A halo-based galaxy group finder: calibration and application to the 2dFGRS

Xiaohu Yang,[1] H. J. Mo,[1] Frank C. van den Bosch[2] and Y. P. Jing[3]★

[1]*Department of Astronomy, University of Massachusetts, Amherst, MA 01003-9305, USA*
[2]*Department of Physics, Swiss Federal Institute of Technology, ETH Hönggerberg, Zurich CH-8093, Switzerland*
[3]*Shanghai Astronomical Observatory; the Partner Group of MPA, Nandan Road 80, Shanghai 200030, China*

ABSTRACT

We use the halo occupation model to calibrate galaxy group finders in magnitude limited redshift surveys. Because, according to the current scenario of structure formation, galaxy groups are associated with cold dark matter (CDM) haloes, we make use of the properties of the halo population in the design of our group finder. The method starts with an assumed mass-to-light ratio to assign a tentative mass to each group. This mass is used to estimate the size and velocity dispersion of the underlying halo that hosts the group, which in turn is used to determine group membership (in redshift space). This procedure is repeated until no further changes occur in group memberships. We find that the final groups selected this way are insensitive to the mass-to-light ratio assumed. We use mock catalogues, constructed using the conditional luminosity function (CLF), to test the performance of our group finder in terms of completeness of true members and contamination by interlopers. Our group finder is more successful than the conventional friends-of-friends (FOF) group finder in assigning galaxies in common dark matter haloes to a single group. We apply our group finder to the 2-degree Field Galaxy Redshift Survey (2dFGRS) and compare the resulting group properties with model predictions based on the CLF. For the ΛCDM concordance cosmology, we find a clear discrepancy between the model and data in the sense that the model predicts too many rich groups. In order to match the observational results, we have to either increase the mass-to-light ratios of rich clusters to a level significantly higher than current observational estimates, or to assume $\sigma_8 \simeq 0.7$, compared with the concordance value of 0.9.

Key words: methods: statistical – galaxies: haloes – dark matter – large-scale structure of the Universe.

1 INTRODUCTION

It is common practice to apply a group finder to large galaxy redshift surveys in order to assign galaxies to groups.[1] The clustering properties of galaxies can then be studied by analysing the spatial clustering of the groups and the distribution functions of the groups with respect to their internal properties, such as luminosity, velocity dispersion, mass, shape and galaxy population. Such analyses have been carried out using various galaxy redshift surveys, most noticeably the CfA redshift survey (e.g. Geller & Huchra 1983), the Las Campanas Redshift Survey (e.g. Tucker et al. 2000), the 2-degree Field Galaxy Redshift Survey (hereafter 2dFGRS; Merchán &

Zandivarez 2002; Eke et al. 2004a,b) and the Sloan Digital Sky Survey (hereafter SDSS; Bahcall et al. 2003; Lee et al. 2004).

One of the problems in such analyses is that the properties of the groups depend on the group finder used to identify groups and it is in general quite hard to judge whether the groups identified are true associations of galaxies in space. There are several reasons for such complication. First of all, galaxies are observed in redshift space, not in the real space, and so every group finder has to contend with the redshift distortion in the apparent clustering patterns of galaxies. Secondly, redshift surveys are usually apparent magnitude limited and so any criteria for clustering based on the distances between galaxies has to take into account the variation of the mean intergalaxy separation with distance. Thirdly, because only a finite number of galaxies can be observed for each group, shot noise can also affect the accuracy of membership assignment, especially for small groups that contain only a small number of observable galaxies.

★E-mail: xhyang@astro.umass.edu
[1] In this paper, we refer to a system of galaxies as a group regardless of its richness, including isolated galaxies (i.e. groups with a single member).

One may argue that this is less a real problem than a matter of definition for galaxy groups. However, because the ultimate goal of such analyses is to compare with theory, the definition must have a sound physical basis, in order to make a meaningful comparison between observation and theory. From the point of view of current theory of galaxy formation, the most natural reference for defining galaxy groups is dark matter haloes. These are quasi-equilibrium systems of dark matter particles, formed through non-linear gravitational collapse. In the standard Λ cold dark matter (CDM) model favoured by current observations, most mass at any given time is bound within dark haloes; galaxies and other luminous objects are assumed to form by cooling and condensation of the baryons within haloes. Thus, it is extremely useful to have a group finder that can group galaxies according to their common dark haloes. Unfortunately, dark haloes are not directly observable and so it is not possible to define groups directly from dark matter distribution. On the other hand, because the current CDM scenario is very successful in explaining a very large range of observational data (e.g. Spergel et al. 2003), we may use the properties of dark haloes in current CDM models as a guide in our design of a group finder. Much progress has been made in recent years in understanding the relationship between CDM haloes and galaxies, based on numerical simulations (e.g. Katz, Weinberg & Hernquist 1996; Pearce et al. 2000) and semi-analytic models (e.g. Kauffmann et al. 1999; Somerville & Primack 1999; Cole et al. 2000). However, our understanding of the details about galaxy formation in CDM haloes is still quite poor and so calibrations of group finders have not been carried out in a reliable, model-independent way. Furthermore, to produce a mock catalogue that can not only cover a volume as large as current large redshift surveys of galaxies, such as 2dFGRS and SDSS, but also have the ability to resolve faint galaxies, is not trivial for both numerical simulation and semi-analytical modelling. What one really needs is an empirical model that can correctly partition the galaxy population into dark haloes but does not depend on uncertain details about galaxy formation in dark haloes.

The halo occupation model recently developed has this spirit. In this model, one simply specifies halo occupation numbers, $\langle N(M) \rangle$, which describe how many galaxies on average occupy a halo of mass M. Many recent investigations have used such halo occupation models to study various aspects of galaxy clustering (Jing, Mo & Börner 1998; Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; White 2001; Berlind & Weinberg 2002; Bullock, Wechsler & Somerville 2002; Jing, Börner & Suto 2002; Kang et al. 2002; Marinoni & Hudson 2002; Scranton 2002; Zheng et al. 2002; Kochanek et al. 2003; Magliocchetti & Porciani 2003; Yan, Madgwick & White 2003; Yan, White & Coil 2004). In two recent papers, Yang, Mo & van den Bosch (2003, hereafter Paper I) and van den Bosch, Yang & Mo (2003, hereafter Paper II) have taken this halo occupation approach one step further by considering the occupation as a function of galaxy luminosity and type. They introduced the conditional luminosity function (hereafter CLF) $\Phi(L \mid M) \, dL$, which gives the number of galaxies with luminosities in the range $L \pm dL/2$ that reside in haloes of mass M. The advantage of this CLF over the halo occupation function $\langle N(M) \rangle$ is that it allows one to address the clustering properties of galaxies as function of luminosity. In particular, such a model can be used to populate dark matter haloes in high-resolution N-body simulations to construct realistic mock galaxy redshift surveys that automatically have the correct galaxy abundances and correlation lengths as function of galaxy luminosity and type. As shown in Yang et al. (2004), the mock galaxy catalogues constructed in this way can recover many of the properties of galaxy clustering in redshift

space. Further analyses with such mock catalogues show that the model is also successful in matching the environmental dependence of galaxy luminosity function (Mo et al. 2004), the kinematics of satellite galaxies around galaxies (van den Bosch et al. 2004a,b) and the three-point correlation function (Wang et al. 2004).

The aim of this paper is two-fold. First, we intend to develop a group finder that is based on the consideration given above and can group galaxies according to their common haloes. We test the reliability of the method with the extensive use of various mock samples constructed from the CLF model. Secondly, we apply the group finder to the 2dFGRS and compare the properties of the groups so obtained with model predictions. The outline of the paper is as follows. In Section 2, we describe our halo-based group finder, which we test in Section 3 with the extensive use of detailed mock samples constructed from the CLF model. In Section 4, we apply our group finder to the 2dFGRS and compare the properties of the groups thus obtained with model predictions. We summarize our results in Section 5.

Unless stated otherwise, we consider a flat ΛCDM cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $h = H_0/(100 \, \text{km s}^{-1} \, \text{Mpc}^{-1}) = 0.7$ and with initial density fluctuations described by a scale-invariant power spectrum with normalization $\sigma_8 = 0.9$. These cosmological parameters are in good agreement with a wide range of observations, including the recent *WMAP* results (Spergel et al. 2003), and in what follows we refer to it as the concordance cosmology.

2 THE GROUP FINDER

Our aim here is to develop a group finder that assigns galaxies in a common halo to a single group. The properties of the halo population in the standard ΛCDM model are well understood, largely as a result of a combination of N-body simulations and analytical models. Dark matter haloes are defined as virialized structures with a mean over-density of approximately 180. Their density profiles are well described by the so-called NFW profile (Navarro, Frenk & White 1997):

$$\rho(r) = \frac{\bar{\delta}\bar{\rho}}{(r/r_s)(1 + r/r_s)^2}, \tag{1}$$

where r_s is a characteristic radius, $\bar{\rho}$ is the average density of the Universe and $\bar{\delta}$ is a dimensionless amplitude that can be expressed in terms of the halo concentration parameter $c = r_{180}/r_s$ as

$$\bar{\delta} = \frac{180}{3} \frac{c^3}{\ln(1 + c) - c/(1 + c)}. \tag{2}$$

Here, r_{180} is the radius within which the halo has an average over-density of 180. Numerical simulations show that the halo concentration depends on halo mass, we use the relation given by Bullock et al. (2001) and properly rescaled to our definition.

Both observations and numerical simulations suggest that the brightest galaxy in each dark matter halo resides at rest at the centre, while the number density distribution of the fainter, satellite galaxies matches that of the dark matter particles (e.g. Carlberg et al. 1997; van der Marel et al. 2000; Berlind et al. 2003; Diemand, Moore & Stadel 2004; Lin, Mohr & Stanford 2004; Rines et al. 2004). Furthermore, although velocity bias may exist in the sense that galaxies move with different velocities as the dark matter particles at the same location, such bias is found to be not large (Berlind et al. 2003; Yoshikawa, Jing & Börner 2003). These findings suggest that it may be possible to use information regarding the spatial distribution and velocity dispersion of galaxy group members to estimate halo masses and vice versa. Motivated by these considerations, we design a group finder that consists of the following steps.

(i) Step 1: we combine two different methods to identify the centres of potential groups. First, we use the traditional Friends-Of-Friends (FOF) algorithm to assign galaxies into groups. Because we are working in redshift space, we separately define linking lengths along the line of sight (ℓ_z) and in the transverse direction (ℓ_p). Because the purpose here is only to identify the group centres, we use relatively small linking lengths: $\ell_z = 0.3$ and $\ell_p = 0.05$, both in units of the mean separation of galaxies. Note that for an apparent magnitude limited survey the mean separation of galaxies is a function of redshift, which we take into account. The geometrical centres of all FOF groups thus identified with more than two galaxies are considered as centres of potential groups. Next, from all galaxies not yet linked together by these FOF groups, we select bright, relatively isolated galaxies, which we also associate with the centres of potential groups. Following McKay et al. (2002), Brainerd & Specian (2003), Prada et al. (2003) and van den Bosch et al. (2004a), we identify a galaxy as central and thus as the centre of a potential group, when it is the brightest galaxy in a cylinder of radius $1\,h^{-1}$ Mpc and velocity depth $\pm 500\,\mathrm{km\,s^{-1}}$.

(ii) Step 2: we estimate the luminosity of a selected potential group using

$$L_{\mathrm{group}} = \sum_i \frac{L_i}{f_c(L_i)}, \tag{3}$$

where L_i is the luminosity of each galaxy in the group and f_c is the luminosity-dependent incompleteness of the survey (which is relevant only when the group finder is applied to real data, such as the 2dFGRS). The total luminosity of the group is approximated by

$$L_{\mathrm{total}} = L_{\mathrm{group}} \frac{\int_0^\infty L\phi(L)\,\mathrm{d}L}{\int_{L_{\mathrm{lim}}}^\infty L\phi(L)\,\mathrm{d}L}, \tag{4}$$

where L_{lim} is the minimum luminosity of a galaxy that can be observed at the redshift of the group and $\phi(L)$ is the galaxy luminosity function

$$\phi(L)\mathrm{d}L = \phi_* \left(\frac{L}{L_*}\right)^\alpha \exp\left(-\frac{L}{L_*}\right) \frac{\mathrm{d}L}{L_*}. \tag{5}$$

Throughout this paper, galaxy luminosities are defined in the photometric b_J band. For the luminosity function, we take the parameters from Norberg et al. (2002): $(\phi_*, M_*, \alpha) = (0.0161, -19.66, -1.21)$.

(iii) Step 3: from L_{total} and a model for the group mass-to-light ratio (see below), we compute an estimate of the halo mass associated with the group in consideration. From this estimate, we also compute the halo radius r_{180}, the virial radius r_{vir}[2] and the virial velocity $V_{\mathrm{vir}} = (GM/r_{\mathrm{vir}})^{1/2}$. The line-of-sight velocity dispersion of the galaxies within the dark matter halo is assumed to be $\sigma = V_{\mathrm{vir}}/\sqrt{2}$.

(iv) Step 4: once we have a group centre, and a tentative estimate of the group size, mass and velocity dispersion, we can assign galaxies to this group according to the properties of the associated haloes. If we assume that the phase-space distribution of galaxies follows that of the dark matter particles, the number density contrast of galaxies in redshift space around the group centre (= centre of dark matter halo) at redshift z_{group} can be written as

$$P_{\mathrm{M}}(R, \Delta z) = \frac{H_0}{c} \frac{\Sigma(R)}{\bar{\rho}} p(\Delta z). \tag{6}$$

Here, $\Delta z = z - z_{\mathrm{group}}$ and $\Sigma(R)$ is the projected surface density of a (spherical) NFW halo:

$$\Sigma(R) = 2r_s \bar{\delta} \bar{\rho} f(R/r_c), \tag{7}$$

with

$$f(x) = \begin{cases} \frac{1}{x^2-1} \left[1 - \frac{\ln(1+\sqrt{1-x^2})/x}{\sqrt{1-x^2}} \right] & \text{if } x < 1 \\ 1/3 & \text{if } x = 1 \\ \frac{1}{x^2-1} \left(1 - \frac{\mathrm{atan}\sqrt{x^2-1}}{\sqrt{x^2-1}} \right) & \text{if } x > 1 \end{cases} \tag{8}$$

The function $p(\Delta z)\,\mathrm{d}\Delta z$ describes the redshift distribution of galaxies within the halo for which we adopt a Gaussian form

$$p(\Delta z) = \frac{c}{\sqrt{2\pi}\sigma(1+z_{\mathrm{group}})} \exp\left[\frac{-(c\Delta z)^2}{2\sigma^2(1+z_{\mathrm{group}})^2}\right], \tag{9}$$

where σ is the rest-frame velocity dispersion.

Thus defined, $P_{\mathrm{M}}(R, \Delta z)$ is the three-dimensional density contrast in redshift space. In order to decide whether a galaxy should be assigned to a particular group, we proceed as follows. For each galaxy we loop over all groups, and compute the corresponding distance $(R, \Delta z)$ between galaxy and group centre. Here, R is the projected distance at the redshift of the group. If $P_{\mathrm{M}}(R, \Delta z) \geqslant B$, with B an appropriately chosen background level, the galaxy is assigned to the group. If a galaxy can be assigned to more than one group, it is only assigned to the group for which $P_{\mathrm{M}}(R, \Delta z)$ has the highest value. Finally, if all members of two groups can be assigned to one group according to the above criterion, the two groups are merged into a single group. The background level B defines a threshold density contrast in redshift space whose value we calibrate using mock galaxy catalogues (see Section 3.2).

(v) Step 5: using the group members thus selected, we recompute the group centre and go back to step 2, iterating until there is no further change in the memberships of groups. Note that, unlike with the traditional FOF method, this group finder also identifies groups with only one member.

The basic idea of this group finder is similar to that of the matched filter algorithms developed by Postman et al. (1996; see also Kepner et al. 1999; Kim et al. 2002; White & Kochanek 2002; Kochanek et al. 2003), although we also make use of the galaxy kinematics in our analyses. The crucial ingredient of our method is to somehow obtain a reliable estimate of the halo mass associated with the galaxy group from observable properties of its selected member galaxies. Ideally, this mass estimate should be model-independent. For example, one might hope to estimate the mass from the velocity dispersion of the selected members, using the fact that $M \propto \sigma^3$. Unfortunately, we found that this is not a reliable method, simply because the estimate of σ is too noisy when only a few group members are available. Therefore, we decided to use a model for the mass-to-light ratio to estimate the group mass from the total luminosity of the selected galaxies. Because total group luminosity is dominated by the few brightest galaxies, the estimated mass is much less sensitive to the absence of faint group members. A downside of this method, however, is that it requires a model for the mass-to-light ratios. Fortunately, we found that the memberships of the selected groups are remarkably insensitive to the adopted model. The reason for this is that, even if the estimated mass is wrong by a factor of 10, the implied radius and velocity dispersion, used in the membership determination, change only by a factor of 2.15.

---

[2] The virial radius is defined as the radius inside of which the average density is Δ_{vir} times the critical density, with Δ_{vir} given by Bryan & Norman (1998).

In what follows, we use the average mass-to-light ratios obtained by Paper II from the CLF formalism

$$\frac{M}{L} = \begin{cases} \frac{1}{2} \left(\frac{M}{L}\right)_0 \left[\left(\frac{M}{M_1}\right)^{-\gamma_1} + \left(\frac{M}{M_1}\right)^{\gamma_2} \right] & \text{if } M_{14} \leqslant 1 \\ (M/L)_{cl} & \text{if } M_{14} > 1 \end{cases}, \quad (10)$$

where we set the model parameters to those of their model D: $M_1 = 10^{10.94}\, h^{-1}\, M_\odot$, $\gamma_1 = 2.02$, $\gamma_2 = 0.30$, $(M/L)_0 = 124\, h\, (M/L)_\odot$ and $(M/L)_{cl} = 500\, h\, (M/L)_\odot$. M_{14} is the halo mass in units of 10^{14} $h^{-1}\, M_\odot$. As we demonstrate in the next section, even if we adopt a constant mass-to-light ratio in step 3 of $M/L = 400\, h\, (M/L)_\odot$ for all haloes, the selected groups are virtually identical to those selected when using these CLF-based mass-to-light ratios.

3 TESTING THE GROUP FINDER

3.1 Constructing mock samples

We construct mock galaxy samples by populating dark matter haloes in numerical simulations with galaxies of different luminosities, using the CLF formalism developed in Papers I and II. In what follows, we focus on CLF model D defined in Paper II, which is valid for the concordance cosmology considered here, and which yields excellent fits to the observed LFs and the observed correlation lengths as a function of both luminosity and type. The same CLF has also been used in Yang et al. (2004) to construct large, mock galaxy redshift surveys, which they used to compare various large-scale structure statistics with those obtained from the 2dFGRS.

In this paper, we use the same mock samples to test our group finder and to construct detailed mock galaxy redshift surveys for comparison with the 2dFGRS. Here, we briefly summarize the ingredients, but refer the reader to Yang et al. (2004) for details. The mock samples are constructed using a set of *N*-body simulations carried out by Y.P. Jing and Y. Suto (see Jing 2002; Jing & Suto 2002) on the VPP5000 Fujitsu supercomputer of the National Astronomical Observatory of Japan using a vectorized-parallel P³M code. The set consists of a total of six simulations, each of which uses $N = 512^3$ particles to evolve the distribution of dark matter from an initial redshift of $z = 72$ down to $z = 0$ in a ΛCDM concordance cosmology. All simulations consider boxes with periodic boundary conditions; in two cases $L_{\text{box}} = 100\, h^{-1}$ Mpc, while the other four simulations all have $L_{\text{box}} = 300\, h^{-1}$ Mpc. Different simulations with the same box size are completely independent realizations and are used to estimate uncertainties resulting from cosmic variance. The particle masses are $6.2 \times 10^8\, M_\odot$ and $1.7 \times 10^{10}\, M_\odot$ for the small and large box simulations, respectively. Dark matter haloes are identified using the FOF algorithm with a linking length of 0.2 times the mean interparticle separation. For each individual simulation, we construct a catalogue of haloes with 10 particles or more, for which we store the mass, the position of the most bound particle, and the mean velocity and velocity dispersion of the halo. Haloes that are unbound are removed from the sample. In Yang et al. (2004), we have shown that the resulting halo mass functions are in excellent agreement with the analytical halo mass function given by Sheth & Tormen (1999) and Sheth, Mo & Tormen (2001). The mock galaxy distributions are constructed by populating the dark matter haloes in these *N*-body simulations with galaxies according to the CLF. The brightest galaxy in each halo is located at rest at the centre of the halo, while the other galaxies follow a number density distribution that is identical to the mass distribution of the dark matter and with an isotropic velocity dispersion that is identical to that of the dark matter. Finally, when using these mock samples to create

a mock 2dFGRS (hereafter MGRS), we stack simulations with different box sizes and resolutions in order to cover a large redshift range (to match the depth of the 2dFGRS) and to properly sample the faint galaxies. Observational errors and selection effects in the 2dFGRS, such as position-dependent completeness and magnitude limit variations are all taken into account. From our set of six simulation boxes, we construct eight independent MGRSs, which we use to estimate the impact of cosmic variance.

3.2 Volume limited samples

Before applying our group finder to the MGRSs and the 2dFGRS, we first test and calibrate it using a volume limited sample of mock galaxies in one of our simulation boxes with $L_{\text{box}} = 100\, h^{-1}$ Mpc. To mimic observation, we simply assume redshifts to be along one side of the simulation box.

In what follows, a true group is defined as the set of galaxies that reside in the same halo. In order to quantify the group finder performance, we introduce the completeness, F_t, defined as the ratio between the number of true members selected by the group finder and the total number of true group members, and the contamination, F_i, defined as the ratio between the number of false members (interlopers) selected by the group finder and the total number of true members. A perfect group finder has $F_t = 1$ and $F_i = 0$. Therefore, we optimize our group finder by simultaneously maximizing F_t and minimizing F_i. Note that these two criteria may sometimes be in conflict. For example, a certain group finder may select all true members but at the expense of including many interlopers. Other group finders may yield zero interlopers but miss a large fraction of true members. Whether one gives more weight to large F_t or small F_i depends on the question one wishes to address. In this paper, we give equal weight to both criteria and we tune our parameter in the group finder, i.e. the background level B, so that the value of $\Psi \equiv F_t \times (1 - F_i)$ is maximized. The background level B defined in this way is approximately 10 for M_* haloes and Ψ has a weak dependence on the mass of the system.

The performance of our group finder is not very sensitive to the exact value of B used. This is evident from the various curves in Fig. 1, which show that similar values of Ψ are obtained using a fairly large range of background levels. To put the absolute values of B in perspective, we recall that it is used as a threshold for the redshift-space density contrast of groups. Ideally, B should therefore correspond roughly to the redshift-space density contrast at the edge of a halo, i.e.,

$$B \approx \frac{\rho_{\text{red}}(r_{180})}{\bar{\rho}} \approx \frac{\rho(r_{180})}{\bar{\rho}} \frac{(4\pi/3)r_{180}^3}{\pi r_{180}^2 \sigma/H_0}. \quad (11)$$

Using that $\rho(r_{180})/\bar{\rho} \sim 30$ and $\sigma/H_0\, r_{180} \sim 4$, we obtain $B \sim 10$, in excellent agreement with the value obtained by maximizing Ψ. Therefore, in what follows we adopt $B = 10$, independent of halo mass.

Fig. 2 gives a more detailed overview of the completeness and contamination for our fiducial group finder with $B = 10$. Dots correspond to individual groups (haloes), while the big symbols with error bars indicate the average F_t (solid circles) and F_i (open triangles). Only galaxies brighter than a certain absolute-magnitude limit (indicated in the upper right corner of each panel) are used to identify groups. The incompleteness and contamination are also defined with respect to all group members brighter than that limit. Overall the group finder is quite successful, with an average completeness of approximately 90 per cent and a contamination of approximately 20 per cent for groups in dark matter haloes with masses
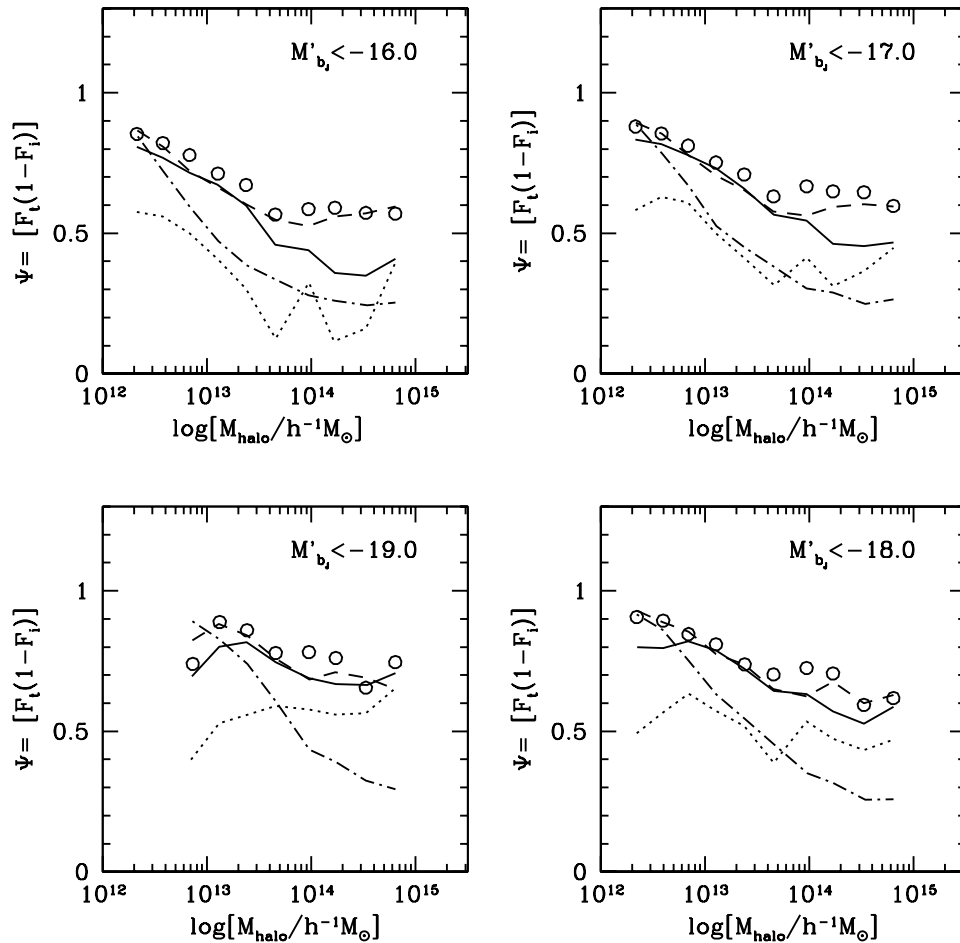
Figure 1. The value of $\Psi \equiv F_t(1 - F_i)$ (where F_t is the completeness factor and F_i is the contamination factor) as a function of halo mass for different choices of the background level B. The circles show the result for our fiducial group finder with $B = 10$. The solid and dashed lines correspond to $B = 5$ and 20, respectively, while the dotted and dot-dashed lines outline two extreme cases with $B = 0$ and 100, respectively. The four different panels show results using different luminosity cuts: only galaxies brighter than the absolute magnitude limit indicated (where $M'_{b_J} = M_{b_J} - 5 \log h$) are used to identify groups and to define F_t and F_i.

spanning the entire range from $\sim 10^{12} \, h^{-1} \, \mathrm{M}_\odot$ to $\sim 10^{15} \, h^{-1} \, \mathrm{M}_\odot$. Note also that the completeness and contamination do not depend significantly on the luminosity limit of the tracer galaxies. This implies that our group finder can make a fairly uniform identification of galaxy groups even in an apparent magnitude limited sample. As we shall see in Section 3.3, this is indeed the case.

In the group finder used thus far, we have used the mass-to-light ratio predicted by the CLF (equation 10) to estimate the sizes and velocity dispersions of the groups. In order to test the sensitivity of the group finder to this model assumption, we performed a number of tests with different mass-to-light ratios. One of the more extreme examples tested is a model with constant mass-to-light ratio of $M/L = 400 \, h(\mathrm{M}/L)_\odot$ independent of halo mass. As shown in Fig. 3, the resulting completeness and contamination are very similar to those obtained using our fiducial mass-to-light ratios of equation (10). We obtained similar results for all mass-to-light ratios tested, indicating that the completeness and contamination levels of our group finder are extremely insensitive to the exact mass-to-light ratios assumed.

Fig. 3 also shows the results for groups identified using the standard FOF algorithm (i.e. only using step 1 in Section 2). We use exactly the same linking lengths as in Eke et al. (2004a):

$$\ell_p = 0.13(\Delta/5)^{0.04} n^{-1/3} \tag{12}$$

in the transverse direction and

$$\ell_z = 1.43(\Delta/5)^{0.16} n^{-1/3} \tag{13}$$

along the line of sight, with n the number density of galaxies. The quantity Δ is the galaxy density contrast relative to the background at the redshift considered in a cylinder of radius $1.5 \, h^{-1}$ Mpc and velocity depth $\pm 1650 \, \mathrm{km \, s^{-1}}$. Although the completeness given by the FOF algorithm is similar to our group finder, the contamination is significantly higher. In addition, the contamination in poor groups increases systematically when only using brighter galaxies. Therefore, the FOF method does not give a uniform identification in an apparent magnitude limited sample (e.g. Diaferio et al. 1999).

To further test the performance of our group finder, we compare properties of the identified groups with those of the true groups. Fig. 4 shows the predicted group luminosity L_{group} (the sum of the luminosities of all members assigned by the group finder) versus the true group luminosity L_{true} (the sum of the luminosities of all true members). We only show results for groups with three or more selected members. Each of the four panels shows the result where galaxies brighter than a certain absolute magnitude limit are used in selecting the groups and in calculating the group luminosities. Overall the correlation is fairly tight, indicating that the incompleteness
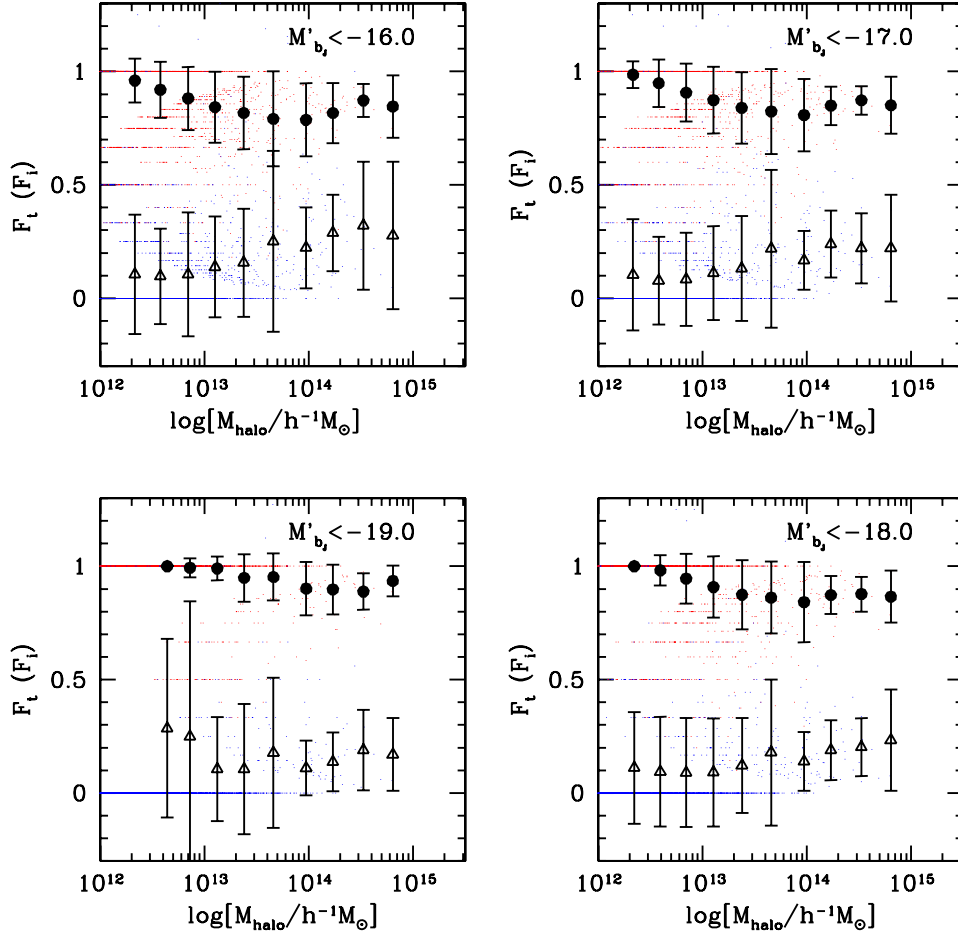
Figure 2. Red and blue points indicate the completeness and contamination, respectively, of individual groups as function of halo mass. Solid dots and open triangles indicate the corresponding averages with the error bars indicating the 1-σ variance. Results correspond to our fiducial group finder with $B = 10$. Different panels correspond to different absolute magnitude limits as in Fig. 1.

and the presence of interlopers does not strongly impact on total group luminosity.

For a system in dynamical equilibrium, its mass may be estimated from its velocity dispersion through the virial theorem. Because the velocity dispersion of a group can be measured from the redshifts of member galaxies, it is interesting to examine how the velocity dispersion among selected members compares to that of the true members. We use the gapper estimator described by Beers, Flynn & Gebhardt (1990) to estimate the group velocity dispersions. The method involves ordering the set of recession velocities $\{v_i\}$ of the N group members and defining gaps as

$$g_i = v_{i+1} - v_i, \quad i = 1, 2, \ldots, N - 1. \tag{14}$$

The rest-frame velocity dispersion is then estimated by

$$\sigma_{\text{gap}} = \frac{\sqrt{\pi}}{(1 + z_{\text{group}})N(N - 1)} \sum_{i=1}^{N-1} w_i g_i, \tag{15}$$

where the weight is defined as $w_i = i(N - i)$. Because there is a central galaxy in each group (which is true in our mock samples and which we also assume to hold in the real Universe), the estimated velocity dispersion has to be corrected by a factor of $\sqrt{N/(N - 1)}$. In addition, both in the 2dFGRS and in our MGRSs there is a redshift measurement error of $\sigma_{\text{err}} = 85\,\text{km}\,\text{s}^{-1}$, which needs to be removed in quadrature. Thus, the final velocity dispersion of group members

is given by

$$\sigma = \sqrt{\max\left(0, \frac{N\sigma_{\text{gap}}^2}{N - 1} - \sigma_{\text{err}}^2\right)} \tag{16}$$

(Eke et al. 2004a). Fig. 5 shows the velocity dispersions of groups obtained from the identified galaxies, $\sigma_{\text{group}}^{\text{los}}$, versus those obtained from the true members, $\sigma_{\text{true}}^{\text{los}}$. As in Fig. 4, only groups with three of more members are shown. Contrary to the luminosities, the scatter here is quite large and there is a clear trend that $\sigma_{\text{group}}^{\text{los}} < \sigma_{\text{true}}^{\text{los}}$. This discrepancy is a result of the fact that galaxies with the highest peculiar velocities in a group are the most likely to be missed by the group finder as a result of the incompleteness. We can reduce the fraction of missed members by lowering the background level B, but at the expense of a significant increase in the fraction of interlopers. This indicates that it is in general quite difficult to accurately estimate the depth of the gravitational potential well of a dark matter halo from the velocities of its member galaxies identified in redshift space.

The above tests are based on the properties of selected groups, but they do not address whether or not the group finder can find all the existing groups. To check the completeness of the group identification, we estimate the group luminosity function, which gives the number density of groups as a function of group luminosity. Fig. 6 shows the group luminosity functions estimated from selected groups with
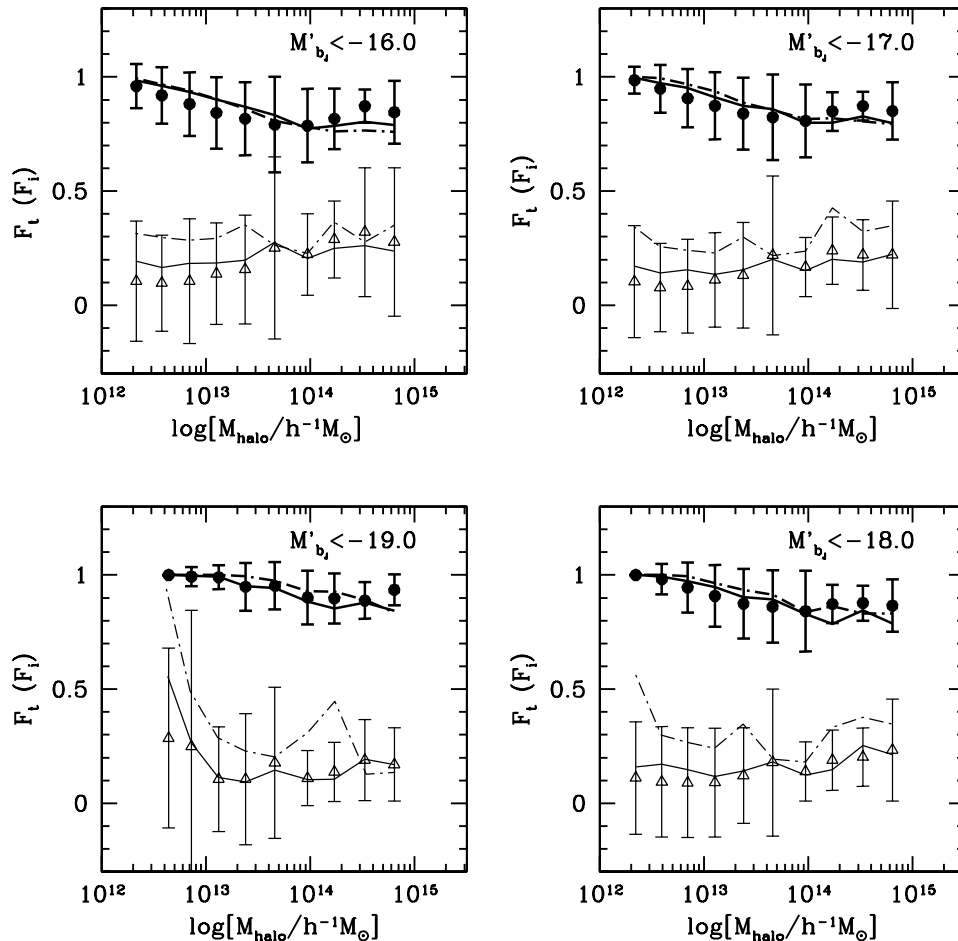
Figure 3. Same as in Fig. 2, but here we compare different group finder methods. The thick (thin) lines and symbols show the completeness (contamination) of the groups. The symbols with error bars are the same results as in Fig. 2. The solid lines show the results in which we assume a constant mass-to-light ratio $M/L = 400$ ($h\,\mathrm{M}_\odot/L_\odot$). The dot-dashed lines show the results where we use the traditional FOF group finder method with the same linking lengths as in Eke et al. (2004a).

at least one (thick lines) or at least three (thin lines) members. The histogram indicates the luminosity function of all true groups (i.e. of all dark matter haloes) and is shown for comparison. If we include all systems selected by our groups finder (single galaxies, binary systems and groups with more than two members), the group catalogue is roughly complete down to a luminosity of approximately $10^{10.5}\,h^{-2}\,\mathrm{L}_\odot$ (corresponding to a typical halo mass of $\sim 10^{12.5}\,h^{-1}\,\mathrm{M}_\odot$). If, on the other hand, we exclude single galaxies and binary systems, the sample is only more or less complete down to a group luminosity of approximately $10^{10.8}\,h^{-2}\,\mathrm{L}_\odot$. Finally, we tested the impact of the assumed mass-to-light ratio on the completeness of the group finder by using a constant $M/L = 400\,h\,\mathrm{M}_\odot/\mathrm{L}_\odot$. This leads to a group luminosity function that differs only very slightly from the fiducial case (dashed lines in Fig. 6), indicating once again that our group finder is insensitive to the assumptions regarding the mass-to-light ratios.

3.3 Flux limited samples

Having tested our group finder on volume limited mock samples, we now turn to more realistic, flux-limited samples. To that extent, we use our MGRS, the construction of which is described in Section 3.1 and, in more detail, in Yang et al. (2004).

Applying the group finder to these MGRSs yields the completeness and contamination shown in the left-hand panel of Fig. 7 (symbols with error bars). The results are very similar to those for our volume limited mock samples shown in Fig. 2, indicating that our group finder works almost equally well for flux limited samples as for volume limited samples. For comparison, the dashed lines indicate the completeness and contamination obtained using the standard FOF group finder of Eke et al. (2004a), with the linking lengths given by equations (12) and (13). Clearly, this group finder yields significantly more interlopers than our method, especially in small groups. The right-hand panel of Fig. 7 compares the estimated group luminosity (obtained using our fiducial group finder) with the true group luminosity. Again the results are very similar to those obtained using volume limited samples (cf. Fig. 4) and indicate that group luminosities can be obtained fairly robustly. The scatter in $\log L$ is approximately 0.1 in both directions, corresponding to an error of $\sim$30 per cent when the luminosity of a selected group is used to infer the true luminosity.

As a final demonstration of the performance of our group finder, Fig. 8 plots the number of groups identified as a function of redshift. In the left-hand panel, we plot the redshift distribution of all identified groups (including those with only one member) with luminosity $L_{18} > 10^{10.5}\,h^{-2}\,\mathrm{L}_\odot$ (where L_{18} is a scaled group luminosity, which
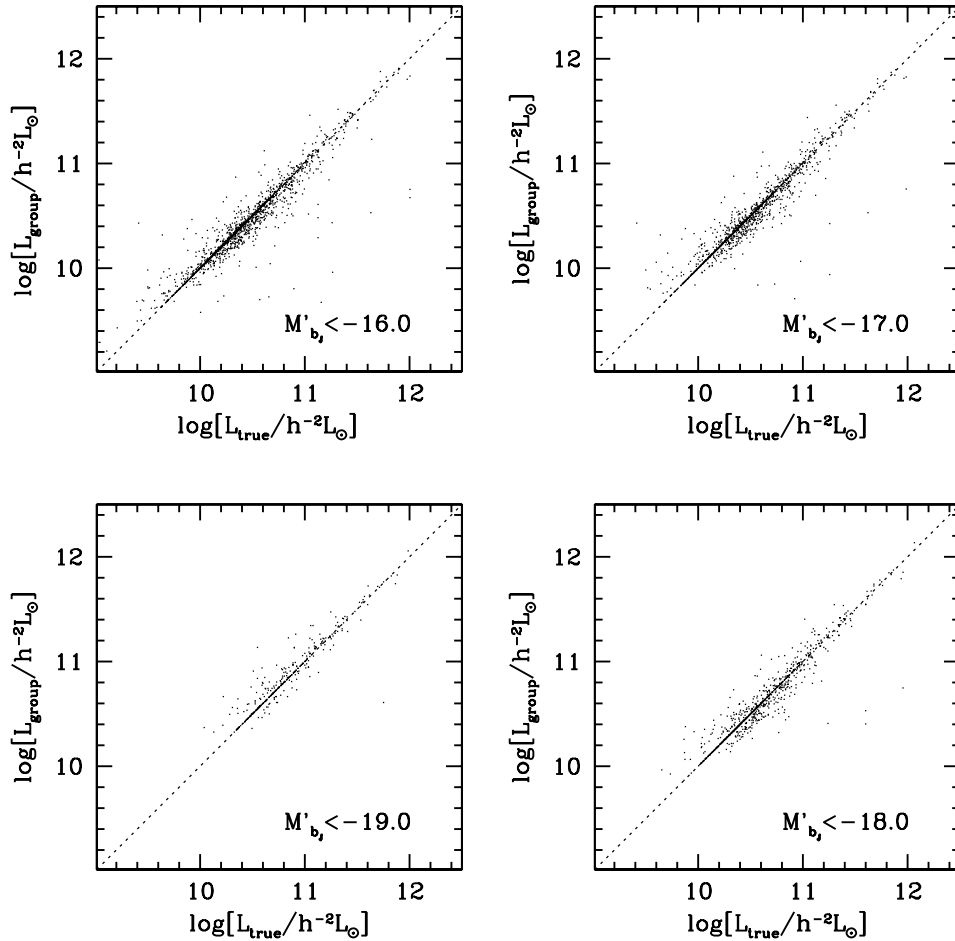
Figure 4. Comparison of the group luminosity L_{group} (the total luminosity of all identified group members) with the true group luminosity L_{true} (the total luminosity of all true group members). Only results for groups with more than two members are shown. The four different panels correspond to different absolute magnitude cuts (as indicated), where $M'_{b_J} = M_{b_J} - 5 \log h$.

will be defined later). Comparing these result with the solid line, which corresponds to a constant number density, indicates that the group finder works remarkably uniformly over the entire redshift range probed: the group completeness is virtually independent of redshift. This success is partly a result of the fact that our group finder can also identify systems that contain only one or two galaxies. If we impose a richness threshold, then the number of groups will decline with redshift at high z, as shown in the right-hand panel of Fig. 8. However, the change with redshift of the number of the groups selected by our group finder matches that of the true groups remarkably well. In contrast, the FOF method of Eke et al. (2004a) yields far too many groups. For example, for groups with $L_{18} > 10^{10.5} \, h^{-2} \, L_{\odot}$, our group finder selects 7091 groups with richness $N \geqslant 2$, close to the true number of 7040, while the FOF method selects 9665 groups. For $N \geqslant 3$, the true number of groups is 5231, while the number of groups selected are 4889 by our method and 6786 by the FOF method; for $N \geqslant 4$, the corresponding numbers are 3820, 3567 and 4969.

Although the completeness of the group identification is virtually redshift independent, the individual completeness of the identified groups does depend on redshift. Because in an apparent magnitude limited sample the mean number density of galaxies decreases with increasing redshift, groups associated with similar haloes will be richer at lower redshift. Caution is therefore required when using

groups selected from an apparent magnitude limited sample to study the intrinsic group properties. For example, if we consider groups with similar richness, we may in fact mix systems with different masses. One can circumvent this problem by using volume limited samples to identify groups. This has the advantage that the identified groups are uniform in redshift, but the disadvantage that it only makes limited use of the full observational data set.

An alternative method, which we adopt in this paper, is to use the full apparent magnitude limited sample to identify galaxy groups. In this case, however, before comparing groups at different redshifts, one needs to bring their intrinsic properties as determined from the detected member galaxies to a common scale. Here, we focus on the group luminosity and investigate how to scale L_{group}. One possibility is to compute L_{total} using equation (4). In fact, many earlier analyses have used this approach to calculate the total luminosity or richness of the groups (Tucker et al. 2000; Merchán & Zandivarez 2002; Kochanek et al. 2003; Eke et al. 2004b). This method is based on the assumption that the galaxy luminosity function in groups is similar to that of field galaxies. However, as shown in our earlier analyses, the galaxy luminosity function in different haloes (Paper I; Paper II) and different environments (Mo et al. 2004) may be very different. Therefore, it is not reliable to use equation (4) to estimate L_{total} from L_{group}. Here, we suggest a more empirical approach. A nearby group selected in an apparent magnitude limited survey should contain all
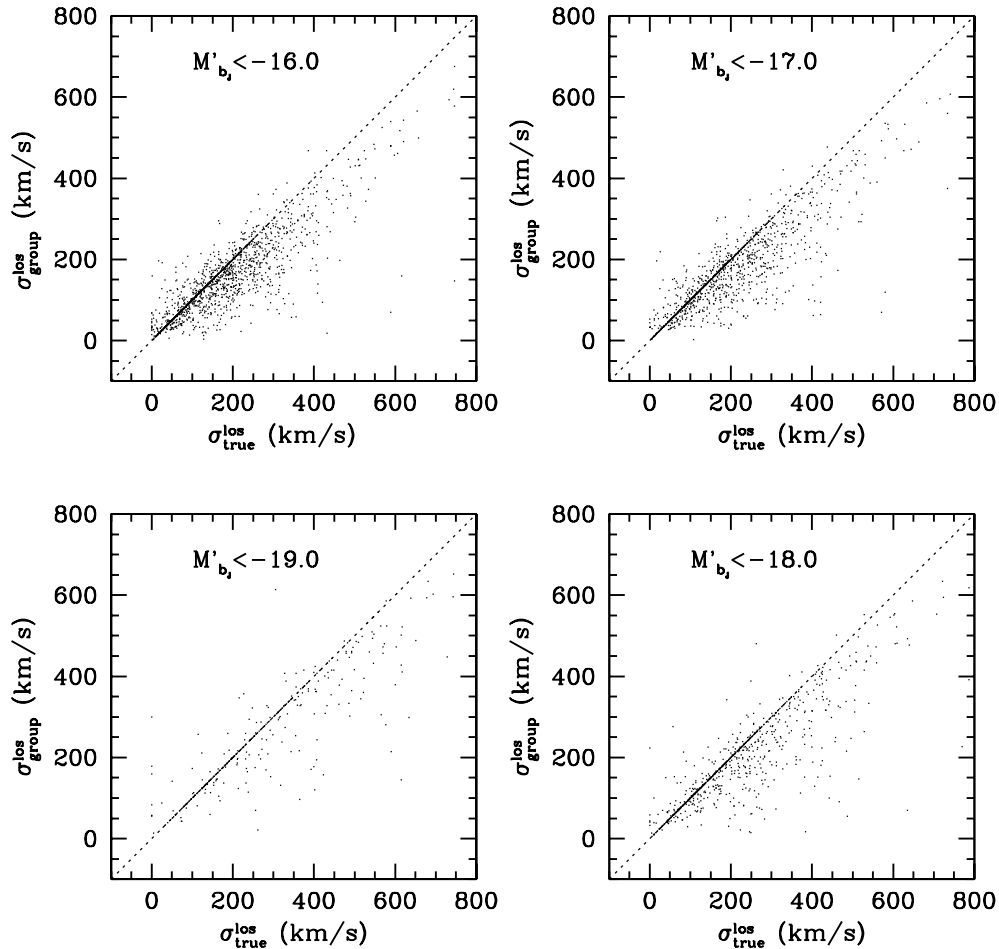
Figure 5. Same as Fig. 4, except that here we compare the line-of-sight velocity dispersion of identified group members with the line-of-sight velocity dispersion of true members. All velocity dispersions are estimated using the gapper estimator described in the text.

of its members down to a faint luminosity. We can therefore use these nearby groups to determine the relation between the group luminosity obtained using only galaxies above a bright luminosity limit and that obtained using galaxies above a fainter luminosity limit. Assuming that this relation is redshift-independent, one can correct the luminosity of a high-z group, where only the brightest members are observed, to an empirically normalized luminosity scale.

As a common luminosity scale, we use L_{18}, defined as the luminosity of all group members brighter than $M_{b_J} = -18 + 5 \log h$. To calibrate the relation between L_{group} and L_{18}, we first select all groups with $z \leqslant 0.09$, which corresponds to the redshift for which a galaxy with $M_{b_J} = -18 + 5 \log h$ has an apparent magnitude that is equal to the mean limit of the 2dFGRS ($b_J \leqslant 19.3$). Fig. 9 plots the ratio L_{18}/L_{group} obtained from these groups, where L_{group} is now defined as the luminosity of all group members brighter than a given limit. Different symbols correspond to different luminosity limits, as indicated in the right-hand panel. The left- and right-hand panels correspond to our MGRS and the 2dFGRS, respectively. It is clear that, for a variety of luminosity limits, the relation is quite tight.

In what follows, we proceed with two different ways. For all selected groups with $z \leqslant 0.09$, we compute L_{18} directly from the selected members with $M_{b_J} \leqslant -18 + 5 \log h$. For groups at higher redshifts, we compute L_{group} and use the appropriate, average relation between L_{18} and L_{group} to estimate the former.

Fig. 10 plots L_{18} thus obtained from the groups in our MGRS as function of group halo mass. Clearly, L_{18} is quite tightly correlated with the halo mass in the model and so the value of L_{18} can be used to divide groups according to their halo masses. For comparison, we also plot the theoretical prediction of the model group L_{18}: halo mass relation. The excellent match between the model prediction and the mean value confirms the accuracy of our group L_{18} estimation.

The velocity dispersion of galaxies in a group is another mass indicator that can be estimated directly. As shown in Section 3.2, the velocity dispersion obtained directly from the identified members (true plus interlopers) may differ systematically from the true velocity dispersion (obtained from all true members). Moreover, in the 2dFGRS, the measurement error in redshift is typically approximately $85 \, \mathrm{km \, s^{-1}}$, which is quite big in comparison with the typical velocity dispersion of small groups. Therefore, the masses estimated from velocity dispersions are not expected to be accurate. In the upper two panels of Fig. 11, we show the relation between the line-of-sight velocity dispersion of a group (estimated using equation 16) and the mass of the host halo for systems that contain at least three members. The left-hand panel shows the result where all true member galaxies in a halo are used to estimate the velocity dispersion. Even in this case, the scatter is quite large, although there is no noticeable deviation of the mean from the expected relation (solid lines). The right-hand panel shows the result where all selected members (selected true members plus interlopers) are used.
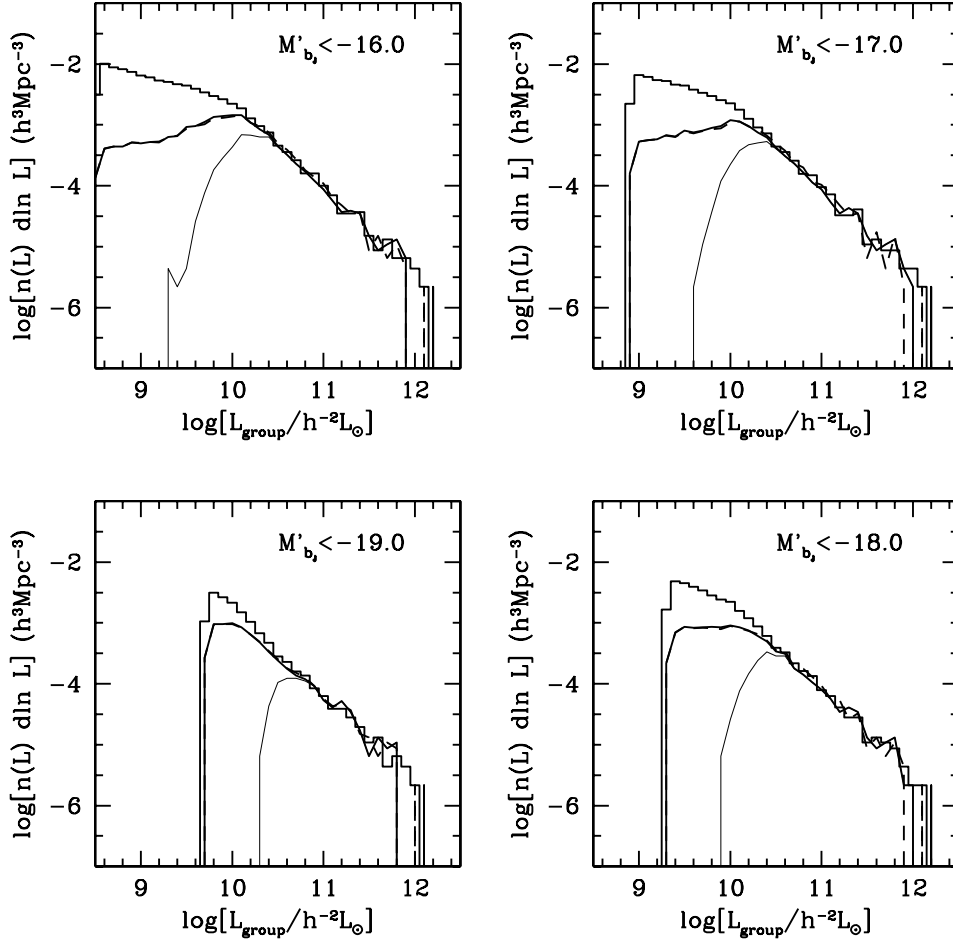
Figure 6. Group luminosity functions. Histograms indicate the true luminosity functions, in which the luminosity of each group is the sum of the luminosities of all true group members brighter than the luminosity cut indicated in each panel. The solid lines correspond to the luminosity functions for all the groups identified with our fiducial group finder, while the thin solid lines indicate the corresponding luminosity functions obtained when only using groups with at least three identified members. The dashed lines, which almost overlap with the thick solid lines, indicate the group luminosity functions obtained when using a constant mass-to-light ratio of $M/L = 400\,[h\,(\mathrm{M/L})_\odot]$ in the group finder.
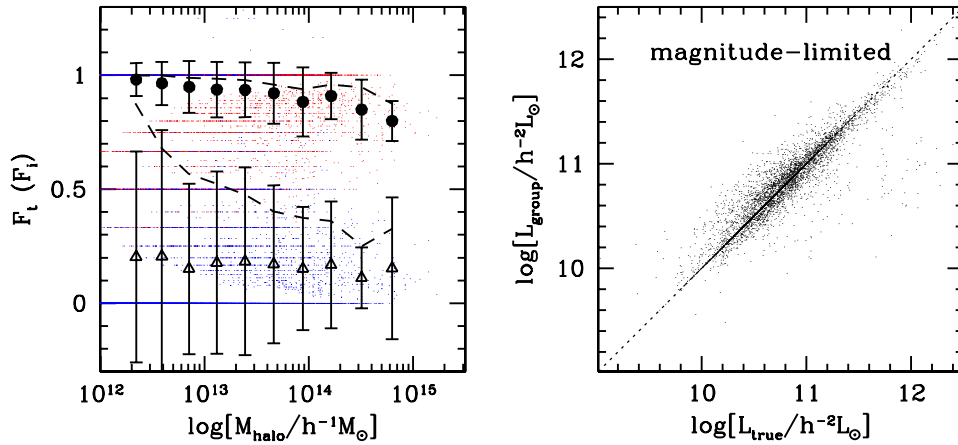
Figure 7. The left-hand panel shows the completeness (F_t) and contamination (F_i) of the groups selected from our MGRS. Symbols are as in Fig. 2, with the dashed lines indicating F_t and F_i obtained when using the traditional FOF method with the linking lengths of equations (12) and (13). Note how this method yields much larger interloper fractions than our group finder, especially in low-mass haloes. The right-hand panel plots the group luminosity, L_{group}, for groups with $N > 2$ obtained using our fiducial group finder, versus the true group luminosity L_{true} (cf. Fig. 4).
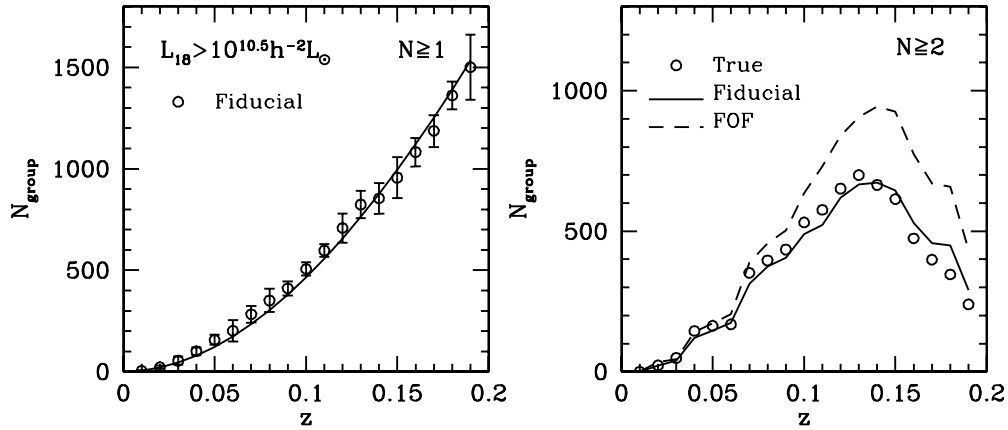
Figure 8. The number of groups detected in our MGRS as function of redshift. In the left-hand panel, we plot results for all groups with $N \geqslant 1$ and $L_{18} > 10^{10.5} \, h^{-2} \, L_\odot$, with the open circles indicating the mean number of groups identified using our fiducial group finder. The error bars indicate the 1-σ variance obtained from all eight MGRSs. The solid line indicates the expected relation for a constant group number density (i.e. for a group detection completeness that is redshift independent) and is shown for comparison. In the right-hand panel, we show the same $N_{\text{group}}(z)$ but this time only for groups with $N \geqslant 2$. The open circles indicate the true $N_{\text{group}}(z)$, with the solid and dashed lines indicating the numbers of groups detected using our fiducial group finder and the traditional FOF method, respectively.
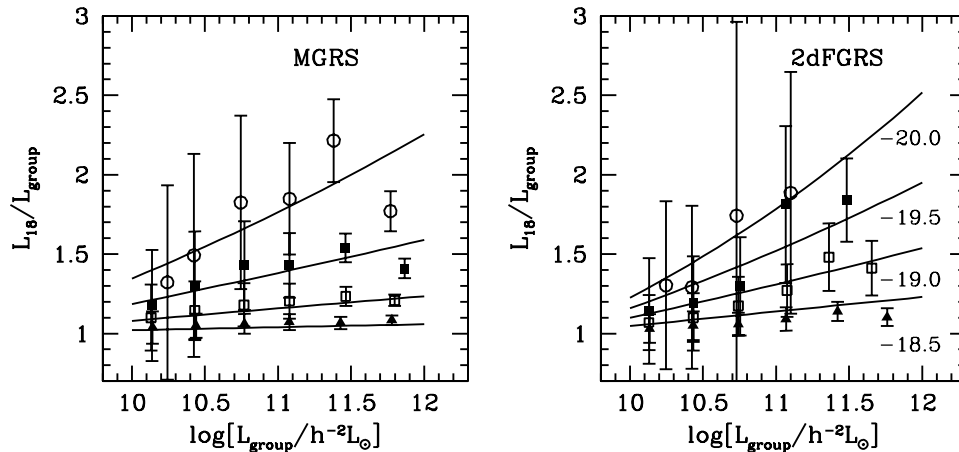

Figure 9. The ratio L_{18}/L_{group} as a function of L_{group}. Here, L_{18} is the total luminosity of all group galaxies brighter than $M_{b_J} - 5 \log h = -18.0$, while L_{group} is defined as the total luminosity of all group galaxies with $M_{b_J} - 5 \log h \leqslant -20.0$ (open circles), -19.5 (filled squares), -19.0 (open squares) and -18.5 (filled triangles). The error bars indicate 1-σ scatter of the ratios within different group luminosity bins, while the solid lines are our fits to these ratios, used to compute L_{18} from an observed L_{group} (see text for details). Results in the panels on the left and right correspond to groups identified in our MGRS and the 2dFGRS, respectively.

In this case, the presence of interlopers contaminates the mass–velocity dispersion relation and the velocity dispersion is slightly underestimated with respect to the expected value. In the lower panels of Fig. 11, we plot the same relations, but for groups with at least eight members. Here, the scatter in the relation is greatly reduced. In the case where all true members are used, the velocity dispersion of galaxies in a group is a good indicator of mass (see the lower left-hand panel), but bias still exists when using the selected group members (lower right-hand panel). This is a result of the fact that member galaxies with the highest velocities are the easiest to miss by the group finder.

4 APPLICATIONS TO THE 2DFGRS

Having calibrated and tested our group finder to both volume limited mock samples and to flux limited mock galaxy redshift surveys, we now apply it to the 2dFGRS. We use the final public data release, which contains approximately 250 000 galaxies with redshifts and is complete to an extinction-corrected apparent magnitude of $b_J \approx 19.45$. The survey covers an area of ~ 1500 deg^2 selected from the extended Automated Plate Measurement (APM) Survey (Maddox, Efstathiou & Sutherland 1996). The survey geometry consists of two separate declination strips in the North Galactic Pole (NGP) and the South Galactic Pole (SGP), respectively, together with 100 2-degree fields spread randomly in the southern Galactic hemisphere.

In what follows, we restrict ourselves only to galaxies with redshifts $0.01 \leqslant z \leqslant 0.20$ in the NGP and SGP subsamples with a redshift quality parameter $q \geqslant 3$ and a redshift completeness >0.8. This leaves a grand total of 151 280 galaxies with a typical rms redshift error of 85 km s^{-1} (Colless et al. 2001). Absolute magnitudes for galaxies in the 2dFGRS are computed using the type-dependent K-correction of Madgwick et al. (2002). For those galaxies without a reliable estimate of η, we adopt the average K-correction (see Madgwick et al. 2002 for details). We use exactly the same
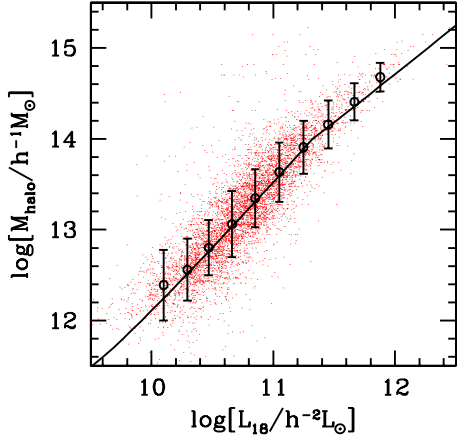
Figure 10. The luminosity L_{18} (see text for definition) of groups selected from the MGRS versus the associated halo mass. Open circles with error bars indicate the mean and 1-σ variance of the distribution of halo mass for groups with constant L_{18}, while the solid line corresponds to the prediction computed from the CLF.

restrictions when selecting galaxies from each of our eight MGRSs, yielding samples with $152\,000 \pm 4\,000$ mock galaxies.

4.1 Groups in the 2dFGRS

Applying our fiducial group finder to our 2dFGRS sample yields a group catalogue containing 78 708 systems. Among these are 7251 binaries, 2343 triplets and 2502 groups with four members or more. Using the FOF group finder, Eke et al. (2004a) identified 7020 groups with richness $N \geqslant 4$. Therefore, our group finder seems to give a much smaller number of $N \geqslant 4$ systems than does the FOF method adopted by Eke et al. There are two reasons for this discrepancy. First of all, as discussed in Section 3.3, the FOF method tends to select too many groups at high $z \gtrsim 0.1$. Secondly, our sample is approximately 25 per cent smaller than theirs, because we have imposed a more strict cut in sky coverage to get rid of regions where observational incompleteness is relatively high. When applying the FOF group finder adopted by Eke et al. to our 2dFGRS sample, we find 5749 groups with richness $N \geqslant 4$, which is consistent with the 25 per cent difference in sample size. Taking these two effects into account, the number of groups we obtain is roughly
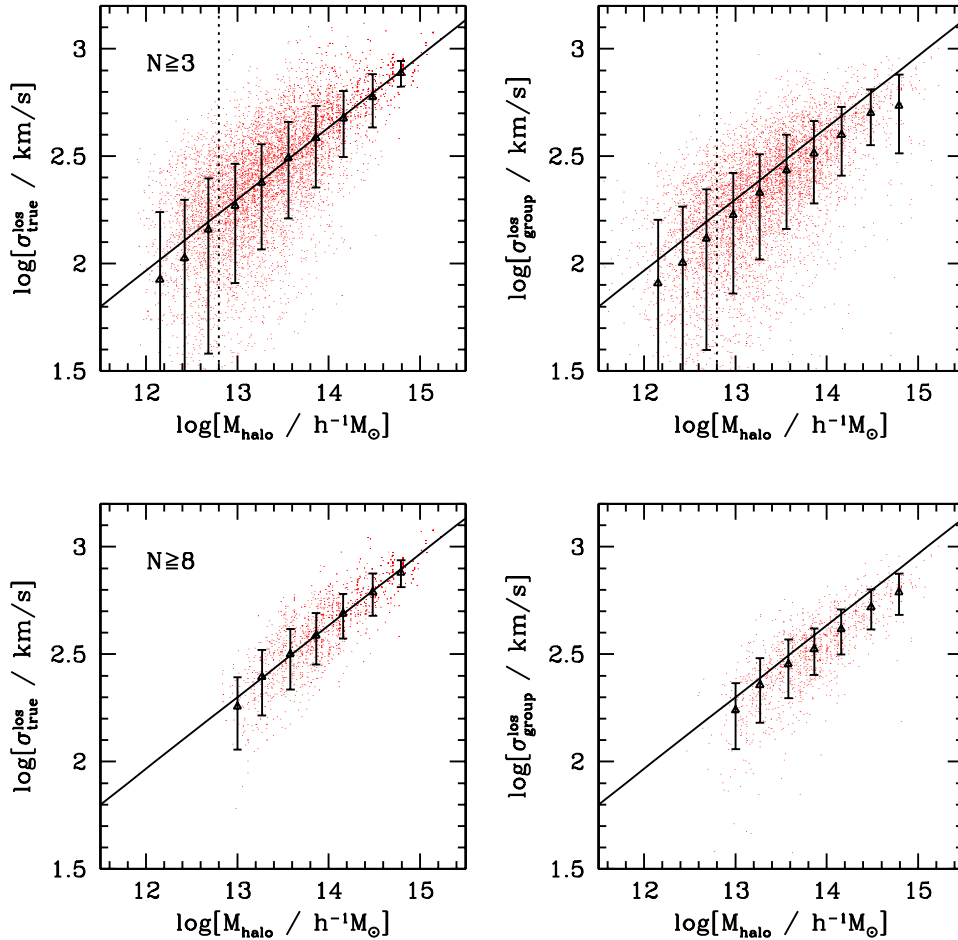

Figure 11. The left-hand panels show the relation between the line-of-sight velocity dispersion of (all) true members versus the halo mass in an apparent magnitude limited sample. The right-hand panels show that for groups with identified members (true plus interloper). The upper and lower panels correspond to results of groups with richness at least 3 and 8, respectively. The symbols and error bars are the mean and 1-σ scatter of the line-of-sight velocity dispersion in each mass bin. The solid line is the mass-velocity relation of the dark matter particles within haloes and $M \propto \sigma^3$. The dotted lines in the two upper panels roughly correspond to the completeness limit of the groups with at least three members.
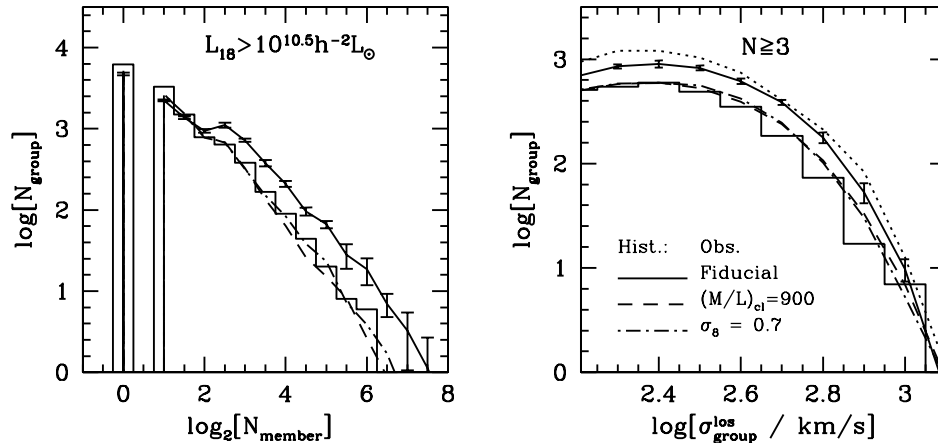
Figure 12. The richness (left-hand panel) and group line-of-sight velocity dispersions (right-hand panel) of the groups found in 2dFGRS observations and mock samples. The histograms are results of 2dFGRS and different lines are the results of different model mock samples. The error bars are 1-σ variances of eight mock samples. Obviously, there are more rich systems in the fiducial mock 2dFGRS samples than in the observations. As a comparison, the dotted line in the right-hand panel shows the velocity dispersion of dark matter particles within the host haloes.

comparable to that obtained by Eke et al. Of course, because we use a different group finder, the systems we select are expected to be different from theirs and so we do not expect to have an exact match in group number.

The histograms in Fig. 12 show the number of groups selected from the 2dFGRS as a function of richness (left-hand panel) and line-of-sight velocity dispersion (right-hand panel). In the left-hand panel, we only show results for groups with luminosity $L_{18} > 10^{10.5} h^{-2} L_\odot$, which approximately corresponds to the completeness limit of the group luminosity function (see Fig 6), while the right-hand panel only shows results for groups with $N \geqslant 3$. The various curves in both panels correspond to model predictions obtained from various mock catalogues, which will be discussed in the next subsection. As a comparison, we plot in the right-hand panel of Fig. 12 the velocity dispersion function (which is equivalent to the mass function) of the dark matter particles within the mock haloes. There is significant difference between the group velocity dispersion function and the halo velocity dispersion function, which is as a result of the bias already shown in Fig 11 and discussed in Section 3.3.

The histograms in Fig. 13 show the luminosity functions of the 2dFGRS groups. The upper left-hand panel is the result for the full magnitude limited sample. In this case, the observed group luminosities are corrected to L_{18} using the relations given in Fig. 9. The other three panels are the results for volume limited samples. Comparing the group luminosity functions (GLFs) for the different volume limited samples, we see that groups with more than three members are complete to a luminosity $L_{\mathrm{group}} \sim 10^{11} h^{-2} L_\odot$ for the volume limited sample with $M_{b_J} < -20.0 + 5 \log h$ and to a luminosity $L_{\mathrm{group}} \sim 10^{10.5} h^{-2} L_\odot$ for the volume limited sample with $M_{b_J} < -18.0 + 5 \log h$.

4.2 Comparison between model and observation

The results obtained above for the 2dFGRS groups can be compared with model predictions to constrain theories of structure formation. Because our MGRSs have been constructed to mimic the 2dFGRS in detail and because we have applied exactly the same selection criteria as for the 2dFGRS, we can make such a comparison in a straightforward way.

The solid curves in Figs 12 and 13 show the predictions obtained from our MGRSs. The error bars indicate the scatter from among our eight independent MGRSs and illustrate the expected scatter resulting from cosmic variance. There is a pronounced discrepancy between these MGRSs and the 2dFGRS: the model predicts too many rich systems that have high-velocity dispersions and high luminosities. This is consistent with Yang et al. (2004), who found that the same mock galaxy redshift surveys predict too high amplitudes for the pairwise peculiar velocity dispersion and for the real space two-point correlation function on small scales. As shown in Yang et al., these discrepancies can be attributed to the fact that this model predicts too many galaxies in massive clusters.

Before preceding to seek potential solutions to the problem, we first test whether the assumption of satellite galaxy distribution in individual haloes has a significant impact on our results. First, we consider a case in which satellite galaxies have a velocity bias relative to the dark matter. As in Yang et al. (2004), we assume the velocity bias parameter $b = \sigma_{\mathrm{gal}}/\sigma_{\mathrm{DM}} = 0.6$. As an example, we show the result of the group luminosity function given by this model as the thin solid line in the upper left-hand panel of Fig. 13. Secondly, we consider a case where the distribution of satellite galaxies in a halo is assumed to be shallower than that of the dark matter. Note that this assumption is in fact consistent with the result obtained by van den Bosch et al. (2004b). As a simple model, we assume the number density distribution of satellites in a halo to be

$$n_{\mathrm{sat}}(r) \propto \left(\frac{r}{\mathcal{R} r_s}\right)^{-\alpha} \left(1 + \frac{r}{\mathcal{R} r_s}\right)^{\alpha-3}, \qquad (17)$$

with $\alpha = 0.0$ and $\mathcal{R} = 2.0$. The group luminosity function predicted by this model is plotted as the thin dotted line in the upper left-hand panel of Fig. 13. Thirdly, we consider a case where we introduce 20 per cent scatter around the mean M/L predicted by the CLF when populating dark matter haloes with galaxies. Note that even without this scatter, there is scatter in the M/L among haloes of the same mass, because satellite galaxies are assigned to individual haloes in a stochastic way (see Yang et al. 2004). The group luminosity function predicted by the galaxies of this model is shown as the thin long-dashed line in the upper left-hand panel of Fig. 13. Obviously, all these changes have only negligible impact on our final results. We can therefore rule out the possibility that the discrepancy between
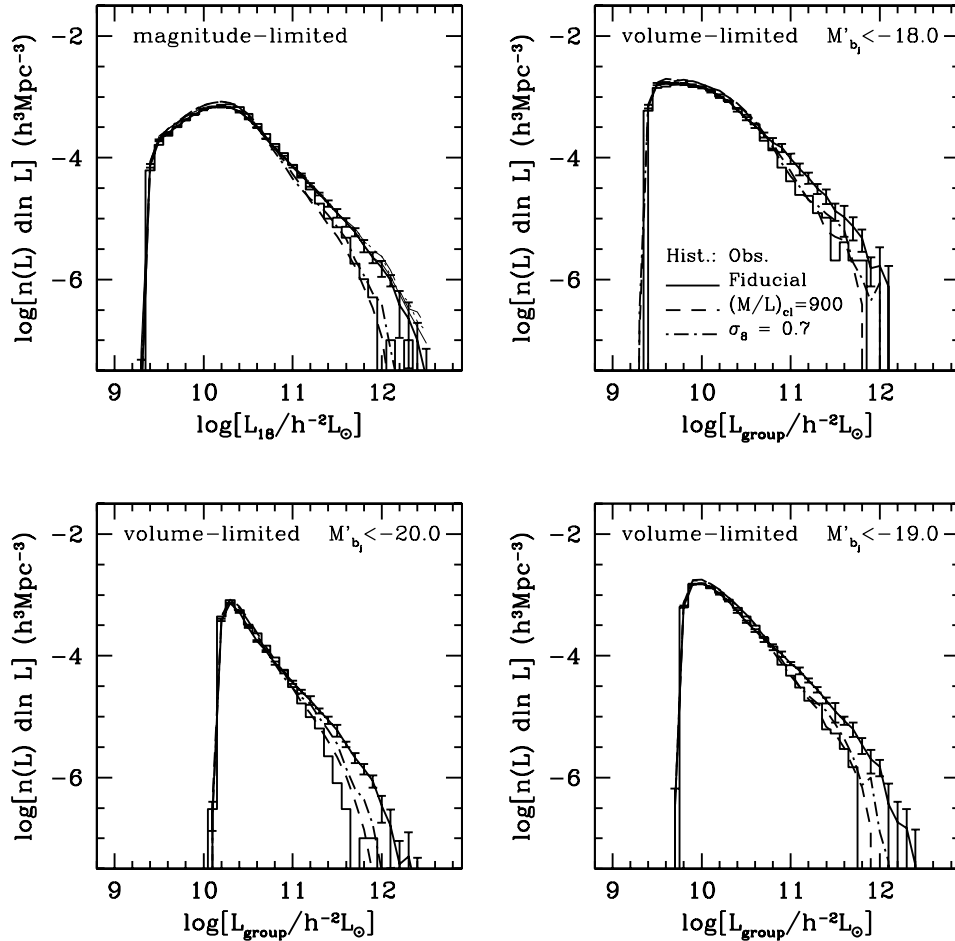
Figure 13. Similar to Fig. 6, but for 2dFGRS observations and mock samples. The histograms are the results of 2dFGRS observations and different lines are the results of different model mock samples. The error bars are 1-σ variances of eight mock samples. The upper left-hand panel shows the group luminosity function for magnitude limited 2dF samples. The upper right, lower right and lower left-hand panels are the group luminosity functions for volume limited 2dF samples with absolute magnitude cut, $M_{b_j} - 5 \log h < -18.0$, -19.0 and -20.0, respectively. Again, we find more bright (rich) groups in the fiducial mock 2dFGRS samples than in the observations. To test the impact of satellite galaxy distribution in individual haloes, we also plot in the upper left-hand panel the results for the MGRSs that assume velocity bias (thin solid line), spatial bias (thin dotted line) and scatter in the *M/L* (thin long-dashed line). Details about these tests are described in the text.

our fiducial model and observation is the result of our assumption about the distribution of satellite galaxies in individual haloes.

Next, we seek potential solutions to the discrepancy between our MGRSs and the 2dFGRS. The CLF used to construct these MGRSs was constrained to yield an average mass-to-light ratio for clusters of $(M/L)_{cl} = 500\,h\,(M/L)_\odot$. Either the true mass-to-light ratios of clusters are significantly higher (i.e. fewer galaxies per cluster), or there are fewer massive haloes, which implies a reduction of the power-spectrum normalization σ_8. Yang et al. (2004) were able to reproduce the clustering properties of 2dFGRS galaxies if either the cluster mass-to-light ratio is increased to approximately $(M/L)_{cl} = 900\,h\,(M/L)_\odot$, or σ_8 is lowered to approximately 0.7 (compared to 0.9 in the concordance cosmology adopted so far). The dashed and dot-dashed curves in Figs 12 and 13 show the group properties obtained using MGRSs with $(M/L)_{cl} = 900\,h\,(M/L)_\odot$ and $\sigma_8 = 0.7$,[3] respectively. Both very nicely match the multiplicity

function, the velocity dispersion function and the luminosity function of the 2dFGRS groups, thus strengthening the conclusions reached by Yang et al. (2004) that either clusters have high mass-to-light ratios, which seems incompatible with a number of independent observational constraints (e.g. Carlberg et al. 1996; Fukugita, Hogan & Peebles 1998; Bahcall et al. 2000), or that the power-spectrum normalization is significantly lower than typically assumed. As discussed in van den Bosch, Mo & Yang (2003b), this latter option cannot be ruled out by current observations.

5 CONCLUSIONS

The current CDM scenario, in which galaxies are assumed to form in dark matter haloes, is extremely successful in explaining a large range of observational data. Motivated by this success, and aided by our detailed knowledge of the abundances and properties of dark matter haloes within this CDM scenario, we have developed a group finder that can successfully group galaxies in redshift surveys according to their common haloes. Using detailed mock catalogues, constructed using large numerical simulations combined with the CLF of galaxies, we carefully tested the performance of this group

[3] As in Yang et al. (2004), the model results for $\sigma_8 = 0.7$ are not based on new simulations, but on the rescaling of the halo number density from the simulations of the $\sigma_8 = 0.9$ model.

finder. Individual groups selected using our group finder have an average completeness of approximately 90 per cent and with only ~20 per cent interlopers. The group luminosities agree with the true values to better than the 70 per cent level and the overall group luminosity function matches the real one well for groups with $L_{group} \gtrsim 10^{10.5} h^{-2} L_\odot$.

We have applied our group finder to the 2dFGRS and compared the properties of the 2dF groups with those extracted from detailed mock galaxy redshift surveys. Although the 2dF groups have similar properties as the mock groups, we find a clear discrepancy between mock and 2dF groups, in the sense that the model predicts too many rich systems. In order to match the observational results, we have to either increase the mass-to-light ratios of rich clusters to a level significantly higher than the typical observational value, or assume that $\sigma_8 \simeq 0.7$ compared with the concordance value of 0.9. This result is in perfect agreement with our previous findings based on the redshift-space clustering of galaxies (Yang et al. 2004) and enforces the conclusion that the concordance ΛCDM model may have too high clustering power on small scales.

The groups identified by our group finder are closely related to the underlying dark matter haloes. Given a uniform group catalogue constructed in this way, one can do many interesting things that can provide pivotal information about how galaxies form and evolve in CDM haloes. We plan to return to some of the problems in forthcoming papers.

ACKNOWLEDGMENT

REFERENCES

Bahcall N. A., Cen R., Davé R., Ostriker J. P., Yu Q., 2000, ApJ, 541, 1
Bahcall N. A. et al., 2003, ApJS, 148, 243
Beers T. C., Flynn K., Gebhardt K., 1990, AJ, 100, 32
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Berlind A. A. et al., 2003, ApJ, 593, 1
Brainerd T. G., Specian M. A., 2003, ApJ, 593, L7
Bryan G., Norman M., 1998, ApJ, 495, 80
Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559
Bullock J. S., Wechsler, R. H., Somerville R. S., 2002, MNRAS, 329, 246
Carlberg R. G., Yee H. K. C., Ellingson E., Abraham R., Gravel P., Morris S., Pritchet C. J., 1996, ApJ, 462, 32
Carlberg R. G. et al., 1997, ApJ, 485, L13
Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Colless M. et al., 2001, MNRAS, 328, 1039
Diaferio A., Kauffmann G., Colberg J. M., White S. D. M., 1999, MNRAS, 307, 537
Diemand J., Moore B., Stadel J., 2004, MNRAS, 352, 535
Eke V. R. et al., 2004a, MNRAS, 348, 866
Eke V. R. et al. (the 2dFGRS Team), 2004b, MNRAS, 355, 769
Fukugita M., Hogan C. J., Peebles P. J. E., 1998, ApJ, 503, 518
Geller M. J., Huchra J. P., 1983, ApJS, 52, 61
Jing Y. P., 2002, MNRAS, 335, L89

Jing Y. P., Suto Y., 2002, ApJ, 574, 538
Jing Y. P., Mo H. J., Börner G., 1998, ApJ, 494, 1
Jing Y. P., Börner G., Suto Y., 2002, ApJ, 564, 15
Kang X., Jing Y. P., Mo H. J., Börner G., 2002, MNRAS, 336, 892
Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19
Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188
Kepner J., Fan X., Bahcall N., Gunn J., Lupton R., Xu G., 1999, ApJ, 517, 78
Kim R. J. S. et al., 2002, AJ, 123, 20
Kochanek C. S., White M., Huchra J., Macri L., Jarrett T. H., Schneider S. E., Mader J., 2003, ApJ, 585, 161
Lee B. C. et al., 2004, AJ, 127, 1811
Lin Y.-T., Mohr J. J., Stanford S. A., 2004, ApJ, 610, 745
McKay T. A. et al., 2002, ApJ, 571, L85
Maddox S. J., Efstathiou G., Sutherland W. J., 1996, MNRAS, 283, 1227
Madgwick D. S. et al., 2002, MNRAS, 333, 133
Magliocchetti M., Porciani C., 2003, MNRAS, 346, 186
Marinoni C., Hudson M. J., 2002, ApJ, 569, 101
Merchán M., Zandivarez A., 2002, MNRAS, 335, 216
Mo H. J., Yang X., van den Bosch, F. C., Jing Y. P., 2004, MNRAS, 349, 205
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Norberg P. et al., 2002, MNRAS, 336, 907
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Pearce F. R., Thomas P. A., Couchman H. M. P., Edge A. C., 2000, MNRAS, 317, 1029
Postman M., Lubin L.M., Gunn J. E., Oke J. B., Hoessel J. G., Schneider D. P., Christensen J. A., 1996, AJ, 111, 615
Prada F. et al., 2003, ApJ, 598, 260
Rines K., Geller M. J., Diaferio A., Kurtz M. J., Jarrett T. H., 2004, ApJ, 128, 3
Seljak U., 2000, MNRAS, 318, 203
Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
Scranton R., 2002, MNRAS, 332, 697
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Spergel D. N. et al., 2003, ApJS, 148, 175
Tucker D. L. et al., 2000, ApJS, 130, 237
van den Bosch F. C., Yang X., Mo H. J., 2003a, MNRAS, 340, 771 (Paper II)
van den Bosch F. C., Mo H. J., Yang X., 2003b, MNRAS, 345, 923
van den Bosch F. C., Norberg P., Mo H. J., Yang X., 2004a, MNRAS, 352, 1302
van den Bosch F. C., Yang X., Mo H. J., Norberg P., 2004b, MNRAS, in press (doi:10.1111/j.1365-2966.2004.08407.x) (astro-ph/0406246)
van der Marel R. P., Magorrian J., Carlberg R. G., Yee H. K. C., Ellingson E., 2000, AJ, 119, 2038
Wang Y. Yang X., Mo H. J., van den Bosch F. C., Chu Y. Q., 2004, MNRAS, 353, 287
White M., 2001, MNRAS, 321, 1
White M., Kochanek C. S., 2002, ApJ, 574, 24
Yan R., Madgwick D. S., White M., 2003, ApJ, 598, 848
Yan R., White M., Coil A. L., 2004, ApJ, 607, 739
Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057 (Paper I)
Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., Chu Y. Q., 2004, MNRAS, 350, 1153
Yoshikawa K., Jing Y. P., Börner G., 2003, ApJ, 590, 654
Zheng Z., Tinker J. L., Weinberg D. H., Berlind A. A., 2002, ApJ, 575, 617

This paper has been typeset from a TeX/LaTeX file prepared by the author.