Probing the Large-Scale Homogeneity of the Universe with Galaxy Redshift Surveys

Cristiano Giovanni Sabiu, BSc

Thesis
submitted to the
University of Glasgow
for the degree of
MSc

UNIVERSITY
of
GLASGOW

Astronomy & Astrophysics Group
Department of Physics & Astronomy
University of Glasgow
Scotland
September 2006

dedicato alla

mia nonna

&

al mio nonno

Probing the Large-Scale Homogeneity of the Universe with Galaxy Redshift Surveys

Cristiano Sabiu BSc

Submitted for the degree of MSc
September 2006

Abstract

Modern cosmological observations clearly reveal that the universe contains a hierarchy of clustering. However, recent surveys show a transition to homogeneity on large scales. The exact scale at which this transition occurs is still a topic of much debate. There has been much work done in trying to characterise the galaxy distribution using multifractals. However, for a number of years the size, depth and accuracy of galaxy surveys was regarded as insufficient to give a definitive answer. One of the main problems which arises in a multifractal analysis is how to deal with observational selection effects: *i.e.* 'masks' in the survey region and a geometric boundary to the survey itself.

In this thesis I will introduce a *volume* boundary correction which is rather similar to the approach developed by Pan and Coles in 2001, but which improves on their *angular* boundary correction in two important respects: firstly, our volume correction 'throws away' fewer galaxies close the boundary of a given data set and secondly it is computationally more efficient.

After application of our volume correction, I will then show how the underlying generalised dimensions of a given point set can be computed. I will apply this procedure to calculate the generalised fractal dimensions of both simulated fractal point sets and mock galaxy surveys which mimic the properties of the recent IRAS PSCz catalogue.

Acknowledgments

Firstly, I must thank my two supervisors Dr. Martin Hendry & Dr. Luis Teodoro who have kept me on the right track throughout my research.

I express my gratitude to my parents, Sheena & Antonio, and my grand parents Maria-Elena & Giovanni Sabiu, who have been a constant support to me through my whole life: This thesis is for you.

To my friend (& office-mate) Luis, thank you for teaching me something new each day. I hope we will work on interesting things together in the future.

Thanks also to some great guys, Paul Holt & Eamon Scullion; what can I say? Thanks for letting me holiday in the surreal world you two live in. Alejandro 'Speedy' Gonzalez, Russell 'Render' Johnston & Robert McKay.

Finally, in the spirit of leaving the best till last, I wish to thank my fiancée (at the time of writing) Georgie Mooney for sharing her life with me.

Declaration

The work in this thesis is based on research carried out at the Department of Physics & Astronomy within the Astronomy & Astrophysics Group of the University of Glasgow, Scotland. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Contents

List of Figures

Abbreviations

2dF	2 degree field (galaxy redshift survey)
2PCF	Two-Point Correlation Function
BAO	Baryon Acoustic Oscillations
CDM	Cold Dark Matter
CfA	Center for Astrophysics
CMB	Cosmic Microwave Background
CP	Cosmological Principle
Gpc	Gigaparsec
ISM	Interstellar Medium
LG	Local Group
LSS	Large Scale Structure
Mpc	Megaparsec
MST	Minimal Spanning Trees
PSCz	Point Source Catalogue (redshift)
RWM	Robertson Walker Metric
SDSS	Sloan Digital Sky Survey
WMAP	Wilkinson Microwave Anisotropy Probe

Chapter 1

Introduction

This thesis is devoted to a study of the large scale structure (LSS) of the universe and as such belongs in the field of Cosmology. Cosmology is the study of structure and evolution in the universe. The main constituents of LSS are individual galaxies and clusters of galaxies up to Gigaparsec (Gpc) scales.

1.1 Cosmology through the ages

The ancient Greeks were undoubtedly the leaders in astronomical understanding of their time. Around the 4th century BC a general consensus emerged, from the combined ideas of many philosophers, including Plato and Aristotle, which put our spherical Earth at the centre of the universe. They speculated that the Sun, Moon and planets were carried around the Earth on concentric spheres, arranged: Moon, Sun, Venus, Mercury, Mars, Jupiter, Saturn and the fixed stars beyond. Aristotle was to later elaborate on this *geocentric* model by trying to explain the lunar cycle.

An updated *geocentric* theory of the heavens was put together by the astronomer Claudius Ptolemy from many works in Greek astronomy. His *Ptolemaic* model was penned in the 2nd century AD and stood as the standard theory for more than a millennium. Ptolemy made extensive use of epicycles to explain many aspects of planetary motion. In particular, his epicyclic explanation of retrograde motion in the planets helped elevate this theory to the forefront of astronomical thinking. This picture of our universe stood solid until the 16th century when Nicolaus Copernicus changed forever our view of the cosmos.

Copernicus, the famous Polish astronomer, showed that a model, with the sun at its centre, could explain the motion of the planets in a very simple way, with no need for complicated orbits and epicycles. However this *Heliocentric* model was not new at all: its origins dated back many centuries BC to the workings of an Indian philosopher, Yajnavalkya. He had the vision to see that the sun, being the most important of the heavenly bodies, should be at the centre of our universe. However he lacked any real observational or scientific evidence. This Heliocentric idea was also present in ancient Greece, held strong by the Pythagoreans. The first to propose this was Aristarchus of Samos (c. 270 BC) and later Archimedes, the Greek scientist, was swayed by this line of thinking.

1.2 The Cosmological Principle

Despite the greater simplicity involved with the Heliocentric system, it would not come to dominate the astronomical community. This was, at least in part, not for scientific reason but religious prejudices. Many religions held the false belief that the Earth was somehow special and therefore must be the centre of our observable universe. The Roman Catholic Church had a strong hold on science and particularly astronomy (due to its close connection to the heavens). Any theories

which did not conform to the teachings of the Bible were deemed 'heretical' and were hidden away from public knowledge. Despite this dogma, Copernicus in the 16th century managed to garner much support for the Heliocentric model, mainly due to his scientific writings, *De Revolutionibus* (1543) and Galileo's supporting observations. Galileo later opposed the Catholic Church by his strong support for the *Copernican* ideas. While on trial for heresy he famously said of the Earth, "Eppur si muove" - and yet it moves[1].

So why are we dwelling in the past here? The reason is this; Copernicus and Galileo did not see themselves at the centre of the solar system. They had displaced everyone from a special location in space and this trend would not stop with the Earth. It would later lead to the whole solar system being placed in the outer rim of our Milky way galaxy. Then, in the early 20^{th} century our galaxy became one of many. Now we are but a mere speck of dust in a vast, ever expanding universe, a fact that was first realised by the American astronomer Edwin Hubble in the 1920's.

Edwin Hubble

Edwin Hubble studied the systematic variations of the red-shift in the 'spiral nebulae', as they were known. The redshift z is defined as,

$$z = \frac{\lambda_O}{\lambda_E} - 1, \tag{1.1}$$

where λ_O and λ_E are the observed and emitted wavelengths of light, respectively. Hubble used this technique to investigate populations of similar objects, usually galaxies of a particular morphological type, and examined the relationship between the red-shifts and their relative brightnesses. What he found is now known as

[1]Galileo probably never spoke these exact words, however they stand as a symbol of his support for scientific truth.

Hubble's Law: the red-shift in the spectra of the objects grow as the objects became more distant. The farther away an object, the faster it is receding from us. This, Hubble concluded, is because the universe itself is expanding – a fact that (as we will see in the next Chapter) was consistent with the theoretical predictions of Einstein's General Theory of Relativity

In fact Hubble found this connection to be a linear relation between recessional velocity and distance. The usual Hubble law is written as,

$$v = H_0 D, \tag{1.2}$$

where v is the recessional velocity due to redshift, typically expressed in km/s. H_0 is Hubble's constant and corresponds to the value of H (often termed the Hubble parameter which is a value that is time dependent) in the Friedmann equations (*cf.* eq.2.13) taken at the time of observation denoted by the subscript 0. This value is the same throughout the universe for a given conformal time. D is the proper distance that the light had travelled from the galaxy in the rest frame of the observer, measured in megaparsecs (Mpc).

For relatively nearby galaxies (*i.e.* $z \ll 1$), the velocity v can be estimated from the galaxy's redshift z using the formula $v = zc$ where c is the speed of light. For more distant galaxies, the relationship between recession velocity and distance becomes more complicated and requires general relativity (see Ch.2).

In using Hubble's law to determine distances, only the velocity due to the expansion of the universe should be used. Since gravitationally interacting galaxies move relative to each other independent of the expansion of the universe, these relative velocities, called peculiar velocities, need to be accounted when applying Hubble's law. So more generally the Hubble law is,

$$v_{rec} = \mathrm{H}_0 D + v_{pec}. \tag{1.3}$$

In this case, v_{pec} is the radial component of the peculiar motion of the object. As an example the local group of galaxies has a $v_{pec} \approx$ 600 km s^{-1} in the direction of the constellation Hydra.

It is straightforward to show that the observation of Hubble's Law is consistent with what we would expect in a universe which is homogeneous (*i.e.* looks the same everywhere) and isotropic (*i.e.* looks the same in all directions). A number of modern cosmological observations support the properties of homogeneity and isotropy, including the distribution of galaxies on large scales (which will be the main topic of this thesis) and the smoothness of the cosmic background radiation. Together, assumptions of homogeneity and isotropy are known as the Cosmological Principle (CP).

1.3 Our view of the Universe

If one is to understand anything about the large scale structure of the universe, it is generally advisable to know where the galaxies that make up that structure are. Mapping and understanding the spatial galaxy distribution is a prerequisite for constructing a viable model of structure formation in the universe. This effort reached its first step with the Abell, Zwicky & Lick catalogues, which eventually documented the angular positions of around a million galaxies. The second step then, was to expand the 2-D galaxy information by including the distances. This is distance to the galaxies via redshift surveys.

Abell, Zwicky & Lick

Prior to the 80's, knowledge of the large scale structure of the universe was limited to only the angular distributions of galaxies, and a very uniform microwave

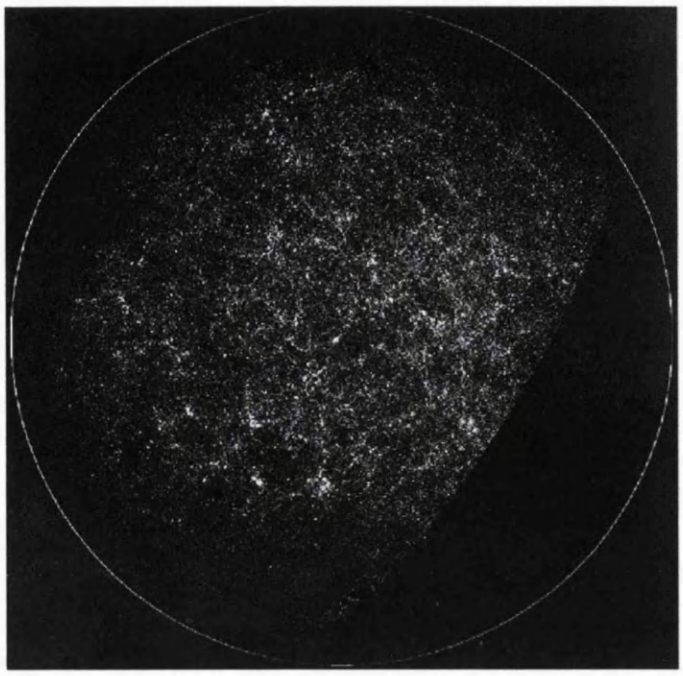

Figure 1.1: Lick galaxy survey, adapted from Peebles (1993)

background. In the late 50's, G. Abell (1958) collected several thousand angular positions of galaxies from the Palomar Sky Survey. This catalogue did not contain any information concerning the distance to the galaxies – it was essentially a projection of the true galaxy distribution onto a sphere. Then in the 60's Fritz Zwicky and collaborators visually scanned thousands of photographic plates from the same survey, obtaining positions of over 30,000 galaxies in the northern sky (Zwicky*et al.* 1968). After that, in the early 80's, Schechtman accumulated a catalogue of 1 million galaxies in the northern sky from the Lick astrographic survey, see figure 1.1.

The Cambridge APM survey followed in the early 90's cataloguing about 2 million galaxies in the Southern Galactic Cap. Maddox *et al.* 1990. See figure 1.2

Figure 1.2: The APM galaxy survey. Maddox *et al.* 1990.

Redshift Surveys

Redshift measurements involve determining the spectrum of the object to be measured. Once that is known, recognisable spectral lines can be found, and their deviation from their normal positions used to find the object's redshift. The Hubble Law then allows one to turn that redshift into a radial distance from our galaxy.

When redshifts were first being measured, it would typically take a few hours on a large telescope to collect enough photons to obtain the required spectrum. Once telescopes with enough light gathering power became available and spectroscopic detectors became sophisticated enough to allow many redshifts to be taken simultaneously in a reasonable amount of time, astronomers started using these instruments to make maps of the 3 dimensional locations of galaxies and galaxy clusters.

Figure 1.3 is a representation of some of the measured 3 dimensional galaxy positions in redshift space. The radial coordinate in this plot is the measured redshift (essentially indicating the distance from us) and the angular coordinates

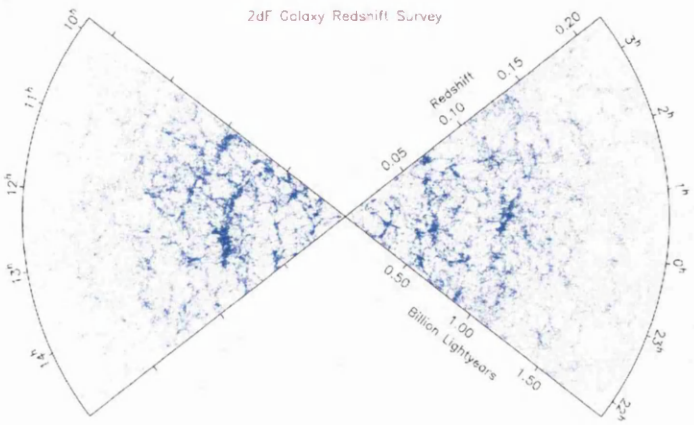

Figure 1.3: 2dF galaxy redshift survey.

represent the angular position of the objects in the sky. Thus the 3-D mapping of the universe began:

- In the late 70's red-shift surveys finally became a reality with the very successful CfA (Center for Astrophysics) survey (Huchra *et al.* 1983). They managed to record 1100 spectroscopic red-shifts.

- In the early 90's the PSCz red-shift survey (Saunders *et al.* 2000) mapped 15,000 spiral galaxies from $\sim 83\%$ of the sky. This still stands as the largest survey in terms of sky coverage.

- From 1998 to 2003, the Two Degree Field Galaxy Redshift Survey, using the Anglo-Australian telescope, accumulated 220,000 galaxy red-shifts. See Colless *et al.* 1999. This survey is illustrated in figure 1.3.

- The Sloan Digital Sky Survey (SDSS) is an ongoing attempt to collect 1 million galaxy red-shifts. The current release, DR5, contains 674,749 galaxies. See Percival *et al.* 2006 (DR5) and Abazajian *et al.* 2005 (DR4).

1.3.1 Problems with this view

The information directly available to us on the observed spatial distribution of galaxies is systematically biased compared to the true galaxy distribution.

Angular positions are trivial, but precise distance estimates are more difficult to obtain. More serious, however, is the problem that the sampling rate of observable galaxies depends strongly on redshift. If one considers the universe to be homogeneously distributed with matter, then we would expect the number of objects, in some observed solid angle of the sky, to grow as r^3. However this is not what we actually observe, instead we see distributions like figure (5.5), where the reference curve initially has the form of r^3 but soon drops off. This effect is due the diminished flux of faraway objects and our inability to collect enough photons from them to make valid observations. This 'selection effect' can be visualised in figure (1.3), as the fall off in the apparent number density of sampled galaxies at larger redshift. This is accounted for by weighting each observed object with the inverse of the *selection function*, $\phi(r)$, this increases the contribution of counted objects. The functional form of the *selection function* is,

$$\phi(r) = \left(\frac{r_\circ}{r}\right)^{2\alpha} \left(\frac{r_\star^2 + r_\circ^2}{r_\star^2 + r^2}\right)^{\beta}. \tag{1.4}$$

In the above expression $r_\star, \alpha, \beta$ and r_0 are set parameters and r is the proper distance to the observed galaxy.

Apart from the red-shifts, distances can also be estimated from the flux, F_υ, received from a source. To accurately use this method, the intrinsic luminosity, L_υ, of the source must be well known and that it radiates in a particular fashion, *i.e.* beamed or uniform. Considering the source to radiate uniformly, the distance d can be calculated from,

$$d_L^2 = \frac{L_\upsilon}{4\pi F_\upsilon}. \tag{1.5}$$

The main problem with this distance estimator is that the vacuum of space is not so empty. It is filled with dust particles and gas, collectively termed the Interstellar Medium (ISM). The ISM can cause extinction of light (diminished flux) and leads to eq.(1.5) giving the wrong answer. The extinction of light is maximal at low galactic latitudes due to the dust content of our own galaxy. In fact at its worst this effect can completely block out the light coming from distant sources.

In figure 5.4, the masked regions (black) are shown for the *PSCz* survey. The majority of it is due to the extinction of light through the galaxy. The sweeping arcs (north and south) are due to cryogenic problems in the satellite near the end of the survey period, it was not completed. In figure 4.1, some masked regions for the 2dF survey are shown. These masks are placed over the survey because local object may be blocking the view in a particular direction.

Chapter 2

The Universe at Large

Of the four fundamental forces of nature, the universe on large scales is governed by a single force: Gravity. In the proceeding section we will employ the use of Einstein's General Relativity to construct the main equations of cosmological evolution. From there we will relax our use of complicated tensor algebra and work in a Newtonian approximation to study the departures from the homogeneous Friedmann equations, due to the growth under gravity of tiny density inhomogeneities in the Universe. This is the 'first order' Universe.

2.1 General Relativity

In 1915 Einstein developed the theory of General Relativity to explain gravity as a consequence of the fundamental connection between the geometry and matter content of space-time. This can be summed up in the neat phrase 'Space-time tells matter how to move and matter tells space-time how to curve'. More specifically, Einstein developed a set of equations which balanced the curvature of space-time and the matter it contained. It is this curvature which generates the force of gravity

and determines how matter moves within it. Generally, this can be written as,

$$8\pi G T_{ik} = G_{ik} = R_{ik} - \frac{1}{2} g_{ik} R - \Lambda g_{ik}. \tag{2.1}$$

To explain the ideas and physical motivation of the above equation is beyond the scope of this thesis (for further details see Misner *et al.* 1973 and Weinberg 1972 for a more physical description). In eq.(2.1), G_{ik} is the Einstein tensor, R_{ik} and R are the Ricci tensor and scalar respectively. g_{ik} is the space-time metric and Λ is the Cosmological Constant. Although its inclusion in eq.(2.1) was described by Einstein to be his "Greatest blunder", it is nowadays a very crucial parameter in cosmology. We will describe in more detail each of the above parameters as we go along.

In our study of cosmology we need a theory of gravity, which we have, and a metric to describe the space-time. A metric is a function that measures the distance between events in space-time. Assuming that the CP is true, the form of this metric must conserve the CP as described in §1.2; *i.e.* it must be homogeneous and isotropic. To construct the metric with the CP in mind, we slice up the 4-D space-time along $x^0 = const.$ hypersurfaces, *i.e.* constant time.

To impose isotropy it is easier to tackle this problem in spherical coordinates since there is no preferred direction. For homogeneity we insist that the Ricci scalar is constant at all points on the hypersphere. And since we are dealing with a universe which is expanding it would be to our advantage to use a coordinate system which reflects this property. So we will work in a comoving system and therefore preserve the positions of galaxies relative to one another. Under the above assumptions it can then be shown (e.g. Carroll S.M, 2004) that the metric takes the following form,

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right]. \tag{2.2}$$

This is known as the Roberson-Walker metric (hereafter RWM). In the above expression $a(t)$ is the cosmic scale factor, k is a constant which determines the spatial curvature and we are working in units where $c = 1$. k may take any real value but by suitable re-scaling of our coordinates it is only necessary to consider 3 possibilities.

- $k = 0$, corresponds to zero curvature and thus a flat universe. Initially parallel trajectories remain parallel.
- $k = +1$, has positive curvature and the universe is closed. Its geometry is like the surface of a sphere. Initially parallel trajectories will converge.
- $k = -1$, has negative curvature and leads to an open universe. Its geometry can be represented by the surface of a saddle. Initially parallel trajectories will diverge.

The RWM in component notation is,

$$g_{\alpha\beta} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t)/(1-kr^2) & 0 & 0 \\ 0 & 0 & a^2(t)r^2 & 0 \\ 0 & 0 & 0 & a^2(t)r^2\sin^2\theta \end{pmatrix}. \tag{2.3}$$

With the above information about the metric we can go about obtaining the non-vanishing components of the Christoffel Symbols, $\Gamma^{\gamma}_{\beta\mu}$. The components are calculated via,

$$\Gamma^{\gamma}_{\beta\mu} = \frac{1}{2} g^{\alpha\gamma}(g_{\alpha\beta,\mu} + g_{\alpha\mu,\beta} + g_{\beta\mu,\alpha}), \tag{2.4}$$

where α is a summation index and a comma denotes partial derivative. Now it is merely a case of turning the handle to derive the essential non-zero elements of

the Christoffel Symbols for the metric given by eq.(2.2),

$$\begin{aligned}
\Gamma^0_{11} &= \frac{a\dot{a}}{1-kr^2} & \Gamma^1_{11} &= \frac{kr}{1-kr^2} \\
\Gamma^0_{22} &= a\dot{a}r^2 & \Gamma^0_{33} &= a\dot{a}r^2\sin^2\theta \\
\Gamma^1_{22} &= r(kr^2-1) & \Gamma^1_{33} &= r(kr^2-1)\sin^2\theta \\
\Gamma^2_{33} &= -\sin\theta\cos\theta & \Gamma^3_{23} &= \cot\theta \\
\Gamma^1_{01} &= \Gamma^2_{02} = \Gamma^3_{03} = \frac{\dot{a}}{a} & \Gamma^2_{12} &= \Gamma^3_{13} = \frac{1}{r}
\end{aligned} \tag{2.5}$$

From the above set of equations we can now determine the nonzero components of the Ricci tensor, $R_{\alpha\beta}$.This is constructed by contraction of the Riemann tensor, $R^{\lambda}_{\mu\nu\gamma}$.

$$R_{\alpha\beta} \equiv R^{\lambda}_{\alpha\lambda\beta}. \tag{2.6}$$

Summation is implied on the index λ according to the *Einstein summation convention.* The components of the Ricci tensor are therefore related to the Christoffel symbols, above by,

$$R_{\alpha\beta} = \Gamma^{\tau}_{\alpha\beta}\Gamma^{\rho}_{\tau\rho} - \Gamma^{\tau}_{\alpha\rho}\Gamma^{\rho}_{\tau\beta} + \Gamma^{\rho}_{\alpha\beta,\rho} - \Gamma^{\rho}_{\alpha\rho,\beta}. \tag{2.7}$$

In the above expression summation is implied over τ and ρ. As an example this leads to, for the tt component:

$$\begin{aligned}
R_{tt} &= \Gamma^{\tau}_{tt}\Gamma^{\rho}_{\tau\rho} - \Gamma^{\tau}_{t\rho}\Gamma^{\rho}_{\tau t} + \Gamma^{\rho}_{tt,\rho} - \Gamma^{\rho}_{t\rho,t} \\
&= 0 - \Gamma^{\tau}_{t\rho}\Gamma^{\rho}_{\tau t} + 0 - \Gamma^{\rho}_{t\rho,t} \\
&= -3\left(\frac{\dot{a}}{a}\right)^2 - 3\frac{d}{dt}\left(\frac{\dot{a}}{a}\right) \\
&= -3\frac{\ddot{a}}{a}.
\end{aligned} \tag{2.8}$$

The rest of the nonzero components are:

$$\begin{aligned} R_{rr} &= \frac{a\dot{a} + 2\dot{a}^2 + 2k}{1 - kr^2}, \\ R_{\theta\theta} &= r^2(a\dot{a} + 2\dot{a}^2 + 2k), \\ R_{\phi\phi} &= r^2(a\dot{a} + 2\dot{a}^2 + 2k)\sin^2\theta. \end{aligned} \tag{2.9}$$

The Ricci scalar, R, is then obtained via the contraction of $R_{\alpha\beta}$ i.e,

$$R = g^{\alpha\beta}R_{\alpha\beta}. \tag{2.10}$$

Therefore,

$$\begin{aligned} R &= g^{tt}R_{tt} + g^{rr}R_{rr} + g^{\theta\theta}R_{\theta\theta} + g^{\phi\phi}R_{\phi\phi}, \\ &= 6\left[\left(\frac{\dot{a}}{a}\right)^2 + \frac{\ddot{a}}{a} + \frac{k}{a^2}\right]. \end{aligned} \tag{2.11}$$

Now looking back at equation (2.1), we can compute everything on the RHS of this equation. To evaluate the LHS we must define our energy-momentum tensor, T_{ik}. This tensor describes the matter and energy content of the universe. A perfect fluid approximation is completely defined by two quantities: the rest frame energy density ρ and the isotropic rest frame pressure p.

$$T_{ik} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{2.12}$$

Now using eq.(2.12) with eq.(2.1) and assuming that $\Lambda = 0$, we can obtain two different equations. From the tt component we get,

$$\left(\frac{\dot{a}}{a}\right)^2 = H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \tag{2.13}$$

and from the other components we obtain

$$\left(\frac{\ddot{a}}{a}\right) = -\frac{4\pi G}{3}(\rho + 3p)\,. \tag{2.14}$$

These expressions are known as the Friedmann equations. H is the Hubble parameter and is defined as,

$$H = \frac{\dot{a}}{a}. \tag{2.15}$$

Apart from the global expansion, we can now investigate other aspects of the RWM universe. The *density parameter*, Ω is defined as,

$$\Omega = \frac{8\pi G}{3H^2}\rho \equiv \frac{\rho}{\rho_c}, \tag{2.16}$$

with ρ_c being the critical density required to produce a flat universe. Combining the above equation with (2.13) gives,

$$H^2 = \frac{8\pi G}{3}\rho_c\Omega - \frac{k}{a^2} = H^2\Omega - \frac{k}{a^2}, \tag{2.17}$$

and rearranging this gives,

$$\Omega - 1 = \frac{k}{a^2H^2}. \tag{2.18}$$

The special case of $\Omega = 1$ implies that $k = 0$, and since k is a fixed constant it must be concluded that $\Omega = 1$ for all time.

2.2 Structure Formation

There are two crucial observations that any model of structure formation has to explain: the quadrupole anisotropy of CMB, as measured by WMAP, is one part in 10^5 (Hinshaw *et al.* 2003) suggesting that the amplitude of the fluctuations was very small at the epoch of recombination; while redshift surveys of the Local

universe show highly inhomogeneous matter distributions over galactic and cluster scales. The *gravitational instability* is believed to be the physical mechanism which amplifies the small primeval fluctuations into the structure that we observe today.

We now have the main ingredients for analysing the evolution of structure, at least within a linear approximation, these are contained within equations (2.13)-(2.15). Combining the Friedmann equations and rearranging gives us the *continuity* equation,

$$\dot{\rho} = -3H\left(\rho + p\right), \tag{2.19}$$

which shows the conservation of energy as the universe expands. To proceed we will also make the assumption that the matter content of the universe is well described by an ideal gas equation of state, *i.e.*,

$$p = w\rho. \tag{2.20}$$

In standard cosmology we assume that the majority of the matter is in the form of cold dark matter (CDM), which has the equation of state, $w = 0$. So the pressure component is assumed to be negligible which greatly simplifies the following calculations.

In eq.(2.2) the effect of the curvature is small for distances much less than the Hubble radius $cH_0^{-1} = 3000\,h^{-1}$Mpc (a variety of observations clearly favour $\Omega_{tot} \approx 1$ and $\Omega_\Lambda \approx 0.7$ therefore $|k| < H_0^2$). Hence, the RWM is well approximated by $ds^2 = cdt^2 - a^2(t)(dx^2 + dy^2 + dz^2)$ (similar to the flat Minkowski metric of special relativity but with a time varying rescaling.), where (x, y, z) denote the comoving Cartesian coordinates. Assuming a conformal Newtonian gauge[1] (*cf.* Ma & Bertschinger 1995) the Einstein field equations applied to the first-order

[1]Also known as the longitudinal gauge. This transform is very useful when considering scalar perturbations.

perturbations ϕ of such a metric yield the Poisson equation of Newtonian gravity:

$$\nabla^2\phi = 4\pi a^2 G\delta\bar{\rho}, \tag{2.21}$$

where $\delta\bar{\rho} \equiv \rho(\mathbf{x},t) - \bar{\rho}(t)$ indicates the fluctuation of the mass density about the mean density $\bar{\rho}(t)$ and ϕ is interpreted as the Newtonian potential. Note that eq.(2.21) does not assume that $\delta\bar{\rho}$ is small.

2.2.1 The Eulerian Formalism

The Eulerian Formalism considers the large scale universe as a continually expanding fluid, whereby momentum, energy and mass conservations are encapsulated within the equations below:

$$a\dot{\delta} + \nabla\cdot[(1+\delta)\mathbf{v}] = 0 \tag{2.22}$$

$$a\dot{\mathbf{v}} + (\mathbf{v}\cdot\nabla)\,\mathbf{v} + \dot{a}\mathbf{v} = -\nabla\phi - \bar{\rho}^{-1}\nabla p, \tag{2.23}$$

$$\nabla^2\phi = 4\pi a^2 G\delta\bar{\rho}, \tag{2.24}$$

Equations (2.22)-(2.24) are the equations of motion of a non-relativistic perfect fluid in comoving coordinates. Eq.(2.22) is the continuity equation (expressing mass continuity) and eq. (2.23) is the Euler equation (conservation of the linear momentum). In this system of differential equations the over-density field, $\delta(\mathbf{x},t)$, appears rather than the usual density field,

$$\rho(\mathbf{x},t) \equiv \bar{\rho}(t)[1+\delta(\mathbf{x},t)],$$

with $\bar{\rho}$ being the spatially averaged mean density and the peculiar velocity, $\mathbf{v}(\mathbf{x}, t)$, is defined as

$$\mathbf{v} \equiv \frac{d\mathbf{r}}{dt} - \frac{\dot{a}}{a}\mathbf{r}. \tag{2.25}$$

Over-dots indicate partial time derivatives. The pressure p is related to the density ρ through equation (2.20). For *adiabatic perturbations* there are no spatial variations in the equation of state, therefore $\nabla p = w\nabla\rho = w\bar{\rho}\nabla\delta$.

The Linear regime of Adiabatic perturbations

Linearising our system of equation, we obtain

$$a\dot{\delta} + \nabla \cdot \mathbf{v} \approx 0, \tag{2.26}$$

$$a\dot{\mathbf{v}} + \dot{a}\mathbf{v} \approx -\nabla\phi - c_s^2\nabla\delta, \tag{2.27}$$

$$\nabla^2\phi = 4\pi G a^2 \delta\bar{\rho}. \tag{2.28}$$

Where c_s represents the adiabatic sound speed, $c_s^2 \equiv (\partial p/\partial\rho)_s = w$, and the subscript S indicates constant entropy throughout the space ($\nabla S = 0$).

A general vector field may be decomposed into a (potential) longitudinal and a (rotational) transversal part:

$$\mathbf{v}(\mathbf{x}, t) = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp}, \qquad \nabla \times \mathbf{v}_{\parallel} = \nabla \cdot \mathbf{v}_{\perp} = 0, \tag{2.29}$$

From the curl of eq. (2.27) it follows that

$$\frac{\partial}{\partial t}(a\nabla \times \mathbf{v}) = \frac{\partial}{\partial t}(a\nabla \times \mathbf{v}_{\perp}) = 0. \tag{2.30}$$

This implies that rotational modes are not coupled to density perturbations and decay as a^{-1}. Combining the time derivative of the linearised continuity with the

divergence of the linearised Euler (*cf.* eq. 2.27) we yield the equation of motion for the longitudinal density perturbations

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = \left(\frac{c_s}{a}\right)^2 \nabla^2\delta + 4\pi G\delta\rho. \tag{2.31}$$

Since the coefficients are spatially homogeneous (independent of $\mathbf{x}$) this equation may be solved by expanding $\delta(\mathbf{x},t)$ in plane waves, $\delta(\mathbf{x},t) = \delta_{\mathbf{k}}(t)e^{i\mathbf{k}\cdot\mathbf{x}}$, $\lambda = 2\pi a(t)/k$, where λ is the proper wavelength. After some straightforward calculations, it is easy to show that the dynamical behaviour of $\delta_{\mathbf{k}}(t)$ obeys the following differential equation:

$$\ddot{\delta}_{\mathbf{k}} + 2\frac{\dot{a}}{a}\dot{\delta}_{\mathbf{k}} = -c_s^2\left(k^2 - k_J^2\right)\delta_{\mathbf{k}}, \tag{2.32}$$

where we have defined the comoving Jeans wavenumber k_J by

$$k_J = a\left(\frac{4\pi G\bar{\rho}}{c_s^2}\right)^{1/2}. \tag{2.33}$$

Two qualitative behaviours of the solutions can be easily discerned from eq. (2.32). For wavenumbers larger than k_J pressure dominates the right hand term and perturbations do not grow, merely oscillate. For $k < k_J$ self-gravity dominates so that gravitational instability can take place. Exact solutions to eq. (2.32) exist for a variety of cases (see, for instance, Peebles 1980). Since the dynamical behaviour of $\delta_{\mathbf{k}}(t)$ is governed by a second order differential equation, in general, there is one monotonically growing solution and one monotonically decaying solution. In the limit $k \ll k_J$ the effects of the pressure p are negligible and thus all modes grow at the same rate. In this regime, the general solution to eq. (2.31) is given

$$\delta(t,\mathbf{x}) = A(\mathbf{x})D_+(t) + B(\mathbf{x})D_-(t) \approx A(\mathbf{x})D_+(t) \tag{2.34}$$

where $D_+(t)$ and $D_-(t)$ are the growing and decaying modes, respectively, while

$A(\mathbf{x})$ and $B(\mathbf{x})$ are time independent functions (Heath 1977). The decaying solution is a perturbation with initial over-density and peculiar velocity arranged so its initial velocity quickly becomes negligible (Peebles 1980). Thus for most of the history of the universe the growing solution quickly comes to dominate. In an Einstein-de Sitter universe $D_+(t) \propto a$ and $D_-(t) \propto a^{-3/2}$. For a dust universe with Ω <1, the growing mode D_+ is

$$D_+(t) = \frac{3\sinh\eta(\sinh\eta - \eta)}{(\cosh\eta - \eta)^2} - 2. \tag{2.35}$$

η indicates conformal time $\eta \equiv (-k)^{1/2}\int^t dt'/a(t')$.

Given a solution for the density perturbation field $\delta(\mathbf{x}, t)$, the velocity, gravitational potential and gravity field follow. For the longitudinal modes $\mathbf{v} = \mathbf{v}_{\|} = -\nabla\phi_v/a$. The gravity field is $\mathbf{g} \equiv -\nabla\phi/a$. Thus, from the system of differential equations, eq's. (2.26)–(2.28), for $k \ll k_J$, we obtain

$$\phi_v = \frac{a^2 H f}{4\pi}\int \frac{\delta(\mathbf{x}')}{|\mathbf{x}' - \mathbf{x}|}d^3x', \qquad \phi = \frac{3}{2}\frac{\Omega H}{f}\phi_{\mathbf{v}}, \qquad \mathbf{g} = \frac{3}{2}\frac{\Omega H}{f}\mathbf{v}. \tag{2.36}$$

Where f is defined by

$$f(\Omega, z) \equiv \frac{d\log D_+}{d\log a} = \frac{\dot{D}_+(t)}{H(t)D_+(t)}. \tag{2.37}$$

The behaviour of $f(\Omega, z)$ at the present epoch (z=0) is very well described by $f \approx \Omega^{0.6}$ in the case of universes with negligible space curvature or rather small cosmological constant (Peebles 1980).

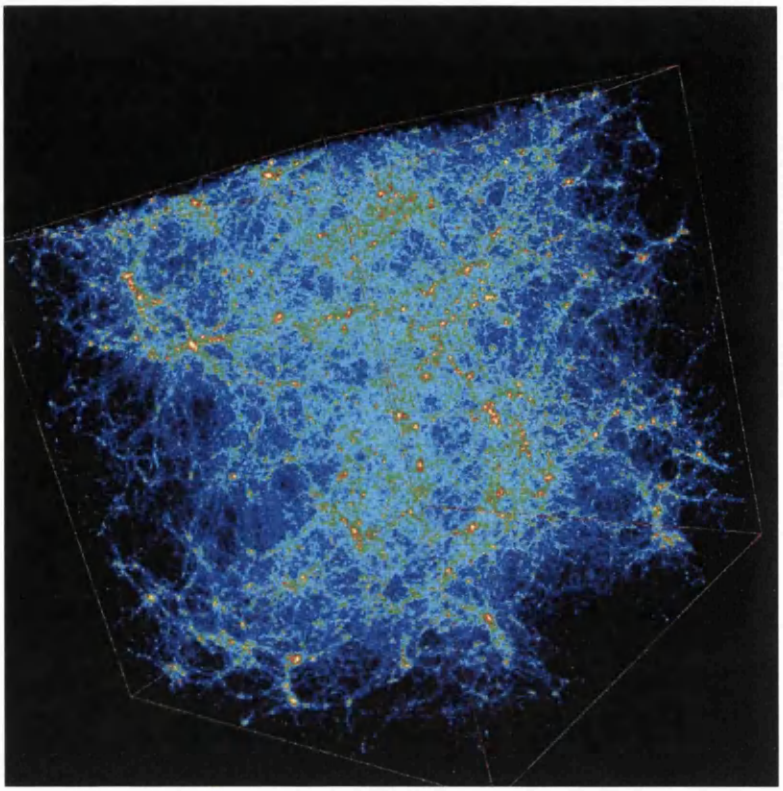

Figure 2.1: ΛCDM simulation, described briefly in § 5.2 (see Warren *et al.* 2006 for details).

2.3 N-body Simulations

So far we have only been concerned with the linear regime. To tackle the non-linear regime we must leave our analytical tools behind and instead buy a very big computer to run N-body simulations. This is exactly what people have been doing since Sverre Aarseth wrote the first astrophysical N-body codes (Aarseth, 1978)[2].

In the early days of this field, cosmological simulations could only handle roughly 1,000 particles (Goth & Turner, 1979). But with the exponential increase in computing power it was not long until very large N-body simulation were being

[2]Aarseth helped greatly in the early days of N-body simulations by making his codes readily available and easily adaptable.

carried out. By the mid-80's there were 30,000 particles being used (Efstathiou, 1985), this rose steeply to 10^6 in the 90's (Bertschinger & Gelb, 1991) and now the most recent effort by the "Virgo Consortium" has seen the first billion particle cosmological simulation (Evrard *et al.* 2002).

Over the last two decades cosmological N-body simulations have played a crucial role in the study of the formation and evolution of cosmic structure. Primarily they have been used to match theory with observations. However, simulations like the recent "millennium run" (Springel *et al.* 2005) use only dark matter particles which form the gravitational potentials for structure growth. Therefore, there is a crucial step to go from dark matter particles to dark matter halos and finally individual galaxies. The basic technique for doing so is discussed in §5.3.1. One would hope that in this process we are not masking some underlying physics.

The impressive progress achieved in the observational front with the completion of very large surveys such as 2dF and SDSS poses a clear challenge to the numerical work in cosmology: the precision of the predictions provided by the current N-body experiments have to be of the order of a few percent.

N-body simulations, however, do not include all relevant physics like magnetic fields and gas dynamics. They also have resolution issues, which must play a roll on small scales. So while modern simulations, like that of Warren *et al.* (2006) in figure 2.1, are undoubtedly very useful tools, they must be used with care.

Chapter 3

Quantifying The Large Scale Structure

The issue of quantifying structure is not confined to cosmology, it is a complex matter which stretches across many areas of science. Patterns are there to be identified and exploited in order to find subtle connections and correlations between observables. This method of analysis can be seen in fields as diverse as studying the stock market and modelling biological systems. A wide variety of different statistical tools have been employed to quantify structure, but in studies of the large-scale distribution of galaxies perhaps the most common has been the 2-point correlation function

3.1 The Two Point Correlation Function: ξ

The two point correlation function (hereafter 2PCF), $\xi(r)$, is defined as the excess probability, with respect to a Poisson distribution, of finding two galaxies in

volumes dV_1 and dV_2 separated by a distance r. The joint probability is then,

$$dP_{1,2} = n^2[1 + \xi(r)]dV_1 dV_2, \tag{3.1}$$

where n is the average density of the sample. This is however not a straight forward working definition since we know that density is varying through the space. Consequently it is not straightforward to identify how large a volume of space must be sampled in order to reliably measure the average density of the galaxy population. Equally importantly, even if such a 'fair sample' volume can be identified, the number count of observable galaxies within that volume will not generally be a reliable estimate of the 'true' number of galaxies in the volume because of observational selection effects - *i.e.* a radial selection function and angular masked regions. Also, measuring accurately the distance of galaxies in a survey is not straightforward: redshifts are distorted by peculiar motions and redshift-independent distance indicators (*e.g.* using some form of 'standard candle' assumption) are rather noisy.

For this work we proceed to measure the mean density $\bar{n}$ as the number of galaxies inside the survey, each weighted by the selection function, $\phi(r_i)$; in practice we perform a *Monte Carlo* over the sample space (see Strauss & Willick, 1995 and references therein). We can also write the conditional probability, dP_c, of finding a galaxy in a volume dV_1, at a distance r from another galaxy,

$$dP_c = \bar{n}[1 + \xi(r)]dV_1. \tag{3.2}$$

From this equation it is straightforward to see the properties of the function $\xi(r)$. For $\xi(r) = 0$, we recover a uniformly random point set, such that the probability of finding a galaxy in volume dV is simply proportional to dV. The case of excess clustering is $\xi(r) > 0$, where we have more galaxies than in the Poisson case. Then there is also the case where $\xi(r) < 0$. This case corresponds to anti-clustering, and could be relevant *e.g.* to some models of galaxy formation where the formation of

a galaxy may inhibit the formation of other galaxies in its vicinity.

In general the observed $\xi(r)$ is well described by a power law scaling with distance and it is standard practice in the literature to write the 2PCF as,

$$\xi(r) = \left(\frac{r_0}{r}\right)^{\gamma}. \tag{3.3}$$

Here r_0 is a characteristic scale length usually evaluated when $\xi = 1$. This description is however an over simplification as the physics of structure formation is richer in complexity than a simple power law.

One great advantage of using this estimator, is that the Fourier Transform of $\xi(r)$ gives the Power Spectra, $\mathcal{P}(k)$, of density anisotropies (see §3.1.4). It is also a statistic which is very easy to visualise and very easy to compute. From $\xi(r)$ it is also possible to determine the correlation dimension, D_2, of the discrete point distribution. D_2 is calculated via,

$$D_2 = 3 + \frac{d[\log \xi(r)]}{d[\log(r)]}. \tag{3.4}$$

For a homogeneous distribution in 3-D we would expect $D_2 \rightarrow 3$. However, at scales around r_0, $\gamma \sim 1.77$ (Davis-Peebles, 1983), corresponding to a dimensional value, $D_2 \sim 1.23$. We will soon see where expression (3.4) comes from when we discuss generalised dimensions in §3.4.

3.1.1 ξ-Correlation Estimators

There are a few different ways to measure $\xi(r)$, in the literature but the basic computational structure is more or less the same. The main differences between estimators are usually the way in which they deal with 'edge effects' and the 'shot noise'. We are obviously looking for correlations between galaxies so we begin by

centering on a galaxy, and proceed to count the number of galaxies within spherical shells of different radii around the central object. This procedure is repeated by centering on all, or a randomly chosen subset of all, the galaxies in the catalogue to obtain a statistical average. The number of galaxies counted in each cell is then normalised by a Poisson term which is related to the mean density. This estimator is written as,

$$\xi_j(r) = \frac{dP_c}{\bar{n}dV} - 1 = \frac{DD}{DR} - 1 \tag{3.5}$$

$$= \sum_i^N \frac{\Theta(|r_j - r_i| - r).\Theta(r + dr - |r_j - r_i|)}{\bar{n}(r).dV(r).\phi(r_i).\phi(r_j)} - 1, \tag{3.6}$$

where DD are data-data pairs, DR are data-random pairs and N is the number of galaxies. ϕ is the selection function and

$$\Theta(x) = \begin{cases} 0, & x < 0, \\ 1, & x > 0, \end{cases} . \tag{3.7}$$

The problem with eq.(3.6) lies within the mean density term. To obtain $\bar{n}$, the most accurate method should be to sum over all the galaxies while weighting each by the inverse of the selection function. This is calculated as follows,

$$\bar{n} = \frac{1}{V} \sum_i^N \phi(r_i)^{-1}. \tag{3.8}$$

In the last equation, V is the volume and the sum is over all galaxies. To reduce the variance associated with $\xi(r)$, the average can be taken,

$$\langle \xi(r) \rangle = \frac{1}{N} \sum_j^N \xi_j(r). \tag{3.9}$$

The angled brackets represent a statistical average, which usually invokes the use of the *cosmological ergodic theorem* (see §3.1.2). The result of applying eq.(3.6) to a mock galaxy catalogue can be seen in figure 3.1. The red line is a straight line

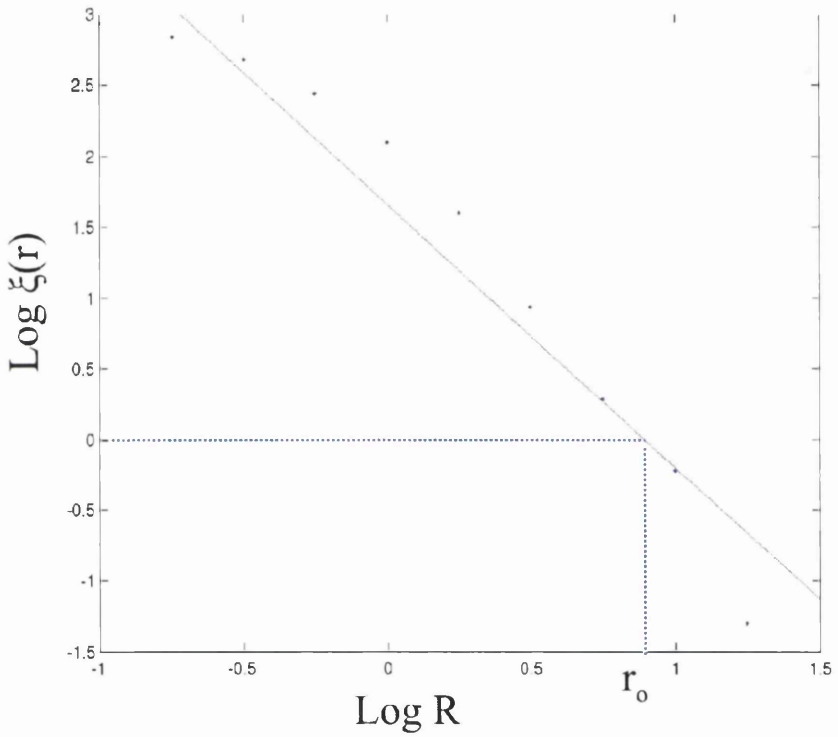

Figure 3.1: The 2PCF measured using (3.6) and applied to the PSCz mock galaxy catalogue. The gradient gives $\gamma \approx 1.6$ and therefore $D_2 \approx 1.4$ on scales up to 30Mpc

fit to the data, giving $\gamma \approx 1.6$.

So far we have used DD and DR, however RR pairs are also useful in calculating the 2PCF. Other estimators like the minimum variance estimator, by Landy & Szalay (1993), use the RR pairs to help correct for boundary effects. Their estimator takes the form,

$$\xi(r) = \frac{M(M-1)}{N(N-1)}\frac{DD}{RR} - \frac{(M-1)}{N}\frac{DR}{RR} + 1, \tag{3.10}$$

where the M is the number of random points. There are many more estimators for the 2PCF. Kerscher, Szapudi, & Szalay (2000) show that equation (3.10) is strongly preferred over other methods.

3.1.2 Ergodicity

A short note on the Ergodic Principle:
The observed universe is unique. This implies that averages have to be spatial ones. Such averages will be equal to those obtained if instead we were to average over an ensemble of universes if the *Cosmic Ergodic Theorem* holds. *Ergodicity* in the cosmological context means that ensemble averaging and spatial averaging are equivalent. Note that, in contrast with the common practice in statistical mechanics, the cosmological Ergodic Hypothesis refers to the spatial distribution of a random field at a fixed time rather than to the time evolution of the system. Thus, for instance, the ensemble average of the random field $\delta(\mathbf{x})$ at a point $\mathbf{x}$, $\langle\delta(\mathbf{x})\rangle$, is simply the expectation value of the random variable $\delta(\mathbf{x})$.

3.1.3 Gaussianity

Let us introduce some of the statistics used by cosmologists to characterise the spatial distribution of matter. We define *r.m.s.* σ_δ fluctuations of a continuous density field $\delta(\mathbf{x})$ as

$$\sigma_\delta^2 \equiv \left\langle \delta(\mathbf{x})^2 \right\rangle, \tag{3.11}$$

and the *correlation function* by

$$\xi(r_{12}) \equiv \left\langle \delta(\mathbf{x_1})\delta(\mathbf{x_2}) \right\rangle. \tag{3.12}$$

(Note that for a homogeneous and isotropic random process ξ only depends on the distance between the two points $r_{12} = |\mathbf{x}_1 - \mathbf{x}_2|$.) The correlation function is a measure of the spatial correlation of the field $\delta(\mathbf{x})$.

A random field is said to be *Gaussian* if all N-point multivariate probability distribution functions are multivariate Gaussian distributions defined by their mean vector $\langle \delta(\mathbf{x}_i) \rangle$ (which the ergodicity implies to be identically zero) and their covariance matrix $M_{ij} = \xi(\mathbf{x}_i, \mathbf{x}_j)$. Gaussianity is a very popular assumption for two reasons. The first one is that the calculations are "easy" to perform. The second reason is that the CMB seems to support a Gaussian initial density field, Komatsu *et al.* 2003.

3.1.4 Power spectrum

If we expand the $\delta(\mathbf{x})$ field in plane waves as

$$\delta(\mathbf{x}) = \frac{V_u}{(2\pi)^3} \int e^{i\mathbf{k}\cdot\mathbf{x}} \delta_\mathbf{k} d^3k, \tag{3.13}$$

we see that its Fourier transform $\delta_{\mathbf{k}}$ is given by

$$\delta_{\mathbf{k}} = \frac{1}{V_u} \int e^{-i\mathbf{k}\cdot\mathbf{x}} \delta(\mathbf{x}) d^3x, \tag{3.14}$$

where V_u may be thought of a "fair sample" of the universe. The *power spectrum* of the density field $\delta(\mathbf{x})$ is defined as the expectation of the two-point function in Fourier space, as follows:

$$\left\langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \right\rangle \equiv P(k_1) \delta_D(\mathbf{k}_1 - \mathbf{k}_2) \tag{3.15}$$

where δ_D is the well-known Dirac delta function. This implies that even if $\delta_{\mathbf{k}}$ is not a Gaussian distribution, the random variable $\delta(\mathbf{x})$, being an infinite sum of independent random variables, will still be Gaussian by the Central Limit Theorem for some well-behaved power spectra. We can see that the Dirac function in eq. (3.15) is required because of the translational invariance, $\langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \rangle = \xi(|\mathbf{x}_1 - \mathbf{x}_2|)$. Similarly, we can also see that isotropy implies that $P(k)$ depends only the magnitude of the wave-vector $\mathbf{k}$.

3.1.5 Window Functions

For some calculations it may be necessary to apply a cutoff at high spatial frequencies, this is due to non-linearities on small scales. The smoothed field $\tilde{\delta}(\mathbf{x})$ that may be obtained by convolution of the "raw" field with some weighting function W (called *window function*) having a characteristic scale r_W is given by

$$\tilde{\delta}(\mathbf{x}) \equiv \int \delta(\mathbf{x}') W(\mathbf{x}' - \mathbf{x}, r_W) d^3x', \tag{3.16}$$

has *r.m.s.* $\sigma_{\tilde{\delta}}$ fluctuations given by

$$\sigma_{\tilde{\delta}}^2 = \langle|\tilde{\delta}(\mathbf{x})|^2\rangle = \frac{1}{(2\pi)^3}\int |W(\mathbf{k})|^2 P(k) d^3k. \tag{3.17}$$

Where $W(\mathbf{k})$ is the representation in Fourier space of $W(\mathbf{y}, r_W)$. The window function has the following properties: $W(\mathbf{x}' - \mathbf{x}, r_W) = const. \simeq r_W^{-3}$ if $|\mathbf{x} - \mathbf{x}'| \ll r_W$, $W(\mathbf{x}'-\mathbf{x}, r_W) = 0$ if $|\mathbf{x}-\mathbf{x}'| \gg r_W$, satisfying the relation $\int W(\mathbf{x}'-\mathbf{x}, r_W)d\mathbf{y} = 1$. One of the most common window functions is the "*top hat*" (TH) window function which is defined by the relation

$$W_{TH}(|\mathbf{x} - \mathbf{x}'|; r_{TH}) = \frac{3}{4\pi r_{TH}^3} H(1 - \frac{|\mathbf{x} - \mathbf{x}'|}{r_{TH}}), \tag{3.18}$$

where H denotes the Heaviside step function ($H(y) = 0$ if $y \leq 0$, and $H(y) = 1$ if $y > 0$). Another commonly used window function is the *Gaussian* kernel:

$$W_G(|\mathbf{x} - \mathbf{x}'|; r_G) = \frac{1}{(2\pi r_G^2)^{2/3}} \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^3}{2r_G^2}\right). \tag{3.19}$$

3.2 Higher Order Correlations

It is a natural question then to ask whether there are higher order correlations than the simple 2PCF, generally defined as, $\xi_2(r)$. The answer is most definitely YES. A Gaussian random field would in principle be completely defined by the 2PCF (the initial density field is thought to have this property), however due to non-linear structure formation we now have local non-gaussianity which means we must look to higher order moments to accurately quantify the galaxy distribution.

These higher order correlations are defined as, $\xi_n(r_1, ..., r_n)$, where n is the order of the correlation function. As an example the 3 point correlation function is defined as the joint probability of there being a galaxy in volume elements dV_1

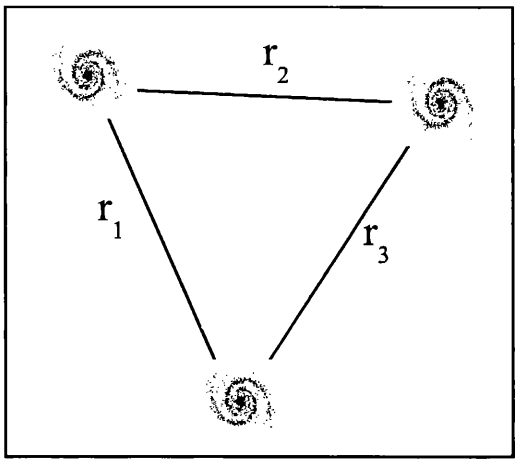

Figure 3.2: The distances $\mathbf{r}_1$ $\mathbf{r}_2$ and $\mathbf{r}_3$ separated the three galaxies. The triangular configuration can be fixed at the outset or it could be included as a variable at the cost of much more computation.

and dV_2 given that these elements have displacements $\mathbf{r}_1$ and $\mathbf{r}_2$ from the galaxy which is being investigated. This is illustrated in fig.(3.2). The joint probability can be written as,

$$dP = n^3[1 + \xi(r_a) + \xi(r_b) + \xi(r_c) + \zeta(r_a, r_b, r_c)]dV_1 dV_2 dV_3, \tag{3.20}$$

where r_a,r_b and r_c are the sides of the triangle. Assuming homogeneity and isotropy means leads to ζ being a symmetric function of the three lengths. Equation (3.20) is the full three-point correlation function, and ζ is known as the reduced part.

The conditional probability of finding two objects to complete the triangular configuration given that we are centring on a galaxy is,

$$dP = n^2[1 + \xi(r_a) + \xi(r_b) + \xi(r_c) + \zeta(r_a, r_b, r_c)]dV_2 dV_3. \tag{3.21}$$

Then the conditional probability of finding a galaxy to complete the triangle, given

that we have a pair of galaxies with separation r_a, is,

$$dP = n\frac{[1+\xi(r_a)+\xi(r_b)+\xi(r_c)+\zeta(r_a,r_b,r_c)]}{1+\xi(r_a)}dV_3, \tag{3.22}$$

In extending to N-point correlation functions the computational load is increased exponentially. This limits our ability to determine accurately $N > 4$, however there may be a way to avoid this problem by making some approximations.

Given that $\xi(r)$ can be represented by a power law,

$$\xi(r) = Br^{-\gamma}, \qquad \gamma \simeq 1.77, \tag{3.23}$$

the 3PCF is then found to be well described by a combination of ξ's (Peebles, 1980), *i.e.*,

$$\zeta(r_a,r_b,r_c) = Q[\xi(r_a)\xi(r_b)+\xi(r_b)\xi(r_c)+\xi(r_a)\xi(r_c)], \tag{3.24}$$

with $Q \simeq 1.0 \pm 0.2$ (Meiksin, Szapudi & Szalay, 1992).

3.3 Minimal Spanning Trees

One draw back with the 2PCF, as we have defined it in eq.(3.1), is that it is insensitive to filamentary structure. This is due to it being a function only of distance and not direction; thus all angular information is lost through the averaging. The Universe does appear to contain filaments, walls and other such features, but whether these are real or due to chance alignments has of course been a topic of debate over recent decades.

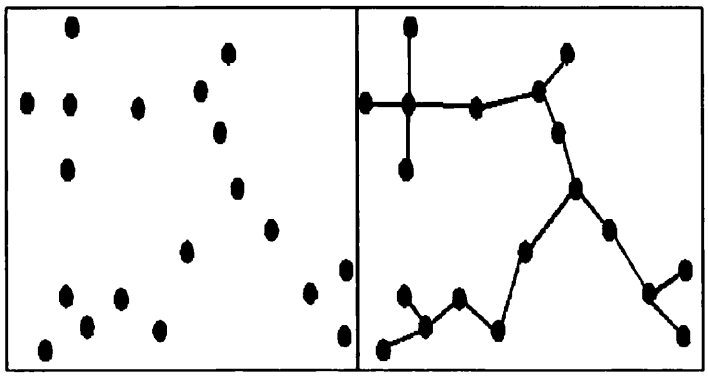

Figure 3.3: On the left hand panel we see a distribution of points. As we apply the method from 1-5 the Minimal Spanning Tree is constructed on the right hand panel.

To quantify this kind of structure Barrow, Bhavsar and Sonoda (1985) introduced a method from Graph Theory, called Minimal Spanning Trees (MST). The procedure to implement this is as follows:

1 A galaxy A is chosen as a starting point within a 3-D galaxy distribution.

2 The nearest galaxy to A is labelled B and a straight line is drawn to connect the two. This line is known as a **path**.

3 The closest galaxy to the set of previous galaxies (in this case A and B) is added to the set and is connected to the closest galaxy in the set. This produces a branching behaviour, and this step is repeated.

4 After some time we are left with many paths and **circuits** (closed paths). If there are no circuits the graph is open and this is known as a **tree**.

5 To then transform this abstract visualisation into a numerical representation of structure we can do a few things, e.g.:

   i Calculate the number of lines in angular bins of $d\theta$ w.r.t an adjoining line.

   ii Calculate the number of lines in distance bins, dr. Obtaining dN vs dr.

This method connects all galaxies in some fashion. But we know that not all galaxies are physically connected, as a galaxy in one cluster has little to do with another galaxy in a distant cluster.

To account for this, we adjust the previous method by only joining a galaxy to a pre-existing tree if its distance to the closest member is less than some threshold distance. This technique is known as separating and was introduced by Clark & Miller (1966). This method was recently applied to the SDSS DR1 by Doroshkevich *et al.*(2004). They found that groups and clusters are more likely to be found close to walls rather than filamentary structure.

3.4 The Fractal Universe

Fractal patterns can be thought of as the place where chaos and order meet. This is because self-similarity (fractality) seems to be an eventual by-product of chaotic systems. Fractal pictures are usually associated with "Julia" and "Mandelbrot" sets named after the French mathematicians.

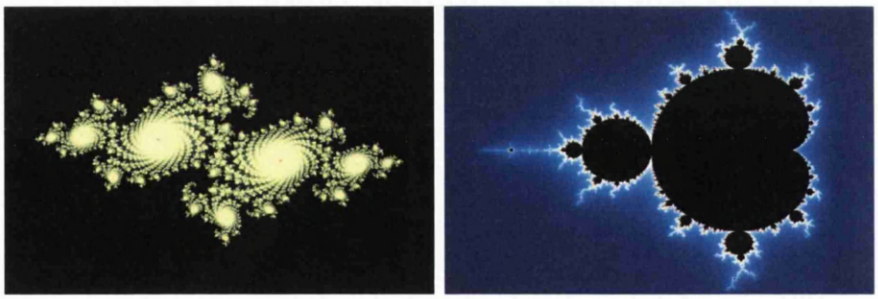

Figure 3.4: *Left:* The Julia set. *Right:* The Mandelbrot set. In both the Julia & the Mandelbrot sets the self-similarity is clearly apparent *i.e.* successive enlargements of areas will show similar patterns to the picture as a whole.

Fractals belong to a branch of mathematics known as Fractal Geometry. Unlike the usual definition of Geometry, where regular shapes and patterns are studied,

fractals can be highly irregular and the dimensions which are explored are not confined to integer values like $D = 2$ for surfaces and $D = 3$ for volumes. The term fractal was originally coined by Mandelbrot himself, who is widely considered the father of fractal studies. In fig.(3.4) the Mandelbrot and Julia sets are shown. These are classical examples of fractal patterns, we can easily see that zooming in on certain areas will yield patterns which are reproductions of the whole picture. It is this self-similarity in fractal patterns which make them scale-invariant objects. Multifractals on the other hand are not scale invariant, since their spatial dimension can vary with scale length. In these terms monofractals can be thought of as a special case of a more general multifractal.

Fractals were merged with the physical sciences through their intimate connection to nonlinear physics and chaotic dynamics. In configuration space we may have chaotic motion for an unstable system, however, in phase space dynamics may be more ordered and the system may evolve towards an attractor. This is rather like a chaotic analogy to an equilibrium state. Phase space trajectories are, however, by no means an easy-to-use tool for analysing structures since, in cosmology, we only have access to spatial coordinates[1]. Nevertheless, this kind of ordered phase space gives rise to a self similar pattern in real space. This is, more specifically, a fractal pattern.

Now it is known that on large scales, the main ingredient to structure formation is the force of gravity. However the $1/r$ potential for gravity leads to highly nonlinear motions and also to cross talking between different spatial scales. So it is not so great a jump to consider the galaxy distribution as a fractal of some kind. In fact this line of analysis is not a new one. The distribution of galaxies in the universe has already been shown to be well described using a multifractal framework, see *e.g.* Jones *et al.* 1988.

[1]This is not entirely true as we do have limited velocity information as well.

3.4.1 Multifractal Formalism

In this analysis we will adopt the procedure layed out in Henschel *et al.* (1983) to determine the Rényi (Generalised) dimensions of a point set embedded in a three-dimensional Euclidean space. The probability of a galaxy, j, being within a sphere of radius r centred on galaxy i is,

$$\begin{aligned} p_i(r) &= \frac{n_i(r)}{N}, \\ &= \frac{1}{N}\sum_{j=1}^{N}\Phi(|r_i - r_j| - r). \end{aligned} \tag{3.25}$$

Here $n_i(r)$ is the number of galaxies within radius r, N is the total number of galaxies and

$$\Phi(x) = \begin{cases} 1, & x < 0, \\ 0, & x > 0, \end{cases} \tag{3.26}$$

Equation (3.25) can then be related to the partition sum via Grassberger and Procaccia's (1983) correlation algorithm,

$$Z(q,r) = \frac{1}{M}\sum_{i=1}^{M}[p_i(r)]^{q-1} \propto r^{\tau(q)}. \tag{3.27}$$

In this case M is the number of counting spheres and q defines the generalised dimension we are investigating. $\tau(q)$ is the scaling exponent, which is then related to the infinite set of dimensions through,

$$D_q = \frac{\tau(q)}{q-1}, \qquad q \neq 1. \tag{3.28}$$

Clearly the special case of $q = 1$, the information dimension, cannot be determined using the above expression but can be found approximately in the limit $q \to 1$. This is an important dimension to calculate as it gives equal weighting to voids and clusters. Voids are enhanced for $q < 1$ and clusters are enhanced for $q > 1$,

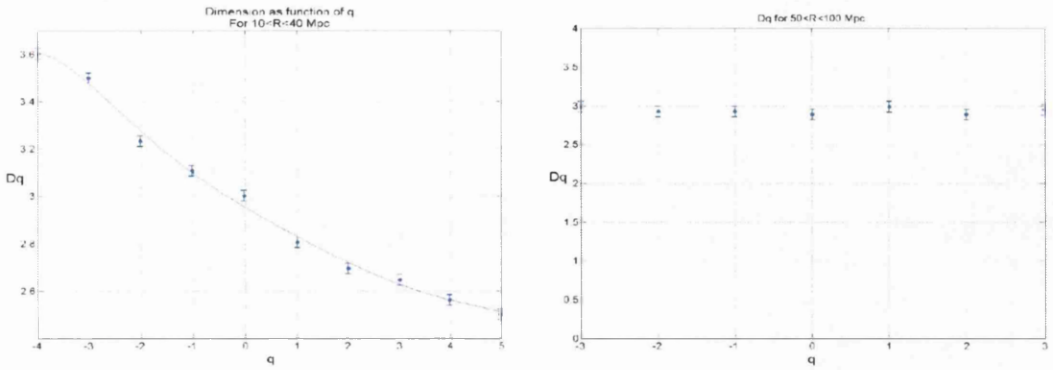

Figure 3.5: Applying the multifractal analysis to ΛCDM halo catalogue, we obtain the usual D_q curve. A linear χ^2 fit to the partition sum, $Z(q,r)$, over the two distance scales produces each point in the plots above. The catalogue has approximately 1.5 million halo positions.

so $q = 1$ is in this sense the most unbiased dimension in the set. To determine D_1 more accurately we must calculate,

$$S(r) = \frac{1}{M}\sum_{i=1}^{M} \log\left[p_i(r)\right] \propto r^{D_1}, \tag{3.29}$$

where $S(r)$ is the *partition entropy* of the point set.

In §3.1 the 2PCF was related to the correlation dimension, D_2, through a differential operation. However, in Multifractal terms the correlation dimension is only one of an infinite number of generalised dimensions which we can use to characterise the distribution. Other important dimensions include, D_1 - *information* dimension, D_0 - *Capacity* dimension, and D_q as $q \to \pm\infty$. Multifractal distributions are usually defined by a D_q curve which generally decreases with q (see figure 3.5). Monofractals on the other hand produce a flat D_q curve.

The methodology we have constructed is used to calculate the D_q curve for the ΛCDM N-body simulations. The two plots in figure 3.5 show D_q curves for different distance scales. A minimum χ^2 approach was applied to the partition sum, $Z(q,r)$ as illustrated in figure 3.6(a).

The χ^2 minimisation was performed on a straight line model fit to the data at two different distance scales; $10 < R_1 < 40$ Mpc and $50 < R_2 < 100$ Mpc. Both R_1 and R_2 were chosen as they appeared to have constant gradient in these regions. At small scales the data seems to be supporting a multifractal distribution, whereas on large scales the universe appears to reach homogeneity on scales considerably smaller than the size of the box.

3.4.2 Other Estimators

The partition sum can also be calculated using a counts-in-cells approach (Mandelbrot, 1982). Define a new measure μ of a discrete point set as,

$$\mu_i = \frac{N_i}{N_{tot}}, \tag{3.30}$$

which is just the fraction of all galaxies N_{tot} contained within cell i. Also if the relation $\sum_i \mu_i = 1$ is satisfied, then it must follow that the cells cover the entire space (of topological dimension, d) and that the cubes have volume r^d. We can now construct a new measure,

$$M(q,r) = \sum_{i=1}^{N(r)} \mu_i^q . r^a = N(q,r).r^a \underset{r \to 0}{\longrightarrow} \begin{cases} 0, & a > \tau(q), \\ \infty, & a < \tau(q), \end{cases} \tag{3.31}$$

with $N(r)$ being the number of occupied cells and $N(q,r)$ is the number of occupied cells, weighted by q. The measure is dominated, for large positive values of q when the cells are more populated, and for large negative q when the cells are sparsely occupied. This is a very important property of multifractal analysis, since under-dense and over-dense regions are probed by different values of q.

In the limit $a \to \tau(q)$, it follows that M will be finite and nonzero. Then we

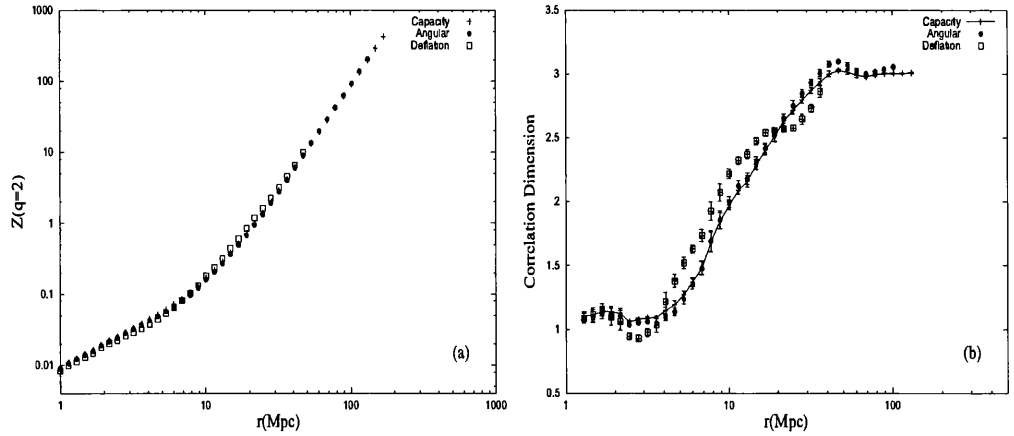

Figure 3.6: Results of the mock *PSCz* catalogue from a ΛCDM model. *Left:* The partition sum Z_2 varying with distance. This is very closely related to the 2PCF. The direct derivative of this plot can determine the correlation dimension varying with scale. [Figures from Pan & Coles, 2001]

can see from eq.(3.31) that,

$$N(q,r) = \sum_{i=1}^{N(r)} \mu_i^q \propto r^{-\tau(q)}, \tag{3.32}$$

and

$$\tau(q) = \lim_{r\to 0} \frac{\ln N(q,r)}{\ln(1/r)}. \tag{3.33}$$

3.4.3 The $f(\alpha)$ Curve

Grassberger *et al.* (1988) show that we can rewrite the usual fractal measure as,

$$p_i \sim r^{\alpha_i} \tag{3.34}$$

The distribution of the scaling indices α_i characterise the dimensionality of the survey. This is evaluated using the α-spectrum,

$$n(\alpha)d\alpha \sim N|\ln r|^{1/2} r^{\alpha - f(\alpha)d\alpha}, \tag{3.35}$$

where $n(\alpha)d\alpha$ is the number of times that α takes values in the range $(\alpha, \alpha + d\alpha)$. For a homogeneous fractal distribution the $f(\alpha)$ curve reduces to a single point: $\alpha_0 = f(\alpha_0) = D_0$. In any case the statistical properties of a distribution are equally described by either the generalised dimensions, D_q or by the $f(\alpha)$ curve since they are a Legendre pair, as shown below. The only main drawback, as we will soon see, is that the latter strategy requires an extra differentiation of the data.

We can show this by considering the integral version of equation (3.27),

$$Z(q,r) = \frac{1}{N}\sum_{i=1}^{N} p_i(r)^{q-1} = \frac{1}{N}\int n(\alpha) r^{(q-1)\alpha} d\alpha,$$
$$= \int |\ln r|^{1/2} r^{\alpha q - f(\alpha)} d\alpha. \tag{3.36}$$

A solution to the above expression can be found using the Laplace integral approximation (Martinez *et al.* 1990), giving,

$$Z(q,r) = r^{\alpha(q) - f\{\alpha(q)\}} \left\{ \frac{\pi}{2f''[\alpha(q)]} \right\}^{1/2}. \tag{3.37}$$

The conditions of this theorem defining the function $\alpha(q)$ are,

$$\frac{d\, f(\alpha')}{d\alpha'}\Big|_{\alpha' = \alpha(q)}, \tag{3.38}$$

and

$$\frac{d^2 f(\alpha')}{d\alpha'} < 0. \tag{3.39}$$

Using $Z(q,r) = const \times r^{\tau(q)}$ and (3.37) we get,

$$\tau(q) = \alpha(q)q - f(\alpha), \tag{3.40}$$

and using (3.38) with (3.39) leads to,

$$\alpha(q) = \frac{d\tau}{dq}. \tag{3.41}$$

Equations (3.40) and (3.41) relate the variable pairs, (q, τ) and (α, f): a Legendre transform. So we can see that the distribution is equivalently characterised using either method. However, in practical terms the Generalised Dimension approach may prove to be more accurate, since the $f(\alpha)$ curves require a further differentiation of the data. i.e. $f(\alpha)$ curves (fig.3.7) are related to the derivative of the D_q's (fig.3.5), through eq.(3.41)

In figure 3.7, two realisations of a multiplicative random fractal are shown with their corresponding $f(\alpha)$ curves. The method for constructing these distributions is discussed in Jones *et al.* (1988).

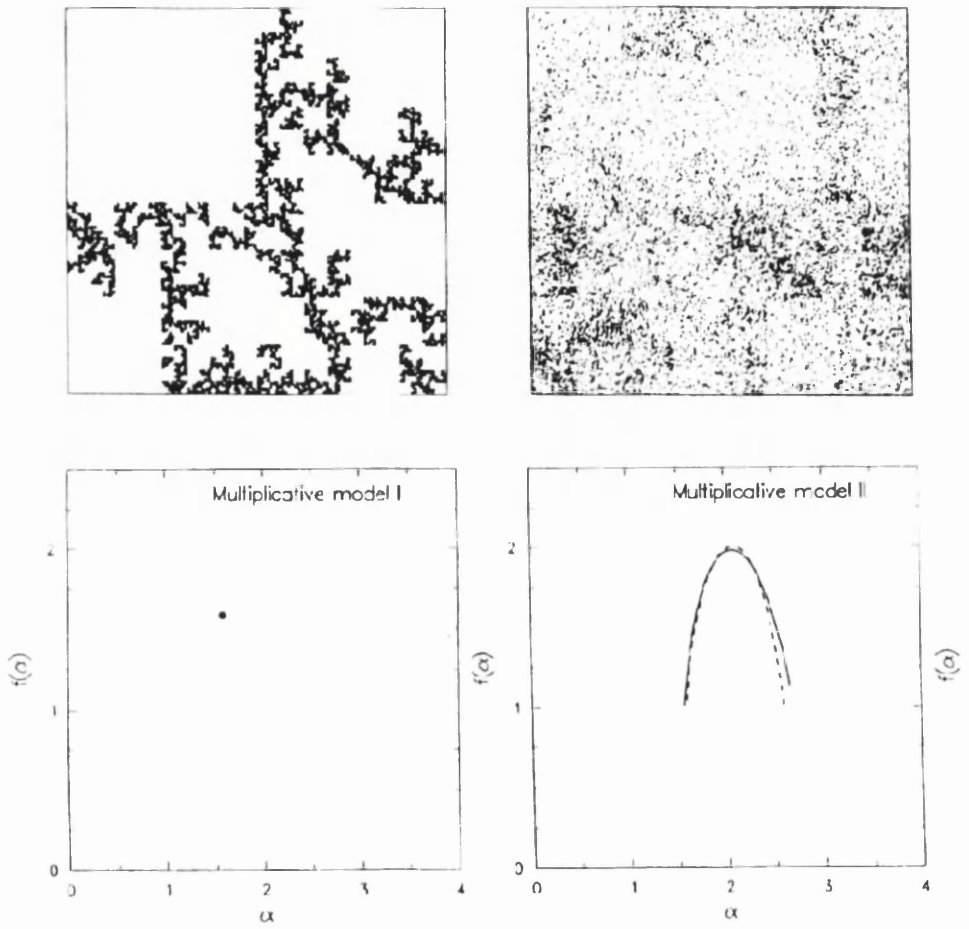

Figure 3.7: Two fractal distributions with their corresponding $f(\alpha)$ curves below. *left:* A monofractal distribution produces a single point in the $f(\alpha) - \alpha$ plane. *right:* A genuine multifractal point set produces a curve in the $f(\alpha) - \alpha$ plane. [Figures from Jones *et al.* 1988]

Chapter 4

Boundary Corrections

As we touched on in §1.3.1, the analysis of redshift surveys is prone to many problems, some subtle and some not so subtle. The presence of large 'holes' in a survey – regions where no galaxies have been observed – is clearly an example of a serious problem which needs to be appropriately corrected for.

To analyse real (or indeed mock) galaxy surveys we must, therefore, deal with the practical issue of incomplete sky coverage. This can arise firstly because of the geometry of the survey, which is usually a thin beam or a fan. Figure 4.1 illustrates the latter case for the recent example of the 2dF galaxy redshift survey (2dFGRS, Colless *et al.* 1999). The figure shows the projected distribution of galaxies on the plane of the sky, from which we see that the coverage of the survey is not all-sky – *i.e.* we do not sample galaxies over 4π steradians. Also, we notice that there are small patches and strips within the geometrical area of the survey which were not sampled.

Secondly, incomplete sky coverage can be caused by the extinction of light through parts of our own galaxy, or regions being obscured by local objects. Hence, for example, even redshift surveys such as the IRAS PSCz (Saunders *et al.* 2000) which set out to be all-sky are missing galaxies within a 'mask' close to the plane

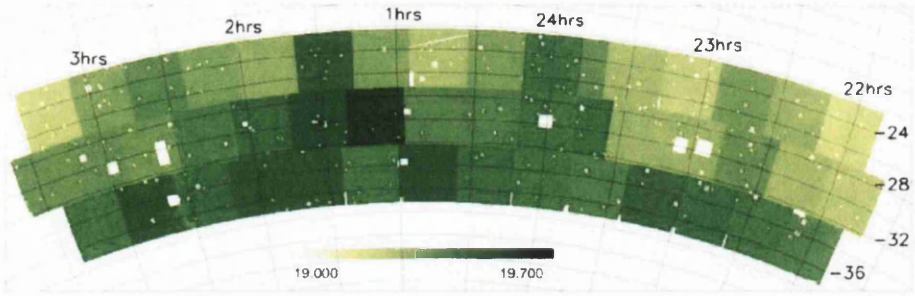

Figure 4.1: The 2dFGRS southern region shown in projection. The colour scale gives the varying magnitude limit from plate to plate, white denotes regions which lie outside the survey. [Figure from the website: 2dF www.mso.anu.edu.au/2dFGRS]

of the Milky Way galaxy.

Thirdly, of course, as we already discussed in §1.3.1, redshift surveys are affected by radial incompleteness, which we describe in terms of a selection function, caused by the flux limit below which distance galaxies are too faint to be observed. We can see the influence of this flux limit in Figure 4.1, which shows the variation in magnitude limit from plate to plate in the 2dFGRS.

So, in summary, information about the 'true' population of galaxies is hidden or distorted by the effect of a flux limit, by the presence of masked regions and by the boundary of the survey itself.

There has been a number of methods created to account for the problems mentioned above. In the following sections a few of these methods will be reviewed and in §4.6 a new correction technique is introduced, whose benefits include computational and statistical efficiency.

4.1 Deflation Method

As we consider placing spherical shells of increasing radius around a galaxy, even allowing for our 'weighting' of the number count of galaxies in each sphere by the radial selection function of the survey, it is clear that we will eventually reach the edge of the survey (which of course we can think of as the distance beyond which the selection function is equal to zero). Therefore, unless some form of edge correction is applied, further increases in the shell radius will result in an estimate of the density for shells that are systematically underpopulated relative to the mean density of the underlying galaxy population, because of the volume of each shell that lies outside the survey. This effect is illustrated in the upper panel of Figure 4.2. For the outer (shaded) shell in this diagram, the estimated density will be systematically lower than the true density since the shell includes a region that lies entirely outside the survey volume and so, by definition, will contain no galaxies.

The *deflation* method is, perhaps, the simplest and crudest form of boundary correction. It simply restricts the sum in equation (3.25) to include only those counting spheres which lie completely within the survey region. However, this drastically reduces the distance out to which the density estimator can reliably probe, leading effectively to a 'cosmic variance'[1] problem at larger radii.

The maximum scale, R_{surv} which is probed using the *deflation* method is defined as (Hatton, 1999),

$$R_{surv} = \frac{d \sin \theta_{surv}}{1 + \sin \theta_{surv}}, \tag{4.1}$$

where θ_{surv} is the opening angle of the survey and d the volume limit.

As an example, the Stromlo-APM redshift survey (Loveday *et al.* 1992) had an

[1] Cosmic variance is the cosmologist's definition of sample variance. If the scales being sampled are comparable in size to the survey, then only a few independent measurements can be taken, *i.e.* only a few values available for averaging.

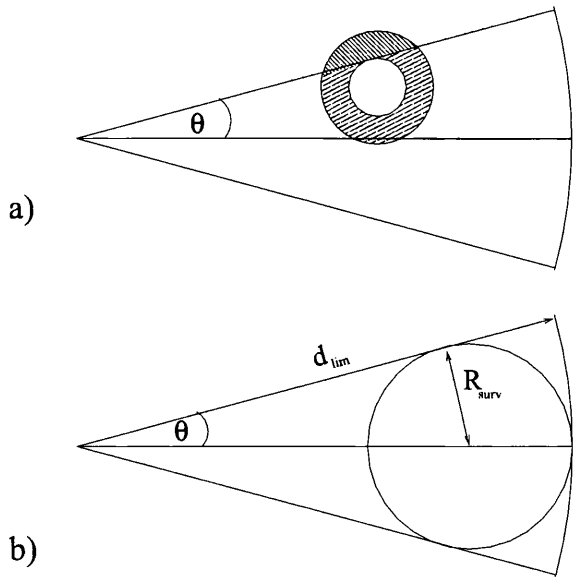

Figure 4.2: Panel (a) illustrates the state of concentric shells around a galaxy in the survey. As the radius of the shell increases, eventually density is measured for shells that are partially external to the survey, such as the shaded shell above. The lower panel (b) demonstrates the maximum distance to which this survey can be probed: the radius of the largest sphere which can totally be contained within the survey. [Figure from Hatton, 1999]

opening angle, $\theta_{surv} = 22.6^o$ and a volume limit, $d \sim 110h^{-1}$Mpc, this leads to a maximum sampling radius of only, $R_{surv} = 30.5h^{-1}$Mpc.

4.2 Capacity Correction

The *capacity* correction can be thought of as the next step up from the *deflation* method. Here we allow all counting spheres while using equation (3.25), even those which cross the boundary or contain masks. The missing volume is accounted for by re-weighting the contribution of the (incomplete) counting sphere, essentially equivalent to filling it with a distribution of mock galaxies. The main problem is in deciding which distribution should be used to fill the void. Borgani *et al.* (1994) chose to weight each cell by a factor $f_i(r)$ which is determined by the missing volume.

This at first glance may seem to be a valid choice. On the other hand the weighting factor should be, more correctly, proportional to some measure of the average density in the counting sphere. This highlights the potential problem with the capacity correction: even though it can, in principle, be applied to considerably larger counting spheres than the deflation correction, its form is fundamentally flawed since it is assuming an answer to the question being posed – *i.e.* that $N \sim r^3$, which is a statement of exactly the homogeneity that we are trying to test.

In practise the capacity correction is applied using the expression,

$$p_i(r) = \frac{n_i(r)}{N_{tot}} = \frac{n_i^*(r) + \bar{\rho}.V_{missing}}{N_{tot}}, \tag{4.2}$$

$$= \frac{1}{N_{tot}} \left\{ \sum_{j=1}^{N} \frac{\Phi(|r_j - r_i| - r)}{\phi(r_j)} + 4\pi.\bar{\rho}(r^2.\Delta r + r.\Delta r^2 + \Delta r^3).\frac{M}{M_{tot}} \right\}, \tag{4.3}$$

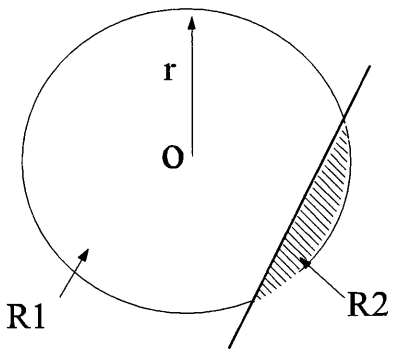

Figure 4.3: A counting sphere centred on galaxy O with radius r. Regions R1 & R2 are inside and outside the survey respectively.

to obtain the corrected number of objects within radius r from a given particle. The RHS of eq.(4.2) contains two main terms: the reduced number count, n_i, and a term accounting for galaxies which are missing. This is done by calculating the missing volume of the survey and filling it with a density corresponding to the average density of the survey.

To determine the missing volume of the sphere one can place random points within it, or equivalently one can shoot off vectors in random directions from the central point in the sphere, and count how many of these lie within the survey. This approach is known as a *Monte Carlo* technique. The expression for the missing volume, therefore, requires us to determine two numbers:

1. the number M of random vectors emanating from the central galaxy which fall outside the survey

2. the total number M_{tot} of random vectors emanating from the central galaxy.

Of course choosing to fill the missing portion of the counting cell with any presupposed density should be considered *bad science.* What we would ultimately

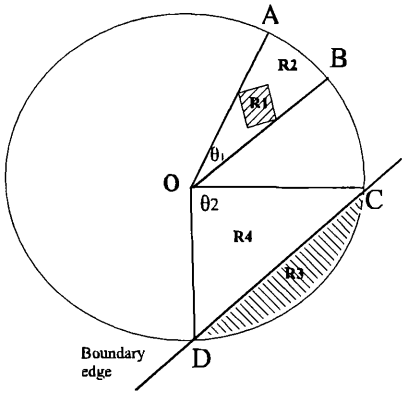

Figure 4.4: This is a counting cell within the survey, it has a masked region (R1) and a missing portion due to the intersection with the boundary (R3). The slices AOB and COD encompass both of these missing parts.

like to do is fill the missing regions with the right amount of particles. To do this we need information which is hidden from us, behind the mask, but the Cosmological Principle may come to our rescue.

4.3 Angular Correction

In 2002 Pan & Coles, used the the assumption of isotropy[2] to infer properties of unknown regions, masks and boundaries, from the well known survey space. Essentially they proposed that you can average over the part of the sphere which is observed and use this average to fill in for the missing regions. Of course this has to be done in an angular fashion, since $\xi(r)$ is assumed not to depend on direction.

Assuming that the universe is statistically the same in all directions, they concluded that the number of galaxies in a given solid angle should be comparable to the number of galaxies in the same solid angle but in a different direction. The

[2]The Isotropy of the universe is a cornerstone of modern cosmology so in this case it is not a bad assumption to make.

average density of galaxies per steradian can then be defined. In order to achieve this, they introduce the weighting factor $f_i(r)$ which has values $0 < f_i < 1$. The value of the weight determines how much of the counting sphere is missing from the survey, and therefore needs to be accounted for. Then the corrected number of galaxies in a cell is computed from,

$$n_i(r) = \frac{1}{f_i(r)} \sum_{j=1}^{N} \frac{\Phi(|r_j - r_i| - r)}{\phi(r_j)}, \tag{4.4}$$

where ϕ is the selection function for a flux limited sample. The weighting factor appears in the above expression to increase the number count accordingly, just as the selection function does.

To implement the angular correction one would start as usual and centre on a galaxy within the survey. At a given radius the counting sphere may contain a part of a masked region (indicated by region R1 in fig.4.4) or the survey boundary (R3 in fig.4.4). If this were to occur, the solid angles which contains these features are cut out (slice AOB & COD in fig.4.4) and replaced by an average over the rest of the cell. As an example, for figure 4.4 the weighting factor in this 2-D analogy is $f_0 = 1 - (\theta_1 + \theta_2)/2\pi$. For computational purposes, the method to calculate equation (4.4) is as follows:

1 From the set of all galaxies, choose a galaxy as the centre of a counting sphere of radius R and determine its position relative to the mask and boundary.

2 'Shoot off' random vectors (r, θ, ϕ) from the centre, where:

- r is sampled uniformly $\in (0, R)$. R - sphere radius.
- ϕ is sampled uniformly $\in (0, 2\pi)$.
- $\theta = \sin^{-1}(U)$, U is sampled uniformly from $\in (-1, 1)$.

3 Determine which vectors lie outside the survey region. Any vector which does so, save its components.

4 Using the scalar product expressed in terms of vector components, calculate the angle between each pair of vectors which lie inside the survey. This allows us to define a reference direction, and a maximum angle, θ_{max} between vectors which lie inside the survey. (θ_{max} is the opening angle of the cone which points in the reference direction).

5 Now begin counting those galaxies in the counting sphere which make an angle, $\theta > \theta_{max}$ relative to our reference direction.

6 The weighting factor, $f_i(r)$, is then simply related to θ_{max}.

This is essentially the algorithm developed by Pan & Coles (2002) which they tested on simulated catalogues and applied to analyse the fractal clustering of the IRAS PSCz survey.

The angular correction, although in principle a very successful method to correct for boundary effects, is very slow and inefficient. Moreover, and worse still, it throws away potentially useful data. We can see this from Figure 4.4, where the galaxies in regions R2 and R4 are excluded by the angular correction.

In the next section we consider a method which has the potential to improve further upon the angular correction.

4.4 Volume Correction

The basic idea of our new, volume, correction can be illustrated in figure 4.3. Here the counting sphere has exceeded the geometrical boundary of the survey. The number of galaxies counted in the sphere of radius r is depleted which leads to $p_i(r)$ being reduced through equation (3.25). As we have seen, to solve this problem we could either add galaxies to the missing region, as is the case with the *capacity* correction, or or equivalently we could somehow modify our definition of the volume itself (hence the name for our new correction!). Of course you may

notice that eq.(3.25) does not contain any explicit reference to the volume, but we can cast this equation as,

$$p_i(r) = \frac{V_i(r)\rho^\star(r)}{N_{tot}}, \tag{4.5}$$

$$= \frac{V_i(r)}{V_i^\star(r)} \cdot \frac{n_i^\star(r)}{N_{tot}}, \tag{4.6}$$

with V being the true volume of the sphere and $V^\star$ being what we can term the reduced volume. We have also introduced the reduced density, $\rho^\star(r)$, as an intermediary step which need not be calculated, and related this with a reduced volume and number count, $n_i^\star$. On its own this method can be visualised in figure (4.3), as assigning to the missing region (R2) the same density as that of region (R1). This would be wrong if density varies with distance, so that $\rho_{R1} \neq \rho_{R2}$. To overcome this problem we assume *only* that the density does not vary with θ or ϕ *i.e.* the universe is isotropic and hence equation (4.6) will hold for fixed r. So to apply this method to a galaxy survey we must count in spherical shells, correcting our estimate of the density in each shell as we go along, and then integrate up the shells at the end. This method is illustrated in figure (4.5). The shells are individually corrected and summed according to,

$$p_i(r) = \sum_{r=0}^{r} \alpha_i(r) \frac{n_i^\star}{N}, \tag{4.7}$$

where $\alpha_i(r) \equiv \frac{V}{V^\star}$; this is the enhancement factor of the i^{th} shell at radius r and has value ≥ 1.

The main advantage of this method is that makes the maximum use of the data. Specifically, if the boundary edge cuts across a counting sphere at a particular radius, the method still makes full use of galaxies at smaller radii from the centre of the counting sphere, even if they lie in the solid angle subtended by the boundary edge.

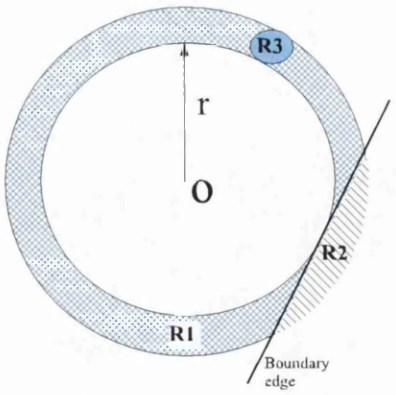

Figure 4.5: A counting shell centred on galaxy O with radius r. Region R1 is inside the survey, R2 & R3 are outside the survey. The missing parts of the shell (R2 & R3) are replaced by the average over the rest of the shell.

The deflation method and the angular correction, of Pan & Coles, on the other hand throw away a lot of potentially useful data, which limits the counting sphere radius within which the density may be reliably estimated.

4.5 Practical Computing Issues

One significant drawback faced when implementing the *angular* correction is that it must fill the counting spheres with random vectors to determine the missing regions. To obtain the required resolution means placing many random vectors at the centre of the sphere. For a cell with $R \approx 100\,h^{-1}$Mpc we have found that this requires the generation of $\sim 100,000$ random points.This is not a trivial computational task given that the vectors must be computed and angles stored at every step.

To see where this comes from consider figure 4.6. Here a simulation has been set up to mimic what happens when computing the angular correction. A counting

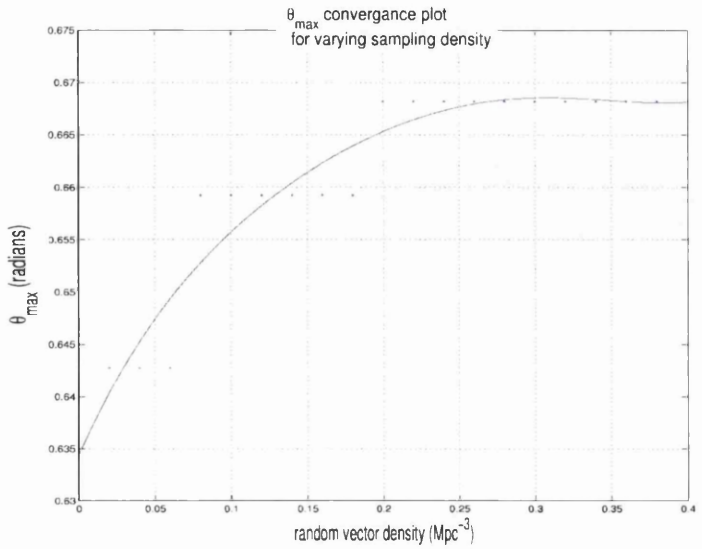

Figure 4.6: The convergence of θ_{max} for increasing vector density. The graph seems to converge at a random vector density of $\approx 0.25\mathrm{Mpc}^{-3}$.

sphere is placed close to the survey boundary and θ_{max} is calculated for varying numbers of random vectors. From the plot we can see the convergence of the opening angle at an approximate vector density of $\rho_{vec} \sim 0.25h^3\mathrm{Mpc}^{-3}$. Taking this value we can make a back of the envelope calculation of the required number of random vectors to accurately constrain θ_{max} at a typical scale of $100h^{-1}\mathrm{Mpc}$. The number of vectors required is then,

$$\begin{aligned} N(r) &= \rho_{vec}.V(r) \\ N(r) &= \rho_{vec}.\frac{4\pi}{3}.r^3 \\ N(r=100) &\approx 100,000 \end{aligned}$$

To make the comparison with the new volume correction, the same calculation is performed. In this case however we are calculating the number of vectors required

to populate a spherical shell with thickness $dr = 1$Mpc, as a typical value.

$$\begin{aligned} V(r) &\approx 4\pi r^2 dr \\ N(r) &= 4\pi\rho r^2 dr \\ N(r=100) &\approx 30,000 \end{aligned}$$

The volume correction is definitely more computationally efficient, especially when you consider that the missing volume must be calculated for every galaxy and at every distance iteration. One should also bear in mind however that these codes could be used on much bigger surveys, SDSS is now pushing 1 million galaxies.

4.6 Error Analysis

Error estimation in fractal analysis has been largely swept under the carpet by many in this field. One of the reasons for this is the computational costs of using, *e.g.*, a set of mock galaxy catalogues to estimate error estimates via a Monte Carlo approach. Nevertheless, we can obtain an approximate expression for the error on an estimate of the galaxy number density via some remarkably simple mathematics.

We follow Grassberger and Procaccia's (1983) suggestion and use the partition function $Z(q,r)$,

$$Z(q,r) = \lim_{r\to 0}\lim_{N\to\infty}\sum_{i}^{M} p_i^{q-1} \sim r^{\tau(q)}, \tag{4.8}$$

to estimate the generalised dimensions of a set of N galaxies. Here p_i is the probability,

$$p_i = \frac{n_i}{N}, \tag{4.9}$$

of a cell, centred on the i^{th} galaxy, having an occupation number n_i. r is the cell

size while $\tau(q)$ is a scaling exponent.

$$\langle Z(q)\rangle = \langle \sum_{i=1}^{M} p_i^{q-1} \rangle \tag{4.10}$$

By construction the mean of the partition function is positive as there are always cells with galaxies within. However, there is a subtlety when handling astronomical data: in expression (4.8) we assume that any set contains an infinite number of elements and the size of the cells vanish. Redshift surveys, however, contain a finite number of objects in a finite volume. This will lead to configurations where none of the cells have galaxies other than the central one when $\lim r \to 0$.

We can then construct the second moment of the distribution and rearrange as below,

$$\langle Z(q)Z(q')\rangle = \langle \sum_i p_i^{q-1} \sum_j p_j^{q'-1} \rangle \tag{4.11}$$

$$= \langle \sum_{i,j} p_i^{q-1}\, p_j^{q'-1} \rangle \tag{4.12}$$

$$= \sum_{i,j} \langle p_i^{q-1}\, p_j^{q'-1} \rangle \tag{4.13}$$

$$= \frac{1}{N^{q+q'-2}} \sum_{i,j} \langle n_i^{q-1}\, n_j^{q'-1} \rangle \tag{4.14}$$

$$= \frac{1}{N^{q+q'-2}} \sum_{i=1}^{M} \langle n_i^{q+q'-2} \rangle + \frac{2}{N^{q+q'-2}} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \langle n_i^{q-1}\, n_j^{q'-1} \rangle. \tag{4.15}$$

Where $\langle \cdots \rangle$ represents an ensemble average. Applying the *Cosmological Ergodic Theorem* to the above equation leads to sums over the cells. Thus, for $q = q'$ the

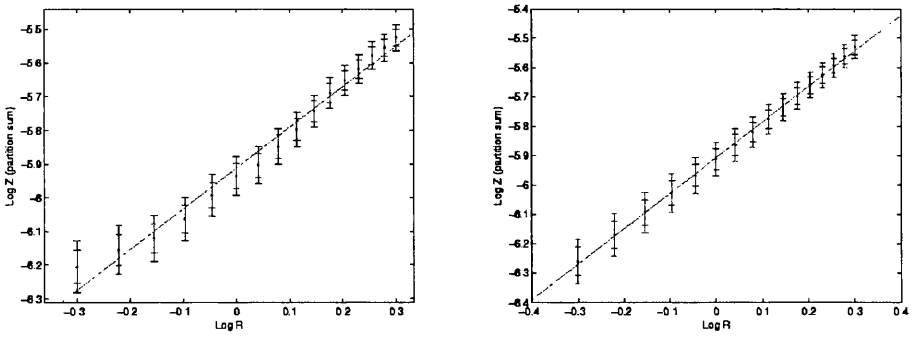

Figure 4.7: This is a log-log plot of the partition function varying with distance. The larger error bars are estimated from the prescription discussed in § 4.6, whereas the smaller errors are obtained from averaging the results of 100 Lévy Flight fractal simulations. This fractal has 1,000,000 particles with $D_2 = 1.2$ and is contained in a $400 Mpc^3$ box. *left:* The points in this plot have not been averaged. *right:* Points have been averaged over 100 distributions.

$\langle Z(q)Z(q')\rangle$ can be cast as

$$\langle Z(q)Z(q)\rangle = \sum_i \langle p_i^{2(q-1)}\rangle + \text{cross terms} \tag{4.16}$$

$$= \langle Z(2q-1)\rangle + \text{cross terms} \tag{4.17}$$

$$= \sum_i \langle \left(p_i^2\right)^{(q-1)}\rangle + \text{cross terms} \tag{4.18}$$

$$\leq 2 \sum_i \langle \left(p_i^2\right)^{(q-1)}\rangle. \tag{4.19}$$

We can then estimate the standard deviation directly from,

$$\sigma_q^2 \simeq 2.Z(2q-1) - Z(q)^2 \tag{4.20}$$

We show a comparison between our error estimation, from eq.(4.20), and an error estimation from 100 simulated Lévy Flight fractals (see figure 4.7).

The left hand plot, of figure (4.7), shows the partition function calculated as a function of distance. To each point is attached a larger error bar estimated from the prescription discussed in §4.6 and a smaller error bar obtained via an

average over the 100 simulations. We can see from the plot that the gradient of Z is constant and the points are all very close to the fitted straight line. The fit was obtained by minimising the χ^2 function. The right hand plot shows similar points but now the partition function calculated at each distance is also averaged over the 100 simulations, with error bars computed as before. The excellent agreement in the fitted slopes of the left and right hand plots indicates that the partition function is an unbiased estimator.

Chapter 5

Analysis & Results

We will begin in the proceeding section by applying the different corrections to a toy fractal model. This model can be used to compare and contrast the differing methods. Then the *volume* correction will be used to analyse in detail the distribution of particles produced from an N-body simulation.

5.1 The Lévy Flight

An initial test for the different corrections is to analyse a simple fractal distribution, the Lévy Flight. This fractal is very easy to construct and has an analytical determined dimension (see Meakin 1998).

The Lévy flight fractal is finding its way into many areas of physics due to its close connection to Brownian Random Motion. For example it has been used to explain Interstellar scintillation (Boldyrev & Gwinn 2003) and even modelling the financial market (Chowdhury & Stauffer 1999)
The Lévy Flight is constructed as follows:

1 A point A is chosen at random, maybe the origin, in Cartesian space.

2 A displacement is given to A by a vector (θ,ϕ,R) to give a second point B. The angular direction is uniformly sampled and the Probability of R exceeding a value r is given by,

$$P(R \geq \frac{r}{r_0}) = \begin{cases} (r/r_0)^{-D}, & r/r_0 > 1 \\ 1, & r/r_0 \leq 1 \end{cases}, \tag{5.1}$$

3 This procedure is repeated many times to 'fill' the 3-D space, which was restricted to a cubic box of side 400 Mpc.

In expression (5.1), D is the fractal dimension and r_0 is a characteristic scale length, both of which we can adjusted to produce different features. The resulting distribution is not so dissimilar to a true galaxy survey (*cf.* figure 5.1). r_0 can be related to the average inter-cluster separation.

5.1.1 Multifractal Analysis

A multifractal analysis, as described in § 3.4.1, is performed on a Lévy Flight distribution of particles (*cf.* fig 5.1). The boundary is corrected by considering each of the different methods from Ch.4. This setup should give a fair comparison of the different correction methods.

Since the correlation dimension is known analytically, the multifractal analysis will be restricted to the D_2 dimension. Now from the partition sum $Z(2,r)$, equation (3.28) will provide the D_2 value through differentiation. This differentiation was performed directly on the data by applying a linear fit to every three consecutive data points. The errors, as described in § 4.6, were considered when minimising the χ^2 function.

The results of this procedure are plotted in figure 5.2. There are a few points to note in this plot. Firstly, the Lévy Flight is highly anisotropic, which is a cause

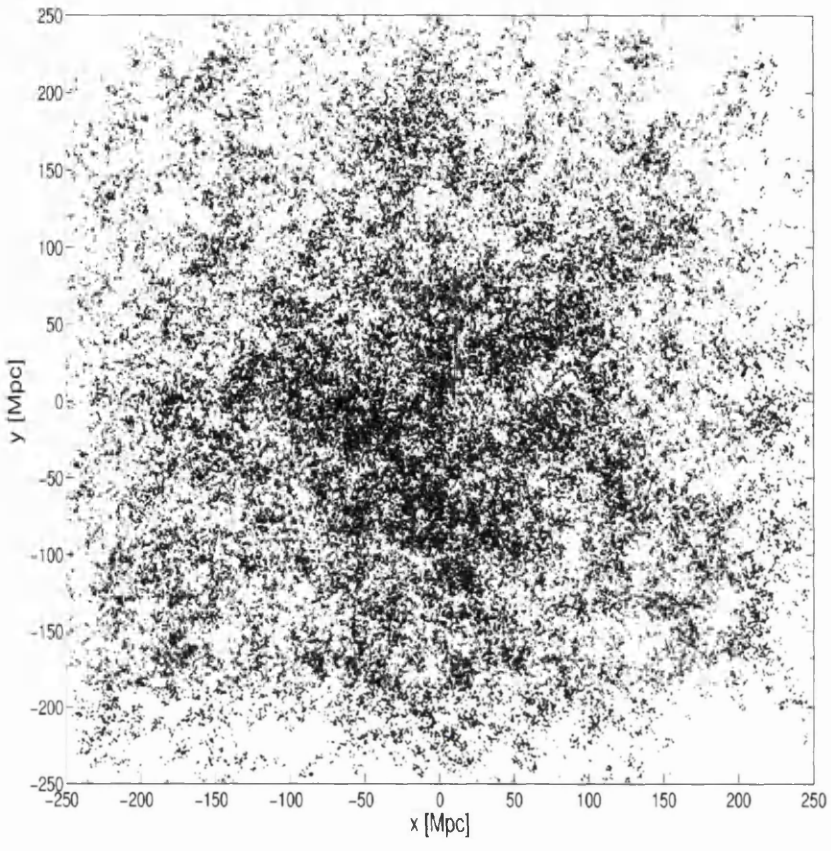

Figure 5.1: This is a 2-D view of a 3-D Lévy Flight distribution of particles, with parameters: $r_0 = 0.2$, $D = 1.2$

for concern when applying a statistic which does not consider angular information. Although this may be a problem, following the suggestion of Martínez *et al.* (1990), the correct result from the simulation seems to be confined to the range between the average nearest neighbour separation and the mean inter-particle spacing. In figure 5.2 it can be easily seen that the volume correction lies closer to the analytical answer of $D_2 = 1.2$, than either the capacity or angular corrections.

Secondly, the volume correction is more or less always closer to the true value, even as the methods begin to over estimate on larger scales. This over estimation

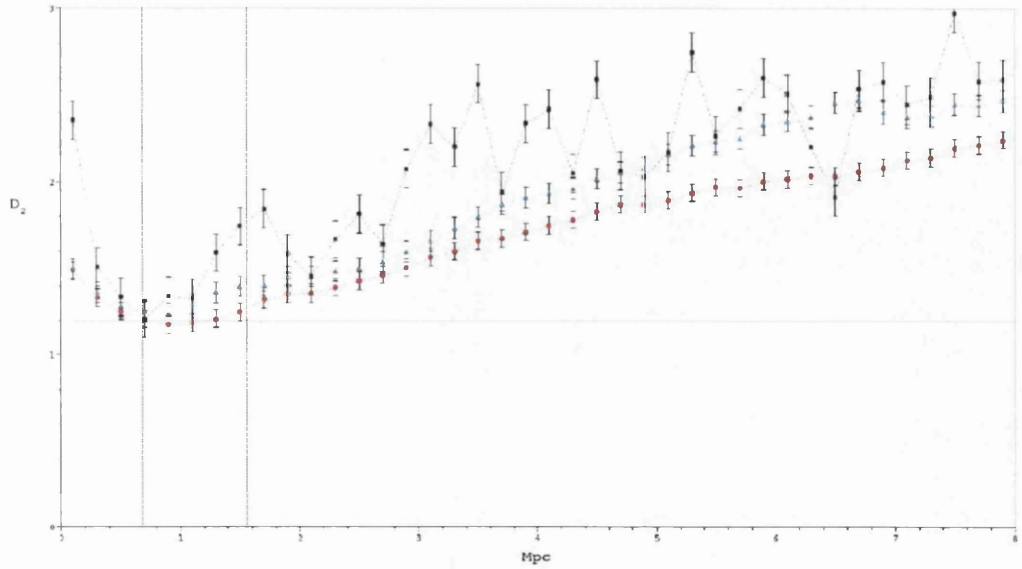

Figure 5.2: D_2 evaluated for a Lévy Flight distribution of particles. The volume (red), capacity (blue) and angular (black) corrections are used to correct for the boundary of the simulation. The error bars are from our prescription as described in §4.6. The solid blue line corresponds to the analytically determined D_2 value and the two vertical dashed lines represent the average nearest neighbour separation and the mean inter-particle spacing respectively.

by all the methods is due to the anisotropic nature of the Lévy Flight distribution.

Thirdly, the large errors and visual noisiness of the angular correction are not present in either the capacity or volume corrections. This, I can only conclude, is due to the angular method throwing away data, leading to low number statistics, *i.e.* intrinsic noise.

5.2 ΛCDM Simulation

In this section we will analyse the distribution of dark matter halos from a ΛCDM simulation. Since there are no galaxies in this analysis, it is only the underlying dark matter distribution which is being probed. This simulation was performed

by Warren *et al.* (2006), see reference for a detailed description. For this work we are using a $384h^{-1}$ Mpc box with a flat geometry and cosmological parameters,

$$\mathbf{p} = (\Omega_M, \Omega_b, n, h, \sigma_8) = (0.3, 0.04, 1, 0.7, 0.9). \tag{5.2}$$

Initial conditions were derived from the transfer functions as calculated by CMB-FAST (Seljak & Zaldarriaga, 1996). The final catalogue has approximately 1.5 million halo positions.

5.2.1 Multifractal Analysis

The multifractal analysis as presented in §3.4.1, must be applied over a range of different distance scales. In figure (3.5) the results of this analysis are shown over two different ranges; $10 < R_1 < 40h^{-1}$Mpc and $50 < R_2 < 100h^{-1}$Mpc. These scales were chosen almost arbitrarily[1]. The left hand plot of fig.(3.5) shows a clear sign of multifractality on small scales whereas the right hand plot appears to signal a transition to homogeneity. However, since the partition function (like figure 3.6 (a)) is generally smooth, we can expect a smooth transition to homogeneity.

Instead of applying this analysis in certain ranges, the plots in figure 3.5 could be extended by adding another axis: distance. This would give us a $D_{q,R}$ *Surface.*

The $D_{q,R}$ Surface

Figure 5.3 shows the result of extending the multifractal analysis to include varying scales. The high peak and dip on low scales ($< 10h^{-1}$Mpc) corresponds to a multifractal distribution. The surface then levels off to a constant value of three

[1]The only reason was that the partition sum, $Z(q, R)$, seemed to have a constant gradient in these regions.

advocating a transition to homogeneity at a scale of $\approx 30h^{-1}$Mpc. Another interesting feature is that homogeneity is not reached at the same scale for all q.

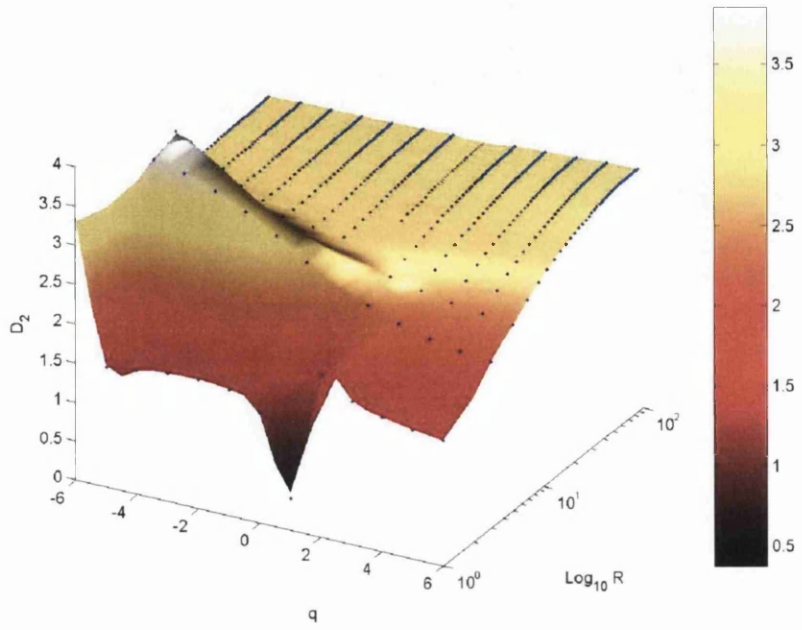

Figure 5.3: This is a 3-d surface fit of $D_q(r)$. The data used is the halo positions from a ΛCDM cosmology. Described in Warren *et al.* 2006.

5.3 PSCz Mock Catalogue

In the previous section the whole halo catalogue was used, thus the analysis is only relevant for the underlying dark matter distribution. In this section the galaxies are under investigation. The galaxy mock catalogue is extracted from the N-body simulation (*cf.* § 5.3.1).

5.3.1 Making Mock Catalogues

Mock *PSCz* catalogues have been extracted from the N-body simulations performed by Cole *et al.* (1997). They used the AP^{3}M code of Couchman (1991) loaded with 192^3 particles in a box of comoving size of $345.6\,h^{-1}$Mpc. The particle mass is $1.62\times10^{12}\,\Omega_M h^{-1} M_\odot$. Further details can be found in Cole *et al.* (1997). For the analysis performed in § 5 we have considered two different cosmologies (*cf.* Table 5.3.1): a flat model $\Omega_M = 0.3$ and cosmological constant term, $\Omega_\Lambda c^2/3H_0^2 = 0.7$ and a critical density universe ($\Omega_M = 1.0$) with power spectrum shape parameter, $\Gamma = 0.25$. The relevant details of the two cosmological models explored are summarised in the table below.

Model	Ω_M	Ω_Λ	Γ	σ_8
LCDM	0.3	0.7	0.25	1.13
SCDMG	1.0	0.0	0.25	0.55

Table 5.1: Cosmological Models.

Ten different mock catalogues, which we will refer to as LCDM0i, SCDMG0i, $i = 0, ..., 9$, have been extracted from each of the above models.

Although in this work we have not created an N-body code or extracted the mock catalogues from it, it is still worth discussing the procedure for doing so.

- A population of particles with properties similar to those of the Local Group (LG) is identified. A LG-like observer is defined by implementing two observationally based constraints: the peculiar velocity of the point must be $\mathbf{v}_{LG} = 625\pm25\,\mathrm{km\,s^{-1}}$, and the particle must be located in a region for which

the fractional overdensity δ averaged in a radius of $5\,h^{-1}$Mpc is in the range $-0.2 < \delta < 1.0$.

- A sphere of $120\,h^{-1}$Mpc radius is drawn around the LG-like observer and the whole frame is rotated so that the motion of the observer is in the direction $(l = 276°, b = 30°)$, the direction of the LG peculiar velocity with respect to the CMB frame (e.g. Wilkinson 1988).

- A friends-of-friends algorithm is implemented to find galaxy clusters. (see Frenk *et al.* 1988 for details)

- The number density of particles in the simulation is $\sim 0.039\ h^3\mathrm{Mpc}^{-3}$ while the number density given by the *PSCz* selection function exceeds this density closer than some critical distance. Thus, the simulations are volume-limited for distances less than $10.9\,h^{-1}$Mpc, where the *PSCz* number density (Saunders *et al.* 1999) matches the N-body one. For distances greater than this the simulated surveys follow the *PSCz* number density.

- A Monte Carlo rejection was used to choose particles according to the *PSCz* selection function (Saunders *et al.* 2000):

$$\phi(r) = \left(\frac{r_\circ}{r}\right)^{2\alpha} \left(\frac{r_\star^2 + r_\circ^2}{r_\star^2 + r^2}\right)^{\beta}. \tag{5.3}$$

  The optimal parameters are listed in Table 5.3.1. For $r \leq 10.9\,h^{-1}$Mpc the mock catalogues are volume-limited and thus $\phi(r) = 1$. A random flux consistent with the *PSCz* selection function is then attributed at each selected galaxy.

- Despite the large sky coverage, *PSCz* is not a full-sky catalogue. Unsurveyed regions are present both at high and low galactic latitudes that need to be accounted for to properly reproduce the existing observational biases. All

α	β	$r_\circ$ $[h^{-1}\text{Mpc}]$	$r_\star$ $[h^{-1}\text{Mpc}]$
0.53	1.90	10.90	86.40

Table 5.2: Selection Function Parameters for Eq. (5.3)

the galaxies which fall within masked regions have been rejected leaving the final sky coverage to be $\approx 84\%$ complete.

The final mock catalogue contains the positions of the galaxies in redshift-space and their "observed" flux. The galaxy redshifts are assigned by adding the line-of-sight component of the peculiar velocity to the recession velocity.

5.3.2 Multifractal Analysis

Figure 5.3 shows the $D_{q,R}$ surface for the halo positions in an ideal and complete $(400\ h^{-1}\text{Mpc})^3$ box. It has a standard ΛCDM cosmology with no galaxies, so only the underlying dark matter distribution was investigated. Overall, it is a very smooth surface, which tends towards homogeneity at scales $> 30h^{-1}\text{Mpc}$. A clear peak, at low q, and dip, at high q corresponds to multifractality for $R \approx 10h^{-1}\text{Mpc}$.

The same $D_{q,R}$ surface analysis is repeated, this time for the Mock galaxy catalogues mentioned above. The main difference is, now there are a lot less particles to analyse, $\sim 15,000$. However, given that we have 10 mock realisations to average over, the noise should not be much worse. Figure 5.6 shows the averaged $D_{q,R}$ surfaces of the ΛCDM and the SCDMG cosmologies.

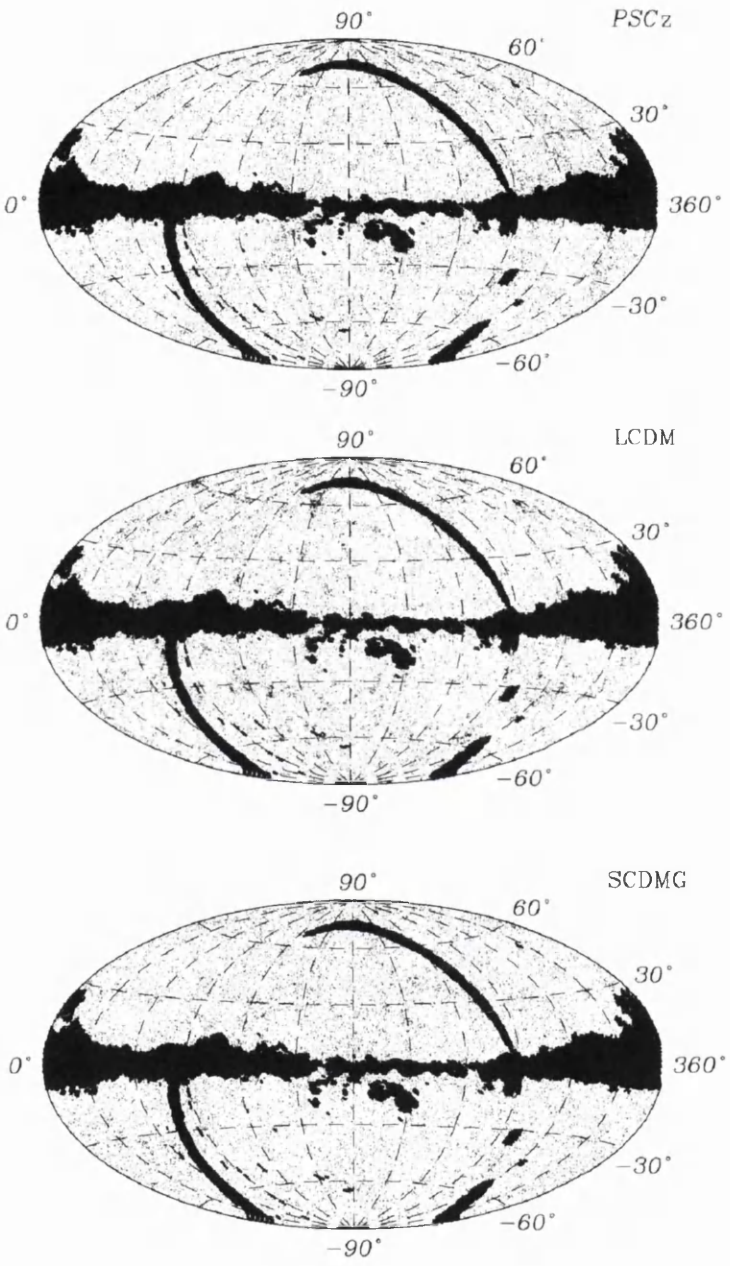

Figure 5.4: Sky distribution of galaxies in the *PSCz* and in two N-body mock-catalogues. From the top to the bottom we illustrate mocks drawn from the LCDM and SCDMG cosmologies, respectively. The Aitoff projection is in Galactic coordinates.

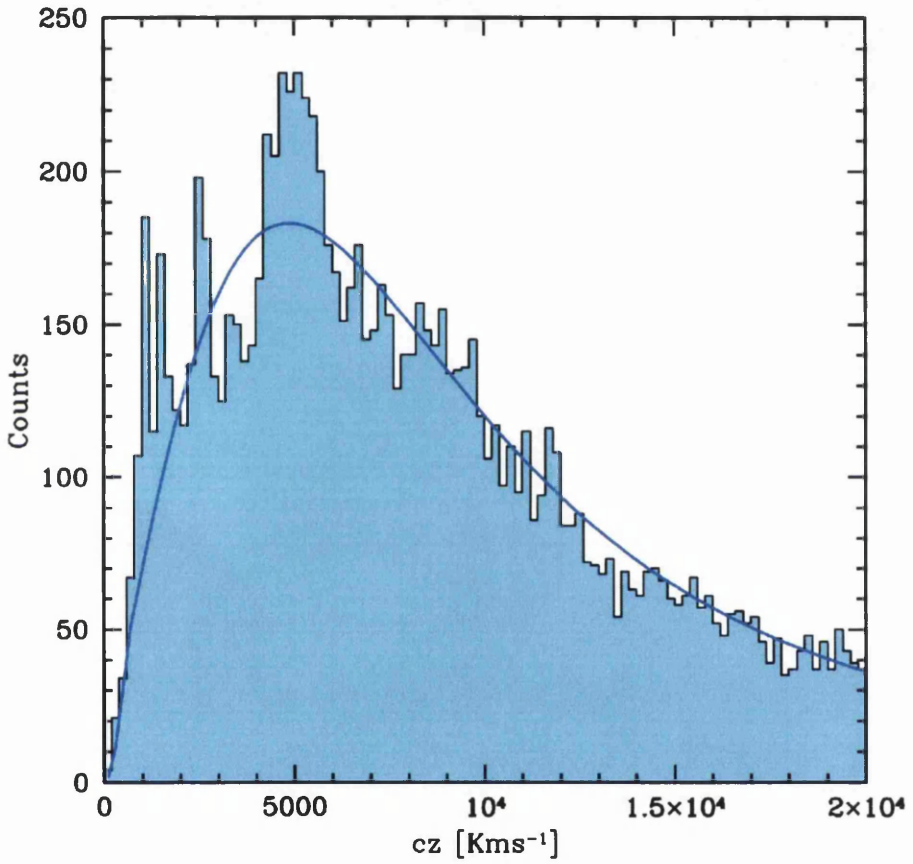

Figure 5.5: Histogram of the radial number count, dN, in the PSCz survey. The solid line represents the expected number in each bin after the inclusion of the selection function,ϕ. *i.e.* $dN = 4\pi\bar{n}\phi(r)r^2dr$.

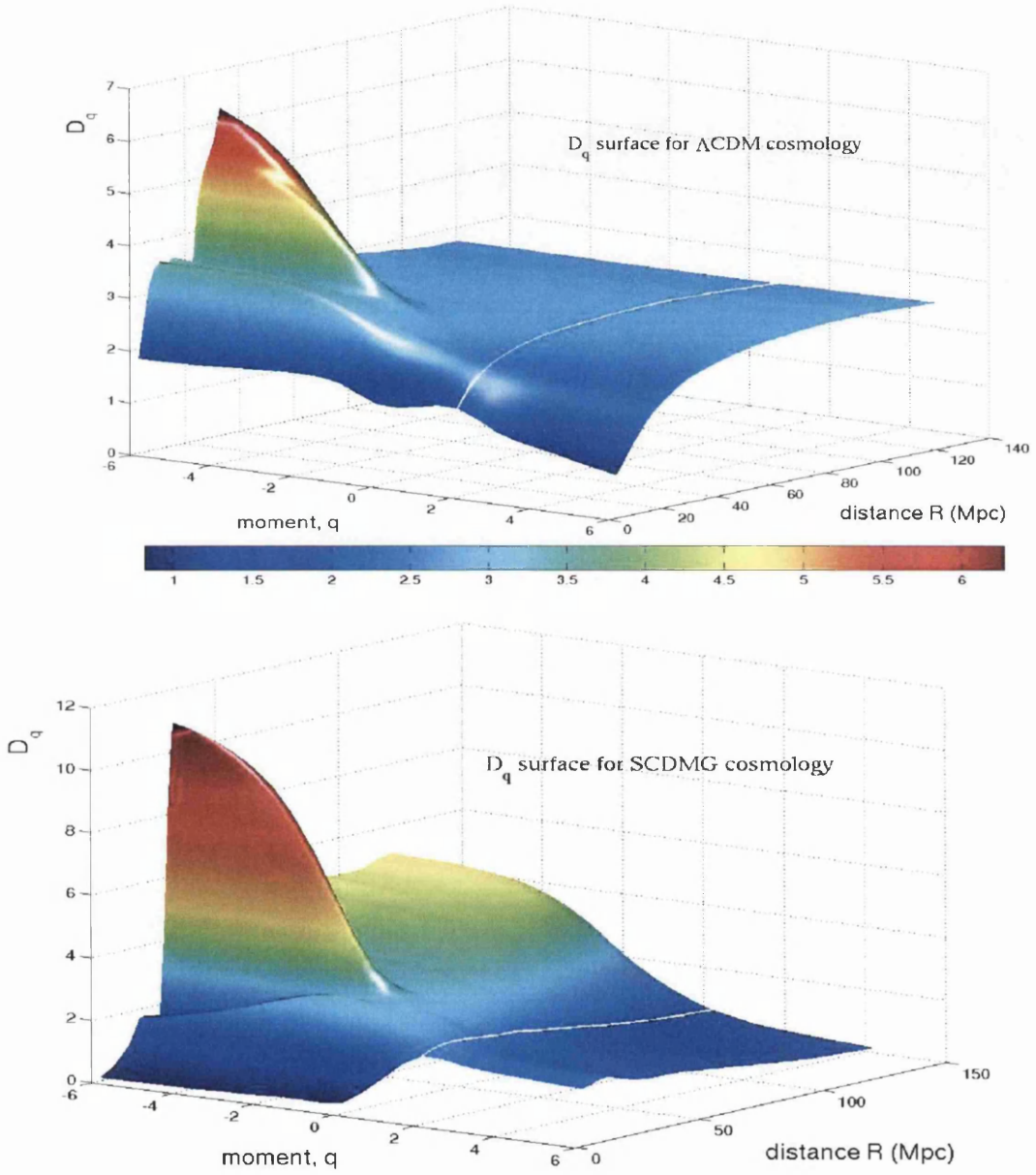

Figure 5.6: $D_{q,R}$ surfaces for two different cosmologies. The white reference line corresponds to the correlation dimension, D_2, which is usually obtained from the derivative of the 2PCF. *Top*: A ΛCDM cosmology. *Bottom*: An SCDMG cosmology. The two white reference curves are very similar on small scales in that they give the usual value of $D_2 \sim 2$. On larger scales, however, it is evident that the $D_{q,R}$ surface can in principle differentiate between different cosmological models.

$D_{q,R}$ Surface - ΛCDM

There are some points we should note regarding the $D_{q,R}$ surfaces of figure 5.6. The ΛCDM cosmology yields a plot quite similar to that of figure 5.3, this should not be too surprising given that they have the same cosmological parameters. There is however clearly a few features on small scale where the two differ. In fig.(5.6) at $\sim 10h^{-1}$Mpc there is a multifractal feature peaking at $D_{q=-6} \approx 3.5$. A little further out at $\sim 20h^{-1}$Mpc there is a second multifractal signature with a much higher peak $D_{q=-6} \approx 6.5$. Recalling that figure 5.6 is an average taken over 10 realisations, it is interesting that the two multifractal features on small scales have not merged into one *i.e.* they are probably not statistical anomalies. It would be premature to suggest that these distinct features have any physical significance[2] but there is definitely room here for further study.

On scales larger than $30h^{-1}$Mpc the surface flattens to homogeneity, $D_q \rightarrow 3$. The white reference line shows the correlation dimension, D_2, varying with scale. This is related to the derivative of the 2PCF. Visually this shows that all the information which the 2PCF can provide is but a very small part of the $D_{q,R}$ surface.

$D_{q,R}$ Surface - SCDMG

In the lower panel of figure 5.6, the $D_{q,R}$ surface is plotted for the SCDMG cosmology. It clearly shows on small scales a similar multifractal peak to that of the ΛCDM model. However, there seems to be only one very sharp distinct peak at $\sim 20h^{-1}$Mpc and for $q = 6$, $D_{q=6} \approx 11$. Casting our mind back to §3.4.1, this is telling us that on scales $\sim 20h^{-1}$Mpc void under-dense regions are more clustered.

[2]This same point has been made by Bernard Jones in many of his papers, most recently in the review article (Jones *et al.* ,2005).

Looking beyond the first peak it can be seen that the surface does not flatten off like the ΛCDM model. In fact on scales beyond $30h^{-1}$Mpc there is no transition to homogeneity in this universe, on scales up to the size of the simulation it is entirely multifractal.

It is evident, therefore, that on large scale the $D_{q,R}$ surface can in principle discriminate between different cosmological models.

Chapter 6

Summary

We have shown that our *volume* correction recovers the true fractal dimension, when tested with the extreme case of an anisotropic Lévy Flight. The increase in accuracy over Pan & Coles angular correction estimator is marginal, however the computational load required is much less. This fact will become increasingly more important as red-shift surveys are now pushing 1,000,000+ galaxies.

We present the $D_{q,R}$ surface as a possible unique descriptor of a discrete point distribution. Whether this is strictly true or not does not differ from the fact that there is much more information contained in our multifractal analysis than can be extracted from the usual 2PCF. In fact to go from a 2PCF to our multifractal measure, is of no significant computational cost.

The 2PCF in our formalism, is represented by the integral along the $q = 2$ line in figures (5.6) & (5.3). It is clearly apparent that $q = 2$ is confined to a rather boring and flat part of the surface and generally for $q > 1$ the surface is very smooth and D_q tends towards a constant value without any interesting features.

There has been a lot of interest, since the dawn of cosmological simulations, to apply statistics to the resulting distributions so that a comparison can be made

with real surveys. Much of this effort has involved the 2PCF (also Minimal Spanning Trees and other geometrically motivated descriptors). However it must be made clear that the trade off between statistical robustness and visual interpretation is of prime importance when quantifying structure. Neither should be favoured too heavily.

As we have presented here, our $D_{q,R}$ surface gives varying weights to under dense (void) regions and over dense (clustered) regions, this obviously has a significant advantage over calculating higher order correlations. The major task now, regarding the $D_{q,R}$ surface, is to make full use of it and to extracted as much information as is statistically possible.

Our methodology as presented in §3.4.1 also has the possibility to contribute to parameter estimation. Especially in the area of Baryon Acoustic Oscillations (BAO). BAO's are difficult to measure accurately due to the low amplitude on scale above 100Mpc. To help increase the power in this range, clusters, instead of individual galaxies, could be used as they show a higher clustering amplitude. This technique has some drawbacks The biggest problem it faces is in accurately determining what is a cluster. Another technique for measuring BAO's would be to use the $D_{q,R}$ surface as we introduced in §5.2.1. Since this method treats dense and under-dense regions differently, it suggests that the usual 2PCF (or $Z(2,r)$) may not be the best suited to observing and measuring BAO's. In future work we hope to explore the use of the $D_{q,R}$ surface as a cosmological tool.

Bibliography

[1] S. J. Aarseth. Computer simulations of galaxy clustering. In M. S. Longair and J. Einasto, editors, *IAU Symp. 79: Large Scale Structures in the Universe*, pages 189–194, 1978.

[2] K. Abazajian *et al.*. The Third Data Release of the Sloan Digital Sky Survey. *Astro. J.*, 129, March 2005.

[3] G. O. Abell. The Distribution of Rich Clusters of Galaxies. *ApJ Suppl.*, 3(211), May 1958.

[4] J. D. Barrow, S. P. Bhavsar, and D. H. Sonoda. Minimal spanning trees, filaments and galaxy clustering. *MNRAS*, 216:17–35, September 1985.

[5] E. Bertschinger and J. M. Gelb. Cosmological N-body simulations. *Computers in Physics*, 5:164–175, April 1991.

[6] S. Boldyrev and C. R. Gwinn. Lévy Model for Interstellar Scintillations. *Physical Review Letters*, 91(13):131101, September 2003.

[7] S. Borgani, V. J. Martinez, M. A. Perez, and R. Valdarnini. Is there any scaling in the cluster distribution? *ApJ*, 435, November 1994.

[8] S. M. Carroll. *Spacetime and geometry. An introduction to general relativity.* Spacetime and geometry / San Francisco, CA, USA: Addison Wesley, ISBN 0-8053-8732-3, 2004, 2004.

[9] D. Chowdhury and D. Stauffer. A generalized spin model of financial markets. *Euro. Phys Journal B*, 8:477, 1999.

[10] R. Clark and W.F. Miller. *Math. Comp. Phy*, 1966.

[11] S. Cole, D. H. Weinberg, C. S. Frenk, and B. Ratra. Large-scale structure in COBE-normalized cold dark matter cosmogonies. *MNRAS*, 289, July 1997.

[12] M. Colless. First results from the 2dF Galaxy Redshift Survey. In G. Efstathiou, editor, *Large-Scale Structure in the Universe*, 1999.

[13] H. M. P. Couchman. Mesh-refined P3M - A fast adaptive N-body algorithm. *ApJ Lett.*, 368, February 1991.

[14] M. Davis and P. J. E. Peebles. A survey of galaxy redshifts. V - The two-point position and velocity correlations. *ApJ*, 267:465–482, April 1983.

[15] A. Doroshkevich, D. L. Tucker, S. Allam, and M. J. Way. Large scale structure in the SDSS galaxy survey. *A&AP*, 418:7–23, April 2004.

[16] G. Efstathiou, M. Davis, S. D. M. White, and C. S. Frenk. Numerical techniques for large cosmological N-body simulations. *ApJ Suppl.*, 57:241–260, February 1985.

[17] C. S. Frenk, S. D. M. White, M. Davis, and G. Efstathiou. The formation of dark halos in a universe dominated by cold dark matter. *ApJ*, 327, April 1988.

[18] J. R. Gott, III and E. L. Turner. An extension of the galaxy covariance function to small scales. *ApJ Lett.*, 232:L79–L81, September 1979.

[19] P. Grassberger and I. Procaccia. Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A*, 28, October 1983.

[20] S. Hatton. Approaching a homogeneous galaxy distribution: results from the Stromlo-APM redshift survey. *MNRAS*, 310:1128–1136, December 1999.

[21] D. J. Heath. The growth of density perturbations in zero pressure Friedmann-Lemaitre universes. *MNRAS*, 179:351–358, May 1977.

[22] H. G. E. Hentschel and I. Procaccia. Fractal nature of turbulence as manifested in turbulent diffusion. *Phys. Rev. A*, February 1983.

[23] G. Hinshaw, D. N. Spergel, L. Verde, R. S. Hill, S. S. Meyer, C. Barnes, C. L. Bennett, M. Halpern, N. Jarosik, A. Kogut, E. Komatsu, M. Limon, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: The Angular Power Spectrum. *ApJ Suppl.*, 148:135–159, September 2003.

[24] J. Huchra, M. Davis, D. Latham, and J. Tonry. A survey of galaxy redshifts. IV - The data. *ApJ Suppl.*, 52, June 1983.

[25] B. J. Jones, V. J. Martínez, E. Saar, and V. Trimble. Scaling laws in the distribution of galaxies. *Reviews of Modern Physics*, 76:1211–1266, February 2005.

[26] B. J. T. Jones, V. J. Martinez, E. Saar, and J. Einasto. Multifractal description of the large-scale structure of the universe. *ApJ Lett.*, 332, September 1988.

[27] M. Kerscher, I. Szapudi, and A. S. Szalay. A Comparison of Estimators for the Two-Point Correlation Function. *ApJ Lett.*, 535:L13–L16, May 2000.

[28] E. Komatsu, A. Kogut, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, L. Verde, E. Wollack, and E. L. Wright. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Tests of Gaussianity. *ApJ Suppl.*, 148, September 2003.

[29] S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *ApJ*, 412:64–71, July 1993.

[30] J. Loveday, B. A. Peterson, G. Efstathiou, and S. J. Maddox. The Stromlo-APM Redshift Survey. I - The luminosity function and space density of galaxies. *ApJ*, 390, May 1992.

[31] C.-P. Ma and E. Bertschinger. Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges. *ApJ*, 455, December 1995.

[32] S. J. Maddox, G. Efstathiou, W. J. Sutherland, and J. Loveday. The APM galaxy survey. I - APM measurements and star-galaxy separation. *MNRAS*, 243:692–712, April 1990.

[33] B. B. Mandelbrot. *The Fractal Geometry of Nature.* The Fractal Geometry of Nature, San Francisco: Freeman, 1982, 1982.

[34] V. J. Martinez, B. J. T. Jones, R. Dominguez-Tenreiro, and R. van de Weygaert. Clustering paradigms and multifractal measures. *ApJ*, 357:50–61, July 1990.

[35] P. Meakin. *Fractal, scaling and growth far from equilibrium.* Cambridge University Press, 1998.

[36] A. Meiksin, I. Szapudi, and A. Szalay. Higher order correlations of IRAS galaxies. *ApJ*, 394:87–90, July 1992.

[37] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation.* San Francisco: W.H. Freeman and Co., 1973, 1973.

[38] J. Pan and P. Coles. Large-scale cosmic homogeneity from a multifractal analysis of the PSCz catalogue. *MNRAS*, 318:L51–L54, November 2000.

[39] J. Pan and P. Coles. Boundary corrections in fractal analysis of galaxy surveys. *MNRAS*, March 2002.

[40] P. J. E. Peebles. *The large-scale structure of the universe.* Princeton University Press, 1980. 435 p., 1980.

[41] P. J. E. Peebles. *Principles of physical cosmology.* Princeton University Press, 1993.

[42] W. J. Percival *et al.* . The shape of the SDSS DR5 galaxy power spectrum. *ArXiv Astrophysics e-prints*, August 2006.

[43] W. C. Saslaw. *The Distribution of the Galaxies.* The Distribution of the Galaxies. ISBN 0521394260. Cambridge University Press, October 1999.

[44] W. Saunders, S. Oliver, O. Keeble, M. Rowan-Robinson, S. Maddox, R. McMahon, G. Efstathiou, W. Sutherland, H. Tadros, S. D. M. White, and C. S. Frenk. The IRAS Point Source Catalog Redshift (PSCz) Survey. In *ASP Conf. Ser.*, 2000.

[45] U. Seljak and M. Zaldarriaga. A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *ApJ*, 469, October 1996.

[46] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *nature*, 435, June 2005.

[47] M. A. Strauss and J. A. Willick. The density and peculiar velocity fields of nearby galaxies. *Phys. Rep.*, 261:271–431, 1995.

[48] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S-PLUS.* Springer, New York, 1999.

[49] J.-M. Virey, P. Taxil, A. Tilquin, A. Ealet, C. Tao, and D. Fouchez. Determination of the deceleration parameter from supernovae data. *Phys. Rev. D*, 72(6):061302, September 2005.

[50] M. S. Warren, K. Abazajian, D. E. Holz, and L. Teodoro. Precision Determination of the Mass Function of Dark Matter Halos. *ApJ*, 646, August 2006.

[51] S. Weinberg. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity.* ISBN 0-471-92567-5. Wiley-VCH, July 1972.

[52] D. T. Wilkinson. Recent Measurements of the Cosmic Microwave Radiation. In J. Audouze, M.-C. Pelletan, and S. Szalay, editors, *IAU Symp. 130: Large Scale Structures of the Universe*, 1988.

[53] F. Zwicky *et al.*. Catalogue of Galaxies and of Clusters of Galaxies. vol. 6. In *California Inst. Techn.*, 1968.