



HAL
open science

Large Scale Structure with SDSS-III/BOSS

Mariana Vargas-Magaña

► **To cite this version:**

Mariana Vargas-Magaña. Large Scale Structure with SDSS-III/BOSS. Cosmology and Extra-Galactic Astrophysics [astro-ph.CO]. Université Paris-Diderot - Paris VII, 2012. English. NNT: . tel-00726113

HAL Id: tel-00726113

<https://theses.hal.science/tel-00726113>

Submitted on 29 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE PARIS DIDEROT

UFR de Physique

DOCTORAT de Physique

Mariana Vargas Magaña

Large Scale Structure with SDSS III- BOSS

Thèse dirigée par Eric Aubourg,
effectuée de septembre 2008-février 2012
au Laboratoire AstroParticule et Cosmologie
soutenue le 16 février 2012.

JURY

M. Jean-Paul KNEIB
M. Vincent BRETON
M. Stephanie ESCOFFIER
M. Pierre BINETRUY
M. Jim RICH
M. Eric AUBOURG

Rapporteur
Rapporteur
Membre de Jury
Président de Jury
Membre de Jury
Directeur de Thèse

*A mi papá, mi mamá,
Rosita , Germáncito y Sebastien*

La mejor forma que tengo para describir mi experiencia al hacer investigación en física "...es como un viaje a través de una oscura y desconocida mansión . Entrás en la primera habitación y está completamente a oscuras. Tropezas con el mobiliario que te hace caer, pero con el tiempo acabas sabiendo donde están los muebles. Finalmente luego de unos seis meses encuentras el interruptor de la luz, lo enciendes y todo se ilumina. Entonces descubres donde estabas exactamente. Luego te trasladas a la siguiente estancia y te pasas otros seis meses en la oscuridad. De manera que todos estos momentos de iluminación, algunos muy rápidos, otros de un día o dos, son la culminación de los muchos meses de tropiezos y caídas en la oscuridad que los precedieron, y sin los cuales no podían existir."

Andrew Wiles

En mi caso yo añadiría la mágica presencia de luciérnagas que iluminaron mi camino.

Remerciements

Il y a une "boule" des personnes a qui je doit dire merci...

D'abord, je tiens à remercier mes directeurs de thèse, Jean-Christophe Hamilton et Eric Aubourg, pour m'avoir guidé, encouragé, conseillé et m'avoir fait beaucoup voyager. A Jean-Christophe qui a pris la grande responsabilité de faire sortir cette thèse quand à la fin de l'épineuse deuxième année j'ai commencé a désespérer et grâce à qui j'ai beaucoup appris sur l'étude des grandes structures. A Eric Aubourg qui a toujours répondu à la minute à mes mails, et grâce à qui nous faisons parti de dans cette aventure de l'étude BAO dans la très grand collaboration SDSS-BOSS, avec les avantages et défis que cela implique. A Nicolas qui depuis son arrivée dans le groupe a apporté beaucoup de part ses connaissances a cette recherche et qui lorsqu'il s'est consacré à l'analyse Lyman-alpha a toujours suivi mon travail. A Antoine Labatie avec qui j'ai pu travailler, discuter et apprendre à propos de l'analyses des BAO. Je remercie les membres du jury, Stéphanie Escoffier, Jean-Paul Kneib, Vincent Breton, Jim Rich et Pierre Binetruy. Notamment Stephanie et Jean Paul qui m'ont apporte enormement de corrections pour améliorer mon manuscrit.

Je remercie le laboratoire APC qui m'a accueilli pour faire ma thèse, et plus particulièrement Pierre Binetruy, directeur du laboratoire APC, qui m'a fait l'honneur d'être président de mon jury de thèse. A Yannick, responsable du groupe de Cosmologie et Gravitation et qui a fait un suivi de mon parcours durant la thèse. A Cécile et Jean, pour leurs remarques et suggestions durant la préparation de la soutenance et avec qui j'ai aussi partagé beaucoup des discussions lors des cafés-preprint. Aux autres membres du groupe Cosmologie et Gravitation ainsi que les membres de l'administration du Laboratoire, Sabine, Martine Laird-Bardissa, Martine Piochaud, Aurélia et Sarodia. Je remercie les thésards qui ont toujours fait partie de mon chemin dans la recherche et le doctorat, qui sont aussi mes amies; a ceux avec qui j'ai commencé, Sébastien ,Romain, Silvia, Gaël et Clément qui mon fait apprendre beaucoup avec les cafés-preprints et avec qui j'ai toujours rigolé et ceux qui ont arrivés après comme Joseph, Guillaume et Josquin avec qui j'ai partagé beaucoup de choses, comme le bureau, les réunion et bien sûr le projet Cosmovan. Aux autres thésards des autres groupes, Ixandra, Massimo, a les thésards qui sont arrivés après comme Marie-Anne, Loïc, Alexandre, et mêmes ceux qui viennent d'arriver comme Julian, Fabien et Ivan avec qui j'ai passé de bons moments a Paris.

Je remercie mes amis de Paris et d'ailleurs qui m'ont toujours encouragé à finir la thèse. Tout d'abord mes inséparables amis Sandra, Yareli, Vero, Toño, Yessica, Ricardo, los Albertos, Ale, Lesdy, qui sont resté jusqu'à la fin, aux autres très chers amis qui ont été du debut Emilio, Eloy, Aleyda, Laura ma voisine, Laura et Aldo.

Je remercie ma famille, ma mère, mon père, ma soeur Rosita et mon frère Germancito qui sont toujours presents dans ma vie. Je remercie Sébastien qui m'a toujours accompagné, encouragé, soutenu et qui m'a fait le chemin plus heureux....

Analyse des structures à grande échelle avec SDSS-III/BOSS

Le travail de ma thèse s'est concentré sur l'extraction du signal laissé par les BAO dans la distribution des galaxies. Celui se présente sous forme d'un pic autour de 150 Mpc dans la fonction de corrélation à deux point de la matière. Cette échelle correspond à la distance parcourue par l'onde acoustique dans le fluide matière-radiation entre la période d'égalité matière-radiation et leur découplage à $z \sim 1100$. Il en résulte une sur-densité à cette échelle autour de chaque perturbation primordiale. Le but de mon travail était d'effectuer une analyse complète sur les données prises par la collaboration SDSS III/BOSS jusqu'à l'été 2011 afin d'apporter des contraintes sur les paramètres caractérisant l'énergie noire, w_0 et w_a . On veut savoir si ces observations sont consistantes avec la fameuse constante cosmologique ($w_0 = 1$ et $w_a = 0$) ou si l'équation d'état de l'énergie noire évolue avec le temps ($w_a \neq 0$) impliquant alors des scénarios bien plus complexes pour cette composante. Pour y arriver, j'ai étudié et testé les outils d'analyse avec des simulations log-normal ainsi qu'avec les données publiques du DR7 de SDSS. J'ai étudié les effets de distorsion des redshifts dans le régime linéaire avec des simulations et j'ai développé une méthode d'optimisation d'estimateur de fonction de corrélation. Dans un second temps j'ai réalisé l'analyse complète des données BOSS-CMASS, qui utilise les galaxies elliptiques très lumineuses (LRGs), depuis la construction du catalogue jusqu'à l'obtention des contraintes cosmologique en passant par la correction des effets systématiques.

Large scale structure with SDSS-III/BOSS

My research focuses on the extraction of the signal imprinted by the BAO on the galaxy distribution. The BAO feature is a bump at ~ 150 Mpc in the two-points correlation function of the matter in the Universe. This corresponds to maximum distance travelled by acoustic waves in the matter-radiation fluid during the period from matter/radiation equality to their decoupling at $z \sim 1100$. An excess density at a radius of 150 Mpc was left after decoupling around each dark-matter density peak. The final goal of my work is to perform a complete analysis of the data available up to summer 2011 in SDSS III/BOSS in order to give constraints on isotropised distance D_V and constraint the dark energy equation of state parameters w_0 and w_a . With such observations, one wants to check if the data is consistent with the so-called cosmological constant (corresponding to $w_0 = 1$ and $w_a = 0$) or if the dark energy equation of state shows a time evolution ($w_a \neq 0$) pointing to more complex dark energy scenarios. I studied and tested data analysis tools with log normal simulations and the public DR7 to achieve this goal. I studied the redshift distortions with simulations in the linear regime and I developed a optimization method for the correlation function estimator. The second part of my my work was focused in the analysis of BOSS data. There are to main programs in the SDSS III/BOSS project: the Luminous Red Galaxies catalogue and a new technique based on quasars using the Lyman-alpha forest. I have focused my research on the LRG science. I performed a complete analysis of BOSS-CMASS sample from catalog construction, systematic corrections and cosmological constrains.

Contents

I	Cosmology in a nutshell	5
I.1	Overview of Observational Cosmology	6
I.1.1	Hubble Law: Evidence of the expansion of the Universe	6
I.1.2	Light Elements Abundances validation of the Big Bang Nucleosynthesis	8
I.1.3	CMB: Evidence of thermal evolution and homogeneity	9
I.1.4	Beyond the Big Bang Model	11
I.2	Homogeneous Universe	16
I.2.1	General Relativity	16
I.2.2	Lemaître-Friedman-Robertson Walker Metric	17
I.2.3	Einstein’s Equation	18
I.2.4	Friedman equations	20
I.2.5	Contents of the Universe	22
I.2.6	Distances	25
I.2.7	Brief Thermal History of the Universe	28
I.3	Inhomogeneous Universe and structure formation	30
I.3.1	Inflation	31
I.3.2	Evolution of Density Perturbations	35
I.3.3	Statistical description of Density Fluctuations	38
I.3.4	CDM Power Spectrum	40
I.3.5	Spherical collapse model	41
I.4	Cosmological Parameters	42
II	Baryonic Acoustic Oscillations	45
II.1	Basic Framework	45

II.1.1	A not technical introduction to BAO	45
II.1.2	Configuration Picture of Acoustic Oscillations	47
II.1.3	Baryon Signatures in the CDM Transfer Function	49
II.2	Defining an Statistical Standard Ruler	52
II.3	Systematic Uncertainties from Theory	55
II.3.1	Non linearities	55
II.3.2	Bias	57
II.4	Approaches	58
II.4.1	Perturbations Theory	60
II.4.2	Reconstruction	66
II.4.3	Halo Model	67
II.5	State-of-the-art	69
III	From Sloan Digital Sky Survey to Baryonic Oscillation Spectroscopic Survey (SDSS-III/BOSS)	71
III.1	Sloan Digital Sky Survey	71
III.1.1	Original Project	72
III.1.2	SDSS II	74
III.1.3	Legacy of SDSS I and II	75
III.2	SDSS III	76
III.2.1	SEGUE-II	77
III.2.2	APOGEE	78
III.2.3	MARVELS	79
III.2.4	BOSS	80
III.3	Baryonic Oscillations Spectroscopic Survey in detail	81
III.3.1	Scientific Goal	81
III.3.2	Coverage and Schedule	82
III.3.3	Hardware:Camera, Spectroscope, Telescope	83
III.3.4	Between observations until reduced data	86
III.3.5	Target Selection Algorithms	88
IV	Study of correlation function estimators with simulations	93
IV.1	Log Normal Simulations	94

IV.2	Estimators	97
IV.2.1	Correlation Function Estimators	98
IV.2.2	Cosmic Error and Cosmic Bias	99
IV.2.3	Analytic expressions of mean and variance of estimators	103
IV.2.4	Analytical expression for bias and variance of the LS Estimator	105
IV.3	Random Generation	107
IV.3.1	Radial Selection Function	107
IV.3.2	Angular Selection Function	111
IV.4	Correlation function from Realistic Simulations	112
IV.4.1	Completeness of the Survey	113
IV.4.2	Radial Weights	116
IV.5	Error Estimation	118
IV.5.1	Errors from Simulations	119
IV.5.2	Error from Data	120
IV.5.3	Comparison of Covariance matrices	123
V	Large Scale Correlation Function from BOSS Data	133
V.1	Measuring the Correlation Function from BOSS Data	133
V.1.1	Samples for Large Scale Structure	135
V.1.2	Angular Selection Functions	138
V.1.3	Catalog definition	139
V.1.4	Targets Definition	139
V.1.5	North/South Offset	144
V.1.6	Target Failure Correction.	145
V.1.7	Fiber Collision Correction	148
V.1.8	Legacy Data.	151
V.1.9	Summary:Final Catalog	152
V.2	Observational Angular Systematic Errors	153
V.2.1	Stars	154
V.2.2	Correction of others sources of angular systematic errors	156
V.2.3	Iterative Weights	158
V.2.4	Linear Regression Method	159

V.2.5	Effect of angular corrections in correlation function	161
V.2.6	Summary of Correlation Function Corrections	162
V.3	Covariance Matrix	165
V.3.1	Simulations Description	165
V.3.2	Covariance Matrix Comparison	168
V.4	Summary	170
VI	Studying Redshift Distortions with simulations and more	175
VI.1	Introduction	175
VI.1.1	Linear Regime Description	177
VI.1.2	Multipolar developpement of 2PCF	178
VI.2	Estimation of the Linear Redshift parameter β	178
VI.2.1	Study Redshift Distortions with lognormal Random Fields	180
VI.2.2	Study non linear behavior effects HORIZON Simulation	182
VI.3	Measurement of Linear Redshift Distortion Parameter with BOSS	184
VII	Constraining Cosmology with BOSS	191
VII.1	Fitting procedure and Confidence Limits	191
VII.1.1	χ^2 Method	191
VII.1.2	Likelihood	192
VII.1.3	BAO detection and Significance	193
VII.2	Modeling the correlation function	194
VII.2.1	Dilatation Parameter	194
VII.2.2	Non linear effects	195
VII.2.3	Model of the correlation function the Official Version	196
VII.2.4	Extracting the BAO signal	196
VII.3	Parameter Constraints from BOSS	197
VII.3.1	Bin effect	198
VII.3.2	South and South(DR10)	201
VII.3.3	Final constraints $\Omega_m - \alpha$, $\Omega_m - D_V$ and discussion	203
VII.3.4	Significance of BAO detection with BOSS	206
VII.4	Conclusions (preliminary) and comments	210
VII.5	Perspectives	210

VIII	Optimization of correlation function estimator for BOSS survey	213
VIII.1	Motivation	213
VIII.2	Methodology	214
VIII.2.1	Simulations	214
VIII.2.2	Estimator Definition	215
VIII.2.3	Optimization criterium	216
VIII.3	Setting the bar for the optimization	217
VIII.3.1	Minimum Variance	217
VIII.3.2	Minimum Bias	217
VIII.3.3	Mean Square Errors	218
VIII.3.4	Landy-Szalay	218
VIII.3.5	Optimization Performance as a function r_{min}	219
VIII.4	Best Compromise $\lambda = 0.5$	220
VIII.4.1	Orders contribution to optimize estimator	220
VIII.4.2	Varying Geometry	222
VIII.4.3	Covariance Matrices	223
VIII.5	Testing if over-fitting	223
VIII.6	Conclusions and Perspectives	224

List of Figures

I.1	Hubble Diagram showing the relation between recession velocity of galaxies and distances.	6
I.2	A grid showing the comoving coordinates.	7
I.3	Constraints to the baryon density from the Big Bang Nucleosynthesis(^4He , deuterium, ^3H and lithium)	9
I.4	A view of the sky as would have been seen by the microwave receiver of Penzias and Wilson. The COBE (Cosmic Background Explorer) satellite measurements of the shape of the spectrum of this emission	10
I.5	The bullet cluster (1E 0657-558) consists of two colliding clusters of galaxies at $z=0.296$. A recent example of a rotation curve for M33 (21 cm) compared with best fitting-model.	12
I.6	CMB map seen by COBE Satellite in 1992. CMB seen 18 years later by WMAP created from 7 years of observations at 0.2 degree resolution.	13
I.7	The angular spectrum of the fluctuations in the WMAP full-sky map.	14
I.8	The first large scale survey, CfA, appears in 1985	14
I.9	Hubble diagram for 42 high-redshift Type Ia supernovae from the Supernova Cosmology Project. Best-fit confidence regions in the $\Omega_M - \Omega_\Lambda$ plane.	16
I.10	The total fractional density of the Universe is related to its local geometry.	21
I.11	Evolution of the energy density as function of the scale factor for different constituents is shown: radiation, matter and cosmological constant.	23
I.12	The scale factor varies with time and this variation is determined by the energy density.	24
I.13	Photons emitted at t_1, χ_1 by an object of size $l = a(t_1)\chi_1\Delta\theta$. The photons follows trajectories of constant (θ, ϕ) and they would be observed at (t_0, χ_0) with and angular separation given by $\Delta\theta = l/a(t_1)\chi_1$	27
I.14	The luminosity distance, comoving distance and angular distance for two cosmological models.	28
I.15	Illustration of Hubble radius.	33

I.16	The evolution of Hubble radius with time.	34
I.17	Evolution of radius of an spherical overdensity as a function of the θ parameter within the framework of the spherical collapse model	43
II.1	Propagation of a spherical wave.	46
II.2	Snapshot of evolution of the mass profile vs comoving radius of an initial point-like over-density located at the origin.	47
II.3	Snapshot of evolution of the mass profile vs comoving radius of an initial point-like over-density located at the origin.	48
II.4	Power spectra of matter and Angular power spectrum from CMB.	49
II.5	Examples of Transfer function of CDM with baryons for deferents set of parameters Ω_m and Ω_b from simulations and the fitting formula of Eisenstein and Hu.	51
II.6	Effect of baryons in power spectrum and correlation function.	52
II.7	The comoving size of this ruler in the transverse ($r_{ }$) and perpendicular (r_{\perp}) directions are related with the observables $\Delta\theta$ and Δz	53
II.8	First Detection of BAO in 2005.	54
II.9	Effets of non linearities in power spectra.	56
II.10	Effets of non linearities in correlation function.	57
II.11	Wiggles from power spectrum from N-body simulations considering 2 bias schemes	59
II.12	Derivates from biased power spectrum $dLnP/dLnk$ from N-body simulations considering 2 bias schemes.	59
II.13	The power spectrum from SPT is shown for 1-loop and 2-loop for Λ CDM model.	61
II.14	The $P_{PT}^{(1)}$ ($P_{RPT}^{(1)}$) denotes the 1-loop correction in PT (RPT)	61
II.15	The power spectrum and in right panel the two-point correlation function described by RPT at $z=0$	62
II.16	The 2 point correlation function in real space, in blue the linear one in red the non linear. The points corresponds to results from N-body simulations. We observe this shift in the peak location towards smaller scales . From	63
II.17	The correlation function is shown for various redshifts obtained with simulations and linear theory and the gaussian filtered linear theory power spectrum.	65
II.18	The wiggle part of the power spectrum from N-body-simulations in open circles, the linear theory power spectrum, the LPT power spectrum, the 1-loop SPT and the phenomenological model of the redshift power spectrum	65

II.19	The power spectrum from RPT, the closure theory for 1-loop and 2-loops (cyan) at redshift $z = 0$ and $z = 1$. The power spectrum from lagrangian perturbation theory, Large-N theory, time-RG theory, RGTP theory is shown at redshift $z = 0$ and $z = 1$ and the linear theory.	66
II.20	The non linear correlation function at $z=0$ and the reconstructed correlation function.	68
II.21	The power spectrum of the dark matter density field at the present time ($z = 0$) and the contributions to the power from the single and two halo terms.	69
III.1	Schlegel Extinction Maps and Stripes of SDSS.	73
III.2	Data Release 5 Coverage, in red footprint of photometric survey in green footprint of spectroscopic survey.	73
III.3	Data Release 7 Coverage, in red footprint of photometric survey and in green footprint of spectroscopic survey.	76
III.4	Coverage of SEGUE-I and SEGUE-II.	78
III.5	Coverage of APOGEE.	79
III.6	A science exposure from MARVELS.	80
III.7	BOSS Coverage.	83
III.8	SDSS-III Schedule.	83
III.9	SDSS Telescope.	84
III.10	A schematic view of CCD array and the Response Function of Filters	85
III.11	Scheme of BOSS spectrograph.	86
III.12	Illustration of Target Selection using AGES.	90
III.13	Comovil number density of Galaxy samples.	91
III.14	Comovil number density of QSO samples.	92
IV.1	In left panel an slice of the density field generated with a log normal simulation using a power spectrum as the input. In right panel the same slide populated with galaxies following the density field distribution of matter	96
IV.2	The power spectrum introduced to the simulation as input (in solid black line) and the power spectrum generated from the log normal density field (in dashed yellow line).	97

IV.3	In right panel, the correlation function using the five popular estimators from a realization from a lognormal simulation following a Λ CDM model with 3000 mock galaxies in the left and with 30000 in the right. In black, the input correlation function of the simulation. When the number of objects are small there are large fluctuations as we will see later dominated by the shot noise. As the number of objects increases the estimators converge to the input of the simulation.	99
IV.4	In left panel, the empiric mean of five popular estimators of the correlation function from 100 simulations log normal following a Λ CDM model is shown. In black, the input correlation function of the simulation. As we observe from the plot the less biased estimators are the Hamilton and Landy Szalay. In right panel the bias of the different estimators is shown, as defined by Eq.[IV.15] . . .	100
IV.5	Sample variance of five popular estimators of the correlation function from 100 simulations log normal following a Λ CDM model. In black, the input correlation function of the simulation. The points have been translated 0.5 Mpc for purposes. To isolate the effect of the variance in left panel I show the errors using the mean value of the LS estimator. In left panel the result using the mean value correspondent to each estimator is shown.	101
IV.6	The correlation function of one realization of 100 000 mock galaxies with a random set of same size and with a random 2 and 3 times greater than the data set. As we observe the shot noise decreases with the size of the random set used to compute the correlation function.	102
IV.7	1- σ error from log normal simulations for three different sizes of random set, in black a random set equal to data set, in red a random set 2 times greater than the data set and in blue three times. Average over 100 simulations of Landy Szalay in dashed lines, average of 100 realizations of Hamilton in solid lines. . .	103
IV.8	In black the input correlation function using the error bars (in red) given the analytical expression IV.19 compare with the error bars given by the standard deviation from simulations in yellow.	106
IV.9	The blue points is an example of a flux limited sample from SDSS DR7. The points constrained to the dashed lines corresponds to a Volume limited sample that could be defined from the first one. As we observe from the plot a volume limited catalog implies the loss of objects and thus a loss of information.	108
IV.10	The luminosity function of SDSS-Early Data Release in the r^* band. The data points are commissioning data and the black line corresponds to the Schechter fit. In left panel the redshift distribution of SDSS (histogram) compared with the expected average distribution. Figures from	109
IV.11	In left panel the number density of <i>Full Sample</i> and <i>Bright Sample</i> from samples defined by E.Kazin et al 2010. In right panel the correlation function using the samples	110

IV.12	The average correlation function over 100 realizations of 30,000 objects following different radial functions. In blue the parabolic, in yellow the gaussian and in red the uniform. The input correlation function is shown in black.	111
IV.13	An example of a section of an angular mask from SDSS DR7. In left panel the mangle polygons defining the mask , in the right side the same polygons showing the completeness weight of each region correspondent to the value of the completeness of the sample in each region.	112
IV.14	Left panel. The geometry of SDSS LRG sample DR7 in galactic coordinates. In the right panel the simplified mask used to generate the mock catalog is shown. It was generated using the tiles from DR7 concatenated with the photometric stripes.	113
IV.15	In left panel the redshift distribution of data fitted by a spline line to eliminate the noise is shown. This distribution is normalized and used to generate the radial selection function. In the right panel the comoving number density of data and mocks is shown.	114
IV.16	The average correlation function over 100 realizations of the sample in a cubic geometry compared with a set of samples following DR7 geometry. As we can observe from the figure the estimator shows a slight bias at the scales of interest and a increment in the variance in the case of DR7 geometry compared with the cube.	114
IV.17	The completeness of a region of DR7 Sample using the file generated by NYE value added Catalog	115
IV.18	The correlation function of DR7 Sample data	116
IV.19	In correlation function computed with no weights (black) and with FKP weight (red) fore single realizations.	117
IV.20	In right panel the mean correlation function computed with no weights (black) with FKP weight (red) .In right panel the relative effect of applying the FKP weights.	118
IV.21	The covariance matrix (left) and the correlation coefficients matrix (right) using 100 log normal simulations using the radial selection function from DR7 data and the angular selection function with weights equal to the completeness of the survey using files from the NYE Value Added catalog.	120
IV.22	In blue the mean correlation function from 100 bootstrap subsamples of 10,000 each. The error bars are given by the standard deviation. In black the correlation function from the whole sample.	121
IV.23	The covariance matrix (left) and the correlation matrix (right) calculated using the bootstrap method. The whole data sample is subsampled with repetitions allowed generating 100 realizations of 10,000 objets.	121

IV.24	In the left panel, the regions correspondent to each subsample used to estimated the error via the subsampling method. In right the correlation function of each subsample.	122
IV.25	In the left side the mean over the 10 subsamples with the error bars given by the standard deviation. In the right side the mean compared with the correlation function computed with the whole sample	123
IV.26	The covariance matrix (left) and the correlation matrix (right) using Subsampling method. The whole data-set has been divided in 10 non overlapping regions and the empiric variance have been computed from the samples.	124
IV.27	In the left panel, the regions correspondent to each subsample used to estimated the error via the jackknife method. In right, the correlation function of each subsample. As we observe compared with the Fig.[IV.24] the fluctuations of the correlation function are smaller given the size of the subsamples.	125
IV.28	In the left side the mean over the 10 jackknife subsamples with the error bars given by the standard deviation. In the right side the mean compared with the correlation function computed with the whole sample	126
IV.29	The covariance matrix multiplied by r^2 has been computed following the Eq.[IV.38]. The whole data-set has been divided in 10 non overlapping regions, and each subsample is generated removing the region i (Jackknife method).	127
IV.30	The diferent covariance matrix at the same scale. In botton from left to right, we have the result from simulation and the subsample. In the top the bootstrap and the jackknife.	128
IV.31	The diferent correlation coeficients matrix at the same scale. In botton from left to right, we have the result from simulation and the subsample. In the top the bootstrap and the jackknife.	129
IV.32	The errors given by the different methods from the diagonal terms of the covariance matrix.	130
IV.33	In left panel the errors coming from the diagonal of the covariance matrix compared with errors from the simulations lognormal. In right the $R(r)$ term defined as Eq. V.10 to compare quantitatively the errors coming from the off-diagonal terms.	130
IV.34	The means of the correlation function for each method. with which the covariance matrix is computed. As a reference de result from E. Kazin 2010 in black dashed line.	131
V.1	BOSS spectra separated by chunks. Includes Chunks 1-19	135
V.2	Diagram showing the concepts of sector, tile, mask. From http : //cas.sdss.org/dr7/135	
V.3	Comoving number density of BOSS Galaxy samples	137

V.4	Targets from the target run MAIN008 and the combined target list <i>GATH</i> for each	140
V.5	Targets from the target run <i>MAIN008</i> and the combined target list <i>GATH</i> for each chunk from 1-12.	142
V.6	Targets from the target run <i>MAIN008</i> and the combined target list <i>GATH</i> for each chunk from 13-14.	143
V.7	A) Comoving density from the target run MAIN008 for each chunk. B) Comoving density from the target run <i>GATH</i> for each chunk1-14.	143
V.8	Correlation function using the combined photometric files <i>GATH</i> compared with sample generated using only one reduction <i>MAIN008</i>	144
V.9	Left panel. Mean number density of <i>GATH</i> catalog separating North and South, with and without offset applied. Right panel. Mean number density of <i>MAIN008</i> and <i>GATH</i> catalogs, with and without corrections for whole sample.	145
V.10	The targets from <i>GATH</i> catalog in the color space. In blue the north hemisphere CMASS targets in red the South hemisphere with offset. As we observe this offset implies more restrictive cuts for south targets.	146
V.11	The correlation function of applying this offset in both catalogs <i>GATH</i> (right) and <i>MAIN</i> (left).	147
V.12	The correlation function of applying this offset in catalog <i>GATH</i> separating North and South in left in right the correlation function of south after and before offset correction.	147
V.13	The "target success rate" as a function of the galactic latitude.	148
V.14	The completeness as a function of the galactic latitude for the 3 methods of correcting the bad targets.	149
V.15	Correlation function computed with <i>GATH</i> catalog using different methods for correcting the bad targets. Top left for the hole sample, top right the north hemisphere and bottom south.	150
V.16	Coverage of Legacy Data, in cyan all data in red legacy data	152
V.17	Correlation function weighted with Legacy	153
V.18	The relation between mean CMASS target density and stellar density and the right panel shows the mean CMASS spectra versus stellar density.	154
V.19	The relation between mean CMASS target density and stellar density for the bins in galactic latitude: [-90, -40] [-40, 0] [0, 40]	155
V.20	Left panel. The relation between mean CMASS target density and surface masked by stars for the bins in galactic latitude: [-90, -40] [-40, 0] [0, 40]. Right panel. The slope of the linear fit for the relationship between mean galaxy density and surface masked by the stars as a function of the bin in galactic latitude	156
V.21	The correlation function is shown after and before stellar correction	157

V.22	I-band sky flux (Sky) and I band PSF FWHM (Seeing) maps with HEALPIX pixelization of nside=256.	157
V.23	Galactic extinction (Ar) and Airmass maps with HEALPIX pixelization of nside=256.	158
V.24	Stellar Density and CMASS galaxy density with HEALPIX pixelization of nside=256.	158
V.25	The dependency of the mean density with the systematic is shown before correction as well as the spline line fit for the airmass and sky	159
V.26	The dependency of the mean density with the systematic is shown before correction as well as the spline line fit for the dust and seeing.	159
V.27	Correlation function with the iterative weights using A. Ross files and applying an iterative correction with spline for the dependance in the systematic as a function of the mean density.	160
V.28	The dependency of the mean density with the systematic is shown before and after correction as well as the linear line fit for dust and seeing.	161
V.29	The dependency of the mean density with the systematic is shown before and after correction as well as the linear line fit for dust and seeing.	162
V.30	Weights computed with the iterative method compared with weights computed with the linear regression.. . . .	163
V.31	Correlation function with weights computed separately for each systematic . . .	163
V.32	The correlation function computed with the different different target failure correction before and after correcting from angular systematics for the hole sample	164
V.33	The correlation function computed with the different different target failure correction before and after correcting from angular systematics separating North and South	164
V.34	Left panel. Coverage of LAS DAMAS CMASS simulation. Right panel. The projected correlation function for the first semester of BOSS Data used to calibrate the simulations	166
V.35	Mean correlation function of 600 PTHALOS simulations	167
V.36	Boss Correlation matrices for the North (left) and South(right) hemispheres generated from PTHALOS simulations.	168
V.37	Boss Covariance matrices for the North (left) and South(right) hemispheres generated from PTHALOS simulations.	169
V.38	Boss Covariance matrix and the correlation matrix for the all sample generated from LOG NORMAL simulations.	169
V.39	Boss Covariance matrix and the correlation matrix for the all sample generated from LAS DAMAS CMASS simulations for regions A,B,C	171
V.40	The correlation function with the final catalog for the North and the South. The errors bars given by the diagonal terms of the covariance matrices from PTHALOS.	172

V.41	The correlation function with the final catalog for combined sample and weighted sample. The errors bars given by the diagonal terms of the combined covariance matrices from PTHALOS.	173
VI.1	Illustration of Redshift Distortions from Hamilton	176
VI.2	Theoretical prediction of $\xi(r_{\parallel}, r_{\perp})$ in the linear regime approximation in real and redshift space	179
VI.3	Theoretical monopole, quadrupole and hexadecapole for linear redshift distortions	179
VI.4	Slide of lognormal simulation in real and redshift espace	181
VI.5	Left Panel $\xi(\theta, s)$ in real space (without redshift distortions), Top Right panels corresponds to $\xi_0(s)s^2$, $\xi_2(s)s^2$ and $\xi_4(s)s^2$ estimate from 200 Monte Carlo realizations of a lognormal simulation in real espace (red points), the blue dotted line is the linear correlation function without Redshift distortions and the green dotted line the linear correlation function with redshift distortions	182
VI.6	Left Panel $\xi(\theta, r)$ with redshift distortions from lognormal simulations, Top Right panels corresponds to ξ_0r^2 , ξ_2r^2 and ξ_4r^2 estimate from 200 Monte Carlo realizations of a lognormal simulation in redshift espace (red points), the blue dotted line is the linear correlation function without Redshift distortions and the green dotted line the linear correlation function with redshift distortions	183
VI.7	Estimation of β parameter using the $\xi_s(r)/\xi_0(r)$ with redshift distortions from lognormal simulations. The input corresponds to the dotted line.	184
VI.8	The power spectrum from the simulation Horizon – 4π measured with powergrid	185
VI.9	Left Panel $\xi(r, \theta)$ in real space (without redshift distortions) from HORIZON – 4π simulation, Top Right panels corresponds to $\xi_0(s)s^2$, and $\xi_2(s)s^2$ estimate from 200 Monte Carlo realizations of halos from HORIZON – 4π simulation in real espace (red points).	185
VI.10	Left Panel $\xi(r_r, \theta)$ with redshift distortions from HORIZON – 4π simulation, Top Right panels corresponds to $\xi_0(s)s^2$ and $\xi_2(s)s^2$ estimate from 200 Monte Carlo realizations of from HORIZON – 4π simulation in redshift espace (red points) . . .	186
VI.11	Left Panel $\xi(r_{\parallel}, \pi)$ in real space (without redshift distortions) from HORIZON – 4π simulation, Top Right panels corresponds to $\xi_0(s)s^2$ and $\xi_2(s)s^2$ estimate from 200 Monte Carlo realizations of halos from HORIZON – 4π simulation in real espace (red points).	186
VI.12	Left Panel $\xi(r_{\parallel}, \pi)$ with redshift distortions from HORIZON – 4π simulation, Top Right panels corresponds to $\xi_0(s)s^2$ and $\xi_2(s)s^2$ estimate from 200 Monte Carlo realizations of from HORIZON – 4π simulation in redshift espace (red points) . . .	187
VI.13	Estimation of β parameter using the $\xi_s(r)/\xi_0(r)$ with redshift distortions from HORIZON – 4π simulations, where $\xi_0(r)$ is the input correlation function generating the HORIZON – 4π simulation. The input in $\beta = 1$ parameter.	188

VI.14	$\xi(r_{ }, r_{\perp})$ for BOSS CMASS galaxies in redshift space	189
VI.15	$\xi_2(r)$ for BOSS CMASS galaxies in black, best fit model in blue	189
VI.16	$Q(s)$ for BOSS CMASS galaxies in black (PRELIMINARY PLOT). In left panel the result for the estimation of $Q(s)$ for SDSS-DR6 from	190
VII.1	Calibrating the σ_{NL} with simulations PTHALOS. In black the mean of the 600 realizations of the PTHALOS simulations using only the north hemisphere. The $\sigma_{NL} = 6.25$	198
VII.2	North: Fitting alpha, bias . WMAP as fiducial cosmology-bin 7Mpc.	199
VII.3	Combined: Fitting alpha, bias ,bin 7Mpc width.	199
VII.4	North: Fitting alpha, bias bin 7Mpc-width	200
VII.5	Combined: Fitting alpha, bias bin 4Mpc width.	200
VII.6	South _{DR9} : <i>Fitting alpha, bias. Bin 7Mpc</i>	201
VII.7	South DR10: Fitting alpha, bias . Bin 7 Mpc	202
VII.8	South DR9: Fitting alpha, bias . Bin 4 Mpc	202
VII.9	South DR10: Fitting alpha, bias . Bin 4 Mpc	203
VII.10	1-2 σ Contours for North (blue), South(red), Combined (black) and the South DR10(green). Bin 4 Mpc	204
VII.11	Contours confidence $\alpha - \Omega$ North	205
VII.12	Contours confidence $\alpha - \Omega$ North/South	205
VII.13	Contours confidence $\alpha - \Omega$ South DR9	206
VII.14	Contours confidence $\alpha - \Omega$ South DR10	206
VII.15	Contours confidence the physical matter density and the distance scale $\Omega_m h^2 -$ $D_V(z = 0.55)$	207
VII.16	Best fit parameters for model no wiggles	208
VII.17	Best fit parameters for model no baryons	209
VIII.1	Relative performance of estimators depends on geometry. In top left panel the mean of 300 log normal simulations in a cubic box and in top right panel following the BOSS geometry for different estimators. In bottom panels, in left side the bias (scaled by r^2) and in right the variance (scaled by r^2) for the different estimators in a Cubic and in BOSS geometries.	214
VIII.2	In left panel the average correlation function of the 300 simulations in the Cubic and BOSS geometries, compared with the input model and the correlation func- tion of the lognormal density field. In right panel the deviations of the average correlation function of simulations in the box and BOSS geometry from the input correlation function (scaled by r^2) with the 1σ intervals are plotted also.	215

VIII.3	Minimal Variance estimator ($\lambda = 0$). In left panel the bias (solid line) and error (dotted line) for the optimal (red) and the LS (blue) estimators. In right panel, the signal to noise ratio for the bias corrected (dotted line) and non corrected (solid lines) for LS(red) and BOSS (blue) estimators.	217
VIII.4	Minimal Bias estimator ($\lambda = 0$). In left panel the bias (solid line) and error (dotted line) for the optimal (red) and the LS (blue) estimators. In right panel, the signal to noise ratio for the bias corrected (dotted line) and non corrected (solid lines) for LS(red) and BOSS (blue) estimators.	218
VIII.5	Mean Square Error estimator ($\lambda = 0.5$). In left panel the bias (solid line) and error (dotted line) for the optimal (red) and the LS (blue) estimators. In right panel, the signal to noise ratio for the bias corrected (dotted line) and non corrected (solid lines) for LS(red) and BOSS (blue) estimators.	219
VIII.6	Landy-Szalay estimator ($\lambda = 0.96$). In left panel the bias (solid line) and error (dotted line) for the optimal (red) and the LS (blue) estimators. In right panel, the signal to noise ratio for the bias corrected (dotted line) and non corrected (solid lines) for LS(red) and BOSS (blue) estimators.	219
VIII.7	Mean Bias and mean RMS evolution with λ for different minimal distance considered in the fit (r_{min}).	220
VIII.8	The bias of Landy Szalay and the optimal estimator is shown and the ratio of the optimal error bars respect to Landy-Szalay	221
VIII.9	The bias (solid lines) and variance (dotted lines) are shown for the three cases, first order (green), adding the second order (yellow), using all terms (red) and compared to the reference the Landy Szalay estimator (blue). The yellow is so close to red that it is almost indistinguishable.	221
VIII.10	Effect of the optimization in a given geometry and applied to a simpler/complex one.	222
VIII.11	In top panel, Covariance matrices of Landy-Szalay estimator (left) and new estimator (right). In Bottom panel correlation matrices for the LS estimator (left) and new estimator (right)	223
VIII.12	Testing if over-fitting. In solid lines the bias and in dotted lines the rms. We plot 2 cases, in green the optimization and test are done with the training sample in red the optimization is done with the training sample and the test with another sample	224

List of Tables

I.1	Timeline of the Universe	29
I.2	Evolution of large scale perturbations.	37
I.3	Evolution of large scale perturbations.	40
II.1	Measurements of Baryonic Acoustic Feature.	70
III.1	Error estimation for BAO measurement.	82
III.2	Forecast of parameter constraints over the near future and current BAO.	82
VII.1	Constrains for 4Mpc and 7Mpc bins North , South and Combined Samples	201
VII.2	Constrains for 4Mpc and 7Mpc bins South DR9 and South DR10	203
VII.3	Constrains for D_V and $\Omega_m h^2$	204
VII.4	Constrains for D_V and $\Omega_m h^2$ from other studies	207

Introduction

The current paradigm of cosmology is the Λ CDM model. A number of observational evidences place this model as the best description we have of the universe up to now. However, our understanding of the origin and evolution of the universe is not completely decrypted. There are still fundamental open questions. Among these open questions the nature of the dark energy is one of the most important issues to be explored by modern observational cosmology. There are several probes that permit to extract some informations about the equation of state of this unknown component of the universe: clusters, weak lensing, supernovae and the baryonic acoustic oscillations (BAO). My research focuses on the extraction of the signal imprinted by the Baryonic Acoustic Oscillations on the galaxy distribution. This thesis is part of the experimental efforts to learn more about the nature of this elusive component of the universe: the dark energy.

The BAO feature is a bump at $\sim 150Mpc$ in the two-points correlation function of the matter in the Universe. This corresponds to the maximum distance travelled by acoustic waves in the matter-radiation fluid during the period from matter/radiation equality to their decoupling at $z \sim 1100$. An excess density at a radius of 150 Mpc was left after decoupling around each dark-matter density peak. We therefore expect today an excess probability to find matter separated by this distance (in comoving coordinates). This feature can be seen as a standard ruler allowing to study the history of the expansion of the Universe and infer cosmological information. The physics underlying the acoustic oscillations is well understood and the value of this standard ruler has been measured with exquisite accuracy at a redshift $z \sim 1100$ using WMAP data. The apparent BAO distance scale measured from observations at different redshift leads to measurements of the Hubble parameter $H(z)$ and the angular diameter distance $D_A(z)$ which are related to the cosmological parameters, especially to the dark energy parameters (See Eisenstein 2003 [23]). One can look for the imprint of the BAO feature in the galaxy distribution as galaxies are good tracers of the dark matter density distribution. The first detection of the BAO in the galaxy distribution was obtained with the Sloan Digital Sky Survey (SDSS) in 2005 (Eisenstein et al 2005[4]) with a sample of 40000 Luminous Red Galaxies in a volume of $0.72(Gpc/h)^3$. Many analysis have been performed with different methods using both the Fourier space power spectrum (Percival 2006 [16] and 2009 [17]; Padmanabhan 2006[14]) or configuration space two-points correlation \mathfrak{D} the inverse Fourier transform of the former (Eisenstein 2005 [4]; E.Kazin 2008 [7]). Both the Luminous Red Galaxy sample or the main one were used with successive releases of the SDSS data. Moreover, some analysis have used combined dataset (SDSS and 2DFGRs) or larger samples using photometric redshifts instead of spectroscopic ones. All of these analysis have now led to a convincing detection of the BAO feature. The main difficulty for detecting and analyzing the BAO scale in the large scale structure of the Universe is its low statistical significance due to the sparse sampling of the underlying dark matter distribution by

luminous galaxies. One can only expect to be able to observe the BAO feature in the widest galaxy surveys and it has only been unambiguously detected in the most extended samples like Luminous Red Galaxies.

The final goal of my work is to perform a complete analysis of the data available up to summer 2011 in SDSS III/BOSS in order to give constraints on the dark energy equation of state parameters w_0 and w_a . With such observations, one wants to check if the data is consistent with the so-called cosmological constant (corresponding to $w_0 = 1$ and $w_a = 0$) or if the dark energy equation of state shows a time evolution ($w_a \neq 0$) pointing to more complex dark energy scenarios. BOSS is one of the 4 programs composing the third version of the Sloan Digital Sky Survey (SDSS-III). There are two main programs in the SDSS III/BOSS project: the Luminous Red Galaxies (LRG) catalog and a new technique based on quasars using the Lyman-alpha forest as a tracer of matter in the line of sight. I have focused my research on the LRG science.

In order to constrain theories against observations the accurate estimation of clustering statistics from galaxies surveys is necessary. The most widely used statistic in large scale structure analysis is the two-point correlation function in real and Fourier space. In this work I use the correlation function analysis, which is more adapted to explain the physics involved in BAO. My work could be separated in 2 levels : a preparatory work done with simulations to study in details the estimators and the error analysis that will be used in further chapters to analyze the large scale structure in BOSS data. The second level was to perform a complete analysis of BOSS galaxies, from catalog generation to cosmological constraints, using this tools previously developed, tested and once the estimators were completely understood from the study with mocks.

Thus, in a first time I concentrated on the properties of the estimators of the correlation function using mock catalogs. I studied the different estimators available in the literature in terms of the bias and variance. As we will see, the estimators of the correlation function are based on the comparison of the galaxy distribution with a random distribution that mimics the observational conditions of data samples. I therefore spent time understanding the elements that have to be taken into account in random generation to have an accurate estimation of clustering statistics: the radial and angular selection functions. I also explored the weights.

A second important ingredient for doing clustering analysis is an accurate estimation of the errors. Measurements in galaxy catalogs are subject to uncertainties that must be properly addressed in order to compare with theoretical predictions. These uncertainties can be separated according to their nature in statistical and systematic. The statistical errors are the shot noise and the cosmic variance. As part of this preliminary work to prepare the BOSS data analysis I studied in this first stage the systematic and statistical uncertainties using mock catalogs and some data based methods.

Concerning the systematic errors, they could be separated in theoretical and observational. Among the uncertainties that comes from the theoretical side we have the effect of peculiar velocities. The fact that we only have access to redshift as an estimator of the radial distance of galaxies introduces distortions in the direction parallel to the line of sight due to peculiar velocities of the galaxies. Although, these distortions introduce an extra-complication in the physical interpretation of the results, they also provide extra information about cosmological parameters and the mass-luminosity bias of the tracers we use. I devoted a chapter to study with lognormal simulations the effect in the correlation function. There are other uncertainties related with theory as the

galaxy bias. The galaxies distribution do not follows exactly the matter distribution and finally the deviations from the linear description. These issues will be presented in the BAO chapter and they will be discussed again in the chapter devoted to cosmological constraints. Finally the observational errors arise from technical limitations due to telescopes and instruments. As an example we have not perfect sampling (close pairs of galaxies), effects of dust extinction, contamination from sources (stars), among others. In the chapter devoted to analysis of BOSS data I will talk about the observational bias and errors specific to BOSS .

Chapter 5 is devoted to the heart of my thesis the BOSS correlation function. In this chapter, I use the first two years of BOSS data to compute the correlation function of galaxies. These data constitute the Data Release 9 (DR9) that will be public in summer 2012. The goal of my work was to generate a complete chain of analysis from catalog generation to the cosmological constraints. The first step to do clustering analysis is to generate a 'clean' catalog from noisy data. In the first section I discuss this process. The definition of a catalog is not a simple task because we must guaranty that the sample selected is homogeneous and there is no systematic errors or bias originated by observational issues. The second part is devoted to the computation of the correlation function and the correction from angular systematic effects. The main systematic error found in BOSS data is the stellar density (Ross et al., 2012). I explore in detail this source of systematic error. I study the effects on the correlation function of other observational variables as the seeing, extinction, airmass and sky flux. Finally, in the third part I discuss the error estimation. As we will see the common method to estimate the errors is computing the covariance matrix from simulations. In this section I explore three different kind of simulations that would permit to validate the error estimation. The chapter 7 is the culmination of the chain in the context of a large scale structure analysis: the cosmological constraints generated for the fitting of the correlation function from BOSS data. The final goal of the observational cosmology is to confront the the results generated from the statistical data analysis with the predictions from theory. This chapter is devoted to the cosmological constrains we can get from the analysis of BOSS data.

Finally the chapter 8 is like the *cerise sur le gateau*, it is inspirited from the work done in chapter 4 for the estimators and as well is closely linked with the heart of the thesis the BOSS data correlation function, this chapter presents a methodology to define a new optimized estimator that seems to perform better that the universal estimator for the correlation: the Landy-Szalay estimator.

A summary of global conclusions is presented at the end, it does not mean that the research is finished by contrary the idea is to present a *bilan* of the results generated of this first stage, the graduate studies, and to enumerate the perspectives that results from this first exploration.

I. Cosmology in a nutshell

In the current state of cosmology, there is an almost universal consensus about which is the best description of our Universe. That model is the Standard Cosmological Model. The most exciting feature of modern cosmology is that now we have an enormous amount of observational data that enable us to test predictions from theory. In that sense, it is usually said that cosmology has passed to the era of precision measurements. This chapter introduces the basic notions of modern cosmology in order to provide the context of my research. Before I describe the standard cosmological model, also known as Λ CDM model, I discuss the basic picture in cosmology: the Big Bang theory. I begin with an overview of the observational evidences that supports the Big Bang cosmology. Then, I summarize observational evidence that required to go beyond this model, such as the evidence of existence of non baryonic matter, density perturbations and an extra component of the energy density.

The second part of the chapter is devoted to the description of the homogeneous Universe. I develop the fundamental framework of standard cosmology established by the Einstein equation in a homogeneous and isotropic metric, the so called Lemaitre-Friedman-Robertson-Walker metric, giving place to the Friedman equations. The Friedman equations describe the dynamics of the evolution of the Universe. I present the components of the Universe and their evolution. Afterwards I discuss the notion of distances, that is a key element related to my research subject. I finish this section giving a brief description of the thermal history of our Universe. The third part is devoted to the perturbed Universe. We have observed that the Universe is not exactly homogeneous and isotropic, we observe structures (i.e. voids, clusters and filaments). In order to explain the transition from a description of an homogeneous Universe to an inhomogeneous one, I begin the third section discussing inflation as a way to solve some of the problems of the Big Bang model. I present the statistical description of density fluctuations and their evolution followed by a brief discussion on the instability paradigm, using the model of spherical collapse that, even if it is an approximation, illustrates the formation of structures from initial density fluctuations. Finally, I summarize the cosmological parameters of the Λ CDM model and discuss some of the current unanswered questions that will drive the further research in this domain.

I.1 Overview of Observational Cosmology

The modern cosmology emerges with the Big Bang paradigm in the 30's. The fundamental features of the standard cosmological model come from this original model. Usually three observational evidences are recognized as the pillars of the Big Bang Model: the Hubble diagram that exhibits the expansion of the Universe, the light elements abundance in agreement with Big Bang Nucleosynthesis(BBN) predictions, and the measurement of the Cosmic Microwave Background (CMB) evidence of the thermal evolution (hot and dense past) and homogeneity of our Universe (from the black body radiation spectrum). In fact the BBN provides too an evidence of the hot and dense past of our Universe.

I.1.1 Hubble Law: Evidence of the expansion of the Universe

The key piece of modern cosmology is the evidence that our Universe is expanding. Edwin Hubble measured in 1929 (Hubble, 1929) and 1931(Hubble & Humason, 1931) the velocities of Cepheids and estimated their distances using an empirical period-luminosity relation (see Fig. [I.1]).

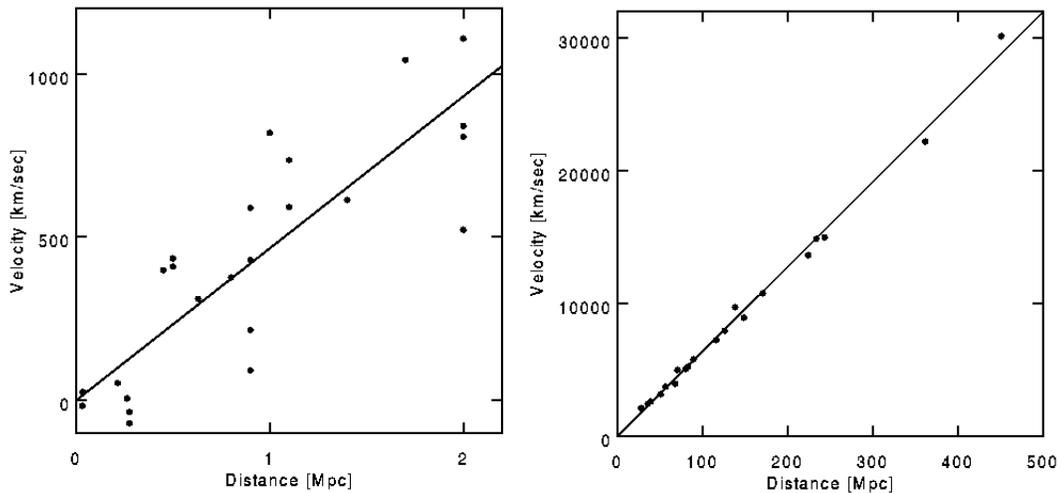


Figure I.1: Hubble Diagram showing the relation between recession velocity of galaxies and distances. Left panel: original data from Hubble. (Hubble, 1929). Right panel: recent data with more distant galaxies. Both figures from Edward. L. Wright <http://www-cosmosaf.iap.fr/Cours-cosmo-1.htm>

Hubble found that the extragalactic galaxies appears to be moving away from us. He also noticed that the velocity increases with distance. If we believe in the cosmological principle, that tell us that we do not have a preferred place in the Universe, Hubble's findings lead immediately to the conclusion of an expanding Universe: each observer in the Universe detects galaxies receding at a velocity proportional to their distance with respect to them by a universal factor. This phenomenon is what we would observe in an expanding Universe. A convenient way to describe this effect, in the context of general relativity, is to introduce a scale factor $a(t)$ that evolves with time. As the time passes, the expansion increases the physical distances by a factor $a(t)$. An illustration of expansion

is shown in Fig. [I.2]. In the figure, a grid representing the comoving coordinates is shown. A comoving frame include directly the expansion effect in the coordinate system. The expansion increase the physical distances by a factor $a(t)$, the points on the grid maintain their coordinates and the comoving distance remains constant. The value of the scale parameter is set to one now ($a_0 = 1$). The index 0 in cosmology refers to quantities evaluated at present time.

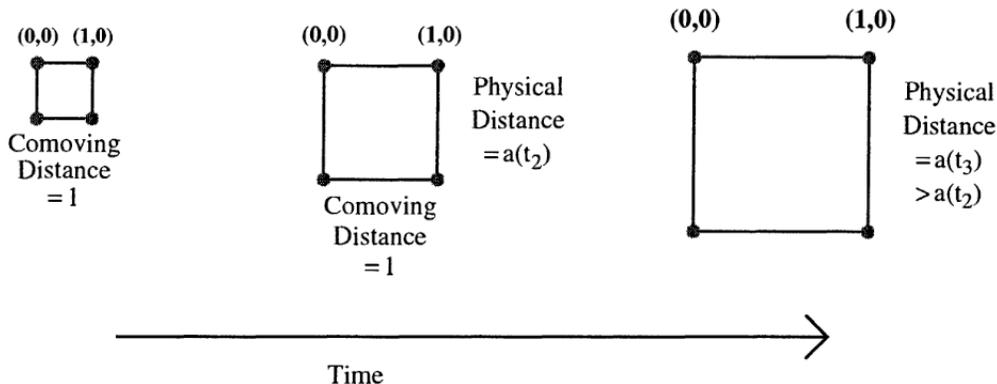


Figure I.2: A grid showing the comoving coordinates, the expansion increase the physical distances by a factor $a(t)$, the points on the grid maintain their coordinates and the comoving distance remains constant. From (Dodelson, 2003).

A useful quantity defined from scale the parameter is the Hubble parameter $H(a)$:

$$H \equiv \frac{\dot{a}}{a} \quad (\text{I.1})$$

The Hubble parameter tells us the expansion rate of the Universe. The present value of H denoted by H_0 is called the Hubble constant and its value is $H_0 = 100h \text{ km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1}$ where $h=0.738 \pm 0.024$ (Riess et al., 2011). Hubble's findings are known as the Hubble Law and is expressed as follows:

$$v \simeq Hd \quad (\text{I.2})$$

As this equation shows, the slope of the curve in Fig. I.1 corresponds to the Hubble parameter. I will now introduce the basic concept of redshift. In cosmology, velocities are measured via their redshift which is basically the Doppler effect applied to light waves. If galaxies recede from us, the expansion of the Universe stretches the wavelength and thus the frequency is lower, spectral lines go to the red frequencies so the effect is known as redshift. The standard non relativistic Doppler formula give us the relation between velocity and the redshift :

$$1 + z \equiv \frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} = \frac{1}{a} \quad (\text{I.3})$$

with $z = \frac{v}{c}$. This interpretation of cosmological redshift in terms of a Doppler shift that indicates a recession velocity is known as the kinematic interpretation. This interpretation is clear when taking the limit $z \ll 1$, but when redshift becomes large it could lead to confusion and temptation of using the relativistic expression. In fact, the redshift must be interpreted as an accumulation

of infinitesimal Doppler shifts caused by photons passing between a family of comoving observers separated by small distances along the photons paths. This interpretation is based on the fact that in an expanding universe the curvature is small in nearby regions, thus we can apply special relativity and interpret the shift in the frequency of photons as a recession velocity.

This recession velocity is in fact a relative velocity between the observer and the object, and in general relativity there is not a unique way of comparing vectors for widely separated points. In order to define a relative velocity the vectors must be transported parallelly, and depending on the chosen path it would lead to different results.

When carrying the vector along the light path corresponds to the interpretation that I presented. As we see, the interpretation of redshift is not trivial, for a discussion about interpretation see Bunn & Hogg (Bunn & Hogg, 2009) (Chodorowski, 2011). Even if the interpretation is complex, the physics is clear. The light from distant galaxies is redshifted, and this provides the observational evidence for the idea of the expanding Universe.

This notion of an evolving Universe is the fundamental idea at the base of Big Bang theory. If the Universe is expanding now, going to the past, the distance between objects was smaller than it is today. Extrapolating this idea back in time we find a time t_{\min} where scale parameter becomes zero $a(t_{\min}) = 0$, thus giving us the notion of the first instant of the Universe, an initial spatial singularity at a time t_{\min} . A second implication of this idea on expansion is that in the past the density of energy was higher, the energy content of Universe contained in a contracted volume extrapolating to the initial time will give us a singularity in the energy density. This fact gave the name of Hot Big Bang model or "great explosion" in a sarcastic manner when this model emerged in the times where stationary models were still in vogue. Even if this image of explosion is misleading, it represents well the idea of the first instants of our Universe as hot and dense. An important remark that permit us to avoid this problem of an initial singularity is that when the Universe exceeds certain density, quantum effects must be important even for gravitational physics. This should happen at Planck energy scale $E_{\text{Planck}} = (hc^5/G)^{1/2} = 1.22 \times 10^{19}\text{GeV}$. The density had achieved this value at the Planck time. The Planck time delimits the validity of our current theoretical framework, its value is $t_{\text{Planck}} = (hG/c^5)^{1/2} = 5.39 \times 10^{-44}\text{s}$. It is from this initial time that we begin the description of the Universe avoiding to deal with these singularities.

I.1.2 Light Elements Abundances validation of the Big Bang Nucleosynthesis

In the context of Big Bang model, the early Universe was hot and dense. At the first seconds of the Universe, the vast amount of energetic photons prevented nuclei to be formed. As the Universe expands, it gets cooler. When Universe achieve an energy below the binding energy of nucleus, the first nuclei begin to form, these process is called nucleosynthesis.

Knowing the conditions at early Universe and using our knowledge of nuclear physics about relevant cross sections, it is possible to estimate the primordial abundances of elements. These predictions depend on the density of protons and neutrons at the time of nucleosynthesis. The Fig. I.3 shows the predictions of light elements (^4He , deuterium, ^3H and lithium) for a range of baryon densities. The measurements of primordial abundances are shown in form of squares. As the figure shows, the measurements of light elements are consistent with the Big Bang nucleosynthesis. Notably deuterium measurements ((Burles & Tytler, 1998)) constraint the baryon density to only

a few percent of the critical density.

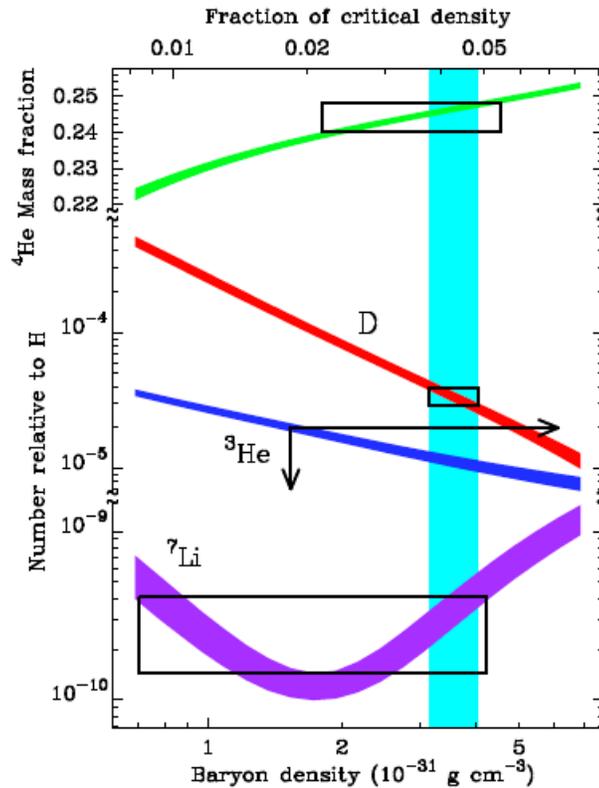


Figure I.3: Constraints to the baryon density from the Big Bang Nucleosynthesis [^4He , deuterium, ^3H and lithium]. Solid vertical is fixed by measurements of primordial deuterium (Kirkman et al., 2003)

Thus, the Big Bang Nucleosynthesis provides an estimate of the baryon density in the Universe. Baryons contribute at most to a 5% to the density matter. We will see later that the matter density is estimated to be $\sim 20 - 30\%$, thus nucleosynthesis provides an argument to suspect of existence of non baryonic matter. Big-bang nucleosynthesis (BBN) provided a key test of the hot Big-Bang cosmology, the agreement between the light element abundances measurements and predictions, confirmed the basic idea of Big Bang: the thermal evolution of our Universe.

I.1.3 CMB: Evidence of thermal evolution and homogeneity

The most direct evidence for a hot and dense past comes from the cosmic microwave background (CMB). Penzias and Wilson measured in 1965 (Penzias & Wilson, 1965) the cosmic background radiation from the cosmos finding a $T = 3.5 \pm 1\text{K}$. This result was in agreement with predictions from the Big Bang model done by Dicke et al. (Dicke et al., 1965) in 1965 of a background radiation at this temperature. In fact, as we will see later, photons of the CMB were emitted when the

temperature of the Universe was about $T \sim 3000\text{K}$. These photons, because of the expansion, are redshifted by a factor and now we measured this low temperature.

In 1992 the COBE (Cosmic Background Explorer) satellite measured the shape of the spectrum of this emission. It matched a perfect blackbody spectrum at a temperature of $T = (2.736 \pm 0.017)\text{K}$. (Smoot et al., 1992). In Fig. [I.4], it is shown the intensity of the cosmic microwave radiation as a function of the wavelength, measured by many experiments in a later time. It shows a perfect black body spectrum that corresponds to $T = 2.726\text{K}$. This spectrum provided a direct evidence of an homogeneous temperature in all directions at 3 decimal precision (Mather et al., 1990).

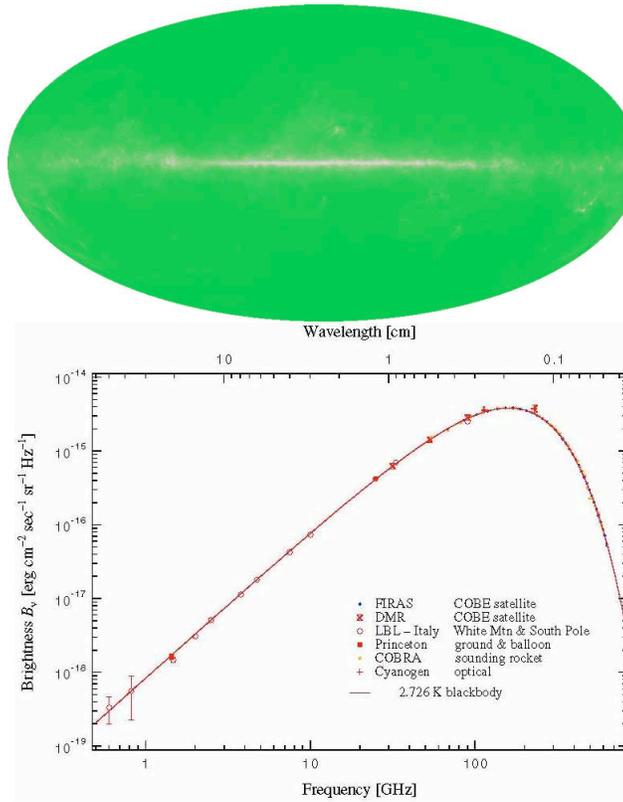


Figure I.4: Left panel. A view of the sky as would have been seen by the microwave receiver of Penzias and Wilson, if it could have surveyed the whole sky. From <http://map.gsfc.nasa.gov/>. On the right panel, the COBE (Cosmic Background Explorer) satellite measurements of the shape of the spectrum of this emission, showing a perfect blackbody at a temperature of 2.726K ⁽¹⁾

We will see later that this fact cause some problems, because an homogeneous temperature implies thermal equilibrium at the early Universe and therefore a causally connected Universe. I will discuss later how Inflation envisages to solve this problem of the Big Bang model.

In the framework of the Big Bang Model, the origin of the Cosmic Microwave Background goes back to early times. After the nucleosynthesis, energy of photons was still high and atoms could not be formed, the Universe continued expanding, decreasing the energy density. At a temperature

equivalent to $T = 3 \times 10^3$ K, the photons not having enough energy to pull off electrons from atoms, travel freely from this time on, this time is called the decoupling time. These photons from the decoupling time constitute the CMB. In that sense, CMB offers a snapshot of the Universe at $z \approx 1000$ when it was 400,000 years old. This snapshot of the Universe supports the use of the cosmological principle corresponding to an homogeneous and isotropic Universe. In fact it only supports isotropy then we still need the copernican principle to imply homogeneity.

I.1.4 Beyond the Big Bang Model

While the Big Bang Model was well settled on the three fundamental observational pillars described above, the observations from the last two decades of the 20th century point out to the need of going beyond this picture. The Big Bang picture of the Universe implies a smooth, expanding Universe composed of baryonic matter and radiation. Nowadays, the evidence of temperature fluctuations in the CMB and the existence of large scale structure, leads to a Big Bang picture of a not exactly homogeneous Universe. As well as the need to introduce further components, as dark matter to explain the gravitational field seen in the dynamics of galaxies and required to explain the CMB anisotropies and structure formation, or the dark energy that was introduced with the discovery of the accelerated expansion of the Universe. These observations required to go beyond this model, the addition of the cold dark matter (CDM) give place to the CDM model and the subsequent inclusion of dark energy give place to the Λ CDM model. The Λ CDM model is the most accepted description of the Universe that we have today. In the following lines I summarize briefly these observational evidences.

- **Dark Matter.** In the late 1960s and early 1970s, from rotational curves of galaxies, the gravitational field was measured enabling to infer the mass. However, when comparing the result with visible matter a mismatch exist (Rubin & Ford, 1970)(Rubin et al., 1980). This finding led to the necessity of introducing a new component to account for the missing matter: the dark matter. The concept of dark matter was originally evoked by Fred Zwicky in 1937 (Zwicky, 1937) to explain the velocity dispersion of galaxies in clusters. The best evidence up to date of Cold Dark Matter (8σ significance (Clowe et al., 2006)) is provided by the Bullet cluster observation Fig. I.5. The bullet cluster consist of two colliding cluster of galaxies. The weak-lensing mass map reveals a dark matter clump lying ahead of the collisional gas bullet, but coincident with the effectively collisionless galaxies (Markevitch et al., 2004). This new component of the Universe must be non collisional, that means it neither interacts strongly nor electromagnetically. We will see that this component is fundamental to explain the CMB anisotropies and structure formation (see Fig. [I.6]).
- **Anisotropies in the CMB.** Precise measurements of the Cosmic background radiation in 1992 (Smoot et al., 1992) made possible the determination of the temperature fluctuation at the level of the one part in 10^5 . These measurements indicate that the Universe was not completely smooth. The temperature fluctuations found in the CMB imply the existence of primordial fluctuations at early times. Moreover, such primordial fluctuations are required to explain the origin of the structures we observe today in galaxy surveys. In Fig.[I.6] we see the all-sky map of temperature fluctuations seen by the COBE Satellite in 1992, with 7 degree resolution. The temperature range of the image is $\pm 100 \mu\text{K}$. In addition, the all-sky map

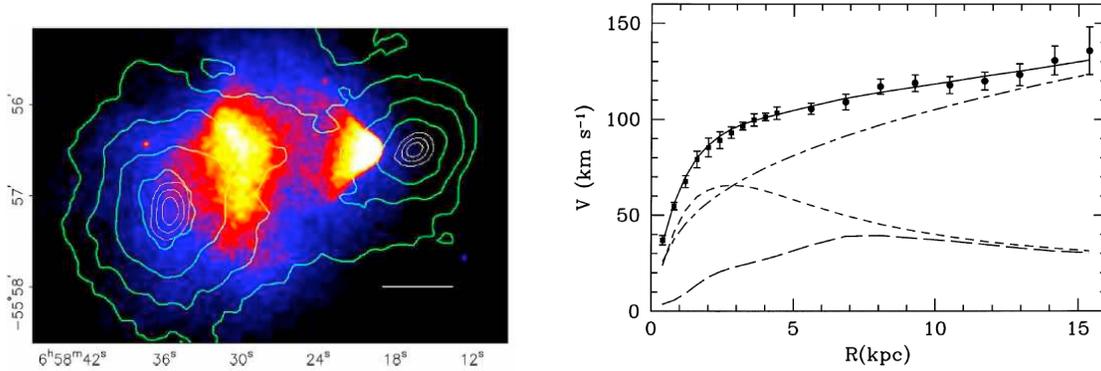


Figure I.5: Left panel, the bullet cluster (1E 0657-558) consists of two colliding clusters of galaxies at $z=0.296$. It provides the best evidence up to date of Cold Dark Matter with 8σ significance (Clowe et al., 2006). An offset of the center of total mass (by gravitational lensing) from the center of baryonic matter (Hot Gas seen in X-rays) peaks. Right panel, a recent example of a rotation curve for M33 (21 cm) compared with best fitting-model (continuous line). Also shown are the contribution from the Halo (dotted dashed line), the stellar disk contribution (short dashed line) and the gas contribution (long-dashed line) from reference (Corbelli & Salucci, 2000)

of temperature fluctuations seen by WMAP 18 years later with data from 7 years. The correspondent angular power spectrum is shown in Fig.[I.7]. Even though I will discuss in more detail the CMB in a further section, I want to underline some important informations that come from the anisotropies of the CMB 1) They provide us a way of determining the total energy density Ω_0 and thus the curvature of the Universe (the first peak in the CMB power spectrum corresponds to the angular size of horizon size at last scattering). The analyses indicate $\Omega_0 = 1 \pm 0.2$ (Spergel et al., 2003). 2) From the anisotropies, fixing the geometry, the matter density of the Universe could be inferred giving a value of $\Omega_m \sim 0.3$. That is only a third of the critical value 3) Fixing the geometry, anisotropies constrain as well the baryon density². Current CMB measurements are consistent with the nucleosynthesis baryon density estimation.

- Large Scale Structure. In the last decade of the 20th century a number of large scale surveys appeared as CfA(1985), APM Galaxy Survey (1990), IRAS(1988), SDSS (2000) and 2dF(1997). The distribution of galaxies found by these surveys was clearly non random. They indicated the existence of large scale structures as filaments, voids and clusters (see Fig. [I.8]). These observations proved that the Universe was not exactly smooth in large scales. In order to explain the structure observed in the Univers, it seemed necessary to introduce a dark matter component to the standard model, as well as to allow deviation from the smooth condition. I will discuss this subject in the section devoted to the perturbed Universe.
- Acceleration of the expansion. A similar study to Hubble's work was done in 1998 with Type Ia supernovae by Perlmutter (Perlmutter et al., 1999) and Riess (Riess et al., 1998)

²In fact, the height of the first acoustic peak rises with the baryon density.

simultaneously. The supernovas are the result of a collapse explosion mechanism³ that implies they are rare and really luminous objects. To have an idea of how rare they are, the local rate of Type Ia supernovas measured recently is about 5×10^{-5} SNe Ia Mpc.⁻³ yr⁻¹ (Li et al., 2011). A supernova is about 10 magnitudes more luminous than a regular novae (their absolute magnitude is about $M \sim -19.5$, which is comparable to the brightness of the entire host galaxy in which they appear). Thus, their large luminosity allows us to detect them even at high redshift ($z \sim 1$). The supernovas Type Ia⁴ are particularly useful because from all

³In a system consistent of two component stars; a white dwarf primary, and a mass transferring secondary. The stars are so close to each other that the gravity of the white dwarf distorts the secondary, and the white dwarf accretes matter from the companion until achieving the Chandrasekhar limit. Novae is a nuclear explosion caused by the accretion of hydrogen onto the surface of a white dwarf star, which ignites and starts nuclear fusion.

⁴Supernova classification is based on their spectral appearance near maximum light, the Type II show hydrogen in their maximum-light spectrum, while Type I lack hydrogen, Type Ib are related with Helium, etc

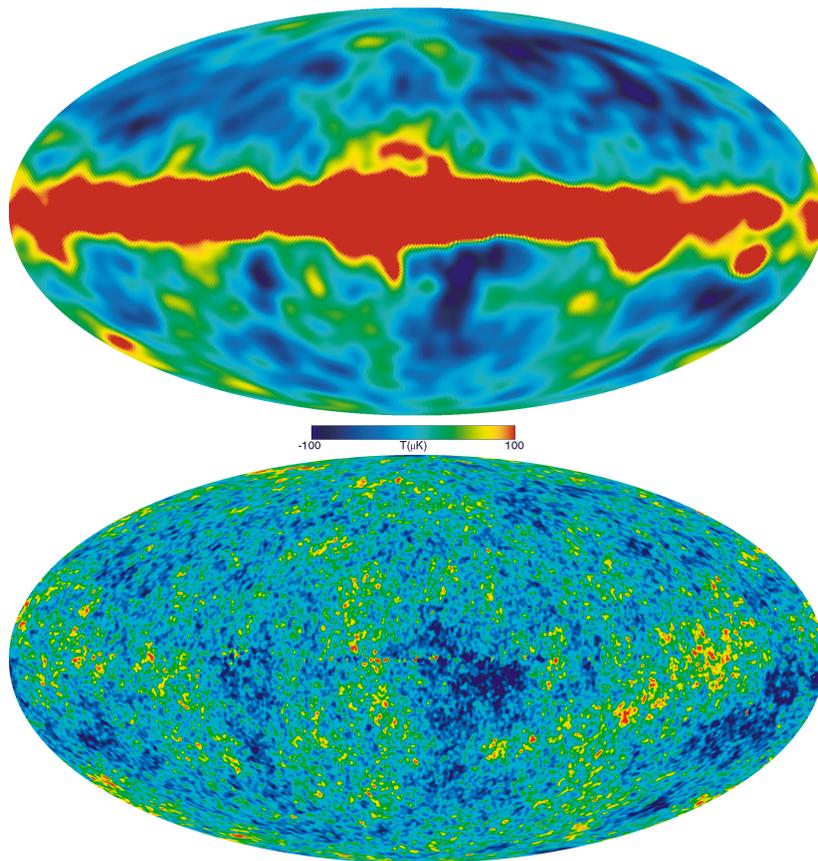


Figure I.6: Top panel. CMB map seen by COBE Satellite in 1992, with 7 degree resolution, the temperature range of the image is $\pm 100 \mu\text{K}$. Bottom panel, CMB seen 18 years later by WMAP created from 7 years of observations at 0.2 degree resolution. The signal of our galaxy was subtracted using multifrequency data. The temperature range of image is $\pm 200 \mu\text{K}$. Image from <http://map.gsfc.nasa.gov/>

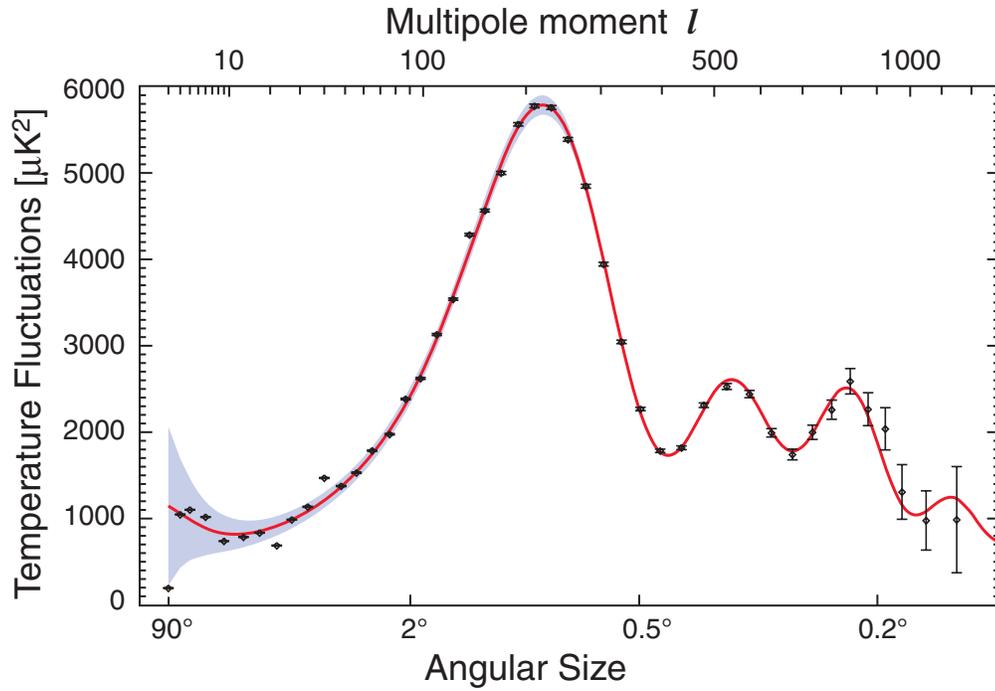
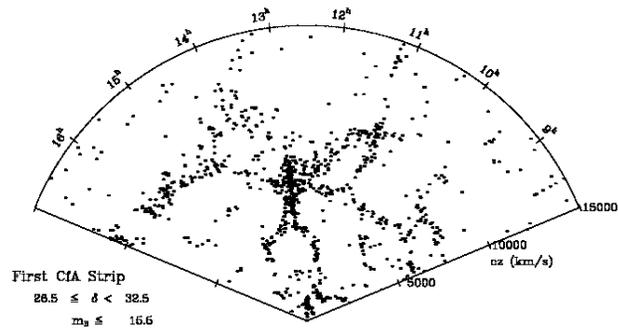


Figure I.7: The angular spectrum of the fluctuations in the WMAP full-sky map. Based on the 7 years data release. From <http://map.gsfc.nasa.gov>.



Copyright SRO 1998

Figure I.8: The first large scale survey, CfA, appears in 1985 (From <https://www.cfa.harvard.edu/~dfabricant/huchra/zcat/>).

types of supernovae, they are almost "uniform" in their appearance⁵. Thus, in principle they can be calibrated fairly reliably in order to provide accurate distances, based on their intrinsic brightness. It is important to underline that current theoretical models are not yet able to accurately reproduce all of the important features of the observed events. The supernovae measurements extended the interval of distances of previous works, by one order of magnitude. As Fig. [I.9] shows, Riess and Perlmutter found fainter luminosities, so farther objects, than predicted in a Universe dominated by matter, and thus possible an acceleration in the expansion of the Universe.

These results could be explained with other hypothesis that an acceleration of the expansion: 1) questioning the use of supernovae as standard candles, would mean that distant supernovas behave different than close supernovae, indicating a variation with time of physical process of supernovae as function of metallicity which would produce systematically a lower luminosity peak. Based on the fact that the luminosity curves of far and close supernovae are really similar it is probable that their luminosities were the same too, thus excluding this possibility. 2) Possible absorption by the interstellar medium, indicating less luminous curves 3) A possible experimental bias. Even if these alternative explanations were not completely excluded, the preferred explanation was the acceleration of expansion. To understand this choice we have to understand that the interpretation of the supernovae results arrived in a moment where there was evidence of the matter density inferior to the critical value ($\Omega_M \sim 0.4$) and results from CMB indicated a value for the total density $\Omega_0 \sim 1$.

Those measurements were in apparent conflict at that time. In this context cosmologists were more inclined to believe the SNe type Ia results, because of this preexisting evidence for a missing-energy component, which predicted an accelerated expansion. With supernova results interpreted in terms of a cosmological constant, overall measurements matched with the preferred theoretical model. In Fig. [I.9] the best-fit confidence regions in the $\Omega_M - \Omega_\Lambda$ plane are shown. These results excluded with high confidence a flat and matter dominated Universe ($\Omega_m = 1$) and even an open and matter dominated Universe ($\Omega_m = 0.3$) was strongly disfavored. In fact, using prior constraints, these results favoured a current acceleration of the expansion of the Universe, thus an open universe dominated by something else, a cosmological constant or dark energy ($\Omega_\Lambda > 0$)⁶.

A Universe in accelerated expansion was beyond the Big Bang model that led to the reintroduction to the Einstein equation of the famous blunder of Einstein, the cosmological constant. The search for an explanation for the acceleration of the expansion gave rise to the emergence of new physics, possibly a new density component, the dark energy, that would represent 73% of the total energy density, or in other perspective, modifications to our gravity theory, or the abandon of the cosmological principle. The preferred theoretical framework to describe acceleration of expansion is to interpret this new term in the Einstein equation as a new component in the energy density. This new component, the Dark Energy, very recently ($z \sim 0.4$) begins to dominate the density of Universe and it exerts a repulsive force accelerating

⁵In fact supernovas shows some spread in its luminosity and spectral evolution. But the majority are quite homogeneous, few are true outliers (70% in a volume-limited sample and %77 in a flux limited-sample)(Goobar & Leibundgut, 2011).

⁶In the words of Riess *The distances of the high-redshift SNe Ia are, on average, 10%-15% farther than expected in a low mass density ($\Omega_M = 0.2$) universe without a cosmological constant... prior constraints unanimously favor eternally expanding models with positive cosmological constant (i.e., $\Omega_\Lambda > 0$) and a current acceleration of the expansion...*

the expansion of the Universe. The nature of this Dark energy is considered one of the most important open questions in modern cosmology. In a further section I will discuss in more detail about the current understanding and the trails.

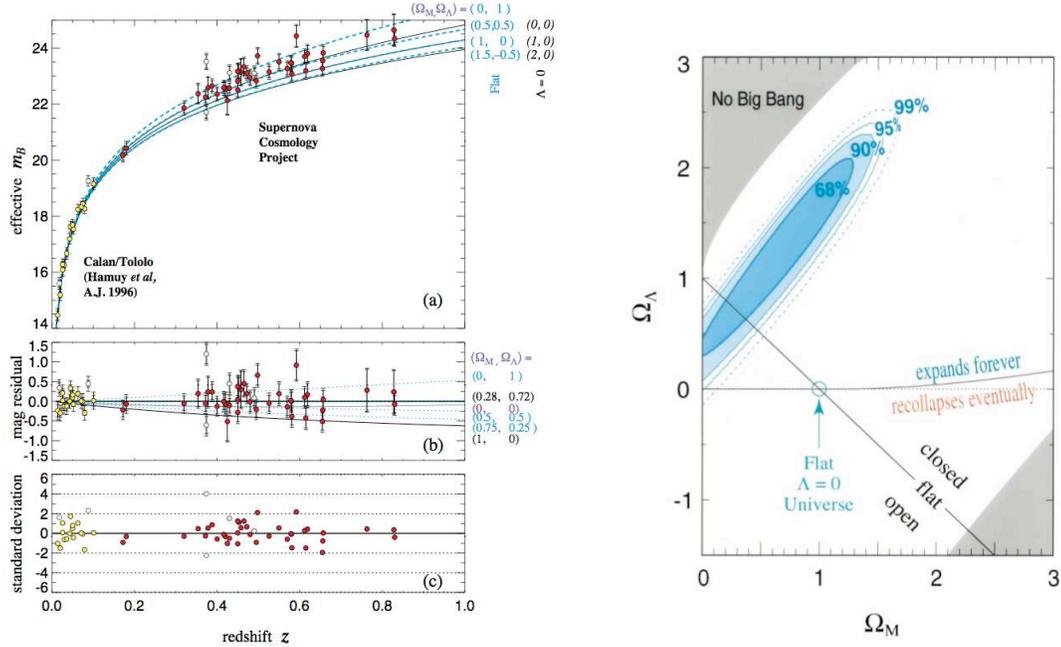


Figure I.9: Original plots from (Perlmutter et al., 1999) Left panel. Hubble diagram for 42 high-redshift Type Ia supernovae from the Supernova Cosmology Project, and 18 low-redshift Type Ia supernovae from the Calan/Tololo Supernova Survey, plotted on a linear redshift scale to display details at high redshift. Right panel. Best-fit confidence regions in the $\Omega_M - \Omega_\Lambda$ plane. The SN results exclude with high confidence a flat and matter dominated Universe ($\Omega_m = 1$) and even an open and matter dominated Universe ($\Omega_m = 0.3$) was strongly disfavored. In fact, using prior constraints, this results favored an open universe dominated by something else, a cosmological constant or dark energy ($\Omega_\Lambda > 0$)

I.2 Homogeneous Universe

This section is devoted to the theoretical framework that allows us to study an homogeneous and isotropic Universe. The theoretical framework used to describe the evolution of the Universe is done by the integration of cosmological principle in General Relativity.

I.2.1 General Relativity

The General Relativity allows to study a Universe with non trivial geometry (non Euclidean), evolutive and dependent of the energy content of the Universe. General relativity is based on the

equivalence principle. The equivalence principle states that all laws in physics take the same form in freely falling frame as they would in the absence of gravity. In this sense, the equivalence principle guaranties that we can always find a reference system in which the gravitational force is erased (locally), but gravitation do not disappears. It is hidden in a new property of the space-time: curvature. In this frame, the objects in the absence of non gravitational forces, follow the minimum action in the curved space-time called geodesics. The innovative feature of general relativity in relation with the Newtonian framework, is that gravity is a manifestation of the curvature of space-time.

The geometry of space time is described by its metric. It provides the connection between values of coordinates and physical distance ds^2 . The invariant element of distance give us the variation between two events in space time

$$ds^2 = \sum_{\mu, \nu=0}^3 g_{\mu\nu} dx^\mu dx^\nu \quad (\text{I.4})$$

where $g_{\mu\nu}$ is the metric, and dx^μ is a differential space-time coordinate interval. By now, I have enunciated the first aspect of general relativity that permits to relate gravity to the metric. The second aspect of general relativity is that it relates the geometrical properties of the space-time with the energy content. This element is contained in the Einstein Equation. Let us begin describing the metric that we will use to describe the Universe and then discuss to the relation between these geometrical properties and the energy content of the Universe.

I.2.2 Lemaître-Friedman-Robertson Walker Metric

It seems clear that in order to study the Universe, we must define an adapted metric that describes the geometrical properties of the Universe. This lead us to a basic principle of cosmology itself. Cosmology is based on the believe that we do not occupy a special place in the Universe, this principle is known as the cosmological principle or Copernican principle. From this principle it follows that, since the Universe appears isotropic around us, it should be isotropic around every point. Moreover, isotropy around every point implies homogeneity. Thus, we arrive to the conclusion that our Universe must be homogeneous and isotropic. At this point it is important to clarify the sense of these terms. Isotropy is the claim that the Universe looks the same in all direction (invariant under rotations). Homogeneity is the claim that the Universe looks the same at every point (invariant under translations). Then, we have to precise that the sense of isotropy and homogeneity is statistical. I have already said we have evidences about non homogeneities. To conciliate these evidences with the cosmological principle we have to think in our Universe as a realization of a set of possible representations. Thus, the sense of isotropy and homogeneity acts over the ensemble average of energy density, which is the same in every point of the space at a fixed time.

The most general space time metric consistent with homogeneity and isotropy in an expanding Universe is the Robertson-Walker metric, that could be defined in terms of the invariant element of distance ds^2 :

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \right] \quad (\text{I.5})$$

where t is time measured for a comoving observer with constant spatial coordinates, a(t) is the scale

factor and k denotes the curvature of space-time. The curvature is an intrinsic geometrical property of the space-time.

$$k = \begin{cases} +1 & \text{close} \\ 0 & \text{flat} \\ -1 & \text{open} \end{cases} \quad (\text{I.6})$$

Thus, three possible geometries for the Universe are compatible with this theoretical framework in a trivial topology: flat, spherical and hyperbolic, corresponding to a close, flat and open Universes (Fig. Geometries). The flat geometry corresponds with $k=0$, is the easiest and also the more familiar. It is described by Euclidean geometry. That is the parallel lines remain at a constant distance and the sum of angles is equal to 180° . This geometry implies an infinite universe. The spherical geometry corresponds to $k < 0$. It is described by Riemmanian geometry. This geometry is usually represented by spherical surface to understand the basic features. In this geometry, parallel lines will converge, and sum of angles is greater than 180° . In a spherical universe an object travelling in straight line eventually will return where it started, as in the 2D analogue in a spherical surface. This geometry implies a finite Universe without boundary. Finally, the hyperbolic geometry corresponds to a $k < 0$, is the more unfamiliar. It is represented as a saddle-like surface. It is characterized by the fact that parallel lines diverge away from each other and the sum of angles is less than 180° . The hyperbolic geometry implies an infinite Universe.

The metric tensor corresponding to the metric of Equation I.5 in Cartesian coordinates for a flat Universe is (with $c=1$):

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix} \quad (\text{I.7})$$

I.2.3 Einstein's Equation

As described above, General Relativity not only allow us to relate the gravity with the geometry but also allows us to relate the geometrical properties of the space-time with the energetic content. This is contained in the Einstein's equation:

$$G_{\mu\nu} \equiv R_{\mu\nu} + \frac{1}{2}Rg_{\mu\nu} = 8\pi GT_{\mu\nu} \quad (\text{I.8})$$

Einstein's equation is defined in terms of Einstein's tensor $G_{\mu\nu}$ describing the geometry and the energy momentum tensor $T_{\mu\nu}$ describing the energy content. Einstein's tensor is written in terms of $g_{\mu\nu}$ the metric tensor, $R_{\mu\nu}$ the Ricci Tensor that depends on the metric and its derivatives⁷, R the Ricci scalar that is the contraction of the Ricci tensor with the metric tensor $R \equiv g^{\mu\nu}R_{\mu\nu}$ and G the Newton constant. As we can observe, the left side of the equation is a function of the metric, and the

⁷The Ricci tensor is expressed in terms of the Christoffel symbol:

$$R_{\mu\nu} = \Gamma_{\mu\nu,\alpha}^\alpha - \Gamma_{\mu\alpha,\nu}^\alpha + \Gamma_{\beta\nu}^\alpha \Gamma_{\mu\nu}^\beta - \Gamma_{\beta\nu}^\alpha \Gamma_{\mu\alpha}^\beta \quad (\text{I.9})$$

where $\Gamma_{\mu\nu,\alpha}^\alpha \equiv \partial\Gamma_{\mu\nu}^\alpha/\partial x^\alpha$. and $\Gamma_{\alpha\beta}^\mu = \frac{g_{\mu\nu}}{2} \left[\frac{\partial g_{\alpha,\nu}}{\partial x^\beta} + \frac{\partial g_{\beta,\nu}}{\partial x^\alpha} - \frac{\partial g_{\alpha,\beta}}{\partial x^\nu} \right]$

right side function of the energy, so the Einstein's equation relates geometry of the Universe with energy content. At the time when Einstein proposed this equation, the common believe between cosmologist was a static Universe. This equation does not have a stable static solution. This conducted initially Einstein to add the famous cosmological constant Λ for philosophical reasons enabling the possibility of a static Universe⁸. The cosmological constant provided a way of balancing the gravitational contraction caused by matter.

With Hubble's evidence of a non static Universe, this constant disappeared and was reintroduced recently for a completely different reason: the evidence of the acceleration of expansion. In fact, the general form of Einstein's equation is the following:

$$R_{\mu\nu} + \frac{1}{2}Rg_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu} \quad (\text{I.10})$$

It is time now to discuss the left side of the equation. The Energy-Momentum tensor is a symmetric tensor describing the constituents of the Universe. For the energy-momentum tensor $T_{\mu\nu}$, it is assumed a perfect fluid form, thus it is simply expressed in term of the density ρ and pressure p :

$$T_{\mu\nu} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \quad (\text{I.11})$$

To get the temporal evolution of the components of the energy-momentum tensor, we translate the usual mass and momentum conservation to a most general criterium of conservation of energy momentum tensor, in terms of the covariant derivative, it is expressed by:

$$\nabla_{\mu}T^{\mu\nu} = 0 \quad (\text{I.12})$$

Applying this to the LFRW metric and the perfect fluid energy momentum tensor yields:

$$\dot{\rho} + 3H(\rho + p) = 0 \quad (\text{I.13})$$

This equation describes how the expansion of the Universe (as specified by H) leads to local changes in the energy density. When applied to the different components of the Universe, we get how the expansion scales the different components of the Universe.

As I mentioned before, from the acceleration of the expansion follows the reintroduction of the cosmological constant. Adding a cosmological constant to the Einstein equation is equivalent to add a component of the energy density in the Universe, thus implying to include an energy-momentum tensor of the form:

$$T_{\mu\nu} = \frac{\Lambda}{8\pi G}g_{\mu\nu} \quad (\text{I.14})$$

This simple act of placing the Λ term in the left or right side of the equation involve deep conceptual differences. Written in the left side must be interpreted as a geometrical property. It represents thus the curvature of the empty space. In the right side, the interpretation is in energetic terms. It

⁸In fact add a constant term is the simpler modification that conserves the null divergence

represents the energy momentum tensor of a component of the Universe, possibly the vacuum or a new one. This new component in the energy-momentum tensor is interpreted as a fluid characterized by its density and pressure. The simplest case is considering a cosmological constant. In terms of the equation of state of a fluid $p = w\rho$, a cosmological constant (with $\Lambda > 0$) corresponds to $w = -1$. This particular case gives us constant density and negative pressure as the properties of this unknown fluid⁹ given by the following expressions:

$$\begin{aligned}\rho_\Lambda &= \frac{\Lambda}{8\pi G} \\ p_\Lambda &= -\rho_\Lambda\end{aligned}\tag{I.15}$$

As we will see later the Friedman equations, when we specify the equation of state of a cosmological constant, enables us to interpret the effect of a negative pressure as a repulsive gravity, generating the acceleration of the expansion.

The supernovae data, fixing the flat geometry, favoured a cosmological constant (see Fig. [??](Perlmutter et al., 1999)) but it is not the only possibility. In general, a fluid with an equation of state with $w < -1/3$ generates an acceleration of expansion. Thus, in order to consider other possibilities than a cosmological constant, we get the general expression for the evolution of the density integrating eq. (I.13) and using the general expression for the state equation,

$$\rho_{DE} \propto \left\{ 3 \int^a \frac{da'}{a} [1 + w(a')] \right\}\tag{I.16}$$

This general case is known as dark energy. In what follows I consider this choice to present the theoretical framework. It is common in literature to refer as dark energy to the energy component responsible for the acceleration of the expansion in a general way, thus including the particular case of a cosmological constant. For this text I will separate the two cases for clarity.

I.2.4 Friedman equations

At this moment we have the tools to develop the equations that describe the dynamics of the Universe. The dynamics are described by the scale parameter. To understand the history of the Universe, we need to determine the evolution of the scale factor with time. General Relativity provides the connection between this evolution and the energy content of the Universe.

The Friedman equations come from applying the Einstein equation to the Lemaître-Friedman-Robertson-Walker (LFRW) metric. They describe the evolution of the scale parameter in an homogeneous Universe. In fact from Equations I.7 and I.10 only two equations survive. The first Friedman equation comes from the time-time component of Einstein equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \sum_i \rho_i - \frac{k}{a^2}\tag{I.17}$$

the overdot denotes a derivative with respect to the cosmic time t , and i indexes all possible gravitational sources in the Universe. The Friedmann equation relates the rate of increase of the scale

⁹Unknown in terms of the Standard Model of Particle Physics

factor, as encoded by the Hubble parameter, to the total energy density of all matter in the Universe. It is convenient to define critical density δ_c at any given time for which $k=0$

$$\rho_c \equiv \frac{3H^2}{8\pi G} \quad (\text{I.18})$$

The current value of $\rho_{c,0} = 1.88 \times 10^{-29} h^2 \cdot g \cdot cm^{-3}$. This allows to relate the total energy density of the Universe to its local geometry $\Omega \equiv \frac{\rho}{\rho_{crit}}$:

$$\begin{aligned} \Omega > 1 &\iff k = +1 \\ \Omega = 1 &\iff k = 0 \\ \Omega < 1 &\iff k = -1 \end{aligned} \quad (\text{I.19})$$

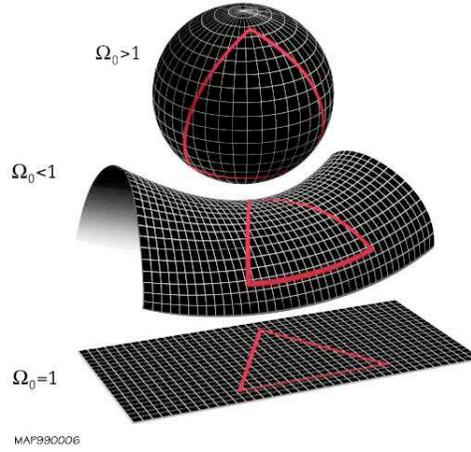


Figure I.10: The total fractional density of the Universe is related to its local geometry. A value $\Omega_0 > 1$ indicates a spherical geometry, a value $\Omega_0 < 0$ would indicate an hyperbolic geometry and $\Omega_0 = 1$ corresponds to a flat universe. Images from <http://map.gsfc.nasa.gov>

In fact, the usual way to express the first Friedman equation is in terms of the fractional densities of each component of the critical density $\Omega_i \equiv \frac{\rho_i}{\rho_{crit}}$ ¹⁰. That is:

$$H^2 = H_0^2 \left(\frac{\rho}{\rho_{c,0}} + \frac{k}{a^2 H_0^2} + \frac{\Lambda}{3H_0^2} \right) = H_0^2 (\Omega_m + \Omega_k + \Omega_\Lambda) \quad (\text{I.21})$$

In this expression Ω_k accounts for a nonzero curvature, it is a convenient way to keep track of how much energy density is lacking in comparison to a flat Universe.

The second Friedman equation is:

$$\frac{\ddot{a}}{a} + \frac{1}{2} \left(\frac{\dot{a}}{a} \right)^2 = -4\pi G \sum_i p_i - \frac{k}{2a^2} \quad (\text{I.22})$$

¹⁰Evaluating a current time, we find that the geometry of the universe (Ω_k) is a function of Ω_m and Ω_Λ , that is :

$$\Omega_m + \Omega_k + \Omega_\Lambda = 1 \quad (\text{I.20})$$

It is useful to re-express this equation combining with the first equation to get the acceleration equation:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \sum_i (\rho_i + 3p_i) \quad (\text{I.23})$$

A few comments about this equation: first, this equation shows us that fluids with pressure increases the gravitational force decreasing the acceleration. Second, this equation is, as we have said before when we specify the equation state of dark energy (or particularly a cosmological constant), enables us to interpret the effect of dark energy as a repulsive gravity, generating the acceleration of expansion $\ddot{a} > 0$.

I.2.5 Contents of the Universe

One final piece of information is required to solve the equations: how the pressure and energy density are related to each other. This question leads to the discussion about the nature of the different components of the Universe. At this point we will consider three possibilities: matter, radiation and dark energy (considering a cosmological constant as a particular case in the dark energy framework). By matter we mean non relativistic matter, in fact it refers to any type of particles that exert negligible pressure $p \sim 0$. The second type, radiation, includes all relativistic particles ($v \sim c$) notably photons and neutrinos will be included. Their radiation pressure is given by $p = \rho c^2/3$. To deal with the different components of the Universe, we must specify the relationship between the energy density and the pressure. Within the fluid approximation used here, we may assume that the pressure is a single-valued function of the energy density $p = p(\rho)$. It is often convenient to define the equation of state in terms of the state parameter, w , by

$$p = w\rho \quad (\text{I.24})$$

This expression for the equation of state describes the different components of the Universe, setting a value for the equation of state parameter w , $w = 0$ corresponds to massive non relativistic particles. Similarly, $w = 1/3$ corresponds to relativistic particles as photons or other highly relativistic species, and for the dark energy we only have an upper constraint $w < -1/3$, the particular case $w = -1$ corresponds to a cosmological constant.

Using this parametrization of the equation of state I.24 is possible to solve the evolution equation for the density of the different components of the Universe I.13 giving the following expressions for the evolution of densities as function of scale parameter for a flat universe ($k=0$):

$$\rho(a) \propto \frac{1}{a(t)^{3(1+w)}} = \begin{cases} a^{-3} & \text{matter} \\ a^{-4} & \text{radiation} \\ cte & \text{cosmological constant} \\ 3 \int^a \frac{1+w(a)}{a} da & \text{dark energy} \end{cases} \quad (\text{I.25})$$

Let us now explain these results in a more heuristic manner. For the case of non relativistic matter, we know that energy equals to rest mass energy times number density of matter. The rest mass energy is constant in time and the number density evolves with expansion inversely to the volume, thus $\propto a^{-3}$, that means energy density of matter would evolves as $\propto a^{-3}$. In the case

of relativistic matter, photons, the expansion of the Universe results in a redshift of the photons. Because of the redshift, the energy density in a fixed number of photons in a fixed comoving volume, drops with the physical volume ($\propto a^{-3}$) and by an extra factor of the scale factor as the expansion of the Universe stretches the wavelengths of light. Thus, the energy density of radiation will scale as a^{-4} .

In the case of dark energy, the scenario is more uncertain. The evidence from supernovae suggest that this new form of energy is compatible with a cosmological constant. It would indicate that the energy density remains constant with time, but as we have said it is not the only possibility, it could be more complex. At the present time, a lot of effort is being invested trying to discern this puzzle. Many current and future experiments envisage, by different methods, to understand the nature of dark energy. The current approach is to try to constraint the equation of state of dark energy, looking for a time dependence or a verification of a cosmological constant (parametrization Chevalier-Palarski-Linder CPL).

$$w_{\Lambda} = w_0 + w_a \left(\frac{z}{1+z} \right) \quad (\text{I.26})$$

In Fig. I.11 the evolution of the energy density as function of the scale factor for different constituents is shown: radiation, matter and cosmological constant. We observe how the different constituents have dominated at different times, today it seems that cosmological constant dominates since very recently, before it was the matter and at early times the radiation.

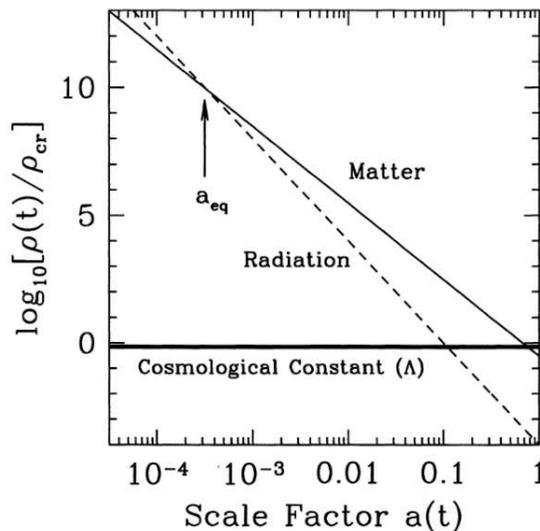


Figure I.11: Evolution of the energy density as function of the scale factor for different constituents is shown: radiation, matter and cosmological constant. We observe how the different constituents have dominated at different times, today it seems that cosmological constant dominates since very recently, before was the matter and at early times radiation. Figure from (Dodelson, 2003)

From the equations of state of the different components of the Universe and the Friedman equation I.17, the evolution of the scale factor as a function of time considering only the dominant

fluid is given by the following expressions:

$$a(t) = \left(\frac{t}{t_0}\right)^{\frac{2}{3(1+w)}} = \begin{cases} a(t) = \left(\frac{t}{t_0}\right)^{2/3} & \text{matter dominated} \\ a(t) = \left(\frac{t}{t_0}\right)^{1/2} & \text{radiation dominated} \\ a(t) = \exp(Ht) & \text{cosmological constant dominated} \\ a(t) = \exp\left(3 \int \frac{1+w(a)}{a} da\right) & \text{dark energy dominated} \end{cases} \quad (\text{I.27})$$

These equations express how the scale factor varies with time, and this variation is determined by the energy density. At early times $a \propto t^{1/2}$ when the Universe was radiation dominated, at later times switches to $a \propto t^{2/3}$ during the matter dominated era. As we will see in the next section, this transition is very important from the point of view of structure formation. Notice that the Universe expands more slowly in a radiation dominated than in a matter dominated Universe. This is a consequence of the deceleration that pressure applies (I.12). At this point we must be convinced that a way to explore the energy content of the Universe is to measure changes in the scale factor. Very recently, the scale factor has stopped growing as $t^{2/3}$, a signal that a new energy has come to dominate the cosmological constant or dark energy.

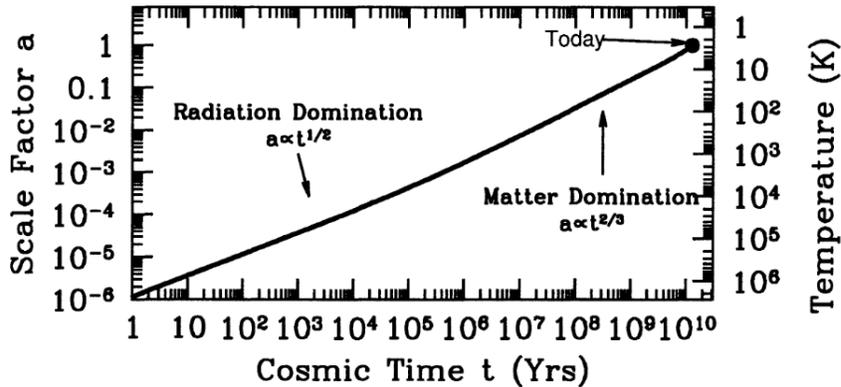


Figure I.12: The scale factor varies with time and this variation is determined by the energy density. At early times $a \propto t^{1/2}$ when the Universe was radiation dominated, at later times switches to $a \propto t^{2/3}$ when matter dominated era. Very recently, the scale factor has stopped growing as $t^{2/3}$. Figure from (Dodelson, 2003)

Before going to the next section let us summarize which are the specific components of the Universe:

- **Baryons.** It refers to the ordinary matter. In particle physics the baryons are the particles composed of three quarks. Of the jungle of particles the only stable baryons are neutrons and protons. Even if electrons are not baryons, in cosmology the use of baryonic matter traditionally includes electrons. In fact, nuclei vastly outweigh the electrons. They are considered as non relativistic particles. There are four established ways of measuring Ω_b : 1)

From nucleosynthesis calculations $\Omega_b h^2 = (0.0214 \pm .002)$ (Kirkman et al., 2003). 2) The luminous matter in the Universe, consisting of stars and radiation-emitting gas, accounts for only $\Omega_b h^2 \sim 0.02$. 3) From the quasar spectra, the amount of light absorption by the interstellar gas gives a measurement of hydrogen, thus gives an estimate of the baryon density $\Omega_b h^2 > 0.021$ (Rauch et al., 1997). 4) From CMB anisotropies, as we have already said, giving a value of $\Omega_b h^2 \simeq 0.02258_{-0.00056}^{+0.00057}$ (Larson et al., 2011)

- Dark matter. From the theoretically favoured value of $\Omega_0 = 1$, that agrees with the observed value and from fractional baryonic density estimations, it follows that a non baryonic matter component must account for the lacking matter: dark matter. We know few things about the dark matter: it must be cold, that is, non-relativistic, non baryonic and should interact very weakly with ordinary matter to explain why it has not been detected until now. Historically, the way of measure this density component was from the light-to-mass ratio estimated from the gravitational field of certain systems, providing a value of $\Omega_m \sim 0.3$ (Bahcall et al., 2000). A second method comes from the power spectrum of matter, where the inferred value is $\Omega_m \sim 0.3$ (Percival et al., 2002) or from the bispectrum (Verde et al., 2002). A third method is provided by the velocity field (Strauss & Willick, 1995). As we have said before, from the CMB anisotropies we get an estimate $\Omega_m h^2 \sim 0.16 \pm 0.004$ (Netterfield et al., 2002). There are other ways of measuring the total mass density taking for example observations sensitive to the Ω_b/Ω_m , like measurements of X-ray emission or features in power spectrum. All of the measurements indicated a value around $\Omega_m \sim 0.3$
- Photons. The photon density is dominated by CMB photons. The energy density associated to CMB photons is measured via temperature, it gives a value $\Omega_\gamma \equiv \rho_\gamma/\rho_{crit} = 2.47 \times 10^{-5}/h^2$. These particles are modelled as a relativistic fluid.
- Neutrinos. Are extremely weakly interacting particles. They are modelled as a relativistic fluid as photons. The contribution to the energy density of neutrinos considering three species of massless neutrinos today would be $\Omega_{\nu,0} = \rho_\nu/\rho_c = 1.68 \times 10^{-5}/h^2$. In reality neutrinos are not massless, thus this consideration leads to another expression that is function of the neutrinos mass.
- Dark energy. Let us summarize what has been said until now about dark energy: There are two sets of arguments pointing toward the existence of a new energetic component in the Universe. The first one, we observe a total density of the Universe close to the critical value and our account from all the components is only 1/3 of it. Second argument, from supernovae we infer an acceleration of the expansion, indicating that the energy density of the universe now is dominated by another energetic component. From supernovas, the constraints on the fractional densities using prior constraints are $\Omega_m \sim 0.3$ and $\Omega_\Lambda \sim 0.7$. The current constraints combining different observations (SN+BAO+CMB) are $\Omega_m = 0.278 \pm 0.015$ and $\Omega_\Lambda = 0.722 \pm 0.015$

I.2.6 Distances

In the last sections the basic framework for describing an evolving Universe has been presented, we have now the tools to discuss about the fundamental concept of distances. Measuring distances

in an expanding Universe is not a trivial exercise, because distances between comoving objects are constantly changing. Furthermore in FR Cosmology there is no a unique definition of distance. As we will see, there are three distances that provide us different ways for testing the cosmological model. The fundamental distance measurement is the comoving distance between two objects (two points of the grid following expansion). To define the comoving distance we will use the photon trajectory. Let us suppose that an object emits a signal that arrive to us. To compute the trajectory we use the equation of the invariant interval eq.[I.5]; the photons not having mass will follow the geodesic, hence $ds^2 = 0$. The photons will follow radial trajectories ($d\phi = 0$ and $d\theta = 0$), then considering a flat Universe give us $dr^2 = \frac{dt}{a(t)}$. Finally, to get comoving distance we must integrate between the emission time and now, giving the following expression:

$$\chi = c \int_{t_{\text{emis}}}^{t_0} \frac{dt}{a(t)} = \int_{a(t_{\text{emis}})}^{a(t_0)=1} \frac{da}{a^2 H(a)} = \int_0^{z_{\text{emis}}} \frac{dz}{H(z)} \quad (\text{I.28})$$

The last equation includes the expressions in terms of scale parameter and the redshift that follows from simple variable changes. The dependency on cosmological parameters is enclosed in Hubble parameter or scale factor¹¹. As we see, in a evolving Universe the notion of distance is complex, within the time the photons were emitted and we receive the signal, the universe have expanded. The rate of expansion is determined by the total energy density, thus, it is continuously evolving too, as a consequence the distances provides a direct probe for the cosmological model. The comoving distance is not an observable, we have to define others distances.

Angular Diameter Distance

A second definition of distance comes from the measure of the angle subtended by an object θ of known physical size l . The distance d_A to that object in a non expanding Universe, assuming the angle subtended is small ($\sin\theta \sim \theta$) is given by:

$$d_A = \frac{l}{\theta}. \quad (\text{I.30})$$

d_A is known as the angular diameter distance. To explain the notion of angular diameter distance in a expanding Universe let us see Fig. [I.13]. We represent an object by two points at a radial coordinate χ_1 separated by a distance l that corresponds to the physical size of the object at t_1 . To get the expression for l we have to use the metric at the time the light was emitted, that is the expression of the invariant element of distance ds^2 eq.[I.5] giving the following expression for l :

$$l = a(t_1)\chi_1\Delta\theta. \quad (\text{I.31})$$

For convenience, we place the observing time at present t_0, χ_0 , the photons will propagate radially, preserving the angle. Thus, the angular size we perceive is $\Delta\theta = \frac{l}{a(t_1)\chi_1}$, which give us the final

¹¹This result consider only the case of a flat geometry, adding the curvature density as $\Omega_k = 1 - \Omega_0$ we can explore other kind of geometries as spherical and hyperbolic. The expressions used in this chapter could be generalized to others geometries replacing χ by the correspondent term of the geometry accordingly the following table:

$$\begin{aligned} k > 0 & \quad \sin(\sqrt{-\Omega_k}H_0\chi/(H_0\sqrt{|\Omega_k|}) \\ k < 0 & \quad \sinh(\sqrt{\Omega_k}H_0\chi/(H_0\sqrt{|\Omega_k|}) \end{aligned} \quad (\text{I.29})$$

expression for the diameter angular distance:

$$d_A = a(t_1)\chi = \frac{\chi}{1+z}. \quad (\text{I.32})$$

The angular diameter distance is a measure of how large objects appear to be. As we observe in Fig.[I.14] the angular diameter distance matches with physical distance for nearby objects, but for farther objects the angular diameter distance appears to be smaller. Furthermore, as $z \rightarrow \infty$, $d_A \rightarrow 0$ meaning that distant objects appears to be nearby.

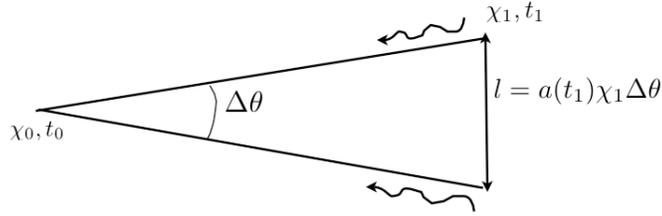


Figure I.13: Photons emitted at t_1, χ_1 by an object of size $l = a(t_1)\chi_1\Delta\theta$. The photons follows trajectories of constant (θ, ϕ) and they would be observed at (t_0, χ_0) with and angular separation given by $\Delta\theta = l/a(t_1)\chi_1$

Luminous distance

The second definition of distance in FR cosmology consist in using the flux from an object of known luminosity. The observed flux at F a distance d from a source of known luminosity L in an non expanding Universe is given by :

$$F = \frac{L}{4\pi d^2} \quad (\text{I.33})$$

To generalize the result to a expanding Universe, we notice that the luminosity is given by the number of photons passing through a comoving spherical shell (radius= χ) per unit of time multiplied by their energy ($E = \hbar\nu = \frac{\hbar c}{\lambda}$). The expansion affects the luminosity in two ways, first, the photons emitted a λ_{emit} will be redshifted ($\lambda_{obs} = \lambda_{emit}/a(t)$) thus their energies will be lower by a factor a . The second effect is in the number of photons crossing the shell in a unit of time. Because of the expansion, at earlier times the physical distances traveled by photons are smaller that at later times. Thus the number of photons crossing the shell by unit of time at emission will be greater than at later times by a factor a . The total effect is that observed luminosity will be smaller by a factor a^2 . Therefore we get the following expression for the flux:

$$F = \frac{La^2(t_1)}{4\pi\chi^2(a)} \quad (\text{I.34})$$

The luminosity distance is defined so that the flux remains defined as eq.[I.33], hence the following expression for d_L :

$$d_L \equiv \frac{\chi}{a(t_1)} = \chi(1+z) \quad (\text{I.35})$$

The luminosity distance is a way of expressing the quantity of light received from a source. As we observe in Fig.[I.14], for closer objects, the physical distance coincide with luminosity distance as the angular diameter distance does. For farther objects, as a consequence of the expansion they will appear farther away than they physical distance indicates.

Finally, the three distances are plotted in Fig [I.14] as a function of z for a flat Universe for two different cosmological models. The dashed lines corresponds to a universe with only matter and the solid lines for a model with a cosmological constant Λ and matter. In a Λ -dominated Universe, at high redshifts distances are larger than in a matter dominated one. As we have seen the distances are function of the cosmological parameters and it is precisely this feature that enable us to study the expansion of Universe using standard patrons as we will see in a further chapter.

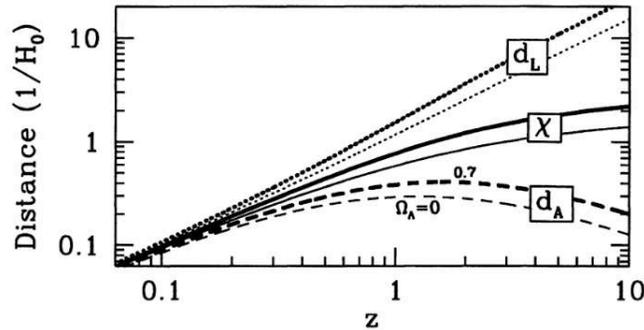


Figure I.14: The luminosity distance, comoving distance and angular distance for two cosmological models. The solid lines corresponds to a Λ -dominated Universe and the light curves to a flat matter dominated Universe. As we observed the distances are function of the cosmological parameters and it is precisely this feature that enable us to study the expansion of Universe using standard patrons. Figure from (Dodelson, 2003)

I.2.7 Brief Thermal History of the Universe

Within the framework of the Standard Model of cosmology that has been presented in the previous sections, the Universe is a dynamic entity that have passed for subsequent epochs. Let us gather all the elements together to reconstruct the thermal history of the Universe in the framework of the Λ CDM model.

In Table.[I.1] I show a time line with the principal events in the history of the Universe. I separate the table in 2 blocks. The first block refers to the first 3 minutes of our Universe. This part is beyond the purposes of this section, even thought, to have a complete time-line I will give some comments about this period. The first three minutes of Universe are intimately related to particle physics. In this short period the fundamental particles were generated. The physics that governs this era is dictated by theoretically well settled model known as the Standard Model of Particle Physics. This model, based on the Quantum Field Theory, enables us to describe the fundamental particles and understand their interactions but still has many open questions. At this moment, the range of energy explored in accelerators is of the order of 7TeV (CERN). This value is below the energies achieved in the early Universe thus this range of energies lacks for direct

validation. Even though within the theoretical framework of Standard Model of Particle Physics there is a belief that some events must have taken place. Actually, it is in this range of energies where cosmology could contribute to the understanding of the Standard Model of Particles. In the following, I will enumerate rapidly the events that mark the first instants of the Universe.

Table I.1: Timeline of the Universe

The first 3 minutes				
Time (s)		Temperature (K)	Energy	Epoch
$10^{-36} - 10^{-33}$		10^{32}	10^{16} GeV	Inflation
10^{-35}		10^{28}	10^{19} GeV	Grand Unification Transition
10^{-10}		10^{15}	10^3 GeV	Electroweak phase transition
10^{-6}		10^{12}	1 GeV	quark-hadron phase transition.
After the first 3 minutes				
Time (s)	redshift	Temperature (K)	Energy	Epoch
3min	10^{16}	10^{10}	1 MeV	Nucleosynthesis
10^4 yrs	3100	10^4	1 eV	Radiation-Matter Equality
10^6 yrs	1100	10^3	0.3 eV	Recombination and Decoupling
10^9 yrs	~ 10	10	$\sim 10^{-3}$ eV	Galaxy formation

As we have noticed, within the framework of the Standard model the notions of time and temperature are intimately related, being $t \propto 1/T$, thus, to relate the history of the Universe, it becomes equivalent to use the time or the temperature. Let us use the time for summarizing the first 3 minutes. As we have said, we situate the "beginning" of the Universe at the Planck time, that sets the limits of validity of gravity theory. At this time it is believed that happens a symmetry breaking splitting the GUT forces (Grand Unification Theory of non gravitational forces) and gravity. At a time of 10^{-36} s (10^{32} K.), the Universe is supposed to follow a period of exponential expansion known as Inflation that I will introduce in the next section. The Inflation, as we will see, is necessary to solve some of the problems of the Big Bang Theory. At the end of inflation period, we suppose the particles were created from the disintegration of the inflaton. Then, the beginning of the radiation dominated era is set, it is believed that at this early times the energy density of relativistic particles dominates the energy density of the Universe. In this standard picture as time passes, the Universe continues expanding, and cooling. At a time $\sim 10^{-35}$ (10^{28} K.), it is predicted that the Universe underwent a phase transition, called GUT transition where GUT forces split into the strong and electroweak. At this time, it begins the quark era, the image of the Universe we have is that it consisted of a plasma of quarks and gluons, photons, electrons and neutrinos that were interacting. Later on, at a time 10^{-10} (10^{15} K.) we predict the Electroweak phase transition has occurred. It corresponds to the split of electroweak force in the electromagnetic and weak forces. At this moment it is supposed that particles acquire mass throughout the Higgs mechanism. Finally at a time 10^{-6} (10^{12} K.) the Quark Hadron transition must have happened, that epoch defines the formation of Hadrons, among which protons and neutrons, are the only stable particles. After the first 3 minutes of the Universe than this standard picture we have a plasma of particles, composed by free electrons, protons, neutrons, photons and neutrinos interacting.

The second part, after the 3 minutes is what concerns us directly. To summarize the second part of the Λ CDM time-line is convenient to use the redshift z and the energy instead of time and temperature. We begin this second part of the thermal history a little before the first 3 minutes. At this time, we suppose the fundamental blocks of nuclei were compound, protons and nucleons, but the interactions with other energetic particles prevented for nuclei formation. At $z_{nucleo} \sim 10^{16}$ (~ 3 min), the particles energy was of the order of MeV ($\sim 10^{16}K$). These energies are below the binding energy of nucleus thus, the first nuclei begin to form. After nucleosynthesis, again we think the basic blocks of atoms were ready but the interactions with energetic photons prevented atom formation. The Universe continues cooling to achieve the time where radiation energy density equals to matter energy density, this time is called t_{equiv} correspondent to a $z_{eq} = 3100$. At this moment, it is thought that the matter dominated era begins. As we will see in the following section, this time would be important since the point of view of fluctuations, it marks the moment where dark matter density contrast starts to grow. The Universe continues expanding, finally at redshift $z_{recom} \sim 1100$, we suppose the particle energies were of the order of $10^{-1}eV$, thus the photons not having enough energy to pull of electrons from atoms, allows atoms to be formed. This phenomena is known as recombination. As a result of recombination, the photons do not interact anymore with atoms and could traveled freely from this time to now, this time is called decoupling time. To finish the Λ CDM thermal history, we suppose the galaxies were formed very recently at redshift ($z \sim 10$).

Before passing to the next section devoted to the inhomogeneous Universe. I underline the events from this Λ CDM time-line that would be important from the point of view of evolution of density fluctuations and structure formation. Three events would be important: the inflation period where presumably the primordial fluctuations were generated, the transition time between a radiation and a matter dominated epochs and the decoupling time.

I.3 Inhomogeneous Universe and structure formation

The precedent description of the Universe is based on the homogeneity and isotropy of the Universe. Nowadays we know that it is not strictly true. As I anticipated, the CMB anisotropies and the large scale structure shows the existence of primordial fluctuations from which the structures were formed. This fluctuations in the primordial plasma are supposed to by tinny $\Delta\rho/\rho \sim 10^{-5}$ and observationally we have verified this fact. Presumably this density fluctuations were generated by quantum fluctuations during the inflation. As I will describe in this section, the standard picture of structure formation is that structures originate from initial fluctuations in the density field and grew through gravitational attraction. It is widely accepted that baryonic matter would not provide enough gravitational attraction to form the observed structures. Thus, a fundamental ingredients in this description of structure formation would be the dark matter. It would provides the extra gravitational force to allow structures to form quicker and not inhibit by pressure effects. I will begin the section describing the Inflationary paradigm that is believed to have set the initial conditions for the study of density fluctuations and structure formation.

I.3.1 Inflation

The Inflationary paradigm emerges in 1981 to solve some conceptual problems of Big Bang Cosmology: the flatness problem, the Horizon problem and the relics problem. As an extra bonus, inflationary paradigm provides us with a physical mechanism to explain the deviations from homogeneity. Before passing to the basic description of inflation I will discuss briefly problems of the Big Bang Model to understand how Inflation is inserted in the current Standard Cosmological Model.

- *Flatness*. In order to explain the flatness problem, let us take eq.[I.17] and reexpressed in terms of the fractional densities of the critical density.

$$\Omega - 1 = \frac{k}{a^2 H^2} \quad (\text{I.36})$$

Observationally we know that $\Omega_0 \simeq 1$ today. Indicating that geometry of our local Universe is flat. The idea is that, if our Universe is flat now, it must be flat since the beginning because any deviation from this flatness condition would cause that the Universe become more curved with time. To understand this point we must notice that $|\Omega - 1|$ is an increasing function of time, meaning that flat geometry is unstable. In consequence, a value now close to zero ($|\Omega_0 - 1| \sim 0$) would indicate that the primordial value would be infinitely close to zero, indicating a "fine tuning" problem. To get an order of magnitude of this value, if we consider a Universe where there is only radiation and dust then the value of Ω_0 observed would require at early times (t_{PLANCK}):

$$0 \leq 1 - \Omega \leq 10^{-60} \quad (\text{I.37})$$

- *Horizon*. To understand this problem we have to introduce the concept of the horizon. As the Universe has a finite age, we know that even photons can travel only a finite distance within the age of Universe because of the finite value of speed of light. This gives rise to the concept of observable Universe. The horizon is defined by the maximum distance an observer can see. It delimits the space-time events which are in causal contact with observer. By the same reason then there exist a cone of past space-time events that could influence the observer denoted as the event horizon distance or past horizon. The horizon is one of most crucial concepts of FRW model.

To calculate the horizon distances we consider a photon traveling in a radial trajectory ($\Delta\theta = 0$ and $\Delta\phi = 0$), and we compute the comoving distance traveled by these photon (eq.[I.28]). To get the physical distances that would measure an observer at a time, t we have to multiply by the factor $a(t)$, giving finally the following expression for the horizon distance:

$$d_H \equiv a(t) \int_{t_e}^t \frac{dt'}{a(t')} = a(t) \int_0^a \frac{da'}{a' a' H(a')} \quad (\text{I.38})$$

The comoving horizon is the logarithmic integral of the comoving Hubble radius $d_{Hubble} = 1/aH = \frac{1}{\dot{a}}$. The Hubble radius is the distance over which particles can travel in the course of one expansion if the expansion rate was constant. Hubble radius is another form of measuring where particles are in causal contact, if they are separated more than the Hubble radius they cannot communicate now.

In a matter or radiation dominated Universe, the comoving horizon is approximately equivalent to comoving Hubble radius ($\chi_H = \frac{d_H}{a} \sim d_{Hubble} = \frac{1}{aH}$). Even though, to explain the horizon problem we must have in mind the difference between both terms. In the first case, particles separated more than the horizon distance could never have communicated; in the case of the Hubble distance, they could not communicate now. To understand this, let us think in a Hubble radius fixed at present time $d_{Hubble,0} = (a_0 H_0)^{-1}$. We must notice that we can make the comoving horizon distance χ_H as large as we want if d_{Hubble} is a decreasing function of time. In this way we get an increasing χ_H as we go back in time ($a = 0$). This would mean that particles that could not communicate today (i.e. outside Hubble radius at present) were in causal contact in the past (i.e. inside the Hubble radius in the past). In this case the Hubble radius in the past would be larger than now (see Fig.[I.15]).

Now, we have defined the horizon concept, let us return to the horizon problem, that is the most important problem of Big Bang. The CMB measurements indicate a homogeneous temperature. A homogeneous temperature implies thermal equilibrium, but widely separated points on the last scattering surface are completely outside each other's horizons. As I said before, this might happen if Hubble radius was greater in the past, but it could not happen at the matter or radiation era where the Hubble radius is an increasing function of the time.

- Homogeneity: Within the Standard Model there is no physical mechanism to generate the fluctuations of density, these fluctuations are considered only as initial conditions. Inflation gives us a plausible physical mechanism that would be responsible for both the large scale homogeneity of the Universe and the small fluctuations that were seeds of structure. Inflation paradigm gives place to adiabatic, gaussian initial fluctuations.

The main idea of Inflation is that the Universe undergoes a period of accelerated exponential expansion at early times. Inflation is triggered by the domination of a scalar field ϕ . It is characterized by two phases: the first one generates the accelerated expansion, known as slow-roll, where perturbations are generated and the second phase is called reheating where the scalar field decays into the standard particles. This scalar field behaves as a perfect fluid with pressure and density given in terms of ϕ :

$$\begin{aligned}\rho_\phi &= \frac{\dot{\phi}^2}{2} + V(\phi) \\ p_\phi &= \frac{\dot{\phi}^2}{2} - V(\phi)\end{aligned}\tag{I.39}$$

The condition of slow roll implies neglecting kinetic compared to potential energy $\dot{\phi}^2 \ll V(\phi)$, thus the potential energy is the dominant contribution to the energy density and pressure, resulting in an equation of state $p \simeq -\rho$ as the cosmological constant implying acceleration of expansion [see section 1.2.4]. From eq.[I.13] we find the evolution of energy density $\rho_\phi = cte$ and the Friedman equation with this field as the only source of energy density is:

$$H^2 \simeq \frac{8\pi G}{3} \rho_\phi = \frac{8\pi G}{3} V = cte\tag{I.40}$$

Finally solving we get the final expression for the scale factor:

$$a(t) \propto \exp\left(\sqrt{\frac{8\pi G}{3}} \rho_\phi t\right)\tag{I.41}$$

As a consequence of this acceleration small regions expand to a huge size rapidly. This accelerated expansion thus diminish spatial curvature making the Universe extremely flat. The horizon problem is solved because of the horizon size increase assuring the causal contact. This means that the astronomical regions observed today were microscopically small in the past. Since the point of view from the comoving Hubble radius, the exponential expansion of scale parameter generates a dramatical increase in the comoving Hubble radius at early times. Meaning that the principal contribution to the horizon would come from the early times and not from now . The Fig [I.15] helps us to understand this idea, the regions now separated by several Hubble radius (outside the circle in the bottom panel) were in the past in causal contact (inside the circle of top panel). In

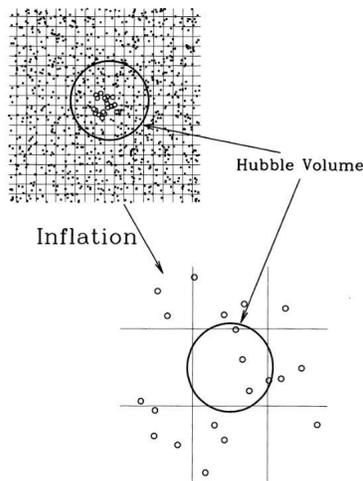


Figure I.15: The regions now separated by several Hubble radius (outside the circle in the bottom panel) were in the past in causal contact (inside the circle of top panel).

figure [I.16] the evolution of Hubble radius with time is shown. The left side shows the inflationary period and the right side shows the evolution after inflation. Before inflation started the comoving Hubble radius was larger than any cosmological distance today, so all observable scales were within the horizon. Inflationary epoch reduces the comoving Hubble radius.

The scales of interest are shown in dark shadow. As we see at the beginning all of them were within the Hubble radius thus in causal contact, as the Universe inflates, they exit the horizon. When a perturbation exits the horizon, no causal physics can affect it and it continues frozen with constant amplitude until it reenters the horizon at later times during the conventional (no accelerating) big bang expansion.

Finally, Inflation provides a physical mechanism to understand primordial inhomogeneities. Inflation generates two kinds of perturbations: scalar and tensor fluctuations in the metric. For purposes of structure formation we are interested in the scalar ones because they are related to the density and matter and are responsible for the inhomogeneities. The tensor fluctuations are related to gravitational waves and they are not coupled with density but they induce fluctuations in the CMB, in particular B-modes which are therefore a specific observable signature of inflation.

Now I will sketch how these fluctuations were generated. During inflation one assumes that the Universe consists of an almost homogeneous field and a uniform metric. This field follows quantum fluctuations as any other quantum field. Thus, the field could be expressed as $\phi = \phi_0 + \delta\phi$ where ϕ_0 is the zero-order homogeneous part and $\delta\phi$ are the quantum fluctuations. These fluctuations are the same for all wavelengths, $\delta\phi_k = T_{GH}$ where T_{GH} is the Gibbons-Hawking temperature of the vacuum state ($T_{GH} = H/2\pi$). These fluctuations average to zero, but their variance is different from zero. They are characterized by their power spectrum:

$$P_\phi(k) \propto \frac{1}{M_{Plank}^6} \frac{V^3}{V'^2} \quad (\text{I.42})$$

These fluctuations in the field could be related to those of the density by:

$$\delta\rho = \frac{dV}{d\phi} \delta\phi \quad (\text{I.43})$$

Therefore, inflation produces density perturbations in all scales with amplitude nearly equal for all the wavelengths (slight deviations because of the gradual change of V). These density fluctuations are adiabatic, that means fluctuations in the density of all species are correlated. The fluctuations are also Gaussian, in the sense that the phases of the Fourier modes describing fluctuations at different scales are uncorrelated. As a consequence of accelerated expansion, these microscopic quantum fluctuations transformed into macroscopic fluctuations. Eventually, these fluctuations associated to cosmological structures exit the horizon at the inflation period and reenter the horizon just before decoupling. Thus, the fluctuations during inflation determine, modulus what happens before the reentry parametrized through a transfer function (?), the power spectrum of fluctuations. Inflation leads to the Gaussian adiabatic and scale-invariant density perturbations. To relate this to observations usually the power spectrum is parametrized by:

$$P_s^2(k) \propto k^{n_s-1} \quad (\text{I.44})$$

where n_s is the spectral index. The next section will discuss whether or not such a spectrum can evolve so as to reproduce LSS observations.

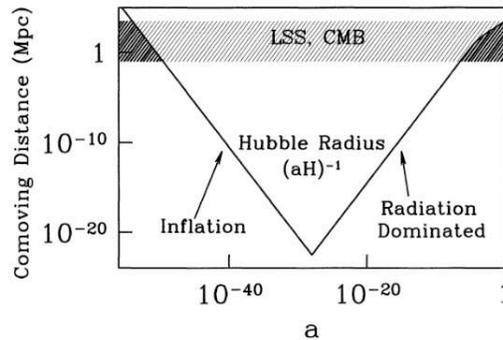


Figure I.16: The evolution of Hubble radius with time is shown. The left side shows the inflationary period and the right side shows the evolution after inflation. Before inflation started the comoving Hubble radius was larger than any cosmological distance today, so all observable scales were within the horizon. Inflationary epoch reduces the comoving Hubble radius

I.3.2 Evolution of Density Perturbations

Now that we have the initial conditions, we will concentrate on the evolution of these inhomogeneities in the density under the action of gravity in an expanding Universe. Study the evolution of density perturbations in the framework of general relativity is not a trivial exercise, fortunately, most of the essential physics can be extracted from a Newtonian approach. This approach is valid for the modes within the Hubble radius ($\lambda < d_H$), for *superhorizon* modes we have to use a general relativistic perturbation theory approach.

For the following it is convenient to define the density contrast δ in place of using the density:

$$\delta \equiv (\rho - \bar{\rho})/\bar{\rho} \quad (\text{I.45})$$

In the Newtonian framework, the cosmological gravitational potential $\phi = \phi(\vec{x}, t)$ is sourced by the density fluctuations $\delta = \delta(\vec{x}, t)$ ¹², this relation is expressed in the Poisson equation (the last equality comes from Friedman equation eq.[I.17]):

$$\nabla^2 \phi = 4\pi G \bar{\rho} \delta = \frac{3}{2} \Omega_m (Ha)^2 \delta \quad (\text{I.46})$$

The basic equations to describe a perfect fluid are the continuity equation that ensures the conservation of mass:

$$\dot{\delta} + \vec{\nabla} \cdot [(1 + \delta)\vec{v}] = 0 \quad (\text{I.47})$$

where $\vec{v} = \vec{v}(\vec{x}, \tau)$ is the velocity flow, and the Euler equation that ensures the momentum conservation:

$$\dot{\vec{v}} + H\vec{v} + \frac{1}{a}(\vec{v} \cdot \vec{\nabla})\vec{v} = -\frac{1}{a}\vec{\nabla}\phi - \frac{1}{a\bar{\rho}}\vec{\nabla}p \quad (\text{I.48})$$

with p the pressure. A fourth equation would be necessary to close the system, the equation of state $p = w\rho = w\bar{\rho}(1 + \delta)$, of the fluid in consideration. In order to get an analytic solution of this system, we are going to consider only the large scales where we can assume the fluctuations are small ($\delta \ll 1$), thus, we can neglect all the terms $\delta\vec{v}$ and v^2 , giving the following expressions:

$$\dot{\delta} + \frac{1}{a}\vec{\nabla} \cdot \vec{v} = 0 \quad (\text{I.49})$$

$$\dot{\vec{v}} + H\vec{v} = -\frac{1}{a}\vec{\nabla}\Phi - \frac{1}{a\bar{\rho}}\vec{\nabla}p \quad (\text{I.50})$$

These two equations combined with the Poisson equation give us a second order equation in terms only of the density contrast in the linear regime approximation:

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - \frac{c_s^2}{a}\nabla^2\delta = 4\pi G\bar{\rho}\delta \quad (\text{I.51})$$

where $c_s^2 = w$. Before passing to the large scale solutions in the linear regime, let us analyze this equation in a qualitative way. If we translate to Fourier space, the laplacian operator becomes $\nabla^2 \rightarrow -k^2$, leading to the following expression:

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} + c_s^2 \left[\frac{k^2}{a} - \frac{4\pi G\bar{\rho}}{c_s^2} \right] \delta = 0 \quad (\text{I.52})$$

¹²Positions of particles are described in comoving coordinates \vec{x} , physical coordinates are $r = a(t)\vec{x}$

This is the equation of an harmonic oscillator with a damping term $2H\dot{\delta}$ called *Hubble Drag*. This equation expresses the competition between the density fluctuations infalling into the gravitational potential and the effect of expansion which works like a friction term opposing the clumping, as well as the effect of pressure that opposes compression. We can identify a physical wavenumber k/a ¹³ where the term in square brackets vanishes, this wavenumber is called the Jeans wavenumber and is defined as follows:

$$k_J = \frac{\sqrt{4\pi G\bar{\rho}}}{c_s^2} \quad (\text{I.53})$$

Even if the analytic expressions are quite complicated, with the qualitative behavior would be enough to understand in rough terms the evolution of perturbations. The Jeans scale allows us to separate the two behaviors of solutions. For small scales, $k > k_J$ we have oscillatory solutions and for $k < k_J$ we have growing or decaying solutions. The Jeans length characterize the scales when evolution of the perturbations is dominated by gravity ($\lambda > \lambda_J$) and when the gravitation is negligible and dominant contribution comes from pressure that prevents growth of perturbations ($\lambda < \lambda_J$). In the following table the evolution of Jeans length for the different epochs is presented in terms of scale parameter:

$$\lambda_J \propto \begin{cases} a & a < a_{eq} \\ a^{1/2} & a > a_{eq} \end{cases} \quad (\text{I.54})$$

From the eq.[I.54], we observe that the Jeans length in a radiation dominated universe is comparable to the Hubble radius, implying that only the perturbations outside the horizon grow in this way, perturbations do not grow inside horizon. In the matter dominated epoch, the Jeans length is enormous, thus growth applies on all scales. In the next section, we will see how these large scale fluctuations grow in the different epochs.

Large Scale Limit

In the large scale limit ($k \rightarrow 0$), the solution to this equation can be expressed in terms of two functions depending only on time, one is an increasing function denoted D_+ and the other is a decreasing function denoted by D_- .

$$\delta(\vec{x}, t) = D_+(t)(\tau)A(\vec{x}) + D_-(t)(\tau)B(\vec{x}) \quad (\text{I.55})$$

where $A(\vec{x})$ and $B(\vec{x})$ are two arbitrary functions of positions related the initial field configuration, D_+ and D_- can be thought as respectively describing the overdense regions that increased their density contrast with time thanks to the action of gravity and the other describes the underdense regions that become less dense with time. The growing mode D_+ is known as the linear growth factor. Let us calculate the evolution of matter density fluctuations in a simple case, a Universe dominated by a single fluid. If we consider only the dominant components of the energy density in each epoch, we would get an idea of the growth of perturbation in the successive periods. We begin with the simpler case, a matter dominated Universe. In this case there is no pressure term ($w = 0$), thus, assuming a solution $\delta = D\delta_0(\vec{x})$, eq.[I.51] takes the following form:

$$\ddot{D} + 2H\dot{D} + 3/2H^2D = 0 \quad (\text{I.56})$$

¹³The wavenumber k is expressed in comoving coordinates so the physical wavenumber is k/a

In a Universe dominated by matter, the evolution of scale factor is $a(t) \propto t^{2/3}$, thus $\dot{a}/a(t) = 2/3t$. In this case the growing solution and the decaying solution are $D_+ \propto t^{2/3} \propto a$ and $D_- \propto t^{1/2} \propto H$ respectively. For calculating the evolution of matter perturbations in a radiation or a Cosmological constant dominated Universes¹⁴, we have to take into account multiple components that contribute to the gravitational potential and the result is a more complicated expression for each component that reads as follows:

$$\ddot{\delta}_m + 2\frac{\dot{a}}{a}\dot{\delta}_m - 4\pi G \sum_j \bar{\rho}_j \delta_j = 0 \quad (\text{I.57})$$

where the subscript labels the species and the sum is over all species.

From this expression, the behaviors of the growth factor of matter density fluctuations can be calculated for the radiation or a cosmological constant dominated epochs. The results are shown in the following Table(for the derivation see ??). Summarizing the evolution of large scale perturbations, we have seen that the over dense regions accretes matter from surroundings increasing the density contrast. This growing of the density fluctuations follows different laws of growing depending of the epoch in question. In the radiation dominated era the growing is slow only logarithmic because of radiation pressure. In the matter dominated era by contrary, the fluctuations grows proportional to a . Finally at the cosmological constant era there is no growing.

Table I.2: Evolution of large scale perturbations.

Epoch	$a(t)$	$D(t)$	$D(a)$	$d_{Hubble}(a)$
Radiation	$\propto t^{1/2}$	$\text{Ln}(t)$	$\propto \text{Ln}(a)$	$\propto a^2$
Matter	$\propto t^{2/3}$	$\propto t^{2/3}$	$\propto a$	$\propto a^{3/2}$
Cosmological Constant	$\propto e^{Ht}$	$\propto cte$	$\propto cte$	$\propto cte$

Before passing to the next topic, let us put all the pieces together for studding the evolution of a perturbation of wavelength $\lambda \propto a$ that enters in the Hubble radius in the radiation dominated epoch. In the radiation dominated epoch, for the wavelengths bigger than λ_J , there is no pressure support that prevents collapse, but at this stage, the rapid expansion driven by the energy density of radiation supresses the growth, therefore, the fluctuations grow only logarithmically. The wavelengths smaller than λ_J are dominated by the radiation pressure and they oscillates as acoustic waves. The fluctuations continue evolving until achieve matter dominated era. The wavelength is bigger than λ_J and the expansion is matter dominated thus any process could suppressed the growth, the matter fluctuations growths proportional to the scale parameter.

Velocity Field

As we have seen in previous sections, a direct observable from surveys is the redshift and it is related directly to the velocities. Thus, we are interested in the solution for the velocity field in the linear regime approximation. In fact, from the system of equations I.46 and I.49, we get the solution for

¹⁴In this expression we ignores the pressure terms because we are worried about the density perturbations where $w = c^2 = 0$

the divergence of the velocity field, done by:

$$\theta(k, z) \equiv \vec{\nabla} \cdot \vec{v} = -Ha[fA(x)] \quad (\text{I.58})$$

In the solution for the velocity field appears a new function f , called the growth factor rate that is defined as the logarithmic derivate of D in terms of scale factor ¹⁵

$$f \equiv \frac{dLnD_+}{dLna} \quad (\text{I.61})$$

Before passing to the next section I give some final remarks about the approach described here. The above treatment is a simplified view the full treatment must:

- Be relativistic, implying as the starting point the perturbed equation of Einstein This means a difference from a smooth Universe that is characterized only by one function depending only of time $a(t)$. In a perturbed Universe two functions are required both of which depends of time and space. And it must deal with gauge selection.
- This treatment models the components of the Universe as perfect fluids. But in reality the Universe consist of a mixture of fluids, and between them we have collision particles (photons, baryons, electrons) thus interactions between differents species must be taken into account. Imposing the use of Boltzman equation to treat the problem adequately. The metric which determine the gravitational forces is influenced by all species dark matter, neutrinos photons and baryons.
- This approach does not deal with non linear behavior, required to explain the collapsed structures we observe. Others approaches would be necessary as N-body simulations, Perturbations Theory or approximative analytical approaches as the spherical collapse model that will be descrided in a further section.

I.3.3 Statistical description of Density Fluctuations

In order to compare the model described until now with observations, we must define the statistical tools we will be using. As I said before, the most widely considered models are based on the inflation paradigm that gives place to adiabatic, gaussian initial fluctuations. The fact that the origin of fluctuations comes from quantum fluctuations implies stochasticity. Thus, the observable Universe is modeled as a stochastic realization of a statistical ensemble of possibilities. The goal will be to make statistical predictions that depend on the statistical properties of the primordial

¹⁵In fact in the solution for the velocity field appears two new functions f and g , that are defined in terms of D_+ and D_- .

$$\nabla \cdot u(x, \tau) = -Ha(\tau)[fA(x) + gB(x)] \quad (\text{I.59})$$

where:

$$\begin{aligned} f(\Omega_m, \Omega_\Lambda) &\equiv \frac{dLnD_+}{dLna} \\ g(\Omega_m, \Omega_\Lambda) &\equiv \frac{dLnD_-}{dLna} \end{aligned} \quad (\text{I.60})$$

perturbations. We describe the density field $\rho(x)$ as a scalar random field. In fact it is preferable to work with the contrast of density $\delta(x)$ whose statistical properties we want to describe.

$$\delta(\vec{x}) = \frac{\rho(\vec{x}) - \langle \rho \rangle}{\langle \rho \rangle} \quad (\text{I.62})$$

Its counterpart in Fourier space is:

$$\langle \delta(\vec{k}) \rangle = \frac{1}{V} \int \delta(\vec{x}) e^{i\vec{k}\cdot\vec{x}} d\vec{x} \quad (\text{I.63})$$

We will assume that the field is statistically homogeneous and isotropic as predicted by cosmological theories. For the moment, we will consider that $\delta(\vec{x})$ is a gaussian field.

The first kind of statistics that we use to characterize a random field is the 2 point statistics. In the direct space, it is the 2 point correlation function (2PCF). The 2PCF is defined as the joint ensemble average of the density at two locations separated by a distance r ,

$$\xi(r) = \langle \delta(x)\delta(x+r) \rangle \quad (\text{I.64})$$

The physical interpretation of the 2 point correlation function is that it measures the excess of probability of finding two particles in volume elements dV_1 and dV_2 separated by distance $x_{12} \equiv |\vec{x}_1 - \vec{x}_2|$, where n is the mean density is:

$$dP_{12} = n^2 [1 + \xi(x_{12})] dV_1 dV_2 \quad (\text{I.65})$$

For a random distribution $\xi = 0$, and the probability of having a pair of particles is given by the mean density squared independently of distance. If the objects are clustered, the points are correlated $\xi > 0$ the probability is enhanced, if the objects are anti-correlated $\xi < 0$ the probability is suppressed with respect to a random distribution.

The Fourier transform of the two-point correlation function is the power spectrum:

$$\langle \delta(\vec{k})\delta(\vec{k}') \rangle = \delta_D(\vec{k} + \vec{k}') P(\vec{k}) \quad (\text{I.66})$$

$$\xi(r) = \int d^3k P(k) e^{i\vec{k}\cdot\vec{r}} \quad (\text{I.67})$$

In case of gaussian fluctuations the statistical properties of random variables $\delta(\vec{k})$ are completely determined by the power spectrum.

A convenient way to express the power spectrum to give an intuition of its meaning is:

$$\frac{d\sigma^2}{dLnk} = \frac{V}{2\pi^2} k^3 P(k) = \Delta^2(k) \quad (\text{I.68})$$

which expresses that the contribution to the density variance from each logarithmic bin in k , that is the variance¹⁶ as a function of the scale. Thus $\Delta(k_0) = 1$ would mean fluctuations of the order of unity implying that linear theory does not work.

¹⁶The expression of the variance σ^2 is

$$\sigma^2 = \frac{V}{2\pi} \int k^3 P(k) d^3k \quad (\text{I.69})$$

I.3.4 CDM Power Spectrum

Now that we have in hands the statistical tools to describe the fluctuations of the density field, let us pass to discuss what are the predictions of the model described until this moment for the power spectrum of matter. Theories of structure formation predicts $P(k)$ splits in two parts, the initial or primordial power spectrum $P_i(k)$ when the Universe was very young, the second one describes how this spectra is altered by various process to produce the $P(k)$ we observe today. It is specified using the Transfer function $T(k)$ so that:

$$P(k) = T(k)^2 P_i(k) \quad (\text{I.70})$$

For the primordial power spectrum, as we have seen in a previous section, a simple power law is assumed for some constant spectral index n_s . This assumption implies that there is no characteristic scale.

$$P(k) \propto k^{n_s-1} \quad (\text{I.71})$$

To Compute the Transfer function, the standard way is to solve numerically the linearized Boltzman equations for all particle species in the model (baryons, photons, neutrinos and dark matter) for $\delta \ll 1$ and using N body numerical simulations when this condition is broken. Even though, the main features of power spectrum of matter can be understood qualitatively with the approaches developed until now. In Table I.3 the results from previous sections are summarized.

Table I.3: Evolution of large scale perturbations.

Epoch	Inside	Outside
Radiation	$\text{Ln}(a)$	a^2
Matter	a	a

Let us combine this results to get the power spectrum from matter fluctuations. First one must notice that in the absence of a physical process that imprints a characteristic scale, the power spectrum must continue being a power law $T(k) = 1$. As the Universe expands the horizon keeps increasing, so the different wavelengths will enter to the horizon scales at different moments. At horizon entry the power spectrum continue being a power law. Before recombination, the Universe was dominated by radiation, the perturbations that enter to the horizon did not grow much, logarithmically at best. Jeans length was enormous compared with horizon, thus pressure support prevents growing for a whole range of wavelengths. As the Universe became matter dominated, the matter perturbations began to grow. Recombination caused Jeans scale to drop drastically down size of galaxies today, therefore, for the most of scales we can ignore pressure support. Thus, fluctuations that entered to the horizon after matter radiation equality ($\lambda \gg \lambda_{eq}$) were never affected by microphysics, which means $T(k) \approx 1$ for $k \ll \lambda_{eq}^{-1}$. Fluctuations with $k \gg \lambda_{eq}^{-1}$ entered while contribution of dark matter to the overall density was not dominant, and they did not start to grow properly until matter domination era. In this case, the suppression of power comes from rapid background expansion rather than the pressure support. Thus, $T(k)^2$ must be a different power law, is found to be $m = -4$. Thus, the modes $k \gg \lambda_{eq}^{-1}$ have grown by the same factor since the entered to the horizon. As a consequence, the slope of the power spectrum is down by a factor k^4 .

$$P(k) \propto \begin{cases} k^{n_s} & k \ll \lambda_{eq}^{-1} \\ k^{n_s-4} & k \gg \lambda_{eq}^{-1} \end{cases} \quad (\text{I.72})$$

This last expression give us the overall trail of power spectrum of CDM, we know that CDM plays a predominant role in determining the power spectrum of matter. We will see in a further chapter how this transfer function is modified with the inclusion of baryons.

I.3.5 Spherical collapse model

The linear perturbation theory developed in the last sections fails when density contrast $\delta > 1$ but most of the observed structures in Universe have density contrast greater than this value. In order to confront observations with theory there is still a barrier to overcome, the objects that we observe, galaxies, follows from non linear gravitational evolution. Therefore is crucial to understand the process involved to extract adequately the information from observations. To get exact solutions for structure formation we must use simulations N-corps. Even though is possible to illustrate the physics involved with a simple model called spherical collapse model. The model of spherical collapse was first studied by Gunn & Gott.

The basic idea is that in the past, which we set at the time of matter dominated Universe to not have to deal with pressure, we suppose the Universe contains a ideal fluid which average density was close to critical density. A cause of fluctuations in the density there would be regions which density is over and under the critic value, this regions are supposed to behave as closed Universe that is, even if they participate in expansion , they would decelerate expansion a cause of the over density in relation to the universal expansion of background Universe. This slowing down of expansion would increase the contrast of density and gravitational potential of the local mass concentration would be more dominant. Thus this overdensities would increase their radius until achieve a maximum radius after which they would collapse to form a gravitational structure.

Let's develop a little bit more the model. In the simplified picture of the spherical collapse model, we consider an spherical Universe homogeneous and isotropic except for an over density at the origin with spherical symmetry and $\rho_{crit} + \delta\rho$. As we see in previous section a Universe with $\rho > \rho_{crit}$, the scale parameter will increase until achieve a maximal value $r = r_{max}$ and then collapse. The Friedman equation for a closed Universe with only matter is:

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0(\Omega_{m,0}a^{-3} + (1 - \Omega_{m,0})a^{-2}) \quad (\text{I.73})$$

The solution of this equation is usually parametrized in terms of $\theta = H_0\tau(\Omega_{m,0} - 1)^{1/2}$, thus:

$$\begin{aligned} r(\theta) &= A(1 - \cos\theta) \\ t(\theta) &= B(\theta - \sin\theta) \end{aligned} \quad (\text{I.74})$$

with

$$A = r_0 \frac{\Omega_{m,0}}{2(\Omega_{m,0} - 1)} \quad B = \frac{1}{H_0} \frac{\Omega_{m,0}}{2(\Omega_{m,0} - 1)^{3/2}} \quad (\text{I.75})$$

The θ parameter is scaled form of the conformal time. We have that the parameter θ increase with time. The motion equation $\ddot{r} = -GM/r^2$ relate the parameters A and B ($A^3 = GMB^2$), where M is the shell enclosed mass in r given by:

$$M = \rho_{back} \left(\frac{4\pi}{3} r_i^3 \right) (1 + \bar{\delta}_i) \quad (I.76)$$

where r_i is the initial radius of the shell and $\bar{\delta}_i$ is the average contrast of density within r_i at time t_i . Now we have all the elements to describe the simplified picture of formation of a structure. Let's see Fig. [I.4] where the evolution of r as a function of θ is shown. As we can see three epochs can separated from above solution:

- *Turnround* The spherical overdensity expands and progressively slows down until reach a maximal value. The maximal value is achieved at $\theta = \pi$ ($dr/dt = 0$) and this point is called the *turn around*. It is said that at r_{max} the sphere decouple from the expansion of the Universe.
- *Collapse* If only gravity operates, after reaching this maximum value, the radius would decrease and the sphere will collapse at $\theta = 2\pi$ that is achieved at a time known as t_{crit} , the density contrast predicted at collapse by linear theory is $\delta \simeq 1.69$.
- *Virialization* Collapse will never occur because other mechanism acts, the virialization, this mechanism dissipative convert the kinetic energy into thermic movement (random motions) the sphere will achieve an stationary state before collapse, the radius at which equilibrium is reached is denoted r_{vir} . Virialization is defined by the condition of equilibrium $V = -2K$. It is supposed to occur at $\theta = 3\pi/2$ corresponding to a $\delta_{lin} \simeq 1.58$.

I.4 Cosmological Parameters

As a summary of the chapter I finish presenting the parameters that defines the Standard cosmological model as well as the better observational constraints at the this moment from (WMAP 7 years+BAO+supernovae) As we have seen many independent observations are in agreement of Λ CDM model, even if I did not give an exhaustive list Even if until now there is no clear evidence against standard cosmological model, a number of questions are still open to speculation. The most important objection against this model is that we do not know the 96% of the Universe components. Notably, the dark energy between the open questions, is a fundamental issue that most be investigated in the following years. A lot of effort is done to get more observational data to investigate the nature of dark energy. Fourth probes are considered as privileged to developed this research: lensing, bao, supernovae and clusters. The Baryonic acoustic Oscillations, that is the subject of this thesis is considered as the least affected by systematics uncertainties at the actual stage of experiments(DETF).

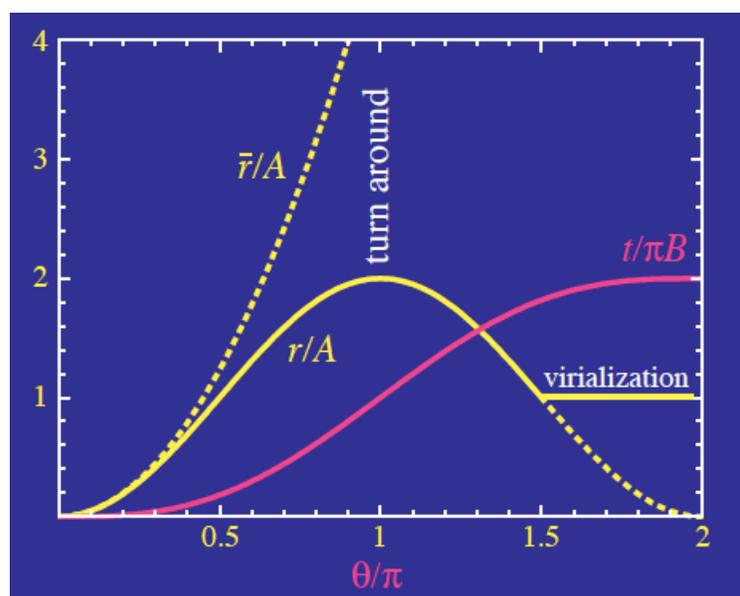


Figure I.17: Evolution of radius of an spherical overdensity as a function of the θ parameter within the framework of the spherical collapse model

WMAP Cosmological Parameters			
Model: Λ cdm+sz+lens			
Data: wmap7+bao+snsalt			
$10^2\Omega_b h^2$	2.249 ± 0.053	$1 - n_s$	$0.040^{+0.013}_{-0.012}$
$1 - n_s$	$0.015 < 1 - n_s < 0.064$ (95% CL)	$A_{\text{BAO}}(z = 0.35)$	0.472 ± 0.010
C_{220}	5757^{+38}_{-37}	$d_A(z_{\text{eq}})$	14231 ± 124 Mpc
$d_A(z_*)$	14065^{+124}_{-125} Mpc	$\Delta_{\mathcal{R}}^2$	$(2.461^{+0.088}_{-0.091}) \times 10^{-9}$
h	0.699 ± 0.013	H_0	69.9 ± 1.3 km/s/Mpc
k_{eq}	0.00990 ± 0.00025	ℓ_{eq}	139.3 ± 2.4
ℓ_*	302.52 ± 0.75	n_s	$0.960^{+0.012}_{-0.013}$
Ω_b	0.0461 ± 0.0015	$\Omega_b h^2$	0.02249 ± 0.00053
Ω_c	0.232 ± 0.013	$\Omega_c h^2$	0.1131 ± 0.0033
Ω_Λ	0.722 ± 0.015	Ω_m	0.278 ± 0.015
$\Omega_m h^2$	0.1356 ± 0.0034	$r_{\text{hor}}(z_{\text{dec}})$	284.2 ± 1.8 Mpc
$r_s(z_d)$	$152.6^{+1.3}_{-1.2}$ Mpc	$r_s(z_d)/D_v(z = 0.2)$	0.1889 ± 0.0035
$r_s(z_d)/D_v(z = 0.35)$	0.1135 ± 0.0019	$r_s(z_*)$	146.1 ± 1.1 Mpc
R	1.7275 ± 0.0093	σ_8	0.811 ± 0.023
A_{SZ}	$0.95^{+0.69}_{-0.94}$	t_0	13.78 ± 0.11 Gyr
τ	0.086 ± 0.014	θ_*	0.010385 ± 0.000026
θ_*	0.5950 ± 0.0015 °	t_*	376985^{+3003}_{-2996} yr
z_{dec}	1088.4 ± 1.1	z_d	1020.3 ± 1.3
z_{eq}	3249^{+82}_{-83}	z_{reion}	10.4 ± 1.2
z_*	$1091.10^{+0.71}_{-0.69}$		

II. Baryonic Acoustic Oscillations

In this chapter I do a review of the Baryonic Acoustic Oscillations (BAO), from a theoretical point of view. The chapter consist of 5 sections. The first section is devoted to present the basic framework. I begin with an brief non technical introduction to the BAO physics. Then I describe in a more formal way the picture in the configuration space. To finish the basic framework, I discuss how the presence of baryons modifies the transfer function presented in the first chapter. In the second section I explain motivation for studying BAO physics. This physical phenomena understood from first principles and calibrated precisely with CMB provides a good *Standard Ruler* to study the expansion of the universe and decrypt the equation of state of dark energy as well to constraint cosmology in a complementary way as others methods. In the third part of the chapter I discuss the limitations of this description: the bias problem originated by the fact the galaxies do not trace exactly the matter density and the deviations from the linear description¹. In the fourth section I talk about the current understanding of these problems and the standard way to deal with them. I present briefly the perturbation theory and reconstruction scenarios originated to deal with non linearities and consequent mode coupling. I present the halo model a semi-analytic approach to model small scales based on our understanding of structure formation. To finish I talk about the current state of the art in the measurements of BAO feature and the precision achieved.

II.1 Basic Framework

II.1.1 A not technical introduction to BAO

In order to explain the Baryonic Acoustic Oscillations, we place ourselves at the time just after inflation era. As we have seen in the first chapter, the universe was filled with an almost homogeneous baryon-photon plasma. This plasma showed tinny fluctuations generated presumably by quantum fluctuations during the inflation. This over-dense regions of the fluid initiates the propagation of acoustic waves. A way to understand this acoustic waves is thinking in the propagation of sound wave through a medium. Following the analogy, the over-densities attracted matter from under dense regions, generating over-pressed regions, this compression in the fluid was accompanied by an

¹The redshift distortions generated by fact we have access only to the redshift and not the radial coordinate will be discussed in a separate chapter.

increase of density and temperature, the photons were pushed to higher energies and the increase in radiation pressure opposes the compression driving to a consequent expansion. The expansion implies a shoot in the density, temperature and pressure that would be balanced by the surroundings generating the acoustic wave that propagates in the fluid as a sound wave. Because of the radiation pressure of photons and energy density of photon field were enormous compared with baryons the speed of propagation was very high nearly $1/3$ of speed of light².

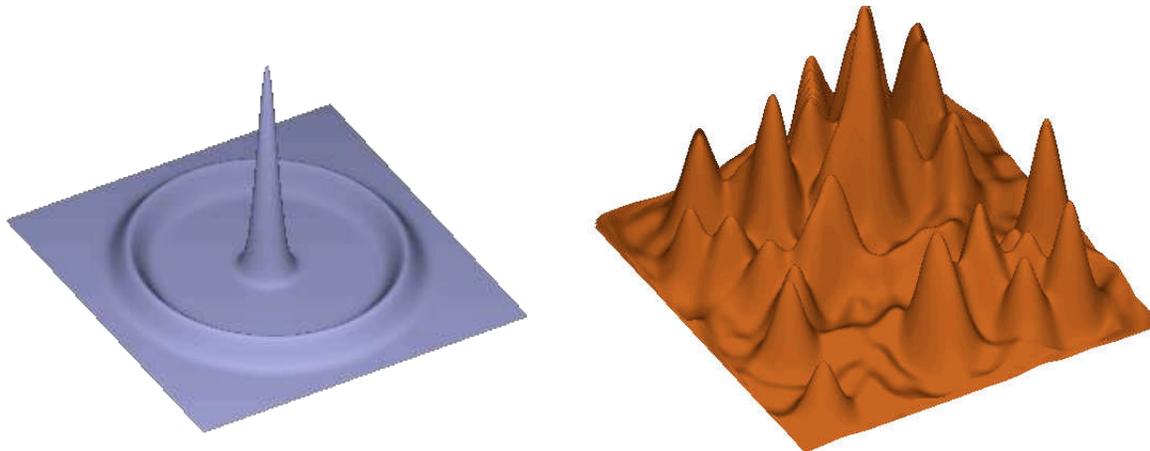


Figure II.1: Propagation of a spherical wave.

Propagation of a spherical wave. In left panel the superposition of several spherical waves.

With the expansion of the universe the fluid gets cooler and less dense. At the decoupling time, the baryon fluid pressure is released, the sound speed goes to zero and the wave is frozen at a radius equal to the sound horizon. The baryon perturbation originally at the same location as the dark matter perturbation was carried out to form a spherical shell around initial over-densities. Then the gravitational interaction between over-density of dark matter and the shell of baryons leads to the final configuration an over-density surrounded by a shell at the acoustic scale. This configuration of the matter will be the seed of the structure. The photons do not interact anymore with baryons and could travel freely until now giving us a snapshot of the fluid at the last scattering. The temperature of the photons reflects the density fluctuations (ρ proportional T^4) giving place to temperature fluctuations. The spectra of this fluctuations exhibits an harmonic series of acoustic peaks that signals a preferred angular scale. This preferred length reflects a preferred scaled in the last scattering surface ($\lambda \sim r_{LS}$ or $l \propto kr_{LS}$), this length, called the acoustic scale, is the distance that a sound wave would travel since the radiation-matter equality era until the decoupling time, it corresponds to the sound horizon at last scattering $s = c_s t_{LS}$. Because of baryons represents a fraction of total matter density, we observe the same acoustic oscillations in matter power spectrum suppressed by a factor Ω_b/Ω_m .

²Sound speed scales as square root of restoring force per unit of inertia and the radiation pressure of an electromagnetic field is $1/3$ of its energy density

II.1.2 Configuration Picture of Acoustic Oscillations

For a more formal explanation let us follow (Eisenstein et al., 2007). We consider a over-density of this plasma located at the center and suppose the rest of the universe is homogeneous and isotropic except for this over-density. Because the fluctuations are small the effect of all over-densities will be a linear superposition. The components of the universe are dark matter, the plasma photon-baryons and neutrinos. The perturbations are adiabatic implying that all species are equally perturbed³. In Figs. [II.2] and [II.3] a set of snapshots describing the evolution of a mass profile of the perturbations are shown. The profile represents the density fraction times the square of radius, the area under the curve is the mass of the perturbation in comoving coordinates. In Fig.[II.2.a) we see this initial

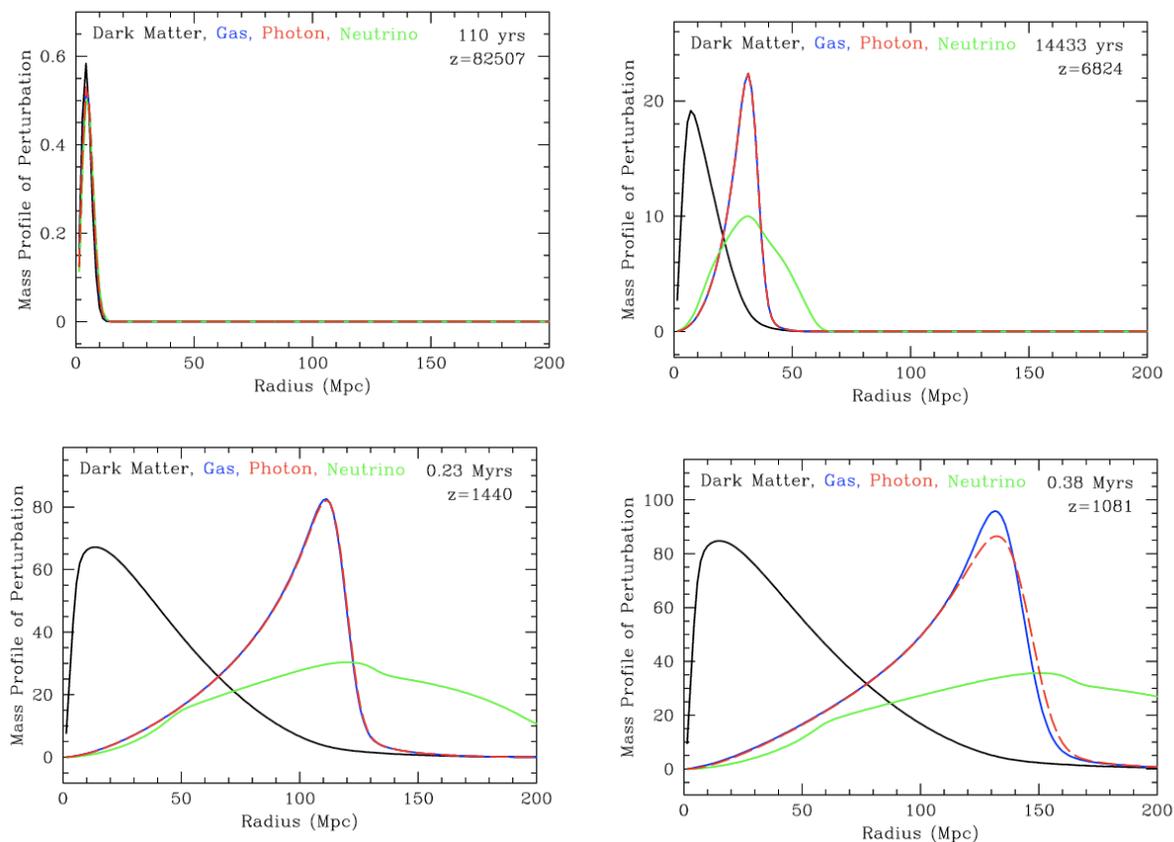


Figure II.2: Snapshot of evolution of the mass profile vs comoving radius of an initial point-like over-density located at the origin, the black, blue, red and green are respectively dark matter, baryons, photons and neutrinos rescaled to be at same scale. From (Eisenstein et al., 2007).

configuration of mass profile of the perturbations. At this moment, the baryons are coupled with photons in a single fluid, the pressure of radiation of photons generates an spherical wave around the over-density that begins to propagate Fig.[II.2.b]. The dark matter interact only gravitationally

³In fact relativistic species (photons and neutrinos) are perturbed 4/3 bigger than non-relativistic (baryons and dark matter)

in consequence it stays at the center. Neutrinos are not bounded gravitationally, they stream away from initial point-like perturbations. In panel Figs. [II.2.b] and [II.2.c], the spherical shell continue expanding, fractional density drops because energy is being spread in a large area. The neutrinos spread out and the dark matter remains at the center collecting in the overall density perturbation increasing the width of the peak.

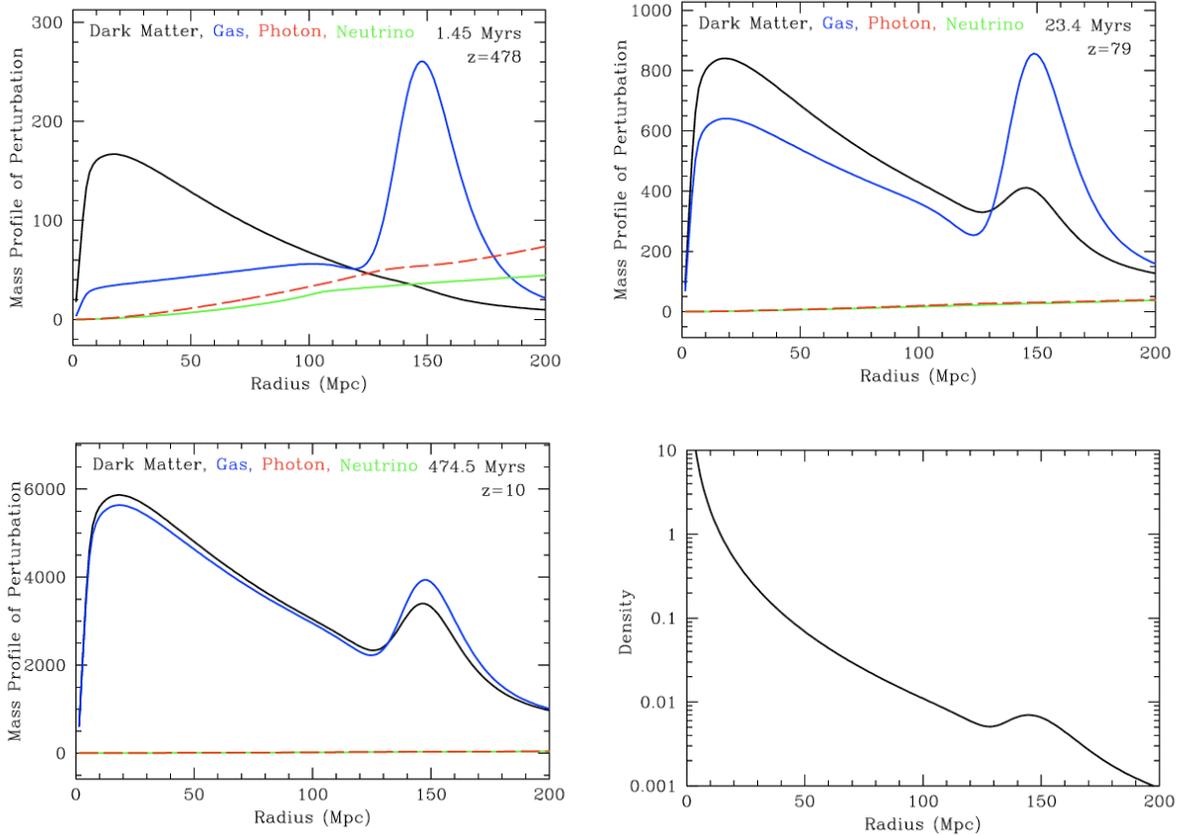


Figure II.3: Snapshot of evolution of the mass profile vs comoving radius of an initial point-like overdensity located at the origin, the black, blue, red and green are respectively dark matter, baryons, photons and neutrinos rescaled to be at same scale. From (Eisenstein et al., 2007).

The decoupling of the photons leaves the baryons frozen forming a spherical shell around the overdensity Fig.[II.2.d]. Now the configuration of a dark matter perturbation at the center surrounded by a shell of baryons begins to interact gravitationally giving place to the final configuration of an over-density at the center surrounded by a shell at the sound horizon distance Figs. [II.3.a], [II.3.b]. Without the pressure of radiation the accretion of matter of over dense regions by gravity is done even more rapidly than Hubble expansion, increasing the contrast between over-dense and under-dense regions. The final configuration in the original pic at the center surrounded by a shell of baryons at roughly 150 Mpc in radius (Fig. [II.3. c]). Finally in Fig.[II.3.d] we have the density as function of the comoving radius.

Each initial perturbation have generated a spherical wave around each over density, the superposition of this waves hide the preferred sound scale, but statically is still possible to get a signal. The galaxies will form at $z \sim 10$ preferably at the over-densities , generating a preferred separation between galaxies at the acoustic radius. Statistically we can detect this preferred scale in the correlation function as a peak at the sound horizon. Equivalently, a pic in the correlation function traduces to a serie of harmonic oscillations.

As we have seen the BAO and CMB are both a consequence of same phenomena the acoustic oscillations. The differences are that CMB give us a two dimensional snapshot of density fluctuations and BAO a three dimensional picture In Fig.[II.4] the angular power spectrum of CMB and the power spectrum of matter are shown.

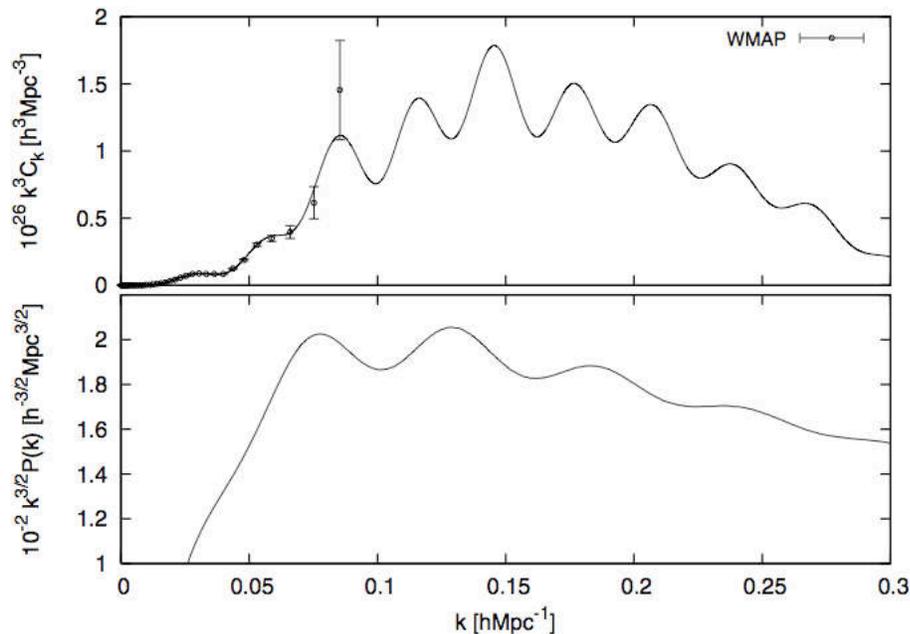


Figure II.4: Power spectra of matter and Angular power spectrum from CMB. From (White, 2005).

II.1.3 Baryon Signatures in the CDM Transfer Function

I have introduced in chapter 1 the pure CDM transfer function and the phenomenology that determines its form. Even if CDM plays a predominant role in determining the power spectrum, when the baryons are taken into account we observe that the coupling between baryons and photons generates additional structure in the transfer function compared with a pure CDM scenario.

We have seen that between epoch of equality radiation-matter (z_{eq}) and the recombination (z_d) acoustic oscillations were generated in the plasma photon-baryons. During this phase the amplitude of baryons can not grow, they undergoes harmonic motion with amplitude and velocity decaying.

Once the photons are decoupled from baryons we get the the solutions for perturbations in terms of growing and decaying modes for a pressureless fluid.

The baryons have two effects in the transfer function: the suppression of power and the oscillations below the sound horizon. Let us separate the baryons and CDM modes to understand the total effect. The baryons and CDM modes that enter before matter-radiation equality grow logarithmically. The baryon modes that enter to the horizon after matter-radiation equality but before decoupling oscillate with decaying amplitude so they can not contribute to the gravitational potentials. For the case of CDM modes that enter after equality and before decoupling, they grow as a universe with matter density without baryons $\Omega_c = \Omega_m - \Omega_b$. As a consequence all modes within the period $z_{eq} - z_d$ are suppressed.

To fix the ideas I use the fitted form proposed by (Eisenstein & Hu, 1998a). This fitting form of the transfer function models the baryons and dark matter using a effective transfer function for each species. In a CDM-baryon model universe $\Omega_b + \Omega_c = \Omega_0$. Thus , T_c is the transfer function of CDM, T_b is the transfer function of baryons pondered by the fractional density.

$$T(k) = \frac{\Omega_b}{\Omega_m} T_b(k) + \frac{\Omega_c}{\Omega_m} T_c(k) \quad (\text{II.1})$$

In a CDM model with baryons, there is an additional scale playing a role in the evolution of perturbations: the sound horizon at the time of recombination s . The transfer function of dark matter is affected by baryons only at scales below the sound horizon s where the effect of baryons is a suppression power. This is modeled using a generalized transfer function $\sim T_0(k, \alpha_c, \beta_c)$ and a factor f that smoothes the transition between a transfer function almost without baryons before s and a baryon loaded solution.

$$T_c(k) = f T_0(k, 1, \beta_c) + (1 - f) T_0(k, \alpha_c, \beta_c)^4 \quad (\text{II.2})$$

In the case of baryons, T_b is given by two oscillatory terms, $T_w(k)$ and $T_s(k)$. The term T_w

show a series of declining oscillations ⁵ suppressed by the decay of potentials between z_{eq} and z_d that corresponds to the limiting case in the absence of CDM. The other term T_s is a rapidly decreasing function and it is necessary to describe the transition from large to small scales, it comes from the fact that the suppression of power is for scales smaller that s .

$$T_b = T_w(k) + T_s(k)^6 \quad (\text{II.5})$$

In Fig.[II.5], we show the transfer function from simulations and the fitting form of Einsenstein and Hu, for different parameters Ω_b and Ω_m . The dotted line signals a pure CDM model , showing

⁴The factor is given by $f = 1/[1 + (ks/5.4)^4]$

⁵The declining of oscillations is originated by the silk damping not treated in this explanation

⁶The expressions for the terms $T_w(k)$ and $T_s(k)$ are:

$$T_s(k) = \frac{j_0(ks)T_0(k, 1, 1)}{1 + (ks/5.2)^2} \quad (\text{II.4})$$

$$T_w(k) = \frac{j_0(ks)e^{(k/k_{silk})^{1.4}}}{1 + (\beta_b/ks)^3} \quad (\text{II.5})$$

the two effects of baryons in the transfer function, suppression of power and oscillations. We observe on the top purely baryonic models without CDM, in this case the oscillations have alternating sign in the transfer function.

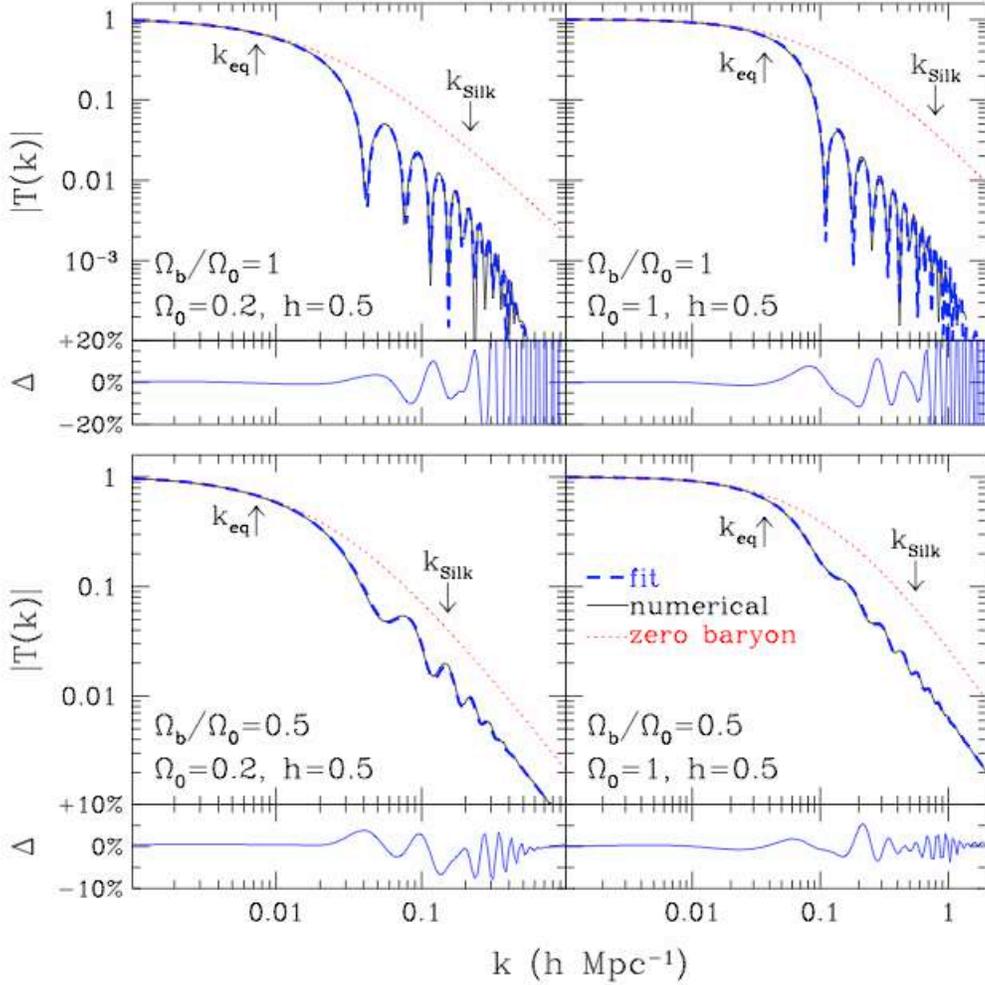


Figure II.5: Examples of Transfer function of CDM with baryons for different set of parameters Ω_m and Ω_b from simulations and the fitting formula of Eisenstein and Hu. On top pure baryonic models. Figure from (Eisenstein & Hu, 1998a).

Phenomenologically, to understand the effect of baryons let us imagine a baryonic model without dark matter. The perturbations still are present, and the subsequent launching of spherical waves on each over-density. The spherical shells are frozen at recombination time but in this case all the matter is displaced in the shells, there is no perturbation at the center, so there is no peak in the correlation function, because it is generated by the correlation between the center and the shell, indeed there is a change in the slope at twice the sound horizon corresponding to the correlation of

two points in the sphere. So the presence of baryons is one of the strongest arguments in favour of non baryonic matter that has decoupled before dark matter.

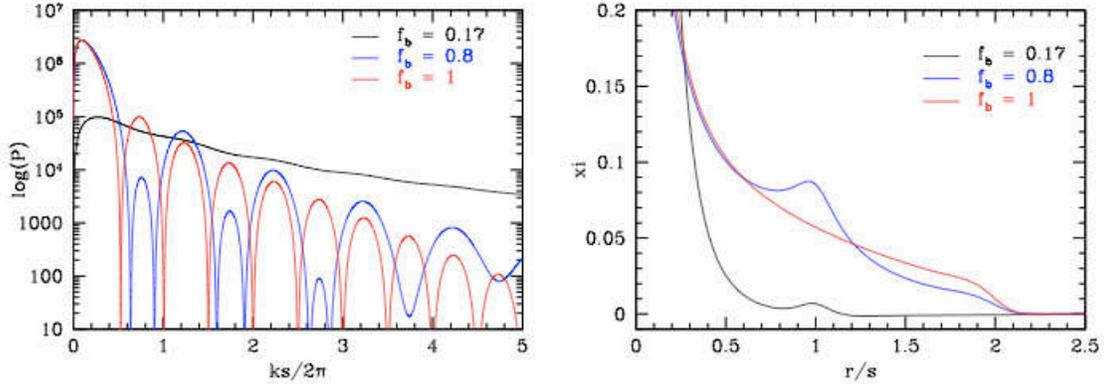


Figure II.6: Effect of baryons in power spectrum and correlation function. From <http://scholar.harvard.edu/deisenstein/book/baryon-acoustic-oscillations>.

II.2 Defining an Statistical Standard Ruler

In first chapter we have seen that distance measurements as a function of scale factor or redshift directly map the expansion history. As we have seen in the previous section the acoustic phenomena imprinted a characteristic scale in the distribution of matter: the sound horizon. This scale is completely computable theoretically since first principles. The sound horizon s is given by:

$$s = \int_0^{t_{rec}} c_s(1+z)dt = \frac{1}{\sqrt{\Omega_m H_0^2}} \frac{2c}{\sqrt{3z_{eq}r_{eq}}} Ln \frac{\sqrt{1+r_{eq}} + \sqrt{r_{eq}+r_{rec}}}{1 + \sqrt{r_{eq}}} \quad (\text{II.6})$$

The aim of BAO measurements is to learn about the expansion using this characteristic scale imprinted in matter distribution as a standard ruler.

The concept of *Standard Ruler* consist of using an object of known size at a single redshift (or wich size change in a well-known way with redshift) and use this patron to compare with the apparent size at different redshifts. This technique is similar in nature to the supernovae that uses an standard "candle" to calibrate the distances, the difference is the kind of standard pattern used to map the expansion history. The comoving size of this ruler in the transverse ($r_{||}$) and perpendicular (r_{\perp}) directions are related with the observables $\Delta\theta$ and Δz . If we know the true dimensions of $r_{||}$ and r_{\perp} we can access to $d_A(z)$ and $H(z)$ respectively. Even if we do not know the comoving size of an object using the Alcock-Paszhisky it would be possible to access to the product $D_A(z)H(z)$.

$$r_{||} = \frac{c\Delta z}{H(z)}. \quad (\text{II.7})$$

$$r_{\perp} = (1+z)D_A(z)\Delta\theta. \quad (\text{II.8})$$

In the case of BAO we know the size of the acoustic feature thus we can use it as an standard ruler to study expansion. Using this apparent distances (angular size or redshift extent) we can constrain

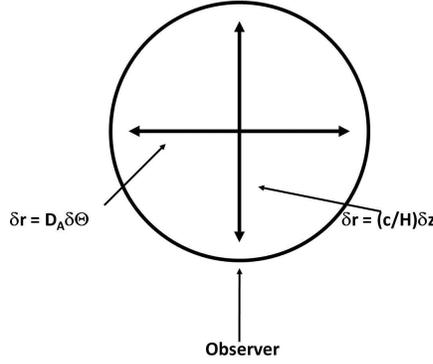


Figure II.7: The comoving size of this ruler in the transverse ($r_{||}$) and perpendicular (r_{\perp}) directions are related with the observables $\Delta\theta$ and Δz .

the parameters of dark energy via their cosmological dependence on the Hubble parameter $H(z)$ ⁷.

Thus, to study the expansion history of the universe we will measure the Hubble parameter $H(z)$ and $D_A(z)$ using the baryonic acoustic feature at different redshifts.

This characteristic scale has a statistical nature, even though, it could be measured using the two point statistics (see Fig. II.8).

7

$$H(z) = H_0^2 [\Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_X \exp\left(3 \int_0^z \frac{1+w(z)}{1+z} dz\right)]^{1/2}. \quad (\text{II.9})$$

$$D_A(z) = \frac{c}{1+z} \int_0^z \frac{dz}{H(z)}. \quad (\text{II.10})$$

where Ω_X is the present day dark energy density and $H_0 = h \times 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, the present day Hubble constant. This expression implies an equation of state of dark energy $w_X(z)$ given by the following expression (?):

$$w_X(z) = \frac{p_X(z)}{\rho_X(z)} \quad (\text{II.11})$$

The energy density of dark energy thus writes as follows:

$$\rho_X(z) = \rho_X(0) \exp\left(3 \int_0^z \frac{dz'}{1+z'} (1+w(z'))\right). \quad (\text{II.12})$$

In the case of a cosmological constant, ie. $w = -1$, the expression of $H(z)$ simplifies .

$$H(z) = H_0^2 [\Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_X(1+z)^{3(1+w)}]^{1/2}. \quad (\text{II.13})$$

where Ω_X represents the cosmological constant if $w=-1$.

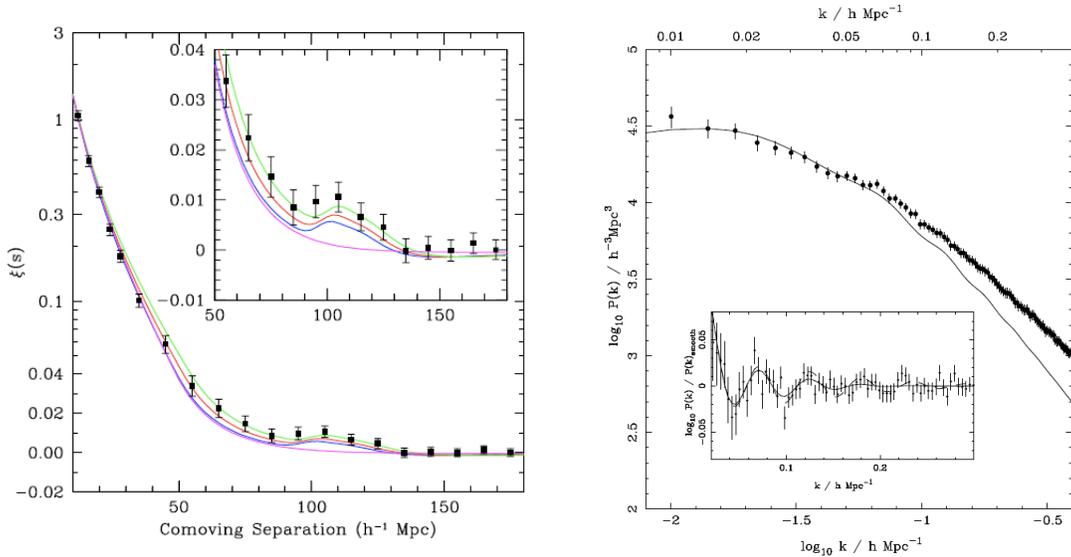


Figure II.8: First Detection of BAO in 2005. Left panel correlation function of LRG's in SDSS (Eisenstein et al., 2005). Right Panel power spectrum of main galaxy sample and LRG sample (Percival et al., 2007). The line corresponds to fitting model from Eisenstein & Hu (Eisenstein & Hu, 1998b).

The ideal case is to compute the 2 point statistics separating the line-of-sight and parallel directions, locate the feature corresponding to the sound horizon, and measure $\Delta\theta$ and Δz correspondent to this sound horizon at variety of redshifts. Then, using the constraints based on CMB results, compare to the value at $z \approx 10^3$ to get d_A and $H(z)$, and infer expansion history. Nowadays, when data sets do not have enough statistics, the angle average statistics is used, what implies analyzing the transverse and perpendicular direction together.

A final point must be taken in account before inferring cosmological consequences. As we have seen to define the distances a fiducial cosmology must be assumed. To deal with the distortions introduced by taking the wrong cosmology. The standard approach is define a scale parameter called "dilatation parameter" that takes account of changes in scale issues of taking wrong cosmology⁸. In the case of the angle average correlation function, the dilatation parameter is a combination of dilation in transverse and perpendicular directions as defined in (Eisenstein et al., 2005):

$$D_V = \left(D_M(z)^2 \frac{cz}{H(z)} \right)^{1/3}. \quad (\text{II.14})$$

This definition takes into account that d_A and $H(z)$ changes differently. In addition his quantity was defined in such a way that the ratio $D_V(z)/s(z)$ is approximately independent of choice of fiducial cosmology. Thus, measuring $s(z)$ with a fiducial cosmology and given the acoustic scale r_s

⁸When the redshifts are close enough the distortion could be thought simply as a dilatation effect, changing the spherical features for ellipsoids.

from CMB we can determine the distance D_V to that redshift using the following expression:

$$\frac{D_V(z)}{r_s} = \frac{D_{V, fid}(z)}{\gamma s_p}. \quad (\text{II.15})$$

As we will see in the following section non linearities, bias and redshift distortions modifies the acoustic feature. Fortunately, the effect of taking a wrong cosmology is not degenerated with the theoretical systematics as non linearities, redshift distortions and bias enabling to use this acoustic feature to investigate the expansion history (Seo & Eisenstein, 2005).

II.3 Systematic Uncertainties from Theory

At the actual state of BAO measurements, with the increasing volume at redshift precision of spectroscopic surveys the systematic error in the BAO method are more likely to arise from theoretical side (Seo & Eisenstein, 2005). Unfortunately we do not have access to the linear theory matter power spectrum in real space. Non linear structure formation, redshift distortions and galaxy clustering bias degrades the acoustic signature (Meiksin et al., 1999) (White, 2005). The correct modeling of this phenomena is necessary to achieve a precision better than 1% in current and future experiments. In this section the bias and non linearities will be discussed. Concerning the redshift distortions a further chapter will be devoted to this important subject.

II.3.1 Non linearities

As we have said in chapter 1, on large scales the matter distribution is well described by linear theory, however on smaller scales (about 10 Mpc) dynamics are highly non linear thus, N-body simulations are required to understand astrophysical process involved. There is a range between where the behavior is quasi linear and for the BAO phenomena is necessary to model accurately to achieve the precision required by current and future experiences.

Basically we can distinguish two effects of the non linear regime: first, the gravitational growth of perturbations in one mode is increasingly coupled with perturbations in others modes, giving place to higher order contributions. This mode coupling blurs the initial features at given mode, erasing the contrast. The second effect is the increase of power in large wave numbers (small scales) due to non linear growth. This additional power modifies the slope of power spectrum as a function of wave number, the resulting change in slope shifts the position of the peaks.

In Fig.[II.9] the non linear power spectrum issue from N-body simulations from (Seo & Eisenstein, 2005) is shown for different redshifts ($z=0.3$, $z=1$, $z=3$). The solid lines indicates the power spectrum divided by a zero baryon power spectrum giving the wiggled part of spectrum, the vertical lines indicate the k_{max} defining where the non linearities become important. We observe the two effects of non linear behavior in power spectrum: loss of contrast and excess power from non linear collapse . We observe that , as the amplitude of the density perturbations grows with time, these nonlinear effects become stronger and proceed to larger scales.

The figure shows also how at higher harmonics the peaks are slightly shifted in relation to the initial power spectrum. In the right panel of Fig.[II.9] the derivate $dLnP/dLnk$ is shown. This plot

permit us discern the degree of degradation of acoustic feature.

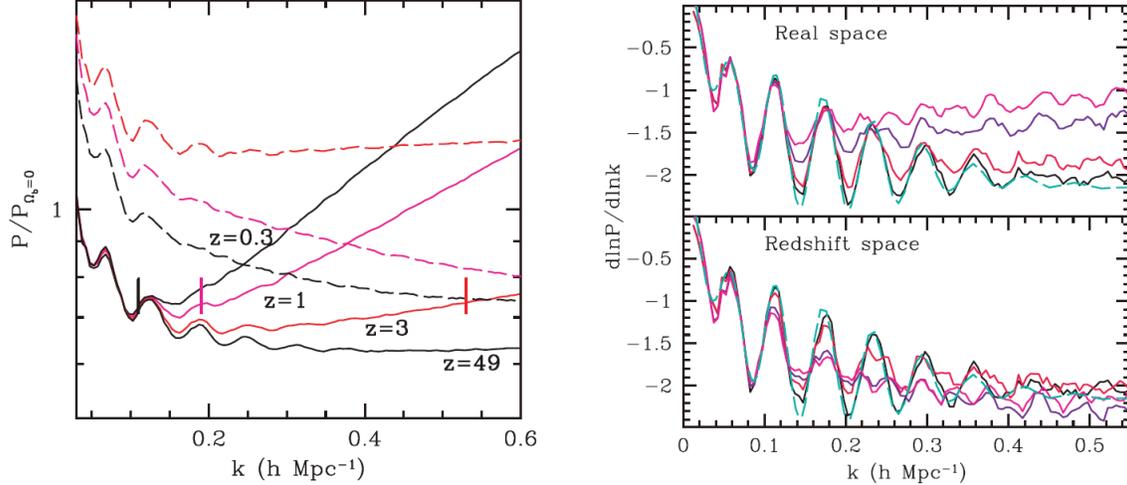


Figure II.9: Left panel. Matter power spectra at various redshifts from Seo and Eisenstein 2005. The solid lines are the power spectrum divided by no wiggle power spectra. The vertical lines indicate the scale k_{max} where linear theory validity for rms $\sigma \sim 0.5$, the value of $k_{max} = 0.11hMpc^{-1}$ for $z = 0.3$, $k_{max} = 0.19hMpc^{-1}$ for $z = 1$ and $k_{max} = 0.53hMpc^{-1}$ for $z=3$. We observe that oscillations are visible at $k < k_{max}$. The dashed lines are in the redshift space. Right panel. $dLnP/dLnk$ from the matter power spectrum at various redshifts in real [top panel] and redshift spaces [bottom panel]. Green line corresponds to input power spectrum from CMBFAST, black the linear power spectrum at $z=49$, the non linear power spectrum at $z=3$ red, violet $z=1$ and pink $z=0.3$. Dashed lines shows the correspondent power spectrum in the redshift space. Both figures from (Seo & Eisenstein, 2005).

In Fig.[II.10] the result of non linear behavior is shown in configuration space. The oscillatory feature traduce to a peak at scale sound horizon, the information in fourier space is carried in the contrast between peaks and troughs, thus the effect of non linearities generates a broadening of the peak and a shift generated by mode coupling.

To explain the phenomenology behind non linearities in configuration space, (Eisenstein et al., 2007) builds a simple model for the non linearity in terms of the differential motion of pairs. In this model, the pairs initially separated preferentially a 150 Mpc, have moved around 10 Mpc from its initial condition due to the velocity flows and non linear collapse. Thus the peak in the correlation function initially placed at 150 will be broadened by this displacement of the matter. This behavior is modeled using the lagrangian approach up to first order (Zeldovich approximation). The validation of this analytic approach has been done using N-body simulations. This simple approach suggests to model the non linear effect as a smoothing gaussian given by

$$P(k) = exp \left[-\frac{1}{2} \left(\frac{k_{\perp}^2}{\sigma_{\perp}^2} - \frac{k_{\parallel}^2}{\sigma_{\parallel}^2} \right) \right] \quad (\text{II.16})$$

where σ_{\parallel} and σ_{\perp} are given by $\sigma_{\parallel} = s_0 D$ and $\sigma_{\perp} = s_0 D(1 + f)$ with s_0 that characterizes the scale

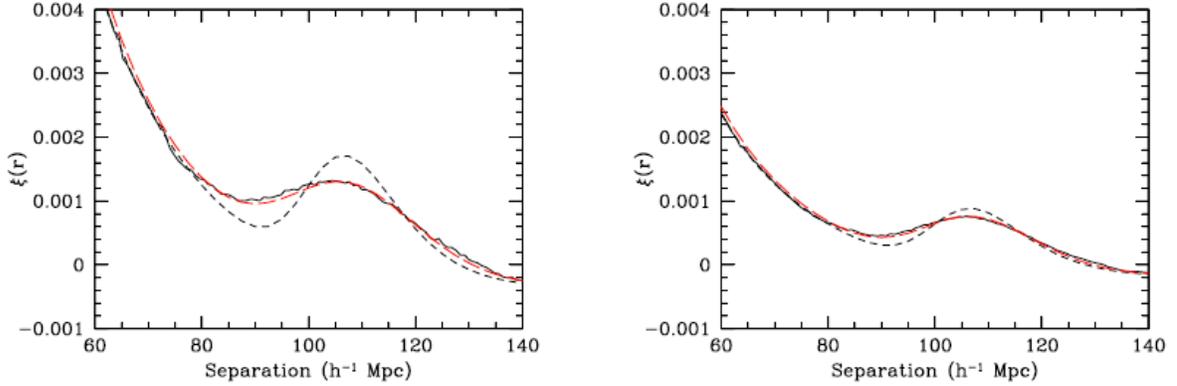


Figure II.10: Correlation functions from (Eisenstein et al., 2007). The solid lines is the non linear correlation issue from simulations , the dotted line is the linear correlation function and the red line is the model correlation function. Left panel for $z=0.3$, right panel for $z=1$

of radial displacements. A variant of this model will be used to model the correlation function in chapter 8.

II.3.2 Bias

Up to now we have supposed that galaxies map the distribution of matter. However, the matter distribution do not correspond to the galaxy distribution. The galaxies are biased tracers of the matter. The common models of the correlation function uses a linear and deterministic bias:

$$\delta_g(x) = b\delta(x). \quad (\text{II.17})$$

This bias prescription implies that

$$P_g(k) = b^2 P(k) \quad (\text{II.18})$$

Even if usually this first approximation is used, the bias enclose all the complex galaxy formation dynamics thus is could be expected to be non linear, scale dependent or even stochastic [33, 34] . There is not accepted framework for bias.

The bias prescriptions available in literature could be described in terms of the following assumptions:

- Linear vs Nonlinear (Scale independent vs scale dependent). At large scales is assumed that a linear bias is a good description, implying scale independence. Some analyses () suggest that bias is the sum of two terms one of them k independent and the other one k dependent.
- Lagrangian vs Eulerian. The time where the bias is established determines the lagrangian and eulerian approaches.

In the case of lagrangian approach, the bias is established at initial conditions, thus, non linear evolution affects the bias adding non linearities, scale dependency and stocasticity. In Eulerian approach, the bias is established at present time thus is not subject to non linear behavior.

- Deterministic vs Stochastic . Deterministic or local means that density of galaxy is completely determined by local density field. Stochastic comes from the hypothesis that galaxy formation depends on other variables besides density field and in consequence the relation between δ_g and δ are stochastic.

There are three major lines of investigations to study galaxy biasing: hydrodynamic simulations with simplified astrophysics, 2) phenomenological mapping and 3) halo model formalism. Within Halo model formalism, the bias problem is split in two separated problems: formation and clustering of dark matter halos, neglecting non gravitational effects, and the second, distribution of galaxies within halos that includes all non gravitational effects. It is believed that the spatial distribution of galaxy clusters has little effect from non gravitational effects. In the next section I will explain briefly the Halo model. Even if a variety of bias prescriptions are available, N-body simulations analysis concludes that effect of bias is an excess of clustering that appears as an additional constant term in the biased power spectrum even in non liner regime. This excess of power varies smoothly with scale at least at large scales implying that bias do not erase the acoustic feature (Seo & Eisenstein, 2005). In Fig.[II.11], I show an example from (Seo & Eisenstein, 2005) of bias effect in non linear power spectrum issue from simulations using two simple prescriptions of bias denoted by *NUM* and *MASS*⁹. The different bias corresponds to different minimum group multiplicity for particles of N-body simulation. The solid lines corresponds to real space, dashed lines to redshift space. The multiplicities $m=4, 10$ and 30 reproduces the results for bias of $1.7, 2$ and 2.5 for *NUM* and $2.4, 2.7$ and 3.1 for *MASS* at $z=1$. From the figure we observe an increase in the amplitude of the power as a function of the bias for both prescriptions of bias. The amplitude depends of the bias prescription, comparing similar bias in the 2 schemes we observe that ($m=4$ *MASS* to $m=10$ *NUM* in $z=1$) the *MASS* presents less anomalous power up to $k \sim 0.3hMpc^{-1}$.

Finally in Fig.[II.12] the logarithmic derivate of the power spectrum $dLnP/dLnk$ is shown, from the figure we verify that the power spectrum preserves the oscillatory features at least $k \sim 0.2hMpc^{-1}$

II.4 Approaches

Several approaches have been developed to model the quasi non linear range. In the following subsections I discuss briefly some of the approches. I present the variants of perturbation theory, reconstruction scenario and finally the Halo model. As we will see, the different approaches would introduce the problem of bias and redshift distortions within their framework.

⁹The two prescriptions follows the formalism of Halo Occupation Distribution (HOD). In this formalism the distribution of galaxies are deterministic models based on halos and biased tracers. The 2 prescriptions are : *NUM* where halos above a threshold have a galaxy and the other *MASS* where halos above a threshold has as many galaxies as linear with the mass.

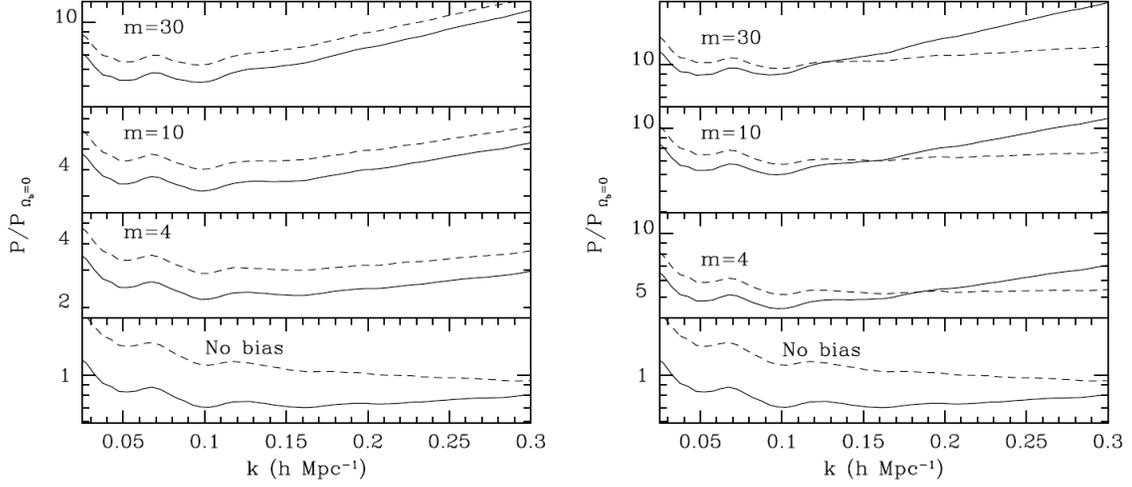


Figure II.11: Wiggles from power spectrum from N-body simulations considering 2 bias schemes: a number weighted scheme denoted NUM, and mass weighted scheme denoted MASS. The solid lines corresponds to real space, dashed lines to redshift space. The multiplicities $m=4, 10$ and 30 reproduces the results for bias of $1.7, 2$ and 2.5 for NUM [right] and $2.4, 2.7$ and 3.1 for CMASS [left] at $z=1$. From (Seo & Eisenstein, 2005)

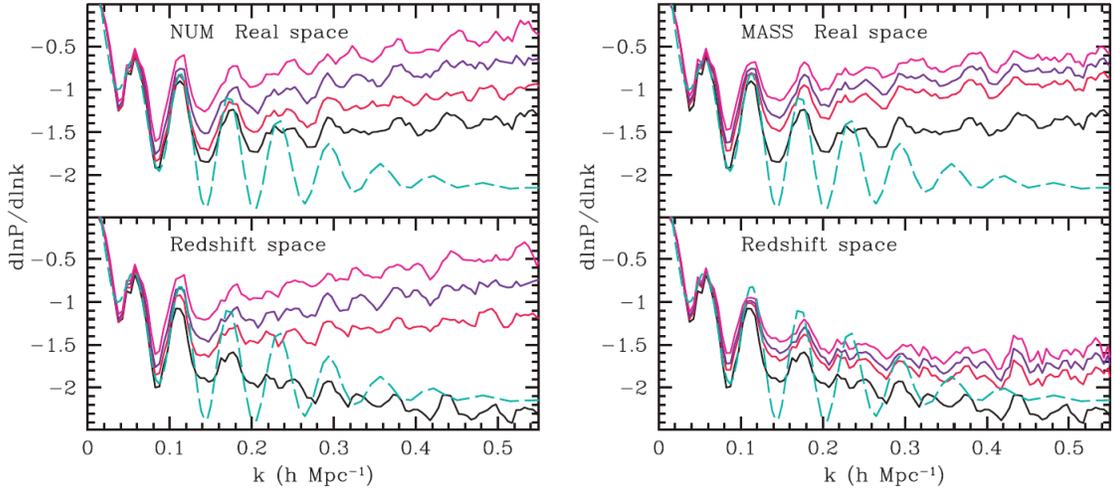


Figure II.12: Derivates from biased power spectrum $d \ln P / d \ln k$ from N-body simulations considering 2 bias schemes. In top results for $z=1$ in bottom $z=0.3$. At each plot real space plot is in top and redshift space at bottom. In right panel the number weighted case (NUM) in left side the mass weighted scheme MASS. The colors corresponds different multiplicities $m=4$ (red), 10 (blue), 30 (magenta), that corresponds to bias of $1.7, 2$ and 2.5 for NUM [right] and $2.4, 2.7$ and 3.1 for CMASS [left] at $z=1$. The black line the non linear matter power spectrum in green the input power spectrum. From (Seo & Eisenstein, 2005).

II.4.1 Perturbations Theory

The fact that the acoustic feature falls in the large scale, enables using perturbation theory as a correction to linear result.

In this section I presented some of the recent analytical approaches for computing the non linear matter power spectrum using perturbation theory. I discuss briefly how these analytical approaches predict the behavior of 2 point statistics of dark matter improving validity range upon a pure linear theory description. I present the advantages and disadvantages of each one. To finish, I will give some comments comparing the differents approches.

Standard Perturbation Theory (Eulerian)

The standard approach of perturbation theory (SPT) in Eulerian coordinates is based on the fact that at large scales we can suppose fluctuations are small, an expansion of density and velocity around a linear solution provides and accurate description of the cosmological fields:

$$\delta(x, \tau) = \sum_{n=1}^{\infty} \delta^{(n)}(x, \tau) \quad (\text{II.19})$$

$$\theta(x, \tau) = \sum_{n=1}^{\infty} \theta^{(n)}(x, \tau) \quad (\text{II.20})$$

In chapter 1, the equations of motion for the general case were introduced and they have been linearized to solve the system. The idea of Perturbations theory is to extend this result and include the non linear phenomena.

Advantages and Disadvantages

- Standard perturbation theory taken to infinite order provides an exact solution. In Fig.[II.13] the power spectrum from SPT is shown for 1-loop and 2-loop for Λ CDM model at $z=1$, and $z=0$ compared with N-body simulations. As we can observe from figure, SPT works well at high redshift and large scales but is badly behaved for scales and redshift smaller ($z \sim 1$ agrees with simulation to 1% out to $k = 0.2hMpc^{-1}$).
- The positivity of the perturbative corrections is not guaranty and SPT has a poor convergence even adding the higher order corrections. The Fig.[II.13] shows the predictions from 2-loop improve the result from 1-loop for $z=1$, however the same trend is not observed at redshift $z=0$, where the result from 2-loop is worst that the predictions from 1-loop. In order to explain the problems of convergence in Fig.[II.14] is shown the comparison between SPT and RPT loop expansion for the nonlinear power spectrum. The $P_{PT}^{(1)}$ ($P_{RPT}^{(1)}$) denotes the 1-loop correction in PT (RPT). Dashed lines denotes negatives contributions and solid lines positive ones. At large scales the linear contribution dominates as expected but as the nonlinear scale approaches, the different perturbative corrections become increasingly important , in addition some contributions cancels among them. By contrast in RPT all contributions are positive

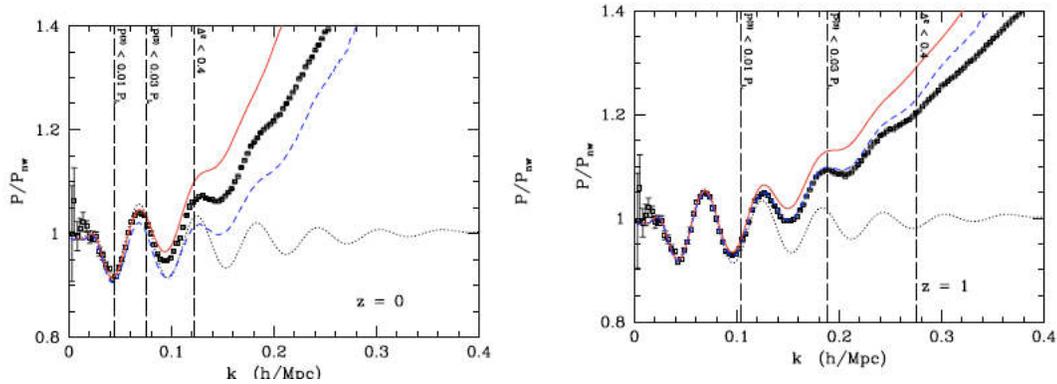


Figure II.13: The power spectrum from SPT is shown for 1-loop (red solid line) and 2-loop (blue dashed line) for Λ CDM model at $z=1$, and $z=0$ compared with N-body simulations (points). In black solid line the linear power spectrum. From (Carlson et al., 2009)

and successive terms dominate at smaller scales and there are no cancellations. The partial re-summation method of RPT change completely the behavior of perturbation expansion guarantying the convergence.

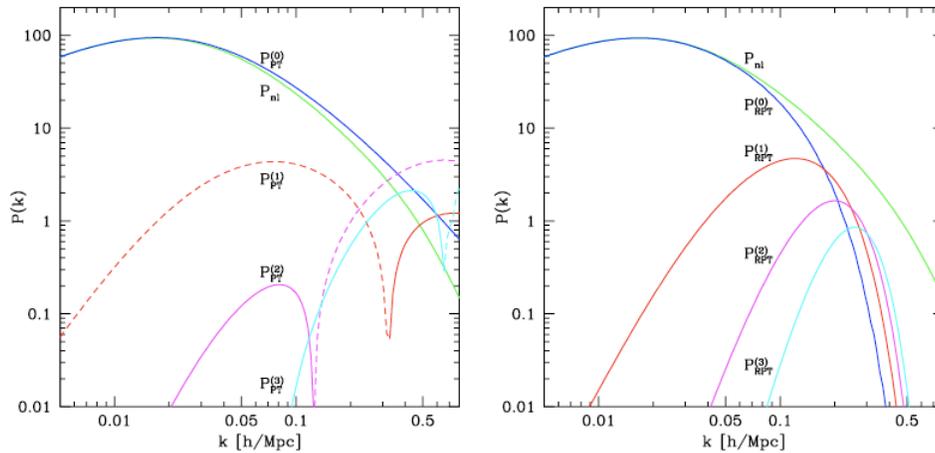


Figure II.14: The $P_{PT}^{(1)}$ ($P_{RPT}^{(1)}$) denotes the 1-loop correction in PT (RPT). Dashed lines denotes negatives contributions and solid lines positive ones. From (Crocco & Scoccimarro, 2006)

Renormalized Perturbation Theory (RPT)

Renormalized Perturbation Theory (RPT) reorganizes the expansions of SPT in terms of a quantity called the non linear propagator $G(k, z)$. In terms of this quantity a partial re-summation is done and the convergence is dramatically improved.

This rearrangement in summation terms originates the following power spectrum in RPT:

$$P(k, z) = G^2(k, z)P_0(k) + P_{MC}(k, z) \quad (\text{II.21})$$

where all contributions proportional to $P_0(k)$ (original power spectrum) are included in the first term. The propagator G describes the evolution of a mode k taking into account all the Fourier modes (all scales), the second term P_{MC} represents the power generated by mode coupling.

Advantages and Disadvantages

- RPT provides a physical understanding of different contributions. In the Fig.[II.15] the 2 ingredients of RPT are plotted, where is shown the smoothing of the oscillations due to de-correlation of Fourier modes in G^2P_0 , and the newly generated power of mode coupling predominant at small scales. The mode coupling power shows a strong scale dependence.

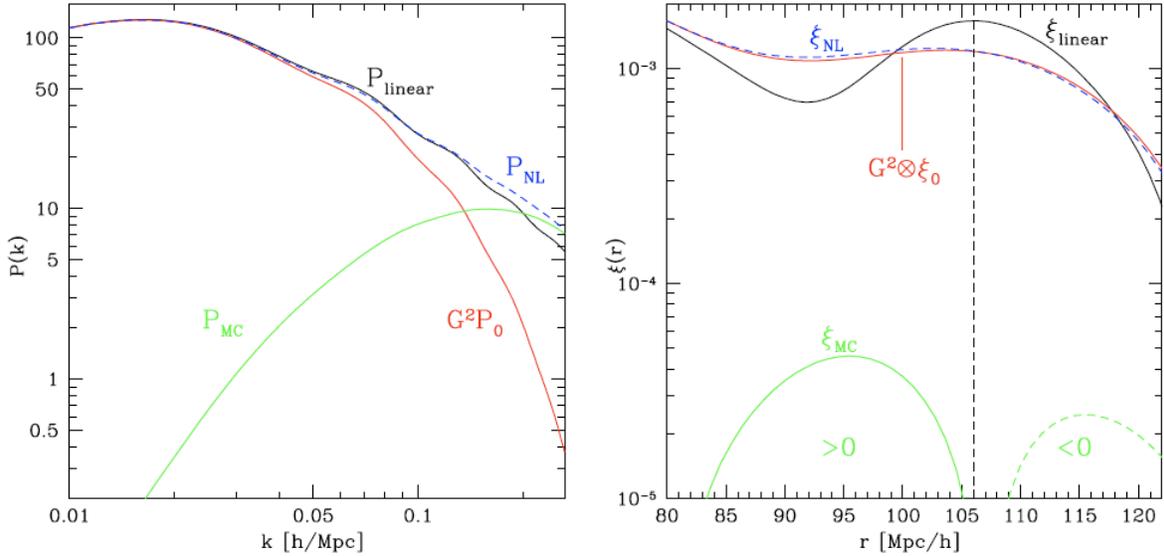


Figure II.15: From (Crocco & Scoccimarro, 2008), in left panel the power spectrum and in right panel the two-point correlation function described by RPT at $z=0$. The two ingredients of RPT are plotted: de-correlation of Fourier modes in G^2P_0 (in red), newly generated power of mode coupling (in green), and the final non linear power spectrum in blue. In black the linear power spectrum.

- In addition, it is possible to derive an expression for the non linear propagator in the fully non linear regime. ie. high- k limit ($k\sigma_b D \gg 1$):

$$G(k, z) \approx D(z) \exp\left[-\frac{1}{2} k^2 \sigma_v^2 (D(z) - 1)^2\right] \quad (\text{II.22})$$

Using this expression we can estimate the validity of linear perturbation theory. The characteristic scale determining the breakdown of linear perturbation theory is thus given by:

$$\sigma_v^2 \equiv \frac{1}{3} \int \frac{P_0(q)}{d^3q} \quad (\text{II.23})$$

One of the conclusions of RPT is that the validity of RPT is much limited than commonly thought. Another important result from RPT is the evidence of the shift generated by non linear behavior. In Fig.[II.16] I show the 2 point correlation function in real space. We observe this shift in the peak location towards smaller scales. This shift comes from two sources: first the convolution with G^2 generates a shift in the acoustic peak due to break of symmetry of acoustic peak in linear regime, another source of shift is the mode coupling contributions describing coherent infall velocities¹⁰

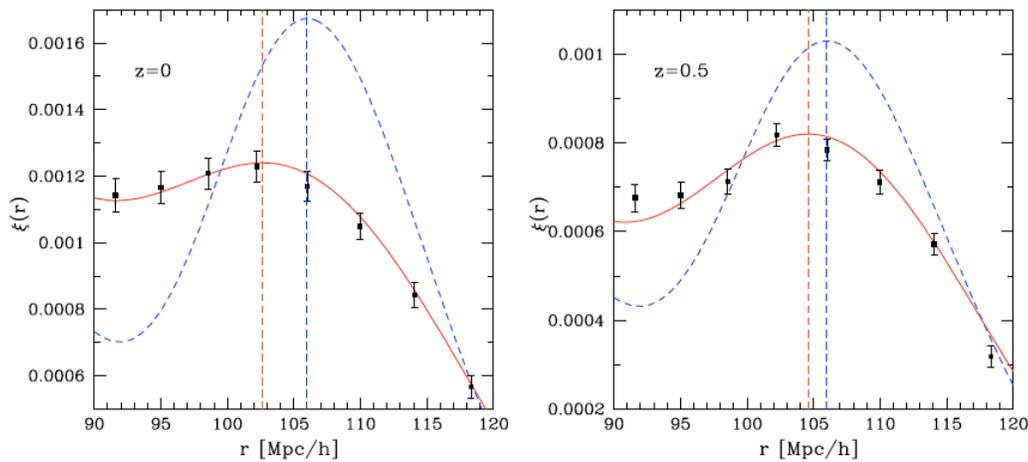


Figure II.16: The 2 point correlation function in real space, in blue the linear one in red the non linear. The points corresponds to results from N-body simulations. We observe this shift in the peak location towards smaller scales. From (Crocco & Scoccimarro, 2008)

- Renormalized perturbation theory provides a good description of two point statistics in real space. The results are in agreement with N-body simulations (see Fig. II.16]. Nowadays this formalism do not include the redshift distortions and bias

Lagrangian Perturbation Theory

In the lagrangian picture the fundamental variable is the displacement field Ψ , that maps the original position q of a particle to the eulerian coordinate x :

$$x(q, t) = q + \psi(q, t). \quad (\text{II.24})$$

¹⁰The mode coupling contribution contains oscillations out of phase with the acoustic oscillations in the linear power spectrum. This comes from the fact that non linear corrections are a decreasing function of local spectral index, thus maximal corrections occurs at wave numbers where derivative of acoustic oscillations is most negative.

In the lagrangian picture the displacement field is expanded in series:

$$\Psi = \Psi^{(1)} + \Psi^{(2)} + \Psi^{(3)} + \dots \quad (\text{II.25})$$

where $\Psi^{(N)} \equiv (\Psi^{(1)})^N$, is considered small quantity. The displacements are related to the overdensities, by the following expression in Fourier space derived from the continuity equation¹¹;

$$\delta(k) = \int d^3q (e^{-ik \cdot \Psi(q)} - 1) \quad (\text{II.26})$$

Advantages and Disadvantages

- LPT provides a good description of two point statistics. Lagrangian picture is intrinsically non linear in the density fields and a small perturbation in density fields carries a considerable amount of non linear information. Second order Lagrangian PT (2LPT) provide a remarkable improvement over linear order. In the Fig.[II.17] the correlation function from LPT (green) is shown compared with simulations (black) and linear theory (red). The LPT approach coincides well with simulations.
- Redshift distortions effects could be easily included in the Lagrangian approach . SPT deals with RD only in the linear regime, it does not reproduce non linear effects in small scales (fingers of god), commonly the fingers of god effect must be introduced by hand including a gaussian damping factor. In contrast in LPT the redshift distortions could be included in a natural way and this damping factor is derives naturally. The power spectrum computed in linear case (Kaiser formula) is compared with LPT model and the phenomenological model (Scoccimarro, 2004) with simulations are shown in the Fig.[II.18] , the power spectrum calculated with LPT describes the results from simulations and behaves better than 1-loop SPT.
- Another advantage is that this approach could be expanded adding biasing prescriptions , this analysis have been done by (Matsubara, 2008). The nature of bias included in this approach is lagrangian, as we have already said it depends on the initial density and velocity fields.

Comparison of perturbation theory approaches

I finish this section showing a figure from (Carlson et al., 2009) where a comparison between different perturbation approaches is done with N-body simulations and giving some final remarks. In (Carlson et al., 2009), the power spectrum was computed using different perturbations approaches at two redshifts and for different cosmologies. In Figs. [II.19] I show only the results for a Λ CDM cosmology¹². These figures show some trends that I have exposed along section even if not all the models appearing in the figure were described in the section. In Fig.[II.19] we observe in left panels the power spectrum from RPT¹³ for 1-loop and 2-loops at redshift $z = 0$ (top) and $z = 1$ (bottom).

¹¹From the continuity equation that relates the eulerian and lagrangian coordinates we have that $\rho(x)d^3x = \bar{\rho}d^3q$ where ρ is the mean density, this relation traduces in fourier space to $\delta(x) = \int d^3q \delta^3[x - q - \psi(q)] - 1$ where $\delta = \rho/\bar{\rho} - 1$

¹²($\Omega_M = 0.25$, $\Omega_b h^2 = 0.0224$, $h=0.72$, $n=0.97$ and $\sigma_8 = 1$)

¹³closure theory for 1-loop and 2-loops at redshift $z = 0$ (top) and $z = 1$ (bottom)

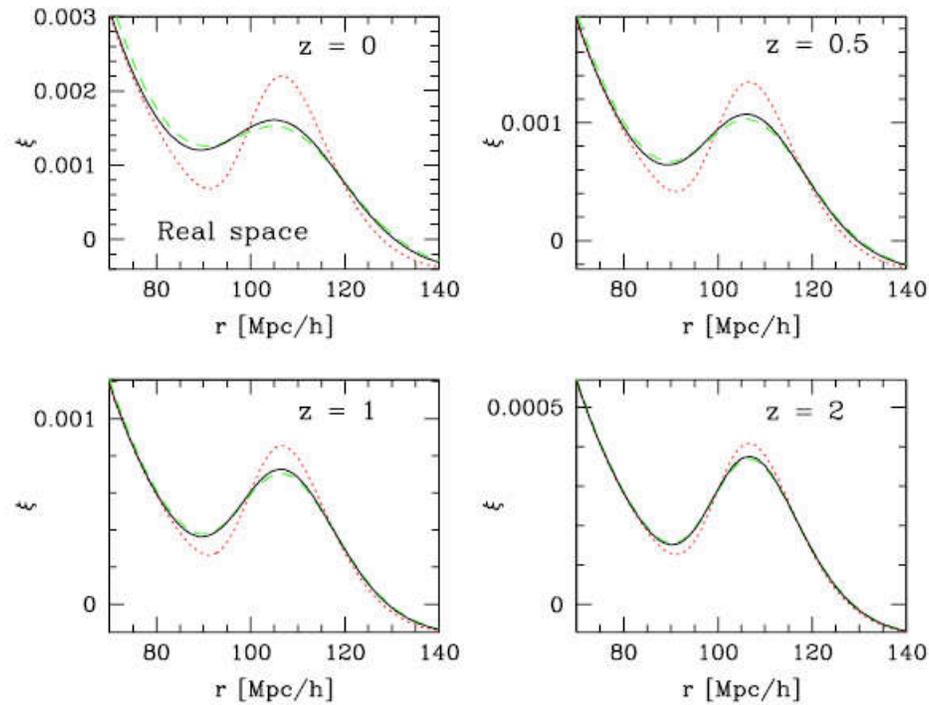


Figure II.17: The correlation function is shown for various redshifts obtained with simulations (black) and linear theory (red) and the gaussian filtered linear theory power spectrum (green). From (Matsubara, 2008)

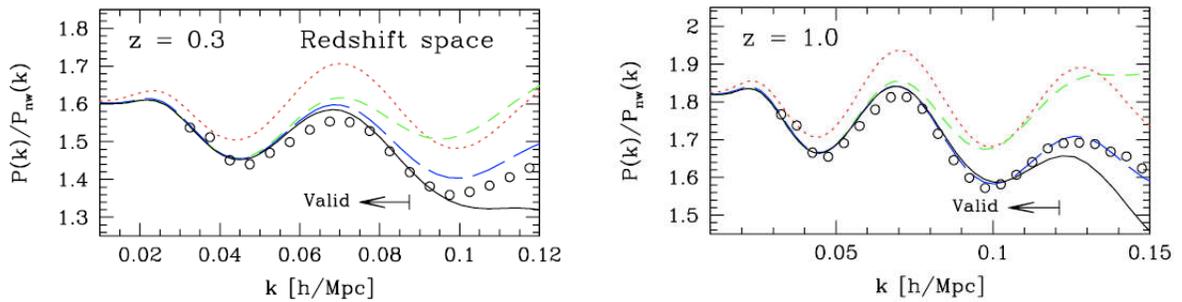


Figure II.18: The wiggle part of the power spectrum from N-body-simulations in open circles, in red dotted line the linear theory power spectrum, in black LPT power spectrum, in green dashed line 1-loop SPT and in blue dashed line phenomenological model of the redshift power spectrum (Scoccimarro, 2004). From (Matsubara, 2008)

In right panels the power spectrum from lagrangian perturbation theory, Large-N theory, time-RG theory, RGTP theory is shown at redshift $z = 0$ (top) and $z = 1$ (bottom).

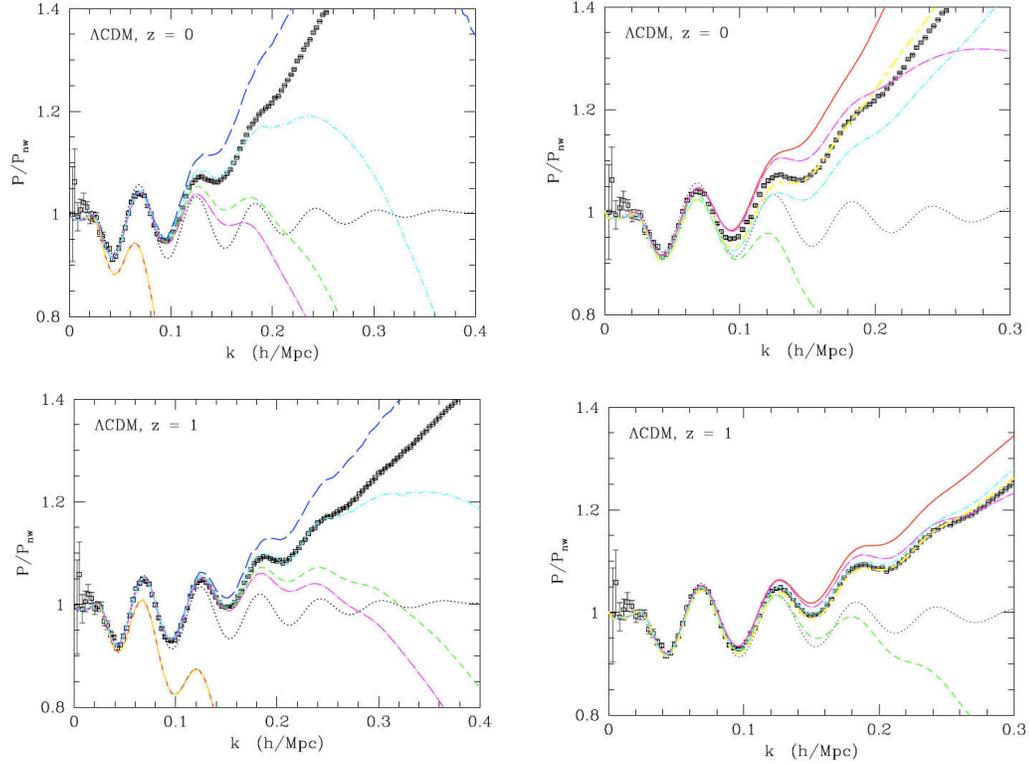


Figure II.19: In left panels the power spectrum from RPT (1-loop in green and 2-loop in blue) and closure theory for 1-loop (magenta) and 2-loops (cyan) at redshift $z = 0$ (top) and $z = 1$ (bottom). In right panels the power spectrum from Lagrangian perturbation theory (green), Large-N theory (magenta) (Valageas, 2007), time-RG theory (yellow) (Pietroni, 2008), RGTP theory (cyan) (McDonald, 2007) is shown at redshift $z = 0$ (top) and $z = 1$ (bottom). The black dotted line is the linear theory. From (Carlson et al., 2009)

We observe that the different perturbations approaches are in well agreement with N-body simulations at large scales and part of the quasi linear regime. All the approaches give a better prediction for power spectrum than linear regime, but at sufficient small scales all of them fail. None of the methods appear to be accurate beyond $k \sim 0.1 hMpc^{-1}$ at $z=0$. As is expected the methods perform better at smaller scales and at higher redshift. We observe that going to higher orders not always guaranty to achieve a better prediction.

II.4.2 Reconstruction

In recent years another method to deal with non linear degradation of baryonic feature has been proposed by Eisenstein et al (Eisenstein et al., 2007). The idea of reconstruction is that the degradation of the peak of BAO is a correctable process. We have seen that non linear gravitational structure formation and redshifts distortions move galaxies away from original locations, blurring the peak (or erasing higher harmonics in power spectrum). The main argument is that blurring effect is

largely due to bulk flows and super cluster formation effects generated by gravitational forces, thus running backward the gravitational flow will restore the acoustic peak to its linear regime shape.

Bulk flows are generated by density field, and the connection between velocity field and density fields at that scales are almost linear, thus predicting the velocity field and remove it. This approach is based on LPT. As we have seen the linear theory motion in LPT is $\nabla \cdot q = -\delta$. Thus moving the particles $-q$. To introduce the redshift distortions the same procedure is followed but in redshift space and this takes into account linear redshift distortions. It is important to said that the idea is not to move all particles back rather is to move densities back to their initial position.

Algorithm of Reconstruction

- Generate a smooth density field to filter high harmonics that are difficult to model.
- Compute the negative Zeldovich displacement q from the smoothed density field.
- Then to shift the original particles by s and compute displaced density field δ_d .
- Shift an initially uniform distributed particles by s to form the shifted density field δ_s .
- The reconstructed density field is defined as $\delta_r = \delta_d - \delta_s$. The power spectrum is $P_r(k) \propto \langle |\delta_r^2| \rangle$.

The reconstruction cannot fully cancel the non linearities because the correction to $\delta^{(n)}$ implies only first order. Thus it generates a density field with second order corrections not the the linear density field. The algorithm is inherently non linear and an effort to understand analytically has been done in term of LPT (Padmanabhan et al., 2009) (Noh et al., 2009). The method of reconstruction has been tested by N-body simulations following Λ CDM model in (Noh et al., 2009). In Fig.[II.20] I show the results from (Noh et al., 2009) for the correlation functions. In solid line the non linear correlation function from simulations at $z=0$. In dotted line the reconstruction correlation function using a smoothing of $R = 5h^{-1}Mpc$ and in short dashed line $R = 10h^{-1}Mpc$. In long dashed the initial correlation function before the non linear evolution. As we can observed from the figure the non-linear evolution has partially washed out the peak in the matter correlation function and applying reconstruction restores much of the original signal.

It is possible to extend the results to biased tracers. LPT provides a good framework for thinking in reconstruction. It explains in a natural way how reconstruction works. Reconstruction does not generate the initial power spectrum but it serves to sharpen the peak and reduce the change in the peak position generated by mode coupling¹⁴.

II.4.3 Halo Model

Another route to understand nonlinear gravitational clustering is done by the Halo Formalism (Cooray & Sheth, 2002), this semi analytical tool enables to understand statistical properties of

¹⁴The peak location comes from the second order corrections that generates an the out of phase component. The structure of the mode coupling terms is modified in such a way as to reduce the amplitude of the out of phase contribution.

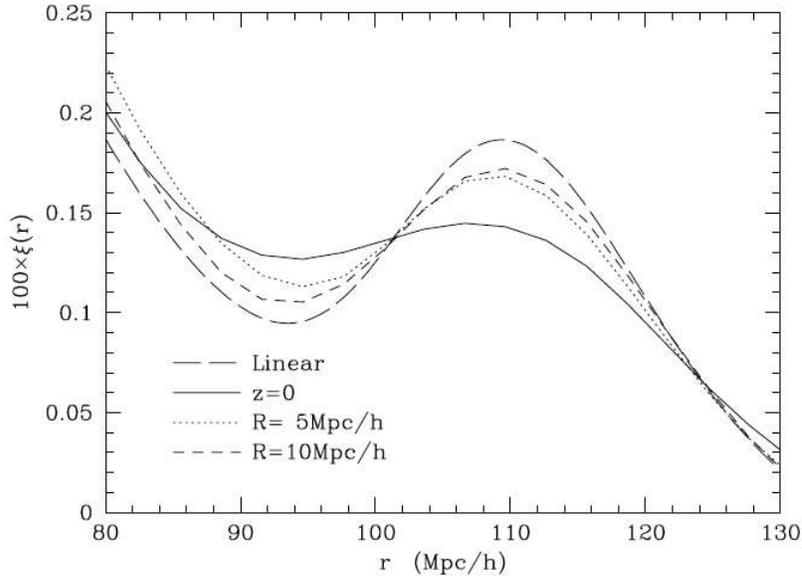


Figure II.20: The solid line corresponds to the non linear correlation function at $z=0$. In dotted line the reconstruction correlation function using a smoothing of $R = 5h^{-1}Mpc$ and in short dashed line $R = 10h^{-1}Mpc$. In long dashed the initial correlation function before the non linear evolution. From (Noh et al., 2009)

large scale density and velocity field and at the same time provides a natural way to model the bias. The Halo model is based on the assumption that matter is located in virialized halos of different mass, as a consequence the power spectrum of matter could be separated in two contributions: one coming from the correlation of matter within the halo called 1-halo term, and the contribution coming from the matter belonging to two different halos, known as the 2-halo term. The size of halos being smaller than the typical distances between halos enables that statistic properties could be separated. Thus the specific distribution within halos do not affect large scale structures, only the spatial distribution of halos. Inversely statistics of the mass density field on small scales are determined by the spatial distribution within the halos, the precise way in which the halos themselves may be organized into large scale structures is not important. This separation permit to study distribution of mass in two steps. Distribution of halos and spatial distribution within halos. A particular assumption is that the physics not described by perturbations theory is confined within halos, and that halos can be approximated as being in virial equilibrium. Thus the halo model permits to separate the spatial statistics and the physics involved.

The halo description requires the knowledge of three ingredients: the distribution of halos as a function of size known as abundance, the spatial distribution within each halo known as mass profile, and the spatial distribution of halos (clustering). These elements are related to the initial density field. I refer to the reader to (Cooray & Sheth, 2002), for a detailed description of the basic equations of this model and the common prescriptions used in the literature for the different ingredients that goes beyond the scope of this section. The halo model provides a physically mo-

tivated means of estimating the two-point statistics of the dark matter density field. However, it has several limitations originated from the simplifying assumptions related to the prescriptions for the abundance and mass profile. To finish I show some results obtained from this formalism. In Fig.[II.21] the power spectrum of the dark matter density field at the present time ($z = 0$) is shown in solid line. The dotted and short dashed lines show the contributions to the power from the single and two halo terms. Their sum corresponds to the solid line. In dashed line the power spectrum measured in numerical simulations (labeled “PD”). The figure shows the agreement between the predictions of the model and the simulations.

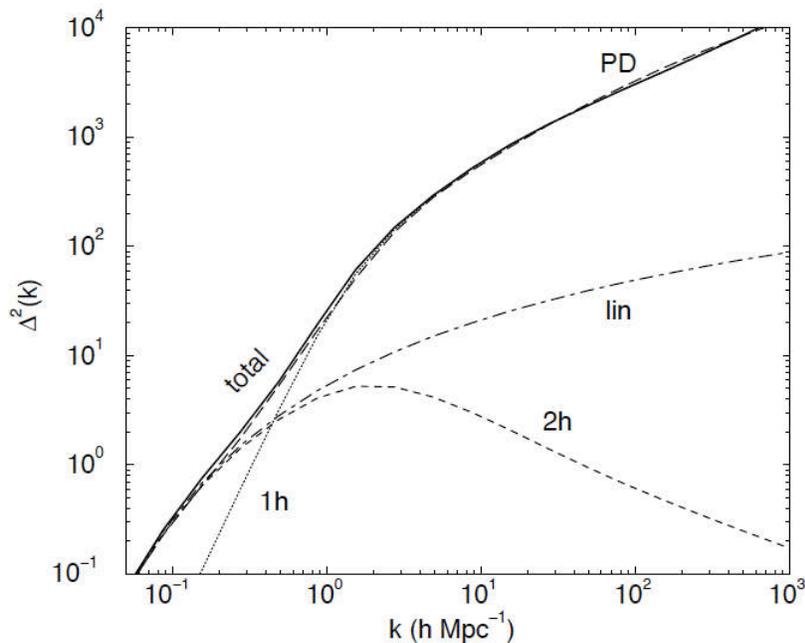


Figure II.21: The power spectrum of the dark matter density field at the present time ($z = 0$) is shown in solid line. The dotted and short dashed lines show the contributions to the power from the single and two halo terms. Their sum corresponds to the solid line. In dashed line the power spectrum measured in numerical simulations. In From (Cooray & Sheth, 2002)

II.5 State-of-the-art

To conclude the chapter in Table. II.1 I give a non exhaustive list of the measurements to date of the baryonic acoustic feature. This summary serves to show the current state of the measurements and their precision. From the table we observe the context in which the Baryonic Oscillations Sky Survey emerged.

Table II.1: Measurements of Baryonic Acoustic Feature.

Reference	Type Data	Survey	# Galaxies	V (Gpc)	err%	z
Eisenstein 2005	3D	SDSS	46,000	0.72	4	spec
Cole 2005	3D	2dFGRS	221,000	0.2	5	spec
Hutsi 2005	3D	SDSS	46000	0.72	4	spec
Padmanabhan 2007	2D	SDSS	600,000	1.5	6	photo
Blake 2007	2D	SDSS	600,000	1.5	6	photo
Blake 2011	3D	WIGGLEZ	200000	1	4	spec

III. From Sloan Digital Sky Survey to Baryonic Oscillation Spectroscopic Survey (SDSS-III/BOSS)

This chapter is dedicated to the Baryonic Oscillations Spectroscopic Survey that gives rise the data used along this thesis to analyze the large scale structure of the universe. This survey is one of the four programs composing the third version of the Sloan Digital Sky Survey (SDSS-III). Recognizing the great impact that Sloan Digital Sky Survey has made in the history of astronomy I devote the first part of the chapter to give a brief historical review of the Sloan Digital Sky Survey. I begin describing the original project SDSS I (2000-2005), then I review the second version SDSS II (2005-2008) emphasizing the evolution on the science "design goals" and I finish this section giving a summary of the scientific achievements in the whole period. The second part is devoted to the third version of SDSS-III (2008-2014). I describe the evolution of project trying to explain the motivation for doing a third version. I present the four programs forming this third phase focusing on the science goals. Finally in third part I concentrate in the Baryonic Spectroscopic Survey, I give an overview of the survey, I present the science goals, I describe the instruments, the data reduction process, the coverage and schedule and finally a section to explain the target selection algorithm. The idea of this part is to provide to the reader a synthesis of the important information of the survey and data reduction, essential for the data analysis.

III.1 Sloan Digital Sky Survey

The Sloan Digital Sky Survey is considered to be one of the most influential astronomical surveys in the last years evaluated by the number of scientific articles that it supported and the number of references to them. By 2008, to give an idea of the productivity, the number of SDSS-data based papers reached 1500, 16 of them were within the 100 most cited papers in the past five years. This characteristic places SDSS as one of the most productive astronomical observatory in this decade (Madrid & Macchetto, 2009).

The Sloan Digital Sky Survey accomplished in 2010, 10 years of operations in its different

versions: SDSS-I between 2000-2005, SDSS-II 2005-2008 and now in its third edition, SDSS-III is planned to operate between July 2008 until July 2014. In this section the original project of this survey would be presented.

III.1.1 Original Project

The Sloan Digital Sky Survey (York et al., 2000), (Stoughton et al., 2002), (Gunn & Weinberg, 1995) born in 1978 when a consortium of astronomers gathered to design the next generation survey with aim of providing to the scientific community a data base suitable to do large scale structure analysis. In that time the lack of a uniform galaxy sample in wide area of sky prevented to do statistical analysis, the galaxy surveys available consisted of ≈ 20000 galaxies and covers about 700 deg^2 (Las Campanas Redshift Survey (LCRS)(Shectman et al., 1996)), in consequence even the lowest order of the statistical properties of distribution of galaxies at large scale were unknown. Large scale structure analysis requires a database with high density of objects, so an increase in number of galaxies and volume surveyed was necessary. In that context SDSS set out to increase two orders of magnitude over the largest surveys existing in 1995. SDSS proposed to cover the largest fraction of sky for a digital photo and spectrometric survey at that time. The purpose was set to take spectra of 5×10^7 and 10^6 QSO, spectra with signal to noise, resolution and wavelength coverage until that moment only available for small samples.

Performing this uniform spectroscopic sample required a photometric survey with quality and uniformity of photo data. These requirements since the beginning profiled the project to the design of a dedicated telescope destined to generate a photometric survey that serves as the input data to the target selection algorithms guarantying the uniformity and the redshift completeness of the spectroscopy sample.

One crucial factor that enabled the SDSS project was the digitalization of photometric data. At this moment it was recognized that the CCD had better sensitivity, uniformity and linear behavior, but for large area due to technologic limitations it was preferred the photographic plates. The technological advances in solid state detectors and the increase in computational processing and instrument control capabilities enabled to figure out the first large scale imaging survey using CCD detectors. In this sense SDSS marked the beginning of the era of large area digital surveys.

The strategy of observation was defined to do photometry and spectrometry simultaneously to optimize the strategy of scanning. The plan was to take photometry in the best conditions of seeing, sky brightness and sky transparency, and spectroscopy to be done the rest of the time with less good conditions. This decision was motivated based on the more consuming time of spectrometry together with the fact that time available in optimal conditions is less frequent. To accomplish this program in the telescope design it was considered to make easy the change of instruments (camera et spectrograph).

The design converges to a special purpose 2.5 m telescope (Gunn et al., 2006) in Apache Point Observatory (APO), New Mexico with a wide angle 3° diameter. For photometry a camera (Gunn et al., 1998) with a large focal plane of a 30 CCD array of 2048 pixels. The CCDs arranged in six columns and 5 rows corresponding to the 5 wavebands filters. An additional array of 24 more CCDs serving to astrometric calibration. The telescope would operate in drift scan mode, permitting almost simultaneously image in five wavebands (Fukugita et al., 1996). An auxiliary 0.7

m telescope with its respective CCD camera would monitor stars for calibration. And a multi-fiber spectrograph with two arms, would take 640 spectra at a time. Plates would be manually drilled with holes corresponding to the target coordinates. This manual plugged plate system was chosen for the spectrograph considering that it accomplish the requirements with less cost than building a robot to automatize the task.

The coverage of the survey area was chosen to minimize the galactic extinction, sky brightness and differential refraction, optimizing observing season. The survey area consisted of a slightly elliptical region centered in the north galactic pole, and when the north galactic cap was inaccessible, three stripes in the southern would be scanned repeatedly. These stripes gave place when coadded to observe objects 2 magnitudes fainter opening the possibility of study variable objects. In Fig.[III.1], the stripes defining the photometric survey are shown and the extinction maps of Schlegel that determines the survey area (Schlegel et al., 1998).

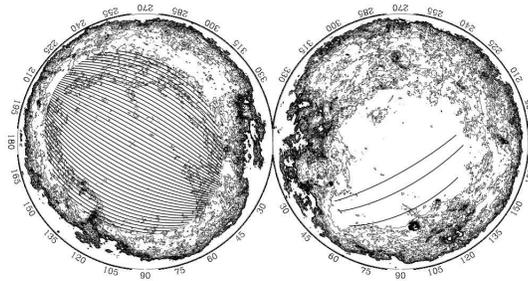


Figure III.1: Schlegel Extinction Maps and Stripes of SDSS. From (Schlegel et al., 1998).

The Sloan Digital Sky Survey in its first version successfully accomplished the science requirements fixed at the science project. It legged a catalogue of $\sim 7 \times 10^5$ spectra of galaxies, 9×10^4 spectra of quasars and 2×10^5 spectra of stars, and a photometric survey of a contiguous region of sky covering $\sim 8000 \text{ deg}^2$ toward the Northern Galactic Cap, and 700 deg^2 in three stripes in the South Galactic Cap. The Fig.[III.2] shows the coverage of SDSS. This powerful database traduced in a unprecedented scientific productivity. The scientific researches went beyond the science goals predicted at the initial project. The next section will give a summary of the results obtained with SDSS within the first two phases.

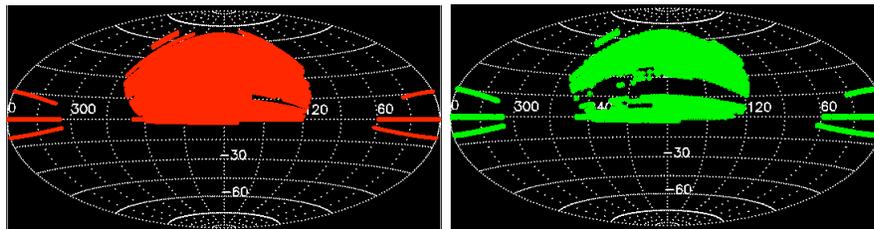


Figure III.2: Data Release 5 Coverage, in red footprint of photometric survey in green footprint of spectroscopic survey. From <http://www.sdss.org/>.

III.1.2 SDSS II

Based on the scientific achievements and taking advantage of the potential of the facilities that promised continuous increasing the power of the database, before the end of operations of SDSS , a second version begin to cooking. It used the facilities of SDSS, the telescope, camera and spectroscopy. This second phase denominated SDSS-II designates a set of 3 programs: Legacy Survey, Supernova Survey and SEGUE. The Legacy survey aimed to complete SDSS initial photometric and spectroscopic survey, the supernovae program was dedicated to complete the Hubble diagram in the low and intermediate redshift interval and the SEGUE was devoted to explore the stellar populations and dynamics in the Milky Way. The beginning of operations was august 2005 and it was planned to continue in operation until july 2008.

Legacy Survey

This survey continued the SDSS initial project, its aim was to complete the spectroscopic and photometric survey in the footprint of initial project, covering uniformly the survey area using the same target selection algorithm that SDSS. The imaging coverage of the North Galactic Cap had been almost completed as of the end of SDSS. The last Legacy imaging data was obtained the second semester of 2006. The spectroscopic survey of SDSS covered only 5700 deg² of the imaging survey, to cover the rest of the imaging area 524 new plates were used including only the North Galactic Cap. The coverage was finished in May 2008. Cumulatively between SDSS and SDSS-II, the total area surveyed was 7646 square degrees in the North Galactic Cap, compared to the goal of 7700 square degrees. In total as part of Legacy survey a set of 250,000 new objects were classified as galaxies, and 28,000 new objects were classified as quasars. The total spectroscopic footprint for Legacy, adding SDSS and SDSS-II, is 8032 square degrees. Of this area, about 750 square degrees are in the South Galactic Cap.

Supernova Survey

The supernovae is a recognized technique to measure the distance redshift relation using standard candles whose luminosities are calibrated by objects in the local universe. At that moment in the Hubble diagram there was a gap in the redshift interval ($0.1 \leq z \leq 0.3$). Only 6 of the 157 high quality SN Ia light curves fell in this interval. Furthermore, in the low redshift range the measurements came from different surveys, implying different systematics. The goal of supernovae survey was to address this gap in redshift coverage. The Survey had as objective to achieve a catalogue of 200 intermediate redshift ($0.1 < z < 0.4$) type Ia supernovae. Another goal of Supernovae Survey was the minimization an evaluation of SN systematics. In SN surveys systematics errors are comparable with statistical errors, improved control on the sample was required to acquire better accuracy. In that sense the requirement imposed densely sampled multi-band SN light curves with well understood photo calibrated data.

The Supernova Survey carried out repeated imaging on Stripe 82 (southern equatorial stripe) covering a total of 300-square-degree. These observations took place during September through November in 2005, 2006, and 2007. The stripe was scanned in total 20 times in SDSS, and an additional 40 complete coverages of this area were obtained in SDSS-II. A total of 498 spectroscop-

ically confirmed Type Ia supernovae was found for the three-season survey. This number is beyond predictions. In combination with higher redshift samples (ESSENCE, SNLS), SDSSII SN Survey lead to more robust estimates of cosmological constraints.

SEGUE

From the photometric results from SDSS, the existence of substructures in the stellar halo in the Milky Way was evidenced . At that moment persisted the idea that our galaxy consisted of an axially symmetric spheroid with a smooth density profile. The discovery of this substructure motivated the design of a photometric and spectroscopic survey permitting to study the structure of our galaxy. In that context appeared in the scenary the Sloan Extension for Galactic Understanding and Exploration (SEGUE) which science goal was understanding the galaxy structure and testing models of galaxy formation and evolution. It consisted of a imaging and spectroscopic survey of the Milky Way and its surrounding halo.

It would expand the photometric survey of SDSS, adding 3500 deg², with the aim of sampling all the visible sky from Apache Point Observatory. The galaxy extinction restricts the coverage to regions not close to the bulb. Nowadays, the coverage of this survey enabled to study the transition between stellar halo and galaxy disk. The survey was designed to take 200 pencil beams spaced around sky, arranged to sample the sky at intervals of 15° . Two observations of each pencil beam was done.

It aimed to map stellar populations and their kinematics over large volume of the galaxy covering distances from inner halo out to galactocentric distances. The targets covered a variety of ages and metallicities, permitting to enable a whole range of astrophysics projects. In solar neighborhood the sample includes stars M, L, and T dwarfs BHB and K/M giant stars. A subset of targets are devoted to unusual stars based in their colors and velocities like low metallicity or high velocity stars.

III.1.3 Legacy of SDSS I and II

The combined results of the SDSS I-II (Final Report SDSS-II (Richard et al., 2008)) covered a imaging survey in five optical bands of 8400 deg² in total, the most part of the survey area (7646 deg²) is at the northern galactic cap (NGC) and it has 3 equatorial stripes in southern galactic cap (777 deg²). The spectroscopic survey covers a total of 8032 deg² and contains spectra of 900,000 galaxies, 100,000 quasars and 400,00 stars. It also comprehends a supernova survey of 500 Type Ia supernovas.

This powerful database was gradually released to public since the first years of operation of SDSS. Over the eight years of operation of SDSS, seven data released were done almost one per year, for the 5 years period of SDSS the final data for release was DR5 and for SDSS-II the final data release version was DR7 (summer 2008) covering the whole period.

SDSS-I-II database have supported a enormous quantity of researches in a broad spectra of astronomical disciplines many findings were part of the scientific goals but many others were completely unexpected . An analysis of researches shows that only half of the numbers of publications were

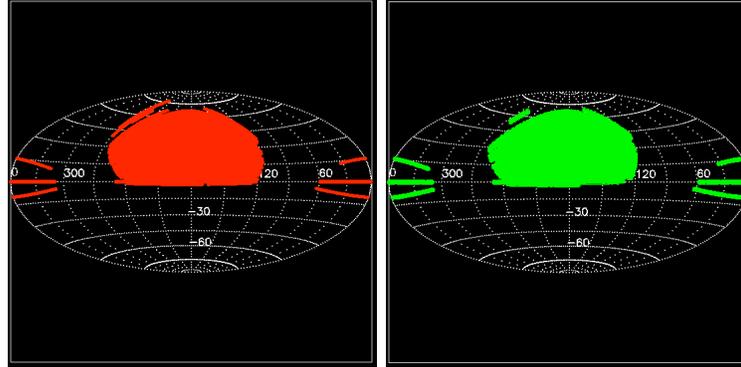


Figure III.3: Data Release 7 Coverage, in red footprint of photometric survey and in green footprint of spectroscopic survey. From <http://www.sdss.org/>

generated within the collaboration, showing that legacy of SDSS contributes to the fundamental work within all scientific community.

As non exhaustive list of scientific achievements during this first 8 years of operation is:

- Large Scale Structure: precision measurement of large scale clustering and cosmological constraints, detection of the BAO signatures in clustering, gravitational lensing.
- Galaxy Formation Clustering in intermediate scales, structure of stellar populations, substructures in the outer Milky Way, discovery of new companions of the Milky Way and Andromeda.
- Science with QSO. Precision measurements of luminosity distribution of QSO, measurement of clustering, precision measurements of early structure with Lyman-alpha forest, discovery of the most distant QSO
- Astrophysics: Properties of galaxies, stellar astrophysics, substellar objects, common origin of asteroid families, discovery of large populations of substellar objects, discovery of the stars escaping galaxy

The power of the SDSS database is far from been over, is evident that for the next years It will continue generating scientific results, and the SDSS will conserve this status as one of the most influential survey in the history of astronomy, only supered by the next generation wide surveys as LSST.

III.2 SDSS III

Basically two reasons have motivated a third phase of Sloan, the potential productivity of SDSS-database and the fact the SDSS facilities remain been a powerful resource for wide field spectroscopic surveys. At the moment when the project of a third version began to take form, the camera was still among the most powerful (if not unique), the technical infrastructure (telescope, instruments asset ,

etc) and the organization of the collaboration was a strength of SDSS that maintained competitive for doing wide field spectroscopy even if the development of the spectrographs for 3 programs were required.

The third version of Sloan, SDSS-III (Eisenstein et al 2010 (Eisenstein et al., 2011a), Proposal SDSS-III (Proposal, 2008)), called the Massive Spectroscopic surveys of the Distant Universe, the Milky Way Galaxy and the Extrasolar Planetary Systems, was proposed for 6 years. It is composed by a program of four spectroscopic surveys: the Sloan Extension for Galactic Understanding and Exploration in his second version (SEGUE-2), the Apache Point Observatory Galactic Evolution Experiment (APOGEE), the Multi-object APO Radial Velocity Exoplanet Large-area Survey (MARVELS) and the Baryon Oscillation Spectroscopic Survey (BOSS). The first two dedicated to explore the structure, dynamics and chemical evolution of our galaxy. the third one to study the population of giants planets around stars and finally the last one to study the dark energy and cosmological constraints. BOSS and SEGUE-2 are scheduled for dark time (moonless) while MARVELS and APOGEE for the bright time survey. These four programs will use SDSS infrastructure with minor changes but each program will use a different spectrograph, SEGUE is the only one that continued using original SDSS spectrograph.

III.2.1 SEGUE-II

Like its predecessor, SEGUE-II aimed of study the structure , dynamics and chemical composition of our Galaxy to contribute to the understanding of the galaxy formation models and stellar evolution.

SEGUE-II continue mapping of stelar populations with almost the same strategy and quality of its predecessor. It add 1317 deg² to the imaging survey and total of 140 000 spectra of a diversity of targets. As the first version the targets were selected from SDSS imaging along individual 7 deg² line-of-sight spread out over imaging survey. Nevertheless, a change in the observational strategy was done with the aim of maximizing the coverage of the deepest regions of galaxy. In SEGUE II only deep pointings along the same line-of sight were done, a difference of the first version where pointings were divided in two categories: short exposures of bright objets and deep pointings. The Fig.[III.4] shows the coverage of SEGUE-I and II.

The targets were basically the same: Halo main Sequence turnoff stars, blue horizontal branch stars, K-giants and M-giants stars, hight velocities and Hypervelocity stars, ultra-subdwarfs , ultra cool white dwarfs low metallicity stars. A exception of the ultra cool white dwarfs, all target selection algorithms were adjusted based of the lessons from SEGUE I to achieve a higher success rate or to achieve larger mean distances. The targets serve to different science goals. The Halo main Sequence turnoff stars, the blue horizontal branch stars, K-giants and M-giants stars were selected to study the kinematics and chemical structure of the outer galaxy. Halo high velocity stars and Hypervelocity stars were selected to study this ejection mechanism originated by interactions with the Galactics central black hole. The ultra-subdwarfs and ultra cool white dwarfs were selected for being the most metal poor cool stars in the neighborhood . Finally, the low metallicity candidates were chosen as future candidates for high resolution spectroscopy to probe nucleosynthesis in the first generation of metal poor stars.

The first year of SDSS-III, all dark time was dedicated to survey because of the upgrading of BOSS spectrograph. The spectrograph used in all version of SEGUE was the original one of SDSS

with 640 fibers, spectral range of $3800 < \lambda < 9200$ and a resolution $R = \lambda / \Delta\lambda = 1800$. The final number of spectra measured was 118,151, and the exposures are accumulated until a fix ($S/N \approx 10$ per pixel) threshold at $r=19.5$ as BOSS does.

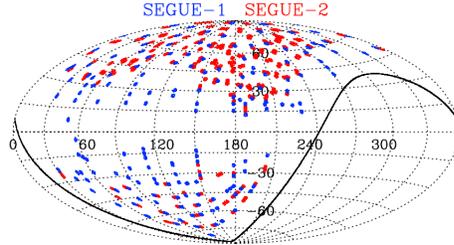


Figure III.4: Coverage of SEGUE-I and SEGUE-II. From (Eisenstein et al., 2011b)

III.2.2 APOGEE

The Apache Point Observatory Galactic Evolution Experiment (APOGEE) is a survey designated to cover the lack of a systematic survey that maps the Milky way. The principal obstacle to realize this kind of survey is the presence of intergalactic dust that obscures principally the optical band. Until present the studies of the structure and dynamics of our galaxy are based in different surveys with a mix of tracers and techniques that generates uncertainties. The program APOGEE emerges as the first large scale systematic spectroscopic survey. This survey is focused on getting the chemistry composition and dynamical properties of *all* populations of stars in our galaxy. In a certain way it is designed to complement and extend SEGUE. Nevertheless, it goes beyond capabilities of this survey, because of his high resolution to get abundances and high signal to noise.

Regarding infrastructure it uses the telescope and camera original from SDSS, and the same plug plate system as BOSS. It uses an infrared 300 fiber cryogenic spectrograph operating in near infrared (NIR) H - band ($1.6 \mu m$) where extinction is lower. The resolution of the spectrograph is of $R = \lambda / \Delta\lambda \approx 30000$ and (S/N) of 100 per pixel (compared with SEGUE $R = \lambda / \Delta\lambda = 1800$, and $S/N \approx 10$ per pixel). The high resolution and high signal-to-noise enables to get abundances of numerous chemical species. The H-band contains lines of molecules OH, CN and CO giving place a measurements the abundances of H, O and N that are the more abundant elements in the universe produce by SN II. H-band posseses also lines O, Mg, Si, S, Ca and Ti elements produced in the explosions of short lived stars, their abundances reflect the shape of initial masse function. Finally H-band contains elements in the iron peak (Cr V Mn Fe Co and Ni) produced in supernovae type Ia longer lived lower mass stars. It is designed to provide precise radial velocities across the galactic budge, disk. The combination between precision velocities and abundances patterns will enable to test models of galaxy formation

APOGEE plans to take spectra for 10^5 late type evolved star. The photometric database that supports the target selection is taken from 2MASS Survey, the selection output is a uniform sample of giant stars to a flux limited of $H \approx 13.5$. This tracer is luminous and all found in all the ages assuring the uniformity of the survey. The resulting homogenous database will surpass by two orders

of magnitude the total number of high resolution high signal to noise stellar spectra even taken on all of the world telescopes. The Fig.[III.5] shows the planned coverage for APOGEE.

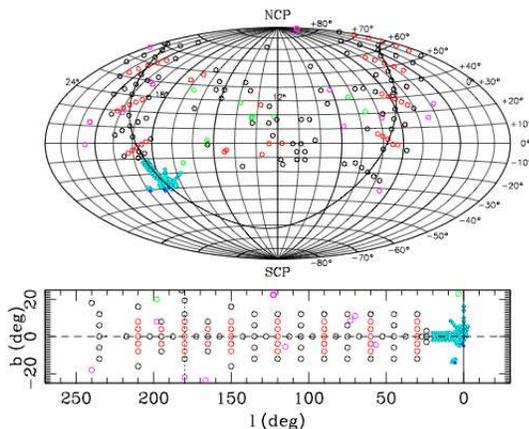


Figure III.5: Coverage of APOGEE. From (Eisenstein et al., 2011b).

III.2.3 MARVELS

The Multi-Object APO Radial Velocity Exoplanet Large-area Survey (MARVELS) is devoted to the formation and dynamical evolution of giant planetary systems. The scientific motivation behind MARVELS emerges from diversity of masses, semi major axes and eccentricities in extrasolar planets. Within the standard scenario of planetary system formation, the formation of giant planets like Jupiter is explained in terms of the core accretion scenario. In principle this kind of planets was generated in the region beyond the snow line in the protoplanetary disk (where ice is stable) by the collision of ice bodies achieving masses from 5-10 Earth masses. The subsequent accretion of gas give places to the final masses of the order of Jupiter mass. These planets are predicted to have nearly circular orbits with periods of the order of years. Nevertheless, giant planets have been found with periods of days and with highly eccentric orbits. The high eccentricities are explained in terms of dynamic interactions with the disk and the shorter periods behavior is explained in terms of migration models but is not clear which of these mechanism is predominant. This size of the surveys existing to date do not have the statistical power for studies of planetary system formation and dynamical evolution. (Over 15 years of observations around 200 planets have been found). MARVELS aims to provide a large sample of short to intermediate period of giant planets suitable for statistical analysis. It predicts the discovery of 150 new planets with 0.5 to 10 Jupiter masse.

It uses a infrared fiber fed interferometric spectrograph to measure radial velocities with a precision and cadence needed to detect giant planets with period lying between days to years. The detection of planets using doppler spectroscopy involves the observation of doppler shifts in the spectrum of the star. The fixed delay interferometry method consists in split the incoming beam and fed an interferometer type Michelson with a fixed optical delay. This uncorrelated fringes are separated by a disperser and recorded in a two dimensional detector. This configuration adds

fringes the interference in the stellar absorption lines. Once the delay of the interferometer is fixed, the shift in the fringes corresponds to doppler velocity shifts of a detected planet. The Fig.[III.6] shows a science exposure prevenient of MARVELS. This technique reduce photon noise by measuring multiple fringes in a broad band.

Marvels aims to survey of ~ 8400 bright stars (FGK stars ¹) in a apparent magnitude V8-12, visiting each star approximately 24 times over 4 years. The stars are selected from cross-matched combination of NOMAD (Zacharias et al., 2004) UCAC3 (Zacharias et al., 2010), GSCC2.3 (Lasker et al., 2008) and 2MASS (Skrutskie et al., 2006) catalogs.

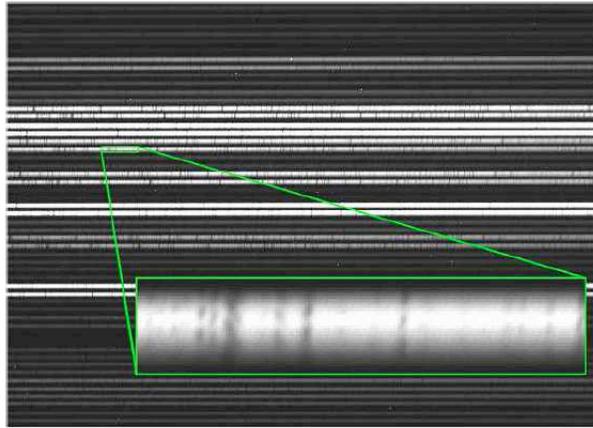


Figure III.6: The figure shows a science exposure from MARVELS. A zoom shows the region of the fringed spectrum of an object. The horizontal axis is the wavelength direction. Each spectral line produces a a modulated fridge pattern in vertical direction. Fitting the shifts in the vertical position of the fringes over all the spectral lines allows to extract precise measurements of radial velocities. From (Eisenstein et al., 2011b).

III.2.4 BOSS

The Baryon Oscillations Spectroscopy Survey (BOSS) as its name indicates, is an spectroscopic survey that uses the Baryonic Oscillation technique to study the expansion history of the universe. The expansion of universe is considered as one of the profound question in fundamental physics (DETF, Albrecht et al (Albrecht et al., 2006)). Four observational probes have been recognized to disentangle this mystery: Supernovae, BAO, Clusters and Weak lensing. All of these techniques repose over 2 approaches : a purely geometrical approach of studying the expansion history via standard patterns: *standard rulers* in the case of BAO or *standard candles* in the case of supernovae or investigating growth of structure over a broad range of redshift (Clusters and Weak Lensing). The BAO technique is among these techniques the least affected by systematics errors in the phase of experiments of current phase. BOSS aims to do precise measurements of this characteristic scale known as BAO distance. It was inspired in the previous work done in SDSS I-II that give place

¹Stars belonging to spectral types F, G, or K. Being most similar to the Sun (type G2), these are considered the most likely to be capable of supporting life on worlds that circle around them within their habitable zones.

to the first detection of BAO scale in 2005. This detection was done with SDSS and 2dF galaxy Redshift Survey (Cole et al., 2005), (Eisenstein et al., 2005) and yields a measure of the distance at $z \approx 0.275$ with 5 %. Using the last data release of SDSS, the measure of this distance at $z \approx 0.275$ was estimated with $\sim 2\%$ precision with the galaxy sample. BOSS aims to measure this distance, $d_A(z)$ and $H(z)$ at $z=0.35$ and $z \approx 0.6$ with $\approx 1\%$ precision . BOSS is in this sense the natural extension of SDSS I-II with proven hardware and acquired experience on the previous versions.

In terms of survey it planned to build a catalogue consisting of 1.5 million galaxies in the range of redshift $0.2 < z < 0.8$. This quantity will permit to achieve precision of 1 % in the measure of the absolute BAO scale separating the radial and transverse direction, giving place to constraints to Hubble parameter $H(z)$ and the diameter distance $d_A(z)$. It planes to innovate a new technique of BAO measurement using the Lyman-alpha forest of quasars as a tracer of matter. To explore this new technique it will devote 20 % of the fibers to quasar spectra. The science goal is fixed to 160 000 QSO in the redshift interval of $2.3 < z < 2.8$. The precision of the measure with quasars is predicted to be of the order of $\sim 4\%$.

III.3 Baryonic Oscillations Spectroscopic Survey in detail

III.3.1 Scientific Goal

As it has been explain in chapter 2, the oscillations of primordial plasma of baryon-photon previous to the combination era imprinted a characteristic scale in the matter distribution. This phenomena called Baryonic Acoustic Oscillation (BAO), appears as a bump in the 2-point correlation function of galaxies at a comoving scale corresponding to the sound horizon d_{BAO} . The value of this characteristic scale could be computed , in absolute units using straightforward physics and the tight constraints that Cosmic Microwave Background (CMB) measurements impose to the cosmological parameters.

The BAO technique consists in using this characteristic scale as a standard ruler to measure cosmological distances over a wide range of redshift, permitting to study the history of Cosmic expansion. Separating the measure of the BAO scale in the Line of Sight and the perpendicular directions, permits to constraint the Hubble parameter $H(z)$ and the angular distance $d_A(z)$ independently. The final goal of this measurements is to constraint the equation of the dark energy parametrized by w_0 and w_a ² and constraint the cosmological parameters H_0 and Ω_k .

BOSS could be seen as of two spectroscopic surveys that are executed simultaneously, the galaxy survey and the quasar survey. The galaxy survey was planned to extend the results of BAO obtained in previous versions with the well establish technique using the galaxies as tracers of the underlying matter. It consist of spectra of 1.5 million massive galaxies in a redshift range of $0.45 \leq z \leq 0.7$. The target selection is designed to give a uniform sample with a constant comoving density between $z=0.45$ and $z=0.6$ and then consisting of a magnitude limited sample up to $z=0.7$. It is the largest spectroscopic galaxy survey in effective volume and for BAO measurement at $z=0.6$. the error is only within a factor two of the cosmic variance limit. These characteristics make BOSS the most powerful galaxy survey for BAO measurements for the decade .

²Where the equation of dark energy is parametrized as $w(a) = w_0 + w_a(1 - a)$ with $a \equiv (1 + z)^{-1}$

The quasar survey will explore a new method of BAO measurement using the quasar spectra. It uses the Ly α forest as a tracer of structure at high redshift. The Lyman- α forest is the section in the quasar spectra between alpha emission line and beta emission line. When redshifted photons emitted by the quasar pass throughout the intergalactic medium they are redshifted and then absorbed by hydrogen gas producing a series of absorption line in the quasar spectra. This absorption spectra maps the distribution of matter along the line of sight direction. The quasar sample consist of 1.5×10^5 quasars in redshift range of $2.2 \leq z \leq 4$, this sample was designed to permit that Ly α forest absorption spectra be in the spectral range of the spectrographs. This set of quasars will sample a sparsely grid of light-of-sight permitting to test a large volume and giving place to a measure of cosmological distance at high redshift ($z=2.5$).

In the table III.1 the error estimation on the measurement of the BAO scale are shown for the case of galaxy of low and high redshift and for the case of QSO. The estimation for the case of QSO assumes 15 QSO for deg² over the whole survey area.

Table III.1: Error estimation for BAO measurement. From Project Description ³.

	% d_A	% H(z)
Galaxies z=0.35	1	1.8
Galaxies z=0.6	1	1.7
Ly α F z=2.5	4.5	2.6

The table III.2 shows the forecast of parameter constraints over the near future and current BAO experiments.

Table III.2: Forecast of parameter constraints over the near future and current BAO. From the Project Description⁴.

	h	Ω_K	w_0	w_p	w_a	FoM
BOSS LRG	0.008	0.0028	0.089	0.032	0.366	86
BOSS LRG +QSO	0.08	0.0019	0.076	0.029	0.279	122
WiggleZ	0.012	0.0028	0.099	0.035	0.430	66
HETDEX	0.015	0.0021	0.098	0.034	0.417	70
WF MOS	0.011	0.0017	0.083	0.033	0.323	95

III.3.2 Coverage and Schedule

BOSS plans to cover a total of 10 000 deg². The footprint in the NGC corresponds to the boundary defined in the original project: an elliptical region centered at ra = 15 h 20 m dec= +32 ° It covers a total area of 7646 deg² and corresponds to region least affected for galactic extinction, following the Schlegel maps. (Schlegel et al., 1998).

By contrary in the south galactic cap the survey original region covered only three stripes, one along the celestial equator (20h.7 to 4h, d=0), the other two north and south. It summed a total 777 deg². For BOSS the idea was to produce a compact and contiguous footprint in the SGC in

order to do that an additional 2500 deg^2 was imaged. The boundaries of this region was chosen trying to reach fields of CFHT Legacy Survey. The program was completed in January 2010. The Fig.[III.7] shows the footprint of BOSS and the plates observed during the first year of operation.

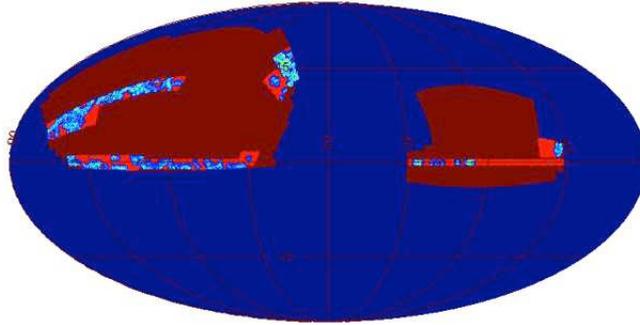


Figure III.7: BOSS Coverage. In color vino footprint of BOSS, the red regions represents tiled regions and in blue plates exposed. From (Eisenstein et al., 2011b).

BOSS plans to operate a total of 5 years using dark and grey time. Dark time is assumed to be 50% and grey time 10% of total time. It began operations in August 2009 and it will finish at the summer 2014. In Fig.[III.8] the SDSS schedule is presented, a data release is planned for each year of data. The first data release, DR8 was delivered in January 2010. It includes all the BOSS photometry and SEGUE-II spectroscopic data.

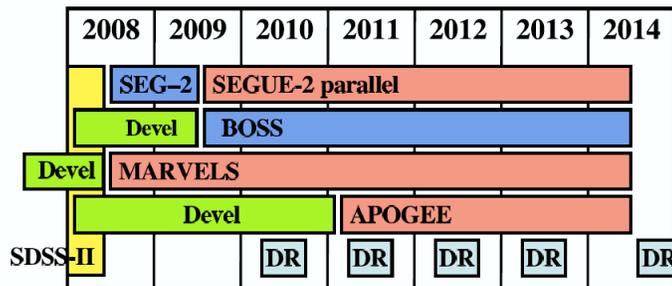


Figure III.8: SDSS-III Schedule. From Project Description in <http://www.sdss3.org/>

III.3.3 Hardware: Camera, Spectroscope, Telescope

BOSS benefits of the facilities of previous version of SDSS. Nevertheless some upgrades were necessary because of SDSS-III/BOSS was designed to observe 1-2 magnitudes fainter in relation with its predecessor. In this section I describe the instruments used in BOSS specifying in each case were they are the original ones.

- Telescope. The Sloan Survey uses a 2.5m f/5 modified Ritchey Chretien ⁵ wide field altitude azimuth telescope (Gunn et al., 1998). It is located at Apache Point Observatory (APO) in Southern New Mexico. It images the sky by scanning along great circles at the sidereal rate or TDI mode that means time delay and integrate. TDI mode involves keeping the telescope stationary and letting the stars drift across the CCD camera (or one telescope that moves at a rate that is slowed from the sidereal rate). The objective is to synchronize the readout rate of the CCD image detector with the rate stars moves across that column. The sky drift rate varies as the cosine of the declination angle. The readout of each line of CCD camera must be timed to coincide with the drift rate of a single star across one column of the CCD camera. It poses a wide field of view of 7 deg² (Gunn et al., 2006). The TDI mode of the telescope requires an almost flat field with zero distortion. These requirements are satisfied using a optical configuration consisting of large secondary mirror (half of the primary) and two corrector lenses. The first corrector lens is a classical Gascoigne astigmatism corrector, the second is a aspheric plate close to the focus which corrects distortion and the lateral chromatic aberration introduced by the first element. The result is a almost flat field of 3 deg with zero distortion. The quality of telescope is better than 0.5 arcsec rms images over the camera field in all bands. Resolution is seeing limited. The seeing is about 0.8 arcsec in the APO site. It is equipped by a imaging camera that mounts on Cassegrain focus and two double spectrographs permanently mounted on the rotator. The telescope was designed to change between photometric mode and spectrometric mode by removing the camera and muting at the Cassegrain focus a fiber plug plate. The Fig.[III.9] shows a view of SDSS telescope and scheme of the instrument.

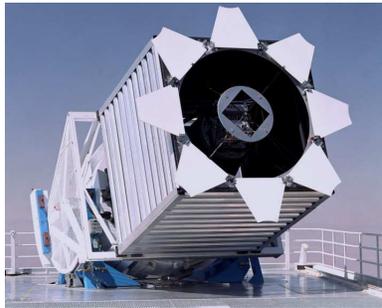


Figure III.9: SDSS Telescope. From <http://www.sdss3.org/>

- Camera and Filters. For imaging the sky it possesses a drift-scanning mosaic CCD camera (Gunn et al., 1998). It contains 2 sets of CCD arrays, one set for the astrometric measures and the other set for the photometric measurements. The imaging array consists of 30 CCD of 2048 × 2048 pixels. The pixel size is 24 μm (0.396 arcsec sky). These CCD are arranged in a six columns array and 5 rows (Fig.[III.10]). Each row observes the sky with a different filter in a temporal sequence r, i, u, z and g. The imaging survey is taken in a drift scan mode

⁵ A telescope Ritchey-Chretien is a telescope type Cassegrain design to eliminate the coma aberration. A Cassegrain telescope is an optic dispositive compound of two mirrors. A primary concave and parabolic mirror and a convex et hyperbolic secondary mirror. In the Ritchey-Chretien telescope both mirrors are hyperbolic

(TDI mode). ⁶The effective wavelength of the 5 filters are 3560 for u, 4680 for g, 6180 for r 7500 for i and 8870 for z. The response functions are shown in the Fig.[III.10].

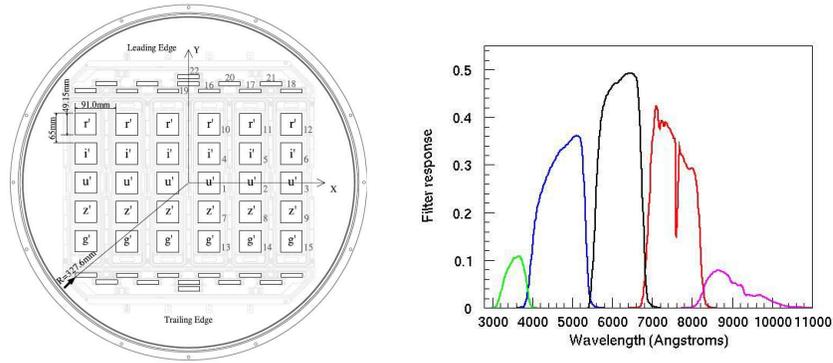


Figure III.10: Left panel corresponds to a schematic view of CCD array . Right panel the Response Function of Filters (Gunn et al., 1998).

Finally the array for astrometry measurements consist of an array of 128×2048 CCD with an r filter covering the entire camera.

- Spectrograph. Each spectrograph ⁷ is a device of two arms one for the blue and one for the red sides of the spectra. Each spectrograph is fed with optical fibers using aluminum plates. The plates are drilled with holes corresponding to the positions of the targets and then the fibers are manually plugged into plates. Each spectroscopic plugged plate has a circular field-of-view with radius 1.49 degrees. The range of spectrograph is from $3600 < \lambda < 10\ 000$ with a resolution of $R = \lambda/\Delta\lambda = 1300 - 3000$ with a $(S/N)^2 \approx 22$ per pix at magnitude limit average over (7000-8500 and $(S/N)^2 \approx 10$ per pix average over (4000-5500). The principal update of BOSS in relation with the predecessor was done in terms of the number of fibers . The spectrograph passes from 640 fibers with 3 arcsec of diameter to 1000 fibers of 2 arcsecs of diameter (500 fibers each spectrograph). The increase in fibers yields a rise in survey efficiency and the smaller diameter traduce in a higher signal to noise spectra. The light from the fibers is sent to a beamsplitter with a coating that reflects the blue part of the spectrum and allows the red part to pass through it. Then, 2 cameras, one for the red and the other for the blue bands , register the 2-dimensional spectra for each one of the objets. An upgrade in cameras system was done permitting to achieve the requirements in terms of sensitivity for the Lyman α forest measurements. The SDSS CCD of 2048^2 (24 micron pixel) for the red and blue channels were replaced by a LBNL CCDs of 4128×4114 in red channel (15 micron pixel), and a 4096^2 (15 micron) pixel e2v CCDs in blue channel. The red side CCD has high quantum efficiency in the near-infrared and the blue side is adapted to have better sensitivity in this range of the spectra. This smaller pixel size of CCDs enables the update in fiber system . The Fig.[III.11] shows a scheme of BOSS spectrograph.

⁶TDI mode involves keeping the telescope stationary and letting the stars drift across the chip. The camera surveys the sky in great circles and a given point on the sky passes through the five filters in temporal succession. The effective

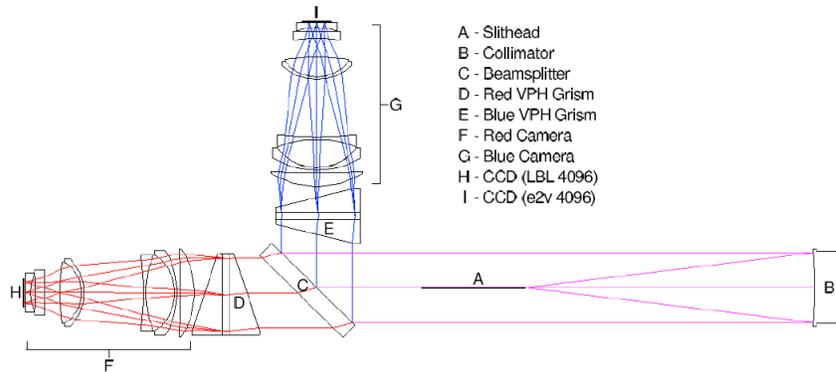


Figure III.11: Scheme of BOSS spectrograph. From <http://www.sdss3.org/>.

III.3.4 Between observations until reduced data

The process to give place to final data follows different stages that I will briefly explain in the next enumeration. For detail description of each one of the stages, I referred the lector to the technical papers that describe in detail the procedures, in this section I only pretend to give an overview of the complete procedure that give place to the reduced data that I analyze,

- Photometric Measurements. The camera surveys the sky in great circles, each point on the sky passes through the five filters in temporal succession. The effective integration time per filter is 54.1 s, the time to pass over the entire photometric array is 54.1 s. The result of the scanning is a long strip consistent of 6 scanlines. To cover a region two strips are overlapped in 8 % (1 arc min) forming an stripe 2.54 centered on the great circle coordinates ⁸Each scanline comprehends information of the 5 bands and the 2 astrometric CCD. The information of each CCD is broken in overlapping series of $10' \times 13.5'$ for easy processing, each series is called frame and contains 161 lines. The frames than corresponds to the same sky location in each 5 bands are grouped together for processing as a field.
- Photo Reduction. The imaging data follows a series of pipelines that perform astrometric calibration (Pier et al., 2003), photometric reduction (Lupton et al., 2001) and photometric calibration (Padmanabhan et al., 2008).
 - Astrometric calibration This pipeline relies pixel position to celestial coordinates, it follows a 3 steps process. First a catalogue is generated in r band, the stars selected from r-band are matched with USNO CCD Astrograph Catalog if available, then bright stars detected in astrometric CCD's are matched with Tycho 2 Catalog. Using this matched objects the set of transformations defined in terms or r band are translated to the others

integration time per filter is 54.1 s, the time to pass over the entire photometric array is 54.1 s.

⁷BOSS posses two spectrographs

⁸The natural system for processing a drift scan is the great circle coordinates system ($\mu \nu$).

bands matching the others bands centroid positions. The transformations are defined including chromatic corrections.

- Photometric Reduction The photometric reduction consists in two separated pipelines postage and frames.

- * Postage Pipeline. This pipeline characterizes the behavior of the PSF as a function of time and localization in the focal plane. It calculates global sky for a field, the flat field vector and bias level.

- * Frames. This pipeline finds deblends, and measures the properties of the objets. It analyses one field by a time integrating the information contained in 5 different frames corespondents to 5 filters. It performs at each frame removal of instrumental signatures (flat field, bias, cosmic rays, bad columns). Then it substrates sky and finally the objets are detected and their properties measured.

The detection of the objets is done in a 4-steps process for each band. First step is run the objet-finder that detects BRIGHT objets looking for pixels 200σ above the noise. Second step is the estimation of sky level by median smoothing the frame on scale $100''$ and proceed to the subtraction. Third step is searching for peaks 5σ over smoothed sky. This set of objets are flagged as binned1 and then subtracted. The frame is rebinned in a 2×2 image, the object-finder runs again, resulting in the sample flagged binned2 and then subtracted, again after rebinning 4×4 , the objet-finder is run a last time, and the set is called binned 4.

Then, the objets are deblended looking for peaks in individual objets. The final list of objets in done merging the list all of bands and adapting the profile of the image. Each objet is identified by run, rerun, camcol, field id objectid identifier within each field, and the genealogy. Finally the pipeline measures properties of each objet and position between others.

- Photometric Calibration. Photometric calibration consists on relaying outputs of the CCD to physical fluxes received above the Earth atmosphere. In SDSS I and II the calibration process implied the comparison of 3 difference telescopes and filters: the auxiliary telescope that observed a set of standard stars to determine photometricity and atmospheric extinction, the SDSS telescope that surveyed a set of fields to calibrate and finally the Naval Observatory 1m telescope used to extract the standard stars and place 3 photometric system on a uniform system. Because of the systematics introduced for using different photometric systems, SDSS III met in place a new calibration method called ubercalibration (Padmanabhan et al., 2008). It consist on using overlapping regions to realize an internal calibration. In fact the photometric calibration could be separated in relative calibration and absolute calibration. The relative calibration pretends to establish a consistent calibration system across the entire survey, the absolute calibration links it to physical fluxes. The uber-calibration system decouples the two problems. It relays on repeated observations to constraint the photometric calibrations. To implemented to SDSS III : 1)Two additional sets data were taken, one set of short scans across scan directions to get temporal variations in the flat field , the other set consists of long scans designed to connect every part to survey. ⁹ 2) A photometric model is defined by a set of parameters (optical response of the telescope $a(t)$, flat field $f(i,j, t)$, and atmospheric

⁹This data was observed 7 times faster to not affect science goals of the survey.

extinction $k(t)$) and 3) Repeated observations of stars are used to constrain the set of parameters using a likelihood analysis.

- **Target Selection** The targets to perform spectroscopy are selected based on the photometric imaging in the 5 bands applying some cuts that generate homogenous samples for the case of galaxies or using statistical methods in the case of quasars to achieve a high efficiency. The next subsection details the target selection algorithm.
- **Tiling Algorithm** Based on the target list the tiling algorithm (Blanton et al., 2003) assigns targets to the plates, the aim of the algorithm is to assign targets to fibers with a maximum efficiency (number of targets assigned in relation to fibers available) and ensuring uniform completeness in the survey. A practical problem that this algorithm deal with is the fiber collision. Due to the finite size of fiber plugs a separation of the fibers centers of minimum 62 arcsec is necessary so two targets separated by less than this distance could be covered only in the regions of overlapping of plates. These targets are denominated collision targets, so the survey is designed to allocated 99 % of the maximal set of targets that do not have such collisions and 92 % of all targets.
- **Spectra measurement** For each observation the aluminum plates previously drilled with holes corresponding to targets are mounted on a cartridge fixed to the telescope. The observations of BOSS spectra are done with a criterium of constant signal-to-noise rate to assure completeness in redshift. It consist of a base of 15 minutes exposures with additional exposures until arrive to a fix $(S/N)^2$ criterium, a rate $(S/N)^2 \geq 22$ per wavelength pixel at $i=21$ (2 arc sec fiber magnitude) in red cameras and $(S/N)^2 \geq 10$ per wavelength pixel at $g=22$ in blue cameras with magnitudes corrected for galactic extinction. To give an order of magnitude of the time required to take a spectra is about 45 minutes in good weather conditions.
- **Spectroscopic Reduction** After taking spectra, the data follows 2 pipelines `spec2d` and `spec1d`.
 - `spec2d` This pipeline will reduce the 2 dimensional spectrograms to a flux and wavelength calibrated spectra
 - `spec1d` This pipeline analyses the combined merged spectra output by `spec2d` and determines the emission and absorption redshifts , classifies the spectra and measures the emission lines and fit a continuum with galaxy templates.

III.3.5 Target Selection Algorithms

Galaxy Target Selection

As I already mentioned, BOSS galaxy targets were selected from imaging in the 5 bands. The selection was designed to give place a sample of luminous and massive galaxies with an uniform distribution of stellar masses from $z \sim 0.2$ to $z \sim 0.6$. The basis of target selection is that this kind of luminous galaxies have old stellar systems and in consequence they have a prominent 4000 break in their spectral energy distributions, so the presence of this feature in their spectral energy distribution allow us to detect them and also determine the redshift considering the band where this 4000 lies. This criterium was used for the LRG selection in SDSS I-II.

The target selection algorithm (Padmanabhan et al in preparation) consists of two sets of cuts that give rise to two samples at different redshift interval. The Cut I generates the sample called LOW redshift in the interval $0.2 \leq z \leq 0.4$ and the Cut II generates the sample called CMASS. The CMASS sample consists of high mass galaxies and covers redshift greater than $z=0.4$. The cuts are defined in terms of the rotated combination of colors, $c_{\perp}, c_{\parallel}, d_{\perp}$ that are defined to track the passive evolution with redshift of the stellar population:

$$c_{\perp} = (r - i) - (g - r)/4 - 0.18 \quad (\text{III.1})$$

$$c_{\parallel} = 0.7(g - r) + 1.2(r - i - 0.18) \quad (\text{III.2})$$

$$d_{\perp} = (r - i) - (g - r)/8 \quad (\text{III.3})$$

$$(\text{III.4})$$

The definition of the LOW sample is as follows:

$$r < 13.5 + \frac{c_{\parallel}}{0.3} \quad (\text{III.5})$$

$$c_{\perp} < 0.2 \quad (\text{III.6})$$

$$16 < r < 19.6 \quad (\text{III.7})$$

The first cut is responsible for setting the luminosity threshold as a function of redshift (Eisenstein et al., 2001), (Cannon et al., 2006), the cut in c_{\perp} restricts the sample of galaxies to LOW redshift ($z < 0.4$) and the cut in r ensure good spectroscopic performance.

The definition of the CMASS sample is as follows:

$$d_{\perp} > 0.55 \quad (\text{III.8})$$

$$i < 19.86 + 1.6(d_{\perp} - 0.8) \quad (\text{III.9})$$

$$i < 19.9 \text{ and } i_{fiber} < 21.5 \quad (\text{III.10})$$

$$r - i < 2 \quad (\text{III.11})$$

$$z_{psf} - z > 9.125 - 0.45z \quad (\text{III.12})$$

$$r_{psf} - r > 0.3 \quad (\text{III.13})$$

The first three cuts are defined to separate high redshift galaxies and to account for passively evolution of massive galaxies accordingly the population synthesis Maraston model (Maraston, 2005). The second cut is a cut in absolute magnitude or stellar mass where d_{\perp} tracks for redshift (Panmanabhan et al in preparation). The Fig.[III.12] shows an illustration of the 3 cuts in d_{\perp} and i magnitude using the AGES red sample. This plot was generated using the SDSS photometry and the AGES redshifts. The black line represents the evolution in the frame of colors using the passive evolution

model of Maraston. The plot shows the performance of the algorithm, the cut in d_{\perp} performs the separation of low and high redshifts and the combination with i cuts guaranties the constance of the masses in the sample. The last 2 cuts are to reduce the stellar contamination. In relation to its precessor SDSS I-II, BOSS goes fainter and bluer than LRG sample.

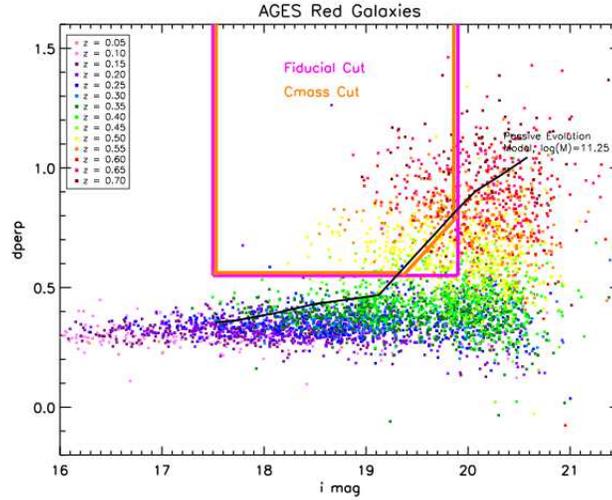


Figure III.12: Illustration of Target Selection using AGES. The plot shows the frame color coded by galaxy redshift for red and blue galaxies from AGES red sample, the CMASS cut is shown and the prediction of the Passive Evolution model of galaxies for a $\log(M)=11.25$

These magnitude cuts use model magnitudes and colors are defined with model magnitudes except for i_{fiber2} which is the magnitude in the 2 arcsec fiber (Stoughton et al., 2002)¹⁰ In case of SDSS I-II Petrosian magnitudes have been used because the sample was stronger, in case of BOSS the cmodel magnitudes are the best choice.¹¹

¹⁰In SDSS we have different kind of magnitudes estimates: Petrosian, model magnitudes and cmodel magnitudes. The Petrosian (1976) system, measures galaxy fluxes within a circular aperture whose radius is defined by the shape of the azimuthally averaged light profile. The model magnitudes (...to be completed). The cmodel magnitudes are based on the better-fitting of a particular model in the r band as a matched aperture to calculate the flux in all bands. The models used are , the exponential and or de Vaucouleurs model, the expressions of the 2 profiles are de following, the first one corresponds to a pure de Vancouleurs and the second equation an exponential profile.

$$I(r) = I_0 \exp(-7.67[(r/r_{eff})^1/4]) \quad (\text{III.14})$$

$$I(r) = I_0 \exp(-1.68r/r_{eff}) \quad (\text{III.15})$$

$$(\text{III.16})$$

¹¹For bright galaxies the measurements of the Petrosian magnitudes are preferred in relation with the others because they have high signal-to-noise ratio and because they are model independent yield a large fraction of the total flux, roughly constant with redshift. For the faintest galaxies, estimates of the Petrosian magnitudes become very noisy, so it is preferred the modelMag is usually a more reliable estimate of the galaxy flux galaxies.

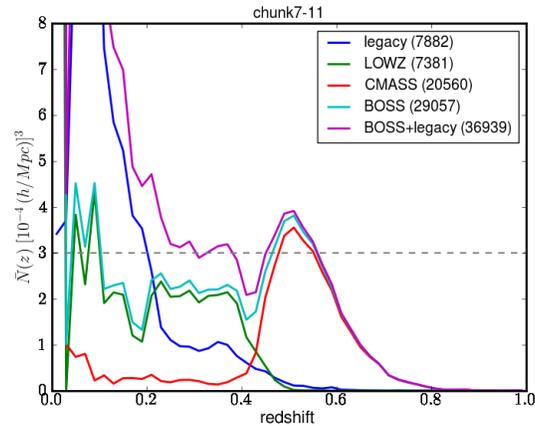


Figure III.13: Comoving number density of Galaxy samples. From (Eisenstein et al., 2011b).

Quasar Target selection

The principal goal of BOSS is to provide cosmological constraints using the BAO physics. Because quasars are only used as a backlight it is not necessary to define an homogeneous sample, the unique criteria for doing cosmology is to maximize the surface density of quasars with redshift $z \approx 2.2$ in order to have access to the Ly α forest in the wavelength range of the spectrograph (having the required S/N within the magnitude limits ($g \approx 22$)).

However, in addition to large scale structure analysis, BOSS plans to provide a quasar sample suitable for doing quasar science. So even if for BAO measurements the uniformity of the sample is not required, for the other quasar science is important to guaranty this characteristic of the sample. The quasar target selection (Ross et al in preparation) reflects this double motivation. It is defined in terms of two samples: the CORE and the BONUS samples. The CORE aims to achieve a uniform sample and the BONUS is concerned to achieve the greatest surface density. The CORE quasars are selected only from Single Epoch photometry with the same algorithm along the survey to guaranty a uniform selection function. They possess the confidence of being QSO based on the highest probability from one or more different methods. The BONUS are selected using a combination of methods in order to achieve largest surface density (≈ 20), the target selection could vary along the survey and also could use information from other sources when available. The 20% of BOSS fibers are reserved to QSO targets, this percentage corresponds to a total of 40 quasar targets per deg^2 . From them a half is dedicated to CORE and the other to BONUS targets. The recovering of quasars is challenging because 2 reasons: the stellar locus at $z \approx 2.7$ crosses quasar locus and the photometric error at this depth is important (near flux limits of SDSS imaging). In consequence the efficiency estimated is about 50 %.

The first year of BOSS data 3 methods of target selection have been used: Kernel density estimator (Richards et al., 2009), the Artificial Neural Network (Yèche et al., 2010) and the Likelihood method (Kirkpatrick et al 2011 in preparation). A variation of Neuronal Network have been used too "Combined Neural Network". The first method denominated KDE uses a statistical

technique known as kernel density estimation, KDE . The kernel density estimate is a generalization of the concept of a histogram where instead of discrete bins whose locations are defined by a grid the are substituted by continuous kernel function. Using the KDE Kernel Density estimation on a training set of known quasars a probability distribution function (PDF) is generated. For the PDF estimation the 4 colors are used t(u-g, g-r, r-i, i-z) and then for a given object the probability of being a quasar or a star is computed, and depending of the position of the objet in the KDE density space it acquires a clasification. An additional cut is used to assure not being close to the stelar locus.

The likelihood method, is based on the generation of a probaility estimation based in Monte Carlo simulations. Classification of the objets is done based on a rate of likelihood of being a quasar and the likelihood of being other objet. The calculations are performed in the flux espace.

$$\mathcal{P} = \frac{\mathcal{L}_{QSO}/A_{QSO}}{\mathcal{L}_{EE}/A_{EE} + \mathcal{L}_{QSO}/A_{QSO}} \quad (\text{III.17})$$

where \mathcal{L}_{QSO} is the likelihood that an object is a quasar and \mathcal{L}_{EE} is the likelihood that an object is in the ‘‘Everything Else’’ catalog. \mathcal{P} is the probability for a single target. The normalization A_{QSO} is the sky area of the QSO catalog. The normalization A_{EE} is the sky area of the Everything Else catalog.

Finally the neural network method works directly with the fluxes, a neural network is constructed based on 10 inputs, the g magnitud, 5 error magnitudes and the 4 colors. The neural network is trained with a set of known objets. The output of the NN provides a function valued between 0 for non quasar an 1 for quasar that enables to classify the objets.

The target algorithm of quasars have been evolving within the first year of BOSS operation, globally for this year in the nominal area the efficiency achieve $\sim 35\%$ that corresponds to $\sim 14 z \geq$ quasars deg^{-2} . The ‘‘CORE’’ selection efficiency is just over $\sim 50\%$. The science requirement is set to 15 quasars per deg^2 . The surface density is now achieving the science requirement goal. In the Fig.[III.14] the number density of quasars is shown with the results of one year of BOSS operations.

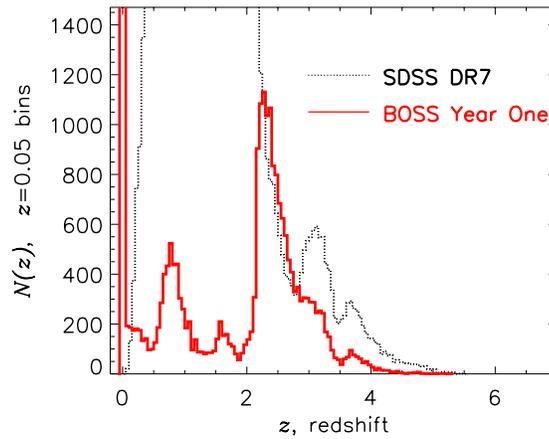


Figure III.14: Comovil number density of QSO samples. From (Eisenstein et al., 2011b).

IV. Study of correlation function estimators with simulations

In order to constrain theories against observations the accurate estimation of clustering statistics from galaxies surveys is necessary. This chapter is devoted to study the estimators and the error analysis that will be used in further chapters to analyze the large scale structure in BOSS data.

The most widely used statistic in large scale structure analysis is the two-point correlation function in real and Fourier space. If initial fluctuations are Gaussian, the power spectrum, or its Fourier transform, the correlation function, provides a complete description of fluctuations. Even if the initial fluctuations are not exactly Gaussian, the two point statistics provide a good starting point for higher order analysis. We have seen in the first chapter that the power spectrum is supplied by theories that describe the state of universe at early times and in the chapter devoted to the BAO we have seen how this power spectrum had evolved. In that sense, a direct comparison between theories and observations could be done. Even if $\xi(x)$ and $P(k)$ are equivalent in theory, the same statement applied to the estimates of $\hat{\xi}(x)$ and $\hat{P}(k)$ derived from finite and noisy samples, is not true (Feldman et al., 1994). Historically, the Baryon Acoustic Oscillations were detected in real space, even though, the Fourier analysis have also been used by many authors for data analysis of large surveys in the last decade. There are advantages and disadvantages of both analyses. In this work I use the correlation function analysis, which is more adapted to explain the physics involved in BAO.

For this chapter, I decide to follow an approach that corresponds to the process followed during my thesis. That is, I study the properties of the estimators of the correlation function and the systematic and statistical uncertainties using mock catalogs. The structure of the chapter is the following. In the first section, I first describe the mock catalogs used to study the estimators. The mock catalogs were generated from log normal simulations. These simulations are not demanding in computing time and are accurate enough to study the two point statistics in large scales, permitting to generate many realizations and study precisely the effects of cosmic variance. These mocks will serve to illustrate the estimators and related concepts and to test different statistical and systematic effects in the estimator.

In the second section I present the various correlation function estimators available. Briefly I will define the concepts of cosmic error and cosmic bias that would allow us to compare the

different estimators. Using the lognormal simulations I show that Landy-Szalay estimator is the most adequate to analyze large scale structures. In the last part of the section I will concentrate in the Landy-Szalay estimator and its properties, I present some approached analytical expressions for the variance and the bias available in the literature. I compare theoretical errors with the error estimation from simulations. As we will see, the estimators of the correlation function are based in the comparison of the galaxy distribution with a random distribution which mimics the observational conditions of data samples. In the fourth section I discuss the random generation that is a crucial ingredient of the estimators. I discuss the elements that have to be taken into account in random generation to have an accurate estimation of clustering statistics: the radial and angular selection functions. I study separately the effect of radial selection function and the geometry. In the fourth part I join all the pieces to compute the correlation function from realistic simulations. The simulations mimics the properties of the LRG sample of SDSS-DR7, survey geometry, redshift distribution, bias and mean density. In this section I study how the completeness of the sample affects the correlation function. I introduce the radial weights designed to minimize the cosmic variance. I study the effect of applying them in the estimation of the correlation function.

Finally, the last part is devoted to the error analysis. Measurements in galaxy catalogs are subject to uncertainties that must be properly addressed in order to compare with theoretical predictions. These uncertainties can be separated according to its nature in statistical and systematic. The statistical errors are the shot noise and the cosmic variance. The systematic errors could be separated in theoretical and observational. Within the uncertainties that comes from the theoretical side we have the effect of peculiar velocities, projection effects, galaxy bias (galaxy distributions do not follow exactly the matter distribution), magnitude limited catalogs, etc. The observational errors arise from technical limitations due to telescopes and instruments. As an example we have not perfect sampling (close pairs of galaxies), effects of dust extinction, contamination from sources (stars), among others. In this chapter I will concentrate in the statistical uncertainties, in the chapter devoted to analysis of BOSS data I will talk about the observational bias and errors proper to BOSS Survey and in the chapter devoted to cosmological constraints I will talk about the uncertainties coming from theoretical side.

IV.1 Log Normal Simulations

We have said in the last chapter that to describe the density field $\rho(\vec{r})$ one uses an scalar random field. To introduce the lognormal random field let us remind the definition of the contrast of density $\delta(\vec{r})$ of a continuous density field $\rho(\vec{r})$:

$$\rho(\vec{r}) = \bar{\rho}[1 + \delta(\vec{r})] = \bar{\rho}\Delta \tag{IV.1}$$

where the fluctuations must have $\langle \Delta \rangle = 1$ and $\Delta > 0$, given a value of the mean density field $\bar{\rho}$. The idea is to generate a random field Δ whose variance follows the right power spectrum. One possibility is to use a gaussian random field for Δ , motivated by the physical reason we have introduced in the last chapter supposedly generated within Inflation. A gaussian random model is only valid for a density field with linear evolution (valid in the limit of zero fluctuation amplitudes). Assuming a finite rms (root mean square) fluctuations σ some problems arise because of the non positivity of the density field. In particular, as a result from this, we assign a non zero probability

to regions of meaningless negative density. A solution is to use lognormal density fields to simulate matter distributions. The use of lognormal (LN) random fields to model the distribution of matter of the universe was proposed and studied by Coles and Jones (Coles & Jones, 1991). This model which is completely specified statistically as the gaussian one and do not violate the common sense conditions $\rho > 0$. Furthermore, the possibility to use this models to study the clustering statistics was shown in (Coles & Jones, 1991). The lognormal model is physically motivated as well by a kinematic argument stating that if initial peculiar velocity are gaussian and extrapolating into non linear regime leads to a lognormal distribution. Overall arguments there is the simplicity of this model in comparison with others non gaussian random fields.

A lognormal density field is defined as:

$$Y = Ae^X \quad (\text{IV.2})$$

where $A = e^{-\frac{\sigma_X^2}{2}}$ is the normalization allowing $\langle Y \rangle = 1$ and $X(\vec{x})$ is a gaussian random field which probability density function (PDF) given by the normal distribution with some mean and variance $N(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(x - \mu)^2/2\sigma^2] \quad (\text{IV.3})$$

The resultant random field Y respect the desired properties of the density field, that is $\langle Y \rangle = 1$ and $Y > 0$. As well as gaussian field, a log normal gaussian field is completely described by its correlation function, and has the advantage that the positivity of the field is guaranteed and its proximity with a gaussian random field allows us to calculate many properties in the same way as for a gaussian field.

To simulate a lognormal field the starting point is an input power spectrum that we want to imprint in the random field. We need an expression that relates the power spectrum that must be applied to a gaussian field in order to get the desired power spectrum for the lognormal random field. This step is required because what we know how to generate are the gaussian random fields. This relation is given in terms of the respective correlations functions. The following expression relates the correlation function of the lognormal gaussian field $\xi_{LN}(k)$ with the correlation function of the gaussian random field $\xi_G(k)$ (Percival et al., 2010):

$$\xi_G(k) = Ln[1 + \xi_{LN}(k)] \quad (\text{IV.4})$$

Once we have the expression for the power spectrum. We have to generate a discrete gaussian random field $\delta(\vec{r})$ in configuration space with a variance $\sigma = 1$ and mean $\mu = 0$.

$$\rho(x) = N(1, 0) \quad (\text{IV.5})$$

Then we go to the Fourier space where the different \vec{k} modes are independent. We get the Fourier coefficients $\hat{\rho}_k$, then, each Fourier mode $\vec{k} = (k_x, k_y, k_z)$, must have a variance given by the value of the power spectrum at $k = |(k_x, k_y, k_z)|$. To do it in Fourier space we only have to multiplying the gaussian random field $\hat{\rho}_k$ which variance is 1 by the desired variance $\sqrt{P_G(k)}$ with the power spectrum given by Eq.[IV.4] multiplied by a factor $\frac{(\Delta k)^3}{(2\pi)^3} = \frac{1}{L^3}$. And compute the inverse Fourier

transform again to get the gaussian density field that we will introduce in expression Eq.[IV.2] to get the final lognormal field.

The density field generated with the lognormal random field is then sampled by a discrete point distribution, we assume that the galaxies follows a Poisson distribution. A complication comes when working with galaxies, as we will see later, the distribution of galaxies does not trace exactly the distribution of matter. The galaxies are more clustered in overdense regions, thus a bias must be included to account for the amplification of fluctuations, this factor is called *galaxy bias*.

$$\rho_b(x) = b\rho_m(x) \quad (\text{IV.6})$$

Thus, to generate the mock galaxies from the resultant density field, the biased density field in each pixel is used as a probability distribution function. In Fig.[IV.1] a slice of the density field generated in box of 1600 Mpc^3 in a grid of 200^3 pixels.

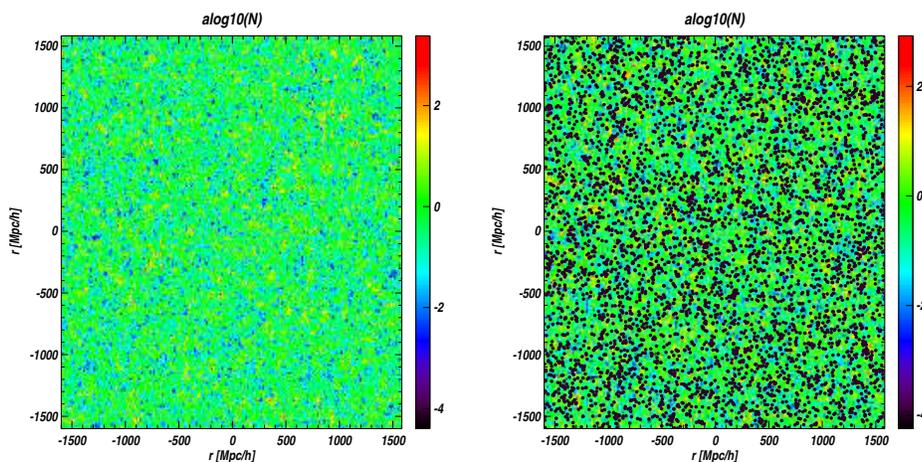


Figure IV.1: In left panel an slice of the density field generated with a log normal simulation using a power spectrum as the input. In right panel the same slide populated with galaxies following the density field distribution of matter

In panel Fig.[IV.2] a plot of the input power spectrum and the power spectrum from mock galaxies are shown. To estimate the power spectrum from galaxies the density contrast $\delta(\vec{r}) = \frac{n_g(\vec{r}) - \bar{n}_g}{\bar{n}_g}$ is directly Fourier transformed:

$$\delta_k = \frac{1}{V} \int \left[\frac{n_g(\vec{r})}{\bar{n}_g} - 1 \right] e^{i\vec{k}\cdot\vec{r}} d^3r = \frac{1}{N_g} \sum_{j=1}^{N_g} e^{i\vec{k}\cdot\vec{x}_j} - W_k \quad (\text{IV.7})$$

where the galaxy number density is $n_g(\vec{r})$, N_g is the galaxy number and W_k is the Fourier transform of the window function of the survey:

$$W_k = \frac{1}{V} \int e^{i\vec{k}\cdot\vec{r}} d^3r \quad (\text{IV.8})$$

In this case, computing the power spectrum of the galaxy distribution is trivial because we compute it in the periodic box.

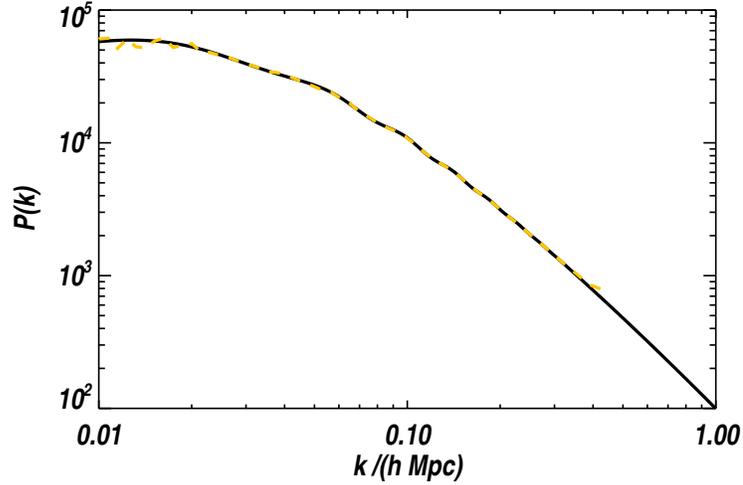


Figure IV.2: The power spectrum introduced to the simulation as input (in solid black line) and the power spectrum generated from the log normal density field (in dashed yellow line).

IV.2 Estimators

The theory of estimators of large scale structure was developed by Peebles in the 70's in real and Fourier space. From that moment a lot of work have been done to improve the estimators, and to do an accurate estimation of error and bias beyond the simplistic Poisson errors.

Before going to the correlation function estimators let us discuss the assumptions that are implicit in the measurements of structure from galaxy catalogs. There are two assumptions, the so called *fair sample hypothesis* and *local Poisson approximation*. The fair sample hypothesis states that the finite part of the observable is a fair sample of the whole universe. Thus separated regions of the universe could be considered as independent realizations of the same physical process, enabling to replace ensemble average by spatial averages. The second hypothesis comes from the fact that we do not have access to the density field directly rather we have a discrete sample, a galaxy catalogue. Thus, we assume that this point distribution results from a Poisson realization of a underlying continuous field. Where the probability of finding N objects in a volume v at a location \vec{x} , with a mean $\bar{N} = \bar{n}v$ is given by:

$$P_N^{Poisson}(\bar{N}) \equiv \frac{\bar{N}^N}{N!} e^{-\bar{N}} \quad (\text{IV.9})$$

In the first chapter we presented the correlation function as the second order statistics that describes the clustering of a point process. Based on these 2 hypothesis, the calculation of the two-point correlation function relies on the fact that it can be defined in terms of the excess of probability of finding a pair of galaxies in volume elements dV_1 and dV_2 separated a distance $r_{12} \equiv |\vec{x}_1 - \vec{x}_2|$ compared with a random distribution as we have said before (Chapter 1).

$$dP_{12} = \bar{n}_g^2 [1 + \xi(\vec{r}_{12})] dV_1 dV_2 \quad (\text{IV.10})$$

For a random distribution ($\xi = 0$), the probability of having a pair of particles is given by the mean density squared independently of distance. If the objects are clustered, the points are correlated ($\xi > 0$) the probability is enhanced, if the objects are anti-correlated ($\xi < 0$) the probability is suppressed with respect to a random distribution.

IV.2.1 Correlation Function Estimators

Due to the discrete nature of studied sample, the traditional estimators are based on pairs counting in the data sample compared with a random sample in the same volume. In what follows we assume a 3-D data catalogue D of volume V and containing N_d objects, the average number density is defined as $\bar{n}_d = N_d/V$. In order to compute its correlation function we have to generate a random catalogue R , in the same volume V , with N_r random objects ($\bar{n}_r = N_r/V$).

The natural or simplest estimator that we could define is compare the pair counting of our data sample with the random sample. This estimator was proposed by Peebles-Hauser and is given by:

$$\hat{\xi}_{PH}(r) = \frac{DD}{RR} - 1 \quad (\text{Peebles\&Hauser, 1974}) \quad (\text{IV.11})$$

where DD is the normalized number of pairs belonging to a bin r in the galaxy catalogue, RR is the normalized number of pairs belonging to a bin r from the random catalogue.

$$\begin{aligned} DD &= \frac{P_{DD}(r)}{N_d(N_d-1)} \\ RR &= \frac{P_{RR}(r)}{N_r(N_r-1)} \end{aligned} \quad (\text{IV.12})$$

$P_{DD}(r)$ is the number of pairs belonging to a bin r in the D catalogue and $P_{RR}(r)$ is the number of pairs belonging to a bin r from R catalogue. The idea of this simple estimator is to deal with geometry and to avoid the need of calculate separately the edge corrections. Various alternatives have been proposed to improve the estimator. I summarize the most popular:

$$\begin{aligned} \hat{\xi}_{DP}(r) &= \frac{DD}{RD} - 1 && (\text{Davis\&Peebles, 1983}) \\ \hat{\xi}_H(r) &= \frac{DD \times RR}{RD^2} - 1 && (\text{Hamilton, 1993}) \\ \hat{\xi}_{Hew}(r) &= \frac{DD - DR}{RR} && (\text{Hewett, 1982}) \\ \hat{\xi}_{LS}(r) &= \frac{DD - 2RD + RR}{RR} && (\text{Landy\&Szalay, 1993}) \end{aligned} \quad (\text{IV.13})$$

RD is the normalized number of pairs belonging one to the galaxy catalogue one to the random catalogue, given by:

$$RD = \frac{P_{RD}(r)}{N_r N_d} \quad (\text{IV.14})$$

In Fig.[IV.3] the different estimators are shown using a single realization of a log normal field with 3,000 and 30,000 mock galaxies within a sphere and a random catalog of the same number of objects. As we can see from the plot, for the smaller set the estimators have large fluctuations. As the number of data points increases these estimators converge to the input of the simulation. When applied to a larger sample the Hewett estimator and Davis-Peebles gives similar results. The same applies for Landy-Szalay and Hamilton estimators. We will see later that this effect is related to the fact that we are sampling a density field by a discrete distribution of points. Let us introduce some definitions that will allow to us select the best estimator for large scale structure analysis.

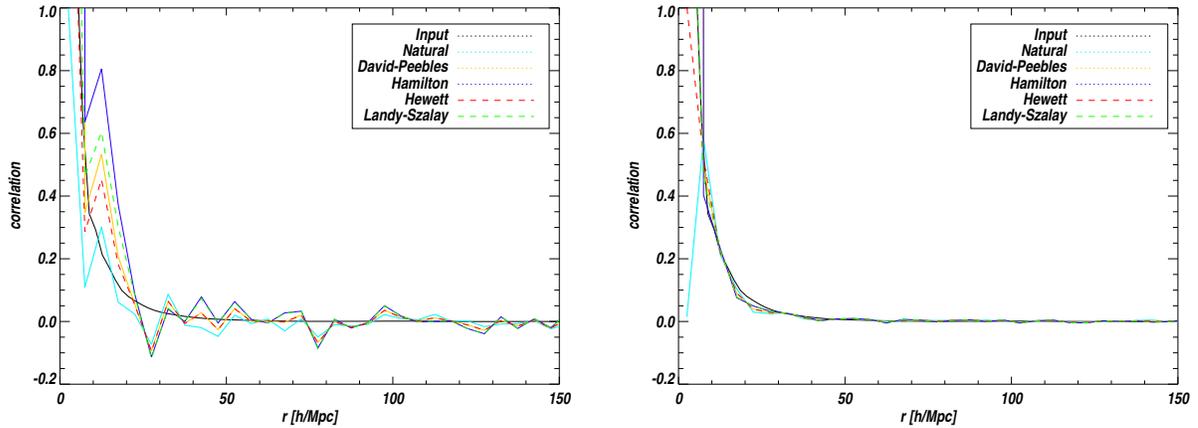


Figure IV.3: In right panel, the correlation function using the five popular estimators from a realization from a lognormal simulation following a Λ CDM model with 3000 mock galaxies in the left and with 30000 in the right. In black, the input correlation function of the simulation. When the number of objects are small there are large fluctuations as we will see later dominated by the shot noise. As the number of objects increases the estimators converge to the input of the simulation.

IV.2.2 Cosmic Error and Cosmic Bias

If we have a statistic A , and \hat{A} is its estimator, the bias¹ of the estimator is defined as:

$$b_A = \frac{\langle \hat{A} \rangle - A}{A} \quad (\text{IV.15})$$

The cosmic bias expresses the difference between the expected value and the true value of the physical quantity concerned. With real data there is no way of evaluate the bias of an estimator if it exist, nowadays, we can study the bias with simulations.

To study the bias of the estimators 100 lognormal simulations were generated with 30 000 mock galaxies. The distribution of galaxies follows a power spectrum that we introduce as an input of the simulation. The power spectrum used corresponds to a Λ CDM model. For the moment, we use a simple geometry a sphere centered at origin of radius of $800 \text{ Mpc}/h$. For each realization the estimator is computed. The empiric mean over 100 realizations is compared with the input correlation function. In the left panel Fig.[IV.4], the means for each estimator are shown as well as the input function. In the right panel the bias of the different estimators as defined by Eq.[IV.15] is plotted. As we can observe from the figures, the estimators that are in better agreement with the input correlation function are Hamilton and Landy-Szalay. The estimators Davis-Peebles and Hewett converge to the same mean underestimating the correlation function at large scales. Finally the naive estimator is the most biased estimator at least on the large scales.

¹This term must not be confused with the galaxy bias that takes account of the difference between matter distribution and galaxy distribution

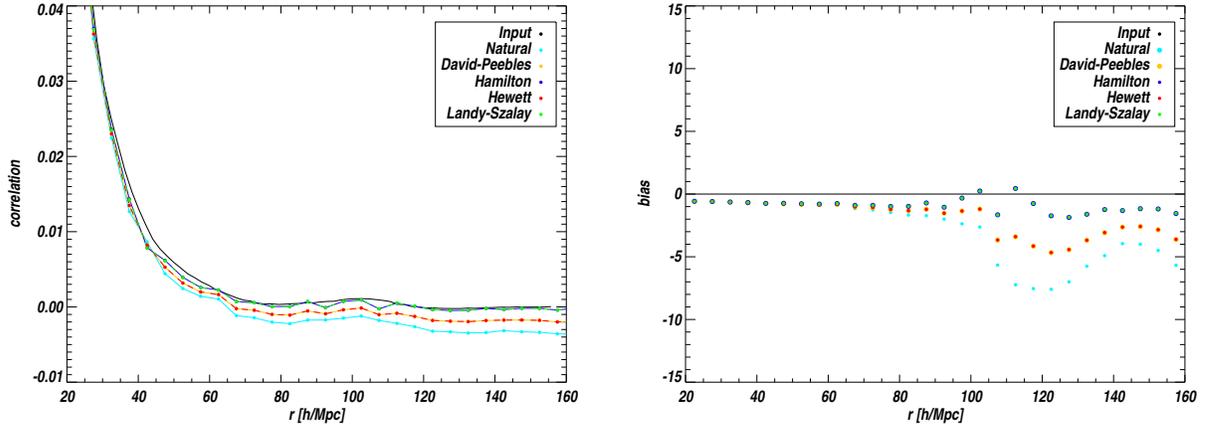


Figure IV.4: In left panel, the empiric mean of five popular estimators of the correlation function from 100 simulations log normal following a Λ CDM model is shown. In black, the input correlation function of the simulation. As we observe from the plot the less biased estimators are the Hamilton and Landy Szalay. In right panel the bias of the different estimators is shown, as defined by Eq.[IV.15]

Error

The error is defined as the variance of the estimator ΔA^2 . That is :

$$\Delta A^2 = \left\langle \left(\hat{A} - \langle \hat{A} \rangle \right)^2 \right\rangle = \int \left(\hat{A} - \langle \hat{A} \rangle \right)^2 P(\hat{A}) d\hat{A} \quad (\text{IV.16})$$

where $P(\hat{A})$ is the probability of measure the value of \hat{A} in a galaxy catalog given a theory, also called *Cosmic Distribution function*. Usually it is assumed $P(\hat{A})$ is gaussian.

To study the variance of the estimators we follow the same procedure, we estimate the sample variance of the estimators for the 100 realizations using the corresponding sample mean for each estimator. In left panel of fig IV.5 I show the variance of the different estimators. As we have seen in the Fig.[IV.4], the mean of the different estimators do not converge exactly to the real value of the correlation function, thus in Fig.[IV.5] I plot the mean of Landy-Szalay estimator with the error bars measured for the different estimators. In the right plot I show the result using the mean value correspondent to each estimator to show the combined effect.

This simple test permits us to compare the different estimators and statistical uncertainty in the estimation. A good estimator must minimize the cosmic bias and the cosmic error. We are interested in the scales related with the bump of BAO, that we call the large scale. As we observe from the plot at this scales the four estimators seem to give similar variances but taking into account that the less biased estimators are the Hamilton and Landy Szalay. It seems that the H and LS show better properties.

Before concluding which is the best estimator let us introduce more definitions. The statistical error or cosmic error is usually separated in three contributions, even if in practice the three effects are correlated with each other.

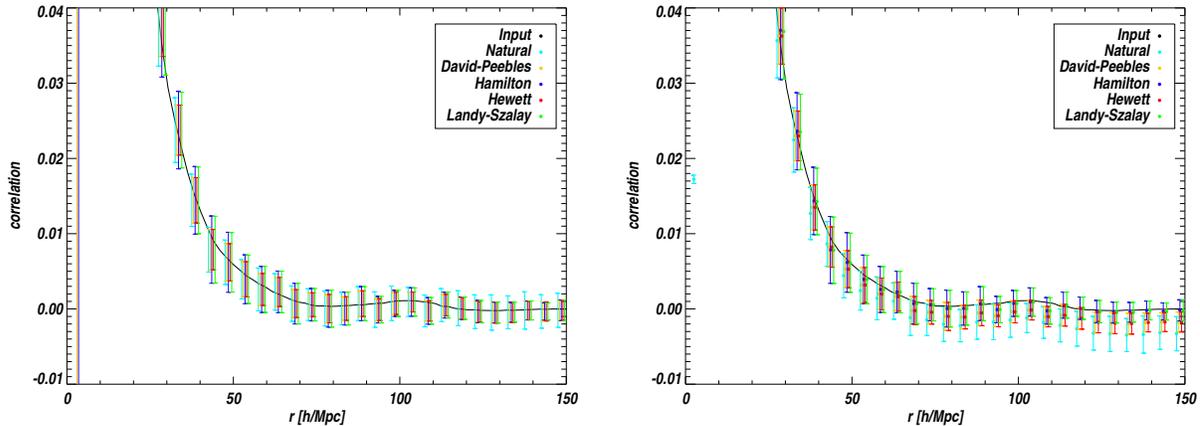


Figure IV.5: Sample variance of five popular estimators of the correlation function from 100 simulations log normal following a Λ CDM model. In black, the input correlation function of the simulation. The points have been translated 0.5 Mpc for purposes. To isolate the effect of the variance in left panel I show the errors using the mean value of the LS estimator. In left panel the result using the mean value correspondent to each estimator is shown.

- *Cosmic Variance.* This error comes from the fact that we have access only to finite volumes, so given a size L of a survey we have access to a limited number of modes of a given size. In particular the mean density is not well determined. Another way to understand the cosmic variance is to relate it to the number of independent representations of the observable universe. As we measure a statistical quantity averaged over the number of realizations. As the number of realizations is small the error is large. The cosmic variance is roughly proportional to the average of two point correlation function $\bar{\xi}$ over the whole survey.
- *Sample Variance* Also called *Shot Noise* is originated because the galaxy distribution is a discrete Poisson representation of an underlying continuous field whose statistical properties we want to describe. This error is proportional to $1/N_d$, where N_d is the number of data points. It becomes negligible for large N_d .
- *Edge Effect* This error comes from the geometry of the survey, in the borders of the survey the objects do not have many neighbors as the points in the inner regions do, thus the pair counts tend to be smaller than they are. The general idea of edge corrections is to give less weight to the galaxies near the edges than those far away from the boundaries. In fact the edges effect could be partly corrected for scales $r \ll L$.

A characteristic from these estimators that we must notice is that we can reduce the part of the discreteness effect that comes from the random catalogs using a number of objects N_r several times larger than N_d . In this way the calculations of DR and RR can be arbitrarily improved. To test this we are going to use simulations with a greater number of objects 100,000 to guaranty we have enough statistical power. In figure IV.6 the correlation function is shown using only the Landy-Szalay estimator from the same data sample calculated using a random of the same size and

a random two and three times the size of the mock galaxy catalog. This figure shows that using a random sample larger than the data sample we can decrease the sample error.

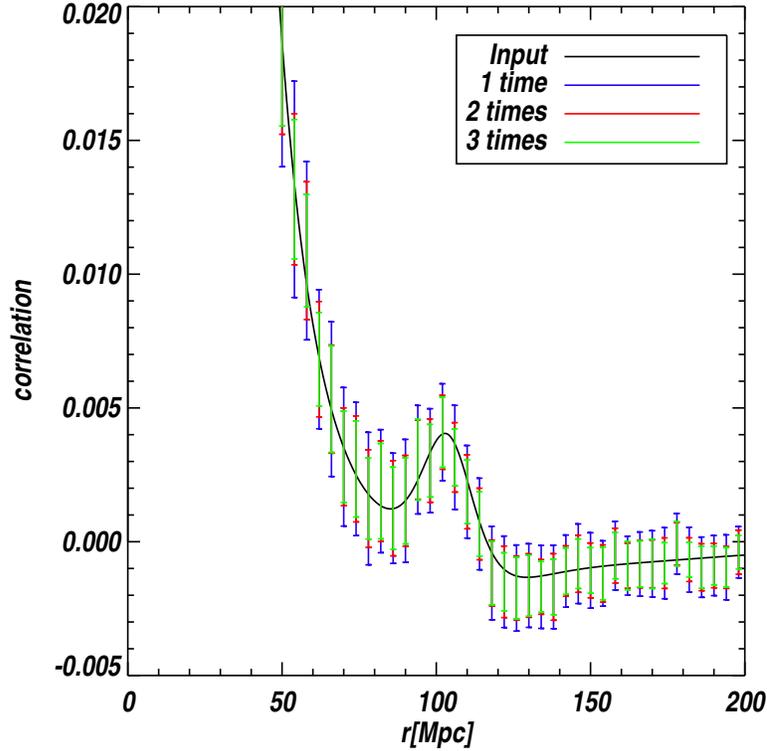


Figure IV.6: The correlation function of one realization of 100 000 mock galaxies with a random set of same size and with a random 2 and 3 times greater than the data set. As we observe the shot noise decreases with the size of the random set used to compute the correlation function.

Until now we have shown that Landy-Szalay and Hamilton have similar properties, it is not surprising due to the similarity of the estimator, being Landy Szalay the symmetric version of Hamilton. In (Pons-Bordería et al., 1999) it is shown that the Hamilton estimator is more sensitive to the number of random thus is preferable to use the Landy-Szalay estimator. We will test this statement comparing the results from Hamilton and Landy-Szalay. The dispersion has been computed over 100 simulations using Landy-Szalay estimator in dashed lines and Hamilton in solid lines. In Fig.[IV.7] I show 1- σ error from log normal simulations for three different sizes of random set (a random set equal to data set, in red and blue, a random set 2 and 3 times greater than the data). From the plot we observe that the difference between the dispersion of Landy Szalay and Hamilton estimators is indistinguishable. Thus, these statements must be tested with more realistic simulations.

We can conclude that Hamilton and Landy-Szalay show the better properties, less bias and small variance. Some references claim the superiority of Landy Szalay over Hamilton but we have not found a difference in their behavior. In fact the difference between these estimators roughly is the

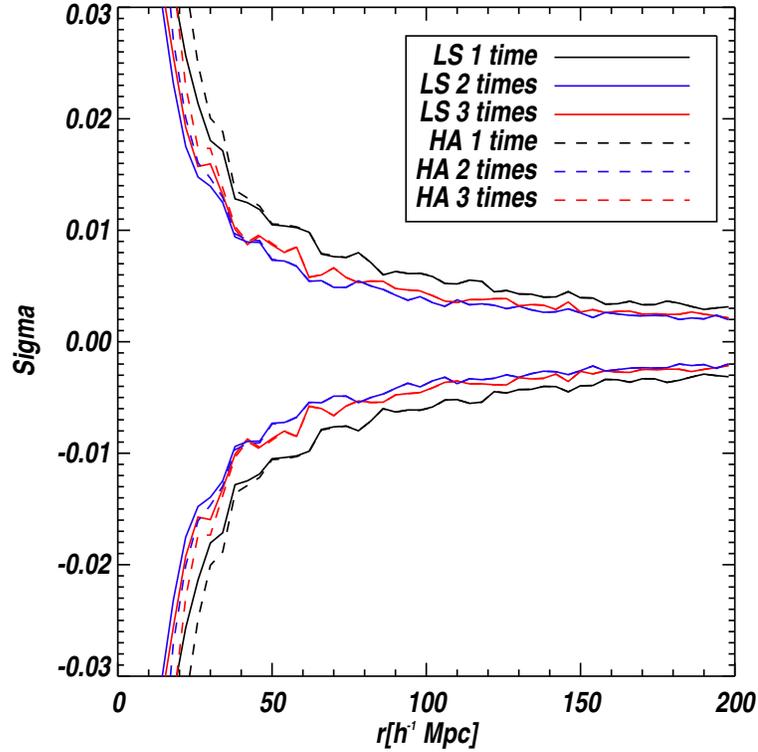


Figure IV.7: 1-*sigma* error from log normal simulations for three different sizes of random set, in black a random set equal to data set, in red a random set 2 times greater than the data set and in blue three times. Average over 100 simulations of Landy Szalay in dashed lines, average of 100 realizations of Hamilton in solid lines.

way of dealing with edge effects, thus the differences between them would be enhanced with more complex geometries where the edges of the sample volume are an important source of uncertainty. In a further section we will test Hamilton and Landy-Szalay estimators in a more realistic scenario taking into account radial and angular selection functions, to test if Landy-Szalay show better properties. I present in the next section the approximated analytical expressions for mean and variance of the estimators that point out Landy Szalay as the most adequate estimator of the correlation function.

IV.2.3 Analytic expressions of mean and variance of estimators

Landy & Szalay (Landy & Szalay, 1993) developed a formalism to calculate cosmic bias and variance of any estimator based on the pair counts in terms of fluctuations about their mean:

$$\begin{aligned}
 DD &= \langle DD \rangle (1 + \alpha) \\
 RD &= \langle RD \rangle (1 + \beta) \\
 RR &= \langle RR \rangle (1 + \gamma)
 \end{aligned}
 \tag{IV.17}$$

where α, β, γ are the fluctuations about the mean for a given realization of the data or random set. By definition, $\langle \alpha \rangle = \langle \beta \rangle = \langle \gamma \rangle = 0$. As we have said before, the variance of γ could be made arbitrary small choosing a random set sufficiently large, thus we can suppose $\text{var}(\gamma) = \langle \gamma^2 \rangle = 0$. Using these expressions (Eq.[IV.17]) we can express the variance of any estimator based on combinations of pair counts of data and random sets in terms of the mean of pair counts and their fluctuations. In order to do that we must calculate the variances of pair counts.

$$\begin{aligned}\langle \alpha^2 \rangle &= \frac{\langle DD \cdot DD \rangle - \langle DD \rangle^2}{\langle DD \rangle^2} \\ \langle \beta^2 \rangle &= \frac{\langle RD \cdot RD \rangle - \langle RD \rangle^2}{\langle RD \rangle^2} \\ \langle \alpha\beta \rangle &= \frac{\langle DD \cdot RD \rangle - \langle DD \rangle \langle RD \rangle}{\langle DD \rangle \langle RD \rangle}\end{aligned}\tag{IV.18}$$

Landy-Szalay calculated the mean and variance for the different estimators in terms of these fluctuations in the weak correlation limit ($|\bar{\xi}(L)|, \left| \frac{\xi(L)}{\bar{\xi}} \right| \ll 1$). We will present directly the results avoiding the cumbersome calculations that could be consulted in reference (Landy & Szalay, 1993). They found that the mean and variances are given by:

$$\begin{aligned}\langle \hat{\xi}_{PH} \rangle &= \left\langle \frac{DD}{G_p} \right\rangle & \langle \Delta \hat{\xi}_{PH}^2 \rangle &= \left\langle \frac{DD}{G_p} \right\rangle^2 (4t + p) \\ \langle \hat{\xi}_{DP} \rangle &= \left\langle \frac{DD}{G_p} \right\rangle (1 + t) & \langle \Delta \hat{\xi}_{DP}^2 \rangle &= \left\langle \frac{DD}{G_p} \right\rangle^2 (t + p) \\ \langle \hat{\xi}_H \rangle &= \left\langle \frac{DD}{G_p} \right\rangle (1 + t) & \langle \Delta \hat{\xi}_H^2 \rangle &= \left\langle \frac{DD}{G_p} \right\rangle^2 (p) \\ \langle \hat{\xi}_{LS} \rangle &= \left\langle \frac{DD}{G_p} \right\rangle & \langle \Delta \hat{\xi}_{LS}^2 \rangle &= \left\langle \frac{DD}{G_p} \right\rangle^2 (p)\end{aligned}\tag{IV.19}$$

where

$$\begin{aligned}t &= \frac{1}{N_d} \left[\frac{G_t}{G_p^2} - 1 \right] \\ p &= \frac{2}{N_d(N_d-1)} \left[\frac{1}{G_p} - 2 \frac{G_t}{G_p^2} + 1 \right] \approx \frac{2}{N_d(N_d-1)G_p}\end{aligned}\tag{IV.20}$$

G_p and G_t are geometrical factors. G_p is the fraction of unique pairs separated a distance $r+dr/2$ (in practical terms is equal to RR), G_t is the fraction of triplets, that is, given a point the set of other two points within a radius $r + dr/2$. Thus, p is approximately the Poisson noise that is the inverse of the pair counts. As we can observe the H and LS estimators are of second order in the $(1/N_d)$ thus close to the Poisson variance while PH and DP are of first order in $(1/N_d)$. We observe also that there is an additional bias in estimators DP and H. This bias comes from the integral constraint (Landy & Szalay, 1993) as we would see in a further section when we introduce this concept.

The results of this section suggest that the best estimator presented is the Landy-Szalay (LS) estimator. In fact the Hamilton estimator has proven to be good in large scales too, but the analytical expressions show that Hamilton is a biased estimator contrary to Landy-Szalay. We will come back to compare Hamilton and Landy-Szalay in a further section. Meanwhile, it is important to add that detailed studies have shown that the LS estimator is optimal with respect to the cosmic bias and cosmic error in the weak correlation limit (Kerscher et al., 2000) and for a random

distribution (Poisson sample with no correlation) (Landy & Szalay, 1993). Thus, for the following we will use only this estimator to study in a more realistic scenario the correlation function. Before going to the random generation we will present the analytic expressions for the cosmic bias and variance for the Landy-Szalay estimator and how they compare with the errors estimated from log normal simulations.

IV.2.4 Analytical expression for bias and variance of the LS Estimator

Landy Szalay Cosmic Error

In the last section we have already presented the analytic expression of the cosmic variance for Landy-Szalay estimator in the weak correlation limit ($(|\xi|, |\bar{\xi}(L)| \ll 1)$), neglecting edge effects ($r/L \ll 1$) and supposing the average density \bar{n}_g is perfectly determined.

The missing pieces are the expressions of the geometrical factors G_p and G_t which could be computed using Monte Carlo realizations of N_r points randomly distributed within a geometry and computing the following expression:

$$\begin{aligned} G_p(r) &= \frac{\langle n_p(r) \rangle}{N_r(N_r - 1)/2} \\ G_t(r) &= \frac{\langle n_t(r) \rangle}{N_r(N_r - 1)(N_r - 2)/2} \end{aligned} \quad (\text{IV.21})$$

where $\langle n_p \rangle$ is the average number of unique pairs over the ensemble of data sets, and $\langle n_t \rangle$ is the average number of triplets within the sample geometry given one point at the center. Let us compare this estimation of the errors with the results from lognormal simulations. In Fig.[IV.8] the analytical estimation for the error is compared with the result from simulation. The error bars are estimated using expression IV.19 that corresponds essentially to the Poisson noise. We observe that the error bars from Landy-Szalay estimated analytically underestimates the errors of measurements. This is due to the fact that the analytical expression assumes the complete knowledge of \bar{n}_g .

Bernstein (Bernstein, 1994) developed a more general expression considering a fluctuating \bar{n}_g but assuming also the weak correlation limit and neglecting edge effects. The diagonal term is given by the following expression:

$$\begin{aligned} \left(\frac{\Delta\xi}{\xi} \right)_B^2 &\simeq 2 \frac{\bar{\xi}^2}{\xi^2} + 4(1 - 2Q_3 + Q_4)\bar{\xi}(L) + \frac{4}{N_d} \left[\frac{\xi_{ring}(1 + 2Q_3\xi)}{\xi^2} + Q_3 - 1 \right] \\ &\dots + \frac{2}{N_d^2} \left[\left(\frac{1}{G_p} - 1 \right) \frac{1 + \xi}{\xi^2} - \frac{1}{\xi} - 1 \right] \end{aligned} \quad (\text{IV.22})$$

where $\bar{\xi}$ is the average two-point correlation function in the whole survey, $\bar{\xi}^2$ is the average of the square of the two-point correlation function over the survey, ξ_{ring} is the average of the two-point correlation function for pairs inside the shell of radius r and thickness Δr , Q_3 and Q_4 are the 3 and 4 point correlation functions. For the derivation of these expressions we refer to the reader to the references. As we observe from Eq.[IV.22], the more general expression for the cosmic errors relies on prior assumptions about the behavior of statistics involved thus in general they are not commonly used to estimate the errors in clustering statistics.

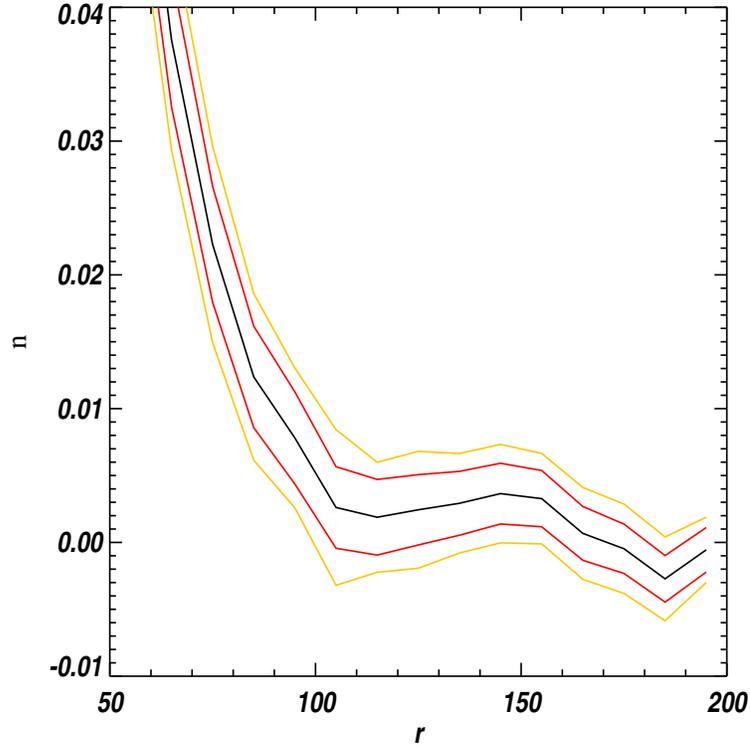


Figure IV.8: In black the input correlation function using the error bars (in red) given the analytical expression IV.19 compare with the error bars given by the standard deviation from simulations in yellow.

Landy Szalay Cosmic Bias

The expression for the cosmic bias for Landy-Szalay estimator in the weak correlation limit ($|\xi|, |\bar{\xi}(L)| \ll 1$), and neglecting the edge effects ($r/L \ll 1$) has a dominant contribution given by.

$$b_{\xi} \simeq \frac{-\bar{\xi}(L)}{\xi} \quad (\text{IV.23})$$

An important concept must be introduced here, the so called *Integral constrain*. The fact that we are computing the average density and the correlation function using the same sample has as a consequence that fluctuations must vanish at the scale of the survey.

$$\frac{1}{V} \int_{V^2} \hat{\xi}(r) dr = 0 \quad (\text{IV.24})$$

As a consequence a bias is introduced, meaning that the correct value of ξ could be obtained adding an unknown constant to the measured value. This bias cannot be corrected unless doing priori assumptions of the form of correlation function at scales larger than those proved by the

survey, so it must be considered as a source of uncertainty.

IV.3 Random Generation

As we have said before, a crucial ingredient in the correlation function is the random sample. The estimators defined to now suppose statistically homogeneous catalogs, i.e. with constant $\bar{n}_d(r)$, but when we deal with real data samples usually we have to deal with non homogeneous samples. The way to deal non uniform samples is to correct for observational selection effects applying the same selection function to random catalogs. The selection function represents the fraction of the total population of galaxies satisfying the limitation criterium at a certain distance or angular position. The uncertainty in the knowledge of this selection function could influence the result of the correlation function.

When a random catalogue is created in practice each random point would have a probability given by the selection function $f(\hat{\theta}, z)$ of being included in the random set. In some cases it is possible to factorize this probability into two independent functions, $f(\hat{\theta}, z) = C(\hat{\theta})\phi(z)$, where $C(\hat{\theta})$ is the angular and $\phi(z)$ the radial selection functions respectively. We will suppose for the following that it is the case, we will see in the analysis of BOSS data that is a reasonable hypothesis.

IV.3.1 Radial Selection Function

In general due to observational constraints at different distances, the sample do not necessary have uniform properties. The apparent brightness of a source depends on the observed flux, since flux decreases with square of distance, only intrinsically luminous sources at large distances will be observed. In consequence we have two effects: number density decreases with distances and a tendency of farther objects to be more luminous. These samples are called *flux limited samples*. A simple way to correct from this observational selection effect is to define a *volume limited sample* that consists in extracting from the parent catalog a subsample characterized by a maximum redshift z_{max} and a minimal luminosity L_{min} . The values of z_{max} and L_{min} are defined such that the apparent magnitude of objects in these catalogs at distance z_{max} would be larger than the magnitude limit. Such a selection criterion renders the number density of galaxies in the subsamples independent of distance. The problem with such samples is a significant information loss (the sample size is significantly smaller than the parent sample). In Fig.[IV.9] we give an example from SDSS DR4 of a flux limited catalog and a volume limited catalog.

Unless we work with volume limited samples we must correct for observational selection effects when analyzing a sample. Thus when working with flux limited samples we must model the expected number of galaxies in the absence of clustering as a function of redshift what is called the radial selection function.

$$\bar{n}_d(z) = \bar{n}_g\phi(z) \tag{IV.25}$$

where $\phi(z) \leq 1$. The radial selection functions $\phi(z) = dN/dz$ gives the probability that an object within a given luminosity range will be observed as a function of the distance and redshift. It is obtained by integrating the luminosity function $\Phi(M, z)$. The luminosity function provides the

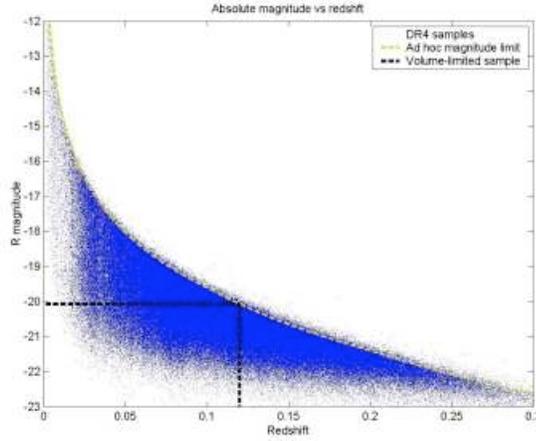


Figure IV.9: The blue points is an example of a flux limited sample from SDSS DR7. The points constrained to the dashed lines corresponds to a Volume limited sample that could be defined from the first one. As we observe from the plot a volume limited catalog implies the loss of objects and thus a loss of information.

density number of galaxies for each range of absolute magnitude (M). In the case of flux limited samples the selection function is unity for redshift close to zero and it tends to zero for large redshift, while for volume limited samples the radial selection function is unity throughout.

$$\phi(z) = \frac{\int_{M_{min}(z)}^{M_{max}(z)} dM\Phi(M)}{\int_{Min}^{Max} dM\Phi(M)} \quad (IV.26)$$

For example, SDSS LRG Early Data Release sample is limited in apparent magnitudes $14.5 \leq r^* \leq 17.6$ in absolute magnitudes corresponding to $-22 \leq M_{r^*} \leq -19$. Let us introduce some terms. We observe an object with an apparent magnitude. The absolute magnitude of this object refers to the apparent magnitude of the object if it was observed at rest at 10 pc. using an aperture that contains its total flux. In order to convert apparent magnitudes m to absolute magnitudes M you have to take into account two effects: the cosmological effects and the k corrections. The cosmological effects refer to angular diameter distance and the cosmological surface-brightness dimming. These effects are enclosed in the distance modulus $DM(z)$. The distance modulus accounts for the difference in magnitude due to the difference in distance. DM is the distance modulus defined as $DM(z) = m - M$. The K-correction takes in account that the measurement (or the spectrum) in the observed frame corresponds to a narrower and bluer rest frame passe band that depends of the redshift of the object. This k correction is dependent on the galaxy spectral energy distribution (SED). The expression of the absolute magnitude is given by:

$$M = m - DM(z) - K(z) \quad (IV.27)$$

Thus, for the example:

$$\begin{aligned} M_{min}(z) &= \max[-22, 14.5 - DM(z) - K(z)] \\ M_{max}(z) &= \min[-19, 17.6, -DM(z) - K(z)]. \end{aligned} \quad (IV.28)$$

Based on luminosity function computed in (Blanton et al., 2001) for the SDSS bands². The expected redshift distribution was computed given the r^* luminosity function and the flux limits using the k corrections for a "typical" galaxy color of $g^* - r^* = 0.65$. In Fig.[IV.10] the redshift distribution of SDSS galaxies is compared with the average distribution expected giving a reasonable agreement.

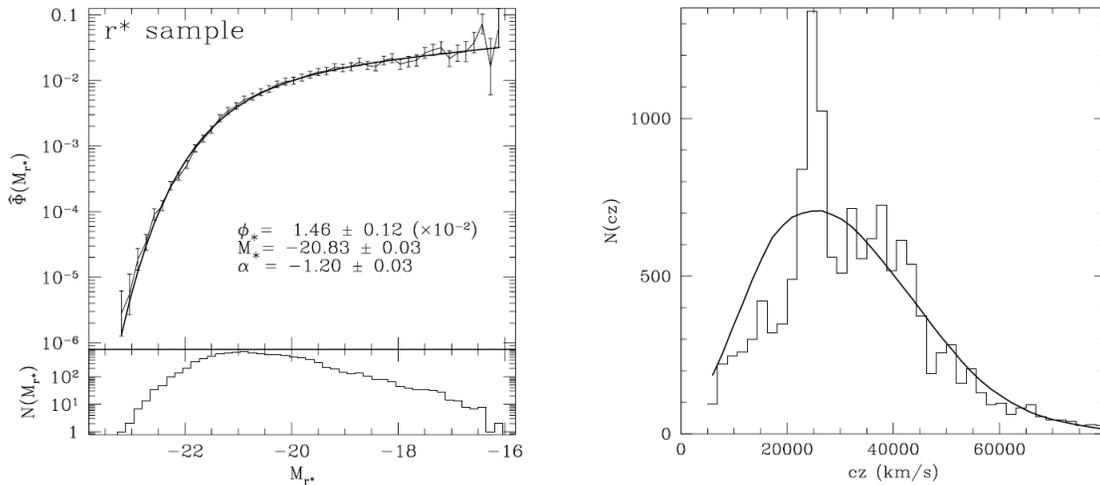


Figure IV.10: The luminosity function of SDSS-Early Data Release in the r^* band. The data points are commissioning data and the black line corresponds to the Schechter fit. In left panel the redshift distribution of SDSS (histogram) compared with the expected average distribution. Figures from (Blanton et al., 2001)

In some cases it is difficult to implement this standard way to estimate the expected number density, as for example in SDSS DR3 LRG sample the luminosity function is very steep, thus really sensitive to changes in k corrections. Thus instead, the procedure was to model the selection function, using the average spectrum observed convolved with the response function and then to apply a low pass filter and finally to fit this model to the observed redshift (Zehavi et al., 2005).

Instead of using these approximative models for the expected number density commonly, the radial selection function to be applied to the random catalogue is estimated directly from data. The method consists in using the redshift distribution from data and assign the redshift to the random set following the same redshift distribution. In practice, there are two ways commonly used: 1) redistribute the redshift of the data sample to the random sample. 2) resample the redshift distribution of data so that the overall $n(z)$ shape matches that of the data (smoothed to eliminate fluctuations from data).

As an example I take the samples used in (Kazin et al., 2010) to do large scale structure analysis. I take the *Full Sample* and the *Bright Sample*. The Full Sample is a quasi flux limited sample.

²The luminosity function is presented using a fit of the Schechter function characterized by three parameters $/\phi_i^* [10^{-2} h^3 Mpc^{-3}]$, M_* and α :

$$\Phi(M)dM = 0.4 \ln(10) / \phi_i^* 10^{-0.4(M-M_*)(\alpha+1) \exp[-10^{-0.4(M-M_*)}]} dM \quad (IV.29)$$

It consists of a set of ~ 100000 galaxies in the range in redshift of $z=[0.16, 0.36]$ and limits in absolute magnitude are $M_g = [-23.2, -21.8]$. The Bright Sample is a volume limited sample which consists in ~ 30000 LRG's in the redshift range $z=[0.16, 0.44]$, with $M_g = [-23.2, -21.8]$. In Fig.[IV.11]. I show in the left panel the number density of both samples . In the right panel I show the correspondent correlation function. The random catalog in the first case was generated using as the radial selection function the redshift distribution of the Full sample while the radial function is constant in the volume limited case .

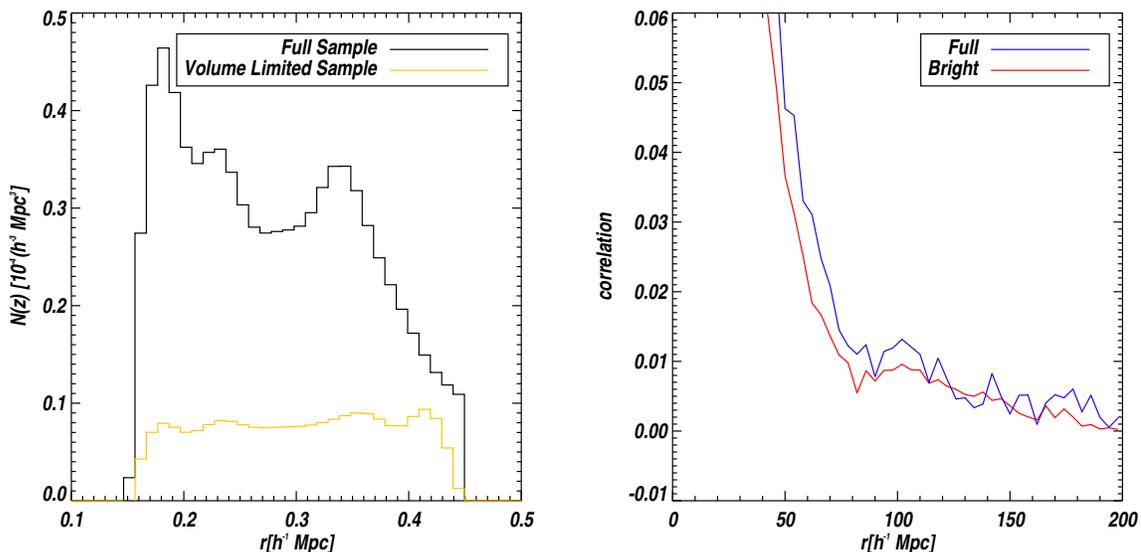


Figure IV.11: In left panel the number density of *Full Sample* and *Bright Sample* from samples defined by E.Kazin et al 2010. In right panel the correlation function using the samples .

In order to study the effect of radial selection in the estimator 100 realizations were generated taking different functions for the radial selection and comparing the resultant correlation function with a uniform selection function. We do not applied geometry effects. To sample the points accordingly to the desired radial distribution a rejection method was implemented. Let us begin comparing samples with the same number of objects in an spherical volume where the radial distribution follows different radial selection functions. In the left panel of Fig (IV.12) the comoving density of the samples using different radial selections functions (uniform, gaussian and parabolic) is shown. In the right panel of Fig.[IV.12] the average over 100 realizations following the different selection functions is shown. As we can observe from the fig IV.12 when the radial selection function from data is correctly applied to the random, the estimator is no sensible about the form of the selection function.

Now, let us compare the effect of the radial selection function with samples with the effective volume, defined as (Feldman et al., 1994):

$$V = (2\pi)^3 \int d^3r \bar{n} w^4 (1 + 1/\bar{n}P)^2 \quad (\text{IV.30})$$

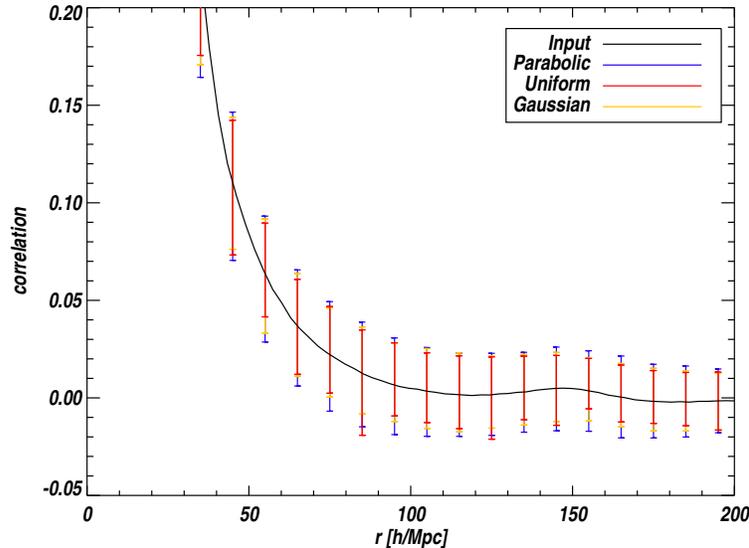


Figure IV.12: The average correlation function over 100 realizations of 30,000 objects following different radial functions. In blue the parabolic, in yellow the gaussian and in red the uniform. The input correlation function is shown in black.

IV.3.2 Angular Selection Function

In the real large scale surveys usually the geometry of the sample is complex, and because of observational and instrumental issues, the angular selection function is not trivial. The angular selection function keeps track of the survey parameter that introduces variations of the mean density as a function of the direction in the sky, completeness magnitude limit, seeing, dust extinction and geometry. Thus in order to get an accurate estimate of the correlation function these effects must be introduced in the random samples.

A scheme to deal with complex angular mask has been developed by Hamilton and Tegmark (Hamilton & Tegmark, 2004). A software package called **MANGLE**³ which implements this scheme is publically available. We have used this package to deal with the complex angular selection functions from SDSS data. In the context of **MANGLE** the angular selection function is replaced by the concept of *angular mask*. An *angular mask* as arbitrary union of arbitrarily weighted angular regions bounded by arbitrary number of edges. Each subregion of the angular mask is called a *polygon*. A polygon is a non overlapping region of the angular mask characterized by a constant weight. Each polygon is defined as the intersection of caps, where a cap is an spherical disk. The precise definition of a cap is a region on unitary sphere above some line of constant latitude with respect to an arbitrary polar axis. In Fig.[IV.13] an illustration of a region of an angular mask correspondent to BOSS survey defined in terms of polygons, the different colors show the weight of each region.

³The software and documentation are available at <http://casa.colorado.edu/~jsh/mangle/>

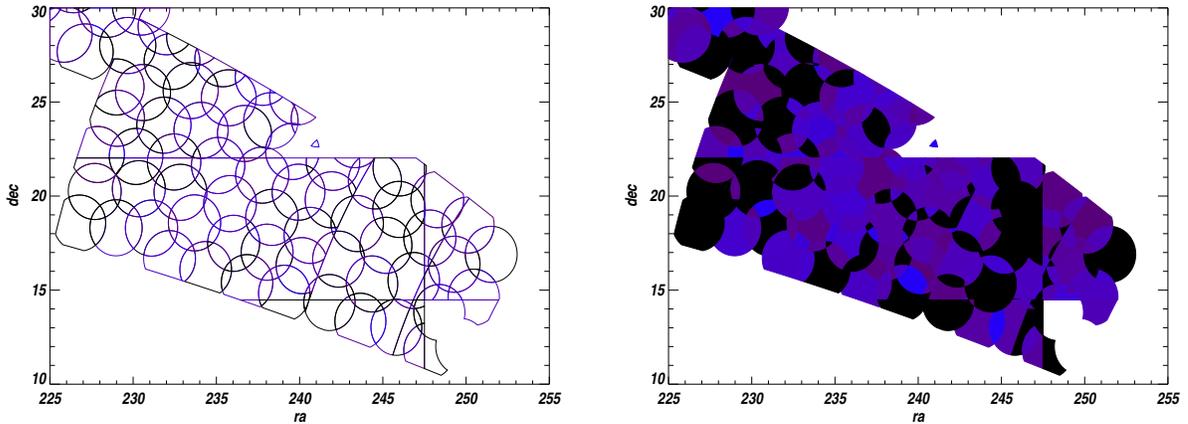


Figure IV.13: An example of a section of an angular mask from SDSS DR7. In left panel the mangle polygons defining the mask, in the right side the same polygons showing the completeness weight of each region correspondent to the value of the completeness of the sample in each region.

IV.4 Correlation function from Realistic Simulations

Until now we have explored separately the different aspects that could affect the estimator of the correlation function, in this section the idea is to mimic a realistic simulations. In order to do it we use the Luminous Red Galaxies LRG Sample of SDSS-DR7. This sample was used for large scale structure (LSS) analyses in (Kazin et al., 2010). The data set consists of a flux limited catalog of $\sim 100,000$ LRG. The survey covers a total area of 7908 deg^2 but we are going to use only the northern hemisphere. The total volume covered is $1.6h^{-3} \text{ Gpc}^3$.

The first problem that must be confronted is the volume simulated. In order to achieve a fine binning and a large volume, the periodicity property of log normal simulations is exploited. A grid of 256^3 with a physical size of 1.6 Gpc^3 is used giving a pixel size about $\sim 6 \text{ Mpc}$. To reproduce the volume and geometry of DR7, a cubic volume of 8 Gpc^3 would be generated, to achieve this volume the simulated box of 1.6 Gpc^3 will be replicated 8 times. The correlation function used to generate de log normal simulation follows a Λ CDM model with cosmological parameters of WMAP 5 generated with the public code CAMB (Lewis et al., 2000). A galaxy bias of $b = 2$ was applied, that is approximately the bias of LRG ($b=2.5$ from (Labatie et al., 2010), $b=2.3$ (Kazin et al., 2010).) In a first time, I am not going to use the completeness estimation for each region, only the survey geometry. An approximative angular window was constructed from the stripes from photometric survey (tiling geometry) and the plates used in DR7 concatenated using MANGLE tools. This angular mask considers a constant weight everywhere. It is possible to use the accurate angular mask prepared by (Blanton et al., 2005) for the DR7 of SDSS that is provided in terms of MANGLE polygons. Nowadays for the purposes of studying the effects of the geometry is enough to construct a simple mask what makes fast the mocks generation. A realistic mask for DR7 taking into account all details of the geometry (bad fields, mask of bright stars, etc) is described with about 200,000 polygons while in the simple approach the number of polygons reduced by 1 order of magnitude

($\sim 10,000$ polygons). To minimize time I decided to work only with the Northern hemisphere. In the left panel of Fig.[IV.14] the geometry of DR7 sample is shown in galactic coordinates. In the right panel I show the simplified mask for DR7 only considering the north hemisphere in equatorial coordinates.

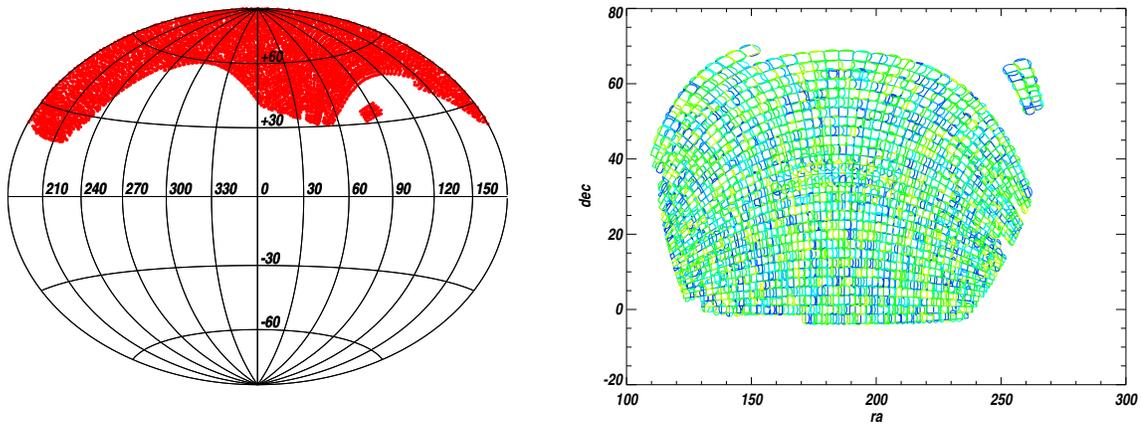


Figure IV.14: Left panel. The geometry of SDSS LRG sample DR7 in galactic coordinates. In the right panel the simplified mask used to generate the mock catalog is shown. It was generated using the tiles from DR7 concatenated with the photometric stripes.

For the radial selection function we use the redshift distribution of data. In Fig.[IV.15] the redshift distribution of data is shown. To construct the radial function the redshift distribution is normalized and then fitted by a spline to eliminate the noise to generate the selection function. Finally, a rejection method is applied to select from the mock the galaxies following the redshift distribution. In the right panel of Fig.[IV.15] the redshift distribution of the data set is shown and in left the comoving number density of mocks.

The first purpose is to verify if the estimator is sensitive to the effects of the geometry and edges. For that we use a set of 100 lognormal realizations for the same input power spectra in 2 cases, the cube that we have been using until now and the complex geometry of SDSS-DR7. In the Fig.[IV.16] the average correlation function over 100 realizations in a cube compared with 100 realizations mimic the geometry of DR7 is shown. As we can observe from the figure the estimator shows a slight bias of % at the scales of interest and a increment in the variance of the variance of % in the case of DR7 geometry compared with the cube.

IV.4.1 Completeness of the Survey

When a survey of galaxies is designed a set of objects is selected following certain criteria via a target selection algorithm to ensure a complete sample. But even if the design of the survey is optimal, for different reasons there would be certain rate of success in the spectra measured that would be function of the angular direction. The completeness of the survey is defined as the rate between good spectra in relation to the targets assigned by region of sky. In surveys in ongoing data-

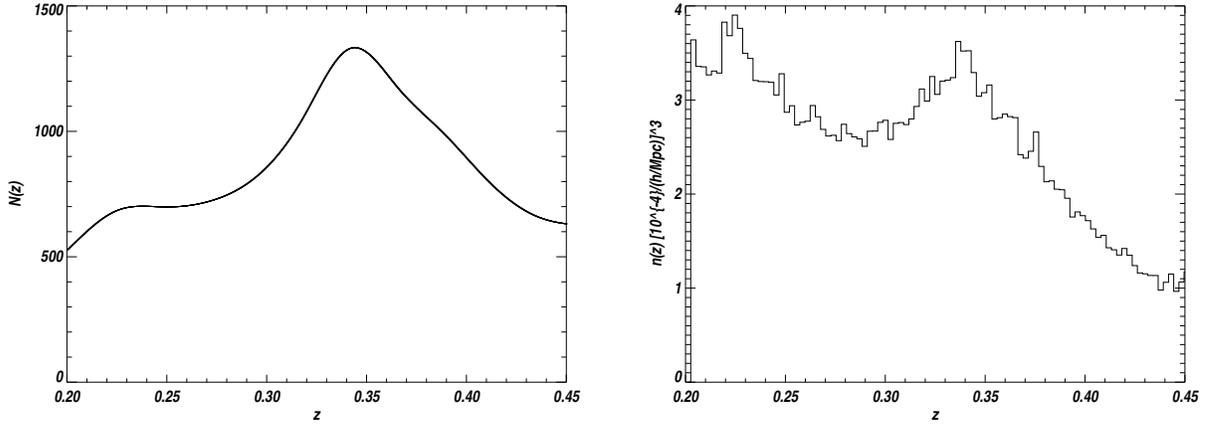


Figure IV.15: In left panel the redshift distribution of data fitted by a spline line to eliminate the noise is shown. This distribution is normalized and used to generate the radial selection function. In the right panel the comoving number density of data and mocks is shown.

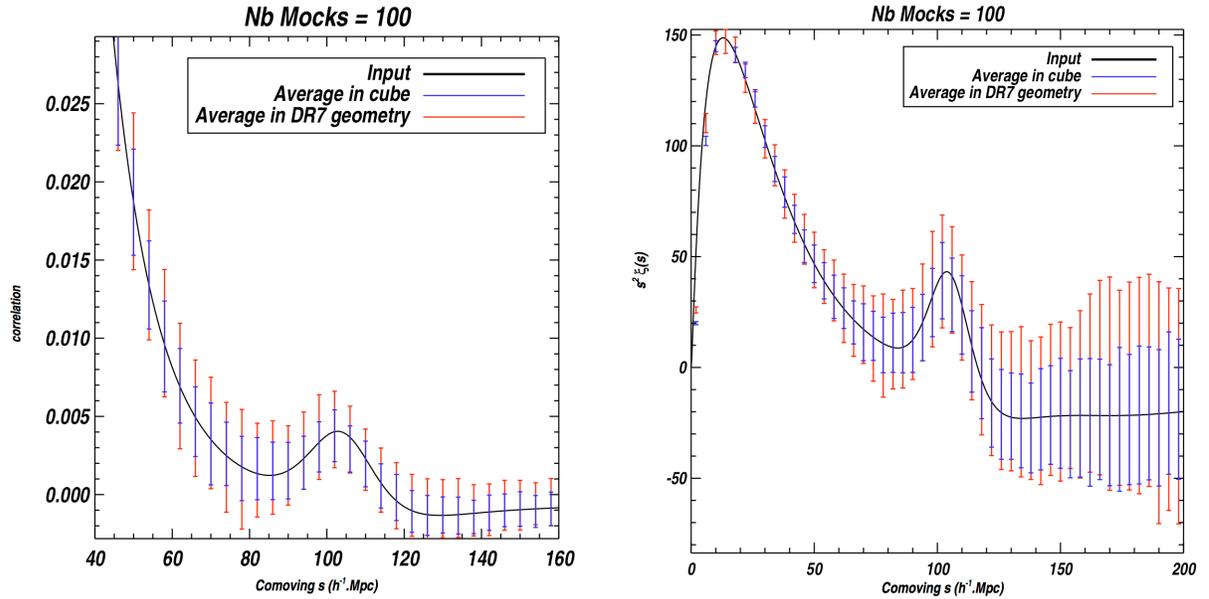


Figure IV.16: The average correlation function over 100 realizations of the sample in a cubic geometry compared with a set of samples following DR7 geometry. As we can observe from the figure the estimator shows a slight bias at the scales of interest and a increment in the variance in the case of DR7 geometry compared with the cube.

taking, the completeness of the survey must be taken into account for doing correlation function analysis. The way to account for the incompleteness of the survey is to assign a weight to each polygon proportional to the completeness on that region and to sample the random set following these weights. We will show with the simulations how the incompleteness affects the correlation

function. In Fig.[IV.17] a region of DR7 is shown, the colors corresponds to completeness range [0.97,1.]. As we can observe in this region the completeness is high , in fact the average completeness over the survey is 98% (Kazin et al., 2010).

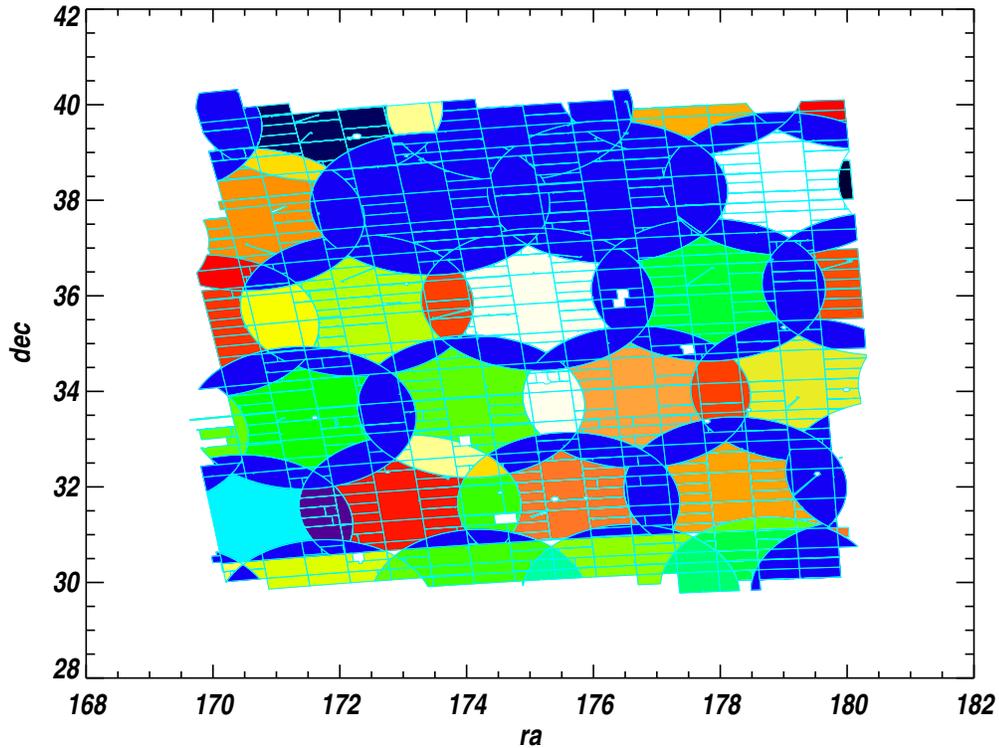


Figure IV.17: The completeness of a region of DR7 Sample using the file generated by NYE value added Catalog (Blanton et al., 2005).

Thus as LRG sample is almost a complete sample in this case the effect of taking into account the completeness in the random generation is negligible. To verify this fact let us compute the correlation function using a random sample which completeness is set to one in all polygons and a correlation function using the correct value of the completeness. In order to do that we will use the mask prepared by Blanton et al (Blanton et al., 2005) publicly available in the NYE value added catalog. As we observe in Fig.[IV.18] the effects of incompleteness of LRG sample for DR7 is not significative . It is within the error bars computed by Kazin from N-body simulations in (Kazin et al., 2010). By contrary, we will see in the further chapter, if we do not considerer completeness of the survey in BOSS it would have an important effect in the correlation function estimator.

Finally before passing to the next section let us compare in this most realistic scenario the two estimators that seemed to behaved similarly that is the Hamilton and the Landy Szalay estimator. It has been said that the superiority of Landy-Szalay has been proven to be nearly of minimal variance for a random distribution (i.e. Poisson with no correlation) since the original paper in 1993 (Landy & Szalay, 1993). A similar study have been done by(Labatie et al., 2010) with lognormal

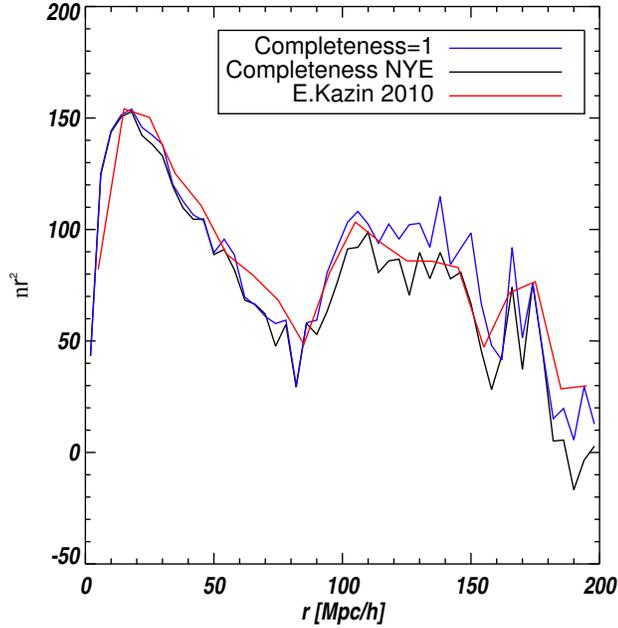


Figure IV.18: The correlation function of DR7 Sample data (Kazin et al., 2010) using the angular mask generated by NYe value added Catalog (Blanton et al., 2005) considering the weights given by the completeness in black, considering a completeness equal to 1 everywhere in the window in blue and as a reference the results from E. Kazin (Kazin et al., 2010).

mocks that mimics SDSS main samples conditions (volume limited samples), probing the superiority of Landy-Szalay estimator over the others estimators. In Fig.[??] both estimators LS (black) and H (red) are computed for a realistic scenario, including radial and angular mask of DR7. The two estimators are indistinguishable.

IV.4.2 Radial Weights

It is possible to optimize measurements giving to each galaxy or fraction of space an specific statistical weight chosen to minimize the cosmic error. If we neglect edge effects, considering a constant weight only function of r it has been shown in (Hamilton, 1993) that the optimal weight is done by:

$$w(r) \propto \frac{1}{\sigma^2(r)} \quad (\text{IV.31})$$

where $\sigma(r)$ is the cosmic error of the considered statistic, assuming the gaussian limit that is valid only in the weakly nonlinear regime leads to the following weight for the two point correlation function:

$$w_i(r) = \frac{1}{[1 + \bar{n}_d(r)J(r)]^2} \quad (\text{IV.32})$$

where

$$J \sim \int_0^r \xi(r) dV \quad (\text{IV.33})$$

These weights depends of the pair separation. In (White et al., 2011) this weighting scheme has been used and the ξ has been approximated by a power law.

Another common scheme of weights to optimize the correlation function estimator was proposed by Feldman-Kaiser-Peacock (FKP)(Feldman et al., 1994):

$$w_i(r) = \frac{1}{[1 + \bar{n}_d(r)P(k)]^2} \quad (\text{IV.34})$$

This weighting scheme is used in Eisenstein et al. (2005). These weights are independent of the pair separation, they only depend on the redshift. Thus, they have the advantage of not introducing any scale-dependence to the weights. For this reason I consider only the FKP weights for the following.

In the Fig.[IV.19] I show some examples of the correlation function of single realizations of the log normal simulations following the properties of DR7. It is difficult to conclude with only one realization as the fluctuations related to the shot noise and cosmic variance dominate.

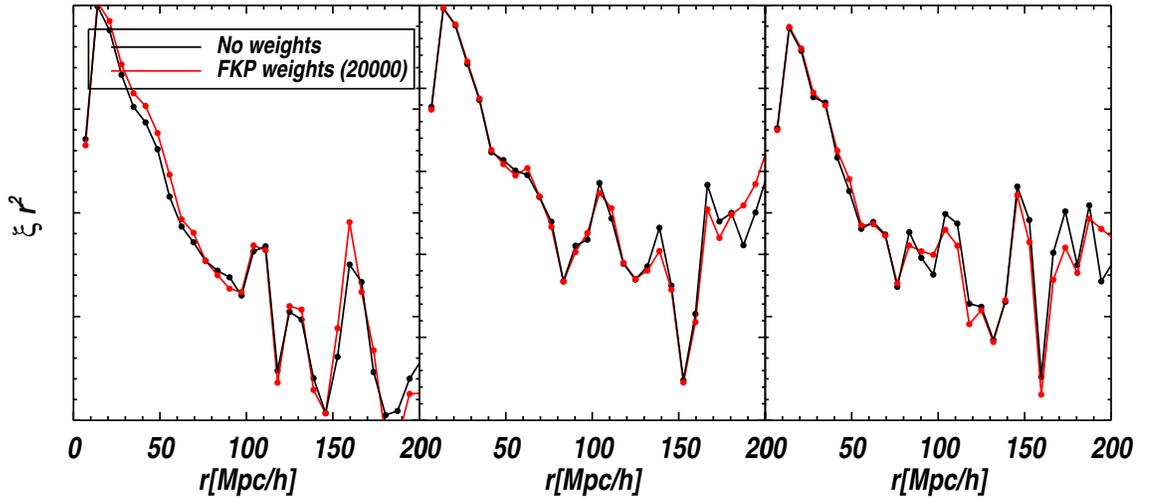


Figure IV.19: In correlation function computed with no weights (black) and with FKP weight (red) fore single realizations.

In Fig.[IV.20] I present the effect of the weights for the case of 100 simulations following DR7 geometry and completeness. The correlation function is computed with no weights (black) and with FKP weights (red) The FKP weights are calculated fixing $p(k)=20000$. As we observe from the figure the FKP weights affects only the large scales giving less than 15% difference for scales larger than 100 Mpc/h. Our results are consistent with results found by (Kazin et al., 2010) where it was found the radial weights has a marginal effect in the correlation function for DR7.

To understand the behavior of the weights, let us think in density fluctuations at a scale $\lambda = \pi/k$. As $\bar{n}(r)$ is a decreasing function of r , thus we can separate 2 limit behaviors. At small scales (r

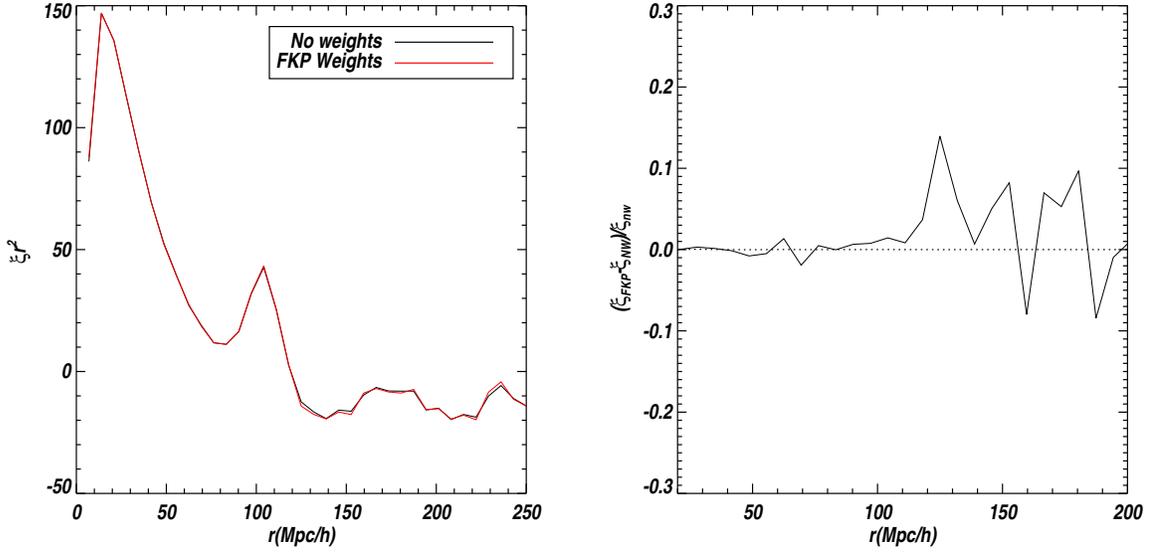


Figure IV.20: In right panel the mean correlation function computed with no weights (black) with FKP weight (red). In right panel the relative effect of applying the FKP weights.

small), we have many galaxies per volume, so the error will be dominated by cosmic variance (finite number of volumes λ^3). The idea is to give equal weight per volume what is equivalent to give a weight proportional to $1/n_d$. For large scales we have less galaxies, we are dominated by Poisson noise so we want to weight galaxies equally.

IV.5 Error Estimation

In the case of correlation function we will see that the different scales are correlated, neglecting off diagonal terms could lead to substantial overestimate of the constraining power of observations. Thus the treatment of covariance between measurements at different scales would be necessary to properly test theoretical prediction. Let us remind the definition of covariance matrix. Considering two estimators \hat{A} and \hat{B} we can generalize the concept of variance to introduce the concept of the covariance between two estimators:

$$\text{Cov}(\hat{A}, \hat{B}) = \left\langle \left(\hat{A} - \langle \hat{A} \rangle \right) \left(\hat{B} - \langle \hat{B} \rangle \right) \right\rangle \quad (\text{IV.35})$$

Applying this concept to the same estimator at different scales we get the covariance between different scales of an estimator, for example, in the case of ξr we get:

$$C_{ij} \equiv C(r_i, r_j) = \left\langle \xi(\hat{r}_i) \hat{\xi}(r_j) \right\rangle - \left\langle \hat{\xi}(r_i) \right\rangle \left\langle \hat{\xi}(r_j) \right\rangle \quad (\text{IV.36})$$

The cross-correlation coefficients are defined as:

$$r_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} \quad (\text{IV.37})$$

In principle given an estimator and a model of galaxy clustering it is possible to predict analytically the statistical errors for any geometry and selection function. However, as we have seen in the previous sections, such error estimates can be cumbersome to compute and they depend on the assumed clustering model itself. Thus alternative ways of estimate the error are implemented in clustering analysis. In this section I show the error estimation using different approaches applied to DR7 data. The first approach that I present is using simulations. The simulations are the preferred way to assess the consistency of the model with the data. These estimations of the errors depend on the assumed clustering model itself. For some purposes it is desirable to have estimates of statistical errors and their covariances that depend only on the data set itself. Thus I present some of the methods commonly used to estimate errors from data as bootstrap, jackknife, etc. The purpose would be to compare the different methods in order to define the error estimation that would be used when analyzing BOSS data. The structure of the section is the following, I present the different methods showing at the same time the results obtained for each one. In the final part of the section these results are compared.

IV.5.1 Errors from Simulations

One possible way to estimate the errors consists in, given a clustering model, generating a large number of realistic realizations of mock catalogs using Monte Carlo methods and measuring the dispersion of the measurements. In this way errors automatically take account of non gaussianity of the galaxy distribution, cosmic variance, shot noise and the geometry of the survey. To generate samples with realistic statistics we have many possibilities more or less approximative the most exact would be to use N-body simulations. N-body simulations have the advantage to follow the fully non linear and thus non gaussian evolution of structure. However this method would be constrain by the computer limitations . Another alternative is to use lognormal simulations that have the advantage to be generated in a fraction of time in relation to the exact N-body simulations. This kind of simulations allow us to study accurately the clustering properties of galaxies. This errors automatically take into account cosmic variance, shot noise, and the geometry of the survey and some systematics. Nowadays they do not take account of the fully non linear behavior (i.e. they do not take into account the non gaussianity of the galaxy distribution). There exists another kind of simulations based in second order perturbation theory (2LPT) conjugate with semi-analytical models of structure formation, so called PTHalos created by (Scoccimarro & Sheth, 2002). This kind of simulations approximates very well the fully non-linear, thus non-Gaussian, evolution of structure in a very small fraction of the time and cost of a full n-body simulation. In this chapter we will only present the result of the log-normal simulations. In the following chapter devoted to BOSS data we will use a set of simulations based on PTHalos ⁴ and a set of realizations of N-body simulation *Lasdamas*.

Let us begin showing the result with log normal simulations. The covariance matrix and the correlation matrix have been computed using 100 lognormal simulations with the simplified angular

⁴These simulations have been generated within BOSS coloboration *Manera et al in prep*

mask of DR7 and the radial selection function already described in the last section. In Fig.[IV.21] the resultant covariance matrix and the correlation matrix coefficients are shown. The covariance matrix has a enhanced diagonal and some correlation between off-diagonal elements as expected. as we have said the simulations method provide the most reliable estimation of the errors. We will compare this result with the data-based methods that will be implemented in the following subsection.

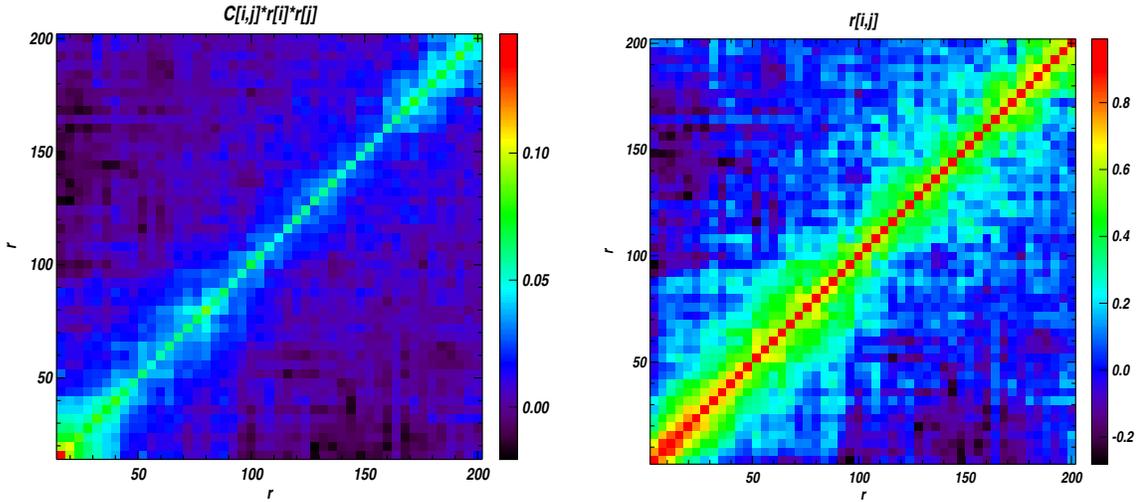


Figure IV.21: The covariance matrix (left) and the correlation coefficients matrix (right) using 100 log normal simulations using the radial selection function from DR7 data and the angular selection function with weights equal to the completeness of the survey using files from the NYE Value Added catalog.

IV.5.2 Error from Data

In this section I present some common recipes based on the data itself for the error estimation in clustering statistics. I summarize in the following list the methods and results obtained in each case.

- *Bootstrap*. It is a resampling method consisting in drawing randomly original data set , N points with the possibility of re-emplacment. The random resample is defined on the set of all variations of data repetitions allowed. In (Snehlage, 2000) it has been shown that this method does not lead to reliable estimates of cosmic error for point process. To apply this method 100 subsamples of 10,000 objects with repetitions allowed have been generated. The error bars are computed using the empiric standard deviation. In plot IV.22 is shown the average of the 100 realizations with the error bars given by the standard deviation. We observe the mean is in agreement with the measured correlation function from (Kazin et al., 2010). In Fig.[IV.23] I show the resultant correlation matrix and the covariance matrix multiplied by r^2 .

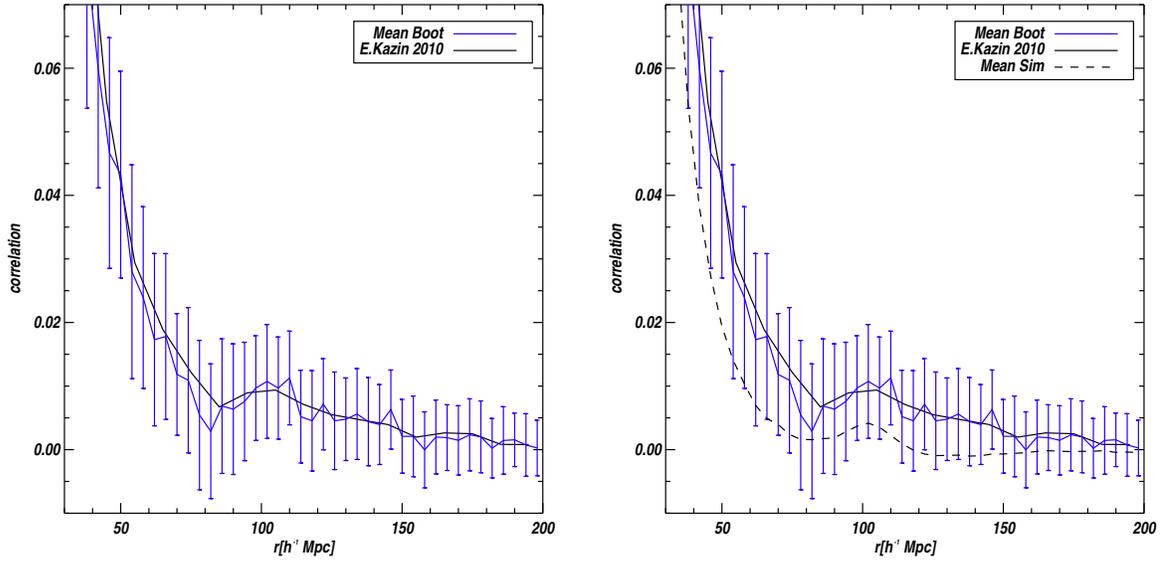


Figure IV.22: In blue the mean correlation function from 100 bootstrap subsamples of 10,000 each. The error bars are given by the standard deviation. In black the correlation function from the whole sample.

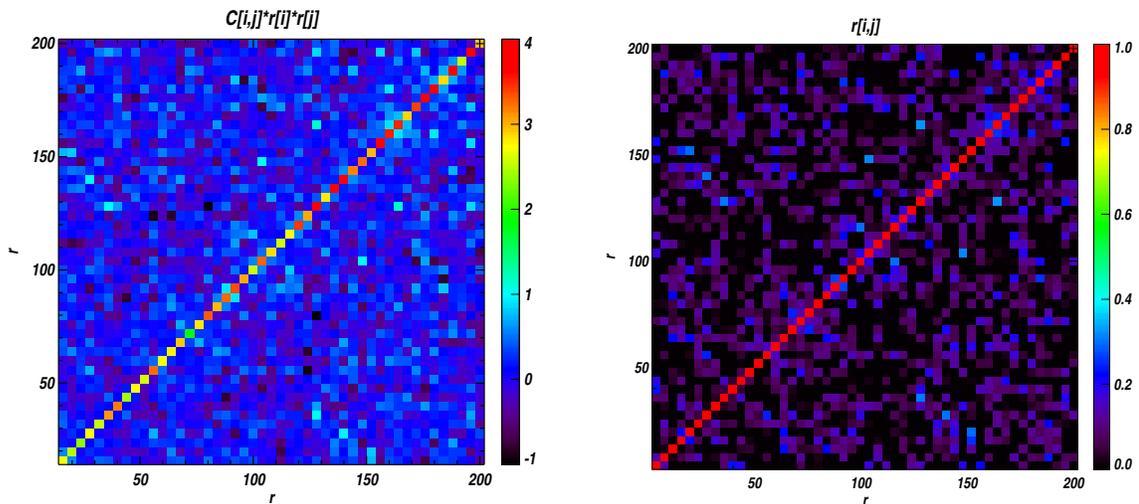


Figure IV.23: The covariance matrix (left) and the correlation matrix (right) calculated using the bootstrap method. The whole data sample is subsampled with repetitions allowed generating 100 realizations of 10,000 objects.

- *Subsample Method.* This method consists in dividing the catalog in a number N of smaller subsamples of the same volume and compute the dispersion measurements. This method has been proven to overestimate the errors (). To implement this method I divide the sample region in 10 subregions of approximately same area. I show the regions in left panel of Fig.[refsubs]. The regions are not spatially separated thus the border points must be correlated. In right panel of Fig.[IV.24] I show the correlation function for each subsample. As we can observe from the figure, given the small size of subregions there are large fluctuations. In the Fig.[IV.25]

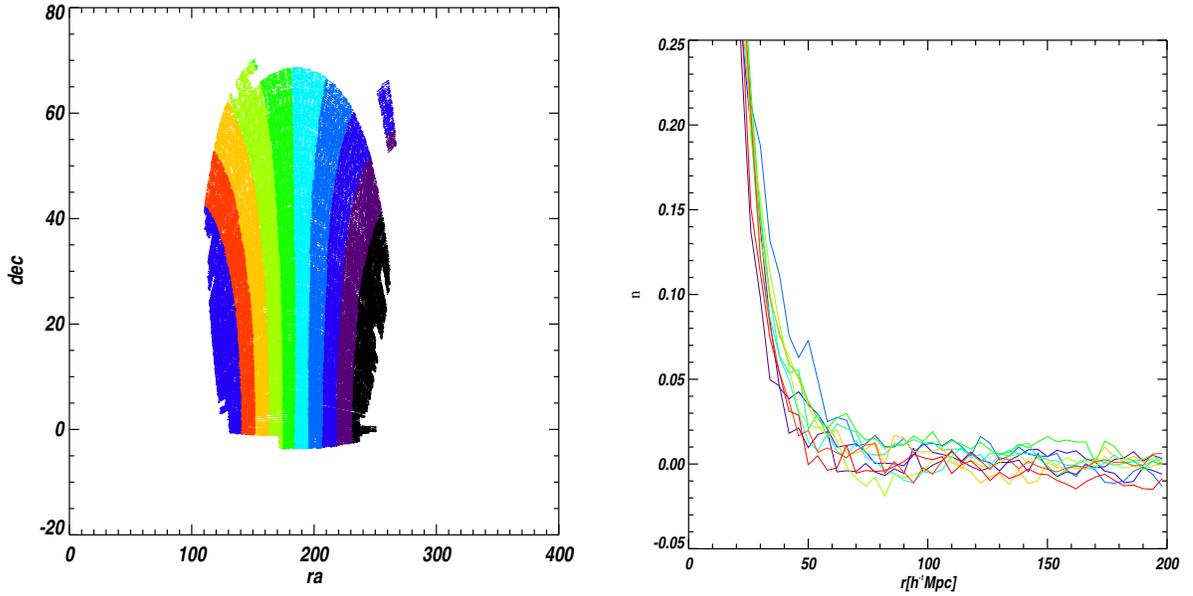


Figure IV.24: In the left panel, the regions correspondent to each subsample used to estimated the error via the subsampling method. In right the correlation function of each subsample.

I show in the left panel the mean of the 10 subsamples with the error bars given by the standard deviation and in the right panel how it compares with the correlation function of the whole data-set. By contrary to the bootstrap method, in this case, compared with the (Kazin et al., 2010) the mean of the subsamples is under-estimated. Finally in Fig.[IV.26] I show the resultant correlation matrix and the covariance matrix multiplied by r^2 .

- *Jackknife* This method is a variant of the subsample method. The complete sample is divided into m separate regions on the sky of approximately equal area. Then, the m subsamples are analyzed removing the i^{th} subsample. The covariance is then estimated:

$$C(\xi_i, \xi_j) = \frac{N-1}{N} \sum_{l=1}^N (\xi(r_i) - \bar{\xi}(r_i))(\xi(r_j) - \bar{\xi}(r_j)) \quad (\text{IV.38})$$

The error estimated using this method is reasonably well if the sub-sample measurements are not strongly correlated. Accordingly to (Scranton et al., 2002) and (Zehavi et al., 2002) this method gives a good estimate of the cosmic error at large scales compared to results

from mock catalogs in the context of angular clustering in SDSS galaxy sample. To applied this method I used the regions defined for the sample method but this time each region will be removed giving the 10 jackknife subsamples. In Fig.[IV.28] the mean of the correlation function is shown with the error bars given by the diagonal terms of the covariance matrix. In the right panel the mean compared with the correlation function compute in (Kazin et al., 2010). The mean over the jackknife subsamples is slightly under-estimated. Finally in left panel of Fig.[IV.29] I show the resultant the covariance matrix multiplied by r^2 and correlation matrix (right).

These methods are not accurate for estimate of finite volume effects at the scale of the survey (Cosmic variance) since they are based on only a single realization of the density field at this volume is available. This can only be achieved by using the analytic expressions of the cosmic error for the estimator or using simulations.

IV.5.3 Comparison of Covariance matrices

We have computed the covariance matrix using four different methods, log normal simulations, bootstrap, subsample and jackknife methods. We have seen that all methods produce covariance matrices that have off diagonal elements as expected. In Fig.[IV.30] the covariances matrices generated from the different methods are plotted at the same scale. In botton from left to right, we have the result from simulation and the subsample. In the top the bootstrap and the jackknife. We

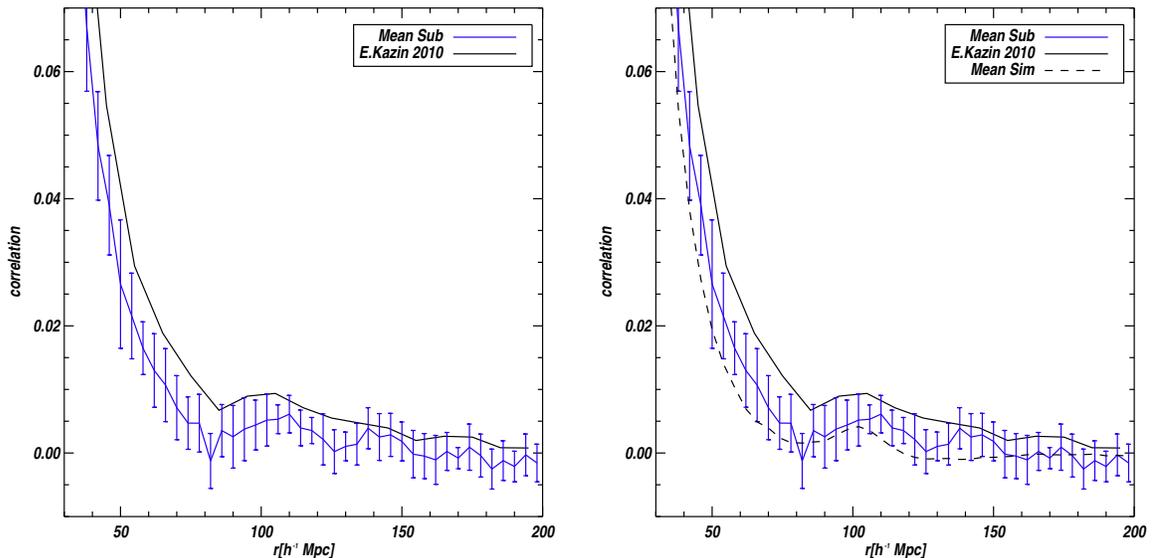


Figure IV.25: In the left side the mean over the 10 subsamples with the error bars given by the standard deviation. In the right side the mean compared with the correlation function computed with the whole sample (Kazin et al., 2010)

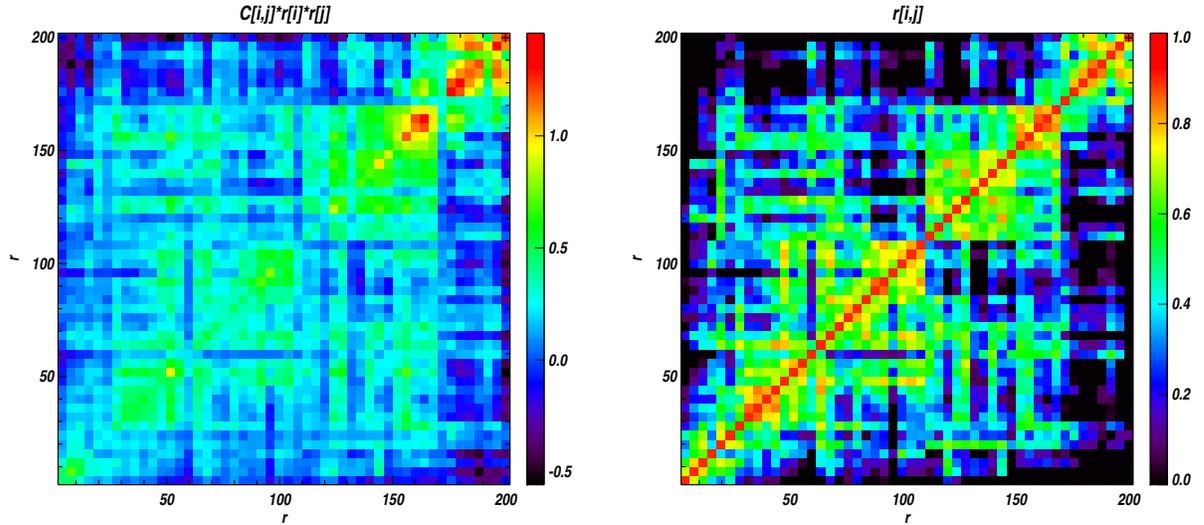


Figure IV.26: The covariance matrix (left) and the correlation matrix (right) using Subsampling method. The whole data-set has been divided in 10 non overlapping regions and the empiric variance have been computed from the samples.

observe that the bootstrap and the subsampling method generates noisy covariances matrices. In the case of jackknife even if it is the less noisy of data error methods presents additional correlation in the covariance matrix compared with simulations.

To do a detailed comparison of the covariance matrices there is no a single well established method. I will follow (Scranton et al., 2002) where a number of test have been done to compare the shape and amplitude of covariance matrices in the context of SDSS data. From the tests proposed I decided to applied the following three:

- Correlation test. The correlation coefficients permit us to qualitatively compare the shape of the covariance matrices . In Fig.[IV.31], the diferent correlation coefficients matrix at the same scale. In botton from left to right, we have the result from simulation and the subsample. In the top the bootstrap and the jackknife. The same trend observed in the covariance matrices, the bootstrap and subsamples methods give noisy correlations matrices even if they show a clear diagonal structure. The jackknife approximates the result from simulations but presents extra structure
- Diagonal Test. A more quantitative test consists in comparing the diagonal terms of the covariance matrix, $\Delta\xi(r)$ and $\Delta\xi/\xi$. In Fig.[IV.32] the different estimation of the errors is shown given by the diagonal terms of the covariance matrix. In Fig.[IV.33] the errors $\Delta\xi(r)/\Delta_{sim}\xi(r)$ given by the different methods from the diagonal terms of the covariance matrix compared with the errors given by simulations are shown. We observe the bootstrap method present the larger error bars followed by the subsample method and finally the jackknife gives very

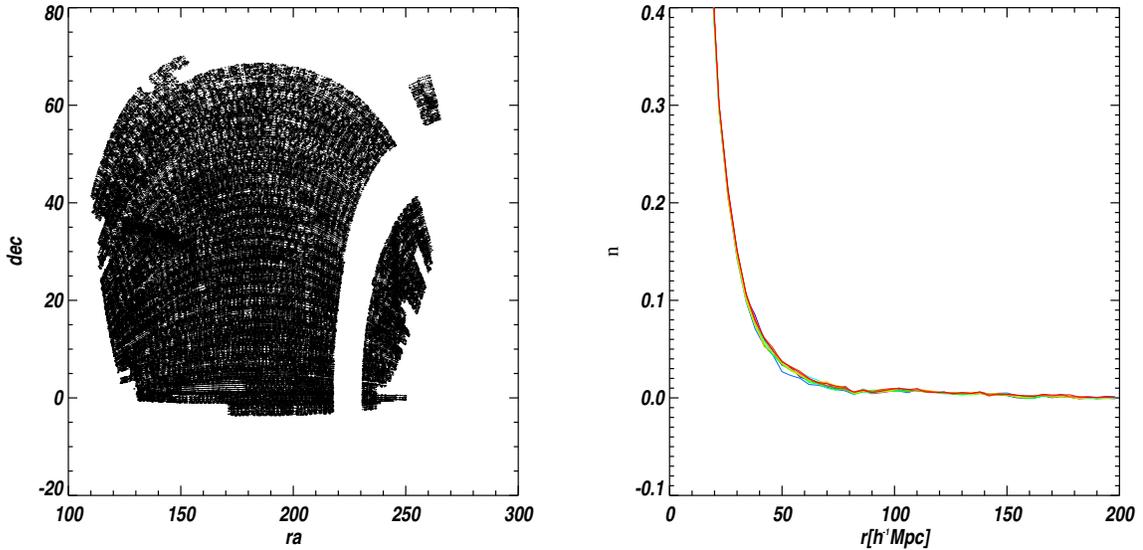


Figure IV.27: In the left panel, the regions correspondent to each subsample used to estimated the error via the jackknife method. In right, the correlation function of each subsample. As we observe compared with the Fig.[IV.24] the fluctuations of the correlation function are smaller given the size of the subsamples.

similar values to the simulations that gives the smaller error bars.

- Product Test. A more complete comparison of the covariance matrix is to include in the comparison the off diagonal terms. In order to include the off-diagonal terms the quantity $R(r)$ defined as follows :

$$R(r) \equiv \left(\prod_{\beta}^N |r_{(i,j)}| \right)^{\frac{1}{N}}$$

(IV.39)

This quantity for a perfect diagonal matrix will be zero and will be equal to 1 for perfect correlation (or anti-correlation) between each bin. In right panel of Fig.[IV.33] the quantity $R(r)$ is plotted for the different methods. We observe that the jackknife and the subsamples present the larger off diagonal correlations compared with bootstrap and simulations that signals a strong diagonal structure.

From the test performed to the different error estimations I conclude that methods based on data give noisy estimations of the covariance matrices. The extra structure found could be generated by the correlation between the borders of the regions defining the subsamples for the case of jackknife and subsamples methods. The jackknife is the method that presents the best properties among

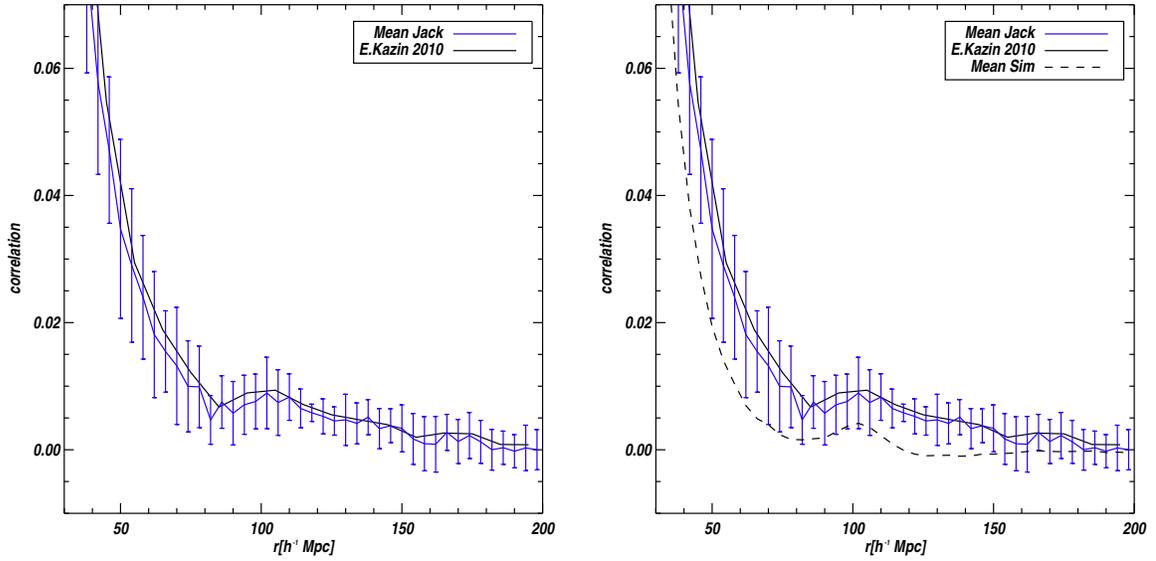


Figure IV.28: In the left side the mean over the 10 jackknife subsamples with the error bars given by the standard deviation. In the right side the mean compared with the correlation function computed with the whole sample (Kazin et al., 2010)

the data based error methods but even though it shows some structure. In general the data error methods overestimate the errors. In the case of simulations, the mean of simulations is underestimated. An improvement to the input correlation function should be applied to get a mean closer to the data mean or a rescaling should be performed.

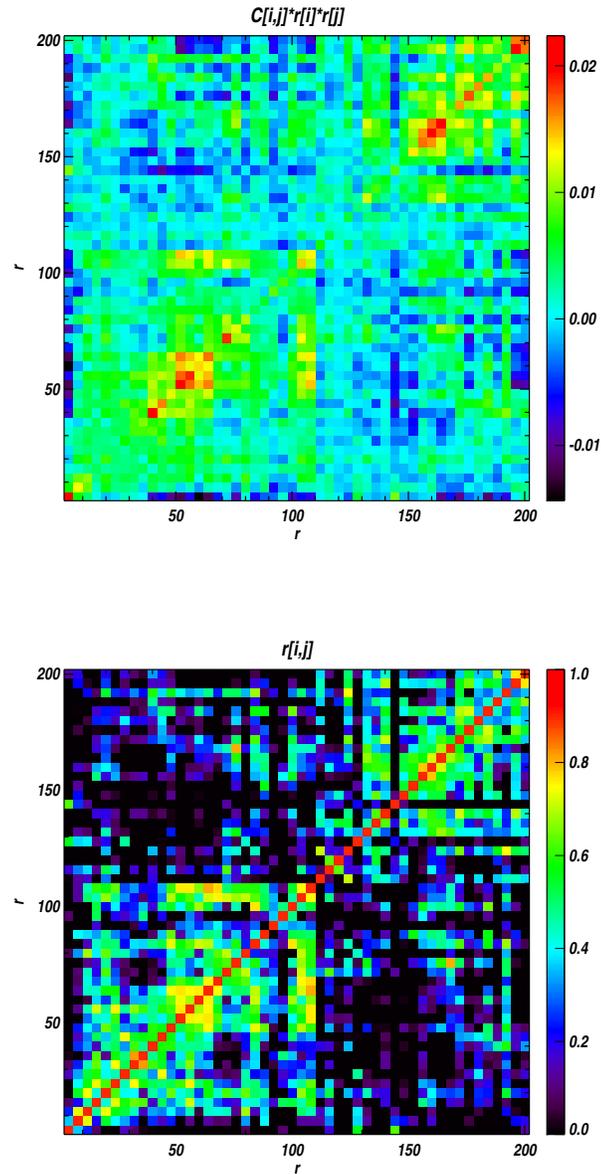


Figure IV.29: The covariance matrix multiplied by r^2 has been computed following the Eq.[IV.38]. The whole data-set has been divided in 10 non overlapping regions, and each subsample is generated removing the region i (Jackknife method).

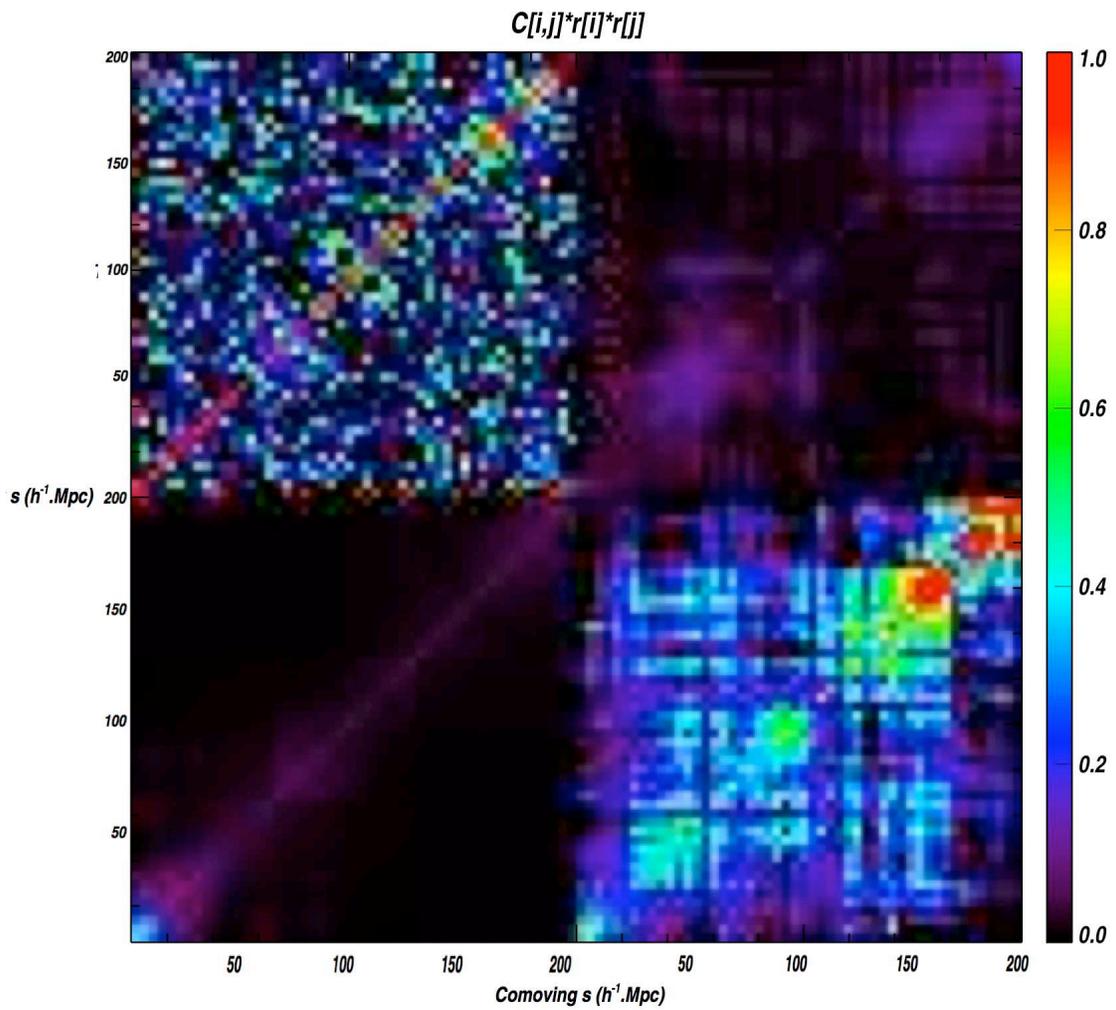


Figure IV.30: The different covariance matrix at the same scale. In bottom from left to right, we have the result from simulation and the subsample. In the top the bootstrap and the jackknife.

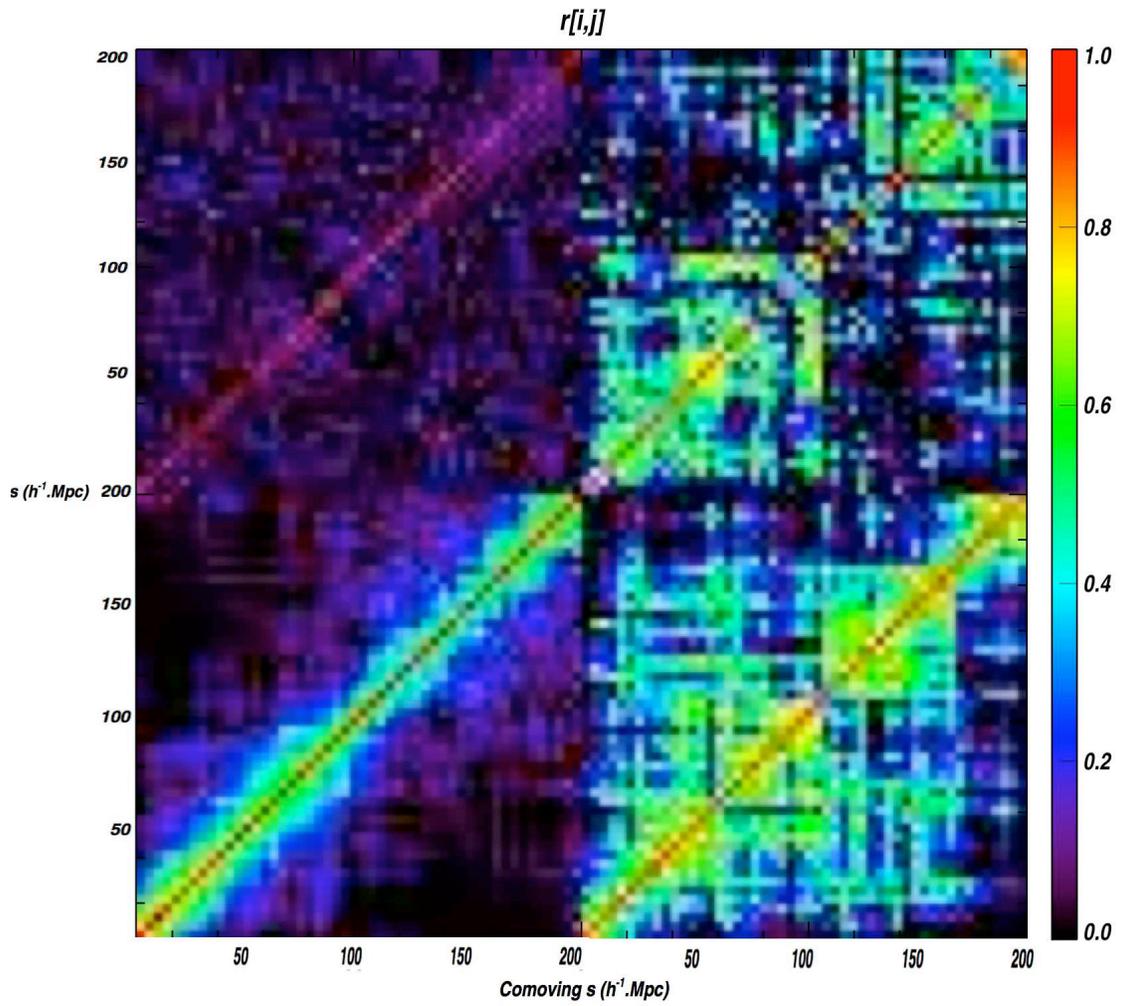


Figure IV.31: The different correlation coefficients matrix at the same scale. In bottom from left to right, we have the result from simulation and the subsample. In the top the bootstrap and the jackknife.

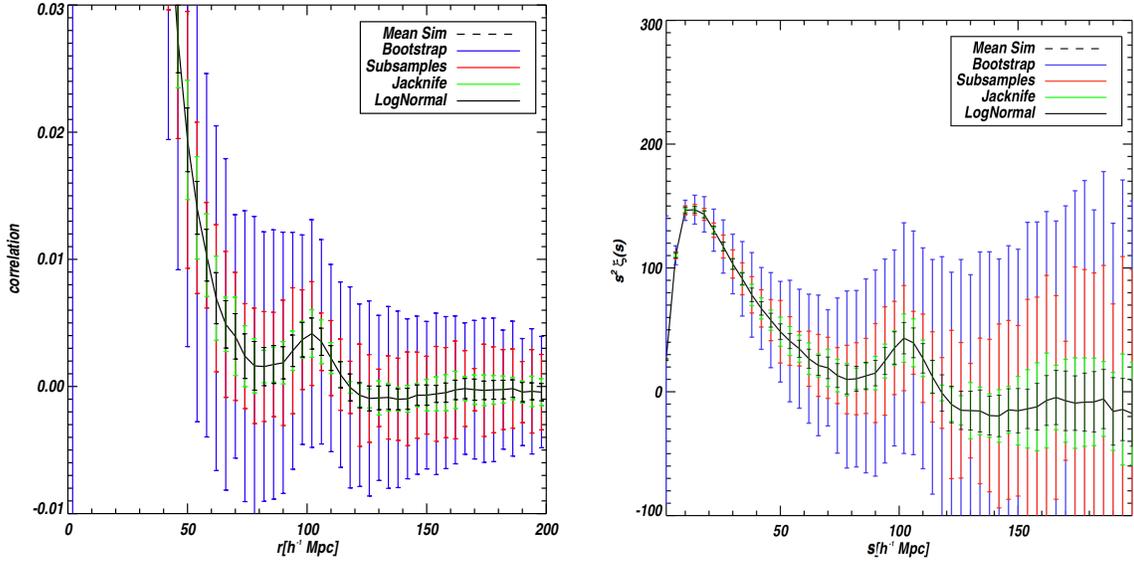


Figure IV.32: The errors given by the different methods from the diagonal terms of the covariance matrix.

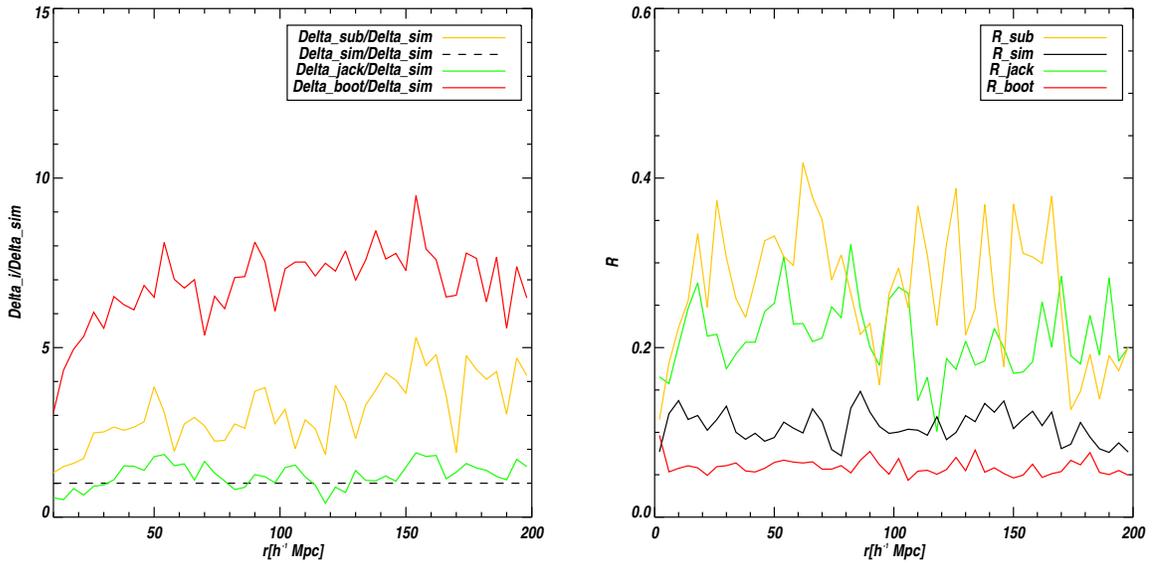


Figure IV.33: In left panel the errors coming from the diagonal of the covariance matrix compared with errors from the simulations lognormal. In right the $R(r)$ term defined as Eq. V.10 to compare quantitatively the errors coming from the off-diagonal terms.

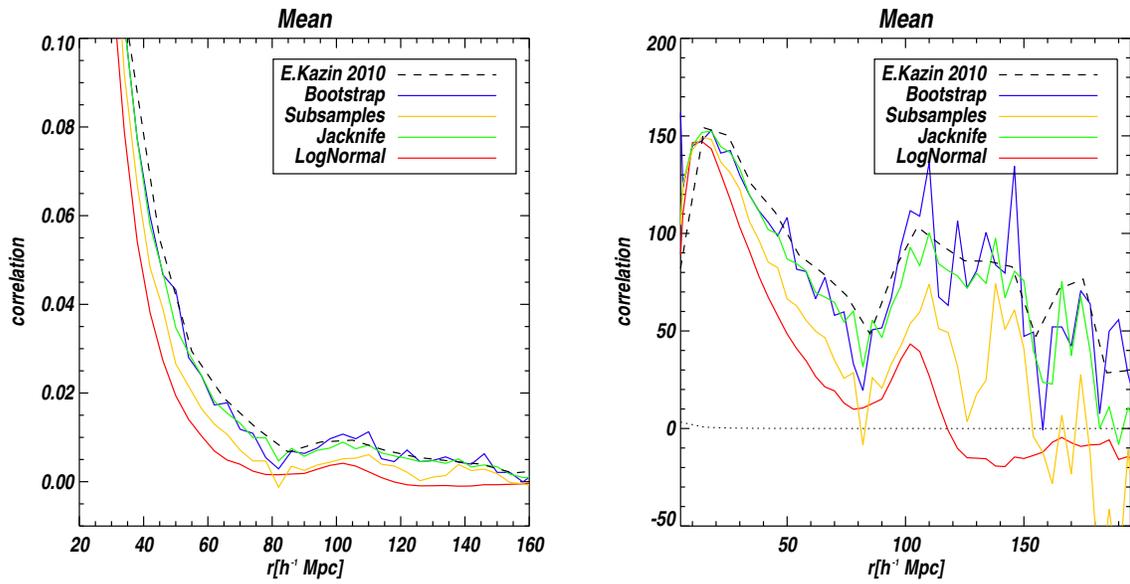


Figure IV.34: The means of the correlation function for each method, with which the covariance matrix is computed. As a reference de result from E. Kazin 2010 in black dashed line.

V. Large Scale Correlation Function from BOSS Data

V.1 Measuring the Correlation Function from BOSS Data

In this chapter, I use the first two years of BOSS data to compute the correlation function of galaxies. These data constitute the Data Release 9 (DR9) that would be public in summer 2012. It contains $\sim 500,000$ galaxies with good spectra, in a redshift range of $[0,0.8]$, covering 3,500 square degrees. Two different samples are generated using the BOSS data. The LOWZ sample, that is essentially an extension of SDSS-II LRG sample and the CMASS sample, a mass limited sample. This chapter focus on the CMASS sample which has ~ 350000 massive galaxies. The mean density is $\bar{n} = 3.5 \times 10^{-4} h/Mpc$ at $z=0.52$. The CMASS sample represents the largest sample of the Universe surveyed at this density [Paper 1 in prep].

The first step to do clustering analysis is to generate a 'clean' catalog from noisy data. In the first section I discuss this process. The definition of a catalog is not a simple task because we must guaranty that the sample selected is homogeneous and there is no systematic errors or bias originated by observational issues. The basic catalog generation consists of several stages beginning with the target list determination, spectra selection, correct accounting of legacy data (data from SDSS I-II), correction from fiber collision and redshift failures and finally the completeness determination. This catalog would be compared with the official catalog issue from the clustering galaxy group where I was involved.

The second part is devoted to the computing of the correlation function and the correction from angular systematic effects. The principal systematic error found in BOSS data is the stellar density (Ross et al., 2012). I explore in detail this source of systematic error. I study the effects on the correlation function of other observational variables as the seeing, extinction, airmass and sky flux. Some of these variables have been found to contribute to the systematic error of the correlation function, and even if their contribution is minor in respect to the stellar density, they are taken into account for computing the weights. In this section two methods to estimate the corrections and weights have been explored: the iterative weighting scheme suggested within the working group and the linear regression method that I implemented. The weights computed with these two methods are compared giving very similar results. Finally I compare the methods with the official method

to compute the weights based on MCMC developed by A.Sanchez within the working group.

Finally, in the third part I discuss the error estimation. As we have seen in the last chapter the common method to estimate the errors is computing the covariance matrix from simulations. In this section I explore three different kinds of simulations that would permit to validate the error estimation. The first kind of simulations, the log normal mocks, have been presented in the chapter 4. The second type of simulations, called PTHALOS simulations, are generated using second order Lagrangian Perturbation theory. They were generated within working group and they are documented in Manera et al and finally the Nbody simulations, Las Damas, that covers the geometry of the first semester of BOSS data only. They are documented in Cameron et al. Finally to close this chapter an analytical approach documented in Roland Putter et al in prep would be compared with the result of error estimations from simulations.

Data

The catalog was constructed using the data available from January 2010 up to the summer shut down of 9 July 2011. This spectroscopic data set will form the Data Release 9 in July 2012. There are 808 survey-quality *survey* plates, and 20 survey-quality *commissioning* plates containing approximately 487,000 unique galaxy survey targets, at least ~ 490000 with good redshifts. The total region covers $\sim 3500 \text{ deg}^2$. The coverage of the survey is shown in Fig.(V.1). The geometry of the survey is determined by the intersection of the photometric and the spectroscopic surveys. The photometric survey consists of a series of overlapping stripes across the sky defined by the scanning of the two cameras. The spectroscopic survey consists of a set of plates covering the regions that had passed by the tiling algorithm. The target algorithm assigns the targets optimally to the plates. I begin defining some terms that would be useful when working with BOSS data.

- *Chunk*. It is a region in the sky targeted and tiled at the same time, i.e. all targets coming from the same target selection code. Usually it covers about 250 deg^2 (Chunks 1-7) and typically 50 plates. The current BOSS data consist of 19 Chunks that are shown in the figure V.1.
- *Tile*. It is a circular region of radius of 1.49 deg which contains the objects for a given observation, correspondent to the 900 fibers. The tiles are not uniformly distributed on the space to optimize efficiency and completeness (Blanton et al., 2003). The tiles overlap to recover the colliding objects. They are trimmed to the chunk boundaries giving place to the non overlapping polygons (See Fig. V.2).
- *Sector*. A sector is the region of the sky covered by an unique set of tiles (within the chunk edges). A sector could be made up by many polygons. The polygons do not have to be necessarily contiguous but these regions must have the same properties. They are natural place to define completeness (See Fig. V.2). We will see later that it is possible to define completeness in polygons as well, however, given the small size of polygons it could be subject to statistical fluctuations.

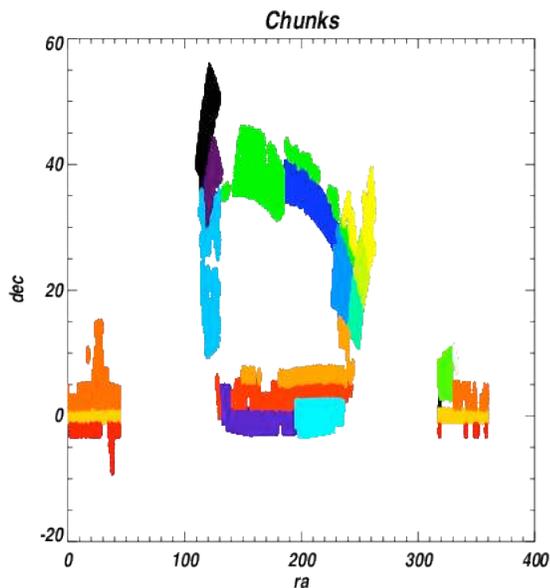


Figure V.1: BOSS spectra separated by chunks. Includes Chunks 1-19

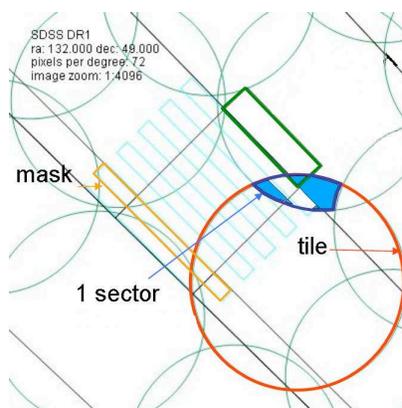


Figure V.2: Diagram showing the concepts of sector, tile, mask. From <http://cas.sdss.org/dr7/>

V.1.1 Samples for Large Scale Structure

There are two complete galaxy samples that we can generate from BOSS data: the so-called *LOWZ* sample and the *CMASS* sample. For the clustering analysis I will concentrate in *CMASS* sample. The *CMASS* sample consists of high mass galaxies and covers the redshift range between $z=0.43$ and $z=0.7$. Let us remind the definition of the *CMASS* sample:

$$d_{\perp} > 0.55 \quad (\text{V.1})$$

$$i < 19.86 + 1.6(d_{\perp} - 0.8) \quad (\text{V.2})$$

$$17.5 < i < 19.9 \quad (\text{V.3})$$

$$i_{fiber} < 21.5 \quad (\text{V.4})$$

$$r - i < 2 \quad (\text{V.5})$$

$$z_{psf} - z > 9.125 - 0.46z \quad (\text{V.6})$$

$$i_{psf} - 1 > 0.2 + 0.3(20.0 - i) \quad (\text{V.7})$$

$$(\text{V.8})$$

The cuts are defined in terms of the rotated combination of colors, c_{\perp} , c_{\parallel} , d_{\perp} :

$$d_{\perp} = (r - i) - (g - r)/8 \quad (\text{V.9})$$

$$(\text{V.10})$$

The first three cuts are defined to separate high redshift galaxies and to account for passively evolution of massive galaxies accordingly the population synthesis Maraston model (Maraston et al 2009 (Maraston, 2005)) In particular, the second equation is a cut in absolute magnitude or stellar mass where d_{\perp} tracks for redshift (Padmanabhan et al in preparation) . The cut in d_{\perp} performs the separation of low and high redshifts and the combination with i cuts guaranties the constance of the masses in the sample. The last two cuts are designed to reduce the stellar contamination. The middle cuts are quality cuts. These magnitude cuts uses *CMODEL* magnitudes and the colors are defined with *MODEL* magnitudes except for i_{fiber2} which is the magnitude in the 2 arcsec fiber. (Stoughton et al 2002 (Stoughton et al., 2002)) . All magnitudes are corrected from galactic extinction. The cmodel magnitudes are defined as follows

$$\begin{aligned} cmodelflux &= I_{dev} * fracpsf + I_{exp} * (1.0 - fracpsf) \\ cmodelmag &= 22.5 - 2.5 * alog10(cmodelflux) - extintion \end{aligned} \quad (\text{V.11})$$

The cmodel magnitudes are based on the better-fitting of a particular model in the r band as a matched aperture to calculate the flux in all bands. The models used are: the exponential flux (I_{exp}) and or de Vaucouleurs model flux (I_{dev})¹. The terms $fracpsf$ is the best fit linear combination of the respective profiles.

There are some objets that satisfy the selection from CMASS and LOWZ , thus a minimal redshift is set to separate both samples, $z_{min} = 0.43$.

¹The expressions of the 2 profiles are the following, the first one corresponds to a pure de Vancouleurs and the second equation an exponential profile.

$$I_{van}(r) = I_0 exp-7.67[(r/r_{eff})^{1/4}] \quad (\text{V.12})$$

$$I_{exp}(r) = I_0 exp(-1.68r/r_{eff}) \quad (\text{V.13})$$

$$(\text{V.14})$$

Comoving Density

In Fig. (V.3) the comoving number density of BOSS data are shown, to select the samples the flag *BOSSTARGET1* is used. This flag indicates for a given version of the photometry file and a given version of the target selection the nature of the targets. In Table 1 BOSS survey primary target selection flags for galaxies are shown. We will see later that due to the changes in the target selection and in the photometric reduction it would be necessary to homogenize the sample, thus the cuts would be reapplied as this flag is not used anymore for the sample generation. However for getting an idea of the number galaxy density it is accurate enough. In fact the comoving density would be used to monitor and warranty the homogeneity of the sample.

Table 1. BOSS survey primary target selection flags for galaxies

	Masktype	BOSS_TARGET1
0	GAL_LOZ	low-z lrgs
1	GAL_CMASS	$d_{\text{perp}} > 0.55$, color-mag cut
2	GAL_CMASS_COMM	$d_{\text{perp}} > 0.55$, commissioning color-mag cut
3	GAL_CMASS_SPARSE	GAL_CMASS_COMM & (!GAL_CMASS) & ($i < 19.9$) sparsely sampled
7	GAL_CMASS_ALL	GAL_CMASS and the entire sparsely sampled region
6	SDSS_KNOWN	Matches a known SDSS spectra.

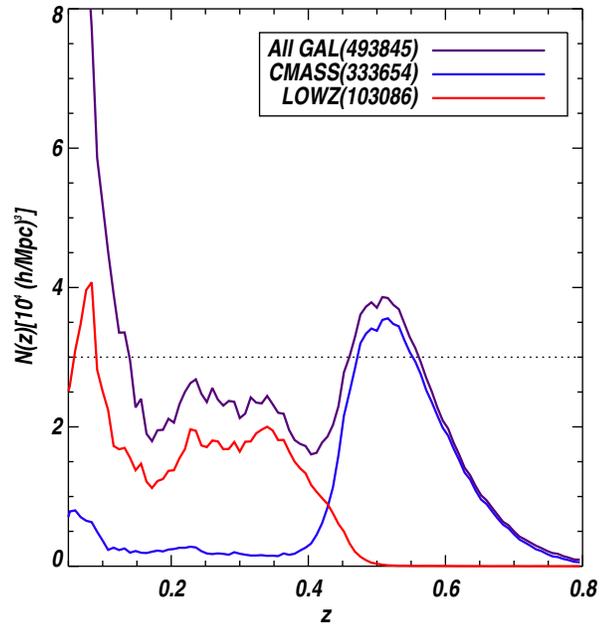


Figure V.3: Comoving number density of BOSS Galaxy samples

V.1.2 Angular Selection Functions

The angular selection function of BOSS could be separated into two functions in term of MANGLE polygons, the veto mask accounting for the regions we do not have access and the window that defines the geometry and accounts for the completeness of the survey. The completeness of the survey is used as a weight for each region (sector, polygon) that should be applied to random sample in the correlation function analysis.

Veto Mask

The veto was generated within the working group of BOSS galaxy clustering and they are documented in Blanton and Tinker in prep. They consist of the concatenation of different components:

- Bad fields of photometric survey. The holes generated by bad fields where no photometry is available to apply the target algorithm.
- Bright star mask. This mask hides the neighboring around bright stars. It is generated with the Tycho catalog of stars.
- Center post mask. This mask hides the centers of the plates where is not possible to place a fiber. It consists of a circular region of 92 arcsec at the center of each tile. Of the N? tiles used up to now, the total area of the mask is ?, so even if it is small it should be considered for precise clustering analysis .
- Collision priority mask. This mask accounts for the objects that received a fiber and had a higher priority than the LRG's. It puts a circle with radius 62 arcsec.

Completeness

As we have seen in the last chapter the completeness is fundamental to estimate the angular selection function of data, permitting to compensate the incompleteness of the sample via the random set. The basic definition of the completeness is the rate of successful spectra taken to the targets for each region.

$$Completeness = \frac{spectras}{targets} \tag{V.15}$$

The completeness defined in this way will account for the missing redshifts independently of the reason why they were not measured (sky area not yet survey, redshift failure, observational issues, etc). Some analysis (Ross et al in prep) separate the completeness in two kinds of completeness angular and redshift. This practice is useful when looking to understand the systematics, but in order to do clustering analysis it is more pragmatic to include all the possible effects in only one quantity that will be apply to the random generation.

In principle for computing the completeness for each sector (polygon) we only need thus a target list and spectra list. Nevertheless in BOSS survey some considerations must be taken into account for the correct computing of the completeness and the catalog generation. Before beginning with the detailed description I enumerate the global procedure to permit the reader to follow the further subsections.

V.1.3 Catalog definition

1. We define a target list which by construction generates a complete sample. The targets lying in the veto mask as well as the targets corresponding to legacy data (data from previous versions of SDSS), are subtracted. This data are flagged UNKNOWNCOLL. Regarding the targets I discuss in detail in further subsections the changes in this target list definition and the photometric offset between hemispheres.
2. We identify the legacy data, which passes the CMASS cuts. In a further subsection we explore some ways to deal with this sample. These objects are flagged as KNOWN.
3. We select a spectra list applying some cuts. We select the non repeated, with good redshift, CMASS spectra.²
4. The failure targets are identified and corrected as explained in a further subsection.
5. The spectra taken are matched to this target list and to the legacy list. Some legacy data have BOSS spectra. These data are included in the catalog, we flag these objects as DUPLICATE
6. The fiber collision objects are identified and corrected as explained in a further subsection.
7. The completeness is computed and to homogenize sample a cut in completeness is applied ($C > 0.7$).
8. We restrict the sample to the redshift range $z=[0.43, 0.7]$, the lower limit is set to separate from LOWZ catalog because there is an overlap in this region.

V.1.4 Targets Definition

During the first chunks up to Chunk 15 the photometry reduction as well as the target selection changed. In Table 2 these changes are enumerated. The changes in the target selection generates inhomogeneities in the sample. The changes in photometric reduction implies some ambiguity in the definition of the targets and the completeness calculation. The first step to generate the target list is to homogenize the sample freezing the target selection to the final version. The consequences of reapplying the cuts depends on the *sense* of the changes in the target selection:

- *Early target selection more RESTRICTIVE.* In this case we loose targets The missing objects lie in the borders of the color cuts, what makes difficult the possibility to compensate up-weighting. However as the latest data is more permissive, an estimation of $n(z)$ with the earlier and the final target selection could be achieved to evaluate the impact of the changes. This is the case of LOWZ sample that is not treated in this work.
- *Early target selection more PERMISSIVE.* In this case the only effect would be the loose of objects. This is the case of CMASS sample where the target selection changes imply a loss of objects of 3.5% for MAIN008 (7.8% for GATH, for the entire geometry)

²That is using BOSS flags ((BOSSTARGET1&2 =0)& (PRIMARY_OBJECT=1) & (ZWARNING_NOQSO = 0)).

In relation to the photometric reduction and the ambiguity in the completeness it is possible to proceed in two ways:

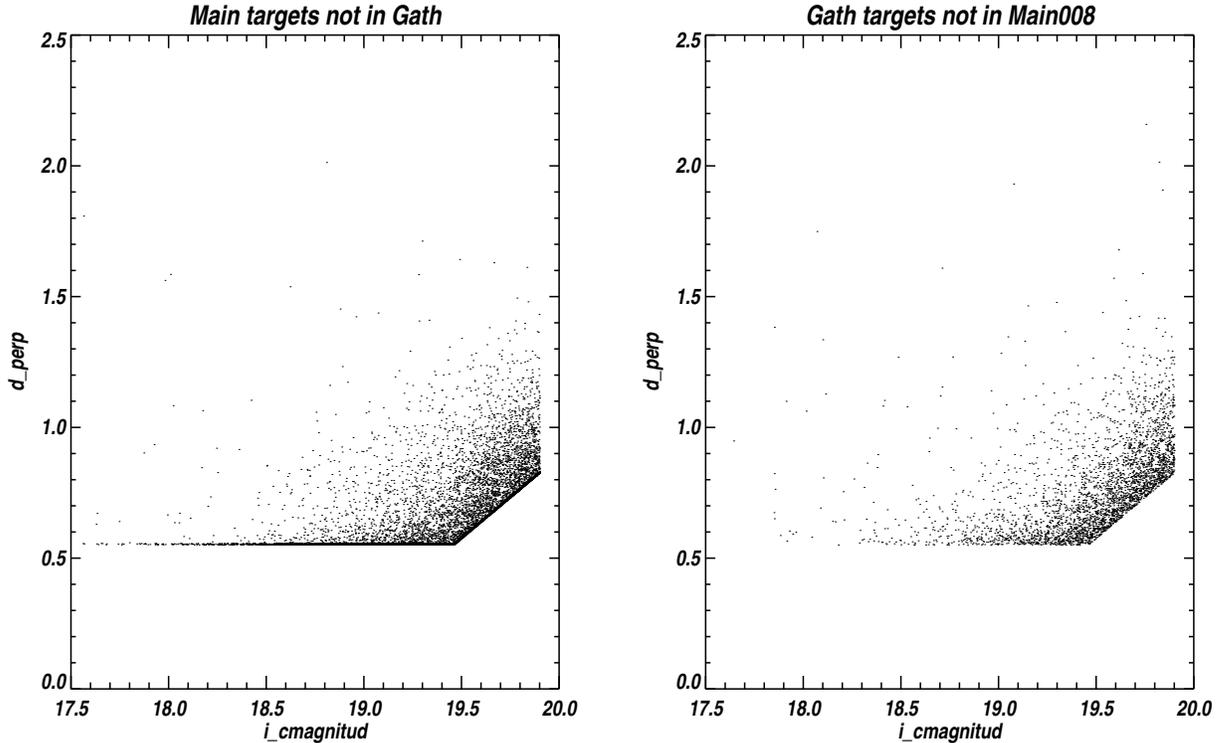


Figure V.4: Targets from the target run MAIN008 and the combined target list *GATH* for each .

- The first possibility is to take the latest photometric reduction and redefine the sample. As a consequence of the changes in photometry, some galaxies pass the selection cuts in the final photometry but they were not targeted at the time of observing and inversely. The result is that many objects scatter across the boundary as we see in figure V.4 where we explored the implications of using a singular photometric reduction and a combination of them.
- The second choice is to take the selected targets at the time of targeting. This evolving photometric reduction implies the completeness definition is fixed but the changes in the photometric reduction could generate an evolving comoving density. If the difference between reductions do not generate changes in the comoving density this option is preferable as the completeness definition is safe.

In order to choose the better option to define the target list I will explore two choices to generate the catalog. I test using photometry reduction MAIN008 . The catalog generated with this photometric reduction will be call *MAIN008* . I generate another catalog using this time the photometric reduction correspondent to each chunk. For the generation of the combined photometry file, I use

Table 2. Changes in photometric reduction and target selection

chunk	targetrun	target version	Photo_Calib	Photo_Resolve	Photo_sweep
boss1	comm	default0	v1_0_4	2009-06-14	2009-06-14
boss2	comm2	default0	v1_1_2	2009-06-14	2009-09-28
boss3-4	main002	default0	v2_0_1	2009-06-14	2009-11-16
boss5-6	main005	fall09i	v2_0_3	2009-11-16	2009-11-16.v2
boss7-11	main007	fall09i	v2_0_5	2009-11-16	2009-11-16.v2
boss12-14	main008	dr8final	v2_0_9	2010-05-23	dr8final
boss15-*	main009	dr8final	v2_0_13	2010-05-23	dr8final

Summary of Changes in photometry and target selection

I Fiber magnitude limit i_{fib2} has changed from 21.7 to 21.5 since chunk 15.
 100 % of LOWZ and 90.5% of CMASS was targeted using the DR8 Photometry

the first routine of the mini-pipeline *MKSAMPLE* created for this purpose within collaboration by Berkeley Group. This routine begins with the latest photometric reduction *MAIN009* and for each chunk within 1-15, the targets from the respective chunk are replaced for their respective reduction following the table 2. The output file is then used as the target list. The catalog generated with this file will be called *GATH*. In both cases, I re-apply the final target selection cuts defined in V.1.1 to the target list and the veto mask.

In figures V.5 and V.6 the targets corresponding to each chunk are shown using the different target list, in black the targets common to both target list *MAIN008* and *GATH*, in red the targets only present in *MAIN008* and in blue the targets present only in *GATH*. As we can observe from the figure, even if the differences between target list are small, the spatial distribution is not random, there are particular regions as the borders between chunks and the borders of plates. Now I compare the distribution of missing targets in the color space. In figure V.4, in left panel, the targets from *MAIN008* not matched with any target in *GATH* as well as the targets from *GATH* not found in *MAIN008* are shown in right panel. As we observe from the figure the scattered objects from *MAIN008* are not random they are preferentially from the boundaries in color-mag space, i.e., at preferred redshifts.

Then I followed the prescriptions described in section 1.1.3 to generate the two catalogs. The Legacy is excluded, the fiber collision corrected as method 1 and the target failures corrected as method 2 explained in further sections. The cut in completeness is $C=0.7$ and the cut in redshift is $[0.43,0.7]$. Some statistics of the catalogs are shown in TABLE 3.

Table 3. Statistics from GATH and MAIN008 catalogs

Starting with	2,288,550	1,591,029
Targets After BossTarget1 cut	1410081	990958
Targets After Color cuts	1299425	956689
CMASS galaxies matched with targets	289004	297010
Number of NonCMASS in SpAll	15734	15452
non spectra BOSS, known objects:	14588	14225

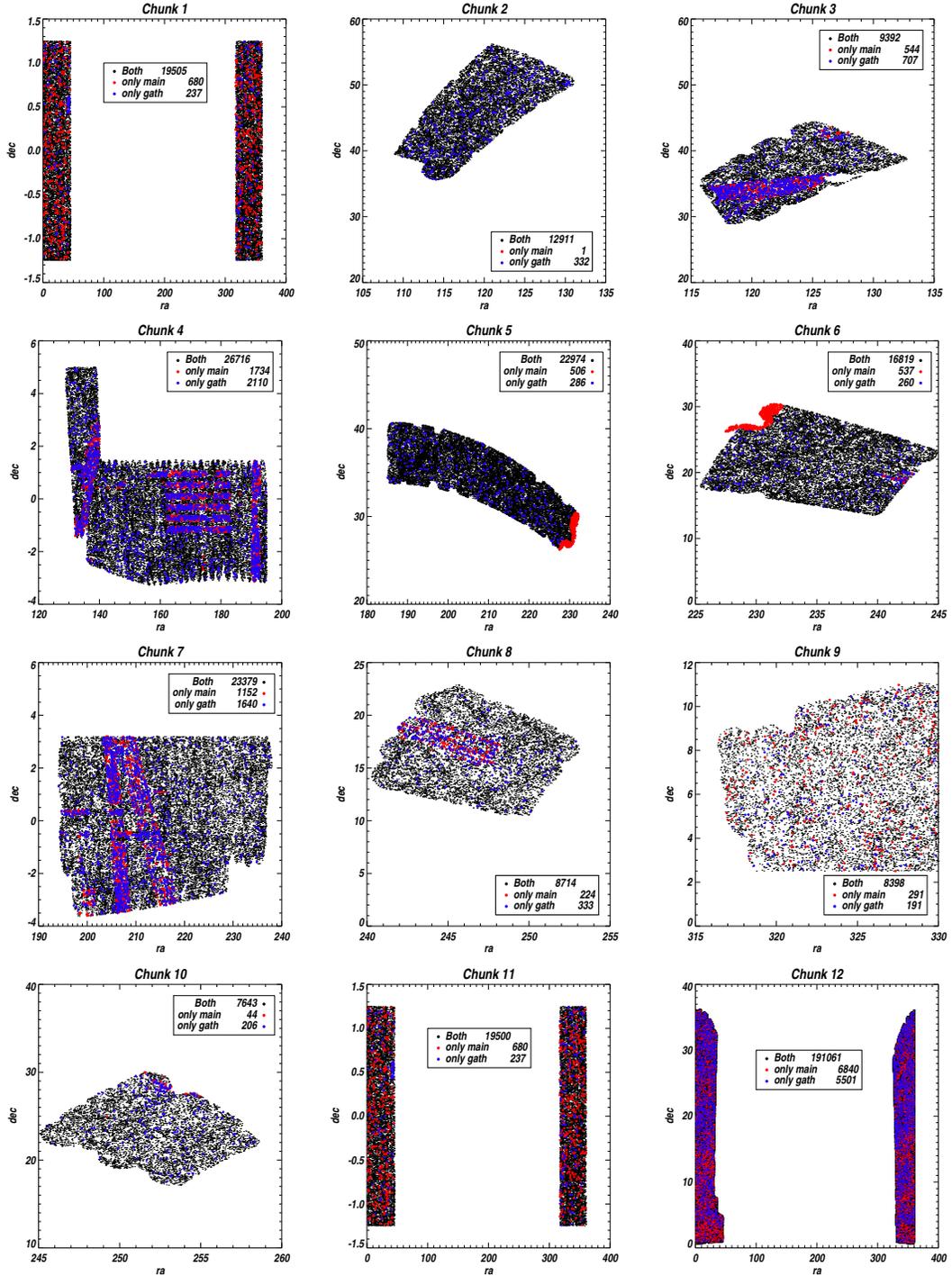


Figure V.5: Targets from the target run *MAIN008* and the combined target list *GATH* for each chunk from 1-12.

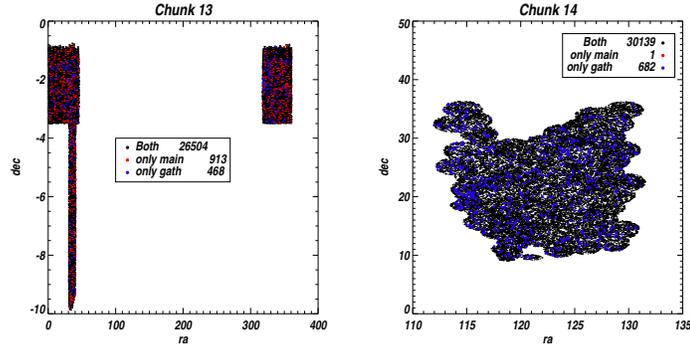


Figure V.6: Targets from the target run *MAIN008* and the combined target list *GATH* for each chunk from 13-14.

The first test I apply is to study how the comoving density evolves for each chunk in both catalogs. In figure V.7 the comoving density is shown for each chunk in panel A for *MAIN008* and in panel B for catalog *GATH*. As we can observe from the figure in panel B, the comoving density does not seem to be affected by the changes in photometric reduction within the statistical fluctuations. Thus, the *GATH* catalog seems to provide the best characteristics to generate the catalog.

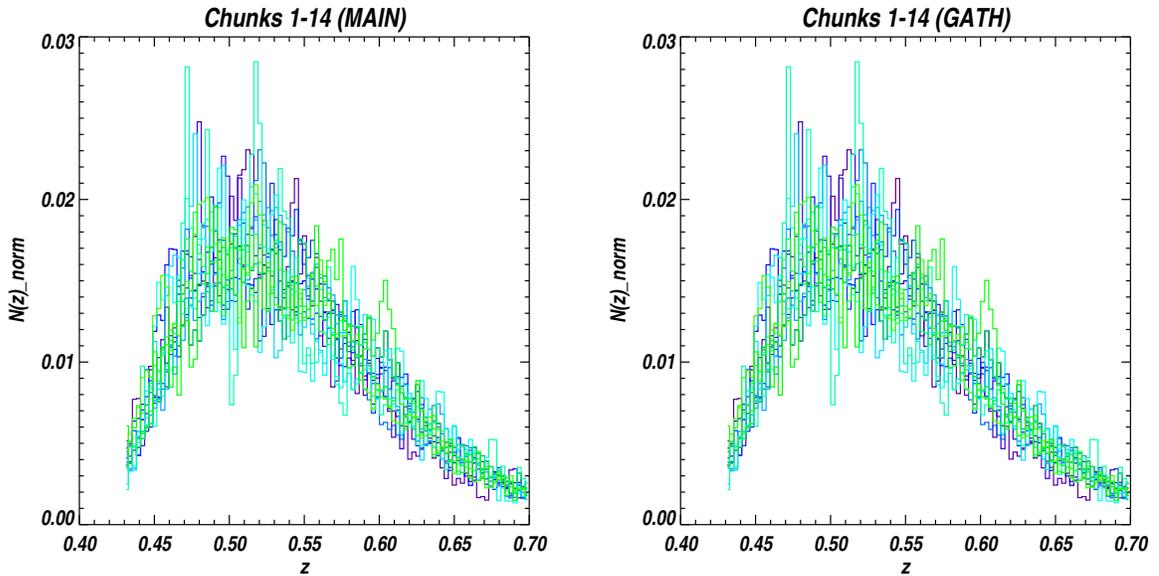


Figure V.7: A) Comoving density from the target run *MAIN008* for each chunk. B) Comoving density from the target run *GATH* for each chunk 1-14.

Finally, I compute the correlation function using both catalogs. For each case the random size is 5 times the size of data set. The random follows the redshift distribution of data and the completeness. In figure V.8 the correlation function using each one of the catalogs. The correlation

functions are completely compatible, the small fluctuations are probably generated for the size of the random.

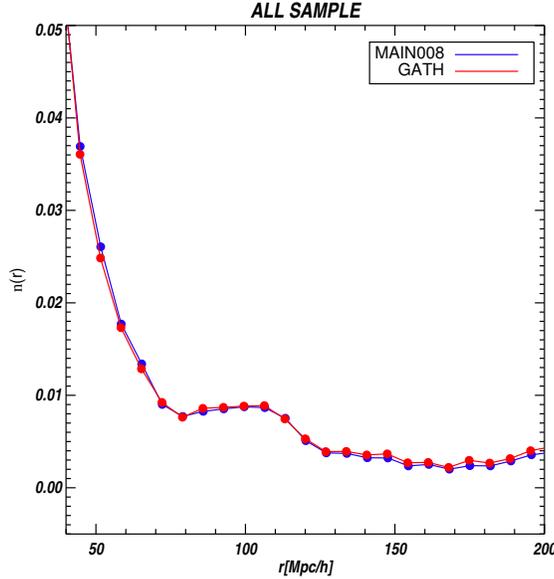


Figure V.8: Correlation function using the combined photometric files *GATH* compared with sample generated using only one reduction *MAIN008*

V.1.5 North/South Offset

One of the big concern at early analysis of DR9 data within the clustering group was that it has been found a clear difference in the mean density of south (SGC) and north (NGC) hemisphere in Galactic coordinates. In figure V.9 the mean density of CMASS galaxy sample for all sample and separating North South is shown. The galaxy number density for $0.43 < z < 0.7$ CMASS sample of the south hemisphere shows a $\sim 3.4\%$ higher number density than the NGC.³

Schlafly & Finkbeiner (2010) have found an offset in the colors as a function of the position. They explained the origin of the offsets by a mixing between calibration errors and errors in the correction of the galactic extinction. This offset affects the target selection because the CMASS selection is sensitive to d_{perp} color (and LOWZ sample is sensible to $c_{||}$ color). Their results shows an offset of 0.0064 between North and South in d_{perp} . Ross et al found that this offset in d_{perp} between the Northern and southern hemisphere is consistent with the 2% difference in the number density of the CMASS targets (Ross et al., 2012).

Following the results of Schlafly et al and Ross et al, this offset in d_{perp} was applied to the south hemisphere targets before applying the color-cuts of the target selection. The figure V.10 shows the effect in the target selection of applying this offset. In blue the north hemisphere CMASS targets in red the South hemisphere with offset. As we observe this offset implies more restrictive cuts for

³For the LOWZ sample in the redshift range $0.2 < z < 0.4$ the difference in number density is larger $\sim 12\%$

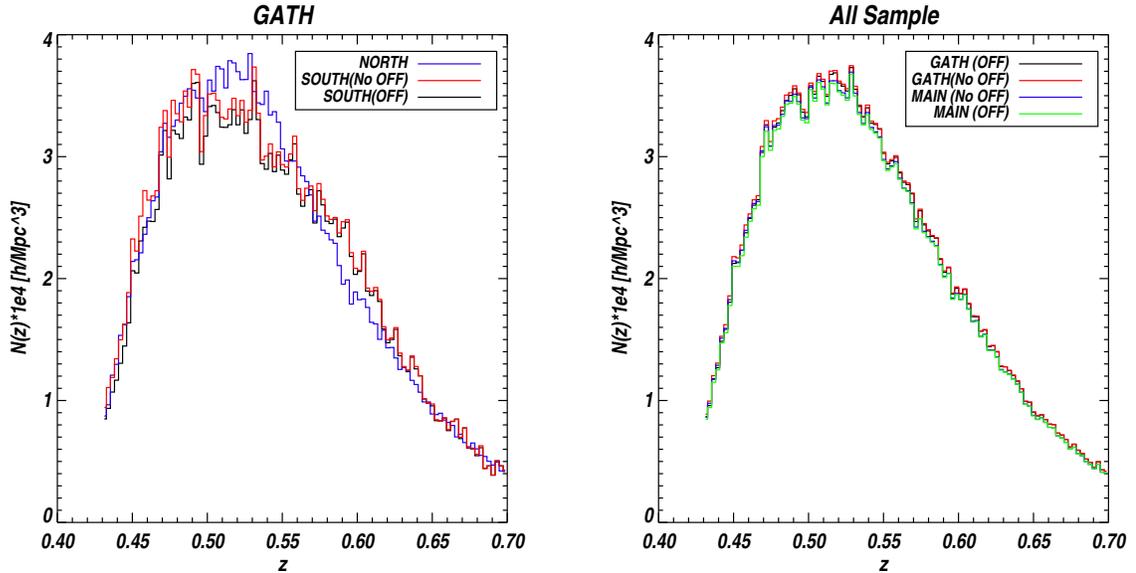


Figure V.9: Left panel. Mean number density of *GATH* catalog separating North and South, with and without offset applied. Right panel. Mean number density of *MAIN008* and *GATH* catalogs, with and without corrections for whole sample.

south targets. This implies a loss of objects of % that corresponds to difference in the mean density as we can observe from the panel B of the figure V.9.

I study the effect in the correlation function of applying this offset in both catalogs *GATH* and *MAIN*. As we observe from Fig V.11 the effect in the correlation function is negligible, it only adds some small correlation in large scale.

Now in panel A of figure V.12 I show the correlation function separating the North and South hemisphere with the offset corrected and not for the south targets. From the figure we observe the correlation function of North and South seems to follow different behaviors. From the same figure, panel B we observe that the correction of the offset do not change significantly the south correlation function. We will see later that after all systematic corrections the discrepancy between North and South is within the statistical fluctuations.

V.1.6 Target Failure Correction.

The last correction that we must apply to the targets is related to the *target failures*. There are some *CMASS* targets that results in spectra of no *CMASS* objects, even if this a small fraction when accounting for the completeness of the survey it is important to deal correctly with these objects we called *target failures*. As all the spectra with no *CMASS* nature are eliminated, a correction must be done in the targets list to account for this failure rate. If we do not correct from this no *CMASS* objects in the denominator, it generates a completeness dependent on galactic latitude related with the principal polluter the stars. It provokes an inverse weight in the completeness related with the

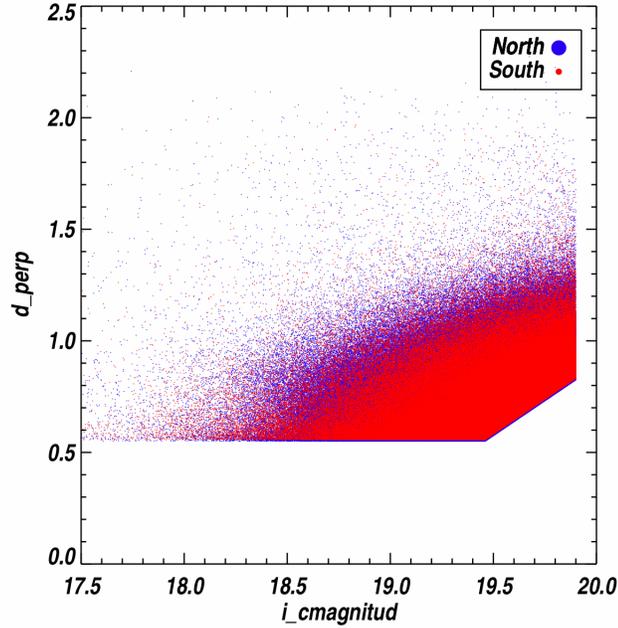


Figure V.10: The targets from GATH catalog in the color space. In blue the north hemisphere CMASS targets in red the South hemisphere with offset. As we observe this offset implies more restrictive cuts for south targets.

stars contamination in the targets. Heretofore I implemented method 2, in this section I study different methods to correct for this failure rate. I tested three methods:

- Method 1. We eliminate the objects that were found not to be CMASS spectra (i.e. stars, qso) from the target list from each polygon.
- Method 2. We estimated the proportion of (targets failures)/spectra and we applied this weight to the targets in each polygon (reducing the number of remaining targets by this fraction). This option is justified by the fact that the survey is not complete, thus it is natural to estimate that the proportion of failures would be constant for the rest of the targets not yet observed. The "target success rate" τ is defined as :

$$\tau = targets * \frac{spectras_{NoCMASS}}{spectras_{CMASS}} \quad (V.16)$$

- Method 3. Given that there are few objects per polygon, doing (2) is still subject to large statistical fluctuations so , as this rate of failures would be function of the galactic latitude, we applied a global correction function of the galactic latitude . To implemented this method we separated north and south see figure V.13. As we notice the correction seems to be different for both hemispheres thus we treat each hemisphere separately. The dependency on the galactic latitude is not exactly linear, thus, we fit with a spline (see Fig V.13.).

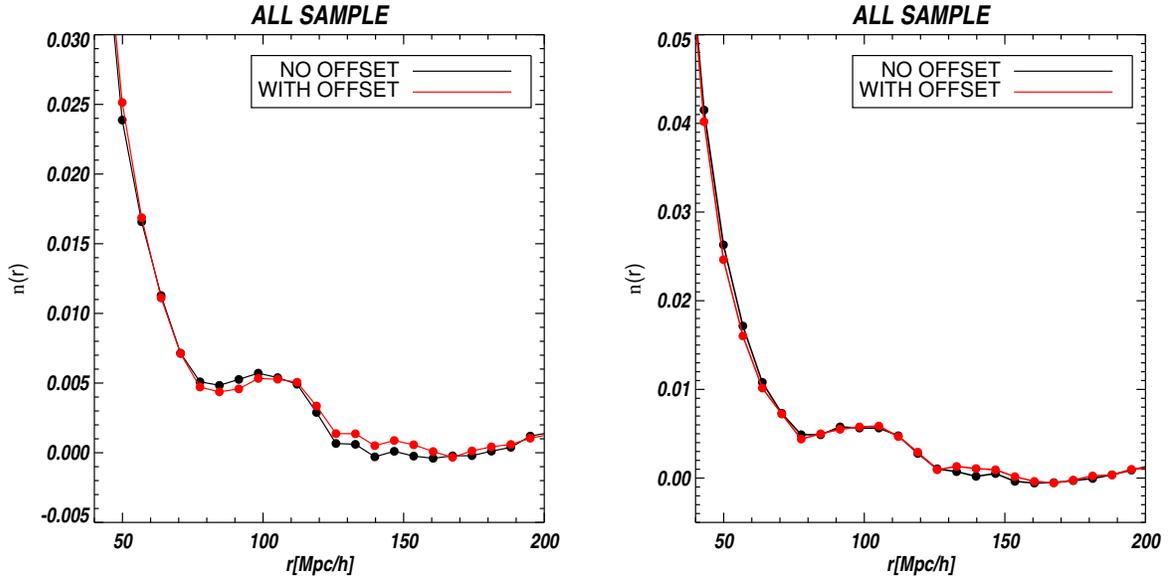


Figure V.11: The correlation function of applying this offset in both catalogs GATH (right) and MAIN (left).

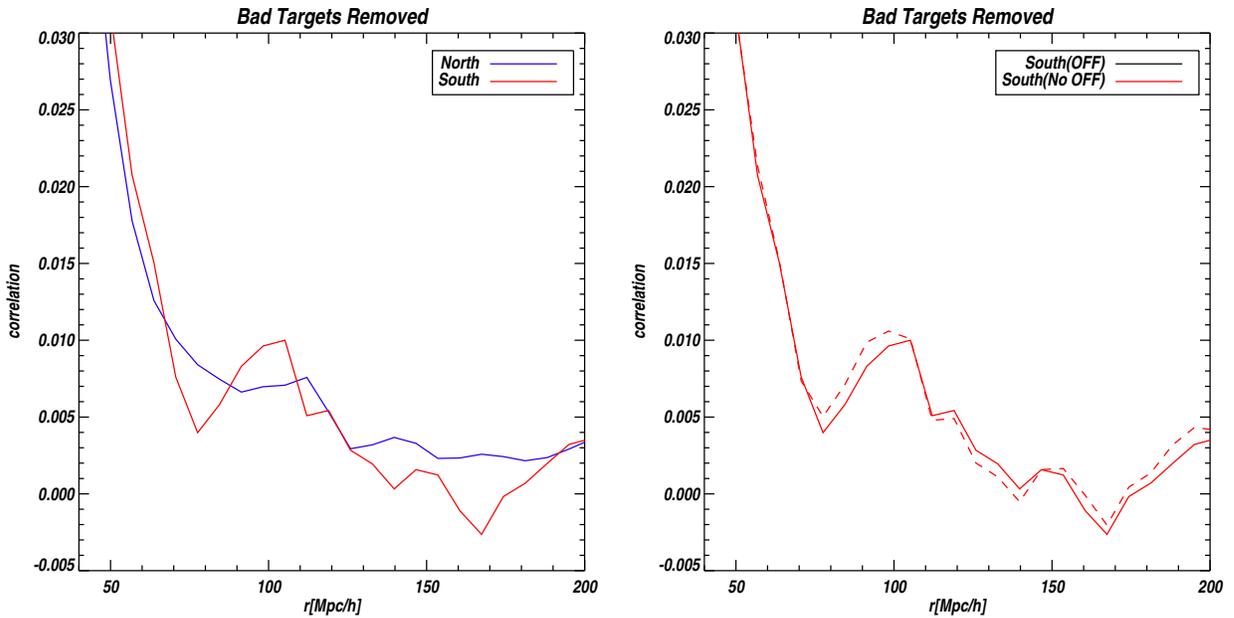


Figure V.12: The correlation function of applying this offset in catalog GATH separating North and South in left in right the correlation function of south after and before offset correction.

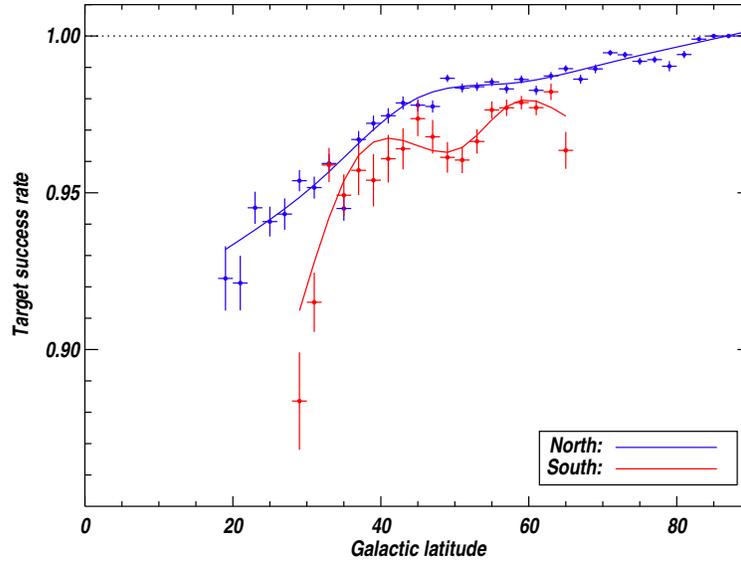


Figure V.13: The “target success rate” as a function of the galactic latitude.

We analyze the results in terms of the completeness and the correlation function and we compare with the case of not applying any correction. The results are shown in figure V.14 and V.15. The three methods give rather similar results but not exactly:

- Method 1: In the completeness there seems to remain a very small evolution with galactic latitude. The resulting correlation function is higher than erroneous initial calculation without correction.
- Method 2: Completeness does not depend on galactic latitude. The correlation function is higher than the case with no corrections, and even higher to method 1.
- Method 3: Completeness does not depend on galactic latitude. The correlation function is very close to that of method 2.

We compare in a further section the different methods for correcting bad targets with the angular weights included and we compare with the results from the official catalog of BOSS clustering group. For the moment, we fixed the method 2 to correct for the target failure rate as we considered it performs a better correction as it considers the rate of failure in a statistical way. As we will correct in a subsequent section from angular systematics, the method 3 could affect the estimation of the angular corrections.

V.1.7 Fiber Collision Correction

The finite size of the plugins for the fibers imposes a limit in the selection of targets of 61 arcsec around which is not possible to observe at the same plate other targets, thus is not possible to get

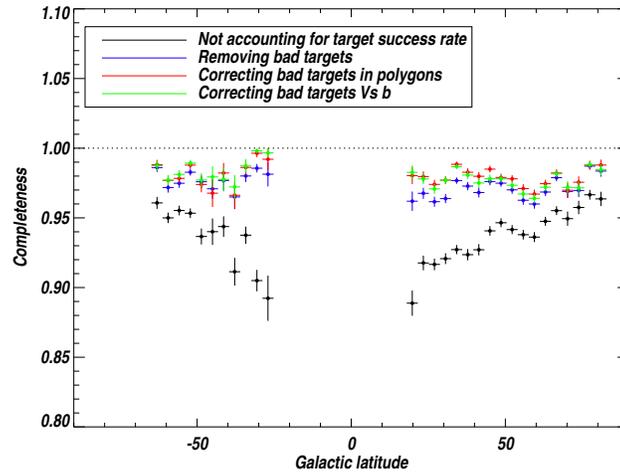


Figure V.14: The completeness as a function of the galactic latitude for the 3 methods of correcting the bad targets.

spectra of two galaxies which angular separation is less than 62 arcsec. The targets that are within this radius are called colliding targets. The tiling algorithm is optimized to assign targets to fibers, trying to optimize the non colliding targets. Even though, to retrieve a % of this colliding targets the tiles are designated to overlap. The % of the total area are covered by more than one plates. Even doing this a % of targets are lost. The fiber collision affects the small scales essentially however it has an effect in the large scales as we will see. There are two ways to correct from fiber collision.

- Option 1. The first approach to account for the fiber collision used in SDSS-I-II is to include a fiber correction weight in the completeness. The idea is to identify the targets having close pairs and not having a fiber assigned, the so called *colliding targets* and add a weight to the target with spectra associated to compensate the fiber collision. This method is equivalent to include the close pairs in the catalog and assigning them the redshift of the closer pair from legacy data or boss data. We applied the correction in the second form. Taking into account this correction the final definition of the completeness is:

$$completeness = \frac{spectra + closepairs}{targets + closepairs} \quad (V.17)$$

As the assigned redshift is essentially random the global effect is a kind of noise originated from the random of the redshift assigning.

- Option 2 Hong in preparation have proposed an alternative to this approach that gives a more precise correction for small scales. The method consist in separating the sample in two populations. A non colliding population with N_1 elements that consist in all galaxies that are separated by a distance larger than the fiber collision scale. The second sample includes all the N_2 collide objects. Thus, there are N galaxies targeted and only N' would have assigned a fiber, N' is composed of the N_1 , non collide targets plus a fraction of the collide targets

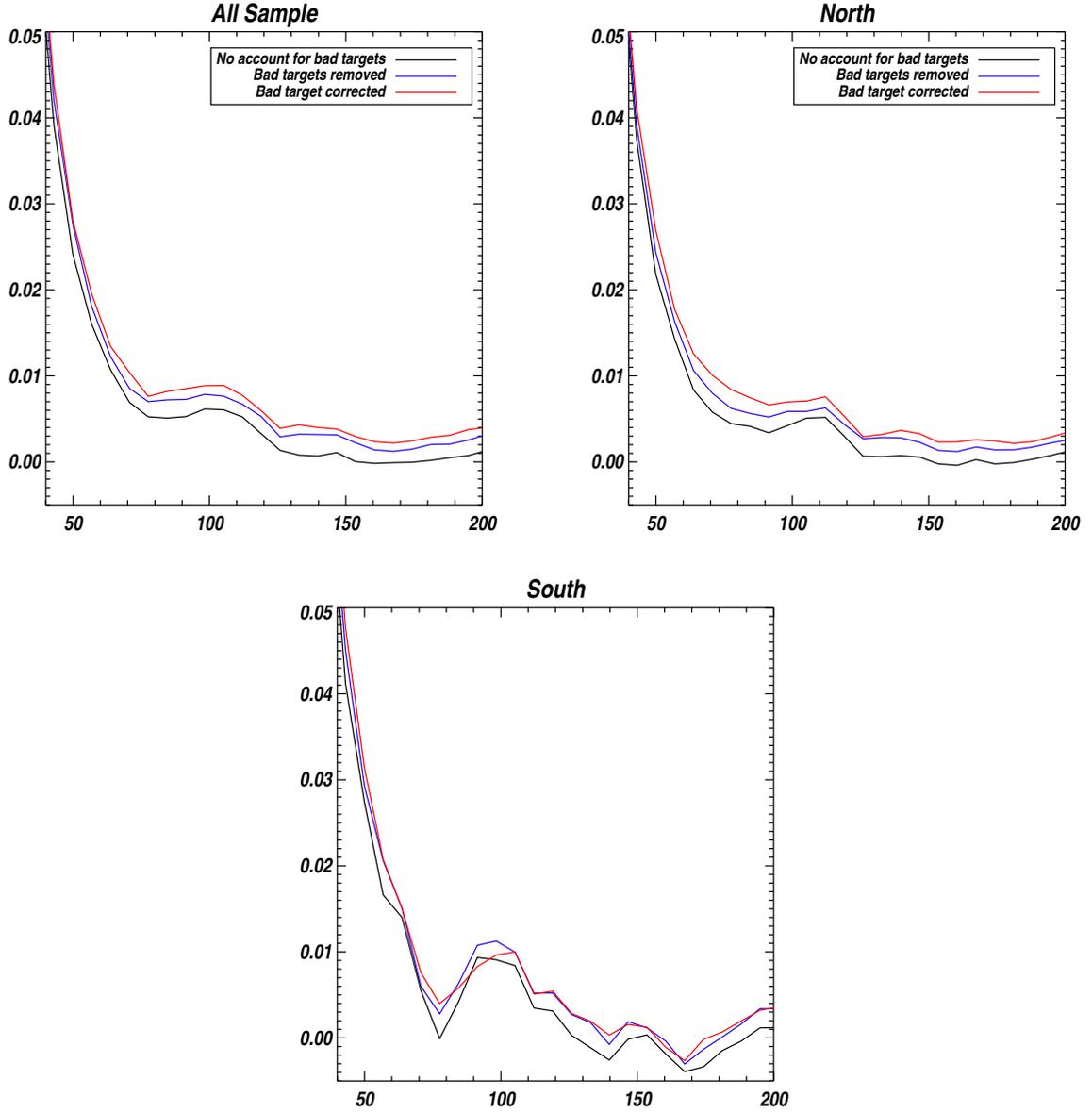


Figure V.15: Correlation function computed with GATH catalog using different methods for correcting the bad targets. Top left for the hole sample, top right the north hemisphere and bottom south.

$N'_2 = f(N_2)$, that is :

$$N = N_1 + N'_2 \quad (\text{V.18})$$

Following Hong et al the data pairs would be given by:

$$\begin{aligned} DD &= D_1 D_1 + D_1 D'_2 (N_2/N'_2) + D'_2 D'_2 (N_2/N'_2)^2 \\ DR &= D_1 R + D'_2 R * (N_2/N'_2) \end{aligned} \quad (\text{V.19})$$

This expression assumes that the data counts of the collide terms with spectra are simply proportional to all the colliding objects.

Up to now we have applied the method 1 for correcting for the fiber collision.

V.1.8 Legacy Data.

Up to now we have worked with BOSS data exclusively. However, we have to take into account that for the targets where a spectra is available in Legacy data, they do not have an assigned fiber in the BOSS survey. Thus BOSS data should be combined with the Legacy Data. In fact the Legacy data and BOSS data are in principle two separated samples, each one with a completeness, angular and radial selection function. Thus being strict we must treat as two separated samples and duplicate all analysis. But because Legacy data represents a small percentage of the targets (1.5%) we preferred to combine them in a unique sample. Merging both samples is not a trivial issue, the problem arise when dealing with the random set. In fact the random set must mimic the data set, but because the data is a combination of two samples, the random generation must take this into account. In this section we study the ways to deal with this issue and the effects in the correlation function. I study four different options to deal with legacy with different levels of accuracy, the idea is to study the effect in the correlation function and select the best choice for the catalog generation.

- Option 1. As the existence or not of known objects do not affect targeting of the object, the idea is that we can treat the completeness for both samples separately. The legacy data is taken as a known sample and a completeness equal to 1 is assigned. The BOSS completeness is computed separately using only the targets with a fiber assigned. It implies that some sectors will have to values from the completeness. What is proposed, is to generate random points following boss completeness and weight data with their respective completeness.

$$Completeness = \frac{fibersassigned}{targets} \quad (V.20)$$

- Option 2. The completeness is computed in the same way that option 1. The Legacy objects are included subsampled to mimic the completeness of BOSS . The close pair are subsampled to follow BOSS fractions (Boss-known, known-known). It implies a lose of data of ?%. This is the method implemented in the BOSS clustering Group.
- Option 3. Include the known in the completeness calculation. This option is not exact as LEGACY data is different in luminosity, redshift and colors. Thus, completeness of SDSS could not be an indicator of BOSS completeness. But the idea is to compare the error associated with the incorrect accounting of legacy data

$$completeness = \frac{spectra + known}{targets + known} \quad (V.21)$$

- Option 4. Not including legacy data. Loss of data (1.5% of target list) negligible at the current state.

Before presenting the results I begin describing the known sample. It consists of about 14588 objects (1.5% of target list). In the data available up to now, 914 have been re-observed and from them 517 are CMASS objects. The legacy data is not randomly distributed within BOSS coverage.

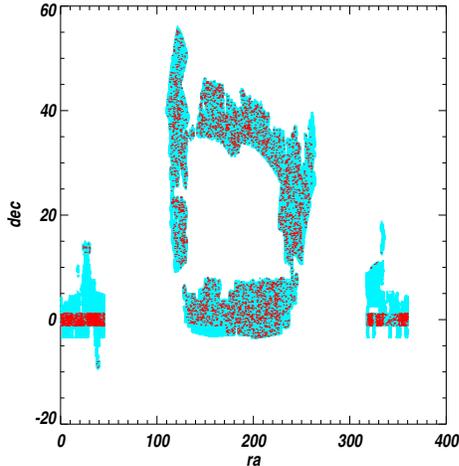


Figure V.16: Coverage of Legacy Data, in cyan all data in red legacy data

In figure V.17 the results are shown. As we can observe from the figure including the legacy has not a big effect in the correlation function, including the legacy data in the completeness seems to decrease the correlation globally thus it does not seem to be a solution, the method 1 seems to give similar results than not include the legacy data, The subsampling method do not seems to give similar results. The correlation function seems to be displaced by constant noise, typically generated when the random do not follows the same selection function than the data. For the following I decide to exclude the legacy data, this sample at this stage do not represent a large fraction.

V.1.9 Summary:Final Catalog

In this section I have discussed the choices taken to construct the catalog and I explored the implications and justify the choices. To conclude the section I present a summary of the choices taken to define the final catalog and I present the choices taken in the official catalog. As we will see at the end of the following section, once corrected from the systematics, the results obtained with our catalog are pretty similar to the results obtained with the official catalog.

1. The combine photometric target files has been used as the target list. The official catalog uses this version as well
2. An offset to the rotated color dperp have been applied to the south hemisphere. The official catalog applies the offset too.

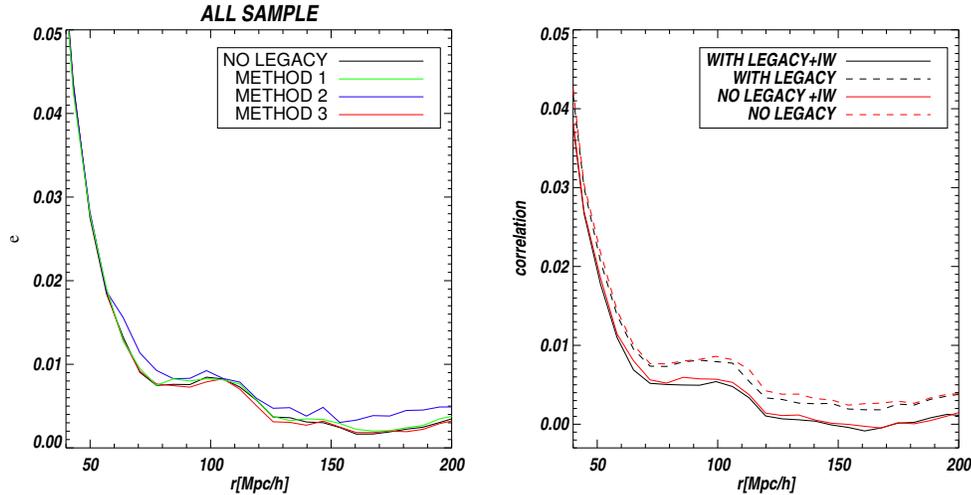


Figure V.17: Correlation function weighted with Legacy

3. The correction for bad targets is given, statistically, multiplying the targets by the target failure, given by the rate defined as V.16. The official catalog removes the bad targets and treats the failures in redshift independently in a similar way to the fiber collision, assigning a weight w_f .
4. The fiber collision correction is done with the nearest neighbor approach. The official catalog assign a weight w_{fc} to the nearest neighbor, the total weight is given by $w_z = w_{fc} + w_f - 1$
5. The Legacy data is not included. The official catalog includes the legacy data subsample to the BOSS completeness.
6. The completeness cut is $C=0.7$ and the redshift range= $[0.43,0.7]$

V.2 Observational Angular Systematic Errors

In this section I test possible sources of angular systematic errors. I describe the way of correct these systematics defining a set of weights and finally I study the effect in the correlation function. The general idea for looking of sources of systematics is to study the fluctuations of the mean density as a function of the variables that are believed to have an effect in the sample. In Ross et al (2010) the fluctuations of the mean density as function of five variables have been studied: airmass, sky, seeing, dust and stellar density. Ross found that stellar density is the principal source of systematic errors in BOSS galaxies. I begin this section studying closely this variable before taken into account the other variables responsible of angular systematics. Then, I study the 5 variables considered by Ross and I add to the analysis one more variable, the offset in photometry.

V.2.1 Stars

As we have said before an anti-correlation between the mean density of galaxies as a function of stellar density have been found. The explanation proposed for explaining this anti-correlation is that where stellar density is high, the surface masked by the stars begins to have an effect in the correlation function because of the missing CMASS targets that could not be observed. Thus, inducing an anti-correlation between density of stars and density of galaxies. The effect is visible in BOSS and not in previous stages of SDSS because we have enough area that effect of stars become important. In fact in order to apply a global correction in Ross in prep the dependency of mean density and the stellar density has been calculated with the whole sample. Even if we agreed that the best way to compute the weights is in a global treatment, in order to explore and understand the origin of the systematic error I have done the same kind of exercise separating North and South hemispheres. For the stars catalog we use all stars in the magnitude range [17.5,20].

The left panel of figure V.18 shows the relation between mean CMASS target density and stellar density and the right panel shows the mean CMASS spectra versus stellar density. The first remark is that we observe a different dependency between targets and CMASS spectra densities. The difference in the slope can be explained because the percentage of targets that are stars has a positive relationship with stellar density, so the target density vs. stellar density has a more shallow slope than the spectra density vs. stellar density.

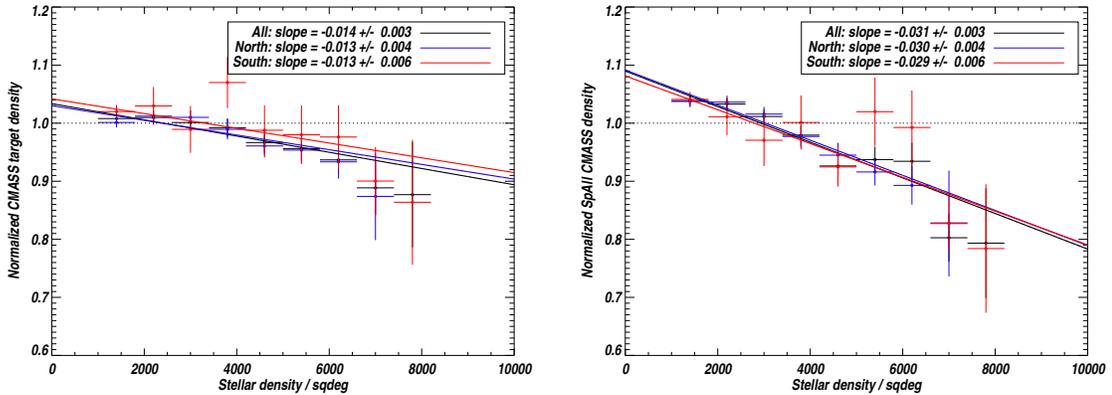


Figure V.18: The relation between mean CMASS target density and stellar density and the right panel shows the mean CMASS spectra versus stellar density.

The second remark is that we observe a different dependency between North and South in each case, however this difference is compatible with the statistical uncertainty. As the dependencies are different we cut in four regions [-90, -40] [-40, 0] [0, 40] and [40, 90], the results are shown in figure V.19. The results indicate that in the high latitude regions (North/South) the CMASS density is compatible with no dependency with stellar density. In the regions of low latitude, we observe the dependency of North region is two times greater than South region.

The last result seems to indicate some atmospheric effects as the seeing is systematically worse in south, thus enhancing the effect of stellar density. We explore this hypothesis, for that we use a pixelised Healpix map of the seeing with Npix=256. For each star we compute a quantity

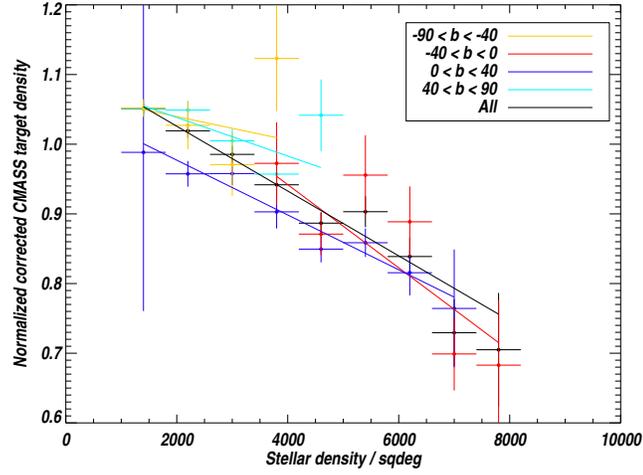


Figure V.19: The relation between mean CMASS target density and stellar density for the bins in galactic latitude: [-90, -40] [-40, 0] [0, 40]

proportional to the surface masked :

$$S_* = \sum_i \frac{seeing(p)^2}{\langle seeing \rangle^2} \quad (\text{V.22})$$

In the left panel of Figure V.20 I present the relation between the normalized CMASS targets and the surface masked normalized. As we observe for the figure clearly the slope of the fits evolves with the galactic latitude bins indicating the existence of a non considered systematic effect in the South no related to the surface masked by the stars. In the right panel of figure V.20 I show the evolution of the slope of the linear fit as a function of the galactic latitude.

Stellar Correction

Independently of understanding the origin of the systematic error the standard procedure to correct from a systematic is to estimate the relation of the mean CMASS density as a function of the possible source of systematic error. In principle there is no reason to find a correlation between CMASS density and each one of the systematics, and what we must observe would be a flat distribution. A deviation of a horizontal line indicates a correlation or anti-correlation with a systematic. To correct a systematic the standard procedure is to apply a weight calculated using the inverse of the relation between the mean density and the systematic.

The first question that we must overcome is which estimator for mean density of galaxy we will use for computing the systematic corrections. Both estimators the mean CMASS spectra density or the CMASS target density are biased, the density of CMASS spectra is biased because the survey is not complete and CMASS targets density is biased because the target failure rate is correlated with galactic latitude. As what we want is to correct the galaxy sample, the natural choice is to estimate

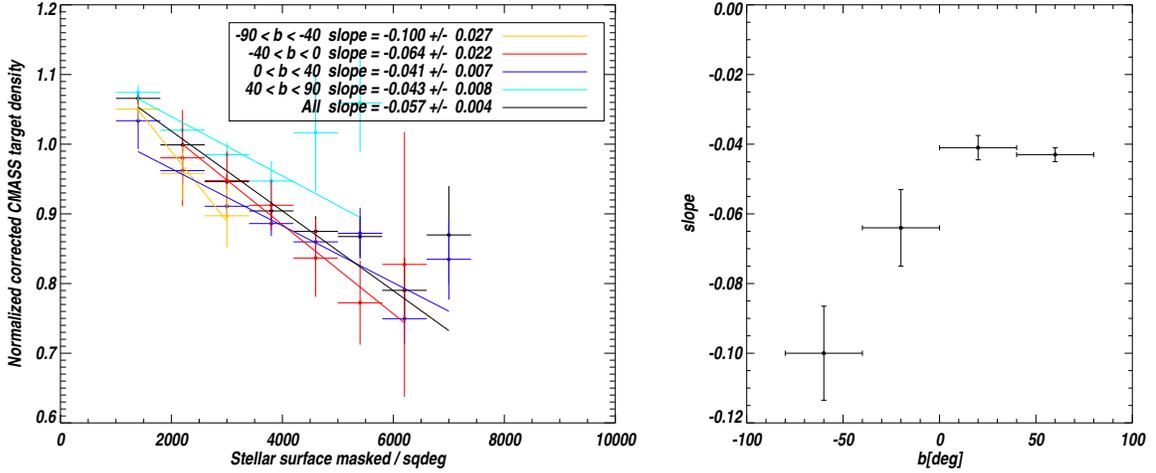


Figure V.20: Left panel. The relation between mean CMASS target density and surface masked by stars for the bins in galactic latitude: $[-90, -40]$ $[-40, 0]$ $[0, 40]$. Right panel. The slope of the linear fit for the relationship between mean galaxy density and surface masked by the stars as a function of the bin in galactic latitude

the corrections using the CMASS spectra density. In the left panel of figure V.21 the CMASS spectra density normalized to the mean density as a function of the stellar density is shown after and before corrections. In right panel of figure V.21, the correlation function is shown after and before stellar correction. As we observe the stellar density has an important effect in the correlation function, as the effect of the stars is not accounted in the random, the global effect is to add a constant noise in the correlation function, when the weights are applied this effect is taken into account and the correlation function tends to zero as predicted.

V.2.2 Correction of other sources of angular systematic errors

Ross et al found significant fluctuations of the mean density as a function of other variables. In this section I study the way to correct for multiple sources of systematic error and the effect of the corrections in the correlation function. The variables explored are: I-band sky Flux (we will denote as sky flux), the galactic extinction A_r (Extinction), I band PSF FWHM (Seeing), the airmass, the Schafly offset $r-i$ and $g-r$ (that combined gives place to the d_{perp} offset). In the figures V.22, V.23 and V.24 the pixelized HEALPIX maps with $n_{\text{pix}}=256$ show the angular distribution of the systematics explored. These maps were produced within the clustering working group by A. Ross and S. Ho. In an analogous way as stellar density, we calculate the relation between density of CMASS galaxies as a function of the possible source of systematic error. I exclude for the analysis the pixels without BOSS data, and the pixels outside the ranges defined in the table (outliers of distribution). In figure V.25 and V.26 the CMASS density normalized to the mean density as a function of the possible source of systematic error is shown. As we can observe from the figures, we verify that the main source of systematic error is the stellar density, we find a smooth dependence on the seeing and dust and an almost null effect of sky flux and airmass.

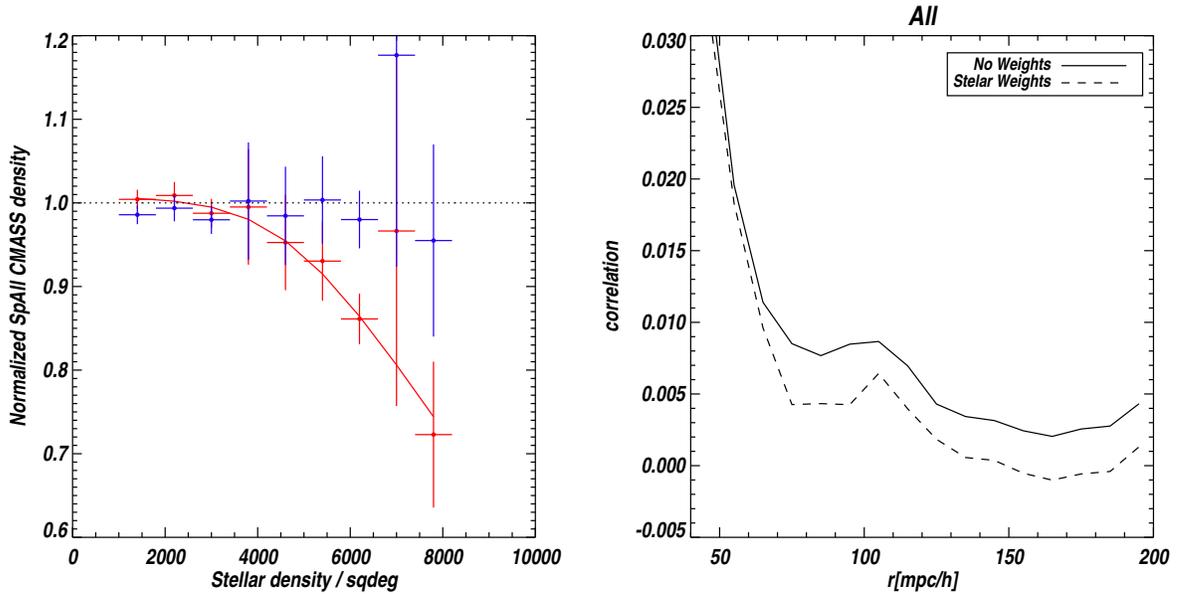


Figure V.21: The correlation function is shown after and before stellar correction

When we have more than one source of systematic errors we can not apply directly the standard procedure described before to correct them, we must account for the possible correlations between systematics. If variables are completely independent we should correct for each one separately and take the product of the weights derived to correct each systematic. To take into account of these correlations between variables we will use two methods: the iterative weights and the linear regression.

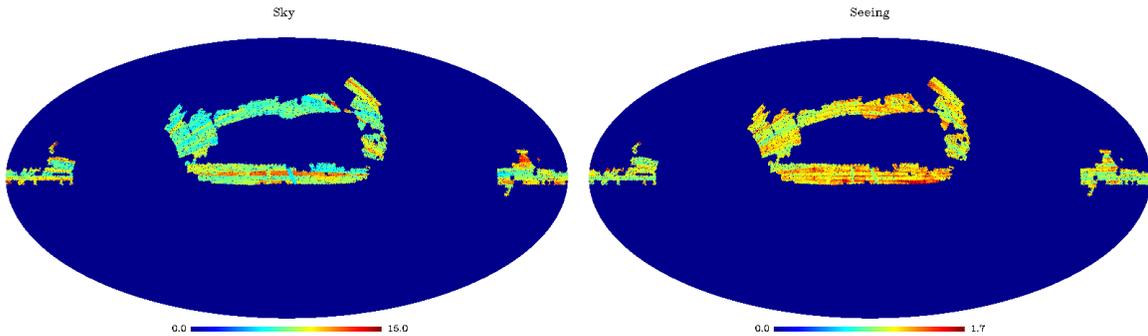


Figure V.22: I-band sky flux (Sky) and I band PSF FWHM (Seeing) maps with HEALPIX pixelization of $n_{\text{side}}=256$.

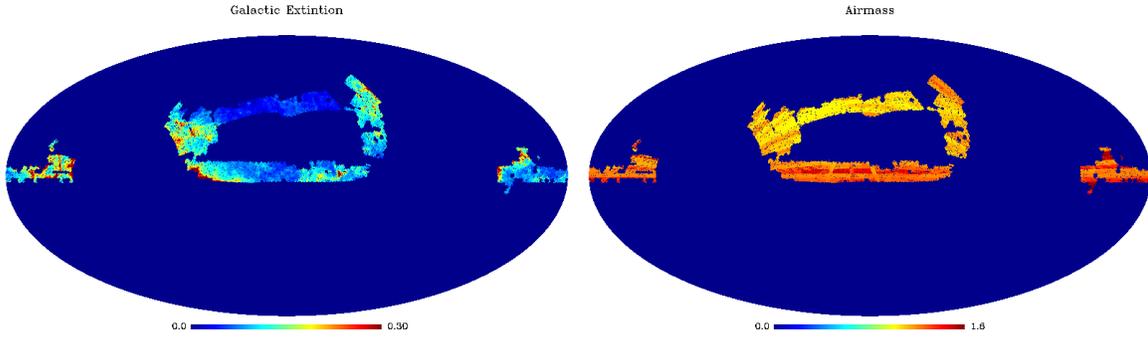


Figure V.23: Galactic extinction (A_r) and Airmass maps with HEALPIX pixelization of $n_{\text{side}}=256$.

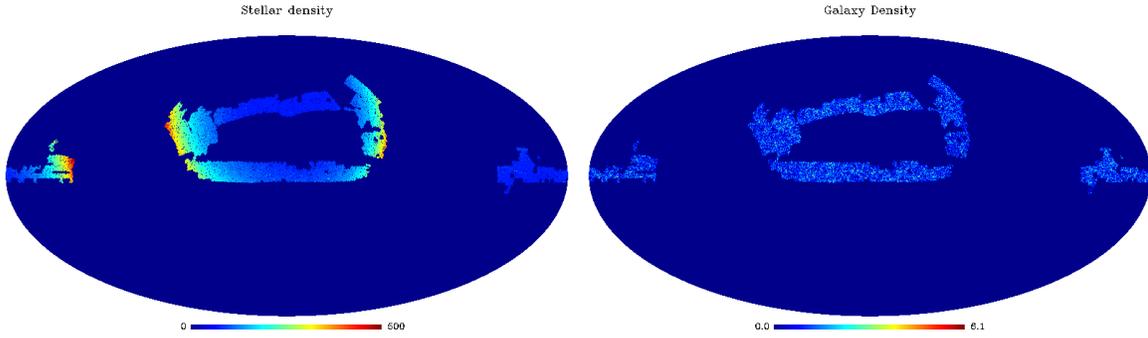


Figure V.24: Stellar Density and CMASS galaxy density with HEALPIX pixelization of $n_{\text{side}}=256$.

V.2.3 Iterative Weights

The iterative method has been proposed within the BOSS collaboration. It consists of dealing with each systematic separately in an iterative way, that is, to compute the dependency of the mean density with one systematic, correct it applying a weight to each galaxy proportional to the inverse of this relation and recompute the relation of the mean density for next systematic and correct it in an iterative way. The final weight computed in this way takes into account the correlations between variables. To apply this method I fit a spline of 5 nodes to the relation between the mean density and the systematic. The disadvantages of this method is that the fluctuations pixel to pixel are taken into account in the corrections what could result in an extra noise. We can try to avoid this over-fitting using a linear function instead, but even if we use a linear function, the resultant weight will be an expression non linear on the systematics and the problem is that we have no reason to predict a more complex dependency than linear. In Figures V.25 and V.26, the dependency of the mean density with the systematic is shown before correction as well as the spline line fit.

Now, in figure V.27 I show the correlation function computed with/without the iterative weight for both samples *MAIN008* and *GATH*. As we see the choice of catalog follows the same pattern

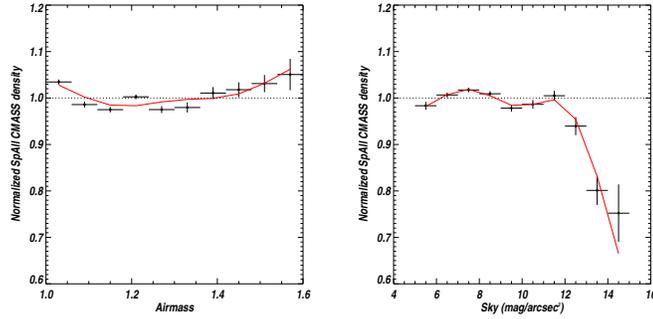


Figure V.25: The dependency of the mean density with the systematic is shown before correction as well as the spline line fit for the airmass and sky

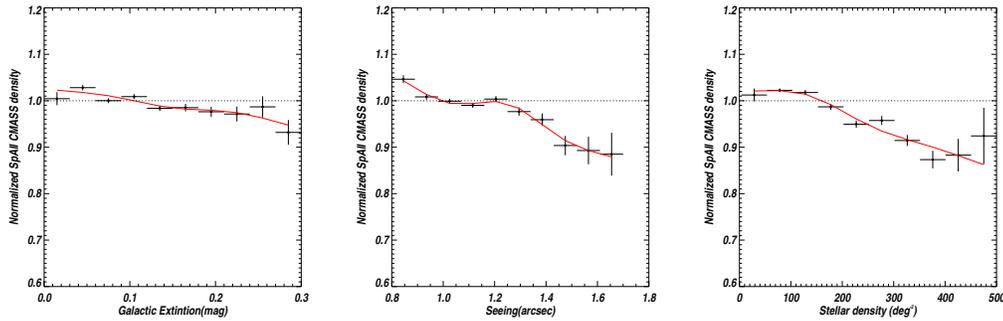


Figure V.26: The dependency of the mean density with the systematic is shown before correction as well as the spline line fit for the dust and seeing.

than without weights, using both catalogs gives very similar results, the small fluctuations could be explained by the size of the random set used to compute the correlation function. The choice to use a linear fit or a spline do not affect the correlation function.

V.2.4 Linear Regression Method

The second method explored to correct angular systematics was a linear regression applied pixel by pixel. We suppose that the value of the normalized density $\delta_g(p) = n_g(p) / \langle n_g \rangle$ in each pixel p is a linear regression of 5 independent variables. These variables are defined as the deviations from the mean value of the systematic. Then, we compute the best linear combination of parameters that follows the relation.

$$\delta_g = \delta_{g0} \left(1 + \sum_{i=0}^5 a_i * \delta_{sys}^i(p) \right) \quad (\text{V.23})$$

where, $n_g(p)$ is the value of the measured mean density at the pixel p , $\delta_{sys}^i(p)$ is the value of the fluctuation of systematic in relation with the mean value of the systematic, where i runs over the

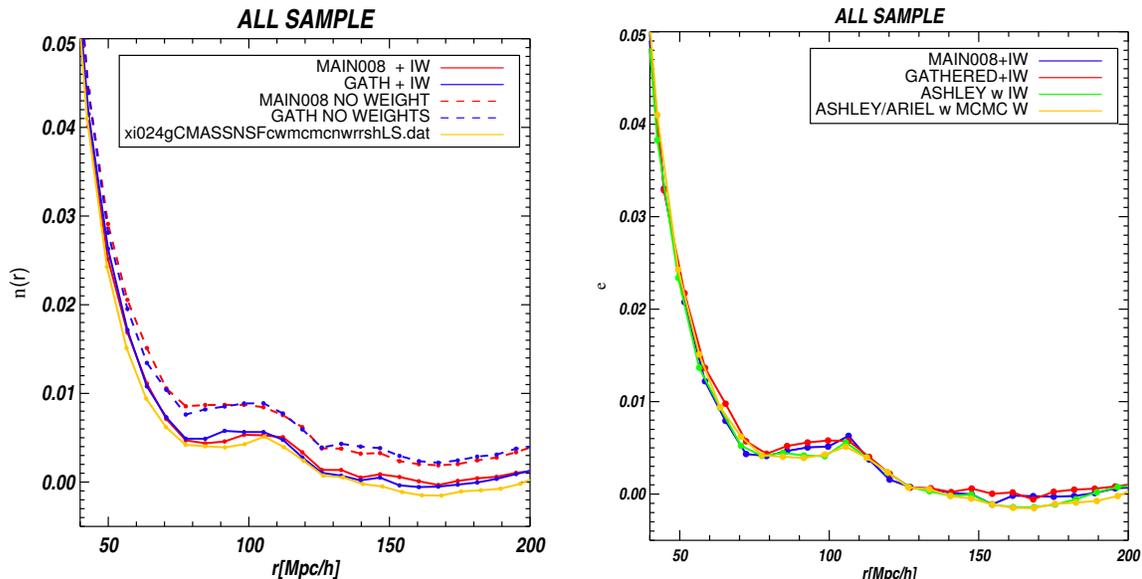


Figure V.27: Correlation function with the iterative weights using A. Ross files and applying an iterative correction with spline for the dependance in the systematic as a function of the mean density.

number of systematics studied and δ_{g0} is the real normalized mean density .

$$\delta_{sys}^i(p) = sys^i(p) - \langle sys^i \rangle \quad (V.24)$$

Under the assumption that $\delta_0 = n_0 / \langle n_0 \rangle = 1$. Then, using a linear regression we find the parameters a that better agrees with the observed density. In this case the weights will be given by:

$$w^{-1}(p) = \sum_{i=0}^5 a_i * \delta_{sys}^i(p) + 1 \quad (V.25)$$

In figure V.28 and V.29 the relation between the mean density before and after corrections is shown. The lines corresponds to the best linear fit to the data points. As we can observe from the values of the linear fits, after corrections the mean galaxy density does not depend any more on the systematics. From the plots we confirm that the principal sources of systematics errors are given by the stellar density followed by the dust and the seeing. The airmass and the sky seems to have negligible effect.

Now we will compare the results from the iterative method with the results from linear regression. In principle the weights from linear regression must be less agressive given their linear nature. In figure V.30 the correlation function computed with the same sample and same random set but using in one case the weights from the iterative method and in the other case the weights from the linear regression is shown. We observe that both methods give similar results showing the robustness of the results face to the weighting scheme.

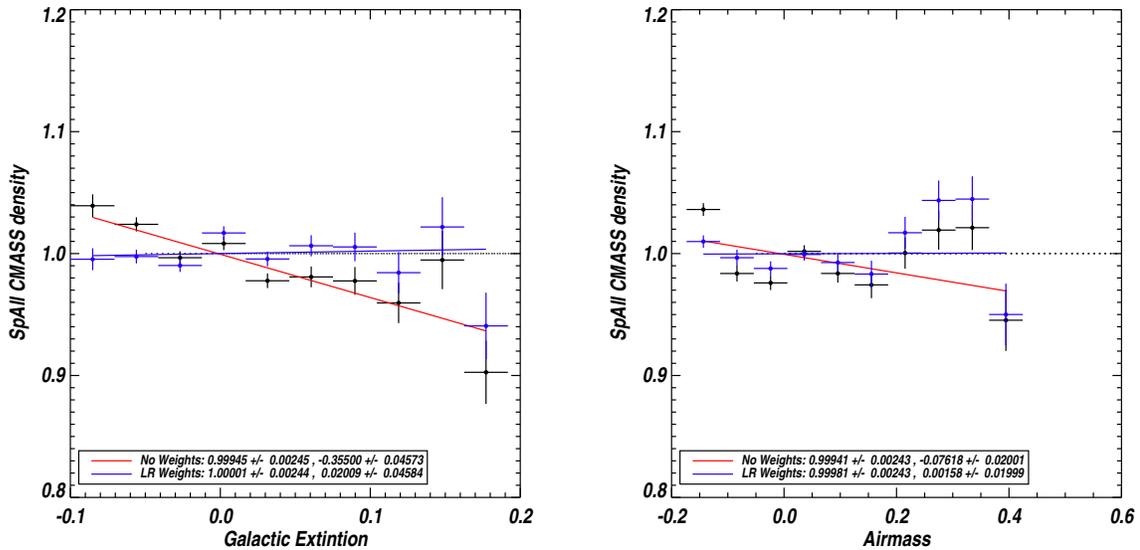


Figure V.28: The dependency of the mean density with the systematic is shown before and after correction as well as the linear line fit for dust and seeing.

V.2.5 Effect of angular corrections in correlation function

In the last subsection I tested the two methods to compute the weights and I show they give similar results. Now we will explore which is the effect of each systematic in the correlation function. For this test I used the iterative method. I correct for each systematic independently and I study the effect in the correlation function. Then I test the effect of applying the four systematics together excluding stellar correction. In figure V.31 the results are shown, as we can see from the figure the larger effect is generated by stellar density, that decrease globally the correlation in $\sim 1\%$, the others systematics decrease the correlation only in $\sim 0.5\%$, only the seeing increase the correlation value globally.

Finally I test the correction from angular systematics with the catalogs generated with the different methods to correct the target failure. As we have seen before, this choice has an important effect in the correlation function. In figure V.32 and in V.33 the correlation function computed with the different different target failure correction before and after correcting from angular systematics is shown. I present the results for the whole sample and separating North and South. As we can observe from the figures for the whole and north samples the behaviors of the correlation function seems to follow the same patterns than without corrections, the south hemisphere seems to present an atypical behavior with the method. This result seems to point out that the best way to correct from bad targets is the method 2, which gives consistent results between both hemispheres.

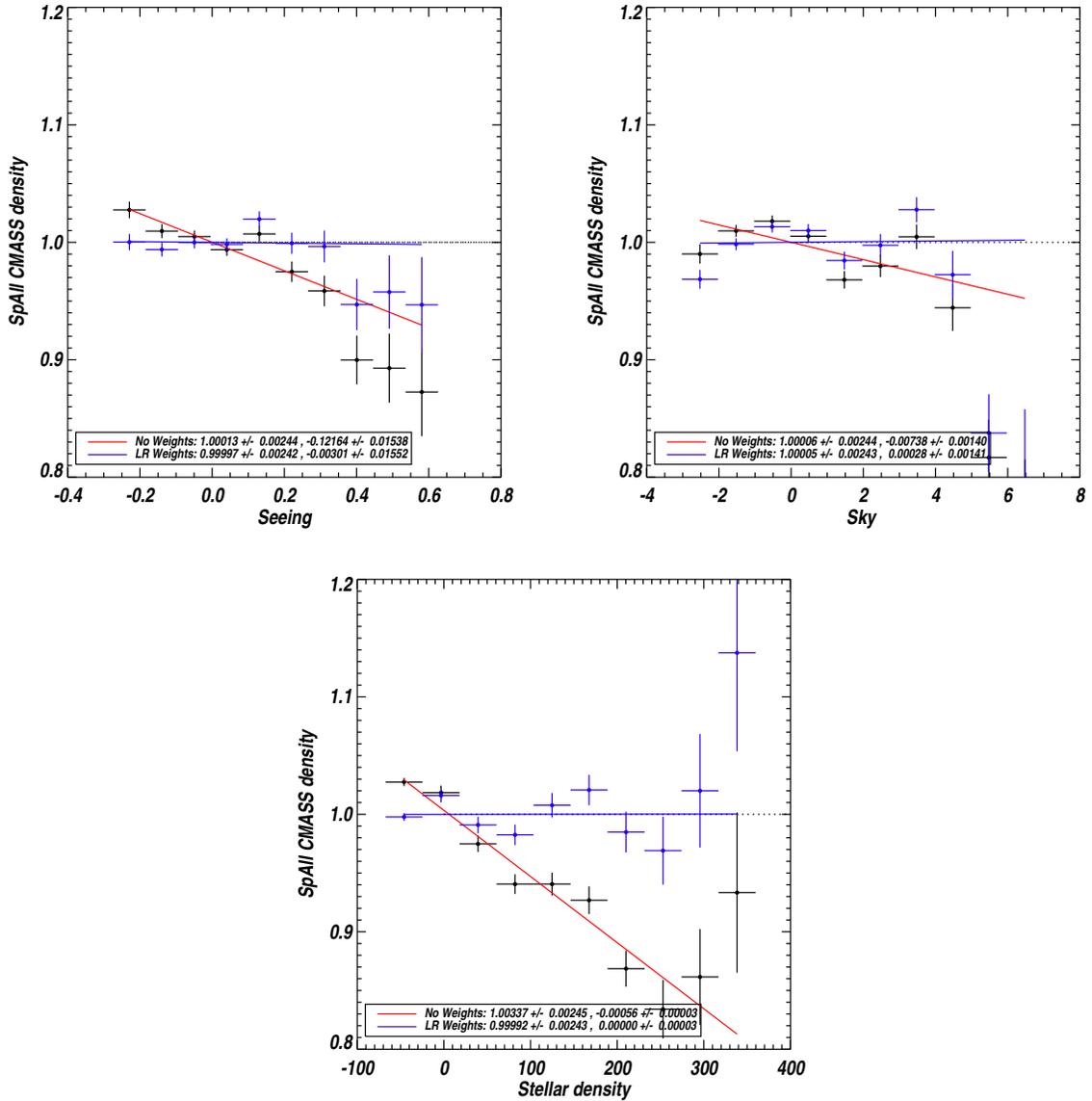


Figure V.29: The dependency of the mean density with the systematic is shown before and after correction as well as the linear line fit for dust and seeing.

V.2.6 Summary of Correlation Function Corrections

In this section I described the correction of the correlation function from angular systematics effects. I verify that the main systematic is the stellar density. I study the stellar density in detail, separating the dependency of the mean density and stellar density in bins of galactic latitude. I found a different dependency between South and North at low galactic latitude. I explored the explanation for this systematic effect calculating the surface masked by stars and I studied its dependency with stellar

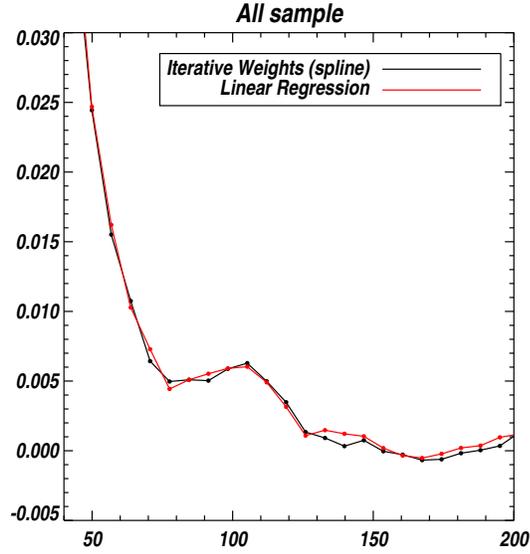


Figure V.30: Weights computed with the iterative method compared with weights computed with the linear regression..

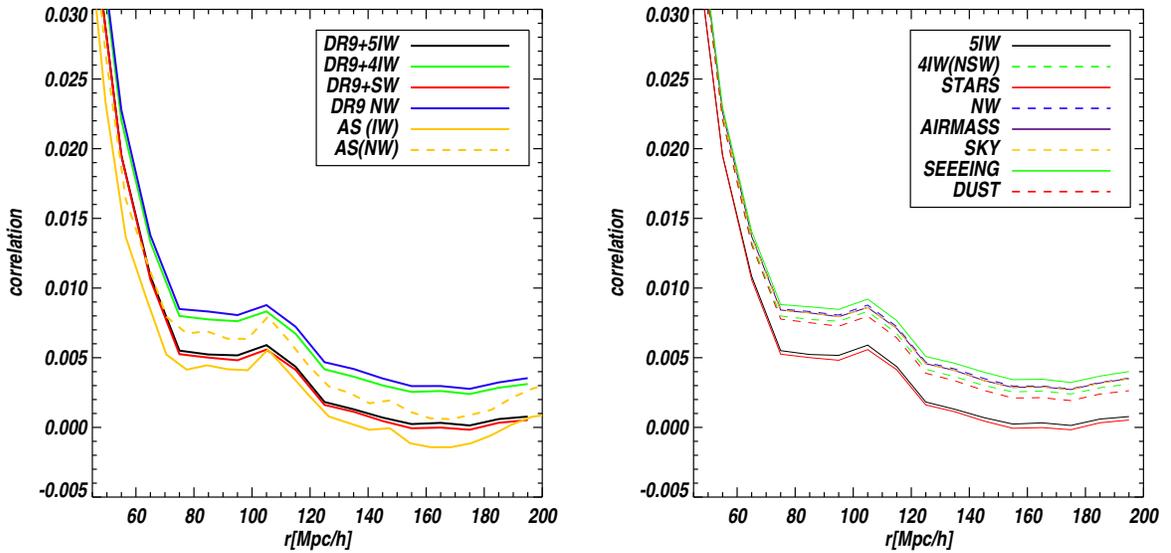


Figure V.31: Correlation function with weights computed separately for each systematic

density. An evolution of this dependency in terms of galactic latitude was found which indicates an unknown systematic affecting South hemisphere. I tested different sources of systematics effects (seeing, sky, extinction, stars, air mass, dperp offset) and 2 methods (Iterative weights and Linear

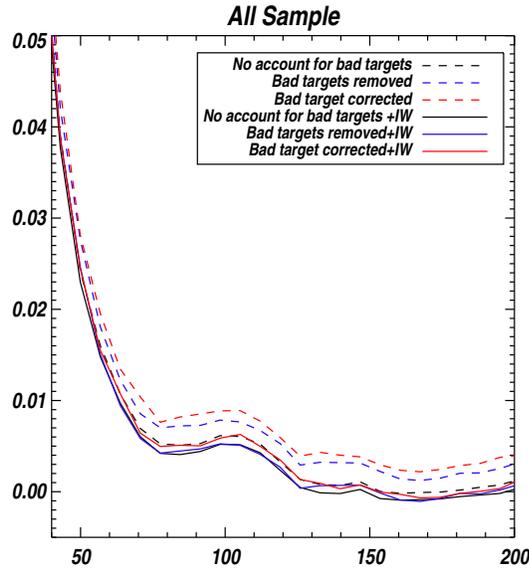


Figure V.32: The correlation function computed with the different different target failure correction before and after correcting from angular systematics for the hole sample

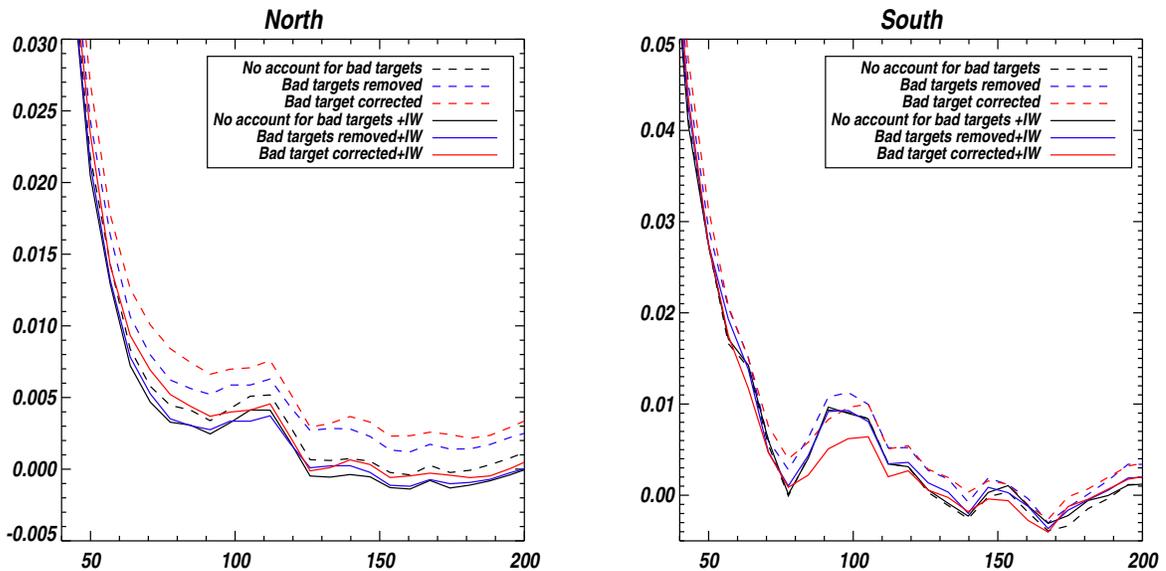


Figure V.33: The correlation function computed with the different different target failure correction before and after correcting from angular systematics separating North and South

regression) to estimate the weights to correct from this systematic error. Both methods give similar results in terms of the correlation function. The main effect of correcting from angular systematics is

to decrease the correlation function level globally. I tested the effect of the correction from angular systematics and the method from correcting bad targets. I found that method 2 is consistent between both hemispheres applying the weights.

V.3 Covariance Matrix

In the last section I have developed the procedure that gives rise to the correlation function from BOSS data corrected from angular systematic effect. In order to do cosmological analysis with this correlation function a fundamental ingredient is necessary, the error estimation. In order to compute the errors three different kind of simulations would be tested. I begin describing briefly each kind of simulation. Then, I present the covariance matrix compute with them. In a later time we compare results between simulations to select and validate the method to be used to compute the covariance matrix. This covariance matrix would be used to do the cosmological parameter estimation in the final chapter.

V.3.1 Simulations Description

Lasdamas

Lasdamas simulations is a set of cosmological N-body simulations that follow the evolution of dark matter in the universe. The cosmological model used is $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_b = 0.04$, $h = 0.7$, $\sigma_8 = 0.8$ and $n_s = 1.0$. The power spectrum used to compute the density fluctuations was generated with CMBFAST (Seljak & Zaldarriaga 1996). The initial positions and velocities were computed for the particles using the 2LPT code (Scoccimarro). The gravitational evolution was performed using the publicly available Gadget-2 code (Springel et al. 2005). The halos are identified using a parallel friends-of-friends (FOF) code with a linking length of $b=0.156$ in units of the inter-particle separation. The simulations were generated with the same initial power spectrum but with random phases. The version generated for BOSS analyses is called CMASS *Gamma Weak*. It consists of a set of 40 realizations for 3 regions each one denoted A,B,C, defined as in (?) correspondent to the early data coverage, that is, the first semester of BOSS data (See Fig. V.34). The definition of the regions are in table N. The coverage is shown in figure V.34

The simulations do not include fiber collisions and they include the redshift distortions from velocities. The galaxy mocks have uniform radial selection function in the range $z=[0.4, 0.6]$. The halos of dark matter were populated using the HALO model, a HOD was applied, consistent with a central galaxy placed in the deepest potential particle and the mean velocity was assigned to it, the satellite galaxies are placed randomly selecting dark matter particles from the halo using their position and velocities. The 5 parameters of the HOD were fitted (Zheng, Coil, and Zehavi 2007) globally in three regions to reproduce the galaxy number density and the projected correlation function panel $w_r(r_p)$ (See Fig. V.34).

The simulations were created from The Carmen simulation at a static redshift $z=0.52$, each one of them consists of a box of $1Gpc/h$ sideleng. The simulations of regions A,B and C were generated using the same boxes thus they must be analyzed separately.

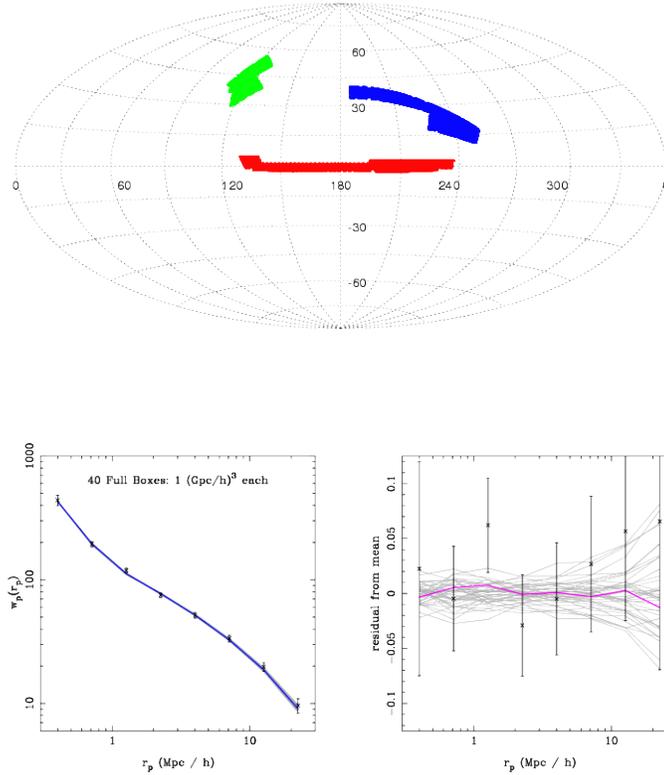


Figure V.34: Left panel. Coverage of LAS DAMAS CMASS simulation. Right panel. The projected correlation function for the first semester of BOSS Data used to calibrate the simulations

PTHALOS Simulations

PTHALOS (Roman Scoccimarro & Ravi Sheth (2001) (?)) is a fast algorithm which generates point distributions with similar clustering statistics to those of the observed galaxy distributions. This approach reproduce the results of numerical simulations requiring minimal computational resources. The motivation for using PTHALOS simulations is that this kind of simulations are faster in relation to n-body simulations and provide results similar to them. The gaussian simulations do not correspond exactly to the distributions generated with N-body simulations. The advantage of PTHALOS is that in term of the statistics of 2 and 3 orders the distribution generated with PTHALOS are correct in large scale even in the non linear regime. The PTHALOS algorithm consists in the generation of a density field of dark matter at large scale using lagrangian perturbation theory at second order and it follows some corrections at small scale. These corrections consist in partition the density field in halos based on the spherical collapse model and the mass function and then using the distribution of dark matter within halos generated from N-body simulations (Profiles NFW). Finally the galaxies are generated using a HOD prescription that is fitted to the data at

small scales using the projected correlation.

The simulations PTHalos are documented in Manera et al 2012. They consist of 600 mocks generated separating North and South Hemisphere. The simulations are generated at $z=0.52$. The mass of the halos have been calibrated with Las damas simulations (Mc Bride 2012) The mocks follows the mask and completeness from 18 July 2011 for the North and HOD parameters have been set to fit the projected $w_p(r_p)$, the redshift distortions are included. The mocks are not accurate at small scales, below 2.5 Mpc/h. The mocks follows the radial distribution from data.

In figure V.35 I show the mean correlation function of the 600 simulations for the north and south mocks.

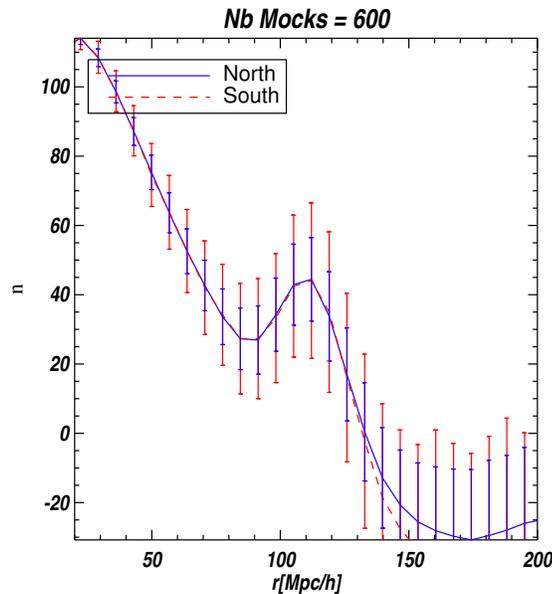
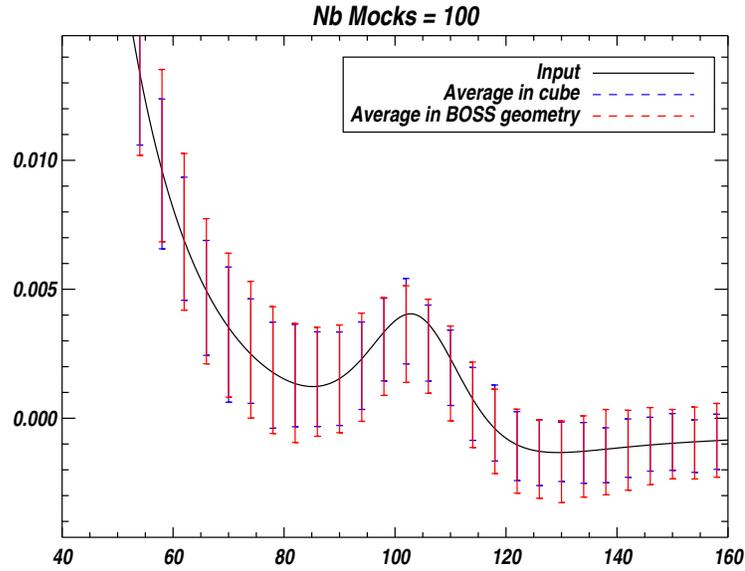


Figure V.35: Mean correlation function of 600 PTHALOS simulations

LogNormal

The log normal simulations are generated using a non linear spectra from CAMB (?) with the same cosmological parameters as LAS DAMAS simulation. The non linear spectra in generated at redshift $z=0.52$. The non linear behavior in CAMB is done using fitted expression from N body simulations documented in (?). This non linear behavior do no include the BAO. To include the non linear behavior to the BAO scale, we use the standard procedure explained in last chapter that is to take linear power spectrum and we apply a gaussian factor to mimic the non linear degradation of the BAO oscillations. The redshift distortions are included as explained in last chapter. The values of the bias and beta parameters are fitted to the observed clustering. The mocks follows the geometry and completeness of Boss, North and South hemisphere included. The radial selection function function is given by number density of BOSS data. We generate 1000 simulations with a random set 5 times greater than the sample.



V.3.2 Covariance Matrix Comparison

In figure V.36 I show the correlation matrices for the North (left) and South(right) hemispheres generated from PTHALOS simulations. In figure V.37 I show the covariance matrices for the North (left) and South(right) hemispheres generated from PTHALOS. In figure V.38 the Boss Covariance

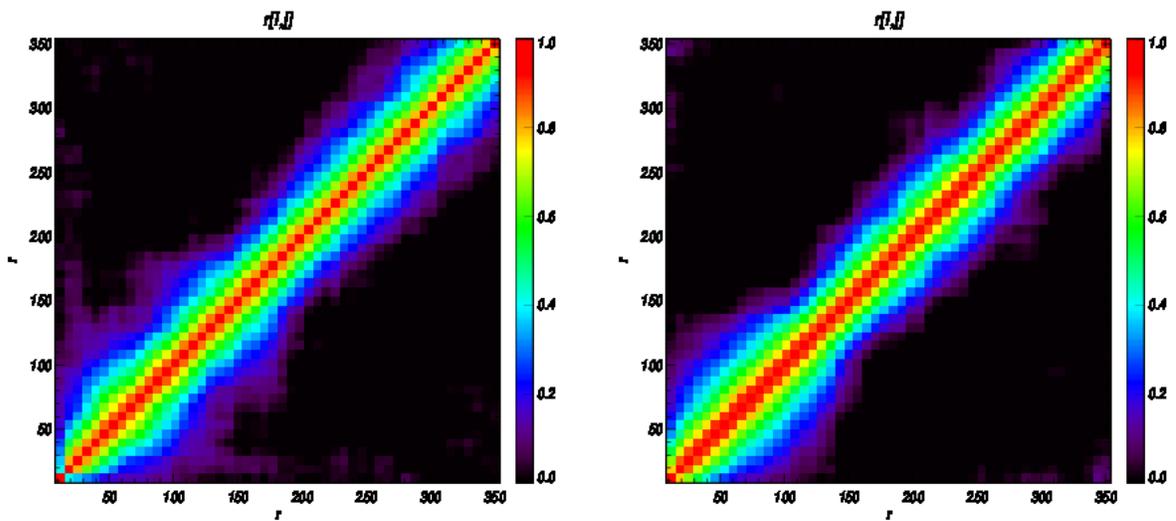


Figure V.36: Boss Correlation matrices for the North (left) and South(right) hemispheres generated from PTHALOS simulations.

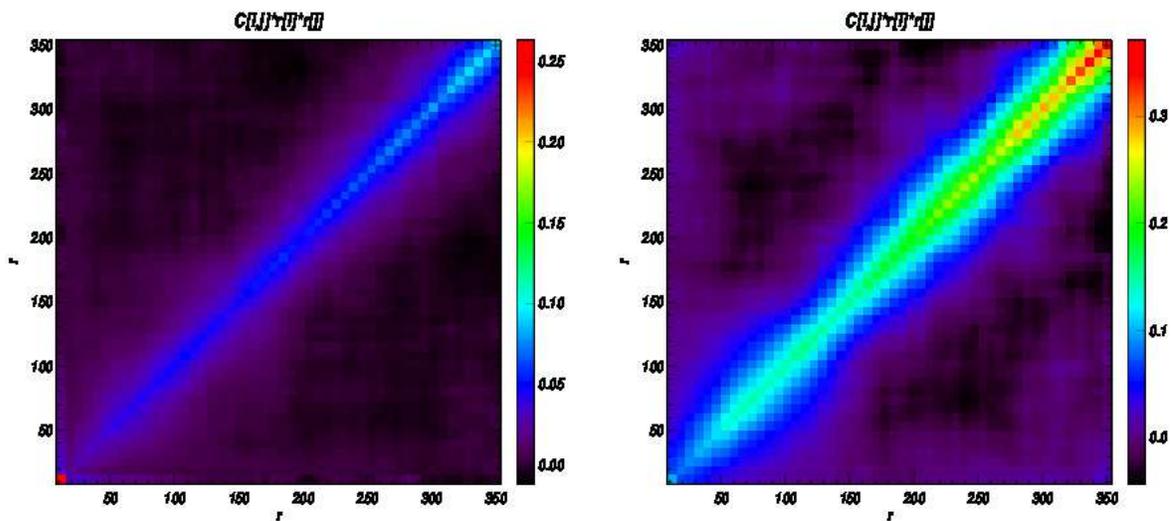


Figure V.37: Boss Covariance matrices for the North (left) and South(right) hemispheres generated from PTHALOS simulations.

matrix and the correlation matrix for the all sample generated from LOG NORMAL simulations are shown. Finally in figure V.39 I show the covariance matrices from Lasdamas for each region

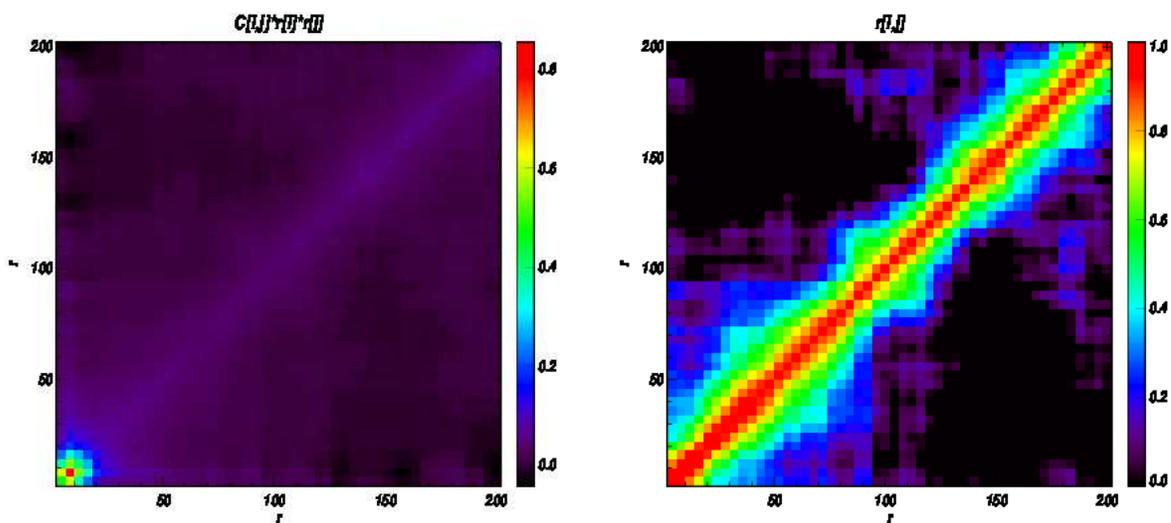


Figure V.38: Boss Covariance matrix and the correlation matrix for the all sample generated from LOG NORMAL simulations.

A,B and C and their respective correlation matrices.

As we can verify from the figures V.36, V.38 and V.39, all correlation matrices present an enhanced diagonal as expected. In the case of log normal simulations we observe an additional structure with respect of the result of Lasdamas and PTHALOS. Probably this extra structure comes from the replication of the box to generate the mocks catalogs. In the case of Lasdamas we observe some fluctuations between regions probably derived of their small size.

V.4 Summary

To conclude the chapter I present the results with the final catalog with errors given by the diagonal terms of the covariance matrices from PTHALOS for the North and the South V.40 and the combined Sample ?? as well as the pondered mean of North and South. As we observe from the figures our results are completely compatible with the official catalog. For the next section we will use the official catalog to generate the constraints in cosmology. The aim of this section was to generate our own pipeline from the beginning and to show we are able to reproduce all steps to do large scale analysis. This procedure permit me to understand very well the sample we are using to constrain the cosmology to understand the corrections involved in the catalog generation and in the weighting scheme.

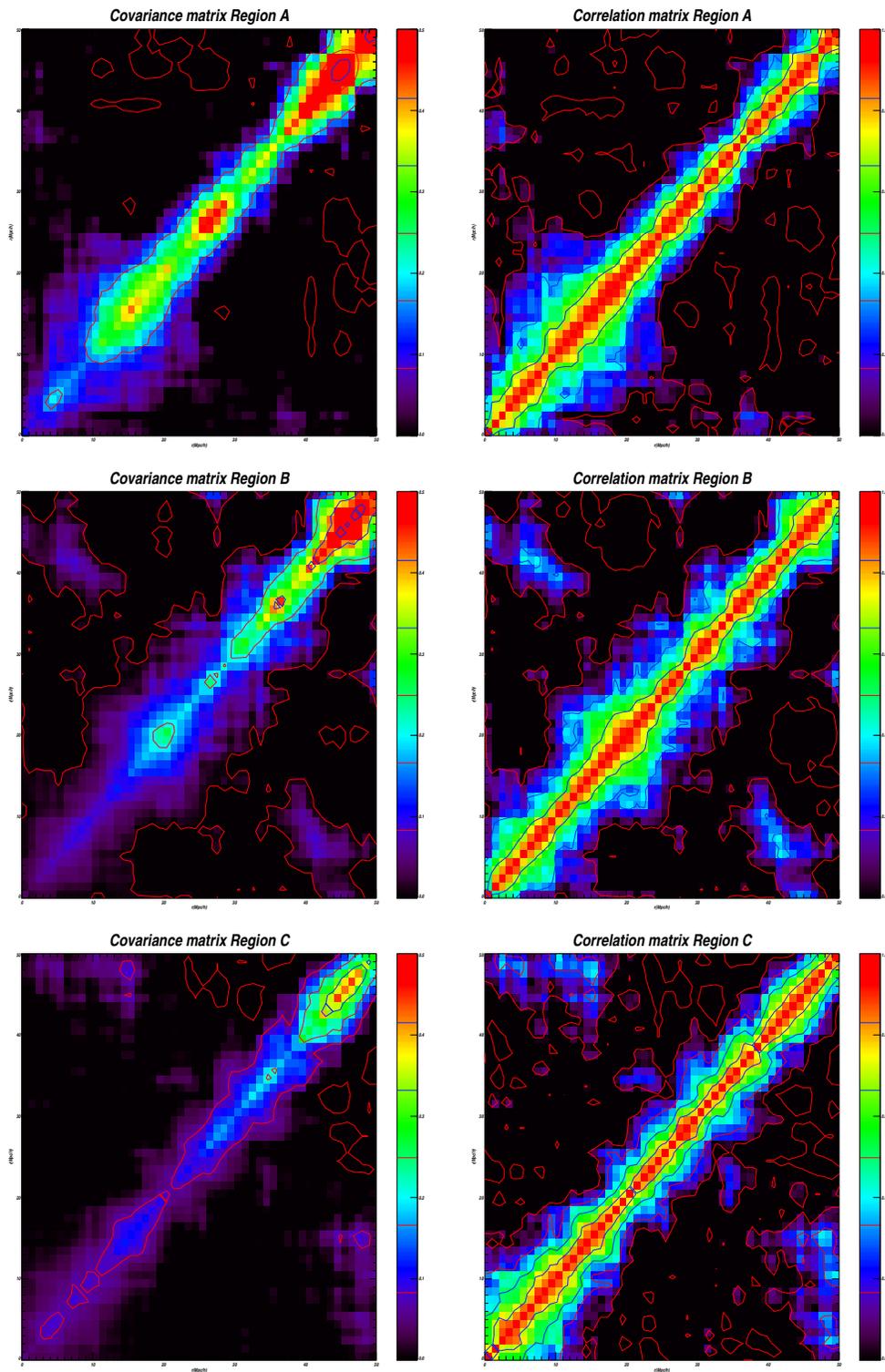


Figure V.39: Boss Covariance matrix and the correlation matrix for the all sample generated from LAS DAMAS CMASS simulations for regions A,B,C

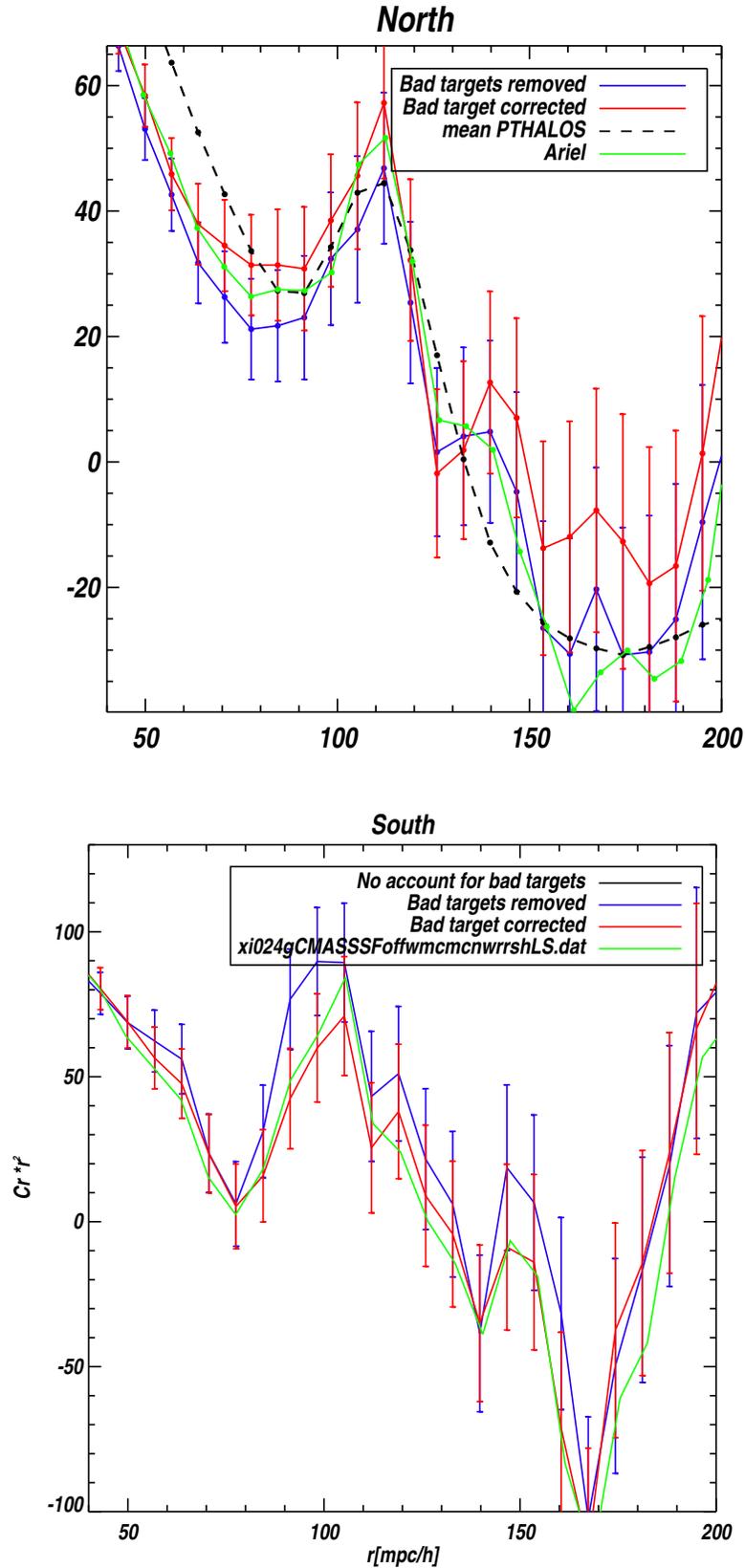


Figure V.40: The correlation function with the final catalog for the North and the South. The errors bars given by the diagonal terms of the covariance matrices from PTHALOS.

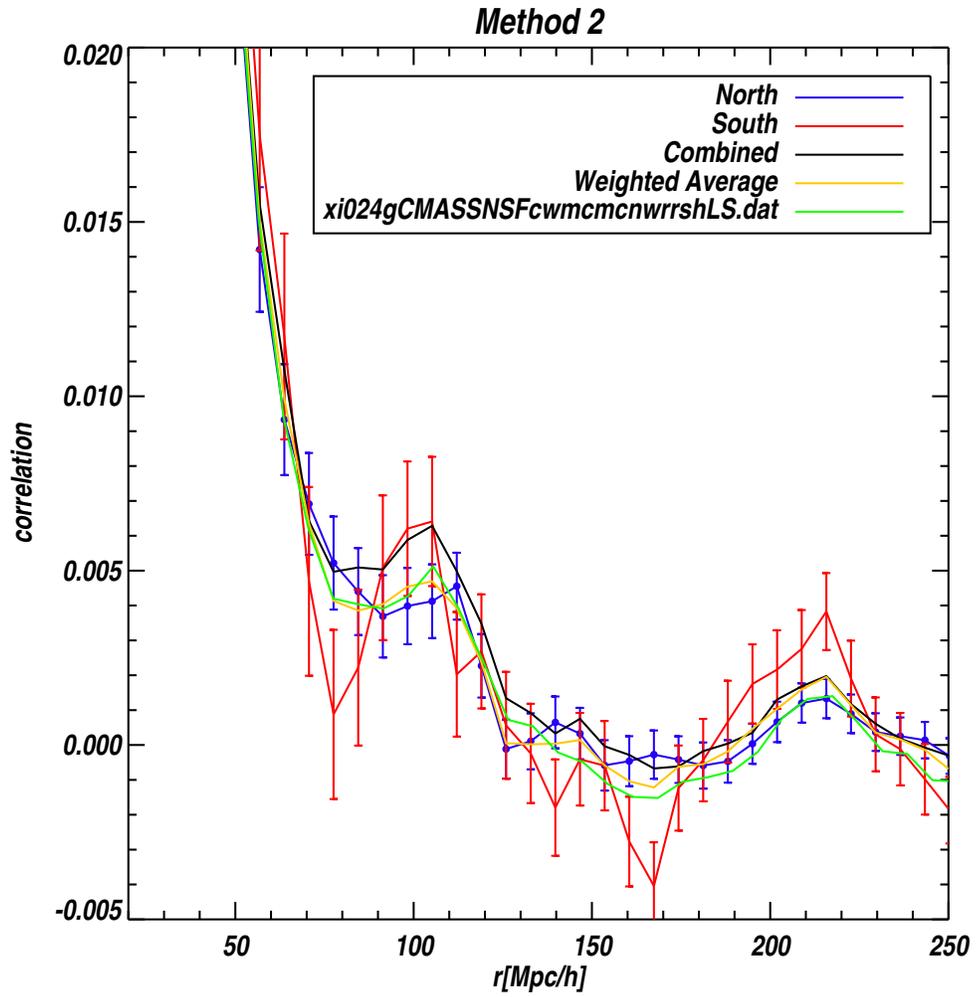


Figure V.41: The correlation function with the final catalog for combined sample and weighted sample. The errors bars given by the diagonal terms of the combined covariance matrices from PTHALOS.

VI. Studying Redshift Distortions with simulations and more

One of the most promising tools to investigate modified gravity is redshift-space distortion effects caused by peculiar velocities in galaxy redshift surveys. In this chapter I study the redshift distortions using simulations. The chapter is organized as follows: in the first section I introduce the redshift distortions theory. First I describe schematically the redshifts effects, then I present the linear regime description formulated by Kaiser (1987) in Fourier space under the plane parallel approximation (Hamilton, 1998) and the decomposition in spherical harmonics. As we will see, the power spectrum in redshift space would appear amplified by an anisotropic factor depending of the angle μ_k and the linear redshift distortion parameter β . The β parameter is related to the bias and the matter density via the growth factor rate $f(\Omega)$. The β parameter enables to study two different physical effects of the growth of structure, the expansion history encoded in Ω_m and the gravitational theory encoded in the growth index γ following the parametrization of (Linder, 2005) of the growth of structure $f(\Omega_m) = \Omega_m^\gamma$. Finally, I will focus on the way of measuring the linear redshift distortion parameter β . In the second section I describe the studies I realize with simulations to study the possibility to measure β parameter from mock catalogs of galaxies. The first study is done using lognormal simulations where the density field is linear and the input correlation function is known. I used another kind of mock catalogs. The second kind of mocks catalogs is issued from N-body simulations. In N-body simulations the non linear evolution is taken into account but are very demanding in computing-time. Thus I used a single realization and resampling methods were applied to generated several realizations. These simulations were used to compare to results obtained with lognormal simulations for a realistic scenario. In both cases, I estimated the β parameter from the ratio of real to redshift space monopole. In section four I present some preliminary results with real data, BOSS. In this case we do not have access directly to the real space correlation function, thus the estimation of the β parameter must be computed from the normalized quadrupole.

VI.1 Introduction

In redshift surveys the redshift of the objects is used to estimate the radial coordinate. Because the galaxies have peculiar velocities, the real distance is distorted by the component parallel to the line-of-sight of peculiar velocities. The relation between true distance r and the redshift distance s is:

$$\vec{s} = \vec{r} + \vec{v} \cdot \vec{r} \quad (\text{VI.1})$$

These displacements are called redshift distortions. The redshift distortions could be separated in two effects that occurs at different scales. In the large scales, the dominant effect is provoked by the bulk motions of galaxies and the description could be done in terms of the linear regime. By contrary in the small scale regime the random motions of virialized galaxies will dominate and the description have to take into account the non linearities.

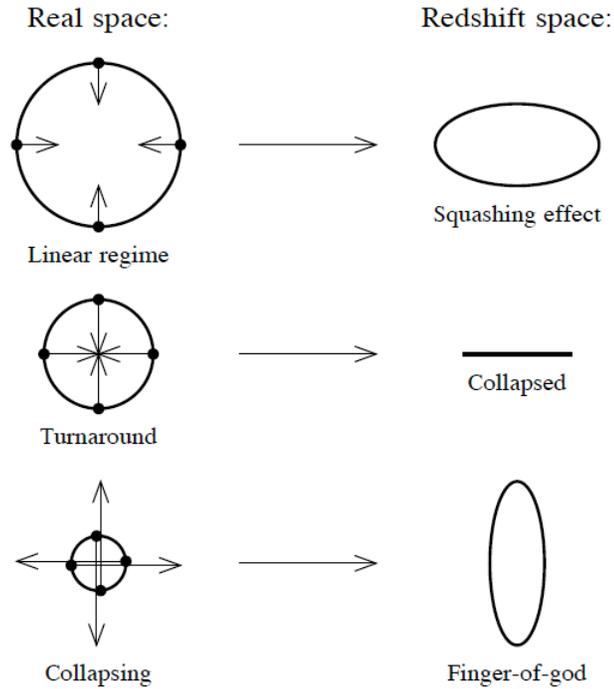


Figure VI.1: Illustration of Redshift Distortions from Hamilton (Hamilton, 1992)

In order to illustrate the effects let us take the scheme from Hamilton (Hamilton, 1992) in figure VI.1. In this scheme a spherical overdensity in the center is shown surrounded by a uniform distribution of matter, represented by the circle where the dots are galaxies and the arrows their peculiar velocities. The three levels represent the effect of redshift distortions at different scales, from large (top) to small (bottom) scales. As we observe from the figure in the large scale regime illustrated by the first block, the main effect would be an squashing effect in the direction of the light

of sight, breaking the spherical symmetry of the shell in real space (left) when passing to the redshift space (right). We notice that in large scale regime the velocities are small compared with the scales that we are studying (radius of the sphere). Descending to the smaller scales, the velocities are larger compared with the radius. There is a particular point where peculiar velocities cancels the Hubble expansion, this point is known as *turnaround*, in this particular point all the points in the redshift space will collapse to a single point. In the third block the small scales are represented, in this case we observe that the velocities are larger than the shell radius, thus, the mixing between different collapsing shells generated by the virialization process will produce an elongated pattern known in the literature as the Fingers Of God (FOG). The main effect is a radial stretching pointing to the observer.

In the next section I present the theory behind the redshift distortions in the linear regime I describe the relation of the distortions with the growth of structures and finally I will focus on the way of measuring the linear redshift distortion parameter β .

VI.1.1 Linear Regime Description

Even if redshift distortions add complexities to the analysis of large scale structure, they are interesting because they encode important information about the growth of structure. The linear regime description was formulated by Kaiser (1987) in Fourier space under the plane parallel approximation that means the observer is sufficiently far from the objects that the displacement is effectively parallel. A review of linear regime could be found in (Hamilton, 1992). In this approximation the power spectrum in redshift space takes the following form:

$$P_s(k) = (1 + \beta\mu_k^2)^2 P(k) \quad (\text{VI.2})$$

with $P(k)$ the power spectrum in real space, the subscript s denotes the analogue in redshift space, μ is the cosine of the angle between the line of sight and k and β is the linear redshift distortion parameter. Thus this equation indicates that power spectrum in redshift space would appear amplified by an anisotropic factor depending of the angle μ_k . This parameter is related to the matter density and gives information of the growth factor (growth of the newtonian gravitational potential $f(\Omega)$) and bias.

$$\beta = f(\Omega_m) b \delta_m = \frac{1}{b} \frac{d \ln D}{d \ln a} \quad (\text{VI.3})$$

where D is the linear density growth factor and f the velocity growth factor, that could be written as:

$$f(\Omega_m) = \Omega_m(a)^\gamma \quad (\text{VI.4})$$

where γ is the gravitational growth index (Linder, 2005), $\Omega_m(a)$ is the matter density.

This parametrization of the growth factor comes from a fitting formula for linear perturbation growth from (Linder, 2005). The idea is to separate the cosmic expansion history of the cosmic growth history. This fitting formula is accurate to 0.05%–0.2%. In this parametrization γ quantifies deviations from standard framework. In this way the β parameter enables to study two different physical effects of the growth of structure, the expansion history encoded in Ω_m and the gravitational theory encoded in γ . For standard gravity the value of gamma is $\gamma = 0.55$ and depending of the gravity theory it takes other values. The equivalent analysis in configuration space was done by

Hamilton (Hamilton, 1998). The correlation function is the Fourier transform of the power spectrum, giving the following expression for the correlation function in redshift space:

$$\xi_s(r) = (1 + \beta(\partial/\partial z)^2(\nabla^2)^{-1})^2 \xi(r) \quad (\text{VI.5})$$

where $(\nabla^2)^{-1}$ is the inverse laplacian.

VI.1.2 Multipolar developpement of 2PCF

Let us consider for the following the separation between pairs in two components, the component parallel to the line of sight r_{\parallel} and the componenet perpendicular r_{\perp} . The solution of VI.5 could be expressed in terms of spherical harmonics:

$$\xi(r_{\parallel}, r_{\perp}) = \xi_0(r)P_0(\mu) + \xi_2(r)P_2(\mu) + \xi_4(r)P_4(\mu) \quad (\text{VI.6})$$

with:

$$\begin{aligned} \xi_0(s) &= (1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2)\xi(r) \\ \xi_2(s) &= (\frac{4}{3}\beta + \frac{4}{7}\beta^2)[\xi(r) - \xi(r)] \\ \xi_4(s) &= \frac{8}{35}\beta^2[\xi(r) + \frac{5}{2}\bar{\xi}(r) - \frac{7}{2}\bar{\bar{\xi}}(r)] \end{aligned} \quad (\text{VI.7})$$

where $\mu = \hat{r} \cdot \hat{x}$ is the cosine of the angle between line-of-sight r and the pair separation x (Fig.??), P_l are the Legendre polynomials ¹ and the following expression for $\bar{\xi}(r)$ and $\bar{\bar{\xi}}(r)$:

$$\begin{aligned} \bar{\xi}(r) &\equiv 3r^{-3} \int_0^r \xi(s)s^2 ds \\ \bar{\bar{\xi}}(r) &\equiv 5r^{-5} \int_0^r \xi(s)s^4 ds \end{aligned} \quad (\text{VI.9})$$

In Fig.VI.2 I show the theoretical prediction for $\xi(r_{\parallel}, r_{\perp})$ for a LCDM model in the linear regime in real and redshift spaces. From figure we observe the effect of redshift distortions in linear regime is an squashing effect in the line of sight direction changing the circular patterns for ellipsoid due to the bulk motions of the galaxies.

In the Fig.VI.3 I show the theoretical prediction for the monopole, quadrupole and hexadecapole again in the linear regime approximation. The monopole term ξ_0 corresponds to the average angle correlation function usually used to study large scale structure. In the absence of redshift distortions the terms quadrupole and hexadecapole must be null. The negativity of the quadrupole term reflects the line-of sight compression of clusters causing by this infalling velocities. The hexadecapole and all terms β^2 are related to the dispersion of velocities.

¹The Legendre polynomials l=0, 2, 4 are:

$$\begin{aligned} P_0 &= 1 \\ P_2 &= 3(\mu^2 - 1)/2 \\ P_4 &= (35\mu^4 - 30\mu^2 + 3)/8 \end{aligned} \quad (\text{VI.8})$$

Figure VI.2: Theoretical prediction of $\xi(r_{\parallel}, r_{\perp})$ in the linear regime approximation in real and redshift space

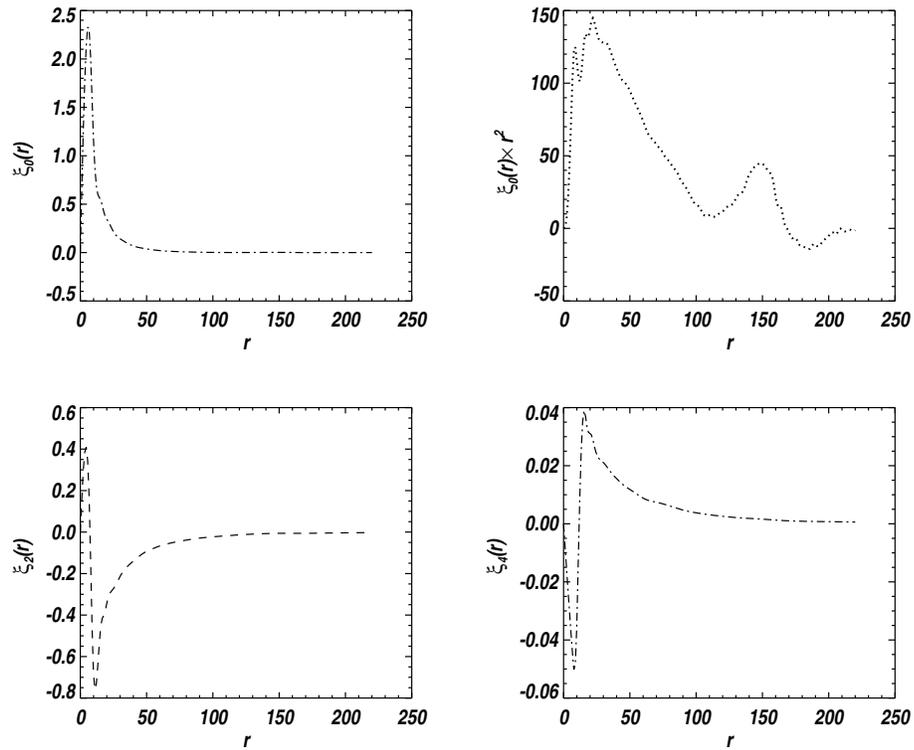


Figure VI.3: Theoretical monopole, quadrupole and hexadecapole for linear redshift distortions

VI.2 Estimation of the Linear Redshift parameter β

The expressions shown in Eq. VI.7 permit us to access to the β information in the case of simulations when we have access to real $\xi(r)$ and redshift correlation function (monopole) $\xi_0(s)$:

$$\frac{\xi_0(s)}{\xi(r)} = \left(1 + 2\frac{\beta}{3} + \frac{\beta^2}{5} \right) \quad (\text{VI.10})$$

we will use this prescription to find the β parameter with simulations. Unfortunately, when working with real data we do not have access to the real space correlation function. Commonly the way to estimate the β parameter with real data is using the normalized quadrupole defined as:

$$Q(s) = \frac{\xi_2(s)}{\xi_0(s) - \int_0^s \xi_0(s')s'^2 ds'} \quad (\text{VI.11})$$

in the linear approximation valid in the large scale regime, the quadrupole is related directly to the β parameter as follows:

$$Q(s) = \frac{\frac{4}{3}\beta + \frac{4}{7}\beta^2}{1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2} \quad (\text{VI.12})$$

In (Cabr e & Gazta naga, 2009) has been shown that the ξ_0 and ξ_2 depends strongly on the shape of the correlation function (i.e . Ω_m, Ω_b, n_s , the non linear bias and the amplitude), while the normalized quadrupole depends only on β and σ_v .

VI.2.1 Study Redshift Distortions with lognormal Random Fields

The first goal I investigated is how well we can reconstruct the linear amplification redshift parameter β with mock catalogs issues from lognormal simulations. In the chapter 4, I explained the generation of galaxy mock catalogs from lognormal simulations, for this section an extra ingredient is considered: the velocities. In the linear regime approximation, the velocity field is related with the matter density field by :

$$v = -\frac{\partial}{\partial r} \nabla^{-2} \theta \quad (\text{VI.13})$$

where ∇^{-2} is the inverse laplacian operator and $\theta = \nabla \cdot \vec{v}$. In Fourier space this expression simplifies as the inverse laplacian operator becomes $\partial/\partial r \nabla^{-2} = (k \cdot r/k)^2 = \mu^2$.

Using this expression for the velocities in the linear regime we assign the velocities to the mock galaxies generated using the lognormal density field. The simulations were done considered uniquely a bias $b = 1$.

The mocks were generated in a cubic geometry, I generate a set of 200 simulations with 30,000 galaxies each realization, and a random with the same number of objects. No completeness effects were introduced. In figure VI.4 a slide of the galaxy density issued from a log-normal simulation of the density field in real (left) and redshift (right) space is shown in the linear regime. We notice that the global effect of redshift distortions is to make the voids appear emptier and bigger and a squashing effect along the line of sight. As we observe from the figure, the effects of redshift

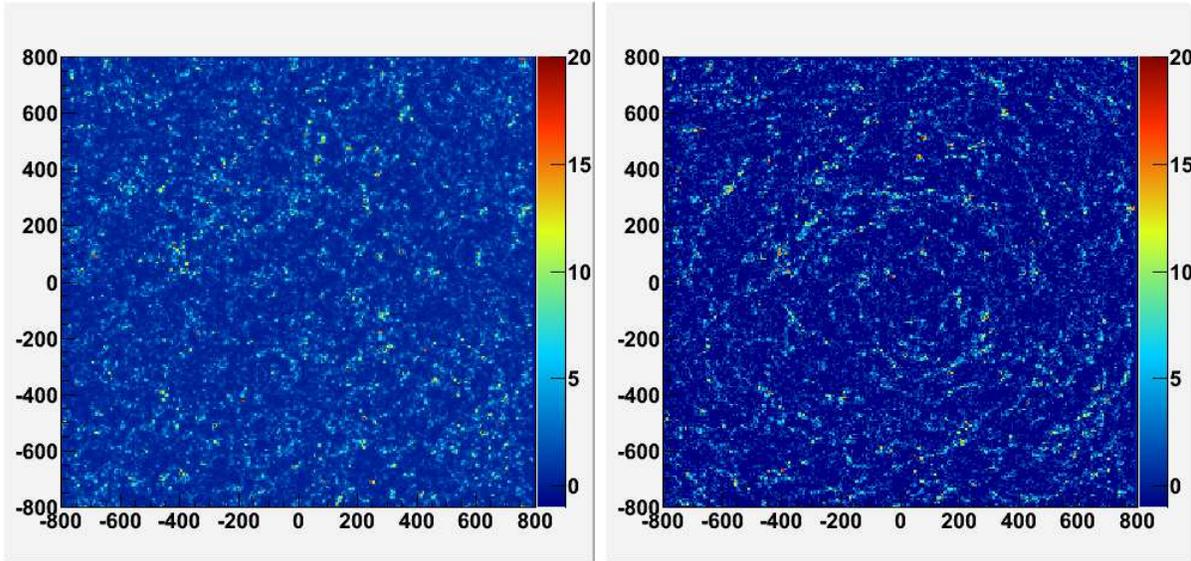


Figure VI.4: Slide of lognormal simulation in real and redshift space

distortions in real data are subtle, are not easy to observe by eye, but we will see that statistically the effect is notorious in the 2 point statistics VI.6.

From each mock catalog we compute the 2 point correlation function using the Landy-Szalay estimator. In practice there exist two equivalent ways of representing the 2D correlation function. The first one is separating the distances into 2 components, the radial direction $r_{||}$ (usually denoted π) along the line-of-sight (LOS) and the perpendicular direction r_{\perp} (usually denoted σ) (VI.2). An equivalent representation of the correlation function in 2 dimensions is $\xi(r, \theta)$ where θ is the angle between the line of sight \vec{s} and \vec{r} . In figure VI.5 and VI.6 the real and redshift space correlation function are shown. As we can observe from the plot the distortions due to the redshift distortions in the space $r - \theta$ generates deviations from vertical lines .

For the moment we are only considering the large scale effect, i. e. the linear regime. From the 2D correlation function we estimate the spherical harmonics fitting for each bin in r_i a quadratic function for the dependency in $\xi(r_i, \mu^2 = \cos^2\theta)$ for each simulation. In figures VI.5 and VI.6 I show the results in real and redshift space respectively. In the left panel the 2d correlation function. In the right panel I show the results for the harmonics spheric rescaled by the distance s^2 ($\xi_0(s)s^2$, $\xi_2(s)s^2$ and $\xi_4(s)s^2$) estimated from the average of the 200 Monte Carlo realizations of a lognormal simulation. In figure VI.5 the red points corresponds to the estimation in real espace taking the galaxies without applying the velocities. In the figure VI.6 the dots corresponds to the redshift space, i.e. applying the peculiar velocities. The blue dotted line is the linear correlation function without Redshift distortions and the green dotted line the linear correlation function with redshift distortions. As we observe from the fits the fitted $\xi_0(s)s^2$, $\xi_2(s)s^2$ and $\xi_4(s)s^2$ are in agreement with the linear prediction in both cases up to 50 Mpc. In the case of $\xi_4(s)s^2$ we found is very noisy.

Now we will test if we recover the β parameter set in the lognormal simulation. In the case of lognormal simulations we have access to the correlation function in real space thus we can use

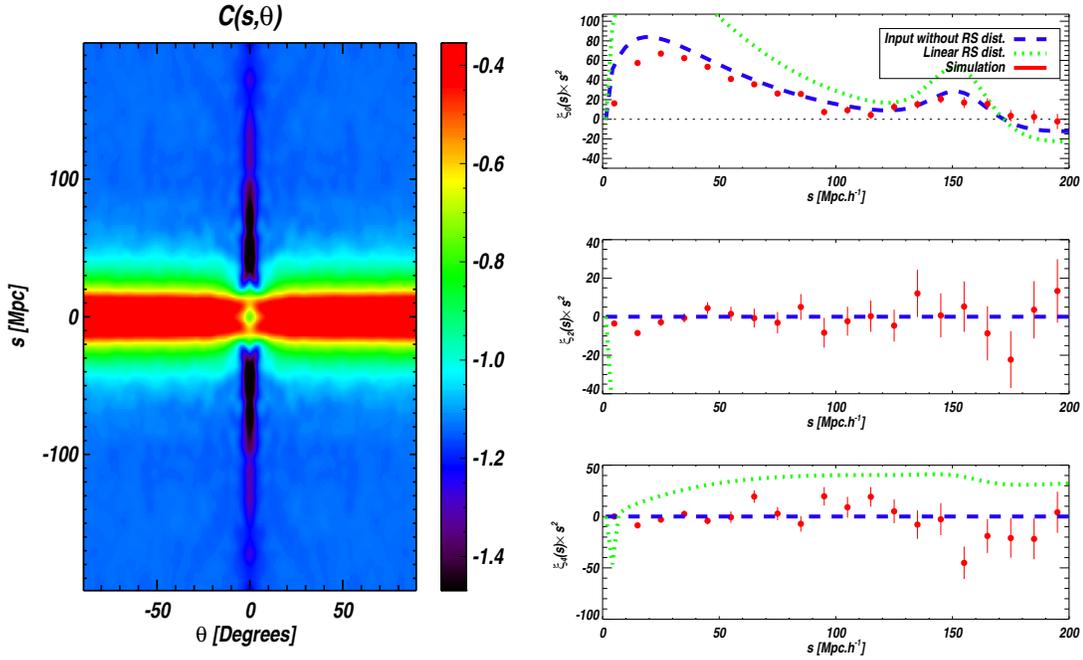


Figure VI.5: Left Panel $\xi(\theta, s)$ in real space (without redshift distortions), Top Right panels corresponds to $\xi_0(s)s^2$, $\xi_2(s)s^2$ and $\xi_4(s)s^2$ estimate from 200 Monte Carlo realizations of a lognormal simulation in real space (red points), the blue dotted line is the linear correlation function without Redshift distortions and the green dotted line the linear correlation function with redshift distortions

expression VI.7 to estimate the β parameter. In figure VI.7 the red points corresponds to the estimation of β parameter from the rate $\xi_s(r)/\xi_0(r)$. The black dotted line correspond to the input β parameter from the simulation. We observe than the reconstruction of the beta parameter is possible.

VI.2.2 Study non linear behavior effects HORIZON Simulation

The HORIZON – 4 π Simulation (Prunet et al., 2008) is an N-body simulation using 4096^3 dark matter particles in a periodic box of side 2 Gpc/h. This simulation uses the initial conditions given by the parameters of WMAP 3 years [$\Omega_m = 0.24, \Omega_\Lambda = 0.76, H_0 = 73.0, \Omega_b = 0.042, n_s = 0.96, \sigma_8 = 0.92$]. The particles were evolved using particle mesh scheme of the RAMSES code on a adaptively refined grid (AMR) with about 140 billions cells permitting a resolution of about $7kpc h^{-1}$ in comoving coordinates. The power spectrum from simulation is shown in figure VI.8. Using one realization of the N-body simulation HORIZON – 4 π I generated a set of 200 realizations using the Bootstrap method. The bootstrap method is a resampling method consisting in drawing randomly from original data set, N points with the possibility of re-emplacement. In figure ?? a slide showing the fof particles.

Up to now we have considered only the linear regime, now we will explore what happens when we

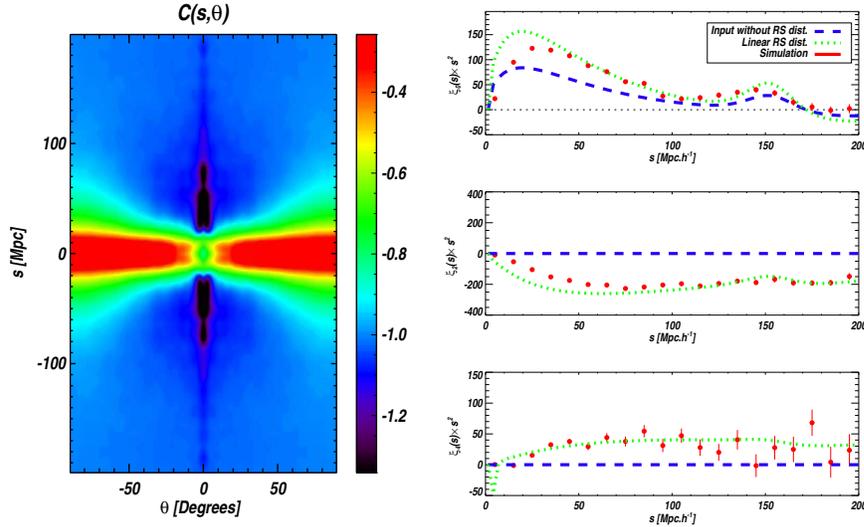


Figure VI.6: Left Panel $\xi(\theta, r)$ with redshift distortions from lognormal simulations, Top Right panels corresponds to $\xi_0 r^2$, $\xi_2 r^2$ and $\xi_4 r^2$ estimate from 200 Monte Carlo realizations of a lognormal simulation in redshift space (red points), the blue dotted line is the linear correlation function without Redshift distortions and the green dotted line the linear correlation function with redshift distortions

introduce the non linearities. From the 200 hundred realizations I computed the correlation function as before in the plane $r - \theta$. In figure VI.9 in left Panel the average $\xi(r, \theta)$ of the simulations in real space (without redshift distortions) from **HORIZON** – 4π simulation. The right panels corresponds from top to bottom to the average harmonics rescaled by s^2 ($\xi_0(s)s^2$, $\xi_2(s)s^2$ and $\xi_4(s)s^2$) estimated from 200 Monte Carlo realizations of halos from **HORIZON** – 4π simulation in real space (red points). In figure VI.10 the analog plot with redshift distortions from **HORIZON** – 4π simulation. In this case as we do not have the input correlation function, we can not plot the estimation of the linear theory.

As an illustration I present the analog representation of the correlation function in 2D in the plane $r_{\parallel} - r_{\perp}$ where r_{\parallel} is the component of the distance parallel to the line-of-sight (LOS) and r_{\perp} is the component perpendicular to the line of sight. In figure VI.11 in left panel the average $\xi(r_{\parallel}, \pi)$ in real space (without redshift distortions) from **HORIZON** – 4π simulation, in the top right panels corresponds to the average $\xi_0(s)s^2$ and $\xi_2(s)s^2$ estimated from 200 Monte Carlo realizations of halos from **HORIZON** – 4π simulation in real space (red points). In figure VI.12 the analog plot with redshift distortions from **HORIZON** – 4π simulation. In the case of the of the plane $r_{\parallel} - r_{\perp}$ the redshift distortions effect is modify the circular pattern of the correlation function adding an squashing effect in the radial direction.

Finally, in figure VI.13 we show the estimation of β parameter using the $\xi_s(r)/\xi_0(r)$ with $\xi(s)$ redshift distortions from **HORIZON** – 4π simulation, where $\xi_0(r)$ is the correlation function in real space and $\xi_s(r)$ in redshift space. The bias of the simulation is $b = 1$ by definition as we are working with halos of dark matter. The black dotted line corresponds to the input parameter from the simulation. We observe than the reconstruction of the beta parameter is possible.

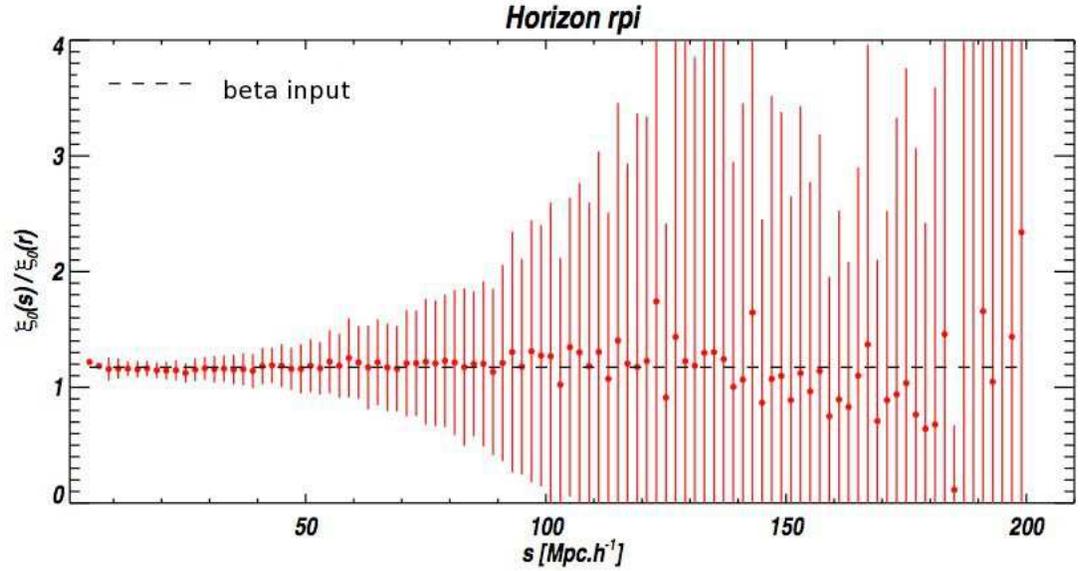


Figure VI.7: Estimation of β parameter using the $\xi_s(r)/\xi_0(r)$ with redshift distortions from lognormal simulations. The input corresponds to the dotted line.

VI.3 Measurement of Linear Redshift Distortion Parameter with BOSS

The final goal of working with simulations is prepared the way to applied to real data from BOSS, in this section I show some preliminar results.

Continuing in the same line of the exercise done with simulations we would estimate the redshift distortion parameter β from the 2D correlation function. We have tested with simulations the extraction of the beta parameter from the redshift-to-real-space monopole ratio . However with real data we do not have acces to the real space correlation function, we can estimated from the projected correlation function but the result would be subject to error. In this case we use the normalized quadrupole-to-monopole ratio which we have seen is related also to β in linear theory. I will use the current DR9 catalog.

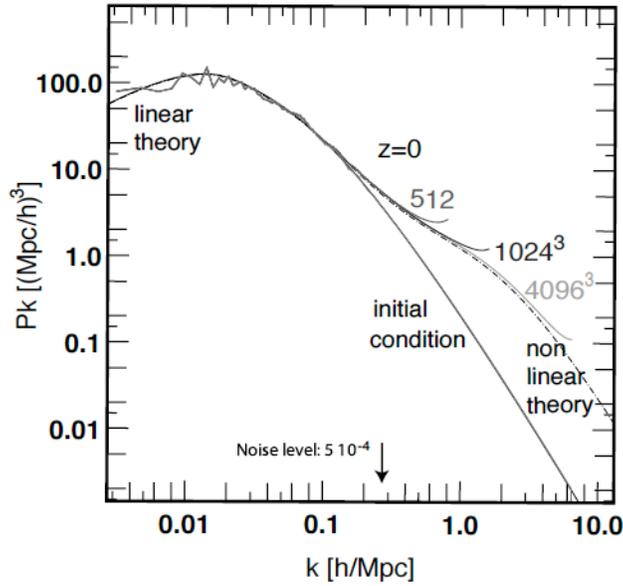


Figure VI.8: The power spectrum from the simulation Horizon-4 Π measured with `powergrid` (Prunet et al., 2008). The power spectrum is compared with the linear power spectrum and the non linear prediction using the Smith et al 2003. (Smith et al., 2003). Figure taken from (Prunet et al., 2008)

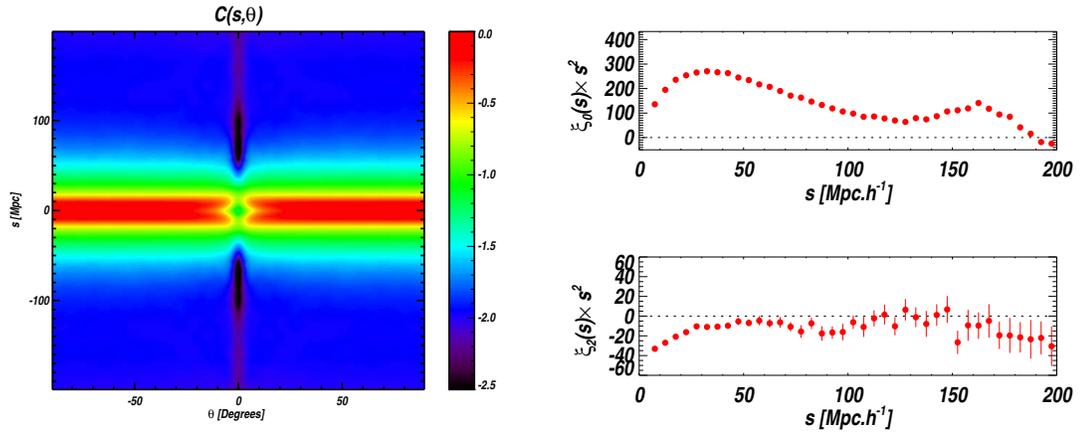


Figure VI.9: Left Panel $\xi(r, \theta)$ in real space (without redshift distortions) from HORIZON-4 Π simulation, Top Right panels corresponds to $\xi_0(s)s^2$, and $\xi_2(s)s^2$ estimate from 200 Monte Carlo realizations of halos from HORIZON-4 Π simulation in real space (red points)

In figure VI.14 in left panel I show the 2 point correlation function as a function of $r_{||}$ and r_{\perp} measured from BOSS data (scaled by r^2). We can observe the circular ring correspondent to the BAO feature at a scale of 100 Mpc/h. In right panel a zoom of the correlation function in the region

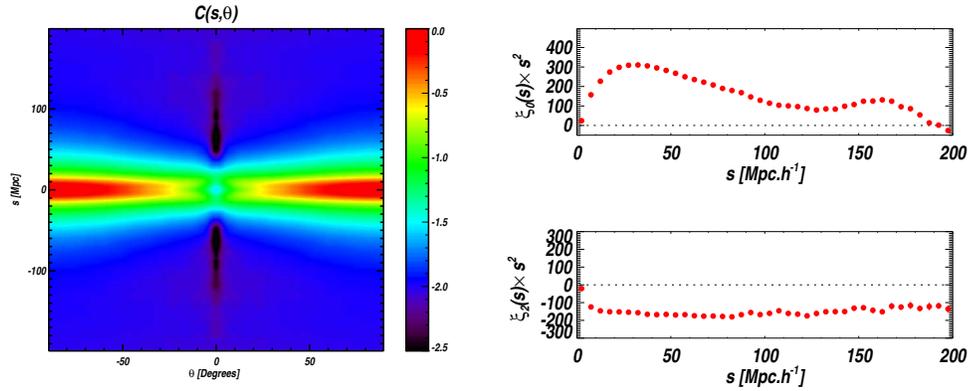


Figure VI.10: Left Panel $\xi(r_r, \theta)$ with redshift distortions from HORIZON – 4 Π simulation, Top Right panels corresponds to $\xi_0(s)s^2$ and $\xi_2(s)s^2$ estimate from 200 Monte Carlo realizations of from HORIZON – 4 Π simulation in redshift space (red points)

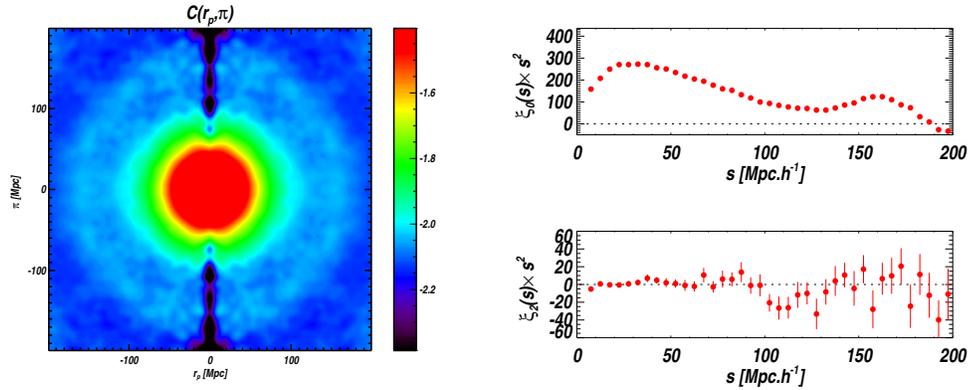


Figure VI.11: Left Panel $\xi(r_{\parallel}, \pi)$ in real space (without redshift distortions) from HORIZON – 4 Π simulation, Top Right panels corresponds to $\xi_0(s)s^2$ and $\xi_2(s)s^2$ estimate from 200 Monte Carlo realizations of halos from HORIZON – 4 Π simulation in real space (red points).

of 25Mpc/h to distinguish the fingers of god, as we can appreciate the effect in small scales is an elongation along the line of sight.

Following the same procedure as before I compute the spherical harmonics from the 2D correlation function. In figure VI.15 the black line shows the estimation from the results from data, and the error bars are the diagonal terms of the covariance matrix computed from PTHALOS simulations from *Ross in prep*. More details about this covariance matrix could be found in the BOSS chapter. The lines corresponds to a fit done using the correlation function model explained in the same chapter.

Finally I figure VI.16 I show in top panel the beta parameter estimated from BOSS data using expression VI.12. In order to analyze the result I show in the bottom panel of figure VI.16 a result from (Cabr e & Gazta aga, 2009) where this procedure to estimate the beta parameter have been

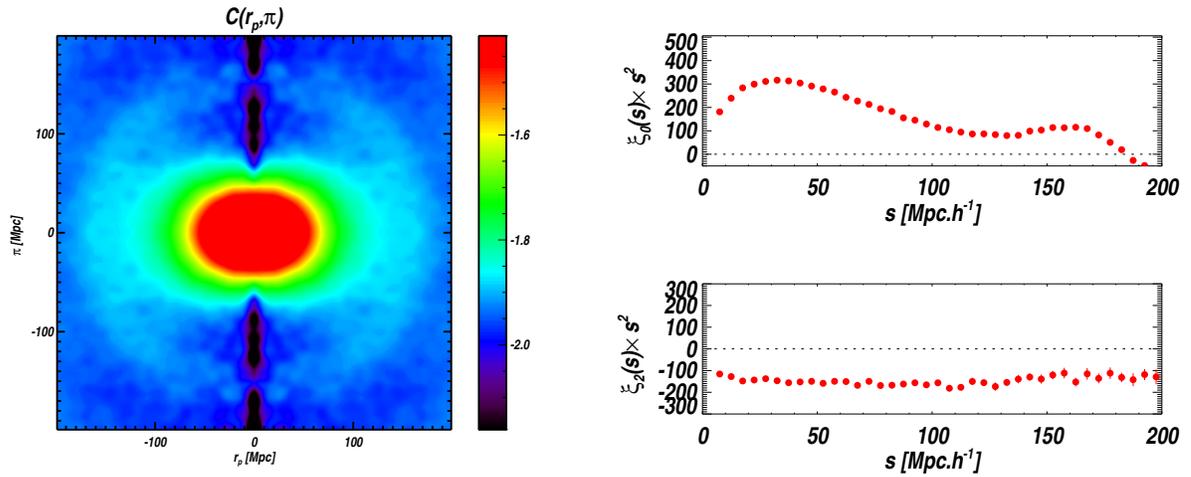


Figure VI.12: Left Panel $\xi(r_{\parallel}, \pi)$ with redshift distortions from HORIZON – 4 Π simulation, Top Right panels corresponds to $\xi_0(s)s^2$ and $\xi_2(s)s^2$ estimate from 200 Monte Carlo realizations of from HORIZON – 4 Π simulation in redshift espace (red points)

used with SDSS-DR6 sample. We can measure *beta* using the asymptotic large scale value of the quadrupole. In the case of (Cabr e & Gazta naga, 2009) they found $\beta = 0.34 \pm 0.06$ in the range 40-80Mpc/h.

To finish this chapter I would like to say that this is a very preliminary work., In a future work I would be interested to continue/conclude this work introducing to this simplified picture, the non linear behavior and wide angle effects as well as the fingers of god modeling. In Samushia 2011 (Samushia et al., 2011) these effects have been considered in the context of DR7 probing to be important at a level of 20 per cent.

In a second time I would be interested in the fitting of the 2D correlation function, that would connect this chapter with chapter 7 of cosmological constrains.

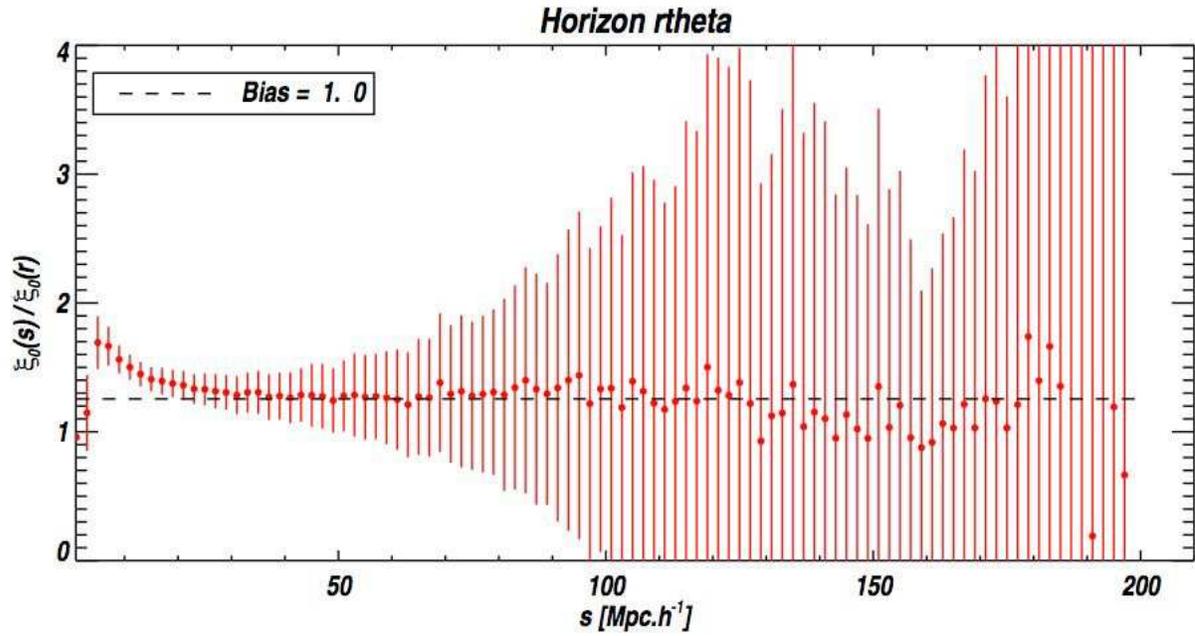


Figure VI.13: Estimation of β parameter using the $\xi_s(r)/\xi_0(r)$ with redshift distortions from HORIZON – 4 Π simulations, where $\xi_0(r)$ is the input correlation function generating the HORIZON – 4 Π simulation. The input in $\beta = 1$ parameter.

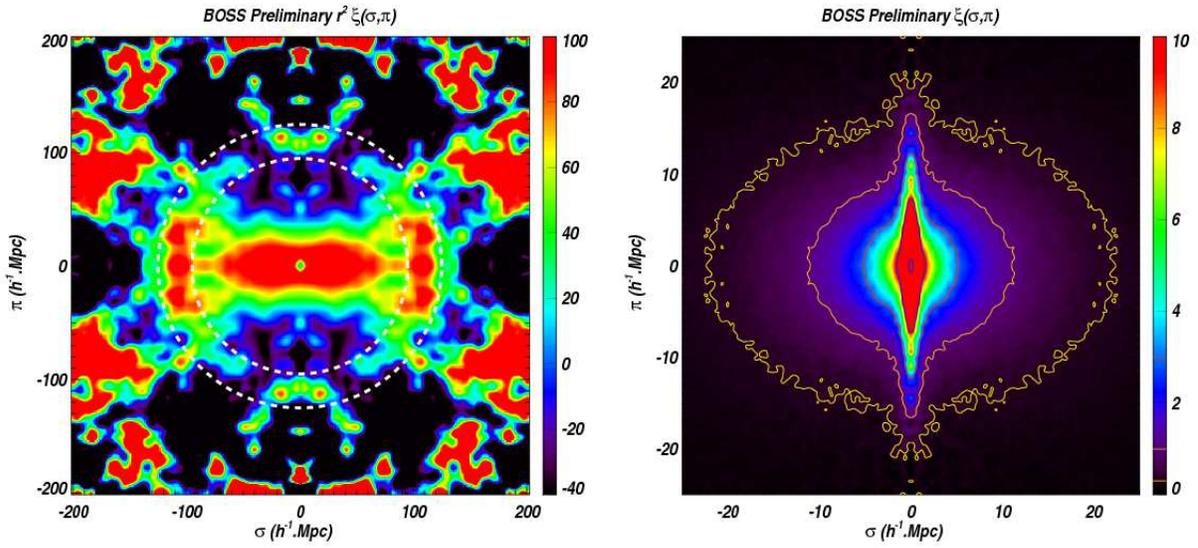


Figure VI.14: $\xi(r_{||}, r_{\perp})$ for BOSS CMASS galaxies in redshift space

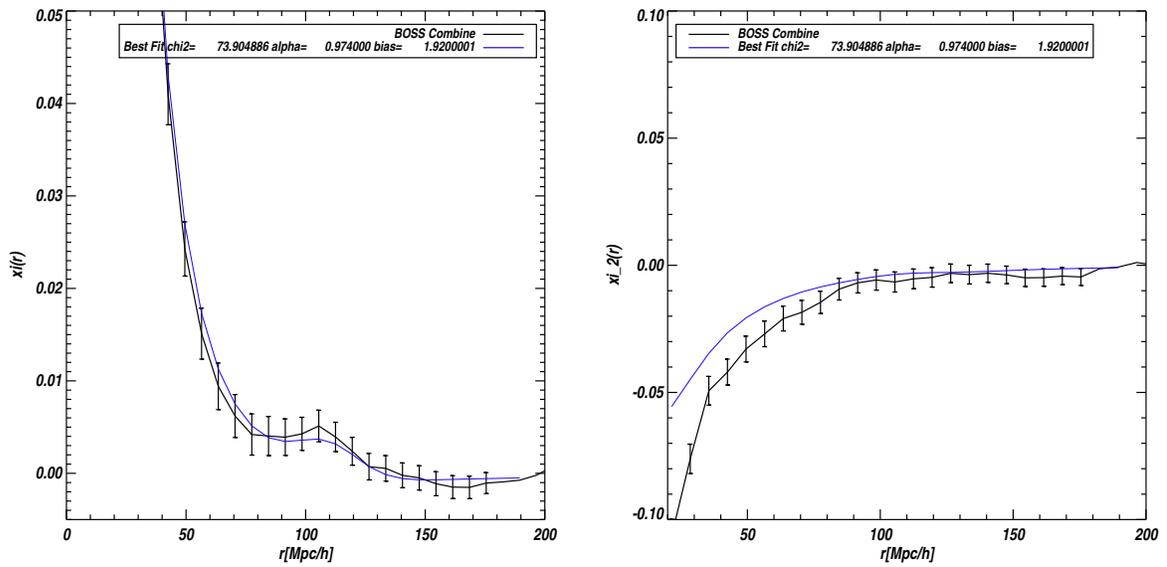


Figure VI.15: $\xi_2(r)$ for BOSS CMASS galaxies in black, best fit model in blue

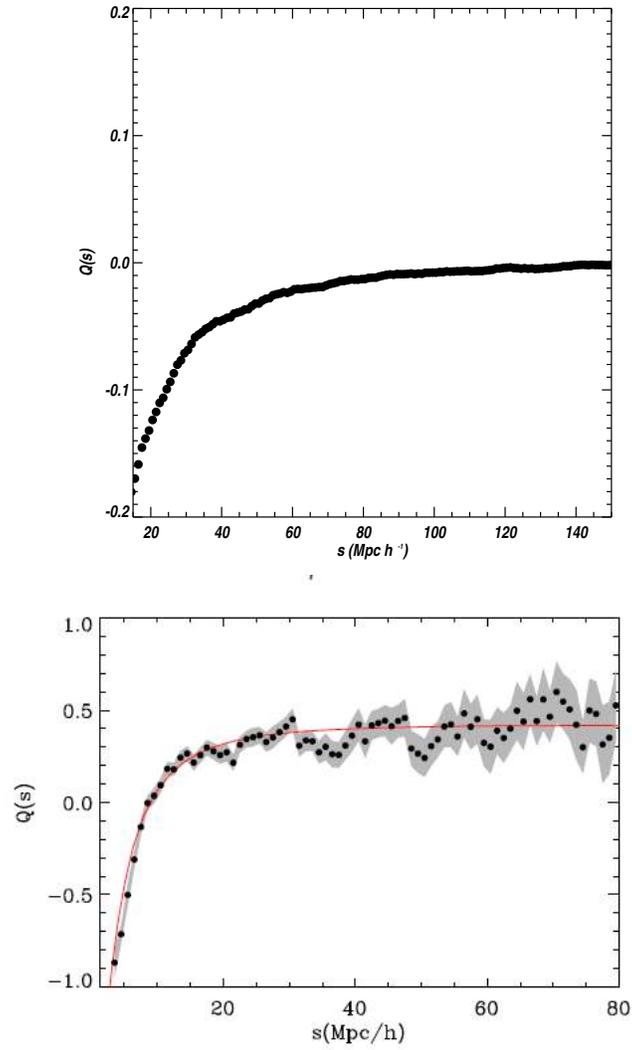


Figure VI.16: $Q(s)$ for BOSS CMASS galaxies in black (PRELIMINARY PLOT). In left panel the result for the estimation of $Q(s)$ for SDSS-DR6 from (Cabr e & Gazta naga, 2009). Measured $Q(s)$ in dots with error in gray and in red the best fit

VII. Constraining Cosmology with BOSS

This chapter is devoted to the cosmological constraints we can get from the analysis of BOSS data. In the first section, I describe briefly the fitting procedure even if it is a standard procedure. In the second section, I describe the model used to fit the correlation function. This work has been done with the collaboration of A. Labatie who has developed a code which implements the model correlation function. In the third section I present and discuss the results obtained with BOSS data and finally I give some conclusions and perspectives.

VII.1 Fitting procedure and Confidence Limits

The fitting procedure¹ consists in obtaining the best set of parameters, characterizing a given correlation function model, corresponding to the data. Thus, we have a vector of the observed correlations in various bins ξ_o . We have a model of $\xi(r)$ with various free parameters $\vec{\theta}$ to be fitted. We construct a vector $\xi_{\vec{\theta}}$ of the expected results for the measurement given the model. We construct the covariance matrix $C = C_{ab}(\xi_a, \xi_b)$ that encodes the errors of the estimator in each bin as well as the correlated errors between bins. Adjusting the parameters to maximize the agreement one obtains the best fit parameters. A fitting procedure provides us the best fit parameters, the error estimates on the parameters, and possibly a statistical measure of the goodness of fit. There are two different approaches in order to obtain these results. One is frequentist and the other is bayesian.

VII.1.1 χ^2 Method

The frequentist approach is based on χ^2 (least squares) method for which a goodness of fit can be established. The χ^2 statistics is written as:

$$\chi^2(\vec{\theta}) = (\xi_o - \xi_{\vec{\theta}})^T C^{-1} (\xi_o - \xi_{\vec{\theta}}). \quad (\text{VII.1})$$

¹For this section a follow closely references (Wall & Jenkins, 2003), (Bevington & Robinson, 1992), (Verde, 2010), (Press et al., 1992), (Andrae, 2010), (Labatie et al., 2011),

The best set of parameters is found by minimizing the χ^2 value. The χ^2 method is a classical method for BAO detection and parameter constraints. The χ^2 method is fully model dependent. It requires a full modeling of the correlation function including the systematics effects. Within the definition is also included the noise measurement, i. e the covariance measurement of the estimator $\hat{\xi}$. The classical approach of χ^2 simplifies the problem assumes the covariance is model independent, that means a constant covariance matrix. The χ^2 is designed for hypothesis where $\hat{\xi}$ is gaussian. If we suppose the model is correct we suppose that the estimator $\xi \sim N(\xi_{mod}, C)$, in this case the χ^2 statistics follows a chi square distribution χ_n^2 with n degrees of freedom. The last statement is based over the fact that the sum of n independent standard normal variables of $\sum_{i=0}^n X_i^2$ follows a χ^2 distribution.

Goodness of the fit

If the measurements errors are gaussianly distributed and the model is a linear function of the parameters, the probability distribution of the χ^2 at the minimum follows Chi-square distribution of $\eta = n - m$ degrees of freedom where m is the number of parameters and n is the number of data points. The probability that the observed χ^2 even for a correct model is less than a value $\hat{\chi}^2$ is :

$$P(\chi^2 < \hat{\chi}^2, \nu) = \Gamma(\nu/2, \hat{\chi}^2/2), \quad (\text{VII.2})$$

where Γ is the incomplete Gamma function. The complement gives us the probability that the observed χ^2 exceed by chance $\hat{\chi}^2$ even for a correct model. The Q quantity provides a quantitative measure of the goodness of the fit when evaluated at the best parameters (i.e. χ_{mu}^2)

$$Q = 1 - P(\hat{\chi}^2/2, \nu/2) \quad (\text{VII.3})$$

Chi-Square Boundaries as Confidence levels

Confidence regions are the n dimensional parameter space that contains a certain percentage of the total probability distribution. It is customary to choose 1σ (68.3%), 2σ (95.4%) and 3σ (99.7%). In general, if μ is the number of fitted parameters for which we want to plot the join confidence region and p is the confidence limit desired, thus you find the $\Delta\chi^2$ such that the probability of a chi-square variable with μ degrees of freedom being less than p . A table extracted from numerical recipes is shown in table???. For other values is necessary to look for the $\Delta\chi^2$ such that

$$\Gamma(\nu/2, \Delta\chi^2) = 1 - p. \quad (\text{VII.4})$$

VII.1.2 Likelihood

While for frequentists events are just frequencies of occurrence, bayesians interpret probabilities as the degree of belief in an hypothesis. Bayesians consider hypothesis as events. We want to know the probability of the hypothesis given the data:

$$P(H|D) = \frac{P(H)P(D|H)}{P(D)}, \quad (\text{VII.5})$$

where $P(H|D)$ is the probability of the model given the data is called the posterior, $P(D|H)$ the probability of the data given the model is the likelihood, that is the probability of the data given the hypothesis, and $P(H)$ is the prior. The hypothesis H could be a model, in our case the Λ CDM model, which is characterized by a set of parameters. We are interested in the probability distribution for this model characterized by a set of parameters $\vec{\theta}$ given the data $P(\theta|D)$. From this distribution we can extract the most likely value for the parameters and their confidence limits. Thus we will compute the Likelihood $P(D|H)$ and $P(D) = 1$, if the data is already taken, and assuming a particular form for the prior $P(H)$ we have the posterior probability. The prior may come from a previous measurement or from your belief. The choice for the prior affects the final result thus particular attention must be paid. Usually flat priors are used in such a case the bayesian and frequentist approaches gives the same result.

If we ignore the prior, in maximizing the likelihood we will find the most likely hypothesis (parameters of a given model). The general approach relies on the approximation of the likelihood \mathcal{L} by a multivariate Gaussian. This approximation is justified by the central limit theorem. The Central Limit theorem states that the sum of many independent random variables will be approximately gaussianly distributed:

$$\mathcal{L} = \frac{1}{(2\pi)^{n/2} |\det C|^{1/2}} \exp \left[-\frac{1}{2} \chi(\theta)^2 \right]. \quad (\text{VII.6})$$

For bayesians, confidence regions are found as regions R in model space such that

$$\int P(\theta|D) d\theta = p, \quad (\text{VII.7})$$

where $p = 0.68, 0.95, \text{etc.}$ This implies the prior information, to report results independently of the prior the ratio of likelihood of the points in the parameter space with the \mathcal{L}_{max} is used:

$$-2Ln \left[\frac{L(\theta)}{L_{max}} \right] \leq \text{threshold}. \quad (\text{VII.8})$$

This threshold is calibrated by calculating the distribution of the likelihood ratio in the case where a particular model is the true model.

VII.1.3 BAO detection and Significance

In BAO analysis usually before of doing the parameter confidence levels a preliminary test is done to establish the detection of the baryonic feature. The BAO detection is done by hypothesis testing, in general we have:

$$\begin{aligned} H_0 &: \text{no BAO hypothesis} \\ H_1 &: \text{BAO hypothesis} \end{aligned}$$

The common procedure is to design a test statistic to assess the truth to the null hypothesis H_0 (Labatie et al., 2011), compute the statistics with the measurement, and calculate its p value and a significance. If the hypothesis is found to be less likely than a given threshold under H_0 , H_0 is rejected and H_1 is accepted. The data measurement is the correlation function. The BAO models ξ_{BAO} and no-BAO models ξ_{NO_BAO} are parametrized by a set of parameters.

The χ^2 method for BAO detection consists in computing the χ^2 statistic for the two class of models corresponding to the two hypothesis.

$$\begin{aligned}\chi_{BAO,\theta}^2 &= (\xi_o^- \xi_{BAO,\theta}) C^{-1}(\xi_o - \xi_{BAO,\theta}) \\ \chi_{noBAO,\theta}^2 &= (\xi_o - \xi_{noBAO,\theta}) C^{-1}(\xi_o - \xi_{noBAO,\theta})\end{aligned}\quad (\text{VII.9})$$

Then we will compare the best fit for both classes of models given by the minimum value of χ^2 . Thus, the minimum $\chi_{BAO,\theta}^2$ value could be used to the compatibility of the data with the hypothesis H_1 . The best fit value follows a χ_{n-k}^2 where k is the number of parameters of the fit. For the rejection of the null hypothesis H_0 it is necessary to add an artificial parameter η in the fit. This parameter is introduced as a weight factor of ξ_{BAO} and ξ_{noBAO} :

$$\xi_{\eta,\theta} = \eta \xi_{BAO,\theta} + (1 - \eta) \xi_{noBAO,\theta}. \quad (\text{VII.10})$$

In this case, the χ^2 value follows a distribution with the numbers of freedom equal to the difference of parameters number between the 2 classes. Because there is a difference of one parameter between the two classes, the best fit values follows a chi_1^2 distribution:

$$\Delta\chi^2 = \min\chi_{noBAO,\theta}^2 - \min\chi_{\beta,\theta}^2 \sim \chi_1^2. \quad (\text{VII.11})$$

In practice $\min\chi_{\beta,\theta}^2$ is replaced by a $\min\chi_{BAO,\theta}^2$, giving:

$$\Delta\chi^2 = \min\chi_{noBAO,\theta}^2 - \min\chi_{BAO,\theta}^2 \sim \chi_1^2. \quad (\text{VII.12})$$

VII.2 Modeling the correlation function

To model the correlation function on large scales we follow the procedure of (Eisenstein et al., 2005)(Blake et al., 2011). This model considers the plane parallel approximation, valid for small angles of separation. In this classical approach we define a fiducial cosmology and we define a distortion parameter α introduced to account of deviations of fiducial cosmology. Concerning the non linearities this model corrects the linear prediction by including the global non linear correction to the full shape correlation function and a damping of the oscillations generated by the bulk velocities. There exists another approaches based on RPT (Crocco & Scoccimarro, 2008) implemented in the literature (Sánchez et al., 2009)(Beutler et al., 2011) that for further works it would interesting to compare with.

VII.2.1 Dilatation Parameter

Essentially, the correlation function implies the computing of distances and thus, suppose the choice of certain cosmology to traduce the redshift in distance measurements. The classical approach for BAO modeling is considering a fiducial cosmology to compute the correlation function. A dilatation parameter α is introduced to take into account the possibility of a wrong cosmological model:

$$\xi(r) = b^2 \times \xi_{fid}(\alpha r), \quad (\text{VII.13})$$

where b is the bias.

The probability distribution of the scale of distortion parameter α , after marginalizing over $\Omega_m h^2$ and b^2 , give us the probability distribution of the distance scale $D_V(z_{effect}) = \alpha D_{V, fid}(z_{effect})$ variable (Eisenstein et al., 2005) explained in more detail in a further subsection. To get the correlation function from power spectrum we perform the Hankel transform:

$$\xi(r) = \frac{1}{2\pi} \int dk k^2 P_{fid}(k) \sin(kr)/kr. \quad (\text{VII.14})$$

The fiducial cosmology is one adopted in the (White et al., 2011) corresponding to a flat Λ CDM model, $h = 0.71, \Omega_h^2 = 0.0226, n_s = 0.96, \sigma_8 = 0.8, \Omega_b h^2 = 0.0224$. To construct $P_{fid}(k)$, we are going to consider the non linear effects following the procedure of (Eisenstein et al., 2005) and (Blake et al., 2011).

VII.2.2 Non linear effects

The precision we have achieved is such that we cannot rely entirely on linear theory even at $r > 10h^{-1}$. There are two effects. We have seen in the BAO chapter that the non linearities erase the higher harmonics generated by the displacement of the matter due to the bulk flows. A degradation of the peak in the correlation function appears as a damping of the oscillations in the power spectrum. The second effect is related to the scale dependence of the growth of structure, it accounts for the alterations in the power from non linear gravitational collapse.

In order to construct the non linear power spectrum we start from a $P_{lin}(k)$. We generate a non wiggle power spectrum P_{nw} . To approximate the suppression of the higher harmonics, a gaussian $g(k)$ with a non linear radius σ_v is convolved with the power spectrum :

$$P_{damp}(k) = g(k)P_{lin}(k) + [1 - g(k)]P_{nw}(k) \quad (\text{VII.15})$$

where

$$g(k) = \exp[-k^2 \sigma_{NL}^2] \quad (\text{VII.16})$$

In fact, it is possible to derivate an expression of the σ_{NL} using perturbations theory giving the following expression (Crocco & Scoccimarro, 2008):

$$\sigma_{NL} = \frac{1}{6\pi^2} \int P_{lin}(k) dk. \quad (\text{VII.17})$$

In general, instead of using the last expression, the non linear radius is typically fitted to simulations.

We use ICOSMO (Refregier et al., 2011) to compute the linear and the non wiggle power spectrum. There are different ways to get these power spectrum giving similar results as using the fitting formulas of E&H (Eisenstein & Hu, 1998b) that is equivalent to using CMBFast (Seljak & Zaldarriaga, 1996); (Zaldarriaga, 1998)).

To take into account the non linear enhancement from the clustering, we apply the corrections derived from N-body simulations and from halo fit model (Smith et al., 2003). These corrections are calibrated with simulations considering pure CDM models.

Finally, the first correction affects the wiggles while the second affect the global shape of the power spectrum so, the non linearities in the BAO must be included separately:

$$P_{damped,NL}(k) = P_{NL}(k)/P_{nw} \times P_{damp}. \quad (\text{VII.18})$$

For the $P_{NL}(k)$ the common procedure is to use the non linear formula of (Smith et al., 2003) (provided by CAMB or ICOSMO (Refregier et al., 2011)). I used the code generated by Antoine Labatie that generates the grid of xi models. This code uses the ICOSMO software for generating the $P_L(k)$, $P_{NL}(k)$ and P_{nw} .

VII.2.3 Model of the correlation function the Official Version

For the collaboration paper, the model used to fit the average angle correlation function is documented in Xiaoying Xu in preparation. The model is slightly similar to the model that I used. The xi_m is composed by:

$$\xi(r) = B^2 \xi_m(\alpha r) + A(r), \quad (\text{VII.19})$$

where $A(r)$ count for the unknown systematics:

$$A(r) = a_1/r^2 + a_2/r + a_3 \quad (\text{VII.20})$$

and $\xi_m(r)$ is the Fourier transform of:

$$P_m(k) = [P_{lin}(k) - P_{nw}(k)] * \exp[-(\sigma_{nl}^2 k^2)/2] + P_{nw}(k), \quad (\text{VII.21})$$

where $P_{nw}(k)$ is the de-wiggled power spectrum from (Eisenstein & Hu, 1998b). The considerate range of fitting is $\sim 30 - 200 \text{Mpc/h}$ in r with $\sigma_{NL} = 8 \text{Mpc/h}$. The fitting was done over 5 free parameters (a_1, a_2, a_3, B, α), the interval of measure is $[40, 200]$ which corresponds to 45 bins, giving place to 40 degrees of freedom. The model is normalized to the data at $r = 50 \text{Mpc/h}$ before fitting ($N \sim 2.5$).

VII.2.4 Extracting the BAO signal

In the case of 2D correlation function, along the line of sight, the BAO directly constrains $H(z)$ at a redshift z_{effect} . For each redshift shell of thickness dz , the 2D correlation function constrains the angular diameter distance. In the case of the angular average correlation function (1D), a combination of tangential and radial components of the signal are mixed. Commonly, the quantity D_V introduced by (Eisenstein et al., 2005), called the dilatation parameter defined as follow, is used:

$$D_V(z) = \left[(1+z)^2 D_A^2(z) \frac{cz}{H(z)} \right]^{1/3}. \quad (\text{VII.22})$$

It consists in two parts, the physical angular diameter distance D_A and the radial distance $cz/H(z)$. It reflects the importance relative of the tangential and radial modes in the BAO measurement (Padmanabhan et al., 2008).

The dilatation parameter is related to $\Omega_m h^2$ that controls the standard ruler scale. The probability distribution of the scale of distortion parameter α , after marginalizing over $\Omega_m h^2$ and b^2 , give us the probability distribution of the $D_V(z_{effect})^2$ as follows:

$$D_V(z_{effect}) = \alpha D_{V, fid}(z_{effect}) \quad (\text{VII.24})$$

For the fiducial cosmology we have $D_{V, fid}(z = 0.5) = 2044.26$. This relation implies that the measured value of D_V resulting from the fitting process will be independent of the fiducial cosmology up to first order. A change in the fiducial cosmology generates a shift in the D_V which is compensated by α and so let the relation VII.3.4 unchanged. The probability distribution of alpha is calculated as:

$$P(\alpha_i) = \frac{\exp[-\chi^2(\alpha_i)/2]}{\int \exp[-\chi^2(\alpha_i)/2] d\alpha}. \quad (\text{VII.25})$$

VII.3 Parameter Constraints from BOSS

In order to constrain the cosmological parameters, we fit the model, presented in the last section, to the galaxy correlation function measured on the BOSS data. There are 3 free parameters $\Omega_M h^2$, b , α . We fit on the interval [40, 200] Mpc. As we have said before to perform the fit, the parameter σ_{NL} must be calibrated using the simulations. We used the average correlation function generated from the six hundred simulations PTHALOS and we fit on the interval [60, 150] correspondent to the peak of BAO. For this fit, we applied a normalization to the overall correlation function before performing the fitting. This normalization gives a factor of 2.25. In figure VII.1 is shown the mean of PTHALOS North simulations and the fitted function in the respective range. The resultant radius was $\sigma_{NL} = 6.25$.

As we have seen in the chapter of BOSS correlation function, there seems to be a tension between North and South. Thus, for the observed correlation function we perform the analysis for 4 samples: the North and South galactic caps, the combined sample and the South DR10 which consists of data available up to date for the South galactic cap. I will discuss in a subsection the discrepancy North/South in terms of constraints extracted from the two regions. The correlation function used for the fits are the official versions to be able to compare with the people in the working group. We have shown in the last chapter that our catalogs give similar results to the official version, thus the results of the analyses apply to our catalogs. The official ξ_o include stellar density weights as a function of the magnitude band i for one fiber i_{fiber} , and FKP weights using a normalization of 20,000. To generate the correlation function of the combined sample, the North and South samples are treated separately. The South catalog do not have color offset applied. The random set are generated following the redshift distribution of each region. The pair counts are computed separately (i.e DD , DR and RR) for each hemisphere. The combine correlation function is generated combined the pair counts (i.e. $DD_{comb} = DD_{north} + DD_{south}$, etc.)

²The effective redshift of the sample is:

$$z_{eff} = \sum_i^{N_b} \sum_j^{N_b} \frac{w_i w_j}{2N_b^2} (z_i + z_j) \quad (\text{VII.23})$$

N_b number of galaxies within a bin. For the case of CNMASS sample $z_{effect} = 0.55$

The covariance matrices were computed with the six hundred PTHALOS simulations for the North and South separately, as explained in the chapter of the BOSS Correlation function. For the combined sample, the covariance matrix is generated from the covariance computed from North C_N and South C_S as:

$$C_T^{-1} = C_N^{-1} + C_S^{-1}. \quad (\text{VII.26})$$

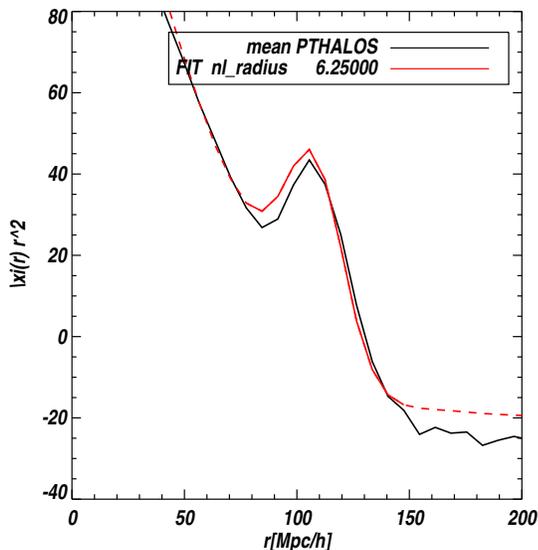


Figure VII.1: Calibrating the σ_{NL} with simulations PTHALOS. In black the mean of the 600 realizations of the PTHALOS simulations using only the north hemisphere. The $\sigma_{NL} = 6.25$.

We perform the fits using

the minimal model described in section 2 which consists of 3 free parameters, only the bias as nuisance parameter.

VII.3.1 Bin effect

Before testing the two models, we will perform a test of the binning. We perform the fits using two different sizes of bin, one with bin of 7Mpc equivalent to 25 bins for the range desired and one with bin of 4Mpc corresponding to 43 bins. For this test we fix the value of Ω_m to the fiducial value of $\Omega_m = 0.274$ (White et al., 2011). For performing the fit we construct a grid 200^2 bins of ξ_{mod} with the 2 free parameters α and b varying in a range $[0.8, 1.2]$, and $[1.5, 3]$. We compute the χ^2 for each model and we look for the minimum. I will compare the results between the 2 binning scheme in terms of χ^2 results. In figure VII.2, ?? and VII.3 I show the results of the fits using the minimal model for the bin of 7Mpc. In the left panels, I show in black the correlation function with the errors bars given by the diagonal terms of the covariance matrix and in red the best fit function as well as the parameters that minimizes the χ^2 . In the right panels, I show the contours for $1 - 5\sigma$ for the plane $\alpha - b$.

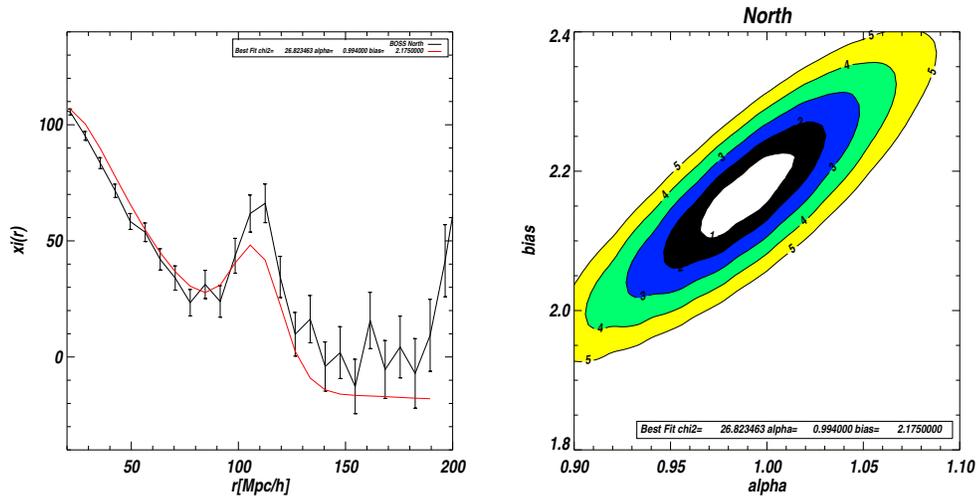


Figure VII.2: North: Fitting α , bias . WMAP as fiducial cosmology-bin 7Mpc.

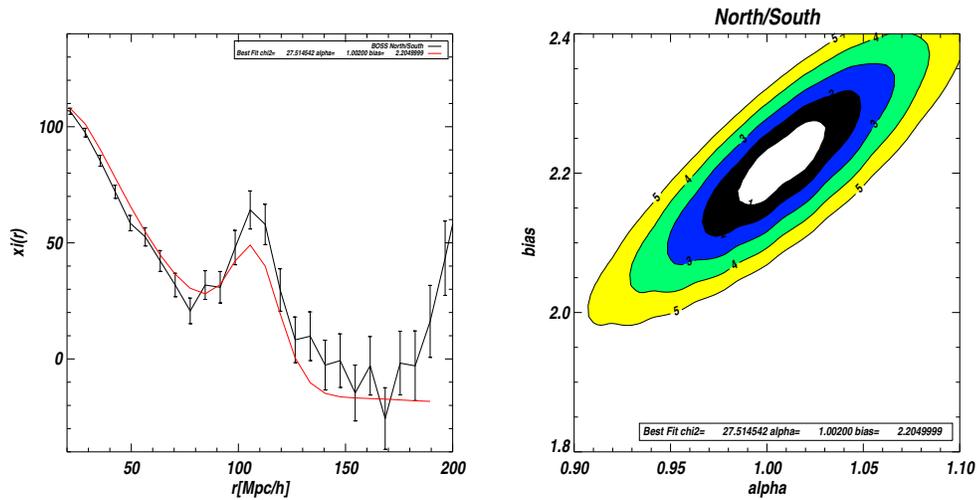


Figure VII.3: Combined: Fitting α , bias ,bin 7Mpc width.

In figure VII.4, ?? and VII.5 I show the results of the fits using the minimal model for the binning of 4Mpc. In the left panels, I show in black the correlation function with the error bars given by the diagonal terms of the covariance matrix and in red the best fit function, with the parameters that minimizes the χ^2 . In the right panels, I show the contours for 1 – 5 σ for the plane $\alpha - b$

In table VII.1, the constraints obtained from the different samples using the 2 binning schemes are shown. In the second column is given the χ^2 par degree of freedom and in the third column is given the value of the alpha parameter and 1 σ -errors obtained marginalizing over the bias. The

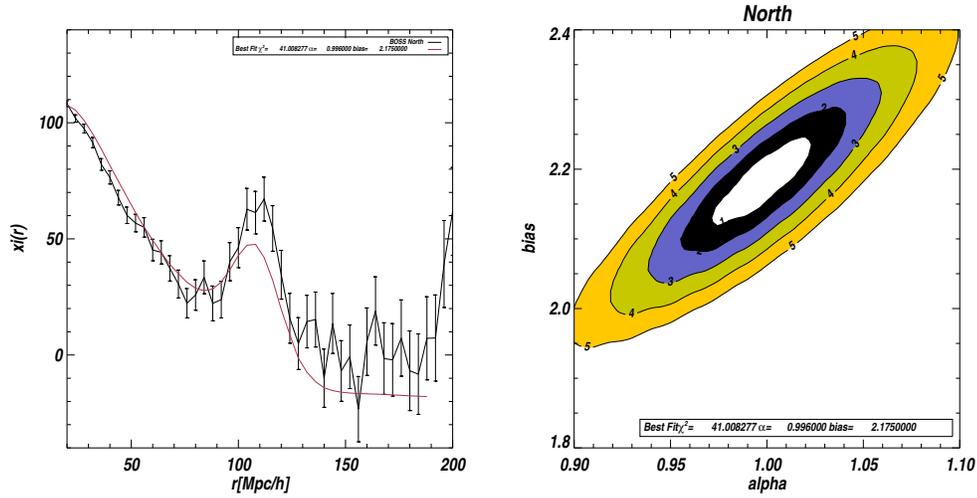


Figure VII.4: North: Fitting alpha, bias bin 7Mpc-width

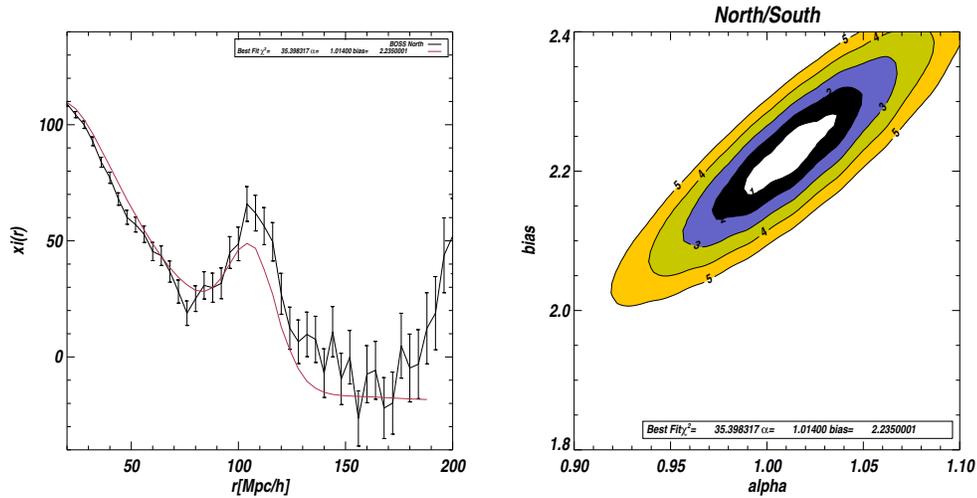


Figure VII.5: Combined: Fitting alpha, bias bin 4Mpc width.

fourth column is the value of the bias that minimize the χ^2 and finally the Q-value. As we can see from the figures there is a disagreement between South and North hemispheres of about 1σ , as a consequence the constraints generated from the combined sample deviate from the North constraints. In terms of χ^2/dof and Q-values they indicate acceptable fits to the data. The north gives better fits compared to South. There is a clear discrepancy in the constraints derived for North and South that generates the fitted parameters of the combined sample to be pushed to higher values of α and σ . Before discussing this tension between North/South hemispheres let us concentrate in North hemisphere (and the Combined sample) to analyze the effect of the binning in the constraints.

Table VII.1: Constrains for 4Mpc and 7Mpc bins North , South and Combined Samples

Sample	χ^2	α	bias	Q
North 4 Mpc	41.008/41=1.002	0.9960±0.016	2.1750	0.529
North 7 Mpc	26.823/25=1.073	0.9940±0.016	2.1750	0.635
Comb 4 Mpc	35.396/41=0.863	1.0140±0.015	2.2350	0.283
Comb 7Mpc	27.514/25=1.101	1.0000±0.018	2.2049	0.669
South 4 Mpc	35.4498/41=0.865	1.0640± 0.023	2.4150	0.283
South 7 Mpc	26.6456/25=1.066	1.0620± 0.030	2.4000	0.626

It seems the 7 Mpc bin gives better Q-values. This could be generated by more noisy data for the case of smaller bins width. In the case of the North the results with both bins with are almost the same, with a small difference in the alpha fitted parameter inferior to 0.2%. In the case of the combined sample the different binning schemes generates a change in the fitted parameters of $\sim 1\%$ in α and bias but still within $1\text{-}\sigma$ errors. In the case of South hemisphere the difference in the binning is less than 0.2% and 0.6% for α and the bias respectively. The official binning has been chosen to be 4 Mpc bins for the rest of the chapter.

VII.3.2 South and South(DR10)

In order to disentangle the disagreement between the North and South samples we will include in the analysis the data available up to now for the south even if for the conclusions we will try to restrict the analysis to the DR9 catalog. In figure VII.6, VII.7 , VII.8 and VII.9, the correlation function of South hemisphere for DR9 and DR10 samples with the best fit parameters is shown as well as the contours $1 - 5\sigma$ for the 2 binning scheme (7Mpc and 4 Mpc bin size).

Let us summarize the results before discussing the discrepancy North/South for the DR9. In figure VII.10 the contours a $1\sigma, 2\sigma$ for the North South, and the Combined sample are shown. As we can see the disagreement between North and South is within 2σ for the DR9. By contrast the contours considering the new data (DR10) seems to reduce the disagreement between measurements to a 1σ level. This fact seems to indicate that the correlation function observed at South hemisphere in DR9 presents an atypically strong correlation and when including more data this behavior average the weird early result. In the fits we can observe this tendency , when comparing the fits between south DR9 and south DR10. In the case of DR9 the strong correlation observed pushed the fit to higher values of alpha and bias, while the fits corespondents to South(DR10) seems to follow the same behavior of North DR9. The error bars are $\sim \sqrt{2}$ smaller for DR10 The estimated errors indicate that the fluctuation is significant?.

In table VII.2 the results of the fits are shown , the minimum χ^2 values and the Q-values indicates again acceptable fits to the data. The value for DR10 is strangely low compared with the others binning schemes and samples.

In the case of the South DR10 the results in the fitted parameters with both bin widths present larger differences than South DR9. The α parameter is in agreement within $\sim 1\%$ (compared to

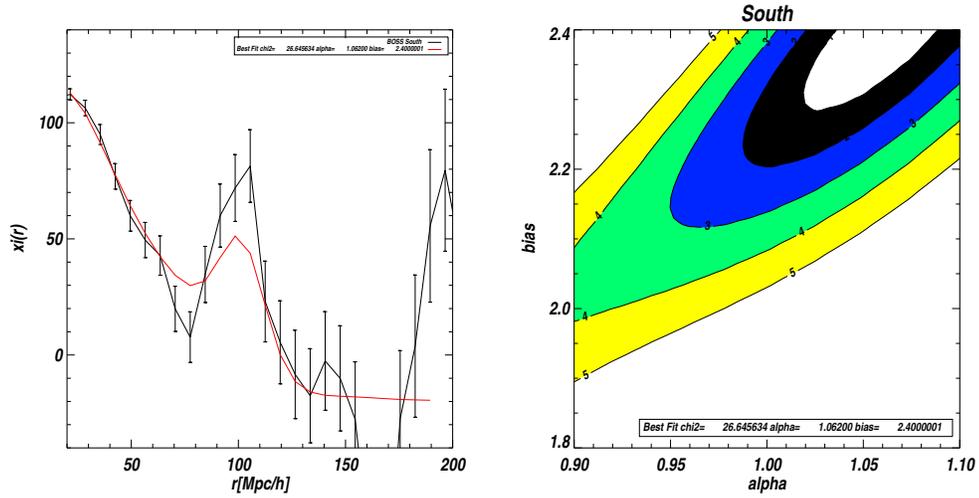


Figure VII.6: South_DR9 : Fitting α , bias. Bin 7 Mpc

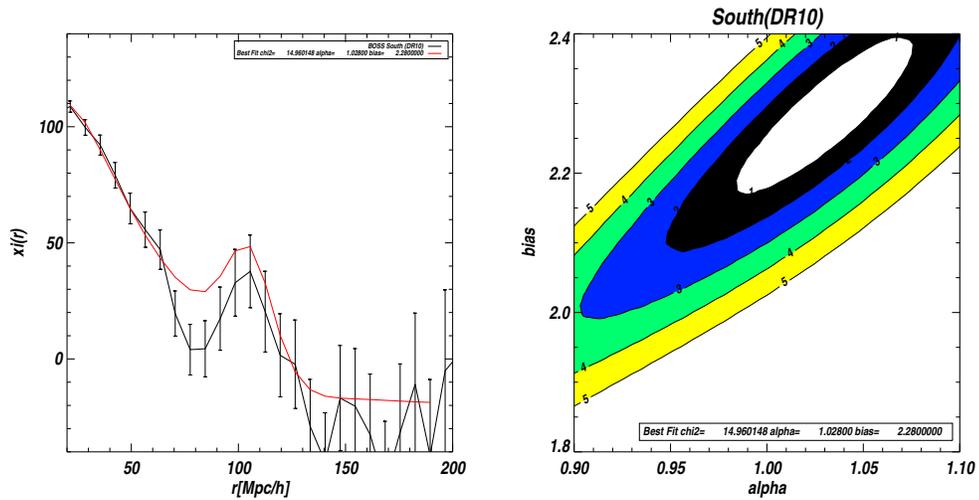


Figure VII.7: South DR10: Fitting α , bias . Bin 7 Mpc

0.2%) and the bias within 2% (compared to 0.6% of DR9). If we consider only the official binning the difference of the fitted parameters comparing North/South and North/South DR10, the difference in the fitted α pass from 6% to 4%. and in bias from 10% to 6%.

VII.3.3 Final constraints $\Omega_m - \alpha$, $\Omega_m - D_V$ and discussion

In order to get the final constraints we vary the 3 parameters $\Omega_m h^2$, b , and α , to get the conjoint confidence levels $\alpha - \Omega$ we marginalize over the bias. We continue separating the NGC (Fig. VII.11),

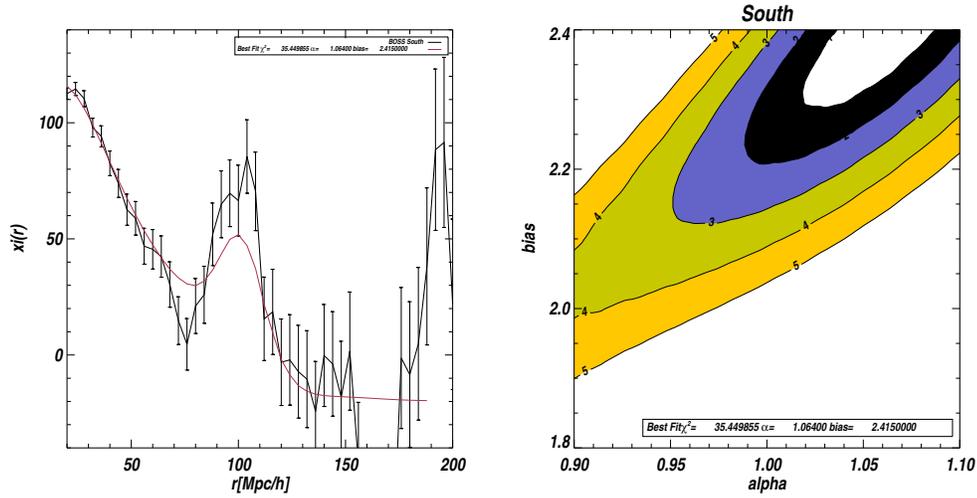


Figure VII.8: South DR9: Fitting alpha, bias . Bin 4 Mpc

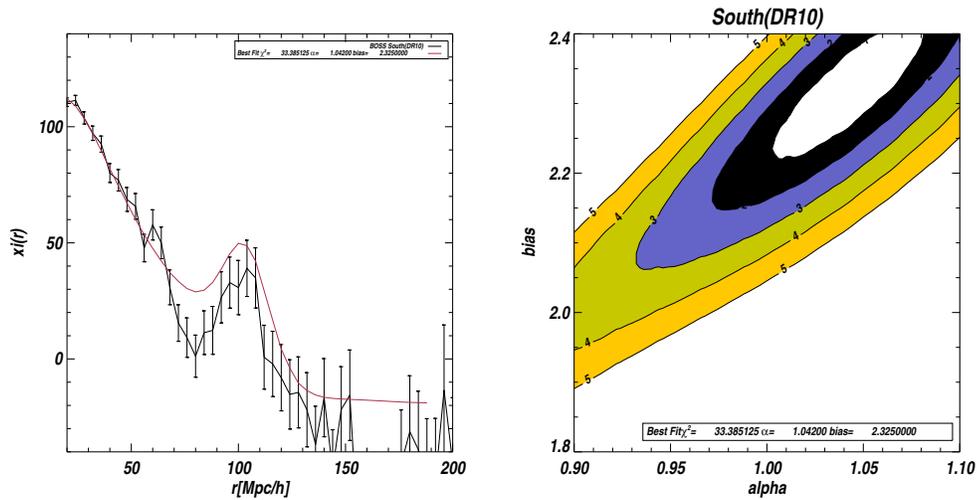


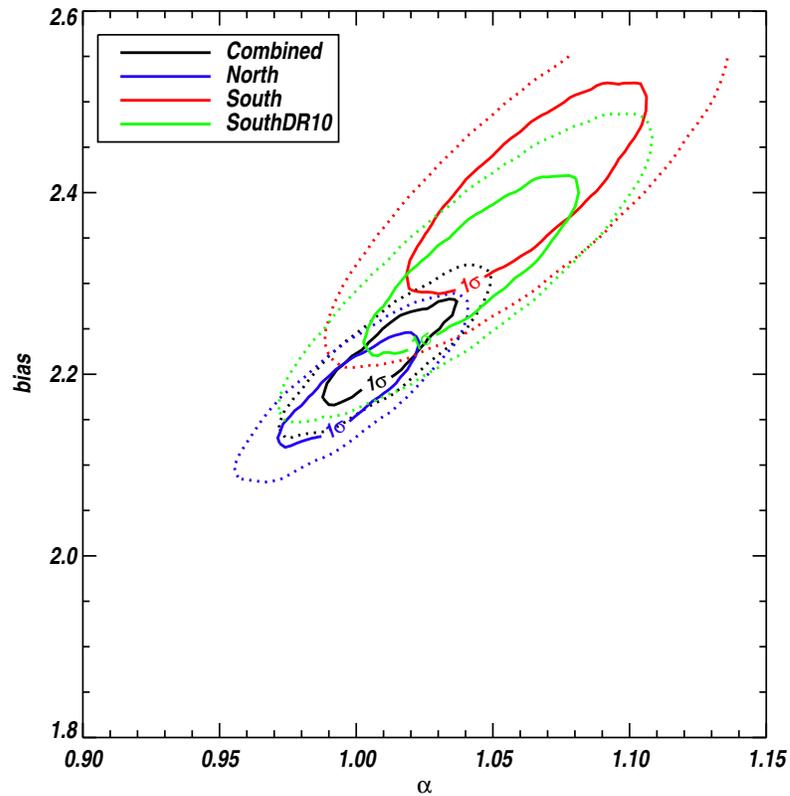
Figure VII.9: South DR10: Fitting alpha, bias . Bin 4 Mpc

SGC (Fig. VII.13) and the Combined Sample (Fig. VII.12) as well as the South DR10 (Fig. VII.14). A summary of the best fitting parameters are presented in the Table VII.3.4 , the error bars were predicted marginalizing over the remaining parameters.

Now we are going to translate the results of the distortion parameter α in terms of dilatation scale $D_V = \alpha D_{V, fid}$. In figure VII.15 the confidence contours $D_V = \alpha D_{V, fid}$ are shown as well as the 2 degeneracy directions as defined by Eisenstein et al 2005. The first direction corresponds to a constant measured acoustic peak separation: $r_s(z_d)/D_V(z = 0.55) = cte$. The second degeneracy direction corresponds to a constant measured shape of CDM spectrum $D_V(z = 0.55)\Omega_m h^2 = cte$.

Table VII.2: Constrains for 4Mpc and 7Mpc bins South DR9 and South DR10

Sample	χ^2	α	bias	Q
South(DR9) 4 Mpc	35.4498/41=0.865	1.0640 ± 0.023	2.4150	0.283
South(DR9) 7 Mpc	26.6456/25=1.066	1.0620 ± 0.030	2.4000	0.626
South(DR10) 4Mpc	33.3851/41=0.814	1.0420 ± 0.028	2.3250	0.204
South (DR10) 7Mpc	14.9601/25=0.598	1.0280 ± 0.030	2.2800	0.057

**Figure VII.10:** 1-2 σ Contours for North (blue), South(red), Combined (black) and the South DR10(green). Bin 4 Mpc

The orientation of the ellipses with respect to these axes indicates when the fits are influenced by shape of a pure CDM model and when the acoustic pic exerts an influence in the fit. For computing r_s the fitting formula of Eisenstein& Hu is used.

In table VII.3.4 the constraints in terms of D_V and $\Omega_M h^2$ are summarized. The distances measurements of $D_V(z = 55)$ and $\Omega_m h^2$ using the 4 samples gives broadly consistent results to 1σ level.

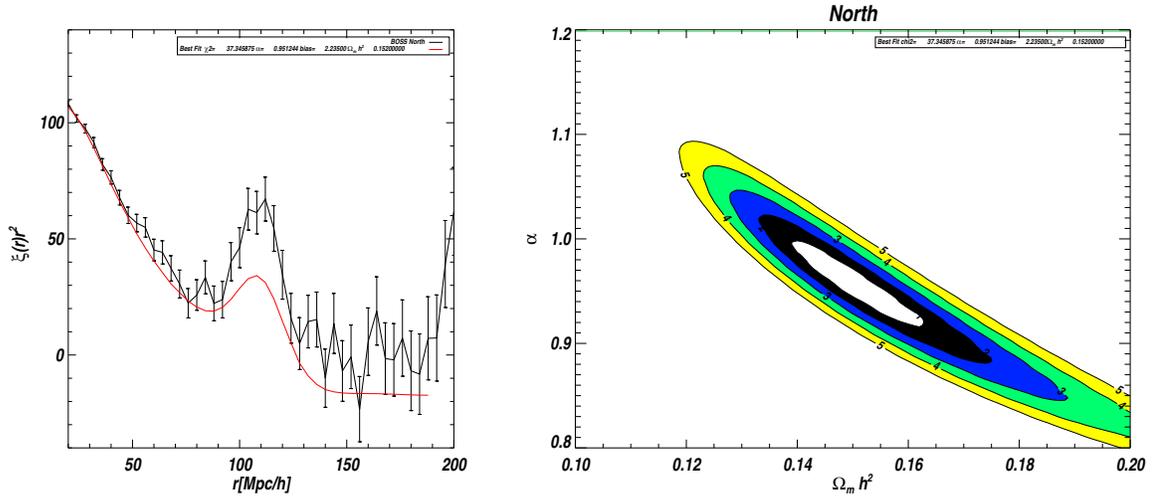


Figure VII.11: Contours confidence $\alpha - \Omega$ North

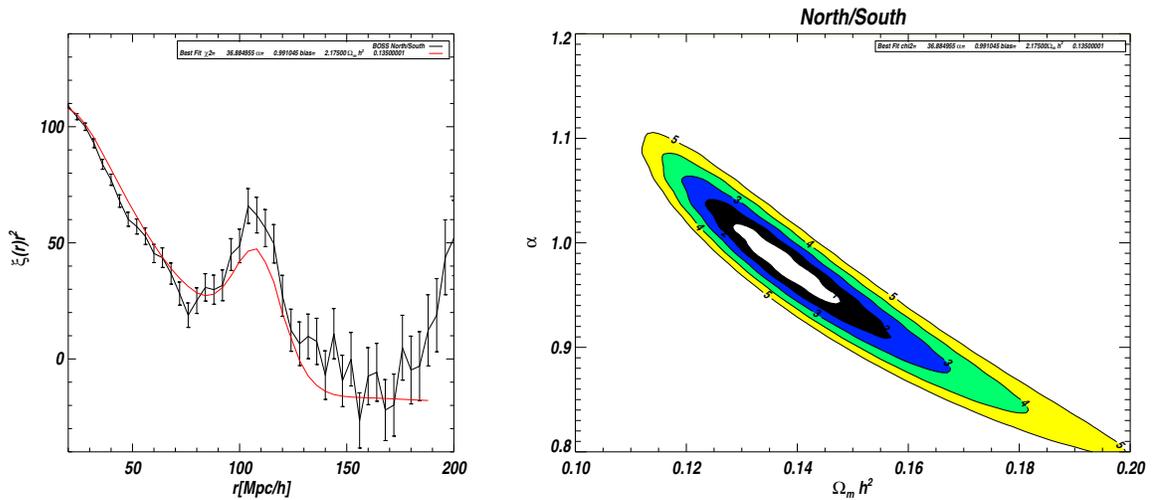
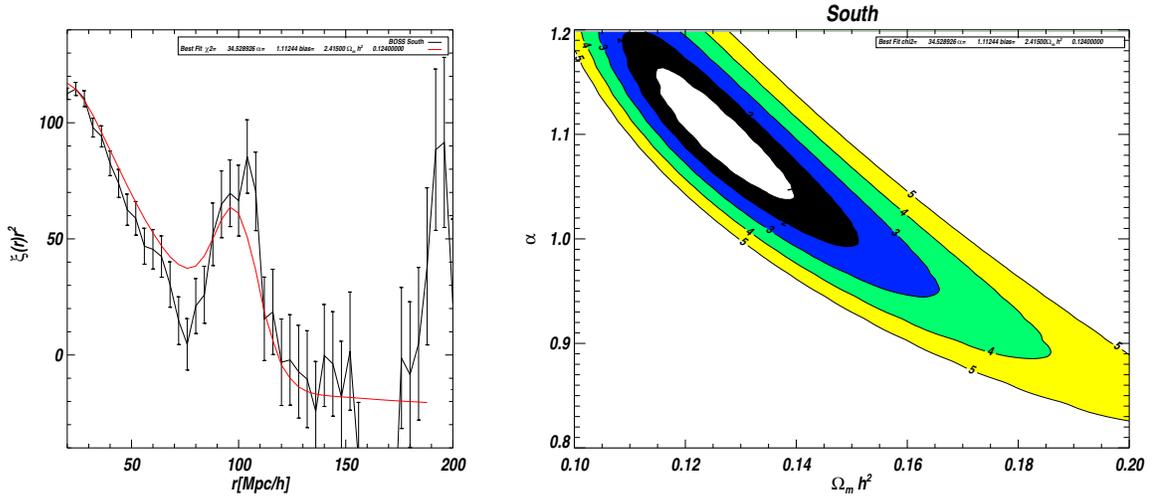
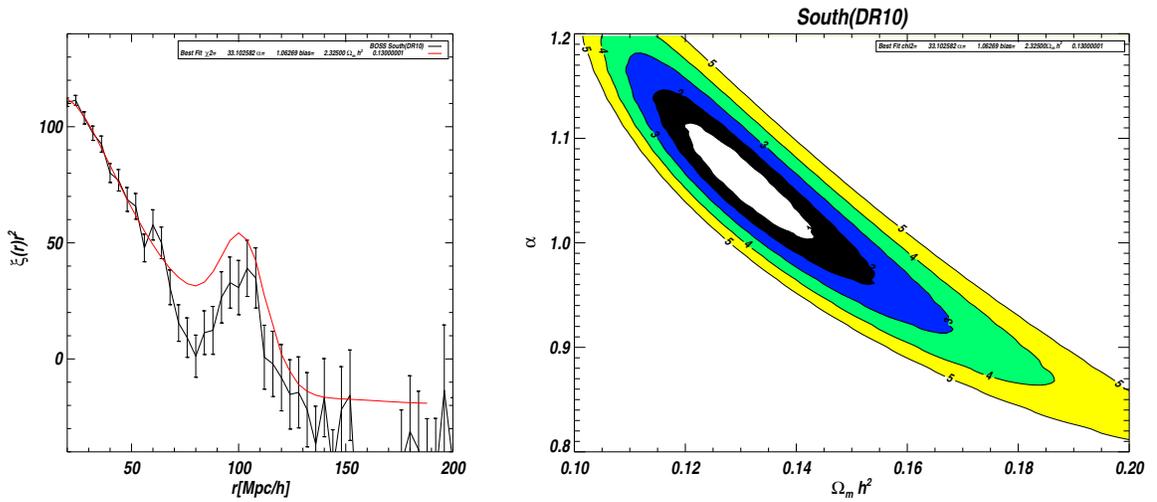


Figure VII.12: Contours confidence $\alpha - \Omega$ North/South

Finally I will compare our results to others studies and discuss the cosmological implications. In table ?? an non exhaustive list of the results from different samples /surveys is shown. Our results are in agreement with previous studies performed in other samples/surveys.

Within the working group, other results in cosmology are available, Xiaoying with an analog χ^2 analysis constraint the α parameter and Kushal using DR9 Combined sample data and using Maximum Likelihood covariance matrix and P(alpha) from Xiaoying's BAO fits performed a MCMC to constrain between others $\Omega_m h^2$. I present these preliminary results in a table to compare with them. As we can verify in the table our results concerning the α parameter are consistent with

Figure VII.13: Contours confidence $\alpha - \Omega$ South DR9Figure VII.14: Contours confidence $\alpha - \Omega$ South DR10

Xiaoying's results, the differences are generated because we are not using exactly the same model. In the case of $\Omega_m h^2$ our results are compatible with the results generated from MCMC also.

VII.3.4 Significance of BAO detection with BOSS

Finally we assess the question of the significance of the detection. We consider two perspectives commonly used, the first is considering the set of models with no acoustic oscillations, thus replacing $\xi_{mod} = \xi_{nowiggles}$ and the second perspective is considering the set of models with no baryons. In

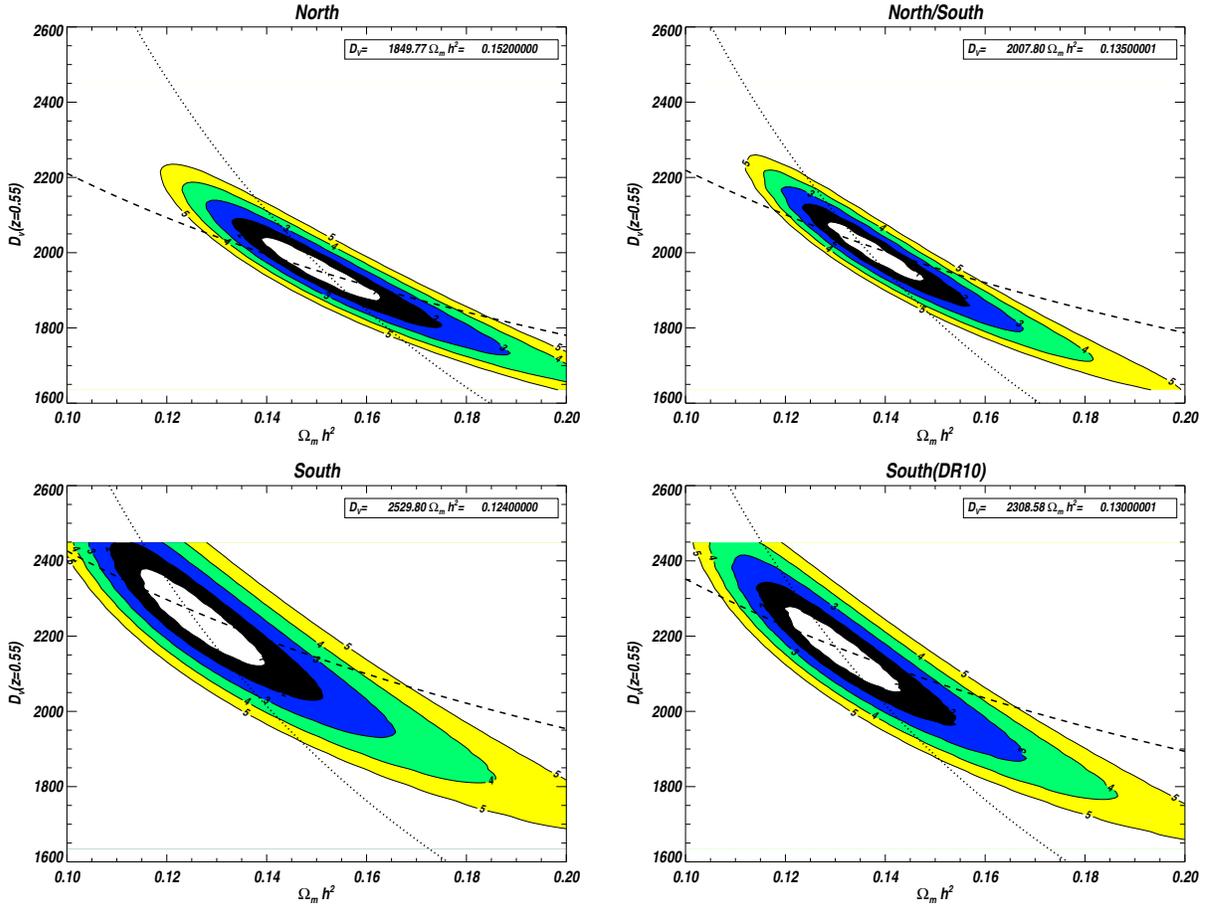


Figure VII.15: Contours confidence the physical matter density and the distance scale $\Omega_m h^2 - D_V(z = 0.55)$

Table VII.3: Constrains for D_V and $\Omega_m h^2$

Sample	α	D_V	$\Omega_m h^2$
North(DR9)	0.952 ± 0.028	1849.77 ± 57.24	0.1520 ± 0.10
Combined(DR9)	0.992 ± 0.026	2007.80 ± 53.15	0.1440 ± 0.007
South(DR9)	$1.11244 \pm$	$2529.80 \pm$	$0.1240 \pm$
South (DR10)	$1.06269 \pm$	$2317.23 \pm$	$0.1300 \pm$

figure VII.16 the best fit parameters for the models *no wiggles* for each region is shown.

$$\Delta\chi^2_{non-wiggles}(s) = \min(\chi^2_{nowiggles}) - \min(\chi^2_{BAO}) \quad (\text{VII.27})$$

$$\Delta\chi^2_{non-wiggles}(s) = \min(\chi^2_{noBaryons}) - \min(\chi^2_{BAO}) \quad (\text{VII.28})$$

In figure VII.17 the best fit parameters for the models without baryons. We combine these

Table VII.4: Constrains for D_V and $\Omega_m h^2$ from other studies

Sample	range fitting	D_V	$\Omega_m h^2$
WiGGLEZ Blake et al 2011	$[30, 180]h^{-1}$	$D_V(z=0.6)=2127.7\pm 127.9$	0.166 ± 0.14
WiGGLEZ Blake et al 2011	$[10, 180]h^{-1}$	$D_V(z=0.6)=2234.9\pm 115.2$	0.132 ± 0.11
SDSS-LRG Kazin et al 2009	$[60,150]h^{-1}$	$D_V(z=0.287) = 1132 \pm 40$ Mpc	---
SDSS-LRG Eisenstein et al 2005	$[?, ?]h^{-1}$	$D_V(z=0.35)=1370\pm 64$ Mpc	0.130 ± 0.01
BOSS DR9 <preliminar)< td=""> <td>$[20,180]h^{-1}$</td> <td>$D_V(z=0.55)=2007.80\pm 53.15$</td> <td>$0.1440\pm 0.007$</td> </preliminar)<>	$[20,180]h^{-1}$	$D_V(z=0.55)=2007.80\pm 53.15$	0.1440 ± 0.007

Table. Constrains for α and $\Omega_m h^2$ from other BOSS collaborators

Covariance	α (Xiaoying)	Constrains MCMC
Sample	$\alpha = 1.028 \pm 0.019$	$\Omega_m h^2 = 0.1378 \pm 0.0039$
ML	$\alpha = 1.025 \pm 0.021$	$\Omega_m = 0.289 \pm 0.018$
ANALYTIC	$\alpha = 1.026 \pm 0.019$	$H_0 = 69.1 \pm 1.5 km/s/Mpc$

results with the models with acoustic oscillations considered before to compute the significance of the detection. In table VII.3.4 the results are summarized. The combined sample gives a detection with a significance of 4.6σ over non wiggles models.

Sample	BAO Significance of Detection				
	$\chi^2_{nowiggles}$	$\chi^2_{nobaryons}$	χ^2_{BAO}	$\Delta\chi^2_{NW}$	$\Delta\chi^2_{NB}$
North(DR9) 4 Mpc	55.2057	56.8878	37.3140	17.9510(4.23 σ)	19.5738(4.42 σ)
South(DR9) 4 Mpc*	47.7540	47.4059	34.5289	13.2252(3.6 σ)	12.8770(3.6 σ)
Combined(DR9) 4Mpc	58.1823	59.4024	33.7167	25.6824(4.95 σ)	24.4600(5.07 σ)
South (DR10) 4Mpc*	47.8985	41.7652	33.1025	14.7960(3.8 σ)	8.66271(2.9 σ)

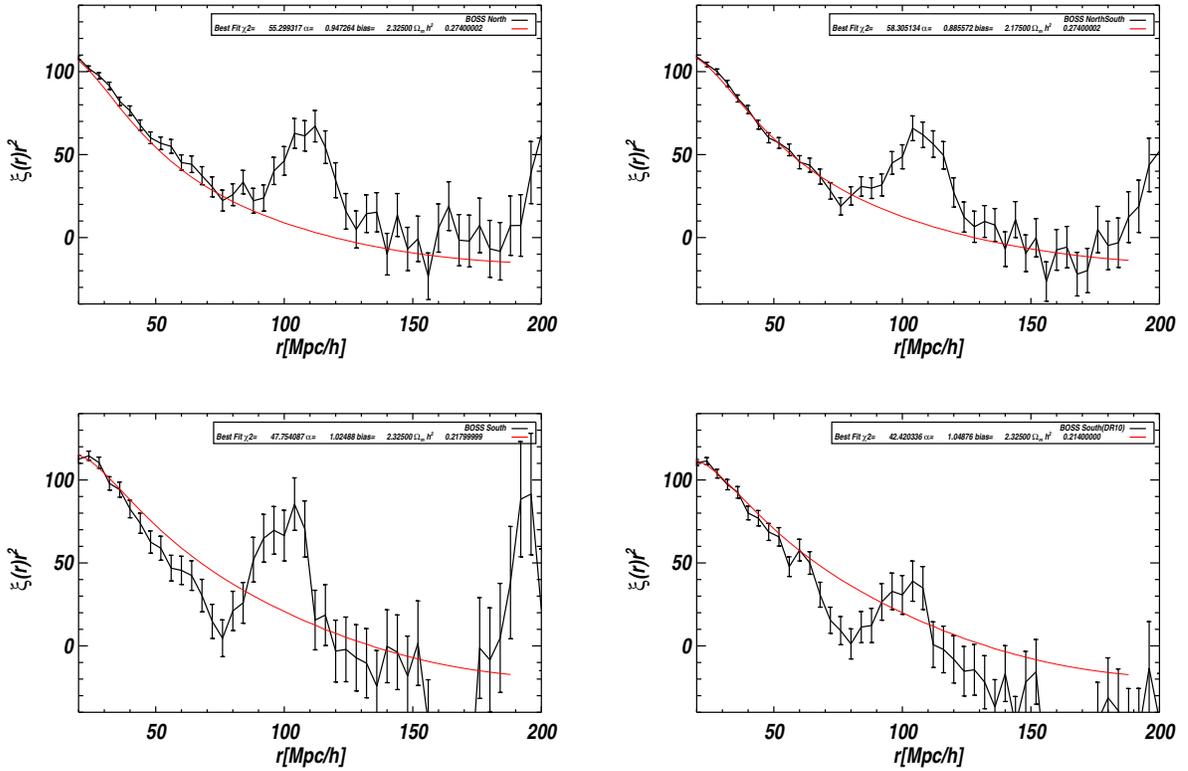


Figure VII.16: Best fit parameters for model no wiggles

Significance from other studies			
Sample	statistic	method	Significance
SDSS-LRG DR3 Eisenstein 2005	correlation function	no baryons	3.4σ
2dFGRS Colles 2001	power spectrum	no baryons	2.5σ
SDSS LRG DR4 Hutshi 2006	power spectrum	no wiggles	3.3σ
SDSS-LRG DR5 Percival 2005	power spectrum	smoothed ps	3.0σ
SDSS-LRG DR7Percival 2010	power spectrum	smoothed ps	3.6σ
SDSS-LRG DR7 E.Kazin 2010	correlation function	pic position	1.5σ
WiGGLEZ Blake 2011	correlation function	no wiggles	3.2σ
BOSS DR9	correlation function	no wiggles	4.95σ
BOSS DR9	correlation function	no baryons	5.07σ

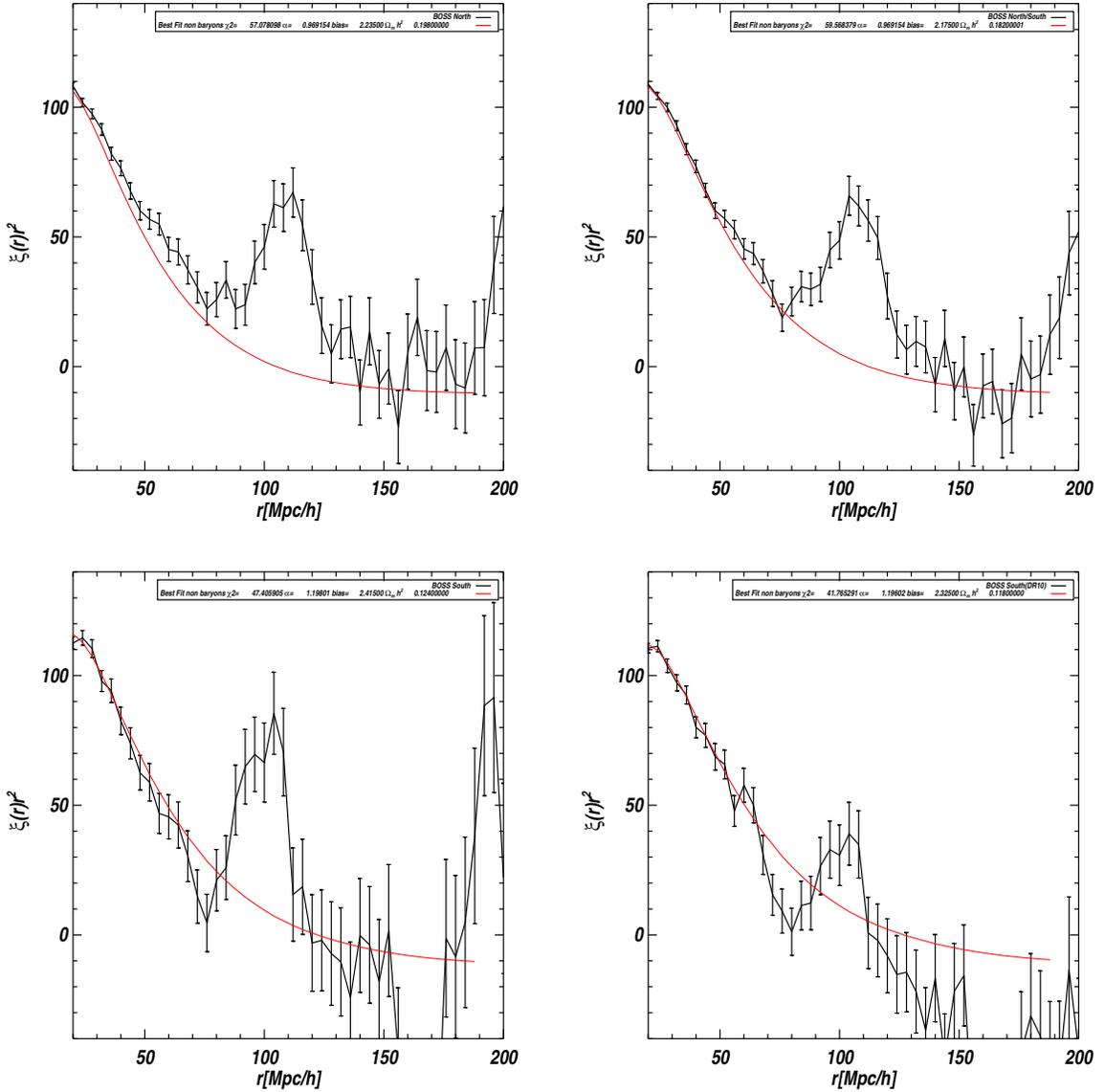


Figure VII.17: Best fit parameters for model no baryons

VII.4 Conclusions (preliminary) and comments

This analysis is very preliminary we are working in parallel with the working group to crosscheck the results and the cosmological interpretation. At this moment our results are compatible with the results of other members of the clustering group. I enumerate some preliminary conclusions.

1. The tension of South in DR9 sample that is more than $1 - \sigma$ from North seems to be less significant when including more data (i.e DR10 data). This fact suggests the South in DR9

presents a large fluctuation.

2. The preliminary results indicate a $\Omega_m h^2 = 0.144 \pm 0.007$ and $\alpha = 0.992 \pm 0.026$. The constraints in α traduce in constraints in distance scale $D_V = 2007.80 \pm 53.15$. Our results are in agreement with previous studies.
3. The combined sample gives a detection with a significance of 4.6σ over non wiggles models for the combined sample. The south has added a $\Delta\chi^2 \sim 7$ versus an expected $\Delta\chi^2 \sim 4$.

VII.5 Perspectives

- A further work should consider the wide angle effects,. Actually, in the past, as the surveys were smaller, the angle of separation was within this approximation. Some references about this effect could be consulted in Papai & Szapudi 2008 (Pápai & Szapudi, 2008).
- In the modeling of the correlation function, it is possible to include a scale dependent bias $B(s)$.

$$\xi_{fid}(r) = B(s) * \xi_{damp,NL}(\alpha * r) \quad (\text{VII.29})$$

where $B(s)$ is fitted from N-body simulations using the halo catalogs. That is done comparing the non linear redshift space halo correlation function issue from N-body simulations of halos and the non linear dark matter correlation function issue from our model correlation function.

- For modeling the correlation function there exists another approach derived from renormalized perturbations theory. This procedure was already used by (Sánchez et al., 2009) and (Beutler et al., 2011)

$$\xi_{NL} = b^2[\xi_L(r) * e^{(-k \cdot r)^2} + A_{MC} \xi'_L(r) \xi_L^{(1)}(r)] \quad (\text{VII.30})$$

where, b , k and A_{MC} are nuisance parameters and $*$ denotes convolution, $\partial\xi_L/\partial r$ is the derivative of linear correlation function and $\xi_L^{(1)}$ is given by:

$$\xi_L^{(1)}(r) \equiv \hat{r} \cdot \nabla^{-1} \xi_L(r) = 4\pi \int P_L(k) j_1(kr) k dk \quad (\text{VII.31})$$

this model is inspired from RPT where the matter power spectrum is :

$$P = G^2 P_L + P_{MC} \quad (\text{VII.32})$$

where G is the non linear growth factor and P_{MC} the power generated by the mode coupling.

G is well approximated by a gaussian form, and the leading order contribution to P_{MC} is the term $\xi'_L(r) \xi_L^{(1)}(r)$

- The next step is to use techniques as MCMC to combine different datasets as CMB - WMAP7, SNe Ia data, Direct Hubble constant measurement and/or other BAO data.

VIII. Optimization of correlation function estimator for BOSS survey

VIII.1 Motivation

Along the chapter 4 we have studied different estimators of the correlation function (Hamilton, Peebles-Hauser, etc) and we have concluded that among them the best properties for an estimator of the correlation function for large scales is given by the Landy-Szalay estimator. Similar studies have proven that the Landy-Szalay is the best estimator for the case of no correlations (Landy & Szalay, 1993), for the weak correlation limit (Kerscher et al., 2000) and in SDSS I-II geometry (Labatie et al., 2010). However we have shown that the best estimator depends on the geometry of the survey. We have seen that the best estimator most satisfy minimum bias and variance. The fact that the bias and variance of different estimators depends of the geometry was the motivation to asked ourselves if it was possible to construct an estimator optimized for a survey. In particular we would be interested in optimizing the estimator for BOSS survey. In figure VIII.1 the performance of the estimator as a function of the geometry is shown. In top panel the mean of 300 log normal simulations in a cubic box and in the left panel following the BOSS geometry. In bottom panels, in right side the bias (scaled by r^2) and in right the variance (scaled by r^2) for the the different estimators in a Cubic and in BOSS geometries. As we observe in the plot the bias and errors for the estimators have different performances as a function of the geometry. An analog plot for DR7 was shown in chapter 4.

As a natural extension of chapter 4, having in hand the tools as the lognormal simulations and knowing the properties of a survey, in this chapter I study the possibility to optimize the estimator for the particular case of BOSS. The structure of the chapter is the following in the first section I describe the methodology followed to optimize the estimator. I describe the simulations characteristics and performance. Then I describe the estimator definition and the optimization criterium. As we will see there is a compromise between the minimization of the variance and the bias. In the second part I discuss the limit cases, to help to set the bar of the optimization. We will find that the best compromise is the so called mean square error corresponding to $\lambda = 0.5$. Once the optimized estimator is defined I study the properties of this estimator. First, I show what is the improvement in the optimized estimator to other

geometries. Then, I study what is the contribution of the different orders. Finally, to validate the optimization process we test if there is no over-fitting. The last section is devoted to the conclusions and perspectives.

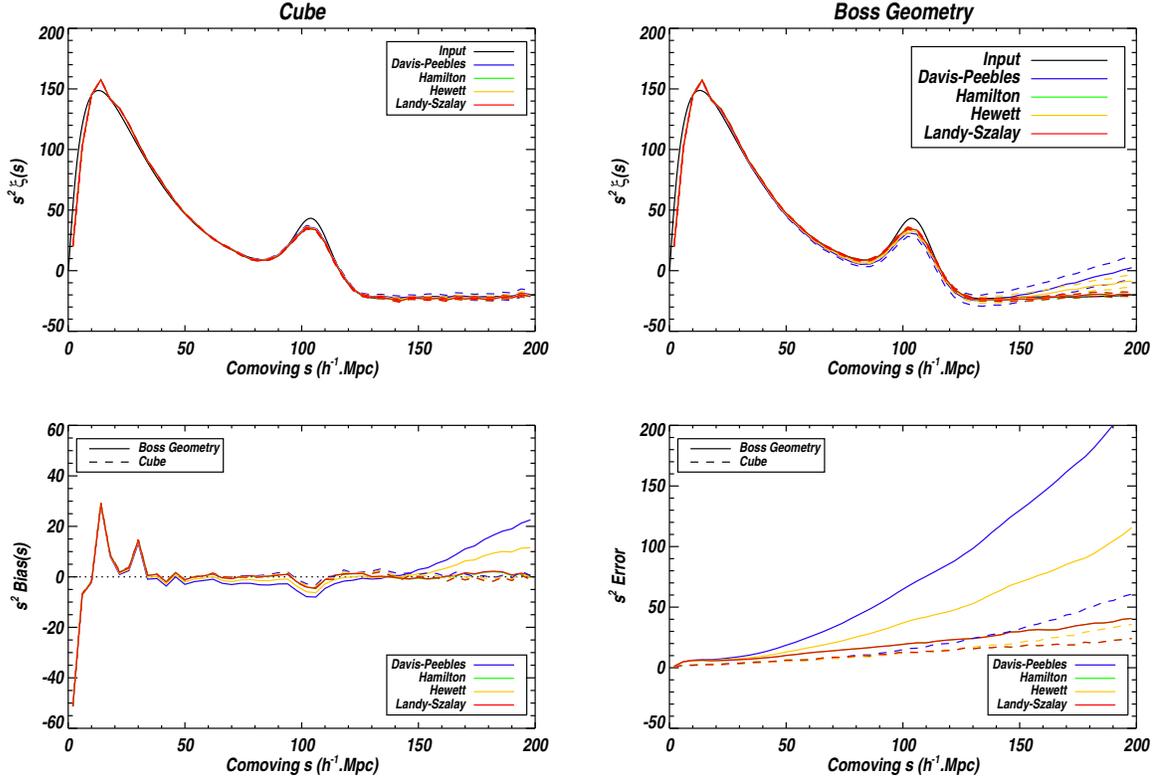


Figure VIII.1: Relative performance of estimators depends on geometry. In top left panel the mean of 300 log normal simulations in a cubic box and in top right panel following the BOSS geometry for different estimators. In bottom panels, in left side the bias (scaled by r^2) and in right the variance (scaled by r^2) for the different estimators in a Cubic and in BOSS geometries.

VIII.2 Methodology

VIII.2.1 Simulations

In order to optimize the 2 point correlation function for a given geometry we will use the lognormal simulations described in chapter 4. The advantages of this kind of simulations is that we know the input correlation function thus the bias could be defined without ambiguity. For this study we generated a set of 300 log normal simulations following Λ CDM cosmology without velocities. The simulations were generated in a 1 Gpc periodic Box with 256^3 pixels. The box was replicated as explained in chapter 4 to cover the entire BOSS geometry. The galaxies are drawn from the lognormal density field using this as a PDF. The generated mock

catalogs mimic BOSS geometry and number density of CMASS sample (not updated to DR9). The random used to compute the correlation function are three times larger than the data.

Before implementing the optimization we tested our simulations to determine the intrinsic bias of simulations, for this we compare the input correlation function with the correlation function from Log-normal density field, and the average of 300 galaxy mocks in the box and in BOSS geometries. The results are shown in figure VIII.2 in left panel.

In right panel, the deviations of the average correlation function of the simulations in the Cubic and BOSS geometries from the input correlation function (scaled by r^2) is shown as well as the 1 sigma intervals. The results shown the minimum reliable distance is $r=40$ Mpc/h. We observe that mean over the simulations do not reproduce exactly the input correlation function. There is a slight under estimation of the correlation at the level of the BAO pic. For the following we will use the mean over density field to avoid introducing an extra bias from the simulations. In right panel the deviations of the average correlation function of

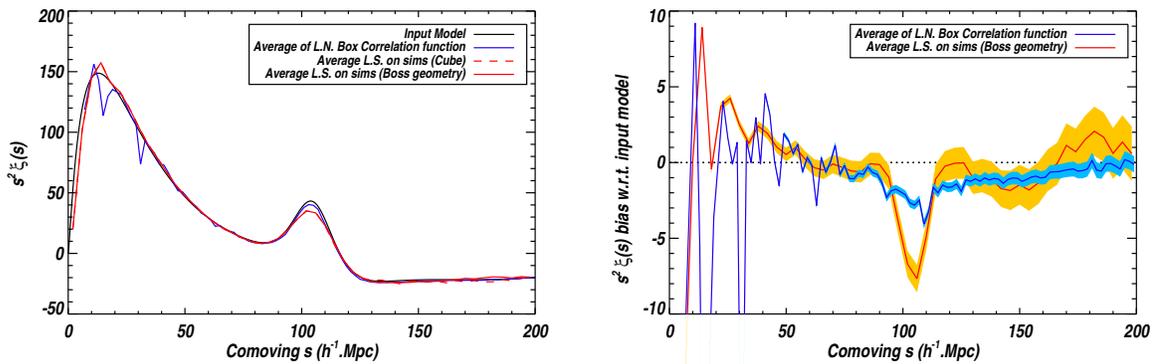


Figure VIII.2: In left panel the average correlation function of the 300 simulations in the Cubic and BOSS geometries, compared with the input model and the correlation function of the lognormal density field. In right panel the deviations of the average correlation function of simulations in the box and BOSS geometry from the input correlation function (scaled by r^2) with the 1σ intervals are plotted also.

simulations in the box and BOSS geometry from the input correlation function (scaled by r^2) with the 1σ intervals are plotted. As we observe from the figure there is apparently an issue with the replication and/or drawing of galaxies because the correlation function of galaxies (yellow region) is significantly more biased than that of the LN field in the box (blue region).

VIII.2.2 Estimator Definition

As we have seen in chapter 4, the different estimators are a combination of the pairs counts data-data(DD), random-random(RR), and random-data (RD). We can generalize the estimator taking into account all possible combinations of these terms, linear, quadratic and higher orders and we can look for the parameters that optimize the estimator for a given sample.

$$\xi_{\text{new}}(r_i) = \sum_t \alpha_t R_t(r_i) \quad (\text{VIII.1})$$

The list of possible terms are :

$$\begin{array}{l}
0^\circ \text{ order} \quad \text{cte} \\
1^\circ \text{ order} \quad \left\{ \begin{array}{l} \frac{RR}{DD}, \frac{DR}{DD}, \frac{DD}{RR}, \frac{DR}{RR}, \frac{RR}{DR}, \frac{DD}{DR} \\ \frac{DR \times RR}{DD^2}, \frac{RR^2}{DD^2}, \frac{DR^2}{DD^2}, \frac{DR \times DD}{RR^2} \end{array} \right. \\
2^\circ \text{ order} \quad \left\{ \begin{array}{l} \frac{DD^2}{RR^2}, \frac{DR^2}{RR^2}, \frac{DD^2}{DR^2}, \frac{RR^2}{DR^2} \\ \frac{DD \times RR}{DR^2}, \frac{RR^2}{DD \times DR}, \frac{DR^2}{DD \times RR}, \frac{DD^2}{DR \times RR} \end{array} \right.
\end{array} \quad (\text{VIII.2})$$

Thus for each mock we will compute and store the pair counts (DD, RR, RD).

VIII.2.3 Optimization criterium

In chapter 4 we have said a good estimator must minimize the variance and the bias, these two criteria will guide the optimization. To find best parameters that optimized the BOSS estimator we use the method of maximum likelihood ¹. To consider both criteria, minimum variance and bias, we need to set separately the χ^2 for each criteria. For minimizing the variance the χ^2 is defined as the difference between each measure of the correlation function and the input estimation:

$$\chi_v^2 = \sum_i \sum_s \left(\frac{\xi_{new,s}(r_i) - \xi_{th}(r_i)}{\sigma_{LS}} \right)^2 \quad (\text{VIII.3})$$

For minimizing the bias we must compare instead of each realization the difference between the mean of the whole set of simulations and the real correlation function, thus we must minimize the χ^2 defined as:

$$\chi_b^2 = \sum_i \left(\frac{\langle \xi_{new,s}(r_i) \rangle_{sim} - \xi_{th}(r_i)}{\sigma_{LS}} \right)^2 \quad (\text{VIII.4})$$

where:

$$\langle \xi_{new,s}(r_i) \rangle_{sim} = \frac{1}{N_s} \sum_s \xi_{new,s}(r_i) \quad (\text{VIII.5})$$

The estimator that satisfies both criteria is given by $\chi^2(\lambda)$:

$$\chi^2(\alpha) = (1 - \lambda)\chi_v^2 + \lambda\chi_b^2 \quad (\text{VIII.6})$$

The errors for the χ^2 are given by the Landy-Szalay errors this choice implies that the optimization is defined with respect to Landy-Szalay errors.

¹If we supposing the PDF is gaussian, the Maximum Likelihood is equivalent to minimizing the χ^2 .

VIII.3 Setting the bar for the optimization

As we see from equation VIII.6 there is a compromise between both criteria: minimum bias and minimum variance. At first approach it is not clear what is the best choice to optimize the estimator, we will test the limiting cases to have a feeling of the behavior.

VIII.3.1 Minimum Variance

The first case we explored is the minimum variance estimator that corresponds to $\lambda = 0$. In figure VIII.3 the results are shown. In the left panel the bias and error (RMS) is shown for the optimized estimator and the Landy-Szalay estimator. As we observe from the figure the variance for the optimized estimator is reduced for all scales (being at the BAO scales a factor of 2 smaller) by contrary the bias increases by a factor of 3, resting within the variance level. In fact the idea of the minimal variance estimator is that if it is possible to know the bias and correct it, this option will reduce dramatically the variance. To show the effects in the left panel of figure VIII.3 the signal to noise ratio for the bias corrected and non corrected is shown for both cases, optimal estimator and Landy-Szalay. We can see the gain in signal/noise passe for the optimal estimator compared to Landy-Szalay. This option could be interesting if the bias is calculated from Monte Carlo simulations.

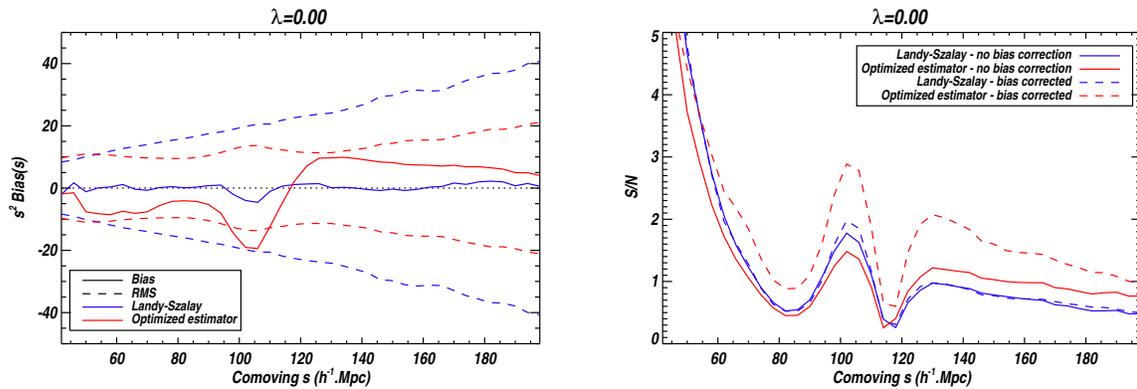


Figure VIII.3: Minimal Variance estimator ($\lambda = 0$). In left panel the bias (solid line) and error (dotted line) for the optimal (red) and the LS (blue) estimators. In right panel, the signal to noise ratio for the bias corrected (dotted line) and non corrected (solid lines) for LS (red) and BOSS (blue) estimators.

VIII.3.2 Minimum Bias

Now lets study the other case extreme:the minimum bias, where $\lambda = 1$. In figure VIII.4 in left panel the bias and error (RMS) is shown for the optimized estimator and the Landy-Szalay estimator (scaled by r^2). As we observe from the figure the variance for the optimized estimator is huge with respect to Landy-Szalay for all scales, while the bias is smaller in

respect to Landy Szalay estimator. This limit is not interesting to optimize the estimator. In terms of signal to noise, the correction of the bias do not improve the estimator.

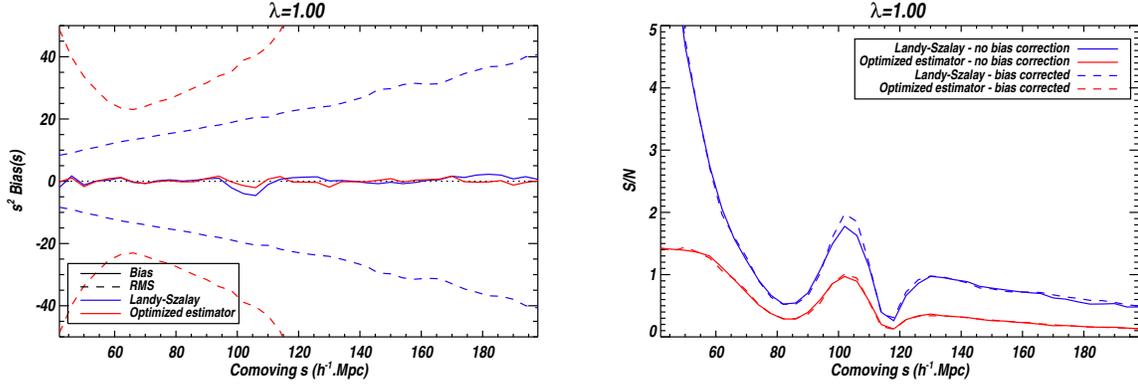


Figure VIII.4: Minimal Bias estimator ($\lambda = 0$). In left panel the bias (solid line) and error (dotted line) for the optimal (red) and the LS (blue) estimators. In right panel, the signal to noise ratio for the bias corrected (dotted line) and non corrected (solid lines) for LS (red) and BOSS (blue) estimators.

VIII.3.3 Mean Square Errors

The *minimum square error* estimator is given by $\lambda = 0.5^2$. In figure VIII.5 the results are shown. In the left panel the bias and error (RMS) is shown for the optimized estimator and the Landy-Szalay estimator. As we observe from the figure there is an improvement of the variance for the optimized estimator in respect to Landy Szalay, and the bias is reduced with respect to the minimal variance but is still larger than Landy Szalay however within the variance error. The argument used in the first case is again applicable in this case. If we can calculate the bias using Monte Carlo realizations we observe an improvement in terms of signal to noise from 1 to 2.5 at the scale BAO.

VIII.3.4 Landy-Szalay

Finally, we look for λ that give us an equivalent to Landy Szalay estimator, and we find that $\lambda = 0.96$ give us the Landy Szalay estimator. In figure VIII.6 the results are shown.

²The MSE of an estimator in respect to the estimated parameter θ is defined as

$$\text{MSE}(\hat{\theta}) = \text{E} [(\hat{\theta} - \theta)^2]. \quad (\text{VIII.7})$$

The MSE is equal to the sum of the variance and the squared bias of the estimator

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \left(\text{Bias}(\hat{\theta}, \theta)\right)^2. \quad (\text{VIII.8})$$

The MSE thus assesses the quality of an estimator in terms of its variation and unbiasedness.

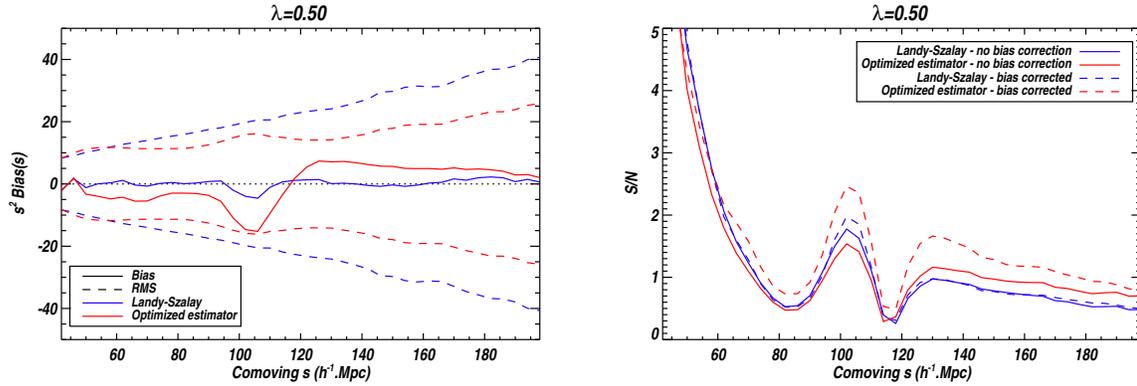


Figure VIII.5: Mean Square Error estimator ($\lambda = 0.5$). In left panel the bias (solid line) and error (dotted line) for the optimal (red) and the LS (blue) estimators. In right panel, the signal to noise ratio for the bias corrected (dotted line) and non corrected (solid lines) for LS (red) and BOSS (blue) estimators.

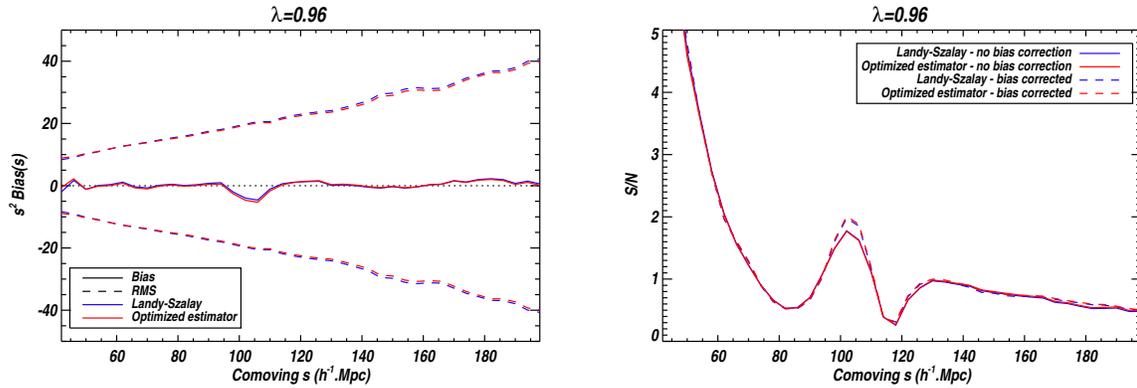


Figure VIII.6: Landy-Szalay estimator ($\lambda = 0.96$). In left panel the bias (solid line) and error (dotted line) for the optimal (red) and the LS (blue) estimators. In right panel, the signal to noise ratio for the bias corrected (dotted line) and non corrected (solid lines) for LS (red) and BOSS (blue) estimators.

VIII.3.5 Optimization Performance as a function r_{min}

Looking to answer the question of the best criteria of optimization we will study the behavior of the mean error and mean bias in respect to Landy -Szalay as a function of the minimal distance used in the fit r_{min} . In top panel of figure VIII.7 the results are shown, each color represents a value for r_{min} . For each r_{min} we study how the mean error and mean bias varies as a function of λ values. Let us discuss the bias first, the dotted line represents the case where mean bias of the optimal estimator is equal to Landy-Szalay. We observe for a fixed value of r_{min} , the bias decreases as we approach to the limit of minimum bias limit ($\lambda = 1$), but only in the case of $r_{min} = 50$ we achieve a average level below Landy Szalay for λ values over

0.9. We observe that for increasing values of r_{min} the mean bias increases also. At low r_{min} we are biased by the fact that the LN simulations is bad: in a pixel, the galaxies are drawn uniformly, so there is no power at small scales. Concerning the RMS, for increasing values of r_{min} we have in general mean error smaller with respect to Landy-Szalay for ($\lambda < 0.9$). We observe also that in all cases for λ values larger than 0.9 the variance explodes.

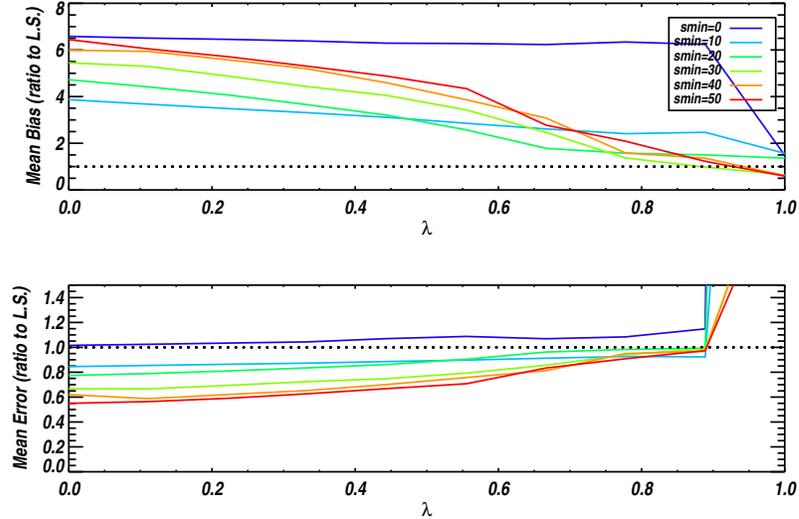


Figure VIII.7: Mean Bias and mean RMS evolution with λ for different minimal distance considered in the fit (r_{min}).

VIII.4 Best Compromise $\lambda = 0.5$

Before the exploration it seems that the mean square error estimator has ($\lambda = 0.5$) presented the best compromise. In figure VIII.8 the bias of Landy-Szalay and the optimal estimator is shown and the ratio of the optimal error bars respect to Landy-Szalay. The bias is smaller than 1σ and presents an improvement in term of the variance between 20% and 40% with respect to Landy-Szalay. Now that we have set the bar we will perform some tests to the new estimator: we will compute the covariance matrices, the effect of applying to other geometry and the different contributions for the different orders.

VIII.4.1 Orders contribution to optimize estimator

Now we will explore what are the contributions of the different terms in the optimized estimator, we compare the contributions by using only first order terms, adding second order terms, using all terms. In figure VIII.9, the bias (solid lines) and the variance (dotted lines) are shown for the 3 cases, first order (green), adding the second order (yellow), using all terms (red), and compared to the reference the Landy Szalay estimator (blue). As we observe from

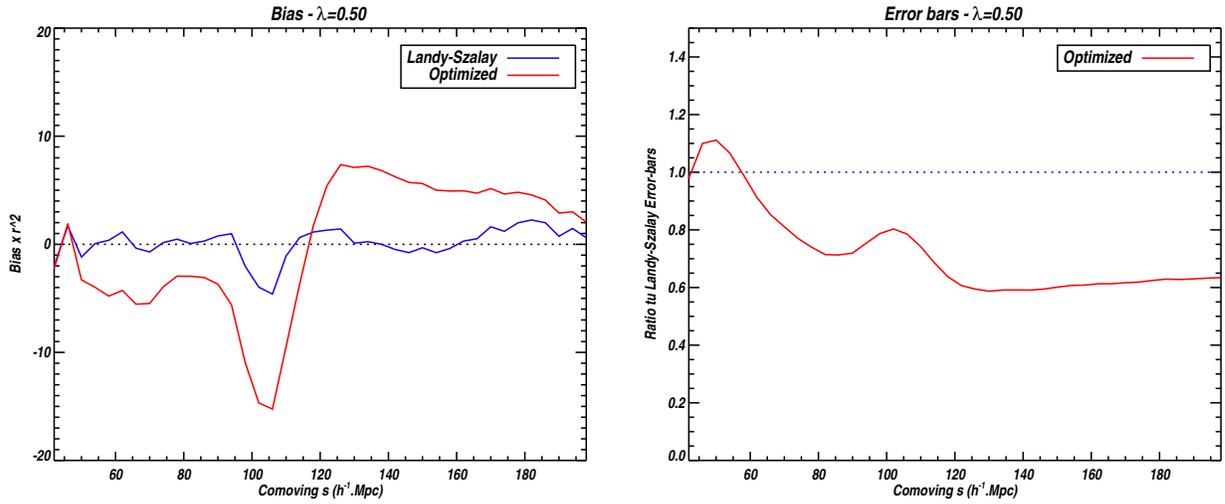


Figure VIII.8: The bias of Landy Szalay and the optimal estimator is shown and the ratio of the optimal error bars respect to Landy-Szalay

the figure in terms of the variance the best case is always using all terms. But even using only the first order terms we observe a considerable improvement in terms of the variance. The difference in term of bias is less significant.

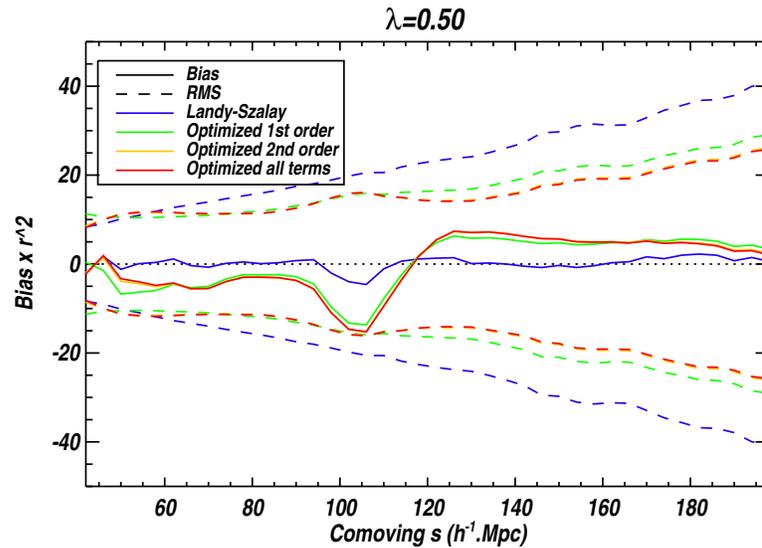


Figure VIII.9: The bias (solid lines) and variance (dotted lines) are shown for the three cases, first order (green), adding the second order (yellow), using all terms (red) and compared to the reference the Landy Szalay estimator (blue). The yellow is so close to red that it is almost indistinguishable.

VIII.4.2 Varying Geometry

We have seen that in the cube the estimators presented similar behaviors (bias and variance) and with more complex geometries the performance varies. Now we study how the optimized estimator for a complex geometry behaves when it is applied to a simpler geometry. In figure VIII.10 we explored different combinations, we compare the behavior of the optimized estimator for BOSS applied to BOSS (in red solid line) to the case where the optimization is performed over a simpler geometry as a cube and then applied to BOSS (green solid line). As we observe from the figure there is an considerable increase of the variance of the estimator for the case where the optimization was performed in a simpler geometry and then applied to complex one, the errors become larger than Landy-Szalay in this case. Let us explore the opposite case where the estimator is optimized for BOSS and applied to simpler geometry (green dotted line), in this case the optimized estimator conserves their properties even if it is applied to a simpler geometry.

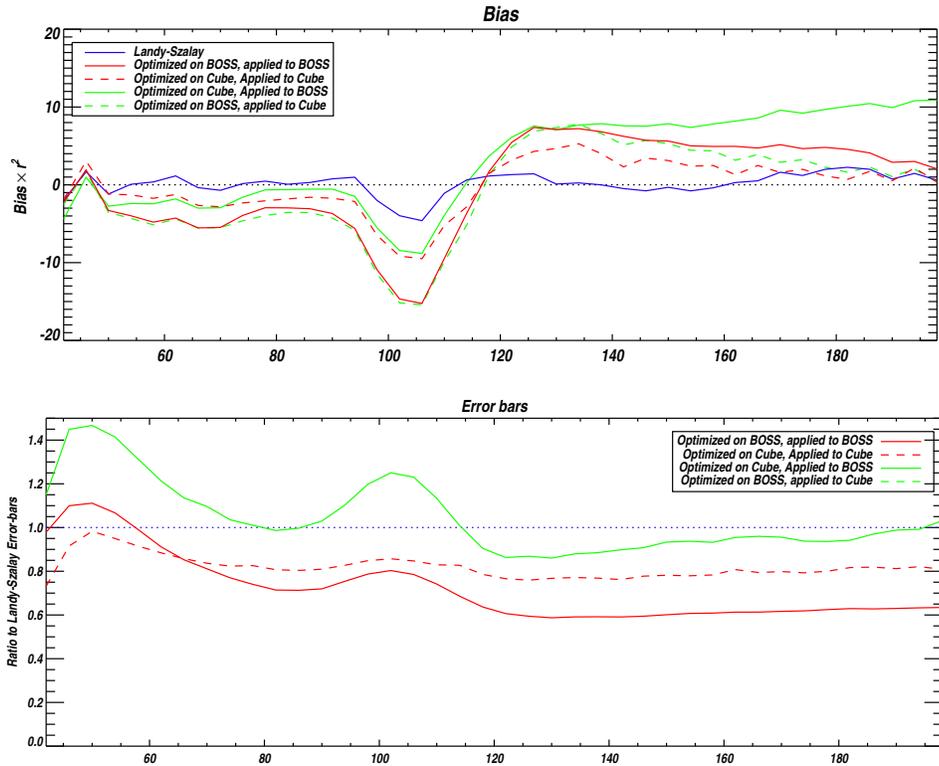


Figure VIII.10: Effect of the optimization in a given geometry and applied to a simpler/complex one.

VIII.4.3 Covariance Matrices

The first test we perform is to assure that the estimator do not introduce correlation between bins. We calculate the covariance matrices and correlation matrices to compare the Landy-Szalay and the BOSS estimator. In figure VIII.11 I show the covariances matrices using the Landy -Szalay estimator and the BOSS estimator as well as the correlation matrices. As we observe from the figure the improvement in terms of variance is remarkable. Furthermore if we compare the correlation coefficients matrices we notice there is no extra extra correlation between bins introduced by the new estimator.

VIII.5 Testing if over-fitting

Finally we perform a test to warranty we are not over-fitting. In fact for performing the optimization we used the same sample to optimize the estimator and to test the estimator. To test we are not over-fitting the sample we divide the samples in a sample to training and

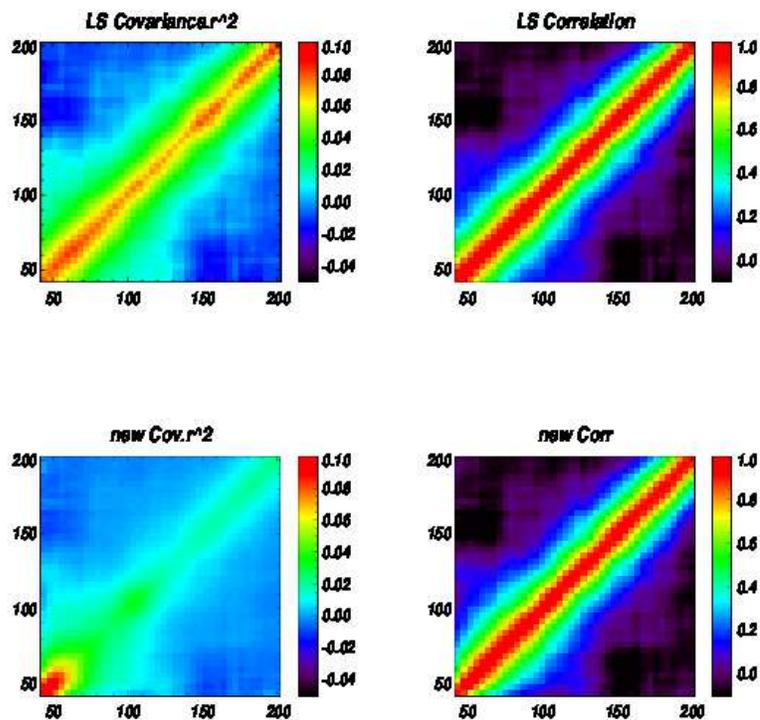


Figure VIII.11: In top panel, Covariance matrices of Landy-Szalay estimator (left) and new estimator (right). In Bottom panel correlation matrices for the LS estimator (left) and new estimator (right)

a sample to test it. In figure VIII.12 the results are show, in solid lines the bias and in dotted lines the variance. We plot two cases, in green the optimization and test are done with the training sample and in red the optimization is done with the training sample and the test with another sample. As we observe the results are still valid for both cases.

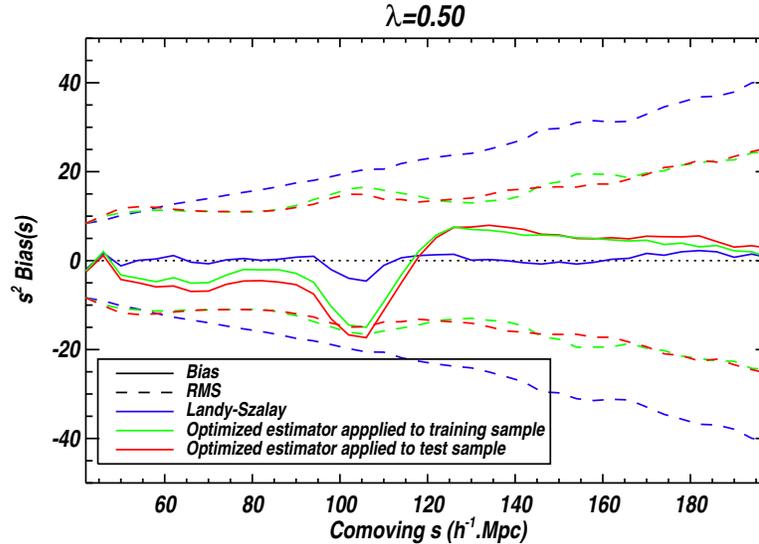


Figure VIII.12: Testing if over-fitting. In solid lines the bias and in dotted lines the rms. We plot 2 cases, in green the optimization and test are done with the training sample in red the optimization is done with the training sample and the test with another sample

VIII.6 Conclusions and Perspectives

We have developed a simple and straightforward way to optimize the correlation function estimator for a given survey. The extra bias introduced can be controlled via the λ parameter to be of the order of the variance. The bias can be exactly calculated via the Monte Carlo simulations and corrected. We get a gain of the order of 20% to 40% in signal to noise.

The final goal of this work would be to applied this optimized estimator to BOSS data. In order to validate this estimator there is still a lot of work do be done. As we have seen the bias present a pic, one of the issue that must be studied is that the optimization is not generating an artificial peak that biases the measure.

We need to use more realistic simulations as PTHALOS and study the behavior when the non linear behavior is introduced. Within the collaboration there are available a set of PTHALOS simulations with BOSS geometry thus the next step is to applied the optimization procedure on these simulations. With PTHALOS mocks we would test if we get the performance of the optimized estimator found with log normal simulations. Then, the next test would be to do is to use the N-body simulations , Las DAMAS CMASS. Even if they cover only a small region of the DR9 , this kind of simulations is closer to the real galaxy catalogs. In further stage,

the study could be extended to test the consequences in terms of parameter constraints.

We want to introduce FKP weights that could reduce the gain in S/N as this weighting scheme is supposed to improve optimally the estimator.

Another possible axis to develop would be the optimization for each bin as we have seen that the performance of the optimization depends on the range of the optimization.

Finally, this brute force approach to optimize an estimator implies that for each survey a similar optimization procedure must be performed giving different coefficients in each case. In order to really have an impact an interesting question would be to find the *magic numbers* that gives the universal best estimator. This kind of study could be performed using techniques as principal components to find which are the leading terms that determine the optimization. It would also avoid the huge degeneracy that exist among the coefficients used here, removing just a few of them does not change the results significantly.

Conclusions

To conclude this work I present a list of the global conclusions we can get from my research during this 3 years. These enumeration includes some very preliminary cosmological results from the BOSS correlation function even if the analysis is not finished (as the survey is still taking data) and as the cosmological consequences are still being discussed within the working group.

I studied the properties of the estimators of the correlation function and the systematic and statistical uncertainties using mock catalogs. Using lognormal simulations I showed that the performance of the estimators in terms of the variance and the bias depends of the properties of the survey as the angular mask (geometry) . I showed that within the estimators commonly used in clustering analysis Hamilton and Landy-Szalay show the better properties, less bias and small variance. We generate realistic simulations that mimics the properties of the LRG sample of SDSS-DR7, survey geometry, redshift distribution, bias and mean density. With these simulations we studied the effect of the completeness of the sample and the radial selection function on the correlation function. I also studied the effect of radial weights in the correlation function.

I studied the error estimation using different approaches applied to DR7 data : log-normal simulations, bootstrap, subsample and jackknife methods. I showed that all methods produce covariance matrices that have off-diagonal elements as expected. Even though, from the test performed with the different error estimation, we found that data-based methods give noisy estimations of the covariances matrices, the extra correlation found could be generated by the correlation between the borders of the regions defining the subsamples. The jackknife is the method that presents the best properties among the data-based error methods but even it still shows some extra correlation. In general we conclude the data error methods overestimates the errors. In the case of lognormal simulations, the mean of simulations is underestimated an improvement to the input correlation function must be applied to get a mean closer to the data mean or a rescaling must be performed.

I studied the redshift space distortions in the linear regime using lognormal simulations. In particular I explored the possibility to measure β parameter from mock catalogs of galaxies using the spherical harmonics extracted from the 2d correlation function. I also explored also the estimation of the β parameter using mocks catalogs issued from N-body simulations that includes the non linear effects. I observed that the reconstruction of the beta parameter is possible in both cases. I presented some preliminary results with real data from BOSS. In this case the estimation of the β parameter must be computed from the normalized quadrupole and no conclusion could be drawned yet.

Concerning the Analysis of BOSS data I performed a complete analysis of the CMASS sample using the data available up to the summer 2011. This analysis ranges from the catalog generation up to the cosmological constrains: I studied the choices taken to construct the catalog and I explored the implications and justify the choices. I described the correction of the correlation function from angular systematics effects. I verified that the main systematic is the stellar density. I studied the stellar density in detail, separating the dependency of the mean density and stellar density in bins of galactic latitude. I found a different dependency between south a north at low galactic latitude. I explored the explanation for this systematic effect calculating the surface masked by stars and I studied its dependency with stellar density. An evolution of this dependency in terms of galactic latitude was found indicating an unknown systematic affecting south hemisphere. I tested the effect in the correlation function from others sources of systematics effects (seeing, sky, extinction, stars, air mass, dperp offset) and I implemented 2 methods (Iterative weights and Linear regression) to estimate the weights needed to correct from this systematic error. Both methods give similar results in terms of the correlation function. The main effect of correcting from angular systematics is to decrease the correlation function level globally.

To finish the conclusions concerning the correlation function from BOSS I showed that the correlations functions issued from my final catalog and corrected from the angular systematics are completely compatible with the official catalog. The results validate my pipeline, showing that I was able to reproduce all steps to do large scale analysis. This procedure permit me to understand very well the sample we are using to constraint the cosmology, in particular to understand the corrections involved in the catalog generation and in the weighting scheme.

In collaboration with A. Labatie we fitted the correlation function following the method of (Eisenstein et al., 2005) with 3 free parameters $\Omega_M h^2$, b , α . For the fits we used the covariance matrix computed from the 600 simulations PTHALOS and the official correlation function to be able to compare with the people in the working group. This analysis is very preliminary we are working in parallel with the working group to crosscheck the results and the cosmological interpretation, at this moment our results are compatible with the results of other members of the clustering group. I enumerate some preliminary conclusions.

- (a) We observe a disagreement between North and South within 2σ for the DR9. The tension of South in DR9 sample that is more than $1-\sigma$ from North seems to be less significative when including more data (i.e DR10 data). This fact suggest the South in DR9 present a large fluctuation.
- (b) The preliminary results indicates a $\Omega_m h^2 = 0.144 \pm 0.007$ and $\alpha = 0.992 \pm 0.026$. The constrains in α traduces to constrains in distance scale $D_V = 2007.80 \pm 53.15$. Our results are in agreement with previous studies.
- (c) The combined sample gives a detection with a significance of 4.6σ over non wiggles models for the combined sample. The south has added a $\Delta\chi^2 \sim 7$ versus an expected $\Delta\chi^2 \sim 4$.

Finally, I have developed a simple and straightforward way to optimize the correlation function estimator for a given survey. The preliminary results from simulations indicates a gain of the order of 20% to 40% in signal to noise and the price is to get a larger bias. The extra bias introduced can be controlled via the λ parameter to be of the order of the variance. The bias can be exactly calculated via the Monte Carlo simulations and corrected. The final goal

of this work would be to apply this optimized estimator to BOSS data. In order to validate this estimator there is still a lot of work to be done. I need to use more realistic simulations as PTHALOS and study the behavior when the non linear effects are introduced and finally apply it to N-body simulations, LasDamas CMASS. In a further stage, the study could be extended to test the consequences in terms of parameter constraints. We want to introduce FKP weights that could reduce the gain in S/N as this weighting scheme is designed to improve optimally the estimator. We want also to perform optimization for each bin as we have seen that the performance of the optimization depends on the range of the optimization. We plan also to use techniques as principal components analysis to find which are the leading terms that determine the optimization. It would also avoid the huge degeneracy that exist among the coefficients used here, removing just a few of them does not change the results significantly. We hope to converge very soon on this topic and prepare a publication.

Bibliography

- Albrecht A., Bernstein G., Cahn R., Freedman W. L., Hewitt J., Hu W., Huth J., Kamionkowski M., Kolb E. W., Knox L., Mather J. C., Staggs S. & Suntzeff N. B., 2006, *ArXiv Astrophysics e-prints*
- Andrae R., 2010, *ArXiv e-prints*
- Bahcall N. A., Cen R., Davé R., Ostriker J. P. & Yu Q., 2000, *ApJ* **541**, 1
- Bernstein G. M., 1994, *ApJ* **424**, 569
- Beutler F., Blake C., Colless M., Jones D. H., Staveley-Smith L., Campbell L., Parker Q., Saunders W. & Watson F., 2011, *MNRAS* **416**, 3017
- Bevington P. R. & Robinson D. K., 1992, *Data reduction and error analysis for the physical sciences*
- Blake C., Davis T., Poole G. B., Parkinson D., Brough S., Colless M., Contreras C., Couch W., Croom S., Drinkwater M. J., Forster K., Gilbank D., Gladders M., Glazebrook K., Jelliffe B., Jurek R. J., Li I.-H., Madore B., Martin D. C., Pimblett K., Pracy M., Sharp R., Wisnioski E., Woods D., Wyder T. K. & Yee H. K. C., 2011, *MNRAS* **415**, 2892
- Blanton M. R., Dalcanton J., Eisenstein D., Loveday J., Strauss M. A., SubbaRao M., Weinberg D. H., Anderson, Jr. J. E., Annis J., Bahcall N. A., Bernardi M., Brinkmann J., Brunner R. J., Burles S., Carey L., Castander F. J., Connolly A. J., Csabai I., Doi M., Finkbeiner D., Friedman S., Frieman J. A., Fukugita M., Gunn J. E., Hennessy G. S., Hindsley R. B., Hogg D. W., Ichikawa T., Ivezić Ž., Kent S., Knapp G. R., Lamb D. Q., Leger R. F., Long D. C., Lupton R. H., McKay T. A., Meiksin A., Merelli A., Munn J. A., Narayanan V., Newcomb M., Nichol R. C., Okamura S., Owen R., Pier J. R., Pope A., Postman M., Quinn T., Rockosi C. M., Schlegel D. J., Schneider D. P., Shimasaku K., Siegmund W. A., Smee S., Snir Y., Stoughton C., Stubbs C., Szalay A. S., Szokoly G. P., Thakar A. R., Tremonti C., Tucker D. L., Uomoto A., Vanden Berk D., Vogeley M. S., Waddell P., Yanny B., Yasuda N. & York D. G., 2001, *AJ* **121**, 2358
- Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I. & Loveday J., 2003, *AJ* **125**, 2276
- Blanton M. R., Schlegel D. J., Strauss M. A., Brinkmann J., Finkbeiner D., Fukugita M., Gunn J. E., Hogg D. W., Ivezić Ž., Knapp G. R., Lupton R. H., Munn J. A., Schneider D. P., Tegmark M. & Zehavi I., 2005, *AJ* **129**, 2562
- Bunn E. F. & Hogg D. W., 2009, *American Journal of Physics* **77**, 688
- Burles S. & Tytler D., 1998, *ApJ* **507**, 732
- Cabré A. & Gaztañaga E., 2009, *MNRAS* **393**, 1183
- Cannon R., Drinkwater M., Edge A., Eisenstein D., Nichol R., Outram P., Pimblett K., de Propris R., Roseboom I., Wake D., Allen P., Bland-Hawthorn J., Bridges T., Carson D.,

- Chiu K., Colless M., Couch W., Croom S., Driver S., Fine S., Hewett P., Loveday J., Ross N., Sadler E. M., Shanks T., Sharp R., Smith J. A., Stoughton C., Weibacher P., Brunner R. J., Meiksin A. & Schneider D. P., 2006, *MNRAS* **372**, 425
- Carlson J., White M. & Padmanabhan N., 2009, *Phys. Rev. D* **80**(4), 043531
- Chodorowski M. J., 2011, *MNRAS* **413**, 585
- Clowe D., Bradač M., Gonzalez A. H., Markevitch M., Randall S. W., Jones C. & Zaritsky D., 2006, *ApJ* **648**, L109
- Cole S., Percival W. J., Peacock J. A., Norberg P., Baugh C. M., Frenk C. S., Baldry I., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., Collins C., Couch W., Cross N. J. G., Dalton G., Eke V. R., De Propris R., Driver S. P., Efstathiou G., Ellis R. S., Glazebrook K., Jackson C., Jenkins A., Lahav O., Lewis I., Lumsden S., Maddox S., Madgwick D., Peterson B. A., Sutherland W. & Taylor K., 2005, *MNRAS* **362**, 505
- Coles P. & Jones B., 1991, *MNRAS* **248**, 1
- Cooray A. & Sheth R., 2002, *Phys. Rep.* **372**, 1
- Corbelli E. & Salucci P., 2000, *MNRAS* **311**, 441
- Crocce M. & Scoccimarro R., 2006, *Phys. Rev. D* **73**(6), 063519
- Crocce M. & Scoccimarro R., 2008, *Phys. Rev. D* **77**(2), 023533
- Davis M. & Peebles P. J. E., 1983, *ApJ* **267**, 465
- Dicke R. H., Peebles P. J. E., Roll P. G. & Wilkinson D. T., 1965, *ApJ* **142**, 414
- Dodelson S., 2003, *Modern cosmology*, Academic Press, San Diego, Calif
- Eisenstein D. J., Annis J., Gunn J. E., Szalay A. S., Connolly A. J., Nichol R. C., Bahcall N. A., Bernardi M., Burles S., Castander F. J., Fukugita M., Hogg D. W., Ivezić Z., Knapp G. R., Lupton R. H., Narayanan V., Postman M., Reichart D. E., Richmond M., Schneider D. P., Schlegel D. J., Strauss M. A., SubbaRao M., Tucker D. L., Vanden Berk D., Vogeley M. S., Weinberg D. H. & Yanny B., 2001, *AJ* **122**, 2267
- Eisenstein D. J. & Hu W., 1998a, *ApJ* **496**, 605
- Eisenstein D. J. & Hu W., 1998b, *ApJ* **496**, 605
- Eisenstein D. J., Seo H.-J. & White M., 2007, *ApJ* **664**, 660
- Eisenstein D. J., Weinberg D. H., Agol E., Aihara H., Allende Prieto C., Anderson S. F., Arns J. A., Aubourg É., Bailey S., Balbinot E. & et al., 2011b, *AJ* **142**, 72
- Eisenstein D. J., Weinberg D. H., Agol E., Aihara H., Allende Prieto C., Anderson S. F., Arns J. A., Aubourg E., Bailey S., Balbinot E. & et al., 2011a, *ArXiv e-prints*
- Eisenstein D. J., Zehavi I., Hogg D. W., Scoccimarro R., Blanton M. R., Nichol R. C., Scranton R., Seo H., Tegmark M., Zheng Z., Anderson S. F., Annis J., Bahcall N., Brinkmann J., Burles S., Castander F. J., Connolly A., Csabai I., Doi M., Fukugita M., Frieman J. A., Glazebrook K., Gunn J. E., Hendry J. S., Hennessy G., Ivezić Z., Kent S., Knapp G. R., Lin H., Loh Y., Lupton R. H., Margon B., McKay T. A., Meiksin A., Munn J. A., Pope A., Richmond M. W., Schlegel D., Schneider D. P., Shimasaku K., Stoughton C., Strauss M. A., SubbaRao M., Szalay A. S., Szapudi I., Tucker D. L., Yanny B. & York D. G., 2005, *ApJ* **633**, 560
- Feldman H. A., Kaiser N. & Peacock J. A., 1994, *ApJ* **426**, 23
- Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K. & Schneider D. P., 1996, *AJ* **111**, 1748
- Goobar A. & Leibundgut B., 2011, *ArXiv e-prints*
- Gunn J. & Weinberg D., 1995, dans S. J. Maddox & A. Aragon-Salamanca (ed.), *Wide Field*

- Spectroscopy and the Distant Universe*, pp 3–+
- Gunn J. E., Carr M., Rockosi C., Sekiguchi M., Berry K., Elms B., de Haas E., Ivezić Ž., Knapp G., Lupton R., Pauls G., Simcoe R., Hirsch R., Sanford D., Wang S., York D., Harris F., Annis J., Bartozek L., Boroski W., Bakken J., Haldeman M., Kent S., Holm S., Holmgren D., Petravick D., Prosapio A., Rechenmacher R., Doi M., Fukugita M., Shimasaku K., Okada N., Hull C., Siegmund W., Mannery E., Blouke M., Heidtman D., Schneider D., Lucinio R. & Brinkman J., 1998, *AJ* **116**, 3040
- Gunn J. E., Siegmund W. A., Mannery E. J., Owen R. E., Hull C. L., Leger R. F., Carey L. N., Knapp G. R., York D. G., Boroski W. N., Kent S. M., Lupton R. H., Rockosi C. M., Evans M. L., Waddell P., Anderson J. E., Annis J., Barentine J. C., Bartoszek L. M., Bastian S., Bracker S. B., Brewington H. J., Briegel C. I., Brinkmann J., Brown Y. J., Carr M. A., Czarapata P. C., Drennan C. C., Dombeck T., Federwitz G. R., Gillespie B. A., Gonzales C., Hansen S. U., Harvanek M., Hayes J., Jordan W., Kinney E., Klaene M., Kleinman S. J., Kron R. G., Kresinski J., Lee G., Limmongkol S., Lindenmeyer C. W., Long D. C., Loomis C. L., McGehee P. M., Mantsch P. M., Neilsen, Jr. E. H., Neswold R. M., Newman P. R., Nitta A., Peoples, Jr. J., Pier J. R., Prieto P. S., Prosapio A., Rivetta C., Schneider D. P., Snedden S. & Wang S., 2006, *AJ* **131**, 2332
- Hamilton A. J. S., 1992, *ApJ* **385**, L5
- Hamilton A. J. S., 1993, *ApJ* **417**, 19
- Hamilton A. J. S., 1998, dans D. Hamilton (ed.), *The Evolving Universe*, Vol. 231 of *Astrophysics and Space Science Library*, pp 185–+
- Hamilton A. J. S. & Tegmark M., 2004, *MNRAS* **349**, 115
- Hewett P. C., 1982, *MNRAS* **201**, 867
- Hubble E., 1929, *Proceedings of the National Academy of Science* **15**, 168
- Hubble E. & Humason M. L., 1931, *ApJ* **74**, 43
- Kazin E. A., Blanton M. R., Scoccimarro R., McBride C. K., Berlind A. A., Bahcall N. A., Brinkmann J., Czarapata P., Frieman J. A., Kent S. M., Schneider D. P. & Szalay A. S., 2010, *ApJ* **710**, 1444
- Kerscher M., Szapudi I. & Szalay A. S., 2000, *ApJ* **535**, L13
- Kirkman D., Tytler D., Suzuki N., O’Meara J. M. & Lubin D., 2003, *ApJS* **149**, 1
- Labatie A., Starck J.-L. & Lachièze-Rey M., 2011, *ArXiv e-prints*
- Labatie A., Starck J.-L., Lachièze-Rey M. & Arnalte-Mur P., 2010, *ArXiv e-prints*
- Landy S. D. & Szalay A. S., 1993, *ApJ* **412**, 64
- Larson D., Dunkley J., Hinshaw G., Komatsu E., Nolta M. R., Bennett C. L., Gold B., Halpern M., Hill R. S., Jarosik N., Kogut A., Limon M., Meyer S. S., Odegard N., Page L., Smith K. M., Spergel D. N., Tucker G. S., Weiland J. L., Wollack E. & Wright E. L., 2011, *ApJS* **192**, 16
- Lasker B. M., Lattanzi M. G., McLean B. J., Bucciarelli B., Drimmel R., Garcia J., Greene G., Guglielmetti F., Hanley C., Hawkins G., Laidler V. G., Loomis C., Meakes M., Mignani R., Morbidelli R., Morrison J., Pannunzio R., Rosenberg A., Sarasso M., Smart R. L., Spagna A., Sturch C. R., Volpicelli A., White R. L., Wolfe D. & Zacchei A., 2008, *AJ* **136**, 735
- Lewis A., Challinor A. & Lasenby A., 2000, *Astrophys. J.* **538**, 473
- Li W., Chornock R., Leaman J., Filippenko A. V., Poznanski D., Wang X., Ganeshalingam M. & Mannucci F., 2011, *MNRAS* **412**, 1473
- Linder E. V., 2005, *Phys. Rev. D* **72**(4), 043529

- Lupton R., Gunn J. E., Ivezić Z., Knapp G. R. & Kent S., 2001, dans F. R. Harnden Jr., F. A. Primini, & H. E. Payne (ed.), *Astronomical Data Analysis Software and Systems X*, Vol. 238 of *Astronomical Society of the Pacific Conference Series*, pp 269–+
- Madrid J. P. & Macchetto D., 2009, *ArXiv e-prints*
- Maraston C., 2005, *MNRAS* **362**, 799
- Markevitch M., Gonzalez A. H., Clowe D., Vikhlinin A., Forman W., Jones C., Murray S. & Tucker W., 2004, *ApJ* **606**, 819
- Mather J. C., Cheng E. S., Eplee, Jr. R. E., Isaacman R. B., Meyer S. S., Shafer R. A., Weiss R., Wright E. L., Bennett C. L., Boggess N. W., Dwek E., Gulkis S., Hauser M. G., Janssen M., Kelsall T., Lubin P. M., Moseley, Jr. S. H., Murdock T. L., Silverberg R. F., Smoot G. F. & Wilkinson D. T., 1990, *ApJ* **354**, L37
- Matsubara T., 2008, *Phys. Rev. D* **77(6)**, 063530
- McDonald P., 2007, *Phys. Rev. D* **75(4)**, 043514
- Meiksin A., White M. & Peacock J. A., 1999, *MNRAS* **304**, 851
- Netterfield C. B., Ade P. A. R., Bock J. J., Bond J. R., Borrill J., Boscaleri A., Coble K., Contaldi C. R., Crill B. P., de Bernardis P., Farese P., Ganga K., Giacometti M., Hivon E., Hristov V. V., Iacoangeli A., Jaffe A. H., Jones W. C., Lange A. E., Martinis L., Masi S., Mason P., Mauskopf P. D., Melchiorri A., Montroy T., Pascale E., Piacentini F., Pogosyan D., Pongetti F., Prunet S., Romeo G., Ruhl J. E. & Scaramuzzi F., 2002, *ApJ* **571**, 604
- Noh Y., White M. & Padmanabhan N., 2009, *Phys. Rev. D* **80(12)**, 123501
- Padmanabhan N., Schlegel D. J., Finkbeiner D. P., Barentine J. C., Blanton M. R., Brewington H. J., Gunn J. E., Harvanek M., Hogg D. W., Ivezić Ž., Johnston D., Kent S. M., Kleinman S. J., Knapp G. R., Krzesinski J., Long D., Neilsen, Jr. E. H., Nitta A., Loomis C., Lupton R. H., Roweis S., Snedden S. A., Strauss M. A. & Tucker D. L., 2008, *ApJ* **674**, 1217
- Padmanabhan N., White M. & Cohn J. D., 2009, *Phys. Rev. D* **79(6)**, 063523
- Pápai P. & Szapudi I., 2008, *MNRAS* **389**, 292
- Peebles P. J. E. & Hauser M. G., 1974, *ApJS* **28**, 19
- Penzias A. A. & Wilson R. W., 1965, *ApJ* **142**, 419
- Percival W. J., Cole S., Eisenstein D. J., Nichol R. C., Peacock J. A., Pope A. C. & Szalay A. S., 2007, *MNRAS* **381**, 1053
- Percival W. J., Reid B. A., Eisenstein D. J., Bahcall N. A., Budavari T., Frieman J. A., Fukugita M., Gunn J. E., Ivezić Ž., Knapp G. R., Kron R. G., Loveday J., Lupton R. H., McKay T. A., Meiksin A., Nichol R. C., Pope A. C., Schlegel D. J., Schneider D. P., Spergel D. N., Stoughton C., Strauss M. A., Szalay A. S., Tegmark M., Vogeley M. S., Weinberg D. H., York D. G. & Zehavi I., 2010, *MNRAS* **401**, 2148
- Percival W. J., Sutherland W., Peacock J. A., Baugh C. M., Bland-Hawthorn J., Bridges T., Cannon R., Cole S., Colless M., Collins C., Couch W., Dalton G., De Propriis R., Driver S. P., Efstathiou G., Ellis R. S., Frenk C. S., Glazebrook K., Jackson C., Lahav O., Lewis I., Lumsden S., Maddox S., Moody S., Norberg P., Peterson B. A. & Taylor K., 2002, *MNRAS* **337**, 1068
- Perlmutter S., Aldering G., Goldhaber G., Knop R. A., Nugent P., Castro P. G., Deustua S., Fabbro S., Goobar A., Groom D. E., Hook I. M., Kim A. G., Kim M. Y., Lee J. C., Nunes N. J., Pain R., Pennypacker C. R., Quimby R., Lidman C., Ellis R. S., Irwin M., McMahon R. G., Ruiz-Lapuente P., Walton N., Schaefer B., Boyle B. J., Filippenko A. V., Matheson T., Fruchter A. S., Panagia N., Newberg H. J. M., Couch W. J. & The Supernova

- Cosmology Project, 1999, *ApJ* **517**, 565
- Pier J. R., Munn J. A., Hindsley R. B., Hennessy G. S., Kent S. M., Lupton R. H. & Ivezić Ž., 2003, *AJ* **125**, 1559
- Pietroni M., 2008, **10**, 36
- Pons-Bordería M.-J., Martínez V. J., Stoyan D., Stoyan H. & Saar E., 1999, *ApJ* **523**, 480
- Press W. H., Teukolsky S. A., Vetterling W. T. & Flannery B. P., 1992, *Numerical recipes in FORTRAN. The art of scientific computing*
- Proposal S.-C., 2008, *SDSS-III: Massive Spectroscopic Surveys of Distant Universe, the Milky Way, and Extra-Solar Planetary Systems*
- Prunet S., Pichon C., Aubert D., Pogosyan D., Teyssier R. & Gottloeber S., 2008, *ApJS* **178**, 179
- Rauch M., Miralda-Escude J., Sargent W. L. W., Barlow T. A., Weinberg D. H., Hernquist L., Katz N., Cen R. & Ostriker J. P., 1997, *ApJ* **489**, 7
- Refregier A., Amara A., Kitching T. D. & Rassat A., 2011, *A&A* **528**, A33
- Richard K., Gunn J. E., Weinberg D. H., Boroski W. N. & Evans M. L., 2008, *Final Report to the Alfred P. Sloan Foundation*
- Richards G. T., Myers A. D., Gray A. G., Riegel R. N., Nichol R. C., Brunner R. J., Szalay A. S., Schneider D. P. & Anderson S. F., 2009, *ApJS* **180**, 67
- Riess A. G., Filippenko A. V., Challis P., Clocchiatti A., Diercks A., Garnavich P. M., Gilliland R. L., Hogan C. J., Jha S., Kirshner R. P., Leibundgut B., Phillips M. M., Reiss D., Schmidt B. P., Schommer R. A., Smith R. C., Spyromilio J., Stubbs C., Suntzeff N. B. & Tonry J., 1998, *AJ* **116**, 1009
- Riess A. G., Macri L., Casertano S., Lampeitl H., Ferguson H. C., Filippenko A. V., Jha S. W., Li W. & Chornock R., 2011, *ApJ* **730**, 119
- Ross A. J., Percival W. J., Sanchez A. G., Samushia L., Ho S., Kazin E., Manera M., Reid B., White M., Tojeiro R., McBride C. K., Xu X., Wake D. A., Strauss M. A., Montesano F., Swanson M. E. C., Bailey S., Bolton A. S., Montero Dorta A., Eisenstein D. J., Guo H., Hamilton J.-C., Nichol R. C., Padmanabhan N., Prada F., Schlegel D. J., Vargas Magana M., Zehavi I., Blanton M., Bizyaev D., Brewington H., Cuesta A. J., Malanushenko E., Malanushenko V., Oravetz D., Parejko J., Pan K., Schneider Alaina Shelden D. P., Simmons A., Snedden S. & Zhao G.-b., 2012, *ArXiv e-prints*
- Rubin V. C. & Ford, Jr. W. K., 1970, *ApJ* **159**, 379
- Rubin V. C., Ford W. K. J. & Thonnard N., 1980, *ApJ* **238**, 471
- Samushia L., Percival W. J. & Raccanelli A., 2011, *ArXiv e-prints*
- Sánchez A. G., Crocce M., Cabré A., Baugh C. M. & Gaztañaga E., 2009, *MNRAS* **400**, 1643
- Schlegel D. J., Finkbeiner D. P. & Davis M., 1998, *ApJ* **500**, 525
- Scoccimarro R., 2004, *Phys. Rev. D* **70(8)**, 083007
- Scoccimarro R. & Sheth R. K., 2002, *MNRAS* **329**, 629
- Scranton R., Johnston D., Dodelson S., Frieman J. A., Connolly A., Eisenstein D. J., Gunn J. E., Hui L., Jain B., Kent S., Loveday J., Narayanan V., Nichol R. C., O'Connell L., Scoccimarro R., Sheth R. K., Stebbins A., Strauss M. A., Szalay A. S., Szapudi I., Tegmark M., Vogeley M., Zehavi I., Annis J., Bahcall N. A., Brinkman J., Csabai I., Hindsley R., Ivezić Z., Kim R. S. J., Knapp G. R., Lamb D. Q., Lee B. C., Lupton R. H., McKay T., Munn J., Peoples J., Pier J., Richards G. T., Rockosi C., Schlegel D., Schneider D. P., Stoughton C., Tucker D. L., Yanny B. & York D. G., 2002, *ApJ* **579**, 48

- Seljak U. & Zaldarriaga M., 1996, ApJ **469**, 437
- Seo H.-J. & Eisenstein D. J., 2005, ApJ **633**, 575
- Shectman S. A., Landy S. D., Oemler A., Tucker D. L., Lin H., Kirshner R. P. & Schechter P. L., 1996, ApJ **470**, 172
- Skrutskie M. F., Cutri R. M., Stiening R., Weinberg M. D., Schneider S., Carpenter J. M., Beichman C., Capps R., Chester T., Elias J., Huchra J., Liebert J., Lonsdale C., Monet D. G., Price S., Seitzer P., Jarrett T., Kirkpatrick J. D., Gizis J. E., Howard E., Evans T., Fowler J., Fullmer L., Hurt R., Light R., Kopan E. L., Marsh K. A., McCallon H. L., Tam R., Van Dyk S. & Wheelock S., 2006, AJ **131**, 1163
- Smith R. E., Peacock J. A., Jenkins A., White S. D. M., Frenk C. S., Pearce F. R., Thomas P. A., Efstathiou G. & Couchman H. M. P., 2003, MNRAS **341**, 1311
- Smoot G. F., Bennett C. L., Kogut A., Wright E. L., Aymon J., Boggess N. W., Cheng E. S., de Amici G., Gulkis S., Hauser M. G., Hinshaw G., Jackson P. D., Janssen M., Kaita E., Kelsall T., Keegstra P., Lineweaver C., Loewenstein K., Lubin P., Mather J., Meyer S. S., Moseley S. H., Murdock T., Rokke L., Silverberg R. F., Tenorio L., Weiss R. & Wilkinson D. T., 1992, ApJ **396**, L1
- Snethlage M., 2000, *ArXiv Mathematics e-prints*
- Spergel D. N., Verde L., Peiris H. V., Komatsu E., Nolta M. R., Bennett C. L., Halpern M., Hinshaw G., Jarosik N., Kogut A., Limon M., Meyer S. S., Page L., Tucker G. S., Weiland J. L., Wollack E. & Wright E. L., 2003, ApJS **148**, 175
- Stoughton C., Lupton R. H., Bernardi M., Blanton M. R., Burles S., Castander F. J., Connolly A. J., Eisenstein D. J., Frieman J. A., Hennessy G. S., Hindsley R. B., Ivezić Ž., Kent S., Kunszt P. Z., Lee B. C., Meiksin A., Munn J. A., Newberg H. J., Nichol R. C., Nicinski T., Pier J. R., Richards G. T., Richmond M. W., Schlegel D. J., Smith J. A., Strauss M. A., SubbaRao M., Szalay A. S., Thakar A. R., Tucker D. L., Vanden Berk D. E., Yanny B., Adelman J. K., Anderson, Jr. J. E., Anderson S. F., Annis J., Bahcall N. A., Bakken J. A., Bartelmann M., Bastian S., Bauer A., Berman E., Böhringer H., Boroski W. N., Bracker S., Briegel C., Briggs J. W., Brinkmann J., Brunner R., Carey L., Carr M. A., Chen B., Christian D., Colestock P. L., Crocker J. H., Csabai I., Czarapata P. C., Dalcanton J., Davidsen A. F., Davis J. E., Dehnen W., Dodelson S., Doi M., Dombeck T., Donahue M., Ellman N., Elms B. R., Evans M. L., Eyer L., Fan X., Federwitz G. R., Friedman S., Fukugita M., Gal R., Gillespie B., Glazebrook K., Gray J., Grebel E. K., Greenawalt B., Greene G., Gunn J. E., de Haas E., Haiman Z., Haldeman M., Hall P. B., Hamabe M., Hansen B., Harris F. H., Harris H., Harvanek M., Hawley S. L., Hayes J. J. E., Heckman T. M., Helmi A., Henden A., Hogan C. J., Hogg D. W., Holmgren D. J., Holtzman J., Huang C., Hull C., Ichikawa S., Ichikawa T., Johnston D. E., Kauffmann G., Kim R. S. J., Kimball T., Kinney E., Klaene M., Kleinman S. J., Klypin A., Knapp G. R., Korienek J., Krolik J., Kron R. G., Krzesiński J., Lamb D. Q., Leger R. F., Limmongkol S., Lindenmeyer C., Long D. C., Loomis C., Loveday J., MacKinnon B., Mannery E. J., Mantsch P. M., Margon B., McGehee P., McKay T. A., McLean B., Menou K., Merelli A., Mo H. J., Monet D. G., Nakamura O., Narayanan V. K., Nash T., Neilsen, Jr. E. H., Newman P. R., Nitta A., Odenkirchen M., Okada N., Okamura S., Ostriker J. P., Owen R., Pauls A. G., Peoples J., Peterson R. S., Petravick D., Pope A., Pordes R., Postman M., Prosapio A., Quinn T. R., Rechenmacher R., Rivetta C. H., Rix H., Rockosi C. M., Rosner R., Ruthmansdorfer K., Sandford D., Schneider D. P., Scranton R., Sekiguchi M., Sergey G., Sheth R., Shimasaku

- K., Smee S., Snedden S. A., Stebbins A., Stubbs C., Szapudi I., Szkody P., Szokoly G. P., Tabachnik S., Tsvetanov Z., Uomoto A., Vogeley M. S., Voges W., Waddell P., Walterbos R., Wang S., Watanabe M., Weinberg D. H., White R. L., White S. D. M., Wilhite B., Wolfe D., Yasuda N., York D. G., Zehavi I. & Zheng W., 2002, *AJ* **123**, 485
- Strauss M. A. & Willick J. A., 1995, *Phys. Rep.* **261**, 271
- Valageas P., 2007, *A&A* **465**, 725
- Verde L., 2010, dans G. Wolschin (ed.), *Lecture Notes in Physics, Berlin Springer Verlag*, Vol. 800 of *Lecture Notes in Physics, Berlin Springer Verlag*, pp 147–177
- Verde L., Heavens A. F., Percival W. J., Matarrese S., Baugh C. M., Bland-Hawthorn J., Bridges T., Cannon R., Cole S., Colless M., Collins C., Couch W., Dalton G., De Propris R., Driver S. P., Efstathiou G., Ellis R. S., Frenk C. S., Glazebrook K., Jackson C., Lahav O., Lewis I., Lumsden S., Maddox S., Madgwick D., Norberg P., Peacock J. A., Peterson B. A., Sutherland W. & Taylor K., 2002, *MNRAS* **335**, 432
- Wall J. V. & Jenkins C. R., 2003, *Practical Statistics for Astronomers*
- White M., 2005, *Astroparticle Physics* **24**, 334
- White M., Blanton M., Bolton A., Schlegel D., Tinker J., Berlind A., da Costa L., Kazin E., Lin Y.-T., Maia M., McBride C. K., Padmanabhan N., Parejko J., Percival W., Prada F., Ramos B., Sheldon E., de Simoni F., Skibba R., Thomas D., Wake D., Zehavi I., Zheng Z., Nichol R., Schneider D. P., Strauss M. A., Weaver B. A. & Weinberg D. H., 2011, *ApJ* **728**, 126
- Yèche C., Petitjean P., Rich J., Aubourg E., Busca N., Hamilton J., Le Goff J., Paris I., Peirani S., Pichon C., Rollinde E. & Vargas-Magaña M., 2010, *A&A* **523**, A14+
- York D. G., Adelman J., Anderson, Jr. J. E., Anderson S. F., Annis J., Bahcall N. A., Bakken J. A., Barkhouser R., Bastian S., Berman E., Boroski W. N., Bracker S., Briegel C., Briggs J. W., Brinkmann J., Brunner R., Burles S., Carey L., Carr M. A., Castander F. J., Chen B., Colestock P. L., Connolly A. J., Crocker J. H., Csabai I., Czarapata P. C., Davis J. E., Doi M., Dombeck T., Eisenstein D., Ellman N., Elms B. R., Evans M. L., Fan X., Federwitz G. R., Fiscelli L., Friedman S., Frieman J. A., Fukugita M., Gillespie B., Gunn J. E., Gurbani V. K., de Haas E., Haldeman M., Harris F. H., Hayes J., Heckman T. M., Hennessy G. S., Hindsley R. B., Holm S., Holmgren D. J., Huang C., Hull C., Husby D., Ichikawa S., Ichikawa T., Ivezić Ž., Kent S., Kim R. S. J., Kinney E., Klaene M., Kleinman A. N., Kleinman S., Knapp G. R., Korienek J., Kron R. G., Kunszt P. Z., Lamb D. Q., Lee B., Leger R. F., Limmongkol S., Lindenmeyer C., Long D. C., Loomis C., Loveday J., Lucinio R., Lupton R. H., MacKinnon B., Mannery E. J., Mantsch P. M., Margon B., McGehee P., McKay T. A., Meiksin A., Merelli A., Monet D. G., Munn J. A., Narayanan V. K., Nash T., Neilsen E., Neswold R., Newberg H. J., Nichol R. C., Nicinski T., Nonino M., Okada N., Okamura S., Ostriker J. P., Owen R., Pauls A. G., Peoples J., Peterson R. L., Petravick D., Pier J. R., Pope A., Pordes R., Prosapio A., Rechenmacher R., Quinn T. R., Richards G. T., Richmond M. W., Rivetta C. H., Rockosi C. M., Ruthmansdorfer K., Sandford D., Schlegel D. J., Schneider D. P., Sekiguchi M., Sergey G., Shimasaku K., Siegmund W. A., Smee S., Smith J. A., Snedden S., Stone R., Stoughton C., Strauss M. A., Stubbs C., SubbaRao M., Szalay A. S., Szapudi I., Szokoly G. P., Thakar A. R., Tremonti C., Tucker D. L., Uomoto A., Vanden Berk D., Vogeley M. S., Waddell P., Wang S., Watanabe M., Weinberg D. H., Yanny B. & Yasuda N., 2000, *AJ* **120**, 1579
- Zacharias N., Finch C., Girard T., Hambly N., Wycoff G., Zacharias M. I., Castillo D., Corbin

- T., DiVittorio M., Dutta S., Gaume R., Gauss S., Germain M., Hall D., Hartkopf W., Hsu D., Holdenried E., Makarov V., Martinez M., Mason B., Monet D., Rafferty T., Rhodes A., Siemers T., Smith D., Tilleman T., Urban S., Wieder G., Winter L. & Young A., 2010, *AJ* **139**, 2184
- Zacharias N., Monet D. G., Levine S. E., Urban S. E., Gaume R. & Wycoff G. L., 2004, dans *American Astronomical Society Meeting Abstracts*, Vol. 36 of *Bulletin of the American Astronomical Society*, pp 1418–+
- Zaldarriaga M., 1998, *Fluctuations in the Cosmic Microwave Background* %J *Ph.D. Thesis*, M. I.T astro-ph/9806122., *Thèse de doctorat*, , M. I.T astro-ph/9806122., (1998)
- Zehavi I., Blanton M. R., Frieman J. A., Weinberg D. H., Mo H. J., Strauss M. A., Anderson S. F., Annis J., Bahcall N. A., Bernardi M., Briggs J. W., Brinkmann J., Burles S., Carey L., Castander F. J., Connolly A. J., Csabai I., Dalcanton J. J., Dodelson S., Doi M., Eisenstein D., Evans M. L., Finkbeiner D. P., Friedman S., Fukugita M., Gunn J. E., Hennessy G. S., Hindsley R. B., Ivezić Ž., Kent S., Knapp G. R., Kron R., Kunszt P., Lamb D. Q., Leger R. F., Long D. C., Loveday J., Lupton R. H., McKay T., Meiksin A., Merrelli A., Munn J. A., Narayanan V., Newcomb M., Nichol R. C., Owen R., Peoples J., Pope A., Rockosi C. M., Schlegel D., Schneider D. P., Scoccimarro R., Sheth R. K., Siegmund W., Smee S., Snir Y., Stebbins A., Stoughton C., SubbaRao M., Szalay A. S., Szapudi I., Tegmark M., Tucker D. L., Uomoto A., Vanden Berk D., Vogeley M. S., Waddell P., Yanny B. & York D. G., 2002, *ApJ* **571**, 172
- Zehavi I., Eisenstein D. J., Nichol R. C., Blanton M. R., Hogg D. W., Brinkmann J., Loveday J., Meiksin A., Schneider D. P. & Tegmark M., 2005, *ApJ* **621**, 22
- Zwicky F., 1937, *ApJ* **86**, 217