Power Spectrum Analysis of Redshift Surveys

HELEN TADROS-ATTALLA

St John's College, Oxford

August 1996

A thesis submitted to the University of Oxford in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Declaration

I declare that no part of this thesis has been accepted, or is currently being submitted for any degree or diploma or certificate or any other qualification in this University or elsewhere. This thesis is the result of my own work. Some work in Chapters 2, 4 and 6 involved collaboration with my supervisor, Professor G. Efstathiou and work in Chapter 7 involved a collaboration with Dr. A. Heavens, Dr. A. Taylor and Mr. W. Ballinger at the Royal Observatory in Edinburgh and Professor G. Efstathiou at Oxford. The work in Chapter 2 will be published in *Monthly Notices of the Royal Astronomical Society* as Tadros & Efstathiou (1996) and the work in Chapter 4 is published in *Monthly Notices of the Royal Astronomical Society* as Tadros & Efstathiou (1995).

Power Spectrum Analysis of Redshift Surveys

Helen Tadros-Attalla
St John's College, Oxford

Submitted for the Degree of Doctor of Philosophy
Trinity Term, 1996

Abstract

This thesis describes a study of the clustering properties of galaxies and clusters of galaxies as measured by the power spectrum ($P(k)$) and the counts in cells statistic. The samples used are the optical Stromlo-APM galaxy survey, the APM cluster survey and the IRAS $1.2Jy$, QDOT and PSCz surveys. Throughout, N-body simulations, for a variety of cosmological models, are used to test methods and to supplement analytic error estimates.

For the Stromlo-APM sample the amplitude of the power spectrum is dependent on galaxy morphology. Early–type galaxies show a higher clustering amplitude than late–type galaxies by a factor of ~ 1.8. There is also tentative evidence for some dependence of the clustering amplitude on galaxy luminosity. The parameter $\Omega^{0.6}/b$ is estimated via a comparison with the real-space power spectrum of the two-dimensional APM galaxy survey.

For APM clusters the power spectrum is measured to very small wave numbers, with a possible detection of the expected turn-over. The results are inconsistent with the standard cold dark matter model. The shape of $P(k)$ for clusters is approximately the same as that for Stromlo-APM galaxies but amplified by a factor of ~ 3.5.

The power spectrum of the QDOT sample depends sensitively on the galaxy weighting scheme, probably due the manner in which the region of the Hercules supercluster is sampled. A best estimate of the power spectrum of IRAS galaxies is computed by combining the IRAS 1.2Jy and QDOT samples.

The PSCz galaxy power spectrum is also computed. The PSCz galaxies have a clustering amplitude twice that of optical galaxies. A similar result is found from a joint counts in cells analysis. Redshift-space distortions in the PSCz sample are analysed using a spherical harmonic decomposition of the density field. The value of $\Omega^{0.6}/b = 1$ is ruled out by this analysis at the 2σ significance level.

Acknowledgements

First and foremost I would like to thank my supervisor George Efstathiou, for his patience, help, advice and friendship.

Many other people have advised and guided me over the course of this thesis, and I would especially like to thank Carlton Baugh, Jon Loveday, Will Saunders, Will Sutherland, Andrew Hamilton and Cedric Lacey. Thanks also to the rest of the PSCz team (Will and Will, George, Seb Oliver, Oliver Keeble, Steve Maddox, Carlos Frenk, Michael Rowan-Robinson, Simon White, and Richard McMahon) for their very hard work on the survey and for letting me put results from the survey in my thesis.

I am indebted to my other collaborators, Alan Heavens, Andy Taylor and Bill Ballinger for, well for Chapter 7, but also for their generous hospitality during my stay in Edinburgh, and for helping me appreciate the fact that pubs south of the border have a closing time.

Thanks to the other inmates of my office, Rup Doggy Dog (for always having a different outlook on things and filling our office with toys), Jamboree Stu of Arabia (for being a bringer of food) and Snaussage Serjeant (for being shorter than me). These people made my time in Oxford particularly enjoyable. Somehow, other office mates will never match up. Everyone in the Oxford astrophysics department is to be commended on making it a great place to work. Thanks to all the organising types for the poker evenings, cricket matches, meals, pub quizzes, punting trips and general light relief.

Thanks are also due to The Astronomy Centre at Sussex University, where this thesis was completed and I acknowledge financial support from SERC/PPARC.

Wendy Groom, Andrew Ratcliffe and Gordon Squires helped make a month long summer school seem like 3 weeks and 7 days, but go on admit it, we enjoyed it really. If only Gordon hadn't lost that film....

As for computing matters, Paul Collison, the Starlink manager at Oxford, deserves a medal for his running of the system at Oxford. I also thank Gavin Dalton for miraculously having the answer to almost any computing question, and for his general help and advice. Tony Lynas-Gray is also to be commended on ensuring that computing support is provided even during the unconventional working hours that some of us keep.

I thank my parents for their constant love and support and for not asking (too often) when I was going to finish my thesis. My friend Rashmi Tank deserves a special mention for always being the best friend a person could wish for. I hope I can repay all his kindness.

Finally, I thank Stephen Warren for dragging me kicking and screaming to the end of my thesis and for all his generosity, love and friendship over the past three years.

This thesis is for my parents.

Contents

Chapter 1

Introduction

The origin and nature of the structure which we see in the Universe today is still one of the unanswered questions in the field of cosmology. There have been major advances both observationally and theoretically since the birth of modern cosmology early this century. We still, however, do not know how much matter there is in the Universe, what kind of matter it is, or how the structures we see grew out of the young Universe. The framework within which these questions are made more specific is that of the standard Hot Big Bang model. In this introduction I describe the salient features of this model and I indicate the aspects of the model to which the results of this thesis are relevant.

The introduction is organized as follows: In Section 1.1, I describe the theoretical framework of the Hot Big Bang model and in Section 1.2 the observational support for it is reviewed. In Section 1.3 I briefly outline the stages in the thermal history of the Universe in the standard Hot Big Bang model, moving on to the formation and growth of structure in Section 1.4. The power spectrum, as a measure of the amplitude of density fluctuations, is also introduced in Section 1.4. In Section 1.5, I discuss how the power spectrum we measure from galaxies or clusters in the Universe today differs from that predicted from the theoretical considerations presented in the preceding sections. Section 1.6 contains a summary of the analysis carried out in the rest of this thesis.

1.1 The standard Hot Big Bang model

The standard model assumes that the *Cosmological Principle* applies i.e. that the Universe is homogeneous and isotropic. The metric for such a universe is the Robertson-Walker metric, wherein the line element is given by:

$$ds^2 = c^2dt^2 - a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + sin^2\theta(d\phi^2)\right)\right) \tag{1.1}$$

(see Weinberg 1972). In the above equation, $a(t)$ is the expansion factor of the Universe; physical distances (i.e. proper distances) $\mathbf{r}(r, \theta, \phi)$ are related to comoving distances $\mathbf{x}(r, \theta, \phi)$ by this factor: $\mathbf{r} = a(t)\mathbf{x}$. The geometry of the Universe is expressed in the curvature constant k, which can be either -1, 0 or 1; I will come back to k below. The evolution of the scale factor $a(t)$ is governed by the Friedmann equations which result from solution of Einstein's field equations of General Relativity. The Friedmann equations are

$$\frac{1}{a}\frac{d^2a}{dt^2} = -\frac{4}{3}\pi G\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \tag{1.2}$$

$$H^2 = \frac{8}{3}\pi G\rho - \frac{kc^2}{a^2} + \frac{\Lambda}{3}, \tag{1.3}$$

where Λ is the cosmological constant, p the pressure and ρ the density. The quantity $H(t) = \frac{1}{a}\frac{da}{dt}$ is the 'Hubble constant' (which is a function of time). For Λ equal to zero in equation (1.3) the relation between the curvature constant k and the density of the Universe is:

$$\frac{k}{a^2} = \frac{1}{c^2}\left(\frac{\dot{a}}{a}\right)^2\left(\frac{\rho}{\rho_c} - 1\right), \tag{1.4}$$

where ρ_c is the *critical density*:

$$\rho_c = \frac{3}{8\pi G}\left(\frac{\dot{a}}{a}\right)^2. \tag{1.5}$$

The density parameter, Ω, may now be defined in terms of ρ_c

$$\Omega = \frac{\rho}{\rho_c}. \tag{1.6}$$

Re-writing equation (1.4) we can now relate k to the density parameter,

$$\Omega = 1 + \frac{kc^2}{H^2a^2}. \tag{1.7}$$

Thus, if the Universe has critical density, $k = 0$. In this case, the Universe is flat (Euclidean) and will continue to expand for ever, with the rate of expansion asymptotically approaching zero. For $k = -1$, the Universe has less than the critical density and is open; space is infinite, has uniform negative curvature, and will continue to expand forever. If $k = +1$ the density of the Universe is super-critical, space is closed and at some point in the future the expansion of the Universe will halt and it will collapse back on itself producing the 'Big Crunch'. If there is a cosmological constant Λ, then this can contribute to the closure density and $\Omega_{tot} = \Omega_{matter} + \Omega_\Lambda$, where Ω_Λ is $\Lambda/3H^2$.

We do not know which of these scenarios represents reality due to our ignorance of the density parameter Ω. The value of Ω is one of the questions addressed in this thesis.

1.2 Observational evidence for the standard model

In the standard Hot Big Bang model (hereafter SHBBM), the Universe began at an extremely high density and temperature and uniformly expanded and cooled thereafter. In this section I will briefly outline the observational evidence in support of this scenario. The evidence comes in three main forms, the cosmic microwave background, the Hubble expansion and Big Bang nucleosynthesis. Each of these phenomena are expanded upon in the following sections.

1.2.1 The cosmic microwave background

The main competitor of the SHBBM was the 'Steady State' theory (Bondi & Gold 1948, Hoyle 1948). This was disproved, at least to most people's satisfaction, by the discovery of the relic radiation field of the Big Bang. The cosmic microwave background (CMB) was first observed serendipitously by Penzias and Wilson (Penzias & Wilson 1965) in the same year as it was predicted by Dicke *et al.* (1965) (it was also predicted earlier than this by Gamow 1949). Much theoretical and observational work on the CMB has been carried out since 1965. The most impressive experimental result was the measurement, by the COBE FIRAS (far infrared absolute spectrophotometer) instrument, of the microwave background spectrum. The spectrum of the CMB is a black-body spectrum with a temperature of $T = 2.726 \pm 0.01$K (Mather *et al.* 1994). The deviations from a black-body spectrum are less than 0.03% of the peak intensity over a frequency range 2 to 20cm^{-1}. Small fluctuations in the CMB temperature across the sky, for example those detected by the COBE DMR (differential microwave radiometer) instrument (Smoot *et al.* 1992), provide us with information about the structures in the early Universe.

Photons decouple from baryons at a redshift $z \sim 1000$ and, provided there is no early re-ionization of the inter-galactic medium, they contain information about the structures in the Universe at the time of recombination. The pattern of temperature fluctuations on the sky is determined from so called 'primary' and 'secondary' mechanisms. The primary causes are connected with the state of the density, potential and peculiar velocity fields in the Universe at the time of decoupling. Secondary fluctuations are determined by what happens to a photon on its way to us from the last scattering surface. The Sachs-Wolfe effect (a primary effect, Sachs & Wolfe 1967) determines the power spectrum of temperature fluctuations on large angular scales. The Sachs-Wolfe effect is a gravitational redshift effect; if a photon is last scattered when in a potential well, it will lose energy climbing out and consequently acquire a gravitational redshift. For adiabatic perturbations the Sachs-Wolfe effect is described by $\Delta T/T = \Delta\phi/3$, where $\Delta T/T$ is the temperature difference along two lines of sight, and $\Delta\phi$ is the gravitational potential difference, at the time of last scattering, along the two lines of sight. For isocurvature perturbations the Sachs-Wolfe effect is six times larger than this. So called 'Doppler Peaks' in the power spectrum of temperature fluctuations, on smaller angular scales than the Sachs-Wolfe effect, are also a primary effect. They originate from the acoustic oscillations present in the matter and radiation fluid before recombination (Peebles & Yu 1970, Doroshkevich, Zeldovich & Sunyaev 1978, Bond & Efstathiou 1984).

Secondary effects include the gravitational redshifting of photons caused when they pass through a changing potential field. If we consider the growing mode of the density perturbations in the matter dominated era, we find that the Newtonian gravitational potential field is time independent. When this is the case, the photons suffer no gravitational redshift when they pass through overdense or underdense regions of the Universe on their way to the observer. However, the Newtonian potential field is not always strictly time independent, for reasons detailed below, and the gravitational potential of a region may change while a photon is traversing it. This causes temperature fluctuations in the CMB. All fluctuations caused by varying potential fields are collectively known as the integrated Sachs-Wolfe effects (ISW, Sachs & Wolfe 1967). There are three main ISW effects. Firstly, after recombination, the photons still make a non-negligible contribution to the density of the Universe for a short time. This results in the potential decaying as a function of time, causing the so called early ISW effect. Secondly, a time varying potential will result if the Universe were eventually to become vacuum dominated (which would be the case if there were a cosmological constant Λ) or if the Universe became curvature dominated ($\Omega+\Lambda \neq 1$). Both of these scenarios are only likely to be important at low redshift, and consequently the resulting fluctuations are known as late ISW effects. Thirdly, another important ISW effect is the Rees-Sciama effect (Rees & Sciama 1968) and is important as soon as non-linear structures form. In this case, linear perturbation theory breaks down and it is no longer the case that the Newtonian potential is time independent. The Rees-Sciama effect causes fluctuations in the CMB which are too small to be detected using current experiments, but as sky maps become more sensitive, we should see fluctuations in the CMB corresponding to known structures.

On their way from the last scattering surface, the photons also travel through regions which have been locally re-ionized, such as the hot gaseous atmospheres of rich clusters of galaxies. Here, the photons are inverse Compton scattered off the hot electrons causing a transfer of photons from the low energy Rayleigh-Jeans region of the black-body spectrum into the Wien tail. This causes the Sunyaev-Zeldovich effect (Sunyaev & Zeldovich 1970, Sunyaev & Zeldovich 1980), where the temperature of the CMB in the direction of the cluster is lowered at wavelengths longer than 1.34mm and raised at shorter wavelengths (see Rephaeli 1995 for a review of theoretical and observational work on the Sunyaev-Zeldovich effect). If there was global re-ionization in the Universe the fluctuations on small angular scales would be erased.

There are of course other, local, sources of fluctuations which must be distinguished from cosmological fluctuations. Examples are the contribution to the microwave sky of the Galaxy itself. This can be due to, for example, emission from dust, free-free emission and synchrotron radiation. Also problematic is the emission from extra-galactic sources.

An accurate measurement of the angular power spectrum of temperature fluctuations will provide strong constraints on many of the important cosmological parameters e.g. the density parameter Ω, the baryon fraction Ω_b, the cosmological constant Λ, the spectral index of primordial density fluctuations n and the Hubble constant H_0, provided re-ionization does not occur very early (see Jungman *et al.* 1995 for discussion of determination of cosmological parameters). If the Universe was re-ionized at very early times, our ability to infer these parameters from the observations would be compromised. Recent analyses of CMB data suggest, however, that early re-ionization is unlikely to have occurred (Scott, Silk & White 1995).

Most attention so far has been focussed on the determination of two parameters. The first is n, the index of the power spectrum of initial density fluctuations ($P(k) \propto k^n$, see Section 1.4.1 for the definition of the power spectrum). The second is a quantity related to the amplitude of the sky-averaged quadrupole moment of the angular power spectrum of temperature fluctuations. The latter, generally referred to as Q_{rms} is defined as follows

$$Q_{rms}^2 = \frac{5T_0^2}{4\pi} \left\langle |a_{2m}|^2 \right\rangle , \tag{1.8}$$

where the angle brackets denote a sky average, T_0 is the temperature of the microwave background and the a_{lm} are the coefficients in a spherical harmonic expansion of the pattern of temperature fluctuations on the sky (with $l = 2$ for the quadrupole). Q_{rms} is used to define the normalisation of density fluctuations in the Universe today via a conversion of

Q_{rms} to a value of σ_8, the *rms* mass fluctuation in spheres of radius 8 h^{-1}Mpc[1]. The most recent determinations of these parameters are from the four year COBE data (Bennet *et al.* 1996) and give $n = 1.2 \pm 0.3$, consistent with a Harrison-Zeldovich power spectrum ($P(k) \propto k$), and $Q_{rms} = 15.3^{+3.7}_{-2.8}\mu K$ (Gorski *et al.* 1996).

For a recent review of microwave background observations, see White, Scott & Silk (1994). A review of the determination of cosmological parameters from such observations can be found in Hu & Sugiyama (1995)

1.2.2 The Hubble law and the age of the Universe

Direct observational evidence for the expansion of the Universe was discovered by Hubble (Hubble 1929, see also Sandage 1972), who measured the recession velocity of nearby galaxies to be proportional to their distance from us. The proportionality constant is consequently called the Hubble constant, H_0. Measurement of the value of the Hubble constant is an area of intense research. It is thought to lie between 30km s^{-1}Mpc^{-1} and 100km s^{-1}Mpc^{-1}, although there has been a degree of consensus in the literature of late that H_0 falls in the region 55km s^{-1} to 70km s^{-1}. Most methods for the determination of H_0 now agree to within 1.5 - 2σ of their quoted errors.

Methods using type Ia supernovae as standard candles (after correction for the correlation of absolute magnitude with the shape of the light curve see Riess, Press & Kirshner 1995) give, for example, $H_0 = 65 \pm 10$km s^{-1} (Hamuy *et al.* 1995). Freedman *et al.* (1994) obtain $H_0 = 80 \pm 17$km s^{-1} using Hubble Space Telescope Cepheid distances to M100 in the Virgo cluster. Much of the uncertainty in this measurement is due to the complicated spatial structure of this cluster (Binggeli, Tammann & Sandage 1987, Fukugita, Okamura & Yasuda 1993), and the resulting uncertainty in the position of M100 relative to the core of the cluster. Tanvir *et al.* (1995) use the more compact group of galaxies Leo-I and obtain a Cepheid distance to the galaxy M96. They derive $H_0 = 69 \pm 8$km s^{-1}. Schmidt *et al.* (1994) have used the expanding photosphere method on type II supernovae to obtain $H_0 = 73 \pm 6 \pm 7$km s^{-1} (statistical and systematic errors respectively). Recent Sunyaev-Zeldovich effect measurements give, for example, 65 ± 25km s^{-1} (Birkinshaw & Hughes 1994). Measurements of H_0 derived from time delays between luminosity variations in multiple images of a gravitationally lensed source are also giving consistent results, the greatest uncertainty here being the modelling of the mass distribution along the line of sight to the

[1]Throughout this thesis I write the Hubble constant as $H_0 = 100h$ km s^{-1} Mpc^{-1}. Where relevant, absolute magnitudes are quoted assuming $h = 1$ unless otherwise stated.

source (e.g. see Grogin & Narayan 1996). The method of surface brightness fluctuations (Tonry & Schneider 1988) yields, for example, $H_0 = 78 \pm 6$ (Tonry, Ajhar & Luppino 1990), although the values from the surface brightness fluctuation method generally fall on the high side of the range of H_0 values in the literature.

The implications of a high value of H_0 are disturbing to those who would like to think that $\Omega = 1$. In an Einstein-de Sitter universe the age is given by $t_0 = 2/3H_0 \sim 6.533/h\ Gyr$. Using $H_0 = 70\text{km s}^{-1}$ gives $t_0 = 9.3\ Gyr$. The Universe is constrained, however, to be at least as old as the objects of which it is comprised. Consequently, we would like to measure the age of the Universe directly. Considerable effort has been concentrated on determining the ages of globular clusters, mainly via isochrone fitting (Sandage 1982, VandenBerg 1983). The resulting age determinations lie between 14 Gyr (VandenBerg 1988) and 19 Gyr (Salaris, Cheiffi & Straniero 1993). In an Einstein-de Sitter universe, this upper bound is at odds with a Hubble constant $> 34\text{km s}^{-1}\text{Mpc}^{-1}$, and even the lower bound constrains the Hubble constant to be $< 47\text{km s}^{-1}\text{Mpc}^{-1}$. The situation is alleviated by allowing Ω to be less than unity, or by introducing a cosmological constant. However, it is difficult to find cosmological parameters which would allow both a high value of the Hubble parameter, and globular clusters to be as old as is currently estimated.

Recently, appreciably lower values for the lower limit on globular cluster ages have appeared in the literature. Re-analyses of the globular cluster ages (Chaboyer *et al.* 1996) have attempted to account for all the uncertainties in the models of stellar structure. These include the lack of a complete theory of convection in stellar interiors as well as uncertainties in some nuclear reaction rates, in stellar opacities and in the abundance of elements such as He, O, Mg, Si, S and Ca. Chaboyer *et al.* (1996) tackle this problem by using Monte-Carlo simulations and varying the input parameters within the allowed ranges. They find that the lower bound on the age of the oldest globular clusters is 12.07 Gyr (95% confidence). Jimenez *et al.* (1996) have developed a new method for the estimation of globular cluster ages, using the morphology of the horizontal branch in the Hertzprung-Russell diagram of globular clusters. They find that the lower limit for the oldest clusters is 9.7 Gyr.

1.2.3 Big Bang nucleosynthesis

Because the Universe began at formally infinite density and temperature in the SHBBM, it passed through a stage in its thermal history ($0.1s \lesssim t_{BBN} \lesssim 10^3 s$) in which it became a primordial nuclear reactor. The SHBBM thus makes verifiable predictions concerning the relative primordial abundances of elements in the Universe. These ideas were first formulated by Alpher, Bethe & Gamow (1948) and a detailed network of nuclear reactions was used by Wagoner, Fowler & Hoyle (1967) to calculate abundances of the light elements (for more recent work, see Boesgaard & Steigman 1985 and Walker *et al.* 1991).

Big Bang nucleosynthesis (BBN) models have had considerable success up to now in matching the observed abundances of the light elements D, ^3He, ^4He and ^7Li. The abundances of these elements relative to hydrogen span a range of nine orders of magnitude, and the BBN model has essentially only one free parameter, η, the ratio of baryon to photon number density. The predicted abundances of elements depend on the rate of expansion of the Universe during nucleosynthesis, which in turn is proportional to the square root of the density, ρ, of the Universe. During the epoch of nucleosynthesis

$$\rho = \rho_\gamma + \rho_e + N_\nu \rho_\nu, \tag{1.9}$$

where ρ_γ is the density in photons, ρ_e is that in electron-positron pairs and ρ_ν is the contribution from one species of light neutrino (Steigman, Schramm & Gunn 1977). If N_ν, the number of massless (or very light) neutrino species (or other light particles) increases, the rate of expansion increases. The ratio of the number of neutrons to protons maintains its equilibrium value as long as the relevant reaction rates ($n+e^+ \rightleftharpoons p+\overline{\nu}_e$, $p+e^- \rightleftharpoons n+\nu_e$, $n \rightleftharpoons p+e^-+\overline{\nu}_e$) are fast compared with the expansion rate of the Universe. If the expansion rate increases, then these weak interactions are taken out of equilibrium at a higher temperature, resulting in a higher ratio of residual neutrons to protons. Consequently, the mass fraction of ^4He in the Universe increases as N_ν increases, as most residual neutrons are incorporated into this element.

It is not straightforward to compare the predicted primordial abundance values with those measured in the Universe, as the abundance of elements will have been altered by astrophysical processes. In the following paragraphs I briefly outline how the primordial abundances of the light elements are inferred from present day observations.

Deuterium is not produced by any stellar nuclear process after the BBN era, and consequently any measurement of the abundance of deuterium provides a lower limit on the primordial value. Due to the strong monotonic decrease of the abundance of deuterium as a function of η, an observational uncertainty of 40% on the value of its abundance would lead to a measurement of η with an uncertainty of only $\sim 25\%$. Direct observations of the amount of deuterium in high redshift systems would be a useful measurement of primordial abundance. The deuterium abundance can be obtained by measuring the column density of deuterium in a metal poor gas cloud between us and a distant quasar. In some systems such measurements have given quite high values for the D/H ratio e.g. 2.5×10^{-4} (Carswell *et al.* 1994, Songaila *et al.* 1994). However, there is a worry that observational uncertainties such as contamination by hydrogen absorption in intervening clouds causes these measurements to be spurious. More recent measurements in well understood systems are finding a much lower value for the primordial deuterium abundance of $\sim 2.4 \times 10^{-5}$ (Tytler, Fan & Burles 1996, Burles & Tytler 1996).

Since deuterium is burnt in stars to form ^3He, and some of this ^3He survives stellar processing, the observed abundance of $^3He + D$ should be correlated with the primordial value of this sum. In measurements of this quantity, the uncertainties in the chemical evolutionary models are parameterised by the quantity g_3, the survival fraction of ^3He.

Measurements of ^7Li are obtained from old, very metal poor, population II halo stars (Spite & Spite 1982) and require little extrapolation to obtain primordial values. An upper limit on the amount of ^7Li provides both an upper and lower limit on η because the predicted ^7Li abundance as a function of η has a minimum. ^4He observations also require little extrapolation since they are made in low metallicity extragalactic HII regions by observing absorption lines at radio frequencies (Wilson & Rood 1994) and in the optical (e.g. Skillman *et al.* 1994).

Using the inferred values for the primordial abundances of the light elements, two main tests are applied to the BBN models. Firstly, the upper limit placed on η (η_{MAX}) from the measurements of primordial abundances of ^4He, ^7Li and $D+^3He$ must be higher than the lower limit placed on this quantity (η_{MIN}). The second test is that the observed primordial abundance of ^4He should agree with the predicted value, or at least that the upper limit of the primordial value from observations should be greater than the predicted lower limit from BBN.

At first, the measurements and predictions were in good agreement (Yang *et al.* 1984, Walker *et al.* 1991). However, more recently the situation has changed. Allowance for the fact that ^3He is made in low mass stars (Iben 1967, Rood 1972, Iben & Truran 1978), which could dominate over the primordial production (Rood, Steigman & Tinsley 1976), as well as uncertainty in the value of g_3, has raised the lower bound on η. It is still the case that $\eta_{MAX} > \eta_{MIN}$ but, increasing η_{MIN} increases the lower limit on the predicted primordial amount of ^4He. The amount of primordial ^4He inferred from measurements is $^4He/H = 0.232 \pm 0.003$ (Olive & Steigman 1995), whereas the predicted value is now $^4He/H > 0.241$ for $\eta \geq 3.1 \times 10^{-10}$ (Geiss 1993, see Hata *et al.* 1995 for detailed statistical analysis of this problem). These values of the ^4He abundance disagree seriously.

Possible explanations for the discrepancy are three-fold. Firstly, it could be due to our ignorance concerning the parameter g_3. The disagreement would be removed if $g_3 \sim 0.1$. The current theoretical lower bound on this number is $g_3 \geq 0.25$ (Dearborn, Schramm & Steigman 1986). A second possibility is that the primordial value of ^4He, inferred by extrapolation of observations made at the present epoch, is in error. Even though the regions in which the ^4He abundance measurements are made have metallicities $\sim 1/40$ of the solar value, a correction must still be applied for the amount of ^4He synthesized since BBN. This is done by correlating the amount of ^4He observed with the metallicity and extrapolating to zero metallicity. There are sources of possible systematic error in the measurement of the ^4He abundance in these regions, however. These include uncertainties in ionization corrections, collisional excitation corrections, corrections for dust and corrections for stellar absorption, and have been discussed recently by Copi, Schramm & Turner (1995), Sasselov & Goldwirth (1995) and Olive & Steigman (1995). The third possibility is more exciting. Kawasaki *et al.* (1994) have shown that if the τ neutrino had a mass $5 \lesssim m_{\nu_\tau} \leq 24 Mev$, the upper limit supplied by the ALEPH collaboration experiments, then the effective number of neutrinos N_ν in equation (1.9) is less than 3. Decreasing N_ν decreases the predicted amount of ^4He. It should be noted, however, that Kernan & Sarkar (1996) have suggested that there is no crisis for BBN if attention is restricted only to direct measurements of primordial abundances rather than those which rely on models of chemical evolution.

Thus, BBN has had considerable success in explaining the origin of the elements and may even be making predictions concerning the mass of the τ neutrino. These predictions may be checked by modern day particle physics experiments.

The three phenomena described above, the CMB, the Hubble velocity field and Big Bang nucleosynthesis provide the pillars on which the standard model rests.

1.3 Thermal history of the Universe

This section briefly outlines the main stages in the thermal history of the Universe from the Planck time up to the epoch, described in detail in Section 1.4, at which structure starts to form. The Planck time ($t \sim 10^{-43}s, T \sim 10^{31}K$) represents the temporal limit of present theories of particle physics. Previous to this time, Quantum Gravity, for which we currently have no compelling description, is of vital importance. At very early times the Universe consisted of a plasma of relativistic particles. These included quarks, leptons, gauge bosons, and Higgs bosons. The Universe evolved by undergoing a series of phase transitions, one of the earliest being the transition which marked the end of Grand Unification. This occurred at $t \sim 10^{-34}s$, $T \sim 10^{26}$K, and it is possibly at this time that the matter-antimatter asymmetry originated, although there has been recent interest in the scenario where this occurs at a later time ($t \sim 10^{-12}s$, $T \sim 10^{15}$K), during the epoch in which the weak and electromagnetic fields become decoupled (see Cohen, Kaplan & Nelson 1993 for a review). At $t \sim 10^{-6}$, $T \sim 10^{13}$K, the quark-hadron transition occurs in which quarks are confined and baryons (protons and neutrons) and mesons are formed. The era of primordial nucleosynthesis is the next significant stage ($t \sim 0.1s$ - $10^{3}s$, $T \sim 10^{9}$ - 10^{11}K), when the ratio of neutrons to protons freezes out at a value $\sim 1/6$ and the light elements are synthesised. Neutrinos also decouple from the plasma and consequently form a neutrino background which should be at a temperature of $\sim$ 2K today. By $t \sim 1000yr$, $T \sim 10^{5}$K, the Universe passes into a matter dominated era. The recombination of electrons and protons, which results in the cosmic microwave background of photons occurs at $t \sim 10^{6}yr$, $T \sim 3000$K.

1.4 The formation and evolution of structure

In the following sections I outline how structure grows in the Universe by gravitational amplification of small initial density perturbations. In Section 1.4.1 I summarize ideas on the origin of the initial density perturbations. I concentrate on the hypothesis that the fluctuations originated from quantum fluctuations generated during a period of near-exponential expansion of the Universe (inflation). Section 1.4.2 contains an outline of how these small fluctuations are amplified by gravitational instability and how non-gravitational processes affect growth. The dependence of the growth of structure on the matter content in the Universe is described in Section 1.4.3. In Section 1.4.4 I discuss the present day shape of the power spectrum of density fluctuations predicted in inflationary models of the early Universe.

1.4.1 Inflation

There are a few other ingredients which need to be added to the standard model to form the structure we observe today. Although the Universe appears almost homogeneous on the largest scales there must have been small initial fluctuations in the density field which thereafter grew by gravitational instability into the galaxies and clusters that we see. The origin of these fluctuations must be specified.

At present, the theory of inflation provides a popular model for the origin of fluctuations. The inflationary theory was originally formulated by Guth (1981), but since then many flavours of inflation have been proposed (see Peacock 1996 for a recent pedagogical review of inflation and the resulting structure formation, also Liddle & Lyth 1993 and Brandenberger 1990). In simple inflationary models there is a potential field $V(\phi)$ which determines the evolution of the scalar field ϕ. This scalar field is generally referred to as the inflaton field. The Universe passes through an epoch where this is the dominant contributor to the energy density. During this epoch, the scale factor increases exponentially with the number of e-foldings depending on the exact details of the model.

The structure in the Universe, in inflationary models, originates via gravitational amplification of quantum fluctuations of the inflaton field. These microscopic fluctuations become inflated to classical fluctuations. The resulting spectrum of density fluctuations is nearly scale-invariant i.e. $P(k) \propto k$, where $P(k)$, the power spectrum, is defined as

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

The angle brackets in equation (1.10) denote an ensemble average, and the $\delta_{\mathbf{k}}$ are the Fourier components of the overdensity field (see Section 1.4.2).

In generic models of inflation, the spectrum of initial perturbations is Gaussian (i.e. the $\delta_{\mathbf{k}}$ are distributed as $\delta_{\mathbf{k}} = A(k)e^{i\theta_k}$, where θ_k is a uniform random variate between 0 and 2π). The perturbations are also adiabatic. There exist models, however, which produce scale-invariant *entropy* perturbations in a natural way (see e.g. the axion dominated model, Efstathiou & Bond 1986). A directly testable result of inflation is the production of a gravitational wave background (Rubakov, Sazhin & Veryakin 1982, Starobinskii 1983).

As well as providing density fluctuations in the early Universe, inflation solves a number of other worrying problems. One of these is the homogeneity of the microwave background — the 'horizon problem'. The temperature of the CMB is the same in all directions to within 1 part in 10^5. In the standard (no-inflation) model however, the horizon size of the last scattering surface now subtends an angle of about $1°$ on the sky. The question then arises, how can regions which could not have been in causal contact at the time of recombination now have the same temperature to such a high degree of accuracy. In the inflationary scenario, the whole of the last scattering surface observed by COBE originated in one causally connected region. A second problem solved by inflation is the so-called 'fine-tuning' (or flatness) problem. The problem can be appreciated as follows. From the definition of Ω we have

$$\Omega(t) = \frac{\rho(t)}{\frac{3H^2}{8\pi G}}. \quad (1.11)$$

The Friedmann equation (1.3) may be rearranged to give

$$\frac{3H^2(t)}{8\pi G} = \rho(t) - \frac{3kc^2}{a^2 8\pi G} \quad (1.12)$$

(setting $\Lambda = 0$). Substituting equation (1.12) into equation (1.11) and re-arranging we obtain

$$a^2(t)\rho(t)\left(\Omega^{-1} - 1\right) = \frac{-3kc^2}{8\pi G} = \text{constant} = \left(\Omega_0^{-1} - 1\right)\rho_0 a_0^2. \quad (1.13)$$

We then use the relations between the density ρ and the scale factor a, which are

$$\rho(t) = \rho_{eq}\left(\frac{a_{eq}}{a}\right)^4 \quad (z > z_{eq}) \quad (1.14)$$

and

$$\rho(t) = \rho_0\left(\frac{a_0}{a}\right)^3 \quad (z < z_{eq}), \quad (1.15)$$

where the subscript eq refers to the epoch at which the Universe becomes matter dominated. We are now in a position to find a relationship between Ω at a very early time and Ω today (Ω_0). This relationship is

$$\left(\Omega^{-1} - 1\right) = \left(\Omega_0^{-1}\right)(1 + z_{eq})^{-1}\left(\frac{T_{eq}}{T}\right)^2 = \left(\Omega_0^{-1} - 1\right)10^{-60}\left(\frac{T_p}{T}\right)^2 \quad (1.16)$$

with T being temperature, and p signifying the Planck time. If, today, $|\Omega_0^{-1} - 1|$ is of order unity, then at early times the value of Ω must have been very finely tuned to almost exactly unity. If we go back to the Planck time, t_p, then Ω must have been tuned to one part in 10^{60}. This is not very satisfactory as it seems highly unlikely.

Having an inflationary epoch where the Universe expands exponentially solves this problem. During the inflationary epoch the Universe has an equation of state with a negative pressure, i.e. it behaves like a universe which is dominated by a cosmological constant. In this case the scale factor grows as follows:

$$a(t) \propto sinh(Ht) \qquad k = -1 \quad (1.17)$$

$$a(t) \propto cosh(Ht) \qquad k = +1 \quad (1.18)$$

$$a(t) \propto exp(Ht) \qquad k = 0. \quad (1.19)$$

So, equation (1.17) and (1.18) evolve to look like equation (1.19) i.e. all universes tend to the $k = 0$ case. To solve the flatness problem (and the horizon problem) there must be at least some 60 e-folds of expansion (see e.g. Peacock 1996).

From this discussion it is then clear that generic inflation models produce a universe which is spatially flat i.e. $\Omega = 1$, or if there is a cosmological constant, $\Omega + \Lambda = 1$. It is possible however, to construct inflationary models in which this is not the case, but again a degree of fine tuning is generally required, for example the exact number of e-folds may need to be tuned. I do not discuss these models here, but examples can be found in Gott & Statler (1984), Linde & Mezhlumian (1995), Bucher & Turok (1995).

A period of exponential expansion in the Universe would also dilute away the predicted, but unobserved, relics of the end of Grand Unification, for example magnetic monopoles.

Other possible origins for the initial seeds required for the growth of structure include topological defects such as monopoles, domain walls, cosmic strings and textures. A feature of models in which topological defects seed structure formation is that the initial density field is non-Gaussian. The signatures of non-Gaussianity predicted by these models can be searched for in the density field of galaxies obtained from redshift surveys (Weinberg & Cole 1992, Nusser & Dekel 1993, Stirling & Peacock 1996). I do not attempt to compare observational data with the predictions of defect models in this thesis, mainly due to a lack of suitable numerical simulations. More details on these types of cosmological scenarios may be found in Vilenkin & Shellard (1993).

In the following section I describe the growth of structure from initial adiabatic fluctuations. If the initial fluctuations are entropy fluctuations, structure formation proceeds in a different manner (see, for example, Section 6 of Efstathiou 1990).

1.4.2 The growth of structure

The growth rate of the initial perturbations in the density field depends on many factors. For example, it depends on whether we consider the radiation dominated era or the mass dominated era. It also depends crucially on the nature of the dark matter. In this section I outline the growth of structure using a Newtonian approach. We can follow the growth of perturbations with the Newtonian equations which govern fluid-flow, provided that the sizes of the perturbations are smaller than the Hubble radius (ct). If we ignore pressure, the relevant equations are,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \qquad \text{The continuity equation,} \tag{1.20}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\nabla \Phi \qquad \text{The Euler equation,} \tag{1.21}$$

$$\nabla^2 \Phi = 4\pi G \rho \qquad \text{The Poisson equation.} \tag{1.22}$$

To proceed, we make four modifications to the above equations. Firstly, we convert from proper to comoving coordinates,

$$\mathbf{r} = a(t)\mathbf{x} \tag{1.23}$$

$$\mathbf{v} = \dot{a}\mathbf{x} + a\dot{\mathbf{x}}. \tag{1.24}$$

Secondly, we write the spatially varying density as:

$$\rho(\mathbf{x}, t) = \overline{\rho}(t)\,(1 + \delta\,(\mathbf{x}, t)) \tag{1.25}$$

where $\delta(\mathbf{x}, t)$ is the fractional density perturbation at the space-time point $(\mathbf{x}, t)$. The third step is to Fourier transform equations (1.20) to (1.22) using relations of the form

$$\delta\,(\mathbf{x}, t) = \sum_{\mathbf{k}} \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{1.26}$$

$$\delta_{\mathbf{k}}(t) = \frac{1}{V}\int \delta e^{-\mathbf{k}\cdot\mathbf{x}} d^3\mathrm{x}. \tag{1.27}$$

The last step is to neglect second order terms. For the linear evolution of $\delta_{\mathbf{k}}$, this results in the equation

$$\frac{d^2\delta_{\mathbf{k}}}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta_{\mathbf{k}}}{dt} - 4\pi G\overline{\rho}\delta_{\mathbf{k}} = 0. \tag{1.28}$$

In an $\Omega = 1$ matter dominated universe the scale factor a varies as $t^{\frac{2}{3}}$, and the solution to equation (1.28) is

$$\delta_{\mathbf{k}}(t) = A_k t^{\frac{2}{3}} + B_k t^{-1}. \tag{1.29}$$

If $\Omega \ll 1$ then $a(t) \propto t$ for $z \ll \frac{1}{\Omega_0} - 1$, and the solution to equation (1.28) becomes

$$\delta_{\mathbf{k}}(t) = A_k + B_k t^{-1}. \tag{1.30}$$

In this case then, the clustering 'freezes' at $z \sim 1/\Omega_0 - 1$.

When the Universe is in the radiation dominated era, the relevant equation for the evolution of $\delta_{\mathbf{k}}$ is slightly different (see Section 2.3 of Efstathiou 1990) and the evolution of density perturbations follows

$$\delta \propto 1 + \frac{3}{2}\eta \qquad \text{Growing mode,} \tag{1.31}$$

$$\delta \propto \left(1 + \frac{3}{2}\eta\right) ln \left[\frac{(1+\eta)^{\frac{1}{2}} + 1}{(1+\eta)^{\frac{1}{2}} - 1}\right] - 3(1+\eta)^{\frac{1}{2}} \qquad \text{Decaying mode,} \tag{1.32}$$

where η is the ratio of the density in matter to the density in relativistic particles. Thus, the fluctuations cannot grow significantly in the radiation dominated era (i.e. when $\eta < 1$). This phenomenon is sometimes called the Meszaros effect and is physically caused by the fact that the time scale for the collapse of a perturbation is longer than the characteristic time scale of expansion.

For density perturbations of size exceeding the Hubble radius ($= ct$) a general relativistic approach must be used (see for example Efstathiou 1990). The result is as follows. In the radiation dominated era, we find

$$\delta(t) = At + Bt^{-1}, \tag{1.33}$$

where $\delta(t)$ is defined in the synchronous gauge and A and B are constants. In the matter dominated era

$$\delta(t) = At^{\frac{2}{3}} + Bt^{-1}, \tag{1.34}$$

as in equation (1.29).

Some further work is now required as the Universe does not behave like a perfect fluid. as we have assumed up until now. The baryon component will be subject to various dissipative processes which we must consider. A detailed account of these process is given in Peebles (1981). Here I provide a brief outline.

By including a pressure term in equation (1.21) we would find that equation (1.28) is modified to

$$\frac{d^2\delta_{\mathbf{k}}}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta_{\mathbf{k}}}{dt} = \left[4\pi G\bar{\rho} - \left(\frac{c_s k}{a}\right)^2\right]\delta_{\mathbf{k}}, \tag{1.35}$$

where c_s is the sound speed. The 'Jeans length' is defined in terms of this speed as follows,

$$\lambda_J = c_s\left(\frac{\pi}{G\rho_m}\right)^{\frac{1}{2}}. \tag{1.36}$$

For perturbations larger than this length, the first term in square brackets in equation (1.35) is dominant and the analysis described previously is applicable. However perturbations smaller than the Jeans length (for which the second term in square brackets in equation 1.35 dominates) will oscillate acoustically. The sound speed depends on the nature of the medium. Before recombination, the Jeans length is of order supercluster scales, while the post-recombination Jeans length falls to values appropriate for globular cluster sized objects (see e.g. Efstathiou 1990).

Another important process is that of Silk damping (Silk 1968). Before recombination, the photons are tightly coupled to the electrons by the process of Thomson scattering. However, the coupling is not perfect and the photons can execute a random walk of length $\lambda_D \sim c(tt_c)^{\frac{1}{2}}$ where c is the speed of light, t_c is the time between collisions and t is the duration of the random walk. The photons will thus be able to leak away from overdense regions into underdense regions, taking the matter with them (since they are coupled to it). This process damps the amplitude of adiabatic perturbations in the pre-recombination era. Numerical computations of the Silk damping mass give $M_D \sim 1.3 \times 10^{12} \left(\Omega h^2\right)^{-3/2} M_{\odot}$ (Peebles 1981). This mass corresponds to that of large clusters of galaxies, for small Ω, so perturbations of this size and smaller would be damped away. It should be noted that entropy perturbations are not subject to Silk damping (see Efstathiou 1990). After recombination, when the baryons are no longer coupled to the photons, they fall into the potential wells defined by the dark matter fluctuations.

It is not clear how the luminous objects which subsequently form trace the underlying dark matter. For instance we may expect luminous galaxies only to form at the highest peaks in the density field. High peaks in the density field have an enhanced clustering amplitude; the amount of enhancement depends on the height of the peak (Bardeen *et al.* 1986). This type of 'natural bias' has been discussed for example by White *et al.* (1987a). These authors use numerical simulations of structure formation in a cold dark matter universe and find that bright galaxies (i.e. galaxies with high circular velocities) have a correlation function which is a constant factor larger than the mass correlation function, with this factor increasing with the mass of the galaxies. It is also possible that the morphological type of the galaxy that forms at a particular point depends on the nature of the environment.

We generally assume that (at least on large scales) the density field of galaxies is proportional to that of the mass,

$$\delta_{gal}(r) = b\delta_{mass}(r) \tag{1.37}$$

where b is the 'linear biasing parameter' and the power spectra of the galaxies and the mass are related by $P(k)_{gal} = b^2 P(k)_{mass}$. The quantity b becomes another free parameter in large scale structure measurements. Of course, it is not necessarily the case that the galaxy density field bears a simple, linear relation to that of the mass. The efficiency of galaxy formation at any point may be influenced by the state of the inter-galactic medium at that point. For instance the presence of a quasar may enhance or suppress the probability of galaxy formation. It is not inconceivable that effects of this kind influence galaxy formation over scales of order 100Mpc (Bower *et al.* 1993). Some of the work in this thesis will address the problem of the bias parameter.

Dark matter perturbations are subject to the process known as 'free-streaming'. The details of this process depend on the nature of the dark matter, as this determines the free-streaming length. For example, if the dark matter is composed of neutrinos with rest masses of a few eV, then perturbations that enter the Hubble radius during the time when neutrinos are still relativistic are damped away, since the neutrinos can move out of the perturbations at the speed of light without interaction. The damping length λ_{fs} in this cosmological scenario is $13(\Omega h^2)^{-1}$Mpc (Bond & Szalay 1983), and the first structures to form are larger than rich clusters of galaxies. If the Universe is composed mainly of cold dark matter (CDM) particles, the damping length is much shorter, since by definition the CDM particles are assumed to become non-relativistic at a very early epoch. Thus if the dark matter is hot (e.g. neutrinos) structure forms in a 'top-down' manner while if it is cold, the structure forms 'bottom up'.

1.4.3 The transfer function

The linear transfer function $T(k)$ encapsulates the dependence of the growth of structure on the type of matter in the Universe and it can generally be calculated numerically for a given form of matter. The power spectrum of density fluctuations in the linear regime at a time t_0 is related to the power spectrum at an earlier time t_i as follows,

$$P(k,t_0) = T^2(k)P(k,t_i). \tag{1.38}$$

Transfer functions have been calculated for most cosmologically interesting models. For example, the baryon dominated model has been studied by Peebles (1981), Wilson & Silk (1981), and Kaiser (1983). Scale-invariant adiabatic baryon dominated models are not compatible with observations; fluctuations in the matter distribution on large scales, much larger than those measured, are predicted by this model, as well as microwave background fluctuations in excess of those observed. Bond & Szalay (1983) calculate the transfer function for the adiabatic massive neutrino dominated universe and find $P(k,t) \propto k^n 10^{-2\left(\frac{k}{k_\nu}\right)^{1.5}}$. The free-streaming length is $\lambda_{fs} = 2\pi/k_\nu$.

The transfer function has also been calculated for a cold dark matter universe (see for example Bond & Efstathiou 1984):

$$T(k) = \frac{1}{\left[1 + \left(ak + (bk)^{\frac{3}{2}} + (ck)^2\right)^{\nu}\right]^{\frac{1}{\nu}}} \tag{1.39}$$

where $a = 6.4\,(\Omega h^2)^{-1}$ Mpc, $b = 3.0\,(\Omega h^2)^{-1}$ Mpc, $c = 1.7\,(\Omega h^2)^{-1}$ Mpc and $\nu = 1.13$. This formula is valid if the baryons make a negligible contribution to the density of the Universe. Thus calculation of the transfer function for any cosmological model gives us the shape of the linearly evolved power spectrum.

1.4.4 The turn-over in the power spectrum

If the initial power spectrum of density fluctuations has the Harrison-Zeldovich form: $P(k) \propto k$, then at some point in k space the evolved power spectrum changes smoothly from the initial $P(k)$ to one which is decreasing as a function of k. This is an important feature in the spectrum of evolved density perturbations. This so called 'turn-over' is quite gentle, but the position of the maximum is a crucial indicator of the nature of the dark matter and the quantity Ωh (which will sometimes be referred to as Γ in this thesis). The reason for a turn-over can be seen as follows. Short wavelengths, which enter the Hubble radius during the epoch of radiation domination experience no growth (see equation 1.31). Their amplitudes remain frozen until the Universe passes into the matter dominated era. For these wave numbers we get $\delta = A(k)\,(k\lambda_i)^{-2}$ where λ_i is the initial wavelength. So, for these small wavelengths, the power spectrum scales as $P(k) \propto k^{n-4}$. Longer wavelengths, which do not enter the Hubble radius until the epoch of matter domination, experience growth according to equation (1.29) and the shape of the initial power spectrum at these wavelengths is preserved, $P(k) \propto k^n$. The scale of the peak in $P(k)$ occurs at about the wavelength which is equal to the Hubble radius at the time of matter-radiation equality, i.e.

$$\lambda_{turn-over} = \frac{ct_{equ}a_0}{a(t_{equ})} \simeq 10\left(\Omega h^2\right)^{-1} \text{Mpc}. \tag{1.40}$$

1.5 Measuring the power spectrum

It is clear from the previous section that measuring the point of the turn-over in the power spectrum of the galaxy distribution would provide strong constraints on the cosmological scenario. The turn-over scale has not yet been unambiguously identified. To obtain an accurate measurement of the power spectrum on the scale of the turn-over requires a very large data set. One needs to survey a large volume of the Universe to reduce cosmic variance in the measurement of $P(k)$, as well as requiring many galaxy redshifts to reduce the effect of Poisson noise. There are also systematic biases in the measurement of $P(k)$ which cause us to underestimate the power spectrum at the largest wavelengths (see Peacock & Nicholson 1991 and the appendices of this thesis). Thus, artifacts of our analyses can masquerade as a turn-over in the power spectrum.

One of the main aims of this thesis is to constrain the power spectrum of galaxy clustering using several large data sets. The power spectrum measured from a redshift survey differs, in a number of ways, from the linear power spectrum of the mass distribution described in the previous sections. The processes which modify the linear theory power spectrum are described in the following sections.

1.5.1 The bias parameter

Firstly, as alluded to earlier, we measure the power spectrum of luminous objects, which may or may not trace the underlying mass. We parameterised this uncertainty in terms of the bias parameter b. Knowing the details of the process of galaxy formation would give us a better idea about the way in which galaxies trace the mass distribution. It would be hoped however that clusters of galaxies trace the underlying mass in a linear fashion since they correspond to rare, massive agglomerations of the dark matter. The power spectrum of rich APM clusters is calculated in Chapter 3. The power spectrum of APM clusters is then compared with that of galaxies in order to place some constraints on the biasing parameter. To assume that the bias parameter b does not depend on scale, at least on large scales. is probably fair. This is what most models, in which galaxies form at the high peaks in the density field, would predict (Bardeen *et al.* 1986). It is also suggested by various analyses of higher order moments of the density field. Cooperative models of galaxy formation (Babul & White 1991, Rees 1986, Efstathiou 1992, Bower *et al.* 1993), where the formation of objects influences the later collapse of structure on a large scale, would result in a more complicated function for the bias parameter. I comment on the validity of such models in Chapter 3.

1.5.2 Redshift-space distortions

A second difference between the power spectrum, as measured from galaxy redshift surveys, and that predicted by models of structure formation is the effect of peculiar velocities (Kaiser 1987). In linear theory, the galaxy peculiar velocity field is coherent and results from infall of matter onto overdensities and outflow of matter from voids. The effect this has on the power spectrum on large scales is to increase the amplitude of $P(k)$. Kaiser (1987) derived a simple expression, valid for wavelengths in the linear regime which subtend a small angle at the position of the observer, which relates the real-space power spectrum $P_r(k)$ to that in redshift-space $P_s(k)$

$$P_s\left(\mathbf{k}\right) = P_r\left(\mathbf{k}\right)\left(1+\beta\mu^2\right)^2. \tag{1.41}$$

In this equation $\beta = \Omega^{0.6}/b$, and μ is the angle between the vector $\mathbf{k}$ and the line of sight. After angle averaging over μ this gives

$$P_s\left(k\right) = P_r\left(k\right)\left(1+\frac{2}{3}\beta+\frac{1}{5}\beta^2\right). \tag{1.42}$$

The assumption that the mode in question must be in the linear regime, as well as subtending a small angle to the observer, is somewhat restrictive and can compromise the usefulness of equation (1.41) for the measurement of β. Heavens & Taylor (1995) (see also Fisher, Scharf & Lahav 1994) derive an alternative formulation of the problem utilizing a spherical harmonic decomposition of the density field. The latter formulation avoids the small angle restriction. The methods of Heavens & Taylor (1995) are used in the analysis of the IRAS PSCz redshift survey in Chapter 7. Due to the explicit dependence of the redshift-space power spectrum on β in equation (1.41), redshift-space distortions have been extensively used to measure the value of β. There have been several approaches to this problem.

One approach is to use equation (1.42) directly, together with some measure of the real-space power spectrum. Power spectra measured by inversion of angular clustering statistics (see Baugh & Efstathiou 1993, Baugh & Efstathiou 1994) are free from redshift-space distortions. This approach is used in Chapter 2 to measure β. I use the redshift-space power spectrum from the Stromlo-APM redshift survey and the real-space power spectrum, derived by Baugh and Efstathiou, from the APM galaxy survey.

Another way to determine β from redshift-space distortions is by the detection of anisotropies in the redshift-space power spectrum (or correlation function). In this case, one does not require a separate measurement of the power spectrum in real-space. This approach has been adopted by many authors, for example Hamilton (1993a) measures the correlation function for optical galaxies in the CfA redshift survey as a function of r and μ, $\xi(r,\mu)$. This function can then be expanded in terms of Legendre polynomials. The same procedure may be applied to the measurement of the power spectrum as a function of two variables, $P(k,\mu)$. Theoretically, the ratio of the quadrupole moment (Q) to the monopole moment (M) tends to a constant on large scales (i.e. scales in the linear regime), so in terms of the power spectrum,

$$\frac{P_Q(k)}{P_M(k)} = \frac{\frac{4}{3}\beta + \frac{4}{7}\beta^2}{1 + \frac{2}{3}\beta + \frac{1}{5}f\beta^2}. \tag{1.43}$$

Thus, by measuring $P(k,\mu)$ (or $\xi(r,\mu)$) on a large enough scale, an estimate of β may be obtained (see Loveday *et al.* 1995b, Fisher *et al.* 1994, Cole, Fisher & Weinberg 1994 and Cole, Fisher & Weinberg 1995 for other applications of this method).

The methods which measure the intrinsic anisotropies in the redshift-space power spectrum as well as those which compare a real-space clustering estimate with a redshift-space estimate are based on Kaiser's formula (equation 1.41) and are thus subject to the small angle approximation. As mentioned above, the method of Heavens & Taylor (1995) does not rely on this approximation but, in their particular implementation, does require a real-space power spectrum as an input to the calculations (e.g. a CDM-like linear theory model). An extension to this method is detailed in Ballinger, Heavens & Taylor (1995) and results in a maximum likelihood estimate, not only for β, but also for the real-space power spectrum, calculated in discrete bins in k-space.

Another approach to the measurement of β is that used by the POTENT analysis (Bertschinger & Dekel 1989, Dekel, Bertschinger & Faber 1990, Bertschinger *et al.* 1990). The POTENT analysis employs the linear relation (or a quasi-linear extension thereof)

$$\nabla \cdot \mathbf{v} = -\beta\delta(r), \tag{1.44}$$

where $\mathbf{v}$ is the peculiar velocity field. The density field, δ, is obtained from redshift surveys and the peculiar velocities of galaxies are measured directly using independent distance indicators. These fields are then compared to obtain an estimate of β. I briefly discuss the POTENT method again in Chapter 5.

We do not know the exact relation between the clustering of galaxies and the clustering of mass. This is a problem for the determination of the density parameter Ω. The methods described so far allow a determination of the parameter $\beta = \Omega^{0.6}/b$, but cannot independently determine the quantities Ω and b. At present there is no way to directly measure b, although higher order moments of the galaxy distribution can provide some further clues as to whether there is biasing between the galaxies and the mass and whether this biasing is a function of scale (see e.g. Szalay 1988, Gaztanaga & Frieman 1994, Frieman & Gaztanaga 1994). The higher order moments of the mass density field, $S_J = \bar{\xi}_J/\bar{\xi}_J^2$, (e.g. the skewness $J = 3$, and kurtosis - $J = 4$) can be predicted by non-linear perturbation theory. These predicted amplitudes are approximately independent of time, scale, density or geometry. Thus by measuring the corresponding moments of the galaxy density field, limits can be put on the difference between the two fields. At present, scale-dependent biasing schemes are disfavoured (see e.g. Gaztanaga & Frieman 1994).

Some measurements of Ω, obtained using measured galaxy peculiar velocity data, are independent of b since the galaxy velocity field responds to the underlying mass distribution. For example Dekel & Rees (1994) derive a limit $\Omega > 0.3$ (at 2.4σ) from consideration of the peculiar velocity fields around voids, the velocity field being recovered by a POTENT type analysis. The method of Dekel & Rees (1994) does involve making the assumption that the structure we see has grown via gravitational instability, but requires no assumptions to be made about the biasing between galaxies and mass. The procedure utilizes the fact that the fractional overdensity, δ, is overestimated if a particular quasi-linear relation (Nusser *et al.* 1991) is used to relate the measured derivatives of the velocity field, outflowing from the void, to the value of Ω. A value of Ω is assumed and this value is lowered until the recovered (and overestimated) value of δ drops below -1. This signals the onset of unphysical behaviour and thus provides a lower limit to the value of Ω. By making the assumption that the initial fluctuations were Gaussian distributed (as discussed in Section 1.4.1), it is also possible to use the peculiar velocity field observed today, trace it back in time, and recover the initial probability distribution function (PDF) of the density fluctuations. The recovered PDF, which depends on Ω is then compared with the assumed initial PDF (which does not) to obtain a value of Ω (see Nusser & Dekel 1993).

Thus far, I have concentrated on the linear peculiar velocity field and how it affects measures of large scale structure. On small scales, where linear theory is no longer valid, the effect of the peculiar velocity field on the power spectrum is different. The velocity field on these scales is incoherent. For example, the thermal velocities of galaxies in the central regions of a virialised cluster will suppress the amplitude of the power spectrum on small scales. This incoherent velocity field has been modelled by assuming a form for the velocity distribution function (e.g. a Gaussian or exponential form). The effect of the incoherent peculiar velocity field can then be included in the redshift-space distortion analysis. It is necessary, however, to be able to estimate the magnitude of small scale galaxy pairwise velocities. This quantity is not well constrained by observations (Bean *et al.* 1983, Mo, Jing & Börner 1993, Marzke *et al.* 1995). Redshift-space distortion analyses which incorporate the effect of small scale incoherent velocity fields include those of Peacock & Dodds (1994), Cole, Fisher & Weinberg (1995), and Heavens & Taylor (1995).

Notwithstanding the discussion above, the breakdown of the linear theory of redshift-space distortions does depend on more than just one effect. Cole, Fisher & Weinberg (1994) draw attention to the main problems. One of these is the small scale velocity dispersion as already mentioned. But of course, we also rely on the fact that the density and velocity perturbations are related by linear theory and the fact that the fractional overdensity is much less than unity, as well as the assumption that the gradient in the large scale peculiar velocity field is small. Cole, Fisher & Weinberg (1994) discuss ways in which we might be able to disentangle Ω and b via a detailed study of the onset of non-linearities in the theory of redshift-space distortions. This would require much larger data sets than are currently available, although such data sets will be forthcoming (the Sloan Digital Sky Survey and the AAT 2dF survey for instance). Fisher & Nusser (1996) suggest that the limiting factor in the analysis of redshift-space distortions in the linear regime is not, in fact, the small scale peculiar velocity field, but the gradient in the large scale peculiar velocity field. They use the Zeldovich approximation (Zeldovich 1970) to predict the ratio of the quadrupole to monopole moment of the redshift-space power spectrum and find that it declines as a function of scale. Denoting the ratio between the quadrupole and monopole moments by R, Fisher & Nusser (1996) find a universal relation between the ratio R/R_{lin}, where R_{lin} is the linear theory value of R, and k/k_{ln}, where k_{ln} is the wave number determined by a characteristic non-linear scale (it is in fact where the ratio R crosses zero). The idea is then to measure R (and k_{ln}) from the data, and use the universal relation to predict the value of R_{lin}, which is simply related to β (equation 1.43). This makes use of the data even in the non-linear regime.

1.5.3 Non-linear evolution

The description of the growth of $P(k)$ afforded by the transfer function is only valid in the linear regime. Thus the power spectrum on scales where the variance in the density field is greater than of order unity will not be accurately predicted by equation (1.38). The wave number at which the power spectrum measured from an evolved density field begins to differ from the linear theory prediction depends on the cosmological model. Analytic approximations for the power spectrum in the mildly non-linear regime have been studied (e.g. Hamilton *et al.* 1991, Peacock & Dodds 1994, Jain, Mo & White 1995, Peacock & Dodds 1996, Baugh & Gaztanaga 1996). In particular these formulae have been used to transform the measured power spectra of a variety of objects (cluster of galaxies, optical galaxies, IRAS galaxies and radio galaxies) to their corresponding initial, linear power spectra (Peacock & Dodds 1994). This was done under the assumption that structure grows from an initially random Gaussian field by gravitational instability, and using corrections for the effects of redshift-space distortions. Peacock & Dodds (1994) found excellent agreement in the shapes of the reconstructed linear theory power spectra of all the classes of object examined (for $\lambda \gtrsim 10\ h^{-1}$Mpc), and obtained the relative shift between the amplitudes of these spectra. The latter is then to be interpreted as the linear bias between the objects. Values obtained were: $b_A : b_R : b_O : b_I = 4.5 : 1.9 : 1.3 : 1$, where A signifies Abell clusters (of richness $R \geq 1$), R radio galaxies, O optical galaxies and I IRAS galaxies. The approach, of correcting the data to find a representation of the initial power spectrum differs from most earlier work. Generally, the theoretical power spectrum is manipulated to account for redshift-space distortions and non-linearities and this is then compared with the measured power spectrum.

Evolution of density perturbations in the non-linear regime, as well as the effects of redshift-space distortions, have been investigated by using numerical simulations. N-body simulations are used extensively in this thesis to test for systematic biases in the statistical analyses and for comparison of theoretical models with data. In the N-body simulations used in this thesis, the mass is represented by a large number of points which are displaced from their initial positions on a grid, according to the Zeldovich approximation. The Newtonian equations of motion are then solved in a discrete set of time steps to move the particles to their new positions.

There are several possible methods for calculating the force on each particle at any one time. Early simulations used the PP (particle-particle) method, where the force on every particle due to the effect of every other particle is calculated explicitly. This is very time consuming and cannot be used for simulations with a large number of particles. Much larger numbers of particles can be evolved using the PM (particle-mesh) method. In this case the particles are assigned to a grid and this density field is convolved in Fourier space with the Green's function of the Laplacian operator. This gives the Fourier transform of the gravitational potential. This potential is then used to calculate the force on each particle. The disadvantage of the PM method is that the gain in calculation speed is at the expense of spatial resolution. Efstathiou & Eastwood (1981) implemented a third improvement, where the forces on a particle from other particles within 2.7 cell radii were calculated by direct summation. This greatly increases the force resolution of the code over the PM method. The method of Efstathiou & Eastwood (1981) is known as P^3M (particle-particle-particle-mesh) and it has been employed, by George Efstathiou, to run the simulations used in this thesis. The particle-particle part of the P^3M calculation becomes slow if there are many particles within one mesh cell, i.e. at late times or in the centre of large density concentrations (e.g. clusters). Couchman (1991) has adapted the P^3M method (the adaptive particle-particle-particle-mesh AP^3M) to re-grid the density distribution on a finer grid for high density regions.

Using N-body simulations we can evolve a set of mass points to the present epoch from a given set of linear initial conditions. The power spectrum measured from the evolved mass distribution then has non-linearity and redshift-space distortions included, and we can compare directly with power spectra measured from observed data sets. The only remaining question is that of the bias. How do we relate the mass points (which represent the collisionless dark matter component only) to the galaxies that would have formed? Throughout this thesis I assume a linear bias. I use mass points in the simulations to represent the positions of galaxies and, if required, the final results may be multiplied by an arbitrary factor to change the bias. There are schemes to assign weights to the mass points which give the probability of finding a galaxy at a particular point (for example, see the scheme used by White *et al.* 1987b), however, these methods introduce new parameters which complicate the analysis.

Simulations which model dissipative effects as well as gravity are now being conducted. The most widely used method is that of SPH (Smooth Particle Hydrodynamics see Monaghan 1992 for a review). Development of these techniques, paralleled with an increase of our understanding of the important physical processes in galaxy formation, will hopefully lead to realistic simulations of the galaxy and mass distributions in any cosmological model of structure formation.

1.6 Summary of thesis contents

This thesis describes measurements of the redshift-space evolved power spectrum of the density field in the Universe as traced by various objects – optical galaxies, IRAS galaxies and rich APM clusters of galaxies. I also use the counts in cells method, which is closely related to the power spectrum, to examine the optical and IRAS galaxy density field. I make attempts to measure and use redshift-space distortions in the galaxy power spectra to determine the parameter β. Via inter-comparison of the clustering statistics of the various objects under consideration, I also draw some conclusions about the biasing of galaxies and clusters with respect to each other and the matter density field. The remainder of this introduction describes, in a little more detail, the analysis in each of the following chapters.

Chapter 2 describes an analysis of the Stromlo-APM optical galaxy redshift survey (Loveday *et al.* 1992a). I calculate the power spectrum of flux limited and volume limited subsets of the data. I also examine the dependence of the measured power spectrum on galaxy luminosity and morphology. The biases in the estimation of the power spectrum caused by uncertainty in the mean number density of objects in the survey is also discussed in detail. By comparing the redshift-space power spectrum of this survey with the real-space power spectrum of galaxies in the APM galaxy survey (Maddox *et al.* 1990a, Baugh & Efstathiou 1993), I estimate the value of the parameter β. I use simulated Stromlo-APM surveys drawn from large N-body simulations of low density cold dark matter (LCDM), mixed dark matter (MDM) and standard cold dark matter (SCDM) to estimate the random and systematic errors in this measurement of β. I also present a comparison of the Stromlo-APM power spectrum with that of other surveys and with the predictions of various cosmological models.

In Chapter 3 I review the status of measurements of large scale structure from catalogues of galaxy clusters. I present a power spectrum analysis of the APM rich cluster redshift survey, which is a complete survey of 364 rich clusters, free of many of the selection biases that affect many other cluster catalogues. I estimate the effect on the power spectrum of the convolution with the survey window function. I also estimate the effect of uncertainty in the mean number density of clusters. A comparison with the results of simulated APM cluster surveys drawn from N-body simulations of LCDM, MDM and SCDM universes is presented. Finally, in Chapter 3, I compare the power spectrum of rich clusters with that of galaxies in the Stromlo-APM survey and discuss the implications of this result for the biasing parameter of galaxies.

Chapter 4 presents an analysis of the power spectrum of IRAS galaxies. I use the $1.2Jy$ sample (Fisher *et al.* 1995a) and the QDOT sample (Lawrence *et al.* 1995). The power spectrum of the QDOT survey (Feldman, Kaiser & Peacock 1994, Chapter 4) is significantly different from that of the IRAS $1.2Jy$ survey. I examine whether this discrepancy depends on the way in which the galaxies are weighted in the analysis. The IRAS $1.2Jy$ and QDOT surveys are combined to yield a 'best' estimate of the power spectrum for IRAS galaxies (pre the PSCZ survey see Chapter 5).

In Chapter 5 I review the major scientific results that have resulted from surveys using the IRAS satellite data base, before describing the construction of the new IRAS PSCz redshift survey. This has the same flux limit as the QDOT survey but is fully sampled, making it the deepest and largest all-sky IRAS redshift survey to date. I present the power spectrum of this survey in Chapter 5. I examine the dependence of the power spectrum on Galactic latitude, b, to assess the effect of any possible variation in detector sensitivity as a function of b. I also compare the clustering properties of IRAS and optical galaxies.

In Chapter 6 I carry out a counts in cells analysis following the techniques of Efstathiou (1995). In this chapter, I divide the IRAS PSCz galaxy redshift survey into subsamples as a function of apparent flux and test for excess variance in the counts uncorrelated with galaxy distribution. This would result from a variation in the depth of the catalogue as a function of position on the sky (Hamilton 1995). Also in this chapter, I perform a joint counts in cells analysis of the galaxies in the PSCz survey with those in the Stromlo-APM galaxy survey and compare the results with those found in Chapter 5.

Chapter 7 describes some work in progress, in collaboration with Dr. A Heavens, Dr. A Taylor and Mr. B. Ballinger at the Royal Observatory in Edinburgh, and Professor G. Efstathiou at Oxford. In this chapter the redshift-space distortions in the PSCz redshift survey are examined using the spherical harmonic framework of Heavens & Taylor (1995). We test the analysis on simulated PSCz surveys drawn from large N-body simulations and apply the technique to the real survey.

I summarize the conclusions of this thesis in Chapter 8.

Chapter 2

Power spectrum and redshift-space distortion analysis of the Stromlo-APM galaxy redshift survey

2.1 Previous analysis

The Stromlo-APM redshift survey is a 1 in 20 sparsely sampled subset of 1787 galaxies selected from the APM galaxy survey (Maddox *et al.* 1990a) to a magnitude limit of $b_J = 17.15$. The survey has been used in several investigations of the large-scale clustering of galaxies: Loveday *et al.* (1992a) have investigated galaxy counts in nearly cubical cells with sizes in the range 5 h^{-1}Mpc – 60 h^{-1}Mpc; Loveday *et al.* (1995a) have computed the two-point galaxy correlation function of the Stromlo-APM survey and Loveday *et al.* (1995b) have investigated anisotropies in the two-point correlation function caused by redshift-space distortions (see Kaiser 1987). In this chapter I present an analysis of the power spectrum, $P(k)$, of galaxies in the Stromlo-APM survey. The power spectrum of the galaxy distribution has been estimated from a number of other surveys (see Efstathiou 1996 for a review). Power spectra for optically selected samples have been computed by *e.g.* Baumgart & Fry (1991), Vogeley *et al.* (1992), Park *et al.* (1994) and Baugh & Efstathiou (1993), Baugh & Efstathiou (1994). Power spectra for various IRAS redshift surveys have been computed by Fisher *et al.* (1993), Feldman, Kaiser & Peacock (1994) (hereafter FKP) and Tadros & Efstathiou (1995).

The Stromlo-APM redshift survey is described in Section 2.2. In Sections 2.3 and 2.4, I discuss estimators of the power spectrum and their biases and test the techniques against simulated Stromlo-APM redshift surveys constructed from large N-body simulations. In

Section 2.5 I present estimates of the power spectrum for volume limited and flux limited samples selected from the Stromlo-APM survey and investigate the sensitivity of the results to the volume limit and the weights assigned to the galaxies. A number of authors have claimed that luminous galaxies are more strongly clustered than less luminous galaxies (*e.g.* Hamilton 1988, Santiago & daCosta 1990, Iovino *et al.* 1993, Park *et al.* 1994, da Costa *et al.* 1994). I investigate the dependence of clustering strength with luminosity in Section 2.5.2. In Section 2.5.3 I analyse the clustering strength as a function of galaxy type. In Section 2.6 I investigate the distortion of the power spectrum, measured in redshift-space, caused by galaxy peculiar motions, and attempt to measure the value of β by comparing the redshift-space estimates of $P(k)$ from the Stromlo-APM survey with real-space estimates of $P(k)$ inferred from the parent APM catalogue (see Baugh & Efstathiou 1993). The main conclusions are presented in Section 2.7, where I also compare my measurement of $P(k)$ with those for the CfA-2 survey and IRAS surveys, and with the predictions of various cold dark matter models. I summarize in Section 2.8.

2.2 Data set

The Stromlo-APM redshift survey is described in detail by Loveday *et al.* (1992a). The survey covers an area approximately defined by the equatorial coordinates $21^h \lesssim \alpha \lesssim 5^h, -72.5° \lesssim \delta \lesssim -17.5°$. In the analysis described below, I analyse flux limited and volume limited samples. The volume limited samples are constructed by removing all galaxies with redshifts $z > z_{max}$ and absolute magnitudes $M > M_{crit}$, where

$$M_{crit} = m_{lim} - 25 - 5log(d_L(z_{max})) - k(z_{max}), \tag{2.1}$$

$m_{lim} = 17.15$ is the magnitude limit of the survey and $d_L(z_{max})$ is the luminosity distance at redshift z_{max}. I assume a spatially flat universe with $\Omega = 1$ throughout, thus

$$d_L(z) = \frac{2c}{H_0}\left[1 - \frac{1}{\sqrt{1+z}}\right](1+z). \tag{2.2}$$

The median redshift of the Stromlo-APM survey is $z = 0.05$, hence the results presented below are insensitive to the assumed cosmological model. I adopt a k correction of $3z$ in equation (2.1) which is appropriate for the median morphological type in the Stromlo-APM survey in the b_J passband (Efstathiou, Ellis & Peterson 1988).

There are some advantages in estimating power spectra from volume limited rather than flux limited samples. Firstly, every galaxy in a volume limited sample carries equal weight, whereas weighting factors that are a function of the power spectrum $P(k)$ are required to determine a minimum variance estimate of $P(k)$ from a flux limited sample if the underlying density field is assumed to be Gaussian (see FKP and Section 2.3). Secondly, the analysis of flux limited samples requires a model for the mean galaxy density $\overline{n}(r)$ as a function of radial distance. Thirdly, by estimating power spectra of progressively larger volume limited samples it is possible, given a large enough sample, to test whether the clustering amplitude is a function of luminosity. These points will be discussed in further detail in Sections 2.4 and 2.5.

In Figure 2.1 I show the number of galaxies in volume limited subsamples of the Stromlo-APM survey as a function of the limiting redshift z_{max} (solid line). The total number of galaxies peaks at $z_{max} = 0.06$, corresponding to an absolute magnitude limit of -19.34. For most of the analysis in this chapter I use a subsample of the Stromlo-APM survey with a volume limit of $z_{max} = 0.06$ (coordinate distance of $x_{max} = 172\ h^{-1}$Mpc) containing 469 galaxies. The dashed curve in Figure 2.1 shows the number of galaxies predicted by integrating over the galaxy luminosity function:

$$N\left(z_{max}\right) = \phi_\star V\left(z_{max}\right) \int_{L_{min}(z_{max})}^{\infty} \left(\frac{L}{L_\star}\right)^{\alpha} exp\left(-\frac{L}{L_\star}\right) d\left(\frac{L}{L_\star}\right) \tag{2.3}$$

where $V\left(z_{max}\right)$ is the comoving volume of the survey to redshift z_{max}, and L_{min} is the luminosity of a galaxy with absolute magnitude M_{crit} (equation 2.1). I have used the luminosity function parameters from Loveday *et al.* (1992b) $\phi_\star = 1.12 \times 10^{-2}\ h^3$Mpc^{-3}, $\alpha = -1.11$ and $M_\star = -19.73$; these are the best fitting Schechter function (Schechter 1976) parameters for galaxies with absolute magnitudes in the range $-22 < M < -15$, uncorrected for magnitude errors.

2.3 Measurement of the power spectrum

In this section I establish the notation and describe the estimator for $P(k)$. Imagine that the Universe is divided into infinitesimal cells of volume δV containing n_i galaxies such that the n_i are either 0 or 1. The mean and variance of n_i are therefore,

$$\langle n_i \rangle = \overline{n}\delta V \tag{2.4}$$

$$\left\langle n_i^2 \right\rangle = \langle n_i \rangle . \tag{2.5}$$

The observed galaxy count $n_o(\mathbf{x}_i)$ at point $\mathbf{x}_i$ is related to the real galaxy count n_i by

$$n_o(\mathbf{x}_i) = n_i W(\mathbf{x}_i), \tag{2.6}$$

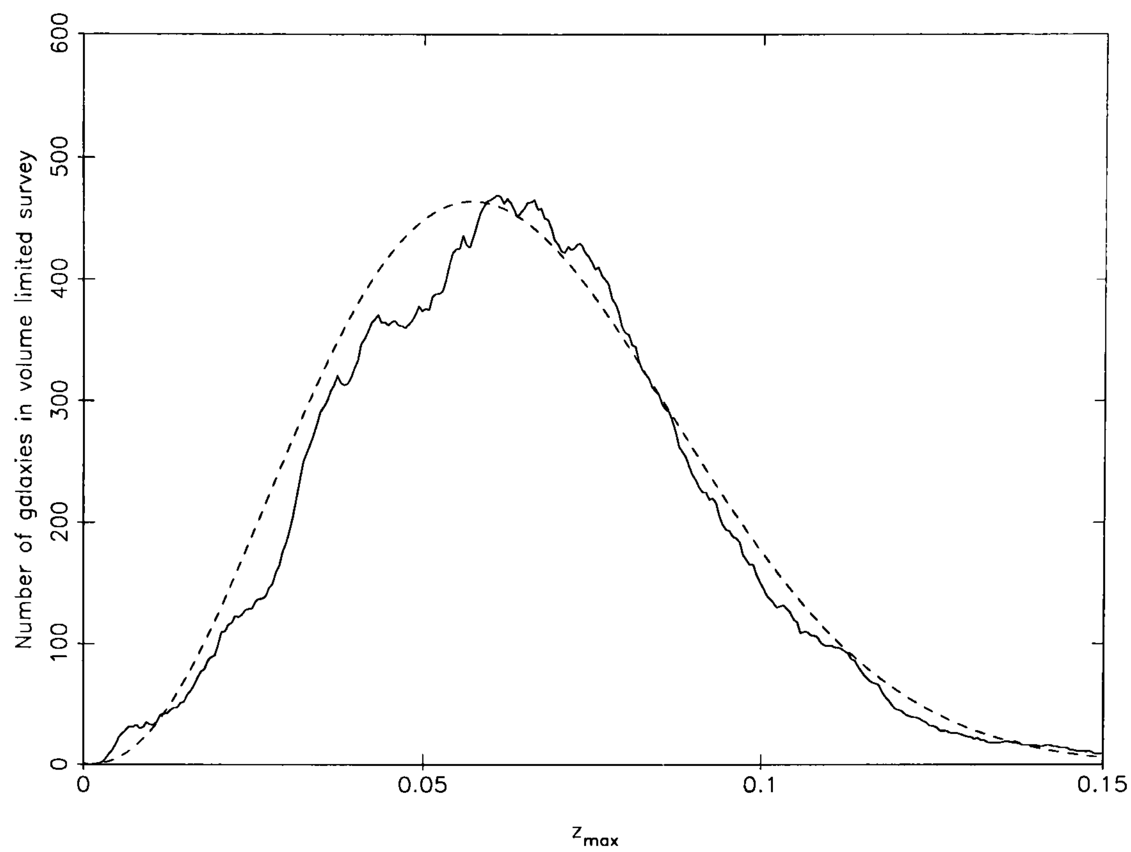


Figure 2.1: The solid line shows the number of galaxies in volume limited subsamples of the Stromlo-APM survey as a function of limiting redshift z_{max}. The dashed curve shows the expected distribution derived by integrating the luminosity function from Loveday *et al.* (1992) using the Schechter function parameters given in the text.

where $W(\mathbf{x}_i)$ defines the window function of the survey. In general, the window function depends on the radial distribution of galaxies in the survey, the survey boundaries on the celestial sphere, and the weights applied to each galaxy. In this section, it is assumed that we are analysing a volume limited sample in which each galaxy is assigned an equal weight. In this case, the window function $W(\mathbf{x})$ is equal to unity or zero according to whether a particular patch of the Universe is included or excluded from the survey. It is straightforward to generalize the equations in this section to more complex window functions and to include radially dependent galaxy weights (see FKP).

For the Stromlo-APM survey, the window function is defined by an angular mask consisting of the APM galaxy survey boundaries together with rectangular holes around regions on the plates where there are bright stars, satellite trails, globular clusters, step-wedges *etc.* (see Plate 2 of Maddox, Efstathiou & Sutherland 1990b). I estimate the Fourier transform of the window function

$$\hat{W}(\mathbf{k}) = \frac{1}{V}\int W(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}\,\mathrm{d}^3\mathbf{x} \tag{2.7}$$

by generating a large number of random points with mean density $\overline{n}_R$ within the survey volume and computing

$$\hat{W}_e(\mathbf{k}) = \frac{1}{\overline{n}_R V}\sum_i e^{i\mathbf{k}\cdot\mathbf{x}_i}, \tag{2.8}$$

where the sum extends over all random points located at positions $\mathbf{x}_i$. The volume V is a normalizing volume which encloses the survey.

I compute the Fourier transform of the observed density field

$$\hat{n}_o(\mathbf{k}) = \frac{1}{V}\sum_i n_i e^{i\mathbf{k}\cdot\mathbf{x}_i}, \tag{2.9}$$

and define a variable with zero mean

$$\Delta(\mathbf{k}) = \hat{n}_o(\mathbf{k}) - \overline{n}\hat{W}_e(\mathbf{k}), \tag{2.10}$$

where $\overline{n}$ is the mean galaxy density. From this we can derive the relation between the variance of $\Delta(\mathbf{k})$ and the power spectrum $P(k)$ of the galaxy distribution.

$$|\Delta(k)|^2 = \left(\hat{n}_o(k) - \overline{n}\hat{W}_e(k)\right)\left(\hat{n}_o^\star - \overline{n}\hat{W}_e^\star(k)\right)$$
$$= \hat{n}_o(k)\,\hat{n}_o^\star(k) - \overline{n}\hat{n}_o(k)\,\hat{W}_e^\star - \overline{n}\hat{n}_o^\star\hat{W}_e(k) + \overline{n}^2\left|\hat{W}_e(k)\right|^2. \tag{2.11}$$

From equation (2.9)

$$\hat{n}_o(k) = \frac{1}{V}\int n(x)\,\hat{W}(x)\,e^{i\mathbf{k}\cdot\mathbf{x}}d^3x, \tag{2.12}$$

and since $n(x) = \sum_i \delta(x - x_i)$ then

$$\hat{n}_o(k) = \frac{1}{V}\int\sum_i \hat{W}(x_i)\,e^{i\mathbf{k}\cdot\mathbf{x_i}}$$
$$= \frac{1}{V}\sum_i\sum_{k'}\hat{W}(k')\,e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x_i}}, \tag{2.13}$$

and so the first term in equation (2.11) is

$$\hat{n}_o(k)\,\hat{n}_o^\star(k) = \frac{1}{V^2}\sum_{ij}\sum_{k',k''}\hat{W}(k')\,\hat{W}^\star(k'')\,e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x_i}}e^{-i(\mathbf{k}-\mathbf{k}'')\cdot\mathbf{x_j}}. \tag{2.14}$$

Upon taking the ensemble average of equation (2.14),

$$\langle\hat{n}_o(k)\,\hat{n}_o^\star(k)\rangle = \frac{1}{V^2}\left\langle\sum_i\sum_{k',k''}\hat{W}(k')\,\hat{W}^\star(k'')\,e^{i(\mathbf{k}''-\mathbf{k}')\cdot\mathbf{x_i}}\right\rangle$$
$$+\frac{1}{V^2}\left\langle\sum_{i\neq j}\sum_{k',k''}\hat{W}(k')\,\hat{W}^\star(k'')\,e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x_i}}e^{-i(\mathbf{k}-\mathbf{k}'')\cdot\mathbf{x_j}}\right\rangle$$
$$= \frac{1}{V^2}\sum_i\sum_{k'}\left|\hat{W}(k')\right|^2 + \frac{1}{V^2}\sum_{k'}\left|\hat{W}(k')\right|^2\sum_{i\neq j}e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x_i}}e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x_j}}$$
$$= \frac{\overline{n}}{V}\sum_{k'}\left|\hat{W}(k')\right|^2 + \frac{\overline{n}^2}{V}\sum_{k'}\left|\hat{W}(k-k')\right|^2 P(k). \tag{2.15}$$

The second and third terms in equation (2.11) disappear upon ensemble averaging,

$$\left\langle \overline{n}\hat{n}_o(k)\,\hat{W}_e^\star(k)\right\rangle = \left\langle \overline{n}\hat{n}_o^\star(k)\,\hat{W}_e(k)\right\rangle = 0 \tag{2.16}$$

so,

$$\left\langle |\Delta(k)|^2\right\rangle = \frac{\overline{n}}{V}\sum_{k'}\left|\hat{W}(k')\right|^2 + \frac{\overline{n}^2}{V}\sum_{k'}\left|\hat{W}(k-k')\right|^2 P(k') + \overline{n}^2\left|\hat{W}_e(k)\right|^2. \tag{2.17}$$

If we assume that $\hat{W}(k-k')$ is small unless $k \simeq k'$, $P(k)$ can be removed from the sum in equation (2.17) and we find

$$\left\langle |\Delta(k)|^2\right\rangle \simeq \frac{\overline{n}}{V}\sum_{k'}\left|\hat{W}(k')\right|^2 + \frac{\overline{n}^2}{V}P(k)\sum_{k'}\left|\hat{W}(k')\right|^2 + \overline{n}^2\left|\hat{W}_e(k)\right|^2 \tag{2.18}$$

i.e.

$$P(k) \simeq \frac{\frac{V}{\overline{n}^2}\left\langle |\Delta(k)|^2\right\rangle - \frac{1}{\overline{n}}\sum_{k'}\left|\hat{W}(k')\right|^2 - V\left|\hat{W}_e(k)\right|^2}{\sum_{k'}\left|\hat{W}(k')\right|^2}. \tag{2.19}$$

This is a fairly good assumption at high wave numbers for most survey geometries (except for odd geometries, for example pencil beam surveys and the Las Campanas survey (Shechtman *et al.* 1995)). I have computed the sum in equation (2.8) by fast Fourier transforming the random density field within a cubical volume of side 840 h^{-1}Mpc, and Figure 2.2 shows the window function of the Stromlo-APM survey region volume limited at $z = 0.06$. This figure shows that $|W(k)|^2$ is sharply peaked in **k**-space, falling off approximately as k^{-4}. Also plotted are the components of $|W(\mathbf{k})|^2$ along the principal axes of the cube, showing that the window function is anisotropic and is narrower in the x-direction, which is aligned along the radial direction of the centre of the survey.

From equation (2.7), we see that

$$\begin{aligned}\sum_k\left|\hat{W}(k)\right|^2 &= \frac{1}{V^2}\sum_k\int d^3\mathbf{x}d^3\mathbf{x}'W(\mathbf{x})\,W^\star(\mathbf{x}')\,e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\\ &\frac{1}{V^2}\int d^3\mathbf{x}d^3\mathbf{x}'W(\mathbf{x})\,W(\mathbf{x}')\,\delta(\mathbf{x}-\mathbf{x}')\end{aligned} \tag{2.20}$$

i.e.,

$$\sum_k\left|\hat{W}(k)\right|^2 = \frac{1}{V}\int d^3\mathbf{x}W^2(\mathbf{x}) = \frac{V_s}{V}, \tag{2.21}$$

where V_s is the survey volume ($V_s = \int W(\mathbf{x})\,d^3\mathbf{x}$). Inserting equation (2.21) into equation (2.19) finally results in the equation:

$$P_e(\mathbf{k}) = \frac{\frac{V^2}{N_G^2}|\Delta(\mathbf{k})|^2 - \frac{1}{N_G} - \frac{1}{N_R}}{\frac{V}{V_s^2}\sum_{\mathbf{k}'}|W(\mathbf{k}')|^2} \tag{2.22}$$

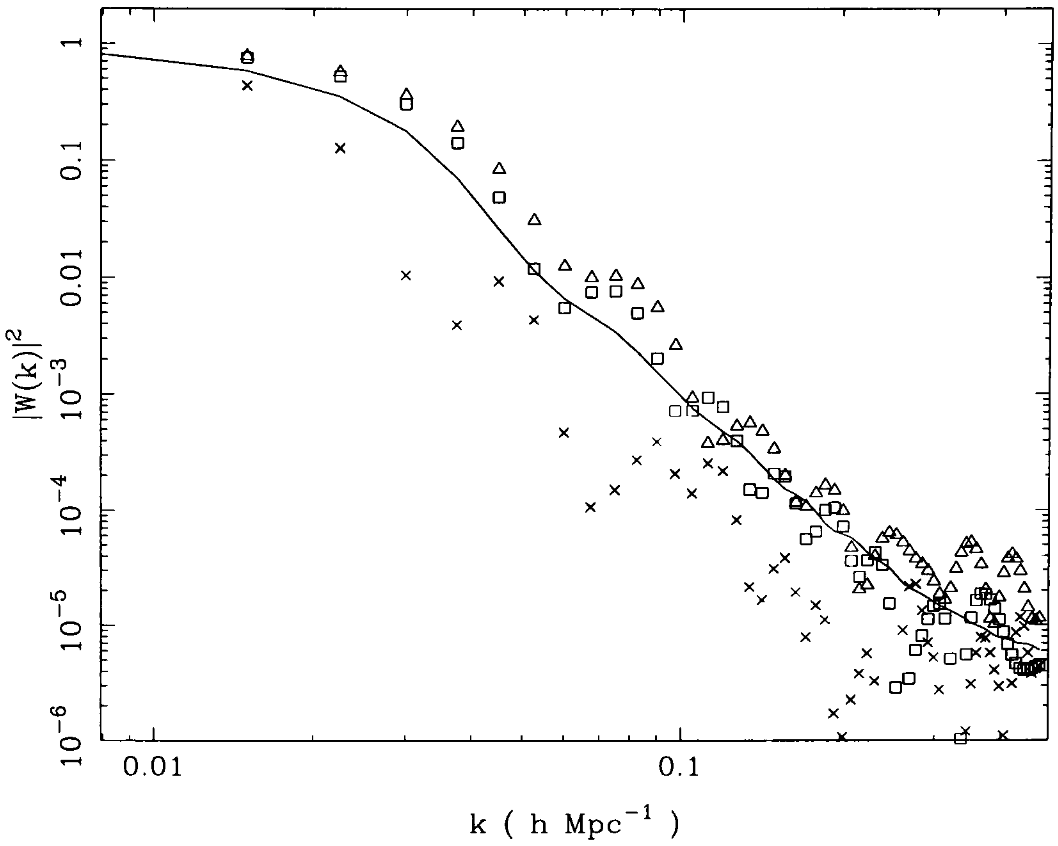


Figure 2.2: The Fourier transform of the Stromlo-APM survey volume limited at $z = 0.06$. The solid line shows the window function averaged in radial shells in k-space while the crosses, squares and triangles show $|W(k_x, 0, 0)|^2$, $|W(0, k_y, 0)|^2$, and $|W(0, 0, k_z)|^2$ components respectively. The x-axis is aligned approximately along the radial direction pointing to the centre of the survey on the celestial sphere. The curves are normalized so that they are equal to unity at $k = 0$.

for the final estimator of the power spectrum[1]. In equation (2.22), N_G and N_R are the total number of galaxy and random points within the survey volume V_s.

The negative terms in equation (2.22) correct for Poisson shot noise in the galaxy and random number distributions. Notice that in equation (2.22) it is assumed that the mean galaxy density is equal to N_G/V_s. The estimate $P_e(k)$ will therefore be biased low (at small wave numbers) because the estimate of the mean galaxy density is affected by galaxy clustering. We estimate the mean galaxy number density from

$$\overline{n} = \frac{1}{V_s} \sum_i n_i W(\mathbf{x}_i), \tag{2.23}$$

where the sum extends over all infinitesimal cells of galaxy count n_i. Inserting equation (2.23) into equation (2.18), we find that the expectation value of $|\Delta(k)|^2$ is given by,

$$\langle|\Delta(\mathbf{k})|^2\rangle = \frac{\overline{n}}{V} \sum_{\mathbf{k}'} |\hat{W}(\mathbf{k}')|^2 + \frac{\overline{n}^2}{V} \sum_{\mathbf{k}'} |\hat{W}(\mathbf{k} - \mathbf{k}')|^2 P(\mathbf{k}') - \frac{\overline{n}}{V_s} |\hat{W}(\mathbf{k}')|^2$$

[1] Note that the quantity $|\hat{W}(k)|^2$ relates to our estimate $|\hat{W}_e(k)|^2$ as follows,

$$\left|\hat{W}(k)\right|^2 = \left|\hat{W}_e(k)\right|^2 - \left(\frac{V_s}{V}\right)^2 \frac{1}{N_R}.$$

$$+\frac{\overline{n}^2}{V_s^2}\int\int \xi(x_{12})W(\mathbf{x}_1)W(\mathbf{x}_2)|\hat{W}(\mathbf{k}')|^2\, dV_1\, dV_2$$
$$-\frac{\overline{n}^2}{VV_s}\int\int \xi(x_{12})W(\mathbf{x}_1)W(\mathbf{x}_2)\hat{W}(\mathbf{k}')e^{i\mathbf{k}\cdot\mathbf{x}_1}\, dV_1\, dV_2$$
$$-\frac{\overline{n}^2}{VV_s}\int\int \xi(x_{12})W(\mathbf{x}_1)W(\mathbf{x}_2)\hat{W}(\mathbf{k}')^*e^{-i\mathbf{k}\cdot\mathbf{x}_2}\, dV_1\, dV_2. \qquad (2.24)$$

The last four terms in equation (2.24) account for the biases in estimating $P(k)$ when the mean galaxy density is determined from the sample itself. The first double integral in equation (2.24) measures the excess variance in the galaxy fluctuations above Poisson noise averaged over the survey volume,

$$\sigma_s^2 = \frac{1}{V_s^2}\int\int \xi(x_{12})W(\mathbf{x}_1)W(\mathbf{x}_2)\, dV_1\, dV_2. \qquad (2.25)$$

If this variance is dominated by fluctuations on scales smaller than the scale of the survey, it is a good approximation to set the other two integrals in (2.24) equal to equation (2.25). An approximate expression for the bias in $|\Delta(k)|^2$ is therefore,

$$-\frac{\overline{n}}{V_s}|\hat{W}(\mathbf{k}')|^2 - \overline{n}^2\sigma_s^2|\hat{W}(\mathbf{k}')|^2 \equiv -\frac{\overline{n}}{V_s}|\hat{W}(\mathbf{k}')|^2 - \overline{n}^2\left(\sum_{\mathbf{k}'}|W(\mathbf{k}')|^2P(\mathbf{k}')\right)|\hat{W}(\mathbf{k}')|^2 \quad (2.26)$$

where in equation (2.26) I have written σ_s^2 in terms of the power spectrum $P(k)$. I have used these equations to compute the dashed lines shown in Figure 2.3.

The first term in equation (2.26) introduces a bias in the estimate of $P(k)$ (equation 2.22) that depends on the number of galaxies in the survey, but is independent of the shape of the power spectrum; the second term introduces a bias that is independent of the number of galaxies in the sample, but depends on the shape of the power spectrum. The full derivation of these corrections is given in Appendix A (see also Peacock & Nicholson 1991).

The sampling errors on the power spectrum are estimated in two different ways. The first method is based on the error analysis of FKP, who assume that the galaxy density field is a Gaussian point process. In the case of a volume limited survey, equation (2.4.6) of FKP for the error on the point with wave vector k is, in my notation,

$$\sigma_P(k) = \frac{V_s}{N_G}\frac{(1+\frac{N_G}{N_R}+\overline{n}P(k))}{\sum_{\mathbf{k}'}\left|\hat{W}(\mathbf{k}')\right|^2}$$
$$\times\left(\frac{1}{N_{sum}}\sum_{\mathbf{k}'}\sum_{\mathbf{k}''}\left|\hat{W}(\mathbf{k}'-\mathbf{k}'')\right|^2\right)^{\frac{1}{2}}. \qquad (2.27)$$

In equation (2.27), $\mathbf{k}'$ and $\mathbf{k}''$ are constrained to lie in the same shell with modulus k, and N_{sum} is the total number of terms in the double summation.

The second method is based on the variance of $P(k)$ derived from mock Stromlo-APM surveys constructed from N-body simulations. The mock surveys, and a comparison with the FKP errors derived from equation (2.27) are described in Section 2.4.2. This is a useful test because equation (2.27) assumes Gaussian statistics, whereas the true galaxy density field is non-Gaussian at least on scales $\lesssim 5\ h^{-1}$Mpc where the *rms* fluctuations exceed unity. By comparing the errors from mock catalogues which are non-Gaussian on small scales, we can check the accuracy of equation (2.27) over the full range of wave numbers presented in this chapter.

In the above analysis, I have assumed that we are analysing volume limited samples and assigning an equal weight to each galaxy. The analysis of a flux limited sample is more complicated because we must assign radially dependent weights to the galaxies to minimise the variance on $P(k)$, and hence the window function $W(x)$ will depend on radial distance. These points are discussed in detail by FKP and I follow their analysis in this chapter. If the underlying density fluctuations are Gaussian, FKP show that the variance on the estimated $P(k)$ is minimised if each galaxy is assigned a weight

$$w(r) = \frac{1}{1+\overline{n}(r)P_w(k)}. \tag{2.28}$$

In equation (2.28) $\overline{n}(r)$ is the mean galaxy density as a function of radial distance r, which is computed from the luminosity function of the Stromlo-APM survey with the parameters given in Section 2.2, and $P_w(k)$ is the value of the power spectrum used in the weight function. From equation (2.28) we see that the minimum variance weighting for each wave number k, depends on the true value of the power spectrum at wave number k. Rather than applying different weights at each wave number, I have estimated the power spectrum from the flux-limited survey for four values of $P_w(k)$ in equation (2.28), $P_w(k) = 4000$, 8000, 16000 and 32000 $(h^{-1}\mathrm{Mpc})^3$, which span the range of interest at wave numbers $\lesssim 0.3h\mathrm{Mpc}^{-1}$. This analysis of flux limited surveys follows that of FKP except that I compute α (the ratio of the space densities in the real catalogue to that in the random catalogue, equation 2.1.3 of FKP) from the ratio of the sums

$$\sum_i \frac{1}{(1+4\pi\overline{n}(r_i)J_3)}\ , \tag{2.29}$$

instead of $\alpha = N_G/N_R$. The summations are over all galaxies and random points. I have set $4\pi J_3 = 10000(h^{-1}\mathrm{Mpc})^3$ (see Tadros & Efstathiou 1995, and Chapter 4).

2.4 Tests of the estimators using N-body simulations

In this section, I investigate the accuracy with which we can recover the power spectrum from the Stromlo-APM survey using the methods described in the previous section. There are three key aspects of the analysis that must be tested: (i) the assumption that the convolution of the window function with the power spectrum in equation (2.17) can be replaced by a product as in equation (2.18); (ii) the bias in the estimate of $P(k)$ caused by estimating the mean galaxy density from the survey itself (Appendix A); (iii) the accuracy of the FKP error estimates (equation 2.27). I investigate these points by analysing mock Stromlo-APM surveys constructed from N-body simulations.

2.4.1 Numerical simulations

The numerical simulations that are used here are described by Croft & Efstathiou (1994b). They consist of three ensembles of 10 simulations each containing 10^6 particles within a periodic computational box of length $\ell = 300\ h^{-1}$Mpc. The simulations were run with the particle-particle-particle-mesh (P^3M) code described by Efstathiou *et al.* (1985). The simulations model gravitational clustering in a cold dark matter (CDM) dominated universe with scale invariant initial density fluctuations. The three ensembles are as follows: the standard CDM model (Davis *et al.* 1985), *i.e.* a spatially flat universe with $\Omega_0 = 1$ and $h = 0.5$ (SCDM); a spatially flat low density CDM universe with $\Omega_0 = 0.2$ and a cosmological constant $\Omega_\lambda = \frac{\Lambda}{3H_0^2} = (1 - \Omega_0) = 0.8$ (LCDM); a spatially flat mixed dark matter model in which CDM contributes $\Omega_{CDM} = 0.6$, baryons contribute $\Omega_b = 0.1$ and massive neutrinos contribute $\Omega_\nu = 0.3$ (MDM).

For the SCDM and LCDM simulations, the initial power spectrum is given by

$$P(k) \propto \frac{k}{\left[1 + \left(ak + (bk)^{\frac{3}{2}} + (ck)^2\right)^\nu\right]^{\frac{2}{\nu}}} \tag{2.30}$$

where $\nu = 1.13$, $a = \frac{6.4}{\Gamma}\ h^{-1}$Mpc, $b = \frac{3.0}{\Gamma}\ h^{-1}$Mpc and $c = \frac{1.7}{\Gamma}\ h^{-1}$Mpc. Equation (2.30) is a good approximation to the linear power spectrum of scale-invariant CDM models with low baryon density, $\Omega_b \ll \Omega_0$ (Bond & Efstathiou 1984). The parameter Γ in equation (2.30) is equal to $\Omega_0 h$. Thus $\Gamma = 0.5$ for the SCDM ensemble and $\Gamma = 0.2$ for the LCDM ensemble. The initial conditions for the MDM simulations are generated from the power spectrum given by equation (1) of Klypin *et al.* (1993). In the MDM models, the thermal motions of the neutrinos are ignored and so follow the evolution of a collisionless cold component with

$\Omega_0 = 1$.

The final output times of the models are chosen to approximately match the microwave background anisotropies measured in the first year COBE maps (Smoot *et al.* 1992) ignoring any contribution from gravitational waves. Thus the *rms* mass fluctuations in spheres of radius 8 h^{-1}Mpc are $\sigma_8 = 1$ for the SCDM and LCDM ensembles and $\sigma_8 = 0.67$ for the MDM ensemble. The amplitude of the temperature anisotropies measured from the four year COBE maps (Wright *et al.* 1996) is only slightly higher than from the first year results and the small differences implied for the values of σ_8 are unimportant in the discussion below.

2.4.2 Mock Stromlo-APM surveys

From the simulations described above I have constructed mock Stromlo-APM surveys. I assume that galaxies are distributed like the mass points in the simulations and make no attempt to introduce biasing. The APM angular mask is applied to the simulations and the periodic box replicated where necessary to generate distant points. I then generate fully sampled volume limited catalogues to $z_{max} = 0.06$ that contain typically 100,000 mass points. I also generate volume limited samples containing on average 469 points by random sampling the mass points. These mock surveys thus each have a similar number of points to the number of galaxies in the Stromlo-APM survey at this volume limit. Flux limited surveys, containing on average 1787 galaxies, are generated by selecting mass points with the radial selection function of the Stromlo-APM survey (see Section 2.2) and then randomly sampling the distribution to obtain the required number of points.

Figure 2.3 shows three estimates of the power spectrum in real-space measured for each ensemble. The solid lines show the mean of the power spectra determined from the full cubical N-body simulations. These are computed by Fast Fourier transforming the particle distribution on a 128^3 grid using the nearest grid point assignment scheme, as described by Efstathiou *et al.* (1985). The open squares show the power spectra derived from equation (2.22) for the fully sampled mock Stromlo-APM surveys volume limited at $z_{max} = 0.06$. The symbols show the mean value of $P(k)$ from the ten members of each ensemble and the error bars show the standard deviation on the mean derived from the scatter in the $P(k)$ estimates. The open squares are in very good agreement with the power spectra of the full simulations at wave numbers $k \gtrsim 0.04h\mathrm{Mpc}^{-1}$, showing that replacing the convolution over $P(k)$ in equation (2.17) by a product is an excellent approximation.

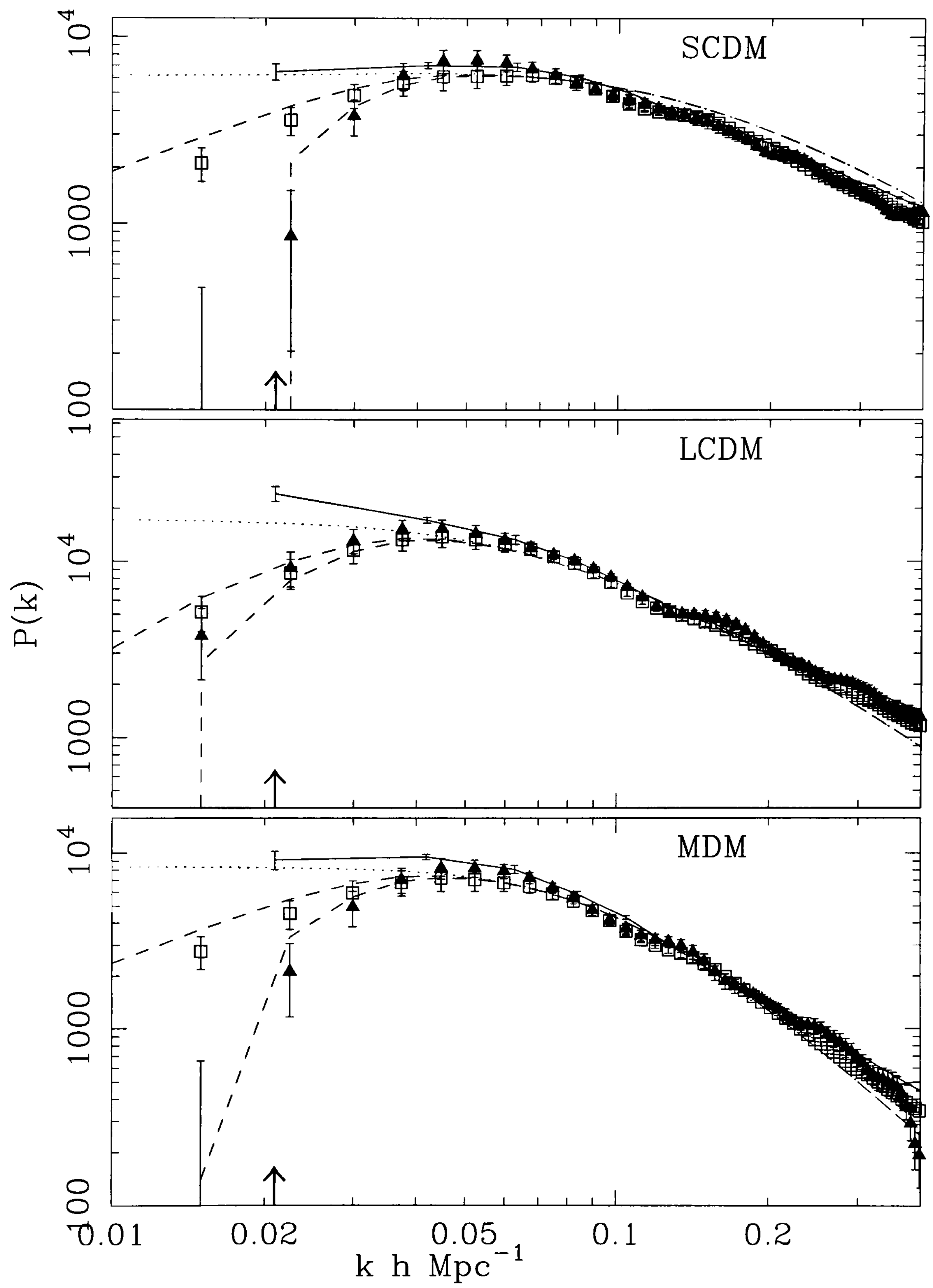


Figure 2.3: Estimates of the power spectra in real-space for the three ensembles of N-body simulations described in the text. The solid lines show the average of the power spectra derived from the full cubical volume of the simulations. The open squares show the average of $P(k)$ derived from equation (2.22) applied to 10 fully sampled realizations of the Stromlo-APM survey, volume limited at $z_{max} = 0.06$. The filled triangles show the average of $P(k)$ over 30 realizations of the sparse-sampled Stromlo-APM survey volume limited at $z_{max} = 0.06$. The error bars in each case show the standard deviation of the mean. The dashed lines are derived from equation (A12) of Appendix A, assuming the linear theory form of the power spectrum for each ensemble, and illustrate how the mock surveys are affected by the bias associated with the estimate of the mean galaxy density. The dotted lines show the convolution of the linear power spectrum (equation 2.30) with the window function for the Stromlo-APM survey. The arrows show the smallest wave number represented in the simulations, $k = 2\pi/\ell_b$, where $\ell_b = 300\ h^{-1}$Mpc is the computational box length.

At smaller wave numbers, the power spectrum is systematically underestimated. This is caused by the bias associated with the mean density estimate described in Section 2.3 and Appendix A. The dashed curves in Figure 2.3 show the linear theory power spectra for each model including an approximate correction for the bias derived from equation (A12) of Appendix A. These curves are in excellent agreement with the data points, showing that the bias can be removed if the form of the power spectrum is known. The filled triangles show the power spectra for volume limited ($z_{max} = 0.06$) realizations of the sparse sampled Stromlo-APM survey. As described above, these simulations have approximately the same number of objects as in the real catalogue and accurately model the window function of the real survey. Because the number of points in each realization is small (and hence the errors on $P(k)$ are large), I have plotted the averages over 30 realizations from each ensemble. The results are very similar to those from the fully sampled Stromlo-APM simulations, except that the bias associated with the mean density estimate is larger, as expected from equation (A12) of Appendix A (shown by the dashed lines passing through the points).

In Figure 2.4 I plot the results for flux limited sparse-sampled Stromlo-APM surveys for the SCDM and LCDM ensembles using the techniques of FKP. Each panel shows the power spectra averaged over 10 realizations for four values of $P_w(k)$ in the weighting scheme of equation (2.28), as indicated in each panel. The results are qualitatively similar to those of Figure 2.3 and show that the FKP estimator provides an unbiased estimate of $P(k)$ at wave numbers $k \gtrsim 0.04h\mathrm{Mpc}^{-1}$ for this survey. Furthermore, Figure 2.4 shows that the estimates of $P(k)$ and the errors on $P(k)$ are relatively insensitive to the weighting scheme. This is as expected since a minimum variance estimator should not be sensitive to small departures from the minimum variance weighting scheme. The errors on $P(k)$ at wave numbers $k \gtrsim 0.1h\mathrm{Mpc}^{-1}$ are noticeably larger if we adopt a constant value of $P_w(k) \geq 16000(h^{-1}\mathrm{Mpc})^3$ in the weighting scheme equation (2.28), but it is clear that this is an overestimate of the amplitude of the power spectrum at these wave numbers. Again, I also show here the effect of the bias in the power spectrum estimate discussed in Section 2.3. For a flux limited survey we estimate the mean galaxy density, using equation (2.29), from the survey itself. The resulting bias in the case of the flux limited survey is somewhat smaller than in the volume limited case, and the relevant expressions are derived in Appendix B. The solid lines in Figure 2.4 show the power spectra from the full N-body boxes i.e. no window functions, the dashed line shows the effect of the convolution, while the dot-dash line shows the predicted power spectra after taking account of both the convolution and

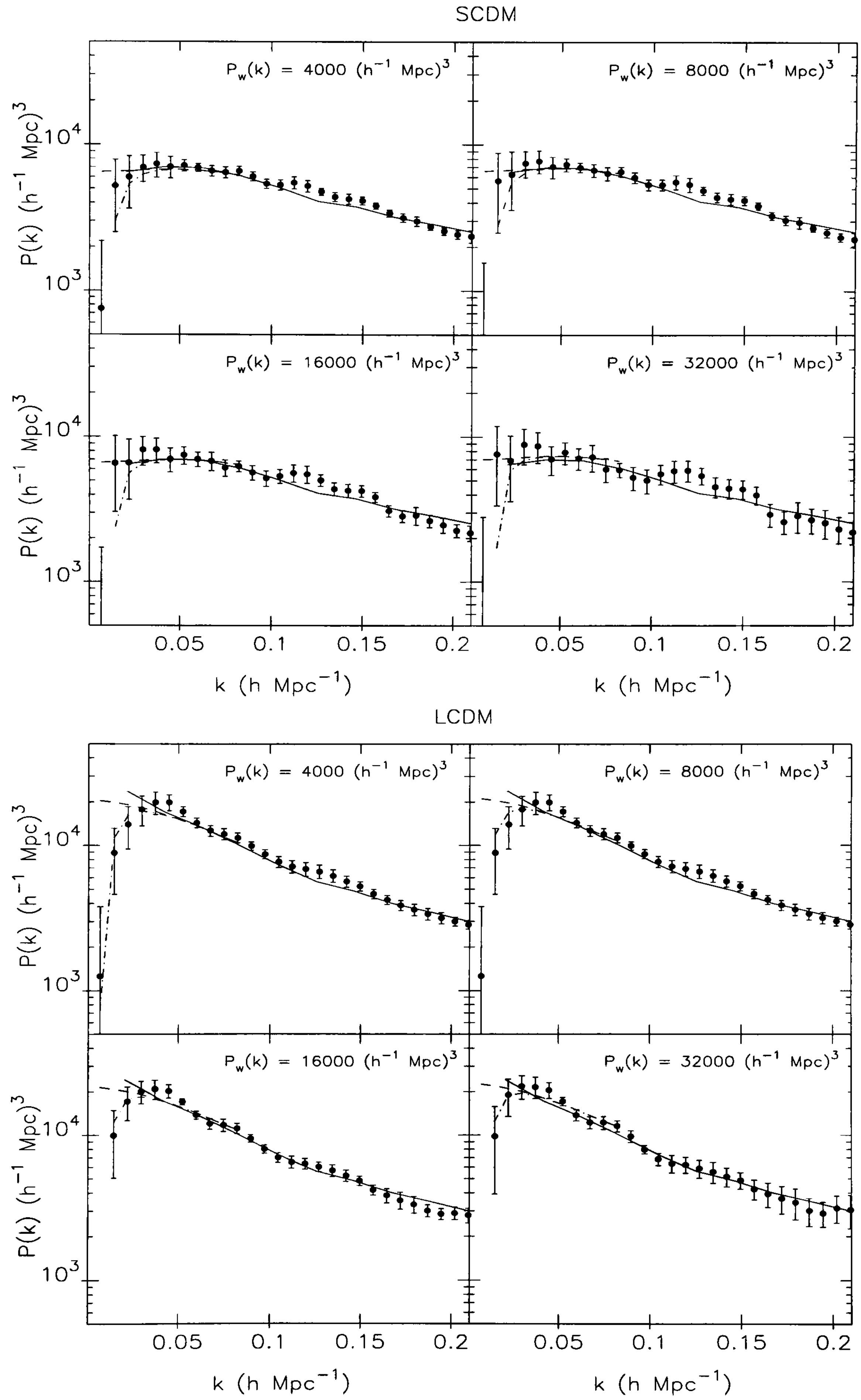


Figure 2.4: Flux limited power spectra determined from mock sparsely sampled Stromlo-APM surveys. The upper figure shows results for the SCDM ensemble and the lower figure shows results for the LCDM ensemble. The solid line in each panel shows the mean power spectrum determined from the full N-body simulations, as plotted in Figure 2.3. The points show the average over ten mock Stromlo-APM surveys for four values of $P_w(k)$ in the weighting scheme of equation (2.28), $P_w(k)$ = 4000, 8000, 16000, 32000$(h^{-1}\mathrm{Mpc})^3$, as indicated in each panel. The error bars show one standard deviation on the mean. The dashed and dot-dashed lines show the effects of convolution with the window function, and uncertainty in the mean number density of the survey as explained in the text.

the bias due to the estimate of the mean galaxy density from the data.

I have also investigated the accuracy of the error estimates derived from equation (2.27) and its generalization to flux limited samples by comparing with the dispersion in the $P(k)$ estimates from the mock surveys. Equation (2.27) provides an accurate estimate of the errors for wave numbers $k \lesssim 0.1h\mathrm{Mpc}^{-1}$, but tends to underestimate the errors by a factor of ~ 2 at larger wave numbers.

In summary, the results of this section show that these methods provide nearly unbiased estimates of the power spectrum at wave numbers $k \gtrsim 0.04h\mathrm{Mpc}^{-1}$. At smaller wave numbers, the estimates are biased low because we determine the mean galaxy density from the sample itself. The power spectra derived from flux limited samples are insensitive to the weighting scheme, provided we adopt a reasonable value of $P_w(k)$ in equation (2.28).

2.5 The power spectrum of the Stromlo-APM survey

2.5.1 Comparison of results from volume limited and flux limited samples

In Figure 2.5, I show the power spectra for three volume limited subsets of the Stromlo-APM survey, together with 1σ errors derived from the FKP formula (equation 2.27). The power spectra are consistent with each other, with no obvious dependence on the volume limit. The results are qualitatively similar to those of the simulations plotted in Figure 2.3. The power spectrum of the real survey has an approximately power-law behaviour $P(k) \propto k^{-1.8}$ at wave numbers $k \gtrsim 0.05h\mathrm{Mpc}^{-1}$, and flattens off and declines at smaller wave numbers. The decline is almost certainly caused by the bias discussed in Appendix A and so is not a real feature of the galaxy distribution. The Stromlo-APM survey contains little information at wave numbers $k \lesssim 0.05h\mathrm{Mpc}^{-1}$, since these scales are comparable to or greater than the size of the survey.

Figure 2.6 shows the flux limited power spectrum from the Stromlo-APM survey for four values of $P_w(k)$ in the weighting function of equation (2.28). As in the analysis of the flux limited mock Stromlo-APM simulations (Figure 2.4) we see that the power spectra in Figure 2.6 are insensitive to the weighting scheme. The power spectra in Figures 2.5 and 2.6 peak at a maximum value of $\sim 30000(h^{-1}\mathrm{Mpc})^3$ at a wave number of $\sim 0.05h\mathrm{Mpc}^{-1}$ and decline to a value of $\sim 4000(h^{-1}\mathrm{Mpc})^3$ at wave numbers $\sim 0.15h\mathrm{Mpc}^{-1}$. Thus, over most of the wave number range plotted in Figure 2.6, a value of $P_w(k) \sim 10^4(h^{-1}\mathrm{Mpc})^3$ in equation (2.28) should result in an estimate of $P(k)$ that is close to the minimum variance estimate. In the remainder of this chapter, unless otherwise stated, I will use the flux

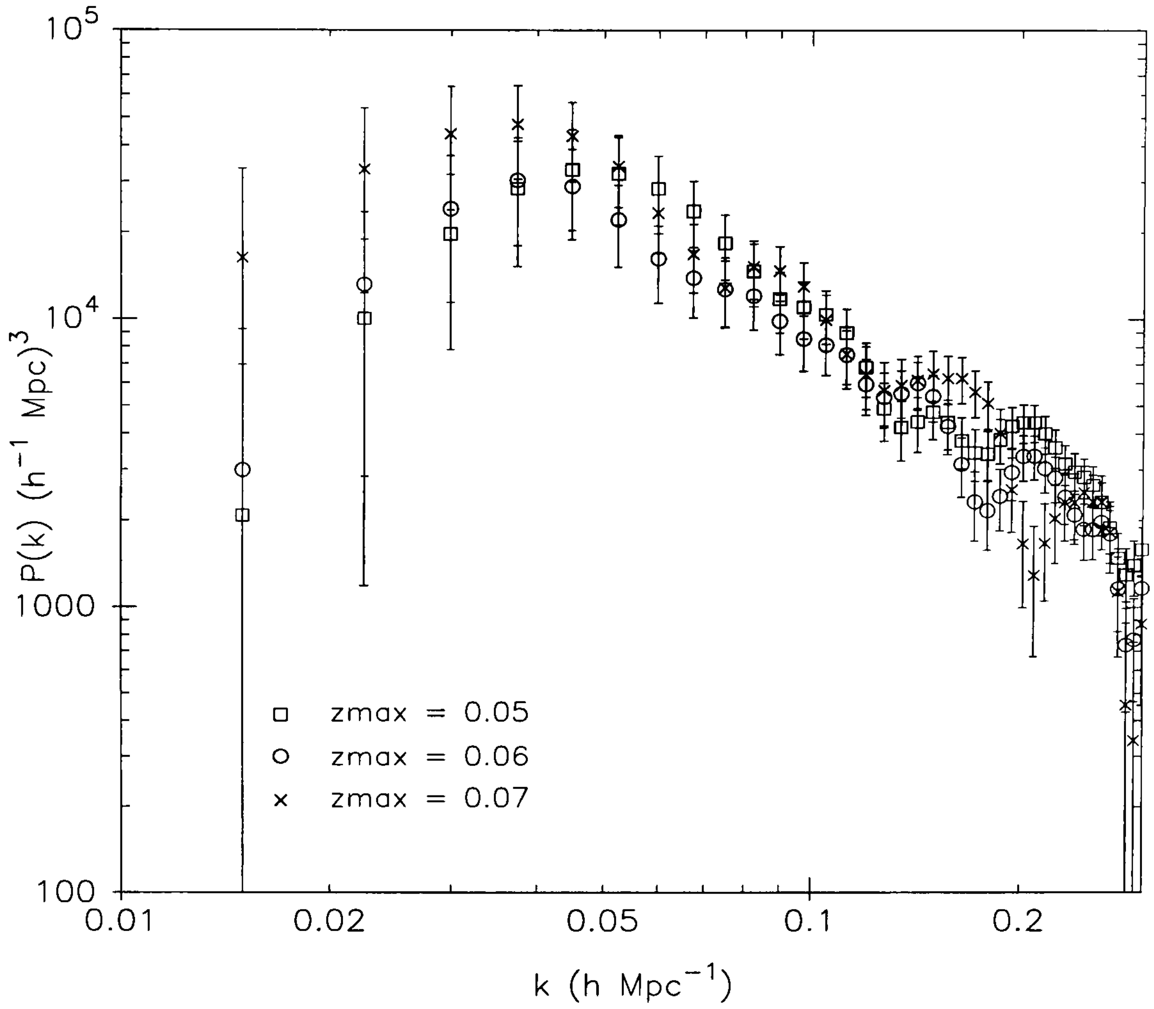


Figure 2.5: Volume limited power spectra of the Stromlo-APM survey, volume limited at $z = 0.05, 0.06, 0.07$ (coordinate distances of 145 h^{-1}Mpc, 172 h^{-1}Mpc, and 200 h^{-1}Mpc).

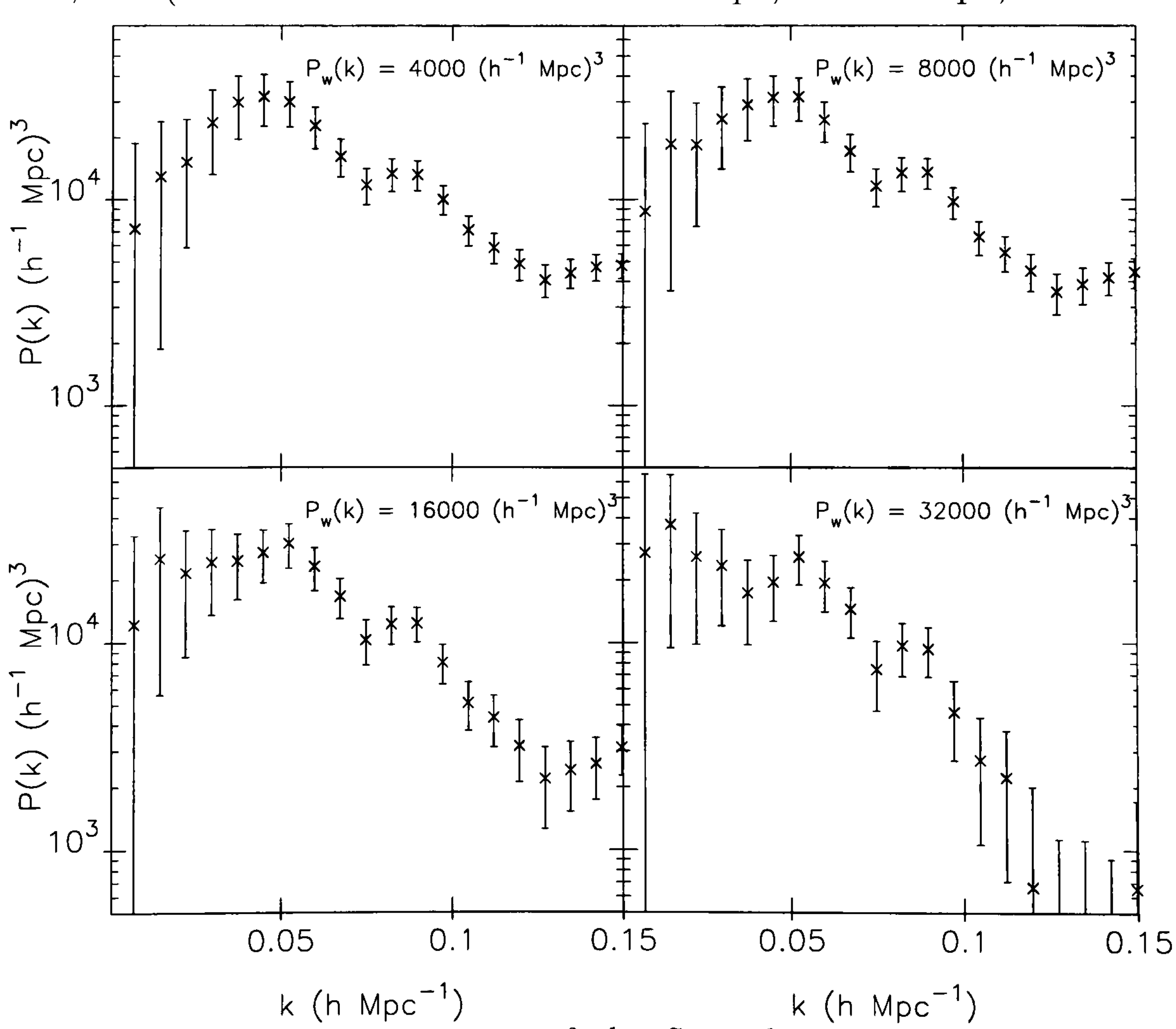


Figure 2.6: Flux limited power spectra of the Stromlo-APM survey for four values of $P_w(k)$ in the weighting scheme of equation (2.28), $P_w(k) = 4000$, 8000, 16000, and $32000(h^{-1}\text{Mpc})^3$ as indicated in each panel.

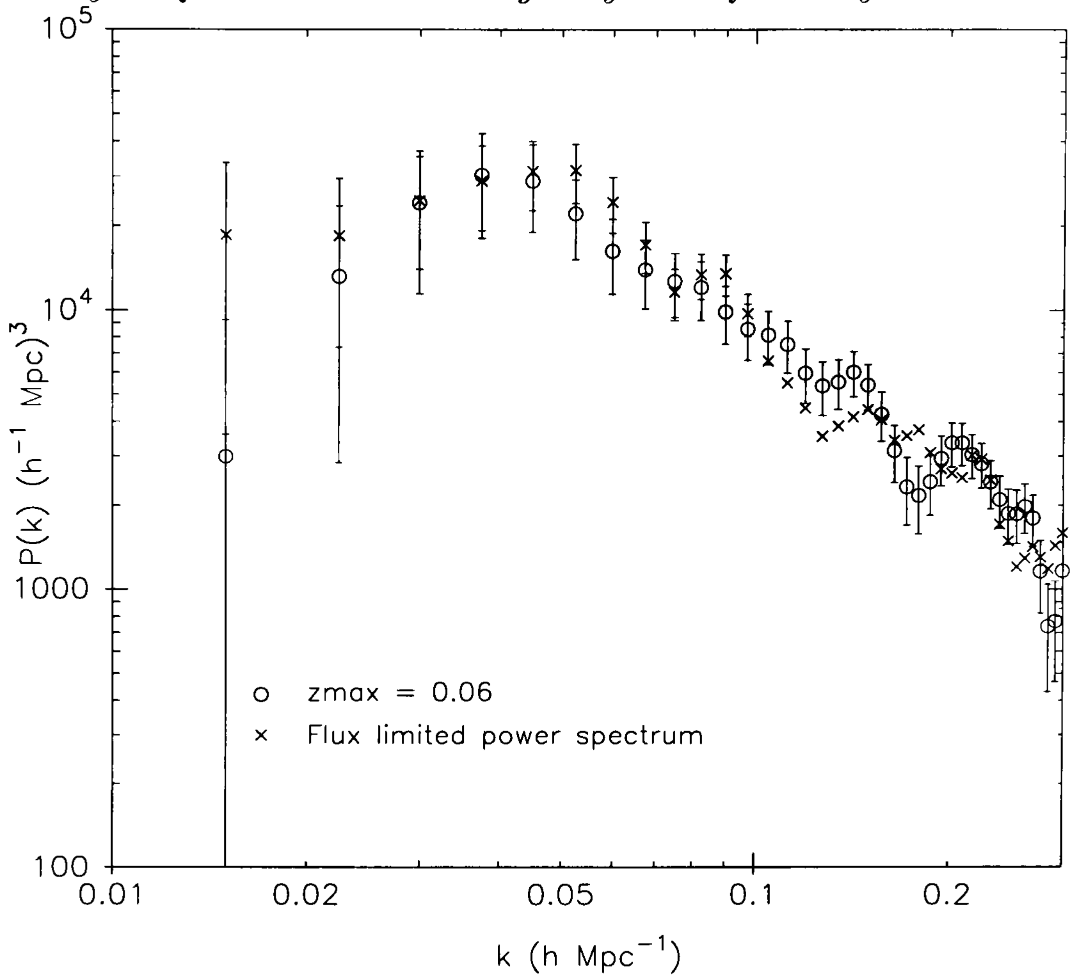


Figure 2.7: Comparison of the volume and flux limited power spectra estimated from the Stromlo-APM survey. The circles show the volume limited power spectrum for the $z_{max} = 0.06$ sample as plotted in Figure 2.5, and the crosses show the flux limited power spectrum with $P_w(k) = 8000(h^{-1}\mathrm{Mpc})^3$ in equation (2.28), as plotted in Figure 2.6. I have plotted 1σ error bars, but suppressed the errors on the flux limited estimates at high wave numbers for clarity.

limited power spectrum with $P_w(k)$ set to $8000(h^{-1}\mathrm{Mpc})^3$ in equation (2.28).

In Figure 2.7, I compare the volume limited power spectrum estimate for $z_{max} = 0.06$ with the flux limited estimate derived with $P_w(k) = 8000(h^{-1}\mathrm{Mpc})^3$ in equation (2.28). The estimates are consistent with each other to within the 1σ error bars. This figure, and the results shown in Figures 2.4 to 2.6, show that the power spectrum estimates from the Stromlo-APM survey are robust. In summary, no obvious differences are found between the power spectra of the volume limited samples plotted in Figure 2.5, or between the power spectra measured from flux limited samples with different weighting functions. This places some constraints on variations of the power spectrum as a function of galaxy luminosity which I discuss in further detail in the next subsection.

2.5.2 Variations of the clustering strength with galaxy luminosity

As mentioned in Section 2.1, several groups have claimed to find various correlations between the strength of the clustering pattern and galaxy luminosity. In this section I investigate this possibility by comparing the power spectrum measured from volume limited subsamples of the Stromlo-APM data as a function of volume limit. We have already seen from Figure 2.5 that there is little evidence for any systematic dependence on the amplitude of $P(k)$ with volume limit over the range $z_{max} = 0.05$–0.07, corresponding to cuts in absolute magnitude of $M_{crit} = -18.91$ at $z_{max} = 0.05$ and $M_{crit} = -19.71$ at $z_{max} = 0.07$ (equation 2.1). As we increase the volume limit beyond $z_{max} = 0.07$, the number of galaxies in the Stromlo-APM survey declines exponentially and hence the errors on $P(k)$ become large. I have therefore averaged the power spectrum estimates over a range of wave numbers to test for a linear bias in $P(k)$ (*i.e.* a change in amplitude independent of wave number) as a function of volume limit. This has been done as follows: I have computed the power spectra of six volume limited samples with $z_{max} = 0.04, 0.05, 0.06, 0.07, 0.08, 0.09$. These power spectra were averaged to produce a mean power spectrum $P_{av}(k)$. For each volume limit. I compute the ratio $P(k)/P_{av}(k)$, which I then averaged over the wave number range $0.052 < k < 0.14h\mathrm{Mpc}^{-1}$ to produce an estimate of the relative bias factor $b^2(M_{crit})$ for each volume limit. As can be seen from Figure 2.6, the error bars on $P(k)$ depend only weakly on wave number, hence it is reasonable to compute $b^2(M_{crit})$ as an unweighted average over wave number. These bias factors are plotted against M_{crit} in Figure 2.8. If there were no dependence of the clustering strength with luminosity, we would expect the resulting bias factors to be close to unity for all values of M_{crit}. In fact, the relative bias factor is close to unity for most of the magnitude range, but is lower than unity for the faintest magnitude cut at $M_{crit} = -18.4$ and higher than unity for the brightest magnitude cut at $M_{crit} = -20.3$.

To assess the statistical significance of these results, I have computed linear bias factors in exactly the same way for each of the 30 mock Stromlo-APM surveys constructed from the LCDM ensemble. The results are plotted as the crosses in Figure 2.8, together with the standard deviation for a single Stromlo-APM survey. The results from the mock surveys are close to unity over the entire magnitude range, as expected, since there is no luminosity dependence of the clustering pattern in the simulations. These results show that there is no firm evidence for any luminosity dependence of the amplitude of the power spectrum determined from the Stromlo-APM survey. At the faintest magnitude cut, 7 out of 30 simulated catalogues gave a relative bias factor lower than that measured for the real data. For the brightest magnitude cut, only 2 out of 30 simulations gave a relative bias factor larger than that determined from the Stromlo-APM survey. Thus, there is tentative evidence that the amplitude of the power spectrum may increase at high luminosities. These

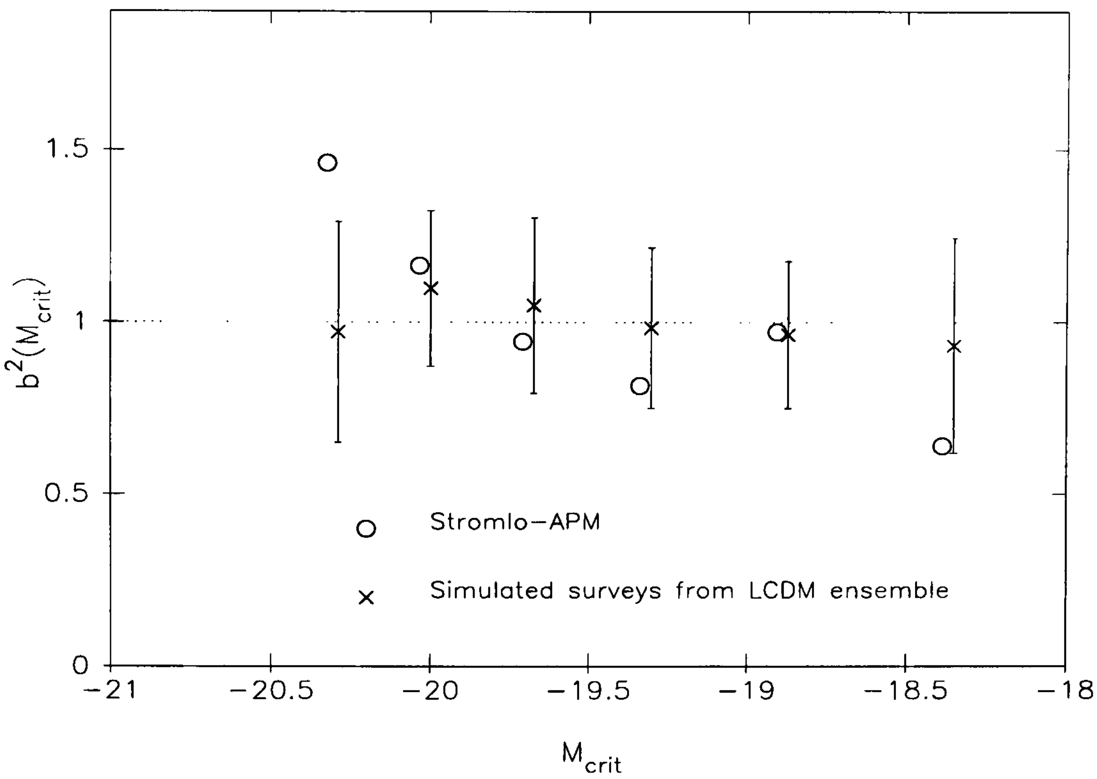


Figure 2.8: The relative bias factor b^2 defined in the text plotted against limiting absolute magnitude M_{crit}. Circles show the result for the Stromlo-APM data, crosses show the results for simulated Stromlo-APM surveys drawn from the LCDM ensemble. The error bars on these points show the standard deviation of one simulation.

results seem consistent with previous work. Loveday *et al.* (1995a) found some evidence that the spatial two-point correlation function for galaxies in the Stromlo-APM survey in the magnitude range $-19 < M < -15$ has a lower amplitude than that measured for brighter galaxies. The results of Figure 2.8 suggest that this may be caused by sampling fluctuations rather than by a real luminosity dependence of the clustering pattern. Park *et al.* (1994) have analysed the power spectra of volume limited subsets of the CfA-2 redshift survey and find evidence that the amplitude of the power spectrum for galaxies brighter than M^* is about 40% higher than the amplitude measured for fainter galaxies. This is consistent with the results plotted in Figure 2.8 and suggests that the rise in the relative bias factor at $M_{crit} < -20.3$ may be a real feature of the galaxy distribution. Evidently, larger redshift surveys are required to firmly establish whether this effect is real but the results of this section show that over most of the luminosity range, the power spectra measured from the Stromlo-APM survey are consistent with the null hypothesis that the clustering strength is independent of luminosity.

2.5.3 Variations of the clustering strength with galaxy morphology

The variation in clustering strength with morphological type may give us information about the way in which galaxies form. If galaxies of different morphological types cluster in different ways then not all of them can be unbiased tracers of the underlying mass in the

Universe. It is already known that on scales $\lesssim 5\ h^{-1}$Mpc the correlation function of early type galaxies has a steeper slope and higher amplitude than that of late type galaxies (e.g. Davis & Geller 1976, Giovanelli, Haynes & Chincarini 1986, Iovino *et al.* 1993, Loveday *et al.* 1995b) but what happens on larger scales is unclear.

The galaxies in the Stromlo-APM redshift survey have all been assigned morphological types by visual inspection of photographic plates (Loveday 1996), so it is possible to find the power spectrum of galaxies according to morphological type. I have divided the sample into galaxies of class 1 and 2 (elliptical and S0 galaxies) and those of class 3 and 4 (spirals and irregular galaxies), as classified by Jon Loveday. The power spectrum analysis of these samples is carried out exactly as described above, except that the selection function is now derived in a different way. Assigning a morphological type to galaxies by visual inspection becomes difficult at faint magnitudes. Thus the morphological classifications of galaxies in the Stromlo-APM survey are inaccurate at faint magnitudes. This precludes a reliable derivation of the type-specific luminosity function. To derive the selection function of the two subsamples of galaxies in the survey then, I fit the observed $N(r)$, via a maximum likelihood method, with a function applied to the QDOT survey by Feldman, Kaiser & Peacock (1994):

$$N(r) = A\left(\frac{r}{r_{max}}\right)^{x}\left[1+\left(\frac{r}{r_{max}}\right)^{y}\right]^{-\frac{y+x}{y}} \tag{2.31}$$

where $A = 2^{\frac{y+x}{y}} n_{max}$. I have used this formula as an empirical match to the observed histogram of $N(r)$ and fit the parameters x , y , r_{max} and n_{max} by maximising the likelihood function

$$\mathcal{L} = -2ln(p) \tag{2.32}$$

where p is the probability of the observed distribution of galaxy radial distances,

$$ln(p) = \sum_i ln\left[N\left(r_i\right)\right] - \int N(r)dr, \tag{2.33}$$

the sum being over all galaxies in the subsample under consideration. The derived parameters for early type, late type and all galaxies are shown in Table 2.1, and Figure 2.9 shows these functions plotted with the observed $N(r)$ histograms. All the analysis in this section is performed on galaxies with radial distances between 20 and 400 h^{-1}Mpc.

Figure 2.10 shows the flux limited power spectrum of all the galaxies in the Stromlo-APM survey using a weighting of $P_w(k) = 8000(h^{-1}\text{Mpc})^3$ both using the selection function

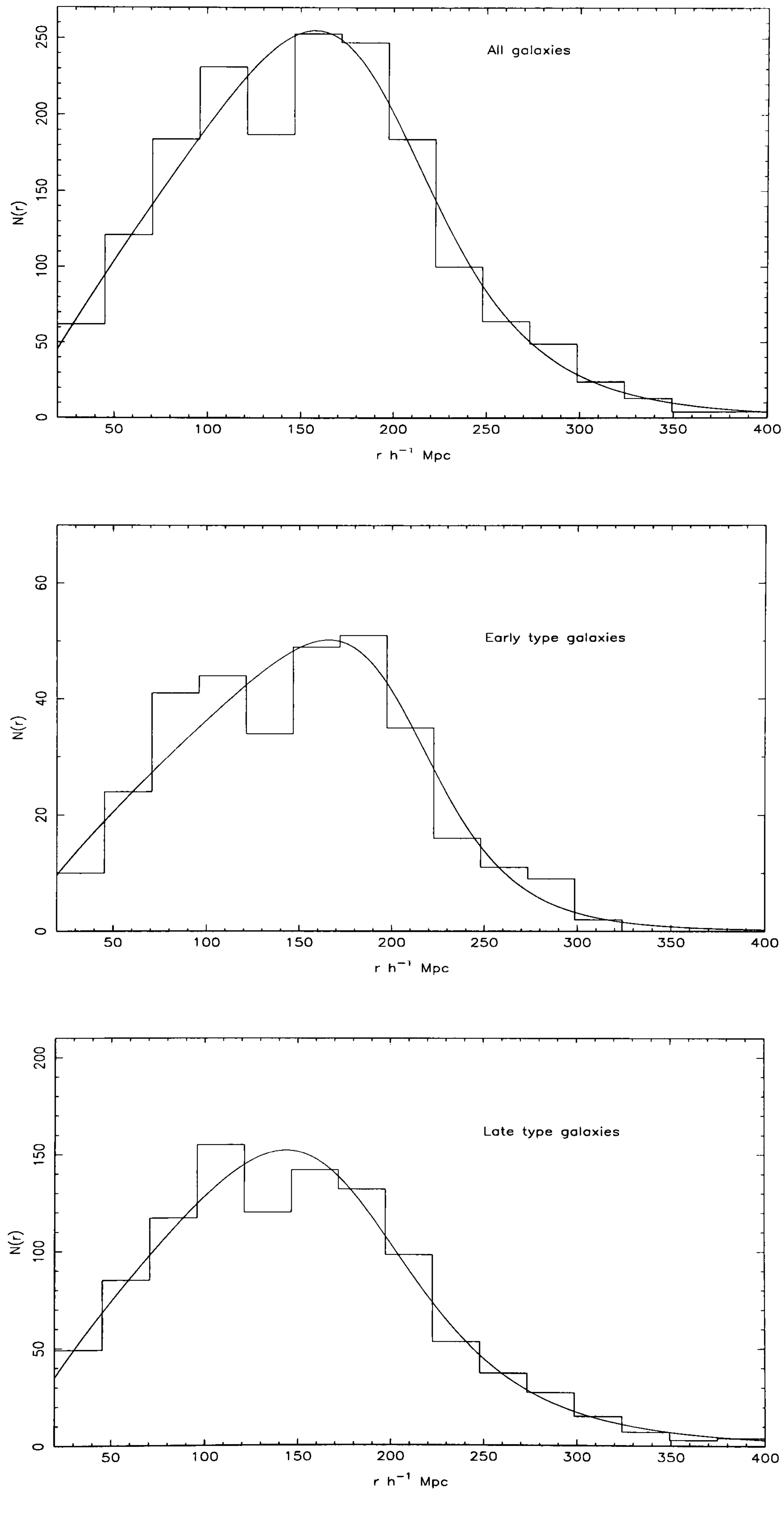


Figure 2.9: Histograms of the distribution of radial distances of galaxies in the Stromlo-APM survey. The top panel shows the histogram for all galaxies, the middle panel is early type galaxies only and the bottom panel is late type galaxies only. The solid curve shows the fitted $N(r)$ using equation (2.31) with parameters listed in Table 2.1.

	n_{max}	r_{max}	x	y
Early type	1.27	215.73	0.82	9.19
Late type	4.12	199.73	0.82	6.17
All	6.84	211.12	0.90	7.16

Table 2.1: Parameters for the fit to $N(r)$ using equation (2.31) for early type, late type and all galaxies in the Stromlo-APM survey.

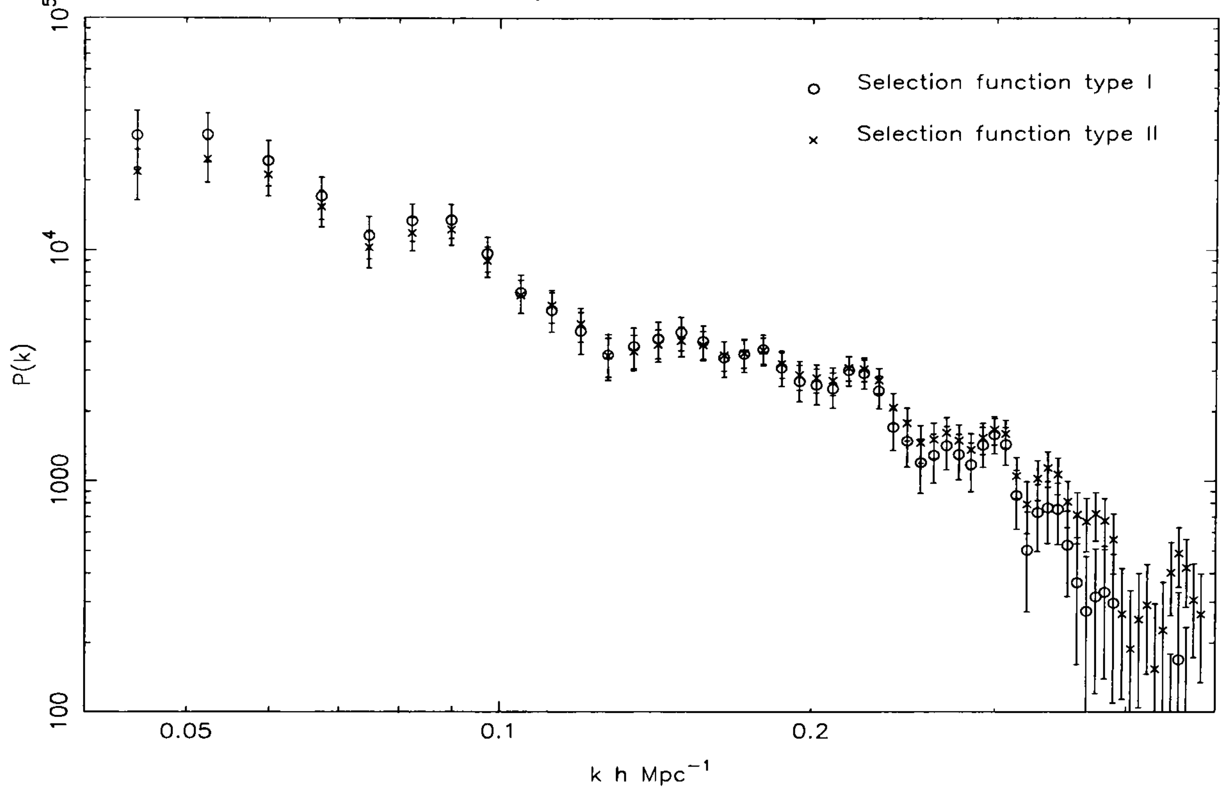


Figure 2.10: Figure shows measurements of the power spectrum of all galaxies in the Stromlo-APM redshift survey. The circles show results for a selection function derived from the Schechter luminosity function with parameters given in Section 2.2. The crosses show results using a selection function derived from equation (2.31) with the parameters listed in Table 2.1.

derived from the Schechter luminosity function (as in Section 2.5.1), and using the selection function derived from the fit to equation (2.31) with the parameters listed in Table 2.1. It can be seen that the power spectrum measured using the selection function derived from equation (2.31) is compatible with that measured previously to within the calculated error bars.

Using these selection functions I have measured the power spectra of early and late type galaxies separately. The 362 untyped galaxies were not included in the analysis. (Note, the conclusions of this section are unchanged if we include these galaxies and randomly assign them a class such that the numbers of early and late type galaxies are in the ratio 1 : 4). This leaves 326 early type galaxies and 1044 late type galaxies for the power spectrum analysis. Figure 2.11 shows a comparison of the power spectra measured from the two subsamples. The power spectrum of early type galaxies (shown by crosses) shows a higher amplitude than that of late type galaxies (shown by circles) by about a factor ~ 1.8

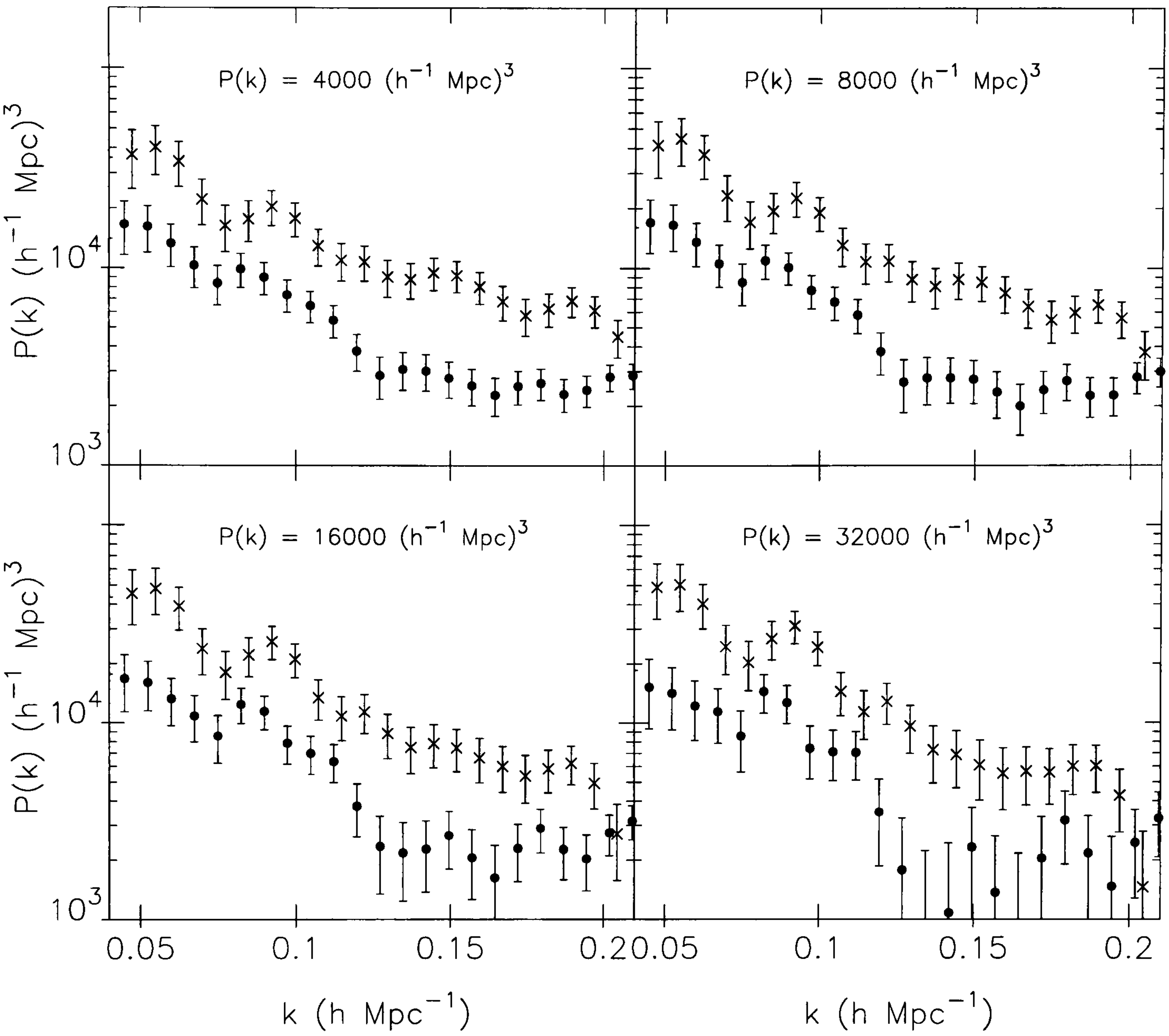


Figure 2.11: The power spectrum of galaxies in the Stromlo-APM survey as a function of morphological type. The crosses show the power spectra of early type galaxies in the Stromlo-APM survey, which are amplified over the power spectra of the late type galaxies in the survey (shown by circles) by a factor of about 1.8.

over the wave number range $0.05 \lesssim k \lesssim 0.1\ h\ \mathrm{Mpc}^{-1}$. This difference is very stable to the value of $P_w(k)$ applied in the weighting function. The power spectrum of all galaxies in the survey is almost identical (just very slightly higher) to that of late type galaxies alone, due to the fact that the power spectrum of the whole survey is heavily weighted to that of late type galaxies.

Clustering properties dependent on morphological type may be expected in hierarchical models of galaxy formation. For instance, if early type galaxies formed at the highest peaks in the density field then we would expect them to have an enhanced clustering amplitude (Bardeen *et al.* 1986).

Blumenthal *et al.* (1984) discuss galaxy formation in CDM-like models and derive the occupied region in the $logM - logT$ plane, where M is the total mass of an object and

T is the virial temperature, for objects corresponding to 0.5, 1, 2 and 3σ fluctuations in $\delta M/M$. They assume a cold dark matter spectrum with values of $n = 0,\ 1,\ 2$, $\Omega = 0.2,\ 1$ and $h = 0.5,\ 1$. The observed positions of spirals, ellipticals, clusters and groups of galaxies are then plotted in the $logM - logT$ plane (assuming that the total mass is ten times the luminous mass for galaxies). Elliptical galaxies are seen to lie along the 2σ fluctuation line while spiral galaxies tend to lie along the 1σ line. This is consistent with the observed statistics of the numbers of spirals versus the number of ellipticals ($\sim$ 50%:15%).

There are many plausible suggestions as to why elliptical galaxies should form preferentially at high sigma peaks in the density field. For instance, at a fixed mass, a higher peak has a higher binding energy. It will therefore collapse at a high redshift, and have more time to undergo mergers with other galaxies. If mergers produce elliptical galaxies (see e.g. Toomre 1977, Icke 1985), then the elliptical galaxies will naturally be associated with the high peaks in the density field.

Finally, it would seem likely that the results presented in the previous section (Section 2.5.2) are in part explained by the findings of this section, as the average absolute magnitude of the ellipticals in the sample is brighter than that of the spirals.

2.6 Redshift-space distortions of the power spectrum

Galaxy peculiar velocities cause a distortion of the clustering pattern measured in redshift-space compared to the true pattern in real-space (see *e.g.* Kaiser 1987). In this section, I analyse the distortions to the shape of the power spectrum measured in redshift-space using the N-body simulations described in Section 2.4. This allows us to quantify the effects of linear and non-linear peculiar velocities and hence to establish the range in wave number over which linear perturbation theory can be used to model the distortions. (For a similar analysis of redshift-space distortions in N-body simulations, see Gramman, Cen & Bahcall 1993). I then compare the redshift-space estimates of $P(k)$ for the Stromlo-APM survey with the real-space estimates of $P(k)$ for the APM galaxy survey, derived by Baugh & Efstathiou (1993), to quantify the effects of redshift-space distortion on large spatial scales.

2.6.1 Redshift-space distortions of $P(k)$ determined from N-body simulations

Figure 2.12 shows the power spectra measured in real-space and redshift-space for each of the ensembles of N-body simulations described in Section 2.4. In each case, the amplitude of the power spectrum measured in redshift-space is enhanced on large scales compared to the power spectrum measured in real-space. In linear perturbation theory, the power spectrum in redshift-space, $P_s(k)$, is related to the power spectrum in real-space, $P_r(k)$, according to the formula

$$P_s(\mathbf{k}) = P_r(\mathbf{k})\left(1+\beta\mu^2\right)^2, \qquad \beta = \frac{\Omega^{0.6}}{b}, \tag{2.34}$$

(Kaiser 1987). In this equation, μ is the angle between the vector $\mathbf{k}$ and the line-of-sight, b is a linear bias factor relating fluctuations in the galaxy distribution to fluctuations in the mass distribution, $(\delta\rho/\rho)_{gal} = b(\delta\rho/\rho)_{mass}$, and it is assumed that the power spectrum is determined from a patch of the Universe that subtends a small solid angle at the position of the observer. Averaging equation (2.34) over the angle μ, we obtain

$$P_s(k) = P_r(k)\left(1+\frac{2}{3}\beta+\frac{1}{5}\beta^2\right). \tag{2.35}$$

Thus the ratio $P_s(k)/P_r(k)$ depends on the parameter $\beta = \Omega^{0.6}/b$ in linear perturbation theory[2].

On small scales (wave numbers $k \gtrsim 0.1h\mathrm{Mpc}^{-1}$), the redshift-space power spectra in Figure 2.12 are suppressed in amplitude compared to the real-space spectra. This is caused by the small-scale peculiar velocities in high density regions that produce so called 'fingers of God' in redshift-space. The solid lines in Figure 2.12 show a simple model derived by Peacock & Dodds (1994) that incorporates both the linear theory distortion of equation (2.34) with a model for the distortions caused by fingers of God. The small scale peculiar velocities are assumed to be uncorrelated with position and drawn from a Gaussian distribution with one-dimensional dispersion σ_v. The combined effects of the linear redshift-space distortion and the non-linear small scale peculiar velocities lead to the following formula, relating the angle averaged power spectra in redshift-space and real-space:

$$P_s(k) = P_r(k)\, G(\beta, k\sigma_v) \tag{2.36}$$

where the function G is given by

$$G(\beta, k\sigma_v) = \frac{\sqrt{\pi}}{8}\frac{erf\,(k\sigma_v)}{(k\sigma_v)^5}\left[3\beta^2 + 4\beta\,(k\sigma_v)^2 + 4\,(k\sigma_v)^4\right]$$

$$-\frac{exp\left(-\,(k\sigma)^2\right)}{4\,(k\sigma)^4}\left[\beta^2\left(3+2\,(k\sigma)^2\right) + 4\beta\,(k\sigma)^2\right]. \tag{2.37}$$

[2]This is true even for low density spatially flat models, see Peebles (1984) and Lahav *et al.* (1991).

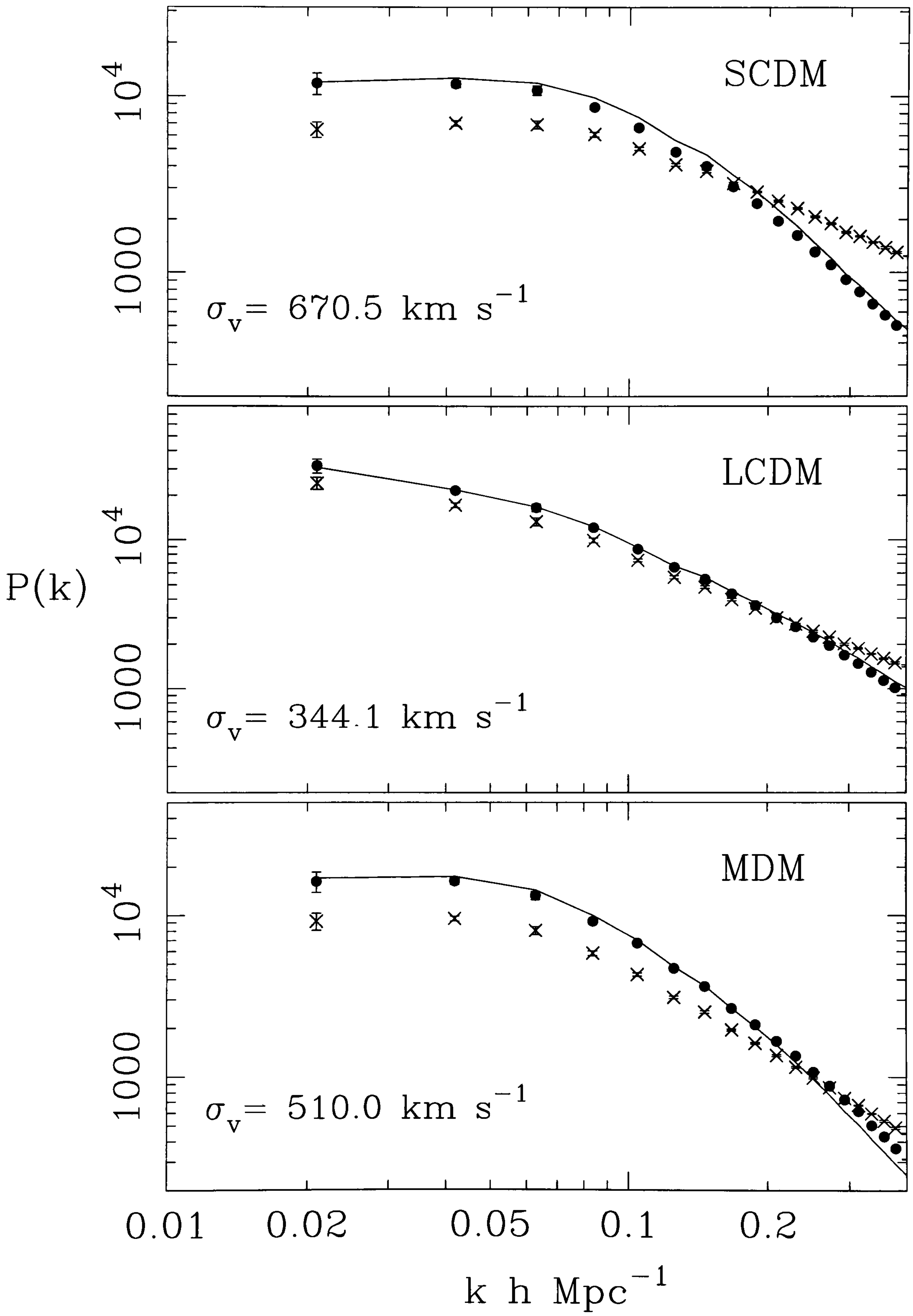


Figure 2.12: Power spectra evaluated in real and redshift-space for each ensemble. Redshift-space power spectra are shown by the circles and real-space spectra are shown by the crosses. The solid lines show the predicted redshift-space power spectrum, using equation (2.37) and the measured real-space power spectrum. Values for σ_v are shown in each panel.

The solid lines in Figure 2.12 show equation (2.37) for each ensemble, where I have used the one-dimensional velocity dispersion determined from the simulations $\sigma_v = 670.5, 344.1$, and $510.0\,\mathrm{km\ s^{-1}}$ for the SCDM, LCDM and MDM ensembles respectively. These curves provide a good match to the results from the N-body simulations over the range of wave numbers plotted in Figure 2.12 and are more accurate than the model described by Gramman, Cen & Bahcall (1993). Nevertheless, the model of equation (2.37) is simplistic, as it ignores non-linear streaming between galaxy pairs (see e.g. Fisher & Nusser 1996).

Figure 2.13 shows the ratio of the redshift-space and real-space power spectra plotted in Figure 2.12. The solid lines show the linear theory relation of equation (2.35) and the dotted lines show equation (2.37). These results show that at wave numbers $k \gtrsim 0.1h\mathrm{Mpc}^{-1}$, small-scale peculiar velocities cause a significant reduction in the amplitude of the power spectrum measured in redshift-space. One must therefore be cautious in determining β from redshift distortions using the linear theory relations in equations (2.34) and (2.35). The range of validity of linear theory is strongly dependent on the amplitude of the small-scale peculiar velocity field, as expected from equation (2.37). The biases can be large, for example, even at $k = 0.1h\mathrm{Mpc}^{-1}$, the ratio of the power spectra in the MDM model is equal to 1.57 compared to the linear theory prediction of 1.87.

The amplitude of the small-scale peculiar velocity field is relatively poorly constrained by observations. For example, Davis & Peebles (1983) found an *rms* relative peculiar velocity of $v_{12} = 340 \pm 40\mathrm{kms}^{-1}$ between galaxy pairs separated by $\sim 1\ h^{-1}\mathrm{Mpc}$ from an analysis of the CfA-1 survey and Bean *et al.* (1983) found $v_{12} = 250 \pm 50\mathrm{kms}^{-1}$ from an analysis of a smaller, but deeper survey. However, values as high as 600–800kms^{-1} have been derived in the literature (Hale-Sutton *et al.* 1989, Mo, Jing & Bórner 1993). The best constraints on v_{12} for optical galaxies come from an analysis of the CfA-2 and Southern Sky Redshift Survey by Marzke *et al.* (1995) who find $v_{12} = 540 \pm 180\mathrm{kms}^{-1}$. The error on v_{12} is large and the Marzke *et al.* (1995) results are consistent with the small-scale peculiar velocities ($\sigma_v \sim v_{12}/\sqrt{2}$) for the mass points in the LCDM and MDM models but are lower than those of COBE normalized SCDM models. In principle, with a larger survey, we could use equation (2.36) (or a more complicated non-linear model) to simultaneously determine the parameters β and σ_v (see Cole, Fisher & Weinberg 1995, for an application to redshift surveys of IRAS galaxies).

In addition to the effects of non-linear peculiar velocities, biases in the estimates of $P(k)$ derived from equation (2.22) will also introduce departures from the linear theory

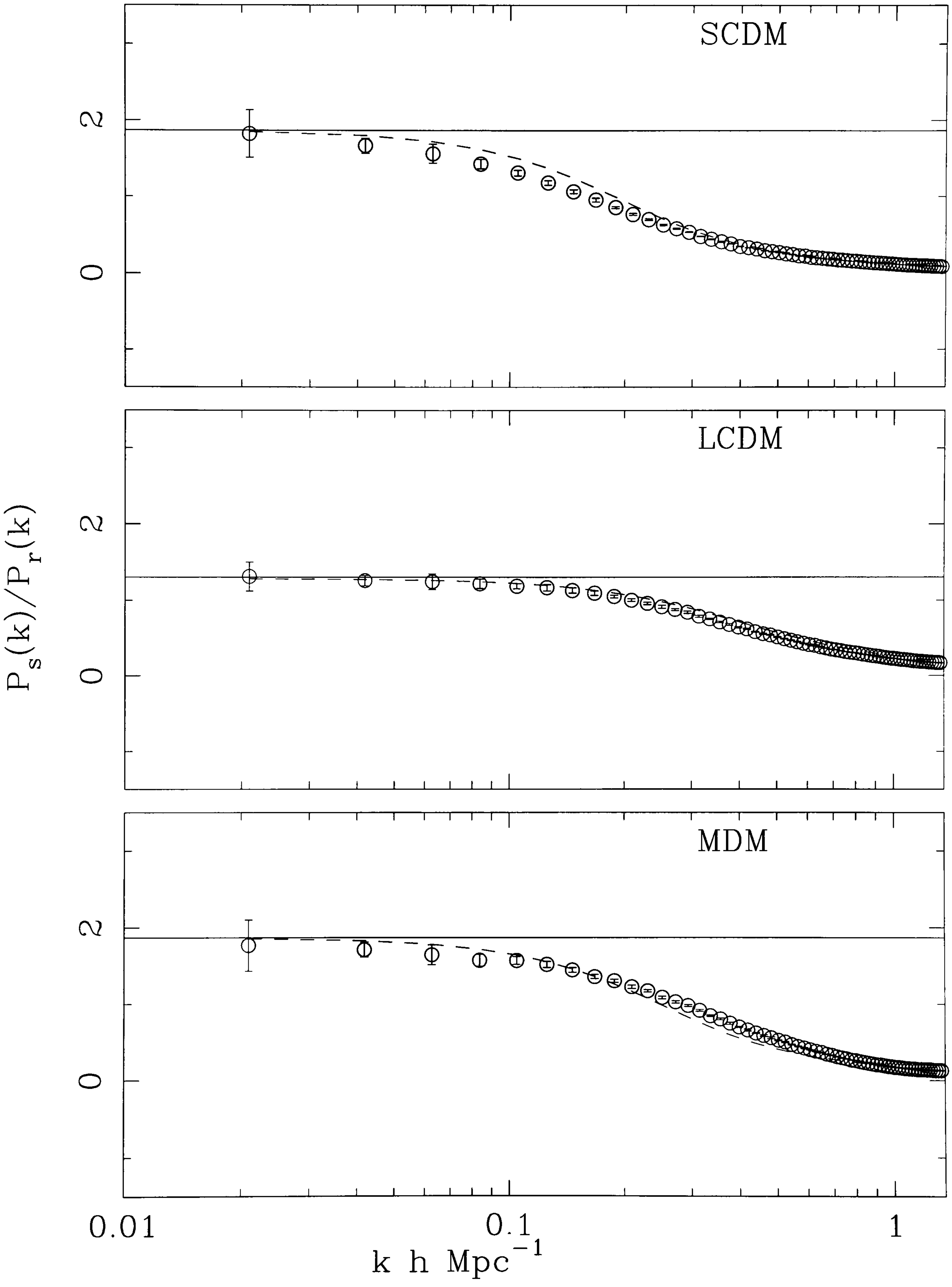


Figure 2.13: The ratio of redshift-space to real-space power for the full N-body simulations (box size 300 h^{-1}Mpc). The solid lines shows the linear theory predictions and the dashed lines show equation (2.37) using the same values of σ_v as in Figure 2.12

predictions for the ratio $P_s(k)/P_r(k)$ at small wave numbers. This is illustrated in Figure 2.14, where I have plotted the ratio of the redshift-space power spectra of the mock Stromlo-APM simulations to the real-space power spectra measured from the full N-body simulations. The decline in the ratio at $k \lesssim 0.05h\mathrm{Mpc}^{-1}$ is caused by the biases in the power spectrum estimates discussed in Appendix A. Even at wave number $k \sim 0.05h\mathrm{Mpc}^{-1}$, the ratios for the LCDM and MDM ensembles fail to reach the linear predictions because the convolution of the redshift-space power spectrum with the window function, and the effect of uncertainties in the mean number density of the survey, cause a slight depression in its amplitude (*cf* Figure 2.3).

2.6.2 Estimates of β from redshift-space distortions

In this section, I attempt to estimate β by comparing the Stromlo-APM power spectrum with real-space estimates of $P(k)$ for the APM survey. Baugh and Efstathiou (1993, hereafter BE) have described a method for recovering the three-dimensional power spectrum from the angular two-point correlation function $w(\theta)$. The inversion requires a knowledge of the redshift distribution of the galaxies used to estimate $w(\theta)$, of the geometry of the Universe and of the evolution of the power spectrum. The redshift distribution of APM galaxies is well constrained by observations (see Maddox, Efstathiou & Sutherland 1996). However, BE show that uncertainties in the evolution of $P(k)$ and the cosmological model introduce uncertainties of $\sim 20\%$ in the amplitude of the real-space power spectrum. As an illustration of these uncertainties, I will use two estimates of the real-space power spectrum derived from APM galaxies in the magnitude range $17 < b_J < 20$, plotted in Figure 2.15. In these inversions an Einstein-de Sitter universe is assumed and a power spectrum that evolves as $P(k,z) = P(k)/(1+z)^{\alpha}$ with $\alpha = 0$ (filled circles) and $\alpha = 1.3$ (open circles) corresponding to a clustering pattern that is stable in comoving and physical coordinates respectively (see BE for further details). Uncertainties in the redshift distribution and cosmological model introduce similar systematic differences in the real-space power spectra, but since these errors are considerably smaller than the random errors in the redshift-space estimates of $P(k)$, I do not discuss them in detail here.

The open squares in Figure 2.15 show the $z_{max} = 0.06$ redshift-space estimate of $P(k)$ for the Stromlo-APM survey as calculated in Section 2.5. Qualitatively, the behaviour is similar to that seen in the simulations; at wave number $k \lesssim 0.03\ h\mathrm{Mpc}^{-1}$ the redshift-space power spectrum declines and falls below the real-space estimates. Over the wave number

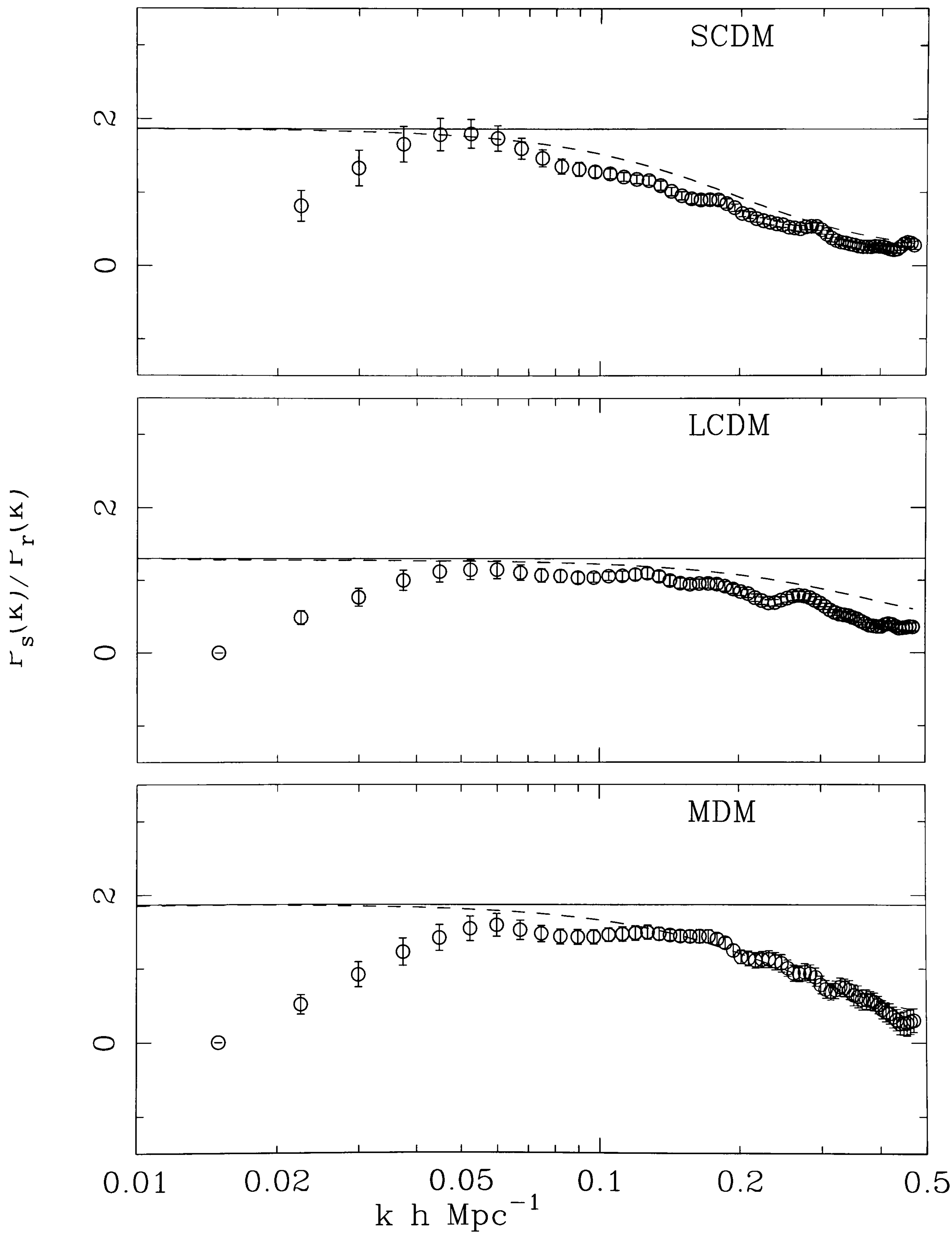


Figure 2.14: The ratios of the redshift-space power spectra, measured for sparsely sampled mock Stromlo-APM surveys, to the real-space power spectra for the full N-body simulations. The solid and dashed lines are as in Figure 2.13.

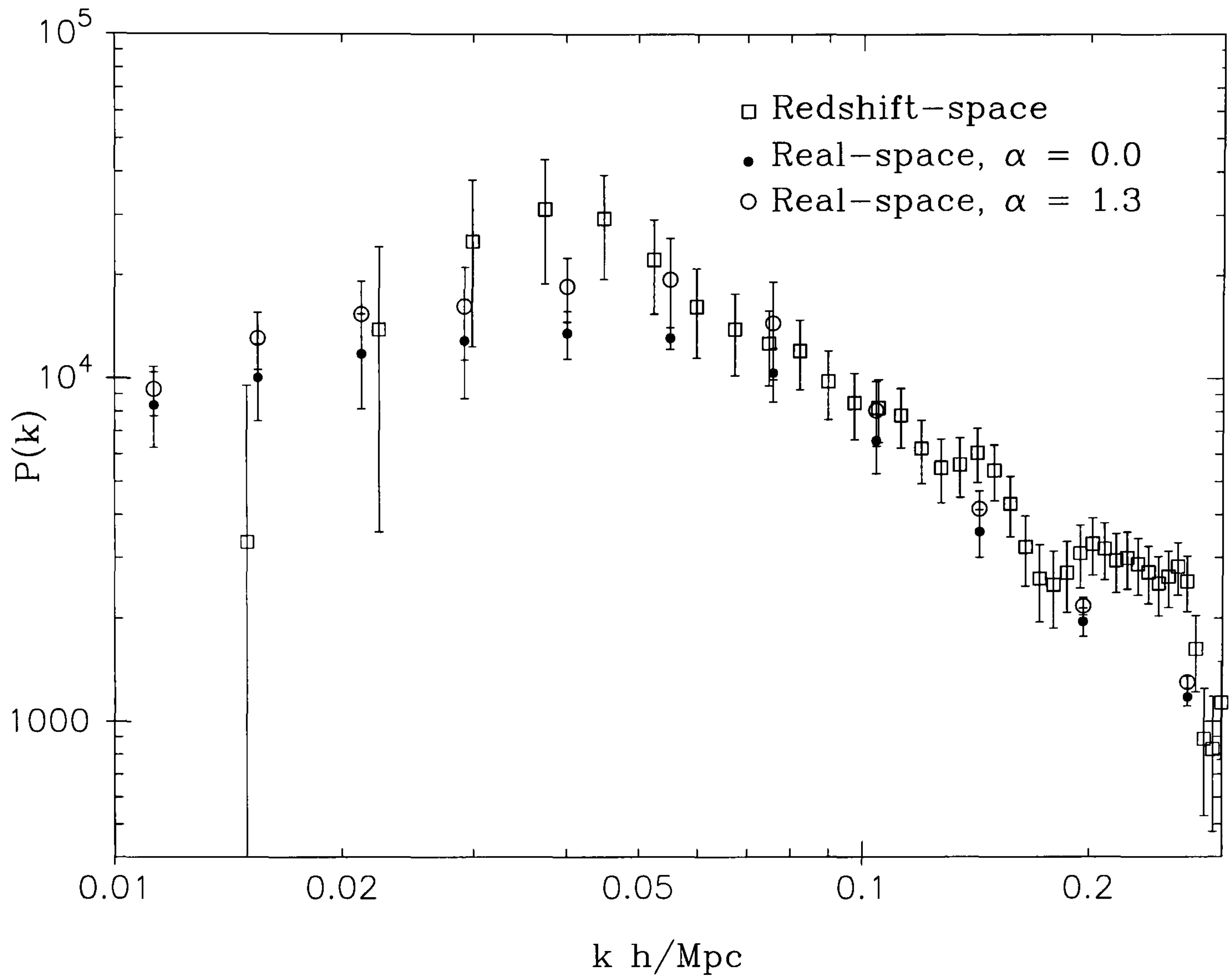


Figure 2.15: The open squares show the redshift-space power spectrum for a volume limited sample from the Stromlo-APM survey limited at $z_{max} = 0.06$. The circles show real-space power spectra, determined by BE, via inversion of the angular correlation function of APM galaxies in the magnitude slice $17 < b_J < 20$. The filled and open circles show the inversion assuming $\alpha = 0$ and $\alpha = 1.3$ respectively, where the power spectrum is assumed to evolve according to $P(k,z) = P(k)/(1+z)^{\alpha}$. I have plotted 1σ errors on the Stromlo-APM estimates, as in Figure 2.7. The error bars on the real-space power spectra show the standard deviation determined from the scatter in the estimates derived from 4 roughly equal area zones of the APM survey as discussed in BE.

range $0.03 \lesssim k \lesssim 0.07$ $h\mathrm{Mpc}^{-1}$ the redshift-space amplitude is enhanced suggesting a significant distortion of the clustering pattern in redshift-space. At larger wave numbers, the errors in the Stromlo-APM power spectrum become large because it is a sparse sampled redshift survey and so contains little information on the small-scale clustering of galaxies.

I estimate the amplitude of the redshift-space distortions in the Stromlo-APM survey as follows. The ratio $P_s(k)/P_r(k)$ is computed by linearly interpolating the real-space power spectra plotted in Figure 2.15 to the k values used in the redshift-space distortion analysis, and I average this ratio over the wave number range $0.05 < k < 0.1$ $h\mathrm{Mpc}^{-1}$. This leads to an estimate of β from equation (2.35). To estimate the error in β, I compute the scatter in the ratio of the redshift-space to real-space power spectra from the mock

	α=0.0	α=1.3
	β	β
Flux limited $P_s(k)$	0.74 ± 0.48	0.20 ± 0.44
Volume Limited $P_s(k)$	0.38± 0.67	-0.13 ± 0.64

Table 2.2: Estimates of β from the Stromlo-APM survey.

Stromlo-APM surveys plotted in Figure 2.14 averaged over the same wave number range as the data. This provides an estimate of errors associated with sampling fluctuations in the redshift survey. I estimate the error associated with uncertainties in the real-space estimate of $P(k)$ from the scatter in the ratio $P_s(k)/P_r(k)$ using $P_r(k)$ determined from four roughly equal area zones of the APM survey (see BE). The final error estimate on β is derived by adding these two error terms in quadrature on the assumption that they are uncorrelated. Table 2.2 shows the estimates of β derived using the two real-space estimates of $P(k)$ plotted in Figure 2.15, and two estimates of the redshift-space power spectrum for the Stromlo-APM survey; the $z_{max} = 0.06$ volume limited results plotted in Figure 2.15 and the flux-limited results from Figure 2.6, with a weight function (equation 2.28) in which I have set $P_w(k) = 8000(h^{-1}\mathrm{Mpc})^3$.

These results suggest a positive value of β, though the errors are large. The differences in the flux-weighted and volume limited estimates of $P(k)$ over the wave number range $0.05 < k < 0.1$ $h\mathrm{Mpc}^{-1}$ lead to an uncertainty of nearly a factor of two in the derived value of β. A similar uncertainty is introduced by the difference in the real-space estimates of $P(k)$ derived from the APM survey for the two adopted values of the evolution parameter α.

The inferred errors on β are larger than those quoted in other analyses of redshift-space distortions. For example, Hamilton (1993a) finds $\beta = 0.66^{+0.34}_{-0.22}$ from an analysis of angular moments of the two-point correlation function of the $2Jy$ IRAS redshift survey of Strauss *et al.* (1992a); Cole, Fisher & Weinberg (1995) analyse angular harmonics of the power spectrum and find $\beta = 0.52 \pm 0.13$ and $\beta = 0.54 \pm 0.3$ for the $1.2Jy$ and QDOT IRAS surveys respectively. Loveday *et al.* (1995b) find $\beta = 0.48 \pm 0.12$ from an analysis of anisotropies in the redshift-space correlation function of the Stromlo-APM survey. At face value the estimates of β in Table 2.2 appear to disagree with those of Loveday *et al.* (1995b), who analyse the same survey. However, it seems likely that Loveday *et al.* (1995b) have underestimated the error on β. As a further check of whether these errors are realistic, Table 2.3 lists the errors and biases in β derived by applying the above analysis to the mock

Model	β	$\delta\beta$	Linear β
SCDM	0.64	0.54	1.00
LCDM	0.13	0.55	0.38
MDM	0.62	0.52	1.00

Table 2.3: Estimates of β derived from the ratio P_s/P_r over the wave number range $0.05 < k < 0.1\ h\mathrm{Mpc}^{-1}$ from mock Stromlo-APM surveys.

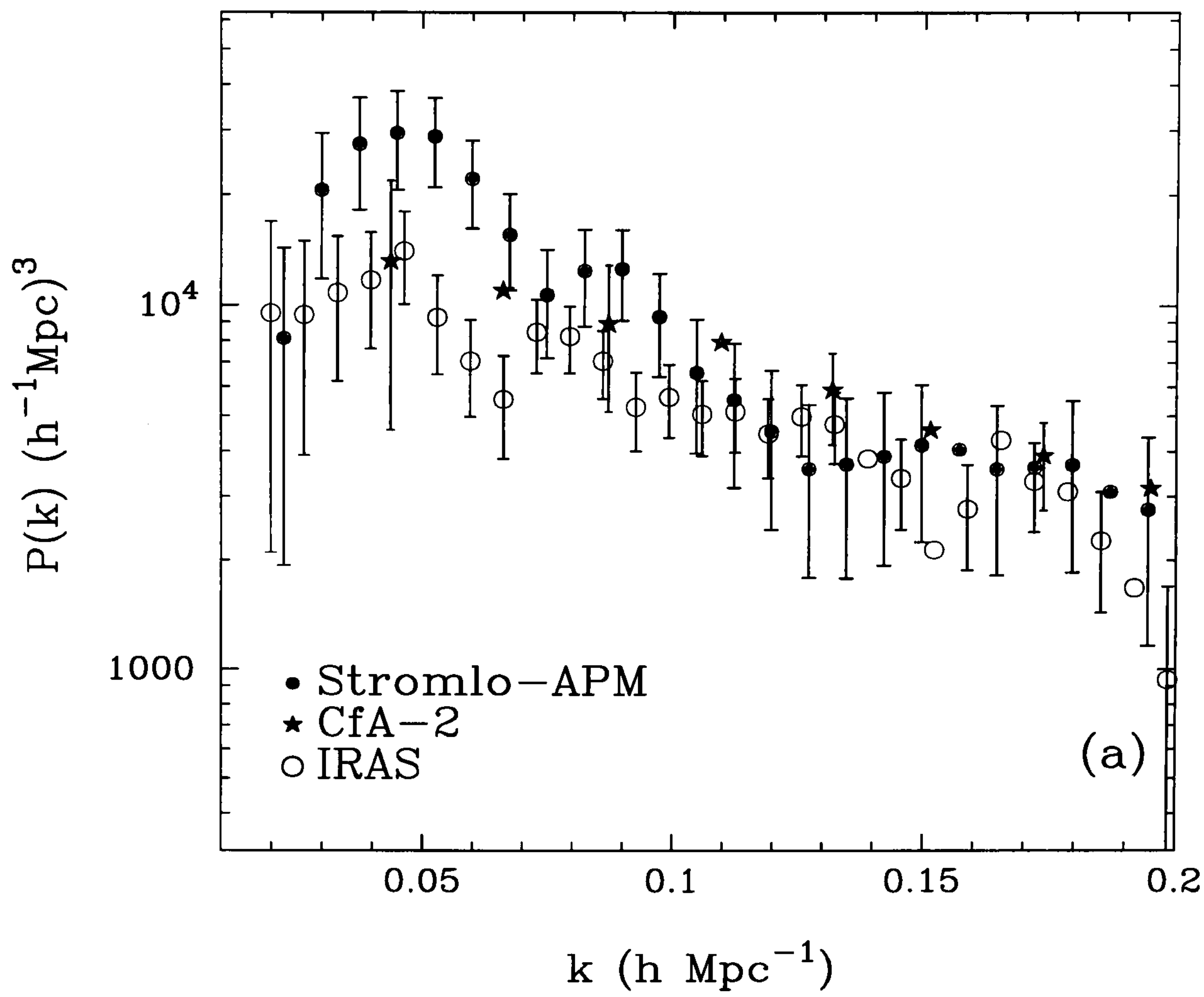
Stromlo-APM surveys constructed from the simulations. Here I have computed β from the ratio of the redshift-space power spectra of the mock surveys to the mean real-space power spectrum for each ensemble, and have neglected errors in the real-space estimate $P(k)$ since these are negligible. In each case, the errors in β are close to $\delta\beta \approx 0.5$ in good agreement with the error estimates for the real data. In addition, Table 2.3 shows that β is underestimated in each case because of the non-linear corrections described in the previous subsection and because of the biases in the redshift-space estimates of $P(k)$ described in Appendix A.

The results given in Table 2.2 are in good agreement with those determined by Baugh (1996) from a comparison of the redshift-space two-point correlation function of the Stromlo-APM survey and the real-space correlation function inferred by inverting the angular correlation function of the APM survey.

This analysis suggests that the results of Table 2.2 should be considered as lower bounds on the true value of β and hence that $\beta = 1$ is compatible with the observations. With larger redshift surveys it should be possible to fit a more complicated model to the data so extending the useable range of wave numbers and reducing the biases in β, *e.g.* equation (2.37) with β and σ_v as free parameters, or the non-linear model of Fisher & Nusser (1996) based on the Zeldovich approximation.

2.7 Comparison with other surveys and theoretical models

In Figure 2.16, I compare the Stromlo-APM power spectrum with power spectra estimated from other surveys. The filled circles show the flux limited estimates of $P(k)$ for the Stromlo-APM survey using $P_w(k) = 8000\,(\,h^{-1}\mathrm{Mpc})^3$ in the weighting scheme of equation (2.28). The open symbols show $P(k)$ for the combined $1.2Jy$ and QDOT IRAS redshift surveys, as analysed in Chapter 4 (or see Tadros & Efstathiou 1995); these estimates are for a flux limited sample with $P_w(k) = 8000\,(\,h^{-1}\mathrm{Mpc})^3$ in the weighting scheme of equation (2.28). The filled stars show $P(k)$ from Park *et al.* (1994) for a volume limited subset of



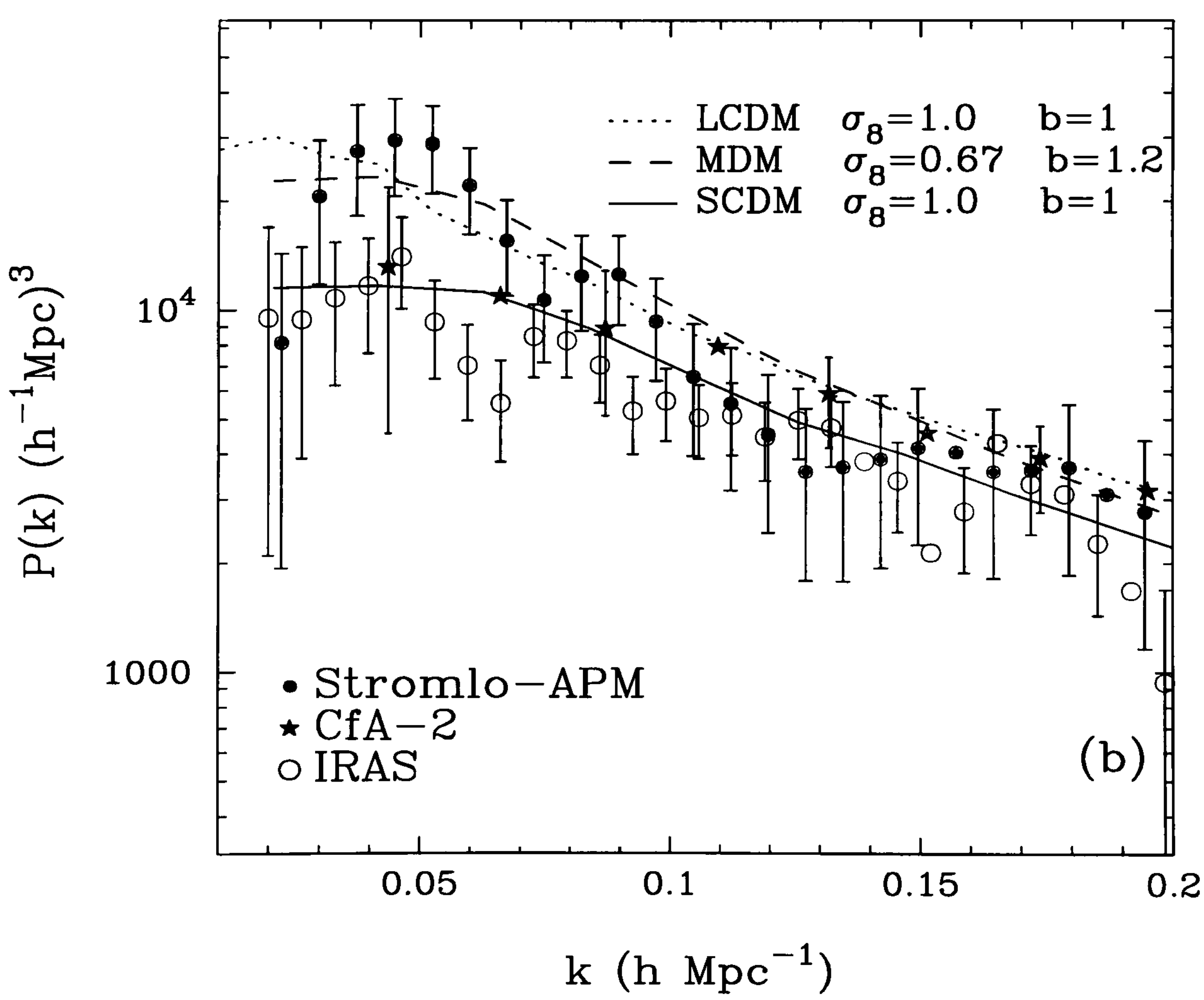


Figure 2.16: Estimates of the power spectra for various redshift surveys. The filled circles show the flux-limited estimate of $P(k)$ for the Stromlo-APM survey, as plotted in Figure 2.7. The open circles show a flux limited estimate of $P(k)$ for IRAS galaxies from Tadros and Efstathiou (1995). The filled stars show estimates of $P(k)$ from Park *et al.* (1994) for a volume limited subset of the CfA-2 survey. The three curves in Figure 2.16b show redshift-space power spectra determined from the mass distributions in N-body simulations of the three CDM-like models described in the text; σ_8 gives the *rms* amplitude of the mass fluctuations in spheres of radius $8h^{-1}$Mpc and I have multiplied the power spectrum of the MDM model by $b^2 = (1.2)^2$ to approximately match the power spectra of the Stromlo-APM and CfA-2 surveys.

the CfA-2 survey consisting of 1509 galaxies to a coordinate distance of $101h^{-1}Mpc$. The power spectra for the two optically selected catalogues are consistent with each other within the errors, but the amplitude of the IRAS power spectrum is slightly lower indicating that IRAS galaxies are less strongly clustered than optically selected galaxies. The results shown in Figure 2.16 are consistent with a linear relative bias between optical and IRAS galaxies of $b_{opt}/b_{IRAS} \sim 1.2$.

Recently, Landy *et al.* (1996) have computed a two-dimensional power spectrum for the Las Campanas redshift survey and find evidence for a strong peak in the power spectrum at wave numbers $k \sim 0.06h\mathrm{Mpc}^{-1}$. Their results are not easily comparable to the ones presented in this chapter and it is unclear whether there is any inconsistency with the three-dimensional power spectrum inferred for the APM galaxy survey (Baugh & Efstathiou 1993) or with the redshift-space power spectrum of the Stromlo-APM survey presented here.

In Figure 2.16, I show the empirical estimates of $P(k)$ together with redshift-space estimates of $P(k)$ determined from N-body simulations of the LCDM, MDM, and SCDM models. The shape of the SCDM curve is in good agreement with the power spectra measured for the IRAS and CfA-2 surveys, though it lies low compared to the Stromlo-APM data points at wave number $k \sim 0.05h\mathrm{Mpc}^{-1}$. The high amplitude for the SCDM model implied by the COBE temperature anisotropies results in a high amplitude for the peculiar velocity field. This in turn leads to a significant difference in the shape of the redshift-space power spectrum compared to the real-space power spectrum (see also Bahcall, Cen & Gramman 1993) and so the model can provide an acceptable match to the redshift-space $P(k)$ although it fails to match the shape of the real-space power spectrum (see *e.g.* BE, Maddox, Efstathiou & Sutherland 1996). As mentioned in Section 2.6.1, the small scale ($\sim 1h^{-1}\mathrm{Mpc}$) relative peculiar velocities predicted by the COBE normalised SCDM model are considerably higher than the relative motions between galaxy pairs measured by Marzke *et al.* (1995). Both the LCDM and MDM models provide acceptable fits to the power spectra measured from the redshift surveys and distinguishing between these models would require accurate estimates of the effects of redshift-space distortions. As we have seen from the analysis of the Stromlo-APM survey, larger redshift surveys, such as the AAT 2-degree field redshift survey (see Efstathiou 1996) and the Sloan Digital Sky Survey (Gunn & Weinberg 1995) are required to determine β accurately and so distinguish between these models.

2.8 Summary

The conclusions of this chapter are as follows:

[1] I have tested estimators of $P(k)$ using simulated redshift surveys constructed from N-body simulations. These tests show that estimates of $P(k)$ are biased at low wave numbers if the surveys are used to estimate the mean galaxy density. Formulae for these biases are given in Appendix A.

[2] I have estimated $P(k)$ for the Stromlo-APM redshift survey using volume limited and flux limited samples. The power spectra are insensitive to the volume limit and to the galaxy weights applied in the analysis of flux limited samples.

[3] I have investigated whether the amplitude of $P(k)$ depends on galaxy luminosity and used N-body simulations to assess the statistical significance of the analysis. No evidence for any significant luminosity dependence is found except possibly at absolute magnitudes brighter than $M = -20.3$, where some evidence is seen for a higher amplitude. This is broadly consistent with the analysis of the CfA-2 redshift survey by Park *et al.* (1994) (and with the analysis of the SSRS2 optical redshift survey by da Costa *et al.* 1994), who find that the amplitude of $P(k)$ for galaxies brighter than M^* is about 40% higher than the amplitude measured from fainter galaxies.

[4] I have analysed numerical simulations to determine the effects of redshift-space distortions on the shape of $P(k)$. The distortions can be well approximated by the formula of Peacock & Dodds (1994), equation (2.37), which depends on two parameters, $\beta = \Omega^{0.6}/b$ and a measure of the small-scale rms peculiar velocity, σ_v.

[5] I estimate the redshift-space distortions in the Stromlo-APM survey by comparing redshift-space power spectra with estimates of the real-space power spectrum determined by inverting the angular correlation function measured for the parent APM galaxy survey. The results indicate a positive value of β, consistent with other work, but the uncertainties are large and do not exclude $\beta = 1$.

[6] The power spectra measured for the Stromlo-APM survey are in good agreement with power spectra measured for other optically selected redshift surveys. The amplitude of the Stromlo-APM power spectrum is systematically higher than that measured for IRAS galaxies, consistent with a linear relative bias between optical and IRAS galaxies of $b_{opt}/b_{IRAS} \sim 1.2$.

[7] The Stromlo-APM power spectrum has a higher amplitude at wave numbers $k <$

0.08 h Mpc^{-1} than expected for the COBE normalized standard CDM model but is well fitted by COBE normalized MDM and LCDM models.

Chapter 3

Power spectrum analysis of the APM rich cluster catalogue

3.1 Analysis of cluster redshift surveys

Clusters of galaxies are useful tools in the study of large scale structure in the Universe. Cluster redshift surveys probe much greater volumes of the Universe than galaxy redshift surveys, and the results from cluster surveys can provide strong constraints on models of structure formation. The construction of complete and reliable cluster catalogues is, however, a non-trivial task. Clusters are rare objects and are difficult to identify optically, although the situation has been much improved with the advent of automatic plate measuring machines (see Section 3.1.4). It is also possible to identify clusters via their X-ray emission and this is now being applied to construct X-ray flux limited cluster catalogues (see Section 3.1.4). Obtaining the redshift of an identified cluster is also a difficult task due to contamination by foreground and background galaxies along the line of sight to the cluster. Nevertheless, measurements of large scale structure from cluster surveys now provide reliable constraints on cosmological models and can also provide some insight into the relation between the clustering of galaxies relative to that of the mass (see Section 3.4.2).

In Section 3.1.1 of this chapter I review previous analyses of cluster-cluster correlations, performed using the Abell catalogue of galaxy clusters (Abell 1958) and its extension to the southern Hemisphere (Abell, Corwin & Olowin 1989). In Section 3.1.2 I discuss the systematic biases likely to be present in these catalogues and how they affect the correlation function analyses. Previous power spectrum analyses of cluster surveys are described in Section 3.1.3. Section 3.1.4 describes the new generation of cluster redshift surveys which have been convincingly shown to be free of the selection biases affecting the Abell catalogue. The new surveys are those constructed using automatic plate measuring machines to identify clusters, and those using X-ray emission to identify clusters. The large scale structure results from these new catalogues are summarized in Section 3.1.4. The rest of this chapter (Sections 3.2 to 3.4) describes my analysis of the power spectrum of the APM rich cluster redshift survey. I present a comparison with the results expected from LCDM, SCDM and MDM models and discuss the biases and uncertainties in the measurement of the power spectrum of this survey. I show that the power spectrum can be reliably measured to very small wave numbers ($k \sim 0.03\, h^{-1}$Mpc) using this sample. Finally, I present a comparison of the power spectrum of the APM rich clusters with that of the Stromlo-APM galaxies. I summarize in Section 3.5.

3.1.1 Correlation function analysis of cluster catalogues

One of the first measurements of the clustering properties of clusters was presented by Hauser & Peebles (1973) who examined the angular correlation function for clusters of galaxies in Abell's catalogue (Abell 1958). Hauser and Peebles concluded that clusters were more strongly clustered than galaxies and that the richer clusters seemed to show a higher clustering amplitude than the poorer ones. This agrees with more recent work by Bahcall & Soneira (1983) (see also Klypin & Kopylov 1983). Bahcall & Soneira (1983) used the Abell catalogue to determine the angular correlation function $w(\theta)$ for clusters of richness classes $R = 1$ and $R \geq 2$[1]. They found that the amplitude of $w(\theta)$ was strongly dependent on richness with the richer class ($R \geq 2$) being more clustered by a factor of ~ 3 than the poor class (R = 1), the amplitude of the total cluster correlation function being dominated by the more numerous $R = 1$ clusters. The spatial correlation function they found could be fit by a power law

$$\xi(r) = 300\,(rh)^{-1.8} \quad 5 \lesssim r \lesssim 150\, h^{-1}\mathrm{Mpc} \quad (R \geq 1) \tag{3.1}$$

with $\xi(r) \sim 10$ for $r \lesssim 10\ h^{-1}$Mpc and $\xi(r) \sim 1$ at $r \sim 25\ h^{-1}$Mpc. They observed no significant correlation for $r \gtrsim 150\ h^{-1}$Mpc. Bahcall & Soneira (1983) also drew attention to the fact that the cluster correlation function has approximately the same power law form

[1] Abell classified clusters into richness classes by using a metric radius of 1.5 h^{-1}Mpc around the centre of each cluster and counting the number of galaxies, in excess of the background count, within this radius. He counted galaxies with magnitudes between m_3 and $m_3 + 2$ where m_3 is the magnitude of the third brightest galaxy in the cluster. Clusters with $30 - 49$ galaxies were assigned richness class $R = 0$, if they had $50 - 79$ galaxies they were assigned richness class $R = 1$ and those with $80 - 129$ were assigned richness class $R = 2$. He also assigned the clusters distance classes D on the basis of the magnitude of the tenth brightest cluster galaxy. To give an idea of distance, the outer boundary of distance class $D = 4$ is at $\sim 300\ h^{-1}$Mpc. Abell stated that his catalogue was likely to be incomplete for richness class $R = 0$ clusters.

as the galaxy correlation function, but with an amplitude 15 – 18 times larger. Bahcall & Soneira (1983) also attempted to quantify the magnitude of the peculiar velocities of clusters by comparing the observed distribution of redshift differences between pairs of galaxies, at various angular separations, with the distribution predicted by integrating the spatial correlation function. These two quantities agreed for large angles but at small angular separations the observed distribution appeared broader than that expected for pure Hubble flow. This was interpreted as being due to peculiar velocities of order 2000km s^{-1} (which includes the possible measurement errors). These are very large peculiar velocities and alternative explanations for these observations have been proposed (see Section 3.1.2).

All of these results were controversial and difficult to reconcile with most theories of galaxy formation where structure grew from initially Gaussian perturbations. For instance, SCDM predicts that the mass correlation function will become negative on scales $\gtrsim 20(\Omega h^2)^{-1}$Mpc (Davis *et al.* 1985). At these large scales the cluster correlation function is expected to be proportional to the mass correlation function (Croft & Efstathiou 1994a), implying a discrepancy between the SCDM model and the results of Bahcall & Soneira (1983). The high measured amplitude at both intermediate and large scales cannot be matched in N-body simulations of the standard CDM model, whether the fluctuations are normalized to the microwave background or to the galaxy distribution (Bardeen, Bond & Efstathiou 1987, White *et al.* 1987b and Blumenthal, Dekel & Primack 1988).

Since Bahcall and Soneira's pioneering work, the richness dependence of the clustering strength of clusters has been investigated further, both theoretically and with larger data sets. In most theories of galaxy formation, the more massive objects are expected to be more clustered than the less massive objects, so a dependence of the clustering strength on richness is expected. After investigation of this effect using catalogues mainly drawn from the Abell catalogue (Bahcall & Burgett 1986, Bahcall 1988, Bahcall & West 1992, Szalay & Schramm 1985) a scaling law for the correlation length as a function of richness was proposed:

$$r_0 \sim 0.4 d_c \qquad 30\, h^{-1}\text{Mpc} \lesssim d_c \lesssim 90\, h^{-1}\text{Mpc} \tag{3.2}$$

where d_c is the mean inter-cluster distance and equals $n_c^{-1/3}$, n_c being the mean space density of clusters of a given richness class. Bahcall & West (1992) showed that the above relation appeared to hold for all cluster catalogues from which $\xi(r)$ had been estimated (e.g. Bahcall & Soneira 1983, Postman, Geller & Huchra 1986, Lahav *et al.* 1989, Nichol *et al.* 1992, Dalton *et al.* 1992).

The clustering dependence on richness, or equivalently number density, has also been examined in N-body simulations. It is easier to simulate cluster catalogues than to simulate galaxy catalogues since the clusters are large mass concentrations which are hopefully insensitive to the details of galaxy formation. We can therefore hope to identify rich clusters directly from the mass distributions in N-body simulations using objective cluster finding algorithms (see for example Gaztanaga, Croft & Dalton 1995, but see also Eke *et al.* 1996 who point out possible large scale structure sensitivities caused by different cluster selection algorithms). Croft & Efstathiou (1994a) simulated APM rich cluster surveys from three cosmological models, LCDM, SCDM and MDM using the N-body simulations described in Chapter 2. They simulated sets of clusters with inter-cluster separations between $d_c \sim 10\ h^{-1}$Mpc and $d_c \sim 70\ h^{-1}$Mpc increasing d_c in steps of 5 h^{-1}Mpc, and found that the correlation length r_0 of the clusters was extremely insensitive to d_c for $d_c \gtrsim 30\ h^{-1}$Mpc. They showed that this behaviour was well matched by the results from the automated plate measuring surveys of clusters (see Section 3.1.4) which had been carried out (Nichol *et al.* 1992, Dalton *et al.* 1992, Dalton 1992, Dalton *et al.* 1994). Bahcall & Cen (1992) also analysed simulated clusters in a low density CDM-like universe. They found a strong dependence of clustering strength with richness (or inter-cluster separation). Their result approximately matched the prediction of equation (3.2). However, Croft & Efstathiou (1994a) argue that this result was a statistical fluctuation.

The richness dependence of clustering strength has been studied analytically using the statistics of Gaussian density fields (Kaiser 1984, Bardeen *et al.* 1986, Bardeen, Bond & Efstathiou 1987, Holtzman & Primack 1993, Mann, Heavens & Peacock 1993, Croft & Efstathiou 1994a). For instance Croft & Efstathiou (1994a) find a theoretical dependence of r_0 with d_c quantitatively similar to that seen in simulations (i.e. little dependence beyond about $d_c = 30\ h^{-1}$Mpc), although the theoretical predictions of the correlation lengths do not match in detail the results from numerical simulations. Mann, Heavens & Peacock (1993) find a similarly weak trend in their theoretical predictions, in contrast to the strong dependence of r_0 on d_c shown in equation (3.2).

3.1.2 Errors in the measurement of $\xi(r)$

As mentioned previously, clusters of galaxies are difficult to identify from photographic plates. Cluster catalogues have generally been afflicted by incompleteness and inhomogeneity. Catalogues constructed by visual examination of photographic plates are subject to two main selection biases. The first is due to the fact that the limiting magnitude of the uncalibrated photographic plates from which the cluster catalogues are selected is not uniform. More clusters will be identified on deeper plates than on shallower plates. This causes a modulation of the detection efficiency correlated with position on the sky. This effect can be seen clearly in the Southern Abell catalogue (see Dalton 1992). The second bias is due to the fact that clusters overlap on the plane of the sky. The cluster-galaxy cross correlation function is positive out to fairly large scales ($\sim 20\ h^{-1}$Mpc) (Seldner & Peebles 1977, Lilje & Efstathiou 1988). This means that clusters of galaxies are surrounded by large halos of galaxies. A cluster near to a previously identified cluster would thus have its richness artificially boosted by the halo galaxies of its neighbour. Consequently, a cluster near another cluster is more likely to be included in the cluster catalogue.

This type of 'projection bias' can be tested for in cluster redshift surveys (Sutherland 1988, Soltan 1988). It manifests itself as anisotropies in the clustering pattern. If the spatial correlation function is calculated as a function of two variables, the separation of pairs perpendicular to the line-of-sight, σ, and the separation of pairs parallel to the line-of-sight, π, then projection biases will cause the contours of $\xi(\sigma,\pi)$ to be elongated along the π axis. This has the same appearance as large 'fingers of God' but is not caused by peculiar velocities. The peculiar velocities required would be $\gtrsim 2000$km s^{-1} (see Section 3.1.1, Bahcall & Soneira 1983). Velocities this large are are not consistent with cluster peculiar velocities measured directly using independent distance indicators, which indicate velocities $\lesssim 500$km s^{-1} (Aaronson *et al.* 1986, Lynden-Bell *et al.* 1988 and Lucey & Carter 1988). Sutherland (1988) examined the behaviour of the function

$$\xi\,(\delta z, \theta) = \frac{N_\theta^c}{N_\theta^r} - 1 \tag{3.3}$$

where c signifies clusters, r random points, and N_θ is the number of pairs with angular separation $\theta \rightarrow \theta + d\theta$ and redshift separation $\delta z = |z_1 - z_2|$. This function would drop to zero as δz increased if peculiar velocities were the significant effect, whereas Sutherland (1988) finds that it remains constant out to very large redshifts.

Lucey (1983) examined the systematic effects of the process of identifying clusters from a projected galaxy distribution while Dekel *et al.* (1989) and Olivier *et al.* (1990) have modelled the possible effect it has on the measured cluster-cluster correlation function. They concluded that the amplitude of the cluster-cluster correlation function can be significantly enhanced by projection biases. The effect is clearly visible in the $\xi(\sigma, \pi)$ contour plots of the sample of 351 Abell clusters used by Postman, Huchra & Geller (1992) (see Efstathiou *et al.* 1992). For this sample the value of $\xi(\sigma, \pi)$ is unity out to separations $\pi \sim 70\ h^{-1}$Mpc for projected separations $\sigma \sim 20\ h^{-1}$Mpc. Even at $\pi \sim 100\ h^{-1}$Mpc the correlation function is ~ 0.5. Sutherland (1988) and Soltan (1988) also see similar anisotropies in the Abell cluster sample used by Bahcall & Soneira (1983), and Sutherland & Efstathiou (1991) detect the anisotropies in a deeper redshift survey of Abell clusters (Huchra *et al.* 1990) and in a sample of Lick clusters (Shectman 1985).

Direct observations of anisotropies have also been reported. For instance, Postman, Geller & Huchra (1986) comment on the existence of 'voids' behind nearby rich clusters. For example A1367 is a richness class 2, distance class 1 cluster (velocity $\sim$ 6400km s^{-1}) and subtends an area of about 15 square degrees on the sky. Behind it there appears to be a void extending to a distance of $\sim$ 18000km s^{-1}. A similar phenomenon is seen behind the Coma cluster ($R = 4$, $D = 1$, subtended area of 22 square degrees). This could be a selection effect, due to the fact that Abell failed to detect objects behind these clusters. Postman, Geller & Huchra (1986) also find that the number of clusters as a function of Galactic latitude is not independent of richness, with richer clusters at low Galactic latitude being under-represented. The Galactic latitude distributions of richness class $R = 1$ clusters and $R = 3$ clusters are different at the 98% confidence level as evidenced by a KS test. Postman, Geller & Huchra (1986) also examined the Zwicky cluster catalogue and found evidence for large variations in the depth of the plates used to select clusters. They do not detect this in the Abell catalogue (selected from the same POSS plates), probably due to the lower density of objects, but the Abell catalogue is likely to be affected by the same bias. Some authors attempt to correct for the various biases in the Abell catalogues (Sutherland 1988, Sutherland & Efstathiou 1991, Efstathiou *et al.* 1992) and conclude that a more accurate value for the correlation length of clusters is $r_0 \sim 14\ h^{-1}$Mpc.

3.1.3 Power spectrum analysis of cluster catalogues

Although the correlation function and the power spectrum form a Fourier transform pair and should thus be equally useful, the measured correlation function from a finite and noisy sample is often less informative than the power spectrum. The uncertainty caused by the fact that we must estimate the mean number density of objects from the sample itself is more problematic for the measurement of the correlation function. This uncertainty affects the measurement of $\xi(r)$ at every value of r, causing the correlation function to shift up and down. Since the value of the correlation function is very small at large wavelengths, it becomes very difficult to measure $\xi(r)$ beyond a few tens of h^{-1}Mpc. Hamilton (1993b) gives a discussion of this problem and suggests alternative correlation function estimators which reduce the effect of the uncertainty in the mean number density. The power spectrum is, in principle, less affected by the uncertainty in the value of the mean number density (see Peacock & Nicholson 1991 for discussion of this point). If there were no window function, the uncertainty would only affect the $k = 0$ mode of the power spectrum. Even when we take account of the correlations introduced by the window function, the uncertainty in $\overline{n}$ only affects a few of the small wave number modes in the power spectrum analysis (see Chapter 2 and Appendices A and B). It is also more convenient to have the measured power spectrum of a sample when one compares with the predictions of various cosmological models as this is the quantity which naturally results from the theoretical framework.

The power spectrum of the cluster distribution has been measured by a number of authors, although there are no measurements to date of power spectra from the automated surveys except that presented below. Peacock & West (1992) measured the power spectrum of galaxies in a sample of Abell clusters and they detected power on scales $r > 100\ h^{-1}$Mpc. The amount of power detected was less than that which would be expected if the power law on small scales were extrapolated to the largest scales measured. This provides support for the galaxy power spectrum having a break on wavelengths $\gtrsim 100\ h^{-1}$Mpc. The power spectrum of clusters has also been measured by Einasto *et al.* (1993) and Gramman & Einasto (1992).

3.1.4 New generation cluster surveys

Many of the problems associated with the old cluster catalogues have been overcome by the new generation of cluster samples which have been objectively selected by automatic plate measuring machines from the photographic plates (Dalton *et al.* 1992, Dalton 1992, Lumsden *et al.* 1992) or from samples selected on the basis of X-ray emission (Romer *et al.* 1994, Nichol, Briel & Henry 1994). X-ray surveys of clusters of galaxies should be free of projection type effects because the X-ray emission is concentrated in the core of a cluster and thus the area subtended on the sky is small.

Results on the measurement of the correlation function from these new generation surveys are in good agreement and are well approximated by a power law $\xi(r) \sim (r/r_0)^{-2}$

with r_0 in the range $\sim 12 - 16\ h^{-1}$Mpc. In this chapter I analyse the APM cluster redshift survey (Dalton *et al.* 1994, see Section 3.2 for a description of this survey). The cluster correlation function measured from this survey is

$$\xi_{cc} \sim \left(\frac{r}{r_0}\right)^{-2.0} \tag{3.4}$$

with $r_0 = 14.3 \pm 2.35\ h^{-1}$Mpc (Dalton *et al.* 1992). The space density of clusters in the APM cluster survey is $\overline{n} \simeq 2.4 \times 10^{-5} h^3 Mpc^{-3}$ which is about four times the space density of $R \geq 1$ Abell clusters. Bahcall & West (1992) argue that this result fits onto the relation between correlation length and number density (equation 3.2), as does the result from the Edinburgh-Durham automated survey of clusters (Nichol *et al.* 1992) which gives $r_0 = 16.4 \pm 4\ h^{-1}$Mpc for clusters with a number density of $\overline{n} \simeq 1.3 \times 10^{-5} h^3 Mpc^{-3}$. However, the error bars on the values of r_0 also allow no dependence of clustering amplitude with richness (see Figure 6.4 of Efstathiou 1996 where, in fact, in can be seen that the results from the new cluster redshift surveys are more consistent with no dependence). All support for equation (3.2) comes from analysis of the Abell cluster catalogue which has been shown to suffer seriously from selection biases which affect the measurement of large scale structure in the catalogue.

The new automated surveys described above have been shown to be free of anisotropies caused by selection biases (Dalton *et al.* 1992, Nichol *et al.* 1992, Romer *et al.* 1994). Efstathiou *et al.* (1992), for example, show plots of the correlation function of clusters as a function of separation perpendicular and parallel to the line of sight, $\xi(\sigma, \pi)$, for the APM cluster catalogue. The small anisotropies in $\xi(\sigma, \pi)$ are consistent with a Gaussian convolution with dispersion ~ 700km s^{-1}. The measurement error on the velocity of a cluster is $\sim 300 - 400$km s^{-1}, giving an expected dispersion of $\sim \sqrt{2} \times 350$km s^{-1} ~ 500km s^{-1}. This is consistent with the observed anisotropies being caused by the measurement errors alone as the error on the value of 700km s^{-1} is large. Nichol *et al.* (1992) also show a $\xi(\sigma, \pi)$ plot for their cluster sample and again it is free of anisotropies.

In summary, the new X-ray selected and automated cluster catalogues, which seem to be free of selection inhomogeneities, are consistently showing a correlation length of $r_0 \sim 12 - 16\ h^{-1}$Mpc. The results of analyses using the Abell catalogue (especially the richer clusters) should be regarded as unreliable until there is an independent check.

In the rest of this chapter I present the power spectrum analysis of the APM rich cluster redshift survey. In Section 3.2 I describe the sample used and the simulated surveys extracted from N-body simulations which are used for comparison. In Section 3.3 I review the method of power spectrum analysis and test the performance of the power spectrum estimator. In Section 3.4 I present the power spectrum results for the APM cluster survey and I examine the sensitivity of the power spectrum to the weighting scheme and the selection function. The results from the APM cluster survey are compared with the predictions of various cosmological models in Section 3.4.1. In Section 3.4.2 I discuss the comparison of the clustering of galaxies and the clustering of clusters, and I summarize in Section 3.5.

3.2 Survey data and cosmological models

3.2.1 The APM rich cluster survey

The APM cluster redshift survey was constructed in two parts. The first is described in Dalton *et al.* (1992) and Dalton (1992). To summarize, the clusters were selected from the APM galaxy survey (Maddox *et al.* 1990a, Maddox, Efstathiou & Sutherland 1990b, Maddox *et al.* 1990c). Cluster candidate centres were identified by applying a percolation algorithm to the APM galaxy catalogue, limited at $b_J = 20.5$, to link galaxies with angular separations less than 0.7 times the mean separation. Groups with more than 20 members were chosen as cluster candidate centres. For each candidate cluster a characteristic magnitude m_X and richness $\mathcal{R}$ (not the same as Abell's richness) were defined via an iterative procedure, as follows. The starting value of the redshift of the cluster was taken as 0.1 and a starting richness of $\mathcal{R}$ of 20 was chosen. Then, around each cluster candidate centre a circle of metric radius 0.75 h^{-1}Mpc was defined. Galaxies within this circle were counted after having been ordered in apparent magnitude and each galaxy was weighted in the count by $(1 + 2r_p/r_c)^{-1}$ where r_p is the projected distance of the galaxy from the cluster centre and r_c is the metric radius i.e. 0.75 h^{-1}Mpc. The count was stopped when it reached a value of $\mathcal{R}/2$ above the mean background count. The value of the apparent magnitude of the galaxy at this point in the count was defined as the characteristic magnitude m_X for that cluster. The richness $\mathcal{R}$ is the weighted galaxy count in the magnitude range $(m_X - 0.5, m_X + 1.0)$ above the mean background count in the magnitude range $(m_X - 0.5, m_X + 1.5)$. Any cluster whose centre fell within the counting circle of a richer cluster was deleted and the process was repeated, starting with more accurate cluster centres. The procedure typically converged to a stable list of clusters in $3 - 5$ iterations.

The APM cluster selection algorithm has several advantages over that used by Abell. Firstly, it is objective rather than subjective. Also the variations in plate depth between different plates are removed by careful matching of the plates. The search radius used ($0.75\ h^{-1}$Mpc) is half as big as that used by Abell and thus alleviates the projection effects which may have contributed to the anisotropies in the Abell catalogues. In fact the effective search radius is even smaller since galaxies near the centre of the cluster are given more weight than those further out. Also, the characteristic magnitude m_X used for APM clusters should be almost independent of richness unlike m_{10} used by Abell, which decreases with increasing cluster richness (the 'Scott effect', Scott 1957). The above procedure produced a sample of 240 clusters with richness $\mathcal{P} > 20$ and $z_x < 0.1$ (photometric redshift). Out of these, 211 comprise the final sample of clusters with spectroscopically measured redshifts. The richness of the APM cluster sample is estimated to correspond to Abell richness class $R = 0$ (Dalton *et al.* 1992). The cluster velocities are measured to an accuracy of 512km s^{-1}.

This catalogue was later supplemented by an extension of the survey (Dalton *et al.* 1994). This survey was designed to have a greater depth than its predecessor, which was limited by the magnitude range used to define the richness of the clusters. The catalogue described above is limited at the point where $m_X + 1.5$ becomes equal to the magnitude limit of the plates, i.e. when $m_X = 19.0$. For the selection of the second sample, the magnitude slice for which the background count was calculated was changed to $(m_X - 0.5, m_X + 1.0)$ and X was changed from $X = \mathcal{P}/2$ to $X = \mathcal{P}/2.1$. The clusters from the old catalogue had richnesses re-calculated according to

$$\mathcal{P} = 24.8 + 1.2\mathcal{P}_{old} \tag{3.5}$$

where $\mathcal{P}_{old}$ is the richness as defined in the old cluster catalogue. This relation was found to hold with a dispersion of 16 in $\mathcal{P}$. All clusters with $\mathcal{P} > 50$ were selected and redshifts were measured for those which were not already known (about 223 clusters). The final sample, which is the one analysed in this chapter consists of 364 clusters and the sample mean density is $3.4 \times 10^{-5} h^3 Mpc^{-3}$. Again this sample was shown to be free of selection biases in a plot of $\xi(\sigma, \pi)$ (Dalton *et al.* 1994). The median redshift of clusters in this survey is ~ 0.09, and there is useful information out to velocities $\sim$ 55000km s^{-1}.

Dalton *et al.* (1994) present the results of a correlation function analysis of this sample and find $r_0 = 14.3 \pm 1.75\ h^{-1}$Mpc ($2\sigma$ errors) which is consistent with the result from the earlier sample (Dalton *et al.* 1992). This was shown to be inconsistent with SCDM but consistent with LCDM (with $\Gamma = \Omega h = 0.2$) and an MDM model.

3.2.2 Simulated APM cluster surveys

Simulated APM cluster surveys from LCDM, SCDM and MDM models are used in this chapter both to test the reliability of the power spectrum measurement and to compare against the predictions for $P(k)$ for these cosmological models. The simulated surveys were constructed by Rupert Croft and the details of the simulations and the method for creating mock APM cluster surveys are described in Croft & Efstathiou (1994a) and Croft & Efstathiou (1994b). The mock surveys are designed to cover the same volume as the real APM cluster survey (the same sky-mask is used), and the clusters in the simulated catalogues have a density similar to that of the APM clusters themselves. Cluster centres are selected by using a friends of friends algorithm to link particles with separation less than 0.1 times the mean separation. The cluster centres are then used as the centres of spheres of size $r_c = 0.5\ h^{-1}$Mpc within which the mass of the cluster is calculated. The centre of mass of the cluster was found and any cluster which was less than r_c from the centre of another, richer cluster was deleted. The procedure was then repeated. The list of clusters was ordered by mass and a lower mass cut off was applied to obtain a cluster sample of a chosen mean number density. The APM clusters selection function was also imposed on the cluster catalogues.

Croft & Efstathiou (1994a) describe tests of the cluster selection algorithm and find that the resulting cluster-cluster correlation function is insensitive to the algorithm used. The LCDM, SCDM and MDM cosmological N-body simulations that the catalogues were selected from are those described in Chapter 2. For each of the cosmological models we have several realizations of mock APM cluster surveys (20 each for the two CDM type universes and 8 for the MDM universe).

3.3 Power spectrum analysis

Since the cluster redshift survey has a radially varying selection function (i.e. it is not accurately volume limited), we must treat it as we would a flux limited survey. I have applied the formalism of Feldman, Kaiser & Peacock (1994) exactly as implemented in Chapter 2 to find the power spectrum and the errors on each point. The radial selection function is obtained by smoothing the velocity histogram of the data with a Gaussian. Figure 3.1 shows the histogram of the velocities of APM clusters together with the selection function obtained by convolving with a Gaussian of width 4000km s^{-1}. Sensitivity of the

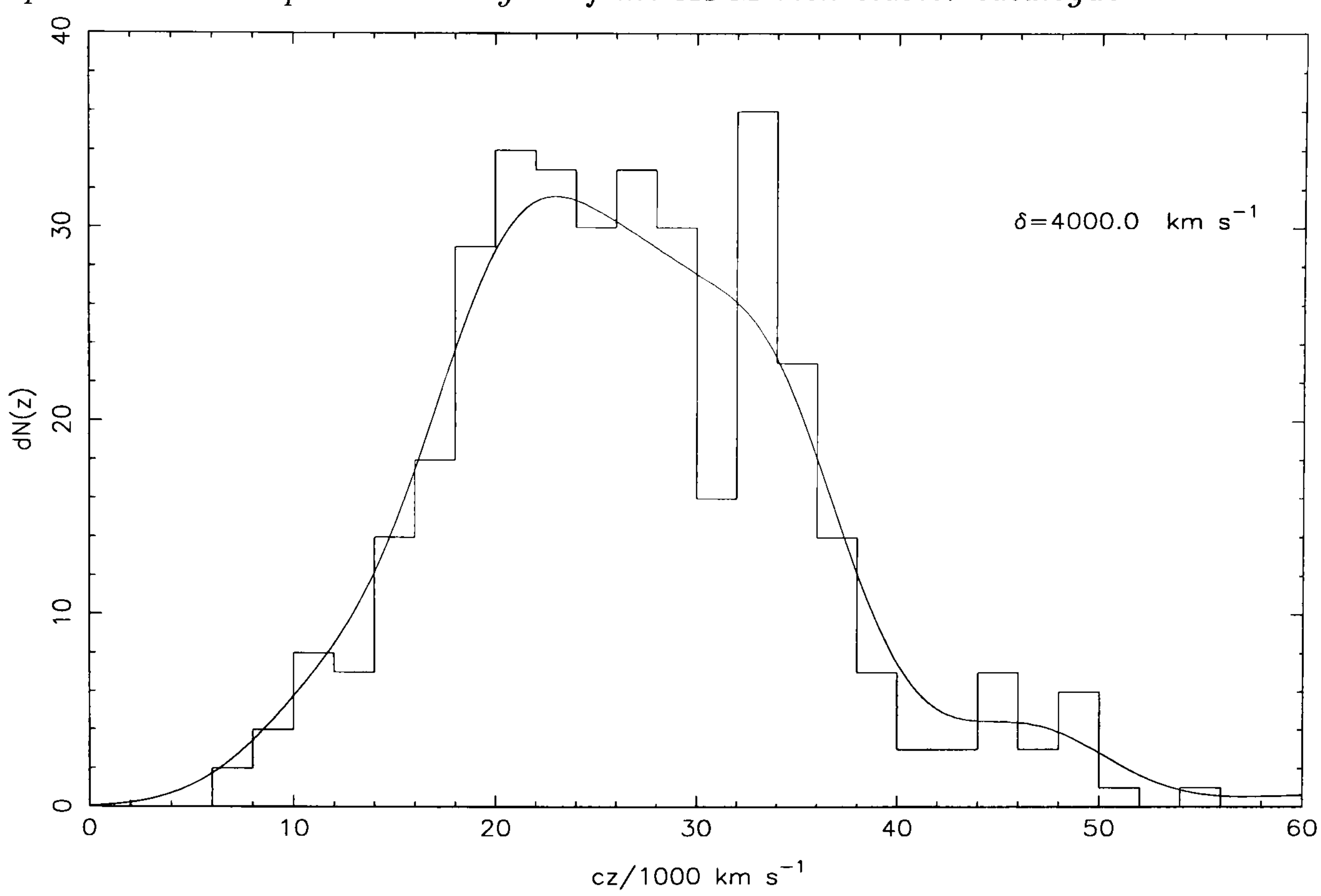


Figure 3.1: Figure shows a histogram of the velocities of the clusters in the APM rich cluster survey. The solid line shows the histogram convolved with a Gaussian of width 4000km s^{-1}.

power spectrum to the assumed selection function is discussed in Section 3.4.

To test the accuracy and precision of the estimation of the power spectrum, I have computed the power spectrum of the simulated cluster catalogues described above. I have also computed for each model the power spectrum of clusters in the entire N-body box, i.e. with no window function or selection function. We can test how well we can expect to recover the power spectrum by comparing these two. All catalogues are in redshift-space. Again, as in Chapter 2, we must choose a value of $P_w(k)$ in the weighting function $w(r) = 1/(1+\overline{n}P_w(k))$. I have chosen weights with $P_w(k) = 50000\,(h^{-1}\mathrm{Mpc})^3$, $16000\,(h^{-1}\mathrm{Mpc})^3$ and $40000\,(h^{-1}\mathrm{Mpc})^3$ for the LCDM, SCDM and MDM models respectively. The recovered power spectrum is relatively insensitive to the exact value of $P_w(k)$ used in the weighting scheme. The resolution with which we measure the power spectra of clusters in the full box is different to the resolution used for the simulated cluster catalogues. The full box is of side 300 h^{-1}Mpc and a 64^3 grid was used to do the fast Fourier transform analysis. The simulated APM clusters, however, were embedded in a box of side 1600 h^{-1}Mpc and a 256^3 grid (as large as is computationally feasible) was used. This means that the resolution is better when measuring the power spectra of the clusters in the full box. The result of this is that the power spectra of the simulated cluster catalogues will be slightly underestimated relative to the power spectra of the full box of clusters at high wave numbers. It would be fairly easy to overcome this problem by using a 'slow Fourier transform' i.e. an explicit

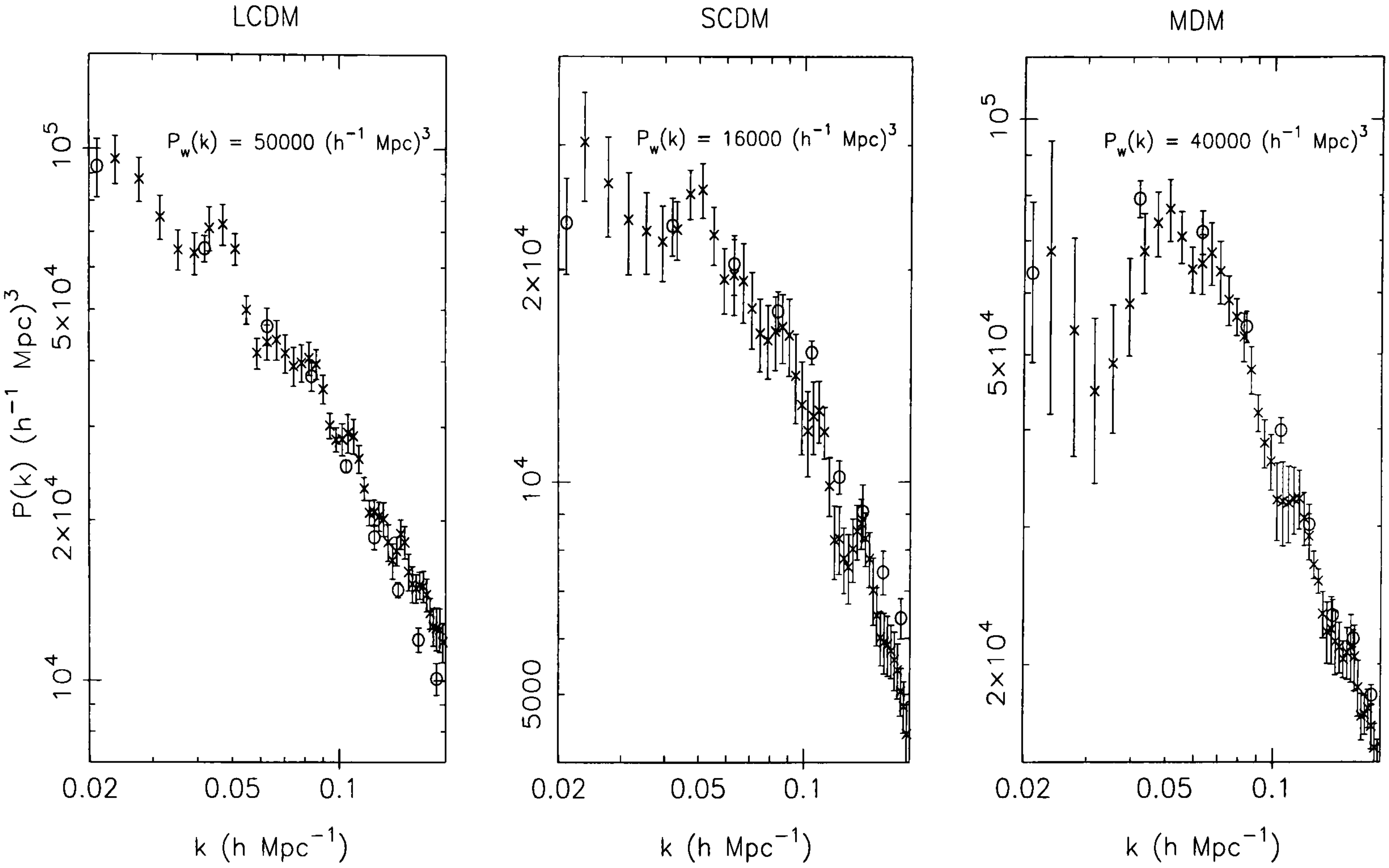


Figure 3.2: Estimates of the redshift-space power spectra of simulated cluster surveys compared to the true redshift-space power spectra of the clusters. In each case, circles show the power spectra obtained by averaging over ten realizations of the distribution of clusters in the whole N-body simulation, while crosses show the power spectra obtained by averaging over 20 realizations of simulated APM cluster surveys, with the same selection function and sky coverage as the real APM cluster survey described in the text. The three panels show results from LCDM, SCDM and MDM respectively and in the case of the simulated surveys, the power $P_w(k)$ used in the weighting function was 50000, 16000 and 40000 $(h^{-1}\mathrm{Mpc})^3$ respectively. Error bars show errors on the mean (note that the MDM simulations only had 8 realizations, hence the larger fluctuations and errors).

sum over the $e^{\mathbf{k}\cdot\mathbf{x}_i}s$ instead of using a fast Fourier transform, but for most of this chapter we will be primarily interested in the small wave number region of the spectrum.

In Figure 3.2 I show, for each of the three cosmological models, a comparison of the power spectrum of clusters measured from the full N-body simulation box (circles) with the power spectrum recovered from simulated APM rich cluster surveys. The power spectra from the whole boxes are averages over 10 realizations in each case, while the power spectra from the mock surveys are averaged over 20 realizations in the cases of LCDM and SCDM, and over 8 realizations for MDM. The error bars show the error on the mean value of $P(k)$. It can be seen that the power spectrum is recovered accurately from the mock APM cluster surveys over the range of wave numbers $0.02 \lesssim k \lesssim 0.1\ h\ \mathrm{Mpc}^{-1}$. We cannot test the power spectrum estimator on scales with wave number $\lesssim 0.02\ h\ \mathrm{Mpc}^{-1}$ with these simulations, as they are performed in a box of size 300 $h^{-1}\mathrm{Mpc}$.

Note that the bias terms discussed in Chapter 2 which cause one to underestimate the power spectrum at small wave numbers are relatively small in the case of this survey. In Figure 3.3 I compare the percentage correction to the power spectrum as a function of wave number k for the APM cluster survey, a flux limited Stromlo-APM survey (weighted with $P_w(k) = 8000\,(h^{-1}\mathrm{Mpc})^3$ and assuming a $\Gamma = 0.2$ CDM model for the underlying power spectrum) and a volume limited Stromlo-APM survey (volume limited at $z_{max} = 0.06$, containing about 450 galaxies, again assuming a $\Gamma = 0.2$ CDM underlying power spectrum). As expected, the predicted corrections are largest for the volume limited Stromlo-APM survey, slightly smaller for the flux limited Stromlo-APM survey and smallest in the case of the APM cluster survey. In the latter case, the percentage correction at $k = 0.02\,h\ \mathrm{Mpc}^{-1}$ is predicted to be only $\sim 15\%$. However, in practice (see Figure 3.2) no correction appears to be necessary and equation (B14) of Appendix B appears to overestimate the bias in the measured power spectrum. This is not so surprising as a number of approximations are involved in the derivations of the corrections in Appendix A and Appendix B, and if the amount of bias is small it is very hard to predict the correction accurately. Small corrections will also be very difficult to detect relative to the larger sampling errors in the power spectrum. It can also be seen that in the cases of the two flux limited surveys the correction becomes negative at high wave numbers. However, for a number of reasons, the bias corrections derived in the appendices are only valid at small wave numbers. For example one of the approximations used in Appendix B was that the two weighting functions $w(r) = 1/(1 + \overline{n}P_w(k))$ and $w'(r) = 1/(1 + 4\pi\overline{n}J_3)$ are equal. There will be some wave number at which this is accurate, and it will be just an approximation for other wave numbers; in particular it will be a bad approximation at higher wave numbers. (Note that this assumption is not made in the case of the correction to volume limited surveys, hence the dot-dash line in Figure 3.3 levels off at zero).

Figure 3.4 shows the average power spectrum measured from the 20 mock APM cluster surveys from the LCDM model, compared to the linear theory power spectrum of the mass distribution (where the linear theory power spectrum has been scaled up by a factor of 4). The errors show the error on the mean over the 20 power spectra multiplied by $\sqrt{20}$ i.e. the error expected in the power spectrum measured from a single simulated catalogue. It can be seen that we are accurately measuring the power spectrum from this simulated catalogue to very large scales, in this case, to wave numbers $k \sim 0.01\,h\ \mathrm{Mpc}^{-1}$.

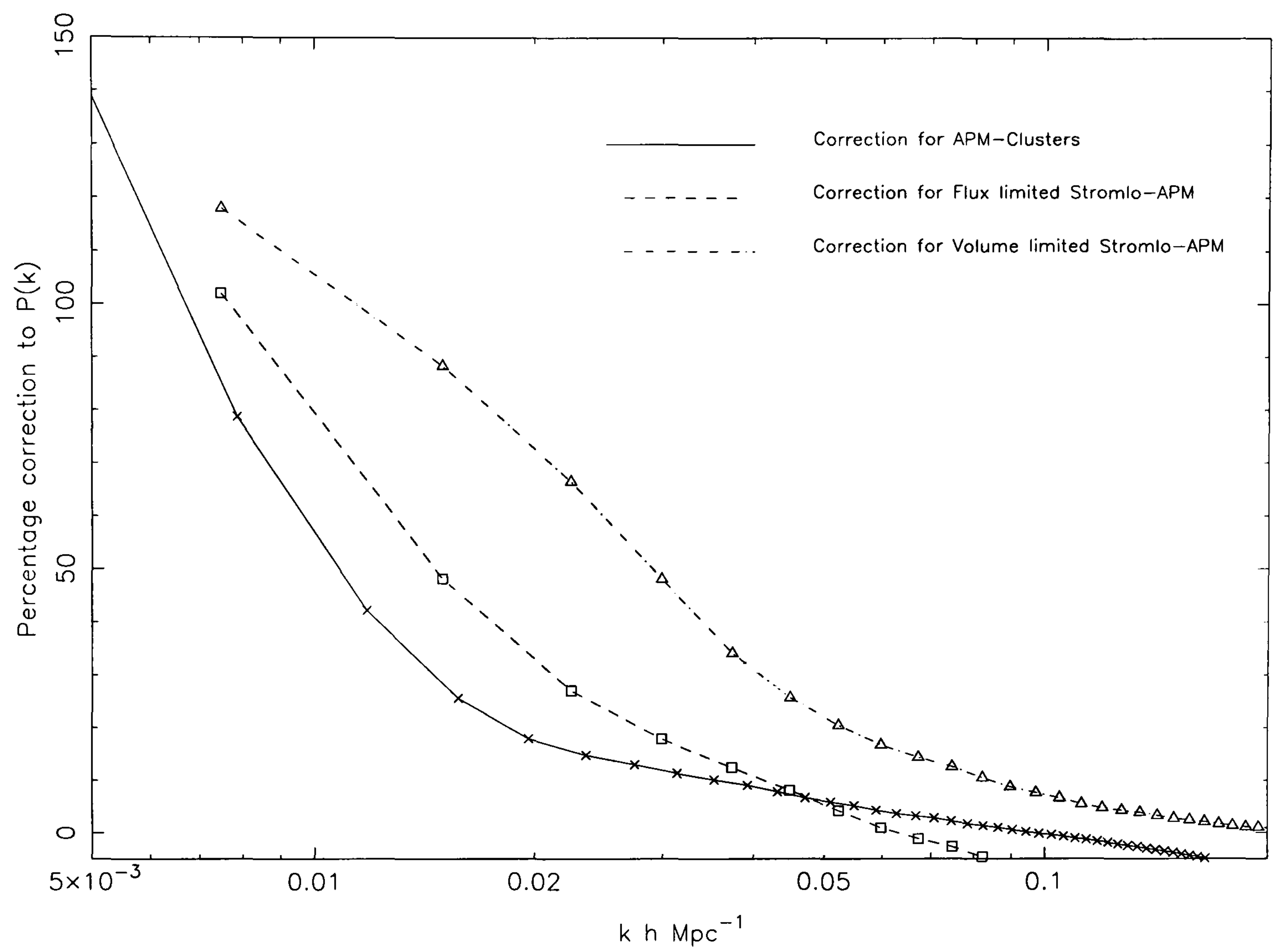


Figure 3.3: This figure illustrates the degree to which the measured power spectrum is biased low at small k due to the fact that we must estimate the mean density of galaxies (or clusters) from the survey which we have. The y axis shows the percentage correction to the power spectrum where the correction is calculated using the equations in Appendix A in the case of volume limited surveys and the equations in Appendix B in the case of flux limited surveys. The solid line show the calculated correction for the APM clusters sample discussed in this chapter, assuming that the true underlying power spectrum is a $\Gamma = 0.2$ CDM-like power spectrum (multiplied by a factor of 4) and using $P_w(k) = 50000\ (h^{-1}\text{Mpc})^3$ in the weighting function. The dashed line shows the calculated correction to a flux limited Stromlo-APM survey, again assuming an underlying $\Gamma = 0.2$ CDM-like power spectrum and using $P_w(k) = 8000\ (h^{-1}\text{Mpc})^3$ in the weighting function. The dot-dash line shows the same thing for a volume limited Stromlo-APM survey ($z_{max} = 0.06$). The number of galaxies in such a survey is taken to be ~ 450, as in the real data (the correction term depends on the number of galaxies - see Appendix A), and again the $\Gamma = 0.2$ CDM spectrum is assumed. The corrections, as expected, are largest for the volume limited Stromlo-APM survey, and smallest for the cluster survey. The smaller the true correction, the more difficult it is to accurately predict what the correction should be. In practice it appears that the correction to the cluster power spectrum is overestimated by the calculations in Appendix B.

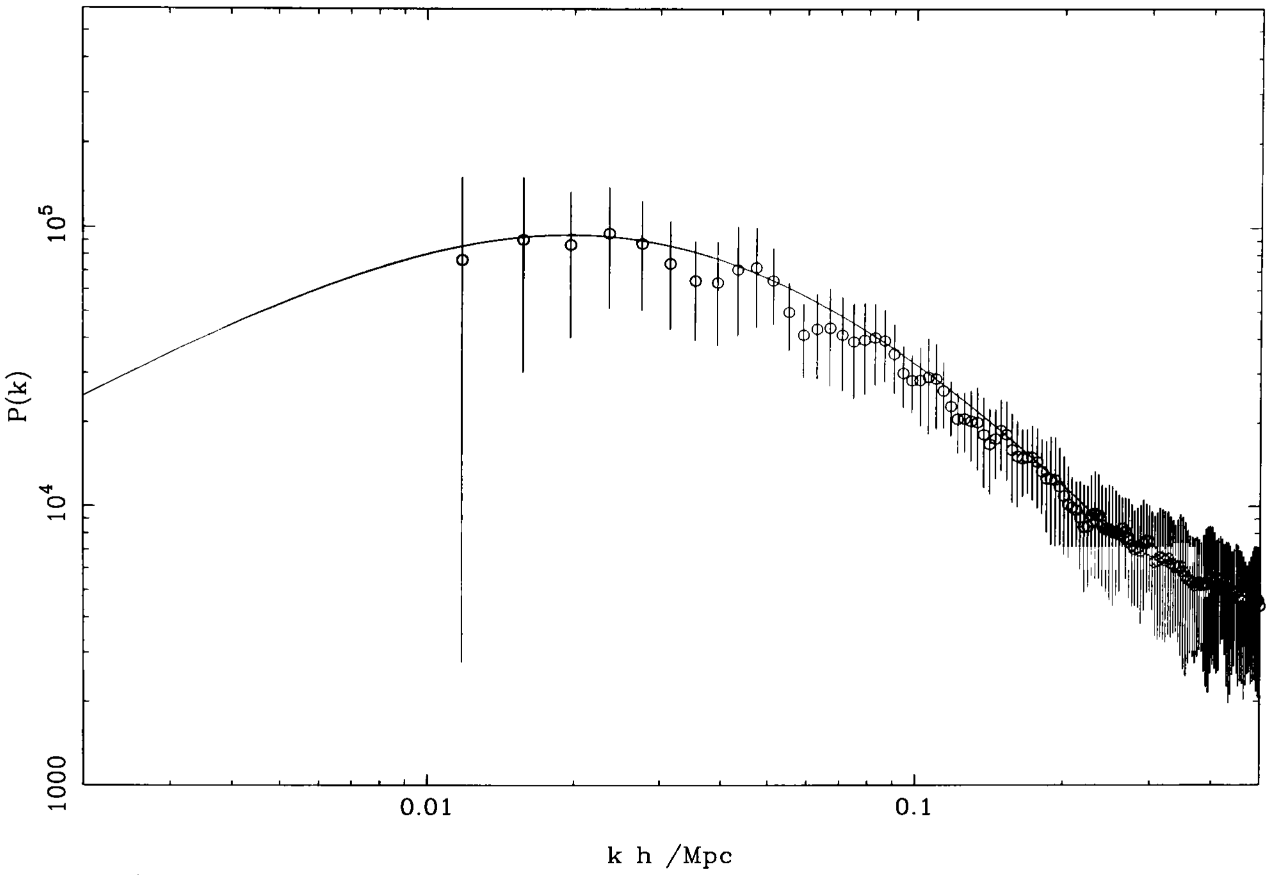


Figure 3.4: This figure illustrates the range in wave number over which we can measure the power spectrum of clusters using a survey like the APM cluster survey. The circles show the power spectrum (again averaged over 20 realizations) of the simulated APM cluster surveys taken from the LCDM model. The solid line shows the linear theory power spectrum appropriate to this model, with normalization such that $\sigma_8 = 1$ and scaled up by a factor of 4. The error bars show the error on the mean multiplied by $\sqrt{20}$ and therefore give an idea of the error expected on $P(k)$ from one realization of the APM cluster survey.

3.4 Power spectrum results from the APM rich cluster survey

In this section I present the results from the APM cluster survey. The power spectrum analysis is carried out as described above and in Chapter 2, with the survey embedded in a box of size 1600 h^{-1}Mpc, and a Fourier transform grid of size 256^3 is used. From the analysis of the previous section, we see that the results are expected to be accurate at least to wave numbers $k \sim 0.02$ h Mpc^{-1}. The bias caused by estimating the mean density from the survey itself (see Appendix B and discussion above) is also expected to be negligibly small. The results from the power spectrum analysis of the APM cluster survey are presented in Figure 3.5 where I have used $P_w(k) = 50000\,(h^{-1}\text{Mpc})^3$ in the weighting function (equation 2.28 of Chapter 2). Error bars are calculated in the same way as in Chapter 2 following the analysis of Feldman, Kaiser & Peacock (1994), who provide an analytic expression, derived under the assumption that the power distribution in the k modes follows a Gaussian distribution. The power spectrum of APM clusters is consistent with a power law in k of index -1.6 on scales $k \gtrsim 0.04$ h Mpc^{-1}. On larger

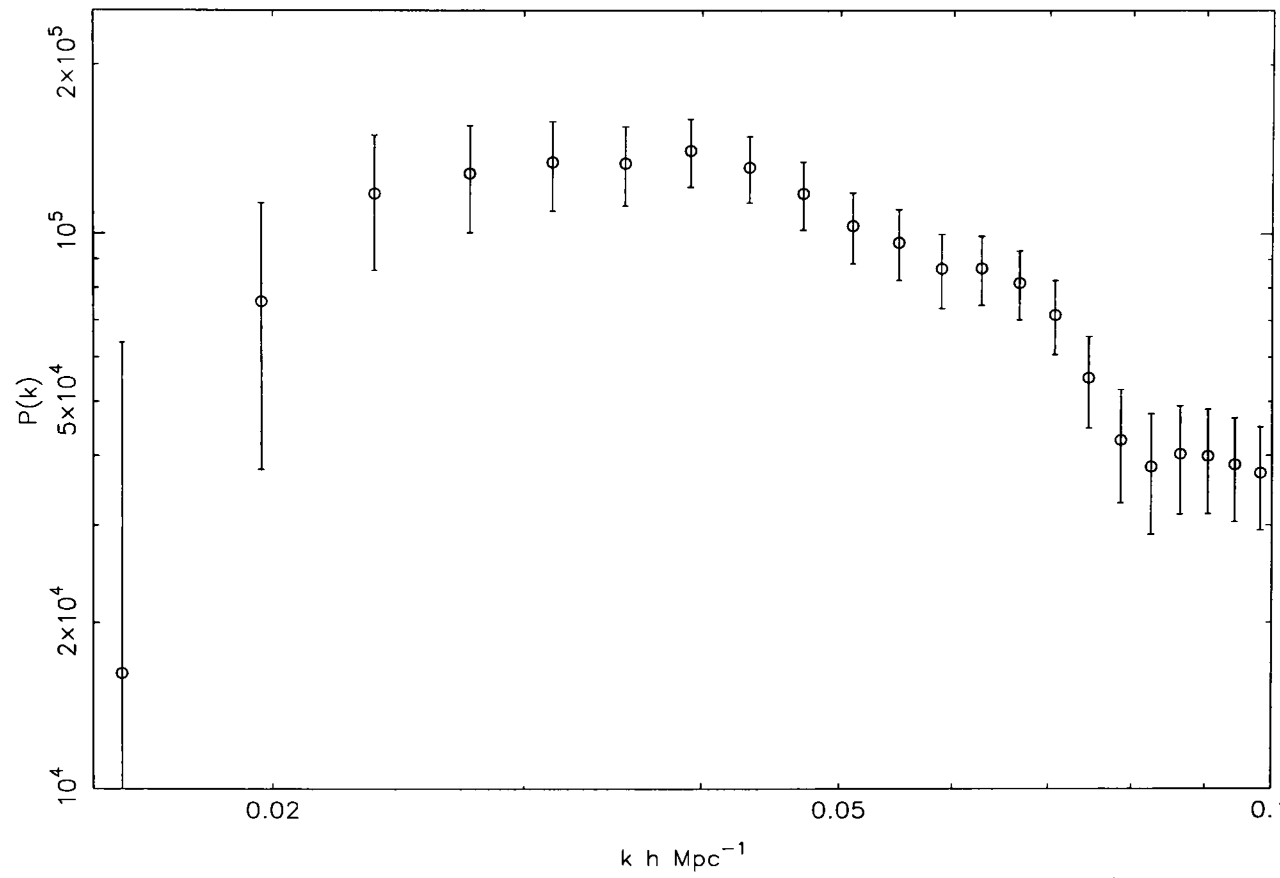


Figure 3.5: The power spectrum of APM clusters. The plot shows the power spectrum as a function of wave number calculated from the data by embedding the survey in a Fourier transform box of size 1600 h^{-1}Mpc. The value of $P_w(k)$ in the weighting function was 50000 $(h^{-1}\mathrm{Mpc})^3$ and the selection function was obtained by smoothing the velocity histogram with a Gaussian of width 4000km s^{-1}. All clusters with velocity less than 45000km s^{-1} are used in this analysis. The power spectrum is accurately measured to wave numbers at least as small as $k = 0.02\,h$ Mpc^{-1}.

scales than this, the power law flattens and the power spectrum turns over at a wave number $k \sim 0.024\,h$ Mpc^{-1}. From the test carried out above on simulated surveys we have reason to believe that this turn-over is real and not an artifact of the way in which the power spectrum analysis is performed.

Figure 3.6 shows the sensitivity of the power spectrum to different values of $P_w(k)$ in the weighting function (equation 2.28 of Chapter 2). The circles show the power spectrum using $P_w(k) = 16000\,(h^{-1}\mathrm{Mpc})^3$ while the crosses show the power spectrum if we use $P_w(k) = 92000\,(h^{-1}\mathrm{Mpc})^3$ in the weighting function. The squares show the resulting power spectrum when a variable weight is using in the weighting function. In this case, each wave number is weighted separately with a power appropriate to a $\Gamma = 0.2$ CDM linear theory power spectrum, again multiplied by a factor of 4. It can be seen that the value of $P_w(k)$ in the weighting function is not crucial as long as a sensible value is chosen. This is in agreement with what we found in Chapter 2 when analysing the flux limited simulated and real Stromlo-APM surveys. Indeed we would not expect the result of a minimum variance estimate to be sensitive to small deviations from the minimum. Hereafter, unless otherwise stated, the power spectrum results from the APM cluster survey are calculated

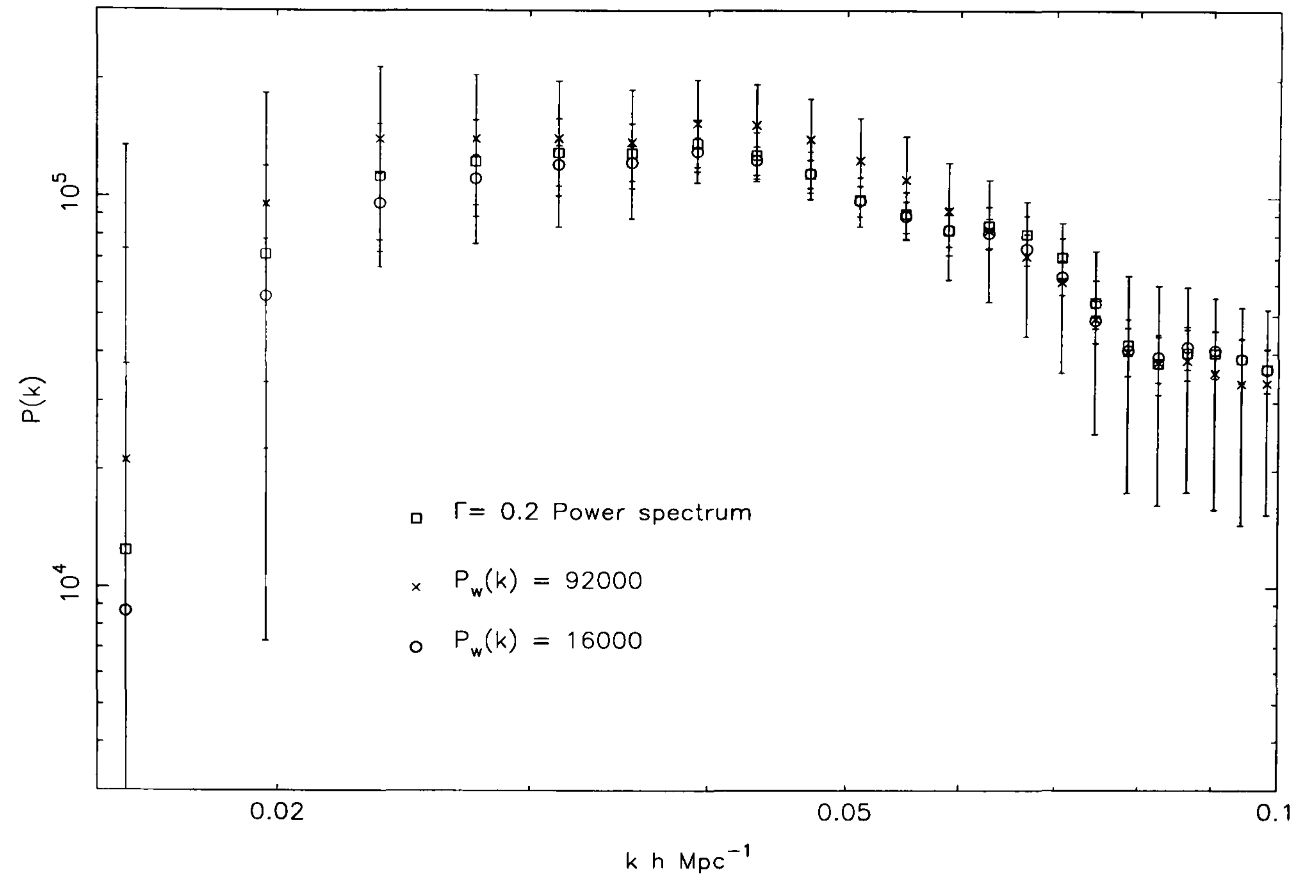


Figure 3.6: Figure shows the sensitivity of the measured power spectrum of APM clusters to the value of $P_w(k)$ used in the weighting function.

using $P_w(k) = 50000\ (h^{-1}\text{Mpc})^3$ in the weighting function.

The measured power spectrum is however slightly sensitive to the value of the smoothing velocity used to derive the selection function. As described above we obtain the selection function by taking the histogram of the velocities of the clusters and smoothing it with a Gaussian width δ km s^{-1}. The resulting power spectrum depends somewhat on the value of δ used. Figure 3.7 shows the power spectrum of APM clusters for 3 different values of the smoothing velocity: 3000km s^{-1} and 5000km s^{-1}, which span the nominal value of 4000km s^{-1} used for the power spectrum in this chapter, and a value 10000km s^{-1}. For wave numbers $k \gtrsim 0.025\ h$ Mpc^{-1} the power spectrum is insensitive to the value of the smoothing velocity to within the calculated error bars. For smaller wave numbers, the power spectrum calculated using a smoothing velocity of 10000km s^{-1} is significantly higher than those calculated using smoothing velocities of 3000km s^{-1} and 5000km s^{-1}. Most of the sensitivity to the selection function is caused by a bad fit to the tail of high redshift clusters beyond a velocity of 40000km s^{-1} when using a large smoothing velocity. If we limit the analysis to clusters with a velocity less than 40000km s^{-1} then the power spectrum is independent of the smoothing velocity to within the calculated error bars even for δ = 10000km s^{-1}, as illustrated in Figure 3.8. Hereafter, unless otherwise stated, the power spectra of the APM cluster survey are calculated using a smoothing velocity of 4000km s^{-1} and including all clusters with velocity less than 45000km s^{-1}.

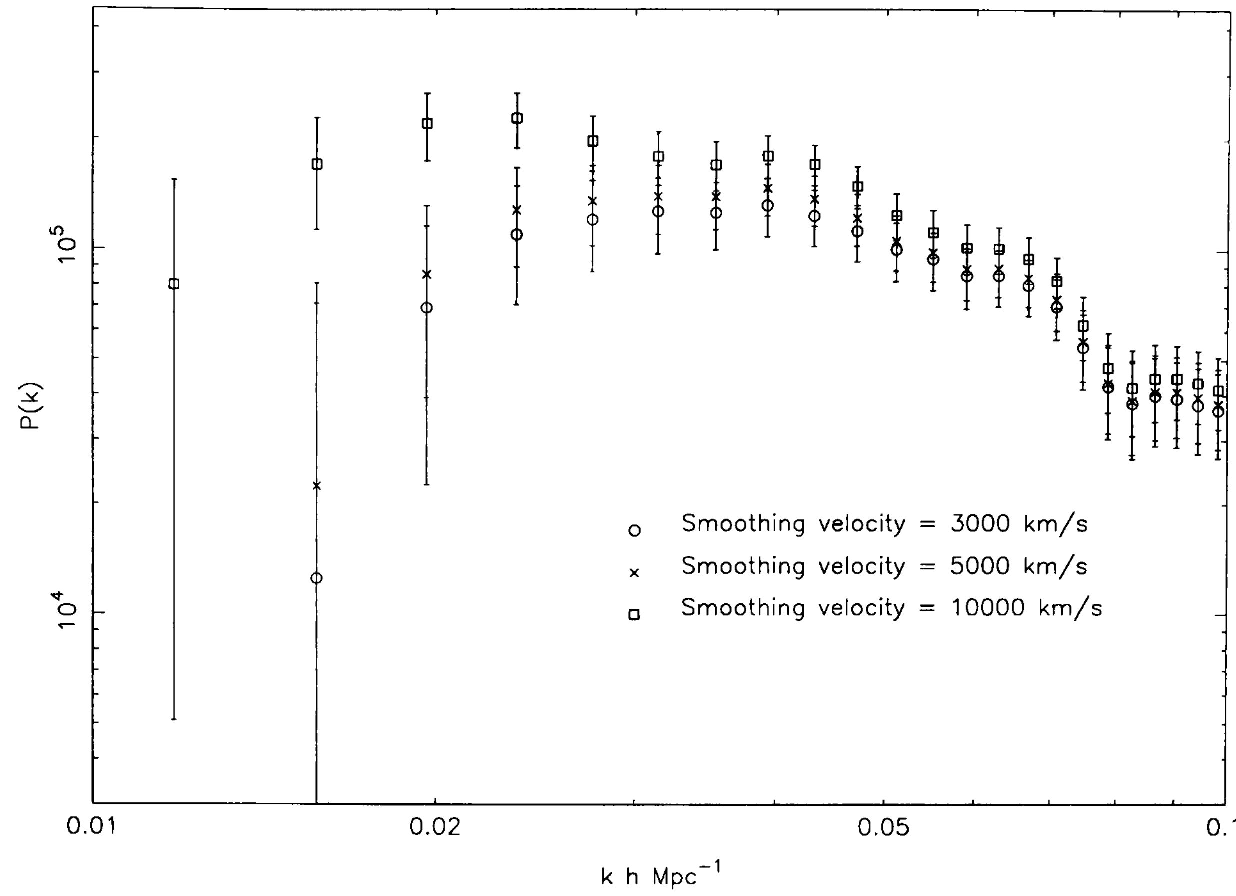


Figure 3.7: This figure illustrates the sensitivity of the measured power spectrum to the selection function. The selection function is obtained by smoothing the velocity histogram of the data with a Gaussian of a certain velocity width. The figure shows the measured power spectrum for three values of this velocity, 3000km s^{-1} and 5000km s^{-1}, which span the actual velocity used for the final cluster power spectrum (4000km s^{-1}), and 10000km s^{-1}. The power spectrum measured is stable to the selection function for wave numbers $k \gtrsim 0.025\,h$ Mpc^{-1}.

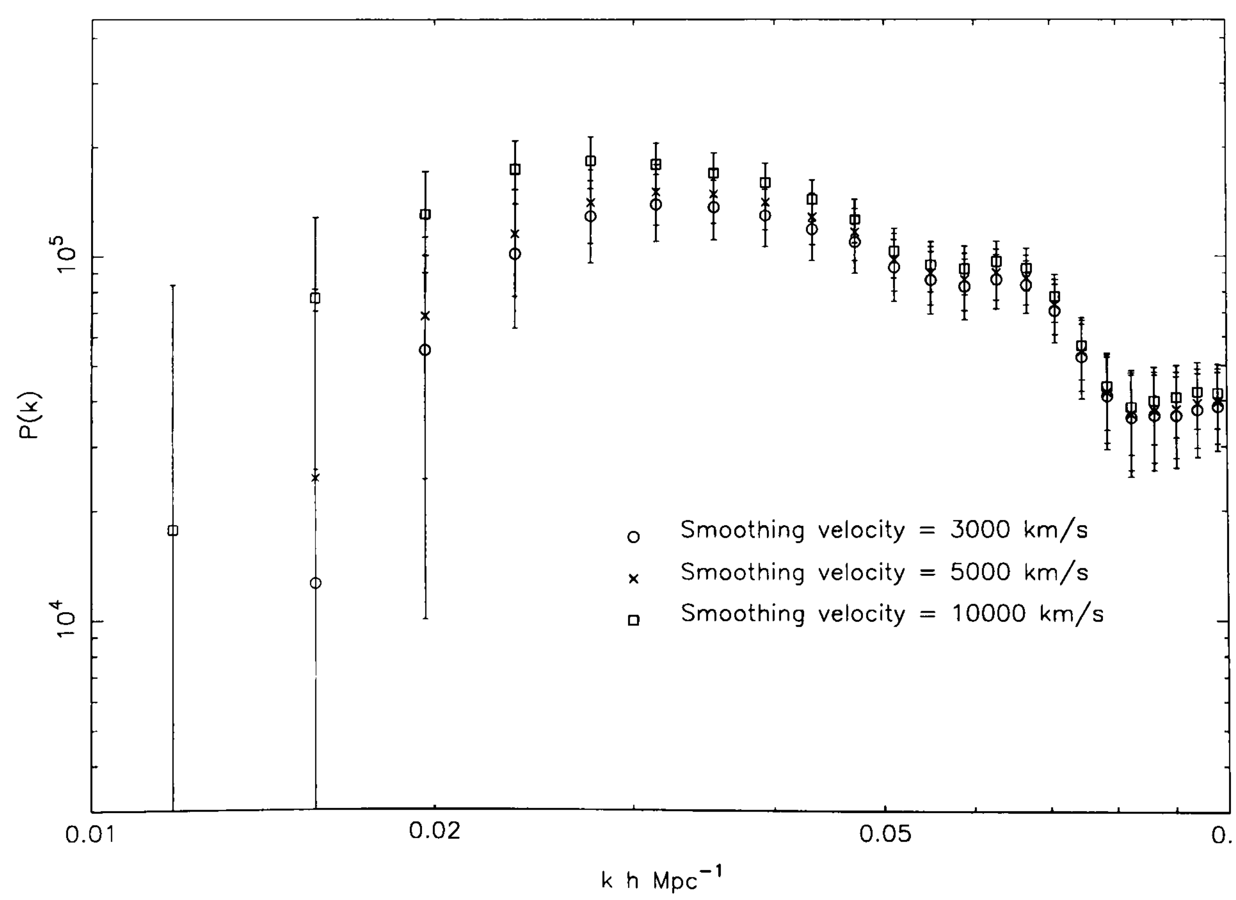


Figure 3.8: As Figure 3.7 but limiting the analysis to clusters with velocities less than 40000km s^{-1}.

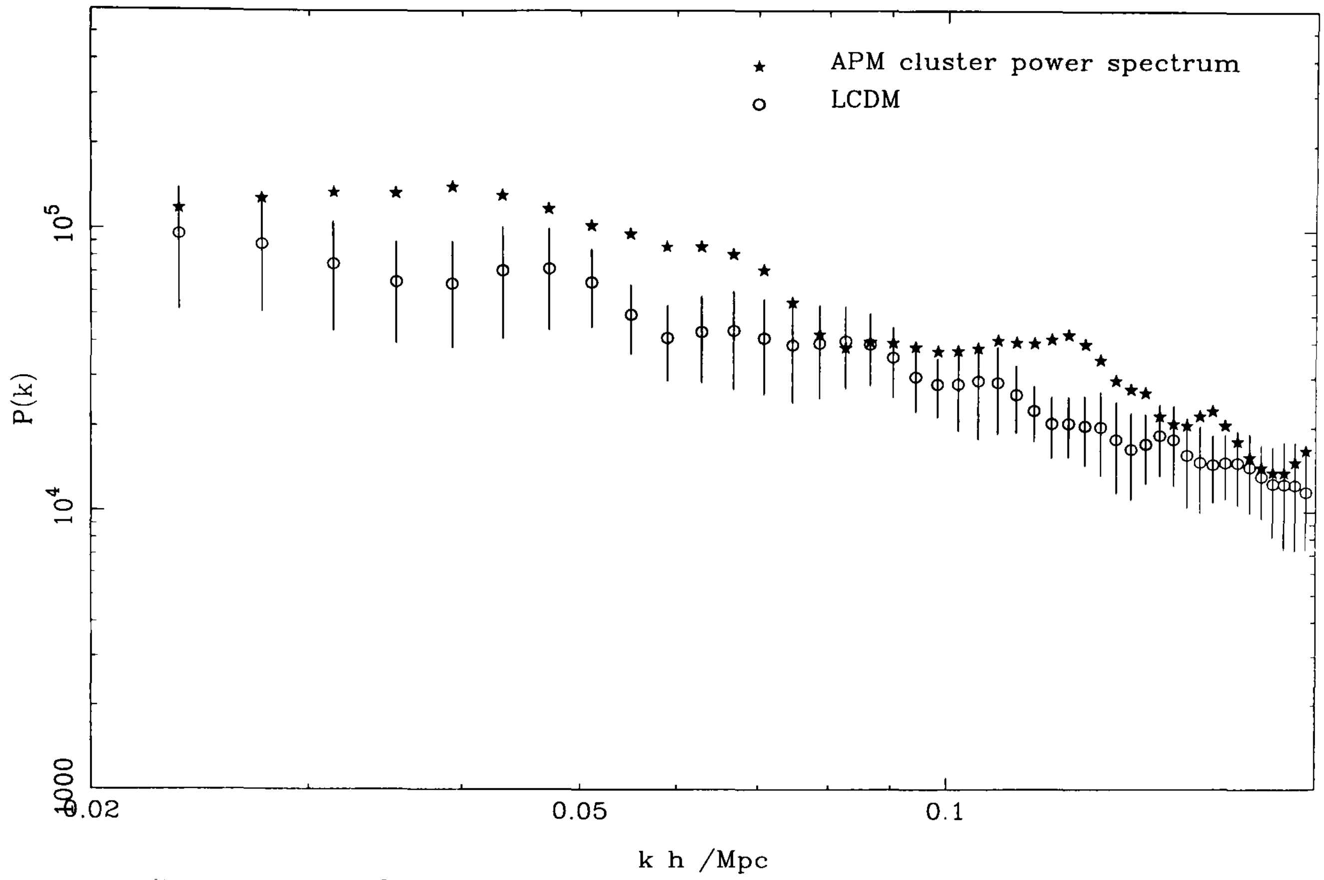


Figure 3.9: Comparison of the power spectrum of APM clusters (stars) with the prediction of the LCDM model (circles). The LCDM power spectrum is obtained by averaging over 20 realizations of simulated APM cluster surveys. The errors show the error on the mean multiplied by $\sqrt{20}$ and thus are appropriate to the power spectrum measured from a single simulated survey.

3.4.1 Comparison with different cosmological models

In this section I compare the power spectrum from the APM cluster survey (Figure 3.5) with the expected results using the simulated cluster surveys for the 3 cosmological models described in Section 3.2.2. Figures 3.9, 3.10 and 3.11 show the power spectra from the simulated APM cluster surveys for the LCDM, SCDM and MDM models respectively, averaged over 20 realizations of each model (except for the MDM model which has 8 realizations), compared with the true APM cluster power spectrum. In each case the model power spectrum is shown by the circles and the real power spectrum is shown by stars. Error bars show the 1σ error appropriate to one simulated APM cluster survey. The error bars on the real APM cluster survey have been suppressed. The value of $P_w(k)$ used for each model is the same as in Figure 3.2, i.e. $P_w(k) = 50000\,(h^{-1}\mathrm{Mpc})^3$ for LCDM, $P_w(k) = 16000\,(h^{-1}\mathrm{Mpc})^3$ for SCDM and $P_w(k) = 40000\,(h^{-1}\mathrm{Mpc})^3$ for MDM. Figure 3.10 shows that the SCDM model fails to match the real data, the difference on all scales, being at least 3σ (with the discrepancy at some wave numbers being $> 9\sigma$). The LCDM model and the MDM model both provide a better match to the data. Each however is still marginally inconsistent with the power spectrum of the APM clusters on very large scales.

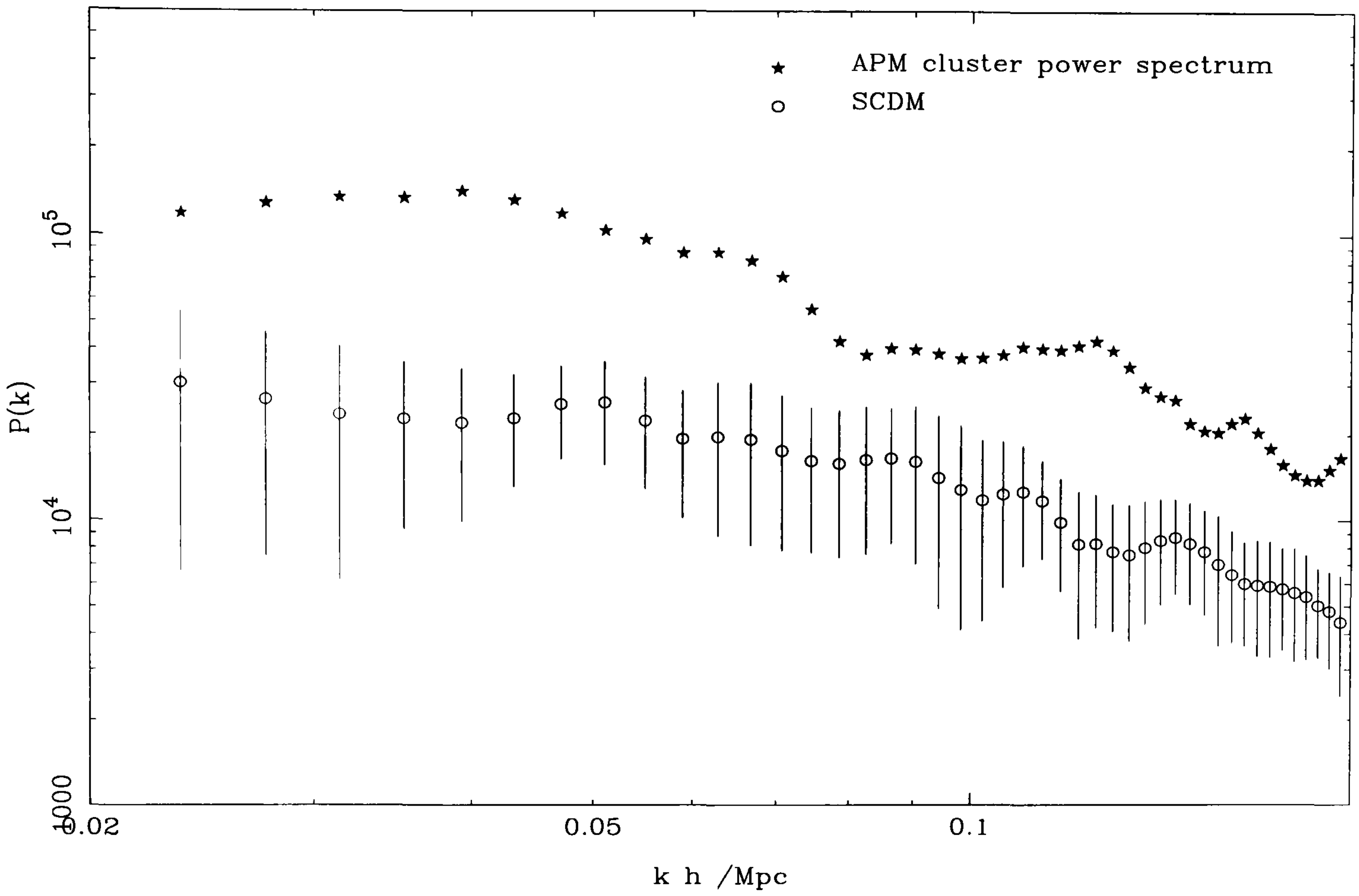


Figure 3.10: As Figure 3.9 for the SCDM model.

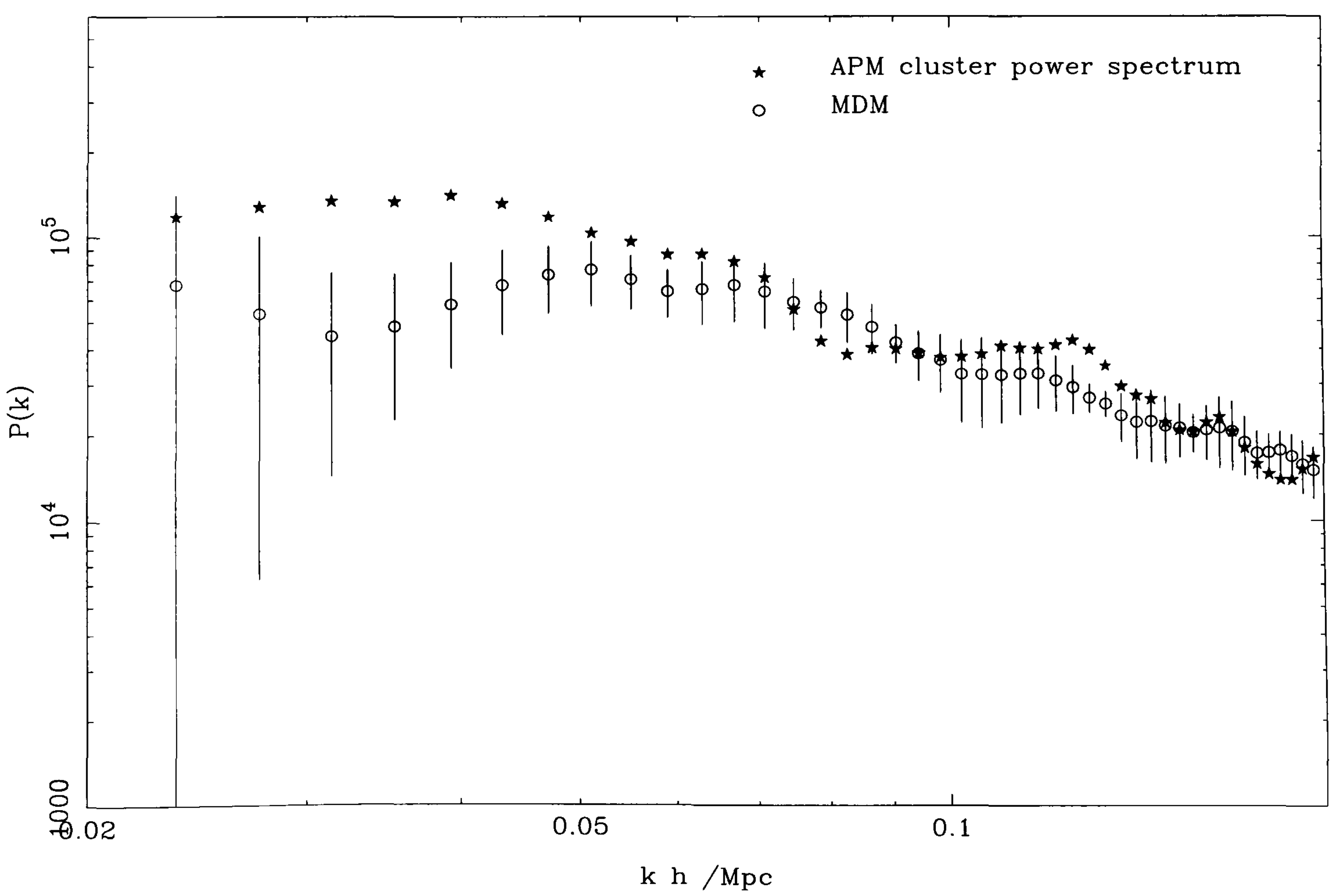


Figure 3.11: As Figure 3.9 for the MDM model.

3.4.2 Comparison with the Stromlo-APM galaxy survey

It is interesting to compare the power spectrum of clusters and galaxies. It has long been known, as discussed in Section 3.1.1, that clusters of galaxies show a higher clustering amplitude than that of galaxies. This would arise naturally in theories of structure formation where the structure grows via gravitational instability of initially Gaussian density fluctuations, if the clusters form at high peaks in the density field (Kaiser 1984, Bardeen *et al.* 1986). Given that rich clusters of galaxies are the most massive collapsed objects and are very rare (only 2 – 5% of galaxies are within an Abell radius of a $R \geq 1$ cluster, (Bahcall & Soneira 1983)) it would not be an unreasonable conjecture that they formed at the rare high σ peaks in the density field. The exact dependence of the degree of enhancement of the clusters relative to the mass would depend on the details of the cosmological model of structure formation (Barnes *et al.* 1985).

Figure 3.12 shows the power spectrum of the flux limited Stromlo-APM survey, as shown in Figure 2.7 of Chapter 2 compared with the power spectrum of the APM clusters, as shown in Figure 3.5. The two power spectra accurately have the same shape on all scales measured, with just a constant offset in amplitude between them. Figure 3.13 illustrates this point by plotting the ratio of the cluster power spectrum to that of the Stromlo-APM galaxies as a function of wave number. The cluster power spectrum is about 3.5 times higher than that of the galaxies on all scales (the value of b^2 stated on the plot is the weighted mean fit to the data points, the error σ is the one sigma error on this weighted mean).

Clusters of galaxies are very likely to be linear tracers of the underlying mass distribution. It would be very difficult to think of physical mechanisms that could significantly alter the positions or probability of formation of galaxy clusters. Figure 3.13 therefore provides some constraints on 'co-operative models' of galaxy formation. In such models (e.g. Babul & White 1991, Rees 1986, Efstathiou 1992, Bower *et al.* 1993) the efficiency of galaxy formation at a particular point is influenced by the result of galaxy formation at some distant point. For example the presence of a quasar in the vicinity may reduce or enhance the probability of a galaxy being formed. If these effects were dominant, the distribution of galaxies in the Universe would be correlated in a rather complicated and non-linear way with the distribution of mass. The fact that the power spectrum of galaxy clustering has the same shape as that of clusters of galaxies over a large range of wave numbers (Figure 3.13) means that these effects are probably not significant on such scales,

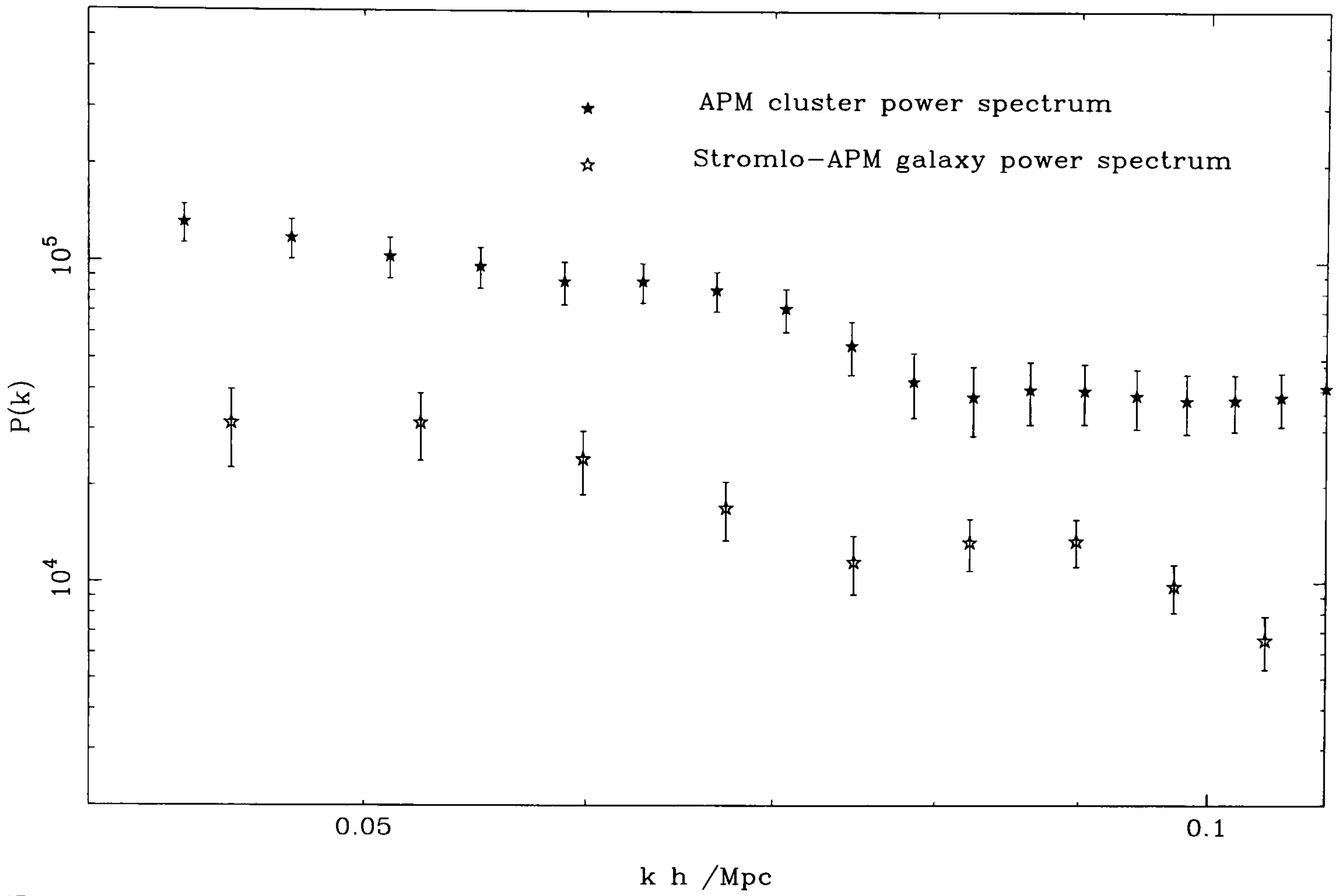


Figure 3.12: Here I compare the power spectrum of APM clusters, shown by filled stars with the power spectrum of galaxies in the Stromlo-APM survey, shown by open stars (Figure 2.7, Chapter 2). The plot is shown to a wave number of $k \sim 0.04\ h\ \mathrm{Mpc}^{-1}$ which is about the limiting wave number to which we can measure the power spectrum of Stromlo-APM galaxies.

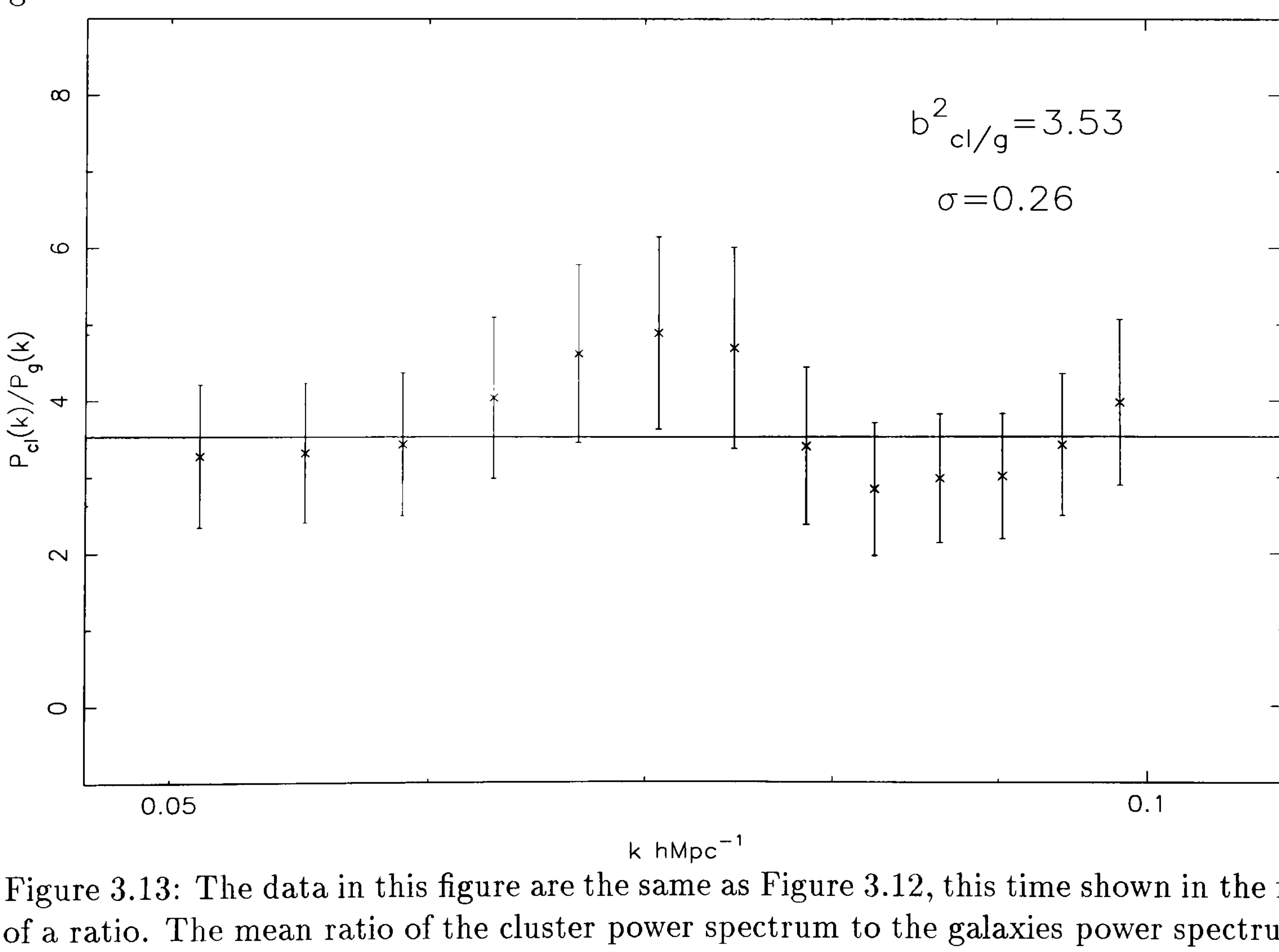


Figure 3.13: The data in this figure are the same as Figure 3.12, this time shown in the form of a ratio. The mean ratio of the cluster power spectrum to the galaxies power spectrum is ~ 3.5.

and that on scales with wave numbers $0.05 \lesssim k \lesssim 0.1$ galaxies are linearly biased with respect to the mass distribution.

3.5 Summary

I have presented a power spectrum analysis of the APM cluster redshift survey, which has previously been shown to be free of all significant selection effects and biases. The power spectrum of galaxy clusters can be fit by a power law $P(k) \propto k^{-1.6}$ on scales $k \gtrsim 0.04\ h\ \mathrm{Mpc}^{-1}$, flattening on larger scales and perhaps turning over on a scale of about $k \sim 0.024\ h\ \mathrm{Mpc}^{-1}$. This power spectrum has been shown to be inconsistent with the predictions of the standard CDM model, at very high confidence levels. The spectrum is better fit by either a LCDM model or a MDM model, the data not being adequate to distinguish convincingly between the two. This is consistent with earlier results from an analysis of the correlation function of this survey (Croft & Efstathiou 1994a). The measured power spectrum of APM clusters has been shown to be stable to the exact details of the power spectrum estimation. I have also compared the cluster power spectrum with that of galaxies from the Stromlo-APM redshift survey. The two power spectra have the same shape on all measurable scales, lending support to a linear biasing model. The power spectrum of APM clusters has an amplitude about 3.5 times larger than that of galaxies in the Stromlo-APM survey.

Chapter 4

The power spectrum of IRAS galaxies

The power spectrum of galaxy clustering provides one of the strongest constraints on theories of the formation of structure in the Universe (see Efstathiou 1996 and Strauss & Willick 1995 for recent reviews). The three-dimensional power spectrum of optically selected galaxies has been estimated by a number of authors (see *e.g.* Vogeley *et al.* 1992, Park *et al.* 1994, Baugh & Efstathiou 1993, Baugh & Efstathiou 1994 and Chapter 2). The power spectrum of IRAS galaxies has been estimated by Fisher *et al.* (1993) from the $1.2Jy$ redshift survey of Fisher *et al.* (1995a) and by Feldman, Kaiser & Peacock (1994) (FKP) from the QDOT survey of Lawrence *et al.* (1995). These estimates of the IRAS power spectrum differ in shape and amplitude (see *e.g.* Figure 5.7 of Efstathiou 1996); the power spectrum for the QDOT survey is steeper and has a higher amplitude at wave numbers $k \sim 0.05\ h\mathrm{Mpc}^{-1}$ than the power spectrum derived from the $1.2Jy$ sample. However, different techniques and weighting schemes were used by the authors and so it is unclear whether the differences in the power spectra are attributable to the estimators or to the galaxy catalogues.

In this chapter I address this problem by applying the techniques of FKP to the $1.2Jy$ and QDOT IRAS redshift surveys, using identical weighting schemes and sky coverage. The results presented here are closely related to a statistical analysis of galaxy counts-in-cells in the $1.2Jy$ and QDOT surveys described by Efstathiou (1995, hereafter E95). By comparing the counts on a cell-by-cell basis, E95 found that the two surveys were compatible with the hypothesis that they sample the same underlying density field. The variances in the cell counts of the QDOT survey on scales $\ell \sim$ 30–40 $h^{-1}\mathrm{Mpc}$ are, however, systematically higher than those derived from the $1.2Jy$ survey. Most of these differences are caused by a small number of galaxies in the QDOT survey that lie in the region of the Hercules supercluster. E95 concluded that the QDOT variances were biased high because Hercules is over-represented in the QDOT survey. In this chapter I show that similar conclusions apply to the power spectra of IRAS galaxies.

The surveys are described briefly in Section 4.1. The IRAS satellite and resulting surveys are described in more detail in Sections 5.1 and 5.2 of the next chapter. I review and test the power spectrum estimator in Section 4.2. In Section 4.3, I present power spectrum estimates for the QDOT and $1.2Jy$ surveys. I examine the sensitivity of the power spectrum estimates to the weights assigned to the galaxies and and show that the QDOT power spectrum is sensitive to galaxies in the Hercules supercluster and, consequently, to the weighting scheme. In contrast, the power spectra derived from the $1.2Jy$ survey are insensitive to galaxies in Hercules and the weighting scheme. I also examine the effect on the measured power spectrum of possible evolution in the IRAS galaxy density field. In Section 4.4 I combine the two surveys to produce the best estimate of the power spectrum of IRAS galaxies. Section 4.5 discusses alternative possibilities for the differences between the IRAS $1.2Jy$ and QDOT surveys. In Section 4.6 I compare the power spectrum of IRAS galaxies with the predictions for different cosmological models, using N-body simulations, and I summarize the results of this chapter.

4.1 IRAS galaxy redshift surveys

Both the QDOT survey and the $1.2Jy$ survey are based on version 2 of the IRAS Point Source Catalogue (Joint IRAS Science Working Group 1988, hereafter PSC). The PSC is a catalogue of the point sources seen by the IRAS satellite which flew in 1983 and surveyed the infrared sky in four wavelength bands: 12, 25, 60 and $100\mu m$. Most IRAS galaxy redshift surveys are based on point source catalogues flux limited in the $60\mu m$ band. The QDOT survey (Lawrence *et al.* 1995) is a 1 in 6 sparse-sampled redshift survey of galaxies with 60μm fluxes greater than $0.6Jy$, while the $1.2Jy$ survey is fully sampled at this brighter flux limit (Fisher *et al.* 1995a). The QDOT catalogue used here has been corrected for 211 erroneous redshifts in the original version. Both surveys exclude regions of the sky (defined by a sky-mask) at low Galactic latitudes and regions of high Galactic emission at 60μm (see Strauss *et al.* 1990, Rowan-Robinson *et al.* 1991a, also see Chapter 5). For the power spectrum analysis presented here I have used a concatenated QDOT and $1.2Jy$ sky-mask and used galaxies within a radial distance of 450 h^{-1}Mpc that lie above a Galactic latitude

$|b| = 10°$. This leaves 2024 galaxies in the QDOT sample and 4384 galaxies in the $1.2Jy$ sample.

4.2 Estimation of the power spectrum

I have followed the methods described in FKP (see also Chapter 2 and Chapter 3) to compute the power spectrum of a flux limited redshift survey. To simplify the discussion I use the same notation as FKP and Chapter 2. In summary, a weighted density field is computed

$$F(\mathbf{r}) = \frac{w(r)\left[n_g(\mathbf{r}) - \alpha n_s(\mathbf{r})\right]}{\left[\int \overline{n}^2(\mathbf{r})\, w^2(r)\, d^3r\right]^{\frac{1}{2}}}, \tag{4.1}$$

where the subscript g denotes the galaxy density in the real catalogue and s denotes the density field for a random catalogue with the same angular and radial selection functions as the galaxy survey, computed from equations (4.4) and (4.5) below. In equation (4.1), $\overline{n}(\mathbf{r})$ is the expected mean density of galaxies in a catalogue with the same angular and radial selection functions as the data. The function $\overline{n}(\mathbf{r})$ is therefore the mean galaxy density $\overline{n}(r)$ multiplied by the angular mask. The factor α is the ratio of the space densities in the real catalogue to that in the random catalogue. In the analysis presented here I use several hundred times as many points in the random catalogue as there are galaxies in the real catalogue, and compute α from the ratio of the sums

$$\sum_i \frac{1}{(1 + 4\pi\overline{n}(r_i)J_3)}, \tag{4.2}$$

where I have set $4\pi J_3 = 10000(h^{-1}\mathrm{Mpc})^3$, and the sums run over all galaxies and random points. Equation (4.2) provides a minimum variance estimate of the mean galaxy density and the specific choice for J_3 is based on the power spectra of CDM models, though the results presented below are insensitive to this value (see Efstathiou 1996, Section 5.3 for details).

The weight function $w(r)$ that minimises the variance of the power spectrum $P(k)$, under the assumption that the fluctuations are Gaussian, is given by

$$w(r) = \frac{1}{1 + \overline{n}(r)\, P_w(k)} \tag{4.3}$$

(FKP) and so depends on the value of $P_w(k)$ at each wave number k. As in FKP, and previous chapters, I have decided to set $P_w(k)$ in equation (4.3) equal to a constant value and to show how the estimates of the power spectrum change for four values of $P_w(k) =$ 2000, 4000, 8000 and 16000 $(h^{-1}\mathrm{Mpc})^3$, which span the range of interest. Changing the value of $P_w(k)$ used in equation (4.3) changes the effective depth of the catalogue. A larger value of $P_w(k)$ results in a larger effective depth. The power spectrum is derived by Fourier transforming equation (4.1), removing shot noise (equation 2.1.9 of FKP) and averaging the power spectrum estimates over shells in k-space of volume V_k. The power spectrum estimates will be correlated over a range in k space given approximately by $\delta k \sim D^{-1}$, where D is the characteristic depth of the survey.

The mean galaxy density is estimated from the luminosity function $\phi(L)$ for IRAS galaxies

$$\overline{n}(r) = \int_{L_{min}(r)}^{\infty} \phi(L)\; dL, \tag{4.4}$$

where $L_{min}(r)$ is the luminosity of a galaxy at distance r with flux equal to the flux limit of the survey. Galaxy distances are calculated from redshifts using the relativistic formula and assuming the Hubble constant is 100 km s^{-1} Mpc^{-1} and Ω equals 1. I use the parametric form of the luminosity function given by Saunders *et al.* (1990),

$$\phi(L) = C\left(\frac{L}{L_\star}\right)^{(1-\alpha)} exp\left[-\frac{1}{2\sigma^2} log_{10}^2\left(1+\frac{L}{L_\star}\right)\right] \tag{4.5}$$

with parameters

$C = \frac{2.6}{log(10)} \times 10^{-2}\, h^3\mathrm{Mpc}^{-3}$
$\alpha = 1.09$
$\sigma = 0.724$
$L_\star = 10^{8.47} h^{-2} L_\odot$.

Because of the relatively small median depth of the samples, evolutionary effects in $\phi(L)$ are unimportant compared to the random errors in $P(k)$. The parameters given above provide a very good fit to the $\overline{n}(r)$ for both the QDOT and the $1.2Jy$ surveys as shown in Figure 4.1. The effects of evolution in the IRAS density field are discussed in Section 4.3.2.

4.2.1 Tests of the power spectrum estimator

The power spectrum estimator used here is the same as that used in previous chapters and in those chapters it was tested extensively. Here then, I only briefly present a comparison

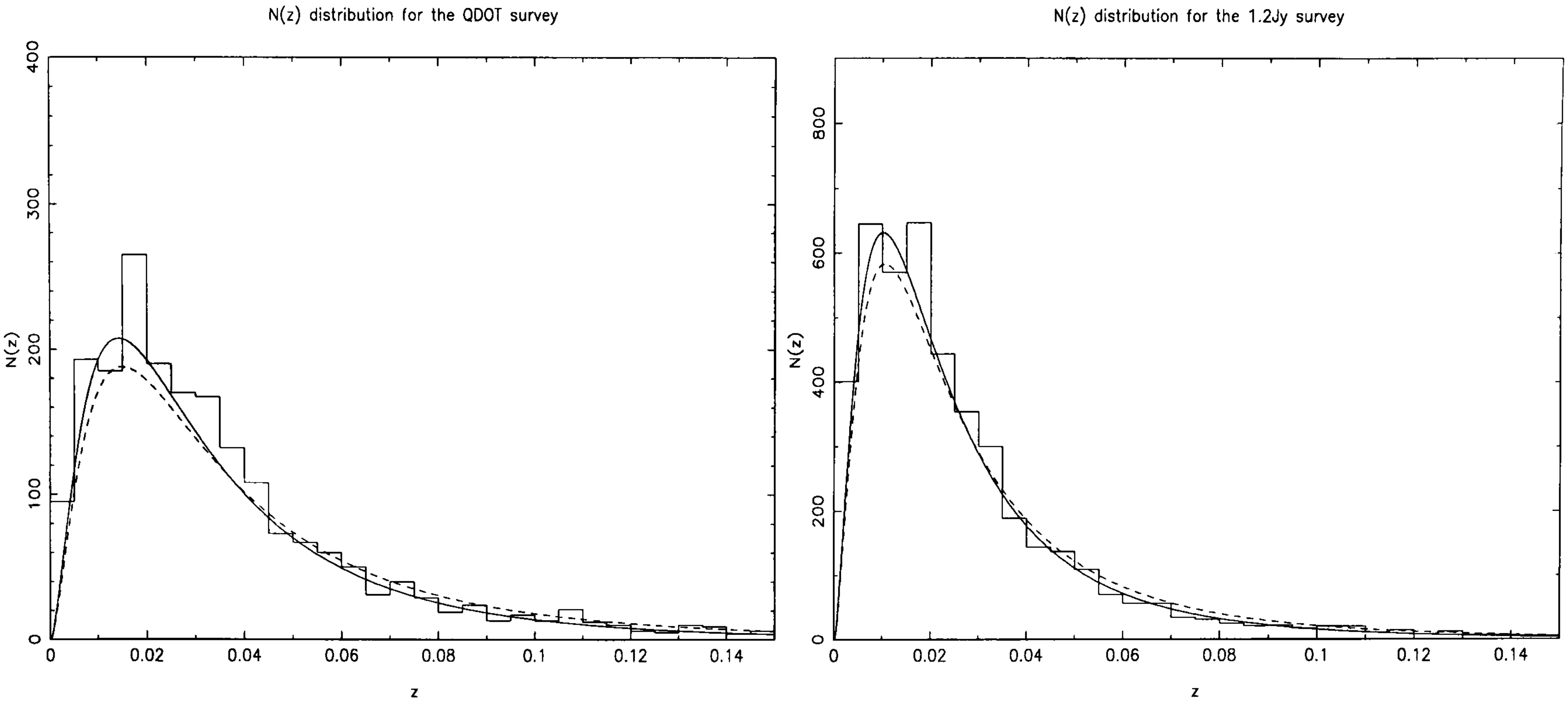


Figure 4.1: The left panel shows a histogram of the $N(z)$ distribution for the galaxies in the QDOT sample. The solid line shows the predicted $N(z)$ obtain using the selection function, equations (4.4) and (4.5). The right panel shows the same thing for the galaxies in the $1.2Jy$ sample. The dashed lines in each case show the predicted $N(z)$ distribution for a luminosity function with evolution included (see Section 4.3.2).

with N-body simulations. Shown in Figure 4.2 is the power spectrum derived by averaging over twenty realizations of simulated combined IRAS $1.2Jy$/QDOT surveys with the same sky-mask as that used in this chapter (a concatenated QDOT – $1.2Jy$ mask). The selection function of the simulated surveys is that of the $1.2Jy$ survey above a flux of $1.2Jy$ and is that of the QDOT survey from $0.6Jy$ to $1.2Jy$. The power spectrum from these simulated surveys is shown by the circles in Figure 4.2 and the error bars show the error on the mean $P(k)$. The solid line shows the power spectrum from the whole N-body box with no selection function or sky-mask. The dashed line shows the effect of the convolution with the window function and the dot-dash line shows the power spectrum after removing the bias correction described in Chapter 2 and derived in Appendix B. It can be seen that we accurately recover the power spectrum to fairly small wave numbers and that the bias due to the estimation of the quantity α from the data itself is small, but non-negligible. We expect the power spectrum results of this chapter to be reliable to wave numbers as small as $k \sim 0.04\, h\ \mathrm{Mpc}^{-1}$.

4.3 Results

In Figure 4.3 I show the power spectrum of the QDOT survey (circles) and the $1.2Jy$ survey (crosses) for each of the four values of $P_w(k)$ used in the weighting function. The

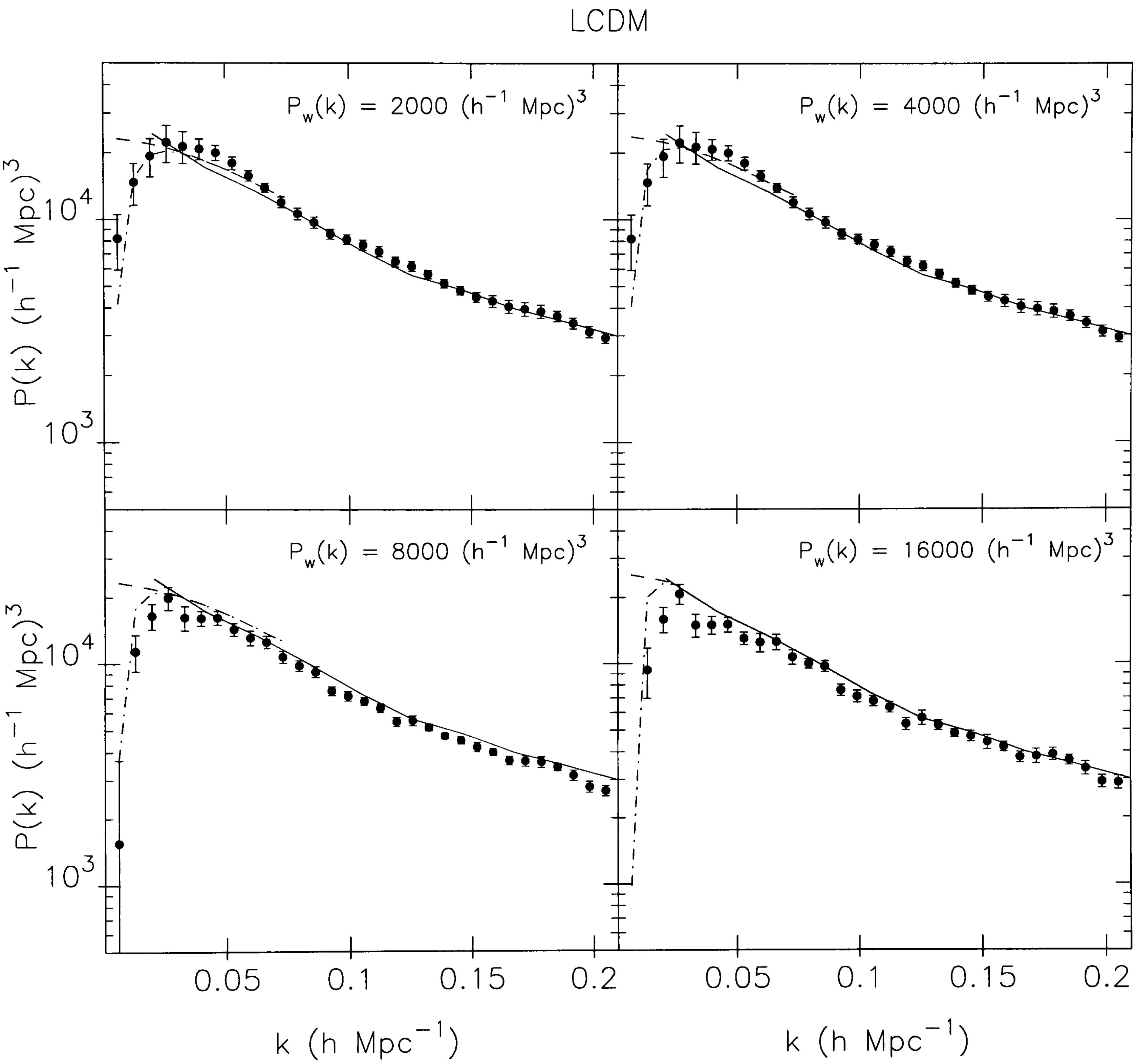


Figure 4.2: Four panel figure illustrates the accuracy of the power spectrum measured from a simulated IRAS data set, where the data set is a combined IRAS $1.2Jy$/QDOT survey. The circles show the power spectra averaged over twenty realizations and the error bars show the error on the mean $P(k)$. The simulated surveys are constructed from the low density CDM model described in Chapters 2 and 3 which has $\Omega = 0.2$ and a cosmological constant to make the model flat. These power spectra are computed at the output time when the rms mass fluctuations in spheres of radius 8 h^{-1}Mpc is unity. The solid line shows the power spectrum computed using the entire N-body box, with no window function or selection function. The dashed line (visible only at very small k) shows the effect of convolution with the IRAS window function, and the dot-dash line shows the effect of the bias due to the fact that we estimate the mean number density of the sample from the sample itself.

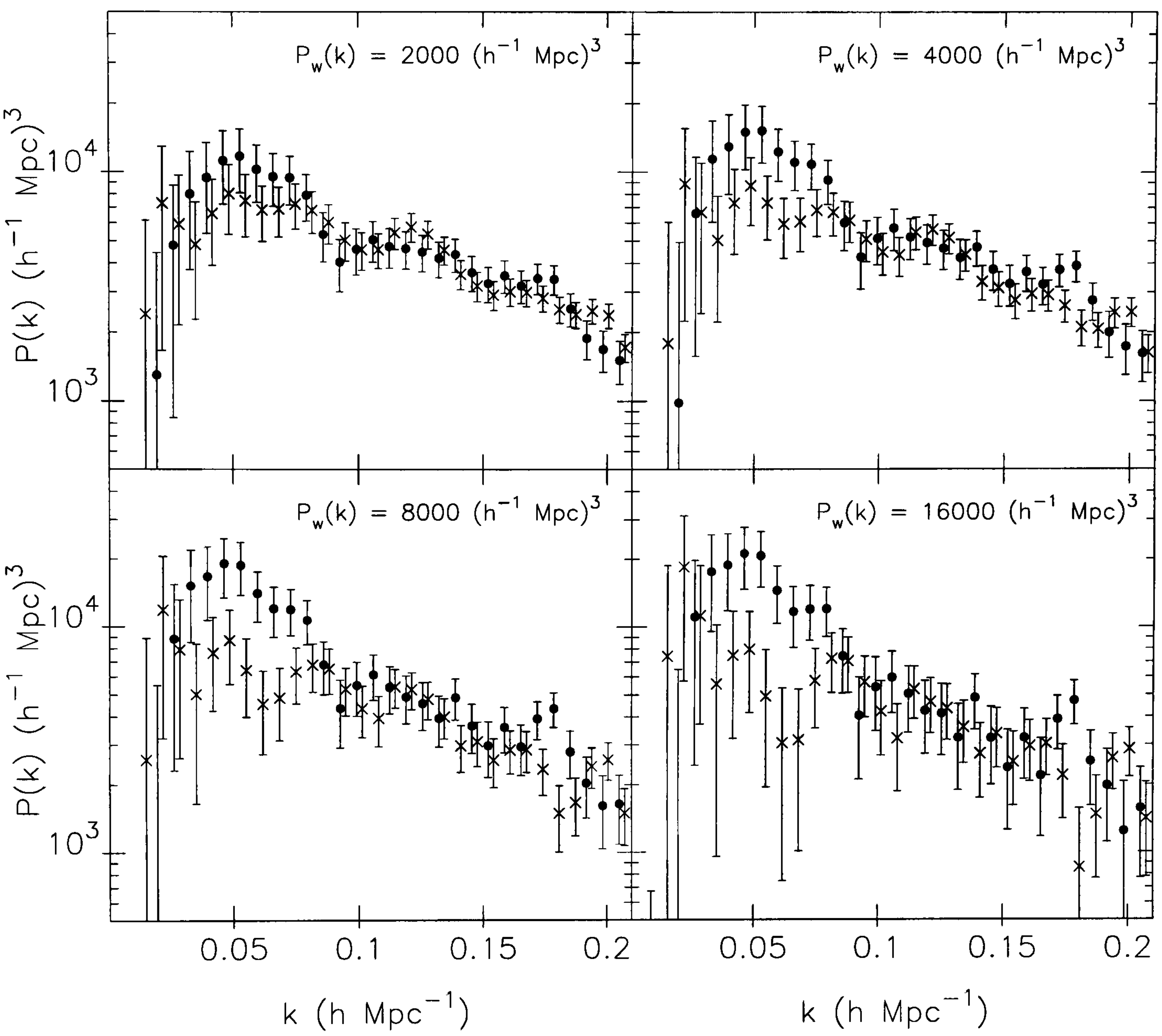


Figure 4.3: Power spectra of IRAS galaxies determined with four weighting schemes, $P_w(k) = 2000$ –16000 $(h^{-1}\text{Mpc})^3$ in equation (4.3), as indicated in each panel. The crosses show the power spectra measured from the $1.2Jy$ survey and the filled circles show the power spectra measured from the QDOT survey. The error bars show one standard deviation, computed from equation (2.4.6) of FKP. For clarity, the IRAS $1.2Jy$ power spectra have been shifted to the right of the plot by a third of the spacing between points in k-space.

QDOT spectra shown in Figure 4.3 are in very good agreement with those published by FKP. The estimates of the power spectrum for the $1.2Jy$ survey are also in good agreement with those of Fisher *et al.* (1993), but a point-by-point comparison is not straightforward because they use a different weighting scheme and an estimator which returns the power spectrum convolved with a cylindrical window function.

Figure 4.3 illustrates the sensitivity of the estimates to the weighting scheme. As we increase the value of $P_w(k)$ in equation (4.3), we assign progressively higher weight to more distant galaxies. Since the weights are designed to give a minimum variance estimate of $P(k)$, we would not expect the results to be sensitive to small changes in the weighting, and indeed the results for the $1.2Jy$ survey shown in Figure 4.3 are insensitive to changes in the weighting. However, the power spectrum measured from the QDOT survey is much more sensitive to the weights and increases systematically as we give proportionately higher weight to more distant galaxies. The estimates from the two surveys are consistent with each other when we weight with $P_w(k) = 2000(\ h^{-1}\mathrm{Mpc})^3$, as noted by FKP. However, when we weight with $P_w(k) \geq 8000(\ h^{-1}\mathrm{Mpc})^3$, the QDOT power spectrum lies about a factor of 2 or 3 higher than that of the $1.2Jy$ survey over the wave number range $0.08 \gtrsim k \gtrsim 0.03h\mathrm{Mpc}^{-1}$. From the amplitudes of the power spectra plotted in Figure 4.3, we would expect that weighting with $P_w(k) \sim 8000(\ h^{-1}\mathrm{Mpc})^3$ would be close to optimal over the wave number range 0.05–$0.1h\mathrm{Mpc}^{-1}$. It is surprising, therefore, that the power spectra estimated from the QDOT survey are so sensitive to the weights and appear to differ systematically from those of the $1.2Jy$ survey when we weight with $P_w(k) \geq 8000(\ h^{-1}\mathrm{Mpc})^3$.

The systematic difference is significant, even allowing for the fact that the points on the plot are correlated over some range of wave number. FKP show that the covariance between the power spectrum estimates at two points in k space, with wave numbers k_i and k_j, is

$$C_{ij} = \langle \delta P(k_i)\, \delta P(k_j) \rangle = \frac{2}{N_k N_{k'}} \Sigma_{\mathbf{k}} \Sigma_{\mathbf{k}'} \left| PQ\left(\mathbf{k} - \mathbf{k}'\right) - S\left(\mathbf{k} - \mathbf{k}'\right) \right|^2 \tag{4.6}$$

where the functions Q and S are given by

$$Q\left(\mathbf{k}\right) = \frac{\int d^3 r \overline{n}^2\left(\mathbf{r}\right) w^2\left(\mathbf{r}\right) e^{i\mathbf{k}\cdot\mathbf{r}}}{\int d^3 r \overline{n}^2\left(\mathbf{r}\right) w^2\left(\mathbf{r}\right)} \tag{4.7}$$

and

$$S\left(\mathbf{k}\right) = \frac{(1+\alpha) \int d^3 r \overline{n}\left(\mathbf{r}\right) w^2\left(\mathbf{r}\right) e^{i\mathbf{k}\cdot\mathbf{r}}}{\int d^3 r \overline{n}^2\left(\mathbf{r}\right) w^2\left(\mathbf{r}\right)}. \tag{4.8}$$

In Figure 4.4 I show the degree of correlation, $C_{ij}/(C_{ii})(C_{jj})$, of a point with index j (where $k = (2\pi(j-1)/950)\ h\ \mathrm{Mpc}^{-1}$) with the nearby points i, where $i - j$ has been plotted along the x axis. The covariance is plotted for j of 6, 15, 25, 35, and 55. At the wave numbers where we see a difference between the QDOT and $1.2Jy$ surveys (i between about 5 and 14) the points are only significantly correlated over a distance of about 1.5 spacings in k-space. For larger wave numbers the correlation coefficient quickly drops to zero; for $j \gtrsim 25$ (corresponding to wave numbers $k \gtrsim 0.16\ h\ \mathrm{Mpc}^{-1}$) neighbouring points in the power spectrum are virtually independent.

4.3.1 Effects of removing the Hercules supercluster

The analysis of the $1.2Jy$ and QDOT surveys described in E95, shows that the cell statistics of the QDOT survey are extremely sensitive to a small number of galaxies in the region of the

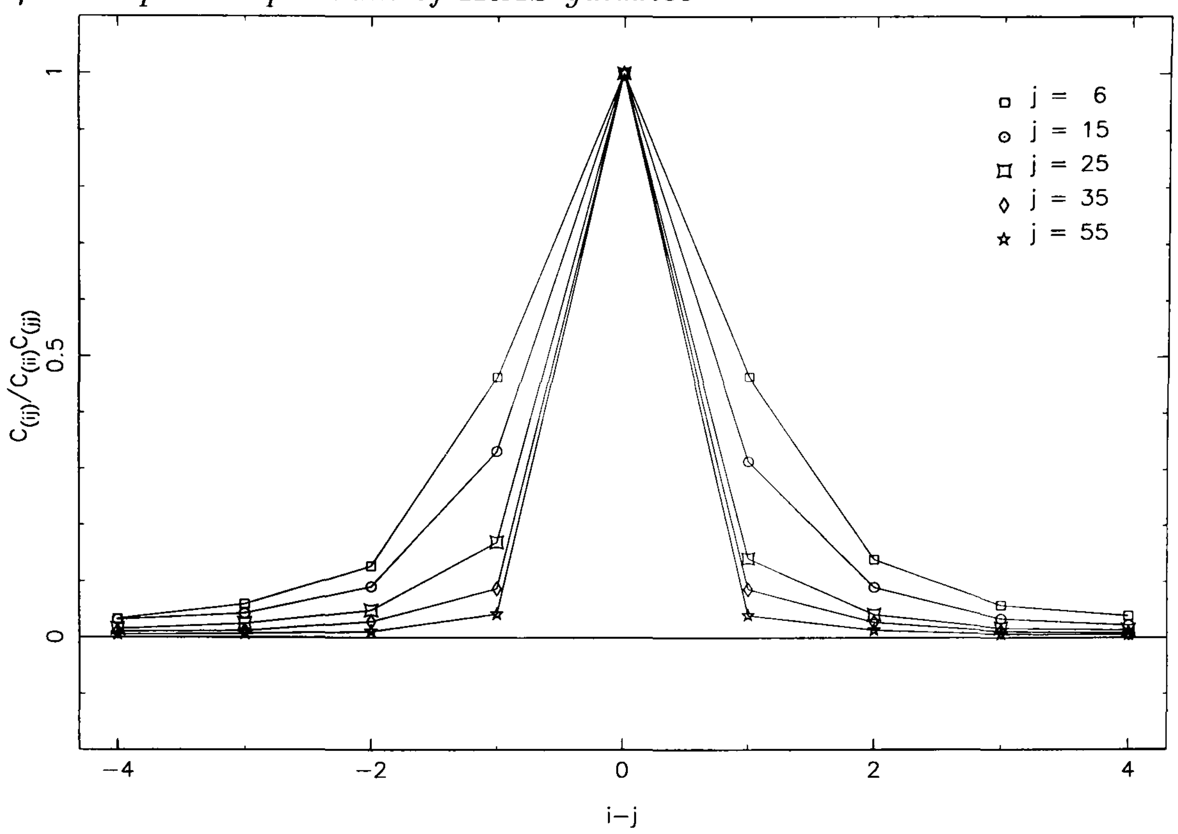


Figure 4.4: The degree of correlation between points in k-space for the power spectra in this chapter. I plot the correlation coefficient on the y axis against $i-j$ on the x axis, where j is the index of the wave number of the point in k-space ($k = (2\pi(j-1)/950)\ h\ \mathrm{Mpc}^{-1}$) and i is the index of the neighbouring points. The degree of correlation is shown for wave numbers with indices j of 6, 15, 25, 35 and 55.

Hercules supercluster ($l = 48°$, $b = 43°$, $v = 11000\mathrm{km\,s}^{-1}$). A similar conclusion applies to the power spectrum analysis described here. I illustrate this point by recomputing the power spectra after removing galaxies in the Hercules region. The QDOT survey was divided into 54 regions of the sky called q-sectors (see Lawrence *et al.* 1995). The Hercules supercluster is located largely in q-sector 7 which covers the area $40° \leq b \leq 60°$, $0° \leq l \leq 60°$. I exclude Hercules by removing all galaxies that lie in q-sector 7; this removes 59 galaxies from the QDOT survey and 74 galaxies from the $1.2Jy$ survey.

Figure 4.5 shows the power spectra after this is done. The power spectra for the $1.2Jy$ survey are consistent within the calculated errors with those plotted in Figure 4.3, showing that Hercules has little effect on this survey. In contrast, the results for the QDOT survey differ significantly from those plotted in Figure 4.3. With the Hercules region removed, the QDOT power spectra are in excellent agreement with those of the $1.2Jy$ survey and are insensitive to the weighting scheme. There is little doubt, therefore, that the differences between the $1.2Jy$ and QDOT power spectra apparent in Figure 4.3, and the sensitivity of the QDOT power spectra to the weighting scheme, are caused by a small number of QDOT galaxies in the Hercules region. This is consistent with the results of the cell count analysis described in E95. Galaxies in the Hercules region are over-represented in the QDOT survey (see E95) and so bias the power spectrum to high values. As the factor $P_w(k)$ is increased

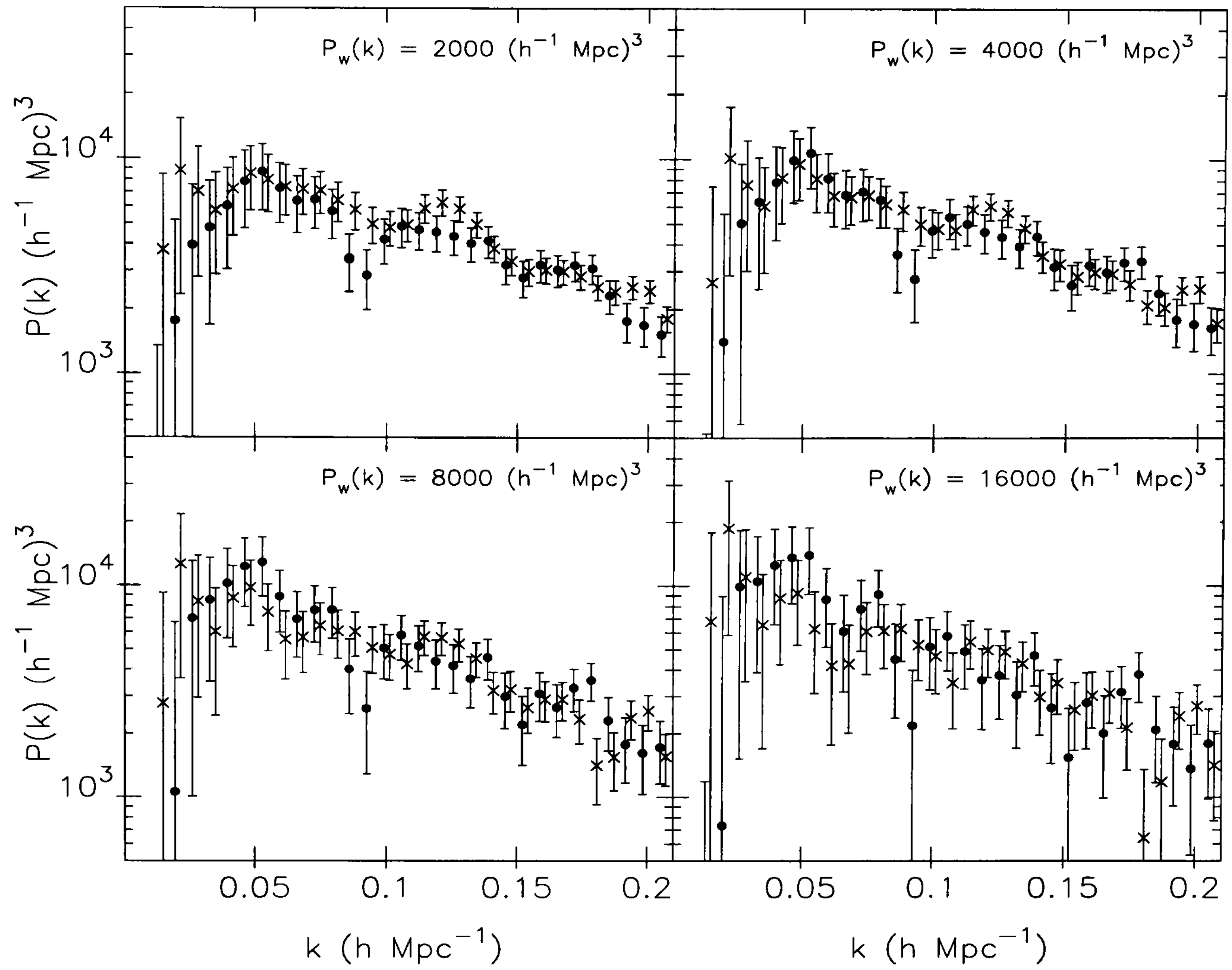


Figure 4.5: As Figure 4.3 except that here I have removed 59 galaxies from the QDOT survey and 74 galaxies from the $1.2Jy$ survey that lie in the region of the Hercules supercluster.

in the weighting function, the Hercules region is given higher weight and so the discrepancy with the $1.2Jy$ power spectrum becomes more pronounced.

4.3.2 Effects of evolution in the IRAS galaxy population

In this section, I show that the difference between the QDOT and $1.2Jy$ surveys is not remedied by using a selection function corrected for evolution in the number density of galaxies as a function of redshift. The evolution of IRAS galaxies is uncertain and there is some disagreement about whether or not there is strong evolution in the IRAS galaxy population. As discussed in the next chapter, the QDOT team (Saunders *et al.* 1990) parameterise the evolution by writing

$$\phi_z(L) = g(z)\,\phi_0\left(\frac{L}{f(z)}\right). \tag{4.9}$$

If pure density evolution is assumed, $f(z) = 1$ and $g(z) = (1+z)^p$. The parameter p is measured to be 6.7 ± 2.3. If pure luminosity evolution is assumed, and we suppose that luminosities decay exponentially, then $f(z) = exp(\frac{2}{3}q[1-(1+z)^{-3/2}])$ and q is measured to be 3.2 ± 1.

With the current data sets, we cannot tell the difference between luminosity and density evolution, although in reality the true evolution is unlikely to be simply one or the other. The evolution measured in the $1.2Jy$ survey is at odds with the result of Saunders *et al.* (1990), quoted above. Fisher *et al.* (1992) find that the distribution of IRAS galaxy redshifts in the $1.2Jy$ catalogue is consistent with no evolution (see next chapter for more discussion). Recently the QDOT survey has been reanalysed, as described in the next chapter (Lawrence *et al.* 1995), and the evolution parameter p has been revised down slightly. The catalogue used by Saunders *et al.* (1990) to calculate $p = 6.7$ was subject to various errors and Lawrence *et al.* (1995) have repeated the analysis on a corrected QDOT catalogue, finding $p = 5.6 \pm 2.3$. If an approximate correction is made for a possible non-linearity in the PSC fluxes then Lawrence *et al.* (1995) estimate the evolution parameter to be $p = 4.5 \pm 2.3$ (see Section 4.5 and Chapter 5 for discussion of PSC fluxes).

For the calculation of the power spectrum, the question of whether the evolution is in luminosity or density is not important, as long as the clustering is independent of luminosity. For power spectrum analysis all we require is a model for the selection function, i.e. the important information is contained in the shape of the $N(z)$ distribution. The power spectra shown in Figure 4.6 are computed assuming there is some evolution in the IRAS galaxy number densities. A value of $p = 4.5$ is adopted, so the number density as a function of radial distance $\overline{n}'(r)$ is now given by

$$\overline{n}'(r) = \int_{L_{min}(r)}^{\infty} \phi'(L)dL \tag{4.10}$$

with

$$\phi'(L) = \phi(L)\,(1+z)^p\,, \tag{4.11}$$

where $\phi(L)$ is given in equation (4.5). I have normalized the selection function for the case of evolution by setting

$$\int_{V_{surv}} \overline{n}(r)dV = \int_{V_{surv}} \overline{n}'(r)dV. \tag{4.12}$$

The resulting functions $N(z)$ are shown as the dotted lines in Figure 4.1. Returning to Figure 4.6, it can be seen that even with this much evolution assumed, the relative difference between the power spectra changes little. (I have also run the analysis for an assumed evolution parameter $p = 6.7$, the results of which are not presented here as they look very similar to the results shown in Figure 4.6).

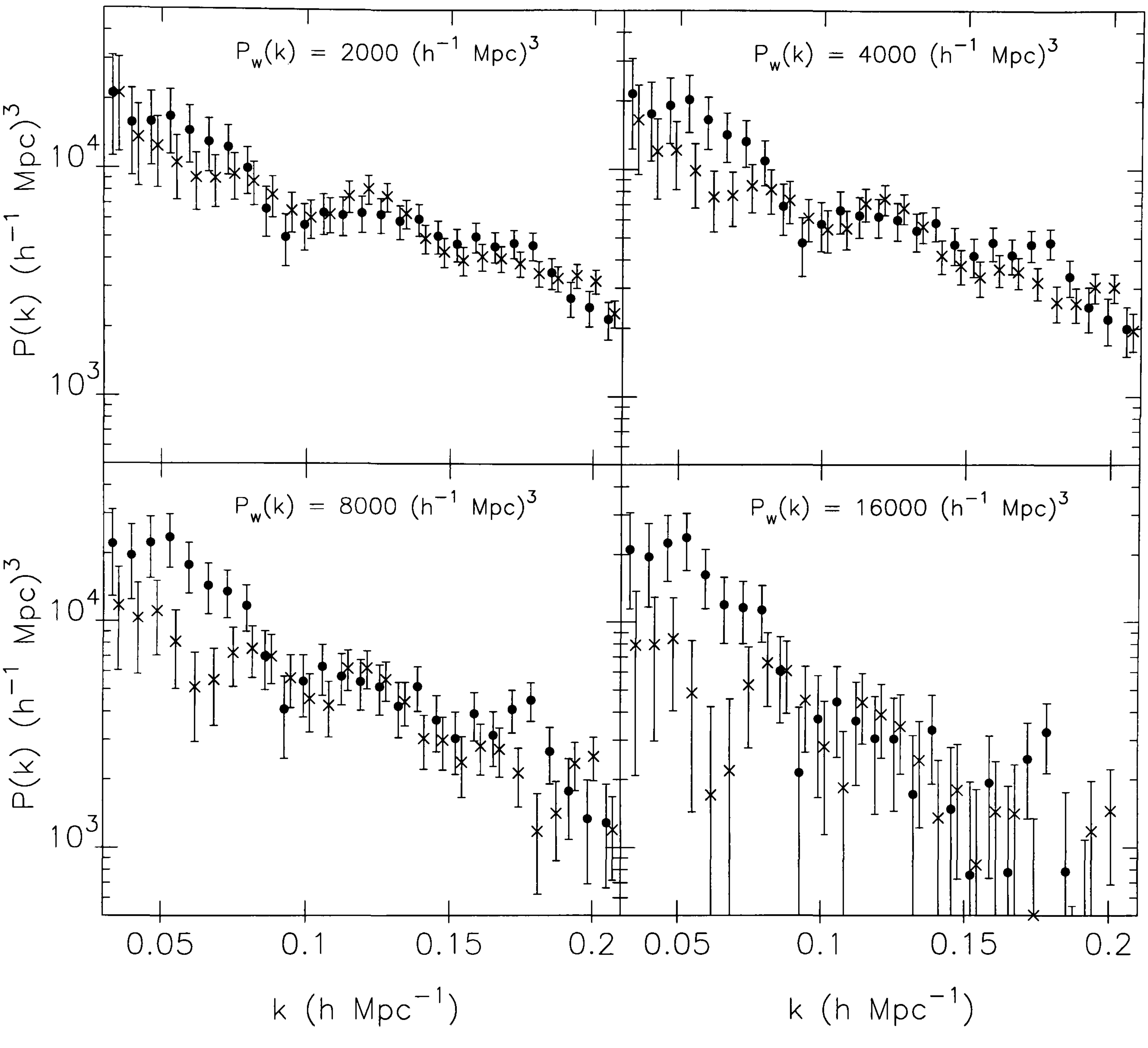


Figure 4.6: Power spectra of IRAS galaxies determined with four weighting schemes, $P_w(k) = 2000$ -16000 $(h^{-1}\text{Mpc})^3$ in equation (4.3), as indicated in each panel. The crosses show the power spectra measured from the $1.2Jy$ survey and the filled circles show the power spectra measured from the QDOT survey. The error bars show one standard deviation, computed from equation (2.4.6) of FKP. For clarity, the IRAS $1.2Jy$ power spectra have been shifted to the right of the plot by a third of the spacing between points in k-space. In both cases the selection function used in the power spectrum analysis is that appropriate in the case that IRAS galaxies are undergoing pure density evolution, $\phi_z(L) = (1+z)^p\phi_0(L)$ with $p = 4.5$.

4.4 P(k) from the combined QDOT and $1.2Jy$ IRAS surveys

Although the analysis in E95 shows that the Hercules region is over-represented in the QDOT survey, most likely because of a statistical fluctuation in the random numbers used to construct the survey, the $1.2Jy$ and QDOT surveys were found to be consistent with each other within Poisson statistics. To obtain an estimate of $P(k)$ from the two surveys it therefore seems reasonable simply to add them together without excluding galaxies from any particular region. In practice the $1.2Jy$ survey carries so much weight in the analysis of the combined surveys that it makes little difference whether or not we exclude Hercules from QDOT.

Figure 4.7 shows the power spectrum derived for the combined surveys. In this figure I remove QDOT galaxies with 60μm fluxes above $1.2Jy$, since these are already included in the $1.2Jy$ survey. I have used the concatenated QDOT-$1.2Jy$ mask, as in the estimates shown in Figure 4.3, and a selection function for the combined surveys computed from equations (4.4) and (4.5). I have set $P_w(k) = 8000(h^{-1}\mathrm{Mpc})^3$ in the weighting function of equation (4.3); varying $P_w(k)$ in the weight function over the range 2000 –16000 $h^3\mathrm{Mpc}^{-3}$ shifts the points by less than the 1σ error bars. The power spectrum plotted in Figure 4.7 is about a factor of two lower in amplitude over the wave number range $0.03 \lesssim k \lesssim 0.08h\mathrm{Mpc}^{-1}$ than the QDOT power spectrum plotted in Figure 7 of FKP. In Table 1 (at the end of this chapter) I give the values of the power spectra and the one sigma errors for the combined survey, weighted with $P_w(k) = 4000, 8000, \mathrm{and}\ 16000(\ h^{-1}\mathrm{Mpc})^3$

4.5 The 'Hamilton effect'

Hamilton (1995) (hereafter H95) has also examined the differences between the QDOT and IRAS $1.2Jy$ surveys. He uses the $2Jy$, $1.2Jy$ and QDOT surveys to calculate the correlation function as a function of two variables $\xi(\sigma, \pi)$, where, as before, σ is the separation of galaxies perpendicular to the line of sight to the galaxies and π is the separation of the galaxies parallel to the line of sight. In real-space this function is isotropic, i.e. if we plot contours in the σ, π plane we would get circular contours. Redshift-space distortions cause the function to become anisotropic. On small scales the incoherent peculiar velocities (from galaxies in the centres of clusters for instance) cause the contours to be elongated along the π direction, while on large scales the peculiar velocities caused by infall of galaxies into clusters and outflow from voids cause elongation of the contours along the σ axis. H95

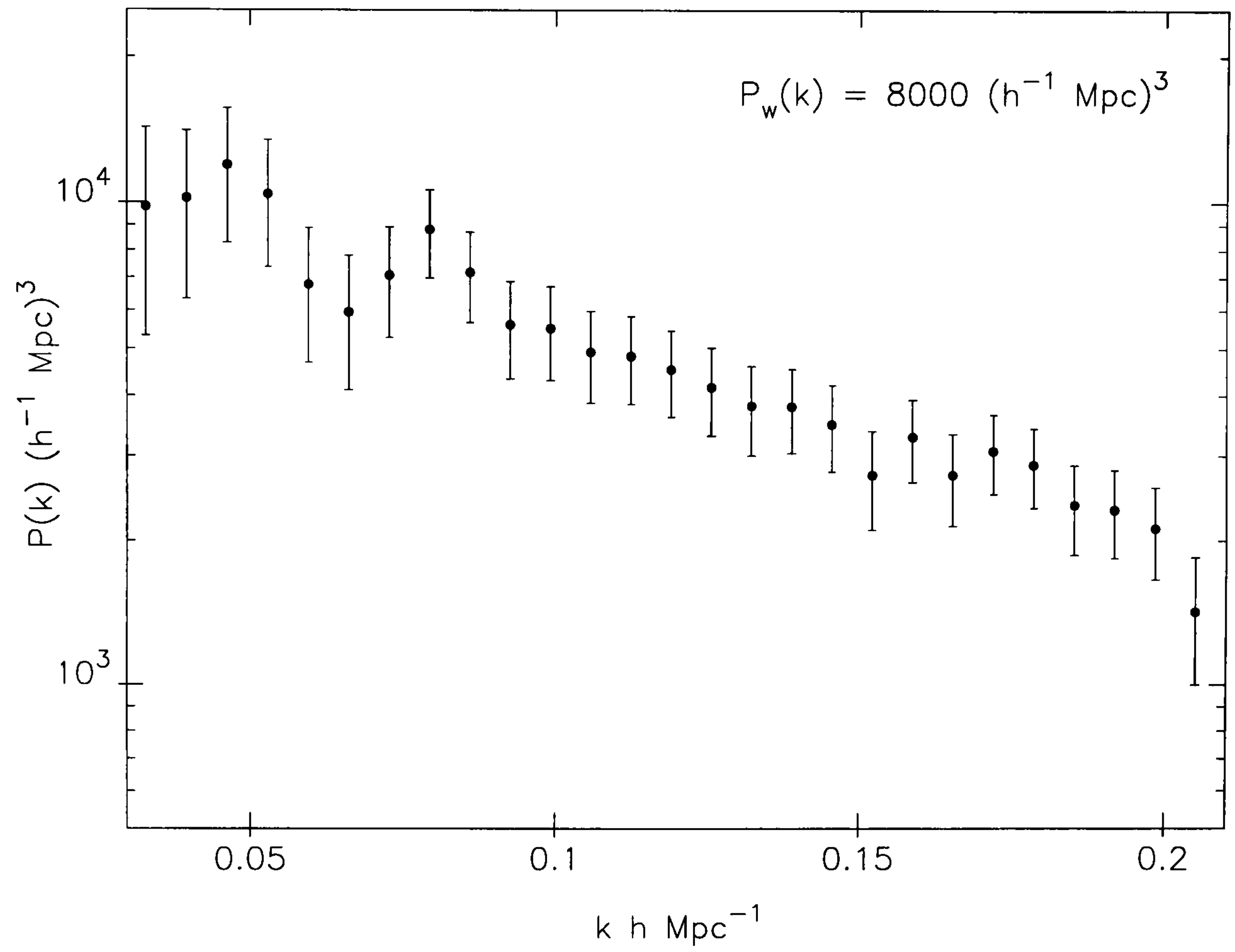


Figure 4.7: Figure shows the power spectrum from the combined QDOT and IRAS $1.2Jy$ survey. A value of $P_w(k) = 8000(\ h^{-1}\mathrm{Mpc})^3$ is used in the weighting function.

analyses the three IRAS redshift surveys for the regions $r > 80\ h^{-1}$Mpc and $r < 80\ h^{-1}$Mpc. For the QDOT survey especially, he finds that the region beyond $80\ h^{-1}$Mpc has contours which are elongated along the π axis. This is unexpected. The influence of small scale peculiar velocities is not expected to extend to such a large scale. The effect is seen to a much lesser extent in the IRAS $1.2Jy$ survey and not at all in the $2Jy$ survey.

H95 also measures the ratio of the quadrupole to monopole power in the IRAS surveys as a function of wave number. In linear perturbation theory the redshift-space power $P_s(k)$ is given in terms of the real-space power spectrum $P_r(k)$ by

$$P_s(k) = P_r(k)\left(1 + \beta\mu^2\right)^2 \tag{4.13}$$

(Kaiser 1987). If we expand this in Legendre polynomials, then according to linear theory the redshift-space power spectrum has a monopole (M), quadrupole (Q) and hexadecapole component. The ratio of the quadrupole to monopole moment can be used to measure the value of $\Omega^{0.6}/b$ (H95, Cole, Fisher & Weinberg 1994,Cole, Fisher & Weinberg 1995) :

$$\frac{P_Q(k)}{P_M(k)} = \frac{\frac{4}{3}\beta + \frac{4}{7}\beta^2}{1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2}. \tag{4.14}$$

Thus the ratio of the quadrupole to monopole power in equation (4.14) is a constant in the linear regime. Non-linearities suppress the quadrupole moment, and eventually it be-

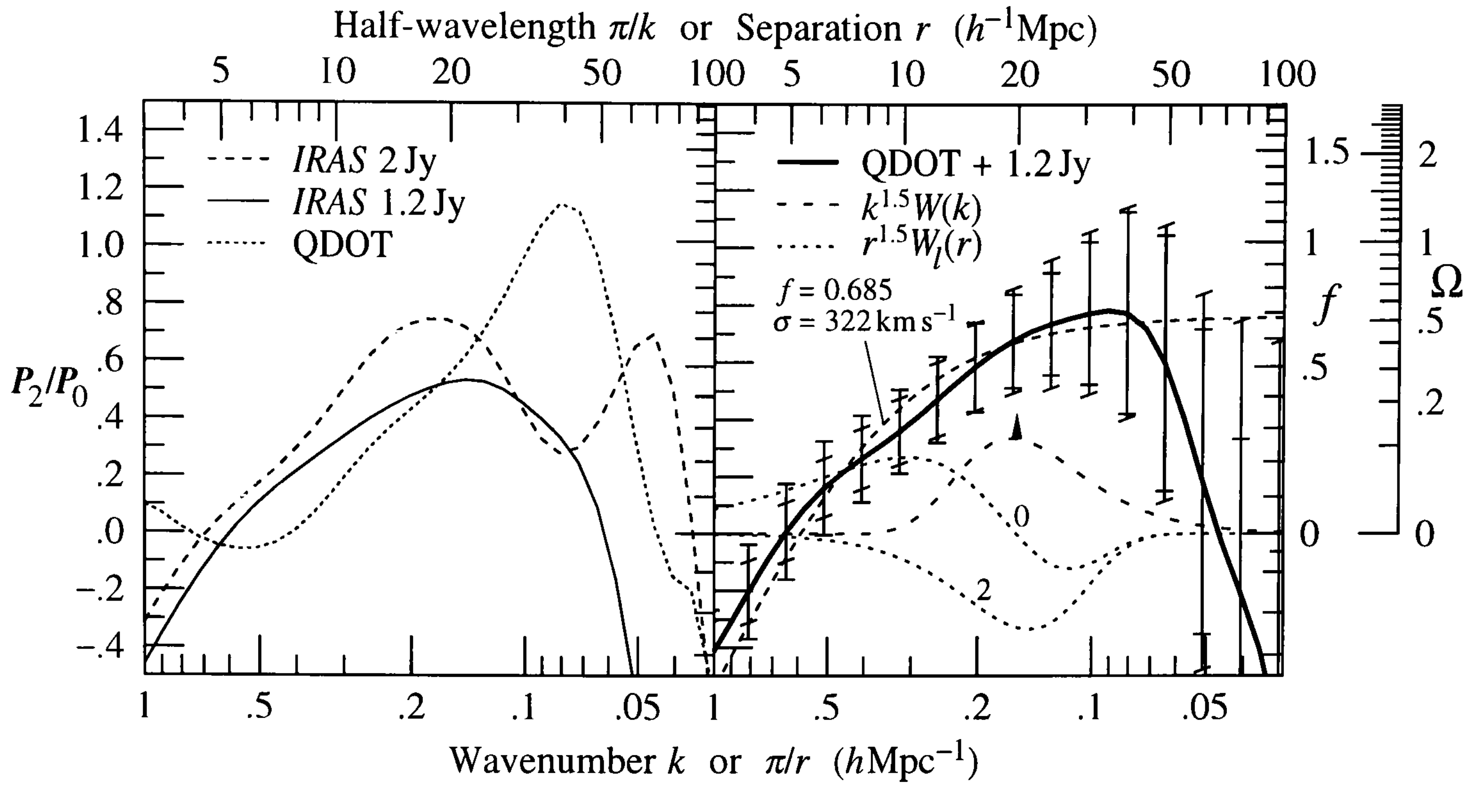


Figure 4.8: The ratio of the quadrupole to monopole power from Hamilton (1995). The left panel of the figure shows this ratio for the QDOT, IRAS $1.2Jy$ and IRAS $2Jy$ surveys. The right hand panel shows the ratio for the combined QDOT and $1.2Jy$ surveys. In the right hand panel the lower set of dashed lines show information about the window function, explained in Hamilton (1995). The uppermost dashed line shows the predicted ratio of the quadrupole to monopole power for $\beta = 0.685$ and a one-dimensional pairwise exponential velocity dispersion of $\sigma = 322\text{km s}^{-1}$.

comes negative. H95 finds that for the QDOT survey especially the ratio of quadrupole to monopole power does not tend to a constant at large scales but turns down again as shown in Figure 4.8 (kindly supplied by Andrew Hamilton). This is another reflection of the large scale 'finger of God' type behaviour (although not due to small scale peculiar velocities). H95 also sees that the monopole power spectrum of QDOT is too large, as found in this chapter. If it were the small scale peculiar velocities which were causing the observed anisotropies the monopole power would be too low rather than too high.

As an explanation for all of the above observations H95 advocates that there are systematic variations in the depth of the PSC over the sky. Such an effect would be more marked in the QDOT survey beyond a distance of 80 h^{-1}Mpc because this is where the selection function of QDOT begins to steepen, so small variations in flux cause relatively large variations in density. Also, the flux limit of QDOT ($0.6Jy$) is nearest the flux limit of the PSC ($0.5Jy$). H95 carries out an analysis to estimate the size of such an effect if it were to be the cause of the observed anomalies. He concludes it could be caused by a variation in the flux limit across the sky on scales of $\sim 7°$, with dispersion $0.1Jy$.

As I have just mentioned, H95 suggests that the effect is seen preferentially in the

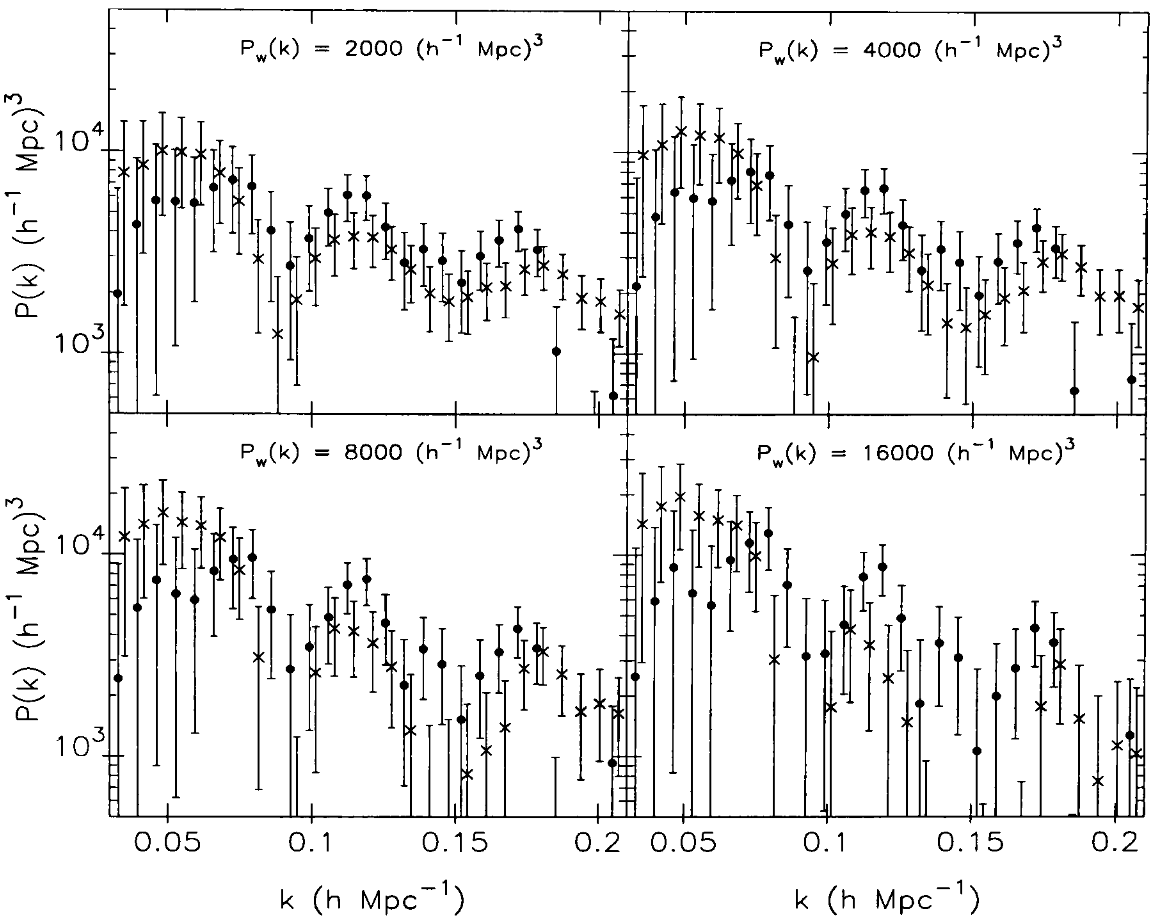


Figure 4.9: Power spectra of QDOT IRAS galaxies determined with four weighting schemes, $P_w(k) = 2000 - 16000\ (h^{-1}\mathrm{Mpc})^3$ in equation (4.3), as indicated in each panel. The crosses show the power spectra measured from galaxies above a flux limit of $1.2Jy$ and the circles show the power spectra of galaxies between 0.6 and $0.8Jy$. The error bars show one standard deviation, computed from equation (2.4.6) of FKP. For clarity, the crosses have been shifted to the right of the plot by a third of the spacing between points in k-space.

QDOT sample because of its fainter flux limit. In Figure 4.9 I show the power spectrum of the galaxies in the QDOT survey with flux above $1.2Jy$ (718 galaxies) together with the power spectrum for galaxies in the QDOT survey with fluxes near to the flux limit (0.6 to $0.8Jy$, 703 galaxies). I have removed the galaxies in q-sector 7 before calculating these power spectra, since the majority of them are faint. The crosses show the power spectra from the brighter flux slice and the circles show the power spectra from the $0.6 - 0.8Jy$ slice. The two spectra are the same to within the calculated errors, with the fainter flux slice showing no more power than the brighter flux slice.

The fact that the faint subsample of the QDOT survey shows no excess of power over the bright sample is encouraging, but does not directly test the hypothesis that there are flux errors which vary as a function of position on the sky. This idea must be considered seriously and there are two possibilities. The first of these is that the error bars on the ratio of the quadrupole to monopole moment (P_2/P_0, shown in Figure 4.8) are underestimates of the true error. The error bars on this ratio calculated by Cole, Fisher & Weinberg (1995), for instance, via Monte-Carlo simulation of the IRAS QDOT survey, are much larger. The second possibility is that the error bars shown in Figure 4.8 are realistic and the effect is significant, i.e. there is an unknown systematic error in the IRAS fluxes. This problem will need further investigation using the new PSCz redshift survey (described in Chapter 5).

The actual fluxes stated in the PSC are accurate to about 10%. Recently Lawrence *et al.* (1995) have tried to quantify the accuracy of the PSC fluxes using sources which are believed to be stellar photospheres. These have a definite S_{60}/S_{12} ratio and a negligible error in the measurement of S_{12}. By investigating the scatter in the ratio S_{60}/S_{12} plotted against S_{12} for these sources, Lawrence *et al.* (1995) quantified three types of error:

1. A fractional error $\sigma_f = KS_{60}$.

2. An absolute error σ_a.

3. A non-linearity in the flux scale. They allowed the mean ratio S_{60}/S_{12} to vary with S_{12} such that the mean ratio, $\overline{r}$ is given by $\overline{r}_0 = \gamma log(S_{12}/1Jy)$.

Maximum likelihood values for the parameters K, σ_a and γ are 0.096 ± 0.006, 0.056 ± 0.006 and -0.019 ± 0.006 respectively. Thus the fractional error measured agrees well with that quoted in the IRAS Explanatory Supplement (Beichman *et al.* 1988), but there is also a three sigma detection of non-linearity in the PSC fluxes. Lawrence *et al.* (1995) warn that this result is not conclusive due to the systematic problems caused by incomplete removal of non-photospheric sources from the analysis. However, if the result is to be believed, it represents a 4.5% systematic change in the $60\mu m$ flux over a factor of ten in flux. How such an error would affect the analysis of large scale structure has yet to be evaluated.

The above analysis does not specifically address the question of variation of flux levels on the sky. One possible position dependent source of error is the so called 'hysteresis effect' described in the next chapter, due to a change in the detector sensitivity as the satellite scans through the Galactic plane. Strauss *et al.* (1990) show that this is a negligible effect, at least to a flux limit of $1.2Jy$. As far as I know it has not been investigated to the deeper flux limit of $0.6Jy$. To test the power spectrum dependence on Galactic latitude, I have divided the QDOT sample into those galaxies above a Galactic latitude of 35° (1122 galaxies, excluding q-sector 7) and those below this latitude (902 galaxies, also excluding the region $|b| < 10°$). Figure 4.10 shows the result. The two power spectra are consistent to within the calculated error bars except for three points at $k \sim 0.06 - 0.08\ h$ Mpc^{-1} which are discrepant by about $1 - 2\sigma$. Since these points are highly correlated, this is probably not very significant. Figure 4.11 shows the same test, but this time with q-sector 7 included in the analysis. As expected the power spectra for the region above $b = 35°$

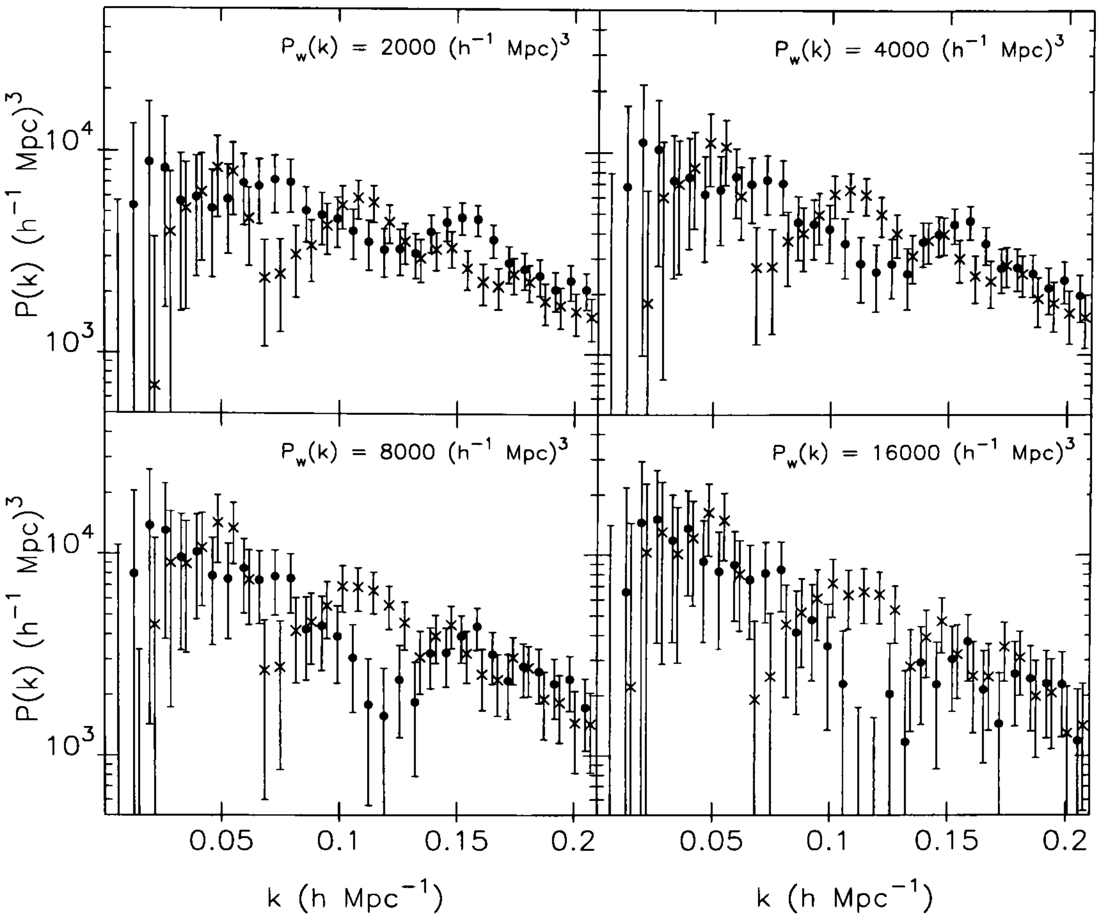


Figure 4.10: Power spectra of QDOT IRAS galaxies determined with four weighting schemes, $P_w(k)$ = 2000 –16000 $(h^{-1}\text{Mpc})^3$ in equation (4.3), as indicated in each panel. The crosses show the power spectra measured from galaxies above a Galactic latitude of 35° (excluding the Hercules cluster) and the circles show the power spectra for galaxies below a latitude of 35°. The error bars show one standard deviation, computed from equation (2.4.6) of FKP. For clarity, the crosses have been shifted to the right of the plot by a third of the spacing between points in k-space.

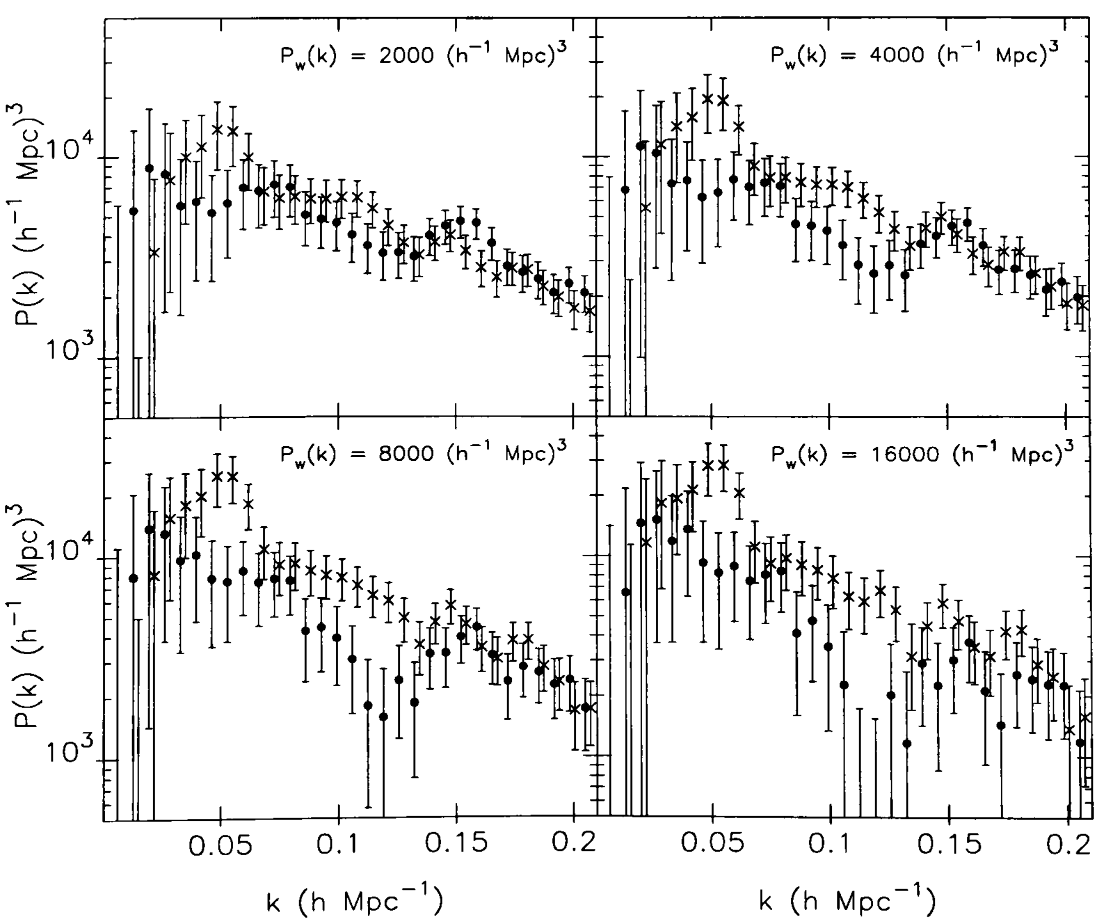


Figure 4.11: As Figure 4.10 but with Hercules included.

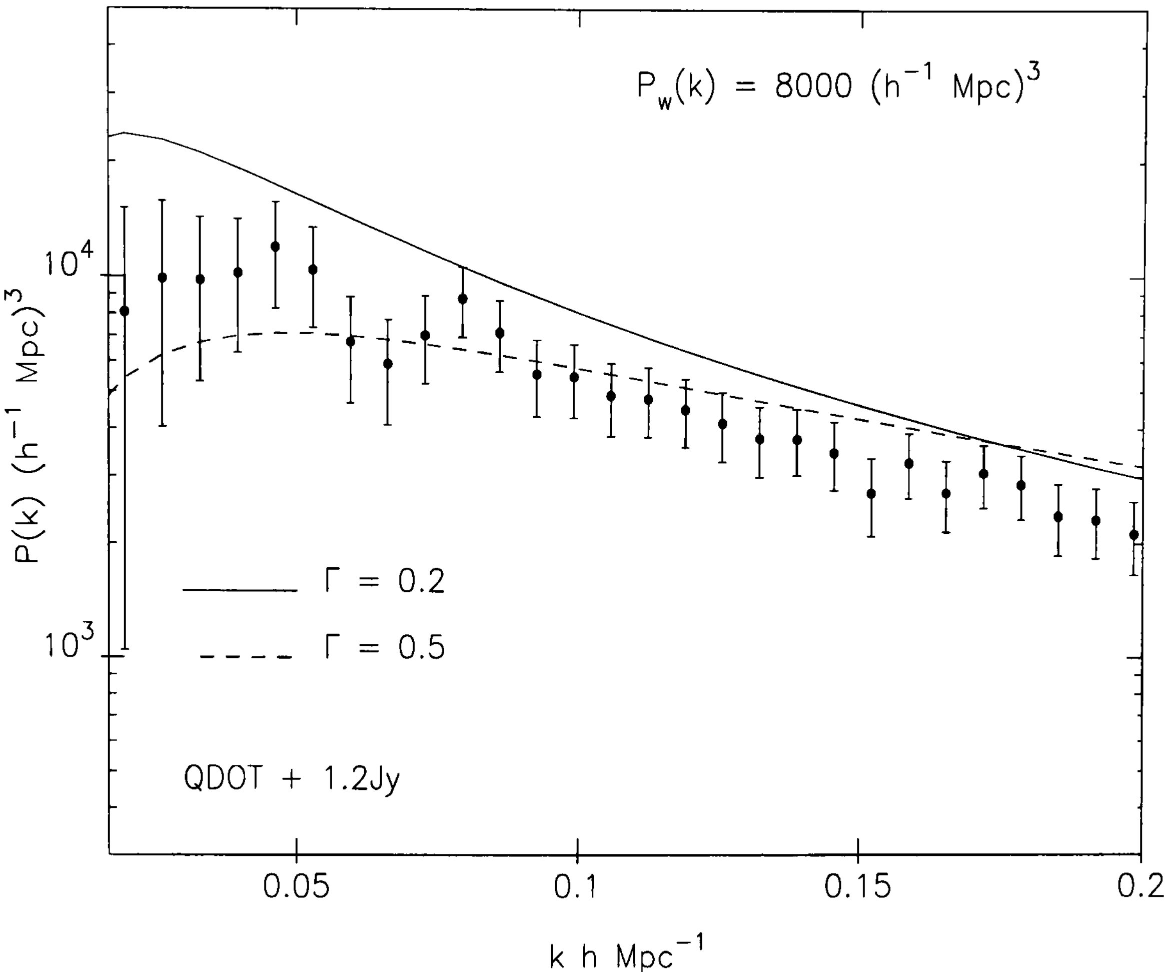


Figure 4.12: The power spectrum of the combined QDOT and $1.2Jy$ surveys. The error bars show one standard deviation and I have adopted $P_w(k) = 8000(h^{-1}\mathrm{Mpc})^3$ in the weight function. The two lines show the linear theory power spectra for two scale-invariant CDM models with $\Gamma = \Omega h = 0.2$ and 0.5. The linear theory curves have been normalised so that the *rms* fluctuation in spheres of radius 8 h^{-1}Mpc is equal to unity.

are now discrepant with the power spectra below $b = 35°$. In conclusion, from Figure 4.10, there is no obvious dependence of the power spectrum on Galactic latitude. A similar test is performed on the new fully sampled IRAS PSCz redshift survey in Chapter 5.

4.6 Summary

The main goal of this chapter has been to compare the power spectra of the $1.2Jy$ and QDOT surveys using the same estimators and weighting schemes. The results show that the QDOT power spectrum is sensitive to a small number of galaxies in the Hercules supercluster. The QDOT power spectrum lies systematically higher than the power spectrum of the $1.2Jy$ survey over the wave number range $0.03 \lesssim k \lesssim 0.08h\mathrm{Mpc}^{-1}$. The power spectrum for the combined surveys plotted in Figure 4.7 is insensitive to galaxies in Hercules and to the weighting scheme and lies well within the 1σ errors of the power spectra for the $1.2Jy$ survey plotted in Figure 4.3. These results are in accord with the counts-in-cells analysis of the $1.2Jy$ and QDOT surveys discussed in E95.

In Figure 4.12 I show the IRAS $1.2Jy$/QDOT combined power spectrum again. The two lines in Figure 4.12 show the linear theory power spectra of two scale invariant CDM models convolved with the window function of the combined surveys (convolution reduces the amplitude of the theoretical power spectra by $\sim 15\%$ at wave numbers $\lesssim 0.1h\mathrm{Mpc}^{-1}$). These curves are computed from equation (7) of Efstathiou, Bond & White (1992), for two values of the parameter $\Gamma = \Omega h$, and are normalised so that the *rms* fluctuation in spheres of radius $8h^{-1}\mathrm{Mpc}$, σ_8, is equal to unity. I do not attempt a detailed comparison with theoretical models here since this requires an analysis of the distortions to the power spectrum caused by peculiar motions (see Kaiser 1987) and of the relative bias between the clustering of IRAS galaxies and the mass distribution. It should be noted, however, that the general shape of the power spectrum plotted in Figure 4.12 resembles that of a low density CDM model with $\Gamma = 0.2$ and $\sigma_8 \approx 0.84$ and that there is some evidence for excess power at wave numbers $k \sim 0.04h\mathrm{Mpc}^{-1}$ compared to the standard CDM model with $\Gamma = 0.5$ normalised to $\sigma_8 \approx 1$. These results are in qualitative agreement with the conclusions of FKP, and with those deduced from the power spectra of optically selected redshift surveys (*e.g.* Park *et al.* 1994). A comparison of the power spectrum of IRAS galaxies with optical galaxies is discussed in Chapters 2 and 5.

Weight	4000	8000	16000
k	$P(k)$	$P(k)$	$P(k)$
0.0198	4605 ± 4459	8066 ± 7020	13322 ± 10908
0.0265	6181 ± 3946	9891 ± 5841	14665 ± 8467
0.0331	7657 ± 3481	9814 ± 4495	12854 ± 6148
0.0397	9240 ± 3325	10235 ± 3901	10640 ± 4852
0.0463	11159 ± 3287	11981 ± 3698	12076 ± 4430
0.0529	10240 ± 2763	10437 ± 3059	10854 ± 3797
0.0595	7335 ± 1907	6784 ± 2106	6118 ± 2657
0.0661	6784 ± 1655	5939 ± 1830	4001 ± 2244
0.0728	7628 ± 1641	7094 ± 1822	5520 ± 2219
0.0794	8302 ± 1574	8822 ± 1825	9117 ± 2308
0.0860	7007 ± 1298	7180 ± 1520	7027 ± 1954
0.0926	5725 ± 1073	5604 ± 1276	5851 ± 1731
0.0921	5609 ± 1006	5502 ± 1208	5886 ± 1661
0.1058	5164 ± 887	4918 ± 1062	4675 ± 1434
0.1124	5040 ± 826	4828 ± 995	4678 ± 1354
0.1190	4747 ± 754	4525 ± 918	4038 ± 1240
0.1257	4452 ± 700	4160 ± 859	4065 ± 1198
0.1323	4048 ± 638	3807 ± 796	3858 ± 1124
0.1389	3796 ± 584	3792 ± 747	4358 ± 1080
0.1455	3389 ± 539	3491 ± 709	4237 ± 1044
0.1521	2797 ± 476	2738 ± 635	2850 ± 930
0.1587	3102 ± 475	3283 ± 636	3699 ± 930
0.1653	2823 ± 437	2738 ± 580	2893 ± 851
0.1720	2940 ± 427	3072 ± 574	3359 ± 839
0.1786	2754 ± 404	2873 ± 547	3153 ± 805
0.1852	2355 ± 368	2368 ± 505	2831 ± 761
0.1918	2238 ± 351	2316 ± 487	2530 ± 728
0.1984	2111 ± 334	2120 ± 463	2197 ± 693
0.2050	1501 ± 296	1422 ± 422	1696 ± 652

Table 4.1: Numerical values for the power spectrum of the combined QDOT and $1.2Jy$ survey for three different values of $P_w(k)$ in the weighting scheme. In the first line, weight refers to the value of $P_w(k)$ used in the weighting scheme (Equation 4.3). Units of k are $h\text{Mpc}^{-1}$ and units of $P(k)$ are $(h^{-1}\text{Mpc})^3$. The points in the power spectrum estimate are correlated, (see equation 2.5.2 of FKP) over a range $\delta k \approx 0.011$ for the weightings shown in this table, i.e. about one and a half times the spacing between points.

Chapter 5

The IRAS PSCz redshift survey and its power spectrum

In this chapter I use the new IRAS PSCz redshift survey to measure the power spectrum of IRAS galaxies. In Section 5.1 I describe the IRAS satellite and its operation. In Section 5.2 I briefly summarize the major scientific results from the redshift surveys of galaxies constructed from the IRAS Point Source Catalogue (PSC). Section 5.3 introduces and describes the newly completed IRAS PSCz redshift survey which is the largest and deepest all-sky IRAS redshift survey conducted to date. Section 5.4 describes the power spectrum analysis of this survey and Section 5.5 presents a comparison between the power spectrum of PSCz galaxies and the power spectrum of optical galaxies in the Stromlo-APM survey. The conclusions of this chapter are presented in Section 5.6.

5.1 The IRAS satellite

The IRAS satellite flew in 1983 and is described in detail in the IRAS Explanatory Supplement (Beichman *et al.* 1988, hereafter ES). Details can also be found in Neugebauer *et al.* (1984) and subsequent Letters in the same journal. The infrared telescope was a f/9.6 Ritchey-Chrétien design with a focal length of 5.5m and an aperture diameter 57cm (the primary mirror had a diameter of 60cm and the field of view had a diameter $63.6'$). The 62 infrared detectors at the Cassegrain focus of the telescope were arranged so that every source crossing the field of view was seen by at least two of the detectors in each of the wavebands (see Figure 1 of Neugebauer *et al.* 1984). The satellite scanned the sky at a constant rate ($3.85' s^{-1}$) along lines of constant ecliptic longitude. A zero-sum filter was used on the raw data (i.e. the flux from each detector as a function of time). This procedure identifies sources while subtracting the background; any source detected with a signal to noise ratio of more than or equal to 3.7 was convolved with a point source template. If the correlation coefficient of this fit to a point source was greater than or equal to 0.87 then the source was retained. The reality of a source is described in terms of 'seconds', 'hours', 'weeks' and 'months' confirmation. Seconds confirmation was achieved if the source was seen by two rows of detectors in the same wavelength band. If the source appeared in two adjacent overlapping scans then it was said to be hours confirmed. Each pair of hours confirming (HCON) scans was repeated within 7 – 11 days to get a weeks confirmation. If there was a third pair of hours confirming scans then the source was months confirmed. In all 72% of the sky was observed with 3 or more hours confirming scans and 95% with two or more hours confirming scans. The minimum requirement, in terms of confirmation, for a source to get into the PSC was that it had two hours confirmations.

The satellite surveyed the sky in four wavelength bands centred at wavelengths 12, 25, 60 and $100\mu m$, with an angular resolution of $\sim 1'$. The galaxies detected by the satellite are consequently mostly spiral galaxies, as ellipticals and SOs do not emit strongly in the far infrared. The sources in the IRAS survey comprise the two-dimensional parent catalogue of many subsequent redshift surveys. The parent catalogue derived from the IRAS data is generally referred to as the Point Source Catalogue (PSC) and contains a list of point sources detected by the satellite along with the flux of the object in each of the four wavelength bands. This list includes many Galactic sources in addition to galaxies, for example, stars, planetary nebulae, and cirrus (cirrus is the name for the emission from interstellar dust on all spatial scales which characterizes the infrared sky at $100\mu m$, and is caused mainly by emission from small grains of graphite). By using the colour information derived from the flux of the object in all four pass bands, we can distinguish galaxies from other objects. In this way galaxy catalogues covering nearly the whole sky have been constructed (see Section 5.3.1 for more details of object selection).

Some areas of the sky are excluded from the catalogue; these areas define the sky-mask of the survey. Regions of the sky that are heavily contaminated by Galactic objects, or that have large amounts of cirrus are excluded from the catalogue, as well as regions that do not have enough hours confirming scans by the satellite (see Section 5.3.1 for definition of the sky-mask for the PSCz redshift survey). Objects which are extended with respect to the IRAS beam may have their fluxes underestimated as the flux is derived via comparison with a point source template. This problem is discussed in Section 5.3.3 and can be solved by a process known as add-scanning.

The catalogues derived from the IRAS satellite survey data have proved invaluable in the field of large scale structure because of their uniform and almost complete sky coverage. Galaxy redshift surveys resulting from the IRAS survey include:

1. Soifer *et al.* (1987) who measured redshifts for 234 galaxies with $60\mu m$ flux $f_{60} > 5.4Jy$ and $|b| > 30°$ in the North.

2. Strauss *et al.* (1992a) who completed a near whole sky redshift survey to a flux limit of $f_{60} > 1.936Jy$, consisting of 2658 galaxies, covering 11.06 sr of the sky.

3. Fisher *et al.* (1995a) who extended the Strauss *et al.* (1992a) survey to the deeper flux limit of $f_{60} > 1.2Jy$. This survey consists of 5339 galaxies with a median redshift of 5800km s^{-1}.

4. Lawrence *et al.* (1995) who went to the deeper flux limit of $f_{60} > 0.6Jy$ (giving a median redshift of 8400km s^{-1}). They followed the sparse sampling strategy suggested by Kaiser (1986) and obtained redshifts for a random sample of 1 in 6 of the galaxies to the above flux limit. The number of galaxies with measured redshifts in their survey was 2184. The sky-mask for this survey was slightly more conservative than for the IRAS $1.2Jy$ survey mentioned above and the unmasked region of the sky was 10.3 sr. This survey is generally referred to as the QDOT redshift survey.

5.2 Major scientific results from the IRAS survey

There is a wealth of published results from galaxy redshift surveys derived from the IRAS satellite survey database and I do not attempt to describe them all here. Below I summarize some of the important results from these surveys.

5.2.1 Luminosity function of IRAS galaxies

Saunders *et al.* (1990) calculated the $60\mu m$ and far infrared luminosity functions of IRAS galaxies. These authors used 2818 galaxies with known redshifts taken from various IRAS surveys and estimated the $60\mu m$ luminosity function using a clustering independent maximum likelihood method. They found that the shape of the luminosity function was not well fit by the usual Schechter form, or by the Gaussian form (Sandage, Binggeli & Tammann 1985). It was well fit by a Gaussian in log(luminosity), which becomes a very flat power-law at low luminosities. The form of the luminosity function and its parameters are those given in Chapter 4 (equation 4.5). At low luminosities, the results of Saunders *et al.* (1990) are in disagreement with previous findings (Yahil 1988, Lawrence *et al.* 1986).

Saunders *et al.* (1990) suggested two reasons for why the luminosity function that they derived is not in agreement with the two power-law luminosity functions found by Yahil and Lawrence *et al.* . The first of these is that the Schmidt estimator (Schmidt 1968), used by Lawrence *et al.* (1986), is biased by large local overdensities. The second reason is that the two power-law fit is not an appropriate parametric form. Saunders *et al.* (1990) found that the curvature in the luminosity function is spread over a large range of luminosities and this cannot be accommodated by two power-laws. This, together with the fact that there are few galaxies of luminosity below $\sim 10^8 h^{-1} L_\odot$, forces the faint end of the luminosity function upwards in the two power-law fit in order to fit the knee of the luminosity function.

Saunders *et al.* (1990) also calculated the luminosity functions of two subsamples of their IRAS galaxies, subdivided by temperature. A dust emission temperature may be assigned on the basis of the $60\mu m$ and $100\mu m$ fluxes and the sample was split into galaxies above and below an assigned temperature of $T = 36$K. Physically, the 'warm' galaxies ($T > 36$K) are active galaxies (starbursts or Seyferts) while the 'cool' galaxies are normal galaxies (see Rowan-Robinson & Crawford 1989 and Section 5.3.1 for discussion of the physical origin of the spectra of IRAS galaxies). The distinction between the two subsamples of galaxies is, of course, not sharp and $T_{em} = 36$K is used only as a nominal boundary between the samples. Saunders *et al.* (1990) found that the same functional form (equation 4.5) provided acceptable fits to both subsamples and that the overall $60\mu m$ luminosity function was dominated by normal cool galaxies, for luminosities below $\sim 5 \times 10^{10} h^{-2} L_\odot$, and by warm galaxies above this luminosity.

5.2.2 Evolution in the density of IRAS galaxies

Saunders *et al.* (1990) presented a method for determining the variation of density with distance in a sample, under the assumption that the form of the luminosity function is independent of distance. From this analysis, they concluded that there is strong evolution in the population of IRAS galaxies. The luminosity function is assumed to be

$$\phi_z = g(z)\phi_0 \left[\frac{L}{f(z)}\right], \tag{5.1}$$

where $g(z)$ parameterises the density evolution and $f(z)$ parameterises the luminosity evolution. The samples used by Saunders *et al.* (1990) were not deep enough to distinguish between these two forms of evolution so solutions were calculated in terms of pure density evolution ($f(z) = 1$, $g(z) = (1 + z)^p$). Luminosity evolution may be added by assuming that luminosities decay exponentially with time:

$$L_z = L_0 exp\frac{2}{3}q\left[1 - (1+z)^{-\frac{3}{2}}\right]. \tag{5.2}$$

Saunders *et al.* (1990) found that

$$p + 2.1q = 6.7 \pm 2.3. \tag{5.3}$$

Evolution is therefore explained, either in terms of pure density evolution with $p = 7 \pm 2$, or pure luminosity evolution with $q = 3 \pm 1$.

These results are in strong contrast with the results of Fisher *et al.* (1992), who found no evidence for evolution when analysing the $1.2Jy$ IRAS redshift survey. Fisher *et al.* (1992) pointed out, however, that the analysis of the evolution of the galaxy population is very sensitive to a number of unknowns. These unknowns include the model for the spectral energy distribution of IRAS galaxies, the background cosmological model, the assumption of homogeneity in the sample, and possible errors in the IRAS fluxes. Fisher *et al.* (1992) estimated that if the flux errors are Gaussian distributed with an absolute mean error of $\sim 0.15Jy$, then the evolution found by Saunders *et al.* (1990) would disappear. However as described in Section 4.5 of Chapter 4, Lawrence *et al.* (1995) have estimated the absolute error to be no more than $0.06Jy$.

The galaxy data set used in the above analysis of Saunders *et al.* (1990) was a preliminary version of the QDOT redshift survey. The final version of the survey has been used by Lawrence *et al.* (1995), who have re-calculated the evolution parameter p. The preliminary version of the QDOT catalogue was afflicted by a number of errors, the most serious being a wavelength calibration error which affected about 200 redshifts. These 200 redshifts were too large by 2800km s^{-1}. Lawrence *et al.* (1995) also removed some objects at high redshift/high luminosity from the final catalogue because of unconvincing redshift measurements. The Lawrence *et al.* (1995) analysis of the final QDOT redshift survey leads to a value of $p = 5.6 \pm 2.3$. If a rough correction is made for non-linearity of PSC fluxes (see Section 4.5, Chapter 4), this becomes $p = 4.5 \pm 2.3$. The result is therefore still indicative of strong evolution in the IRAS galaxy population.

5.2.3 Topology of the IRAS density field

Saunders *et al.* (1991) analysed the density field mapped by the IRAS QDOT survey and identified the dominant superclusters and voids. They also found definite evidence for non-Gaussianity in the density field on scales as large as 20 h^{-1}Mpc. Moore *et al.* (1990) analysed the topology of the QDOT redshift survey to a depth of 200 h^{-1}Mpc and concluded that the galaxy distribution is consistent with having grown from initially Gaussian density fluctuations. The topology of the density field of the survey was found to be sponge-like, i.e. the high and low density regions are topologically similar. The authors also measured the effective slope of the power spectrum of galaxy fluctuations on scales 10 h^{-1}Mpc – 50 h^{-1}Mpc. and found $P(k) \propto k^n$ with $n \sim -1$.

5.2.4 Power spectrum and counts in cells analyses

The power spectrum of IRAS galaxies has been determined from both the IRAS $1.2Jy$ catalogue (Fisher *et al.* 1993) and from the QDOT survey (Feldman, Kaiser & Peacock 1994). These two estimates were not compatible, probably due to a statistical fluctuation in the QDOT catalogue (see Chapter 4). Other explanations have been put forward, however. Hamilton (1995) suggested that the discrepancy is due to a flux calibration error which varies across the sky, causing the IRAS PSC to have a varying depth, dependent on angular position. This hypothesis is discussed in detail in Section 4.5 of Chapter 4. Later in this chapter (Section 5.3.3) I discuss the flux accuracy of the PSC. In Chapter 6, I show that there is no detectable source of excess variance, uncorrelated with the galaxy distribution in the PSCz redshift survey (as would be the case if the PSC fluxes where non-uniform across the sky). Of course, if the problems with the QDOT survey were caused by variations in the effective depth of the survey across the sky, then we would expect to detect the excess variance much more clearly when using the fully sampled PSCz survey.

Efstathiou *et al.* (1990) analysed the counts in cells for the QDOT survey. They found results which disagreed with the standard CDM (SCDM) model predictions, for cells of size 30 and 40 h^{-1}Mpc. The values of σ^2 predicted by SCDM, for cells of this size, were found to be too small at the 95% confidence level. Some of the excess variance over the predicted SCDM values was probably caused by the statistical fluctuation in the QDOT survey referred to above (see Efstathiou 1995 and Chapter 6 for more discussion of this point).

Bouchet *et al.* (1993) analysed the counts in cells probability distribution function of galaxies in the IRAS $1.2Jy$ redshift survey. They derived from this the 2, 3, 4 and 5 point correlation functions for IRAS galaxies. They found some evidence, on small scales, that the brighter IRAS galaxies are more strongly clustered than the fainter ones. The measured skewness of the distribution ($S_3 = \overline{\xi_3}/\overline{\xi_2^3}$, where the bar signifies a volume averaged quantity) was approximately scale independent over scales where $0.1 < \overline{\xi_2} < 10$, and on large scales is consistent with the hypothesis that structure grew via gravitational instability of initially Gaussian density fluctuations. Bouchet *et al.* (1993) found that the measured correlations are consistent with the scaling relation $\overline{\xi_N} \propto \overline{\xi_2^{N-1}}$.

5.2.5 Dipole analyses

Rowan-Robinson *et al.* (1990) analysed the IRAS dipole using the data from the QDOT redshift survey. They found that all the matter causing our motion is enclosed within 100 h^{-1}Mpc and that the direction of our motion as determined from the IRAS dipole is consistent with that of the microwave background dipole. They inferred a value of $\beta = \Omega^{0.6}/b = 0.7^{+0.3}_{-0.2}$. Strauss *et al.* (1992c) also carried out a dipole analysis using the $1.2Jy$ survey. They found that the direction of the acceleration of the Local group lies 18° to 28° from the CMB dipole and that the value of β lies between 0.4 and 0.85 (1σ range). Strauss *et al.* (1992c) found that the dipole direction had converged within 4000km s^{-1} of the Local Group, although there was an additional contribution to the dipole moment from galaxies between 17000km s^{-1} and 20000km s^{-1} which was aligned with the measured dipole. The significance of this contribution is only 2σ and they concluded that deeper samples would be necessary to confirm its existence, in which case β would decrease.

5.2.6 Peculiar velocity fields

Kaiser *et al.* (1991) used the QDOT redshift survey to make predictions of the peculiar velocity field of the galaxies from the density field which they measured. This peculiar velocity field was compared with the data that exists for galaxy peculiar velocities (Burstein 1990), and assuming the density and velocity fields are governed by gravitational instability they estimated a value for the parameter β of 0.86 ± 0.16.

The IRAS redshift survey data has also been used extensively in POTENT analysis. POTENT analysis (Bertschinger & Dekel 1989, Dekel, Bertschinger & Faber 1990, Bertschinger *et al.* 1990) is a method for reconstructing the density field of galaxies from the measured radial peculiar velocity field. The radial peculiar velocity is obtained via distance independent means (e.g. the Tully – Fisher relation for spirals (Tully & Fisher 1977) or the $D_n - \sigma$ relation for ellipticals (Djorgovski & Davis 1987, Dressler *et al.* 1987b, Lynden-Bell *et al.* 1988)). The radial peculiar velocity field is integrated to obtain the velocity potential field

$$\Phi(r) = -\int_0^r u\,(r', \theta, \phi)\, dr', \tag{5.4}$$

where $u\,(r', \theta, \phi)$ is the measured peculiar velocity field. With knowledge of the potential field $\Phi(r)$, the three-dimensional peculiar velocity field can now be obtained

$$v(r) = -\nabla\Phi. \tag{5.5}$$

Equation (5.5) is valid if the velocity field is curl free. Linear theory relates the peculiar velocity field $v(r)$ to the density field $\delta(r)$, and this relation has been modified to account for mild non-linearities (see Dekel 1994 for an account of the different quasi-linear approximations that may be made). The density field can be reconstructed, for instance, via the equation

$$\delta(r) = \left(\left|\mathbf{I} - \left(\frac{\Omega_0}{b}\right)^{-1} \frac{\partial \mathbf{v}}{\partial \mathbf{r}}\right| - 1\right), \tag{5.6}$$

$\mathbf{I}$ being the unit matrix (Nusser *et al.* 1991). The derived density field is then compared with the smoothed density field measured directly from an IRAS redshift survey and the ratio of the two gives a measure of the value of β. These methods are however subject to a number of statistical and systematic uncertainties, the most serious being sampling gradient bias (inhomogeneous sampling of the velocity field) and Malmquist bias (homogeneous and inhomogeneous). Detailed discussion of these effects and how to account for them can be found in Dekel *et al.* (1993) and Strauss & Willick (1995).

5.2.7 Redshift-space distortion analyses

The IRAS redshift surveys have also been used to analyse the redshift-space distortions caused by galaxy peculiar motions. Hamilton (1993a) used the $2Jy$ survey to measure the correlation function $\xi(\sigma, \pi)$ as a function of two variables, σ – the separation of galaxy pairs perpendicular to the line of sight, and π – the separation of the pair parallel to the line of sight. The redshift-space distortions show up as the departures of $\xi(\sigma, \pi)$ from isotropy. On small scales, random peculiar motions cause the correlation function to be enhanced along the line of sight, so that contours of $\xi(\sigma, \pi)$ are elongated along the π axis. On large scales the (linear) coherent velocity field, caused by infall towards clusters and outflow from voids, causes the correlation function to be enhanced perpendicular to the line of sight. In this case, the contours of $\xi(\sigma, \pi)$ are elongated along the σ axis. Hamilton (1993a) used angular moments of $\xi(\sigma, \pi)$ to determine a value for the parameter $\beta = 0.66^{+0.34}_{-0.22}$, from the $2Jy$ survey. Fisher *et al.* (1994) have also measured $\xi(\sigma, \pi)$. They fitted for the value of the small scale peculiar velocity dispersion between pairs of galaxies, σ, finding $\sigma = 317^{+40}_{-49}$ km s^{-1} at a pair separation of $r = 1\ h^{-1}$Mpc. They also derived the mean relative streaming velocity between pairs of galaxies, $\langle v_{12} \rangle$ and found $\langle v_{12} \rangle = 167^{+99}_{-67}$ km s^{-1} at a pair separation $r = 4\ h^{-1}$Mpc, and $\langle v_{12} \rangle = 109^{+64}_{-67}$ km s^{-1} at $r = 10 - 15\ h^{-1}$Mpc. Linear perturbation theory was then used to derive a value of the parameter $\beta = 0.45^{+0.27}_{-0.18}$, measured on scales $10 - 15\ h^{-1}$Mpc.

Both the QDOT and the IRAS $1.2Jy$ redshift surveys have been used by Cole, Fisher & Weinberg (1995) to estimate the parameters β and the one-dimensional small scale velocity dispersion of IRAS galaxies. They used the relation between real and redshift-space power spectra given in equation (2.37) of Chapter 2, which includes the effects of the small scale incoherent velocity field. They measured $\beta = 0.52 \pm 0.13$ and $\beta = 0.54 \pm 0.3$ from the $1.2Jy$ and QDOT surveys respectively. The methods used to obtain these measurements are discussed in more detail in Chapter 7.

Heavens & Taylor (1995) used a spherical harmonic decomposition of the density field of the $1.2Jy$ survey and formulated a theory of redshift-space distortions which is exact to first order in the density perturbation δ. This formulation correctly accounts for the mixing of modes in k space caused by the distortion, and for the finite extent of the survey. The amplitude of the power spectrum and the value of β are free parameters obtained via maximum likelihood methods. The amplitude is expressed in terms of σ_8 (which in this case is not a linear theory extrapolation, but the present day value). They obtained a maximum likelihood value of $\beta = 1.1 \pm 0.3$ and a $\sigma_8 = 0.68 \pm 0.05$. Ballinger, Heavens & Taylor (1995) performed a similar analysis, but obtained maximum likelihood values for β and for the power spectrum in real-space, characterized by the value of $P(k)$ at six discrete bins in k. In this case, the authors obtained a maximum likelihood value of $\beta = 1.0 \pm 0.3$ using the $1.2Jy$ data. The approach of Heavens & Taylor (1995) is described in more detail in Chapter 7 and is applied to the PSCz redshift survey.

Of course, much useful work has also been done with the two-dimensional distribution of IRAS galaxies on the sky, but I do not describe this in any detail here (see, for example, Scharf *et al.* 1992, Lahav, Nemiroff & Piran 1990, Harmon, Lahav & Meurs 1987, Scharf & Lahav 1993).

5.3 The new IRAS PSCz redshift survey

The work in this and subsequent chapters is based on data from the new IRAS PSCz redshift survey. This survey was conducted as a follow on from the QDOT redshift survey by a collaboration of people from many institutes (the people involved in this project apart from myself are Will Saunders (ROE), Will Sutherland, George Efstathiou (Oxford), Steve Maddox (RGO), Richard McMahon (IoA), Simon White (MPI Munich), Sebastian Oliver, Oliver Keeble, Michael Rowan-Robinson (Imperial College) and Carlos Frenk (Durham)) . The aim was to determine redshifts for all galaxies in the Point Source Catalogue (PSC) to the flux limit of $f_{60} > 0.6Jy$. The survey is now basically complete and an outline of its construction is given in Saunders *et al.* (1995). About 4600 new redshifts were measured in the course of the project, the rest being taken from the literature or obtained via private communications.

Strategies for completing a large all-sky redshift survey must be carefully considered. Telescope time is at a premium, and we must decide how to make the best use of it given that large redshift surveys traditionally take many nights to complete (although the recent advent of multi-fibre systems such as FLAIR (Parker & Watson 1995), the 2dF system at the AAT (Ellis 1993), and the Sloan Digital Sky Survey (Gunn & Weinberg 1995) will speed the process up considerably). There is a compromise between going to a faint flux limit and collecting a large number of redshifts over a wide area. Pencil beam surveys (e.g. Broadhurst *et al.* 1990) are an extreme example of the former. Because of the geometry of these surveys, they are not generally useful for studies of large scale structure or dynamics. Shallower, all-sky surveys such as the IRAS $2Jy$ and $1.2Jy$ surveys are more useful and by sparsely sampling the galaxy distribution we can make a deeper sample for a given amount of telescope time. Kaiser (1986) first described how sparse sampling could be used and calculated the optimum sampling rate (see discussion below). This strategy has been implemented in several surveys (Loveday *et al.* 1995a, Lawrence *et al.* 1995, Ratcliffe *et al.* 1996). The argument for sparse sampling is as follows. Peebles (1980) showed that the uncertainty, ϵ_ξ, in the estimate of the correlation function ξ in the regime where ξ is small compared to unity, is given by

$$\epsilon_\xi(r) \simeq \frac{1 + 4\pi\overline{n}J_3}{\sqrt{N_p}}, \tag{5.7}$$

where $\overline{n}$ is the mean number density of objects in the catalogue,

$$J_3 = \int_0^r r'^2 \xi(r')\, dr', \tag{5.8}$$

and N_p is the number of independent pairs of objects used in the estimation of the correlation function. This equation shows that even though one may have a large number of galaxies, N, in the survey, the number of independent points (which govern the degree of uncertainty in the measured statistic) is less than N, and is given by N/N_c where $N_c = 1 + 4\pi\overline{n}J_3$. If we sample the distribution at a rate f then the error is independent of the sampling rate if $4\pi\overline{n}fJ_3 \gtrsim 1$. Thus for a give amount of telescope time, the best strategy for measuring the correlation function or power spectrum, is to sparsely sample the distribution such that $4\pi\overline{n}fJ_3 \simeq 1$ and to sample a larger volume of space. It must be kept in mind, however, that since the selection function of a flux limited survey is radially dependent, the optimal rate at which to sample the galaxy distribution is not well defined.

Of course, a fully sampled survey is desirable for studies of the local density field and topology. In topological studies it is necessary to smooth the density field. On large scales the intrinsic clustering pattern is weak and shot noise makes it difficult to detect. However, if we smooth heavily to alleviate the shot noise, we will Gaussianise the distribution.

The new PSCz fully sampled survey will of course have a larger effective volume than QDOT as the shot noise will be considerably reduced. The power spectrum will be reliable to smaller wave numbers, and we may hope to detect the turn-over in the power spectrum. We must be careful in interpreting the power spectrum results at small wave numbers because of the various biases that were first discussed in Chapter 2 (see Section 5.4). A turn-over in the power spectrum, expected in CDM-like models, has not been convincingly detected to date in the power spectrum of galaxies. There are various limits on the behaviour of the power spectrum in this regime. For instance, the Ballinger, Heavens & Taylor (1995) analysis referred to in Section 5.2.7 uses maximum likelihood methods to constrain the value of β as well as the IRAS galaxy power spectrum in real-space. They do not detect power on scales with wave number $k \lesssim 0.03\ h\ \mathrm{Mpc}^{-1}$ but have upper limits on the power at larger scales. If all other parameters in the analysis are held fixed, the fit probability of a power spectrum which continues as a power-law is ten times less than that for the Peacock & Dodds (1994) fit to the power spectrum which turns over to a Harrison-Zeldovich spectrum at wave number $\sim 0.04\ h\ \mathrm{Mpc}^{-1}$. The results from the Las Campanas survey (Shechtman *et al.* 1995) also show no indication of a turn-over. The power spectrum (Lin 1995) shows a definite flattening on scales with wave numbers less than $k \sim 0.07\ h\ \mathrm{Mpc}^{-1}$ but to within the 1σ errors, the turn-over is not seen.

In addition to analysis of the power spectrum, we will use the PSCz survey to measure redshift-space distortions (see Chapter 7), and to compare the clustering distributions of galaxies of different luminosities and morphological types (see Chapter 6). Other projects planned by members of the PSCz team include using the data set for a detailed comparison of the peculiar velocity and density fields, topological studies, dipole analysis, application of the cosmic virial theorem, and analysis of redshift-space distortions in terms of $\xi(\sigma, \pi)$.

5.3.1 Selection criteria and sky-mask for the PSCz redshift survey

I will now briefly review the selection criteria for the PSCz redshift survey. The initial parent catalogue is the QMW IRAS galaxy catalogue (QIGC) (Rowan-Robinson *et al.* 1991b), derived from the PSC. This was also used as the parent catalogue for the QDOT redshift survey but has now been modified in a number of ways to produce the parent catalogue for the PSCz survey. For the purposes of the IRAS sky survey the sky was divided into 41,167 $1^\circ \times 1^\circ$ 'lune' bins defined by ecliptic coordinates. Certain regions of the sky (i.e. some of these bins) are excluded from the redshift survey due to reasons enumerated below. The bins which are excluded define a sky-mask and the PSCz mask (hereafter MASK I) is defined as follows:

1. IRAS coverage gaps i.e. areas of the sky which failed to achieve 2 hours confirmed coverage (about 3% of the sky see ES).

2. Areas which have a high source density at 12 or $25\mu m$. These areas are the ones where there is a high density of stellar sources and so galaxies cannot be identified reliably. Areas which have a high source density at $60\mu m$ present problems for the PSC extraction software, and consequently the parent catalogue is also incomplete in these regions. The QDOT mask also excluded areas of high source density at $100\mu m$, we now include these regions as we can check on the photographic plates or APM raster scans of the plates to see if the source in the region is a galaxy or not.

3. Areas with $I_{100} > 22.5MJy/sr$. Above this threshold, dust extinction is very large, making galaxy identification and spectroscopy very uncertain.

4. Areas within 2.5° of the SMC with $I_{100} > 5MJy/sr$, and within 5° of the LMC with $I_{100} > 10MJy/sr$.

5. Areas suffering from noise lagging. The noise estimator used in the IRAS observations lagged behind any change in noise. The lag was approximately $25'$, $25'$, $50'$ and $50'$

for the $12\mu m$, $25\mu m$, $60\mu m$ and $100\mu m$ bands respectively. However in regions where there was a very steep gradient in the noise (i.e. near the Galactic plane) the noise estimator performed badly. The noise was severely underestimated as the scan approached the plane and overestimated as the scan left the plane, resulting in incompleteness and unreliability in these regions. Some small high Galactic latitude regions are also adversely affected by noise lagging (see ES).

The catalogue covers about 84% of the sky (10.56 sr). The completeness of the catalogue in the 2HCON regions of the sky is not guaranteed to the flux limit that we use ($0.6Jy$). This is basically because an object has a greater chance of satisfying the flux quality criterion (e.g. two hours confirmed detections) if it lies in a region of sky which was covered three times than if it lies in a region only covered twice. We try to correct for this incompleteness by adding in to the two-dimensional parent catalogue 1HCON sources with galaxy like colours, where there is a corresponding source in the Faint Source Survey (Moshir *et al.* 1992). The estimated completeness in the 2HCON regions is 98.5% (Saunders *et al.* 1995). The distribution of masked bins on the sky is shown in Figure 5.1 (this plot was produced by Will Saunders).

To extract the galaxies from the parent two-dimensional catalogue, there are several criteria which are examined for each source. These are based on the work of Rowan-Robinson & Crawford (1989) who model the spectrum of IRAS galaxies with three components, a cool disc component (peaking at $100\mu m$), a warm starburst component (peaking at $60\mu m$) due to the optically thick dust clouds surrounding regions of star formation, and a hot Seyfert component (peaking at $25\mu m$) and due to emission from dust clouds surrounding a power-law continuum source. If the source satisfies the following criteria it is classified as a galaxy:

1. $Log(S_{60}/S_{25}) > -0.3$. This is not satisfied for the majority of stellar objects, but no catalogued galaxy violates this condition. If the source is not detected in the $25\mu m$ band then the quoted upper limit is used.

2. $Log(S_{25}/S_{12}) < 0.1$. If the source does not satisfy this condition, but is identified with a catalogued galaxy, the source is retained. Again, if the object is not detected at $25\mu m$ then the upper limit is used. This condition mostly excludes planetary nebulae.

3. $Log(S_{100}/S_{25}) > -0.3$. This excludes some stars and planetary nebulae but not galaxies.

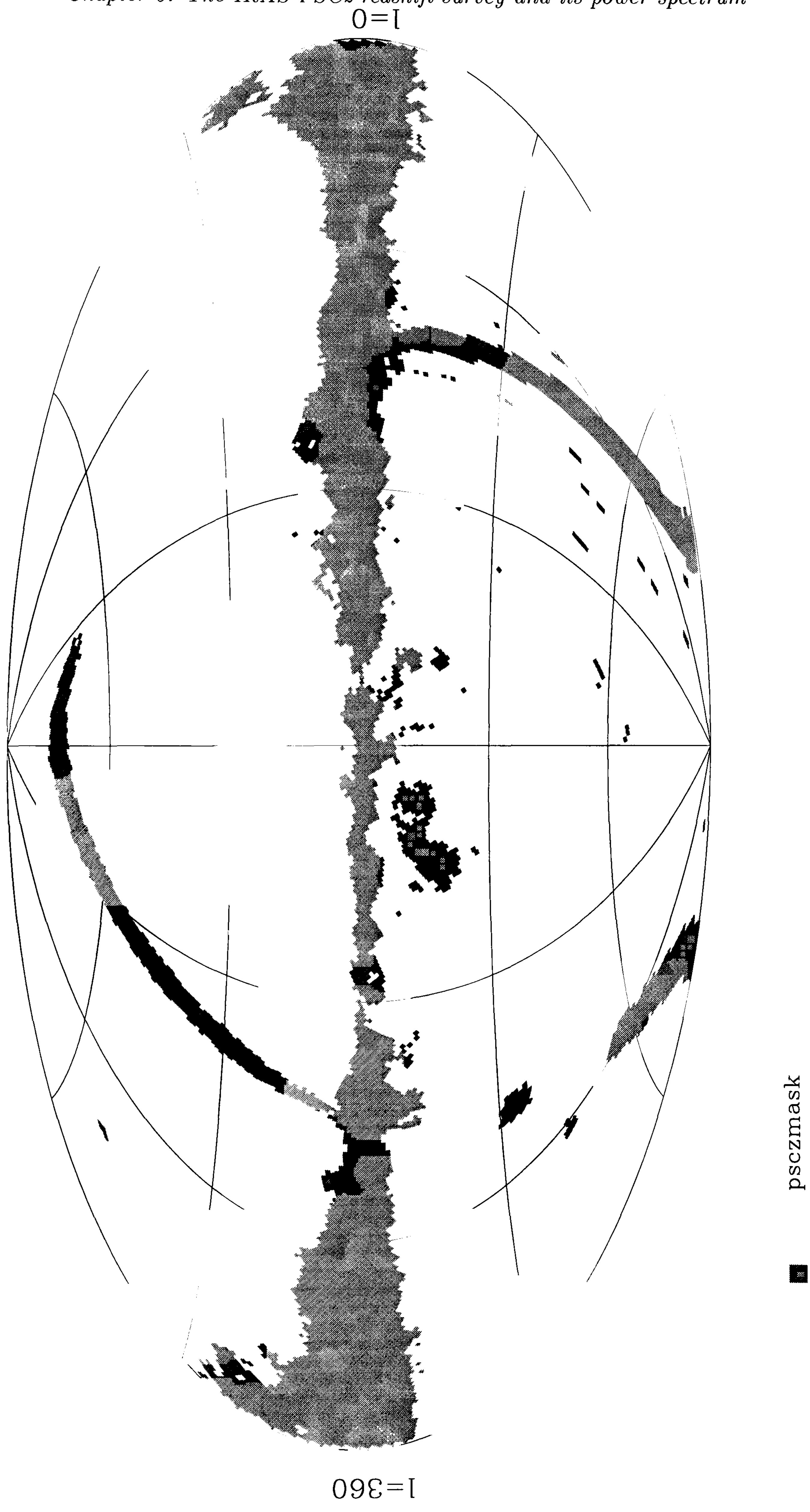


Figure 5.1: The sky-mask (MASK I) used for the PSCz redshift survey, as defined in the text.

4. $Log\,(S_{60}/S_{12}) > 0$. This removes more stars.

5. $Log\,(S_{100}/S_{60}) < 0.6$. If the source is identified with a catalogued galaxy then it is retained regardless of whether it satisfies this criterion. Again if there is no measured flux at $100\mu m$ then S_{100} is replace by its upper limit. This removes mainly cirrus sources. We must be careful with this criterion since a number of nearby low-luminosity galaxies have colours similar to cirrus and we do not want to exclude these galaxies. This point is discussed in Rowan-Robinson *et al.* (1991b). From previous work (Rowan-Robinson, Helou & Walker 1987), which notes a correlation between the ratio of far infrared luminosity to optical luminosity and the $100\mu m/60\mu m$ colour, Rowan-Robinson *et al.* (1991b) estimate that no cool normal galaxies of the type which would be excluded by the criterion above would have a magnitude $m_{pg} > 15.5$. This means that the excluded galaxies will be found by cross-correlation with optical galaxy catalogues (the University of Lyons LEDA database was used for this purpose), and can thus be retained in the sample. The only exception to this is at low Galactic latitudes, where optical galaxy catalogues are affected by extinction, and in the region $-2.5° > \delta > -17.5°$ where the optical galaxy positions where taken from the *Morphological Catalogue of Galaxies* (MCG, Vorontsov-Velyaminov & Krasnogorskaja 1962, Vorontsov-Velyaminov & Arhipova 1963 to 1974), and are not accurate. In this case an IRAS source may fail to be matched with its corresponding optical source. Also, there are cases where, due to cirrus contamination, a galaxy may be assigned a spuriously high value of $S_{100\mu m}$. Saunders (1987) estimates that the incompleteness due to all of these effects is $< 0.7\%$.

6. The last criterion is that individual structures within or associated with Local Group galaxies are excluded from the catalogue.

The distribution of the PSCz galaxies on the sky is shown in Figure 5.2.

5.3.2 Redshift measurement and errors

The redshifts measured in the course of the project were taken mainly at the INT, AAT and CTIO 1.5m telescopes. Because the main consideration was speed ($\sim$ 100 redshifts could be measured in one night) we used low dispersion spectrographs. For example, the Faint Object Spectrograph on the INT has a wavelength range of 5000 – 10500Å and a dispersion 10.7Å/pixel, with a resolution of $\sim$ 20Å. The resolution of the spectrograph

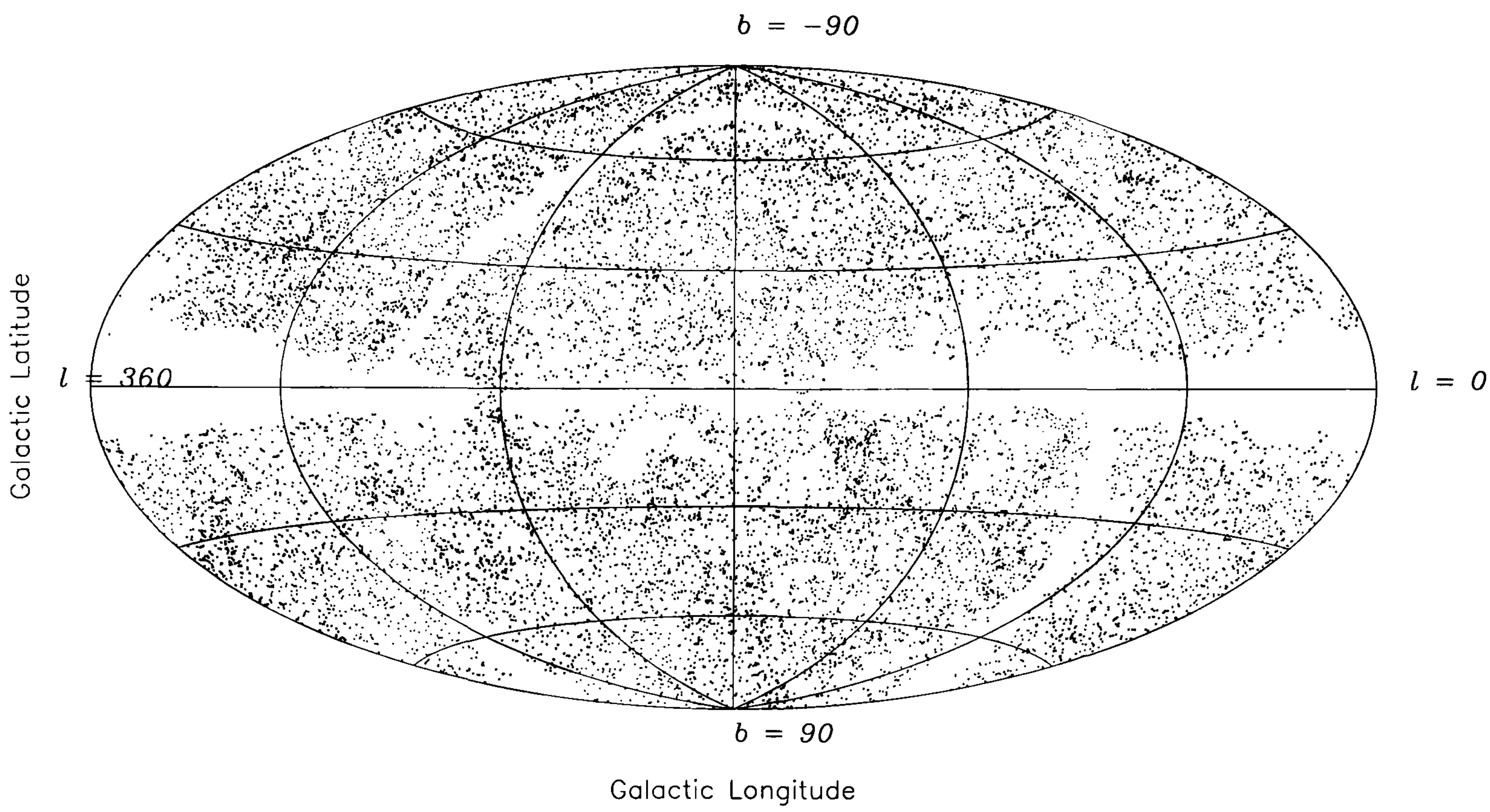


Figure 5.2: The distribution of the PSCz galaxies on the sky.

used on the AAT was also of this order and that on the CTIO telescope was slightly better, $\sim$ 8Å. Each spectrum took approximately 100 – 200s to acquire. The data reduction was undertaken by Sebastian Oliver and Oliver Keeble at Imperial College. The redshift was obtained from the spectrum by a fit of a blend of the H_α and NII lines and the SII line in the spectrum. The total error on the measured velocities is $\sim$ 100 – 150km s^{-1}. Figure 5.3 shows the $N(z)$ distribution for the PSCz redshift survey. The median redshift of galaxies in the survey is $\sim$ 0.03, but there is a tail of galaxies out to redshifts as large as $\sim$ 0.2. The solid curve shows the predicted $N(z)$ using the selection function described in Section 5.4.

5.3.3 Fluxes and flux accuracy

The IRAS satellite measures the flux of a source F in any given band. This is a multiplication of the spectral energy distribution of the source, f_ν, by the response of the detector, R_ν, in the band:

$$F = \int f_\nu R_\nu d\nu. \tag{5.9}$$

The point source catalogue quotes flux densities $f(\nu_0)$ in Jansky, derived by integrating the above equation assuming a spectral energy distribution $f_\nu \propto \nu^{-1}$. The quoted errors on the flux densities from ES are log-normally distributed and have a width $\sigma \sim \pm 10\%$.

Lawrence *et al.* (1995) have also measured the flux errors (both absolute and frac-

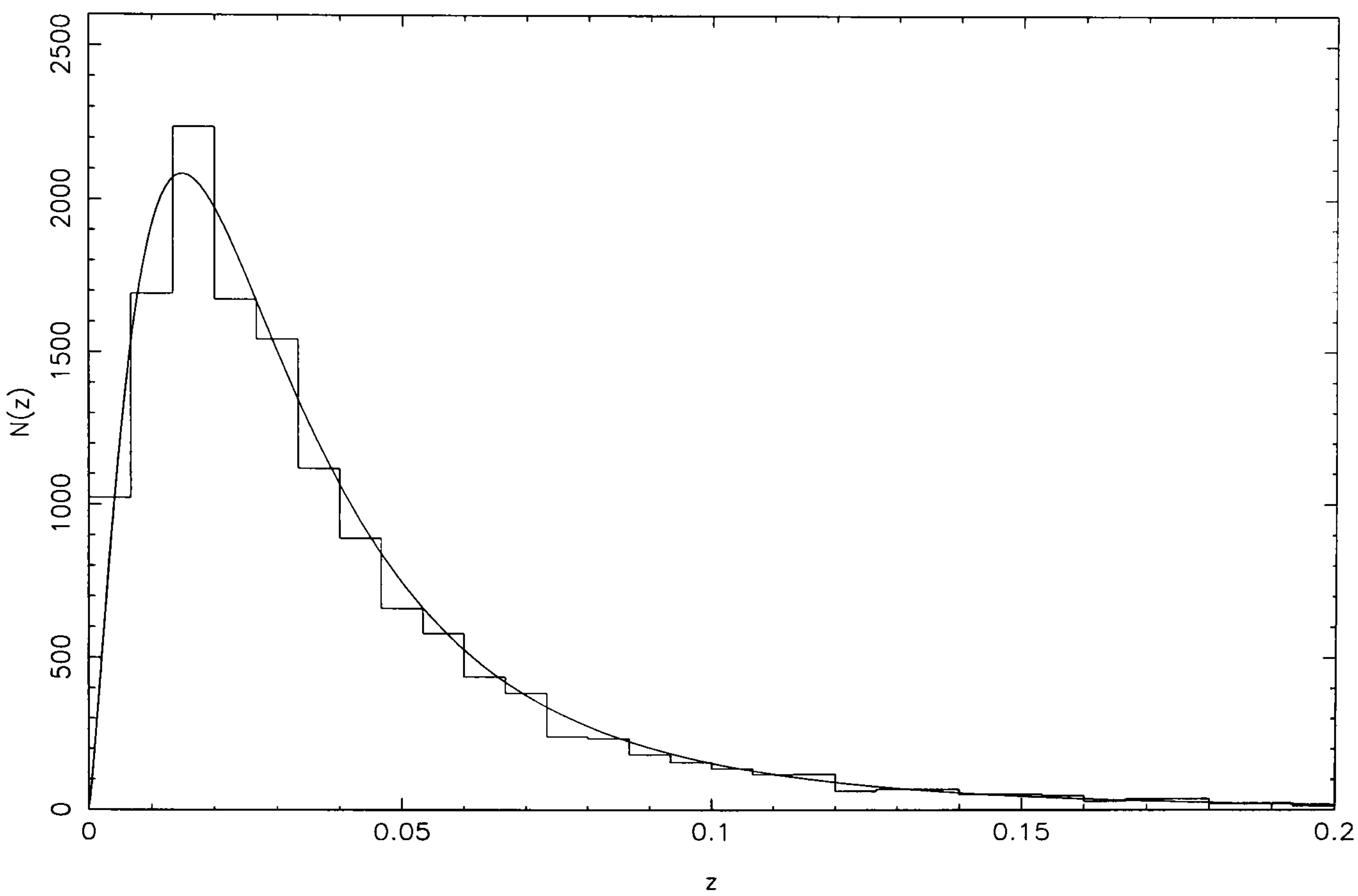


Figure 5.3: The $N(z)$ distribution of the PSCz galaxies. The solid curve shows the predicted $N(z)$ using the selection function of Mann, Saunders & Taylor (1996) with parameters derived to fit the PSCz redshift survey data – see Section 5.4.

tional) as described in the previous chapter. They find that the fractional flux error is of the order of 10% and the absolute error is $\sim 0.05 - 0.06Jy$. They also find some evidence (a 3σ detection) for non-linearities in the PSC fluxes. They conclude that faint sources may have their fluxes overestimated by about 5%.

As a default measure of the flux density, we use the quoted PSC flux. However, there are circumstances in which this is an underestimate of the true flux. If the observed object is larger than the IRAS beam, the estimate of the flux obtained by using a point source template will be an underestimate of the true flux of the object. Saunders *et al.* (1995) estimate the expected ratio of the measured PSC flux of a galaxy to the real flux of the source as a function of galaxy scale length (assuming an exponential distribution of surface brightness). The PSC flux is a good measure of the flux provided that the scale length is less than $0.2'$. Rice *et al.* (1988) show that the infrared scale length is about half of the optical scale length. From this we can expect the PSC flux to be an adequate measure if the optical diameter of the source is less than $2.4'$ (since stated optical diameters generally correspond to about 6 scale lengths).

Galaxies which are larger than $\sim 2.4'$ have been 'add-scanned' to obtain a more accurate flux measure. Add-scanning is a process in which all the individual scans in a given wavelength band, which passed within 1.5′ of the object, are combined and a median scan is calculated. A quadratic baseline is then fitted to this median and the new estimate of the flux density is the total flux density between the zero crossings of the baseline. Saunders *et al.* (1995) show a plot of the ratio of the PSC flux to add-scan flux as a function of the optical diameter of the galaxies. This ratio tends to a constant at small optical diameters but the constant is not unity because of another intrinsic problem with the measurement of flux from the IRAS galaxies. Nearby pairs of galaxies will be resolved by the IRAS beam, however at larger distances, pairs of galaxies will appear as a single extended object and at still larger distances, will appear as one unresolved object. There is no obvious way to correct this problem. However, provided the measure of flux used is uniform over the sky, then the catalogue will be perfectly adequate for the purposes of measures of large scale structure. Problems would arise if the flux were non-uniform, in the sense that it depended on position on the sky or radial depth.

Another potential worry, discussed in detail by Strauss *et al.* (1990), is the hysteresis effect. The detectors on board the IRAS satellite respond to high flux density by becoming temporarily more sensitive. This is especially problematic when the IRAS beam scans across the Galactic plane. It is expected that a source observed before the satellite scans across the plane will have a smaller measured flux than if the same source had been observed after the satellite scans across the Galactic plane. Strauss *et al.* (1990) demonstrated that to their limiting flux ($1.9Jy$) this effect was small enough to be entirely negligible. As far as I know it has not been examined for the fainter flux limit of $0.6Jy$. In the previous chapter (Chapter 4, Figures 4.10 and 4.11) I showed plots of the power spectrum of the QDOT survey for various cuts in Galactic latitude. The power spectrum amplitude is fairly robust to the cut taken. The same test is applied in Section 5.4 of the current chapter for the PSCz survey. These tests indicate that any variations in homogeneity as a function of Galactic latitude are probably not large enough to cause any perturbation in the measurement of large scale structure.

As mentioned above we have supplemented the regions of the sky which have 2HCON with additional sources to boost the completeness of the catalogue in these areas, but there is still a worry that incompleteness in these regions will bias the large scale structure statistics. We can test whether the 2HCON regions are introducing any problems by defining a more restrictive mask (hereafter MASK II) than the one plotted in Figure 5.1 (MASK I). MASK II leaves 72% of the sky unmasked (compared to 84% for MASK I) and has the same definition as MASK I except that bins are excluded if they have $I_{100} > 12.5 MJy$ per steradian (instead of 22.5 for MASK I). 2HCON regions with $|b| < 10°$ where there is no corresponding Faint Source Catalogue from which to select supplementary sources are also excluded.

5.4 Power spectrum analysis of the PSCz redshift survey

The power spectrum analysis in this chapter follows the methods used in Chapters 2, 3 and 4 (see Feldman, Kaiser & Peacock 1994). Figure 5.3 shows the $N(z)$ distribution for the PSCz survey, together with the predicted $N(z)$ from the selection function. In this chapter the selection function is not derived from the luminosity function. Mann, Saunders & Taylor (1996) have summarized a method for calculating the selection function of a flux limited survey, the full details of which will be found elsewhere (Saunders, in preparation). They have used the non-parametric estimator

$$\phi\left(r_j\right) = \frac{n\left(r_j\right)\overline{\rho}\left(r_j\right)}{V\left(r_j\right)\rho\left(r_j\right)} \tag{5.10}$$

where $n(r_j)$ is the number of galaxies in the bin centred at r_j, $V(r_j)$ is the volume of this bin, $\rho(r_j)$ is the radial density estimated via the methods of Saunders *et al.* (1990) and $\overline{\rho}(r_j)$ is the evolution factor, put in to correct for the fact that dividing by the radial density estimate removes the density evolution of the sources, which may be important for IRAS galaxies. The selection function estimated this way is unbiased by galaxy clustering and is independent of k-corrections. The selection function is normalized by setting

$$\int dV \frac{\phi(r)}{1 + 4\pi J_3 \phi(r)}, \tag{5.11}$$

equal to

$$\sum_{j=1}^{no.\ of\ sources} \frac{1}{(1 + 4\pi J_3 \phi(r_j))}. \tag{5.12}$$

The selection function calculated in this way for IRAS galaxies was found to be well fit by a double power-law, so the parametric form

$$\phi(\Delta) = \phi_\star \frac{10^{(1-\alpha)\Delta}}{(1 + 10^{\gamma\Delta})^{\frac{\beta}{\gamma}}} \tag{5.13}$$

is adopted, where $\Delta = log_{10}(r/r_0)$ and the parameters $\phi_\star$, α, r_0, γ and β are estimated using maximum likelihood. These parameters have been estimated for the PSCz survey, and are used in the calculations of this chapter. The best fitting values are:

$\phi_\star = 0.0085368(\,h^{-1}\mathrm{Mpc})^{-3}$

$r_0 = 2.029328\,h^{-1}\mathrm{Mpc}$

$\alpha = 1.7369$

$\beta = 4.834672$

$\gamma = 1.2339$

(Will Saunders, private communication).

Figure 5.4 show the power spectrum of galaxies in the PSCz redshift survey, again for the same four values of $P_w(k)$ in the weighting function (see equation 2.28 of Chapter 2) as were used for the IRAS power spectra in Chapter 4. Galaxies with redshift less than $z = 0.15$ were included in the analysis. No explicit cuts in Galactic latitude were imposed for the power spectra calculated in this chapter, unless otherwise stated.[1] The power spectrum is extremely stable, and is insensitive to the weighting used. The effect of using the more restrictive mask (MASK II) is illustrated in Figure 5.5. There is no significant difference between the power spectra calculated using the two different masks to within the calculated error bars. The most encouraging result is shown in Figure 5.6 and is that the power spectrum of the PSCz galaxy survey is in excellent agreement with that of the combined QDOT/IRAS $1.2Jy$ survey computed in Chapter 4, showing no excess power at any wave number. In contrast, as expected, the power spectrum of the PSCz galaxies is significantly different from that of the QDOT galaxy survey (Figure 5.7). If there was indeed an error in the flux calibration of the IRAS satellite as suggested by Hamilton (1995) then the effect of this in the PSCz survey would be more noticeable than it was in QDOT since the number of galaxies is now six times larger.

In Figure 5.8, I show the possible effect on the power spectrum of the window function and the biases described in previous chapters. As mentioned in these chapters the method of estimating the power spectrum employed here fails at small wave numbers (as will any method, since given a survey of finite size we cannot expect to measure the power spectrum on arbitrarily large scales). To gauge the possible size of the effect of the convolution with the window function and the uncertainty in the mean number density for the PSCz power spectrum, I have chosen a CDM-type power spectrum with $\Gamma = 0.2$, $\Omega_{tot} = 1$ and a normalization which matches the PSCz power spectrum in the wave number range $0.05 \lesssim k \lesssim 0.1\ h\ \mathrm{Mpc}^{-1}$. At these wave numbers we would expect to recover the power

[1] The analysis of the PSCz redshift survey in this, and following, chapters is based on a preliminary version of the catalogue.

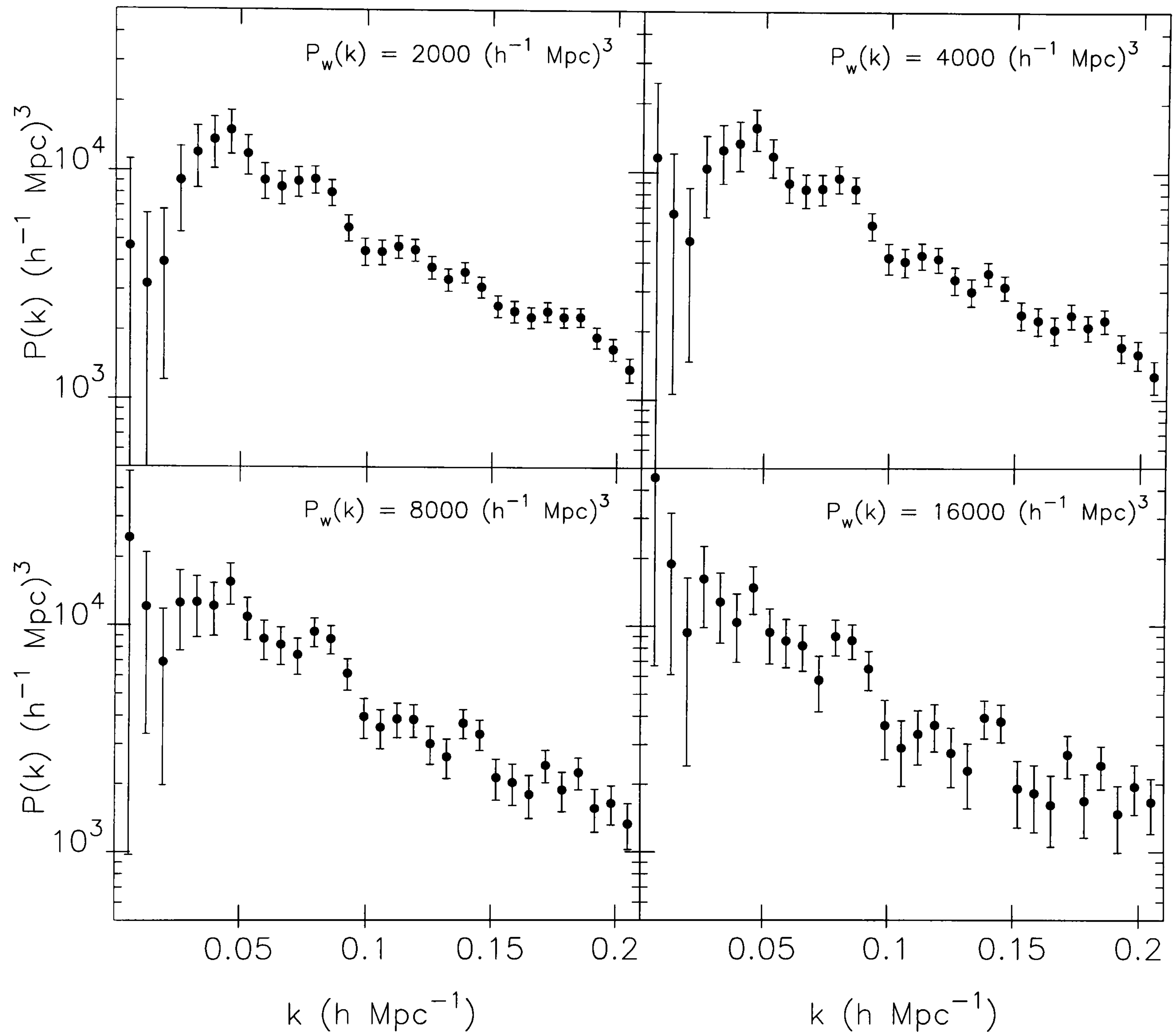


Figure 5.4: The power spectrum of IRAS galaxies from the PSCz redshift survey. The four panels show the flux limited power spectra for four different values of $P_w(k)$ in the weighting function (equation 2.28 of Chapter 2).

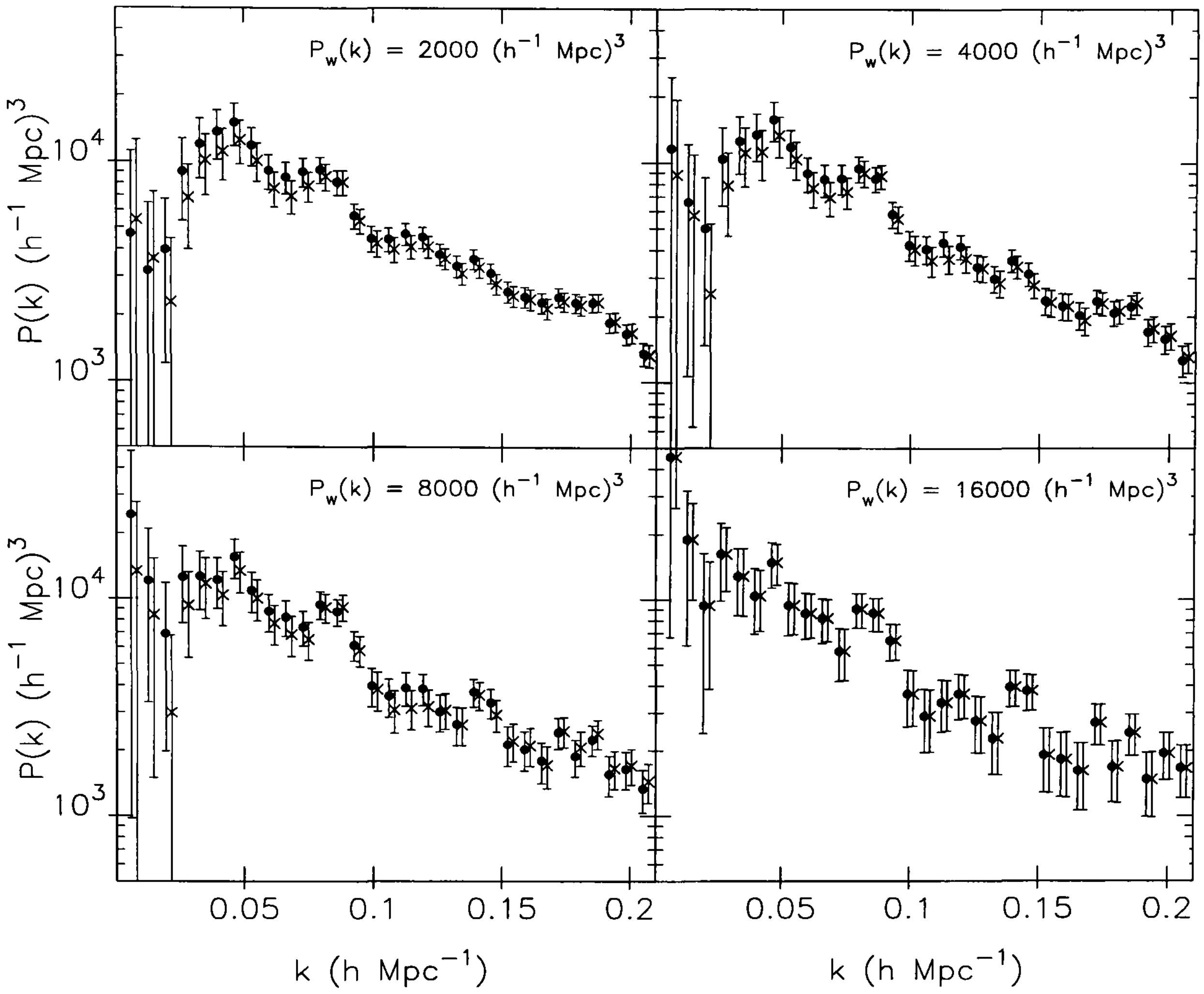


Figure 5.5: The power spectra for the PSCz galaxy survey. The crosses show the power spectra of galaxies with the exclusion zones being defined by MASK I, while the circles show the power spectra using the more restrictive MASK II. See text for definition of masked areas.

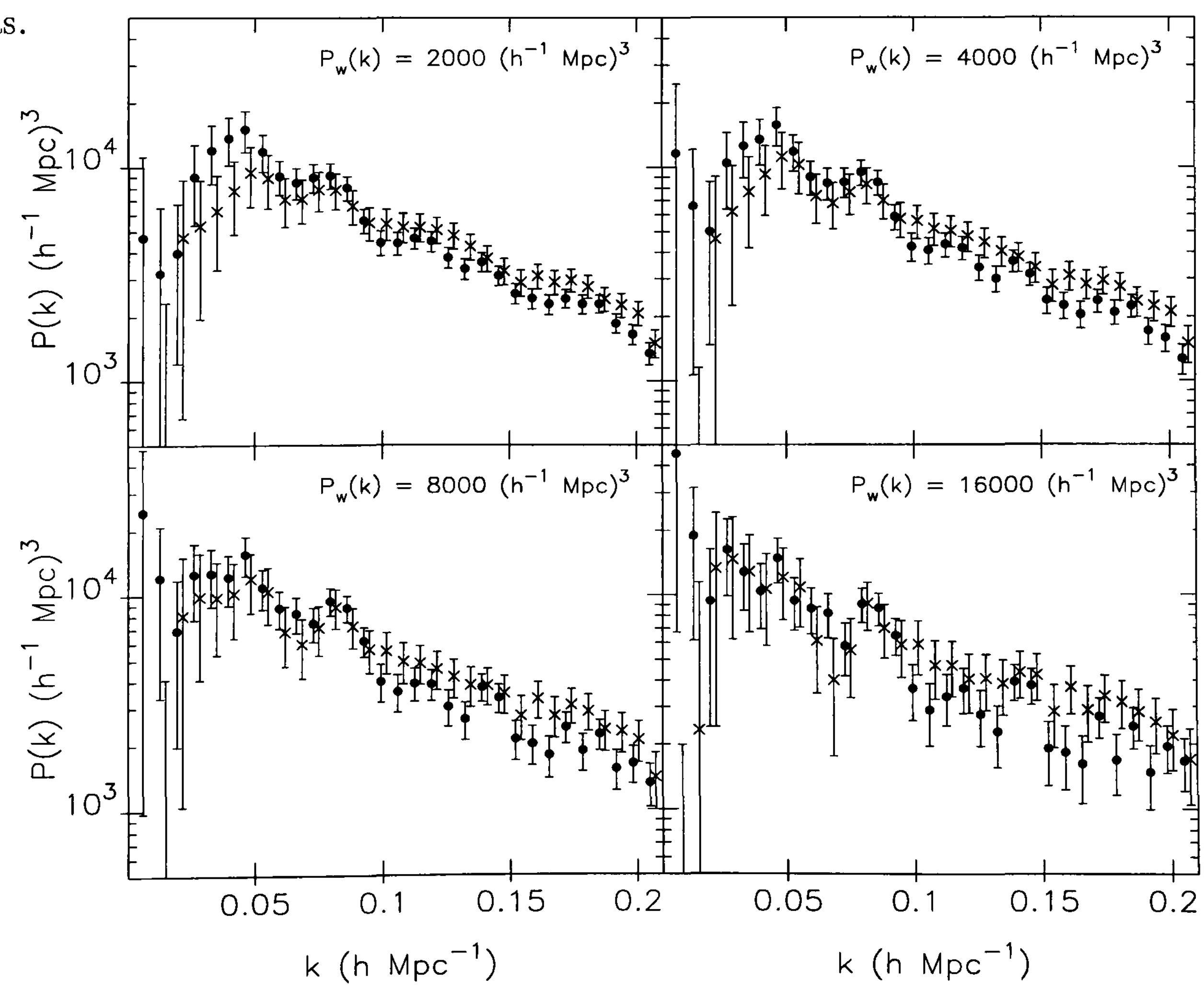


Figure 5.6: Figure shows a comparison of the power spectrum of galaxies in the PSCz redshift survey (circles) with that of the combined QDOT/IRAS $1.2Jy$ survey (crosses) as calculated in Chapter 4.

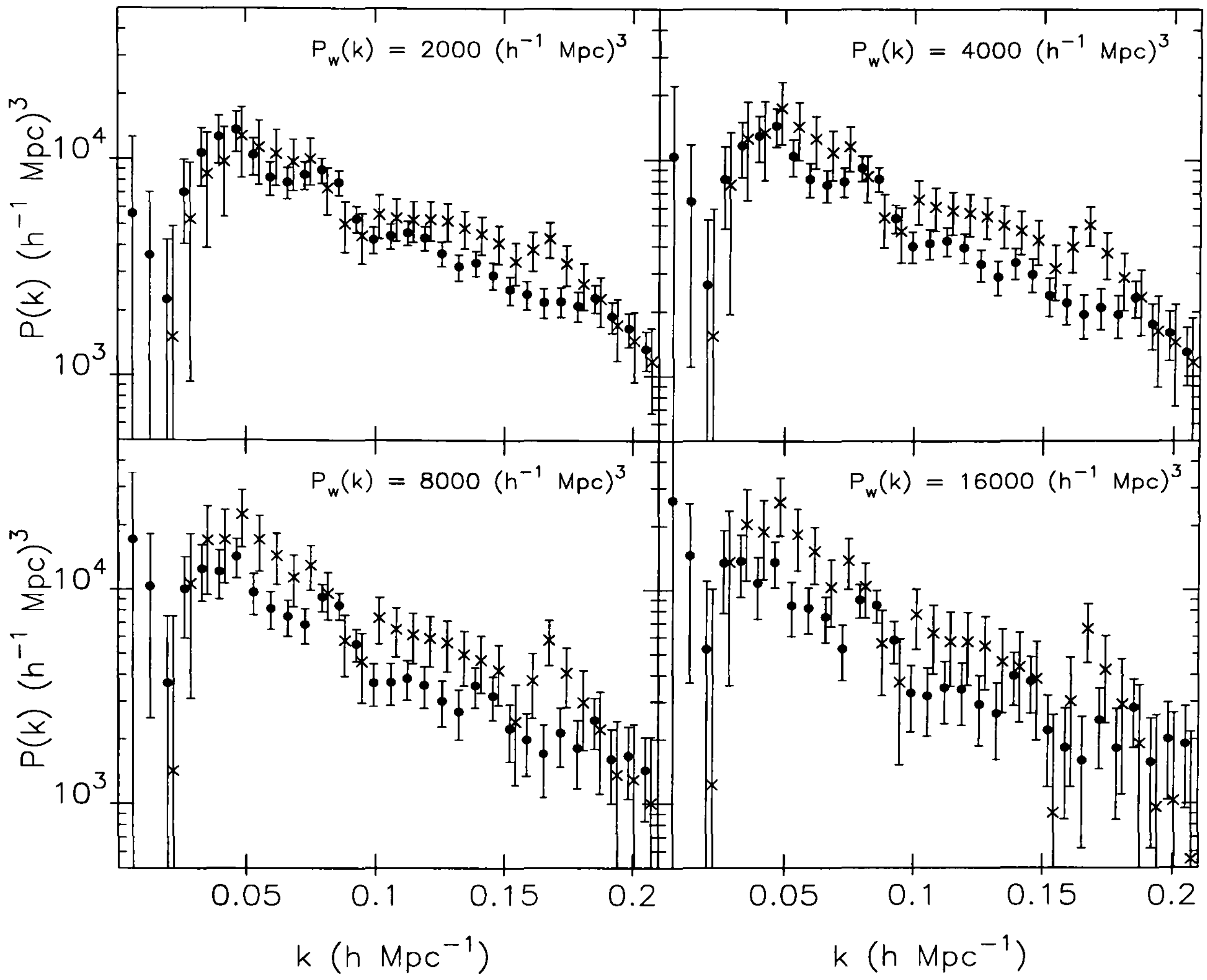


Figure 5.7: Figure shows a comparison of the power spectrum of the PSCz redshift survey (circles) with that of the QDOT redshift survey (crosses) as calculated in Chapter 4.

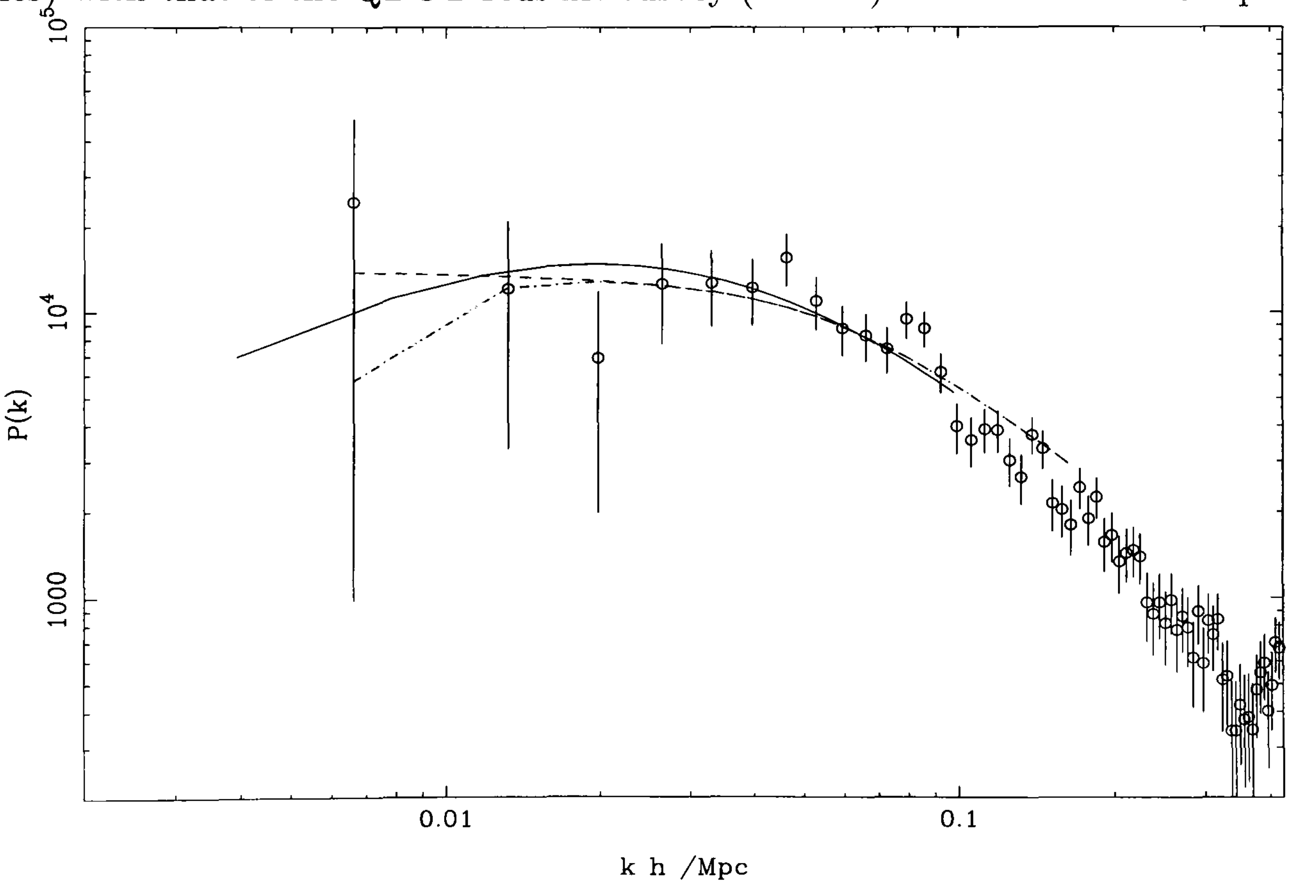


Figure 5.8: Figure shows the power spectrum of the PSCz redshift survey with one sigma errors, using a value $P_w(k) = 8000\,(h^{-1}\mathrm{Mpc})^3$ in the weighting function. The solid line shows a $\Gamma = 0.2$, CDM-like linear theory power spectrum in redshift-space, normalized such that it provides a good fit to the PSCz power spectrum in the wave number range $0.05 \lesssim k \lesssim 0.1\ h\ \mathrm{Mpc}^{-1}$. The dashed line shows the effect on the linear theory power spectrum of the convolution with the window function of the PSCz survey, and the dot-dashed line shows the further bias in the measured power spectrum due to the estimation of the mean density of the survey from the survey itself.

spectrum very accurately, and to be unaffected by non-linearities, assuming this model. With this theoretical power spectrum I have calculated the effect of the convolution with the window function and the effect of the bias correction described in Appendix B. In Figure 5.8 this linear theory power spectrum (in redshift-space) is shown as the solid line, while the dashed line shows the effect of the convolution with the window function of the survey. The dot-dashed line shows the effect of the bias due to the estimation of the mean number density from the data set itself. We can see that if the $\Gamma = 0.2$ linear theory power spectrum is representative of the true galaxy power spectrum then it is the convolution with the window function, rather than the effect of the bias derived in Appendix B which determines the shape of the power spectrum at wave numbers $k \simeq 0.02\ h\ \mathrm{Mpc}^{-1}$. This suggests that it may be worthwhile to try to deconvolve the power spectrum using, for instance, a Lucy deconvolution algorithm (Lucy 1974). I have already experimented with this idea, however the deconvolution process is at present too unstable to be useful. This will be the subject of future work. The power spectrum of the PSCz survey is well described by a power-law of index ~ -1.5 on scales with wave number $k \gtrsim 0.04$, flattening on scales with smaller wave numbers. The power spectrum does not turn over to within the calculated errors.

Figure 5.9 shows the power spectrum of PSCz galaxies as a function of Galactic latitude. I have split the PSCz sample into 6 equal area slices in b. These have $-90° < b < -42°$. $-42° < b < -20°$, $-20° < b < 0°$, $0° < b < 20°$, $20° < b < 42°$, and $42° < b < 90°$. The power spectra from the six slices show reasonable agreement. The most discrepant slices are those with the largest $|b|$ (especially the slice with negative Galactic latitude), not the slices nearest to the Galactic plane which we may suspect of being incomplete. The differences in the power spectra are apparent at wave numbers $k \gtrsim 0.1\ h\ \mathrm{Mpc}^{-1}$. These differences may not be significant as the error bars are calculated from equation (2.4.6) of Feldman, Kaiser & Peacock (1994) under the assumption that the galaxy density field is a Gaussian point process. As was discussed in Chapter 2 these error bars are expected to be underestimated by about a factor of 2 at wave numbers $k \gtrsim 0.1\ h\ \mathrm{Mpc}^{-1}$. Figure 5.9 suggests that any possible variation in the detector sensitivity as a function of Galactic latitude has little effect on the large scale structure as measured by the power spectrum.

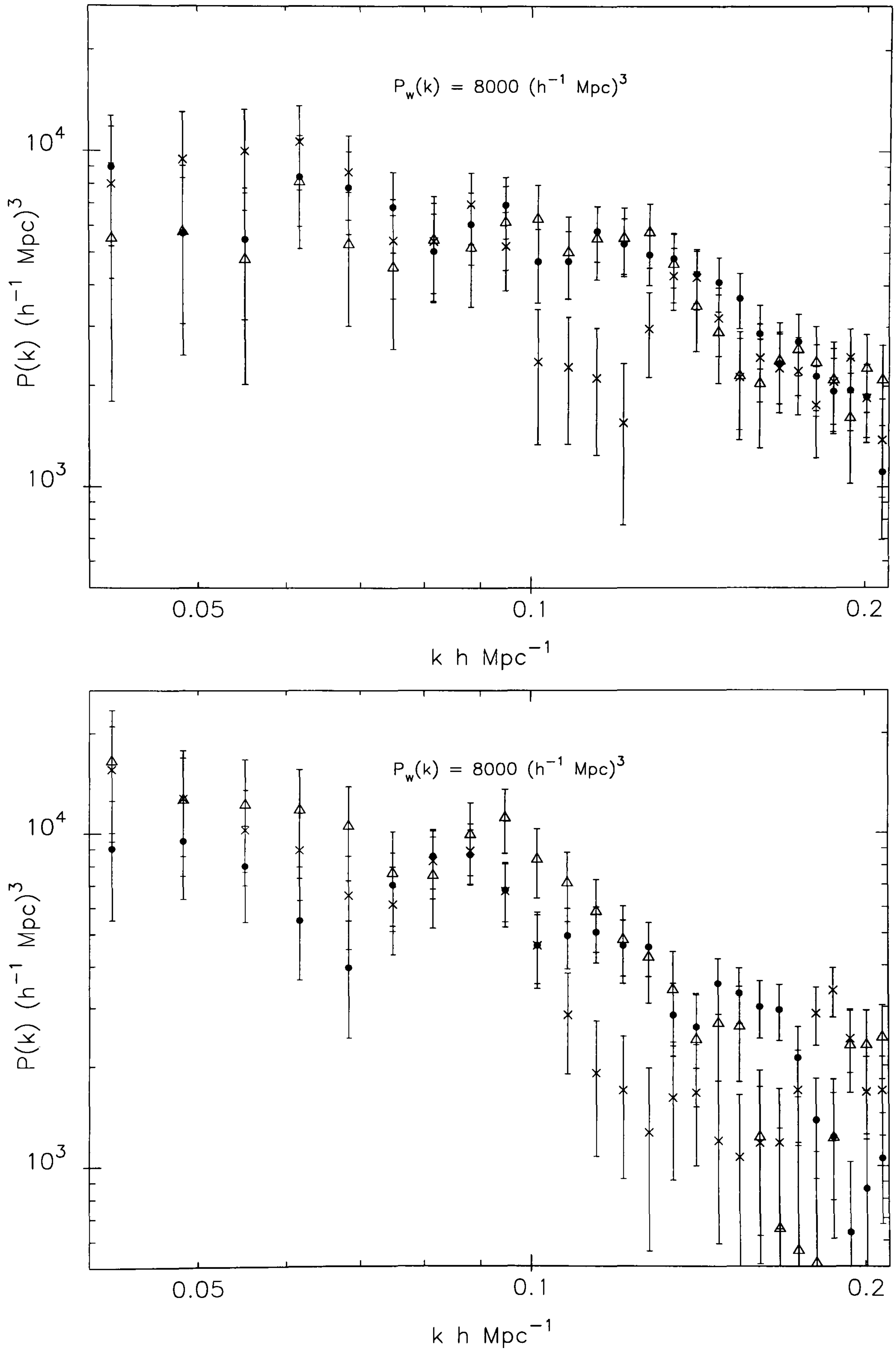


Figure 5.9: The power spectrum of the PSCz redshift survey as a function of Galactic latitude. The top panel shows the power spectrum for galaxies with positive b and the bottom panel shows the power spectrum for galaxies with negative b. In each case, the crosses show the spectra from the slice with the largest $|b|$ (i.e. $-90° < b < -42°$ for the bottom panel and $42° < b < 90°$ for the top panel), the triangles show the power spectra from the smallest $|b|$ slice (i.e. $-20° < b < 0°$ for the bottom panel and $0° < b < 20°$ for the top panel) and the circles show the power spectra from the middle $|b|$ slice (i.e. $-42° < b < -20°$ for the bottom panel and $20° < b < 42°$ for the top panel). All the estimates use a value of $P_w(k) = 8000\,(h^{-1}\mathrm{Mpc})^3$ in the weighting function.

5.5 Comparison with the Stromlo-APM galaxy power spectrum

In Figure 5.10 I show the power spectrum of the PSCz redshift survey, weighted with $P_w(k) = 8000\,(h^{-1}\mathrm{Mpc})^3$ plotted with the power spectrum of the Stromlo-APM galaxies survey, also weighted with $P_w(k) = 8000\,(h^{-1}\mathrm{Mpc})^3$ (see Chapter 2). The power spectra are plotted only in the region of k space where we expect non-linearities to be minor, and where we expect to recover the power spectrum accurately from the data. The IRAS galaxies show consistently lower power than the optical galaxies on the scales shown in Figure 5.10. The ratio of the two spectra (where I have linearly interpolated the PSCz power spectrum points to the k values of the Stromlo-APM power spectrum) is shown as the inset in Figure 5.10. An unweighted average of the points in the bottom panel gives ~ 2.1 for the ratio of the Stromlo-APM power spectrum to that of the PSCz survey. Thus Figure 5.10 suggests a relative bias of $b_{opt}/b_{iras} \sim 1.4$. Reassuringly this value agrees well with the bias parameter obtained from a counts in cells analysis of the two surveys presented in Chapter 6, where the optical/IRAS bias parameter is also estimated to be $b_{opt}/b_{iras} \sim 1.4$. The relative bias of ~ 1.4 is slightly higher than the ratio $b_{opt}/b_{iras} \sim 1.2$ quoted in Chapter 2. This is probably because the value 1.2 there resulted from a comparison of the Stromlo-APM, CfA and APM surveys with the IRAS power spectrum; using just the Stromlo-APM survey gives a slightly higher bias because the power spectrum of this survey is higher than that of the other two.

5.6 Summary

In this chapter I have described the construction of a new IRAS galaxy redshift survey containing $\sim 14,000$ galaxies. I have calculated the power spectrum of IRAS galaxies and found it is well described by a power-law of index ~ -1.5 on scales with wave number $k \gtrsim 0.04$, flattening on scales with smaller wave numbers. It seems worthwhile looking into ways of deconvolving the window function of the survey to attempt to measure the power spectrum reliably to smaller wave numbers. I have illustrated the stability of the power spectrum measurement to the value of $P_w(k)$ used in the weighting function, and also to changes in the sky-mask. The IRAS PSCz power spectrum is significantly different from that of the QDOT survey, but agrees very well with that of the IRAS $1.2Jy$ redshift survey. I have also measured the power spectrum of the PSCz galaxies as a function of Galactic

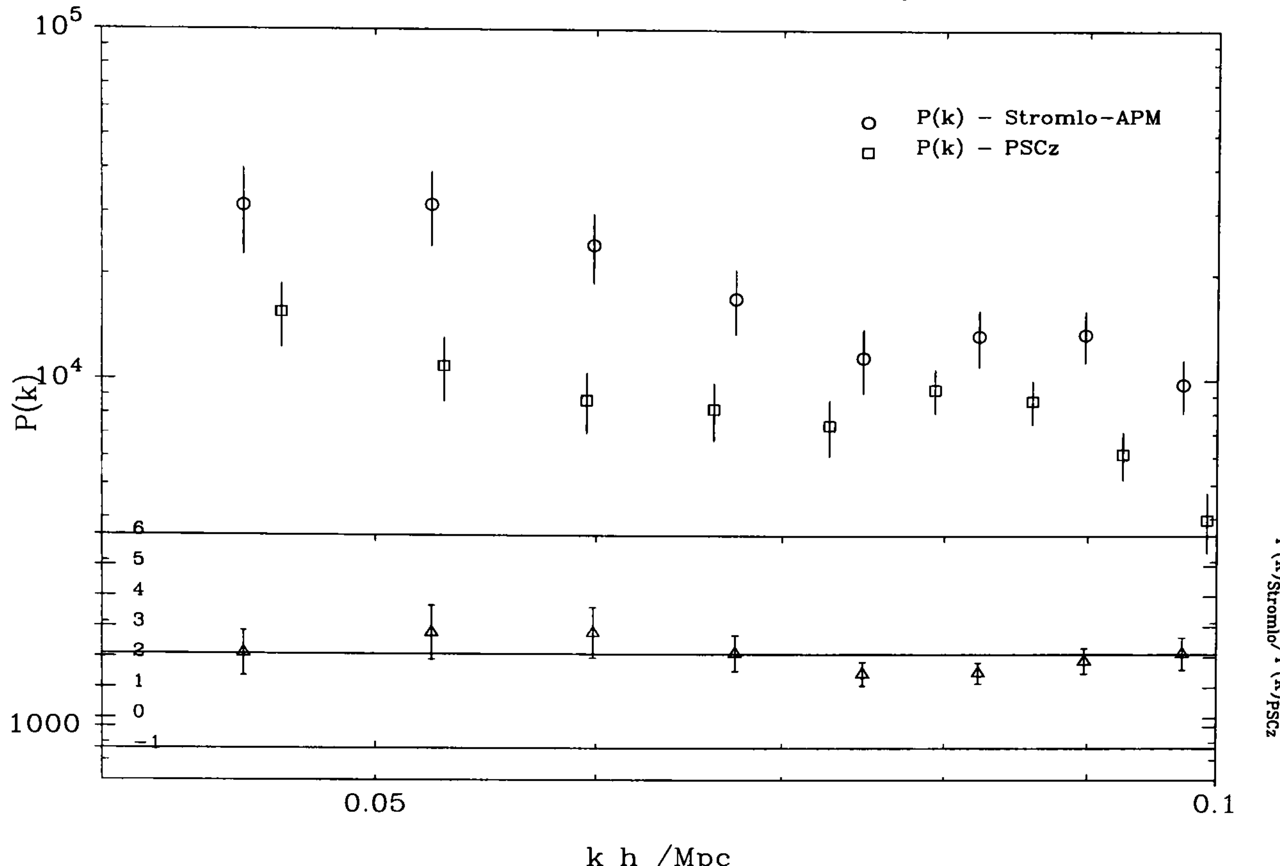


Figure 5.10: A comparison of the power spectrum of optical and IRAS galaxies. The circles show the power spectrum of the Stromlo-APM galaxies survey as computed in Chapter 2. The squares show the power spectrum of the PSCz redshift survey, computed in this chapter. Both spectra are calculated using $P_w(k) = 8000\,(h^{-1}\mathrm{Mpc})^3$ in the weighting function. The inset shows the ratio of the two power spectra, illustrating the fact that the optical galaxies have a clustering amplitude higher than that of IRAS galaxies by a factor of about 2.1.

latitude by splitting the sample into six equal-area Galactic latitude slices, three above the plane and three below the plane. The power spectra of these slices are in reasonable agreement, with the slices with the largest $|b|$ showing a slight deficit of power on scales $k \gtrsim 0.1\ h\ \mathrm{Mpc}^{-1}$. This result is probably insignificant as the error bars at this wave number range are underestimated.

Finally I have compared the IRAS power spectrum with that of galaxies in the Stromlo-APM optical redshift survey and estimated that the relative bias between the two sets of galaxies, b_{opt}/b_{iras} is $\simeq 1.4$, and is approximately constant over the wave number range 0.045 – 0.1 h Mpc^{-1}.

Chapter 6

Counts in cells analysis of galaxy redshift surveys

6.1 Method

For a statistical analysis of large scale structure in galaxy redshift surveys one must think carefully about how to weight the information carried by each galaxy. In flux limited surveys the number density of galaxies declines as a function of radial distance and consequently we would like to give more relative weight to the few galaxies we see at large distances. An obvious weighting function to apply to each galaxy would be $1/\phi(r)$ i.e. the inverse of the selection function. However, heavily weighting distant regions of the survey, where shot noise dominates over the signal, generally gives unreliable results. For the measurement of the power spectrum the optimum weight function is $w(r) = 1/(1 + \overline{n}P_w(k))$, as derived by Feldman, Kaiser & Peacock (1994) and discussed in previous chapters. This scheme is optimum in the sense that it results in the minimum variance in the measured power spectrum. It does, however, require one to know, or at least estimate, the power spectrum in advance of the measurement.

Using a volume limited redshift survey is more straight forward as each galaxy carries an equal weight. We can make volume limited surveys from flux limited ones, and sometimes there are good reasons for doing so (see Chapter 2) but this typically involves discarding around three quarters of the galaxies in the survey, which is wasteful, and means that the noise in the measured statistic increases.

We can avoid the question of how to weight galaxies in a redshift survey by using a counts in cells analysis. The galaxies are automatically weighted in an appropriate way without any assumption as to the form of the power spectrum or correlation function, as we will see below. Counts in cells analysis has been performed on many galaxy redshift surveys

(Efstathiou *et al.* 1990, Loveday *et al.* 1992a, Bouchet *et al.* 1993) and the counts in cells method has recently been extended so that one can compare two different redshift surveys covering the same volume of space (Efstathiou 1995). We can then test whether the two sets of galaxies sample the same underlying fluctuations and/or whether there are any errors (for example flux errors) in the parent catalogues. The statistical information recovered from a counts in cells analysis is the same as that provided by the power spectrum or the correlation function. We measure the fluctuations of the counts of galaxies in roughly cubical cells. The variance, σ^2, of the counts in these cells (above the Poisson level) is related to the correlation function by an integral,

$$\sigma^2(l) = \frac{1}{V}\int_{V=l^3} \xi(r_{12})\, dV_1 dV_2, \tag{6.1}$$

where l is the size of the cell and $V = l^3$ is the volume of the cell. The relation between the variance and the power spectrum $P(k)$ is given by

$$\sigma^2(l) = \int_0^\infty P(k)\, \hat{W}^2(kr)\, d^3k, \tag{6.2}$$

where $\hat{W}(kr)$ is the Fourier transform of the cell window function. To apply counts in cells statistics to a single redshift survey, we divide up the volume of the survey into N_{shell} concentric shells. Each shell is then divided into N_c cubical cells of volume V. The number of galaxies in cell i is N^i, so

$$\langle N \rangle = nV \tag{6.3}$$

and

$$\left\langle (N - nV)^2 \right\rangle = nV + n^2V^2\sigma^2 \tag{6.4}$$

give the expectation values of the mean and the variance of the counts in cells, where n is the mean density of the redshift survey, and σ^2 is defined in equation (6.1). Following Efstathiou *et al.* (1990) and Efstathiou (1995) we compute,

$$\overline{N} = \frac{1}{N_c}\sum_i N^i \tag{6.5}$$

and

$$S = \frac{1}{N_c - 1}\sum_i \left(N^i - \overline{N}\right)^2 - \overline{N}, \tag{6.6}$$

the expectation values of these being

$$\left\langle \overline{N} \right\rangle = nV \tag{6.7}$$

and

$$\langle S \rangle = n^2 V^2 \sigma^2. \tag{6.8}$$

If the cells are approximately independent, then

$$Var(S) = < S^2 > - < S >^2 = \frac{\mu_4 - 2\mu_3 + \mu_2 - \mu_2^2}{N_c}, \tag{6.9}$$

where the μ_2, μ_3, μ_4 are the second, third and fourth moments of the cell counts. The central limit theorem may be applied when the number of cells N_c is large, and in this case the distribution of S approaches a Gaussian distribution with variance given by equation (6.9). The moments of the cell counts μ_2, μ_3 and μ_4 are given by

$$\mu_2 = nV + n^2 V^2 \sigma^2 \tag{6.10}$$

$$\mu_3 = nV + 3n^2 V^2 \sigma^2 + n^3 V^3 \alpha \tag{6.11}$$

$$\mu_4 = nV + n^2 V^2 \left(3 + 7\sigma^2\right) + 6n^3 V^3 \left(\sigma^2 + \alpha\right) + n^4 V^4 \left(3\sigma^4 + 2\beta\right), \tag{6.12}$$

where α and β are integrals over the reduced three and four point correlation functions and are zero for Gaussian density fluctuations. Setting α and β to zero is a fairly good approximation for scales $l \gtrsim 20\ h^{-1}$Mpc, in which case we find

$$Var(S) = \frac{2n^2 V^2 \left(1 + \sigma^2\right) + 4n^3 V^3 \sigma^2 + 2n^4 V^4 \sigma^4}{N_c}. \tag{6.13}$$

This formalism may be extended to the joint analysis of two redshift surveys. We now have cell counts for the two surveys: N_1^i and N_2^i for survey 1 and 2 respectively. The expectation values of these counts are denoted by $\langle N_1^i \rangle = \lambda$ and $\langle N_2^i \rangle = \mu$.

For each of the surveys we estimate the mean and the variance of the counts from equations (6.5) and (6.6). We can also measure the mean square difference in the cell counts between the two surveys:

$$S_{12} = \frac{1}{N_c - 1} \sum_i \left(\overline{N_2} N_1^i - N_2^i \overline{N_1}\right) - \left(\overline{N_2} + \overline{N_1}\right) \overline{N_2 N_1}. \tag{6.14}$$

The expectation value of this statistic is

$$\langle S_{12} \rangle = \lambda^2 \mu^2 \left(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}^2\right), \tag{6.15}$$

where σ_{12} is the equivalent of equation (6.1) for the cross correlation function ξ_{12}. The variance of S_{12} is given by

$$Var(S_{12}) \quad = \quad \lambda\mu \frac{N_c}{(N_c - 1)^2} \{\lambda^3 + \mu^3 + 4\lambda^2 \mu^2$$

$$
\begin{aligned}
&+ \quad 2\left(\lambda^3\mu + \lambda\mu^3\right) + \lambda\mu^3\left(7\sigma_2^2 - 4\sigma_{12}^2\right) + 6\lambda^2\mu^2\sigma_{12}^2 \\
&+ \quad 4\left(\lambda^3\mu^2 + \lambda^2\mu^3\right)\left(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}^2\right) \\
&+ \quad 2\lambda^3\mu^3\left[\left(\sigma_1^2 + \sigma_2^2\right)^2 - 4\sigma_{12}^2\left(\sigma_1^2 + \sigma_2^2 - \sigma_{12}^2\right)\right]\},
\end{aligned} \tag{6.16}
$$

where it is assumed that the mean values μ and λ are known. If the two galaxy samples are identical tracers of the same underlying density distribution, then the value of S_{12} should be zero (to within the errors estimated from equation 6.16). However, the galaxies may be biased tracers of the mass distribution. Higher luminosity galaxies may be more clustered than those of lower luminosity, or galaxy clustering strength may depend on morphological type. If we assume a linear bias model (Bardeen *et al.* 1986), we can see that

$$\frac{\langle S_1 \rangle}{\langle \lambda^2 \rangle} = b_1^2 \sigma_m^2, \tag{6.17}$$

$$\frac{\langle S_2 \rangle}{\langle \mu^2 \rangle} = b_2^2 \sigma_m^2 \tag{6.18}$$

and

$$\frac{\langle S_{12} \rangle}{\langle \lambda^2 \mu^2 \rangle} = \left(b_1 - b_2\right)^2 \sigma_m^2, \tag{6.19}$$

where the subscript m refers to the actual underlying density field. So from these statistics we can measure the relative linear bias parameter between galaxies in different surveys (for instance an IRAS galaxy survey and an optical galaxy survey), and between galaxies of different luminosities in the same survey. The strength of this method is that the statistics of the two populations of galaxies are calculated within the same volume of space.

We may again appeal to the central limit theorem when the number of cells N_c is large. In this case, the distribution of S_{12}, S_1 and S_2 will become a multivariate Gaussian

$$p\left(S_1, S_2, S_{12}\right) dS_1 dS_2 dS_{12} = \frac{1}{(2\pi)^{\frac{3}{2}} \left(det\mathbf{V}\right)^{\frac{1}{2}}} \times exp\left(-\frac{1}{2} S_i S_j V_{ij}^{-1}\right) dS_1 dS_2 dS_{12}, \tag{6.20}$$

where V_{ij} is the covariance matrix, $Cov\left(S_i S_j\right)$. We can therefore compute the variances σ_1^2, σ_2^2 and σ_{12}^2 by forming a likelihood function

$$L \propto \prod_{k=1}^{N_{shell}} p\left(S_1^k, S_2^k, S_{12}^k\right), \tag{6.21}$$

and maximising it with respect to these variables. This is how we avoid having to weight the individual galaxies. The likelihood function automatically gives a weight to each shell.

Nearby shells get little weight as they contain few cells, and distant shells get little weight as they contain few galaxies.

Implementation of the statistical machinery described above must be carried out very carefully, as in practice there are a number of complications. Firstly, redshift surveys are generally not all-sky. There are always incomplete cells because of the regions of the sky which are masked. Efstathiou *et al.* (1990) describe a correction for this effect. The correction is approximate in the sense that it assumes that the main effect of having an incomplete cell is to alter the mean count only, leaving the variance unaffected. To apply the correction, equation (6.6) is replaced by

$$S_i = \frac{\sum_j \left(N_j - \frac{V_j \sum_k N_k}{\sum_k V_k}\right)^2 - \left(1 - \frac{\sum_k V_k^2}{\left(\sum_k V_k\right)^2}\right) \sum_j N_j}{\left(\frac{\sum_j N_j}{\sum_j V_j}\right)^2 \left[\sum_k V_k^2 - \frac{2\sum_k V_k^3}{\sum_k V_k} + \frac{\left(\sum_k V_k^2\right)^2}{\left(\sum_k V_k\right)^2}\right]}, \tag{6.22}$$

where the sums are over all cells in the i^{th} radial shell. In the analysis in this chapter the correction for incomplete volumes has been applied and cells are not used unless the filled (i.e. unmasked) volume is $\geq 70\%$ of the total volume of the cell.

The statistical properties of the estimators described also rely heavily on the applicability of the central limit theorem. The behaviour of the statistics approaches that expected from the central limit theorem as the number of cells in a given radial shell, N_c, becomes large. In the analysis presented in this chapter, the statistics are calculated as a function of radial distance, and I have only used radial shells containing $\gtrsim 20$ filled cells (in fact the actual limit is 19 cells; filled is taken to mean $\geq 70\%$ unmasked volume). This number is a compromise between the demands of the central limit theorem that N_c is large, and the need for enough radial shells to be included in the analysis for the statistics to be worthwhile. Whether or not there are enough cells for convergence to central limit theorem behaviour is only really a consideration for the optical/IRAS counts in cells comparison (Section 6.4), where most of the sky area is masked and there are few cells per radial shell.

A third consideration is the assumption, made in deriving equation (6.13), that the density fluctuations follow a Gaussian distribution. This is manifestly not true for scales $\lesssim 20\ h^{-1}$Mpc; below this cell size the calculated error bars on the measured statistics may be overestimated. We will also overestimate the error bars by assuming the cells are independent (equation 6.9) when they are in fact not. This will also affect the error bars for cells sizes up to about $20\ h^{-1}$Mpc.

6.2 Previous analyses

The statistics of Section 6.1 have already been applied in a comparison between the IRAS QDOT and $1.2Jy$ surveys to test whether they are compatible with being drawn from the same underlying density field (Efstathiou 1995). One possible cause of the 'Hamilton effect' in the QDOT survey (first discussed in Chapter 4) is that there are errors in the flux calibration of the data from the IRAS satellite, causing the survey to be effectively deeper in some areas of the sky than in others.

The counts in cells analysis described above can test this hypothesis. Suppose we add, to one of the two surveys which we are comparing, some extra variance σ_e^2 which is uncorrelated with the galaxy distribution but caused, for example, by the flux errors. We would then expect the statistic S_{12} to give us

$$\frac{\langle S_{12}\rangle}{\langle \lambda^2 \mu^2 \rangle} = \sigma_e^2. \tag{6.23}$$

When the QDOT and IRAS $1.2Jy$ surveys were compared (Efstathiou 1995), it was found that the variances measured from the QDOT survey for $l \sim 30 - 40\ h^{-1}$Mpc size cells were significantly higher than those from the $1.2Jy$ survey. The statistic S_{12}, however, was measured to be consistent with zero. The statistics are therefore consistent with the hypothesis that the two surveys are drawn from the same distribution, and that the results from the QDOT survey are a sampling fluctuation due to an unlucky draw of galaxies in one particular region of the survey. Efstathiou (1995) identified this region as being that around the Hercules supercluster in the QDOT survey. When this region was excluded from the analysis the cell count variances for the two surveys became consistent (see also Chapter 4 for the effect on the power spectrum). In the next section, Section 6.3, I come back to the question of possible flux errors in the PSC by using the IRAS PSCz catalogue in the counts in cells analysis described above.

In Section 6.4 I compare the clustering of the galaxies in the IRAS PSCz survey with that of galaxies in the optical Stromlo-APM survey.

6.3 Counts in cells comparison of galaxies in the IRAS PSCz redshift survey as a function of measured flux

The IRAS PSCz redshift survey (described in detail in Chapter 5) approximately encompasses all the other surveys based on the IRAS Point Source Catalogue that have been done to date; the QDOT, $1.2Jy$, and $2Jy$ surveys would be subsets of the PSCz redshift survey except for the small differences in the procedures for galaxy selection and the definition of the sky-mask employed in the construction of these surveys. If there are any flux errors in the PSC we can attempt to isolate them by taking appropriate subsets from the PSCz survey and performing a counts in cells analysis. Throughout the analysis a cut at $|b| = 10°$ is applied to the PSCz catalogue. I have first split the PSCz survey into two subsets, galaxies with $f_{60} > 1.2Jy$ (4763 galaxies) and those with flux between 0.6 and $1.2Jy$ (8498 galaxies). I have analysed counts in cells for cells of size: 5, 10, 20, 30, 40 and 60 h^{-1}Mpc. This is like a comparison of a fully sampled QDOT survey with the IRAS $1.2Jy$ survey. The results are shown in Figure 6.1 in the form of likelihood contours for the counts in cells variances of the bright and faint subsamples. Only galaxies within a radial distance of 200 h^{-1}Mpc are included in the analysis.

The distribution of the quantity $2ln\,(L/L_{max})$, where L_{max} is the maximum likelihood and L is given by equation (6.21), should follow a χ^2 distribution with two degrees of freedom, as we maximize only with respect to σ_1 and σ_2, setting $\sigma_{12}^2 = \sigma_1\sigma_2$ (see Efstathiou 1995). The contours are therefore plotted at the places where $2ln\,(L/L_{max})$ is equal to 1.39, 5.99, 9.21 and 13.81. These correspond to the 50, 5, 2 and 0.1 percentile points of the χ^2 distribution with two degrees of freedom. The bold contour is the 5 percentile contour (or the 95% confidence level). It can be seen that, unlike in the case of the QDOT/$1.2Jy$ comparison, the variances for galaxies with fluxes $0.6 < f_{60} < 1.2Jy$ are the same (at the 95% confidence level) as those for galaxies with $f_{60} > 1.2Jy$, on all measured scales.

The statistic $S_{12}/\left(\lambda^2\mu^2\right)$ for the 6 cells sizes is shown in Figure 6.2 as a function of radial distance (note, hereafter I shall drop the subscript '12' on S_{12}). The x axis is the radial distance of the centre of the ith radial shell in h^{-1}Mpc. We can now test the hypothesis that $\sigma_b^2 = \sigma_f^2$ as in this case $\langle S\rangle / \left(\lambda^2\mu^2\right)$ is expected to be zero. The probability of the fit $\langle S\rangle / \left(\lambda^2\mu^2\right) = 0$ to the statistic $\langle S\rangle / \left(\lambda^2\mu^2\right)$, for each cell size, is shown in Figure 6.2. The null hypothesis is acceptable for all cell sizes, showing that the counts in cells statistics are compatible with the two galaxy subsets being drawn from the same distribution. The fit is 'too good' for the smallest cell sizes (5 h^{-1}Mpc and 10 h^{-1}Mpc), where the errors on $\langle S\rangle / \left(\lambda^2\mu^2\right)$ are overestimated by equation (6.16) as the cells are not independent. The χ^2 for the fit $\langle S\rangle / \left(\lambda^2\mu^2\right) = 0$ is also shown.

Figure 6.3 shows the variances, σ_l^2, as a function of the radial distance of the centre of the shell for the two galaxy subsets. The maximum likelihood values are shown by the

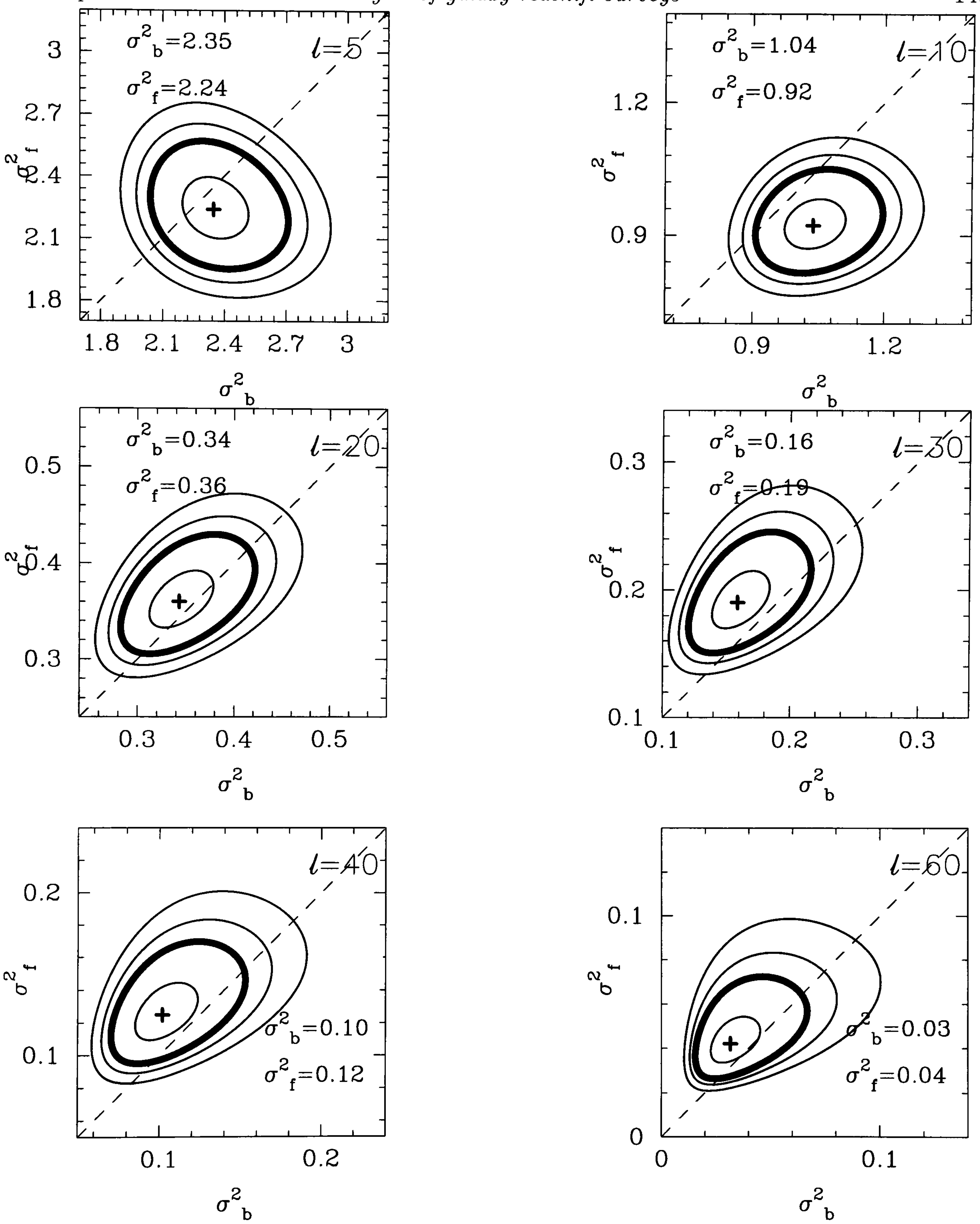


Figure 6.1: Likelihood contours for a counts in cells comparison of galaxies in the PSCz redshift survey. The subscript f refers to galaxies with flux $0.6Jy < f_{60} < 1.2Jy$, and b refers to those galaxies with flux $f_{60} > 1.2Jy$. Each panel shows contours for cells of size l in h^{-1}Mpc. The contours are plotted at the levels corresponding to $\Delta 2ln\,(L/L_{max}) =$ 1.39, 5.99, 9.21 and 13.81, which are the 50, 5, 2 and 0.1 percentile points of the χ^2 distribution with two degrees of freedom. The bold contour illustrates the 95% confidence region in $\sigma_b^2 - \sigma_f^2$ space, and the cross shows the maximum likelihood values of σ_b^2 and σ_f^2.

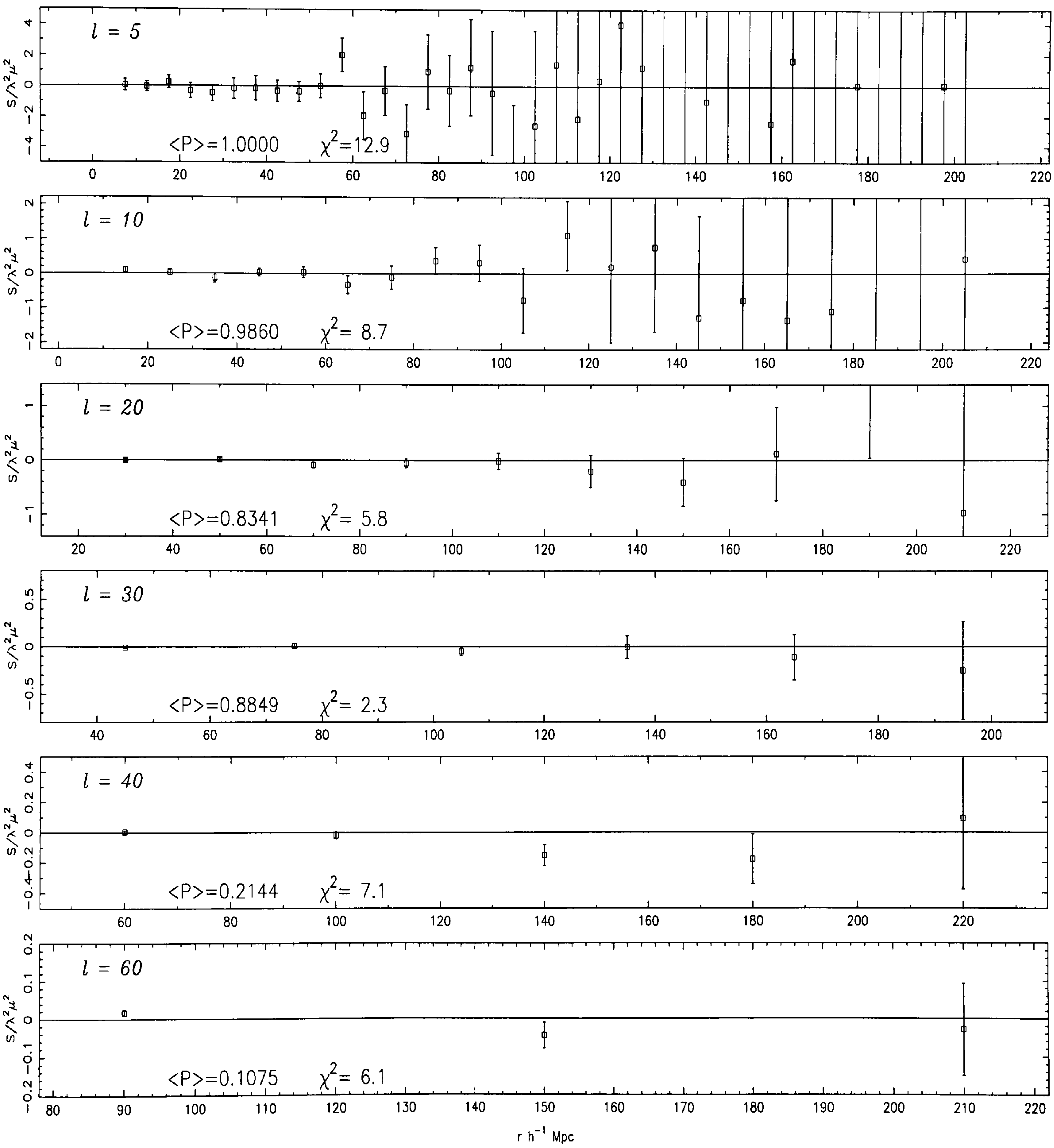


Figure 6.2: The statistic $S/\left(\lambda^2\mu^2\right)$ as a function of the radial distance of the shell for the comparison of bright ($60\mu m$ flux $> 1.2Jy$) and faint ($0.6 < 60\mu m$ flux $< 1.2Jy$) galaxies in the PSCz redshift survey. The cell size, l in h^{-1}Mpc, is shown in the top left hand corner of each panel. Also shown is the probability $\langle P\rangle$ and the χ^2 for the fit $S/\left(\lambda^2\mu^2\right) = 0$.

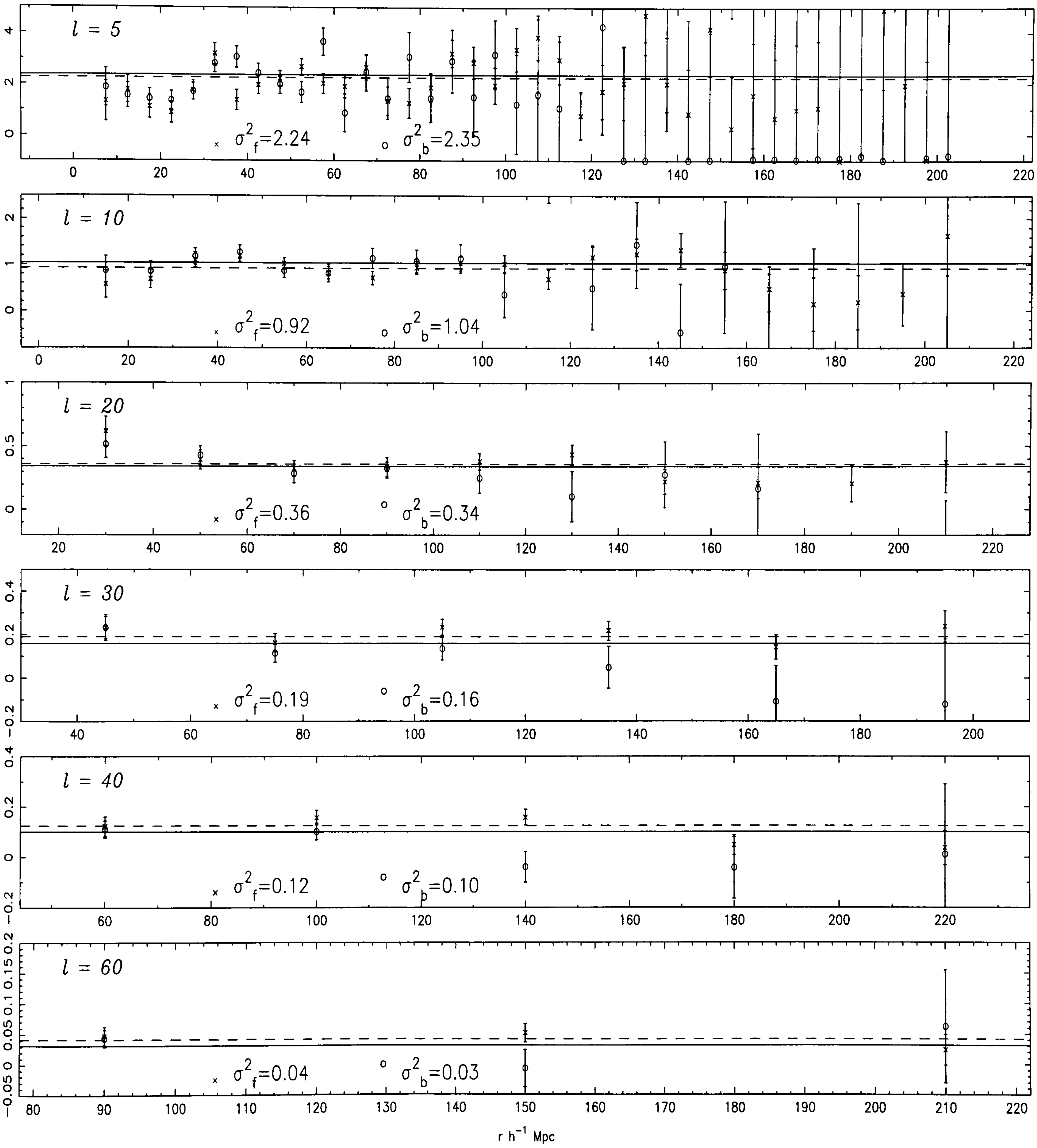


Figure 6.3: The variances of the counts in 5, 10, 20, 30, 40 and 60 h^{-1}Mpc size cells (from top to bottom) as a function of radial distance. Circles show the result for the bright ($60\mu m$ flux $> 1.2Jy$) galaxies while crosses show the result for faint ($0.6 < 60\mu m$ flux $< 1.2Jy$) galaxies in the PSCz redshift survey.

dashed and solid lines. The two values for the variance, for every cell size, are consistent with each other.

In Figures 6.4, 6.5 and 6.6 the likelihood contours, S statistic and variances as a function of radial distance are shown for the PSCz galaxy subsets $f_{60} > 1.9Jy$ (2416 galaxies) and 0.6 to $1.9Jy$ (10805 galaxies). This is like comparing a fully sampled QDOT survey with the IRAS $2Jy$ survey. Again, in this comparison there is no significant difference between the two subsets. The probability for the fit $\langle S \rangle / (\lambda^2 \mu^2) = 0$ is shown in Figure 6.5 (and again the error bars are overestimated for cells of size 5 and 10 h^{-1}Mpc).

As a further test, I have divided the PSCz sample into galaxies with flux between 0.6 and $0.8Jy$ (4828 galaxies) and those above $0.8Jy$ (8540 galaxies). These results are shown in Figures 6.7, 6.8 and 6.9. This is the most stringent test, as we are splitting the survey into a small flux slice near the flux limit, which would be most affected by flux errors, and comparing this slice with the rest of the survey. Again it can be seen that there is no evidence at any scale for an excess variance that is uncorrelated with the galaxy distribution. The $\langle S \rangle / (\lambda^2 \mu^2)$ statistic (Figure 6.8) is consistent with zero and the maximum likelihood values of σ_b^2 and σ_f^2 are almost identical on all scales. It appears unlikely that there is a problem with the IRAS fluxes. If this were the case we would expect the effect to be more visible in the PSCz catalogue than in QDOT, but the effect seems to disappear, giving weight to the suggestion that the differences between the QDOT and $1.2Jy$ surveys were caused by a sampling fluctuation in the QDOT survey. We have not, of course, examined the angular dependence of the statistics. In his original analysis, Hamilton (1995) looked at both an angle average statistic – the monopole power spectrum (which seemed to increase if the part of the sample beyond 80 h^{-1}Mpc was analysed) – as well as the dependence of the power spectrum $P(k)$ on the angle between the line of sight and the direction of the wave vector $\mathbf{k}$. He found the quadrupole component of the power decreased beyond 80 h^{-1}Mpc (signifying a 'finger of God' type effect). The correlation function of the QDOT survey, $\xi(\sigma, \pi)$, also showed large, unexpected anisotropies. The region in QDOT beyond 80 h^{-1}Mpc was elongated along the line of sight (π) axis. However, the possibility that these effects are caused by flux errors in the parent catalogue seems to be unlikely in the light of the counts in cells analysis of the PSCz redshift survey performed here. The correlation function $\xi(\sigma, \pi)$ for the new PSCz redshift survey is currently being investigated by Oliver Keeble at Imperial College.

For completeness, and for comparison with the combined IRAS $1.2Jy$/QDOT results

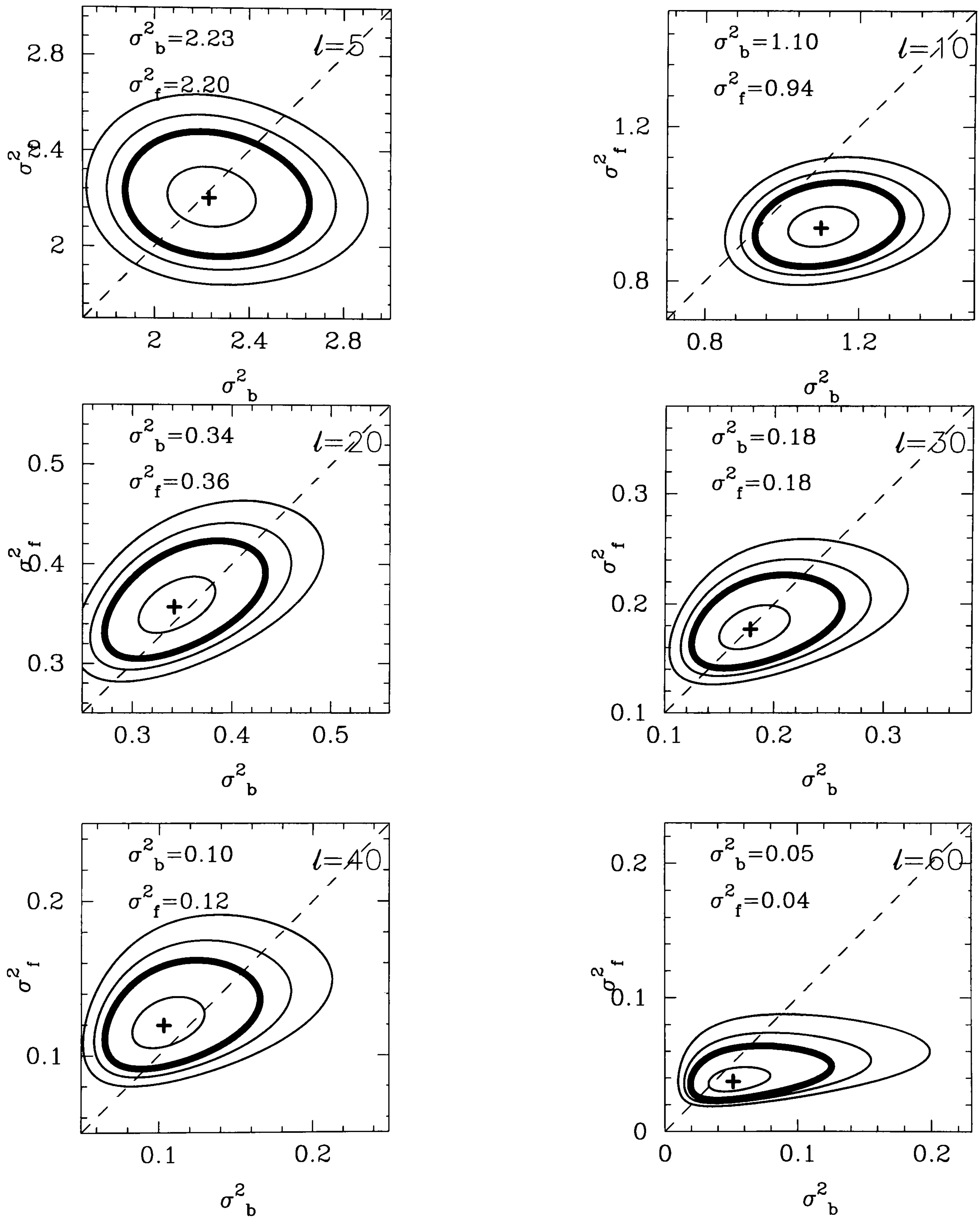


Figure 6.4: Likelihood contours for the bright/faint comparison of the PSCz redshift survey, this time split at a $60\mu m$ flux of $1.9Jy$.

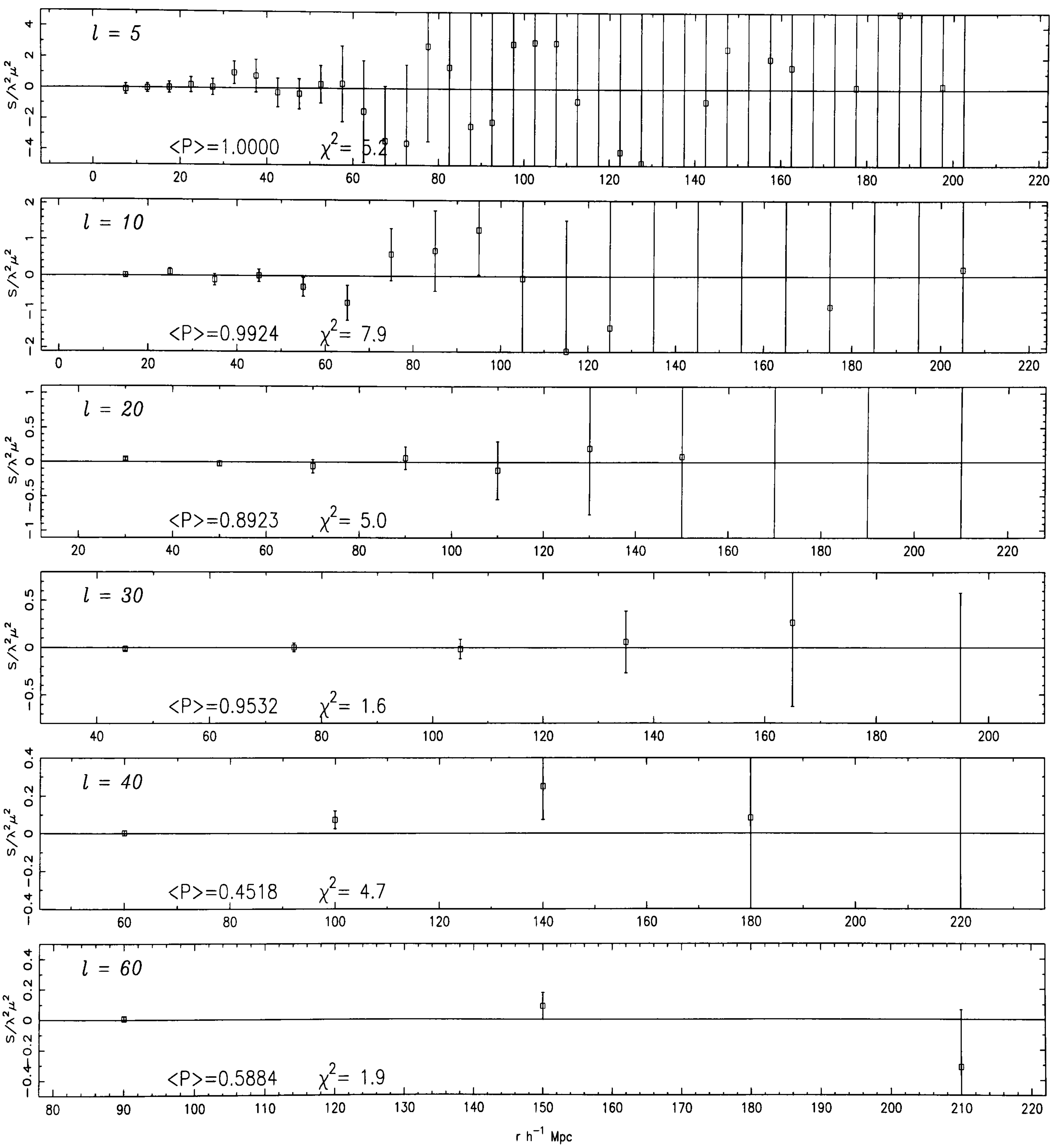


Figure 6.5: From top to bottom, figure shows the statistic $S/\lambda^2\mu^2$ (equation 6.14) as a function of radial distance plotted for cells of size 5, 10, 20, ,30, 40, and 60 h^{-1}Mpc in the comparison of PSCz galaxies with $60\mu m$ flux between 0.6 and $1.9Jy$ and those with $60\mu m$ flux $> 1.9Jy$. The χ^2 and $\langle P\rangle$ values appropriate to the fit $S/\left(\lambda^2\mu^2\right) = 0$ are also shown.

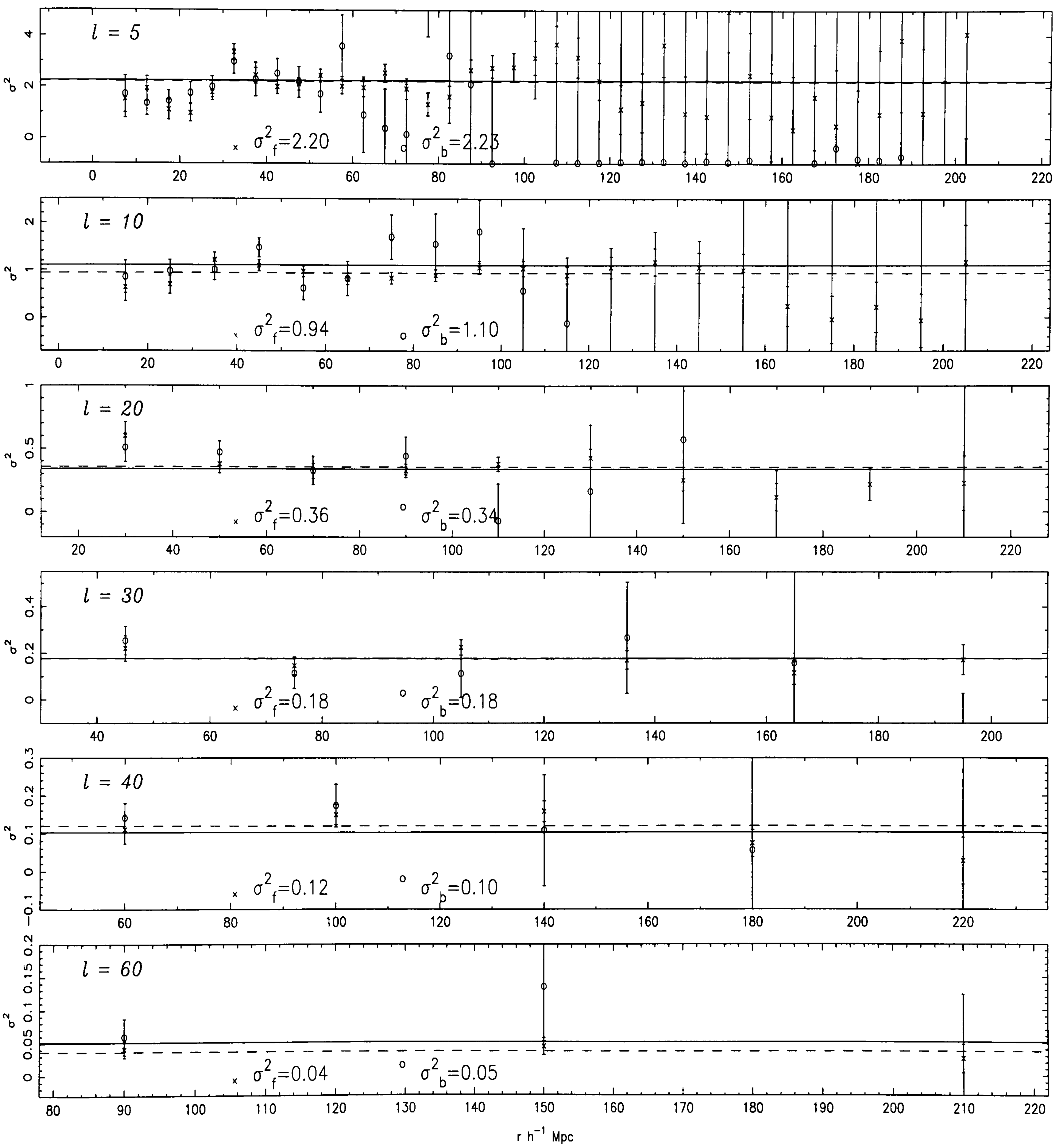


Figure 6.6: Variances of counts in cells of size 5, 10, 20, 30, 40, and 60 h^{-1}Mpc, from top to bottom, as a function of radial distance r. Circles show the variances for the PSCz galaxies with $60\mu m$ flux $> 1.9Jy$ and crosses show the variances for galaxies with $60\mu m$ flux between 0.6 and $1.9Jy$.

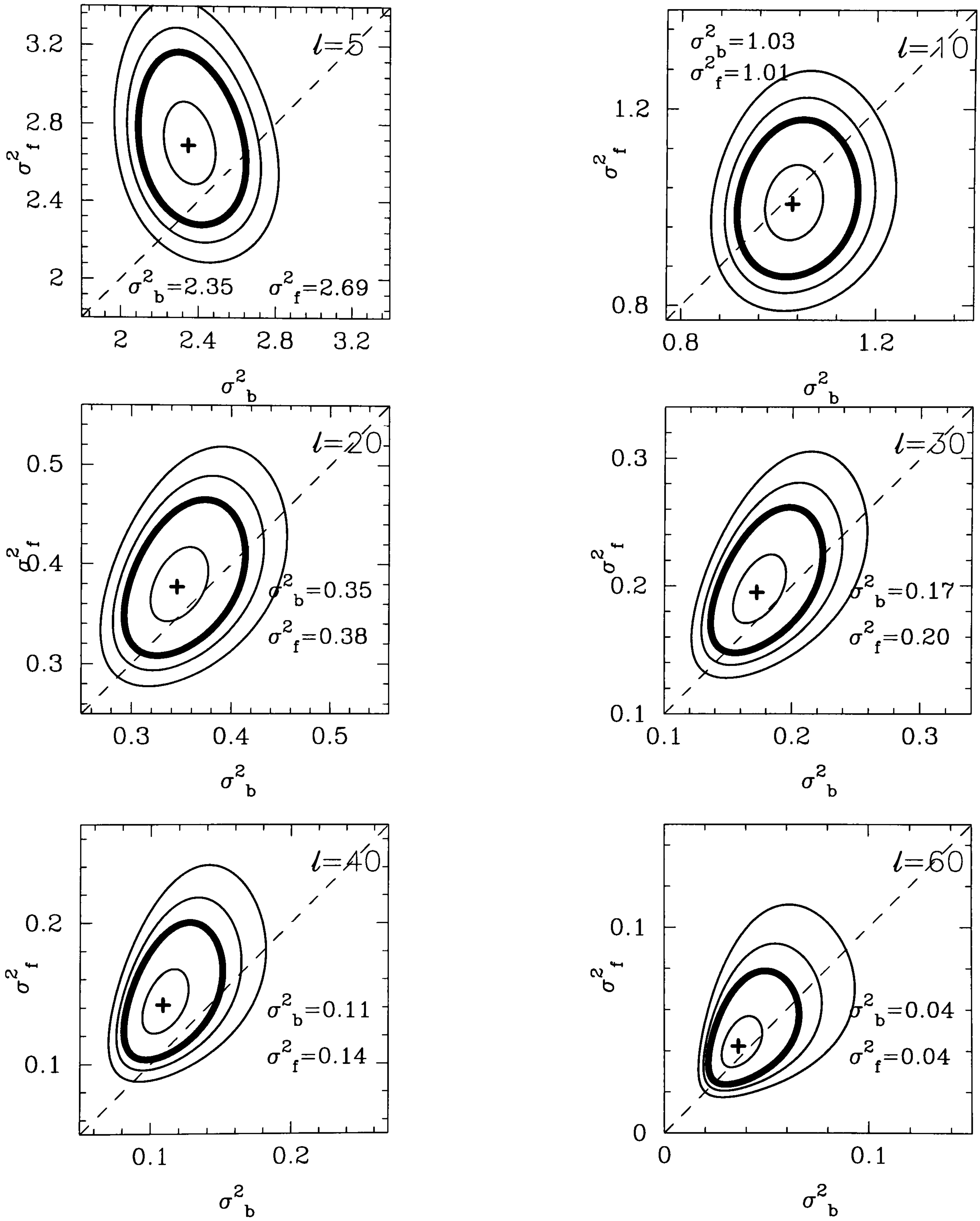


Figure 6.7: Likelihood contours as in Figures 6.1 and 6.4 but for the comparison of PSCz galaxies with $60\mu m$ fluxes between 0.6 and $0.8Jy$ and those with $60\mu m$ fluxes above $0.8Jy$.

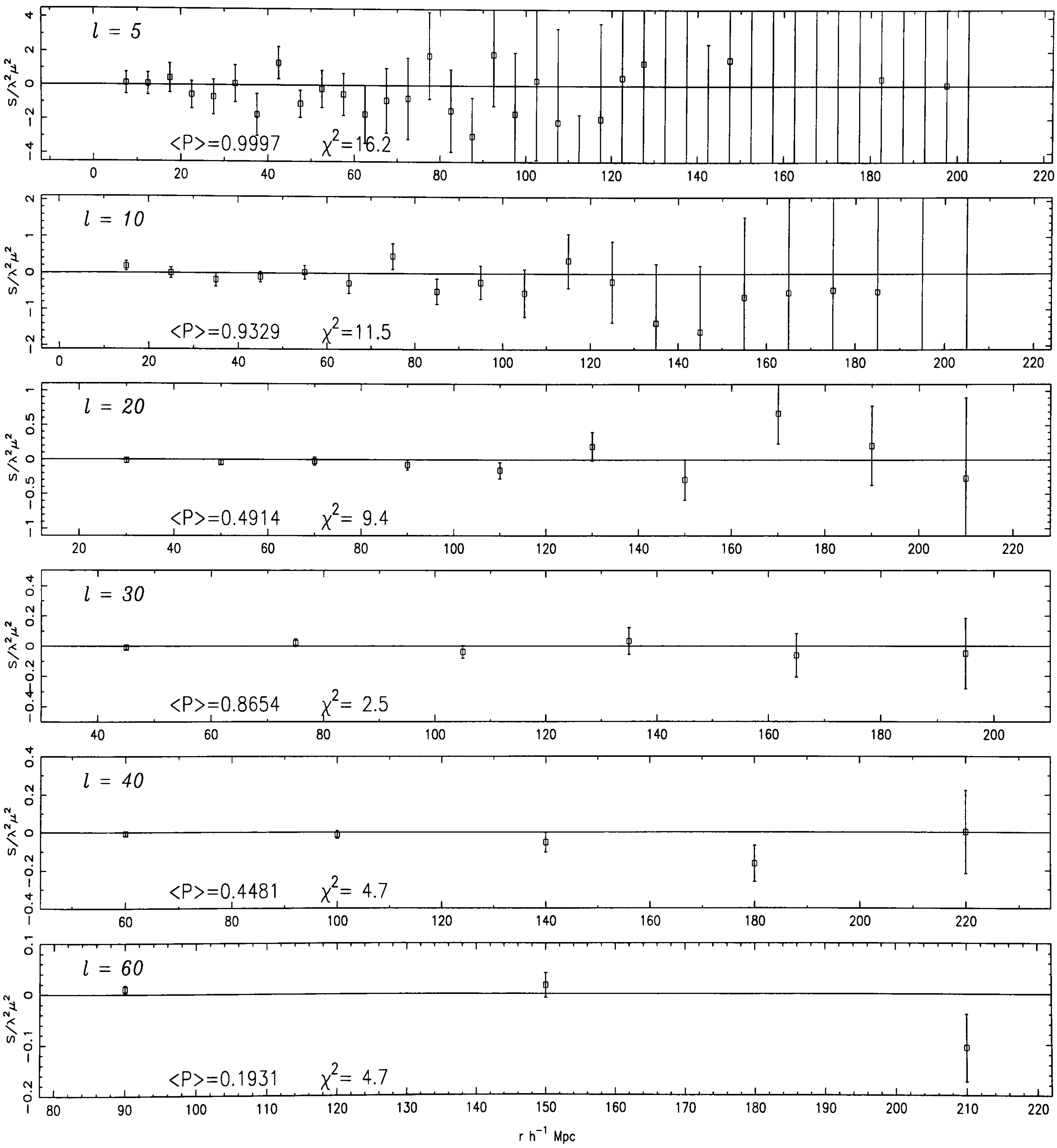


Figure 6.8: As Figures 6.2 and 6.5 but for the comparison of PSCz galaxies with $60\mu m$ fluxes between 0.6 and $0.8Jy$ and those with $60\mu m$ fluxes above $0.8Jy$. Also shown is the probability $\langle P\rangle$ and the χ^2 for the fit $S/\left(\lambda^2\mu^2\right) = 0$.

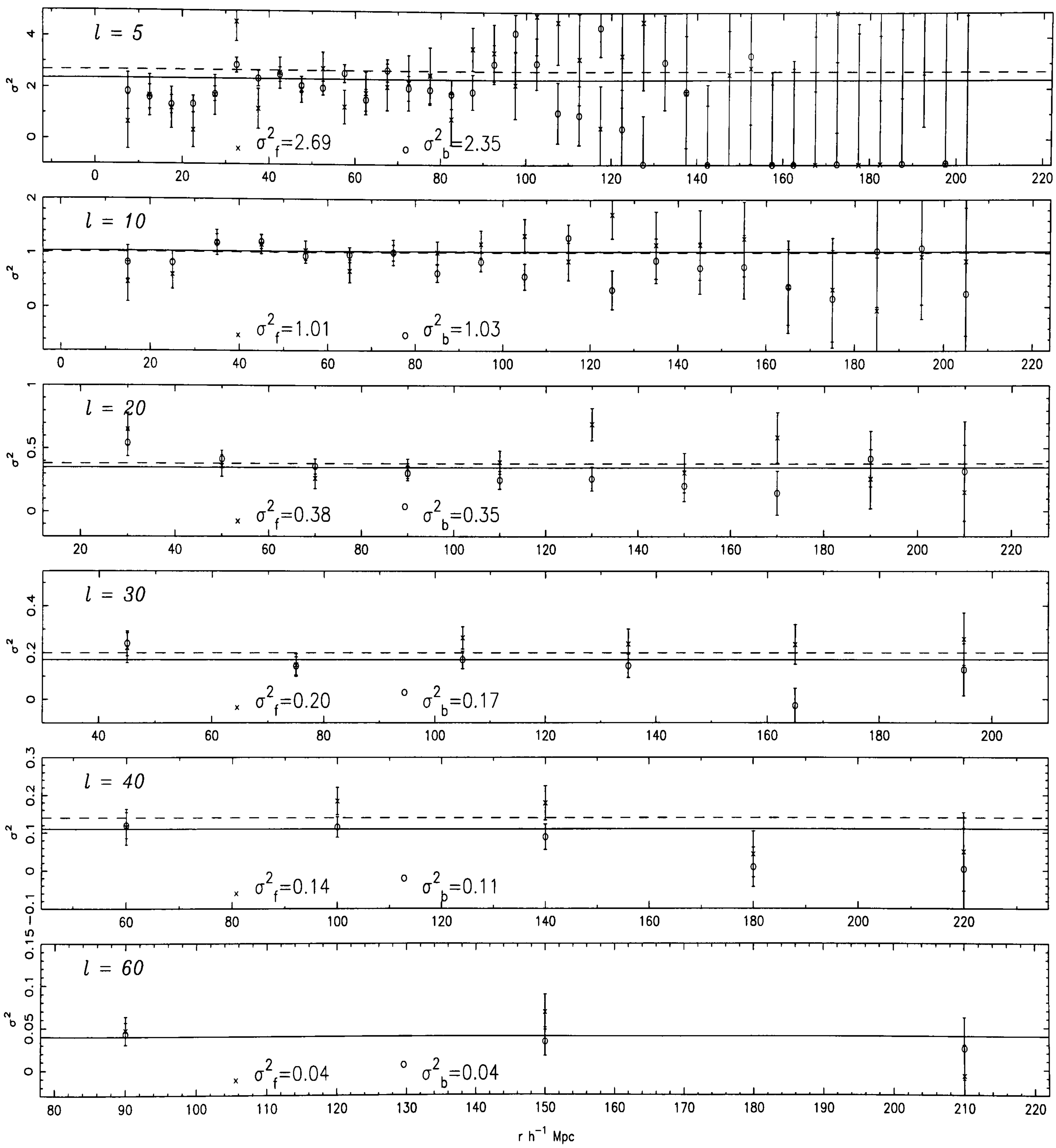


Figure 6.9: As Figures 6.3 and 6.6 but for the comparison of PSCz galaxies with $60\mu m$ fluxes between 0.6 and $0.8Jy$ and those with $60\mu m$ fluxes above $0.8Jy$.

(Efstathiou 1995), I have shown the variances as a function of cell size for the entire PSCz redshift survey in Table 6.1. In this case we just maximise the likelihood

$$L\left(\sigma^2\right) = \prod_i \frac{1}{(2\pi Var\,(S_1^i))^{\frac{1}{2}}} exp\left[\frac{-\left(S_1^i - \overline{N}_1 V \sigma_1^2\right)^2}{2Var\,(S_1^i)}\right] \tag{6.24}$$

with respect to σ_1^2. From Table 6.1 we can see that the variances measured from the PSCz redshift survey are in very good agreement with those measured from the joint IRAS $1.2Jy$ and QDOT surveys.

$l\,h^{-1}$Mpc h^{-1}Mpc	σ^2	68% confidence interval	95% confidence interval	$\sigma^2_{1.2Jy/QDOT}$
10	0.97	0.93 1.00	0.90 1.05	0.96
20	0.35	0.33 0.37	0.31 0.40	0.37
30	0.18	0.16 0.20	0.15 0.22	0.20
40	0.12	0.11 0.13	0.10 - 0.15	0.13
60	0.04	0.03 0.05	0.02 - 0.06	0.04

Table 6.1: Table shows the counts in cells variances, as a function of cell size, for the PSCz redshift survey. Results are shown for cells sizes 10, 20, 30, 40 and 60 h^{-1}Mpc and the 68% and 95% confidence intervals are given. The last column in the table gives the variances, measured by Efstathiou (1995), from the combined IRAS $1.2Jy$ and QDOT surveys.

6.4 Comparison of optical and IRAS galaxy counts in cells

An important issue for our understanding of the formation and evolution of galaxies is whether all types of galaxies are clustered in the same way. This question has been tackled observationally in several ways, using various classes of objects. In the following section I briefly review previous work on the comparison of the clustering strength of galaxies as a function of galaxy properties.

As early as 1936, Hubble noted that the sky distribution of elliptical and spiral galaxies appeared to be different (Hubble 1936). Many people have since investigated this effect (Oemler 1974, Dressler 1980, Whitmore, Gilmore & Jones 1993) and it was quantified to an extent by Postman & Geller (1984) who pointed out that although elliptical galaxies made up only about 15% of the galaxy population in the field, they are the dominant galaxy type in the highest density regions i.e. in the centres of galaxy clusters.

Differential clustering has also been searched for in galaxies of different surface brightness (Davis & Djorgovski 1985, Bothun *et al.* 1986, Mo, McGaugh & Bothun 1994). Davis & Djorgovski (1985) used the Uppsala General Catalogue of Galaxies (UGC, Nilson 1973) to investigate the clustering of low and high surface brightness galaxies. They examined the angular two point correlation function of subsamples divided by surface brightness and concluded that low surface brightness galaxies (LSBGs) have a lower clustering amplitude than high surface brightness galaxies. By matching the sample against a redshift catalogue compiled by John Huchra (consisting mainly of the CfA redshift survey, Huchra *et al.* 1983) they inferred the amplitude of the spatial two point correlation function and again found that LSBGs are less clustered than those of high surface brightness. Davis & Djorgovski (1985) also noted that LSBGs populate regions which are voids in the spatial distribution of high surface brightness galaxies. Bothun *et al.* (1986) on the other hand, using the same catalogue (UGC), but examining the velocity distributions of high and low surface brightness galaxies in cells on the sky of width 10° in declination and 4^h in right ascension, found no discernible difference between the two populations.

Mo, McGaugh & Bothun (1994) examined the spatial cross correlation function of LSBGs with galaxies in both the CfA optical redshift survey (Huchra *et al.* 1983) and the IRAS $2Jy$ survey (Strauss *et al.* 1992a). The LSBG sample used was that of Schombert *et al.* (1992) consisting of LSBGs discovered by visual inspection of plates in the Second Palomar Sky Survey; 171 of these galaxies have measured redshifts. On large scales (i.e. larger than 2 h^{-1}Mpc) they found that the cross correlation function of the LSBGs with the CfA and IRAS galaxies has the same shape as the auto-correlation function of the CfA and IRAS galaxies. However, the amplitude of the cross correlation between the LSBGs and CfA galaxies is a factor ~ 2 lower than the auto correlation of CfA galaxies, and similarly for the ratio of the amplitude of the cross correlation of LSBGs with IRAS galaxies and the auto-correlation amplitude of IRAS galaxies. Mo, McGaugh & Bothun (1994) concluded that LSBGs trace the same distribution as that of the normal galaxies but are less clustered. They also examined the clustering statistics on small scales ($< 2\ h^{-1}$Mpc) and found that there is a dearth of LSB galaxy pairs on these scales. The authors put forward a model in which LSBGs only form within isolated dark matter halos. In halos which interact with other close halos, the LSBGs either do not form or are destroyed in the interactions; they showed that this model could explain the observations in their paper. However, as pointed out by Bothun *et al.* (1986) and Davis & Djorgovski (1985), these analyses are dogged by problems of incompleteness and, in particular, differential incompleteness between the two populations being considered.

Mo & Lahav (1993) investigated the relation between the internal velocities of galaxies (and thus presumably their halo velocities) and the strength of clustering. For spiral galaxies the internal velocity, v_c, was taken to be the measured circular velocity and for elliptical galaxies the velocity dispersion was used. Mo & Lahav (1993) used 1000 galaxies with measured intrinsic velocities as 'targets', and cross-correlated these with the spatial distribution of 'tracer' galaxies for which they used both the optical CfA sample and the $2Jy$ IRAS sample. The local overdensity, δ, of the tracer galaxies around each of the target galaxies was estimated. A weak positive correlation of the correlation function strength with circular velocity was found, and Mo & Lahav (1993) quote $\Delta\delta(r \lesssim 3.6h^{-1}Mpc)/\Delta v_c \lesssim 0.02$. This is a much smaller effect than that predicted by high-biasing models for galaxy formation. Mo & Lahav (1993) concluded that either the circular velocities of the galaxies are not simply related to the circular velocities of the halos in which they reside, or that most dark halos were formed at high redshift, or that the galaxy distribution does not trace the dark matter distribution in a simple way.

White, Tully & Davis (1988) (hereafter WTD) attempted to measure clustering as a function of mass. CDM and related theories of structure formation predict that clustering strength should scale with the circular velocity of the dark halo in which the galaxy resides (White *et al.* 1987a). This would also imply a dependence on luminosity and mass as these scale with the circular velocity. WTD used the *Nearby Galaxies Catalogue* (Tully 1988) which provides information on 2367 galaxies with velocity $v < 300\text{km s}^{-1}$. This information consists of a redshift, magnitude, angular size, inclination angle and velocity width, where available, for each galaxy. WTD split their sample into subsets of galaxies with circular velocities v_c in the ranges: $v_c < 75\text{km s}^{-1}$, $75 < v_c < 150\text{km s}^{-1}$, $150 < v_c < 300\text{km s}^{-1}$, and $v_c > 300\text{km s}^{-1}$. They analysed the percentage of galaxies in regions of density higher than ρ as a function of ρ. They found a definite trend for galaxies with high circular velocity to be situated in the high density regions while those with low circular velocity are in low density regions. The same trend, with similar amplitude, was seen in their CDM simulations of structure formation (where they assigned the 'galaxy' the circular velocity of the dark halo in which it would form).

Clustering differences between radio and optical galaxies, and between IRAS and radio galaxies have also been examined (Shaver 1991, Peacock & Nicholson 1991, Mo, Peacock & Xia 1993). Previous work on the comparison of optical and IRAS galaxies includes Babul & Postman (1990), Lahav, Nemiroff & Piran (1990), Strauss *et al.* (1992b) and Peacock & Dodds (1994). Babul & Postman (1990) analysed the angular and spatial correlation functions and the void probability functions for IRAS (sample limited at $0.7Jy$, see Babul & Postman 1990 for details) and CfA galaxies (the CfA slice of 1091 galaxies – de Lapparent, Geller & Huchra 1986). They concluded that, except for the regions in the centres of clusters, the clustering statistics of the two populations were the same i.e. there is no relative biasing between the optical and IRAS galaxies. Lahav, Nemiroff & Piran (1990), on the other hand, conducted a similar analysis comparing the UGC and ESO optical catalogues with an IRAS sample (flux limited at $0.7Jy$, Meurs & Harmon 1988), and obtained a relative bias $b_{opt}/b_{IRAS} \sim 1.7$. Lahav, Nemiroff & Piran (1990) also pointed out that a relative bias of $b_{opt}/b_{IRAS} \sim 1.4 - 2.3$ would explain the difference in the measured values of Ω from the IRAS dipole studies (which are high, $\Omega : 0.5 - 0.8$) and the values measured from the optical dipole studies (which are low, $\Omega : 0.2 - 0.3$). These dipole studies include: Yahil, Walker & Rowan-Robinson (1986), Meiksin & Davis (1986), Villumsen & Strauss (1987), Lahav (1987), Lahav, Rowan-Robinson & Lynden-Bell (1988), Lynden-Bell, Lahav & Burstein (1989), Kaiser & Lahav (1989), Strauss & Davis (1989).

Peacock & Dodds (1994) analysed many surveys of galaxies, both optical and IRAS, as well as cluster surveys and surveys of radio galaxies. They attempted to account for redshift-space distortion, non-linear evolution and the relative bias between the different objects in the data sets in order to transform the present day power spectra of these objects to the initial power spectrum, assuming gravitational instability and a Gaussian density field. They found a best fit value for the optical to IRAS relative bias of $b_{opt}/b_{IRAS} = 1.3$. The values of the relative bias between optical and IRAS galaxies derived in the previous chapters are consistent with $b_{opt}/b_{IRAS} = 1.3$ (see Chapters 2 and 5).

In this section I compare the clustering properties of galaxies in the Stromlo-APM optical redshift survey (see Chapter 2) with galaxies in the IRAS PSCz redshift survey (see Chapter 5). This is done via a counts in cells analysis, and compares the clustering of optical and IRAS galaxies *within the same volume of space.*

The analysis is done to a radial distance of 300 h^{-1}Mpc, for cells of size 10 h^{-1}Mpc, 20 h^{-1}Mpc, 30 h^{-1}Mpc and 40 h^{-1}Mpc, and a joint APM/PSCz mask is used. The Stromlo-APM survey covers 1.31 sr of the sky, and the further small amount of masking in this region from the PSCz mask reduces the area of sky used to 1.24 sr. The number of galaxies in the Stromlo-APM survey in this region (between absolute magnitudes -23 and -14, and with distances between 0 and 300 h^{-1}Mpc) is 1618. The number in the PSCz survey in this region (also with distances between 0 and 300 h^{-1}Mpc, and fluxes between 0.6 and $2000Jy$) is 1369. The two surveys therefore have roughly similar surface densities.

We must now exclude, from the PSCz survey, the galaxies which are in common to both surveys so that we have two independent samples.

As the Stromlo-APM survey is sparsely sampled, and roughly consists of 50% spiral and 50% elliptical galaxies, we expect very roughly about 1 in 40 of the galaxies in the PSCz survey (in the Stromlo-APM region of sky) to appear in the Stromlo-APM survey as well. This is about $30 - 40$ galaxies. To select these common galaxies I proceed as follows. For every galaxy in the Stromlo-APM survey which has an absolute magnitude between -23 and -14, and a distance between 0 and 300 h^{-1}Mpc, and is included by the joint APM/PSCz mask, I calculate the angular separation of every galaxy in the PSCz survey. The separation of the two galaxies is then projected onto the coordinate axes defined by the major and minor axes and the position angle of the IRAS 1σ error ellipse for the PSCz galaxy in question. The measured separation along the major and minor axes is then divided by the length of the major and minor axes respectively and these two distances are added in quadrature to obtain the number of sigma that separates the two galaxies. The velocity difference $v_{apm} - v_{pscz}$ of the pair is also calculated. For each Stromlo-APM galaxy information is stored about the nearest (in terms of number of σ) PSCz galaxy.

Figure 6.10 shows a plot of the velocity differences of the Stromlo-APM/PSCz pairs, against the number of sigma that the pairs of galaxies are separated by. Some of these points will represent real matches, i.e. the pair of galaxies are actually the same galaxy, and we must exclude the galaxy from the PSCz survey when doing a joint counts in cells analysis. In Figure 6.10 we can see that there is a dense clump of pairs between 0 and 5σ which have velocity differences of between about -600 and 600km s^{-1}. These almost certainly represent real matches. A maximum velocity difference of 600km s^{-1} is reasonable given the random errors in the velocity measurements of the Stromlo-APM survey ($\sim$ 50km s^{-1}) and the PSCz survey ($\sim 100 - 150$km s^{-1}). These are the galaxies which I remove from the PSCz survey. The smaller group of pairs between 5 and 7σ, with similar separation in velocity space, may or may not be matches but for this analysis I assume that 5σ is a reasonable upper limit on the separation of true matched pairs. Since the number of galaxies in this smaller group is only 7 the analysis is not going to be affected by inclusion or exclusion of these galaxies. The galaxies lying at very large sigma but small velocity difference are presumably neighbouring clustered PSCz galaxies. The above procedure removes 39 galaxies from PSCz sample leaving 1330 galaxies for the counts in cells analysis.

In Figure 6.11 the information is shown in the form of a histogram of the number of

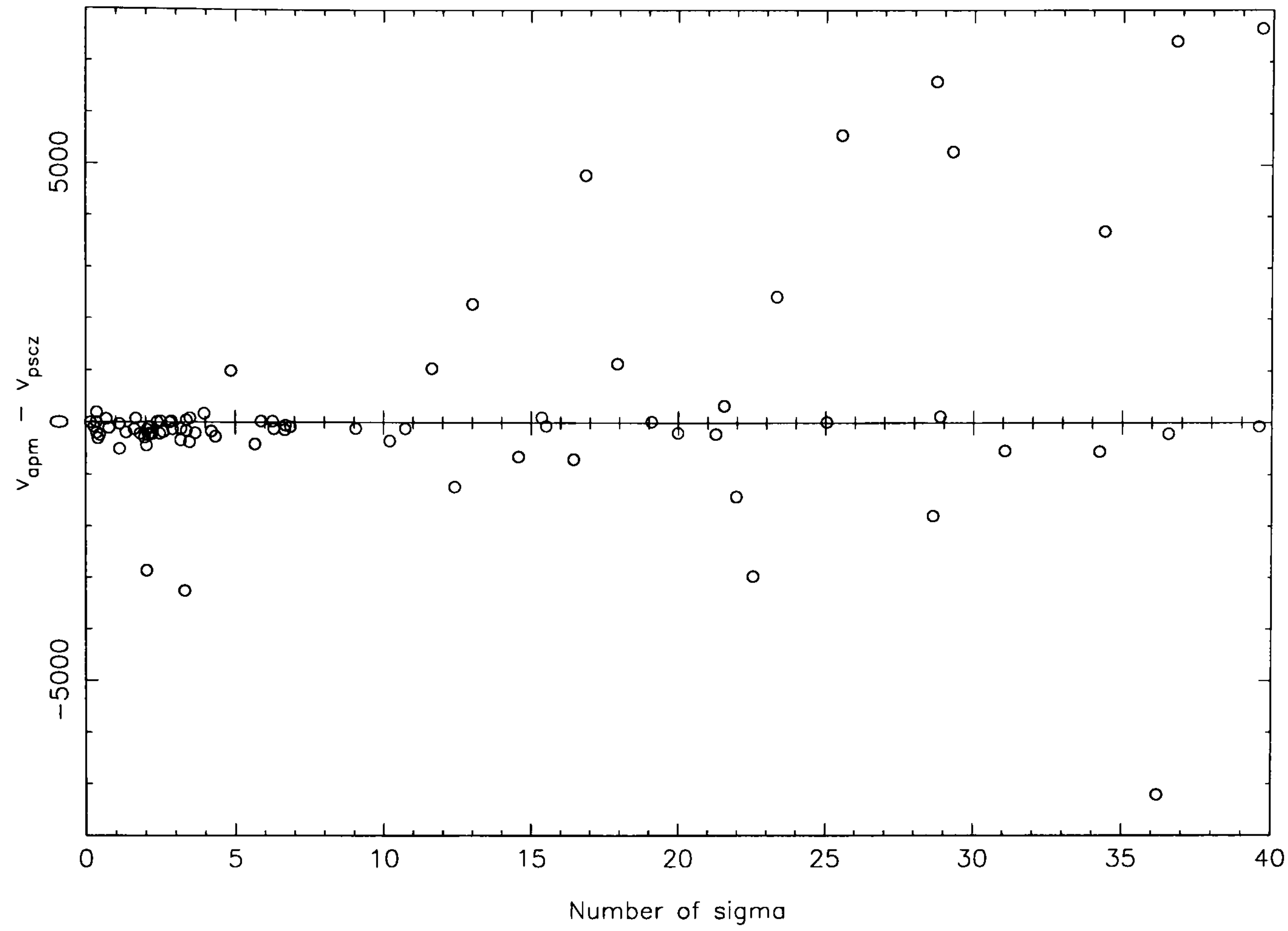


Figure 6.10: The difference in velocity between a Stromlo-APM galaxy and its nearest neighbour in the PSCz galaxy, plotted as a function of the separation of these two galaxies in terms of σ, the one sigma positional error on the IRAS error ellipse.

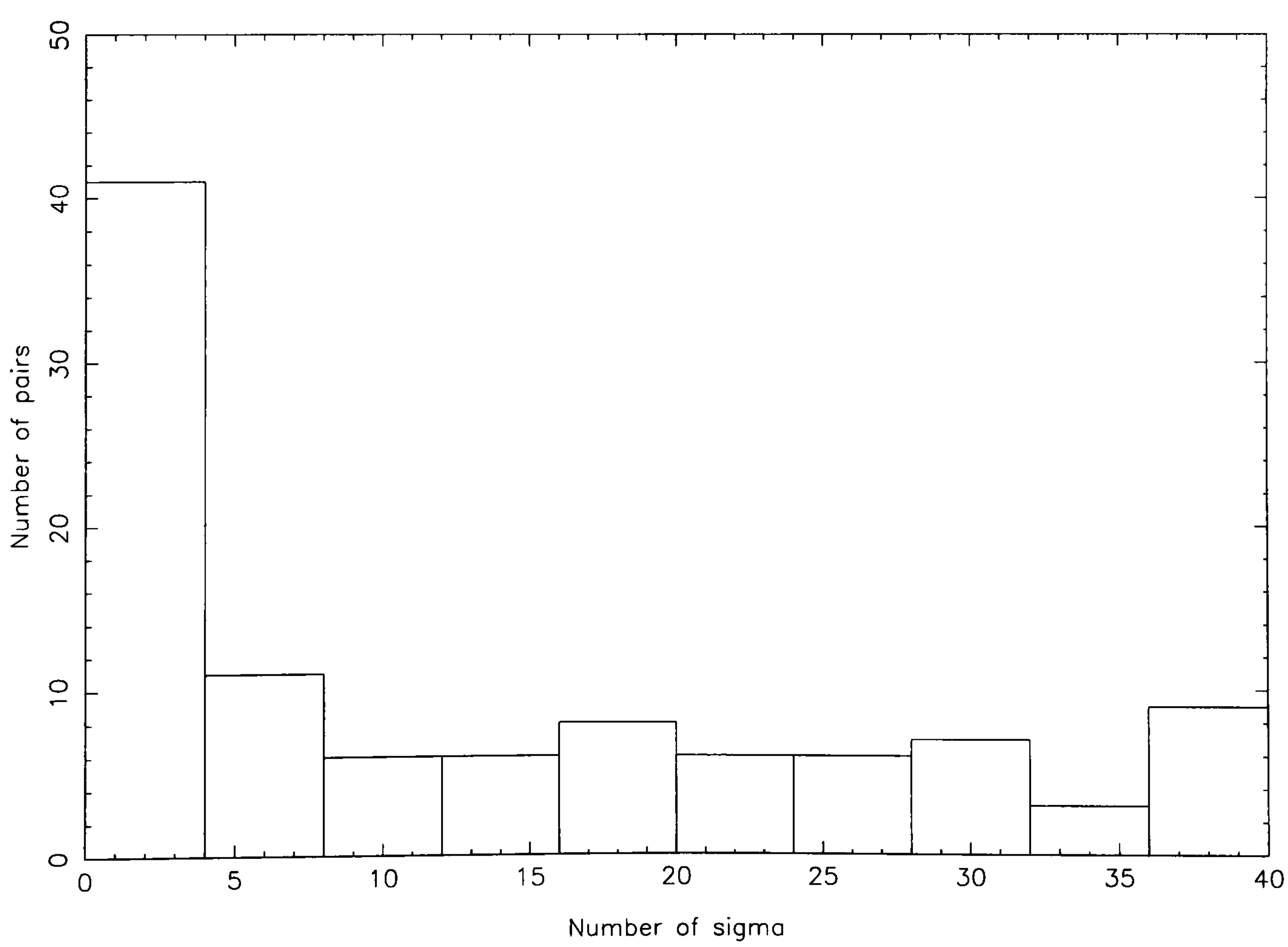


Figure 6.11: The figure shows a histogram of the number of Stromlo-APM/PSCz pairs against the separation in terms of σ, the one sigma positional error of the IRAS error ellipse.

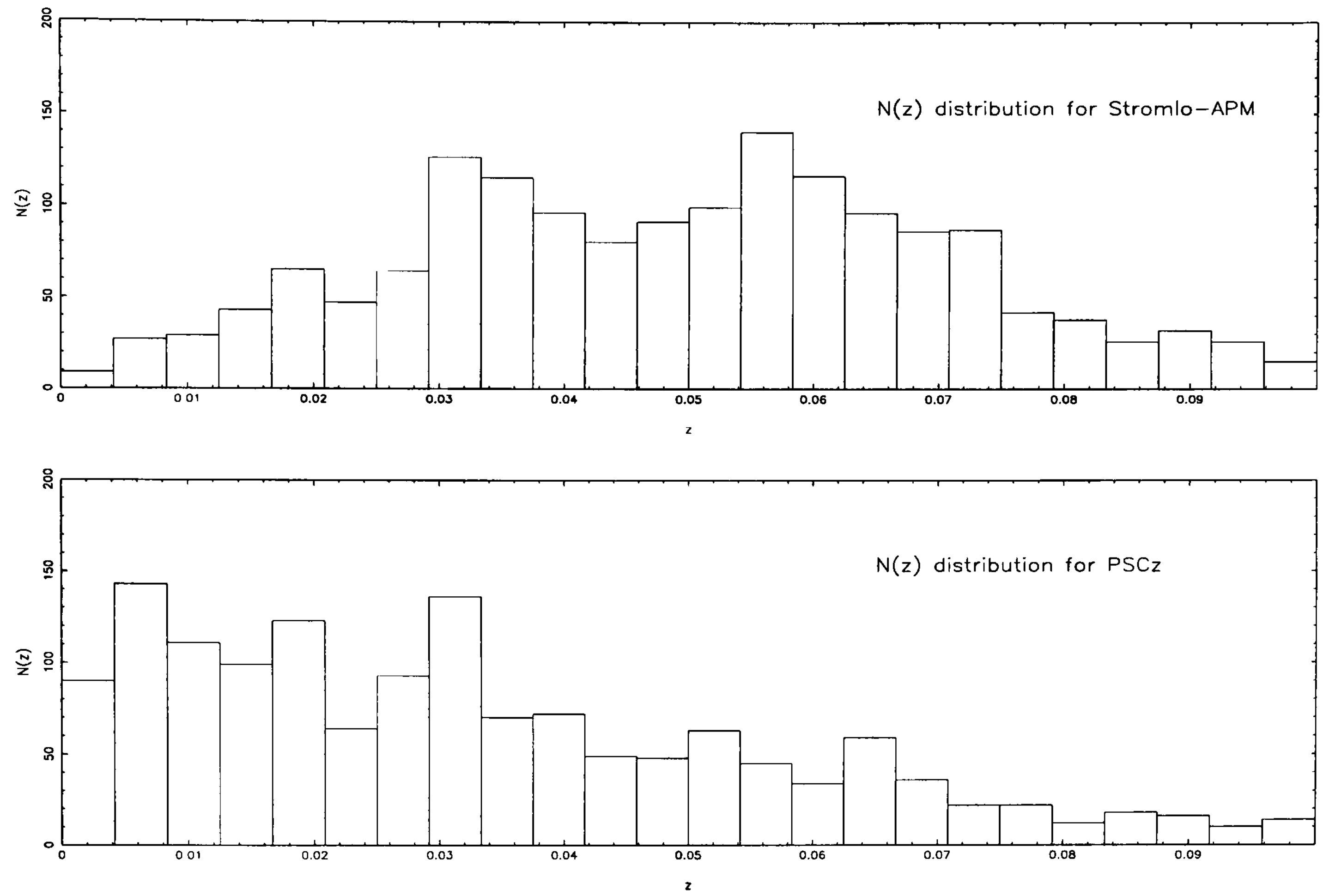


Figure 6.12: Histograms of the number of galaxies as a function of redshift, z, for the two samples used in the comparison of IRAS and optical galaxies.

pairs against the separation in terms of sigma. Again, this shows that a separation of order 5σ should find most of the real matches.

Figure 6.12 shows the $N(z)$ distributions for the Stromlo-APM and the PSCz galaxies used in the comparison. It can be seen that the Stromlo-APM survey goes deeper than the PSCz survey, because of the sparse sampling strategy employed in the construction of that survey.

In the joint counts in cells analysis I shall only use radial shells which contain more than $\sim$ 20 filled cells (again defining filled as $\geq$ 70% unmasked). Table 6.2 shows the number of filled cells as a function of the radial shell number i, for the joint Stromlo-APM/PSCz mask. For each of the cell sizes, the first 4 radial shells are unusable making the 5^{th} radial shell the first to be analysed. To a radial distance of 300 h^{-1}Mpc, this gives 26, 11, 6 and 3 radial shells in the analysis of the 10 h^{-1}Mpc, 20 h^{-1}Mpc, 30 h^{-1}Mpc and 40 h^{-1}Mpc size cells respectively.

The results of the counts in cells comparisons are shown in Figure 6.13 again as likelihood contours. Figure 6.14 and Figure 6.15 show the $S/\left(\lambda^2\mu^2\right)$ statistic and the run of σ^2 with radius. The χ^2 and the probability $\langle P\rangle$ shown in Figure 6.14 are appropriate for the case $\langle S\rangle / \left(\lambda^2\mu^2\right) = 0$.

i	Number of active cells	i	Number of active cells
1	0	16	281
2	1	17	316
3	5	18	356
4	9	19	407
5	19	20	455
6	32	21	501
7	47	22	551
8	61	23	612
9	81	24	665
10	9	25	716
11	121	26	778
12	149	27	838
13	178	28	903
14	208	29	968
15	247	30	1042

Table 6.2: This table shows the number of cells per radial shell (i) which have an unmasked volume $> 70\%$ of their total volume for the joint APM/PSCz mask. For the analysis in this chapter, only radial shells with at least 19 active cells are used.

The data in Figure 6.13 suggest that the optical galaxies are more clustered than the IRAS galaxies on scales 10, 20 and 30 h^{-1}Mpc at $\gtrsim 95\%$ confidence. For a cell size of 40 h^{-1}Mpc the two distributions are consistent with $\sigma^2_{pscz} = \sigma^2_{apm}$. The S statistic (Figure 6.14) shows no definite difference in the two distributions except possibly for the 30 h^{-1}Mpc cells. For the smaller cell sizes (10 and 20 h^{-1}Mpc) this apparent discrepancy between the likelihood contours and the S statistic could be due to overestimation of the errors (see discussion at the end of Section 6.1). We may also worry, in this analysis, that the number of cells analysed is small (see Table 6.2) and that convergence to the central limit theorem is not achieved. To test whether the differences between the clustering properties of optical and IRAS galaxies apparent in Figure 6.13 are real, or just sampling fluctuations (as Figure 6.14 would suggest), it would be worthwhile carrying out this analysis using an optical survey which covered a much greater area on the sky. We could then be more confident of convergence to central limit theorem behaviour.

As well as maximising the likelihood as a function of σ^2_{apm} and σ^2_{pscz} we can look at the problem slightly differently and maximise with respect to one of the variances, say σ^2_{pscz}, and the relative linear bias parameter between the Stromlo-APM and PSCz surveys. Figure 6.16 shows the result of this comparison. I have plotted the contours in σ^2_{pscz}, b space where b is defined:

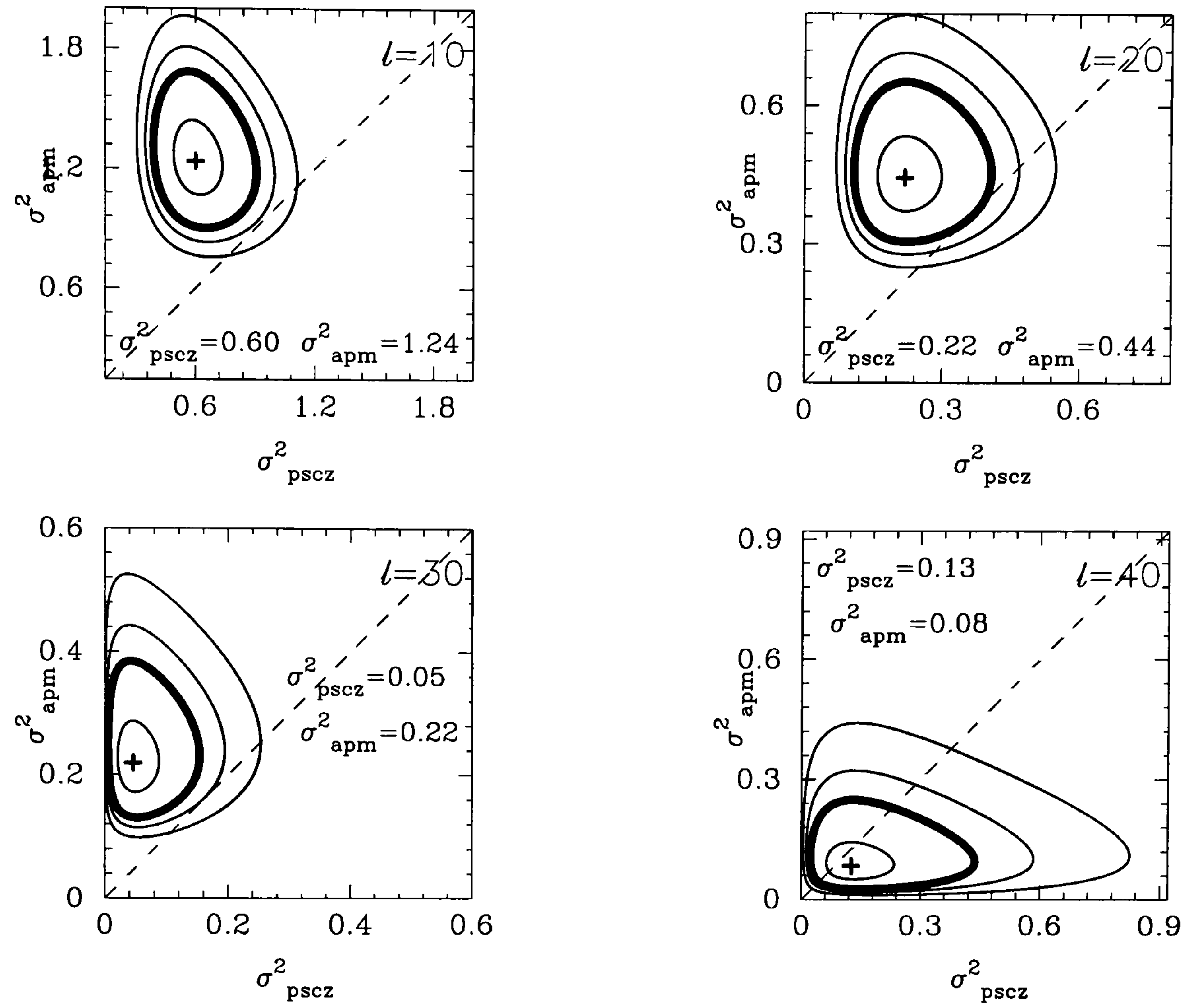


Figure 6.13: Likelihood contours for variances of counts in cells for a comparison of the optical Stromlo-APM redshift survey with the IRAS PSCz redshift survey.

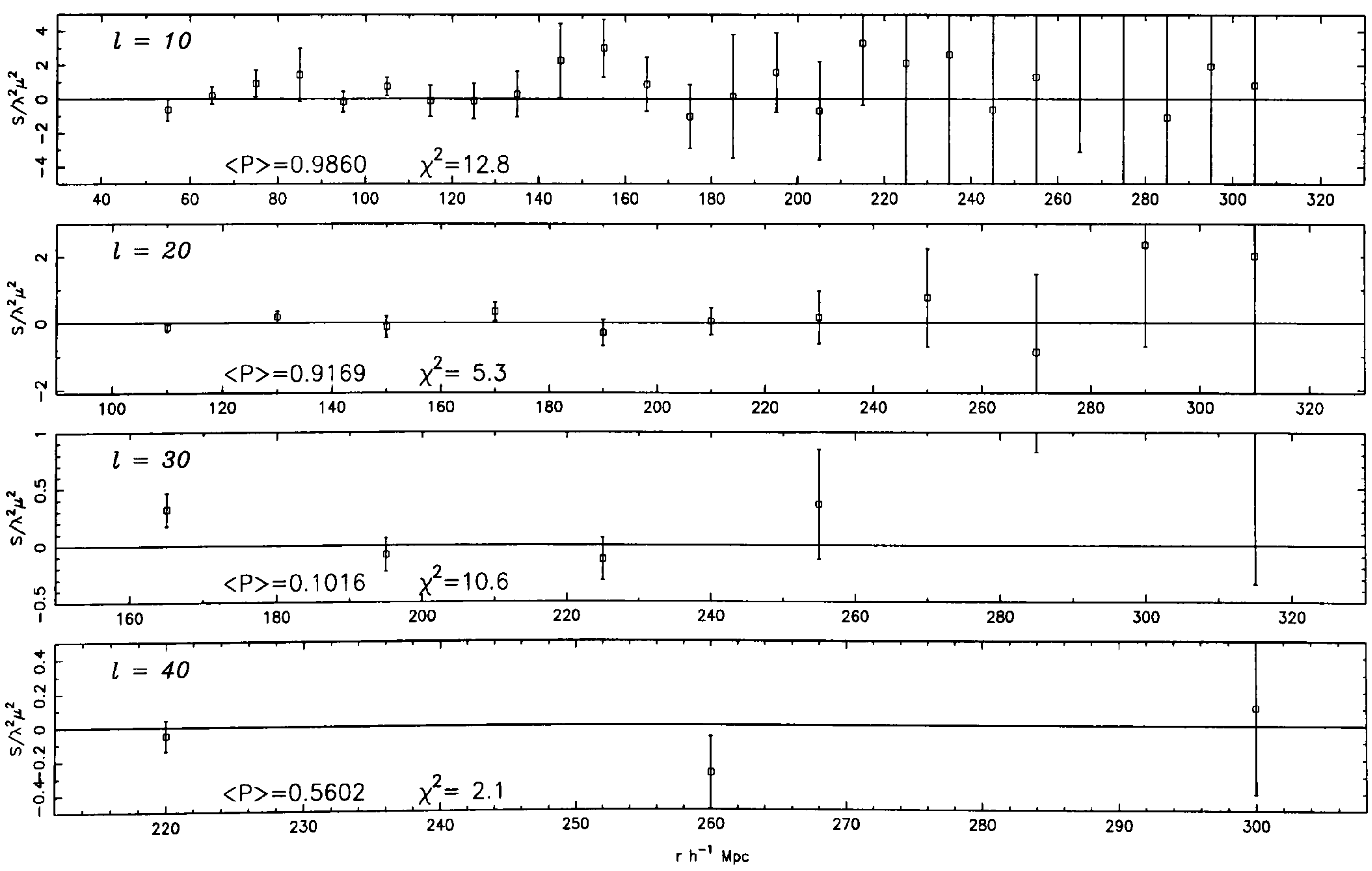


Figure 6.14: The statistic $S/\left(\lambda^2\mu^2\right)$ (equation 6.14) plotted as a function of radial distance for the comparison of the Stromlo-APM survey with the PSCz survey. The panels show results for cells of size 10, 20, 30 and 40 h^{-1}Mpc from top to bottom. The values of $\langle P\rangle$ and χ^2 are for the fit $S/\left(\lambda^2\mu^2\right)=0$

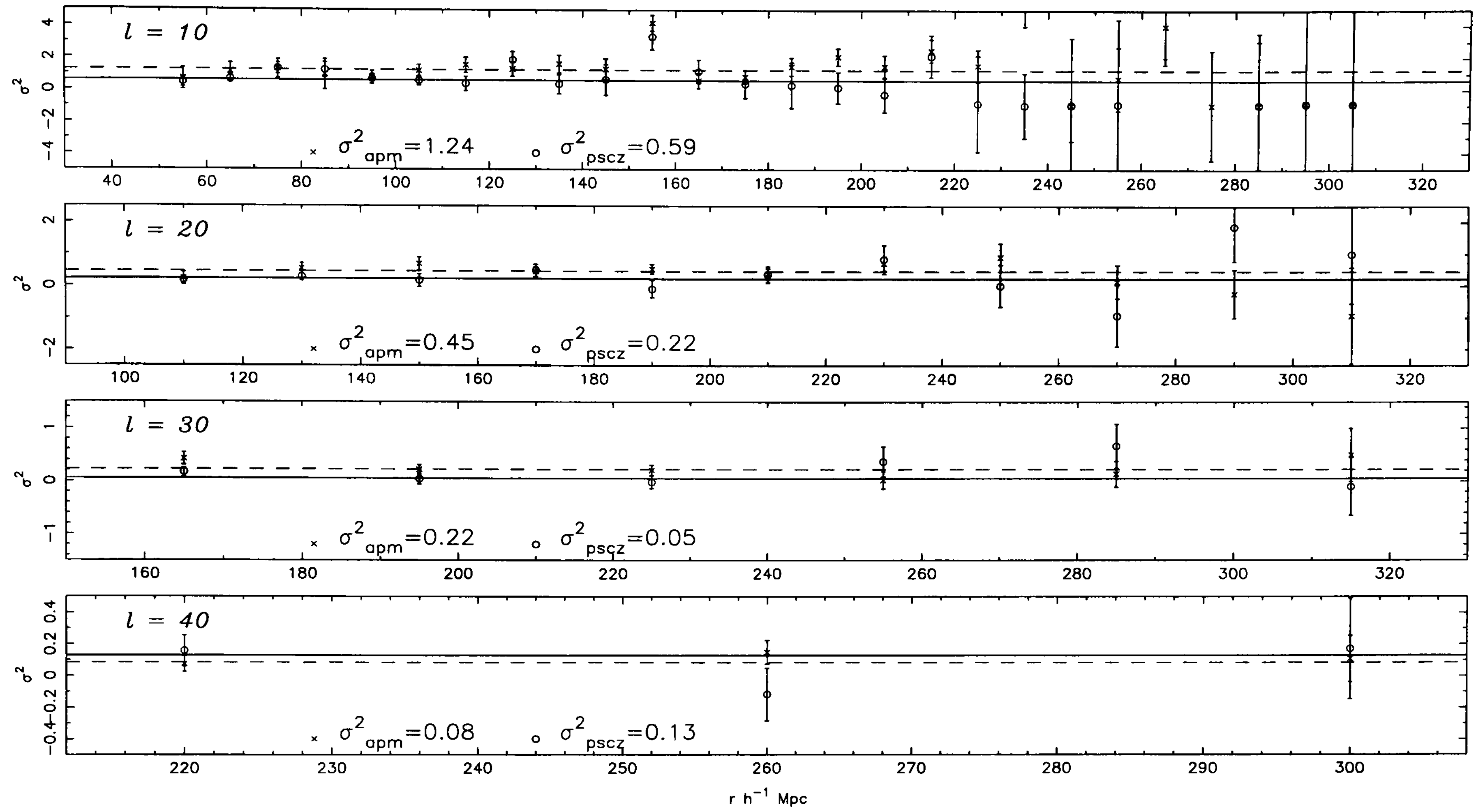


Figure 6.15: The variances of counts in cells as a function of radial distance r for cells of size 10, 20, 30 and 40 h^{-1}Mpc (top to bottom) in the comparison of the Stromlo-APM redshift survey with the PSCz redshift survey.

$$\sigma^2_{opt} = b^2\sigma^2_{iras}. \tag{6.25}$$

The results shown in Figure 6.16 are in broad agreement with the results in Chapter 5 which suggested, from the power spectrum analysis, that the relative bias parameter between the optical and IRAS galaxies is ~ 1.4. This lends support to the idea that the differences between the IRAS and optical galaxy distributions are real and not sampling fluctuations. In fact, if we remove the ellipticals and S0s from the Stromlo-APM survey and correlate just the spirals and irregulars with the PSCz survey, the relative bias decreases substantially. On every scale it becomes consistent with no bias. The likelihood contours for the Stromlo-APM and PSCz variances in this case are shown in Figure 6.17 and the likelihood contours in terms of σ^2_{pscz} and b are shown in Figure 6.18. For each cell size the value $b = 1$ is allowed within the 95% contour level.

For completeness Tables 6.3 and 6.4 show the maximum likelihood values of the variances if we analyse the 1618 Stromlo-APM galaxies within the joint masked region independently of the PSCz galaxies, and similarly, analysing the 1369 PSCz galaxies independently of the Stromlo-APM galaxies. The Stromlo-APM values are in agreement with

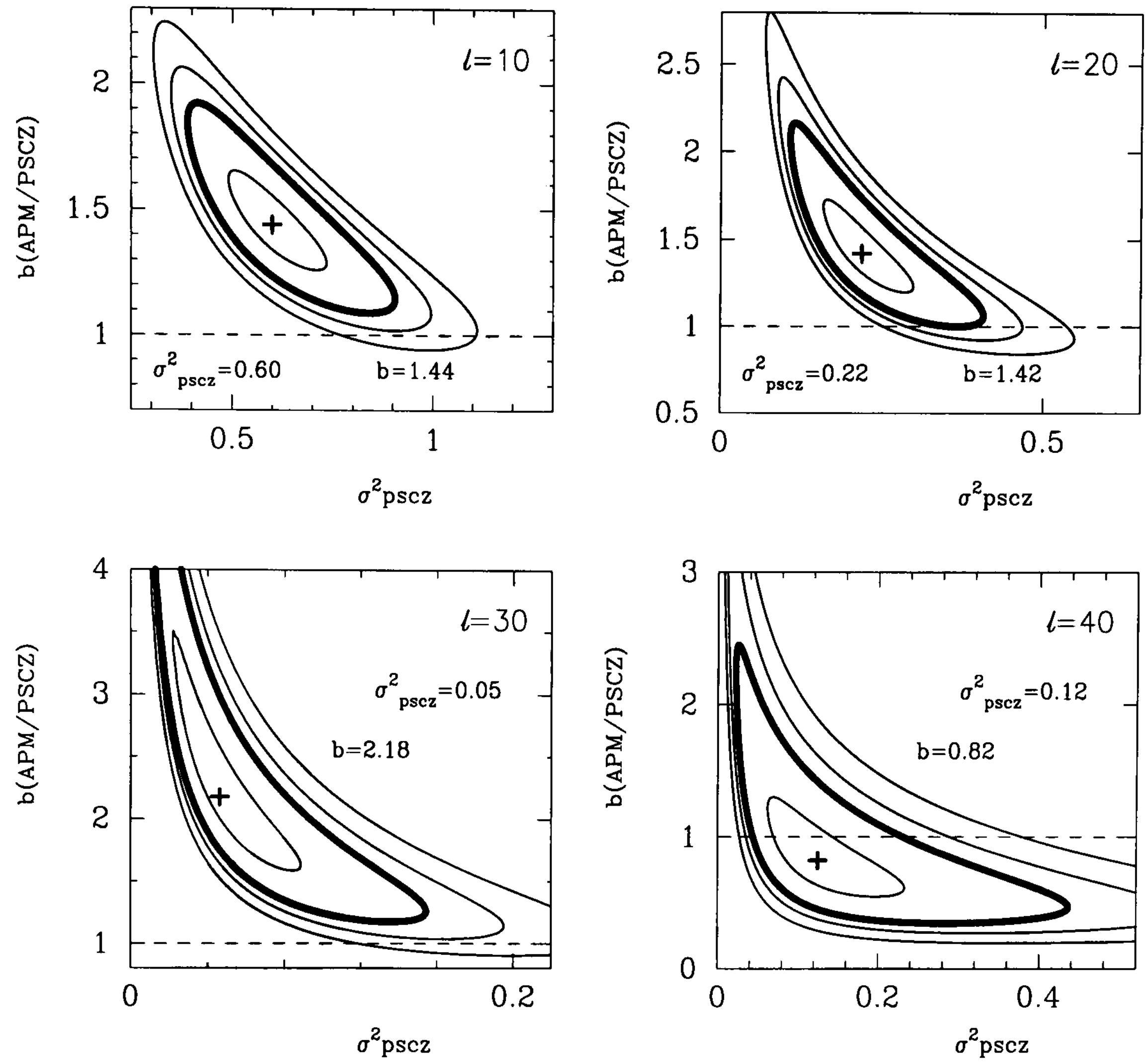


Figure 6.16: Likelihood contours for the optical-IRAS comparison plotted in terms of the PSCz variance and the bias parameter b ($\sigma^2_{opt} = b^2\sigma^2_{iras}$).

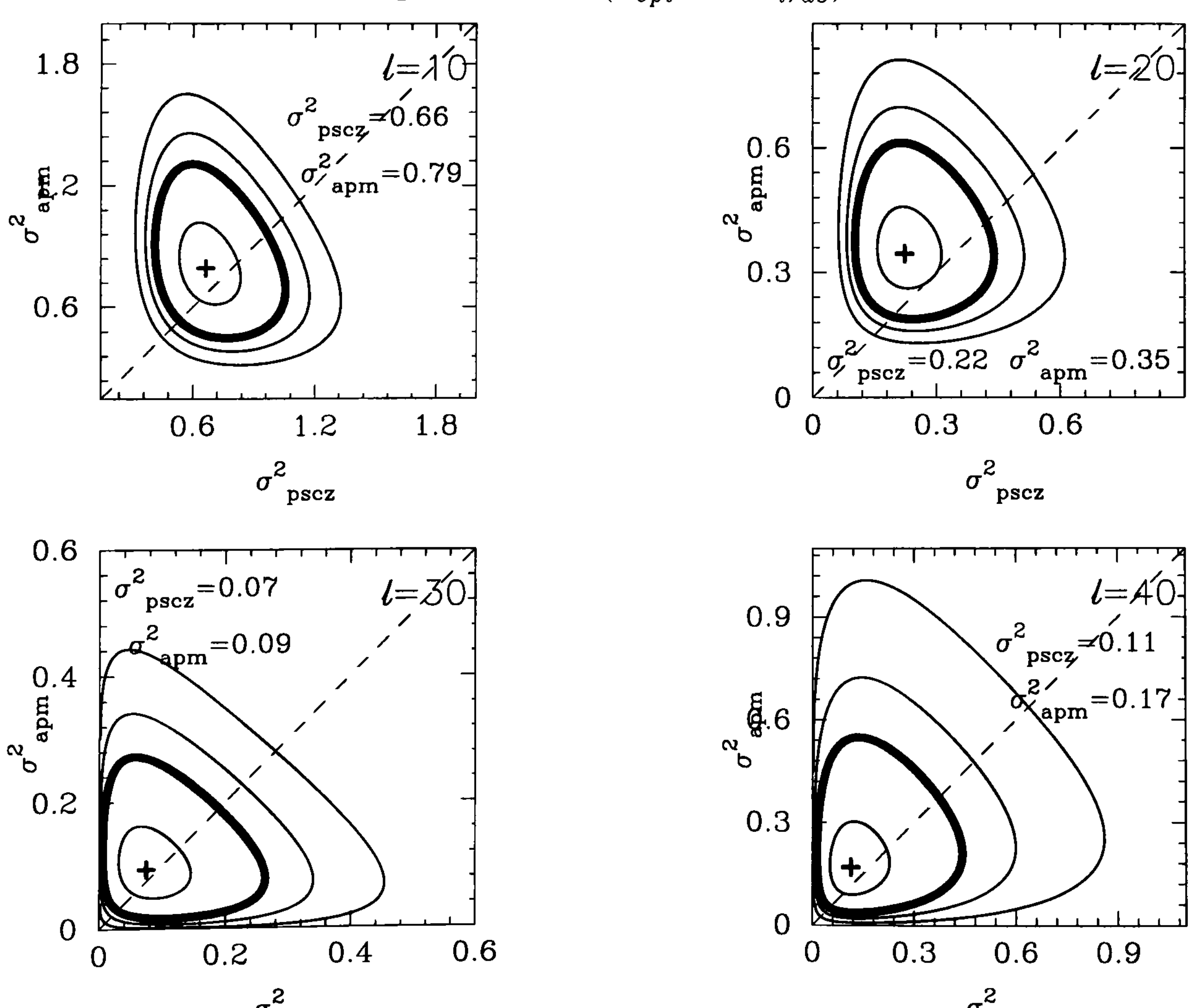


Figure 6.17: Likelihood contours for the optical-IRAS comparison. In this case only the spirals and irregulars in the Stromlo-APM survey are used in the analysis.

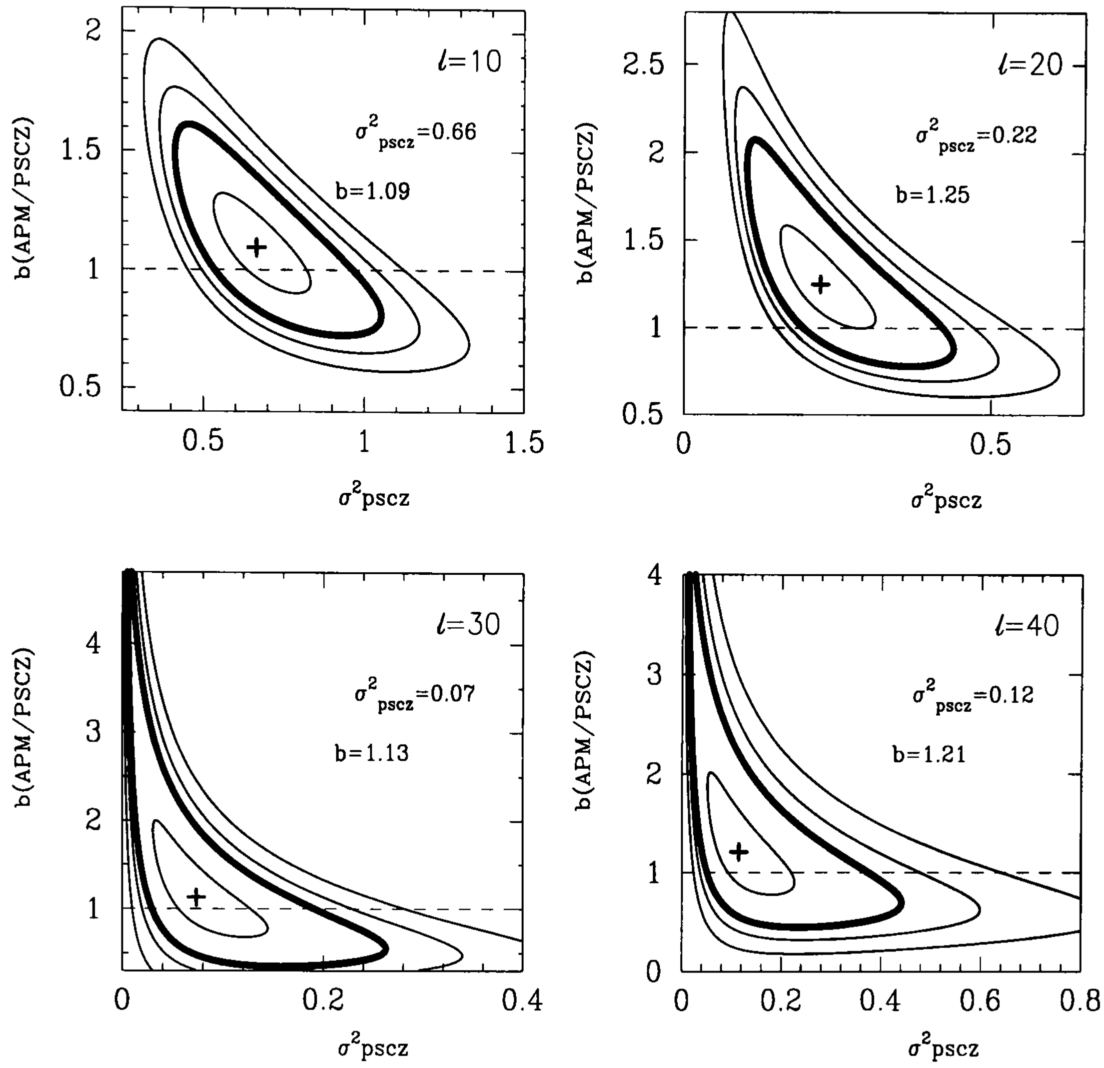


Figure 6.18: Likelihood contours for the relative bias parameter b between optical and IRAS galaxies. In this case only the spirals and irregulars in the Stromlo-APM survey are used in the analysis.

$l\,h^{-1}$Mpc	σ^2	68% confidence interval	95% confidence interval	No. shells analysed
10	1.56703	1.43513 - 1.70796	1.31721 1.85676	26
20	0.47122	0.41245 - 0.53891	0.36227 0.61459	11
30	0.23222	0.19163 - 0.28294	0.15940 - 0.34507	6
40	0.08552	0.05446 - 0.13039	0.03297 0.19496	3

Table 6.3: Values of σ^2 for analysis of the 1618 Stromlo-APM galaxies in the joint Stromlo-APM/PSCz region.

those of Loveday *et al.* (1992a). The PSCz values are compatible, to within the error bars, with those estimated previously in this chapter (see Table 6.1).

6.5 Summary

I have analysed counts in cells for the newly completed IRAS PSCz galaxy redshift survey, particularly to test whether there is any sign of a flux error in the parent catalogue e.g. a variation in the effective depth of the catalogue over the sky caused by calibration errors in the IRAS satellite software. There is no evidence for such an effect at any scale to a high degree of confidence.

$l\,h^{-1}$Mpc	σ^2	68% confidence interval	95% confidence interval	No. shells analysed
10	0.82903	0.71484 0.95885	0.61718 - 1.10229	26
20	0.23376	0.17205 - 0.31011	0.12359 0.40154	11
30	0.09060	0.04525 0.15243	0.01193 0.23434	6
40	0.10635	0.05856 0.17770	0.02585 - 0.28610	3

Table 6.4: Values of σ^2 for analysis of the 1369 PSCz galaxies in the joint Stromlo-APM/PSCz region.

In the comparison between the IRAS PSCz redshift catalogue and the optical Stromlo-APM redshift catalogue differences in the clustering strength of optical and IRAS galaxies were detected. In particular on scales of 10, 20, and 30 h^{-1}Mpc the distributions were different at $> 95\%$ confidence. These results are consistent with those in Chapter 5 for the relative clustering strength of IRAS and optical galaxies, and with the results of Lahav, Nemiroff & Piran (1990) and Peacock & Dodds (1994).

Chapter 7

Spherical harmonic analysis of redshift-space distortions

As described in earlier chapters, peculiar motions of galaxies provide us with a probe of the cosmological density parameter Ω or, more precisely, the parameter $\beta = \Omega^{0.6}/b$ where b is the bias parameter between the galaxy distribution and the mass distribution, $P(k)_{gal} = b^2 P(k)_{mass}$. In most analyses of the effect of peculiar velocities on clustering statistics (i.e. redshift-space distortion analyses) the bias parameter b is assumed to be scale independent and independent of galaxy type and luminosity. These assumptions may not be true. We would hope that it is a reasonable assumption that the bias parameter is scale independent on large scales however (Kauffman, Nusser & Steinmetz 1996, Bardeen *et al.* 1986), otherwise the value of β obtained from measurements of redshift-space distortions is ambiguous.

We have seen evidence in previous chapters that not all types of galaxy can be unbiased tracers of the underlying mass distribution (see Chapters 2, 5 and 6). Larger data sets than are currently available will be required in order to perform redshift-space distortion analysis as a function of galaxy type and/or luminosity. In Chapter 1 (Section 1.5.2) I briefly discussed some techniques to recover the quantity β using redshift-space distortion analyses. In the next section I expand upon this discussion and review some recent results.

There have been several approaches to utilizing redshift-space distortions to measure the value of β. One approach is to use the linear theory relation between the power spectrum (or correlation function) in redshift-space and the power spectrum (or correlation function) in real-space (Kaiser 1987). This can be used with or without a correction for the non-linearities introduced by small scale incoherent peculiar velocity fields. It is necessary, however, to obtain a measure of the clustering in real-space. This is a non trivial requirement as galaxy surveys (except two-dimensional surveys) are done in redshift-space. This approach has been used, for example, by Baugh (1996) and Loveday *et al.* (1995b).

Baugh (1996) measured the real-space correlation function of galaxies in the APM galaxy survey in the magnitude range $17 \leq b_J \leq 20$ by Fourier transforming the measured real-space power spectrum (Baugh & Efstathiou 1993, Baugh & Efstathiou 1994). The real-space correlation function was then compared with the redshift-space estimate of the correlation function of the Stromlo-APM galaxies (Loveday *et al.* 1995a). This is similar to what was done in Chapter 2 except in the latter case the power spectrum was used instead of the correlation function. In the course of the analysis assumptions must be made about the evolution of clustering and the cosmological model in order to invert the two-dimensional statistics to three dimensions. Baugh (1996) found, assuming $\Omega = 1$ and that clustering is fixed in comoving coordinates (this assumption corresponds to $\alpha = 0$ in Table 2.2 of Chapter 2), that $0.38 \leq \beta \leq 0.81$. If instead it is assumed that clustering evolves according to linear perturbation theory, then $-0.02 \leq \beta \leq 0.39$ (1σ error ranges). These results are consistent with the values of β measured from the Stromlo-APM survey in Chapter 2. Loveday *et al.* (1995b) also used the Stromlo-APM redshift survey to measure β but obtained the estimate of the real-space correlation function via a cross-correlation of galaxies in the Stromlo-APM survey with galaxies in the parent APM survey (see Saunders, Rowan-Robinson & Lawrence 1992 for a description of this technique). Loveday *et al.* (1995b) obtained a result $-0.03 \leq \beta \leq 0.75$ (2σ error range), again consistent with the results previously stated.

Because we only measure the radial component of a galaxy's velocity we introduce a preferred direction into the Universe, and the measured clustering statistics become spatially anisotropic. The degree of anisotropy depends on the magnitude of the peculiar velocities which in linear theory is related to the parameter β. We can therefore avoid the problem of having to obtain a real-space estimate of galaxy clustering altogether if we analyse the anisotropies in the redshift-space power spectrum or correlation function to determine the value of β. As mentioned in Chapter 4 we can expand the redshift-space power spectrum $P^s(k)$ in terms of Legendre polynomials i.e.

$$P^s(k,\mu) = \sum_{l=0}^{\infty} P_l^s(k)\mathcal{P}_l(\mu), \qquad (7.1)$$

where μ is the angle between the line of sight and the wave vector $\mathbf{k}$. The coefficients $P_l^s(k)$

are given by

$$P_l^s(k) = \frac{2l+1}{2}\int_{-1}^{+1} d\mu P^s(k,\mu)\mathcal{P}_l(\mu), \tag{7.2}$$

and the redshift-space power spectrum in linear theory is related to the real-space power spectrum $P^r(k)$ by

$$P^s(k) = P^r(k)\left(1+\beta\mu^2\right)^2. \tag{7.3}$$

Equation (7.3) contains no powers of μ higher than 4, thus the angular moments above $l = 4$ will be zero. Also, since equation (7.3) contains no odd powers of μ, all odd moments will be zero. A linear theory estimator for β, which is independent of the value of the real-space power spectrum, can thus be formed by computing the ratio of the quadrupole (Q) to the monopole (M) moment (ratios involving the hexadecapole are generally too noisy to be of use),

$$\frac{P_Q^s(k)}{P_M^s(k)} = \frac{\frac{4}{3}\beta+\frac{4}{7}\beta^2}{1+\frac{2}{3}\beta+\frac{1}{5}\beta^2}. \tag{7.4}$$

Both of the methods described so far were used in the analysis of the Durham/UKST galaxy survey (Ratcliffe 1996). Ratcliffe *et al.* (1996) use an inversion of the projected correlation function to obtain a measure of the real-space correlation function, and compare this with the redshift-space correlation function to obtain $\beta = 0.48 \pm 0.18$. Ratcliffe *et al.* (1996) also use the measured anisotropies in the redshift-space correlation function to obtain $\beta = 0.55 \pm 0.12$.

The analysis of the anisotropies in the redshift-space correlation function or power spectrum can be modified to include the non-linear effects of a small scale peculiar velocity field by assuming a functional form for the velocity distribution function. A Gaussian or an exponential form is generally used. Using an exponential distribution, Peacock & Dodds (1994) showed that the redshift-space power spectrum becomes

$$P^s(k,\mu) = P^r(k)\left(1+\beta\mu^2\right)^2 exp\left(-k^2\sigma_v^2\mu^2\right), \tag{7.5}$$

which can then be inserted into equation (7.2) to re-calculate the ratio of the quadrupole to monopole moment (see Cole, Fisher & Weinberg 1995). This technique has been applied by to the IRAS $1.2Jy$ and QDOT surveys by Cole, Fisher & Weinberg (1994), who also describe a practical method for the measurement of the power spectrum as a function of k and μ. They also derive an empirical correction for the effects of the finite angle subtended, by the survey, at the position of the observer. From their analysis, $\beta = 0.52 \pm 0.13$ is measured from the $1.2Jy$ survey, and $\beta = 0.54 \pm 0.3$ from the QDOT survey. The same process can be applied to the correlation function (see e.g. Hamilton 1992, Hamilton 1993a, Fisher *et al.* 1994, Hamilton 1995).

Of course the success of the methods described above depends on whether or not the model for the small scale peculiar velocity field is a good description of reality. Also, there are several factors which cause the breakdown of linear theory, the small scale incoherent velocity field being just one of them. As discussed in Chapter 1 (Section 1.5.2), Fisher & Nusser (1996) examined the redshift-space power spectrum predicted using the Zeldovich approximation (Zeldovich 1970). They drew attention to the fact that it is the *coherent* peculiar velocity field that causes the decline in the ratio of the power spectrum quadrupole to monopole moment, not the small scale incoherent peculiar velocities (these are not included in their model). Thus the situation as regards redshift-space distortions in the quasi-linear regime and the modelling thereof is not straightforward.

Another approach to the measurement of redshift-space distortions is the use of a spherical harmonic expansion. The galaxy density field from a redshift survey is expanded in orthogonal radial and angular functions, these being spherical Bessel functions and spherical harmonics. The advantage of using this method to examine the redshift-space distortions in a survey is that the peculiar velocities affect only the radial functions. Using the Fourier approach is less straightforward as we would expect there to be mixing of modes (Zaroubi & Hoffman 1996). Spherical harmonic decomposition of the density field was first suggested by Peebles (1973) as a way to account for the curved nature of the celestial sphere. However, the elegance of the method was largely wasted on the existing galaxy surveys at that time as they had very poor sky coverage, and an extensive sky-mask makes a spherical harmonic analyses more cumbersome. It has been the advent of the IRAS near all-sky surveys that has provided the impetus to re-examine the spherical harmonic approach. Infrared (IR) radiation is less affected by galactic obscuration, thus IR surveys can cover $\gtrsim 80\%$ of the sky. Recent analyses utilizing spherical harmonic methods include Regos & Szalay (1989), Fabbri & Natale (1990), Scharf *et al.* (1992), Scharf & Lahav (1993), Fisher, Scharf & Lahav (1994), Fisher *et al.* (1995b) and Heavens & Taylor (1995). In this chapter I use spherical harmonic methods to measure the redshift-space distortions in the IRAS PSCz redshift survey (described in Chapter 5).

This chapter describes work in progress as a collaboration with Dr. A. Heavens, Dr. A. Taylor and Mr. W. Ballinger at the Royal Observatory, Edinburgh, and Professor G. Efstathiou at Oxford. The approach used is that described by Heavens & Taylor (1995), and the reader is referred to that paper for details. In Section 7.1 I briefly outline the theory behind the technique used to measure β. These methods are tested on simulated PSCz surveys drawn from large N-body simulations in Section 7.2. In Section 7.3 the analysis is applied to the PSCz redshift survey. Future improvements to the analysis are described in the relevant sections. I summarize in Section 7.4.

7.1 Methods

The methods used in this chapter are exactly those of Heavens & Taylor (1995) with minor improvements. The density field of the galaxy distribution is expanded in terms of spherical harmonics Y_{lm} and spherical Bessel functions j_l,

$$\hat{\rho}_{lmn} = c_{ln} \int w(s)\rho(s) j_l(k_{ln}s)\, Y^{\star}_{lm}(\theta,\phi)\, d^3s, \tag{7.6}$$

where $w(s)$ is an adjustable weighting function. On applying the continuity equation $\rho(s)d^3s = \rho(r)d^3r$, equation (7.6) becomes

$$\hat{\rho}_{lmn} = c_{ln} \int w(s)\rho(r) j_l(k_{ln}s)\, Y^{\star}_{lm}(\theta,\phi)\, d^3r. \tag{7.7}$$

The c_{ln} are normalization constants which are calculated in Appendix C of Heavens & Taylor (1995) via the application of orthogonality relations for spherical Bessel functions, for given boundary conditions. The boundary conditions used for this analysis are Neumann boundary conditions (i.e. that the first derivative of the potential field normal to the boundary is zero). This implies that the velocity field is zero on the boundary i.e. the boundary is undistorted. We therefore need include no surface effects. The boundary condition gives us the wave numbers k_{ln}:

$$\frac{d}{dr} j_l(k_{ln}r)\,|_{r_{max}} = 0, \tag{7.8}$$

and the normalization constants

$$c_{ln} = \frac{2k_{ln}^{\frac{3}{2}}}{\sqrt{\pi\left(\left[\frac{1}{4} + k_{ln}^2 r_{max}^2 - \left(l+\frac{1}{2}\right)^2\right] J^2_{l+\frac{1}{2}}(k_{ln}r_{max})\right)}}. \tag{7.9}$$

Heavens & Taylor (1995) then expand $w(s)j_l(k_{ln}s)$ in equation (7.7) to first order in $s - r$,

$$w(s)j_l(k_{ln}s) \simeq w(r)j_l(k_{ln}r) + u(\mathbf{r})\frac{d}{dr}\left(w(r)j_l(k_{ln}r)\right), \tag{7.10}$$

where $u(r)$ is the radial component of the peculiar velocity field, $v(r)$. Linear theory provides a relation between the peculiar velocity field $v(r)$ and the gradient of the potential field Ψ,

$$v(r) = -\nabla\Psi \tag{7.11}$$

where

$$\nabla^2\Psi \simeq H_0\Omega_0^{0.6}\delta_\rho. \tag{7.12}$$

In Appendix A of Heavens & Taylor (1995) it is shown that the radial component of this field is

$$u(\mathbf{r}) = -\hat{\mathbf{r}}\cdot\nabla\left(\frac{\Omega^{0.6}}{\nabla^2}\delta_\rho(\mathbf{r})\right) = \Omega_0^{0.6}\sum_{lmn} c_{ln}k_{ln}^{-1}\delta_{lmn}j_l'(k_{ln}r)Y_{lm}\left(\theta,\phi\right), \tag{7.13}$$

where $j'(z) = dj_l/dz$ and δ_{lmn} is the transform of the galaxy number overdensity field $\delta(r)$. Substituting equation (7.13) into equation (7.10) gives

$$w(s)j_l(k_{ln}s) \simeq w(r)j_l(k_{ln}r) + \beta\sum_{l'm'n'} c_{l'n'}\frac{\delta_{l'm'n'}}{k_{l'n'}}j_{l'}'(k_{l'n'}r)\frac{d}{dr}\left(w(r)j_l(k_{ln}r)\right)Y_{l'm'}. \tag{7.14}$$

We also know

$$\rho(r) = \rho_0(r)\left(1+\delta(r)\right) = \rho_0(r)\left(1+\sum_{lmn} c_{ln}\delta_{lmn}j_l\left(k_{ln}r\right)Y_{lm}\left(\theta,\phi\right)\right). \tag{7.15}$$

Equations (7.14) and (7.15) are then substituted into equation (7.7) and the resulting expression, to first order in δ, is

$$\hat{\rho}_{lmn} = (\rho_0)_{lmn} + \sum_{l'm'n'}\left(\Phi_{ll'nn'}^{mm'} + \beta V_{ll'nn'}^{mm'}\right)\delta_{l'm'n'}, \tag{7.16}$$

where

$$(\rho_0)_{lmn} = c_{ln}\int \rho_0(r)w(r)j_l\left(k_{ln}r\right)Y_{lm}^\star\left(\theta,\phi\right)d^3r, \tag{7.17}$$

$$\Phi_{ll'nn'}^{mm'} = c_{ln}c_{l'n'}\int \rho_0(r)w(r)j_l\left(k_{ln}r\right)Y_{lm}^\star\left(\theta,\phi\right)j_{l'}\left(k_{l'n'}r\right)Y_{l'm'}\left(\theta,\phi\right)d^3r \tag{7.18}$$

and

$$V_{ll'nn'}^{mm'} = \frac{c_{ln}c_{l'n'}}{k_{l'n'}^2}\int \rho_0(r)\frac{d}{dr}\left[w(r)j_l\left(k_{ln}r\right)\right]Y_{lm}^\star\left(\theta,\phi\right)j_{l'}'\left(k_{l'n'}r\right)Y_{l'm'}\left(\theta,\phi\right)d^3r. \tag{7.19}$$

Assigning $D_\mu = (\hat{\rho}_{lmn} - (\hat{\rho}_0)_{lmn})/\overline{\rho}$, equation (7.16) can be written in a compact form as

$$D_\mu = \sum_\nu \Lambda_\mu^\nu\delta_\nu \tag{7.20}$$

with

$$\Lambda_\mu^\nu = \Phi_\mu^\nu + \beta V_\mu^\nu. \tag{7.21}$$

The quantity $\rho_0(r)$ may now be separated into angular and radial components

$$\rho_0(r) = M(\theta, \phi)\, \rho_0(r) \tag{7.22}$$

with $M(\theta, \phi)$ being the sky-mask of the survey and $\rho_0(r)$ containing the radial selection function. The 'transition matrices' Φ and V now become

$$\Phi^{mm'}_{ll'nn'} = W^{mm'}_{ll'} \Phi^{nn'}_{ll'} \tag{7.23}$$

$$V^{mm'}_{ll'nn'} = W^{mm'}_{ll'} V^{nn'}_{ll'} \tag{7.24}$$

where

$$\Phi^{nn'}_{ll'} = c_{ln} c_{l'n'} \int \rho_0(r) w(r) j_{l'}(k_{l'n'} r)\, j_l(k_{ln} r)\, r^2 dr, \tag{7.25}$$

$$V^{ll'}_{nn'} = \frac{c_{ln} c_{l'n'}}{k^2_{l'n'}} \int \rho_0(r) \frac{d}{dr} \left[w(r) j_l(k_{ln} r) \right] \frac{d}{dr} j_{l'}(k_{l'n'} r)\, r^2 dr \tag{7.26}$$

and

$$W^{mm'}_{ll'} = \int_\Omega Y_{l'm'} M(\Omega) Y^*_{lm} d\Omega. \tag{7.27}$$

Heavens & Taylor (1995) assumed that the sky-mask was azimuthally symmetric; this makes the mask mixing matrices, W, real. We now drop this assumption as the real sky-mask for the PSCz survey is not symmetric (see Figure 5.1 in Chapter 5). In the case of the simulated surveys (Section 7.2) we do not mask the surveys, so the formalism is exactly that given in Heavens & Taylor (1995). In the case of an azimuthally asymmetric sky-mask, the covariance matrix $C =< D_\mu D_\nu >$ changes slightly from that in Heavens & Taylor (1995). There are four terms to the covariance matrix C,

$$\begin{aligned} C_{\Re\Re} &= \langle \Re(D_\mu) \Re(D_\nu) \rangle, \\ C_{\Re\Im} &= \langle \Re(D_\mu) \Im(D_\nu) \rangle, \\ C_{\Im\Re} &= \langle \Im(D_\mu) \Re(D_\nu) \rangle, \\ C_{\Im\Im} &= \langle \Im(D_\mu) \Im(D_\nu) \rangle, \end{aligned}$$

where $\Re$ signifies the real part and $\Im$ the imaginary part. These terms are given by the following summations:

$$\begin{aligned} C_{\Re\Re} &= \frac{1}{2} \sum_\alpha \left[\Re(\Phi^\alpha_\mu + \beta V^\alpha_\mu) \Re(\Phi^\alpha_\nu + \beta V^\alpha_\nu) + \Im(\Phi^\alpha_\mu + \beta V^\alpha_\mu) \Im(\Phi^\alpha_\nu + \beta V^\alpha_\nu) \right] P(k_\alpha) \\ &\quad + \Lambda^0_{\mu\nu}, \end{aligned} \tag{7.28}$$

$$C_{\Im\Im} = C_{\Re\Re}, \tag{7.29}$$

$$C_{\Re\Im} = \frac{1}{2}\sum_\alpha \left[\Re(\Phi_\mu^\alpha + \beta V_\mu^\alpha)\Im(\Phi_\nu^\alpha + \beta V_\nu^\alpha) - \Im(\Phi_\mu^\alpha + \beta V_\mu^\alpha)\Re(\Phi_\nu^\alpha + \beta V_\nu^\alpha)\right] P(k_\alpha) + \Lambda_{\mu\nu}^0, \tag{7.30}$$

$$C_{\Im\Re} = -C_{\Re\Im}. \tag{7.31}$$

The shot noise is given by the equation

$$\Lambda_{\mu\nu}^0 = c_{ln}c_{l'n'}\int \rho_0(r)w^2(r)j_l(k_{ln}r)j_{l'}(k_{l'n'}r)r^2dr\int \mathcal{P}_\mu(Y_{lm}(\Omega))M(\Omega)\mathcal{P}_\nu(Y_{l'm'}^*(\Omega))\,d\Omega \tag{7.32}$$

as in Heavens & Taylor (1995), where the $\mathcal{P}_{\mu,\nu}$ represent the real or imaginary parts of the spherical harmonics according to whether we are using the real or imaginary parts of $D_{\mu,\nu}$. Now we have the covariance matrix, a likelihood functional is constructed and we use an assumed real-space power spectrum $P(k)$ in equations (7.28) to (7.31), and maximise the likelihood for the value of β and the amplitude of the power spectrum. The amplitude of the power spectrum is defined in terms of $\Delta(k=0.1)$ where $\Delta^2(k) = k^3P(k)/2\pi^2$. The likelihood function is

$$\mathcal{L}[D|\beta, P(k_{ln})] = (det(C_{\Re\Re}))^{-\frac{1}{2}}(det(C_{\Re\Im}))^{-\frac{1}{2}}(det(C_{\Im\Re}))^{-\frac{1}{2}}(det(C_{\Im\Im}))^{-\frac{1}{2}} \tag{7.33}$$

$$\times exp\left(-\frac{1}{2}\sum_{lmn}\Re(D_{lmn})C_{\Re\Re}\Re(D_{lmn})\right) \tag{7.34}$$

$$\times exp\left(-\frac{1}{2}\sum_{lmn}\Im(D_{lmn})C_{\Im\Re}\Re(D_{lmn})\right) \tag{7.35}$$

$$\times exp\left(-\frac{1}{2}\sum_{lmn}\Re(D_{lmn})C_{\Re\Im}\Im(D_{lmn})\right) \tag{7.36}$$

$$\times exp\left(-\frac{1}{2}\sum_{lmn}\Im(D_{lmn})C_{\Im\Im}\Im(D_{lmn})\right). \tag{7.37}$$

There remains the question of what to use as the weighting function $w(r)$. The minimum variance weighting function has been derived by Heavens & Taylor (1995); in fact it is the same function as that derived by Feldman, Kaiser & Peacock (1994) for flux limited power spectrum analysis, i.e.

$$w(r) = \frac{1}{1+\overline{n}(r)P_w(k)}. \tag{7.38}$$

For the analysis in this chapter we have modified the Heavens & Taylor (1995) analysis by including the weight function in its full k-dependent form, instead of assuming a constant value for $P_w(k)$. We could in principle maximise the likelihood function for both β and the real-space power spectrum $P(k)$, and indeed this has been implemented by Ballinger, Heavens & Taylor (1995) for the $1.2Jy$ redshift survey; this analysis will be applied to the PSCz redshift data in the course of future work.

7.2 Tests on simulated surveys

7.2.1 Construction of simulated PSCz surveys

In this section I have applied the methods described above to simulated PSCz redshift surveys drawn from large N-body simulations. The simulations are the same as those described in previous chapters (see e.g. Chapter 2) except they are in a box of comoving size 600 h^{-1}Mpc (instead of 300 h^{-1}Mpc previously) and have 160^3 particles (compared to 100^3 before). For this test I use two of the three available cosmological models to simulate the surveys. The first model is the low density CDM universe (LCDM), with $\Omega_{matter} = 0.2$ and $\Gamma = \Omega h = 0.2$. The contribution to Ω in the form of a cosmological constant, Ω_Λ, is 0.8, therefore this is a flat universe. The simulated surveys are drawn from the simulation at the output time when σ_8, the *rms* mass fluctuations in spheres of radius 8 h^{-1}Mpc, is unity. The second model is the standard CDM model (SCDM), with $\Omega = 1.0$ and $\Gamma = 0.5$. Again, the output time used is that for which $\sigma_8 = 1.0$. This model has very high small scale peculiar velocities, the three-dimensional velocity dispersion of particles is 1160km s^{-1}.

The simulated surveys are all-sky (i.e. no mask is imposed) and extend to a radial distance of 300 h^{-1}Mpc. The simulations are thus larger enough that no replication of the box is required in order to simulate the PSCz redshift survey. These mock surveys are flux limited, using a selection function:

$$\phi(\Delta) = \phi_\star \frac{10^{(1-\alpha)\Delta}}{(1+10^{\gamma\Delta})^{\frac{\beta}{\gamma}}}, \tag{7.39}$$

where $\Delta = log_{10}(r/r_0)$ and the parameters $\phi_\star$, α, r_0, γ and β are those estimated by Mann, Saunders & Taylor (1996) to be appropriate to the QDOT survey. The values of these parameters are $\phi_\star = 0.0132(\ h^{-1}\mathrm{Mpc})^{-3}$, $r_0 = 1.81\ h^{-1}$Mpc, $\alpha = 1.61$, $\beta = 3.90$ and $\gamma = 1.64$. I have ten realizations of the PSCz redshift survey drawn from the LCDM simulations and nine realizations drawn from the SCDM simulations. There has been no attempt made to select 'galaxies' from the simulations, mass points have been chosen to represent galaxies, i.e. the mock surveys have a linear bias parameter of unity. The number of points representing galaxies that are used in the analysis of the simulated PSCz surveys is $\sim$ 17700.

7.2.2 Results from analysis of simulated PSCZ surveys

Figure 7.1 shows likelihood contours for the values of β and $\Delta(k = 0.1)$ for the nine SCDM mock PSCz surveys. No correction has been applied for the effect of small scale peculiar velocities, but only modes with wave number $0.01 < k < 0.073$ h Mpc^{-1} are used in the analysis. This range corresponds to the modes with $2 < l < 19$ and $1 < n < 7$. For the matrices where there is mixing of modes (the Φ, V, W and S matrices), $2 < l < 19$ and $1 < n < 20$ were included. Because the simulated surveys are all-sky, there is no mixing between modes of different l.

In all of the following figures the likelihood contours for β and $\Delta(k = 0.1)$ are plotted in steps of $\Delta(ln\mathcal{L}) = 0.5$. The cross shows the maximum likelihood value of β and $\Delta(k = 0.1)$, and the dot shows the true value of these parameters. The value of β in the case of the simulated surveys is just $\Omega^{0.6}$ since $b = 1$. Taking the nine maximum likelihood values of β and averaging, we get $\beta_{av}^{SCDM} = 0.73 \pm 0.13$, and for the amplitude, $\Delta_{av}^{SCDM} = 0.55 \pm 0.02$, where I have stated the 1σ error on the mean value. The true values for β and $\Delta(k = 0.1)$ are 1.0 and 0.48 respectively. Thus for the SCDM model the average value of β recovered is low; the recovered value is about 2σ from the true value. That we underestimate the value of β is to be expected due to the very high small scale peculiar velocities, the effects of which extend to large scales. This can be seen in Figure 2.13 of Chapter 2, where the linear theory result is not reached except on very large scales, in fact too large to be measured by a PSCz redshift survey. The average value of $\Delta(k = 0.1)$ over the nine realizations is more than 3σ away from the true value. The reason for this discrepancy is not clear. It may be due to the fact that the model used for the real-space power spectrum $P(k)$ (a linear theory SCDM power spectrum) is not accurate at wave numbers even as small as $k = 0.1$ h Mpc^{-1} as there has been considerable non-linear evolution of the power spectrum.

The situation is more encouraging when we examine the LCDM model. The results of the analysis on the ten realizations of the simulated all-sky PSCz redshift surveys are shown in Figure 7.2. We get $\beta_{av}^{LCDM} = 0.37 \pm 0.06$ and $\Delta_{av}^{LCDM} = 0.63 \pm 0.02$, to be compared with the true values of $\beta = 0.38$ and $\Delta(k = 0.1) = 0.62$. For the LCDM model we therefore recover the correct values of the parameters to within the error bars.

Heavens & Taylor (1995) describe an approximate correction for the effect of the small scale non-linear peculiar velocity field (Appendix B of Heavens & Taylor 1995). It is assumed that the real-space position of a galaxy is perturbed slightly by a peculiar velocity such that the redshift-space position becomes

$$s(r) \rightarrow s'(r) = s(r) + \epsilon(r). \tag{7.40}$$

The quantity $\epsilon(r)$ is a random variable drawn from a Maxwellian distribution. This distribution is decomposed into radial and angular parts, and the effect of the angular part is allowed to asymptote to zero as the small scale peculiar velocities provide only a radial distortion. Heavens & Taylor (1995) derive a scattering matrix S

$$S^{mm'}_{ll'nn'} = \frac{c_{ln}c_{l'n'}\delta^K_{ll'}\delta^K_{mm'}}{V\pi} \times \int\int \frac{e^{-\frac{(r-y)^2}{(2\sigma)^2}}}{\sqrt{2\pi}\sigma} j_l(k_{ln}r)\, j_{l'}(k_{l'n'}y)\; r dr\, y dy, \tag{7.41}$$

where σ is the small scale three-dimensional velocity dispersion and the quantity D_{lmn} now becomes

$$D_{lmn} = S^{mm''}_{ll''nn''} W^{mm'}_{l''l'} \left(\Phi^{n''n'}_{l''l'} + \beta V^{n''n'}_{l''l'}\right) \delta_{l'm'n'}. \tag{7.42}$$

From the SCDM simulations the measured small scale velocity dispersion of particles, σ, is 1160km s^{-1}. Using this value of σ and applying the correction we obtain the results shown in Figure 7.3. Averaging the results over the nine realizations, we now recover $\beta^{SCDM}_{av} = 0.83 \pm 0.18$ and $\Delta^{SCDM}_{av} = 0.61 \pm 0.03$. The value of β has thus been restored to the correct value to within the 1σ error bar, but we have exacerbated the problem with the overestimation of the amplitude of the power spectrum.

For the LCDM model the three-dimensional velocity dispersion of particles, σ, is 587km s^{-1}. Figure 7.4 shows the likelihood contours for β and $\Delta(k = 0.1)$ for this model when we include the effects of the scattering matrix. Averaging the maximum likelihood results for β and $\Delta(k = 0.1)$ over the ten realizations we obtain $\beta^{LCDM}_{av} = 0.41 \pm 0.07$ and $\Delta^{LCDM}_{av} = 0.64 \pm 0.02$. We can see that for this model the results are relatively unchanged by inclusion of the effects of the small scale peculiar velocity field. This is to be expected since the non-linearities introduced by these velocity fields are expected to be very small at a wave number of 0.07 h Mpc^{-1} (see Figure 2.13 in Chapter 2).

The spherical harmonic analysis of simulated PSCz surveys performed here shows that if the real Universe has similar properties to the LCDM model, then we can be confident of recovering the true value of β. The analysis of the SCDM model was not quite so successful, especially in the recovery of the amplitude of the power at $k = 0.1$ h Mpc^{-1}. Improvements that could be made include, for example, using a more realistic power spectrum $P(k)$ in equations (7.28) to (7.31), i.e. using a non-linear power spectrum as the linear SCDM power spectrum is likely to be a bad fit to the evolved power spectrum even on scales with wave numbers $k = 0.1\ h\ \mathrm{Mpc}^{-1}$. This problem will of course be side-stepped by the use of the Ballinger, Heavens & Taylor (1995) method where the likelihood function is maximised as a function of β and $P(k)$ (the power spectrum being characterized by its value in number of discrete bins in k space). This latter method will be tested on these simulated PSCz surveys in the near future, before being applied to the real data. Another point that needs to be made is that three-dimensional small scale velocity dispersion of galaxies in the real Universe is probably not as large as that in the SCDM simulation used here. Its actual value is uncertain, but the highest values of σ measured from galaxy surveys are of the order $500 - 800\mathrm{km\,s}^{-1}$ (Hale-Sutton *et al.* 1989, Mo, Jing & Bórner 1993, Guzzo *et al.* 1995). For the real PSCz survey we adopt a value of $420\mathrm{km\,s}^{-1}$ for the three-dimensional velocity dispersion. This is appropriate to the assumption that the small scale peculiar velocities are incoherent and that the pairwise velocity dispersion of galaxies is $340\mathrm{km\,s}^{-1}$ (Huchra *et al.* 1983, Fisher *et al.* 1994). A future improvement would be to determine the appropriate value of σ from the PSCz redshift survey itself.

In summary, the tests on the simulated surveys show that we are likely to be able to recover the correct value of β using a survey the size of the PSCz redshift survey. The amplitude $\Delta(k = 0.1)$ is more uncertain and the success of its recovery is not guaranteed. The internal estimates of the error on the parameters β and $\Delta(k = 0.1)$, i.e. the 1σ errors obtained by projecting the first contour onto the coordinate axes, are consistent with the spread in the results over the nine/ten realizations.

7.3 Spherical harmonic redshift-space distortion analysis of the PSCz survey

In this section I describe the application of the methods, described and tested above, to the PSCz redshift survey. The analysis is performed on galaxies within a radial distance of 200 h^{-1}Mpc and the sky-mask used is MASK II, described in Chapter 5. This is the more restrictive of the two available sky-masks, masking 28% of the sky thus leaving 72% for analysis. The number of PSCz galaxies used in this analysis is 10186, and the selection function used is that described in Chapter 5.

For the PSCz data we do not know the correct real-space power spectrum to use in equations (7.28) to (7.31). Consequently, I have analysed the survey using the methods described above for a family of CDM-like power spectra characterized by their value of

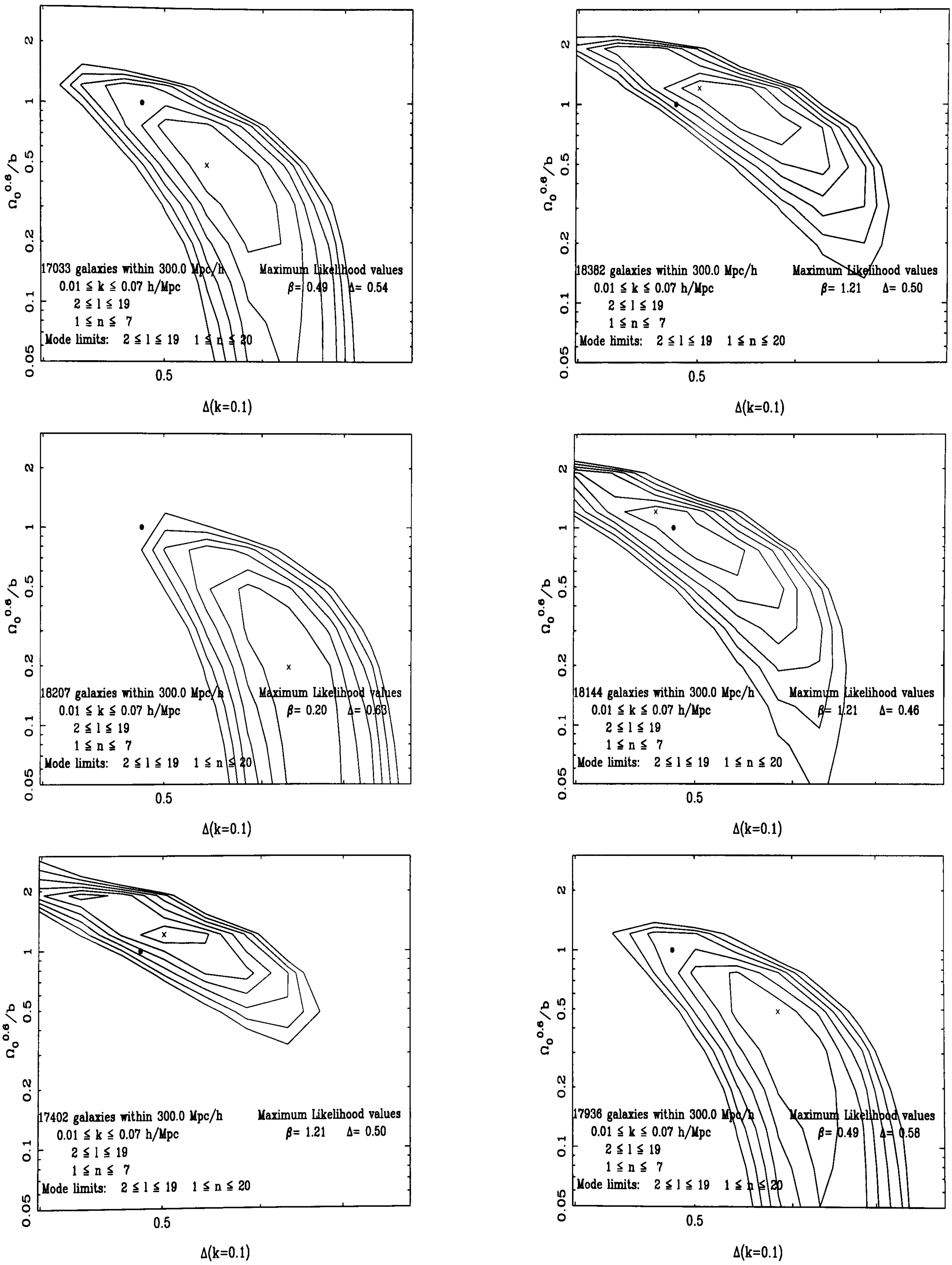


Figure 7.1: Figure shows contours of likelihood in the β $\Delta(k = 0.1)$ plane for 9 realizations of the PSCz redshift survey drawn from SCDM N-body simulations. The contours are plotted at intervals of $\delta ln\mathcal{L} = 0.5$. The x axis is plotted in steps of 0.1 in $\Delta(k = 0.1)$.

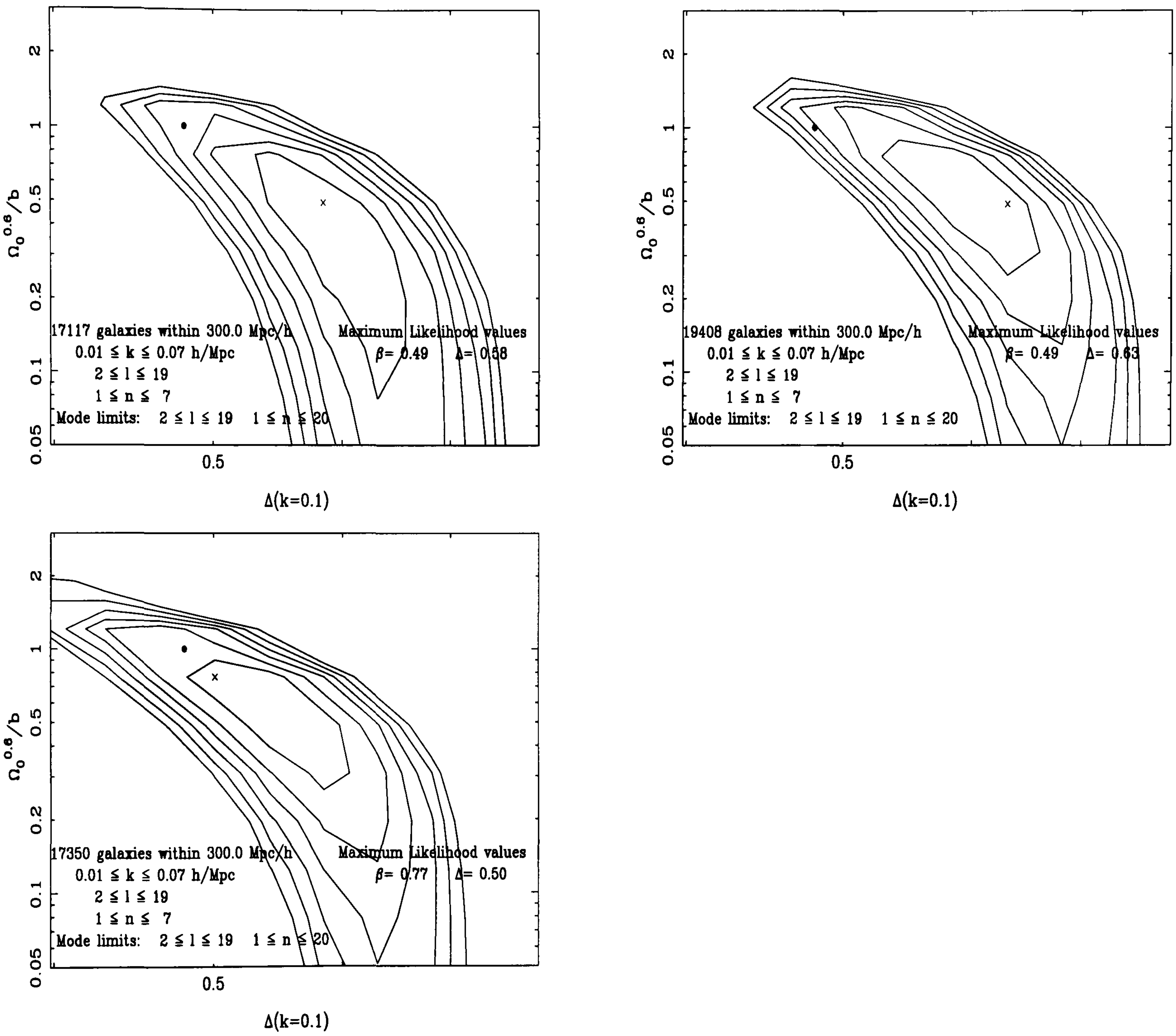


Figure 7.1: continued.

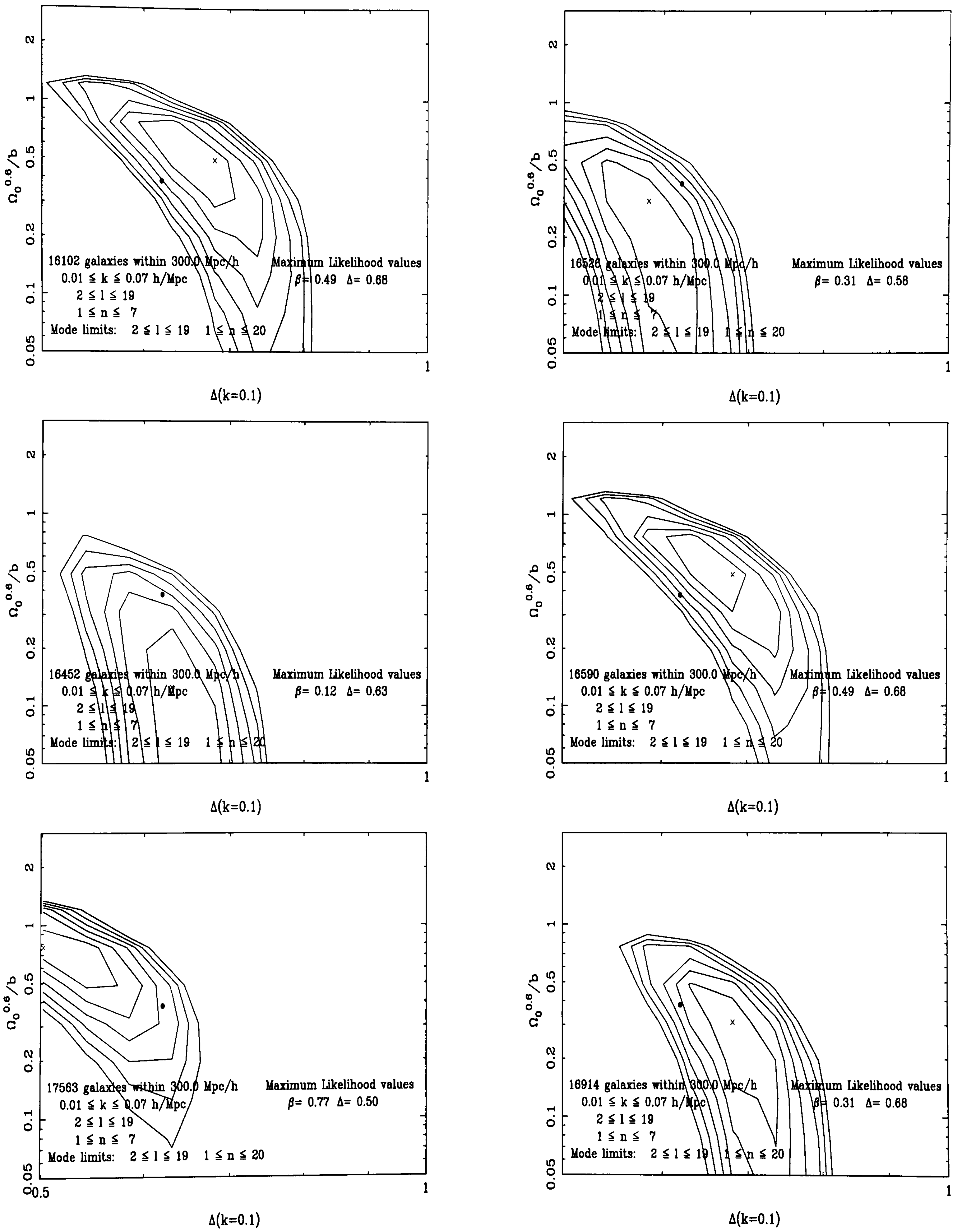


Figure 7.2: Figure shows contours of likelihood in the β $\Delta(k = 0.1)$ plane for 10 realizations of the PSCz redshift survey drawn from LCDM N-body simulations. Again, as in all the following figures in this chapter, contours are plotted at intervals of $\delta ln\mathcal{L} = 0.5$ and the x axis is plotted in steps of 0.1 in $\Delta(k = 0.1)$.

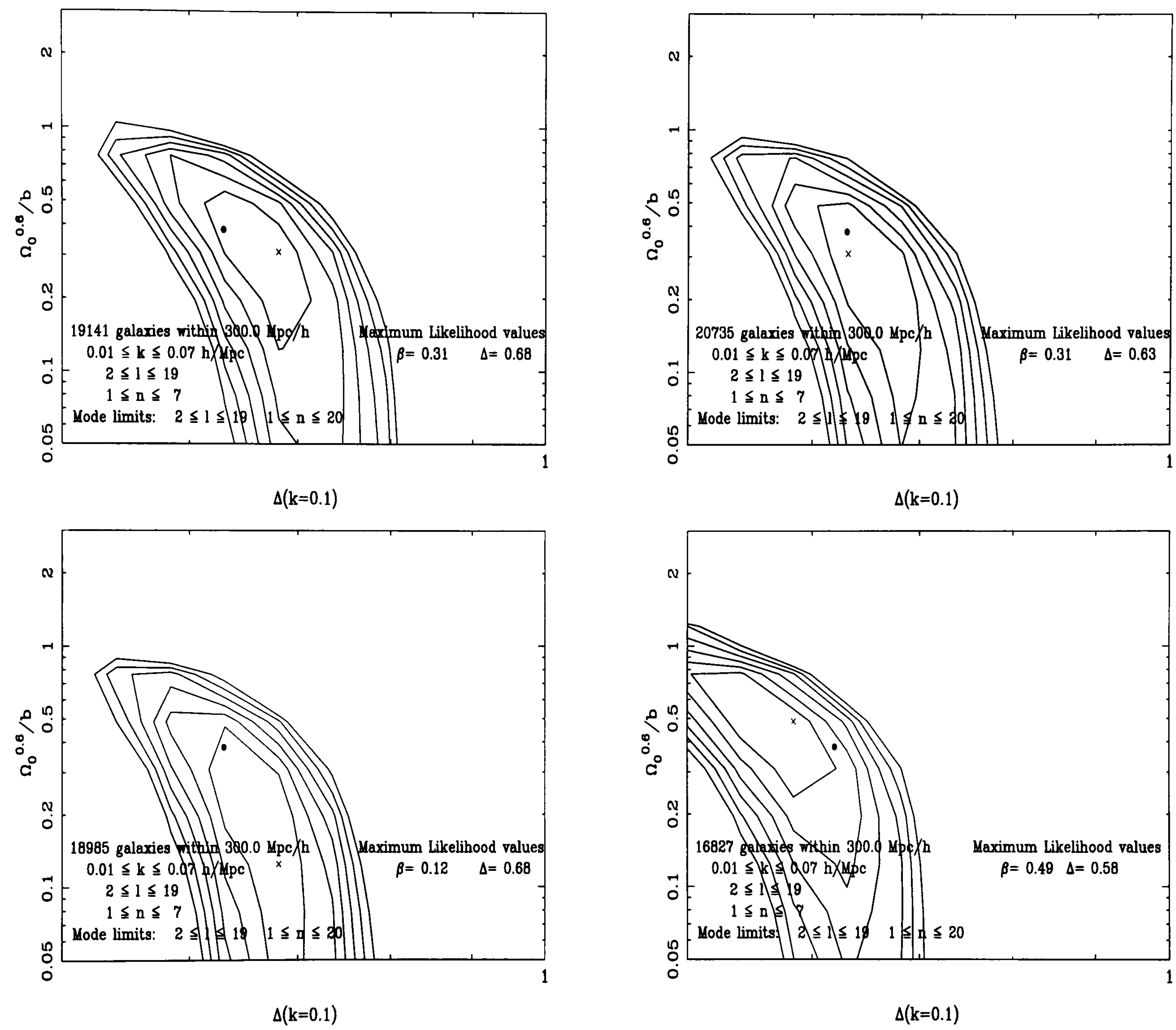


Figure 7.2: continued.

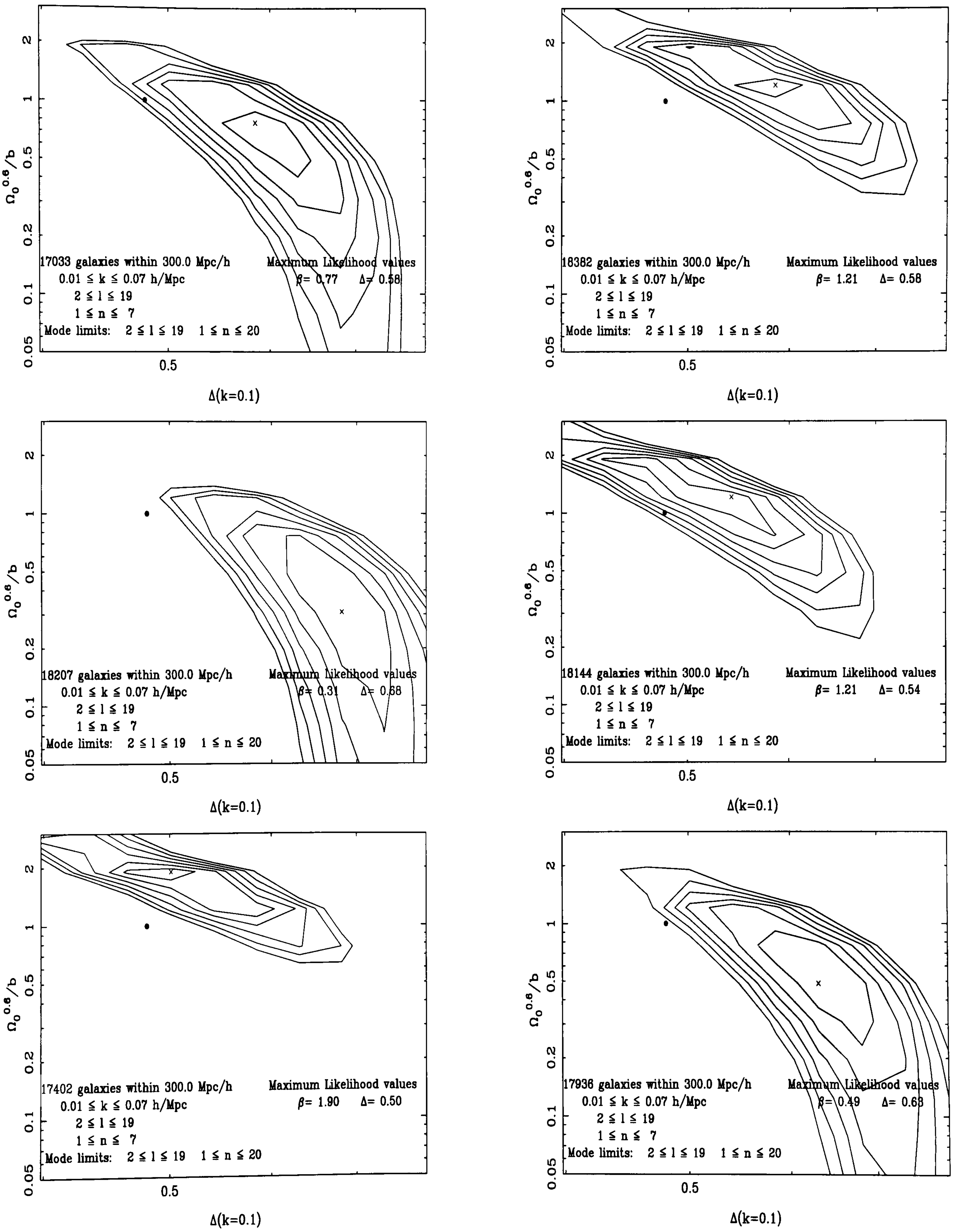


Figure 7.3: As Figure 7.1 except now the effects of the small scale velocity field have been included (see text).

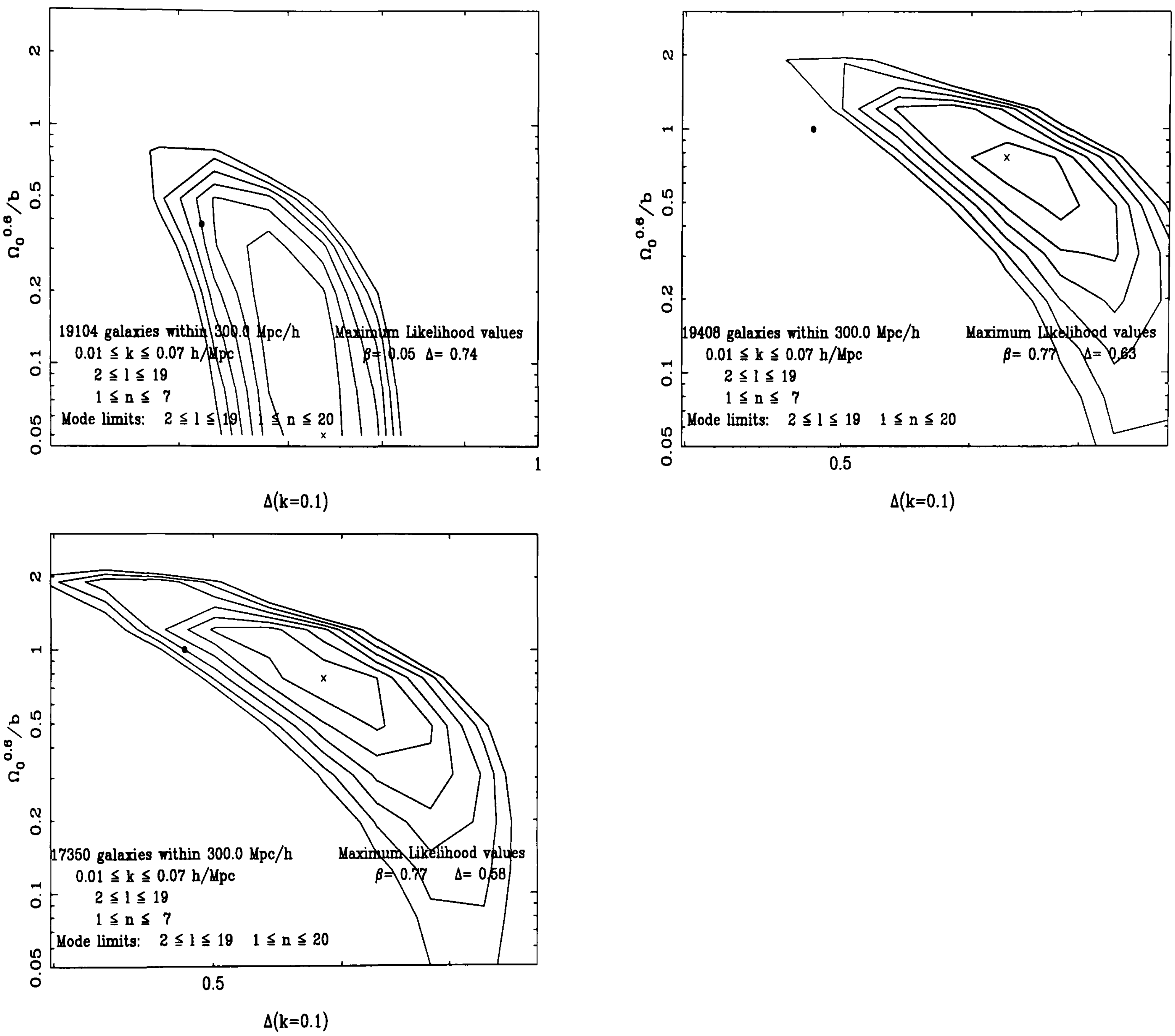


Figure 7.3: continued.

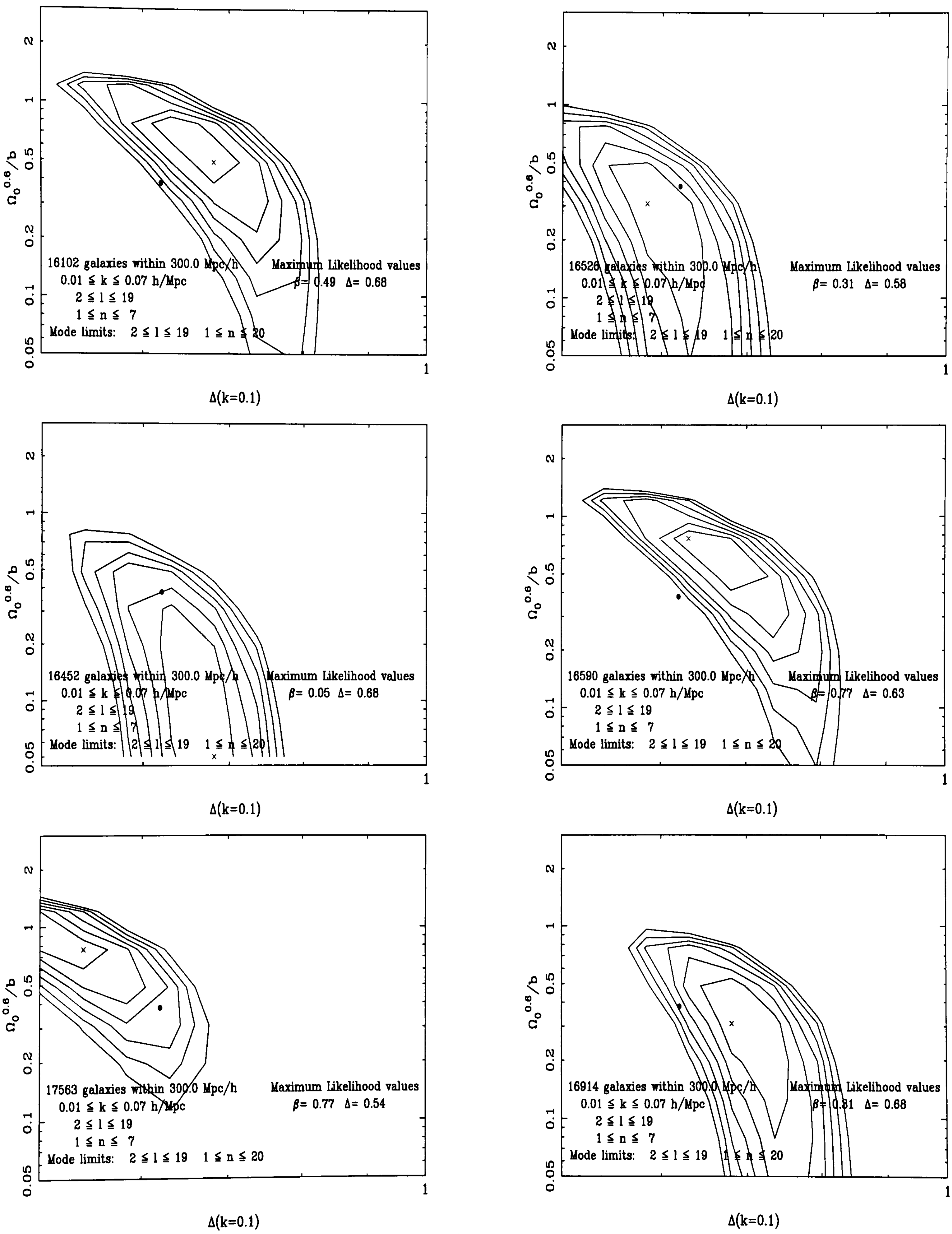


Figure 7.4: Likelihood contours for β and $\Delta(k = 0.1)$ for 10 realizations of the PSCz redshift survey drawn from the LCDM simulations. The effects of the small scale peculiar velocity field have been included.

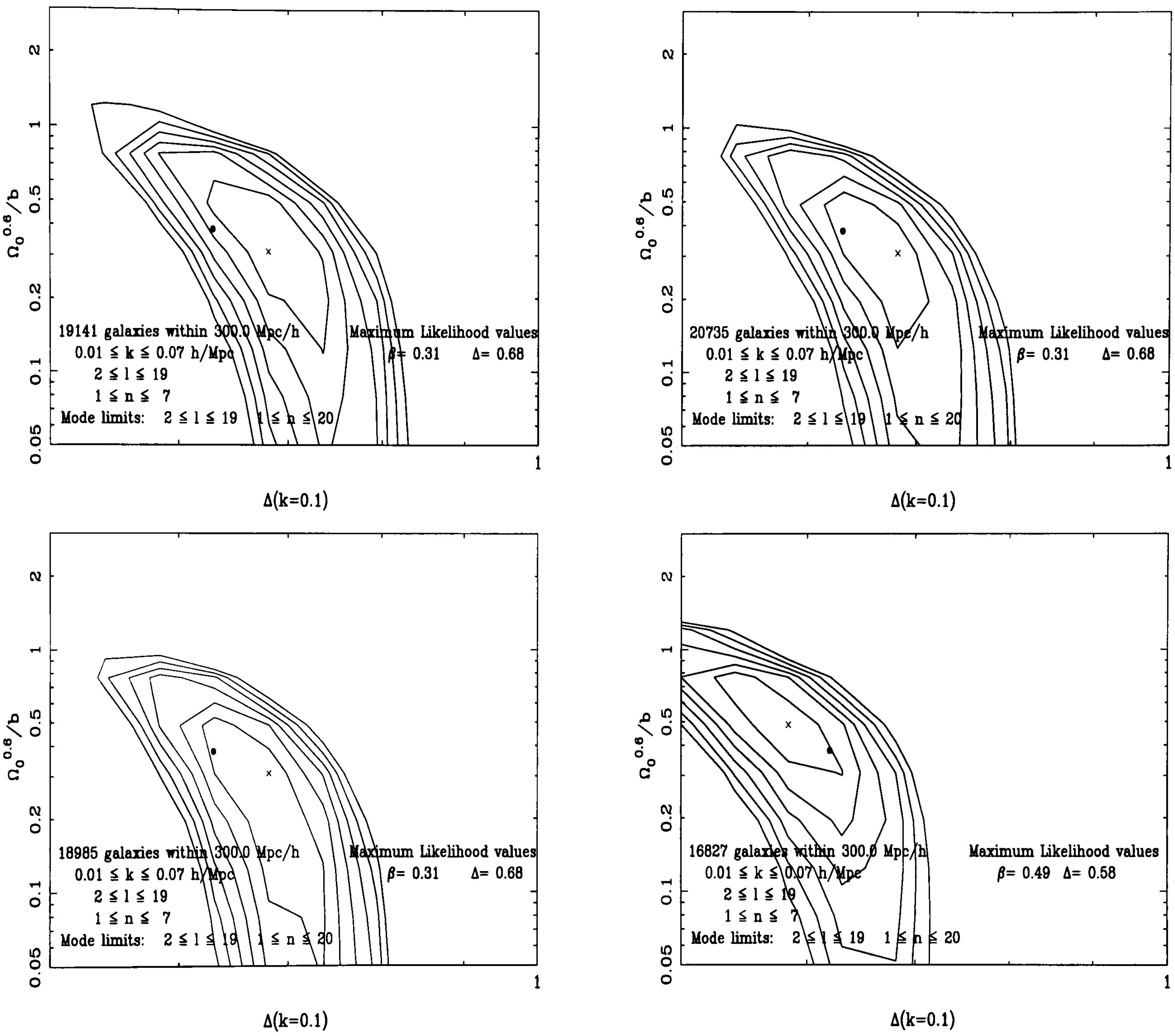


Figure 7.4: continued.

$\Gamma = \Omega h$. I have performed the analysis for values of Γ ranging from 0.1 to 0.55 in steps of 0.05 in Γ. Modes with $2 < l < 17$ and $1 < n < 6$ have been included in the analysis and for the cross terms, $0 < l < 30$ and $1 < n < 20$. The analysis extends from wave numbers $k = 0.01\ h\ \mathrm{Mpc}^{-1}$ to $k = 0.1\ h\ \mathrm{Mpc}^{-1}$. Again, a wave number dependent weighting has been used, however this is not optimised to the value of Γ used in the likelihood analysis. The weighting function is always used with a $\Gamma = 0.2$ CDM power spectrum, however it is reasonable to argue that this should not have a large effect if one is close to the minimum variance weighting.

Figure 7.5 shows likelihood contours of β and $\Delta(k = 0.1)$ for the PSCz redshift survey, without inclusion of the scattering matrix. The values of β and $\Delta(k = 0.1)$ recovered are fairly insensitive to the value of Γ used for the real-space power spectrum in the likelihood analysis. The maximum likelihood value of β obtained is 0.31 for every case, except when $\Gamma = 0.35$ in which case it is slightly higher at 0.49. The value of the amplitude $\Delta(k = 0.1)$ ranges between 0.4 and 0.46 depending on Γ. Taking all the results in Figure 7.5 together, the absolute maximum likelihood values of β and $\Delta(k = 0.1)$ occur for the $\Gamma = 0.2$ model. Hereafter I shall adopt these as the estimates of β and $\Delta(k = 0.1)$ from the PSCz survey (in the case of no correction for the small scale velocity field). In Figure 7.5 the second contour level approximately gives the joint 1σ error region for the two parameters. To obtain an estimate of the error on each of the parameters separately I have projected the contour at $\delta ln\mathcal{L} = -0.5$ onto each of the coordinate axes. The final estimate of the parameters β and $\Delta(k = 0.1)$ is:

$$\beta = 0.31^{+0.29}_{-0.20}$$
$$\Delta(k = 0.1) = 0.43^{+0.03}_{-0.04},$$

or,

$$\Omega = 0.14^{+0.29}_{-0.12}\ b^{1.67}.$$

The value $\beta = 1$ is excluded at the 2σ significance level.

Figure 7.6 shows the likelihood contours for β and $\Delta(k = 0.1)$ in the case where the effects of the small scale peculiar velocity dispersion of galaxies are included. As mentioned above, when the scattering matrix S is included in the analysis, a value $\sigma = 420\mathrm{km\ s}^{-1}$ is employed. This modification has some effect on the likelihood contours. Formally, the maximum likelihood result now occurs for $\Gamma = 0.25$ and is $\beta = 0.49$, $\Delta(k = 0.1) = 0.43$. As expected the value of β has risen slightly, while $\Delta(k = 0.1)$ is unaffected. However, $\beta = 1$ is still excluded at the level of almost 2σ. The formal result for this case is:

$$\beta = 0.49^{+0.10}_{-0.37}$$
$$\Delta(k = 0.1) = 0.43^{+0.05}_{-0.02},$$

or,

$$\Omega = 0.30^{+0.12}_{-0.27}\ b^{1.67}.$$

The results presented here can be compared with those of Heavens & Taylor (1995) who analysed the IRAS $1.2Jy$ survey and measured $\beta = 1.1 \pm 0.3$. This is high compared to the value measured here. It is 2σ away from the second result quoted above. This prompts a cautionary note. If the PSCz galaxy redshift survey is found to have the anisotropies from which the QDOT redshift survey suffered (Hamilton 1995), then we would expect to obtain a lower value of β, just as is found here. We would also expect to find a high value for the amplitude Δ, which is not the case. However, this possibility must be investigated and there are several tests which may be applied. Firstly, the analysis presented here will be repeated with the higher flux limits of $0.7Jy$ and $1.2Jy$. Effects due to flux errors may be expected to have a greater impact near the survey flux limit. Secondly, we can use the add-scanned flux (see Chapter 5) instead of the point source flux for galaxies in the PSCz redshift survey. Using the add-scanned flux, the anisotropies seen in the QDOT survey were considerably alleviated (Hamilton, private communication). Of course, as mentioned in Chapter 6, the Hamilton (1995) analysis is being repeated on the PSCz redshift survey by Oliver Keeble at Imperial College, London. Other tests to directly measure anisotropies not connected with the peculiar velocity field will also be investigated.

7.4 Summary

In this chapter I have employed the methods described by Heavens & Taylor (1995) to analyse the redshift-space distortions in the IRAS PSCz galaxy redshift survey. These methods have been tested on simulated surveys and found to be adequate to recover the correct value of the parameter β. From the analysis of the PSCz redshift survey a value of $\beta = 0.31^{+0.29}_{-0.20}$ is measured, or $\beta = 0.49^{+0.10}_{-0.37}$ if the effects of the small scale peculiar velocity field are included. The formal maximum likelihood result is obtained using $\Gamma = 0.2$ in the former case and $\Gamma = 0.25$ in the latter, although the analysis is fairly insensitive to the

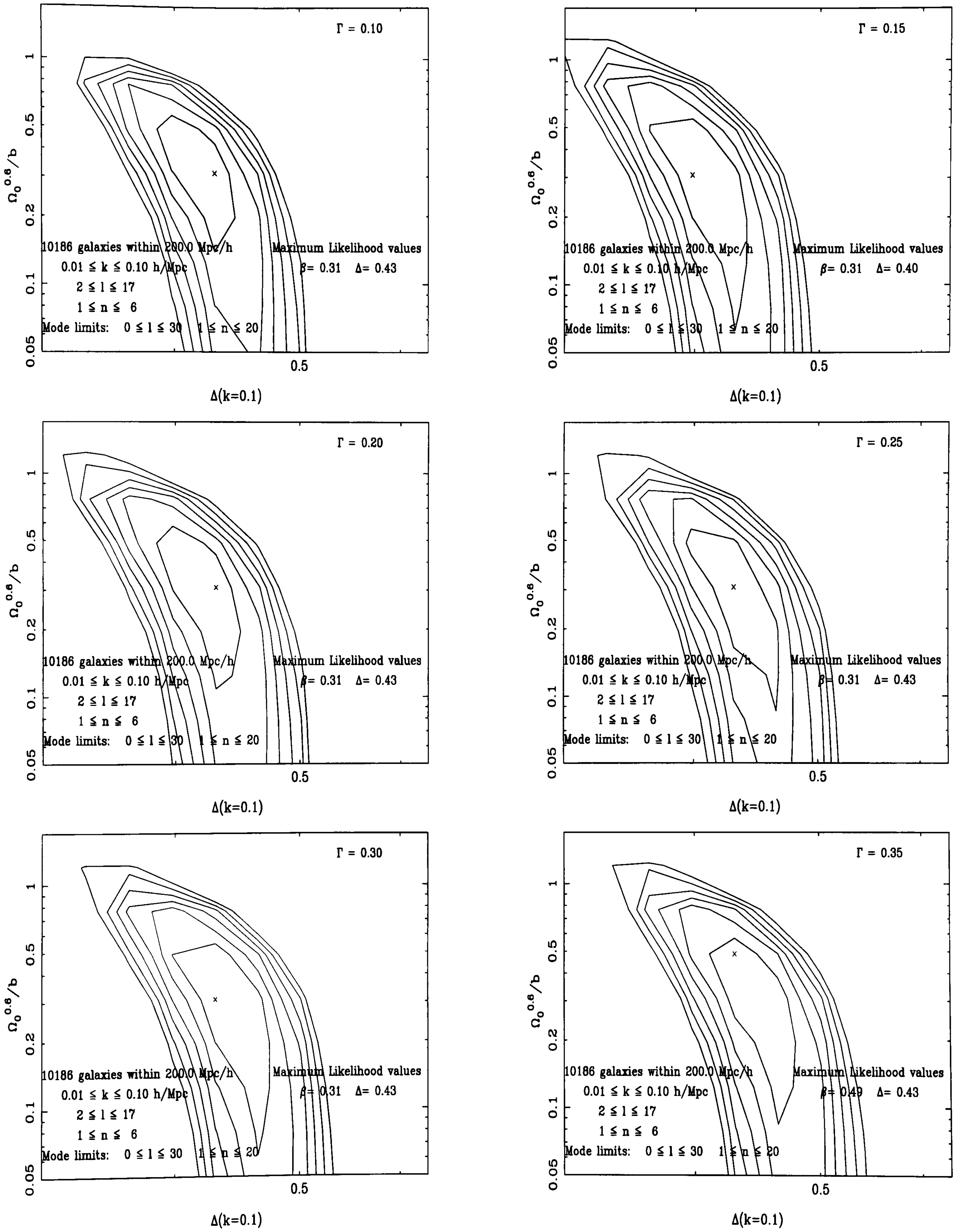


Figure 7.5: Likelihood contours for β and $\Delta(k = 0.1)$ for the PSCz redshift survey. Contours are plotted at intervals of $\delta ln\mathcal{L} = 0.5$ and the x axis is labeled at intervals of 0.1. The value of Γ used for the real-space power spectrum in the likelihood analysis is shown in the top right hand corner of each figure.

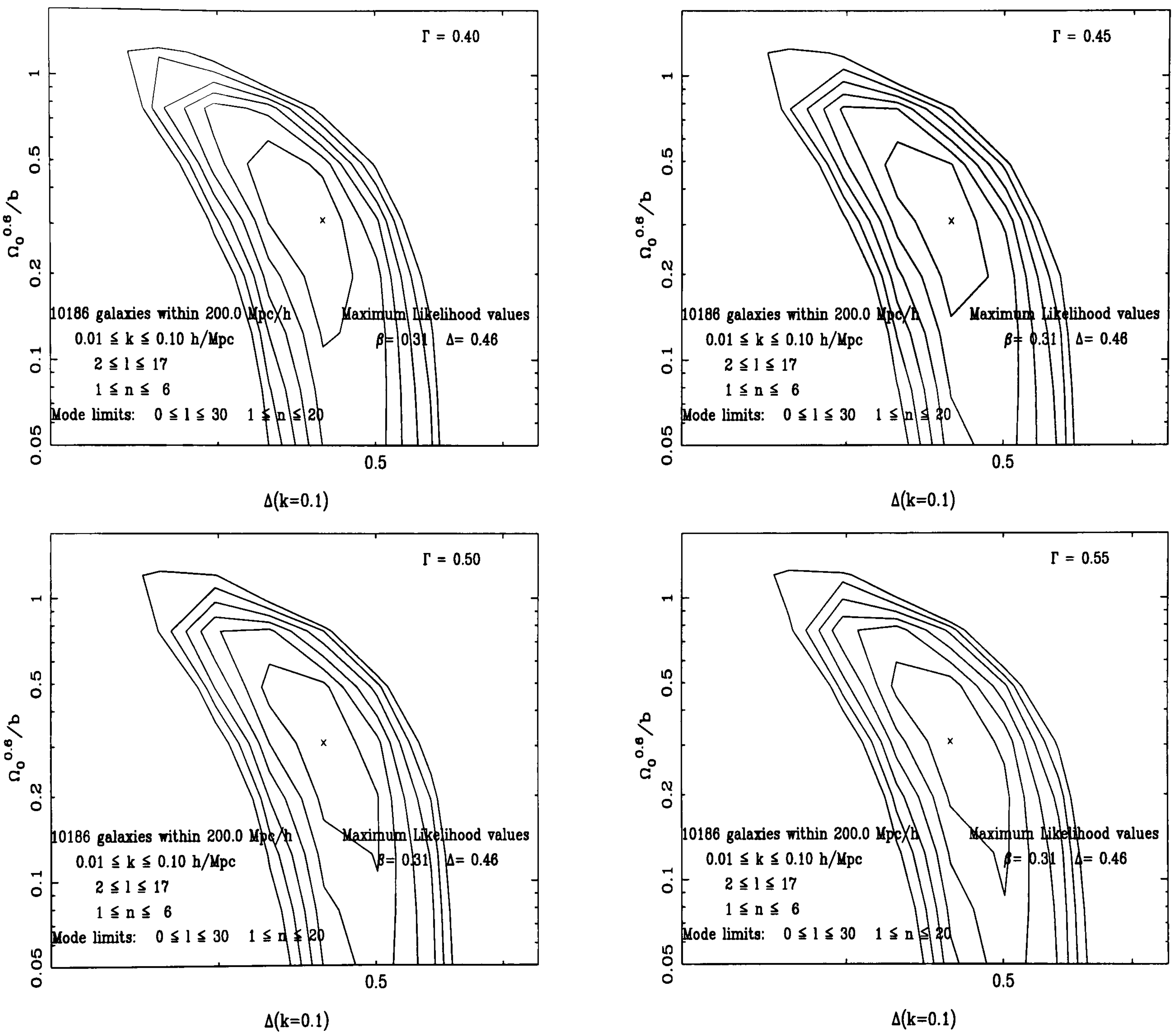


Figure 7.5: continued.

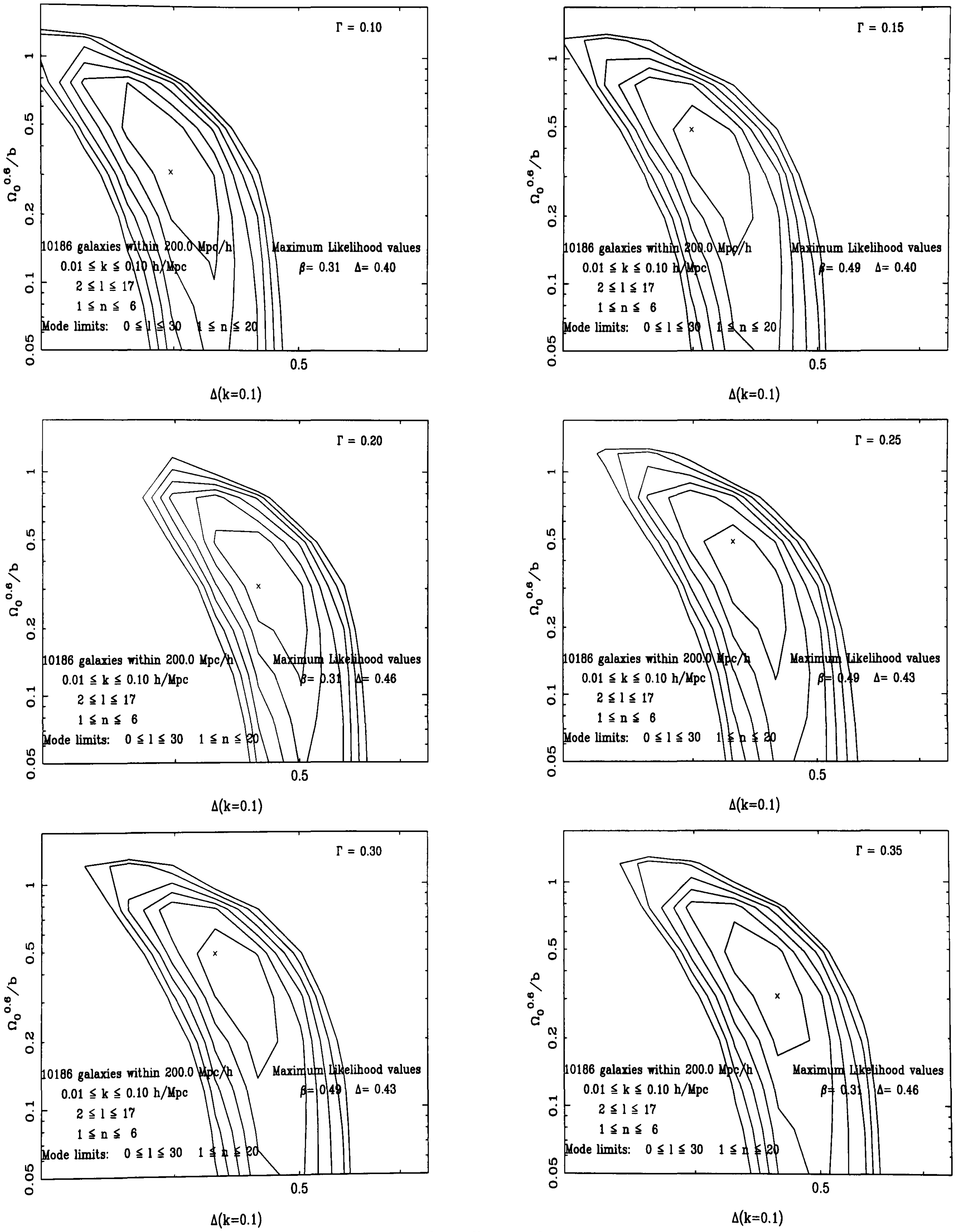


Figure 7.6: As Figure 7.5 except with the effects of the small scale peculiar velocity field included.

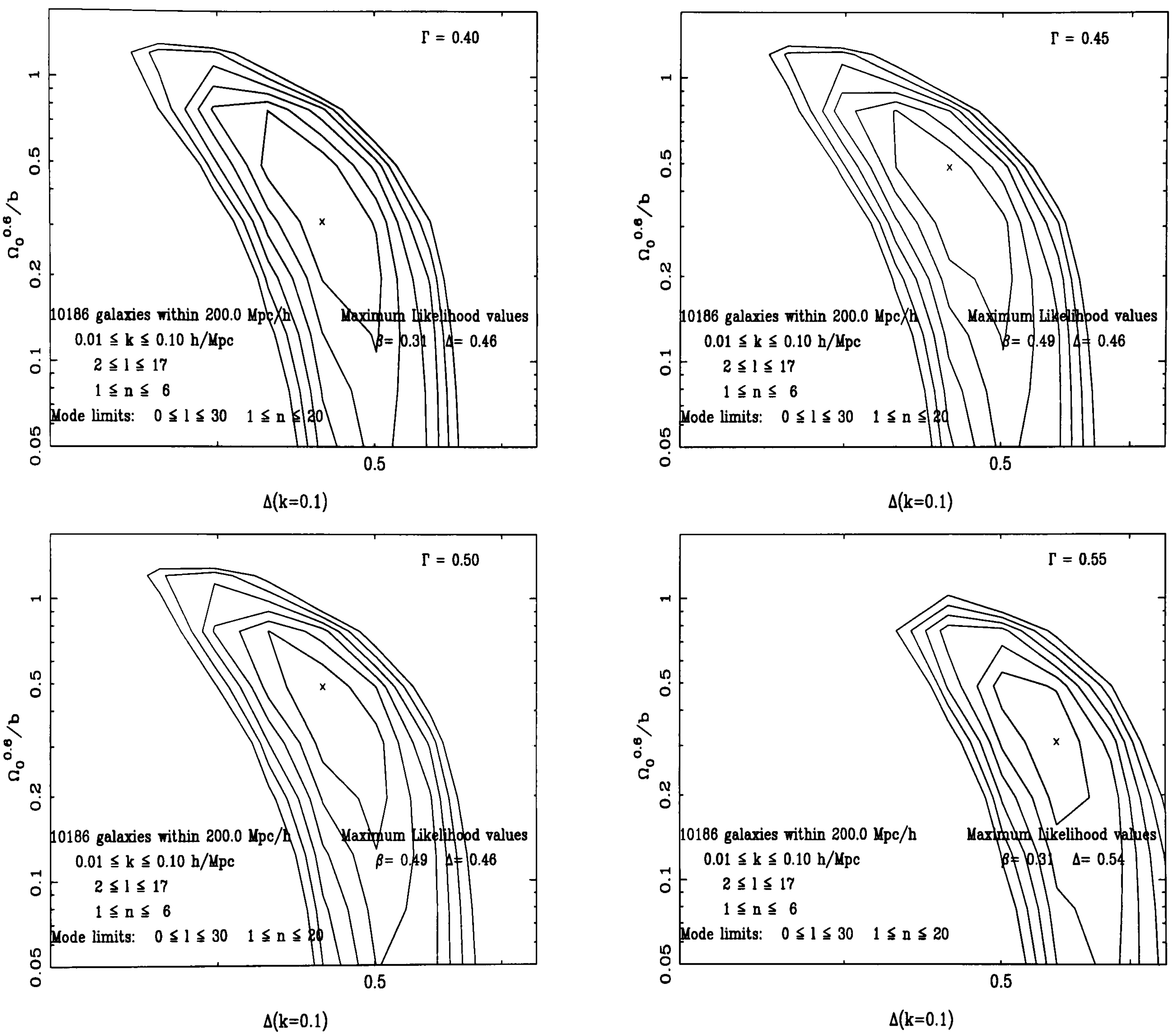


Figure 7.6: continued.

value of Γ assumed for the real-space power spectrum in the likelihood analysis. A value of $\beta = 1$ is ruled out at the 2σ significance level.

Chapter 8

Conclusions

8.1 Introduction

In the course of this thesis I have used several data sets to examine the spatial clustering properties of galaxies and clusters of galaxies in the Universe. I have examined sets of galaxies which are detected by their far infrared emission as well as galaxies detected in the optical. These two sets of galaxies generally have different physical characteristics and, as argued in these conclusions, have different clustering properties.

I have also investigated the dependence of the clustering amplitude of optical galaxies as a function of their luminosity. There is tentative evidence for some dependence of clustering strength on galaxy luminosity.

Clusters of galaxies have long been known to exhibit an enhanced clustering amplitude over that of galaxies themselves and this has been further investigated in this thesis using the deep, complete sample of APM rich galaxy clusters.

All of the measurements made of clustering statistics in this thesis are in redshift-space. Although working in redshift-space gives us a distorted view of the Universe, we can use the degree of distortion as an indication of the value of the density parameter of the Universe. I have attempted to measure $\beta = \Omega^{0.6}/b$, both from the optical Stromlo-APM redshift survey, as well as from the IRAS PSCz redshift survey, using different methods in each case.

In the following sections I discuss the conclusions which may be drawn from the work presented in this thesis. I also discuss the areas in which improvements to the analysis may be made, and I point to further work that could be undertaken in the future.

8.2 The clustering properties of galaxies and galaxy clusters

8.2.1 Clustering as a function of galaxy type

I have measured the power spectrum of three classes of object in this thesis. The first of these was optical galaxies from the Stromlo-APM redshift survey (Loveday *et al.* 1992a). I have also measured the power spectrum of galaxy clusters from the APM rich cluster redshift survey (Dalton *et al.* 1994). The power spectrum of IRAS galaxies was measured in Chapter 4 from the combined IRAS $1.2Jy$ and QDOT surveys, and in Chapter 5 from the new IRAS PSCz galaxy redshift survey. In Figure 8.1, I have compared all the power spectra measured in this thesis. The following paragraphs explain how the figure was produced.

As discussed in previous chapters, the points in the power spectra calculated in this thesis are correlated because of the convolution with the survey window function. From Figure 4.4 in Chapter 4 it can be seen that points in the power spectra are only significantly correlated with their nearest neighbouring points. Therefore, to obtain approximately independent measurements, I have averaged together the points on each of the power spectra in groups of three. Each new point is plotted at the wave number corresponding to the second of the three original points. The error bar placed on this point is the average error of the three original points divided by $\sqrt{3}$. The power spectra used are the ones weighted with the standard values of $P_w(k)$ in the weighting function, i.e. $P_w(k) = 8000\,(\,h^{-1}\mathrm{Mpc})^3$ for the PSCz and Stromlo-APM power spectra, and $P_w(k) = 50000\,(\,h^{-1}\mathrm{Mpc})^3$ for the APM cluster power spectrum. After producing a spectrum of almost independent points for the three power spectra, I have performed a weighted mean fit for the ratio of the cluster power spectrum to the PSCz power spectrum, as well as for the ratio of the Stromlo-APM power spectrum to the PSCz power spectrum. The value of b^2_{opt}/b^2_{IRAS} resulting from this fit is 1.54 ± 0.09 and $b^2_{clus}/b^2_{IRAS} = 7.00 \pm 0.35$.

Using these values for the relative linear bias, the Stromlo-APM power spectrum and the APM cluster power spectrum (and their error bars) are scaled to the amplitude of the PSCz power spectrum. I have then fit the function, $P(k) = Ak^n$, to each power spectrum. A least χ^2 fit for A and n was performed. The data points, and the fits are shown in Figure 8.1, where the whole analysis has been performed over a range in wave number of $0.06 \leq k \leq 0.2\,h\,\mathrm{Mpc}^{-1}$. The squares, together with the dot-dash line, show the result for the APM rich cluster survey, while the crosses and the dashed line show the result for the

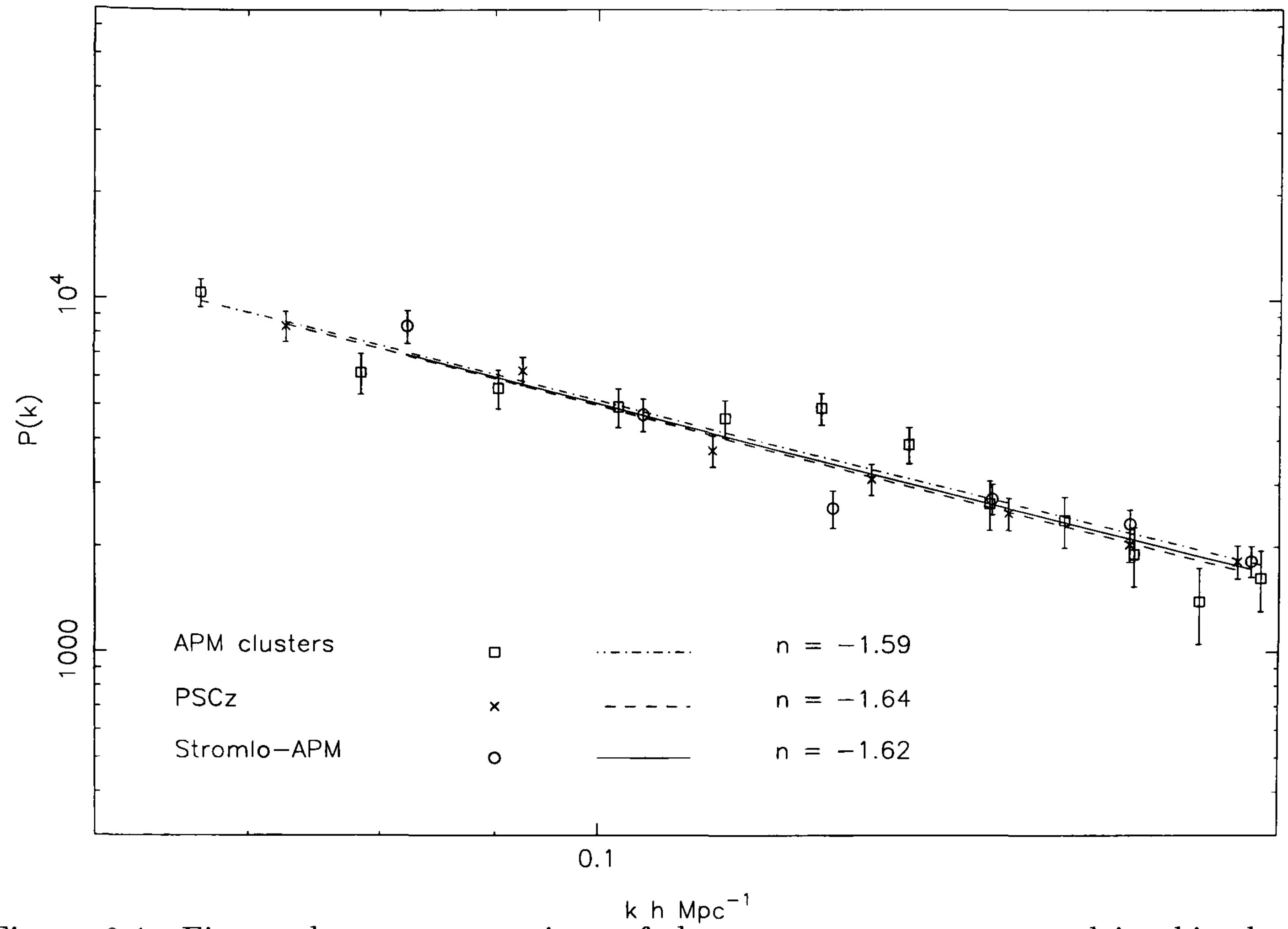


Figure 8.1: Figure shows a comparison of the power spectra measured in this thesis for optical galaxies, IRAS galaxies and galaxy clusters. The optical Stromlo-APM power spectrum and the APM cluster power spectrum have been scaled, using a linear relative bias, to the amplitude of the IRAS PSCz power spectrum.

Sample	n	1σ range	A	1σ range
PSCz	1.64	-1.76 - 1.52	114	88 - 148
Stromlo-APM	1.62	-1.79 - 1.45	122	87 - 172
APM clusters	1.59	-1.70 - 1.48	135	105 - 172

Table 8.1:

IRAS PSCz redshift survey. The circles and the solid line are for the optical Stromlo-APM redshift survey. The errors on the fit parameters have also been evaluated by calculating the tangent points to the $\Delta\chi^2 = 1$ contour in the $A - n$ plane. Table 8.1 shows the resulting slopes, amplitudes and 1σ error ranges.

The slopes (and amplitudes) of the three power spectra are consistent with each other to within the measured errors.

The clustering properties of IRAS and optical galaxies were also compared in Chapter 6, this time using a counts in cells analysis. In Chapter 6, I performed a joint counts in cells analysis of the IRAS PSCz redshift survey with the optical Stromlo-APM survey, and evaluated the linear bias parameter b_{opt}/b_{IRAS} on a number of scales. This comparison showed a definite difference in the clustering properties of IRAS and optical galaxies. On scales of 10 h^{-1}Mpc, 20 h^{-1}Mpc and 30 h^{-1}Mpc the values of the variance of the counts in cells for the IRAS and optical galaxies were different at more than 95% confidence. The actual values of b_{opt}/b_{IRAS} on these three scales were 1.4, 1.4 and 2.2 respectively, in agreement with $b_{opt}/b_{IRAS} \sim 1.4$ found in Chapter 5, although the value $b_{opt}/b_{IRAS} = 2.2$ is slightly higher. Even for this scale (30 h^{-1}Mpc), a value of 1.4 is allowed within the 95% confidence interval. On a scale of 40 h^{-1}Mpc the distributions of IRAS and optical galaxies were consistent. There are some improvements which could be made to this analysis. In particular, one would like to examine the sensitivity of the results to the gridding of galaxies. Preliminary results suggest that moving the grid by half a grid spacing introduces differences in the results of order the 1σ errors. A significant improvement of course would be to have an all-sky optical redshift survey with which to compare the PSCz redshift survey, as there was a worry in Chapter 6 that there were not enough unmasked cells in the analysis for the statistics to converge to central limit theorem behaviour. However, it was reassuring to find that if the elliptical and SO galaxies were removed from the Stromlo-APM survey, then the counts in cells statistics of IRAS PSCz galaxies and optical spirals and irregulars were compatible with no relative bias.

In Chapter 2 I also directly compared the power spectrum of the spiral and irregular galaxies with that of the elliptical and SO galaxies in the Stromlo-APM redshift survey. Again, the two power spectra have the same shape, but there is an offset in the amplitude. The early type galaxies have a power spectrum with an amplitude 1.8 times higher than that of the late type galaxies.

In summary, the statistical analysis in this thesis provides support for the hypothesis that not all types of galaxy can be unbiased tracers of the mass distribution. However, the difference between the power spectra of galaxies of different types, as well as the difference between the power spectra of galaxies and galaxy clusters, is consistent with a linear relative biasing parameter over a large range in wave number. The slopes of the power spectra measured in this thesis are all consistent with $P(k) \propto k^{-1.6}$ over the wave number range $0.06 < k < 0.2\, h$ Mpc^{-1}.

8.2.2 Clustering as a function of galaxy luminosity

In Chapter 2, I examined the power spectrum of sub-volumes of the Stromlo-APM survey volume limited at different redshifts. Subsets of the data volume limited at higher redshift contain galaxies of higher luminosity as the critical absolute magnitude M_{crit} that a galaxy requires to appear in the sample is brighter. It was found in Chapter 2 that the power spectrum amplitude as a function of M_{crit} was relatively flat, except at very bright critical magnitudes (brighter than ~ -20) where there was some evidence from an enhanced clustering amplitude of the brighter galaxies. The fact that a subset of the data with a high value of z_{max} samples a greater volume than one with a low value of z_{max} was accounted for by simulating the procedure using mock Stromlo-APM surveys drawn from an LCDM N-body simulation. Thus the error bars on each of the points included the effects of differing sampling volumes (to the extent that the LCDM model provides a good description of the clustering of Stromlo-APM galaxies). However, the technique used in Chapter 2 does underestimate the size of any possible effect since the subsets of data are not independent. The subsamples with low z_{max} will still contain the high luminosity galaxies, and thus boost the power spectrum amplitude of this sample (if there is a differential clustering amplitude with luminosity). An alternative way to measure the size of any possible effect is to use the technique of Hamilton (1988). Hamilton (1988) uses the CfA survey (Huchra *et al.* 1983) to measure the correlation function of a volume limited subsample of CfA galaxies, limiting at some critical magnitude M_{crit}. He then removes the faintest galaxies from the volume limited sample to leave galaxies brighter than $M_{crit} - \Delta M_{crit}$, and re-measures the correlation function. A relative correlation function amplitude as a function of absolute magnitude is built up by multiplying the ratios $\xi(< M_{crit} - \Delta M_{crit})/\xi(< M_{crit})$ of successive volume limited subsamples. This method has the advantage that the comparison for galaxies of different luminosity is conducted within the same volume. This particular advantage is not so great when we can model the effects of cosmic variance, as done in Chapter 2, using N-body simulated redshift surveys. It does have the second advantage, however, that the galaxy samples are independent, and we do not dilute the effect we are trying to measure. Unfortunately, the method also has the serious drawback that the errors propagate cumulatively through the analysis. In future work I hope to use the method of Hamilton (1988), with a refined error analysis, to re-examine the luminosity dependence of galaxy clustering. Of course, as mentioned in Chapter 2, the dependence of the amplitude of the power spectrum on galaxy luminosity is probably connected with its dependence on galaxy type. Larger optical samples will be required to separate these two effects.

Another future project will be to use the counts in cells analysis of Chapter 6 to examine the clustering of PSCz galaxies as a function of their absolute luminosity.

8.2.3 The power spectrum on large scales

What happens to the galaxy power spectrum on scales larger than those shown in Figure 8.1, i.e. scales with $k < 0.06\ h^{-1}$Mpc? Does the spectrum flatten or turn over? As discussed in Chapter 1, the position of the turn-over in the galaxy power spectrum is a strong constraint on the cosmological model. To be more precise, it is the position of the turn-over in the power spectrum of the mass fluctuations that provides this constraint. We can only measure the power spectrum of galaxy fluctuations, which we assume to be related to the power spectrum of the underlying dark matter through the bias parameter b, where $P_{gal}(k) = b^2 P_{mass}(k)$. If the parameter b is a scale independent parameter, then the position of the turn-over in the galaxy power spectrum tells us the position of the turn-over in the mass power spectrum. Figure 8.1 gives some support to the idea of a scale independent bias parameter, at least on scales $0.06 < k < 0.2\ h\ \mathrm{Mpc}^{-1}$. The simple linear relation between the mass power spectrum and that of the galaxy distribution would break down, however, if environmental effects (caused, for instance, by the feedback processes of galaxy formation) could have a long-range influence on the probability of galaxy formation. The galaxy power spectrum may then be modified, in a non-trivial way, over that of the mass power spectrum on large scales.

At the moment, this is an academic question since the turn-over is not unambiguously detected in the power spectrum of any galaxy (as opposed to cluster) survey to date. Future work will include attempts to improve the measurement of the power spectrum, especially on the very large scales where we may expect to see a turn-over. For example, we could try to truly deconvolve the window function of the survey using a Lucy deconvolution algorithm (Lucy 1974). We will also be measuring the real-space power spectrum of the PSCz galaxies using the methods of Ballinger, Heavens & Taylor (1995).

If we turn our attention to galaxy clusters, the situation is somewhat different. The clustering amplitude of galaxy clusters is much higher than that of galaxies. Any modulation effect would therefore have to be relatively strong in order to distort the shape of the power spectrum of galaxy clusters. We believe that the clustering strength of galaxy clusters is relatively insensitive to cluster richness (see e.g. Croft & Efstathiou 1994a). If the richness of a cluster is modified slightly by environmental effects, the effect on the power spectrum will be small. The power spectrum of galaxy clusters on large scales will thus be likely to trace the shape of the power spectrum of mass fluctuations. In Chapter 3, I have suggested that the turn-over seen in the power spectrum measured from the rich APM cluster redshift survey, at wave number $k \sim 0.024\ h\ \mathrm{Mpc}^{-1}$, is real. Further work on the power spectrum of rich APM clusters is planned. In particular, the APM cluster redshift survey will be modelled using the N-body simulations described in Chapter 7, which have the larger volume of $(600\ h^{-1}\mathrm{Mpc})^3$ (instead of the $(300\ h^{-1}\mathrm{Mpc})^3$ simulations used in Chapter 3). The power spectrum estimator can then be tested, to sufficiently small wave number k, on realistic mock APM cluster redshift surveys. The reality of the turn-over in the power spectrum of rich APM clusters can then be checked.

8.3 The value of β from redshift-space distortions

The other main theme of this thesis has been the measurement of the value of $\Omega^{0.6}/b$ from redshift-space distortions. In Chapter 2, I used the Stromlo-APM galaxy redshift survey to estimate a value of $\Omega^{0.6}/b$. In that chapter, I used a Fourier approach and a measure of the clustering of APM galaxies in real-space (Baugh & Efstathiou 1994) to compare directly the real and redshift-space power spectra in the linear regime. I used N-body simulations to assess the random and systematic errors, which in fact are particularly large. A major conclusion of this work was that there is only a very small region in k space where the linear theory formula of Kaiser (1987) may be used in this type of analysis. Non-linearities introduce systematic errors even at relatively small values of k, depending on the magnitude of small scale peculiar velocities. At very small k, where linear theory is valid, the random and systematic errors in the estimation of the redshift-space power spectrum become problematic.

A different approach was used in Chapter 7, with the PSCz redshift survey. The spherical harmonic analysis of Heavens & Taylor (1995) was implemented with some small refinements. This method was tested on simulated all-sky PSCz surveys drawn from LCDM and SCDM N-body simulations. The method was shown to be successful, especially in the recovery of the correct value of β, although the amplitude of the power spectrum was not recovered so successfully for the SCDM model. Some possible reasons for this failure were mentioned in Chapter 7 and we hope to quantify and correct this problem in future work. For the PSCz redshift survey, a low value of β resulted from this analysis, $\beta = 0.31^{+0.29}_{-0.20}$. If we include the estimated effects of a small scale peculiar velocity field, with a three-dimensional velocity dispersion of 420km s^{-1}, we obtain $\beta = 0.49^{+0.10}_{-0.37}$. In both cases $\beta = 1$ is excluded at the level of 2σ. The maximum likelihood value of β occurs for $\Gamma = 0.2$ in the former case, and $\Gamma = 0.25$ in the latter.

There are several points, concerning this work, which will be addressed in the future. Firstly, as discussed in Chapter 7, we must investigate the so-called 'Hamilton effect' described in various chapters of this thesis. This effect would naturally lead to a low value of β and we must check that this is not the cause of our result. We will test this by repeating the analysis using different flux cuts for the data. If the result is stable to the flux cut used, then we may have some confidence that the result is a true reflection of the value of β. As mentioned in previous chapters, the 'Hamilton effect' is also being investigated directly via the measurement of anisotropies in the correlation function, $\xi(\sigma, \pi)$.

The sensitivity of the spherical harmonic analysis to the position of the boundary at which the survey is truncated must also be investigated. Preliminary results suggest that this is not a problem. If we move the boundary to 240 h^{-1}Mpc and analyse modes with l up to 21 and n up to 8 (i.e. modes with $k = 0.01\ h\ \mathrm{Mpc}^{-1}$ to $k = 0.1\ h\ \mathrm{Mpc}^{-1}$), we again obtain a maximum likelihood value of $\beta = 0.49$ for $\Gamma = 0.25$, including the effects of a small scale peculiar velocity field.

The next project will be to use a Ballinger, Heavens & Taylor (1995) type analysis with the PSCz redshift survey to recover not only the value of β but also the real-space power spectrum of IRAS galaxies. Again, a major project will be the detailed testing of this method against mock surveys extracted from N-body simulations.

It would also seem a worthwhile project to undertake a comparative study of all of the different redshift-space distortion analysis techniques, using a simulated PSCz redshift survey. By performing each type of analysis on the same simulated survey we could compare the techniques in terms of the size of the random and systematic errors, the ranges of validity and the actual values of β recovered. The most successful analysis could then be applied to the PSCz redshift survey.

Power spectrum and redshift-space distortion analysis is likely to be able to provide very strong constraints on the cosmological model in the near future, particularly in the light of the forthcoming AAT 2dF and Sloan Digital Sky redshift surveys.

References

Aaronson M., Bothun G., Mould J., Huchra J., Schommer R., Cornell M. E., 1986. *Ap. J.*, **302**, 536.

Abell G. O., Corwin H. G., Olowin R. P., 1989. *Ap. J. Suppl.*, **70**, 1.

Abell G. O., 1958. *Ap. J. Suppl.*, **3**, 211.

Alpher R. A., Bethe H. A., Gamow G., 1948. *Phys. Rev.*, **73**, 803.

Babul A., Postman M., 1990. *Ap. J.*, **359**, 280.

Babul A., White S. D. M., 1991. *Mon. Not. R. astr. Soc.*, **253**, 31P.

Bahcall N. A., Burgett W. S., 1986. *Ap. J. Lett.*, **300**, L35.

Bahcall N. A., Cen R., 1992. *Ap. J. Lett.*, **398**, L81.

Bahcall N. A., Soneira R. M., 1983. *Ap. J.*, **270**, 20.

Bahcall N. A., West M., 1992. *Ap. J.*, **392**, 419.

Bahcall N. A., Cen R., Gramman M., 1993. *Ap. J. Lett.*, **408**, L77.

Bahcall N. A., 1988. *Ann. Rev. Astron. Astrophys.*, **26**, 631.

Ballinger W. E., Heavens A. F., Taylor A. N., 1995. *Mon. Not. R. astr. Soc.*, **276**, 59P.

Bardeen J. M., Bond J. R., Efstathiou G., 1987. *Ap. J.*, **321**, 28.

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

Barnes J., Dekel A., Efstathiou G., Frenk C. S., 1985. *Ap. J.*, **295**, 368.

Baugh C. M., Efstathiou G., 1993. *Mon. Not. R. astr. Soc.*, **265**, 145.

Baugh C. M., Efstathiou G., 1994. *Mon. Not. R. astr. Soc.*, **267**, 323.

Baugh C. M., Gaztanaga E., 1996. *Mon. Not. R. astr. Soc.*, **280**, 37P.

Baugh C. M., 1996. *Mon. Not. R. astr. Soc.*, **280**, 267.

Baumgart D. J., Fry J. N., 1991. *Ap. J.*, **375**, 25.

Bean A. J., Efstathiou G., Ellis R. S., Peterson B. A., Shanks T., 1983. *Mon. Not. R. astr. Soc.*, **205**, 605.

Beichman C., Habing G. N. H. J., Clegg P. E., Chester T. J., 1988. *IRAS Explanatory Supplement, NASA RP-1190, Vol 1*, U. S. Government Printing Office, Washington D. C.

Bennet C. L., Banday A. J., Gorski K. M., Hinshaw G., Kogut A., Wright E. L., 1996. *submitted*, .

Bertschinger E., Dekel A., 1989. *Ap. J. Lett.*, **336**, L5.

Bertschinger E., Dekel A., Faber S. M., Dressler A., Burstein D., 1990. *Ap. J.*, **364**, 370.

Binggeli B., Tammann G. A., Sandage A., 1987. *Astron. J.*, **94**, 251.

Birkinshaw M., Hughes J. P., 1994. *Ap. J.*, **420**, 33.

Blumenthal G. R., Dekel A., Primack J. R., 1988. *Ap. J.*, **326**, 539.

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984. *Nature*, **311**, 11.

Boesgaard A. M., Steigman G., 1985. *Ann. Rev. Astron. Astrophys.*, **23**, 318.

Bond J. R., Efstathiou G., 1984. *Ap. J. Lett.*, **285**, L45.

Bond J. R., Szalay A. S., 1983. *Ap. J.*, **274**, 443.

Bondi H., Gold T., 1948. *Mon. Not. R. astr. Soc.*, **108**, 252.

Bothun G. D., Beers T. C., Mould J. R., Huchra J. P., 1986. *Ap. J.*, **308**, 510.

Bouchet F., Strauss M. A., Fisher K. B., Yahil A., Huchra J. P., 1993. *Ap. J.*, **417**, 36.

Bower R. G., Coles P., Frenk C., White S., 1993. *Ap. J.*, **405**, 403.

Brandenberger R. H., 1990. In: *Physics of the Early Universe*, 281, eds Peacock J. A., Heavens A. F., Davies A. T., S. U. S. S. P.

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

Bucher M., Turok N., 1995. *Phys. Rev. D*, **52**, 5538.

Burles S., Tytler D., 1996. *preprint astro-ph/9603070*, .

Burstein D., 1990. *Rep. Prog. Phys.*, **53**, 421.

Carswell R. F., Rauch M., Weymann R. J., Cooke A. J., Webb J. K., 1994. *Mon. Not. R. astr. Soc.*, **268**, 1P.

Chaboyer B., Demarque P., Kernan P. J., Krauss L. M., 1996. *Science*, **271**, 957.

Cohen A., Kaplan A., Nelson A., 1993. *Ann. Rev. Nucl. Part. Sci.*, **43**, 27.

Cole S., Fisher K. B., Weinberg D. H., 1994. *Mon. Not. R. astr. Soc.*, **267**, 785.

Cole S., Fisher K. B., Weinberg D. H., 1995. *Mon. Not. R. astr. Soc.*, **275**, 515.

Copi C., Schramm D. N., Turner M. S., 1995. *Science*, **267**, 192.

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

Croft R. A. C., Efstathiou G., 1994a. *Mon. Not. R. astr. Soc.*, **267**, 390.

Croft R. A. C., Efstathiou G., 1994b. *Mon. Not. R. astr. Soc.*, **268**, 23P.

da Costa L. N., Vogeley M. S., Geller M. J., Huchra J. P., Park C., 1994. *Ap. J. Lett.*, **437**, L1.

Dalton G. B., Efstathiou G., Maddox S. J., Sutherland W. J., 1992. *Ap. J. Lett.*, **390**, L1.

Dalton G. B., Croft R. A. C., Efstathiou G., Sutherland W. J., Maddox S. J., Davis M., 1994. *Mon. Not. R. astr. Soc.*, **271**, 47P.

Dalton G. B., 1992. *PhD thesis*, Oxford University.

Davis M., Djorgovski S., 1985. *Ap. J.*, **299**, 15.

Davis M., Geller M. J., 1976. *Ap. J.*, **208**, 13.

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

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985. *Ap. J.*, **292**, 371.

de Lapparent V., Geller M. J., Huchra J. P., 1986. *Ap. J. Lett.*, **302**, L1.

Dearborn D. S. P., Schramm D. N., Steigman G., 1986. *Ap. J.*, **302**, 35.

Dekel A., Rees M. J., 1994. *Ap. J. Lett.*, **422**, L1.

Dekel A., Bertschinger E., Faber S. M., 1990. *Ap. J.*, **364**, 349.

Dekel A., Blumenthal G. R., Primack J. R., Olivier S., 1989. *Ap. J. Lett.*, **338**, L5.

Dekel A., Bertschinger E., Yahil A., Strauss M., Davis M., Huchra J., 1993. *Ap. J.*, **412**, 1.

Dekel A., 1994. *Ann. Rev. Astron. Astrophys.*, **32**, 371.

Dicke R. H., Peebles P. J. E., Roll P. G., Wilkinson D. T., 1965. *Ap. J.*, **142**, 414.

Djorgovski S., Davis M., 1987. *Ap. J.*, **313**, 59.

Doroshkevich A. G., Zeldovich Y. B., Sunyaev R., 1978. *Sov. Astron.*, **22**, 523.

Dressler A., Lynden-Bell D., Burstein D., Davies R. L., Faber S. M., Terlevich R. J., Wegner G., 1987b. *Ap. J.*, **313**, 42.

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

Efstathiou G., Bond J. R., 1986. *Mon. Not. R. astr. Soc.*, **218**, 103.

Efstathiou G., Eastwood J., 1981. *Mon. Not. R. astr. Soc.*, **194**, 503.

Efstathiou G., Bond J. R., White S. D. M., 1992. *Mon. Not. R. astr. Soc.*, **258**, 1P.

Efstathiou G., Ellis R. S., Peterson B. A., 1988. *Mon. Not. R. astr. Soc.*, **232**, 431.

Efstathiou G., Davis M., Frenk C. S., White S. D. M., 1985. *Ap. J. Suppl.*, **57**, 241.

Efstathiou G., Kaiser N., Saunders W., Lawrence A., Rowan-Robinson M., Ellis R. S., Frenk C. S., 1990. *Mon. Not. R. astr. Soc.*, **247**, 10P.

Efstathiou G., Dalton G. B., Sutherland W. J., Maddox S. J., 1992. *Mon. Not. R. astr. Soc.*, **257**, 125.

Efstathiou G., 1990. In: *Physics of the Early Universe*, 361, eds Peacock J. A., Heavens A. F., Davies A. T., S. U. S. S. P.

Efstathiou G., 1992. *Mon. Not. R. astr. Soc.*, **256**, 43P.

Efstathiou G., 1995. *Mon. Not. R. astr. Soc.*, **276**, 1425.

Efstathiou G., 1996. In: *Les Houches Lectures 1995, in press*, ed. Schaeffer R., Elsevier Science Publishers, Netherlands.

Einasto J., Gramman M., Saar E., Tago E., 1993. *Mon. Not. R. astr. Soc.*, **260**, 705.

Eke V., Cole S., Frenk C. S., Navarro J. F., 1996. *preprint astro-ph/9602091*, .

Ellis R. S., 1993. In: *Sky Surveys: Protostars to Protogalaxies*, 165, ed. Soifer B. T., ASP Conference Series V43.

Fabbri R., Natale V., 1990. *Ap. J.*, **363**, 3.

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

Fisher K. B., Nusser A., 1996. *Mon. Not. R. astr. Soc.*, **279**, 1P.

Fisher K. B., Strauss M. A., Davis M., Yahil A., Huchra J. P., 1992. *Ap. J.*, **389**, 188.

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

Fisher K. B., Davis M., Strauss M. A., Yahil A., Huchra J. P., 1994. *Mon. Not. R. astr. Soc.*, **267**, 927.

Fisher K. B., Huchra J. P., Strauss M. A., Davis M., Yahil A., Schlegel D., 1995a. *Ap. J. Suppl.*, **100**, 69.

Fisher K. B., Lahav O., Hoffman Y., Lynden-Bell D., Zaroubi S., 1995b. *Mon. Not. R. astr. Soc.*, **272**, 885.

Fisher K. B., Scharf C., Lahav O., 1994. *Mon. Not. R. astr. Soc.*, **266**, 219.

Freedman W. L., Maddore B. F., Mould J. R., Hill R., Ferrarese L., Kennicutt L., Jnr. R. C., Saha A., Stetson P. B., Graham J. A., Ford H., Hoessel J. G., Huchra J., Hughes S. M., Illingworth G., 1994. *Nature*, **371**, 757.

Frieman J. A., Gaztanaga E., 1994. *Ap. J.*, **425**, 392.

Fukugita M., Okamura S., Yasuda N., 1993. *Ap. J. Lett.*, **412**, L13.

Gamow G., 1949. *Phys. Rev.*, **74**, 505.

Gaztanaga E., Frieman J. A., 1994. *Ap. J. Lett.*, **437**, L13.

Gaztanaga E., Croft R. A. C., Dalton G. B., 1995. *Mon. Not. R. astr. Soc.*, **276**, 336.

Geiss J., 1993. In: *Origin and Evolution of the Elements*, 89, eds Prantzos N., Vangioni-Flam E., Casse M., Cambridge University Press.

Giovanelli R., Haynes M. P., Chincarini G. L., 1986. *Ap. J.*, **300**, 77.

Gorski K., Banday A. J., Bennet C. L., Hinshaw G., Kogut A., Smoot G. F., Wright E. L., 1996. *Ap. J. Lett.*, **464**, L11.

Gott J. R., Statler T., 1984. *Phys. Lett.*, **136B**, 157.

Gramman M., Einasto J., 1992. *Mon. Not. R. astr. Soc.*, **254**, 453.

Gramman M., Cen R., Bahcall N. A., 1993. *Ap. J.*, **419**, 440.

Grogin N., Narayan R., 1996. *Ap. J.*, **464**, 92.

Gunn J. E., Weinberg D. H., 1995. In: *Wide Field Spectroscopy and the Distant Universe*, *The* 35^{th} *Herstmonceux Conference*, eds Maddox S. J., Aragon-Salamanca A.

Guth A., 1981. *Phys. Rev. D*, **23**, 347.

Guzzo G., Fisher K. B., Strauss M. S., Giovanelli R., Haynes M. P., 1995. In: *Astroph. Lett. & Communications: Proceedings of the 3rd Italian Cosmology Meeting.*

Hale-Sutton D., Fong R., Metcalfe N., Shanks T., 1989. *Mon. Not. R. astr. Soc.*, **237**, 569.

Hamilton A. J. S., Kumar P., Lu E., Matthews A., 1991. *Ap. J. Lett.*, **374**, L1.

Hamilton A. J. S., 1988. *Ap. J. Lett.*, **331**, L59.

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

Hamilton A. J. S., 1993a. *Ap. J. Lett.*, **406**, L47.

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

Hamilton A. J. S., 1995. In: *Clustering in the Universe*, 143, eds Maurogordato S., Balkowski C., Tao C., Tran Thanh Van J., Editions Frontieres.

Hamuy M., Phillips M. M., Maza J., Suntzeff N., Schommer R., Aviles R., 1995. *Astron. J.*, **109**, 1.

Harmon R. T., Lahav O., Meurs E. J. A., 1987. *Mon. Not. R. astr. Soc.*, **228**, 5P.

Hata N., Scherrer R. J., Steigman G., Thomas D., Walker T. P., Bludman S., Langacker P., 1995. *Phys. Rev. Lett.*, **75**, 3977.

Hauser M. G., Peebles P. J. E., 1973. *Ap. J.*, **185**, 757.

Heavens A. F., Taylor A. N., 1995. *Mon. Not. R. astr. Soc.*, **275**, 483.

Holtzman J. A., Primack J. R., 1993. *Ap. J.*, **405**, 428.

Hoyle F., 1948. *Mon. Not. R. astr. Soc.*, **108**, 372.

Hu W., Sugiyama N., 1995. *Ap. J.*, **436**, 456.

Hubble E., 1929. *Proc. Nat. Acc. Sci.*, **15**, 168.

Hubble E. P., 1936. *The Realm of the Nebulae*, New Haven: Yale University Press.

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

Huchra J. P., Henry P., Postman M., Geller M. J., 1990. *Ap. J.*, **365**, 66.

Iben I., Truran J. W., 1978. *Ap. J.*, **220**, 980.

Iben I., 1967. *Ap. J.*, **147**, 624.

Icke V., 1985. *Astron. Astrophys.*, **144**, 115.

Iovino A., Giovanelli R., Haynes M., Chincarini G., Guzzo L., 1993. *Mon. Not. R. astr. Soc.*, **265**, 21.

Jain B., Mo H. J., White S. D. M., 1995. *Mon. Not. R. astr. Soc.*, **276**, 25P.

Jimenez R., Thejll P., Jorgensen U., MacDonald J., Pagel B., 1996. *preprint astro-*

ph/9602132, .

Joint IRAS Science Working Group, 1988. *IRAS Point Source Catalogue, Version 2*, U. S. Government Printing Office, Washington D. C.

Jungman G., Kamionkowski M., Kosowsky A., Spergel D. N., 1995. *preprint astro-ph/9512139*, .

Kaiser N., Lahav O., 1989. *Mon. Not. R. astr. Soc.*, **237**, 129.

Kaiser N., Efstathiou G., Ellis R. S., Frenk C. S., Lawrence A., Rowan-Robinson M., Saunders W., 1991. *Mon. Not. R. astr. Soc.*, **252**, 1.

Kaiser N., 1983. *Mon. Not. R. astr. Soc.*, **202**, 1169.

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

Kaiser N., 1986. *Mon. Not. R. astr. Soc.*, **219**, 785.

Kaiser N., 1987. *Mon. Not. R. astr. Soc.*, **227**, 1.

Kauffman G., Nusser A., Steinmetz M., 1996. *preprint astro-ph/9512009*, .

Kawasaki M., Kernan P., Kang H. S., Scherrer R. J., Steigman G., Walker T. P., 1994. *Nucl. Phys. B*, **419**, 105.

Kernan P. J., Sarkar S., 1996. *preprint astro-ph/9603045*, .

Klypin A. A., Kopylov A. I., 1983. *Sov. Astron. Lett.*, **9**, 41.

Klypin A., Holtzman J., Primack J., Regos E., 1993. *Ap. J.*, **416**, 1.

Lahav O., Edge A., Fabian A. C., Putney A., 1989. *Mon. Not. R. astr. Soc.*, **238**, 881.

Lahav O., Lilje P. B., Primack J. R., Rees M. J., 1991. *Mon. Not. R. astr. Soc.*, **251**, 128.

Lahav O., Nemiroff R. J., Piran T., 1990. *Ap. J.*, **350**, 119.

Lahav O., Rowan-Robinson M., Lynden-Bell D., 1988. *Mon. Not. R. astr. Soc.*, **234**, 677.

Lahav O., 1987. *Mon. Not. R. astr. Soc.*, **225**, 213.

Landy S. D., Schectman S. A., Lin H., Kirshner R. P., Oemler A. A., Tucker D., 1996. *Ap. J. Lett.*, **456**, L1.

Lawrence A., Walker D., Rowan-Robinson M., Leech K. J., Penston M. V., 1986. *Mon. Not. R. astr. Soc.*, **219**, 687.

Lawrence A., Rowan-Robinson M., Saunders W., Parry I. R., Xiaoyang X., Ellis R. S., Frenk C. S., Efstathiou G., Kaiser N., Crawford J., 1995. *in press*, .

Liddle A. R., Lyth D., 1993. *Phys. Rep.*, **231**, 1.

Lilje P. B., Efstathiou G., 1988. *Mon. Not. R. astr. Soc.*, **231**, 635.

Lin H., 1995. *PhD thesis*, University of Harvard.

Linde A. D., Mezhlumian A., 1995. *Phys. Rev. D*, **52**, 6789.

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

Loveday J., Peterson B. A., Estathiou G., Maddox S. J., 1992b. *Ap. J.*, **390**, 338.

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

Loveday J., Efstathiou G., Maddox S. J., Peterson B. A., 1995b. *to appear: Astron. J.*, **468**.

Loveday J., 1996. *Mon. Not. R. astr. Soc.*, **278**, 1025.

Lucey J. R., Carter D., 1988. *Mon. Not. R. astr. Soc.*, **235**, 1177.

Lucey J. R., 1983. *Mon. Not. R. astr. Soc.*, **204**, 33.

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

Lumsden S. L., Nichol R. C., Collins C. A., Guzzo L., 1992. *Mon. Not. R. astr. Soc.*, **258**, 1.

Lynden-Bell D., Faber S., Burstein D., Davies R. L., Dressler A., Terlevitch R., Wegner G., 1988. *Ap. J.*, **326**, 19.

Lynden-Bell D., Lahav O., Burstein D., 1989. *Mon. Not. R. astr. Soc.*, **241**, 325.

Maddox S. J., Efstathiou G., Sutherland W. J., 1990b. *Mon. Not. R. astr. Soc.*, **246**, 433.

Maddox S. J., Efstathiou G., Sutherland W. J., 1996. *preprint*, .

Maddox S. J., Sutherland W. J., Efstathiou G., Loveday J., 1990a. *Mon. Not. R. astr. Soc.*, **243**, 692.

Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990c. *Mon. Not. R. astr. Soc.*, **242**, 43P.

Mann R. G., Heavens A. F., Peacock J. A., 1993. *Mon. Not. R. astr. Soc.*, **263**, 798.

Mann R. G., Saunders W., Taylor A. N., 1996. *Mon. Not. R. astr. Soc.*, **279**, 636.

Marzke R. O., Geller M. J., da Costa L. N., Huchra J. P., 1995. *Astron. J.*, **110**, 477.

Mather J. C., Cheng E. S., Cottingham D. A., Eplee R. E., Fixsen D. J., Hewagama T., Isaacman R. B., Jensen K. A., Meyer S. S., Noerdlinger P. D., Read S. M., Rosen L. P., Shafer R. A., Wright E. L., Bennet C. L., Bogess N. W., Hauser M. G., Kelsall R., Moseley S. H., Silverberg R. F., Smoot G. F., Weiss R., Wilkinson D. T., 1994. *Ap. J.*, **420**, 439.

Meiksin A., Davis M., 1986. *Astron. J.*, **91**, 191.

Meurs E. J. A., Harmon R. T., 1988. *Astron. Astrophys.*, **206**, 53.

Mo H. J., Lahav O., 1993. *Mon. Not. R. astr. Soc.*, **261**, 895.

Mo H. J., Jing Y. P., Börner G., 1993. *Mon. Not. R. astr. Soc.*, **264**, 825.

Mo H. J., McGaugh S., Bothun G., 1994. *Mon. Not. R. astr. Soc.*, **267**, 129.

Mo H. J., Peacock J. A., Xia X. Y., 1993. *Mon. Not. R. astr. Soc.*, **260**, 121.

Monaghan J. J., 1992. *Ann. Rev. Astron. Astrophys.*, **30**, 543.

Moore B., Frenk C. S., Weinberg D. H., Saunders W., Lawrence A., Ellis R. S., Kaiser N., Efstathiou G., Rowan-Robinson M., 1990. *Mon. Not. R. astr. Soc.*, **256**, 477.

Moshir M., Kopan G., Conrow T., McCallon H., Hacking P., Gregorich D., Rohrbach G., Melnyk M., Rice W., Fullmer L., White J., Chester T., 1992. *Explanatory Supplement to the IRAS Faint Source Survey, Version 2*, JPLD-10015 8/92 (Pasadena: JPL).

Neugebauer G., Habing H. J., Duinen R. V., Aumann H. H., Baud B., Beichman C. A., Beintema D. A., Boggess N., Clegg P. E., Jong T. D., Emerson J. P., Gautier T. N., Gillett F. C., Harris S., Hauser M. G., Houck J. R., Jennings R. E., Low F. J., Marsden P. L., Miley G., Olnon F. M., Pottasch S. R., Raimond E., Rowan-Robinson M., Soifer B. T., Walker R. G., Wesselius P. R., Young E., 1984. *Ap. J. Lett.*, **278**, L1.

Nichol R. C., Briel U. G., Henry J. P., 1994. *Mon. Not. R. astr. Soc.*, **267**, 771.

Nichol R. C., Collins C. A., Guzzo L., Lumsden S. L., 1992. *Mon. Not. R. astr. Soc.*, **255**, 21P.

Nilson P., 1973. *Uppsala Astr. Obs. Ann.*, **6**.

Nusser A., Dekel A., 1993. *Ap. J.*, **405**, 437.

Nusser A., Dekel A., Bertschinger E., Blumenthal G. R., 1991. *Ap. J.*, **379**, 6.

Oemler A., 1974. *Ap. J.*, **194**, 1.

Olive K. A., Steigman G., 1995. *Ap. J. Suppl.*, **97**, 49.

Olivier S., Blumenthal G. R., Dekel A., Primack J. R., Stanhill D., 1990. *Ap. J.*, **356**, 1.

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

Parker Q. A., Watson F. G., 1995. In: *Wide Field Spectroscopy and the Distant Universe, The* 35^{th} *Herstmonceux Conference*, eds Maddox S. J., Aragon-Salamanca A.

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

Peacock J. A., Dodds S. J., 1996. *Mon. Not. R. astr. Soc.*, **280**, 19P.

Peacock J. A., Nicholson D., 1991. *Mon. Not. R. astr. Soc.*, **253**, 307.

Peacock J. A., West M. J., 1992. *Mon. Not. R. astr. Soc.*, **259**, 494.

Peacock J. A., 1996. In: *EADEN School, Leiden: The Structure of the Universe*, *in press* astro-ph/9601135.

Peebles P. J. E., Yu J. T., 1970. *Ap. J.*, **162**, 815.

Peebles P. J. E., 1973. *Ap. J.*, **185**, 413.

Peebles P. J. E., 1980. *The Large Scale Structure of the Universe*, Princeton University Press.

Peebles P. J. E., 1981. *Ap. J.*, **248**, 885.

Peebles P. J. E., 1984. *Ap. J.*, **284**, 439.

Penzias A. A., Wilson R. W., 1965. *Ap. J.*, **142**, 421.

Postman M., Geller M. J., 1984. *Ap. J.*, **281**, 95.

Postman M., Geller M. J., Huchra J. P., 1986. *Astron. J.*, **91**, 1267.

Postman M., Huchra J. P., Geller M. J., 1992. *Ap. J.*, **384**, 404.

Ratcliffe A., Shanks T., Broadbent A., Parker Q. A., Watson F. G., Oates A. P., Fong R., Collins C. A., 1996. *Mon. Not. R. astr. Soc.*, **281**, 47P.

Ratcliffe A., 1996. *PhD thesis*, University of Durham.

Rees M. J., Sciama D. W., 1968. *Nature*, **517**, 611.

Rees M. J., 1986. *Mon. Not. R. astr. Soc.*, **218**, 25P.

Regos E., Szalay A. S., 1989. *Ap. J.*, **345**, 627.

Rephaeli Y., 1995. *Ann. Rev. Astron. Astrophys.*, **33**, 541.

Rice W., Lonsdale C., Soifer B. T., Neugebauer G., Kopan E. L., Lloyd L. A., de Jong T., Habing H. J., 1988. *Ap. J. Suppl.*, **68**, 91.

Riess A. G., Press W. H., Kirshner R. P., 1995. *Ap. J. Lett.*, **438**, L17.

Romer A. K., Collins C. A., Bohringer H., Cruddace R. G., Ebeling H., MacGillivray H. T., Voges W., 1994. *Nature*, **372**, 75.

Rood R. T., Steigman G., Tinsley B. M., 1976. *Ap. J. Lett.*, **207**, L57.

Rood R. T., 1972. *Ap. J.*, **177**, 681.

Rowan-Robinson M., Crawford J., 1989. *Mon. Not. R. astr. Soc.*, **238**, 523.

Rowan-Robinson M., Lawrence A., Saunders W., Crawford J., Ellis R., Frenk C. S., Parry I., Xiaoyang X., Allington-Smith J., Efstathiou G., Kaiser N., 1990. *Mon. Not. R. astr. Soc.*, **247**, 1.

Rowan-Robinson M., Hughes J., Jones M., Leech K., Vedi K., Wallke K., 1991a. *Mon. Not. R. astr. Soc.*, **249**, 729.

Rowan-Robinson M., Saunders W., Lawrence A., Leech K., 1991b. *Mon. Not. R. astr. Soc.*, **253**, 485.

Rowan-Robinson M., Helou G., Walker D., 1987. *Mon. Not. R. astr. Soc.*, **227**, 589.

Rubakov V. A., Sazhin M., Veryakin A., 1982. *Phys. Lett. B*, **115**, 189.

Sachs R. K., Wolfe A. M., 1967. *Ap. J.*, **147**, 73.

Salaris M., Cheiffi A., Straniero O., 1993. *Ap. J.*, **414**, 580.

Sandage A., Binggeli B., Tammann G. A., 1985. *Astron. J.*, **90**, 1759.

Sandage A., 1972. *Quart. J. R. astr. Soc*, **13**, 282.

Sandage A., 1982. *Ap. J.*, **252**, 553.

Santiago B. X., da Costa L. N., 1990. *Ap. J.*, **362**, 386.

Sasselov D., Goldwirth D., 1995. *Ap. J. Lett.*, **444**, L5.

Saunders W., Rowan-Robinson M., Lawrence A., Efstathiou G., Kaiser N., Ellis R. S., Frenk C. S., 1990. *Mon. Not. R. astr. Soc.*, **242**, 318.

Saunders W., Frenk C., Rowan-Robinson M., Efstathiou G., Lawrence A., Kaiser N., Ellis R., Crawford J., Xia X., Parry I., 1991. *Nature*, **349**, 32.

Saunders W., Sutherland W., Efstathiou G., Tadros H., Maddox S., McMahon R., White S., Oliver S., Keeble O., Rowan-Robinson M., Frenk C., 1995. In: *Wide Field Spectroscopy and the Distant Universe, The* 35^{th} *Herstmonceux Conference*, eds Maddox S. J., Aragon-Salamanca A.

Saunders W., Rowan-Robinson M., Lawrence A., 1992. *Mon. Not. R. astr. Soc.*, **258**, 134.

Saunders W., 1987. *PhD thesis*, University of London.

Scharf C., Lahav O., 1993. *Mon. Not. R. astr. Soc.*, **264**, 439.

Scharf C., Hoffman Y., Lahav O., Lynden-Bell D., 1992. *Mon. Not. R. astr. Soc.*, **256**, 229.

Schechter P. L., 1976. *Ap. J.*, **203**, 297.

Schmidt B. P., Kirshner R. P., Eastmann R. G., Philips M. M., Suntzeff N. B., Hamuy M., Maza J., Aviles R., 1994. *Ap. J.*, **432**, 42.

Schmidt M., 1968. *Ap. J.*, **151**, 393.

Schombert J. M., Bothun G. D., Schneider S. E., McGaugh S. S., 1992. *Astron. J.*, **103**, 1107.

Scott D., Silk J., White M., 1995. *Science*, **268**, 829.

Scott E. L., 1957. *Ap. J.*, **62**, 248.

Seldner M., Peebles P. J. E., 1977. *Ap. J.*, **215**, 703.

Shaver P. A., 1991. *Aust. J. Physics*, **44**, 759.

Shechtman S. A., adn A. A. Oemler S. L., Tucker D., Kirshner R. P., Lin H., Schechter P. L., 1995. In: *Wide Field Spectroscopy and the Distant Universe, The* 35^{th} *Herstmonceux Conference*, eds Maddox S. J., Aragon-Salamanca A.

Shectman S. A., 1985. *Ap. J. Suppl.*, **57**, 77.

Silk J., 1968. *Ap. J.*, **151**, 459.

Skillman E. D., Terlevich R. J., Kennicutt R. C., Garnett D. R., Terlevich E., 1994. *Ap. J.*, **431**, 172.

Smoot G. F., Bennett C. L., Kogut A., Wright E. L., Aymon J., Boggess N. W., Cheng E. S., De Amici G., Gulkis S., Hauser M. G., Hinshaw G., Jackson P. D., Janssen M., Kaita E., Kelsall T., Keegstra P., Lineweaver C., Loewenstein K., Lubin P., Mather J., Meyer S. S., Moseley S. H., Murdock T., Rokke L., Silverberg R. F., Tenorio L., Weiss R., Wilkinson D. T., 1992. *Ap. J. Lett.*, **396**, L1.

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

Soltan A., 1988. *Mon. Not. R. astr. Soc.*, **231**, 309.

Songaila A., Cowie L. L., Hogan C. J., Rugers M., 1994. *Nature*, **368**, 599.

Spite F., Spite M., 1982. *Astron. Astrophys.*, **115**, 357.

Starobinskii A. A., 1983. *Sov. Astron. Lett.*, **9**, 302.

Steigman G., Schramm D. N., Gunn J. E., 1977. *Phys. Rev. Lett.*, **B66**, 202.

Stirling A. J., Peacock J. A., 1996. *preprint astro-ph/9608101*, .

Strauss M. A., Davis M., 1989. In: *Large Scale Motions in the Universe, Proc. of the Vatican Study Week.*, eds Coyne G., Rubin V. C., Princeton University Press.

Strauss M. A., Willick J. A., 1995. *Physics Reports*, **261**, 271.

Strauss M. A., Davis M., Yahil A., Huchra J. P., 1990. *Ap. J.*, **361**, 49.

Strauss M. A., Huchra J. P., Davis M., Yahil A., Fisher K. B., Tonry J., 1992a. *Ap. J. Suppl.*, **83**, 29.

Strauss M. A., Davis M., Yahil A., Huchra J. P., 1992b. *Ap. J.*, **385**, 421.

Strauss M. A., Yahil A., Davis M., Huchra J. P., Fisher K., 1992c. *Ap. J.*, **397**, 395.

Sunyaev R. A., Zeldovich Y. B., 1970. *Astroph. Sp. Sci.*, **7**, 3.

Sunyaev R. A., Zeldovich Y. B., 1980. *Ann. Rev. Astron. Astrophys.*, **18**, 537.

Sutherland W. J., Efstathiou G., 1991. *Mon. Not. R. astr. Soc.*, **248**, 159.

Sutherland W. J., 1988. *Mon. Not. R. astr. Soc.*, **234**, 159.

Szalay A. S., Schramm D. N., 1985. *Nature*, **314**, 718.

Szalay A. S., 1988. *Ap. J.*, **333**, 21.

Tadros H., Efstathiou G., 1995. *Mon. Not. R. astr. Soc.*, **276**, 45P.

Tadros H., Efstathiou G., 1996. *Mon. Not. R. astr. Soc.*, **282**, 1381.

Tanvir N. R., Shanks T., Ferguson H. C., Robinson D. R. T., 1995. *Nature*, **377**, 27.

Tonry J. L., Schneider D. P., 1988. *Astron. J.*, **96**, 807.

Tonry J. L., Ajhar E. A., Luppino G. A., 1990. *Astron. J.*, **100**, 1416.

Toomre A., 1977. In: *The Evolution of Galaxies and Stellar Populations*, 410, eds B.M.Tinsley, R.B.Larson, Yale University Observatory.

Tully R. B., Fisher J. R., 1977. *Astron. Astrophys.*, **54**, 661.

Tully R. B., 1988. *Nearby Galaxies Catalogue*, Cambridge: Cambridge University Press.

Tytler D., Fan X. M., Burles S., 1996. *Nature*, **381**, 207.

VandenBerg D. A., 1983. *Ap. J. Suppl.*, **51**, 29.

VandenBerg D. A., 1988. In: *The Harlow-Shapley Symposium on Globular Cluster Systems in Galaxies*, eds Grindlay J. E., Philip A. G. D.

Vilenkin A., Shellard E. P. S., 1993. *Cosmic Strings and Other Topological Defects*, Cambridge University Press.

Villumsen J. V., Strauss M. A., 1987. *Ap. J.*, **322**, 37.

Vogeley M. S., Park C., Geller M. J., Huchra J. P., 1992. *Ap. J. Lett.*, **391**, L5.

Vorontsov-Velyaminov B. A., Arhipova V. P., 1963 to 1974. *Morphological Catalogue of Galaxies: Vols 2-5*, Moscow State University.

Vorontsov-Velyaminov B. A., Krasnogorskaja A. A., 1962. *Morphological Catalogue of Galaxies: Vol 1*, Moscow State University.

Wagoner R. V., Fowler W. A., Hoyle F., 1967. *Ap. J.*, **148**, 340.

Walker T. P., Steigman G., Schramm D. N., Olive K. A., Kang H. S., 1991. *Ap. J.*, **376**, 51.

Weinberg D. H., Cole S., 1992. *Mon. Not. R. astr. Soc.*, **259**, 652.

Weinberg S., 1972. *Gravitation and Cosmology*, Wiley.

White S. D. M., Davis M., Efstathiou G., Frenk C. S., 1987a. *Nature*, **330**, 451.

White S. D. M., Frenk C. S., Davis M., Efstathiou G., 1987b. *Ap. J.*, **313**, 505.

White M., Scott D., Silk J., 1994. *Ann. Rev. Astron. Astrophys.*, **32**, 319.

White S. D. M., Tully R. B., Davis M., 1988. *Ap. J. Lett.*, **333**, L45.

Whitmore B. C., Gilmore D. M., Jones C., 1993. *Ap. J.*, **407**, 489.

Wilson T. L., Rood R. T., 1994. *Ann. Rev. Astron. Astrophys.*, **32**, 191.

Wilson M. L., Silk J., 1981. *Ap. J.*, **243**, 14.

Yahil A., Walker D., Rowan-Robinson M., 1986. *Ap. J. Lett.*, **301**, L1.

Yahil A., 1988. In: *Large Scale Motions in the Universe, Proceedings of the Pontifical Academy of Sciences Study Week No. 27.*, eds Coyne G., Rubin V. C., Princeton University Press.

Yang J., Turner M. S., Steigman G., Schramm D. N., Olive K. A., 1984. *Ap. J.*, **281**, 493.

Zaroubi S., Hoffman Y., 1996. *Ap. J.*, **462**, 25.

Zeldovich Y. B., 1970. *Astron. Astrophys.*, **5**, 84.

Appendix A

Biases in estimating P(k) I - Volume limited surveys

As described in Chapter 2, the quantity $\Delta(k)$ defined in equation (2.10) requires an estimate of the mean galaxy density $\overline{n}$ and we must derive this estimate from the sample itself, *e.g.* for a volume limited sample,

$$\overline{n} = \frac{1}{V_s}\sum_i n_i W(\mathbf{x}_i), \qquad (A1)$$

where V_s is the volume of the survey and the sum extends over all infinitesimal cells of galaxy count n_i. Equation (2.11) of Chapter 2 states

$$|\Delta(k)|^2 = \hat{n}_o(k)\,\hat{n}_0^\star(k) - \overline{n}\hat{n}_o(k)\,\hat{W}_e^\star(k) - \overline{n}\hat{n}_o^\star\hat{W}_e(k) + \overline{n}^2\left|\hat{W}_e(k)\right|^2. \qquad (A2)$$

Inserting (A1) into (A2) gives

$$\begin{aligned}
|\Delta(k)|^2 \;=\; & \frac{1}{V^2}\sum_{ij} n_i n_j W(x_i)\,W(x_j)\,e^{i\mathbf{k}\cdot(\mathbf{x_i}-\mathbf{x_j})} \\
& -\frac{1}{VV_s}\sum_i n_i W(x_i)\sum_j n_j W(x_j)\,e^{i\mathbf{k}\cdot\mathbf{x_j}}\hat{W}_e^\star(k) \\
& -\frac{1}{VV_s}\sum_i n_i W(x_i)\sum_j n_j W(x_j)\,e^{-i\mathbf{k}\cdot\mathbf{x_j}}\hat{W}_e(k) \\
& +\frac{1}{V_s^2}\sum_{ij} n_i n_j W(x_i)\,W(x_j)\left|\hat{W}_e(k)\right|^2. \qquad (A3)
\end{aligned}$$

Now we split all the terms on the RHS of equation (A3) into the case where $i=j$, and the case where $i \neq j$,

$$\begin{aligned}
|\Delta(k)|^2 \;=\; & \frac{1}{V^2}\sum_i n_i^2 W^2(x_i) + \frac{1}{V^2}\sum_{i\neq j} n_i n_j W(x_i)\,W(x_j)\,e^{i\mathbf{k}\cdot(\mathbf{x_i}-\mathbf{x_j})} \\
& -\frac{1}{VV_s}\sum_i n_i^2 W^2(x_i)\,e^{i\mathbf{k}\cdot\mathbf{x_i}}\hat{W}_e^\star(k) - \frac{1}{VV_s}\sum_{i\neq j} n_i n_j W(x_i)\,W(x_j)\,e^{i\mathbf{k}\cdot\mathbf{x_j}}\hat{W}_e^\star(k)
\end{aligned}$$

$$-\frac{1}{VV_s}\sum_i n_i^2 W^2(x_i)\, e^{-i\mathbf{k}\cdot\mathbf{x_i}}\hat{W}_e(k) - \frac{1}{VV_s}\sum_{i\neq j} n_i n_j W(x_i)\, W(x_j)\, e^{-i\mathbf{k}\cdot\mathbf{x_j}}\hat{W}_e(k)$$
$$+\frac{1}{V_s^2}\sum_i n_i^2 W^2(x_i)\left|\hat{W}_e(k)\right|^2 + \frac{1}{V_s^2}\sum_{i\neq j} n_i n_j W(x_i)\, W(x_j)\left|\hat{W}_e(k)\right|^2. \qquad (A4)$$

Upon ensemble averaging these terms become

$$\begin{aligned}
\left\langle |\Delta(k)|^2 \right\rangle &= \frac{1}{V^2}\sum_i \overline{n} dV_i W^2(x) \\
&+\frac{1}{V^2}\sum_{i\neq j}\overline{n}^2\left(1+\xi(x_{ij})\right) dV_i dV_j W(x_i)\, W(x_j)\, e^{i\mathbf{k}\cdot(\mathbf{x_i}-\mathbf{x_j})} \\
&-\frac{1}{VV_s}\sum_i \overline{n} dV_i W^2(x_i)\, \hat{W}_e^{\star} e^{i\mathbf{k}\cdot\mathbf{x_i}} \\
&-\frac{1}{VV_s}\sum_{i\neq j}\overline{n}^2\left(1+\xi(x_{ij})\right) W(x_i)\, W(x_j)\, e^{i\mathbf{k}\cdot\mathbf{x_i}} dV_i dV_j \hat{W}_e^{\star}(k) \\
&-\frac{1}{VV_s}\sum_i \overline{n} dV_i W^2(x)\, \hat{W}_e(k)\, e^{-i\mathbf{k}\cdot\mathbf{x_i}} \\
&-\frac{1}{VV_s}\sum_{i\neq j}\overline{n}^2\left(1+\xi(x_{ij})\right) W(x_i)\, W(x_j)\, e^{-\mathbf{k}\cdot\mathbf{x_j}} dV_i dV_j \hat{W}_e(k) \\
&+\frac{1}{V_s^2}\sum_i \overline{n} dV_i W^2(x)\left|\hat{W}_e(k)\right|^2 \\
&+\frac{1}{V_s^2}\sum_{i\neq j}\overline{n}^2\left(1+\xi(x_{ij})\right) W(x_i)\, W(x_j)\left|\hat{W}_e(k)\right|^2 dV_i dV_j, \qquad (A5)
\end{aligned}$$

where $x_{ij} = (\mathbf{x_i} - \mathbf{x_j})$. Taking each of the terms on the RHS of equation (A5) individually,

$$1^{st} \text{ term} = \frac{\overline{n} V_s}{V^2}, \qquad (A6.1)$$

$$\begin{aligned}
2^{nd} \text{ term} &= \frac{1}{V^2}\sum_{i\neq j}\overline{n}^2 dV_i dV_2 W(x_i)\, W(x_j)\, e^{i\mathbf{k}\cdot(\mathbf{x_i}-\mathbf{x_j})} \\
&+\frac{1}{V^2}\sum_{i\neq j}\overline{n}^2 \xi(x_{ij})\, dV_i dV_j W(x_i)\, W(x_j)\, e^{i\mathbf{k}\cdot(\mathbf{x_i}-\mathbf{x_j})},
\end{aligned}$$

and since

$$\hat{W}(k) = \frac{1}{V}\sum_i W(x_i)\, e^{i\mathbf{k}\cdot\mathbf{x_i}} dV_i,$$

$$\begin{aligned}
2^{nd} \text{ term} &= \overline{n}^2\left|\hat{W}(k)\right|^2 + \frac{\overline{n}^2}{V^2}\sum_{i\neq j}\xi(x_{ij})\, dV_i W(x_i)\, e^{i\mathbf{k}\cdot\mathbf{x_i}}\sum_j dV_j W^{\star}(x_j)\, e^{-i\mathbf{k}\cdot\mathbf{x_j}} \\
&= \overline{n}^2\left|\hat{W}(k)\right|^2 + \frac{\overline{n}^2}{V^2}\sum_{i\neq j}\left(V\sum_{k'} P(k')\, e^{-i\mathbf{k}'\cdot(\mathbf{x_i}-\mathbf{x_j})}\right) dV_i W(x_i)\, e^{i\mathbf{k}\cdot\mathbf{x_i}}\sum_j dV_j W^{\star}(x_j)\, e^{-i\mathbf{k}\cdot\mathbf{x_j}}
\end{aligned}$$

$$
\begin{aligned}
&= \overline{n}^2 \left|\hat{W}(k)\right|^2 + \frac{\overline{n}^2}{V}\sum_{k'} P(k') \left(\sum_i dV_i W(x_i) e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x_i}}\right)\left(\sum_j dV_i W(x_j) e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x_j}}\right) \\
&= \overline{n}^2 \left|\hat{W}(k)\right|^2 + \frac{\overline{n}^2}{V}\sum_{k} P(k') \left|W(k-k')\right|^2, \qquad (A6.2)
\end{aligned}
$$

$$
\begin{aligned}
3^{rd}\text{ term} &= -\frac{1}{VV_s}\sum_i \overline{n} dV_i W^2(x_i)\hat{W}_e^\star(k) e^{i\mathbf{k}\cdot\mathbf{x_i}} \\
&= -\frac{1}{VV_s}\sum_i \overline{n} dV_i W(x_i) e^{i\mathbf{k}\cdot\mathbf{x_i}}\hat{W}^\star(k) \\
&= \frac{\overline{n}\left|\hat{W}(k)\right|^2}{V_s}, \qquad (A6.3)
\end{aligned}
$$

$$
\begin{aligned}
4^{th}\text{ term} &= -\frac{1}{VV_s}\sum_{i\neq j}\overline{n}^2 W(x_i) W(x_j) e^{i\mathbf{k}\cdot\mathbf{x_j}} dV_i dV_j \hat{W}^\star(k) \\
&\quad -\frac{1}{VV_s}\sum_{i\neq j}\xi(x_{ij})\overline{n}^2 W(x_i) W(x_j) e^{i\mathbf{k}\cdot\mathbf{x_i}} dV_i dV_j \hat{W}^\star(k) \\
&= -\frac{\overline{n}^2}{VV_s}\hat{W}^\star(k)\sum_i W(x_i) dV_i \sum_j W(x_j) e^{i\mathbf{k}\cdot\mathbf{x_j}} dV_j \\
&\quad -\frac{1}{VV_s}\sum_{i\neq j}\xi(x_{ij})\overline{n}^2 W(x_i) W(x_j) e^{i\mathbf{k}\cdot\mathbf{x_i}} dV_i dV_j \hat{W}^\star(k) \\
&= -\frac{\overline{n}^2}{VV_s}\hat{W}^\star(k) V_s V \hat{W}(k) \\
&\quad -\frac{1}{VV_s}\sum_{i\neq j}\xi(x_{ij})\overline{n}^2 W(x_i) W(x_j) e^{i\mathbf{k}\cdot\mathbf{x_i}} dV_i dV_j \hat{W}^\star(k) \\
&= -\overline{n}^2\left|\hat{W}(k)\right|^2 - \frac{1}{VV_s}\sum_{i\neq j}\xi(x_{ij})\overline{n}^2 W(x_i) W(x_j) e^{i\mathbf{k}\cdot\mathbf{x_j}} dV_i dV_j \hat{W}^\star(k) \\
&= -\overline{n}^2\left|\hat{W}(k)\right|^2 - \frac{\overline{n}^2}{VV_s}\int\int \xi(x_{ij}) W(x_i) W(x_j) e^{i\mathbf{k}\cdot\mathbf{x_i}}\hat{W}^\star(k), \qquad (A6.4)
\end{aligned}
$$

$$
\begin{aligned}
5^{th}\text{ term} &= -\frac{1}{VV_s}\sum_i \overline{n} dV_i W^2(x_i)\hat{W}(k) e^{-i\mathbf{k}\cdot\mathbf{x_i}} \\
&= -\frac{\overline{n}}{VV_s}\sum_i W(x_i) dV_i W(x_i) e^{-i\mathbf{k}\cdot\mathbf{x_i}}\hat{W}(k) \\
&= -\frac{\overline{n}}{VV_s}\hat{W}(k) V\hat{W}^\star(k) \\
&= -\frac{\overline{n}}{V_s}\left|\hat{W}(k)\right|^2, \qquad (A6.5)
\end{aligned}
$$

$$\begin{aligned}
6^{th} \text{ term} &= -\frac{1}{VV_s}\sum_{i\neq j}\overline{n}^2 W(x_i)W(x_j)e^{-i\mathbf{k}\cdot\mathbf{x_j}}dV_i dV_j\hat{W}(k) \\
&\quad -\frac{1}{VV_s}\sum_{i\neq j}\xi(x_{ij})\overline{n}^2 W(x_i)W(x_j)e^{-i\mathbf{k}\cdot\mathbf{x_j}}dV_i dV_j\hat{W}(k) \\
&= -\frac{1}{VV_s}\sum_{i}\overline{n}^2 W(x_i)dV_i\hat{W}(k)\sum_{j}W(x_j)e^{-i\mathbf{k}\cdot\mathbf{x_j}}dV_j \\
&\quad -\frac{1}{VV_s}\sum_{i\neq j}\xi(x_{ij})\overline{n}^2 W(x_i)W(x_j)e^{-i\mathbf{k}\cdot\mathbf{x_j}}dV_i dV_j\hat{W}(k) \\
&= -\overline{n}^2\left|\hat{W}(k)\right|^2 - \frac{\overline{n}^2}{VV_s}\int\int\xi(x_{ij})W(x_i)W(x_j)e^{-i\mathbf{k}\cdot\mathbf{x_j}}dV_i dV_j\hat{W}(k), \qquad (A6.6)
\end{aligned}$$

$$\begin{aligned}
7^{th} \text{ term} &= \frac{1}{V_s^2}\sum_i \overline{n}dV_i W^2(x_i)\left|\hat{W}_e(k)\right|^2 \\
&= \frac{\overline{n}}{V_s}\left|\hat{W}_e(k)\right|^2, \qquad (A6.7)
\end{aligned}$$

$$\begin{aligned}
8^{th} \text{ term} &= \frac{1}{V_s^2}\sum_{i\neq j}\overline{n}^2 W(x_i)W(x_j)dV_i dV_j\left|\hat{W}_e(k)\right|^2 \\
&\quad +\frac{1}{V_s^2}\sum_{i\neq j}\overline{n}^2\xi(x_{ij})W(x_i)W(x_j)\left|\hat{W}_e(k)\right|^2 dV_i dV_j \\
&= \overline{n}^2\left|\hat{W}(k)\right|^2 + \frac{1}{V_s^2}\int\int\overline{n}^2\xi(x_{ij})W(x_i)W(x_j)\left|\hat{W}_e(k)\right|^2 dV_i dV_j. \qquad (A6.8)
\end{aligned}$$

So, summing the terms (A6.1) to (A6.8) gives

$$\begin{aligned}
\left\langle|\Delta(k)|^2\right\rangle &= \frac{\overline{n}V_s}{V^2} + \frac{\overline{n}^2}{V}\sum_{k'}\left|\hat{W}(k-k')\right|^2 P(k') - \frac{\overline{n}}{V_s}\left|\hat{W}(k)\right|^2 \\
&\quad -\frac{\overline{n}^2}{VV_s}\int\int\xi(x_{ij})W(x_i)W(x_j)e^{i\mathbf{k}\cdot\mathbf{x_i}}\hat{W}^{\star}(k)dV_i dV_j \\
&\quad -\frac{\overline{n}^2}{VV_s}\int\int\xi(x_{ij})W(x_i)W(x_j)e^{-i\mathbf{k}\cdot\mathbf{xx}}\hat{W}(k)dV_i dV_j \\
&\quad +\frac{1}{V_s^2}\int\int\overline{n}^2\xi(x_{ij})W(x_i)W(x_j)\left|\hat{W}(k)\right|^2 dV_i dV_j, \qquad (A7)
\end{aligned}$$

and using

$$\sum_k\left|\hat{W}(k)\right|^2 = \frac{V_s}{V},$$

we obtain

$$\begin{aligned}
\left\langle|\Delta(k)|^2\right\rangle &= \frac{\overline{n}}{V}\sum_k\left|\hat{W}(k)\right|^2 + \frac{\overline{n}^2}{V}\sum_k\left|\hat{W}(k-k')\right|^2 P(k') \\
&\quad -\frac{\overline{n}}{V_s}\left|\hat{W}(k)\right|^2 - \frac{\overline{n}^2}{VV_s}\int\int\xi(x_{ij})W(x_i)W(x_j)e^{i\mathbf{k}\cdot\mathbf{x_i}}\hat{W}^{\star}(k)dV_i dV_j
\end{aligned}$$

$$-\frac{\overline{n}^2}{VV_s}\int\int \xi\left(x_{ij}\right) W\left(x_i\right) W\left(x_j\right) e^{-\mathbf{k}\cdot\mathbf{x_j}} \hat{W}\left(k\right) dV_i dV_j$$
$$+\frac{\overline{n}^2}{V_s^2}\int\int \xi\left(x_{ij}\right) W\left(x_i\right) W\left(x_j\right) \left|\hat{W}\left(k\right)\right|^2 dV_i dV_j. \qquad (A8)$$

This is equation (2.24) of Chapter 2, which as discussed in that chapter simplifies to

$$-\frac{\overline{n}}{V_s}|\hat{W}(\mathbf{k}')|^2 - \overline{n}^2\sigma_s^2|\hat{W}(\mathbf{k}')|^2 \qquad (A9)$$

$$\equiv -\frac{\overline{n}}{V_s}|\hat{W}(\mathbf{k}')|^2 - \overline{n}^2\left(\sum_{\mathbf{k}'}|W(\mathbf{k}')|^2P(\mathbf{k}')\right)|\hat{W}(\mathbf{k}')|^2 \qquad (A10)$$

if the variance

$$\sigma_s^2 = \frac{1}{V_s^2}\int\int \xi(x_{ij})W(\mathbf{x}_i)W(\mathbf{x}_j)\,dV_i\,dV_j \qquad (A11)$$

is dominated by fluctuations on scales smaller than the scale of the survey. Equation (A10) gives the bias in the estimate of $< |\Delta(k)|^2 >$; the bias in the power spectrum is given by equation (A10) multiplied by $\frac{1}{V_s}\left(\frac{V}{\overline{n}}\right)^2$ (see equation 2.19 of Chapter 2). Thus the bias terms in the estimate of the power spectrum are

$$-\frac{V^2}{N_G V_s}\left|\hat{W}\left(k'\right)\right|^2 - \left(\frac{V^2}{V_s}\right)\left|\hat{W}\left(k\right)\right|^2\sum_{k'}\left|\hat{W}\left(k'\right)\right|^2 P\left(k'\right), \qquad (A12)$$

the first of which depends on the number of galaxies in the survey N_G while the second term is independent of the number of galaxies but depends on the shape of the power spectrum $P(k)$. An approximate expression for the bias in $P(k)$, equivalent to equation (A10), has been derived by Peacock & Nicholson (1991).

Appendix B

Biases in estimating P(k) II - Flux limited surveys

We can generalise the derivation of the bias terms caused by estimating the mean galaxy density of a redshift survey from the survey itself to the case of flux limited samples. I start with equation (2.4.1) of Feldman, Kaiser & Peacock (1994) which is

$$\left\langle |F(k)|^2 \right\rangle = \frac{\int d^3r \int d^3r' w(r)\, w(r') \left\langle [n_g(r) - \alpha n_s(r)]\,[n_g(r') - \alpha n_s(r')] \right\rangle e^{i\mathbf{k}(\mathbf{r}-\mathbf{r}')}}{\int d^3r \overline{n}^2(r)} \qquad (B1)$$

where s signifies the galaxies in the random catalogue, and r the galaxies in the real catalogue. Expanding out the terms in brackets,

$$\begin{aligned}
\left(\int d^3r \overline{n}^2(r)\, w^2(r)\right)\left\langle |F(k)|^2 \right\rangle &= \int\int d^3r d^3r' w(r)\, w(r') \left\langle n_g(r)\, n_g(r') \right\rangle e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \\
&- \int\int d^3r d^3r' w(r)\, w(r') \left\langle n_g(r)\, \alpha n_s(r') \right\rangle e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \\
&- \int\int d^3r d^3r' w(r)\, w(r') \left\langle \alpha n_s(r)\, n_g(r') \right\rangle e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \\
&+ \int\int d^3r d^3r' w(r)\, w(r') \left\langle \alpha n_s(r)\, \alpha n_s(r') \right\rangle e^{i\mathbf{k}(\mathbf{r}-\mathbf{r}')}. \qquad (B2)
\end{aligned}$$

In practice, we estimate the quantity α in the calculation of flux limited power spectra as follows

$$\begin{aligned}
\alpha = \frac{n'}{n'_s} &= \frac{1}{n'_s}\sum_i w'(r_i) \\
&= \frac{1}{n'_s}\int d^3r n_g(r)\, w'(r). \qquad (B3)
\end{aligned}$$

The sum in the first line of equation (B3) is over the galaxies in the survey, and n'_s is the same sum over points in the random catalogue. For the rest of this analysis I assume that the uncertainty in this quantity is negligible. The weight function $w'(r)$ in equation (B3) is given by

$$w'(r) = \frac{1}{1 + 4\pi\overline{n}(r)J_3} \qquad (B4)$$

(see equation 2.29 of Chapter 2). Inserting equation (B3) into equation (B2) gives

$$\begin{aligned}
A\left\langle |F(k)|^2 \right\rangle &= \int\int d^3r d^3r' w(r) w(r') \left\langle n_g(r) r n_g(r') \right\rangle e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \\
&- \int\int d^3r d^3r' w(r) w(r') \left\langle \frac{n_s(r')}{n_s'} \int d^3r n_g(r) n_g(r'') w'(r'') \right\rangle e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \\
&- \int\int d^3r d^3r' w(r) w(r') \left\langle \frac{n_s(r)}{n_s'} \int d^3r'' n_g(r'') w'(r'') n_g(r') \right\rangle e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \\
&+ \int\int d^3r d^3r' w(r) w(r') \left\langle \frac{n_s(r) n_s(r')}{(n_s')^2} \int\int d^3r'' d^3r''' n_g(r'') w'(r'') n_g(r''') w'(r''') \right\rangle e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}
\end{aligned} \qquad (B5)$$

where the normalizing factor $A = \int d^3r\overline{n}^2(r)w^2(r)$. We have the following identities relating the number densities to the correlation function (see Feldman, Kaiser & Peacock 1994, with $\overline{n}_s(r) = \alpha^{-1}\overline{n}(r)$):

$$\begin{aligned}
\left\langle n_g(r) n_g(r') \right\rangle &= \overline{n}(r)\overline{n}(r') \left[1 + \xi(r - r')\right] + \overline{n}(r)\delta(r - r') && (B6) \\
\left\langle n_s(r) n_s(r') \right\rangle &= \overline{n}_s(r)\overline{n}_s(r') + \overline{n}_s(r)\delta(r - r') && (B7) \\
\left\langle n_g(r) n_s(r') \right\rangle &= \left\langle n_g(r) \right\rangle \left\langle n_s(r) \right\rangle = \overline{n}_s(r)\overline{n}_g(r). && (B8)
\end{aligned}$$

Putting equations (B6) to (B8) into equation (B5),

$$\begin{aligned}
A\left\langle |F(k)|^2 \right\rangle &= \int\int d^3r d^3r' w(r) w(r') \left\langle n_g(r) n_g(r') \right\rangle e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \\
&- \int\int d^3r d^3r' w(r) w(r') \frac{\overline{n}_s(r')}{n_s'} \int d^3r'' \left[\overline{n}(r)\overline{n}(r'')\left(1 + \xi(r - r'')\right) + \overline{n}(r)\delta(r - r'')\right] w'(r'') e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \\
&- \int\int d^3r d^3r' w(r) w(r') \frac{\overline{n}_s(r)}{n_s'} \int d^3r'' \left[\overline{n}(r)\overline{n}(r'')\left(1 + \xi(r'' - r')\right) + \overline{n}(r)\delta(r'' - r')\right] w'(r'') \\
&+ \Bigg(\int\int d^3r d^3r' w(r) w(r') \frac{1}{n_s'} \left[\overline{n}_s(r)\overline{n}_s(r') + \overline{n}_s(r)\delta(r - r')\right] \\
&\times \int\int d^3r'' d^3r''' \left[\overline{n}(r'')\overline{n}(r''')\left(1 + \xi(r'' - r''')\right) + \delta(r'' - r''')\overline{n}(r'')\right] w'(r'') w'(r''') e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\Bigg).
\end{aligned} \qquad (B9)$$

After regrouping terms,

$$\begin{aligned}
A\left\langle |F(k)|^2 \right\rangle &= \int\int d^3r d^3r' w(r) w(r') \left\langle n_g(r) n_g(r') \right\rangle e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \\
&- \left[\int d^3r' w(r') \frac{\overline{n}_s(r)}{n_s'} e^{i\mathbf{k}\cdot\mathbf{r}}\right] \left[\int d^3r' d^3r'' w(r) w'(r'') \left[\overline{n}(r)\overline{n}(r'')\left(1 + \xi(r - r'')\right) + \overline{n}(r)\delta(r - r'')\right] e^{i\mathbf{k}\cdot\mathbf{r}}\right]
\end{aligned}$$

$$- \left[\int d^3r w(r) \frac{\overline{n}_s(r)}{n_s'} e^{i\mathbf{k}\cdot\mathbf{r}}\right] \left[d^3r' d^3r'' w(r') w'(r'') \left[\overline{n}(r') \overline{n}(r'') (1 + \xi(r'' - r')) + \overline{n}(r') \delta(r' - r'')\right] e^{-i\mathbf{k}\cdot\mathbf{r}'}\right]$$
$$+ \left(\left[\int \int d^3r d^3r' w(r) w(r') \frac{1}{(n_s')^2} (\overline{n}_s(r) \overline{n}_s(r') + \overline{n}_s(r) \delta(r - r')) e^{i\mathbf{k}(\mathbf{r}-\mathbf{r}')}\right]\right.$$
$$\left. \times \left[\int \int d^3r'' d^3r''' w'(r'') w'(r''') \left[\overline{n}(r'') \overline{n}(r''') (1 + \xi(r'' - r''')) + \overline{n}(r'') \delta(r'' - r''')\right]\right]\right), \qquad (B10)$$

so,

$$A\left\langle |F(k)|^2 \right\rangle = \int \int d^3r d^3r' w(r) w(r') \left\langle n_g(r) n_g(r') \right\rangle e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}$$
$$- \left[\int d^3r' w(r') \frac{\overline{n}_s(r')}{n_s'} e^{-i\mathbf{k}\cdot\mathbf{r}}\right] \left[\left(\int d^3r'' \overline{n}(r'') w'(r'')\right) \left(\int d^3r \overline{n}(r) w(r) e^{i\mathbf{k}\cdot\mathbf{r}}\right)\right]$$
$$- \left[\int d^3r w(r) \frac{\overline{n}_s(r)}{n_s'} e^{i\mathbf{k}\cdot\mathbf{r}}\right] \left[\left(\int d^3r'' \overline{n}(r'') w'(r'')\right) \left(d^3r' \overline{n}(r') w(r') e^{-i\mathbf{k}\cdot\mathbf{r}''}\right)\right]$$
$$+ \left(\left[\int d^3r d^3r' w(r) w(r') \frac{1}{(n_s')^2} (\overline{n}_s(r) \overline{n}_s(r') + \overline{n}_s(r) \delta(r - r')) e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\right]\right.$$
$$\left. \times \left[\left(\int d^3r'' w'(r'') \overline{n}(r'')\right) \left(\int d^3r''' w'(r''') \overline{n}(r''')\right)\right]\right)$$
$$- \left[\int d^3r' w(r') \frac{\overline{n}_s(r')}{n_s'} e^{-i\mathbf{k}\cdot\mathbf{r}'}\right] \left[\int \int d^3r d^3r'' w(r) w'(r'') \left[\overline{n}(r) \overline{n}(r'') \xi(r - r') + \overline{n}(r) \delta(r - r'')\right] e^{i\mathbf{k}\cdot\mathbf{r}}\right]$$
$$- \left[\int d^3r w(r) \frac{\overline{n}_s(r)}{n_s'} e^{i\mathbf{k}\cdot\mathbf{r}}\right] \left[\int \int d^3r' d^3r'' w(r') w'(r'') \left[\overline{n}(r') \overline{n}(r'') \xi(r'' - r') + \overline{n}(r') \delta(r'' - r')\right] e^{-i\mathbf{k}\cdot\mathbf{r}'}\right]$$
$$+ \left(\left[\int \int d^3r d^3r' w(r) w(r') \frac{1}{(n_s')^2} (\overline{n}_s(r) \overline{n}_s(r') + \overline{n}_s(r) \delta(r - r')) e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\right]\right.$$
$$\left. \times \left[\int \int d^3r'' d^3r''' w'(r'') w'(r''') \left(\overline{n}(r'') \overline{n}(r''') \xi(r'' - r''') + \overline{n}(r'') \delta(r'' - r''')\right)\right]\right). \qquad (B11)$$

The first four terms in equation (B11) will together produce the RHS of equation (2.1.6) of Feldman, Kaiser & Peacock (1994), which is the expectation value of $|F(k)|^2$ if we knew α exactly. The remaining three terms thus form the bias correction. Hereafter, I only consider these three terms, and denote the correction to $< |F(k)|^2 >$ by $F_{bias}(k)$. Using

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

we obtain

$$AF_{bias}(k) =$$
$$- \left[\int d^3r' w(r') \frac{\overline{n}_s(r')}{n_s'} e^{-i\mathbf{k}\cdot\mathbf{r}'}\right] \left(\left[\frac{1}{(2\pi)^3} \int \int \int d^3k' d^3r d^3r'' w(r) w'(r'') \overline{n}(r) \overline{n}(r'') P(k') e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} e^{i\mathbf{k}'\cdot\mathbf{r}''}\right]\right.$$
$$\left. + \left[\int d^3r w(r) w'(r) \overline{n}(r) e^{i\mathbf{k}\cdot\mathbf{r}}\right]\right)$$
$$- \left[\int d^3r w(r) \frac{\overline{n}_s(r)}{n_s'} e^{i\mathbf{k}\cdot\mathbf{r}}\right] \left(\left[\frac{1}{(2\pi)^3} \int \int \int d^3k' d^3r' d^3r'' w(r') w'(r'') \overline{n}(r') \overline{n}(r'') P(k') e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}'} e^{-i\mathbf{k}\cdot\mathbf{r}''}\right]\right.$$

$$
\begin{aligned}
&+\left[\int d^3r' w(r')w'(r')\overline{n}(r')e^{-i\mathbf{k}\cdot\mathbf{r}'}\right]\Big) \\
+&\left[\int\int d^3r d^3r' w(r)w(r')\frac{1}{(n_s')^2}\left(\overline{n}_s(r)\overline{n}_s(r') + \overline{n}_s(r)\delta(r-r')\right)e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\right] \\
&\times\left(\left[\frac{1}{(2\pi)^3}\int\int\int d^3k' d^3r'' d^3r''' w'(r'')\, w'(r''')\, \overline{n}(r'')\overline{n}(r''')\, P(k')e^{-i\mathbf{k}'\cdot\mathbf{r}''}e^{i\mathbf{k}'\cdot\mathbf{r}'''}\right]\right. \\
&\left.+\left[\int d^3r'' w'(r'')^2\overline{n}(r'')\right]\right). \qquad (B12)
\end{aligned}
$$

Since the bias correction is only large at small wave numbers, I shall set $w(r) = w'(r)$ to make the formulae tractable. These two weight functions will be equal at some value of k. At other wave numbers, this will only be an approximation. In particular it will be a bad approximation at high wave numbers. Using this approximation and defining the following (in the notation of Feldman, Kaiser & Peacock 1994):

$$
\begin{aligned}
G(k) &= \frac{\int d^3r\overline{n}(r)w(r)e^{i\mathbf{k}\cdot\mathbf{r}}}{[\int d^3r\overline{n}(r)^2w(r)^2]^{\frac{1}{2}}}, \\
G(0) &= \frac{\int d^3r\overline{n}(r)w(r)}{[\int d^3r\overline{n}(r)^2w(r)^2]^{\frac{1}{2}}}, \\
\frac{G(k)}{G(0)} &= \frac{\int d^3r\overline{n}(r)w(r)e^{i\mathbf{k}\cdot\mathbf{r}}}{\int d^3r\overline{n}(r)w(r)} = \frac{\int d^3r\overline{n}_s(r)w(r)e^{i\mathbf{k}\cdot\mathbf{r}}}{\int d^3r\overline{n}_s(r)w(r)} \\
&= \int d^3r\frac{\overline{n}_s(r)}{n_s'}w(r)e^{i\mathbf{k}\cdot\mathbf{r}}, \\
S(k) &= (1+\alpha)\frac{\int d^3r\overline{n}(r)w(r)^2e^{i\mathbf{k}\cdot\mathbf{r}}}{\int d^3r\overline{n}(r)^2w(r)^2},
\end{aligned}
$$

we get

$$
\begin{aligned}
F_{bias} &= -\frac{G^\star(k)}{G(0)}\left(\left[\frac{1}{(2\pi)^3}\int d^3k'\hat{W}^\star(k'-k)\,\hat{W}(k')P(k')\right] + \frac{S(k)}{1+\alpha}\right) \\
&-\frac{G(k)}{G(0)}\left(\left[\frac{1}{(2\pi)^3}\int d^3k'\hat{W}(k'-k)\,\hat{W}^\star(k')P(k')\right] + \frac{S^\star(k)}{1+\alpha}\right) \\
&+\left[\frac{|G(k)|^2}{|G(0)|^2} + \left(\frac{\alpha}{1+\alpha}\right)\frac{S(0)}{|G(0)|^2}\right]\left(\left[\frac{1}{(2\pi)^3}\int d^3k'\left|\hat{W}(k')\right|^2P(k')\right] + \frac{S(0)}{1+\alpha}\right). \qquad (B13)
\end{aligned}
$$

I will now use the same simplification as used in Appendix A for the case of volume limited surveys in setting the first two terms of (B13) equal to the third term. This will only be valid at small wave numbers (we can see that at $k = 0$ the three terms are identical except for a small shot noise term from the random catalogue which is the second term in the first set of square brackets in the third term, and may be made very small by generating a large number of random points). We now arrive at the final approximate expression for the correction to the flux limited power spectrum,

$$F_{bias} = -\frac{1}{|G(0)|^2}\left(|G(k)|^2 + \left(\frac{\alpha}{1+\alpha}\right)S(0)\right)\left[\left(\frac{1}{V}\sum_{k'}\left|\hat{W}(k)'\right|^2 P(k')\right) + \frac{S(0)}{1+\alpha}\right]. \qquad (B14)$$