. E. 1965, ApJ, 142, 1677

Bi, H. 1993, ApJ, 405, 479

Bi, H., & Davidsen, A. F. 1997, ApJ, 479, 523, astro-ph/9611062

Bi, H., Ge, J., & Fang, L.-Z. 1995, ApJ, 452, 90, astro-ph/9504061

Bolton, A. S., et al. 2012, AJ, 144, 144, 1207.7326

Busca, N. G., et al. 2013, A&A, 552, A96, 1211.2616

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

Cole, S., et al. 2005, MNRAS, 362, 505, arXiv:astro-ph/0501174

Cowie, L. L., Songaila, A., Kim, T.-S., & Hu, E. M. 1995, AJ, 109, 1522

Croft, R. A. C., Weinberg, D. H., Bolte, M., Burles, S., Hernquist, L., Katz, N., Kirkman, D., & Tytler, D. 2002, ApJ, 581, 20

Croft, R. A. C., Weinberg, D. H., Katz, N., & Hernquist, L. 1997, ApJ, 488, 532, astro-ph/9611053

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

Croft, R. A. C., Weinberg, D. H., Pettini, M., Hernquist, L., & Katz, N. 1999, ApJ, 520, 1

Davé, R., Hellsten, U., Hernquist, L., Katz, N., & Weinberg, D. H. 1998, ApJ, 509, 661, astro-ph/9803257

Dawson, K. S., et al. 2013, AJ, 145, 10, 1208.0022

Eisenstein, D. J., et al. 2011, AJ, 142, 72, 1101.1529

Eisenstein, D. J., et al. 2005, ApJ, 633, 560

Font-Ribera, A., McDonald, P., & Miralda-Escudé, J. 2012, JCAP, 1, 1, 1108.5606

Font-Ribera, A., McDonald, P., Mostek, N., Reid, B. A., Seo, H.-J., & Slosar, A. 2014, JCAP, 5, 23, 1308.4164

Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, AJ, 111, 1748

Gamow, G. 1948, Physical Review, 74, 505

Gough, B. 2009, GNU Scientific Library Reference Manual - Third Edition (3rd ed.) (Network Theory Ltd.)

Gunn, J. E., et al. 1998, AJ, 116, 3040

Gunn, J. E., & Peterson, B. A. 1965, ApJ, 142, 1633

Gunn, J. E., et al. 2006, AJ, 131, 2332, arXiv:astro-ph/0602326

Guth, A. H. 1981, *Phys. Rev.* **D** , 23, 347

Hernquist, L., Katz, N., Weinberg, D. H., & Jordi, M. 1996, ApJ, 457, L51

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Hui, L., Gnedin, N. Y., & Zhang, Y. 1997, ApJ, 486, 599, astro-ph/9608157

Kim, T., Viel, M., Haehnelt, M. G., Carswell, R. F., & Cristiani, S. 2004, MNRAS, 347, 355, arXiv:astro-ph/0308103

Lemaître, G. 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Lynds, R. 1971, ApJ, 164, L73

Margala, D., Kirkby, D., Dawson, K., Bailey, S., Blanton, M., & Schneider, D. P. 2015, ArXiv e-prints, 1506.04790

McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., Cen, R., & Ostriker, J. P. 2000, ApJ, 543, 1

McDonald, P., et al. 2006, ApJS, 163, 80, arXiv:astro-ph/0405013

Meyer, D. M., & York, D. G. 1987, ApJ, 315, L5

Noterdaeme, P., et al. 2012, A&A, 547, L1, 1210.1213

Oppenheimer, B. D., & Davé, R. 2006, MNRAS, 373, 1265, astro-ph/0605651

Osmer, P. S. 1982, ApJ, 253, 28

Palanque-Delabrouille, N., et al. 2013, A&A, 559, A85, 1306.5896

Pâris, I., et al. 2014, A&A, 563, A54, 1311.4870

Penzias, A. A., & Wilson, R. W. 1965, ApJ, 142, 419

Perlmutter, S., et al. 1999, ApJ, 517, 565, arXiv:astro-ph/9812133

Planck Collaboration, et al. 2014, A&A, 571, A16, 1303.5076

Reisenegger, A., & Miralda-Escude, J. 1995, ApJ, 449, 476, astro-ph/9502063

Richards, G. T., et al. 2006, AJ, 131, 2766, astro-ph/0601434

Riess, A. G., et al. 1998, AJ, 116, 1009, arXiv:astro-ph/9805201

Ross, N. P., et al. 2012, ApJS, 199, 3, 1105.0606

Rubin, V. C., & Ford, W. K., Jr. 1970, ApJ, 159, 379

Sánchez-Ramírez, R., et al. 2015, ArXiv e-prints, 1511.05003

Sargent, W. L. W., Young, P. J., Boksenberg, A., & Tytler, D. 1980, ApJS, 42, 41

Schaye, J., Aguirre, A., Kim, T., Theuns, T., Rauch, M., & Sargent, W. L. W. 2003, ApJ, 596, 768

Schmidt, M., Schneider, D. P., & Gunn, J. E. 1995, AJ, 110, 68

Seljak, U. . 1998, ApJ, 506, 64

Simcoe, R. A., Sargent, W. L. W., & Rauch, M. 2004, ApJ, 606, 92, astro-ph/0312467

Slosar, A., et al. 2013, JCAP, 4, 26, 1301.3459

Smee, S. A., et al. 2013, AJ, 146, 32, 1208.2233

Smoot, G. F., et al. 1992, ApJ, 396, L1

Starobinsky, A. A. 1982, Physics Letters B, 117, 175

Tegmark, M. 1997, *Phys. Rev.* **D** , 55, 5895

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

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

Zhang, Y., Anninos, P., & Norman, M. L. 1995, ApJ, 453, L57

Zwicky, F. 1937, ApJ, 86, 217

Appendix A

SMOOTHING

The concept behind the Power Spectrum fitting is as follows. We need to account for smoothing caused by the pixel size and resolution, namely a top-hat function and Gaussian kernel, respectively. In real space, we need to convolve our delta field by a top-hat function and a Gaussian kernel. Assuming $W(r)$ corresponds to the convolution of a top-hat and Gaussian kernel in real space, we have the following relationships:

$$\delta_s(r) = \delta(r) \star W(r) \tag{A.1}$$

$$W(r) = W_{pixel}(r) \star W_{res}(r) \tag{A.2}$$

where $\delta_s(r)$ denotes the smoothed delta field at position r, $\delta(r)$ is the underlying delta field, and $W(r)$ is the smoothing function. Here, the $\star$ denotes a convolution. In Fourier space, the convolution is simplified to multiplication, therefore we can write $\delta_s(r)$ in Fourier space:

$$\tilde{\delta_s} = \tilde{\delta} W(k) \tag{A.3}$$

where $\tilde{\delta}$ denotes the Fourier transform of the delta field. The smoothing function will be the product of the Fourier transform of a top hat and Gaussian kernel. The top hat function has been transformed using a cosine transformation as follows:

$$W_{pixel}(k) = l^{-1} \int_{-l/2}^{l/2} \cos(kr)\, dr = \frac{\sin(kl/2)}{kl/2} \tag{A.4}$$

where l is the pixel width. The Fourier transform of the Gaussian kernel gives us:

$$W_{res}(k) = \frac{1}{R\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{r^2}{2R^2}}\ e^{-ikr}\ dr = e^{-\frac{k^2R^2}{2}} \tag{A.5}$$

where R is the resolution.

We then see that the covariance matrix, $\mathbf{C}$, or the correlation function, $\xi(r)$ can be defined as the Fourier transform of the Power spectrum:

$$\langle \tilde{\delta}_i \tilde{\delta}_j \rangle = P(k)W^2(k) \tag{A.6}$$

$$\mathbf{C} = \xi(r) = \langle \delta_i \delta_j \rangle = \int_{-\infty}^{\infty} P(k)W^2(k)\frac{\cos(kr)}{2\pi}\ dk \tag{A.7}$$

The diagonal values of the covariance matrix, $\mathbf{C}$ are given by $\xi(r=0)$. For our simple constant P case:

$$\sigma^2 = \xi(r=0) = \frac{P}{2\pi} \int_{-\infty}^{\infty} \frac{\sin^2(kl/2)}{(kl/2)^2}\ e^{-k^2R^2}\ dk \tag{A.8}$$

In this simple case of $P =$ constant, the derivative, $d\xi(r)/dP$ is simple:

$$\frac{d\xi(r)}{dP} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin^2(kl/2)}{(kl/2)^2}\ e^{-k^2R^2}\ \cos(kr)\ dk \tag{A.9}$$

Appendix B

MEASURED ERROR ON POWER SPECTRUM

To assess the accuracy of our resulting power spectrum, we can compare the error on our measurement to that which we expect from theory. Our measured power spectrum can be written as:

$$P_{meas}(k) = \frac{1}{N_k} \sum_{i=1}^{N_k} |\delta(\mathbf{k}_i)|^2 \Bigg|_{|\mathbf{k}_i - k| \leq k_{fun}} \tag{B.1}$$

where k_{fun} is the fundamental Fourier mode and N_k is the number of independent Fourier modes available per bin. Taking the expectation value of this, we obtain the underlying power spectrum,$P(k)$,

$$\langle P_{meas}(k) \rangle = \frac{1}{N_k} \sum_{i=1}^{N_k} \langle |\delta(\mathbf{k}_i)|^2 \rangle \Bigg|_{|\mathbf{k}_i - k| \leq k_{fun}} = \langle |\delta(k)|^2 \rangle = P(k). \tag{B.2}$$

We can solve for the variance of our measured power spectrum by computing

$$\left\langle \left(\frac{P_{meas}(k) - P(k)}{P(k)} \right)^2 \right\rangle = 1 - 2\frac{\langle P_{meas} \rangle}{P(k)} + \frac{1}{N_k^2 P(k)^2} \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} \langle \delta^*(\mathbf{k}_i)\delta(\mathbf{k}_i)\delta^*(\mathbf{k}_j)\delta(\mathbf{k}_j) \rangle \tag{B.3}$$

Since we are assuming our density field is Gaussian with variance described by $P(k)$, the following relationship applies:

$$\langle \delta_i^* \delta_j \rangle = P(k)\delta_{ij}. \tag{B.4}$$

We use Wick's theorem to solve the double summation.

$$\sum_{i=1}^{N_k}\sum_{j=1}^{N_k}\langle\delta_i^*\delta_i\delta_j^*\delta_j\rangle = \sum_{i=1}^{N_k}\sum_{j=1}^{N_k}[\langle\delta_i^*\delta_i\rangle\langle\delta_j^*\delta_j\rangle + \langle\delta_i^*\delta_j\rangle\langle\delta_j^*\delta_i\rangle + \langle\delta_i^*\delta_j^*\rangle\langle\delta_i\delta_j\rangle] \quad \text{(B.5)}$$

$$= N_k^2[P(k])^2 + N_k[P(k)]^2 \quad \text{(B.6)}$$

Solving for the variance:

$$\sigma^2_{P(k)} = \langle[P_{meas}(k) - P(k)]^2\rangle = \frac{[P(k)]^2}{N_k} \quad \text{(B.7)}$$

and the standard deviation:

$$\sigma_{P(k)} = \langle[P_{meas}(k) - P(k)]^2\rangle^{1/2} = \sqrt{\frac{1}{N_k}}P(k). \quad \text{(B.8)}$$

Since our power spectrum is symmetric such that $\delta^*(\mathbf{k}) = \delta(-\mathbf{k})$, the number of independent k-modes is half of the number of available modes at a given k.

$$N_k = \frac{1}{2}\frac{\Delta k}{k_{fun}} \quad \text{(B.9)}$$

where $k_{fun} = 2\pi/L$ where L is the length of the Lyα forest region and Δk is the size of the bin.

To extend this test to multiple spectra, whether mocks or spectra, we can multiply $N_k * N_q$, the number of quasars, for the number of total modes available in the data set. We then must define what $P(k)$ will be.

$$P(k) = (P_F + P_N)W^2 \quad \text{(B.10)}$$

$$= P_{damp} + P_N W^2 \quad \text{(B.11)}$$

Here, P_F is our input power, P_N is noise power and W is the smoothing function given by the product of Eqns.A.4 - A.5. P_{damp} is the product of our input power and the smoothing function and noise power is defined

$$P_N = l\sigma_N^2 \tag{B.12}$$

where l is the pixel width and σ_N is noise on a pixel, whether its derived from the pipeline for real data or an assigned value for mocks. For our white mocks, this is defined by a constant inverse variance on each pixel.

Error on our measurement is:

$$\sigma_{P(k)} = \frac{1}{N_{modes}} P(k) \tag{B.13}$$

Appendix C

TESTING ξ TRANSFORMATION CODE

Using our white noise mocks, we can do a variety of tests. First, we will test whether our ξ transformation code works. We transform a constant P with only pixel smoothing to ξ. The following plots, Figs. C.1 and C.2, show the comparison of transformations done with different dk and k_{max} values for the integration using a cosine transformation.

For our constant power case, we expect $\xi(r)$ to be 0 for $r > w$. We can take a look at the ringing that occurs, as the transformation will not return exactly zero. Fig. C.3 compares the ringing for various values of dk and k_{max}.

The next step would be to transform the power spectrum that has both pixel smoothing by a top-hat function and resolution smoothing by a Gaussian kernel.

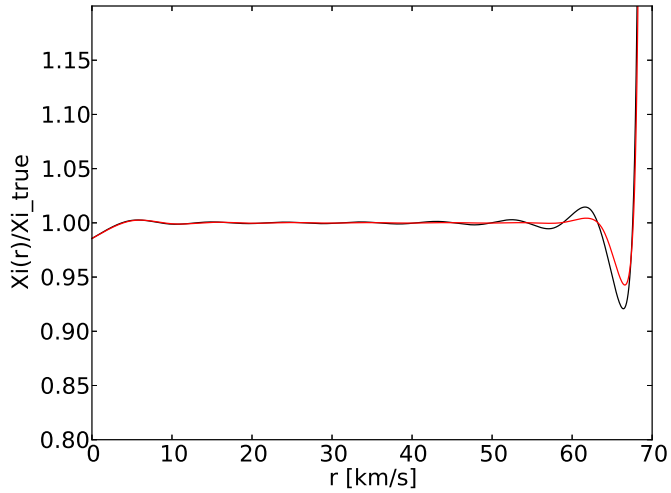

Figure C.1 We compare the ratio of $\xi(r)/\xi_{true}$ for $dk = 1e-5$ (black) and $dk = 0.01$ (red) with $k_{max} = 1.0$.

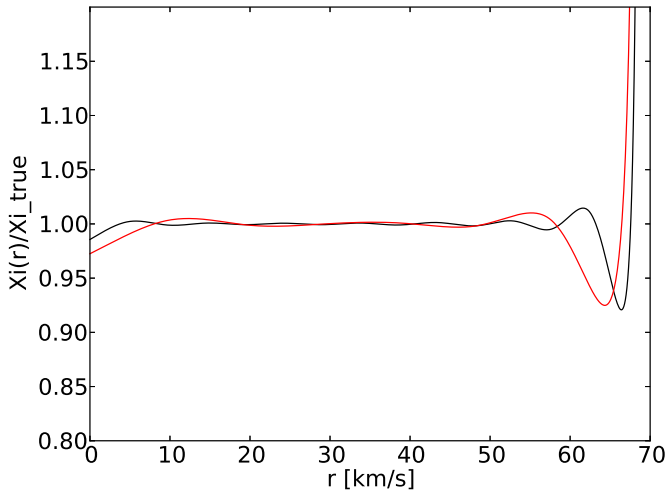

Figure C.2 We compare the ratio of $\xi(r)/\xi_{true}$ for $k_{max} = 1.0$ (black) and $k_{max} = 0.5$ (red) with $dk = 1.e-6$.

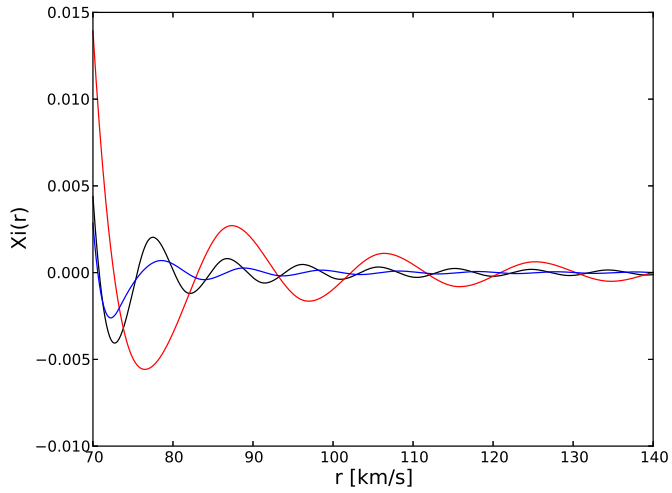

Figure C.3 We can see the ringing of the computed $\xi(r)$ for values of $r > w$. The black curve represents $dk = 1e-5$ and $k_{max} = 1.0$, blue curve plots $dk = 0.01$ and $k_{max} = 1.0$ and the red curve plots $dk = 1.e-5$ and $k_{max} = 0.5$.

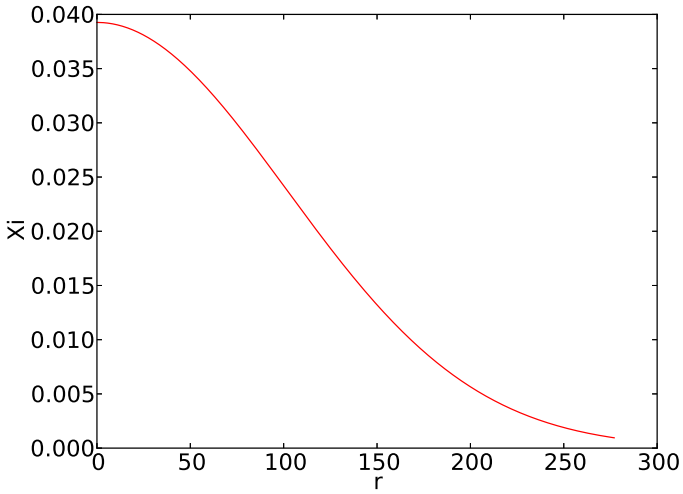

Figure C.4 For a pixel width of $l = 69$ km/s and resolution of $R = 69$ km/s, transformation of $P(k)$ to $\xi(r)$.

VITA

Vaishali Bhardwaj was born in Mountain View, California on July 24, 1985.