UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MEASURING COSMOLOGICAL PARAMETERS USING MATTER POWER SPECTRUM FROM GALAXY REDSHIFT SURVEY DATA

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

PRASAD HEMANTHA MADDUMAGE DON
Norman, Oklahoma
2013

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/7902

MEASURING COSMOLOGICAL PARAMETERS USING MATTER POWER SPECTRUM FROM GALAXY REDSHIFT SURVEY DATA

A DISSERTATION APPROVED FOR THE
HOMER L DODGE DEPARTMENT OF PHYSICS AND ASTRONOMY

BY

Dr. Yun Wang, Chair

Dr. Karen Leighly

Dr. Edward Baron

Dr. Ronald Kantowski

Dr. S. Lakshmivarahan

© Copyright by PRASAD HEMANTHA MADDUMAGE DON 2013
All Rights Reserved.

This dissertation is dedicated to my late father, Hemachandra Madduamge Don,

and my mother, Rupa Perera, for their unconditional love and support

Acknowledgements

First, I thank my adviser, Dr. Yun Wang, for giving me the opportunity to work on this challenging project and for her support. I thank my adviser for encouraging me to explore beyond my research topic and learn new concepts and technologies: My life today would be much different without this opportunity. Also, I thank the Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma for providing me with this opportunity and the support. I would like to express my gratitude to Dr. Karen Leighly for providing me the opportunity to explore and learn new ideas and technologies and for supporting me during various stages of my graduate student career. I also thank Dr. Edward Baron and Dr. Ronald Kantowski for letting me expand my knowledge even further. I thank my outside committee member, Dr. S. Lakshmivarahan, for his time serving on my committee. I also thank my colleague, Dr. Chia-Hsun Chuang, for his friendship and help when I really needed it.

A special thank goes to Dr. Will Percival, University of Portsmouth, UK for his guidance and support during the initial stage of my project, although I never got the opportunity to meet him in person. Also, I am grateful for the discussions I had with Dr. Gert Hütsi, Tartu Observatory, Estonia.

I thank Dr. Henry Neeman and the staff of the Oklahoma Supercomputing Center for Education and Research (OSCER) for helping me when needed and for making many learning opportunities freely available.

On a personal note, I am grateful to my parents for who I am today and for

making my success their dream, the only dream, and the only thing they ever wanted in their life: I promise I never let you down. I thank the love of my life, my wife, Dhrshani Bopege, for her support and understanding during the good and the bad, the successes and the failures of my life. Siyara, thank you for sacrificing valuable father-daughter time for the completion of this project.

I wholeheartedly thank my motherland, Sri Lanka, and the people for the free education I received during the entire time I was a student. I will never be able to graduate from college without this support. I sincerely thank my life long friends Suneth Fernado and Tharaka Gamage for their friendship and understanding. I also thank all of my Sri Lankan friends in Norman, Oklahoma for simply being there for each other. I very much appreciate the spiritual guidance provided by Rev. Higulwala Piyarathana and other monks at Oklahoma Buddhist Center. Last but not least, I am grateful to the United States for giving students like me who would never be able to afford graduate education without help, the opportunity of making their dreams a reality.

This work was supported in part by DOE grants DE-FG02-04ER41305 and DE-SC0009956.

Table of Contents

List of Tables **vii**

List of Figures **ix**

1 Introduction **1**
1.1 Effect of Matter Density Variations on the Power Spectrum 2
1.2 Galaxy Redshift Surveys 5
1.2.1 Luminous Red Galaxies (LRGs) 8

2 One Dimensional Galaxy Power Spectrum using Direct Fourier Method **14**
2.1 Theory 14
2.1.1 Matter Overdensity Field 14
2.1.2 Relationship between Correlation Function and Power Spectrum . 15
2.1.3 Selection Functions 17
2.1.4 Estimation of Power Spectrum 17
2.1.5 Proof of FKP Estimator of the Power Spectrum 19
2.1.6 Finding Optimum Weights $w(\mathbf{r})$ 24
2.1.7 Window Function 29
2.1.8 Summary of the FKP Method 30
2.2 Obtaining Galaxy Power Spectrum using 2dFGRS and SDSS Data 31
2.2.1 Method 31
2.3 Results and Discussion 33

3 Measurement of $H(z)$ and $D_A(z)$ from the Two Dimensional Power Spectrum of Sloan Digital Sky Survey luminous red galaxies **39**
3.1 Introduction 39
3.2 Data 41
3.3 Methodology 42
3.3.1 2D Galaxy Power Spectrum Estimation 42
3.3.2 Theoretical Model 44
3.3.3 Window Matrix 46
3.3.4 Covariance Matrix 50
3.3.5 Likelihood 52
3.4 Results 53
3.4.1 Validating the Method Using Mock Data 54
3.4.2 Constraints on Parameters from SDSS Data 56
3.5 Conclusion and Discussion 59

4 Summary **69**

List of Tables

1.1 LRGs are defined using colors. The definition consists of two color cuts depending on redshift. 9

3.1 LasDamas mock catalog fitting results. Each mock catalog is divided into five patches, and each patch is analyzed separately. The estimated parameters from each mock is the weighted average of the estimates from the patches. The mean and standard deviation are obtained by averaging over 80 mock catalogs. 55
3.2 Same as Table 3.1, but for dividing each mock into 10 patches. 56
3.3 Results from our analysis of SDSS DR7 LRGs. The mean values and standard deviations are calculated from the mean parameter values and covariance matrices obtained by fitting parameters for the 5 patches. . . . 58
3.4 Normalized average covariance matrix corresponding to Table 3.3. 58

List of Figures

1.1 The effect of cold dark matter (Ω_c) on the power spectrum is shown here. These three power spectra were generated for $\Omega_c = 0.1, 0.2$, and 0.3 as depicted by dashed line, dotted line, and dot dashed line, respectively. The maximum of the spectrum shifts to higher k as Ω_c increases. $\Omega_b = 0.05$ is assumed. 3

1.2 The baryon density, Ω_b, determines the BAO peak positions on the power spectrum. BAOs become stronger as the baryon density increases. $\Omega_c = 0.2$ is assumed. 4

1.3 A summary of different galaxy redshift surveys. Diagonal lines represent the total number of redshifts measured in the survey. Magnitude limited surveys are plotted with squares. Surveys that use photographic redshifts are shown with circles and highly selected surveys such as BOSS where only LRGs are targeted are plotted with triangles. Adapted from Ivan K. Baldry (`http://www.astro.ljmu.ac.uk/\nobreakspace{}ikb/research/galaxy-redshift-surveys.html`) 7

1.4 This two dimensional projection of SDSS shows that it has a broader sky coverage and deeper redshift coverage compared to 2dFGRS. 10

1.5 Top: Redshift coverage map of the 2dFGRS. Bottom: Sky coverage of 2dFGRS. 11

2.1 The radial selection function of 2dFGRS showing the number of galaxies vs. redshift. The smooth curve is a best fit (Colless et al. (2001)). The galaxy count decreases with redshift as the probability of detection reduces. 18

2.2 Comparison of 2dFGRS and SDSS window functions. The SDSS window is more compact compared to 2dFGRS window as SDSS has more redshift coverage (deep). 35

2.3 2dFGRS power spectrum obtained by direct Fourier technique compared to the 2dF group published galaxy spectrum. Two spectra deviate from each other at large scale due to the effect of window function as Cole et al. (2005) corrected the window effect by division of a factor determined by model power spectrum and its convolution. This step was omitted in our study as it was only effective for plotting. 36

2.4 Comparison of SDSS power spectrum obtained by direct Fourier technique and published results by SDSS consortium. They have divided the power spectrum by a smooth theoretical spectrum which does not have BAO oscillations and we also used the same method in this comparison. 37

3.1 This is a plot of SDSS DR7 LRG galaxy sample using a Sanson-Flamsteed projection. The five patches we use are shown. Note that the coordinates are not equatorial (RA, Dec). From left to right, patches 1-3 are the lower panels, and patches 4 and 5 are the upper panels. 43

3.2 Top: Comparison of the average of 160 LasDamas 2D galaxy power spectra (solid lines) and our model 2D power spectrum convolved with the appropriate window function (dotted lines). Model parameters are set to the LasDamas input values. Bottom: Average 2D power spectrum from SDSS DR7 LRGs (solid lines). All five power spectra from different patches were averaged to obtain a smooth plot. The best fit model corresponding to the parameters listed in Table 3.3, convolved with window functions of five patches and averaged together, is plotted with dashed lines. 47
3.3 Two dimensional window fuction obtained by combining five different window functions with different bin sizes (note that different contours start at different values as a resullt). The contour levels are logarithmic from 10^7 to 10. Also note that this combining was done only for visualization purposes. 49
3.4 Covariance matrices for SDSS data set(top) and LasDamas mock data(bottom). Both covariance matrices are calculated for the patch 1. These matrices are created by unrolling the actual 2D array of points; There are 154 points inside the area of interest. 51
3.5 LasDamas fitting results for the parameters $D_A^{\rm fid}(0.35)/D_A(0.35)$ (top left), $D_A(0.35)/r_s(z_d)$ (top right), $H(0.35)/H^{\rm fid}(0.35)$ (lower left), $H(0.35)r_s(z_d)/c$ (lower right). Dashed lines represent mean values and 1σ error bars and input parameter values are plotted with dotted lines. 60
3.6 The 1D marginalized probability distribution functions and 2D joint confidence contours of the primary parameters in our analysis of SDSS DR7 LRG sample at patch 1. 61
3.7 Same as Fig.3.6, but for patch 2 62
3.8 Same as Fig.3.6, but for SDSS patch 3 63
3.9 Same as Fig.3.6, but for SDSS patch 4 64
3.10 Same as Fig.3.6, but for SDSS patch 5 65

Abstract

The galaxy power spectrum is a powerful tool used to study cosmic large scale structure and dark energy. A method to obtain the one dimensional galaxy power spectrum using the direct Fourier method is presented and this technique was applied to two degree field Galaxy Redshift Survey (2dFGRS) data and Sloan Digital Sky Survey (SDSS) data separately. Although this method is slow compared to fast Fourier transform, it is more accurate at small scales. The two dimensional matter power spectrum was only obtained using surveys with narrow sky coverage until now. The method presented here is applicable to any current or future survey with wider sky coverage. A method to measure the Hubble parameter $H(z)$ and the angular diameter distance $D_A(z)$ simultaneously from the two-dimensional matter power spectrum is also presented. The method is validated by applying it to the LasDamas mock galaxy catalogs. Then, this method is applied to Sloan Digital Sky Survey (SDSS) Data Release 7 to measure two dimensional galaxy power spectrum and obtain measurements of $\Omega_m h^2 = 0.1268 \pm 0.0085$, $H(z = 0.35) = 81.3 \pm 3.8$km/s/Mpc, $D_A(z = 0.35) = 1037 \pm 44$Mpc, without assuming a dark energy model or a flat universe. The derived parameters $H(0.35)r_s(z_d)/c = 0.0431 \pm 0.0018$ and $D_A(0.35)/r_s(z_d) = 6.48 \pm 0.25$ are also measured and these are in excellent agreement with similar measurements from the two-dimensional correlation function of the same data.

Chapter 1

Introduction

Studying the large scale structure of the Universe is one of the key goals of contemporary cosmology. One possible path for large scale structure studies is through studying galaxy clustering. First we assume that galaxy distribution is a good indicator of the underlying matter distribution of the Universe, i.e. galaxy rich areas of the Universe can be treated as high matter density areas and voids as low density regions. Then, the galaxy power spectrum can be used to study fractional matter density contributions on different scales. Peebles (Yu & Peebles (1969), Peebles (1973), Peebles & Hauser (1974), and Peebles (1980)) pioneered the statistical analysis of galaxy catalogs in order to obtain matter density correlation and the power spectrum. With the information from increasingly large galaxy surveys such as Sloan Digital Sky Survey (SDSS), cosmologists can infer much more information today. As we will see, the power spectrum contains all the information about large scale structure and therefore is useful in measuring the properties of the Universe.

This chapter presents a brief introduction to the galaxy power spectrum followed by an overview of different galaxy redshift surveys with emphasis on the data used in this work. Chapter 2 contains a detailed proof of some of the key ideas and techniques as well as the results obtained by applying these techniques to obtain the one dimensional power spectrum. The next chapter presents the results obtained using the two dimensional power spectrum of the SDSS LRG sample.

1.1 Effect of Matter Density Variations on the Power Spectrum

The density variations we see today in the Universe can be traced back to the quantum fluctuations produced during inflation in the very early Universe, according to the standard theory of the structure formation in the Universe. These density perturbations created during inflation are very small. Gravity slows the expansion of high density regions causing the density contrasts to grow larger as the Universe evolves (Lyth, Liddle & Ma (2010)). This subsequent evolution of tiny density changes through gravity, eventually gave the Universe its present structure. Therefore, understanding the structure of the Universe will also reveal clues about its early stage. As long as these density perturbations evolve linearly, inflation theory predicts they are Gaussian distributed (Linde & Mukhanov (1997); Mukhanov & Chibisov (1981)). So far, there has been no conclusive evidence to contradict the assumption that observed matter density variations are Gaussian (Komatsu et al. (2003)). If this condition holds, all the information about matter density variations are contained in its power spectrum.

Stability of matter fluctuations is closely related to the Jeans scale: Perturbations smaller than the Jeans scale do not collapse due to pressure support while those larger grew due to gravity at the same rate, independent of scale (Sunyaev & Zeldovich (1970)). In a Universe with dark matter and radiation only, the Jeans scale grows to the size of the horizon at matter-radiation equality, and then reduces to zero when matter dominates. Therefore, the horizon scale at matter-radiation

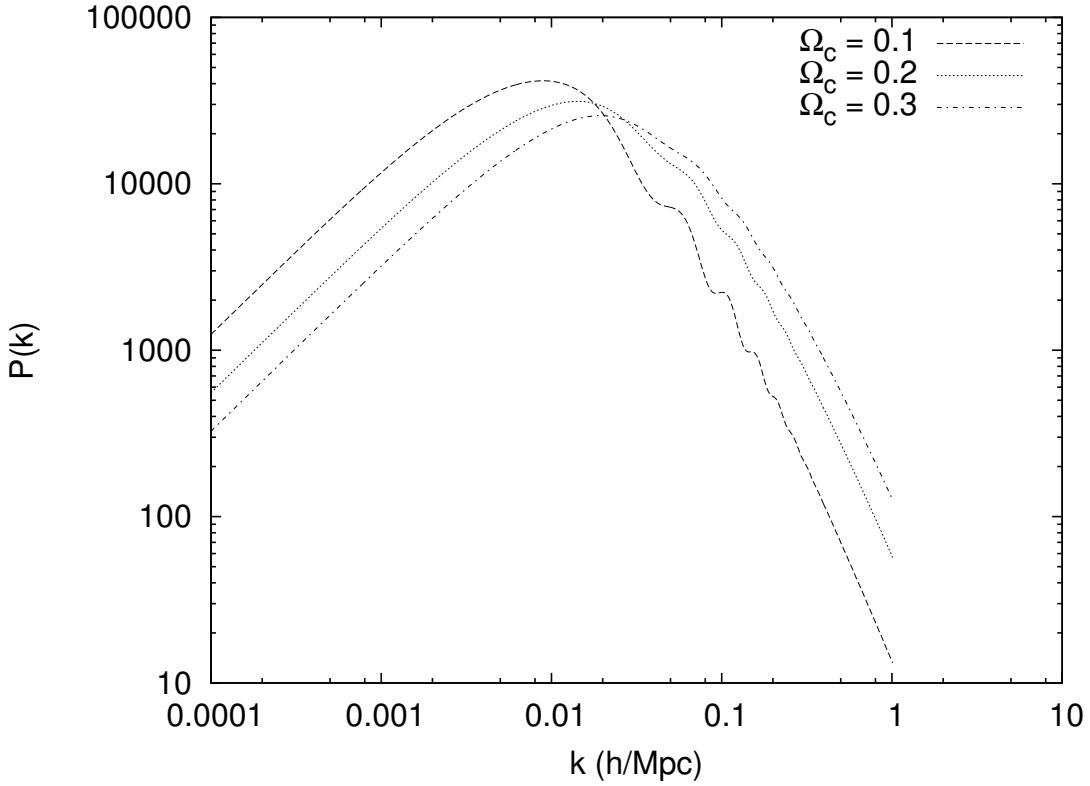

Figure 1.1: The effect of cold dark matter (Ω_c) on the power spectrum is shown here. These three power spectra were generated for $\Omega_c = 0.1, 0.2$, and 0.3 as depicted by dashed line, dotted line, and dot dashed line, respectively. The maximum of the spectrum shifts to higher k as Ω_c increases. $\Omega_b = 0.05$ is assumed.

equality will be imprinted in the power spectrum as a cutoff of small scale power and this allows a measurement of $\Omega_m h$ where $h = H_0/100$ (Longair (2008)). In other words, as Fig. 1.1 shows, the position of the maximum of power spectrum is sensitive to cold dark matter density Ω_c.

Although small, baryons too have perturbations and therefore leave a measurable imprint on the power spectrum. Eisenstein & Hu (1998) studied models

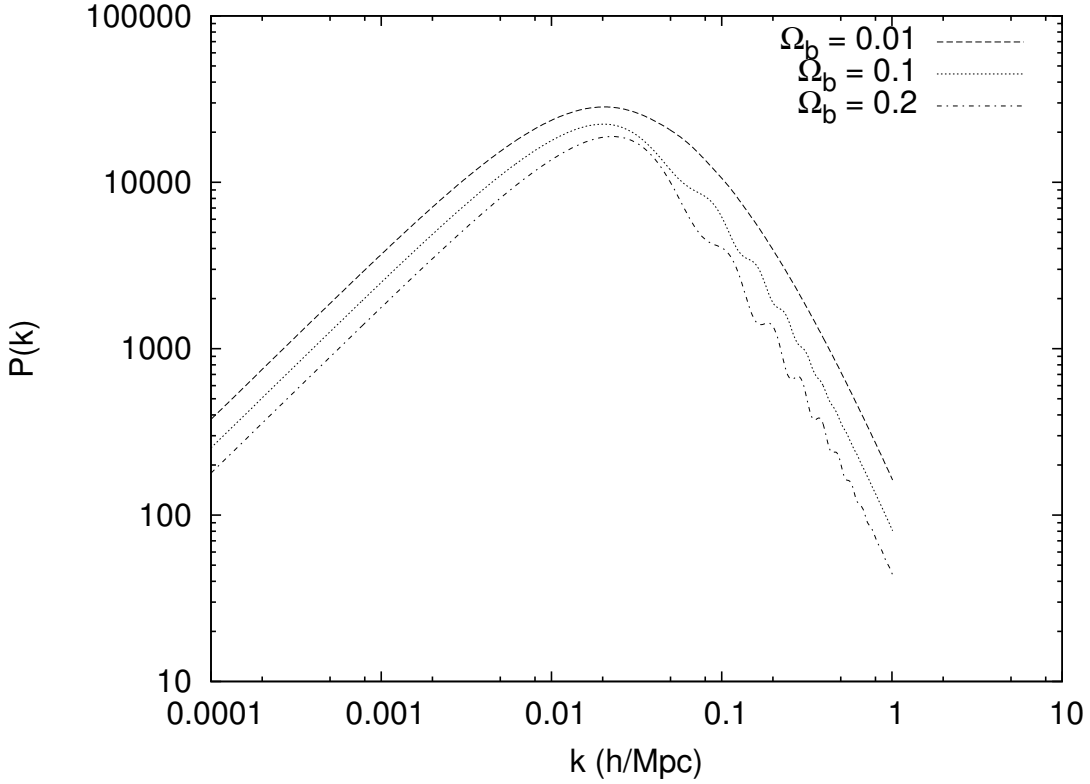

Figure 1.2: The baryon density, Ω_b, determines the BAO peak positions on the power spectrum. BAOs become stronger as the baryon density increases. $\Omega_c = 0.2$ is assumed.

with dark matter and baryons which provided greater insight on finding baryonic signature in the power spectrum. Decoupled from photons there remains a shell of baryonic matter around dark matter density fluctuations. Later they grow together due to gravity and these shell structures appear as oscillations in the power spectrum, which are known as baryon acoustic oscillations. The effect of various baryon content on the power spectrum is shown in Fig. 1.1.

The simplest theory of inflation predicts that the matter power spectrum is of the form $P(k) \propto k^n$ where $n \approx 1$ (Harrison (1970); Zeldovich (1972)). Combining the effect of dark matter, baryons, and neutrons on the power spectrum, various cosmological models can be fitted to the power spectrum to find parameters.

It can be shown that the power spectrum $P(\mathbf{k})$ and the galaxy correlation function $\xi(\mathbf{s})$ are Fourier pairs (see section 2.1.2). Therefore, studying the power spectrum is mathematically equivalent to studying the correlation function. However, as we shall see, our understanding of both the power spectrum and the correlation function depends on galaxy redshift surveys. These are volume limited samples of the Universe and hence result in $P(\mathbf{k})$ and $\xi(\mathbf{s})$ that are not mathematically equivalent. Therefore, it is important to study both the correlation function and the power spectrum, because they have different systematic uncertainties.

1.2 Galaxy Redshift Surveys

Galaxy redshift survey data are essential for contemporary precision cosmology as they provide a method to study large-scale structure of the Universe with increas-

ing accuracy as the number of galaxies included grow exponentially. Early surveys such as Canada-France Redshift Survey(CFRS) contained only 591 galaxies (Lilly et al. (1995)), Harvard-Smithsonian Center for Astrophysics 2 (CfA2) survey contained 19,369 galaxies (Falco et al. (1999)), Las Campanas Redshift Survey (LCRS) consists of 26,418 redshifts of galaxies (Shectman et al. (1996)), and Point Source Catalog redshift (PSCz) survey measured redshifts of 15,411 galaxies (Saunders et al. (2000)) using Infra-Red Astronomical Satellite(IRAS). Most of these are all sky surveys. Recent efforts such as the 2dF Galaxy Redshift Survey (2dFGRS) measured redshifts of 221,414 galaxies (Colless et al. (2003)), WiggleZ survey measured 238,770 galaxy redshifts (Parkinson et al. (2012)), and SDSS obtained redshift of 929,555 galaxies in the seventh data release, DR7, (Abazajian et al. (2009)). The SDSS-III Baryon Oscillation Sky Survey (BOSS) is targeting 1.5 million Luminous Red Galaxies (LRGs) (Dawson et al. (2013)), while the Euclid mission will obtain redshifts of approximately 50 million galaxies (Cimatti et al. (2009); Wang et al. (2010)). A summary of different galaxy surveys is shown in Fig.1.2.

The most complete galaxy surveys to date are 2dFGRS and SDSS. The 2dF redshift survey used the two-degree field spectroscopic facility on the Anglo-Australian Telescope to measure redshifts reliably from 1997 to 2002. The galaxies covered approximately 1500 square degrees of three regions: NGP (North Galactic Pole) strip, SGP (South Galactic Pole) strip, and random fields around SGP strip. All the data was made publicly available through 2dF website after its final data release on July 2003. The most important cosmological result from the 2dFGRS was the measurement of the galaxy power spectrum, initially by Percival et al. (2001),

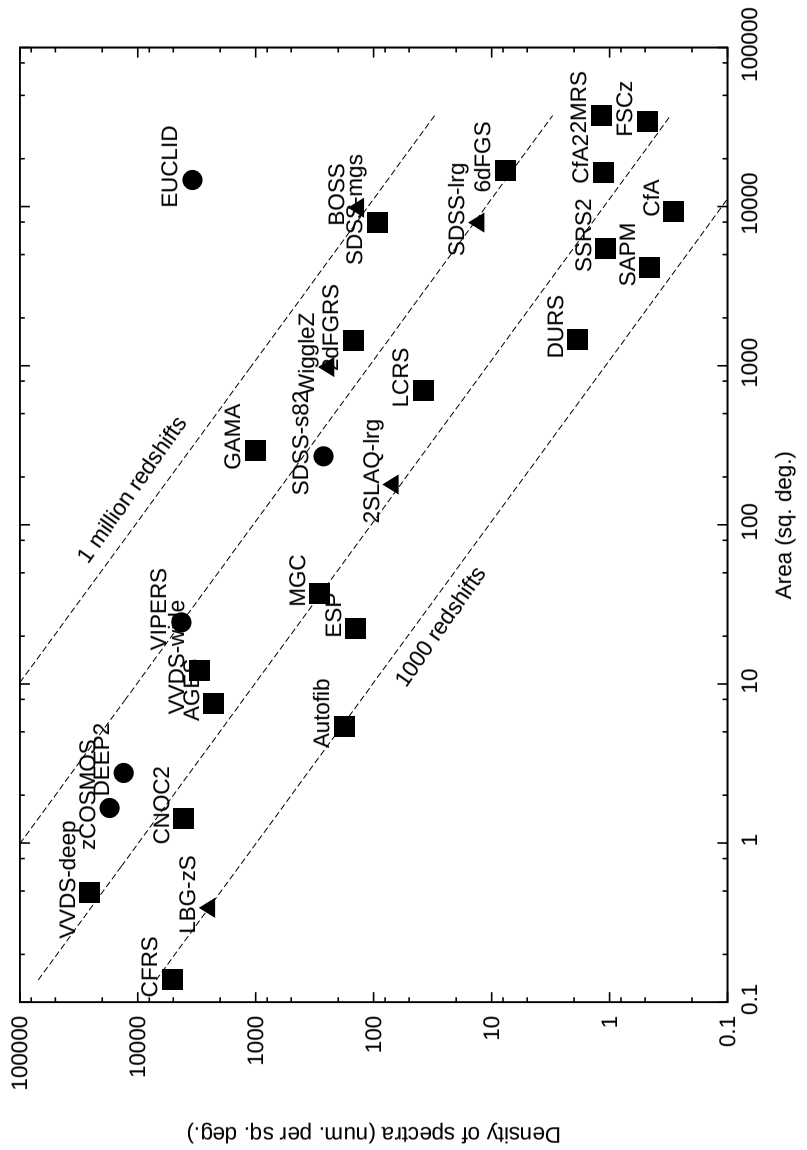

Figure 1.3: A summary of different galaxy redshift surveys. Diagonal lines represent the total number of redshifts measured in the survey. Magnitude limited surveys are plotted with squares. Surveys that use photographic redshifts are shown with circles and highly selected surveys such as BOSS where only LRGs are targeted are plotted with triangles. Adapted from Ivan K. Baldry (`http://www.astro.ljmu.ac.uk/~ikb/research/galaxy-redshift-surveys.html`)

with a final refined estimate by Cole et al. (2005). There is evidence that the first and second BAO peaks of the power spectrum have been detected at wavenumbers 0.06h Mpc^{-1} and 0.12h Mpc^{-1}, respectively. This corresponds to 100h^{-1} in physical space. Assuming standard ΛCDM cosmological model, the overall density parameter was derived as $\Omega_m h = 0.168 \pm 0.016$ and the baryon fraction was estimated to be $\Omega_b/\Omega_m = 0.185 \pm 0.046$. The redshift and sky coverage of 2dFGRS is shown in Fig.1.2.1. I have used 2dFGRS data to obtain the one dimensional galaxy power spectrum in chapter 2.

In contrast, SDSS covered more galaxies and has a larger redshift coverage. SDSS is a major multi-filter imaging and spectroscopic redshift survey using a dedicated 2.5m wide-angle optical telescope at Apache Point Observatory in New Mexico. The project was named after Alfred P. Sloan. It began its goal of mapping 25% of the sky in 2000 and has covered more than 7,500 square degrees of the North Galactic Cap, and three stripes in the South Galactic Cap totaling 740 square degrees containing redshifts of 929,555 galaxies. The redshift coverage of SDSS is shown in Fig.1.2.1. The Luminous Red Galaxy (LRG) data, a subset of SDSS data, was used to obtain one and two dimensional galaxy power spectrum in chapter 2 and 3.

1.2.1 Luminous Red Galaxies (LRGs)

Although LRGs form a quantitatively small sample, they are scattered over a wide range of redshifts. This makes LRGs a perfect candidate for large scale structure studies. The most luminous galaxies in galaxy clusters are a very homogeneous

Cut I for z $\lesssim 0.4$	Cut II for z $\gtrsim 0.4$
$r_{\rm Petro} < 13.1 + c_{\parallel}/0.3$	$r_{\rm Petro} < 19.5$
$\lvert c_{\perp}\rvert < 0.2$	$c_{\perp} > 0.45 - (g-r)/6$
$\mu_{50} < 24.2$ mag/arcsec^{-2}	$\mu_{50} < 24.2$ mag/arcsec^{-2}
$r_{\rm PSF}$ - $r_{\rm model} > 0.3$	$r_{\rm PSF} - r_{\rm model} > 0.5$

Table 1.1: LRGs are defined using colors. The definition consists of two color cuts depending on redshift.

population (Postman & Lauer (1995)) as they have a very narrow range of color and intrinsic luminosity. Because these objects are intrinsically very luminous, they can be observed to great distance. Therefore, these galaxies can be treated as a volume limited sample (Eisenstein et al. (2001)). LRGs are defined using two different cuts as shown in Table 1.1 based on colors (Eisenstein et al. (2001)) due to the shifting of Balmer break at $4000\mathring{A}$ from the g band to r band around redshift z$\sim$ 0.4. In the table 1.1,

$$c_{\perp} = (r-i) - (g-r)/4 - 0.18, \tag{1.1}$$

$$c_{\parallel} = 0.7(g-r) + 1.2[(r-i) - 0.18], \tag{1.2}$$

and g, r, i are SDSS green, red, and infrared magnitudes, respectively through the corresponding filter. $r_{\rm Petro}$ is the Petrosian corrected red magnitude (Petrosian (1976)) and μ_{50} is the Petrosian corrected red surface brightness.

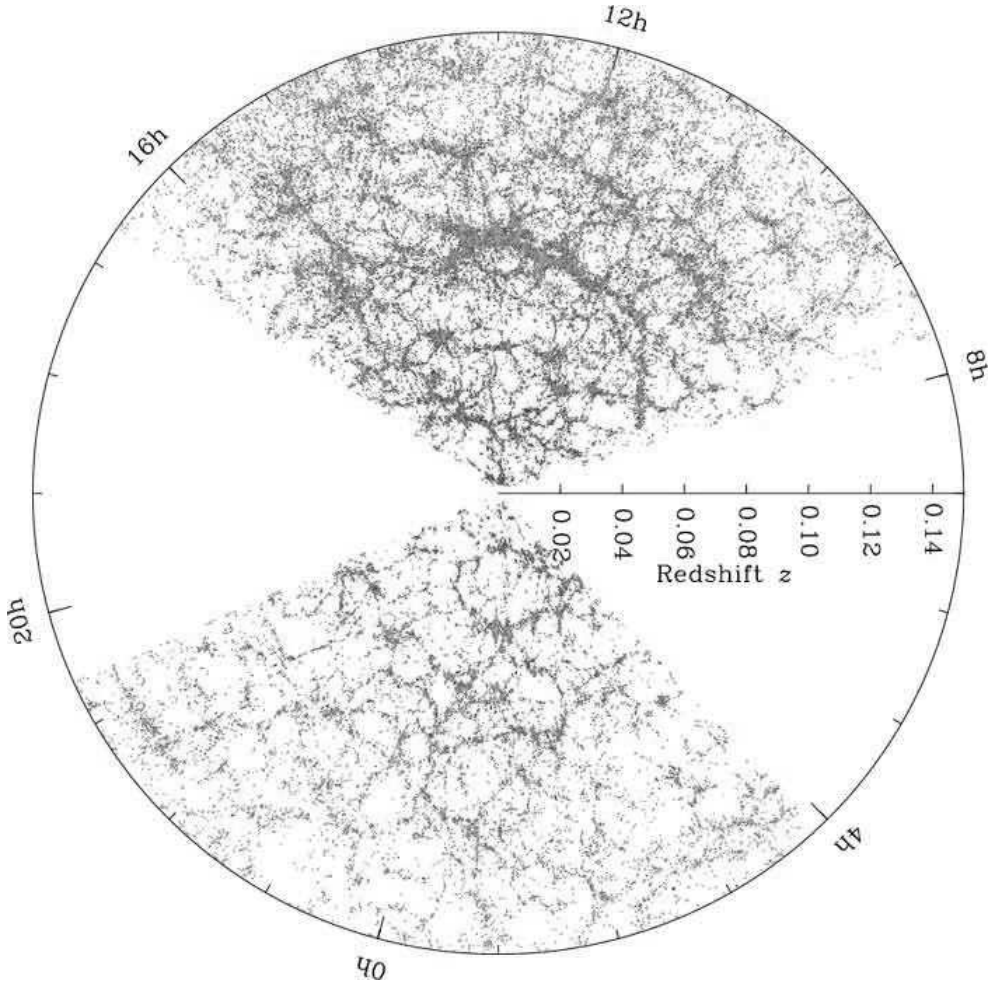

Figure 1.4: This two dimensional projection of SDSS shows that it has a broader sky coverage and deeper redshift coverage compared to 2dFGRS.

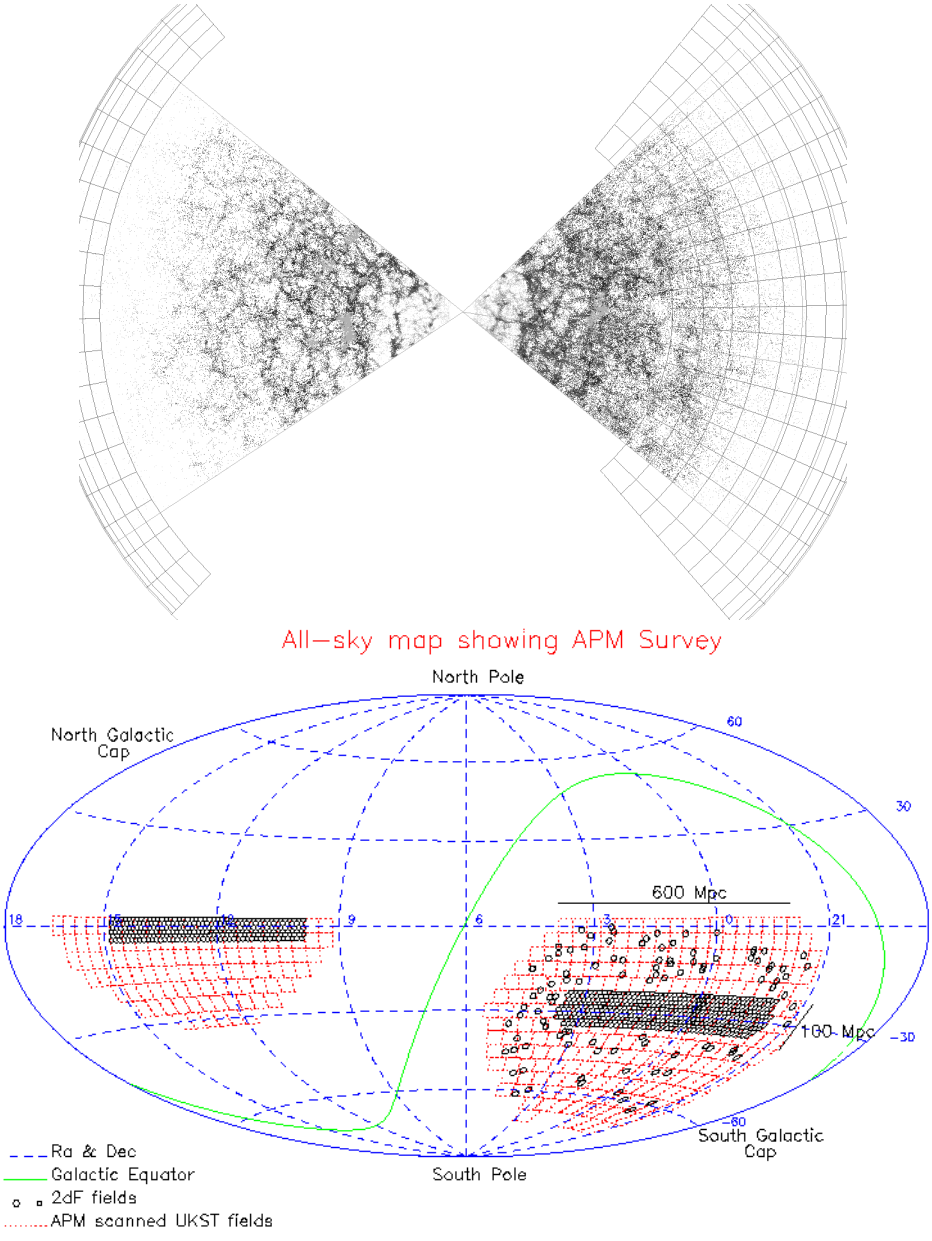

Figure 1.5: Top: Redshift coverage map of the 2dFGRS. Bottom: Sky coverage of 2dFGRS.

References

Abazajian K. N. et al., 2009, Astrophys. J. Suppl. Ser., 182, 543

Cimatti A. et al., 2009, Experimental Astronomy, 23, 39

Cole S. et al., 2005, Mon. Not. R. Astron. Soc., 362, 505

Colless M. et al., 2003, ArXiv Astrophysics e-prints

Dawson K. S. et al., 2013, Astron. J., 145, 10

Eisenstein D. J. et al., 2001, Astrophys. J., 122, 2267

Eisenstein D. J., Hu W., 1998, Astrophys. J., 496, 605

Falco E. E. et al., 1999, Publ. Astron. Soc. Pac., 111, 438

Harrison E. R., 1970, Phys. Rev. D, 1, 2726

Komatsu E. et al., 2003, Astrophys. J., 148, 119

Lilly S. J., Tresse L., Hammer F., Crampton D., Le Fevre O., 1995, Astrophys. J., 455, 108

Linde A., Mukhanov V., 1997, Phys. Rev. D, 56, R535

Longair M. S., 2008, Galaxy Formation

Lyth D. H., Liddle A. R., Ma C.-P., 2010, Physics Today, 63, 070000

Mukhanov V. F., Chibisov G. V., 1981, ZhETF Pisma Redaktsiiu, 33, 549

Parkinson D. et al., 2012, Phys. Rev. D, 86, 103518

Peebles P. J. E., 1973, Astrophys. J., 185, 413

Peebles P. J. E., 1980, The large-scale structure of the universe

Peebles P. J. E., Hauser M. G., 1974, Astrophys. J., Suppl. Ser., 28, 19

Percival W. J. et al., 2001, Mon. Not. R. Astron. Soc., 327, 1297

Petrosian V., 1976, Astrophys. J. Lett., 209, L1

Postman M., Lauer T. R., 1995, Astrophys. J., 440, 28

Saunders W. et al., 2000, Mon. Not. R. Astron. Soc., 317, 55

Shectman S. A., Landy S. D., Oemler A., Tucker D. L., Lin H., Kirshner R. P., Schechter P. L., 1996, Astrophys. J., 470, 172

Sunyaev R. A., Zeldovich Y. B., 1970, Astrophys. Space. Sci., 7, 3

Wang Y. et al., 2010, Mon. Not. R. Astron. Soc., 409, 737

Yu J. T., Peebles P. J. E., 1969, Astrophys. J., 158, 103

Zeldovich Y. B., 1972, Mon. Not. R. Astron. Soc., 160, 1P

Chapter 2

One Dimensional Galaxy Power Spectrum using Direct Fourier Method

2.1 Theory

2.1.1 Matter Overdensity Field

The matter overdensity field is defined as,

$$\delta(\mathbf{r}) = \frac{\rho(\mathbf{r}) - \langle \rho(\mathbf{r}) \rangle}{\langle \rho(\mathbf{r}) \rangle} \tag{2.1}$$

where $\rho(\mathbf{r})$ is the density at position $\mathbf{r}$ and $\langle \rho(\mathbf{r}) \rangle$ is the expected matter density. Assuming that galaxies represent the underlying continuous matter density field, it is possible to use the observed discrete galaxy distribution to estimate the density field. The number density of galaxies is given by,

$$n(\mathbf{r}) = \sum_i \delta_D(\mathbf{r} - \mathbf{r}_i) \tag{2.2}$$

where $\delta_D(\mathbf{r})$ is the Dirac delta function. This field is modeled as an inhomogeneous Poisson process (Cox process) (Martínez & Saar (2002)) with intensity (rate), $\langle n(\mathbf{r}) \rangle$, itself a Poisson density field. Here, angle brackets denote expectation values over the ensemble of the point processes. The intensity (rate) of $\langle n(\mathbf{r}) \rangle$ is denoted by $\bar{n}$ which would be a constant for the usual Cox processes. However, as mentioned above, $\bar{n}$ is a function of the position due to selection effects induced by observations. By combining Eq. 2.1 and Eq. 2.2, one can obtain an estimator

for overdensity of the sample.

$$D(\mathbf{r}) = \sum_i \frac{\delta_D(\mathbf{r} - \mathbf{r}_i)}{\bar{n}(\mathbf{r})} - 1. \tag{2.3}$$

Fourier transform of this would give a simple estimator for Fourier amplitude:

$$F(\mathbf{k}_i) = \sum_j \frac{\psi(\mathbf{r}_j)}{\bar{n}(\mathbf{r}_j)} e^{\mathbf{i}\mathbf{k}_i \cdot \mathbf{r}} - \widetilde{\psi}(\mathbf{k}_i) \tag{2.4}$$

where $\psi(\mathbf{r})$ is a weight function which will optimize our estimator, and $\widetilde{\psi}(\mathbf{k})$ is its Fourier transform. This weight function is normalized as,

$$\int_V \psi^2(\mathbf{r}) d^3\mathbf{r} = 1 \tag{2.5}$$

where V is the sample volume.

2.1.2 Relationship between Correlation Function and Power Spectrum

The autocorrelation function is defined as,

$$\xi(\mathbf{r}) \equiv \left\langle \delta(\mathbf{r} + \mathbf{x}) \delta(\mathbf{x}) \right\rangle \tag{2.6}$$

Here, the average runs over $\mathbf{x}$. This will also be referred as the correlation function elsewhere in this text. One can rewrite the overdensity using the Fourier series as

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3} \sum_k \delta_k e^{-\mathbf{i}\mathbf{k}\cdot\mathbf{x}}. \tag{2.7}$$

Therefore,

$$\delta(\mathbf{r} + \mathbf{x}) = \frac{1}{(2\pi)^3} \sum_{k'} \delta_{k'} e^{-\mathbf{i}\mathbf{k}'\cdot(\mathbf{r}+\mathbf{x})}.$$

Note that the overdensity given by Eq. 2.1 is a real quantity. Therefore, it can be replaced by its complex conjugate without loss of generality. Using this and

substituting Eq. 2.7 to Eq. 2.6

$$\xi(\mathbf{r}) = \frac{1}{(2\pi)^6}\Big\langle \sum_k \sum_{k'} \delta_k^* \delta_{k'} e^{-\mathbf{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} e^{-\mathbf{i}\mathbf{k}'\cdot\mathbf{r}} \Big\rangle.$$

Since the angle brackets denote the average over all $\mathbf{x}$ within the volume the overdensity is defined, the above summation vanishes unless $\mathbf{k} = \mathbf{k}'$. This gives

$$\xi(\mathbf{r}) = \frac{1}{(2\pi)^3} \sum_k |\delta_k|^2 e^{-\mathbf{i}\mathbf{k}\cdot\mathbf{r}}. \tag{2.8}$$

Using the definition of Fourier series,

$$|\delta_k|^2 = \int_V d^3\mathbf{r}\xi(\mathbf{r}) e^{\mathbf{i}\mathbf{k}\cdot\mathbf{r}}.$$

Therefore,

$$\xi(\mathbf{r}) = \frac{1}{(2\pi)^3} \int_{V_k} d^3\mathbf{k} |\delta_k|^2 e^{-\mathbf{i}\mathbf{k}\cdot\mathbf{r}}. \tag{2.9}$$

Eq. 2.9 shows that the autocorrelation function of overdensities is the Fourier transform of its power spectrum since we define the power spectrum as

$$P(k) \equiv |\delta_k|^2, \tag{2.10}$$

and our Fourier transform convention is

$$\delta(\mathbf{r}) = \frac{1}{(2\pi)^3} \int d^3\mathbf{k} \delta_k e^{-\mathbf{i}\mathbf{k}\cdot\mathbf{r}}. \tag{2.11}$$

The power spectrum defined in Eq. 2.10 has units of volume. Alternatively, another form of unitless power spectrum is sometimes constructed as

$$\Delta^2(k) = \frac{1}{2\pi^2} k^3 P(k). \tag{2.12}$$

However, the former definition of power spectrum is used throughout this text unless otherwise specified.

2.1.3 Selection Functions

$\langle \rho(\mathbf{r}) \rangle$ can be expressed in terms of the position dependent mean number density or the selection function ($\bar{n}(\mathbf{r})$) times a small volume δV around the point $\mathbf{r}$ (Eq. 2.16). The selection function accounts for the observational limitations of galaxy surveys and decreases with distance. There are two components in the selection function: radial and angular. Galaxies which are far away from us are too faint and therefore have a small chance of detection compared to the ones closer to us. Thus, the radial galaxy distribution decreases with distance and this is called the radial selection function. The radial selection function of the 2dFGRS is shown in Fig.2.1.3. The angular selection function is also known as the mask and it contains information about the areas avoided or not well-observed in the survey. This is due to bright stars, dust clouds, or some other observational limitation. The radial function is obtained from the survey itself. Usually the mask information is provided with the survey data or one can also make the mask from the data set (eg.Hütsi (2006)). Theoretically, the selection function can be obtained by multiplying the radial and angular selection functions together and counting number of galaxies in each volume element.

2.1.4 Estimation of Power Spectrum

The most widely used power spectrum estimator was first introduced by Feldman, Kaiser & Peacock (1994) (hereafter FKP). Most research groups use the FKP method or a variant of this technique. Our results presented in the next section are based on the original FKP method without using the fast Fourier transform.

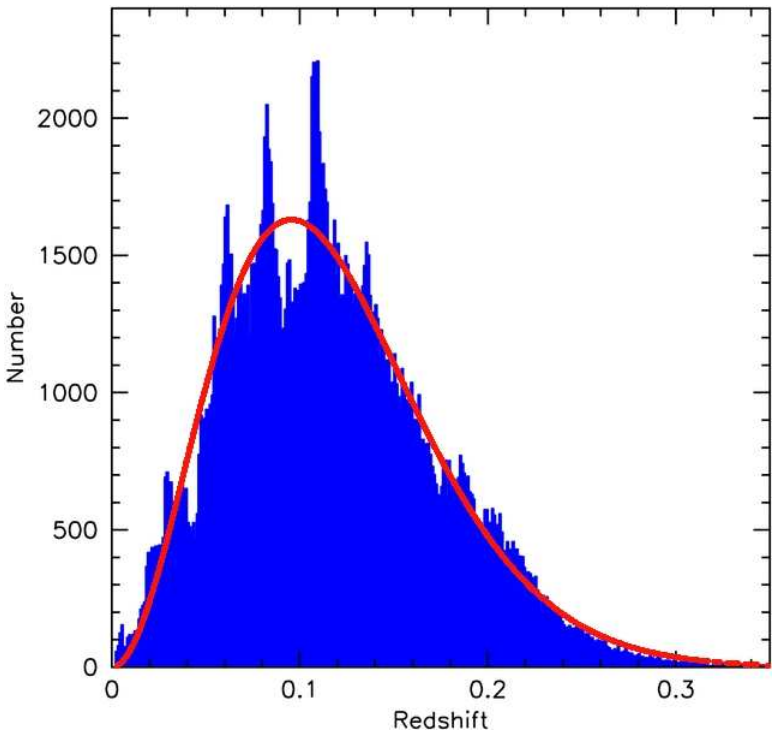

Figure 2.1: The radial selection function of 2dFGRS showing the number of galaxies vs. redshift. The smooth curve is a best fit (Colless et al. (2001)). The galaxy count decreases with redshift as the probability of detection reduces.

They start by defining an estimator,

$$F(\mathbf{r}) \equiv \lambda w(\mathbf{r}) \frac{[n_g(\mathbf{r}) - \alpha n_s(\mathbf{r})]}{\bar{n}(\mathbf{r})} \qquad \text{where} \quad \lambda = \frac{1}{[\int_V d^3\mathbf{r} w^2(\mathbf{r})]^{\frac{1}{2}}}, \tag{2.13}$$

$w(\mathbf{r})$ is the weight function and $n_g(\mathbf{r}), n_s(\mathbf{r})$ are the number density of actual and random synthetic galaxy catalogs, respectively. $\bar{n}$ is the selection function of the chosen galaxy survey. Synthetic catalog contains $1/\alpha$ times galaxies as the real catalog. The synthetic catalog has all the properties as the observed catalog except clustering. i.e., synthetic galaxies are randomly distributed over the survey volume with the same selection function and mask as the real galaxies. This is done by generating points randomly and then using the redshift distribution and angular mask of the actual survey to make its geometry the same as the survey. Details of how to do this will be discussed in the next section. The purpose of using a random galaxy catalog is to quantify the survey volume as done by the second term of Eq. 2.4. The larger the number of random galaxies the better it represents the volume. Therefore, α is a small number ($\alpha \ll 1$).

2.1.5 Proof of FKP Estimator of the Power Spectrum

FKP defined a new estimator for the power spectrum given by Eq. 2.13. We present the motivation for defining $\left\langle \left| F(\mathbf{k}) \right|^2 \right\rangle$ as an estimator for the power spectrum in this section and the following derivation given here is supplementary to Feldman, Kaiser & Peacock (1994).

The Fourier transform of $F(\mathbf{r})$ is

$$F(\mathbf{k}) = \lambda \int d^3\mathbf{r} w(\mathbf{r}) \frac{(n_g(\mathbf{r}) - \alpha n_s(\mathbf{r}))}{\bar{\mathbf{n}}(\mathbf{r})} e^{\mathbf{i}\mathbf{k}\cdot\mathbf{r}}.$$

Therefore,

$$\begin{aligned}
\big|F(\mathbf{k})\big|^2 =& F(\mathbf{k})\cdot F^*(\mathbf{k}) \\
=& \lambda^2 \int d^3\mathbf{r} \int d^3\mathbf{r}' w(\mathbf{r}) w(\mathbf{r}') \frac{(n_g(\mathbf{r}) - \alpha n_s(\mathbf{r}))(n_g(\mathbf{r}') - \alpha n_s(\mathbf{r}'))}{\bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}')} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \\
\Big\langle \big|F(\mathbf{k})\big|^2 \Big\rangle =& \lambda^2 \int d^3\mathbf{r} \int d^3\mathbf{r}' \frac{w(\mathbf{r})w(\mathbf{r}')}{\bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}')} \big\langle (n_g(\mathbf{r}) - \alpha n_s(\mathbf{r}))(n_g(\mathbf{r}') - \alpha n_s(\mathbf{r}')) \big\rangle e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}.
\end{aligned} \tag{2.14}$$

The angle brackets denote expectation value over the galaxies.

In order to proceed further, term inside the angle brackets of Eq. 2.14 needs to be simplified. For this, let $g(\mathbf{r}, \mathbf{r}')$ be any function and consider the expectation value

$$\begin{aligned}
\left\langle \int d^3\mathbf{r} \int d^3\mathbf{r}' g(\mathbf{r}, \mathbf{r}') \mathbf{n}(\mathbf{r}) \mathbf{n}(\mathbf{r}') \right\rangle &= \int d^3\mathbf{r} \int d^3\mathbf{r}' g(\mathbf{r}, \mathbf{r}') \big\langle n(\mathbf{r}) n(\mathbf{r}') \big\rangle \\
&= \sum_i \sum_j g(\mathbf{r}_i, \mathbf{r}_j) \big\langle \mathbf{n}_i \mathbf{n}_j \big\rangle.
\end{aligned} \tag{2.15}$$

The last equality follows from the fact that the integration is done over a discrete rather than a continuous galaxy field. If this galaxy field is divided into infinitesimal microcells of volume δV such that each cell contains maximum of one galaxy (i.e., $n_i = 1$ or 0) then

$$\big\langle \mathbf{n}_i^2 \big\rangle = \big\langle \mathbf{n}_i \big\rangle = \bar{\mathbf{n}}(\mathbf{r}_i)\delta V, \tag{2.16}$$

where $\bar{\mathbf{n}}(\mathbf{r}_i)$ is the selection function at point $\mathbf{r}_i$. Now use Eq. 2.1 and Eq. 2.6 to find a relation between the galaxy density and the correlation function. Here, the

galaxy density is replaced by the number density:

$$\begin{aligned}
\xi(\mathbf{r}) =& \Big\langle \Big(\frac{\mathbf{n}(\mathbf{r}+\mathbf{x})}{\big\langle \mathbf{n}(\mathbf{r}+\mathbf{x}) \big\rangle} - 1 \Big) \Big(\frac{\mathbf{n}(\mathbf{x})}{\big\langle \mathbf{n}(\mathbf{x}) \big\rangle} - 1 \Big) \Big\rangle \\
=& \frac{\big\langle (\mathbf{n}(\mathbf{r}+\mathbf{x}) - \big\langle \mathbf{n}(\mathbf{r}+\mathbf{x}) \big\rangle)(\mathbf{n}(\mathbf{x}) - \big\langle \mathbf{n}(\mathbf{x}) \big\rangle) \big\rangle}{\big\langle \mathbf{n}(\mathbf{r}+\mathbf{x}) \big\rangle \big\langle \mathbf{n}(\mathbf{x}) \big\rangle} \\
\xi(\mathbf{r}) =& \frac{\big\langle \mathbf{n}(\mathbf{r}+\mathbf{x})\mathbf{n}(\mathbf{x}) - \big\langle \mathbf{n}(\mathbf{x}) \big\rangle \mathbf{n}(\mathbf{r}+\mathbf{x}) + \big\langle \mathbf{n}(\mathbf{r}+\mathbf{x}) \big\rangle \mathbf{n}(\mathbf{x}) - \big\langle \mathbf{n}(\mathbf{r}+\mathbf{x}) \big\rangle \big\langle \mathbf{n}(\mathbf{x}) \big\rangle \big\rangle}{\big\langle \mathbf{n}(\mathbf{r}+\mathbf{x}) \big\rangle \big\langle \mathbf{n}(\mathbf{x}) \big\rangle} \\
=& \frac{\big\langle \mathbf{n}(\mathbf{r}+\mathbf{x})\mathbf{n}(\mathbf{x}) \big\rangle - \big\langle \mathbf{n}(\mathbf{r}+\mathbf{x}) \big\rangle \big\langle \mathbf{n}(\mathbf{x}) \big\rangle}{\big\langle \mathbf{n}(\mathbf{r}+\mathbf{x}) \big\rangle \big\langle \mathbf{n}(\mathbf{x}) \big\rangle}
\end{aligned}$$

This simplifies to,

$$\big\langle \mathbf{n}(\mathbf{r}+\mathbf{x})\mathbf{n}(\mathbf{x}) \big\rangle = \big\langle \mathbf{n}(\mathbf{r}+\mathbf{x}) \big\rangle \big\langle \mathbf{n}(\mathbf{x}) \big\rangle (1 + \xi(\mathbf{r}))$$

$$= \bar{\mathbf{n}}(\mathbf{r}+\mathbf{x})\bar{\mathbf{n}}(\mathbf{x})\delta V^2 (1 + \xi(\mathbf{r})) \qquad \text{(using Eq. 2.16).}$$

Therefore,

$$\big\langle \mathbf{n}(\mathbf{r_1})\mathbf{n}(\mathbf{r_2}) \big\rangle = \bar{\mathbf{n}}(\mathbf{r_1})\bar{\mathbf{n}}(\mathbf{r_2})\delta V^2 (1 + \xi(\mathbf{r_1} - \mathbf{r_2})). \tag{2.17}$$

By substituting Eq. 2.16 to Eq. 2.15, and using Eq. 2.17

$$\int d^3\mathbf{r} \int d^3\mathbf{r}' g(\mathbf{r},\mathbf{r}') \big\langle \mathbf{n}(\mathbf{r})\mathbf{n}(\mathbf{r}') \big\rangle = \sum_i \sum_{i \neq j} g(\mathbf{r}_i,\mathbf{r}_j) \big\langle \mathbf{n}_i, \mathbf{n}_j \big\rangle + \sum_i g(\mathbf{r}_i,\mathbf{r}_j) \big\langle \mathbf{n}_i^2 \big\rangle$$

$$= \sum_i \sum_{i \neq j} g(\mathbf{r}_i,\mathbf{r}_j) \bar{\mathbf{n}}(\mathbf{r}_i)\bar{\mathbf{n}}(\mathbf{r}_j)(1 + \xi(\mathbf{r}-\mathbf{r}'))\delta V^2 + \sum_i g(\mathbf{r}_i,\mathbf{r}_i)\bar{\mathbf{n}}(\mathbf{r}_i)\delta V$$

$$= \int d^3\mathbf{r} \int d^3\mathbf{r}' g(\mathbf{r},\mathbf{r}') \bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}')(1 + \xi(\mathbf{r}-\mathbf{r}')) + \int d^3\mathbf{r} g(\mathbf{r},\mathbf{r}')\bar{\mathbf{n}}(\mathbf{r})$$

$$= \int d^3\mathbf{r} \int d^3\mathbf{r}' g(\mathbf{r},\mathbf{r}') \{ \bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}')(1 + \xi(\mathbf{r}-\mathbf{r}')) + \bar{\mathbf{n}}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}') \}. \tag{2.18}$$

By comparing the left-and right-hand sides of Eq. 2.18,

$$\big\langle \mathbf{n}(\mathbf{r})\mathbf{n}(\mathbf{r}') \big\rangle = \bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}')(1 + \xi(\mathbf{r}-\mathbf{r}')) + \bar{\mathbf{n}}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}'). \tag{2.19}$$

Random galaxies are uncorrelated. Therefore, $\xi(\mathbf{r}-\mathbf{r}')=0$ for the random galaxy catalog. Using this and Eq. 2.19, one can obtain following expressions:

$$\left\langle \mathbf{n_g}(\mathbf{r})\mathbf{n_g}(\mathbf{r}')\right\rangle = \bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}')(1+\xi(\mathbf{r}-\mathbf{r}')) + \bar{\mathbf{n}}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}') \tag{2.20}$$

$$\begin{aligned}\left\langle \mathbf{n_g}(\mathbf{r})\mathbf{n_s}(\mathbf{r}')\right\rangle =& \bar{\mathbf{n}}(\mathbf{r})\frac{\bar{\mathbf{n}}(\mathbf{r}')}{\alpha} + \bar{\mathbf{n}}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}') \\ =& \frac{1}{\alpha}\bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}').\end{aligned} \tag{2.21}$$

Now, substitute the above expressions into Eq. 2.14 to obtain

$$\begin{aligned}\left\langle \left|F(\mathbf{k})\right|^2\right\rangle = \lambda^2 \int d^3\mathbf{r}\int d^3\mathbf{r}'\frac{w(\mathbf{r})w(\mathbf{r}')}{\bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}')}e^{\mathbf{ik}\cdot(\mathbf{r}-\mathbf{r}')}\{\bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}')(1+\xi(\mathbf{r}-\mathbf{r}'))+ \\ \bar{\mathbf{n}}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}') + \bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}') + \alpha\bar{\mathbf{n}}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}') - 2\bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}')\},\end{aligned}$$

which simplifies to

$$\begin{aligned}\left\langle \left|F(\mathbf{k})\right|^2\right\rangle =& \lambda^2 \int d^3\mathbf{r}\int d^3\mathbf{r}' w(\mathbf{r})w(\mathbf{r}')e^{\mathbf{ik}\cdot(\mathbf{r}-\mathbf{r}')}\left\{\xi(\mathbf{r}-\mathbf{r}') + (1+\alpha)\frac{1}{\bar{\mathbf{n}}(\mathbf{r}')}\delta(\mathbf{r}-\mathbf{r}')\right\} \\ =& \lambda^2 \int d^3\mathbf{r}\int d^3\mathbf{r}' w(\mathbf{r})w(\mathbf{r}')\xi(\mathbf{r}-\mathbf{r}')e^{\mathbf{ik}\cdot(\mathbf{r}-\mathbf{r}')} + (1+\alpha)\lambda^2\int d^3\mathbf{r}\frac{w^2(\mathbf{r})}{\bar{\mathbf{n}}(\mathbf{r})}.\end{aligned} \tag{2.22}$$

The galaxy power spectrum is defined as the Fourier transform of the correlation function:

$$P(\mathbf{k}) \equiv \frac{1}{(2\pi)^3}\int d^3\mathbf{r}\xi(\mathbf{r})e^{\mathbf{ik}\cdot\mathbf{r}}. \tag{2.23}$$

Therefore,

$$\xi(\mathbf{r}) = \frac{1}{(2\pi)^3}\int d^3\mathbf{k}P(\mathbf{k})e^{-\mathbf{ik}\cdot\mathbf{r}},$$

which gives,

$$\xi(\mathbf{r}-\mathbf{r}') = \frac{1}{(2\pi)^3}\int d^3\mathbf{k}P(\mathbf{k})e^{-\mathbf{ik}\cdot(\mathbf{r}-\mathbf{r}')}. \tag{2.24}$$

Also define

$$W(\mathbf{k}) \equiv \lambda \int d^3\mathbf{r} w(\mathbf{r}) e^{\mathbf{i}\mathbf{k}\cdot\mathbf{r}}, \tag{2.25}$$

which gives,

$$W(\mathbf{k}-\mathbf{k}') = \lambda \int d^3\mathbf{r} w(\mathbf{r}) e^{\mathbf{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}.$$

Therefore,

$$\left|W(\mathbf{k}-\mathbf{k}')\right|^2 = \lambda^2 \int d^3\mathbf{r} \int d^3\mathbf{r}' w(\mathbf{r}) w(\mathbf{r}') e^{-\mathbf{i}(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{r}-\mathbf{r}')}. \tag{2.26}$$

Substituting Eq. 2.24 and Eq. 2.26 into Eq. 2.22,

$$\left\langle \left|F(\mathbf{k})\right|^2 \right\rangle = \int \frac{d^3\mathbf{k}'}{(2\pi)^3} P(\mathbf{k}') \left|W(\mathbf{k}-\mathbf{k}')\right|^2 + (1+\alpha)\lambda^2 \int d^3\mathbf{r} \frac{w^2(\mathbf{r})}{\bar{\mathbf{n}}(\mathbf{r})}. \tag{2.27}$$

The function $W(\mathbf{k})$ defined in Eq. 2.25 is a mask which is also related to the window function. Eq. 2.27 states that our estimator for power, $F(\mathbf{k})$, is in fact the convolution with this window function. This is unavoidable due to the survey geometry effects discussed earlier. However, practically speaking, this window function is quite compact. Therefore, it is fair to say

$$\left\langle \left|F(\mathbf{k})\right|^2 \right\rangle \simeq P(\mathbf{k}) + P_{\rm shot}, \qquad \text{where } P_{\rm shot} = \lambda^2 \int d^3\mathbf{r} \frac{w^2(\mathbf{r})}{\bar{\mathbf{n}}(\mathbf{r})}. \tag{2.28}$$

Therefore, the estimator of power $P(\mathbf{k})$ is

$$\widehat{P}(\mathbf{k}) = \left|F(\mathbf{k})\right|^2 - P_{\rm shot}. \tag{2.29}$$

The final estimator of power spectrum $P(k)$ is obtained by averaging $\widehat{P}(\mathbf{k})$ over a shell in k space:

$$\widehat{P}(k) \equiv \frac{1}{V_{\mathbf{k}}} \int_{V_{\mathbf{k}}} d^3\mathbf{k}' \widehat{P}(\mathbf{k}'). \tag{2.30}$$

2.1.6 Finding Optimum Weights $w(\mathbf{r})$

The mean square fluctuation of $\widehat{P}(k)$ is

$$\sigma_P^2 \equiv \left\langle [\widehat{P}(k) - P(k)]^2 \right\rangle = \left\langle \left| \delta \widehat{P}(\mathbf{k}) \right|^2 \right\rangle = \frac{1}{V_{\mathbf{k}}^2} \int_{V_{\mathbf{k}}} d^3\mathbf{k} \int_{V_{\mathbf{k}}} d^3\mathbf{k}' \left\langle \delta \widehat{P}(\mathbf{k}) \delta \widehat{P}(\mathbf{k}') \right\rangle. \quad (2.31)$$

Now consider

$$\begin{aligned} \left\langle \delta \widehat{P}(\mathbf{k}) \delta \widehat{P}(\mathbf{k}') \right\rangle =& \left\langle (\widehat{P}(\mathbf{k}) - P(k))(\widehat{P}(\mathbf{k}') - P(k')) \right\rangle \\ =& \left\langle \widehat{P}(\mathbf{k}) \widehat{P}(\mathbf{k}') + P(k)P(k') - P(k)\widehat{P}(\mathbf{k}') - \widehat{P}(\mathbf{k})P(k') \right\rangle. \end{aligned} \quad (2.32)$$

By assuming the Fourier coefficients $F(\mathbf{k})$ are Gaussian variables (FKP) one can prove that,

$$\left\langle P(k)P(k') \right\rangle = \left\langle P(k) \right\rangle \left\langle P(k') \right\rangle \text{ and } \left\langle P(k)\widehat{P}(\mathbf{k}') \right\rangle = \left\langle P(k) \right\rangle \left\langle \widehat{P}(\mathbf{k}') \right\rangle.$$

Also, $\left\langle \widehat{P}(\mathbf{k}') \right\rangle = \left\langle P(k) \right\rangle$ since $\widehat{P}$ is an estimator of P. By substitution into Eq. 2.32,

$$\left\langle \delta \widehat{P}(\mathbf{k}) \delta \widehat{P}(\mathbf{k}') \right\rangle = \left\langle \widehat{P}(\mathbf{k}) \widehat{P}(\mathbf{k}') \right\rangle - \left\langle P(k) \right\rangle \left\langle P(k') \right\rangle. \quad (2.33)$$

Realizations of Gaussian processes in k-space can be obtained by Fourier transforming a set of independent Gaussian random variables in real space. i.e.,

$$F(\mathbf{k}) = \sum_i g_i e^{\mathbf{i}\mathbf{k}\cdot\mathbf{r}_i}.$$

Therefore,

$$\begin{aligned} \left\langle F(\mathbf{k}) F^*(\mathbf{k}') \right\rangle =& \left\langle \sum_i g_i e^{\mathbf{i}\mathbf{k}\cdot\mathbf{r}_i} \sum_j g_i e^{\mathbf{i}\mathbf{k}'\cdot\mathbf{r}_j} \right\rangle \\ =& \sum_i \left\langle g_i^2 \right\rangle e^{\mathbf{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}_i}. \end{aligned} \quad (2.34)$$

The last step results from the fact that $\langle g_i g_j \rangle = 0$ for two independent Gaussian variables if $i \neq j$. Now consider

$$\begin{aligned}
\langle \widehat{P}(\mathbf{k})\widehat{P}(\mathbf{k}')\rangle =& \langle F(\mathbf{k})F^*(\mathbf{k})F(\mathbf{k}')F^*(\mathbf{k}')\rangle \\
=& \Big\langle \sum_i g_i e^{\mathbf{i}\mathbf{k}\cdot\mathbf{r}_i} \sum_j g_j e^{-\mathbf{i}\mathbf{k}\cdot\mathbf{r}_j} \sum_l g_l e^{\mathbf{i}\mathbf{k}'\cdot\mathbf{r}_l} \sum_m g_m e^{-\mathbf{i}\mathbf{k}'\cdot\mathbf{r}_m} \Big\rangle \\
=& \sum_i \langle g_i^4 \rangle + \Big\langle \sum_i g_i^2 \sum_{j\neq i} g_j^2 \Big\rangle + \Big\langle \sum_i g_i^2 e^{\mathbf{i}(\mathbf{k}+\mathbf{k}')\cdot\mathbf{r}_i} \sum_{j\neq i} g_j^2 e^{-\mathbf{i}(\mathbf{k}+\mathbf{k}')\cdot\mathbf{r}_j} \Big\rangle \\
&+ \Big\langle \sum_i g_i^2 e^{\mathbf{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}_i} \sum_{j\neq i} g_j^2 e^{-\mathbf{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}_j} \Big\rangle \\
=& \sum_i \langle g_i^4 \rangle + \sum_i \sum_{j\neq i} \langle g_i^2 \rangle \langle g_j^2 \rangle (1 + e^{\mathbf{i}(\mathbf{k}+\mathbf{k}')\cdot(\mathbf{r}_i-\mathbf{r}_j)} + e^{\mathbf{i}(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{r}_i-\mathbf{r}_j)}).
\end{aligned} \tag{2.35}$$

The third step results from the fact that the product does not vanish only when pairs of the indices are equal (i.e., $i = j$ and $l = m$, $i = l$ and $j = m$, $i = m$ and $j = l$) or when all indices are equal. FKP neglected the $e^{\mathbf{i}(\mathbf{k}+\mathbf{k}')\cdot(\mathbf{r}_i-\mathbf{r}_j)}$ term as it oscillates rapidly and therefore, the sum is negligible compared to other terms. Also, for a Gaussian variable, $\langle g_i^4 \rangle = 3\langle g_i^2 \rangle^2$. Therefore, Eq. 2.35 can be simplified further:

$$\begin{aligned}
\langle \widehat{P}(\mathbf{k})\widehat{P}(\mathbf{k}')\rangle =& \sum_i \langle g_i^2 \rangle^2 + \sum_i \sum_j \langle g_i^2 \rangle \langle g_j^2 \rangle (1 + e^{\mathbf{i}(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{r}_i-\mathbf{r}_j)}) \\
=& \langle P(k) \rangle \langle P(k') \rangle + \big| \langle F(\mathbf{k})F^*(\mathbf{k}') \rangle \big|^2.
\end{aligned} \tag{2.36}$$

By substituting Eq. 2.36 to Eq. 2.33

$$\langle \delta\widehat{P}(\mathbf{k})\delta\widehat{P}(\mathbf{k}') \rangle = \big| \langle F(\mathbf{k})F^*(\mathbf{k}') \rangle \big|^2. \tag{2.37}$$

Now, a similar approach used to derive Eq. 2.27 is used to find an expression

for $\big\langle F(\mathbf{k})F^*(\mathbf{k}')\big\rangle$:

$$\begin{aligned}\big\langle F(\mathbf{k})F^*(\mathbf{k}')\big\rangle =&\lambda^2\int d^3\mathbf{r}\int d^3\mathbf{r}'\frac{w(\mathbf{r})w(\mathbf{r}')}{\bar{\mathbf{n}}(\mathbf{r})\bar{\mathbf{n}}(\mathbf{r}')}\big\langle(n_g(\mathbf{r})-\alpha n_s(\mathbf{r}))(n_g(\mathbf{r}')\\ &-\alpha n_s(\mathbf{r}'))\big\rangle e^{\mathbf{i}(\mathbf{k}\cdot\mathbf{r}-\mathbf{k}'\cdot\mathbf{r}')}\\ =&\lambda^2\int d^3\mathbf{r}\int d^3\mathbf{r}'w(\mathbf{r})w(\mathbf{r}')\Big(\frac{1}{(2\pi)^3}\int d^3\mathbf{k}''P(\mathbf{k}'')e^{-\mathbf{i}\mathbf{k}''\cdot(\mathbf{r}-\mathbf{r}')}\Big)\\ &\times e^{\mathbf{i}(\mathbf{k}\cdot\mathbf{r}-\mathbf{k}'\cdot\mathbf{r}')}+(1+\alpha)\lambda^2\int d^3\mathbf{r}\frac{w^2(\mathbf{r})}{\bar{\mathbf{n}}(\mathbf{r})}e^{\mathbf{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}\\ \big\langle F(\mathbf{k})F^*(\mathbf{k}')\big\rangle =&\int\frac{d^3\mathbf{k}''}{(2\pi)^3}P(\mathbf{k}'')W(\mathbf{k}-\mathbf{k}'')W^*(\mathbf{k}-\mathbf{k}'')+S(\mathbf{k}'-\mathbf{k}),\end{aligned}\tag{2.38}$$

where

$$S(\mathbf{k})\equiv(1+\alpha)\lambda^2\int d^3\mathbf{r}\frac{w^2(\mathbf{r})}{\bar{\mathbf{n}}(\mathbf{r})}e^{\mathbf{i}\mathbf{k}\cdot\mathbf{r}}.$$

As before, restricting k, k' to be on the same thin spherical shell and treating the mask W to be compact, one can obtain

$$\big\langle F(\mathbf{k})F^*(\delta\mathbf{k})\big\rangle\simeq P(\mathbf{k})Q(\mathbf{k})+S(\delta\mathbf{k}),\tag{2.39}$$

where

$$Q(\mathbf{k})=\lambda^2\int d^3\mathbf{r}w^2(\mathbf{r})e^{\mathbf{i}\delta\mathbf{k}\cdot\mathbf{r}}.$$

By substituting Eq. 2.39 into Eq. 2.37,

$$\big\langle\delta\widehat{P}(\mathbf{k})\delta\widehat{P}(\mathbf{k}')\big\rangle=\big|P(\mathbf{k})Q(\mathbf{k})+S(\delta\mathbf{k})\big|^2.\tag{2.40}$$

Now, the variance can be obtained with Eq. 2.40 and Eq. 2.31:

$$\begin{aligned}
\sigma_P^2 =& \frac{1}{V_{\mathbf{k}}^2}\int_{V_{\mathbf{k}}} d^3\mathbf{k}\int_{V_{\mathbf{k}}} d^3\mathbf{k}'\Big|P(\mathbf{k})Q(\mathbf{k}'-\mathbf{k})+S(\mathbf{k}'-\mathbf{k})\Big|^2 \\
\simeq& \frac{1}{V_{\mathbf{k}}}\int d^3\mathbf{k}'\Big|P(\mathbf{k})Q(\mathbf{k}')+S(\mathbf{k}')\Big|^2 \text{ since } \mathbf{k} \text{ is small compared to } \mathbf{k}' \qquad (2.41)\\
=& \frac{1}{V_{\mathbf{k}}}\int d^3\mathbf{k}'\Big|P(\mathbf{k})\Big(\lambda^2\int d^3\mathbf{r}w^2(\mathbf{r})e^{i\mathbf{k}'\cdot\mathbf{r}}\Big)+(1+\alpha)\lambda^2\int d^3\mathbf{r}\frac{w^2(\mathbf{r})}{\bar{\mathbf{n}}(\mathbf{r})}e^{i\mathbf{k}'\cdot\mathbf{r}}\Big|^2 \\
=& \frac{\lambda^2}{V_{\mathbf{k}}}\int d^3\mathbf{k}'\Big|d^3\mathbf{r}\Big(w^2(\mathbf{r})P(\mathbf{k})+(1+\alpha)\frac{w^2(\mathbf{r})}{\bar{\mathbf{n}}(\mathbf{r})}\Big)e^{i\mathbf{k}'\cdot\mathbf{r}}\Big|^2 \\
=& \frac{(2\pi)^3}{V_{\mathbf{k}}}\lambda^2\int d^3\mathbf{r}\Big|w^2(\mathbf{r})P(\mathbf{k})+(1+\alpha)\frac{w^2(\mathbf{r})}{\bar{\mathbf{n}}(\mathbf{r})}\Big|^2.
\end{aligned}$$

The last step of the above derivation is based on the Parseval's theorem. The assumption that $|\mathbf{k}|$ is small compared to $|\mathbf{k}'|$ is justified by choosing a shell width that is larger than the coherence length. Also, the synthetic catalog is much larger than the actual catalog. Therefore, α is very small and we have

$$\begin{aligned}
\sigma_P^2(\mathbf{k}) =& \frac{(2\pi)^3}{V_{\mathbf{k}}}\lambda^2\int d^3\mathbf{r}w^4(\mathbf{r})\Big(P(\mathbf{k})+\frac{1}{\bar{\mathbf{n}}(\mathbf{r})}\Big)^2 \\
\frac{\sigma_P^2(\mathbf{k})}{P^2(\mathbf{k})} =& \frac{(2\pi)^3}{V_{\mathbf{k}}}\lambda^2\int d^3\mathbf{r}w^4(\mathbf{r})\Big(1+\frac{1}{\bar{\mathbf{n}}(\mathbf{r})P(\mathbf{k})}\Big)^2.
\end{aligned}$$

In order to find optimum weights w, use the fact that $\sigma_P^2(\mathbf{k})$ should be stationary for arbitrary variations of w:

$$\begin{aligned}
\frac{d\sigma_P^2}{dw} =& 0 \\
\frac{\int d^3\mathbf{r}w^3(\mathbf{r})\Big(1+\frac{1}{\bar{\mathbf{n}}(\mathbf{r})P(\mathbf{k})}\Big)^2}{\int d^3\mathbf{r}w^4(\mathbf{r})\Big(1+\frac{1}{\bar{\mathbf{n}}(\mathbf{r})P(\mathbf{k})}\Big)^2} =& \frac{\int d^3\mathbf{r}w(\mathbf{r})}{\int d^3\mathbf{r}w^2(\mathbf{r})}.
\end{aligned}$$

This is satisfied by

$$w(\mathbf{r}) = \frac{\bar{\mathbf{n}}(\mathbf{r})}{1+\bar{\mathbf{n}}(\mathbf{r})P(\mathbf{k})}. \qquad (2.42)$$

Note that this derivation is carried out by assuming k is much larger than the coherence length. Therefore, $w(\mathbf{r})$ is optimum only for $k \gg 1/L$ where L is

the depth of the survey. Although this weighting requires prior knowledge of $P(k)$, weights remain unbiased as they are normalized in Eq. 2.13 using the normalization constant λ and hence an approximate value of average power spectrum within the range under consideration is used. Error in P will only increase the variance of the estimates. According to this weighting scheme, equally dense points are weighted equally while more-clustered points receive less weight.

As shown in Eq. 2.29, the power spectrum is estimated by subtracting a shot noise term from $\left\langle \left| F(\mathbf{k}) \right|^2 \right\rangle$. Origin of this contribution is the delta function in the density correlation function at zero separation (Eq. 2.19). Since Fourier transform of correlation function is the power spectrum, and the Fourier transform of the delta function is a constant for all wavenumbers, we need to subtract a constant term when estimating power. This is called shot noise because we model the density field with a Poisson distribution and the noise of such distributions is usually known as shot noise in physics.

Measuring more galaxies in the same volume does not improve the estimated power substantially. However, it does reduce shot noise. Also, a sample with higher density (i.e. more galaxies) allows us to measure the power spectrum on larger scales (smaller k). The scale which we can measure with significant accuracy is $k \approx 1/d$, where d is the average separation between galaxies. Therefore, it is important to conduct new redshift surveys for understanding the large scale structure of the universe. Increasing the depth of the survey L reduces the size of coherence cells and hence increases the resolution of the power spectrum. This is the advantage of SDSS over 2dFGRS. Also, SDSS has considerably more data and

sky coverage.

2.1.7 Window Function

The power estimated by FKP estimator is actually convolved with the window function (Eq. 2.27), not the actual power spectrum. The effect of the convolution is to increase the large scale power. The window function is the Fourier transform of a function that depends only on mean density. Mean density is a small and slowly varying function and therefore the window function decays very fast with the wavenumber. Thus, one can treat the FKP power spectrum as the true power spectrum in small scales. The exact form of the window function of the 2dF survey will be presented in section 3.4. The window function is usually treated as spherically symmetric and therefore spherically averaged for mathematical simplicity. Although this is not entirely true for a highly non-symmetric survey volume, most groups who extracted power spectrum from survey data have justified the use of such a window to correct small deviations caused by convolution. Percival et al. (2001) approximated the normalized window function as,

$$|W(k)|^2 = \frac{A}{1 + ak^2 + bk^4} \tag{2.43}$$

where a, b are fitting constants and A is the normalization constant. Although the estimated power spectrum can be corrected by deconvolving $|W(k)|^2$ with the raw power spectrum (Lucy (1974)), this does not produce accurate results due to noise in the observed spectrum.

2.1.8 Summary of the FKP Method

Eq. 2.13 and Eq. 2.42 can be rewritten as,

$$F(\mathbf{r}) \equiv \lambda w(\mathbf{r})[n_g(\mathbf{r}) - \alpha n_s(\mathbf{r})], \qquad \text{where} \quad \lambda = \frac{1}{[\int d^3\mathbf{r}\bar{n}^2(\mathbf{r})w^2(\mathbf{r})]^{\frac{1}{2}}} \tag{2.44}$$

$$\text{and } w(\mathbf{r}) = \frac{1}{1 + \bar{n}(\mathbf{r})P(k)}. \tag{2.45}$$

This is done purely for convenience in expressing integrations as summations as shown below. The Fourier transform of the fluctuation field is then

$$\begin{aligned} F(\mathbf{k}) = & \int d^3\mathbf{r} w(\mathbf{r})[n_g(\mathbf{r}) - \alpha n_s(\mathbf{r})] \exp(i\mathbf{k}\cdot\mathbf{r}) \\ = & \sum_g w(\mathbf{r}_g) \exp(i\mathbf{k}\cdot\mathbf{r}_g) - \alpha \sum_s w(\mathbf{r}_s) \exp(i\mathbf{k}\cdot\mathbf{r}_s). \end{aligned} \tag{2.46}$$

The conversion of integration into summation above is done by using Eq. 2.2. The normalization of $w(\mathbf{r})$ can be expressed in a similar way by using $\int_V d^3\mathbf{r}\bar{n}(\mathbf{r})... \rightarrow \alpha \sum_s ...$. This follows from the fact that $\bar{n}(\mathbf{r})$ is the mean density and the random catalog has α times as many galaxies as real catalog. It is better to express the summation over the synthetic catalog here because we measure mean density using the random catalog. This gives a better estimate simply because there are more galaxies in it. In a similar way, the normalization constant λ is obtained from Eq. 2.13:

$$\lambda^{-1} = \int d^3\mathbf{r}\bar{n}^2(\mathbf{r})w^2(\mathbf{r}) = \alpha \sum_s \bar{n}(\mathbf{r}_s)w^2(\mathbf{r}). \tag{2.47}$$

Finally the estimated power spectrum is given by

$$P(k) = \frac{1}{N_k} \sum_{k<|\mathbf{k}|<k+\delta k} [|F(\mathbf{k})|^2 - S(0)], \qquad S(0) = \alpha(1+\alpha) \sum_s w^2(\mathbf{r}_s) \tag{2.48}$$

where $S(0)$is the shot noise term. As showed in section 2.1.6, the uncertainty of the power estimate is obtained as

$$\sigma_P^2(k) = \frac{2}{N_k} \sum_{\mathbf{k'}} \sum_{\mathbf{k''}} |PQ(\mathbf{k'} - \mathbf{k''}) + S(\mathbf{k'} - \mathbf{k''})|^2, \tag{2.49}$$

$$\text{where} \quad Q(\mathbf{k}) = \alpha \sum_s \bar{n}(\mathbf{r}_s) w^2(\mathbf{r}_s) e^{\mathbf{ik}\cdot\mathbf{r}_s}, \tag{2.50}$$

$$\text{and} \quad S(\mathbf{k}) = \alpha(1+\alpha) \sum_s w^2(\mathbf{r}_s) e^{\mathbf{ik}\cdot\mathbf{r}_s}. \tag{2.51}$$

2.2 Obtaining Galaxy Power Spectrum using 2dFGRS and SDSS Data

2.2.1 Method

The FKP method described in the previous section was used to obtain the galaxy power spectrum. 2dFGRS final data release (Colless et al. (2003)) was used as the data set. There are 186908 galaxies used from this data set. Some galaxies are omitted because the redshifts are poorly determined (according to 2dF group). The NGP region has 77870 total galaxies while the SGP region has 109038 galaxies with accurately determined redshifts. Galaxies in random fields around these regions were skipped as this volume contains many more galaxies than have been observed and this leads to a poor estimate on mean density. The same method was used to measure galaxy power spectrum from SDSS DR7 LRG data (further details about this data set is provided in the next chapter).

The 2dF group provides a code to generate the selection function and completeness at a given point within the survey volume as part of their data set. This code

was used to generate synthetic galaxies. The completeness contains angular mask information. If a given point has a completeness of 1, it means that redshifts have been measured accurately on all the galaxies observed around this point. A fractional completeness is an indication of incomplete redshift data at a point. When a random galaxy is created for the synthetic catalog, the completeness is obtained through this code. It is then probabilistically determined whether to keep this galaxy using the completeness at its position. Resultant synthetic galaxies need to be assigned a redshift. The first step of achieving this is to obtain the redshift distribution of the real data. Following Cole, Sánchez & Wilkins (2007), redshift distribution was fitted with the function

$$\frac{dn}{dz} \propto z^{\alpha} \exp(-(z/z_0)^{\beta}). \tag{2.52}$$

This distribution function is normalized and used to retain or reject the galaxy using the rejection method. i.e., generate a random number and if that number is smaller than the distribution value at its randomly generated redshift, then the galaxy is retained in the catalog. This process was repeated until we obtain five times as many random galaxies as the real data set.

The raw data downloaded from 2dF website contains multiple entries for the same object which correspond to multiple observations. The best observations were filtered for each galaxy by using the "`quality`" field of the data: `quality=3` or higher means they are sufficiently accurate for distance measurements. The galaxy density falls rapidly at the outer boundary of the survey and therefore an upper limit was imposed on redshift. Following Percival et al. (2001), the $0.003 <$

$z < 0.25$ range was selected. The lower limit was chosen for the same reason. The next task was to convert redshifts into distances. Assuming a flat $\Omega_m = 0.3$ cosmological model, redshifts of real and synthetic galaxies were converted into distances (Carroll, Press & Turner (1992)). All catalogs were recorded in a Cartesian coordinates making the dot products in Eq. 2.46 easy to evaluate.

The next step was to set up a three-dimensional Cartesian wave-vector grid. Theoretical considerations of the power spectrum shows that it is a rapidly decreasing function of wavenumber k. Also, using a large k grid is computationally expensive as one needs to evaluate Eq. 2.46 at each and every grid point. Therefore, the chosen grid size was limited to $|\mathbf{k}| \leq 0.5\text{hMpc}^{-1}$ with $\delta k = 0.002\text{hMpc}^{-1}$. Using symmetry in Fourier space, one can show that it is sufficient to consider only the positive side of the z axis (in fact, this is true for any axis and the choice of z axis is arbitrary). This gives more than 8×10^6 grid points. the Fourier transform was performed by evaluating the equations listed in the previous section at each of these grid points. This requires summation over real and synthetic catalogs at each grid point and hence takes a large amount of computational time. The Fourier transform was then used to estimate the power spectrum by calculating the power in each thin shell.

2.3 Results and Discussion

Fig. 2.3.a shows the resultant power spectrum and the 2dF group result (Cole et al. (2005)). Two results agree with each other within 1σ except for small k.

2dF group has deconvolved the measured power spectrum to remove any effects from the window function and therefore our unconvolved result deviates from their power spectrum in large scale. The error bars shown are obtained by evaluating Eq. 2.49 using the same scheme described above. Although error reduces with increasing k, it is worth noting that the constant shot noise term becomes comparable to the power for small k. The shot noise contribution becomes 50% of the power at $k \approx 0.2$. The wiggles extending towards the small scale are baryon oscillations. In particular, the first two peaks at 0.06h Mpc^{-1} and 0.12h Mpc^{-1} are visible and these values agree with the 2dF group findings. Cole et al. (2005) used a slightly different technique: They used fast Fourier transform to do the transformation instead of direct summation which expedite the calculation immensely. However, this also adds some noise to the power spectrum which can not be removed completely at small scale (large k range). The 2dF survey window function is also obtained from the synthetic catalog and shown in Fig 2.2 along with the SDSS window function. This shows that the window function is very compact and hence would only contribute in the large scale limit.

This comparison proves the potential of direct FKP method for one dimensional power spectrum calculations although it is rather computationally expensive. The advantage is that it does not need any modification to obtain the correct power spectrum for all scales. However, required computational time increases linearly with the increase of sample size. For example, SDSS survey currently has $\sim 10^6$ galaxies. This is about five times as many galaxies as 2dFGRS. The SDSS synthetic catalog has $\sim 5 \times 10^6$ galaxies and hence the time required to obtain the

Fourier transform increases five-fold. This is not impossible given the computing techniques such as distributed computing. The data set described in the next chapter is used to generate one dimensional SDSS power spectrum shown in Fig.2.4 with a comparison of published results.

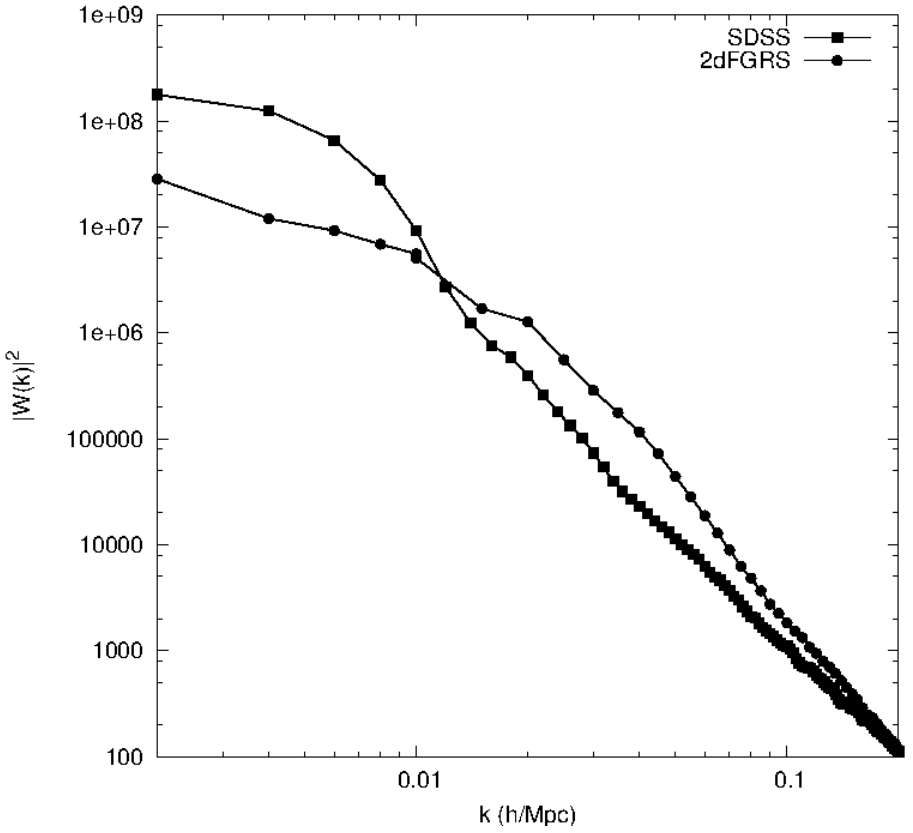

Figure 2.2: Comparison of 2dFGRS and SDSS window functions. The SDSS window is more compact compared to 2dFGRS window as SDSS has more redshift coverage (deep).

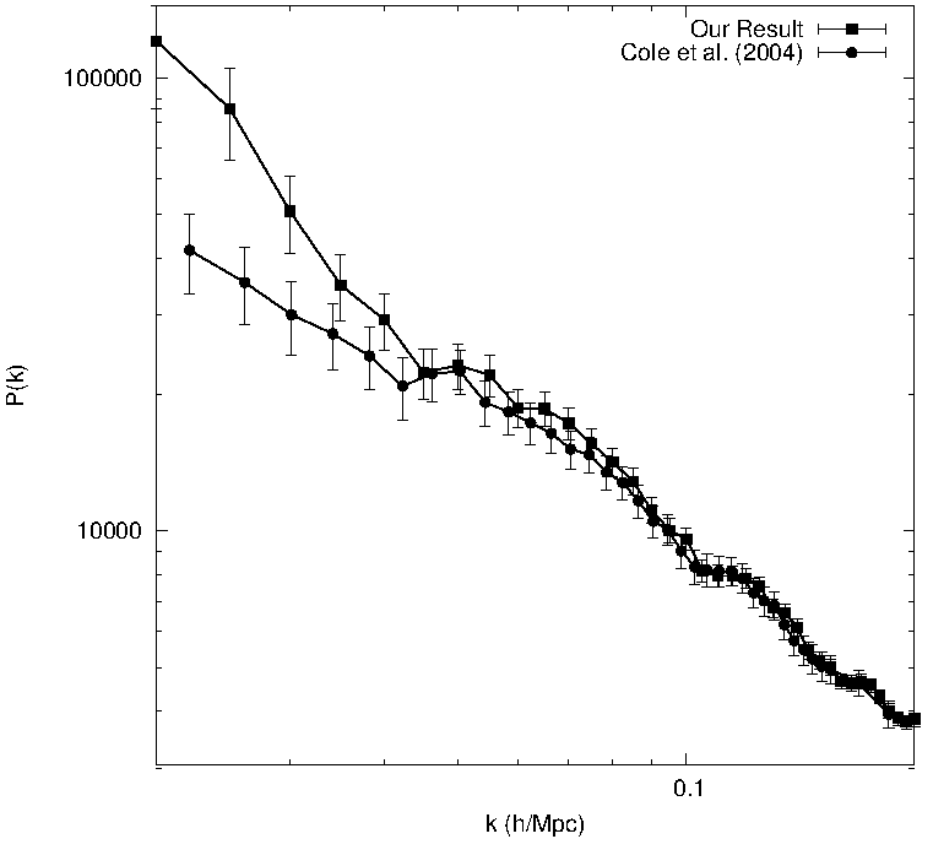

Figure 2.3: 2dFGRS power spectrum obtained by direct Fourier technique compared to the 2dF group published galaxy spectrum. Two spectra deviate from each other at large scale due to the effect of window function as Cole et al. (2005) corrected the window effect by division of a factor determined by model power spectrum and its convolution. This step was omitted in our study as it was only effective for plotting.

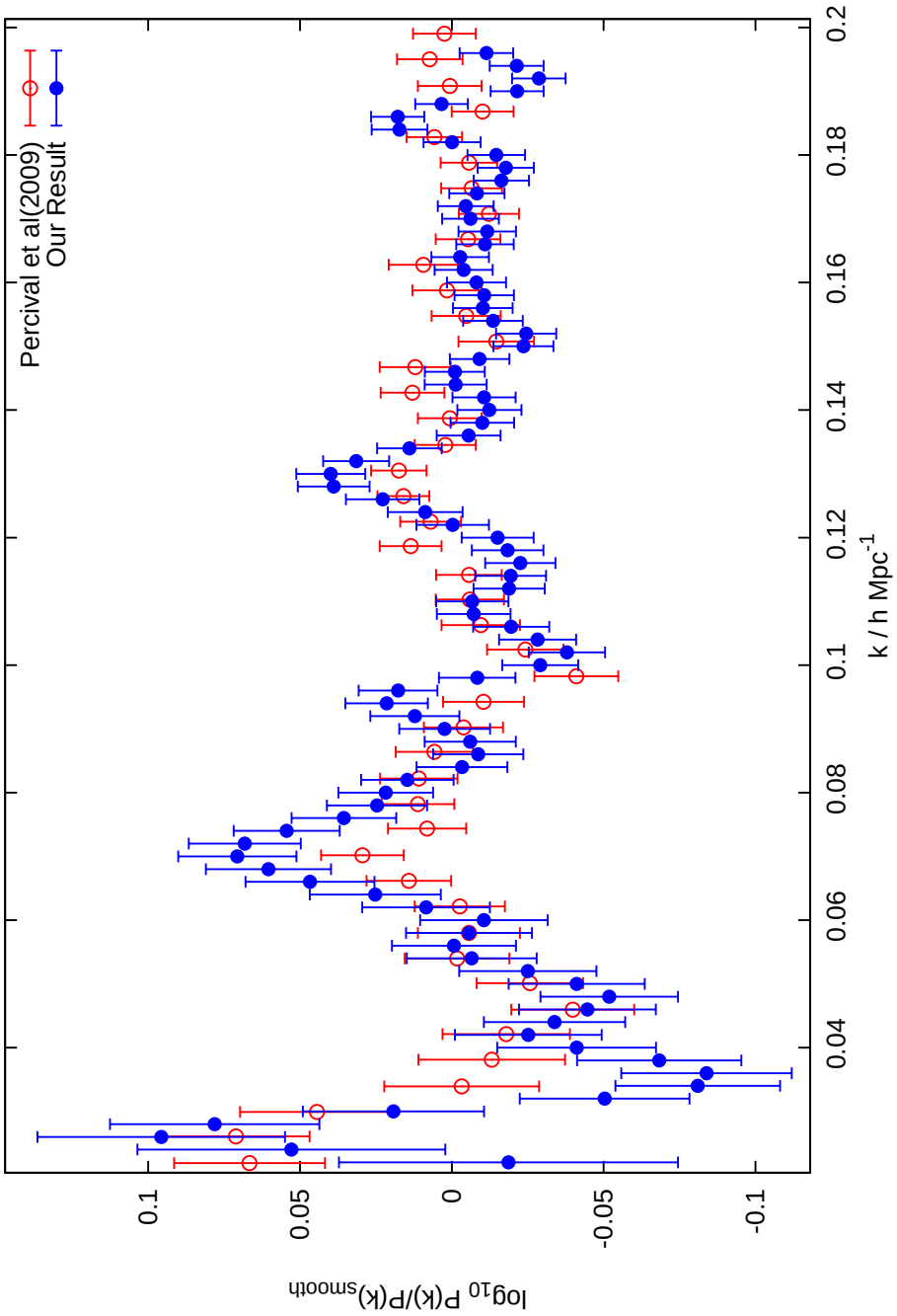

Figure 2.4: Comparison of SDSS power spectrum obtained by direct Fourier technique and published results by SDSS consortium. They have divided the power spectrum by a smooth theoretical spectrum which does not have BAO oscillations and we also used the same method in this comparison.

References

Carroll S. M., Press W. H., Turner E. L., 1992, Ann. Rev. Astron. Astrophys., 30, 499

Cole S. et al., 2005, Mon. Not. R. Astron. Soc., 362, 505

Cole S., Sánchez A. G., Wilkins S., 2007, in Astronomical Society of the Pacific Conference Series, Vol. 379, Cosmic Frontiers, Metcalfe N., Shanks T., eds., p. 57

Colless M. et al., 2001, Mon. Not. R. Astron. Soc., 328, 1039

Colless M. et al., 2003, ArXiv Astrophysics e-prints

Feldman H. A., Kaiser N., Peacock J. A., 1994, Astrophys. J., 426, 23

Hütsi G., 2006, Astron. Astrophys., 449, 891

Lucy L. B., 1974, Astron. J., 79, 745

Martínez V. J., Saar E., 2002, Statistics of the Galaxy Distribution. Chapman & Hall/CRC, Boca Raton

Percival W. J. et al., 2001, Mon. Not. R. Astron. Soc., 327, 1297

Chapter 3

Measurement of $H(z)$ and $D_A(z)$ from the Two Dimensional Power Spectrum of Sloan Digital Sky Survey luminous red galaxies

3.1 Introduction

The galaxy power spectrum is obtained through Fourier transforming the observed galaxy sample. One dimensional power spectrum formed by spherically averaging the Fourier space has been studied well (eg: Cole et al. (2005); Percival et al. (2001, 2010); Reid et al. (2010)) to estimate cosmological parameters including matter density and Hubble's constant. In chapter 2, an analysis of one dimensional galaxy power spectrum from the same data (Chuang, Wang & Hemantha (2012)) was presented. Similar studies have used different data sets such as Eisenstein et al. (2005), Cabré & Gaztañaga (2009), and Kazin et al. (2010). However, it is not possible to measure both Hubble parameter $H(z)$ and angular diameter distance $D_A(z)$ from one dimensional power spectrum or 2PCF alone. The first simultaneous measurement of both of these quantities was obtained by Chuang & Wang (2012) using the SDSS DR7 two-dimensional two point correlation function (2D2PCF). Although the power spectrum and the 2PCF are a Fourier pair, they provide information complementary to each other as redshift surveys cover a limited volume of the Universe. Therefore, an analysis of two-dimensional galaxy power spectrum is presented here.

The two-dimensional galaxy power spectrum has been studied from different

redshift surveys: Las Campanas survey (Landy et al. (1996)), WiggleZ survey (Blake et al. (2010) and Blake et al. (2011a)), HETDEX project (Chiang et al. (2012)), for example. However, the estimation of the full set of cosmological parameters was not carried out. Jing & Börner (2001) measured 2D galaxy power spectrum for $0.25 \leq k \leq 2.5h\mathrm{Mpc}^{-1}$ using LCRS data. However, their limited data set prevented them from measuring the complete set of cosmological parameters. Hu & Haiman (2003) explored the possibility of extracting the Hubble parameter, $H(z)$, and angular diameter distance, $D_A(z)$, from future surveys and noted that curvature of the sky needs to be handled correctly for a broad sky survey such as SDSS. The WiggleZ data was used to obtain 2D power spectrum and estimate bias and growth rate as well as cosmic expansion rate at several redshifts (Blake et al. (2010, 2011a), Blake et al. (2011b)). However, the underlying cosmological model used throughout that analysis was fixed to *Wilkinson Microwave Anisotropy Probe* (WMAP) best fit parameters. Our study aims to measure the main cosmological parameters in addition to $H(z)$ and $D_A(z)$ from the two-dimensional power spectrum.

In section 3.2, we describe the data set used. The method used to obtain the two-dimensional power spectrum is presented in section 3.3. In section 3.4, we validate our method using simulated data and then present the results obtained from real data. We also compare the parameter values with similar work in section 3.4 and summarize our findings in section 3.5.

3.2 Data

The SDSS-II project was finished in October 2008 and this final public data release included spectroscopic observations of 9380 square degrees of sky. These observations were carried out with 2.5 m telescope (Gunn et al. (2006)) at Apache Point Observatory in New Mexico, United States. The luminous red galaxy (LRG) sample (Eisenstein et al. (2001)) used in this work was extracted from `dr72full0` the New York University-Value Added Galaxy Catalog (NYU-VAGC) (Blanton et al. (2005)) by setting the flag `primTarget = 32`. The K-correction was applied to NYU-VAGC data assuming a ΛCDM fiducial model with $\Omega_m = 0.3, h = 1$. We have selected LRGs located within the redshift range $0.16 - 0.47$ and excluded Southern Galactic Cap region, resulting in an LRG sample of 89,599.

Spectra of individual galaxies are obtained by placing fibres on the focal plane of the telescope to guide the light from individual objects to spectrometers. The finite size of these fibres makes it impossible to measure galaxies closer than 55", a problem known as "fibre collisions". Although the overlapping of spectroscopic tiles (Blanton et al. (2003)) alleviates this issue partially through multiple observations, some galaxies in crowded regions were not observed. Zehavi et al. (2002) showed that assigning the redshift of the nearest galaxy with measured redshift is sufficient for large scale structure studies. VAGC used this procedure to correct for fibre collisions.

The angular selection function is generated from the geometry and completeness information provided by VAGC in terms of spherical polygons. We have used

the MANGLE (Swanson et al. (2008)) software package to apply the angular selection function to the data and random galaxies. The radial selection function was constructed by binning the galaxy sample with redshift bins of size $\Delta z = 0.01$.

3.3 Methodology

3.3.1 2D Galaxy Power Spectrum Estimation

In this section, we describe the power spectrum estimation method, which is a two-dimensional extension of the FKP estimator (Feldman, Kaiser & Peacock (1994)). The first step is tiling the SDSS sky coverage into equal area patches as shown in Fig.3.1. This is necessary as the flat sky approximation will not hold for a survey with extended sky coverage such as SDSS. We used the Sanson-Flamsteed projection (Wall & Jenkins (2012)) where a given Right Ascension (α), Declination (δ) pair is mapped such that,

$$\alpha' = \alpha \cos \delta, \qquad \delta' = \delta \tag{3.1}$$

to generate equal area patches.

Choosing too small patches decreases the number of galaxies inside each patch, thus increasing the shotnoise. Choosing patches that are too big will lead to deviation from the flat sky approximation. We have tested dividing the entire survey area into 2, 5, and 10 patches. We find that the 5 patch division yields the lowest bias on estimated parameters, based on application to the SDSS DR7 LRG mocks from the LasDamas (Large suit of Dark matter simulations) collaboration (McBride et al., in preparation) (see section 3.4.1 for further details). Therefore,

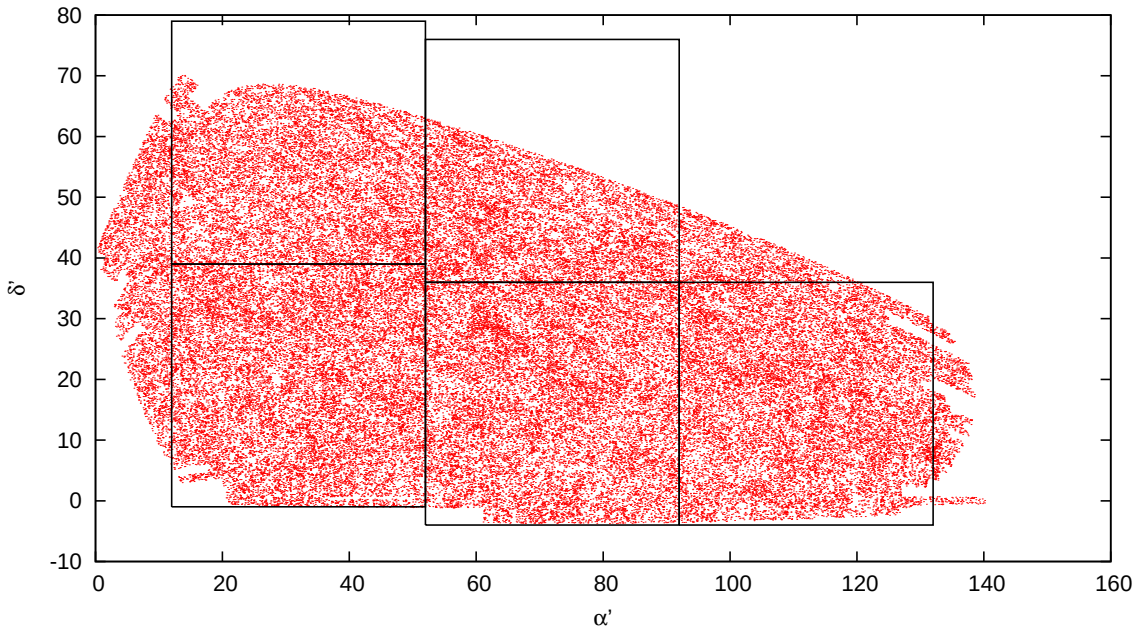

Figure 3.1: This is a plot of SDSS DR7 LRG galaxy sample using a Sanson-Flamsteed projection. The five patches we use are shown. Note that the coordinates are not equatorial (RA, Dec). From left to right, patches 1-3 are the lower panels, and patches 4 and 5 are the upper panels.

we divide the sky into five patches throughout this paper. Galaxies inside each patch were converted to a cartesian coordinate system such that x axis is pointed towards the center of each patch. Distances to galaxies are calculated from redshifts assuming a ΛCDM fiducial model (the same as used by LasDamas in making the LRG mocks) with matter density fraction, $\Omega_m = 0.25$. Each patch is then Fourier transformed as described below. Our choice of axes means that $k_{\parallel} = k_x$ and $k_{\perp} = \sqrt{k_y^2 + k_z^2}$.

We enclose each patch individually in a cube of side $2000h^{-1}$ Mpc, and use the Nearest Grid Point (NGP) scheme (Hockney & Eastwood (1988)) to interpolate

weighted galaxy positions to a regular grid of size 512^3. We use the standard FKP optimal weights (for minimum variance), $w(\mathbf{r}) = \bar{n}(\mathbf{r})/(1 + \bar{n}(\mathbf{r})\bar{P})$, where $\bar{n}(\mathbf{r})$ is the expected number density of galaxies and $\bar{P} = 10000h^{-3}\mathrm{Mpc}^3$ is the average amplitude of the power spectrum. We tested the robustness of this choice by using $\bar{P} = 40000h^{-3}\mathrm{Mpc}^3$ instead, and verified that the exact value of $\bar{P}$ has virtually no effect on the shape of the power spectrum. The FKP estimator described in Eq.2.1.3 of FKP is calculated at each grid point, and then the fast Fourier transform of the grid was obtained. A random galaxy set was generated using MANGLE with the same sky coverage and angular selection function as the real LRG sample. We have used approximately one hundred times more random galaxies than real LRGs to minimize the shot noise. The random galaxies are also divided into the same five patches described above before being used. The Fourier space was then cylindrically summed with bin size $\Delta k = 0.01h\mathrm{Mpc}^{-1}$ in each direction and the shot noise term is subtracted to obtain 2D power spectrum with z axis pointed in $k_\parallel$ direction. We retain only the region $0.02h$ Mpc$^{-1} \leq k \leq 0.16h$ Mpc^{-1} where $k = \sqrt{k_\parallel^2 + k_\perp^2}$ to minimize the effects from aliasing (Jing (2005)).

3.3.2 Theoretical Model

A theoretical model power spectrum is necessary for extracting cosmological parameters from the measured 2D power spectrum. We use the model,

$$P^{\mathrm{s}}_{\mathrm{dw}}(k, \mu, z_0) = P_{\mathrm{dw}}(k, \mu, z_0)\frac{(1 + \beta\mu^2)^2}{1 + (k\mu\sigma_v)^2} \tag{3.2}$$

(Kaiser (1987); Peacock & Dodds (1994); Hamilton (1998)), where β is the redshift distortion parameter, σ_v is the pairwise peculiar velocity dispersion divided by H_0, and μ is the cosine of the angle between the line of sight and wave vector $\mathbf{k}$. $P_{\rm dw}(k, \mu, z_0)$ is the dewiggled linear galaxy power spectrum given by,

$$P_{\rm dw}(k, \mu, z_0) = G^2(z_0) P_0 k^{n_s} T^2_{\rm dw}(k, \mu, z_0), \tag{3.3}$$

where $G(z_0)$ is the linear growth factor and n_s is the power-law index of the primordial matter power spectrum. Anisotropicaly dewiggled transfer function, $T_{\rm dw}(k, \mu, z_0)$, is constructed from the linear transfer function, $T_{\rm lin}(k, z_0)$, and the "no wiggle" transfer function, $T_{\rm nw}(k, z_0)$ from Eq.(29) of Eisenstein & Hu (1998) as in Wang, Chuang & Hirata (2013),

$$T^2_{\rm dw}(k, \mu, z_0) = T^2_{\rm lin}(k, z_0) \exp\left(-g_\mu k^2/k^2_\star\right) + T^2_{\rm nw}(k, z_0)(1 - \exp\left(-g_\mu k^2/k^2_\star\right)), \tag{3.4}$$

where g_μ is given by

$$g_\mu = G^2(z_0)[1 - \mu^2 + \mu^2(1 + f_g(z_0))^2] \tag{3.5}$$

We use $z_0 = 0.35$ as the average redshift in this paper, following previous work on the same data. We use CAMB (Lewis, Challinor & Lasenby (2000)) to calculate linear transfer functions. For the efficient calculation of $T_{\rm lin}(k, z_0)$ for parameters $(\Omega_b h^2, \Omega_c h^2)$, where Ω_b and Ω_c are the baryon and dark matter density fractions respectively, and h is the dimensionless Hubble constant ($H_0 = 100h$km/s/Mpc), we create an evenly spaced grid of transfer functions with spacing 0.001 and 0.005 respectively in each parameter. Cubic spline interpolation is then used to find the linear theory transfer function for a given set of parameter values. This process

is much faster than running CAMB and was rigorously tested and found to be accurate for fitting purposes in this paper. However, linear theory power spectrum does not adequately describe the galaxy power spectrum due to non linear effects. We use a modified version (Sánchez, Baugh & Angulo (2008)) of the semi-analytic model introduced by Cole et al. (2005) to correct the linear matter power spectrum, and modify the galaxy power spectrum as follows:

$$P^{\rm s}_{\rm nl} = \frac{1+Qk^2}{1+Ak+Bk^2} P^{\rm s}_{\rm dw}(k,\mu,z_0), \tag{3.6}$$

where, A, B, Q are constants. Following Sánchez, Baugh & Angulo (2008), we fix $B = Q/10$ and this seem to fit the observed galaxy power spectrum on the range of interest ($0.02h\mathrm{Mpc}^{-1} \leq |\mathbf{k}| \leq 0.16h\mathrm{Mpc}^{-1}$).

Fig.3.2 (top panel) shows a comparison of our theoretical model and the average of 2D power spectra obtained from 160 LasDamas mock catalogs. As discussed in the next section, the model spectrum is convolved with the window function of each of the five patches and then averaged to obtain a smooth plot. This shows the non-linear correction model is able to approximate the observed galaxy power spectrum within our range of interest.

3.3.3 Window Matrix

The observed galaxy power spectrum, $P_{\rm obs}(\mathbf{k})$, is given by convolving the true galaxy power spectrum, $P_{\rm t}(\mathbf{k})$, with the survey window function, $W(\mathbf{k})$, as follows:

$$P_{\rm obs}(\mathbf{k}) = \int d^3\mathbf{k}\prime P_{\rm t}(\mathbf{k}')|W(\mathbf{k}-\mathbf{k}')|^2, \tag{3.7}$$

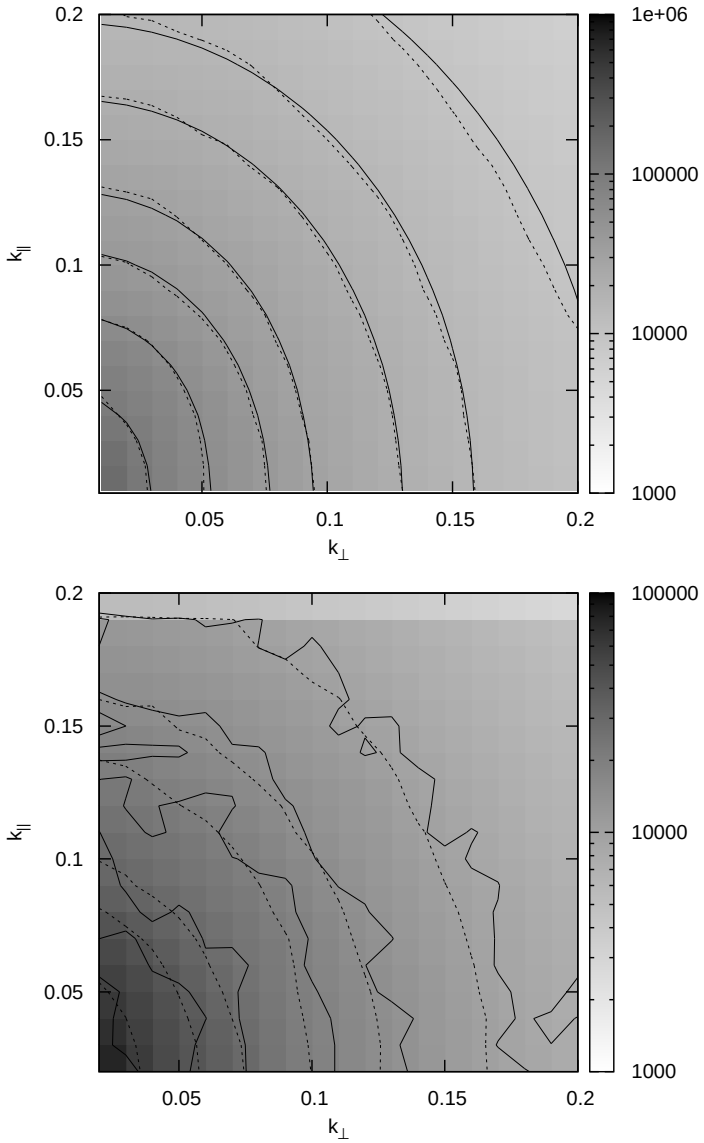

Figure 3.2: Top: Comparison of the average of 160 LasDamas 2D galaxy power spectra (solid lines) and our model 2D power spectrum convolved with the appropriate window function (dotted lines). Model parameters are set to the LasDamas input values. Bottom: Average 2D power spectrum from SDSS DR7 LRGs (solid lines). All five power spectra from different patches were averaged to obtain a smooth plot. The best fit model corresponding to the parameters listed in Table 3.3, convolved with window functions of five patches and averaged together, is plotted with dashed lines.

where the window function is given by

$$W(\mathbf{k}) = \int d^3\mathbf{r}\bar{n}(\mathbf{r})w(\mathbf{r})\exp(i\mathbf{k}\cdot\mathbf{r}). \tag{3.8}$$

As cylindrical coordinate system is a natural choice for 2D power spectrum, Eq.(3.7) can be rewritten as,

$$P_{\rm obs}(\mathbf{k}) = \int dk'_{\parallel}dk'_{\perp}d\phi' k'_{\perp}P_t(\mathbf{k}')|W(\mathbf{k}-\mathbf{k}')|^2. \tag{3.9}$$

The survey window function in configuration space, $w(\mathbf{r})$, is obtained from the random galaxy catalog by using NGP scheme on weighted random catalog alone on the previously mentioned 512^3 size grid. In theory, one can deconvolve the observed power spectrum with the window function to obtain the underlying true galaxy power spectrum. However, deconvolution is susceptible to noise degradation. Thus, we convolve the model with the window window function instead, and compared the convolved model with the observed galaxy power spectrum.

Starting with a cube of size $8000\,{\rm Mpc}h^{-1}$ and successively dividing the size by a factor of 2 until the size is $500\,{\rm Mpc}h^{-1}$ (similar to Cole et al. (2005)), we construct a full three dimensional survey window by only keeping the range 25% - 50% of Nyquist frequency from each box. We use periodic boundary conditions to map points that lie outside boxes. It is necessary to use multiple boxes to obtain a window function with sufficiently wide range ($0.0004\,h{\rm Mpc}^{-1} \leq |\mathbf{k}| \leq 0.7979\,h{\rm Mpc}^{-1}$). We repeat this procedure for each of our five patches, and obtain five window functions. These five windows were used to generate a composite window function for graphical purposes alone and shown in Fig.3.3.3. As the convolution process given by Eq.(3.9) is numerically expensive, we do this integration one time and cast the

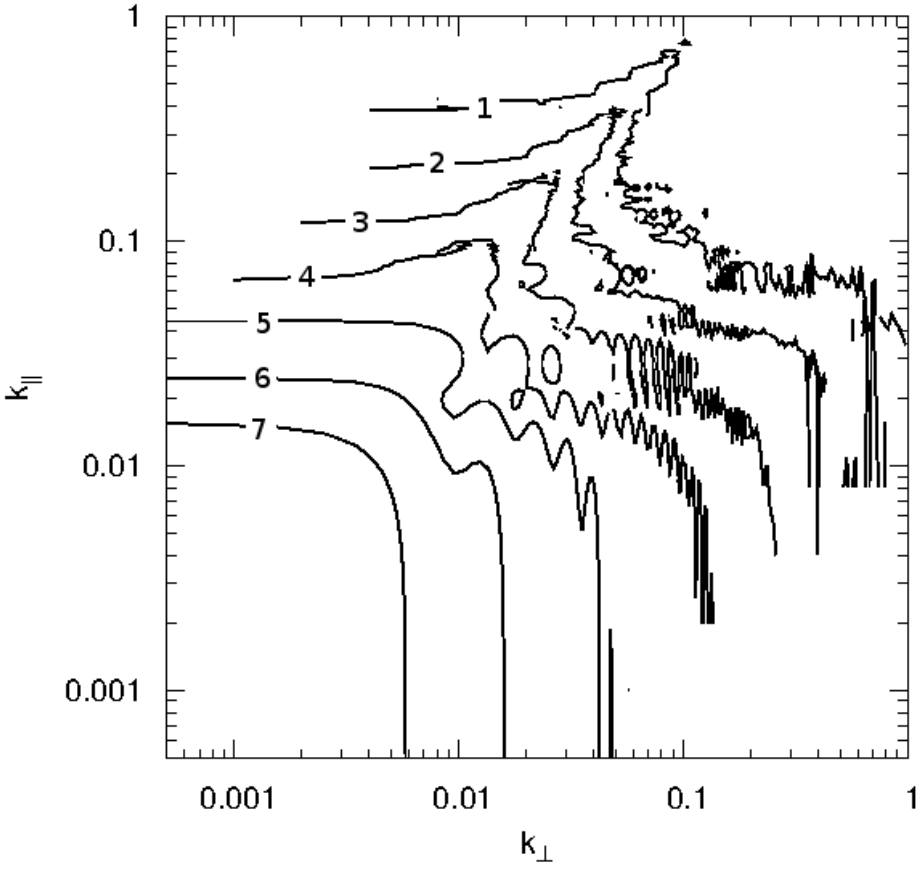

Figure 3.3: Two dimensional window fuction obtained by combining five different window functions with different bin sizes (note that different contours start at different values as a resullt). The contour levels are logarithmic from 10^7 to 10. Also note that this combining was done only for visualization purposes.

result into a window matrix $W_{i,j}$. $P_t(\mathbf{k})$ is replaced by a set of unit basis vectors and the contribution of the window is calculated on each basis vector. For a fixed set of $i \equiv (k_\parallel, k_\perp)$ and $j \equiv (k'_\parallel, k'_\perp)$,

$$W_{i,j} = k'_\perp \int_0^{2\pi} d\phi |W(\mathbf{k} - \mathbf{k}')|^2. \tag{3.10}$$

The window matrix terms are normalized such that $\sum_j W(i,j) = 1$ for each i.

Pre calculated 3D window is spline-interpolated(Press et al. (1992)) to carry out the integration. Now, using Eq.(3.9), a 2D model galaxy power spectrum given by Eq.(3.6) can be convolved with the SDSS window function as follows:

$$P_{\mathrm{th},i} = \sum_j P_{\mathrm{gal},j} W_{i,j}. \tag{3.11}$$

We construct window matrices for each patch separately, and convolve each with the model 2D power spectrum, to obtain the model power spectrum for each patch. The model for each patch can be compared with the observed power spectrum of that patch in a likelihood analysis.

3.3.4 Covariance Matrix

We estimate the covariance matrix as follows

$$C_{ij} = \frac{1}{N-1} \sum_k (\bar{P}_i - P_i^k)(\bar{P}_j - P_j^k), \tag{3.12}$$

where N is the number of mocks catalogs, $\bar{P}_i$ is the mean power spectrum at the ith bin, and P_i^k is the power spectrum at the ith bin in the kth mock catalog. We construct a total of five covariance matrices (one each for the five patches shown in the Fig.3.4). For convenience, we unroll the 2D array of points inside the mask $0.02 \leq |\mathbf{k}| \leq 0.16$ and construct a 1D array of 154 points. This allows us to express the covariance matrix as a 2D matrix.

We use 160 LasDamas mocks to generate covariance matrix for SDSS data. As the galaxy density of the volume limited LasDamas mocks are different from luminosity limited SDSS real galaxy sample, we dilute the mock catalog using the rejection method so that both SDSS and mock data have the same radial selection

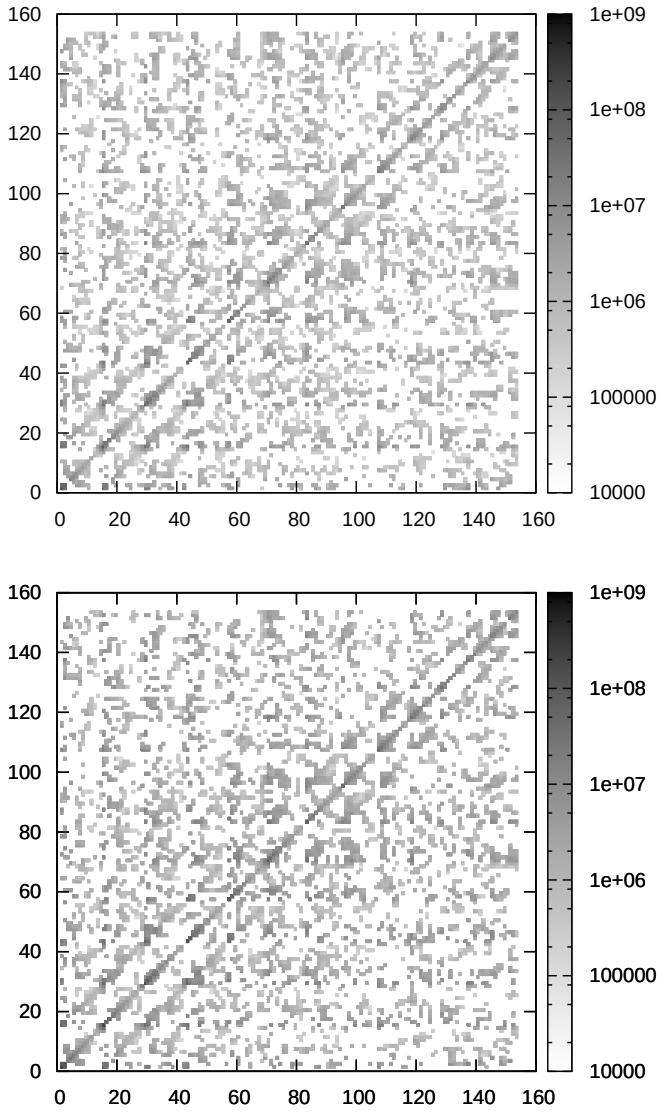

Figure 3.4: Covariance matrices for SDSS data set(top) and LasDamas mock data(bottom). Both covariance matrices are calculated for the patch 1. These matrices are created by unrolling the actual 2D array of points; There are 154 points inside the area of interest.

function. These covariance matrices need smoothing due to the fact that there are only 160 mock catalogs available, and the diluting process described above further reduces the number of galaxies by about 20% in each catalog. We use the same method as described in Chuang & Wang (2012) to make covariance matrices smooth. We use their Eq.(A1) with $p = 0.01$, $\Delta s = \Delta k = 0.01h\mathrm{Mpc}^{-1}$ and repeat the process ten times. The diagonal elements are smoothed using their Eq.(A2) with the same parameter choices.

3.3.5 Likelihood

We derive constraints on estimated parameters in a Markov Chain Monte Carlo (MCMC) likelihood analysis. The likelihood is proportional to $\exp(-\chi^2)$, with

$$\chi^2 = \sum_{i,j} (P_{\mathrm{obs},i} - P_{\mathrm{th},i}) C_{ij}^{-1} (P_{\mathrm{obs},j} - P_{\mathrm{th},j}). \tag{3.13}$$

To use this equation in its original form, one needs to recalculate the covariance matrix and the observed 2D power spectrum for each set of cosmological parameters under consideration (e.g.,Cole et al. (2005)). We use the scaling method from Chuang & Wang (2012), which has the advantage that the observed 2D power spectrum and its covariance matrix only need to be calculated once. The scaling operator T is defined as,

$$P_{\mathrm{obs}}(k_\parallel, k_\perp) = \mathrm{T}(P_{\mathrm{obs}}^{\mathrm{fid}}(k_\parallel, k_\perp)), \tag{3.14}$$

where $P_{\mathrm{obs}}^{\mathrm{fid}}(k_\parallel, k_\perp)$ is the observed power spectrum obtained using a fiducial cosmological model for distance estimation. Now, Eq.(3.13) can be written as,

$$\chi^2 = \sum_{i,j} [\mathrm{T}^{-1}(P_{\mathrm{th},i}) - P_{\mathrm{obs},i}^{\mathrm{fid}}] C_{\mathrm{fid},ij}^{-1} [\mathrm{T}^{-1}(P_{\mathrm{th},j}) - P_{\mathrm{obs},j}^{\mathrm{fid}}]. \tag{3.15}$$

The scaling operator T can be constructed by considering the size of an object of observed size $\Delta z, \Delta\theta$ in the line of sight and transverse directions respectively. Using this argument, Seo & Eisenstein (2003) found,

$$k_{\perp}^{\rm fid} = k_{\perp}\frac{D_A(z)}{D_A^{\rm fid}(z)};\ k_{\parallel}^{\rm fid} = k_{\parallel}\frac{H^{\rm fid}(z)}{H(z)}. \tag{3.16}$$

We define our scaling operator using the above relations, and apply it to the theoretical power spectrum as follows,

$$\mathrm{T}^{-1}(P_{\rm th}(k_{\parallel}, k_{\perp})) = P_{\rm th}\left(\frac{D_A^{\rm fid}(z)}{D_A(z)}k_{\perp}, \frac{H(z)}{H^{\rm fid}(z)}k_{\parallel}\right), \tag{3.17}$$

which we use to calculate $\exp(-\chi^2)$ (see Eq.[3.15]).

We use COSMOMC (Lewis & Bridle (2002)), a publicly available package for MCMC likelihood analysis. Cosmological parameters $\Omega_b h^2$ and n_s are fixed at WMAP 7 values as these parameters are not well constrained by power spectrum alone, and $k_{\star} = 0.11h\mathrm{Mpc}^{-1}$ is used as results are found to be insensitive to small changes of $k_{\star}$. We use the data to extract constraints on $\{\Omega_m h^2$, $H(0.35)/H^{\rm fid}(0.35)$, $D_A^{\rm fid}(0.35)/D_A(0.35)\}$, and marginalize over parameters $\{\beta$, σ_v, Q, A, $N\}$ where N is the normalization of the power spectrum. We use flat priors $\beta = [0.0, 0.9]$, $\sigma_v = [0.0, 700.0]$km/s, $Q = [5.0, 30.0]h^2\mathrm{Mpc}^{-2}$, and $A = [0.5, 10.0]h$Mpc throughout this work.

3.4 Results

We will first present the results from appying our method to mocks (which establish the validity of our method), then the results from the analysis of SDSS DR7 LRGs.

3.4.1 Validating the Method Using Mock Data

We use 80 LasDamas mock catalogs (1a through 40a and 1b through 40b) to validate the method discussed in section 3.3. Each mock catalog is divided into five patches, and each patch is individually analyzed to obtain constraints on the parameters $\{\Omega_m h^2,\ H(0.35)/H^{\rm fid}(0.35),\ D_A^{\rm fid}(0.35)/D_A(0.35)\}$. The estimated parameters from each mock is the weighted average of the estimates from the patches, with the weight proportional to the galaxy count in each patch. The parameters $\Omega_b h^2$ and n_s were fixed at the simulation input values, 0.0196 and 1.0 respectively. Table 3.1 summarizes the results. All the estimated parameters are consistent within 1σ with their input values; this provides validation of our method. We also include derived parameters $H(0.35)r_s(z_d)/c$ and $D_A(0.35)/r_s(z_d)$, where $r_s(z_d)$ is the sound horizon at the drag epoch. As shown in Fig.3.1, not all tiles are entirely full. This reduces the galaxy count in some patches and hence induces more noise compared to other patches. Therefore, we have weighted each tile appropriately before averaging and obtaining standard deviations. Fig.3.5 shows the distributions of the mean values of $H(0.35)r_s(z_d)/c$ and $D_A(0.35)/r_s(z_d)$, as well as $H(0.35)r_s(z_d)/c$ and $D_A(0.35)/r_s(z_d)$, from the 80 mocks. For reference, it also shows the standard deviation of the distributions, as well as the input values of the parameters.

In order to optimize the choice for the number of patches that the survey area is divided into, we have applied our method with different patch sizes, corresponding to 2, 5, and 10 patches respectively. Estimated parameters from the division

Parameter	Mean	σ	Input Value
$\Omega_m h^2$	0.1271	0.0049	0.1225
$D_A^{\rm fid}(0.35)/D_A(0.35)$	1.007	0.033	1.0
$H(0.35)/H^{\rm fid}(0.35)$	1.002	0.035	1.0
$D_A(0.35)/r_s(z_d)$	6.41	0.17	6.48
$H(0.35)r_s(z_d)/c$	0.0425	0.0012	0.0434

Table 3.1: LasDamas mock catalog fitting results. Each mock catalog is divided into five patches, and each patch is analyzed separately. The estimated parameters from each mock is the weighted average of the estimates from the patches. The mean and standard deviation are obtained by averaging over 80 mock catalogs.

into two patches deviate by more than 2σ from the input values; we believe this is due to the breakdown of the flat sky approximation as each patch is about $60^\circ \times 60^\circ$. When the survey region is divided into ten patches, the number of galaxies in each patch is significantly lower and hence the power spectrum is noisy. Therefore, the covariance matrix is very noisy, and the estimated parameters have significantly larger error bars, although mean parameter values are consistent with input parameters, as shown in Table 3.2. We conclude that dividing the survey area into five patches is the optimal choice for this work.

Parameter	Mean	σ	Input Value
$\Omega_m h^2$	0.124	0.010	0.1225
$D_A^{\rm fid}(0.35)/D_A(0.35)$	1.017	0.086	1.0
$H(0.35)/H^{\rm fid}(0.35)$	1.032	0.075	1.0
$D_A(0.35)/r_s(z_d)$	6.39	0.29	6.48
$H(0.35)r_s(z_d)/c$	0.0431	0.0017	0.0434

Table 3.2: Same as Table 3.1, but for dividing each mock into 10 patches.

3.4.2 Constraints on Parameters from SDSS Data

We now present our results from the analysis of SDSS DR7 LRGs. Table 3.3 lists the mean and standard deviation for measured parameters $\{\Omega_m h^2,\ H(0.35),\ D_A(0.35)\}$, and derived parameters $H(0.35)r_s(z_d)/c$ and $D_A(0.35)/r_s(z_d)$ that we have obtained from the 2D power spectrum of the SDSS DR7 LRGs. The mean parameter values are calculated as follows,

$$p = \sum_{i=1}^{5} \frac{p_i}{\sigma_i^2} \Big/ \sum_{i=1}^{5} \frac{1}{\sigma_i^2}, \tag{3.18}$$

where, p, p_i are mean parameter value and the mean parameter value for the $i^{\rm th}$ patch, respectively. The standard deviations are the square roots of the diagonal elements of the covariance matrix, which is obtained by inverting the matrix sum of the inverse covariance matrices from the five patches. Table 3.4 gives the normalized covariance matrix. The covariance matrix can be reconstructed as follows:

$$C_{i,j} = \sigma_i \sigma_j C_{i,j}^{\rm norm}, \tag{3.19}$$

where $C_{i,j}^{\rm norm}$ is the normalized covariance matrix, and the σ_i's are given in Table 3.3. Figs.3.6-3.10 show the one dimensional probability distribution functions and 2D joint confidence contours of the primary parameters in our analysis. In this analysis, we have fixed $\Omega_b h^2$ and n_s to the WMAP 7 year cosmological parameter values (Larson et al. (2011)), 0.02258 and 0.963 respectively, and $k_\star = 0.11 h{\rm Mpc}^{-1}$. Fixing $\Omega_b h^2$ and n_s is justified by the fact that neither parameter is well constrained by power spectrum data alone (eg. Percival et al. (2010)), and both are well determined by WMAP data. Both of these parameters were fixed in similar studies (eg. Reid et al. (2010)).

Chuang & Wang (2012) simultaneously measured $H(0.35) = 82.1^{+4.8}_{-4.9}$km/s/Mpc, $D_A(0.35) = 1048^{+60}_{-58}$Mpc for the first time using two-dimensional two point correlation function. Our results from using the same data set are within 1σ of their measurements. The differences in mean values and errors can be attributed to the different methods used (correlation function versus power spectrum). Our results are also comparable with Xu et al. (2013), where they measured $H(0.35) = 84.4 \pm 7.0\,{\rm km/s/Mpc}$, $D_A(0.35) = 1050 \pm 38\,{\rm Mpc}$ assuming WMAP7 cosmology from correlation function analysis of SDSS DR7 data. Their measurements are within 1σ of our measurements. They used the multipole method to carry out an anisotropic analysis similar to Chuang & Wang (2013). However, it should be noted that their theoretical model is different from Eq.(3.2): They used a different FoG model such that the denominator of Eq.(3.2) is squared. This may explain the difference in the magnitude of errors for each parameter, as the additional damping of radial power they applied is expected to result in increased uncertainty on the

measured $H(z)$.

Parameter	Mean	σ
$\Omega_m h^2$	0.1268	0.0085
$D_A(0.35)$	1037	44
$H(0.35)$	81.3	3.8
$D_A(0.35)/r_s(z_d)$	6.48	0.25
$H(0.35)r_s(z_d)/c$	0.0431	0.0018

Table 3.3: Results from our analysis of SDSS DR7 LRGs. The mean values and standard deviations are calculated from the mean parameter values and covariance matrices obtained by fitting parameters for the 5 patches.

$\Omega_m h^2$	$D_A(0.35)$	$H(0.35)$	$D_A(0.35)/r_s(z_d)$	$H(0.35)r_s(z_d)/c$
1	-0.4535	0.4936	0.1746	-0.0915
-0.4535	1	-0.4009	-0.2772	0.9270
0.4936	-0.4009	1	0.9420	-0.2435
0.1746	-0.2772	0.9420	1	-0.2384
-0.0915	0.9270	-0.2435	-0.2384	1

Table 3.4: Normalized average covariance matrix corresponding to Table 3.3.

3.5 Conclusion and Discussion

We present the first measurement of $H(z)$ and $D_A(z)$ from the two-dimensional galaxy power spectrum from SDSS DR7 LRG data. This method can be applied to any future survey with a broad sky coverage. The basic concept is to divide the sky into patches of roughly equal area and calculate individual power spectra for each patch. We find that the optimum number of patches for SDSS DR7 data is five, so that enough number of galaxies are included in each patch and the flat sky approximation is also valid. We have measured $\{\Omega_m h^2,\ H(0.35),\ D_A(0.35)\}$ and derived parameters $H(0.35)r_s(z_d)/c$ and $D_A(0.35)/r_s(z_d)$ from the SDSS DR7 LRGs, as shown in Table 3.3. Note that we have analyzed the full two-dimensional power spectrum, and not the Baryon Acoustic Oscillation (BAO) alone.

To validate our method, we applied it to LasDamas mock data and constrained cosmological parameters. The results shown in Table 3.1 are consistent with the LasDamas input parameters, thus establishing the validity of our method.

Our measurements of $H(0.35)$ and $D_A(0.35)$ from the SDSS DR7 LRGs, with errors of 4.67% and 4.29% respectively, are comparable with the values reported in similar work. We also find that the derived parameters $H(0.35)r_s(z_d)/c$ and $D_A(0.35)/r_s(z_d)$ are more tightly constrained, with errors of 4.18% and 3.87% respectively. A survey such as BOSS which is currently ongoing with more galaxies and deeper than SDSS would enable the utilization of this method to further tighten the constraints on these parameters, as well as the matter density and index of the primordial power spectrum.

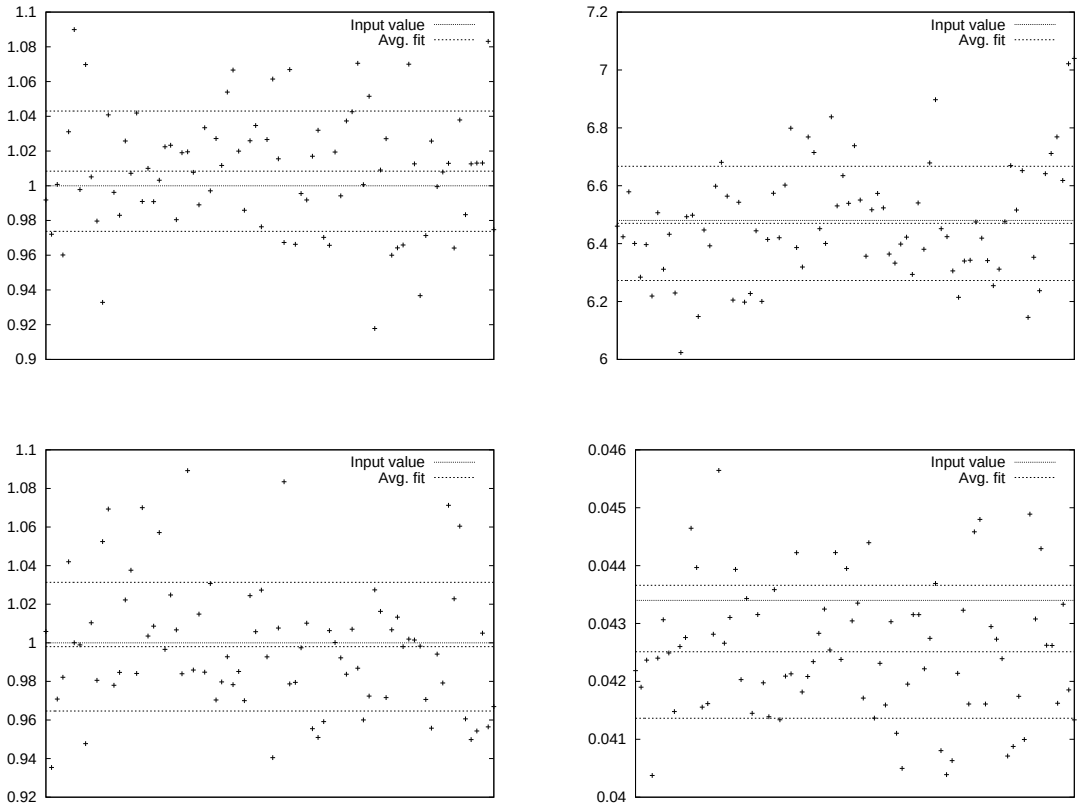

Figure 3.5: LasDamas fitting results for the parameters $D_A^{\rm fid}(0.35)/D_A(0.35)$ (top left), $D_A(0.35)/r_s(z_d)$ (top right), $H(0.35)/H^{\rm fid}(0.35)$ (lower left), $H(0.35)r_s(z_d)/c$ (lower right). Dashed lines represent mean values and 1σ error bars and input parameter values are plotted with dotted lines.

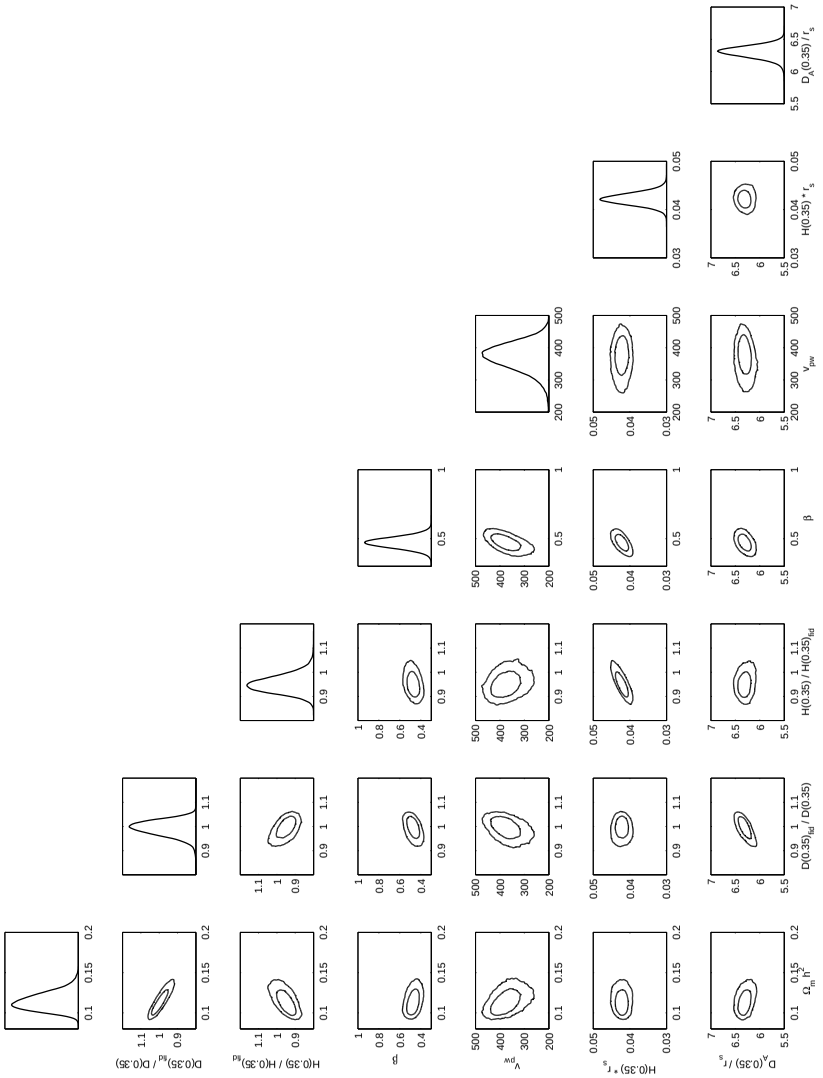

Figure 3.6: The 1D marginalized probability distribution functions and 2D joint confidence contours of the primary parameters in our analysis of SDSS DR7 LRG sample at patch 1.

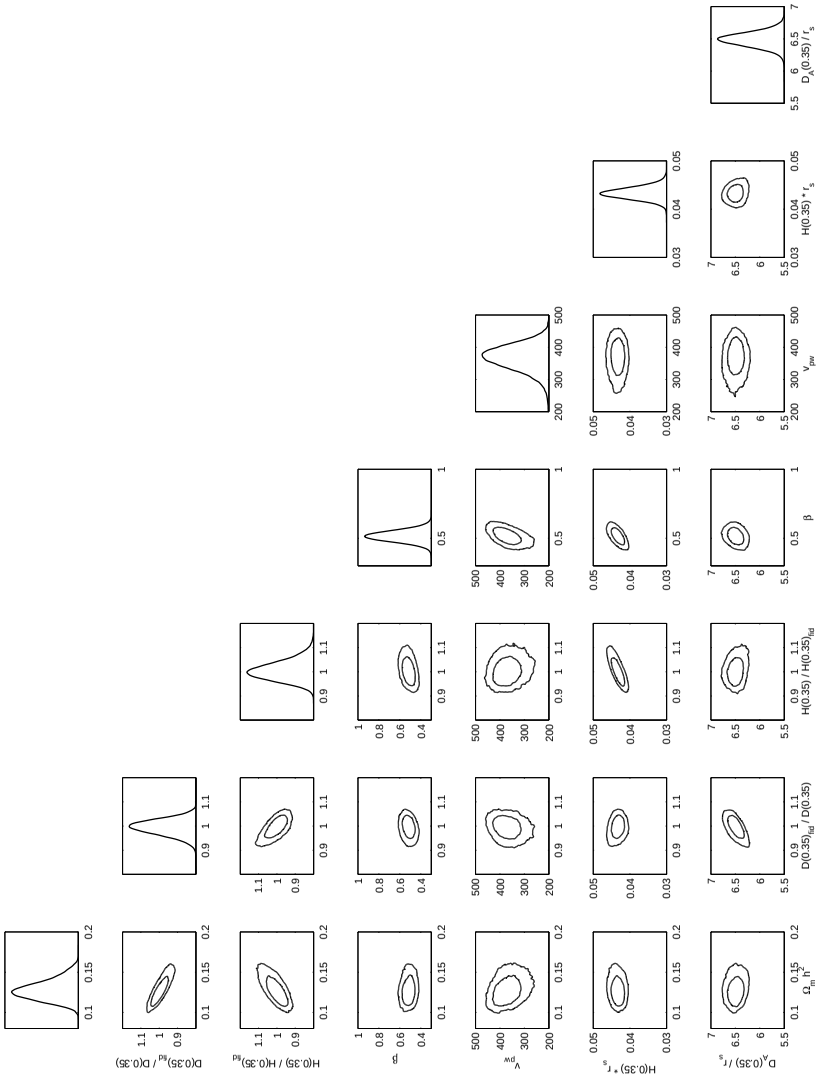

Figure 3.7: Same as Fig.3.6, but for patch 2

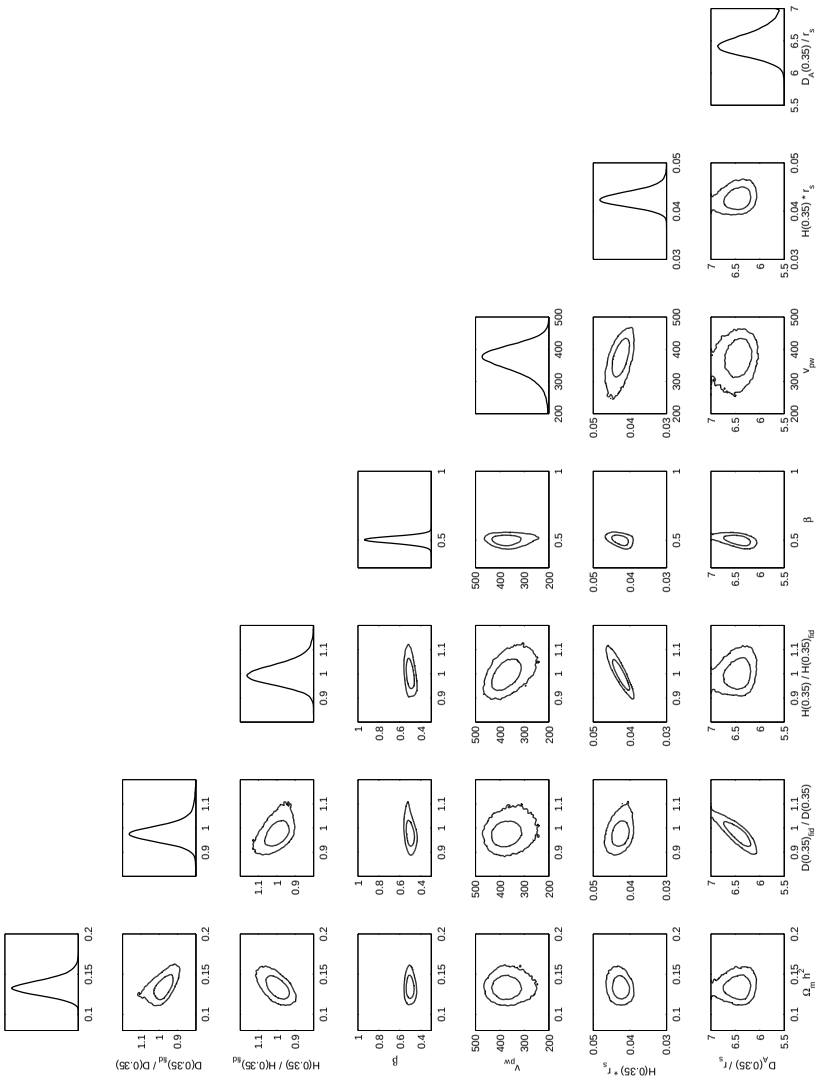

Figure 3.8: Same as Fig.3.6, but for SDSS patch 3

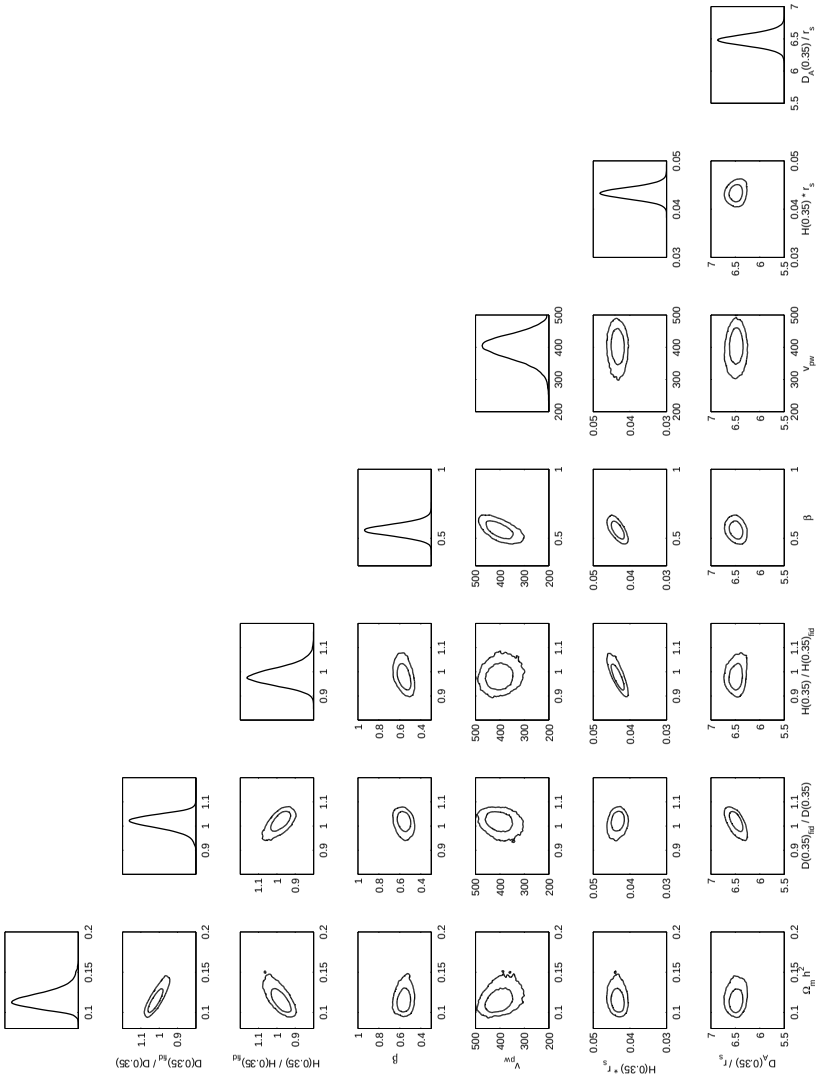

Figure 3.9: Same as Fig.3.6, but for SDSS patch 4

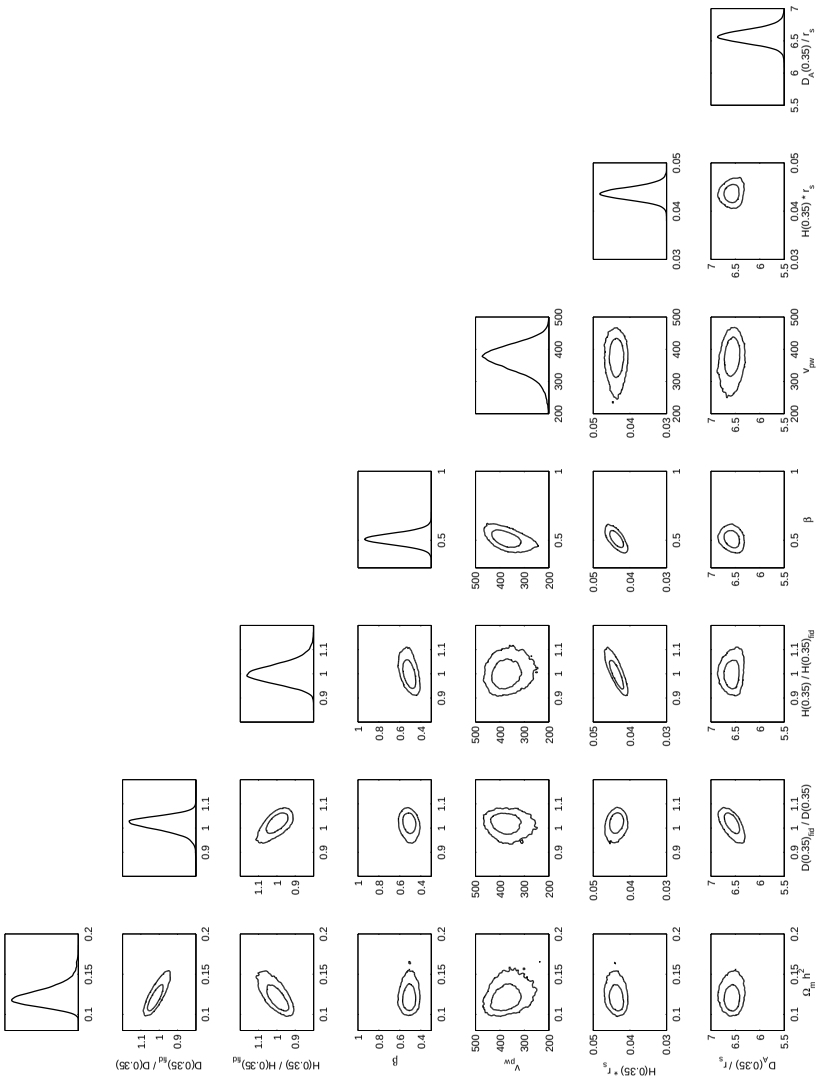

Figure 3.10: Same as Fig.3.6, but for SDSS patch 5

References

Blake C. et al., 2011a, Mon. Not. R. Astron. Soc., 415, 2876

Blake C. et al., 2010, Mon. Not. R. Astron. Soc., 406, 803

Blake C. et al., 2011b, Mon. Not. R. Astron. Soc., 418, 1725

Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003, Astron. J., 125, 2276

Blanton M. R. et al., 2005, Astron. J., 129, 2562

Cabré A., Gaztañaga E., 2009, Mon. Not. R. Astron. Soc., 393, 1183

Chiang C.-T., Komatsu E., Jeong D., Hill G. J., Gebhardt K., HETDEX Collaboration, 2012, in American Astronomical Society Meeting Abstracts, Vol. 219, American Astronomical Society Meeting Abstracts #219, p. 424.14

Chuang C.-H., Wang Y., 2012, Mon. Not. R. Astron. Soc., 426, 226

Chuang C.-H., Wang Y., 2013, Mon. Not. R. Astron. Soc., 431, 2634

Chuang C.-H., Wang Y., Hemantha M. D. P., 2012, Mon. Not. R. Astron. Soc., 423, 1474

Cole S. et al., 2005, Mon. Not. R. Astron. Soc., 362, 505

Eisenstein D. J. et al., 2001, Astron. J., 122, 2267

Eisenstein D. J., Hu W., 1998, Astrophys. J., 496, 605

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

Feldman H. A., Kaiser N., Peacock J. A., 1994, Astrophys. J., 426, 23

Gunn J. E. et al., 2006, Astron. J., 131, 2332

Hamilton D., ed., 1998, Astrophysics and Space Science Library, Vol. 231, The evolving universe. Selected topics on large-scale structure and on the properties of galaxies

Hockney R. W., Eastwood J. W., 1988, Computer simulation using particles

Hu W., Haiman Z., 2003, Phys. Rev. D, 68, 063004

Jing Y. P., 2005, Astrophys. J., 620, 559

Jing Y. P., Börner G., 2001, Mon. Not. R. Astron. Soc., 325, 1389

Kaiser N., 1987, Mon. Not. R. Astron. Soc., 227, 1

Kazin E. A. et al., 2010, Astrophys. J., 710, 1444

Landy S. D., Shectman S. A., Lin H., Kirshner R. P., Oemler A. A., Tucker D., 1996, Astrophys. J. Lett., 456, L1

Larson D. et al., 2011, Astrophys. J. Suppl. Ser., 192, 16

Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511

Lewis A., Challinor A., Lasenby A., 2000, Astrophys. J., 538, 473

Peacock J. A., Dodds S. J., 1994, Mon. Not. R. Astron. Soc., 267, 1020

Percival W. J. et al., 2001, Mon. Not. R. Astron. Soc., 327, 1297

Percival W. J. et al., 2010, Mon. Not. R. Astron. Soc., 401, 2148

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical recipes in FORTRAN. The art of scientific computing

Reid B. A. et al., 2010, Mon. Not. R. Astron. Soc., 404, 60

Sánchez A. G., Baugh C. M., Angulo R. E., 2008, Mon. Not. R. Astron. Soc., 390, 1470

Seo H.-J., Eisenstein D. J., 2003, Astrophys. J., 598, 720

Swanson M. E. C., Tegmark M., Hamilton A. J. S., Hill J. C., 2008, Mon. Not. R. Astron. Soc., 387, 1391

Wall J. V., Jenkins C. R., 2012, Practical Statistics for Astronomers

Wang Y., Chuang C.-H., Hirata C. M., 2013, Mon. Not. R. Astron. Soc., 430, 2446

Xu X., Cuesta A. J., Padmanabhan N., Eisenstein D. J., McBride C. K., 2013, Mon. Not. R. Astron. Soc., 431, 2834

Zehavi I. et al., 2002, Astrophys. J., 571, 172

Chapter 4

Summary

The main idea of this work is to investigate the potential of galaxy power spectrum as a tool for extracting properties of the Universe. I have showed that using direct Fourier transform (DFT) to obtain one dimensional power spectrum is more efficient than finding ad-hoc methods to recover original spectrum from fast Fourier transform (FFT) methods. Although I have used FFT to obtain two dimensional power spectrum, it is possible to use DFT given the enough computing power. Also, I have developed a method to extract two dimensional galaxy power spectrum from any redshift survey with a broad sky coverage. I have applied this method successfully to obtain SDSS 2D galaxy power spectrum and to measure $H(0.35)$ and $D_A(0.35)$ simultaneously, for the first time.

These techniques can be utilized to measure properties of the Universe from future galaxy surveys that contain more galaxies. For example, BOSS has made its initial data release already with a goal of measuring redshifts of over one million LRGs. This is a ten fold increase of data points which should translate to much tighter constraints on parameter values. Also, Eulid mission to be launched in 2020 with a goal of measuring 50 million galaxies will lead to much stringent constraints. With large data sets, it is possible to measure the evolution of H and D_A accurately by dividing the data sample into broad redshift slices. This is very important for understanding dark energy as the form of dark energy can be distinguished from this information. Although I have marginalized over several

cosmological parameters, accurate power spectrum measured from these future surveys will provide better estimates of parameters including the linear redshift space distortion parameter β.

Cosmology has evolved from a purely theoretical science to an active research area strongly coupled with observational data that can estimate various properties of the Universe with few percent accuracy. For example, my estimates are 4-5% accurate. Also, the ability to obtain model independent measurements from power spectrum as well as correlation function helps to determine the accuracy of the standard ΛCDM model. This increased accuracy will eventually lead to better understanding of the dark energy, the biggest problem in contemporary cosmology today.