Mon. Not. R. Astron. Soc. **368,** 21–36 (2006) doi:10.1111/j.1365-2966.2006.10066.x

The dependence of clustering on galaxy properties

Cheng Li,[1,2,3]★ Guinevere Kauffmann,[3] Y. P. Jing,[1] Simon D. M. White,[3] Gerhard Börner[3] and F. Z. Cheng[2]

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

Accepted 2006 January 9. Received 2005 December 13; in original form 2005 September 30

ABSTRACT

We use a sample of ∼200 000 galaxies drawn from the Sloan Digital Sky Survey (SDSS) with $0.01 < z < 0.3$ and $-23 < M_{0.1r} < -16$ to study how clustering depends on properties such as stellar mass (M_*), colour ($g-r$), 4000-Å break strength (D_{4000}), concentration index (C), and stellar surface mass density (μ_*). Our measurements of $w_{\rm p}(r_{\rm p})$ as a function of the r-band luminosity are in excellent agreement with the previous two-degree Field Galaxy Redshift Survey and SDSS analyses. We compute $w_{\rm p}(r_{\rm p})$ as a function of stellar mass and we find that more-massive galaxies cluster more strongly than less-massive galaxies, with the difference increasing above the characteristic stellar mass M^* of the Schechter mass function. We then divide our sample according to colour, 4000-Å break strength, concentration and surface density. As expected, galaxies with redder colours, larger 4000-Å break strengths, higher concentrations and larger surface mass densities cluster more strongly. The clustering differences are largest on small scales and for low-mass galaxies. At fixed stellar mass, the dependences of clustering on colour and 4000-Å break strength are similar. Different results are obtained when galaxies are split by concentration or surface density. The dependence of $w_{\rm p}(r_{\rm p})$ on $g-r$ and D_{4000} extends out to physical scales that are significantly larger than those of individual dark matter haloes ($>5\,h^{-1}$ Mpc). This large-scale clustering dependence is not seen for the parameters C or μ_*. On small scales ($<1\,h^{-1}$ Mpc), the amplitude of the correlation function is constant for 'young' galaxies with $1.1 < D_{4000} < 1.5$ and a steeply rising function of age for 'older' galaxies with $D_{4000} > 1.5$. In contrast, the dependence of the amplitude of $w_{\rm p}(r_{\rm p})$ on concentration on scales less than $1\,h^{-1}$ Mpc is strongest for disc-dominated galaxies with $C < 2.6$. This demonstrates that different processes are required to explain environmental trends in the structure and in the star formation history of galaxies.

Key words: galaxies: clusters: general – galaxies: distances and redshifts – cosmology: theory – dark matter – large-scale structure of Universe.

1 INTRODUCTION

Our understanding of the large-scale structure of the Universe has come primarily from the studies of redshift surveys of nearby galaxies. The two-point correlation function (2PCF) of galaxies has long served as the primary way of quantifying the clustering properties of galaxies in these surveys (e.g. Peebles 1980). As the fundamental lowest order statistic, the 2PCF is simple to compute and provides a full statistical description for the Gaussian fields. It can also be easily compared with the predictions of theoretical models. Such comparisons have led to the conclusion that the observations are not consistent with the predictions of the standard Λ cold dark matter (ΛCDM) 'concordance' model, unless there is a scale-dependent bias in the distribution of galaxies relative to the dark matter (Gross et al. 1998; Jenkins et al. 1998; Jing, Mo & Börner 1998).

Benson et al. (2000a) clarified how the dependence of the galaxy formation efficiency on halo mass could lead to just such a scale-dependent bias. On large scales, the bias in the galaxy distribution is related in a simple way to the bias in the distribution of dark haloes. On small scales, the amplitude and slope of the correlation function are determined by the interplay of a number of different effects, including the distribution of the number of galaxies that occupy a halo of given mass and the fact that the brightest galaxy in each halo is always located near the halo centre. These ideas have been further developed into the so-called 'halo occupation distribution' (HOD)

★E-mail: leech@ustc.edu.cn

approach by many different authors (e.g. Jing et al. 1998; Peacock & Smith 2000; Seljak 2000; Berlind & Weinberg 2002; Cooray & Sheth 2002; Yang, Mo & van den Bosch 2003).

The HOD approach enables one to understand why the correlation function of the L^* galaxies is close to a power law over nearly four orders of magnitude in amplitude in a flat, $\Omega_0 = 0.3$ CDM universe. However, an important corollary is that the clustering properties of galaxies ought to depend strongly on the galaxy colour, star formation rate and morphology, because the halo occupation distributions of galaxies are predicted to depend sensitively on these properties (see e.g. Kauffmann, Nusser & Steinmetz 1997; Kauffmann et al. 1999; Benson et al. 2000b).

The fact that the measured correlations of galaxies differ according to type has been known for almost three decades. Davis & Geller (1976) computed angular correlations for the galaxies in the Uppsala catalogue and showed that the elliptical–elliptical correlations were characterized by a power law with steeper slope than the spiral–spiral correlations. Dressler (1980) quantified this as a relation between galaxy type and local galaxy density, with an increasing elliptical and S0 population and a corresponding decrease in spirals in the densest environments.

The large redshift surveys assembled in recent years, for example, the two-degree Field Galaxy Redshift Survey and Sloan Digital Sky Survey (SDSS), have provided angular positions and redshifts for samples of hundreds of thousands of galaxies and have allowed the dependence of clustering on the galaxy properties to be studied with unprecedented accuracy. These studies have established that the clustering of galaxies in the local Universe depends on a variety of factors, including luminosity (Norberg et al. 2001; Zehavi et al. 2002, 2005), colour (Zehavi et al. 2002, 2005), concentration (Zehavi et al. 2002; Goto et al. 2003), and spectral type (Norberg et al. 2002; Budavári et al. 2003; Madgwick et al. 2003). These studies have revealed that the galaxies with red colours, bulge-dominated morphologies and spectral types indicative of old stellar populations reside preferentially in the dense regions (Zehavi et al. 2005, and references therein, hereafter Z05). Furthermore, the luminous galaxies cluster more strongly than the less-luminous galaxies, with the luminosity dependence becoming more significant for the galaxies brighter than L^* [the characteristic luminosity of the Schechter (1976) function]. When galaxies are divided by colour, redder galaxies show a higher amplitude and steeper correlation function at all luminosities.

In order to interpret these clustering dependencies in the framework of galaxy formation models, it is useful to express the clustering results in terms of physical quantities, such as galaxy mass, size and mean stellar age, instead of more traditional quantities, such as luminosity or colour. Galaxy luminosity does not necessarily correlate very closely with stellar mass (the dominant baryonic component in all but the smallest galaxies). Both luminosity and colour are subjects to strong dependences on the fraction of young stars in the galaxy and on its dust content. These effects also complicate comparisons between the clustering of low- and high-redshift galaxies. It is now known that the star formation rates in galaxies evolve very strongly as a function of redshift. As a result, if one measures a change in clustering amplitude at fixed luminosity, it is not simple to ascertain which part of the effect is caused by the evolution in the stellar mass-to-light ratio (M_*/L) and which part by a change in the halo occupation distributions at higher redshift.

In this paper, we study the dependence of galaxy clustering on both luminosity and stellar mass, using a large sample of galaxies drawn from the SDSS. We then probe the dependence on other physical parameters, including colour ($g-r$), 4000-Å break strength (D_{4000}), concentration parameter (C) and stellar surface mass density (μ_*). The first two quantities, that is, $g-r$ and D_{4000}, are parameters associated with the recent star formation history of the galaxy (D_{4000} is expected to be less sensitive to the dust attenuation effects than colour), whereas the other two are related to galaxy *structure*. We first describe the observational samples used for the analysis. In Section 3, we outline our method of measuring the 2PCF from large redshift surveys. The results are described in Section 4 and summarized in the final section.

Throughout this paper, we assume a cosmological model with the density parameter $\Omega_0 = 0.3$ and the cosmological constant $\Lambda_0 = 0.7$. To avoid the $-5\log_{10} h$ factor, the Hubble constant $h = 1$, in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, is assumed throughout this paper when computing absolute magnitudes. In this paper, the quantities with a superscript asterisk are those at the characteristic luminosity/mass (e.g. characteristic luminosity L^*), whereas the quantities with a subscript asterisk refer to quantities associated with the stars in a galaxy (e.g. stellar mass M_*).

2 OBSERVATIONAL SAMPLES

2.1 NYU-VAGC

The SDSS is the most ambitious optical imaging and spectroscopic survey to date. The survey goals are to obtain the photometry of a quarter of the sky and spectra of nearly one million objects. Imaging is obtained in the *u, g, r, i, z* bands (Fukugita et al. 1996; Smith et al. 2002; Ivezić et al. 2004) with a special purpose drift scan camera (Gunn et al. 1998) mounted on the SDSS 2.5-m telescope at Apache Point Observatory. The imaging data are photometrically (Hogg et al. 2001) and astrometrically (Pier et al. 2003) calibrated, and used to select stars, galaxies, and quasars for the follow-up fibre spectroscopy. Spectroscopic fibres are assigned to objects on the sky, using an efficient tiling algorithm designed to optimize completeness (Blanton et al. 2003b). The details of the survey strategy can be found in (York et al. 2000) and an overview of the data pipelines and products is provided in the Early Data Release paper (Stoughton et al. 2002).

The large areal coverage and moderately deep survey limit (a mean redshift of $\sim$0.1 for the galaxies in the main spectroscopic sample) make the SDSS ideal for studying the large-scale structure and the characteristics of galaxy populations in the local Universe. The SDSS covers two regions on the sky, one in the northern Galactic cap (NGC) and another in the southern Galactic cap (SGC). In the SGC, three stripes are observed, one along the celestial equator and the other two north and south of the equator. The NGC lies mostly above the Galactic latitude 30°, but its footprint is adjusted slightly to lie within the minimum of the Galactic extinction contours (Schlegel, Finkbeiner & Davis 1998), resulting in an elliptical survey region (York et al. 2000). Currently, the survey in the NGC consists of two separate regions, one along the celestial equator (hereafter NGCE) and another off the equator (hereafter NGCO).

In this paper, we use the New York University Value Added Catalog (NYU-VAGC),[1] which is a catalogue of local galaxies (mostly below $z \approx 0.3$) constructed by Blanton et al. (2005a) based on the SDSS Data Release Two (DR2, Abazajian et al. 2004). Earlier proprietary versions of this catalogue have formed the basis of many SDSS investigations of the power spectrum, correlation function, and luminosity function of galaxies. The current version of the

[1] http://wassup.physics.nyu.edu/vagc/.

NYU-VAGC consists of 693 319 photometric objects (3514 deg^2); 343 568 of these have redshift determinations (2627 deg^2), with about 85 per cent completeness. This small subset of the full SDSS catalogue contains all of the information necessary for analysing the SDSS spectroscopic survey at the catalogue level. Compared with the catalogues distributed by the SDSS DR2 Archive Servers, the NYU-VAGC is photometrically calibrated in a more consistent way, reducing the systematic calibration errors across the sky from $\sim$2 to about $\sim$1 per cent. It is therefore more appropriate for the statistical studies of galaxy properties, galaxy clustering, and galaxy evolution. The NYU-VAGC is described in detail in Blanton et al. (2005a).

2.2 Physical quantities

The rich stellar absorption-line spectrum of a typical SDSS galaxy provides unique information about its stellar content and dynamics. Kauffmann et al. (2003a) presented a method for using this information to estimate the stellar masses of galaxies. The amplitude of the 4000-Å break [the narrow version of the index defined in Balogh et al. (1999)] and the strength of the Hδ absorption line (the Lick Hδ_A index of Worthey & Ottaviani 1997) were used as diagnostics of the stellar populations of the galaxies. Both the indices were corrected for the observed contributions of the emission lines in their bandpasses. From a library of 32 000 model star formation histories, the measured D_{4000} and Hδ_A indices were used to obtain a maximum-likelihood estimate of the z-band M_*/L for each galaxy. By comparing the colour predicted by the best-fitting model to the observed colour of the galaxy, the attenuation of the starlight due to dust could be estimated.

The SDSS imaging data provide the basic structural parameters that are used in this analysis. The z-band absolute magnitude, combined with the estimated values of M_*/L and dust attenuation A_z yield the stellar mass (M_*). The half-light radius in the z band and the stellar mass yield the effective stellar surface mass-density ($\mu_* = M_*/2\pi r^2_{50,z}$, in units of $h^2 \mathrm{M}_\odot\ \mathrm{kpc}^{-2}$). As a proxy for the Hubble type, we use the SDSS 'concentration' parameter C, which is defined as the ratio of the radii enclosing 90 and 50 per cent of the galaxy light in the r band (see Stoughton et al. 2002). Strateva et al. (2001) found that galaxies with $C > 2.6$ are mostly early-type galaxies, whereas spirals and irregulars have $2.0 < C < 2.6$.

The reader is referred to Kauffmann et al. (2003a) for a more detailed description of the methodology used to derive the stellar masses used in this paper. An analysis of how the physical properties of galaxies correlate with mass is presented in Kauffmann et al. (2003b). All the parameters used in this paper are available publically at http://www.mpa-garching.mpg.de/SDSS/ (see also Brinchmann et al. 2004).

2.3 Sample selection

In this paper, all the three regions in the NYU-VAGC, that is, NGCE, NGCO and SGC, are considered. Statistics are measured separately for the three regions but the results are always presented for the whole survey by combining the results in these regions.

We first select all the NYU-VAGC galaxies with the extinction-corrected Petrosian magnitude $14.5 < r < 17.77$. The bright limit is so chosen because the SDSS becomes incomplete for the bright galaxies with large angular size, whereas the faint limit corresponds to the magnitude limit of the Main galaxy sample in the SDSS. Further criteria for the galaxies to be included in our analysis are: (i) they are identified as galaxies from the Main sample (see Blanton et al. 2005a for a detailed description); and (ii) they lie within the redshift range $0.01 \leqslant z \leqslant 0.3$ and the absolute magnitude range $-23 < M_{0.1_r} < -16$. Here, $M_{0.1_r}$ is the r-band absolute magnitude corrected to its $z = 0.1$ value using the K-correction code (KCORRECT v3_1b) of Blanton et al. (2003a) and the luminosity evolution model of Blanton et al. (2003c). Our resulting sample includes a total of 196 238 galaxies.

The galaxies are then divided into a variety of different subsamples. We create 13 subsamples according to absolute magnitude, ranging from $M_{0.1_r} = -16$ to -23. Each sample includes galaxies in an absolute magnitude interval of 1 mag, with successive subsamples overlapping by 0.5 mag. Details are given in Table 1 (Samples L1–L13).

Similarly, the galaxies are divided into six subsamples according to $\log_{10} M_*$ (M1–M6 in Table 1). We do not consider galaxies with $\log_{10} M_* < 9$, because the volume of the survey over which such systems can be detected is extremely small. This is illustrated in Fig. 1, where we plot the maximum redshift out to which a galaxy of mass M_* with maximal M_*/L would be detected in the survey. This calculation assumes a 13-Gyr single-age stellar population and is based on the Bruzual & Charlot (2003) models. At stellar masses below $10^9\ \mathrm{M}_\odot$, the oldest galaxies are only visible at $z < 0.03$. For the galaxies with masses less than $10^8\ \mathrm{M}_\odot$, the maximum redshift is well below 0.02.

To compare our results to previous work, we have also constructed volume-limited subsamples (see Table 2), including subsamples that are volume limited in luminosity (Samples VL1–VL7) and in stellar mass (Samples VM1–VM5). The absolute magnitude ranges and redshift ranges used for selecting subsamples VL1–VL7 are the same as in Z05.

As will be described in Section 4.4, we further divide each luminosity and stellar mass subsample into red and blue, high- and low-D_{4000}, low- and high-concentration, low- and high-density subsamples by fitting the distributions of these parameters, using the bi-Gaussian functions. These subsamples are also listed in Table 1. It is also interesting to investigate how clustering varies as a function of colour/D_{4000}/concentration/surface density at fixed stellar mass. To this end, we select a sample of galaxies with stellar masses in the range of $10 < \log_{10} M_* < 11$, and divide the galaxies into subsamples according to their $g - r$ colours (Samples c1–c8), D_{4000} values (Samples D1–D12), concentrations (Samples C1–C10) and surface mass densities (Samples μ1–μ6). The details of these subsamples are given in Table 3.

3 CLUSTERING MEASURES

In this section, we outline our method for measuring the galaxy 2PCF for a flux-limited sample of galaxies. We begin by describing our methods for constructing random samples. We then describe how we correct for the effect of fibre collisions. Finally, we describe the 2PCF estimator and how measurement errors are calculated.

3.1 Constructing random samples

In order to use galaxy surveys in a statistically meaningful way, we need to have complete knowledge of their selection effects. A detailed account of the observational selection effects accompanies the NYU-VAGC release. The survey geometry is expressed as a set of disjoint convex spherical polygons, defined by a set of 'caps'. This methodology was developed by Andrew Hamilton to deal accurately and efficiently with the complex angular masks of galaxy surveys

Table 1. Flux-limited samples selected according to luminosity/stellar mass.

Sample	$M_{0.1_r}$	Number of galaxies SGC	NGCO	NGCE	Total	Percentage in subsamples[a] $g-r$ (per cent)	D_{4000} (per cent)	C (per cent)	$\log_{10}\mu_*$ (per cent)
L1	[−17.0, −16.0)	458	548	608	1614	27.2	41.7	90.7	48.0
L2	[−17.5, −16.5)	735	1261	1115	3111	26.2	32.2	82.9	43.4
L3	[−18.0, −17.0)	1257	2301	1695	5253	27.3	27.6	74.2	41.4
L4	[−18.5, −17.5)	2130	3808	2693	8631	31.0	26.3	67.0	44.2
L5	[−19.0, −18.0)	3657	6391	4366	14 414	36.6	29.3	60.5	49.3
L6	[−19.5, −18.5)	6532	10754	8582	25 868	43.9	36.4	55.9	56.3
L7	[−20.0, −19.0)	10 349	16 788	15 740	42 877	49.1	42.6	54.3	61.5
L8	[−20.5, −19.5)	14 804	24 688	22 879	62 371	51.9	46.9	54.9	63.7
L9	[−21.0, −20.0)	18 460	31 997	27 530	77 987	52.9	51.2	56.6	61.9
L10	[−21.5, −20.5)	17 717	31 010	25 376	74 103	53.4	56.4	58.9	54.7
L11	[−22.0, −21.0)	12 140	20 647	16 252	49 039	55.9	62.7	64.3	41.1
L12	[−22.5, −21.5)	5384	8895	6876	21 155	61.1	70.5	72.6	23.3
L13	[−23.0, −22.0)	1267	2097	1674	5038	64.8	78.0	76.3	7.20
Sample	$\log_{10} M_*$								
M1	[9.0, 9.5)	1686	3230	2325	7241	14.7	14.5	55.8	26.3
M2	[9.5, 10.0)	4086	6695	5237	16 018	23.8	19.8	44.3	36.5
M3	[10.0, 10.5)	9757	15 528	14 275	39 560	43.4	38.8	49.5	56.3
M4	[10.5, 11.0)	17 340	30 519	26 423	74 282	55.3	53.1	58.9	63.8
M5	[11.0, 11.5)	12 475	21 213	16 671	50 359	65.5	70.0	70.5	51.7
M6	[11.5, 12.0)	1183	2082	1603	4868	71.7	83.7	78.1	17.3

[a]Percentage of objects in subsample with larger value of physical quantities.

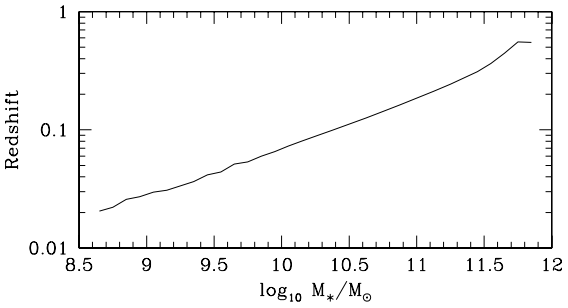

Figure 1. The results of a computation that we carried out to see out to what redshift a given mass in stars with maximum possible M_*/L would be detected. This assumes a 13-Gyr stellar population and the Bruzual & Charlot models.

(Hamilton & Tegmark 2002).[2] The advantage of using this method is that it is easy to determine whether a point is inside or outside a given polygon (Tegmark, Hamilton & Xu 2002). The redshift sampling completeness is then defined as the number of galaxies with redshifts divided by the total number of spectroscopic targets in the polygon. The completeness is thus a dimensionless number between 0 and 1, and it is constant within each of the polygons. The limiting magnitude in each polygon is also provided (it changes slightly across the survey region).

We have constructed separate random catalogues for each of the three regions of sky. These catalogues are designed to include all observational selection effects and are constructed as follows. First, we select a spatial volume that is sufficiently large to contain the survey sample. Then, we randomly distribute points within the volume and eliminate the points that are outside the survey boundary. Adopting the same magnitude limits as in the observational sample, we select random galaxies and we use the luminosity function derived by Blanton et al. (2003c) to assign to each of these galaxies an apparent and an absolute magnitude (appropriately, K and E corrected, see Section 2.3).

Since we will estimate the correlation function as a function of stellar mass, we also need to assign a mass to each point in the random sample. One way to do this is to use the observed relation between luminosity and stellar mass derived directly from our sample. The black lines in the top panel of Fig. 2 show contours of the number density of galaxies in the plane of absolute magnitude versus stellar mass. It can be seen from the histograms in the bottom panel of this figure that at fixed luminosity, the distribution of the stellar mass of galaxies is well described by a Gaussian, with the width of the Gaussian decreasing at higher luminosities.

We have divided the galaxies in our sample into 282 subsamples separated by 0.03 mag in $M_{0.1_r}$. The bin size was chosen so that each subsample contained at least 500 galaxies. The stellar mass distribution in each subsample is fitted with a Gaussian and the solid lines in the bottom panel of Fig. 2 show examples of these fits for several luminosity intervals. To test the quality of the fits, we randomly assign each galaxy a *new* stellar mass using the Gaussian fits. The red lines in the top panel of Fig. 2 show contours of the number density distribution that is predicted by this parametrization. The recovered distribution is a good match to the observations except in the region corresponding to the luminous galaxies with low M_*/Ls, where the method tends to overpredict the masses.

We now introduce a more general method, which should still be applicable even when the relation between galaxy luminosity and the physical property under investigation is not well fitted by a Gaussian and is subject to redshift-dependent selection biases.[3]

[2]http://casa.colorado.edu/∼ajsh/mangle/.

[3]One example of such a property would be the emission-line luminosity of a central active galactic nucleus (AGN). The line-detection limit is a strong function of redshift, because increasing contamination by light from the surrounding host galaxy makes extraction of weak lines more difficult for more distant AGN.

Table 2. Volume-limited samples.

Sample	$M_{0.1_r}$	z	Number of galaxies SGC	NGCO	NGCE	Total
VL1.....	[−18.0, −17.0)	(0.01, 0.03)	675	1011	1016	2702
VL2.....	[−19.0 −18.0)	(0.02, 0.04)	986	1776	1433	4195
VL3.....	[−20.0, −19.0)	(0.03, 0.07)	4510	7135	4514	16 159
VL4.....	[−21.0, −20.0)	(0.04, 0.07)	2202	3275	2076	7553
VL5.....	[−21.0, −20.0)	(0.04, 0.10)	6886	10 021	10 772	27 679
VL6.....	[−22.0, −21.0)	(0.07, 0.16)	5566	9446	8335	23 347
VL7.....	[−23.0, −22.0)	(0.10, 0.23)	705	1146	874	2725
Sample	$\log_{10} M_*$	z	SGC	NGCO	NGCE	Total
VM1.....	[9.0, 9.5)	(0.015, 0.045)	1195	2190	1624	5009
VM2.....	[9.5, 10.0)	(0.020, 0.075)	3368	5622	4020	13 010
VM3.....	[10.0, 10.5)	(0.025, 0.100)	8039	12 530	11 865	32 434
VM4.....	[10.5, 11.0)	(0.040, 0.140)	14 324	24 403	22 035	60 762
VM5.....	[11.0, 11.5)	(0.070, 0.200)	10 343	17 393	14 065	41 801

Table 3. Flux-limited samples selected according to physical quantities.

Sample	$g-r$	Number of galaxies SGC	NGCO	NGCE	Total
c1.....	[0.2, 0.5)	2348	3674	3383	9405
c2.....	[0.3, 0.6)	5798	9511	8411	23 720
c3.....	[0.4, 0.7)	9350	15 713	13 647	38 710
c4.....	[0.5, 0.8)	11 464	19 502	16 925	47891
c5.....	[0.6, 0.9)	14 206	23 717	21 363	59 286
c6.....	[0.7, 1.0)	16 557	28 504	25 476	70 537
c7.....	[0.8, 1.1)	13 188	22 679	20 229	56 096
c8.....	[0.9, 1.2)	6979	12 617	10 714	30 310
Sample	D_{4000}	SGC	NGCO	NGCE	Total
D1.....	[1.0, 1.3)	3807	6245	5309	15 361
D2.....	[1.1, 1.4)	7755	12 868	11 050	31 673
D3.....	[1.2, 1.5)	10 333	17 565	15 065	42 963
D4.....	[1.3, 1.6)	9900	17 010	14 561	41 471
D5.....	[1.4, 1.7)	8281	14 296	12 192	34 769
D6.....	[1.5, 1.8)	7762	13 298	11 527	32 587
D7.....	[1.6, 1.9)	9090	15 716	13 874	38 680
D8.....	[1.7, 2.0)	9902	17 143	15 622	42 667
D9.....	[1.8, 2.1)	8048	13 776	12 805	34 629
D10.....	[1.9, 2.2)	4274	7044	6916	18 234
D11.....	[2.0, 2.3)	1108	1654	1766	4528
Sample	C	SGC	NGCO	NGCE	Total
C1.....	[1.5, 2.1)	3372	5677	4783	13 832
C2.....	[1.7, 2.3)	7457	12 720	10 708	30 885
C3.....	[1.9, 2.5)	11 027	19 257	16 399	46 683
C4.....	[2.1, 2.7)	12 504	22 439	19 232	54 175
C5.....	[2.3, 2.9)	13 007	23 332	20 616	56 955
C6.....	[2.5, 3.1)	12 841	22 119	19 967	54 927
C7.....	[2.7, 3.3)	10 620	17 116	15 881	43 617
C8.....	[2.9, 3.5)	6533	9827	9227	25 587
C9.....	[3.1, 3.7)	2647	3752	3580	9979
Sample	$\log_{10}\mu_*$	SGC	NGCO	NGCE	Total
μ1.....	[8.00, 8.50)	2006	3415	2818	8239
μ2.....	[8.25, 8.75)	5430	9447	7989	22 866
μ3.....	[8.50, 9.00)	9960	17 419	15 139	42 518
μ4.....	[8.75, 9.25)	14 166	24 865	21 974	61 005
μ5.....	[9.00, 9.50)	13 593	23 073	20 834	57 500
μ6.....	[9.25, 9.75)	7015	10 908	10 044	27 967
μ7.....	[9.50, 10.0)	1495	2065	1819	5379

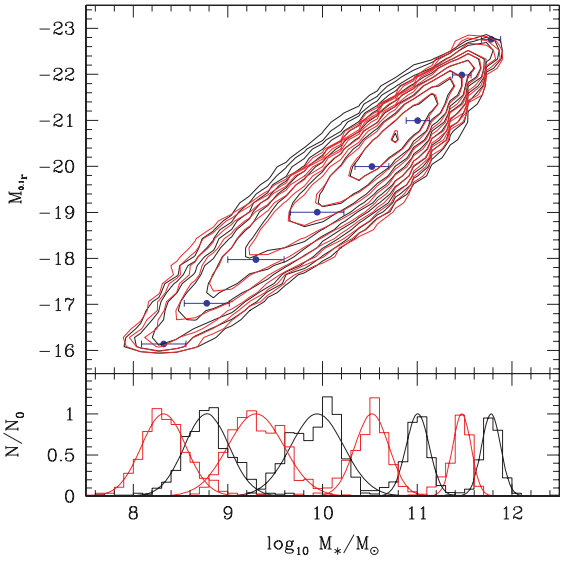

Figure 2. Shown in the upper panel are the contours of number density of galaxies in the plane of stellar mass versus luminosity. The black lines are for the data and the red are reconstructed using the Gaussian functions that best fit the stellar mass distribution in 282 different luminosity intervals. The contour levels are increased by factors of 2 from the lowest (15 [0.2 mag]$^{-1}$ [0.2 $\log_{10}$ M$_\odot$]$^{-1}$) to the highest (7680 [0.2 mag]$^{-1}$ [0.2 $\log_{10}$ M$_\odot$]$^{-1}$). The lower panel shows examples of the Gaussian distributions. Histograms show the data, and solid lines, the best fits. N_0 is the Gaussian height. The corresponding centres (blue points) and widths (error bars) of the Gaussians are shown in the upper panel.

Our method takes the observed sample and randomly re-assigns the position of each galaxy on the sky, while keeping the redshift, absolute magnitude, stellar mass, and any other physical quantities fixed. The spectroscopic incompleteness at each sky position is imposed for the random points as in the observed sample. To get a random catalogue as large as possible, we repeat the above procedure for 20 times using different random number seeds. In this way, all possible redshift-dependent selection biases are automatically taken into account, and it is only the sky position that is randomized. This method is valid only when the sample is a wide-angle survey and the variation of its limiting magnitudes is small across the survey region, both of which are valid in the SDSS. For very large-area

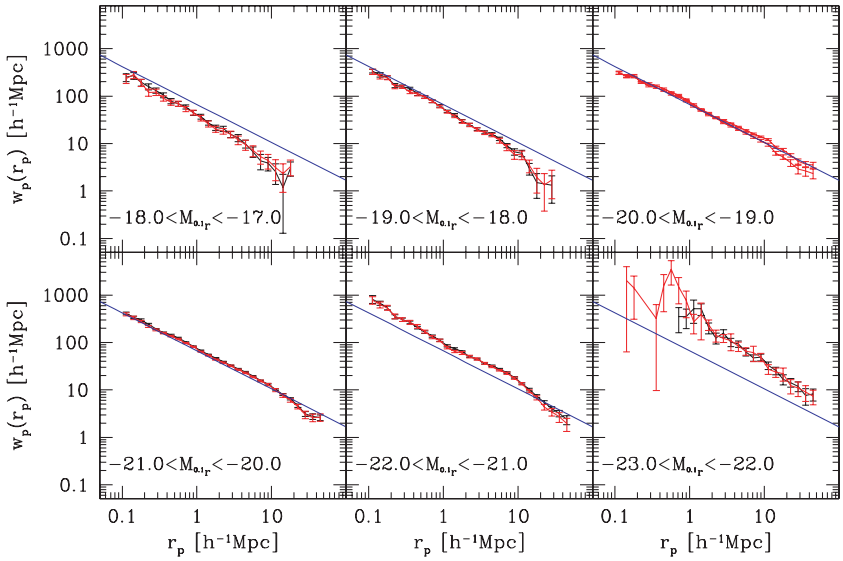

Figure 3. Projected 2PCF $w_p(r_p)$ in different luminosity intervals (Samples L3, L5, L7, L9, L11 and L13, in Table 1). When measuring 2PCFs, two methods are used to construct random samples (see Section 3.1). The black lines are for the standard method and the red lines are for the method in which the sky positions are randomized (see the text, for a detailed description). In each panel, the blue line is the line corresponding to $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-1.8}$.

surveys such as the SDSS, randomizing the sky positions should be sufficient to break the coherence of the large-scale structures in the survey. In the next section, we will use random catalogues constructed using both the methods and we will show that the measured projected correlation functions are in good agreement (see Fig. 3).

3.2 Volume corrections

When computing correlation functions as a function of stellar mass, it is important to note that at a given stellar mass M_*, galaxies with lower M_*/L will be detected out to higher redshifts. A mass-selected sample will thus be biased to galaxies with younger populations and this may lead to systematic errors when computing the correlation function at fixed M_*. In this paper, we correct for this M_*/L bias by computing a weighted correlation function: each galaxy pair is weighted by the inverse of the volume over which *both* the galaxies can be detected in the survey. This is similar to the $1/V_{\rm max}$ correction that one makes when computing a mass function or luminosity function. The same volume weighting must also be applied to the random catalogue. It is very simple to apply the same technique to the catalogues constructed by randomizing the sky positions, so this will be our method of choice when estimating correlations as a function of stellar mass.

In order to compute the volumes over which galaxies can be detected, we have computed $z_{\rm min}$ and $z_{\rm max}$ for each galaxy in the sample, where $z_{\rm min}$ is defined as the redshift where the galaxy has an r-band magnitude of 14.5 and $z_{\rm max}$ is the redshift where the galaxy has an r-band magnitude of 17.77. These are derived using the KCORRECT code of Blanton et al. (2003c).

3.3 Correction for fibre collisions

In the SDSS survey, two galaxies closer than 55 arcsec (corresponding to $\sim$100 h^{-1} kpc at the median redshift of our sample) cannot be assigned fibres simultaneously on one spectroscopic plate. If these fibre 'collisions' are not taken into account, the real-space (or projected) 2PCF will be systematically underestimated at small separations. In earlier work (e.g. Zehavi et al. 2002; Tegmark et al. 2004), a correction was made by simply assigning to each galaxy affected by a collision the same redshift as its nearest spectroscopically targeted neighbour on the sky. Zehavi et al. (2002) have performed extensive tests of this procedure and have shown that it works well for $r_p > 0.1\,h^{-1}$ Mpc. Tegmark et al. (2004) also found no evidence that fibre collisions are boosting their measured power spectrum on the smallest scales they probe ($k \sim 0.3\,h$ Mpc^{-1}).

Table 4. Coefficients for the formula of dividers in physical quantities.

Quantity	A	B
$g-r$	-0.788 ± 0.028	-0.078 ± 0.001
D_{4000}	-0.563 ± 0.038	-0.108 ± 0.002
C	-0.498 ± 0.193	-0.150 ± 0.009
$\log_{10}\mu_*$	3.738 ± 0.213	-0.256 ± 0.011

In this paper, we use a different method for correcting for fibre collisions. We measure the angular 2PCF both for the spectroscopic samples and for the parent photometric sample from which they were drawn; the effect of fibre collisions can then be estimated and corrected for by comparing the two correlation functions. A similar method has been used in 2dF clustering analyses by Hawkins et al. (2003).

Here we briefly summarize our method, which will be described in more detail in a separate paper (Li et al., in preparation). We calculate the angular 2PCF for the photometric sample ($w_p(\theta)$) and for the spectroscopic sample [$w_z(\theta)$]. The quantity

$$F(\theta) = \frac{w_z(\theta)+1}{w_p(\theta)+1}, \qquad (1)$$

can then be used to account for the effect of fibre collisions. For each data–data pair, we calculate the angular distance θ between the two members of the pair and weight this pair by $1/F(\theta)$ when estimating the pair counts. If this correction is not applied, both

Table 5. $w_{\rm p}(r_{\rm p})$ for the galaxies in different luminosity intervals and with different properties.

$r_{\rm p}$	All	$g-r$		D_{4000}		C		$\log_{10}\mu_*$	
(h^{-1}Mpc)		Red	Blue	Red	Blue	Red	Blue	Red	Blue
$-17.0 \leqslant M_{0.1_r} < -16.0$									
0.11	319.6/111.5	321.6/172.1	–	1075.4/439.2	184.8/153.1	1912.7/1330.0	94.7/114.3	710.8/377.0	200.2/147.6
0.14	221.0/62.9	235.5/79.2	–	800.2/319.6	127.8/104.8	1859.3/706.8	32.4/52.2	547.8/212.1	9.7/84.2
0.18	247.2/64.9	246.5/102.1	–	446.5/191.9	158.1/130.1	845.8/382.8	112.5/76.0	365.1/142.4	180.7/150.8
0.23	204.7/56.4	180.1/58.4	–	1430.1/734.4	105.2/60.6	1014.5/409.5	363.2/608.3	244.1/104.5	1461.9/1130.4
.....									

$w(\theta)$ and $w_{\rm p}(r_{\rm p})$ exhibit a strong 'rollover' in amplitude on small scales. Once the correction is applied, this feature disappears. In the rest of our analysis, we will always include the $1/F(\theta)$ weighting in the measurements of the correlation functions. Since the effect of fibre collisions is expected to be independent of galaxy property, we will not derive the correction function $F(\theta)$ for each individual galaxy sample, but choose to derive it from the whole sample and then apply it to our subsamples.

3.4 Estimator of the correlation function and errors

In this paper, the 2PCFs are measured in equal logarithmic bins of $r_{\rm p}$ and in equal linear bins of π, using the Hamilton (1993) estimator,

$$\xi(r_{\rm p},\pi) = \frac{4DD(r_{\rm p},\pi)RR(r_{\rm p},\pi)}{[DR(r_{\rm p},\pi)]^2} - 1. \quad (2)$$

Here $r_{\rm p}$ and π are the separations perpendicular and parallel to the line of sight; $DD(r_{\rm p},\pi)$ is the count of data–data pairs with perpendicular separations in the bins $\log_{10} r_{\rm p} \pm 0.5\Delta \log_{10} r_{\rm p}$ and with radial separations in the bins $\pi \pm 0.5\Delta\pi$; $RR(r_{\rm p},\pi)$ and $DR(r_{\rm p},\pi)$ are the counts of random–random and data–random pairs, respectively. The reason why we choose different bins for $r_{\rm p}$ and π is the fact that $\xi(r_{\rm p},\pi)$ decreases rapidly as a function of $r_{\rm p}$, but remains constant as a function of π on small scales. Following standard practice, we estimate the projected 2PCF $w_{\rm p}(r_{\rm p})$ by

$$w_{\rm p}(r_{\rm p}) = 2\int_0^\infty \xi(r_{\rm p},\pi){\rm d}\pi = 2\sum_i \xi(r_{\rm p},\pi_i)\Delta\pi_i. \quad (3)$$

Here the summation for computing $w_{\rm p}(r_{\rm p})$ runs from $\pi_1 = 0.5\,h^{-1}$ Mpc to $\pi_{40} = 39.5\,h^{-1}$ Mpc, with $\Delta\pi_i = 1\,h^{-1}$ Mpc. The projected correlation function $w_{\rm p}(r_{\rm p})$ is directly related to the real-space 2PCF $\xi(r)$ by a simple Abel transform of $\xi(r)$. Commonly, $w_{\rm p}(r_{\rm p})$ is modelled by a power law

$$w(r_{\rm p}) = Ar_{\rm p}^{1-\gamma}. \quad (4)$$

Then $\xi(r)$ is also a power law

$$\xi(r) = (r_0/r)^\gamma \quad (5)$$

with

$$r_0^\gamma = \frac{A\Gamma(\gamma/2)}{\Gamma(1/2)\Gamma[(\gamma-1)/2]}, \quad (6)$$

where $\Gamma(x)$ is the Gamma function. However, the parametrization of the correlation function using only r_0 and γ does not provide sufficient information to recover the full observational results, unless the correlation function is a pure power law on all scales. Our results (see below) show that this is not the case. The departures of $w_{\rm p}(r_{\rm p})$ from a pure power law have also been discussed in previous papers (e.g. Zehavi et al. 2004). We have thus chosen to present our results in terms of the measured *amplitude* of $w_{\rm p}(r_{\rm p})$ on different physical scales. We also tabulate the correlation functions so that our readers can recover them accurately. A detailed description of these tables (Tables 5 and 6) is given in Appendix A. The tables themselves are available in electronic form at http://www.mpa-garching.mpg.de/∼leech/papers/clustering/.

The errors on the clustering measurements are estimated using the bootstrap resampling technique (Barrow, Bhavsar & Sonoda 1984). We generate 100 bootstrap samples from the observations and compute the correlation functions for each sample using the weighting scheme (but not the approximate formula) given by Mo, Jing & Börner (1992). The errors are then given by the scatter of the measurements among these bootstrap samples. The tests in Jing et al. (1998) using mock samples showed that the bootstrap errors are comparable (within a factor of 2) to the scatter among different mock samples, thus proving that the error estimates are robust.

4 DEPENDENCE OF CLUSTERING ON GALAXY PROPERTIES

4.1 Luminosity

Fig. 3 shows the projected 2PCF $w_{\rm p}(r_{\rm p})$ in different luminosity intervals (Samples L3, L5, L7, L9, L11 and L13, in Table 1). The red and black lines on the figure compare the results obtained for the two different methods of constructing random samples described in Section 3.1. Black lines are for the 'standard' method in which the selection function is explicitly modelled. Red lines are for the method in which the sky positions of the observed galaxies are randomly re-assigned. The agreement between the two methods is very encouraging, suggesting that the latter method does work well for analysing large redshift surveys like the SDSS and can be applied in the case of more complicated selection by physical parameters with redshift-dependent biases.

To guide the eye, we have plotted the relation $\xi(r) = (r/5\,h^{-1}\ {\rm Mpc})^{-1.8}$ in blue in every panel in Fig. 3. In general, we see that the amplitude of the correlation function increases with luminosity, but the strength of this effect is different on different scales. For the galaxies fainter than L^* ($M_{0.1_r} = -20.44$), the clustering amplitude stays nearly constant on very small scales ($r_{\rm p} \sim 0.1\,h^{-1}$ Mpc), but on larger scales there is a much stronger luminosity dependence. For bright galaxies, the correlation amplitude increases strongly with luminosity at all scales. It is also interesting that the *slope* of the correlation function gets flatter with increasing luminosity for the galaxies fainter than L^*, but then increases for the galaxies brighter than L^*. In another word, L^* galaxies exhibit the flattest correlation functions.

These trends are illustrated more clearly in Fig. 4, where we plot the amplitude of the projected correlation function $w_{\rm p}(r_{\rm p})$ as

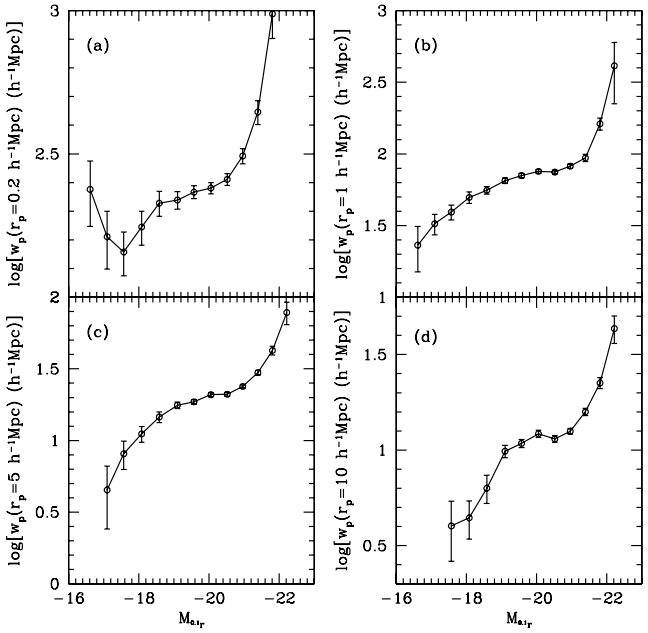

Figure 4. Amplitude of the projected 2PCF $w_{\rm p}(r_{\rm p})$ as a function of luminosity (Samples L1–L13) at $r_{\rm p} = 0.2$, 1, 5 and 10 h^{-1} Mpc.

a function of luminosity at $r_{\rm p} = 0.2$, 1, 5 and 10 h^{-1} Mpc. At $r_{\rm p} = 0.2\,h^{-1}$ Mpc, the correlation function probes galaxy pairs that reside within a common dark matter halo. At $r_{\rm p} = 10\,h^{-1}$ Mpc, the correlation function should only be sensitive to pairs of galaxies in separate haloes. This figure confirms that the luminous galaxies cluster more strongly than the faint galaxies, with the difference becoming more marked above L^*. However, the luminosity dependence of galaxy clustering is different on different scales. On small scales, the clustering amplitude does not vary with luminosity for the galaxies fainter than L^*, but increases steeply for the galaxies brighter than L^*. In contrast, the amplitude on large scales rises more continuously as a function of luminosity. It is also interesting that the dependence of $w_{\rm p}(r_{\rm p})$ on luminosity appears to change slope at $M_{0.1r} \sim -20$. One possible reason for this switch in behaviour is that a significant fraction of the faint galaxies are 'satellite' systems orbiting within a common dark matter halo, whereas bright galaxies are mainly 'central' galaxies located at the centres of their dark matter haloes. We intend to explore this in more detail in future work.

To compare our results to previous studies, we have also computed correlation functions using samples that are volume limited in luminosity (Samples VL1–VL7). The results are shown in Fig. 5. Black lines show the correlation functions for Samples VL1–VL7. For comparison, the measurements provided by Z05 are shown in red and the correlation functions computed from the corresponding magnitude-limited subsamples are shown in green. The agreement between the magnitude-limited analysis and the volume-limited one indicates that our results are robust and reliable. Furthermore, it can be seen that our measurements are in good agreement with those carried out by Z05, although there are some small differences. These are probably due to the different 2PCF estimators or the different methods of constructing random samples. We note that the magnitude-limited sample of galaxies with $-20 < M_{0.1r} < -19$ (Sample L7) and the magnitude-limited and volume-limited samples with $-21 < M_{0.1r} < -20$ (Sample L9 and VL5) all exhibit anomalously high $w_{\rm p}(r_{\rm p})$ values at large separations ($r_{\rm p} \gtrsim 5\,h^{-1}$ Mpc). As pointed out by Z05, this anomalous behaviour is a 'cosmic variance' effect caused by an enormous supercluster at $z \sim 0.08$, which overlaps these three samples. When the Sample VL5 is restricted to redshifts below 0.07, its projected correlation functions drop and steepen, (blue line in Fig. 5), coming into good agreement with that of Z05.

Following Z05, we calculate the relative bias factor b/b^* as a function of the normalized luminosity L/L^*. The relative bias factor is defined by the amplitude of $w_{\rm p}(r_{\rm p})$ measured at a fixed separation $r_{\rm p} = 2.7\ h^{-1}$ Mpc relative to the value measured for the $-21 < M_{0.1r} < -20$ subsample (Sample L9, which has $L \approx L^*$). This fiducial separation of 2.7 h^{-1} Mpc was chosen because it is well out of the very non-linear regime, but still small enough so that the correlation functions are very accurately measured in all surveys. The solid circles in Fig. 6 show our results and the triangles show

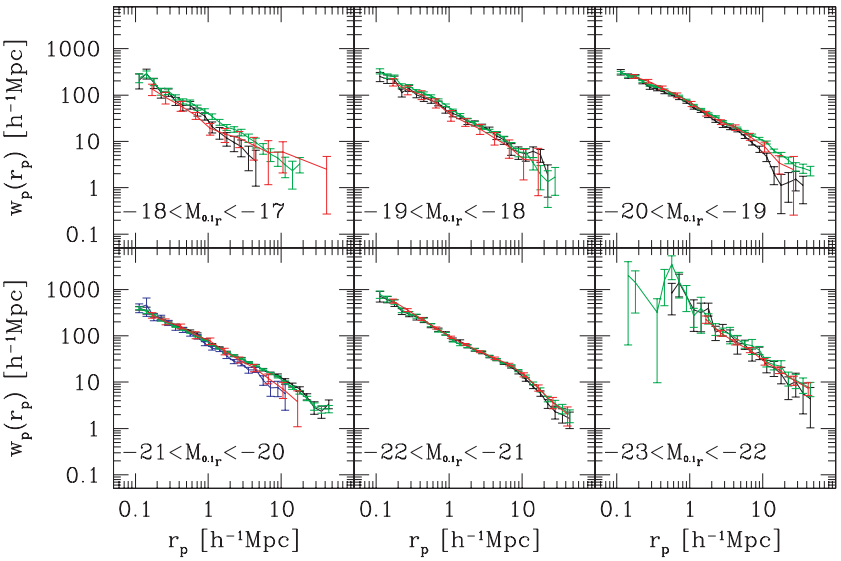

Figure 5. Comparison of our $w_{\rm p}(r_{\rm p})$ measurements with Zehavi et al. (2005, red). The black and green are, respectively, for the volume-limited and magnitude-limited samples. The blue line in the left-hand bottom panel is for the volume-limited sample with the redshift threshold reduced from 0.10 to 0.07. See the text for a more detailed description.

the SDSS results from Z05. The long-dashed curve is taken from Tegmark et al. (2004), where bias factors are derived from the galaxy power spectrum $P(k)$ at the wavelength $2\pi/k \sim 100\,h^{-1}$ Mpc. The results in this paper and in Z05, both of which are derived from $w_{\rm p}(r_{\rm p})$ measurements, agree very well (as they should). Our results are also in quite good agreement with those of Tegmark et al. The dashed curve in Fig. 6 shows the result of Norberg et al. (2001), based on $w_{\rm p}(r_{\rm p})$ measurements of somewhat more luminous galaxies in the 2dF survey ($\log_{10} L/L^* \gtrsim -0.6$). The agreement is again very good over the range of luminosities where the different analyses overlap.

4.2 Stellar mass

In this section, we present the measurements of the projected 2PCF $w_{\rm p}(r_{\rm p})$ as a function of stellar mass. As discussed in Section 3.2, when computing $w_{\rm p}(r_{\rm p})$ as a function of mass, we weight each galaxy pair by the inverse of the volume over which both galaxies can be detected in the survey. The effect of this correction can be seen in Fig. 7 by comparing the black lines (no volume weighting) with the red lines (with volume weighting). As can be seen, the volume correction steepens the correlation function of low-mass galaxies. Recall that if no volume correction is applied, the sample is biased towards the galaxies with low M_*/Ls. As we will show in detail in the following section, the slope of the correlation function is very sensitive to the colour (and hence the young stellar content) of galaxies, particularly for low-mass systems. This is why the volume corrections make the most difference for the galaxies in our two lowest-mass bins. We have also compared our results with the measurements obtained using samples that are volume limited in stellar mass (blue lines). As can be seen, the results obtained for the volume-limited samples agree very well with the volume-corrected $w_{\rm p}(r_{\rm p})$.

The bottom panel of Fig. 6 shows the relative bias factor b/b^* at $r_{\rm p} = 2.7\,h^{-1}$ Mpc as a function of stellar mass, with points showing the results from our $w_{\rm p}(r_{\rm p})$ measurements based on Samples M1–M6, and dashed lines showing the fit to the measurements $b/b^* = 0.90 + 0.10M/M^*$. The value M^* is determined by fitting a Schechter function to the stellar mass function of the galaxies in our sample. We obtain $M^* = (4.11 \pm 0.02) \times 10^{10}\,h^{-2}\,{\rm M}_\odot$, $\alpha = -1.073 \pm 0.003$ and $\phi^* = 0.0204 \pm 0.0001\,h^3\,{\rm Mpc}^{-3}$ (Wang et al., in preparation). Qualitatively, the behaviour of the relative bias as a function of M_* is very similar to the results obtained as a function of L. This is not surprising, because luminosity and stellar mass are reasonably tightly correlated (see Fig. 2). What is of interest, however, is that these measurements can be used to set constraints on the fraction of baryons that have been turned into stars in dark matter haloes of different mass. We will come back to this in future work.

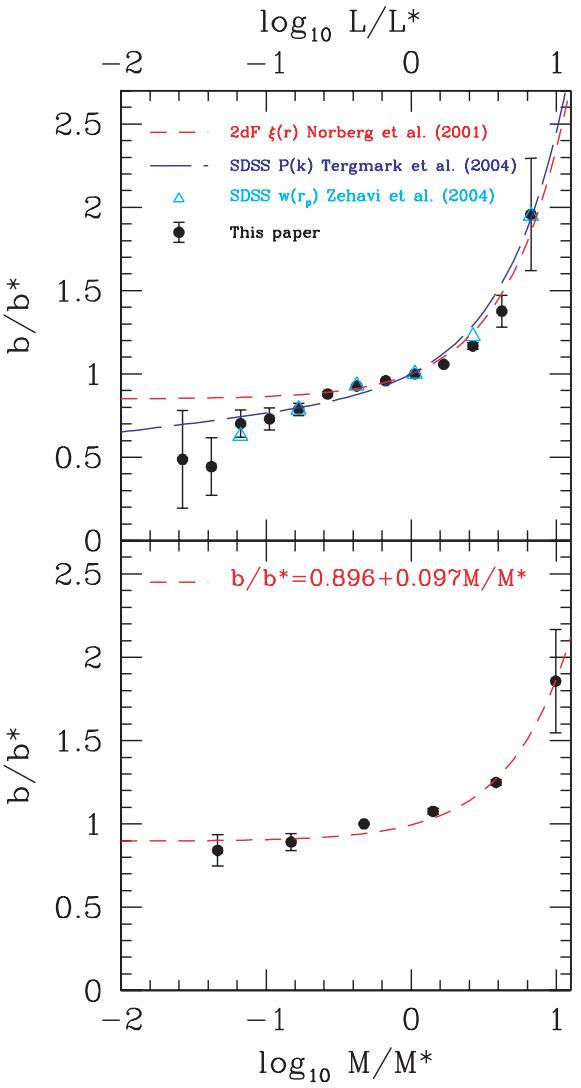

Figure 6. Top panel: relative bias factors for luminosity subsamples (Samples L1–L13). Bias factors are defined by the relative amplitude of the $w_{\rm p}(r_{\rm p})$ estimates at a fixed separation of $r_{\rm p} = 2.7\,h^{-1}$ Mpc and are normalized by the $-21 < M_{0.1r} < -20$ sample (Samples L9, $L \approx L^*$). The dashed curve is a fit to $w_{\rm p}(r_{\rm p})$ measurements in the 2dF survey $b/b^* = 0.85 + 0.15L/L^*$ (Norberg et al. 2001), and the long-dashed curve is a fit obtained from measurements of the SDSS power spectrum, $b/b^* = 0.85 + 0.15L/L^* - 0.04(M - M^*)$ (Tegmark et al. 2004; note that here the symbols M and M^* are for absolute magnitudes, but not for stellar mass.). The triangles are obtained from the $w_{\rm p}(r_{\rm p})$ measurements of Zehavi et al. (2005). Bottom panel: relative bias factors for stellar mass subsamples (Samples M1–M6). Bias factors are normalized by Sample M3, where the mean stellar mass is close to the characteristic stellar mass of the Schechter mass function.

4.3 Division by physical parameters

We now investigate how the clustering of galaxies of given luminosity (or stellar mass) depends on properties such as colour, 4000-Å break strength, concentration and surface mass density. Z05 performed such an analysis in the space of luminosity versus $g - r$ colour. They adopted a tilted colour cut motivated by the colour–magnitude diagram. A similar colour division is presented in Baldry et al. (2004), who found that the distribution of galaxy colour could be well approximated using the bi-Gaussian functions (see Baldry et al. 2004, also see Fig. 8 here). Fig. 8 shows that other physical quantities, such as D_{4000}, C and μ_* also exhibit bimodal distributions. We thus fit bi-Gaussian functions to the distribution of $g - r$, D_{4000}, C and $\log \mu_*$ for each of the 282 luminosity subsamples described in Section 3.1. These are shown in Fig. 8 for three representative luminosity intervals. In Fig. 9, we illustrate how well these fits recover the true distribution of these parameters as a function of luminosity. Black lines show contours of the actual number density of galaxies and red lines show the predicted number densities from the bi-Gaussian fits. As can be seen, the bi-Gaussian model does a reasonable job of reproducing the observations.

The division of the luminosity subsamples into red and blue, high D_{4000} and low D_{4000}, high concentration and low concentration, high

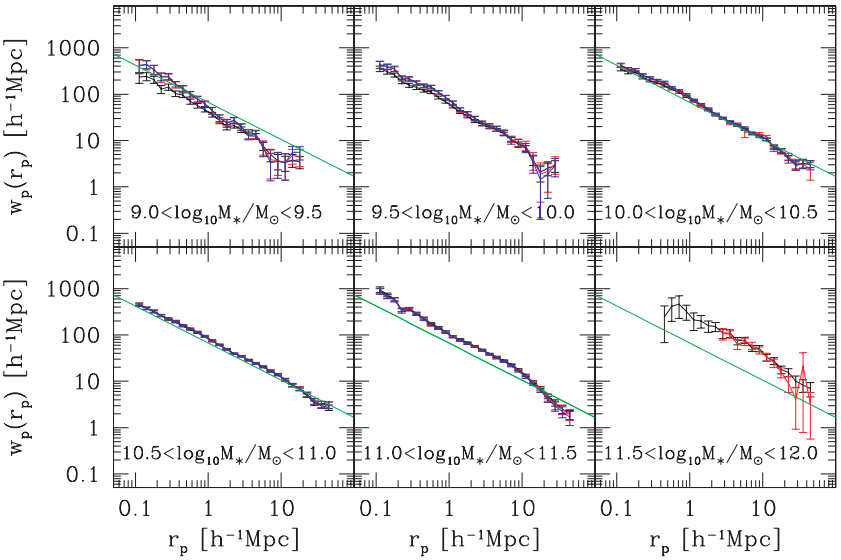

Figure 7. Projected 2PCF for stellar mass subsamples, as indicated. The red lines are for the results obtained by applying volume corrections, compared with those without applying the corrections (black). The blue lines in some panels are for the samples that are volume limited in M_*. In each panel, the green line is the line corresponding to $\xi(r) = (r/5\ h^{-1}\ \mathrm{Mpc})^{-1.8}$.

surface density and low surface density, is defined as the mean of the two Gaussian centres in each luminosity bin. In Fig. 9, triangles indicate the two Gaussian centres and the crosses are the mean of these centres. We fit the dividing point as a function of luminosity using a linear equation of the form (see Fig. 9, blue lines),

$$P = A + B\,M_{0.1r}, \tag{7}$$

where P is the physical parameter under investigation, and A and B are the best-fitting linear coefficients. These are listed in Table 4 for reference. Using these best-fitting cuts (equation 7), we divide the galaxies in each of the 13 luminosity samples (Samples L1–L13) and the five stellar mass samples (Samples M1–M5) into two further subsamples. For simplicity, we use 'red' to denote the subsamples with larger values of the physical quantity and 'blue' for the subsamples with the smaller value. The percentages of the galaxies in the 'red' subsamples are listed in the last four columns of Table 1.

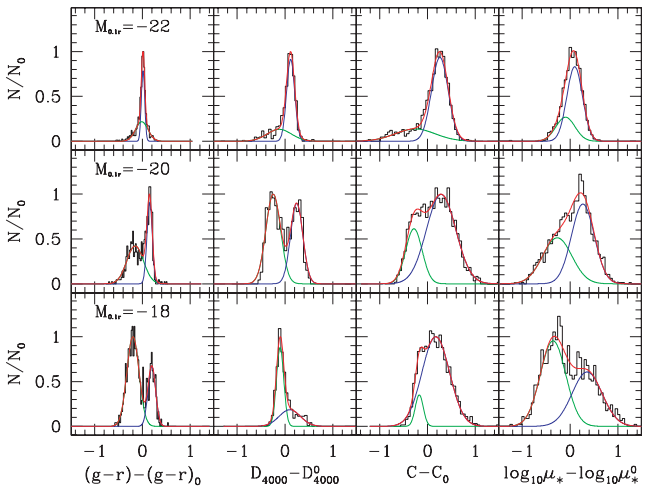

Figure 8. Examples of the bimodal distribution of physical quantities in different luminosity intervals, as indicated. In each panel, the histogram is for the data, while the green and blue lines are the best-fitting Gaussians and the red is the total. N_0 is the maximum of the total fit, and the quantities with a zero give the median of the two Gaussian centres.

4.3.1 *In luminosity bins*

The projected 2PCFs in the space of luminosity versus colour, D_{4000}, concentration and surface density are presented in Fig. 10. Red

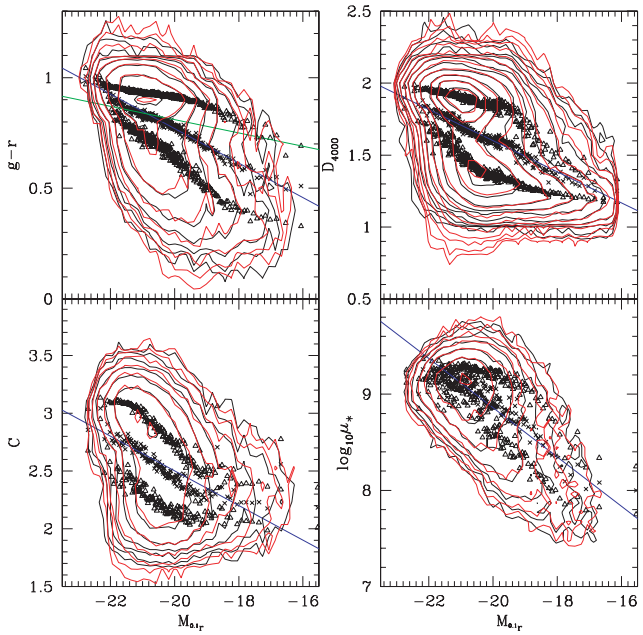

Figure 9. Contours of number density of galaxies in the planes of luminosity versus physical quantities. The black lines are for the data, while the red are reconstructed according to the best-fitting bi-Gaussians (see Fig. 8; also see the text for a detailed description). The blue lines are the best linear fits to the median Gaussian centres as a function of luminosity (see Fig. 8). These are the luminosity-dependent cuts that we adopt for dividing galaxies according to a given physical property. The green line in the top left-hand panel is the $g - r$ cut adopted by Zehavi et al. (2005).

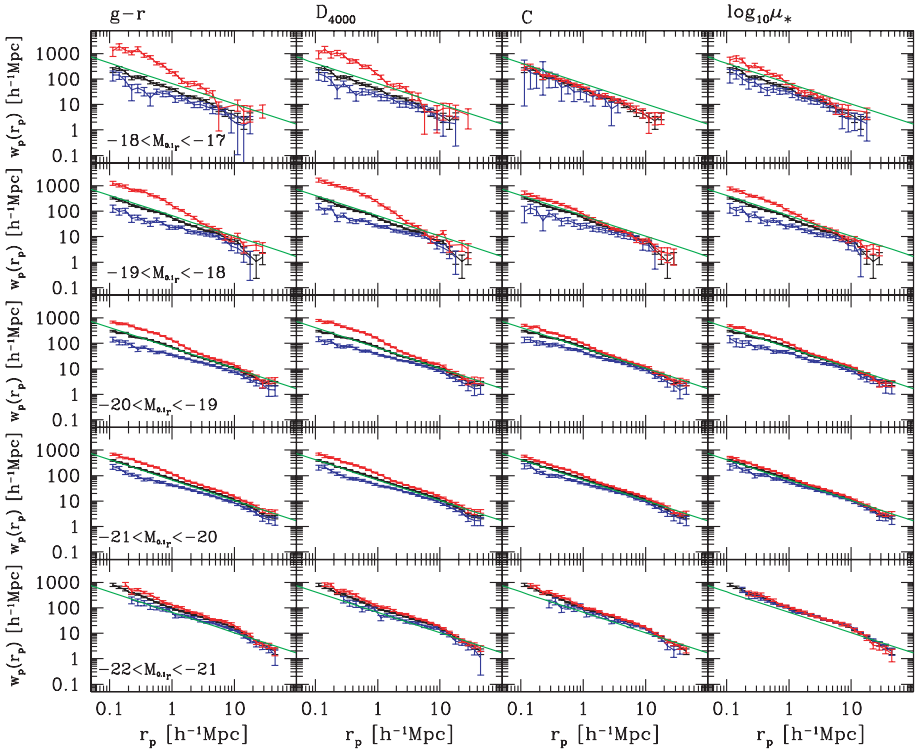

Figure 10. Projected correlation function $w_{\rm p}(r_{\rm p})$ for the galaxies in different luminosity intervals and with different properties (from the left-hand to right-hand panel: $g-r$ colour, D_{4000}, concentration and log stellar surface mass density $\log_{10}\mu^*$). The panels in each column are for different luminosity subsamples, with the range of absolute magnitude indicated in the left-hand column. In each panel, the black is for the full sample, the red (blue) is for the subsample with larger (smaller) value of the corresponding physical parameter. In each panel, the green line is the line corresponding to $\xi(r) = (r/5\ h^{-1}\ {\rm Mpc})^{-1.8}$.

(blue) lines correspond to the 'red' ('blue') subsamples. Black lines are for the sample as a whole. Fig. 11 shows the measurements of the amplitude of $w_{\rm p}(r_{\rm p})$ at $r_{\rm p} = 0.2$, 1, 5 and 10 h^{-1} Mpc.

When the sample is divided by $g-r$ colour, redder galaxies of all luminosities are more strongly clustered and have steeper correlation functions than their blue counterparts. This colour dependence is much stronger for the faint galaxies than for the bright galaxies, particularly on small scales. Fig. 11 shows that the clustering amplitude of the blue galaxies increases as a function of luminosity at all scales. However, the situation is more complicated for red galaxies. On small scales, faint red galaxies are clustered more strongly than bright red galaxies. On large scales, however, the trend reverses and the clustering amplitude increases with luminosity. These results are all consistent with the findings of Z05. The behaviour of the slope of the correlation function as a function of luminosity is also different for red and blue galaxies. The correlation function of faint red galaxies is very steep and the slope flattens systematically as luminosity increases. In contrast, the slope of the correlation function of blue galaxies exhibits rather little change with luminosity. All these trends are qualitatively consistent with a picture in which faint red galaxies are primarily 'satellite' systems in massive dark matter haloes, but faint blue galaxies occupy haloes of smaller mass (Z05; Berlind et al. 2005; Li et al. 2005).

Our results show that the dependence of clustering on D_{4000} is very similar to what is obtained for $g-r$ colour. On the other hand, rather different results are obtained for the structural parameters C and μ_*. Fig. 11 clearly shows that the dependence of $w_{\rm p}(r_{\rm p})$ on $g-r/D_{4000}$ is considerably stronger than the dependence on C/μ_* at all physical scales.

4.3.2 *In stellar mass bins*

The projected 2PCFs in the space of stellar mass versus the same set of physical parameters are presented in Fig. 12. The measurements at $r_{\rm p} = 0.2$, 1, 5 and 10 h^{-1} Mpc are plotted in Fig. 13.

Qualitatively, the results shown in Figs 11 and 13 appear very similar. However, careful comparison of these two figures shows that interesting quantitative differences do exist between the clustering of the 'red' and 'blue' subsamples at fixed luminosity and at fixed stellar mass. On small scales, the dependences are stronger when evaluated at fixed mass, particularly for low-mass galaxies. We also note that there is a small difference in the clustering amplitude of the 'red' and 'blue' subsamples at projected radii as large as 10 h^{-1} Mpc in Fig. 11. This difference is seen both in $g-r$ colour and D_{4000} and more weakly in the structural parameters C and μ_*. Fig. 13 shows, however, that at fixed stellar mass there is no longer any significant difference in the clustering amplitude of the high- and low-concentration galaxies or high surface density and low surface density galaxies on scales larger than 5 h^{-1} Mpc. The clustering differences in $g-r$ and D_{4000} do persist, however. This is a rather surprising result, because at scales larger than a few Mpc, galaxies inside the same dark matter halo no longer contribute to the clustering signal. Our result thus indicates that at fixed stellar mass, the clustering properties of the surrounding *dark matter haloes* are somehow correlated with the colour of the selected galaxies.

To investigate this effect further, we have computed the 2PCF as a function of $g-r$, D_{4000}, C and μ_* for the galaxies spanning a narrow range in stellar mass (10^{10}–10^{11} $M_\odot$). In Fig. 14, we plot the amplitude of the correlation function as a function of these quantities measured on four different physical scales ($r_{\rm p} = 0.2$,1, 5 and

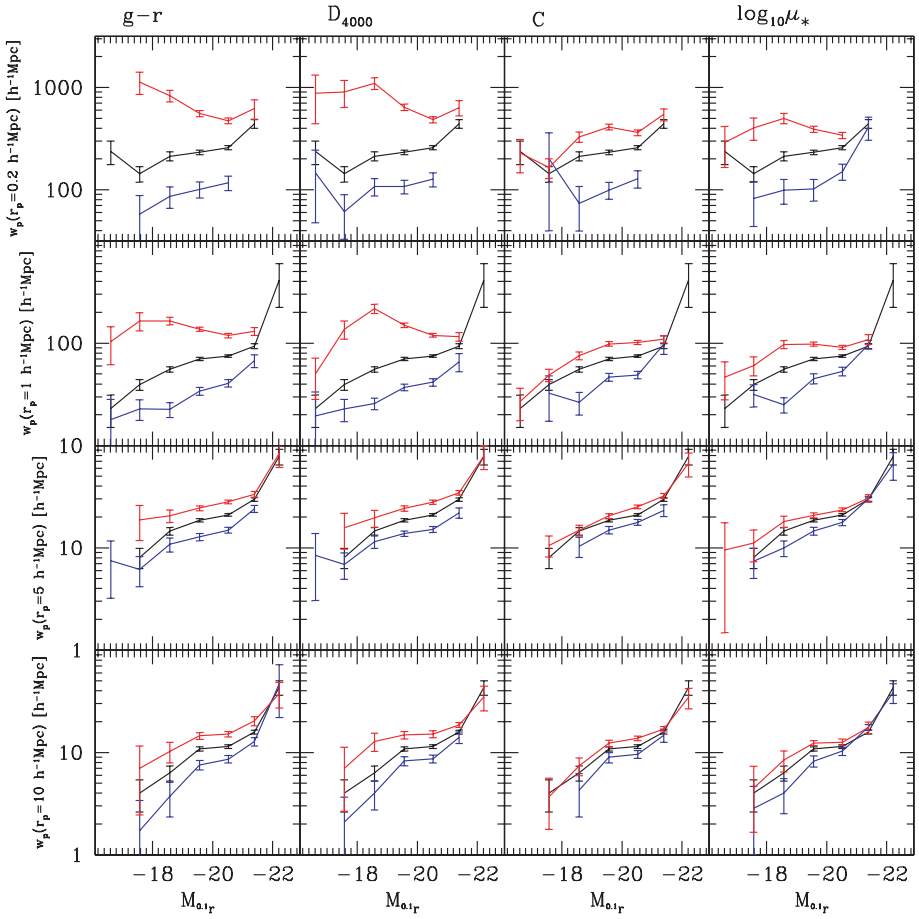

Figure 11. $w_p(r_p)$ measured at $r_p = 0.2$, 1, 5 and $10\,h^{-1}$ Mpc, as a function of luminosity. The different columns show the dependence on different physical quantities, as indicated. In each panel, the black is for the full sample, the red (blue) is for the subsample with larger (smaller) value of the corresponding physical parameter.

$10\,h^{-1}$ Mpc). This figure confirms that the dependence of $w_p(r_p)$ on $g - r/D_{4000}$ extends out to larger physical scales than the dependence of $w_p(r_p)$ on C/μ_*. The figure also shows that the dependence of $w_p(r_p)$ on C and μ_* is also qualitatively quite different on small scales. On scales less than $<1\,h^{-1}$ Mpc, the amplitude of the correlation function is constant for 'young' galaxies with $1.1 < D_{4000} < 1.5$ and a steeply rising function of age for 'older' galaxies with $D_{4000} > 1.5$. In contrast, the dependence of the amplitude of $w_p(r_p)$ on concentration is strongest for disc-dominated galaxies with $C < 2.6$ on these same scales. This demonstrates that different physical processes are required to explain environmental trends in star formation and in galaxy structure.

5 SUMMARY AND DISCUSSION

In this paper we present our determinations of the projected 2PCF $w_p(r_p)$ for different classes of galaxies in order to study the dependence of clustering on the physical properties of these systems. We use the NYU-VAGC which is constructed from the SDSS DR2.

The conclusions of this paper can be summarized as follows.

(i) We confirm previous findings that the luminous galaxies cluster more strongly than the faint galaxies, with the difference becoming larger for the galaxies with $L > L^*$, where L^* is the characteristic luminosity of the Schechter (1976) function. The dependence of galaxy clustering on luminosity is different on different physical scales. On small scales ($r_p \sim 0.2\,h^{-1}$ Mpc), the correlation amplitude is almost constant for the galaxies fainter than L^*, but the amplitude increases sharply above L^*. On large scales, the correlation amplitude increases more continuously as a function of luminosity. Around L^* there appears to be a shoulder, with $w_p(r_p)$ increasing more steeply with L for higher luminosity galaxies. Our results are in good agreement with previous studies of clustering as a function of luminosity in the SDSS.

(ii) We present $w_p(r_p)$ as a function of stellar mass. In analogy with previous results obtained as a function of luminosity, we find that more-massive galaxies cluster more strongly than less-massive galaxies, with the difference increasing above the characteristic stellar mass M^* of the Schechter mass function.

(iii) When galaxies are divided according to their physical properties, we find that galaxies with redder colours, larger 4000-Å break strengths, more concentrated structure, and higher surface mass densities cluster more strongly and have steeper correlation functions at all luminosities and masses. The differences in clustering strength are larger on small scales and for low-luminosity and less-massive galaxies.

(iv) We have found that the dependence of $w_p(r_p)$ on $g - r$ or D_{4000} extends out to larger physical scales ($r_p > 5\,h^{-1}$ Mpc) than the dependence of $w_p(r_p)$ on C or μ_*. On small scales ($\sim 0.2\,h^{-1}$ Mpc), the behaviour of $w_p(r_p)$ as a function of $g - r$ or D_{4000} and as a function of C are qualitatively different.

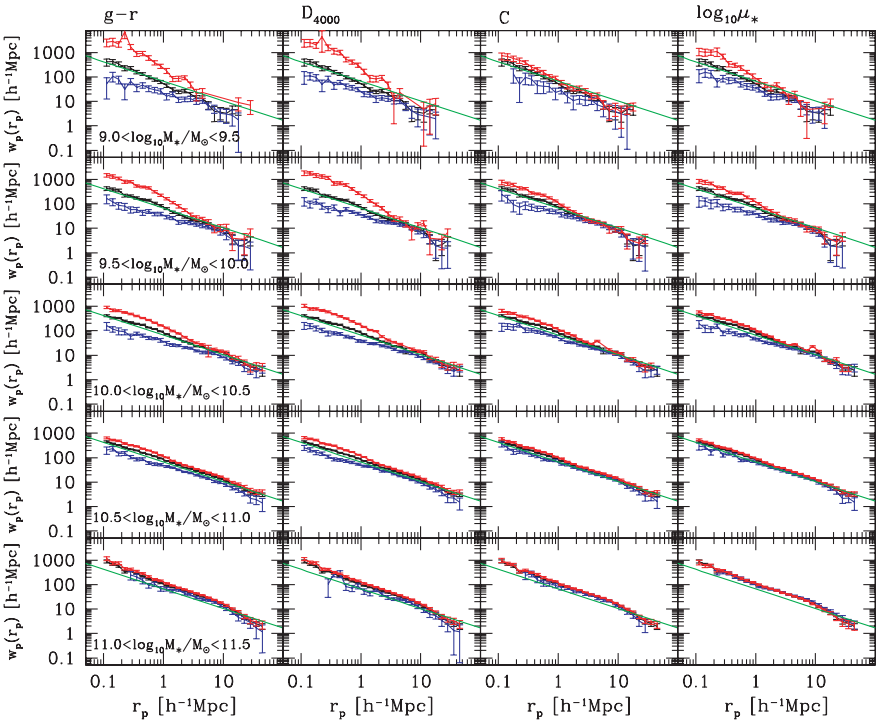

Figure 12. Projected correlation function $w_{\rm p}(r_{\rm p})$ for the galaxies in different stellar mass intervals and with different physical properties (from the left-hand to right-hand panel: $g-r$ colour, D_{4000}, concentration and log stellar surface mass density $\log_{10}\mu^*$). The panels in each column are for different stellar mass subsamples, with the range of stellar mass indicated in the left-hand column. In each panel, the black is for the full sample, the red (blue) is for the subsample with larger (smaller) value of the corresponding physical parameter, and the green line is the line corresponding to $\xi(r) = (r/5\,h^{-1}\,{\rm Mpc})^{-1.8}$.

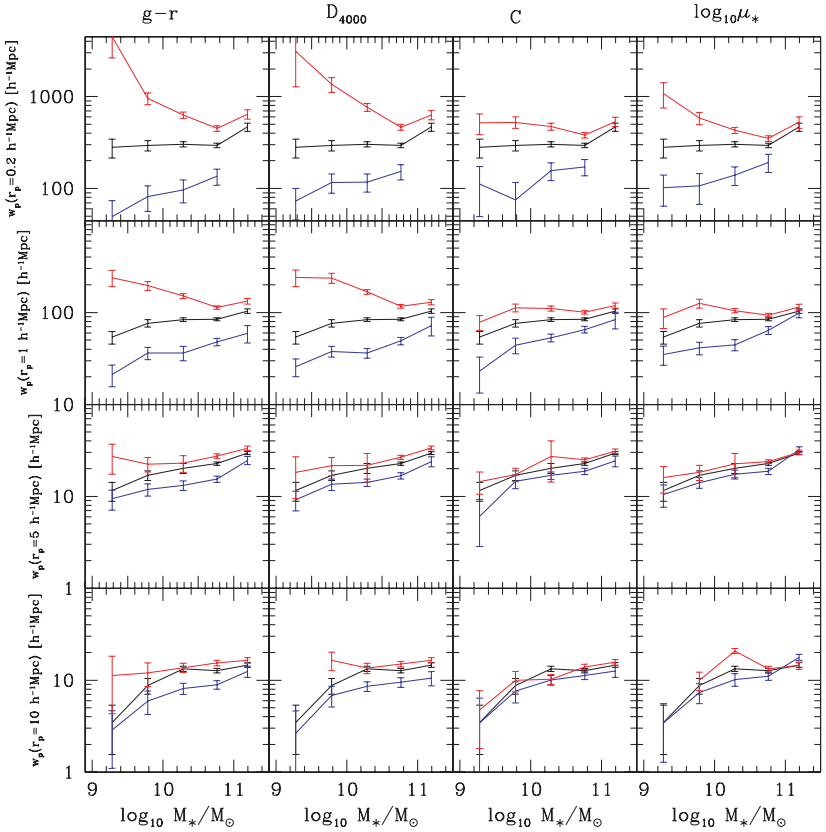

Figure 13. $w_{\rm p}(r_{\rm p})$ measured at $r_{\rm p} = 0.2$, 1, 5 and $10\,h^{-1}$ Mpc, as a function of stellar mass. The different columns show the dependence on different physical quantities, as indicated. In each panel, the black is for the full sample, the red (blue) is for the subsample with larger (smaller) value of the corresponding physical parameter.

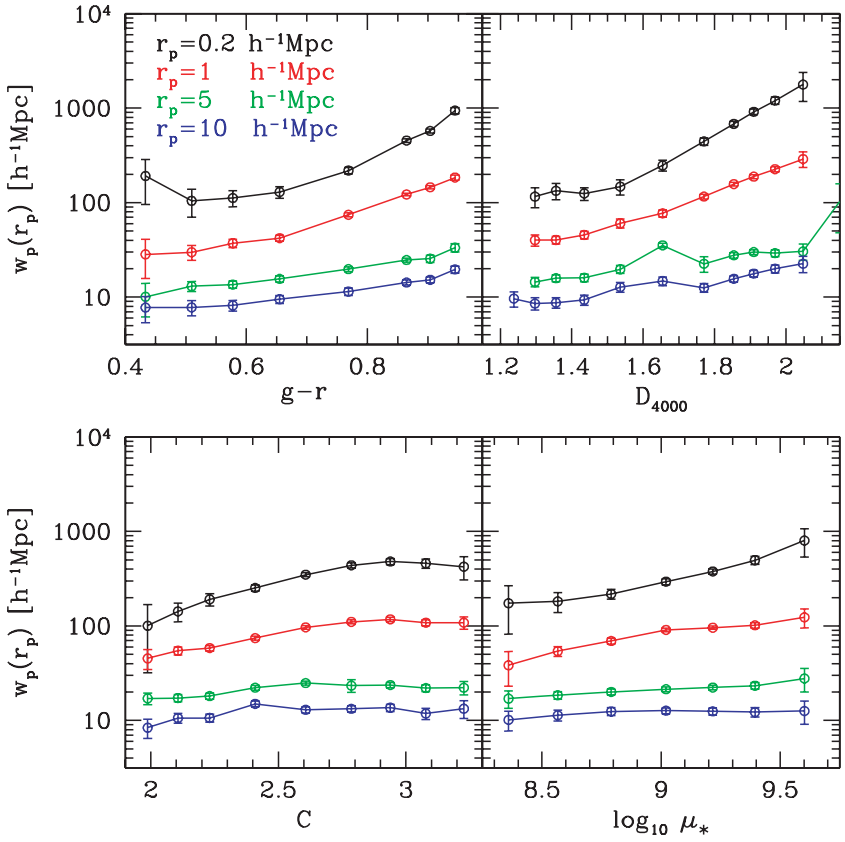

Figure 14. Projected correlation functions measured at $r_{\rm p} = 0.2$, 1, 5 and $10h^{-1}$ Mpc, as a function of different physical parameters. In all panels, the stellar masses are limited to the range $10 < \log_{10} M_*/{\rm M}_\odot < 11$.

We have chosen not to express our results in terms of power-law fits to our $w_{\rm p}(r_{\rm p})$ measurements. We have tabulated the measurements of our correlation functions so that they can be accurately recovered. As discussed by Z05, a single power law is a poor description of the data and as we expand our exploration of physical parameter space, it is important not to place unnecessary restrictions on the way in which the observational results are described. In this paper, we have chosen to plot trends in clustering *amplitude* evaluated on a variety of different physical scales. This leads to a number of interesting insights that have not received much attention up to now: (i) the dependence of the clustering amplitude (or equivalently, the relative bias factor) on luminosity is qualitatively different on small scales and on large scales; and (ii) there is a different scale dependence in the amplitude of the correlation function for parameters that measure the star formation histories of galaxies and for parameters that measure galaxy structure, suggesting that the trends in star formation and in galaxy structure are governed by different physical processes.

Finally, it is worth comparing our results with the many studies that have examined the correlations between galaxy properties and the local environment. One of the most fundamental correlations between the properties of galaxies in the local Universe is the so-called morphology–density relation. Oemler (1974) and Dressler (1980) pioneered the quantification of this relation, showing that spheroidal systems reside preferentially in dense regions. Since the standard morphological classification scheme mixes elements that depend on the structure of a galaxy with elements related to its recent star formation history, it is by no means obvious that these two elements should depend on environment in the same way.

Recent studies using large surveys such as the SDSS have revealed that galaxy colour is the galaxy property most predictive of the local environment (e.g. Blanton et al. 2005b; Kauffmann et al. 2004). Hogg et al. (2003) showed that the local density increases strongly with luminosity for the brightest galaxies. For the faint galaxies, local density is sensitive mainly to colour, with faint red galaxies occupying highest-density regions. Blanton et al. (2005b) found that at fixed luminosity and colour, density is not closely related to surface brightness or to the Sérsic index (a quantity related to galaxy structure), so that the morphological properties of galaxies are less closely related to galaxy environment than their luminosities and star formation histories. Kauffmann et al. (2004) obtained very similar results. They found that at fixed stellar mass both star formation and nuclear activity depend strongly on local density, while structural parameters such as size and concentration are almost independent of it.

Our analyses of $w_{\rm p}(r_{\rm p})$ as a function of luminosity, colour and structural parameters are consistent with these conclusions. The power of the $w_{\rm p}(r_{\rm p})$ statistic is that it encapsulates information about how galaxy properties depend on environment over a wide range of physical scales. Kauffmann et al. (2004) found no evidence for a significant dependence of galaxy structure on local density. However, their local densities are calculated in a fixed aperture of $2\,h^{-1}$ Mpc, whereas our plots (see Fig. 14) show clearly that the dependence of structural parameters on environment becomes significant on scales that are smaller than this value.

The other advantage of $w_{\rm p}(r_{\rm p})$ is that it can be very easily compared with the predictions of galaxy formation simulations. It probes the physical processes occurring inside individual dark matter haloes

as well the masses of the dark matter haloes that host galaxies of given mass, luminosity, size, age and concentration, thus placing strong constraints on theoretical models. This will be the focus of future work.

ACKNOWLEDGMENTS

We are grateful to Dr Idit Zehavi for providing her $w_p(r_p)$ measurements and for her detailed comments on our paper, and to Dr Michael Blanton for his help with the NYU-VAGC. We thank the SDSS teams for making their data publicly available. This work is supported by NKBRSF (G19990754), by NSFC (Nos.10125314, 10373012, 10073009), by Shanghai Key Projects in Basic research (04jc14079, 05xd14019), by the Max Planck Society, and partly by the Excellent Young Teachers Programme of MOE, P.R.C. CL acknowledges the financial support of the exchange programme between Chinese Academy of Sciences and the Max Planck Society.

Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS website is http://www.sdss.org/. The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, the Korean Scientist Group, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

REFERENCES

Abazajian K. et al., 2004, AJ, 128, 502
Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681
Balogh M. L., Morris S. L., Yee H. K. C., Carlberg R. G., Ellingson E., 1999, ApJ, 527, 54
Barrow J. D., Bhavsar S. P., Sonoda D. H., 1984, MNRAS, 210, 19
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000a, MNRAS, 311, 793
Benson A. J., Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 2000b, MNRAS, 316, 107
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Berlind A. A., Blanton M. R., Hogg D. W., Weinberg D. H., Davé R., Eisenstein D. J., Katz N., 2005, ApJ, 629, 625
Blanton M. R. et al., 2003a, AJ, 125, 2348
Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003b, AJ, 125, 2276
Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003c, ApJ, 592, 819
Blanton M. R. et al., 2003d, preprint (astro-ph/0310453)
Blanton M. R. et al., 2005a, AJ, 129, 2562
Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005b, ApJ, 629, 143
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Budavári T. et al., 2003, ApJ, 595, 59
Cooray A., Sheth R., 2002, PhR, 372, 1
Davis M., Geller M. J., 1976, ApJ, 208, 13
Dressler A., 1980, ApJ, 236, 351
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Goto T., Yamaguchi C., Fujita Y., Okamura S., Sekiguchi M., Smail I., Bernardi M., Gomez P., 2003, MNRAS, 346, 601
Gross M. A. K., Somerville R. S., Primack J. R., Holtzman J., Klypin A., 1998, MNRAS, 301, 81
Gunn J. E. et al., 1998, AJ, 116, 3040
Hamilton A. J. S., 1993, ApJ, 417, 19
Hamilton A. J. S., Tegmark M., 2002, MNRAS, 330, 506
Hawkins E. et al., 2003, MNRAS, 346, 78
Hogg D. W., Finkbeiner D. P., Schlegel D. J., Gunn J. E., 2001, AJ, 122, 2129
Hogg D. W. et al., 2003, ApJ, 585, L5
Ivezić Ž. et al., 2004, AN, 325, 583
Jenkins A. et al., 1998, ApJ, 499, 20
Jing Y. P., Börner G., 2004, ApJ, 617, 782
Jing Y. P., Mo H. J., Börner G., 1998, ApJ, 494, 1
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., 2003a, MNRAS, 341, 33
Kauffmann G. et al., 2003b, MNRAS, 341, 54
Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, MNRAS, 353, 713
Li C., Jing Y. P., Kauffmann G., Boerner G., White S. D. M., Cheng F. Z., 2006, MNRAS, this issue (doi: 10.1111/j.1365.2966.2006.10177.x)
Madgwick D. S. et al., 2003, MNRAS, 344, 847
Miller C. J., Nichol R. C., Gomez P. L., Hopkins A. M., Bernardi M., 2003, ApJ, 597, 142
Mo H. J., Jing Y. P., Börner G., 1992, ApJ, 392, 452
Norberg P. et al., 2001, MNRAS, 328, 64
Norberg P. et al., 2002, MNRAS, 332, 827
Oemler A., 1974, ApJ, 194, 1
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Pier J. R., Munn J. A., Hindsley R. B., Hennessy G. S., Kent S. M., Lupton R. H., Ivezić Ž., 2003, AJ, 125, 1559
Schechter P., 1976, ApJ, 203, 297
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Seljak U., 2000, MNRAS, 318, 203
Smith J. A. et al., 2002, AJ, 123, 2121
Stoughton C. et al., 2002, AJ, 123, 485
Strateva I. et al., 2001, AJ, 122, 1861
Tegmark M., Hamilton A. J. S., Xu Y., 2002, MNRAS, 335, 887
Tegmark M. et al., 2004, ApJ, 606, 702
Worthey G., Ottaviani D. L., 1997, ApJS, 111, 377
Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057
York D. G. et al., 2000, AJ, 120, 1579
Zehavi I. et al., 2002, ApJ, 571, 172
Zehavi I. et al., 2004, ApJ, 608, 16
Zehavi I. et al., 2005, ApJ, 630, 1 (Z05)

APPENDIX A: DESCRIPTION OF ONLINE TABLES

Tables 5 and 6 contain our $w_p(r_p)$ measurements for the galaxies with different luminosities/stellar masses. Results are also given as a function of the physical quantities $g - r$, D_{4000}, C and μ_*. The tables are available in electronic form from http://www.blackwell-synergy.com and at http://www.mpa-garching.mpg.de/~leech/papers/clustering/.

Tables 5 and 6 list the data points in Figs 10 and 12 respectively, that is, the measured projected 2PCF $w_p(r_p)$ for the galaxies with different luminosity or stellar mass intervals and with different

properties. Table 5 consists of 13 separate parts corresponding to the 13 luminosity samples (Samples L1–L13, in Table 1). Likewise, Table 6 consists of five parts corresponding to the five stellar mass samples (Samples M1–M5, in Table 1). To make the description clearer, we present here an abridged version for the first part in Table 5, listing only the first several rows for Sample L1. The first column is the projected separation $r_{\rm p}$ in units of h^{-1} Mpc, ranging from $\sim$0.1 to $\sim$45 h^{-1} Mpc. The other columns give the $w_{\rm p}(r_{\rm p})$ measurements and errors for the full luminosity/stellar mass sample (Column 2) and the 'red' and 'blue' subsamples divided by $g-r$ (Columns 3–4), D_{4000} (Columns 5–6), C (Columns 7–8) and $\log_{10}\mu_*$ (Columns 9–10). Short straight lines denote the points that have no measurements, either because of low signal-to-noise ratio or for any other reason.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this paper online.

This material is available as part of the online paper from http://www.blackwell-synergy.com.

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