Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 8 October 1997 (MN LaTeX style file v1.4)

Genus Statistics of the Virgo N-body simulations and the 1.2-Jy Redshift Survey

Volker Springel[1], Simon D. M. White[1], Jörg M. Colberg[1],
Hugh M. P. Couchman[2], George P. Efstathiou[3], Carlos S. Frenk[4],
Adrian R. Jenkins[4], Frazer R. Pearce[4], Alistair H. Nelson[5],
John A. Peacock[6], and Peter A. Thomas[7] (The Virgo Consortium)
[1] *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Straße 1, 85740 Garching bei München, Germany*
[2] *Dept. of Physics & Astronomy, Univ. of Western Ontario, London, Ontario, N6A 3K7, Canada*
[3] *Dept. of Astrophysics, Nuclear & Astrophysics Laboratory, Keble Road, Oxford, OX1 3RH, UK*
[4] *Dept. of Physics, South Road, Durham, DH1 3LE, UK*
[5] *Dept. of Physics, College of Cardiff, P.O.Box 913, Cardiff, CF4 3TH, UK*
[6] *Royal Observatory Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK*
[7] *Astronomy Centre, University of Sussex, Falmer, Brighton, BN1, 9QH, UK*

8 October 1997

ABSTRACT

We study the topology of the Virgo N-body simulations and compare it to the 1.2-Jy redshift survey of *IRAS* galaxies by means of the genus statistic. Four high-resolution simulations of variants of the CDM cosmology are considered: a flat standard model (SCDM), a variant of it with more large-scale power (τCDM), and two low density universes, one open (OCDM, $\Omega_0 = 0.3$) and one flat (ΛCDM, $\Omega_0 = 0.3$, $\Lambda = 0.7$). In all cases, the initial fluctuation amplitudes are chosen so that the simulations approximately reproduce the observed abundance of rich clusters of galaxies at the present day. The fully sampled N-body simulations are examined down to strongly nonlinear scales, both with spatially fixed smoothing, and with an adaptive smoothing technique. While the τCDM, ΛCDM, and OCDM simulations have very similar genus statistics in the regime accessible to fixed smoothing, they can be separated with adaptive smoothing at small mass scales. In order to compare the N-body models with the 1.2-Jy survey, we extract large ensembles of mock catalogues from the simulations. These mock surveys are used to test for various systematic effects in the genus analysis and to establish the distribution of errors of the genus curve. We find that a simple multivariate analysis of the genus measurements is compromised both by non-Gaussian distributed errors and by noise that dominates the covariance matrix. We therefore introduce a principal components analysis of the genus curve. With a likelihood ratio test we find that the 1.2-Jy data favours the ΛCDM, τCDM and OCDM models compared to SCDM. When genus measurements for different smoothing scales are combined, the SCDM model can be excluded at a 99 per cent confidence level, while the other three models fit the 1.2-Jy data well. These results are unlikely to be significantly modified if galaxies are biased tracers of the mass, provided that biasing preserves a monotonic relation between galaxy density and mass density.

Key words: cosmology: large-scale structure of Universe – cosmology: observations – methods: statistical

1 INTRODUCTION

The observed large-scale structure of the Universe represents one of the most important constraints for theories of cosmic structure formation. In the past, the clustering of galaxies was mainly studied with statistics like the two-point correlation function, the power spectrum, or the counts-in-cells analysis. These statistical measures have been routinely applied to more and more powerful redshift surveys, leading to significant advances in our understanding of cosmic history.

Cosmological N-body simulations have played a vital part in this development, and they are already strongly constrained by the available data.

However, the 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. Furthermore, the low-order statistics are also quite insensitive to the geometrical aspects of the clustering which the human eye can so easily detect in pictures of the matter distribution. For example, it is still unclear whether the galaxy distribution is best described as filamentary, cellular or sheet-like.

In order to develop a measure of the geometrical and morphological aspects of the galaxy distribution a number of measures have been proposed, among them percolation (Klypin 1988), level-crossing statistics (Ryden 1988), genus statistics (Gott, Melott & Dickinson 1986), minimal spanning tree (Pearson & Coles 1995), shape statistic (Luo & Vishniac 1995), and Minkowski functionals (Mecke, Buchert & Wagner 1994).

In this paper we focus on the genus statistic which was first proposed by Gott, Melott & Dickinson (1986) and has become a widely accepted statistical tool in cosmology since then. The genus probes the topology of isodensity surfaces of a smoothed mass density field. It is therefore sensitive to global aspects of the density field that manifest themselves in higher order correlations, which reflect the connectedness and morphology of the structure in the Universe. Such features are not revealed by standard measures like the power spectrum or the two-point correlation function. For example, galaxy distributions that are wall-like, bubble-like or filamentary would all lead to different signatures in the genus statistic. Thus the topology has potentially strong discriminative power and might be used to rule out or support certain models for structure formation.

A particularly interesting application of the genus statistic is a test of the random phase hypothesis for the initial density fluctuation field. Because the topology is invariant during the linear growth of structure, the topology of the present galaxy distribution can be related to the topology of the initial density field, which in turn allows a test of the random phase hypothesis. Any departure 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.

In contrast to the linear regime, the topology on small, strongly non-linear scales has hardly been studied to date. In particular, there are no theoretical predictions for the genus statistics in this regime. Hence it is presently unclear whether the genus on small scales is a useful statistic to discriminate between different models of structure formation.

In this work we examine the topology of a suite of large, high-resolution N-body simulations carried out by the Virgo collaboration (Jenkins et al. 1997). The suite of models consists of a standard cold dark matter model (SCDM), and three variants of the cold dark matter cosmology, which have more power on large scales. These are a flat $\Omega_0 = 1$ model (τCDM), and two low density universes with $\Omega_0 = 0.3$, one open model (OCDM), and one closed by means of a cosmological constant (ΛCDM).

In the first of a series of papers, Jenkins et al. (1997) measured the correlation function, the power spectrum and various statistics of the velocity field of these simulations. As Jenkins et al. (1997) demonstrate, all three models with the power spectrum shape $\Gamma = 0.21$ can fit the APM two-point galaxy correlation function reasonably well, if one allows for a moderate scale-dependent bias. However, the differences *between* these models are rather small, at least when only the distribution at $z = 0$ is examined with these statistics.

In the second paper of the series, Thomas et al. (1997) showed that the internal structure of halos of rich clusters is also similar in all the models and would be difficult to distinguish in practice.

Here we use the genus statistic to analyse these N-body simulations down to strongly non-linear scales. In this regime, differences between the models can be detected.

In a second thread we compare the genus of the N-body models to the 1.2-Jy redshift survey of *IRAS* galaxies. Our particular aim is to improve the statistical methodology of such a comparison. For this purpose we work with ensembles of mock catalogues to assess the properties of new smoothing techniques and to derive accurate estimates for errors and systematic effects. This Monte-Carlo technique also allows the derivation of formal exclusion levels for the N-body models. New larger redshift surveys may be used subsequently to further tighten these constraints.

This paper is organized as follows. In section 2 we briefly review the genus statistic which we apply in section 3 to the fully sampled Virgo N-body simulations, both with fixed and adaptive smoothing. Section 4 describes the 1.2-Jy redshift survey and the construction of ensembles of mock catalogues. In section 5 we introduce different methods to compute smoothed density fields from the redshift survey data, and in section 6 we discuss various systematic effects that affect the genus statistic. We then turn in section 7 to the statistical methodology we adopt for the comparison with the 1.2-Jy redshift survey, and we present the results of this comparison in section 8. Finally we summarize and conclude in section 9.

2 GENUS STATISTIC

We first review briefly the genus statistic as introduced by Gott et al. (1986). A number of redshift surveys have been analysed with it, starting with a compilation of relatively small samples examined by Gott et al. (1989). Further surveys that have been studied include the SSRS (Park, Gott & da Costa 1992), QDOT (Moore et al. 1992), Abell Clusters (Rhoads et al. 1994), CfA (Vogeley et al. 1994), and most recently PSCz (Canavezes et al. 1997). There have also been a number of theoretical studies of the genus in the mildly non-linear regime (Matsubara 1994; Matsubara 1996; Matsubara & Yokoyama 1996; Matsubara & Suto 1996), and the genus has been applied in two dimensions to the microwave background (Colley, Gott & Park 1996) and to slice surveys (Park et al. 1992; Colley 1997).

Given an isodensity contour of a smoothed mass density field we define the genus as

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

where

$$\kappa = \frac{1}{r_1 r_2} \tag{2}$$

is the local Gaussian curvature. Here r_1 and r_2 denote the principal radii of curvature, and the integration extends over the whole surface. The Gauss-Bonnet theorem shows that this definition of genus makes G equal to the number of topological holes (like the one in a doughnut) minus the number of isolated regions of the surface.

Assuming an ergodic universe, we define a genus

$$g = \frac{G}{V} \tag{3}$$

per unit volume, where V is finite, but large enough to be a representative patch of the universe. The genus depends on the density threshold used to construct the isodensity surface. As a function of threshold we therefore obtain a *genus curve*, which is the central object of this investigation.

We parameterize the genus curve by the fraction f of the volume above the density threshold value or by the quantity

$$\nu = \sqrt{2}\,\mathrm{erf}^{-1}(1-2f) \tag{4}$$

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

The definition (4) is chosen such that for a Gaussian random field the quantity $\nu = \delta_\mathrm{t}/\sigma$ just measures the threshold value δ_t in units of the dispersion σ. However, we always define ν in terms of the volume fraction via equation (4), because this definition has the advantage of making the genus curve invariant under arbitrary monotonic one-to-one transformations of the density field. For example, a simple linear biasing transformation would not affect it. Also, it is insensitive to the skewness of the one-point probability distribution function, that quickly develops in the mildly non-linear regime.

There is a theoretical prediction for the expected genus curve of a Gaussian random field (Hamilton, Gott & Weinberg 1986). It is given by

$$g(\nu) = N\,(1-\nu^2)\exp\left(-\frac{\nu^2}{2}\right), \tag{5}$$

where the amplitude

$$N = \frac{1}{(2\pi)^2}\left(\frac{\langle k^2\rangle}{3}\right)^{\frac{3}{2}} \tag{6}$$

is determined by the second moment

$$\langle k^2\rangle = \frac{\int k^2 P(k)\mathrm{d}^3k}{\int P(k)\mathrm{d}^3k} \tag{7}$$

of the (smoothed) power spectrum. Interestingly, only the amplitude of the genus curve depends on the shape of the power spectrum. Apart from that it exhibits a universal, symmetric w-shape that we will use as a benchmark to detect non-Gaussian features of the density field.

For a given threshold value we compute the genus with the algorithm proposed by Gott et al. (1986). The method tesselates space in small cubes that allow the isodensity surface to be defined as the boundary between the volume elements above and below threshold. If the cubes are sufficiently small this approximation does not change the topology of the smooth isodensity surface. The curvature of the resulting polygonal surface is compressed into the vertices of the cubes. This property allows a computer to rapidly sum up the appropriate angle deficits and to compute the genus per unit volume. The method also allows arbitrarily shaped survey volumes. Here one just counts those vertices that are surrounded by eight volume elements that all lie inside the actual survey region.

Based on Weinberg's (1988) code CONTOUR for computing the genus we have written a new version in C that is optimized for a high execution speed, since we need to compute several thousand genus curves in this work. A simple sorting of the density field prior to the genus computation led already to a major speed-up because it is then possible to instantly find the threshold value corresponding to a desired volume fraction. This also allows the efficient computation of high resolution genus curves.

Table 1. Parameters of the examined CDM models. The simulations have been done by the Virgo collaboration.

	SCDM	τCDM	ΛCDM	OCDM
Number of particles	256^3	256^3	256^3	256^3
Box size[h^{-1}Mpc]	239.5	239.5	239.5	239.5
z_start	50	50	30	119
Ω_0	1.0	1.0	0.3	0.3
Ω_Λ	0.0	0.0	0.7	0.0
Hubble constant h	0.5	0.5	0.7	0.7
Γ	0.5	0.21	0.21	0.21
σ_8	0.60	0.60	0.90	0.85

3 FULLY SAMPLED VIRGO SIMULATIONS

3.1 The N-body models

In this section we compute genus curves for four fully sampled N-body simulations of variants of the CDM cosmology. These simulations, part of the Virgo project (Jenkins et al. 1997) to examine structure formation in the universe at very high resolution, follow 256^3 dark matter particles in boxes of comoving size of $239.5\,h^{-1}$Mpc. The simulations have been carried out with the parallel adaptive grid code HYDRA (Pearce & Couchman 1997).

The linear theory power spectrum used to generate the initial conditions of the models is parameterized by the generic fitting form

$$P(k) = \frac{Bk}{(1+[ak+(bk)^{3/2}+(ck)^2]^\nu)^{2/\nu}}, \tag{8}$$

where $a = 6.4\,\Gamma^{-1}\,h^{-1}$Mpc, $b = 3.0\,\Gamma^{-1}\,h^{-1}$Mpc, $c = 1.7\,\Gamma^{-1}\,h^{-1}$Mpc, $\nu = 1.13$, and Γ is a shape parameter (Efstathiou, Bond & White 1992).

While the SCDM model has the shape parameter $\Gamma = 0.5$, the other three models have the same linear power spectrum with $\Gamma = 0.21$. All four models are normalized so as to give the observed abundance of rich clusters of galaxies at the present day. Further simulation parameters are listed in table 1 and may be found in Jenkins et al. (1997).

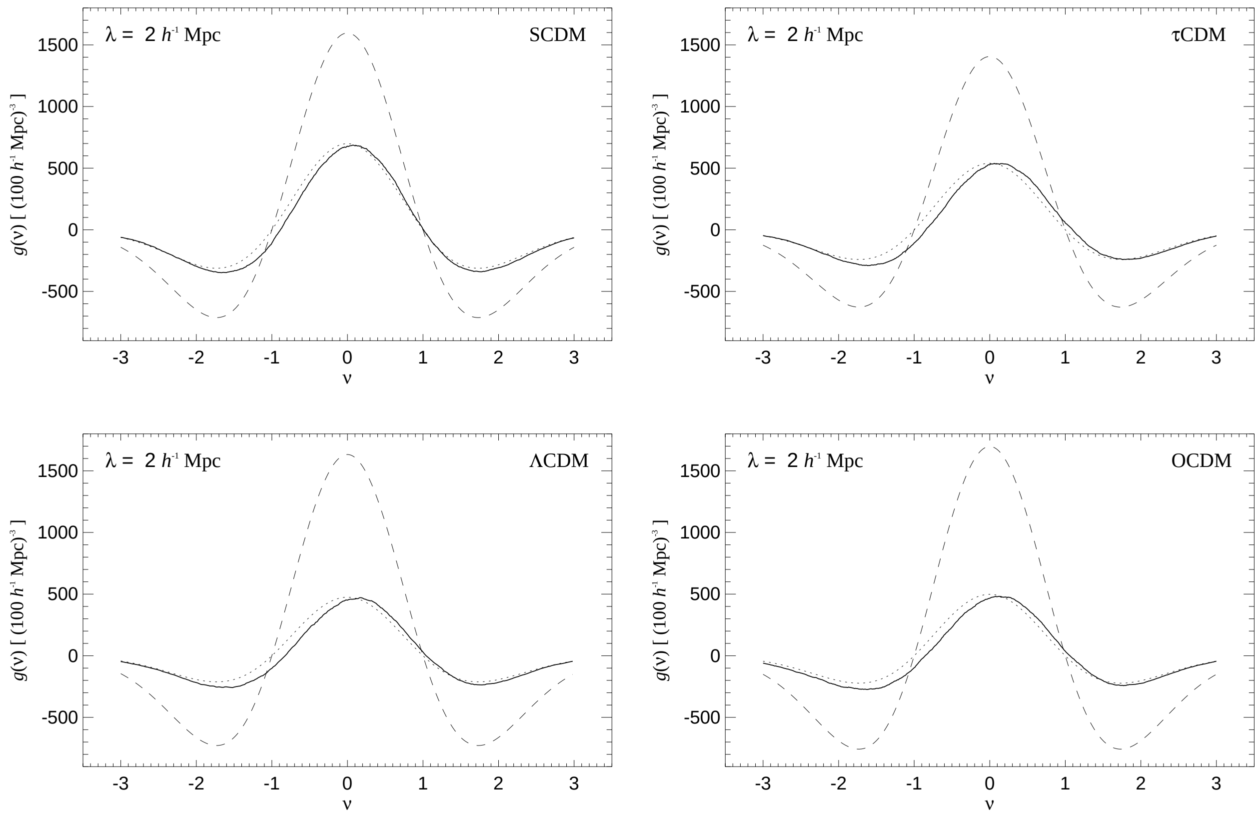

Figure 1. Genus curves of four variants of the CDM cosmology. The solid lines in each panel show the genus of the evolved density fields of the Virgo simulations for a smoothing scale of $2\,h^{-1}$Mpc. The dotted line is a fit to the random phase genus curve, while the dashed curve gives the genus of the corresponding *Gaussianized* field.

3.2 Fixed smoothing

We start with the ordinary genus statistic, i.e. we assume a spatially constant smoothing kernel. In order to construct a smooth density field we bin the 16.7 million particles of one simulation onto a mesh using the cloud-in-cell assignment, and we smooth the resulting density field with a Fast Fourier convolution. We use a Gaussian kernel of the form

$$W(\boldsymbol{x}) = \frac{1}{\pi^{3/2}\lambda^3} \exp\left(-\frac{\boldsymbol{x}^2}{\lambda^2}\right). \tag{9}$$

Note that this differs from an ordinary normal distribution by a factor of $\sqrt{2}$ in the definition of the smoothing scale. We stick to this convention which is used in the majority of the literature on the subject.

Typically we employ a 128^3 grid to represent the density field. Only for smoothing lengths below $5\,h^{-1}$Mpc we do find that a smaller mesh is indicated. We then use a 256^3 grid. For test purposes, we repeated one of our calculations on a 512^3 mesh.

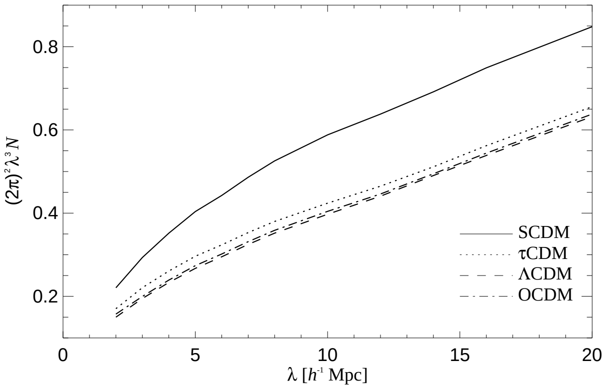

Figure 2. Genus amplitude of the Virgo simulations as a function of smoothing scale (fixed smoothing). The dimensionless vertical axis gives the genus amplitude N times the factor $(2\pi)^2\lambda^3$.

3.3 Results

In figure 1 we show the genus curves of the four simulations at a smoothing scale of $2\,h^{-1}$Mpc, the smallest scale considered here. We will come back to the question of what happens on still smaller ones later on. It is evident that the genus curves retain their universal w-shape, even at this small smoothing scale where the density fields are already fairly non-linear. Only a small *bubble* shift to the right has developed, which seems somewhat weaker for the SCDM model than for the other cosmologies.

We obtain qualitatively similar genus curves for larger smoothing scales. For $\lambda > 8\,h^{-1}$Mpc the small bubble shift

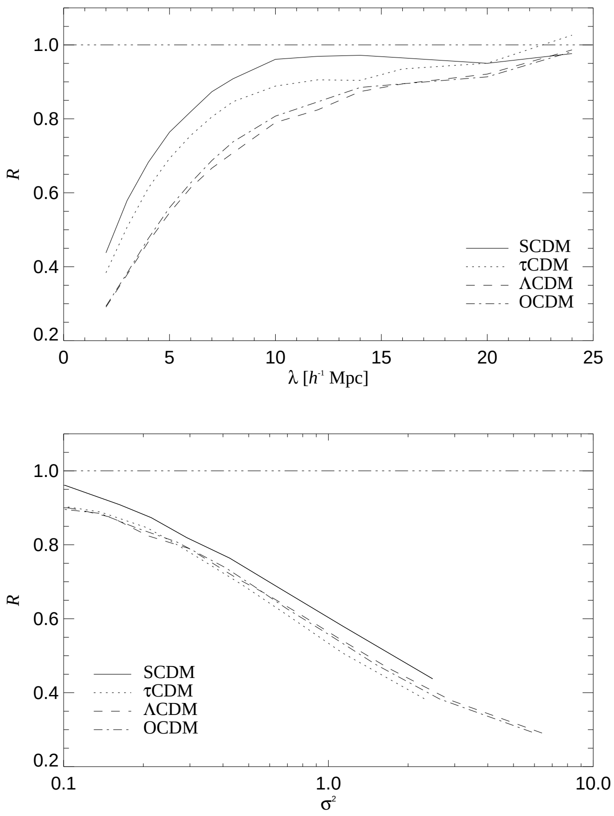

Figure 3. Amplitude drop for the Virgo simulations, i.e. the ratio R of the genus amplitude N to the corresponding amplitude of the Gaussianized density field. The top panel shows the amplitude drop versus the smoothing scale, while the lower panel displays it against the variance of the smoothed fields.

has vanished for all models and the shape of the genus curves is fit perfectly by the random phase form of equation (5). In figure 2 we show the measured genus amplitudes as a function of smoothing scale. The SCDM model exhibits a significantly higher amplitude, reflecting the different shape of its power spectrum, while the other three models show very similar amplitudes. This demonstrates that the genus amplitude in the linear and mildly non-linear regime is determined by the shape of the power spectrum alone.

Even if the genus curve is well described by equation (5), the underlying density field does not have to be Gaussian. In fact, Vogeley et al. (1994) and Canavezes et al. (1997) pointed out that the amplitude of the genus curve is suppressed on small scales compared to the expected amplitude based on the power spectrum alone. This amplitude drop is a direct consequence of phase correlations that develop during the nonlinear growth of density perturbations. The phase drop may be measured for a periodic N-body simulation by *Gaussianizing* the evolved density field, i.e. by taking it to Fourier space, randomizing the phases of all modes constrained by the reality condition $\delta_{\boldsymbol{k}} = -\delta^{\star}_{\boldsymbol{k}}$, and transforming back to real space. In this way a Gaussian field with identical power spectrum as the evolved density field is obtained, and a measurement of its genus allows an estimate of the amplitude drop.

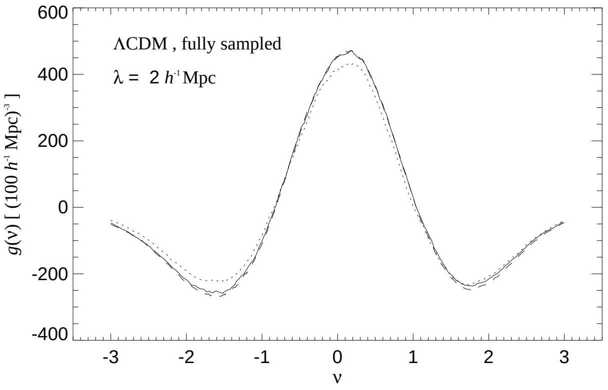

Figure 4. Finite grid size effect. The lines are genus curves for the ΛCDM simulation, smoothed with $\lambda = 2\,h^{-1}$Mpc. The solid line is computed with a 512^3 grid, the dashed line with 256^3, and the dotted line with 128^3.

We have measured the genus amplitude drop in this way and show the results in figure 3. Here we reach smoothing scales as small as $2\,h^{-1}$Mpc, thereby extending the work of Canavezes et al. (1997). Note that the amplitude drop becomes very substantial at small scales, showing that the genus is indeed quite sensitive to higher order correlations in this regime. If we plot the amplitude drop against the variance of the smoothed density fields, we can approximately take out the differences between the models due to their slightly different normalizations. As the lower panel in figure 3 shows, we again find that the models with the same shape of the power spectrum exhibit a nearly degenerate behaviour. However, the horizontal axis of this plot is affected by biasing. Since the bias required to match the observed two-point correlation function (Jenkins et al. 1997) is different for the four models, the amplitude drop may still be a useful measure to discriminate between different CDM variants. Future galaxy redshift surveys should allow an accurate measurement of the amplitude drop by combining genus statistics with an independent measure of the power spectrum or clustering strength as outlined by Canavezes et al. (1997). Comparing these measurements with the results obtained in figure 3 for the dark matter may then allow a direct measure of the bias $b^2 = \sigma^2_{\rm gal}/\sigma^2_{\rm DM}$.

We now examine systematic limitations of the genus analysis performed above. We will consider finite grid size effects first and then examine the resolution limit of the Virgo simulations in terms of the genus statistic.

3.3.1 Finite grid size

Our method for computing the genus curve relies on an approximate representation of isodensity contours as polygonal surfaces that are made up of faces of small cubes used to tesselate space. As Hamilton et al. (1986) show the genus curve is expected to be unaffected by this approximation, if $d/\lambda \ll 1$, where d is the size of the cubes. Typically we use 128^3-grids, and for $\lambda \leq 5\,h^{-1}$Mpc 256^3 grids. We find that for $d \leq 0.5\lambda$ there is hardly any finite grid size effect. For example, in figure 4 we show a comparison of the genus curves for the ΛCDM simulation, smoothed at $\lambda = 2\,h^{-1}$Mpc, using

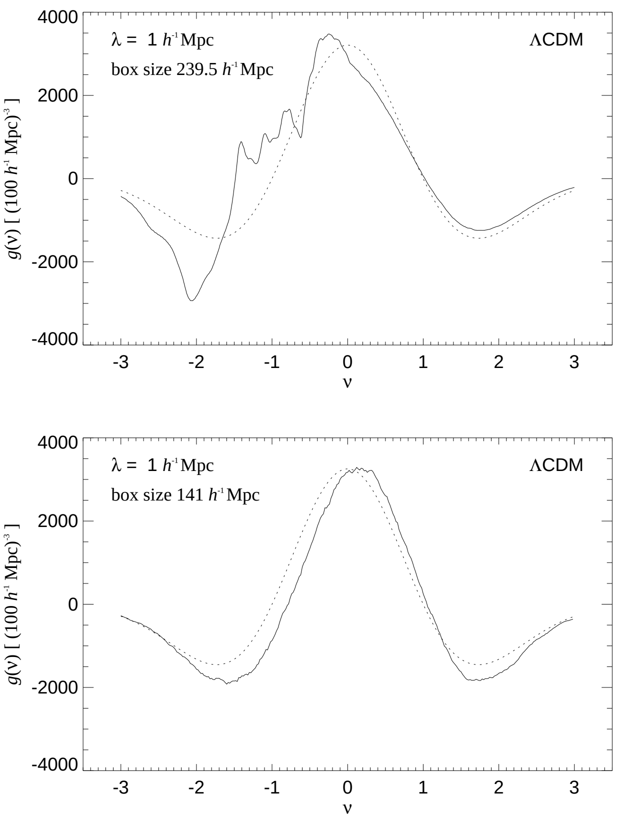

Figure 5. Resolution limit of the Virgo simulations. The solid line of the top panel shows the genus of the ΛCDM model smoothed with $\lambda = 1\,h^{-1}$Mpc, and the dotted curve shows a best-fit random phase genus curve. In this case the smoothing reaches the level of the inter-particle separation and the features in the genus curve are in fact artifacts. We show that this is the case by computing the genus for a second ΛCDM simulation with smaller box size of $141\,h^{-1}$Mpc, but with the same number of particles and hence significantly higher spatial resolution. As seen in the bottom panel, the genus curve remains featureless.

128^3, 256^3, and 512^3 grids. There is only a small depression of the amplitude and the minima of the genus curve, when the 128^3 mesh is used, but for 256^3 the asymptotic behaviour is clearly reached. It is therefore not necessary to use costly computations at 512^3 resolution.

3.3.2 Resolution limit

If the smoothing length is reduced below $2\,h^{-1}$Mpc one suddenly starts to see features in the genus curve, as exemplified in figure 5, where we show the genus for the ΛCDM simulation at a smoothing scale of $1\,h^{-1}$Mpc. However, the 'ringing' on the low density side is just an artifact due to the fact that the resolution of the Virgo simulations is limited. We demonstrate that this is indeed the case by analysing yet another simulation of the Virgo consortium, the same ΛCDM model, but with a smaller box size of $141\,h^{-1}$Mpc. This model has roughly twice the spatial resolution of the original simulation. As the bottom panel of figure 5 demonstrates, it still gives a 'normal' genus curve for $\lambda = 1\,h^{-1}$Mpc.

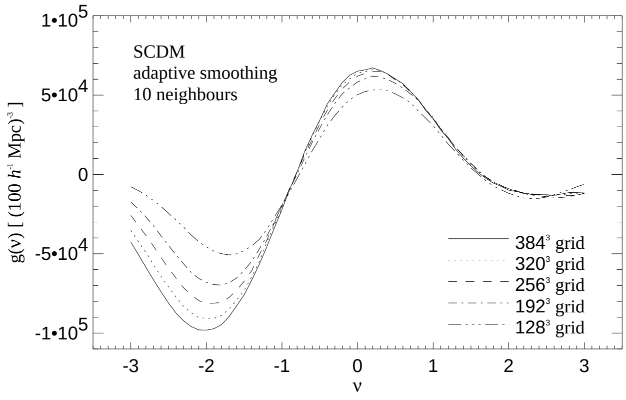

Figure 6. Finite grid size effects in the adaptive smoothing technique. Shown are five genus curves for one of the SCDM subvolumes, computed at different grid resolutions. For 10 neighbours, the number considered here, a 256^3 mesh is sufficient to resolve accurately the genus of the density field, at least in the regions of positive genus and for $\nu > 0$. However, the minimum on the negative side is still not fully resolved by a grid as fine as 384^3. Comparing the size distribution of the isolated regions at the two minima of the genus curve, we find that this is due to a larger relative abundance of very small regions at the minimum on the negative side. This population of very small structures is difficult to resolve; those regions with volume smaller than a mesh cell can be lost, leading to a suppression of the genus amplitude.

The smoothing scale of $1\,h^{-1}$Mpc is already close to the mean inter-particle separation of $\approx 0.94\,h^{-1}$Mpc for the large-box Virgo runs. Due to the strong clustering of matter, the voids apparently contain too few particles to prevent discreteness effects becoming visible when a fixed smoothing kernel with $\lambda = 1\,h^{-1}$Mpc is used. On the other hand, the regions with high particle density allow a much higher resolution in principle. In order to take full advantage of this spatially varying resolution an adaptive smoothing technique needs to be developed. We will introduce such a scheme in the next section.

3.4 Adaptive Smoothing

As we saw above, a spatially fixed smoothing is not able to take full advantage of the information content on small scales. Borrowing an idea from smoothed particle hydrodynamics (SPH), we can improve on this by following the particles in a Lagrangian sense and varying the smoothing scale with local density. This application of SPH-smoothing to the dark matter has first been proposed by Thomas & Couchman (1992).

Of course, a difficulty with such an approach is that the resulting density field cannot be studied analytically. In particular, the power spectrum of the smoothed field is not related to that of the underlying field in a simple way. Because of that, it is also not obvious what one can expect for the genus statistic. However, a comparison between theory and observation is always possible if ensembles of mock surveys are used to calibrate the adaptively smoothed genus statistic.

As Hernquist & Katz (1989) point out there are two different approaches to defining a smoothed density estimate

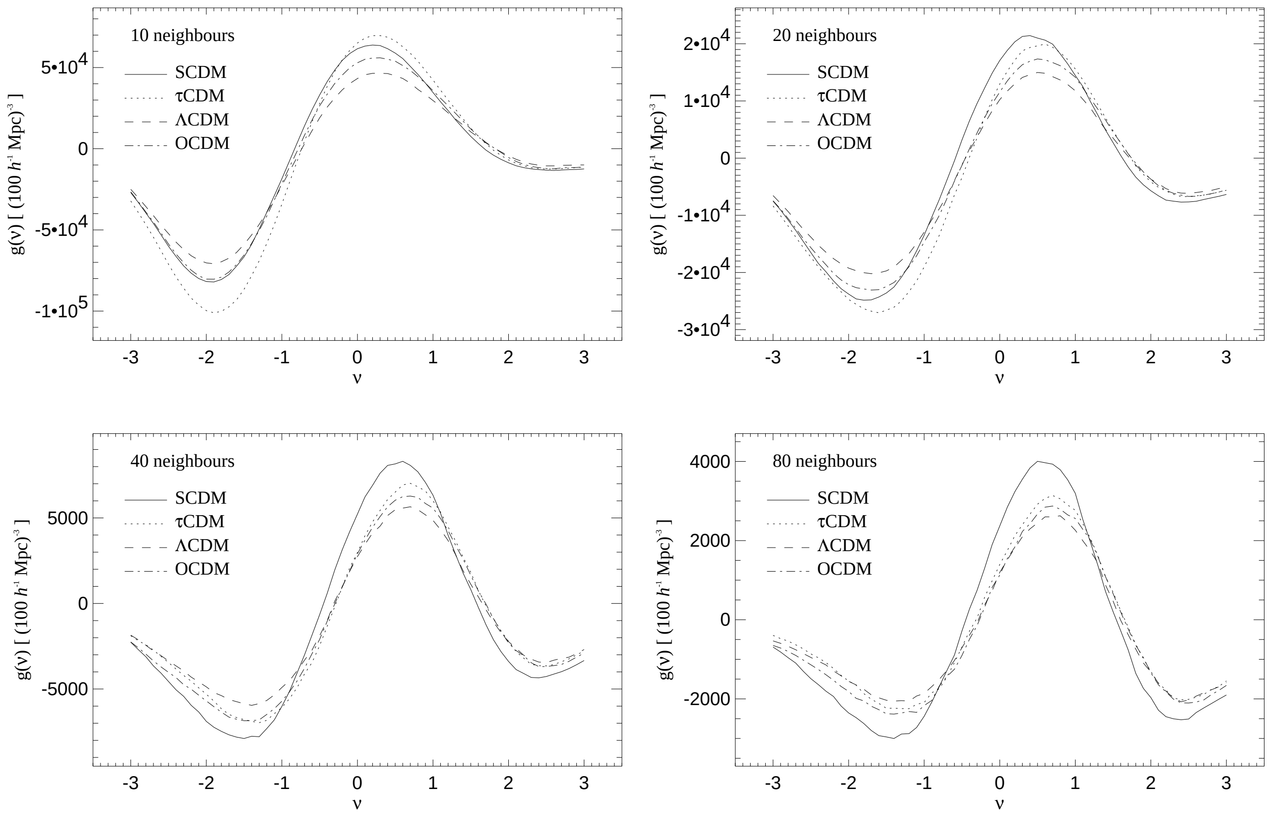

Figure 7. Genus curves for the adaptively smoothed Virgo simulations. The four panels show results when 10, 20, 40, and 80 neighbours are used to determine the local smoothing scale. Each curve is the average of 4 subvolumes, each being 64 times smaller than the full simulation volume. Since the variance between these curves is quite small, it is not necessary to extend the computation to a larger fraction of the total volume.

with variable smoothing scale. In the *scatter* approach the mass of each particle is distributed in space, and the density estimate at a particular point $\boldsymbol{x}$ follows from the overlap of the individual smoothing spheres. Alternatively, one can define a smoothing radius for each point $\boldsymbol{x}$ and weight all particles in its neighbourhood by the resulting kernel. Only for a fixed smoothing scale does this *gather* approach become identical to the scatter method.

We want to compute the density field on a fine mesh at a large number of points. Because the number of tracer particles is also large, the gather approach is computationally less costly in this case. Hence we will adopt it in the following.

As a smoothing kernel we choose the spherically symmetric spline kernel

$$W(r;h) = \frac{8}{\pi h^3} \begin{cases} 1 - 6\left(\frac{r}{h}\right)^2 + 6\left(\frac{r}{h}\right)^3, & 0 \le \frac{r}{h} \le \frac{1}{2}, \\ 2\left(1-\frac{r}{h}\right)^3, & \frac{1}{2} < \frac{r}{h} \le 1, \\ 0, & \frac{r}{h} > 1, \end{cases} \tag{10}$$

which is frequently used in SPH calculations (Monaghan & Lattanzio 1985). This kernel has the advantage of compact support, which again simplifies the computational task.

For simplicity, we choose the smoothing scale $h(\boldsymbol{x})$ as the distance to the N-th nearest neighbour. In this way the smoothing scale is set to a fixed fraction of an estimate of the local mean inter-particle separation. We then estimate the density field as

$$\rho(\boldsymbol{x}) = \sum_i W(\boldsymbol{r}_i - \boldsymbol{x}; h(\boldsymbol{x})), \tag{11}$$

where the sum extends over all particles. Due to the compact kernel only the nearest N particles give a nonzero contribution. Hence the smoothing scheme heavily relies on a fast algorithm for finding the nearest neighbours for arbitrary points in space. We utilize a coarse search grid and link-lists to match this requirement.

Nevertheless, the computational cost for adaptive smoothing is much higher than for fixed smoothing, because Fourier techniques can no longer be applied and one has to work in real space. Furthermore, the adaptive smoothing will also be able to resolve very small structures that demand a fine mesh in order to allow a proper resolution of their topology.

Because of this it is not possible to consider the whole simulation box in one pass. Instead, we compute adaptively smoothed density fields for subvolumes of the simulation box of size one quarter of the total box size. In this way, we divide the simulation in 64 subvolumes, each containing still 64^3 particles on average. For the small scales we are considering here, we anticipate that each of these subvolumes will already contain so many structural elements that the cosmic

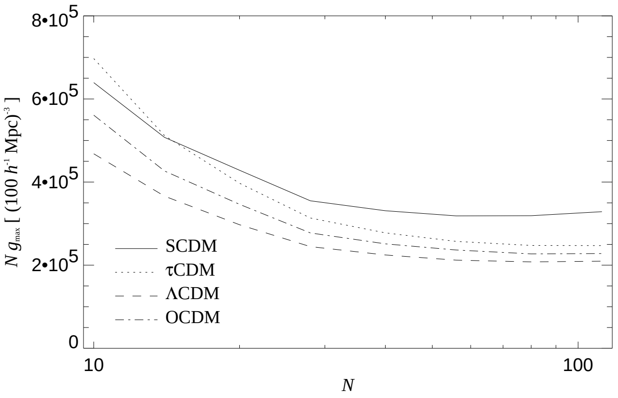

Figure 8. Peak genus densities in the Virgo models as a function of the number N of neighbours used in the adaptive smoothing scheme. We have measured the genus for $N = 10$, 14, 20, 28, 40, 56, 80, and 112 neighbours, and plot the maximum genus density times the number N of smoothing neighbours.

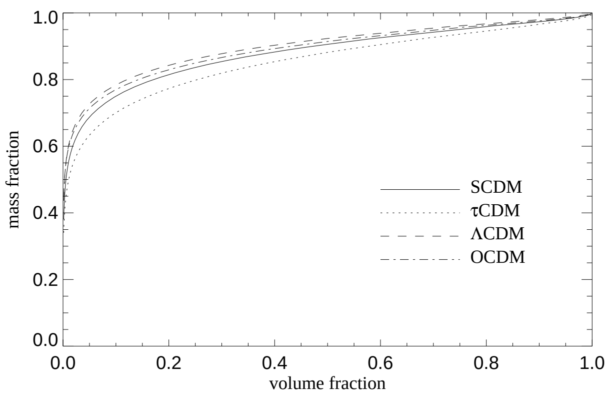

Figure 9. Mass fraction above threshold versus volume fraction above threshold for the adaptively smoothed Virgo simulations. Here $N = 10$ neighbours have been used in the smoothing process. At a given volume fraction, the ΛCDM model contains a higher fraction of its mass in the high density regions than the other three models.

variance between them will be small. Hence we assume that we will have to consider at most a few of the subvolumes in order to accurately reproduce the genus of the whole simulation box.

We compute adaptively smoothed density fields on a regular grid for each of the subvolumes. In the smoothing process, we also consider the particles that lie outside the subvolume, hence boundary smoothing effects are not present. In figure 6 we show the dependence of the result on the chosen grid resolution for the extreme case of $N = 10$ neighbours. A 128^3 grid clearly leads to a systematic underestimate of the genus, showing that small features are missed due to the coarseness of the grid. On the other hand, the result for the 256^3 grid is already very close to the finer meshes, at least for $\nu \geq -1$. However, the minimum of the genus on the negative side is seen to be very difficult to resolve accurately. By comparing the distributions of the sizes of the isolated regions at the two minima, we find that the number density of very small regions is much larger at the minimum on the negative side than on the positive side. Since these very small voids are so abundant, the loss of the smallest of them due to the finite grid size leads to a suppression of the genus. Fortunately this effect is less severe at larger smoothing scales. As a compromise between computational cost and accuracy, we decided to use 256^3 grids for $N = 10$ neighbours, 192^3 for $14 \leq N \leq 20$, and 128^3 for $N \geq 28$. This ensures that the genus density is accurately resolved for $\nu \geq -1$. However, the depth of the minimum on the negative side may be slightly diminished due to finite grid size effects.

In figure 7 we show genus curves for adaptive smoothing with $N = 10$, 20, 40, and 80 neighbours. In each panel we plot the results for SCDM, τCDM, ΛCDM, and OCDM. Interestingly, there are marked differences between the models, particularly at small smoothing scales. The SCDM model shows usually the highest genus amplitude, as expected for the larger amount of small-scale power in this model. Only on the smallest scale considered here, it is surpassed by the τCDM model. While the genus amplitudes of the three models τCDM, ΛCDM, and OCDM were practically degenerate in the regime accessible to fixed smoothing, they can now be used to discriminate between the models.

Note that the number density of structure elements resolved with adaptive smoothing is really much larger than the one accessible to fixed smoothing. Compared to the $\lambda = 2\,h^{-1}$Mpc fixed smoothing, the $N = 10$ adaptive scheme reaches a density of structure elements which is approximately two orders of magnitude larger. This really opens up a new regime for topological analysis.

In figure 8 we plot the peak genus densities as a function of the number of neighbour particles used for the adaptive smoothing. It is interesting that the relative difference between τCDM, ΛCDM, and OCDM grows with decreasing smoothing length. ΛCDM consistently shows the smallest amplitude, which demonstrates that its structure has a higher degree of coherence with fewer small-scale features than the other models.

However, for the $N = 10$ smoothing, all four simulations give roughly the same genus signal at the minimum on the positive side of the genus curves. This minimum occurs at $\nu \simeq 2.4$, corresponding to a volume fraction above threshold of only 0.0082. For a Gaussian random field, the minimum occurs at $\nu = \sqrt{3}$, equivalent to the much larger volume fraction of 0.042. At the minimum the genus is completely dominated by the number density of very dense, isolated clumps. While this number density of roughly $0.012\,h^3\mathrm{Mpc}^{-3}$ is similar for the four models, the mass fraction contained in the clumps is different. For the $N = 10$ smoothing, we plot in figure 9 the mass fraction above threshold as a function of the volume fraction above threshold. Note the very high concentration of the mass in these adaptively smoothed density fields. In the SCDM model 50 per cent of the mass is contained in only 0.65 per cent of the volume. The τCDM model requires almost twice the volume fraction, 1.23 per cent, to include half of its mass. On the other hand, the ΛCDM and OCDM simulations have half-mass volume fractions of 0.33 and 0.31 per cent, respectively. At high densities a given volume fraction therefore contains a higher fraction of their mass than in the high density models. This reflects the normalization of the models which ensures that they all have the same abundance of virialized clusters of any given mass.

In contrast to the minimum on the positive side, the minimum on the negative side of the genus curve is not shifted significantly. However, here the models show stronger differences in their genus signal. Again, at the minimum the genus is dominated by a high number density of isolated regions. These voids are significantly more abundant in the τCDM model than in the ΛCDM model. This means that the underdense regions are choppier in the τCDM model than in ΛCDM. The latter model has voids which are more coherent and larger on the average. Presently it is unclear, whether this can be understood merely as a consequence of the smaller amount of mass left in the ΛCDM model to fill the voids.

The maxima of the genus densities occur at a smaller volume fraction than 0.5, i.e. the adaptive smoothing results in a substantial bubble-shift to the right. At the maxima, the isodensity surfaces have the topology of a sponge, with an interlocking high density region and a complex system of tunnels and voids. The smaller amplitude of ΛCDM can again be interpreted as a larger degree of coherence; there are not so many topological holes as in τCDM, for example, and the typical size of tunnels and cavities is expected to be larger.

It should be noted that the two low-Ω_0 models show only small differences when their two-point correlation functions or their velocity fields are considered (Jenkins et al. 1997). This demonstrates that the genus can indeed reveal additional information about the morphological properties of the matter distribution. Here we conclude that the voids in the ΛCDM model are in some sense 'emptier' than in the OCDM model.

We also note that only the adaptive smoothing allowed an extension of the genus statistics essentially down to the mass resolution of the Virgo simulations. Future large redshift surveys feature a large number of galaxies, possibly in the range 10^6 for the Sloan and 2dF surveys. Adaptive smoothing techniques should prove very powerful for this kind of data. We now work out a first test of this idea for the 1.2-Jy redshift survey.

4 THE 1.2-JY REDSHIFT SURVEY AND VIRGO MOCK CATALOGUES

4.1 The 1.2-Jy redshift survey data

The data of the 1.2-Jy redshift survey (Strauss et al. 1990) of *IRAS* galaxies has been published (Fisher et al. 1995) and can be retrieved electronically from the Astronomical Data Center (ftp://adc.gsfc.nasa.gov). The 5321 galaxies of the survey are selected from the PSC catalogue above a flux limit of 1.2 Jy in the 60 μm band. The sky coverage is 87.6 per cent, excluding only the zone of avoidance for $|b| < 5^\circ$ and a few unobserved or contaminated patches at higher latitude.

There have been numerous studies of the 1.2-Jy survey, essentially covering all standard methods for analysing large-scale structure. Very recently the catalogue has also been examined with topological methods. Yess, Shandarin & Fisher (1997) used a percolation analysis, Kersher et al. (1997) employed Minkowski functionals, and Protogeros & Weinberg (1997) used genus statistic.

Table 2. Parameters of the selection function of the 1.2-Jy survey.

α	β	γ
$0.741^{+0.128}_{-0.135}$	$4.210^{+0.419}_{-0.344}$	$1.582^{+0.237}_{-0.214}$
$z^\star$	ψ $[h^3\mathrm{Mpc}^{-3}]$	
$0.0184^{+0.00213}_{-0.00167}$	$(486.5 \pm 13.0) \times 10^{-6}$	

Our approach is similar to that of the latter authors. We also work with mock catalogues to derive the statistical properties of the genus statistics. However, in contrast to Protogeros & Weinberg (1997) we do not volume-limit our catalogues but use instead a selection function weighting to derive the density fields. Additionally, we introduce adaptive smoothing techniques. We also differ in some conclusions; e.g. we find no confirmation of their claim of a finite volume bias, and we explicitly show that the errors of the genus curve are not multivariate normally distributed.

In our analysis of the 1.2-Jy survey we convert the redshifts to the Local Group frame and use them to infer distances without further corrections for peculiar velocities. Redshift space distortions have only a negligible effect on the genus, as has been shown in a number of studies.

We will assume an Einstein-de-Sitter model for the background cosmology throughout. The results will not be sensitive to this choice because the 1.2-Jy density field maps only the very local Universe.

We define the selection function $S(z) = \langle m(\boldsymbol{r})\rangle$ of the survey as the mean expected comoving number density of sources at redshift $z = |\boldsymbol{r}|$. We employ the fitting form

$$S(z) = \frac{\psi}{z^\alpha \left[1 + \left(\frac{z}{z^\star}\right)^\gamma\right]^{\frac{\beta}{\gamma}}}, \tag{12}$$

and adopt the parameters (table 2) determined by Springel & White (1997). Note that the selection function includes a correction for the strong evolution seen in *IRAS* galaxies.

4.1.1 *Depth of maps*

The galaxy density of a flux limited sample falls off quickly with distance. As a consequence, the uncertainty in the density estimate grows rapidly with redshift. It is desirable, of course, to use a survey volume that is as large as possible in order to beat down statistical noise and cosmic variance. According to a useful rule of thumb (Weinberg, Gott & Melott 1987) discreteness effects are approximately negligible if

$$\lambda \geq d = S^{-\frac{1}{3}}, \tag{13}$$

where d is the mean inter-galaxy separation. Adopting this criterion we choose a maximal radius $R_{\max}$ by $\lambda = S(R_{\max})^{-\frac{1}{3}}$ and use it to delimit the usable survey volume V_{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.

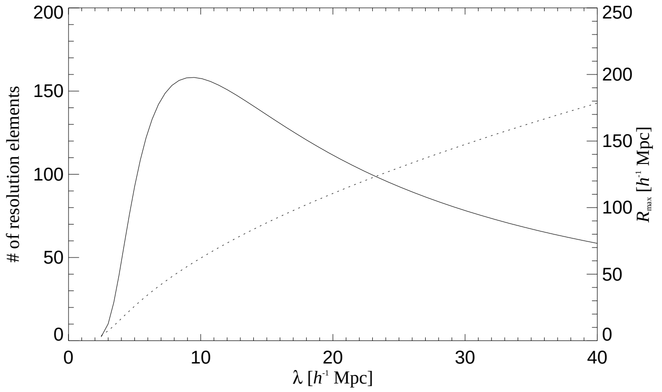

Figure 10. The number of resolution elements (solid) for the 1.2-Jy survey when the maximal survey volume is used. Also shown is the radius (dashed) of the usable survey volume.

Table 3. The smoothing lengths adopted for the topological analysis of the 1.2-Jy survey. Listed are the chosen survey depth $R_{\rm max}$, the resulting number $N_{\rm res}$ of resolution elements and the number $N_{\rm gal}$ of galaxies inside the survey volume.

λ [h^{-1}Mpc]	$R_{\rm max}$ [h^{-1}Mpc]	$N_{\rm res}$	$N_{\rm gal}$
5	26.00	92.6	1030
7	42.23	144.7	1783
10	62.07	157.6	2820
14	83.73	140.9	3582
20	110.63	111.5	4199

4.1.2 Resolution elements

The notion of number of resolution elements provides a useful way to compare roughly the statistical power of genus measurements. Because the smoothing extends over an effective volume $V_{\rm sm} = \pi^{3/2}\lambda^3$ the number of independent structures that can be present in a finite survey volume is limited. This number is of order

$$N_{\rm res} = \frac{V_{\rm s}}{V_{\rm sm}} = \frac{\omega R_{\rm max}^3}{3\pi^{3/2}\lambda^3}, \qquad (14)$$

where ω is the solid angle covered by the survey.

The number $N_{\rm res}$ indicates the power of a data set used for topological analysis. With the QDOT survey Moore et al. (1992) reached a maximum of about $N_{\rm res} = 80$ whereas the CfA survey allowed Vogeley et al. (1994) to achieve $N_{\rm res} = 260$ for their best subsample. The most powerful dataset examined so far is PSCz, making it possible for Canavezes et al. (1997) to reach $N_{\rm res} = 415$. For the 1.2-Jy redshift survey we have to be content with $N_{\rm res} = 158$. This already indicates that one can hardly expect very tight constraints from the genus statistics of this data set. Instead, the sparse sampling of the density field can be expected to severely limit the power of the genus test.

For the genus analysis we have examined the smoothing lengths 5, 7, 10, 14, and 20 h^{-1}Mpc in approximately logarithmic spacing. Table 3 lists some relevant parameters for the different cases.

4.2 Construction of mock surveys

We use the Virgo N-body simulations to obtain artificial redshift surveys that mimic the statistical properties of the 1.2-Jy survey with respect to sky coverage, selection function and luminosity distribution. The suites of mock catalogues are then analysed in exactly the same way as the 1.2-Jy redshift survey data. In this way sampling noise, cosmic variance, and systematic biases can be reliably modeled, which allows a fair assessment of the viability of the models, even if strong biases in our analysis existed.

To construct a mock catalogue we first select an arbitrary observer position in the periodic simulation box, which we replicate periodically in order to allow the construction of catalogues with sufficient depth. For computational convenience, we adopt a depth of 239.5 h^{-1}Mpc, corresponding to the size of the simulation box. We do not attempt to restrict the observers to positions that match certain properties of the immediate neighbourhood of the solar system because we want to allow for a realistic degree of cosmic variance. Because we have only one simulation at our disposal, not all the mock catalogues are independent. However, the survey volumes considered are much smaller than the simulation volume itself.

We identify every dark matter particle with a possible galaxy site, i.e. we construct only unbiased galaxy catalogues. We then compute the cosmological redshifts z_i of the potential galaxy sites with respect to the observer and draw uniformly distributed random numbers $x_i \in [0,1]$ for them. Only those galaxies with $x_i \le S(z_i)$ are kept, where $S(z)$ is the selection function scaled such that $S(z_0) = 1$ for some very small z_0. This selection results in a catalogue with selection function proportional to $S(z)$.

We finally discard the galaxies behind the angular mask of the 1.2-Jy survey and degrade the source density randomly such that the mock surveys contain the same number of sources (5083) as the 1.2-Jy survey in a sphere of radius 239.5 h^{-1}Mpc.

In a second step (optional for this work) we assign observed fluxes to the galaxies according to the luminosity function that results as a consequence of the adopted selection function (Springel & White 1997). For this purpose we draw a random number $q_i \in [0,1]$ from a uniform distribution, compute a maximum redshift

$$z_{\rm max} = S^{-1}\left((1-q_i)S(z_i)\right), \qquad (15)$$

and assign an observed flux

$$f_i = f_{\rm min}\frac{r_{\rm max}^2}{r_i^2}\left(\frac{1+z_{\rm max}}{1+z_i}\right)^{1-\alpha} \qquad (16)$$

for each source. Here S^{-1} denotes the inverse selection function, and $f_{\rm min} = 1.2$ Jy is the flux limit. Note we here assumed a straight K-correction with $\alpha = -2$, a non-evolving luminosity function, and an Einstein-de-Sitter universe. For $\Omega_0 \neq 1$, the latter may be replaced by the appropriate cosmology.

We usually neglect peculiar velocities in the construction of mock catalogues. As pointed out above, redshift space distortions are generally found to have only a minor effect on the genus statistics, at least on the relatively large scales accessible so far. We test that this is indeed the case by producing an additional set of mock catalogues in redshift space, where we adopt the velocities of the dark matter particles themselves as peculiar velocities of the mock galaxies.

4.2.1 The mock samples

We construct two main suites of mock catalogues; one is drawn from the SCDM simulation, the other from the ΛCDM simulation. Each of them contains 500 1.2-Jy mock surveys, that feature 5083 galaxies on 87.6 per cent of the sky, flux limited at 1.2-Jy and volume limited at a depth of $239.5\,h^{-1}$Mpc.

These samples are used to develop the statistical methodology for the comparison of the 1.2-Jy data to the N-body models. The large number of mock catalogues is necessary to determine the statistical distribution of the genus reliably. We restrict ourselves to the SCDM and ΛCDM simulations because we do not expect the genus of such a sparse sample as 1.2-Jy to be able to discriminate between ΛCDM, τCDM, and OCDM.

For the adaptive smoothing technique that we outline below, we fill the masked regions of the sky with a small set of fake galaxies. For simplicity, we construct for each catalogue 733 points with random angular positions in the masked regions, and with a redshift distribution sampled from the catalogue itself.

To examine a number of systematic effects we construct additional mock catalogues. For example, we use 100 mock catalogues in redshift space to assess the influence of redshift space distortion on the genus. We further compute additional catalogues that are full-sky 1.2-Jy surveys, i.e. which have no angular mask. These are used to test the influence of the mask. Finally, we compute a number of different survey realizations for fixed observer positions in order to separately determine the relative importance of sampling noise and cosmic variance.

5 SMOOTHING TECHNIQUES

We now outline our procedures to construct smoothed density fields from the 1.2-Jy redshift survey and the mock catalogues. These fields are then used as input to the genus computation.

5.1 Fixed smoothing

Assuming a universal luminosity function an unbiased estimate of the galaxy density field can be obtained by weighting the discrete point distribution $m(\boldsymbol{r})$ of the observed galaxies with the inverse of the selection function $S(r)$:

$$\rho(\boldsymbol{r}) \propto \frac{m(\boldsymbol{r})}{S(r)}. \tag{17}$$

We obtain an estimate of the density field smoothed on some scale λ by convolving with a filter $W(\boldsymbol{r})$, which we choose as the Gaussian of equation (9).

However, due to the absence of galaxies in the regions of the angular mask, the density would be systematically underestimated at locations close to unobserved patches of the sky if the smoothing were just done by a straightforward use of the kernel of equation (9). In order to avoid this problem we employ the ratio method proposed by Melott & Dominik (1993), who have shown in a systematic study that a smoothing according to

$$\hat{\rho}(\boldsymbol{r}) = \frac{\int W(\boldsymbol{r}-\boldsymbol{r}')\rho(\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}'}{\int W(\boldsymbol{r}-\boldsymbol{r}'')M(\boldsymbol{r}'')\,\mathrm{d}\boldsymbol{r}''}, \tag{18}$$

leads to the smallest loss or distortion of topological information compared to a number of alternative schemes that treat the mask differently. Here $M(\boldsymbol{r})$ is a mask field defined to be equal to 0 for $\boldsymbol{r}$ lying behind the angular mask and to be 1 otherwise. For this choice the denominator of equation (18) essentially renormalizes the smoothing kernel to the survey volume visible from the reference point $\boldsymbol{r}$.

In the actual computation of the genus curve we only use the volume with $M(\boldsymbol{r}) = 1$ which is not hidden by the mask. Additionally, we restrict the genus computation to a sphere carved out of the smoothed density field. Note that there is no boundary smoothing effect due to the outer surface of this sphere since we also include the sources outside this final region in the smoothing process.

We compute the convolutions that appear in the numerator and denominator of equation (18) with the help of a Fast Fourier Transform (FFT) on a 128^3 mesh. We choose a grid size of $b = \lambda/8$, which ensures that the genus is free of finite mesh size effects (Hamilton, Gott & Weinberg 1986), as we will demonstrate below. The final depth $R_{\max}$ of the density field we use for the topological analysis is always small enough to avoid wrap around effects due to the periodic FFT smoothing.

5.2 Adaptive smoothing

As we saw in the analysis of the Virgo simulations, the smoothing scale is limited by the poor sampling of underdense regions. To make use of the additional resolution in high density regions a variable smoothing length can be employed. We have already implemented an adaptive smoothing scheme for the fully sampled N-body simulations. We now define such a scheme for the analysis of flux limited redshift surveys as well. Again, the hope is that the overall effect of adaptive smoothing is an increase of the number of structure elements visible in a given density field.

5.2.1 Spherically symmetric kernel

We start by making the kernel a function of position, i.e. the smoothed density field is computed as

$$\hat{\rho}(\boldsymbol{r}) = \int \rho(\boldsymbol{r}')W(\boldsymbol{r}-\boldsymbol{r}';\boldsymbol{r}')\mathrm{d}\boldsymbol{r}'. \tag{19}$$

This definition corresponds to the *scatter* approach of SPH. Since this scheme strictly conserves 'mass' (if W is properly normalized), it seems more natural to use for a small number of tracer particles than the *gather* formulation, which we employed for the densely sampled N-body simulations.

In a first variant of the adaptive smoothing, we stay with a spherical Gaussian and allow only the smoothing scale to vary with the local density. Later we will generalize the technique to triaxial Gaussians.

For a given survey volume $V_{\rm s}$ and a prescribed smoothing scale λ_0 we first compute the average mass M_0 in a Gaussian sphere of radius λ_0, that is

$$M_0 \equiv \frac{\pi^{\frac{3}{2}}\lambda^3}{V_s}\int_{V_s}\rho(\boldsymbol{r})\mathrm{d}\boldsymbol{r} = \pi^{\frac{3}{2}}\lambda_0^3, \qquad (20)$$

where the last equality holds (at least on average) due to our normalization of the selection function.

We then compute for every galaxy site $\boldsymbol{r}_i$ an individual smoothing radius λ_i such that

$$\int \rho(\boldsymbol{r})\exp\left[-\frac{(\boldsymbol{r}-\boldsymbol{r}_i)^2}{\lambda_i^2}\right] = M_0. \qquad (21)$$

This definition implies that λ_i will be smaller than λ_0 if the density around $\boldsymbol{r}_i$ is higher than the mean, and it will be larger if the local density is lower than the mean.

5.2.2 *Triaxial kernel*

Up to now we have only varied the volume of the kernels. In an attempt to improve the flexibility of the smoothing we may also allow the shape of the kernel to vary. For this purpose we adopt triaxial Gaussians

$$W(\boldsymbol{x};\boldsymbol{r}) = \frac{1}{\pi^{\frac{3}{2}}(\det\boldsymbol{\Lambda})^{\frac{1}{2}}}\exp\left(-\boldsymbol{x}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{x}\right) \qquad (22)$$

as kernels, where the quadratic form $\boldsymbol{\Lambda}$ is a function of $\boldsymbol{r}$. We now need to specify the matrices $\boldsymbol{\Lambda}_i$ for every particle. For simplicity, we set $\boldsymbol{\Lambda}_i$ proportional to the local moment of inertia tensor around the site $\boldsymbol{r}_i$, i.e.

$$\boldsymbol{\Lambda}_i \propto \int (\boldsymbol{r}-\boldsymbol{r}_i)(\boldsymbol{r}-\boldsymbol{r}_i)^T\rho(\boldsymbol{r})\exp\left[-\frac{(\boldsymbol{r}-\boldsymbol{r}_i)^2}{\lambda_i^2}\right]\mathrm{d}\boldsymbol{r}, \qquad (23)$$

and we keep the original smoothing volume fixed by requiring $(\det\boldsymbol{\Lambda}_i)^{\frac{1}{2}} = \lambda_i^3$. Once the matrices $\boldsymbol{\Lambda}_i$ are determined we ‘only’ need to compute

$$\hat{\rho}(\boldsymbol{r}) = \sum_i m_i W(\boldsymbol{r}-\boldsymbol{r}_i;\boldsymbol{\Lambda}_i) \qquad (24)$$

with $m_i = [S(\boldsymbol{r}_i)]^{-1}$ to arrive at the triaxially smoothed density field.

Because the adaptive smoothing has to be carried out in real space it requires much more CPU time than the fixed smoothing described above. For this reason we construct the density fields only in spheres of radius $R_{\max}$, each of them inscribed in a 64^3 mesh. For simplicity we deal with the mask by filling the empty region with fake galaxies as described above. Note that the masked volume is not used in the calculation of the genus curve.

The adaptive smoothing schemes we employ here are by no means unique and many other variants are conceivable. In fact, we have tried a number of alternatives ourselves. However, the suggested procedure is fairly intuitive and allows a study of the effects of volume and shape adaptivity separately. Currently we see no alternative to Monte-Carlo experiments in determining the performance and properties of such adaptive smoothing techniques.

6 SYSTEMATIC EFFECTS

Before turning to the results for the 1.2-Jy survey and the Virgo mock catalogues, we first examine various systematic effects that can affect the genus measurements. In particular, we are interested in the amount of bias present in the genus curve derived from a 1.2-Jy-like sample compared to the genus of the underlying density field.

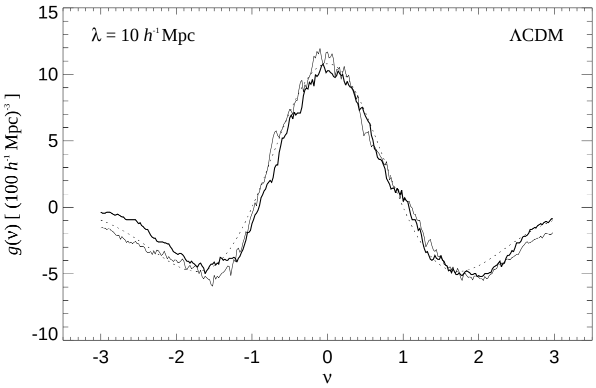

Figure 11. Finite volume effect. The thin line shows the average genus curve of 10 spherical subvolumes of radius $60\,h^{-1}$Mpc extracted from the SCDM simulation, while the thick line gives the genus of the full simulation volume. The genus curves have been computed for a resolution of $\Delta\nu = 0.01$. Comparing the two curves we find no evidence for the strong amplitude bias claimed by Protogeros & Weinberg (1997). They claim an increase of the genus amplitude for the finite volume curves by a factor of 1.5-2.0.

6.1 Finite volume effect

Recently, Protogeros & Weinberg (1997) have claimed that the genus curve is severely biased high if it is computed from subvolumes carved out of large N-body simulations or out of Gaussian random fields. In particular, for a roughly spherical volume of radius $R_{\max} = 60\,h^{-1}$Mpc and smoothing scale $\lambda = 10\,h^{-1}$Mpc they detected an increase of the genus amplitude by a whopping factor of 1.5-2. They further found that this volume effect becomes somewhat smaller for larger volumes, yet it seems to be present independently of the shape of the survey volume.

Their result is puzzling since the volumes used in their analysis seem large enough that any correlation of the curvature between adjacent points on the surface of the volume is expected to average out, i.e. the mean curvature of the subvolume should allow an unbiased estimate of the genus of the full volume.

We have searched for this volume effect ourselves, but we could not find it. For example, in figure 11 we compare the average genus curve for 10 spherical subvolumes of radius $R_{\max} = 60\,h^{-1}$Mpc extracted from the SCDM simulation with the genus of the full simulation box. Reassuringly, the average genus curve of the subvolumes appears to be largely unbiased; there is no trace of the claimed strong finite volume effect.

6.2 Mask effects

We dealt with the limited sky coverage of the 1.2-Jy survey by employing the ratio smoothing method of Melott & Dominik (1993). Here we look for systematic biases inflicted on the genus curve because of that. For this purpose we have

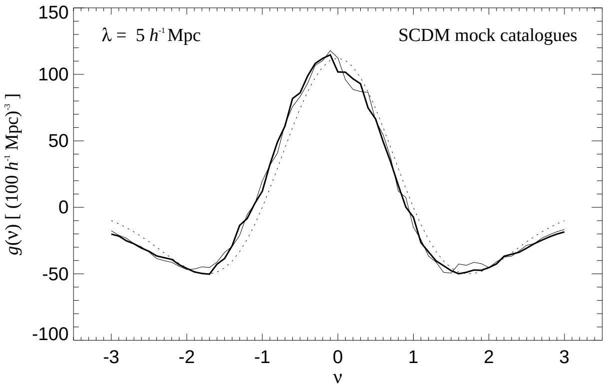

Figure 12. Effects due to the angular mask. The thick curve shows the average genus curve for 100 SCDM mock catalogues before they are subjected to the 1.2-Jy mask, i.e. their genus is computed for an all-sky density field. The thin line shows the genus for the masked mock catalogues, where the genus computation involves the ratio method. The curves shown are for fixed smoothing with $\lambda = 5\,h^{-1}$Mpc. Larger smoothing scales or the adaptive smoothing techniques result in similar small effects.

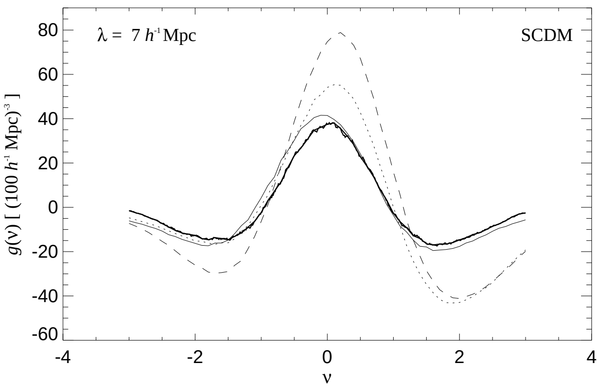

Figure 13. Genus curves for the SCDM simulation, smoothed at $7\,h^{-1}$Mpc with different smoothing techniques. The thick solid line gives the result for the fully sampled SCDM simulation, while the thin line gives the corresponding average genus curve with a fixed smoothing kernel for the SCDM mock catalogues. The dotted line shows the result for the spherical adaptive smoothing, and the dashed curve is for triaxial adaptive smoothing.

constructed an additional suite of 100 mock catalogues for the SCDM model. These catalogues have full sky coverage, with 5816 galaxies to a depth of $239.5\,h^{-1}$Mpc.

In figure 12 we compare the average genus curve of these full sky mock catalogues to the average curve of the masked catalogues, where the computation involves the ratio method. The good agreement between the two results gives us confidence that any bias due to our treatment of the mask is very small. The result shown is for fixed smoothing with $\lambda = 5\,h^{-1}$Mpc, but we observe a similar small influence of the mask for other smoothing scales and for the adaptive smoothing techniques.

6.3 Genus curves

Figure 13 shows the average genus curve of the SCDM suite of mock catalogues compared to the genus of the fully sampled simulation. Obviously the genus curve derived from the mock catalogues is biased high compared to full sampling. This can be explained by the influence of shot noise on the genus amplitude, as demonstrated by Canavezes et al. (1997). Furthermore, the average curve exhibits a slight *meatball* bias, that is a shift of the peak to the left. This effect can also be understood as a discreteness error that results from the sparse sampling of the density field.

Also shown in figure 13 are the mean genus curves resulting from the adaptive smoothing techniques we tried. Both of them show a strong enhancement of the genus amplitude. The spherically symmetric smoothing also results in a strong asymmetry between the minima of the genus curve, with the minimum on the positive side being lowered strongly, while the minimum on the negative side remains practically at the level obtained for the fixed smoothing technique. Note that the genus curves for the different smoothing techniques are expected to be different since the smoothing procedures are sensitive to different properties of the density field.

Interestingly, the triaxial smoothing technique gives a genus very close to the spherical smoothing around the minimum on the positive side, while it increases the genus signal for smaller values of ν. Presumably the amplitude of the minimum on the positive side is just set by the number density of individual, isolated clumps. Apparently, the triaxial technique does not change the topology of these isolated regions.

For fixed smoothing, the measured genus is biased compared to the fully sampled simulations, albeit by a small amount. Instead of trying to correct for it as attempted by Canavezes et al. (1997), we compare the 1.2-Jy measurements only to the results obtained for the mock catalogues, and not to the fully sampled simulations themselves. This is a viable procedure, even if strong biases are present.

7 STATISTICAL METHODOLOGY

Ultimately we want to use the genus statistic to compare theory with observation, i.e. to quantify the level of agreement of the 1.2-Jy survey with the Virgo N-body models. A prerequisite to derive formal exclusion levels is a precise understanding of the distribution of errors of the genus measurement.

Perhaps the most general method to assess random and systematic errors is to work with ensembles of mock galaxy surveys that mimic the statistical properties of the actual observed data set. When the mock catalogues and the redshift survey are analysed in the same way, systematic biases that might be present in the adopted analysis enter in the same way.

7.1 Distribution of errors

For a suite of n mock catalogues we measure the genus curve at k values $\nu_1, \nu_2, \ldots, \nu_k$ of the filling factor. In what follows, we compute the genus with spacing $\Delta\nu = 0.1$ in the range $[-3.0, 3.0]$, i.e. at $k = 61$ positions. We now use these measurements to estimate the distribution of errors in the genus. The mean genus curve and its covariance matrix may be estimated as

$$\overline{\boldsymbol{g}} = \frac{1}{n}\sum_{l=1}^{n} \boldsymbol{g}^{(l)} \tag{25}$$

and

$$\mathbf{V} = \mathrm{cov}(g_i, g_j) = \frac{1}{n-1}\sum_{l} (\boldsymbol{g}^{(l)} - \overline{\boldsymbol{g}})(\boldsymbol{g}^{(l)} - \overline{\boldsymbol{g}})^T, \tag{26}$$

where $\boldsymbol{g}^{(l)} = (g_1^{(l)}, \ldots, g_k^{(l)})$ denotes the measured genus curve for the catalogue l.

Recently Protogeros & Weinberg (1997) conjectured that the distribution of errors is well described by a multivariate Gaussian

$$f(\boldsymbol{g}) = \frac{1}{(2\pi)^{\frac{k}{2}}|\det\mathbf{V}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{g} - \overline{\boldsymbol{g}})^T\mathbf{V}^{-1}(\boldsymbol{g} - \overline{\boldsymbol{g}})\right]. \tag{27}$$

Then the quantity

$$\chi^2(\boldsymbol{g}) = (\boldsymbol{g} - \overline{\boldsymbol{g}})^T\mathbf{V}^{-1}(\boldsymbol{g} - \overline{\boldsymbol{g}}) \tag{28}$$

would exhibit a χ^2 distribution with k degrees of freedom, and one could estimate the probability

$$p = 1 - \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}\int_{\chi^2}^{\infty} x^{\frac{k}{2}-1}\mathrm{e}^{-\frac{x}{2}}\,\mathrm{d}x \tag{29}$$

of finding a mock survey in the ensemble that differs from the mean of the mock catalogues by more than a particular observation with $\chi^2 = \chi^2(\boldsymbol{g}^{(\mathrm{obs})})$. In this way a formal exclusion level could be derived.

This procedure seems attractive, yet we find that it can fail miserably in practice. As we show in the appendix, this failure is partly due to the fact that the errors are at best approximately distributed as a multivariate Gaussian. More importantly, we will use a principal components analysis (PCA) to demonstrate that there are only a small number of principal components that can be determined with some confidence. The rest of them are dominated heavily by noise; it is therefore not a good idea to invert the noisy 61×61 covariance matrix of equation (26). Instead we will regularize the problem by means of a PCA. Note that a simple smoothing of the genus curve can remove some of the noise. However, as a side effect this will make the covariance matrix close to singular. Of course, this causes trouble if one naively goes ahead and tries to compute V_{ij}^{-1}, so a PCA cannot be avoided in this way.

7.2 Principal components analysis

The principal components analysis (PCA, Murtagh & Heck 1987) is frequently applied in astronomy to extract the most relevant features from data sets that may be strongly contaminated by noise. Since PCA is a linear method it works best for uncorrelated noise, a situation not really valid for the genus. However, we can still expect that it allows the construction of a clean multivariate analysis of the measured genus curves.

Each measured genus curve can be viewed as a point in a 61-dimensional space. The idea of PCA is to transform to a new set of coordinates which correspond to the directions of maximum extension of the cloud of measured genus points. These principal axes are just the eigenvectors of the covariance matrix (26). It is convenient to order these components in descending order of their eigenvalues, i.e. the first principal component shows the highest variance. Usually one then considers only the first few principal components, which are the ones that describe the most prominent features of the signal. In this way the method allows to efficiently filter out the noisy contributions to the signal and to concentrate on its essential features.

The principal axes and their eigenvalues can be conveniently found by a singular value decomposition (SVD) of the covariance matrix. Having ordered the principal components in descending order we consider only the first m of them. We can then construct a $m \times 61$ matrix $\mathbf{P}$ that contains the m eigenvectors in its rows and that projects a genus curve $\boldsymbol{g}$ onto new coordinates $\boldsymbol{h} = \mathbf{P}\boldsymbol{g}$. One can also transform back to the original space, giving rise to a PCA-filtered genus curve

$$\boldsymbol{g}_{\mathrm{PCA}}(m) = \mathbf{P}^T\mathbf{P}\boldsymbol{g}. \tag{30}$$

In figure 14 we display the first six principal components of the genus measurement. Actually these principal components have been derived from an averaged covariance matrix, obtained by adding up matrices corresponding to several smoothing scales which we have scaled to a common highest eigenvalue. The curves resulting from just one suite of genus curves look very similar, although they are not quite so smooth. We employ the averaging procedure to establish a generic set of smooth orthogonal principal components that we subsequently apply to all the different samples on an equal footing. Of course, depending on the covariance matrices used in the averaging procedure the derived principal components may differ slightly in detail. However, in all cases the first few components (we will use $m = 6$ of them) span very nearly the same region in the full 61-dimensional space of measured genus curves. Hence the specific choice of covariance matrices is uncritical as long as all the information of these principal components is used in a multivariate analysis.

Some of the principal components shown in figure 14 are easy to interpret. The first clearly measures the amplitude, while the next two can be seen to be sensitive to a horizontal shift and a broadening of the genus curve. Hence the first three principal components are similar in meaning to the genus meta-statistics introduced by Vogeley et al. (1994). However, here these measures are not postulated in an ad-hoc way, but they suggest themselves naturally as the most relevant features of the measured genus curves.

How many principal components should we reasonably take? To answer this question we examine a global error function

$$E(m) = \sum_{l} (\boldsymbol{g}_{\mathrm{PCA}}^{(l)}(m) - \overline{\boldsymbol{g}})^2, \tag{31}$$

where the sum is over the mock ensemble and $\boldsymbol{g}_{\mathrm{PCA}}^{(l)}(m)$ denotes the genus curve of catalogue l, treated with a PCA filter of order m. Figure 15 shows a minimum of $E(m)$, when $m \approx 6$ principal components are used. Higher principal components lead to additional noise, in the sense that the reconstructed genus curves differ more and more from

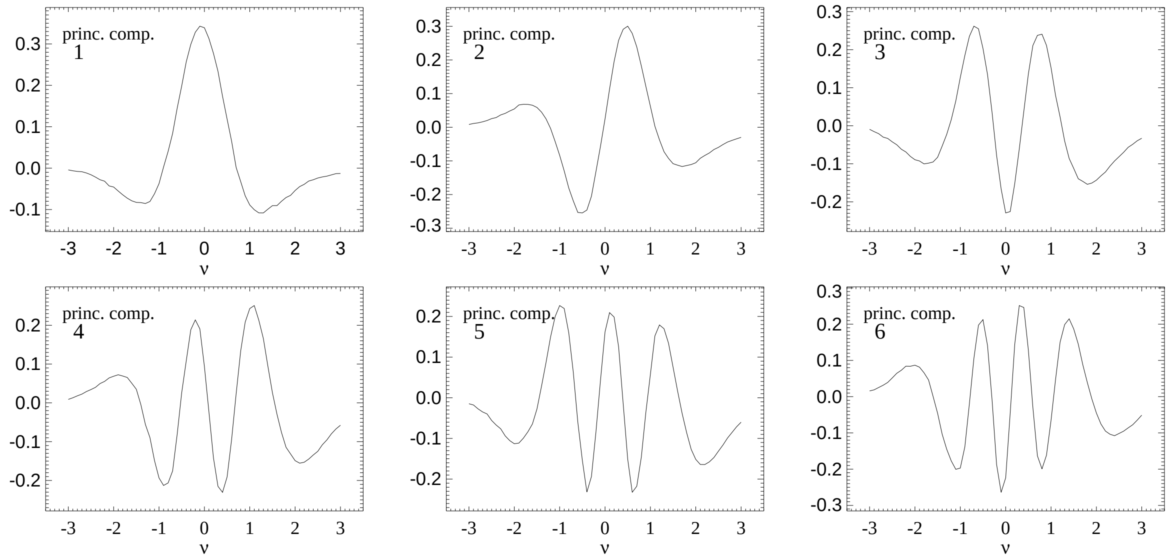

Figure 14. Principal components of the genus measurement. The displayed curves are based on an average of covariance matrices corresponding to different smoothing scales. The matrices have been scaled to a common highest eigenvalue before the averaging. The resulting curves are similar to the ones obtained for a single covariance matrix, but they are smoother due to the reduction of noise by the averaging. We use these curves as generic principal components in the PCA of all examined samples. Note that the principal components are normalized to unity and mutually orthogonal to each other.

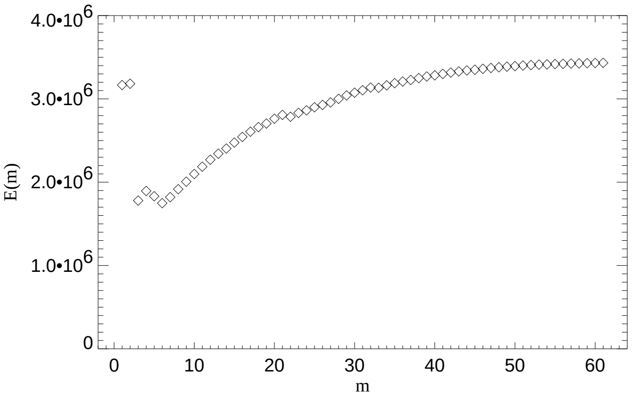

Figure 15. Global approximation error E as a function of the number m of included principal components. The points shown are for the SCDM mock catalogues, smoothed with $7\,h^{-1}$Mpc. Other smoothing scales show similar results. For the adaptive smoothing techniques the minimum occurs also at $m \approx 6$. However, the higher components lead to a somewhat smaller increase of E.

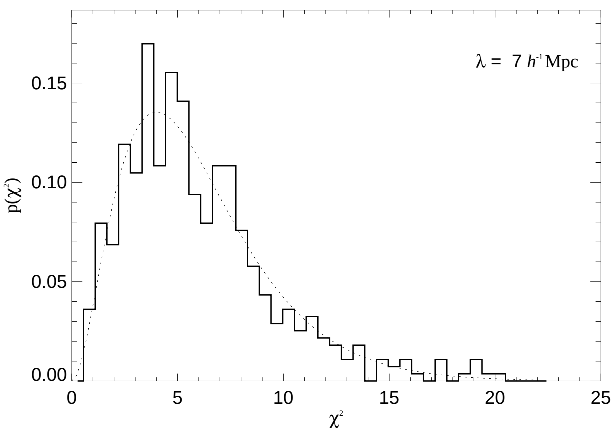

Figure 16. Distribution of $\chi^2(\boldsymbol{h})$, when only the first 6 principal components are included. The variance of the distribution is 12.66, close to the expected value of 12.0 for a χ^2 distribution with $m = 6$ degrees of freedom (dotted line). The data shown here are for the SCDM catalogues with fixed smoothing of $7\,h^{-1}$Mpc. All the other cases show equally good agreement with the theoretical χ^2 distribution. Note that the test curve $\boldsymbol{g} = 2\overline{\boldsymbol{g}}$, which demonstrates the failure of the method applied in the appendix (figure A1), here gives a χ^2 of 32.4, indicating correctly a terrible fit.

the ensemble average. For this reason we will restrict ourselves to the first 6 principal components.

As is seen in figure 14 the first six principal components give only small weight on the genus at low and high values of ν, where we know that the errors are not normally distributed. Because of that, one can hope that the distribution of the principal components $\boldsymbol{h}$ is consistent with a Gaussian, at least if we only use the first $m = 6$ of them. A Kolmogorov-Smirnov test (appendix A) reveals that this is indeed the case. We will therefore attempt an ordinary multivariate analysis based on the fixed set of principal components displayed in figure 14. Since we keep them fixed for all suites of mock catalogues, we do not expect that the correlations between the different principal components will exactly vanish, although they should be small. For this reason we compute the covariance matrix $\mathbf{C}$ of the measured values of $\boldsymbol{h}$, and consider the statistic

$$\chi^2(\boldsymbol{h}) = (\boldsymbol{h} - \overline{\boldsymbol{h}})^T \mathbf{C}^{-1} (\boldsymbol{h} - \overline{\boldsymbol{h}}). \tag{32}$$

In figure 16 we show an example of the distribution of this quantity. It is indeed very well fit by a χ^2 distribution with $m = 6$ degrees of freedom. Note that for this statistic

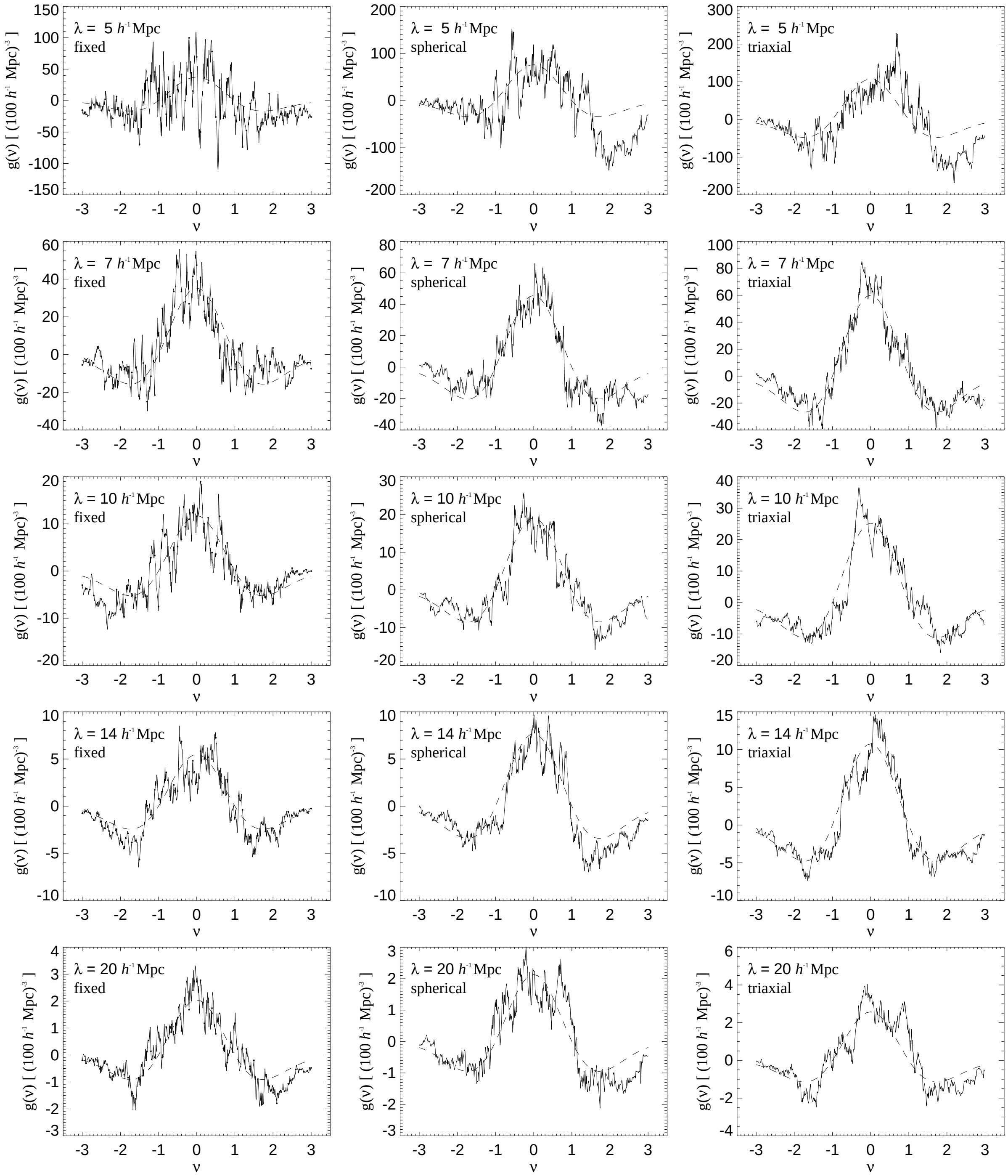

Figure 17. Genus curves of the 1.2-Jy redshift survey of *IRAS* galaxies. The left column shows high resolution genus curves obtained with fixed smoothing, while the middle and right columns give the results for the two adaptive smoothing schemes. In each row the same smoothing scale is considered, rising from $5\,h^{-1}$Mpc to $20\,h^{-1}$Mpc. Note however that the vertical scale is different for individual panels. The adaptively smoothed genus curves exhibit a considerably higher genus amplitude. The dashed lines are random phase genus curves with amplitudes determined from the first principal component of the measured genus curves. In the case of adaptive smoothing, λ sets the mass per smoothing volume via equation (20).

the pathological result obtained in appendix A for the curve $\boldsymbol{g} = 2\overline{\boldsymbol{g}}$ is gone; here this curve gives $\chi^2 = 32.4$, implying a terrible disagreement, as it should.

8 COMPARISON OF THE 1.2-JY SURVEY WITH Virgo

8.1 1.2-Jy genus curves

Figure 17 shows the genus curves we obtain for the 1.2-Jy redshift survey with three different smoothing schemes.

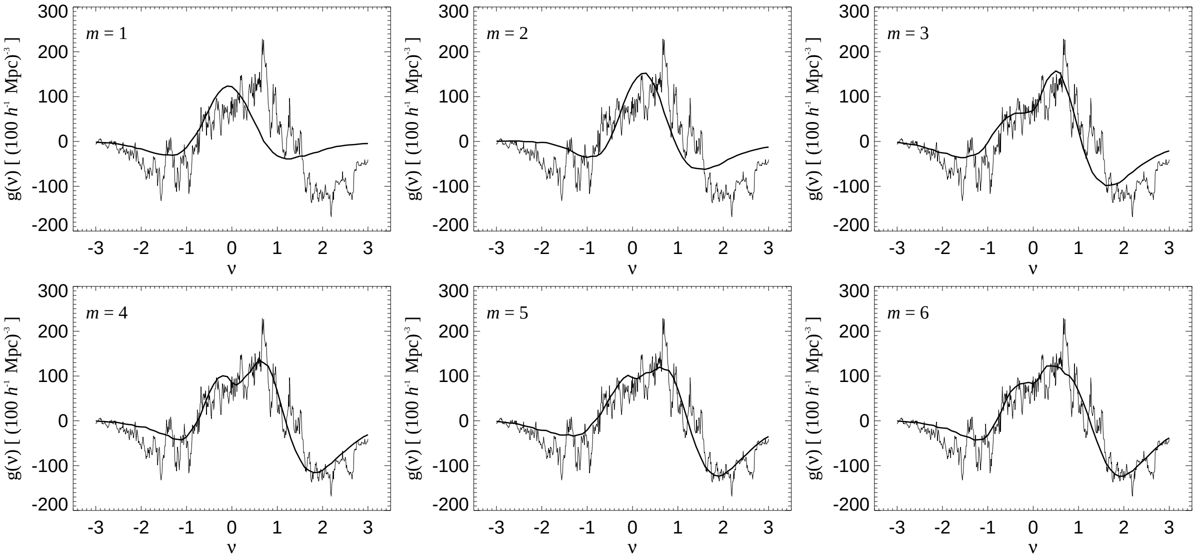

Figure 19. Example for a PCA-filtered genus curve as a function of the number m of included principal components. In each panel the thin line shows the 1.2-Jy genus curve resulting for triaxial adaptive smoothing with $\lambda = 5\,h^{-1}$Mpc. The thick line gives the corresponding PCA-filtered curve for the specified order m of the filter.

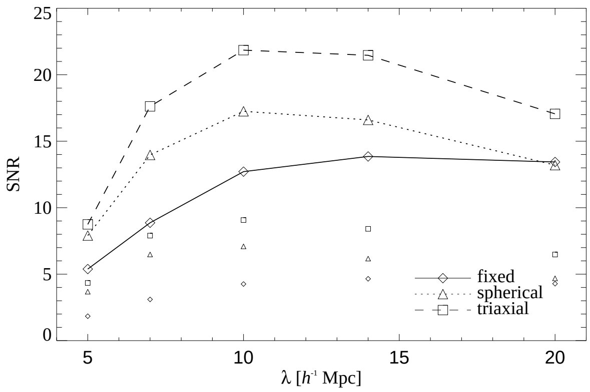

Figure 18. Average signal-to-noise ratio of the measured genus curves for different smoothing techniques as a function of smoothing scale. The results are for the SCDM mock catalogues and PCA-filtered ($m = 6$) genus curves. The small symbols show the corresponding SNR for the raw genus curves. The ΛCDM catalogues give a similar result.

The left column displays results for fixed smoothing, while the middle and right columns give the corresponding curves for the two variants of adaptive smoothing. The genus curves in the figure are computed with a uniform spacing of $\Delta\nu = 0.01$ in the interval $-3 \le \nu \le 3$. Note that the genus curves that are used for the comparison with the Virgo simulations represent only a subsampling of this data with a spacing of $\Delta\nu = 0.1$.

The large amount of jitter in the curves indicates that there is substantial noise in the measurement. Interestingly, the adaptively smoothed genus curves show less scatter and seem to be smoother than with fixed smoothing. Clearly, the triaxial smoothing technique produces the smoothest genus curves. It might therefore be hoped that this reflects an increase of the number of resolved structural elements which in turn allows a measurement of the genus curve with less error.

In order to test this expectation we define an average signal-to-noise ratio (SNR)

$$\langle \mathrm{SNR} \rangle \equiv \left\langle \frac{\overline{g}^2}{\sigma^2} \right\rangle = \frac{1}{k} \sum_{i=1}^{k} \frac{\overline{g}_i^2}{\sigma^2(g_i)} \qquad (33)$$

of the measured genus curves. Here $\overline{g}_i$ denotes the mean genus density at each of the $k = 61$ measured positions, and $\sigma^2(g_i)$ is the variance of the corresponding measurements.

In figure 18 we show the SNR for the three different smoothing techniques as a function of smoothing scale. As expected, the adaptive smoothing techniques can significantly improve the SNR compared to the fixed smoothing scheme. Also, the triaxial method clearly performs better than the spherical adaptive smoothing. Note that figure 18 shows the SNR for genus curves that are filtered with a PCA-filter of order $m = 6$. The SNR for the raw genus curves is also plotted; it is generally smaller, showing that the PCA-filter can indeed reduce the contamination with noise.

A further illustration of the effect of the PCA-filtering is given in figure 19. Here we show the 1.2-Jy genus curve resulting for the triaxial adaptive smoothing technique with $\lambda = 5\,h^{-1}$Mpc together with PCA-filtered versions of it. As the number m of included principal components is increased, more features of the measured genus curve can be faithfully reproduced by the filtered curve. However, adding in more principal components will eventually only lead to a reproduction of the noise inherent in the measured curve.

8.2 Comparison with Virgo

As outlined in section 7 we base our statistical comparison between Virgo and the 1.2-Jy survey on 6 principal components of the measured genus curves. The amplitudes of the principal components are just the projections of the measured genus on the curves shown in figure 14.

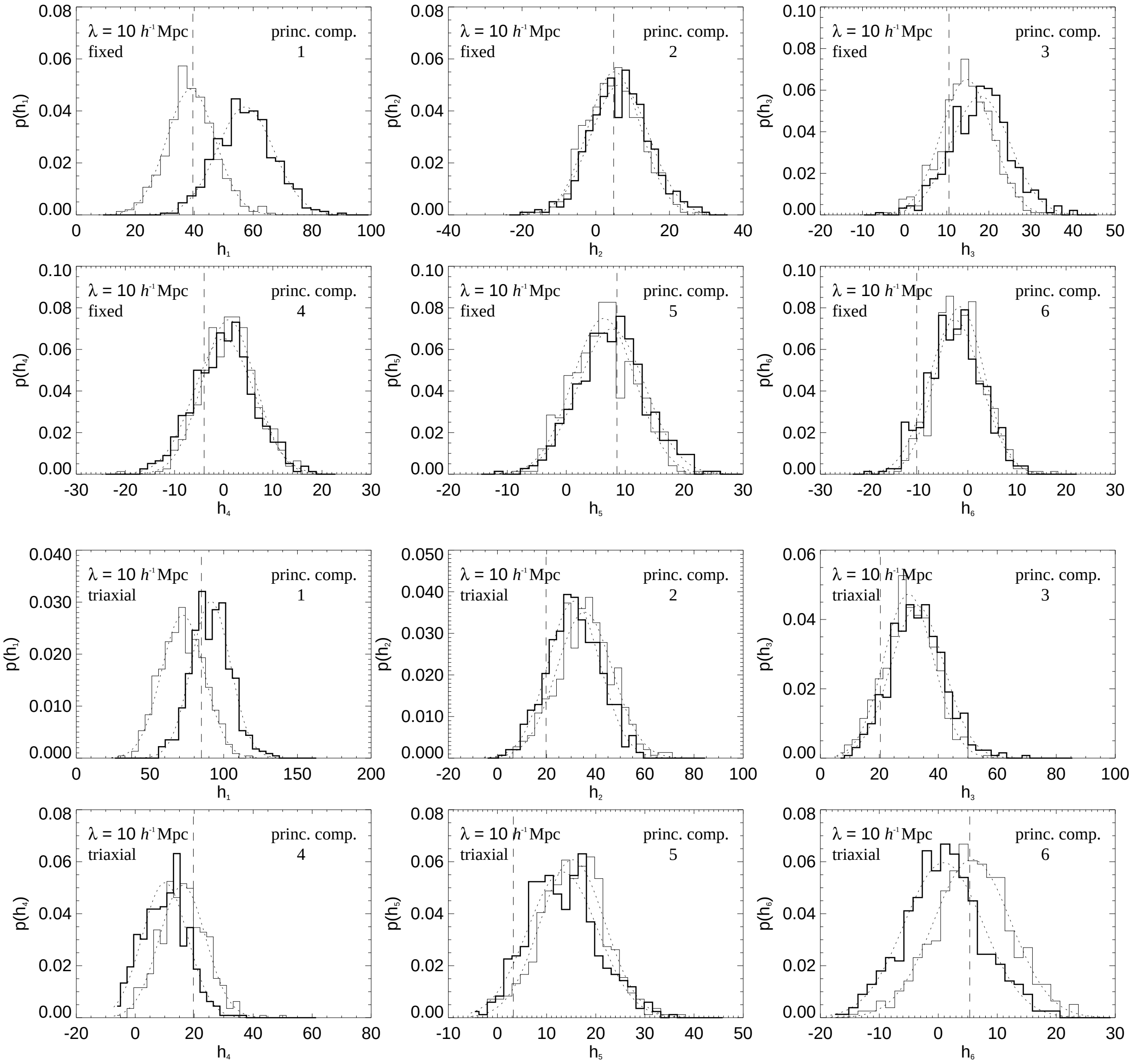

Figure 20. Distribution of the measured principal components for the SCDM (thick histogram) and ΛCDM (thin histogram) models. In each panel, the dashed vertical line marks the result for the 1.2-Jy survey. The dotted lines are normal distributions with the mean values and variances of the individual histograms. The data shown compares the distribution of the principal components for the fixed and triaxial adaptive smoothing techniques with $\lambda = 10\,h^{-1}$Mpc. Other smoothing scales show qualitatively similar results.

In figure 20 we show examples of the distribution of the principal components for fixed smoothing with $\lambda = 10\,h^{-1}$Mpc. In each panel, the solid histogram shows the result for the SCDM suite of mock catalogues, while the thin line gives the ΛCDM distribution. The dotted lines show the normal distributions that result from the mean and the variances of these histograms, and the dashed vertical line marks the measurement for the 1.2-Jy survey itself. The panels of the bottom half of figure 20 show the corresponding plots for the triaxial adaptive smoothing technique.

It is immediately apparent, that the discrimination between SCDM and ΛCDM is poor with a dataset as sparse as 1.2-Jy. The amplitude h_1 proves to be the most sensitive measure of differences between the models. However, the large uncertainties in the genus measurement manifest themselves in broad, overlapping distributions of the principal components. Nevertheless we can still compute the exclusion probability of the 1.2-Jy measurements for each of the models. For this purpose we calculate the $\chi^2(\boldsymbol{h})$ value of the 1.2-Jy measurement with respect to the distributions of the SCDM and ΛCDM models, and the probability p that a mock catalogue shows a higher χ^2 than the 1.2-Jy survey. Here we assume that the principal components are multivariate normally distributed.

Additionally we give the likelihood ratio

$$\mathcal{L}\left(\tfrac{\Lambda\mathrm{CDM}}{\mathrm{SCDM}}\right) = \frac{|\det \mathbf{C}_{\mathrm{S}}|^{\frac{1}{2}}}{|\det \mathbf{C}_{\Lambda}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\chi^2_{\Lambda} - \chi^2_{\mathrm{S}}\right)\right] \tag{34}$$

between the ΛCDM and SCDM models. Here χ^2_{S} and χ^2_{Λ} denote the χ^2-values of the 1.2-Jy data with respect to the SCDM and ΛCDM samples, and $\mathbf{C}_{\mathrm{S}}$ and $\mathbf{C}_{\Lambda}$ are the 6×6

Table 4. Comparison of SCDM and ΛCDM with the 1.2-Jy survey assuming a multivariate normal distribution for the principal components. Listed are the χ^2 values ($m = 6$ degrees of freedom) for the 1.2-Jy survey when it is compared to either the SCDM or the ΛCDM suite of mock catalogues, and the resulting exclusion levels p. We also compute the relative likelihood of ΛCDM compared to SCDM.

Fixed smoothing

	SCDM		ΛCDM		
λ [h^{-1}Mpc]	χ^2	p	χ^2	p	$\mathcal{L}\left(\frac{\Lambda\mathrm{CDM}}{\mathrm{SCDM}}\right)$
5	7.22	0.30	3.91	0.69	8.122
7	7.06	0.32	5.43	0.49	4.316
10	8.05	0.23	4.64	0.59	10.790
14	5.46	0.49	5.15	0.52	2.429
20	9.04	0.17	7.30	0.29	4.465

Adaptive smoothing (spherical)

	SCDM		ΛCDM		
λ [h^{-1}Mpc]	χ^2	p	χ^2	p	$\mathcal{L}\left(\frac{\Lambda\mathrm{CDM}}{\mathrm{SCDM}}\right)$
5	5.52	0.48	2.10	0.91	5.208
7	12.97	0.04	10.72	0.10	2.860
10	5.76	0.45	6.98	0.32	0.637
14	4.98	0.55	6.82	0.34	0.382
20	3.62	0.73	3.90	0.69	0.911

Adaptive smoothing (triaxial)

	SCDM		ΛCDM		
λ [h^{-1}Mpc]	χ^2	p	χ^2	p	$\mathcal{L}\left(\frac{\Lambda\mathrm{CDM}}{\mathrm{SCDM}}\right)$
5	4.87	0.56	2.25	0.90	3.478
7	11.31	0.08	8.97	0.18	2.852
10	7.26	0.30	8.14	0.23	0.633
14	7.34	0.29	8.67	0.19	0.524
20	11.88	0.06	10.40	0.11	2.545

covariance matrices of the measured principal components. We list a summary of our results in table 4.

All the measured probabilities p are so high, that the 1.2-Jy genus data are consistent with being drawn from either of the two CDM models, i.e. the genus test cannot rule out the SCDM or ΛCDM model with high significance when only one smoothing scale is considered. However, we can still make a statement about their relative likelihoods given the measured 1.2-Jy data. For the fixed smoothing, the ΛCDM model is more likely than the SCDM model for all smoothing scales. This preference of ΛCDM is also found with the adaptive smoothing, however at a weaker level since the results actually favour SCDM at $10\,h^{-1}$Mpc and $14\,h^{-1}$Mpc.

To improve the discriminative power of the genus test we can try to combine the measurements corresponding to different smoothing scales. As we have demonstrated above, the genus amplitude (i.e. the first principal component h_1) is most sensitive to differences between the models, essentially because it measures the shape of the power spectrum. Hence we will focus on it in the following.

In figure 21 we show the average genus amplitudes of the SCDM and ΛCDM mock catalogues together with the measurements for the 1.2-Jy survey. The models follow approximately parallel lines, albeit at different heights, when we plot dimensionless genus densities $(2\pi)^2\lambda^3 N$. To quantify the overall agreement between the 1.2-Jy survey and the models we consider two simple statistics. First we compute a χ^2-like quantity by adding up the amplitude measurements corresponding to the different smoothing scales, viz.

$$Q = \sum_i \frac{(A_i^{\mathrm{m}} - \overline{A}_i)^2}{\sigma_i^2}, \tag{35}$$

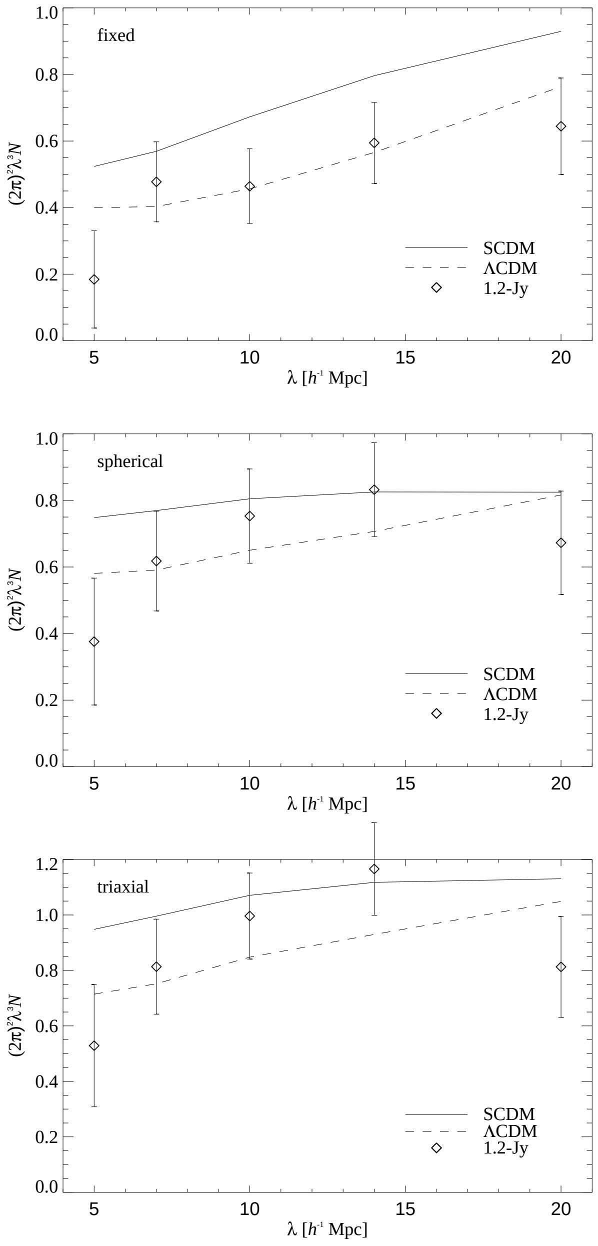

Figure 21. Measured genus amplitudes for the SCDM and ΛCDM mock catalogues and the 1.2-Jy survey. Each panel shows results for one of the three different smoothing schemes employed. The mean of the SCDM mock catalogues is displayed as solid line, the ΛCDM result is the dashed curve, and the diamonds mark the 1.2-Jy measurements. The attached error bars indicate the rms scatter of the SCDM sample. Note that we have normalized the first principal component such that the measured amplitude N corresponds to that of a best-fitting Gaussian curve (equation 5).

Table 5. Comparison of the SCDM and ΛCDM models with the 1.2-Jy survey using a combination of genus amplitudes for different smoothing scales. Listed are the fractions $p_{\rm Q}$, $p_{\rm R}$ of mock catalogues that fit the ensemble mean worse than the 1.2-Jy measurements. The $p_{\rm Q}$ statistic is just a sum over the χ^2-deviations at individual smoothing scales, and the $p_{\rm R}$ statistic is based on a fit of the run of genus amplitude with smoothing scale.

		SCDM	ΛCDM
fixed	$p_{\rm Q}$	0.012	0.522
	$p_{\rm R}$	0.002	0.696
spherical	$p_{\rm Q}$	0.330	0.590
	$p_{\rm R}$	0.242	0.992
triaxial	$p_{\rm Q}$	0.174	0.384
	$p_{\rm R}$	0.132	0.944

where $A_i^{\rm m}$ is the measured 1.2-Jy amplitude at smoothing scale i, and $\overline{A}_i$ and σ_i give the mean and dispersion of one of the two mock ensembles. Because the survey volumes corresponding to different smoothing scales are not independent, the quantity Q cannot be expected to follow a χ^2 distribution. Instead, we calibrate its distribution with the mock ensembles themselves. In particular, we compute the fraction $p_{\rm Q}$ of mock catalogues that give a higher value for Q than 1.2-Jy. This quantity can be interpreted as an exclusion probability.

The above test makes no assumption about the run of genus amplitude with smoothing scale. We can further strengthen the constraints by assuming that the power spectrum of the 1.2-Jy survey is reasonably close to a CDM spectrum. Then the measured genus amplitudes should lead to an approximately straight line as well. Using linear regression we therefore fit a straight line to the measured amplitudes in the $(2\pi)^3\lambda^3 N$–λ diagram and consider the distribution of the amplitude of this fit at the intermediate smoothing scale of $\lambda = 12.5\,h^{-1}$Mpc. Again we calibrate this statistic with the mock catalogues. In particular, we compute for the measured 1.2-Jy value the fraction $p_{\rm R}$ of mock catalogues that deviate more from the mean than 1.2-Jy itself.

Table 5 summarizes the results of these tests. As already suggested by figure 21 the fixed smoothing technique clearly favours the ΛCDM model, and rules out the SCDM model with high significance. In fact, the combined genus test excludes the SCDM model at a 99 per cent confidence level.

The adaptive smoothing techniques, however, disappoint the hopes for stronger constraints. While they also clearly favour the ΛCDM model, they cannot exclude SCDM with reasonable significance. This is mainly due to 1.2-Jy measurements that are quite high at intermediate smoothing scales and which fit the SCDM model better than ΛCDM in this regime. This may well be a fluctuation due to cosmic variance. We also want to argue that this should not be viewed as a failure of the adaptive smoothing techniques. As we have shown, adaptive smoothing does increase the signal-to-noise ratio of the measured genus curves. Hence it is able to measure more properties of the examined density fields. The results we obtain just mean that these additional properties of the 1.2-Jy density field agree with both models and are not able to discriminate more strongly between them.

In summary the genus statistic of the 1.2-Jy survey is consistent with the ΛCDM model, while the SCDM cosmology is ruled out with high significance. This is in accordance with the expectation that a model with power spectrum shape $\Gamma \approx 0.2$ should do substantially better than SCDM, since the observed galaxy distribution has been repeatedly shown to exhibit more large-scale power than SCDM.

9 CONCLUSIONS

In this work we have examined the genus statistics of N-body simulations down to the smallest scales examined so far. With the conventional smoothing technique of a spatially fixed kernel we showed that the genus curves of CDM cosmologies retain their random phase shape far into the non-linear regime. However, the genus amplitude is strongly reduced by phase correlations in the density field on scales below $10\,h^{-1}$Mpc. At $2\,h^{-1}$Mpc the suppression reaches a factor of 4. While it is not obvious at the moment how this amplitude drop is related to more traditional measures of higher order correlations, it might be an interesting quantity to characterize non-linearity in future investigations. The genus in the fixed smoothing regime fails to show strong differences between the τCDM, ΛCDM, and OCDM models. This suggests that the genus is only sensitive to the shape of the power spectrum in this regime.

We have shown that an adaptive smoothing is required to use the smallest resolved mass scales of the Virgo simulations. Because of that we have computed genus curves with a novel adaptive smoothing technique for the fully sampled simulations. With the adaptive scheme we can clearly separate the four models at the smallest scales we examined, i.e. the 'degeneracy' between the three models τCDM, ΛCDM, and OCDM can be lifted. In addition, on these scales the genus statistics show very strong departures from 'quasi-Gaussian' behaviour.

We have also performed a large Monte-Carlo experiment in order to establish the statistical properties of the genus statistic when it is applied to a redshift survey like the *IRAS* 1.2-Jy catalogue. For this purpose we extracted a large number of 1.2-Jy mock catalogues from the simulations.

We found that the genus statistic of the 1.2-Jy survey is well consistent with the ΛCDM simulation, while the SCDM model is ruled out at a 99 per cent confidence level. We have not explicitly examined τCDM and OCDM, since we expect them to show at most marginal differences from ΛCDM at the resolution of the 1.2-Jy data.

In this work we also proposed two variants of adaptive smoothing techniques for flux limited redshift surveys. We demonstrated that these techniques can improve the signal-to-noise ratio of the measured genus curves. Hence they are able to extract more topological information from a given redshift survey. Using the 1.2-Jy catalogue, however, we could not achieve a stronger discrimination between the SCDM and ΛCDM models. Since adaptive smoothing is sensitive to additional properties of the density field, we conclude that these properties of the 1.2-Jy survey are consistent with both models. We remain convinced that adaptive smoothing will exhibit a clear advantage, if redshift surveys are used that allow a reconstruction of the density field on strongly clustered scales. Since very large redshift surveys like the Sloan survey are underway, this regime will be accessible in the next few years.

Due to its sensitivity to higher order correlations, the genus statistic remains a useful tool to test the random phase hypothesis, and to compare theoretical models with observations. In this work we confirm that the topology of the 1.2-Jy redshift survey is consistent with current models of cold dark matter universes that grow structure out of random phase initial conditions. A particular advantage of the genus is that these results should be largely independent of a possible bias between the galaxy density and the mass density, at least if this bias relation is monotonic.

REFERENCES

Canavezes A., Springel V., Oliver S. J., Rowan-Robinson M., Keeble O., White S. D. M., Saunders W., Efstathiou G., Frenk C., McMahon R. G., Maddox S., Sutherland W., Tadros H., 1997, MNRAS, submitted
Colley W. N., 1997, astro-ph/9612106
Colley W. N., Gott J. R., Park C., 1996, MNRAS, 281, L82
Efstathiou G., Bond J. R., White S. D. M., 1992, MNRAS, 258, 1P
Fisher K. B., Huchra J. P., Strauss M. A., Davis M., Yahil A., Schlegel D., 1995, ApJS, 100, 69
Gott J. R., Melott A. L., Dickinson M., 1986, ApJ, 306, 341
Gott J. R., Miller J., Thuan T. X., Schneider S. E., Weinberg D. H., Gammie C., Polk K., Vogeley M., Jeffrey S., Bhavsar S. P., Melott A. L., Giovanelli R., Haynes M. P., Tully R. B., Hamilton A. J. S., 1989, ApJ, 340, 625
Hamilton A. J. S., Gott J. R., Weinberg D., 1986, ApJ, 309,1
Hernquist L., Katz N., 1989, ApJ, 70, 419
Jenkins A. R., Frenk C. S., Pearce F. R., Thomas P. A., Colberg J. M., White S. D. M., Couchman H. M. P., Peacock J. A., Efstathiou G. P., Nelson A. H., 1997, submitted, astro-ph/9709010
Kendell M. G., Stuart A., 1973, The advanced theory of statistics. Hafner, New York. Vol. 2, 3rd ed.
Kersher M., Schmalzing J., Buchert T., Wagner H., 1997, astro-ph/9704028
Klypin A. A., 1988, Acta Cosmologica, 15, 101
Luo S., Vishniac E., 1995, ApJ, 96, 429
Matsubara T., 1994, ApJ, 434, L43
Matsubara T., 1996, ApJ, 463, 409
Matsubara T., Suto Y., 1996, ApJ, 460, 51
Matsubara T., Yokoyama, J., 1996, ApJ, 457, 13
Mecke K. R., Buchert T., Wagner H., 1994, A&A, 288, 697
Melott A. L., Dominik K. G., 1993, ApJS, 86, 1
Monaghan J. J., Lattanzio J. C., 1985, Astr. Ap., 149, 135
Moore B., Frenk C. S., Weinberg D. H., Saunders W., Lawrence A., Ellis R. S., Kaiser N., Efstathiou G., Rowan-Robinson M., 1992, MNRAS, 256, 477
Murtagh F., Heck A., 1987, Multivariate Data Analysis. Reidel, Dordrecht
Park C., Gott J. R., da Costa L. N., 1992, ApJ, 392, L51
Park C., Gott J. R., Melott A. L., Karachentsev I. D., 1992, ApJ, 387, 1
Pearce F. R., Couchman H. M. P., 1997, New Astronomy, in press, astro-ph/9703183
Pearson R. C., Coles P., 1995, MNRAS, 272, 231
Protogeros Z. A. M., Weinberg D. H., 1997, astro-ph/9701147
Rhoads J. E., Gott J. R., Postman M., 1994, ApJ, 421, 1

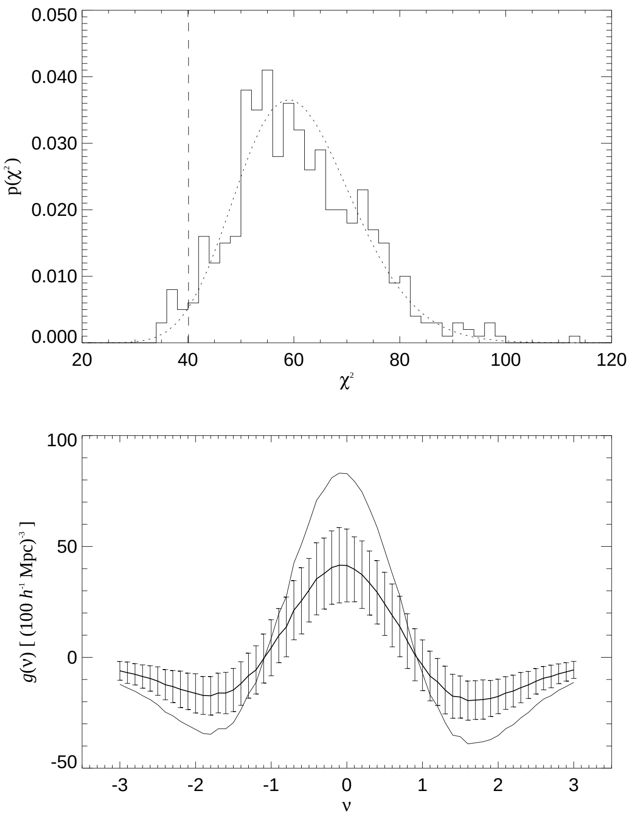

Figure A1. The top panel shows the distribution of χ^2 defined by equation (28) for $k = 61$ measured points of the genus curve (SCDM mock catalogues, fixed smoothing of $7\,h^{-1}$Mpc). The dotted curve is the theoretical χ^2 distribution for 61 degrees of freedom. The dashed line marks the χ^2-value obtained if one checks the fit of a genus curve that is everywhere twice the mean, i.e. $\boldsymbol{g} = 2\overline{\boldsymbol{g}}$. The resulting low value of χ^2 suggests a perfect fit; yet this curve is highly inconsistent with the mock ensemble, as we demonstrate in the bottom panel. Here the thick line gives $\overline{\boldsymbol{g}}$, while the thin line shows $\boldsymbol{g} = 2\overline{\boldsymbol{g}}$. The error bars are the rms deviations of the genus at individual points ν_i of the curve.

Ryden, B. S., 1988, ApJ, 333, L41
Springel V., White S. D. M., 1997, MNRAS, submitted
Strauss M. A., Davis M., 1990, ApJ, 361, 49
Thomas P. A., Couchman H. M. P., 1992, MNRAS, 257, 11
Thomas P. A., Colberg J. M., Couchman H. M. P., Efstathiou G. P., Frenk C. S., Jenkins A. R., Nelson A. H., Hutchings R. M., Peacock J. A., Pearce F. R., White S. D. M., 1997, submitted, astro-ph/9707018
Vogeley M. S., Park C., Geller M. J., Huchra J. P., Gott J. R., 1994, ApJ, 420, 525
Weinberg D. H., Gott J. R., Melott A. L., 1987, ApJ, 321, 2
Weinberg D. H., 1988, PASP, 100, 1373
Yess C., Shandarin S. F., Fisher K. B., 1997, ApJ, 474, 553

APPENDIX A: DISTRIBUTION OF ERRORS

In section 7.1 we defined a multivariate analysis of the measured genus curves using the full 61×61 covariance matrix of equation (26). Here we demonstrate that this method fails in practice, and we examine reasons for this failure.

In figure A1 we show the distribution of χ^2 as defined

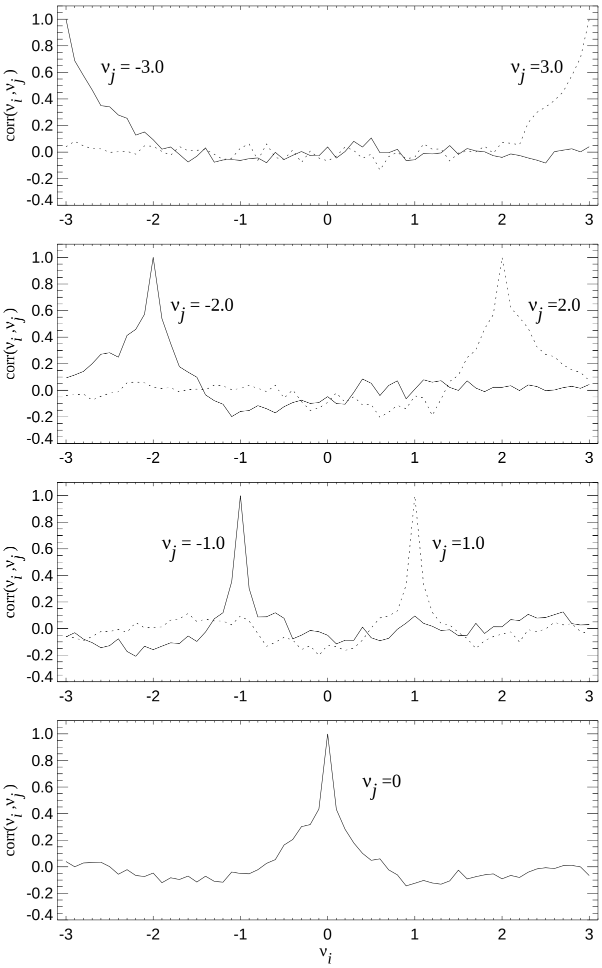

Figure A2. Pearson's correlation coefficient for the genus measurement. The labelled curves show the correlation coefficient for a number of different points on the genus curve. Adjacent points are quite strongly correlated over a range $\Delta\nu \approx 1$. The result shown here is for the SCDM catalogues with fixed smoothing of $5\,h^{-1}$Mpc; it is very similar for all other smoothing scales we consider in this work.

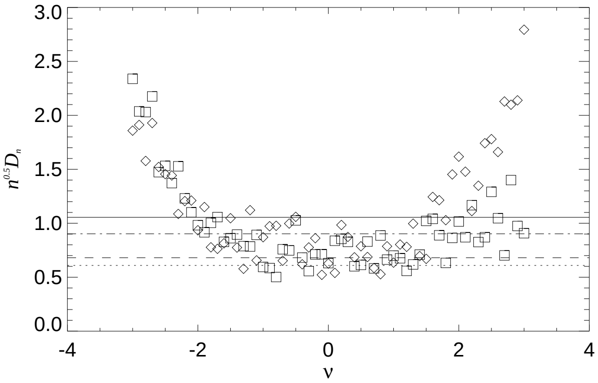

Figure A3. Kolmogorov-Smirnov test of normality of the genus measurements. The symbols give the test measure $\sqrt{n}D_n$ for 61 points on the genus curve of the SCDM mock sample. Diamonds refer to fixed smoothing with $5\,h^{-1}$Mpc, boxes are for adaptive smoothing. The horizontal lines delimit probability regions for various exclusion levels. If the distribution were normal, a point should be found with probability 50% below the lowest line, with 68% below the dashed line, with 95% below the dot-dashed line, and with 99% below the solid line. We obtain a qualitatively similar result for other smoothing scales and mock ensembles.

in equation (28) for the SCDM mock catalogues, smoothed with $\lambda = 7\,h^{-1}$Mpc. The dotted line shows a χ^2 distribution for $k = 61$ degrees of freedom, which is apparently able to fit the observed distribution reasonably well. However, this does not necessarily mean that the distribution of errors is in fact well approximated by a multivariate Gaussian. A first hint that a problem is lurking here may be obtained by computing the χ^2 value for a genus curve with $\boldsymbol{g} = 2\overline{\boldsymbol{g}}$, i.e. one that deviates by 100 per cent from the mean. For this curve the χ^2 comes out as 40.1, suggesting a perfect fit, although the curve $\boldsymbol{g}$ is actually discrepant at individual points on the genus curve with high significance level, as is shown in the bottom panel of figure A1.

Clearly, this peculiar result needs to be understood. In figure A2 we show Pearson's correlation coefficient for a number of places on the genus curve. In general, adjacent points on the genus curve are correlated over roughly a range $\Delta\nu \approx 1$. The main effect of these correlations is to introduce negative values in the inverse V_{ij}^{-1} of the covariance matrix in the elements just off the diagonal. These terms reduce the weight of deviations from the mean if adjacent points exhibit deviations of the same sign.

As a result, coherent deviations (as in $\boldsymbol{g} = 2\overline{\boldsymbol{g}}$) are hardly penalized at all. This partly explains the low χ^2 we obtained for the test curve $\boldsymbol{g} = 2\overline{\boldsymbol{g}}$; yet it does not fully account for why the test fails so badly. We think that there are two reasons for this. First, the statistical properties of the genus errors cannot be fully described by the covariance matrix alone, because the distribution of errors is not consistent with a multivariate Gaussian. This will be shown below. Second, the noise in the covariance matrix compromises its inversion; although the inversion is mathematically possible and stable, the result is not necessarily meaningful, because it is strongly affected by the noise. The principal components analysis developed in section 7.2 provides a solution to this problem.

We now demonstrate that the distribution of the genus measurements is in fact not a multivariate Gaussian. It is sufficient to show that even the distribution of just the genus at a single value of ν_i is not normal. For this purpose we employ a Kolmogorov-Smirnov (KS) test. Based on the $n = 500$ genus measurements for a mock ensemble we can estimate the mean and the variance at individual points on the genus curve from the sample itself. We can then evaluate the KS test

$$D_n = \sup |S_n(g) - F(g)|\,, \tag{A1}$$

where $S_n(g)$ denotes the cumulative distribution function of the measurements and $F(g)$ is the cumulative probability distribution function of the presumed Gaussian.

In general the KS test is fully distribution-free only if the test distribution F is known beforehand. However, here we estimate the parameters of the normal distribution from the sample itself. Because mean and variance are only scale factors of the normal distribution, the KS test remains applicable, although the distribution of D_n is changed (Kendell & Stuart 1973). We calibrate the latter with a Monte-Carlo experiment.

In figure A3 we show measurements of $\sqrt{n}D_n$ for the SCDM mock ensemble, smoothed at $5\,h^{-1}$Mpc with the spherical and, alternatively, the triaxially adaptive method. If the distribution of the g_i were Gaussian, in half of the cases we should measure $\sqrt{n}D_n > 0.609$, and in only 1 per cent of the cases $\sqrt{n}D_n > 1.055$. Hence the KS test shows that the genus distribution is inconsistent with a Gaussian at low and high values of ν. Close to $\nu = 0$ the points are consistent with a normal distribution, although there seems to be a lack of points with low values of $\sqrt{n}D_n$.

We think that it is not too surprising that the distribution of errors is not well described by a multivariate Gaussian. For example, sampling fluctuations can affect the whole volume-fraction/density-threshold relation and thus result in coherent shifts of parts of the genus curve along the x-axis. Similarly, the small survey volumes examined here can lead to large irregular fluctuations in the genus due to particular density features of the patch under examination. As a result, we expect that the genus curve exhibits complicated higher order correlations which make it rather difficult to take full advantage of its information content.