GALAXY CLUSTERING IN LARGE REDSHIFT SURVEYS

ED HAWKINS

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

December 2002

"There is a theory which states that if ever anyone discovers exactly what the universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable.

There is another that states that this has already happened."

—– **Douglas Adams, *The Hitchhikers Guide to the Galaxy***

Supervisor: Dr. Steve Maddox

Examiners: Dr. Reynier Peletier
Dr. Jon Loveday

Contents

Abstract

This thesis analyses the large scale distribution of galaxies in two recently completed redshift surveys and develops models to infer estimates of redshift-space distortions. These models demonstrate that coherent infall cannot be neglected in analyses of peculiar velocities as previous works have claimed.

The clustering statistics of the largest infra-red (IR) redshift survey to date, the PSCz, are measured for the whole survey and also for various sub-samples, split by absolute luminosity and colour. It is found that IR selected galaxies are less strongly clustered than optically selected samples. A weak dependence of clustering on IR colour is found, along with a consistent dependence on IR absolute luminosity.

An analysis of the 2dF Galaxy Redshift Survey (2dFGRS), containing $\sim 220\,000$ galaxy redshifts, is presented. Detailed, high precision measurements are made of both the redshift- and real-space correlation functions. A comprehensive analysis of the measured 2-d correlation function leads to accurate estimates of the redshift-space distortions, including the infall parameter, β. This measurement is then combined with other recent works to constrain the matter density of the Universe, $\Omega_{\rm m}$.

The 2dFGRS is also split by absolute luminosity and spectral type to look for dependence of clustering on these galaxy properties. A strong dependence on both luminosity and spectral type is found.

The 2dF Quasar Redshift Survey (2QZ) is used, along with the 2dFGRS, to look for periodicites in the redshift distributions of QSO's in projected close galaxy-QSO pairs. Recent works have reported such an effect in small samples. No such periodicity is found in the largest, most well defined sample tested to date.

These results are entirely consistent with the current cosmological paradigm of a Universe dominated by a cosmological constant and cold dark matter, and also the notion that the large scale distribution of galaxies depends on galaxy type.

Chapter 1

Introduction

This chapter introduces some background information and describes some of the basic cosmological concepts, definitions and observations which underpin the work in subsequent chapters.

1.1 The beginning

The generally accepted picture of the Universe is based on three fundamental principles: that (i) distant objects are receding, implying that the Universe is expanding, (ii) the Universe is isotropic and (iii) the galaxies trace the mass.

1.1.1 The expanding Universe

In the late 1920's it was postulated that an object's distance was related to the velocity with which it was receding. By measuring the Doppler shift of spectral lines in nearby galaxies it had been discovered previously that most were moving away from us. Using galactic apparent magnitudes and Cepheid variable stars in nearby galaxies as 'standard candles', Edwin Hubble and others were able to estimate their distances and discover what has become known as Hubble's Law,

$$v = H_0 d \tag{1.1}$$

where a galaxy's receding velocity, v, is proportional to its distance, d, from the observer. The constant of proportionality, H_0, is known as Hubble's constant, and initial estimates for this parameter were around 500 km s^{-1}Mpc^{-1}.

Until the late 1990's the exact value of the parameter was still uncertain although most agreed that the value was in the range 50 - 100 km s^{-1}Mpc^{-1}. One of the primary goals of the Hubble Space Telescope, launched in 1990, was to obtain an accurate estimate of Hubble's constant. Being above the atmosphere and therefore avoiding the absorption and disturbances that ground-based telescopes have to suffer, allowed much more distant 'standard candles' such as supernovae to be observed. This allowed a more accurate determination of Hubble's constant and the value obtained was $H_0 = 72 \pm 8$ km s^{-1}Mpc^{-1} (Freedman et al. 2001). The uncertainty in this quantity is parameterized by h, such that $H_0 = 100\,h$ km s^{-1}Mpc^{-1}.

In whichever direction we look it seems as though distant objects are speeding away at high velocities. This universal expansion of the Universe was the first strong evidence for the theory that was to become known as the Big Bang.

To estimate the velocity with which an object is receding you first need a spectrum of the emitted light. It is then straight-forward to measure the redshift, z, by comparing the wavelengths of spectral lines in the observed object to those measured in a laboratory,

$$\frac{\lambda_{\rm obs}}{\lambda_{\rm lab}} = 1 + z \tag{1.2}$$

and then the velocity, $v = cz$ for small z. And so, by measuring redshifts we can infer distances via Hubble's Law, and so produce a three dimensional picture of the surrounding Universe.

1.1.2 The isotropic Universe

The philosophy that we do not live in a special region of the Universe combined with the observations of the almost completely homogeneous nature of the Cosmic Microwave Background (CMB) on large scales leads to the Cosmological Principle which can be simply stated as:

> *The Universe is homogeneous throughout space and isotropic about every point.*

A glance at the night sky shows a great many inhomogeneities; stars, galaxies and clusters, meaning that this principle is not true on all scales. It is merely a statement about the global properties of the Universe and our (non-special) place in it.

1.1.3 The Universe in equations

By solving Einstein's field equations, assuming the Universe behaves as a perfect fluid, Friedmann calculated the equations which describe the geometry, energy content and evolution of the Universe. In terms of the pressure, p, mass density, ρ_m, and curvature, K, the scale factor $a(t)$, evolves like,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho_m + \frac{3p}{c^2}\right) + \frac{\Lambda}{3} \tag{1.3}$$

and

$$H(t)^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho_m}{3} - \frac{Kc^2}{a^2} - \frac{\Lambda}{3} \tag{1.4}$$

where $a(t)$ describes the scale of the Universe and $H(t)$ is the time-varying form of Hubble's constant (Equation 1.1) and Λ is (for now) just a constant. In this notation,

$$1 + z = \frac{1}{a(t)} \tag{1.5}$$

where the scale factor at the present day, $a(t_0) = 1$.

The measurement of the density of the Universe, and the relative proportions of its contents, including the Λ term, is a major target of current cosmological research. Using Equation 1.4 and rewriting $\Lambda = -8\pi G\rho_\Lambda$ and defining a total density, $\rho_t = \rho_m + \rho_\Lambda$, we can define a 'critical density' so that the curvature, $K \equiv 0$, ensuring a 'flat' Universe,

$$\rho_c = \frac{3H^2}{8\pi G} \tag{1.6}$$

and to compare to the actual value the density parameters are defined relative to the critical density,

$$\Omega = \frac{\rho_t}{\rho_c} \qquad \Omega_{\rm m} = \frac{\rho_m}{\rho_c} \qquad \Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c} \tag{1.7}$$

Historically there have been many arguments over whether or not there is enough mass in the Universe to reverse the overall expansion, i.e. whether $\Omega_{\rm m} > 1$. There now seems to be a consensus over this debate and this has led to a generally accepted cosmological model.

1.2 The fashionable cosmology

The comparison of observations with theory seems to have broadly agreed on the fundamental cosmological parameters. Although there is much detail to work out it now seems as though the structure is in place.

1.2.1 The mass

Total: Measurements of the large scale structure of the Universe, and particularly the measurement of the gravitational infall of galaxies into clusters (e.g. Peacock et al. 2001), have concluded that the total mass density $\Omega_{\rm m} \sim 0.3$. This will be a topic discussed at length in later chapters. The completion of future large redshift surveys will refine the measurements from this method.

Baryons: Big Bang Nucleosynthesis (BBN) is the study of how the primordial protons and neutrons combined to form the light elements of hydrogen, deuterium, isotopes of helium and lithium. By comparing predictions with observations of their relative abundances in stars today it is calculated that the density of baryons is $\Omega_{\rm b} \approx 0.04$ (e.g. O'Meara et al. 2001), not nearly enough to 'close' the Universe.

Stars: The stars that we see are all made of baryons, generally burning hydrogen to form helium and the heavier elements. By measuring the total mass of stars it is thought that $\Omega_{\rm stars} \sim 0.004$ (Fukugita, Hogan & Peebles 1998).

Dark Matter: By observing the rotation curves of nearby galaxies it is seen that the outer regions are rotating too fast for the apparent mass, which is derived from the number of stars.

This all leads to the conclusion that there must be mass that we cannot see, leading to the term 'dark matter', and that there are both baryonic and non-baryonic components.

1.2.2 The cosmological constant

One of the most important pieces in the cosmic jigsaw is from measurements of the CMB, the relic light from the earliest moments of the Universe, which has now cooled to around 2.7 K. The observations that the CMB is uniform across

the whole sky to better than 1 part in 10^5, and that the radiation has the form of a black-body, is a strong argument that this is indeed the light from the Big Bang. The tiny temperature fluctuations that do exist in the CMB can be used to probe the physics of the early Universe - their strength and position tells us how large the initial mass density fluctuations were. Latest data indicate that $\Omega \approx 1$ (e.g. Netterfield et al. 2002). With the upcoming data from the *MAP* satellite the accuracy of these results will be much improved.

If the total density is close to unity but the mass content is ~ 0.3 then there must be something providing the rest. Recent observations of distant supernovae (e.g. Perlmutter et al. 1999) show tantalising evidence that the Universe is not only expanding, but accelerating. This is consistent with the existence of a 'cosmological constant' which was first proposed by Einstein to promote the steady-state theory of the Universe. The cosmological constant, or 'dark energy', Λ, is a form of vacuum energy that is repulsive, and has a density Ω_Λ. Future planned space missions (e.g. *SNAP*) dedicated to searching for distant supernovae will revolutionise this field.

1.2.3 The concordance values?

The combination of these observations, along with others (e.g. gravitational lensing), suggest a near-flat, Big Bang Universe with, at the present day[1],

$$\begin{aligned} H_0 &\approx 70 \text{ km s}^{-1}\text{Mpc}^{-1}, \\ \Omega_\text{m} + \Omega_\Lambda = \Omega_0 &\approx 1, \\ \Omega_\text{m} &\sim 0.3, \\ \Omega_\Lambda &\sim 0.7, \\ \Omega_\text{b} &\sim 0.04, \end{aligned}$$

with (possibly) neutrinos and other non-baryons to make up the mass.

Most current observations are consistent with this paradigm, which is known as ΛCDM, but it is worth reflecting that (i) there are some discrepancies (e.g. satellite galaxies) and (ii) in adopting this concordance model we find that $\approx 95\%$ of the Universe is in an unknown form!

Work is well under way to determine the nature of dark matter from a wealth

[1] denoted by a subscript $_0$

of exotic possibilities (e.g. WIMP's and MACHO's). The 'cosmological constant' is perhaps a larger mystery with predictions from particle physics being 60-120 orders of magnitude too large. Whether we have got the right model is still a matter of debate.

1.3 Structure formation

1.3.1 Density fluctuations

The first significant detection of the primordial density fluctuations in the CMB, mentioned above, were taken by the *COBE* satellite, launched in the early 1990's. Since then subsequent experiments have refined these observations and they are now performed by a multitude of instruments all over the world. These tiny density fluctuations, of size δ, in the mean density of the Universe, $\langle\rho\rangle$,

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \langle\rho\rangle}{\langle\rho\rangle} \tag{1.8}$$

are interpreted as the 'seeds' of structure formation which grew to become the stars, galaxies and clusters we see today.

In a continuous density field, $\rho(\mathbf{x})$, with a large enough volume it can be assumed that the Universe is periodic and δ can be expressed as a Fourier series,

$$\delta(\mathbf{x}) = \sum_k \delta_{\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{x}) \tag{1.9}$$

The advantage of this approach is that, while linear theory is valid, the Fourier modes evolve independently. The Power Spectrum is then defined as,

$$P(k) = \langle|\delta_{\mathbf{k}}|^2\rangle \tag{1.10}$$

It is assumed that $P(k) \propto k^n$ so that there is no preferred physical scale to explain.

1.3.2 Growth of structure

In linear theory the growth of perturbations is given by,

$$\frac{\partial^2\delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial\delta}{\partial t} = 4\pi G\rho\delta \tag{1.11}$$

and combining this equation with the Friedmann equations you obtain a relation between δ, a and the cosmological parameters (Lahav et al. 1991).

From calculating the expected peculiar velocities caused by gravitational infall into an overdense region, it was calculated (Peebles 1980) that the growth of structure is governed by the equation,

$$f(\Omega_{\rm m}, \Omega_\Lambda, z) = \frac{d \ln \delta}{d \ln a}. \tag{1.12}$$

A useful approximation to f, at all redshifts, is (Lahav et al. 1991)

$$f = \Omega_{\rm m}^{0.6} + \frac{\Omega_\Lambda (2 + \Omega_{\rm m})}{140} \tag{1.13}$$

which is nearly independent of Λ, and in a flat Universe becomes,

$$f = \Omega_{\rm m}^{0.6} + \frac{(2 - \Omega_{\rm m} - \Omega_{\rm m}^2)}{140} \tag{1.14}$$

which is often approximated further to $f \approx \Omega_{\rm m}^{0.6}$.

1.3.3 Bias

Up to this point properties of the underlying mass density field have been described. When observations of galaxies are made, it is the *galaxy* density field which is examined, not that of the mass. Although it seems natural to assume that the light will trace the mass, the observation that different types of galaxy cluster in different ways (e.g. Dressler 1980) suggests that this is not the whole story.

This leads to the suggestion that galaxies are *biased* tracers of the mass distribution and the definition of a linear bias parameter, b,

$$\delta_{\rm g} = b \delta_{\rm m} \tag{1.15}$$

where the subscripts refer to *galaxies* and *mass*. It needs to be remembered, therefore, that any measurements performed on the galaxy distribution need not imply the same result for the mass distribution. This is only the simplest form for the bias, and more complicated versions exist, such as a bias that varies with scale.

Kaiser (1984) suggested a high peak bias model where galaxies only form at peaks in the mass density field above a certain level. Bardeen et al. (1986) showed that

the highest (implying more massive) peaks are more strongly clustered than the lower (less massive) peaks. This is discussed further in later chapters.

Assuming a linear bias, we define a parameter which describes the growth of galaxy structure,

$$\beta = \frac{f(\Omega_{\rm m}, \Omega_\Lambda, z)}{b} \tag{1.16}$$

and we attempt to measure this parameter for observed data in later Chapters and use it to infer the value of the matter density of the Universe.

1.4 Galaxy redshift surveys

If we want to examine the large scale structure of the Universe we would prefer to know a galaxy's position in three dimensions. Early galaxy surveys could only measure positions on the sky for large numbers of galaxies. Hubble's law (Equation 1.1) gives us an estimate of the all important third dimension measurement, distance, via the redshift. Due to the sheer enormity of the Universe and the number of galaxies it contains it is still only possible to observe a small fraction of the whole. It is therefore important to make the observed sample as representative as possible by having large numbers of objects in a large volume.

The first attempt at this was the Centre for Astrophysics (CfA) redshift survey (Huchra et al. 1983), containing about 2000 galaxies. An order of magnitude larger was the Las Campanas Redshift Survey (LCRS, Shectman et al. 1996) containing about 25 000 galaxies. The infra-red selected PSCz survey (Saunders et al. 2000b), contains about 15 000 galaxies, covering about 84% of the sky - the largest area to date - and is used in Chapters 3 and 4.

As telescopes have grown larger, and as technology has advanced, the taking of redshift measurements has become more and more routine. Multi-fibre spectroscopy now allows several hundred spectra to be taken simultaneously allowing redshifts to be obtained at a far faster rate. The numbers of galaxy redshifts in two new surveys have therefore increased by a further order of magnitude, leading to more accurate and ever more representative results. Far fainter objects can now be observed easily allowing the depths of the surveys to increase also.

The recently completed 2-degree Field Galaxy Redshift Survey (2dFGRS, Colless et al. 2001) contains $\sim$ 220 000 redshifts and is described and analysed in

Chapters 5, 6 and 7. The 2dF Quasar Survey (2QZ, Croom et al. 2001) was observed simultaneously with the 2dFGRS and is also used in Chapter 7. The high redshifts of the nearly 25 000 quasars in this survey give a much larger scale picture of the Universe than has been possible before.

The on-going Sloan Digital Sky Survey (SDSS, Stoughton et al. 2002) has the ambitious goal of detecting 100 million celestial objects, along with measuring 1 million galaxy and 100 000 quasar redshifts, and will be complete in a few years.

The description of how the measurements of galaxy clustering statistics using these surveys can help pin down some of the uncertainties raised in this introduction is included in the following Chapter.

1.5 Layout of the thesis

The work in this thesis is laid out as follows:

- Chapter 2 includes details of the clustering statistics that will be employed in later chapters and describes the models and assumptions which go into these statistics.
- Chapter 3 details work done on measuring the galaxy clustering in the largest infra-red galaxy redshift survey to date, the PSCz.
- Chapter 4 divides the PSCz survey into various sub-samples and analyses the dependence of galaxy clustering on infra-red luminosity and colour.
- Chapter 5 moves into optical clustering, namely the detailed measurements of galaxy clustering and redshift-space distortions in the 2-degree Field Galaxy Redshift Survey.
- Chapter 6 briefly describes some of the other work I have been involved with on the 2dF Team - the dependence of galaxy clustering on luminosity and spectral type.
- Chapter 7 breaks away from the clustering measurements and analyses claims of redshift periodicities in projected close galaxy-quasar pairs, using the 2dFGRS and the 2QZ - the largest sample analysed to date.
- and finally in Chapter 8 I briefly summarise this work.

Chapter 2

Clustering Statistics

This chapter gives an overview of the main statistical methods employed in the later chapters and discusses the models and assumptions which are used. Sections 2.3 to 2.5 are based on work from Hawkins et al. (2002).

2.1 The two-point correlation function, ξ

2.1.1 Definition

One of the most widely used measures of clustering, $\xi(r)$, is defined as the excess probability, over random, that a galaxy is found in a volume dV at a distance r from another galaxy,

$$dP = n[1 + \xi(r)]dV \tag{2.1}$$

where n is the mean number density of galaxies. Thus a uniform random distribution of galaxies will have $\xi = 0$, and large values of ξ indicate a highly clustered distribution. ξ is calculated by counting pairs of galaxies at particular separations and comparing this to a random distribution. In an infinite distribution of points the number of random pairs would be $4\pi n r^2 dr$. Real galaxy surveys are far from this ideal case because of their limited extent and other observational factors, and this is discussed later.

The two-point correlation function is also the Fourier transform of the Power Spectrum defined in Equation 1.10,

$$\xi(r) = \frac{V}{(2\pi)^3} \int P(k) \frac{\sin(kr)}{kr} 4\pi k^2 dk \tag{2.2}$$

If $P(k) \propto k^n$ (i.e. a power law), this implies that $\xi(r) \propto r^{-(3+n)}$, also a power law.

2.1.2 Application to redshift surveys

To measure the distances between pairs of galaxies we need to know their location in three dimensions; their position on the sky gives us two and their redshift gives us a measure of their distance from us.

Redshift surveys therefore start as a source catalogue of object positions, each of which is a potential galaxy. It is then attempted to measure the redshift of each source. To convert between redshift, z, and redshift distance, d, in a flat Universe ($\Omega_{\rm m} + \Omega_\Lambda = 1$) with a cosmological constant, (Kantowski, Kao & Thomas 2000)

$$d = \frac{c}{H_0} \int_0^z \frac{dz'}{1 + \Omega_{\rm m} z'(1 + 3z' + z'^2)} \tag{2.3}$$

Armed with the full 3-d position of each galaxy the distances, s, between pairs of galaxies in redshift space can be easily calculated.

Real redshift surveys are far from ideal due to various observational factors including incompleteness, varying magnitude limits and edge effects. There has therefore been considerable interest in analysing the best way to estimate ξ from such redshift surveys (e.g. Landy & Szalay 1993; Vogeley & Szalay 1996; Hamilton 1997a; Hamilton 1997b). Reliable and unbiased estimates can be obtained by cross-correlating the real data with an artificial unclustered (random) catalogue which has the same selection functions in angle on the sky and in redshift.

2.1.3 Constructing a random catalogue

The random catalogue is created by generating random positions uniformly over the sky, and rejecting positions outside the boundaries of the real data. Many more positions are assigned than real data to try and reduce the shot noise. For each accepted position a redshift is assigned from a redshift distribution which matches the observed data.

The exact methods for generating the redshift distribution for each dataset vary and the particular techniques used for the redshift surveys considered in this work are discussed in the appropriate chapters.

2.1.4 Weighting

Real redshift surveys are restricted at the high redshift range due to a magnitude (or flux) limit, depending on the instrument used and integration times. Thus, at higher redshifts only the intrinsically brightest galaxies can be seen and so the galaxy population is not sampled completely. Due to this under-sampling these high redshift galaxies are given an increased weight. This weighting is also defined to give the minimum variance in the estimate of ξ (Efstathiou 1988; Loveday et al. 1995).

Each galaxy in a pair, separated by a distance, s, is weighted by the factor

$$w_i = \frac{1}{1 + 4\pi n(z_i) J_3(s)}, \tag{2.4}$$

where $n(z_i)$ is the space density of the random galaxies at the redshift z_i for galaxy i, and

$$J_3(s) = \int_0^s \xi(s') s'^2 ds' \tag{2.5}$$

Again, the detailed use of this weighting scheme varies for the survey in question and the details are deferred to the relevant chapter.

2.1.5 Estimators

Various estimators for ξ exist from the obtained pair counts,

(Traditional) $$\xi(s) = \frac{DD}{RR} - 1 \tag{2.6}$$

(Hamilton 1993) $$\xi(s) = \frac{DD\ RR}{DR^2} - 1 \tag{2.7}$$

(Landy & Szalay 1993) $$\xi(s) = \frac{DD - 2DR + RR}{RR} \tag{2.8}$$

where DD is the normalised galaxy-galaxy weighted pair counts at separation, s, RR is the normalised random-random weighted pair counts and DR is the normalised galaxy-random weighted pair counts.

In all the analysis which follows the Landy & Szalay (1993) estimator is adopted but it is tested that our results are consistent with the other estimators. Landy

& Szalay claim that their estimator has the least variance of all the suggested estimators.

2.1.6 A functional form for ξ

Early in the studies of galaxy clustering (e.g. Peebles 1980) it became clear from the measurements that $\xi(s)$ seemed to follow a particularly simple functional form, namely a power law, (cf. Equation 2.2)

$$\xi(s) = \left(\frac{s}{s_0}\right)^{-\gamma_s} \tag{2.9}$$

where s_0 is known as the clustering scale length, and γ_s is the slope of the power law. The measurement of these parameters allows a direct comparison with models, simulations and other surveys or samples.

But it is also known that this power law behaviour of $\xi(s)$ breaks down at small ($\lesssim 1\ h^{-1}$Mpc) and large ($\gtrsim 10\ h^{-1}$Mpc) scales where ξ becomes less than the power law, which is predicted, for instance, by the popular ΛCDM model.

2.2 The 2-d correlation function, $\xi(\sigma, \pi)$

The measurement of $\xi(s)$ is not the end of the story as there is far more information available in the galaxy distribution.

The total separation between galaxy pairs, s, is the sum of two components; the separation perpendicular to (σ) and parallel to (π) the line-of-sight, so that $s^2 = \pi^2 + \sigma^2$.

To estimate $\xi(\sigma, \pi)$ it is necessary to count pairs in bins of both σ and π and weight them using the J_3 scheme as before.

2.2.1 Redshift-space distortions

Of course, when measuring the redshifts of galaxies, we are measuring the total velocity at which the galaxy is receding. This total velocity has at least two components; the main component being of a cosmological origin, due to the expansion of the Universe. The (usually) smaller component will be due to the galaxy's local potential, the 'peculiar velocities', which give rise to the so called

'finger-of-god' effect in redshift cone plots. When converting from redshift to distance it must be remembered that this is the redshift-space distance (s) and not the true real-space distance (r).

There are also two components for the peculiar velocities; that due to the coherent infall around clusters and that due to the random motions of the galaxy. These two types of peculiar velocities distort $\xi(\sigma, \pi)$ in different ways and allow their properties and amplitudes to be estimated and the methods for this are discussed later. Again, these measurements can be compared to models and other surveys to try and understand any differences encountered.

These distortions also affect $\xi(s)$ and so to obtain the real-space clustering, $\xi(r)$, it is necessary to go further.

2.2.2 The projected correlation function, $\Xi(\sigma)$

As the peculiar velocities can only move galaxies along the line of sight it is possible to integrate out this effect and so remove their influence. The first step to measuring the real-space clustering is the estimation of $\Xi(\sigma)$,

$$\Xi(\sigma) = 2 \int_0^\infty \xi(\sigma, \pi)\ d\pi \tag{2.10}$$

though in practice the upper limit of the integral becomes $\pi_{\max}$ (which varies for the survey in question), beyond which it is assumed that $\xi = 0$ and so the integral has converged.

The projected correlation function can also be written as an integral over the real-space correlation function (Davis & Peebles 1983)

$$\frac{\Xi(\sigma)}{\sigma} = \frac{2}{\sigma} \int_\sigma^\infty \frac{r\xi(r)dr}{(r^2 - \sigma^2)^{\frac{1}{2}}} \tag{2.11}$$

Assuming a power law for the real-space correlation function, $\xi(r) = (r/r_0)^{-\gamma_r}$, this integral can be performed analytically,

$$\frac{\Xi(\sigma)}{\sigma} = \left(\frac{r_0}{\sigma}\right)^{\gamma_r} \frac{\Gamma(\frac{1}{2})\Gamma(\frac{\gamma_r - 1}{2})}{\Gamma(\frac{\gamma_r}{2})} = \left(\frac{r_0}{\sigma}\right)^{\gamma_r} A(\gamma_r) \tag{2.12}$$

where Γ is the Gamma Function.

A best-fit method allows the parameters γ_r and r_0 to be calculated from the measured $\Xi(\sigma)$, giving a estimate of the real-space clustering independent of any peculiar motions.

2.2.3 The real-space correlation function

It is possible to estimate $\xi(r)$ by inverting $\Xi(\sigma)$ without making the assumption that it is a power law (Saunders, Rowan-Robinson & Lawrence 1992, hereafter S92). They recast Equation 2.11 into the form,

$$\xi(r) = -\frac{1}{\pi}\int_r^\infty \frac{(d/d\sigma)\Xi(\sigma)}{(\sigma^2 - r^2)^{\frac{1}{2}}} d\sigma \tag{2.13}$$

and assuming a step function for $\Xi(\sigma) = \Xi_i$ in bins centered on σ_i, and interpolating between values,

$$\xi(\sigma_i) = -\frac{1}{\pi}\sum_{j\geqslant i} \frac{\Xi_{j+1} - \Xi_j}{\sigma_{j+1} - \sigma_j} \ln\left(\frac{\sigma_{j+1} + \sqrt{\sigma_{j+1}^2 - \sigma_i^2}}{\sigma_j + \sqrt{\sigma_j^2 - \sigma_i^2}}\right) \tag{2.14}$$

for $r = \sigma_i$. Fitting this $\xi(r)$ with a power law yields the real-space clustering parameters and provides a way of checking the power law assumption. S92 claimed that their method was only good for scales $r \lesssim 30\ h^{-1}$Mpc in the QDOT survey because r became comparable to the scale out to which they estimated Ξ.

2.3 Modelling $\xi(\sigma, \pi)$

A far more powerful technique could be available by attempting to model all the effects simultaneously. The 2-d correlation function can be modelled by starting with a model of the real-space correlation function, $\xi(r)$, inputting the effects of the coherent infall, parameterized by β (Equation 1.16), and convolving this with the form of the random peculiar motions to produce the final model. The machinery for doing this is now described.

2.3.1 The coherent infall

Kaiser (1987) published an equation relating the redshift-space power spectrum to the real-space power spectrum,

$$|P_s(k)|^2 = (1 + \beta\mu_k^2)^2 |P_r(k)|^2 \tag{2.15}$$

where μ_k is the cosine of the angle between k and the line-of-sight. Hamilton (1992) transformed this equation from Fourier-space into real space,

$$\xi'(\sigma, \pi) = [1 + \beta(\partial/\partial z)^2(\nabla^2)^{-1}]^2 \xi(r) \tag{2.16}$$

which reduces to

$$\xi'(\sigma,\pi) = \xi_0(r)P_0(\mu) + \xi_2(r)P_2(\mu) + \xi_4(r)P_4(\mu) \tag{2.17}$$

where $\mathrm{P}_\ell(\mu)$ are Legendre polynomials, $\mu = \cos(\theta)$ and θ is the angle between r and π, and

$$\xi_0(r) = \left(1 + \frac{2\beta}{3} + \frac{\beta^2}{5}\right)\xi(r) \tag{2.18}$$

$$\xi_2(r) = \left(\frac{4\beta}{3} + \frac{4\beta^2}{7}\right)[\xi(r) - \overline{\xi}(r)] \tag{2.19}$$

$$\xi_4(r) = \frac{8\beta^2}{35}\left[\xi(r) + \frac{5}{2}\overline{\xi}(r) - \frac{7}{2}\overline{\overline{\xi}}(r)\right] \tag{2.20}$$

where

$$\overline{\xi}(r) = \frac{3}{r^3}\int_0^r \xi(r')r'^2 dr' \tag{2.21}$$

$$\overline{\overline{\xi}}(r) = \frac{5}{r^5}\int_0^r \xi(r')r'^4 dr' \tag{2.22}$$

In the case of a pure power law for the real-space correlation function, these equations are simple to integrate analytically,

$$\overline{\xi}(r) = \frac{3\xi(r)}{3-\gamma_r} \tag{2.23}$$

$$\overline{\overline{\xi}}(r) = \frac{5\xi(r)}{5-\gamma_r} \tag{2.24}$$

and hence

$$\xi_0(r) = \left(1 + \frac{2\beta}{3} + \frac{\beta^2}{5}\right)\xi(r) \tag{2.25}$$

$$\xi_2(r) = \left(\frac{4\beta}{3} + \frac{4\beta^2}{7}\right)\left(\frac{\gamma_r}{\gamma_r - 3}\right)\xi(r) \tag{2.26}$$

$$\xi_4(r) = \frac{8\beta^2}{35}\left(\frac{\gamma_r(2+\gamma_r)}{(3-\gamma_r)(5-\gamma_r)}\right)\xi(r) \tag{2.27}$$

In the case of non-power law forms for the correlation function the integrals in Equations 2.21 and 2.22 must be performed numerically.

2.3.2 The random peculiar velocities

$\xi'(\sigma,\pi)$ can then be convolved with a model for the random peculiar velocities,

$$\xi(\sigma,\pi) = \int_{-\infty}^{\infty} \xi'(\sigma, \pi - v/H_0)f(v)dv \tag{2.28}$$

to produce the final model $\xi(\sigma,\pi)$ (Peebles 1980), where the form of the random peculiar velocities is traditionally given by,

$$f(v) = \frac{1}{a\sqrt{2}} \exp\left(-\frac{\sqrt{2}|v|}{a}\right) \tag{2.29}$$

where a is the pairwise peculiar velocity dispersion (often known as σ_{12}).

2.3.3 Model limitations

In this model there are three main assumptions. Firstly, a form for the correlation function has to be assumed and a power law is normally chosen. This is a good fit at scales $\lesssim 30\ h^{-1}$Mpc but not so accurate at larger scales. This will impose an upper limit on the scales that can be probed using this method. Any other model for the correlation function would need the numerical integration of Equations 2.21 and 2.22. The second assumption is that this linear theory model will hold in the non-linear regime of small scales. A lower limit to the scales may need to be imposed to overcome this problem. The final assumption is the choice of an exponential distribution of peculiar velocities (Equation 2.29). This has been shown by various authors (e.g. Ratcliffe et al. 1998) to be a reasonable fit to the data. It is also discussed in Section 2.5, as well as being tested, and justified, on observed data in Chapter 5.

2.3.4 Testing the models

To show the effect of these distortions on the $\xi(\sigma,\pi)$ plot, Figure 2.1 shows four model $\xi(\sigma,\pi)$'s. If there were no redshift-space distortions then the contours would be circular, as shown in the top left panel due to the isotropy of the real-space correlation function. On the small σ scales the random peculiar velocities cause an elongation of the contours in the π direction (the bottom left panel). On larger scales there is the flattening of the contours (top right panel) due to the coherent infall. The bottom right panel is a model with both distortion effects included and both are clearly distinguishable. A comparison of the models to the data allows the four free parameters, r_0, γ_r, β and a to be determined simultaneously. This method is used on observed data in Chapter 5.

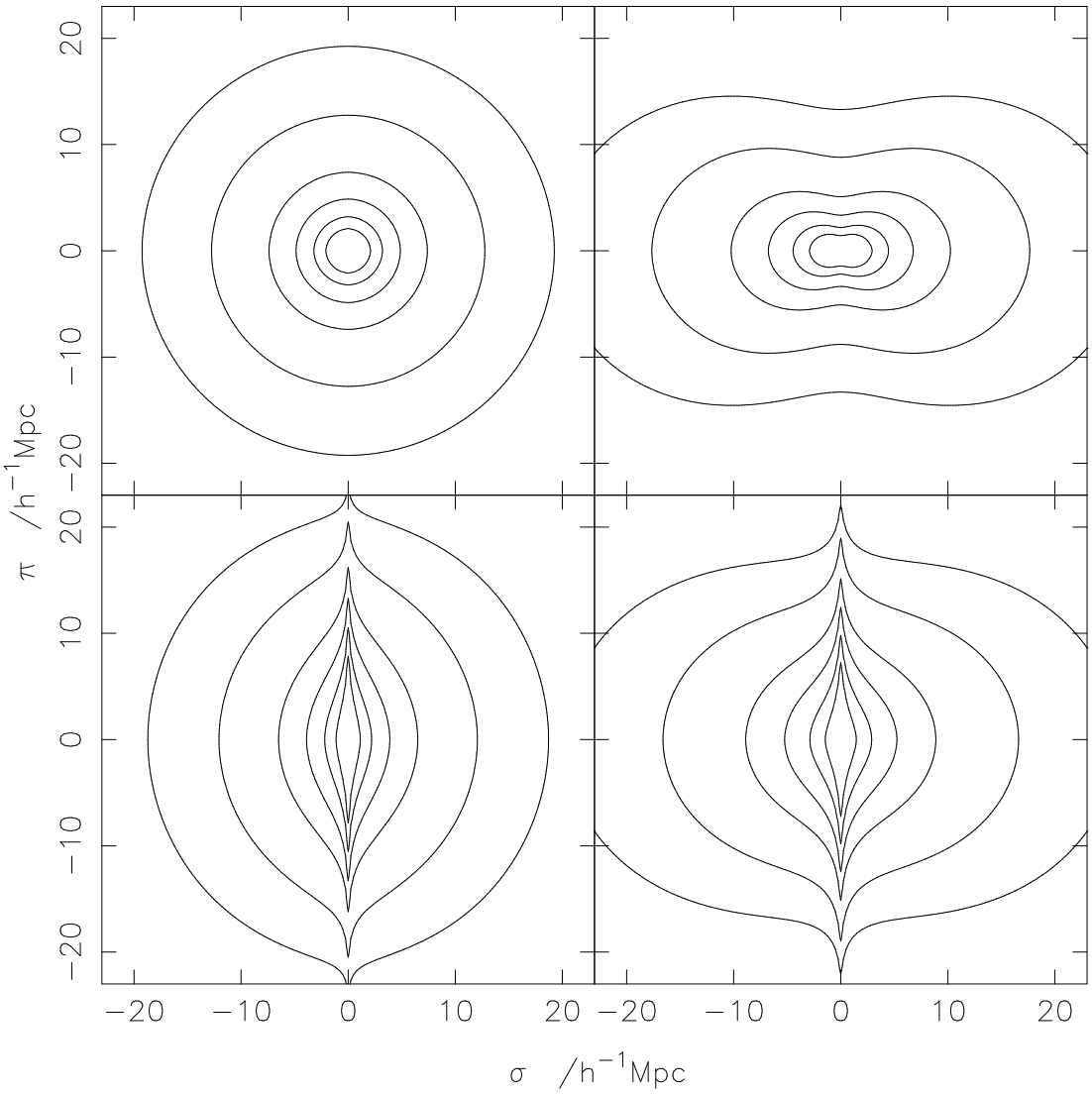


Figure 2.1: Plot of four different model $\xi(\sigma,\pi)$'s. The lines represent constant contours of ξ = 4.0, 2.0, 1.0, 0.5, 0.2 and 0.1 for different models. The top left figure represents an undistorted correlation function ($a = 0,\ \beta = 0$), the top right figure is a model with coherent infall added ($a = 0,\ \beta = 0.4$), the bottom left figure is a model with just random peculiar velocities added ($a = 500\,\mathrm{km\,s^{-1}},\ \beta = 0$) and the bottom right figure has both infall and random motions added ($a = 500\,\mathrm{km\,s^{-1}},\ \beta = 0.4$). All of these models have a power law $\xi(r)$ with $r_0 = 5.0\ h^{-1}$Mpc and $\gamma_r = 1.7$.

2.4 Estimating β

The models just described provide a new method of estimating β, but previous surveys have lacked the signal-to-noise to perform such a fit. They have used simpler methods to estimate the coherent infall into clusters (parameterized by β) which provides information about the total mass density in the Universe (Equation 1.12). Because we are measuring properties of the galaxy distribution the bias parameter must be taken into account. The definition of the bias, b (defined in Equation 1.15), leads to

$$\xi(r)_{\rm galaxies} = b^2 \xi(r)_{\rm mass} \tag{2.30}$$

Some recent results (Verde et al. 2002; Lahav et al. 2002) suggest that this linear bias scheme is consistent with observations.

A measurement of β is then an important step in measuring the matter density of the Universe. There are two common ways of estimating β from the correlation functions and these are now described.

2.4.1 Method 1: ratio of ξ's

The ratio of $\xi(s)$ to $\xi(r)$ in the linear regime gives an estimate of β through the equation (Hamilton 1992, cf. Equation 2.25)

$$\frac{\xi(s)}{\xi(r)} = 1 + \frac{2\beta}{3} + \frac{\beta^2}{5}. \tag{2.31}$$

A high signal to noise ratio is needed to make this a sensitive measure of β as the errors on the data become relatively large in the range of scales ($\gtrsim 8\ h^{-1}$Mpc) where the equation is valid.

2.4.2 Method 2: Q

Hamilton (1992) used the quadrupole moment of the correlation function as a measure of β (cf. Equations 2.18 and 2.19),

$$Q = \frac{\frac{4}{3}\beta + \frac{4}{7}\beta^2}{1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2} = \frac{\xi_2(s)}{\frac{3}{s^3}\int_0^s \xi_0(s') s'^2 ds' - \xi_0(s)} \tag{2.32}$$

where (with $P_\ell(\mu)$ and μ defined as in Equation 2.17)

$$\xi_\ell(s) = \frac{2\ell+1}{2} \int_{-1}^{+1} \xi(\sigma,\pi) P_\ell(\mu) d\mu \tag{2.33}$$

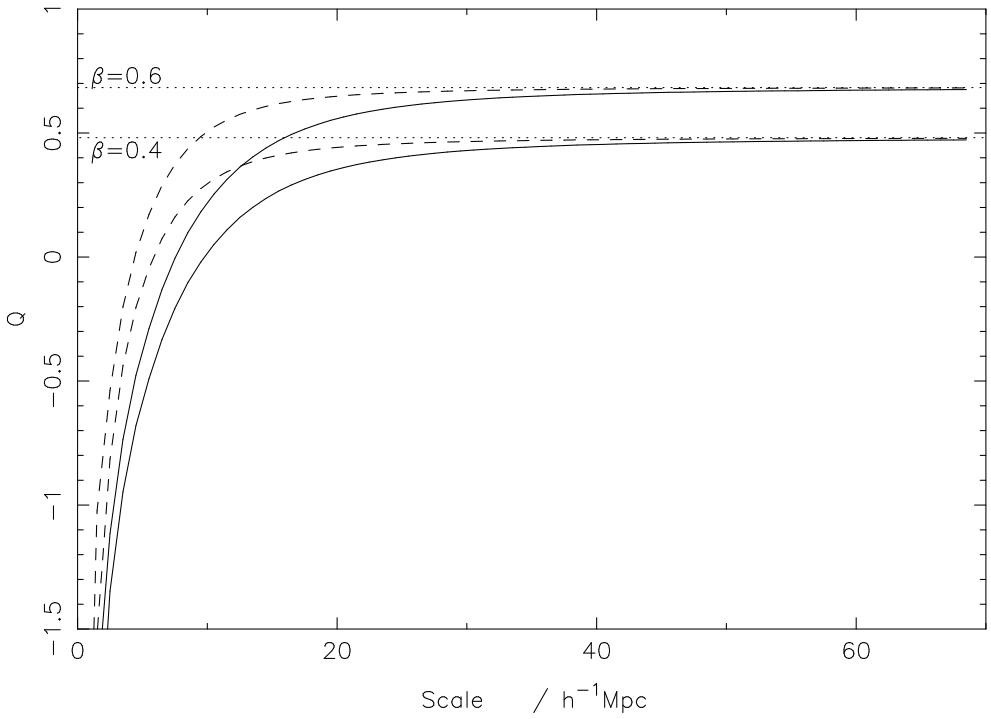


Figure 2.2: The Q parameter for four model $\xi(\sigma,\pi)$'s. The solid lines are for models with $a = 500$ km s^{-1} and the dashed lines are for $a = 300$ km s^{-1}. There are two sets of these lines for $\beta = 0.4$ and $\beta = 0.6$, denoted by the dotted lines.

and so Q is a constant if the random peculiar velocities are ignored. The models described in Section 2.3 are subjected to this method and the results are shown in Figure 2.2. Two models with $\beta = 0.4$ and $\beta = 0.6$ (corresponding to the dotted lines), are used, each with velocity dispersions of $a = 300$ km s^{-1} and $a = 500$ km s^{-1}. It can be seen that on small scales, where the signal is highest, the random peculiar velocities cause Q to go negative. On large scales the value of Q in the models asymptotes towards the true value, but even at scales $>$ 40 h^{-1}Mpc the estimate is biased slightly low, especially if the velocity dispersion is high (solid lines). Thus a direct estimate of β is possible but there will be a slight bias, though this bias is likely to be smaller than the random errors, which become quite large where the method is valid. A two parameter fit to all scales could be performed to overcome these problems.

2.5 Random peculiar velocities

When attempting to estimate $f(v)$ (Equation 2.29) previous authors (e.g. Davis & Peebles 1983; Jing, Mo & Börner 1998) have assumed a simple one-dimensional form for the effects of coherent infall velocities, rather than using β, and the results depend on the precise choice of model. As will be shown later, the infall velocities have a large effect on measurements of $f(v)$.

Landy, Szalay & Broadhurst (1998, hereafter LSB98) presented a method for recovering the form of the random peculiar velocities. Landy (2002, hereafter L02) used this method on the 100k 2dFGRS Public Release data. To extract the peculiar velocity distribution you need to deconvolve the real-space correlation function from the redshift-space correlation function.

2.5.1 The method

Firstly, you take the 2-d Fourier transform of the $\xi(\sigma,\pi)$ grid[1] to give $\hat{\xi}(k_\sigma,k_\pi)$ and then take cuts along the k_σ and k_π axes which are denoted by $\Sigma(k)$ and $\Pi(k)$ respectively, so $\Sigma(k)=\hat{\xi}(k_\sigma=k,k_\pi=0)$ and $\Pi(k)=\hat{\xi}(k_\sigma=0,k_\pi=k)$. These cuts are equivalent to the Fourier transforms of the real-space projections of $\xi(\sigma,\pi)$ onto the σ and π axes (by the so-called slicing-projection theorem). The projection of $\xi(\sigma,\pi)$ onto the σ axis is a distortion free measurement of the correlation function, but the projection onto the π axis gives us the real-space correlation function convolved with the peculiar velocity distribution (ignoring for now the effects of large-scale coherent infall). Since a convolution in real space is a multiplication in Fourier-space, the ratio of $\Sigma(k)$ to $\Pi(k)$ is the Fourier transform of the velocity distribution that we want to estimate. All that is left is to inverse Fourier transform this ratio to give us the peculiar velocity distribution.

2.5.2 Testing the models

The effectiveness of this method can be tested on our model $\xi(\sigma,\pi)$'s by (hopefully) recovering the input peculiar velocity distribution (Equation 2.29). Because the resulting $f(v)$ distribution comes from integrating over the whole grid, fitting

[1]This version of the method differs slightly from that described in LSB98 and L02, who cut the data at 32 h^{-1}Mpc and smoothed it with a Hann window, before performing the Fourier transform. We choose to use all the raw data, and the reason for this will become clear.

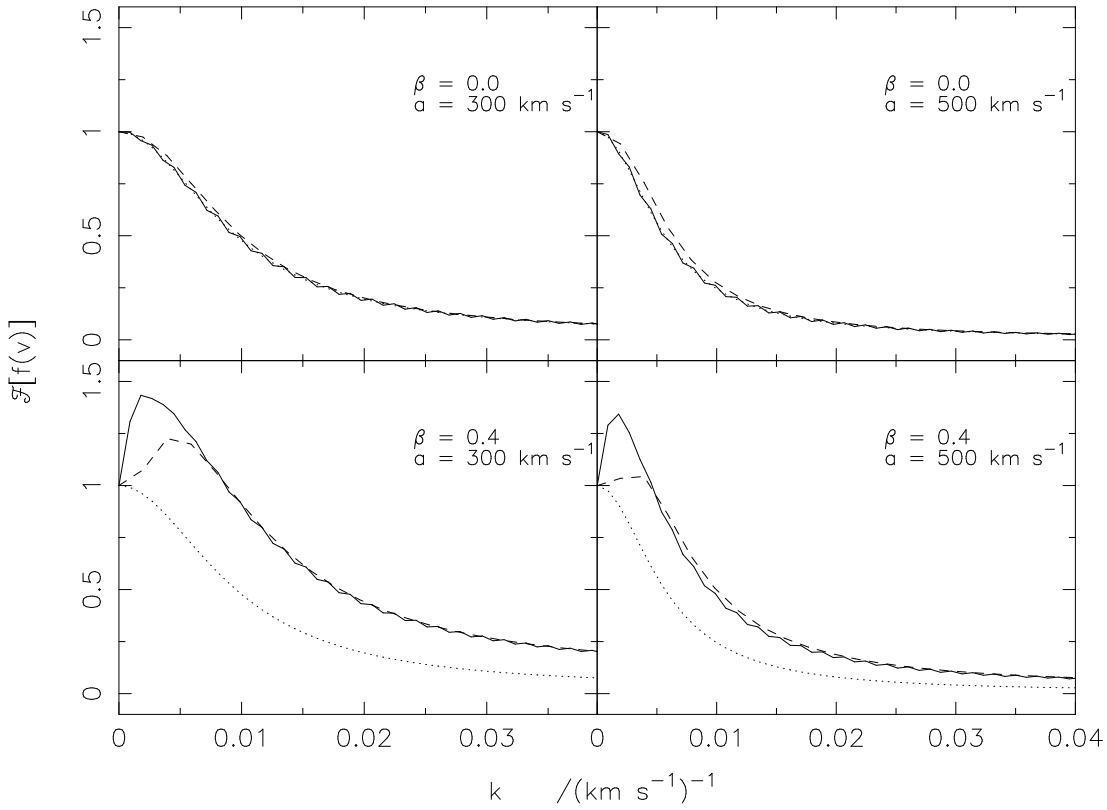


Figure 2.3: The Fourier transform of the peculiar velocity distribution for various models (see labels). The solid line is for a model with no smoothing and using all scales $< 70\ h^{-1}$Mpc. The dashed line is for a model cut at $32\ h^{-1}$Mpc and smoothed with a Hann window. The dotted line is the Lorentzian equivalent of the input exponential peculiar velocity distribution (this coincides with the solid line in the top panels).

an exponential to the resulting curve gives a value for a assuming that the infall contribution to the velocity distribution is negligible.

The models described in Section 2.3 are tested with and without a $\beta = 0.4$ infall factor and using the LSB98 smoothing method and our non-smoothing method. In Figure 2.3 the resulting Fourier transforms of the peculiar velocity distribution are shown and in Figure 2.4 the resulting distribution functions are shown. The solid line in each plot is with no smoothing using scales out to $70\ h^{-1}$Mpc. The dashed line is a model cut at $32\ h^{-1}$Mpc and smoothed with a Hann window (as in LSB98, L02). The dotted line is the assumed exponential velocity distribution.

LSB98 claim that their method is not sensitive to the infall velocity. Comparing the top panels with the bottom panels in Figure 2.3 for low k clearly shows that

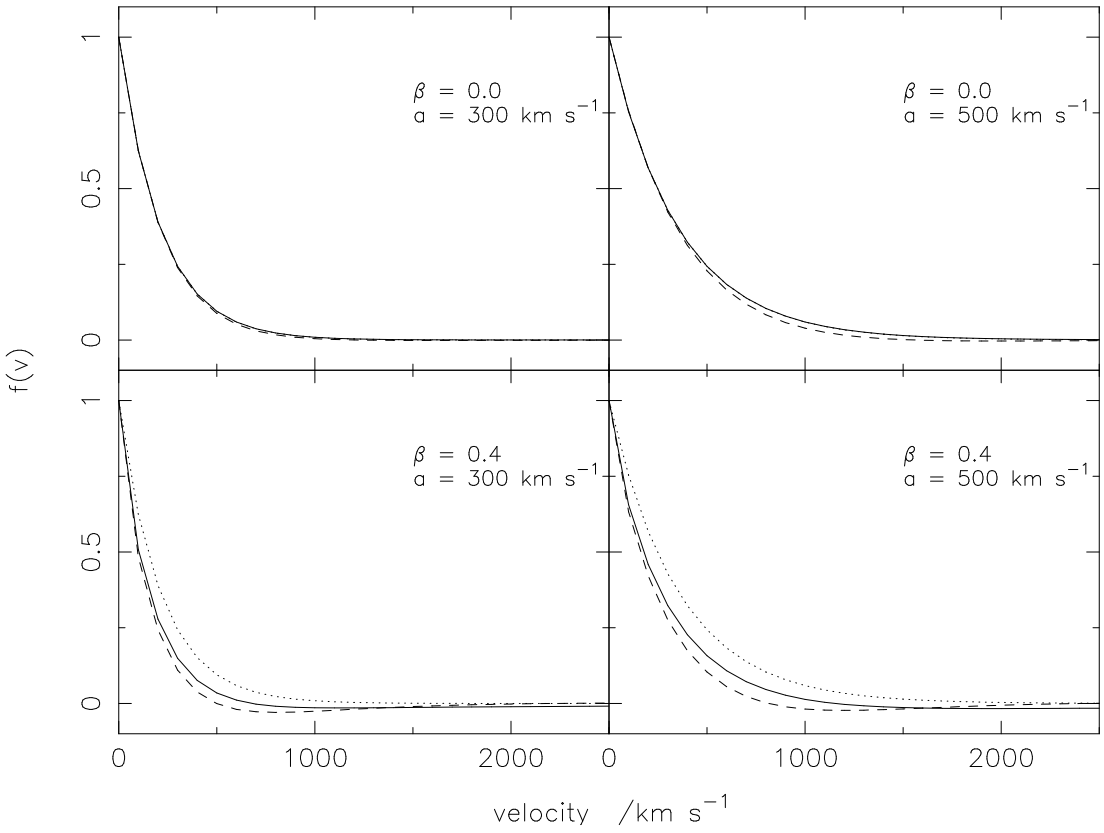


Figure 2.4: The peculiar velocity distribution for various models (see labels). The solid line is for a model with no smoothing and using all scales $< 70\ h^{-1}$Mpc. The dashed line is for a model cut at 32 h^{-1}Mpc and smoothed with a Hann window. The dotted line is the input exponential peculiar velocity distribution (this coincides with the solid line in the top panels).

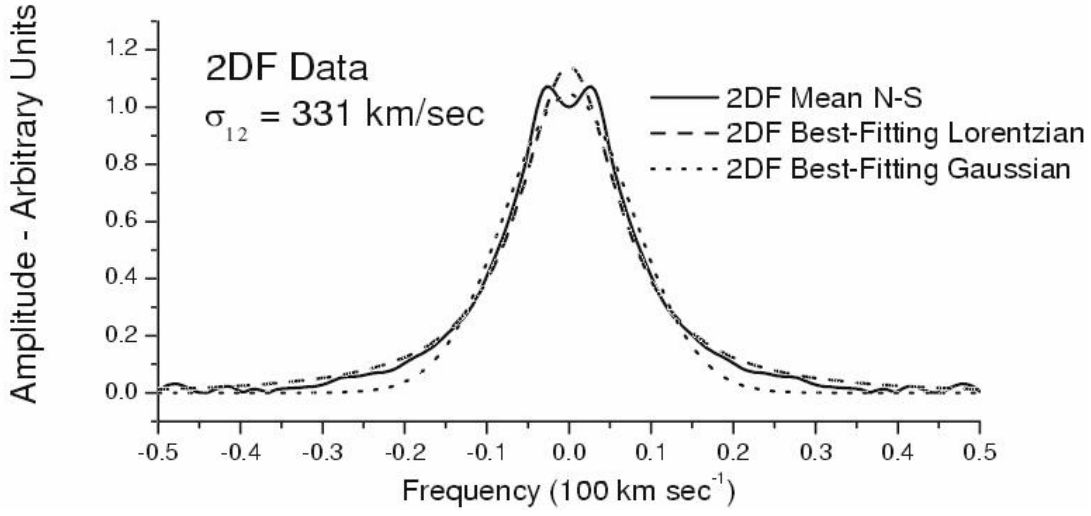


Figure 2.5: From L02: The discovery of additional structure (solid line) in the Fourier Transform of the peculiar velocity distribution.

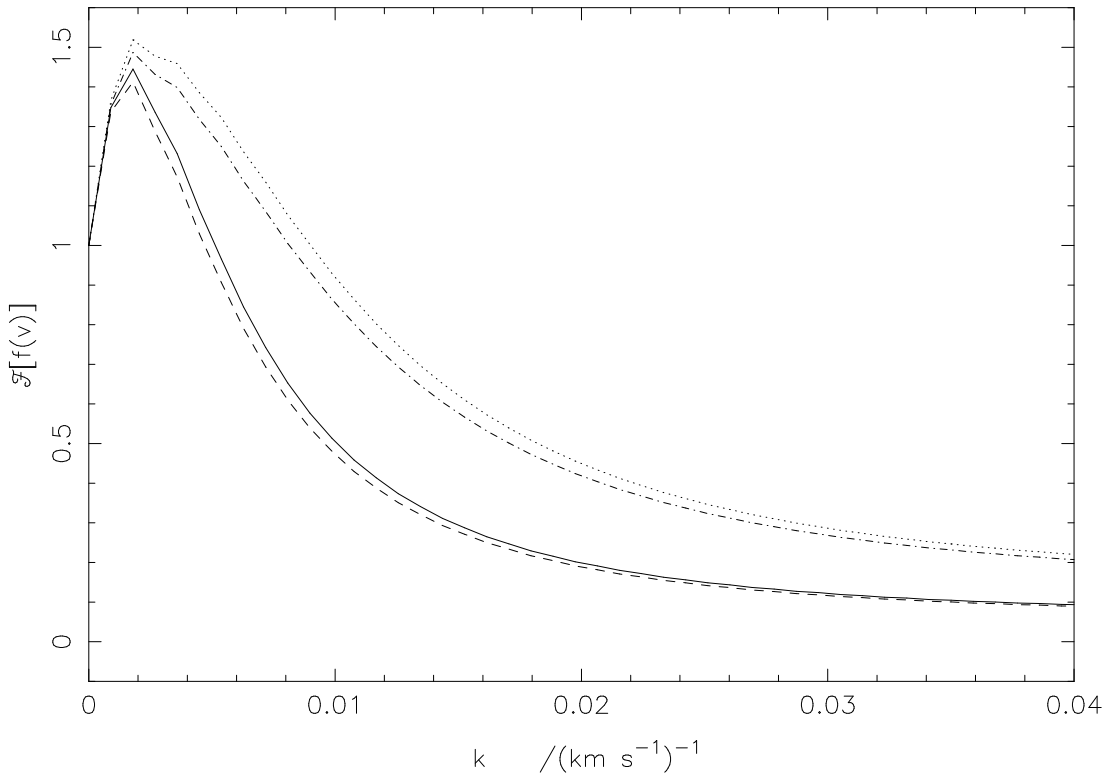


Figure 2.6: The Fourier transform of the peculiar velocity distribution for a model with $a = 500$ km s^{-1} (dashed line), $a = 300$ km s^{-1} (dotted line) and a model with a decreasing linearly from 500 km s^{-1} to 300 km s^{-1} from $\sigma = 0$ to $\sigma = 20$ h^{-1}Mpc (solid line) and a model with a increasing linearly from 300 km s^{-1} to 500 km s^{-1} from $\sigma = 0$ to $\sigma = 20$ h^{-1}Mpc (dot-dashed line). $\beta = 0.5$ for all three models.

this is not the case. The additional structure in the Fourier transform of the velocity distribution found in the models (and in the 2dFGRS data by L02 - see Figure 2.5) is a direct consequence of the infall velocities, parameterized by β.

Figure 2.4 also shows the effect of β on the peculiar velocity dispersion measurement. A fit to the resulting velocity distribution will be biased low if β is ignored (and this is found to be the case in Section 5.4.2).

There are other possible sources of error in the LSB98 method which may bias the method, namely only using scales < 32 h^{-1}Mpc and smoothing the data with a Hann window in the analysis (as described in LSB98). Figure 2.4 shows the effects of these assumptions and the difference between the solid and dashed lines in Figure 2.4 shows the result is biased.

What the models show is that this method is clearly affected by the infall velocities

and the measured peculiar velocity dispersions are biased low, especially in the case where a smoothing window is used and only using limited scales. Both types of peculiar velocity must be considered when analysing the results of this method.

2.5.3 Why an exponential distribution?

The choice of an exponential distribution for the random peculiar velocities may seem a bit arbitrary but there have been theoretical works which show that there are reasonable physical arguments for such a distribution.

Sheth (1996) used an extended Press-Schechter model to show that an exponential pairwise peculiar velocity distribution is easily explained by the Press-Schechter (Press & Schechter 1974) theory of non-linear clustering and is fairly independent of power-spectrum and the assumed density profile in clusters.

Diaferio & Geller (1996) also found that an exponential pairwise peculiar velocity distribution is a result of well-defined gravitational processes. They concluded that the exponential pairwise peculiar velocity distribution occurs because of the presence of a large fraction of unrelaxed galaxy systems in the universe.

It has been suggested that the pairwise velocity dispersion, a, may vary as a function of the projected separation, σ. Figure 2.6 compares models with a constant a to models with a varying a. It is clear that the value of a at small σ is the main influence on the shape of $\mathcal{F}[f(v)]$. This is considered for observed data in Chapter 5.

2.6 Uncertainty estimates

Measurements mean very little without the errors associated with them. Obtaining reliable error estimates is crucial to the understanding and interpretation of the results, allowing useful and valid conclusions to be made.

2.6.1 Correlated errors

The statistics that are measured in galaxy clustering analyses are all based on pair counting. This causes large correlations in the results because the same galaxy affects the measurements over all scales, and so each data point is not

independent. This must be taken into account when estimating parameters of interest from the data, and this can be done using mock catalogues.

2.6.2 Mock catalogues

Throughout this work mock redshift surveys are used to determine errors on the real redshift surveys. These catalogues are extracted from N-body simulations and are analysed as if they were real. Their exact construction is described in later chapters where they are used.

The idea of using mock catalogues, which have similar clustering properties to the data, is that they mimic the possible variance in the real data and hence the spread of results from the mock catalogues gives a good estimate of the errors on the real data. The bootstrap or jack-knife methods used by some authors do a good job as long as the volume sampled is representative, but cannot take cosmic variance into account.

We analyse N mock catalogues in the same way as the real data, so that we have N mock measurements for every measurement made on the real data. The standard deviation between the N mock measurements gives a robust estimate of the uncertainty on the real data. This approach is used to estimate the uncertainties for direct measurements from the data, such as the individual points in the correlation function, and for best-fit parameters such as s_0.

When fitting parameters this standard deviation is used as a weight for each data point and a minimum χ^2 analysis is performed to obtain the best-fit parameter. The errors that are quoted for any particular parameter are the rms spread between the N best fit parameters obtained in the same way from the mock catalogues, and not the error from the likelihood analysis. This simple way of estimating the uncertainties avoids the complications of dealing directly with correlated errors in measured data points, while still providing an unbiased estimate of the real uncertainties in the data, including the effects of correlated errors. Although this approach gives reliable estimates of the uncertainties, the simple weighting scheme is not necessarily optimal in the presence of correlated errors. Nevertheless, for all statistics that are considered, it is found that the means of the mock estimates agree well with the values input to the parent simulations. We are confident that our measurements and uncertainty estimates are robust and unbiased.

Chapter 3

Clustering in the PSCz survey

Much of the work in this chapter was presented in the paper Hawkins et al. (2001).

3.1 Motivation

The relation between the distribution of galaxies and the distribution of mass is now one of the most important problems in large-scale structure. Empirically, it has been found that different galaxy types have different clustering amplitudes and hence that galaxies are biased tracers of the mass distribution. Evidently, we need to understand the physical mechanisms responsible for these biases if we are to establish the connection between galaxy tracers and the mass. This problem can be understood in terms of two related questions: (i) How does the clustering of a galaxy sample depend on the properties of the galaxies? and (ii) How do the properties of galaxies depend on their local environment?

Many ideas have been put forward as to how the environment may modify galaxy properties, from schematic ideas concerning feedback in the formation process (e.g. White & Rees 1978; Dekel & Rees 1987) to hydro-dynamical simulations of galaxy formation incorporating cooling and dissipation (e.g. Katz & Gunn 1991; Navarro & White 1993; Pearce et al. 1999). Over the next few years we can expect major advances in hydrodynamic simulations employing parallel computers and there is a real prospect of understanding the relationship between galaxy morphologies and environment within the context of specific theories of cosmological fluctuations. We can hope that such numerical simulations will eventually be able to provide detailed predictions of the spatial distribution of

galaxies as a function of their stellar content, rotation velocity, star-formation rate, and morphological type.

The analysis of the PSCz survey, based on an infra-red catalogue, will allow comparisons with studies of optically selected surveys and provide insights into how the galaxies are related to the mass distribution.

3.2 The PSC redshift survey

3.2.1 Description

Galaxies for the PSC Redshift (PSCz) Survey were selected from the Infra-Red Astronomical Satellite (*IRAS*) Point Source Catalogue (PSC; Beichmann, Helou & Walker 1988) by positional, identificational flux and colour criteria designed to accept almost all galaxies while keeping contamination by Galactic sources to an acceptable level. All sources with a $60\mu m$ flux > 0.6Jy were included.

Previous *IRAS* redshift surveys (e.g. The 1.2Jy survey, QDOT) had not included all the known galaxies. The QDOT survey was a random 1-in-6 sampling of the PSC catalogue, while the 1.2Jy survey was complete down to a flux of 1.2Jy. Thus the PSCz is the largest 'all-sky' redshift survey to date.

The significant differences compared with the QDOT survey (Lawrence et al. 1999), were that the colour criteria were relaxed to avoid excluding previously uncatalogued galaxies with unusual (especially cool) colours and the sky coverage is increased to 84% by including areas of high cirrus contamination. These changes increase the contamination of the galaxy catalogue from local galactic sources (stars, planetary nebulae, cirrus etc) but ensure the sample has a higher completeness. The contaminating Galactic sources were excluded by a combination of *IRAS* and optical properties (from sky survey plates), and where necessary spectroscopy. The Behind The Plane (BTP; Saunders et al. 2000a) project aims to increase the sky coverage by including areas of high galactic extinction around the Galactic Plane.

Redshifts were taken from all available published or unpublished sources; principally Huchra's ZCAT, the LEDA database, the 1.2Jy and QDOT surveys. Also 4500 new redshifts were taken on the Isaac Newton Telescope, the Anglo-Australian Telescope, the Cerro-Tololo 1.5-m and Cananea 2.1-m telescopes.

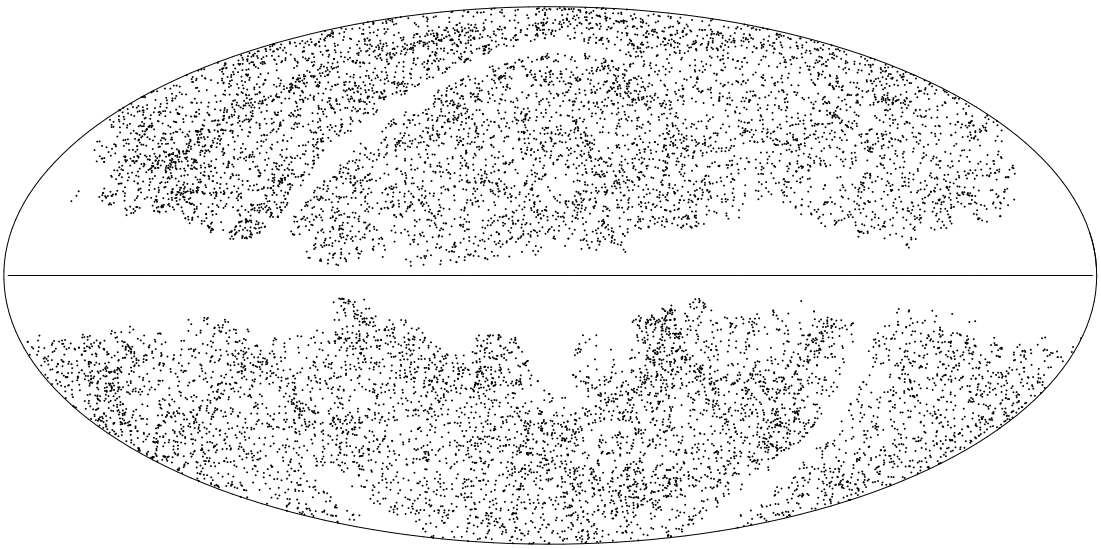


Figure 3.1: The sky distribution of the PSCz survey galaxies. The masked areas around the galactic equator (horizontal line) are clearly visible.

In total the catalogue contains 14 677 redshifts and the median redshift $z_{\rm med} \approx 0.028$. The distribution of galaxies on the sky is shown in Figure 3.1, where the horizontal line is the galactic equator and the masked regions due to galactic extinction are clearly visible. The two curved arcs masked out are due to incompleteness in the *IRAS* data. The uniformity and completeness of the catalogue are discussed by Saunders et al. (2000b).

The redshift distribution of the galaxies is shown in Figure 3.2. It can be seen that the peak of the distribution is near $z = 0.02$, and there is a tail which extends out to beyond $z = 0.1$. This tail means that there are over 1200 galaxies in the survey with a redshift > 0.08. For this analysis only galaxies with a redshift > 0.004, corresponding to a recession velocity of 1200 km s^{-1}, are considered so that local effects and the velocity uncertainties of around 120 km s^{-1} were less significant. The highest redshift is $z_{\rm max} = 0.17$. We also assume a $\Omega_{\rm m} = 1$ geometry to convert between redshift and distance. At the redshifts considered, this choice is unlikely to make much difference to the results.

3.2.2 PSCz mock catalogues

Large N-body simulations were used to create mock PSCz surveys (Cole et al. 1998). The simulations used the AP^3M code of Couchman (1991) loaded with

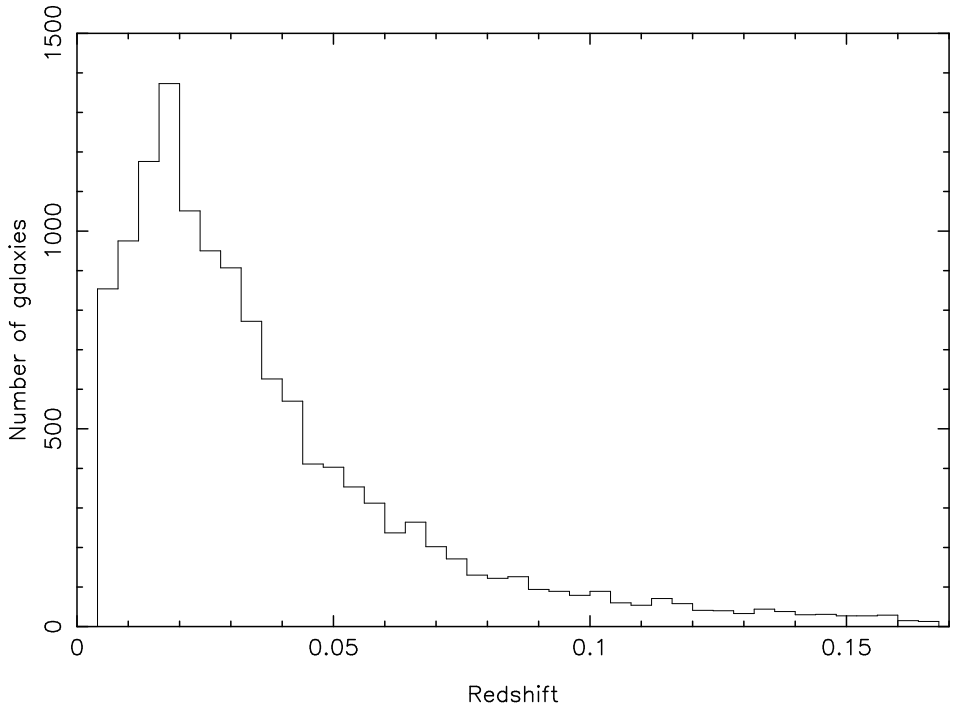


Figure 3.2: The redshift distribution for the complete PSCz survey.

192^3 particles of mass $1.64 \times 10^{12}\Omega_{\rm m}\ h^{-1}M_{\odot}$ in a box with a co-moving size of 345.6 h^{-1}Mpc. More details can be found in Cole et al. (1998). For this analysis three different cosmological models have been considered. They are a spatially flat low density universe model (ΛCDM) with $\Omega_{\rm m} = 0.3$, $\Omega_{\Lambda} = 0.7$ and $\Gamma = 0.25$, a spatially flat universe model (SCDM$\Gamma_{0.25}$) with $\Omega_{\rm m} = 1.0$ and $\Gamma = 0.25$, and finally a spatially flat universe model (SCDM$_{COBE}$) with $\Omega_{\rm m} = 1.0$ and $\Gamma = 0.5$, normalised to match the amplitude of the *COBE* data. For pure CDM-type models the shape parameter, $\Gamma = \Omega h$.

The galaxy catalogues were extracted from the numerical simulations by first identifying a population of objects with similar properties to the Local Group (LG). Then an observer is placed by considering the observational constraints of peculiar velocity, local over-density and shear. All points within a sphere of 120 h^{-1}Mpc radius around the observer are included, and the frame is rotated so that the motion of the observer is the same as the LG peculiar velocity. Note that this sphere radius is less than the size of the PSCz survey. A luminosity was randomly assigned to each galaxy in the mock realizations to mimic the corresponding observational properties. The PSCz density is reproduced by us-

ing the PSCz selection function to reject points and also to randomly assign a flux to each point. Unsurveyed regions are rejected using the PSCz mask to give the same sky coverage as the real catalogue. Each mass point is associated with a galaxy so the linear bias parameter, $b = 1.0$. Lastly the bias against early-type galaxies in the original *IRAS* survey is introduced by rejecting an appropriate fraction of galaxies in the cores of clusters, which is meant to reproduce Dressler's morphology-density relation (Dressler 1980). The resulting mock surveys have approximately the same correlation function, redshift distribution and flux distribution as the real PSCz sample, though we quote all our errors from the ΛCDM mock catalogues as these are the closest to the real data.

3.2.3 Considerations when estimating ξ

Several methods were investigated for generating the redshift distribution for the random catalogues: random shuffling of the positions on the sky relative to the redshifts; random shuffling followed by adding a further random velocity from a Gaussian distribution with $\sigma = 500, 1000$ or 1500 km s^{-1}; and generating a random redshift distribution to match an analytic fit to the selection function (Saunders et al. 2000b). For the full catalogue, random shuffling with 1000 km s^{-1} smoothing and the analytic fit gave indistinguishable results for ξ. For the random catalogue, 12 times the number of random points as real galaxies were generated.

To calculate the correlation function, $\xi(s)$ was first calculated using the Landy-Szalay estimator (Equation 2.8), with a standard pair weighting of the sample (Equation 2.4 with $J_3 = 0$) and a power law was fitted to the results for $1 < s < 10$ h^{-1}Mpc. The correlation function was then recalculated using the J_3 calculated from this best fit power law, truncated so that the maximum J_3 is 1500. It was found that only one iteration was needed to produce a stable result with a consistent J_3 and ξ. The calculation of $\xi(s)$ is relatively insensitive to the precise form of the J_3 weighting employed.

The correlation function for the 10 realizations of each mock were calculated in the same way and the standard deviation about the mean of s_0, r_0, γ_s and γ_r was used to estimate the uncertainties in these measurements for the real sample.

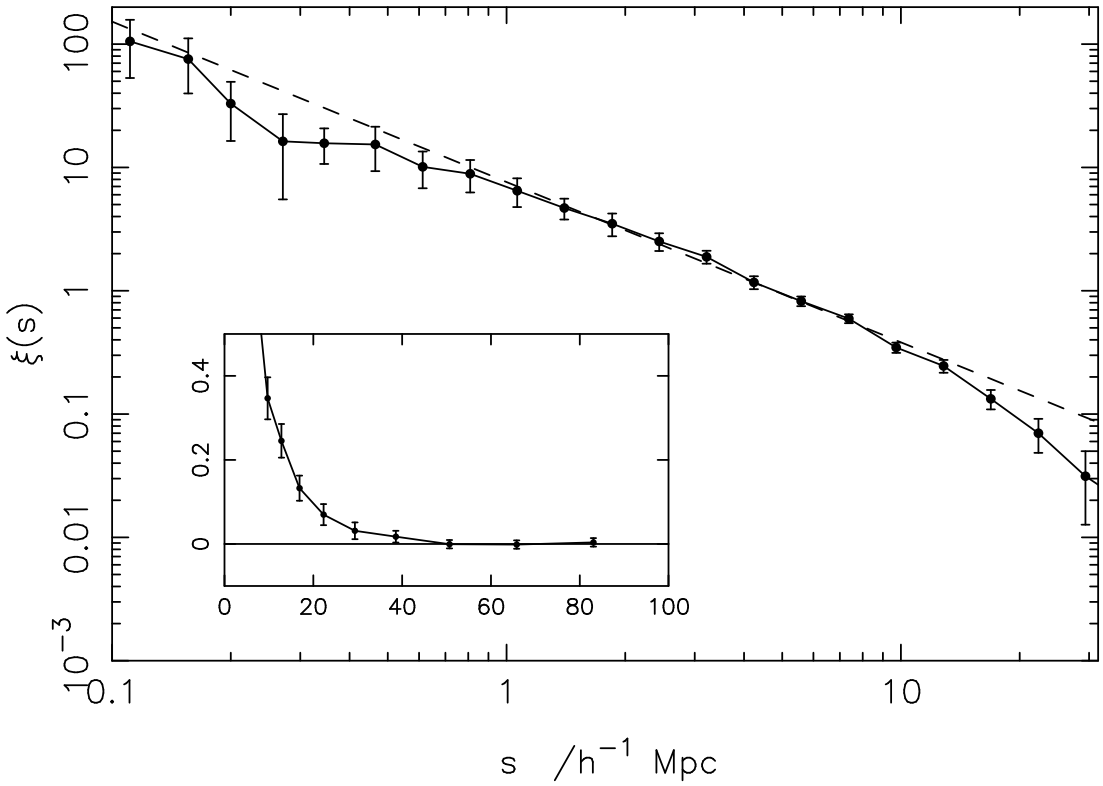


Figure 3.3: $\xi(s)$ for the full PSCz survey (solid line) with error bars from the rms scatter in mock catalogue results. The dashed line is the best-fit power-law on scales $1 < s < 10\ h^{-1}$Mpc. The inset is plotted on a linear scale.

3.3 Results

3.3.1 The redshift-space correlation function

The estimate of $\xi(s)$ for the full sample is shown in Figure 3.3. The power law fit gives $s_0 = 4.77 \pm 0.20$ and $\gamma_s = 1.30 \pm 0.04$. The errors are estimated from the standard deviation between the ten realizations of the ΛCDM mock catalogues. The value of J_3 was calculated and it was found to be 1500 ± 400 at a scale of $75\ h^{-1}$Mpc. The error on the J_3 result is from the standard deviation of the mock realization results.

The estimate of $\xi(s)$ becomes consistent with zero at $\sim 50\ h^{-1}$Mpc, and this can be taken as the scale at which the Universe looks homogeneous for infra-red selected galaxies. Pan & Coles (2000) also found that the PSCz survey becomes homogeneous on these scales from a fractal analysis.

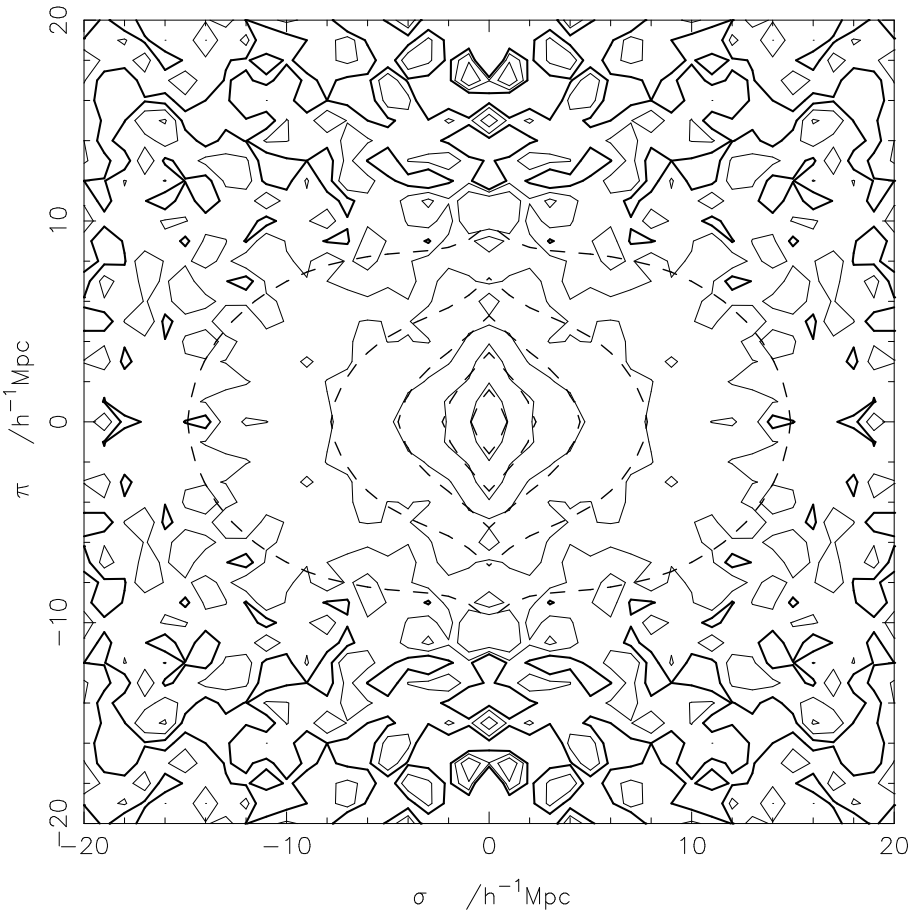


Figure 3.4: $\xi(\sigma, \pi)$ for the full PSCz survey out to 20 h^{-1}Mpc (solid lines), and a model (dashed lines). Contours at $\xi = 4.0$, 2.0, 1.0, 0.5, 0.2 and 0.0 (thick line).

3.3.2 $\xi(\sigma, \pi)$

The 2-d correlation function is plotted in Figure 3.4 and it can be seen that the two redshift-space distortion effects described in Section 2.2 are visible. The models described in Section 2.3, are used to fit to the data, but the signal-to-noise means the results are not very accurate. Overlaid in Figure 3.4 are contours of a best-fit model with $\beta = 0.66$, $a = 400$ km s^{-1}, $r_0 = 3.98$ and $\gamma_r = 1.84$, which is a reasonable match to the data. The method is discussed further in Section 5.5.1.

Many previous authors have estimated β from the PSCz data and estimates range from 0.39 (Taylor et al. 2001) to 0.75 (Rowan-Robinson et al. 2000), though most lie at the lower end of this range. Jing, Börner & Suto (2002) estimated $a \approx 400$ km s^{-1}, so these results are in broad agreement with other studies.

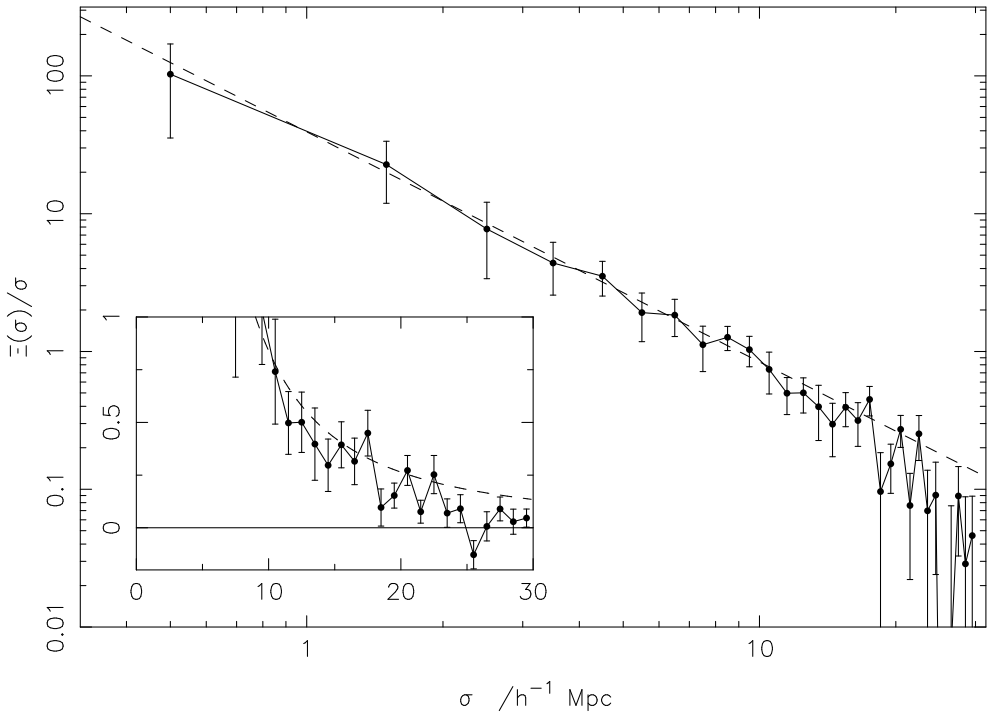


Figure 3.5: The projected correlation function for the full PSCz survey (solid line). The dashed line is the best-fit power-law on scales $1 < \sigma < 10\ h^{-1}$Mpc. Error bars are from the rms spread in ΛCDM mock catalogue results.

3.3.3 The projected correlation function, $\Xi(\sigma)$

The projected correlation function (see Section 2.2.2) estimate for the PSCz survey is plotted in Figure 3.5. The dashed line shows the best fit power law for $1 < \sigma < 10\ h^{-1}$Mpc with parameters $r_0 = 3.82 \pm 0.52$ and $\gamma_r = 1.67 \pm 0.09$, where the errors are the rms spread in mock catalogue results. This estimate of these parameters agrees well with the results of Jing, Börner & Suto (2002) who found that $r_0 = 3.70$ and $\gamma_r = 1.69$. The $\xi(r)$ obtained from an inversion (S92, Section 2.2.3) of $\Xi(\sigma)$ is very noisy and is not shown here.

3.3.4 Comparisons

The value of $s_0 = 4.77 \pm 0.20$ that has been measured is significantly less than that found for the similarly sized, but optically (R-band) selected, Las Campanas Redshift Survey ($s_0 = 6.28 \pm 0.27$, Tucker et al. 1997). The premise that infra-red galaxies tend to avoid the centres of clusters is borne out with this observation.

Chapter 4

PSCz Clustering by Galaxy Type

The work in this chapter was presented in the paper Hawkins et al. (2001).

4.1 Motivation

The simplest bias schemes are those based on the statistics of high density peaks (e.g. Kaiser 1984; Mo & White 1996). Since the optical magnitude of a galaxy is correlated to the depth of the potential via the Tully-Fisher relation for spirals and $D_n - \sigma$ relation for ellipticals, this would suggest that brighter galaxies will be more strongly correlated than fainter galaxies.

Realistic galaxy formation models rely on feedback mechanisms to limit star-formation, either acting internally within each galaxy or involving interactions with other galaxies and the intergalactic medium. So galaxies in regions of high local galaxy density are likely to have a reduced star-formation rate (SFR). Indeed this is confirmed by the observation that later-type optical galaxies, which have higher SFRs, have a much lower clustering amplitude than earlier galaxy types (e.g. Rosenberg, Salzer & Moody 1994; Loveday et al. 1995; Loveday, Tresse & Maddox 1999).

Galaxies with a high SFR tend to be bright in the far infra-red (FIR) because of the thermal emission from dust heated by young stars. Thus the observations that *IRAS* galaxies (selected on their $60\mu m$ flux) tend to avoid rich galaxy clusters, and have a lower clustering amplitude are consistent with this general picture. In this chapter we divide the PSCz sample of *IRAS* galaxies into subsamples based on their FIR colour. Since the galaxies with warmer FIR colours are those

with higher SFR, we can further test the dependence on star-formation rate. Mann, Saunders & Taylor (1996) selected warm and cool sub-samples of *IRAS* galaxies from the QDOT survey based on their $60\mu m/100\mu m$ flux ratio, and found marginal evidence that warmer galaxies are more strongly clustered than cooler galaxies, the opposite to what is expected from the morphology density relation. However they did not consider the result very significant compared to the expected cosmic variance.

Using optical galaxy samples, a number of authors have found indications of changes of the clustering amplitude with galaxy luminosity, (e.g. Valls-Gabaud, Alimi & Blanchard 1989; Park et al. 1994; Moore et al. 1994; Loveday et al. 1995; Benoist et al. 1996) but the samples are small and the observed amplitude shifts do not have a high statistical significance. The expected luminosity dependence for *IRAS* galaxy samples is less clear, since the FIR luminosity of a galaxy is likely to depend on both the mass of its dark halo and its specific SFR. If the FIR luminosity is correlated to the halo mass, then brighter galaxies should have a higher clustering amplitude. On the other hand, a high SFR will make a galaxy more luminous, and galaxies with a high star-formation rate tend to have weaker clustering. It is not easy to predict which of these is the dominant effect. Observationally Szapudi et al. (2000) and Beisbart & Kerscher (2000) have used mark correlation functions with volume limited sub-samples to examine the luminosity dependence of clustering in the PSCz survey. Over the narrow range of luminosities they consider they concluded that there is no significant luminosity dependence.

The analysis described in the previous chapter is now repeated on various sub-samples of the PSCz survey with the aim of answering some of these questions.

4.2 The samples

4.2.1 Selection function

Rather than calculating a best fit selection function for each sub-sample, appropriate sub-samples random $N(z)$'s were generated using the 1000 km s^{-1} smoothed randomized distribution for simplicity (as described in Section 3.2.3). For each random catalogue 20 times the number of random points as real galaxies were generated.

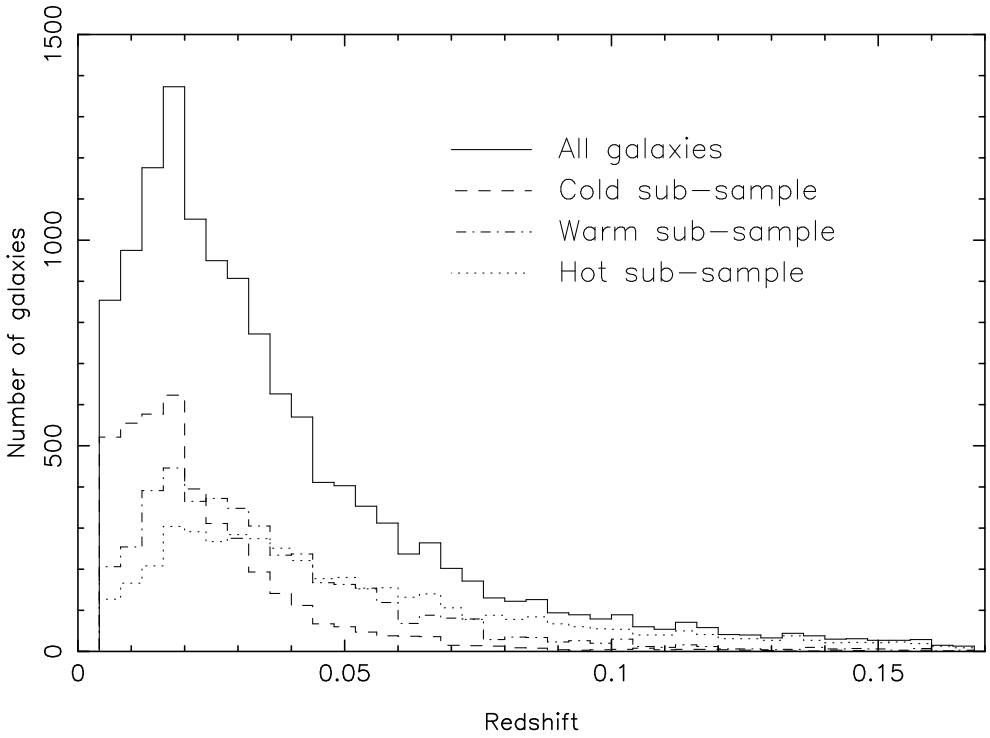


Figure 4.1: Redshift distributions for the colour sub-samples listed in Table 4.1

4.2.2 Colour sub-samples

In order to investigate the dependence of clustering on the temperature of the galaxies, the ADDSCAN $60\mu m$ and $100\mu m$ fluxes (f_{60} and f_{100}) were used to subdivide the parent redshift catalogue into three sub-samples: hot galaxies, with $f_{100}/f_{60} < 1.7$; warm galaxies, with $1.7 < f_{100}/f_{60} < 2.3$; and cold galaxies, with $f_{100}/f_{60} > 2.3$. These boundaries correspond to blackbody temperatures of around 31 and 28K, and were chosen to give roughly equal numbers in each sub-sample. The actual numbers of galaxies in the sub-samples were 4452, 4388 and 4107 respectively. The mean f_{100}/f_{60} colour ratios for the hot, warm and cold sub-samples were 1.31, 1.98 and 2.91 respectively, corresponding to black-body temperatures of about 34K, 30K and 26.5K respectively. The mean for the whole catalogue is 2.05, corresponding to a black-body temperature of around 29.5K.

The redshift distributions for the colour sub-samples are plotted in Figure 4.1. It can be seen that the cooler samples tend to peak at a slightly lower redshift than the hotter samples. This reflects the correlation between colour and absolute luminosity as plotted in Figure 4.2: cooler galaxies tend to be fainter, and so

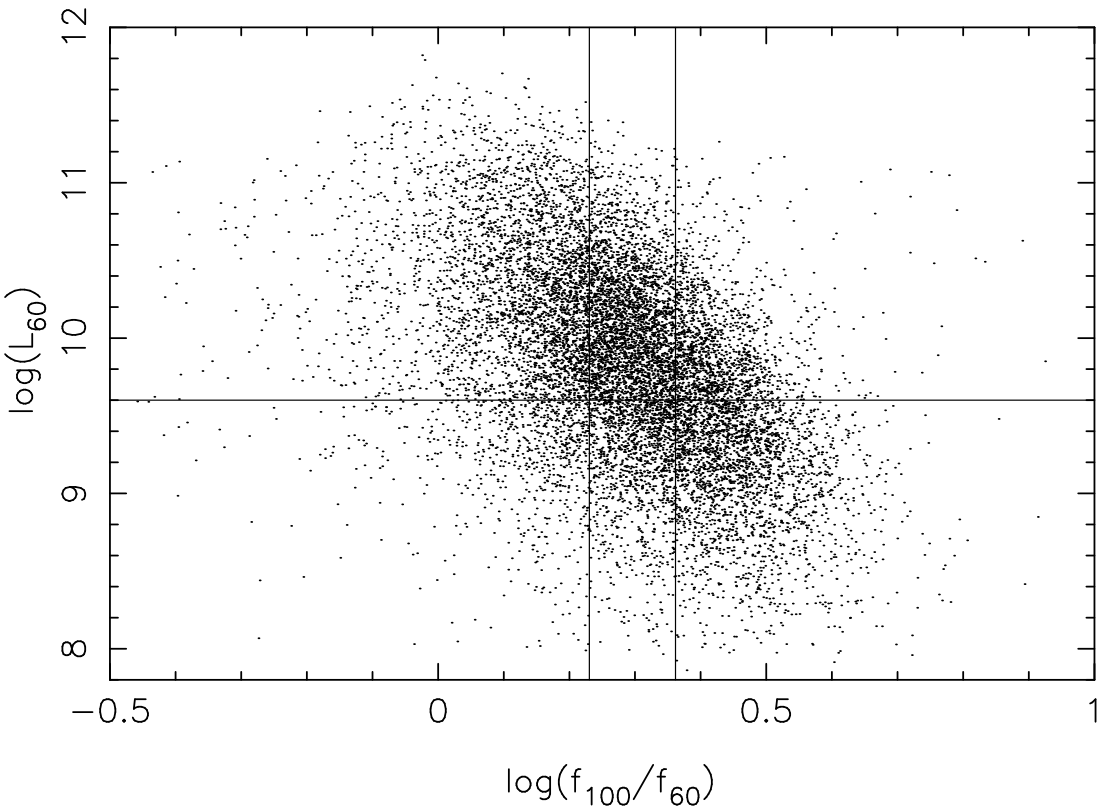
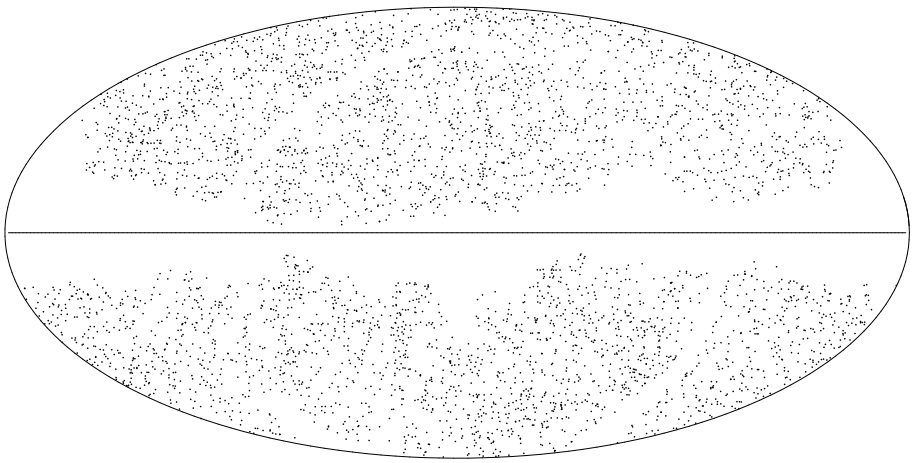


Figure 4.2: Absolute luminosity at $60\mu m$ against the $100\mu m/60\mu m$ colour. The vertical lines show the boundaries used to select our colour sub-samples. The horizontal line shows the boundary between the luminosity sub-samples.

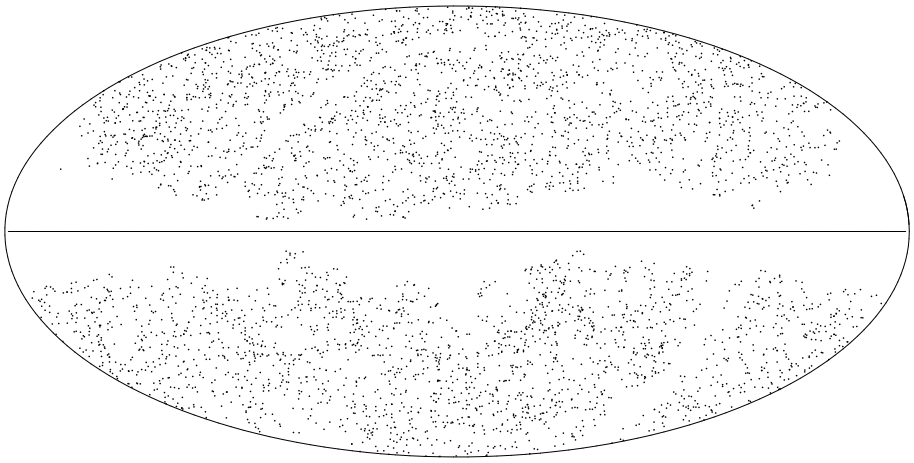
are only seen nearby. Nevertheless, these differences in redshift distributions are smaller than the difference for luminosity selected sub-samples, as discussed in the next section.

The angular distribution of these colour sub-samples are plotted in Figure 4.3. It can be seen that these samples show no obvious gradients as a function of position on the sky, and that the cooler sample appears to be more strongly clustered. Figure 4.4 shows the projection of the galaxies on to a plane along the celestial equator. Again the cooler sample appears more clustered than the warmer samples.

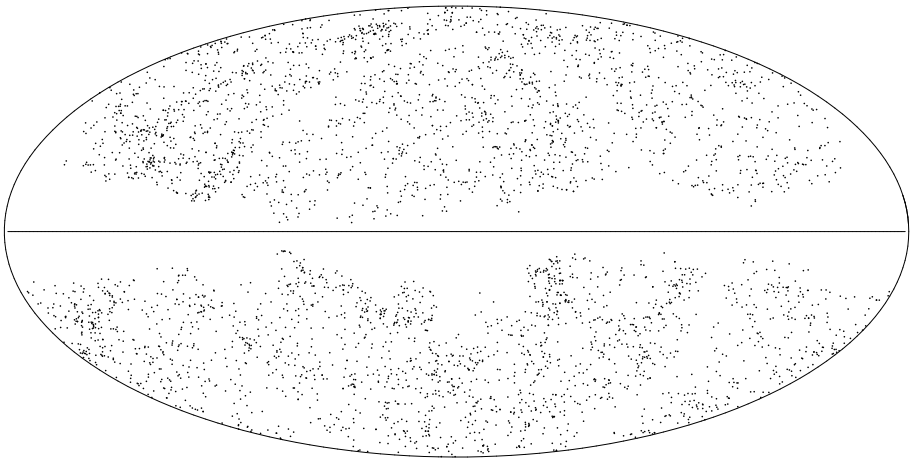
Though apparently quite significant, these visual impressions should be treated with caution. The cooler sample is shallower than the warmer samples, so the angular clustering would be stronger, even if the samples had the same spatial



(a)



(b)



(c)

Figure 4.3: Plots of the sky position of galaxies for the three colour sub-samples listed in Table 4.1: (a) hot, (b) warm, and (c) cold. The horizontal lines show the Galactic equator and the obscuration around this is masked out in the analysis.

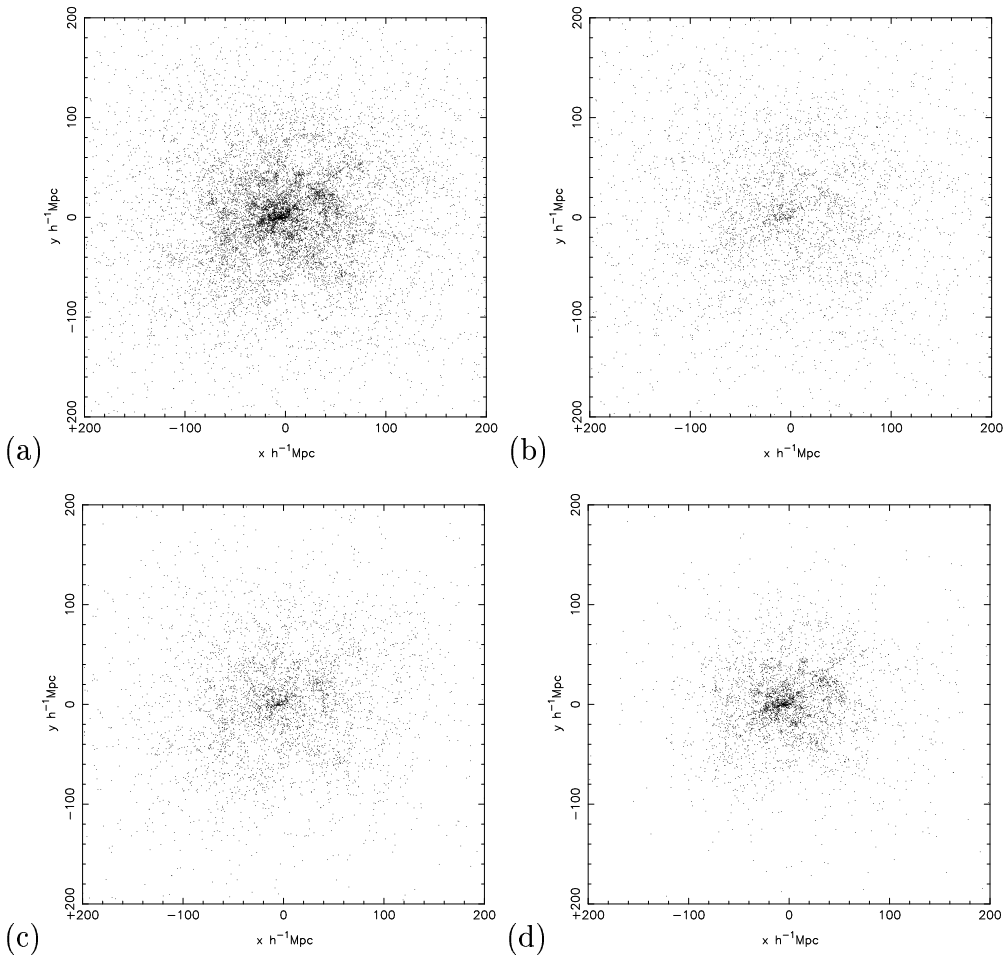


Figure 4.4: The galaxy positions projected onto the celestial equator for: (a) the full survey, (b) hot, (c) warm, and (d) cold sub-samples.

clustering. Also the cool sample has a higher space density than the warmer samples, and this may enhance the visual appearance of clustering.

4.2.3 Absolute luminosity sub-samples

The parent sample was also divided into sub-samples of absolute luminosity at $60\mu m$: faint galaxies, with $\log_{10}(L_{60}) < 9.6$; and bright galaxies with $\log_{10}(L_{60}) > 9.6$, where $L_{60} = 4\pi d^2 f_{60}$ and d is the luminosity distance to the galaxy. The units of L_{60} are $L_{\odot}h^{-2}$. Note that the usual definition of an ultra-luminous *IRAS* galaxy corresponds to $\log_{10}(L_{60}) \gtrsim 11.4$ (Soifer et al. 1987), and so the lower limit of the brighter sample is a factor of ~ 60 fainter than ultra-luminous galaxy

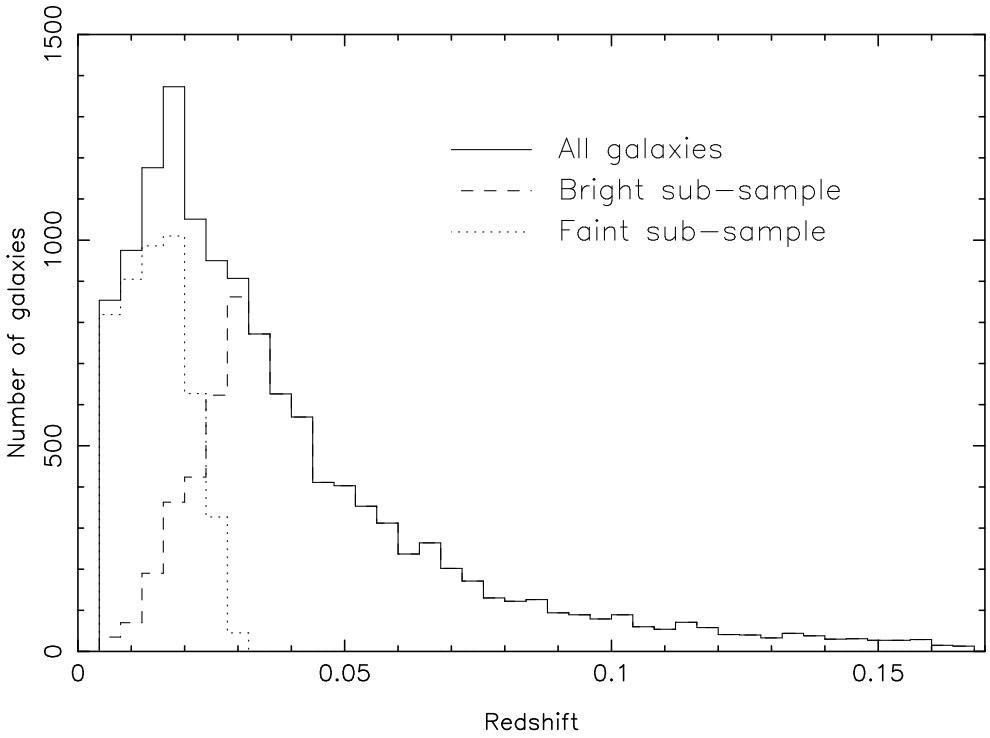


Figure 4.5: Redshift distributions for the luminosity sub-samples (Table 4.2).

samples.

The bright sample has a higher mean redshift, and hence covers a much larger volume, but the galaxies are a more dilute sampling of the density field and so a larger number is needed to give a similar number of small separation pairs. Hence the faint and bright samples were chosen to have 4719 and 8228 galaxies respectively. The redshift distribution for these luminosity sub-samples is shown in Figure 4.5.

Since the parent sample is flux limited, the low luminosity galaxies have a sharply defined upper redshift limit corresponding to the distance where the observed flux equals the flux limit. For the more luminous galaxies, the clustering signal is dominated by the volume at higher redshifts. This means that the clustering measurements for the two samples come from almost independent volumes, and so cosmic variance introduces a large uncertainty in the comparison. The clustering measurements for colour selected sub-samples are less affected by cosmic variance because they cover very similar volumes, and so their comparison should be more reliable.

4.2.4 Volume limited sub-samples

The cosmic variance between the luminosity sub-samples can be partly overcome by using volume limited sub-samples. To do this, the parent catalogue was split into sub-samples limited by distance and also by absolute luminosity. Galaxies within a sphere of radius x, centred on us, are included only if their absolute luminosity (L_{60}) is greater than $4\pi x^2 f_{\rm min}$, where $f_{\rm min}$ is the flux-limit of the survey or sample (0.6 Jy for the full PSCz survey). Unfortunately the numbers of galaxies in these samples, shown in Table 4.3, are rather small. The advantage of using these samples is that, without galaxy evolution or galaxy clustering, these samples' density profiles (i.e. their galaxy number density as a function of radius) will be flat, making them easier to analyse. *IRAS* galaxies are known to show very rapid evolution (Saunders et al. 1990), and this could introduce a radial gradient in space density in the volume-limited samples. Galaxy clustering would produce small fluctuations in the density profile.

Figure 4.6 shows the galaxy number density as a function of radius, normalised so that the maximum radius of the volume is 1, for the samples with radius 100, 150, 200 and 300 h^{-1}Mpc. No significant gradients are apparent, and so a constant density is assumed in our clustering estimates. These sub-samples allow another test for a luminosity trend in the clustering as the larger volumes are limited to brighter galaxies than the smaller volumes. Note that the different volumes have a significant fraction of galaxies in common, so that the results for different volume scales are not completely independent.

For the volume-limited sub-samples a different technique was used to estimate the correlation functions. Following the method of Croft et al. (1997), a maximum likelihood approach was used that assumes that each distinct galaxy pair is an independent object. This technique uses the fact that $1+\xi(s)$ can be considered as the probability distribution for galaxy pairs as a function of separation. Thus a power law ξ can be fitted directly to the galaxy pair distribution without the need to specify arbitrary bins of pair separation. The resulting power law parameters tend to be more robust than fitting to a binned version of ξ, especially for small samples. The homogeneity of volume limited samples means that simple unit pair weighting yields the best results.

If galaxy pairs were independent, the calculated likelihood contours would directly provide 1σ uncertainties in the best-fit parameters, equivalent to the Poisson error

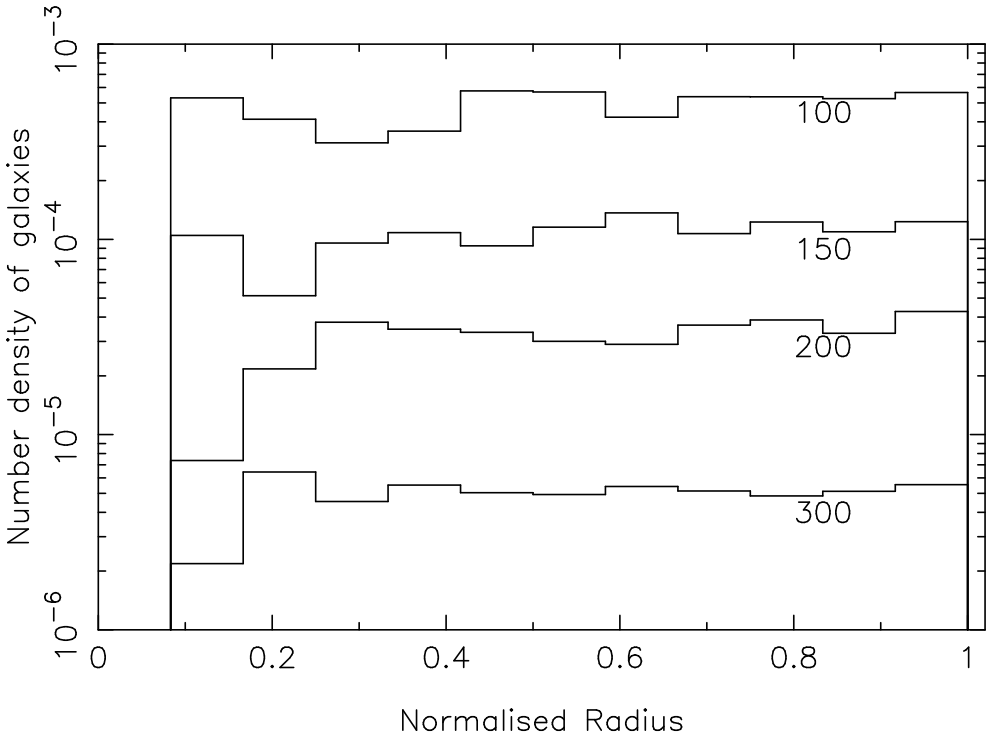


Figure 4.6: Galaxy number density plotted against normalised radius for the volume limited samples with radii labelled.

in the pair counts. Since galaxy pairs are correlated, the effective number of pairs is reduced, and so the uncertainties are increased by a factor $(1 + 4\pi n J_3)$. These corrected Poisson estimates still do not allow for any uncertainty introduced by the variance in mean density on the scale of sample size. A simple approach was used to include this effect, by considering the variance in the mean density within the sample volume, $\mathrm{var}(\delta n/n) = 3J_3/R^3$, where R is the radius of the volume. This introduces an uncertainty in ξ, which we add in quadrature to the corrected Poisson estimates.

For volumes with radius $R \leqslant 100\ h^{-1}$Mpc the ten ΛCDM mock catalogues have been analysed using the same method. The ΛCDM model was chosen because its clustering amplitude is the closest match to the observed amplitude. The standard deviation between results provides an independent estimate of the uncertainties in s_0 for each sample. For the volumes considered (with radius $R \geqslant 50\ h^{-1}$Mpc) the value of J_3 is roughly constant, and we find that the two methods agree to $\sim 10\%$ if we choose $J_3 = 1500$. For volumes with radius $R > 100\ h^{-1}$Mpc the mock catalogues are too small, and so we have to rely on

the analytical approximation to estimate the uncertainties.

4.2.5 Mock catalogues

The mock catalogues described in the previous chapter are used to estimate the errors on the sub-sample results.

Colours can be included in the mocks by considering Figure 4.2, which can be modelled as a linear relation between $\log_{10}(f_{100}/f_{60})$ and $\log_{10}(L_{60})$ with a Gaussian dispersion about the mean. Each galaxy can then be assigned a colour according to its luminosity,

$$\log_{10}(f_{100}/f_{60}) = A \log_{10}(L_{60}) + B + C \tag{4.1}$$

where $A = -0.117$, $B = 1.427$ and C is a random number selected from a Gaussian distribution with variance of 0.2. Then colour sub-samples can be selected and tested in the same way as the real data. Since the luminosities were assigned at random to the mock galaxies, and the colours are based on the luminosities, the clustering in the mocks is independent of luminosity and colour.

The standard deviation of the results between the mock realizations gives an estimate of the uncertainty on the measurements for the real data. We choose to use the ΛCDM mock catalogue results as they match the data the closest.

4.3 Results

4.3.1 Colour sub-samples

The estimates of the redshift-space correlation functions for the three colour sub-samples are shown in Figure 4.7. The error bar for each point shows the standard deviation between 10 realizations of the ΛCDM mock catalogues. It is clear that over the range $1 < s < 10$ h^{-1}Mpc the amplitude of ξ(s) is higher by about a factor of 1.5 for the cooler sub-sample. Fitting a power law to the data over the range of $1 < s < 10$ h^{-1}Mpc gives parameters as shown in Table 4.1, clearly confirming the increase in clustering amplitude for cooler galaxies. Figure 4.8 shows χ^2 contours for the hot and cold sub-sample, and suggests that trend is significant at about the 2σ level.

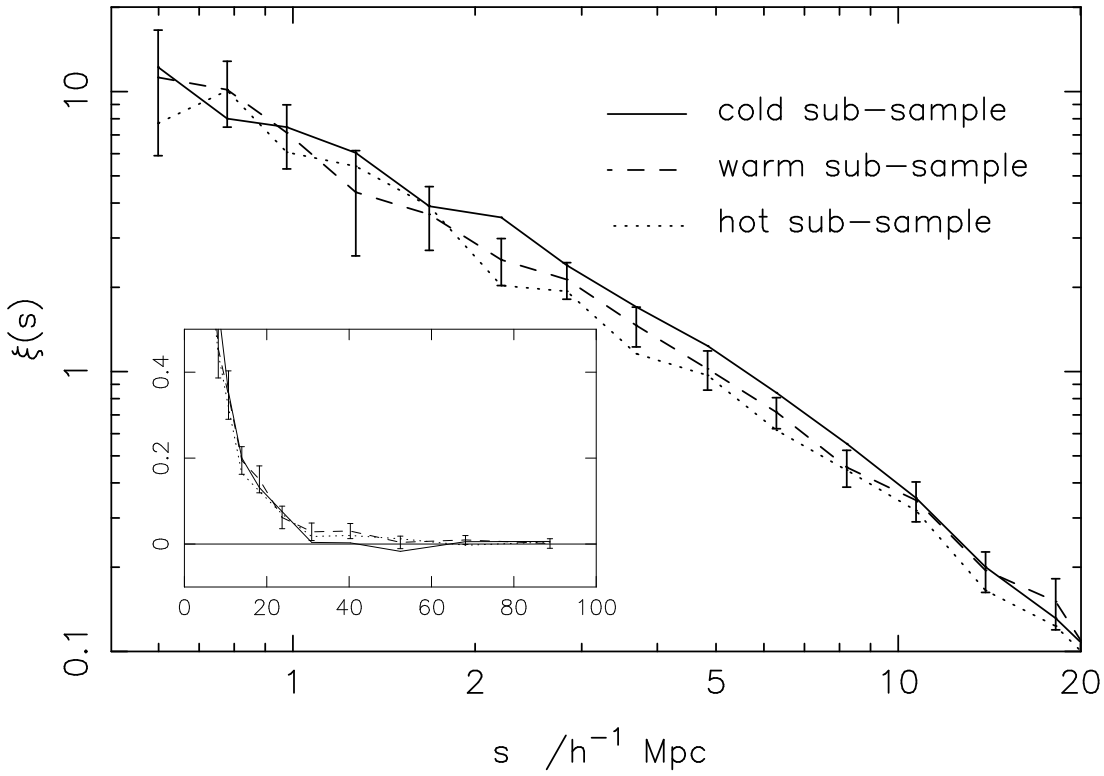


Figure 4.7: $\xi(s)$ for the colour sub-samples with error bars from the scatter in mock catalogue results. For clarity, error bars are plotted only for the warm sub-sample but others are similarly sized. The inset is plotted on a linear scale.

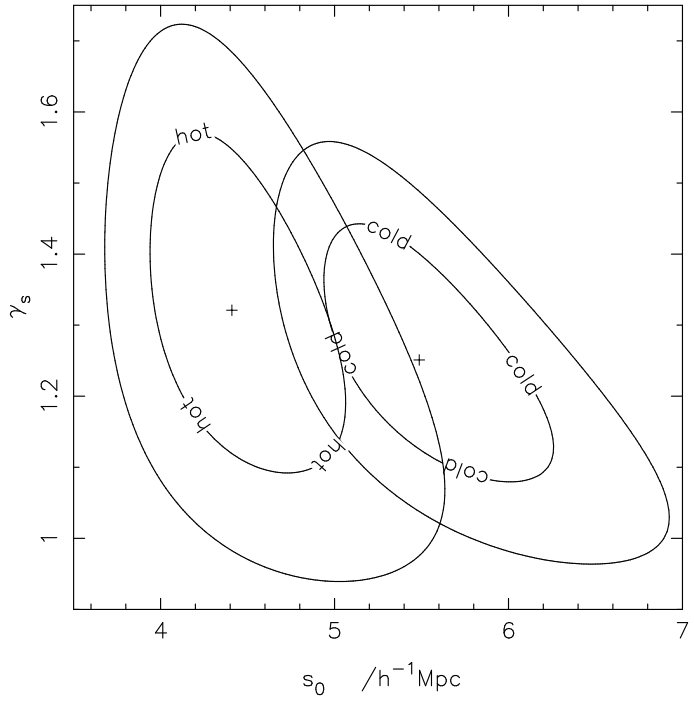


Figure 4.8: 1σ and 2σ likelihood contours for the power-law fits to ξ of the hot and cold sub-samples as in Table 4.1.

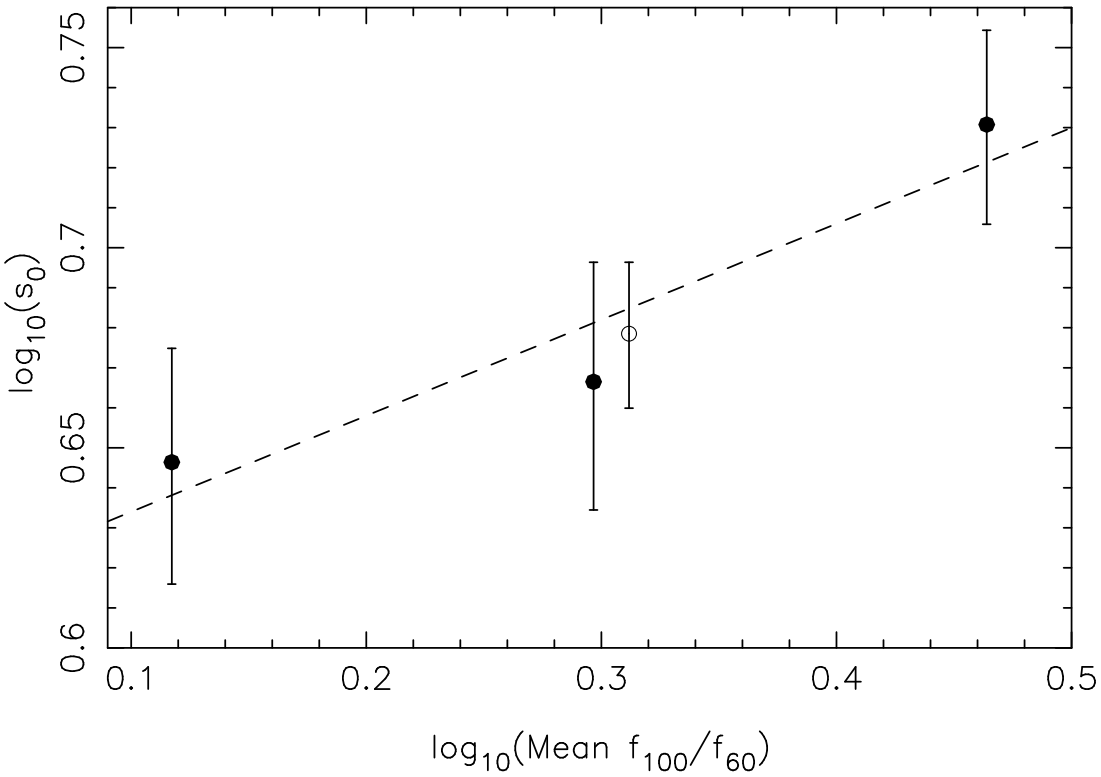


Figure 4.9: Clustering length s_0 (with fixed $\gamma_s = 1.30$) against mean f_{100}/f_{60} colour for the colour sub-samples (filled points) and the full catalogue (open point). The dashed line is the best-fit to the colour sub-samples (Equation 4.2).

Table 4.1: Power law fit parameters to the colour sub-samples

Free parameter best fits:

description	colour range	s_0 / h^{-1}Mpc	γ_s
hot	$f_{100}/f_{60} < 1.7$	4.41 ± 0.29	1.32 ± 0.06
warm	$1.7 < f_{100}/f_{60} < 2.3$	4.76 ± 0.30	1.21 ± 0.06
cold	$f_{100}/f_{60} > 2.3$	5.49 ± 0.29	1.25 ± 0.05

Fixed γ_s at 1.30:

description	colour range	s_0 / h^{-1}Mpc
hot	$f_{100}/f_{60} < 1.7$	4.43 ± 0.30
warm	$1.7 < f_{100}/f_{60} < 2.3$	4.64 ± 0.33
cold	$f_{100}/f_{60} > 2.3$	5.38 ± 0.30

Since the best-fit s_0 and γ_s are correlated, we have also fitted the data using a fixed value for γ_s, set to the best fit value for the full sample, $\gamma_s = 1.30$. The resulting s_0 values for the three different colour sub-samples are listed in Table 4.1 and plotted as a function of mean colour in Figure 4.9. The trend of increasing s_0 for cooler galaxies is clearly apparent at about the 3σ level. A simple linear regression to the points yields

$$\log_{10}(s_0) = 0.24 \log_{10}(f_{100}/f_{60}) + 0.61. \tag{4.2}$$

Analyzing colour selected sub-samples from the mock catalogues shows no significant difference in amplitude between the hot and cold sub-samples. This is as expected since we assigned the colours at random, but the measurements provide an important check that there are no subtle biases in our algorithms.

4.3.2 Luminosity sub-samples

The estimates for the two luminosity sub-samples are shown in Figure 4.10. The error bars are derived from the scatter between ten realizations of the ΛCDM mock galaxy catalogues. It is clear that there is very little difference between the two sub-samples. The power law fit parameters are shown in Table 4.2. These parameters show no significant difference when the errors are taken into account.

For all mock catalogues no difference was found between the results for the faint and bright sub-samples within the errors. This is as expected since we assigned the luminosities at random, but the measurements provide an important check that there are no subtle biases in our algorithms.

4.3.3 Volume limited sub-samples

The results from the volume limited sub-samples are shown in Table 4.3 and displayed in Figure 4.11. Although there is clearly a hint of a trend of reducing amplitude with scale, it can be seen that the error bars are too large to say that this is a statistically significant relation.

The minimum absolute luminosity in the volume limited samples increases with the radius of the volume, so the mean luminosity also increases. Figure 4.13 explicitly shows the clustering amplitude as a function of the mean luminosity in the volume limited samples.

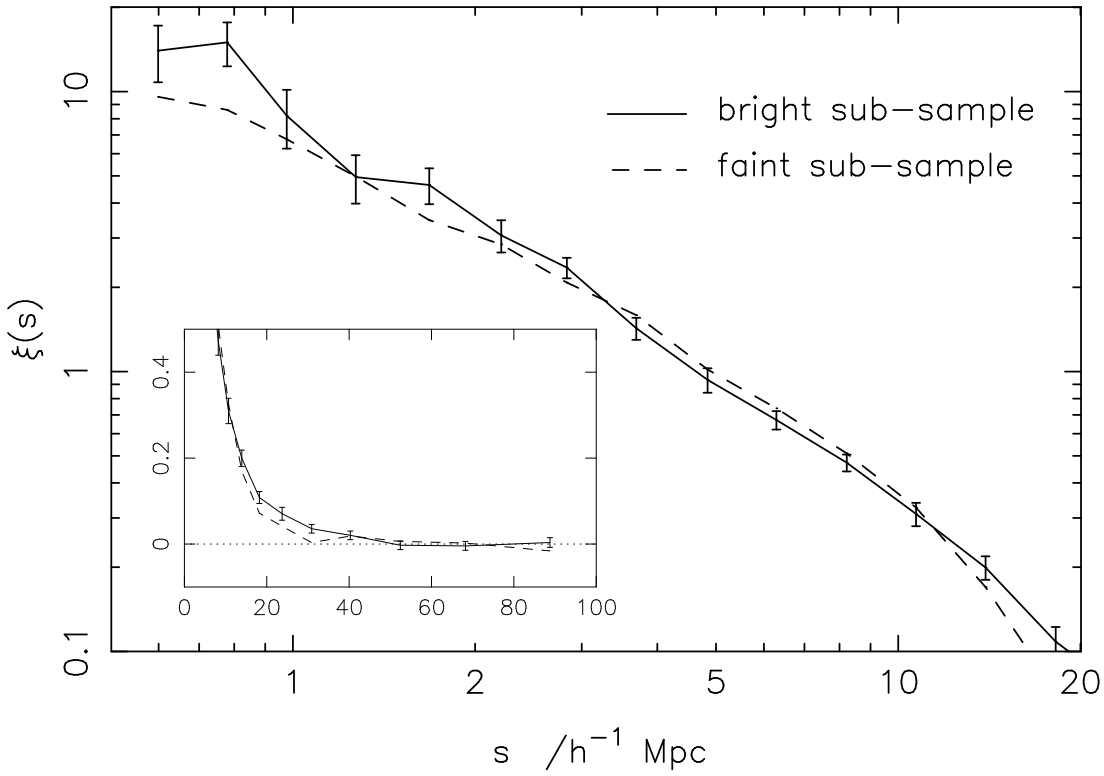


Figure 4.10: $\xi(s)$ for the luminosity sub-samples defined in Table 4.2 with error bars from the scatter in mock catalogue results. The inset is plotted on a linear scale. For clarity, error bars are plotted only for the bright sub-sample.

Table 4.2: Power law fit parameters to the luminosity sub-samples

Free parameter best fits:

description	luminosity range	$s_0\ /h^{-1}$Mpc	γ_s
faint	$\log_{10}(L_{60}) < 9.6$	4.96 ± 0.44	1.21 ± 0.04
bright	$\log_{10}(L_{60}) > 9.6$	4.80 ± 0.31	1.36 ± 0.06

Fixed γ_s at 1.30:

description	luminosity range	$s_0\ /h^{-1}$Mpc
faint	$\log_{10}(L_{60}) < 9.6$	4.82 ± 0.54
bright	$\log_{10}(L_{60}) > 9.6$	4.89 ± 0.35

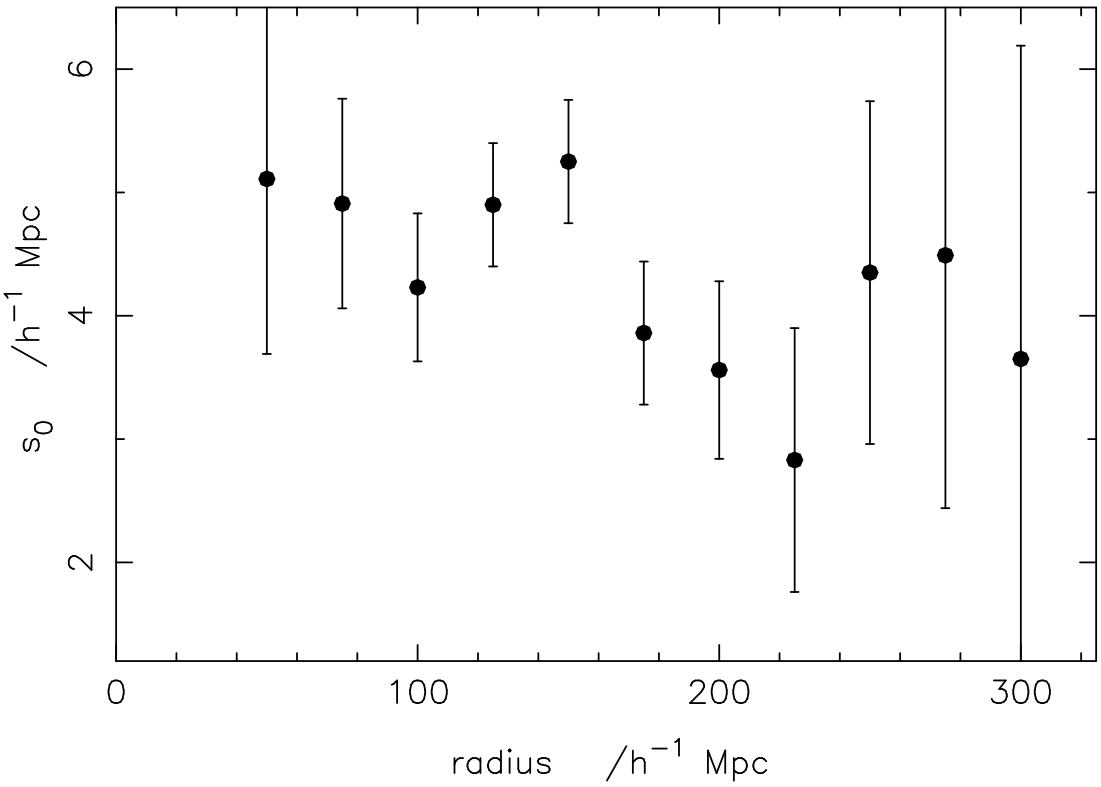


Figure 4.11: s_0 (with fixed $\gamma_s = 1.30$) against volume radius for the volume limited sub-samples with error bars showing the overall uncertainties as described in Section 4.2.4.

Table 4.3: Maximum likelihood fits to the volume limited sub-samples with γ_s fixed at 1.30, and uncertainties estimated as described in Section 4.2.4.

Volume radius (h^{-1}Mpc)	no. of galaxies	$\langle \log_{10}(L/L_\odot) \rangle$	s_0 (h^{-1}Mpc)
50	1699	9.41	5.11 ± 1.19
75	2151	9.71	4.91 ± 0.65
100	2188	9.94	4.23 ± 0.44
125	1961	10.12	4.90 ± 0.38
150	1633	10.28	5.25 ± 0.42
175	1399	10.41	3.86 ± 0.53
200	1202	10.53	3.56 ± 0.74
225	958	10.64	2.83 ± 1.08
250	779	10.74	4.35 ± 1.43
275	653	10.83	4.49 ± 2.00
300	587	10.91	3.65 ± 2.50

4.3.4 Projected correlation functions

So far this analysis has been restricted to the redshift-space correlation functions. This means that any differences in the velocity distributions of the different sub-samples of galaxies will affect the amplitude of $\xi(s)$, particularly on small scales where the velocity separations are comparable to the random peculiar velocities.

Figure 4.12 shows the projected correlation functions (see Section 2.2.2 for a definition) for some of the sub-samples considered already. In the top panel the cold galaxies are marginally more clustered than the hot galaxies, as seen in the redshift-space correlation functions. In the bottom panel, the bright and faint sub-sample results are virtually identical on small scales but on scales $> 8\ h^{-1}$Mpc the faint sub-sample drops away from the bright sub-sample, but given the error bars on the points it is at a low significance. A similar, low significance, trend is also seen in $\xi(s)$, but the data lacks the signal-to-noise to conclude anything meaningful from these trends.

4.4 Conclusions and discussion

A colour dependence is detected in the clustering of PSCz galaxies: the cooler galaxies are more strongly clustered than the hotter galaxies. Although the significance level of amplitude variations is not high, this does not mean that the effects are small: the clustering amplitude differs by a factor of 1.5 between the warm and cool sub-samples. No significant variation is detected in the clustering for luminosity selected subsamples. Volume limited sub-samples show a slight trend corresponding to a weaker clustering for brighter galaxies, but the uncertainties mean it is not significant.

These results are different to those of Mann, Saunders & Taylor (1996) who found marginal evidence that warmer galaxies are more strongly clustered than cooler galaxies. The main reason for the difference is simply the factor of 6 increase in number of galaxies in our sample, which dramatically reduces the uncertainties in our clustering estimates. The uncertainties on the earlier measurements encompass our results, so there is actually no discrepancy.

From a fractal analysis of the same sub-samples, Pan (2002) came to the same conclusions as in this work.

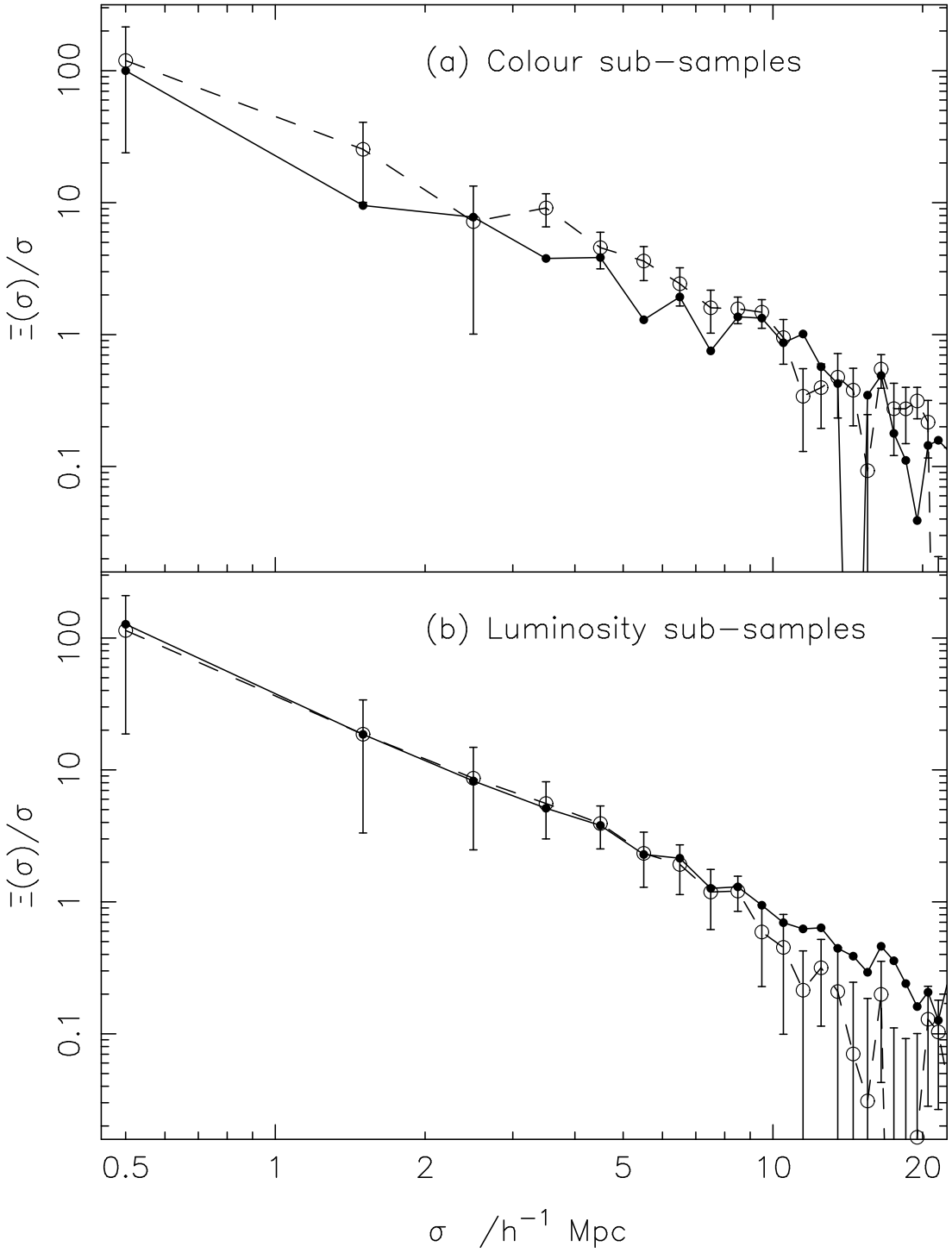


Figure 4.12: The projected correlation function for various sub-samples. (a) hot (solid line) and cold (dashed line) sub-samples. (b) bright (solid line) and faint (dashed line) sub-samples. Error bars are from the rms spread in mock catalogue results and are plotted for just one sub-sample to avoid confusion.

The variation of clustering with FIR colour that is found in the PSCz is consistent with the variation with star-formation rate indicated by Hα and [O II] emission lines in the Stromlo-APM (SAPM) survey (Loveday, Tresse & Maddox 1999). They split the SAPM survey into three samples with low, medium and high Hα equivalent widths (EW), and estimated s_0 for the samples to be 8.7, 5.5 and 4.6 h^{-1}Mpc respectively. A similar split into low, medium and high [O II] EW samples gives s_0 values of 8.6, 4.9 and 4.1 h^{-1}Mpc respectively. The low EW samples are early type galaxies, which do not appear in *IRAS* samples. The medium and high EW samples are actively star-forming galaxies, which are more similar to *IRAS* samples. The clustering amplitude of the medium and high EW galaxies is very similar to that of the *IRAS* galaxies, suggesting that they do indeed trace a similar population. Furthermore, the change in s_0 between the medium and high SAPM emission-line galaxies is very similar to the change between cold and hot PSCz galaxies as seen in Figure 4.9.

The luminosity dependence is a more complex effect, because the FIR luminosity of a galaxy depends on both the mass of dust, and the recent star-formation rate. If the dust mass is correlated with the halo mass, then more luminous galaxies will, on average, be in more massive halos, which are expected to have a higher clustering amplitude. On the other hand, a high star-formation rate will make a galaxy more luminous, and galaxies with a high star-formation rate tend to be less clustered. It is not obvious which will be the dominant effect.

Given the correlation between colour and luminosity seen in Figure 4.2 the luminosity dependence of clustering can be inferred from the colour dependence of clustering seen in Figure 4.9. A fit to the data in Figure 4.2 leads to,

$$\log_{10}(f_{100}/f_{60}) = -0.117 \log_{10}(L_{60}) + 1.427 \tag{4.3}$$

and so together with Equation 4.2, yields the relation

$$\log_{10}(s_0) = -0.028 \log_{10}(L_{60}) + 0.95 \tag{4.4}$$

This relation is plotted as the dashed line on Figure 4.13, and it can be seen that it is consistent with the data points. The luminosity variation seen in optical galaxy samples is approximately of the form $\log_{10}(s_0) = 4.37(0.7 + 0.3L/L^*)$ (Benoist et al. 1996). The dotted line shows this with $L^* = 3.6 \times 10^9 L_\odot$ (Springel & White 1998). This is clearly inconsistent with the trend seen in the PSCz sample.

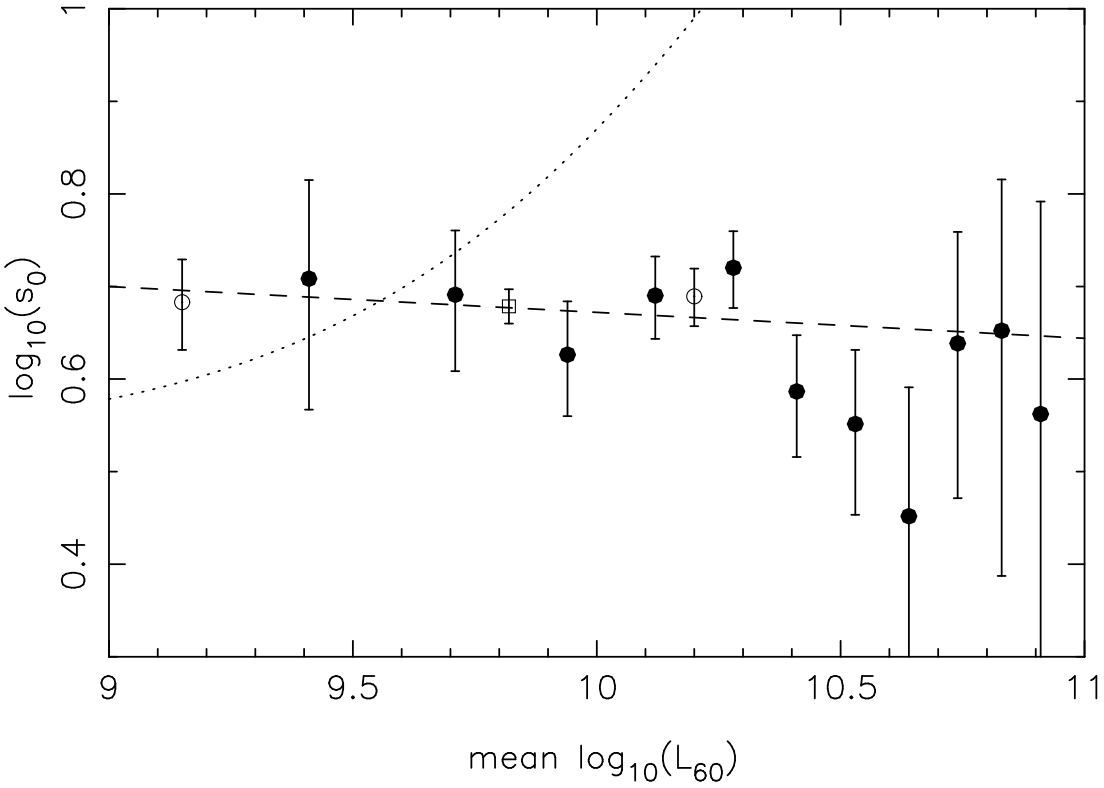


Figure 4.13: Clustering length, s_0, plotted against mean absolute luminosity for the volume limited sub-samples (filled circles), the bright and faint sub-samples (open circles) and the whole sample (open square). The dashed line shows the trend expected from the colour dependence seen in Figure 4.9 and the luminosity-colour correlation in Figure 4.2. The dotted line shows $\log_{10}(s_0) = 4.37(0.7 + 0.3L/L^*)$, which approximates the variation seen in optical galaxy samples.

The small decrease in clustering that is found for more luminous galaxies is consistent with the colour dependence effect alone, with no evidence for an increase in clustering for more luminous galaxies. If the standard picture of high-peak biasing is correct, the conclusion is that there is a very weak correlation between the FIR luminosity of a galaxy and the mass of it's dark-matter halo. However, models which allow a variable number of galaxies per dark halo, and include scatter about the mean mass-luminosity relation lead to a very weak luminosity dependence of clustering (Somerville et al. 2000).

Chapter 5

Clustering in the 2dF Galaxy Redshift Survey

The work in this chapter was presented in Hawkins et al. (2002).

5.1 The 2dF Galaxy Redshift Survey

5.1.1 The 2-degree Field instrument

The 2-degree Field (2dF) instrument (Figure 5.1, Lewis et al. 2002) is a multi-fibre spectrograph on the 3.9m Anglo-Australian Telescope (AAT) situated at Siding Spring Observatory. It is capable of taking up to 400 spectra simultaneously and is shown on the telescope in Figure 5.2. As its name suggests the instrument has a 2° field of view, and the fibres are placed in this field by a robot positioner. There are two of these field plates, and the instrument is constructed so as to allow one field to be positioned, or 'configured', while the other is observed, so maximising the use of telescope time. On a good night it is possible to obtain ~ 3000 spectra, making large redshift surveys perfectly suited to this instrument. To maximise the use of the instrument it was decided to mount both a galaxy redshift survey and a quasar redshift survey, which have been observed simultaneously, with the first observations in 1997, and the last in 2002.

5.1.2 Survey construction

The 2dFGRS is selected in the photometric $b_{\rm J}$ band from the APM galaxy survey (Maddox, Efstathiou & Sutherland 1990), which was observed on the UK Schmidt

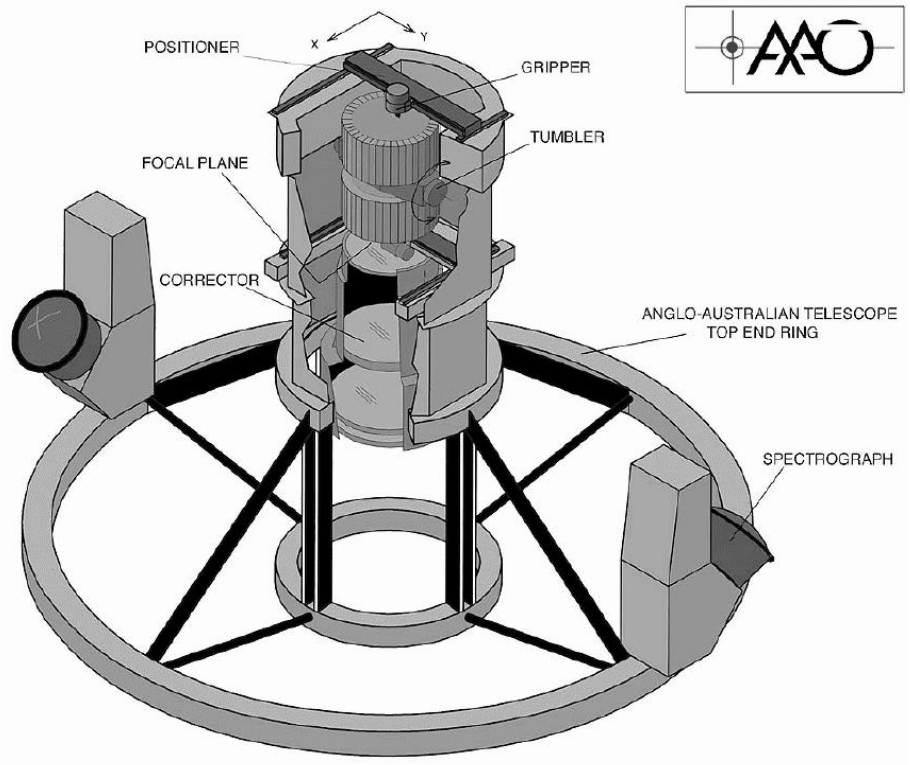


Figure 5.1: A schematic drawing of the 2dF instrument.

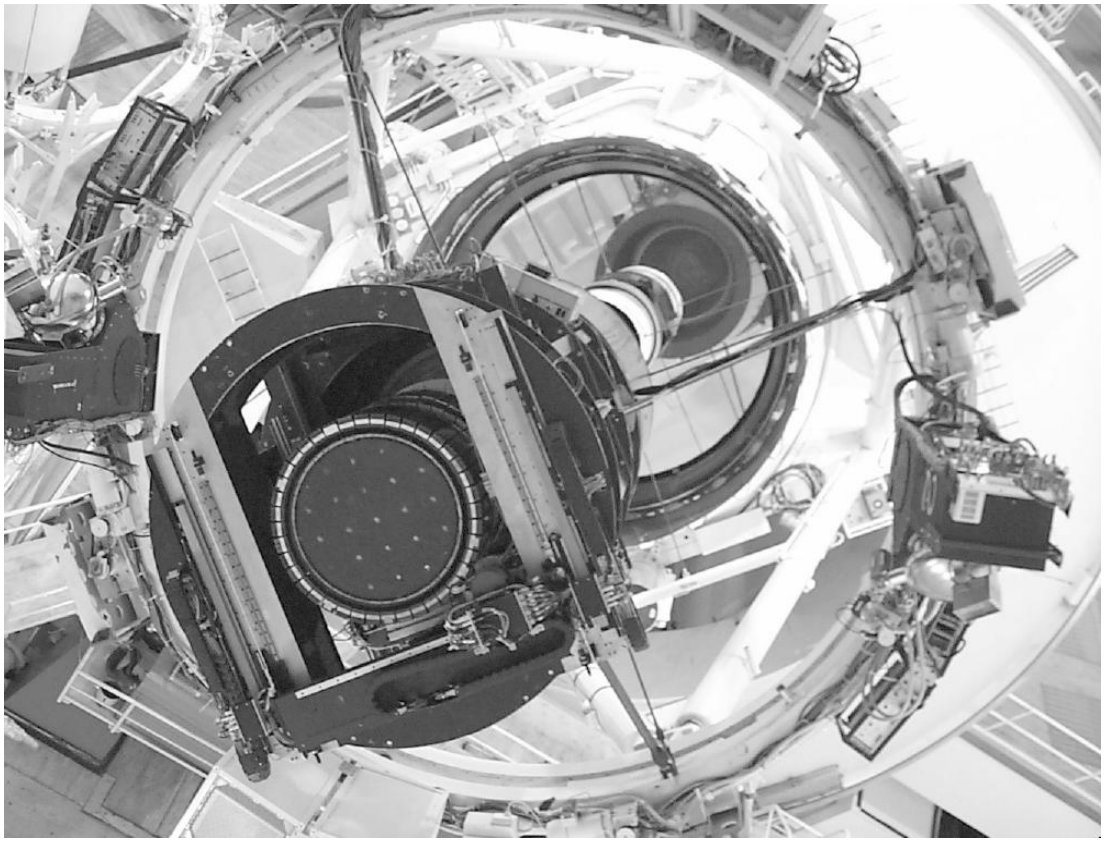


Figure 5.2: The 2dF instrument on the AAT. Courtesy: J.Pogson.

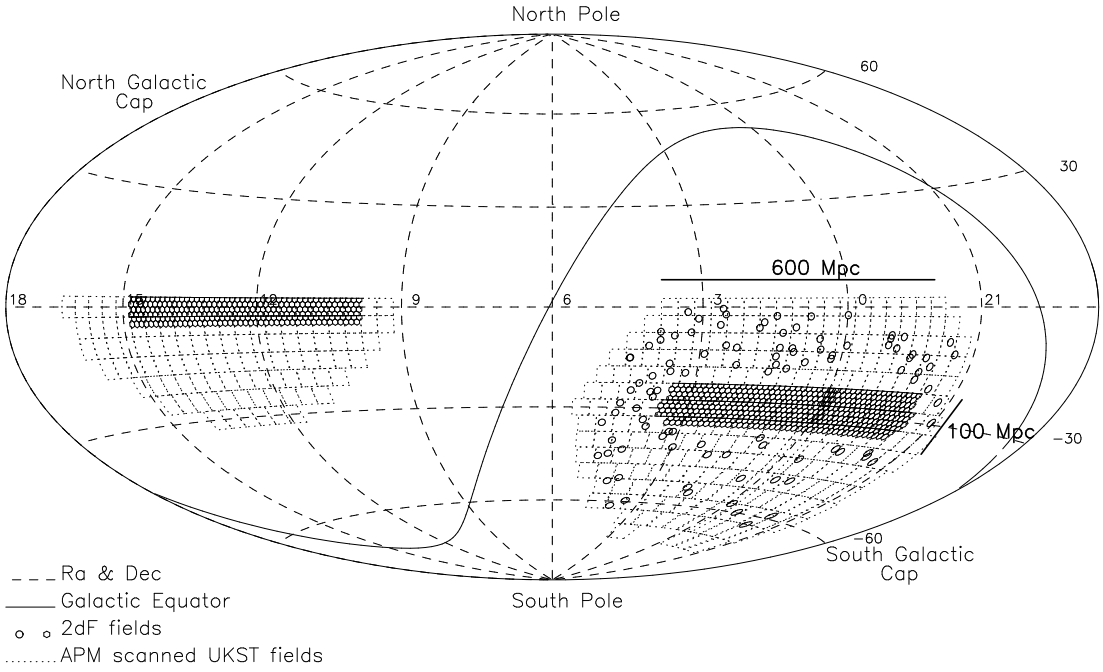


Figure 5.3: The survey map for the 2dFGRS.

Telescope (also at Siding Spring Observatory). Although not possible to observe the entire APM area the survey was designed to take spectra over ~ 2000 deg^2, and observe $\sim 250\,000$ reliable galaxy redshifts.

The bulk of the solid angle of the survey is made up of two broad strips, one in the South Galactic Pole region (SGP) covering approximately $-37^\circ.5 < \delta < -22^\circ.5$, $21^{\rm h}40^{\rm m} < \alpha < 3^{\rm h}40^{\rm m}$ and the other in the direction of the North Galactic Pole (NGP), spanning $-7^\circ.5 < \delta < 2^\circ.5$, $9^{\rm h}50^{\rm m} < \alpha < 14^{\rm h}50^{\rm m}$. In addition to these contiguous regions, there are a number of circular 2-degree fields scattered randomly over the full extent of the low extinction regions of the southern APM galaxy survey (see Figure 5.3).

The magnitude limit at the start of the survey was set at $b_{\rm J} = 19.45$ but both the photometry of the input catalogue and the dust extinction map have been revised since and so there are small variations in magnitude limit as a function of position over the sky. This magnitude limit was chosen so that the number density of objects was similar to the number of fibres in each 2° field and also so that the observation time (~ 1 hour) on each field matched the time to configure the next plate. The field centres were chosen so that the denser regions had more than one pointing, allowing most objects to be targeted.

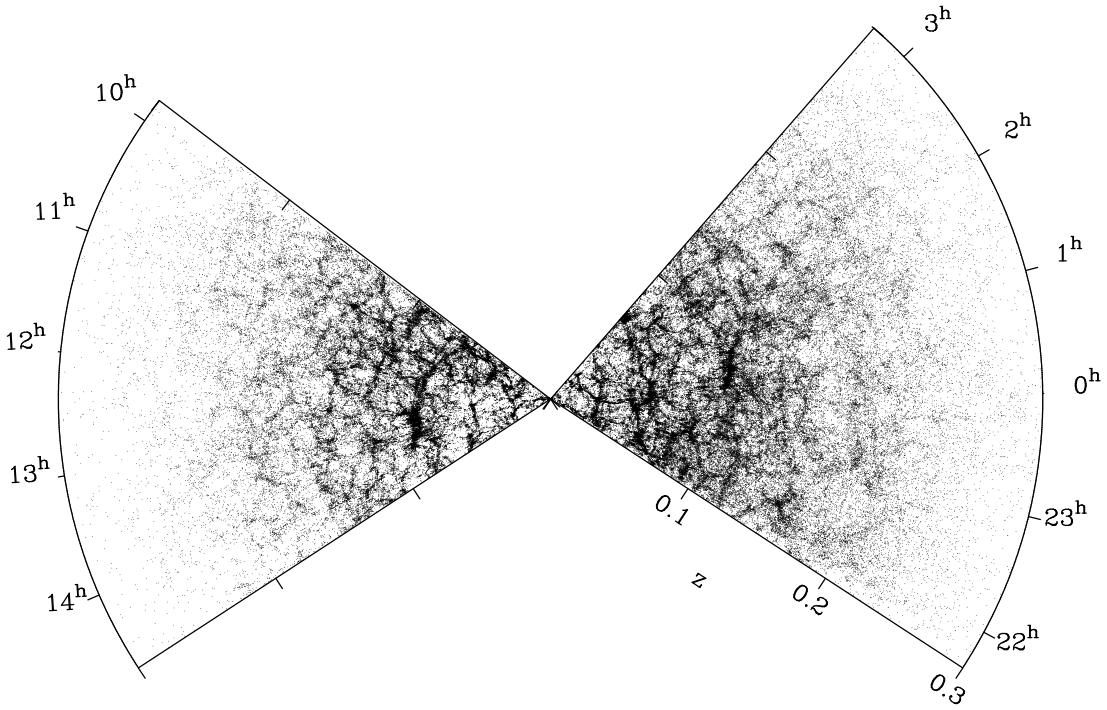


Figure 5.4: The redshift cone plot for the completed 2dFGRS.

5.1.3 Redshift measurements

The spectra are reduced in a pipeline and 'redshifted' at the telescope during the observing. A cross correlation with stellar templates is the primary method for assigning redshifts, but emission lines are also used. Manual redshifts can be applied if the automated methods fail. Each redshift is assigned a quality flag, Q, which can take the values:

$Q = 1$, poor spectrum

$Q = 2$, dubious redshift

$Q = 3$, probable redshift ($\sim 75\%$ confidence)

$Q = 4$, reliable redshift ($\sim 95\%$ confidence)

$Q = 5$, reliable redshift and a high-quality spectrum.

Most scientific analyses use all galaxies with quality flags $Q \geqslant 3$. The final redshift cone plot including all these galaxies is shown in Figure 5.4.

5.1.4 Completeness

The completeness of the survey data varies according to the position on the sky because of unobserved fields (mostly around the survey edges), un-fibred objects in observed fields (due to collision constraints or broken fibres) and observed objects with poor spectra. The variation in completeness is mapped out using a completeness mask (Colless et al. 2001; Norberg et al. 2002a) which is shown in Figure 5.5 for the data used in this chapter. The square holes in the data are due to bright stars which have been masked out.

5.1.5 Data selection

We use the data obtained prior to May 2002, which is virtually the completed survey. This includes 221 283 unique, reliable galaxy redshifts. We analyse a magnitude-limited sample with redshift limits $z_{\rm min} = 0.01$ and $z_{\rm max} = 0.20$, and no redshifts are used from a field with $< 70\%$ completeness. The median redshift $z_{\rm med} \approx 0.11$. The random fields, which contain nearly 25 000 reliable redshifts are not included in this analysis. After the cuts for redshift, completeness and quality we are left with 165 659 galaxies in total, 95 929 in the SGP and 69 730 in the NGP. These data cover an area, weighted by the completeness shown in Figure 5.5, of 647 deg^2 in the SGP and 446 deg^2 in the NGP, to the magnitude limit of the survey, and their positions on the sky are shown in Figure 5.6.

In all of the following analysis we consider the NGP and SGP as independent data sets. Treating the NGP and SGP as two independent regions of the sky gives two estimates for each statistic, and so provides a good test of the error bars we derive from mock catalogues (see below). We have also combined the two measurements to produce our overall best estimate by simply adding the pair counts from the NGP and SGP. The optimal weighting of the two estimates depends on the relative volumes surveyed in the NGP and SGP, but since these are comparable, a simple sum is close to the optimal combination.

It is important to estimate the effective redshift at which all our statistics are calculated. As ξ is based on counting pairs of galaxies the effective redshift is not the median, but a pair-weighted measure. The tail of high redshift galaxies pushes this effective redshift to $z_s \approx 0.15$. Similarly the effective magnitude of the sample we analyse is $M_s \approx -20.0$, corresponding to $L_s \approx 1.4L^*$ (using

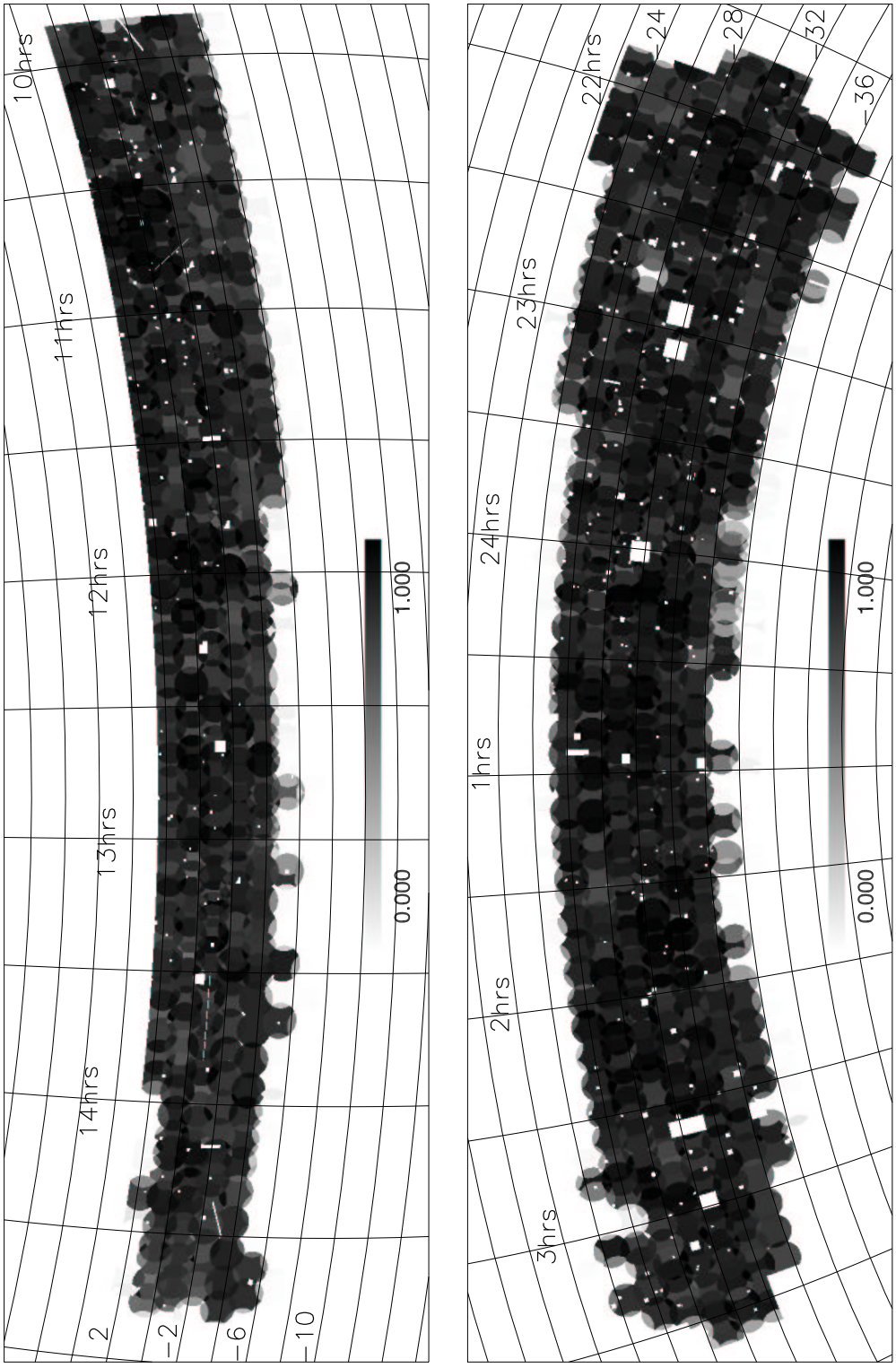


Figure 5.5: The redshift completeness masks for the NGP (top) and SGP (bottom). The greyscale shows the completeness fraction.

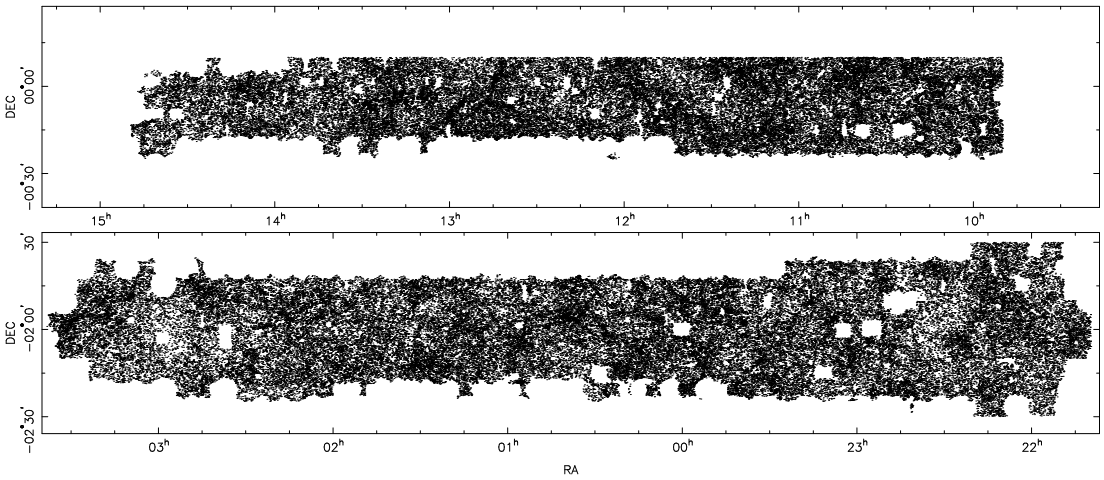


Figure 5.6: The galaxy sky distribution for the 2dFGRS samples used here, NGP (top) and SGP (bottom). The gaps in the data are mainly due to bright stars and low redshift completeness regions.

$M^* = -19.66$, Norberg et al. 2002a).

When converting from redshift to distance it is assumed that the Universe has a flat geometry with $\Omega_\Lambda = 0.7$, $\Omega_\mathrm{m} = 0.3$ and $H_0 = 100\ h$ km s^{-1}Mpc^{-1}, so that all scales are in units of h^{-1}Mpc.

5.1.6 Mock catalogues

For each of the NGP and SGP regions, 22 mock catalogues were generated from the ΛCDM Hubble Volume simulation (Evrard et al. 2002) using the techniques described in Cole et al. (1998). A summary of the construction methods is presented here but for more details see Norberg et al. (2002a).

These simulations used an initial dark matter power spectrum appropriate to a flat ΛCDM model with $\Omega_\mathrm{m} = 0.3$ and $\Omega_\Lambda = 0.7$. The dark-matter evolution was followed up to the present day and then a bias scheme was used to identify galaxies from the dark-matter haloes. The form of this bias was chosen to reproduce the clustering amplitude from observations (Model 2 of Cole et al. 1998, with a smoothing length, $R_S = 2\ h^{-1}$Mpc). The bias scheme used has two free parameters which are adjusted to match the mean slope and amplitude of the correlation function on scales greater than a few megaparsec. On scales smaller than the smoothing length there is little control over the form of the clustering,

but in reality the methods employed work reasonably well (see later Sections).

The resulting catalogues have a bias scheme which asymptotes to a constant on large scales, giving $\beta = 0.47$, but is scale dependent on small scales. Apparent magnitudes were assigned to the galaxies consistent with their redshift, the assumed Schechter luminosity function and the magnitude limit of the survey. The Schechter function has essentially the same parameters as the real data (see Norberg et al. 2002a). Then the completeness mask and variable apparent magnitude limits were applied to the mock catalogues to reproduce catalogues similar to the real data.

In the analysis which follows we make use of the real- and redshift-space correlation functions from the full Hubble Volume simulation. These correlation functions are determined from a Fourier transform of the power spectrum of the full Hubble Volume cube using the real- and redshift-space positions of the mass particles respectively, along with the bias scheme outlined above. This allows us to compare our mock catalogue results with that of the simulation from which they are drawn to ensure we can reproduce the correct parameters. It also allows us to compare and contrast the results from the real Universe with a large numerical simulation.

5.1.7 Error estimates

We analyse each of the mock catalogues in the same way as the real data, so that we have 22 mock measurements for every measurement that we make on the real data. The standard deviation between the 22 mock measurements gives a robust estimate of the uncertainty on the real data. We use this approach to estimate the uncertainties for direct measurements from the data, such as the individual points in the correlation function, and for best-fit parameters such as s_0.

When fitting parameters we use this standard deviation as a weight for each datapoint and perform a minimum χ^2 analysis to obtain the best fit parameter. The errors that we quote for any particular parameter are the rms spread between the 22 best fit parameters obtained in the same way from the mock catalogues. This simple way of estimating the uncertainties avoids the complications of dealing directly with correlated errors in measured data points, while still providing an unbiased estimate of the real uncertainties in the data, including the effects of correlated errors.

Although this approach gives reliable estimates of the uncertainties, the simple weighting scheme is not necessarily optimal in the presence of correlated errors. Nevertheless, for all statistics that we consider, we find that the means of the mock estimates agree well with the values input to the parent simulations. So, we are confident that or measurements and uncertainty estimates are robust and unbiased.

5.2 Estimates of the correlation function

The two point correlation function, ξ, is measured by comparing the actual galaxy distribution to a catalogue of randomly distributed galaxies. These randomly distributed galaxies are subject to the same redshift, magnitude and mask constraints as the real data and we modulate the surface density of points in the random catalogue to follow the completeness variations. We count the pairs in bins of separation along the line of sight, π, and across the line of sight, σ, to estimate $\xi(\sigma, \pi)$. Spherically averaging these pair counts provides the redshift-space correlation function $\xi(s)$. Finally we estimate the projected function $\Xi(\sigma)$ by integrating over all velocity separations along the line of sight and invert it to obtain $\xi(r)$.

5.2.1 Constructing a random catalogue

To reduce shot noise we compare the data with a random catalogue containing ten times as many points as the real catalogue. This random catalogue needs to have a smooth selection function matching the $N(z)$ of the real data. We use the 2dFGRS luminosity function (Norberg et al. 2002a) with $M^*_{b_J} = -19.66$ and $\alpha = -1.21$ to generate the selection function, following the change in the survey magnitude limit across the sky. When analysing the mock catalogues we use the input luminosity function to generate the selection function, and hence random catalogue.

As an alternative method, we also fitted an analytic form for the selection function (Baugh & Efstathiou 1993) to the data, and generated random catalogues using that selection function. We have calculated all of our statistics using both approaches, and found that they gave essentially identical results for the data. When analysing the mock catalogues, we found that the luminosity function

method was more robust to the presence of large-scale features in the $N(z)$ data. Thus, all of our quoted results are based on random catalogues generated using the luminosity function.

5.2.2 Fibre collisions

The design of the 2dF instrument means that fibres cannot be placed closer than approximately 30 arcsec (Lewis et al. 2002), and so both members of a close pair of galaxies cannot be targeted in a single fibre configuration. Fortunately, the arrangement of 2dFGRS tiles means that not all close pairs are lost from the survey. Neighbouring tiles have significant areas of overlap, and so much of the sky is targeted more than once. This allows us to target both galaxies in some close pairs. Nevertheless, the survey misses a large fraction of close pairs. It is important to assess the impact of this omission on the measurement of galaxy clustering and to investigate schemes that can compensate for the loss of close pairs.

To quantify the effect of these so-called 'fibre collisions' we have calculated the angular correlation function[1] for galaxies in the 2dFGRS parent catalogue, $w_p(\theta)$ and for galaxies with redshifts used in our ξ analysis, $w_z(\theta)$. We used the same mask to determine the angular selection and apparent magnitude limit for each sample as in Figure 5.5. Note that the mask is used only to define the area of analysis, and the actual redshift completeness values are not used in the calculation of w. In our ξ analyses we impose redshift limits $0.01 < z < 0.2$, which means that the mean redshift of the redshift sample is lower than the parent sample. We used Limber's equation (Limber 1954) to calculate the scale factors in amplitude and angular scale needed to account for the different redshift distributions. The solid line in Figure 5.7 shows w_p, and the filled circles show w_z after applying the Limber scale factors. The error bars in Figure 5.7 show $w(\theta)$ from the full APM survey (Maddox, Efstathiou & Sutherland 1996), also scaled to the magnitude limit of the 2dFGRS parent sample. On scales $\theta \gtrsim 0.03^\circ$ all three measurements are consistent. On smaller scales w_z is clearly much lower than w_p, showing that the fibre collision effect becomes significant and cannot be neglected.

The ratio of galaxy pairs counted in the parent and redshift samples is given by

[1]This is the angular equivalent of $\xi(s)$ - i.e. using angular separation θ instead of s - thus it uses no redshift information.

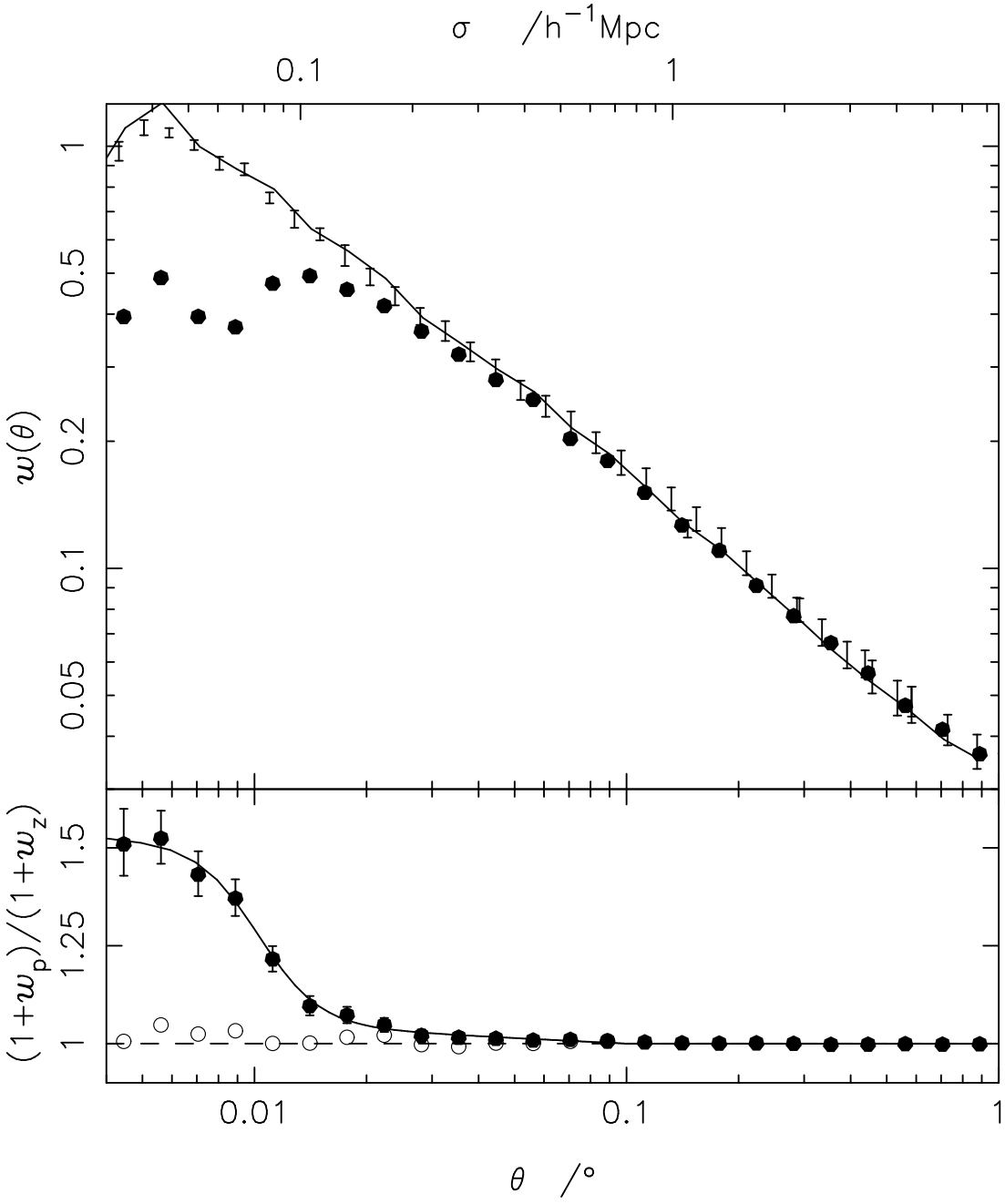


Figure 5.7: Top panel: $w(\theta)$ for the mean of the NGP and SGP redshift catalogues (solid points), the mean of the masked parent catalogues (solid line), and the full APM result (error bars). Bottom panel: The parent catalogue result divided by the redshift catalogue results (uncorrected - solid points; collision corrected (see Section 5.2.3) - open points). The error bars on the solid points are taken from the APM result and are hence only indicative of the errors we might expect. The solid line is the curve used to correct the fibre collisions and has the form shown in Equation 5.1. The top axis converts the angle θ into a projected separation, σ, at the effective redshift of the survey, $z_s = 0.15$.

$(1+w_p)/(1+w_z)$, which is shown by the filled circles in the lower panel of Figure 5.7. As discussed in the next section, we use this ratio to correct the pair counts in the ξ analysis.

5.2.3 Weighting

Each galaxy and random galaxy is given a weighting factor depending on its redshift and position on the sky. The redshift dependent part of the weight is designed to minimize the variance on the estimated ξ (Efstathiou 1988), and is given by $1/(1+4\pi n(z_i)J_3(s))$, where $n(z)$ is the density distribution and $J_3(s) = \int_0^s \xi(s')s'^2 ds'$. We use $n(z)$ from the random catalogue to ensure that the weights vary smoothly with redshift. We find that our results are insensitive to the precise form of J_3 but we used a power law ξ with $s_0 = 13.0$ and $\gamma_s = 0.75$ and a maximum value of $J_3 = 400$. This corresponds to the best-fit power law over the range $0.1 < s < 3\ h^{-1}$Mpc with a cutoff at larger scales.

We also use the weighting scheme to correct for the galaxies that are not observed due to the fibre collisions. Each galaxy-galaxy pair is weighted by the ratio $w_f = (1+w_p)/(1+w_z)$ at the relevant angular separation according to the curve plotted in the bottom panel of Figure 5.7 which has the form,

$$w_f = 1.2 - 0.23\tanh(6.1\log(\theta) + 12.1) - 0.05\log(\theta) \tag{5.1}$$

This corrects the observed pair count to what would have been counted in the parent catalogue. The open points in Figure 5.7, which have the collision correction applied, show that this method can correctly recover the parent catalogue result and hence overcome the fibre collision problem. Since the random catalogues do not have any close-pair constraints, only the galaxy-galaxy pair count needs correcting in this way. We also tried an alternative approach to the fibre-collision correction that we used previously in Norberg et al. (2001, 2002b) where the weight for each unobserved galaxy was assigned equally to its ten nearest neighbours. This produced similar results for $\theta > 0.03°$, but did not help on smaller scales. All of our results are presented using the w_f weighting scheme.

Hence each galaxy, i, is weighted by the factor,

$$w_i = \frac{1}{1+4\pi n(z_i)J_3(s)}, \tag{5.2}$$

and each galaxy-galaxy pair i,j is given a weight $w_f w_i w_j$, whereas each galaxy-random and random-random pair is given a weight $w_i w_j$.

5.2.4 The two-point correlation function, $\xi(\sigma, \pi)$

We use the ξ estimator of Landy & Szalay (1993),

$$\xi(\sigma, \pi) = \frac{DD - 2DR + RR}{RR}, \tag{5.3}$$

where DD is the normalised sum of weights of galaxy-galaxy pairs with particular (σ, π) separation, RR the normalised sum of weights of random-random pairs with the same separation in the random catalogue and DR the normalised sum of weights of galaxy-random pairs with the same separation. To normalise the pair counts we ensure that the sum of weights of the random catalogue equal the sum of weights of the real galaxy catalogue, as a function of scale. We find that other estimators (e.g. Hamilton 1993) give similar results.

The $N(z)$ distributions for the data and random catalogues (scaled so that the area under the curve is the same as for the observed data) are shown in Figure 5.8. It is clear that $N(z)$ for the random catalogues are a reasonable smooth fit to $N(z)$ for the data. Norberg et al. (2002a) showed that large 'spikes' in the $N(z)$ are common in the mock catalogues, and so similar features in the data indicate normal structure.

The resulting estimates of $\xi(\sigma, \pi)$ calculated separately for the SGP and NGP catalogues are shown in Figure 5.9, along with the combined result. The velocity distortions are clear at both small and large scales, and the signal-to-noise ratio is in general very high for σ and π values less than 20 h^{-1}Mpc; it is ≈ 6 in each 1 h^{-1}Mpc bin at $s = 20$ h^{-1}Mpc. At very large separations $\xi(\sigma, \pi)$ becomes very close to zero, showing no evidence for features that could be attributed to systematic photometric errors.

We used an earlier version of the 2dFGRS catalogue to carry out a less detailed analysis of $\xi(\sigma, \pi)$ (Peacock et al. 2001). The current redshift sample has 1.4 times as many galaxies, though more importantly it is more contiguous, and the revised photometry has improved the uniformity of the sample. Nevertheless our new results are very similar to our earlier analysis, demonstrating the robustness of our results. The current larger sample allows us to trace ξ out to larger scales with smaller uncertainties. Also, in the present analysis we analyse mock catalogues to obtain error estimates which are more precise than the previous error approximation (see Section 5.5.3).

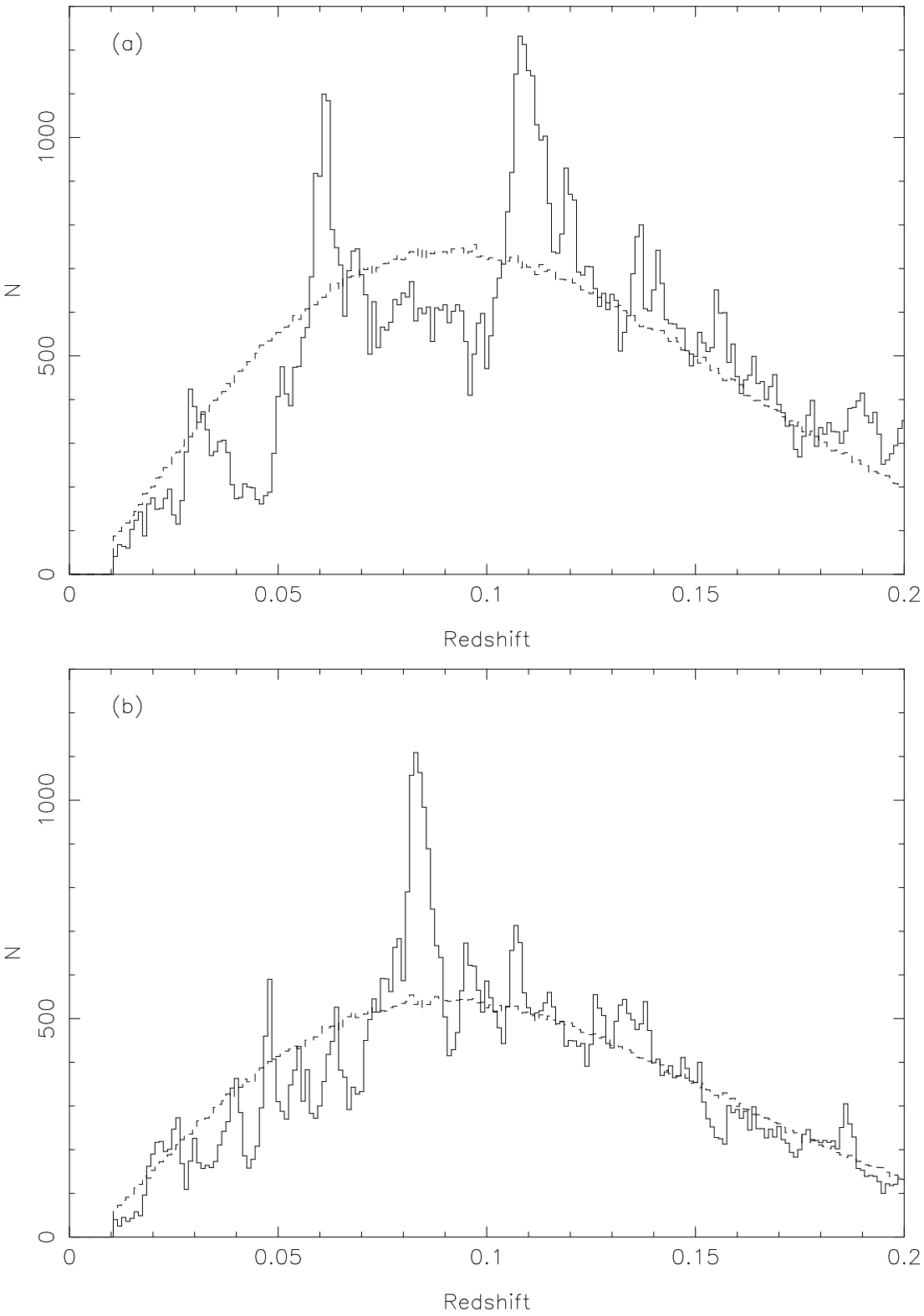


Figure 5.8: Redshift distributions, $N(z)$, for the 2dFGRS data (solid lines) and the normalised random catalogues generated using the survey luminosity function (dashed lines) for the (a) SGP and (b) NGP.

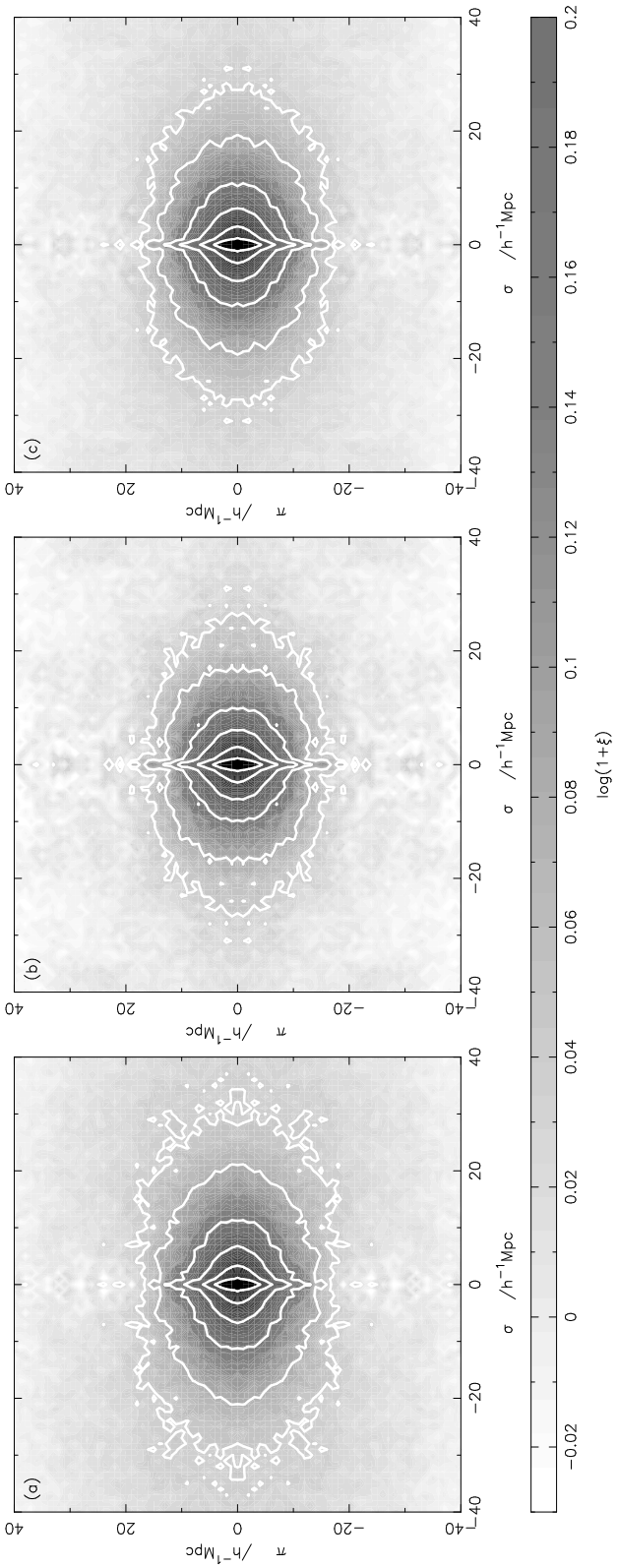


Figure 5.9: Grey-scale plots of the 2dFGRS $\xi(\sigma, \pi)$ (in 1 h^{-1}Mpc bins) for (a) the SGP region, (b) the NGP region and (c) the combined data. Contours are overlaid at $\xi = 4.0, 2.0, 1.0, 0.5, 0.2$ and 0.1.

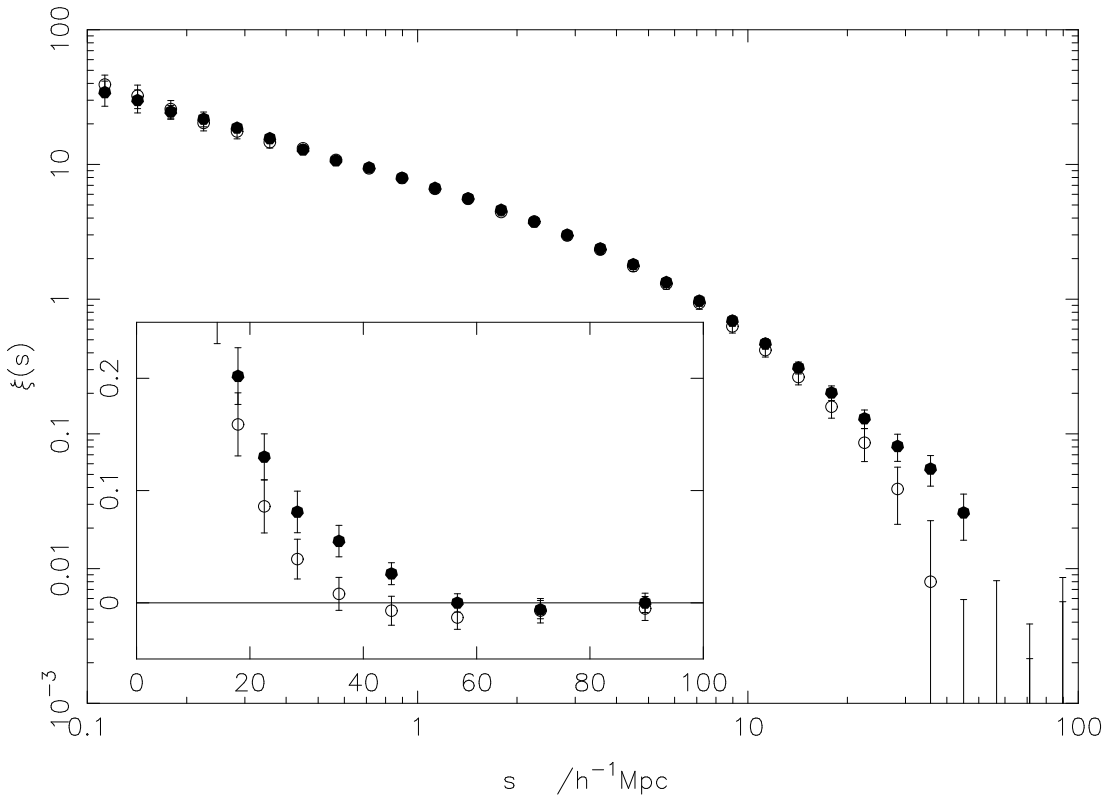


Figure 5.10: The redshift-space correlation function for the NGP (open points) and SGP (solid points) 2dFGRS data with error bars from the rms of mock catalogue results. Inset plotted on a linear scale.

5.2.5 The redshift-space correlation function, $\xi(s)$

Averaging $\xi(\sigma, \pi)$ at constant s gives the redshift-space correlation function, $\xi(s)$, and our results for the NGP and SGP are plotted in Figure 5.10 on both log and linear scales. The NGP and SGP measurements differ by about 2σ between 20 and 50 h^{-1}Mpc, and we find one mock whose NGP and SGP measurements disagree by this much, and so it is probably not significant. We tried shifting M^* by 0.1 mag to better fit the $N(z)$ at $z > 0.15$ in the SGP, and this moved the data points by $\sim 0.2\sigma$ for $20 < s < 50\ h^{-1}$Mpc. The $\xi(s)$ for the combined data is plotted in the top panel of Figure 5.11.

The $\xi(s)$ for the combined data is plotted in the top panel of Figure 5.11 and it is clear that the measured $\xi(s)$ is not at all well represented by a universal power law on all scales, but we do make an estimate of the true value of the

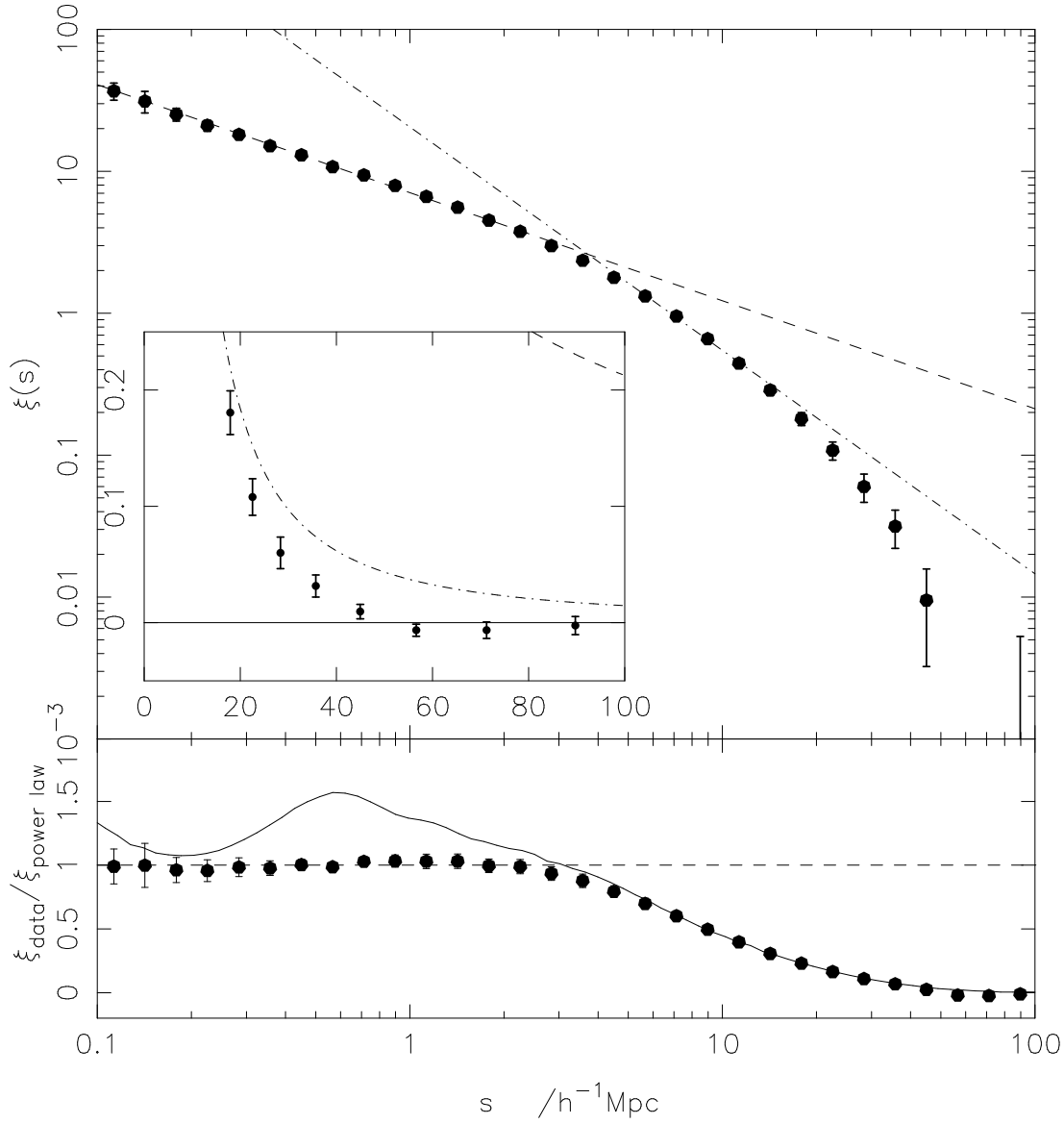


Figure 5.11: Top panel: The redshift-space correlation function for the combined data (points) with error bars from the rms of the mock catalogue results. The dashed line is a small scale power law fit, ($s_0 = 13\ h^{-1}$Mpc, $\gamma_s = 0.75$) and the dot-dashed line is the best fit to points around s_0, ($s_0 = 6.81\ h^{-1}$Mpc, $\gamma_s = 1.56$). Inset is on a linear scale. Bottom panel: As above, divided by the small scale power law. The solid line shows the result from the Hubble Volume simulation.

redshift-space correlation length, s_0, by fitting a localised power law of the form,

$$\xi(s) = \left(\frac{s}{s_0}\right)^{-\gamma_s} \tag{5.4}$$

using a least-squares fit to $\log(\xi)$ as a function of $\log(s)$, using two points either side of $\xi(s) = 1$. This also gives a value for the local redshift-space slope, γ_s. The best fit parameters for the separate poles and combined estimates are listed in Table 5.1.

At large separations, $s > 50\ h^{-1}$Mpc, the correlation function is approximately zero, helping to demonstrate that there are no significant systematic biases in the 2dFGRS. To conserve probability,

$$J_3(\infty) = \int_0^\infty \xi(s) s^2\ ds \equiv 0 \tag{5.5}$$

and so $\xi(s)$ must go negative on large scales to balance out the strong clustering on small scales. We can see this 'zero crossing' in our estimate of $\xi(s)$ (inset of Figure 5.11) at $s \approx 50\ h^{-1}$Mpc, though it has a low significance. Integrating $\xi(s)$ we measure $J_3(100\ h^{-1}\text{Mpc}) = 940 \pm 1720$, which is consistent with zero.

In the bottom panel of Figure 5.11 we examine the shape of $\xi(s)$ more carefully. The points are the data divided by a small scale power law fitted on scales $0.1 < s < 3\ h^{-1}$Mpc (dashed line). The data are remarkably close to the power-law fit for scales $0.1 < s < 3\ h^{-1}$Mpc, and follow a smooth break towards zero for $3 < s < 60\ h^{-1}$Mpc. The measurements from the Hubble Volume simulations are shown by the solid line, and they match the data extremely well on scales $s > 4\ h^{-1}$Mpc. On smaller scales, where the simulation has little control over the form of the clustering amplitude (as discussed in Section 5.1.6) there are discrepancies of order 50%.

The mean $\xi(s)$ determined from the mock catalogues agrees well with the true redshift-space correlation function from the full Hubble Volume. This provides a good check that our weighting scheme and random catalogues have not introduced any biases in the analysis.

5.2.6 Redshift-space comparisons

Redshift-space correlation functions have been measured from many redshift surveys, but direct comparisons between different surveys must be treated with caution because galaxy clustering depends on the spectral type and luminosity of

galaxies (e.g. Guzzo et al. 2000; Norberg et al. 2002b; Madgwick et al. 2002a). Direct comparisons can be made only between surveys that are based on similar galaxy selection criteria. The 2dFGRS is selected using pseudo-total magnitudes in the $b_{\rm J}$ band, and the three most similar surveys are the Stromlo-APM survey (SAPM, Loveday et al. 1992), the Durham UKST survey (Ratcliffe et al. 1998) and the ESO Slice Project (ESP, Guzzo et al. 2000). The Las Campanas Redshift Survey (LCRS, Lin et al. 1996; Jing, Mo & Börner 1998) and Sloan Digital Sky Survey (SDSS, Zehavi et al. 2002) are selected in the R band, but have a very large number of galaxies, and so are also interesting to compare with.

The non-power-law shape of $\xi(s)$ makes it difficult to compare different measurements of s_0 and γ_s, because the values depend sensitively on the range of s used in the fitting procedure. In Figure 5.12(a) we compare the $\xi(s)$ measurements directly for the 2dFGRS, SAPM, Durham UKST and ESP surveys. Our estimate of $\xi(s)$ is close to the mean of previous measurements, but the uncertainties are much smaller. Although we quote uncertainties that are similar in size to previous measurements, we have used the scatter between mock catalogues to estimate them, rather that Poisson or boot-strap estimates that have been used before. These methods seriously underestimate the true uncertainties.

Figure 5.12(b) shows the 2dFGRS measurements together with the LCRS and SDSS measurements. On scales $s \gtrsim 4\ h^{-1}$Mpc there appears to be no significant differences between the surveys, but for $s \lesssim 4\ h^{-1}$Mpc the LCRS and SDSS have a higher amplitude than the 2dFGRS. This difference is likely to be caused by the different galaxy selection for the surveys, though the SDSS results shown are for the Early Data Release (EDR) and have larger errors than the 2dFGRS points. The 2dFGRS is selected using $b_{\rm J}$, whereas the SDSS and LCRS are selected in red bands. Since the red (early type) galaxies are more strongly clustered than blue (late type) galaxies (e.g. Zehavi et al. 2002; and via spectral type, Norberg et al. 2002b) we should expect that ξ will be higher for red selected surveys than a blue selected survey. This is examined further in Chapter 6.

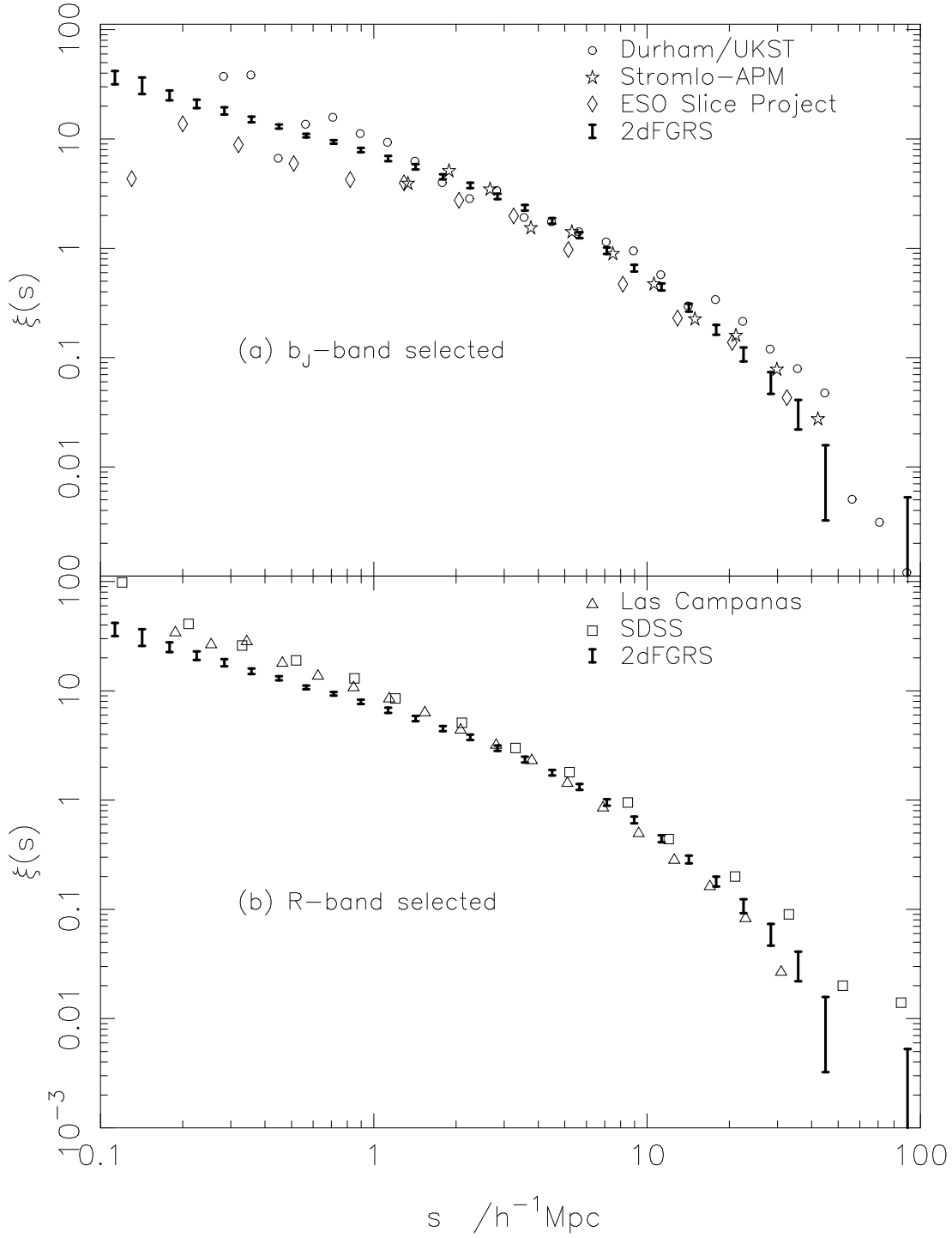


Figure 5.12: Comparison of 2dFGRS $\xi(s)$ with (a) other $b_{\rm J}$ band selected surveys and (b) R band selected surveys as indicated, and discussed in the text.

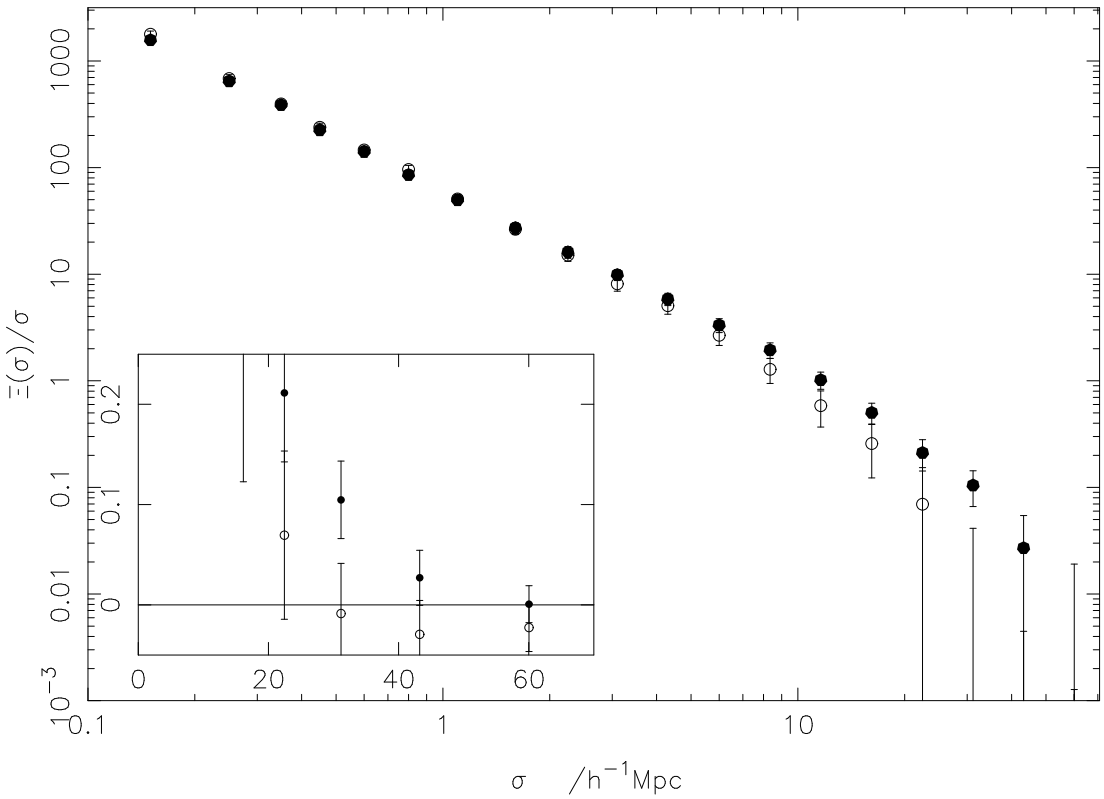


Figure 5.13: The projected correlation functions for the NGP (open points) and SGP (solid points) 2dFGRS data with error bars from the rms spread between mock catalogue results. Inset plotted on a linear scale.

5.2.7 The projected correlation function, $\Xi(\sigma)/\sigma$

The redshift-space correlation function differs significantly from the real-space correlation function because of redshift-space distortions. We can estimate the real-space correlation length, r_0, by first calculating the projected correlation function, $\Xi(\sigma)$ (see Section 2.2.2).

The projected correlation functions for the NGP and SGP are shown in Figure 5.13 and the combined data result is shown in Figure 5.14. The best fit values of γ_r^P and r_0^{P2} for $0.1 < \sigma < 12$ h^{-1}Mpc are shown in Table 5.1. Over this range $\Xi(\sigma)/\sigma$ is an accurate power law, but steepens for $\sigma > 12$ h^{-1}Mpc. This deviation from power-law behaviour limits the scales that can be probed using this approach.

[2]The 'P' superscripts refer to these 'Projected' values, rather than the 'Inverted' values denoted by 'I', as described in the next section

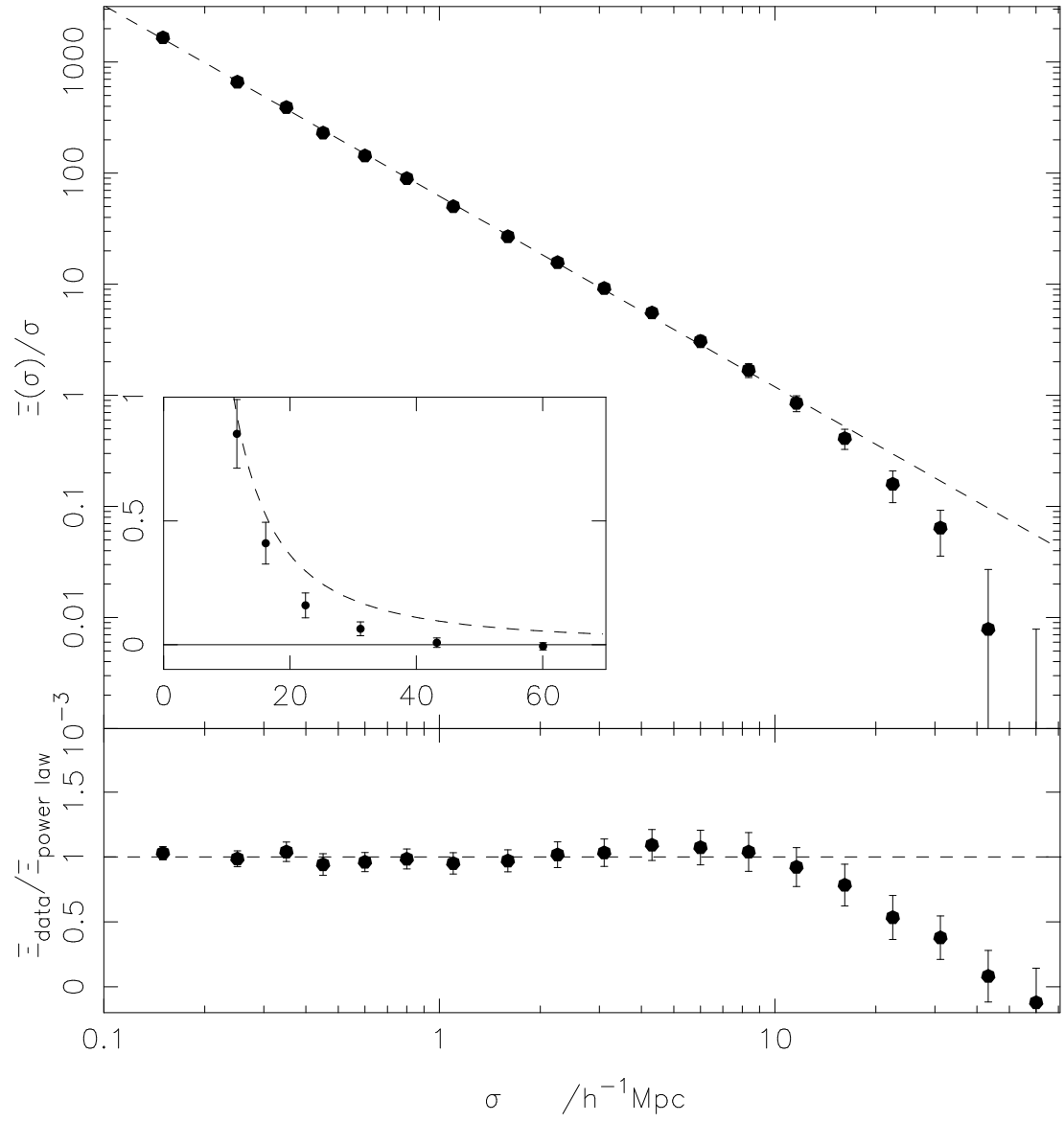


Figure 5.14: Top panel: The projected correlation function of the combined 2dF-GRS data with error bars from the rms spread between mock catalogue results. The dashed line is the best fit power law for $0.1 < \sigma < 12$ h^{-1}Mpc ($r_0 = 4.98$, $\gamma_r = 1.72$, $A = 3.97$). Inset is plotted on a linear scale. Bottom panel: The combined data divided by the power law fit.

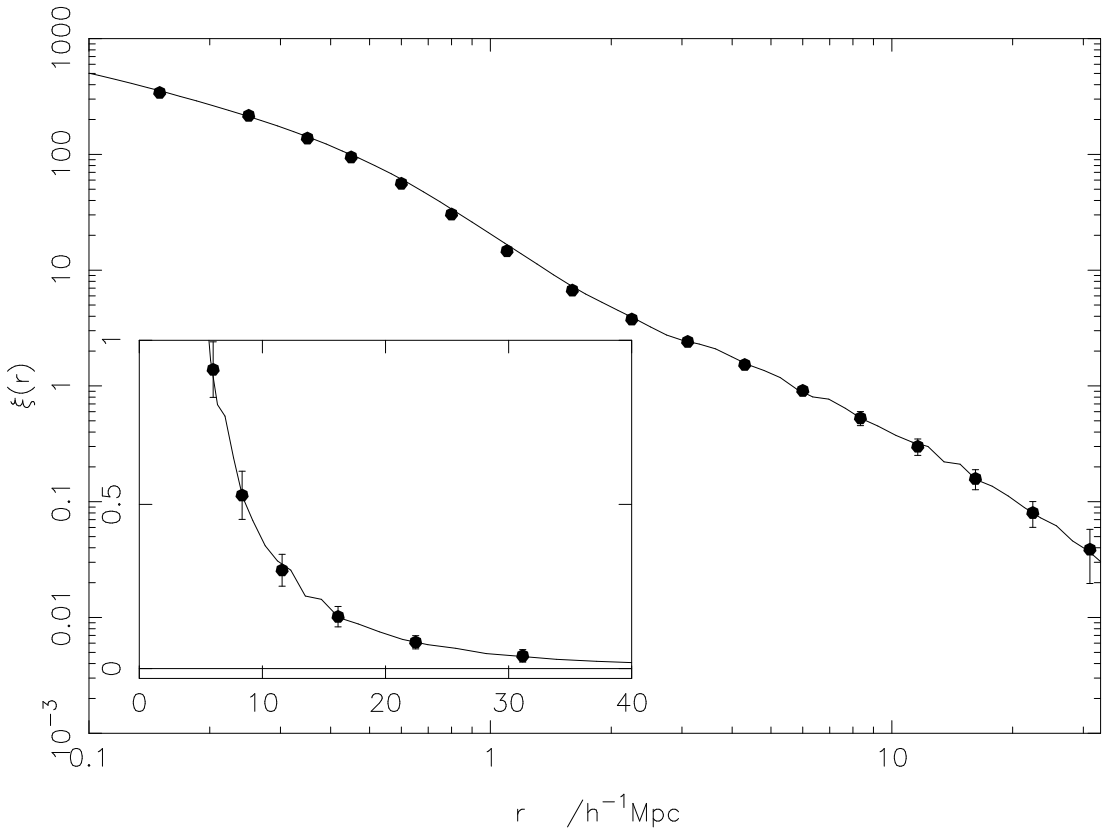


Figure 5.15: The mean real-space correlation function determined from the 22 mock catalogues using the methods of S92. The solid line is the true $\xi(r)$ from the Hubble Volume. The agreement is excellent. Note the changed scale from previous plots.

5.2.8 The real-space correlation function, $\xi(r)$

It is possible to estimate $\xi(r)$ by directly inverting $\Xi(\sigma)$ without making the assumption that it is a power law (S92, Section 2.2.3). We can test the reliability of our inversion of the 2dFGRS data using the mock catalogues.

In Figure 5.15 we show the mean $\xi(r)$ as determined from the mock catalogues using the methods of S92. We compare this to the real-space correlation function determined directly from the Hubble Volume simulation, from which the mock catalogues are drawn. The agreement is excellent and shows that the method works and that we can recover the real-space correlation function out to 30 h^{-1}Mpc.

We have applied this technique to the combined 2dFGRS data and obtain the

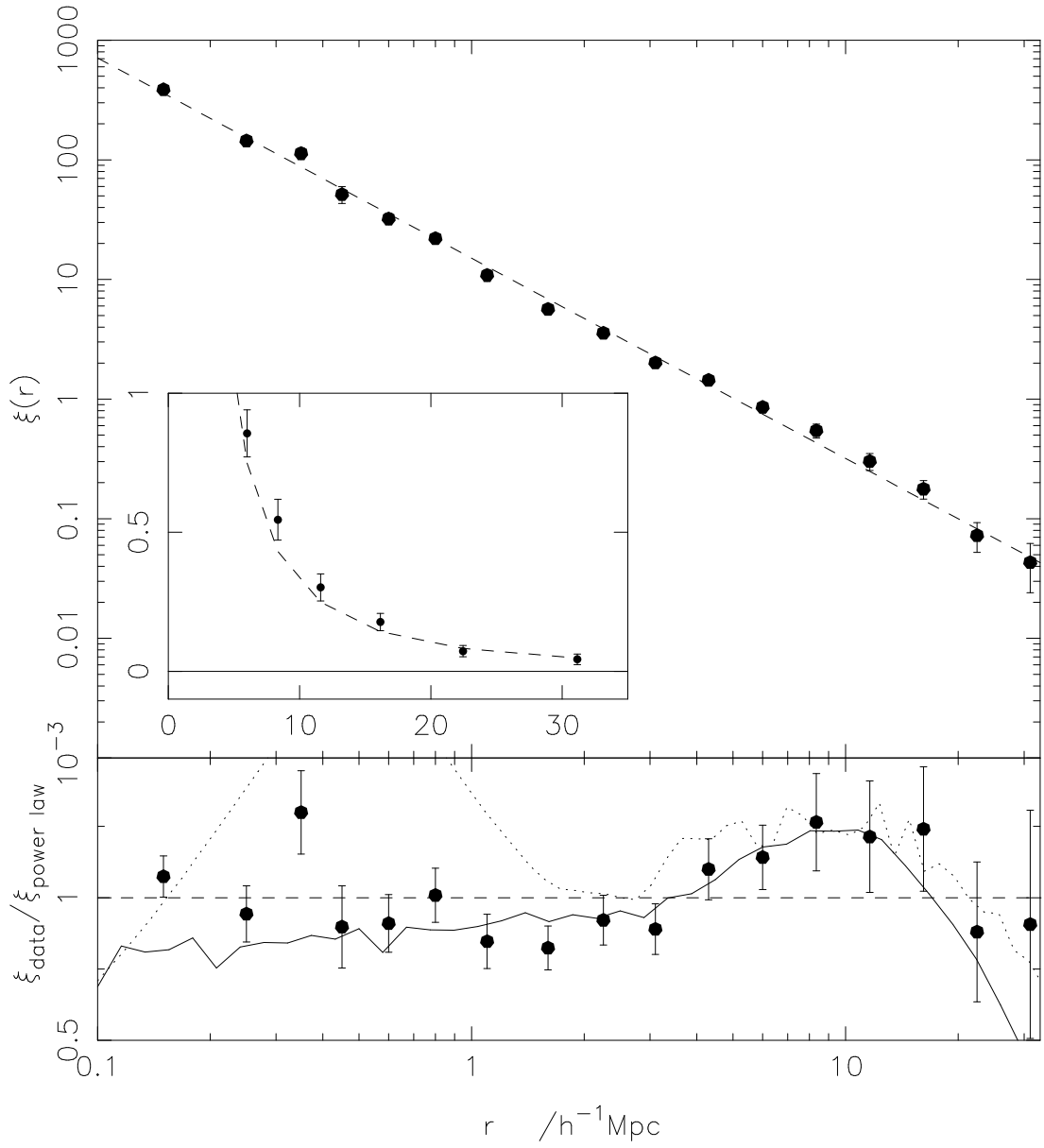


Figure 5.16: Top panel: The real-space correlation function of the combined 2dFGRS with error bars from the rms spread between mock catalogues. The dashed line is the best fit power law ($r_0 = 5.07$, $\gamma_r = 1.67$). Inset is plotted on a linear scale. Bottom panel: The data divided by the power law fit. The solid line is the de-projected APM result (Padilla & Baugh 2003) as discussed in the text. The dotted line is the result from the Hubble Volume.

real-space correlation function as shown in Figure 5.16. The data is plotted out to only 30 h^{-1}Mpc due to the limitations in the method described in Section 2.2.3. On small scales $\xi(r)$ is well represented by a power law, and a best fit over the range $0.1 < r < 12\ h^{-1}$Mpc gives results for r_0^I and γ_r^I as shown in Table 5.1.

The points in the bottom panel of Figure 5.16 show the 2dFGRS data divided by the best fit power law. It can be seen that at scales $0.1 < r < 20\ h^{-1}$Mpc the data $\xi(r)$ is close to the best-fit power-law but does show hints of non power-law behaviour (see also discussion below).

5.2.9 Real-space comparisons

In the inverted $\xi(r)$ (and possibly $\Xi(\sigma)$) there is a weak excess of clustering over the power law for $5 < r < 20\ h^{-1}$Mpc. This has been previously called a 'shoulder' in ξ (see e.g. Ratcliffe et al. 1998). Though the amplitude level of the feature in our data is rather low, it has been consistently seen in different surveys, and probably is a real feature.

The dotted line in the bottom panel of Figure 5.16 shows the Hubble Volume simulation which agrees well with the data for $r > 1\ h^{-1}$Mpc. On smaller scales the Hubble Volume ξ shows significant deviations from a power law. Since the galaxy assignment procedure in the mock catalogues uses the mass distribution smoothed on a scale of $2\ h^{-1}$Mpc the model is not able to produce a reliable ξ at such small scales, as discussed in Section 5.1.6.

The solid line in the bottom panel of Figure 5.16 is the de-projected APM result (Padilla & Baugh 2003), scaled down by a factor $(1 + z_s)^{\alpha}$, with $\alpha = 1.7$, suitable for evolution in a ΛCDM cosmology. There is good agreement between the 2dFGRS and APM results which are obtained using quite different methods.

We have estimated r_0 and γ_r by fitting to the projected correlation function $\Xi(\sigma)/\sigma$, and also by inverting $\Xi(\sigma)/\sigma$ and then fitting to $\xi(r)$. The best fit values from the two methods are shown in Table 5.2, and it is clear they lead to very similar estimates of r_0 and γ_r. This confirms that the power-law assumption in Section 5.2.7 is a good approximation over the scales we consider.

Table 5.2 lists r_0 and γ_r for the 2dFGRS and other surveys estimated using power law fits to the projected correlation function $\Xi(\sigma)$. As mentioned in Section 5.2.5 the SAPM, Durham UKST and the ESP are $b_{\rm J}$ selected surveys, and so should

be directly comparable to the 2dFGRS. The values of r_0 and γ_r for these surveys all agree to within one standard deviation, except that r_0 for the ESP appears to be significantly lower. It is likely that the quoted uncertainties for the ESP and Durham UKST parameters are underestimated since they did not include the effect of cosmic variance. Since they each sample relatively small volumes, this will be quite a large effect. The sparse sampling strategy used in the SAPM means that it has a large effective volume, and so the cosmic variance is small.

The red-selected surveys, LCRS and SDSS, are significantly different from the other surveys. The discrepancies are most likely due to the fact that the amplitude of galaxy clustering depends on galaxy type, and that red-selected surveys have a different mix of galaxy types. The extra surface brightness selection applied to the LCRS may also introduce significant biases. A comparison with the PSCz survey (Section 3.3.3) shows that infra-red selected galaxies seem to have the lowest clustering strength of all the samples.

Each survey is also likely to have a different effective luminosity and as has been shown by Norberg et al. (2001), this will cause clustering measurements to differ. The relation for 2dFGRS galaxies found by Norberg et al. (2001) was,

$$\left(\frac{r_0}{r_0^*}\right)^{\frac{\gamma_r}{2}} = 0.85 + 0.15\left(\frac{L}{L^*}\right), \tag{5.6}$$

which gives, for $L = 1.4L^*$ (see Section 5.1.5), $r_0^* = 4.71 \pm 0.24$, which will allow direct comparisons with other surveys.

5.3 Estimating β

Before using the 2-d modelling described in Chapter 2 to measure the clustering parameters simultaneously we first use methods that have been used in previous studies (see Section 2.4). This allows a direct comparison between our results and previous work.

5.3.1 Ratio of ξ's

The ratio of the redshift-space correlation function, $\xi(s)$, to the real-space correlation function, $\xi(r)$, in the linear regime gives an estimate of the redshift distortion

Table 5.1: Best fit parameters to ξ. For s_0 and γ_s the fit to $\xi(s)$ uses only points around $s = s_0$. For r_0^P, γ_r^P and $\mathrm{A}(\gamma_r^P)$ the fit to $\Xi(\sigma)/\sigma$ uses all points $0.1 < \sigma < 12\ h^{-1}$Mpc. For r_0^I and γ_r^I the fit to the inverted $\xi(r)$ uses all points $0.1 < r < 12\ h^{-1}$Mpc. In each case the errors quoted are the rms spread in the results obtained from the same analysis with the mock catalogues.

Parameter	SGP	NGP	Combined
s_0 (h^{-1}Mpc)	6.92 ± 0.36	6.72 ± 0.41	6.82 ± 0.28
γ_s	1.51 ± 0.08	1.64 ± 0.08	1.57 ± 0.07
r_0^P (h^{-1}Mpc)	5.05 ± 0.32	4.79 ± 0.31	4.95 ± 0.25
γ_r^P	1.68 ± 0.06	1.77 ± 0.07	1.72 ± 0.04
$\mathrm{A}(\gamma_r^P)$	4.17 ± 0.23	3.77 ± 0.28	3.99 ± 0.16
r_0^I (h^{-1}Mpc)	5.09 ± 0.35	5.08 ± 0.28	5.05 ± 0.26
γ_r^I	1.65 ± 0.03	1.70 ± 0.04	1.67 ± 0.03

Table 5.2: Measurements of $\xi(r)$ from 2dFGRS and other surveys as published. The various authors have used very different ways to estimate errors, though none included the effects of cosmic variance, which dominate our estimates because we used the scatter between mock catalogues to take account of this.

Survey	$r_0(h^{-1}\mathrm{Mpc})$	γ_r
2dFGRS (P)	4.95 ± 0.25	1.72 ± 0.04
2dFGRS (I)	5.05 ± 0.26	1.67 ± 0.03
SAPM	5.1 ± 0.3	1.71 ± 0.05
ESP	4.15 ± 0.2	$1.67^{+0.07}_{-0.09}$
Durham UKST	5.1 ± 0.3	1.6 ± 0.1
LCRS	5.06 ± 0.12	1.86 ± 0.03
SDSS	6.14 ± 0.18	1.75 ± 0.03
PSCz	3.82 ± 0.52	1.67 ± 0.09

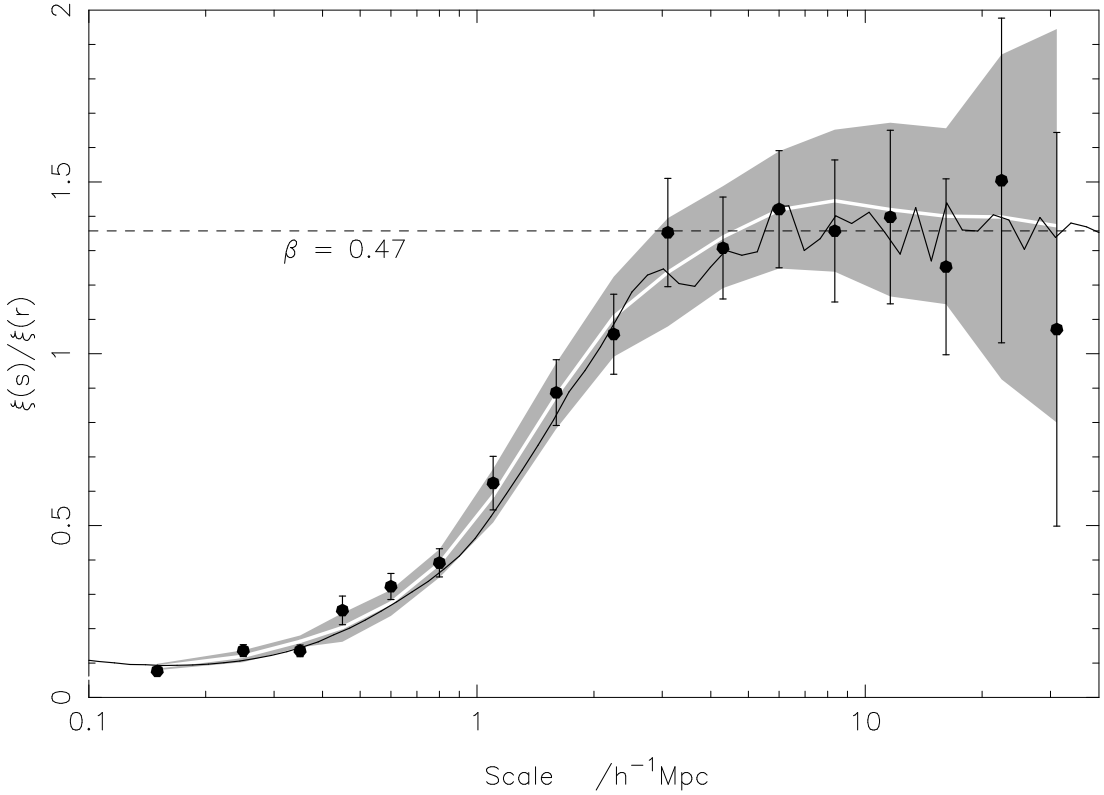


Figure 5.17: The ratio of $\xi(s)$ to $\xi(r)$ for the 2dFGRS combined data (solid points), the Hubble Volume (solid line) and the mean of the mock catalogue results is shown (white line), with the rms errors shaded. Error bars are from the rms spread in mock catalogue results.

parameter, β (see Equation 2.25, Hamilton 1992),

$$\frac{\xi(s)}{\xi(r)} = 1 + \frac{2\beta}{3} + \frac{\beta^2}{5} \tag{5.7}$$

Our results for the combined 2dFGRS data are shown in Figure 5.17 by the solid points, the mean of the mock catalogue results is shown by the white line, with the rms errors shaded and the estimate from the Hubble Volume is shown by the solid line. The data are consistent with a constant value on scales $\gtrsim 4\ h^{-1}$Mpc.

The mock catalogue and Hubble Volume results asymptote to $\beta = 0.47$, as expected. The 2dFGRS data between 8 and 30 h^{-1}Mpc are best fit by a ratio of 1.34 ± 0.13, corresponding to $\beta = 0.45 \pm 0.14$. The maximum scale that we can use in this analysis is determined by the uncertainty on $\xi(r)$ from the S92 inversion method used in Section 5.2.8.

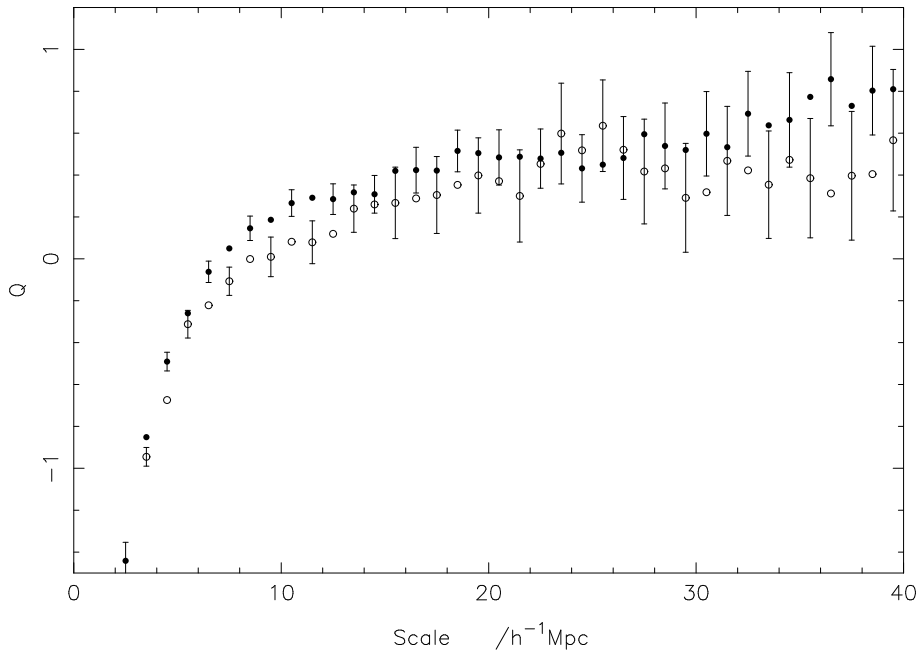


Figure 5.18: The Q factor for the NGP (solid points) and SGP(open points). The error bars are the rms spread in mock catalogue results and are placed alternately to avoid confusion.

5.3.2 The quadrupole moment of ξ

We now measure β using the quadrupole moment of the correlation function, (Hamilton 1992)

$$Q(s) = \frac{\frac{4}{3}\beta + \frac{4}{7}\beta^2}{1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2} = \frac{\xi_2(s)}{\frac{3}{s^3}\int_0^s \xi_0(s')s'^2 ds' - \xi_0(s)} \tag{5.8}$$

where ξ_ℓ is given by,

$$\xi_\ell(s) = \frac{2\ell+1}{2}\int_{-1}^{+1} \xi(\sigma,\pi)P_\ell(\mu)d\mu \tag{5.9}$$

These equations assume the random peculiar velocities are negligible and hence measuring Q gives an estimate for β. The random uncertainties in this method are small enough that we obtain reliable estimates on scales $< 40\ h^{-1}$Mpc, as shown by the mock catalogues (see below), but beyond these scales the data becomes noisy.

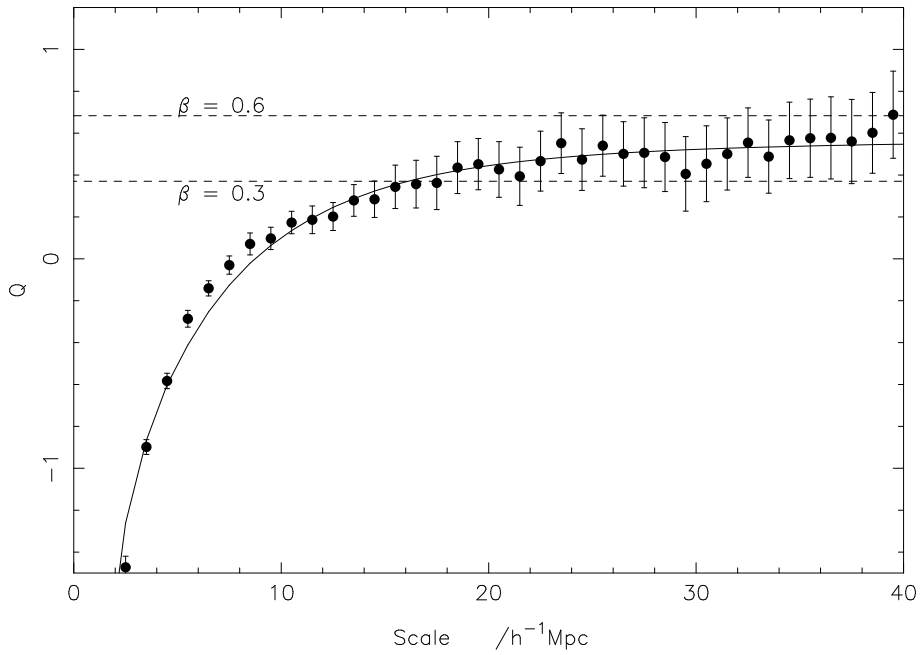


Figure 5.19: The Q factor for the combined 2dFGRS data, restricted to scales not dominated by noise with error bars from the rms spread in mock catalogues results. The two dashed lines show the expected answer for different values of β which approximate the 1σ errors. The solid line shows a model with $\beta = 0.49$ and $a = 506$ km s^{-1} (see Section 5.5.1).

The 2dFGRS data

Figure 5.18 shows Q estimates for the 2dFGRS data in the NGP and SGP separately, and Figure 5.19 shows the estimates for the combined 2dFGRS data. The effect of the random peculiar velocities can be clearly seen at small scales, causing Q to be negative. The best fit value to the combined data for $30 - 40$ h^{-1}Mpc is $Q = 0.55 \pm 0.18$, which gives a value for $\beta = 0.47^{+0.19}_{-0.16}$, where the error is from the rms spread in mock catalogue results. The solid line represents a model with $\beta = 0.49$ and $a = 506$ km s^{-1}, which matches the data well (see Section 5.5.1). Although asymptoting to a constant, the value of Q in the model is still increasing at 40 h^{-1}Mpc. This shows that non-linear effects do introduce a small systematic error even at these scales though this bias is small compared to the random error.

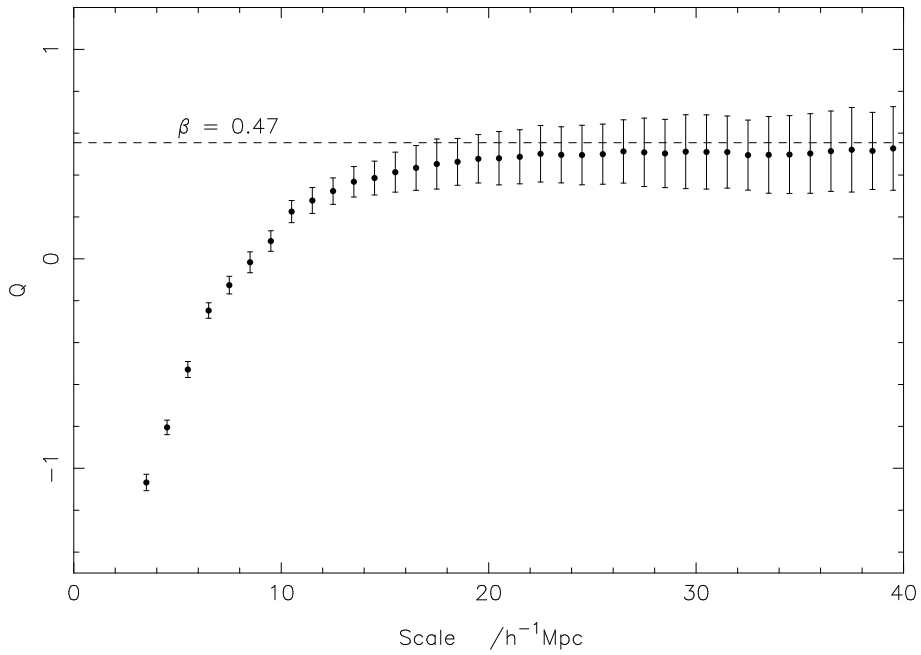


Figure 5.20: The mean Q factor for the mock catalogues with error bars from the rms spread in mock results. The dashed line is the true value of $\beta = 0.47$. The slight bias is caused by the random peculiar velocities as discussed in the text.

The mock catalogues

To check whether this method can correctly determine β we use the mock catalogues. The data points in Figure 5.20 are the mean values of Q from the mock catalogues, and the dashed line is the predicted value of $\beta = 0.47$. The data points seem to converge on large scales to the correct value of Q. Fitting to each mock catalogue in turn for $30 - 40$ h^{-1}Mpc gives a mean $Q = 0.51 \pm 0.18$, corresponding to $\beta = 0.43^{+0.18}_{-0.16}$. As the models showed (Section 2.4.2), the random velocities will cause us to underestimate β even at 40 h^{-1}Mpc, causing the difference between the measured and predicted values. This all shows that we can determine β with a slight bias but the error bars are large compared to the bias.

The Q estimates from the individual mock catalogues show a high degree of correlation between points on varying scales and so the overall uncertainty in Q from averaging over all scales > 30 h^{-1}Mpc is not much smaller than the uncertainty from a single point. It is this fact which makes the spread in results from the mock catalogues vital in the estimation of the errors on our result (also see Section 5.5.3).

5.4 The peculiar velocity distribution

To this point we have assumed that the random peculiar velocity distribution has an exponential form (Equation 2.29). This form has been used by many authors in the past and has been found to fit the data better than other forms (e.g. Ratcliffe et al. 1998). We test this for the 2dFGRS data by following a method similar to Landy, Szalay & Broadhurst (1998). To extract the peculiar velocity distribution we need to deconvolve the real-space correlation function from the peculiar velocity distribution (see Section 2.5 for the description of the methods).

5.4.1 The mock catalogues

The mean of the peculiar velocity distributions for the mock catalogues is shown in Figure 5.21. The distribution is approximately of the form expected for an exponential $f(v)$ with $a = 575\,\mathrm{km\,s^{-1}}$ with the offset caused by the $\beta = 0.47$ infall as shown by the solid line. The exact form of the peculiar velocities in the Hubble Volume simulation, and hence mock catalogues, is not specified and it should not be expected to conform to this model exactly.

5.4.2 The 2dFGRS data

The Fourier transform of the peculiar velocity distribution for the combined 2dFGRS data is shown in Figure 5.22 compared to a best fit model with $\beta = 0.49 \pm 0.05$ and $a = 570 \pm 25\,\mathrm{km\,s^{-1}}$, assuming an exponential form for $f(v)$. Figure 5.23 shows the peculiar velocity distribution itself compared to the same model. We showed in Section 2.5.3 that this was likely to be the value of a at small σ. The distribution of random pairwise velocities do appear to have an exponential form, with a β influence.

Ignoring the infall, L02 find that $a = 331\,\mathrm{km\,s^{-1}}$, using the smaller, publicly available, sample of 2dFGRS galaxies. We made the same approximations and repeated their procedure on our larger sample, and find $a = 370\,\mathrm{km\,s^{-1}}$. Using our data grid out to 70 h^{-1}Mpc, with no smoothing and ignoring β, gives $a = 457\,\mathrm{km\,s^{-1}}$. We have shown that the result in L02 is biased low by ignoring β and that the infall must be properly considered in these analyses.

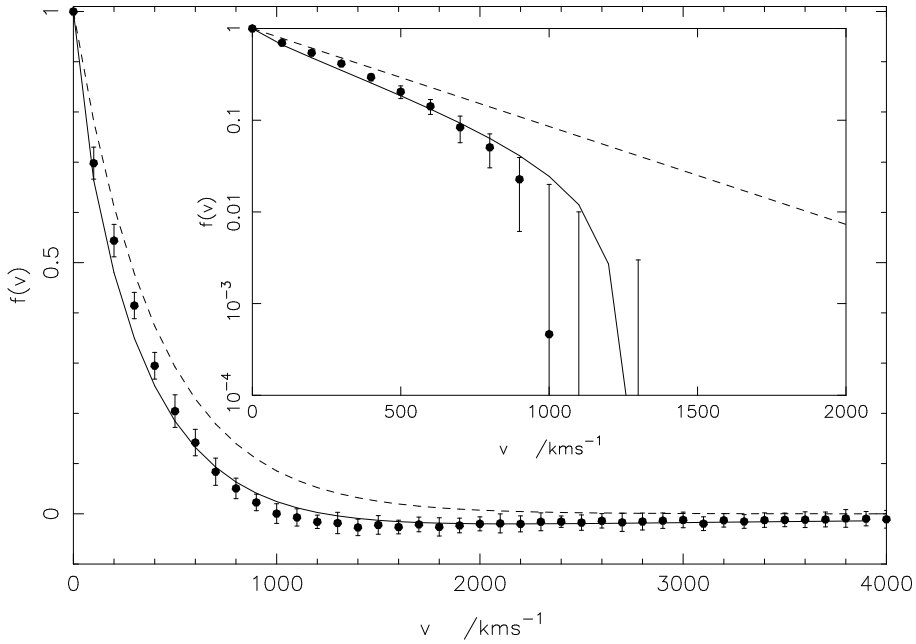


Figure 5.21: The recovered velocity distribution for the mock catalogues. Filled points are the mean result with error bars from the scatter between results. This is compared to an pure exponential distribution with $a = 575$ km s^{-1} (dashed line) and a model with $a = 575$ km s^{-1} and $\beta = 0.47$ (solid line).

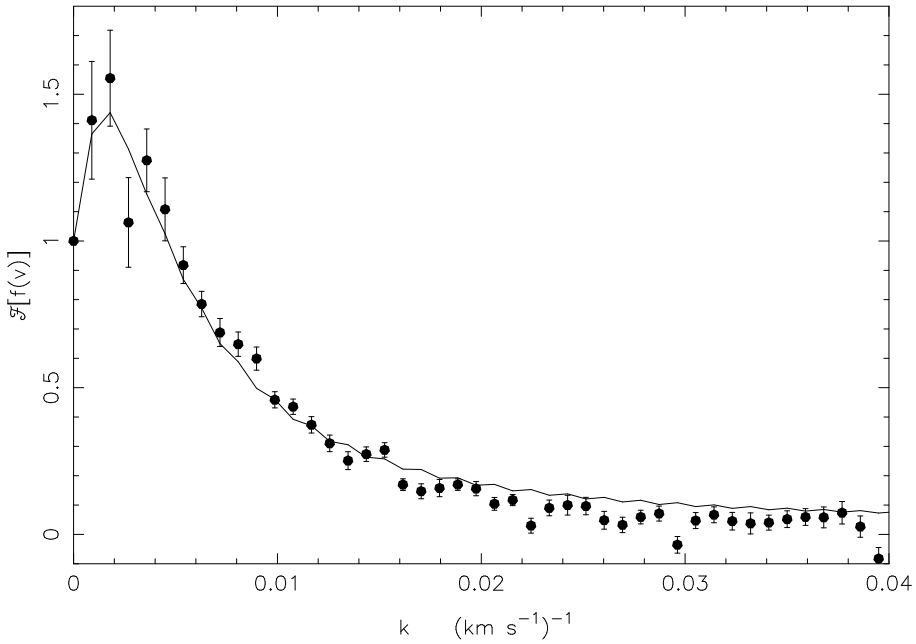


Figure 5.22: The Fourier transform of the peculiar velocity distribution for the combined 2dFGRS data (solid points) compared to a model distribution with $a = 570$ km s^{-1} and $\beta = 0.49$ (solid line).

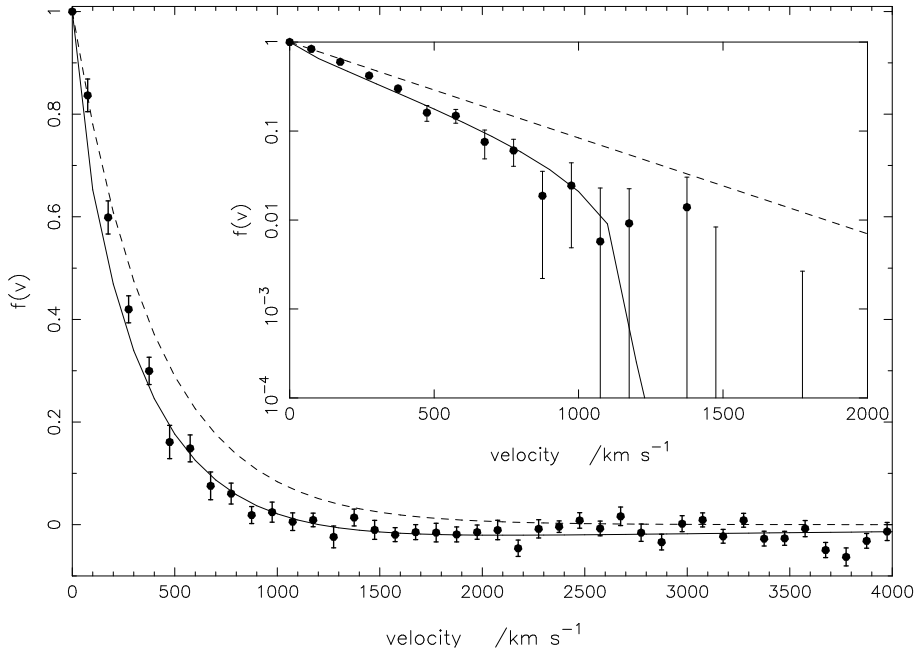


Figure 5.23: The combined 2dFGRS peculiar velocity distribution (solid points), compared to a pure exponential distribution with $a = 570$ km s^{-1} (dashed line) and a model with $a = 570$ km s^{-1} and $\beta = 0.49$ (solid line). The error bars are from the scatter in mock results.

5.4.3 Comparison and discussion

Previous estimates of $f(v)$ (e.g. Davis & Peebles 1983; Jing, Mo & Börner 1998) assumed a simple one-dimensional form for the effects of coherent infall velocities, and the results depend on the precise choice of model. With the 2dFGRS we can include the functional form of β, which we have shown to be essential, and obtain more accurate and robust results. We find that $a = 570 \pm 25$ km s^{-1} at small σ.

Sheth (1996) used an extended Press-Schechter model to show that an exponential pairwise peculiar velocity distribution is easily explained by the Press-Schechter (Press & Schechter 1974) theory of non-linear clustering and is fairly independent of power spectrum and the assumed density profile in clusters. The change of dispersion with scale was also expected to be small in this model.

Diaferio & Geller (1996) also found that an exponential pairwise peculiar velocity distribution is a result of well-defined gravitational processes. They concluded that the exponential pairwise peculiar velocity distribution occurs because of the presence of a large fraction of un-relaxed galaxy systems in the universe.

5.5 Fitting to the $\xi(\sigma, \pi)$ grid

5.5.1 Results

We now fit our $\xi(\sigma, \pi)$ data grid to the models described in Section 2.3, with a power-law form for the real-space correlation function. This model has four free parameters, β, r_0, γ_r and a. The fits to the data are done by minimising

$$\sum \left(\frac{\log[1+\xi]_{\rm model} - \log[1+\xi]_{\rm data}}{\log[1+\xi+\delta\xi]_{\rm data} - \log[1+\xi-\delta\xi]_{\rm data}} \right)^2 \qquad (5.10)$$

for $s < 20\ h^{-1}$Mpc, where $\delta\xi$ is the spread in ξ from the mock catalogues for a particular σ and π. This is essentially a simple χ^2 minimization. We tried a fit to ξ directly but found that it gave too much weight to the central regions and so instead we fit to $\log[1+\xi]$ so that the overall shape of the contours has an increased influence on the fit. The best fit model parameters are listed in Table 5.3.

There are two key assumptions made in the construction of these models. Firstly, although the contours match well at small scales, there are good reasons to believe that our linear theory model will not hold in the non-linear regime for $s < 8\ h^{-1}$Mpc. Secondly, the power law model for the correlation function is assumed and we have seen evidence that this is not completely realistic. Using non-power law forms will also allow us to probe to larger scales.

To test whether our result is robust to these assumptions we firstly reject the non-linear regime of $s < 8\ h^{-1}$Mpc. Then we use the shape of the Hubble Volume $\xi(r)$ instead of a power law, and finally we extend the maximum scale to $s = 30\ h^{-1}$Mpc. We showed in Section 5.2.8 that the Hubble Volume shape is a good match to the data over the range $8 < s < 30\ h^{-1}$Mpc (Equations 2.18-2.22 show how to perform the β infall without assuming a power law).

We find that the best fit parameters change very little with these changes but when using the Hubble Volume $\xi(r)$ the quality of the fit improves significantly. The best fit model is shown compared to the data in Fig. 5.24. Notice the excellent agreement on small scales even though they are ignored in the fitting process. The best-fit parameters are listed in Table 5.3, and we adopt these results as our final best estimates finding $\beta = 0.49 \pm 0.09$.

If we repeat our analysis on the mock catalogues we find a mean value of $\beta =$

Table 5.3: Best fit parameters to the $\xi(\sigma, \pi)$ grids with errors from the rms spread in mock catalogue results.

Parameter	SGP	NGP	Combined
Power law $\xi(r)$: $(0 < s < 20\ h^{-1}\mathrm{Mpc})$			
β	0.53 ± 0.06	0.48 ± 0.08	0.51 ± 0.05
r_0 (h^{-1}Mpc)	5.63 ± 0.26	5.52 ± 0.29	5.58 ± 0.19
γ_r	1.66 ± 0.06	1.76 ± 0.07	1.72 ± 0.05
a (km s^{-1})	497 ± 24	543 ± 26	522 ± 16
Power law $\xi(r)$: $(8 < s < 20\ h^{-1}\mathrm{Mpc})$			
β	0.45 ± 0.10	0.35 ± 0.12	0.49 ± 0.09
r_0 (h^{-1}Mpc)	6.03 ± 0.36	6.06 ± 0.41	5.80 ± 0.25
γ_r	1.74 ± 0.08	1.88 ± 0.10	1.78 ± 0.06
a (km s^{-1})	457 ± 49	451 ± 51	514 ± 31
Hubble Volume $\xi(r)$: $(8 < s < 20\ h^{-1}\mathrm{Mpc})$			
β	0.47 ± 0.12	0.50 ± 0.14	0.49 ± 0.10
a (km s^{-1})	446 ± 73	544 ± 67	495 ± 46
Hubble Volume $\xi(r)$: $(8 < s < 30\ h^{-1}\mathrm{Mpc})$			
β	0.48 ± 0.11	0.47 ± 0.13	0.49 ± 0.09
a (km s^{-1})	450 ± 81	545 ± 85	506 ± 52

0.475 ± 0.090 (cf. the true value of $\beta = 0.47$, Section 5.1.6), showing that we can correctly determine β using this type of fit. When fitting the mock catalogues it became clear that β and a are correlated in this fitting procedure, as we have seen already with other methods. We use the mock catalogues to measure the linear correlation coefficient, r (Press et al. 1992), which quantifies this correlation, and find that, between β and a, $r = 0.66$. If we knew either parameter exactly the error on the other would be smaller than quoted.

We also tried other analytical forms for the correlation function and also different scale limits and found that some combinations shifted the results by $\sim 1\sigma$.

5.5.2 Comparison of methods

We have now estimated the real-space clustering parameters using three different methods. In Section 5.2.9 we saw that the projection and inversion methods gave essentially identical results for r_0 and γ_r whereas using these 2-d fits we get slightly higher values for r_0.

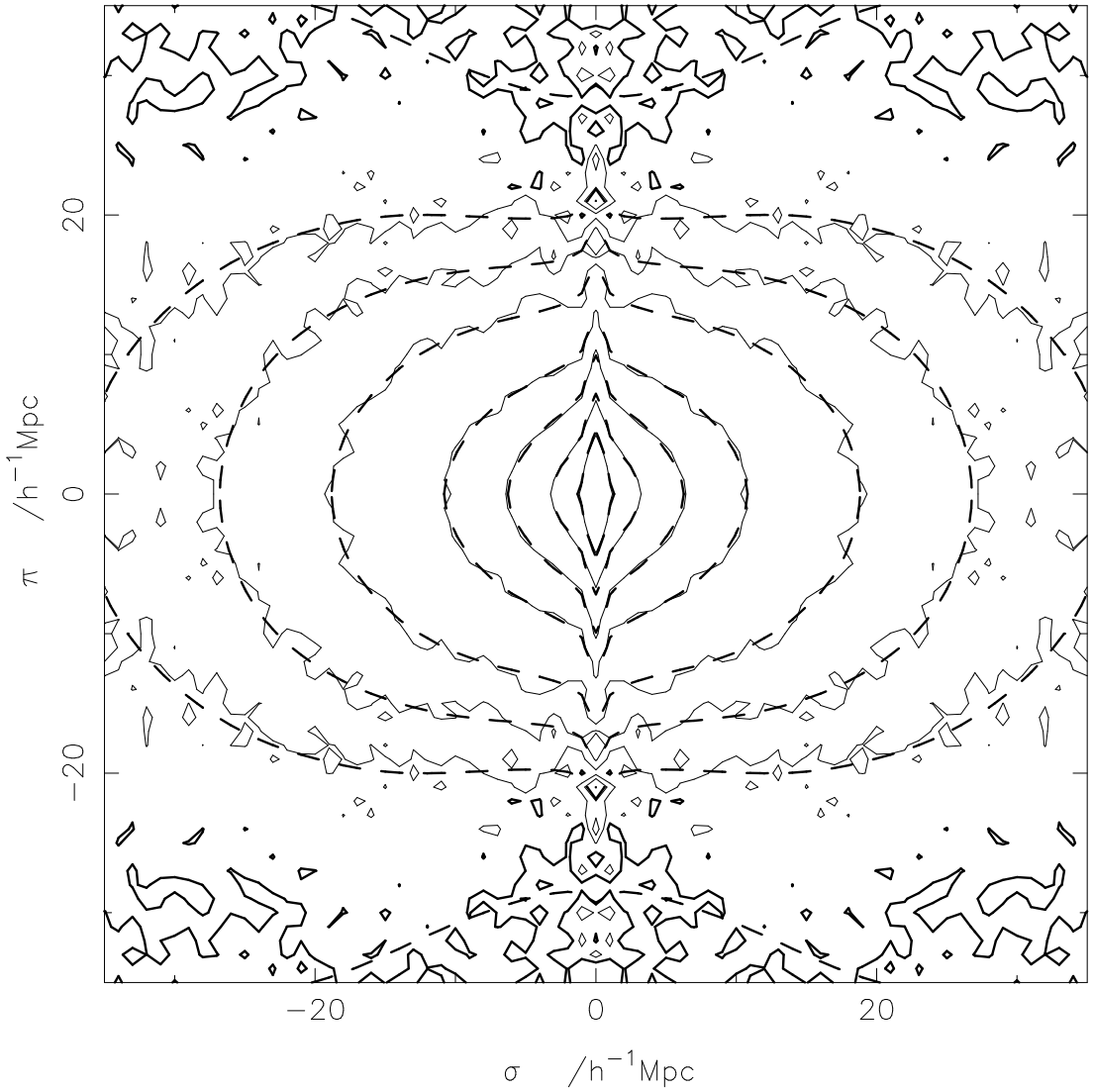


Figure 5.24: Contours of $\xi(\sigma, \pi)$ for the 2dFGRS combined data (solid lines) and the best-fit model (dashed lines). Contour levels are $\xi = 4.0$, 2.0, 1.0, 0.5, 0.2, 0.1, 0.05 and 0.0 (thick line). The model parameters are listed in Table 5.3.

If $\xi(r)$ was a perfect power law the separate methods would give unbiased results for the parameters, but we have seen evidence that this assumption is not true. The methods therefore, give different answers because of the different scales and weighting schemes used, as well as treating the redshift-space distortions in vastly different ways.

5.5.3 Previous 2dFGRS results

It is worth contrasting our present results with those obtained in a previous 2dFGRS analysis (Peacock et al. 2001). This was based on the data available up to the end of 2000: a total of 141 402 redshifts. The chosen redshift limit was $z_{\rm max} = 0.25$, yielding 127 081 galaxies for the analysis of $\xi(\sigma, \pi)$. The present analysis uses 165 659 galaxies, but to a maximum redshift of 0.2. Because galaxies are given a redshift-dependent weight, this difference in redshift limit has a substantial effect on the volume sampled. For a given area of sky, changing the redshift limit from $z_{\rm max} = 0.2$ to $z_{\rm max} = 0.25$ changes the total number of galaxies by a factor of only 1.08, whereas the total comoving volume within $z_{\rm max}$ increases by a factor of 2. Allowing for the redshift-dependent weight used in practice, the difference in effective comoving volume for a given area of sky owing to the difference in redshift limits becomes a factor 1.6. Since the effective area covered by the present data is greater by a factor of $165\,659/(127\,081/1.08) = 1.4$, The total effective comoving volume probed in the current analysis is in fact 15% smaller than the 2001 analysis; this would imply that the errors on clustering statistics should have random errors that are about 7% larger than previously. Of course, the lower redshift limit has several important advantages: uncertainties in the selection function in the tail of the luminosity function are not an issue (see Norberg et al. 2002a); also, the mean epoch of measurement is closer to $z = 0$. Given that the sky coverage is now more uniform, and that the survey mask and selection function have been studied in greater detail, the present results should be much more robust.

The other main difference between the present work and Peacock et al. (2001) lies in the method of analysis. The earlier work quantified the flattening of the contours of $\xi(\sigma, \pi)$ via the quadropole-to-monopole ratio, $\xi_2(s)/\xi_0(s)$. This is not to be confused with the quantity $Q(s)$ from Section 5.3.2, which uses an integrated clustering measure instead of $\xi_0(s)$. This is inevitably more noisy, as reflected in the error bar of $\delta\beta = 0.17$ resulting from that method. The disadvantage of using $\xi_2(s)/\xi_0(s)$ directly, however, is that the ratio depends on the true shape of $\xi(r)$. In Peacock et al. (2001), this was assumed to be known via the deprojection of angular clustering in the APM survey (Baugh & Efstathiou 1993); in this Chapter we have made a detailed internal estimate of $\xi(r)$, and considered the effect of uncertainties in this quantity. Apart from this difference, the previous method of

fitting to $\xi_2(s)/\xi_0(s)$ should in principle give results that are similar to our full fit to $\xi(\sigma,\pi)$ in Section 5.5.1. The key issue in both cases is the errors, which are estimated in a fully realistic fashion using mock samples in this work. The previous analysis used two simpler methods: an empirical error on $\xi_2(s)/\xi_0(s)$ was deduced from the NGP–SGP difference, and correlated data were allowed for by estimating the true number of degrees of freedom from the value of χ^2 for the best-fit model. This estimate was compared with a covariance matrix built from multiple realizations of $\xi(\sigma,\pi)$ using Gaussian fields; consistent errors were obtained. We applied the simple method from Peacock et al. (2001) to the current data, keeping the assumed APM $\xi(r)$, and obtained the marginalized result $\beta = 0.55 \pm 0.075$. The comparison with our best estimate of $\beta = 0.49 \pm 0.09$ indicates that the systematic errors in the previous analysis (from e.g. the assumed $\xi[r]$) were not important, but that the previous error bars were optimistic by about 20%.

5.5.4 Peculiar velocities as a function of scale

There has been much discussion in the literature on whether or not the random pairwise peculiar velocity dispersion, a, is a function of projected separation, σ.

Many authors have used N-body simulations to make predictions of what might be observed. Cen, Bahcall & Gramann (1994) found that a decreased approximately linearly from 2 to 20 h^{-1}Mpc and then stayed constant. Jenkins et al. (1998) found a similar result for the dark matter and this is plotted in Figure 5.25 as the solid line for a ΛCDM cosmology. The dashed line is from Peacock & Smith (2000) who used the halo model to predict the peculiar velocities for the galaxy distribution. Benson et al. (2000) used the GIF simulations combined with semi-analytic models and made the prediction shown by the dotted line.

Observationally, Jing, Mo & Börner (1998) measured the pairwise velocity dispersion in the Las Campanas Redshift Survey and found no significant variation with scale. Zehavi et al. (2002) used the Sloan Digital Sky Survey (SDSS) data and found that a decreased with scale for $\sigma \gtrsim 5\ h^{-1}$Mpc. These observations are also plotted in Figure 5.25.

All these observations have assumed a functional form for the infall (or 'streaming') and not used β directly. We have already shown that proper consideration of the infall parameter is vital in such studies. Indeed, Zehavi et al. (2002) say

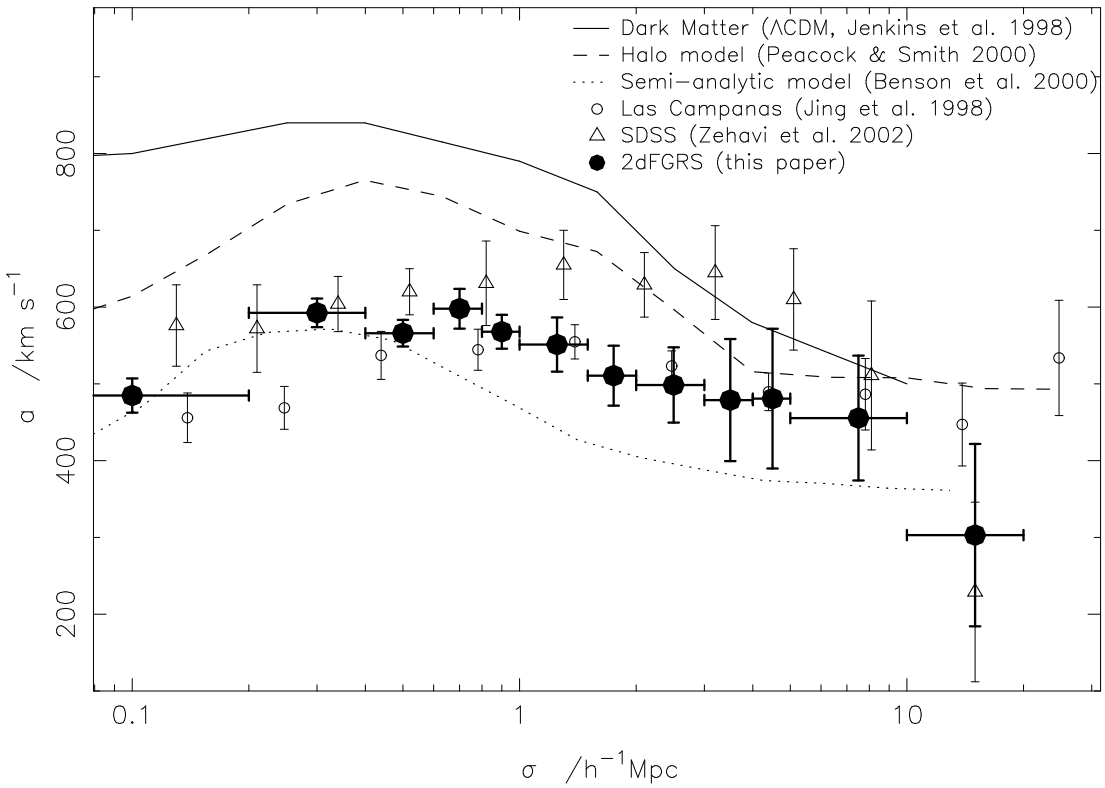


Figure 5.25: The variation of a with projected separation, σ, for various redshift surveys and N-body numerical simulation, as indicated in the legend. The 2dFGRS results use β to calculate the infall velocities whereas the other results assume a functional form (see discussion in Section 5.5.4).

that their estimates of a for $\sigma > 3\ h^{-1}$Mpc depend significantly on their choice of streaming model.

The difference in results from Section 2.5.3 which measured the value of a at small σ (570 km s^{-1}), and from using the $\xi(\sigma, \pi)$ grid (506 km s^{-1}), which measures an average value, hints that there may be such an effect in the 2dFGRS data. We test for variations in a by repeating the fits described in Section 5.5.1 using a global β, r_0 and γ_r but allowing a to vary in each σ slice. The results are shown in Figure 5.25, compared with the results from other surveys, and numerical simulations as discussed above. The value of 506 km s^{-1} obtained from the 2-d fit for scales $> 8\ h^{-1}$Mpc is close to the value at $8\ h^{-1}$Mpc where most of the signal is coming from. The value of 570 km s^{-1} obtained from the Fourier transform technique agrees well with the results found for $\sigma < 1\ h^{-1}$Mpc.

We see that the overall shape of the 2dFGRS results are fairly consistent, though

slightly flatter than with the numerical simulations, but the amplitude is certainly different. The values of β, r_0 and γ_r are essentially unchanged when fitting in this way. We note again that the effects of the infall must be properly taken into account in these measurements. We also note that we have had to use our linear, power law model on all scales, but we have seen that it does a reasonable job on non-linear scales.

5.6 Estimating $\Omega_{\rm m}$

We take the value of β measured from the multi-parameter best fit to $\xi(\sigma, \pi)$,

$$\beta(L_s, z_s) = 0.49 \pm 0.09, \tag{5.11}$$

which is measured at the effective magnitude of our survey sample, $M_s \approx -20.0$, corresponding to $L_s \approx 1.4L^*$ (assuming $M^*_{b_{\rm J}} = -19.66$, Norberg et al. 2002a). The effective pair-weighted redshift of the survey, $z_s \approx 0.15$. These values are the applicable mean values when using the J_3 weighting and redshift cuts employed (see Section 5.1.5). We note here that if we adopt an $\Omega_{\rm m} = 1$ geometry we find that $\beta = 0.55$, within the quoted 1σ errors.

5.6.1 Redshift effects

The redshift distortion parameter can be parameterized by

$$\beta = \frac{f(\Omega_{\rm m}, \Omega_\Lambda, z)}{b}. \tag{5.12}$$

A good approximation for f, at all z, in a flat Universe, was given by Lahav et al. (1991),

$$f = \Omega_{\rm m}^{0.6} + (2 - \Omega_{\rm m} - \Omega_{\rm m}^2)/140. \tag{5.13}$$

Hence, to obtain $\Omega_{\rm m}$ an estimate of $b_s(L_s, z_s)$ is needed, to combine with the value of $\beta(L_s, z_s)$ we have measured in this chapter. There have been two recent attempts at this measurement. Verde et al. (2002) used the bispectrum of 2dFGRS galaxies[3], and Lahav et al. (2002) combined the 2dFGRS data with the latest

[3]This method depends slightly on the assumed values of β and a in the calculation. In the analysis they used the values from Peacock et al. (2001), $\beta = 0.43 \pm 0.07$, $a = 385$ km s^{-1}. This may affect the estimate of b_s.

CMB data[4],

$$\text{Verde et al.:} \quad b_s = 1.04 \pm 0.11 \tag{5.14}$$

$$\text{Lahav et al.:} \quad b_s = 1.20 \pm 0.09 \tag{5.15}$$

We then find, using the two different determinations of b_s,

$$\text{Verde et al.:} \quad \Omega_{\rm m}(z_s) = 0.32 \pm 0.12 \tag{5.16}$$

$$\text{Lahav et al.:} \quad \Omega_{\rm m}(z_s) = 0.41 \pm 0.14 \tag{5.17}$$

for the value of the matter density in the Universe at the effective survey redshift, z_s.

To obtain the present day value, $\Omega_{\rm m}(z = 0)$, we must make a correction for the evolution of $\Omega_{\rm m}$ with redshift. Assuming a flat Universe this correction is (Lahav et al. 1991),

$$\Omega_{\rm m}(z=0) = \frac{\Omega_{\rm m}(z_s)}{\Omega_{\rm m}(z_s) + (1+z_s)^3(1-\Omega_{\rm m}(z_s))}, \tag{5.18}$$

and hence

$$\text{Verde et al.:} \quad \Omega_{\rm m}(z=0) = 0.23 \pm 0.09 \tag{5.19}$$

$$\text{Lahav et al.:} \quad \Omega_{\rm m}(z=0) = 0.31 \pm 0.11 \tag{5.20}$$

for which the Verde et al. result is derived solely from the 2dFGRS, whereas the Lahav et al. result depends on the combination with the CMB.

5.6.2 Comparisons

An interesting comparison is with the results of Percival et al. (2002) who combined the 2dFGRS power spectrum data with the latest CMB data, assuming a flat cosmology and found $\Omega_{\rm m}(z=0) = 0.31 \pm 0.06$. Running the above argument in reverse we find $b_s = 1.20 \pm 0.25$, consistent with both determinations of b_s described above.

These measurements of $\Omega_{\rm m}$ are consistent with a different estimation from the 2dFGRS and CMB (Efstathiou et al. 2002) and from combining the 2dFGRS with cosmic shear measurements (Brown et al. 2002) and also from a recent analysis of X-ray luminous clusters (Allen et al. 2002).

[4]This analysis relies on the assumptions of a 'constant galaxy clustering' model.

5.6.3 Luminosity effects

We note that the above analysis is independent of luminosity as we examine everything at the effective luminosity of the survey, L_s. Using correlation functions for different volume-limited samples of 2dFGRS galaxies, Norberg et al. (2001) found a luminosity dependence of clustering of the form (see Section 6.1.2, cf. Equation 5.6),

$$b/b^* = 0.85 + 0.15(L/L^*) \tag{5.21}$$

which gives an estimate for the bias of the survey galaxies, $b_s = 1.06b^*$ (using $L = 1.4L^*$), where b^* is the bias of L^* galaxies. If this bias relation holds on the scales considered here then β will be increased by the same factor of 1.06,

$$\beta(L^*, z_s) = 0.52 \pm 0.09 \tag{5.22}$$

and evolving β in a 'constant galaxy clustering' model (Lahav et al. 2002) then,

$$\beta(L^*, z = 0) = 0.47 \pm 0.08 \tag{5.23}$$

which we choose as a fiducial point to allow comparisons with other surveys with different effective luminosities and redshifts.

5.7 Summary

In this chapter we have made several measurements based on the $\xi(\sigma, \pi)$ derived using 2dFGRS galaxies.

(i) The spherical average of $\xi(\sigma, \pi)$ gives the redshift-space correlation function, $\xi(s)$, from which we measure the redshift space clustering length, $s_0 = 6.82 \pm 0.28\ h^{-1}$Mpc and local slope $\gamma_s = 1.57 \pm 0.07$. At large and small scales $\xi(s)$ drops below a power law as expected for a low density ΛCDM Universe.

(ii) The projection of $\xi(\sigma, \pi)$ along the π axis gives an estimate of the real-space correlation function, $\xi(r)$, which on scales $0.1 < r < 12\ h^{-1}$Mpc can be fit by a power law $(r/r_0)^{-\gamma_r}$ with $r_0 = 5.05 \pm 0.26\ h^{-1}$Mpc, $\gamma_r = 1.67 \pm 0.03$. At large scales $\xi(r)$ drops below a power law as expected, for instance, in the popular ΛCDM model.

(iii) The ratio of real and redshift-space correlation functions on scales of $8 - 30\ h^{-1}$Mpc gives an estimate of systematic infall velocities which correspond to $\beta = 0.45 \pm 0.14$. The quadrupole moment of $\xi(\sigma, \pi)$ on large scales measures $\beta = 0.47^{+0.19}_{-0.16}$.

(iv) Comparing the projections of $\xi(\sigma, \pi)$ along the π and σ axes gives an estimate of the distribution of random pairwise peculiar velocities, $f(v)$. We find that large scale infall velocities affect the measurement of the distribution significantly and cannot be neglected. Using $\beta = 0.49$ we find that $f(v)$ is well fit by an exponential model with pairwise velocity dispersion, $a = 570 \pm 25$ km s^{-1}, at small σ.

(v) A multi-parameter fit to $\xi(\sigma, \pi)$ simultaneously constrains the shape and amplitude of $\xi(r)$ and both the velocity distortion effects parameterized by β and a. We find $\beta = 0.49 \pm 0.09$ and $a = 506 \pm 52$ km s^{-1}.

(vi) Combining our best estimate of β with two estimates of the linear bias parameter, b_s (Verde et al. 2002; Lahav et al. 2002), we find that $\Omega_{\rm m}(z_s = 0.15) = 0.32 - 0.41$. Allowing for expansion implies that the present day matter density parameter, $\Omega_{\rm m} = 0.23 \pm 0.09$, derived solely from the 2dFGRS or $\Omega_{\rm m} = 0.31 \pm 0.11$ after combination with the CMB. These estimates are consistent with other 2dFGRS studies.

These results show that the clustering of 2dFGRS galaxies as a whole are well matched by a low density ΛCDM model with a non-linear local bias scheme based on the smoothed dark-matter density field. Nevertheless there are features of the galaxy distribution which require more sophisticated models. For example, the mock catalogues use a non-linear bias scheme, whereas the data suggest that a linear bias is a good fit to the data. This is likely to be intimately connected with the observation that the amplitude of clustering depends on both the luminosity and spectral type of galaxies (Norberg et al. 2002b). The dynamical properties of the galaxy distribution also depend on the type of galaxy being studied. The methods presented have also been used on sub-samples of the 2dFGRS, split by their spectral type (see Chapter 6, Madgwick et al. 2002a).

Chapter 6

2dF Clustering By Galaxy Type

Some of my work with the 2dFGRS Team has been on the analysis of the luminosity dependence of galaxy clustering (Norberg et al. 2001), and the spectral type dependence of galaxy clustering (Madgwick et al. 2002a). I am a leading author of both these papers and performed much of the analysis independently.

6.1 Luminosity dependence of clustering

The measurement of the strength of clustering as a function of luminosity (or equivalently absolute magnitude) of the galaxy sample has been a much discussed topic in the literature (see Section 4.1) with several claims of a detection. High peak bias models predict that more luminous galaxies, which are presumably more massive, should cluster more strongly than fainter galaxies.

6.1.1 Methods

By analysing the projected correlation functions (PCF's) of volume limited samples of 0.5 mag and 1.0 mag slices in the 2dFGRS it was possible to obtain estimates of the real-space clustering as a function of absolute magnitude. As peculiar velocities may also depend on the magnitude of the sample, we use PCF's as they are free of any redshift-space distortions (see Section 2.2.2). Volume limited samples are chosen as they are well defined and have a uniform radial selection function and so are easier to deal with, but the number of galaxies in each subsample is much reduced - though each sample has more than 1000 galaxies. When quoting magnitude limits in the following sections, $M = M_{b_J} - 5\log_{10} h$.

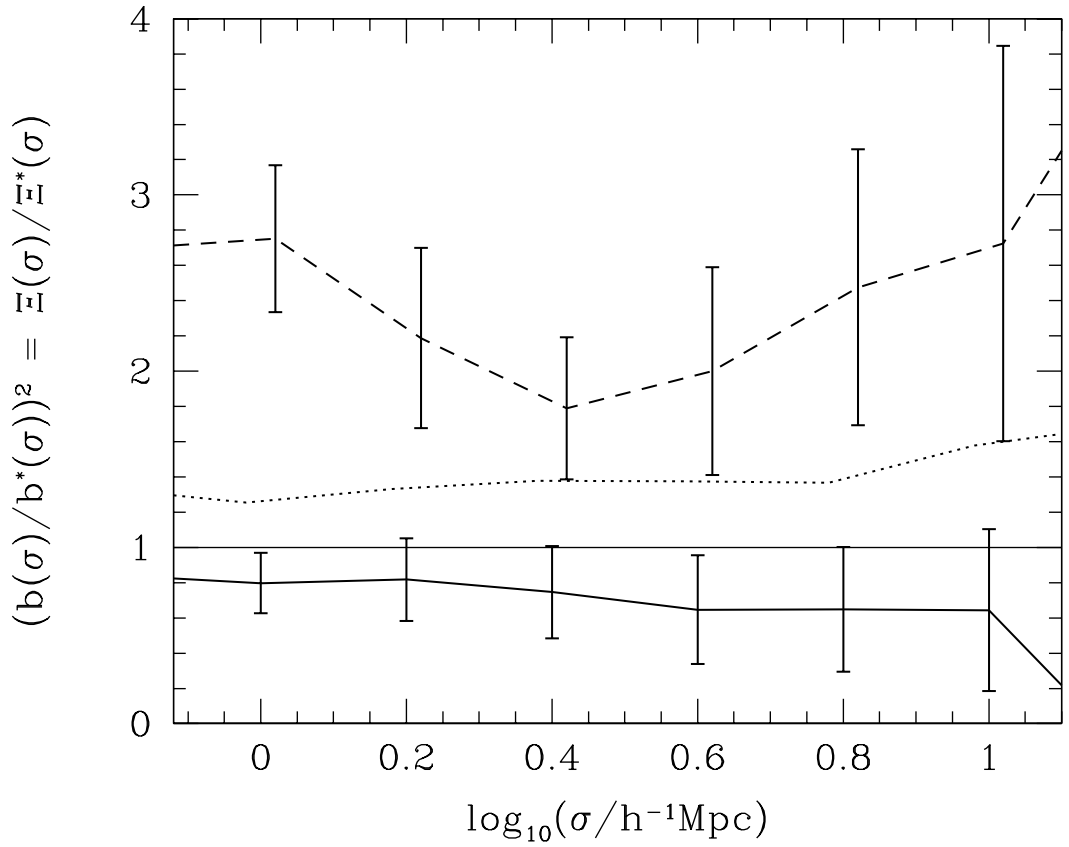


Figure 6.1: From Norberg et al. (2001): The ratio of PCF's for various magnitude slices to the PCF of the galaxies with absolute magnitudes $-20 \leqslant M \leqslant -19$. The solid line shows the ratio for galaxies with absolute magnitudes of $-19 \leqslant M \leqslant -18$, the dotted line for $-21 \leqslant M \leqslant -20$ and the dashed line for $-22 \leqslant M \leqslant -21$. The error bars, derived from the mock catalogues, have been omitted from the dotted line for clarity but are comparable to those on the solid line.

6.1.2 Results

Shape: Plotted in Figure 6.1 is the ratio of PCF's for various magnitude slices to the PCF of galaxies with absolute magnitudes in the range $-20 \leqslant M \leqslant -19$, which is a sample with a median magnitude $\approx M^*$. This plot examines the shape of the PCF's as a function of magnitude. The curves are consistent with being flat, showing that the slope of the correlation function varies remarkably little as a function of magnitude, whereas the very significant vertical shifts show that the amplitude of the correlation function does indeed increase with the magnitude of the sample.

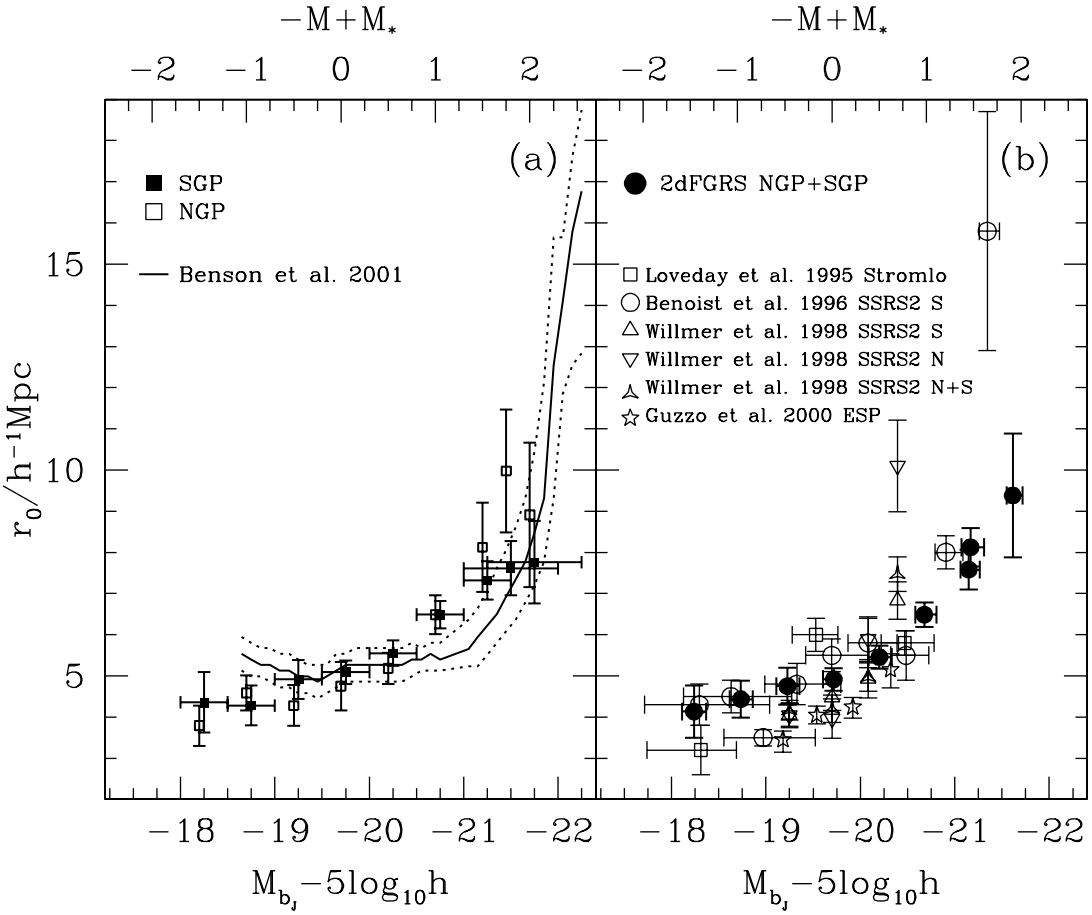


Figure 6.2: From Norberg et al. (2001): (a) r_0 as a function of magnitude for the NGP and SGP, compared to the results of a semi-analytic model. (b) The combined 2dFGRS results compared to the results from other surveys. 2dFGRS error bars are the rms spread in results from the mock catalogues.

Amplitude: From the measured PCF's, and using Equation 2.12, the real-space clustering lengths, r_0, can be measured for each sample. The left panel of Figure 6.2 shows the dependence of r_0 as a function of the median magnitude of the sample, for the two separate regions of the survey and compared to the results from a semi-analytic model (Benson et al. 2001), which matches the data well. The right hand panel compares the combined 2dFGRS results to other surveys. The results show a clear increase in clustering strength for 2dFGRS galaxies as the median magnitude of the sample increases. The advantages of the sheer size of the 2dFGRS is shown in this analysis by being able to make accurate clustering measurements for smaller fractions of the data than earlier surveys and so being able to trace the effect in more detail.

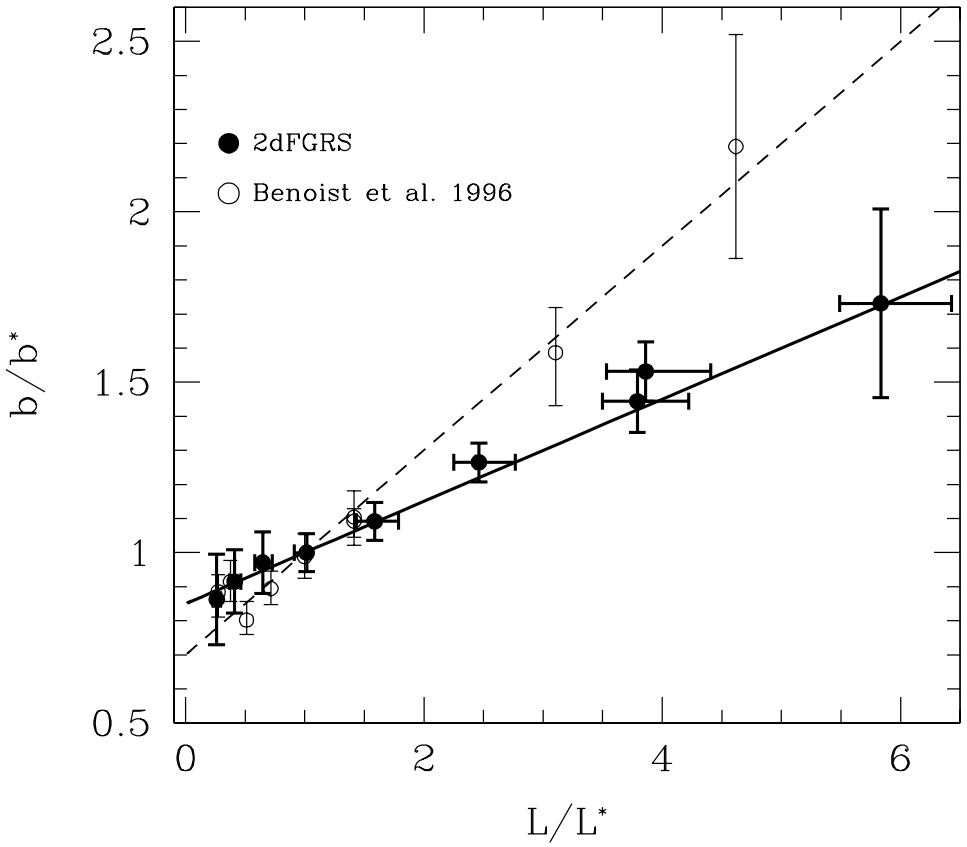


Figure 6.3: From Norberg et al. (2001): Relative bias as a function of luminosity for the 2dFGRS compared to the results of Benoist et al. (1996). The solid line is the best fit to the 2dFGRS data points (Equation 6.1).

Relative bias: These measurements of the real-space correlation length can be converted to a relation in relative bias, $(b/b^*) = (r_0/r_0^*)^{\gamma_r/2}$, as a function of luminosity (Figure 6.3) and the 2dFGRS galaxies are best fit by the relation,

$$\frac{b}{b^*} = 0.85 + 0.15 \left(\frac{L}{L^*}\right) \tag{6.1}$$

shown by the solid line in Figure 6.3, where b^* is the bias of M^* ($= -19.66$, Norberg et al. 2002a) galaxies whose real-space clustering is well described by a power law with $r_0^* = 4.9 \pm 0.3$ and $\gamma_r^* = 1.71 \pm 0.06$ (cf. $r_0^* = 4.71 \pm 0.24$, Section 5.2.9). This relation indicates that, over these luminosity ranges, there is a linear increase of bias with luminosity.

This work has made the most accurate measurement of this effect to date, with an order of magnitude more galaxies per sub-sample than previous surveys.

6.2 Spectral type dependence of clustering

The fact that different morphological types of galaxy cluster in different ways has been long known (e.g. Dressler 1980), but now with new large redshift surveys it is possible to obtain accurate measurements of the relative bias and peculiar velocities of different types of galaxies.

6.2.1 Spectral type, η

As the 2dFGRS galaxy images are taken from scans of photographic plates they are not of the quality needed to accurately determine galaxy morphology. This fact, combined with the sheer number of galaxies, makes a manual 'by-eye' determination of galaxy morphology impossible for the 2dFGRS - an automated process is therefore essential. For studies as a function of galaxy type we have to resort to determining a spectral type for the galaxy.

D. Madgwick performed a Principal Component Analysis (PCA) on the 2dFGRS data. This method reduces the ≈ 1000 channels of spectral data into a few principal components, which can then be used to classify the galaxies. Madgwick et al. (2002b) showed that two-thirds of the variance is contained within the first two principal components. The distribution of galaxies in the $pc_1 - pc_2$ plane is shown in Figure 6.4. It is not possible to accurately flux calibrate the 2dF spectra due to instrumental effects and so the continuum is uncertain. To maximise the effect of the emission/absorption features it is necessary to combine the first two principal components and it is found (Madgwick et al. 2002b) that the linear combination

$$\eta = 0.5 \cdot pc_1 - pc_2 \tag{6.2}$$

of these two components gives a continuous variable describing a galaxy's spectral type. The advantages of this approach is that it allows the data to 'speak for itself', without any ad-hoc interpretation such as morphology. The η spectrum, and the first two principal components are shown in Figure 6.5.

The maximum redshift at which η can be calculated is $z = 0.15$, which is where the Hα emission line, a key diagnostic in the analysis, becomes mixed with sky lines which are not completely removed in the spectral reduction process.

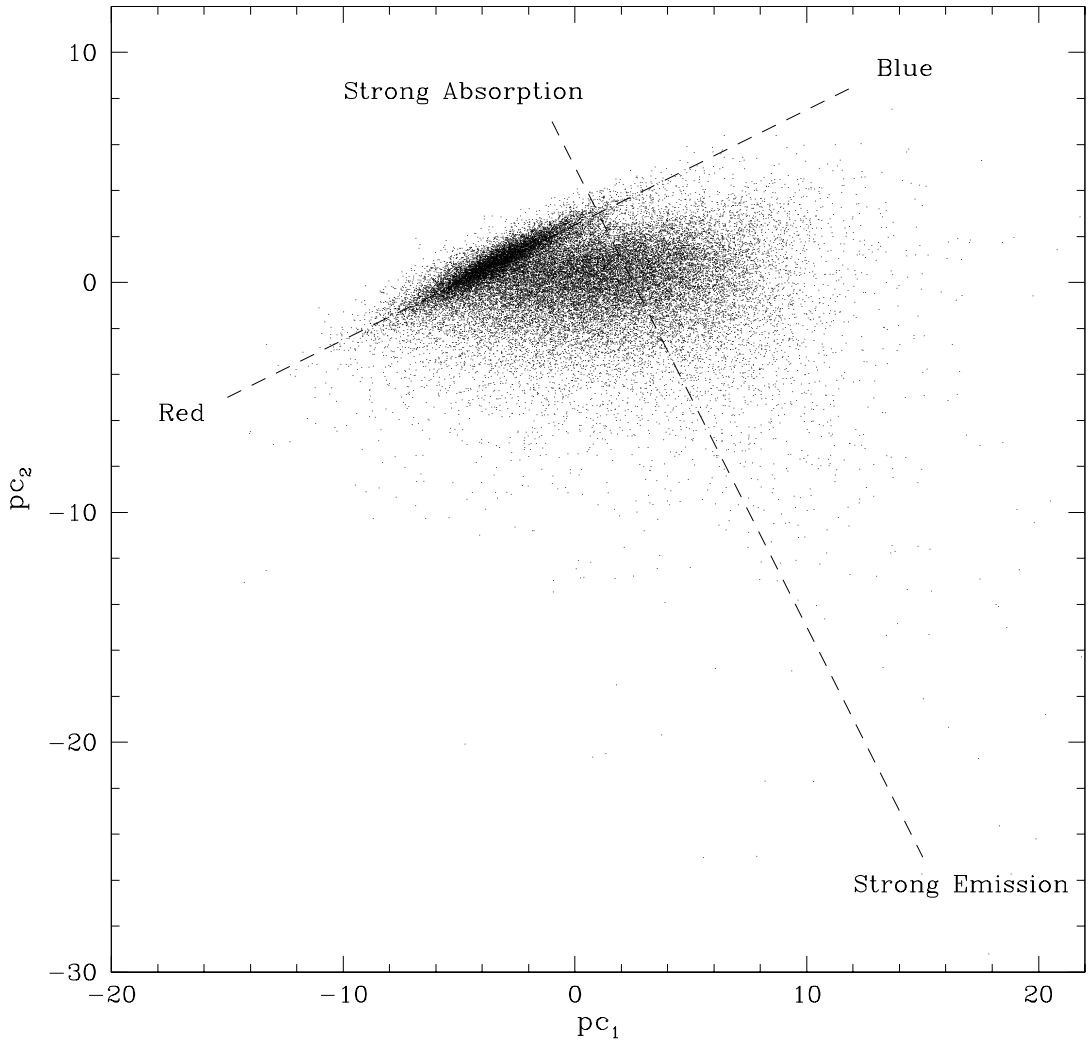


Figure 6.4: From Madgwick et al. (2002b): The distribution of 2dF galaxies in the $pc_1 - pc_2$ plane with features as indicated.

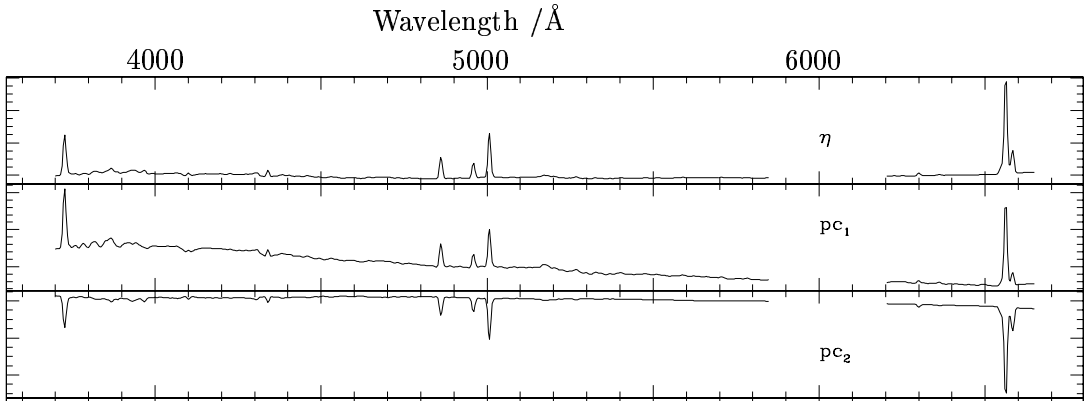


Figure 6.5: The η spectrum and the first two principal components.

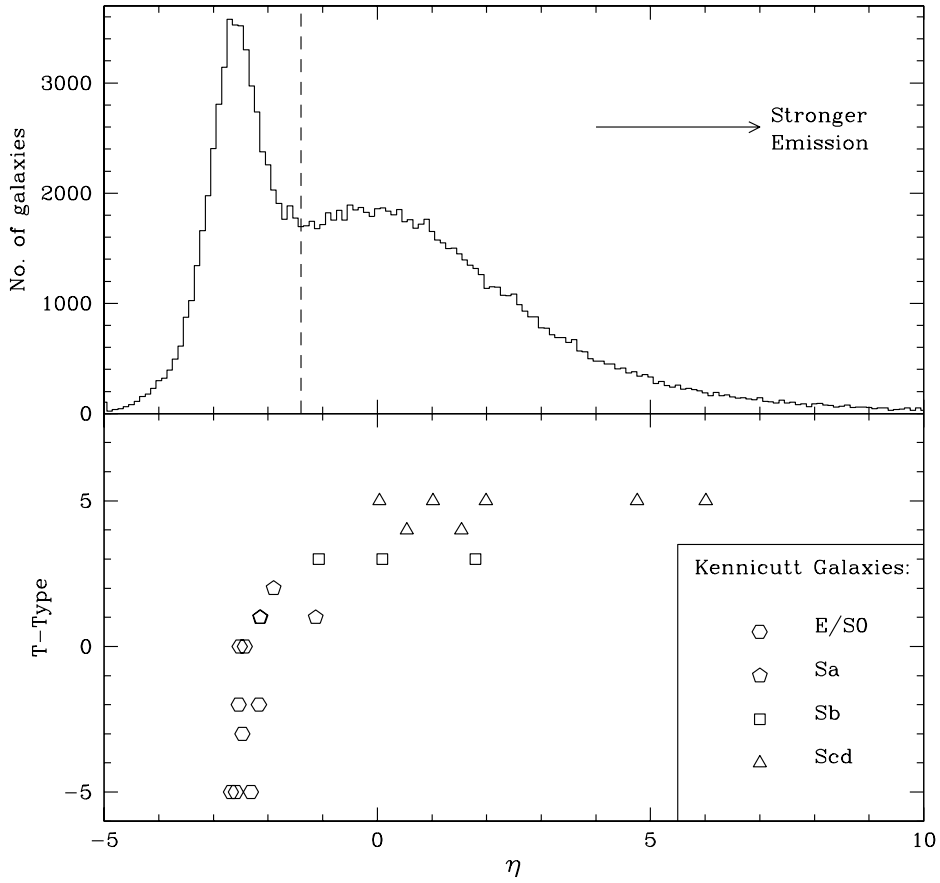


Figure 6.6: From Madgwick et al. (2002a): Top panel: η histogram. The dashed line, at $\eta = -1.4$, is the adopted split between 'passive' and 'active' galaxies. Bottom panel: The correlation of η with morphology for a training set of galaxies.

The top panel of Figure 6.6 shows the distribution of η for the galaxies in the 2dFGRS. It is clearly a bimodal distribution, the shoulder of which, at $\eta \approx -1.4$, is a neat split between relatively 'active' and 'passive' star-forming galaxies, or more crudely, 'late' and 'early' type galaxies as the bottom panel shows. Madgwick et al. (2002a) has shown that η also correlates strongly with the strength of Hα emission, and hence star formation rates.

There has been concern that the small diameter of the 2dF fibres (2 arcsec) will introduce aperture effects by only sampling small regions of large galaxies. Tests performed by Madgwick et al. (2002b) concluded that this effect is not significant because the relatively poor seeing at the observatory (typically 1.5-1.8 arcsec), reduces the effect by smearing out the light sufficiently to sample most of a galaxy.

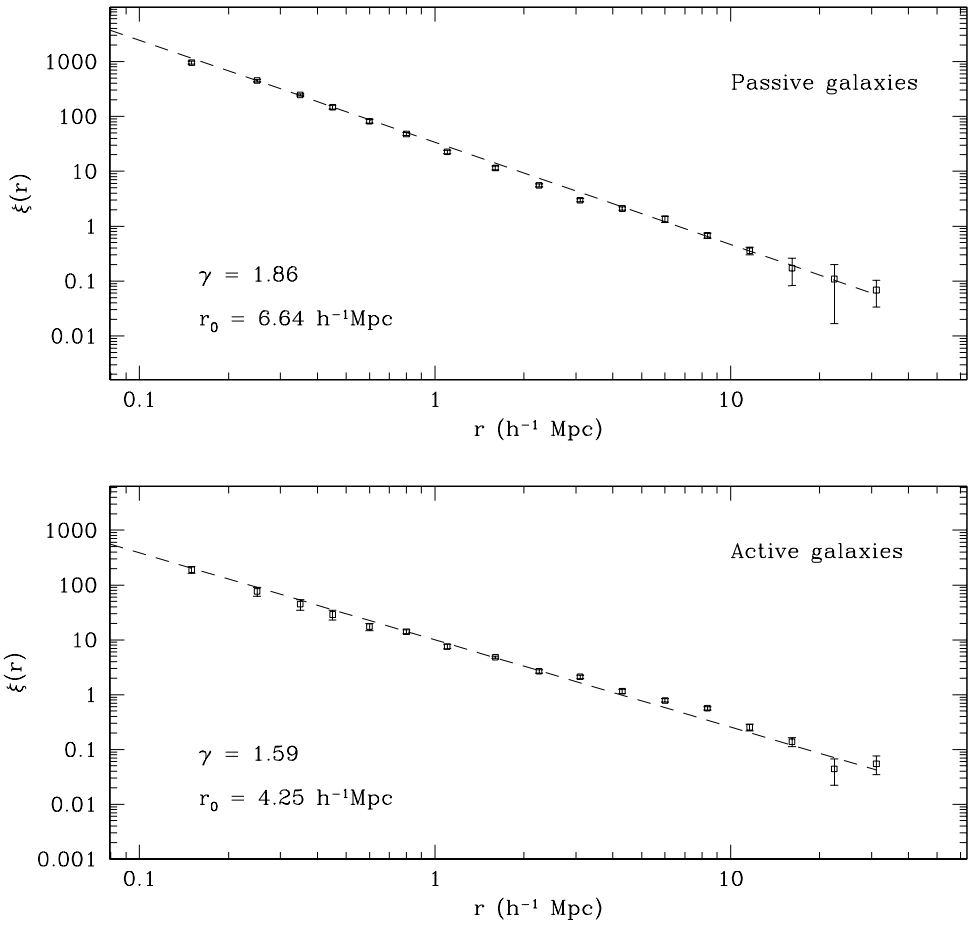


Figure 6.7: From Madgwick et al. (2002a): Real-space correlation functions for the two η sub-samples, with best fit power laws as indicated.

6.2.2 Results

Madgwick et al. (2002a) repeated much of the analysis presented in Chapter 5 on two sub-samples of the 2dFGRS, split at $\eta = -1.4$ as shown in Figure 6.6. This leaves 36 362 galaxies in the 'passive' sample and 60 427 galaxies in the 'active' sample, each with a maximum redshift of $z_{\rm max} = 0.15$.

The real-space correlation functions: A comparison of the real-space correlation functions, obtained by the S92 method (Section 2.2.3), is plotted in Figure 6.7. It can be seen that both types are well represented by power law forms. The clustering strength of the passive galaxies is significantly higher, as shown by the relative r_0 values, meaning that on these scales there is a strong relative bias between the samples. The conclusion is that the passive galaxies are more

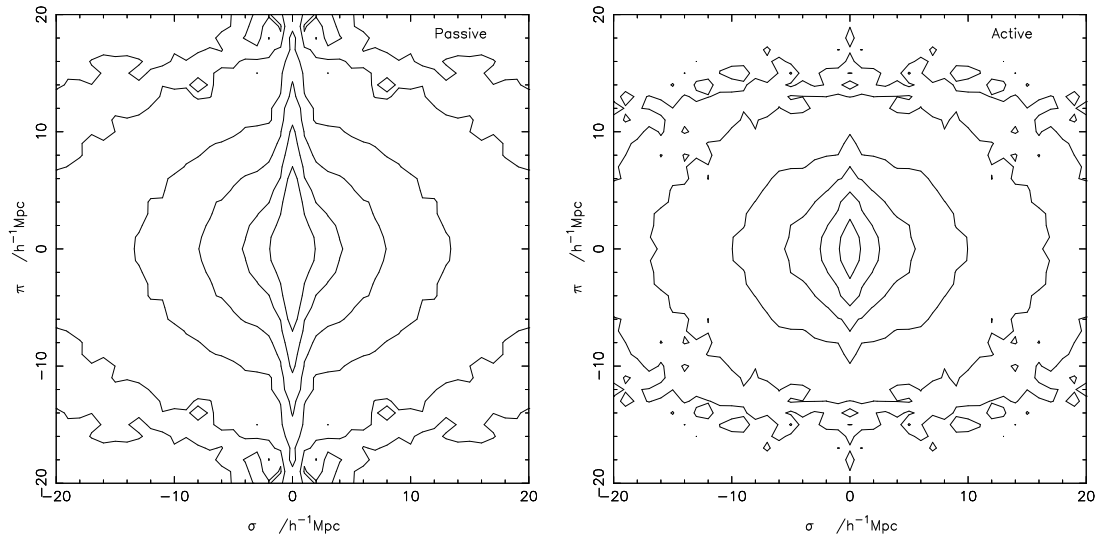


Figure 6.8: From Madgwick et al. (2002a): $\xi(\sigma, \pi)$ for the two η sub-samples as indicated, with contours at $\xi = 4.0$, 2.0, 1.0, 0.5, 0.2 and 0.1.

Table 6.1: From Madgwick et al. (2002a): Best-fit model parameters to $\xi(\sigma, \pi)$ for the two η sub-samples, fitted over the range $8 < s < 20\ h^{-1}$Mpc.

Parameter	Passive	Active
β	0.46 ± 0.13	0.54 ± 0.15
a (km s^{-1})	618 ± 50	418 ± 50
r_0 (h^{-1}Mpc)	6.54 ± 0.28	4.22 ± 0.21
γ_r	1.86 ± 0.07	1.40 ± 0.08

strongly clustered, as has been seen in previous surveys.

The 2-d correlation functions: Figure 6.8 shows $\xi(\sigma, \pi)$ for the two η sub-samples. The clearest difference by-eye is that the passive sample has a significantly higher velocity dispersion, as shown by the elongated contours in the π direction at small σ. It is also possible to also see the difference in correlation lengths by examining the spacing of the contours.

Redshift-space distortions: A quantitative measure of the differences was obtained by fitting the models described in Section 2.3 to $\xi(\sigma, \pi)$. This procedure was similar to that described in Section 5.5.1 and the resulting best fit parameters are shown in Table 6.1. The correlation function parameters show a similar difference to that seen in Figure 6.7.

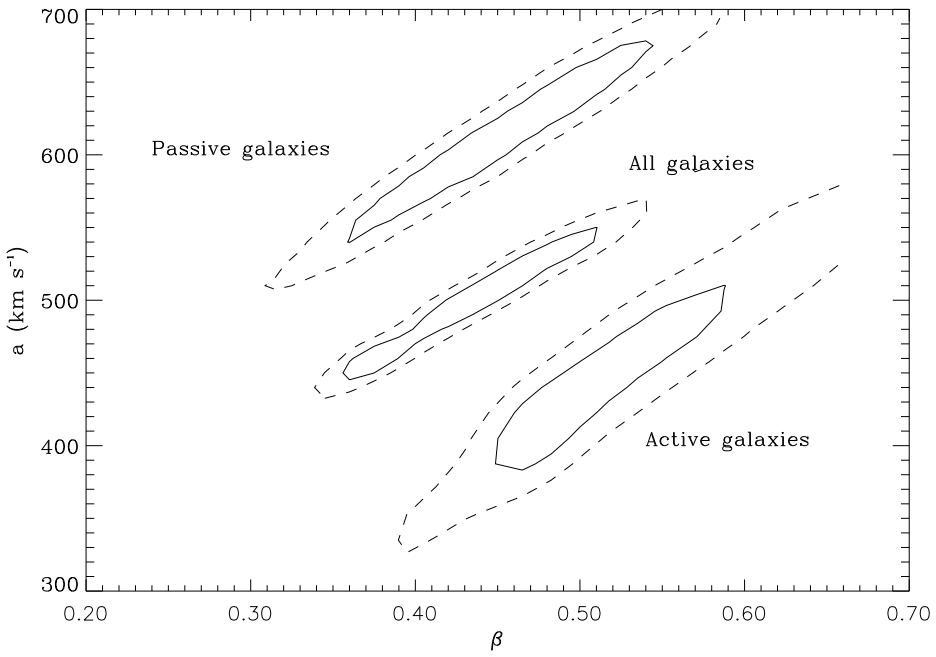


Figure 6.9: From Madgwick et al. (2002a): 1σ and 2σ likelihood contours for β and a, from fits as described in the text. The result for 'All galaxies' refers to the entire η sample, which has a lower redshift cut than the sample in Chapter 5.

Figure 6.9 shows likelihood contours for β and velocity dispersion, a, for the two sub-samples and the entire η-sample. The two β values are within 1σ of eachother, which would indicate that there is little relative bias between the samples. This appears to contradict the result found for r_0, which shows a strong relative bias. The reason is that ξ has a stronger dependence on the relative bias than β,

$$\frac{\beta_{\rm passive}}{\beta_{\rm active}} = \frac{b_{\rm active}}{b_{\rm passive}} = \sqrt{\frac{\xi_{\rm active}}{\xi_{\rm passive}}} \tag{6.3}$$

The effect is further weakened because the correlation function slopes are so different. This causes the two ξ's to converge at the larger scales tested with this fit, making the measurements fully consistent.

The more passive galaxies do have a significantly higher peculiar velocity dispersion, consistent with the notion that they are indeed found in the centre of clusters where the random motions are thought to be greatest.

Chapter 7

Redshift Periodicities

Most of the work in this chapter was presented in Hawkins, Maddox & Merrifield (2002) and was done in collaboration with the other authors.

7.1 Introduction

Claims of periodicities or regularities in redshift distributions of various astronomical objects have been made for many years (e.g. Burbidge & Burbidge 1967; Broadhurst et al. 1990; Karlsson 1990; Burbidge & Napier 2001). This effect, if real, has far-reaching implications for the interpretation of redshift as a cosmological phenomenon, and, indeed, for the nature of objects like quasi-stellar objects (QSOs) that appear to display the periodicities.

One particularly intriguing effect has been explored by Arp et al. (1990) and Karlsson (1990, hereafter K90) and extended to a larger sample by Burbidge & Napier (2001, hereafter BN01). It involves the apparent strong periodicity in $\log(1 + z_{\rm qso})$ for a sample of QSO redshifts, $z_{\rm qso}$, where the QSO appears projected close to a 'foreground' galaxy at a lower redshift. If confirmed, such an effect would be impossible to explain in conventional cosmological terms: it would either require that the QSOs be physically associated with the galaxies in an as-yet unexplained fashion, or that the QSO light passing the galaxy is somehow influenced to quantize its redshift.

The criticism usually levelled at this kind of study is that the samples of redshifts have tended to be rather small and selected in a heterogeneous manner, which makes it hard to assess their significance. The more cynical critics also point out

that the results tend to come from a relatively small group of astronomers who have a strong prejudice in favour of detecting such unconventional phenomena. This small group of astronomers, not unreasonably, responds by pointing out that adherents to the conventional cosmological paradigm have at least as strong a prejudice towards denying such results.

The availability of the data from the 2dF Galaxy Redshift Survey (2dFGRS) and the 2dF QSO Redshift Survey (2QZ) means that for the first time there exists a large homogeneous sample of data to carry out this kind of study. Bill Napier recognized the importance of the study being carried out independent from any of the researchers with vested interests one way or the other and contacted Mike Merrifield with this suggested project. He gave clear instructions (Section 7.2) as to what analysis should be performed and what periodic effect should be seen if the phenomenon is real, but chose to take no part in the subsequent analysis. It has been attempted to carry out this analysis without prejudice. Indeed, either outcome would have been interesting: if the periodicity were detected, then there would be some fascinating new astrophysics for us to explore; if it were not detected, then we would all have the reassurance that our existing work on redshift surveys, etc, has not been based on false premises.

Section 7.3 describes the data sets used, Section 7.4 presents the manner in which they have been analysed and the results are described in Section 7.5.

7.2 Napier's prediction

In describing this project, Bill Napier stated that a strong periodicity had been found in the redshifts of QSOs projected within 30 arcmin of the centres of nearby galaxies (either in the Virgo Cluster or bright galaxies in the Shapley Ames Catalog), corresponding to a physical scale of $\sim 200\,\mathrm{kpc}$ at these galaxies' distances. He therefore suggested that all the redshifts in the 2dFGRS be used to estimate their distance (adopting a Hubble constant of $60\ \mathrm{km\ s^{-1}Mpc^{-1}}$), and find all QSOs from the 2QZ survey projected within a circle whose radius corresponds to $200\,\mathrm{kpc}$ at the galaxy's distance. Then, after transforming the QSOs' redshift to the reference frame of the galaxies that they lie behind, it should be expected to find a strong periodic signal in $\log(1+z)$ at a period $P \sim 0.09$, corresponding to redshifts $z = 0.061$, 0.30, 0.60, 0.96, 1.41, 1.96, 2.63 and 3.45.

7.3 The data

7.3.1 The 2dF parent catalogues

For this study, two large public databases are used. For the galaxies, the publicly available 100k data release from the 2dF Galaxy Redshift Survey (Colless et al. 2001) are used and for the QSOs, the publicly available data from the 10k release from the 2dF QSO Redshift Survey (2QZ, Croom et al. 2001) are used. These two surveys shared observations to measure the redshifts of well-defined samples of galaxies and QSOs in a common region of the sky, making them ideally suited to this analysis.

To ensure the reliability of the sample only galaxies from the 2dFGRS with the two highest quality flags, $Q \geqslant 4$ (see Section 5.1.3) are used, in order to have confidence in the derived redshifts, and only those galaxies in the redshift range $0.01 < z < 0.3$ are used. For the 2QZ sample, only QSOs in the database with the highest quality flag are used, which implies a clear spectral identification of the object as a QSO. This quality control leaves a total of 67 291 galaxies and 10 410 QSOs in the samples. The redshift histogram for the QSO's is shown in Figure 7.1.

7.3.2 Pair selection

As instructed, these datasets have been inter-compared to find all QSO–galaxy pairs with an angular separation corresponding to less than 200 kpc at the distance of the galaxy. For this calculation, the fashionable Λ cosmology is adopted, with parameters $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 60$ km s^{-1}Mpc^{-1}. However, the relatively low redshifts of the foreground galaxies means that the choice of cosmology makes essentially no difference to the sample selection. In a number of cases there is more than one galaxy within the 200 kpc projected distance limit of the QSO; for these objects we take the closest galaxy in projected distance to make up the pair. In a few cases, the same galaxy is used for more than one QSO. This procedure yields a total of 1647 QSO–galaxy pairs.

The predicted periodicity lies in $\log(1+z_{\rm eff})$, where $z_{\rm eff}$ is the redshift of the QSO measured relative to the nearby galaxy, so we define

$$1 + z_{\rm eff} = \frac{1 + z_{\rm qso}}{1 + z_{\rm gal}}, \tag{7.1}$$

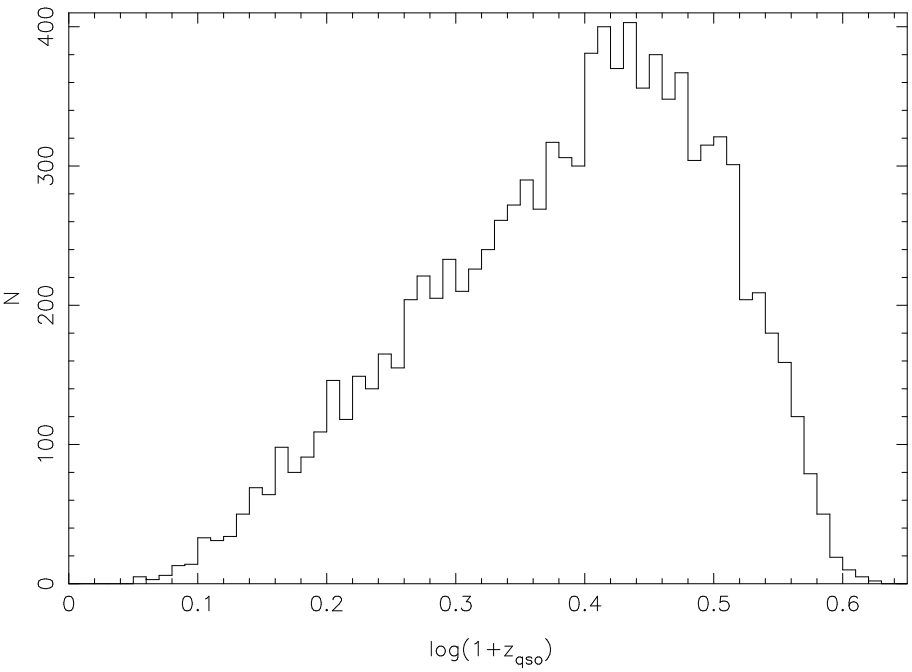


Figure 7.1: The histogram of $\log(1 + z_{\rm qso})$ for the entire 2QZ data sample.

where $z_{\rm qso}$ and $z_{\rm gal}$ are the corresponding heliocentric measurements for the QSO and galaxy. Figure 7.2 shows the distribution of $z_{\rm eff}$ and $z_{\rm qso}$ for the 1647 pairs, with the locations of the predicted periodic peaks indicated. No periodicity leaps off the page, but since the effect is likely to be quite subtle, one would not necessarily expect to be able to pick it out from the raw data, so it is important to carry out a rigorous statistical analysis.

7.3.3 Fibre collisions

In these simultaneous multi-fibre redshift surveys, no two fibres can be placed within 30 arcsec of eachother due to the nature of the instrument (Lewis et al. 2002). At the median distance of the galaxy survey this corresponds to a physical separation of ≈ 50 kpc and at the maximum distance it is ≈ 150 kpc. This will mean that some very close galaxy-quasar pairs are missed (also see Section 5.2.2). If this effect was negligible a uniform density of pairs as a function of separation might be expected, though gravitational lensing might also cause differences.

Figure 7.3 shows the number density of pairs as a function of projected separation. There is an under-density at very close separations, likely to be caused by these

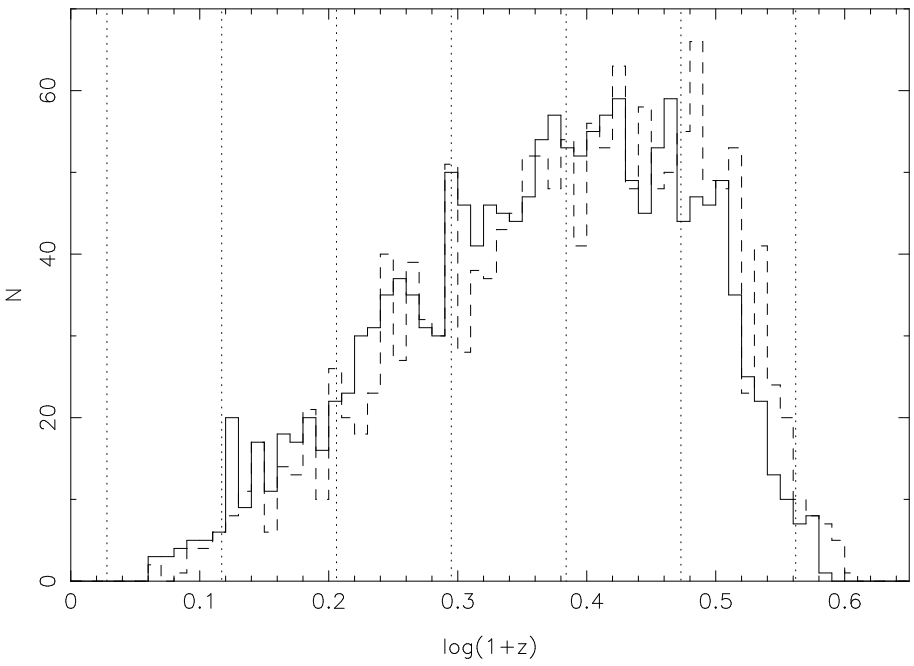


Figure 7.2: Histograms of $\log(1+z)$ for for the 2dF close pairs using $z_{\rm eff}$ (solid line) and $z_{\rm qso}$ (dashed line). The vertical dotted lines are the locations of peaks in the redshift distribution as predicted by BN01.

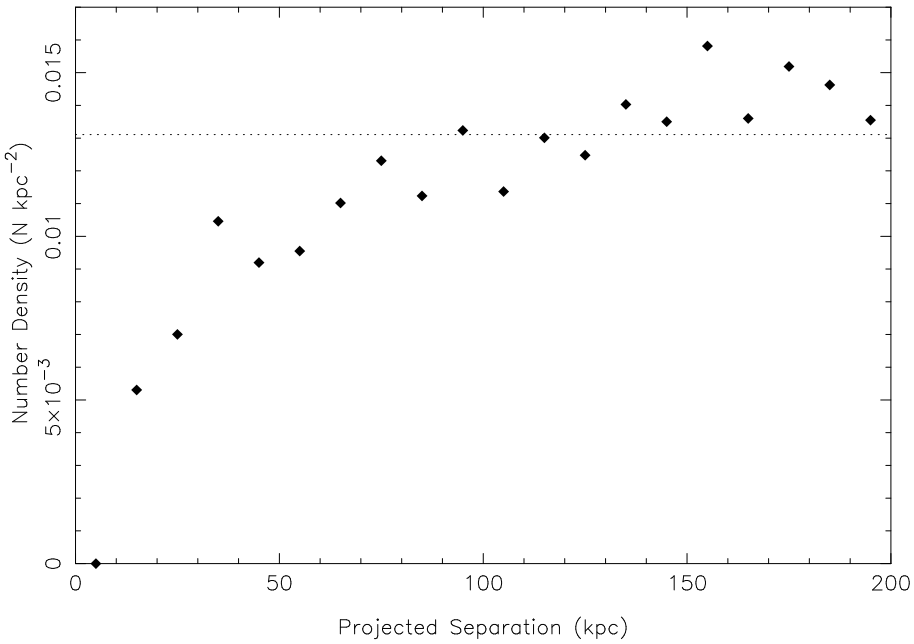


Figure 7.3: Number density of quasars as a function of projected separation from the paired galaxy. The mean number density is shown by the dotted line - there is a clear underdensity at small separations (cf. Figure 5.7).

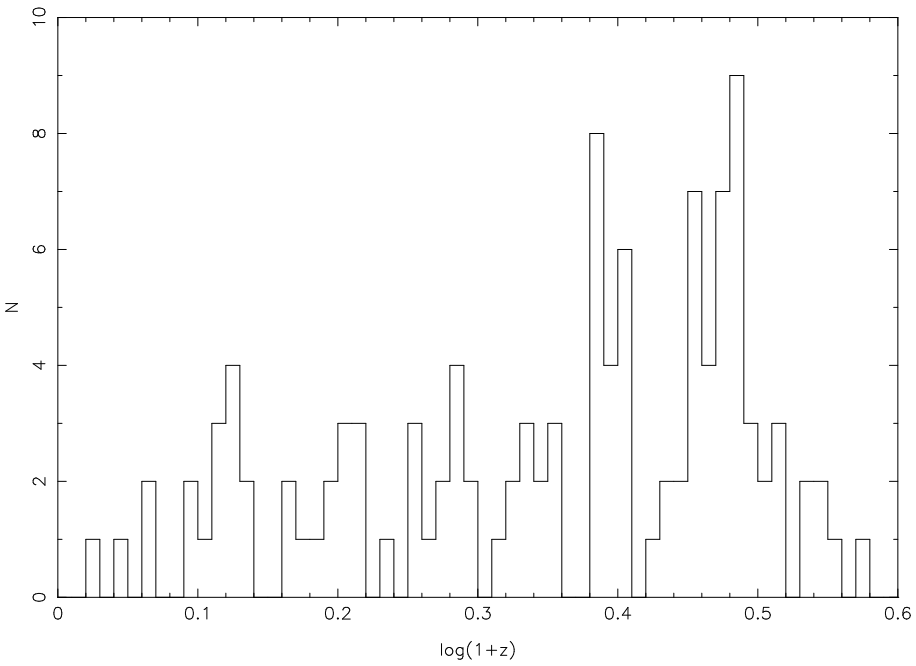


Figure 7.4: Histogram of $\log(1+z)$ for for the K90 data. The periodicity of the data, which is clearly visible, would seem to be significant.

fibre collisions (the effect is qualitatively the same as that seen in Figure 5.7); it would seem that ≈ 60 pairs are missed in this way. This effect will be independent of redshift and although our sample may not be complete, our results will not be compromised. Due to some regions of the sky being observed more than once, especially in dense areas, the other object in the pair is often targeted meaning that not all of these close pairs are missed (cf. Section 5.2.2).

7.3.4 The Karlsson data

As a check of the code developed for this project a re-analysis of the K90 data is undertaken to ensure that we can reproduce the results derived in BN01. This data is one of the main pieces of evidence for the suggested periodicities. The histogram of the 116 QSO redshifts is shown in Figure 7.4 and it clearly shows the apparent periodicity. This data set was compiled from three different surveys and has faced criticism that is not a well defined or homogeneous sample. This fact has, in part, motivated using the 2dF sample described above, which does not have these concerns.

7.4 Analysis

7.4.1 The power spectrum

To analyse the sample the power spectrum is estimated for a set of N measurements of some quantity x_i [in the current case, the value of $\log(1+z_{\rm eff})$ for different QSO–galaxy pairs]. Following the conventions of BN01, the power I at period P is defined via the formulae

$$I(P) = 2R^2 / \sum_{i=1}^{N} w_i^2 \tag{7.2}$$

where

$$R^2 = S^2 + C^2 \tag{7.3}$$

with

$$S = \sum_{i=1}^{N} w_i \sin(2\pi x_i/P), \quad C = \sum_{i=1}^{N} w_i \cos(2\pi x_i/P). \tag{7.4}$$

The quantity w_i is a weighting function to apodize any ill effects from the window function (see below); in the BN01 analysis, $w_i \equiv 1$. With this definition of the power spectrum, an infinite uniform random distribution of values x_i would yield $I \equiv 2$.

Error bars on $I(P)$ can be estimated using the 'jackknife' technique of drawing all possible samples of $N-1$ values from the N data points (without replacement), repeating the power spectral analysis on these resamplings, and calculating the standard deviation in the derived values of I at different periods P, $\sigma_J(P)$. The best estimator for the standard error in the value of I is then just $\sqrt{N-1}\,\sigma_J$.

7.4.2 The window function

The above analysis works perfectly for detecting periodic signals in data sets of infinite extent. However, in practice, such analyses are based on data sets that are finite in extent. In particular, the redshift distribution of QSOs has a cut-off at low redshifts due to the small volume sampled, and one at high redshift due to the colour selection by which QSOs are found. In addition, there may well

be variations in average numbers due to evolution in the quasar population with redshift. Thus, the idealized infinite data series is truncated to a finite series by a 'window function', which varies from a value of unity where no data are missed to zero outside the range sampled.

This truncation can introduce strong spurious features into the power spectrum. Its impact depends quite sensitively on how sharply the cut-off occurs. As an extreme example, the upper panel of Figure 7.5 shows a simulation of a uniform random distribution truncated sharply at 0 and 0.7, corresponding crudely to the distribution in $\log(1+z)$ of the data presented in BN01. As can be seen from this figure, the window function introduces many seemingly-periodic features, so that the resultant spectrum differs greatly from the $I \equiv 2$ that one would expect for an infinite uniform distribution. This figure also shows that the jackknife error analysis does a good job in determining the true root-mean-square uncertainty in the power spectrum.

In practice, the window function is unlikely to cut off this sharply, so the effects will be rather smaller than this extreme case. Nonetheless, particularly for relatively small periodic signals, it is vital that the effects of the window be taken into account [a point, indeed, noted by BN01 in their analysis of the K90 data, where one peak is disregarded as just such an artifact]. However, with relatively small samples taken from heterogeneous data sets, it is very difficult to formally quantify the selection function that specifies the shape of the window, so a rigorous analysis is difficult to implement.

Fortunately, without knowing the exact nature of the window function of the sample, one can manipulate the data in order to specify ones own more optimal window – a procedure that statisticians whimsically refer to as 'carpentry.' This process involves reducing the weighting of data close to the ends of the range observed, thereby smoothing off the sharp edges of the window, or 'apodizing' the function. This apodization is achieved by using the Hann function,

$$w_i = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi x_i}{L}\right)\right] \tag{7.5}$$

as a weighting in Equation 7.4. Here, L is chosen to cover the range over which the data are selected, so that the weighting goes smoothly from unity in the middle of the range to zero at $x_i = 0$ and $x_i = L$. As the lower panel of Figure 7.5 shows, even for the most extreme possibility of an intrinsically instantaneous cut-off in the window function, this procedure effectively restores the power spectrum to

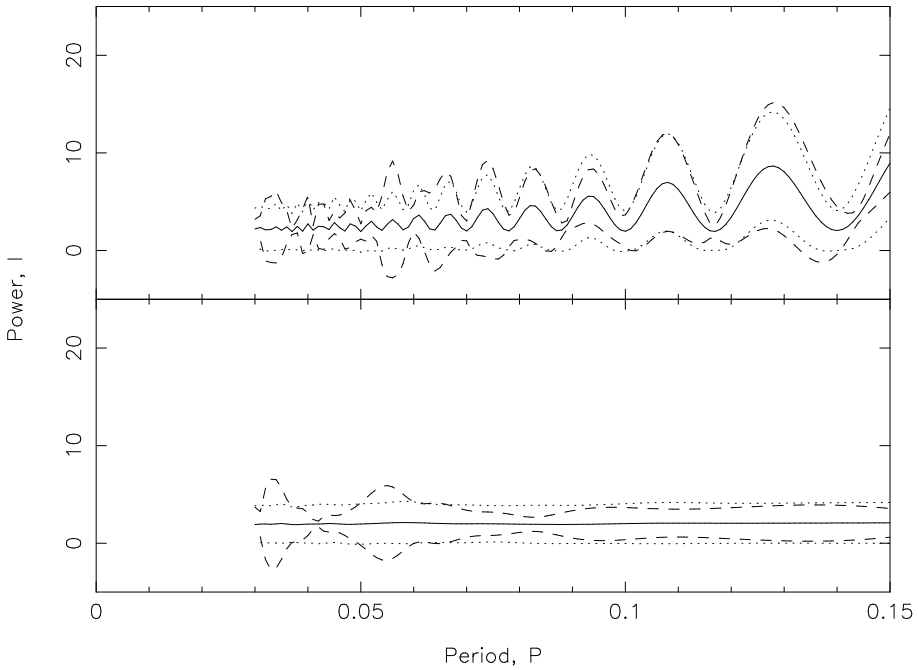


Figure 7.5: Simulation of the power spectrum of a truncated uniform distribution. Each simulation drew 1000 measurements, uniformly distributed between 0 and 0.7. The solid line shows the mean of 1000 realizations, while the dotted lines shows their standard deviation. The dashed lines are the errors as derived by applying the jackknife estimator to a single simulation. The upper panel shows the raw power spectrum, and the lower panel shows the power spectrum derived with the data weighted using a Hann function (Equation 7.5).

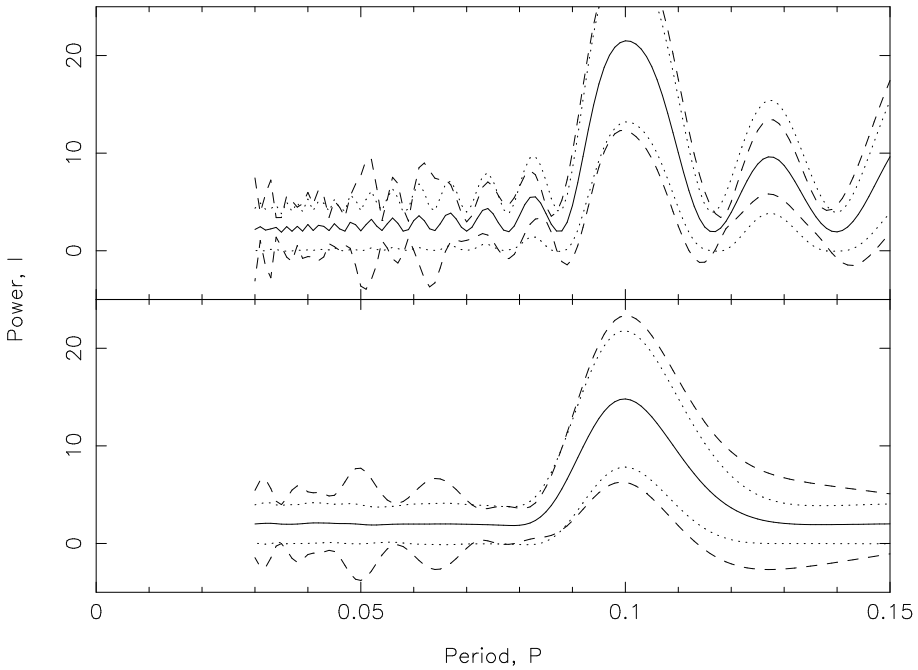


Figure 7.6: As for Figure 7.5, but with a small periodic signal added to the noise.

close to the expected value of $I(P) \equiv 2$, without compromising the errors derived using the jackknife analysis.

One concern with such a process is that it could erase real periodic signals as well as the spurious artifacts. To address this issue, the above simulation was repeated with a small periodic addition to the uniform distribution. As Figure 7.6 shows, the apodization of the data does not erase the periodic signal, and, once again, the jackknife errors provide a good estimate for the true uncertainty.

7.5 Results

7.5.1 The Karlsson data set

As Figure 7.7 shows, the unweighted spectral analysis reveals the peak in the spectrum at $P \sim 0.09$ that BN01 detected, as well as the peak at $P \sim 0.07$ that they attributed to the window function. As expected for this latter artifact, when the data are apodized, its strength is reduced to an insignificant $\sim 1.5\sigma$ above the noise value of $I = 2$. However, the stronger 'real' signal is even more dramatically reduced to a significance of only $\sim 1\sigma$. This analysis would indicate that the peak at $P \sim 0.09$ may well be compromised by the window function in this data set.

7.5.2 The 2dF data set

Figure 7.8 shows the results of the power spectrum analysis for the entire 10k 2QZ sample. A peak does appear but with a significance of only $\sim 1\sigma$ at $P \approx 0.07$.

Figure 7.9 shows the same analysis applied to the sample of 1647 QSO–galaxy pairs drawn from the 2dF surveys, as described in Section 7.3. Here, the raw and apodized power spectra are quite similar – the apodization's lack of major impact presumably reflects the overall smooth distribution in Figure 7.2, which is already quite close to optimal in shape. In any case, it is apparent that there is no significant periodicity in the data at $P \sim 0.09$, or, indeed, at any other frequency. An analysis of the QSOs' heliocentric redshifts revealed a similar absence of significant periodicities.

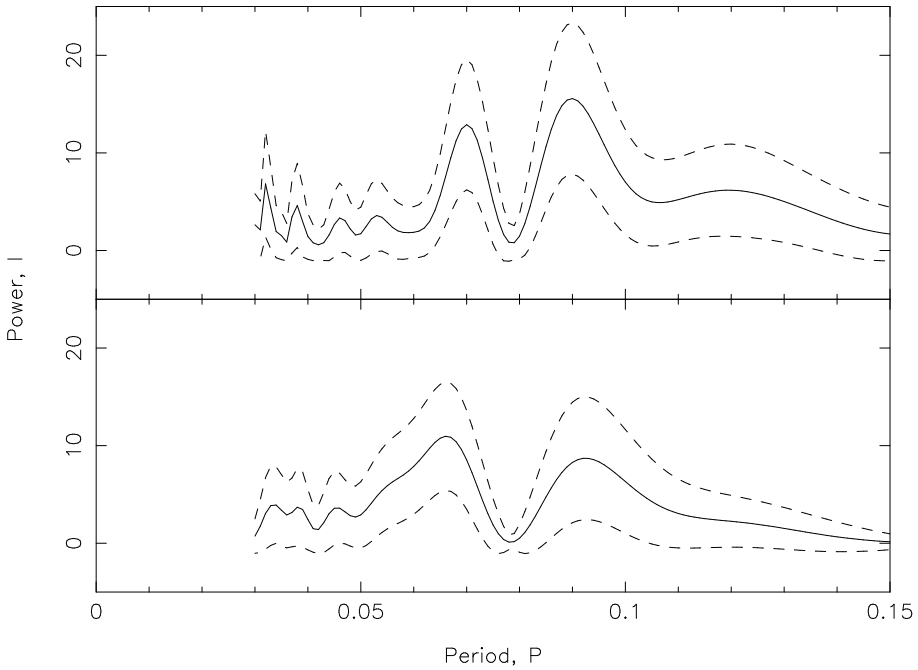


Figure 7.7: Power spectrum analysis of the data from K90. The upper panel shows the raw power spectrum [very similar to that in figure 1 of BN01]; the lower panel shows the spectrum after apodization with a Hann function. Jackknife RMS error estimates are shown by the dashed lines.

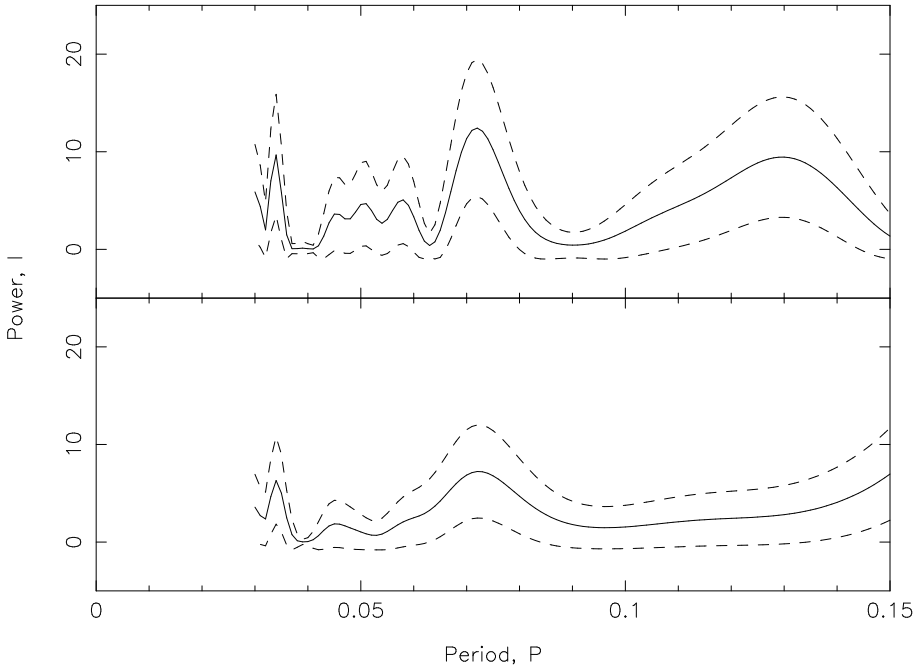


Figure 7.8: Power spectrum analysis of the entire 2QZ QSO data. The upper panel shows the raw power spectrum, and the lower panel shows and the spectrum after apodization with a Hann function. Jackknife RMS error estimates are shown as dashed lines.

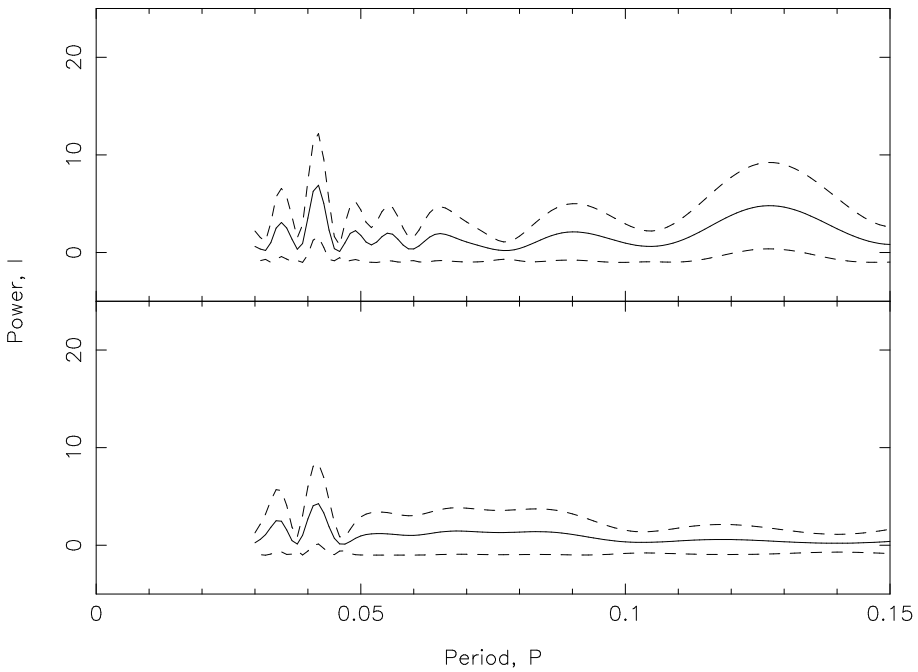


Figure 7.9: Power spectrum analysis of the QSO–galaxy pairs in the 2dF survey data. The upper panel shows the raw power spectrum, and the lower panel shows the spectrum after apodization with a Hann function. Jackknife RMS error estimates are shown by the dashed lines.

7.6 Conclusions

Given that there are almost eight times as many data points in this sample as in the previous analysis by BN01, it must be concluded that the previous detection of a periodic signal arose from the combination of noise and the effects of the window function.

There have also been claims (e.g. Basu 2001) that selection effects in the measurement of quasar redshifts cause the apparent periodicties in datasets such as BN01. The discovery of low significance peaks in the K90 and 2QZ data sets, but not in the close pairs may indicate that this is indeed the case. The small randomization of a foreground galaxy redshift completely washes out any selection effect and shows that there is no fundamental periodicity.

Chapter 8

Conclusions

8.1 The PSCz survey

The correlation function, ξ, of galaxies in the *IRAS* Point Source Catalogue galaxy redshift (PSCz) survey was measured. Its dependence on the far-infrared colour and absolute luminosity of the galaxies was investigated.

It was found that the PSCz survey redshift-space correlation function can be modelled out to a scale of 10 h^{-1}Mpc as a power law with correlation length $s_0 = 4.77 \pm 0.20$ and slope $\gamma_s = 1.30 \pm 0.04$. At a scale of 75 h^{-1}Mpc we find the value of J_3 to be 1500 ± 400. The real-space correlation function could also be modelled by a power law with $r_0 = 3.82 \pm 0.52$ and slope, $\gamma_r = 1.67 \pm 0.09$ - a lower amplitude than many optically selected samples.

It was also found that galaxies with higher $100\mu m/60\mu m$ flux ratio, corresponding to cooler dust temperatures, are more strongly clustered than warmer galaxies. Splitting the survey into three colour subsamples, we find that, between 1 and 10 h^{-1}Mpc, the amplitude of ξ is a factor of 1.5 higher for the cooler galaxies compared to the hotter galaxies. This is consistent with the suggestion that hotter galaxies have higher star-formation rates, and correspond to later-type galaxies which are less clustered than earlier types.

Using volume limited sub-samples, a weak variation of ξ as a function of absolute luminosity was found, in the sense that more luminous galaxies are less clustered than fainter galaxies. The trend is consistent with the colour dependence of ξ and the observed colour-luminosity correlation, but the large uncertainties mean that it has a low statistical significance.

8.2 The 2dF Galaxy Redshift Survey

8.2.1 The flux-limited sample

A detailed analysis of the two point correlation function, $\xi(\sigma, \pi)$, from the 2dF Galaxy Redshift Survey (2dFGRS) was presented. The large size of the catalogue, containing $\sim 220\,000$ redshifts, meant that high precision measurements of various aspects of galaxy clustering could be made.

The redshift-space correlation function, $\xi(s)$, was estimated, from which the redshift-space clustering length was estimated as $s_0 = 6.82 \pm 0.28\ h^{-1}$Mpc. Both the projected correlation function, $\Xi(\sigma)$, and the real-space correlation function, $\xi(r)$, were also measured. The latter can be fit by a power law $(r/r_0)^{-\gamma_r}$ with $r_0 = 5.05 \pm 0.26\ h^{-1}$Mpc, $\gamma_r = 1.67 \pm 0.03$. For $r \gtrsim 20\ h^{-1}$Mpc ξ drops below a power law as, for instance, is expected in the popular ΛCDM model.

The ratio of amplitudes of the real- and redshift-space correlation functions on scales of $8 - 30\ h^{-1}$Mpc gives an estimate of the redshift-space distortion parameter $\beta = 0.45 \pm 0.14$. The quadrupole moment of $\xi(\sigma, \pi)$ on scales $30 - 40\ h^{-1}$Mpc provides another estimate of $\beta = 0.47^{+0.19}_{-0.16}$.

We also estimated the distribution function of random pairwise peculiar velocities and found that the infall velocities affect these measurements significantly and cannot be neglected. The distribution is well fit by an exponential model.

The accuracy of the $\xi(\sigma, \pi)$ measurement is sufficient to constrain a combined model, which simultaneously fits the shape and amplitude of $\xi(r)$ and both the redshift-space distortion effects parameterized by β and velocity dispersion, a. It was found that $\beta = 0.49 \pm 0.09$ and $a = 506 \pm 52$ km s^{-1}. This is the first time that β and $f(v)$ have been estimated in a self consistent model of galaxy velocities.

Using the constraints on bias from two recent estimates, and taking account of redshift evolution, it was concluded that the present day matter density of the Universe, $\Omega_{\rm m} \approx 0.3$, consistent with other recent analyses.

8.2.2 Clustering by...

... absolute luminosity

From the projected correlation function of galaxies in a series of volume-limited samples it was found that the clustering amplitude increases slowly with luminosity for galaxies fainter than L^*, but rises more strongly at higher luminosities. The correlation function amplitude increases by a factor of 3.2 between $M_{b_J} - 5\log_{10} h = -18$ and -21.5. The power law slope of the correlation function shows remarkably little variation for samples spanning a factor of 40 in luminosity. The relative bias is found to be linear in luminosity and the measurements are in excellent agreement with the predictions of hierarchical galaxy formation models.

... spectral type

From the measured real-space correlation functions, $\xi(r)$, for different spectral type sub-samples of the 2dFGRS, it was found that on scales $< 8\ h^{-1}$Mpc there is a strong relative bias with the more passive galaxies being more strongly clustered. On larger scales much more similar clustering strengths are observed. From an analysis of $\xi(\sigma, \pi)$, the different galaxy types display remarkably different peculiar velocity dispersions but similar infall properties. This is all consistent with the notion of more passive galaxies inhabiting the cores of large clusters.

8.2.3 Redshift periodicities

The publicly available data from the 2dF Galaxy Redshift Survey and the 2dF QSO Redshift Survey were used to test the hypothesis that there is a periodicity in the redshift distribution of quasi-stellar objects (QSOs) found projected close to foreground galaxies. These data provide by far the largest and most homogeneous sample for such a study, yielding 1647 QSO–galaxy pairs. There is no evidence for a periodicity at the predicted frequency in $\log(1 + z)$, or at any other frequency.

8.3 Future extensions

Specifically for the projects undertaken in this thesis:

- Use the future extended PSCz survey (the Behind The Plane project) to further examine the dependence of clustering on IR type.
- Use the completed 2dFGRS to further examine the dependence of galaxy clustering as a function of other galaxy properties, such as line strengths, or colour selected samples. Different redshift cuts could be examined to look for possible clustering evolution.
- Use the completed 2dFGRS and 2QZ surveys to repeat the periodicity analysis with an even larger sample to and increase the significance of any findings. The analysis could also be tested as a function of galaxy type and redshift.

The future...

A crucial test of these particular results will come with the release of more data from the Sloan Digitial Sky Survey (SDSS) over the next few years. With several galaxy colours, they will be able to further probe the nature of the relative bias between different types of galaxy, and with more galaxies they will have even higher signal-to-noise to perform more robust modelling.

The continued growth of super-computer speed and power will allow more and more detailed simulations to be created which will be tested against the observations which motivated them. The future planned space missions will help to pin down the cosmological parameters and probe further into the Universe and in more detail.

Bibliography

Allen, S. W., Schmidt, R. W., Fabian, A. C., Ebeling, H., 2002, astro-ph/0208394.

Arp, H., Bi, H. G., Chu, Y., Zhu, X., 1990, Astron. Astrophys., **239**, 33.

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

Basu, D., 2001, Astronomische Nachrichten, **322**, 229.

Baugh, C. M., Efstathiou, G., 1993, MNRAS, **265**, 145.

Beichmann, C. A., Helou, G., Walker, D. W., 1988. *Infrared astronomical satellite (IRAS). Catalogs and atlases.* NASA RP (Reference Publication), Washington.

Beisbart, C., Kerscher, M., 2000, Ap. J., **545**, 6.

Benoist, C., Maurogordato, S., da Costa, L. N., Cappi, A., Schaeffer, R., 1996, Ap. J., **472**, 452.

Benson, A. J., Baugh, C. M., Cole, S., Frenk, C. S., Lacey, C. G., 2000, MNRAS, **316**, 107.

Benson, A. J., Frenk, C. S., Baugh, C. M., Cole, S., Lacey, C. G., 2001, MNRAS, **327**, 1041.

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

Brown, M. L., Taylor, A. N., Bacon, D. J., Gray, M. E., Dye, S., Meisenheimer, K., Wolf, C., 2002, astro-ph/0210213.

Burbidge, G., Napier, W. M., 2001, Astron. J., **121**, 21 (BN01).

Burbidge, G. R., Burbidge, E. M., 1967, Ap. J. Lett., **148**, L107.

Cen, R., Bahcall, N. A., Gramann, M., 1994, Ap. J. Lett., **437**, L51.

Cole, S., Hatton, S., Weinberg, D. H., Frenk, C. S., 1998, MNRAS, **300**, 945.

Colless, M., Dalton, G., Maddox, S., Sutherland, W., Norberg, P., Cole, S., et al. (The 2dFGRS Team), 2001, MNRAS, **328**, 1039.

Couchman, H. M. P., 1991, Ap. J. Lett., **368**, L23.

Croft, R. A. C., Dalton, G. B., Efstathiou, G., Sutherland, W. J., Maddox, S. J., 1997, MNRAS, **291**, 305.

Croom, S. M., Smith, R. J., Boyle, B. J., Shanks, T., Loaring, N. S., Miller, L., Lewis, I. J., 2001, MNRAS, **322**, L29.

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

Dekel, A., Rees, M. J., 1987, Nature, **326**, 455.

Diaferio, A., Geller, M. J., 1996, Ap. J., **467**, 19.

Dressler, A., 1980, Ap. J., **236**, 351.

Efstathiou, G., 1988. Galaxies as Tracers of the Mass Distribution. In *LNP Vol. 297: Comets to Cosmology*, page 312.

Efstathiou, G., Moody, S., Peacock, J. A., Percival, W. J., et al. (The 2dFGRS Team), 2002, MNRAS, **330**, L29.

Evrard, A. E., MacFarland, T. J., Couchman, H. M. P., et al. (The Virgo Consortium), 2002, Ap. J., **573**, 7.

Freedman, W. L., Madore, B. F., Gibson, B. K., et al. (The HST Key Project Team), 2001, Ap. J., **553**, 47.

Fukugita, M., Hogan, C. J., Peebles, P. J. E., 1998, Ap. J., **503**, 518.

Guzzo, L., Bartlett, J. G., Cappi, A., et al. (The ESP Team), 2000, Astron. Astrophys., **355**, 1.

Hamilton, A. J. S., 1992, Ap. J. Lett., **385**, L5.

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

Hamilton, A. J. S., 1997a, MNRAS, **289**, 285.

Hamilton, A. J. S., 1997b, MNRAS, **289**, 295.

Hawkins, E., Maddox, S., Branchini, E., Saunders, W., 2001, MNRAS, **325**, 589.

Hawkins, E., Maddox, S., Cole, S., Madgwick, D. S., Norberg, P., Peacock, J. A., et al. (The 2dFGRS Team), 2002, MNRAS submitted, astro-ph/0212375.

Hawkins, E., Maddox, S. J., Merrifield, M. R., 2002, MNRAS, **336**, L13.

Huchra, J., Davis, M., Latham, D., Tonry, J., 1983, Ap. J. Suppl., **52**, 89.

Jenkins, A., Frenk, C. S., Pearce, F. R., et al. (The Virgo Consortium), 1998, Ap. J., **499**, 20.

Jing, Y. P., Mo, H. J., Börner, G., 1998, Ap. J., **494**, 1.

Jing, Y. P., Börner, G., Suto, Y., 2002, Ap. J., **564**, 15.

Kaiser, N., 1984, Ap. J. Lett., **284**, L9.

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

Kantowski, R., Kao, J. K., Thomas, R. C., 2000, Ap. J., **545**, 549.

Karlsson, K. G., 1990, Astron. Astrophys., **239**, 50 (K90).

Katz, N., Gunn, J. E., 1991, Ap. J., **377**, 365.

Lahav, O., Lilje, P. B., Primack, J. R., Rees, M. J., 1991, MNRAS, **251**, 128.

Lahav, O., Bridle, S. L., Percival, W. J., Peacock, J. A., Efstathiou, G., et al. (The 2dFGRS Team), 2002, MNRAS, **333**, 961.

Landy, S. D., 2002, Ap. J. Lett., **567**, L1 (L02).

Landy, S. D., Szalay, A. S., 1993, Ap. J., **412**, 64.

Landy, S. D., Szalay, A. S., Broadhurst, T. J., 1998, Ap. J. Lett., **494**, L133 (LSB98).

Lawrence, A., Rowan-Robinson, M., Ellis, R. S., et al. (The QDOT Team), 1999, MNRAS, **308**, 897.

Lewis, I. J., Cannon, R. D., Taylor, K., Glazebrook, K., et al., 2002, MNRAS, **333**, 279.

Limber, D. N., 1954, Ap. J., **119**, 655.

Lin, H., Kirshner, R. P., Shectman, S. A., Landy, S. D., Oemler, A., Tucker, D. L., Schechter, P. L., 1996, Ap. J., **471**, 617.

Loveday, J., Efstathiou, G., Peterson, B. A., Maddox, S. J., 1992, Ap. J. Lett., **400**, L43.

Loveday, J., Maddox, S. J., Efstathiou, G., Peterson, B. A., 1995, Ap. J., **442**, 457.

Loveday, J., Tresse, L., Maddox, S., 1999, MNRAS, **310**, 281.

Maddox, S. J., Efstathiou, G., Sutherland, W. J., 1990, MNRAS, **246**, 433.

Maddox, S. J., Efstathiou, G., Sutherland, W. J., 1996, MNRAS, **283**, 1227.

Madgwick, D. S., Hawkins, E., Lahav, O., Maddox, S., et al. (The 2dFGRS Team), 2002a, in preparation.

Madgwick, D. S., Lahav, O., Baugh, C. M., et al. (The 2dFGRS Team), 2002b, MNRAS, **333**, 133.

Mann, R. G., Saunders, W., Taylor, A. N., 1996, MNRAS, **279**, 636.

Mo, H. J., White, S. D. M., 1996, MNRAS, **282**, 347.

Moore, B., Frenk, C. S., Efstathiou, G., Saunders, W., 1994, MNRAS, **269**, 742.

Navarro, J. F., White, S. D. M., 1993, MNRAS, **265**, 271.

Netterfield, C. B., Ade, P. A. R., Bock, J. J., et al. (The Boomerang Team), 2002, Ap. J., **571**, 604.

Norberg, P., Baugh, C. M., Hawkins, E., Maddox, S., Peacock, J. A., Cole, S., Frenk, C. S., et al. (The 2dFGRS Team), 2001, MNRAS, **328**, 64.

Norberg, P., Cole, S., Baugh, C. M., Frenk, C. S., et al. (The 2dFGRS Team), 2002a, MNRAS, **336**, 907.

Norberg, P., Baugh, C. M., Hawkins, E., Maddox, S., Madgwick, D., Lahav, O., Cole, S., Frenk, C. S., et al. (The 2dFGRS Team), 2002b, MNRAS, **332**, 827.

O'Meara, J. M., Tytler, D., Kirkman, D., Suzuki, N., Prochaska, J. X., Lubin, D., Wolfe, A. M., 2001, Ap. J., **552**, 718.

Padilla, N., Baugh, C. M., 2003, astro-ph/0301083.

Pan, J., 2002, Ph.D. Thesis, University of Nottingham.

Pan, J., Coles, P., 2000, MNRAS, **318**, L51.

Park, C., Vogeley, M. S., Geller, M. J., Huchra, J. P., 1994, Ap. J., **431**, 569.

Peacock, J. A., Smith, R. E., 2000, MNRAS, **318**, 1144.

Peacock, J. A., Cole, S., Norberg, P., et al. (The 2dFGRS Team), 2001, Nature, **410**, 169.

Pearce, F. R., Jenkins, A., Frenk, C. S., et al. (The Virgo Consortium), 1999, Ap. J. Lett., **521**, L99.

Peebles, P. J. E., 1980. *"The large-scale structure of the universe"*. Princeton University Press, Princeton N.J.

Percival, W., Sutherland, W., Peacock, J. A., et al. (The 2dFGRS Team), 2002, MNRAS, **337**, 1068.

Perlmutter, S., Aldering, G., Goldhaber, G., et al. (The Supernova Cosmology Project Team), 1999, Ap. J., **517**, 565.

Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1992. *"Numerical Recipes"*. Cambridge University Press, Cambridge.

Press, W. H., Schechter, P., 1974, Ap. J., **187**, 425.

Ratcliffe, A., Shanks, T., Parker, Q. A., Fong, R., 1998, MNRAS, **296**, 191.

Rosenberg, J. L., Salzer, J. J., Moody, J. W., 1994, Astron. J., **108**, 1557.

Rowan-Robinson, M., Sharpe, J., Oliver, S. J., et al. (The PSCz Team), 2000, MNRAS, **314**, 375.

Saunders, W., Rowan-Robinson, M., Lawrence, A., Efstathiou, G., Kaiser, N., Ellis, R. S., Frenk, C. S., 1990, MNRAS, **242**, 318.

Saunders, W., Rowan-Robinson, M., Lawrence, A., 1992, MNRAS, **258**, 134 (S92).

Saunders, W., D'Mellow, K., Tully, B., et al., 2000a, astro-ph/0005604.

Saunders, W., Sutherland, W. J., Maddox, S. J., et al. (The PSCz Team), 2000b, MNRAS, **317**, 55.

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

Sheth, R. K., 1996, MNRAS, **279**, 1310.

Soifer, B. T., Sanders, D. B., Madore, B. F., Neugebauer, G., Danielson, G. E., Elias, J. H., Lonsdale, C. J., Rice, W. L., 1987, Ap. J., **320**, 238.

Somerville, R. S., Lemson, G., Kolatt, T. S., Dekel, A., 2000, MNRAS, **316**, 479.

Springel, V., White, S. D. M., 1998, MNRAS, **298**, 143.

Stoughton, C., Lupton, R. H., Bernardi, M., et al. (The SDSS Collaboration), 2002, Astron. J., **123**, 485.

Szapudi, I., Branchini, E., Frenk, C. S., Maddox, S., Saunders, W., 2000, MNRAS, **318**, L45.

Taylor, A. N., Ballinger, W. E., Heavens, A. F., Tadros, H., 2001, MNRAS, **327**, 689.

Tucker, D. L., Oemler, A., Kirshner, R. P., et al. (The LCRS Team), 1997, MNRAS, **285**, L5.

Valls-Gabaud, D., Alimi, J., Blanchard, A., 1989, Nature, **341**, 215.

Verde, L., Heavens, A. F., Percival, W. J., Matarrese, S., et al. (The 2dFGRS Team), 2002, MNRAS, **335**, 432.

Vogeley, M. S., Szalay, A. S., 1996, Ap. J., **465**, 34.

White, S. D. M., Rees, M. J., 1978, MNRAS, **183**, 341.

Zehavi, I., Blanton, M. R., Frieman, J. A., Weinberg, D. H., Mo, H. J., Strauss, M. A., et al. (The SDSS collaboration), 2002, Ap. J., **571**, 172.

Acknowledgements

These pages represent over three years of my life and would not have been possible without my supervisor Steve - who has had to put up with regular innane ramblings, but always patiently pointed me in the right direction, and Con - who put up with sharing a house with me for all that time.

The friendship of the rest of the astronomy group in Nottingham has been invaluable, especially: Simon - for many nights out, Bo - for endless knowledge of everything, Laura - for her flat, Jesús, Patrick and Pan, though not forgetting, Pepé (the parrot), for providing much amusement all across the world.

Others to have helped keep me sane along the way are far too numerous to name, but include most of the Boat Club - especially Nige, Mitch, Pete B, Pat, and Helen, and fellow astro-peeps Phil and Catherine.

I must also thank the rest of the 2dFGRS Team (though especially Shaun, Peder, Darren, Ofer and John) for all their hard work in constructing the survey and allowing me to be part of such a fantastic project.

Finally, thanks must go to my family for supporting me at every turn.