

Calculating the Topology of Large-Scale Structure

R. J. Massey¹

¹*Department of Physics, University of Durham, South Road, Durham DH1 3LE*

Submitted: 3 May 2000

ABSTRACT

An algorithm has been developed to measure the topology (genus statistic) of the galaxy distribution in mock catalogues designed to mimic upcoming results from the 2-degree field (2dF) redshift survey at the Anglo-Australian Telescope. The precision with which this will be possible with the 2dF survey depends upon the smoothing length but is around 6% of the maximum genus. Sparse sampling of galaxies has been identified as a major source of systematic errors. The results from a flat CDM model show non-linear gravitational evolution on small scales. However, for smoothing lengths above $\sim 10 h^{-1}$ Mpc, they are consistent to a 95% confidence level with a random-phase hypothesis. Deviations from this are measured by several ‘metastatistics’: the width, shift and amplitude drop of the genus curve compared to a Gaussian field with the same power spectrum.

Keywords: methods: statistical – galaxies: clustering – cosmology: theory – large-scale structure of Universe

ACKNOWLEDGEMENTS

Many thanks go to Shaun Cole for supervising this work. Thankyou also for valuable suggestions or helpful discussions with Andrew Benson, Carlton Baugh, Subir Sarkar and Tim Hawarden. The mock 2dF catalogues were supplied by Shaun Cole and the original genus code by Andrew Benson. This work has made extensive use of NASA's ADS and NED databases.

1 CONTENTS

ABSTRACT	1
ACKNOWLEDGEMENTS	1
1 CONTENTS	2
2 INTRODUCTION.....	3
3 THEORY	5
3.1 BACKGROUND TOPOLOGY	5
3.2 RANDOM-PHASE HYPOTHESIS	8
3.3 ALTERNATIVE TOPOLOGIES	9
3.4 METASTATISTICS	13
4 GENUS ALGORITHM TESTS.....	14
4.1 FINITE GRID CELL RESOLUTION	15
4.2 MEAN GALAXY DENSITY	17
4.3 NON-PERIODIC GRID	19
4.4 BOUNDARY CORRECTIONS	21
4.5 GALAXY SELECTION.....	26
5 MOCK 2dF CATALOGUES.....	27
5.1 PERIODIC CUBES	28
5.2 2dF SOUTH GALACTIC POLE.....	30
6 CONCLUSIONS.....	34
REFERENCES.....	35
A1 FORTRAN 77 GENUS ALGORITHM.....	37

2 INTRODUCTION

The AAT 2dF galaxy redshift survey (GRS) will soon obtain a complete quarter of a million optical spectra of galaxies in two contiguous areas of the sky: one around the South galactic pole and one near the North galactic pole. Galaxies will be mapped out in spatial and velocity distributions according to their luminosity and spectral type. This will finally enable cosmologists to investigate the dynamical and star-formation history of galaxies and to quantify the forms and shapes in large-scale structure. Previous surveys have only observed a projection onto the sky, like the APM catalogue (Maddox *et al.* 1990), or measured redshift in small and disjoint slices of declination like the Durham/UKST redshift survey (Radcliffe *et al.* 1996). These separate slices of space are very useful for calculating the power spectrum via a two-point correlation function, but tracings of very large structure become lost in the gaps between data. Not even the CfA redshift survey (Huchra *et al.* 1992) yielded significant data because the surveys were only the size of the structure itself, which can therefore vary greatly from one volume to the next. It is even still unknown (Springel *et al.* 1998) whether the overall galaxy distribution is best described as filamentary, wall-like or cellular. A survey of large, contiguous regions like the 2dF GRS can hope to resolve this long-standing debate and also to constrain cosmological models of galaxy formation.

In the standard cosmological model, large-scale structure arises from primordial zero-point energy-density fluctuations during inflation. These fluctuations have random phases or a *Gaussian* density distribution; which has a known topology. During inflation, the fluctuations grow beyond the scale of quantum effects and for the next 10 billion years evolve just according to gravity, and seed galaxy formation. On small scales around over-dense regions, this gravitational evolution is non-linear: matter densities increase rapidly through accretion, phases become correlated, and the Gaussian distribution is destroyed. However, on scales larger than several times the mass correlation length (such that the local average density is close to the global average: Weinberg, Gott & Melott 1983) accretion is less rapid and the topology of the dark matter distribution is maintained. After smoothing over a suitably large volume and assuming that light faithfully traces mass, redshift surveys of present-day galaxies map out a 3D scalar field of density fluctuations which have evolved from their inflationary precursors in a simple way. This is true even accounting for biased galaxy formation. Any linear transformation of the density field, which would distort the power spectrum, still preserves the ordering of densities in the luminous

matter distribution – from which we can recover the topology of the underlying dark matter unhindered (Springel *et al.* 1998).

Several theoretical alternatives exist to predict the evolution of density fluctuations beside the standard cosmological model. Each proposes a different mechanism for seeding the galaxies we observe today, all or none of which may have made a significant contribution in the real universe. Bottom-up scenarios from zero-point density fluctuations produce a hierarchical clustering of galaxies, but only in a universe dominated by weakly interacting, cold dark matter. If the dark matter is instead mainly relativistic (hot) particles, then small-scale fluctuations are quickly smeared out and very large structures form first. These are themselves distributed as a Gaussian field but later fragment via a top-down process into individual galaxies with filamentary or cellular topologies. Even more esoteric possibilities for seeding galaxies such as monopoles, cosmic strings and textures, or the *energetic* explosions of supernovae, would produce completely different large-scale structure and very non-Gaussian density fields today.

Drawing isodensity contours through the galaxy density field yields a family of surfaces in \mathfrak{R}^3 , whose properties can be studied using the four Minkowski functionals, as introduced by Mecke *et al.* (1994). Together, these give a complete quantitative classification of the contour's morphology and topology (Hadwiger 1957), as a function of the density ν at which it is drawn. They are the volume; the surface area; the integrated mean curvature over the surface and, most importantly, the Euler-Poincaré characteristic $\chi_E(\nu)$ which represents its connectivity. All are analytically calculable for a random-phase density field (Schmalzing & Buchert 1997), and predictions within alternative cosmologies can be obtained via n -body simulations or techniques of semi-analytic and semi-numeric models of galaxy formation. Colley *et al.* (1999) present suggestive evidence that these evolved density fields would be sufficiently distinguishable for the larger Sloan Digital Sky Survey (SDSS) to be able to constrain or rule out some of the speculated processes. It will be interesting to repeat their analysis and find the constraints possible from the more complicated geometry of the 2dF GRS.

This paper is structured in the following way. Background theory from topology and from cosmology is outlined in section 3. Section 4 deals with tests (on a Gaussian density field) of the genus algorithm, to ensure its validity and realise its limitations. The algorithm is then applied to mock 2dF catalogues and the results discussed in section 5. Conclusions are drawn in section 6.

3 Theory

3.1 BACKGROUND TOPOLOGY

The mathematical definition of a surface is the local requirement that a small rubber sheet (onto which may be drawn \mathfrak{R}^2 co-ordinate axes) may be stretched and placed anywhere onto the surface, in the same way as a patch is placed on a bicycle tyre. The stretching and bending are done in the external, embedding space; here \mathfrak{R}^3 . If one of the co-ordinate axes on the sheet is bent down and the other up, we get a ‘Pringle’.

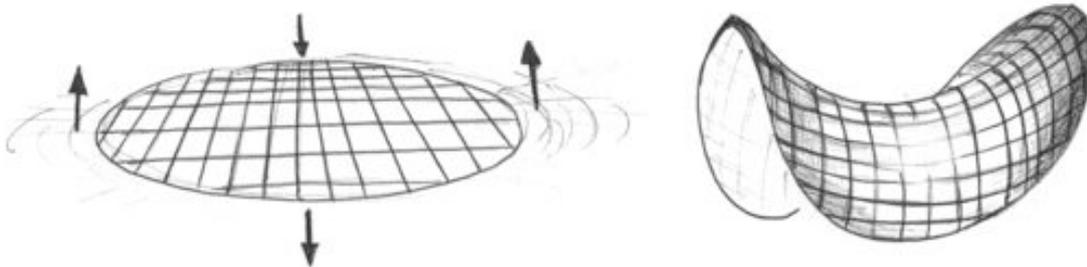


Figure 1. *Hyperbolic point: $\kappa < 0$*

The directions with the minimum or maximum amount of bending or *normal curvature*² κ_n are known as the principal directions. In this case, the principal directions are along the two axes. The mean curvature H is defined at each point to be the average of the two values of κ_n in the principal directions. The Gaussian curvature κ is defined to be their product. Since the principal directions here bend in opposite ways, their normal curvatures have opposite signs and the product of these is $\kappa < 0$. This is a hyperbolic point.

If the local rubber patch is stuck onto something like part of a sphere, it has to bend the same amount in all directions. Since the normal curvature is the same in any direction $\kappa_n^{\min} = \kappa_n^{\max}$, any directions can be called the principal directions and $\kappa = (\kappa_n^{\min})^2 > 0$. This is an elliptic point.

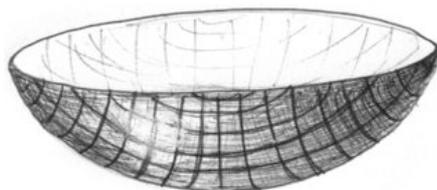


Figure 2. *Elliptic point: $\kappa > 0$*

² κ_n is the reciprocal of the 2nd derivative of position in \mathfrak{R}^3 . On a sphere, $\kappa_n = 1/r$. This is the component of the total curvature in the direction of the normal vector to S – or that which is necessary purely for a curve to stay on S .

κ is a local quantity which can be defined everywhere by properties of the surface at that point alone. But the Gauss-Bonnet theorem in differential geometry makes an incredibly powerful and very impressive link from these local properties to a global statement about the entire surface, S

$$\iint_S \kappa dA + \oint_{\partial S} \kappa_{geo}(\lambda) d\lambda + \sum_{\text{exterior angles of } \partial S} \theta = 2\pi\chi_E(v) \quad (1)$$

where χ_E is the Euler-Poincaré characteristic. χ_E represents an aspect of the curvature across the entire surface in just one number. For a compact surface (continuous and finite in extent) there is no boundary, so the Gauss-Bonnet theorem reduces to

$$\iint_S \kappa dA = 2\pi\chi_E(v) \quad (2)$$

For example, a sphere has everywhere positive curvature κ , so $\chi_E > 0$. It is easy to check that in fact $\kappa = 1/r^2$ and hence $\chi_E = 2$. Integrating over two spheres next to each other gives a total $\chi_E = 4$. A torus consists of elliptical points with $\kappa > 0$ on the outside half, and hyperbolic points with $\kappa < 0$ on the inside half. When the integral is done, these balance out to give $\chi_E = 0$. A double-holed torus has extra hyperbolic points and consequently $\chi_E = -2$. Furthermore, these results apply not only to simple shapes but also to anything *homeomorphic* to them. That is elastically deformable, like a wineglass is to a sphere, a teacup is to a torus and a pair of scissors is to a double-torus.

It is useful to simplify the numbers by classifying shapes by their *genus*

$$g(v) := 1 - \frac{\chi_E(v)}{2} \quad (3)$$

The extra 1 is included following a mathematical tradition. This is generally negligible compared to a measured genus ~ 1000 in a density field and, when definitions vary, it is often omitted. For a single surface, the genus as defined here is just the number of ‘handles’. A wineglass has genus zero, a teacup has genus one and scissors have genus two, etc. For several disjoint surfaces, it must be remembered that χ_E is additive. In a field made of many separate regions, as an isodensity contour may easily be, the genus is a measure of its connectivity

$$g(v) = 1 - (\text{number of isolated regions}) + (\text{number of holes}) \quad (4)$$

Weinberg, Gott & Melott (1987) rather aptly describe a field with negative genus as having a “meatball topology” and a field with a positive genus as having a “swiss-cheese topology”.

A theorem by Poincaré in the related field of integral topology makes it possible to calculate this genus for any arbitrary surface. He used a network of triangles to cover surfaces and form a lattice within them. A complicated proof (see Armstrong 1983) then gives a simple relation between the genus and the number of triangles, vertices and free edges in this lattice. For computer algorithms, it is easier to use a cubic grid, but the principle is no different. Joining up all points contained inside a surface and counting the number of cubes, squares, lines and points,

$$g(v) = (\text{cubes}) - (\text{squares}) + (\text{lines}) - (\text{points}) + 1 \tag{5}$$

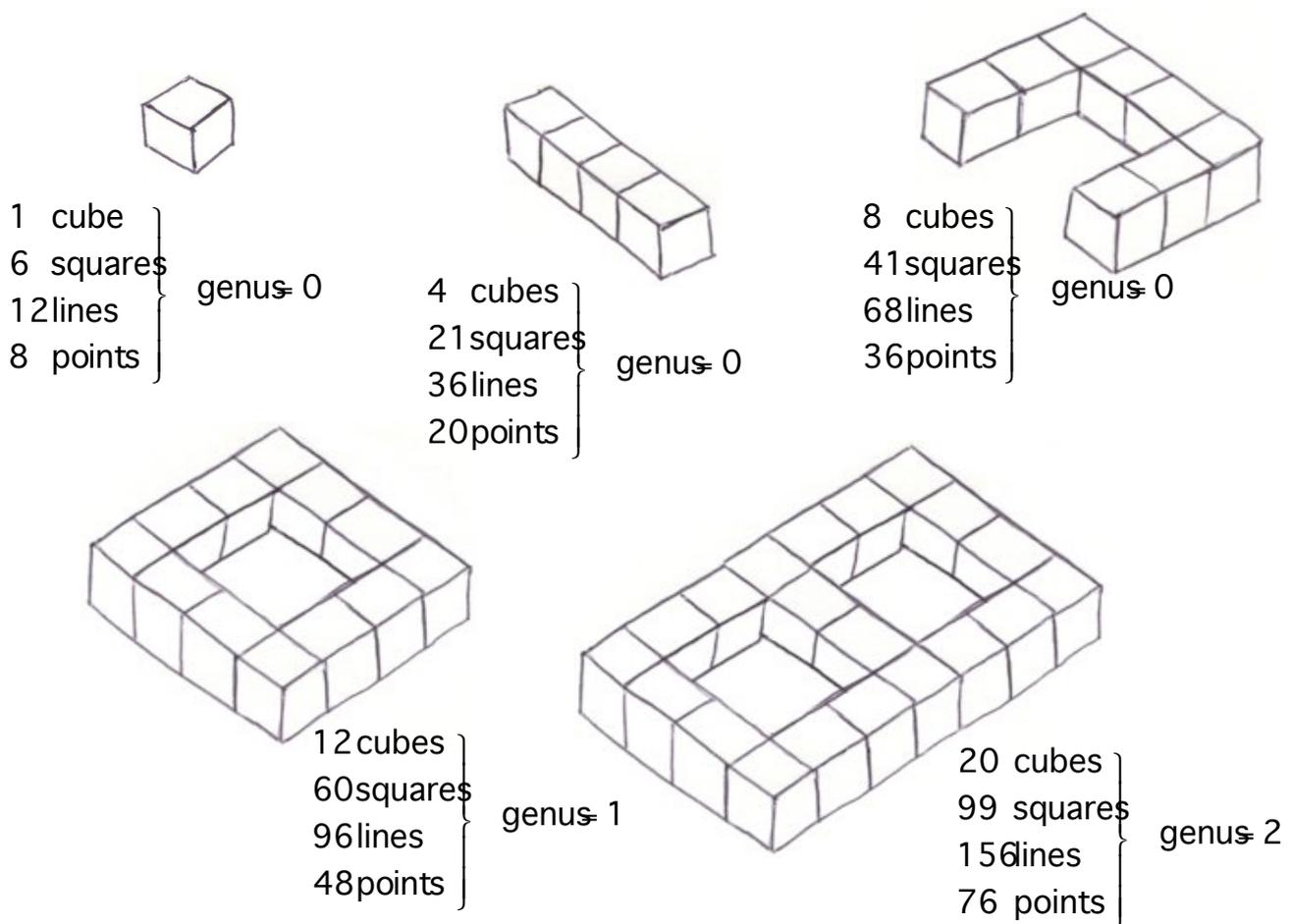


Figure 3. Simple worked examples to demonstrate Poisson's theorem.

The power spectrum of a scalar field contains information about the amplitudes of its Fourier transform components. The genus curve $g(v)$ is complementary to the power spectrum and contains information about their phases. Hence, both measurements can be usefully combined.

3.2 RANDOM-PHASE HYPOTHESIS

In the standard cosmological model, the phases of the Fourier transform density fluctuations are random because they arise via (stochastic) zero point quantum energy density fluctuations during inflation. This leads to a Gaussian distribution of densities in the universe once the field has been smoothed. Data from mock 2dF catalogues and from n -body simulations will throughout be compared to an artificially *Gaussianized* field. This can be manufactured either by randomising the phases of its own Fourier transform; or by starting instead with many random (Poisson) galaxy co-ordinates which also become Gaussian in the limit of a large number of points. By varying the number of points in this latter method, it is also possible to investigate the effects of survey sensitivity and completeness. Random-phase fields are important for the standard cosmological model and because their genus is known analytically. Doroshkevich (1970) shows that for isocontours of density v in dimensionless units of standard deviations away from the mean³, the genus curve of a Gaussian field always has the same shape (see figure 5)

$$g(v) = 1 + A(1 - v^2) \exp\left\{-\frac{v^2}{2}\right\} . \quad (6)$$

The genus reaches a maximum at the mean density $v = 0$ when the contour is multiply connected and becomes negative for $|v| > 1$ as the field separates into isolated regions about individual clusters or voids. The amplitude A is a positive constant determined by the slope of the power spectrum (Weinberg, Gott & Melott 1987)

$$A = \frac{V_{\text{survey}}}{4\pi^2} \left(\frac{\langle k^2 \rangle}{3} \right)^{3/2} , \quad (7)$$

where

$$\langle k^2 \rangle = \frac{\int k^2 P(k) d^3k}{\int P(k) d^3k} , \quad (8)$$

³ Normally with this subject, it is necessary to talk of the density contrast, $(\rho - \langle \rho \rangle) / \langle \rho \rangle$ rather than just the density.

which can be evaluated in general by numerical integration over the power spectrum. Our genus algorithm (see appendix A1) uses a Gaussian window function to first smooth the field.

$$W(k) = \left(\frac{r_{\text{sm}}}{\pi^2} \right) \exp \left\{ \frac{-k^2 r_{\text{sm}}^2}{4} \right\} . \quad (9)$$

The smoothing length⁴ r_{sm} picks out a distance scale at which to quantify the large-scale structure. Small values could be used to isolate individual regions of non-linear galaxy formation for investigation, but this study is interested in large formations with a typical scale (and therefore r_{sm}) generally above $\sim 4h^{-1}$ Mpc. In the case of a simple power law power spectrum $P(k) \propto k^n$ smoothed by convolution with the window function of equation (9), Weinberg, Gott and Melott (1987) show that

$$A = \frac{V_{\text{survey}}}{4\pi^2 r_{\text{sm}}^3} \left(\frac{3+n}{3} \right)^{3/2} . \quad (10)$$

Qualitatively, a field dominated by small-scale power (with high n or smoothed only by r_{sm} much shorter than the mean galaxy-galaxy separation) is corrugated and pockmarked by fluctuations around individual galaxies. The integral for $\langle k^2 \rangle$ diverges in equation (8) and the amplitude of the genus curve increases. Furthermore, since the smoothing length provides the only characteristic length scale for a Gaussian field, A necessarily scales as $1/r_{\text{sm}}^3$ for a fixed survey volume (see figure 7).

3.3 ALTERNATIVE TOPOLOGIES

The preceding section considered only the standard cosmological model in which galaxy formation is seeded by a bottom-up process of accretion onto over-dense regions created from zero point fluctuations during inflation. These fluctuations survive until the epoch of (re)combination in an arena of weakly interacting cold dark matter (CDM) and result in a hierarchical clustering of large-scale structure. This is the most thoroughly studied scenario, because random-phase models can be characterised purely by their power spectrum and are easier to handle mathematically. It is also

This is a trivial scaling to implement, but the convenient definition of v makes even this unnecessary.

⁴ The FWHM has been scaled up from a normal Gaussian by a factor of $\sqrt{2}$ to follow a tradition in the literature.

the most intuitive (and likely?), but not the only model. Other possible seeds of galaxy formation have been suggested from particle physics or from astrophysics. These include the same density fluctuations but immersed in different dark matter; or instead more esoteric topological defects and *energetic* supernovae explosions from a first generation of stars in a chain reaction sweeping across the early universe. These all manufacture large-scale structure with different topologies today and it is the task of the 2dF GRS to constrain these speculative models via departures from Gaussianity in the observed density field.

A species of neutrino with a mass of a few tens of electron volts would be sufficient to account for all the dark matter necessary to gravitationally bind baryons and produce structure in the universe. In the adiabatic hot dark matter (HDM) model, familiar neutrinos replace the unknown and uncertain material of CDM. The term “hot” comes from their relativistic velocities at high redshift $z > z_{\text{eq}}$. High speed free-streaming into large voids and the smoothing effect of radiation pressure on baryons destroys small-scale density fluctuations and maintains coherence at scales up to the Hubble length until z_{eq} . Huge structures with a mass comparable to rich clusters first collapse into proto-walls known as Zel’dovich pancakes, which only then fragment top-down into individual galaxies much later. A local sheet of galaxies known as the fundamental plane has indeed been observed: which does look somewhat like a pancake. Unfortunately, this model faces the problem that peculiar motions of the Local Group are not correlated as if they all emerged from one source, perhaps the Virgo cluster (Kolb & Turner 1990). Their crossing times are also comparable to the Hubble time, suggesting that actually this pancake is being fried up today, by coincidence and out of old galaxies.

The collapse of quantum fields into one of several degrees of freedom during the spontaneous symmetry breaking (SSB) of a phase change can produce more exotic galaxy seeds. All the tension of the field is locked within topological defects such as monopoles or cosmic strings, whose energetic wake can easily be imagined to trigger turbulent instabilities and galaxy formation in a network of filaments or cellular walls respectively. Present-day galaxy distributions may not even trace mass, but rather the passage of these topological instabilities – which makes the theory very flexible and accommodating to observation. Strings are the most acceptable defect (Peacock 1999), created during the SSB of the $U(1)$ abelian Higgs field, and their velocity could determine the mass of the galaxies they seed. Unfortunately, they would presumably also be visible as linear density

steps in the CMBR due to a Doppler shift in front of/behind them, which are not present (Kolb & Turner 1990).

A first generation of AGNs somehow formed before $z \approx 10$ could also shape future structure via non-gravitational forces. Very early supernovae are conjectured to emit shock fronts, which compress the IGM and trigger a chain reaction of more explosions and galaxy formation in outgoing shells. Analysis of the speed of the galactic winds involved successfully predicts the typical galaxy-galaxy separation in the Local Supercluster (Peebles 1993) but fails to explain their peculiar motions. These should all be normal to the fundamental plane but turn out to be mainly within it. This model also fails to explain the apparently self-similar nature of the galaxy distribution. It is not understood (Kolb & Turner 1990) how large-scale formation could be orchestrated by high- λ density waves without destroying the smaller and intermediate structure within it.

Weinberg, Gott & Melott (1987) have calculated genus curves $g(v)$ for simulated density fields representing many different structure models, including the above (see figure 4). However, it is debatable to what scales these would be realistic. All the different mechanisms of galaxy formation are essentially stochastic and produce density fields that are Gaussian to some extent, through the Central Limit Theorem. For example, a network of filaments woven sufficiently tightly may be indistinguishable from hierarchical clustering. However, the fact that the debate is even possible owes itself to the projected ability of the 2dF GRS to measure the topology of large-scale structure with sufficient accuracy to detect any (gross) non-Gaussianity if it *is* present.

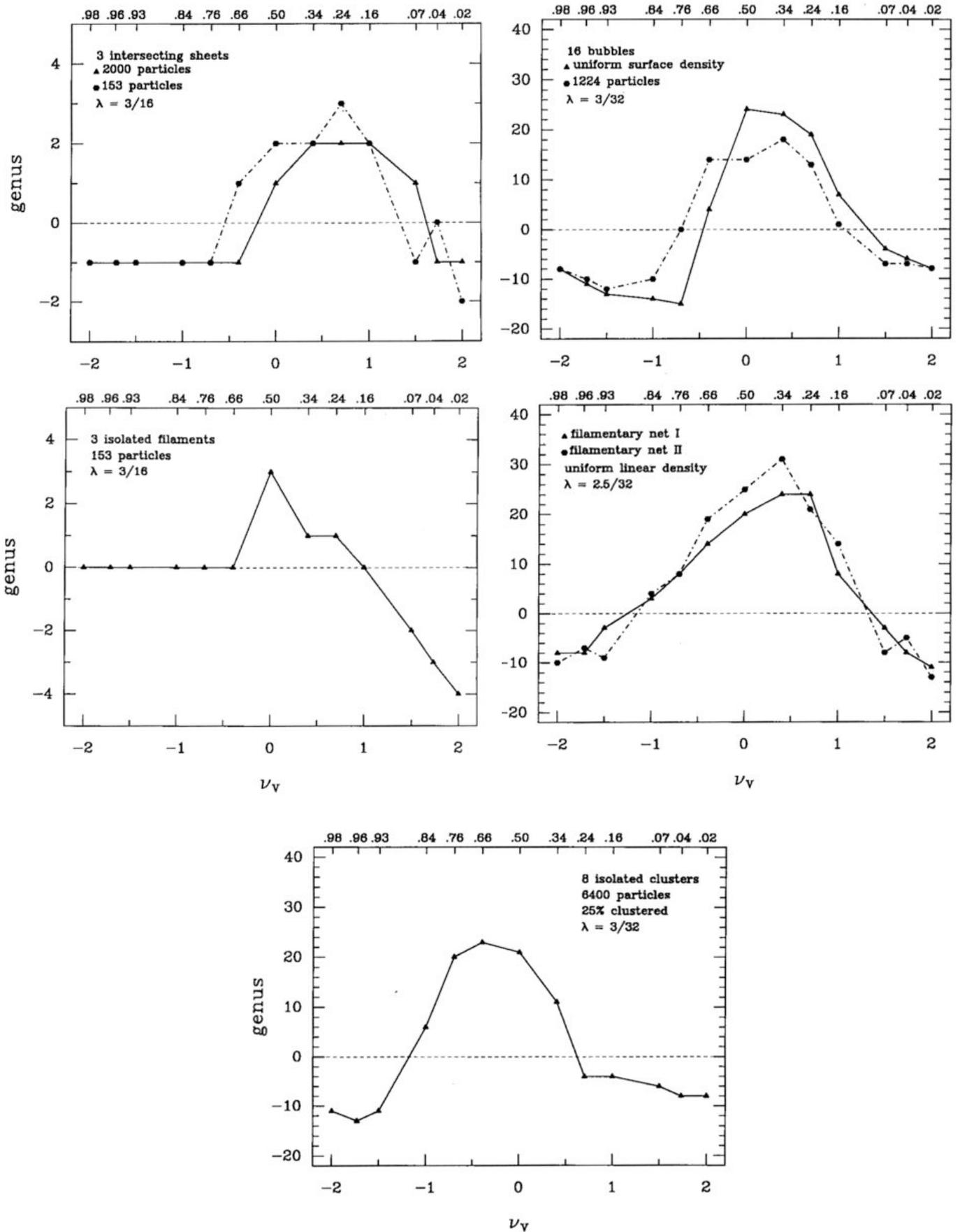


Figure 4. Genus curves of different structure models, from Weinberg, Gott & Melott (1987). All have an asymmetric (non-Gaussian) profile, which would suggest alternative underlying formation processes. Both bubbles and filaments cause a shift to the right, reflecting their “swiss-cheese” topology. Isolated clusters shift $g(\nu)$ to the left, as expected with a “meatball” topology. Notice also the varying widths of the curves.

3.4 METASTATISTICS

In order to measure departures of the genus curve from a random-phase shape using just a few numbers, several “metastatistics” have been suggested by Vogeley *et al.* (1994) and Canavezes *et al.* (1998). Other authors use Principal Component Analysis instead.

3.4.1 Shift

A displacement Δv of the genus curve left or right shows a shift towards respectively a “meatball” or a “swiss-cheese” topology. $g_{\text{obs}}(v)$ is the observed genus curve and if $g_{\text{fit}}(v)$ is the corresponding best fit random-phase analytic curve which minimises χ^2 in the range $-1 < v < 1$, then

$$\Delta v = \frac{\int_{-1}^1 v g_{\text{obs}}(v) dv}{\int_{-1}^1 g_{\text{fit}}(v) dv} . \quad (11)$$

3.4.2 Width

The width W over which the genus curve is positive. This reflects whether the observed density field is more or less sponge-like than a Gaussian random field. If v_{\pm} are the zeros of the observed density field,

$$W = v_{+} - v_{-} . \quad (12)$$

3.4.3 Amplitude

The amplitude A of the best-fitting random-phase genus curve, minimising χ^2 in the range $-1 < v < 1$. This is related to the shape of the power spectrum as described by equations (7) and (10). Note that this topological method of evaluating the amplitude is particularly affected by shot noise due to undersampling of the density field which causes the integrals in equation (8) to diverge and a systematic overestimation of A (see section 4.2).

3.4.4 Amplitude drop

The amplitude of the genus curve is analytically calculable only in the case of a Gaussian density field. In structured fields or after non-linear gravitational evolution, correlations between phases lower the amplitude compared to an estimate based on integration over the power spectrum (Canavezes *et al.* 1998). They define the amplitude drop

$$R = \frac{A}{A_{\text{RP}}} , \quad (13)$$

where A_{RP} is the amplitude of a field with the same power spectrum but random phases. Canavezes *et al.* (1998) discuss several methods to practically achieve this *Gaussianization*; the algorithm used here simply randomises the phases of the field in Fourier space and then transforms back to real space.

4 GENUS ALGORITHM TESTS

Topology is the mathematics of surfaces – and a surface must first be extracted from a list of galaxy co-ordinates. Mecke *et al.* (1994) have one idea to assign a spherical shell of varying radius to each galaxy position and consider the union of these spheres. However, our algorithm employs the more widely used method of placing galaxies into a grid lattice by means of a *cloud-in-cell* routine, smoothing the field with a Gaussian window of varying FWHM r_{sm} , and considering the family of isodensity contours. Adjacent cells inside the surface are then linked together and the number of cubes, squares, lines and points formed by these joins are counted (section 3.1) to calculate the genus by Poincaré’s theorem (5). A listing of the Fortran 77 code used for this is supplied in appendix A1.

Any one individual isodensity contour has two sides and can actually be treated as giving two different surfaces: the incursion and the excursion surfaces. The normal to the incursion surface points into regions of lower density (the excursion set) and the normal to the excursion surface points into the incursion set. N -body simulations are carried out within periodic grids – in which the incursion and excursion surfaces are identical but oppositely oriented. The two values that can be calculated for the genus, $g_{\text{inc}}(\nu)$ and $g_{\text{exc}}(\nu)$ ought to be the same *in a periodic grid*. However, numerical errors are introduced by both the generation of a surface from individual galaxy co-ordinates and by the reduction of this smooth mathematical object into discrete grid cells, which need not necessarily be periodic. The genus algorithm is now tested, to minimise these errors and to ascertain the régime in which it is a valid approximation.

4.1 FINITE GRID CELL RESOLUTION

The (initially smooth) density field is binned into a lattice of some finite grid cell size, l_{cell} . This falsely clumps galaxies together and smudges out structure that really *is* present. Detail is lost on scales smaller than the bin size. If l_{cell} is large, then the magnitude of the genus curve is systematically underestimated: see figure (5). This deviation asymptotically disappears as more grid cells are used and the individual cell size becomes much smaller than the smoothing length, r_{sm} . A practical upper limit has been set on l_{cell} , at least for random fields.

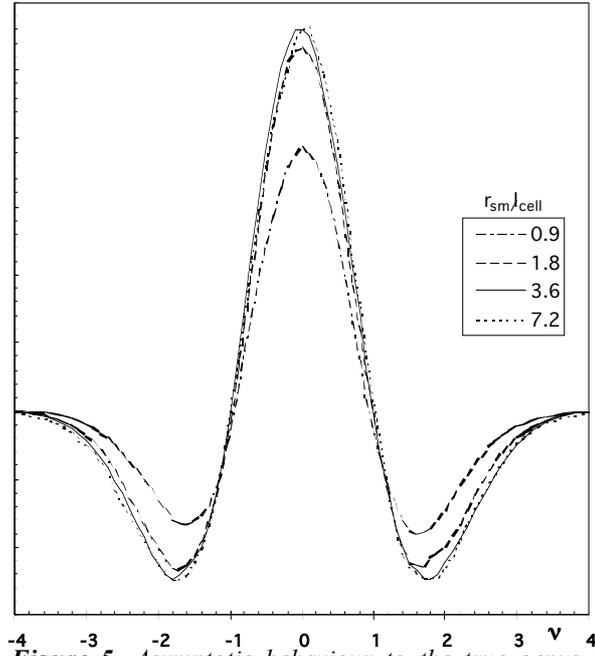


Figure 5. Asymptotic behaviour to the true genus curve with more grid cells. Large cell size results in a systematic underestimation of $g(v)$. Calculated in grids between 64^3 and 512^3 .

Furthermore, using grid cells of comparable size to the smoothing length removes the degeneracy between $g_{\text{inc}}(v)$ and $g_{\text{exc}}(v)$. Adjacent cells are not well correlated and individual concentrations of galaxies can create a pockmarked or corrugated field. $g_{\text{inc}}(v)$ and $g_{\text{exc}}(v)$ differ slightly and artefacts appear in the genus curve (Springel *et al.* 1998) when only diagonally adjacent grid cells delineate the isodensity contour. Figure 6 shows an example.

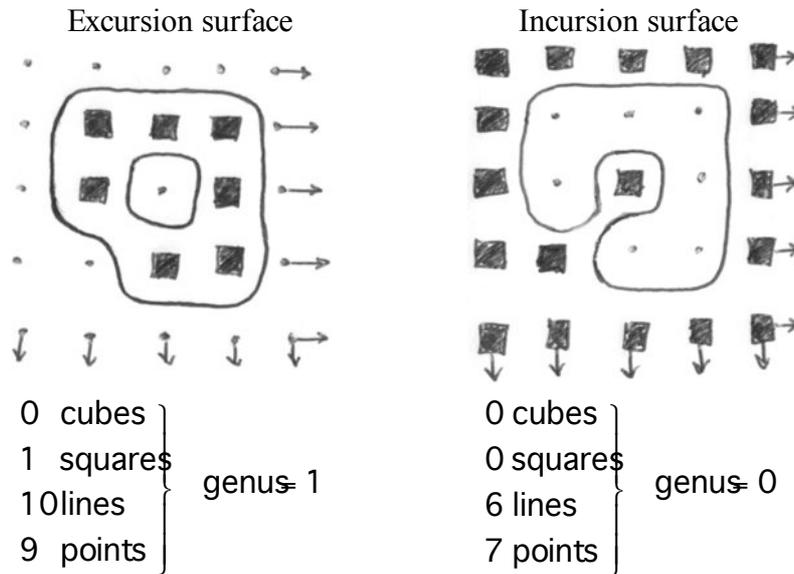


Figure 6. Example of the failure of the genus algorithm to calculate a consistent value for $g(v)$ because of the discrete nature and finite size of grid cells. Genus is calculated using equation (5).

To keep this method in an acceptable régime requires a trade-off between the size of the grid cells and the smoothing length. Using more, smaller grid cells costs processing time. At larger r_{sm} , coherence over several grid cells reduces the number of adjacent diagonals that cause this effect, but (as well as quantifying the structure at a totally different smoothing length) also reduces the amount of independent data. Figure 7 shows genus curves calculated from six iterations of 4 million points distributed randomly in a 128^3 grid and smoothed with increasing r_{sm}/l_{cell} . This field is Gaussian at all scale lengths so $g(v)$ ought to fit a random-phase analytic curve.

The genus curves from the incursion and excursion surfaces are seen to quickly converge to the same value. To reduce processing time these curves were created by holding l_{cell} constant and increasing r_{sm} , so for large r_{sm} there are fewer independent data sets within the cube and larger random errors. Although they tend to the same curve, this is therefore not exactly the random-phase analytic curve for large r_{sm} . As expected, this trend continues to higher r_{sm}/l_{cell} and lessens with a larger ensemble of data sets.

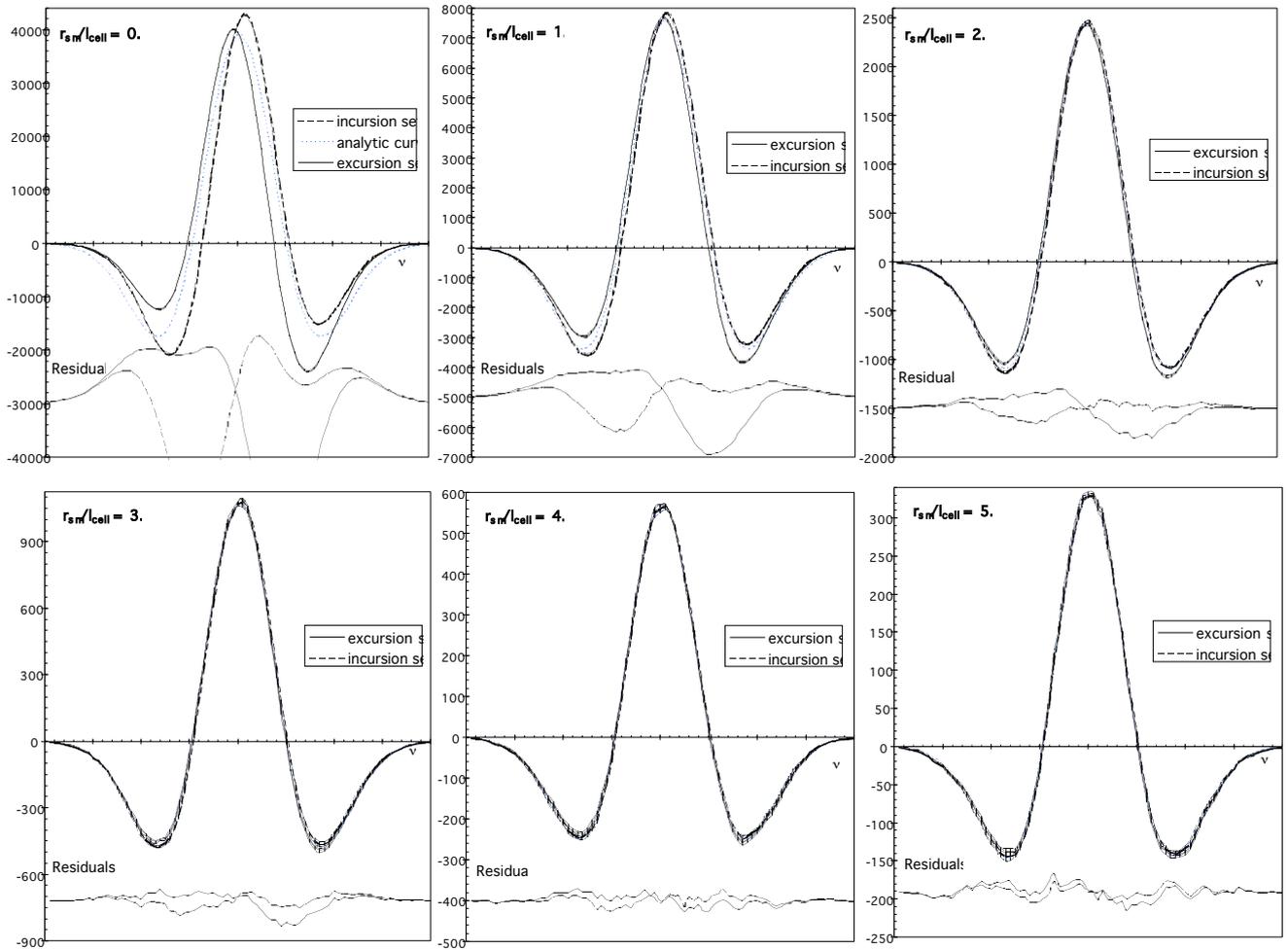


Figure 7. Resolution limit of smoothing length. Error is shown by a hatched area around each curve. Residuals from the best fitting random-phase curve are plotted on an expanded (2x) scale underneath, offset for clarity: finite grid cell effects break the degeneracy between $g_{inc}(v)$ and $g_{exc}(v)$ for small r_{sm}/l_{cell} . Calculated in a fixed 128^3 grid with increasing r_{sm} .

To combat the resolution limit demands a larger $r_{\text{sm}}/l_{\text{cell}}$ than that to ensure no underestimation of the magnitude of the genus, so this alone governs the régime of validity for the algorithm. A quantitative expression for the fractional error e in each curve can be obtained by calculating the rms of the residuals compared to the maximum of the best-fit Gaussian curve. This is shown in figure 8.

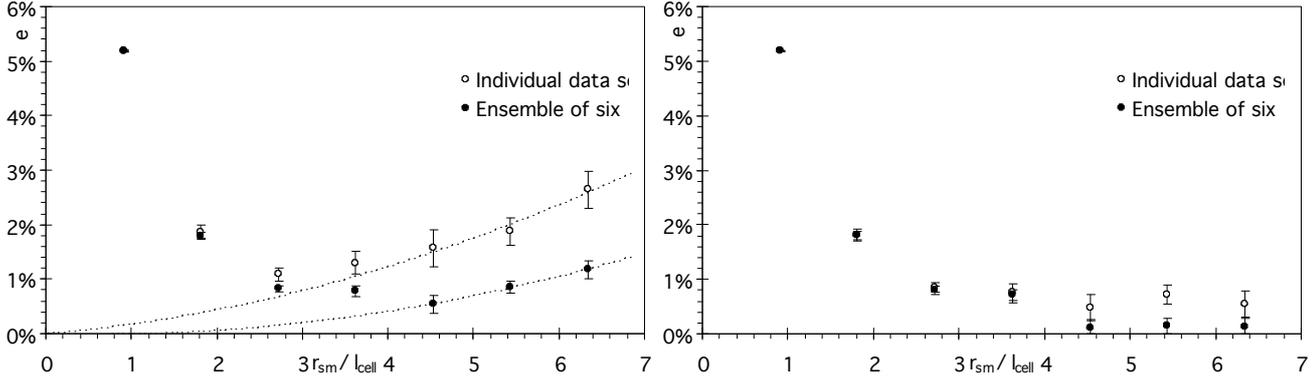


Figure 8. Fractional error due to finite grid cell size. Plot (a) includes a contribution due to the finite size of the sample, which decreases as expected in the averaging of six cubes. This volume effect has then been fitted and removed in quadrature from plot (b).

Springel *et al.* (1998) adopt a minimum value of $r_{\text{sm}}/l_{\text{cell}} \geq 2$; Weinberg, Gott & Melott (1987) require $r_{\text{sm}}/l_{\text{cell}} \geq 2.5$. Clearly this is open to interpretation but of the right order. To keep the contribution to error on the genus curve due to finite grid cell size below 1%, all further calculations will have

$$r_{\text{sm}} \geq 2.5 l_{\text{cell}} \quad (14)$$

4.2 MEAN GALAXY DENSITY

Smoothing a density field alters its power spectrum by introducing a short-wavelength cut-off. Convolved with the window function (9), a power spectrum $P(k) \propto k^n \rightarrow k^n \exp\{-k^2 r_{\text{sm}}^2/4\}$. Smoothing lengths much shorter than the typical galaxy-galaxy separation leave the field corrugated around individual galaxies and the magnitude of the genus curve overestimated, as discussed in section 3.2. Since we are generally keeping $r_{\text{sm}} \sim 4 h^{-1}$ Mpc or larger, the problem is

equivalently one of ensuring that the mean galaxy density $\langle N \rangle$ is sufficiently high that there are always several galaxies per smoothing volume. A sparsely sampled distribution also affects the genus curve in a similar but slightly different manner. As individual galaxy sites are removed, clusters fragment and filaments break up. The isodensity contours separate around new isolated structures and the genus becomes systematically more negative for all densities. The central peak of the genus curve is simultaneously lowered and narrowed by the precise way that $g_{\text{inc}}(\nu)$ and $g_{\text{exc}}(\nu)$ diverge. Under certain conditions, it can even become a double peak as $g_{\text{inc}}(\nu)$ shifts slightly to the right and $g_{\text{exc}}(\nu)$ moves left. This is seen for some smoothing lengths in figure 17.

To accomplish a similar analysis to that in section 4.1, an increasing number of points are randomly distributed in a 128^3 grid and smoothed at $r_{\text{sm}}/l_{\text{cell}} = 4.5$. Fractional errors compared to the maximum genus at $\nu = 0$ are shown in figure 9.

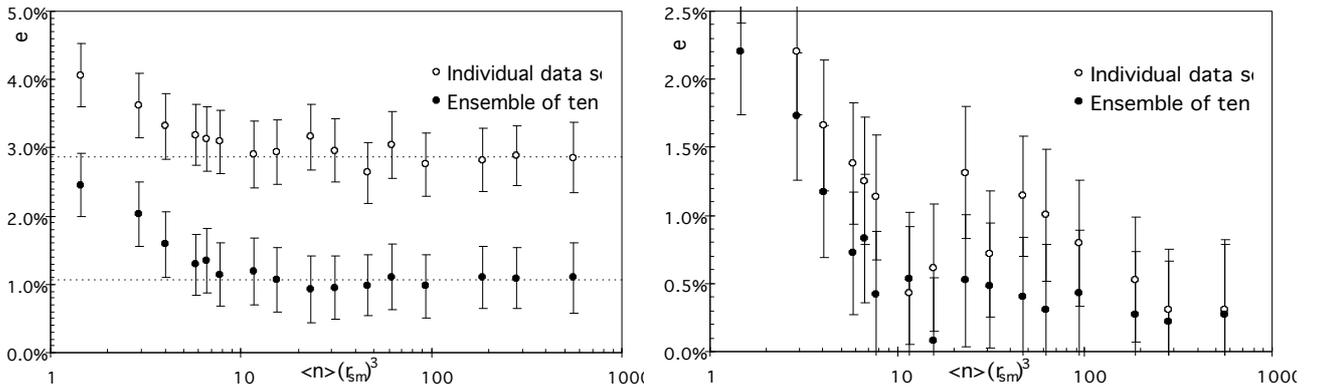


Figure 9. Fractional error due to sparse sampling of galaxy points. Plot (a) includes a contribution due to the finite volume of the sample, which has then been fitted and removed in quadrature from plot (b). Calculated by gradually increasing $\langle N \rangle$ inside a fixed grid.

To keep the contribution to fractional error from sparse sampling below 1%, we find a requirement similar to those of Weinberg, Gott & Melott (1987) $\langle N \rangle r_{\text{sm}}^3 \geq \pi^{3/2} = 5.6$ and Springel *et al.* (1998) $\langle N \rangle r_{\text{sm}}^3 \geq 9.7$

$$\langle N \rangle r_{\text{sm}}^3 \geq 6 \quad (15)$$

This analysis is valid for a white noise power spectrum Poisson random distribution of galaxy sites. The clustering of galaxies in the 2dF sample may make it reasonable to use smaller r_{sm} within the tightly packed and over-dense rich clusters. However, problems occur around very low galaxy

densities in the voids. The algorithm often has difficulty and error bars tend to be large near the under-density minimum at $v \approx -\sqrt{3}$ (see figure 17). Ideally perhaps, an adaptive smoothing algorithm would be designed. However, this has the two disadvantages that adaptive smoothing techniques break the independence of the genus with respect to bias (Springel *et al.* 1998) and also that there is then no analogue of the analytic random-phase genus curve. An analysis of the effect of resampling in clustered density fields has been attempted but proves difficult. Without an analytic genus curve there is no ‘correct’ answer to judge which specific effect is causing an error. Furthermore, there are not as many independent points in the catalogue as there are in a Poisson random field. Seen from large scales, clusters are now the basic building blocks, each containing about $\langle N \rangle \int \rho(\mathbf{r}) d^3\mathbf{r} \sim 10$ individual galaxies. It no longer matters when one or two of these galaxies are randomly removed or shuffled around for bootstrapping: they are still in the same place. Conversely, filaments and sheets consisting of only a few galaxies are very sensitive to changes or missing data points. These easily disintegrate under bootstrapping and the whole interconnected topology breaks down.

The analysis of clustered density fields will use as much data as possible but with a continuing eye to condition (15) and on guard for the kind of global skewing of the genus curve produced by sparse sampling of galaxy points.

4.3 NON-PERIODIC GRID

While remembering the requirements (14) and (15) on grid size and smoothing length, we now drop the periodic boundary conditions. Real data from the 2dF GRS will not wrap around forever in a cubic block, but lie within an (unwieldy) wedge of R.A. and Dec. with a selection function to further complicate the boundary in redshift space. Edge effects are produced in a way that is difficult to predict analytically. In order to be able to compare 2dF data with theory, it is necessary to discard data from unwanted regions of the n -body simulations and imitate these arbitrary boundary conditions.

To achieve this, a mask is produced of the survey region, with a value of unity inside and zero outside. When comparing different theoretical models, this region can simply be one base cube in

the n-body simulations; when the 2dF survey data is introduced this will be the ‘unwieldy’ wedge⁵. Galaxy density data exists inside the mask – then is set to zero in a region of padding around it. Both the density field and the mask are smoothed into this zero padding and the density field is then divided by the mask. This is an attempt to compensate for the distortion of data by averaging with zero padding. However, as densities consequently tend to the mean, voids are closed up and clusters smeared away. Some of the data near the edges of the survey is unreliable and must be thrown away. However, this is obviously not ideal as it lowers the number of independent data points in the survey volume and increases the random errors. There is a trade-off between how much data is discarded and how much systematic error it would introduce, were it left in. This trade-off has been given the usual analysis in figure 10.

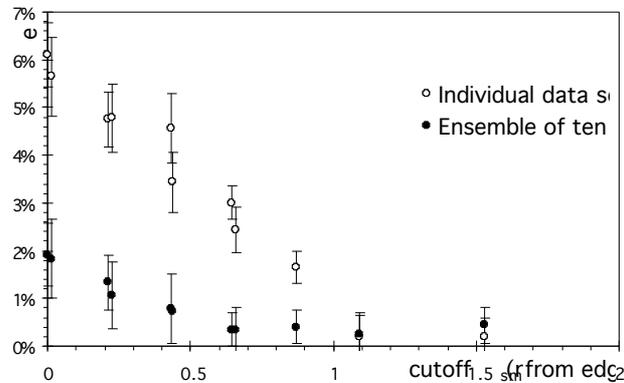


Figure 10. Fractional error due to edge effects when a non-periodic galaxy density field is smoothed into zero padding.

Weinberg, Gott & Melott (1987), working with a slightly different algorithm and the Point Source Catalogue redshift (PSCz) survey, concluded that it was best to ignore data within $\frac{1}{3} r_{sm}$ of the edge of the survey region after smoothing. With the much larger 2dF survey, a slightly more conservative estimate allows a significant reduction in error

$$\text{discard data within } \frac{2}{3} r_{sm} \text{ of the edge of the survey region} \quad (16)$$

This corresponds to the point at which the smoothed mask falls to 80% of its maximum value.

⁵ The mask may also have holes around nearby bright stars (see section 4.5). In redshift space, the survey is truncated into a volume-limited sample extending to the distance at which $\langle N \rangle$ falls below a useful value for criterion (15).

4.4 BOUNDARY CORRECTIONS

As well as corrupting nearby data, inserting a survey boundary breaks even the approximate symmetry between the genus calculated from the incursion and excursion surfaces seen in a periodic grid. Differences between the two arise when the density contour intersects a boundary of the survey region (Coles, Davies & Pearson 1996). For example, a region of high density that wraps from one side of the survey to another is cut in half when periodicity is removed. It now appears as two separate regions in the excursion set, lowering the genus by one. Meanwhile, the incursion set has just suffered two small indents into an otherwise contiguous region: a homeomorphism, which leaves the genus unchanged. The reason for this disparity is hidden in our definition of the genus.

The universe is a continuous medium that extends beyond the boundaries of the survey. The Euler-Poincaré characteristic $\chi_E(\nu)$ of the Gauss-Bonnet theorem (1) has previously been calculated in a boxed-off region, from surfaces which have been artificially closed up if they extend beyond the arbitrary 2dF GRS boundary. This is an unreasonable thing to do. Galaxies certainly exist outside the survey region: and so should the isodensity contours. The only hint of information about the behaviour of the large-scale structure just beyond the boundary of the survey is given by the angle ϕ at which the isosurfaces approach it – and this information should not be ignored. Figure 11 shows two contours that would be artificially capped off by a (cubic) survey boundary to give the same $\chi_E(\nu) = 4\pi$ but are probably edges of larger surfaces with completely different topologies.

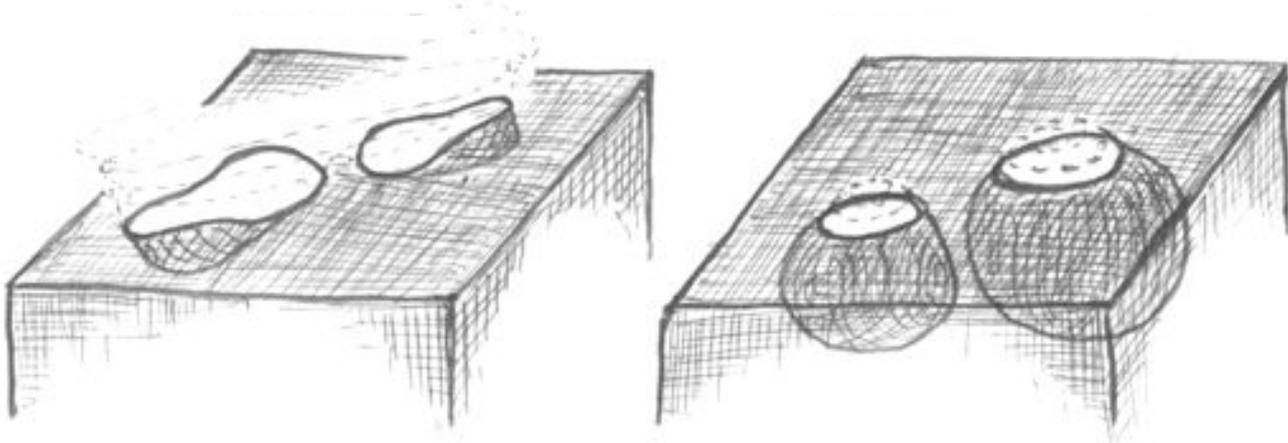


Figure 11. Data is only known within the survey region but galaxies and isodensity contours also exist outside (dotted lines). The boundary corrections in the Gauss-Bonnet theorem destroy any knowledge of this by capping off equally all surfaces which intersect the boundary of the survey region. A cubic survey boundary is illustrated.

What is really needed is not $\chi_E(\mathbf{v})$, but the integrated Gaussian curvature κ over the isodensity surfaces. This is the first term on the LHS of the Gauss-Bonnet theorem (1). The other two terms are integrated boundary corrections. To recover the statistic of real interest in a 2D topology, Davis & Coles (1993) found it necessary to use a complicated and arbitrary weighted averaging of $g_{inc}(\mathbf{v})$ and $g_{exc}(\mathbf{v})$. In 3D, it turns out that the deviation of each of these from the periodic prediction is equal and opposite⁶ and a simple average gives almost exactly the value which would be obtained in the absence of boundaries. ‘Almost’ because on a point-by-point basis, there will still be some slight scatter about the mean (given by the convolution of a cosine and a normal probability curve, as explained below) due to the binning of the surface into finite grid cells. However, over a sufficiently large sample, the total boundary correction tends to zero. This is a fine example of the beauty of Poincaré’s theorem (5).

To show all this, consider the two boundary correction terms in the Gauss-Bonnet theorem (1). If the incursion/excursion surface intersects the boundary of the survey region (the cube in figure 11) in n smooth or *regular*⁷ closed curves $\alpha_i(\square)$ ($i=1\dots n$) parametrised by arclength λ then

$$\oint_{\partial S} \kappa_{geo}(\lambda) d\lambda + \sum_{\substack{\text{exterior angles} \\ \text{of } \partial S}} \theta = \sum_{i=1}^n \oint_{\alpha_i} \kappa_{geo}(\lambda) d\lambda + \iiint_V \frac{\pi}{2} \delta^{(3)}(\text{vertices}) dV, \quad (17)$$

because the curvature at the exterior angles of a cube are compressed into δ -functions at the eight vertices (Canavezes *et al.* 1998). Depending on the density \mathbf{v} , some of these vertices will be included in the excursion set, others in the incursion set. A statistical average suggests that the contribution to $g_{inc}(\mathbf{v})$ from the vertex terms (for any shaped boundary) is typically

$$\sum_{\substack{\text{exterior angles} \\ \text{of } \partial S \text{ (} g_{inc} \text{)}}} \theta = \iiint_{\substack{\text{incursion} \\ \text{set}}} \frac{\pi}{2} \delta^{(3)}(\text{vertices}) dV \approx 2 \frac{V_{inc}(\mathbf{v})}{V_{survey}} = 2 \operatorname{erfc}(\mathbf{v}) \quad (18)$$

and the remaining vertices contribute to $g_{exc}(\mathbf{v})$

$$\sum_{\substack{\text{exterior angles} \\ \text{of } \partial S \text{ (} g_{exc} \text{)}}} \theta \approx 2(1 - \operatorname{erfc}(\mathbf{v})) \quad (19)$$

⁶ It is in fact equal to the 2D genus of the survey boundary from e.g. Davis & Coles (1993), Colley *et al.* (1999).

⁷ A parametrised and differentiable curve $\alpha(\square): I \rightarrow \mathbb{R}^3$ is said to be regular if $\alpha'(\square) \neq 0 \quad \forall \square \in I$, an open set of the reals

These are not equal and opposite but are only a negligible contribution to the genus, compared to that from the integrated curvature. When $g_{\text{inc}}(v)$ and $g_{\text{exc}}(v)$ are combined (see later) this correction will add 1 to the overall genus. Kerscher *et al.* (1997) explicitly keep this term in their analysis because it corresponds to one of the other Minkowski functionals. However, it can be ignored for a similar reason as for the irrelevant 1 in the non-standard definition of genus (3). It does not alter the overall shape of the curve and it is swamped by a typical genus of several thousand.

Now consider the curve $\alpha(\square)$ which lies on the boundary of the survey region. Its geodesic curvature will be shown to create the dominant correction to the genus. $\alpha(\square)$ has a tangent vector

$$\mathbf{t}_{\dot{\alpha}}(\lambda) = \dot{\alpha}'(\lambda) \quad , \quad (20)$$

which is automatically of unit length when $\alpha(\square)$ is parametrised by arclength. It has *normal* and *binormal* vectors (see figure 12) defined by the Serret-Frenet formulae

$$\kappa_{\dot{\alpha}} \hat{\mathbf{n}}_{\dot{\alpha}} = \mathbf{t}'_{\dot{\alpha}} \quad \mathbf{b}_{\dot{\alpha}} = \hat{\mathbf{n}}_{\dot{\alpha}} \wedge \mathbf{t}_{\dot{\alpha}} \quad , \quad (21)$$

where $\kappa_{\alpha}(\square)$ is the extrinsic curvature (seen from the point of view of the embedding space \mathfrak{R}^3). κ_{α} is equal to $1/r_{\text{eff}}$, the radius of a circle tangential to $\alpha(\square)$ and with the same curvature at the point of intersection. The $(\mathbf{b}_{\alpha}, \mathbf{t}_{\alpha})$ plane is called the *rectifying* plane (see figure 12). For a curve on a flat boundary like the cube in figure 11, this simply intersects the boundary at right angles. However, $\alpha(\square)$ lies not only on the survey boundary, but also on the incursion/excursion surfaces. In a co-ordinate basis tied to the surface, $\alpha(\square)$ can be reclassified to have components of curvature κ_n along the unit normal \mathbf{N} to the surface and some extra bending around κ_{geo} within the surface⁸.

⁸ A geodesic has $\kappa_{\text{geo}} = 0$.

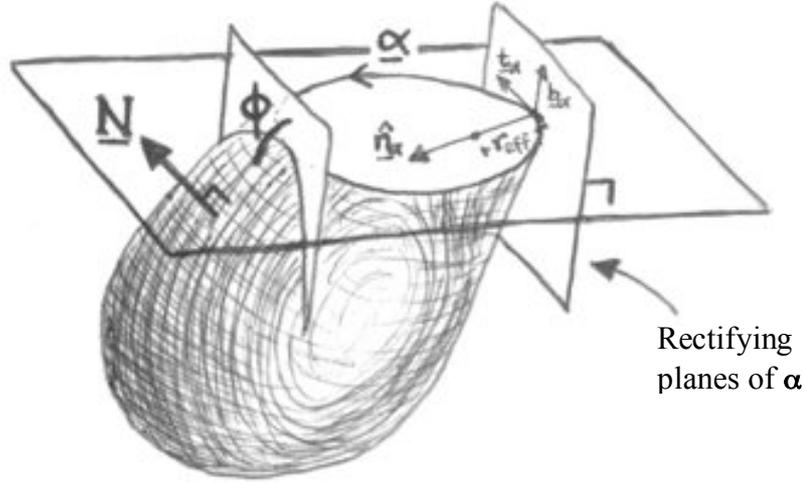


Figure 12. The rectifying planes of $\alpha(\square)$ are at right angles to the survey boundary but at an arbitrary angle $\phi(\square)$ to the incursion/excursion surface. It is this angle which has to be added on to cap off the surfaces and this which is the crucial angle to be removed from the estimation of the genus.

These are defined as

$$\kappa(\lambda)\hat{\mathbf{n}}_{\alpha} = \kappa_{\text{geo}}(\lambda)\mathbf{N} \times \mathbf{t}_{\alpha} + \kappa_n(\lambda)\mathbf{N} \quad . \quad (22)$$

Dotting this with $\mathbf{N} \times \mathbf{t}_{\alpha}$ gives an expression for geodesic curvature

$$\kappa_{\text{geo}}(\lambda) = \kappa(\lambda)\hat{\mathbf{n}}_{\alpha} \cdot (\mathbf{N} \times \mathbf{t}_{\alpha}) \quad (23)$$

but since all of these vectors are of unit length, this reduces to

$$\kappa_{\text{geo}}(\lambda) = \frac{1}{r_{\text{eff}}(\lambda)} \cos(\phi(\lambda)) \quad , \quad (24)$$

where $\phi(\square)$ is the angle between the rectifying plane and the incursion surface. $\alpha(\square)$ can be reparametrised by $\square \in [0, 2\pi)$. Then, with a Jacobian $r_{\text{eff}}(\square)$ from the change of integration variable

$$\begin{aligned} \oint_{\substack{\partial S \\ (g_{\text{inc}})}} \kappa_{\text{geo}}(\lambda) d\lambda &= \sum_{i=1}^n \oint_{\hat{a}_i} \kappa_{\text{geo}}(\lambda) d\lambda = \sum_{i=1}^n \int_{\theta=0}^{2\pi} \frac{1}{r_{\text{eff}}(\theta)} \cos(\phi(\theta)) r_{\text{eff}}(\theta) d\theta \\ &= \sum_{i=1}^n \int_{\theta=0}^{2\pi} \cos(\phi(\theta)) d\theta \quad . \end{aligned} \quad (25)$$

For the excursion surface, $\phi(\square) \square \pi - \phi(\square)$

$$\begin{aligned} \oint_{\substack{\partial S \\ (g_{\text{exc}})}} \kappa_{\text{geo}}(\lambda) d\lambda &= \sum_{i=1}^n \int_{\theta=0}^{2\pi} \cos(\pi - \phi(\theta)) d\theta \\ &= - \oint_{\substack{\partial S \\ (g_{\text{inc}})}} \kappa_{\text{geo}}(\lambda) d\lambda \quad . \end{aligned} \quad (26)$$

The separate corrections from $g_{\text{inc}}(\nu)$ and $g_{\text{exc}}(\nu)$ are equal and opposite, so can be removed by averaging the two values. This is true for any density field: Gaussian or otherwise, and in any shaped boundary. The way forward is open to calculate the ‘true’ genus, which would have been obtained in the absence of boundaries. As an interesting aside (and to verify that this approach is working), these boundary correction terms or *residues* from the ‘true’ genus can be calculated in the limit of large n .

$$\oint_{\partial S} \kappa_{\text{geo}}(\lambda) d\lambda \approx \pm 2\pi n \langle \cos(\phi) \rangle \quad . \quad (27)$$

ϕ will scatter around either 0 or π , depending on the predominant orientation of the incursion/excursion surface at that density. n is proportional to the surface area of this density contour and therefore proportional to the derivative of $V_{\text{inc/exc}}$, which is just a normal distribution curve. Kerscher *et al.* (1997) explicitly use the other three Minkowski functionals as a boundary correction to χ_E . A simple estimate recovers their result by just assuming that the typical length of the intersections between the isodensity contour and the survey boundary is scale-invariant (which is certainly true in the case of white noise power spectrum Poisson random field). Residues are then

$$\oint_{\partial S} \kappa_{\text{geo}}(\lambda) d\lambda \propto \pm \frac{SA_{\text{survey}}}{\sqrt[3]{V_{\text{survey}}}} e^{-\nu/3} (\text{erfc}(\nu) - 1) \quad . \quad (28)$$

This prediction has been plotted in figure 13. Its magnitude could be calculated rigorously or could even be fitted from the data in figure 9 but is unimportant and has just been scaled here to minimise χ^2 . The prediction fits the sample data very successfully.

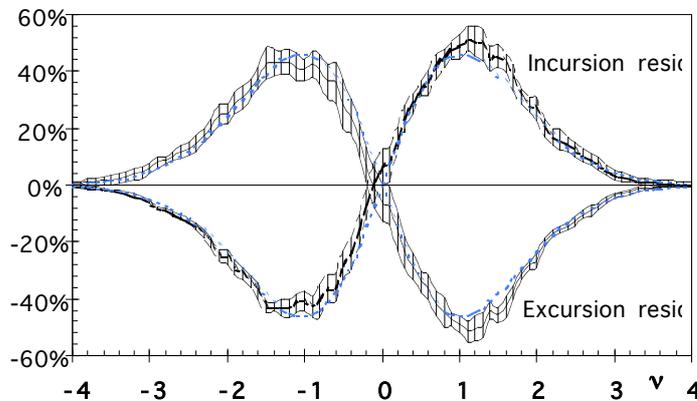


Figure 13. Residuals of $g_{\text{inc}}(\nu)$ and $g_{\text{exc}}(\nu)$ from their mean (assumed to be the ‘true’ genus). Errors are hatched regions around each curve. Data is from a mock 2dF catalogue smoothed at $8h^{-1}$ Mpc, and satisfying the constraints (14), (15) and (16) so that the main source of residual is due to the boundary correction terms. The dotted line is the prediction for this in equation (28), scaled to minimise χ^2 .

4.5 GALAXY SELECTION

The observed galaxy sample will be a Poisson realisation of the underlying true galaxy distribution. The error in the final catalogue due to incomplete sampling by the selection function or because of near neighbours⁹ can be estimated via bootstrapping. The 250,000 galaxies observed by 2dF are replaced at random on any of the sites in the catalogue, resulting in more than one at some sites and none at others. The variance of $g(v)$ from these bootstrapped catalogues describes the rigidity of the algorithm and of the data set. There are also holes of missing data around bright stars in the APM catalogue, which the 2dF GRS sources for galaxy positions. For small holes, data could be interpolated across the gap when the field is smoothed and divided by the value of the mask. For larger holes, subtracting the genus of the mask from the genus of the survey data gives the required result. This just adds more boundaries and boundary corrections.

Cosmic variance can then be emulated by extracting mock catalogues from several independent realisations of the n -body simulations. It is worth noting that the analysis will be performed upon volume-limited mock catalogues. Pushing the limits to larger volumes reduces cosmic variance but tends to select only the very luminous galaxies, which may well be intrinsically more clustered than a complete selection. This is discussed further in section 5.2.

⁹ 2dF cannot measure spectra of galaxy pairs closer than $11.4''$ because of difficulties packing fibres into the focal plane.

5 MOCK 2dF CATALOGUES

A collection of mock catalogues has been obtained via n -body simulations of a flat Λ CDM universe by Cole *et al.* (1998). These catalogues have been extracted with a ‘biasing’ algorithm (29), from each of ten independent evolutions of dark matter in an adaptive particle mesh.

$$P(v) = \begin{cases} \exp(\alpha v + \beta v^{3/2}) & \text{if } v \geq 0 \\ \exp(\alpha v) & \text{if } v < 0 \end{cases} \quad (29)$$

Model parameters have been normalised such that the simulated population reproduces the observed abundance of rich galaxy clusters and the amplitude and slope of the correlation function on scales of ~ 1 - $10 h^{-1}$ Mpc. The catalogues are available in periodic $345.6 h^{-1}$ Mpc cubes, with 128^3 galaxies giving a mean galaxy density of $\langle N \rangle = 5 \times 10^{-2} h^3 \text{ Mpc}^{-3}$. Volume-limited sub-samples have then been extracted from these, in wedges that simulate the geometry and selection function of the southern (SGP) region of the 2dF survey. To maintain a cosmic significance with structures up to $\sim 50 h^{-1}$ Mpc across, these mock catalogues have been prepared to a redshift of $z = 0.1$ which gives only $\langle N \rangle = 1.5 \times 10^{-2} h^3 \text{ Mpc}^{-3}$. This is not really enough and it would be interesting to repeat the analysis to a more densely packed $z \approx 0.05$. However, there must be some effort to overcome the fundamental flaw with the CfA redshift survey (for topology), that large-scale structures are of the same approximate size as the survey itself. Either way, there will be systematic errors: at least with a large but sparsely populated survey, these errors are generally understood and new parts of the universe can be investigated. A snapshot from the simulations at a fixed time is deemed satisfactorily realistic for these small redshifts – i.e. no lookback time is included. Furthermore, analysis is performed in real space, without the added random confusion of peculiar velocities in redshift space.

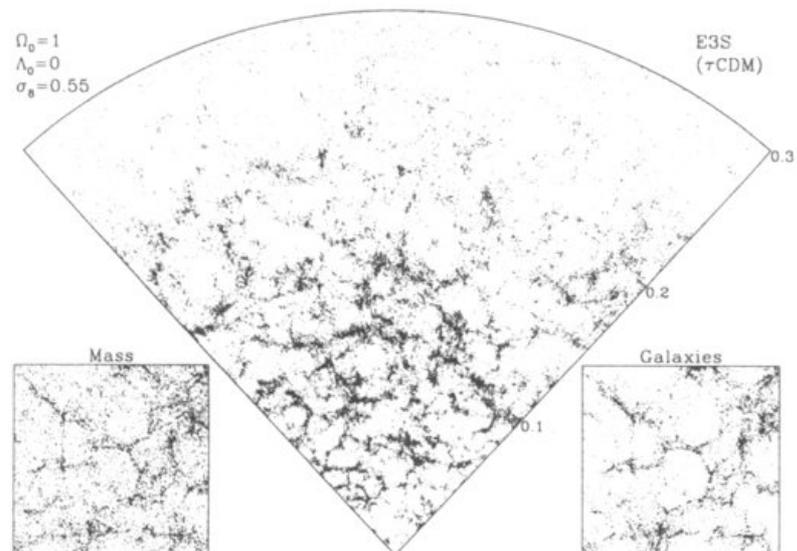


Figure 14. Example τ CDM redshift slice from Cole *et al.* (1998). This shows a magnitude limited catalogue with galaxies in redshift space and out to $z = 0.3$, so is for illustration purposes only

5.1 PERIODIC CUBES

Figure 15 shows the mean of genus curves calculated from ten independent realisations of the Λ CDM model in the n -body simulations described above. Standard error bars between the ten are plotted; plus a contribution from numerical errors, which have been estimated using the analyses of section 4. However, the error bars are of the same size as the (solid) circles used to plot the data and therefore not usually visible. This is no surprise because the grids have been arranged to automatically satisfy criteria (14) and (15)¹⁰ and have no boundaries. In such a large combined survey volume, the other source of scatter is also negligible: cosmic variance.

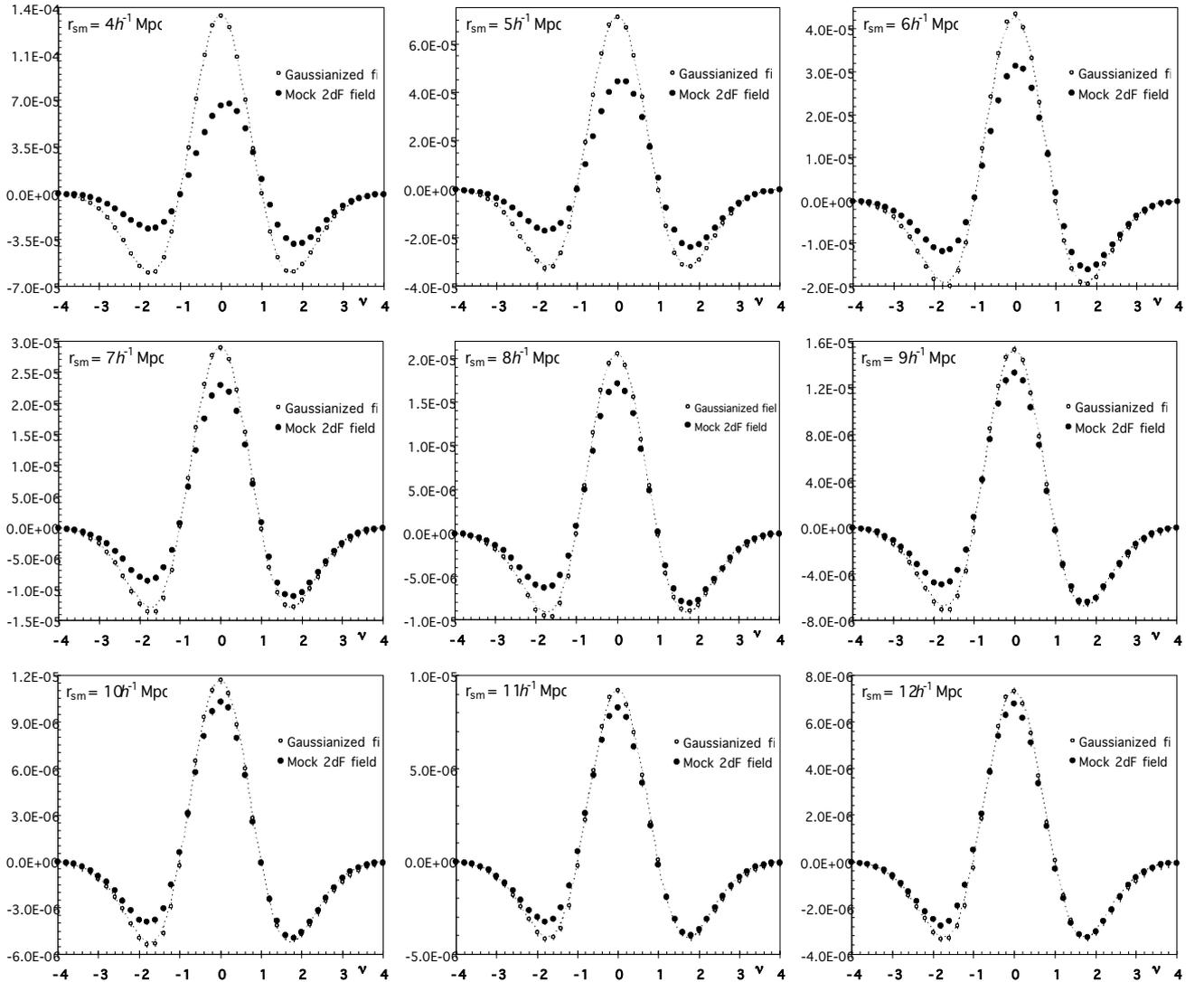


Figure 15. Genus curves per unit volume from an ensemble of 10 mock 2dF catalogues each in a periodic cube of side length $345.6 h^{-1}$ Mpc. Solid circles show the mock 2dF data; open circles the same data after Gaussianization. The dotted line is a fit to the latter. Error bars are shown but are very small: as expected in such a large & periodic grid.

¹⁰ Not quite for $r_{\text{sm}} = 4 h^{-1}$ Mpc. This has been taken into account with bigger error bars.

Figure 15 also shows (open circles) genus curves of corresponding Gaussianized density fields, created by randomising the phases of its Fourier components. As expected, these follow almost exactly the analytic curve (dotted line) given by equation (6). Random-phase fits to both data sets, which minimise χ^2 , are used in the calculation of the metastatistics previously introduced in section 3.4. Summaries of the metastatistics are shown in figure 16.

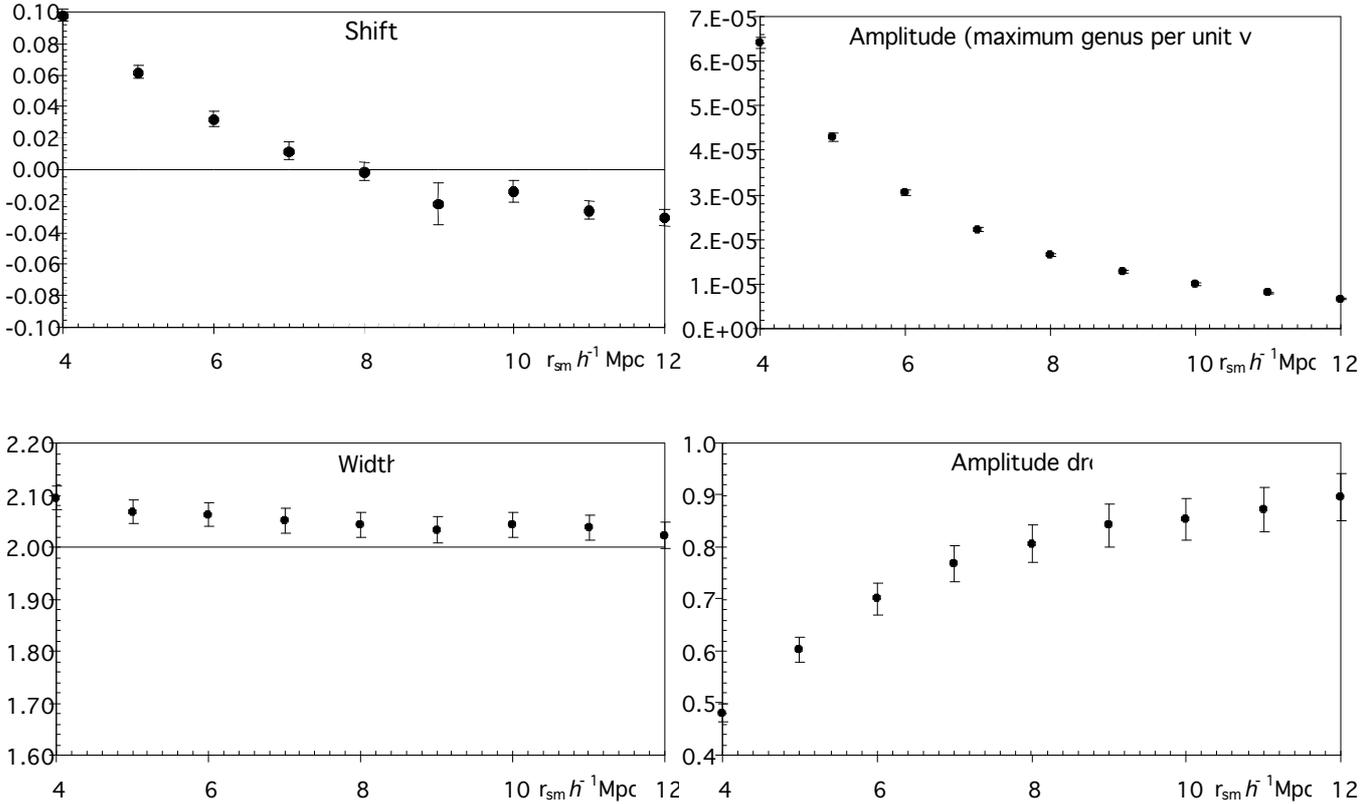


Figure 16. *Metastatistics (see section 3.4) for the mock 2dF genus curves in periodic grids. Error bars are standard errors between the ten catalogues, plus an estimate of systematic numerical inaccuracies from the analyses in section 4.*

The mock 2dF data show an apparent trend towards Gaussianity on larger scales. The amplitude of the genus curve steadily approaches that of the random-phase curve and the asymmetry seen very strongly for $r_{\text{sm}} \sim 6\text{--}8 h^{-1} \text{Mpc}$ begins to even out by $r_{\text{sm}} = 12 h^{-1} \text{Mpc}$. This agrees with other published data, e.g. Canavezes *et al.* (1998). Any unusual width of the curve or coherent shift towards either a “meatball” or “swiss-cheese” topology is very slight and also disappears on longer scale lengths.

On short scales, the density fluctuations are clearly non-Gaussian – with a topology that a comparison with figure 4 demonstrates to be slightly filamentary. This is due to non-linear

gravitational evolution and rapid accretion onto (rich clusters of) galaxies forming around overdensities. However, the immediately interesting information for cosmology comes from the less dramatic non-Gaussian behaviour on very large scales. If this can be measured with sufficient accuracy, it may be useful for constraining cosmological models and parameters.

5.2 2dF SOUTH GALACTIC POLE

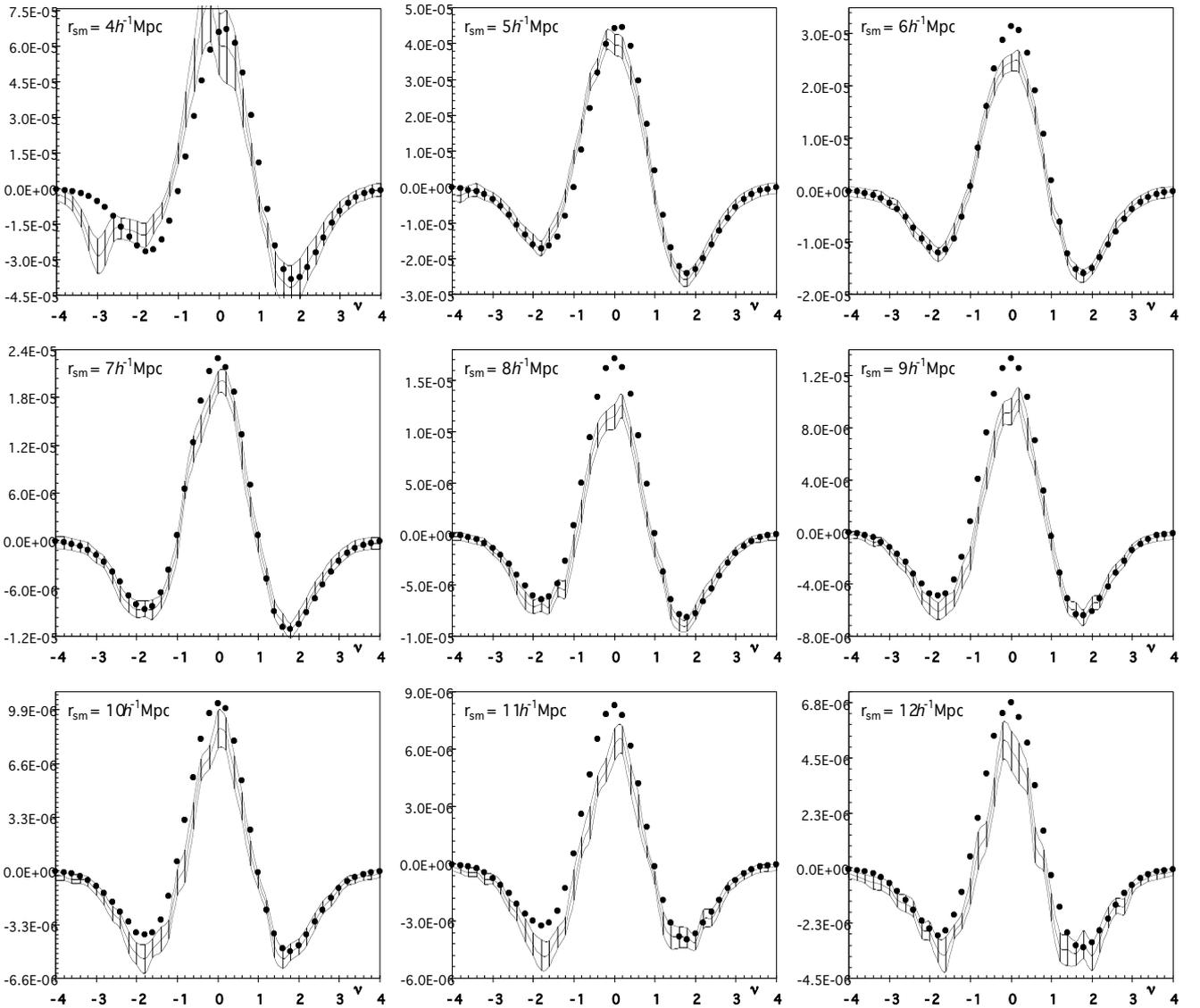


Figure 17. Genus curves from an ensemble of 10 mock volume-limited 2dF SGP catalogues prepared to $z = 0.1$. Hatched areas show standard error between the catalogues, combined in quadrature with estimates of numerical errors from the analyses in section 4. Note the systematic differences between this measurement and that taken in periodic grids (solid circles), mainly due to a low galaxy density in the volume-limited catalogues.

The analysis now moves to tackle volume-limited sub-samples of the same mock 2dF catalogues,

in wedges mimicking the southern region (SGP) of the galaxy redshift survey out to $z = 0.1$. To recap, the errors particularly dominant in this data will be due to sparse sampling of galaxy positions. The volume of the survey is large; binning it into grid cells sufficiently fine to minimise *numerical* errors is simply a matter of computer processing time. For smoothing lengths above $6 h^{-1}$ Mpc, a $288 \times 96 \times 288$ grid was perfectly adequate for the $\sim 360 \times 120 \times 360 h^{-3}$ Mpc³ box needed to contain the $2 \times 10^6 h^3$ Mpc⁻³ survey region. Below that, a $432 \times 144 \times 432$ grid was required. Numerical-type errors from criteria (14) and (16) are therefore easy to control¹¹. Along with an estimate of error due to low galaxy density (from figure 9), these are plotted as hatched areas in figure 17. However, sparse sampling produces systematic, not random, errors. The final genus curves measured from a mock SGP region differ significantly from those previously calculated with the same catalogues in a periodic and completely sampled cube – which, for this purpose, are assumed to be correct. Gaussianized genus curves used for the metastatistics (figure 18) have again been calculated from results in a periodic cube.

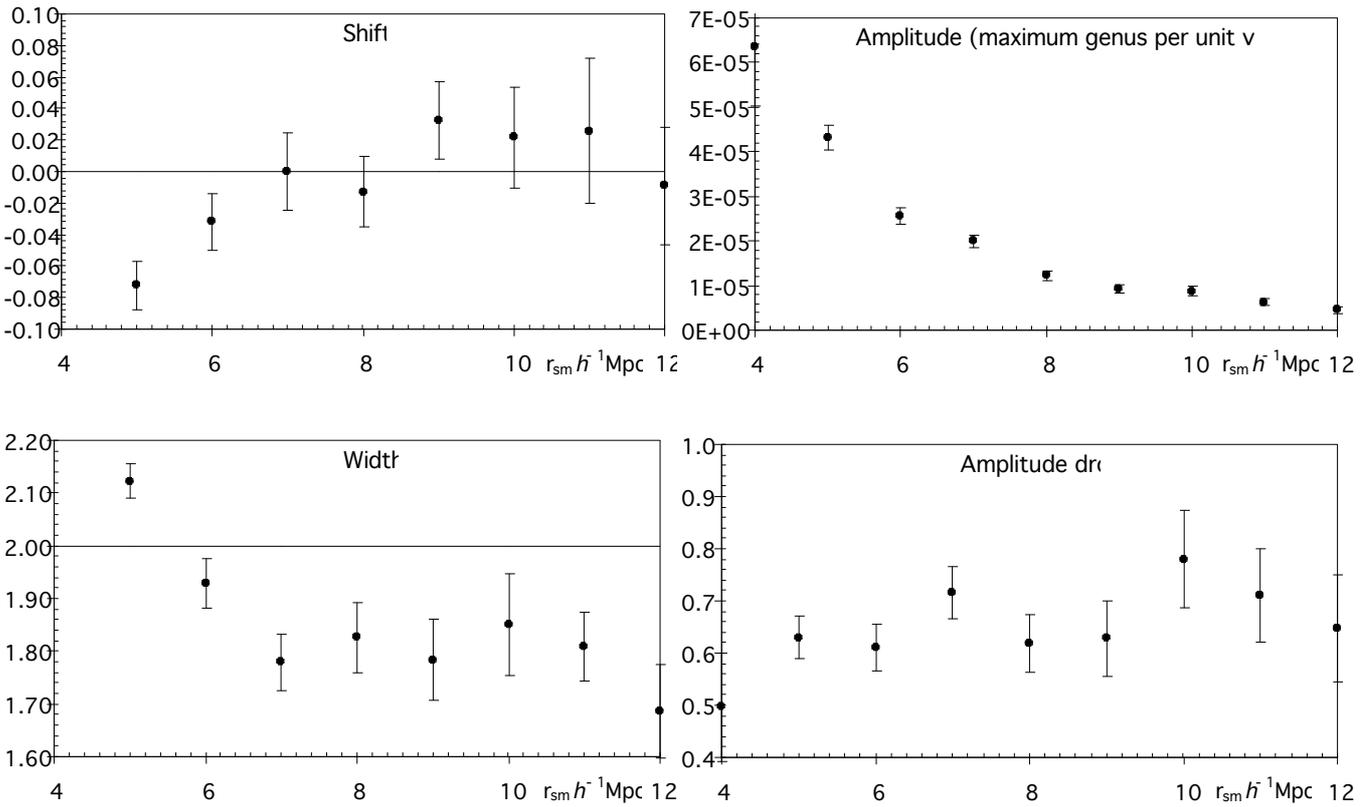


Figure 18. *Metastatistics (see section 3.4) for genus curves of the mock 2dF SGP fields (illustrated in figure 14). Error bars are standard errors between the ten catalogues, plus an estimate of systematic numerical inaccuracies from the analyses in section 4.*

¹¹ The curve for $r_{sm} = 4 h^{-1}$ Mpc shows peculiar artefacts which bring to mind those due to an overly coarse binning of the density field described in section 4.1. Perhaps requirement (14) should also be tightened for a clustered density field.

The systematic differences due to sparse sampling are most pronounced in the width and height of $g(v)$, as discussed in section 4.2. Clusters of galaxies are little affected by one or two missing (unobserved) galaxies. However, the filaments or sheets joining them are often tenuous and poorly defined, and galaxies removed from these can easily break up the whole interconnected topology of the distribution. Only for a limited range of v does an overall “swiss-cheese” topology survive intact. For most densities, the genus suddenly drops as isocontours fragment into separate regions around individual clusters. Under-dense voids fare even worse and the minimum at $v \approx -\sqrt{3}$ is exaggerated because extremely isolated galaxies almost never connect to their neighbours. This corrugation or pock-marking of the density field is particularly evident (in figure 17) for smoothing lengths above $\sim 9 h^{-1}$ Mpc. The particular shape of the periodic genus curve means that this SGP exaggeration happens to mimic Gaussianity; which does question which of the two measurements is truly correct. Low galaxy densities ought to be a problem instead at *short* smoothing lengths. However, given the size of other errors in the SGP genus curves, a reliance upon the periodic results is probably safe.

There have clearly been some difficulties making the SGP measurements, seemingly due to an insufficient galaxy density. This has tended to make the central peak of the genus curve lower and narrower than it was in the fully sampled mock catalogues; which in turn deteriorates the measurement of shift. Qualitatively, however, the algorithm has coped well with a complicated geometry and inevitably sparse sampling. These simulated results are rather narrow but otherwise consistent with PSCz observations from Canavezes *et al.* (1998) and the 1.2 Jy IRAS survey from Protogeros & Weinberg (1997). This method will still be useful so long as that when results are used to constrain cosmologies, they are compared against predictions from similarly degraded simulations.

An alternative approach would be to use magnitude-limited sub-samples (or the whole survey). This quickly and easily raises the galaxy density for most regions of the survey. The redshift selection function would be reflected in a value of the mask decreasing from unity: which would also introduce a volume limit at 80% of $\langle N \rangle$ at $z = 0$ from criterion (16). The problem is that more distant galaxies are biased more luminous because of selection effects: only the brightest galaxies make it into a magnitude-limited catalogue. These typically form through mergers, which have triggered starbursts and built AGNs (Sanders & Mirabel 1996). Bright galaxies are therefore

more likely to be found in very dense rich clusters and their distribution would perhaps be clumpier. Different, anyway. Although simple (linear) biasing does not affect the genus statistic, this selection mechanism by luminosity is not fully understood: particularly since it may involve galaxy mergers and anti-biasing. An investigation to determine it would also tie in nicely with the goal of 2dF to compare the clustering of galaxies as a function of spectral type.

It would finally be productive to repeat the analysis with mock catalogues to only $z \approx 0.05$ and compare these results to CfA data. The 2dF GRS surveys to a fainter magnitude, is more complete and is therefore more accurate. Analysis of how well the CfA redshift survey did, compared to modern 2dF results, can be taken as an indication of how well 2dF data will fare when pushed to its own limits of large volume-limited sub-samples. I suspect that this will warn of large systematic errors in previous analyses. Certainly, results from CfA vary greatly between the Northern and Southern hemispheres (Vogeley *et al.* 1994), due to large error bars and cosmic variance. Just the 2dF data from nearby could also be compared to the 1.2 Jy IRAS (IR-selected) survey. Infrared radiation is probably a better tracer of mass than optical light and may give results more consistent with n -body simulations of dark matter. As mentioned, 2dF galaxy clustering is already being examined as a function of colour. However, since all 2dF coordinates are taken from the (optical) APM survey, it would be interesting to compare the results from differently selected galaxy catalogues.

6 CONCLUSIONS

A Fortran algorithm has been developed and tested for the calculation of the topological ‘genus’ statistic or *interconnectivity* of the galaxy distribution in large-scale structure. The topology of a density field is dependent upon the phases of its Fourier transform and hence a useful way to investigate structure, complementary to the power spectrum, which is determined by their amplitudes. The algorithm has been applied to mock catalogues of the 2dF galaxy redshift survey, which is currently mapping out galaxies in a region of sky around each galactic pole. These regions should be large enough to provide useful constraints on cosmologies, which have not been possible before from redshift surveys only the same size as the structure being investigated. However, this work has not yet attempted to quantify these constraints.

The genus as a function of density is known analytically for a Gaussian (random-phase) distribution, which is favoured by the standard cosmological model. Artificially Gaussianized fields have been manufactured to isolate and then control the typical sources of errors in real galaxy density fields. Several conditions have been imposed on the computing algorithm in order to keep the point-wise contribution from each source below 1% of the maximum genus. These include a sufficiently fine grid mesh to store the galaxy densities; a constraint on the treatment of data near boundaries; and a minimum galaxy density when a volume-limited sub-sample is extracted from the overall catalogue.

$$r_{sm} \geq 2 \cdot 5 l_{cell} \quad (14)$$

$$\langle N \rangle r_{sm}^3 \geq 6 \quad (15)$$

$$\text{discard data within } \frac{2}{3} r_{sm} \text{ of the edge of the survey region.} \quad (16)$$

Measurements of genus curves have been presented from mock catalogues in periodic cubes and in wedges that mimic the geometry and selection function of the Southern region of the 2dF GRS. Both were extracted via a simple biasing algorithm from n -body simulations of a flat Λ CDM model. Results show a strongly non-Gaussian clustering at short scales due to non-linear gravitational evolution and accretion onto over-densities. On scales larger than $\sim 10 h^{-1}$ Mpc, the fields tend to an almost Gaussian behaviour but measurements suffer from a sparse sampling of the galaxy population which fragments the distribution and lowers the genus curve. Even when

satisfying the computational constraints, a possible detection of deviations from Gaussianity using the 2dF GRS will be subject to ~6% random errors and large systematic errors. Any comparisons with theoretical data would need to be performed on a similarly degraded theoretical population. The obvious next step is to use these results to constrain cosmological models and parameters, and see how much these systematic errors matter.

The topological results for the simulated galaxy catalogues are largely consistent with previously published work. When the 2dF GRS is finally complete, it will also be productive to compare this large data set with earlier, well-studied observational data such as the CfA redshift survey or particularly the IRAS 1.2 Jy catalogue. This is IR-selected and may arguably trace the distribution of dark matter more faithfully than the optically selected 2dF data.

REFERENCES

- Armstrong M.A. 1983, *Basic Topology*, Springer-Verlag
- Bailey J. & Glazebrook K. 1996, *2dF User Manual*, Anglo-Australian Observatory
- Beaky M.M., Scherrer R.J. & Villumsen J.V. 1992 *ApJ* 387, 443 Topology of large-scale structure in seeded hot dark matter models
- Canavezes A. *et al.* 1998 *MNRAS* 297, 777 The topology of the IRAS Point Source Catalogue Redshift Survey
- Cole S. *et al.* 1998 *MNRAS* 300, 945 Mock 2dF and SDSS galaxy redshift surveys
- Coles P., Davies A. G. & Pearson R. C. 1996 *MNRAS* 281, 1375 Quantifying the topology of large-scale structure
- Colley *et al.* 1999 *ApJ* 332 Topology from the Simulated Sloan Digital Sky Survey
- Davies A. & Coles P. 1993 *MNRAS* 260, 553 Topology in two dimensions – III. Modelling projected galaxy catalogues
- Doroshkevich A.G. 1970 *Astrofizika* 6, 320
- Gott, J.R. 1991 SSP.conf. 173G Topology of large-scale structure: recent results
- Hadwiger H. 1957 *Vorlesungen uber Inhalt, Oberflache und Isoperimetrie*, Springer-Verlag
- Huchra J.P. *et al.* 1992 *The CfA Redshift Catalogue*, Harvard-Smithsonian Center for Astrophysics, Cambridge MA.
- Kerscher M. *et al.* 1997 *A&A* 333, 1 Fluctuations in the IRAS 1.2 Jy catalogue
- Kolb E.W. & Turner M. S. 1990 *The early universe*, Addison-Wesley Publishing
- Maddox S. *et al.* 1990 *MNRAS* 243, 692 The APM galaxy survey. I – APM measurements and star-galaxy separation
- Melott A.L., Weinberg D.H. & Gott J.R. 1988 *ApJ* 328, 50 The topology of large-scale structure – II. Nonlinear evolution of Gaussian models
- Peacock J. 1999 *Cosmological physics*, Cambridge University Press
- Peebles P.J.E. 1993 *Principles of physical cosmology*, Princeton University Press

- Protopogeros Z.A.M. & Weinberg D.H. 1997 ApJ 489, 457 The topology of large-scale structure in the 1.2 Jy IRAS redshift survey
- Ratcliffe A. *et al.* 1996 MNRAS 281, L47 The Durham/UKST Galaxy Redshift Survey – I. Large-scale structure in the Universe
- Sanders D.B. & Mirabel I.F. 1996 Annu. Rev. Astron. Astrophys. 34, 749 Luminous infrared galaxies
- Schmalzing J. & Buchert T. 1997 ApJ 482, L1 Beyond genus statistics: a unifying approach to the morphology of cosmic structure
- Springel V. *et al.* 1998 MNRAS 298, 1169 Genus statistics of the Virgo *N*-body simulations and the 1.2-Jy redshift survey
- Weinberg D.H., Gott J.R. & Melott A.L. 1987 ApJ 321, 2 The topology of large-scale structure – I. Topology and the random phase hypothesis
- Vogeley *et al.* 1994 ApJ 420, 525 Topological analysis of the CfA redshift survey

A1 FORTRAN 77 GENUS ALGORITHM


```

program genus
c   Calculates the genus statistic of a pre-smoothed density field
c   using the method of Coles et al. (1996), MNRAS, 281, 1375.
c   By Andrew Benson (March 1999) & Richard Massey (February 2000).
c
implicit none
integer NNUMAX           ! # of x positions on genus curve
parameter (NNUMAX=100)  !
real genus (NNUMAX,2)   ! genus statistic (inc/exc)
integer inu,nnu          ! counters
real nu,numin,numax,nustep ! density
integer NGRID,XPAR,YPAR,ZPAR ! size of density and mask grids
parameter (NGRID=144,XPAR=3,YPAR=1,ZPAR=3) !
real grid(NGRID*XPAR,NGRID*YPAR,NGRID*ZPAR) ! density field data
real mgrid(NGRID*XPAR,NGRID*YPAR,NGRID*ZPAR) ! mask data
integer densidx(XPAR*YPAR*ZPAR*NGRID**3) ! index for sorting density field
integer ngx,ngy,ngz,mngx,mngy,mngz ! coordinates in grids
real xmin,xmax,ymin,ymax,zmin,zmax ! dimensions of grids
real mxmin,mxmax,mymin,mymax,mzmin,mzmax !
real lx,ly,lz,vol !
integer inside ! # of cells
integer iargc ! # of arguments on command line
integer i,ix,iy,iz,j,jx,jy,jz ! counters
integer genuscalc ! function to calculate genus
real cutoff ! proximity edge/non-periodic data
real rs,mrs ! smoothing lengths
real biggest,mean,var,erf ! temporary stores for statistics
character outfile*100,mfile*100,infile*100 ! filenames

c
if (iargc().ne.7) then
  write (0,*) 'USAGE: genus.exe infile maskfile outfile numin numax nnu cutoff'
  stop
endif
write (0,*) 'GENUS.EXE initialised for a',NGRID*XPAR,' x',NGRID*YPAR,
           ' x',NGRID*ZPAR,' grid'

c
c READ IN ARGUMENTS:
c
call getstr(1,infile)
call getstr(2,mfile)
call getstr(3,outfile)
call getreal(4,numin)
call getreal(5,numax)
call getint(6,nnu)
call getreal(7,cutoff)
if ((int(float(nnu)/10.)*10).eq.nnu) nnu=nnu+1
if (nnu.gt.numax) stop 'nnu out of range'
if (cutoff.lt.0..or.cutoff.gt.1.) stop 'cutoff out of range'

c
c READ IN GALAXY DATA:
c
write (0,*) 'READING GALAXY DATA'
open (unit=13,file=infile,status='old',form='unformatted')
read (13) ngx,ngy,ngz
read (13) rs
read (13) xmin,xmax,ymin,ymax,zmin,zmax
if (ngx.ne.ngrid*xpar.or.ngy.ne.ngrid*ypar.or.ngz.ne.ngrid*zpar) then
  write (0,*) 'Mismatch: Data requires a ',ngx,' x',ngy,' x',ngz,' grid'
  close (13)

```

```

  stop
endif
read (13) (((grid(ix,iy,iz),ix=1,ngx) ,iy=1,ngy) ,iz=1,ngz)
close (13)
lx=xmax-xmin
ly=ymin-ymin
lz=zmax-zmin
vol=lx*ly*lz/float(ngx*ngy*ngz)
write (0,*) 'Smoothing length =',rs,' Mpc/h'
write (0,*) 'Grid size =',lx,' x',ly,' x',lz,' Mpc/h'
write (0,*) 'Volume per grid cell =',vol,' (Mpc/h)^3'

c
c INITIALISE OBSERVED SKY MASK:
c
write (0,*) 'READING MASK DATA'
open (unit=12,file=mfile,status='old',form='unformatted')
read (12) mngx,mngy,mngz
read (12) mrs
read (12) mxmin,mxmax,mymin,mymax,mzmin,mzmax
if (mngx.ne.ngx.or.mngy.ne.ngy.or.mngz.ne.ngz) then
  write (0,*)
  write (0,*) 'WARNING: data and mask grid size mismatch'
  write (0,*) 'Data in a ',mngx,' x',mngy,' x',mngz,' grid'
  write (0,*) 'Mask in a ',mngx,' x',mngy,' x',mngz,' grid'
  write (0,*)
  close (12)
  stop
endif
read (12) (((mgrid(ix,iy,iz),ix=1,mngx) ,iy=1,mngy) ,iz=1,mngz)
close (12)
write (0,*) mxmax-mxmin,'=?=',xmax-xmin
write (0,*) mymax-mymin,'=?=',ymax-ymin
write (0,*) mzmax-mzmin,'=?=',zmax-zmin
write (0,*) 'Mask smoothing length =',mrs,' Mpc/h'

c
Renormalise mask to unity
biggest=0.
do ix=1,mngx
  do iy=1,mngy
    do iz=1,mngz
      if (mgrid(ix,iy,iz).gt.biggest) then
        biggest=mgrid(ix,iy,iz)
        jx=ix
        jy=iy
        jz=iz
      endif
    enddo
  enddo
enddo
write (0,*) 'Biggest value in mask is ',biggest,' at (',jx,',',jy,',',jz,')'
c
Find average nearby then rescale by this amount.
mean=0.
var=0.
do ix=-xpar+1,xpar
  do iy=-ypar+1,ypar
    do iz=-zpar+1,zpar
      write (0,*) mgrid(jx+ix,jy+iy,jz+iz)
      mean=mean+mgrid(jx+ix,jy+iy,jz+iz)
      var=var+mgrid(jx+ix,jy+iy,jz+iz)**2.
    enddo
  enddo
enddo

```

```

        enddo
    enddo
    mean=mean/float(xpar*ypar*zpar)/8
    var=var/float(xpar*ypar*zpar)/8
    if (var.gt.mean**2) then
        write (0,*) 'Taking mask height as =',mean,' +/-',
            sqrt(var-mean**2.)/float(xpar*ypar*zpar)
    else
        write (0,*) 'Taking mask height as =',mean
    endif
    inside=0
    do ix=1,mngx
        do iy=1,mngy
            do iz=1,mngz
                mgrid(ix,iy,iz)=mgrid(ix,iy,iz)/mean
                if(mgrid(ix,iy,iz).ge.cutoff) then
                    Compensate for edge effects by multiplying densites by 1/mask
                    grid(ix,iy,iz)=grid(ix,iy,iz)/mgrid(ix,iy,iz)
                    inside=inside+1
                else
                    grid(ix,iy,iz)=1.e30
                    This is then sorted to the end of the index in a moment
                endif
            enddo
        enddo
    enddo
    write (0,*) 'Mask occupies ',inside,' = ',nint((float(inside))**(1./3.)),
        '^3 grid cells'
    write (0,*) '
        =',inside*vol,' (Mpc/h)^3'
    write (0,*) 'Genus of mask is
        =',
        genuscalc(cutoff,ngx,ngy,ngz,mgrid,mgrid,cutoff)
c   Surface area (of mask?) and other Minkowski functionals?
c
c   CALCULATE GENUS:
c
c   First sort the density field to find increments of nu
    write (0,*) 'Sorting density field'
    call indexxx(ngx*ngy*ngz,grid,densidx)
    write (0,*) 'CALCULATING GENUS CURVES'
    nustep=(numax-numin)/float(nnu-1)
    do j=1,2
        if(j.eq.1) write (0,*) 'Excursion set'
        if(j.eq.2) write (0,*) 'Incursion set'
        do inu=1,nnu
            nu=numin+float(inu-1)*nustep
            Find the value of nu defined by volume equivalent to nu defined by variance
            i=int(0.5*(1.+erf(nu/sqrt(2.)))*float(inside))
            call lookup(i,ngx,ngy,ngz,densidx,jx,jy,jz)
            nu=grid(jx,jy,jz)
            genus(inu,j)=float(genuscalc(nu,ngx,ngy,ngz,mgrid,grid,cutoff))/inside/vol
            write (0,*) inu,nu,' (' ,i,')',genus(inu,j)
        enddo
    enddo
c   Swap signs within density field to find incursion set
    if(j.eq.1) then
        do ix=1,ngx
            do iy=1,ngy
                do iz=1,ngz
                    grid(ix,iy,iz)=grid(ix,iy,iz)*-1.
                enddo
            enddo
        enddo

```

```

        enddo
    enddo
    endif
enddo
c
c   OUTPUT DATA
c
    write (0,*) 'Outputting data to output//outfile
    open(unit=14,file=outfile,status='unknown',form='formatted')
    write (14,*) 'nu',' excursion',' incursion',nint(cutoff*100),'%'
    do inu=1,nnu
        nu=numin+float(inu-1)*nustep
        write (14,*) nu,(genus(inu,i),i=1,2),0.5*(genus(inu,1)+genus(inu,2))
    enddo
    close(14)
    write (0,*) 'Done'
end
c
c   *****
c   FUNCTIONS AND SUBROUTINES
c   *****
c
function genuscalc(nu,ngx,ngy,ngz,mgrid,grid,cutoff)
c   Uses the method of Coles et al. (1996), MNRAS 281, 1375
implicit none
integer genuscalc           ! returns absolute value of genus
real nu                    ! density of contour (SD from mean)
real cutoff                ! how much data to ignore round edge
integer ngx,ngy,ngz        ! size of mask&data grids
real grid(ngx,ngy,ngz)    ! data
real mgrid(ngx,ngy,ngz)   ! mask
logical lxt,lyt,lzt,sqxyt,sqxzt,sqyzt ! temporary 'yes/no's
integer point,cube,lx,ly,lz,sqxy,sqyz,sqxz ! counters for number of squares etc
integer ix,iy,iz,ixl,iyl,izl ! loop indices
c   Initialise all the counters
point=0
cube=0
lx=0
ly=0
lz=0
sqxy=0
sqxz=0
sqyz=0
c   Loop through each point in the grid
do ix=1,ngx
    ixl=ix+1
    if (ixl.gt.ngx) ixl=1
c   Assures periodicity: the mask is assumed to be zero near edges in a wedge or
    an intentionally non-periodic grid
    do iy=1,ngy
        iyl=iy+1
        if (iyl.gt.ngy) iyl=1
        do iz=1,ngz
            izl=iz+1
            if (izl.gt.ngz) izl=1
c   If this cell is within mask and above the threshold nu then it is a point
            if (grid(ix,iy,iz).ge.nu.and.mgrid(ix,iy,iz).ge.cutoff) then
                point=point+1
            enddo
        enddo
    enddo
enddo

```

```

c   Reset temporary 'yes/no's
      lxt=.false.
      lyt=.false.
      lzt=.false.
      sqxyt=.false.
      sqxzt=.false.
      sqyzt=.false.
c   If the cells along the principal axes are above threshold then we have lines
      if (grid(ix1,iy,iz).ge.nu.and.mgrid(ix1,iy,iz).ge.cutoff.and.ix1.gt.1)
          then
              lx=lx+1
              lxt=.true.
          endif
      if (grid(ix,iy1,iz).ge.nu.and.mgrid(ix,iy1,iz).ge.cutoff.and.iy1.gt.1)
          then
              ly=ly+1
              lyt=.true.
          endif
      if (grid(ix,iy,iz1).ge.nu.and.mgrid(ix,iy,iz1).ge.cutoff.and.iz1.gt.1)
          then
              lz=lz+1
              lzt=.true.
          endif
c   Now check for squares
      if (lxt.and.lyt.and.grid(ix1,iy1,iz).ge.nu.and.mgrid(ix1,iy1,iz).
          ge.cutoff) then
              sqxy=sqxy+1
              sqxyt=.true.
          endif
      if (lxt.and.lzt.and.grid(ix1,iy,iz1).ge.nu.and.mgrid(ix1,iy,iz1).
          ge.cutoff) then
              sqxz=sqxz+1
              sqxzt=.true.
          endif
      if (lyt.and.lzt.and.grid(ix,iy1,iz1).ge.nu.and.mgrid(ix,iy1,iz1).
          ge.cutoff) then
              sqyz=sqyz+1
              sqyzt=.true.
          endif
c   Finally check for the cube
      if (sqxzt.and.sqxyt.and.sqyzt.and.grid(ix1,iy1,iz1).ge.nu.and.
          mgrid(ix1,iy1,iz1).ge.cutoff) cube=cube+1
          endif
      enddo
      enddo
      enddo
c   genus now calculated using a combination of the counters by Poincare's theorem
      genuscalc=cube-(sqxy+sqyz+sqxz)+(lx+ly+lz)-point+1
c   write (0,*) 'cubes =',cube,' squares =',(sqxy+sqyz+sqxz),' lines=',
          (lx+ly+lz),' points =',point
      return
      end
c
c *****
c

```

```

c
c *****
c
      subroutine lookup(ix,nx,ny,nz,densidx,jx,jy,jz)
c   Look up grid coordinates (jx,jy,jz) of ixth density value in the index
      implicit none
      integer ix,iy,nx,ny,nz,jx,jy,jz
      integer densidx(nx*ny*nz)
      if (ix.gt.nx*ny*nz) ix=nx*ny*nz
      if (ix.lt.1) ix=1
      iy=densidx(ix)-1
      jz=int(iy/nz/ny)
      iy=iy-jz*nz*ny
      jy=int(iy/nz)
      iy=iy-jy*nz
      jx=iy
      jx=jx+1
      jy=jy+1
      jz=jz+1
      return
      end
c
c *****
c

```