Topology and Luminosity Function of the PSCz Redshift Survey

Diplomarbeit
der
Fakultät für Physik
an der
Eberhard-Karls-Universität Tübingen

angefertigt am
Max-Planck-Institut für Astrophysik
Garching bei München

betreut von
Prof. Dr. Simon White
Prof. Dr. Hanns Ruder

vorgelegt von
Volker Springel
aus Backnang

Oktober 1996

To My Parents

For their support, patience, and encouragement

Der Versuch einer Allgemeinen Himmelstheorie scheint auf religiöse Vorurteile zu stoßen, da die Weltordnung in dieser Theorie aus rein mechanischen Gesetzmäßigkeiten abgeleitet wird; aber dies ist gerade ein Hinweis auf die Vollkommenheit der Schöpfung, da die allgemeinen Wirkungsgesetze der Materie selbst von Gott schon so geplant wurden, daß daraus die Ordnung und Schönheit des Kosmos erfließt.

Immanuel Kant
Allgemeine Naturgeschichte und
Theorie des Himmels (1755)

Contents

[1] Ich habe diese Diplomarbeit in englisch verfaßt, um allen Mitgliedern der PSCz und Virgo Arbeitsgruppen die Lektüre zu ermöglichen. Da Englisch Publikationssprache in Physik ist, kann auf diese Weise jeder Physiker diese Arbeit verstehen. Für diejenigen Leser, die weniger geübt im Englischen sind, gibt dieses Kapitel einen kurzen Überblick über den Inhalt meiner Arbeit.

Introduction

Beginning with the earliest astronomical surveys of galaxies it has been noticed that the distribution of galaxies in the Universe exhibits remarkable structure. The galaxies are far from being randomly distributed in space. Instead they show a clear tendency to form clusters of galaxies which themselves are arranged in even larger objects, sometimes in the form of superclusters, that extend over several per cent of the observable Universe. On the other hand large regions of space containing only very few galaxies, so-called voids, have been discovered as well.

Since deep redshift surveys of galaxies have become available the large-scale structure revealed by them has been visually assessed as hierarchical (Soneira & Peebles 1977), filamentary (Einasto, Klypin, Saar & Shandarin 1984, Haynes & Giovanelli 1986), wall-like (Geller & Huchra 1989) or cellular (de Lapperent, Geller & Huchra 1986, Broadhurst, Ellis, Koo & Szalay 1990), depending on whether the emphasis laid on prominent regions of high galaxy density or on voids.

In the last decades a number of standard statistical tools have been developed in order to quantify the clustering of galaxies. The most prominent of them are the two-point correlation function, the power spectrum of density fluctuations and the counts-in-cells analysis. These statistical measures are routinely applied to more and more powerful redshift surveys. It turned out that the clustering of galaxies is one of the most important constraints on models of galaxy formation. Any successful theoretical model of large-scale structure and galaxy formation must be able to match the observed clustering data and also exhibit features that are as prominent as the Great Wall (Geller & Huchra 1989) or the Boötes void (Kirshner, Oemler, Schechter & Shectman 1981).

The widely used two-point correlation function and the power spectrum describe the properties of the galaxy distribution only to a limited extent. A full description would involve a hierarchy of three-, four- and higher correlation functions or, alternatively, information on the phase correlations among the different Fourier modes of the density field. Since these quantities are difficult to obtain directly it is desirable to complement the two-point correlation function or the power spectrum with other statistics that are sensitive to higher order correlations.

An important characteristic of the galaxy distribution is its topology. One can hope that a quantitative measure of it can provide additional insight into the process of structure formation and can put constraints on theoretical models.

In the recent past a number of topological measures have been proposed, among them are percolation (Klypin 1988), level-crossing statistics (Ryden 1988), fractals (Mandelbrot 1982), genus statistics (Gott, Melott & Dickinson 1986) and Minkowski functionals (Mecke, Buchert & Wagner 1994, Kerscher et al. 1996). Of course, topology is a major branch of mathematics and I can obviously not discuss all aspects of it in this work. For this reason I will focus on the genus statistics which was first proposed by Gott et al. (1986) and has become a widely accepted statistical tool in cosmology since then.

The genus statistics is particularly interesting since there are useful relations that link it to the traditional measures of clustering. Furthermore it can be used to test an important prediction of the standard models for structure formation in the Universe.

These models hold that galaxies and larger structures develop from initially small amplitude density perturbations in a homogeneous expanding Universe. The density fluctuations are gravitationally unstable and grow in strength until they eventually become nonlinear and collapse to the structures we see today. Most theoretical models assume that the initial density fluctuations had random phases, that is the phases of different Fourier components of the density fluctuation spectrum had been uncorrelated. A random phase distribution is necessarily a Gaussian random field. In inflationary cosmologies the random phase initial conditions can arise naturally from quantum noise.

In such scenarios, the topology of the present galaxy distribution can be related to the topology of the initial density fluctuations. For that reason the genus statistics can be used to test directly the random phase hypothesis.

Departures of the measured topology from the random phase prediction would be evidence for the presence of phase correlations that might reflect non-Gaussian initial conditions. They might arise in alternative scenarios for galaxy formation like the explosive amplification (Ostriker & Cowie 1981) or cosmic string (Vilenkin 1985) models.

In this work I examine the topology of a new large redshift survey, the Point Source Catalogue Redshift Survey (PSCz). This redshift survey of infrared selected galaxies is ideally suited for a topological analysis because it covers almost the full sky and has, with approximately 15000 galaxies, considerably larger statistical power than any previously examined sample. One particular objective of the topological analysis is a test of the random phase hypothesis and a quantification of deviations from the theoretical Gaussian random phase genus. These deviations result from higher order statistics of the density field and provide information that has the potential to constrain theoretical models for structure formation decisively. In particular, the topological analysis probes global aspects of the density field that are not revealed by standard measures like the power spectrum or the two-point function, but which are sensitive to the connectedness and morphology of the structure in the Universe. For example, galaxy distributions that are wall-like, bubble-like or filamentary would all lead to different signatures in the genus statistics. Thus the topology has potentially strong discriminative power and might be used to rule out or support certain models for structure formation.

As an illustration of this discriminative power I will compare the genus statistics of the PSCz survey with very large N-body simulations of three variants of the cold dark matter cosmogony. Interestingly the genus statistics clearly favours one of these models and essentially rules out the standard cold dark matter model.

The topological analysis of the PSCz survey is based on a reconstruction of the mass density field of the Universe. In order to be able to do that an accurate understanding of the mean number density of galaxies in the survey is necessary. Commonly this is described by the so-called selection function which is tightly related to the luminosity function of the galaxies in the catalogue. The entire part II of this study is devoted to an application of various statistical methods in order to determine the far-infrared luminosity function of the galaxies in the PSCz sample. In contrast to optically selected galaxies it turns out that the far-infrared luminosity function exhibits a high rate of evolution. For the sake of an accurate selection function this evolution, which is an interesting problem in its own right, is investigated in detail as well.

Another aspect of the reconstruction of the density field is that this procedure essentially results in maps of the local Universe around our own galaxy. It will therefore be interesting to see whether the features in these maps can be identified with the largest known superclusters and voids.

This work is organized as follows. The topological analysis of the PSCz redshift survey is discussed in part I, whereas I summarize all the aspects concerning selection function, luminosity function and evolution in a separate part II. The selection function represents the essential link between these parts.

With respect to the first part, I will begin with an introduction to the theoretical background of the genus statistics in chapter 1. In chapter 2 the details of my specific implementation of the genus statistics will be further specified. In particular, smoothing and sampling considerations and a treatment of the shot noise influence is given here. Chapter 3 presents and discusses results for the genus statistics and derived statistical measures for the PSCz survey. Here some of the cosmographic aspects of the reconstructed density field are illuminated as well.

The comparison with N-body simulations is treated in chapters 4 and 5. I first directly compare the PSCz survey with three N-body simulations of cold dark matter universes and then perform a thorough test of the formalism applied in the analysis of the PSCz survey by means of a mock observation of a N-body universe.

In part II of this work I start out in chapter 6 with a discussion of the theory of a number of statistical estimators for the determination of the selection function, the luminosity function, the density distribution and the evolution of the PSCz galaxies. This presentation is supplemented with computational details and with a discussion of the spectral energy distribution of the PSCz galaxies in chapter 7. Finally I discuss in chapter 8 results obtained for the luminosity function and related properties, and I conclude and summarize thereafter.

Some reference material can be found in the appendix. In particular, appendix A gives a short review of standard equations in cosmology that are used in this work and also establishes some conventions for Fourier transformations and the like.

Part I

Topology

Chapter 1

Genus statistics

1.1 The topology of isodensity surfaces

I look for a quantitative measure for the topology of isodensity contours of a smoothed mass density field in three-dimensional space.

Mathematically the genus g_{s} of a three-dimensional surface characterizes its global topology by counting its number of holes and isolated regions. The genus number g_{s} of the surface is defined as

$$g_{\mathrm{s}} = (\text{number of holes}) - (\text{number of isolated regions}) + 1, \tag{1.1}$$

where *hole* means a hole like that in a doughnut. For example, a torus has $g_{\mathrm{s}} = 1$ since there is one hole and one connected region. A sphere on the other side is characterized by $g_{\mathrm{s}} = 0$ and a pretzel (three holes) by $g_{\mathrm{s}} = 3$. The genus represents a measure of the connectedness of the surface, not of its geometry. It does not change under continuous deformations of the surface.

Gott et al. (1986) were the first who suggested using the genus as a quantitative tool to measure the topology of the mass distribution in the Universe. For that purpose they proposed to employ the Gauss-Bonnet theorem that relates the genus number g_{s} to the integrated Gaussian curvature of the surface. It states

$$\frac{1}{4\pi}\int \kappa \,\mathrm{d}A = 1 - g_{\mathrm{s}}, \tag{1.2}$$

where

$$\kappa = \frac{1}{r_1 r_2} \tag{1.3}$$

is the local Gaussian curvature of the surface and the integral extends over the entire contour. Later on it will be seen that it is this relation which makes the genus measurement feasible in practice.

The quantity

$$\chi = 2(1 - g_{\mathrm{s}}) \tag{1.4}$$

is the Euler characteristic of the surface. From a more general perspective the Euler characteristic and hence the genus is equivalent to one of the four Minkowski functionals (Mecke et al. 1994) that exist in three-dimensional space. The Minkowski functionals are constructed under the requirement of certain invariance properties (e.g. translational and rotational invariance) as general measures for the topological characterization of subsets of space. By looking at the volume above a certain threshold value a so-called excursion set is examined. The other three Minkowski functionals possible in three-dimensional space are the mean curvature, the volume and the surface area of the excursion set.

Equation (1.2) reveals that the definition of g_{S} is not ideal for my purposes, since the left hand side is proportional to volume for an ergodic universe whereas g_{S} is not due to the additional 1 in its definition (1.1). Following Gott et al. (1986) I will therefore simplify the definition of the genus by calling

$$G = -\frac{1}{4\pi}\int \kappa \,\mathrm{d}A = (\text{number of holes}) - (\text{number of isolated regions}) \tag{1.5}$$

the genus from now on. In principle this is simply one half of the Euler characteristic, nevertheless I will stick to the usual convention and call this quantity *genus* although the definition of the mathematicians is more like that of equation (1.1).

Since G is expected to be proportional to volume in a homogeneous universe it is most interesting to normalize G to unit volume. I will then call $g = \frac{G}{V}$ the genus per unit volume or simply the genus. Since g depends on the threshold value used to construct the isodensity contour one in fact gets a genus curve as a function of this threshold value.

The genus curve contains information on the topology and it will be of prime interest how this curve can be interpreted. A detailed discussion of the expected genus curve and the meaning of possible deviations from it is given in section 1.4.

I already described the surfaces I want to examine as the isodensity surfaces of the mass density field of the Universe. This approach involves a smoothing of the observed discrete galaxy distribution on a scale λ in order to reconstruct the mass density field. Hence the genus curve will also depend on this smoothing length.

It should be noted however, that a smoothing procedure is not the only possible way to derive surfaces from a discrete point distribution that contain information on the topology. For example, Mecke et al. (1994) advocate an Eulerian ball model where each galaxy site is decorated with a ball of radius r and the topology of the union set of these balls is then examined as a function of the ball radius.

In this work I will focus on the approach that involves a reconstruction of the mass density field. This has the valuable advantage of being tightly related to more traditional measures for large-scale structure, like power spectrum and correlation functions. Furthermore it does not require a volume limited catalogue and hence allows a full usage of *all* the sources in the flux limited PSCz survey. This reduces the statistical uncertainty in the sampling of the galaxy field considerably. The Eulerian ball model on the other hand requires a uniform sampling throughout the survey volume which makes it necessary to eliminate a large fraction of the sources in order to volume limit the catalogue. This model is further troubled by the fact that it scales in a non-trivial, so far unknown way with the mean source density.

Given a smoothed density field $\hat{\delta}(\mathbf{x})$ and a threshold value $\delta_{\rm t}$ one can in principle compute the genus $g = G/V$ by means of equation (1.5). How this is actually done will be explained in section 1.2. To construct the genus curve one has to repeat this calculation for a number of threshold values. In principle the result could then be displayed by plotting the function $g = g(\delta_{\rm t})$. But this has the disadvantage that a comparison of genus curves for different smoothing lengths would be quite difficult since the variance of $\hat{\delta}(\mathbf{x})$ and hence the scale of $\delta_{\rm t}$ changes strongly with the smoothing length. For that reason it is better to parameterize the genus curve by the fraction f of the volume above the threshold value or by the quantity

$$\nu = \sqrt{2}\,\mathrm{erf}^{-1}(1 - 2f) \tag{1.6}$$

derived from it. Here erf^{-1} denotes the inverse of the error function $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2} \mathrm{d}t$. I will stick to the usual convention and present my results for the genus curves in the form $g = g(\nu)$.

The definition (1.6) of ν seems slightly odd at first sight, but in fact its motivation is easy to understand if one considers a Gaussian random field. Here the one-point probability function is a Gaussian with dispersion σ. Then the volume fraction above threshold is

$$f = \frac{1}{\sqrt{2\pi}\sigma} \int_{\delta_{\rm t}}^{\infty} \mathrm{e}^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{\nu}^{\infty} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x \tag{1.7}$$

for a threshold value $\delta_{\rm t}$. Equation (1.6) is just the inversion of (1.7) and $\nu = \delta_{\rm t}/\sigma$ measures the threshold value in units of standard deviations.

There is another important reason why it makes sense that the genus curve is plotted as a function of volume fraction. In the reconstruction of the mass density field I will assume that the PSCz galaxies trace the mass exactly. Hence the galaxy density is assumed to be simply proportional to the underlying mass density. However, the galaxy density could be biased with respect to the mass density field leading to a non-trivial relation between the two fields. But biased galaxy formation and moderate dynamical evolution both tend to preserve a monotonic relation between initial mass density and final galaxy density, so a contour defined by volume fraction is the same for the biased and unbiased distributions. Consequently the topology will be the same as well.

Since I will identify the contours *after* smoothing this correspondence does not necessarily hold exactly anymore, but one can expect that contours defined by volume fraction will be very nearly at the same place for the biased galaxy distribution and the underlying mass density field. Thus, by studying the genus curves as a function of volume fraction one can remove most of the uncertainty due to the biasing scheme.

1.2 Computation of the genus curve

So far I have not discussed how the genus curve can be calculated in practice where the density field is derived from a redshift survey or a computer simulation. Clearly, equation (1.5) would be of no practical use if one had to actually carry out the integration over the contour. Instead, one can make use of the fact that any surface can be approximated by a network of polygonal faces. Since this procedure does not change the global topology of the surface the genus is not altered by this replacement.

For the polygonal network one finds that the curvature is compressed into Dirac δ-functions located at the vertices. The curvature at the edges or on the faces is zero since at least one of the curvature radii is infinite there. Parallel transport arguments (Gott et al. 1986) can be used to show that these δ-functions are normalized to the angle deficit

$$D_i = 2\pi - \sum_j \alpha_{ij} \tag{1.8}$$

found at the particular vertex. Here α_{ij} is the angle between a pair j of edges which meet at vertex i. The genus G can then simply be evaluated by summing up all the angle deficits D_i found at the vertices of the polygonal network that approximates the original surface.

A simple example might be helpful to illustrate this. Figure 1.1 shows a torus and its approximation by a rough polygonal network. The eight outer vertices all have angle deficit $\pi/2$, the eight inner vertices have $-\pi/2$ and all other vertices have angle deficit zero. Then $G = \frac{1}{4\pi}\sum_i D_i = 0$ correctly gives the genus of the torus.

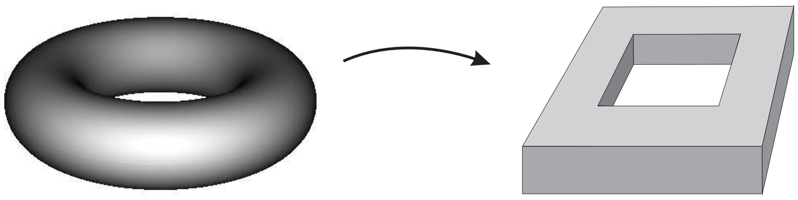


Figure 1.1: A torus and its approximation by a simple polygonal network with the same topology. The curvature is compressed into the 8 inner and 8 outer vertices of the right object.

The above idea can now be extended to a genus measurement of an isodensity contour of an arbitrary three-dimensional field. For that purpose one tessellates space with small volume elements and defines the density field on these voxels. Usually a lattice of small cubes is used for this purpose. The polygonal network that approximates an isodensity contour is then generated as the surface that separates the voxels above a prescribed threshold value from the voxels below it. A computer program can then quickly visit all the vertices and calculate the genus by summing up the appropriate angle deficits.

Weinberg (1988) has given a FORTRAN implementation called CONTOUR of this algorithm which was originally proposed by Gott et al. (1986). Based on this code I translated the algorithm to C and modified it somewhat to match my needs.

1.3 Theoretical genus curve

A number of authors (Doroshkevich 1970, Adler 1981, Hamilton, Gott & Weinberg 1986, Bardeen, Bond, Kaiser & Szalay 1986) have independently derived the expectation value of $g(\nu)$ for a Gaussian random field. It is given by

$$g(\nu) = N(1-\nu^2)\,\mathrm{e}^{-\nu^2/2} \tag{1.9}$$

with the amplitude

$$N = \frac{1}{(2\pi)^2}\left(\frac{-\hat{\xi}''(0)}{\hat{\xi}(0)}\right)^{3/2} = \frac{1}{(2\pi)^2}\left(\frac{\langle k^2\rangle}{3}\right)^{3/2}. \tag{1.10}$$

Here

$$\left\langle k^2 \right\rangle = \frac{\int k^2 \hat{P}(k)\,\mathrm{d}^3k}{\int \hat{P}(k)\,\mathrm{d}^3k} \tag{1.11}$$

is the second moment of the power spectrum $\hat{P}(k)$ of the smoothed density field and $\hat{\xi}(r)$ is the two-point correlation function of this field.

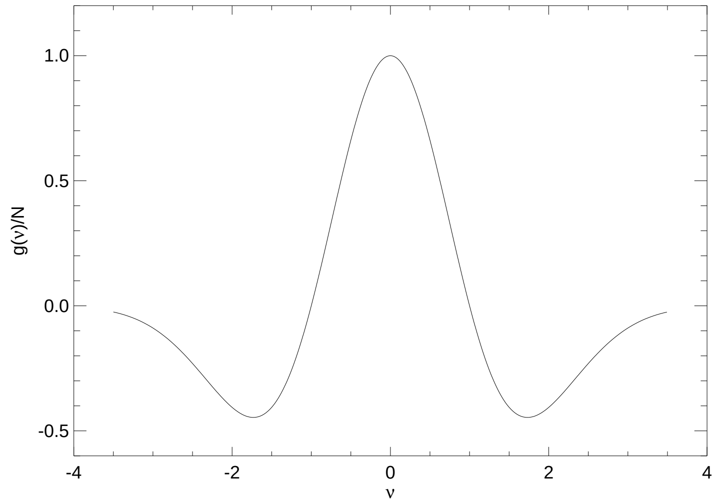


Figure 1.2: The universal w-shape of the genus curve for random phase Gaussian fields. Only the amplitude N depends on the shape of the power spectrum.

Interestingly the genus curve for a Gaussian random field exhibits a universal w-shape $\propto (1-\nu^2)\,\mathrm{e}^{-\nu^2/2}$ (see figure 1.2). Only the amplitude of the genus curve depends via $\langle k^2\rangle$ on the shape of the power spectrum, but it is independent of the normalization of $\hat{P}(k)$. The genus curve is completely symmetric around the mean density, which is a consequence of the perfect symmetry of a Gaussian field with respect to low and high density regions.

For the special case of a smoothed power law spectrum $\hat{P}(k) \propto k^n \, \mathrm{e}^{-k^2\lambda^2/2}$ the second moment evaluates to

$$\left\langle k^2 \right\rangle \lambda^2 = n + 3. \tag{1.12}$$

Here one therefore gets the genus curve

$$g(\nu) = \frac{1}{(2\pi)^2 \lambda^3} \left[\frac{n+3}{3}\right]^{3/2} (1-\nu^2)\, \mathrm{e}^{-\nu^2/2}. \tag{1.13}$$

In the previous section I have discussed how the genus curve can be measured in practice. For that purpose one has to resort to a tessellation of space with volume elements of finite size. Clearly, if these voxels are too large one cannot adequately approximate the topological features of the isodensity contour and will arrive at an erroneous genus result. Hence it is interesting to see whether the final voxel size introduces any error in the genus measurement.

Hamilton et al. (1986) have calculated the expectation value of $g(\nu)$ to second order in tessellation size r using a space filling tessellation of truncated octahedrons. Their result is

$$\begin{aligned} g(\nu) \quad &= \quad \frac{1}{(2\pi)^2} \left(\frac{\langle k^2 \rangle}{3}\right)^{3/2} \mathrm{e}^{-\nu^2/2} \left[1 - \nu^2 + \frac{r^2 \langle k^2 \rangle}{768} \right. \\ &\times \left. \left\{ 19(3 - 6\nu^2 + \nu^4) - 26 \frac{\hat{\xi}(0)\hat{\xi}^{(4)}(0)}{\hat{\xi}''(0)^2}(1-\nu^2) + \left(\frac{\hat{\xi}(0)\hat{\xi}^{(4)}(0)}{\hat{\xi}''(0)^2}\right)^2 \right\} \right]. \end{aligned} \tag{1.14}$$

The prefactor in front of the correction term in curly brackets is of order $\mathcal{O}(r^2/\lambda^2)$. As long as $r^2/\lambda^2 \ll 1$ the finite tessellation will not introduce any error in the genus curve.

Very recently Matsubara & Yokoyama (1996) have calculated the theoretical genus curves for a number of non-Gaussian distributions where the density contrast is characterized by a nontrivial function of Gaussian distributed random numbers. Examples for such fields include the lognormal distribution and the χ^2 distribution. The analytic formulae for these cases are rather lengthy, so I don't give them here.

In another study Matsubara & Suto (1996) have also been able to derive analytic expressions for the evolution of the genus curve of Gaussian random fields in the weakly nonlinear regime. However, the smoothing lengths that have to be used in this work to overcome the sparseness of the sampling are so large that the nonlinear regime can hardly be probed. In the near future when very large redshift surveys become available this regime might become extremely interesting, though.

1.4 Interpretation of the genus statistics

In the discussion of topological characteristics of the galaxy distribution its connectivity plays a primary role. If one imagines that the galaxy distribution is like isolated lumps of matter in an empty or homogeneous background then one has a topology where the voids (the regions of low density) are connected whereas the matter is not. This situation is commonly called a *meatball* topology, since it resembles the isolated meatballs in spaghetti sauce.

Just the opposite situation would be a matter distribution where isolated voids are surrounded on all sides by high density regions. Then the matter is connected in the form of a cell structure and the voids appear as isolated bubbles. For this reason this situation is called a *bubble* topology.

Now one should not think that these two situations mutually exclude each other and that either the matter or the voids are connected but not both. In fact, in three dimensions there is a third possibility which corresponds to the topology of a *sponge*. Here the matter and the voids are connected and surround each other in a mutually interlocking way.

Gott et al. (1986) pointed out that the topology of the mean density contour of a Gaussian random field should be that of a sponge. Because a Gaussian random field is symmetric with respect to low and high density fluctuations one may naturally expect that the median density contour is negatively curved (and consequently has positive genus). Due to the symmetry the radii of curvature tend to point in opposite directions, one towards the high density side and one towards the low density side. Negative and positive fluctuations should be topologically indistinguishable for these fields.

The theoretical result (1.9) for the random phase genus curve is indeed symmetric around the mean density. When the threshold value is increased from large negative values to positive values one first begins to pick up isolated underdense peaks (compare figure 1.2). Since isolated regions contribute negatively the genus is negative. At $\nu = -\sqrt{3}$ corresponding to a low-density volume fraction of 4.1% the genus reaches its negative maximum. If the threshold value is increased further the underdense regions start to merge, holes form and the number of isolated structures decays. The sponge-like region begins at $\nu = -1$ where the topology changes to positive genus and becomes more and more dominated by holes and tunnels forming in fewer and fewer connected regions. At the maximum of the genus curve at $\nu = 0$ the topology is that of a symmetric sponge with a multiply connected structure that exhibits a complex system of tunnels and holes. If the threshold value is then increased further, nothing new happens. Due to the symmetry properties of a Gaussian field one just walks the way back and ends up with a few isolated high density regions and a negative genus.

The genus curve for a Gaussian random field is particularly interesting because many theoretical models for the formation of galaxies and large-scale structure assume random phase fluctuations as initial conditions. For example, in the popular inflationary cosmogonies the initial density perturbations originate naturally as random phase quantum noise.

In the standard theory for structure formation in the Universe the large-scale structure is then the result of the growth of these fluctuations due to their gravitational instability. The growth starts after recombination but first the topology is not affected at all, because the fluctuations are all in the linear regime and grow with the same rate $\delta \propto a$, where a is the cosmological scale factor (Padmanabhan 1993). Hence the map of isodensity contours in comoving coordinates stays fixed. Only when the density contrast in a region reaches $\delta \approx 1$ the region decouples from the expansion of the Universe and starts to undergo nonlinear collapse, eventually forming the galaxies we see today.

At the present time the distribution of matter is nonlinear only on small scales where the density fluctuations have had enough time to grow into nonlinearity and undergo gravitational collapse. However, if the density is smoothed over a sufficiently large scale the remaining fluctuations are still in the linear regime. Hence their topology can be expected to directly mimic the topology of the initial conditions.

By studying the topology of the smoothed density field one can therefore directly examine the topology of the initial density fluctuations that acted as seeds for structure formation. In particular, the Gaussian random phase hypothesis can be tested.

With N-body simulations it was repeatedly shown (Weinberg, Gott & Melott 1987, Melott, Weinberg & Gott 1988, Gott et al. 1989) that this picture is right and that the initial conditions can be very well recovered for smoothing lengths larger than two to three times the mass correlation length r_0 defined by $\xi(r_0) = 1$.

Possible deviations from Gaussianity should show up as departures from the random phase curve. This curve acts therefore as a benchmark to which all genus results obtained from observations can be compared. Such deviations could provide evidence for non-Gaussian perturbations, e.g. cosmic strings, explosions, or for strongly non-linear dynamical evolution.

It is now an important question whether the genus curve can distinguish between various kinds of non-Gaussian distributions. This is a non-trivial problem since for most distributions one can expect that there are isolated peaks at high and low density and one can also expect that they will exhibit a sponge-like behaviour for some range of intermediate threshold values. Similarly, if a density field with non-Gaussian features of some typical coherence scale $r_{\rm c}$ (for example bubbles of radius $r_{\rm c}$) is smoothed over a scale larger than $r_{\rm c}$ the remaining fluctuations will tend to behave like a Gaussian field.

It is therefore not straightforward to predict the genus curves for different models of density distributions. But it can be expected that the genus can discriminate between them if the entire genus curve is used as a diagnostic tool. In order to develop some understanding of the signature of different types of density fields in the genus curve various authors (e.g. Weinberg et al. 1987, Melott et al. 1988, Park, Gott, Melott & Karachentsev 1992b, Beaky, Scherrer & Villumsen 1992, Gott, Cen & Ostriker 1996, Matsubara & Yokoyama 1996) have examined non-Gaussian toy models or N-body simulations of structure formation.

Perhaps the most elementary distribution one can imagine are randomly placed points. After smoothing with a sufficiently large smoothing length one expects to see the topology of a white noise power spectrum. Weinberg et al. (1987) have shown that for a smoothing length

$$\lambda \geq d \tag{1.15}$$

the theoretical prediction for white noise is excellently fit. Here d is the mean interparticle separation of the randomly distributed points. If λ is considerably less than d the genus curve is shifted to the left and appears tilted because the sampling is now so sparse that single points act as isolated lumps of matter. This is called a meatball shift since it corresponds to a situation where the isolated high density regions are more pronounced than in a random phase model. Slight meatball shifts are also obtained for extreme biasing scenarios or for toy distributions that emphasize isolated clusters.

Toy models of bubble distributions, on the other hand, exhibit a genus curve with a shift of the maximum to the right and the region with $g < 0$ on the negative threshold side is more pronounced. This is called a bubble or *Swiss-cheese* shift. This shift, accompanied with other asymmetries of varying degree, also results for toy distributions of isolated filaments, intersecting sheets or filamentary nets.

Experiments with numerical N-body simulations have shown that the genus curve can recover the initial spectrum of the mass density fluctuations if it is smoothed by more than about two times the galaxy correlation length. Cold dark matter (CDM) models show even at smaller scales still Gaussian behaviour, but their amplitude appears somewhat less than expected on the basis of the power spectrum. Models with hot dark matter (HDM), however, give rise to a meatball shift. This shift means that in the HDM model one finds connected clusters at threshold values where the Gaussian distribution would be already disconnected. The CDM models, on the other hand, remain extremely Gaussian, even with biasing. Any observed deviation from this behaviour would be evidence against them.

Finally I should note that even when the shape of a genus curve is very well fitted by the random phase form this should not be taken as conclusive evidence that the underlying density field is a perfect Gaussian random field. I will demonstrate later on that some of the CDM cosmological N-body simulations exhibit genus curves with perfect random phase shapes but with an amplitude that does not match the amplitude predicted by the theoretical formulae (1.9) and (1.10). This must be attributed to phase correlations, or phrased differently, to non-vanishing higher order correlation functions, that develop in the weakly nonlinear regime. This drop in amplitude provides additional information that might be useful to constrain theoretical models with the help of observational data.

Chapter 2

Application to Observations

2.1 The PSCz redshift survey data

The Point Source Catalogue Redshift Survey (PSCz) of galaxies detected in the IRAS[1] point source catalogue (PSC) contains approximately 15000 galaxies over 84.1% of the sky (Saunders, Sutherland, Efstathiou & Tadros 1994, Saunders 1995). The major objective of the survey was to obtain a redshift for every galaxy in the PSC with an infrared flux above 0.6 Jy in the 60 μm band of the IRAS satellite. Particular care has been taken to arrive at a highly complete, statistical uniform sample with large sky coverage. These features make the PSCz an ideal tool to do all sorts of large-scale structure analysis.

The collaboration for PSCz consists of the scientists Will Saunders, Will Sutherland & George Efstathiou (Oxford University), Steve Maddox (Royal Greenwich Observatory), Michael Rowan-Robinson, Seb Oliver & Oliver Keeble (Imperial College, London), Richard McMahon (Institute of Astronomy, Cambridge), Carlos Frenk (University of Durham) and Simon White (Max-Planck-Institute for Astrophysics, Garching). Additionally a number of students are participating in the analysis. I am thankful to these people for kindly granting me access to the data before publication.

Smaller redshift surveys of IRAS galaxies, notably the QDOT survey (Kaiser, Efstathiou, Ellis & Frenk 1991, Lawrence et al. 1996) and the 1.2-Jy survey (Fisher et al. 1995), have already demonstrated how well IRAS galaxies can be used as tracers of the mass density field. As a result they have transformed our understanding of the dynamics of the local universe (Dekel et al. 1993).

The final sky coverage of PSCz reaches 84.1%, only excluding two small strips not covered by the IRAS satellite, a few patches contaminated by galactic cirrus and the galactic plane, defined here by a 100 μm background exceeding 25 MJy/sr. The excluded regions are coded in an angular mask as shown in figure 2.1.

In part II of this work I discuss in detail the far-infrared luminosity function of

[1]The *Infrared Astronomical Satellite* was developed and operated by the U.S. National Aeronautics and Space Administration (NASA), the Netherlands Agency for Aerospace Programs (NIVR), and the U.K. Science and Engineering Research Council (SERC).

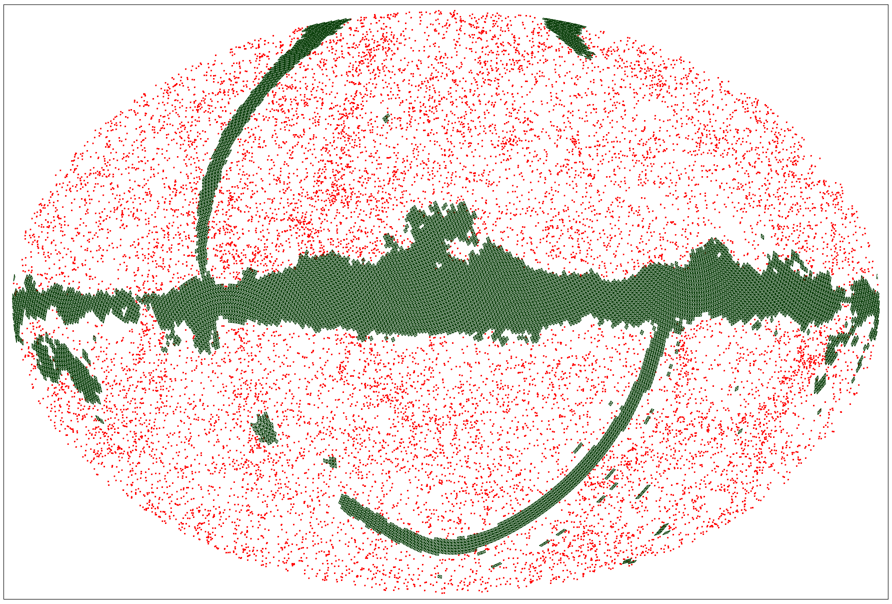


Figure 2.1: An equal area Aitoff projection of the galaxies in PSCz. The shaded regions constitute the area not covered by the survey.

the galaxies in the PSCz redshift survey. There I will also examine the evolution found in this population and show how an accurate selection function for the survey can be derived.

For the topological analysis an accurate determination of the selection function $S(z)$ is essential in order to reconstruct the mass density field. For this purpose I will adopt my best estimate for $S(z)$ as given in equation (8.2). Note that in principal only the shape of $S(z)$ is relevant for the subsequent analysis.

Apart from a transformation of the measured heliocentric redshifts to the barycenter of the Local Group (de Vaucouleurs, de Vaucouleurs, Corwin & Buta 1991) according to

$$v' = v + 300\,\frac{\mathrm{km}}{\mathrm{s}}\,\sin(l)\cos(b), \tag{2.1}$$

where (l, b) are the galactic coordinates of the source, no further correction for peculiar velocities is made. Instead, the redshifts $z = v'/c$ are directly used to infer a distance[2] r assuming an Einstein-de-Sitter cosmology:

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

My results should not be sensitive to this choice of background cosmology since the

[2] The length unit I use is h^{-1}Mpc where h parameterizes the Hubble constant $H_0 = 100\,h\,\mathrm{km/sec}$.

PSCz sample is cosmologically quite local. This feature will also become apparent when I examine in section 8.3 the influence of different background cosmologies on the estimates of evolution in the IRAS luminosity function.

Similarly, the redshift space distortions introduced by neglecting the peculiar velocities should not have a significant influence on the genus measurement because they have been shown theoretically and with N-body simulations to affect the genus curve only very slightly. For example, Matsubara (1996) shows analytically that the three-dimensional genus curve has the same shape in real and in redshift space and that its amplitude is practically unchanged for $0 \leq \Omega_0 b^{-5/3} \leq 1$, where b is the bias parameter.

All the analysis of this study is done with the first science release of the PSCz survey. Its completeness reaches about 98% with 259 redshifts missing at the present time (Saunders 1996). My results are then based on a catalogue with a total number of 14505 galaxies. Future releases of the survey will be further improved by increasing the completeness and by reobserving a small number of redshifts that are considered uncertain. However, the results of this work can be expected to change only a little under these revisions.

2.2 Smoothing procedure

As will be extensively discussed in section 6.2 the discrete point distribution

$$m(\mathbf{r}) = \sum \delta^{(\mathrm{D})}(\mathbf{r} - \mathbf{r}_i) \tag{2.3}$$

of observed galaxies can be transformed in an estimate for the underlying mass density field by weighting them with the inverse of the selection function:

$$\rho(\mathbf{r}) \propto \frac{m(\mathbf{r})}{S(r)}. \tag{2.4}$$

Here the assumption of a universal luminosity function is made and it is also assumed that IRAS galaxies trace the mass in an unbiased way.

With the help of a normalized window function $W(\mathbf{x})$ one can define a smoothed density field $\hat{\delta}(\mathbf{x})$ by the convolution

$$\hat{\rho}(\mathbf{x}) = \int W(\mathbf{x} - \mathbf{x}')\rho(\mathbf{x}')\,\mathrm{d}\mathbf{x}'\,. \tag{2.5}$$

This smoothed field represents an unbiased estimate for the underlying fluctuation field (see also section 2.6).

In practice this convolution can be done efficiently in Fourier space where it translates into a simple multiplication:

$$\hat{\rho}_{\mathbf{k}} = V\,W_{\mathbf{k}}\,\rho_{\mathbf{k}}\,. \tag{2.6}$$

The power spectrum

$$\hat{P}_{\mathbf{k}} = V^2\,|W_{\mathbf{k}}|^2 P_{\mathbf{k}} \tag{2.7}$$

of the smoothed field is then simply given by the spectrum of the original field times the magnitude of the transform of the window function.

In the following I will use a Gaussian window function of the form

$$W(\mathbf{x}) = \frac{1}{\pi^{3/2}\lambda^3}\,\mathrm{e}^{-\frac{\mathbf{x}^2}{\lambda^2}}. \tag{2.8}$$

Of course, different choices of smoothing kernels are possible, but the Gaussian filter is especially convenient because its Fourier transform

$$W_{\mathbf{k}} = \frac{1}{V}\,\mathrm{e}^{-\frac{\mathbf{k}^2\lambda^2}{4}} \tag{2.9}$$

is again a Gaussian. With this choice for $W(\mathbf{x})$ the power spectrum of the smoothed field takes the form

$$\hat{P}(k) = P(k)\,\mathrm{e}^{-\frac{k^2\lambda^2}{2}}. \tag{2.10}$$

Later on the variance σ^2 of the smoothed density field will be of interest, too. It is given by the correlation function at the origin and can be expressed in terms of the power spectrum:

$$\sigma^2 = \left\langle \hat{\delta}(\mathbf{x})^2 \right\rangle = \hat{\xi}(0) = \int \hat{P}(k)\,\mathrm{d}^3k = \int P(k)\,\mathrm{e}^{-\frac{k^2\lambda^2}{2}}\,\mathrm{d}^3k. \tag{2.11}$$

If $P(k)$ can be approximated by a power law of the form

$$P(k) = P_0\left(\frac{k}{k_0}\right)^n \tag{2.12}$$

the variance becomes explicitly

$$\sigma^2 = \frac{1}{4\pi}\Gamma\left(\frac{n+3}{2}\right)\frac{1}{(k_0\lambda)^n}\frac{P_0}{\lambda^3}. \tag{2.13}$$

In N-body simulations one usually works with periodic boundary conditions so the smoothing can be easily and rapidly done with Fourier techniques, where one first Fourier transforms the density field, then multiplies it by the Fourier transformed smoothing kernel and finally Fourier transforms back to obtain the smoothed field.

However, in a practical application to an observational data set one deals with a non-periodic field. Furthermore a typical redshift survey covers only some fraction of the sky, which leads to an odd shaped three-dimensional survey volume that is far from being cubical. It is therefore necessary to specify how the boundaries presented to the observer in this way should be treated. The genus measuring algorithm CONTOUR itself has no problem at all with fancy survey volumes. It can be simply adjusted such that only vertices entirely surrounded by voxels inside the survey volume are used to calculate the genus curve.

But the smoothing procedure affects regions near the survey boundary and cells well inside the survey volume differentially. If the outer regions are just left empty the smoothing would lead to a lower average density near the boundary since some of the mass leaks out by the smoothing.

In order to avoid this one can surround the survey volume by some form of padding that should be chosen such that the contamination of the density field inside the survey volume is minimal with respect to its topological features.

Paddings that come to mind are a homogeneous average density, Poisson noise of density equal to the average density or a Gaussian random field with a similar power spectrum than that of the true density field. The Poisson noise scheme for example was used by Gott et al. (1989).

Melott & Dominik (1993) show in a systematic study that a method that is superior to the three padding schemes mentioned above is obtained if the smoothing is performed such that one sums only over cells inside the survey volume. In this scheme the smoothing kernel is effectively renormalized for each cell to the survey volume visible from it. As Melott & Dominik (1993) show this procedure minimizes the loss or distortion of topological information due to the presence of the boundary.

This smoothing procedure can be defined by

$$\hat{\rho}(\mathbf{r}) = \frac{\int_{V_\mathrm{S}} W(\mathbf{r}-\mathbf{r}')\rho(\mathbf{r}')\,\mathrm{d}\mathbf{r}'}{\int_{V_\mathrm{S}} W(\mathbf{r}-\mathbf{r}'')\,\mathrm{d}\mathbf{r}''}, \tag{2.14}$$

where the integrations extend only over the survey volume V_S and $W(\mathbf{r})$ is the Gaussian smoothing kernel of equation (2.8). One can now define a mask field $M(\mathbf{r})$ which has value 1 for $\mathbf{r} \in V_\mathrm{S}$ and value 0 otherwise. Then one might write

$$\hat{\rho}(\mathbf{r}) = \frac{\int W(\mathbf{r}-\mathbf{r}')\rho(\mathbf{r}')\,\mathrm{d}\mathbf{r}'}{\int W(\mathbf{r}-\mathbf{r}'')M(\mathbf{r}'')\,\mathrm{d}\mathbf{r}''}, \tag{2.15}$$

where the integrations now extend over all space and $\rho(\mathbf{r})$ has zero padding outside the survey volume. The numerator can be recognized as the smoothed density field and the denominator as the smoothed mask field when no boundaries are taken into account. These smoothings are amenable to Fourier techniques.

The smoothed density field can therefore be derived with a ratio method. The density field $\rho(\mathbf{r})$ is smoothed with zero padding outside and then divided by the smoothed mask field. The effective smoothing kernel $w(\mathbf{r}, \mathbf{r}')$ for point $\mathbf{r}$ is therefore given by

$$w(\mathbf{r}, \mathbf{r}') = \frac{\int W(\mathbf{r}-\mathbf{r}')}{\int W(\mathbf{r}-\mathbf{r}'')M(\mathbf{r}'')\,\mathrm{d}\mathbf{r}''}. \tag{2.16}$$

Numerical Fourier transformations on a computer are inevitably discrete and periodic in nature. In order to use them for the non-periodic survey volume one has to embed it in a cube that is large enough to avoid crosstalk between edges of the survey volume that are close to the boundaries of the box. For that purpose the wrap-around distance between the edges should be larger than about 3λ. Given this constraint the actual choice of dimensions represents a compromise between the fineness of the sampling and the necessary computing resources in the form of memory and time.

In my application to the PSCz survey I have chosen 128^3 arrays throughout with double precision arithmetic. Such fields already consume 16 Mbyte of memory but modern workstations like the IBM RS/6000 at the MPA are still able to deal with them without (major) problems.

For a given smoothing length λ I have chosen the physical voxel size $d = \lambda/6$. This ensures that all frequencies of the smoothed field are adequately sampled and also guarantees that the finite voxel size effects in the genus curve measurement are negligible.

The particular choice of $d = \lambda/6$ is also motivated by the fact that then the 128^3 array is optimally filled by the PSCz survey. For every smoothing length I will derive in the following section a maximal distance $R_{\rm max}$ that defines the survey volume which is used for the genus statistics. With $d = \lambda/6$ one has $2\,R_{\rm max}/d \leq 110$, so the 128^3 is used optimally without running in problems with the periodic nature of the Fourier transform. After smoothing a sphere of radius $R_{\rm max}$ is carved out of the density field and used for the topological analysis. Note that there is no boundary influence due to the surface of this sphere since before smoothing the galaxy field is also set up outside this final region.

2.3 Sampling considerations

In its present formulation the genus statistic can only be applied to a continuous density field, which is derived by smoothing the discrete galaxy distribution. This approach makes use of the assumption that the galaxies trace the mass in a Poisson point process, that is, the probability density to find a galaxy at some point is proportional to the mass density at that location.

If this assumption holds the smoothed galaxy field gives an unbiased estimate for the mass density field. Clearly, the error in this estimate will depend decisively on the number of tracers available per smoothing volume. It is therefore an important question how large this smoothing length has to be in order to arrive at a genus curve that really describes the correct genus of the underlying mass density field.

As it turns out there are two aspects to the problem. First, the sampling of the density field must be dense enough to make discreteness effects in the genus curve negligible. A useful rule of thumb (Weinberg et al. 1987) that has been tested with numerical simulations by various authors is

$$\lambda \geq d = \frac{1}{S(r)^{1/3}}\,, \tag{2.17}$$

where d is the mean interparticle separation at distance r. As long as the smoothing length is larger than the mean interparticle separation the genus curve is not affected by discreteness effects. If the smoothing length is smaller than d one risks artificially introducing a meatball bias by identifying lonely tracers as isolated clusters.

Because the galaxy density in a flux limited sample falls off quickly with distance there will be a maximal radius $R_{\rm max}$ for each smoothing length λ up to which the

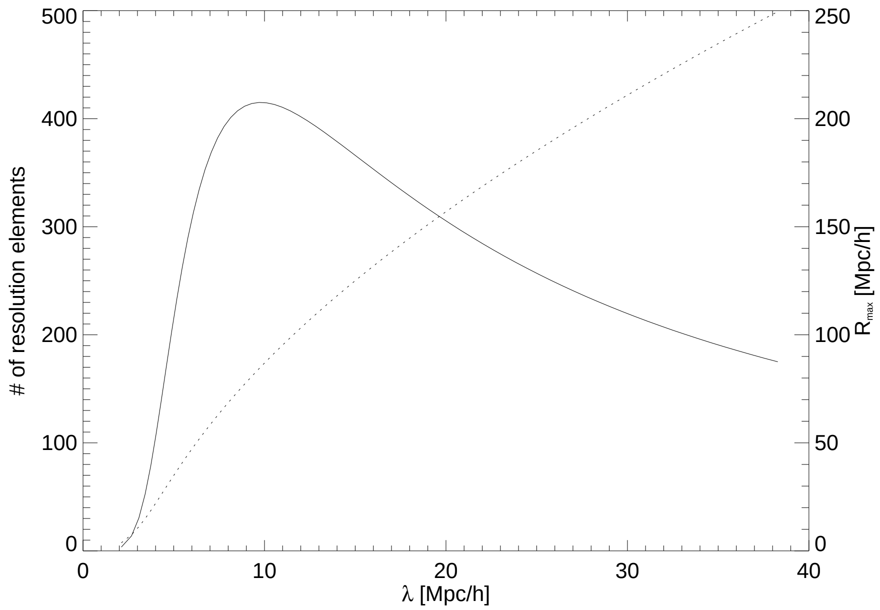


Figure 2.2: The number of resolution elements (solid) for the PSCz survey when the maximal survey volume is used. Also shown is the radius (dashed) of the usable survey volume.

sampling fulfills condition (2.17). This radius is implicitly defined by

$$\lambda = \frac{1}{S(R_{\max})^{1/3}}\,, \tag{2.18}$$

and will be used to delimit the usable survey volume $V_{\rm s}$. This choice ensures that at the far edge of the survey volume the sampling condition is just met, and in the remainder of the volume the sampling is denser.

Second, there is still shot noise in the estimate for the density field even when condition (2.17) is well satisfied. As I will show in section 2.7 this noise can lead to a systematic change of the genus amplitude. Previous studies that have used the genus amplitude directly to infer the second moment of the power spectrum have not corrected for this effect. I will postpone a quantitative estimate of the shot noise influence to section 2.7.

A useful way to predict the statistical power of a genus measurement is the notion of number of resolution elements. Because the smoothing extends over an effective volume

$$V_{\rm sm} = \pi^{3/2}\lambda^3 \tag{2.19}$$

the number of independent structures that can exist in the finite survey volume is

λ $[h^{-1}\text{Mpc}]$	$R_{\max}$ $[h^{-1}\text{Mpc}]$	V_{s} $[(h^{-1}\text{Mpc})^3]$	N_{res}	N_{gal}
5.00	34.92	1.50×10^5	215.5	2295
6.00	47.16	3.70×10^5	307.3	3550
7.00	58.35	7.00×10^5	366.3	4928
8.00	68.58	1.14×10^6	398.5	5909
10.00	86.91	2.31×10^6	415.3	7510
12.00	103.19	3.87×10^6	402.3	8775
14.00	118.04	5.79×10^6	379.2	9681
16.00	131.81	8.07×10^6	353.7	10356
20.00	156.96	1.36×10^7	305.8	11309
24.00	179.79	2.05×10^7	266.0	11968
28.00	200.97	2.86×10^7	233.9	12496
32.00	221.79	3.84×10^7	210.6	12833
40.00	263.44	6.44×10^7	180.7	13339
48.00	305.09	1.00×10^8	162.4	13684
50.00	315.50	1.11×10^8	158.9	13748
56.00	346.73	1.47×10^8	150.2	13924

Table 2.1: The smoothing lengths adopted for the topological analysis of the PSCz survey. Also shown are the resulting survey volume V_{s} and the number N_{res} of resolution elements. N_{gal} gives the number of galaxies inside the survey volume.

limited. This can be quantified by the number of resolution elements

$$N_{\text{res}} = \frac{V_{\text{s}}}{V_{\text{sm}}} = \frac{\omega R_{\max}^3}{3\pi^{3/2}\lambda^3}, \tag{2.20}$$

where ω is the solid angle covered by the redshift survey. If N_{res} is large one can expect that the survey volume represents a fair sample of the Universe in terms of the genus statistics.

Therefore the number N_{res} indicates the power of a data set used for topological analysis. With the QDOT survey Moore et al. (1992) reached a maximum of about $N_{\text{res}} = 80$ whereas the most powerful genus results so far came from Vogeley et al.'s (1994) treatment of the CfA survey. They reached $N_{\text{res}} = 260$ for their best subsample.

The PSCz redshift survey can provide still more resolution elements as is evidenced in figure 2.2 and table 2.1. It also provides a high number of resolution elements over an unprecedented wide range of smoothing scales. In particular there are more than 300 resolution elements in the range $6\,h^{-1}\text{Mpc} \leq \lambda \leq 20\,h^{-1}\text{Mpc}$. This wide dynamic range together with the large volume covered make it more powerful than all previously examined samples.

For the genus analysis I have examined smoothing lengths between $5\,h^{-1}$Mpc and $56\,h^{-1}$Mpc in approximately logarithmic spacing. In particular, all the smoothing lengths considered in the studies by Moore et al. (1992), Vogeley et al. (1994) and Rhoads, Gott & Postman (1994) are contained in my set of smoothing scales. Table 2.1 lists some relevant parameters for the different cases.

2.4 Estimation of errors

There are a number of possible sources of error in the genus curve measurement. The statistical uncertainty arises from the finite sampling of the galaxy distribution and because only a finite volume of the Universe is observed (cosmic variance). Apart from these statistical errors there might be further systematic effects that influence the result, for example, variations in the completeness of the redshift survey or limitations in the knowledge of the selection function.

In order to estimate the statistical uncertainty caused by finite sampling I employ the bootstrap resampling technique (Barrow, Bhavsar & Sonoda 1984) which has also been used in most previous applications of genus statistics to observational data sets (Gott et al. 1989, Park, Gott & da Costa 1992a, Vogeley et al. 1994, Rhoads et al. 1994).

I have assumed that the observed galaxy distribution represents a Poisson sampling of some underlying density field. The uncertainties that arise due to finite sampling could be easily quantified if one had many different sampling realizations of the density field. Since an ensemble of samplings cannot be obtained, the bootstrap resampling strategy uses the observed distribution itself as an approximation to the underlying density field. A bootstrap resampling then consists of a galaxy catalogue with the same number of objects as in the original catalogue but with galaxy positions that are randomly drawn from the set of positions in the original catalogue. That means that the N galaxies are randomly thrown on the N sites in the original sample. Note that the bootstrap sample will contain some positions more than once whereas some of the original positions are left empty. An estimate for the statistical uncertainty of a measurement due to finite sampling is then given by the variance among the measurements for many bootstrap resamplings (Ling, Frenk & Barrow 1986).

In the context of genus statistics the bootstrap method has proven to be a reliable and robust estimator of statistical errors, if anything it is somewhat conservative. Moore et al. (1992) have compared bootstrap errors for the genus statistic with errors obtained for different resamplings of full N-body simulations. Here the bootstrap errors turned out to be 10% larger than the 'true' statistical uncertainty.

As Vogeley et al. (1994) point out one should keep in mind, however, that the statistical properties of the resamplings are not fully identical to the original data. For example, positions in voids are never resampled and since the bootstrap resamplings contain a smaller number of independent positions than the original catalogue the average spacing between locations is higher. Therefore the bootstrap resamplings run earlier in trouble with the discreteness problem. Consequently, they are biased towards a meatball topology.

Thus the average genus curve of the bootstrap resamplings should not be considered a best estimate for the genus curve, although it exhibits less noise than the raw genus curve. Nevertheless, the scatter of the genus curve measurements among bootstrap resamplings still provides a good estimate for the statistical uncertainty of the observed genus curve.

Vogeley et al. (1994) also argue that some smoothing of the genus curve is appropriate since adjacent points on the genus curve are not independent and their relative values are strongly affected by noise. These authors tested a simple boxcar smoothing

$$g'(\nu_i) = \frac{1}{3}\left\{g(\nu_{i-1}) + g(\nu_i) + g(\nu_{i+1})\right\}, \tag{2.21}$$

with $\nu_{i\pm 1} = \nu_i \pm 0.1$, and showed that the boxcar averaged genus curve gave a better agreement (as measured by χ^2) with the true genus curve of N-body simulations than the raw genus curve or the averaged genus curve of bootstrap resamplings.

Unlike the bootstrap average the shape of the boxcar smoothed genus curve is not biased, only its amplitude is slightly damped. For a Gaussian random field the boxcar smoothing (2.21) reduces the amplitude by only 0.46% (Rhoads et al. 1994).

As best estimate for the genus curve of the PSCz survey I will therefore give boxcar averages of the raw genus curve along with errors that are estimated as the dispersion among 15 boxcar averaged bootstrap resamplings. Later on I will also employ the bootstrap resampling technique to estimate the statistical uncertainties for a number of other quantities derived from the survey.

2.5 Genus meta-statistics

In this section I want to define measures that quantify deviations of the measured genus curve from the Gaussian random phase prediction.

2.5.1 Amplitude

The most prominent feature of the genus curve is its amplitude N. I will measure this quantity as the amplitude of a random phase genus curve fitted to the raw genus curve by minimizing χ^2 in the range $-1 \le \nu \le 1$. The error in N is estimated as the variance among the amplitudes of fits to the raw genus curves of bootstrap resamplings of the data.

For a Gaussian random field the amplitude constrains the shape of the power spectrum by directly giving its second moment $\langle k^2\rangle$. Motivated by equation (1.13) I introduce the notion of an effective slope

$$n_{\text{eff}} \equiv 3\left((2\pi)^2\lambda^3 N\right)^{2/3} - 3 \tag{2.22}$$

of the power spectrum as measured from the genus curve. If the underlying field is Gaussian one has

$$\left\langle k^2\right\rangle \lambda^2 = n_{\text{eff}} + 3. \tag{2.23}$$

There are at least two caveats here if one wants to infer information on the shape of the power spectrum via equations (2.22) and (2.23). First, the amplitude N is systematically biased high by shot noise due to finite sampling of the density field. In section 2.7 I will estimate this effect and examine its quantitative consequences. Second, the amplitude might be reduced for small smoothing lengths compared to the Gaussian expectation because of phase correlations that start to develop in the weakly nonlinear regime.

The second effect can be turned into a virtue if a way is found to quantify the amplitude drop due to phase correlations. Let me define for that purpose

$$R_{\rm g} = \frac{N}{N_{\rm rp}}, \tag{2.24}$$

where N is the measured genus amplitude (corrected for shot noise or with negligible shot noise influence) and $N_{\rm rp}$ is the amplitude of a random phase field with identical power spectrum as the examined density field.

How can one measure $N_{\rm rp}$? For a fully sampled N-body simulation on a periodic box I can simply *Gaussianize* the field by Fourier transforming it, randomizing all the phases, and then transforming back and measuring the genus again. I have used this approach for the simulations examined in section 4.3.

Unfortunately, for an odd shaped survey volume like that of PSCz this procedure is spoiled by non-trivial phase correlations between mask and density field. Here one could measure the power spectrum directly and integrate it to find $\langle k^2 \rangle$ and subsequently $N_{\rm rp}$ (Vogeley et al. 1994). However, a power spectrum measurement (Feldman, Kaiser & Peacock 1994) is non-trivial, again in part because of the problems induced by the presence of the mask.

Alternatively, I want to propose to measure the variance $\sigma^2(\lambda)$ of the underlying density field as a function of the smoothing length. Then equations (2.10) and (2.11) show that the second moment of the power spectrum is essentially given by the logarithmic slope of $\sigma^2(\lambda)$:

$$\left\langle k^2 \right\rangle \lambda^2 = -\frac{d\, \log \sigma^2}{d\, \log \lambda}. \tag{2.25}$$

In section 2.6 I demonstrate how the variance $\sigma^2(\lambda)$ can be measured. I will then use this function to estimate $\langle k^2 \rangle$ and finally $R_{\rm g}$.

2.5.2 Shift

Previously some authors have found a meatball shift in their genus results (Gott et al. 1989, Rhoads et al. 1994), that is the peak of the genus curve appeared to be shifted to lower threshold values. A way to quantify this shift is

$$\Delta\nu = \frac{\int_{-1}^{1} \nu \, g_{\rm obs}(\nu)\, {\rm d}\nu}{\int_{-1}^{1} g_{\rm fit}(\nu)\, {\rm d}\nu}, \tag{2.26}$$

where $g_{\rm obs}(\nu)$ is the observed raw genus curve and $g_{\rm fit}(\nu)$ is the minimum χ^2 fit to the Gaussian form. The shift of a Gaussian random field is $\Delta\nu = 0$, whereas $\Delta\nu < 0$ indicates a meatball and $\Delta\nu > 0$ a bubble shift.

Park et al. (1992a) have found that the error in $\Delta\nu$ is approximately Gaussian distributed. For that reason I will estimate the 1σ error in the shift from the dispersion among the shifts of the bootstrap resampled fields.

2.5.3 Width

Another measure for the shape of the genus curve is obtained if one defines a width

$$W_\nu = \nu_+ - \nu_- \tag{2.27}$$

of the measured genus curve, where $\nu_\pm$ are the zero crossings of the genus curve on the positive and negative side, respectively. The random phase genus curve has a width $W_\nu = 2.0$. A value of $W_\nu > 2.0$ means that the measured genus curve is broader than random phase, that is, it remains sponge-like over a wider range of threshold values than a Gaussian random field.

To reduce the dependence on noise in the genus curve I measured the width W_ν with the boxcar averaged genus curves. The error has been estimated as the variance among the widths of boxcar averaged genus curves of bootstrap resamplings of the data.

2.6 Variance of the smoothed density field

In the previous section I proposed to use the amplitude ratio of the measured genus curve and the Gaussian random phase prediction as a diagnostic tool to quantify phase correlations. This ratio can be derived with the help of a measurement of the variance $\sigma^2(\lambda)$ of the smoothed density field.

In this section I demonstrate how the variance can be estimated from the smoothed density field with a counts-in-cells analysis with Gaussian cells. My method is a slightly modified version of the approach proposed by Saunders et al. (1991).

As usual let me assume that the galaxy locations arise as a Poisson point process from some underlying probability density field which is taken to be proportional to the mass density. The probability p to find a galaxy in a small volume δV_i is then $p \propto \rho_i S_i \delta V_i$ where ρ_i is the density in the volume element i and S_i denotes the value of the selection function in the cell. I now identify these cells with the voxels used to construct the density field and I denote the number of galaxies in cell i with m_i.

The mean expected number $\overline{m_i}$ of galaxies averaged over many sampling realizations will be

$$\overline{m_i} = k\rho_i S_i, \tag{2.28}$$

where k is a constant equal for all cells. Here and in the following an over bar will denote the expectation value with respect to different sampling realizations for a fixed density field ρ_i.

Since a Poisson process is assumed the probability density function $p_i(m)$, which describes the distribution of the actual counts in the cell, is given by

$$p_i(m) = \sum_{N=0}^{\infty} \frac{\overline{m_i}^N}{N!} e^{-\overline{m_i}} \delta^{(\mathrm{D})}(m - N). \tag{2.29}$$

Let me now define the moment generating function $M_{m_i}(t)$ for the counts in cell i as

$$M_{m_i}(t) = \int p_i(m)\, \mathrm{e}^{m\,t}\, \mathrm{d}m = \exp\left(\overline{m_i}(\mathrm{e}^t - 1)\right). \tag{2.30}$$

Then a moment $\overline{m_i^n}$ can be calculated as

$$\overline{m_i^n} = \left.\frac{\mathrm{d}^n M_{m_i}}{\mathrm{d}t^n}\right|_{t=0}. \tag{2.31}$$

In practice a smoothed density field

$$d_i = \sum_j w_{ij} \frac{m_j}{S_j} \tag{2.32}$$

is derived from the number counts. Note that once the density field has been specified the m_i are independent random variables with respect to the sampling process. The moment generating function for the d_i is therefore

$$M_{d_i}(t) = \prod_j M_{m_j}\left(\frac{w_{ij}}{S_j} t\right) = \exp\left[\sum_j \overline{m_j}(\mathrm{e}^{t\,w_{ij}/S_j} - 1)\right]. \tag{2.33}$$

One reassuringly finds

$$\overline{d_i} = \left.\frac{\mathrm{d}M_{d_i}}{\mathrm{d}t}\right|_{t=0} = \sum_j w_{ij} \frac{\overline{m_j}}{S_j} = k \sum_j w_{ij} \rho_j = k\hat{\rho}_i, \tag{2.34}$$

so the d_i are an unbiased estimate for the smoothed underlying field which I defined as

$$\hat{\rho}_i = \sum_j w_{ij} \rho_j. \tag{2.35}$$

I now want to determine the moments

$$R_n = \langle (\hat{\rho}_i - 1)^n \rangle \tag{2.36}$$

of the smoothed density field. In particular I am interested in the second moment $\sigma^2 = R_2$.

In a first step I need to estimate the mean density $d = \langle d_i \rangle$. An unbiased estimate for d is given by

$$d = \frac{\sum_i g_i d_i}{\sum_i g_i} \tag{2.37}$$

for arbitrary weights g_i. A minimum variance estimator for d can be obtained if one chooses

$$g_i = \frac{1}{\mathrm{var}(d_i)}. \tag{2.38}$$

The variance

$$\mathrm{var}(d_i) = \left\langle \overline{d_i^2} \right\rangle - \left\langle \overline{d_i} \right\rangle^2 \tag{2.39}$$

needed here is calculated in two steps. First the expectation value for many sampling realizations of a fixed density field is taken (over bar) and then an ensemble average with respect to the density field ρ_i (angular brackets) is performed.

The second moment of the counts is

$$\overline{d_i^2} - \overline{d_i}^2 = \sum_j \overline{m_j} \left(\frac{w_{ij}}{S_j} \right)^2 = k \sum_j \rho_j \frac{w_{ij}^2}{S_j}, \tag{2.40}$$

so that one has

$$\mathrm{var}(d_i) = k\rho \sum_j \frac{w_{ij}^2}{S_j}, \tag{2.41}$$

where $\rho = \langle \rho_i \rangle$ is the mean density of the universe. For convenience I will assume $\rho = 1$ in the following. Once d is determined one can rescale the selection function such that $d = 1$ and $k = 1$, and this rescaling will be assumed in the following.

One can now move to higher moments of the counts. For that purpose it is convenient to introduce a moment generating function

$$M_{d_i - \overline{d_i}}(t) = \exp\left[\sum_j \overline{m_j} (\mathrm{e}^{t\, w_{ij}/S_j} - 1) - \overline{d_i}\, t \right] \tag{2.42}$$

for central moments. They can then be calculated as

$$D_i^{(n)} = \overline{(d_i - \overline{d_i})^n} = \left. \frac{\mathrm{d}^n M_{d_i - \overline{d_i}}(t)}{\mathrm{d}t^n} \right|_{t=0}. \tag{2.43}$$

Note that because of the rescaling one has $\overline{d_i} = \hat{\rho}_i$. With a binomial expansion

$$(d_i - 1)^n = (d_i - \overline{d_i} + \hat{\rho}_i - 1)^n = \sum_{k=0}^{n} \binom{n}{k} (d_i - \overline{d_i})^k (\hat{\rho}_i - 1)^{n-k} \tag{2.44}$$

one can relate the moments of the measured field to the moments of the underlying density field. In particular, for $n = 2$ one finds

$$\overline{(d_i - 1)^2} = D_i^{(2)} + (\hat{\rho}_i - 1)^2. \tag{2.45}$$

An unbiased estimator for the variance $R_2 = \sigma^2 = \langle (\hat{\rho}_i - 1)^2 \rangle$ of the underlying density field is therefore

$$R_2 = \frac{\sum_i h_i \left((d_i - 1)^2 - \left\langle D_i^{(2)} \right\rangle \right)}{\sum_i h_i} \tag{2.46}$$

for arbitrary weights h_i. The expectation value

$$\left\langle D_i^{(2)} \right\rangle = \sum_j \frac{w_{ij}^2}{S_j} \tag{2.47}$$

is just the variance calculated in equation (2.41).
A minimum variance estimator for R_2 can now be obtained if one chooses the weights

$$h_i = \frac{1}{\mathrm{var}((d_i - 1)^2)}. \tag{2.48}$$

To evaluate this variance one can again make use of equation (2.44) but this time one has to keep track of more terms:

$$\overline{(d_i - 1)^4} = D_i^{(4)} + 4D_i^{(3)}(\hat{\rho}_i - 1) + 6D_i^{(2)}(\hat{\rho}_i - 1)^2 + (\hat{\rho}_i - 1)^4. \tag{2.49}$$

Equation (2.43) gives the necessary moments $D_i^{(n)}$ as

$$D_i^{(2)} = \sum_j \rho_j \frac{w_{ij}^2}{S_j}, \tag{2.50}$$

$$D_i^{(3)} = \sum_j \rho_j \frac{w_{ij}^3}{S_j^2}, \tag{2.51}$$

$$D_i^{(4)} = 3\left(\sum_j \rho_j \frac{w_{ij}^2}{S_j}\right)^2 + \sum_j \rho_j \frac{w_{ij}^3}{S_j^2}. \tag{2.52}$$

I will now approximate the sums in the above expressions by replacing the density ρ_j with the averaged (smoothed) density $\hat{\rho}_i$ in the particular region. I then have

$$D_i^{(2)} = \hat{\rho}_i Y_i^{(2)}, \tag{2.53}$$

$$D_i^{(3)} = \hat{\rho}_i Y_i^{(3)}, \tag{2.54}$$

$$D_i^{(4)} = 3\left(\hat{\rho}_i Y_i^{(2)}\right)^2 + \hat{\rho}_i Y_i^{(4)}, \tag{2.55}$$

where

$$Y_i^{(n)} = \sum_j \frac{w_{ij}^n}{S_j^{n-1}}. \tag{2.56}$$

This substitution is strictly correct only for a homogeneous density distribution, but it can be expected to be a reasonably good approximation in the context of this analysis. Now one can take the ensemble average of equation (2.49) and get

$$\left\langle \overline{(d_i - 1)^4} \right\rangle = Y_i^{(4)} + 3(R_2 + 1)\left(Y_i^{(2)}\right)^2 + 4R_2 Y_i^{(3)} + 6(R_2 + R_3)Y_i^{(2)} + R_4. \tag{2.57}$$

One sees that the third and fourth moments of the underlying field enter here, too. This illustrates a well known problem in counts-in-cells analysis. A determination of second and higher moments involves the whole hierarchy of moments, because a minimum variance estimate of R_3 and R_4 requires the knowledge of higher moments which in turn depend on still higher moments, and so on. To resolve this problem one has to adopt some form of closure relation. For simplicity I will assume that the underlying field is approximately Gaussian. Hence one has $R_3 = 0$ and $R_4 = 3R_2^2$. Now I can combine equation (2.57) with

$$\left\langle \overline{(d_i - 1)^2} \right\rangle = R_2 + Y_i^{(2)} \tag{2.58}$$

to find

$$\mathrm{var}((d_i - 1)^2) = Y_i^{(4)} + 2\left(Y_i^{(2)}\right)^2 + \left(3\left(Y_i^{(2)}\right)^2 + 4Y_i^{(3)} + 4Y_i^{(2)}\right) R_2 + 2R_2^2. \tag{2.59}$$

The estimator (2.57) finally takes the form

$$R_2 = \frac{\sum_i h_i \left((d_i - 1)^2 - Y_i^{(2)}\right)}{\sum_i h_i} \tag{2.60}$$

with

$$h_i = \frac{1}{Y_i^{(4)} + 2\left(Y_i^{(2)}\right)^2 + \left(3\left(Y_i^{(2)}\right)^2 + 4Y_i^{(3)} + 4Y_i^{(2)}\right) R_2 + 2R_2^2}, \tag{2.61}$$

and has to be solved for R_2 iteratively.

Now that an estimate for the variance of the underlying density field has been derived it is also desirable to find an estimate of the error in this determination. Unfortunately the d_i are strongly correlated since the many voxels of the data cube heavily oversample the smoothed field. An estimate of the statistical error by directly computing the variance of R_2 with equations (2.59) and (2.61) would therefore seriously underestimate the error.

For that reason I adopt a different approach and estimate the uncertainty again with the bootstrap resampling technique. The bootstrap resamplings are generated in the context of the genus statistic anyway so this procedure offers itself quite naturally without producing much overhead.

I apply the above estimator for each bootstrap resampling of the density field and estimate the uncertainty in σ^2 as the scatter among the bootstrap resampled fields. It should be noted, however, that the mean of the second moments for the bootstrap resampled fields will be systematically higher by a constant amount (for fixed λ) because the two-point correlation function of the bootstrap resamplings picks up a delta function at the origin. The variance around this mean can still be used as a measure for the statistical uncertainty in the determination of the second moment of the underlying density field.

2.7 Shot noise and genus amplitude

An important aspect of the sampling problem is that even when condition (2.17) is fulfilled there is shot noise in the density field. Although the reconstruction of the density field represents an unbiased estimate of the underlying density field higher moments of the field are affected by shot noise. For example, in the previous section it was shown that the second moment is increased additively by a shot noise contribution. It is therefore not surprising that the shot noise also leads to a systematic influence on the genus curve. In particular I will demonstrate below that to leading order the amplitude of the genus curve is changed by shot noise.

It seems that this effect has been largely overlooked so far. Gott et al. (1989) give a rough criterion

$$\frac{\lambda}{d} \geq \frac{1}{\pi^{3/2}} \frac{d^2}{r_0^2} \tag{2.62}$$

for which they expect that the shot noise contamination will not mask the true density fluctuations. Here d is the mean interparticle seperation and r_0 the correlation length. However, a number of subsequent studies have directly used the genus amplitude to infer effective slopes of the power spectrum. I will estimate in this section the shot noise influence more precisely and also calculate its influence on the final genus amplitude. With this result one can then try to correct for the shot noise influence and to give a better estimate for the genus amplitude of the underlying density field.

The genus amplitude depends primarily on the behaviour of the two-point correlation function of the smoothed field near the origin. In the following I will therefore first calculate $\hat{\xi}(r)$ under the influence of shot noise. Later on this will be related to a typical measurement of the genus amplitude where the noise level varies across the survey volume.

2.7.1 Correlation function of the smoothed density field

Let me assume that a density field $\delta_{\rm t}(\mathbf{r})$ is sampled in a Poisson process with discrete points $m(\mathbf{r}) = \sum_i \delta^{(\rm D)}(\mathbf{r} - \mathbf{r}_i)$. I allow for a radial variation of the average expected number density $S(r) = \langle m(\mathbf{r})\rangle$ of tracers, where $S(r) = S(z(r))$ is just the selection function defined in section 6.2.

According to equation (6.6) one can then define the measured density contrast $\delta(\mathbf{r})$ in the number counts by

$$\delta(\mathbf{r}) = \frac{m(\mathbf{r})}{S(r)} - 1, \tag{2.63}$$

and the smoothed density field $\hat{\delta}(\mathbf{r})$, that one ultimately wants to use, becomes

$$\hat{\delta}(\mathbf{r}) = \int W(\mathbf{r} - \mathbf{r}')\delta(\mathbf{r}')\,\mathrm{d}\mathbf{r}. \tag{2.64}$$

If the genus statistic of this smoothed density field is constructed the amplitude of the genus curve will depend on its autocorrelation function

$$\hat{\xi}(\mathbf{r}) = \left\langle \hat{\delta}(\mathbf{r}_0)\hat{\delta}(\mathbf{r}_0 + \mathbf{r}) \right\rangle . \tag{2.65}$$

Usually the ensemble average in this equation eliminates the dependence on $\mathbf{r}_0$. As will be seen this is not the case here, if the shot noise in the sampling exhibits a spatial variation as it is the case for any flux limited redshift survey. I will not bother to write the dependence on $\mathbf{r}_0$ as an additional argument on all of the following results. But one should not forget this dependence. Most of the following results apply only locally at $\mathbf{r}_0$ and eventually one will have to construct suitable averages if a region of space with varying sampling density is observed.

I now calculate in detail the effect of the shot noise on $\hat{\xi}(\mathbf{r})$. Putting in the previous definitions one starts with

$$\begin{aligned} \hat{\xi}(\mathbf{r}) &= \left\langle \int \mathrm{d}\mathbf{r}' \int \mathrm{d}\mathbf{r}''\, W(\mathbf{r}_0 - \mathbf{r}')W(\mathbf{r_0} - \mathbf{r} - \mathbf{r}'') \right. \\ &\quad \left. \times \left(\frac{m(\mathbf{r}')}{S(r')} - 1 \right) \left(\frac{m(\mathbf{r}'')}{S(r'')} - 1 \right) \right\rangle . \end{aligned} \tag{2.66}$$

The ensemble average can be performed by using

$$\langle m(\mathbf{r}) \rangle = S(r) \tag{2.67}$$

and

$$\langle m(\mathbf{r}')m(\mathbf{r}'') \rangle = S(r')S(r'')\left(1 + \xi_{\mathrm{t}}(\mathbf{r}' - \mathbf{r}'')\right) + S(r')\delta^{(\mathrm{D})}(\mathbf{r}' - \mathbf{r}''). \tag{2.68}$$

A derivation of equation (2.68) has been given by Feldman et al. (1994). The above results imply

$$\left\langle \left(\frac{m(\mathbf{r}')}{S(r')} - 1 \right) \left(\frac{m(\mathbf{r}'')}{S(r'')} - 1 \right) \right\rangle = \langle \delta(\mathbf{r}')\delta(\mathbf{r}'') \rangle = \xi_{\mathrm{t}}(\mathbf{r}' - \mathbf{r}'') + \frac{\delta^{(\mathrm{D})}(\mathbf{r}' - \mathbf{r}'')}{S(r')}, \tag{2.69}$$

where $\xi_{\mathrm{t}}(\mathbf{r})$ is the autocorrelation function of the true underlying density field. It is seen that the correlation function of the sampled field picks up a Dirac δ-function at the origin. In the limit of infinite sampling ($S \to \infty$) this shot noise contribution vanishes.

With a few shifts in the integration variables I now get the form

$$\begin{aligned} \hat{\xi}(\mathbf{r}) &= \int \mathrm{d}\mathbf{r}' \int \mathrm{d}\mathbf{r}''\, W(\mathbf{r}')W(\mathbf{r} - \mathbf{r}'' - \mathbf{r}')\xi_{\mathrm{t}}(\mathbf{r}'') + \\ &\quad \int \mathrm{d}\mathbf{r}'\, W(\mathbf{r}_0 - \mathbf{r}')W(\mathbf{r}_0 + \mathbf{r} - \mathbf{r}')\frac{1}{S(r')} \end{aligned} \tag{2.70}$$

for equation (2.66). The second term describes the shot noise contribution which unfortunately varies across the survey volume. Hence the dependence on $\mathbf{r}_0$ does not drop out.

As an approximation to the second integral I can assume that $S(r')$ varies sufficiently slowly over a smoothing length. Because the function $W(\mathbf{r}_0 - \mathbf{r}')$ is strongly peaked around $\mathbf{r}_0$ I can then replace $S(r')$ by $S(r_0)$ in the integrand and move it in front of the integral sign. Now one has

$$\begin{aligned}\hat{\xi}(\mathbf{r}) &= \int \mathrm{d}\mathbf{r}' \int \mathrm{d}\mathbf{r}''\, W(\mathbf{r}')W(\mathbf{r}-\mathbf{r}''-\mathbf{r}')\xi_{\mathrm{t}}(\mathbf{r}'') + \\ &\quad \frac{1}{S(r_0)} \int \mathrm{d}\mathbf{r}'\, W(\mathbf{r}')W(\mathbf{r}-\mathbf{r}'). \end{aligned} \tag{2.71}$$

I now choose the Gaussian smoothing kernel of equation (2.8) for $W(\mathbf{r})$ and find with the convolution theorem

$$\int \mathrm{d}\mathbf{r}' W(\mathbf{r}')W(\mathbf{r}-\mathbf{r}') = \frac{1}{(2\pi)^{3/2}\lambda^3}\, \mathrm{e}^{-\frac{\mathbf{r}^2}{2\lambda^2}}. \tag{2.72}$$

With the help of this relation I can further simplify expression (2.71) and obtain

$$\hat{\xi}(\mathbf{r}) = \hat{\xi}_{\mathrm{t}}(\mathbf{r}) + \hat{\xi}_{\mathrm{shot}}(\mathbf{r}), \tag{2.73}$$

where

$$\hat{\xi}_{\mathrm{t}}(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}\lambda^3} \int \mathrm{d}\mathbf{r}''\xi_{\mathrm{t}}(\mathbf{r}'')\, \mathrm{e}^{-\frac{(\mathbf{r}-\mathbf{r}'')^2}{2\lambda^2}} \tag{2.74}$$

labels the contribution of the true underlying field to the correlation function and

$$\hat{\xi}_{\mathrm{shot}}(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}\lambda^3 S(r_0)}\, \mathrm{e}^{-\frac{\mathbf{r}^2}{2\lambda^2}} \tag{2.75}$$

stands for the shot noise contribution.

As usual I assume statistical isotropy for the Universe. Hence $\xi_{\mathrm{t}}(\mathbf{r})$ has only a radial dependence, $\xi_{\mathrm{t}}(\mathbf{r}) = \xi_{\mathrm{t}}(r)$. In this case the angular integrals in $\hat{\xi}_{\mathrm{t}}(\mathbf{r})$ can be done, but the result

$$\hat{\xi}_{\mathrm{t}}(\mathbf{r}) = \frac{2}{(2\pi)^{1/2}\lambda}\, \mathrm{e}^{-\frac{r^2}{2\lambda^2}} \int \mathrm{d}r'' \frac{r''}{r} \xi_{\mathrm{t}}(r'') \sinh\left(\frac{rr''}{\lambda^2}\right) \mathrm{e}^{-\frac{r''^2}{2\lambda^2}} \tag{2.76}$$

is not very handy. It becomes easier if one recognizes that equation (2.74) reads in Fourier space like

$$\hat{P}_{\mathrm{t}}(k) = P_{\mathrm{t}}(k)\, \mathrm{e}^{-k^2\lambda^2/2}, \tag{2.77}$$

and consequently one finds

$$\hat{\xi}_{\mathrm{t}}(r) = \int \mathrm{d}^3k\, \hat{P}_{\mathrm{t}}(k) \frac{\sin(kr)}{kr}. \tag{2.78}$$

If the density field under consideration is to a good approximation a Gaussian random field the amplitude of the genus curve depends only on $\hat{\xi}(0)$ and $\hat{\xi}''(0)$.

These quantities can then be obtained from equations (2.73), (2.75) and (2.78). The contributions of the underlying field to $\hat{\xi}(0)$ and $\hat{\xi}''(0)$ can be written as

$$\hat{\xi}_{\mathrm{t}}(0) = \sigma_{\mathrm{t}}^2 = \int \mathrm{d}^3 k \hat{P}_{\mathrm{t}}(k) \tag{2.79}$$

and

$$\hat{\xi}_{\mathrm{t}}''(0) = -\int \mathrm{d}^3 k \frac{k^2}{3} P_{\mathrm{t}}(k) = -\frac{\langle k^2 \rangle \sigma_{\mathrm{t}}^2}{3}. \tag{2.80}$$

The shot noise terms give rise to

$$\xi_{\mathrm{shot}}(0) = \sigma_{\mathrm{shot}}^2 = \frac{1}{(2\pi)^{3/2} \lambda^3 S(r_0)} \tag{2.81}$$

and

$$\xi_{\mathrm{shot}}''(0) = -\frac{1}{(2\pi)^{3/2} \lambda^5 S(r_0)} = -\frac{\sigma_{\mathrm{shot}}^2}{\lambda^2}. \tag{2.82}$$

2.7.2 Genus amplitude

With the above results I can now calculate the amplitude of the genus curve under the influence of shot noise. To the extent that the shot noise does not spoil the Gaussian nature of the underlying field the amplitude of the genus curve is given by

$$\begin{aligned} N &= \frac{1}{(2\pi)^2} \left(\frac{-\hat{\xi}''(0)}{\hat{\xi}(0)} \right)^{3/2} \\ &= \frac{1}{(2\pi)^2} \left[\frac{\langle k^2 \rangle}{3} + \frac{\sigma_{\mathrm{shot}}^2}{\sigma_{\mathrm{t}}^2 + \sigma_{\mathrm{shot}}^2} \left(\frac{1}{\lambda^2} - \frac{\langle k^2 \rangle}{3} \right) \right]^{3/2}. \end{aligned} \tag{2.83}$$

If the power spectrum can be approximated on the relevant scales by a power law with slope n one has $\langle k^2 \rangle = (n+3)/\lambda^2$. In this case the amplitude becomes

$$N = \frac{1}{(2\pi)^2 \lambda^3} \left[\frac{n+3}{3} + \frac{\sigma_{\mathrm{shot}}^2}{\sigma_{\mathrm{t}}^2 + \sigma_{\mathrm{shot}}^2} \left(\frac{-n}{3} \right) \right]^{3/2}. \tag{2.84}$$

I have tested relation (2.83) with Monte-Carlo experiments and found it to be fulfilled very well. A detailed description of these tests can be found in the subsequent section. Note that for a Poisson distribution of points simulating a white noise power spectrum ($n = 0$) there will be no change of the genus amplitude due to sparse sampling.

In a realistic application of the genus statistic to the real universe one typically has $-3 < n < 0$ so that the shot noise increases the genus amplitude. The latter is therefore biased towards a white noise power spectrum.

How large might this effect be in the analysis of a flux limited sample? Here the noise

$$\sigma_{\mathrm{shot}}^2(r) = \frac{1}{(2\pi)^{3/2} \lambda^3 S(r)} \tag{2.85}$$

is small nearby, but at the far edge of the survey, where most of the volume contributes, it might be large.

Unfortunately the expectation value of the genus is affected by the noise in a nonlinear fashion. One can therefore not simply subtract the noise contribution and arrive at an unbiased estimate as it was possible for the variance of the field.

Instead, I will give an estimate of the expected increase of the genus amplitude due to the shot noise. Let

$$A_{\mathrm{shot}} = \frac{g_{\mathrm{obs}}(0)}{g_{\mathrm{t}}(0)} \tag{2.86}$$

be the ratio of the observed amplitude $g_{\mathrm{obs}}(0)$ under the influence of noise to the amplitude $g_{\mathrm{t}}(0)$ of the field without noise. Since the CONTOUR algorithm weights volume equally I can estimate A_{shot} by

$$A_{\mathrm{shot}} = \frac{\displaystyle\int_{V_{\mathrm{s}}} \left[\frac{\lambda^2 \langle k^2 \rangle}{3} + \frac{\sigma^2_{\mathrm{shot}}(r)}{\sigma^2_{\mathrm{t}} + \sigma^2_{\mathrm{shot}}(r)} \left(\frac{3 - \lambda^2 \langle k^2 \rangle}{3} \right) \right]^{3/2} \mathrm{d}\mathbf{r}}{\displaystyle\int_{V_{\mathrm{s}}} \left[\frac{\lambda^2 \langle k^2 \rangle}{3} \right]^{3/2} \mathrm{d}\mathbf{r}'}, \tag{2.87}$$

where the integration extends over the entire survey volume. This formula assumes that the underlying density field is Gaussian.

In order to compute A_{shot} I estimate the variance σ^2_{t} as described in section 2.6. The quantity $\lambda^2 \langle k^2 \rangle$ can then be estimated from the logarithmic slope of $\sigma^2_{\mathrm{t}}(\lambda)$ as shown in equation (2.25). Finally the shot noise term is given by equation (2.85), so the integration in equation (2.87) can be done.

In my application to the PSCz survey I typically find $A_{\mathrm{shot}} \sim 1.1 - 1.2$. Hence the shot noise increases the measured genus amplitude by up to 20%. A correction for it is therefore in order if one wants to draw quantitative conclusions from the measured genus amplitude. For this purpose one might simply calculate a corrected amplitude $N' = N/A_{\mathrm{shot}}$ by dividing with the factor A_{shot}.

2.7.3 A Monte Carlo test experiment

The relation (2.83) derived above deserves serious testing before it is applied in further analysis. In order to do that I perform a numerical experiment in a Monte Carlo like way.

A clean test of equation (2.83) should work with an exact Gaussian random field that is then sampled by a Poisson point process for a number of different sampling densities. A measurement of the genus for these realistically sampled fields can then be used to quantitatively check the prediction of equation (2.83).

Setting up a Gaussian random field

In order to set up a realization of a Gaussian random field on the computer I construct the Fourier transform of the desired field by assigning each Fourier mode

$\delta_{\mathbf{k}}$ a magnitude according to a prescribed power spectrum and a randomly chosen phase. Since a real field $\delta(\mathbf{x})$ is desired the Fourier components have to respect the condition $\delta_{-\mathbf{k}} = \delta_{\mathbf{k}}^{\star}$, so I use this relation to set one half of the modes.

Because the field can only be tabulated on a finite grid all frequencies above the Nyquist limit are lost. However, this poses no problem in this context because these frequencies will be wiped out in the smoothing process anyway.

If one samples a periodic box of volume $V = L^3$ with a cubic mesh of N^3 points, the corresponding finite Fourier transform will also be a N^3 array. Some care has to be taken in setting up the Fourier transform in this finite Fourier space. The usual convention stores the positive frequencies in the lower half of this array and the negative frequencies in the upper half. If the array $D(i, j, k)$ with $i, j, k \in \{0, 1, \ldots, N-1\}$ contains the finite Fourier transform then one finds in $D(i, j, k)$ the mode with spatial frequency $k_x = i \times \frac{2\pi}{L}$ for $i < \frac{N}{2}$ and $k_x = -(N - i) \times \frac{2\pi}{L}$ for $\frac{N}{2} \leq i < N$, and similarly for the other indices.

For my test I have chosen a physical box size of $240\,h^{-1}$Mpc sampled by a 128^3 grid. I have set up the finite Fourier space of this field with a power law spectrum of the form

$$P(k) = P_0\, k^n, \tag{2.88}$$

where I have chosen a typical slope of $n = -1.5$. The normalization P_0 has been adjusted such that the resulting field $\delta(x)$ has the small dispersion $\sigma = 0.3$. The dispersion must be small enough in order to allow an interpretation of $\rho = 1 + \delta$ as a physical density field where only negligible probability is assigned to values $\rho < 0$.

Sampling the underlying field

One now has a positive Gaussian random field with known properties that one wants to sample with a Poisson point process. For this purpose I throw randomly points into the volume and take the probability to place one at a certain location proportional to the density $\rho = 1 + \delta$ of the underlying field that has just been constructed.

I parameterize the total number of points that are placed by the number $M = \pi^{3/2}\lambda^3 S$ of points per smoothing volume. A smoothing volume is here defined as $V_{\rm sm} = \pi^{3/2}\lambda^3$ so that the total number N of points in the box is given by $N = M\frac{V}{V_{\rm sm}}$. After I have sampled the Gaussian field with N points I convert the resulting point distribution to density contrast and smooth it with a smoothing length of $\lambda = 11.25\,h^{-1}$Mpc. This length is chosen such that the physical size of one voxel of the data cube is just $\lambda/6$. In this way I mimic the discreteness conditions I will later on encounter in the application to the PSCz survey.

For my test I produced random samplings with the sampling densities $M = 1$, 10, 100 and 1000. Compared to the usual application in redshift surveys these sampling densities are much higher since the underlying density field had to be constructed with much less power in order to arrive at a field that allows the interpretation of $\rho = 1+\delta$ as a (positive) physical density field. For each of the four sampling densities

I generated 5 sampling realizations in order to estimate the random scatter between different realizations.

Resulting genus curves

In Figure 2.3 I show the genus curves resulting for the four different sampling densities M. In each panel the average genus curve of the 5 realizations together with error bars derived from the variance among them is shown. Additionally I plotted the best fit random phase genus curve as a dashed line.

Although the same field was sampled for all these genus curves, the result depends on the sampling density used. For the very small sampling density of $M = 1$ the genus curve deviates from the random phase curve and shows a slight meatball shift.

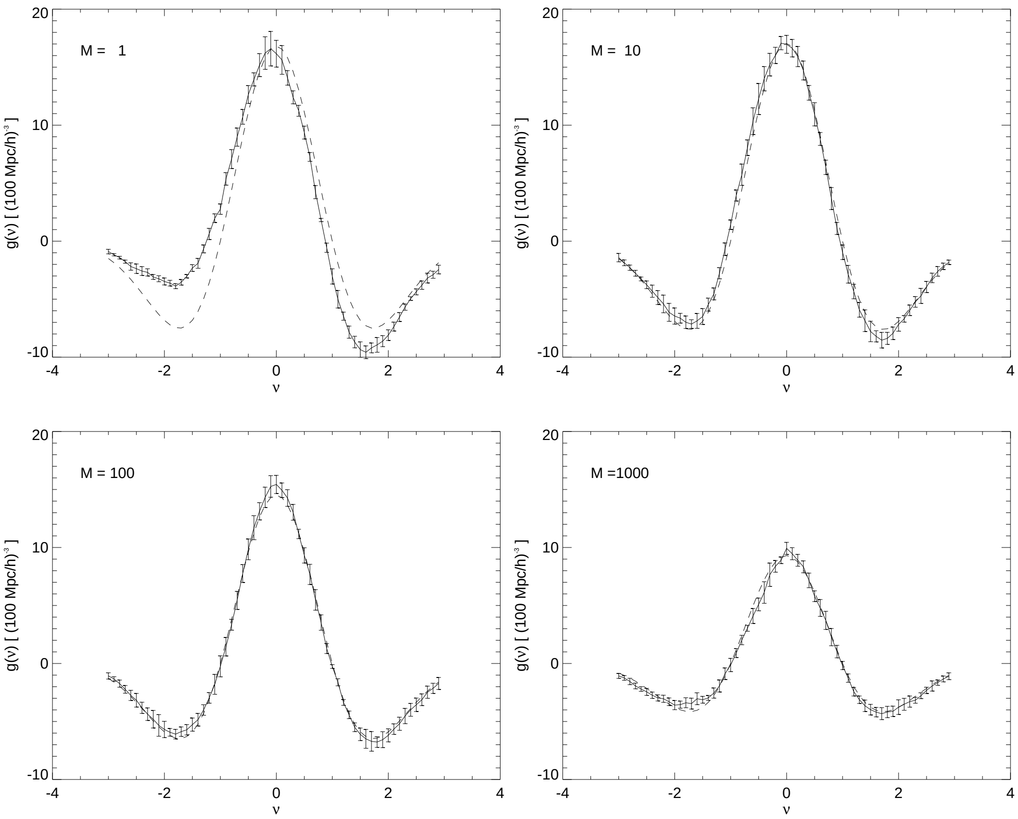


Figure 2.3: The genus curves for Poisson samplings of an underlying Gaussian random field. For each of the four sampling densities the corresponding panel gives the average genus curve (solid line) together with error bars obtained from 5 sampling realizations. The dashed line is the best fit random phase genus curve.

M	$N_{\mathrm{theory}}\ [(100\,h^{-1}\mathrm{Mpc})^{-3}]$	$N_{\mathrm{measured}}\ [(100\,h^{-1}\mathrm{Mpc})^{-3}]$
1	17.74	16.27 ± 0.71
10	17.35	17.25 ± 0.34
100	14.49	14.95 ± 0.63
1000	8.53	9.02 ± 0.30

Table 2.2: The measured genus amplitude N_{measured} compared to the theoretical prediction N_{theory} of equation (2.84).

This discreteness effect quickly vanishes if the sampling density is increased slightly. Then the genus curve follows perfectly the random phase expectation. However, the amplitude of the genus curve still depends on the sampling density.

Of course, this is the effect of the shot noise which is highly emphasized in this set up. In table 2.2 I give a comparison of the measured genus amplitude with the prediction of equation (2.84). Apart from the small discrepancy for $M = 1$, which can be attributed to the discreteness effects, the agreement is very good.

I conclude from this whole exercise that the equations derived for the shot noise influence on the genus amplitude are correct. Hence the shot noise introduced by finite sampling can systematically change the measured genus amplitude.

Chapter 3

Observational Results for PSCz

3.1 Qualitative topology and cosmography

Qualitative topology

The isodensity contours of the smoothed mass density field exhibit a complex appearance. As an illustration I display in figure C.1[1] isodensity contours of the PSCz survey smoothed with $\lambda = 17\,h^{-1}$Mpc. The pictures show surfaces for a number of threshold values in a spherical region of radius $R_{\max} = 150\,h^{-1}$Mpc centered at our own galaxy. From the top left to the bottom right panel in this figure the threshold increases gradually. It is evident that the topological properties vary drastically with this change. It also seen that essential features of the galaxy distribution include coherent high-density structures and large voids.

The images of figure C.1 are actually stills taken from a computer animation[2] that I calculated in order to visualize the change of topology under a continuous variation of the threshold value. Starting at a low density contour the transition from isolated voids to a multiply connected, sponge-like structure for intermediate threshold values is beautifully seen in this movie. If the threshold is increased further the sponge decays and fragments into isolated clusters in a way that perfectly mirrors the low density behaviour. There is no apparent difference between the topology of high and low density regions and as expected a sponge-like topology is seen for threshold values around the mean density.

Cosmography

The smoothed density field might also be used as a map of the universe around us. To demonstrate this cosmographic aspect I have examined the highest local maxima and minima of the smoothed density field in a spherical region of radius $85\,h^{-1}$Mpc around our galaxy. It turns out that most of the significant features in the density field can be identified with known superclusters and voids.

[1] To allow a high quality reproduction of colour plates they had to be moved to appendix C.

[2] A videotape of this movie is available on request from volker@mpa-garching.mpg.de

Density Maxima for $\lambda = 10\,h^{-1}$Mpc, $R_{\max} = 85\,h^{-1}$Mpc

Id	*Name*	l	b	r	x	y	z	δ	*c.l.*
A	Perseus-Pisces	144	-20	50	-39	27	-17	2.69	1.00
B	Coma, A1367	78	86	69	0	4	69	2.44	1.00
C	Local Supercluster	305	24	30	15	-22	12	2.27	0.96
D	Pavo-Indus-Telescopium	337	-24	46	39	-15	-19	2.24	0.88
E	N1600	199	-28	43	-35	-12	-20	1.80	1.00
F	Leo	264	43	76	-5	-55	52	1.79	0.96
G	Hercules	351	57	77	40	-5	65	1.80	0.78
H	Hydra	321	30	67	45	-35	34	1.36	0.97
I	Camelopardalis	144	17	46	-35	25	14	1.28	1.00
J	Cγ	125	-66	53	-12	17	-49	1.41	0.85

Table 3.1: The ten highest density maxima in a sphere of radius $R_{\max} = 85\,h^{-1}$Mpc for smoothing length $\lambda = 10\,h^{-1}$Mpc. The confidence level *c.l.* gives a measure for the statistical significance of the feature. (l, b) are the galactic coordinates of the density peak in degrees and r is its distance in h^{-1}Mpc. The (x, y, z) coordinates are given in h^{-1}Mpc and point into the directions $(l, b) = (0, 0)$, $(l, b) = (90^0, 0)$ and $(l, b) = (0, 90^0)$. *Name* gives an identification of the feature with known superclusters. See figure C.2 for a three-dimensional map of these clusters. The labels *Id* identify the corresponding structures in this map.

In order to assess the statistical significance of these features I have determined the local maxima and minima of 200 bootstrap resamplings of the density field. One cannot expect that the positions of the maxima and minima of these bootstrap resampled fields are at *identical* locations, but for a statistical significant feature they should coincide in some small region of space. I therefore quantify the significance of a feature by giving a confidence level that is set equal to the probability that a bootstrap resampled field exhibits a peak closer than one smoothing length to a local maximum or minimum of the actual density field. Thus, a high confidence level implies that the feature is robust against Poisson sampling noise.

In table 3.1 I list the ten highest density peaks found for $\lambda = 10\,h^{-1}$Mpc out to a distance $85\,h^{-1}$Mpc. The identifications with known superclusters and voids are based on Saunders et al. (1991), Moore et al. (1992), Webster, Lahav & Fisher (1996) and references therein. In figure C.2 I show how these *superclusters* are distributed in space. To simplify the identification of the various structures that can be seen in this three-dimensional map I have chosen different colours for those regions on an isodensity contour that correspond to certain density peaks. In this context the volume of a supercluster was defined as the region of space that can be reached from the peak by following a continuous path on which the density monotonically declines. Note that this definition does not necessarily give disjunct cluster volumes. However, when the threshold value is high enough the clusters will

Density Minima for $\lambda = 10\,h^{-1}$Mpc, $R_{\rm max} = 85\,h^{-1}$Mpc

Id	*Name*	l	b	r	x	y	z	δ	*c.l.*
K	Void γ	233	-47	72	-29	-39	-54	-0.91	0.86
L		116	-19	84	-35	70	-27	-0.87	0.42
M	Tully Void 3	213	19	66	-52	-34	22	-0.85	1.00
N	Eridanus Void	304	-24	65	34	-49	-27	-0.83	0.93
O		332	-28	75	59	-30	-35	-0.81	0.80
P		112	56	60	-12	30	50	-0.79	0.82
Q	Local Void	1	39	50	39	0	32	-0.78	1.00
R	Coma Void	56	38	60	25	39	37	-0.78	0.97
S		160	-74	80	-20	7	-77	-0.76	0.64
T		26	42	67	44	22	45	-0.75	0.85

Table 3.2: The ten lowest density minima in a sphere of radius $R_{\rm max} = 85\,h^{-1}$Mpc for smoothing length $\lambda = 10\,h^{-1}$Mpc. The meaning of the listed quantities corresponds to table 3.1. See figure C.3 for a three-dimensional map of these structures.

appear as isolated structures and the isodensity contour can be split up uniquely into parts that correspond to different density peaks.

On the smoothing scale of $10\,h^{-1}$Mpc our near neighbourhood is dominated by the Local Supercluster, a large overdensity formed by smoothing together the Virgo, Centaurus, and Eridanus-Fornax clusters. The southern hemisphere is dominated by the huge Perseus-Pisces cluster, the relatively sharply located N1600 cluster and the Pavo-Indus-Telescopium structure. Perseus-Pisces and N1600 join to form the mighty Pisces-Cetus complex. In the northern hemisphere the prominent Coma cluster and the beginning of the Hercules supercluster is seen. Together with A1367 and the A2197/2199 clusters they form the northern part of the Great Wall which lies in figure C.2 just off the edge.

The statistical significant identification of voids is somewhat more difficult since they constitute themselves by the very fact that they contain only a few or no galaxies in some region of space. As a consequence very large surveys are needed to pin down their location exactly. Table 3.2 lists the ten deepest underdense peaks detected in the density field out to $85\,h^{-1}$Mpc and figure C.3 provides a map of these structures. The small number of clear identifications mainly reflects the small number of assigned names that can be found in the literature. However, most of the voids seen here (in particular O, L, P and T) can be clearly identified in the contour plots very recently produced by Webster et al. (1996) for the density on shells of four different distances. The agreement found here is remarkable and suggests that voids are indeed a well defined feature of the galaxy distribution. Further evidence for this has been recently provided by El-Ad, Piran & da Costa (1996). They have compiled a list of voids in the IRAS 1.2-Jy redshift survey based on a void finder algorithm and suggest that the Universe is actually void-filled with an average diameter of voids of order $40 \pm 6\,h^{-1}$Mpc.

Let me finally apply a more traditional approach for producing maps of the local cosmography. Here one draws contour plots of the density field in two-dimensional slices through the universe. For this purpose the so-called supergalactic coordinates (de Vaucouleurs et al. 1991) with the north pole at $(l,b) = (47.37^0, 6.32^0)$ and

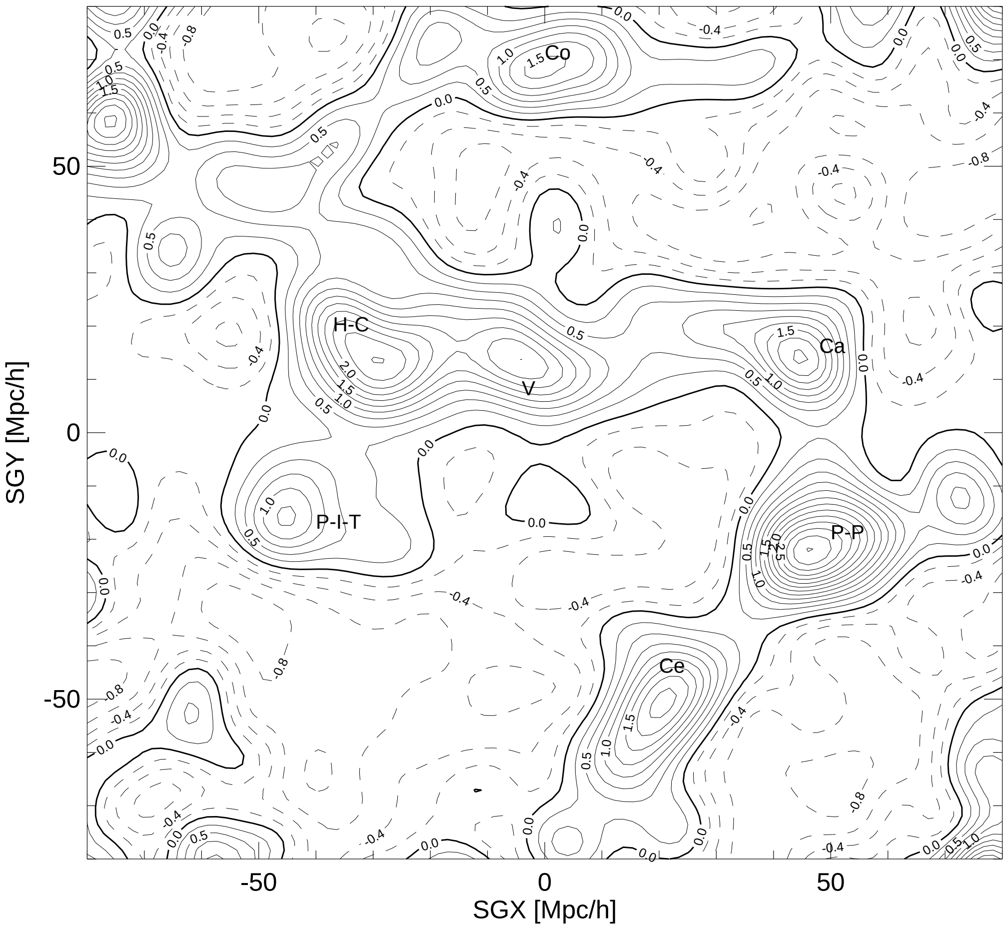


Figure 3.1: The PSCz density field smoothed with $\lambda = 7\,h^{-1}$Mpc in the supergalactic plane. The contour lines are isodensities drawn with spacing $\Delta\delta = 0.3$ for underdense (dashed) and with spacing $\Delta\delta = 0.25$ for overdense (solid) regions. The heavy contour follows the mean density. Identified superclusters are: V=Virgo, P-P=Perseus-Pisces, Ca=Camelopardalis, Ce=Cetus, P-I-T=Pavo-Indus-Telescopium, H-C=Hercules-Centaurus, and Co=Coma.

the origin at $(l, b) = (137.37^0, 0)$ are especially convenient. These coordinates are denoted as (SGL, SGB) and defined such that most of the prominent structure seen nearby lies in the plane defined by $SGB = 0$. This *supergalactic plane* is nearly perpendicular to the ordinary galactic plane which cuts through along the line $SGY = 0$.

In figure 3.1 I show a contour plot of the PSCz density field in the supergalactic plane for a smoothing length of $\lambda = 7\,h^{-1}$Mpc. This map[3] looks quite similar to a corresponding field derived by Strauss & Willick (1995) for the 1.2-Jy survey.

Again, a number of prominent structures can be identified. The nearest significant overdensity is formed by the Virgo supercluster. The large superclusters that lie right next to it at negative supergalactic $SGX \approx -30\,h^{-1}$Mpc are the Hydra-Centaurus and Pavo-Indus-Telescopium structures. Often this system is referred to as the Great Attractor.

The opposite side at positive supergalactic $SGX \approx +50\,h^{-1}$Mpc is dominated by the mighty Pisces-Perseus superclusters with Camelopardalis as its northern extension. The elongated structure in the south of Perseus-Pisces is the Cetus Wall. In the north the Coma cluster shows up as a part of the Great Wall which lies in its main body just beyond the boundary of this plot.

3.2 Previous observational results

In this section I want to give a brief summary of published observational results for the genus statistics.

Gott et al. (1989) have examined the genus curves for a number of observational data sets including a sample of Abell Clusters, versions of the Giovanelli and Haynes sample, the CfA sample, the Thuan and Schneider dwarf sample and the Tully sample. These surveys have been relatively small ($N_{\rm res} \leq 80$) and of varying degree of statistical homogeneity. Gott et al. (1989) conclude from their results that at smoothing scales $\lambda > 12\,h^{-1}$Mpc the topology is fully consistent with the random phase hypothesis. However, at $\lambda = 6\,h^{-1}$Mpc the observed genus curves show a shift towards a meatball topology (that is to the left). In an attempt to explain this shift they pointed out that an extreme biasing procedure together with sparse sampling can cause a meatball asymmetry even when the underlying density distribution is a Gaussian random field. They further find that the genus amplitude as a function of λ is well described by a CDM power spectrum.

Park et al. (1992a) examined the large-scale structure in the southern sky redshift survey. As part of this work they also computed genus curves for a number of smoothing scales. They find that the amplitude is consistent with a $n = -1$ power law power spectrum on the scales they tested. They also observe a slight indication (a $1.3\,\sigma$ effect) for a shift in the meatball direction at $\lambda = 8\,h^{-1}$Mpc. However, their results for the topology remain consistent with being Gaussian.

[3] If you ever happen to get lost in the Universe you better carry a map like this with you...

Moore et al. (1992) have extensively studied the topology of the QDOT redshift survey of IRAS galaxies. The QDOT survey is the direct predecessor of the PSCz survey. In fact, the QDOT survey represents a 1 in 6 sampling of the galaxies selected from the PSC catalogue above 0.6 Jy. It will therefore be especially interesting to compare my PSCz results to Moore et al.'s (1992) findings. Moore et al. (1992) measured the effective slope of the power spectrum on scales between $10\,h^{-1}$Mpc and $50\,h^{-1}$Mpc directly from the amplitude of the genus curves. Over the full range of scales considered they find it to be consistent with a power law power spectrum of slope $n \approx -1$. The authors conclude that the QDOT power spectrum is consistent with a standard CDM power spectrum on scales $< 15\,h^{-1}$Mpc but that at larger scales it falls off less steeply than the CDM spectrum. In particular they claim at a $2\,\sigma$ significance level that there is excess power in QDOT at scales $\approx 30\,h^{-1}$Mpc.

Vogeley et al. (1994) have studied the topology of the CfA redshift survey. With approximately 12000 galaxies this sample had up to 3 times the number of resolution elements of previous studies. (PSCz will be even more powerful.) In the study smoothing scales in the range from 6 to $20\,h^{-1}$Mpc have been examined. To quantify phase correlations in the data the authors compared the measured genus amplitudes to the Gaussian prediction based on a direct measurement of the power spectrum. They find that for smoothing scales $\lambda < 10\,h^{-1}$Mpc the observed genus amplitude is smaller than random phase, which means that the genus amplitude is not a good measure of the power spectrum slope on these scales. For smoothing lengths approaching $20\,h^{-1}$Mpc they find consistency with the random phase hypothesis but there is also weak evidence for a shift towards a bubble topology and the genus curves appear broader than random phase curves. They also compared the genus of the CfA survey with the genus of mock surveys drawn from N-body simulations of three variants of the CDM model. The standard CDM model differed from the observations for $\lambda < 10\,h^{-1}$Mpc whereas an open CDM model and a model with cosmological constant were consistent over the full range of smoothing scales. But all three CDM models weren't able to match the observed broadness of the genus curves.

Rhoads et al. (1994) have examined a recent redshift survey of the Abell cluster catalogue on the smoothing scale $\lambda = 48\,h^{-1}$Mpc. They find that the genus curve shows a slight meatball shift, but remains consistent with Gaussian random phase initial conditions. They determine the effective slope of the power spectrum to be $n = 0.21^{+0.43}_{-0.47}$ on this scale, a result that corresponds to a CDM power spectrum with $\Omega_0 h = 0.36^{+0.46}_{-0.17}$.

3.3 Variance of the density field

In section 2.6 I outlined how the variance $\sigma^2(\lambda)$ of the smoothed underlying density field can be obtained by means of a counts-in-cells analysis in Gaussian spheres. Table 3.3 and figure 3.2 summarize these results. They are still consistent with the findings of Saunders et al. (1991) for the QDOT survey, however, at small smoothing lengths I detect more power for the PSCz survey. At $\lambda = 5\,h^{-1}$Mpc this is a 2σ effect in terms of the error of the QDOT result.

λ [h^{-1}Mpc]	σ^2	λ [h^{-1}Mpc]	σ^2
5.00	0.670 ± 0.036	20.00	0.0752 ± 0.0048
6.00	0.605 ± 0.026	24.00	0.0482 ± 0.0041
7.00	0.517 ± 0.018	28.00	0.0336 ± 0.0031
8.00	0.394 ± 0.015	32.00	0.0240 ± 0.0037
10.00	0.246 ± 0.009	40.00	0.0146 ± 0.0021
12.00	0.188 ± 0.011	48.00	0.0105 ± 0.0019
14.00	0.148 ± 0.008	50.00	0.0102 ± 0.0023
16.00	0.116 ± 0.005	56.00	0.0091 ± 0.0019

Table 3.3: The variance of counts in Gaussian cells for PSCz.

On the scales $5\,h^{-1}\text{Mpc} \leq \lambda \leq 56\,h^{-1}\text{Mpc}$ that I considered the variance σ^2 shows only a slight deviation from a power law. Therefore one might fit a quadratic form

$$\log \sigma^2 = a(\log \lambda)^2 + b \log \lambda + c \tag{3.1}$$

to the data in order to estimate the logarithmic slope

$$\left\langle k^2 \right\rangle \lambda^2 = -\frac{d \log \sigma^2}{d \log \lambda}. \tag{3.2}$$

Tests with the theoretical power spectrum (3.6) show that the form (3.1) is sufficient to recover $\langle k^2 \rangle$ quite accurately on these scales, provided that the values of σ^2 are known well enough. However, the estimate of the amplitude drop crucially depends on the value of $\langle k^2 \rangle \lambda^2$ on small smoothing scales, and tests with mock observations of N-body universes show that the slope of the fit (3.1) at small scales is not sufficiently robust against random errors in the measured variances. It turns out that a direct fit of the CDM power spectrum (3.6) to the emasured variances (see section 3.6 for details) gives a much more reliable estimate of the second moment, so I will employ this alternative procedure in the analysis of PSCz. For the data of the PSCz catalogue both methods give very similar results.

Although the normalization of the power spectrum will not be relevant for the following it is interesting to note that the variance found for IRAS galaxies is smaller than for optically selected galaxies. Optically selected galaxies have variance $\sigma_8 \approx 1$ in *spheres* of radius $8\,h^{-1}$Mpc (Davis & Peebles 1983). Since I used a Gaussian to smooth the galaxy distribution and not top-hat spheres I cannot directly compare my results with σ_8 values. However, if I assume a $n = -1$ power law spectrum for the moment then the variance in a Gaussian window with smoothing length λ will be the same as in spheres of radius $1.56\,\lambda$. Hence the scale $8\,h^{-1}$Mpc corresponds to $\lambda = 5.1\,h^{-1}$Mpc where I find a variance well below unity. IRAS galaxies seem therefore to be less clustered than optically selected galaxies. Based on a model for

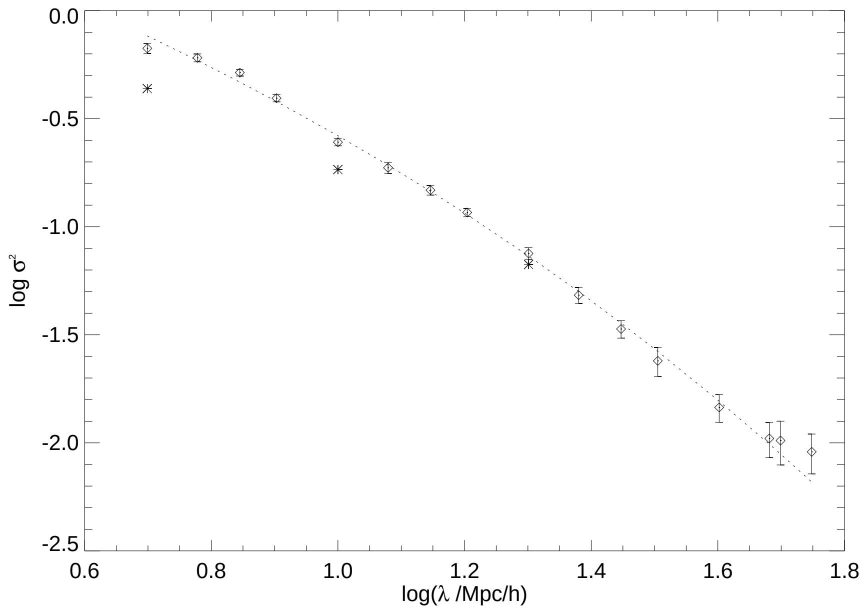


Figure 3.2: The variance $\sigma^2(\lambda)$ found in Gaussian cells in the PSCz survey ($\diamond$ with error bars). For comparison also shown is the result of Saunders et al. (1991) for the QDOT survey ($*$). The dashed line is based on a fit of the data points to the CDM power spectrum and is used to estimate the logarithmic slope of $\sigma^2(\lambda)$.

the power spectrum I will derive in section 3.6 a more accurate estimate for the σ_8 fluctuations found in the PSCz survey.

3.4 Genus curves

In this section I present some of the genus curves derived for the PSCz survey with the methods outlined in sections 1.2 and 2.2. For each smoothing length the unmasked volume in a sphere of radius $R_{\rm max}(\lambda)$ was used as survey volume, where $R_{\rm max}$ was determined as the maximum radius consistent with the sampling condition 2.17. Table 2.1 lists in detail basic characteristics of the adopted survey volumes.

For each of the 16 smoothing lengths in the range $5\,h^{-1}{\rm Mpc} \leq \lambda \leq 56\,h^{-1}{\rm Mpc}$ I calculated a high resolution genus curve with spacing $\delta\nu = 0.01$ and a lower resolution curve with spacing $\delta\nu = 0.1$ in the range of $-3 \leq \nu \leq 3$. Additionally lower resolution curves for 15 bootstrap resamplings of the density field have been calculated.

In figure 3.3 I present results for some selected smoothing lengths. In each case, the left panel shows the high resolution genus curve. The right panel displays my best estimate for the genus curve of PSCz in the form of a boxcar average of the

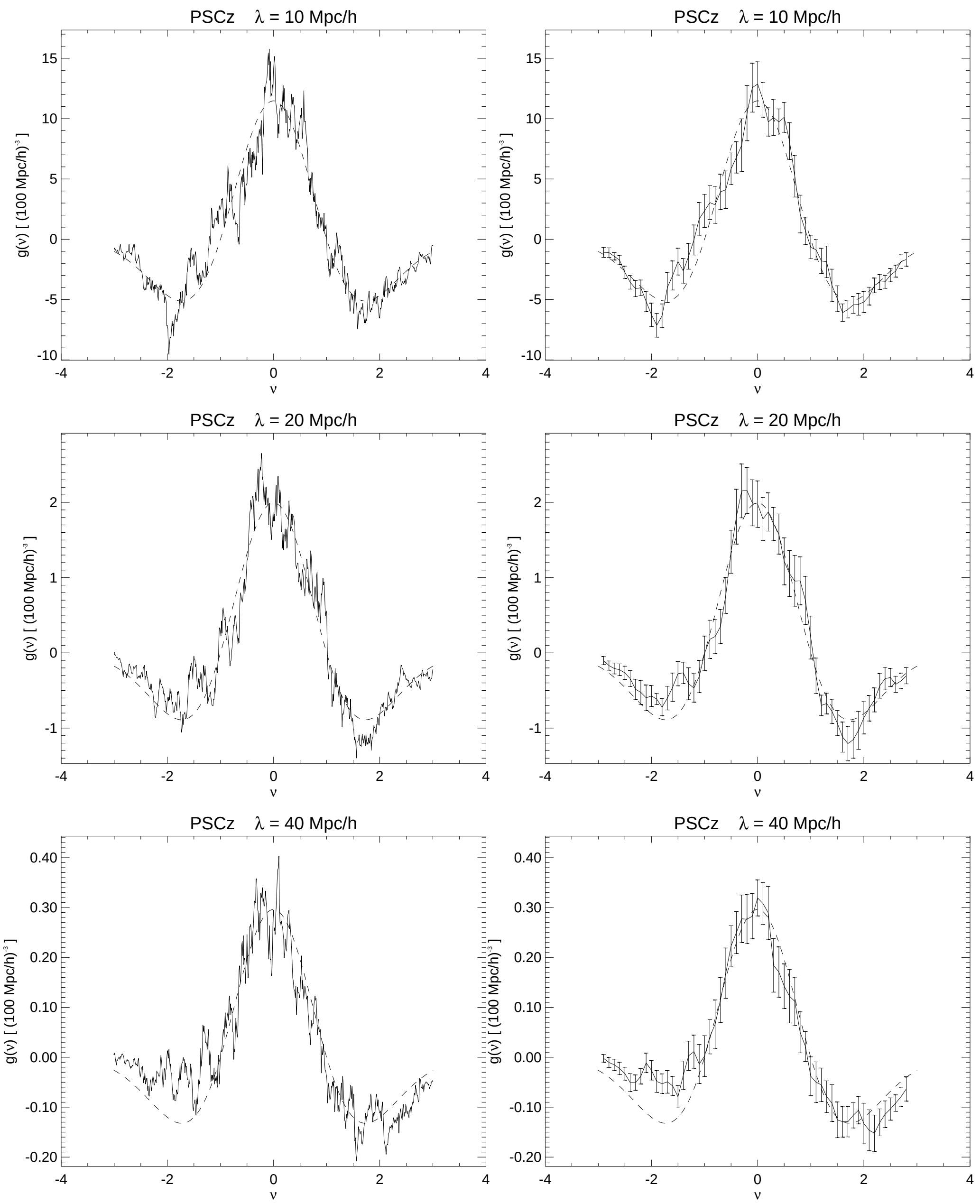


Figure 3.3: Genus curves for the PSCz redshift survey for the smoothing lengths $\lambda = 10$, 20 and $40\,h^{-1}$Mpc. The left panel displays a high resolution genus with $\delta\nu = 0.01$ and the right panel shows the boxcar averaged genus curve together with error estimates derived from bootstrap resamplings of the data. See appendix B for the genus curves of the other smoothing lengths.

raw genus curve with error bars derived with the bootstrap technique. A full listing of the genus curves for the 16 smoothing lengths is given in appendix B.

By visual inspection the genus curves all follow the general form of the random phase curve rather well. The topology is sponge-like over the expected range of threshold values and appears symmetric with respect to high and low density regions. There is no apparent indication for a meatball or bubble shift. However, some of the genus curves seem to be broader than the best fit random phase curves. These observations can be put on a more quantitative basis by means of the genus meta-statistics which is examined in the following section.

3.5 Genus meta-statistics results

The meta-statistics introduced in section 2.5 quantify deviations of the measured genus curve from the Gaussian random phase shape. In table 3.4 I list the most basic of these characteristics in the form of the amplitude N, shift $\Delta\nu$ and width W_ν derived from the genus curves of PSCz. The accompanying confidence intervals have been estimated with the bootstrap technique. A plot of the measured shifts and widths is given in figures 3.4 and 3.5, respectively.

The shift statistics provide no evidence for a meatball or a bubble shift in the genus curves of PSCz, rather the genus curves appear remarkably symmetric over

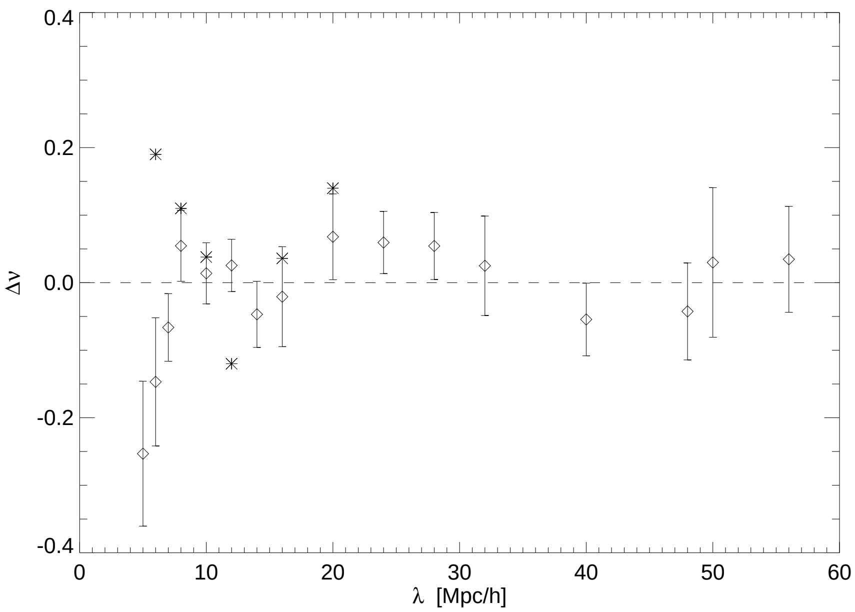


Figure 3.4: The shift $\Delta\nu$ measured for the genus curves of the PSCz survey ($\diamond$ and error bars). Also shown is Vogeley et al.'s (1994) result for the CfA survey ($*$).

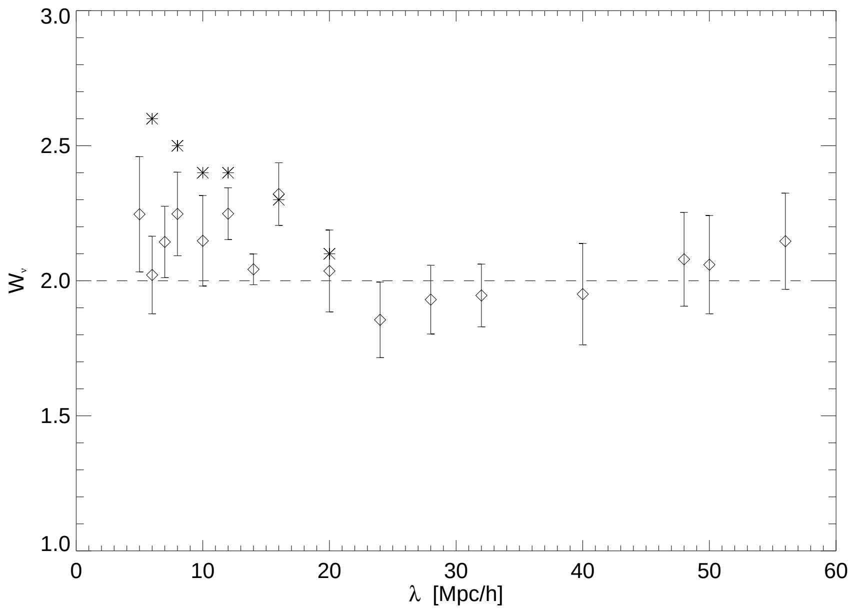


Figure 3.5: The width W_ν measured for the genus curves of the PSCz survey ($\diamond$ and error bars). For comparison Vogeley et al.'s (1994) result for the CfA survey ($*$) is also given.

the full range of scales considered. There is some indication for a meatball shift below $\lambda \leq 7\,h^{-1}$Mpc but the results remain consistent with $\Delta\nu = 0$ if one also takes cosmic variance into account, which is particularly large for $\lambda = 5\,h^{-1}$Mpc due to the small number of resolution elements available here.

The width statistics reveal that the genus curves appear slightly broader than random phase over the range $5\,h^{-1}\text{Mpc} \leq \lambda \leq 20\,h^{-1}$Mpc, but this effect has small statistical significance. Given that there is an additional cosmic variance all the individual results for the width are still consistent with the random phase value $W_\nu = 2.0$. However, the majority of the smoothing scales has $W_\nu \geq 2.0$, so there is some indication of a broadening compared to random phase.

The drop R_g of the genus amplitude compared to the random phase prediction is an interesting quantity because it provides a measure for phase correlations in the density field. Unfortunately, an estimate of this quantity is difficult, since one needs to measure the shape of the power spectrum with an independent method and has to carefully control the systematic influence of the shot noise on the genus amplitude.

In section 2.5 I proposed to use the variance $\sigma^2(\lambda)$ as a handle on the second moment

$$\left\langle k^2 \right\rangle \lambda^2 = -\frac{d \log \sigma^2}{d \log \lambda} \tag{3.3}$$

of the power spectrum. For this purpose I use a value derived from a fit of the CDM

λ [h^{-1}Mpc]	N [$(100\,h^{-1}\text{Mpc})^{-3}$]	Shift $\Delta\nu$	Width W_ν
5.00	50.406 ± 5.487	-0.253 ± 0.107	2.25 ± 0.21
6.00	37.898 ± 3.243	-0.147 ± 0.095	2.02 ± 0.14
7.00	25.312 ± 1.827	-0.066 ± 0.050	2.14 ± 0.13
8.00	19.691 ± 1.329	0.055 ± 0.053	2.25 ± 0.15
10.00	11.468 ± 1.032	0.014 ± 0.045	2.15 ± 0.17
12.00	6.754 ± 0.509	0.026 ± 0.039	2.25 ± 0.10
14.00	4.205 ± 0.260	-0.047 ± 0.049	2.04 ± 0.06
16.00	3.106 ± 0.311	-0.021 ± 0.074	2.32 ± 0.12
20.00	1.993 ± 0.144	0.068 ± 0.064	2.04 ± 0.15
24.00	1.497 ± 0.156	0.060 ± 0.046	1.86 ± 0.14
28.00	1.003 ± 0.084	0.054 ± 0.050	1.93 ± 0.13
32.00	0.641 ± 0.062	0.025 ± 0.074	1.95 ± 0.12
40.00	0.296 ± 0.026	-0.054 ± 0.054	1.95 ± 0.19
48.00	0.144 ± 0.028	-0.043 ± 0.072	2.08 ± 0.17
50.00	0.137 ± 0.024	0.030 ± 0.111	2.06 ± 0.18
56.00	0.097 ± 0.021	0.035 ± 0.078	2.15 ± 0.18

Table 3.4: Basic results for the genus meta-statistics of PSCz. The amplitude N, the shift $\Delta\nu$ and the width W_ν of the raw genus curve is given.

power spectrum to the measured variances (see section 3.6 for further details) as my best estimate of the derivative on the right hand side of equation (3.3). This value together with the measured variance σ^2 itself can then be used to estimate the shot noise influence on the genus amplitude via equation (2.87). I finally apply the resulting correction factor A_{shot} to derive an estimate

$$N^* = \frac{N}{A_{\text{shot}}} \tag{3.4}$$

for the amplitude corrected for shot noise. This amplitude can be also cast into a corrected effective slope n^*_{eff} according to equation (2.22).

Table 3.5 summarizes the results of this procedure. The column with $\langle k^2 \rangle \lambda$ gives the second moment of the power spectrum as estimated from the counts-in-cells analysis. n_{eff} is the effective slope calculated from the measured genus amplitude, n^*_{eff} is the same quantity with the correction factor A_{shot} applied. Finally the amplitude drop is calculated as

$$R_{\text{g}} = \frac{N^*}{N_{\text{rp}}} = (2\pi)^2 N^* \left(\frac{3}{\langle k^2 \rangle}\right)^{3/2} = \left(\frac{n^*_{\text{eff}} + 3}{\langle k^2 \rangle \lambda^2}\right)^{3/2} . \tag{3.5}$$

To estimate the uncertainty of this ratio I assume that the relative error in N_{rp} is of the same order as the relative error in N^* and that the relative uncertainty in the

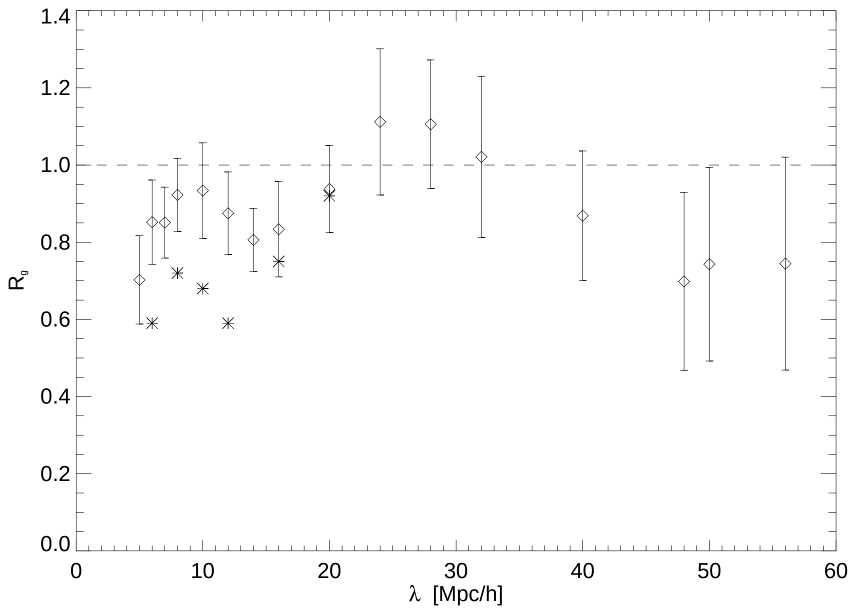


Figure 3.6: The amplitude drop $R_{\rm g}$ for the PSCz survey ($\diamond$ and error bars). A value of $R_{\rm g} = 1.0$ is expected for a Gaussian random field. Vogeley et al.'s (1994) results ($*$) for the CfA survey are also shown.

correction factor $A_{\rm shot}$ is of order of the error of the corresponding determination of the variance σ^2.

Figure 3.6 shows the amplitude drop $R_{\rm g}$ derived in this way. I find that the amplitude of the genus curves is consistent with that predicted for random phases over the full range of scales considered. The agreement is particularly good for $\lambda \geq 20\,h^{-1}$Mpc. At smaller smoothing lengths there is some indication for a drop in genus amplitude, e.g. phase correlations seem to be present there. However, this effect is relatively small. In particular, it is not as strong as in Vogeley et al.'s (1994) results for the CfA survey.

In summary I find that the results for the shift $\Delta\nu$, the width W_ν and the amplitude drop $R_{\rm g}$ are very well consistent with the random phase hypothesis, at least for smoothing scales $\lambda \geq 20\,h^{-1}$Mpc. At smaller smoothing lengths the genus curves continue to follow the Gaussian random phase shape remarkably well. However, in this regime there is an indication for phase correlations in the density field and for a broadening of the genus curves.

λ $[h^{-1}\mathrm{Mpc}]$	$\langle k^2 \rangle \lambda^2$	$A_{\rm shot}$	$n_{\rm eff}$	$n^*_{\rm eff}$	$R_{\rm g}$
5.00	1.40	1.114	-1.81 ± 0.09	-1.90 ± 0.10	0.70 ± 0.11
6.00	1.47	1.110	-1.59 ± 0.08	-1.68 ± 0.09	0.85 ± 0.11
7.00	1.53	1.109	-1.53 ± 0.07	-1.63 ± 0.08	0.85 ± 0.09
8.00	1.58	1.126	-1.38 ± 0.07	-1.50 ± 0.09	0.92 ± 0.09
10.00	1.68	1.157	-1.23 ± 0.11	-1.39 ± 0.11	0.93 ± 0.12
12.00	1.77	1.167	-1.21 ± 0.09	-1.39 ± 0.13	0.87 ± 0.11
14.00	1.84	1.176	-1.22 ± 0.07	-1.41 ± 0.11	0.81 ± 0.08
16.00	1.91	1.187	-1.10 ± 0.13	-1.31 ± 0.13	0.83 ± 0.12
20.00	2.03	1.207	-0.80 ± 0.11	-1.06 ± 0.16	0.94 ± 0.11
24.00	2.13	1.227	-0.38 ± 0.18	-0.71 ± 0.25	1.11 ± 0.19
28.00	2.22	1.232	-0.27 ± 0.15	-0.62 ± 0.26	1.11 ± 0.17
32.00	2.31	1.234	-0.35 ± 0.17	-0.70 ± 0.38	1.00 ± 0.20
40.00	2.45	1.213	-0.53 ± 0.15	-0.83 ± 0.34	0.84 ± 0.16
48.00	2.57	1.178	-0.80 ± 0.28	-1.03 ± 0.44	0.67 ± 0.22
50.00	2.59	1.167	-0.69 ± 0.27	-0.92 ± 0.54	0.72 ± 0.24
56.00	2.67	1.138	-0.69 ± 0.33	-0.88 ± 0.54	0.71 ± 0.26

Table 3.5: Results for the amplitude statistics of PSCz. $\langle k^2 \rangle \lambda^2$ gives the second moment of the power spectrum as estimated from the results of the counts-in-cells analysis. The factor $A_{\rm shot}$ is the expected increase of the genus amplitude due to shot noise. $n_{\rm eff}$ is the effective slope of the power spectrum measured from the raw genus amplitude, $n^*_{\rm eff}$ is the same quantity corrected for the shot noise influence. $R_{\rm g}$ gives the drop of the genus amplitude due to phase correlation as estimated by equation (3.5).

3.6 Shape of the power spectrum

In the previous section it was demonstrated that the genus statistics is fully consistent with the Gaussian random phase hypothesis at least on large scales. The genus amplitude might then be used to constrain the shape of the power spectrum. For this reason I will now compare the measurements with power spectra of theoretical models.

Efstathiou, Bond & White (1992) have given a useful parameterization of the theoretical power spectrum for cold dark matter models in linear theory:

$$P(k) = \frac{Bk}{\left\{1 + \left[ak + (bk)^{3/2} + (ck)^2\right]^\nu\right\}^{2/\nu}}, \tag{3.6}$$

where $a = 6.4\,\Gamma^{-1}\,h^{-1}\mathrm{Mpc}$, $b = 3.0\,\Gamma^{-1}\,h^{-1}\mathrm{Mpc}$, $c = 1.7\,\Gamma^{-1}\,h^{-1}\mathrm{Mpc}$ and $\nu = 1.13$. This form is able to fit a wide class of theoretical models where Γ acts as a shape parameter. The standard, flat, scale-invariant, adiabatic cold dark matter model (SCDM) is accurately fitted for $\Gamma = h$. Also flat cold dark matter models with a cosmological constant are reasonably well described by equation (3.6) if $\Gamma = \Omega_0 h$. Variants of CDM models that exhibit more large scale power due to decaying neutrinos are well approximated if $\Gamma \simeq \Omega_0 h[0.861 + 3.8(m\,\tau)^{2/3}]^{1/2}$, where m is the neutrino mass in units of 10 keV and τ is its lifetime in years.

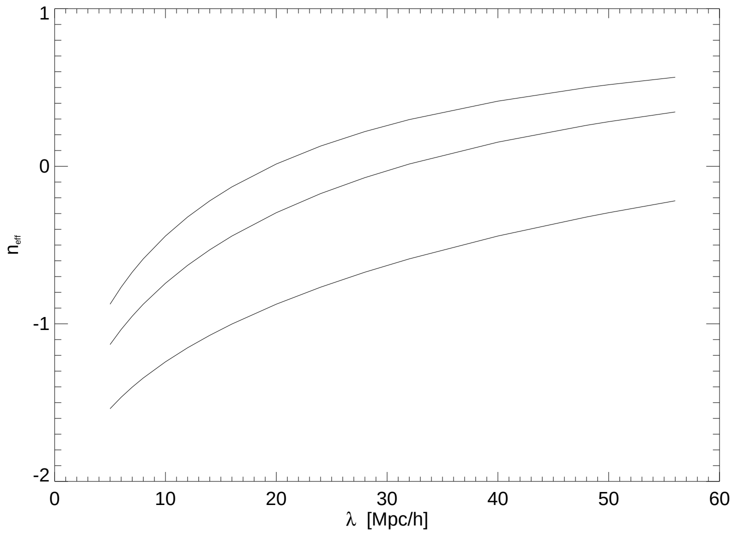


Figure 3.7: The effective slope as predicted by linear theory for cold dark matter models. From top to bottom the curves are computed for the values $\Gamma = 0.8$, 0.5 and 0.2.

The above analytic form (3.6) can be used to calculate the second moment and the effective slope of the smoothed power spectrum $\hat{P}(k) = P(k)\,\mathrm{e}^{-k^2\lambda^2/2}$. Figure 3.7 displays $n_{\mathrm{eff}}(\lambda)$ for a number of different values for Γ. One can see that the curve is sensitive to Γ, hence it should be possible to constrain it with the help of the observed genus amplitudes.

For that purpose I have fitted the effective slope of the CDM power spectrum (3.6) as a function of Γ to the measured values of n^*_{eff} by minimizing χ^2. I have also estimated $1\,\sigma$ confidence intervals by finding the points where χ^2 reached two times its minimum value. In this way I obtained

$$\Gamma = \Omega_0 h = 0.134^{+0.070}_{-0.050}, \tag{3.7}$$

where the fit was only done in the range $\lambda \geq 20\,h^{-1}$Mpc in order to reduce the systematic influence of possible phase correlations at smaller scales. If the fit is performed with the uncorrected slopes n_{eff} a value of $\Gamma = 0.23^{+0.12}_{-0.08}$ results.

Reassuringly the slope resulting from the theoretical power spectrum (3.6) for the value (3.7) for Γ agrees well with the slope derived from the logarithmic derivative of the measured variance $\sigma^2(\lambda)$ (compare the solid and dashed curves in figure 3.8).

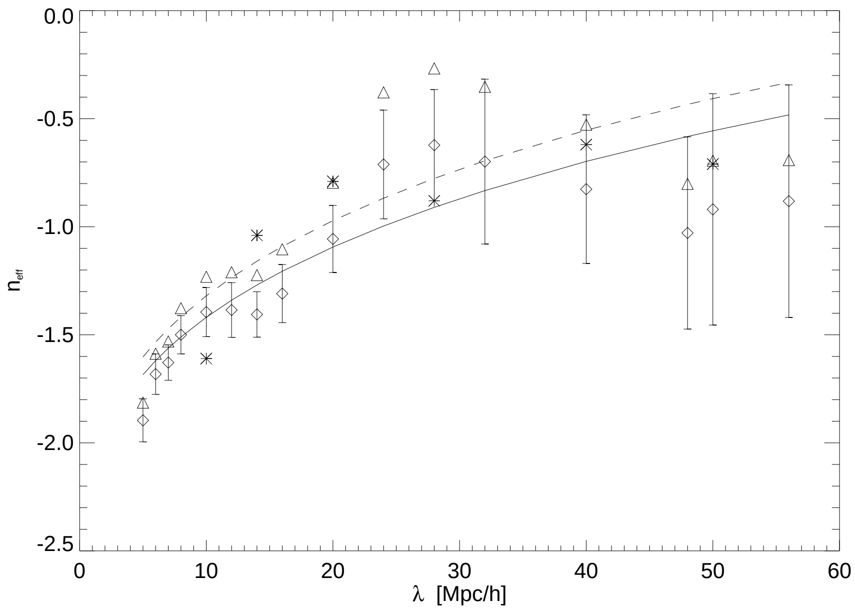


Figure 3.8: The effective slope n^*_{eff} measured from the genus amplitude of the PSCz survey ($\diamond$ with error bars). Also shown is the slope n_{eff} without shot noise correction ($\triangle$). For comparison the slopes measured by Moore et al. (1992) for the QDOT survey ($*$) are also shown. The dashed curve results from a fit of the power spectrum (3.6) to the measured variance $\sigma^2(\lambda)$, whereas the the solid line results from a direct fit of the effective slopes n^*_{eff} to the power spectrum. Not shown in this figure is the measurement of $n_{\text{eff}} = 0.36^{+0.43}_{-0.47}$ by Rhoads et al. (1994) for Abell clusters.

The cold dark matter power spectrum (3.6) can also be used to predict the variance

$$\sigma^2(\lambda) = \int P(k)\,\mathrm{e}^{-k^2\lambda^2/2}\,\mathrm{d}^3k \tag{3.8}$$

of the smoothed density field. Hence these values can be fitted directly to the results of the counts-in-cells analysis. This procedure determines the shape parameter as

$$\Gamma = 0.169^{+0.054}_{-0.041}, \tag{3.9}$$

which is also used to estimate the second moment $\langle k^2 \rangle \lambda^2$ by means of equations (3.3), (3.6), and (3.8).

This value for Γ agrees reasonably well with the value (3.7) derived from the genus amplitudes. Both procedures use entirely different statistical methods and weight the data quite differently. Since the genus analysis is expected to be rather insensitive to biasing whereas the counts-in-cells method might be affected strongly by it, this result suggests that the bias factor is approximately constant over the scales considered.

In contrast to the fits based on the effective slopes the above procedure also determines the normalization of the power spectrum. Conventionally it is expressed in terms of the root mean square fluctuations σ_8 in top-hat spheres of radius $R = 8\,h^{-1}$Mpc. This quantity is related to the power spectrum by

$$\sigma_8^2 = \int P(k) \left(\frac{3 j_1(kR)}{kR} \right)^2 \mathrm{d}^3 k, \tag{3.10}$$

where $j_1(x)$ is a spherical Bessel function arising from the Fourier transform of the top-hat window. My fit to the power spectrum gives rise to

$$\sigma_8 = 0.834^{+0.031}_{-0.034}, \tag{3.11}$$

which is somewhat larger than previous results for smaller surveys of IRAS galaxies. For example, Moore, Frenk, Efstathoiu & Saunders (1994) have found $\sigma_8 = 0.58 \pm 0.14$ for the QDOT survey based on a correlation analysis. Figure 3.2 has already indicated that I find more power on small scales for PSCz than detected in QDOT. Potentially the PSCz result is more accurate and a high value like $\sigma_8 = 0.83$ would reduce the difference in clustering strength between IRAS galaxies and optically selected galaxies.

The value $\Gamma \approx 0.16$ I have found for PSCz means that there is considerably more large-scale power compared to the standard cold dark matter model which has $\Gamma = h$. The latter is ruled out almost certainly by this result, given that the Hubble constant is very likely to be larger than 40 km/s/Mpc. In the framework of CDM models this low value for Γ favours either an open cold dark matter cosmology, a flat universe with a large contribution by a cosmological constant, or a variant of the cold dark matter model with decaying neutrinos or a tilted power spectrum. In any case the ratio of large to small-scale power must be higher than in the standard cold dark matter scenario.

Chapter 4

Comparison with N-body simulations

4.1 The VIRGO project

The Virgo consortium is an international collaboration of scientists who started an ambitious project to model the formation of structure in the Universe with the largest available parallel supercomputers in Europe. In alphabetical order the collaboration is formed by the astrophysicists Hugh Couchman (University of Western Ontario, Canada), George Efstathiou (Oxford University, Britain), Carlos Frenk (University of Durham, Britain), Alistair Nelson (Cardiff University, Britain), John Peacock (Edinburgh University, Britain), Peter Thomas (University of Sussex, Britain), and Simon White (Max-Planck-Institute for Astrophysics, Germany). Additionally Jörg Colberg (Max-Planck-Institute for Astrophysics, Germany), Roger Hutchings (University of Sussex, Britain), Adrian Jenkins (University of Durham, Britain), and Frazer Pearce (University of Sussex, Britain) participate in the study as PhD-students and postdocs, respectively.

As part of the Virgo project (Colberg 1996) the formation of large-scale structure and clusters of galaxies is simulated with unprecedented resolution. The simulations are done with a new AP^{3}M-SPH code named HYDRA (Couchman, Thomas & Pearce 1995). It is parallelized and runs on CRAY T3D/T3E parallel supercomputers based in Edinburgh and Garching. The P^{3}M-code calculates the gravitational force as a sum of two contributions. The long-range part is computed by first assigning the particles to a mesh (PM part) and then solving for the gravitational potential with Fourier techniques. On small scales the force calculation is supplemented by a direct particle-particle (PP part) summation. A further gain in resolution and a considerable improvement in runtime efficiency is obtained by allowing for an adaptive mesh. This means that highly clustered regions in the simulation are computed with a sequence of finer mesh grids. A combination of these techniques leads to the AP^{3}M code, which achieves a wide dynamic range at reasonable cost. HYDRA combines AP^{3}M with smoothed particle hydrodynamics (SPH) to simulate an optional baryonic contribution to the mass content of the model universe.

The Virgo collaboration kindly granted me access to the data of three very large models that contain only cold dark matter (CDM) and exhibit flat background geometries.

These variants of the CDM cosmogenies are specified by the parameters summarized in table 4.1. The standard cold dark matter model (SCDM) contains only cold dark matter and has the power spectrum (3.6) for $\Gamma = \Omega_0 h = 0.5$ in the linear regime. This form is theoretically predicted from the CDM inflationary scenario assuming a primordial scale-invariant Harrison-Zel'dovich spectrum.

	SCDM	τCDM	ΛCDM
Number of particles	256^3	256^3	256^3
Box size	$240\,h^{-1}$Mpc	$240\,h^{-1}$Mpc	$240\,h^{-1}$Mpc
$z_{\rm start}$	50	50	50
Ω_0	1.0	1.0	0.3
Ω_Λ	0.0	0.0	0.7
Hubble constant h	0.5	0.5	0.7
Power spectrum shape Γ	0.5	0.21	0.21
Linear normalization σ_8	0.6	0.9	0.9
Mass per particle $[M_\odot]$	4.55×10^{11}	4.55×10^{11}	9.74×10^{10}
Number of mesh cells	512^3	512^3	512^3

Table 4.1: The specifications of the three CDM models that I analyze subsequently. The simulations have been done by the VIRGO collaboration.

However, recent large-scale structure observations (e.g. Vogeley, Park, Geller & Huchra 1992, Park et al. 1992a, Tadros & Efstathiou 1995) are in serious conflict with the SCDM model, mainly because they measure an excess of power at large scales compared to the SCDM model. In summary these studies find that the power seen in a variety of samples is well fitted by the CDM power spectrum if $\Gamma = \Omega_0 h \approx 0.2$. Note that this result is again supported by my findings for the PSCz redshift survey. Given that the Hubble constant might be as high as $h \approx 0.8$ (Freedman et al. 1994) this therefore implies either an open universe ($\Omega_0 < 0$, single bubble inflation) or a closed universe with a cosmological constant ($\Omega_0 \approx 0.3$, $\Omega_\Lambda \approx 0.7$). Alternatively, models with a contribution of hot dark matter or with a tilted power spectrum can also be made to fit the data.

The τCDM and ΛCDM simulations explore some of these possibilities. The τCDM model is a variant of the cold dark matter scenario with more power on large scales which could arise due to decaying heavy neutrinos, for example. In the ΛCDM model the matter content of the universe is reduced in favour of a vacuum energy density. In many respects the ΛCDM simulation represents an attractive model because it presently can match the broadest range of observational constraints.

The N-body simulations with their huge number of more than 16.7 million particles represent excellent data which in principle allows very accurate genus statistics especially on small scales. Here the number of resolution elements is high, the sampling is excellent, and the genus curves are hardly influenced by cosmic variance.

However, on scales $\lambda \geq 20\,h^{-1}$Mpc the volume probed by PSCz becomes larger than that of the simulations. I therefore restrict my genus analysis of the model universes to the ten smoothing scales between $5\,h^{-1}$Mpc and $24\,h^{-1}$Mpc used in the analysis of PSCz. In this range a direct comparison between the PSCz results and the N-body simulations is most powerful.

The data I used for my analysis is the output of the simulations at redshift $z = 0$. The particle positions and velocities of a single simulation consume even in compressed form an impressive 268 MByte. In order to deal with this amount of information I first constructed a representation of the density field by clouds-in-cell assignment of the particles to a 128^3 data grid. This periodic data cube was then used to generate smoothed density fields with Fourier techniques for the topological analysis. Clearly, a number of complications encountered in the application to the PSCz survey, e.g. angular mask and non-periodicity of the survey volume, are absent here.

4.2 Variance and power spectrum

Once a smoothed density field is calculated, it is straightforward to compute the density contrast $\delta(\mathbf{x})$ and the variance

$$\sigma^2 = \sum_i \delta_i^2 \tag{4.1}$$

of the N-body field. The summation extends here simply over the 128^3 cells of the data cube used to represent the field.

The measurements of $\sigma^2(\lambda)$ can now be used to fit them to the theoretical prediction

$$\sigma_{\mathrm{t}}^2 = \int P(k)\,\mathrm{e}^{-k^2\lambda^2/2}\mathrm{d}^3k \tag{4.2}$$

assuming some model for the power spectrum $P(k)$. In this case the form of equation (3.6) should be able to fit all three models reasonably well, at least on large scales where the modes are still in the linear regime. The parameters of this fit involve the shape Γ and the normalization B. The latter is often expressed in terms of σ_8, e.g. in form of the fluctuations in top-hat spheres of radius $8\,h^{-1}$Mpc, which is the scale of rich clusters of galaxies. In this way the models can be relatively easily normalized to the observed cluster abundances. In principle it is also possible to normalize the spectrum to the power on the very largest scales as observed in the cosmic microwave background by the COBE satellite. However, the simulations of the Virgo collaboration target the clustering of galaxies on small scales. Hence the N-body simulations are actually normalized to the observed abundances of rich clusters of galaxies. White, Efstathiou & Frenk (1993) have shown that σ_8 is almost

λ $[h^{-1}\mathrm{Mpc}]$	σ^2 SCDM	τCDM	ΛCDM
5.00	0.4175	0.4693	1.0858
6.00	0.2922	0.3447	0.7758
7.00	0.2145	0.2647	0.5834
8.00	0.1630	0.2098	0.4548
10.00	0.1012	0.1406	0.2978
12.00	0.0673	0.0999	0.2084
14.00	0.0468	0.0738	0.1525
16.00	0.0336	0.0562	0.1152
20.00	0.0187	0.0346	0.0699
24.00	0.0111	0.0226	0.0448
28.00	0.0068	0.0154	0.0298
32.00	0.0044	0.0109	0.0203
40.00	0.0020	0.0060	0.0102
48.00	0.0010	0.0037	0.0057
50.00	0.0009	0.0033	0.0050
56.00	0.0006	0.0024	0.0035

Table 4.2: Results for the variance of the N-body simulations.

independent of the shape of the power spectrum for this normalization. However, there is a strong dependence on the cosmological parameter in the form

$$\sigma_8 \simeq 0.6\,\Omega_0^{-0.6}. \tag{4.3}$$

Eke, Cole & Frenk (1996) give for a recent calculation of this normalization the value

$$\sigma_8 = (0.52 \pm 0.04)\Omega_0^{-0.46+0.10\Omega_0} \tag{4.4}$$

for models with zero cosmological constant and

$$\sigma_8 = (0.52 \pm 0.04)\Omega_0^{-0.52+0.13\Omega_0} \tag{4.5}$$

for flat models with cosmological constant $\Omega_\Lambda = 1 - \Omega_0$.

In table 4.2 I list my measurements of the variance of the evolved N-body simulations, and table 4.3 gives the resulting parameters of the the power spectra fits. In order to be sensitive only to the linear regime I restricted the fits to $\lambda \geq 2\,r_0$, where r_0 is the correlation length (see below) of the simulation. For the SCDM and τCDM models I used $\lambda \geq 7\,h^{-1}$Mpc and for the ΛCDM model $\lambda \geq 10\,h^{-1}$Mpc. I find that shape of the ΛCDM and τCDM models is given by $\Gamma \approx 0.25$ whereas the SCDM model has $\Gamma = 0.5$.

The normalization of $\sigma_8 = 0.94$ for the ΛCDM model is close to that found for optical galaxies and slightly higher than that of the other two models which are

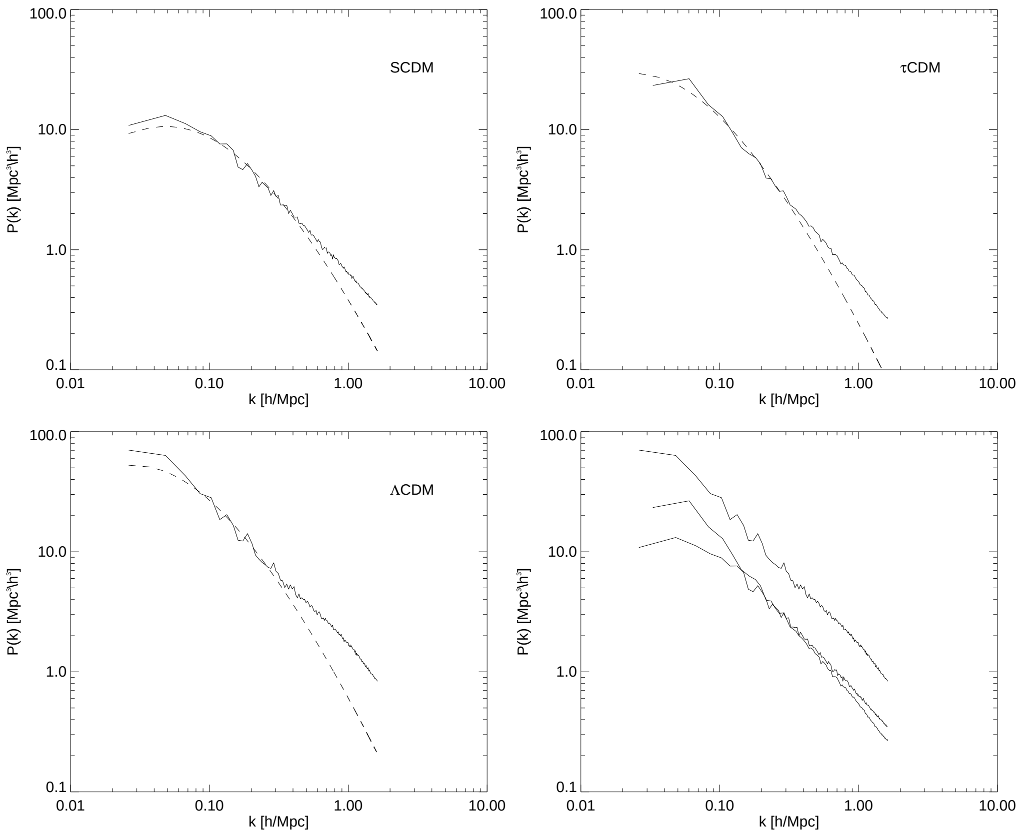


Figure 4.1: The measured power spectra of the three N-body simulations. The panel on the bottom right shows from top to bottom the ΛCDM, τCDM and SCDM models. The other panels compare for each model the measured power spectrum (solid) with the fitted initial power spectrum (dashed). The deviation of the measured spectra from the initial power spectra at small scales is due to the nonlinear evolution of the clustering.

	SCDM	τCDM	ΛCDM
Γ	0.495	0.241	0.282
σ_8	0.604	0.638	0.951

Table 4.3: Parameters for fits of the power spectrum (3.6) to the measured variances of the N-body simulation. For the ΛCDM model the fit was done in the range $10\,h^{-1}\mathrm{Mpc} \leq \lambda \leq 32\,h^{-1}\mathrm{Mpc}$, for the SCDM and τCDM models in the range $8\,h^{-1}\mathrm{Mpc} \leq \lambda \leq 32\,h^{-1}\mathrm{Mpc}$.

close to $\sigma_8 = 0.6$. Note that this normalization strictly applies only to the dark matter component. If galaxies are biased tracers of the mass density field different normalizations for galaxies and the dark matter are entirely possible.

Of course, it is also possible to directly measure the power spectrum from the Fourier transform of the raw data cube. Here one has to carefully collect all the power found in small shells in k-space. Then the power spectrum $P(k)$ can be plotted in the form of a distribution function. Figure 4.1 displays as solid lines the measured power spectra for the three simulations. The corresponding fits to the form of the linear CDM power spectrum are also shown as dashed lines. Because of the nonlinear evolution of the clustering the measured spectra deviate on small scales from the linear prediction.

In these plots one can clearly see the additional power of the ΛCDM and τCDM models at large scales (that is at small k) compared to the SCDM model. Apart from a slightly higher normalization of the Λ-model the shapes of the ΛCDM, SCDM and τCDM power spectra are quite similar on small scales.

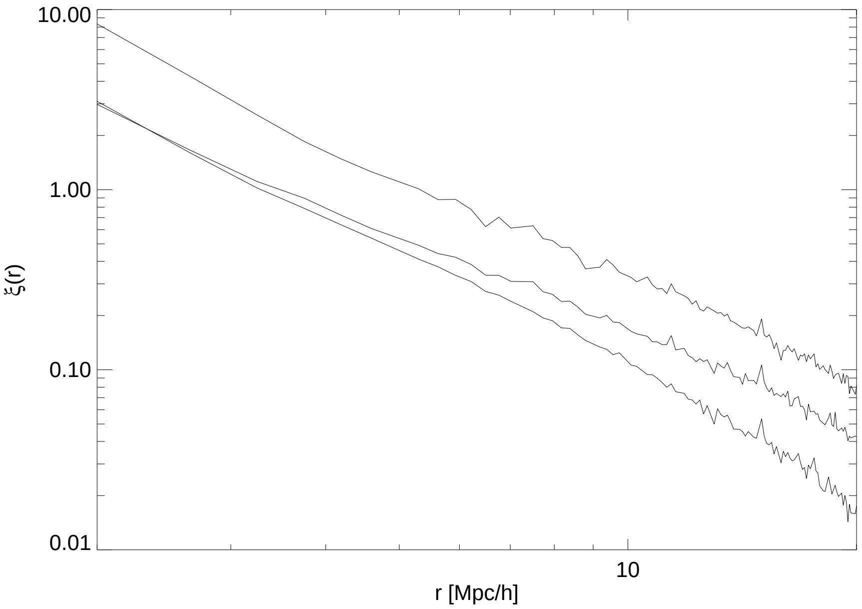


Figure 4.2: The correlation functions of the three N-body simulations. From top to bottom the evolved ΛCDM, τCDM and SCDM models are shown.

Another interesting quantity is the two-point correlation function of the models. In the case of a periodic simulation box it can be computed with relative ease. The correlation field $\xi(\mathbf{r})$ is simply given as inverse Fourier transform of the squared magnitude of the Fourier modes $\delta_\mathbf{k}$ of the density contrast. $\xi(r)$ can then be calculated as angular average over this correlation field.

Figure 4.2 shows the correlation functions I computed in this way for the three simulations. On the scales considered here the correlation functions follow a power law quite accurately. This behaviour is also typically found in correlation analysis of redshift surveys and usually parameterized in the form

$$\xi(r) = \left(\frac{r_0}{r}\right)^{\gamma}. \tag{4.6}$$

	SCDM	τCDM	ΛCDM
r_0 [h^{-1}Mpc]	3.41	3.57	5.50
γ	2.01	1.72	1.90

Table 4.4: Parameters for fits to the two-point correlation function of the N-body simulations.

In table 4.4 I give the parameters of power law fits to the measured correlation functions. The correlation length r_0 sets the scale on which the correlation function reaches unity. Reflecting the higher normalization the correlation length is somewhat larger for the ΛCDM model than for the SCDM and τCDM models.

4.3 Genus curves for variants of CDM models

In order to calculate the genus curves for the three N-body simulations I first generated an underlying mass density field by assigning the particles to a 128^3 grid. This clouds-in-cell assignment represents itself some kind of smoothing, in particular the frequencies above the Nyquist limit are lost. However, this is no problem since the subsequent smoothing with a Gaussian kernel of length λ wipes them out anyway. The smoothing is conveniently done with Fourier techniques, and in this case their periodic nature is quite welcome since the N-body simulations work with a periodic density field.

For each of the simulations I generated smoothed density fields for the ten lowest smoothing lengths used in the analysis of PSCz and I then computed genus curves with spacing $\delta\nu = 0.02$ for them. The number of resolution elements drops from 19861 for $\lambda = 5\,h^{-1}$Mpc to 180 for $\lambda = 24\,h^{-1}$Mpc, hence the limited survey volume will lead to large cosmic variance beyond $\lambda = 24\,h^{-1}$Mpc even for these large simulations.

As discussed in section 2.5 the amplitude drop of the genus curve compared to the genus curve of a Gaussian random field with the same power spectrum might serve as a measure of phase correlations in the density field. In order to measure this drop for the simulations I generated a realization of a Gaussian random field with *exactly* the same power spectrum by *Gaussianizing* the simulations.

For this purpose I computed the Fourier transform of the raw density field, randomized all the phases of the modes $\delta_{\mathbf{k}}$ constrained by the condition $\delta^*_{\mathbf{k}} = \delta_{-\mathbf{k}}$,

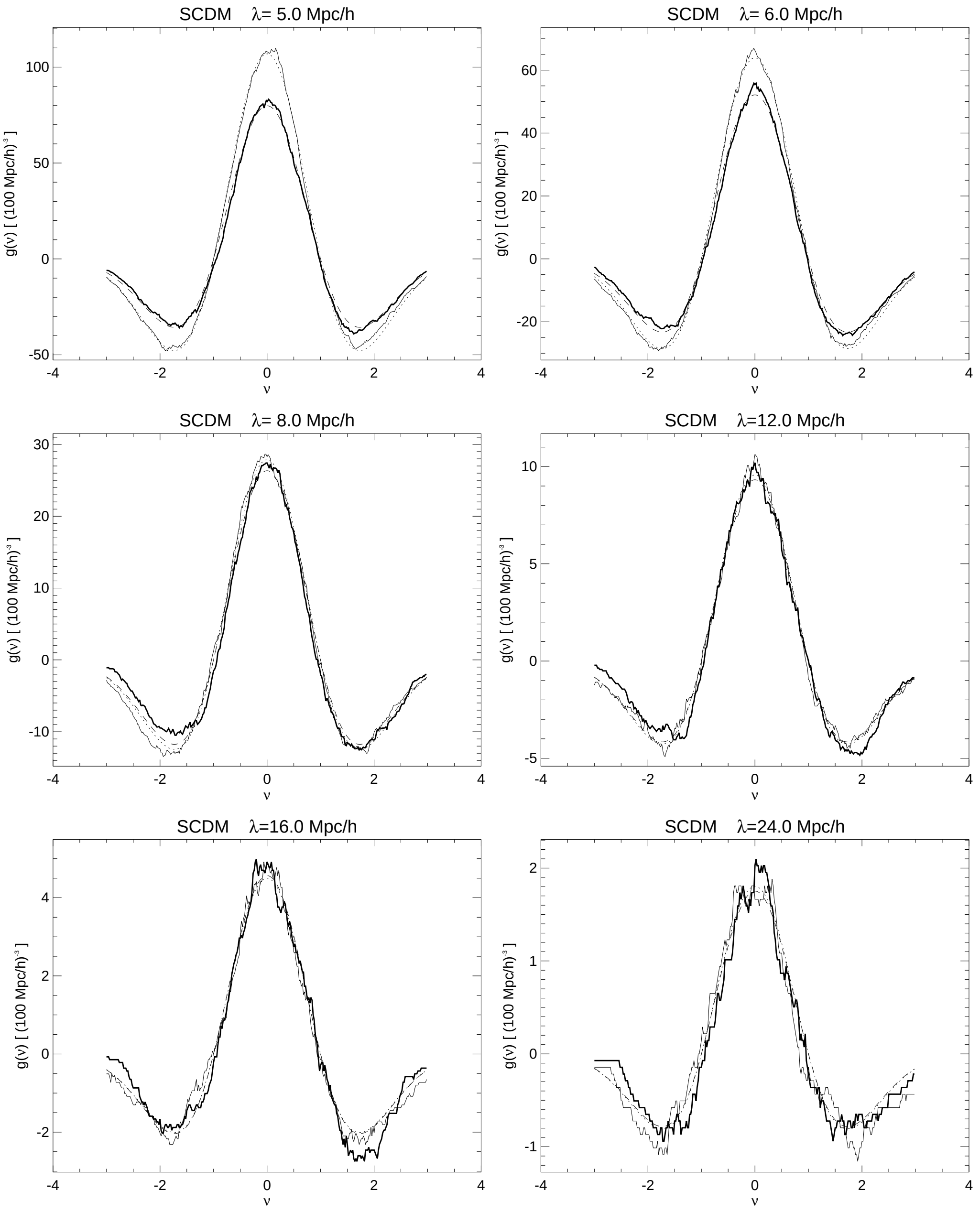


Figure 4.3: Some genus curves computed for the SCDM simulation.

and transformed back. In this way a realization of a Gaussian random field with the same power spectrum as the original field is obtained. This field is then smoothed in the same way as the original one and the genus curves are calculated as before.

In figures 4.3, 4.4 and 4.5 I show some of the genus curves obtained for the

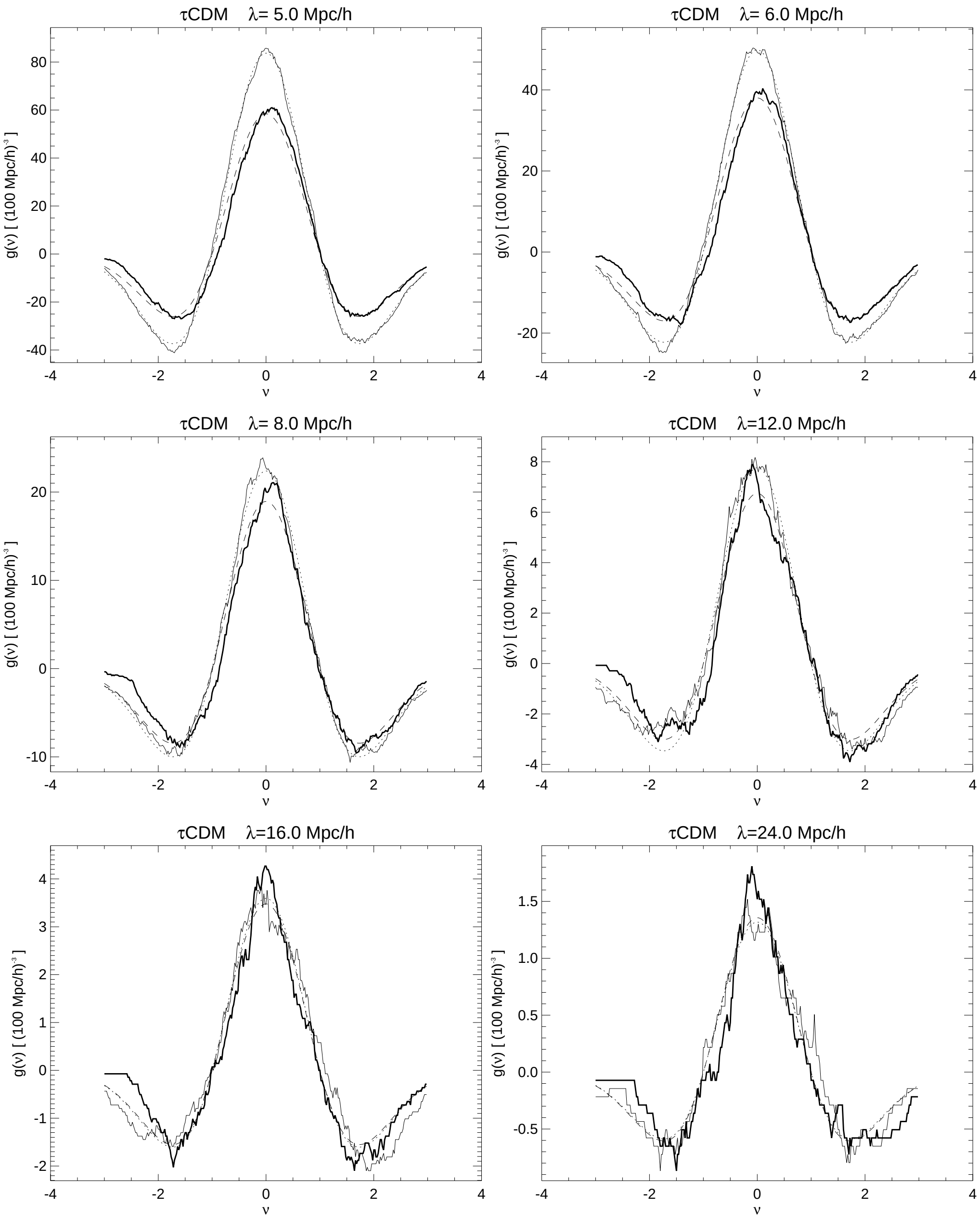


Figure 4.4: Selected genus curves for the τCDM simulation.

SCDM, τCDM and ΛCDM models. In each panel the heavy solid line gives the genus curve of the density field of the N-body simulation, whereas the thin curve arises as the genus of the Gaussianized density field. Additionally, best-fit random phase genus curves for the measurements are shown as dashed and dotted lines, respectively.

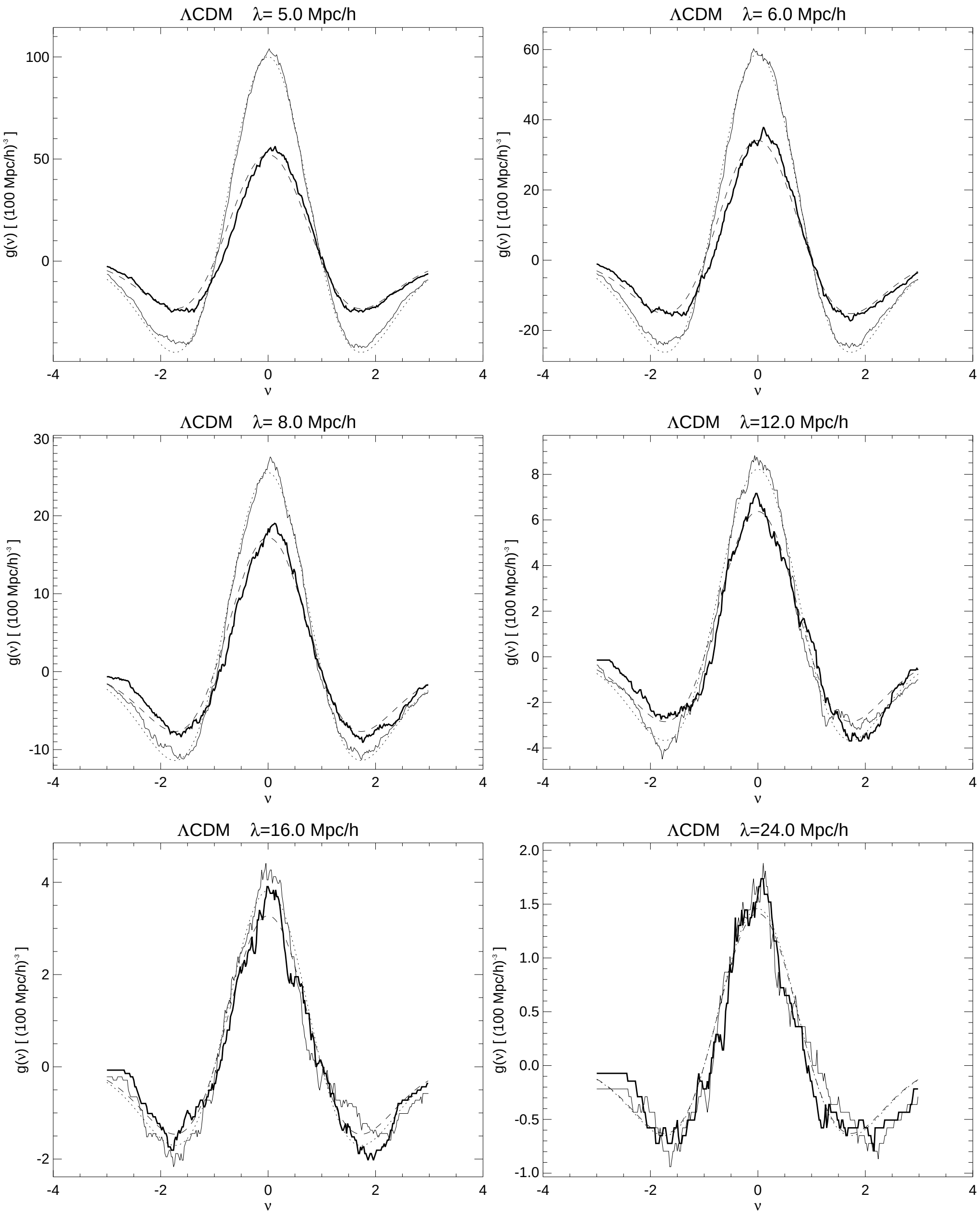


Figure 4.5: Genus curves for the ΛCDM simulation.

There are a number of interesting observations that can be made. In general, all the genus curves follow the w-shape of the random phase curve very well. However, at small smoothing lengths there is a bubble shift present in the ΛCDM model and somewhat weaker in the τCDM model. This shift is absent in the SCDM model.

Apart from that only the comparison with the corresponding genus curves for

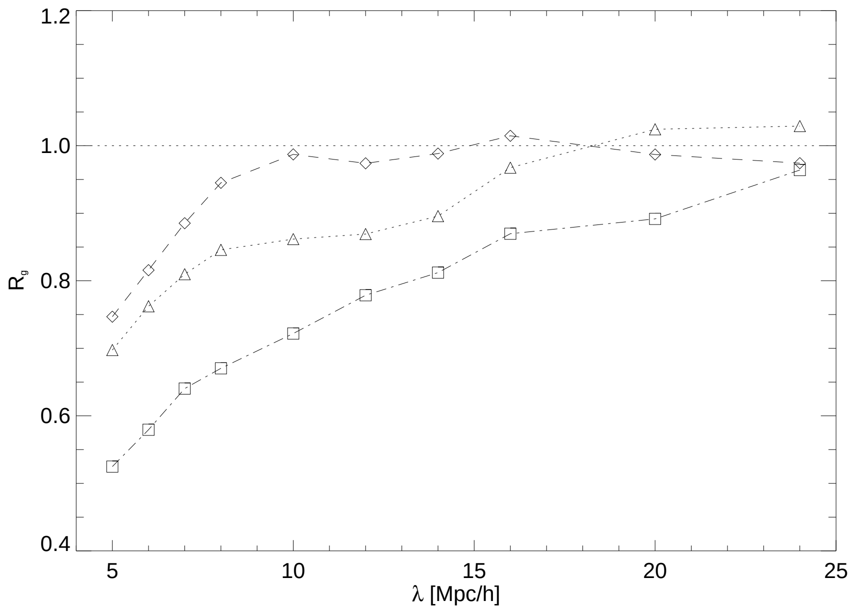


Figure 4.6: Amplitude drop R_{g} for the N-body simulations. The data points are for the SCDM model ($\Diamond$), the τCDM model ($\triangle$), and the ΛCDM model ($\square$).

the Gaussianized fields shows a striking difference for the three models. All of them show a drop of the genus amplitude at $\lambda = 5\,h^{-1}$Mpc. This drop is largest for the ΛCDM model with $R_{\mathrm{g}} \approx 0.5$, whereas it is only $R_{\mathrm{g}} \approx 0.85$ for the SCDM model. The τCDM model shows an intermediate value of $R_{\mathrm{g}} \approx 0.7$. With increasing smoothing length the amplitude drop becomes smaller, but the ΛCDM model continues to show phase correlations up to $24\,h^{-1}$Mpc whereas the SCDM is already quite Gaussian at $10\,h^{-1}$Mpc and the τCDM model at $16\,h^{-1}$Mpc.

In figure 4.6 I show the measured amplitude drop for the three models. It is apparent that they show quite different behaviour. The SCDM and τCDM models have very similar normalizations, so the difference between them cannot simply be attributed to different normalizations.

However, the ΛCDM model is normalized somewhat higher, so one might naturally expect stronger phase correlations in its density field. But the different normalization cannot account for all of the difference in phase correlations seen between the ΛCDM model and the other two CDM variants. This is demonstrated in figure 4.7 which shows the amplitude drop R_{g} as a function of the variance σ^2 of the smoothed density field. At the same σ^2 the ΛCDM model has developed stronger phase correlations than the τCDM and SCDM models. Even the τCDM and SCDM models can be clearly distinguished in this diagram.

One can also try to visualize the effect of the phase correlations that cause the

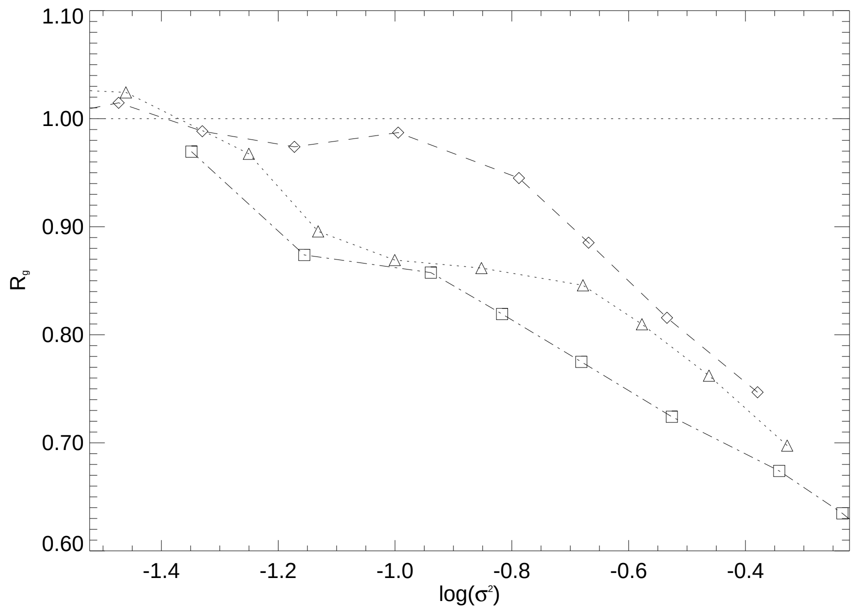


Figure 4.7: Amplitude drop $R_{\rm g}$ for the N-body simulations as a function of the variance σ^2. The data points are for the SCDM model ($\Diamond$), the τCDM model ($\triangle$), and the ΛCDM model ($\Box$).

amplitude drop. In figure C.4 I show isodensity surfaces that enclose the upper 4.1% and lower 4.1% of the ΛCDM density field smoothed at $\lambda = 5\,h^{-1}$Mpc. These volume fractions correspond to the minima of the genus curve at $\nu \pm \sqrt{3}$. For comparison also shown are the contours for the corresponding Gaussianized field. Although the power spectra of both fields are identical, the morphology and topology of the isodensity contours are clearly quite different. The ΛCDM model has developed structures that are much more coherent than in its corresponding Gaussian random field with the same power spectrum.

In the SCDM or τCDM scenarios this difference can be expected to be less pronounced because these models show weaker phase correlations. Especially the SCDM model should therefore produce structures that are choppier than in the ΛCDM model.

4.4 Direct comparison of PSCz with N-body models

It is interesting to compare the results obtained for the genus statistics of the PSCz redshift survey with the corresponding measurements for the N-body simulations. In this way one might hope to be able to constrain the models with the observational data.

Of course, there are numerous ways how such a comparison might be done. Because the theoretical models have many adjustable degrees of freedom (for example, shape and normalization of the power spectrum, matter content, or background cosmology) a complete scan of the parameter space of cold dark matter models in an attempt to find the best-fitting one requires an extensive systematic study. For simplicity I will restrict myself to a direct comparison between the genus statistics of PSCz and the three CDM simulations at hand. In particular, I want to quantify the amount of agreement between PSCz and each of the three variants of the CDM cosmogonies with the aim of finding a best candidate among them. For this purpose I perform for each of the models the following tests:

- A direct comparison of the measured genus amplitudes. The genus amplitude is sensitive to the second moment of the power spectrum and to possible phase correlations in the density field. I quantify the match between PSCz and a model by calculating

$$\chi_N^2 = \sum_i \frac{(N_i^{(\mathrm{PSCz})} - N_i^{(\mathrm{model})})^2}{\mathrm{var}(N_i^{(\mathrm{PSCz})})}. \tag{4.7}$$

  With respect to PSCz I use the measured amplitudes with the shot noise correction applied.

- A comparison of the shape of $\sigma^2(\lambda)$. This test is sensitive to the shape of the power spectrum. I quantify the agreement by evaluating

$$\chi_\sigma^2 = \sum_i \frac{(\sigma_i^{2\,(\mathrm{PSCz})} - \sigma_i^{2\,(\mathrm{model})})^2}{\mathrm{var}(\sigma_i^{2\,(\mathrm{PSCz})})}. \tag{4.8}$$

  Here I first rescale $\sigma^{2\,(\mathrm{model})}$ such that χ_σ^2 is minimized. In this way the test is sensitive only to the shape of $\sigma^2(\lambda)$ and not to its normalization.

- A comparison of the amplitude drop R_g. This test is sensitive to the phase correlations that develop in the density field during the gravitational evolution. I quantify the agreement by computing the quantity

$$\chi_R^2 = \sum_i \frac{(R_{\mathrm{g},i}^{(\mathrm{PSCz})} - R_{\mathrm{g},i}^{(\mathrm{model})})^2}{\mathrm{var}(R_{\mathrm{g},i}^{(\mathrm{PSCz})})}. \tag{4.9}$$

  Again, the measurements of R_g for the PSCz survey involve an application of the shot noise correction.

In figure 4.8 I show the measured genus amplitudes for the three CDM models and the results for PSCz. It is obvious that the SCDM model does a pretty poor job in fitting the observations whereas both the τCDM and ΛCDM model can be reconciled with the observations much better.

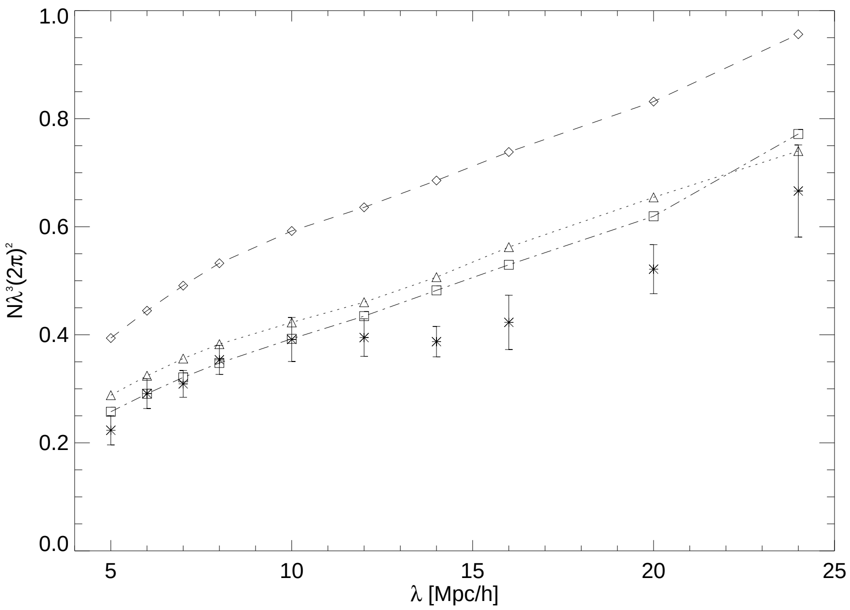


Figure 4.8: Comparison of the amplitude of the genus curves. Shown are the results for the PSCz survey ($*$ and error bars), the SCDM simulation ($\diamond$), the τCDM simulation ($\triangle$) and the ΛCDM simulation ($\square$).

Figure 4.9 shows a comparison of the measured variances of the density fields. Here I allowed a rescaling by a constant factor in order to be less dependent on the different normalizations of the models. It is clearly seen that the SCDM model has not enough power on large scales. It seems therefore that the power spectrum of the SCDM model is in strong disagreement with the power spectrum of PSCz.

Finally I display in figure 4.10 a comparison between the amplitude drop measured for the models and for PSCz. Since the PSCz redshift survey does not exhibit strong phase correlations in the range $5\,h^{-1}\mathrm{Mpc} \leq \lambda \leq 10\,h^{-1}\mathrm{Mpc}$ the ΛCDM model disagrees quite strongly with the observational data whereas the SCDM and τCDM models can be easily reconciled with the measurements.

These observations can be put on a more quantitative basis with the help of the χ^2 values defined in equations (4.7), (4.8) and (4.9). The results for them are listed in table 4.5. It appears that the SCDM model is highly inconsistent with PSCz simply because the shape of its power spectrum doesn't match that of PSCz. Consequently the genus amplitudes and the form of the $\sigma^2(\lambda)$ curve is in disagreement with the data.

Both the τCDM and the ΛCDM model do much better in this respect. They are roughly able to fit the form of the power spectrum of PSCz. The τCDM fits slightly better than the ΛCDM model, but an increase of the Λ-contribution in the latter

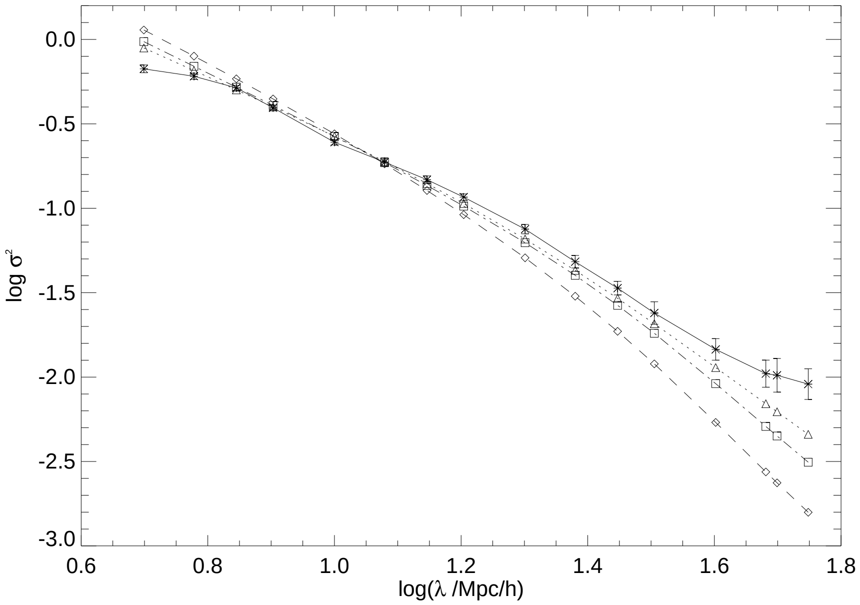


Figure 4.9: Comparison of the shape of $\sigma^2(\lambda)$ between PSCz and the simulations. Shown are the results for the PSCz survey ($*$ and error bars), the SCDM simulation ($\diamond$), the τCDM simulation ($\triangle$) and the ΛCDM simulation ($\square$). The functions $\sigma^2(\lambda)$ for the simulations have been rescaled by a constant factor to give a best fit to PSCz.

	SCDM	τCDM	ΛCDM
Genus amplitude: χ^2_N	451.07	51.02	25.22
Shape of $\sigma^2(\lambda)$: χ^2_σ	549.79	75.00	161.30
Amplitude drop: χ^2_R	9.36	5.03	25.74

Table 4.5: Some quantitative tests for agreement between PSCz and the examined N-body simulation.

case could easily make up for that. However, the test of the amplitude drop favours the τCDM model, simply because such strong phase correlations as in the ΛCDM model are not detected in PSCz.

In summary the τCDM model provides the best match to the observational results for the PSCz survey. A test based on the shape of the power spectrum favours CDM models with $\Gamma \approx 0.2$. The topological analysis with its measure for phase correlations can distinguish between a pure dark matter model or a model with cosmological constant. The latter develops a stronger drop in the genus amplitude than a corresponding model without vacuum energy density.

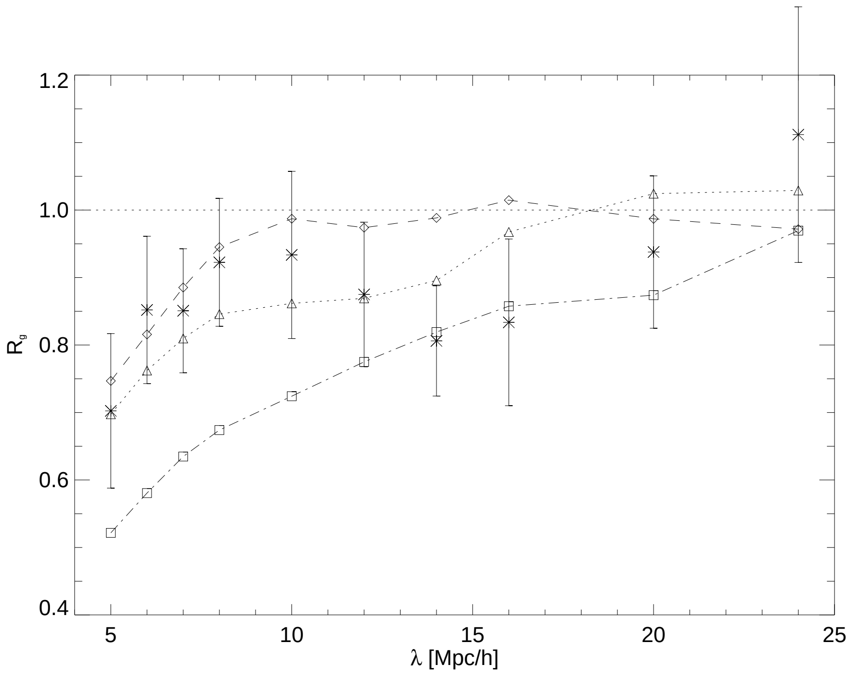


Figure 4.10: Comparison of the amplitude drop due to phase correlations in the density field. Shown are the results for the PSCz survey ($*$ and error bars), the SCDM simulation ($\diamond$), the τCDM simulation ($\triangle$) and the ΛCDM simulation ($\square$).

As of yet I have not examined an open cold dark matter model. This will be an interesting alternative to the τCDM and the ΛCDM models. Currently the Virgo consortium is working on a simulation of such a scenario and its analysis will be left to future work.

The amplitude drop seems to be a new powerful discriminant between different cosmological models. However, I have not demonstrated yet that this amplitude drop can be really measured as precisely as was claimed above. For example, one could imagine that the sparse sampling in a redshift survey wipes out some of the phase information in the topology and that $R_{\rm g} \simeq 0$ is favoured even for an underlying density field with strong phase correlations.

By examining a mock observation of the ΛCDM universe I will show in section 5.2 that the amplitude drop can in fact be measured with a redshift survey as dense as PSCz and would have been most likely detected in the analysis of PSCz if it had been present in the same strength as found for the ΛCDM model. Hence, some trust in the results above is justified and the non-detection of strong phase correlations in PSCz is a potentially serious problem for ΛCDM-like models.

Chapter 5

Observation of a Model Universe

5.1 Mock catalogues

Let me use the term *mock catalogue* for an artificial redshift survey of a model universe with known statistical properties. Typically I will use the result of a N-body simulation, which I *observe* to arrive at a galaxy catalogue that simulates the same selection criteria with respect to sky coverage, selection function and luminosity distribution as encountered in the PSCz survey.

The mock catalogues are ideal tools to examine the power and reliability of the various statistical estimators used in this work and they provide evidence for the correctness of the computer codes and the procedures used in the analysis.

In this section I explain how mock catalogues might be constructed. I will use them in section 5.2 to compare genus curves of a fully sampled N-body simulation with the genus curves of a mock survey derived from it. This will represent a test of the complete analysis scheme applied in chapters 2 and 3 to the PSCz redshift survey. Additionally I have used mock catalogues to test various estimators applied in part II of this work.

My method to construct a mock catalogue proceeds in two steps. First, galaxy positions are generated according to a prescribed selection function $S(z)$ which gives the desired mean density $\langle m(\mathbf{r}) \rangle = S(z)$ of sources. In a second step an observed flux is assigned to each of these galaxy sites. If one wants to make a catalogue that exhibits evolution of the luminosity function or has artificial errors in the flux (e.g. simulation of Malmquist bias) this second step can be easily modified to match these requirements.

I start with a stream of positions $\mathbf{r}_i$ which represent possible galaxy sites. For the N-body simulations I identify these positions with the 16.7 million particles in the simulation box. In this case the galaxies are assumed to directly trace the mass. Biased mock catalogues might be produced if one preselects a subset of these positions according to some biasing scheme. A location for the observer is then selected and the cosmological redshifts z_i of the potential galaxy sites with respect to the observer are calculated. I then draw random numbers x_i uniformly distributed in the interval $[0,1[$ and keep only the galaxies sites with

$$x_i \leq S(z_i), \tag{5.1}$$

where $S(z)$ is the selection function scaled such that $S(z) \leq 1$. If the selection function diverges for $z = 0$ I scale $S(z)$ such that $S(z_0) = 1$ for some very small z_0 and set $S(z) = 1$ for $z < z_0$.

The selection (5.1) leads to a galaxy catalogue with mean density proportional to $S(z)$, as desired. If necessary one can degrade the number of remaining positions randomly to the desired source density or throw out sources covered by a prescribed angular mask.

In a second step I now want to assign observed fluxes to the galaxies in the catalogue. The conditional probability that a galaxy observed at redshift z_i has luminosity L_i smaller than L_0 is

$$p(L_i \leq L_0) = \frac{\int_{L_{\min}(z_i)}^{L_0} \Phi_{z_i}(L)\,\mathrm{d}L}{\int_{L_{\min}(z_i)}^{\infty} \Phi_{z_i}(L')\,\mathrm{d}L'}. \tag{5.2}$$

A detailed discussion of the definition of $\Phi_z(L)$, $L_{\min}(z)$, $S(z)$, and some other quantities used below can be found in section 6.2. Thus, if a luminosity L_i is assigned via the equation

$$q_i = \frac{\int_{L_{\min}(z_i)}^{L_i} \Phi_{z_i}(L)\,\mathrm{d}L}{\int_{L_{\min}(z_i)}^{\infty} \Phi_{z_i}(L')\,\mathrm{d}L'}\,, \tag{5.3}$$

where q_i is a random number drawn from a uniform distribution in the interval $[0,1[$, the correct luminosity distribution (5.2) is reproduced.

Equation (5.3) can be written as

$$q_i = 1 - \frac{1}{S(z_i)} \int_{L_i}^{\infty} \Phi_{z_i}(L)\,\mathrm{d}L. \tag{5.4}$$

If $z_{\max}$ is defined by $L_i = L_{\min}(z_{\max})$ and no evolution in the luminosity function is assumed one finds

$$z_{\max} = S^{-1}\left[(1 - q_i)S(z_i)\right]. \tag{5.5}$$

This equation can be easily generalized to include evolution in the luminosity function, if desired.

Once $z_{\max}$ is determined one can finally assign a flux f_i with equation (7.11). Adopting an Einstein-de-Sitter universe and a straight K-correction (see section 7.1) with $\alpha = -2$ this reads like

$$f_i = f_{\min} \frac{r_{\max}^2}{r_i^2} \left(\frac{1+z}{1+z_i}\right)^{1-\alpha}. \tag{5.6}$$

A galaxy survey exhibits spherical symmetry, whereas N-body simulations usually live on periodic boxes. To arrive at a realistic mock survey I therefore replicated the simulation box periodically at all sides such that the desired survey volume was fully covered by the simulation. Note that the simulation volumes of $(240\,h^{-1}\mathrm{Mpc})^3$ used in the Virgo project are large, but PSCz probes an even larger volume of the real Universe.

5.2 A test observation of a model universe

Mock observations of model universes are interesting for at least two reasons. First, they can be used as a test of the procedures and codes that I used in the analysis of the PSCz redshift survey. By comparing the results for the genus statistics of the artificial redshift survey with the results for the fully sampled N-body simulation one can check that the systematic influences due to boundaries and sampling noise do not affect the outcome significantly.

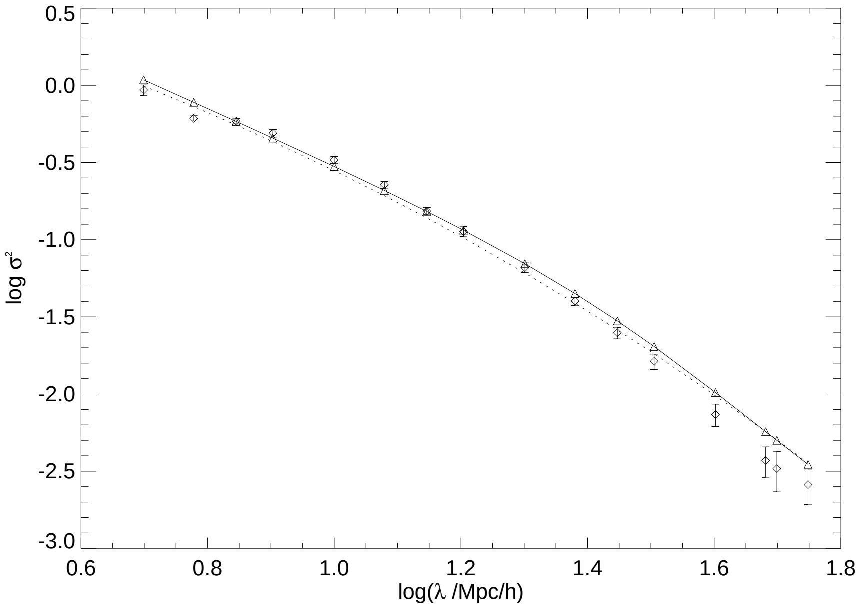


Figure 5.1: Comparison of the measured variance for the ΛCDM simulation and a mock survey of it. Shown is the result for the mock survey (◇ and error bars) and the underlying full simulation (△ and solid line). The dashed line results from a fit of the measured variances to the CDM power spectrum and is used to estimate the logarithmic slope of $\sigma^2(\lambda)$.

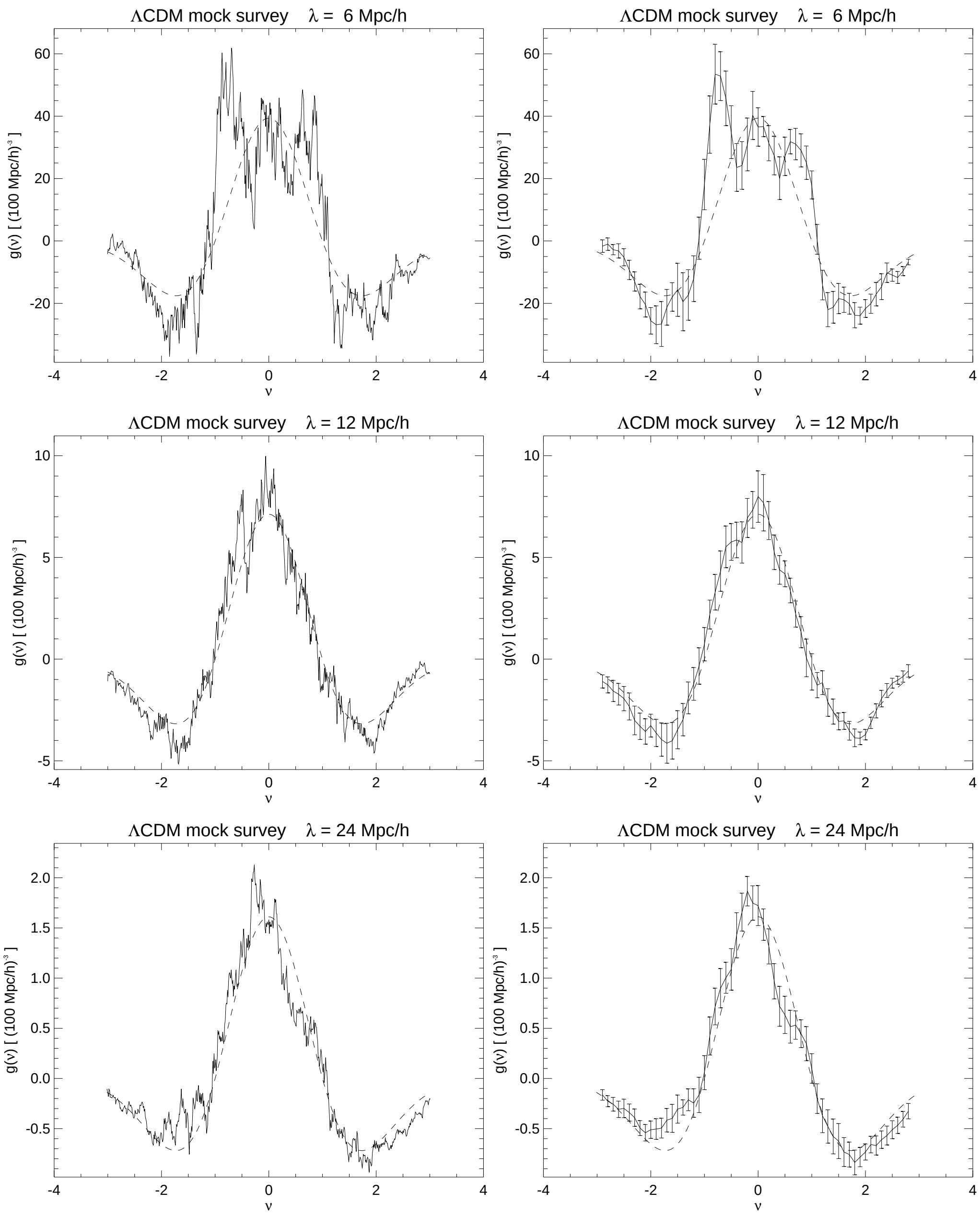


Figure 5.2: Selected genus curves for a mock observation of the ΛCDM simulation. For each of the three smoothing lengths the left panel displays a high resolution genus curve and the right panel gives the best estimate for the genus curve in the form of a boxcar average. The error bars are derived from 15 bootstrap resamplings of the mock survey. Also shown is the best-fit random phase curve (dashed). For brevity the genus curves for the 13 other smoothing lengths that have been computed are not given here.

Second, comparisons between mock observations of N-body universes and real redshift surveys could even be carried out if strong systematic influences existed. This is valid as long as the mock catalogue has the same selection function and source density as the real redshift survey and is analyzed in exactly the same way. All systematic influences are then already folded into this treatment and one can simply test how likely it is to observe the real redshift survey in a population of mock surveys of different cosmological models.

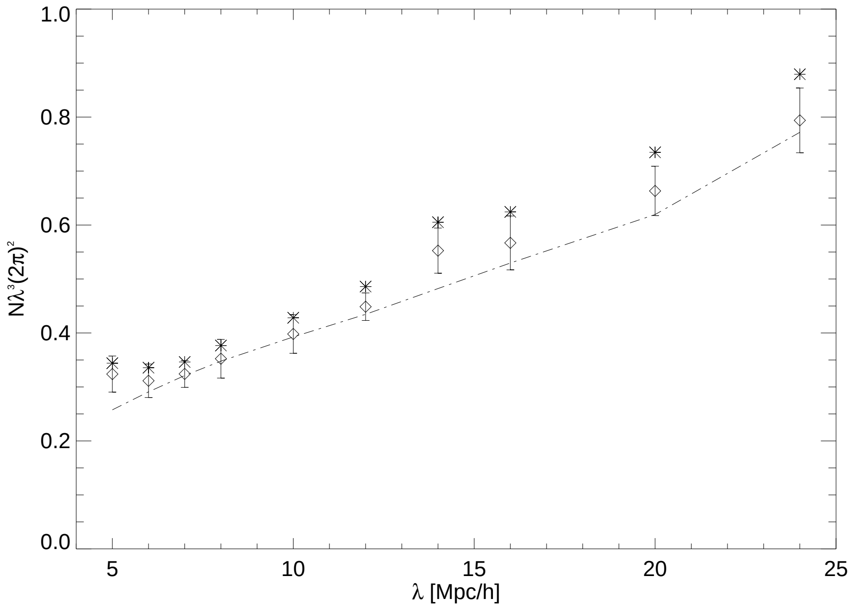


Figure 5.3: Comparison of the genus amplitude N measured for the artificial ΛCDM survey ($\diamond$ with error bars) and the full simulation (dashed line). Also shown are the genus amplitudes without shot noise correction ($*$).

In this section I want to use a mock observation in the first sense in order to demonstrate that the procedures applied to the PSCz survey are really powerful enough to detect phase correlations if they are present in the density field. Recall that I have found for the PSCz survey only a small amplitude drop which is particularly hard to reconcile with the large drop measured for the ΛCDM model. It is therefore especially interesting to check whether a mock observation of the ΛCDM universe with the same sampling density as the PSCz survey can reproduce the results obtained for the fully sampled N-body simulation.

For this purpose I have constructed a mock catalogue with 14120 galaxies to a depth of $360\,h^{-1}$Mpc from the ΛCDM model. The location of the observer was chosen randomly and the selection function used to construct the survey was taken as found for the PSCz survey (equation (8.2)). Hence, the source density was practically

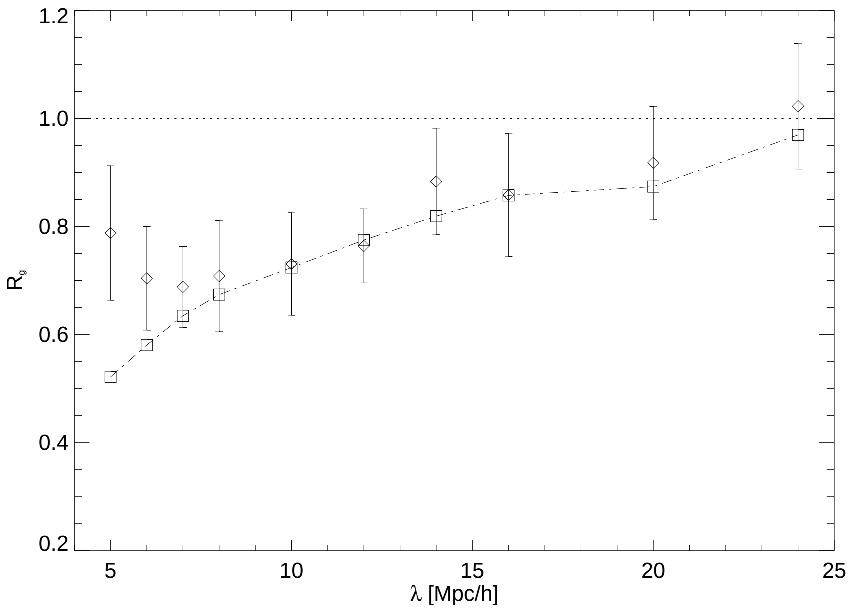


Figure 5.4: Comparison of the amplitude drop $R_{\rm g}$ measured for the ΛCDM mock survey (◇ with error bars) and the fully sampled ΛCDM simulation (□).

identical to PSCz. Additionally the same angular mask as in the PSCz survey was imposed.

I now applied the very same computer codes used in the analysis of PSCz to this mock catalogue. The results can then be compared to the findings for the fully sampled ΛCDM simulation.

In figure 5.1 I show the measured variance $\sigma^2(\lambda)$ for the mock catalogue and compare it to the true variance of the ΛCDM simulation. The agreement is quite good and the differences are in the range of the expected statistical uncertainty.

It should be noted, however, that the error estimates for the measured values of σ^2 do not include cosmic variance. Because of it adjacent points on the $\sigma^2(\lambda)$ curve are not independent, since their computation is largely based on the same survey volume. On the other hand it can be seen that the ΛCDM simulation has very prominent voids and clusters, hence the cosmic variance can be quite large. For example, the somewhat lower power measured in figure 5.1 on very large scales is not a systematic effect but can be attributed to cosmic variance. I have convinced myself of this by examining other mock catalogues with different observer positions.

In principle, an ensemble of such mock observations could be used to reduce the cosmic variance. Here I will refrain from this and see how well the formalism does under the more difficult circumstances of only one observation. After all, this is also the situation we encounter with respect to the real Universe.

A few of the resulting genus curves for the mock observation are shown in figure 5.2. As expected the noise in the curves is low over the range of threshold values where the number of resolution elements is large and it is relatively high at $\lambda = 5\,h^{-1}$Mpc and $\lambda = 56\,h^{-1}$Mpc, where the number of resolution elements is small.

Most interesting perhaps is the question whether one can measure the genus amplitude of the underlying simulation correctly. In figure 5.3 I compare the measured amplitude with the true amplitude of the full N-body simulation. It can be seen that the agreement is quite good, in particular for the range of intermediate smoothing lengths where one expects the cosmic variance to be small. The plot also clearly demonstrates that the shot noise correction does indeed improve the agreement between measured genus amplitude and true genus amplitude.

Finally, I compare the amplitude drop $R_{\rm g}$ measured from the mock survey with the true drop computed from the fully sampled simulation. As is evidenced by figure 5.4 the amplitude decrement is indeed measured with reasonable accuracy. The result is clearly consistent with the true amplitude drop for the ΛCDM simulation and inconsistent with finding no amplitude drop.

It is reassuring that these results compare very well with the fully sampled simulation in the range of the statistical uncertainty. I take this as evidence that the analysis I used for the PSCz survey can correctly recover topological properties of the underlying density field. In particular, a possible amplitude drop due to phase correlations in the density field can be detected.

Part II

Selection Function, Luminosity Function and Evolution

Chapter 6

Methods and Theory

6.1 Introduction

The PSCz redshift survey of IRAS galaxies is a purely flux limited sample. One of the inherent properties of any flux limited catalogue is the fast decay of the observed number density of objects with distance. This is a simple consequence of the $1/r^2$-behaviour of the observed flux density with distance to the source. The further one goes out the more luminous an object has to be in order to be included in the catalogue.

The expected number density of galaxies as a function of distance for a particular redshift survey is described by the so-called selection function. Shortly I will discuss how the selection function can be used to estimate the mass density field of the Universe from the number counts of galaxies. The reconstruction of the three-dimensional density field is one of the principal objectives of this work. Hence it is desirable to understand the selection function of the PSCz redshift survey as well as possible.

Part II of this work is mainly intended to provide a reliable selection function for the topological analysis already treated in part I. In order to determine the selection function I will use maximum likelihood techniques which are described extensively in the following sections. In the course of the analysis of PSCz I found evidence for rather strong evolution in the luminosity function of IRAS galaxies. This is a very interesting problem in its own right. Because of that and for the sake of an accurate selection function I have investigated this aspect in some detail. In particular, a thorough discussion of possible systematic errors that can influence a quantitative analysis of evolution is given as well.

Results for the selection function, for evolution and for the luminosity function are presented in chapter 8.

6.2 Definitions

Let the field $n_z(\mathbf{r}, L)\,\mathrm{d}L$ describe the comoving number density of IRAS galaxies at epoch z, in the luminosity interval $[L, L + \mathrm{d}L]$ and at spatial position $\mathbf{r}$.

In the following I will adopt the assumption that the luminosity distribution of IRAS galaxies is independent of clustering and morphology. Certainly this is not true in detail, since small differences in the luminosity functions of different galaxy classes (for example early and late types) have been detected (Loveday, Peterson, Efstathiou & Maddox 1992). However, the IRAS sample is observed in the far-infrared and these infrared selected sources represent a distinct class of their own which I simply call *IRAS galaxies* in this work. They seem to form a fairly homogeneous class of objects, so that the universal luminosity function hypothesis seems to be reasonably well satisfied. Some support for that will also be provided by the result discussed in section 8.2.

Given the above assumption the luminosity and spatial parts in the distribution $n_z(\mathbf{r}, L)$ factorize. The number density field can then be written as

$$n_z(\mathbf{r}, L) = \frac{n_z(\mathbf{r})}{\bar{n}_z}\Phi_z(L), \tag{6.1}$$

where the luminosity function $\Phi_z(L)$ is introduced to describe the universal distribution of luminosities. Here $n_z(\mathbf{r}) = \int n_z(\mathbf{r}, L)\,\mathrm{d}L$ is the local number density and the ensemble average $\bar{n}_z = \langle n_z(\mathbf{r})\rangle$ over many realizations of the universe gives the average number density of galaxies. Note that the luminosity function can still depend on epoch z. The luminosity function is normalized as

$$\int_0^\infty \Phi_z(L)\,\mathrm{d}L = \bar{n}_z\,. \tag{6.2}$$

I now define the selection function $S(z)$ as the mean comoving number density of galaxies that one *expects* to see in the catalogue at redshift $z = z(|\mathbf{r}|)$. Then $S(z)$ is given by

$$S(z) = \int_{L_{\min}(z)}^\infty \Phi_z(L)\,\mathrm{d}L\,, \tag{6.3}$$

where $L_{\min}(z)$ gives the minimum luminosity a source at redshift z must have in order to be included in the catalogue. The *observed* number density $m(\mathbf{r})$ of galaxies is then given by

$$m(\mathbf{r}) = \int_{L_{\min}(z)}^\infty n_z(\mathbf{r}, L)\,\mathrm{d}L = \frac{n_z(\mathbf{r})}{\bar{n}_z}S(z). \tag{6.4}$$

Again, z is given as $z = z(|\mathbf{r}|)$ in this equation and by definition the relation $\langle m(\mathbf{r})\rangle = S(z)$ holds.

The assumption that galaxies trace the mass without bias means that fluctuations in the number density field are directly proportional to the fluctuations in the underlying mass density field, viz.

$$\frac{\rho(\mathbf{r})}{\bar{\rho}} = \frac{n_z(\mathbf{r})}{\bar{n}_z}. \tag{6.5}$$

What is actually meant by this equation is that the mass density field acts as a probability density for placing galaxies in a Poisson point process. Only after sufficient smoothing equation (6.5) might be taken literally.

If one wants to recover the mass density fluctuations from the galaxy counts equations (6.4) and (6.5) imply that the galaxies have to be weighted with the inverse of the selection function:

$$\frac{\rho(\mathbf{r})}{\bar{\rho}} = \frac{m(\mathbf{r})}{S(z)}. \tag{6.6}$$

This shows clearly why an accurate determination of the selection function is important for the ability to reconstruct the mass density field. Any systematic bias in $S(z)$ has the potential to introduce an equally strong bias in the density field.

As I will demonstrate the IRAS galaxies show evidence for such strong evolution in their far-infrared luminosity function that one can hardly neglect it in the calculation of the selection function. For that reason I will include possible evolution in my formalism for finding the selection function.

Roughly one can distinguish between two types of evolution. Pure density evolution changes only the normalization of the luminosity function but not its shape. Pure luminosity evolution on the other hand describes a possible change of the intrinsic luminosities of the sources while the number density of them stays fixed.

In a realistic scenario for galaxy evolution both forms are likely to occur simultaneously. For example, mergers of galaxies reduce the number density of galaxies and cause density evolution in this way. However, such a merger event will also lead to a different luminosity of the combined system and hence cause luminosity evolution. Other physical mechanisms that give rise to evolution can easily be imagined. For example, sources might fade away with time or starburst activities could lead to episodes of increases infrared brightness.

It will turn out that the PSCz survey is not deep enough to discern between density and luminosity evolution since they have to first order degenerate observational consequences. It will therefore suffice to examine the limiting cases of pure density and pure luminosity evolution separately.

A further discussion of the physical interpretation of the evolutionary results can be found in section 8.3.

Pure density evolution

In the following I will describe density evolution in the luminosity function by

$$\Phi_z(L) = g(z)\Phi_0(L). \tag{6.7}$$

The functional form of $g(z)$ cannot be determined by the PSCz data precisely, so any reasonable form could be used. I will adopt the choice of previous authors and use the scaling

$$g(z) = (1+z)^P \tag{6.8}$$

in this work.

Pure luminosity evolution

For simplicity I will assume in this case that all sources change their luminosity in the same way, that is

$$L_z = f(z) L_0 \,, \tag{6.9}$$

where L_z is the luminosity of a source at epoch z that shines with L_0 at present time. The luminosity function evolves then according to

$$\Phi_z(L) = \frac{1}{f(z)} \Phi_0 \left(\frac{L}{f(z)} \right) . \tag{6.10}$$

The prefactor $1/f(z)$ is needed in order to conserve the comoving number density of sources. Again, any reasonable form for $f(z)$ will suffice to describe the evolution to first order in z. I will typically adopt

$$f(z) = \mathrm{e}^{\frac{2}{3} Q \left(1 - \frac{1}{(1+z)^{3/2}} \right)} \tag{6.11}$$

which corresponds to an exponential decay

$$L(t) = L(t_0) \, \mathrm{e}^{-\frac{t-t_0}{\tau}} \tag{6.12}$$

of luminosity (assuming an Einstein-de-Sitter universe) with time constant

$$\tau = \frac{1}{Q H_0} . \tag{6.13}$$

In previous work the ansatz $f(z) = (1+z)^Q$ has also been used. Since this is identical to equation (6.11) to first order in z this functional form will lead to very similar final results, although they will be somewhat smaller due to the difference in second order.

6.3 Maximum likelihood estimate of the selection function

Each of the galaxies in the PSCz catalogue is essentially described by a redshift z_i with corresponding position $\mathbf{r}_i = \mathbf{r}_i(z_i, \Omega_i)$ and a luminosity L_i, which can be derived from the measured flux(es). I now imagine that the catalogue is drawn from an underlying parent distribution given by

$$m(\mathbf{r}, L) = \begin{cases} n_z(\mathbf{r}, L) & \text{for } L \geq L_{\min}(z) \\ 0 & \text{for } L < L_{\min}(z) \end{cases} , \tag{6.14}$$

where $z = z(\mathbf{r})$.

The conditional probability $p(z_i|L_i)\,\mathrm{d}L$ that a source observed at redshift z_i falls into the luminosity range $[L_i, L_i + \mathrm{d}L]$ can be described by

$$p(L_i|z_i)\,\mathrm{d}L = \frac{m(\mathbf{r_i}, L_i)\,\mathrm{d}L}{\int_{L_{\min}(z_i)}^{\infty} m(\mathbf{r_i}, L')\,\mathrm{d}L'}\,. \tag{6.15}$$

Here one could also replace $m(\mathbf{r}, L)$ by $n_z(\mathbf{r}, L)$. The denominator simply gives the available number of galaxies at that distance and the numerator counts the number of galaxies in the particular luminosity range. Upon insertion of equations (6.1) and (6.3) this boils down to

$$p(L_i|z_i) = \frac{\Phi_{z_i}(L_i)}{S(z_i)}. \tag{6.16}$$

Note that the density fluctuations have dropped out of this expression. This insensitivity to density inhomogeneities makes the maximum likelihood techniques used throughout this work superior compared to older methods like the ordinary $V/V_{\max}$ estimator introduced by Schmidt (1968).

If one now maximizes the likelihood

$$\mathcal{L} = \prod_i p(L_i|z_i) \tag{6.17}$$

of the whole data set with respect to $S(z)$ one gets an estimate of the selection function that is not systematically biased by local density fluctuations.

In order to find this maximum in practice I first express $\Phi_{z_i}(L_i)$ in terms of the selection function. Here the desired model for the evolution in the luminosity function has to be taken into account. For simplicity I will not consider mixed models of evolution but only the limiting cases of pure density and pure luminosity evolution. Because the sample at hand is not deep enough to distinguish between the two forms anyway this is not a serious restriction.

For the purpose of comparison and for possible usage in other surveys I first treat the case with no evolution and then generalize it to include evolution.

No evolution

Here $\Phi_z(L) = \Phi_0(L)$ is assumed independent of epoch. Then the selection function is

$$S(z) = \int_{L_{\min}(z)}^{\infty} \Phi_0(L)\,\mathrm{d}L, \tag{6.18}$$

and its derivative becomes

$$S'(z) = -\Phi_0\left(L_{\min}(z)\right) L'_{\min}(z). \tag{6.19}$$

Now I define z_i^{m} such that

$$L_{\min}(z_i^{\mathrm{m}}) = L_i\,, \tag{6.20}$$

that is, z_i^{m} is equal to the maximal redshift the source could be moved to while still remaining in the sample. If $S'(z)$ is evaluated at z_i^{m} one finds

$$S'(z_i^{\mathrm{m}}) = -\Phi_{z_i}(L_i) L'_{\mathrm{min}}(z_i^{\mathrm{m}}). \tag{6.21}$$

This can be written as

$$p(L_i|z_i) = \frac{-S'(z_i^{\mathrm{m}})}{S(z_i)} \frac{1}{L'_{\mathrm{min}}(z_i^{\mathrm{m}})}. \tag{6.22}$$

It is computationally more convenient to work with the logarithm

$$\Lambda = \log \mathcal{L} \tag{6.23}$$

of the likelihood instead of the likelihood itself. The task is now to find the global maximum of

$$\Lambda = \sum_i \log \left(\frac{-S'(z_i^{\mathrm{m}})}{S(z_i)} \right) + \sum_i \log \left(\frac{1}{L'_{\mathrm{min}}(z_i^{\mathrm{m}})} \right) \tag{6.24}$$

with respect to the function $S(z)$. Note that the second sum can be dropped since it doesn't depend on $S(z)$.

In order to accomplish this maximization in practice one has to model the functional form of $S(z)$ in terms of a number of parameters so that Λ can be optimized with respect to these coefficients.

Generally two basic procedures can be distinguished in this context. In the so-called parametric maximum likelihood estimate (Sandage, Tammann & Yahil 1979, Yahil, Strauss, Davis & Huchra 1991) one assumes an analytic form for $S(z)$ that depends on a few (typically two to four) parameters α_k, which are then found by solving

$$\frac{\partial \Lambda}{\partial \alpha_k} = 0\,. \tag{6.25}$$

The maximum likelihood method also allows an estimate of the statistical uncertainties in the so derived parameters (Kendall & Stuart 1979). Asymptotically the distribution of $\mathcal{L}$ is a multivariate Gaussian around the true values of the parameters. If the information matrix I is defined as

$$I_{ij} = -\frac{\partial^2 \Lambda}{\partial \alpha_i \partial \alpha_j} \tag{6.26}$$

then the covariance matrix of the estimates is given by

$$\mathrm{cov}(\alpha_i, \alpha_j) = (I^{-1})_{ij} \tag{6.27}$$

evaluated at the maximum of Λ. This result holds asymptotically for large sample sizes. As a corollary I note that the $1\,\sigma$ statistical confidence interval can be obtained from the drop of the likelihood function Λ by 0.5 from its peak value if the fit involves only one free parameter.

A disadvantage of the maximum likelihood technique is that there is no built-in measure of goodness of fit. Almost any functional form can be made to 'fit', although the function may provide only a poor description of the data. This technique therefore requires an *a priori* knowledge of a reasonable fitting form for $S(z)$. One can partly work around this problem if the absolute value of the likelihood (which is not meaningful of its own) is used to compare different functional forms. In this way one could try many functional forms and use the one as the best estimate that gives the highest overall likelihood.

Still, the above method is unsatisfactory if one has no precise notion of the expected shape of $S(z)$. In this case one can allow for a very flexible shape of $S(z)$ by describing it by many parameters, for example by a series of step functions. This so called step-wise non-parametric maximum likelihood method (Nicholl & Segal 1983, Efstathiou, Ellis & Peterson 1988) has been used to find luminosity functions (e.g. Efstathiou et al. 1988, Loveday et al. 1992, Lin et al. 1996) and the run of radial density with distance (Saunders et al. 1990, Loveday et al. 1992). I will also apply this technique for an estimate of the run of radial density.

With respect to the selection function I will propose a new non-parametric method, that models $S(z)$ by a series of continuously linked power laws. It turns out that this method can not only very accurately approximate any shape of the selection function but also has a number of computational advantages. Later on I will also extend this estimator to a new non-parametric luminosity function estimator which exhibits significant advantages compared to the step-wise method which is currently used as the standard technique.

But before I begin with that let me first finish up this section and discuss what has to be changed in the estimator (6.24) if one allows for evolution.

Pure density evolution

Here I assume the model

$$\Phi_z(L) = g(z)\Phi_0(L) \tag{6.28}$$

for the evolution of the luminosity function. Then the selection function becomes

$$S(z) = g(z)\int_{L_{\min}(z)}^{\infty} \Phi_0(L)\,\mathrm{d}L, \tag{6.29}$$

so that one gets for its derivative

$$S'(z) = -g(z)\Phi_0\left(L_{\min}(z)\right)L'_{\min}(z) + \frac{g'(z)}{g(z)}S(z). \tag{6.30}$$

Again, I define the maximal redshift z_i^{m} such that

$$L_{\min}(z_i^{\mathrm{m}}) = L_i\,. \tag{6.31}$$

If then $S'(z)$ is evaluated at the position z_i^{m} one gets

$$S'(z_i^{\mathrm{m}}) = -\frac{g(z_i^{\mathrm{m}})}{g(z_i)}\Phi_{z_i}(L_i)L'_{\min}(z_i^{\mathrm{m}}) + \frac{g'(z_i^{\mathrm{m}})}{g(z_i^{\mathrm{m}})}\,S(z_i^{\mathrm{m}})\,, \tag{6.32}$$

so that the probability of equation (6.16) can be written in the form

$$p(L_i|z_i) = \left(\frac{-S'(z_i^{\mathrm{m}})}{S(z_i)} + \frac{g'(z_i^{\mathrm{m}})}{g(z_i^{\mathrm{m}})} \frac{S(z_i^{\mathrm{m}})}{S(z_i)} \right) \frac{g(z_i)}{g(z_i^{\mathrm{m}})} \frac{1}{L'_{\mathrm{min}}(z_i^{\mathrm{m}})}. \tag{6.33}$$

Then one finally has to maximize

$$\begin{aligned} \Lambda &= \sum_i \log \left(\frac{-S'(z_i^{\mathrm{m}})}{S(z_i)} + \frac{g'(z_i^{\mathrm{m}})}{g(z_i^{\mathrm{m}})} \frac{S(z_i^{\mathrm{m}})}{S(z_i)} \right) + \sum_i \log \left(\frac{g(z_i)}{g(z_i^{\mathrm{m}})} \right) \\ & \quad - \sum_i \log \left(\frac{L'_{\mathrm{min}}(z_i^{\mathrm{m}})}{L_{\mathrm{min}}(z_i^{\mathrm{m}})} \right) - \sum_i \log L_i \end{aligned} \tag{6.34}$$

with respect to $S(z)$ and $g(z)$. Note that the constant sums in the second line can be dropped.

The above form suggests that one can maximize Λ simultaneously for $g(z)$ and $S(z)$. Unfortunately this is not a fruitful idea. If one recalls that the maximum likelihood estimator of $S(z)$ is independent of density inhomogeneities it becomes clear that one in fact should not be able to detect pure density evolution with this estimator. The reason for this is that there is no absolute way to distinguish density evolution from density inhomogeneities that arise from large-scale structure.

It is easy to verify this feature in this formalism if one takes a second look at the probability of equation (6.16). For the density evolution model it can be written as

$$p(L_i|z_i) = \frac{g(z_i)\Phi_0(L_i)}{\int_{L_{\mathrm{min}}(z_i)}^{\infty} g(z_i)\Phi_0(L)\,\mathrm{d}L} = \frac{\Phi_0(L_i)}{\int_{L_{\mathrm{min}}(z_i)}^{\infty} \Phi_0(L)\,\mathrm{d}L}. \tag{6.35}$$

So the function $g(z)$ drops out completely and there is no sensitivity to density evolution with this estimator. This means that one can only determine an optimal $S(z)$ for a prescribed model for density evolution. In the limit of large data sets the likelihoods of different solutions are then equal for different $g(z)$. Hence one will have to resort to another method to actually determine the amount of density evolution.

Pure luminosity evolution

Here I assume the model

$$\Phi_z(L) = \frac{1}{f(z)} \Phi_0 \left(\frac{L}{f(z)} \right) \tag{6.36}$$

for the evolution of the luminosity function. The selection function can then be written as

$$S(z) = \int_{\frac{L_{\mathrm{min}}(z)}{f(z)}}^{\infty} \Phi_0(L)\,\mathrm{d}L. \tag{6.37}$$

This implies

$$S'(z) = -\Phi_0 \left(\frac{L_{\mathrm{min}}(z)}{f(z)} \right) \frac{\mathrm{d}}{\mathrm{d}z} \left(\frac{L_{\mathrm{min}}(z)}{f(z)} \right). \tag{6.38}$$

Let me now define the maximal redshift z_i^{m} slightly differently as before:

$$\frac{L_{\min}(z_i^{\mathrm{m}})}{f(z_i^{\mathrm{m}})} = \frac{L_i}{f(z_i)}. \tag{6.39}$$

If $S'(z)$ is evaluated at z_i^{m} one gets this time

$$S'(z_i^{\mathrm{m}}) = -L_i \Phi_{z_i}(L_i) \left(\frac{L'_{\min}(z_i^{\mathrm{m}})}{L_{\min}(z_i^{\mathrm{m}})} - \frac{f'(z_i^{\mathrm{m}})}{f(z_i^{\mathrm{m}})} \right). \tag{6.40}$$

Again this result can be used to express the probability (6.16) entirely in terms of the functions $S(z)$ and $f(z)$:

$$p(z_i|L_i) = \frac{-S'(z_i^{\mathrm{m}})}{S(z_i) L_i \left(\frac{L'_{\min}(z_i^{\mathrm{m}})}{L_{\min}(z_i^{\mathrm{m}})} - \frac{f'(z_i^{\mathrm{m}})}{f(z_i^{\mathrm{m}})} \right)}. \tag{6.41}$$

Then one finally ends up with

$$\Lambda = \sum_i \log \frac{-S'(z_i^{\mathrm{m}})}{S(z_i)} - \sum_i \log \left(\frac{L'_{\min}(z_i^{\mathrm{m}})}{L_{\min}(z_i^{\mathrm{m}})} - \frac{f'(z_i^{\mathrm{m}})}{f(z_i^{\mathrm{m}})} \right) - \sum_i \log L_i\,, \tag{6.42}$$

which I want to maximize with respect to the function $S(z)$. The constant term $\sum \log L_i$ can be dropped. One might suspect that it should be possible here to simultaneously maximize Λ with respect to $f(z)$ and $S(z)$. After all, Λ depends in a non-trivial way on $f(z)$ and $S(z)$. Note that here the redshifts z_i^{m} also depend on $f(z)$. However, it turns out that the maximum in Λ is again degenerate with respect to $S(z)$ and $f(z)$, at least if the bright end of the luminosity function can be approximated by a power law. The somewhat lengthy proof for this is not given here. Therefore one has to find the amount of luminosity evolution with another estimator. Once this has been achieved equation (6.42) can be used to find an optimal estimate for the selection function.

6.3.1 A non-parametric estimator for the selection function

In this section I want to develop a non-parametric maximum likelihood estimator for the selection function. In particular, I am interested in an exact description of the shape of $S(z)$. Once the latter is determined it can be used to justify an empirical fit to $S(z)$. For that purpose I describe the selection function by a series of continuously linked power laws. This description seems rather appropriate since the selection function is a smooth continuous curve that covers a wide range of values and its local behaviour can be very well approximated by a power law. If a large number of steps is chosen this parameterization can follow very flexibly any shape the selection function might have.

So let me describe $S(z)$ by n pieces. Let $\{S_0, S_1, \ldots, S_{n-1}\}$ be the values of $S(z)$ at redshifts $z = \{x_0, x_1, \ldots, x_{n-1}\}$. Then bin 0 covers $0 < z \le x_0$, bin 1 covers $x_0 < z \le x_1$, and so forth.

In each bin k I describe $S(z)$ by a power law of the form

$$S(z) = S_k \left(\frac{z}{x_k}\right)^{m_k} \qquad \text{for } x_{k-1} < z \leq x_k\,, \tag{6.43}$$

where m_k gives the logarithmic slope in that particular bin. Since I want to constrain the shape of the selection function these slopes are precisely the quantities I am looking for.

The different pieces are required to join continuously. Recall that this is necessary because $S(z)$ is the integral over the luminosity function. The latter is a distribution function and doesn't have to be continuous, though. Continuity will be given if the S_k are related by

$$S_k = \begin{cases} S_0 & \text{for } k = 0 \\ S_0 \displaystyle\prod_{j=1}^{k} \left(\frac{x_j}{x_{j-1}}\right)^{m_j} & \text{for } 0 < k < n\ . \end{cases} \tag{6.44}$$

The derivative of $S(z)$ is given by

$$S'(z) = \frac{m_k}{z} S(z) \qquad \text{for } x_{k-1} < z \leq x_k, \tag{6.45}$$

where k is the number of the bin corresponding to redshift z. Let me further define a_i as the number of the bin belonging to the maximal redshift z_i^{m} of galaxy i. Similarly, let b_i be the number of the interval that encloses the redshift z_i of galaxy i.

Everything is now ready to find the maximum likelihood estimates of the slopes m_k. I will first consider the easiest case with no evolution and then generalize it to include density and luminosity evolution.

No evolution

For the purpose of this derivation it is most convenient to insert a term $\sum_i \log S(z_i^{\mathrm{m}}) - \sum_i \log S(z_i^{\mathrm{m}})$ of zero value in the likelihood equation and to rewrite it as

$$\Lambda = \sum_i \log \frac{-S'(z_i^{\mathrm{m}})}{S(z_i^{\mathrm{m}})} + \sum_i \log \frac{S(z_i^{\mathrm{m}})}{S(z_i)}. \tag{6.46}$$

With the above model this likelihood evaluates to

$$\begin{aligned} \Lambda \;&=\; \sum_i \log \frac{-m_{a_i}}{z_i^{\mathrm{m}}} + \sum_i \sum_{j=b_i+1}^{a_i} m_j \log \frac{x_j}{x_{j-1}} \\ &\quad + \sum_i m_{a_i} \log \frac{z_i^{\mathrm{m}}}{x_{a_i}} - \sum_i m_{b_i} \log \frac{z_i}{x_{b_i}}. \end{aligned} \tag{6.47}$$

As expected Λ depends only on the shape of $S(k)$ as described by the slopes m_k and not at all on the normalization given by S_0. The latter has to be determined with another method. The conditions for a maximum in Λ are now

$$\frac{\partial\Lambda}{\partial m_k} = \sum_i \frac{\delta_{k,a_i}}{m_k} + T_k = 0, \tag{6.48}$$

where I have defined

$$T_k = \sum_i \sum_{j=b_i+1}^{a_i} \delta_{j,k} \ln\frac{x_j}{x_{j-1}} + \sum_i \delta_{k,a_i} \ln\frac{z_i^{\mathrm{m}}}{x_{a_i}} - \sum_i \delta_{k,b_i} \ln\frac{z_i}{x_{b_i}}. \tag{6.49}$$

If I further introduce the number n_k of galaxies with $a_i = k$, that is

$$n_k = \sum_i \delta_{k,a_i}\,, \tag{6.50}$$

the slopes are just given by

$$m_k = -\frac{n_k}{T_k}. \tag{6.51}$$

It's perhaps worth pointing out that this is a surprisingly simple result. In particular one can solve for each of the m_k independent of all the others. There is no correlation between the values of the slopes in different bins. This fact makes their calculation rather easy and also greatly simplifies the estimation of errors in the different m_k.

Because the T_k don't explicitly depend on any of the m_k the information matrix is simply found to be

$$I_{lk} = -\frac{\partial^2\Lambda}{\partial m_l \partial m_k} = \delta_{l,k}\frac{n_k}{m_k^2}. \tag{6.52}$$

One can now trivially invert this diagonal matrix and find the error estimates for the m_k as

$$\mathrm{var}(m_k) = \sigma^2(m_k) = \frac{m_k^2}{n_k}. \tag{6.53}$$

As pointed out earlier the normalization of $S(z)$ is lost in this method, so it has to be determined separately. For the reconstruction of the density fluctuation field $\delta(\mathbf{r})$, however, it is not even needed. It makes therefore sense to divide the error in $S(z)$ in a shape part resulting from the uncertainties in the m_k and a normalization part mainly resulting from counting statistics.

With an analytic model for the selection function I primarily want to achieve an optimal description of its shape. Therefore I will suggest in section 6.3.2 a method that directly fits the local slope of a model to the measured m_k. The shape problem becomes strictly separated from the normalization problem in this way.

Density evolution

First I rewrite the likelihood (6.34) as

$$\begin{aligned}\Lambda &= \sum_i \log\left(\frac{-S'(z_i^{\mathrm{m}})}{S(z_i^{\mathrm{m}})} + \frac{g'(z_i^{\mathrm{m}})}{g(z_i^{\mathrm{m}})}\right) + \sum_i \log\frac{S(z_i^{\mathrm{m}})}{S(z_i)} \\ &\quad + \sum_i \log\frac{g(z_i)}{g(z_i^{\mathrm{m}})} - \sum_i \log\frac{L'_{\mathrm{min}}(z_i^{\mathrm{m}})}{L_{\mathrm{min}}(z_i^{\mathrm{m}})}.\end{aligned} \tag{6.54}$$

With an evolutionary ansatz $g(z) = (1+z)^P$ this gives for the power law model a likelihood

$$\begin{aligned}\Lambda &= \sum_i \log\left(\frac{-m_{a_i}}{z_i^{\mathrm{m}}} + \frac{P}{1+z_i^{\mathrm{m}}}\right) + \sum_i \sum_{j=b_i+1}^{a_i} m_j \log\frac{x_j}{x_{j-1}} \\ &\quad + \sum_i m_{a_i} \log\frac{z_i^{\mathrm{m}}}{x_{a_i}} - \sum_i m_{b_i} \log\frac{z_i}{x_{b_i}} \\ &\quad + \sum_i P \log\frac{1+z_i}{1+z_i^{\mathrm{m}}} - \sum_i \log\frac{L'_{\min}(z_i^{\mathrm{m}})}{L_{\min}(z_i^{\mathrm{m}})}. \end{aligned} \tag{6.55}$$

The best estimates for the m_k are then found by solving the equations

$$\frac{\partial\Lambda}{\partial m_k} = \sum_i \frac{\delta_{k,a_i}}{m_k - P\frac{z_i^{\mathrm{m}}}{1+z_i^{\mathrm{m}}}} + T_k = 0, \tag{6.56}$$

where T_k is defined as before. These equations are not fully linear in m_k, but *almost*. Since the redshift bins are narrow, replacing z_i^{m} by x_{a_i} will give a useful first approximation $\tilde{m}_k$ to the true solution m_k. This starting value can be calculated from

$$\tilde{m}_k = -\frac{n_k}{T_k} + P\frac{x_k}{1+x_k} \tag{6.57}$$

and then quickly be improved by iterating equation (6.56) with a Newton method. After 2 to 3 iterations a relative accuracy $< 10^{-10}$ is already reached.

As pointed out earlier the above estimator allows only the determination of an optimal $S(z)$ for a prescribed value of the evolutionary parameter P, but it cannot be used to infer the actual amount of density evolution.

Luminosity evolution

Again I first rewrite the likelihood function as

$$\begin{aligned}\Lambda &= \sum_i \log\left(\frac{-S'(z_i^{\mathrm{m}})}{S(z_i^{\mathrm{m}})}\right) + \sum_i \log\frac{S(z_i^{\mathrm{m}})}{S(z_i)} \\ &\quad - \sum_i \log\left(\frac{L'_{\min}(z_i^{\mathrm{m}})}{L_{\min}(z_i^{\mathrm{m}})} - \frac{f'(z_i^{\mathrm{m}})}{f(z_i^{\mathrm{m}})}\right). \end{aligned} \tag{6.58}$$

For the luminosity evolution I will adopt the model

$$f(z) = \mathrm{e}^{\frac{2}{3}Q\left(1-\frac{1}{(1+z)^{3/2}}\right)}. \tag{6.59}$$

Here one should keep in mind that the maximal redshifts z_i^{m} are calculated differently as before and that their values depend on Q. This is a major difference to the case of density evolution where the z_i^{m} don't depend on P.

I will consider luminosity evolution only together with straight K-corrections of slope α. In this case I show in section 7.3 that

$$\frac{L'_{\mathrm{min}}(z)}{L_{\mathrm{min}}(z)} = \frac{1}{(1+z)^{3/2} - (1+z)} + \frac{1-\alpha}{1+z} \tag{6.60}$$

for an Einstein-de-Sitter universe. The likelihood for the stepwise power law model can then be written as

$$\begin{aligned}
\Lambda \quad = \quad & \sum_i \log \frac{-m_{a_i}}{z_i^{\mathrm{m}}} + \sum_i \sum_{j=b_i+1}^{a_i} m_j \log \frac{x_j}{x_{j-1}} \\
& + \sum_i m_{a_i} \log \frac{z_i^{\mathrm{m}}}{x_{a_i}} - \sum_i m_{b_i} \log \frac{z_i}{x_{b_i}} \\
& - \sum_i \log \left(\frac{1}{(1+z_i^{\mathrm{m}})^{3/2} - (1+z_i^{\mathrm{m}})} + \frac{1-\alpha}{1+z_i^{\mathrm{m}}} - \frac{Q}{(1+z_i^{\mathrm{m}})^{5/2}} \right).
\end{aligned} \tag{6.61}$$

The equations for the slopes are again as easy as in the case of no evolution. One simply has

$$m_k = -\frac{n_k}{T_k} \tag{6.62}$$

with the definitions of n_k and T_k as given in equations (6.49) and (6.50), respectively.

6.3.2 An analytic fit to the selection function

In the last section I have introduced a non-parametric estimator for the selection function that essentially determines the local logarithmic slope of it. The piecewise description as a series of power laws still contains statistical uncertainty in the slopes m_k of the individual pieces. Therefore it is desirable to derive a smooth analytic fit to the selection function in order to simplify subsequent analysis and to reduce the noise in the description of its shape.

Given a theoretical model $S(z)$ with a number s of free parameters one can define a local logarithmic slope as

$$m(z) \equiv \frac{d \log S}{d \log z} = z \frac{S'(z)}{S(z)}, \tag{6.63}$$

and fit it to the m_k by minimizing χ^2:

$$\chi^2 = \sum_k \frac{(m(x_k) - m_k)^2}{\sigma^2(m_k)}. \tag{6.64}$$

In this way an analytic fit to the shape of the selection can be found independent of its normalization.

The quantity (6.64) should follow a usual χ^2 distribution with ν degrees of freedom where ν is the number of steps minus the number s of fitted parameters. If the function $S(z)$ provides an adequate fit to the data the reduced $\chi^2_\nu \equiv \frac{\chi^2}{\nu}$ should be close to one. So the χ^2 test provides a measure for the goodness of the fit itself along with a best estimate of the parameters and confidence intervals of them. The latter can be derived for each parameter by varying it around its best estimate while simultaneously allowing the other parameters to find their optimal values for each of these chosen values. Then the $1\,\sigma$ confidence intervals can be estimated by an increase of 1 in χ^2.

In this work I will fit the selection function to the analytic form used in the QDOT survey:

$$S(z) = \frac{\phi}{z^\alpha \left(1 + (\frac{z}{z^\star})^\gamma\right)^{\beta/\gamma}}. \tag{6.65}$$

This form is slightly more general than the selection function used by Yahil et al. (1991) in the 1.2-Jy survey and the results of section 8.1 show that it indeed provides an accurate description of the data.

Of course, the ansatz (6.65) can also be used to directly fit the parameters by finding the maximum likelihood in equation (6.34). I have also done that to test the overall consistency of my procedures (see section 8.1). However, in this way the ability to directly test the goodness of the fit is lost. And there is also another interesting application of the non-parametric estimator for the selection function. If the estimated slopes m_k are plotted as a function of redshift one can see out to which redshift the shape of the selection function is well constrained by the data. This might serve as an indicator for the maximal survey volume that can be reliably used for further analysis. See also the results in section 8.1 and 8.3 for a further discussion of this point.

6.3.3 Normalizing the selection function

In order to normalize the selection function I write it as

$$S(z) = \phi s(z), \tag{6.66}$$

where $s(z)$ is the shape of the selection function determined with the method of section 6.3.2. Then an unbiased estimate $\tilde{\phi}$ of the normalization ϕ is given by

$$\tilde{\phi} = \frac{\int m(\mathbf{r}) w(\mathbf{r})\, \mathrm{d}\mathbf{r}}{\int s(\mathbf{r}') w(\mathbf{r}')\, \mathrm{d}\mathbf{r}'} \tag{6.67}$$

for an arbitrary weight function $w(\mathbf{r})$. If one takes the expectation value of $\tilde{\phi}$ and recalls $\langle m(\mathbf{r})\rangle = S(z(|\mathbf{r}|)$ one indeed sees that $\left\langle \tilde{\phi} \right\rangle = \phi$, that is, the estimate is unbiased. Note that here and in subsequent formulae I employ the shorthand notation $s(\mathbf{r}) = s(z(|\mathbf{r}|))$.

Following Davis & Huchra (1982) I now want to choose the weight function $w(\mathbf{r})$ such that the estimator minimizes the expected variance

$$\sigma^2 = \left\langle \tilde{\phi}^2 \right\rangle - \left\langle \tilde{\phi} \right\rangle^2. \tag{6.68}$$

The obvious advantage of an estimator with this property is that it maximizes the probability to find an estimate $\tilde{\phi}$ that is close to the true value ϕ.

With the relation

$$\langle m(\mathbf{r}_1)m(\mathbf{m}_2)\rangle = S(\mathbf{r}_1)S(\mathbf{r}_2)\left(1+\xi(\mathbf{r}_1-\mathbf{r}_2)\right) + S(\mathbf{r}_1)\delta^{(\mathrm{D})}(\mathbf{r}_1-\mathbf{r}_2), \tag{6.69}$$

which has been shown by Feldman et al. (1994), one first arrives at

$$\sigma^2 = \frac{\phi^2 \int \mathrm{d}\mathbf{r}_1\,\mathrm{d}\mathbf{r}_2\, s(\mathbf{r}_1)s(\mathbf{r}_2)w(\mathbf{r}_1)w(\mathbf{r}_2)\xi(\mathbf{r}_2-\mathbf{r}_1) + \phi\int \mathrm{d}\mathbf{r}\, s(\mathbf{r})w^2(\mathbf{r})}{\left(\int s(\mathbf{r}')w(\mathbf{r}')\,\mathrm{d}\mathbf{r}'\right)^2}. \tag{6.70}$$

With a shift of the integration variable in the first integral in the numerator this becomes

$$\begin{aligned} I &= \int \mathrm{d}\mathbf{r}_1\,\mathrm{d}\mathbf{r}_2\, s(\mathbf{r}_1)s(\mathbf{r}_2)w(\mathbf{r}_1)w(\mathbf{r}_2)\xi(\mathbf{r}_2-\mathbf{r}_1) \\ &= \int \mathrm{d}\mathbf{r}_1\,\mathrm{d}\mathbf{r}_2 s(\mathbf{r}_1)w(\mathbf{r}_1)s(\mathbf{r}_1+\mathbf{r}_2)w(\mathbf{r}_1+\mathbf{r}_2)\xi(\mathbf{r}_2). \end{aligned} \tag{6.71}$$

If one now expands $s(\mathbf{r}_1+\mathbf{r}_2)w(\mathbf{r}_1+\mathbf{r}_2) = s(\mathbf{r}_1)w(\mathbf{r}_1) + \mathbf{r}_2\cdot\nabla sw|_{\mathbf{r}_1} + \cdots$ one finds that the linear term gives a vanishing contribution since $\xi(\mathbf{r})$ has only a radial dependence. Furthermore, $\xi(\mathbf{r})$ contributes strongly only if $|\mathbf{r}|$ is smaller than about one correlation length. Hence it is a good approximation to write

$$I \approx \int \mathrm{d}\mathbf{r}_1\,\mathrm{d}\mathbf{r}_2\, s(\mathbf{r}_1)w(\mathbf{r}_1)s(\mathbf{r}_1)w(\mathbf{r}_1)\xi(\mathbf{r}_2) = J_3\int \mathrm{d}\mathbf{r}\, s^2(\mathbf{r})w^2(\mathbf{r}), \tag{6.72}$$

where

$$J_3 = \int_V 4\pi r^2\xi(r)\,\mathrm{d}r \tag{6.73}$$

is the second moment of the two-point correlation function. With this approximation one has

$$\sigma^2[w] = \frac{\phi^2 J_3\int \mathrm{d}\mathbf{r}\, s^2(\mathbf{r})w^2(\mathbf{r}) + \phi\int \mathrm{d}\mathbf{r}\, s(\mathbf{r})w^2(\mathbf{r})}{\left(\int s(\mathbf{r}')w(\mathbf{r}')\,\mathrm{d}\mathbf{r}'\right)^2}. \tag{6.74}$$

The optimal weight function is now given by the zero of the functional derivative

$$\frac{\delta\sigma^2[w(\mathbf{r})]}{\delta w(\mathbf{r}_0)} = 0. \tag{6.75}$$

This condition evaluates to

$$(J_3\phi w(\mathbf{r}_0)s(\mathbf{r}_0)+w(\mathbf{r}_0))\int s(\mathbf{r})w(\mathbf{r})\,\mathrm{d}\mathbf{r} = \int (J_3\phi w(\mathbf{r})s(\mathbf{r})+w(\mathbf{r}))s(\mathbf{r})w(\mathbf{r})\,\mathrm{d}\mathbf{r} \tag{6.76}$$

and is solved by

$$w(\mathbf{r}) = \frac{1}{1+J_3\tilde{\phi}s(\mathbf{r})}, \tag{6.77}$$

as is easily checked by direct insertion.

Equation (6.77) gives the weight function that leads to a minimum variance estimate of the normalization. The observed galaxy field $m(\mathbf{r})$ is a sum

$$m(\mathbf{r}) = \sum_i \delta^{(\mathrm{D})}(\mathbf{r} - \mathbf{r}_i), \tag{6.78}$$

over Dirac δ-functions so that the normalization estimate is finally given by

$$\tilde{\phi} = \frac{\sum_i w(\mathbf{r}_i)}{\int_V s(\mathbf{r}) w(\mathbf{r})\,\mathrm{d}\mathbf{r}}. \tag{6.79}$$

Note that this equation has to be solved iteratively since $w(\mathbf{r})$ depends (weakly) on $\tilde{\phi}$. One also needs a model for the two-point correlation function in order to estimate J_3. The observed $\xi(r)$ can be well approximated by a power law

$$\xi(r) = \left(\frac{r_0}{r}\right)^{\gamma} \tag{6.80}$$

with $r_0 = 3.87\,h^{-1}$Mpc and $\gamma = 1.11$ on scales $r \leq 25\,h^{-1}$Mpc (Moore et al. 1994). Unfortunately a direct computation of J_3 with the form (6.80) would lead to a serious overestimation of J_3 because (6.80) is simply not valid for large scales. In fact, it is easy to show that $\xi(r)$ eventually becomes negative and that J_3 is zero for infinite survey volume.

Although the result of the estimator (6.79) is not very sensitive to the actual value of J_3, the error in $\tilde{\phi}$ as estimated by equation (6.74) would be overestimated when J_3 is too high. For that reason I insert the Fourier transform of $\xi(r)$ in equation (6.73) and obtain

$$J_3 = \frac{4\pi}{3} R^3 \int P(k) \frac{3 j_1(kR)}{kR}\,\mathrm{d}^3 k, \tag{6.81}$$

where R denotes the radius of the survey volume. In order to estimate a value for J_3 I assume a CDM power spectrum with $\Gamma = 0.2$ and $\sigma_8 = 0.8$. The final result for the normalization is not sensitive to this choice.

6.4 A non-parametric estimator for the radial density

In the previous sections I have developed a maximum likelihood estimator for the selection function that is *independent of the density distribution.* Because the luminosity distribution was assumed to be universal it is also possible to find the density *independent of the luminosity distribution.*

In particular it is possible to develop a non-parametric estimator for the run of the radial density. Saunders et al. (1990) have first introduced this especially useful probe which can identify the very largest structures in the survey and can indicate evolution if present.

Let me start by neglecting any evolution. This means that a galaxy distribution

$$n_z(\mathbf{r}, L) = n_0(\mathbf{r}, L) = \frac{n_0(\mathbf{r})}{\bar{n}_0} \Phi_0(L) \tag{6.82}$$

is assumed. I now aim at a maximum likelihood determination of the density distribution. Since only a finite number of galaxies are available the density field has to be modeled by a smaller number of parameters in order to be able to extract useful information. One way to do this is to model it as a collection of finite volume bins, each of them having constant density inside. If one is primarily interested in the radial run of density the choice of radial shells is especially convenient. This also avoids any need to deal explicitly with the angular mask of the survey. Of course, other choices of volume bins are in principle possible.

In the following I will use the notation $n_0(r)$ to specify the angular averaged version of the corresponding field $n_0(\mathbf{r})$, e.g. $n_0(r) = \langle n_0(\mathbf{r})\rangle_\Omega$ and similarly for other fields.

The probability $p(r_i|L_i)\,\mathrm{d}r$ that a galaxy with luminosity L_i is located in the distance interval $[r_i, r_i + \mathrm{d}r]$ is then given by

$$p(r_i|L_i)\,\mathrm{d}r = \frac{4\pi r_i^2 n_z(r_i, L_i)\,\mathrm{d}r}{\int_0^{r_i^\mathrm{m}} n_z(r, L_i)\,\mathrm{d}V} = \frac{4\pi r_i^2 n_0(r_i)\,\mathrm{d}r}{\int_0^{r_i^\mathrm{m}} n_0(r)\,\mathrm{d}V}. \tag{6.83}$$

Here r_i^m denotes the maximal radial coordinate the source i with luminosity L_i could be moved to while still being seen above the flux limit. Of course, this coordinate corresponds via $r_i^\mathrm{m} = r(z_i^\mathrm{m})$ to the maximal redshift introduced earlier in equation (6.20).

My goal in this section is to derive a maximum likelihood estimate of the function $n_0(r)$. But in any flux limited redshift survey the number density of sources decays rapidly with increasing distance. For that reason one can only hope to accurately determine the radial density out to some maximal distance R_max. If one doesn't want to be highly influenced by the shot noise in the small number of more distant galaxies the likelihood estimator needs to be formulated in such a way that they are excluded right away. I will accomplish this with the following reformulation of equation (6.83).

The probability $p(r_i|L_i)\,\mathrm{d}r$ that a galaxy that is observed with luminosity L_i and coordinate $r_i < R_\mathrm{max}$ is located in the distance interval $[r_i, r_i + \mathrm{d}r]$ is given by

$$p(r_i|L_i)\,\mathrm{d}r = \frac{4\pi r_i^2 n_z(r_i, L_i)\,\mathrm{d}r}{\int_0^{r_i^\star} n_z(r, L_i)\,\mathrm{d}V} = \frac{4\pi r_i^2 n_0(r_i)\,\mathrm{d}r}{\int_0^{r_i^\star} n_0(r)\,\mathrm{d}V}, \tag{6.84}$$

where I have defined

$$r_i^\star = \min(r_i^\mathrm{m}, R_\mathrm{max}). \tag{6.85}$$

One now wants to maximize

$$\Lambda = \sum_i \log p(r_i|L_i) = \sum_i \log \frac{n_0(r_i)}{\int_0^{r_i^\star} n_0(r)\,\mathrm{d}V} \tag{6.86}$$

with respect to the function $n_0(r)$. Note that I have dropped the constant term $\sum 4\pi r_i^2$ in the expression for Λ and that the sum extends only over those galaxies that have $r_i < R_\mathrm{max}$.

I now model $n_0(r)$ in the range $[0, R_{\max}]$ as a series of step functions of height ρ_k corresponding to the density in the shell k with volume V_k. Let f_{ik} be the volume fraction that galaxy i occupies within shell k, viz.

$$f_{ik} = \begin{cases} 1 & \text{for } r_i^\star > R_k \\ \frac{r_i^{\star\,3} - R_{k-1}^3}{R_k^3 - R_{k-1}^3} & \text{for } R_{k-1} < r_i^\star < R_k \\ 0 & \text{for } r_i^\star < R_{k-1} \end{cases} \tag{6.87}$$

where R_{k-1} and R_k are the inner and outer radii of shell k, respectively. The integral in the denominator of equation (6.86) is then given by

$$\int_0^{r_i^\star} n_0(r)\,\mathrm{d}V = \sum_k \rho_k f_{ik} V_k\,, \tag{6.88}$$

so that Λ becomes

$$\Lambda = \sum_i \log \frac{\rho_{a_i}}{\sum_k \rho_k f_{ik} V_k}. \tag{6.89}$$

Here I introduced a_i to specify the number of the shell that hosts r_i. The maximal likelihood estimate is then given as solution of

$$\frac{\partial \Lambda}{\partial \rho_k} = \sum_i \frac{\delta_{a_i,k}}{\rho_k} - \sum_i \frac{V_k f_{ik}}{\sum_s \rho_s f_{is} V_s} = 0\,. \tag{6.90}$$

These coupled equations can be solved by direct iteration. For this purpose I calculate a new set $\tilde{\rho}_k$ from the ρ_k by

$$\tilde{\rho}_k = \left(\sum_i \frac{V_k f_{ik}}{\sum_s \rho_s f_{is} V_s} \right)^{-1} N_k\,, \tag{6.91}$$

where $N_k = \sum_i \delta_{a_i,k}$ gives the number of galaxies with r_i in bin k. The procedure can be repeated until convergence is achieved.

It should be noted that the absolute values of the ρ_k cannot be determined, but only their relative ratios. This is already seen in equation (6.83) where the overall normalization of $n_0(r)$ drops out. As a consequence exactly one of the equations (6.90) is not independent of the others. In principle one could therefore fix one of the ρ_k at an arbitrary value and only iterate the rest of them. This indeed works in practice but slows down convergence quite a lot. Therefore it is better to vary all ρ_k in a practical implementation of the iteration scheme.

6.5 An estimator for density evolution

In the previous section I have outlined a non-parametric estimator for the run of radial density with distance. The results obtained with it (see section 8.2) show evidence for rather strong evolution.

Let me assume that this evolution is pure density evolution of the form $\bar{n}_z = \bar{n}_0(1+z)^P$. To get an estimate of the strength P of this evolution one can replace $n_z(r)$ in equation (6.84) by the expected mean density $\bar{n}_{z(r)} = \bar{n}_0(1+z)^P$ at epoch $z = z(r)$ (Saunders et al. 1990, Fisher et al. 1992). Then one finds

$$\Lambda = \sum_i \log \frac{(1+z_i)^P}{\int_{z_{\min}}^{z_i^\star} (1+z)^P \frac{\mathrm{d}V}{\mathrm{d}z}\,\mathrm{d}z}\,, \tag{6.92}$$

where I have defined $z_i^\star = z(r_i^\star)$ and also allowed for a lower redshift cut-off $z_{\min}$. The latter can be used to reduce the effects of local density inhomogeneities on this estimator.

To find the maximum likelihood estimate of P the equation $\frac{\mathrm{d}\Lambda}{\mathrm{d}P} = 0$ has to be solved. This translates into

$$\sum_i \log(1+z_i) = \sum_i \frac{\int_{z_{\min}}^{z_i^\star} (1+z)^P \log(1+z) \frac{\mathrm{d}V}{\mathrm{d}z}\,\mathrm{d}z}{\int_{z_{\min}}^{z_i^\star} (1+z)^P \frac{\mathrm{d}V}{\mathrm{d}z}\,\mathrm{d}z}\,, \tag{6.93}$$

where the sums extend only over galaxies with $z_i \in [z_{\min}, z_{\max}]$.

The integral appearing in the denominator of equation (6.92) can be done analytically for an Einstein-de-Sitter universe and I evaluate it with the help of

$$\begin{aligned}\int_0^z (1+z')^P \frac{\mathrm{d}V}{\mathrm{d}z'}\,\mathrm{d}z' \quad &= \quad \frac{4\Delta\Omega}{H_0^3}\left(\frac{(1+z)^{P-1/2}-1}{P-1/2}\right. \\ &\qquad \left. -2\frac{(1+z)^{P-1}-1}{P-1} + \frac{(1+z)^{P-3/2}-1}{P-3/2}\right). \end{aligned} \tag{6.94}$$

For that reason it is easiest to find the maximum in Λ by calculating the likelihood with equation (6.92) for different values of P and then evaluating the peak with a parabolic fit to the computed points.

The estimator presented here is still affected by the large-scale structure present in the survey. Only if one had many universes to observe this cosmic variance could be reduced. Since the deviations of $n(r)$ from the homogeneous case are largest nearby previous authors have excluded the innermost redshift regions in order to reduce the distortions caused by large-scale structure in the estimate of P (Saunders et al. 1990, Fisher et al. 1992, Oliver et al. 1994).

I want to point out that an upper redshift cut-off should also be included. Similar as in the case of the radial density estimator the density field should not be used out to infinite distances because the very sparse sampling out there leads to dominance of shot noise. When no upper redshift cut-off is used this has only a random influence on the above estimator if the redshift survey is perfectly complete. However, if the survey has a problem with completeness at high redshift, which is quite typical, an upper redshift cut-off is strongly indicated since otherwise the result might be systematically underestimated.

Fisher et al.'s (1992) low result of $P = 2\pm3$ for the 1.2-Jy survey represents quite likely an example for this effect. Although the authors have had observed problems with completeness at high redshift and low galactic latitude they have not included an upper cut-off $R_{\max}$ in their estimator.

6.6 An alternative evolutionary estimator

The estimator discussed above has the disadvantage that it is only applicable to density evolution. Here I want to propose an alternative estimator that can be applied to any evolutionary model, e.g. to density evolution, luminosity evolution or mixed models, on an equal footing.

As was pointed out in section 6.3 there is no absolute way to distinguish density evolution from extremely shaped large-scale structure that accidentally takes the form of a spherically symmetric underdensity with our own galaxy at the center. However, the notion that the Universe should be homogeneous and isotropic on very large scales strongly suggests that the most plausible explanation for a radial rise in the density is an evolution in the luminosity function.

Under this hypothesis a determination of the evolutionary rate might be based on the request that the correct selection function should lead to a universe that looks closest to being homogeneous on very large scales. As a quantitative measure for deviations from homogeneity one can take the variance σ^2 of the density field smoothed on some large scale λ. Thus an extremal principle for the determination of the evolution might be formulated: The evolutionary rate can be estimated by finding the value that leads to the smallest large-scale power on very large scales.

Although computationally quite lengthy this scheme is easily feasible with the techniques already developed in this work. As a first step one can calculate the selection function with the methods of section 6.3 for many values of P (or Q). Each of these selection functions give rise to a density field smoothed on some large scale λ (for example, $\lambda \approx 50\,h^{-1}$Mpc). Finally, the variance estimator of section 3.3 can be applied to find the function $\sigma^2(P)$ (or $\sigma^2(Q)$) and the location of its minimum.

In section 8.3 I will discuss some results obtained with this estimator. As it turns out a plot of σ^2 as a function of P (or Q) reveals indeed a well defined minimum which can be taken as an estimate of the evolutionary rate. Reassuringly the results are very similar to those obtained with the density evolutionary estimator of the previous section. However, the alternative estimator can also be used to determine the luminosity evolution Q.

The error in the determination of P (or Q) might be estimated with the help of the statistical standard deviation for the σ^2 measurement which has already been derived with bootstrap methods. I estimate the confidence interval for P (or Q) by taking the range that leads to a variance σ^2 at most one standard deviation higher than the minimum value of σ^2.

6.7 A non-parametric estimator for the luminosity function

The selection function has been defined in equation (6.3) in terms of an integral over the luminosity function. If the selection function is known it is therefore possible to recover the underlying luminosity distribution, or vice versa. In particular one can readily derive a non-parametric luminosity function estimate from the selection function estimator of section 6.3.1.

The usual way to find a non-parametric maximum likelihood estimate of the luminosity function utilizes a parameterization by a series of step functions. The luminosity function is then determined directly (Efstathiou et al. 1988, Loveday et al. 1992, and others).

I believe that my new estimator has a number of advantages compared to this procedure. Let me first describe it in detail and compare its properties with other methods later.

Density evolution

Let me assume that the selection function has been determined with the non-parametric estimator of section 6.3.1 and that the function $L_{\min}(z)$ described in section 7.3 and its inverse $z_{\max}(L)$ is known. The piecewise description of the selection function directly translates into a piecewise description of the luminosity function. For this purpose I define the boundaries L_k of the luminosity intervals by

$$L_k = L_{\min}(x_k), \tag{6.95}$$

where x_k are the redshifts used in the piecewise description of the selection function.

Upon evaluation of equation (6.30) at $z = z_{\max}(L)$ one gets the present day luminosity function for the density evolution model as

$$\Phi_0(L) = \left(\frac{g'(z_{\max})}{g(z_{\max})} S(z_{\max}) - S'(z_{\max})\right) \frac{1}{g(z_{\max}) L'_{\min}(z_{\max})}. \tag{6.96}$$

This translates into the piecewise description

$$\begin{aligned}\Phi_0(L) \quad = \quad & \left(\frac{P}{1+z_{\max}} - \frac{m_k}{z_{\max}}\right) \left(\frac{z_{\max}}{x_k}\right)^{m_k} \\ & \times \frac{S_k}{(1+z_{\max})^P L'_{\min}(z_{\max})} \quad \text{for } L_{k-1} < L < L_k \; .\end{aligned} \tag{6.97}$$

Note that I dropped the argument L of $z_{\max}(L)$ in these equations in order to simplify the notation. Equation (6.97) represents a straightforward and easy way to construct the luminosity function from the non-parametric estimate of the selection function.

I see two main advantages of this method as compared to the non-parametric estimate of the luminosity function introduced by Efstathiou et al. (1988).

First, the shape of each luminosity bin follows approximately a power law, so it is able to adapt to the true shape of the luminosity function in a flexible way. Only very small discontinuities at the boundaries of bins remain and the estimate of $\Phi_0(L)$ of equation (6.97) directly traces the luminosity function. The conventional estimator on the other side uses rectangular bins which lead to large discontinuous jumps in the luminosity function. Efstathiou et al. (1988) show that the heights ϕ_k of these bins are related by $\phi_k \approx \int_{L_k-\Delta L/2}^{L_k+\Delta L/2} \phi \, \mathrm{d}N(L) / \int_{L_k-\Delta L/2}^{L_k+\Delta L/2} \mathrm{d}N(L)$ to the underlying luminosity distribution. The inhomogeneous distribution of galaxies across the bins in this method makes it therefore more complicated to infer the smooth underlying luminosity function from the non-parametric estimate.

Second, the nice feature of the absence of correlations among the m_k simplifies the estimation of errors in the luminosity function. Let me define for that purpose the values

$$\Phi_k \equiv \Phi_0(L_k) = \left(\frac{P}{1+x_k} - \frac{m_k}{x_k}\right) \frac{S_k}{(1+x_k)^P L'_{\mathrm{min}}(x_k)} \tag{6.98}$$

of the luminosity function at the points L_k.

I now want to estimate the fractional uncertainties in the Φ_k which later on might be used to fit an analytic model to the luminosity function by minimizing χ^2.

The value of S_k depends via equation (6.44) on the slopes and on the normalization S_0 of the selection function. The latter in principle also depends weakly on each of the m_k but one can safely neglect this dependence in the error estimate. By using error propagation one can then write

$$\sigma^2(\Phi_k) = \sum_l \left(\frac{\partial \Phi_k}{\partial m_l} \sigma(m_l)\right)^2 . \tag{6.99}$$

This gives rise to the fractional error

$$\sigma^2(\log \Phi_k) = \left(\frac{\sigma(\Phi_k)}{\Phi_k}\right)^2 = \sum_l \left(\frac{\delta_{k,l}}{m_k - P\frac{x_k}{1+x_k}} + \sum_{j=1}^{k} \delta_{j,l} \log \frac{x_j}{x_{j-1}}\right)^2 \sigma^2(m_l). \tag{6.100}$$

Although not of principal importance one might still acknowledge that another advantage of the new estimator for the luminosity function is its relative computational ease. In particular one does not have to resort to lengthy iteration techniques as required by the conventional step-wise non-parametric estimator.

Luminosity evolution

The formulation of equation (6.97) was for density evolution only. In the case of luminosity evolution the scale of the luminosity function changes with time and the normalization stays fixed. Hence one also needs to specify the epoch Z at which it is evaluated.

Again, it is necessary to define the maximal redshift differently as before since the sources effectively brighten when they are moved to higher redshift. Let me define $z^\star_{\max}(L)$ implicitly by

$$\frac{L_{\min}(z^\star_{\max})}{f(z^\star_{\max})} = \frac{L}{f(Z)}. \tag{6.101}$$

One can then evaluate equation (6.38) at $z = z^\star_{\max}(L)$ and finds

$$\Phi_Z(L) = -\frac{S'(z^\star_{\max}(L))}{L}\left(\frac{L'_{\min}(z^\star_{\max})}{L_{\min}(z^\star_{\max})} - \frac{f'(z^\star_{\max})}{f(z^\star_{\max})}\right)^{-1}. \tag{6.102}$$

The mapping of the redshifts x_k to the luminosities L_k proceeds via

$$x_k = z^\star_{\max}(L_k), \tag{6.103}$$

which is equivalent to $L_k = \frac{f(Z)}{f(x_k)}L_{\min}(x_k)$. So one arrives at the final formula

$$\begin{aligned}\Phi_Z(L) &= -\frac{S_k}{L}\frac{m_k}{z^\star_{\max}}\left(\frac{z^\star_{\max}}{x_k}\right)^{m_k} \\ &\quad \left(\frac{L'_{\min}(z^\star_{\max})}{L_{\min}(z^\star_{\max})} - \frac{f'(z^\star_{\max})}{f(z^\star_{\max})}\right)^{-1} \quad \text{for } L_{k-1} < L < L_k\end{aligned} \tag{6.104}$$

for the piecewise description of the luminosity function. Note that I left out the argument L of $z^\star_{\max}(L)$ in these formulae to simplify the notation.

Chapter 7

Preparatory Calculations

7.1 The maximal redshift

Before one can really start to calculate the selection function one needs to specify how the equations (6.20) or (6.39) are solved, respectively. In the case of pure density evolution z_i^{m} gives the maximal redshift a given source could be moved to without falling below the flux limit. In the case of pure luminosity evolution the maximal redshift is calculated in a similar way. The only difference is that the source effectively brightens when it is moved to higher redshift. I will frequently use the generic name $z_{\max}$ for the maximal redshift z_i^{m} of both evolutionary models and come back to differences between them when necessary.

As will be seen the results for the selection function and the evolution crucially depend on this maximal redshift. For an exact calculation of $z_{\max}$ a number of effects have to be observed which complicate things. First the background cosmology lets us observe the intrinsic spectrum of the source at different rest frame frequencies when the source is moved around in redshift space. Recall that the frequency one observes at is fixed, though. Since the source has some spectral energy distribution (SED) the observed flux not only changes because of the difference in distance but also because different pieces of the spectrum are observed at different redshifts. The necessary correction for this effect is called K-correction. At a still finer level the observed flux can show an additional change if the shape of the spectrum inside the finite bandpass of the observer also varies when the source is moved to a different redshift. This colour-correction depends both on the SED of the source and on the responsitivity of the detector. If the SED is simply a power law the colour-correction vanishes.

To calculate the maximal redshift one therefore need a model for the spectral energy distribution and the background cosmology has to be specified.

K- and colour corrections

Let me start by deriving an expression for $z_{\max}$ that incorporates in a general way all necessary K- and colour-corrections.

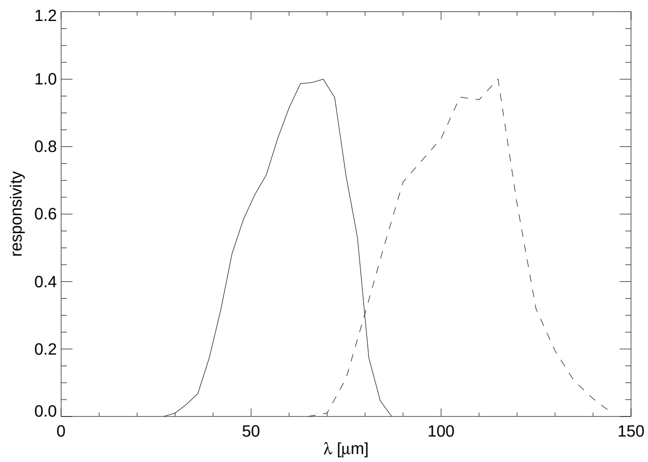


Figure 7.1: Responsivity of the IRAS 60 μm (solid) and 100 μm bands (dashed).

The IRAS satellite has measured the broadband infrared fluxes in four bandpasses centered at the effective wavelengths 12 μm, 25 μm, 60 μm and 100 μm (Beichman et al. 1985). The 60 μm band I am interested in stretches from 30 μm to 80 μm and can be described by a responsivity function $R_{60}(\nu)$ (see figure 7.1) so that the observed flux in the band is given by

$$F = \int R_{60}(\nu) f_\nu(\nu)\, \mathrm{d}\nu. \tag{7.1}$$

In principle one would like to know the true flux density $f_{60} = f_\nu(\nu_0)$ at the center $\nu_0 = c/60\,\mu\mathrm{m} = 5 \times 10^{12}$ Hz of the band. Since one can only measure F and $R_{60}(\nu)$ this information can only be obtained if some spectral shape for $f_\nu(\nu)$ is assumed. The IRAS point source catalog (PSC) takes $f_\nu \propto \nu^{-1}$ and quotes flux density values which are calculated as

$$f_{60}^{\mathrm{PSC}} = \frac{F}{\int \frac{\nu_0}{\nu} R_{60}(\nu)\, \mathrm{d}\nu}. \tag{7.2}$$

Only if the true spectrum follows indeed $f_\nu(\nu) \propto \nu^{-1}$ one has $f_{60} = f_{60}^{\mathrm{PSC}}$. If I combine equations (7.1) and (7.2) the quoted flux density is

$$f_{60}^{\mathrm{PSC}}[f_\nu(\nu)] = \frac{\int R_{60}(\nu) f_\nu(\nu)\, \mathrm{d}\nu}{\int \frac{\nu_0}{\nu'} R_{60}(\nu')\, \mathrm{d}\nu'}. \tag{7.3}$$

From now on I will leave off any additional label 'PSC' so that the flux densities f_{60} and f_{100} always refer to the quoted values in the PSC for the 60 μm and 100 μm bands, respectively.

Let me now consider a particular galaxy in the sample, observed at redshift z with flux density f_{60}. This galaxy has an intrinsic spectral energy distribution $L_\nu(\nu)$

and lies at coordinate r. The spectral distribution $f_\nu(\nu)$ of the observed flux is then given by

$$f_\nu(\nu) = \frac{L_\nu\left(\nu(1+z)\right)}{4\pi a_0^2 r^2 (1+z)}. \tag{7.4}$$

If the source could be moved to the maximal redshift z_{max} the observed flux would be

$$f_\nu^{\mathrm{min}}(\nu) = \frac{L_\nu\left(\nu(1+z_{\mathrm{max}})\right)}{4\pi a_0^2 r_{\mathrm{max}}^2 (1+z_{\mathrm{max}})}, \tag{7.5}$$

where $r_{\mathrm{max}} = r(z_{\mathrm{max}})$ is the coordinate at the maximal redshift. Since the flux limit $f_{\mathrm{min}} = 0.6\,\mathrm{Jy}$ is imposed the redshift z_{max} is just chosen such that the corresponding observed flux density $f_{60}[f_\nu^{\mathrm{min}}(\nu)]$ given by equation (7.3) is equal to f_{min}.

In the subsequent derivation of equation (7.11) I follow closely the treatment of Fisher et al. (1992). From equations (7.3) and (7.5) one finds

$$f_\nu\left(\nu\frac{1+z_{\mathrm{max}}}{1+z}\right) = \frac{r_{\mathrm{max}}^2}{r^2}\frac{1+z_{\mathrm{max}}}{1+z}\, f_\nu^{\mathrm{min}}(\nu). \tag{7.6}$$

With the help of this relation one can express the ratio

$$\frac{f_{\mathrm{min}}}{f_{60}} = \frac{f_{60}[f_\nu^{\mathrm{min}}(\nu)]}{f_{60}[f_\nu(\nu)]} \tag{7.7}$$

as

$$\frac{f_{\mathrm{min}}}{f_{60}} = \frac{r^2(1+z)}{r_{\mathrm{max}}^2(1+z_{\mathrm{max}})}\,\frac{\int R_{60}(\nu) f_\nu\left(\nu\frac{1+z_{\mathrm{max}}}{1+z}\right)\,\mathrm{d}\nu}{\int R_{60}(\nu') f_\nu(\nu')\,\mathrm{d}\nu'}. \tag{7.8}$$

Let me now define

$$\eta = \frac{1+z_{\mathrm{max}}}{1+z} \tag{7.9}$$

and

$$\Psi(\eta) = \frac{\int R_{60}(\nu) f_\nu\left(\nu\eta\right)\,\mathrm{d}\nu}{\int R_{60}(\nu') f_\nu(\nu')\,\mathrm{d}\nu'} \tag{7.10}$$

to simplify the notation. In order to determine z_{max} one then has to solve

$$\frac{f_{\mathrm{min}}}{f_{60}} = \frac{r^2}{r_{\mathrm{max}}^2}\,\frac{\Psi(\eta)}{\eta}. \tag{7.11}$$

Equation (7.11) is the central result of this section and contains K- and colour-corrections and the effects of the background cosmology. Notice that the function $\Psi(\eta)$ encodes the spectral energy distribution of the source as well as the colour-correction and that it is independent of the assumed cosmological background model. The latter affects equation (7.11) only via the function $r(z)$.

In the case of pure density evolution I can directly use equation (7.11) to calculate $z_{\mathrm{max}} = z_i^{\mathrm{m}}$. I will do that for different cosmological backgrounds and for a few different models for the spectral energy distribution.

In the case of luminosity evolution equation (7.11) for the maximal redshift takes the form

$$\frac{f_{\min}}{f_{60}} = \frac{r^2}{r_{\max}^2} \Psi(\eta) \, \frac{f(z_{\max})}{f(z)} \,, \tag{7.12}$$

where $z_{\max} = z_i^{\mathrm{m}}$ gives the maximal redshift defined in equation (6.39). The difference compared to equation (7.11) means that the source is effectively brightened when it is moved to higher redshift.

Let me finally give a few more technical details how I actually solve equation (7.11) or (7.12), respectively. This has to be accomplished for each galaxy specified by a flux f_i and a redshift z_i. Then $\eta = (1 + z_i^{\mathrm{m}})/(1 + z_i)$ and $\Psi(\eta)$ encodes the spectral properties of the source together with the K- and colour-corrections. I either take $\Psi(\eta) = \eta^{\alpha}$ for a power law spectrum or the $\Psi(\eta)$ resulting for the gray body model discussed below.

In the latter case $\Psi(\eta)$ and its derivative have to be calculated for each galaxy individually. Finally $r(z)$ is given by the desired cosmology. I will discuss various choices for it in section 7.4.

To find the solution of equation (7.11) in η I define the auxiliary function

$$g(\eta) = \Psi(\eta) - \eta \frac{f_{\min}}{f_i} \frac{r^2(z_i^{\mathrm{m}}(\eta))}{r^2(z_i)} \tag{7.13}$$

and look for the zero of $g(\eta)$ with a Newton iteration according to

$$\eta_{n+1} = \eta_n - \frac{g(\eta_n)}{g'(\eta_n)}. \tag{7.14}$$

The convergence is fast and easy to achieve. I iterated to a relative accuracy of 10^{-10} in η, which then translates into a solution for z_i^{m}.

In the treatment of luminosity evolution one has to solve equation (7.12) instead. Apart from the minor change that the second term in $g(\eta)$ gets an additional factor of $f(z_{\max})/f(z)$ one can employ the same procedure as above.

7.2 Modeling the spectral energy distribution

In order to determine $z_{\max}$ the function $\Psi(\eta)$ has to be known. This requires a model for the spectral energy distribution (SED) of IRAS galaxies.

Physically the far-infrared radiation emitted by IRAS galaxies, which are in their majority ordinary spirals (Lawrence et al. 1986), is believed to originate from interstellar dust. There are a variety of mechanisms that can cause such dust emission. Dust in interstellar neutral hydrogen clouds can be illuminated by the general interstellar radiation field and subsequently radiate strongly at $100\,\mu$m and $60\,\mu$m. This has been called the *infrared cirrus* (Low et al. 1984) in our own galaxy. To radiate strongly at $25\,\mu$m and $12\,\mu$m small grains or large molecules are needed.

Similarly, the dust in the surface layers of molecular clouds can be heated by the interstellar light or by young OB stars formed from the cloud. Dust in the vicinity of protostars and newly formed stars will also radiate strongly in the far-infrared. Finally, dust shells of high optical depth around hot stars will strongly radiate in the $12\,\mu$m and $25\,\mu$m bands.

Starburst galaxies are expected to differ from ordinary spirals only in the relative proportions of the above ingredients, but galaxies with a quasar like nucleus might produce additional infrared radiation. Here some of the violet and visible light of the nuclear source will be adsorbed and re-emitted in the infrared.

Based on this Rowan-Robinson & Crawford (1989) have proposed that the IRAS spectra in the 12-100$\,\mu$m regime can be understood in terms of three components:

- A cool disc component which originates from interstellar dust illuminated by the interstellar radiation field.
- A warm starburst component that results from newly formed massive stars in optically thick clouds.
- A hot 'Seyfert' component which models the emission from dust with a r^{-1} density distribution around a power-law continuum point source. This component is present only in a small fraction of galaxies.

Clearly a precise description of the infrared emission of an IRAS galaxy requires a detailed individual modeling with several components. But because I am only interested in the K-correction a knowledge of the average behaviour of the spectrum around $60\,\mu$m is sufficient for this purpose.

Empirically the spectra derived with detailed models exhibit a logarithmic slope around $\alpha \approx -2$ at $60\,\mu$m. The simplest choice is therefore the assumption of a straight power law spectrum with this slope for the frequency range probed by the $60\,\mu m$ passband. A more accurate K-correction might be obtained if one tries a more detailed description of the spectra. This is in principle possible if fluxes in other passbands are also available.

Saunders et al. (1990) and most recently Mann, Saunders & Taylor (1996) used a single temperature gray body model based on $60\,\mu$m and $100\,\mu$m fluxes for that purpose. I will apply this model below, too. Fisher et al. (1992) argued that the average K-correction found in this way is too steep and underpredicts in many cases the measured $25\,\mu$m fluxes. They suggested that a polynomial ansatz for the spectrum fitted to the three bands $25\,\mu$m, $60\,\mu$m and $100\,\mu$m might do a better job and give a more realistic description of the spectrum. The resulting average K-correction for their 1.2-Jy redshift survey is somewhat shallower and reduces the observed evolution. But Fisher et al.'s (1992) procedure risks to overestimate the flux between $25\,\mu$m and $60\,\mu$m, especially if the spectrum is indeed comprised of a cool and a warm component (as often suggested) and has a minimum between $25\,\mu$m and $60\,\mu$m.

Since most of the galaxies in PSCz also have $100\,\mu$m fluxes of good or moderate quality but only a small fraction has also good $25\,\mu$m detections I only adopted the scheme used by Saunders et al. (1990) and model the emission in the relevant infrared by single temperature gray body emission.

Gray body model

I assume in this model that the intrinsic spectrum of the source can be described by $L_\nu(\nu) \propto \nu B_\nu(\nu, T_{\rm em})$ where $B_\nu(\nu, T_{\rm em})$ is the Planck spectrum of temperature $T_{\rm em}$. Then the observed flux distribution is given by

$$f_\nu(\nu) \propto \nu B_\nu(\nu, T_{\rm ob}) \tag{7.15}$$

with $T_{\rm ob} = T_{\rm em}/(1+z)$. I now want to fit the temperature $T_{\rm ob}$ such that the IRAS fluxes f_{60} and f_{100} in the $60\,\mu m$ and $100\,\mu m$ passbands are reproduced.

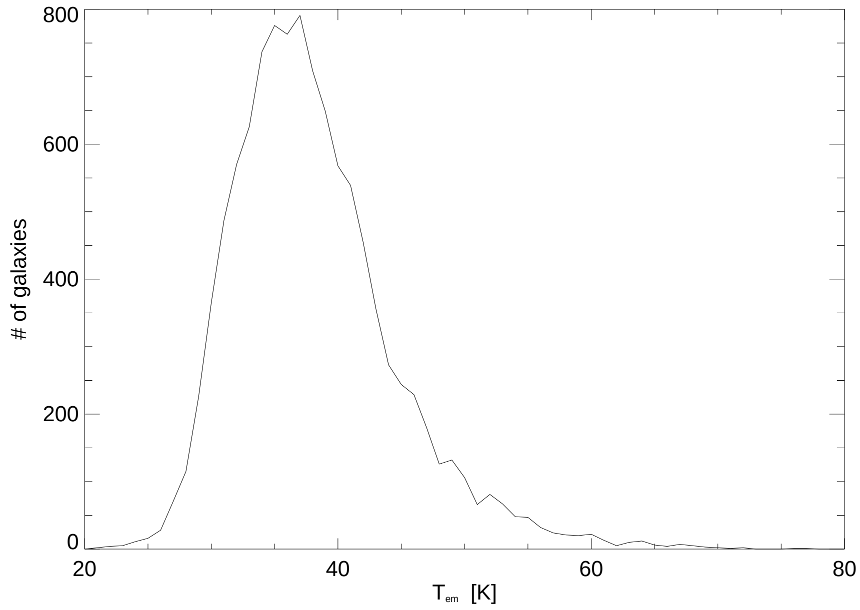


Figure 7.2: Distribution of effective emission temperatures in the single temperature gray body model fitted to the IRAS $60\,\mu$m and $100\,\mu$m fluxes.

By using equation (7.11) the ratio of the measured fluxes can be expressed as

$$\frac{f_{60}}{f_{100}} = \frac{\int R_{60}(\nu)\,\nu B_\nu(\nu, T_{\rm ob})\,{\rm d}\nu}{\int R_{60}(\nu')\,\frac{\nu_{60}}{\nu'}\,{\rm d}\nu'}\;\frac{\int R_{100}(\nu'')\,\frac{\nu_{100}}{\nu''}\,{\rm d}\nu''}{\int R_{100}(\nu''')\,\nu''' B_\nu(\nu''', T_{\rm ob})\,{\rm d}\nu'''}. \tag{7.16}$$

Here ν_{60} and ν_{100} are the frequencies corresponding to $60\,\mu m$ and $100\,\mu m$, respectively. The right hand side of the above equation depends only on the temperature $T_{\rm ob}$. Let me call this function $R(T_{\rm ob})$. I can therefore fit a temperature $T_{\rm ob}$ to the flux ratio f_{60}/f_{100} by numerically inverting the function $R(T)$.

For this purpose I first tabulate

$$R(T) = \left(\frac{\nu_{60}}{\nu_{100}}\right)^4 \frac{\int R_{60}(x)\frac{x^4}{e^{v_1 x}-1}dx}{\int R_{60}(x')x'^{-1}\,{\rm d}x'}\;\frac{\int R_{100}(y')y'^{-1}\,{\rm d}y'}{\int R_{100}(y)\frac{y^4}{e^{v_2 y}-1}\,{\rm d}y}, \tag{7.17}$$

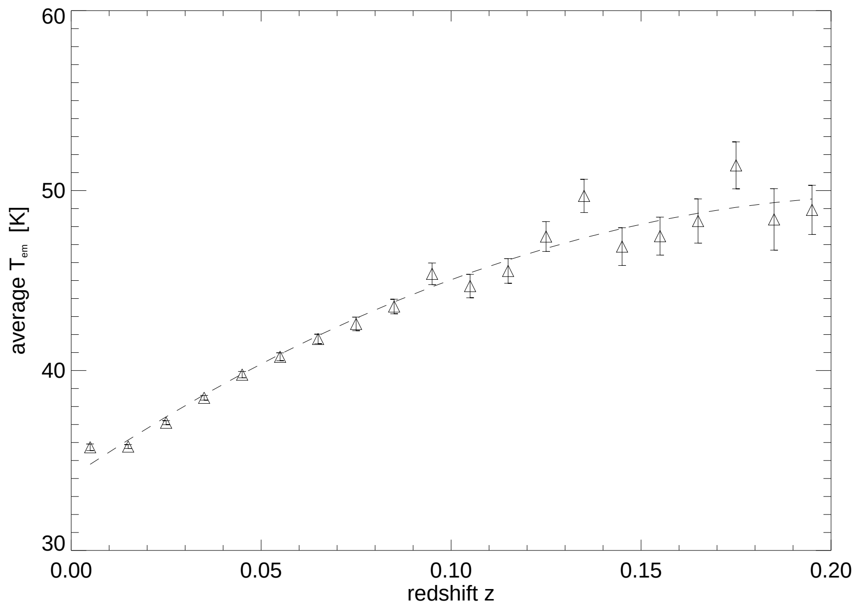


Figure 7.3: The mean emission temperature in PSCz. The dotted curve is an empirical fit used to assign emission temperatures to sources lacking 100μm detections of good quality.

where $x = \nu/\nu_{60}$, $y = \nu/\nu_{100}$, $v_1 = h\nu_{60}/kT$ and $v_2 = h\nu_{100}/kT$ at a large number of different temperatures. By using cubic splines to interpolate between points I can then solve equation (7.16) very accurately.

The distribution of emission temperatures $T_{\rm em} = (1+z)T_{\rm ob}$ derived in this way is shown in figure 7.2. Saunders et al. (1990) showed that the luminosity function of IRAS galaxies is dominated by cool galaxies (defined as galaxies with $T_{\rm em} < 36\,$K) for $L < 5 \times 10^6\, h^{-2} \mathrm{L}_\odot$ whereas warm galaxies ($T_{\rm em} > 36\,$K) dominate for higher luminosity. Mann et al. (1996) pointed out that as a consequence of that there exists a radial gradient in the average emission temperature.

In figure 7.3 I show a plot of the mean emission temperature found as a function of redshift in the PSCz sample. This result is very similar to Mann et al.'s (1996) findings for the QDOT survey.

Among the 14505 galaxies in PSCz are 3842 which have no $100\,\mu m$ fluxes of at least good or moderate flux quality. I have assigned to them the mean emission temperature as a function of redshift given by the empirical fit shown in figure 7.3.

Once $T_{\rm ob} = T_{\rm em}/(1+z)$ is determined for a source one can calculate $\Psi(\eta)$ from its definition (7.10) which can also be used to evaluate the derivative $\Psi'(\eta)$ needed in equation (7.14).

7.3 The minimum luminosity $L_{\min}(z)$

In order to derive the luminosity function of the survey one needs to know the minimum luminosity $L_{\min}(z)$ a source must have at redshift z in order to be seen above the flux limit. The inverse of this function gives the maximal redshift $z_{\max} = z_{\max}(L)$ a galaxy with luminosity L might have.

I define the $60\,\mu$m luminosity L_{60} as

$$L_{60} = \nu L_\nu(\nu) \tag{7.18}$$

at $60\,\mu m$. For simplicity let me assume that the spectral shape in the frequency range around the $60\,\mu$m filter can be well described by a power law with slope α. Then the luminosity density $L_\nu(\nu)$ of the source can be written as

$$L_\nu(\nu) = \frac{L_{60}}{\nu_0}\left(\frac{\nu}{\nu_0}\right)^\alpha, \tag{7.19}$$

which leads to an observed flux distribution of

$$f_\nu(\nu) = \frac{L_{60}}{\nu_0}\left(\frac{\nu}{\nu_0}\right)^\alpha \frac{(1+z)^{\alpha-1}}{4\pi a_0^2 r^2(z)}. \tag{7.20}$$

The minimum luminosity $L_{\min}(z)$ is now the value of L_{60} that gives a quoted flux density $f_{60}[f_\nu(\nu)] = f_{\min} = 0.6\,\mathrm{Jy}$. The quoted flux densities in the PSC are only identical to $f_{\nu_0}(\nu_0)$ for an $\alpha = -1$ spectrum, otherwise one has to correct them by a constant colour-correction factor K according to $f_{60}^{\rm quoted} = K f_{\nu_0}(\nu_0)$. For an $\alpha = -2$ spectrum, K is 0.99 (table VI.C.6 in Beichman et al. 1985). Taking this into account one finds

$$L_{\min}(z) = \frac{\nu_0 f_{\min}}{K} 4\pi a_0^2 r^2(z)(1+z)^{1-\alpha}, \tag{7.21}$$

which gives

$$L_{\min}(z) = \frac{4\pi c^2 \nu_0 f_{\min}}{K H_0^2}\, 4(2+z-2\sqrt{1+z})(1+z)^{-\alpha} \tag{7.22}$$

for an Einstein-de-Sitter universe. The ratio $L'_{\min}(z)/L_{\min}(z)$ evaluates in this case to

$$\frac{L'_{\min}(z)}{L_{\min}(z)} = 2\frac{r'(z)}{r(z)} + \frac{1-\alpha}{1+z}. \tag{7.23}$$

7.4 Distance redshift relations

In the previous chapter I have shown how $\Psi(\eta)$ can be calculated. In order to find $z_{\max}$ and subsequently the selection function I still have to specify the function $r(z)$, that is the distance-redshift relation. In a somewhat sloppy way I call the radial coordinate r of an object its distance. It shouldn't be confused with it's actual *physical* distance. The latter is of little use for this work, though. $r(z)$ depends on the background cosmology and can be calculated from equation (A.19).

In light of the surprising results for the evolution seen in the PSCz I want to examine how the strength of this effect depends on the adopted background cosmology. In the following I will therefore give the distance redshift relations for three distinct cases, the flat Einstein-de-Sitter model, a flat Λ-Model and an open model. Later on I will calculate the evolution for all three of these cosmologies.

Einstein-de-Sitter universe

This model is specified by $\Omega_0 = 1$ and $\Omega_\Lambda = 0$. Hence space is flat with $\Omega_c = k = 0$ and equation (A.19) can be readily integrated to give

$$r(z) = \frac{2}{a_0 H_0}\left(1 - \frac{1}{\sqrt{1+z}}\right). \tag{7.24}$$

Open universe

Here I consider a universe with $\Omega_0 < 1$ and $\Omega_\Lambda = 0$. This therefore leads to an open universe with $\Omega_c = 1 - \Omega_0$, $k = -1$, $a_0 H_0 = \frac{1}{\sqrt{1-\Omega}}$ and $E(z) = (1+z)\sqrt{1+\Omega_0 z}$.

The relevant integral in equation (A.19) evaluates to

$$\int_0^z \frac{dz'}{E(z')} = \frac{1}{\sqrt{1-\Omega_0}} \ln\left(\frac{(\sqrt{1+\Omega_0 z} - \sqrt{1-\Omega_0})(1+\sqrt{1-\Omega_0})}{(\sqrt{1+\Omega_0 z} + \sqrt{1-\Omega_0})(1-\sqrt{1-\Omega_0})}\right), \tag{7.25}$$

so that one finds the expression

$$r(z) = 2\sqrt{1-\Omega_0}\,\frac{\Omega_0 z + (2-\Omega_0)(1-\sqrt{1+\Omega_0 z})}{\Omega_0^2(1+z)} \tag{7.26}$$

for the distance-redshift relation after a few lines of algebra.

For definiteness I will choose the specific model $\Omega_0 = 0.2$ in my later applications. In this case I arrive at the distance-redshift relation

$$r(z) = 4\sqrt{5}\,\frac{z + 9(1-\sqrt{1+z/5})}{1+z}. \tag{7.27}$$

Flat Λ-universe

Different flat background cosmologies are possible if one allows for a cosmological constant Λ. Because of the two conditions $\Omega_0 + \Omega_\Lambda = 1$ and $q_0 = \frac{1}{2}\Omega_0 - \Omega_\Lambda$ (see equations (A.12) and (A.13)) the family of these models can be parameterized by q_0 alone. For that purpose I write

$$\begin{aligned} \Omega_0 &= \frac{2}{3}(1+q_0) \\ \Omega_\Lambda &= \frac{1}{3}(1-2q_0), \end{aligned} \tag{7.28}$$

and express $E(z)$ in terms of q_0. According to equation (A.19) the comoving distance as a function of redshift is then given by

$$r(z) = \frac{1}{a_0 H_0} \int_1^{z+1} \frac{\mathrm{d}y}{\sqrt{\frac{2}{3}(1+q_0)y^3 + \frac{1}{3}(1-2q_0)}}. \tag{7.29}$$

In principal this integral can be expressed in terms of hypergeometric functions:

$$\begin{aligned} r(z) \quad &= \quad \frac{1}{a_0 H_0}\sqrt{\frac{3}{1-2q_0}}\left(-{}_2F_1\left(\frac{1}{3},\frac{1}{2},\frac{4}{3},\frac{2q_0+2}{2q_0-1}\right)\right. \\ &\qquad \left. +(1+z){}_2F_1\left(\frac{1}{3},\frac{1}{2},\frac{4}{3},\frac{2q_0+2}{2q_0-1}(1+z)^3\right)\right). \end{aligned} \tag{7.30}$$

But this result is quite difficult to handle. For this reason I expand it in a series around the points $z=0$ and $q_0=-1$. This leads to

$$\begin{aligned} r(z) \quad &= \quad \frac{1}{a_0 H_0}\left(z - \frac{1}{2}(1+q_0)z^2 + \frac{1}{6}(1+q_0)(1+3q_0)z^3\right. \\ &\qquad \left. +\frac{1}{24}(1+q_0)(1-12q_0-15q_0^2)z^4 + \mathcal{O}(z^5)\right). \end{aligned} \tag{7.31}$$

Since in the application to the PSCz survey the redshift z is always small this expansion is quite sufficient.

Chapter 8

Results for the PSCz Redshift Survey

8.1 Selection function

The most important objective of this chapter is the derivation of a reliable selection function because this quantity is needed in the study of large-scale structure as done in part I of this work.

In order to calculate my best estimate for the selection function I have adopted a $P = 9.6$ density evolution model (see section 8.3 for a discussion of the evolutionary results) and applied the non-parametric estimator of section 6.3.1 to derive the local slope of the selection function. I have then fitted the resulting slopes m_k to the analytic model

$$S(z) = \frac{\phi}{z^{\alpha}\left(1 + (\frac{z}{z^{\star}})^{\gamma}\right)^{\beta/\gamma}}. \tag{8.1}$$

This resulted in a reduced χ^2 of $\chi^2_\nu = 0.86$ which indicates a very reasonable fit. The parameters of this fit are

$$\begin{aligned} \alpha &= 0.991^{+0.068}_{-0.073} \\ \beta &= 3.445^{+0.173}_{-0.158} \\ \gamma &= 1.925^{+0.162}_{-0.153} \\ z^{\star} &= 0.02534^{+0.00130}_{-0.00116}. \end{aligned} \tag{8.2}$$

Note that the relatively large errors on β and γ are mainly a result of the correlation among them.

Figure 8.1 shows the local slopes m_k derived with the non-parametric estimator for the selection function of section 6.3.1. A sudden drop of the slope and a strong increase of the error at redshifts beyond 0.21 can be noted. I take this as evidence that the shape of the selection function is not well constrained at these high redshifts. However, the data seems to be reasonably good out to redshift 0.18. For that reason

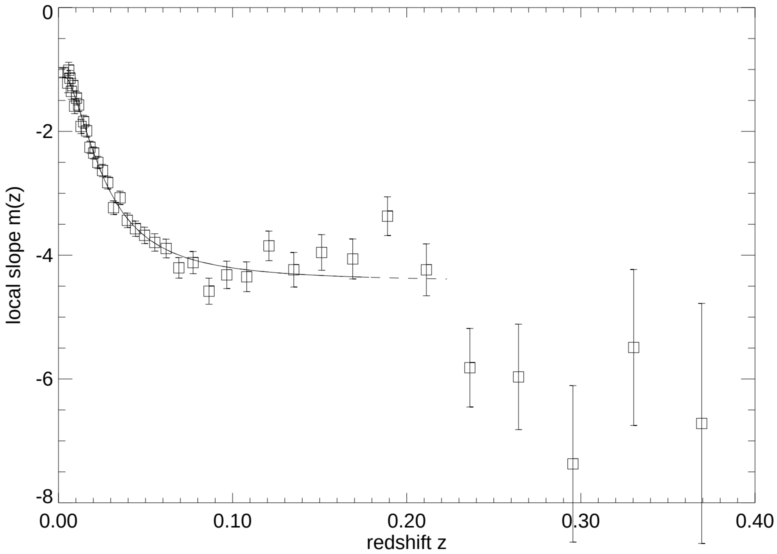


Figure 8.1: Non-parametric estimate of the local slope of the selection function (□ with error bars) for the PSCz survey. The solid line is the slope of an analytic fit to these points. Here an evolutionary model with $P = 9.6$, K-correction $\alpha = -2$ and an Einstein-de-Sitter universe was assumed.

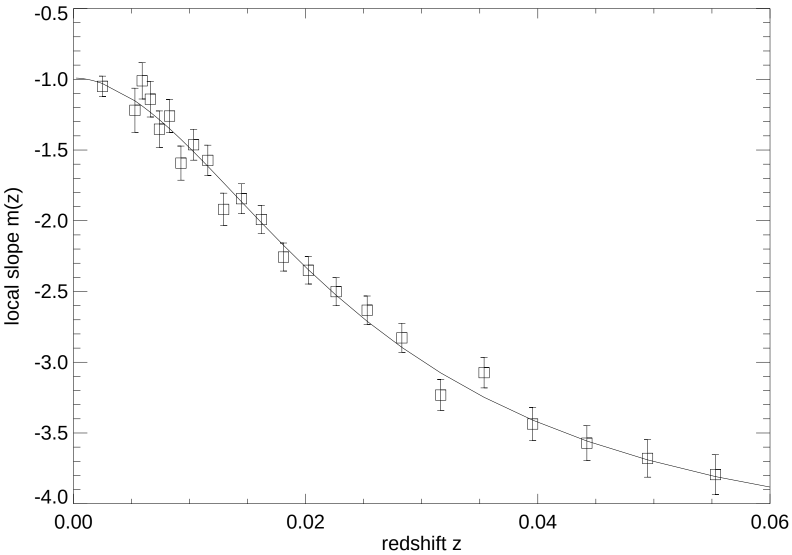


Figure 8.2: The low redshift behaviour of figure 8.1.

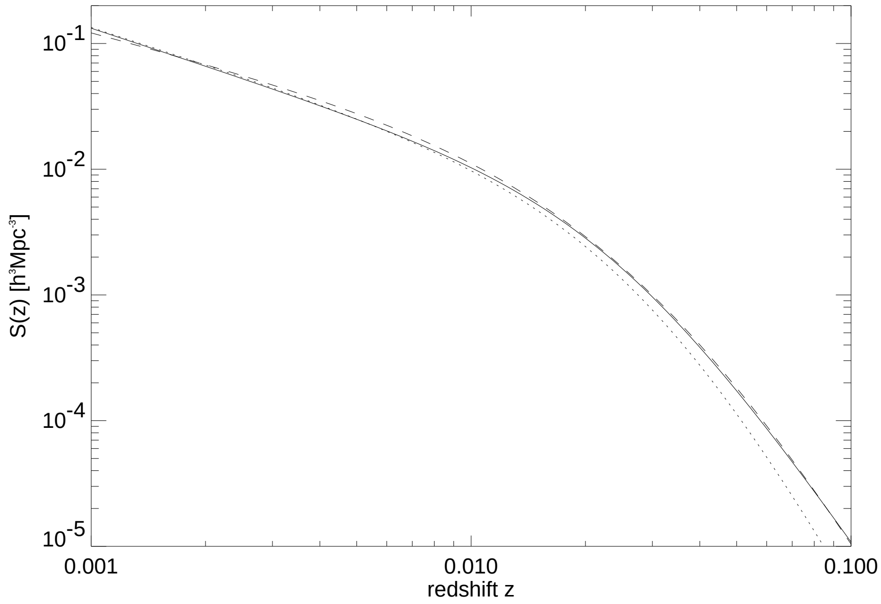


Figure 8.3: Non-parametric estimate of the selection function (solid line) for an evolutionary model with $P = 9.6$, K-correction $\alpha = -2$ and an Einstein-de-Sitter universe. For comparison also shown is Saunders's (1996) selection function (dashed) and the selection function when no evolution is taken into account (dotted).

I have used the redshift interval $[0, 0.18]$ with logarithmically spaced redshifts x_k (40 steps) for the final non-parametric selection function estimate.

The normalization of the selection function can be determined with the method of section 6.3.3. In the application to the PSCz survey I have used the redshift range $[0.00, 0.14]$ containing a total of 14066 galaxies for the normalization. In this way I have obtained

$$\phi = (1.413 \pm 0.024) \times 10^{-4}\, h^3 \mathrm{Mpc}^{-3} \tag{8.3}$$

using the above analytic fit. This translates into a number

$$N = (1368 \pm 23)\ \mathrm{sr}^{-1} \tag{8.4}$$

of galaxies per solid angle on the sky.

Note that this number is not equivalent to the number of sources actually counted on the sky in the *particular realization of the Universe* that we see around us. For example, a slight local overdensity might lead to a higher count as compared to the expected number N. The completeness of the current version of the PSCz catalogue is estimated to be $\approx 98\%$. In this case the normalization of equation (8.4)

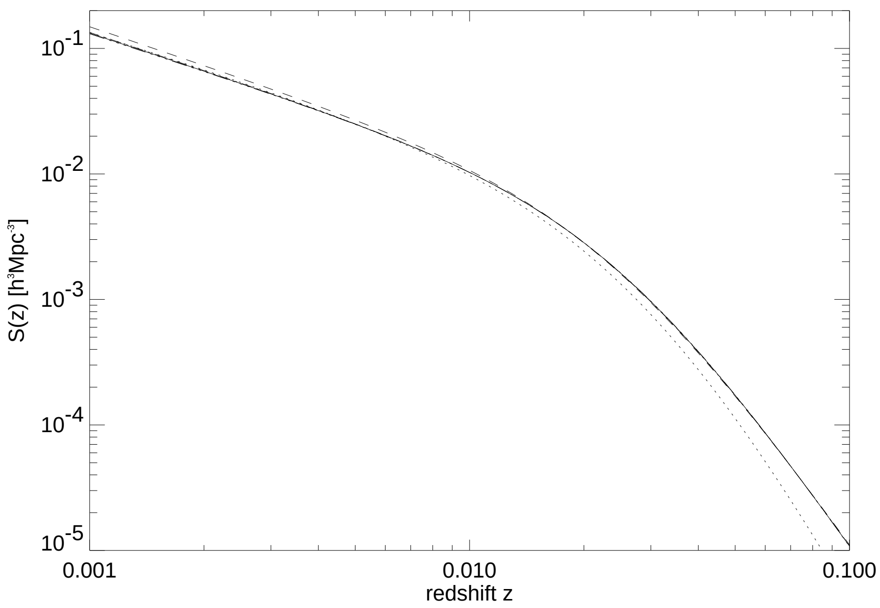


Figure 8.4: Selection Functions. The upper curve actually consists of 4 curves falling practically on top of each other: The $P = 9.6$, $\alpha = -2$ result (solid), the $Q = 4.3$, $\alpha = -2$ result (dashed), the $P = 8.6$, $\alpha = -1.5$ result (dotted), and the $P = 10.6$, $\alpha = -2.5$ result (long dashed). The lower curve (also dotted) gives the selection function when no evolution is assumed.

is systematically underestimated by about 2% with respect to a perfectly complete survey.

In figure 8.3 I plot my best estimate of the selection function of PSCz. In general there is good agreement with the result of Saunders (1996) in most of the survey volume, only at redshifts around $z = 0.005$ my selection function is slightly lower than his. It is also seen that it is crucial to include evolution in the analysis. If it is not taken into account, one arrives at a selection function that is considerably lower at high redshift. At $z = 0.15$ the true mean galaxy density would already be missed by a dramatic factor of ≈ 3. Clearly this difference cannot be neglected.

Despite the uncertainties in the K-correction and the resulting amount of evolution it can be hoped that this uncertainty will affect the final selection function only weakly. In figure 8.4 I show the selection functions resulting for different models of K-correction and evolution. In each case the evolution was estimated with the method outlined in section 6.5 and then folded into the determination of the selection function with the non-parametric maximum likelihood estimator.

It can be seen that the selection functions for the different models are all practically identical. Whereas the quantitative amount of evolution depends on the

K-correction this is fortunately not the case for the selection function itself. If one is only interested in $S(z)$ any reasonable K-correction and evolutionary model will give a good answer. But this is only true if evolution is taken into account in the first place.

8.2 The run of radial density

In this section I will discuss results derived with the non-parametric estimator for the radial density distribution. The theoretical description of this estimator is given in section 6.4.

I applied the method assuming an Einstein-de-Sitter background cosmology and a straight $\alpha = -2$ K-correction. The radial bins were chosen in such a way that the expected number of galaxies (as predicted by the selection function) in each of them was equal. This should give a similar signal-to-noise ratio for each of the bins. In total I used 75 bins out to a radial distance of $464\,h^{-1}$Mpc. Figure 8.5 depicts the result for the full PSCz sample.

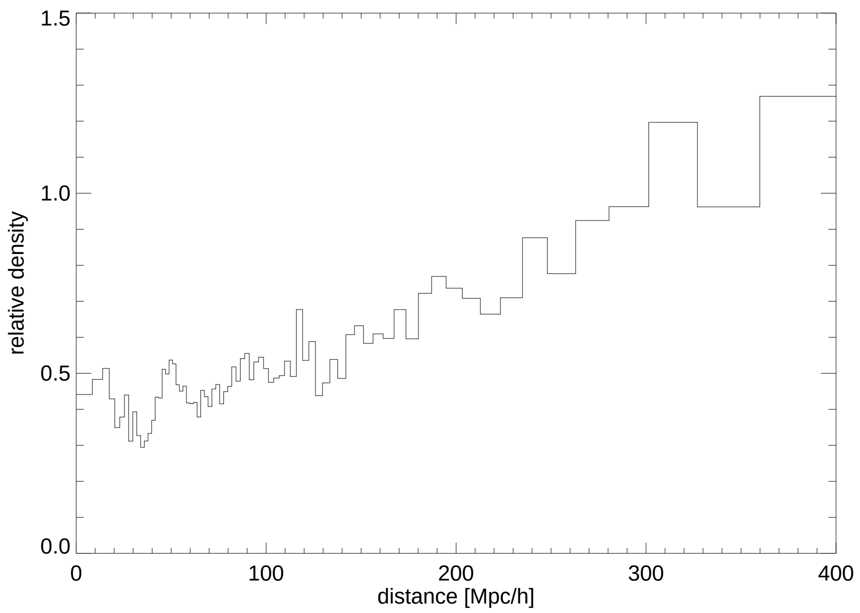


Figure 8.5: The histogram shows the run of the radial density for the PSCz survey derived with the non-parametric maximum likelihood estimator of section 6.4. An Einstein-de-Sitter universe, a straight K-correction with $\alpha = -2$ and no evolution were assumed.

Significant structure can be clearly seen in this plot. At $13\,h^{-1}$Mpc the Virgo cluster shows up as a sharp spike and around $50\,h^{-1}$Mpc the Perseus-Pisces complex (located in the southern hemisphere, compare figure 8.6) can be seen. The overdensity around $100\,h^{-1}$Mpc marks the Hercules region and the underdensity at $140\,h^{-1}$Mpc is a trace of the Boötes void. But perhaps the most striking observation is the apparent rise of the density with increasing distance. Obviously there are substantially more galaxies found at higher redshift than one would expect for a non-evolving luminosity function.

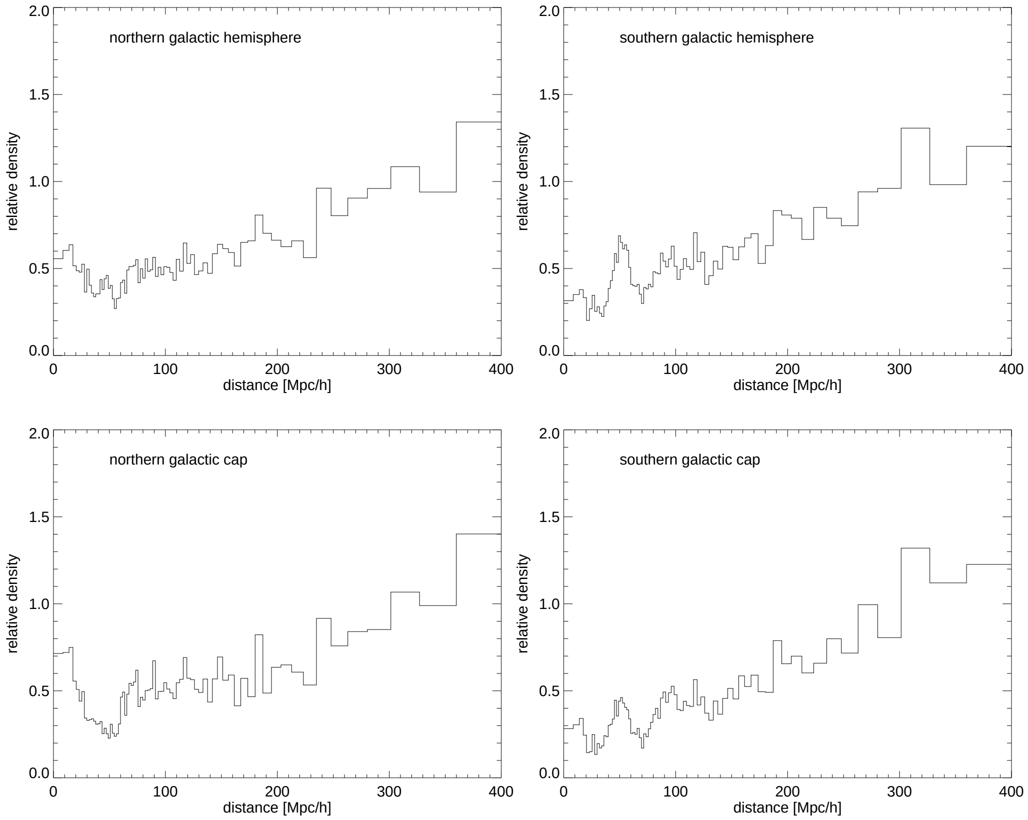


Figure 8.6: The radial density for northern ($b > 0^0$) and southern hemisphere ($b < 0^0$) and for two regions centered around the galactic poles ($b > 25^0$ and $b < -25^0$, respectively). Besides this geometric selection the calculation was as in figure 8.5.

There are a number of possible explanations for this observation. One can invoke either density or luminosity evolution of the luminosity function of IRAS galaxies to account for it. But before this is believed one better examines carefully possible systematic effects that might fake evolution. For example, it turns out that the exact amount of evolution depends rather strongly on the adopted K-correction and slightly on the background cosmology. There might also be systematic errors in the IRAS fluxes that are responsible for a part of the detected evolution.

Of course, another valid explanation is that what is seen in figure 8.5 is indeed

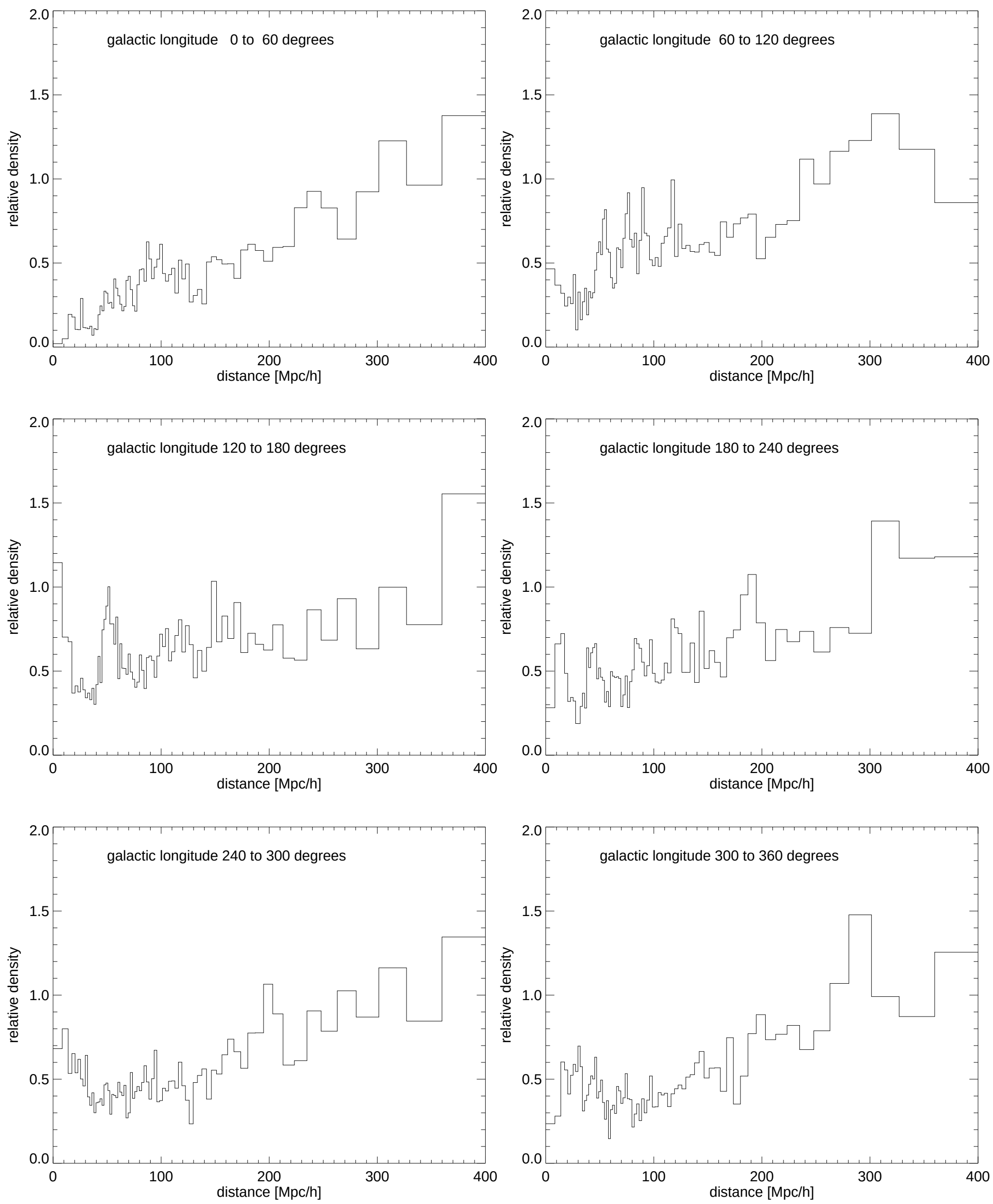


Figure 8.7: The radial density for six slices of 60 degrees galactic longitude. Apart from this geometric selection the calculation was as in figure 8.5.

true large-scale structure. This meant that we lived in a huge underdense hole. Right away, this hypothesis is very unsatisfactory. It becomes even more so if one investigates how the radial density behaves in different directions.

For that purpose I divided the PSCz catalogue in various sub samples. I constructed the radial density estimate for the northern hemisphere ($b > 0^0$), the southern hemisphere ($b < 0^0$), the north galactic cap ($b > 25^0$), the south galactic cap ($b < -25^0$) and finally for six 60^0-slices selected by galactic longitude. The results of these calculations are shown in figures 8.6 and 8.7, respectively.

I conclude from these plots that the characteristic rise of density with distance is present in every direction. Therefore the 'hole' hypothesis required that we lived right at the center of a gigantic hole. This would either violate the cosmological principle or calls for the existence of gigantic voids. Both seems highly unlikely. But perhaps the most convincing argument against it is the fact that local optical redshift surveys haven't detected such a local underdensity out to distance $\approx 200\,h^{-1}$Mpc. In the rest of this analysis I will therefore adopt the working hypothesis that one indeed observes evolution in the IRAS sample.

A consistency check

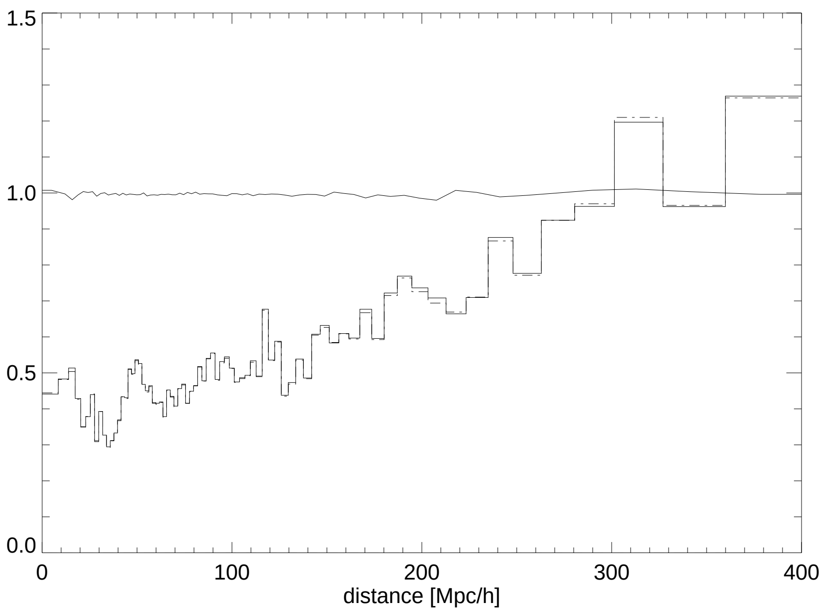


Figure 8.8: A consistency check for the maximum likelihood estimators for the radial density and the selection function. The solid histogram is the radial density derived with the non-parametric estimator for the radial density, the dashed histogram is the density derived with the non-parametric selection function estimate. The horizontal line is the ratio of the two.

The radial density estimates presented in the previous section have been derived with a maximum likelihood estimator which is independent of the luminosity distribution. The selection function estimator on the other hand is independent of density inhomogeneities. Recalling that one can also use the selection function to determine the relative densities in arbitrary bins (one just has to weight the galaxies with $1/S(z_i)$ for that purpose) one can perform an interesting test, if one constructs the radial density distribution again by using the non-parametric selection function estimate.

Figure 8.8 compares the radial density profiles that I obtain with these two methods. In principle they should give the same answer if the underlying assumption of an universal luminosity distribution is well satisfied. As demonstrated in figure 8.8 this test is excellently passed. One can therefore have confidence in the internal consistency of the methods. This result also supports the assumption that a universal luminosity function is sufficient to reconstruct the density field on the (large) scales I consider in this work.

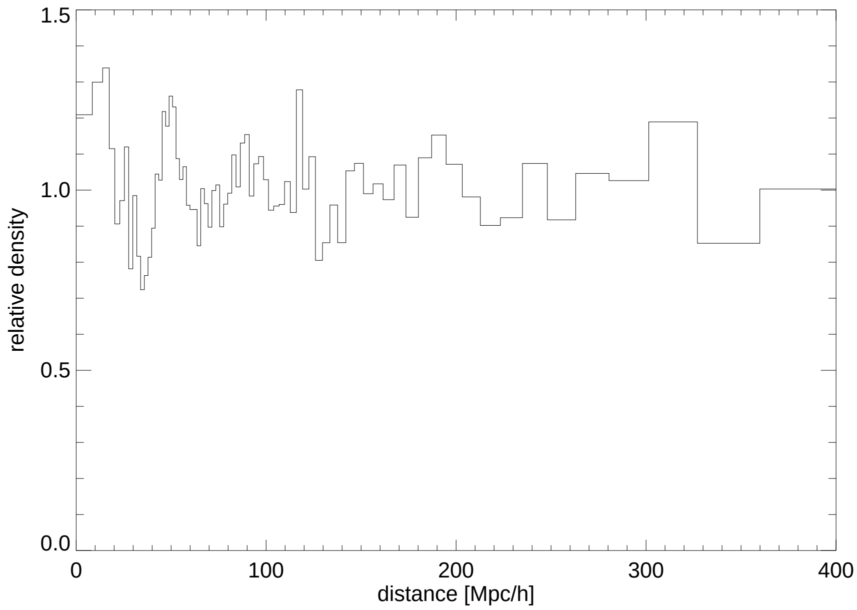


Figure 8.9: The histogram shows the radial density distribution derived with my best estimate of the selection function when evolution was included. This curve resulted from the $\alpha = -2$ model with evolutionary estimate of $P = 9.6$, but it will be nearly the same for other K-corrections or for luminosity evolution with $Q = 4.3$.

It is now also interesting to see what radial density profile one gets if the selection function is determined with an evolutionary correction. My best estimate for density evolution is $P = 9.6$ and that leads to a final selection function that one can use to recompute the radial density distribution. The result is displayed in figure 8.9.

As desired, any systematic trend in the radial density profile has vanished. Instead of living in a giant hole it rather seems that we are in the vicinity of a slight local overdensity.

8.3 Evolution

In this section I discuss quantitative results derived for the evolution seen in the PSCz redshift survey. I will first review previous findings concerning the evolution of IRAS galaxies and then show results derived with the estimator for density evolution of section 6.5 for various redshift intervals. This will demonstrate the uncertainty in the evolutionary result introduced by density inhomogeneities.

I will then examine the dependence of the results on the adopted model for the spectral energy distribution of IRAS galaxies and on the choice of background cosmology. Further I will investigate Malmquist effects as a possible systematic error that might fake evolution. The alternative evolutionary estimator proposed in section 6.6 is also examined and applied to infer the equivalent rate of luminosity evolution. Finally, I discuss the evolution seen in the 1.2-Jy redshift survey and compare it to the PSCz and a 1.2 Jy limited subsample of it.

Estimates for density evolution

Evidence for evolution in IRAS galaxies has been reported by a number of authors using differential source counts (Hacking & Soifer 1991, Hacking & Houck 1987, Hacking, Condon & Houck 1987, Lonsdale, Hacking & Conrow 1990) and redshift surveys (Lonsdale & Hacking 1989, Saunders et al. 1990, Fisher et al. 1992, Oliver et al. 1994).

In particular, the estimator outlined in section 6.5 was introduced by Saunders et al. (1990) in order to quantify this evolution and has subsequently been used by Fisher et al. (1992) and Oliver et al. (1994). However, there has been some controversy about the amount of evolution seen in IRAS galaxies. Saunders et al. (1990) found $P = 7 \pm 2$ for the QCD survey, whereas Fisher et al. (1992) came up with $P = 2 \pm 3$ for the 1.2-Jy survey and claimed that there is no conclusive evidence for evolution in IRAS galaxies. More recently Oliver et al. (1994) found $P = 6.25 \pm 1.5$ for a survey of faint IRAS galaxies and Saunders (1996) gives $P = 9.46 \pm 1.49$ for the PSCz survey.

Fisher et al.'s (1992) result seems to be the only one that does not report a positive detection of evolution in IRAS galaxies. The data of the 1.2-Jy survey has been published (Fisher et al. 1995) recently, so I was able to repeat part of their evolutionary calculations. A discussion of my findings with respect to this catalogue can be found below. In summary I will conclude there that Fisher et al.'s (1992) low result represents most likely an underestimate due to systematic effects. In the framework of my methods I arrive at a better agreement between PSCz and 1.2-Jy. However, the PSCz catalogue gives consistently slightly higher results than the 1.2-Jy survey.

In the discussion of the evolutionary results for the 1.2-Jy survey it will also become apparent that a possible incompleteness at high redshift can severely bias the evolution low. For that reason I introduced an upper redshift cut-off in the estimator for density evolution.

However, it is then also clear that the estimate of the evolutionary parameter will depend rather strongly on the exact window [$z_{\rm min}$, $z_{\rm max}$] in redshift space that is looked at. If $z_{\rm min}$ lies right on a supercluster this will bias the evolution low, if one starts out in a void, this will bias it high. Similar things apply for the upper boundary $z_{\rm max}$, but here the situation is less serious since the fluctuations at $z_{\rm max}$ are generally smaller.

	$\mathbf{z_{min} = 0.00}$	$\mathbf{z_{min} = 0.01}$	$\mathbf{z_{min} = 0.02}$	$\mathbf{z_{min} = 0.03}$
$\mathbf{z_{max} = 0.06}$	9.14 ± 1.33 (11785)	10.66 ± 1.43 (9947)	11.26 ± 1.82 (6788)	9.19 ± 2.67 (4181)
$\mathbf{z_{max} = 0.08}$	9.08 ± 1.09 (12892)	9.99 ± 1.15 (11054)	10.17 ± 1.35 (7895)	9.21 ± 1.72 (5288)
$\mathbf{z_{max} = 0.10}$	9.37 ± 0.97 (13479)	10.10 ± 1.01 (11641)	10.33 ± 1.15 (8482)	9.83 ± 1.37 (5875)
$\mathbf{z_{max} = 0.12}$	9.96 ± 0.87 (13858)	10.60 ± 0.92 (12020)	10.94 ± 1.02 (8861)	10.81 ± 1.18 (6254)
$\mathbf{z_{max} = 0.14}$	9.54 ± 0.84 (14066)	10.08 ± 0.86 (12228)	10.29 ± 0.95 (9069)	10.01 ± 1.07 (6462)
$\mathbf{z_{max}} = \infty$	7.96 ± 0.69 (14505)	8.25 ± 0.71 (12667)	8.18 ± 0.75 (9508)	7.66 ± 0.81 (6901)

Table 8.1: Results for the density evolutionary parameter P derived for the PSCz survey with the estimator of section 8.3 for various redshift intervals. Here a straight K-correction with $\alpha = -2$ and an Einstein-de-Sitter cosmology has been assumed. The errors are the statistical uncertainties as estimated from the drop of the likelihood function. The values in parenthesis give the number of galaxies that lied in the redshift interval and hence contributed to the fit. If one simply takes the average over all these values one gets $P = 9.69 \pm 0.94$ where the error gives the scatter among the measurements.

To demonstrate this scatter I tested several windows in redshift space. Table 8.1 lists results for all combinations resulting from $z_{\rm min} \in \{0.00, 0.01, 0.02, 0.03\}$ and $z_{\rm max} \in \{0.06, 0.08, 0.10, 0.12, 0.14, \infty\}$, where $z_{\rm max} = \infty$ means that no upper redshift cut-off has been employed. The scatter found in the results for P is clearly larger than one expects form the statistical uncertainty. This can be attributed to cosmic variance. Typically the values are somewhat smaller for $z_{\rm min} = 0$, corresponding to the fact that one here starts inside the Local Supercluster which biases the evolution low. For $z_{\rm min} = 0.02$ one begins in the underdense region behind the Perseus-Pisces region which gives larger results. $z_{\rm min} = 0.01$ lies somewhat in between and might be the best choice for this method. Because there is only one representation of the

Universe to look at one more or less has to live with this amount of cosmic error which might only be reduced by deeper observations.

Although the method is troubled by the fact that it is influenced by the large-scale structure present in the Universe the cosmic variance seems to be reasonably low and not to add much more than about $\sigma \approx 1$ to the statistical uncertainty in the determination of P. I will take as my best estimate for P the average of the values that have an upper redshift cut-off applied and add $\sigma = 1$ in quadrature to the statistical error in order to account for the cosmic variance. Subsequently additional systematic uncertainties due to spectral energy distribution, background cosmology and flux errors are examined.

Dependence on K-correction

The amount of measured evolution will depend systematically on the model one takes for the spectral energy distribution of IRAS galaxies. A shallower spectrum will allow a given source to be seen further out and leads therefore to an increase of the number of galaxies one expects to see at larger distance. As a consequence one will have to invoke less evolution to explain the excess that is in fact observed at large redshift.

I have examined various models for the K-correction. I considered in detail the gray body model advocated by Saunders et al. (1990). Here I find that the spectrum is locally rather steep with an average slope of $\alpha = -2.14$. It seems likely that this is too steep, especially for the very luminous galaxies. As pointed out by Fisher et al. (1992) this steep spectrum often underpredicts the $25\,\mu$m fluxes if they are present in the PSC. For that reason they argued that the spectrum is likely to be shallower and that a polynomial fit to the spectrum based on fluxes in three IRAS bands might result in a more accurate description. They derive an average $\Psi(\eta) = 1 - 1.63(1 - 1/\eta) + 7.88(1 - 1/\eta)^2$ for their polynomial model which indeed corresponds to a considerably shallower spectrum that reduced their evolutionary estimates by up to one unit in P.

Besides the gray body model I have examined straight K-corrections for $\alpha = -1.5$, $\alpha = -2.0$ and $\alpha = -2.5$ in order to illuminate the influence of the slope on the evolutionary estimate. The results are listed in table 8.2.

model	**P**
gray body	10.31 ± 0.86
$\alpha = -1.5$	9.04 ± 0.86
$\alpha = -2$	10.07 ± 0.86
$\alpha = -2.5$	11.10 ± 0.87

Table 8.2: The evolutionary parameter P derived with the evolutionary estimator of section 6.5 for different K-correction schemes. Here the redshift interval [0.01,0.14] and an Einstein-de-Sitter cosmology was used for the calculation.

It can be seen that the amount of measured evolution depends on the adopted K-correction scheme and increases when a steeper spectrum is assumed. The physically well motivated gray body model is quite close to $\alpha = -2$ which indicates that it is sufficient to model the average SED by a power law with $\alpha = -2$. If a shallower spectrum in the spirit of Fisher et al.'s (1992) polynomial model is assumed the evolutionary estimate can be reduced by about one unit in P. However, there is hardly enough uncertainty in the SED to bring down the measurements so much that they can be reconciled with a situation without evolution.

Influence of cosmology

In order to test the dependence of the derived evolutionary rate on the adopted background cosmology I calculated the evolution for four different models, an Einstein-de-Sitter universe, two flat Λ-models and an open universe with $\Omega = 0.2$ and $\Lambda = 0$.

cosmology	**P**
flat Einstein-de-Sitter	10.07 ± 0.86
flat Λ-model, $q_0 = 0$	11.01 ± 0.86
flat Λ-model, $q_0 = -0.7$	12.44 ± 0.87
open model, $\Omega_0 = 0.2$	10.95 ± 0.86

Table 8.3: The evolutionary parameter P derived with the estimator of section 6.5 for different background cosmologies. Here the redshift interval $[0.01, 0.14]$ and a K-correction with $\alpha = -2$ have been used.

The results in table 8.3 demonstrate that the resulting differences are rather small, unless the cosmological constant is unreasonably high. An open model or a flat model with cosmological constant give a slightly higher evolutionary rate than the Einstein-de-Sitter universe.

But since the PSCz sample is cosmologically still quite local the measured evolutionary rate is not very sensitive to the background cosmology. For the purposes of this work this is good news since it implies that the analysis of this study will be hardly affected by the convenient choice of an Einstein-de-Sitter cosmology which I will therefore usually adopt.

Malmquist bias and flux errors

The so called 'Malmquist bias' is a potentially serious source for a misleading evolutionary signal. It can arise from random flux errors in the quoted IRAS fluxes. They can in principle be divided into two classes.

First, there can be a constant fractional error in the flux, and some error of this kind certainly *is there*. In this case the absolute error is proportional to the flux itself.

Potentially more serious is a second type of error consisting of a constant baseline flux error that doesn't scale with its absolute value. The first type of error does not lead to substantial Malmquist bias while the second one can in fact fake evolution. This can be understood as follows.

At every distance there is a flux distribution of the objects one sees at that location and this distribution is cut off at the imposed flux limit of 0.6 Jy. Of course, this flux distribution reflects a corresponding piece of the luminosity function, and at different positions the part near the flux limit will correspond to different pieces of the luminosity function. Since the distribution of fluxes near the flux limit has a negative slope a baseline flux error scatters more sources across the flux limit from below than from above. This effect becomes stronger if one looks at a steeper part of the luminosity function at higher redshift. This effectively enhances the density one sees far away compared to the density one sees nearby. In this way one can detect a false evolutionary signal.

catalogue	**P**
unperturbed	0.05 ± 0.26
fractional error 6%	-0.16 ± 0.34
fractional error 10%	-0.09 ± 0.31
baseline error 36 mJy	0.04 ± 0.29
baseline error 50 mJy	0.28 ± 0.31
baseline error 100 mJy	2.64 ± 0.31
baseline error 150 mJy	5.96 ± 0.31

Table 8.4: The evolutionary parameter P derived with the estimator of section 6.5 for seven different mock catalogues of a homogeneous density field. Each of the catalogues contained 100000 galaxies to a depth of redshift 0.169. The surveys have been perturbed with artificial flux errors in order to demonstrate Malmquist bias.

A constant fractional error in the flux on the other side means that all absolute luminosities are measured with the same fractional uncertainty independent of the distance of the source. Consequently, the measured luminosity function will be a convolution of the true luminosity function with some error distribution. This convolution is independent of distance and does not lead to an evolutionary signal for that reason.

In order to quantify the possible effect of Malmquist bias, I have constructed several mock catalogues of 100000 galaxies each. The catalogues consisted of randomly distributed points with an evolution-free luminosity function derived from an analytic fit to the selection function of PSCz. See section 5.1 for a detailed description of the method used to construct the mock catalogues.

In the first catalogue the fluxes remained unperturbed. Two others were constructed with a constant fractional error in the $60\,\mu$m fluxes of 6% and 10%, respectively. Additionally, four catalogues were perturbed with constant baseline flux errors of 36 mJy, 50 mJy, 100 mJy or 150 mJy. For the distribution of these errors I adopted a Gaussian with fixed width for the baseline catalogues and with a dispersion proportional to flux for the fractional catalogues. Unfortunately the distribution of the baseline flux error is not really known, so the Gaussian is only an assumption. The dispersion of 36 mJy is chosen according to Oliver et al.'s (1994) conservative upper bound on the baseline flux error in the 60 μm band.

In table 8.4 I summarize results for the density evolution derived for these mock catalogues. One can see that the baseline flux error can indeed lead to a detection of evolution that is not present in reality. However, at the level of Oliver et al.'s (1994) upper bound of 36 mJy the effect is not very strong and amounts to a systematic increase of the evolutionary estimate by not more than $P \approx 0.5$.

If the baseline flux error is indeed not larger than 36 mJy it is therefore doubtful that a Malmquist bias can explain a significant fraction of the evolution detected in the PSCz survey. Still, P might be biased high by up to $0.5 - 1.0$ due to Malmquist effects, and the systematic uncertainty in the measured values for P or Q is increased accordingly. For my final error estimate I will therefore add ± 1 to the systematic error to account for this uncertainty.

However, Malmquist bias is not the only possible flux error that can lead to an artificial evolutionary signal. For example, the flux scale of the PSC could be slightly nonlinear. More seriously, the PSC (which has a flux limit of 0.5 Jy) might be troubled by a variable depth over the sky due to systematic changes of the effective flux limit. This has been suggested by Hamilton (1995) who finds some subtle but worrisome systematic differences between the QDOT and the 1.2-Jy surveys. The difficult work of trying to disentangle such systematic influences from the true evolutionary rate is still in progress.

Luminosity evolution and alternative estimator

Up to now I have only discussed a model with pure density evolution. Alternatively one can also attribute the evolution entirely to pure luminosity evolution. Because the PSCz sample is not deep enough to differentiate between these two forms any mixed model could also be used. Nevertheless it is interesting to look at luminosity evolution separately. This is also motivated by the fact that it might help in understanding the high rate of evolution in terms of physical processes.

So far all evolutionary results have been computed with the estimator of section 6.5. Here I will now use the alternative method I proposed in section 6.6. It is based on an evaluation of the variance σ^2 of the smoothed mass density field as a function of the evolutionary rate.

For this purpose I have chosen $\lambda = 50\,h^{-1}$Mpc and a survey volume of radius $R = 315\,h^{-1}$Mpc (compare table 2.1). The variance of the smoothed density field was then computed with the unbiased estimator of section 3.3 for selection functions derived for a range of evolutionary rates.

The consistency of this method with the results obtained so far is evidenced by

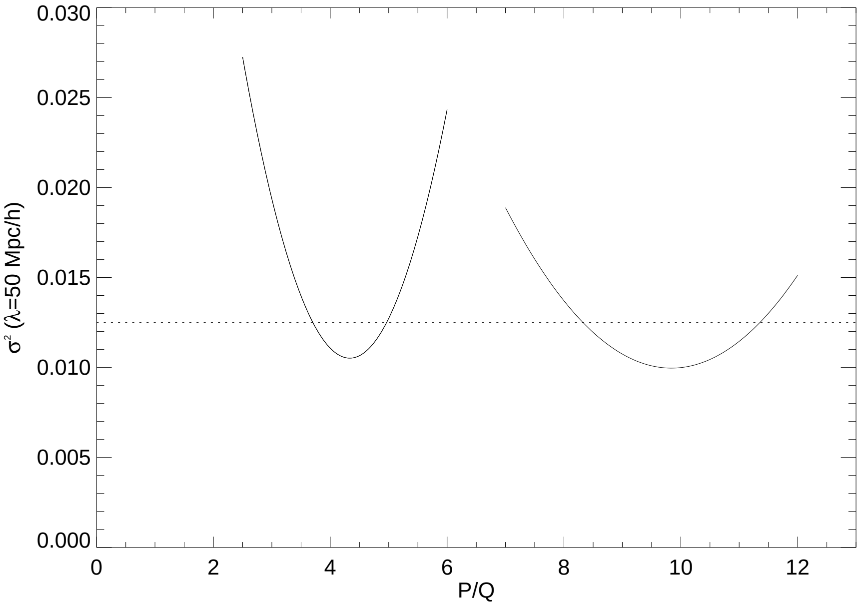


Figure 8.10: The variance σ^2 measured for $\lambda = 50\,h^{-1}$Mpc as a function of the evolutionary rate. The right curve is obtained for density evolution described by the parameter P, whereas the left curve is for luminosity evolution specified by the parameter Q. The dotted line lies one standard deviation (error in σ^2) above the minimum of σ^2.

figure 8.10 where I show the function $\sigma^2(P)$ obtained for density evolution in this way. The minimum is given as $\sigma^2(P_{\min}) = 0.01$ which coincides with the value listed in table 3.3. There the uncertainty of the σ^2-measurement was estimated as $\Delta(\sigma^2) = 0.0023$. This can be translated in an estimate for the confidence interval of P if it is taken as the range for which $\sigma^2(P) < \sigma^2(P_{\min}) + \Delta(\sigma^2)$. In this way I obtain

$$P = 9.84 \pm 1.44 \tag{8.5}$$

as the best estimate for density evolution.

The same method can now also be applied to luminosity evolution. Only the step involving the calculation of the selection function has to be changed. Figure 8.10 also displays the resulting function $\sigma^2(Q)$. Again, one finds a well defined minimum that corresponds to a best evolutionary estimate of

$$Q = 4.33 \pm 0.68. \tag{8.6}$$

Further tests with this alternative estimator for evolution show that it gives very similar results as the conventional estimator when parameters like the survey volume, the SED or the background cosmology are varied. Incidentally, the numerical value for the luminosity evolutionary parameter Q typically comes out as about one half of P and the same is roughly true for the error estimates. Actually this is a consequence of the slope of the luminosity function at the bright end.

Comparison to the 1.2-Jy survey

As mentioned above the result of $P = 2 \pm 3$ found by Fisher et al. (1992) for the 1.2-Jy survey is somewhat puzzling because this value is very low and furthermore still consistent with a non-evolving far-infrared luminosity function.

In order to find the reason for this discrepancy it is certainly interesting to study the 1.2-Jy data (published in Fisher et al. 1995) with the techniques I have applied in this work.

The main difference between my method and that of Fisher et al. (1992) is that these authors have not introduced an upper redshift cut-off in their evolutionary estimator. As a consequence their result is very sensitive to the most distant galaxies in the survey. This can be shown as follows. If I apply their method (with $R_{\rm min} = 6000\,{\rm km/sec}$) to the 1.2-Jy data I get $P = 3.6$. If I then leave out the galaxy with the highest redshift, the result is only 3.1, if I omit the five most distant galaxies, it is only 2.7 and if the 10 most distant galaxies are missing it is only 1.9. In light of this it is also clear that the ≈ 50 galaxies that had been missing in Fisher et al.'s (1992) version of the 1.2-Jy survey compared to the final release (Fisher et al. 1995) can make a difference.

The above tests also demonstrate that the upper redshift cut-off is essential to reduce the shot noise in the *small* number of very distant galaxies. It is also indicated for another reason. Possible incompleteness is most likely to occur for high redshift near the flux limit, and a potential Malmquist bias is also expected to be strongest in this regime. An upper redshift cut-off can reduce both systematic effects.

I now further investigate the dependence on the redshift window. For that purpose I compare the 1.2-Jy survey with a subsample of PSCz that I construct by flux limiting it at 1.2 Jy. In this way a second '1.2-Jy redshift survey' is generated which should have identical statistical properties as the original one (apart from small deviations due to different angular masks).

Table 8.5 shows a comparison of results which I obtained for the density evolution of the PSCz survey and the two 1.2 Jy limited samples. A number of interesting observations can be made. Up to $z_{\rm max} \leq 0.07$ a reasonable agreement within the error bars between all three samples is found, although the PSCz sample gives slightly higher results than the 1.2-Jy sample. Beyond this depth the 1.2-Jy survey exhibits a strong drop of the evolutionary estimate P. A trend of similar strength is not seen in the 1.2 Jy subsample of PSCz, although the values here are somewhat lower than for the full PSCz. However, even at high redshift, where the usage of an 1.2 Jy limited survey seems already quite prudent, the PSCz subsample stays still in line with the full PSCz.

The drop seen in the 1.2-Jy survey for high $z_{\rm max}$ is most likely responsible for the low result found by Fisher et al. (1992). It suggests that the 1.2-Jy survey is not

z_{max}	PSCz(full)	PSCz(≥1.2 Jy)	1.2-Jy survey
0.050	8.05 ± 1.54	7.28 ± 2.80	5.64 ± 2.70
0.055	8.77 ± 1.42	8.32 ± 2.62	6.05 ± 2.54
0.060	9.15 ± 1.33	8.11 ± 2.49	6.33 ± 2.41
0.065	9.16 ± 1.25	7.55 ± 2.40	5.36 ± 2.32
0.070	10.03 ± 1.18	8.12 ± 2.30	5.40 ± 2.23
0.075	9.67 ± 1.13	6.68 ± 2.23	4.46 ± 2.16
0.080	9.08 ± 1.09	6.62 ± 2.16	4.56 ± 2.10
0.085	9.26 ± 1.05	6.12 ± 2.11	4.17 ± 2.05
0.090	9.45 ± 1.02	6.32 ± 2.05	4.32 ± 2.00
0.095	9.35 ± 0.99	5.81 ± 2.02	3.94 ± 1.96
0.10	9.37 ± 0.97	5.47 ± 1.98	4.00 ± 1.93
0.12	9.96 ± 0.87	6.28 ± 1.84	4.56 ± 1.80
0.14	9.54 ± 0.84	6.16 ± 1.75	4.13 ± 1.72
0.16	9.42 ± 0.80	6.35 ± 1.68	4.26 ± 1.65
∞	7.96 ± 0.69	6.59 ± 1.56	3.76 ± 1.58

Table 8.5: The evolutionary parameter P derived with the estimator of section 6.5 for different upper redshift cut-offs and different samples. Here a straight K-correction with $\alpha = -2$, an Einstein-de-Sitter cosmology, and $z_{\min} = 0$ have been used.

perfectly complete near the flux limit. This incompleteness has been suspected by Strauss & Koranyi (1993) and was also noticed by Fisher et al. (1992) themselves for low galactic latitude.

I therefore want to argue that the low result of Fisher et al. (1992) represents most likely an underestimate that has been systematically biased low by incompleteness at high redshift and low galactic latitude. The density estimator can be rendered less sensitive to such effects if an upper redshift cut-off is included. If this is done the evolutionary estimates for the 1.2-Jy survey are found to be reasonably consistent with the PSCz survey at least out to $z_{\max} = 0.08$.

However, there is a strong indication that the PSCz survey gives somewhat higher evolutionary rates than the 1.2-Jy survey. That seems to be the case even when the full PSCz catalogue and a 1.2 limited subsample of it are compared. This difference suggests that a part of the evolutionary signal for the PSCz survey might be caused by Malmquist bias and/or more subtle flux errors (like nonlinearities or variations of the effective flux limit of the PSCz). The small difference between the 1.2-Jy survey and the 1.2 Jy limited PSCz subsample might be due to subtle differences in completeness near the galactic plane. As of yet it is unclear which of these systematic errors are relevant for the problem and it will need more work to eventually track them down. However, evolution of $P = 6.25$ has also been reported by Oliver et al. (1994) for the IRAS Faint Source Survey (FSS) which doesn't work with the PSC. For this reason, $P = 5$ seems to be a firm lower bound for the density evolution of the far-infrared luminosity function.

Summary and discussion

Summarizing the above results I conclude that there is significant evolution seen in the far-infrared luminosity function of IRAS galaxies selected from the PSC catalogue. This evolution can be characterized either as density evolution with

$$P = 9.6 \pm 1.4 \pm 1.5 \tag{8.7}$$

or pure luminosity evolution with

$$Q = 4.3 \pm 0.7 \pm 0.8. \tag{8.8}$$

The error estimates give the statistical (first) and systematic (second) error. For the statistical error I added the fluctuations in different redshift intervals (cosmic variance of $\sigma_P \approx 1.0$) in quadrature to the estimate of the statistical error ($\sigma_P \approx 1$) provided by the evolutionary estimator. The systematic error arises from uncertainties in the K-correction, the cosmological background model and due to the Malmquist bias. I estimate these uncertainties conservatively as $\sigma_P \approx 1.5$. The relative uncertainties in the luminosity evolutionary estimate are taken to be equal to the case of density evolution.

I consider possible flux errors as the major systematic uncertainty in the determination of the strength of the evolution. A nonlinear flux scale, Malmquist bias, or variations of the effective depth of the PSC can easily explain some fraction of the evolutionary rate found for the PSCz survey. It seems likely that the evolutionary results are indeed biased high by such effects, because the full PSCz catalogue gives somewhat higher evolutionary results than a 1.2 Jy limited subsample of it. Despite the uncertainty in the true evolutionary rate the final selection function should reliably give the mean expected number density of galaxies in the catalogue as a function of redshift. Hence the topological analysis of part I should be largely unaffected by the uncertainties in the true evolutionary rate.

If one recalls equation (6.13) it is found that $Q \approx 4$ means that the luminosities seem to decay with an exponential time constant of $\tau \approx \frac{1}{4H_0} = 2.4 \times 10^9\, h^{-1}\mathrm{yr}$. The density evolution model on the other side gives a result that implies that the comoving number density of sources was higher by a factor of 2.4 at redshift 0.1. These evolutionary rates are dramatic and presently difficult to understand theoretically.

Conventional theories of 'normal' galaxy formation hold that spiral galaxies experienced a relatively high and relatively constant star formation rate through most of their lifetimes. Their 'passive' evolution due to the aging of their stellar population and the smooth and slow change of the star formation rate without external influence results in very little luminosity evolution in the recent past.

At first sight the high evolutionary rates found in the infrared selected IRAS galaxies seem therefore to be in conflict with these theories. But as Lonsdale et al. (1990) point out the far-infrared selected galaxies are not normal in the conventional sense. In particular they are biased by construction towards higher infrared luminosity than optical samples. The infrared activity itself is strongly correlated to starbursts, galaxy interactions/collisions (shock excitation) and AGN activities (see also the discussion in section 7.2).

As Lonsdale et al. (1990) further argue this enhanced infrared luminosity is likely to be of highly transient nature. For example, the typical timescale of starbursts is believed to be of order $10^7 - 10^8$ yr (Scalo 1988). These timescales are much shorter than the lookback time in the PSCz sample.

Based on that one comes to a picture where the $60\,\mu$m selected sample is dominated by galaxies which are just undergoing an infrared-bright episode. Any evolution found in this population is then reflecting the evolution in this phenomenon of episodic luminosity enhancements itself and not the passive evolution of individual galaxies. Calling this process 'active' evolution Lonsdale et al. (1990) conclude that the high evolutionary rates, which I confirm in this work, require some form of active evolution and cannot be explained by passive evolution alone.

Whether or not this scenario of active evolution is in conflict with conventional theories of galaxy evolution depends on the question whether the required rate of these episodic phenomena (caused by starbursts, interactions or AGN activity) makes a significant impact on the current theoretical models of the dynamical history of the general galaxy population.

8.4 Luminosity function

The selection function and the luminosity function are related by definition (6.3). In this work I focused on a non-parametric estimate of the selection function, but in section 6.7 it was also outlined how this estimate can be used to derive the luminosity function. I also argued that this approach has some advantages compared to the traditional way of finding a non-parametric estimate of the luminosity function.

In figure 8.11 a comparison of the resulting non-parametric luminosity function of my approach with the more conventional step-wise non-parametric method is given. The latter has been applied by Saunders et al. (1990) to derive the $60\,\mu$m luminosity function of IRAS galaxies and it is their solution which is depicted in 8.11.

One can clearly identify the piecewise nature of the parameterization of the luminosity function in form of the small discontinuous jumps in this plot. Notice that these jumps are much larger in the traditional approach with rectangular luminosity bins than in the new method.

The bins of my new technique directly mimic the shape of the luminosity function. Hence it is easier to find an analytic fit to the non-parametric result. One could also afford to choose larger bins in order to bring down statistical noise in sparse surveys.

I should mention at this point that I silently introduced a slightly different definition of the luminosity function in order to simplify the comparison of my results with the literature. People like to specify the luminosity distribution in terms of

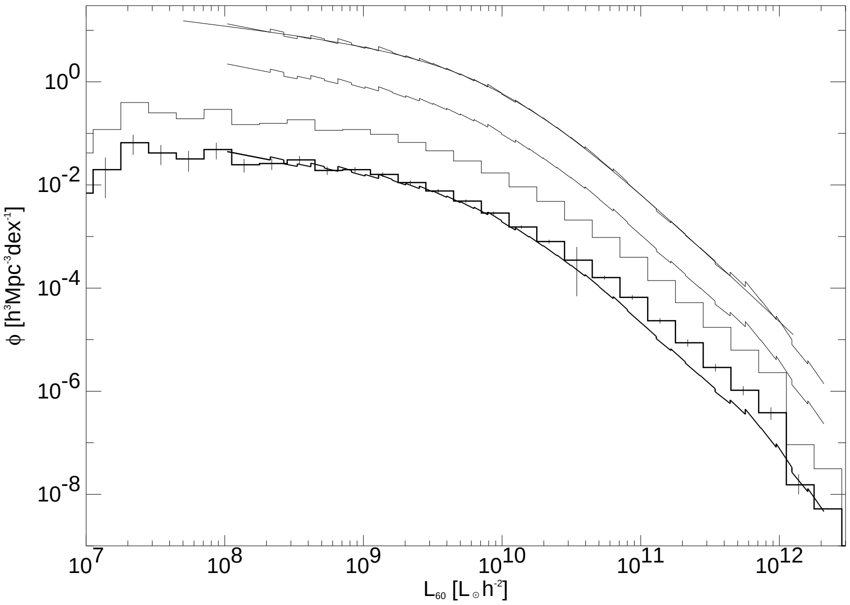


Figure 8.11: Comparison of non-parametric luminosity function estimators. The lower thick line shows my non-parametric estimate of the present day luminosity function (here an Einstein-de-Sitter universe, $\alpha = -2$, and $P = 9.6$ have been assumed), whereas the solid histogram is the non-parametric estimate of Saunders et al. (1990). For graphical clarity these curves have been plotted a second time at vertically displaced positions (two curves in the middle). Finally, the two top curves compare the analytic fit of equation (8.10) with the non-parametric estimate.

luminosity per decimal decade of luminosity. This quantity $\phi(L)$ is related to my definition of $\Phi(L)$ by

$$\phi(L) = \Phi(L)\, L\, \ln 10\,, \tag{8.9}$$

where 'ln' means the natural logarithm as usual in this work.

By minimizing χ^2 I fitted my final non-parametric result to the functional form

$$\phi(L) = C \left(\frac{L}{L_\star}\right)^{1-\alpha} \left(1 + \frac{L}{L_\star \beta}\right)^{-\beta} \tag{8.10}$$

proposed by Lawrence et al. (1986). The reduced χ^2_ν came out as 0.834 in this fit. The resulting parameters are:

$$\begin{aligned} \alpha &= 1.327^{+0.030}_{-0.031} \\ \beta &= 2.214^{+0.029}_{-0.028} \\ L_\star &= 4.534^{+0.381}_{-0.358} \times 10^9\, h^{-2} L_\odot \\ C &= 1.182^{+0.103}_{-0.117} \times 10^{-2}\, h^3 \mathrm{Mpc}^{-3}. \end{aligned} \tag{8.11}$$

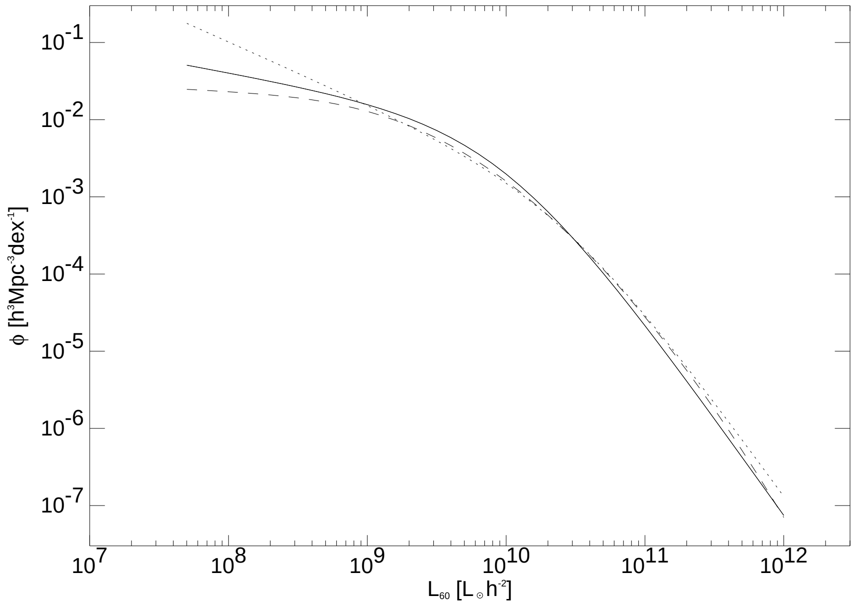


Figure 8.12: Comparison of my luminosity function (solid) with Saunders et al.'s (1990) result (dashed) and Lawrence et al.'s (1986) result (dotted). I changed the normalization of the cited results slightly in order to match my normalization in the luminosity range $10^{10}\,h^{-2}L_\odot < L < 10^{11}\,h^{-2}L_\odot$ and I rescaled Lawrence et al.'s (1986) result to $H_0 = 100\,h^{-1}$km/s/Mpc.

In 8.12 I compare this analytic fit[1] to published results. There is fairly good agreement at high luminosity, only at the faint end I find some disagreement with Saunders et al.'s (1990) almost flat end, although my faint end slope is considerably shallower than that of the early result of Lawrence et al. (1986). Most recently Andreani & Franceschini (1996) seem to have found supporting evidence for a faint end slope that is not as shallow as that of Saunders et al. (1990). But one should keep in mind that my treatment has neglected peculiar velocities entirely. Saunders et al. (1990) on the other side have tried to correct with various simple flow models for the largest systematic effects due to bulk flows in the Local Supercluster. These corrections only affect the faint end of the luminosity function and might reduce the faint end slope of PSCz when applied to it.

As a consequence of the near invariance of the selection function result under the K-correction scheme the same behaviour is encountered for the luminosity function as well. If I plot the luminosity functions for the density evolution model, the luminosity evolution model and a no evolution model together in the same figure then the curves lie right on top of each other. Since the shape of the luminosity

[1] $L_\odot = 3.85 \times 10^{33}$ erg/sec denotes the solar luminosity.

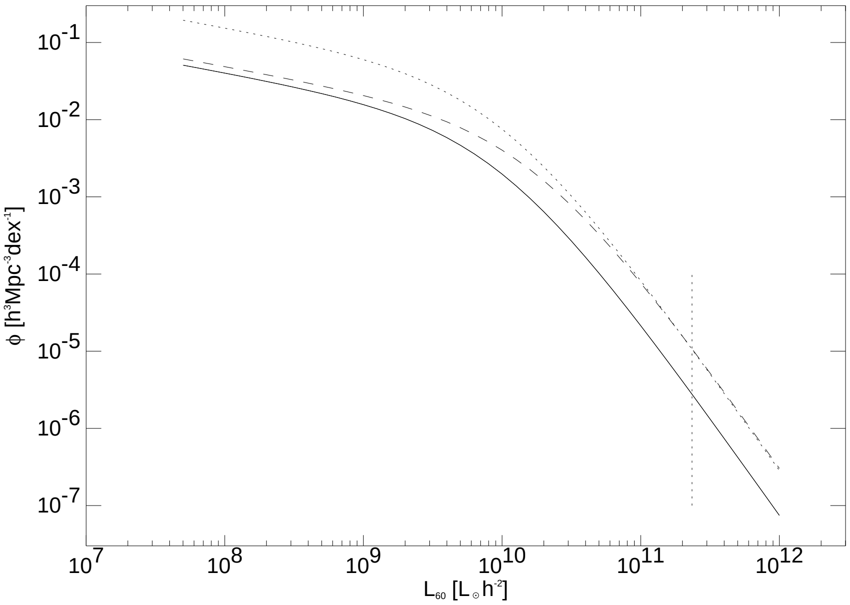


Figure 8.13: Luminosity function at the present epoch (solid) and at epoch $z = 0.15$. The dotted curve shows the density evolution ($P = 9.6$) and the dashed curve the luminosity evolution model ($Q = 4.3$). At epoch $z = 0.15$ only the bright galaxies to the right of the dotted vertical line can be observed.

function is unaffected by density evolution it is also not too surprising that one gets the right present day luminosity function even if evolution is not taken into account.

Only at earlier epochs differences in the luminosity functions of density and luminosity evolution models start to show up. Figure 8.13 displays the present day luminosity function together with the luminosity functions at epoch $z = 0.15$ for density and luminosity evolution, respectively. For the chosen evolutionary models density evolution shifts the $\phi(L)$ curve in a log-log plot in the vertical direction whereas luminosity evolution shifts it in the horizontal direction.

Differences between the two models are present only at low luminosities. However, one can see no galaxies of this kind at redshift $z = 0.15$. In fact, only galaxies with $L_{60} > 2.4 \times 10^{11}\, h^{-2} L_\odot$ at redshift $z = 0.15$ are contained in the PSCz sample, but in this regime there is virtually no difference between the luminosity functions of the density and luminosity evolution models because they behave as power laws in this regime. This clearly demonstrates that the PSCz sample is not deep enough to discern between density and luminosity evolution.

Summary and Outlook

The large-scale structure observed in the Universe represents currently one of the most active fields in cosmology. In this type of research astrophysicists examine the most extended objects that might ever be explored and face the challenge to understand and explain the formation of galaxies, the clustering of galaxies and the cosmological properties of the Universe that give rise to this structure.

It has just been in the last two decades that the clustering of galaxies has been discovered and investigated. Since then enormous observational and theoretical efforts have greatly improved our understanding of structure in the Universe. However, a large number of open questions remain and many aspects of structure formation are unsettled yet or even not understood at all.

Theoretical modeling in cosmology reaches far out in time, in space and in imagination. Physics is not a dogmatic science, hence only observational evidence can be used to support or exclude cosmological theories. In this work I have tried to test one important prediction of a certain class of cosmologies whose most prominent member is *inflation*. These theories hold that the large-scale structure we see today grew out of density perturbations which had the statistical properties of a Gaussian random field. A topological analysis of a redshift survey can be used to test this random phase hypothesis. The main objective of this study has therefore been the analysis of the topology of the mass density field observed with the help of the PSCz redshift survey.

For this purpose a fairly large number of statistical techniques have been implemented. The topology in form of the genus statistics was measured with a standard algorithm called CONTOUR, which I implemented in a new form in C [2]. Along with the genus I used some meta-statistics derived from it to measure deviations of the genus statistics from the random phase expectation. Additionally, I have proposed a new correction for the shot noise influence and a new method for the measurement of phase correlations in the density field.

As a result of this analysis I find that the PSCz topology is consistent with being that of a Gaussian random field, at least on smoothing scales $\lambda \geq 20\,h^{-1}$Mpc. Below this smoothing scale there is some evidence for phase correlations in the density field, but they are small or absent down to $\lambda = 7\,h^{-1}$Mpc. I have also not observed any significant bubble or meatball shift for $\lambda > 7\,h^{-1}$Mpc. Also, only weak evidence for

[2] All the computer codes needed for this thesis have been developed and tested by myself. For this purpose I mainly employed C and the data analysis system IDL.

a broadening of the genus curve has been found. In the regime $\lambda > 7\,h^{-1}$Mpc the genus amplitude follows reasonably well the Gaussian random phase expectation, hence it might be used to derive the second moment of the power spectrum. At smaller smoothing lengths for $\lambda \leq 7\,h^{-1}$Mpc phase correlations are clearly present. These are expected in this regime in the framework of CDM cosmologies that start with Gaussian initial conditions. Qualitatively the results for the topological analysis of the PSCz survey are therefore in agreement with inflationary cold dark matter models. However, they are in disagreement with certain other classes of theories that predict distortions of the genus curves. For example, hot dark matter models would give rise to a strong bubble shift of the genus curve which is not observed in PSCz.

In order to further examine the agreement of PSCz with CDM models I examined three large N-body simulations computed by the Virgo collaboration. These models are a standard cold dark matter cosmology (SCDM), a variant of this model with more large-scale power (τCDM) and a model with cosmological constant (ΛCDM) and similar power spectrum than τCDM. I demonstrated that these variants of CDM cosmologies differ in the amount of phase correlations as measured by the amplitude drop of the genus curve compared to the random phase expectation.

This difference has potentially a strong discriminative power among different cosmological models. In the framework of my analysis the τCDM model turned out to match the phase correlation properties of PSCz better than an ΛCDM model. Both models are approximately able to match the PSCz power spectrum equally well, hence the higher order statistics probed by the topology can lift the degeneracy between them in this case.

As it has been the case in a number of recent studies the standard cold dark matter model (SCDM) has been ruled out by this work with a high significance level. This is mainly due to its power spectrum which simply has the wrong shape. In particular, it has not enough power on large-scales to be able to explain the power seen in the PSCz density field.

Although the τCDM model was found to agree best with PSCz in this study it would be premature to try to rule out models with cosmological constant (ΛCDM) on this basis. All my findings are only valid to the extent that the simple biasing scheme I used ($b = 1$, that is no biasing of galaxies with respect to the dark matter) is a valid approximation. As of yet the details of how galaxies trace the mass are still highly uncertain and more work will be needed in the future to clarify this aspect.

Still, the topological analysis promises to become an even more powerful statistical tool in the near future, especially when it will be possible to probe the nonlinear clustering regime with it as well.

New very large redshift surveys like the Sloan Digital Sky Survey will make this possible by much better statistics and wider dynamic range than achievable with observations so far. With this kind of data much more stringent test of theoretical models and of N-body simulations should be feasible. Recent theoretical advances in the theory of genus statistics will also provide new impulses for the topological analysis. It should be clear, however, that topology is best viewed as an additional tool that complements the family of other statistical methods in the task of quantifying large-scale structure. Only when all of these techniques are applied together they reach their full power.

In part II of this work I have examined properties of the far-infrared luminosity function of IRAS galaxies in order to provide an accurate selection function for the topological analysis of the PSCz survey. In this context I proposed and applied a new non-parametric maximum likelihood estimator to determine the selection function and the luminosity function. These estimators have been generalized to allow a possible evolution in the luminosity function. In fact, it turns out that this evolution is surprisingly strong for the IRAS galaxies and can be quantified by a scaling $\propto (1+z)^{9\pm2.5}$ of the comoving number density of sources.

Hopefully this high evolutionary rate can be explained theoretically in the near future. Alternatively, a dramatic systematic effect might be uncovered that caused this high evolutionary measurement. At the moment only very strong errors in the IRAS fluxes, like a high baseline flux error or a nonlinear flux scale, appear as conceivable sources for such systematic errors. However, there is no evidence so far that a serious experimental error of this kind has occurred in the IRAS experiment. Hence the strong evolution seen in the far-infrared luminosity function must be taken seriously and represents an interesting challenge to theorists of galaxy formation and evolution.

In the near future especially the Infrared Satellite Observatory (ISO) will provide new observational input to the issue of evolution. With its sensitivity of about 1000 times that of IRAS the ISO mission will be able to check the IRAS measurements and also improve the knowledge of the spectral energy distribution of infrared bright galaxies. Since the ISO science mission has just begun recently after a successful launch in November 1995 it can be hoped that we will not have to wait too long for these answers.

Zusammenfassung in deutscher Sprache

Einführung und Motivation

Bereits die ersten astronomischen Galaxienkataloge gaben Hinweise darauf, daß die Verteilung der Galaxien im Universum eine bemerkenswerte Struktur aufweist. Die Galaxien sind nicht einfach zufällig im Raum verteilt, sondern besitzen eine klare Tendenz, sich zu Haufen von Galaxien zusammenzuballen. Diese Galaxienansammlungen vereinigen sich mitunter zu Superhaufen, die Ausdehnungen bis zu einigen Prozent des sichtbaren Universums erreichen können. Neben diesen gewaltigen Strukturen hoher Dichte gibt es auch riesige Leerräume, sogenannte *Voids*, die keine oder nur sehr wenige Galaxien enthalten.

Seitdem die ersten umfangreichen Rotverschiebungskataloge von Galaxien verfügbar wurden und damit eine dreidimensionale Darstellung ihrer Verteilung erlaubten, ist die Galaxienverteilung von verschiedenen Autoren als hierarchisch (Soneira & Peebles 1977), filamentartig (Einasto et al. 1984, Haynes & Giovanelli 1986), mauerförmig (Geller & Huchra 1989) oder blasenartig (de Lapperent et al. 1986, Broadhurst et al. 1990) beschrieben worden, je nachdem ob das Schwergewicht der jeweiligen Untersuchung auf Regionen hoher oder niedriger Galaxiendichte lag. Die Analyse und Erklärung dieser Großraumstruktur im Universum hat sich in jüngster Zeit zu einem der aktivsten Forschungsgebiete in der Kosmologie entwickelt.

In den letzten beiden Jahrzehnten sind dazu eine Reihe von statistischen Werkzeugen entwickelt worden, die eine Quantifizierung der Verklumpungseigenschaften der Galaxien erlauben. Die am häufigsten eingesetzten Größen sind dabei die zwei-Punkt-Korrelationsfunktion (*two-point correlation function*), das Power-Spektrum (*power spectrum*) der Dichtefluktuationen, oder die Varianz der Galaxienanzahl in Volumina fester Größe (*counts-in-cells analysis*). Diese statistischen Methoden werden nun als Standardverfahren auf Rotverschiebungskataloge angewandt, welche ständig an Größe und damit an Aussagekraft zunehmen. Es hat sich dabei herausgestellt, daß die beobachteten Verklumpungseigenschaften zu einer der stärksten Einschränkungen möglicher theoretischer Modelle für Strukturbildung führen. Ein erfolgreiches Modell für die Entstehung von Galaxien und von Galaxienhaufen im Universum muß in der Lage sein, diese Verklumpungseigenschaften hinreichend zu erklären und Strukturen hervorzubringen, die den beobachteten in all ihren Eigenschaften äquivalent sind.

Die zwei-Punkt-Korrelationsfunktion oder das Power-Spektrum beschreiben die statistischen Eigenschaften des Galaxiendichtefeldes allerdings nur bis zu einem gewissen Grad. Da diese beiden Größen zudem durch eine Fourier-Transformation verbunden sind, beschreiben sie lediglich einen identischen Informationsgehalt. Eine vollständige Beschreibung der Galaxienverteilung benötigt auch Korrelationsfunktionen höherer Ordnung oder alternativ dazu eine Kenntnis der Verteilung der Phasen der einzelnen Fourier-Moden. Diese Größen sind mit heutigen Galaxienkatalogen aber nur schwer und nur mit großer Unsicherheit zu messen. Allerdings bedeutet dies nicht, daß es keine anderen Meßgrößen geben kann, die in robuster Weise Information auch über Korrelationsfunktionen höherer Ordnung als zwei beschreiben.

Eine aussichtsreiche Größe für diesen Zweck bietet sich in Form der Topologie der Galaxienverteilung. Eine quantitative Charakterisierung von ihr verspricht, stark von der Morphologie der Galaxienverteilung abhängig zu sein und damit eine klare Unterscheidung zwischen morphologisch verschiedenen Modellen, die etwa blasenartige oder netzartige Strukturen aufweisen können, zu erlauben.

In jüngster Zeit sind eine Reihe topologischer Meßgrößen vorgeschlagen worden, darunter Perkulationsstatistik (Klypin 1988), Niveaukreuzungsstatistik (Ryden 1988), Genus-Statistik (Gott et al. 1986) und Minkowski-Funktionale (Mecke et al. 1994, Kerscher et al. 1996). In dieser Arbeit konzentriere ich mich auf die Genus-Statistik, die von Gott et al. (1986) in einer einflußreichen Veröffentlichung eingeführt wurde und seitdem in der Kosmologie zu einer weithin akzeptierten Technik avanciert ist.

Die Genus-Statistik ist besonders interessant, da sie eng mit traditionellen Meßgrößen für die Galaxienverteilung verbunden ist und es außerdem theoretische Vorhersagen für den erwarteten Genus einfacher Galaxienverteilungen gibt. Vor allem aber kann die Genus-Statistik zur Überprüfung einer wichtigen Vorhersage des Standardmodells für die Strukturbildung im Universum herangezogen werden. Diese Vorhersage bezieht sich auf die Natur der anfänglichen Dichtefluktuationen, die sich schließlich zu der heutigen Struktur entwickelten. Die meisten theoretischen Modelle nehmen an, daß sich Galaxien und Galaxienhaufen aus anfänglich winzigen Störungen in der Energiedichte eines sich expandierenden, homogenen Universums entwickelten. Die Dichtefluktuationen sind dabei gravitativ instabil und wachsen durch die Akkretion weiteren Materials, bis sie eventuell in einem nichtlinearen Kollaps die Strukturen bilden, die wir heute beobachten.

In vielen Modellen wird nun angenommen, daß die anfänglichen Störungen ein Gaußsches Zufallsfeld repräsentierten, das heißt, daß die Phasen der einzelnen Fourier-Moden der Dichtefluktuationen vollständig unkorreliert waren. Solch ein Dichtefluktuationsspektrum kann mit modernen inflationären Theorien zwanglos erklärt werden. Konkurrierende theoretische Modelle für den Ursprung der Struktur führen dagegen häufig zu Dichtefluktuationsspektren, die keine Gaußschen Zufallsfelder sind, sondern Phasenkorrelationen aufweisen. Beispiele für solche Szenarien sind das Explosionsmodell (*explosive amplification*) von Ostriker & Cowie (1981) oder das Modell der kosmischen Fäden (*cosmic strings*) von Vilenkin (1985).

In dieser Arbeit untersuche ich die Topologie eines neuen, sehr großen Rotverschiebungskatalogs, dem *Point Source Catalogue Redshift Survey* (PSCz). Dieser Katalog besteht aus infrarot selektierten Galaxien, ausgewählt aus einer Himmelsdurchmusterung des IRAS-Satelliten (*Infrared Astronomical Satellite*). Die Rotverschiebungen der rund 15000 Galaxien über dem gewählten Flußlimit von 0.6 Jansky im $60\,\mu$m-Band wurden mit verschiedenen optischen Teleskopen in einem mehrjährigen Beobachtungsprogramm bestimmt. Das Ergebnis dieser Arbeit ist ein Galaxienkatalog, der wegen seiner homogenen statistischen Auswahlkriterien, seiner großen Anzahl von Galaxien und der hohen Abdeckung der Himmelskugel von 84% ideal für die Analyse der Großraumstruktur im Universum geeignet ist. Insbesondere hat der PSCz Katalog damit eine größere statistische Aussagekraft in Hinsicht auf die Topologie als alle zuvor untersuchten Rotverschiebungskataloge.

Mit diesem mächtigen Datenmaterial versuche ich in dieser Arbeit, die Voraussage zufälliger Phasen für die anfänglichen Dichtefluktuationen zu prüfen. Jede eventuell gefundene Abweichung hiervon stünde im Widerspruch zu dem derzeitigen Standardmodell der Strukturbildung.

Da die Topologie auch von globalen Eigenschaften des Dichtefeldes, wie zum Beispiel der Morphologie, abhängt, testet sie auch Verklumpungseigenschaften höherer Ordnung. Galaxienverteilungen, die etwa mauerartige oder blasenartige Strukturen aufwiesen, würden zu verschiedenen Signaturen in der Genus-Statistik führen, die mit den konventionellen Methoden der zwei-Punkt-Korrelationsfunktion oder des Power-Spektrums nicht aufgedeckt werden könnten. Die topologische Analyse birgt daher das Potential, stark zwischen verschiedenen Modellen der Strukturbildung unterscheiden zu können und möglicherweise den Spielraum der theoretischen Modellbildung entscheidend einschränken zu können.

Um diese Ideen zu unterstützen, untersuche ich neben der Topologie des PSCz Katalogs in dieser Arbeit auch die Topologie von drei sehr großen numerischen Simulationen zur Strukturbildung. Diese Simulationen erkunden Varianten des Standardmodells der Strukturbildung, welches ein flaches Universum dominiert von kalter dunkler Materie (*cold dark matter*, CDM) annimmt. Es ist daher besonders interessant, die beobachtete Topologie des PSCz Katalogs mit diesen theoretischen Modellen zu vergleichen.

Um die topologische Analyse durchführen zu können, ist es auch notwendig, die mittlere erwartete Galaxiendichte als Funktion des Abstands möglichst genau zu kennen. Diese Abhängigkeit wird durch die Auswahlfunktion (*selection function*) beschrieben und kommt durch die Flußlimitierung des Katalogs zustande, denn mit zunehmendem Abstand müssen die Galaxien immer leuchtkräftiger sein, um über der vorgegebenen scheinbaren Helligkeitsbegrenzung zu liegen und somit in den Katalog aufgenommen zu werden. Die Auswahlfunktion ist eng mit der Leuchtkraftfunktion (*luminosity function*) verwandt, welche die Häufigkeitsverteilung der Leuchtkräfte in der Population der IRAS-Galaxien beschreibt.

Da eine genaue Bestimmung der Auswahl- und Leuchtkraftfunktion essentiell für die topologische Analyse ist, befaßt sich der gesamte Teil II dieser Arbeit mit der

Ableitung und Anwendung verschiedener statistischer Methoden zur Bestimmung dieser Größen. Es stellt sich dabei heraus, daß die Leuchtkraftfunktion eine starke Entwicklung mit zunehmender Rotverschiebung erfährt. Die gemessenen Entwicklungsraten sind dabei so hoch, daß sie in einer Bestimmung der Auswahlfunktion nicht vernachlässigt werden können. Allerdings sind diese hohen Entwicklungsraten physikalisch schwer zu verstehen, weshalb sich der Verdacht aufdrängt, daß möglicherweise systematische Einflüsse zu einem Teil des Entwicklungssignals führen. Die Bestimmung der Entwicklungrate und ihrer systematischen Unsicherheit wird daher ausführlich bearbeitet.

Da die topologische Analyse zunächst auf einer Rekonstruktion des Dichtefeldes des lokalen Universums basiert, kann dieses Feld kann auch zur Erzeugung kosmographischer Karten herangezogen werden. Ich demonstriere diesen Aspekt, indem ich einige solche Karten produziere und in ihnen die bedeutendsten Superhaufen (*Superclusters*) und Leerräume (*Voids*) mit bereits bekannten Strukturen identifiziere.

Im folgenden setze ich die kurze Zusammenfassung meiner Diplomarbeit fort, wobei ich vor allem die wesentlichen Schritte der Analyse und ihre rechentechnische Realisierung in ihren Grundgedanken erläutern will. Ich werde dabei weitgehend auf Formeln verzichten, die im englischen Hauptteil dieser Arbeit detailliert erläutert sind.

Die topologische Analyse

Die topologische Analyse beginnt mit der Rekonstruktion eines geglätteten Massendichtefeldes des Universums. Dieses wird aus den Galaxienpositionen gewonnen, indem sie mit dem Inversen der Auswahlfunktion gewichtet werden und das entstehende Feld geeignet geglättet wird. Der Glättungsprozeß verkompliziert sich dabei durch die Gegenwart einer „Maske“ am Himmel, die Raumwinkelbereiche beschreibt, die nicht im Katalog eingeschlossen sind. Im wesentlichen sind das die Zone starker Infrarotemission der Milchstraße und zwei vom IRAS-Satelliten unbeobachtete schmale Streifen. Die Glättung selbst wird durch Faltung mit einer Gauß-Funktion durchgeführt, wobei Dichtefelder zu verschiedenen Glättungslängen betrachtet werden.

Als topologisches Maß untersuche ich dann den Genus von Flächen konstanter Dichte. Diese Niveauflächen werden als Rand der Menge von Punkten konstruiert, die eine Dichte über einem vorgegebenen Schwellwert aufweisen. Der Genus ist vereinfacht gesagt definiert als die Anzahl der topologischen Löcher (verstanden hier als ein Loch wie das in einem Torus) minus die Anzahl der zusammenhängenden Regionen der Fläche. Um den Genus zu bestimmen, verwende ich den Satz von Gauß-Bonnet, der den Genus der Fläche mit dem Flächenintegral über die Gaußsche Krümmung verknüpft. Um diesen Satz geeignet auszunutzen, kann man die Niveaufläche ohne Änderung ihrer Topologie durch ein Netzwerk aus Polygonen (zum Beispiel durch Dreiecke oder Quadrate) approximieren. In diesem Fall wird die Gaußsche Krümmung in Dirac δ-Funktionen an den Eckpunkten des Netzwerks aus Polygonen konzentriert. Es ergibt sich daher die Möglichkeit, mit Hilfe eines Computer-Algorithmus die integrierte Gaußsche Krümmung der Fläche und damit den Genus zu bestimmen.

Da die Niveaufläche eine Funktion des verwendeten Schwellwerts ist, kann der Genus als Funktion des Schwellwerts gemessen werden. Auf diese Weise erhält man eine Genus-Kurve, welche das entscheidende diagnostische Werkzeug der topologischen Analyse darstellt. Die Genus-Kurve läßt sich am besten als Funktion des Volumenfüllfaktors darstellen, welcher den Bruchteil des Volumens über dem Schwellwert angibt. Dadurch erreicht man weitgehende Unabhängigkeit von einem möglichen *Bias*, das heißt einem nichttrivialen Zusammenhang zwischen Massen- und Galaxiendichte.

Die gemessene Genus-Kurve läßt sich nun gut mit der erwarteten theoretischen Genus-Kurve

$$g(\nu) = \frac{1}{(2\pi)^2} \left[\frac{\langle k^2 \rangle}{3}\right]^{3/2} (1-\nu^2)\ \mathrm{e}^{-\nu^2/2}$$

für ein Dichtefeld mit zufälligen Phasen vergleichen. In dieser Beziehung bezeichnet $g(\nu)$ den mittleren Genus pro Volumeneinheit als Funktion des Volumenfüllfaktors, welcher hier über eine *Error*-Funktion mit dem Parameter ν verknüpft ist. Die Größe $\langle k^2 \rangle$ bezeichnet das zweite Moment des Power-Spektrums und hängt nur von der Form nicht aber von der Normalisierung des Power-Spektrums ab. Interessanterweise hat die Genus-Kurve für Gaußsche Zufallsfelder eine universelle, symmetrische Form. Nur deren Amplitude hängt vom Verlauf des Power-Spektrums ab.

Abweichungen der gemessenen Genus-Kurve von obiger Form deuten auf nicht-Gaußsche Eigenschaften des Dichtefeldes hin. Sie werden in dieser Arbeit durch Messung der Verschiebung der Genus-Kurve, ihrer Verbreiterung und ihrer Amplitude quantifiziert.

In Hinsicht auf die Genus-Amplitude ist besonders aufschlußreich, ob sich für sie genau der theoretische Wert für ein Gaußsches Zufallsfeld ergibt. Im allgemeinen erwartet man, daß mögliche Phasenkorrelationen im Dichtefeld zu einer Reduzierung der Genus-Amplitude führen, da sie Strukturen beschreiben, die stärker zusammenhängend sind.

Um diese Amplitudenunterdrückung zu messen, habe ich die topologische Analyse mit einer Messung der Varianz des Dichtefeldes als Funktion der Glättungslänge verknüpft. Wie ich zeige, läßt sich über die logarithmische Ableitung dieser Funktion die Genus-Amplitude eines Dichtefeldes mit demselben Power-Spektrum wie PSCz aber ohne Phasenkorrelationen bestimmen. Durch Vergleich mit der gemessenen Genusamplitude erhält man dann ein Maß für Phasenkorrelationen im Dichtefeld des PSCz Katalogs, welches ausschließlich den Einfluß von Korrelationsfunktionen höherer Ordnung als zwei mißt. Wie sich durch den Vergleich mit numerischen Simulationen ferner ergibt, hat speziell dieses Maß ein starkes Potential, zwischen verschiedenen Modellvarianten unterscheiden zu können.

Um die topologische Analyse durchführen zu können, ist es auch notwendig zu klären, bis zu welcher Tiefe die Galaxiendichte hoch genug ist, um ein glattes Dichtefeld ohne signifikante Störungen durch den diskreten Charakters der Galaxienpunkte zu erhalten. Das für die topologische Analyse benutzte dreidimensionale Volumen wird dann als Funktion der Glättungslänge stets so angepaßt, daß die Genus-Kurve nicht durch den Punktcharakter der Masseteilchen (hier der Galaxien) verfälscht wird.

Dennoch ist das rekonstruierte Dichtefeld aufgrund der endlichen Anzahl von Punkten natürlich mit Schrotrauschen belastet. Der Einfluß dieses Poisson-Rauschens auf die Topologie wurde bisher nicht beachtet, jedoch zeige ich in dieser Arbeit analytisch, daß es zu einer systematischen Änderung der Genus-Amplitude führt. Wie intuitiv zu erwarten ist, wird die Genus-Amplitude in Richtung der Amplitude für ein Power-Spektrum, das weißes Rauschen beschreibt, verschoben. Die abgeleitete Beziehung für diesen Einfluß wird durch ein Monte-Carlo Experiment bestätigt und benutzt, um die gemessenen Genus-Amplituden bezüglich des Rauschens zu korrigieren.

Die Genus-Kurve hängt auch von der verwendeten Glättungslänge ab. In dieser Arbeit betrachte ich Genus-Kurven für insgesamt 16 Glättungslängen in dem Bereich $5\,h^{-1}\mathrm{Mpc} \leq \lambda \leq 56\,h^{-1}\mathrm{Mpc}$. Die spezifische Wahl dieses Bereiches ergibt sich aus der Forderung, eine möglichst zuverlässige Statistik zu erreichen, was eine große Anzahl von aufgelösten Strukturelementen erfordert. Dies wird allerdings durch die Anzahl von Galaxien und die limitierte Tiefe des Katalogs begrenzt. Die höchste Zahl unabhängiger Strukturelemente und damit die verläßlichste Genus-Messung ergibt sich dabei für $\lambda = 10\,h^{-1}\mathrm{Mpc}$.

Ergebnisse für den PSCz Katalog

Die berechneten Genus-Kurven für den PSCz Katalog zeigen über den ganzen Bereich der verwendeten Glättungsskalen keine signifikanten Abweichungen von der Form, die für Gaußsche Zufallsfelder erwartet wird. Insbesondere habe ich keine bedeutenden Verschiebungen oder Anzeichen von Verbreiterungen der Genus-Kurven für Glättungslängen $\lambda > 7\,h^{-1}\mathrm{Mpc}$ gemessen. Die Messung der Amplitudenunterdrückung ergibt eine leichte Absenkung für $\lambda \leq 10\,h^{-1}\mathrm{Mpc}$, ist für größere Glättungslängen jedoch sehr gut konsistent mit keiner Amplitudenabsenkung. Dies bedeutet, daß die Genus-Kurven für den PSCz Katalog konsistent mit Gaußschen Anfangsbedingungen zu Beginn der Strukturbildung sind.

Die Messung der Varianz des Dichtefeldes ist ebenfalls interessant, da sie eine Einschränkung der Form und Normalisierung des Power Spektrums erlaubt. Wie in einer Reihe anderer statistischer Untersuchungen kleinerer Rotverschiebungskataloge ebenfalls festgestellt, ergibt sich dabei mehr Power auf großen Skalen, als etwa vom Standardmodell mit kalter dunkler Materie (SCDM) vorhergesagt. Dieses wird daher in seiner einfachsten Form von dieser Arbeit erneut ausgeschlossen.

Die Normalisierung des Power-Spektrums auf der Skala typischer Galaxienhaufen erhalte ich zu $\sigma_8 = 0.8$. Interessanterweise ist dieser Wert höher, als er für kleinere

IRAS Kataloge in den letzten Jahren gefunden wurde, und nun deutlich näher an dem Wert von $\sigma_8 \approx 0.95$ für optische Galaxien. Dies setzt einem relativen *Bias* zwischen optisch und infrarot selektierten Galaxien recht enge Grenzen.

Vergleich mit numerischen Simulationen

Um das theoretische Ergebnis für die Genus-Statistik derzeit gängiger kosmologischer Modelle zu erhalten, untersuche ich sehr große numerische Simulationen von drei Modellvarianten, die als dominierenden Anteil am Materieinhalt des Universums kalte dunkle Materie (*cold dark matter*, CDM) annehmen. Im Standardmodell (SCDM) dieses Szenarios betrachtet man ein flaches Universum, das vollständig von kalter dunkler Materie dominiert ist. Unter der Annahme eines anfänglichen skaleninvarianten Harrison-Zel'dovich Spektrums, wie es etwa von der Inflationstheorie vorhergesagt wird, läßt sich das Power-Spektrum zu Beginn der Strukturbildung in linearer Theorie berechnen. Die Computersimulationen entwickeln dann eine Darstellung des Dichtefeldes durch eine große Anzahl von Punktteilchen bis zur Jetztzeit, das heißt auf kleinen Skalen weit in den nichtlinearen Bereich hinein.

Neuere Untersuchungen der Struktur im Universum befinden sich jedoch in ernsthaftem Konflikt mit diesem Modell, da sie mehr Power auf großen Skalen messen, als das SCDM Modell bietet. Um diese Diskrepanz, die ich erneut mit PSCz bestätige, aufzuheben, kann man entweder weitere physikalische Prozesse annehmen, die dem CDM Modell die richtige Anfangsform des Spektrums verleihen oder das kosmologische Hintergrundmodell abändern. Ein Beispiel für die erste Möglichkeit sind zerfallende τ-Neutrinos im sogenannten τCDM Modell. Alternativ dazu kann man ein offenes Universum niedriger Dichte (OCDM) annehmen, oder eine kosmologische Konstante einführen (ΛCDM), so daß das Universum zwar niedrige Dichte hat, aber flach bleibt.

In dieser Arbeit untersuche ich drei Simulationen der Typen SCDM, τCDM und ΛCDM, die von der Virgo-Arbeitsgruppe ausgeführt wurden und zu den größten kosmologischen Simulationen gehören, die je berechnet wurden.

Die Messung der Genus-Kurven für diese Simulationen ergibt zunächst, daß die Genus-Kurven der universellen Form für ein Dichtefeld mit zufälligen Phasen sehr gut folgen, abgesehen von einer sehr kleinen *Bubble*-Verschiebung der ΛCDM und τCDM Modelle bei niedrigen Glättungslängen.

Bei einer Analyse der Amplitudenunterdrückung ergeben sich aber deutliche Unterschiede zwischen den Modellen. Alle drei Varianten beginnen mit zufälligen Phasen, was die Genus-Kurven auch bestätigen, indem sie bei großen Glättungsskalen, bei denen nur Moden, die immer noch in der linearen Entwicklung sind, erhalten bleiben, keine Amplitudenunterdrückung feststellen. Bei geringeren Glättungslängen findet allerdings eine deutliche Absenkung der Genusamplitude statt. Dieser Effekt ist sehr stark für das ΛCDM Modell und weit schwächer für das SCDM Modell, während das τCDM Modell eine mittlere Stellung einnimmt.

Es erweist sich nun als fruchtbar, das Verhalten des PSCz Katalogs mit den Simulationen zu vergleichen. Die Betrachtung der Form des Power Spektrums allein ist bereits in der Lage, das SCDM Modell praktisch auszuschließen, während das τCDM und das ΛCDM Modell in etwa gleich guter Weise in der Lage sind, die Beobachtungen zu reproduzieren. Die Betrachtung der Amplitudenunterdrückung der Genus-Kurven favorisiert allerdings stark das τCDM Modell, da so starke Phasenkorrelationen wie im ΛCDM Modell für den PSCz Katalog nicht beobachtet werden. In dieser Anwendung erweist sich die topologische Analyse daher als machtvolles Instrument, um theoretische Modellbildungen einzuschränken. Es wäre allerdings verfrüht, auf Grund dieses Ergebnisses das ΛCDM Modell bereits als ausgeschlossen zu betrachten. Es ist möglich, daß etwa ein nichttrivialer *Bias* zwischen Galaxien- und Massenverteilung die Übereinstimmung zwischen PSCz und dem ΛCDM Modell verbessern könnte.

Ich ziehe die numerischen Simulationen auch zu einer künstlichen Beobachtung eines Modelluniversums heran. Dabei wird ein Galaxienkatalog konstruiert, der die statistischen Eigenschaften des PSCz Katalogs exakt nachbildet. Mit diesem künstlichen Katalog habe ich dann die Korrektheit und die Robustheit der statistischen Verfahren und Computercodes, die in der Analyse des PSCz Katalogs verwendet werden, getestet.

Auswahlfunktion und Leuchtkraftfunktion

Die Leuchtkraftfunktion beschreibt die Häufigkeitsverteilung von Galaxien bezüglich ihrer Leuchtkraft, während die eng verwandte Auswahlfunktion die mittlere erwartete Anzahldichte von Galaxien als Funktion der Rotverschiebung für einen bestimmten Galaxienkatalog angibt. Da speziell die Auswahlfunktion zur Rekonstruktion des Dichtefluktuationsfeldes notwendig ist, befaßt sich der zweite Teil dieser Arbeit ausführlich mit der Bestimmung von Leuchtkraft- und Auswahlfunktion. Letztere wird dann in der topologischen Analyse des ersten Teils verwendet, was bedeutet, daß dieser zweite Teil in der eigentlichen Anwendung auf den PSCz Katalog zuerst berechnet werden muß.

Die Bestimmung der Leuchtkraft- beziehungsweise Auswahlfunktion erfordert den Einsatz moderner statistischer Verfahren. Eine optimale Methode sollte dabei von Verklumpungseigenschaften unbeeinflußt sein und möglichst wenig Wissen über die Form der erwarteten Leuchtkraftfunktion voraussetzten.

Um der ersten Forderung gerecht zu werden, verwende ich sogenannte *Maximum-Likelihood*-Methoden, welche sich geeignet formulieren lassen, so daß Dichteinhomogenitäten keinen störenden Einfluß nehmen. Falls die analytische Form der zu bestimmenden Funktion bereits bekannt ist, führen diese Verfahren in der Regel auf ein Maximierungsproblem, bei dem eine Wahrscheinlichkeit in Hinblick auf wenige Parameter zu optimieren ist (*parametric maximum likelihood estimator*).

Allerdings ist die Form der Auswahlfunktion beziehungsweise der Leuchtkraftfunktion zunächst unbekannt. In dieser Arbeit entwickle ich daher eine neue, nichtparametrische Maximum-Likelihood-Methode zur Bestimmung der Auswahlfunktion, welche deren Form zunächst fast völlig frei läßt. Diese Technik erlaubt es, genaue analytische Näherungsformen für die Auswahlfunktion und die Leuchtkraftfunktion zu finden. Sie hat ferner einige Vorteile gegenüber einem älteren Verfahren, das bisher als Standardmethode eingesetzt wurde, um Leuchtkraftfunktionen nichtparametrisch abzuschätzen.

Mit ähnlichen Techniken wie zur Bestimmung der Leuchtkraftfunktion kann auch der radiale Verlauf der Dichte gefunden werden. Diese Berechnung ist unabhängig von der Leuchtkraftfunktion, wenn diese als zeitlich konstant angenommen wird. Unter dieser Annahme erhalte ich das zunächst überraschende Ergebnis, daß die Dichte in jeder Richtung anzusteigen scheint.

Dieser Effekt läßt sich jedoch durch eine zeitliche Entwicklung der Leuchtkraftfunktion erklären, wenn diese in der jüngeren Vergangenheit eine höhere Anzahldichte von leuchtetenden Galaxien und/oder hellere Galaxien aufwies. Alternativ sind systematische Einflüsse, etwa Fehler in den Infrarot-Flüssen (*Malmquist*-Effekt), eine nichtlineare Flußskale, eine fehlerhafte K-Korrektur oder ein falsches kosmologisches Hintergrundmodell als Ursache denkbar und entsprechend zu prüfen.

Die Bestimmung der Auswahlfunktion verkompliziert sich durch die starke Entwicklung, die es unabdingbar macht, sie sorgfältig in den Formalismus zur Bestimmung der Auswahl- und Leuchtkraftfunktion einzubinden. Zusätzlich muß die eigentliche Entwicklungsrate mit einem separaten statistischen Verfahren bestimmt werden.

Die gefundenen Entwicklungsraten sind sehr hoch und physikalisch nicht leicht zu erklären. Es muß allerdings betont werden, daß es sich hier um eine Entwicklung in der *Infrarotleuchtkraft* der Galaxien handelt. Selbst hohe Entwicklungsraten hierfür könnten ohne weiteres mit einer sich nicht entwickelnden optischen Leuchtkraftfunktion verträglich sein.

Ich habe verschiedene systematische Ursachen als mögliche Quellen des Entwicklungssignals analysiert. Dabei ergibt sich in Bezug auf K-Korrektur und Fluß-Fehler ein gewisser Spielraum, der die gemessene Evolutionsrate nach unten korrigieren könnte. Allerdings erforderte es extrem starke Fehler in den gemessenen Flüssen oder eine deutlich nichtlineare Flußskala, um die gesamte gefundene Entwicklung einem systematischen Fehler zuzuschreiben. Für solche Probleme mit den IRAS Daten gibt es aber bisher keine überzeugenden Hinweise, daher muß das PSCz Ergebnis als starke Evidenz für eine Evolutionsrate angesehen werden, die vermutlich über derjenigen für optisch ausgewählte Galaxien liegt.

Ausblick

Die heutigen theoretischen Modelle der Kosmologie erstrecken sich bis an die Grenzen unserer physikalischen Erfahrung, bis an den Anfang von Raum und Zeit. Sie unternehmen nichts geringeres als den Versuch, das Universum als solches zu verstehen. Vor allem möchten sie erklären, wie sich die mächtigen Sternsysteme und Galaxienhaufen gebildet haben, die unser eigenes Sonnensystem nur wie ein unbedeutendes Sandkorn in einer riesigen Wüste erscheinen lassen. Manche der hierzu vorgeschlagen Theorien über den Anfang des Universums oder den Ursprung der Struktur sind gewagt und bedürfen der Bestätigung durch Beobachtungen, um sie vom Rang einer physikalischen Spekulation zu einer glaubhaften Theorie zu erheben.

Erst in jüngster Zeit ist es in der Kosmologie möglich geworden, theoretische Modelle der Strukturbildung durch direkte Beobachtungen zu bestätigen oder zu verwerfen. Dies hat bereits zu einer dramatischen Verbesserung unseres Wissens über die Entwicklung des Universums geführt.

In dieser Arbeit konnte ich die Hypothese über die zufälligen Phasen für die anfänglichen Dichtefluktuationen im Rahmen der topologischen Analyse voll bestätigen. Dies bedeutet noch nicht zwingend, daß Modelle mit Phasenkorrelationen ausgeschlossen sind, setzt aber sehr starke Schranken für sie. Insbesondere sollten Modelle, die topologische Defekte als „Saatkörner“ zur Strukturbildung vorschlagen, zu Signaturen in der Genus-Statistik führen, die in dieser Untersuchung nicht beobachtet wurden.

Im Hinblick auf die untersuchten numerischen Simulationen zur Strukturbildung favorisiert die topologische Analyse die τCDM Variante (mehr Power auf großen Skalen als SCDM, zum Beispiel durch zerfallende schwere τ-Neutrinos) gegenüber einer Variante mit einer kosmologischen Konstante (ΛCDM). Es sollte hierbei für eine weiterführende Studie interessant sein, noch weitere Modellvarianten, zum Beispiel mit Anteilen heißer dunkler Materie (*mixed hot dark matter*) zu testen.

In naher Zukunft dürfte sich die diagnostische Kraft der topologischen Analyse noch entscheidend erweitern lassen. Neue, extrem große Rotverschiebungskataloge wie der *Sloan Digital Sky Survey* werden in wenigen Jahren Kataloge liefern, deren Galaxienzahl um mehrere Größenordnungen über der von PSCz liegt. Dadurch eröffnet sich die Aussicht, eine sehr viel geringere statistische Unsicherheit für die Topologie zu erreichen und sie insbesondere weit in den nichtlinearen Bereich der Strukturbildung ausdehnen zu können. Durch Kombination mit großen Simulationen werden dann sehr präzise Tests kosmologischer Modelle möglich werden.

Daneben sind erhebliche Fortschritte mit anderen Methoden zu erwarten, etwa mit dem COBRAS/SAMBA-Satelliten zur Untersuchung des Mikrowellenhintergrunds. Daher scheint die Annahme nicht übertrieben optimistisch, daß wir binnen 10 bis 15 Jahren die wesentlichen Eckpfeiler der Kosmologie wie Hubble-Konstante, Hintergrundgeometrie und Materieinhalt des Universums *genau wissen werden.*

Es bestehen ebenfalls hervorragende Aussichten, schon sehr bald die hohe Evolutionsrate, die für infrarot selektierte Galaxien mit dem IRAS Satelliten gemessen wird, entweder zu bestätigen oder zu falsifizieren. Der neue Infrarotsatellit ISO (*Infrared Space Observatory*) der ESA (*European Space Administration*) besitzt eine mehr als tausendmal so große Empfindlichkeit wie IRAS und wird daher in der Lage sein, viele der wissenschaftlichen Ergebnisse der IRAS Mission zu überprüfen und natürlich auch enorm zu erweitern. Nach dem erfolgreichen Start von ISO im November 1995 dürften neue Daten und Erkenntnisse nicht mehr lange auf sich warten lassen.

Appendix A

Basic Relations and Notation

In this appendix I will first review a number of basic equations of cosmology that will be familiar to every cosmologist. Then I introduce a few standard tools which are widely used to describe the statistics of galaxy clustering. Because some details of the definitions of various quantities differ between authors (in particular the conventions for Fourier transforms) the following sections are also intended to set a consistent notation for the main part of the discussion. Any confusion that might arise in a comparison of this work with different authors due to missing factor of 2π or the like can be hopefully resolved with a reference to the following sections.

References for the following material include most textbooks on cosmology or large-scale structure formation (e.g. Peacock, Heavens & Davies 1990, Padmanabhan 1993, Weinberg 1972, Peebles 1993).

A.1 Basic cosmological equations

In this section I give a brief summary of basic cosmological relations that are needed in this study. I will assume that the cosmological principle holds, that is the Universe is homogeneous and isotropic, at least on large enough scales. Then the spacetime of the Universe can be described by the Robertson-Walker metric. I write it in the form

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\,\mathrm{d}\Omega^2\right). \tag{A.1}$$

Here I use spherical coordinates (r, θ, ϕ) to describe the spatial coordinates and $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$ is a convenient shorthand for the square of the solid angle. There are three different global geometries for the Universe, commonly refered to as closed, flat and open. It is always possible to rescale r such that k adopts one of the following values:

$$k = \begin{cases} +1 & \text{closed} \\ 0 & \text{flat} \\ \text{-1} & \text{open} \end{cases} \tag{A.2}$$

From now on I will set $c = 1$ to simplify the notation. All factors of c can be restored just by dimensional analysis of the equation at hand. Occasionally I will also restore c to emphasize a special form or to make it easier to calculate numerical results.

The Hubble constant is the expansion rate at the present time t_0:

$$H_0 = \left.\frac{\dot{a}}{a}\right|_{t_0}. \tag{A.3}$$

Usually I will hide the present ignorance of the exact value of H_0 by parameterizing it with

$$H_0 = 100\,h\,\frac{\mathrm{km/s}}{\mathrm{Mpc}}. \tag{A.4}$$

The dimensionless fudge factor h appears in most quantitative results in this work. Once one has picked a favorite value for it, like $h = 0.7$ corresponding to $H_0 = 70\,\mathrm{km/s/Mpc}$, it can be used to convert them to final physical values.

The deceleration parameter q_0 is defined by

$$q_0 = -\left.\frac{\ddot{a}a}{\dot{a}^2}\right|_{t_0} \tag{A.5}$$

and is frequently used to parameterize cosmological background models.

Einstein's field equations simplify for the Robertson-Walker metric and the assumption that the contents of the universe can be described as an ideal fluid into the Friedmann equations. They are

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \tag{A.6}$$

and

$$\frac{\dot{a}^2 + k}{a^2} = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3}. \tag{A.7}$$

Here I also allowed for a cosmological constant Λ in Einstein's equations. If the Newtonian limit of them is examined it is easy to see that Λ effectively acts as an energy density $\rho_\Lambda c^2 = c^4\Lambda/8\pi G$ of the vacuum.

In the following I will always neglect the pressure p since the universe is completely matter dominated at the present time ($\rho \gg p$) and this is equally well the case for the lookback times I will encounter in the PSCz redshift survey.

I introduce the following definitions for the critical density ρ_0 and the cosmological Ω's:

$$\rho_c = \frac{3H_0^2}{8\pi G} \tag{A.8}$$

$$\Omega_0 = \frac{\rho}{\rho_c} \tag{A.9}$$

$$\Omega_\Lambda = \frac{\Lambda}{3H_0^2} \tag{A.10}$$

$$\Omega_c = -\frac{k}{a_0^2 H_0^2}. \tag{A.11}$$

Here and in the following ρ_0 and a_0 refer to the values of the corresponding quantities at the present time.

If I divide the Friedmann equations (A.6) and (A.7) by H_0^2 and evaluate them at $t = t_0$ they take on the forms

$$\Omega_0 + \Omega_c + \Omega_\Lambda = 1 \tag{A.12}$$

and

$$\frac{1}{2}\Omega_0 - \Omega_\Lambda = q_0. \tag{A.13}$$

Since $\Omega_c = 0$ implies a flat and $\Omega_c < 0$ a closed universe one reaches closure density for $\Omega_0 + \Omega_\Lambda \geq 1$.

The perhaps simplest cosmological model is a flat universe with zero cosmological constant. It is commonly called Einstein-de-Sitter universe and is characterized by $\Omega_0 = 1$ and $\Omega_\Lambda = \Omega_c = 0$. Modern cosmology assumes that there is a class of yet unidentified particles contributing as dark matter to the mass density. There is plenty of indirect evidence that the dark matter exists, although no one has found such particles so far.

Since it is directly linked to observations the cosmological redshift z is of importance in the study of the Universe. Usually an object some distance away is seen redshifted by $z = (\lambda_{\mathrm{obs}} - \lambda_{\mathrm{em}})/\lambda_{\mathrm{em}}$. If one neglects peculiar motions and solves the geodesic equation of the light ray one finds that the redshift can be expressed as

$$z + 1 = \frac{a_0}{a(t_{\mathrm{emission}})}. \tag{A.14}$$

This redshift arises because space itself expands and is therefore quite different from the ordinary Doppler effect. The above relation characterizes both the distance of the object (see below) and the time of the emission of the light, that is the epoch of its emission. The redshift serves therefore as convenient time variable. The present epoch is given by $z = 0$ and the past extends to higher redshifts.

I just mentioned that the redshift also gives the distance of the object. To find this relation I first rewrite the Friedmann equation (A.7) again. It can be cast into

$$\frac{\dot{a}}{a} = H_0 E(z), \tag{A.15}$$

where

$$E(z) = \sqrt{\Omega_0(1+z)^3 + \Omega_c(1+z)^2 + \Omega_\Lambda}\,. \tag{A.16}$$

Often one needs to follow a light ray backwards to its source. Since it obeys $\mathrm{d}s^2 = 0$ one has $\mathrm{d}l = -\mathrm{d}t = a\ \mathrm{d}r/\sqrt{1 - kr^2}$. Differentiating equation (A.14) gives $\mathrm{d}z/\mathrm{d}t = -(z+1)H_0 E(z)$. Then the physical distance $l(z)$ a light ray coming from epoch z has traveled can be calculated by integrating

$$\frac{\mathrm{d}l}{\mathrm{d}z} = \frac{1}{(1+z)H_0 E(z)}. \tag{A.17}$$

Of greater interest is usually the comoving distance $r(z)$ of the source. It follows from

$$\frac{\mathrm{d}r/\mathrm{d}z}{\sqrt{1-kr^2}} = \frac{1}{H_0 E(z)}. \tag{A.18}$$

Upon integration one finds for the three different values of k the following expressions:

$$\frac{1}{a_0 H_0} \int_0^z \frac{\mathrm{d}z'}{E(z')} = \begin{cases} \sinh^{-1} r(z) & \text{for } k=-1 \\ r(z) & \text{for } k=0 \\ \sin^{-1} r(z) & \text{for } k=+1 \end{cases} . \tag{A.19}$$

In the limit of small distances the three forms become identical and reduce to the linear Hubble law. For the flat universe the integration is straightforward and gives

$$r(z) = \frac{2}{a_0 H_0}\left(1 - \frac{1}{\sqrt{1+z}}\right). \tag{A.20}$$

The physical volume element in the Robertson-Walker metric can be written as

$$\mathrm{d}V_{\mathrm{phys}} = a^3 \frac{\mathrm{d}\Omega\, r^2\, \mathrm{d}r}{\sqrt{1-kr^2}}. \tag{A.21}$$

Often one needs the volume traversed by a bundle of rays of opening solid angle $d\Omega$ at epoch z and redshift interval $\mathrm{d}z$. This is now found to be

$$\mathrm{d}V_{\mathrm{phys}} = a^3 \frac{r^2}{a_0 H_0 E(z)}\, \mathrm{d}\Omega\, \mathrm{d}z. \tag{A.22}$$

Since the physical volume of some coordinate region scales $\propto a^3 = a_0^3(1+z)^3$ I introduce the comoving volume element $\mathrm{d}V$ as

$$\mathrm{d}V = (1+z)^3\, \mathrm{d}V_{\mathrm{phys}} = a_0^3 \frac{r^2(z)}{a_0 H_0 E(z)}\, \mathrm{d}\Omega\, \mathrm{d}z. \tag{A.23}$$

At the present epoch ($z = 0$) one has $\mathrm{d}V = \mathrm{d}V_{\mathrm{phys}}$. The comoving volume doesn't change with the expansion of the universe, so quantities which measure densities (like the number density of galaxies or the mass density) stay constant in time if they are defined in terms of the comoving volume and the absolute number of the objects the density refers to doesn't change. For this reason I will usually employ only comoving number densities along with the notion of the comoving volume.

Let me now finish up this review of cosmographic relations by looking at the important spectral properties of an observed object. Suppose a source with intrinsic spectral energy distribution $L_\nu(\nu)$ and bolometric luminosity L is examined. If it lies at radial coordinate r with corresponding redshift z one will observe at frequency ν a flux density

$$f_\nu(\nu) = \frac{L_\nu\left(\nu(1+z)\right)}{4\pi a_0^2 r^2 (1+z)}. \tag{A.24}$$

The observed total flux l is given by

$$l = \frac{L}{4\pi a_0^2 r^2 (1+z)^2}. \tag{A.25}$$

Sometimes one also uses the luminosity distance d_L and the angular size distance d_A defined by

$$d_L = \sqrt{\frac{L}{4\pi}} = a_0 r(1+z) \tag{A.26}$$

and

$$d_A = a_0 r/(1+z) \tag{A.27}$$

to describe an object but I will not make use of these quantities in this work.

A.2 Density contrast and Fourier transforms

The principal object of this study is the cosmological mass density described by the field $\rho(\mathbf{r})$. Usually I will use the density contrast

$$\delta(\mathbf{r}) = \frac{\rho(\mathbf{r}) - \bar{\rho}}{\bar{\rho}} \tag{A.28}$$

as a convenient dimensionless quantity to describe the density fluctuations.

Fourier representations of this field will turn out to be very useful. In order to introduce a convention for this transform let me assume that the universe is described in terms of a large periodic box of volume $V = L^3$. This assumption is better suited for most applications than a continuous Fourier transform because one knows the field only for a limited volume by observations and in numerical simulations of large-scale structure a periodic box is the usual and necessary assumption. In this periodic box one has the following orthogonality and closure relations:

$$\frac{1}{V}\int_V \mathrm{d}\mathbf{x}\, \mathrm{e}^{i(\mathbf{k}-\mathbf{k}')} = \delta^{(\mathrm{D})}_{\mathbf{k},\mathbf{k}'} \tag{A.29}$$

$$\frac{1}{V}\sum_{\mathbf{k}} \mathrm{e}^{i\mathbf{k}\mathbf{x}} = \delta^{(\mathrm{D})}(\mathbf{x}). \tag{A.30}$$

I use the notation $\delta^{(\mathrm{D})}$ for Dirac δ-functions and Kronecker symbols, respectively, in order to avoid any confusion with the density contrast. The sum over $\mathbf{k}$-vectors extends only over the set $\mathbf{k} \in \left\{\frac{2\pi}{L}(n_1, n_2, n_3) \mid n_1, n_2, n_3 \in \mathcal{N}\right\}$ allowed by the periodic boundary conditions.

I will use the following convention for the Fourier transform, here exemplified for the density contrast:

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}}\, \mathrm{e}^{i\mathbf{k}\mathbf{x}} \tag{A.31}$$

$$\delta_{\mathbf{k}} = \frac{1}{V}\int_V \delta(\mathbf{x})\, \mathrm{e}^{-i\mathbf{k}\mathbf{x}}\, \mathrm{d}\mathbf{x}\,. \tag{A.32}$$

In the continuum limit the k-space volume element $(\Delta k)^3 = \left(\frac{L}{2\pi}\right)^3$ becomes infinitesimally small and one can use the continuous Fourier transform pairs

$$\delta(\mathbf{x}) = \int \delta(\mathbf{k})\, \mathrm{e}^{i\mathbf{k}\mathbf{x}}\, \mathrm{d}\mathbf{k} \tag{A.33}$$

$$\delta(\mathbf{k}) = \frac{1}{(2\pi)^3} \int \delta(\mathbf{x})\, \mathrm{e}^{-i\mathbf{k}\mathbf{x}}\, \mathrm{d}\mathbf{x} \tag{A.34}$$

instead. The values of the discrete and continuous Fourier transforms for a certain mode $\mathbf{k}$ are then related by

$$\delta(\mathbf{k}) = \left(\frac{L}{2\pi}\right)^3 \delta_{\mathbf{k}}. \tag{A.35}$$

Often the convolution theorem can be used to simplify calculations. If $f = g \otimes h$ denotes a convolution of the form $f(\mathbf{x}) = \int g(\mathbf{x} - \mathbf{x}')h(\mathbf{x}')\,\mathrm{d}\mathbf{x}'$ one has

$$f_{\mathbf{k}} = V\, g_{\mathbf{k}}\, h_{\mathbf{k}} \tag{A.36}$$

in the periodic case and

$$f(\mathbf{k}) = (2\pi)^3\, g(\mathbf{k})\, h(\mathbf{k}) \tag{A.37}$$

for the continuum limit.

Since the density contrast is a real quantity its Fourier components satisfy the relation

$$\delta_{-\mathbf{k}} = \delta^*_{\mathbf{k}}. \tag{A.38}$$

A.3 Autocorrelation function and power spectrum

Perhaps the most basic tool to describe the statistics of the density field is the two-point correlation function defined by

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{x})\delta(\mathbf{x} - \mathbf{r}) \rangle\,. \tag{A.39}$$

Another common name for this quantity is autocorrelation function. The angular brackets denote an ensemble average over many realizations of universes with the same statistical properties. I will assume ergodicity, which means that the ensemble average can be replaced by a spatial average over a large enough volume of the Universe we find around us.

In any real observation the accessible volume is always limited so the spatial average can only replace the ensemble average to a limited extend. The resulting uncertainty is called cosmic variance and it will be small if the portion of the Universe we observe is large enough to represent a fair sample for the measured quantity. Usually I will drop any explicit distinction between spatial and ensemble averages and replace ensemble averages by spatial ones silently.

If statistical homogeneity and isotropy of the Universe is assumed the autocorrelation function will not depend on the location $\mathbf{x}$ and the direction of $\mathbf{r}$ but only on the magnitude $r = |\mathbf{r}|$ of the separation.

Upon insertion of the Fourier representation the two-point function can be expressed as

$$\xi(\mathbf{r}) = \sum_{\mathbf{k}} |\delta_{\mathbf{k}}|^2 \, \mathrm{e}^{i\mathbf{kr}}. \tag{A.40}$$

Let me define the power $P_{\mathbf{k}}$ in mode $\mathbf{k}$ as $P_{\mathbf{k}} = |\delta_{\mathbf{k}}|^2$. Then $P_{\mathbf{k}}$ and $\xi(\mathbf{r})$ are Fourier transforms of each other.

This motivates the definition of $P(\mathbf{k})$ for the continuous case as the Fourier transform of $\xi(\mathbf{r})$:

$$P(\mathbf{k}) = \frac{1}{(2\pi)^3} \int \xi(\mathbf{r}) \, \mathrm{e}^{-i\mathbf{kr}} \, \mathrm{d}\mathbf{r}. \tag{A.41}$$

Then one has the correspondence

$$P(\mathbf{k}) = \left(\frac{L}{2\pi}\right)^3 P_{\mathbf{k}}, \tag{A.42}$$

and $P(\mathbf{k})$ has the meaning of power per unit mode in $\mathbf{k}$-space. The commonly used power spectrum $P(k) = \langle P(\mathbf{k}) \rangle$ is the ensemble averaged (or equivalently angular averaged) version of this quantity. Because theoretical predictions are often phrased in terms of $P(k)$ one usually expresses measurements of the power spectrum obtained from redshift surveys or computer simulations with the help of equation (A.42) in terms of $P(k)$ as well.

The volume one considers will always be finite. Hence one cannot guarantee that the angular dependence in $\xi(\mathbf{r})$ vanishes exactly. Therefore I operationally define $\xi(r)$ as the angular average over all directions of $\mathbf{r}$:

$$\xi(r) = \int \frac{\mathrm{d}\Omega}{4\pi} \xi(\mathbf{r}). \tag{A.43}$$

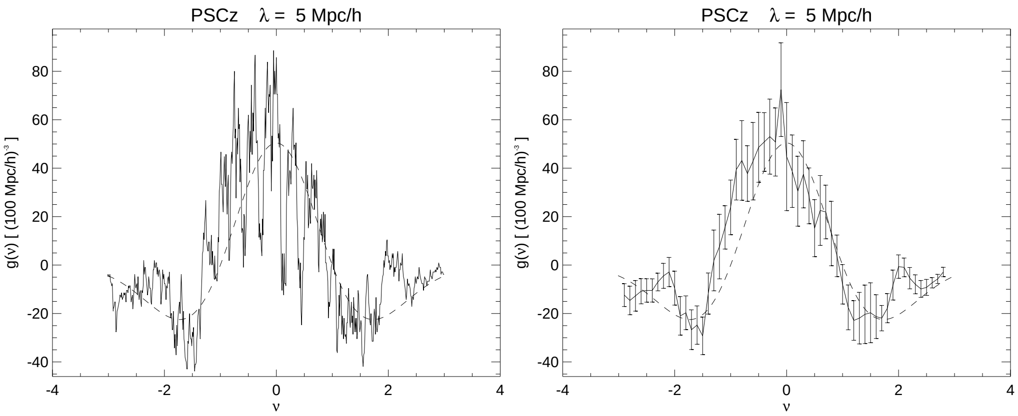
As a consequence of this definition one finds

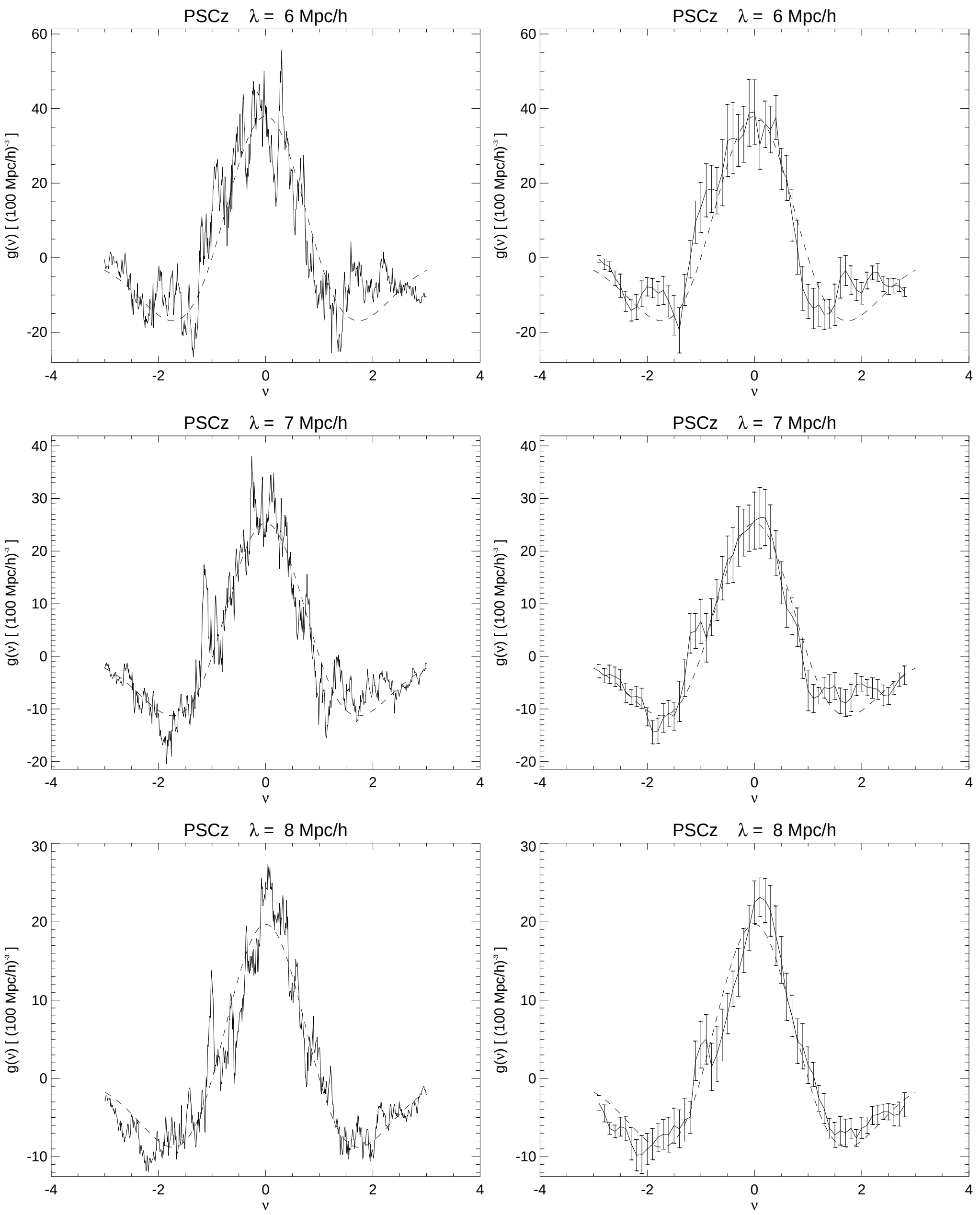
$$\xi(r) = \sum_{\mathbf{k}} \left\langle |\delta_{\mathbf{k}}|^2 \right\rangle \frac{\sin(kr)}{kr} = \int \mathrm{d}^3 k \, P(k) \frac{\sin(kr)}{kr}. \tag{A.44}$$

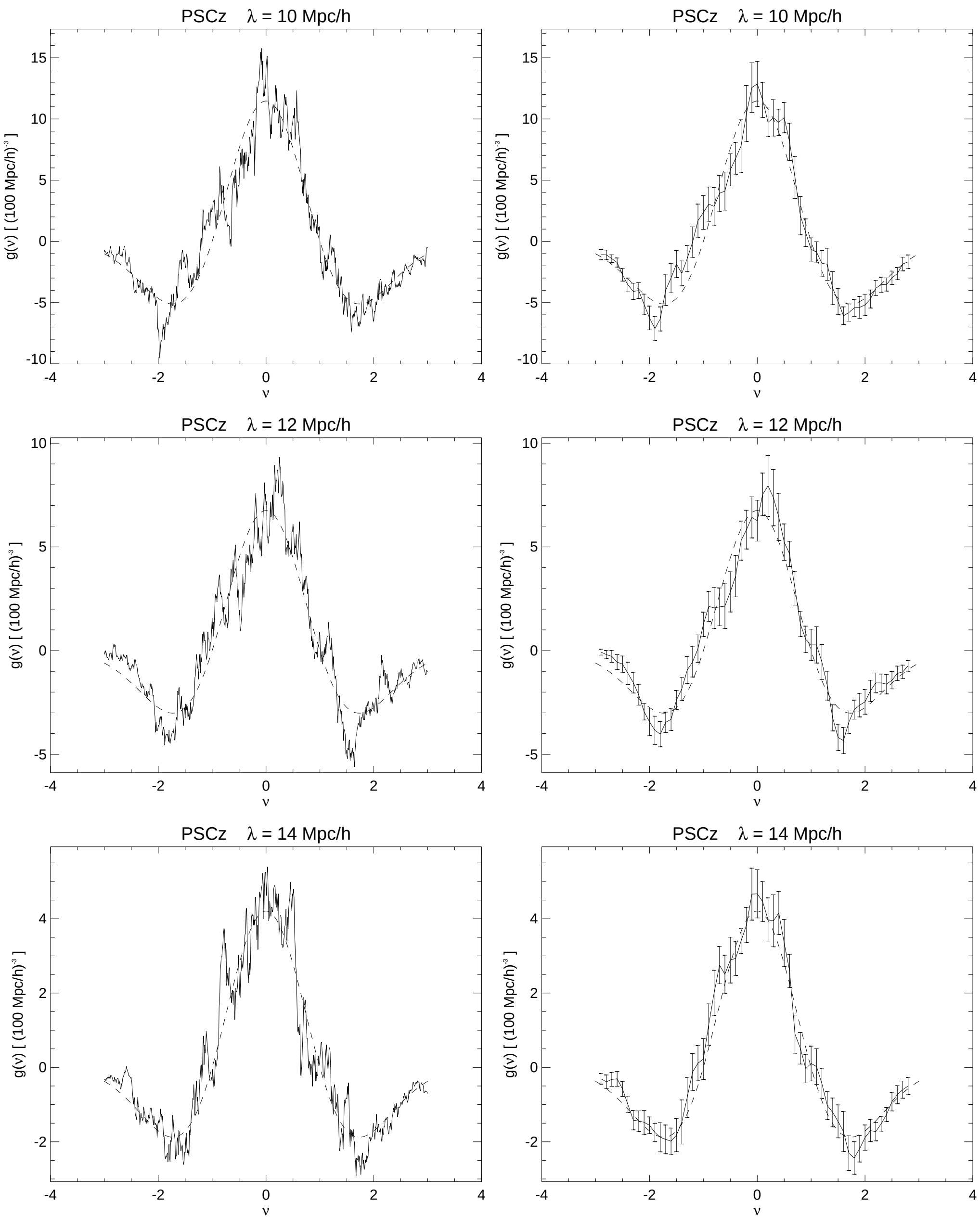
Appendix B

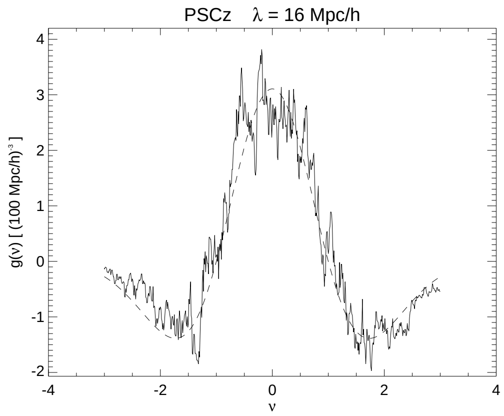
Genus Curves for the PSCz Survey

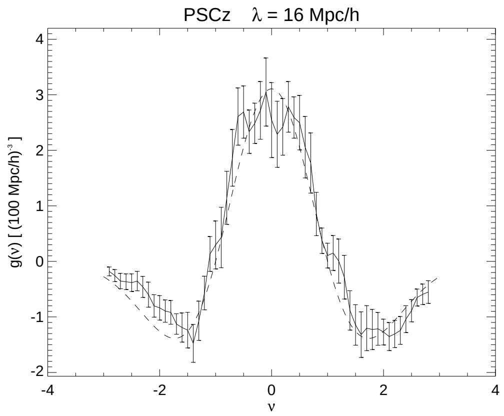
The subsequent figures show my results for the genus curves of the PSCz survey, calculated for 16 different smoothing lengths. For each smoothing length there are two panels. On the left hand side the high resolution raw genus curve is shown, on the right hand side the boxcar averaged curve together with error bars derived from bootstrap resamplings is displayed.

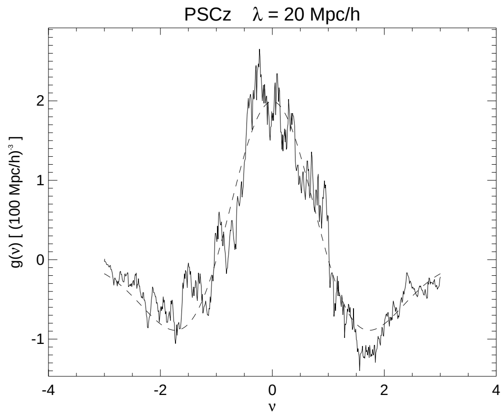


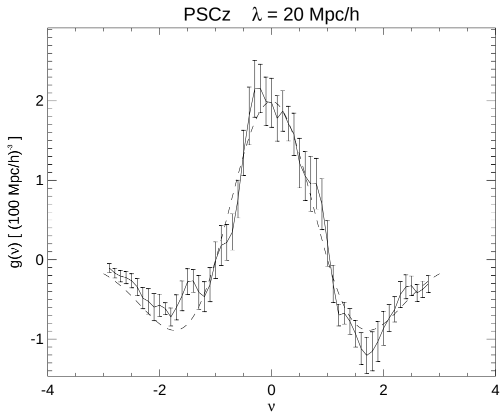

PSCz λ = 6 Mpc/h
PSCz λ = 6 Mpc/h
PSCz λ = 7 Mpc/h
PSCz λ = 7 Mpc/h
PSCz λ = 8 Mpc/h
PSCz λ = 8 Mpc/h
g(ν) [(100 Mpc/h)⁻³]
ν

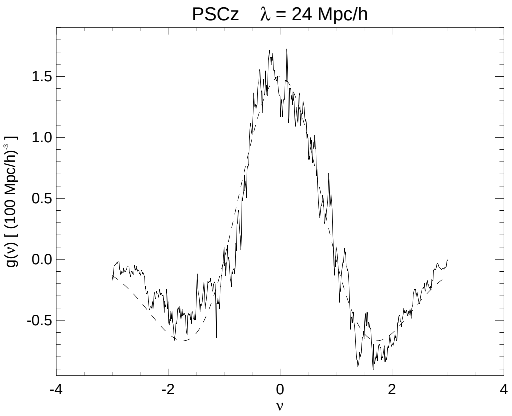

PSCz λ = 10 Mpc/h
PSCz λ = 12 Mpc/h
PSCz λ = 14 Mpc/h
g(ν) [(100 Mpc/h)^3]
ν

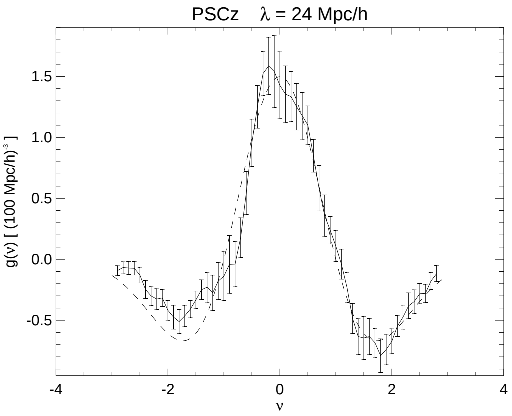

PSCz λ = 16 Mpc/h
g(ν) [(100 Mpc/h)⁻³]
ν

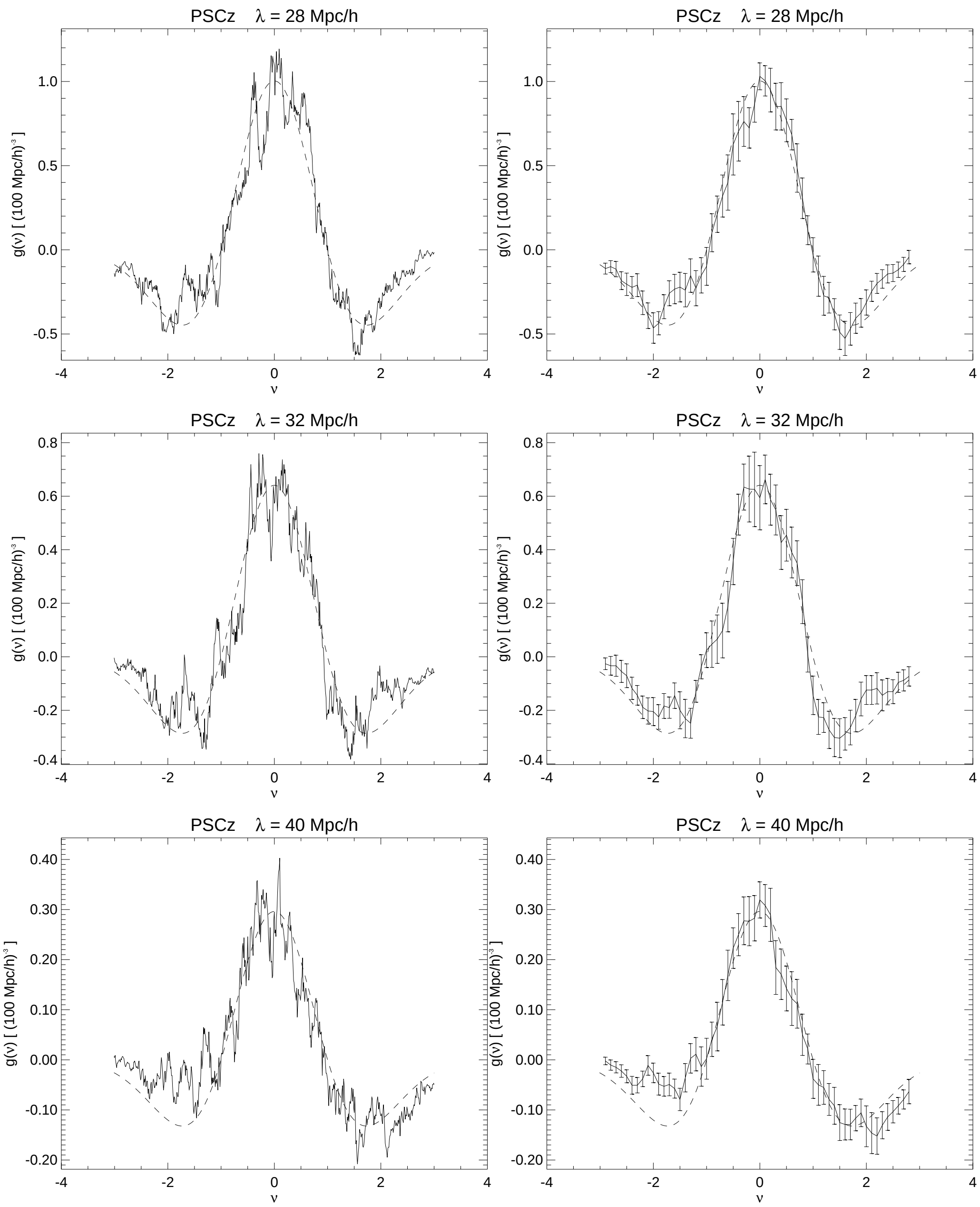

PSCz λ = 16 Mpc/h
g(ν) [(100 Mpc/h)⁻³]
ν

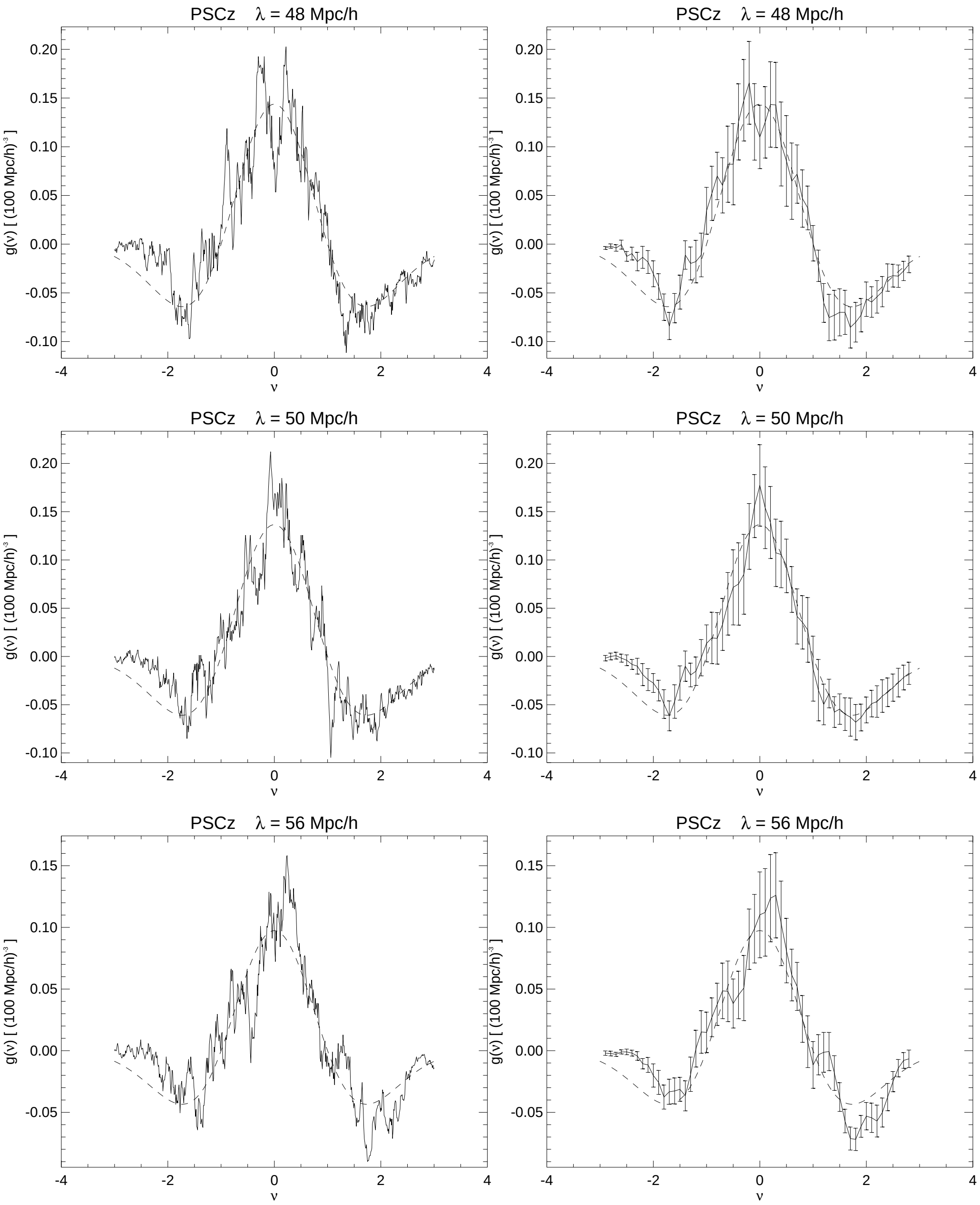

PSCz λ = 20 Mpc/h
g(ν) [(100 Mpc/h)⁻³]
ν


PSCz λ = 20 Mpc/h
g(ν) [(100 Mpc/h)⁻³]
ν


PSCz λ = 24 Mpc/h
g(ν) [(100 Mpc/h)⁻³]
ν


PSCz λ = 24 Mpc/h
g(ν) [(100 Mpc/h)⁻³]
ν


PSCz λ = 28 Mpc/h
PSCz λ = 28 Mpc/h
PSCz λ = 32 Mpc/h
PSCz λ = 32 Mpc/h
PSCz λ = 40 Mpc/h
PSCz λ = 40 Mpc/h
g(ν) [(100 Mpc/h)⁻³]
ν


PSCz λ = 48 Mpc/h
PSCz λ = 48 Mpc/h
PSCz λ = 50 Mpc/h
PSCz λ = 50 Mpc/h
PSCz λ = 56 Mpc/h
PSCz λ = 56 Mpc/h
g(ν) [(100 Mpc/h)$^{-3}$]
ν
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
-4
-2
0
2
4

Appendix C

Colour Plates

Subsequently I provide the figure captions for colour plates that had to be moved to seperate pages to allow a high quality reproduction.

Figure C.1: Isodensity contours of the smoothed mass density field derived from the PSCz survey for varying threshold values. The pictures are stills taken from a computer animation that shows the continuous change of the topology under a variation of the threshold value. From the top-left to the bottom-right panel the threshold varies from low (blue) to high (red) values. For thresholds below the mean density ($d < 0$), the low density region is enclosed, otherwise the high density region.

Figure C.2: Superclusters. The picture shows an isodensity contour of the PSCz density field smoothed at $\lambda = 10\,h^{-1}$Mpc in a sphere of radius $85\,h^{-1}$Mpc. The (x, y, z) axes point into the directions $(l, b) = (0, 0)$, $(l, b) = (90^0, 0)$ and $(l, b) = (0, 90^0)$. Hence the xy-plane is the galactic plane and the northern galactic hemisphere is the $z > 0$ half space. Our galaxy lies at the center of the cube which has a total size of $170\,h^{-1}$Mpc. The ten highest density peaks have been identified with known superclusters. Those patches of the contour that can be identified to them have been displayed with different colours. See table 3.1 for a list of the identifications A-J.

Figure C.3: Voids. This picture is the analogue to figure C.2, but here the ten highest underdense peaks are shown. See table 3.2 for a list of identifications of some of the voids. Note that void S is hidden behind the large void K.

Figure C.4: Some isodensity contours for the ΛCDM simulation smoothed with $\lambda = 5\,h^{-1}$Mpc. The upper two panels show the high (red) and low (blue) density regions that enclose a volume fraction 4.1%. These points correspond to the two minima at $\nu = \pm\sqrt{3}$ of the random phase genus curve. The two lower panels show the correponding surfaces for the Gaussianized version of the ΛCDM density field. Here the phase correlations have been wiped out and as a consequence the number of independent structures is higher.

References

Adler, R. J.: *The geometry of random fields.* John Wiley, Chichester, (1981).

Andreani, P. & Franceschini, A.: MM observations of IRAS galaxies. Preprint, astro-ph/9606002, (1996).

Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S.: The statistics of peaks of Gaussian random fields. *Astrophysical Journal*, **304**, 15, (1986).

Barrow, J. D., Bhavsar, S. P. & Sonoda, D. H.: A bootstrap resampling analysis of galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, **210**, 19P, (1984).

Beaky, M. M., Scherrer, R. J. & Villumsen, J. V.: Topology of large-scale structure in seeded hot dark matter models. *Astrophysical Journal*, **387**, 443, (1992).

Beichman, C. A. et al.: *Infrared astronomical satellite (IRAS) explanatory supplement.* IRAS Project Office, Pasadena, (1985).

Broadhurst, T. J., Ellis, R. S., Koo, D. C. & Szalay, A. S.: Large-scale distribution of galaxies at the galactic poles. *Nature*, **343**, 726, (1990).

Colberg, J.: The VIRGO Project: Cosmological N-body simulations. http://www.mpa-garching.mpg.de/~jgc/sim.html, (1996).

Couchman, H. M. P., Thomas, P. A. & Pearce, F. R.: Hydra: An adaptive-mesh implementation of P^3M-SPH. *Astrophysical Journal*, **452**, 797, (1995).

Davis, M. & Huchra, J.: A survey of galaxy redshifts. III. The density field and the induced gravity field. *Astrophysical Journal*, **254**, 437, (1982).

Davis, M. & Peebles, P. J. E.: A survey of galaxy redshifts. V. The two-point position and velocity correlations. *Astrophysical Journal*, **267**, 465, (1983).

de Lapperent, V., Geller, M. J. & Huchra, J. P.: A slice of the universe. *Astrophysical Journal*, **302**, L1, (1986).

de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H. G. & Buta, R. J.: *Third reference catalogue of bright galaxies*, volume I. Springer, New York, (1991).

Dekel, A. et al.: IRAS galaxies versus POTENT mass: Density field, biasing, and Ω. *Astrophysical Journal*, **412**, 1, (1993).

Doroshkevich, A. G.: Spatial structure of perturbations and origin of galactic rotation in fluctuation theory. *Astrophysics*, **6**, 320, (1970).

Efstathiou, G., Bond, J. R. & White, S. D. M.: COBE background radiation anisotropies and large-scale structure in the Universe. *Monthly Notices of the Royal Astronomical Society*, **258**, 1P, (1992).

Efstathiou, G., Ellis, R. S. & Peterson, B. A.: Analysis of a complete galaxy redshift survey - II. The field-galaxy luminosity function. *Monthly Notices of the Royal Astronomical Society*, **232**, 431, (1988).

Einasto, J., Klypin, A. A., Saar, E. & Shandarin, S. F.: Structure of superclusters and supercluster formation. III Quantitative study of the Ocal supercluster. *Monthly Notices of the Royal Astronomical Society*, **206**, 529, (1984).

Eke, V. R., Cole, S. & Frenk, C. S.: Using the evolution of clusters to contrain Ω. Preprint, astro-ph/9601088, (1996).

El-Ad, H., Piran, T. & da Costa, L. N.: A catalogue of the voids in the IRAS 1.2Jy survey. Preprint, astro-ph/9608022, (1996).

Feldman, H. A., Kaiser, N. & Peacock, J. A.: Power spectrum analysis of three-dimensional redshift surveys. *Astrophysical Journal*, **426**, 23, (1994).

Fisher, K. B. et al.: The density evolution of IRAS galaxies. *Astrophysical Journal*, **389**, 188, (1992).

Fisher, K. B. et al.: The IRAS 1.2Jy survey: Redshift data. *Astrophysical Journal*, **100**, 69, (1995).

Freedman, W. L. et al.: Distance to the Virgo Cluster galaxy M100 from Hubble Space Telescope observations of cepheids. *Nature*, **317**, 757, (1994).

Geller, M. J. & Huchra, J. P.: Mapping the Universe. *Science*, **246**, 897, (1989).

Gott, J. R. et al.: The topology of large-scale structure. III. Analysis of observations. *Astrophysical Journal*, **340**, 625, (1989).

Gott, J. R., Cen, R. & Ostriker, J. P.: Topology of large-scale structure by galaxy type: Hydrodynamic simulations. *Astrophysical Journal*, **465**, 499, (1996).

Gott, J. R., Melott, A. L. & Dickinson, M.: The sponge-like topology of large-scale structure in the Universe. *Astrophysical Journal*, **306**, 341, (1986).

Hacking, P. B., Condon, J. J. & Houck, J. R.: A very deep IRAS survey: Constraints on the evolution of starburst galaxies. *Astrophysical Journal*, **316**, L15, (1987).

Hacking, P. B. & Houck, J. R.: A very deep IRAS survey. *Astrophysical Journal Supplement Series*, **63**, 311, (1987).

Hacking, P. B. & Soifer, B. T.: The number counts and infrared backgrounds from infrared-bright galaxies. *Astrophysical Journal*, **367**, L49, (1991).

Hamilton, A. J. S.: Redshift distortions and omega in IRAS surveys. Invited review to appear in *Clustering in the Universe*, Proceedings XVth Rencontres de Moriond, Les Arcs, France, March, (1995).

Hamilton, A. J. S., Gott, J. R. & Weinberg, D.: The topology of the large-scale structure of the universe. *Astrophysical Journal*, **309**, 1, (1986).

Haynes, M. P. & Giovanelli, R.: The connection between Pisces-Perseus and the Local Supercluster. *Astrophysical Journal*, **306**, L55, (1986).

Kaiser, N., Efstathiou, G., Ellis, R. & Frenk, C.: The large-scale distribution of IRAS galaxies and the predicted peculiar velocity field. *Monthly Notices of the Royal Astronomical Society*, **252**, 1, (1991).

Kendall, M. G. & Stuart, A.: *The advanced theory of statistics*, volume 2. Charles Griffin, London, fourth edition, (1979).

Kerscher, M. et al.: Minkowski functionals of Abell/ACO clusters. Preprint, astro-ph/9606133, (1996).

Kirshner, R. P., Oemler, A., Schechter, P. L. & Shectman, S. A.: A million cubic megaparsec void in Boötes. *Astrophysical Journal*, **248**, L57, (1981).

Klypin, A. A.: Percolation as a descriptive statistics: Prospects and methods. *Acta Cosmologica*, **15**, 101, (1988).

Lawrence et al.: The QDOT all-sky IRAS galaxy redshift survey. In preparation, (1996).

Lawrence, A. et al.: Studies of IRAS sources at high galactic latitudes. *Monthly Notices of the Royal Astronomical Society*, **219**, 687, (1986).

Lin, H. et al.: The luminosity function of galaxies in the Las Campanas redshift survey. *Astrophysical Journal*, **464**, 60, (1996).

Ling, E. N., Frenk, C. S. & Barrow, J. D.: Uncertainties in the cluster-cluster correlation function. *Monthly Notices of the Royal Astronomical Society*, **223**, 21P, (1986).

Lonsdale, C. J. & Hacking, P. B.: Galaxy evolution and large-scale structure in the far-infrared. I. IRAS pointed observations. *Astrophysical Journal*, **339**, 712, (1989).

Lonsdale, C. J., Hacking, P. B. & Conrow, T. P.: Galaxy evolution and large-scale structure in the far-infrared. II. The IRAS Faint Source Survey. *Astrophysical Journal*, **358**, 60, (1990).

Loveday, J., Peterson, B. A., Efstathiou, G. & Maddox, S. J.: The Stromlo-APM redshift survey: The luminosity function and space density of galaxies. *Astrophysical Journal*, **390**, 338, (1992).

Low, F. J. et al.: Infrared cirrus: New components of the extended infrared emission. *Astrophysical Journal*, **278**, L19, (1984).

Mandelbrot, B. B.: *The fractal geometry of nature*. Freeman, New York, (1982).

Mann, R. G., Saunders, W. & Taylor, A. N.: The clustering of warm and cool IRAS galaxies. *Astrophysical Journal*, **279**, 636, (1996).

Matsubara, T.: Statistics of isodensity contours in redshift space. *Astrophysical Journal*, **457**, 13, (1996).

Matsubara, T. & Suto, Y.: Nonlinear evolution of genus in a primordial random Gaussian density field. *Astrophysical Journal*, **460**, 51, (1996).

Matsubara, T. & Yokoyama, J.: Genus statistics of the large-scale structure with non-Gaussian density fields. *Astrophysical Journal*, **463**, 409, (1996).

Mecke, K. R., Buchert, T. & Wagner, H.: Robust morphological measures for large-scale structure in the universe. *Astronomy and Astrophysics*, **288**, 697, (1994).

Melott, A. L. & Dominik, K. G.: Tests of smoothing methods for topological study of galaxy redshift surveys. *Astrophysical Journal Supplement Series*, **86**, 1, (1993).

Melott, A. L., Weinberg, D. H. & Gott, J. R.: The topology of large-scale structure. II. Nonlinear evolution of Gaussian models. *Astrophysical Journal*, **328**, 50, (1988).

Moore, B. et al.: The topology of the QDOT IRAS redshift survey. *Monthly Notices of the Royal Astronomical Society*, **256**, 477, (1992).

Moore, B., Frenk, C. S., Efstathoiu, G. & Saunders, W.: The clustering of IRAS galaxies. *Monthly Notices of the Royal Astronomical Society*, **269**, 742, (1994).

Nicholl, J. F. & Segal, I. E.: An optimal procedure for the non-parametric elimination of observational cutoff in complete samples. *Astronomy and Astrophysics*, **118**, 180, (1983).

Oliver, S. et al.: Evolution of faint IRAS galaxies. In *Wide Field Spectroscopy and the Distant Universe*, edited by S. J. Maddox & A. Aragon-Salamanca, 274, London, The 35th Herstmonceux Conference, World Scientific, (1994).

Ostriker, J. P. & Cowie, L. L.: Galaxy formation in an intergalactic medium dominated by explosions. *Astrophysical Journal*, **243**, L127, (1981).

Padmanabhan, T.: *Structure formation in the Universe*. Cambridge University Press, Cambridge, (1993).

Park, C., Gott, J. R. & da Costa, L. N.: Large-scale structure in the southern sky redshift survey. *Astrophysical Journal*, **392**, L51, (1992a).

Park, C., Gott, J. R., Melott, A. L. & Karachentsev, I. D.: The topology of large-scale structure. VI. Slices of the Universe. *Astrophysical Journal*, **387**, 1, (1992b).

Peacock, J. A., Heavens, A. F. & Davies, A. T., editors: *Physics of the early Universe*, Edinburgh, Edinburgh University Press, (1990).

Peebles, P. J. E.: *Principles of physical cosmology*. Princeton University Press, Princeton, (1993).

Rhoads, J. E., Gott, J. R. & Postman, M.: The genus curve of the Abell clusters. *Astrophysical Journal*, **421**, 1, (1994).

Rowan-Robinson, M. & Crawford, J.: Models for infrared emission from IRAS galaxies. *Monthly Notices of the Royal Astronomical Society*, **238**, 523, (1989).

Ryden, B. S.: The area of isodensity contours as a measure of large-scale structure. *Astrophysical Journal*, **333**, L41, (1988).

Sandage, A. J., Tammann, G. A. & Yahil, A.: The velocity field of bright nearby galaxies. I. The variation of mean absolute magnitude with redshift for galaxies in a magnitude limited sample. *Astrophysical Journal*, **232**, 352, (1979).

Saunders, W.: The IRAS PSC redshift survey. http://www-astro.physics.ox.ac.uk/~wjs/pscz.html, (1995).

Saunders, W.: Private communication, (1996).

Saunders, W. et al.: The 60 μm and far infrared luminosity functions of IRAS galaxies. *Monthly Notices of the Royal Astronomical Society*, **242**, 318, (1990).

Saunders, W. et al.: The density field of the local universe. *Nature*, **349**, 32, (1991).

Saunders, W., Sutherland, W., Efstathiou, G. & Tadros, H.: The point source catalog redshift survey. In *Wide field spectroscopy and the distant universe*, edited by S. J. Maddox & A. Aragon-Salamanca, 88, London, World Scientific, (1994).

Scalo, J.: *Evolution of galaxies*. Astronomical Institute, Prague, (1988).

Schmidt, M.: Space distribution and luminosity functions of quasi-stellar radio sources. *Astrophysical Journal*, **151**, 393, (1968).

Soneira, R. M. & Peebles, P. J. E.: Is there evidence for a spatially homogeneous population of field galaxies? *Astrophysical Journal*, **211**, 1, (1977).

Strauss, M. A. & Koranyi, D. M.: Tests of the Hubble law from the luminosity function of IRAS galaxies. Poster paper at the meeting 'Cosmic velocity fields', Paris, (1993).

Strauss, M. A. & Willick, J. A.: The density and peculiar velocity fields of nearby galaxies. *Physics Reports*, **261**, 271, (1995).

Tadros, H. & Efstathiou, G.: The power spectrum of IRAS galaxies. *Monthly Notices of the Royal Astronomical Society*, **276**, L45, (1995).

Vilenkin, A.: Cosmic strings and domain walls. *Physics Reports*, **121**, 263, (1985).

Vogeley, M. S. et al.: Topological analysis of the CfA redshift survey. *Astrophysical Journal*, **420**, 525, (1994).

Vogeley, M. S., Park, C., Geller, M. J. & Huchra, J. P.: Large-scale clustering of galaxies in the CfA redshift survey. *Astrophysical Journal*, **391**, L5, (1992).

Webster, M., Lahav, O. & Fisher, K.: Wiener reconstruction of the IRAS 1.2Jy galaxy redshift survey: Cosmograhical implications. Preprint, astro-ph/9608021, (1996).

Weinberg, D. H.: XIV. Contour: A topological analysis program. *Publications of the Astronomical Society of the Pacific*, **100**, 1373, (1988).

Weinberg, D. H., Gott, J. R. & Melott, A. L.: The topology of large-scale structure. I. Topology and the random phase hypothesis. *Astrophysical Journal*, **321**, 2, (1987).

Weinberg, S.: *Gravitation and cosmology.* Wiley, New York, (1972).

White, S. D. M., Efstathiou, G. & Frenk, C. S.: The amplitude of mass fluctuations in the Universe. *Monthly Notices of the Royal Astronomical Society*, **262**, 1023, (1993).

Yahil, A., Strauss, M. A., Davis, M. & Huchra, J. P.: A redshift survey of IRAS galaxies. II. Methods for detremining self-consistent velocity and density fields. *Astrophysical Journal*, **372**, 380, (1991).

Acknowledgments

I especially want to thank my advisor Professor Simon White from the Max-Planck-Institute for Astrophysics at Garching for the valuable advice, support and help I enjoyed over the period of this work and also for the suggestion of the topic of my diploma thesis.

I also want to thank my advisor Professor Hans Ruder from the University of Tübingen for proposing and supporting my tight cooperation with the Max-Planck-Institute for Astrophysics.

Many thanks also to numerous other people at the Max-Planck-Institute, most notably to Jörg Colberg and Jan Rehm, for constructive comments, helpful discussions, proof reading and many funny and relaxing conversations.

Special thanks also to Will Saunders from the Royal Observatory at Edinburgh who resolved my initial misunderstandings of the data format of the PSCz catalogue and pointed out a weakness of an early stage of my selection function analysis.

Lots of thanks also to the people of the PSCz and Virgo collaborations who generously granted me access to the data of the PSCz survey and to large N-body simulations before publication.

Finally I want to thank my parents for their love and unconditional support that I enjoyed all my life. Without them neither my study of physics nor this diploma thesis would have been possible.

Erklärung

Hiermit erkläre ich, daß ich diese Diplomarbeit selbständig angefertigt und keine anderen als die angegeben Quellen und Hilfsmittel benutzt habe.

(Volker Springel)

Tübingen, den 1. Oktober 1996