COSMIC STATISTICS ON LINEAR SCALES

Péter Pápai

A dissertation submitted to the graduate division of the University of Hawai'i at Mānoa
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Physics

March 2011

Dissertation Committee:

John Learned, Chairperson
István Szapudi
Xerxes Tata
Peter Gorham
John Mahoney

Acknowledgements

I would like to express my gratitude to my supervisor, István, who supported me financially during the past 4-5 years; to my friends, who made my stay in Hawaii a long vacation, even if it hindered research from time to time; to Yanisa, who made me think of the future, despite the ideal conditions of the present; to my family, without whom my trip to Hawaii would not have been possible.

Abstract

We use the formalism of Szapudi (2004) to derive full explicit expressions for the linear two-point correlation function, including redshift space distortions and large angle effects. We take into account a non-perturbative geometric term in the Jacobian, which is still linear in terms of the dynamics. This term had been identified previously (Kaiser, 1987; Hamilton and Culhane, 1996), but has been neglected in all subsequent explicit calculations of the linear redshift space two-point correlation function. Our results represent a significant correction to previous explicit expressions and are in excellent agreement with our measurements in the Hubble Volume Simulation.

We measure the matter probability distribution function (PDF) via counts in cells in a volume limited subsample of the Sloan Digital Sky Survey (SDSS) Luminous Red Galaxy Catalog on scales from $30\ h^{-1}$Mpc to $150\ h^{-1}$Mpc and estimate the linear integrated Sachs–Wolfe (ISW) effect produced by supervoids and superclusters in the tail of the PDF. We characterize the PDF by the variance, S_3, and S_4, and study in simulations the systematic effects due to finite volume, survey shape and redshift distortion. We compare our measurement to the prediction of ΛCDM with linear bias and find a good agreement. We use the moments to approximate the tail of the PDF with analytic functions. A simple Gaussian model for the superstructures appears to be consistent with the claim by Granett *et al.* (2008) that density fluctuations on $100\ h^{-1}$Mpc scales produce hot and cold spots with $\Delta T \approx 10\mu K$ on the cosmic microwave background.

We calculate the full density and ISW profiles of spherical superstructures. We find that the Gaussian assumptions capable of describing N-body simulations and simulated ISW maps remarkably well on large scales. We construct an ISW map based on locations of superstructures identified previously in the SDSS Luminous Red Galaxy sample. A matched filter analysis of the cosmic microwave background confirms a signal at the $3.2-\sigma$ confidence level and estimates the radius of the underlying structures to be $55 \pm 28 \mathrm{h}^{-1}\mathrm{Mpc}$. The amplitude of the signal, however, is $2-\sigma$ higher than ΛCDM predictions.

Table of Contents

List of Tables

List of Figures

Chapter 1

The largest scales in cosmology

1.1 The cosmological principle

Our intuition about the world has been contradicted on countless occasions since the earliest rigorous scientific observations and experiments. Simply looking further than ever before can have the most profound effect on human knowledge and perception of reality. The discovery of the moons of Jupiter by Galileo provided an irrefutable argument against the central role of the Earth in the mechanics of the universe. Similar significance can be attributed to the observations of Edwin Hubble, who discovered that distances in the universe grew (Hubble, 1929); or to Penzias and Wilson (1965), who first detected the single most important evidence for the Big Bang, the cosmic microwave background radiation.

Owing to observational evidence, modern cosmology exchanged intuition for the cosmological principles: there is no special location in the universe; and there are no special directions. A metric compatible with this is the Friedmann-Robertson-Walker metric. This gives a general parametrization for an ideal, homogeneous and isotropic

universe (see, e.g.,Landau and Lifshitz, 1980):

$$ds^2 = c^2dt^2 - a(t)^2(dr^2 + S_k(r)^2 d\Omega), \tag{1.1}$$

with

$$S_k(r) = \begin{cases} \sqrt{k}^{-1} sin(r\sqrt{k}) & k > 0 \\ r & k = 0 \\ \sqrt{|k|}^{-1} sinh(r\sqrt{|k|}) & k < 0 \end{cases} \tag{1.2}$$

k is called the curvature and $a(t)$ the scaling factor with $a(0) = 1$.

Einstein's equations in this case are reduced to the Friedmann equations which describe the evolution of the parameters of the metric and the physical properties of the substance that fills the universe (see, e.g., Dodelson, 2003):

$$(\frac{\dot{a}}{a})^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2}, \tag{1.3}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + \frac{3p}{c^2}), \tag{1.4}$$

where ρ is the density and p is the pressure. Since this is still underdetermined for $a(t), \rho(t)$, and $p(t)$, in addition, one has to assume an equation of state, e.g.,

$$p = w\rho. \tag{1.5}$$

Obviously, the universe is not homogeneous or isotropic, but as has been justified, the Friedmann equations can form the basis of a perturbative approach for better understanding the clustering of matter.

1.2 The cosmic microwave background radiation

According to Hubble's observation, the universe is expanding. Rewinding the thermal history of such a universe, one encounters epochs with different prevailing physical processes. The change can be fast, phase transition like. The latest, the recombination, the first formation of neutral hydrogens from free protons and electrons, occurred when the temperature of the universe was approximately $3000K$, resulting in the decoupling of photons from matter. Electromagnetic radiation from this epoch is practically unaltered and carries information about the power spectrum of the gravitational potential in the early universe. The existence of the cosmic microwave background (CMB) radiation did not surprise many but it being highly isotropic provided the scientific community with a challenge. Regardless of the origin of the CMB, its fluctuations are small: $\Delta T/T \sim 10^{-5}$ (Bennett *et al.*, 2003); and consistent with Gaussian distribution(Komatsu *et al.*, 2005). This implies that the early universe at recombination was close to homogeneous and perturbation theory can be used to calculate its evolution.

In addition to being the echo of the primordial universe, the CMB acts as a device which explores the space between the last scattering surface and our detector. The CMB only went through subtle changes since decoupling; thus, in order to exploit the secondary anisotropies, the latest instruments and data reduction techniques are required.

Later, in Section 1.4 the results of past and prospects of future CMB measurements are discussed in detail.

1.3 The hydrodynamical model of clustering

The importance of radiation in structure formation decreased dramatically after decoupling. The driving force of clustering is gravity. In the following we restrict ourselves

to the discussion of the clustering of cold dark matter (CDM), which emerged as the winner among the possible scenarios to explain observations. See Bernardeau *et al.* (2002) for a comprehensive review of what is sketched in the following. The CDM is thought of as a collection of non-relativistic invisible particles that interact purely through gravity. The corresponding dynamics are within the regime of Newtonian gravity. In the expanding universe as background, physical coordinates are usually transformed into their comoving forms: $r = a(\tau)x$ and $dt = a(\tau)d\tau$. Where x and τ are the comoving distance and the conformal time. In these coordinates the dynamical quantities of interest are the density contrast $\delta(x, \tau)$, the peculiar velocity $u(x, \tau)$, and the cosmological potential $\Phi(x, \tau)$:

$$\rho(x, \tau) = \overline{\rho}(\tau)(1 + \delta(x, \tau)), \tag{1.6}$$

$$v(x, \tau) = \mathcal{H}(\tau)x + u(x, \tau), \tag{1.7}$$

$$\phi(x, \tau) = -\frac{d\mathcal{H}}{d\tau}x^2 + \Phi(x, \tau), \tag{1.8}$$

where $\mathcal{H} = \dot{a}(\tau)/a(\tau)$. With these variables the Poisson equation takes the following form:

$$\nabla^2\Phi(x, \tau) = \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\delta(x, \tau), \tag{1.9}$$

where $\Omega_m(\tau)$ denotes the fraction of matter in the total density. To arrive at this equation one has to make assumptions about the equation of state parameter, w in Equation (1.5) for the components that fill the universe. Dust like matter has an $w = 0$, relativistic particles have $w = 1/3$, while the cosmological constant can be thought of as an exotic field with $w = -1$. Due to energy conservation, the adiabatic expansion sees the energy density of radiation decrease rapidly and this component plays no role soon after decoupling.

If one knows the density distribution one can express the potential from Equation (1.9). What is left is to derive the equations of motion for the density contrast. Throughout this paper we deal with length scales that are far greater than the displacement of individual non-relativistic particles even on cosmic time scales. This enables us to employ the long wavelength approximation of fluid dynamics, the Euler equations. The first is the simple Newton's equation ($ma = F$) in continuum mechanics for a pressureless fluid, while the second is the conservation of mass:

$$\frac{\partial u(x,\tau)}{\partial \tau} + \mathcal{H}u(x,\tau) + (u(x,\tau)\nabla)u(x,\tau) = -\nabla\Phi(x,\tau), \tag{1.10}$$

$$\frac{\partial \delta(x,\tau)}{\partial \tau} + \nabla\big[(1+\delta(x,\tau))u(x,\tau)\big]. \tag{1.11}$$

By taking the divergence of Equation (1.10) and using Equation (1.9) to eliminate Φ we arrive at a coupled set of equations for δ and $\theta = \nabla u$. It is useful to write them in Fourier space:

$$\frac{\partial \theta(k,\tau)}{\partial \tau} + \mathcal{H}(\tau)\theta(k,\tau) + \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\delta(k,\tau) = -\int dk_1^3 dk_2^3 \delta_D(k-k_{12}) \times\beta(k_1,k_2)\theta(k_1,\tau)\theta(k_2,\tau), \tag{1.12}$$

$$\frac{\partial \delta(k,\tau)}{\partial \tau} + \theta(k,\tau) = -\int dk_1^3 dk_2^3 \delta_D(k-k_{12})\alpha(k_1,k_2)\theta(k_1,\tau)\delta(k_2,\tau), \tag{1.13}$$

where $k_{12} = k_1 + k_2$ and

$$\alpha(k_1,k_2) = \frac{k_{12}k_1}{k_1^2}, \qquad \beta(k_1,k_2) = \frac{k_{12}^2(k_1k_2)}{2k_1^2k_2^2}. \tag{1.14}$$

To get the lowest order deviation from the homogeneous background, one only needs to replace the right sides of Equation (1.12) and (1.13) with 0. To linear order, the time component of the density contrast can be factored out: $\delta(x,\tau) = D(\tau)\delta(x,0)$,

where the growing solution for $D(\tau)$ is the linear growth function. The growing mode gradually drives the system away from the regime where linear perturbation theory is a good approximation. The scales considered linear today are roughly below $k \approx 2\pi/30$Mpc.

The equations above are not useful without the initial conditions. Each cosmological model predicts a probability distribution for the initial density field which can be tested statistically against observations. Because the size of the universe is much larger than the scale of physical processes, the general properties of the probability distribution function (PDF) can be inferred from a single realization. Equations (1.10), (1.11), and (1.9) connect the initial PDF and the one that can be observed today:

$$\begin{array}{ccc} & \text{Eulerian dynamics} & \\ P[\delta(x,0)] & \longrightarrow & P[\delta(x,\tau)] \end{array} \tag{1.15}$$

A PDF for fields can be described by its joint moments, the n-point functions; or by the momentums of local PDF, the one-point PDF of the smoothed continuum PDF. In the case of the latter, e.g., the equivalent of the power spectrum (or the 2-point function) is the variance of the one-point PDF as the function of the smoothing scale. Both approaches have advantages and this dissertation contains examples for both. The CMB shows that the initial conditions were Gaussian (e.g., cumulants above the variance are consistent with 0; see, Komatsu *et al.*, 2005); thus the initial PDF is fully defined by its power spectrum. This picture changes due to the non-linear nature of the Euler equations. From this initial power spectrum using Equations (1.12) and (1.13), it is possible to derive an expansion for the cumulants today. These formulae can be found in reviews, e.g. in Bernardeau *et al.* (2002).

1.4 Precision cosmology

Early observations raised more questions about the universe than they answered. The grand unified models of particle physics generally predict magnetic monopoles and other exotic particles at extremely high temperatures, which have not been observed. The cosmic microwave background exhibits the properties of black body radiation, but thermal equilibrium is impossible between causally unconnected regions according to special relativity. To resolve the contradictions between cosmological observations and the mainstream physical theories, various models were designed.

The most elegant and favored of them is the inflation proposed by Guth (1981). According to his model, after the Big Bang (an initial singularity in space-time) the universe underwent a phase of exponential expansion driven by the inflaton field. This explains why the observed universe is causally connected and why we cannot find cosmic relics, exotic particles. Another consequence of the inflation is the Gaussianity of initial conditions, since all inhomogeneities were ironed out except quantum fluctuations. According to the current state of science, neither inflation nor any event prior to recombination can be tested directly by astronomical observations due to the opacity of the early universe. However, gravitational waves could stream freely from an earlier stage of the expansion. LIGO and similar detectors have been designed to intercept the signal but their sensitivity is still below that required. See Barish and Weiss (1999) for a popular article on this. The physics of the early universe can also be tested in accelerators. The operation of the Large Hadron Collider is scheduled to resume in 2012 and expected to bring us closer to a theoretical understanding of the inflation.

After the inflation ended, the universe still stayed on an expanding course described by the Friedmann equations. As it cooled, first hadrons, then elements formed. After recombination, the universe was filled with mostly neutral matter and photons. The

structure that we see now has grown through gravitational clustering from primordial fluctuations captured in the CMB. This latter, observable phase, provides data for testing cosmological models.

ΛCDM is a model of many parameters: e.g., the primordial nucleosynthesis was determined by the baryon-photon ratio; the spectrum of the inflaton field needs to be parametrized; neutrino mass; to solve the Friedmann equations, one needs the the mass ratios and the equation of state parameters of the components of the universe, etc. ΛCDM correctly predicts the abundance of light elements, the shape of the CMB power spectrum, and the observed large-scale structures. See, e.g., Dodelson (2003) for an introduction to the required calculations.

Next, I give a brief, roughly chronological overview of the most relevant, recent experiments and discoveries in the context of ΛCDM, while discussing pertinent aspects of the model. Until the 1990s, the cosmological parameters were very loosely constrained. The mission of COBE satellite launched the era of precision cosmology by measuring the spectrum of the the CMB (Smoot *et al.*, 1992). They found it to be a nearly perfect black body with a temperature of 2.725 ± 0.002. This temperature tells us the ratio of the size of the universe now and at recombination with 0.1% accuracy. They also discovered the anisotropy of the radiation. Soon, other experiments such as BOOMERanG collected enough data to measure the location of the first peak ($l = 197 \pm 6$) in the power spectrum, separating modes that had grown freely before recombination from modes that were smaller than the causally connected universe (De Bernardis *et al.*, 2000). This put a very tight constraint on the curvature of the universe, making it consistent with a flat universe. It is worth rewriting the first Friedmann equation into a canonical form, which tells less about background physics but highlights the parameters more:

$$\left(\frac{H}{H_0}\right)^2 = \Omega_R a^{-4} + \Omega_m a^{-3} + \Omega_\Lambda + \Omega_k a^{-2}, \tag{1.16}$$

where $H = \dot{a}/a$ and H_0 is the Hubble's constant. This equation can be thought of as the ratio of Equation (1.3) at time t and today; the latter is denoted by subscript 0. The contributions of the components to the total mass change with the scale factor and they scale as $\sim a^{-4}$ for radiation (R) and $\sim a^{-3}$ for matter (m). Ω_Λ represents a mysterious substance that corresponds to Einsteins's cosmological constant, and can be thought of as a component with negative pressure, e.g., $w = -1$. Finally the curvature term from Equation (1.3) can be cast into the remaining term in Equation (1.16). BOOMERanG gave a constraint of $\Omega_k = 0 \pm 0.12$ in 2000.

Due to the numerous parameters of ΛCDM, degeneracy occurs when using only a certain aspect of the model to constrain cosmology. CMB constrains the Hubble parameter only weakly. According to Equation (1.4), the second Friedmann equation, measuring the acceleration of the expansion at present reveals information about the equation of state parameter, constraining Ω_M and Ω_Λ. This scenario was pursued by Riess *et al.* (1998) measuring the redshifts and luminosities of type 1a supernovae. They are standard candles, which means that their absolute magnitudes are constants and their luminosity distances can be measured independently from redshift. The universe appears to be accelerating, as if a substance, referred to as dark energy in the scientific literature, with negative pressure dominated. The dark energy is not necessarily the cosmological constant, its w can be anything below $-1/3$, but this far this constant seems to be consistent with -1. The precise measurement of the equation of state parameter is one of the major goals of cosmology today. At present, we cannot exclude that w depends on the coordinate or the time. Currently, there is no theory that predicts the correct amount of dark energy.

Studying virialized objects, galaxies and clusters, revealed that most of the mass is invisible, not radiating in any particular waveband. This nonbaryonic matter is called dark matter (DM). There are many candidates for DM: all kinds of exotic particles, small black holes, etc. What is common to all is that they do not interact with light. The

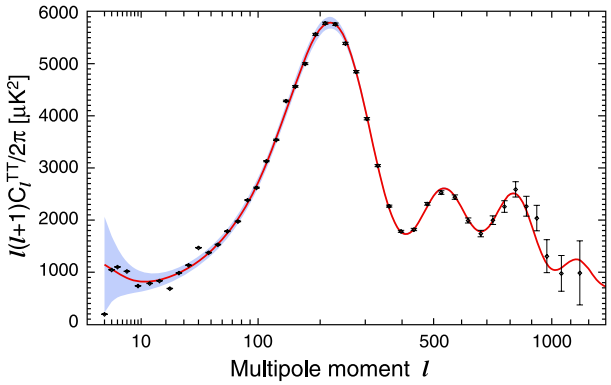

Figure 1.1 As of today, the most accurate temperature power spectrum of the CMB, the WMAP 7-year result (Larson *et al.*, 2011). This plot is reproduced from the WMAP Web site.

properties of acoustic oscillations (the speed of sound) in the primordial photon-baryon plasma depended on baryon density, so this can be inferred from the shape of the CMB power spectrum (Figure 1.1), in particular from the ratio of the odd and even peaks. WMAP (Bennett *et al.*, 2003), the successor of COBE, enhanced the quality of the measured power spectrum to resolve the degeneracy between the DM and the baryonic matter. DM is currently believed to be $\sim 80\%$ of the total matter. It also appears to be non-relativistic, since fast particles cannot be tightly bound to form clumps, clusters; thus the "cold" adjective (CDM).

In general, a model allows more independent parameters if it is more complex. On the other hand, for the same reason the constraints are looser, because the same data are used to constrain more parameters. The degeneracy among the cosmological parameters is the less in the matter distribution than in the CMB power spectrum, due to its numerous features stemming from nonlinear evolution and redshift distortion. The quality of the available data for LSS improved tremendously since the start of the Sloan Digital Sky Survey (Abazajian *et al.*, 2003). This has both imaging and spectroscopic surveys. Spectroscopy is important for good redshift resolution and distance resolution, whereas

gravitational lensing needs images with good quality. Gravitational clustering requires the introduction of new cosmological parameters because the formation of galaxies, the irreducible objects of the analysis, is not fully understood. Galaxies are assumed tracers of an underlying, smooth density field. This can be described by a biasing scheme, which on large scales is given in the form of an expansion:

$$\delta_g(x) = \int d^3x' b_1(x,x')\delta(x') + \int d^3x' b_2(x,x')\delta^2(x') + O(\delta^3) \quad . \tag{1.17}$$

The simplest assumption of this form is the local, linear case, which often gives a good description on linear scales, when δ^2 is small: $\delta_g(x) = b\delta(x)$. (See, e.g., Tegmark *et al.*, 2004, for a measurement of the galaxy power spectra in SDSS and the authors' conclusion about linear and nonlinear bias.) The continuous $\delta_g(x)$ translates into galaxy counts in a particular volume V by Poisson sampling:

$$P(N) = \frac{\lambda^N}{N!} e^{-\lambda}, \tag{1.18}$$

with $\lambda = \overline{\rho}_g(\delta_g + 1)V$ and $\delta_g = \int_V dx^3 \delta_g(x)$ where $\overline{\rho_g}$ is the average galaxy density. In a counts in cells (CIC) analysis of the PDF one needs to correct for the systematics of the Poisson sampling (see, e.g., Szapudi and Szalay, 1993). Similarly, in the case of measuring the two-point function, unbiased estimators can be built with the help of Monte Carlo simulations which span the survey volume (see, e.g., Landy and Szalay, 1993).

Since the PDF is non-Gaussian, its higher order moments cannot be derived simply from the first two; consequently, they add to our ability to constrain cosmology. Additionally, the homogeneity in the measured galaxy distribution breaks down, due to the systematic effects of redshift distortions. It arises from the fact that the distance of a galaxy is calculated from its redshift, which gauges its velocity compared to the observer.

The velocity is converted into distance by the Hubble law. Unfortunately the peculiar velocity of the galaxy, which is its velocity in addition to the Hubble flow, complicates this picture. The coordinates calculated based on redshift are referred to as redshift space coordinates. On small scales, clustering is weaker along the line of sight, because in redshift space a virialized cluster becomes elongated. On large scales, when objects are not virialized yet, gravitational infall squashes the correlation function along the line of sight. The first is called the finger of God effect, the latter is the Kaiser effect (Kaiser, 1987). Since the redshift distortion is bias dependent, the resulting symmetry breaking resolves the degeneracy between the amplitude of the power spectrum and the galaxy bias. Unlike the CMB, the PDF is time dependent, which makes it ideal to study the expansion history of the universe through features like the baryonic peak, which is a manifestation of the sound horizon of the primordial photon-baryon fluid in the matter correlation function (see, e.g., Eisenstein, 2005).

The PDF contains much more information than the CMB. However, extra information cannot be retrieved as easily, due to the nonlinear nature of gravitational clustering. Eulerian perturbation theory breaks down on small scales and redshift distortion makes relatively simple expressions for higher-order moments cumbersome. Due to these problems, only the linear power spectrum or the correlation function have been used for cosmological parameter estimation with significant results (see, e.g., Tegmark *et al.*, 2004; Pope *et al.*, 2004; Okumura *et al.*, 2008).

On small scales, an empirical approach has emerged to grasp the analytically intractable problem of cluster formation. The foundations are due to Press and Schechter (1974) and the formalism has been developing ever since (e.g., Sheth and Tormen, 1999). The Press-Schechter formula gives the number density of collapsed objects as a function of their mass, the mass function. Parameters of the formula are calibrated by simulations or can be approximated by solving the spherical collapse model, while its dependence on

cosmological parameters provides a tool to constrain cosmology. For studying clustering on small scales and the internal structure of halos, various techniques have been developed. X-ray emission and the Sunyaev-Zeldovich effect (SZ; Sunyaev and Zeldovich, 1972) can be used to measure the velocity dispersion of galaxies via the temperature of the cluster. Velocity dispersion is related to the potential and the total mass of the cluster according to the virial theorem. Weak lensing measures the potential directly based on the distortion of the shape of background galaxies. Observations to provide data for these methods are among the current endeavors of research. Pan-STARRS is an ongoing galaxy survey which identifies weak lensing as one of its key project (Kaiser, 2007), whereas the next generation of CMB probes, Planck, will measure the CMB anisotropies on small enough scales for a significant detection of the SZ effect or the polarization (Schäfer *et al.*, 2006; White, 2006; Balbi, 2007).

The CMB and the PDF are not statistically independent. The aforementioned SZ effect, the inverse Compton scattering of CMB photons in hot gas, is a good example of a mechanism that introduces correlation between large scale structure and the CMB. On large-scales, gravity can also have a significant effect on the trajectory or the energy of photons. The statistics of the distortion of images is studied under the title of weak lensing, while the change of CMB temperature due to gravitational effects is the integrated Sachs-Wolfe effect (ISW; Sachs and Wolfe, 1967). The spectrum of the cross correlation between the projected density of galaxies and the CMB temperature depends strongly on the dark energy and its equation of state. The limitation of the cross correlation comes from the large uncertainty of the projected density and small sky coverage of surveys. This will be improved by future surveys, such as Pan-STARRS.

Precision cosmology defines its goal as measuring the parameters of ΛCDM with better and better accuracy. One can ask: why is it important to know the size of the universe with 0.1% accuracy? My answer would be that the size of the universe is not as important

as other parameters of the model; e.g., the curvature, the primordial non-Gaussianity, or the equation of state parameter of the dark energy. They are all close to a special value, 0 or -1, and there is no fundamental theory with explanation. If one day the equation of state is measured to be different from -1 with high significance, that would render models based on the cosmological constant obsolete likewise, or the detection of a nonzero primordial non-Gaussianity would have serious implications for models of inflation.

1.5 Outline of dissertation

In the following Chapters, I present the results of my work between 2007 and 2010. Chapter 2 is basically our paper titled "Non-perturbative effects of geometry in wide-angle redshift distortions" and published in the Monthly Notices of the Royal Astronomical Society (Pápai and Szapudi, 2008). We discuss a full analytic treatment of the correlation function in redshift space. We use the formalism of Szapudi (2004), which is based on spherical harmonics and has the advantage over similar work that it produces a finite number of terms in the expansion and its simplicity enables other researchers to use our formulae without having to dive into the details of the calculation. Recently, Raccanelli *et al.* (2010) carried out a comprehensive study of the significance of the wide-angle corrections for the correlation function. They confirmed the accuracy and the importance of our formulae. Our work makes easier the correct analysis of the correlation function on scales comparable to the distance between the observer and the galaxies in the survey.

Chapters 3 and 4 discuss our paper, "Cosmological Density Fluctuations on 100 Mpc Scales and their ISW Effect", which appeared in The Astrophysical Journal (Pápai and Szapudi, 2010). This paper combines two topics. Clustering on large scales and the ISW effect. In Chapter 3 we present a CIC measurement in the SDSS LRG sample. Our goal is to study the PDF on the largest possible scales and look for deviations from the ΛCDM.

We use these results in Chapter 4 to give an estimate for the temperature shift due to ISW in the direction of the center of the largest superstructures.

In Chapter 5 we deal with the full density and ISW profile of superstructures. We create mock ISW images from N-body simulations to test our theoretical calculations. We create an ISW map based on Granett *et al.* (2008) and our theoretical ISW profile and measure its amplitude in the CMB. This work has been submitted for publication in The Astrophysical Journal with the title of "ISW Imprint of Superstructures on Linear Scales" (Pápai *et al.*, 2010).

In Chapter 6, I summarize my results and their possible extensions in a future work.

References

Abazajian, K., J. Adelman-McCarthy, M. Ag
"ueros, S. Allam, S. Anderson, J. Annis, N. Bahcall, I. Baldry, S. Bastian, A. Berlind, *et al.*, 2003, The Astronomical Journal **126**, 2081.

Balbi, A., 2007, New Astronomy Reviews **51**(3-4), 281, ISSN 1387-6473.

Barish, B. C., and R. Weiss, 1999, Physics Today **52**, 44.

Bennett, C. L., M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, E. Wollack, E. L. Wright, *et al.*, 2003, ApJS **148**, 1.

Bernardeau, F., S. Colombi, E. Gaztañaga, and R. Scoccimarro, 2002, Phys. Rep. **367**, 1.

De Bernardis, P., P. Ade, J. Bock, J. Bond, J. Borrill, A. Boscaleri, K. Coble, B. Crill, G. De Gasperis, P. Farese, *et al.*, 2000, Nature **404**(6781), 955, ISSN 0028-0836.

Dodelson, S., 2003, *Modern cosmology* (Academic Pr), ISBN 0122191412.

Eisenstein, D., 2005, New Astronomy Reviews **49**(7-9), 360, ISSN 1387-6473.

Granett, B. R., M. C. Neyrinck, and I. Szapudi, 2008, ApJ **683**, L99.

Guth, A., 1981, Physical Review D **23**(2), 347, ISSN 1550-2368.

Hubble, E., 1929, Proceedings of the National Academy of Sciences of the United States of America **15**(3), 168.

Kaiser, N., 1987, MNRAS **227**, 1.

Kaiser, N., 2007, in *Bulletin of the American Astronomical Society*, volume 38 of *Bulletin of the American Astronomical Society*, pp. 163–+.

Komatsu, E., D. N. Spergel, and B. D. Wandelt, 2005, ApJ **634**, 14.

Landau, L., and E. Lifshitz, 1980, *The classical theory of fields* (Butterworth-Heinemann), ISBN 0750627689.

Landy, S., and A. Szalay, 1993, The Astrophysical Journal **412**, 64, ISSN 0004-637X.

Larson, D., J. Dunkley, G. Hinshaw, E. Komatsu, M. R. Nolta, C. L. Bennett, B. Gold, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, M. Limon, *et al.*, 2011, ApJS **192**, 16.

Okumura, T., T. Matsubara, D. J. Eisenstein, I. Kayo, C. Hikage, A. S. Szalay, and D. P. Schneider, 2008, ApJ **676**, 889.

Pápai, P., and I. Szapudi, 2008, MNRAS **389**, 292.

Pápai, P., and I. Szapudi, 2010, ApJ **725**, 2078.

Pápai, P., I. Szapudi, and B. R. Granett, 2010, ArXiv e-prints 1012.3750.

Penzias, A., and R. Wilson, 1965, The Astrophysical Journal **142**, 419, ISSN 0004-637X.

Pope, A. C., T. Matsubara, A. S. Szalay, M. R. Blanton, D. J. Eisenstein, J. Gray, B. Jain, N. A. Bahcall, J. Brinkmann, T. Budavari, A. J. Connolly, J. A. Frieman, *et al.*, 2004, ApJ **607**, 655.

Press, W., and P. Schechter, 1974, The Astrophysical Journal **187**, 425, ISSN 0004-637X.

Raccanelli, A., L. Samushia, and W. J. Percival, 2010, MNRAS **409**, 1525.

Riess, A., A. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. Garnavich, R. Gilliland, C. Hogan, S. Jha, R. Kirshner, *et al.*, 1998, The Astronomical Journal **116**, 1009.

Sachs, R. K., and A. M. Wolfe, 1967, ApJ **147**, 73.

Schäfer, B. M., C. Pfrommer, M. Bartelmann, V. Springel, and L. Hernquist, 2006, MNRAS **370**, 1309.

Sheth, R., and G. Tormen, 1999, Monthly Notices of the Royal Astronomical Society **308**(1), 119, ISSN 1365-2966.

Smoot, G., C. Bennett, A. Kogut, E. Wright, J. Aymon, N. Boggess, E. Cheng, G. De Amici, S. Gulkis, M. Hauser, *et al.*, 1992, The Astrophysical Journal **396**, L1, ISSN 0004-637X.

Sunyaev, R. A., and Y. B. Zeldovich, 1972, Comments on Astrophysics and Space Physics **4**, 173.

Szapudi, I., 2004, ApJ **614**, 51.

Szapudi, I., and A. S. Szalay, 1993, ApJ **408**, 43.

Tegmark, M., M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, D. H. Weinberg, I. Zehavi, A. Berlind, T. Budavari, A. Connolly, *et al.*, 2004, ApJ **606**, 702.

White, M., 2006, New Astronomy Reviews **50**(11-12), 938, ISSN 1387-6473.

Chapter 2

Non-perturbative effects of geometry in wide-angle redshift distortions

2.1 Introduction

Redshift distortions represent a curse disguised as a blessing for high precision cosmological applications. Radial coordinates of redshift surveys contain limited phase space information, which in principle can be used to constrain theories more than configuration information alone; moreover, velocities are sensitive to structure outside of the survey boundaries, which potentially translates into a larger "effective volume". On the other hand, redshift distortions are plagued with non-linearities, both on large and small scales, therefore; in the worst case they could amount to poorly understood contamination of the configuration space data. Our aim is to extend the theory of linear redshift distortions such that large angle information could be successfully extracted from galaxy surveys.

The work of Davis and Peebles (1983) and Peebles (1980) showing that redshift distortions affect the power spectrum spawned a lot of activity. The all-important linear, plane-parallel limit was first calculated by Kaiser (1987), showing that the effect on the power spectrum corresponds to "squashing". The other well known "finger of God"

effect dominates small scales, and is irrelevant for our present study. The Kaiser formula has been generalized for real space soon after (e.g., Hamilton, 1993; Cole *et al.*, 1995). These theories have been used to analyze surveys such as the Point Source Catalog Redshift (PSCz) Survey (Tadros *et al.*, 1999), the Two-Degree Field Galaxy Redshift Survey (2dFGRS; Peacock *et al.*, 2001; Hawkins *et al.*, 2003; Tegmark *et al.*, 2002), and the Sloan Digital Sky Survey (SDSS; Zehavi *et al.*, 2002).

The distant observer approximation only holds if pairs are separated by a small angle. This means that a large fraction of pairs needs to be thrown away from modern wide angle redshift surveys when they are analyzed in this limit. These pairs are typically fewer and noisier than close pairs, but if our aim is to extract as much information as possible from a given survey, it would be desirable to include them. Hamilton and Culhane (1996) related the "ω-transform", a complexified Mellin-like transform, of the two-point correlation function to that of redshifted ω-space correlation function. The resulting spherical $\omega\ell m$ expansion is approximately orthogonal to redshift space distortions. This expansion truncated at an appropriate mode was used in Tegmark *et al.* (2002) to analyze data in this transform space. The first explicit perturbation theory calculation in coordinate space was performed in Szalay *et al.* (1998). The result is a simple-to-use finite expression, but only in coordinate space: in Fourier space an infinite series will result for the redshift-distorted analog of the power spectrum. These formulae were later further generalized to include high-z effects in various cosmologies by Matsubara (2004). These calculations provide essential input for the pixel-based Karhunen-Loéve (KL) or quadratic likelihood analyses (e.g., Vogeley and Szalay, 1996; Tegmark *et al.*, 1997), since distant observer approximation is not valid for modern wide angle galaxy surveys. The theory has been applied in several subsequent analyses of wide angle redshift surveys, such as Pope *et al.* (2004) and Okumura *et al.* (2008). Despite its elegance, the theory did not agree well

with dark matter simulations. Scoccimarro (2004) pointed out that this might be due to non-perturbative effects.

Szapudi (2004) reanalyzed the redshift distortion problem from a group-theoretical point of view showing that tripolar spherical harmonics provide an excellent basis for expansion, and result in especially compact formulae. In addition, this work provided specific coordinate systems, one of which recovers the Legendre expansion of Szalay *et al.* (1998), while the other represents the same information in an even simpler two-dimensional Fourier mode expansion. We use this formalism to take into account a term in the Jacobian, previously neglected in all explicit calculations, to derive the full linear redshift-distorted correlation function.

Kaiser's original work starts with the full linear Jacobian. It contains a term negligible for small angles that is linear in terms of the small fluctuations, and is essentially non-linear from the point of view of geometry: it contains a $1/r$ prefactor. Moreover, this term, if expanded in bipolar spherical harmonics (or any other way), would contribute infinite coefficients. Because of the presumed subdominance due to the prefactor, and complexity of the calculation, this term was neglected in all previous coordinate space expressions, although it is represented in the ω-space expansion of Hamilton and Culhane (1996). In this chapter, we introduce a hybrid approach, in which we leave the essentially non-perturbative terms in the expansion intact; our tripolar expansion coefficients will still contain angular variables in a specific way. As we show later, this hybrid procedure results in a finite number of terms, and it provides significant corrections and improvement in the agreement with simulations. In retrospect, the omission of this term, while intuitively reasonable, is not justified, as its contribution can become important on the most interesting scales of tens of $\mathrm{h}^{-1}\mathrm{Mpc}$s.

In Section 2.2, we present the theory of linear redshift distortions, including results from the geometric term in the Jacobian. We follow closely the formalism of Szapudi

(2004), mainly focusing on the new aspects of this calculation. For reference, we print the full result, which has about twice as many terms as previously. In Section 2.3 we compare our results with preliminary measurements in the Hubble Volume Simulations, and present our conclusions.

2.2 Redshift distortion of the two-point correlation function

We use linear perturbation theory to predict the redshift-distorted two-point correlation function in terms of the underlying power spectrum. Our calculation is based directly on the tripolar expansion formalism of Szapudi (2004); therefore our focus will be on the additional terms arising from the Jacobian.

The exact mapping between real and redshift space is $s_i = x_i - f v_j \hat{x}_i \hat{x}_j$, where the "hat" denotes the proper unit vector, $f = \dfrac{\Omega^{0.6}}{b}$ and the velocity has units which provide that its divergence is equal to the density up to linear order. From this, one can calculate the derivative of this matrix: $\partial s_i / \partial x_k = \delta_{ik} + O_{ik}(v)$ where O is linear in v. This results in a linear Jacobian $J = 1 + TrO = 1 - f \hat{x}_i \hat{x}_j \partial_i v_j - 2f \dfrac{x_j v_j}{x^2}$. The last term in the previous expression is usually omitted due to the fact that it scales with $1/x$; i.e., it would tend to zero for large distances, which loosely correspond to large angles as well. Closer examination of this term shows that it is of the same order as the previous term, not only in perturbation expansion (linear), but also in order of magnitude. Our goal is to propagate this new term through the full calculation.

The linear density contrast and the two-point function can be expressed in the usual fashion.

$$\delta_s(x) = \int \frac{\mathrm{d}^3k}{(2\pi)^3} e^{ik_j x_j} \left[1 + f(\hat{x}_j\hat{k}_j)^2 - i2f\frac{\hat{x}_j\hat{k}_j}{xk}\right]\delta(k) \tag{2.1}$$

$$\begin{aligned}\langle\delta_s(x_1)\delta_s^*(x_2)\rangle = \int \frac{\mathrm{d}^3k}{(2\pi)^3} P(k) e^{ik(x_1-x_2)} \\ \left[1 + \frac{f}{3} + \frac{2f}{3}P_2(\hat{x}_1\hat{k}) - \frac{i2f}{x_1k}P_1(\hat{x}_1\hat{k})\right] \\ \left[1 + \frac{f}{3} + \frac{2f}{3}P_2(\hat{x}_2\hat{k}) + \frac{i2f}{x_2k}P_1(\hat{x}_2\hat{k})\right],\end{aligned} \tag{2.2}$$

where P_1 and P_2 are Legendre polynomials and $P(k)$ is the linear power spectrum. The third term in each of the brackets corresponds to the extension of the previous results; these would tend to zero in the plane parallel limit. At wide angles, the separation between the galaxies and the distance between a galaxy and the observer are of the same order; therefore kx is of order unity. This shows explicitly that the order of this term can be as large as the previous, and the detailed calculation confirms this.

Next we express the angular dependence of the correlation function with tripolar spherical harmonics.

$$\begin{aligned}S_{l_1l_2l}(\hat{x}_1, \hat{x}_2, \hat{x}) \\ \equiv \sum_{m_1,m_2,m} \begin{pmatrix} l_1 & l_2 & l \\ m_1 & m_2 & m \end{pmatrix} C_{l_1m_1}(\hat{x}_1) C_{l_2m_2}(\hat{x}_2) C_{lm}(\hat{x})\end{aligned} \tag{2.3}$$

We use x for denoting $x_1 - x_2$. On the right hand side one can find the Wigner $3j$ symbols and we define the normalized spherical functions as $C_{lm} = \sqrt{4\pi/2l+1}Y_{lm}$; these result in simpler expressions.

Eq. (2.2) has become more complex with the additions; x_1 and x_2 appear in the denominator resulting in the following angular dependence

$$x_1 = g_1 x = \frac{\sin(\phi_2)}{\sin(\phi_2 - \phi_1)} x, \tag{2.4}$$

$$x_2 = g_2 x = \frac{\sin(\phi_1)}{\sin(\phi_2 - \phi_1)} x. \tag{2.5}$$

Expanding these terms into tripolar spherical harmonics would yield an infinite number of terms, but simplification arises from the fact that they can be factored out of the integrals. All the rest can be expanded as in Szapudi (2004), resulting in finite expressions. We introduce ϕ_1 to denote the angle between $\hat{x}_1$ and $\hat{x}$ and ϕ_2 for the angle between $\hat{x}_2$ and $\hat{x}$. We emphasize that the coefficients of this (quasi-)tripolar expansion still have an angular dependence in the form of g_1 and g_2:

$$\xi_s = \sum_{l_1 l_2 l} B^{l_1 l_2 l}(x, \phi_1, \phi_2) S_{l_1 l_2 l}(\hat{x}_1, \hat{x}_2, \hat{x}). \tag{2.6}$$

After performing the expansions, only a finite number of coefficients survive. For reference, the ones from Szapudi (2004) are:

$$\begin{aligned}
B^{000}(x) &= (1 + \frac{1}{3} f)^2 \xi_0^2(x) \\
B^{220}(x) &= \frac{4}{9\sqrt{5}} f^2 \xi_0^2(x) \\
B^{022}(x) = B^{202}(x) &= -(\frac{2}{3} f + \frac{2}{9} f^2)\sqrt{5}\xi_2^2(x) \\
B^{222}(x) &= \frac{4\sqrt{10}}{9\sqrt{7}} f^2 \xi_2^2(x) \\
B^{224}(x) &= \frac{4\sqrt{2}}{\sqrt{35}} f^2 \xi_4^2(x);
\end{aligned} \tag{2.7}$$

and the new terms, the main result in this chapter, are

$$
\begin{aligned}
B^{101}(x,\phi_1,\phi_2) &= -(2f+\frac{2}{3}f^2)\frac{\sqrt{3}}{g_1 x}\xi_1^1(x)\\
B^{011}(x,\phi_1,\phi_2) &= (2f+\frac{2}{3}f^2)\frac{\sqrt{3}}{g_2 x}\xi_1^1(x)\\
B^{121}(x,\phi_1,\phi_2) &= \frac{4\sqrt{2}}{\sqrt{15}}f^2\frac{1}{g_1 x}\xi_1^1(x)\\
B^{211}(x,\phi_1,\phi_2) &= -\frac{4\sqrt{2}}{\sqrt{15}}f^2\frac{1}{g_2 x}\xi_1^1(x)\\
B^{123}(x,\phi_1,\phi_2) &= \frac{4\sqrt{7}f^2}{\sqrt{15}g_1 x}\xi_3^1(x)\\
B^{213}(x,\phi_1,\phi_2) &= -\frac{4\sqrt{7}f^2}{\sqrt{15}g_2 x}\xi_3^1(x)\\
B^{110}(x,\phi_1,\phi_2) &= -\frac{4f^2}{\sqrt{3}g_1 g_2 x^2}\xi_0^0(x)\\
B^{112}(x,\phi_1,\phi_2) &= -\frac{4\sqrt{10}f^2}{\sqrt{3}g_1 g_2 x^2}\xi_2^0(x),
\end{aligned}
\tag{2.8}
$$

where $\xi_l^m(x) = \int \mathrm{d}k/2\pi^2 k^m j_l(xk)P(k)$, with j being the spherical Bessel function.

For further elaboration we choose coordinate system a) from Szapudi (2004). This corresponds to our previous choice of angles with ϕ_1, ϕ_2, with which the $S_{l_1 l_2 l}(\hat{x}_1,\hat{x}_2,\hat{x}) = S_{l_1 l_2 l}(\pi/2,\phi_1,\pi/2,\phi_2,\pi/2,0)$ functions can be expressed using cosines and sines only. Using the same notation as Szapudi (2004):

$$
\begin{aligned}
\xi_s(\phi_1,\phi_2,x) = \sum_{n_1,n_2=0,1,2} & a_{n_1 n_2}\cos(n_1\phi_1)\cos(n_2\phi_2)\\
& + b_{n_1 n_2}\sin(n_1\phi_1)\sin(n_2\phi_2).
\end{aligned}
\tag{2.9}
$$

Again, for reference, the previously calculated coefficients are

$$
\begin{aligned}
a_{00} = &\left(1+\frac{2f}{3}+\frac{2f^2}{15}\right)\xi_0^2(x) -\\
&\left(\frac{f}{3}+\frac{2f^2}{21}\right)\xi_2^2(x)+\frac{3f^2}{140}\xi_4^2(x)
\end{aligned}
$$

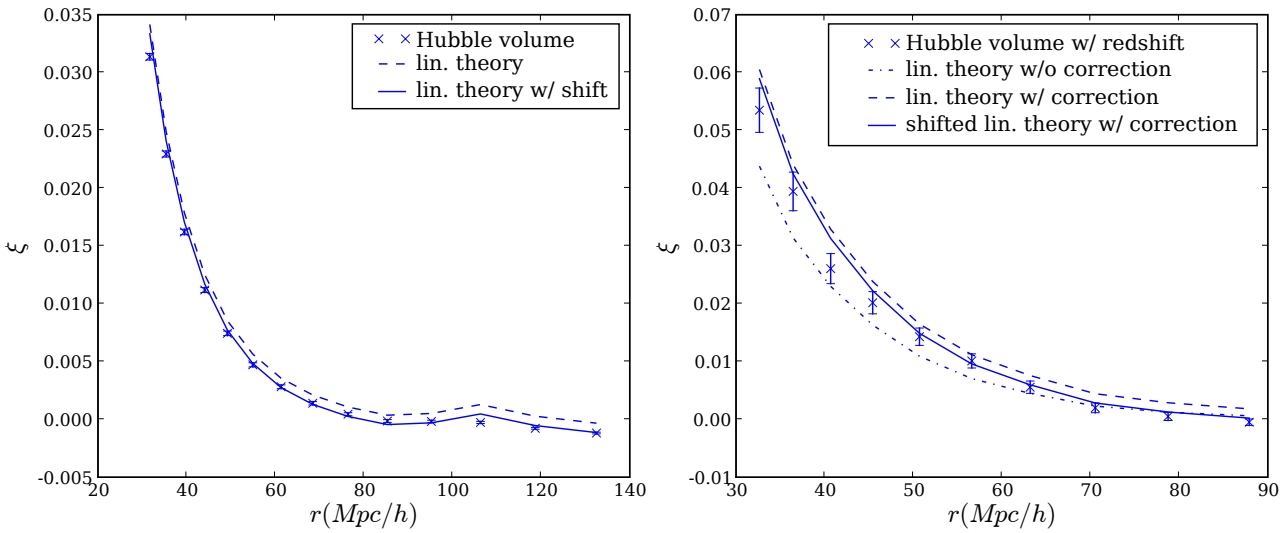

Figure 2.1 (Left) The measurement of the correlation function without redshift distortion of the Hubble Volume Simulation (symbols) compared with linear theory(dashed and solid lines). The error bars were estimated from 9^3 subvolumes of th Hubble Volume. Shifting the theory by 0.00081 downward, motivated by the integral constraint, provides an excellent fit to the data. (Right) Redshift distorted correlation function of the Hubble Volume Simulation (symbols) at constant opening angle (0.71 radian) and while the ratio of the distances of the particles in the pair are kept fixed (at 1.57). The error bars were estimated as before. The lines indicate the linear theories with and without the geometric terms. The solid line is the corrected theory with a downshift of 0.0016. The integral constraint correction is expected to be larger since the average of the two point function is larger.

$$
\begin{aligned}
a_{02} = a_{20} &= \left(\frac{-f}{2} - \frac{3f^2}{14}\right)\xi_2^2(x) + \frac{f^2}{28}\xi_4^2(x) \\
a_{22} &= \frac{f^2}{15}\xi_0^2(x) - \frac{f^2}{21}\xi_2^2(x) + \frac{19f^2}{140}\xi_4^2(x) \\
b_{22} &= \frac{f^2}{15}\xi_0^2(x) - \frac{f^2}{21}\xi_2^2(x) - \frac{4f^2}{35}\xi_4^2(x);
\end{aligned} \tag{2.10}
$$

and the new expressions of this work correspond to

$$
\begin{aligned}
a_{10} = \frac{\tilde{a}_{10}}{g_1} &= (2f + \frac{4f^2}{5})\frac{1}{g_1 x}\xi_1^1 - \frac{1}{5}\frac{f^2}{g_1 x}\xi_3^1 \\
a_{01} = \frac{\tilde{a}_{01}}{g_2} &= -(2f + \frac{4f^2}{5})\frac{1}{g_2 x}\xi_1^1 + \frac{1}{5}\frac{f^2}{g_2 x}\xi_3^1 \\
a_{11} = \frac{\tilde{a}_{11}}{g_1 g_2} &= \frac{4}{3}\frac{f^2}{g_1 g_2 x^2}\xi_0^0 - \frac{8}{3}\frac{f^2}{g_1 g_2 x^2}\xi_2^0 \\
a_{21} = \frac{\tilde{a}_{21}}{g_2} &= -\frac{2}{5}\frac{f^2}{g_2 x}\xi_1^1 + \frac{3}{5}\frac{f^2}{g_2 x}\xi_3^1 \\
a_{12} = \frac{\tilde{a}_{12}}{g_1} &= \frac{2}{5}\frac{f^2}{g_1 x}\xi_1^1 - \frac{3}{5}\frac{f^2}{g_1 x}\xi_3^1 \\
b_{11} = \frac{\tilde{b}_{11}}{g_1 g_2} &= \frac{4}{3}\frac{f^2}{g_1 g_2 x^2}\xi_0^0 + \frac{4}{3}\frac{f^2}{g_1 g_2 x^2}\xi_2^0 \\
b_{21} = \frac{\tilde{b}_{21}}{g_2} &= -\frac{2}{5}\frac{f^2}{g_2 x}\xi_1^1 - \frac{2}{5}\frac{f^2}{g_2 x}\xi_3^1 \\
b_{12} = \frac{\tilde{b}_{12}}{g_1} &= \frac{2}{5}\frac{f^2}{g_1 x}\xi_1^1 + \frac{2}{5}\frac{f^2}{g_1 x}\xi_3^1.
\end{aligned} \tag{2.11}
$$

It is worth emphasizing again that the angular dependence g_1 and g_2 is suppressed for clarity in the above formulae, but it obviously carries through according to the definition of these functions. If the equivalence of the configurations $(\phi_1, \phi_2) \to (\pi - \phi_2, \pi - \phi_1)$ is taken into account (same pairs can be counted twice), the number of independent new coefficients is five; i.e., the number of terms approximately doubled. Next we explore the relevance of these calculations, and compare the theoretical predictions with measurements in dark matter only N-body simulations.

2.3 Discussion and summary

To understand our results, we expanded our formulae to identify leading order corrections to the Kaiser limit.

The leading order corrections to the distant observer approximation are second order. Using the notation $\frac{1}{2}(\phi_1 + \phi_2) = \phi$ and $\frac{1}{2}(\phi_2 - \phi_1) = \Delta\phi$, and keeping leading order terms in $\Delta\phi$ results in

$$\begin{aligned}
&\xi_s(\phi, \Delta\phi, x) \\
&= a_{00} + 2a_{02}\cos(2\phi) + a_{22}\cos^2(2\phi) + b_{22}\sin^2(2\phi) \\
&\quad + \Big[-4a_{02}\cos(2\phi) - 4a_{22} - 4b_{22} \Big] \Delta\phi^2 \\
&\quad + \Big[-4\tilde{a}_{10}\cot^2(\phi) + 4\tilde{a}_{11}\cot^2(\phi) \\
&-4\tilde{a}_{12}\cot^2(\phi)\cos(2\phi) + 4\tilde{b}_{11} - 8\tilde{b}_{12}\cos^2(\phi) \Big] \Delta\phi^2 + \\
&\quad + O(\Delta\phi^4). \qquad (2.12)
\end{aligned}$$

The first line of Equation (2.12) corresponds to the Kaiser formula $(\Delta\phi = 0)$.The next line contains leading order corrections corresponding to previous work only, and the third line collects leading order corrections from the geometric term in the Jacobian. These are all of the same order, reassuring the need of keeping the geometric non-perturbative terms. We speculate that the terms containing the $\cot^2(\phi)$ could be responsible for the reported failure of the linear theory for small angles along the line of sight (Okumura *et al.*, 2008).

As a preliminary test of the validity of our calculations, we measured correlation functions in the Hubble Volume Simulation (Evrard *et al.*, 2002), using cosmological parameters $\sigma_8 = 0.9$, $n_s = 1$, $\Omega_m = 0.3$, $\Omega_\lambda = 0.7$, $h = 0.7$, $\Omega_b h^2 = 0.0196$ and

a volume of $(3000 Mpc h^{-1})^3$, with and without redshift distortions. The volume of the simulation was divided into 9^3 subvolumes to obtain the error bars.

The left panel of the figure shows the measured and the theoretical two-point functions without redshift distortions. The theory agrees with the measurements only after a shift by a constant. This is due to the "integral constraint" problem (e.g., Peebles, 1980), possibly compounded with slight non-linear effects. This constant represents a bias which is approximately equal to the average of the two-point correlation function over the survey area. It can be determined several ways (see discussion below).

Next, an observer was placed at the center of each subvolume and the mapping between real and redshift space was performed using the velocities recorded in the simulation. The correlation function was then measured using brute force counting of pairs in high resolution bins matching our choice of coordinate system described earlier. The right panel of the figure presents wide angle redshift distortion theory both with and without non-perturbative geometric corrections. The latter cannot be made to agree with the measurements even using a constant offset due to the integral constraint. In contrast, the theory presented in this chapter provides excellent agreement with the measurements if the effects of integral constraint are taken into account. Note that this shift is expected to be larger with redshift distortions simply because the two-point function is enhanced on large scales.

Although one can simply fit this constant shift, corresponding to throwing away a constant from the two-point correlation function (Fisher *et al.*, 1993), we have estimated it in two more ways: Monte Carlo integrating the theoretical expression for the correlation function, and empirically measuring the variance of the average density on the scales of the subsamples. All three methods are consistent with each other; the figure uses the empirical variance over subsamples. Note that in applications, the first method, i.e.

discarding a constant from the theory, is the most prudent procedure to follow, since fluctuations on the scale of the full survey are not measurable.

Although these measurements are preliminary in the sense that we did not try to span the full parameter space of wide angle redshift distortions, the results presented in this figure appear to be typical: any other configurations we measured showed similar improvement. Scanning the full parameter space with our present brute force two-point correlation function code would be impractical, since we need a very large number of pairs in each bin to "beat down" the error bars enough that the difference between the two theories can be reliably measured. Although we developed a fast grid-based code, we found that at these small values of the correlation, function the pixel window function effects become important. These are more complex for the redshift-distorted correlation function depending on three variables than in real space. Such an effect should be modeled very accurately before one could fully span the available parameter space.

A few simple extensions and modifications of our theory are needed for practical applications when measuring the two-point function (Okumura *et al.*, 2008), or when using our results to estimate a theoretical covariance matrix for a Karhunen-Loéve (KL) analysis, (see Pope *et al.*, 2004, for details). If the sample is not volume limited, the redshift space density contrast is defined through the redshift space selection function ($\Phi(r)$). The effect of this can be taken into account by $g \to 2g/\alpha$. Where $\alpha = \dfrac{\mathrm{d}\log(r^2\Phi(r))}{\mathrm{d}\log(r)}$. The local bias can be neglected if we only deal with pairs farther away from the observer than the correlation length and motion of the local group can be transformed out by using the frame of the cosmic microwave background. These problems have been discussed in detail by Hamilton and Culhane (1996), and the solutions are exactly analogous in our case.

Note that the integral constraint problem does not appear in KL analysis where only modes orthogonal to the average density are used. This is more elegant than the simple treatment we have given here, but the essence of it is the same: regarding the constant

in the two-point correlation function as a nuisance parameter accomplishes the same for direct applications of our formulae.

With these caveats we conclude that our theory of wide angle redshift distortions yielded simple-to-use explicit formulae, which agree with simulations. The corrections to previous formulae represent a significant improvement at modest cost in complexity. Possible generalizations along the lines of Matsubara (2004) are left for future research.

References

Cole, S., K. B. Fisher, and D. H. Weinberg, 1995, MNRAS **275**, 515.

Davis, M., and P. J. E. Peebles, 1983, ApJ **267**, 465.

Evrard, A. E., T. J. MacFarland, H. M. P. Couchman, J. M. Colberg, N. Yoshida, S. D. M. White, A. Jenkins, C. S. Frenk, F. R. Pearce, J. A. Peacock, and P. A. Thomas, 2002, ApJ **573**, 7.

Fisher, K. B., M. Davis, M. A. Strauss, A. Yahil, and J. P. Huchra, 1993, ApJ **402**, 42.

Hamilton, A. J. S., 1993, ApJ **417**, 19.

Hamilton, A. J. S., and M. Culhane, 1996, MNRAS **278**, 73.

Hawkins, E., S. Maddox, S. Cole, O. Lahav, D. S. Madgwick, P. Norberg, J. A. Peacock, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, *et al.*, 2003, MNRAS **346**, 78.

Kaiser, N., 1987, MNRAS **227**, 1.

Matsubara, T., 2004, ApJ **615**, 573.

Okumura, T., T. Matsubara, D. J. Eisenstein, I. Kayo, C. Hikage, A. S. Szalay, and D. P. Schneider, 2008, ApJ **676**, 889.

Peacock, J. A., S. Cole, P. Norberg, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. D. Cannon, M. Colless, C. Collins, W. Couch, G. Dalton, K. Deeley, *et al.*, 2001, Nature **410**, 169.

Peebles, P., 1980, *The large-scale structure of the universe* (Princeton Univ Pr), ISBN 0691082405.

Pope, A. C., T. Matsubara, A. S. Szalay, M. R. Blanton, D. J. Eisenstein, J. Gray, B. Jain, N. A. Bahcall, J. Brinkmann, T. Budavari, A. J. Connolly, J. A. Frieman, *et al.*, 2004, ApJ **607**, 655.

Scoccimarro, R., 2004, Phys. Rev. D **70**(8), 083007.

Szalay, A. S., T. Matsubara, and S. D. Landy, 1998, ApJ **498**, L1+.

Szapudi, I., 2004, ApJ **614**, 51.

Tadros, H., W. E. Ballinger, A. N. Taylor, A. F. Heavens, G. Efstathiou, W. Saunders, C. S. Frenk, O. Keeble, R. McMahon, S. J. Maddox, S. Oliver, M. Rowan-Robinson, *et al.*, 1999, MNRAS **305**, 527.

Tegmark, M., A. J. S. Hamilton, and Y. Xu, 2002, MNRAS **335**, 887.

Tegmark, M., A. N. Taylor, and A. F. Heavens, 1997, ApJ **480**, 22.

Vogeley, M. S., and A. S. Szalay, 1996, ApJ **465**, 34.

Zehavi, I., M. R. Blanton, J. A. Frieman, D. H. Weinberg, H. J. Mo, M. A. Strauss, S. F. Anderson, J. Annis, N. A. Bahcall, M. Bernardi, J. W. Briggs, J. Brinkmann, *et al.*, 2002, ApJ **571**, 172.

Chapter 3

The dark matter probability distribution function on 100Mpc scales

3.1 Introduction

On large scales, well within the domain of linear perturbation theory, there are hints of tantalizing structure in various reports. Periodic patterns in pencil beams (Broadhurst *et al.*, 1990) and great walls in galaxy surveys (Geller and Huchra, 1989; Gott *et al.*, 2005) have been discovered. Other researchers predict the existence of large coherent structures based on indirect evidence: hot or cold spots in the CMB (Cruz *et al.*, 2005; Granett *et al.*, 2008). Since these findings seemingly reject the simple Gaussian hypothesis, they demand scrutiny from the scientific community. This can mean the recheck of the statistical methods employed, usually a Bayesian calculation of significance (see, e.g., Kaiser and Peacock, 1991, regarding pencil beams) or gathering supporting or contradicting evidence in the newly available data.

All the mentioned anomalies might signal an excess power on large scales or a deviation from a Gaussian PDF. The two-point and three-point functions have already been measured on various scales and compared to predictions in the largest galaxy survey,

the Sloan Digital Sky Survey (Eisenstein *et al.*, 2005; Okumura *et al.*, 2008; Nichol *et al.*, 2006; Kulkarni *et al.*, 2007, SDSS;). In this chapter, we use an alternative characterization of the PDF. If the density field is smoothed on a range of scales, the moments of its local PDF as a function of smoothing-length contain all the information. The technique for measuring the local PDF is termed counts in cells (CIC), which has been neglected recently in favor of the n-point functions. We face practical difficulties that arise due to the survey mask or redshift distortions and solve them in a novel way.

The hydrodynamical model of the universe is based on the assumption that the observed galaxy distribution is a Poisson-sampled version of a continuous field. Furthermore, this continuous field is a realization of a random field (see, e.g., Peebles 1980). Measuring CIC is a well-established method to estimate its PDF (see, e.g., Colombi *et al.* 2000; Szapudi *et al.* 2000). In this chapter, we first give a brief summary of this method and description of the data, the SDSS Luminous Red Galaxy (LRG) Catalog, and the algorithm we use in our analysis. From CIC we derive the first few moments (the variance, S_3, and S_4) of the PDF and compare them to the predictions of ΛCDM.

3.2 CIC and the PDF

In the following we deal with two random fields, one corresponding to dark matter, and the other its biased version, the galaxy field. In our notation the terms that refer to these are "matter PDF", "matter field", "galaxy PDF" and "galaxy field".

The matter PDF for cells is fully given by its cumulants, possibly normalized (Peebles, 1980):

$$\overline{\xi}_n = \frac{1}{V^n} \int_V ... \int_V \left\langle \delta(x_1)...\delta(x_n) \right\rangle_c dx_1...dx_n, \tag{3.1}$$

$$S_n = \frac{\overline{\xi}_n}{\overline{\xi}_2^{n-1}}, \tag{3.2}$$

where V is the volume of a cell and the subscript c refers to connected moments. These formulae show that the cumulants of the local PDF are simply the volume averaged n-point functions. Inverting these equations to get the n-point functions is done via derivation. When $\overline{\xi}_2 << 1$, S_ns are around unity according to perturbation theory. Thus, S_3, the skewness, and S_4, the kurtosis, represent the lowest-order correction to a Gaussian distribution. Estimating these quantities from galaxy counts can be done in two steps: first, by using factorial moments to get the moments of the underlying galaxy PDF (see, e.g., Szapudi and Szalay, 1993); second, by using some galaxy–dark-matter biasing scheme to transform the galaxy PDF into the matter PDF.

Estimators for the variance ($\overline{\xi}_2$), S_3 and S_4 are:

$$\overline{\xi}_2 = \frac{\left\langle N^2 \right\rangle - \left\langle N \right\rangle}{\left\langle N \right\rangle^2} - 1, \tag{3.3}$$

$$S_3 = \frac{(F_3 - 3F_2F_1 + 2F_1^3)/F_1^3}{\overline{\xi}_2^2}, \tag{3.4}$$

$$S_4 = \frac{(F_4 - 4F_3F_1 + 12F_2F_1^2 - 6F_1^4 - 3F_2^2)/F_1^4}{\overline{\xi}_2^3}. \tag{3.5}$$

where

$$F_n = \left\langle N(N-1)...(N-n+1) \right\rangle = V^n \left\langle \rho_V^n \right\rangle. \tag{3.6}$$

The average density of the continuous field in a volume V is denoted by ρ_V. The extra terms next to the naive $\langle N^n \rangle$ in Equation (3.6) cancel out the effects of the Poisson-sampling. This can be demonstrated easily for the variance. If the underlying matter PDF is convolved with a Poisson distribution their variances are to be added: $V^2\langle \rho_V^2 \rangle + \langle N \rangle = \langle N^2 \rangle - \langle N \rangle^2$. Ergodicity ensures that these ensemble averages can be calculated from a single, ideally large, volume-limited survey. The caveats, arising when we depart from the ideal case, are discussed later in this section.

In this chapter, we use cosmological parameters taken at their best-fit *WMAP* values (Spergel *et al.*, 2007). For the bias, we fit the simplest, deterministic, local, linear model:

$$\delta_g = b\delta. \tag{3.7}$$

This is generally a good approximation on quasi-linear scales. Its validity is tested in Section 3.6.

3.3 The data

Among the spectroscopic galaxy surveys available today, the Seventh Data Release of the Sloan Digital Sky Survey (SDSS DR7) covers the largest volume (Abazajian *et al.*, 2009). The LRG sample is generally regarded as a good cosmological probe. Properties of the LRGs can be found in Eisenstein *et al.* (2001). In practice, flags in the Sloan database identify these galaxies. For CIC, one needs a volume-limited sample which can be obtained by magnitude and redshift cuts. We restricted our LRG sample to redshifts between 0.24 and 0.31 and k-corrected absolute magnitudes between -22.3 and -24.3 in the r band. We used values from the Photoz and SpecObjAll tables on the SkyServer Web site. By excluding the three stripes in the Southern Galactic Cap, we were left with

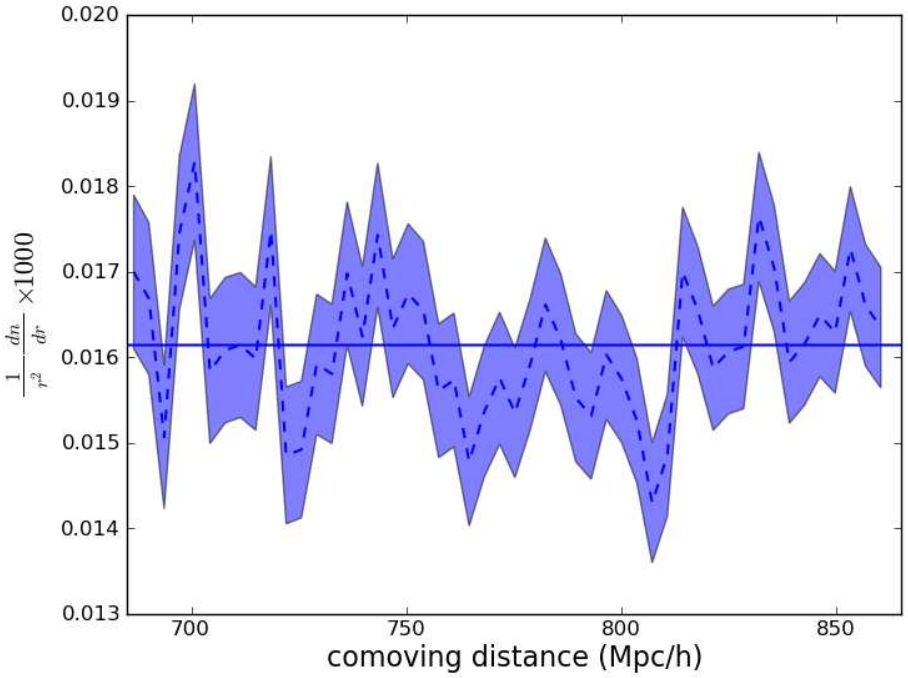

Figure 3.1 The angle-averaged density function of the data we specify in Section 3.3. The shaded area shows the 1σ level of the Poisson noise. The plot supports the assumption that the selection function has an insignificant radial dependence, since no large-scale trend can be observed.

21613 galaxies. After converting redshift into comoving radial distance by using the best *WMAP* cosmological parameters, the angle-averaged density appears to be uniform with fluctuations consistent with Poisson noise, as can be seen in Figure 3.1. The selection function of a similar data set is plotted in Figure 12 in Eisenstein *et al.* (2001).

3.4 The algorithm

From R.A., decl., and z coordinates, we calculated comoving Cartesian coordinates. Then we placed a rectangular grid over the sample. Since in this arrangement cubical cells are readily accessible, we chose to measure CIC in cubes. The survey mask (Figure 3.2),

however, has a complex shape, usually given by spherical polygons. Holes and the irregular boundary cause unwanted edge effects which could bias the results in a complicated way. The mask of the SDSS DR7 spectroscopic survey is plotted in Figure 3.2.

To tackle this problem, first we took a cube-shaped region encompassing the survey. Then we created two negatives by filling the parts in the mask and outside the survey area with dummy galaxies from a Poisson point process; one with the average density of the survey and one with hundred times that density. We added the first negative to the survey to fill the holes. We measured CIC in this and in the second negative in parallel. Since its density is large, the counts from the second negative provide a good measure of the overlap of the cells with the survey geometry. We ignored any cell that had more than 10% of its volume outside, which corresponded to having a galaxy count in the second negative larger than 100 $\times$ average density $\times$ volume of cell $\times$ 0.1. In this work we used MANGLE (Swanson *et al.*, 2008) to check whether an object was inside the mask.

3.5 The systematics

The shape of the survey and diluting the data with a Poisson point process introduce systematic bias into our measurements of CIC. In order to assess its level, we studied simulations. We created mock catalogs with a second-order Lagrangian (2LPT) code (Crocce *et al.*, 2006). We created 100 mock catalogs in $2500\ h^{-1}$Mpc cubes, then we used these to create another set of mocks by applying the mask of the spectroscopic survey. The galaxies were downsampled in every case to match the average density with that of the data. With these and the two negatives described in Section 3.4, we were able to measure CIC in three different arrangements.

(i) Ideal, large, cubic-shaped simulations.

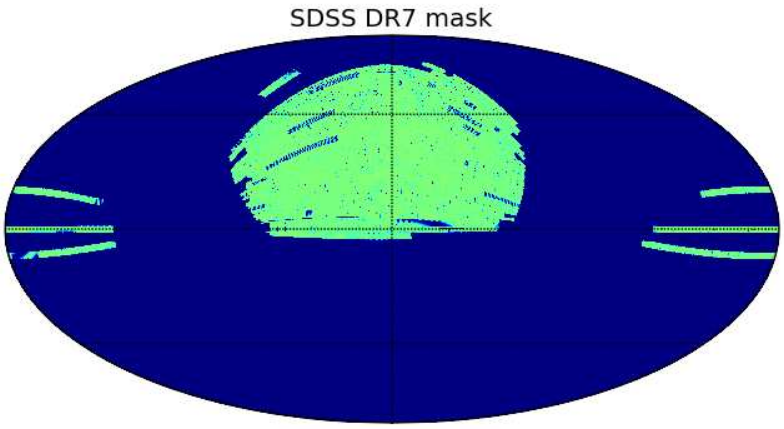

Figure 3.2 Mask, angular selection function, of the SDSS seventh data release spectroscopic sample. The data are presented in equatorial coordinates in HEALPix pixelization (Górski *et al.*, 2005) and retrieved from the MANGLE Web site (Swanson *et al.*, 2008).

(ii) Variant of (i). We only took into account a cell when at least 90% of its volume lay inside the survey area.

(iii) Dropping the galaxies which are outside the survey area and filling the empty space with the dummy galaxies of the negative to preserve the average density. (The 90% rule still applies.)

First, we estimated the cosmic bias due to the survey shape and volume. This question has been studied extensively (see, e.g., Szapudi and Colombi, 1996). We measured the variance, S_3 and S_4 in the first two arrangements (i) and (ii). In Figure 3.3 the ratio, $\delta A/A$, is plotted for each of these quantities, where A is the quantity measured in arrangement (i) . Error bars were estimated from the scatter around the average. The error of the average is plotted, so the error of a single measurement is ten times larger. In the case of the variance, this is the well-understood integral constraint problem and the ratio does not exceed a couple of percent even at the largest scale. For S_3, the ratio is consistent with 1 but for S_4 the difference from 1 is not negligible even at relatively small scales. However, as we show later, this bias is still small compared to the cosmic error.

The bias, caused by the data having been diluted with a random sample, can be understood in the case that evenly distributed holes comprise the mask. The resulting catalog can be considered the linear combination of the Poisson sampled galaxy and a constant field. This constant field is the random sample that fills the mask:

$$\rho = \rho_{galaxy} + \rho_{random}. \tag{3.8}$$

Subsequently the density contrast can be written as

$$\delta = y\delta_{galaxy} + (1-y)\delta_{random} \tag{3.9}$$

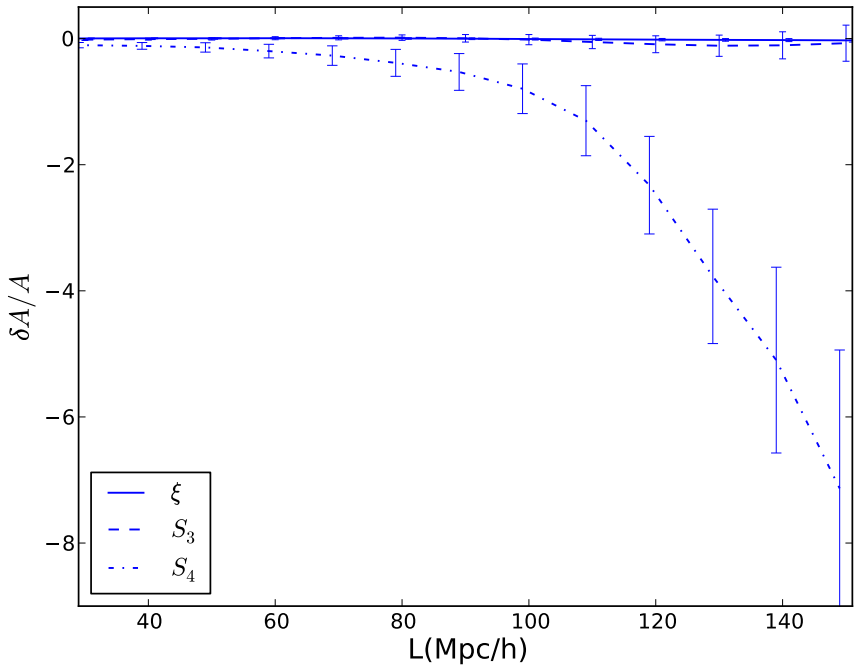

Figure 3.3 Systematic bias due to the finite volume of the survey as a function of the cell size. The difference between $\overline{\xi}$, S_3, and S_4 measured in a $2500^3 h^{-1}$Mpc simulation, arrangement (i), and in one with the volume and shape of the data we described in Section 3.3, arrangement (ii). The differences are normalized by values from arrangement (i). The uncertainty is calculated from 100 realizations. The bias in the case of $\overline{\xi}$ and S_3 is consistent with 0, while for S_4, it becomes more than 100% around $100h^{-1}$Mpc.

with

$$y = \frac{\overline{\rho}_{galaxy}}{\overline{\rho}}. \tag{3.10}$$

Since δ_{random} is zero the moments of δ are proportional to the moments of δ_{data} :

$$\overline{\xi} = y^2 \overline{\xi}_{galaxy}, \tag{3.11}$$

$$S_n = y^{-n+2} S_n^{galaxy}. \tag{3.12}$$

We used arrangement (ii) and (iii) to express y according to Equation (3.11) and Equation (3.12) (Figure 3.4). The measured values are consistent with our assumption, Equation (3.10). The robustness of this simple model is due to the fact that the variance and S_n are insensitive to small changes in the cell shape (e.g., in the case of a power-law correlation function the S_ns are constants, see Peebles 1980 or Boschan *et al.* 1994; Szapudi 1998 for a study of on this). This bias can be corrected for by measuring y directly.

As the next step we added redshift distortions to the mock catalogs. The effect on the variance is expected to be similar to the effect on the monopole of the two-point function. S_3 and S_4 are affected less according to, e.g., Hivon *et al.* (1995). While the Kaiser formula (Kaiser, 1987; Hamilton, 1992) gives a good description of this in the linear regime. The lowest order of the three-point function in Fourier space has been worked out by Scoccimarro *et al.* (1999) but Fourier transforming it back to redshift space is infeasible in general. Higher moments are gradually harder to compute. For these reasons we follow a phenomenological approach. In Figure 3.5, the ratios of redshift and real space values of the variance, S_3, and S_4 are plotted. The thick line is the predicted amplification from

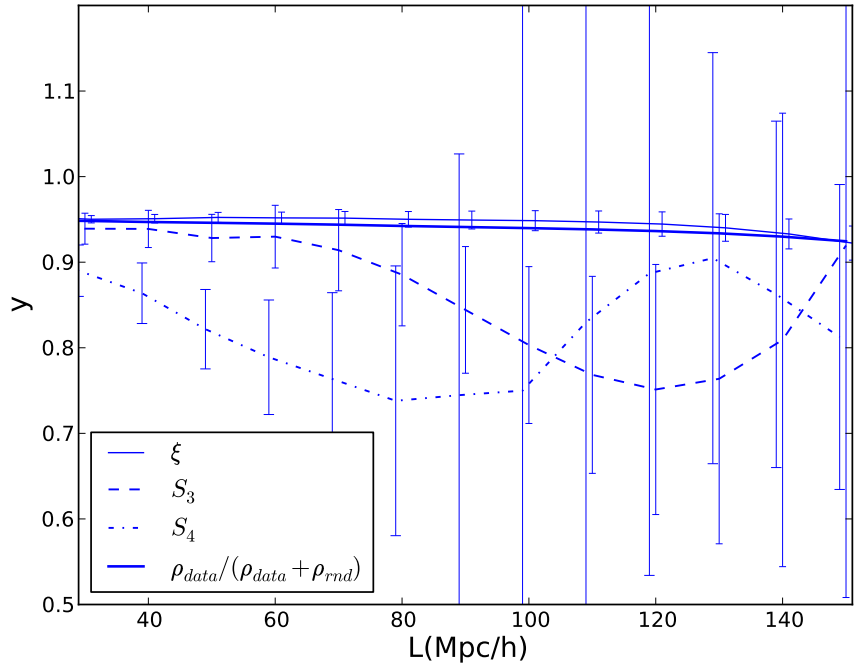

Figure 3.4 y as a function of cell size from Equation (3.11) and (3.12). y is a heuristic value that aims to capture the systematic bias due to using dummy galaxies to fill the holes in the survey. All curves are to match the thick solid line, which is simply the ratio of the dummy galaxies to the total number of galaxies.

the Kaiser formula for the variance:

$$\overline{\xi}_{RS} = (1 + 2f/3 + f^2/5)\overline{\xi}, \tag{3.13}$$

with

$$f = \Omega^{0.6}/b, \tag{3.14}$$

where the subscript RS stands for redshift space. In the simulations b is 1. In this chapter, we assume that the effect of redshift distortion is small compared to the cosmic errors in the case of S_3 and S_4, and that the variance is amplified according to Equation (3.13).

In practice, if the systematic bias is small compared to the cosmic error then it is negligible. In Figure 3.6, the total systematic bias after corrections according to Equation (3.11) and (3.12) and the cosmic error are plotted. It can be concluded that the proposed corrections are sufficient to measure the variance, S_3, and S_4 with an error that is not significantly different from the cosmic error. This plot also tells us that the signal-to-noise ratio drops below 1 around $100\ h^{-1}$Mpc for S_3 and around $50\ h^{-1}$Mpc for S_4, so they cannot be measured reliably beyond these scales.

3.6 The variance, S_3, and S_4

We measured CIC in the SDSS DR7 spectroscopic LRG sample and compared the prediction of ΛCDM to our results.

For measuring the variance, S_3, and S_4 we followed the procedure outlined in Section 3.4 and we corrected for the systematic bias as given by Equation (3.11) and (3.12). We determined the y parameter from the simulations.

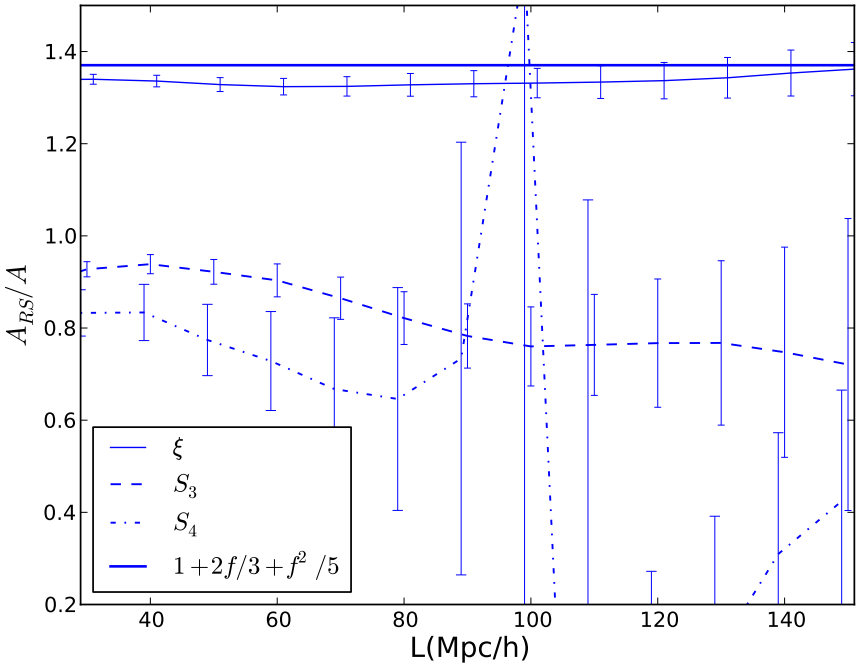

Figure 3.5 Effect of redshift distortion on the variance (ξ), S_3, and S_4 as a function of the cell size. The redshift distortion is added to the 2LPT simulations by hand. The variance agrees with the theoretical values within uncertainty. Due to the large variance of S_4, its inverse can explode.

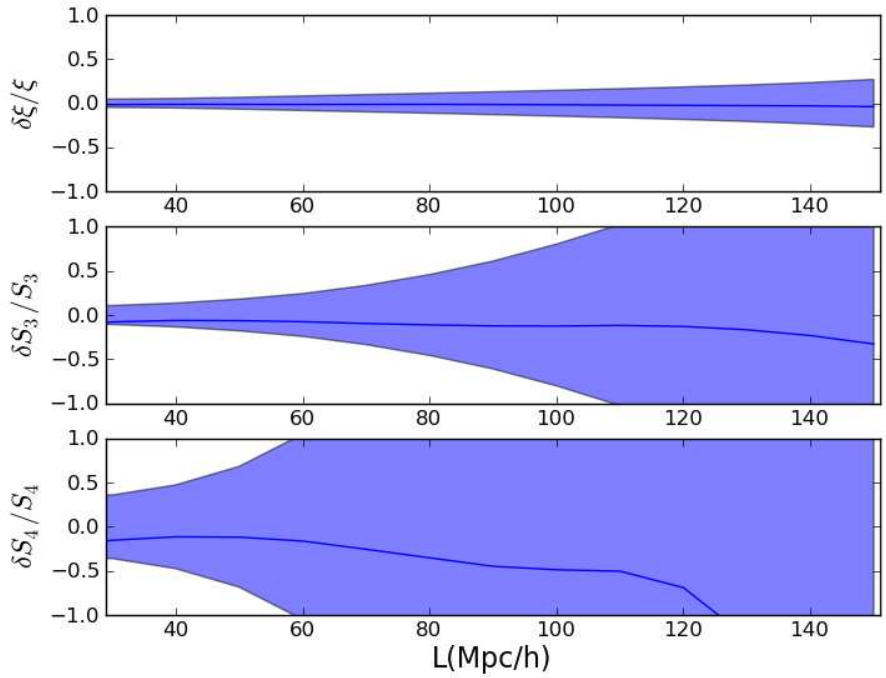

Figure 3.6 Total expected systematic bias in a survey similar to the one discussed in Section 3.3 is plotted after corrections as discussed in Section 3.5. The shaded region represents the cosmic error.

Additionally, in the case of the real data one has to assume a galaxy–dark-matter biasing scheme. We used the simplest linear local model as given by Equation (3.7). We computed the bias parameter from fitting the variance and testing its consistency on S_3 and S_4. We defined the following chi-square:

$$\chi^2(b) = (\overline{\xi}_d - \overline{\xi}_{th}(b))C^{-1}(\overline{\xi}_d - \overline{\xi}_{th}(b)). \tag{3.15}$$

Here $\overline{\xi}_d$ and $\overline{\xi}_{th}$ stand for the measured and the theoretical variance. For the theory, we used the real space linear model and we assumed that it transforms to redshift space as the monopole of the two-point function (see Equation (3.13)). The measured variance was rescaled as in Equation (3.11) and extrapolated to present day ($z = 0$) using the growth function (see, e.g., Dodelson 2003). The covariance matrix was calculated from mock catalogs described in Section 3.5:

$$C_{ij} = \frac{1}{N-1}\sum_n(\overline{\xi}_i^n - \overline{\overline{\xi}_i})(\overline{\xi}_j^n - \overline{\overline{\xi}_j}), \tag{3.16}$$

$$\overline{\overline{\xi}_j} = \frac{1}{N}\sum_n \overline{\xi}_j^n, \tag{3.17}$$

where the superscript n refers to the nth simulation and N is the total number of simulations, in this case 100. These simulations are in redshift space and with bias equal to 1. After finding the minimum of the chi-square, the covariance matrix was rescaled according to Equation (3.13).

For our fit, we used the range from $30\ h^{-1}$Mpc to $150\ h^{-1}$Mpc where the linear theory is generally assumed to be valid. The result is $b = 2.14^{+0.13}_{-0.14}$ with 1σ uncertainty, which is consistent with findings of Okumura *et al.* (2008), who used a very similar data set. The best fitting variance is plotted with the data in the upper panel of Figure 3.7. When we

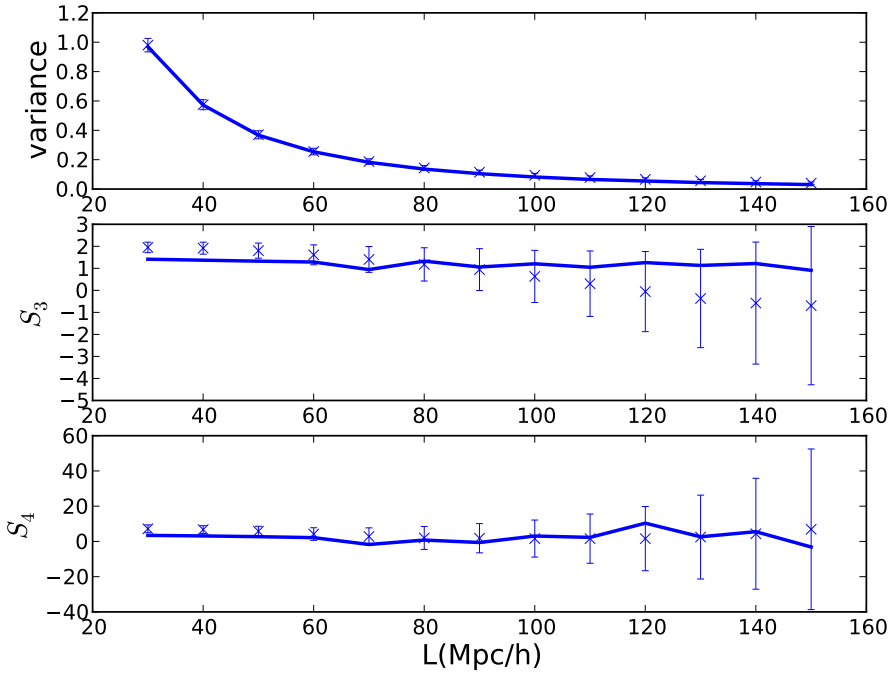

Figure 3.7 Measured variance, S_3, and S_4 from the SDSS spectroscopic LRG subsample (see Section 3.3 for a description of the data) with the theoretical predictions (solid line) vs. the cell size. L is the size of a cubic cell.

changed the boundaries of the range to $50\ h^{-1}$Mpc and $130\ h^{-1}$Mpc, we found that the change in the best b was consistent, only $+0.03$.

We also measured S_3 and S_4 and applied Equation (3.12). In Figure 3.7 these are plotted along with the prediction of linear ΛCDM (Juszkiewicz *et al.*, 1993; Bernardeau, 1994):

$$bS_3 = \frac{34}{7} + \gamma_1, \tag{3.18}$$

$$b^2 S_4 = \frac{60712}{1323} + \frac{62}{3}\gamma_1 + \frac{7}{3}\gamma_1^2 - \frac{2}{3}\gamma_2 \tag{3.19}$$

with

$$\gamma_i = \frac{d\log^i \overline{\xi}}{d\log r^i}. \tag{3.20}$$

These quantities are all derivatives of the two-point function that we calculated from the matter power spectrum generated by CAMB (Lewis *et al.*, 2000) from the proper input cosmological parameters. The average of the correlation function, $\overline{\xi}$, was calculated via Monte Carlo simulations for cells of $30\ h^{-1}$Mpc $+\Delta, 40\ h^{-1}$Mpc $+\Delta, ..., 150\ h^{-1}$Mpc $+\Delta$, where Δ is 0 or ± 3. The γ values were estimated using discrete derivatives of $\overline{\xi}$. The data values from the plot are collected in Table 3.1.

We tested the goodness of the theory by calculating a covariance matrix with the shrinkage technique (Pope and Szapudi, 2008) with a diagonal of the empirical covariance matrix (C) as the target covariance matrix (T):

$$\tilde{C}_{ij} = \lambda T_{ij} + (1-\lambda)C_{ij} \tag{3.21}$$

$r(h^{-1}Mpc)$	ξ	$\Delta\xi$	S_3	ΔS_3	S_4	ΔS_4
30	0.980	0.046	1.951	0.238	7.176	2.225
40	0.574	0.034	1.916	0.273	6.701	2.234
50	0.370	0.027	1.804	0.343	5.821	2.770
60	0.256	0.022	1.613	0.452	4.191	3.581
70	0.187	0.019	1.404	0.586		
80	0.144	0.017	1.179	0.755		
90	0.115	0.015	0.942	0.951		
100	0.094	0.013				
110	0.078	0.011				
120	0.066	0.010				
130	0.056	0.009				
140	0.048	0.008				
150	0.040	0.008				

Table 3.1 The Numerical Values of ξ, S_3, and S_4 with 1σ Uncertainty. We leave the fields blank when the signal-to-noise ratio is less than 1.

A recipe to calculate λ is given in Pope and Szapudi (2008). This method ensures that we get a well-behaving covariance matrix. We estimated the significance for the variance, S_3 and S_4 separately and jointly. The results are in Table 3.2, showing a good overall agreement with our ΛCDM model with linear bias. Visually, S_3 and S_4 appear to be slightly larger than expected on smaller scales; however, this is not statistically significant. On large scales, there appears to be a slight excess power on $130 - 150\ h^{-1}$Mpc scales in the variance (not apparent in Figure 3.7). This is not large enough to influence the ISW effect, and its significance is only 2σ according to Table 3.2. In summary, ΛCDM is a good fit for all the moments we measured. Note that the shrinkage estimator ($\lambda = 0.0006$) gave identical results for the covariance matrix of the variance to that given by Equation (3.17).

data	λ	p
ξ	0.0006	0.05
S_3	0.13	0.80
S_4	1.	0.83
Joint	1.	0.85

Table 3.2 Testing the Linear Model with Shrinkage Techique as in Pope and Szapudi (2008).

3.7 Discussion

We measure CIC in a volume-limited subsample of the SDSS DR7 spectroscopic LRG data. Since there is no clear recipe for how to measure CIC best in a survey with a complicated mask, we study simulations first. We identify three sources of systematic bias: the finite volume effect, survey mask, and redshift distortions. We compare the variance, S_3 and S_4 to their values in ΛCDM. Despite the fact that we use the lowest order approximation, we find agreement between data and prediction (Figure 3.7). However, it cannot be excluded that more complex models can fit the data better, especially on nonlinear scales. As can be seen from Figure 3.7, S_3 and S_4 differ slightly from the lowest order predictions possibly due to nonlinearities, although this difference is not statistically significant.

We find no sign of excess power or a fat tail in the case of the PDF. However, the data still allow large uncertainty in higher-order moments. The useful volume of the survey needs to be increased in order to get tighter error bars. This could be achieved by relaxing the requirements for the selection function, which introduces another source of systematic bias, or with future spectroscopic surveys with larger sky coverage and depth, such as BOSS and BigBOSS.

References

Abazajian, K. N., J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, C. Allende Prieto, D. An, K. S. J. Anderson, S. F. Anderson, J. Annis, N. A. Bahcall, C. A. L. Bailer-Jones, J. C. Barentine, *et al.*, 2009, ApJS **182**, 543.

Bernardeau, F., 1994, A&A **291**, 697.

Boschan, P., I. Szapudi, and A. S. Szalay, 1994, ApJS **93**, 65.

Broadhurst, T. J., R. S. Ellis, D. C. Koo, and A. S. Szalay, 1990, Nature **343**, 726.

Colombi, S., I. Szapudi, A. Jenkins, and J. Colberg, 2000, MNRAS **313**, 711.

Crocce, M., S. Pueblas, and R. Scoccimarro, 2006, MNRAS **373**, 369.

Cruz, M., E. Martínez-González, P. Vielva, and L. Cayón, 2005, MNRAS **356**, 29.

Dodelson, S., 2003, *Modern cosmology* (Academic Pr), ISBN 0122191412.

Eisenstein, D. J., J. Annis, J. E. Gunn, A. S. Szalay, A. J. Connolly, R. C. Nichol, N. A. Bahcall, M. Bernardi, S. Burles, F. J. Castander, M. Fukugita, D. W. Hogg, *et al.*, 2001, AJ **122**, 2267.

Eisenstein, D. J., I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H. Seo, M. Tegmark, Z. Zheng, S. F. Anderson, J. Annis, *et al.*, 2005, ApJ **633**, 560.

Geller, M. J., and J. P. Huchra, 1989, Science **246**, 897.

Górski, K. M., E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, 2005, ApJ **622**, 759.

Gott, J. R., III, M. Jurić, D. Schlegel, F. Hoyle, M. Vogeley, M. Tegmark, N. Bahcall, and J. Brinkmann, 2005, ApJ **624**, 463.

Granett, B. R., M. C. Neyrinck, and I. Szapudi, 2008, ApJ **683**, L99.

Hamilton, A. J. S., 1992, ApJ **385**, L5.

Hivon, E., F. R. Bouchet, S. Colombi, and R. Juszkiewicz, 1995, A&A **298**, 643.

Juszkiewicz, R., F. R. Bouchet, and S. Colombi, 1993, ApJ **412**, L9.

Kaiser, N., 1987, MNRAS **227**, 1.

Kaiser, N., and J. A. Peacock, 1991, ApJ **379**, 482.

Kulkarni, G. V., R. C. Nichol, R. K. Sheth, H. Seo, D. J. Eisenstein, and A. Gray, 2007, MNRAS **378**, 1196.

Lewis, A., A. Challinor, and A. Lasenby, 2000, ApJ **538**, 473.

Nichol, R. C., R. K. Sheth, Y. Suto, A. J. Gray, I. Kayo, R. H. Wechsler, F. Marin, G. Kulkarni, M. Blanton, A. J. Connolly, J. P. Gardner, B. Jain, *et al.*, 2006, MNRAS **368**, 1507.

Okumura, T., T. Matsubara, D. J. Eisenstein, I. Kayo, C. Hikage, A. S. Szalay, and D. P. Schneider, 2008, ApJ **676**, 889.

Peebles, P., 1980, *The large-scale structure of the universe* (Princeton Univ Pr), ISBN 0691082405.

Pope, A. C., and I. Szapudi, 2008, MNRAS **389**, 766.

Scoccimarro, R., H. M. P. Couchman, and J. A. Frieman, 1999, ApJ **517**, 531.

Spergel, D. N., R. Bean, O. Doré, M. R. Nolta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, L. Verde, *et al.*, 2007, ApJS **170**, 377.

Swanson, M. E. C., M. Tegmark, A. J. S. Hamilton, and J. C. Hill, 2008, MNRAS **387**, 1391.

Szapudi, I., 1998, ApJ **497**, 16.

Szapudi, I., and S. Colombi, 1996, ApJ **470**, 131.

Szapudi, I., S. Colombi, A. Jenkins, and J. Colberg, 2000, MNRAS **313**, 725.

Szapudi, I., and A. S. Szalay, 1993, ApJ **408**, 43.

Chapter 4

The integrated Sachs-Wolfe effect of superstructures

4.1 Introduction

After last scattering, photons traveled through mostly neutral media. Although radiation and matter are not strongly coupled, there is still a secondary signal due to large scale structure on top of the primary fluctuations of the cosmic microwave background (CMB) radiation. The integrated Sachs-Wolfe (ISW) effect (Sachs and Wolfe, 1967) accounts for most of the secondary anisotropies for low multipoles (Hu and Dodelson, 2002). As the expansion of the universe accelerates, gravitational potential wells and hills decay. Photons traversing these get blueshifted or redshifted, respectively

Due to its weak signal, ISW detection is challenging. Cross-correlating galaxy surveys with CMB maps yields results from marginally significant (Scranton *et al.*, 2003; Afshordi *et al.*, 2004; Padmanabhan *et al.*, 2005; Raccanelli *et al.*, 2008; Sawangwit *et al.*, 2010) to 4.5σ detections (Giannantonio *et al.*, 2008; Ho *et al.*, 2008). The higher significance was achieved by a joint analysis of surveys. Other techniques focusing on the signal from

discrete objects can reach up to 4.5σ from a single survey (McEwen *et al.*, 2008; Granett *et al.*, 2008).

The ISW effect can be a unique probe of dark energy if well-measured. From cross-correlation measurements and the *Wilkinson Microwave Anisotropy Probe* (*WMAP*) power spectrum (Bennett *et al.*, 2003), it has already been possible to constrain cosmological parameters (Giannantonio *et al.*, 2008; Ho *et al.*, 2008). Despite the fact that the detection of the signal from discrete objects has higher significance, they cannot be used for parameter estimation due to the lack of simple quantitative models.

Further motivation for studying superstructures stems from anomalies in the low l modes of the CMB (Tegmark *et al.*, 2003; Copi *et al.*, 2004). Inoue and Silk (2007) calculate the effect of large, dust filled, compensated voids in the local universe. They were successful in explaining the observed CMB anomalies but these voids, due to their size and depth, do not fit into the widely accepted picture of clustering. They assume extra power on large scales in the matter power spectrum. When subtracting the estimated local ISW signal from CMB maps, Francis and Peacock (2010) found that the significance of the anomalies decreased.

In this chapter, our principal goal is to estimate the ISW signal from large overdense or underdense regions (superclusters or supervoids). We use the results of the CIC measurement from the previous chapter to estimate the PDF focusing on its tail, since an enhanced tail could explain a strong ISW signal from superstructures. Since no significant non-Gaussianity has been found on these scales, we use a simple Gaussian model to derive an expression for their profile and the potential. We compare the estimated ISW signal to the results of Granett *et al.* (2008). More about their ISW measurement can be found in Section 4.2 of this chapter. In Section 4.5, we summarize and discuss our results.

4.2 The ISW signature of SDSS superstructures (a recap of the results of Granett et al.)

Granett *et al.* (2008) identified large underdense and overdense regions (supervoids and superclusters) in the SDSS DR4 photometric LRG sample. They stacked images, cut out from the CMB, centered on the directions of 50 supervoids and 50 superclusters found with the highest significance.

The signal is consistently present in every frequency band, so it is likely to have a cosmological origin. In this section we scrutinize the tail of the density distribution and the density profile of supervoids or superclusters. The question we ask is: what is the expected ISW signal produced by the 50 objects with the highest and lowest densities in a survey similar to the one in Granett *et al.* (2008)?

In order to answer this, we have to have an estimate for tail of the matter PDF. We show that by using simple analytic functions to approximate the galaxy PDF this estimate can be made robustly for our purposes. We also revise the way in which the ISW signal from these density extrema is estimated.

4.3 The matter PDF

If the galaxy PDF is known, a simple convolution with a Poisson-distribution gives the galaxy counts:

$$P(N) = \int \frac{\left\langle N \right\rangle^N (1+\delta_g)^N}{N!} e^{-\left\langle N \right\rangle (1+\delta_g)} P(\delta_g) d\delta_g. \tag{4.1}$$

On large scales one can approximate $P(\delta_g)$ as Gaussian, lognormal or second-order Edgeworth expansion (see, e.g., Kim and Strauss 1998; Szapudi and Pan 2004). The first

two depend on the variance only, whereas for the Edgeworth expansion we need S_3 and S_4 as well:

$$P_G(\delta_g) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\delta_g^2/2\sigma^2} \tag{4.2}$$

$$P_{LN}(\delta_g) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(ln(1+\delta_g)+\tilde{\sigma}^2/2)^2/2\tilde{\sigma}^2}/(1+\delta_g) \tag{4.3}$$

$$\begin{aligned} P_E(\delta_g) = \quad & \frac{1}{\sqrt{2\pi}\sigma} e^{-\delta_g^2/2\sigma^2} \big(1 + \frac{\sigma S_3}{6} H_3(\delta_g/\sigma) \\ & + \frac{\sigma^2 S_4}{24} H_4(\delta_g/\sigma) + \frac{10\sigma^2 S_3^2}{720} H_6(\delta_g/\sigma)\big), \end{aligned} \tag{4.4}$$

where $\sigma^2 = \overline{\xi}$, $\tilde{\sigma}^2 = \sqrt{(1+\overline{\xi})}$ and H_i is the ith Hankel-function. The above approximations are plotted in Figure 4.1 for galaxy counts in cubes with a linear size of $100\ h^{-1}$Mpc from the spectroscopic LRG sample. The best match around the tails is the Edgeworth expansion, whereas around the mean all of these approximations provide qualitatively similar results. The left panel of Figure 4.1 shows the underlying continuous galaxy PDFs, the deconvolved CIC distribution, whereas the right shows them convolved with the Poisson-distribution and the galaxy counts.

To estimate the density of the 50 supervoids and 50 superclusters of a certain size, we generated V_{survey}/V_{cell} random numbers with the PDFs given above, where V_{survey} refers to the volume of the survey and V_{cell} refers to the volume of a superstructure. This approximation slightly overestimates the number of independent cells. Then we stored the lowest and highest 50 of the numbers. Repeating this several times gives the distribution of the extrema. This method, however, is too slow for calculating a covariance matrix from hundreds of simulations. Since we only deal with linear ISW here, as we explain in the next subsection, we need only the mean of the extrema rather than their whole

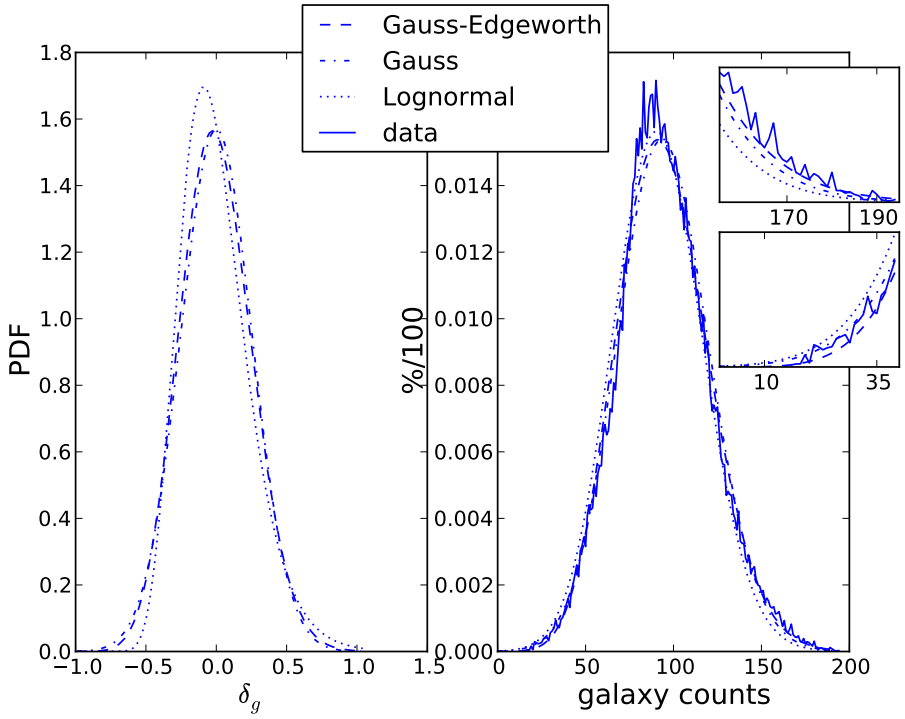

Figure 4.1 In the left panel, three analytic functions approximating the galaxy PDF for cells of $100\ h^{-1}$Mpc are plotted (for details, see the text). In the right panel, these are plotted again convolved with a Poisson noise (Equation (4.1)). The data are the solid line in the right panel.

distribution. A satisfactory approximation for the mean in the case of voids is:

$$\overline{\delta_g} = \int_{-\infty}^{\delta_{max}} P(\delta_g)\delta_g d\delta_g, \tag{4.5}$$

where δ_{max} is given by $50V_{cell}/V_{survey} = \int_{-\infty}^{\delta_{max}} P(\delta_g)d\delta_g$. For clusters, the limits of the integration change to $\int_{\delta_{max}}^{\infty}$. In our tests we found the difference to be about a few percent. As we use the linear bias model, a further division by the bias yields δ.

4.4 The profile of supervoids and superclusters and the ISW Signal

4.4.1 The average profile of superstructures

In general, the ISW effect is determined by an integral along the path of a CMB photon (Sachs and Wolfe, 1967):

$$\frac{\Delta T}{T} = -\frac{2}{c^2}\int d\tau \frac{\partial \Phi(r(\tau),\tau)}{\partial \tau}, \tag{4.6}$$

where τ denotes the conformal time.

According to Rudnick *et al.* (2007), a simple estimate of the linear ISW effect, an underdense or an overdense spherical region at redshift z causes a

$$\frac{\Delta T}{T} \approx \Omega_m \left(\frac{r_c}{c/H_0}\right)^3 (1+2z)(1+z)^{-2}\delta \tag{4.7}$$

temperature shift at its center on the CMB, where r_c is the comoving radius of the sphere. In the derivation, the authors approximated the potential with a top hat which implies a compensated void or cluster. Here, instead, we propose a profile motivated by Gaussian

statistics. For a spherically symmetric object, the Newtonian potential can be calculated easily as:

$$\Phi(r) = -\frac{3\Omega_m}{8\pi}\left(\frac{H_0}{c}\right)^2 \int_r^\infty \frac{M(\tilde{r})}{\tilde{r}^2} d\tilde{r}, \tag{4.8}$$

where $M(r) = 4\pi \int_0^r d\tilde{r}\tilde{r}^2\delta(\tilde{r})$. The density contrast at distance r from the center can be obtained with the condition that the average density inside r_c is known:

$$P(\delta_{in}) = \frac{1}{\sqrt{2\pi\langle\delta_{in}^2\rangle}} \exp\left(-\frac{\delta_{in}^2}{2\langle\delta_{in}^2\rangle}\right), \tag{4.9}$$

$$P(\delta(r), \delta_{in}) = \frac{1}{\sqrt{2\pi|C|}} \exp\left(-\frac{1}{2}\vec{\delta}C^{-1}\vec{\delta}\right), \tag{4.10}$$

$$P(\delta(r)|\delta_{in}) = P(\delta(r), \delta_{in})/P(\delta_{in}) \tag{4.11}$$

where $\langle ... \rangle$ refers to unconditional ensemble averaging and δ_{in} is the average density measured inside r_c. To simplify the notation, we use $\vec{\delta} = (\delta(r), \delta_{in})$ and $C = \langle \vec{\delta} \otimes \vec{\delta} \rangle$. In practice, $\langle ... \rangle$ can be computed with Monte Carlo simulations robustly. From these, the expected density is

$$\langle \delta(r) \rangle_{\delta_{in}} = \frac{\langle \delta(r)\delta_{in} \rangle}{\langle \delta_{in}^2 \rangle} \delta_{in}. \tag{4.12}$$

See the Appendix for details. This profile is not compensated inside a finite radius.

4.4.2 The uncertainty of the profile and the ISW effect

It is useful to calculate the uncertainty of the profile in order to get an estimate of the uncertainty of the potential and the ISW effect. We would like to point out that for a correct treatment of the potential one should drop the assumption of spherical symmetry.

We chose to optimize the accuracy and speed by keeping the spherical approximation. We use the Gaussian model as before:

$$P(\delta(r_1), \delta(r_2), \delta_{in}) = \frac{1}{\sqrt{2\pi|C_3|}} \exp\left(-\frac{1}{2}\vec{\delta}_3 C_3^{-1} \vec{\delta}_3 \right) \quad , \tag{4.13}$$

$$P(\delta(r_1), \delta(r_2)|\delta_{in}) = P(\delta(r_1), \delta(r_2), \delta_{in})/P(\delta_{in}), \tag{4.14}$$

where $\vec{\delta}_3 = (\delta(r_1), \delta(r_2), \delta_{in})$ and $C_3 = \left\langle \vec{\delta}_3 \otimes \vec{\delta}_3 \right\rangle$. From this, the covariance between shells at r_1 and r_2 is:

$$\begin{aligned} Cov(r(1), r(2)) = & \left\langle \delta(r_1)\delta(r_2) \right\rangle_{\delta_{in}} - \left\langle \delta(r_1) \right\rangle_{\delta_{in}} \left\langle \delta(r_2) \right\rangle_{\delta_{in}} \\ = & \frac{1}{\left\langle \delta_{in}^2 \right\rangle} \Big[\left\langle \delta(r_1)\delta(r_2) \right\rangle \left\langle \delta_{in}^2 \right\rangle \\ & - \left\langle \delta(r_1)\delta_{in} \right\rangle \left\langle \delta(r_2)\delta_{in} \right\rangle \Big], \end{aligned} \tag{4.15}$$

In the upper panel of Figure 4.2, we plot $M(r)$ for $r_c = 100/(4\pi/3)^{1/3} \mathrm{h}^{-1}\mathrm{Mpc}$. The curve is normalized so that $M(r_c) = \frac{4\pi r_c^3}{3}$. The error bars are:

$$\Delta M(r) = \sqrt{(4\pi)^2 \int_0^r \int_0^r Cov(r_1, r_2) r_1^2 r_2^2 dr_1 dr_2} \quad . \tag{4.16}$$

On the lower panel of Figure 4.2, we plot the potential of such an object as in Equation (4.8), along with the top hat potential from Rudnick *et al.* (2007).

Since the potential is only Equation (4.8) and in linear theory its time dependence is relatively simple $\Phi(r, \tau) = \Phi(r)\frac{D(\tau)}{a(\tau)}$, it is straightforward to integrate Equation (5.5) numerically. In our calculations we placed the density fluctuation at the median redshift $(z = 0.53)$ of the SDSS photometric LRG survey so that we can compare our result to actual measurements (Granett *et al.*, 2008). In Figure 4.3, we plot ΔT versus the

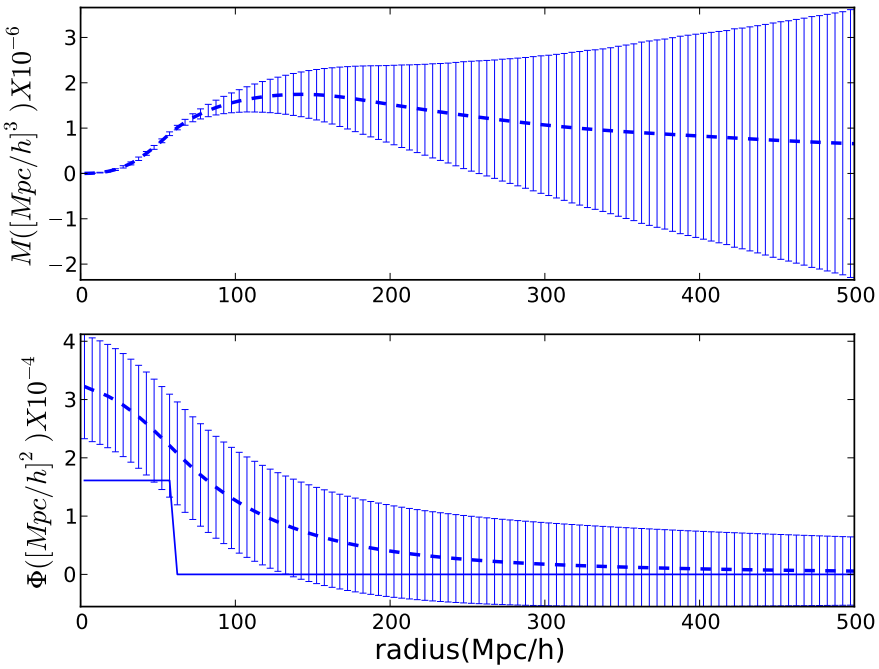

Figure 4.2 In the upper panel, $M(r) = 4\pi \int_0^r drr^2\delta(r)$ is plotted for a supercluster with $\delta_{in} = 1$ for $r_c = 62\ h^{-1}$Mpc. In the lower panel, we plot the potential (dashed line) and the top hat potential (see the text for details).

comoving radius (r_c) of the superstructure, we also plot ΔT according to Equation (4.7). Here, we used $\delta_{in} = 1$. It is clear that Equation (4.7) underestimates the ISW effect.

In Figure 4.4, we used Equation (4.5) to calculate δ_{in}. The PDF was measured in an SDSS LRG subsample at $z = 0.28$ median redshift (see 3.3 for details) and scaled to the subsample described in Granett *et al.* (2008), which is located at $z = 0.53$. This means the scaling of the variance, S_3 and S_4 according to linear dynamics. The value for V_{survey} came from the properties of the survey in Granett *et al.* (2008). All three assumptions in Equations (4.2–4.4) about the density distribution give similar results. We also plot an estimate based on raw data without deconvolution. The dashed line is the ISW effect according to Equation (4.7). We plotted both the supervoids and superclusters. The

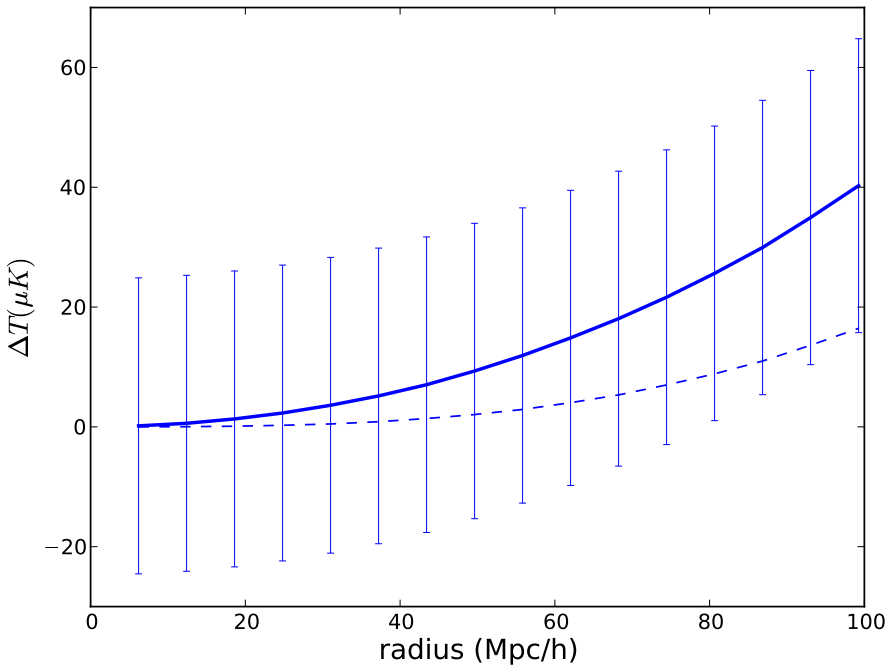

Figure 4.3 ΔT for a photon traveling through the center of a supervoid at redshift 0.52 with $\delta_{in} = 1$ against its radius is plotted (solid line). The dashed line comes from Equation (4.7), an approximation using compensated profile.

intrinsic fluctuations of the matter density (see Figure 4.3) and the uncertainty of the tail of the PDF add up. On the horizontal axis, the scale is the linear size of the cell we measured CIC in.

4.5 Discussion

We used the the first few moments of the matter PDF from Chapter 3 and give an estimate of the linear ISW effect owing to the largest density fluctuations on $100\ h^{-1}$Mpc scales.

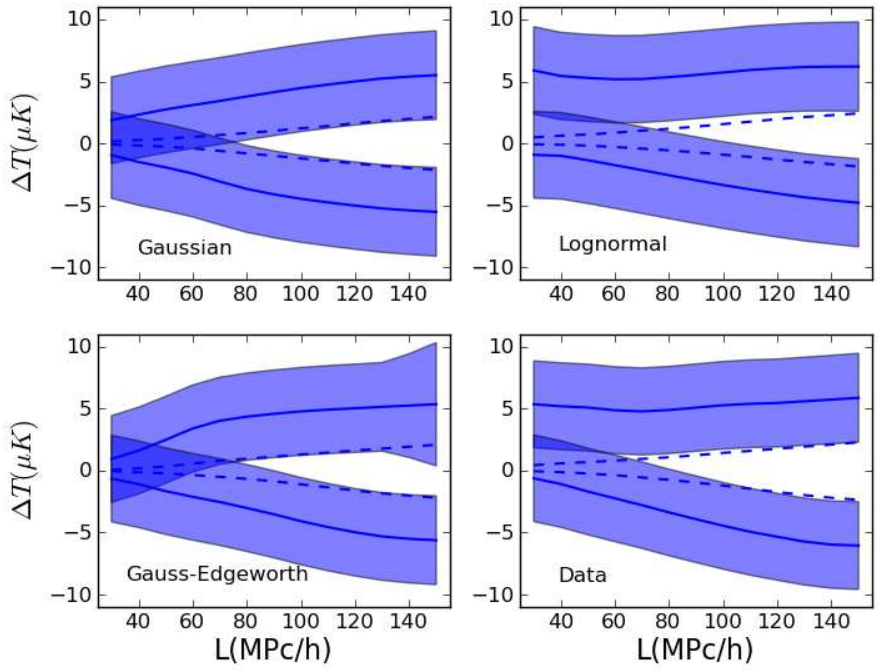

Figure 4.4 Expected ISW effect from the average of 50 supervoids and superclusters from a survey similar to the SDSS DR4 photometric LRG sample. The dashed line is Equation (4.7). In each panel, we use a different approximation of the matter PDF.

We approximate the tail of the matter PDF with analytic functions to estimate the distribution of the density extrema. We calculate the expected radial profile of supervoids and superclusters with the condition that their average density inside a sphere is known. We estimate the average linear ISW signal of the most significant 50 from each in a realistic survey. As can be seen from Figure 4.4, it is plausible that linear ISW can produce the results presented in Granett *et al.* (2008). They used a compensating top hat filter with inner radius of 4° to get $7.9 \pm 3.1 \mu K$ for clusters and $-11.3 \pm 3.1 \mu K$ for voids. In comparison, our estimates for the temperature at the center of the same stacks from Figure 4.4 are 5.5 ± 3.5 and -5.5 ± 3.5 in the case of a Gaussian PDF. We plot the linear size of a cube on the x-axis. The projection of a sphere with the same volume gives the corresponding angle. In case of L $= 150\ h^{-1}$Mpc, this is 3.8°. Although our errors originate from the fluctuations of the ISW signal, theirs come from the primary CMB anisotropies. Thus the errors are independent and our calculation is fully consistent with Granett *et al.* (2008). Our estimate is robust. It is not affected significantly by the details of the matter PDF. The error bars can be tightened if the volume of the survey is larger and if more images are stacked. The former would reduce the cosmic error on the scales we study, whereas the latter would give a more accurate measurement of the average profile.

We take one step toward cosmological parameter estimation with calculating the expected linear ISW signature of supervoids and superclusters. The next step can be to depart from the spherically symmetric model that we use for the sake of simplicity. Anisotropic fluctuations in the matter density around the center of a superstructure might give a quantitatively different error estimate. We also ignore any nonlinearities. We work with linear scales but we also probe the highest and lowest densities. The latter calls for a biasing model more complex than linear. We also ignore the nonlinear ISW, the Rees–Sciama effect (Rees and Sciama, 1968). It has been shown that it is small compared to

the linear part at low redshifts (see Cai *et al.* 2009, 2010). Another possible improvement is to use general relativity instead of Newtonian gravity. A model of compensated voids based on general relativity is discussed in Inoue and Silk (2007) and Inoue *et al.* (2010). We also ignore the correlation between the objects.

Our result suggests that the void needed to produce detectable anomalies on the CMB is smaller than previously estimated. The CMB Cold Spot (Cruz *et al.*, 2005) has been considered consistent with a compensated void having $\delta = -0.3$ and a radius of $200\ h^{-1}$Mpc by Sakai and Inoue (2008) in agreement with the heuristic argument of Rudnick *et al.* (2007). From Figure 4.3, one can see that a void of similar size needs to have a much less significant underdensity in our Gaussian, non-compensated model. The signal from a top-hat potential is less than half of the signal of a realistic void. This makes the detection of such voids harder in today's sparse catalogs (see Granett *et al.*, 2010; Bremer *et al.*, 2010).

References

Afshordi, N., Y. Loh, and M. A. Strauss, 2004, Phys. Rev. D **69**(8), 083524.

Bennett, C. L., M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, E. Wollack, E. L. Wright, *et al.*, 2003, ApJS **148**, 1.

Bremer, M. N., J. Silk, L. J. M. Davies, and M. D. Lehnert, 2010, MNRAS **404**, L69.

Cai, Y., S. Cole, A. Jenkins, and C. Frenk, 2009, MNRAS **396**, 772.

Cai, Y., S. Cole, A. Jenkins, and C. S. Frenk, 2010, MNRAS **407**, 201.

Copi, C., D. Huterer, and G. Starkman, 2004, Physical Review D **70**(4), 43515, ISSN 1550-2368.

Cruz, M., E. Martínez-González, P. Vielva, and L. Cayón, 2005, MNRAS **356**, 29.

Francis, C. L., and J. A. Peacock, 2010, MNRAS **406**, 2.

Giannantonio, T., R. Scranton, R. G. Crittenden, R. C. Nichol, S. P. Boughn, A. D. Myers, and G. T. Richards, 2008, Phys. Rev. D **77**(12), 123520.

Granett, B. R., M. C. Neyrinck, and I. Szapudi, 2008, ApJ **683**, L99.

Granett, B. R., I. Szapudi, and M. C. Neyrinck, 2010, ApJ **714**, 825.

Ho, S., C. Hirata, N. Padmanabhan, U. Seljak, and N. Bahcall, 2008, Phys. Rev. D **78**(4), 043519.

Hu, W., and S. Dodelson, 2002, ARA&A **40**, 171.

Inoue, K. T., N. Sakai, and K. Tomita, 2010, ApJ **724**, 12.

Inoue, K. T., and J. Silk, 2007, ApJ **664**, 650.

Kim, R. S., and M. A. Strauss, 1998, ApJ **493**, 39.

McEwen, J. D., Y. Wiaux, M. P. Hobson, P. Vandergheynst, and A. N. Lasenby, 2008, MNRAS **384**, 1289.

Padmanabhan, N., C. M. Hirata, U. Seljak, D. J. Schlegel, J. Brinkmann, and D. P. Schneider, 2005, Phys. Rev. D **72**(4), 043525.

Raccanelli, A., A. Bonaldi, M. Negrello, S. Matarrese, G. Tormen, and G. de Zotti, 2008, MNRAS **386**, 2161.

Rees, M. J., and D. W. Sciama, 1968, Nature **217**, 511.

Rudnick, L., S. Brown, and L. R. Williams, 2007, ApJ **671**, 40.

Sachs, R. K., and A. M. Wolfe, 1967, ApJ **147**, 73.

Sakai, N., and K. T. Inoue, 2008, Phys. Rev. D **78**(6), 063510.

Sawangwit, U., T. Shanks, R. D. Cannon, S. M. Croom, N. P. Ross, and D. A. Wake, 2010, MNRAS **402**, 2228.

Scranton, R., A. J. Connolly, R. C. Nichol, A. Stebbins, I. Szapudi, D. J. Eisenstein, N. Afshordi, T. Budavari, I. Csabai, J. A. Frieman, J. E. Gunn, D. Johnson, *et al.*, 2003, ArXiv Astrophysics e-prints `arXiv:astro-ph/0307335`.

Szapudi, I., and J. Pan, 2004, ApJ **602**, 26.

Tegmark, M., A. de Oliveira-Costa, and A. J. Hamilton, 2003, Phys. Rev. D **68**(12), 123523.

Chapter 5

Measurement of the full integrated Sachs-Wolfe profile of spherical density fluctuations

5.1 Introduction

Since the first sky maps of the Wilkinson Microwave Anisotropy Probe (WMAP) were published, there have been claims for the existence of circular features (spots and rings) in the cosmic microwave background (CMB) (Cruz *et al.*, 2005; Granett *et al.*, 2008a). Their origin and statistical significance are still debated. Sources, such as foreground contamination, the integrated Sachs-Wolfe (Sachs and Wolfe, 1967; ISW) effect, and the more exotic conformal cyclic cosmology, or cosmic texture, have been considered viable candidates to explain circular CMB anomalies (Rudnick *et al.*, 2007; Inoue *et al.*, 2010; Cruz *et al.*, 2007; Gurzadyan and Penrose, 2010). A notable feature is the cold spot (Cruz *et al.*, 2005), which has a mean temperature of $-70\mu K$ in a 5°-radius aperture. Additionally, $10\mu K$ hot and cold spots have been identified on 4° scales associated with super structures (Granett *et al.*, 2008a).

In this chapter, we focus on the ISW effect and investigate whether it is possible that large spherical fluctuations (supervoids and superclusters) in the dark matter field produce the aforementioned features in the CMB. The large-scale ISW effect is expected in a universe with accelerated cosmic expansion arising either from a dark energy term in flat cosmological models, or from spatial curvature. The effect is sensitive to both the expansion history and the rate of structure formation and provides constraints on alternative cosmological models (e.g. Giannantonio *et al.*, 2008b) as well as initial non-Gaussianity (e.g. Afshordi and Tolley, 2008). Cross correlating a galaxy catalog and a CMB temperature map is the standard way of studying the ISW signal and it has an extensive literature (Scranton *et al.*, 2003; Afshordi *et al.*, 2004; Padmanabhan *et al.*, 2005; Raccanelli *et al.*, 2008; Giannantonio *et al.*, 2008a; Ho *et al.*, 2008; Sawangwit *et al.*, 2010). Our work is similar considering that what we measure is related to the correlation between the dark matter and the CMB. However, we only focus on parts of the sky where the signal is expected to be large, regions corresponding to supervoids and superclusters. Because reports (Cruz *et al.*, 2005; Granett *et al.*, 2008a) indicate that the scale of these regions is beyond nonlinear scales, our model for their average density profile and ISW imprint is derived from the statistics of the linearly evolving primordial Gaussian density field (Pápai and Szapudi, 2010). This is in contrast with Inoue *et al.* (2010), who assumed a top hat density profile, which is the asymptotic final state of a void with steep initial density profile (Sheth and van de Weygaert, 2004).

We use the publicly available Hubble Volume Simulation of the Virgo Supercomputing Consortium (Colberg *et al.*, 2000; HVS) to test the Gaussian model. This is the N-body simulation with the largest volume, which is relevant considering that the gravitational potential is correlated more strongly than the density. We simulate ISW maps by ray tracing through the potential and calculating the linear part of the ISW effect. These

are compared to partial ISW maps generated from sets of spherical regions based on our model.

After gaining some confidence by studying simulations, we create an ISW map from real data. We select locations on the sky based on Granett *et al.* (2008a). They compiled a list of supervoids and superclusters found in the Sloan Digital Sky Survey (SDSS) Luminous Red Galaxy (LRG) sample. The list can be found in Granett *et al.* (2008b). We build an ISW map by placing the theoretical profiles to the given R.A., decl. coordinates. This map is fitted to a WMAP temperature map.

The structure of the Chapter is the following: in Section 5.2 we measure the expected density and ISW profile of spherical dark matter fluctuations in N-body simulations; in Section 5.3 we apply the matched filter technique to detect the signature of superstructures in the CMB; in Section 5.4 we discuss our results and views of the relationship between the linear ISW effect and circular features on the CMB.

5.2 Profiles in simulations

In this section, we calculate the full expected density profiles and ISW imprints of spherical overdensities and underdensities based on Pápai and Szapudi (2010) and compare them with measurements from the HVS.

5.2.1 Density profiles of superstructures in N-body simulations

According to Pápai and Szapudi (2010) and Chapter 4, if the average density contrast in a Gaussian random field inside a sphere of radius R is given, we expect the density

contrast at radius r from its center to be:

$$\left\langle \delta(r) \right\rangle_{\delta_{in}(R)} = \frac{\left\langle \delta(r)\delta_{in}(R) \right\rangle}{\left\langle \delta_{in}^2(R) \right\rangle} \delta_{in}(R), \tag{5.1}$$

where $\left\langle ... \right\rangle$ stands for ensemble averaging, $\delta(r)$ is the density contrast at radius r and $\delta_{in}(R)$ is the average of δ inside radius R. In the appendix of Dekel (1981), the author derived a similar formula for a single, point-like location with given density, which is the $R \to 0$ limit. Equation (5.1) is essentially the two-point function multiplied by a normalization constant. A more general calculation for the density profile around local maxima in a Gaussian random field was carried out by Bardeen *et al.* (1986).

In order to measure $\left\langle \delta(r) \right\rangle_{\delta_{in}(R)}$, one needs to select regions with $\delta_{in}(R)$ and calculate the average over these regions. However, according to Equation (5.1), $\frac{\delta(r)}{\delta_{in}(R)}$ gives an unbiased estimator of $\frac{\left\langle \delta(r)\delta_{in}(R) \right\rangle}{\left\langle \delta_{in}^2(R) \right\rangle}$, the shape of the profile:

$$\begin{aligned} \left\langle \frac{\delta}{\delta_{in}} \right\rangle = & \int \mathrm{d}\delta \mathrm{d}\delta_{in} \frac{\delta}{\delta_{in}} P(\delta|\delta_{in}) P(\delta_{in}) \\ = & \frac{\left\langle \delta\delta_{in} \right\rangle}{\left\langle \delta_{in}^2 \right\rangle} \int \mathrm{d}\delta_{in} P(\delta_{in}) = \frac{\left\langle \delta\delta_{in} \right\rangle}{\left\langle \delta_{in}^2 \right\rangle}, \end{aligned} \tag{5.2}$$

where we shorten $\delta(r)$ and $\delta_{in}(R)$ to δ and δ_{in}. We also use the fact that, by definition, $\left\langle \delta(r) \right\rangle_{\delta_{in}(R)} = \int \mathrm{d}\delta \delta P(\delta|\delta_{in})$ and $\frac{\left\langle \delta\delta_{in} \right\rangle}{\left\langle \delta_{in}^2 \right\rangle}$ is independent of δ_{in}. The argument above shows that it is unnecessary to search for spheres with a certain $\delta_{in}(R)$ to verify Equation (5.1). In the rest of this subsection, we deal with the integral of $\delta(r)$ instead of $\delta(r)$, because the former is measurable directly:

$$M(r) = 4\pi \int_0^r \mathrm{d}r r^2 \delta(r). \tag{5.3}$$

The linear size of the HVS is $3000\mathrm{h}^{-1}\mathrm{Mpc}$. We chose to place spheres at 25^3 evenly spaced coordinates. We counted galaxies in concentric spheres with radii of $10\mathrm{h}^{-1}\mathrm{Mpc}, 20\mathrm{h}^{-1}\mathrm{Mpc}, ..., 160\mathrm{h}^{-1}\mathrm{Mpc}$ and subtracted the average to get $M(r)$. We normalized this as $\frac{M(r)}{M(R)}\frac{4\pi}{3}R^3$ and calculated the average of the 25^3 instances to get an estimate of the integral of Equation (5.2). As R changes, technically, by weighting $M(r)$ differently, one can get the profile of linear superstructures of different sizes. A nuisance that we have to deal with, is that the estimator given by Equation (5.2) is unstable because δ_{in} can be arbitrarily small. The ratio of two Gaussian variables with zero means has a Lorentzian distribution, which has an undefined variance. This problem can be avoided if cases in which $|\delta_{in}|$ is under a threshold are ignored. This constraint leaves the estimator unchanged as can be seen from Equation (5.2). We chose the threshold based on the variance of δ_{in} in the particular catalog, e.q., $2\sqrt{\langle\delta_{in}^2\rangle}$. The result is not sensitive to the exact choice but the right choice can decrease the variance of the estimator substantially. In Figure 5.1 we plotted the measured profiles and their theoretical counterparts for $R = 60\mathrm{h}^{-1}\mathrm{Mpc}$ in the upper panel and for $R = 100\mathrm{h}^{-1}\mathrm{Mpc}$ in the lower panel. We obtained the error bars by repeating the measurement on one hundred mock catalogs generated with a second order Lagrangian (2LPT) code (Crocce *et al.*, 2006). The mocks were set up to have the same cosmological parameters, size and density as the HVS. To calculate the theoretical profiles of Equation (5.1), we used a matter power spectrum calculated by CAMB (Lewis *et al.*, 2000). The measured profiles, both in 2LPT and the HVS, appear to be in good agreement with theory within the uncertainty up to $400\mathrm{h}^{-1}\mathrm{Mpc}$, the largest radius in our measurement.

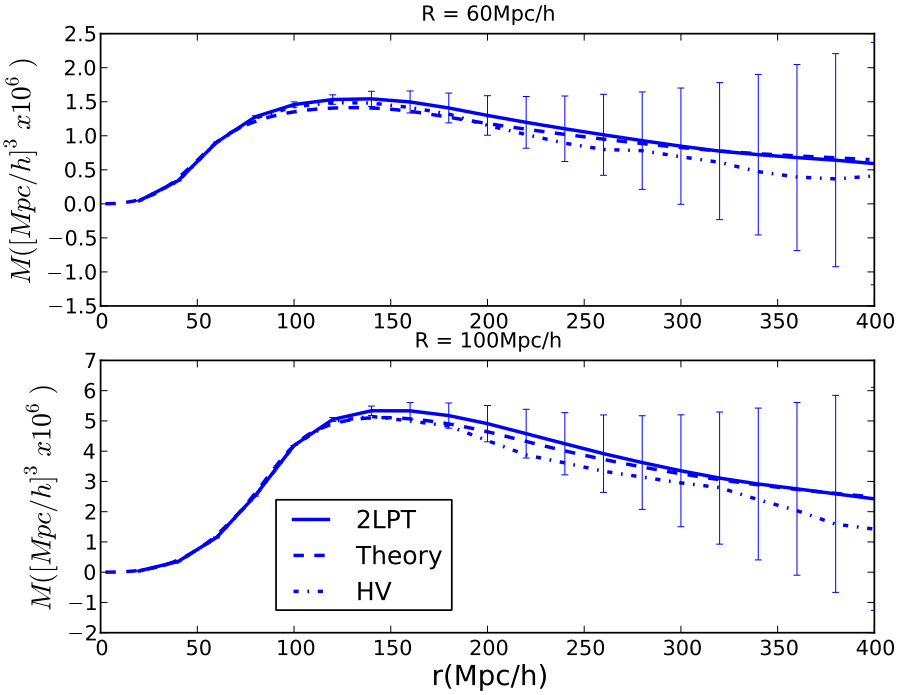

Figure 5.1 The integral of the spherically averaged density profile measured in the HVS and 100 2LPT simulations. Linear theory given by Equation (5.2) appears to be a good approximation.

5.2.2 The ISW imprints of superstructures

In Pápai and Szapudi (2010), the ISW effect in the direction of the center of a superstructure was calculated. This was done simply by calculating the spherically symmetric potential of the superstructure, then computing the following integral along the path of the CMB photon (Sachs and Wolfe, 1967):

$$\Phi(r) = -\frac{3\Omega_m}{8\pi}\left(\frac{H_0}{c}\right)^2 \int_r^\infty \frac{M(\tilde{r})}{\tilde{r}^2}\mathrm{d}\tilde{r}, \tag{5.4}$$

$$\frac{\Delta T}{T} = -\frac{2}{c^2}\int \mathrm{d}\tau \frac{\partial \Phi(r(\tau),\tau)}{\partial \tau}, \tag{5.5}$$

where τ denotes conformal time. Similarly, this calculation can be carried out for any direction easily in order to predict the full ISW profile. In linear theory, the potential at a a particular redshift is just $\Phi(r,z) = \Phi(r,z=0)\frac{D(z)}{a(z)}$. We leave the study of nonlinear corrections for future work. Because these integrals are linear, the linear ISW signal is predicted to be proportional to the average density of the superstructure, δ_{in}. In this chapter as in the previous one, we place the centers of the superstructures at $z = 0.52$, which is the median redshift of the galaxy catalog we use in Section 5.3.

5.2.3 Fluctuations in the potential on the largest scales and the ISW profile

To test the ISW profile, we traced rays through the HV simulation along the z-axis. We calculated δ on an 800^3 grid and calculated Φ at $z = 0$ by using the simple formula:

$$\Phi(k) = -\frac{3\Omega_m}{2}\left(\frac{H_0}{c}\right)^2 \frac{\delta(k)}{k^2}. \tag{5.6}$$

Because we aimed to analyze imprints of superstructures at redshift 0.52, we defined the boundaries of the integral of Equation (5.5) in a way that put the 25^3 preselected locations (see Subsection 5.2.1) at this redshift. We used periodic boundary conditions and we integrated up to $3000\mathrm{h}^{-1}\mathrm{Mpc}$ from starting points ensuring that 25^2 of the locations were at redshift 0.52 each time. This yielded 25 ISW images of the HVS in the $x-y$ plane. Each of these images contains imprints of 25^2 superstructures. After this, one can follow the procedure discussed for density profiles in Subsection 5.2.1. If the ISW profile of a superstructure is normalized by the average density of the superstructure, the result will be an estimator of the shape of the ISW profile:

$$\langle \frac{\Delta T(r_{2D})}{\delta_{in}(R)} \rangle = f(r_{2D}, R). \tag{5.7}$$

The arguments are r_{2D}, the radius in the $x-y$ plane and R, the radius in which we know the galaxy count in the corresponding volume. The interpretation of Equation (5.7) is similar to that of Equation (5.2). The density fluctuations are replaced by the temperature on the left side of the equation; while on the right side, we symbolically refer to the normalized ISW profile which can be derived via Equations (5.3)-(5.5).

Here, we would like to remind the reader that $\langle ... \rangle$ refers to ensemble averaging. For density, averaging over a relatively large volume proved to be a sufficient substitute in Subsection 5.2.1. Despite the fact that the cosmic variance is large on the largest scales, these modes contribute only a little to the density distribution since their amplitudes are small. (The power spectrum goes to zero with k.) For the potential, this picture changes due to the $1/k^2$ factor. As a result, the ISW effect has a large cosmic variance regardless of the large size of the simulation. To demonstrate this, we projected the density to the $x-y$ plane. We split this projection in the middle of the x-axis to create two 1500 X $3000 Mpc^2 h^{-2}$ areas. In Figure 5.2 in the top row we plotted the histograms of these.

We did the same with one of the 25 ISW images, which is essentially the projection of the potential. Those histograms are shown in the middle row. The cosmic variance is significantly larger for the ISW map. To demonstrate that the difference is due to modes on the largest scales, we removed the low k-modes from the potential. After experimenting, we found that removing modes with $|k| < 2\pi\frac{5}{L}$, where $L = 3000\mathrm{h}^{-1}\mathrm{Mpc}$ for the HVS, gave a visually compelling result (Figure 5.2, bottom row). After applying this high-pass filter to a 2-dimensional projection, the effect of the cosmic variance on the modes with the lowest wavenumber will be around 18%. This argument is valid for any survey or simulation. Generally, modes can be combined optimally with the proper covariance matrix. In this paper, we simply use the sum of the modes as an estimator of the ISW signal. In other words, our covariance matrix is diagonal with a truncation at $1/5$ of the linear size. This subjective choice, which we consistently carry along through the paper, is based purely on the histogram of the potential (Figure 5.2).

Having removed the low k-modes from both the density and the potential we computed the average in Equation (5.7). As before, we omitted cases when the denominator was too close to zero. (See Subsection 5.2.1.) The results for $R = 60\mathrm{h}^{-1}\mathrm{Mpc}$ and $R = 100\mathrm{h}^{-1}\mathrm{Mpc}$ are shown in Figure 5.3. The error bars show the uncertainty calculated from the HVS itself due to the long CPU time required to calculate the ISW images. Predictions of the linear model are calculated by removing the low k-modes from the power spectrum when doing calculation according to Equation (5.1).

This approach has to be modified slightly when the CMB is given and late time anisotropies are affected by every mode of the galaxy distribution. The previous exercise still proves that low k-modes are to be ignored if we want to study the imprints of superstructures statistically. This can be achieved by filtering out these modes from the CMB before performing such analysis. In addition, one has to remove the same components from the theoretical ISW profile before fitting the data. This is shown in

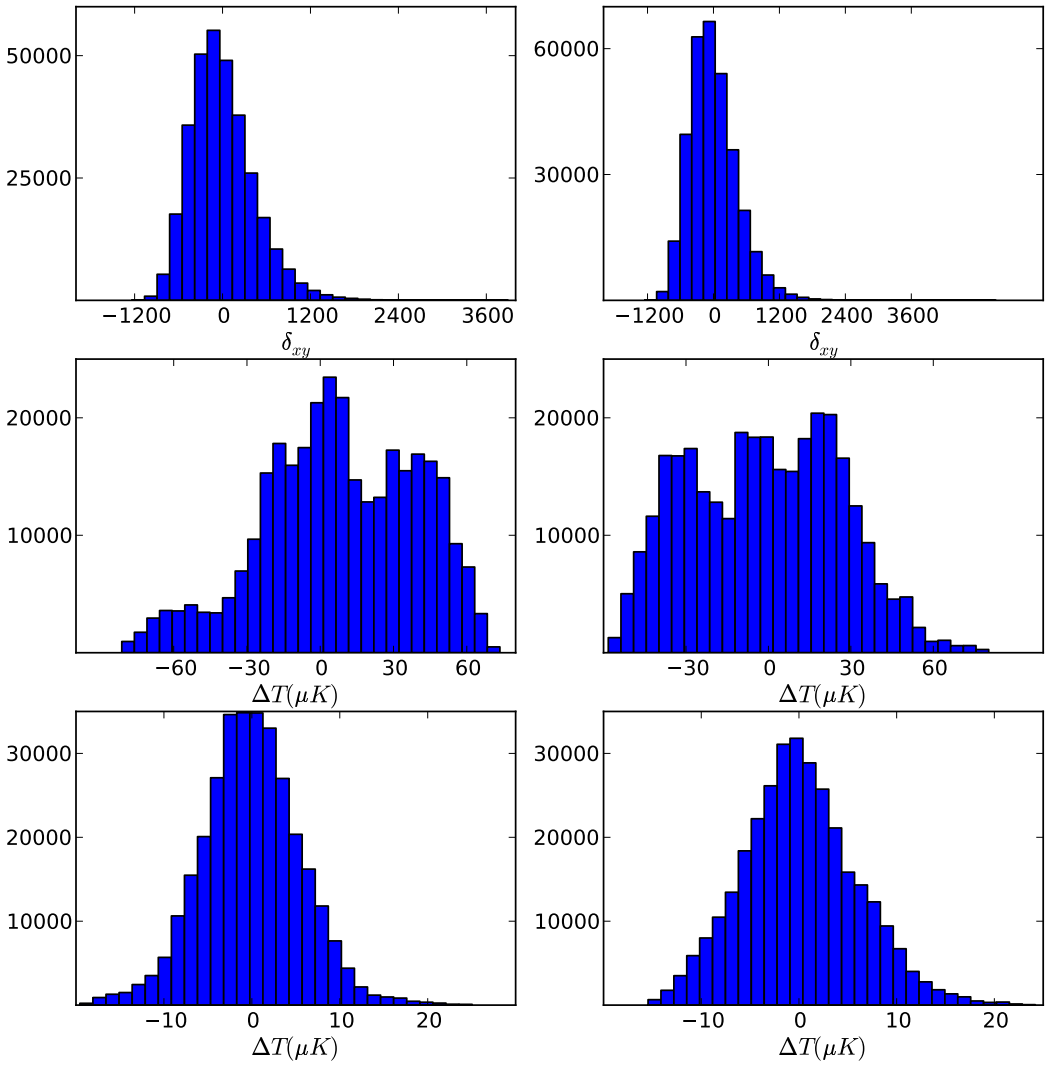

Figure 5.2 Each row shows histograms of a certain projection of density along the z-axis. A simple integral of $\int \delta(x, y, z)\mathrm{d}z$ is presented in the first row, an ISW map in the second and its filtered version in the third. (See text for details.) The $x - y$ plane is halved to test the cosmic variance, hence the two columns. As a result of large fluctuations in the low k-modes in the potential, the ISW projections show a significant cosmic variance (second row) which can be reduced by filtering (third row).

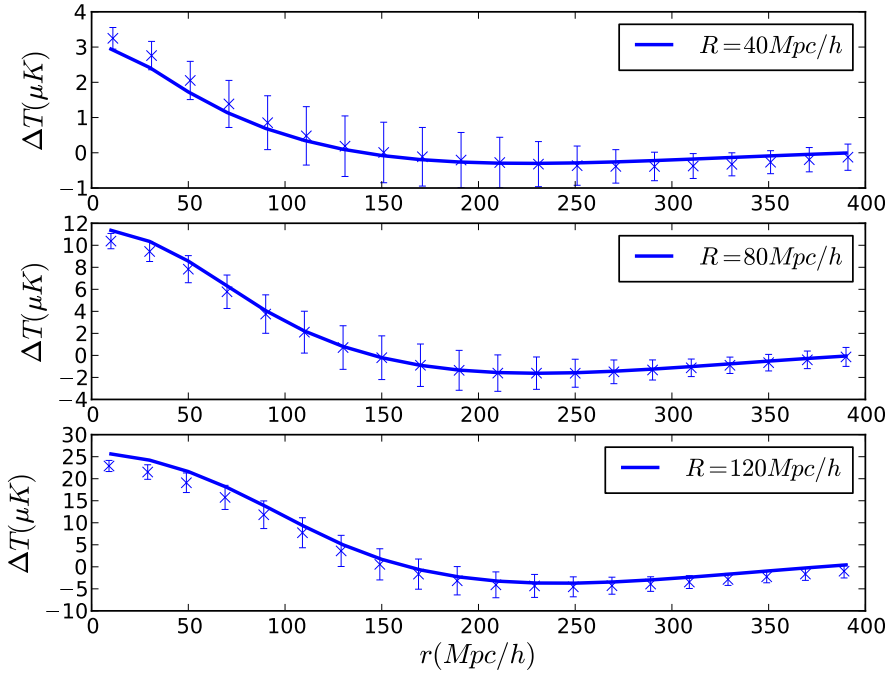

Figure 5.3 The spherically averaged ISW profile from the HVS. This is defined by Equation (5.7). Fourier-modes from the density were filtered out above $\lambda = 0.2L$, where L is the linear size of the HVS. The linear approximation is calculated as described in the text. See Equations (5.2)−(5.5).

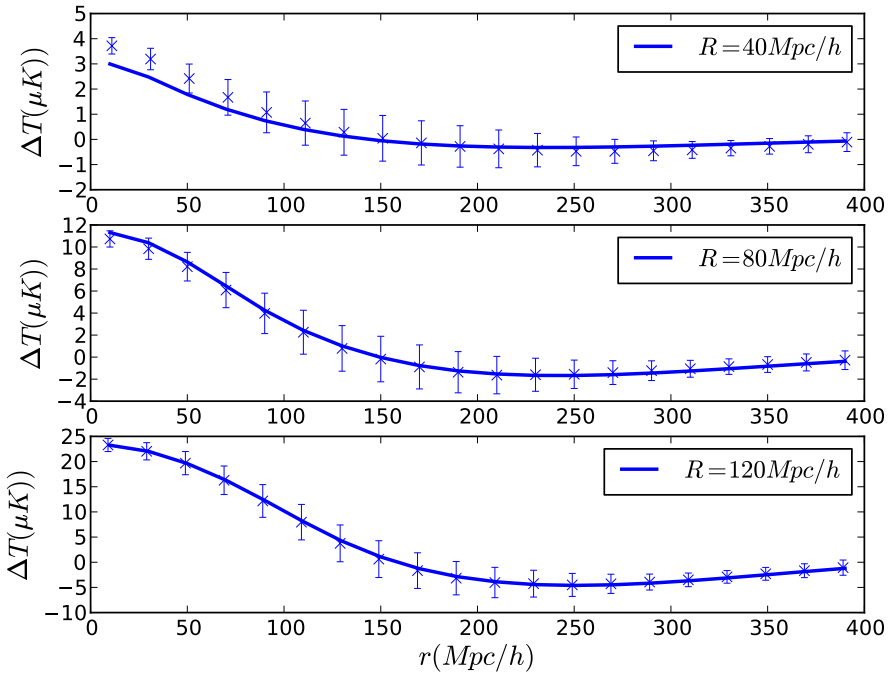

Figure 5.4 The same as Figure 5.3 with the exception that the ISW map was filtered, not the density. The error bars represent the uncertainty of the measurement.

Figure 5.4. We filtered the ISW map and we filtered the ISW profile, which is implicitly expressed as $f(r_{2D}, R)$, on the right side of Equation (5.7), so that only modes with $|k| > 2\pi\frac{5}{L}$ remained. After these manipulations the measurement remained consistent with linear theory.

The success of these tests proves the validity of the Gaussian model of Pápai and Szapudi (2010). We studied Euclidean ISW projections because or their simplicity, but there is no reason to think that our findings are restricted to the Euclidean case. The only major difference is that on a sphere one deals with spherical harmonics and instead of low k-modes we speak of low l-modes analogously.

5.3 ISW map of SDSS Superstructures

Having calculated the full linear ISW profile of superstructures in the dark matter density, we attempt to detect their presence on the CMB. With the locations of 50 superclusters and 50 supervoids found in the SDSS DR6 LRG sample by Granett *et al.* (2008a), we created ISW maps and used the matched filter technique to measure the amplitude of the ISW signal. Our ISW maps are statistical in the sense that the individual ISW profiles were predicted by using statistical properties of Gaussian, spherically symmetric density fluctuations rather than the particular (poorly observed) realization of the density field.

5.3.1 Superclusters and supervoids in the SDSS LRG sample on $100\mathrm{h}^{-1}\mathrm{Mpc}$ scales

Large regions with significantly high overdensities or low underdensities are called here superclusters or supervoids. In a given dataset, these regions can be defined in many different ways. Void finder algorithms were compared in Colberg *et al.* (2008). Each implementation is based on a certain notion of a void. Voids can have spherical or irregular shapes; finders can be parameter free or there can be a hard-coded density threshold or radius to describe what a void is like. It is important to emphasize that despite the similarities, any analysis using the output of a void finder will be sensitive to its particulars. This renders our study qualitative in nature.

Our work is based on supervoids and superclusters found in a subsample of the SDSS DR6 LRG catalog by ZOBOV and VOBOZ, parameter-free void and cluster-finder algorithms (Neyrinck, 2008; Neyrinck *et al.*, 2005). ZOBOV and VOBOZ calculate the Voronoi tessellation of the data in order to approximate the density field with a simple function, a funcion that is constant inside polyhedra. Voids are found by using the concept of drainage basins from the field of geography. Clusters are identified analogously.

A table with the properties of 50 supervoids and 50 superclusters can be found in Granett *et al.* (2008b). Their sizes and densities cannot be quantified simply due to their shapes and survey boundary effects. The bias, redshift error and the shot-noise are also sources of uncertainty for both size and density. This is why we used only the directions of the centers of these superstructures in creating ISW maps in Section 5.3.3. The error in the angular coordinates is relatively small.

5.3.2 CMB and ISW maps

We used the Internal Linear Combination (ILC) Map from the WMAP 7-year release (Hinshaw *et al.*, 2007; Gold *et al.*, 2010) with the KQ75y7 galactic foreground and point source mask (Figure 5.5). We employed HEALPix pixelization at resolution $n_{side} = 64$ (Górski *et al.*, 2005). In general, we strove to carry out the measurement with up to date maps and masks while remaining consistent with Granett *et al.* (2009).

Section 5.2 concluded that a Gaussian model gave a good approximation to the expected ISW signal from a large spherical region in the dark matter field with given density contrast. It also became apparent that in an analysis of statistical ISW maps low l-modes were to be filtered out due to the large cosmic variance of the potential affecting these modes. This also lowers the cosmic variance. The lowest mode to be taken into account is determined by the sky coverage of the ISW map. Experimenting with the HVS gave us $k_{min} \approx 2\pi\frac{5}{L}$ for an Euclidean ISW map, where L is the linear size of the map. This translates into $l_{lmin} \approx \frac{2\pi 5r}{\lambda}$ via the Limber approximation, where λ and r are the angular size and the median distance of the SDSS DR6 LRG sample. A rough estimate yields $l_{min} \approx 12$.

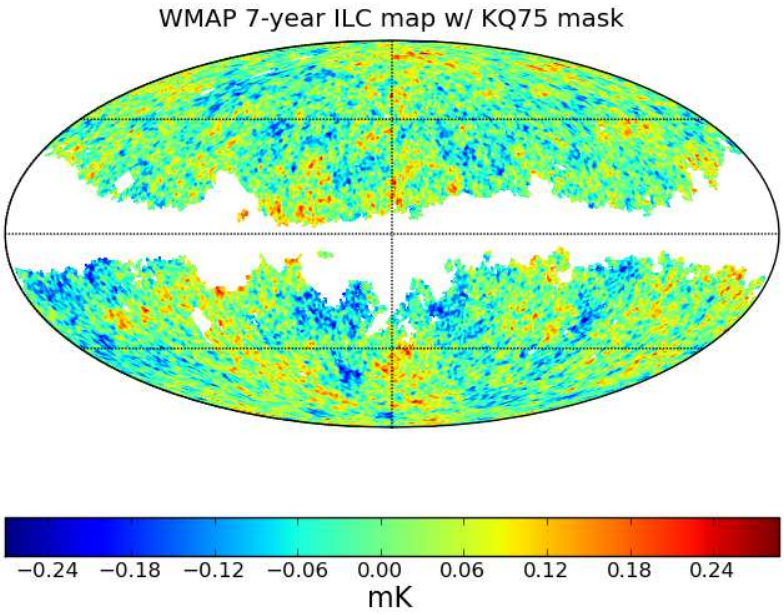

Figure 5.5 WMAP 7-year ILC map with 25% of the sky masked out according to the KQ75y7 foreground mask as explained in Gold *et al.* (2010).

Due to the large uncertainty in their size and density, we assume that each supervoid is a realization of the same statistical entity when we creating ISW maps. We consider superclusters similar in nature but with an opposite density contrast.

An ISW map is made simply by placing the expected ISW profiles of superstructures at the locations given in Granett *et al.* (2008b). We set δ_{in} to 1 for clusters and -1 for voids in formulae of Section 5.2. For every radius there is a corresponding ISW map. We calculated the a_{lm} coefficients of the spherical expansion of both the ISW maps and the CMB map, and zeroed the modes below $l_{min} = 12$ before transforming the a_{lm} coefficients back. These manipulations were carried out with routines from the Healpix package (Górski *et al.*, 2005). Figure 5.6 shows how the ISW profile changes when modes are removed.

5.3.3 The matched filter technique

We continue with measuring the amplitude of the ISW maps created in Section 5.3.2 in the WMAP ILC map. To do this we used the same matched filter technique as Granett *et al.* (2009). A likelihood function is defined as:

$$L(\lambda) = -\frac{1}{2}\big(T_{cmb} - \lambda T_{ISW}\big)C^{-1}\big(T_{cmb} - \lambda T_{ISW}\big), \tag{5.8}$$

where T_{cmb} is a foreground cleaned CMB map, in this case the WMAP ILC map, and T_{ISW} refers to the ISW maps. The pixel covariance matrix, C, is computed from the best fitting ΛCDM power spectrum. This likelihood is based on the assumption that the T_{ISW} and the primary CMB are the realizations of two very different Gaussian random fields

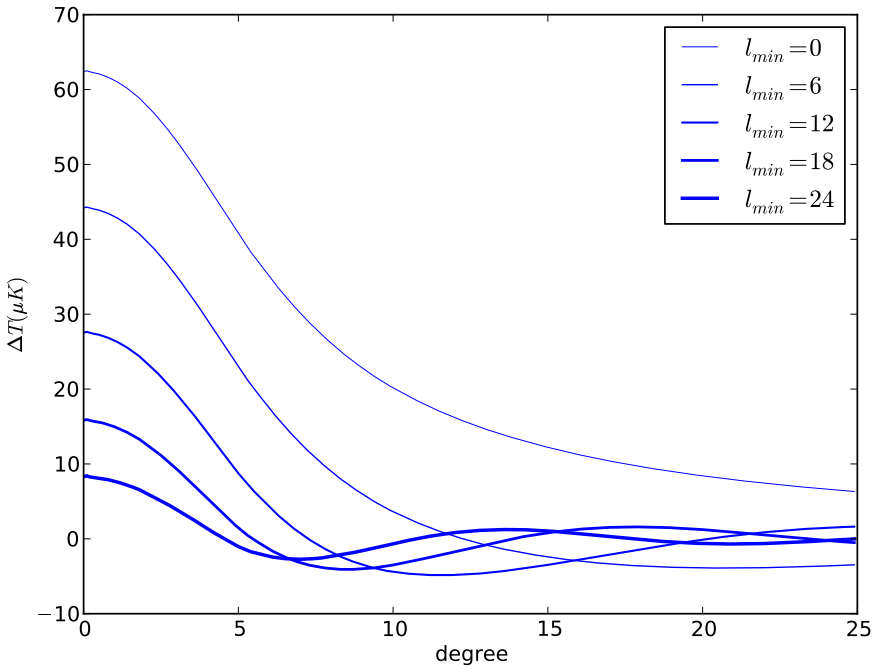

Figure 5.6 The expected ISW imprint of a spherical region with radius of $120\mathrm{h}^{-1}\mathrm{Mpc}$ and an average δ of 1 at redshift 0.52. Each curve shows the same profile without modes under their respective l_{min}.

and the latter is given by the minimum of Equation (5.8):

$$\lambda^* = \frac{T_{ISW} C^{-1} T_{cmb}}{T_{ISW} C^{-1} T_{ISW}}, \tag{5.9}$$

$$T_{prim} = T_{cmb} - \lambda^* T_{ISW}. \tag{5.10}$$

As explained in Section 5.3.2, we filtered out modes up to $l_{min} = 12$ from both T_{ISW} and T_{cmb}, and before building the covariance matrix from C_l's we set $C_0 = C_1 = ... = C_{11} = 0$.

Because the exact inverse of the covariance matrix is unstable for partial sky maps, we regularized C by calculating a pseudo-inverse omitting the noisiest modes. The likelihood function above defines an error on λ^* as

$$\Delta\lambda^2 = \frac{1}{T_{ISW} C^{-1} T_{ISW}}. \tag{5.11}$$

5.3.4 Results

We calculated λ^* according to Equation (5.9). Because we have ISW maps for different void/cluster radii, the result is a function, $\lambda(R)$. This is presented in Figure 5.7 in the bottom row on the left panel. The error bars are given by Equation (5.11). From now on every time we refer to a λ we mean the value at the minimum of the likelihood function, so we drop the superscript *. If we marginalize $\lambda(R)/\Delta\lambda(R)$ over R, we get a 3.24-σ detection. Similarly, the marginalized value of R is $55 \pm 28 \mathrm{h}^{-1}\mathrm{Mpc}$.

$\lambda(R)$ can be interpreted as the average density contrast in one hundred superstructures inside radius R, where the supervoids are taken into account with a minus sign:

$$\lambda(R) \approx \frac{1}{2}\left(\delta_{in}^{50c}(R) - \delta_{in}^{50v}(R)\right), \tag{5.12}$$

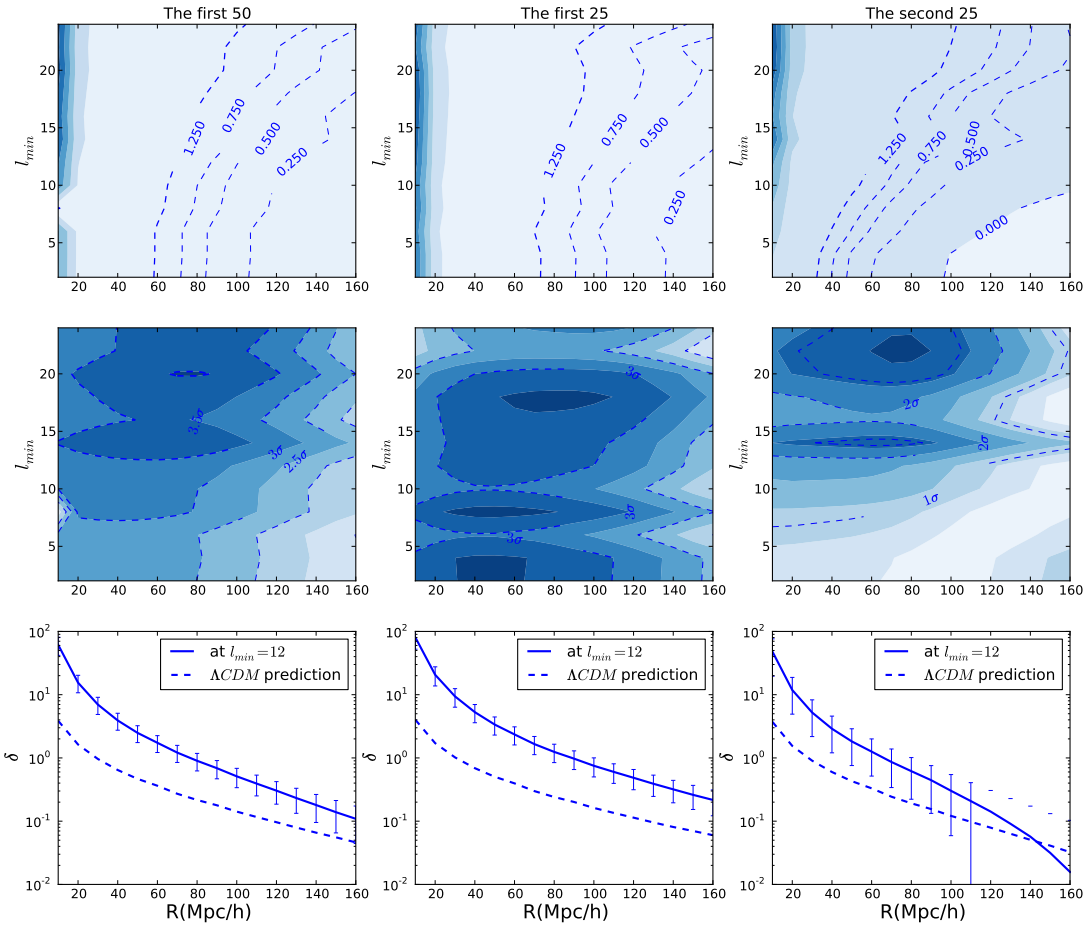

Figure 5.7 The top and the middle rows contain contour plots of the best fitting amplitude of the ISW maps, λ, and their relative uncertainty, σ, (see text for their definition) as functions of l_{min} and R. The bottom row shows cross sections of the contour plots from the top row at $l_{min} = 12$, as well as the prediction for the average density contrast of superstructures of a certain size. In the left column, the fitting is for an ISW map consisting of the 50 most significant supervoids and 50 superclusters. The middle column shows the same for the best 25-25, whereas the right for the remaining 25-25 superstructures which are not included in the middle column.

where superscripts $50c$ and $50v$ refer to the average of 50 superclusters and 50 supervoids. This interpretation originates from the linear relationship between the ISW signal and the density fluctuations, and our choice of $\delta_{in} = \pm 1$ when building ISW maps in Section 5.3.2. In addition to the data curves, in Figure 5.7 we plotted the prediction of linear theory for the average of the 50 largest superclusters. We assumed a Gaussian probability distribution function (PDF), which is reasonable in light of the findings of Pápai and Szapudi (2010). The variance of the PDF was calculated by CAMB (Lewis *et al.*, 2000) using the latest WMAP cosmological parameters (Jarosik *et al.*, 2010) and it was scaled with the growth function to the median redshift of the superstructures $(z = 0.52)$. The number of independent clusters was estimated as $V_{survey}/V_{cluster}$ where V_{survey} is the volume of the LRG sample used by Granett *et al.* (2009). λ is consistently several times larger than the predicted δ_{in}. As a consistency test, we split the voids and clusters into two groups. Group 1 consists of the first 25 voids and 25 clusters with the highest significance in the supplementary tables of Granett *et al.* (2008b). The other 25 voids and 25 clusters make up group 2. We created two ISW maps, one from each of the groups. After repeating the whole procedure for these, we got the center and the right panel in the bottom row of Figure 5.7. As expected, superstructures associated with lower significance produce a lower signal on the CMB and allow larger deviation from the average λ. This supports the hypothesis that the signal is due to the ISW effect.

We also checked the sensitivity of the results to the choice of l_{min}, the lowest spherical harmonics included in the analysis. The result for different choices of l_{min} is shown in the middle and top rows of Figure 5.7. The top row shows $\lambda(R, l_{min})$, while the middle row shows $\sigma(R, l_{min}) = \lambda(R, l_{min})/\Delta\lambda(R, l_{min})$ on contour plots. The result seems robust for wide range of l_{min}.

We performed another consistency check to acquire a better understanding of the uncertainty of λ and its relationship with δ_{in}. We created ISW maps from single

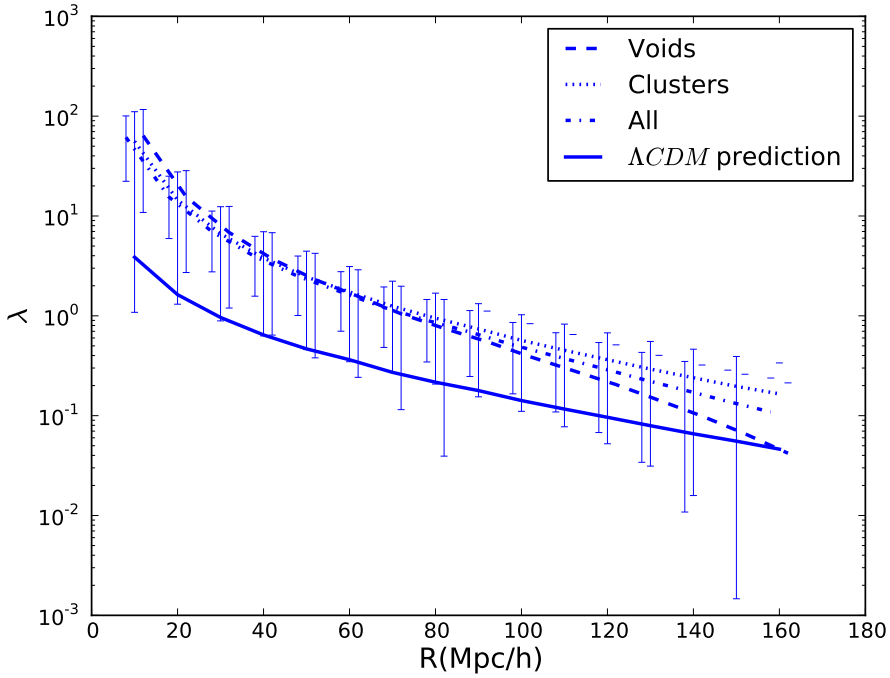

Figure 5.8 After fitting simple ISW maps made from the expected imprint of single superclusters or supervoids we were left with 50-50 best amplitudes, λ. Their averages with the 2σ uncertainty are plotted with the amplitude of a full ISW map consisting of all 100 imprints. These measurements were done with $l_{min} = 12$. The theoretical prediction (continuous line) for the density contrast is also plotted.

superstructures. We used the matched filter technique to determine a λ and its error as before. The average λ for voids and for clusters separately are plotted in Figure 5.8. In this case we defined the uncertainty of λ as the standard deviation of its mean. We also plotted the previous results for the previous combined ISW map. All of this is for $l_{min} = 12$. The three measured curves are consistent with each other, the error bars show the two σ deviation. The single void and cluster fit allow wider range of λ than the combined fit.

5.4 Discussion

This chapter is the continuation of the work started in Pápai and Szapudi (2010). Our goal is to estimate the ISW imprint of large spherical dark matter overdensities and underdensities. In Section 5.2, we first tested a simple Gaussian model for density profiles in the HVS. We created an ISW map by ray tracing through the simulation and computing the linear ISW effect. We neglected the nonlinear part of the ISW effect, as it is small in comparison at $z = 0$ (Cai *et al.*, 2010). Another simplification was that our ISW map was Euclidean. For the purpose of testing this has no relevance.

As it can be seen from Figure 5.1, the density profiles followed linear theory given by Equation (5.1) within 1-σ uncertainty up to $400\mathrm{h}^{-1}\mathrm{Mpc}$, the largest scale in our study. This is not surprising, since the density profile is equivalent of the two-point function and on large scales high-order clustering is not important. When we averaged the density profiles we only used locations where the average density contrast in a certain radius was in the 2-σ (positive or negative) tail of the PDF. This demonstrates that not even the extreme cases are affected by significantly nonlinearities.

We demonstrated that cosmic variance was a much more important factor for the ISW effect than for the density. Because cosmic variance is large for low k-modes and

the potential is non-vanishing as $k \to 0$, these modes have relatively large effect on the ISW profile. Surely large variance modes should be removed from the CMB and the ISW maps if the ISW maps are based on the statistics of density fluctuations and not on the observed density. First, we measured the ISW profiles after filtering out the low k-modes from the HVS density and consequently the ISW map. Filtering the two-point function in the same way gave a good match to the data. (See Figure 5.3.)

If the ISW map is given, as in the CMB, another route has to be taken. We filtered the complete ISW map and compared it to the theoretical profile after removing the same modes. The result is shown in Figure 5.4.

In Section 5.3, we used matched filter technique to detect the ISW signal of superstructures in the CMB. We closely followed the procedure described in Granett *et al.* (2009). The major difference is that our ISW map was built from theoretical assumptions about the shape of ISW profiles, whereas Granett *et al.* (2009) used an analytic curve which fitted the measured profiles the best. Because of this difference, the significance of our measurement can readily be interpreted. We estimated the marginalized significance of our measurement to be $\sigma = 3.24$. (See Subsection 5.3.4 for details.) The interpretation of the best fitting amplitude of the ISW map is still not without difficulties. It is several times higher than the anticipated signal (Figures 5.7 and 5.8), although it appears to be 2-σ higher than ΛCDM predictions, where $sigma$ indicates the uncertainty of the fit and not the intrinsic uncertainty of the ISW signal. We caution, though, that the uncertainty can still be underestimated, because it was derived from data alone. Despite the difference in amplitude, the theoretical and the measured curves run parallel with each other, which supports that the signal is related to the ISW effect. Moreover, traditional cross-correlation methods have also been reported to yield higher signal than expected (Ho *et al.*, 2008; Giannantonio *et al.*, 2008a; Granett *et al.*, 2009).

The ISW effect provides an explanation for another feature of observations. On average, supervoids tend to have a hot ring around a cold center and the opposite is true for superclusters (Granett *et al.*, 2009). When we erased large scale fluctuations in the microwave background, the expected ISW profile changed sign, as seen in Figure 5.6. This means that in certain, not particularly unique, realizations the average profiles of 50 supervoids or superstructures from a finite area of the sky can have rings in the ISW context, simply due to the cosmic variance of low k-modes.

Overall, we can say that the shape of the measured signal follows the predicted ISW profile while its amplitude exceeds expectation. This is a good reason to investigate further by studying galaxy surveys other than the SDSS. A key to a quantitative study is the well-measured galaxy density. Because the model described in this and the previous chapter is a model for density fluctuations around a particular location, which is not necessarily a maximum, it is enough to know the density at this location accurately.

The HVS were carried out by the Virgo Supercomputing Consortium using computers based at the Computing Centre of the Max-Planck Society in Garching and at the Edinburgh parallel Computing Centre. The data are publicly available at http://www.mpa-garching.mpg.de/NumCos.

References

Afshordi, N., Y. Loh, and M. A. Strauss, 2004, Phys. Rev. D **69**(8), 083524.

Afshordi, N., and A. J. Tolley, 2008, Phys. Rev. D **78**(12), 123507.

Bardeen, J. M., J. R. Bond, N. Kaiser, and A. S. Szalay, 1986, ApJ **304**, 15.

Cai, Y., S. Cole, A. Jenkins, and C. S. Frenk, 2010, MNRAS **407**, 201.

Colberg, J. M., F. Pearce, C. Foster, E. Platen, R. Brunino, M. Neyrinck, S. Basilakos, A. Fairall, H. Feldman, S. Gottlöber, O. Hahn, F. Hoyle, *et al.*, 2008, MNRAS **387**, 933.

Colberg, J. M., S. D. M. White, N. Yoshida, T. J. MacFarland, A. Jenkins, C. S. Frenk, F. R. Pearce, A. E. Evrard, H. M. P. Couchman, G. Efstathiou, J. A. Peacock, P. A. Thomas, *et al.*, 2000, MNRAS **319**, 209.

Crocce, M., S. Pueblas, and R. Scoccimarro, 2006, MNRAS **373**, 369.

Cruz, M., E. Martínez-González, P. Vielva, and L. Cayón, 2005, MNRAS **356**, 29.

Cruz, M., N. Turok, P. Vielva, E. Martínez-González, and M. Hobson, 2007, Science **318**, 1612.

Dekel, A., 1981, A&A **101**, 79.

Giannantonio, T., R. Scranton, R. G. Crittenden, R. C. Nichol, S. P. Boughn, A. D. Myers, and G. T. Richards, 2008a, Phys. Rev. D **77**(12), 123520.

Giannantonio, T., Y. Song, and K. Koyama, 2008b, Phys. Rev. D **78**(4), 044017.

Gold, B., N. Odegard, J. L. Weiland, R. S. Hill, A. Kogut, C. L. Bennett, G. Hinshaw, X. Chen, J. Dunkley, M. Halpern, N. Jarosik, E. Komatsu, *et al.*, 2010, ArXiv e-prints 1001.4555.

Górski, K. M., E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, 2005, ApJ **622**, 759.

Granett, B. R., M. C. Neyrinck, and I. Szapudi, 2008a, ApJ **683**, L99.

Granett, B. R., M. C. Neyrinck, and I. Szapudi, 2008b, ArXiv e-prints 0805.2974.

Granett, B. R., M. C. Neyrinck, and I. Szapudi, 2009, ApJ **701**, 414.

Gurzadyan, V. G., and R. Penrose, 2010, ArXiv e-prints 1011.3706.

Hinshaw, G., M. Nolta, C. Bennett, R. Bean, O. Doré, M. Greason, M. Halpern, R. Hill, N. Jarosik, A. Kogut, *et al.*, 2007, The Astrophysical Journal Supplement Series **170**, 288.

Ho, S., C. Hirata, N. Padmanabhan, U. Seljak, and N. Bahcall, 2008, Phys. Rev. D **78**(4), 043519.

Inoue, K. T., N. Sakai, and K. Tomita, 2010, ApJ **724**, 12.

Jarosik, N., C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern, R. S. Hill, G. Hinshaw, A. Kogut, E. Komatsu, D. Larson, M. Limon, *et al.*, 2010, ArXiv e-prints 1001.4744.

Lewis, A., A. Challinor, and A. Lasenby, 2000, ApJ **538**, 473.

Neyrinck, M. C., 2008, MNRAS **386**, 2101.

Neyrinck, M. C., N. Y. Gnedin, and A. J. S. Hamilton, 2005, MNRAS **356**, 1222.

Padmanabhan, N., C. M. Hirata, U. Seljak, D. J. Schlegel, J. Brinkmann, and D. P. Schneider, 2005, Phys. Rev. D **72**(4), 043525.

Pápai, P., and I. Szapudi, 2010, ApJ **725**, 2078.

Raccanelli, A., A. Bonaldi, M. Negrello, S. Matarrese, G. Tormen, and G. de Zotti, 2008, MNRAS **386**, 2161.

Rudnick, L., S. Brown, and L. R. Williams, 2007, ApJ **671**, 40.

Sachs, R. K., and A. M. Wolfe, 1967, ApJ **147**, 73.

Sawangwit, U., T. Shanks, R. D. Cannon, S. M. Croom, N. P. Ross, and D. A. Wake, 2010, MNRAS **402**, 2228.

Scranton, R., A. J. Connolly, R. C. Nichol, A. Stebbins, I. Szapudi, D. J. Eisenstein, N. Afshordi, T. Budavari, I. Csabai, J. A. Frieman, J. E. Gunn, D. Johnson, *et al.*, 2003, ArXiv Astrophysics e-prints arXiv:astro-ph/0307335.

Sheth, R. K., and R. van de Weygaert, 2004, MNRAS **350**, 517.

Chapter 6

Conclusion

In my dissertation, I study aspects of galaxy clustering. The range of related phenomena is wide, since the equations describing the evolution of the universe are coupled, and observables are rarely uncorrelated.

I start out with my work on redshift-distorted two-point functions. The motivation stems from the need of a more precise measurement of the correlation function. In the distant observer approximation, pairs with a large angle of view are discarded in order for the approximation to remain valid. In Chapter 2, I argue that data can be used more efficiently if we use the full redshift-distorted two-point function. By developing a formalism which yields a compact form for the non-perturbative terms, Equation (2.3)-(2.11), we facilitate measurements with lower uncertainty. This improvement depends on properties of the survey, being more significant, when the galaxy pairs are closer to the observer. In Pápai and Szapudi (2008), we tested the formula in the Hubble Volume Simulation and found them accurate for the tested configuration (Figure 2.1). Recently, a more comprehensive study of all configurations was published by Raccanelli *et al.* (2010). They found that our formula gave an excellent description of redshift distortions in simulated data and they could provide a slight improvement over current two-point measurements in the SDSS spectroscopic LRG sample, in agreement with our

own estimates. However, future surveys, such as Euclid and BOSS, will have a larger sky coverage making wide-angle corrections important.

I split my second paper, Pápai and Szapudi (2010) into two parts (Chapter 3 and 4) , because this way the apparent leap from large-scale structure to ISW becomes a natural step along the course of research. Both parts are motivated by the tantalizing observation by Granett *et al.* (2008a), who detected an unexpectedly large correlation between superstructures in the galaxy distribution and the CMB. The ISW effect (Sachs and Wolfe, 1967), the SZ effect (Sunyaev and Zeldovich, 1972), and extragalactic radio contamination (Rudnick *et al.*, 2007, e.g.,) can all contribute to a signal. Since for $l < 200$ the SZ effect is negligible compared to the ISW effect and the signal is consistently present in all wavebands, we focus on the ISW effect. First, we measure CIC, which is an alternative to the n-point functions to describe the PDF of the galaxy distribution. Since CIC has received relatively little attention recently, this is also an opportunity to work out difficulties specific to SDSS and spectroscopic surveys. The mask and the selection function are usually taken care of by adding dummy galaxies to the data, to make the radial and angular selection functions uniform. This solves a problem by creating another; namely, it introduces a systematic bias, because the dummy galaxies are drawn from a Poisson distribution. The redshift distortion also complicates the picture. We study these effects in simulations and develop a phenomenological approach, which is adequate at the current level of noise and cosmic variance (Figure 3.6). The purpose of the measurement is to see whether the tail of the PDF differs that of a Gaussian, because this could explain a higher signal from extreme fluctuations. We measure the variance, skewness and kurtosis and we find a good agreement between them and ΛCDM predictions (see, Table 3.2 and Figure 3.7). We cannot confirm any significant deviation from a Gaussian distribution (Figure 4.1); however, the uncertainty could be decreased with data spanning a larger volume.

The ISW effect is the single most important secondary source of anisotropy of the CMB on large scales (see, e.g., Hu and Dodelson, 2002). It is the temperature shift of CMB photons due to passing through the changing gravitational potential. In a static universe, the net shift would be zero; thus, this effect is a direct measure of the growth function, which is the factor carrying the time dependence of the density field in linear perturbation theory: $\delta(x,\tau) = \delta(x,0)D(\tau)$. In Chapter 4, we estimate the temperature shift of a photon passing through the center of a spherical region with a certain overdensity or underdensity. We first calculate the average density profile for such structures (Figure 4.2) and from the density profile, we calculate the average temperature shift (Figure 4.3). Since the equations are linear, this can be done consecutively. The variance of the density profile is slightly more complicated than its average and it provides an approximation for the uncertainty in the temperature also. Using the results of Chapter 3, we approximate the PDF with analytic functions and calculate the ISW effect owing to the largest fluctuations in a survey similar to the SDSS photometric LRG sample. The result is plotted in Figure 4.4, which is the main result of Chapter 4. We compare our results to previous estimates based on a compensated top-hat potential. This is a valid approximation on nonlinear scales, where perturbations with steep enough density contrast in the initial density field, rapidly converge to a compensated top-hat (see, e.g., Sheth and van de Weygaert, 2004). This model underestimates the ISW effect and is unrealistic on large scales, whereas our uncompensated model appears to reproduce the amplitude of the signal of Granett *et al.* (2008a) within the combined uncertainty of the measured PDF and the intrinsic uncertainty of the ISW effect.

In Chapter 5, we continue with calculating the whole temperature profile of spherical superstructures and test the validity of our model in N-body simulations. The importance of cosmic variance in the case of the potential adds to the complexity of the problem (Figure 5.2), since the largest modes in the potential are dominating. It is necessary to

remove these modes from the CMB. For simulations, the theory works well (Figures 5.3 and 5.4). From the calculated profile, we calculate, we make an ISW map using the coordinates of superstructures identified in the SDSS photometric LRG sample (Granett *et al.*, 2008b). After high-pass filtering both the ISW map and the CMB, we apply the matched filter technique to measure the amplitude of the signal (Figure 5.7). We obtain a positive detection with high significance ($\sigma \approx 3.24$), however, the amplitude of the signal is lower than theoretical prediction from ΛCDM. The contents of this chapter comprise the body of Pápai *et al.* (2010), which we submitted to *The Astrophysical Journal*.

Our goal with this investigation is to create a tool for studying the redshift dependence of the growth function, by studying the ISW imprint of superstructures at different redshifts. As of today, the data are too sparse for a quantitative study, and the previously reported structures correspond to an unexpectedly strong signal.

References

Granett, B. R., M. C. Neyrinck, and I. Szapudi, 2008a, ApJ **683**, L99.

Granett, B. R., M. C. Neyrinck, and I. Szapudi, 2008b, ArXiv e-prints 0805.2974.

Hu, W., and S. Dodelson, 2002, ARA&A **40**, 171.

Pápai, P., and I. Szapudi, 2008, MNRAS **389**, 292.

Pápai, P., and I. Szapudi, 2010, ApJ **725**, 2078.

Pápai, P., I. Szapudi, and B. R. Granett, 2010, ArXiv e-prints 1012.3750.

Raccanelli, A., L. Samushia, and W. J. Percival, 2010, MNRAS **409**, 1525.

Rudnick, L., S. Brown, and L. R. Williams, 2007, ApJ **671**, 40.

Sachs, R. K., and A. M. Wolfe, 1967, ApJ **147**, 73.

Sheth, R. K., and R. van de Weygaert, 2004, MNRAS **350**, 517.

Sunyaev, R. A., and Y. B. Zeldovich, 1972, Comments on Astrophysics and Space Physics **4**, 173.

Appendix A

Gaussian fluctuations (supplement for Chapter 4)

In the main body of the dissertation, I omitted the simple but tedious calculations and I only presented the final results. In Chapter 4 and 5, I used the average of spherical fluctuations from a multivariate Gaussian conditional probability and their covariance. Equations (4.9)-(4.15) summarize the premise and the conclusion. To fill in the gap in the argument, let us consider an n-dimensional Gaussian distribution:

$$P(\vec{\delta}, C) = \frac{1}{\sqrt{(2\pi)^n |C|}} \exp\left(-\frac{1}{2}\vec{\delta}^T C^{-1} \vec{\delta} \right), \tag{A.1}$$

where $\vec{\delta} = (\delta_1, ..., \delta_n)$, and $|C|$ is the determinant of the positive definite symmetric matrix, C. Furthermore, a well known identity, due to which C is referred to as the covariance matrix, is:

$$\begin{aligned} \left\langle \delta_i \delta_j \right\rangle &= \int d\delta_1 ... \int d\delta_n \delta_i \delta_j P(\vec{\delta}, C) \\ &= \frac{-1}{|C|} \frac{\partial}{\partial C_{ij}^{-1}} |C| = C_{ij} \qquad . \end{aligned} \tag{A.2}$$

I made use of following results from linear algebra:

$$d(|CC^{-1}|) = C^{-1}d|C| + CdC^{-1}, \tag{A.3}$$

$$C^{-1}_{ij} = |C|_{ji}/|C|, \tag{A.4}$$

$$\frac{\partial}{\partial C_{ij}}|C| = |C|_{ij}, \tag{A.5}$$

where $|C|_{ij}$ is the subdeterminant with the appropriate sign belonging to the C_{ij} matrix element. One consequence of Equation (A.2) is the relationship between the covariance matrices C_n and C_{n-1} which are defined by:

$$\int d\delta_n P(\vec{\delta}^{(n)}, C^{(n)}) = P(\vec{\delta}^{(n-1)}, C^{(n-1)}). \tag{A.6}$$

$C^{(n-1)}$ is just the $\{1 : n-1\}X\{1 : n-1\}$ submatrix of $C^{(n)}$ and because for its elements

$$C^{(n-1)}_{ij} = C^{(n)}_{ij} = \left\langle \delta_i \delta_j \right\rangle \tag{A.7}$$

must hold for $\forall i, j < n$.

The elements of the vector $\vec{\delta}$ are the average density contrast in $dr \to 0$ thick shells, with δ_1 being δ_{in}, the known average density contrast of the superstructure inside R radius. The constrained probability is given by:

$$P\big(\vec{\delta}(r)|\delta_{in}, C, \sigma^2\big) = P\big(\vec{\delta}(r), \delta_{in}, C\big)/P\big(\delta_{in}, \sigma^2\big)$$
$$= \frac{\sqrt{(2\pi)\sigma^2}}{\sqrt{(2\pi)^n|C|}} \exp\left(-\frac{1}{2}\vec{\delta}^T(r)C^{-1}\vec{\delta}(r)\right)/\exp\left(-\frac{1}{2}\delta^2_{in}/\sigma^2\right) \quad , \tag{A.8}$$

where $\sigma^2 = \left\langle \delta^2_{in} \right\rangle$, a 1-dimensional covariance matrix. If one wants to calculate the average density contrast in a shell at r, every other component of $\vec{\delta}(r)$ (the density contrasts of another shells) will be integrated out of this formula leaving only the $2X2$ covariance

matrix between $\delta(r)$ and δ_{in} as shown by Equations (A.6) and (A.7):

$$\begin{aligned}\left\langle \delta(r) \right\rangle_{\delta_{in}} &= \int d\delta(r)\delta(r)P\big(\delta(r)|\delta_{in}, C^{(2)}, \sigma^2\big) = \\ &= \frac{\sqrt{(2\pi)\sigma^2}}{\sqrt{(2\pi)^2|C^{(2)}|}} \int d\delta(r)\delta(r) e^{-\frac{1}{2}(\delta(r),\delta_{in})^T C^{(2)-1}(\delta(r),\delta_{in})} / e^{-\frac{1}{2}\delta_{in}^2/\sigma^2}, \end{aligned} \tag{A.9}$$

since the expected value of Gaussian variable is the the one with the maximum likelihood, we can write down the result without any calculation:

$$\begin{aligned}\left\langle \delta(r) \right\rangle_{\delta_{in}} &= -\frac{C_{12}^{(2)-1}}{C_{11}^{(2)-1}}\delta_{in}, \\ &= \frac{C_{12}^{(2)}}{C_{22}^{(2)}}\delta_{in} = \frac{\left\langle \delta(r)\delta_{in} \right\rangle}{\left\langle \delta_{in}^2 \right\rangle}\delta_{in}. \end{aligned} \tag{A.10}$$

For the last two steps, I used Equations (A.2) and (A.4). This is the profile used in Chapter 4 and 5. This expression is normalized correctly, since:

$$\int_0^R dr 4\pi r^2 \left\langle \delta(r) \right\rangle_{\delta_{in}} = \frac{4\pi R^3}{3}\frac{\left\langle \delta_{in}^2 \right\rangle}{\left\langle \delta_{in}^2 \right\rangle}\delta_{in} = \frac{4\pi R^3}{3}\delta_{in}. \tag{A.11}$$

The covariance between shells can be obtained similarly:

$$\left\langle \delta_{r_1}\delta_{r_2}\right\rangle_{\delta_{in}} = \int\int d\delta_{r_1}d\delta_{r_2}\delta_{r_1}\delta_{r_2}P\left(\delta_{r_1},\delta_{r_2}|\delta_{in},C^{(3)},\sigma^2\right) =$$

$$= \frac{\sqrt{(2\pi)\sigma^2}}{\sqrt{(2\pi)^3|C^{(3)}|}}\int\int d\delta_{r_1}d\delta_{r_2}\delta_{r_1}\delta_{r_2}e^{-\frac{1}{2}(\delta_{r_1},\delta_{r_2},\delta_{in})^T C^{(3)-1}(\delta_{r_1},\delta_{r_2},\delta_{in})}/e^{-\frac{1}{2}\delta_{in}^2/\sigma^2}$$

$$= -\frac{\sqrt{(2\pi)\sigma^2}}{\sqrt{(2\pi)^3|C^{(3)}|}}e\frac{1}{2}^{\delta_{in}^2/\sigma^2}2\frac{\partial}{\partial C_{12}^{(3)-1}}\int\int d\delta_{r_1}d\delta_{r_2}e^{-\frac{1}{2}(\delta_{r_1},\delta_{r_2},\delta_{in})^T C^{(3)-1}(\delta_{r_1},\delta_{r_2},\delta_{in})}$$

$$= -\frac{\sqrt{(2\pi)\sigma^2}}{\sqrt{(2\pi)^3|C^{(3)}|}}e\frac{1}{2}^{\delta_{in}^2/\sigma^2}2\frac{\partial}{\partial C_{12}^{(3)-1}}\left(\frac{\sqrt{(2\pi)^3|C^{(3)}|}}{\sqrt{(2\pi)\sigma^2}}e^{-\frac{1}{2}\delta_{in}^2/\sigma^2}\right)$$

$$= -\frac{1}{|C^{(3)}|}\frac{\partial|C^{(3)}|}{\partial C_{12}^{(3)-1}} + \frac{1}{\sigma^2}\frac{\partial\sigma^2}{\partial C_{12}^{(3)-1}} - \frac{\delta_{in}^2}{(\sigma^2)^2}\frac{\partial\sigma^2}{\partial C_{12}^{(3)-1}}$$

$$= C_{12} - \frac{C_{13}C_{23}}{C_{33}} + \frac{C_{13}C_{23}\delta_{in}^2}{C_{33}^2}$$

$$= \frac{1}{\left\langle\delta_{in}^2\right\rangle}\left[\left\langle\delta_{r_1}\delta_{r_2}\right\rangle\left\langle\delta_{in}^2\right\rangle - \left\langle\delta_{r_1}\delta_{in}\right\rangle\left\langle\delta_{r_2}\delta_{in}\right\rangle\right] + \frac{1}{\left\langle\delta_{in}^2\right\rangle^2}\left\langle\delta_{r_1}\delta_{in}\right\rangle\left\langle\delta_{r_2}\delta_{in}\right\rangle\delta_{in}^2. \quad \text{(A.12)}$$

In the calculation, I used that $\sigma^2 = C_{33}^{(3)}$ and the easy to prove $\frac{\partial C_{33}^{(3)}}{\partial C_{12}^{(3)-1}} = -C_{13}C_{23}$. The last term in this expression is $\left\langle\delta_{r_1}\right\rangle_{\delta_{in}}\left\langle\delta_{r_2}\right\rangle_{\delta_{in}}$ as can be seen from Equation (A.10). This yields Equation (4.15):

$$\left\langle\delta_{r_1}\delta_{r_2}\right\rangle_{\delta_{in}} - \left\langle\delta_{r_1}\right\rangle_{\delta_{in}}\left\langle\delta_{r_2}\right\rangle_{\delta_{in}} = \frac{1}{\left\langle\delta_{in}^2\right\rangle}\left[\left\langle\delta_{r_1}\delta_{r_2}\right\rangle\left\langle\delta_{in}^2\right\rangle - \left\langle\delta_{r_1}\delta_{in}\right\rangle\left\langle\delta_{r_2}\delta_{in}\right\rangle\right]. \quad \text{(A.13)}$$

A statement from Chapter 5 , that the ratio of two Gaussian variables has a Lorentzian distribution, is addressed for last:

$$\begin{aligned} P(z) &= \int\int dxdy\delta_D(z-x/y)P_1(x)P_2(y) \\ &= \frac{1}{2\pi\sigma_1\sigma_2}\int dy|y|e^{-\frac{1}{2}\frac{z^2y^2}{\sigma_1^2}}e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}} \\ &= \frac{1}{2\pi\sigma_1\sigma_2}\frac{1}{\frac{z^2}{2\sigma_1^2}+\frac{1}{2\sigma_2^2}}2\int_0^\infty d\tilde{y}\tilde{y}e^{-\tilde{y}^2} = \frac{1}{\pi}\frac{1}{\frac{\sigma_2}{\sigma_1}z^2+\frac{\sigma_1}{\sigma_2}}. \end{aligned} \tag{A.14}$$

The Lorentz distribution has a divergent variance, which causes estimators of this kind to have a large uncertainty. This can be mended if one introduces the cut, $y_0 < |y|$:

$$\begin{aligned} P(z) &= \theta(y_0,\sigma_2)\int dx\int dy\delta_D(z-x/y)H(|y|>y_0)P_1(x)P_2(y) \\ &= \frac{\theta(y_0,\sigma_2)}{2\pi\sigma_1\sigma_2}\frac{1}{\frac{z^2}{2\sigma_1^2}+\frac{1}{2\sigma_2^2}}2\int_{\tilde{y}_0}^\infty d\tilde{y}\tilde{y}e^{-\tilde{y}^2} = \frac{\theta(y_0,\sigma_2)}{\pi}\frac{e^{-\tilde{y}_0^2}}{\frac{\sigma_2}{\sigma_1}z^2+\frac{\sigma_1}{\sigma_2}} \\ &= \frac{\theta(y_0,\sigma_2)}{\pi}e^{-\frac{y_0^2}{2\sigma_2^2}}\frac{1}{\frac{\sigma_2}{\sigma_1}z^2+\frac{\sigma_1}{\sigma_2}}e^{-\frac{z^2y_0^2}{2\sigma_1^2}}, \end{aligned} \tag{A.15}$$

where $H(x)$ is the Heaviside function and $\theta(y_0,\sigma_2) = \left(2\int_{y_0}^\infty dye^{-\frac{1}{2}\frac{y^2}{sigma_2^2}}\right)^{-1}$ normalizes the P_2 with the cut-off. The result will have an exponential tail, which guarantees that its every moment exists.

Appendix B

Frequently used abbreviations

WMAP	Wilkinson Microwave Anisotropy Probe
CMB	cosmic microwave background
LIGO	Laser Interferometer Gravitational-Wave Observatory
(C)DM	(cold) dark matter
ΛCDM	most widely accepted cosmological model with cosmological constant (Λ) and cold dark matter
PDF	probability distribution function
SDSS	Sloan Digital Sky Survey
CAMB	software by Anthony Lewis
LRG	luminous red galaxy
HVS	Hubble Volume Simulation
CIC	counts in cells
COBE	Cosmic Background Explorer
ISW	Integrated Sacks-Wolfe (effect)
SZ	Sunyaev-Zeldovich (effect)
Mpc	Mega parsec
R.A., decl.	angles of the equatorial coordinate system