Mon. Not. R. Astron. Soc. **327,** 1041–1056 (2001)

The clustering evolution of the galaxy distribution

A. J. Benson,[1,2]★ C. S. Frenk,[1] C. M. Baugh,[1] S. Cole[1] and C. G. Lacey[3]
[1]*Physics Department, University of Durham, Durham, DH1 3LE*
[2]*California Institute of Technology, MC 105-24, Pasadena, CA 91125-2400, USA*
[3]*SISSA, via Beirut 2-4, 34014 Trieste, Italy*

Accepted 2001 July 12. Received 2001 July 10; in original form 2001 March 15

ABSTRACT

We follow the evolution of the galaxy population in a ΛCDM cosmology by means of high-resolution *N*-body simulations in which the formation of galaxies and their observable properties are calculated using a semi-analytic model. We display images of the spatial distribution of galaxies in the simulations that illustrate its evolution and provide a qualitative understanding of the processes responsible for the various biases that develop. We consider three specific statistical measures of clustering at $z = 1$ and $z = 0$: the correlation length (in both real and redshift space) of galaxies of different luminosity, the morphology–density relation and the genus curve of the topology of galaxy isodensity surfaces. For galaxies with luminosity below L_*, the $z = 0$ correlation length depends very little on the luminosity of the sample, but for brighter galaxies it increases very rapidly, reaching values in excess of $10\,h^{-1}$ Mpc. The 'accelerated' dynamical evolution experienced by galaxies in rich clusters, which is partly responsible for this effect, also results in a strong morphology–density relation. Remarkably, this relation is already well-established at $z = 1$. The genus curves of the galaxies are significantly different from the genus curves of the dark matter, however this is not a result of genuine topological differences but rather of the sparse sampling of the density field provided by galaxies. The predictions of our model at $z = 0$ will be tested by forthcoming data from the 2dF and Sloan galaxy surveys, and those at $z = 1$ by the DEEP and VIRMOS surveys.

Key words: galaxies: evolution – galaxies: formation – dark matter – large-scale structure of Universe.

1 INTRODUCTION

Studies of the large-scale distribution of galaxies have traditionally focussed on problems such as testing hypotheses for the identity of the dark matter, the nature of the initial density perturbations and the mechanism of structure growth. Properties of the observed large-scale structure are also often used to estimate the values of fundamental cosmological parameters. Although none of these issues can be regarded as settled, there is now a growing consensus that cold dark matter (CDM) is the most likely candidate for the dark matter, that cosmic structure grew by the gravitational amplification of random-phase initial density fluctuations of inflationary origin, and that the fundamental cosmological parameters have the following values: density parameter $\Omega_0 \simeq 0.3$, cosmological constant term $\Lambda_0 \simeq 0.7$ and Hubble constant (in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$) $h \simeq 0.7$.

Cosmological constraints reflect only one aspect of the information encoded in the pattern of galaxy clustering. Another, equally interesting, aspect concerns the processes responsible for the formation and evolution of galaxies. To extract this kind of information requires very extensive data sets and these are only now becoming available in the form of a new generation of galaxy surveys, such as 2dF (Peacock et al. 2001), Sloan (Blanton et al. 2001) and 2MASS (Jarret et al. 2000). The expectation is that these new data sets will provide, in addition to further cosmological constraints, some understanding of how the physics of galaxy formation manifests itself in the clustering of galaxies, as a function of internal properties such as morphology, luminosity, colour, or star formation rate. Not only is this important for testing models of galaxy formation, but it is also required for extracting accurate cosmological information from the new surveys. Although it seems plausible that on very large scales the galaxy distribution traces the underlying mass in a simple way (Coles 1993; Cole et al. 1998), complex biases are predicted to be present on small and intermediate scales (Kauffmann et al. 1999a; Benson et al. 2000a).

To extract useful information from observational data of the quality and size of the new surveys, it is necessary to have detailed theoretical predictions. There have been significant advances in this area too in recent years, largely through the development of

★E-mail: abenson@astro.caltech.edu

increasingly realistic *ab initio* calculations of galaxy formation and evolution. Two strategies have been developed for this purpose. In the first one, cosmological *N*-body/gas-dynamics simulations are used to follow the coupled evolution of dark matter and gas, in particular, the cooling of gas in galactic dark matter haloes (e.g. Katz, Hernquist & Weinberg 1992; Evrard, Summers & Davis 1994; Frenk et al. 1996; Weinberg, Hernquist & Katz 1997; Blanton et al. 1999; Pearce et al. 1999). A phenomenological model is employed to decide when and where stars and galaxies form from this cooled gas and to include the associated feedback effects. In the second strategy, only the evolution of the dark matter component is simulated directly, or the assembly history of haloes is obtained with a Monte-Carlo method, and the behaviour of the gas is calculated by solving a simple, analytical, spherically symmetric cooling-flow model. As in the direct simulation approach, star formation and feedback are included in a phenomenological way.

The two strategies offer different advantages. Direct simulations solve the evolution equations for gravitationally coupled dark matter and dissipative gas without imposing any restrictions on geometry. However, limited resolution restricts the range of length and mass scales that can be studied, and the expense of large simulations makes it impractical to carry out extensive parameter space explorations. Because of its simplified treatment of gas dynamics, semi-analytic modelling can follow an essentially unlimited range of length and mass scales, and is sufficiently flexible that the effects of varying assumptions and parameter values can be readily explored. Additional processes that cannot currently be easily investigated at the resolution available in direct simulations, such as those determining galaxy morphology or the effects of dust obscuration, can be readily incorporated into the semi-analytic models by straightforward extensions to the phenomenological model of star formation and feedback. The numerical resolution and physical content of a typical *N*-body/gas-dynamic simulation can be mimicked in a semi-analytic model and Benson et al. (2001) have shown that, at least in the case where only the simplest gas physics are modelled, the two techniques give reassuringly similar statistical results.

In this paper we combine large *N*-body simulations with the semi-analytic model of Cole et al. (2000) to investigate certain properties of the galaxy distribution that are relevant to the new generation of galaxy redshift surveys. We begin by displaying images that illustrate the evolution of the galaxy population in a representative volume of a simulated CDM universe. These images furnish some qualitative understanding of the mechanisms responsible for establishing the spatial distribution of galaxies of different kinds. We then focus specifically on the dependence of the two-point correlation function on galaxy luminosity, the morphology–density relation and the topology of the galaxy distribution as measured by the genus, and on the evolution of these properties with redshift. The dependence of clustering on luminosity and colour have previously been considered, using similar techniques, by Kauffmann, Nusser & Steinmetz (1997), Kauffmann et al. (1999a) and Benson et al. (2000c). The first and last of these studies found a weak increase in the correlation function with luminosity on large scales, but the second failed to detect any effect. These papers used simulations of relatively small volumes and so were unable to investigate the clustering of the brightest galaxies, for which luminosity-dependent effects are expected to be strongest. The new generation of redshift surveys will include large samples of very bright galaxies and may well be able to measure this kind of effect. In this paper we extend earlier work and investigate clustering at the bright end of the galaxy luminosity function. Closely related to the dependence of clustering strength on luminosity is the morphology–density relation which we also quantify in our simulations, both at the present day and at $z = 1$. Finally, we provide the first determination of the genus curves predicted for galaxies in a CDM model; previous simulations had only been able to address the genus curves of the dark matter distribution. Although our model predictions are directed at the new surveys, we carry out limited comparisons with available observational data.

The remainder of this paper is organised as follows. In Section 2, we describe our simulation and modelling techniques. In Section 3 we present colour images of the evolution of a slice of our simulated volume (these images are publically available at http://www.astro.caltech.edu/~abenson/Mocks/mocks.html). In Section 4 we present quantitative estimates of clustering, namely the correlation length–luminosity and morphology–density relations and the genus statistic, and compare our results to observations. Finally, we present our conclusions in Section 5.

2 METHOD

The need for realistic modelling of galaxy formation as a prerequisite for deriving reliable clustering predictions has been emphasised by Benson et al. (2000a) who showed, for example, that the form of the two-point correlation function on small scales is strongly influenced by the physical processes governing galaxy formation. Such processes are readily taken into account when the techniques of semi-analytic modelling are grafted into *N*-body simulations of the dark matter (Kauffmann et al. 1997, 1999a,b; Diaferio et al. 1999; Benson et al. 2000a,c; Somerville et al. 2001). Monte-Carlo implementations of the semi-analytic model can also be used directly for clustering studies, without *N*-body simulations, but they only work well on scales larger than the Lagrangian radii of the dark matter haloes which host galaxies, which in practice means on scales in excess of a few Mpc (although see Seljak 2000), for which the bias can be calculated using the analytic formula of Mo & White (1996); see for example Baugh et al. (1999). Modelling the visible properties of galaxies explicitly allows simulated samples to be selected according to criteria closely patterned on observational selection procedures (e.g. by magnitude, colour, morphology, etc.), thus allowing rigorous comparisons with observations to be made.

In this paper, we use the techniques introduced by Kauffmann et al. (1997) and extended by Benson et al. (2000a) to graft our semi-analytic model of galaxy formation on to *N*-body simulations. Full details of our semi-analytic model and the extensions required to study galaxy clustering are given in Cole et al. (2000) and Benson et al. (2000a,c), respectively. Briefly, dark matter haloes are identified in the simulation at the redshift of interest using the friends-of-friends algorithm with the standard linking length of 0.2 (Davis et al. 1985). After cleaning the halo catalogue in the manner described below, the mass of each halo is input into the semi-analytic model. By means of simple, physically motivated prescriptions, described in detail in Cole et al. (2000), the model calculates the amount of gas that cools in a virialized halo of that mass to make a galaxy, as well as the star formation rate, the reheating of left-over gas by stellar winds and supernovae, and the chemical evolution of the gas and stars. Galaxies are allowed to merge within common dark matter haloes, producing elliptical galaxies and bulges from stellar discs. The spectrophotometric evolution of the galaxies is calculated using a standard stellar

population synthesis model (Bruzual & Charlot 1993; Bruzual & Charlot, in preparation). Extinction by dust is included using the models of Ferrara et al. (1999). The most massive galaxy in each halo is identified as the central galaxy and placed at the centre of mass (and given the peculiar velocity of the centre of mass). Other galaxies (satellites) are assigned the position and peculiar velocity of a randomly chosen dark matter particle within the halo. In this way, satellite galaxies always trace the dark matter within a given halo. Modelling of this kind has been successfully applied to study a large variety of properties of the galaxy population (e.g. Cole et al. 1994, 2000; Kauffmann, Guiderdoni & White 1994; Baugh, Cole & Frenk 1996a,b; Kauffmann 1996; Baugh et al. 1998; Somerville & Primack 1999; Granato et al. 2000.)

We consider only dark matter haloes in the simulation containing 10 or more particles. As we are interested in galaxies of all luminosities, including the faint ones that occupy haloes with masses close to the 10-particle limit, it is important to check that small haloes are actually bound objects. For this, we calculate the total energy of each halo by summing the kinetic energy (measured relative to the centre of mass of the halo) and the gravitational energy (caused by the interaction between all of the particles in the halo). If a halo is found to have positive energy (and so to be unbound), we remove the least bound particle and recompute the total energy. This process is repeated until either the energy becomes negative (in which case we now have a bound halo with a lower mass than the original) or there are fewer than 10 particles left in the halo (in which case we discard it). In this way, we construct a new halo catalogue containing only bound objects.

Typically, approximately 10 per cent of the haloes from the original catalogue fail the binding energy test and are excluded. A slightly smaller fraction have particles removed but remain in the catalogue. Most of the excluded haloes come from the low-mass end of the distribution, with the excluded fraction dropping rapidly as the halo mass increases. We find many examples of haloes contaminated by interlopers for which removal of a small number of the least bound particles results in a bound object. [Note that since the binding energy test only affects haloes near the resolution limit of the *N*-body simulation, it does not alter any of the results of Benson et al. (2000a) who considered only bright galaxies that form in haloes well above the resolution limit.]

We adopt similar values for the parameters of the semi-analytic model as did Cole et al. (2000), except that the parameters describing the normalization and shape of the power spectrum (σ_8 and Γ, see below) are set to the actual values used in the simulation. The parameters of the semi-analytic model are slightly different from those used in Benson et al. (2000a), but the differences in the predictions are negligible. Furthermore, because the dark matter halo mass function in the simulations differs somewhat from the Press–Schechter form assumed by Cole et al. (2000), we find that the model works better if the value of Υ (the ratio of total to visible stellar mass, which depends on the fraction of brown dwarfs) is set equal to 1 rather than 1.4 as in Cole et al. (2000). The recycled fraction in the calculation of the chemical enrichment is modified accordingly (Cole et al. 2000). Benson et al. (2000a) showed that in a ΛCDM cosmology, this model produces a real-space two-point galaxy correlation function which is remarkably similar to that measured in the Automated Plate Measurement (APM) survey by Baugh (1996), in contrast to the τCDM cosmology which fails to match the observed two-point correlation function on all scales. Furthermore, Benson et al. (2000a) showed that clustering predictions are robust to changes in the semi-analytic parameters, provided that the model matches the bright end of the local galaxy luminosity function. The evolution of the galaxy correlation function with redshift is in good agreement with smoothed particle hydrodynamics (SPH) simulations of galaxy formation (Pearce et al. 1999; Benson et al. 2001). In this paper, we consider only the ΛCDM model. We have checked that the parameters we have adopted do produce a correlation function of L_* galaxies identical to that of Benson et al. (2000a).

We use two different *N*-body simulations. The first one is the 'GIF' ΛCDM simulation, a full description of which may be found in Jenkins et al. (1998) and Kauffmann et al. (1999a). This is a 17×10^6 dark matter particle simulation in a cubic volume of side $141.3\,h^{-1}\,\mathrm{Mpc}$, with cosmological parameters $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$, $h = 0.7$, $\Gamma = 0.21$ and $\sigma_8 = 0.9$ (where Γ is the power spectrum shape parameter and σ_8 is the linearly extrapolated rms mass fluctuation in a sphere of radius $8\,h^{-1}\,\mathrm{Mpc}$). The mass of the smallest resolved halo in this simulation is $1.4 \times 10^{11}\,h^{-1}\,\mathrm{M}_\odot$. We also analyse the '$512^3$' simulation described by Jenkins et al. (2001) and Benson et al. (2000c), which has identical cosmological parameters to the GIF simulation (although it has a slightly different transfer function for the input power spectrum). The particle mass in this simulation is larger, by a factor of roughly 5, than in the GIF simulation, so that only dark matter haloes more massive than $7 \times 10^{11}\,h^{-1}\,\mathrm{M}_\odot$ are resolved, but the volume is approximately 40 times larger than in the GIF simulation. The 512^3 simulation is ideal for studying the brightest galaxies which are only found in very massive haloes and have low abundance. Benson et al. (2000c) noted that the dark matter correlation function in the 512^3 simulation differed slightly from that in the GIF simulation because of the large-scale power which is included in the larger volume but is missing from the smaller one. Throughout this paper, we apply a small correction to all the correlation functions determined from the GIF simulation by adding the quantity $\Delta\xi(r) = \xi_{512^3}(r) - \xi_{\mathrm{GIF}}(r)$ (where ξ_{512^3} and ξ_{GIF} are the correlation functions of the dark matter, in real or redshift space as appropriate, in the 512^3 and GIF simulations respectively). The required correction is not necessarily the same for galaxies and dark matter of course. However, the correction is at most 20 per cent in $\xi(r)$ over the range of scales considered in this work and, furthermore, ignoring it does not alter any of our conclusions.

3 IMAGES OF THE GALAXY DISTRIBUTION

Fig. 1 shows the distributions of dark matter and galaxies in slices through the GIF *N*-body simulation volume at six different redshifts. Each slice has dimensions $141.3 \times 141.3 \times 8\,h^{-3}\,\mathrm{Mpc}^3$. The dark matter is represented by the grey-scale, obtained by adaptively smoothing the *N*-body mass distribution. The shade intensity is proportional to the logarithm of the dark matter density in each pixel (the darker the pixel, the greater the projected density of the slice). Each galaxy brighter than $M_B - 5\log h = -19$ that formed in this region is depicted as a circle whose size is proportional to the rest-frame *B*-band absolute magnitude and whose colour indicates the rest-frame $B - V$ colour, as given by the scales at the top of the figure. Fig. 2 displays zoomed images, at three redshifts, of the areas delineated by the coloured boxes in Fig. 1: the region around a supercluster (left), and a more typical region (right).

3.1 The biased galaxy distribution

The images in Fig. 1 illustrate what a realistic distribution of galaxies might look like. At $z = 0$, this patch of universe is

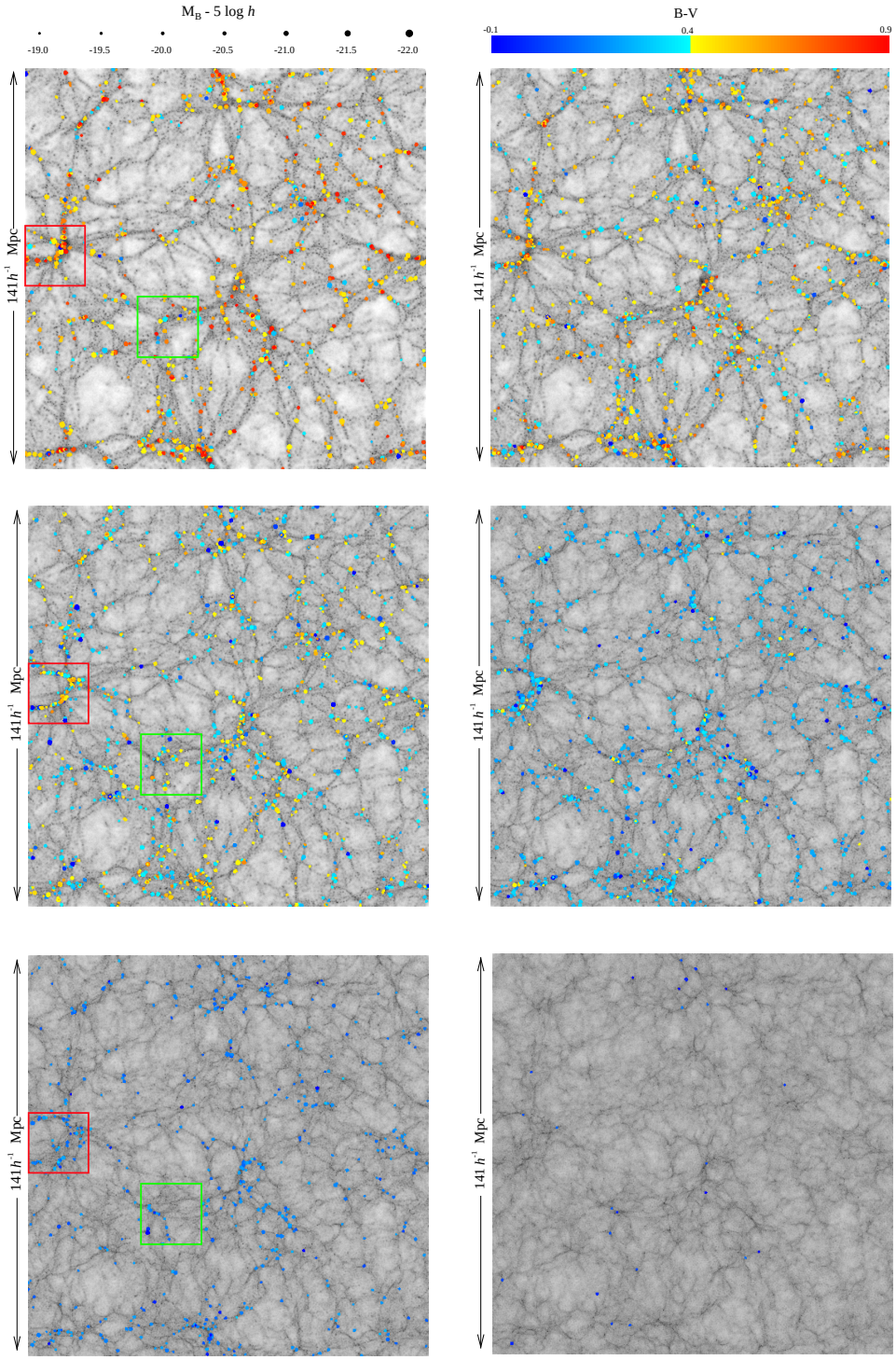

Figure 1. A slice through the N-body simulation volume at six redshifts: $z = 0.0$ (top left), 0.5 (top right), 1.0 (middle left), 2.0 (middle right), 3.0 (bottom left) and 5.0 (bottom right). Comoving coordinates are used. The region displayed has comoving dimensions of $141 \times 141 \times 8\,h^{-3}$ Mpc^3. The dark matter is represented as a grey-scale, with the densest regions darkest. The positions of the model galaxies are indicated by coloured circles whose size corresponds to the rest-frame B-band absolute magnitude of the galaxy, while their colour indicates the rest frame $B - V$ colour (see the key at the top of the figure). The red and green boxes indicate regions that are shown in greater detail in Fig. 2.

populated by galaxies with a wide range of colours. The dark matter has acquired the filamentary appearance characteristic of gravitational growth from cold dark matter initial conditions. The galaxies 'light up' the filaments and superclusters of the dark matter distribution, but are conspicuously absent from regions in which the density of dark matter is low. The emptiness of the voids is quite striking (cf. Peebles 2001). The voids would not be as empty of galaxies as they are if, instead of using the semi-analytic model, 'galaxies' were simply identified with randomly selected dark matter particles. Such a Poisson process inevitably places a

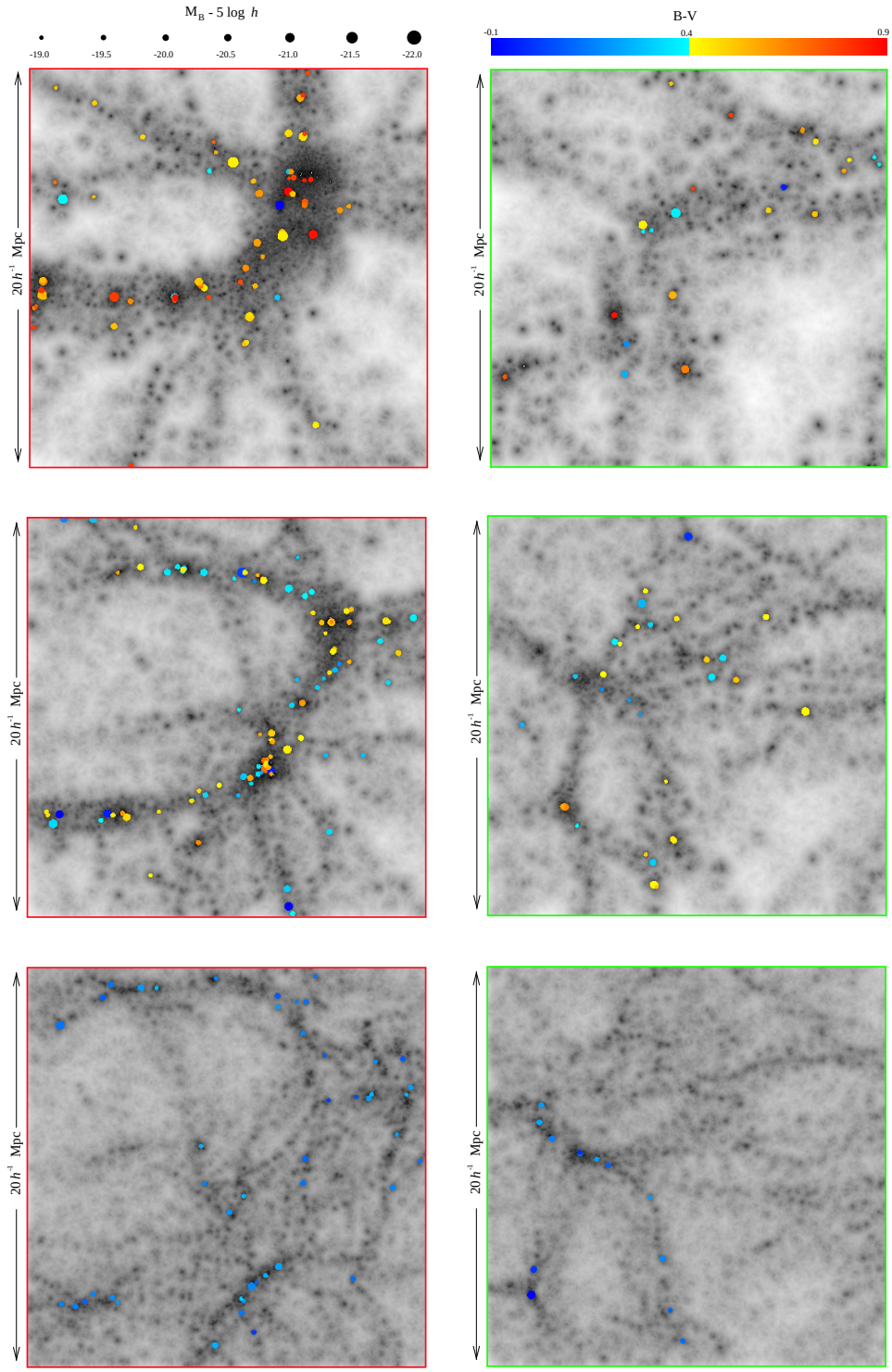

Figure 2. Slices through selected regions of the *N*-body simulation volume at three redshifts: $z = 0.0$ (top two panels), 1.0 (middle two panels) and 3.0 (bottom two panels). Comoving coordinates are used. Each region has comoving dimensions of $20 \times 20 \times 8\,h^{-3}$ Mpc^3. The dark matter is represented as a grey-scale, with the densest regions darkest. The positions of the model galaxies are indicated by coloured circles whose size corresponds to the rest-frame B-band absolute magnitude of the galaxy while their colour indicates the rest-frame $B - V$ colour of the galaxy (see the key at the top of the figure). The left-hand panels show the region around a $z = 0$ supercluster (indicated by the red box in Fig. 1), while the right-hand panels show a more typical region (as indicated by the green box in Fig. 1).

small fraction of galaxies in voids, but the physics of galaxy formation do not allow this: at the magnitude limit appropriate to Figs 1 and 2, voids are truly voids: no galaxies form in them.

While to the eye the distribution of galaxies in the figures may appear to follow the broad features of the dark matter distribution, it is impossible to judge in this way how faithful this tracing really is. That galaxies in CDM models might not be perfect tracers of the mass has been suspected for some time (Davis et al. 1985; Bardeen

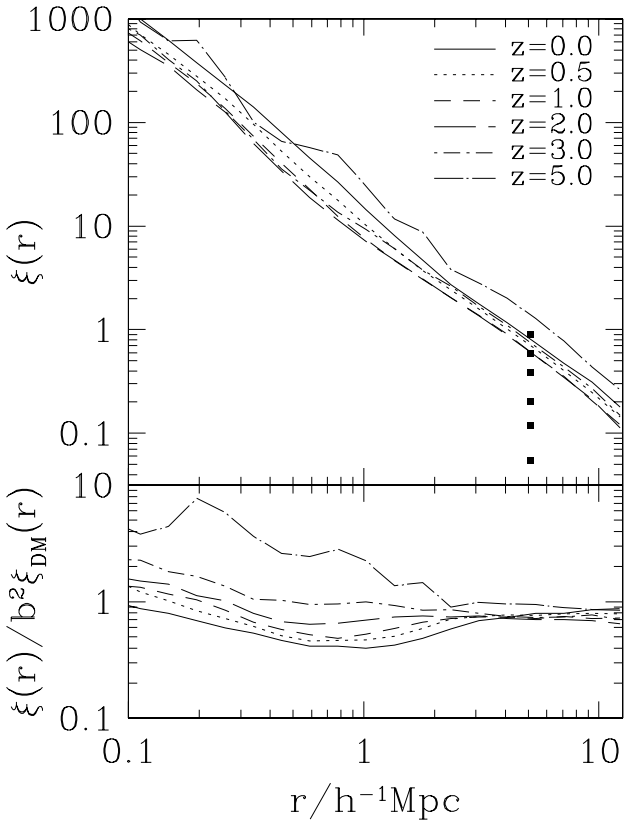

Figure 3. *Top panel:* the real-space correlation functions of galaxies brighter than rest-frame $M_B - 5\log h = -19$ for different redshifts are shown by lines (with redshifts as indicated in the legend). The separation, r, is in comoving coordinates. Filled squares give the value of the correlation function of *dark matter* at $r = 5h^{-1}$ Mpc at the same six redshifts ($z = 0$ at the top, $z = 5$ at the bottom). *Lower panel:* the ratio of the galaxy correlation function to the dark matter correlation function scaled by the analytically derived linear bias term, for the same six redshifts (see text for details).

et al. 1986) and was demonstrated explicitly for the GIF simulations by Kauffmann et al. (1999a) and Benson et al. (2000a). These studies showed, for example, that on scales below a few Mpc, galaxies in the simulations are less strongly clustered than the dark matter (or are 'anti-biased' relative to the mass). The upper panel in Fig. 3 shows the correlation functions of the galaxies for the different redshifts shown in Fig. 1 and demonstrates one aspect of the bias. The correlation functions of these galaxies have remarkably similar amplitudes, in contrast to the correlation function of dark matter (whose value at $r = 5\,h^{-1}$ Mpc is shown as solid points in Fig. 3), the amplitude of which evolves rapidly with redshift. In Fig. 3, we also compare our determinations of the correlation function to those derived from an analytical approach to the bias (cf. Baugh et al. 1999). For this, we compute the effective bias of the galaxy population as:

$$b(z) = \frac{1}{N}\sum_{i=1}^{N} b(M_i, z), \tag{1}$$

where N is the number of galaxies in the simulation, M_i is the mass of the halo in which the i^{th} galaxy is found, and $b(M, z)$ is the bias of dark matter haloes of mass M and redshift z, which we calculate using the fitting formula of Jing (1998) (which is based on the model of Mo & White 1996). An approximation to the galaxy correlation function is then $b(z)^2\xi_{\rm DM}(r, z)$, where $\xi_{\rm DM}$ is the correlation function of dark matter. [For the galaxy samples in our simulation, we find $b(z) = 1.07$, 1.24, 1.47, 2.05, 2.91 and 5.21 for $z = 0.0$, 0.5, 1.0, 2.0, 3.0 and 5.0 respectively.] In the lower panel of Fig. 3 we plot the ratio of the galaxy correlation functions in the simulation to the analytical approximation. On scales above a few Mpc, the analytical bias approach works well (it does slightly underestimate the correlation functions, but given the large biases present in our high-redshift samples, the approximation is actually rather good), but, as expected, it fails on smaller scales where our model predicts a scale-dependent bias.

Certain kinds of bias are readily apparent to the eye in Fig. 1. For example, the largest dark matter clumps at $z = 0$ are preferentially populated by red galaxies, while the field contains a mixture of galaxy colours. Similarly, the brightest galaxies in the region are also preferentially found at the centres of rich clusters. The first of these biases, the 'colour–density' relation is intimately related to a 'morphology–density' relation and is a natural outcome of hierarchical clustering, as we will discuss in more detail later. It has been investigated before in these simulations, in a somewhat different form, by Kauffmann et al. (1999a) and Benson et al. (2000c), who found that the two-point correlation function of elliptical galaxies is higher than that of spirals on small scales. This kind of bias is, of course, known to occur in the real universe (e.g. Davis & Geller 1976; Loveday et al. 1995).

Biases in the distribution of galaxies are the inevitable by-product of the complex physics of galaxy formation. They affect galaxies of different types in different ways and need to be understood before attempting to interpret cosmological observations.

3.2 The evolution of the galaxy population

The evolution of galaxies is driven by a number of processes. The most obvious one is the ageing of the stellar populations. Even if galaxies lived in isolation, stellar evolution would cause their luminosities and colours to evolve. However, no galaxy is an island: accretion, mergers and interactions are common. Many galaxies are observed to be forming stars today, possibly by converting new gas supplied externally (although see Benson et al. 2000b), thus increasing their mass and size. Others are observed to be expelling gas through galactic winds and tidal encounters. Nucleosynthesis in stars causes the metallicity of galactic gas and stars to evolve and this, in turn, affects the integrated stellar spectra.

The net effects of the various processes driving galaxy evolution are readily apparent in Fig. 1. Because young stars produce copious amounts of blue light, galaxies with high star formation rates relative to their total stellar mass appear bluer than galaxies with low relative star formation rates. At $z = 0$, the majority of the galaxies in the simulation are quite red because they have low relative star formation rates and are made primarily of old stars. As we look into the past, the appearance of the population rapidly changes. By $z = 1$, galaxies are much bluer because the typical relative star formation rate is higher than at $z = 0$ and the average stellar age is younger. Beyond $z = 1$, the galaxies remain blue, reflecting the youth of their stellar populations. The apparent star formation rate per unit volume declines at high redshift because fewer galaxies are seen above our magnitude selection, even though those that are seen still have high star formation rates. The detailed star formation history in our model is discussed in Cole et al. (2000). It is broadly in agreement with the star formation

history of the Universe as inferred from observations by, amongst others, Lilly et al. (1996), Madau et al. (1996) and Steidel et al. (1999). Although many quantitative details of the observations remain uncertain (because of complications arising, for example, from dust obscuration), the general behaviour seems to consist of a rapid rise in star formation rate between $z = 0$ and $z = 1$, followed by a slowly declining (or perhaps constant) star formation rate at higher redshifts. This is the kind of behaviour exhibited by our simulations.

As the images in Fig. 1 illustrate, in hierarchical models of galaxy formation the number of galaxies is constantly changing. Galaxies are born as new dark matter haloes form and gas is able to cool in them and turn into stars. The population is depleted when galaxies merge together. Of course, the number of galaxies detected in a particular survey will depend crucially on the selection criteria. All these effects can be seen clearly in the images of our simulations. At $z = 5$, there are very few galaxies present because only a handful of massive dark haloes have had time to collapse. In those that have, the galaxies have had little time to form stars, while feedback from supernovae has strongly suppressed star formation in small haloes. The majority of the galaxies seen in the images at this epoch occur in haloes of mass $10^{11-12}\,h^{-1}\,\mathrm{M}_{\odot}$ and have stellar masses of a few times $10^{9}\,h^{-1}\,\mathrm{M}_{\odot}$; the very brightest galaxies are found in the tail of haloes extending to masses close to $10^{13}\,h^{-1}\,\mathrm{M}_{\odot}$. By $z = 3$ the number of galaxies has increased significantly, as more haloes have collapsed and more galaxies have been able to form. At this epoch several extremely bright (in the *B* band) galaxies are visible. The increased number of galaxies in the image is owing, in part, to our selection in the *B* band which is sensitive to the star formation rate. At $z = 2$ there is a noticeable increase in the abundance of galaxies as structures continue to form. The most obvious change from $z = 2$ to $z = 1$ is a substantial reddening of the galaxies, a trend which continues to $z = 0$ as star formation rates decline and stellar populations age.

Many faint blue galaxies are formed in the filamentary network of the dark matter. In the comoving coordinates of Fig. 1, galaxies move rather little between $z = 5$ and the present. For example, the progenitor of the large supercluster marked by a red box at $z = 0$ is already clearly visible at $z = 3$ as a concentration of young galaxies. In other words, by virtue of forming in the highest density regions, galaxies are strongly biased at birth. This is a fundamental outcome of hierarchical clustering (Kaiser 1984; Davis et al. 1985). It underlies the results of Baugh et al. (1998) and Governato et al. (1998), who argued that Lyman-break galaxies at $z = 3$ would be expected to be strongly clustered, as was subsequently found to be the case observationally (Adelberger et al. 1998).

4 QUANTITATIVE ANALYSIS AND COMPARISONS WITH OBSERVATIONS

In this section, we compare the properties of our model with the limited observational data currently available. More stringent comparisons will be possible with the forthcoming 2dF and Sloan galaxy redshift surveys. We consider, in turn, the variation of the clustering length with the luminosity of a sample, the morphology–density relation and the topology of the galaxy distribution.

4.1 Correlation length versus separation

The evolution of the two-point correlation function of *dark matter*

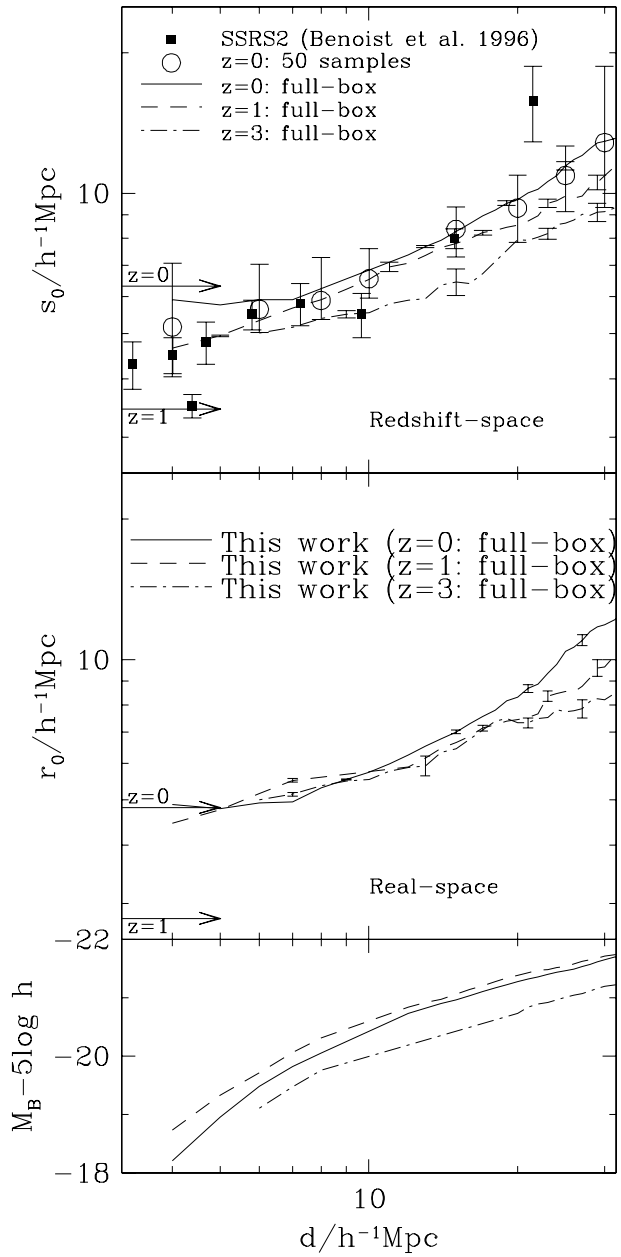

Figure 4. The redshift-space (top panel) and real-space (middle panel) correlation lengths of galaxies as a function of the mean galaxy separation (in redshift and real space respectively). Filled squares in the top panel show the measured correlation lengths in the SSRS2 survey (Benoist et al. 1996). The solid lines show our model results for galaxies selected by their (dust-extincted) *B*-band magnitude at $z = 0$, with error bars indicating the statistical uncertainty. Open circles in the top panel show the median redshift-space correlation lengths estimated from 50 subsamples from the simulations, each of volume equal to that of a volume-limited SSRS2 sample of the same absolute magnitude. Error bars indicate the 10 per cent and 90 per cent intervals of the correlation length distribution. Dashed and dot-dashed lines give our model predictions for galaxies at $z = 1$ and $z = 3$ respectively, selected according to their rest-frame *B*-band magnitude. Horizontal arrows indicate the correlation lengths of the dark matter at $z = 0$ and $z = 1$. The lower panel indicates the dust-extincted *B*-band absolute magnitude corresponding to a given mean galaxy separation at $z = 0$ (solid line), $z = 1$ (dashed line) and $z = 3$ (dot-dashed line).

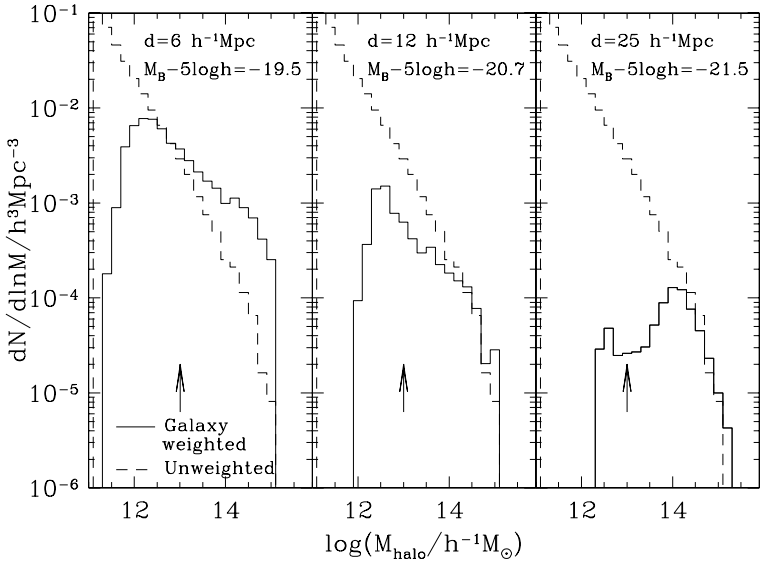

Figure 5. Dark matter halo mass functions weighted by the number of galaxies in each halo at $z = 0$ (solid histograms). Results are shown for galaxies brighter than three different dust-extincted B-band magnitudes, chosen to give a desired mean inter-galaxy separation, d, as indicated in each panel. We show the (unweighted) mass function of dark matter haloes in each panel for comparison (dashed histograms). The vertical arrow in each panel shows the location of M_* [defined by $\sigma(M_*) = \delta_c$].

in CDM models is now well established. In the linear regime, it evolves according to the linear growth factor; in the non-linear regime, its evolution can be calculated accurately using N-body simulations (see, for example, Jenkins et al. 1998). By contrast, the evolution of the two-point correlation function of galaxies has only begun to be investigated in detail recently. In Fig. 4 we plot the redshift-space correlation lengths, s_0 (top panel), and the real-space correlation lengths, r_0 (middle panel), of galaxies brighter than a particular rest-frame (dust-extincted) B-band magnitude, as a function of their mean separation, d, at $z = 0$ (solid line), $z = 1.0$ (dashed line) and $z = 3$ (dot-dashed line). For reference, the lower panel of Fig. 4 shows the relation between mean separation and absolute B-band magnitude. Statistical errors [obtained by assuming Poisson counting statistics to estimate the error in $\xi(r)$ and propagating this error through to the determination of r_0] are shown by the error bars. At small pair separations, we show results from the GIF simulation (applying the small correction for finite volume effects discussed above), and at large pair separations we use the 512^3 simulation which is more accurate on large scales. (The sudden decrease in the size of the errorbars at $\approx 15\,h^{-1}$ Mpc is a result of this change of simulation volume.) For reference, we show the correlation lengths of the dark matter at $z = 0$ and $z = 1$ as horizontal arrows (at $z = 3$ the dark matter correlation length is 1.1 and $1.5\,h^{-1}$ Mpc in real and redshift space respectively).

We compare our model predictions with results from the Southern Sky Redshift Survey 2 (SSRS2) (Benoist et al. 1996), which are shown by filled squares in Fig. 4. These data exhibit a nearly constant redshift-space clustering length for galaxies fainter than $M_B - 5\log h = -20$, followed by a rapid increase at brighter magnitudes. Benoist et al. (1996) compared their data to two simple models of galaxy bias. In the first, they assigned a dark matter halo mass to galaxies in their sample using the Tully–Fisher and Faber–Jackson relations, and then applied the techniques of Mo & White (1996) to compute the bias of these haloes (and hence of the galaxies which occupy them). This model provides a reasonable match to the observed behaviour of the faint galaxies in the survey, but it is unable to reproduce the strong luminosity-dependent bias observed for galaxies brighter than L_*. Their second model is based upon the work of Bernardeau & Schaeffer (1992), who developed a description of bias from the non-linear evolution of the density field. This model is able to match the luminosity-dependent bias for bright galaxies, but predicts too strong a relation for faint galaxies and so is also ruled out by the data. Our model of galaxy clustering, on the other hand, does produce a trend similar to that observed, namely a relatively constant clustering length for small d followed by a rise in correlation length for the rarest objects. From Fig. 4 it appears that our model is not consistent with the data over the whole range of separations. However, this discrepancy may result from sampling variance in the observations, as we discuss below. First, we explain how this trend arises in our models.

It is well known that dark matter haloes are biased relative to the underlying mass, with the most massive haloes being the most strongly clustered (e.g. Frenk et al. 1988; Mo & White 1996). Thus, for the brightest galaxies to be the most strongly clustered, it is necessary that they should preferentially inhabit more massive dark matter haloes than those occupied by their lower-luminosity counterparts. In Fig. 5 we show mass functions of dark matter haloes, weighted by the number of occupant galaxies (solid lines) and unweighted (dashed lines), for three values of d. In each panel, the solid histogram gives the galaxy-weighted halo mass function when galaxies are selected by their dust-extincted B-band magnitude. The vertical arrow in each panel shows the location of M_*, defined by $\sigma(M_*) = \delta_c$, where $\sigma(M)$ is the mass variance in spheres containing a mass M on average, and δ_c is the critical overdensity for collapse in the spherical top-hat model. A simple understanding of the galaxy bias for each sample may be gained from this figure. For the two smaller values of d, the relative numbers of galaxies in highly biased, cluster-sized haloes ($M_{\rm halo} \gtrsim 10^{14}\,h^{-1}\,{\rm M}_\odot$) and in weakly-biased, galaxy-sized haloes ($M_{\rm halo} \sim 10^{12}\,h^{-1}\,{\rm M}_\odot$) are comparable in the two samples. As a result, these two samples have quite similar correlation lengths. However, for the sparser sample with $d = 25\,h^{-1}$ Mpc, the relative number in cluster-sized haloes is much higher than in the two other cases. As

a result, this sample has a larger correlation length than the other two. (For this very bright sample, we find that many galaxies occupying haloes near the peak mass of $M_{\rm halo} \sim 10^{14} h^{-1}\,{\rm M}_{\odot}$ have undergone a recent burst of star formation.)

As noted above, our model is not in perfect agreement with the data of Benoist et al. (1996). The SSRS2 survey, however, covers a relatively small volume and so sample variance may not be negligible. Benoist et al. (1996) estimated the effects of sample variance on their results and concluded that the luminosity dependence of s_0 for faint (sub-L_*) galaxies could well be a result of sample variance, while that for brighter galaxies seemed to be a real effect. We can estimate the size of this uncertainty directly from our simulations, which cover a much larger volume than the real survey. For a given value of d, we extract 50 randomly placed cubic regions of volume equal to that of a volume-limited SSRS2 catalogue cut at the same absolute magnitude from the simulation. We then measure the correlation length in each of the 50 cubes. In Fig. 4 we plot the median correlation lengths from the cubes as open circles, with error bars indicating the 10 per cent and 90 per cent intervals of the distribution. (We note that even for the largest value of d shown, the number of independent SSRS2 volumes that fit within the 512^3 simulation cube is still reasonably large, $\sim$20, and so our estimates of sample variance at these separations should still be accurate.) Evidently, sample variance in a survey the size of the SSRS2 is large, and this can account for the differences with our model. At the smallest values of d, the median value of s_0 from the subsamples is biased low relative to that measured in the full simulation volume, because a large fraction of the clustering signal comes from galaxies in and around a few large clusters and these are often missing from small volumes cut out of the simulation box. At larger separations, the main source of sample variance is the low abundance of the brightest galaxies. Our estimate of sample variance confirm the conclusion reached by Benoist et al. (1996), namely that the luminosity dependence of s_0 observed for sub-L_* galaxies in the SSRS2 is caused by sample variance, while that for brighter galaxies is real.

The evolution of the correlation lengths of galaxies of different abundance is also illustrated in Fig. 4. For samples selected according to our specific criteria (i.e. according to rest-frame B-band absolute magnitude), the redshift-space correlation lengths at $z = 1$ and $z = 3$ are slightly smaller than at $z = 0$. In real space, the evolution is even weaker, with only the brightest galaxies showing a modest reduction in clustering strength at $z \geq 1$. In detail, this behaviour reflects a complex interplay between the growth of the galaxy population, the evolution of its clustering pattern and the selection criteria that we have applied. In any case, the variation in the correlation length of the galaxies is much smaller than the variation in the correlation length of the dark matter, indicated by the horizontal arrows in Fig. 4. Thus, galaxies selected according to our criteria are more strongly biased at $z = 1$ and $z = 3$ than at $z = 0$, even for quite small values of d. The images of the galaxy distribution displayed in Fig. 1 show exactly how this effect arises. The galaxies present at the highest redshift, $z = 3$, have formed in regions that are destined to become incorporated into clusters or superclusters of galaxies by the present day. Such regions are amongst the most overdense at high redshift, and so small-scale density fluctuations tend to collapse earlier there than in less overdense regions (Kaiser 1984; Davis et al. 1985; Bardeen et al. 1986). Forming as they do in the most highly biased regions of the universe, galaxies at high redshift naturally end up being strongly biased themselves. In conclusion, our model predicts a strong luminosity dependence of clustering length for the brightest galaxies from $z = 0$ to 3, an effect which is seen in both real and redshift space.

4.2 Morphology–density relation

The existence of a correlation between galaxy morphology and local environment is a remarkable feature of the galaxy population. Dressler (1980) showed that the fraction of galaxies of different morphological types is strongly correlated with the local galaxy density: elliptical and S0 galaxies are found preferentially in high-density regions while spiral galaxies are found preferentially in low-density regions.

Early N-body simulations suggested that a morphology–density relation is a natural outcome of hierarchical clustering from CDM initial conditions (Frenk et al. 1985, 1988). This was explicitly shown to be the case by Monte Carlo based semi-analytic modelling (Kauffmann 1995; Baugh et al. 1996b). These calculations had no information on the spatial distribution of galaxies and so the relation they established is one between morphology and cluster mass, rather than between morphology and galaxy density. The implementation of semi-analytic techniques in high-resolution N-body simulations allowed the radial distributions of different kinds of galaxies inside large clusters to be calculated for the first time (Springel et al. 2001; Okamoto & Nagashima 2001). This work has shown, for example, that galaxies of different colours are spatially segregated within the cluster.

The emergence of a colour–density relation is clearly illustrated in the images of Fig. 1. Redder galaxies (with $B - V \gtrsim 0.7$) which, in our model, are primarily ellipticals or S0s (see fig. 12 of Cole et al. 2000), are over-abundant in the most overdense regions relative to the field (e.g. the large supercluster in the middle of the left-hand edge, or the large cluster near the centre of the bottom edge). The correlation between galaxy colour (or morphology) and environment is a by-product of the biases discussed in the preceding subsection: the oldest, reddest galaxies form in the highest density regions where the production of elliptical galaxies by mergers is also favoured.

To quantify the morphology–density relation apparent in Fig. 1, we proceed in a manner analogous to the analysis of an observational sample. We use a technique patterned on that employed by Postman & Geller (1984). First, we apply a friends-of-friends group-finding algorithm to the real-space distribution of model galaxies brighter than a particular B-band absolute magnitude, using different values of the linking length. [Postman & Geller (1984) applied their group finder in redshift space, but used an anisotropic linking length to account for distortions in the redshift direction. As we have a galaxy catalogue in real space, we perform group-finding there, avoiding the complications of redshift-space distortions.] In this manner, we build up nested sets of groups as a function of the enclosed density (we consider three or more galaxies linked together to be a 'group'). For very large linking lengths all galaxies will belong to a group, but as the linking length is decreased each galaxy will at some point no longer be a member of a group. We assign each galaxy a local density corresponding to the surface density of the group of which it was last a member. The local density at the surface of a group formed with a linking length r is approximately $n = 3/2\pi r^3$ (e.g. Lacey & Cole 1994). The morphological type of the galaxy is assigned according to our standard definitions based on dust-extincted B-band bulge-to-total luminosity ratios: galaxies with $B/T_B < 0.4$ are labelled as spirals, while those with $B/T_B \geq 0.4$ are labelled as elliptical (E)/S0 (Cole et al. 2000). We constructed

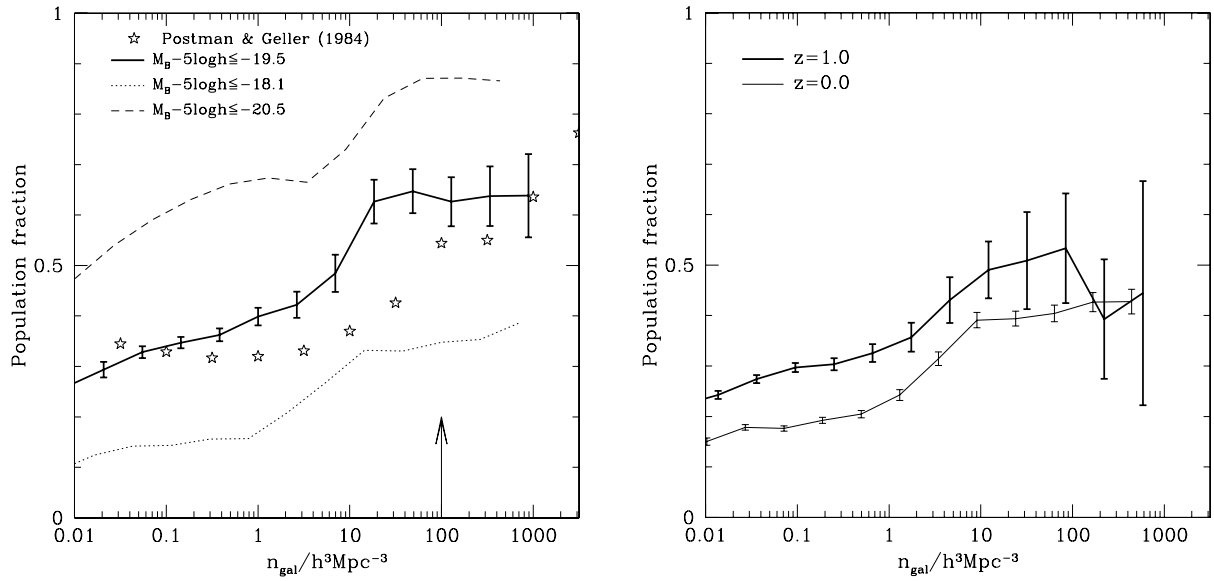

Figure 6. The morphology–density relation for galaxies selected according to their observed **B**-band magnitude. Stars show the fraction of elliptical and S0 galaxies in the CfA redshift survey (Postman & Geller 1984). We plot the group fractions of Postman & Geller (1984) for densities less than $100\,h^3\,\mathrm{Mpc}^{-3}$ (as indicated by the vertical arrow), and their cluster fractions for higher densities (where they define groups and clusters as associations of 3–9 and ≥10 galaxies respectively). In the left-hand panel, the heavy solid line is our model prediction for the elliptical/S0 fraction (i.e. galaxies with a dust-extincted B-band bulge-to-total ratio, $B/T_B > 0.4$) at $z = 0$, obtained from the galaxy distribution in real space. The galaxies themselves are selected to have $M_B - 5\log h \leq -19.5$, but the densities are extrapolated to $M_\mathrm{B} - 5\log h = -17.5$ using the CfA survey luminosity function of Postman & Geller (1984). Dotted and dashed lines are the corresponding model results for galaxies brighter than the completeness limits of the GIF and 512^3 simulations ($M_B - 5\log h = -18.1$ and $M_B - 5\log h = -20.5$ respectively), and are also corrected to $M_B - 5\log h = -17.5$. In the right-hand panel, the heavy solid line shows the model result at $z = 1.0$ for galaxies above the simulation completeness limit, but here the densities are left uncorrected and so correspond to $M_B - 5\log h = -18.4$. The thin solid line shows the relation at $z = 0$ for galaxies brighter than $M_B - 5\log h = -18.4$ for comparison. Error bars show 1σ deviations.

density–morphology relations for samples with different limiting absolute magnitudes. Postman & Geller (1984) measured the morphology–density relation from a magnitude limited sample for which the galaxy density has strong radial variations due to the magnitude limit. To account for this affect they applied a correction to obtain the corresponding galaxy density at $M_\mathrm{B} - 5\log h = -17.5$ using the CfA survey luminosity function. We will therefore correct model galaxy densities in the same way, namely we multiply the densities by a factor

$$f_\mathrm{n} = \frac{\int_{L_\mathrm{ref}}^{\infty} \Phi(L)\,\mathrm{d}L}{\int_{L_0}^{\infty} \Phi(L)\,\mathrm{d}L}, \qquad (2)$$

where L_0 and L_ref are the luminosities corresponding to $M_B - 5\log h \leq -19.5$ and -17.5 respectively and $\Phi(L)$ is the luminosity function of the CfA survey. (Note that this simply shifts the curves of Fig. 6 to the right.)

The left-hand panel of Fig. 6 shows the model morphology–density relation at $z = 0$ for three absolute magnitude cuts: $M_B - 5\log h = -18.1$ (dotted line; the completeness limit of the GIF simulation), -19.5 (thick solid line; close to L_*) and -20.1 (dashed line; the completeness limit of the 512^3 simulation). It is immediately apparent that our model does display a morphology–density relation with the correct trend: E/S0s are more common in high-density environments. For high densities ($\gtrsim 100$ galaxies h^3 Mpc^{-3}) our model shows no relation, because, by construction, no morphology–density relation can exist within individual haloes because of the way in which we assign galaxies to dark matter particles. (We do not plot model results for densities greater than $1000\,h^3\,\mathrm{Mpc}^{-3}$ as these begin to probe single dark matter haloes in our simulation, resulting in a poor determination of the morphology–density relation.) The model relation for $M_\mathrm{B} - 5\log h < -19.5$ shows a distinct jump between densities of 10 and $20\,h^3\,\mathrm{Mpc}^{-3}$. This corresponds to a transition between haloes containing a single galaxy (which have masses $\lesssim 10^{12.5}\,h^{-1}\,\mathrm{M}_\odot$ and dominate the population of galaxies in low-density environments) to those containing more than one galaxy (i.e. groups and clusters) which dominate at high densities. The right-hand panel of Fig. 6 shows that our model predicts a very similar morphology–density relation at $z = 1$ as at $z = 0$.

We can now elaborate a little on the cause of the morphology density relation in our model, focusing on the morphology–density relation at low densities (where the spatial distribution of morphological types within individual haloes is unimportant and the local density is typically determined by averaging over regions containing many dark matter haloes). The morphological mix of galaxies in a single dark matter halo can depend only upon the mass of that halo as this statistically determines the merger history and galaxy formation history in the halo. Therefore, for a morphology–density relation to exist: (i) there must be a relation between dark matter halo mass and local galaxy density and (ii) there must be a dependence of morphological fraction on halo mass. The first of these requirements is naturally met in hierarchical cosmologies as the most massive haloes are preferentially found in the densest environments. The left-hand panel of Fig. 7 shows the local galaxy number density (for galaxies brighter than $M_\mathrm{B} - 5\log h = -19.5$ but corrected to -17.5 as before) at the centres of dark matter haloes as a function of the halo mass (i.e. we plot the density of the three-particle group, as defined above, which contains the central galaxy). Clearly, the densest regions of the

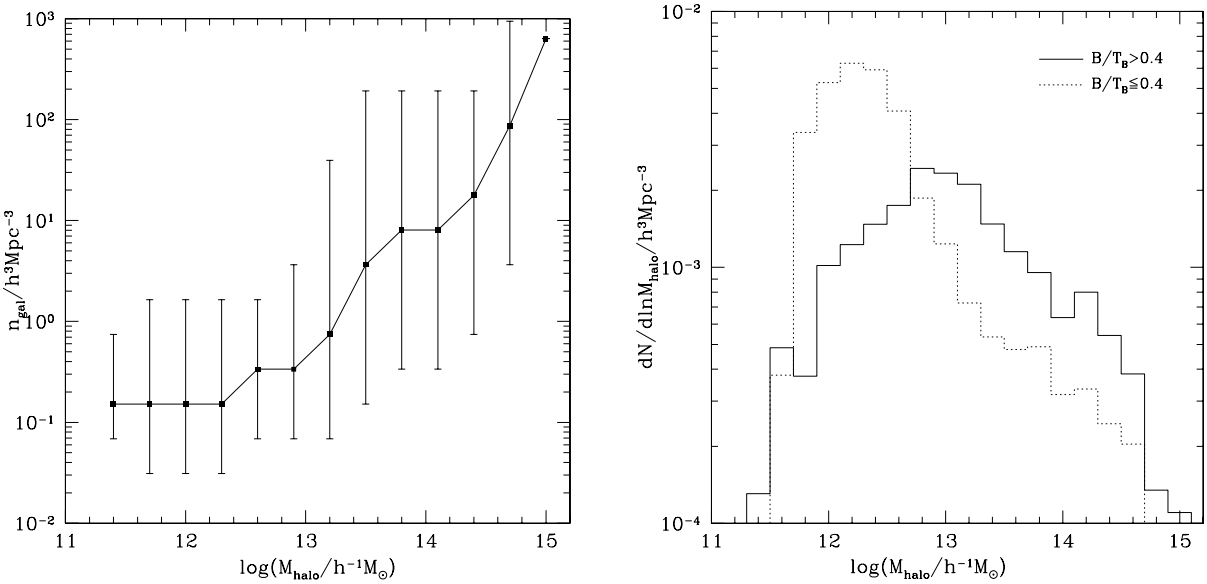

Figure 7. *Left-hand panel:* The local galaxy number density (for galaxies brighter than $M_B - 5\log h = -19.5$, but corrected to $M_B - 5\log h = -17.5$) as a function of dark matter halo mass. Points show the median number density in each halo mass bin, while error bars show the 10 per cent and 90 per cent intervals of the distribution. *Right-hand panel:* dark matter halo mass functions weighted by the number of galaxies brighter than $M_B - 5\log h = -19.5\,\mathrm{halo}^{-1}$. The solid histogram is the mass function for galaxies with $B/T_B > 0.4$, while the dotted histogram shows the mass function for galaxies with $B/T_B < 0.4$.

galaxy population are associated with cluster-sized haloes. The right-hand panel of Fig. 7 shows that our model also meets the second requirement. This figure shows the dark matter halo mass function weighted by the number of galaxies of a particular morphological class per halo. The halo mass function for elliptical/S0 (E/S0) galaxies is shifted to higher mass haloes relative to that for spiral (S) galaxies. The resulting dependence of the morphological mix on the halo mass merely reflects the fact that galaxy formation in clusters is accelerated relative to the field, allowing enough time for galaxy–galaxy mergers to produce a large population of elliptical galaxies.

Fig. 6 also shows that the model morphology–density relation depends upon the absolute magnitude at the which the galaxies are selected. This property is simply a reflection of a morphology–luminosity relation that is present in our model. This property complicates the comparison with observational data because existing analyses are usually based on apparent magnitude limited samples. In the figure we compare our model predictions with the density–morphology relation measured in the CfA survey by Postman & Geller (1984). For densities less than $100\,h^3\,\mathrm{Mpc}^{-3}$ (indicated by the vertical arrow), we plot the E/S0 fraction in 'groups' (defined as associations of three to nine galaxies by Postman & Geller 1984), and for larger densities, we plot the fraction in clusters (associations of 10 or more galaxies). This shows that our model displays qualitatively similar behaviour to the observational data. A more detailed comparison may be possible with the new generation of large redshift surveys.

Finally, another important prediction of our model is that there should be a strong morphology–density relation well-established already at $z = 1$, as shown in the right-hand panel of Fig. 6. At this high redshift, the model relation is qualitatively similar to that at $z = 0$.

4.3 The genus curve for the topology of the galaxy distribution

The two-point correlation function contains only low-order information about the spatial distribution of galaxies. To fully specify this distribution requires determining its higher order clustering properties. Alternatively, the genus, a measure of the topology of a smooth density field, provides a statistic that is sensitive to all of the higher order moments of the distribution (Gott, Melott & Dickinson 1986; Gott, Weinberg & Melott 1987).

The genus is defined as the number of topological holes minus the number of isolated regions of an isodensity surface. By varying the density at which this surface is placed, a genus curve can be constructed. The genus curve has the interesting property that an exact, analytic expression exists for the special case of a Gaussian random density field, namely

$$g(\nu) = A(1 - \nu^2)\exp\left(-\frac{\nu^2}{2}\right), \tag{3}$$

where g is the genus per unit volume and ν is defined by

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

where f is the fraction of the volume above the density threshold and erf^{-1} is the inverse of the error function. (For a Gaussian random field, but not for any other field, this definition implies that ν^2 is the variance of the field.) The amplitude, A, depends only on the second moment of the power spectrum of the smoothed density field. In general, if the field is not Gaussian, the shape of the genus curve may differ from equation (3). With the above definition for ν, the genus curve remains the same under any dynamical evolution or biasing in which the initial and final densities at each Eulerian point are related by a monotonic, one-to-one mapping. Thus, this topological measure provides, in principle, a method for testing whether or not density fluctuations in the early universe were originally random and Gaussian, as predicted in generic inflationary models.

The genus curve has been determined for both N-body simulations of dark matter and for surveys of galaxies in the real Universe by many authors (e.g. (Gott et al. 1989; Moore et al. 1992; Park, Gott & da Costa 1992; Rhoads, Gott & Postman 1994;

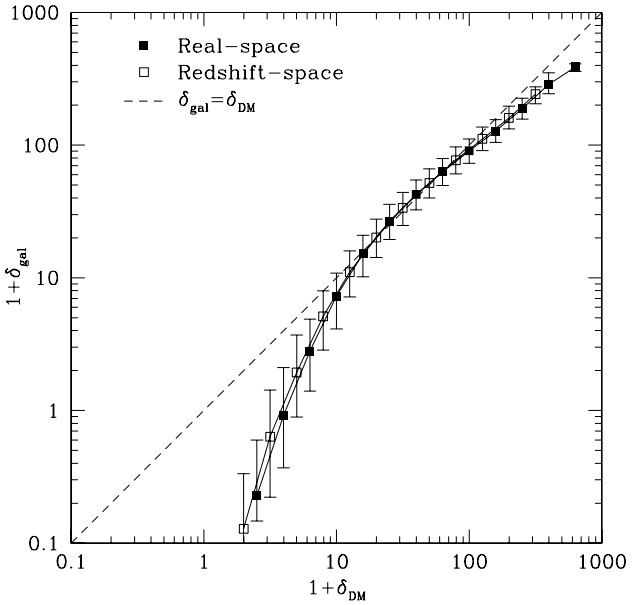

Figure 8. The relation between the present-day galaxy and dark matter overdensities in real space (filled symbols) and redshift space (open symbols). Only galaxies brighter than $M_B - 5\log h = -18.5$ are considered and the two fields have been smoothed with the Gaussian filter of equation (5), with $\lambda_e = 6.0\,h^{-1}$ Mpc. Points show the median galaxy overdensity at each dark matter overdensity, and the error bars show the 10 per cent and 90 per cent intervals of the distribution of $1 + \delta_{gal}$. When smoothed on this scale, redshift-space distortions make little difference to the biasing relation.

Vogeley et al. 1994; Canavezes et al. 1998; Springel et al. 1998; Canavezes & Sharpe 2001.) The genus curve for galaxies would be identical to that of the underlying dark matter if the galaxy and dark matter density fields were related by a monotonically varying factor, i.e. the bias. The bias relation (i.e. the galaxy overdensity as a function of dark matter overdensity) in our model is shown in Fig. 8 for galaxies brighter than $M_B - 5\log h = -18.5$. It is qualitatively similar to the relations found by Somerville et al. (2001) using similar techniques. The symbols show the median relation which is clearly monotonic. However, there is substantial scatter around this relation, as a result of which it is no longer guaranteed that the genus curve for galaxies will be identical to that of the dark matter.

We have used the technique described by Coles, Davies & Pearson (1996) to measure the genus of both dark matter and galaxies in our simulation. Fig. 9 shows the results for galaxies (dashed lines) brighter than $M_B - 5\log h = -18.5$ at $z = 0$ (upper panels) and $z = 1.0$ (lower panels) and also for dark matter (solid lines). These curves were calculated by smoothing the dark matter and galaxies in redshift-space onto a 128^3 grid using a Gaussian filter of the form

$$W(r) = \frac{1}{\pi^{3/2}\lambda_e^3}\exp\left(-\frac{r^2}{\lambda_e^2}\right), \tag{5}$$

as is conventional in the literature on this subject. Smoothing lengths of $\lambda_e = 6.0$ and $\lambda_e = 8.0\,h^{-1}$ Mpc were chosen to match those used by Vogeley et al. (1994) in their analysis of the CfA surveys. These smoothing lengths are over five times larger than the size of the grid cells on which the fields are tabulated and so the finite resolution of the grid has no effect on the calculation of the genus curve (Springel et al. 1998). Note that the mean comoving separation of galaxies in our simulation is 4.3 and $5.6\,h^{-1}$ Mpc at $z = 0$ and $z = 1.0$ respectively (corresponding to 34 000 and 16 000 galaxies in the simulation volume). At $z = 1$ this is just smaller than the minimum smoothing scale, thereby providing the greatest number of independent resolution elements without allowing discreteness effects to become too large (Weinberg, Gott & Melott 1987). We estimate the errors on each genus curve by bootstrap resampling of the galaxy catalogues. Moore et al. (1992) find that this procedure produces slightly larger errors than those estimated by considering several realizations of a mock catalogue.

In a ΛCDM cosmology, *N*-body simulations have shown that the genus curve for the dark matter displays both a 'bubble shift' (i.e. a shift to the right) with respect to the Gaussian random phase genus curve and an amplitude reduction relative to a Gaussian random density field having the same power spectrum (Springel et al. 1998). These two effects can be seen by comparing the dark matter genus curve to the two random phase genus curves (i.e. with the shape given by equation 3), shown in Fig. 9 by dotted and dot-dashed lines. The amplitude of the dotted curve is chosen to best fit (in a least-squares sense) the measured dark matter genus curve, while the amplitude of the dot-dashed curve is that expected for a Gaussian random field with the same power spectrum as the dark matter. The dark matter genus curve is displaced to the right of the dotted curve, showing the bubble shift, and has smaller amplitude than the dot-dashed curve, showing the amplitude drop. The genus curve for galaxies brighter than $M_B - 5\log h = -18.5$ shows definite differences from that for the dark matter. Firstly, no bubble shift exists for the galaxy genus curve which, instead, exhibits a small 'meatball shift' (i.e. shift to the left). The galaxy curve also has a systematically larger amplitude than the dark matter curve. Canavezes et al. (1998) advocate the amplitude drop (defined as the ratio of the amplitudes of the best-fitting random phase genus curved for the actual density field and that of a Gaussian random field with the same power spectrum) as a useful measure of the degree of phase correlation in the galaxy density field. To measure the amplitude drop, we 'Gaussianize' the galaxy density field (i.e. we take it to Fourier space, randomize the phases subject to the reality condition, $\delta_k = \delta^*_{-k}$, and then restore it to real space). The amplitude drop is then simply the ratio of amplitudes of the best-fitting random phase genus curves for the original and Gaussianized density fields. At $z = 0$ we find amplitude drops of $R = 0.70$ and $R = 0.84$ for the dark matter in redshift space, for $\lambda_e = 6.0$ and $8.0\,h^{-1}$ Mpc respectively. (In real space we measure amplitude drops of 0.60 and 0.67 for the same two smoothing lengths, in good agreement with the determinations of Springel et al. 1998.) For galaxies at $z = 0$ we find $R = 0.84 \pm 0.02$ and $R = 0.90 \pm 0.02$ also for these same two smoothing scales. At $z = 1$ (lower panels in Fig. 9) the amplitude drops are somewhat smaller, $R = 0.90 \pm 0.03$ and $R = 0.93 \pm 0.07$ for $\lambda_e = 6.0$ and $8.0\,h^{-1}$ Mpc respectively, as phase correlations caused by non-linear growth of structure have not had as long to develop as at $z = 0$.

As noted above, galaxies would have exactly the same genus curve as the dark matter if there were a one-to-one mapping between dark matter and galaxy density fields which preserved the density ranking. However, we do see significant differences between the galaxy and dark matter genus curves. This must be caused either by the scatter in the biasing relation between galaxies and dark matter, or by systematic biases arising from the relatively small number of galaxies ($\sim 10^4$) compared to dark matter particles ($\sim 10^7$) in our samples. To test this latter possibility, we extracted

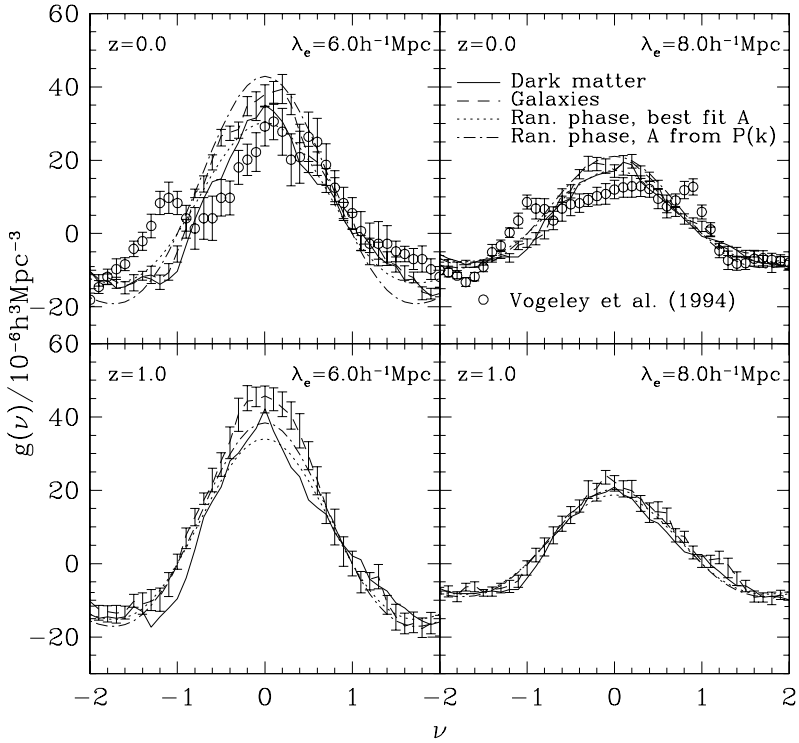

Figure 9. The genus per unit volume for two smoothing lengths, $\lambda_e = 6.0$ and $8.0\,h^{-1}$ Mpc, and two redshifts, $z = 0$ and $z = 1.0$, as indicated in the panels. Open circles show the results from the combined CfA-I and CfA-II redshift surveys (Vogeley et al. 1994). Solid lines show the genus curves of dark matter in our simulations, while the dotted lines show the random phase genus curve which best fits the dark matter and the dot-dashed line shows the genus curve for a Gaussian random field with the same power spectrum as the dark matter. Dashed lines show the genus curves of galaxies brighter than $M_B - 5\log h = -18.5$. All model curves are calculated in redshift space. Error bars on the model galaxy genus curves are the standard deviations from 30 bootstrap resamplings of the galaxy distribution. (Note, however, that the line indicates the genus curve of the actual distribution, not the mean of the bootstrapped samples.)

20 random samples of dark matter particles with the same abundance as the galaxies in our catalogue and computed their genus curves. We find that this sparse sampling is the primary cause of the differences between the galaxy and dark matter genus curves. Just as for the galaxy sample, the sparsely sampled dark matter shows no evidence for a bubble shift (and agrees closely with the galaxy genus curve for $\nu < 0$) and also shows a higher genus curve amplitude compared to the fully sampled dark matter distribution. We conclude therefore that any differences between the genus curves for dark matter and galaxies as a result of the stochastic bias of Fig. 8 are negligible compared to the effects of sparse sampling of the galaxy density field. This sparse sampling at present severely limits the usefulness of the genus statistic for quantifying the Gaussianity of the initial dark matter distribution.

We also show in Fig. 9 the genus curve measured for the CfA surveys by Vogeley et al. (1994). While our model results are in reasonable agreement with these data, the pronounced features in the data suggest that the CfA surveys do not have a sufficiently large volume to avoid significant sample variance effects. Indeed, when we extract 'CfA survey' volumes from random locations in our simulations and measure their genus curves, we find that excursions such as those seen in Fig. 9 are very common.

5 DISCUSSION AND CONCLUSIONS

We have implemented a semi-analytic model of galaxy formation in high-resolution *N*-body simulations of the ΛCDM cosmology in order to study the spatial distribution of galaxies and its evolution. The semi-analytic model requires us to choose values for a number of parameters which describe the physical processes that are modelled, such as gas cooling, star formation, the associated feedback mechanisms and chemical evolution, galaxy merging, the evolution of stellar populations, etc. In keeping with the general philosophy of our work on the subject, we have fixed all the model parameters by requiring a match to a handful of global properties of the local galaxy population, with the largest number of constraints coming from the local *B*-band and *K*-band luminosity functions. No further adjustment to these parameter values was allowed in the clustering study carried out in this paper. Thus, our clustering results are genuine predictions of the model and offer an opportunity to test the validity of the physical assumptions it requires as well as the realism of the ΛCDM model as a whole. In this paper, we have considered three specific statistical measures of clustering: the correlation length (in real and redshift space) of samples of galaxies of different luminosity, the morphology–density relation and the genus curve. At $z = 0$, our model may be tested by forthcoming data from the 2dF and Sloan surveys; at $z = 1$, it may tested by planned surveys such as DEEP (Davis & Faber 1998) and VIRMOS (Le Févre et al. 1999).

The results presented here extend and complement those presented in earlier papers in this series (Benson et al. 2000a,c), as well as in the series of papers by Kauffmann and collaborators (Kauffmann et al. 1999a,b; Diaferio et al. 1999) who analysed one of the *N*-body simulations that we have analysed here, but using their own semi-analytic model. In Benson et al. (2000a), we examined the physical and statistical processes that segregate the

galaxies from the dark matter and we showed that the (real-space) two-point galaxy correlation function in a ΛCDM model that produces an acceptable galaxy luminosity function is in remarkably good agreement with observations. In Benson et al. (2000c), we considered clustering in redshift space and analysed the resulting distortions of the two-point correlation function, as well as its dependence on galaxy luminosity, morphology and colour. Kauffmann and collaborators studied many of these properties too, at the present day (Kauffmann et al. 1999a), and at high redshift (Kauffmann et al. 1999b), as well as the clustering of groups (Diaferio et al. 1999). On the whole, these two independent analyses agree quite well and the differences that do exist can be readily understood in terms of differences in the detailed assumptions for the physics of galaxy formation (see Benson et al. 2000c for a detailed discussion of differences between the two models). The evolution of clustering has also been studied using similar semi-analytic/*N*-body techniques by Governato et al. (1998) and, more recently, by Wechsler et al. (2001).

Images of our simulation clearly illustrate many of the salient features of galaxy growth by hierarchical clustering. They show that galaxies approximately trace the filamentary structure and avoid the lowest density regions of the dark matter distribution, that the redder galaxies tend to predominate in the most massive dark haloes and that the brightest galaxies occur almost exclusively in regions of high dark matter density. A series of time slices shows how the galaxy population changes in abundance and colour with the passage of time and demonstrates the primary effect behind biased galaxy formation: the formation of the first bright galaxies in regions of exceptionally high dark matter density.

From quantitative studies of the galaxy distribution we reach the following three conclusions:

(i) The correlation length of galaxies in real and redshift-space increases rapidly with galaxy luminosity for galaxies brighter than L_*, both at $z = 0$ and at high redshifts, $z \lesssim 3$.

(ii) A strong morphology–density relation, in the same sense as observed, is a natural outcome of hierarchical clustering from CDM initial conditions, and rapidly develops in our simulation. A clear morphology–density relation is predicted to be already in place at least since $z = 1$.

(iii) The topology of the galaxy distribution, as measured by the genus statistic, differs significantly from that of the dark matter. However, the differences are almost entirely caused by sparse sampling effects; the stochastic biasing between galaxies and dark matter is, at most, a minor effect.

We now discuss these points more detail. The variation of the correlation length with luminosity or, equivalently, with mean intergalaxy separation is one of the most striking results of our analysis. The correlation length is virtually insensitive to the mean separation within the sample out to separations of around $10\,h^{-1}$ Mpc (corresponding to luminosities of $M_{\rm B} - 5\log h \approx -20.5$), but for brighter, sparser samples it increases very rapidly. Thus, the redshift-space clustering length of galaxies of luminosity $7L_*$ is predicted to be over twice as large as that of L_* galaxies. The pattern is similar in real and redshift space. Owing to the large volume of our simulations, this is the first time that this rapid increase in correlation length at the brightest luminosities has been unambiguously demonstrated. The main cause of this behaviour is the preponderance of such galaxies at the centres of massive clusters. These galaxies experience enhanced merger rates at early times and are the beneficiaries of late accretion of cool gas which, in our model, is always funnelled onto the central galaxy in each halo. Our predictions for the dependence of clustering strength on galaxy luminosity are in broad agreement with existing data sets, but these are rather small and thus subject to considerable sampling uncertainties. A better test of these predictions should be forthcoming shortly from the 2dF and Sloan surveys.

Our simulations develop a strong morphology–density relation similar to that observed in the local universe: E/S0 galaxies predominate in rich clusters while spirals predominate in the field. We have used a technique patterned after observational procedures to characterise the morphology–density relation in our simulations and find that it quantitatively agrees rather well with observations. The cause of the morphology–density relation is closely related to the reasons behind the luminosity dependence of clustering. In rich clusters, galaxy evolution is 'accelerated' relative to more ordinary regions of space, thus allowing sufficient time for mergers and interactions to build up a large population of bright E/S0 galaxies. Remarkably, a strong morphology–density relation is already well-established by $z = 1$.

Finally, we have investigated the topology of the galaxy distribution, providing the first theoretical prediction for the genus curve of galaxies, rather than merely of dark matter. Of course, if galaxy density were monotonically related to dark matter density, the two curves would be the same. In the simulations there is on average a monotonic relation between the two, but it has such large scatter that it does not preclude differences in the respective genus curves. We do actually find a difference in our simulations: the dark matter genus curve has a 'bubble' shift whereas the galaxy genus curve has a 'meatball' shift and also a higher amplitude. It turns out, however, that the differences are not a result of genuine topological differences but rather of the sparse sampling of the density field provided by galaxies. The confusing effects of sampling have been pointed out by previous authors (e.g. Canavezes et al. 1998; Springel et al. 1998). Our simulations show that, in order to measure an unbiased genus curve for a clustered galaxy distribution (particularly at low overdensity), several hundred galaxies per smoothing volume are required. This is unlikely to be practical even with the new generation of large surveys, and so we will have to live with these biases. The best approach for comparing models and data is therefore to analyse each in identical ways, so as to cancel out any systematic effects. Mock galaxy catalogues such as those presented here will be crucial for this approach.

Some of the trends discussed in this paper, such as the existence of a morphology–density relation, a large correlation length for bright galaxies at high redshift and a rapid increase of clustering strength at very bright magnitudes, are generic outcomes of hierarchical clustering from cold dark matter initial conditions. The detailed predictions presented in this paper depend not only on such fundamental assumptions, but also on the specific galaxy formation processes that we have modelled, such as gas cooling, star formation and feedback, galaxy mergers and so on. The degree to which the predictions depend on the details of the modelling vary from statistic to statistic. For example, in Benson et al. (2000a) we argued that our predicted correlation function for L_* galaxies is robust to changes in model parameters so long as a realistic model luminosity function is preserved, while in Benson et al. (2000c) we showed that other statistics, such as pairwise velocity dispersions, are quite sensitive to the way in which the model apportions galaxies to haloes of different mass. Clustering analyses with the forthcoming generation of large galaxy surveys will test both the generic and specific aspects of the model. Thus,

for example, a failure to detect the luminosity trend seen in Fig. 4 would be a puzzle for the CDM cosmogony in general, even though the exact magnitude at which we expect a strong trend to become apparent does depend on the details of our modelling. Similarly, the clustering of bright galaxies at high redshift is a fundamental test of the CDM theory, but, again, the details are sensitive to the astrophysical modelling. The sheer size of the 2dF and Sloan surveys will allow accurate studies of subsamples defined according to galaxy properties (colour, star formation rate, etc). These will be particularly useful for testing and constraining the details of our modelling of galaxy formation.

To summarize, the combination of high-resolution *N*-body simulations of dark matter and semi-analytic modelling of galaxies provides a powerful technique for turning cosmological and galaxy formation theory into realistic realizations of the galaxy population that can be compared in detail with observations. Tests of this sort will become increasingly common with the new generation of galaxy surveys. In this way, it will be possible to extract not only the cosmological information encoded in the clustering pattern, but also valuable information regarding the physics of galaxy formation.

ACKNOWLEDGMENTS

AJB, SMC and CSF acknowledge receipt of a PPARC Studentship, Advanced Fellowship and Senior Fellowship respectively. CSF also acknowledges a Leverhulme Research Fellowship. CGL acknowledges support at SISSA from COFIN funds from MURST and funds from ASI. This work was supported in part by a PPARC rolling grant, by a computer equipment grant from Durham University and by the European Community's TMR Network for Galaxy Formation and Evolution. We acknowledge the Virgo Consortium and GIF for making available the simulations used in this study.

REFERENCES

Adelberger K. L., Steidel C. C., Giavalisco M., Dickinson M., Pettini M., Kellogg M., 1998, ApJ, 505, 1
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Baugh C. M., 1996, MNRAS, 280, 267
Baugh C. M., Cole S., Frenk C. S., 1996a, MNRAS, 282, L27
Baugh C. M., Cole S., Frenk C. S., 1996b, MNRAS, 283, 1361
Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 1998, ApJ, 498, 504
Baugh C. M., Benson A. J., Cole S., Frenk C. S., Lacey C. G., 1999, MNRAS, 305, L21
Benoist C., Maurogordato S., da Costa L. N., Cappi A., Schaffer R., 1996, ApJ, 472, 452
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000a, MNRAS, 311, 793
Benson A. J., Bower R. G., Frenk C. S., White S. D. M., 2000b, MNRAS, 314, 557
Benson A. J., Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 2000c, MNRAS, 316, 107
Benson A. J., Pearce F. R., Frenk C. S., Baugh C. M., Jenkins A., 2001, MNRAS, 320, 261
Bernardeau F., Schaeffer R., 1992, A&A, 255, 1
Blanton M., Cen R., Ostriker J. P., Strauss M. A., 1999, ApJ, 522, 590
Blanton M. et al. (The SDSS Collaboration), 2001, AJ, 212, 2358
Bruzual A. G., Charlot S., 1993, ApJ, 405, 538
Canavezes A. et al., 1998, MNRAS, 297, 777
Canavezes A., Sharpe J., 2001, MNRAS, 326, 209
Cole S., Aragón-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781
Cole S., Hatton S., Weinberg D., Frenk C. S., 1998, MNRAS, 300, 945
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Coles P., 1993, MNRAS, 262, 1065
Coles P., Davies A. G., Pearson R. C., 1996, MNRAS, 281, 1375
Davis M., Geller M. J., 1976, ApJ, 208, 13
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Davis M., Faber S. M., 1998, Wide Field Surveys in Cosmology. Editions Frontiers, Gif-sur-Yvette, p. 333
Diaferio A., Kauffmann G., Colberg J. M., White S. D. M., 1999, MNRAS, 307, 537
Dressler A., 1980, ApJ, 236, 351
Evrard A. E., Summers F. J., Davis M., 1994, ApJ, 422, 11
Ferrara A., Bianchi S., Cimatti A., Giovanardi C., 1999, ApJS, 123, 423
Frenk C. S., White S. D. M., Efstathiou G., Davis M., 1985, Nat, 317, 595
Frenk C. S., White S. D. M., Efstathiou G., Davis M., 1988, ApJ, 327, 507
Frenk C. S., Evrard A. E., White S. D. M., Summers F. J., 1996, ApJ, 472, 460
Gott J. R., Melott A. L., Dickinson M., 1986, ApJ, 306, 341
Gott J. R., Weinberg D. H., Melott A. L., 1987, ApJ, 319, 1
Gott J. R. et al., 1989, ApJ, 340, 625
Governato F., Baugh C. M., Frenk C. S., Cole S., Lacey C. G., Quinn T., Stadel J., 1998, Nat, 392, 359
Granato G. L., Silva L., Lacey C. G., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, ApJ, 542, 710
Jarret T. H., Chester T., Cutri R., Schneider S., Skrutskie M., Huchra J. P., 2000, AJ, 119, 2498
Jenkins A. et al. (the VIRGO Consortium), 1998, ApJ, 499, 20
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Yoshida N., 2001, MNRAS, 321, 372
Jing Y. P., 1998, ApJ, 503, L9
Kaiser N., 1984, ApJ, 284, 9
Katz N., Hernquist L., Weinberg D. H., 1992, ApJ, 399, 109
Kauffmann G., 1995, MNRAS, 274, 161
Kauffmann G., 1996, MNRAS, 281, 487
Kauffmann G., Guiderdoni B., White S. D. M., 1994, MNRAS, 267, 981
Kauffmann G., Nusser A., Steinmetz M., 1997, MNRAS, 286, 795
Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999a, MNRAS, 303, 188
Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999b, MNRAS, 307, 529
Lacey C. G., Cole S., 1994, MNRAS, 271, 676
Le Févre O. et al., 1999, in Giuricin G., Mezzetti M., Salucci P., eds, Proc. ASP Conf. Ser. Vol. 176, Observational Cosmology: The Development of Galaxy Systems. Astron. Soc. Pac., San Francisco, p. 250
Lilly S. J., Le Févre O., Hammer F., Crampton D., 1996, ApJ, 460, L1
Loveday J., Maddox S. J., Efstathiou G., Peterson B. A., 1995, ApJ, 442, 457
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Madau P., Fergusen H. C., Dickinson M., Giavalisco M., Steidel C. C., Fruchter A., 1996, MNRAS, 283, 1388
Moore B. et al., 1992, MNRAS, 256, 477
Okamoto T., Nagashima M., 2001, ApJ, 547, 109
Park C., Gott J. R., da Costa L. N., 1992, ApJ, 392, L51
Peacock J. A. et al. (The 2dFGRS Team), 2001, Nat, 410, 169
Pearce F. R. et al. (The Virgo Consortium), 1999, ApJ, 521, L99
Peebles P. J. E., 2001, pre-print (astro-ph/0101127)
Postman M., Geller M. J., 1984, ApJ, 281, 95
Rhoads J. E., Gott J. R., Postman M., 1994, ApJ, 421, 1
Seljak U., 2000, MNRAS, 318, 203
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Somerville R. S., Lemson G., Sigad Y., Dekel A., Kauffmann G., White S. D. M., 2001, MNRAS, 320, 289
Springel V. et al., 1998, MNRAS, 298, 1169

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, submitted (astro-ph/0012055)
Steidel C. C., Adelberger K. L., Giavalisco M., Dickinson M., Pettini M., 1999, ApJ, 519, 1
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., Hernquist L., Katz N., 1997, ApJ, 477, 8

This paper has been typeset from a TEX/LATEX file prepared by the author.