Mon. Not. R. Astron. Soc. **371,** 537–547 (2006) doi:10.1111/j.1365-2966.2006.10669.x

Modelling galaxy clustering in a high-resolution simulation of structure formation

Lan Wang,[1,2]★ Cheng Li,[2,3,4] Guinevere Kauffmann[2] and Gabriella De Lucia[2]

[1]*Department of Astronomy, Peking University, Beijing 100871, China*
[2]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany*
[3]*Center for Astrophysics, University of Science and Technology of China, Hefei, Anhui 230026, China*
[4]*The Partner Group of MPI für Astrophysik, Shanghai Astronomical Observatory, Nandan Road 80, Shanghai 200030, China*

Accepted 2006 June 1. Received 2006 May 26; in original form 2006 March 20

ABSTRACT

We use the Millennium Simulation, a 10 billion particle simulation of the growth of cosmic structure, to construct a new model of galaxy clustering. We adopt a methodology that falls midway between the traditional semi-analytic approach and the halo occupation distribution (HOD) approach. In our model, we adopt the positions and velocities of the galaxies that are predicted by following the orbits and merging histories of the substructures in the simulation. Rather than using star formation and feedback 'recipes' to specify the physical properties of the galaxies, we adopt parametrized functions to relate these properties to the quantity $M_{\rm infall}$, defined as the mass of the halo at the epoch when the galaxy was last the central dominant object in its own halo. We test whether these parametrized relations allow us to recover the basic statistical properties of galaxies in the semi-analytic catalogues, including the luminosity function, the stellar mass function and the shape and amplitude of the two-point correlation function evaluated in different stellar mass and luminosity ranges. We then use our model to interpret recent measurements of these quantities from Sloan Digital Sky Survey (SDSS) data. We derive relations between the luminosities and the stellar masses of galaxies in the local Universe and their host halo masses. Our results are in excellent agreement with recent determinations of these relations by Mandelbaum et al. using galaxy–galaxy weak lensing measurements from the SDSS.

Key words: galaxies: distances and redshifts – galaxies: fundamental parameters – galaxies: haloes – cosmology: theory – dark matter – large-scale structure of Universe.

1 INTRODUCTION

According to the current standard paradigm, galaxies form and reside inside extended dark matter haloes. Three different approaches have been used to model the link between the properties of galaxies and the dark matter haloes in which they are found. One approach is to carry out N-body + hydrodynamical simulations that include both gas and dark matter (Katz, Weinberg & Hernquist 1996; Pearce et al. 2001; White, Hernquist & Springel 2001; Yoshikawa et al. 2001). Another approach is to combine N-body simulations with simple prescriptions, taken directly from semi-analytic models of galaxy formation (Kauffmann et al. 1999), to track gas cooling and star formation in galaxies. The third method is the so-called halo occupation distribution (HOD) approach, which aims to provide a purely statistical description of how dark matter haloes are populated by galaxies.

★E-mail: wanglan@mpa-garching.mpg.de

Typical HOD models are constructed by specifying the number of galaxies N that populate a dark matter halo of mass M as well as the distribution of galaxies within these haloes (Kauffmann, Nusser & Steinmetz 1997; Benson et al. 2000; Peacock & Smith 2000; Seljak 2000; Berlind & Weinberg 2002; Berlind et al. 2003). More recent models have concentrated on the so-called conditional luminosity function $\Phi(L|M)\,{\rm d}L$, which gives the number of galaxies of luminosity L that reside in a halo of mass M (Yang, Mo & van den Bosch 2003). Most HOD models also distinguish between 'central' galaxies, located at the centres of dark matter haloes and 'satellite' galaxies, which are usually assumed to have the same density profile as the dark matter within the halo. Physically, this is supposed to reflect the fact that gas cools and accumulates at the halo centres until the halo merges with a larger structure. With this approach, the models can be used to explore the parameters that are required to match simultaneously the galaxy luminosity function as well as the luminosity, colour and morphology dependences of the correlation function (van den Bosch, Yang & Mo 2003; Yang et al. 2005; Zehavi et al. 2005). Other papers have used HOD models to explore

the detailed shape of two-point correlation function (Zehavi et al. 2004) as well as higher order correlation functions (Wang et al. 2004).

N-body simulations can now be carried out with high enough resolution to track the histories of individual substructures (subhaloes) within the surrounding halo (Springel et al. 2001). It is thus becomes possible to specify the positions and velocities of galaxies within a halo in a dynamically consistent way, rather than assuming a profile or form for the velocity distribution. Galaxy clustering statistics that are computed using the full information available from these high-resolution simulations should in principle be more accurate and robust.

Caution must be exercised, however, when only using subhaloes as tracers of galaxies in high-resolution simulations, as has been recently done by Vale & Ostriker (2004, 2005) and Conroy, Wechsler & Kravtsov (2006). In standard models of galaxy formation, when a galaxy is accreted by a larger system such as a cluster, its surrounding gas is shock-heated to high temperatures. Star formation then terminates as the internal gas supply of the galaxy is used up. The stellar masses of satellite galaxies only change by a small amount after they are accreted, while their luminosities dim due to aging of their stars. In contrast, the dark matter haloes surrounding the satellites gradually lose mass as their outer regions are tidally stripped (De Lucia et al. 2004). Near the centres of the haloes, most of the substructures have been completely destroyed. Gao et al. (2004) have shown that the radial distribution of subhaloes is much less centrally concentrated than the radial distribution of galaxies predicted by simulations that follow the full orbital and merging histories of these systems.[1]

In this paper, we make use of the Millennium Simulation, a 10 billion particle simulation of the growth of cosmic structure, to construct a new model of galaxy clustering. We adopt a methodology that falls in between the semi-analytic approach, which tracks galaxy formation *ab initio* within the simulation, and the HOD approach, which only provides a statistical description of how galaxies are related to the underlying dark matter density distribution. In our approach, we adopt the positions and velocities of the galaxies as predicted by following the orbits and merging histories of the substructures in the simulation. Rather than using star formation and feedback 'recipes' to calculate how the physical properties of the galaxies such as their luminosities or stellar masses evolve with time, we adopt parametrized functions to relate these properties to the quantity $M_{\rm infall}$, defined as the mass of the halo at the epoch when the galaxy was last the central dominant object. For central galaxies at the present day, $M_{\rm infall}$ is simply the present day halo mass, but for satellite galaxies, it is the mass of the halo when the galaxy was first accreted by a larger structure.

This approach has the advantage of the semi-analytic models in that it provides very accurate positions and velocities for all the galaxies in the simulation. It maintains the simplicity of the HOD approach, because it bypasses the need to incorporate detailed treatment of star formation and feedback processes. Our aim in developing these models is to use them as a means of constraining the relation between galaxy physical properties and halo mass directly from observational data, not as a means of understanding the physics of galaxy formation.

The paper is organized as follow: we first introduce the Millennium Run and the methodology used for identifying haloes, subhaloes and galaxies in this simulation. We then study the relation between luminosity/stellar mass and $M_{\rm infall}$ in mock galaxy catalogues constructed using these simulations. In Section 4, we introduce a parametrization for these relations and show that we are able to recover basic statistical quantities such as the galaxy luminosity/mass function and the shape and amplitude of the two-point correlation function in different luminosity/mass bins. We also investigate the effect of changing the parameters of the relation on the luminosity function and correlation function. In Section 5, we apply the method to real data on the clustering of galaxies as a function of luminosity and stellar mass (Li et al. 2006) derived from the Sloan Digital Sky Survey (SDSS). Finally, we discuss our results and present our conclusions.

2 THE SIMULATION

The Millennium Simulation (Springel et al. 2005) used in this study is the largest simulation of cosmic structure growth carried out so far. The cosmological parameters values in the simulation are consistent with recent determinations from a combined analysis of the 2dF Galaxy Redshift Survey (2dFGRS) (Colless et al. 2001) and first year *WMAP* (*Wilkinson Microwave Anisotropy Probe*) data (Spergel et al. 2003). A flat Λ cold dark matter (CDM) cosmology is assumed with $\Omega_{\rm m} = 0.25$, $\Omega_{\rm b} = 0.045$, $h = 0.73$, $\Omega_\Lambda = 0.75$, $n = 1$ and $\sigma_8 = 0.9$. The simulation follows $N = 2160^3$ particles of mass $8.6 \times 10^8\ h^{-1}\ {\rm M}_\odot$ from redshift $z = 127$ to the present day, within a comoving box of $500\,h^{-1}$ Mpc on a side.

Full particle data are stored at 64 output times. For each output, haloes are identified using a friends-of-friends (FOF) group finder. Substructures (or subhaloes) within a FOF halo are located using the SUBFIND algorithm of Springel et al. (2001). After finding all haloes and subhaloes at all output snapshots, merging trees are built describing in detail how these systems merge and grow as the universe evolves. Since structures merge hierarchically in CDM universes, for any given halo, there can be several progenitors, but in general each halo or subhalo only has one descendant. Merger trees are thus constructed by defining a unique descendant for each halo and subhalo. Through those merging trees, we are able to follow the history of haloes/subhaloes, as well as the galaxies inside them.

Once a halo appears in the simulation, it is assumed that a galaxy begins to form within it. As the simulation evolves, the halo may merge with a larger structure and become a subhalo, while the galaxy becomes a satellite galaxy. The galaxy's position and velocity are specified by the position and velocity of the most bound particle of its host halo/subhalo. Even if the subhalo hosting the galaxy is tidally disrupted, the position and velocity of the galaxy is still traced through this most bound particle. We will refer to these galaxies without subhaloes as 'orphaned' systems. Galaxies thus only disappear from the simulation if they merge with another galaxy. The time taken for an orphaned galaxy to merge with the central object is given by the time taken for dynamical friction to erode its orbit, causing it to spiral into the centre and merge. The satellite orbits are thus tracked directly until the subhalo is disrupted; thereafter, the time taken for the galaxy to reach the centre is calculated using the standard Chandrasekhar formula.

In this paper, we will parametrize quantities such as galaxy luminosity and stellar mass as a function of the quantity $M_{\rm infall}$, which is defined as the virial mass of the halo hosting the galaxy at the

[1] Note that the simulations analysed by Conroy et al. (2006) are significantly higher resolution than the ones analysed in this paper, but are much smaller in volume. As discussed in their paper, the problem of disrupted subhaloes is not likely to be a significant problem for galaxies in the range of luminosities considered in their analysis.

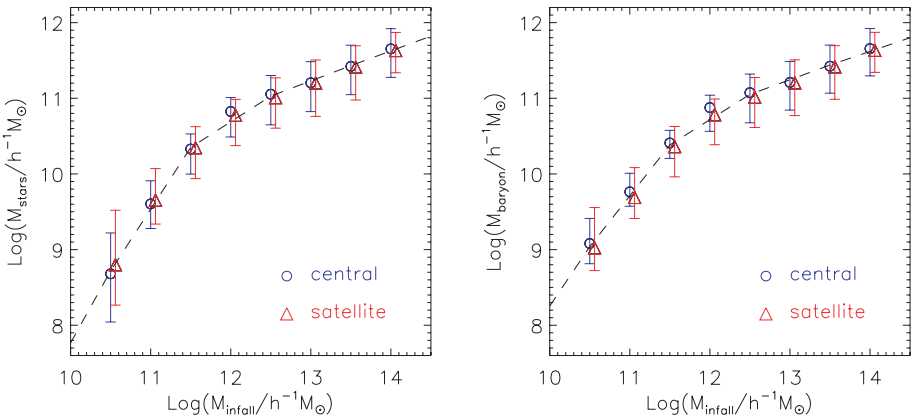

Figure 1. Relations between stellar mass, baryonic mass and $M_{\rm infall}$ calculated from the semi-analytic galaxy catalogues. Open circles represent central galaxies and triangles are for satellite galaxies. Error bars indicate the 95 percentiles of the mass distribution at the given value of $M_{\rm infall}$. Dashed lines show the double power-law parametrized fit to the median value of relations for the galaxy sample as whole.

epoch when it was last the central galaxy of its own halo. The Millennium Simulation catalogues include haloes down to a resolution limit of 20 particles, which yields a minimum halo mass of $2 \times 10^{10}\,h^{-1}\,{\rm M}_{\odot}$. In our study, we only consider galaxies with $M_{\rm infall}$ greater than $10^{10.5}\,h^{-1}\,{\rm M}_{\odot}$. (Note that $M_{\rm infall}$ is simply the virial mass of the host halo for central galaxies at the present day.) This results in a total sample of 11 761 178 galaxies within the simulation volume.

3 THE RELATIONS BETWEEN $M_{\rm infall}$, STELLAR MASS AND LUMINOSITY IN THE SEMI-ANALYTIC GALAXY CATALOGUES

In the following two sections we use the semi-analytic galaxy catalogues constructed from the Millennium Simulation by Croton et al. (2006) (http://www.mpa-garching.mpg.de/galform/agnpaper/) to study how galaxy properties such as stellar mass, baryonic mass (i.e. stellar mass + cold gas mass) and luminosity depend on $M_{\rm infall}$, the mass of the halo in which the galaxy was last a central object. We construct parametrized relations between these quantities and $M_{\rm infall}$ that match the relations found in the mock catalogue. We then show that our parametrization allows us to recover both the luminosity/mass functions of the simulated galaxies and the shape, amplitude and mass/luminosity dependence of the two-point correlation functions. Croton et al. (2006) have shown that their catalogues provide a good match to the observed galaxy luminosity function and the clustering properties of galaxies, so we believe that it is a reasonable to use these catalogues as a way of motivating and testing our simple parametrizations.

In Fig. 1, we plot the relations between $M_{\rm infall}$ and galaxy stellar mass ($M_{\rm stars}$) and baryonic mass ($M_{\rm baryon}$). We show results for present-day central galaxies in blue and satellite galaxies in red. Error bars indicate the 95th percentiles of the distributions. As we will show the relations between $M_{\rm infall}$ and $M_{\rm stars}/M_{\rm baryon}$ are well described by a double power law. The crossover point between the two power laws is at a halo mass of $\sim 3 \times 10^{11}\,h^{-1}\,{\rm M}_{\odot}$, which corresponds to a galaxy with stellar mass of around $10^{10}\,h^{-1}\,{\rm M}_{\odot}$. In less massive haloes, supernova feedback acts to prevent gas from cooling and forming stars as efficiently as in high-mass haloes. In massive haloes, the cooling times become longer and a smaller fraction of the baryons are predicted to cool and form stars. In addition, in the models of Croton et al. (2006), heating from active galactic nucleus (AGN) also acts to suppress cooling on to high-mass galaxies.

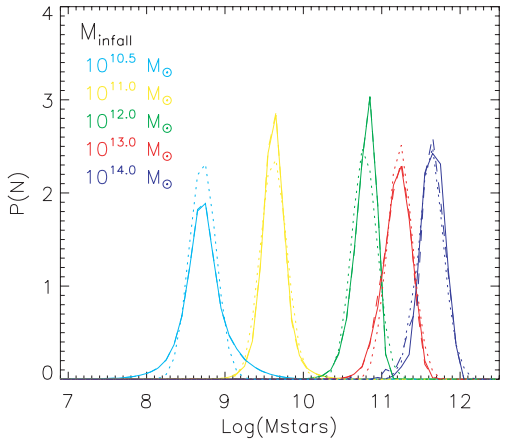

Figure 2. The distribution in stellar mass for different $M_{\rm infall}$ bins increasing from $10^{10.5}\,h^{-1}\,{\rm M}_{\odot}$ (left) to $10^{14.0}\,h^{-1}\,{\rm M}_{\odot}$ (right). Solid and dashed lines are for central and satellite galaxies (note that they lie on top of each other for three lower mass bins). Dotted lines indicate the Gaussian fits to the distributions that are used in our parametrized model.

Fig. 2 shows that the distribution of $M_{\rm stars}$ at a given value of $M_{\rm infall}$ is well described by a lognormal function. The width of the lognormal depends weakly on halo mass with a maximum dispersion σ of 0.2 dex at $M_{\rm infall} = 10^{10.5}\,h^{-1}\,{\rm M}_{\odot}$ and a minimum σ of 0.1 dex at $M_{\rm infall} = 10^{11.5}\,h^{-1}\,{\rm M}_{\odot}$. The relations depend very little on whether the galaxy is a central or satellite system. The dispersion around the relations is also similar for the two types (in Fig. 2, the solid and dashed lines for central and satellite galaxies lie almost on top of each other). The similarity in the $M_{\rm infall}$–$M_{\rm stars}$ relations between satellite and centrals may be regarded as something of a coincidence. Although there is little change in the stellar/baryonic component of the galaxy after it falls into a larger structure, haloes of the same mass at different times have different circular velocities and hence different cooling and star formation efficiencies. As we will show later, we obtain better fits to the observational data if we allow the relations between central and satellites galaxies to differ.

Fig. 3 shows the relation between luminosity and $M_{\rm infall}$. It also can be fit by a double power law, but the difference between

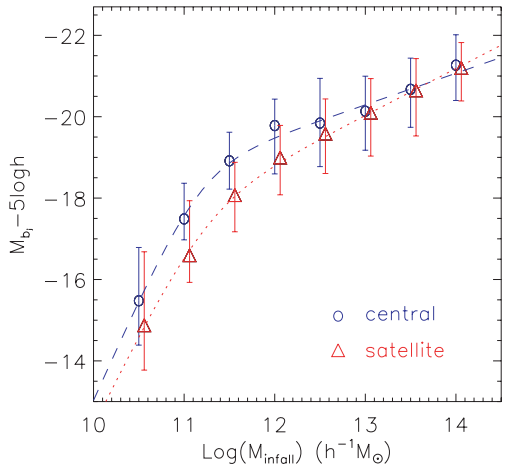

Figure 3. The same as in Fig. 1, but luminosity (represented by magnitude of b_j band) is plotted as a function of $M_{\rm infall}$. Dashed(dotted) lines show the double power-law fits to the relation for central(satellite) galaxies.

central and satellite galaxies is much more obvious. At a given value of $M_{\rm infall}$ central galaxies are more luminous than satellites because they are forming stars at higher rates and their stellar mass-to-light ratios are lower. The difference between central and satellite galaxies becomes very small at large values of $M_{\rm infall}$. This can be understood as a simple consequence of hierarchical structure formation: massive haloes were formed more recently than less massive haloes and subhaloes with large masses are likely to have been accreted relatively recently. Massive satellite galaxies have therefore not been satellites for long and thus have mass-to-light ratios that are more similar to their central counterparts. In addition, the Croton et al. models include a 'radio AGN mode' of feedback, which acts to suppress cooling on to the most massive galaxies. This also acts to reduce the difference between central and satellite galaxy colours and mass-to-light ratios.

4 PARAMETRIZATIONS AND TESTS

4.1 Functional form

We use a two-power-law model of the following form to fit the median value of the relations between $M_{\rm stars}$, $M_{\rm baryon}$, L and $M_{\rm infall}$:

$$x = \frac{2}{(M_{\rm infall}/M_0)^{-\alpha} + (M_{\rm infall}/M_0)^{-\beta}} k,$$

where x denotes $M_{\rm stars}$, $M_{\rm baryon}$ or L, and the relation between luminosity L and b_j band magnitude is given by

$$M_{b_j} - 5\log h = -2.5\log L.$$

We fit these relations for central and satellite galaxies separately, as well as for the galaxy population as a whole. We will later test whether separate fits to the central galaxies and satellites make significant difference to our results. We also assume that the dispersion around the median value has a lognormal form.

Table 1 lists the parameters of the best-fitting models for the relations between $M_{\rm infall}$ and $M_{\rm stars}$, $M_{\rm baryon}$ and L. The models have five parameters.

(i) M_0 is the critical mass/luminosity at which the slope of the relation changes. When we fit satellite and central galaxies separately, we find almost exactly the same values for this parameter (even for luminosity, the difference is less than 20 per cent). We therefore fix M_0 at the best-fitting value for the galaxy sample as a whole.

(ii) α and β describe the slope of the relations at high and low values of $M_{\rm infall}$.

(iii) k is a normalization constant.

(iv) We have calculated the interval in log $M_{\rm stars}$, log $M_{\rm baryon}$ and log L that encloses the central 68 per cent of the probability distribution for eight different values of log $M_{\rm infall}$ from $10^{10.5}$ to 10^{14} h^{-1} $\rm M_\odot$, with step 0.5 dex. We then calculate the average of these eight values and the value of σ quoted in Table 1 is 0.5 times this number.

Table 1. Best-fitting parameter values for the relations between $M_{\rm infall}$ and $M_{\rm stars}$, $M_{\rm baryon}$ and L as derived from the semi-analytic galaxy catalogues of Croton et al. (2006).

		$M_0(h^{-1}\ \rm M_\odot)$	α	β	log (k)	σ	$\tilde{\chi}^2$
$M_{\rm stars}$	Total	3.16×10^{11}	0.39	1.92	10.35	0.156	0.0146
	Central	3.16×10^{11}	0.39	1.96	10.35	0.148	0.0240
	Satellite	3.16×10^{11}	0.39	1.83	10.34	0.167	0.0057
$M_{\rm baryon}$	Total	3.61×10^{11}	0.36	1.59	10.44	0.147	0.0415
	Central	3.61×10^{11}	0.35	1.59	10.46	0.133	0.0542
	Satellite	3.61×10^{11}	0.37	1.59	10.40	0.162	0.0273
$L(M_{b_j})$	Total	1.49×10^{11}	0.36	1.90	7.14	0.215	0.0360
	Central	1.49×10^{11}	0.31	1.99	7.25	0.169	0.1250
	Satellite	1.49×10^{11}	0.46	1.81	6.90	0.189	0.0359

The resulting model fits are plotted as dashed and dotted lines in Figs 1 and 3. The quality of the fit is given in the last column of Table 1 and is calculated as

$$\tilde{\chi}^2 = \sum \left(\frac{x_{\rm fit} - x_{\rm SAM}}{x_{\rm SAM}} \right)^2,$$

where x represents $M_{\rm stars}$, $M_{\rm baryon}$, L for each relation and the sum is over the eight mass bins with $10^{10.5}\,h^{-1}\,\rm M_\odot \leqslant M_{\rm infall} \leqslant 10^{14.0}\,h^{-1}\,\rm M_\odot$.

4.2 Tests

The next step is to see whether these parametrized relations allow us to recover the basic statistical properties of the simulated galaxy catalogue, such as the mass/luminosity function and the mass/luminosity dependence of the two-point correlation function. When fitting to the quantities $M_{\rm star}$ and $M_{\rm baryon}$, we do not distinguish between central and satellite galaxies because the relations are almost the same for both. When fitting to galaxy luminosity, we do allow α, β, k and σ to vary between central and satellite galaxies, but M_0 remains fixed for both. Note that the positions and the velocities of the galaxies are exactly the same as specified in the semi-analytic galaxy catalogues; the parametrized relations between galaxy mass/luminosity and $M_{\rm infall}$ simply provide us with an alternative way of specifying the *properties* of the galaxies.

Figs 4 and 5 show the results of our test. Symbols show results calculated directly from the semi-analytic galaxy catalogues and lines are from our parametrized model. The stellar mass function is well reproduced, and we can also recover the correlation for different stellar mass bins: 10^9, $10^{10.5}$ and $10^{11.5}\,h^{-1}\,\rm M_\odot$ (Fig. 4). For luminosity, the parametrized model is not quite as successful. Although the luminosity function is well reproduced, there are some discrepancies in the dependence of the clustering amplitude on

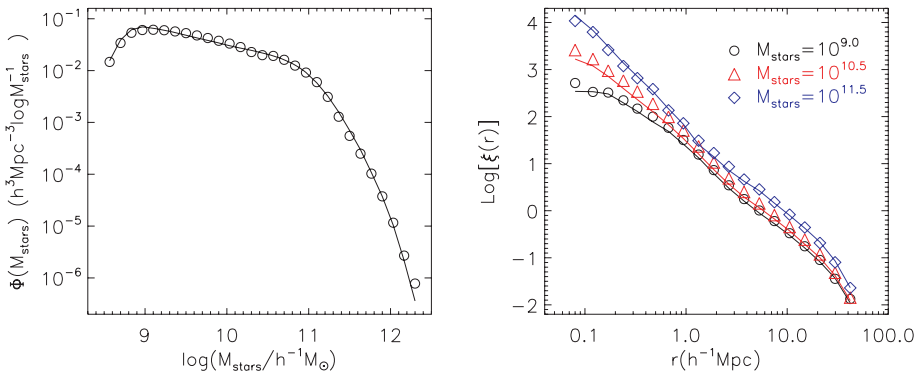

Figure 4. Results from the comparison of the parametrized model and the semi-analytic galaxy catalogue. The left-hand panel shows the stellar mass function. The right-hand panel shows correlation functions for three different stellar mass bins: 10^9, $10^{10.5}$ and $10^{11.5}$ h^{-1} $M_\odot$. Symbols are for the semi-analytic galaxy catalogue. Solid lines are for our parametrized models.

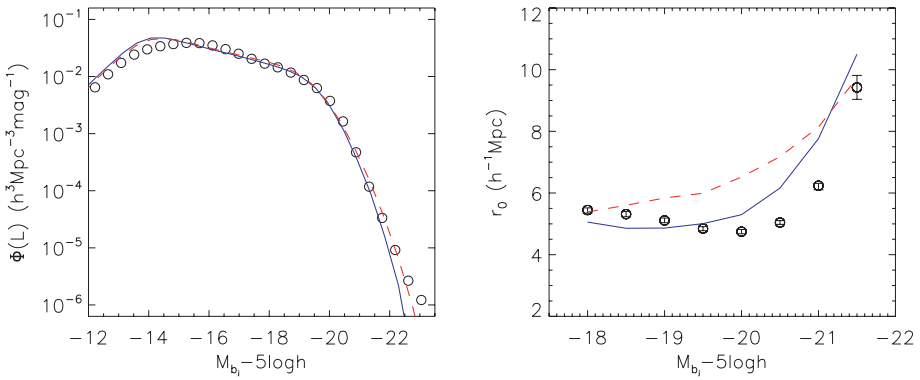

Figure 5. Same as Fig. 4, except for the luminosity function (left) and the correlation function as a function of absolute magnitude (right). Symbols are the semi-analytic results, error bars in the right-hand panel are the bootstrap error of correlation length for the semi-analytic model. Solid lines are for parametrized models where relations for central and satellite galaxies are fit separately. The dashed lines are for models where the fit is for the galaxy population as a whole.

luminosity (solid-blue curve in Fig. 5). Part of the reason for this discrepancy is that our parametrization of the L–$M_{\rm infall}$ relation has somewhat larger χ^2 than the $M_{\rm stars}$–$M_{\rm infall}$ relation (see Table 1). In addition, in order to reproduce the clustering trends as a function of luminosity, it is critical to fit the relation for central and for satellites galaxies separately. If we apply a single relation for both kinds of galaxies, we obtain the red-dashed curve in Fig. 5, which is even more discrepant. Our results suggest that in order to reproduce the clustering dependence on luminosity in a more exact way, one would need to introduce an additional dependence of the L–$M_{\rm infall}$ relation on the parameter $t_{\rm infall}$, the *time* when the galaxy was last the central object of its own halo. This does not appear to be necessary in order to reproduce the stellar mass dependence of galaxy clustering. The reason for this difference is because the optical light from galaxies, unlike their stellar mass, is heavily influenced by the contribution from the youngest stars, which have lifetimes which are short compared to the age of the Universe. Once a galaxy becomes a satellite, it will fade in luminosity even though its stellar mass remains approximately constant. For the sake of simplicity, we will not introduce $t_{\rm infall}$ as an additional parameter in this paper, but we will come back to this in future work in which we consider the colour dependence of galaxy clustering.

4.3 The effect of 'orphan' galaxies

The majority of HOD models in the literature only consider dark matter haloes and subhaloes that can be identified at the present time. Satellite galaxies without surrounding subhaloes are thus omitted from the analysis. We now explore the effect of these 'orphan' satellite galaxies on our results.

The left-hand panel of Fig. 6 compares the L–$M_{\rm infall}$ relation for orphan satellites with the results obtained for central and satellite galaxies that have retained their subhaloes. The right-hand panel of Fig. 6 shows the relative contribution of central galaxies, satellite galaxies with subhaloes and orphan satellites without subhaloes to the luminosity function of the galaxies in the semi-analytic catalogue. As can be seen, orphan satellites have lower luminosities at a given value of $M_{\rm infall}$ than either central galaxies or satellite galaxies with subhaloes – i.e. orphan galaxies are the oldest galaxies with the highest mass-to-light ratios in the simulation. In addition, we see that the contribution of these orphan satellites is highest at the faintest luminosities. Fig. 7 explores the contribution of the orphan satellites to the correlation function in three different bins of absolute magnitude. The solid curves show the result for all the galaxies while the dashed red curves show the result when the orphan satellites are omitted. As can be seen, the orphaned satellites contribute heavily to the correlation function of faint galaxies on scales of less than 1 h^{-1} Mpc. Omission of these systems causes the amplitude of the correlation function to be underestimated by more than an order of magnitude at separations of 0.1 h^{-1} Mpc for galaxies with $-18 < M_{b_j} < -17$.

We note that there are uncertainties in our treatment of orphan galaxies in the simulation. Some of these galaxies may indeed be destroyed or significantly reduced in mass by tidal stripping effects.

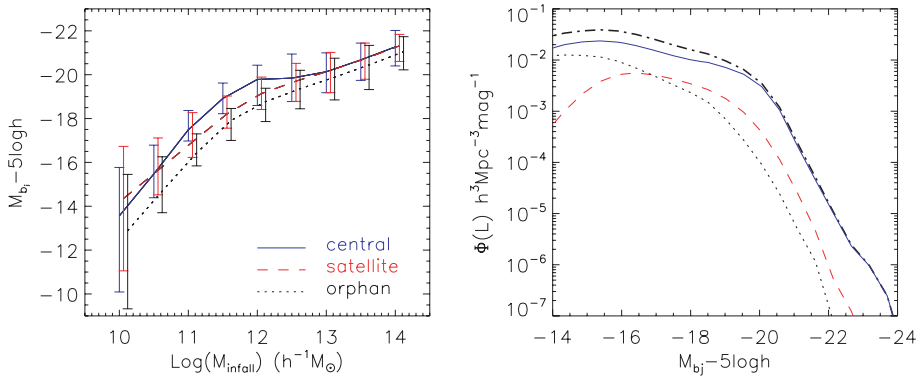

Figure 6. L–$M_{\rm infall}$ relations (left) and luminosity functions (right) for different types of galaxies from the semi-analytic galaxy catalogue: central galaxies (solid lines), satellite galaxies with subhaloes (dashed lines), satellite galaxies without subhaloes (dotted lines). The dashed–dotted line in the right-hand panel shows the total luminosity function for all galaxies.

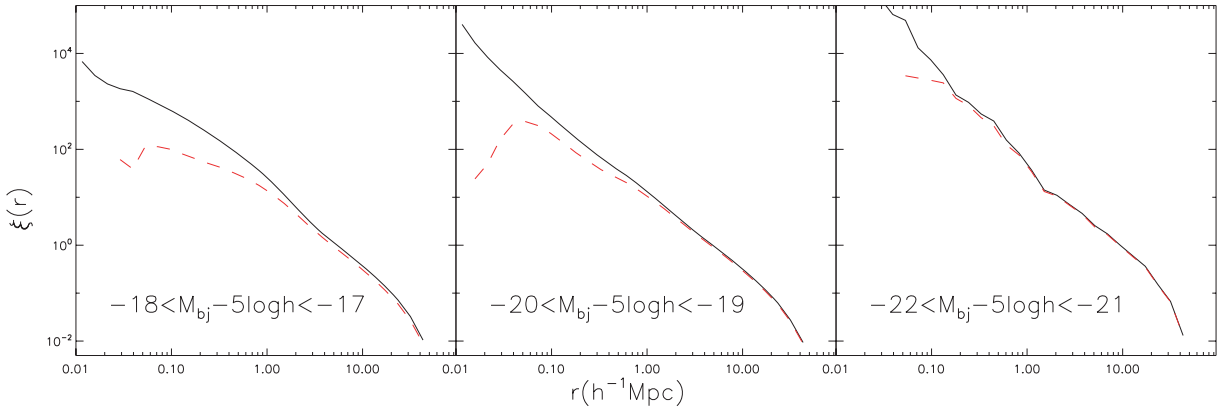

Figure 7. Correlation functions for three luminosity bins including (solid lines) and not including (dashed lines) orphan satellite galaxies.

Indeed, the existence of a significant intracluster light component does suggest tidal effects or mergers do unbind some of the stars in satellite galaxies (Arnaboldi 2004; Feldmeier et al. 2004; Zibetti et al. 2005). In face of these uncertainties, we have chosen to assume that the visible galaxies survive even after their subhalo falls below the resolution limit of the simulation. It is possible that we overestimate the number of these objects because we do not include tidal stripping on the stellar component. However, we believe that 'orphan' galaxies (at least part of them) are needed in order to explain observational results. From Fig. 7, we see that when 'orphan' galaxies are excluded, the correlation signal decreases at small scales, at odds with observational results (see later in Figs 9 and 10). In this work, we consider all the 'orphan' systems as part of satellite subsamples.

4.4 Changes in the input parameters

One advantage of our parametrized approach is that we can understand the effect of changing each different parameter and thus gain intuition about what changes are necessary to bring the models into the closest possible agreement with the observational data. This is different in spirit to exploring parameter space in the semi-analytic models, because the parameters in these models are tied to the physical recipes for star formation and feedback rather than the relation between halo mass and galaxy properties, which is the focus of our approach.

In the upper panel of Fig. 8, we show how changing each of the parameters affects the stellar mass function. Note that the normalization constant k is always adjusted in order to keep the amplitude of the mass function at $M_{\rm stars} = 10^{11}\,\mathrm{M}_\odot$ fixed. Changing M_0 affects the mass scale of the transition between the two power laws as well as the amplitude of the mass function both at low and at high masses. Changes to α affect the shape of the mass function at the high mass end, while changes to β affect the low mass end of the mass function. A change in scatter σ has similar effect to a change in α, and influences the amplitude of the mass function at the high mass end. This is because the mass function is relatively flat at low masses and declines steeply at high masses, so an increasing amount of scatter in the $M_{\rm stars}$–$M_{\rm infall}$ relation will have a strong effect on the number of high-mass galaxies.

The lower panels of Fig. 8 show the effect of the same parameter changes on the amplitude of the correlation function evaluated on scales of $r = 0.33$ and $5.30\,h^{-1}$ Mpc. We see that a parameter change that causes an increase in the number of galaxies of given mass will cause a corresponding *decrease* in the clustering amplitude of these systems. This is easy to understand. In order to have more galaxies of a given mass in the simulation, they must be shifted into lower mass haloes and these low-mass haloes are more weakly clustered.

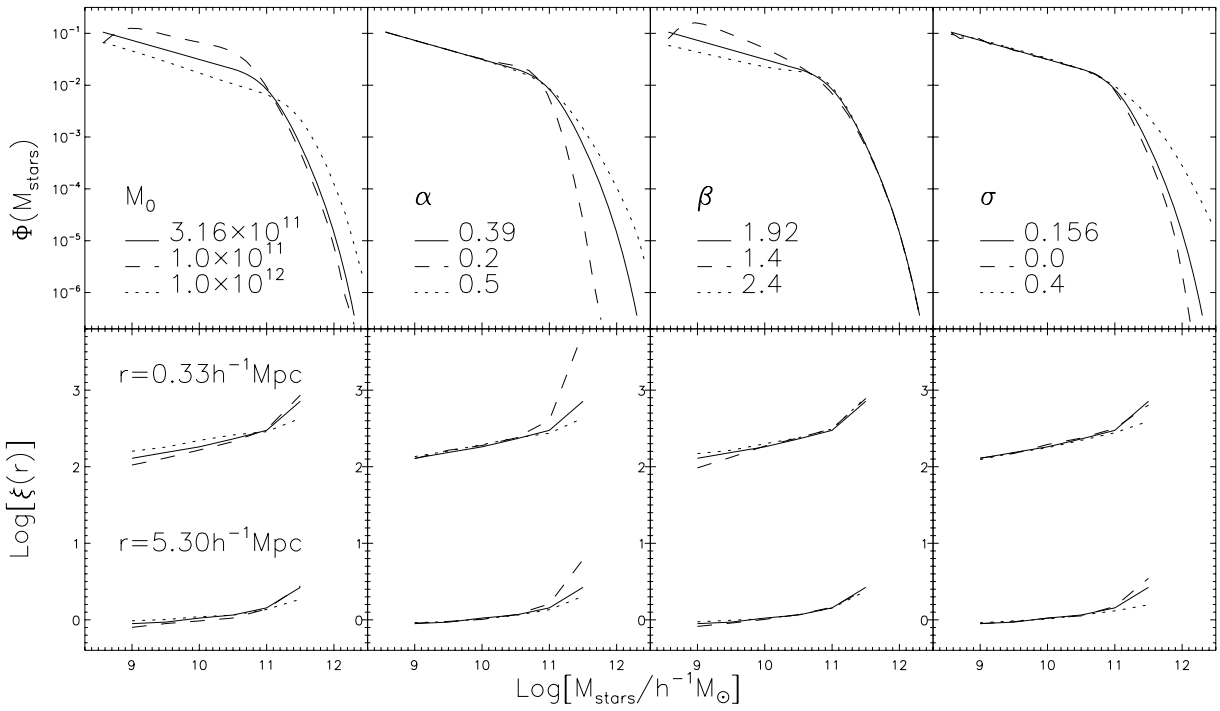

Figure 8. The effect of changing parameters on the stellar mass function (upper panels) and correlation at scales of $r = 0.33$ and $5.30\,h^{-1}$ Mpc (lower panels). The solid lines represent the best-fitting model for the $M_{\rm stars}$–$M_{\rm infall}$ relation.

5 APPLICATION TO SDSS

In this section, we apply our models to observational data from the SDSS. Recent large-scale redshift surveys such as 2dfGRS (Colless et al. 2001) and SDSS (York et al. 2000) provide galaxy samples that are large enough to measure the luminosity dependence of galaxy clustering accurately (Norberg et al. 2002a,b; Zehavi et al. 2005). In this paper, we make use of the recent measurements of the projected correlation function $w(r_{\rm p})$ by Li et al. (2006). These authors calculated $w(r_{\rm p})$ not only as a function of galaxy luminosity, but also stellar mass using a sample of galaxies constructed from the SDSS Data Release 2 (DR2) data. The methods for estimating the stellar masses are described in Kauffmann et al. (2003). Here we make use of these measurements to constrain the relation between galaxy luminosity, stellar mass and $M_{\rm infall}$. To take account the effect of 'cosmic variance' on the observational results, we have constructed a set of 16 mock galaxy catalogues from the simulation with exactly the same geometry and selection function as in the observational sample. The effect of cosmic variance is modelled by placing a virtual observer randomly inside the simulation box when constructing these mock catalogues. For each mock catalogue, we measure $w(r_{\rm p})$ for galaxies in the same intervals of luminosity/stellar mass as in the observations. The $1-\sigma$ variation between these mock catalogues is then added as an additional error in quadrature to the bootstrap errors given by Li et al. (2006). The cosmic variance errors become significant for the low luminosity and low-mass subsamples, particularly at large values of $r_{\rm p}$. The detailed procedure for constructing these mock catalogues will be presented in a separate paper (Li et al., in preparation).

To compare our models with the observations, we need to either convert $w(r_{\rm p})$ to the real-space correlation function $\xi(r)$, or to calculate $w(r_{\rm p})$ from our model galaxy catalogue directly. We tested the method presented by Hawkins et al. (2003) for converting $w(r_{\rm p})$ to $\xi(r)$ on scales less than around $30\,h^{-1}$ Mpc. We find that the conversion amplifies the error and the results for the low luminosity and low-mass bins are then too noisy to provide good constraints on our models. Therefore, we derive $w(r_{\rm p})$ from our catalogue by integrating the real-space correlation function $\xi(r)$:

$$w(r_{\rm p}) = 2\int_0^\infty \xi\left(\sqrt{{r_{\rm p}}^2 + {r_\parallel}^2}\right){\rm d}r_\parallel = 2\int_{r_{\rm p}}^\infty \xi(r)\frac{r\,{\rm d}r}{\sqrt{r^2-{r_{\rm p}}^2}}.$$

We truncate the integration at $r = 60\,h^{-1}$ Mpc and the resulting $w(r_{\rm p})$ is reliable up to a scale of $\sim 10\,h^{-1}$ Mpc.

We now generate a grid of models by systematically varying the five parameters listed in Table 1. We compare each model with the galaxy luminosity function (Blanton et al. 2003b) and the $w(r_{\rm p})$ measurements in different ranges in luminosity. We define the best-fitting model to be the one giving a minimum χ^2 defined as follows:

$$\chi^2 = \frac{\chi^2(\Phi)}{N_\Phi} + \frac{\chi^2_{\rm corr}}{N_{\rm corr}}$$

with

$$\chi^2(\Phi) = \sum_{N_\Phi}\left[\frac{\Phi - \Phi_{\rm SDSS}}{\sigma(\Phi_{\rm SDSS})}\right]^2$$

and

$$\chi^2_{\rm corr} = \sum_{N_{\rm corr}}\left[\frac{w(r_{\rm p}) - w(r_{\rm p})_{\rm SDSS}}{\sigma(w(r_{\rm p})_{\rm SDSS})}\right]^2.$$

N_Φ is the number of points over which the luminosity function is measured ($N_\Phi = 102$ for the r-band absolute magnitude ranging from -18 to -23). $N_{\rm corr}$ is the number of points over which the correlation function is measured ($N_{\rm corr} = 93$, ranging from 0.11 to $8.97\,h^{-1}$ Mpc for luminosity bins $[-19, -18]$, $[-20, -19]$, $[-21, -20]$, $[-22, -21]$ and from 0.57 to $8.97\,h^{-1}$ Mpc for the most luminous bin $[-23, -22]$ in the r band).

Table 2. Best-fitting parameter values for the relations between $M_{\rm infall}$ and $M_{\rm stars}$ and L (M_r) as derived from the SDSS data.

		$M_0(h^{-1}\ {\rm M}_\odot)$	α	β	log (k)	σ	χ^2	$\chi^2(\Phi)/N_\Phi$
$M_{\rm stars}({\rm M}_\odot)$	Total	3.15×10^{11}	0.118	2.87	10.26	0.326	16.96	2.487
	Central	3.33×10^{11}	0.276	2.59	10.27	0.241	5.351	1.850
	Satellite	4.64×10^{11}	0.122	2.48	10.26	0.334		
$L(M_r)$	Central	3.41×10^{11}	0.221	1.67	8.13	0.440	6.115	3.348
	Satellite	2.58×10^{11}	0.345	3.83	7.71	0.742		

To compare with the SDSS observations, where the median galaxy redshift is around 0.1, we correct the r-band absolute magnitude M_r of each model galaxy to its $z = 0.1$ value $M_{0.1r}$ using the K-correction code (KCORRECT v3_1B) of Blanton et al. (2003a) and the luminosity evolution model of Blanton et al. (2003b). To calculate the K- and E-correction, each galaxy is assigned a redshift by placing a virtual observer at the centre of the simulation box. The redshift as 'seen' by the observer is thus determined by the comoving distance to the observer and the peculiar velocity of the galaxy. The corrected r-band magnitude is given by

$$M_{0.1r} = -2.5\log L + K_{\rm correction} + E_{\rm correction} - 5\log h.$$

Our best-fit model has the parameters: $M_0 = 3.41 \times 10^{11}\,h^{-1}\,{\rm M}_\odot$, $\alpha = 0.221$, $\beta = 1.67$, $k = 8.13$ and $\sigma = 0.440$ for the central galaxies and $M_0 = 2.58 \times 10^{11}\,h^{-1}\,{\rm M}_\odot$, $\alpha = 0.345$, $\beta = 3.83$, $k = 7.71$ and $\sigma = 0.742$ for the satellite galaxies (see Table 2). The resulting luminosity function and correlation functions are shown in Fig. 9. $\chi^2(\Phi)/N_\Phi$ is 3.348 and the total χ^2 is 6.115. Also plotted are the results of central and satellite subsamples of our parametrized model, shown by green dashed and dotted lines. The drop in the correlation function on scales larger than $\sim 10\,h^{-1}$ Mpc is not caused by a poor fit; it is due to the truncation of our integration of the real-space correlation function at $r = 60\,h^{-1}$ Mpc^{-1}.

We now carry out the same analysis for stellar mass, rather than luminosity. We have constructed the stellar mass function directly from the SDSS DR2 data (Fig. 10, left-hand panel) and use this, in conjunction with the measurements of $w(r_{\rm p})$ as a function of stellar mass published by Li et al. (2006) to constrain the $M_{\rm stars}-M_{\rm infall}$ relation. In the computation of stellar mass function, we have corrected for the volume effect by weighting each galaxy by a factor of $V_{\rm survey}/V_{\rm max}$, where $V_{\rm survey}$ is the volume for the sample and $V_{\rm max}$ is the maximum volume over which the galaxy could be observed within the sample redshift range ($0.01 < z < 0.3$) and within the range of r-band apparent magnitude ($14.5 < r < 17.77$). A Schechter function provides a good fit to our measurement at stellar mass $M_{\rm stars} < 10^{11.5}\,h^{-2}\,{\rm M}_\odot$. We find best-fitting parameters: $\Phi^* = (0.0204 \pm 0.0001)\,h^3\,{\rm Mpc}^{-3}$, $\alpha = -1.073 \pm 0.003$ and $M^*_{\rm stars} = (4.11 \pm 0.02) \times 10^{10}\,h^{-2}\,{\rm M}_\odot$. This corresponds to a stellar mass density of $(8.779 \pm 0.067) \times 10^8\,h\ {\rm M}_\odot\ {\rm Mpc}^{-3}$.

We fit our models to 30 points along the stellar mass function and 20 points along the correlation function for five different stellar mass bins ranging from 10^9 to $10^{12}\,{\rm M}_\odot$. The parameters of the best-fitting models are listed in Table 2. For the stellar mass function, the errors due to sample size are much smaller than the systematic errors in the stellar mass estimates themselves. We therefore assign the same error to all points at stellar masses less than $10^{11.5}\,h^{-2}\,{\rm M}_\odot$ (the error is equal to the value at that mass). In our first attempt at fitting the data, we assumed that the $M_{\rm stars}$–$M_{\rm infall}$ would be the same for central and satellite galaxies, because the relations are very similar in the semi-analytic galaxy catalogues. The red dashed lines in Fig. 10 show the best-fitting results. The model clearly overpredicts the clustering of the more massive galaxies on small scales. If we allow the relation between $M_{\rm stars}$ and $M_{\rm infall}$ to differ for central and satellite galaxies, we obtain the results shown by the blue dotted lines, which are considerably better.

The best-fitting r-band luminosity–$M_{\rm infall}$ and $M_{\rm stars}$–$M_{\rm infall}$ relations derived from our models are illustrated in Fig. 11. Results are shown separately for central galaxies (blue) and satellite galaxies

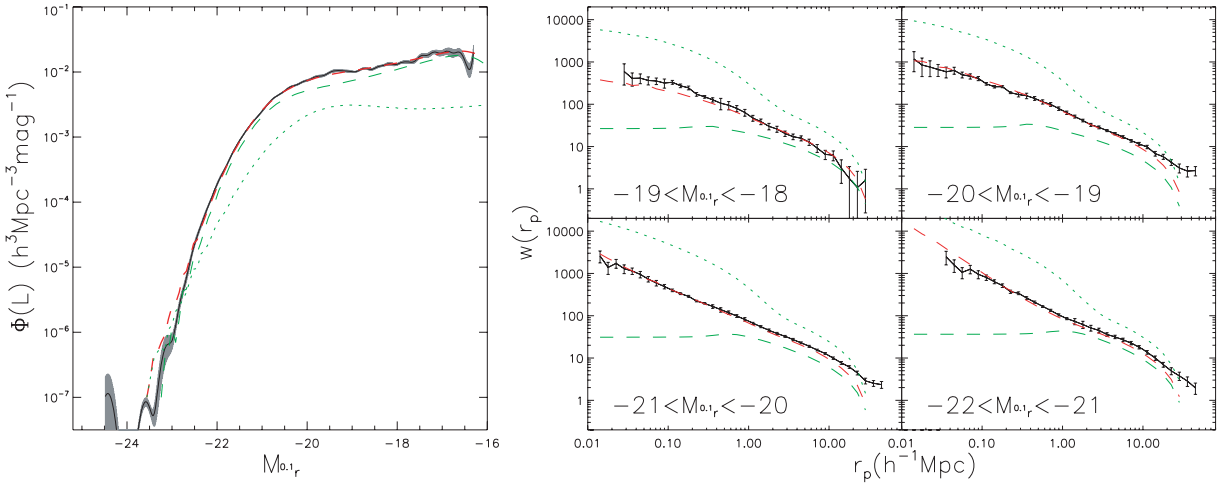

Figure 9. Best-fitting model to the luminosity function and the correlation function evaluated in different luminosity bins using data from the SDSS. Solid lines with error bars are the SDSS results, and red dashed lines are from our parametrized model. Green dashed/dotted lines are results for central/non-central subsamples of our parametrized model.

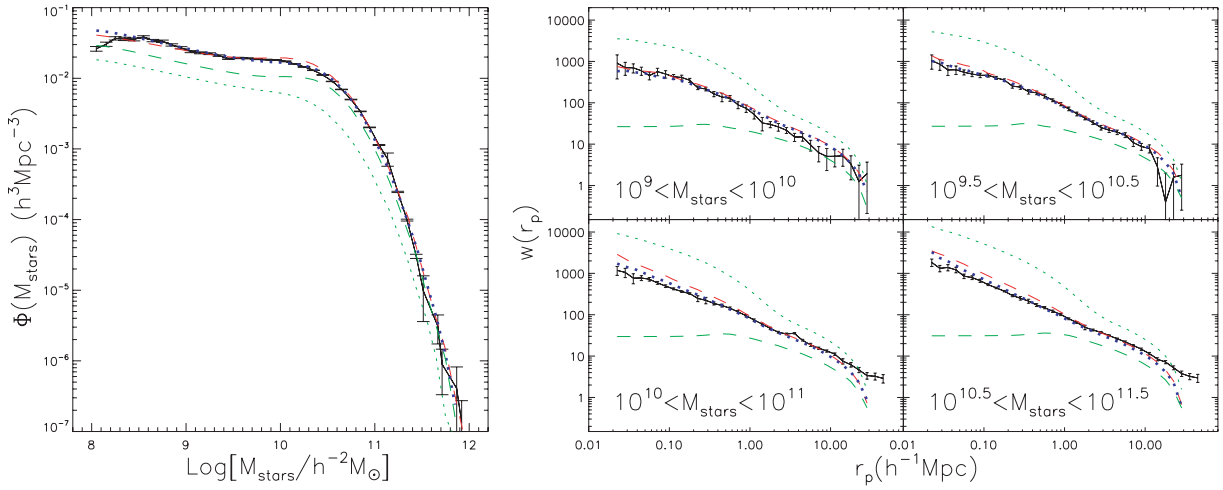

Figure 10. Best-fitting model to the stellar mass function and the correlation function evaluated in different stellar mass bins using data from the SDSS. Symbols with error bars are the SDSS results, and dashed red lines are from our parametrized model. Dotted blue lines show the results obtained when central and satellite galaxies are treated separately. Green dashed/dotted lines are results for central/non-central subsamples of the parametrized model when central and satellite galaxies are treated separately.

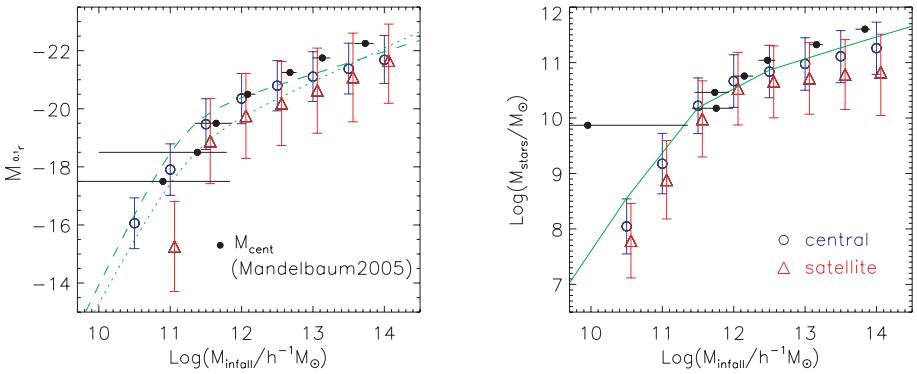

Figure 11. Best-fitting L–$M_{\rm infall}$ and $M_{\rm stars}$–$M_{\rm infall}$ relations as constrained by the SDSS data. Blue circles are the central galaxies and red triangles are satellites. Green lines are the best-fitting relations from the semi-analytic catalogue of Croton et al. (2006). Filled circles show the central halo mass from the galaxy–galaxy lensing results of Mandelbaum et al. (2006); error bars are the 95 per cent confidence limits. The results shown are the combined sample of early- and late-type galaxies (Mandelbaum, private communication).

(red). In our models, satellite galaxies have lower luminosities and smaller stellar masses than central galaxies at a given value of $M_{\rm infall}$. This effect is larger for luminosity than for stellar mass, particularly at low values of $M_{\rm infall}$.

For comparison, we also plot the L–$M_{\rm infall}$ and $M_{\rm stars}$–$M_{\rm infall}$ relations from the semi-analytic galaxy catalogue (Croton et al. 2006). We transform the b_j-band magnitude of semi-analytic catalogue to r band in SDSS according to the luminosity functions of 2dFGRS (Madgwick et al. 2002) and SDSS (Blanton et al. 2003b), and make a shift of 0.9 dex to do the comparison. Because Croton et al. assumed a Salpeter initial mass function, the stellar mass-to-light ratios of the galaxies in their catalogue will be a factor of $\sim$2 higher than in the SDSS data sample. This is because the stellar masses of Kauffmann et al. (2003) have been derived using a Kroupa (2001) initial mass function (IMF). As discussed by Kauffmann et al. (2003), the Salpeter IMF yields stellar masses for elliptical galaxies that *exceed* estimates of their dynamical masses (Cappellari et al. 2006). The Salpeter IMF is clearly unphysical and should be dropped. For the comparison shown in Fig. 11, we have simply scaled the stellar masses in the Croton et al. catalogues by multiplying by a factor 0.5, which should give almost the same results as rerunning the semi-analytic model with the Kroupa IMF. Compared with our results, the semi-analytic catalogue yields systematically higher luminosities and stellar masses at low values of $M_{\rm infall}$, particularly for satellite galaxies. The agreement with the semi-analytic catalogue at $M_{\rm infall} > 10^{11.5}\,h^{-1}\,{\rm M}_{\odot}$ is quite good.

Recently Mandelbaum et al. (2005, 2006) have used galaxy–galaxy weak lensing measurements from SDSS data to explore the connection between galaxies and dark matter. They compare the predicted lensing signal from a halo model constructed using a dissipationless simulation, and extract median/mean halo masses and satellite fractions for galaxies as a function of luminosity, stellar mass and morphology. We plot their estimates of the mean central halo mass as a function of r-band absolute magnitude and stellar mass as filled circles in Fig. 11. The results shown are the combined sample of early- and late-type galaxies (Mandelbaum, private communication). These measurements should be compared with our

blue points, which show the mean halo masses of present-day central galaxies. As can be seen, there is remarkably good agreement between the two methods, both for luminosity and for stellar mass.

6 CONCLUSIONS AND DISCUSSION

We have constructed a new statistical model of galaxy clustering for use in high-resolution numerical simulations of structure formation. Unlike classic HOD models, galaxy positions and velocities are determined in a self-consistent way by following the full orbital and merging histories of all the haloes and subhaloes in the simulation. We believe that this methodology has advantages over the traditional approach. Most HOD models assume that the galaxy content of a halo of given mass is statistically independent of its larger scale environment. Recently, Gao, Springel & White (2005) have shown that there exists an age dependence of halo clustering: haloes that are formed earlier are more clustered than haloes that are assembled more recently, indicating that this assumption may not be as safe as previously thought. Since the positions and the velocities of the galaxies in our model are determined directly from the simulation, we avoid these difficulties.

Our methodology also takes into account the contribution of 'orphaned' galaxies, which have lost their haloes due to tidal stripping. These galaxies contribute significantly to the clustering amplitude of low-mass galaxies on scales less than $\sim 1\ h^{-1}$ Mpc. We have chosen to parametrize the observed properties of galaxies (in particular their luminosity and stellar mass) as a function of the quantity $M_{\rm infall}$, the mass of the halo at the epoch when the galaxy was last the central object in its halo. Using the semi-analytic model results as a reference, we adopt a double power-law form for this relation, and we show that this allows us to recover the mass/luminosity function and the correlation function in different ranges of mass and luminosity with high accuracy.

We then apply our model to measurements of these quantities using data from the SDSS. We find that for a given value of $M_{\rm infall}$, satellite galaxies are required to be less luminous and less massive than central galaxies. This effect is stronger at low values of $M_{\rm infall}$. In the semi-analytic models, satellite galaxies fade in luminosity after they fall into a larger halo because they no longer accrete gas and their star formation rates then decline. The catalogues of Croton et al. (2006) do show differences between satellite and central galaxy luminosities at a fixed value of $M_{\rm infall}$, but the effect is not quite as strong as the data demands, particularly for low-mass haloes. This may indicate that the efficiency with which baryons are converted into stars in low-mass haloes is higher at the present day than it was in the past. The fact that the standard ΛCDM model predicts more low-mass galaxies than observed is very well documented in the literature (Moore et al. 1999). Many authors have tried to invoke mechanisms for 'suppressing' star formation in these systems (Kauffmann, White & Guiderdoni 1993; Somerville 2002) and most of these mechanisms operate more effectively at higher redshifts.

Finally, we compare our relations between galaxy luminosity, stellar mass and host halo mass with similar relations derived using galaxy–galaxy weak lensing measurements. The excellent agreement between these two completely independent methods is very encouraging.

ACKNOWLEDGMENTS

We are grateful to Simon White, Darren Croton and the referee J. S Bullock for their detailed comments and suggestions on our paper. We thank Rachel Mandelbaum for providing their results on Fig 11. CL acknowledges the financial support of the exchange programme between Chinese Academy of Sciences and the Max Planck Society. GDL thanks the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research and the Programme for Investment in the Future (ZIP) of the German Government for financial support.

The simulation used in this paper was carried out as part of the programme of the Virgo Consortium on the Regatta Supercomputer of the Computing Centre of the Max Planck Society in Garching.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, Cambridge University, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington.

REFERENCES

Arnaboldi M., 2004, in Duc P.-A., Braine J., Brinks E., eds, IAU Symp. 217, Intracluster Stellar Population. Astron. Soc. Pac., San Francisco, p. 54
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Berlind A. A. et al., 2003, ApJ, 593, 1
Blanton M. R. et al., 2003a, AJ, 125, 2348
Blanton M. R. et al., 2003b, ApJ, 592, 819
Cappellari M. et al., 2006, MNRAS, 366, 1126
Colless M. et al., 2001, MNRAS, 328, 1039
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, in press (astro-ph/0512234)
Croton D. J. et al., 2006, MNRAS, 365, 11
De Lucia G., Kauffmann G., Springel V., White S. D. M., Lanzoni B., Stoehr F., Tormen G., Yoshida N., 2004, MNRAS, 348, 333
Feldmeier J. J., Mihos J. C., Morrison H. L., Harding P., Kaib N., Dubinski J., 2004, ApJ, 609, 617
Gao L., De Lucia G., White S. D. M., Jenkins A., 2004, MNRAS, 352, L1
Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66
Hawkins E. et al., 2003, MNRAS, 346, 78
Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Kauffmann G., Nusser A., Steinmetz M., 1997, MNRAS, 286, 795
Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188
Kauffmann G. et al., 2003, MNRAS, 341, 33
Kroupa P., 2001, MNRAS, 322, 231
Li C., Kauffmann G., Jing Y. P., White S. D. M., Börner G., Cheng F. Z., 2006, MNRAS, 368, 21
Madgwick D. S. et al., 2002, MNRAS, 333, 133

Mandelbaum R., Tasitsiomi A., Seljak U., Kravtsov A. V., Wechsler R. H., 2005, MNRAS, 362, 1451
Mandelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006, MNRAS, 368, 715
Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19
Norberg P. et al., 2002a, MNRAS, 332, 827
Norberg P. et al., 2002b, MNRAS, 336, 907
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Pearce F. R., Jenkins A., Frenk C. S., White S. D. M., Thomas P. A., Couchman H. M. P., Peacock J. A., Efstathiou G., 2001, MNRAS, 326, 649
Seljak U., 2000, MNRAS, 318, 203
Somerville R. S., 2002, ApJ, 572, L23
Spergel D. N. et al., 2003, ApJS, 148, 175
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2005, Nat, 435, 629
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
Vale A., Ostriker J. P., 2006, MNRAS, in press (doi: 10.1111/j.1365-2966.2006.10605.x) (astro-ph/0511816)
van den Bosch F. C., Yang X., Mo H. J., 2003, MNRAS, 340, 771
Wang Y., Yang X., Mo H. J., van den Bosch F. C., Chu Y., 2004, MNRAS, 353, 287
White M., Hernquist L., Springel V., 2001, ApJ, 550, L129
Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057
Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., 2005, MNRAS, 358, 217
York D. G., Adelman J., Anderson J. E., Anderson S. F., Annis J., Bahcall N. A., Bakken J. A., Barkhouser R. A., 2000, AJ, 120, 1579
Yoshikawa K., Taruya A., Jing Y. P., Suto Y., 2001, ApJ, 558, 520
Zehavi I. et al., 2004, ApJ, 608, 16
Zehavi I. et al., 2005, ApJ, 630, 1
Zibetti S., White S. D. M., Schneider D. P., Brinkmann J., 2005, MNRAS, 358, 949

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