

# UC San Diego

## UC San Diego Previously Published Works

### Title

Measures of galaxy environment - II. Rank-ordered mark correlations

### Permalink

<https://escholarship.org/uc/item/8w39d1tt>

### Journal

Monthly Notices of the Royal Astronomical Society, 429(1)

### ISSN

0035-8711 1365-2966

### Authors

Skibba, R. A

Sheth, R. K

Croton, D. J

et al.

### Publication Date

2013-02-01

### DOI

10.1093/mnras/sts349

Peer reviewed

# Measures of galaxy environment – II. Rank-ordered mark correlations

Ramin A. Skibba,<sup>1,2</sup>★ Ravi K. Sheth,<sup>3,4</sup> Darren J. Croton,<sup>5</sup> Stuart I. Muldrew,<sup>6</sup>  
Ummi Abbas,<sup>7</sup> Frazer R. Pearce<sup>6</sup> and Genevieve M. Shattow<sup>5</sup>

<sup>1</sup>Steward Observatory, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85721, USA

<sup>2</sup>Center for Astrophysics and Space Sciences, Department of Physics, University of California, 9500 Gilman Dr, La Jolla, San Diego, CA 92093, USA

<sup>3</sup>The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, I-34151 Trieste, Italy

<sup>4</sup>Center for Particle Cosmology, University of Pennsylvania, 209 S. 33rd St, Philadelphia, PA 19104, USA

<sup>5</sup>Centre for Astrophysics & Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia

<sup>6</sup>School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD

<sup>7</sup>INAF – Osservatorio Astrofisico di Torino, I-10025 Pino Torinese, Italy

Accepted 2012 November 2. Received 2012 October 30; in original form 2012 August 16

## ABSTRACT

We analyse environmental correlations using mark clustering statistics with the mock galaxy catalogue constructed by Muldrew et al. (Paper I). We find that mark correlation functions are able to detect even a small dependence of galaxy properties on the environment, quantified by the overdensity  $1 + \delta$ , while such a small dependence would be difficult to detect by traditional methods. We then show that rank ordering the marks and using the rank as a weight is a simple way of comparing the correlation signals for different marks. With this we quantify to what extent fixed-aperture overdensities are sensitive to large-scale halo environments, nearest-neighbour overdensities are sensitive to small-scale environments within haloes and colour is a better tracer of overdensity than luminosity.

**Key words:** methods: statistical – galaxies: evolution – galaxies: haloes – large-scale structure of Universe.

## 1 INTRODUCTION

In hierarchical clustering models, clues about the galaxy formation process are encoded in correlations between galaxy properties and their environments. This has motivated measurements of such correlations. Traditional measures are intended to allow one to quantify whether galaxies in dense regions tend to be more luminous, or redder, or older or tend to move faster than average, and so on. These conclusions depend critically on how the density  $N_g/V$  is estimated: fixed-aperture (FA) measurements count the number of galaxies  $N_g$  that are within volume  $V$  of an object (i.e. the numerator of the ratio  $N_g/V$  varies from one object to another), whereas near-neighbour measurements find the  $V$  that contains  $N_g$  nearest neighbours (i.e. the denominator is stochastic). Clearly, the size and shape of  $V$  or the choice of  $N_g$  matters greatly (the Universe is homogeneous on sufficiently large  $N_g$  or  $V$ ). In addition, the choice of three-dimensional or projected surface density matters as well, as do the redshift uncertainties and sample selection. Determining which of the many observed correlations is fundamental, and which is a consequence of others, can be a subtle task, especially since the environment is often the least well determined of a galaxy's attributes. Moreover,

the estimate of the environment is often sufficiently complicated that it cannot be modelled or interpreted analytically.

Mark clustering statistics are fundamentally different in the sense that they are, strictly speaking, statements about pairs, triples, quadruples, etc., of galaxies, rather than about single objects (Stoyan & Stoyan 1994; Martínez et al. 2010). For example, the most commonly used such statistic returns an estimate of how the properties of galaxy pairs (rather than of single galaxies) depend on pair separation. (While this is easily extended to triples, quadruples, etc., such estimates are rarely ever made.) In essence, for each pair separation  $r$ , this statistic weights each galaxy in a pair by its own attribute (e.g. luminosity, colour, etc., expressed in units of the mean across the population) and then divides this weighted pair count by the unweighted one. Symbolically, one may write this statistic as  $WW(r)/DD(r)$ , where  $WW$  and  $DD$  stand for the weighted and unweighted pair counts at separation  $r$ , respectively.

Previous estimates of  $WW/DD$  have shown that close pairs of galaxies are more luminous (Beisbart & Kerscher 2000), redder (Skibba et al. 2006) and older (Sheth et al. 2006) than average ( $WW/DD > 1$  for  $r$  less than a few  $h^{-1}$  Mpc). While these trends are qualitatively the same as those returned by traditional estimates, and they are also in qualitative agreement with galaxy formation models (Sheth 2005), the mark statistics are particularly interesting because a theoretical framework exists for interpreting such measurements quantitatively (Sheth 2005; Skibba et al. 2006). On the other hand,

★ E-mail: rskibba@ucsd.edu

this is also a drawback, because the theoretical framework is almost required if one wishes to draw more than qualitative conclusions from such measurements. This is because the magnitude of the (say) luminosity-weighted signal changes if one weights instead by the log of the luminosity (Sheth, Connolly & Skibba 2005; Skibba et al. 2006). Since the same physics has led to both signals, one would like the measurement not to depend on the ‘units’ in which the measurement was made.

This dependence on ‘units’ derives from the fact that the magnitude of WW/DD depends on the distribution of the weights (e.g. its width, the length of its tails, etc.). Needless to say, it also complicates efforts to determine if one observable correlates more strongly with environment than another. As a case in point, it has long been known that cluster galaxies tend to be redder than average, but there is a wide range in luminosity between the brightest cluster galaxy and the dwarf satellites in a cluster. Since clusters are regions of high density, one naively expects to find that colour correlates more strongly with environment than does luminosity. However, the mark correlation signal appears to have a larger amplitude for luminosity than it does for colour (Skibba & Sheth 2009). One of the main goals of the present work is to show how to remove this effect from the measurement, so that the magnitude of the signal can be compared across different weights.

In the next section, we describe the galaxy catalogues used throughout the paper. In Section 3 we demonstrate that mark correlations are particularly sensitive probes of environmental correlations; they correctly show a strong signal even when traditional estimates based on how one-point statistics vary as a function of environment are unable to see one – a fact that was recently exploited by Paranjape & Sheth (2012). In Section 4 we show how to remove the effect of the mark correlation signal on ‘units’, arguing that for any given weight, one should simply rank order and use the rank as the weight. We then use these rank-ordered mark correlations to compare different mark correlation signals with one another. A final section summarizes our findings.

Throughout this paper we assume a spatially flat cosmology with  $\Omega_m = 0.25$ ,  $\Omega_\Lambda = 0.75$  and  $\sigma_8 = 0.9$ . We write the Hubble constant as  $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

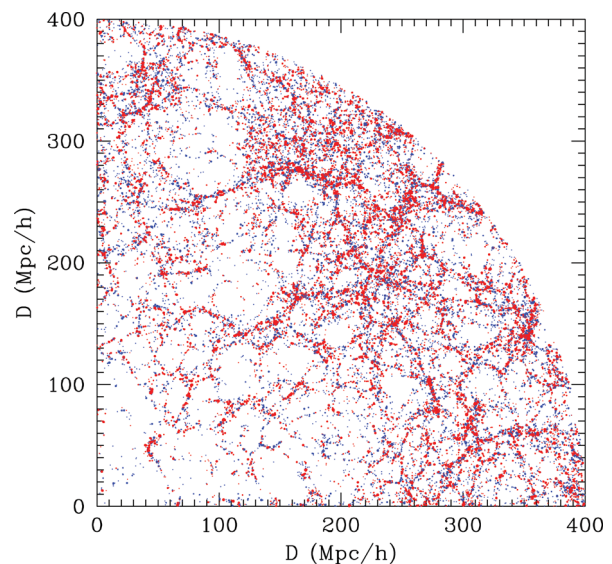
## 2 DATA

### 2.1 Mock galaxy catalogue

To illustrate our methods and to interpret some of our results, we use the mock galaxy catalogue of Muldrew et al. (2012, hereafter M12). We refer the reader to M12 for details about the dark matter simulation, halo-finding algorithm and the procedure for populating the haloes with galaxies.

We begin with the Millennium Simulation (Springel et al. 2005), which is a large  $N$ -body simulation of dark matter structure in a cosmological volume. Dark matter particles are traced in a cubic box of  $500 h^{-1} \text{ Mpc}$  on a side, with a halo mass resolution of  $\sim 5 \times 10^{10} h^{-1} M_\odot$ . Collapsed haloes with at least 20 particles are identified with a friends-of-friends group finder.

The haloes are populated with galaxies with luminosities and colours, following the algorithm described in Skibba et al. (2006) and Skibba & Sheth (2009), which is constrained by the luminosity and colour distribution and clustering in the Sloan Digital Sky Survey (SDSS; York et al. 2000). An important assumption in the model is that all galaxy properties – their numbers, spatial distributions, velocities, luminosities and colours – are determined by halo mass alone. We specify a minimum  $r$ -band luminosity for the galaxies in



**Figure 1.** Slice of the redshift-space mock light cone, showing galaxies within  $\pm 4^\circ$ . The red/blue points denote galaxies on the red/blue sequences of the colour–magnitude diagram, and larger/smaller points represent brighter/fainter galaxies.

the catalogue,  $M_r = -19$ , to stay well above the resolution limit of the Millennium Simulation, avoiding any issues of completeness that may bias our results.

This procedure produces a mock galaxy catalogue containing 1.84 million galaxies, of which 29 per cent are ‘satellite’ galaxies. Galaxies occupy haloes with masses ranging from  $10^{11}$  to  $10^{15.3} h^{-1} M_\odot$ . We show a slice of the mock light cone in Fig. 1, in which more luminous galaxies are identified with larger points, and red and blue sequence galaxies with red and blue points.

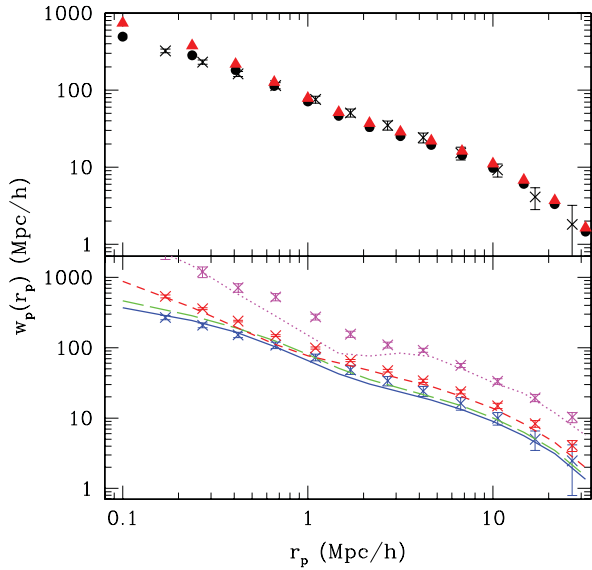
### 2.2 SDSS galaxy catalogue

For comparison, we will also show clustering measurements in the main galaxy sample of the SDSS Seventh Data Release (DR7; Abazajian et al. 2009), in a catalogue volume limited to  $M_r < -19$ , with  $0.02 < z < 0.0642$ .

Clustering measurements of galaxy redshift surveys have traditionally been done by splitting a catalogue into luminosity bins (e.g. Norberg et al. 2002; Zehavi et al. 2005, 2011; Li et al. 2006; Coil et al. 2008). We compare such clustering<sup>1</sup> in the SDSS and in the mock catalogue in Fig. 2. There is generally very good agreement, except at  $0.7 < r_p < 3 \text{ Mpc } h^{-1}$  for the brightest galaxies. Note that the mock catalogue was constrained by earlier SDSS data sets (Zehavi et al. 2005; Skibba et al. 2006), while these new measurements use the full DR7 data set (Zehavi et al. 2011).

Most of the analysis of environmental correlations throughout this paper is based on the mock galaxy catalogue, with additional comparisons to the SDSS galaxy catalogue in Section 4.4.

<sup>1</sup> The clustering measurements in this paper are performed using the NTROPY code developed by Gardner, Connolly & McBride (2007).



**Figure 2.** Upper panel: projected correlation function (circle points) and  $r$ -band luminosity-weighted correlation function (triangles) of the full  $M_r < -19$  mock catalogue. Lower panel: CFs for luminosity bins  $-20 < M_r < -19$  (blue solid line),  $-21 < M_r < -20$  (green long dashed line),  $-22 < M_r < -21$  (red short dashed line) and  $-23 < M_r < -22$  (magenta dotted line). The crosses denote the CFs measured from SDSS DR7 (Zehavi et al. 2011), and the measurement for  $-21 < M_r < -20$  is omitted, for clarity.

### 3 SENSITIVITY TO ENVIRONMENTAL CORRELATIONS

In what follows, we will refer to any property of a galaxy, e.g. its luminosity or its colour, as a ‘mark’. As stated above, the most commonly used mark statistic is the mark correlation function (CF), which is defined as the ratio of the weighted to unweighted CF:

$$M(r) \equiv \frac{1 + W(r)}{1 + \xi(r)} \approx \frac{WW(r)}{DD(r)}, \quad (1)$$

where  $WW/DD$  is the pair count ratio. The mark projected CF is similarly defined:  $M(r_p) \equiv (1 + W_p/r_p)/(1 + w_p/r_p)$ . If the weighted and unweighted clustering are significantly different at a particular separation  $r$ , then the mark is correlated (or anticorrelated) with the environment at that scale; the degree to which they are different quantifies the strength of the correlation.

In this section, we illustrate that mark CFs are particularly sensitive to environmental effects. We do so by introducing a small additional dependence of a galaxy mark  $w$  (luminosity or colour here) on the galaxy’s overdensity (using one of the environment measures defined below). (We use the letter  $w$  because, when we discuss mark correlations below, we treat the mark  $w$  as a ‘weight’.) That is, if a galaxy has mark  $w$ , then we change it to

$$w_\alpha = w(1 + \delta)^\alpha. \quad (2)$$

We then rank order the marks  $w_\alpha$  and rescale them so that they have the same distribution as before the environmental effect was added. That is, we require

$$p(> w_\alpha) = p(> w). \quad (3)$$

Therefore, by construction, there is no trace of the additional correlation with environment in the one-point statistics of  $w_\alpha$ ; it is only by studying how the one-point distribution changes as a function

of environment (the traditional approach), or by measuring spatial correlations (such as mark correlations), that one might discover this correlation. The question is: which approach is more efficient, especially for small  $\alpha$  when  $w_\alpha \approx w + \alpha w\delta$ ?

#### 3.1 Measures of environment

There are many different methods of quantifying the environment, but most of them can be categorized into those that use an FA and others that use nearest-neighbour (NN) finding. A variety of environment measures are analysed in M12, and we use a subset of these, which are briefly described below. Unless stated otherwise, all of the FA and NN overdensities used in this paper are based on redshift-space distances, as they would be in real data. In addition, in all cases the density-defining population (DDP) consists of galaxies brighter than the luminosity threshold,  $M_r < -19$ .

FA measures are often expressed as a local density contrast, determined by counting the number of galaxies within a given radius and taking the ratio with the mean density. The density contrast is typically defined as

$$\delta_g = \frac{N_g - \bar{N}_g}{\bar{N}_g}, \quad (4)$$

where  $N_g$  is the number of galaxies found in the aperture, and  $\bar{N}_g$  is the mean number of galaxies that would be expected in the aperture if the galaxies were randomly distributed. The motivation for using apertures of a particular size is often such that they enclose all of the galaxies within a dark matter halo, while accounting for the effect of redshift-space distortions and redshift uncertainties (Abbas & Sheth 2005; Gallazzi et al. 2009). We use  $1 + \delta_g$ , the overdensity in spherical apertures of radius  $8 \text{ Mpc } h^{-1}$  (Abbas & Sheth 2005; Croton et al. 2005). Cylindrical apertures and annuli (Gallazzi et al. 2009; Wilman, Zibetti & Budavári 2010) yield qualitatively similar results.

NN measures exploit the fact that objects with nearer neighbours tend to be found in denser environments. A value of  $N$  is chosen that specifies the number of neighbours around the point of interest. One can define a projected surface density or a spherical density:

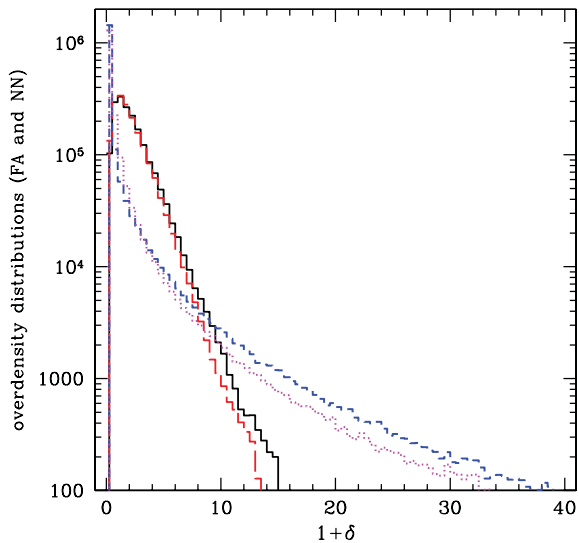
$$\sigma_N = \frac{N}{\pi r_N^2},$$

$$\Sigma_N = \frac{N}{(4/3)\pi r_N^3}, \quad (5)$$

where  $r_N$  is the radius to the  $N$ th nearest neighbour. We use the Baldry et al. (2006) measure, which is an average of  $\log \sigma_N$  for  $N = 4$  and  $5$  with a redshift limit ( $\pm \Delta z c = 1000 \text{ km s}^{-1}$ ) on the DDP, and the  $\Sigma_N$  measure for  $N = 3$  described in M12. In order to use these similarly as the FA overdensities, the NN densities need to be normalized, and we do this by defining  $\delta \equiv (\sigma - \bar{\sigma})/\bar{\sigma}$ , where  $\sigma$  is one of the two density measures described above [i.e.  $\Sigma_3$  or  $(\log \sigma_4 + \log \sigma_5)/2$ ].

Fig. 3 shows the distribution of these  $1 + \delta$  overdensities in the mock catalogue. Note that the FA overdensity distributions have a similar shape (see also de la Torre et al. 2010), as do those of the NN overdensities. For simplicity, throughout the rest of this paper we will usually focus on a single FA environment measure,  $8 \text{ Mpc } h^{-1}$  spheres, and a single NN measure, the combination of  $\sigma_4$  and  $\sigma_5$ , which we will henceforth refer to as  $\sigma_{4,5}$ .

Clearly, the NN measures have longer tails with larger  $1 + \delta$ , consistent with the expectation that they probe smaller scale environments. (That is, if the NN weights trace the environment within



**Figure 3.**  $1 + \delta$  overdensity distributions for FA environment measures [8 Mpc  $h^{-1}$  spheres (black solid histogram) and cylinders (red long dashed histogram)], and for nearest-neighbour (NN) environment measures [ $\sigma_{4,5}$  (blue short dashed histogram) and  $\Sigma_3$  (magenta dotted histogram)], which have longer tails at large  $1 + \delta$ .

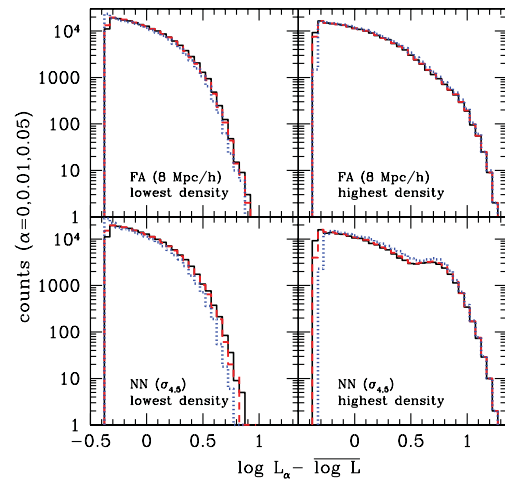
haloes, then we expect them to have  $1 + \delta \sim 200$ , and if the FA weights trace the environment around each halo, they should have  $1 + \delta \sim 1 + \sigma_8$ , where  $\sigma_8$  is the rms variance of the linear density fluctuation field within 8 Mpc  $h^{-1}$  spheres.) As we show below, mark correlations allow us to quantify this expectation. But before doing so, we note that standardizing the distributions by subtracting the mean and dividing by the rms still yields distributions with different shapes.

### 3.2 The traditional approach

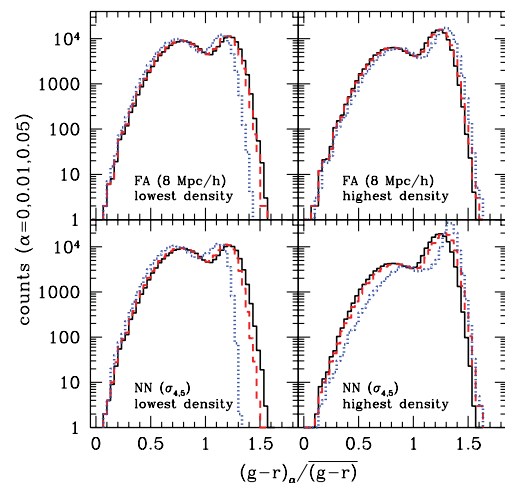
In what follows, we will illustrate our results using these FA and NN overdensities. Specifically, in this section we will insert  $\delta$  in equations (2) and (3) to define  $w_\alpha$  for each galaxy, thus adding an environmental dependence to  $w$  (luminosity or colour), and then we will measure the distribution of (rescaled)  $w_\alpha$  in a number of different overdensity bins.

Fig. 4 shows the results for the luminosity marks. The various histograms in each panel show the overdensity-dependent luminosity distributions,  $p(L_\alpha|\delta)$ , for various choices of  $\alpha$  (0, 0.01 and 0.05). The different panels show different bins in  $\delta$  [lowest and highest 10 per cent;  $p(L_\alpha|\delta)$  of intermediate bins have smaller differences], and there are clearly more luminous galaxies and fewer faint galaxies in dense environments. The question is whether the differences between the  $\alpha = 0$  counts and the others are statistically significant. Obviously, one must be far from  $\delta = 0$  to see a difference; how far is far enough depends on  $\alpha$ . Kolmogorov–Smirnov (KS) tests suggest that large values of  $\alpha$  and/or  $\delta$  far from 0 are required. Since the large  $|\delta|$  tails typically contain a small fraction of the full sample (Fig. 3), this technique, in effect, cannot use the vast majority of the sample to detect the fact that environment matters.

As described earlier in Section 3, to allow for a fair comparison, we have rank ordered and rescaled the luminosities so that they have the same distributions [ $p(>L_\alpha) = p(>L)$ , where  $L_\alpha$  are the luminosities with  $\alpha > 0$ ]. Therefore, the overall luminosity distributions are the same by construction, but at fixed overdensity the rescaled ones (with  $\alpha > 0$ ) may be shifted from the original one ( $\alpha = 0$ ). In



**Figure 4.** Rescaled  $r$ -band luminosity functions for bins of environmental overdensity (upper panels: ranked by 8 Mpc  $h^{-1}$  sphere overdensities; lower panels: ranked by fourth and fifth NN overdensities). Only the distributions for the lowest density (left) and highest density (right) 10 per cent are shown. The black solid, red dashed and blue dotted histograms indicate the distributions for  $\alpha = 0, 0.01$  and  $0.05$ , respectively.

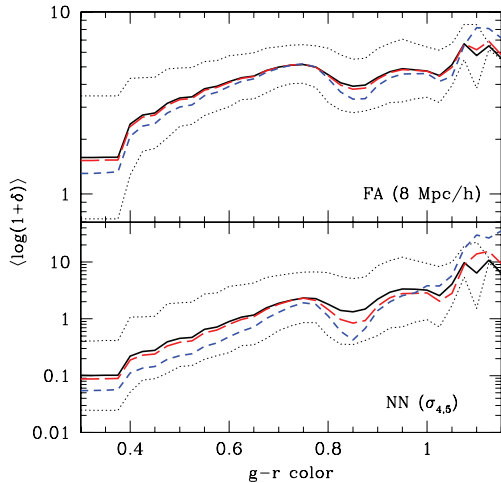


**Figure 5.** Rescaled  $g - r$  colour distributions for bins of environmental overdensity (upper panels: 8 Mpc  $h^{-1}$  spheres; lower panels: fourth and fifth nearest neighbours). Only the distributions for the lowest density (left) and highest density (right) 10 per cent are shown. The black solid, red dashed and blue dotted histograms indicate the distributions for  $\alpha = 0, 0.01$  and  $0.05$ , respectively.

each panel of the figure, we find that the differences between the luminosity distributions  $p(L|\delta)$  and  $p(L_\alpha|\delta)$  appear to be very small, except for  $\alpha = 0$  versus  $\alpha = 0.05$  with the NN overdensity.

Fig. 5 shows a similar analysis of (rescaled)  $(g - r)_\alpha$  colours. Recall that the distributions, when averaged over all  $\delta$ , have been rescaled to be the same. It is evident that there are fewer red galaxies in underdense regions than in very dense ones. In addition, in any overdensity bin, the distributions of  $p(g - r|\delta)$  and  $p[(g - r)_\alpha|\delta]$  are similar, but with significant differences in the red sequence at very low and very high overdensities.

In general, the weak dependence on  $1 + \delta$  appears to produce subtle differences in the luminosity and colour distributions. We quantify the statistical significance of this with KS tests. We find that these in fact yield low KS probabilities, indicating that the



**Figure 6.** The colour–density relation, using  $8 \text{ Mpc } h^{-1}$  sphere overdensities (upper panel) and  $\sigma_{4,5}$  nearest-neighbour overdensities (lower panel). The black, red and blue lines show the running mean of the relation for  $\alpha = 0, 0.01$  and  $0.05$ , respectively. The dotted lines indicate the  $1\sigma$  range between the 16 and 84 percentiles.

distributions  $p(w|\delta)$ , where  $w$  is luminosity or colour, do have statistically significant differences, even for low values of  $\alpha$ . However, the significance depends on the number statistics, and typical SDSS galaxy catalogues are at least 25 times smaller than the mock catalogue used here. When we account for this, at fixed density, we obtain  $P_{\text{KS}} = 0.14$  and  $0.51$  for  $p(L_{\alpha=0.01}|\delta)$  in the lowest and highest density bins for the FA overdensities, respectively, making these distributions statistically indistinguishable, while the corresponding colour distributions are marginally distinguishable. For larger values of  $\alpha$ , and for the NN overdensities, lower probabilities are obtained ( $P_{\text{KS}} < 10^{-3}$ ), indicating statistically significant differences between the luminosity and colour distributions, especially near the peak of the red sequence. Note that for these probability distribution functions, we have normalized by the mean luminosity or colour of the full catalogue; if we normalize by the mean of a given density bin, then  $p(w|\delta)$  become more similar ( $P_{\text{KS}}$  close to unity are obtained, indicating virtually identical distributions), except for  $\alpha = 0.05$  in the highest density bin.

One can also consider the ‘colour–density relation’ or density-dependent red fraction (e.g. Hogg et al. 2003; Balogh et al. 2004; Cooper et al. 2006; Weinmann et al. 2006; Park et al. 2007). We show the  $g-r$  colour–density relation of the mock catalogue in Fig. 6, using the  $8 \text{ Mpc } h^{-1}$  sphere overdensities. Note the dip in the relation between the blue cloud and red sequence. The colour–density relations for the colours modified by  $(1+\delta)^\alpha$  and rescaled following equation (3) are also shown. With  $\alpha > 0$ , the colours have been given an additional environmental dependence, and thus we expect them to have a stronger (i.e. steeper) correlation with overdensity compared to  $\alpha = 0$ . Evidently, the colour–density relation is only slightly steepened for  $\alpha > 0$  and would be difficult to detect, depending on the environment measure used (see also M12). In addition, the relations only cross for the red sequence colours, so the densities in overdense regions must be accurate in order to detect different environmental trends.

The analogous ‘luminosity–density’ relation (not shown) has a similar shape as the luminosity–halo mass relation (e.g. More et al. 2009), except that at luminosities fainter than  $L_*$  (the break in the luminosity function), the relation has increased scatter and is no

longer monotonic. This behaviour is not due to faint satellite galaxies in group/cluster environments, which are outnumbered by faint ‘field’ galaxies; it is instead due to the fact that both FA and NN environment measures do not accurately probe the environments of low-mass haloes (see M12 for details).

### 3.3 Mark correlations

We now turn to the mark correlation measurements as an alternative to the traditional approach to quantifying environmental correlations. The result is shown in Fig. 7, for  $\alpha = 0.01$  and  $0.05$ , using the FA overdensities.

Note that even a dependence as weak as  $(1+\delta)^{0.01}$  results in a significantly stronger signal, while the effect of  $\alpha = 0.05$  is still substantially larger. To clearly demonstrate this, the lower panels of the figures show the ratio of these mark CFs (triangle and square points) to that of the unmodified ( $\alpha = 0$ ) mark CFs (solid curves). The dashed lines show the jack-knife errors<sup>2</sup> of the clustering measurements, indicating that the environmental correlations with  $\alpha = 0.01$  are detectable except for the smallest separations. We have also estimated jack-knife errors of similar measurements of the SDSS catalogue described in Section 2.2 and have found that these are systematically larger than the errors of the mock catalogue’s mark correlations. Nonetheless, we find that a dependence on  $(1+\delta)^{0.01}$  is still detectable at least for colour marks.

Note that the quantitative effect of  $\alpha$  on the mark CFs is similar. In Appendix A, we argue that the  $(1+\delta)^\alpha$  weighting adds a new mark signal which, for  $\alpha \ll 1$ , is proportional to  $2\alpha$ . Therefore, given the mark correlation with a particular value of  $\alpha$  dependence (e.g.  $\alpha = 0.01$ ), we can predict the signal for a different  $\alpha$  (e.g.  $\alpha = 0.05$ ). This expectation is borne out: the predicted mark CFs with  $\alpha = 0.05$  are nearly identical to the measured ones.

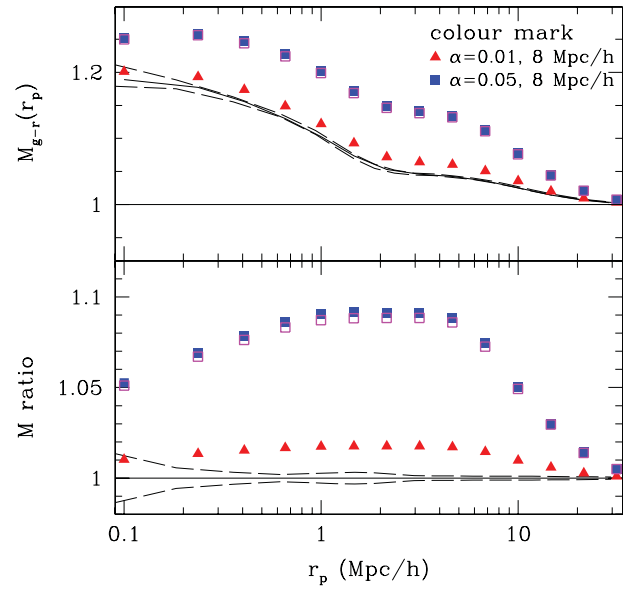
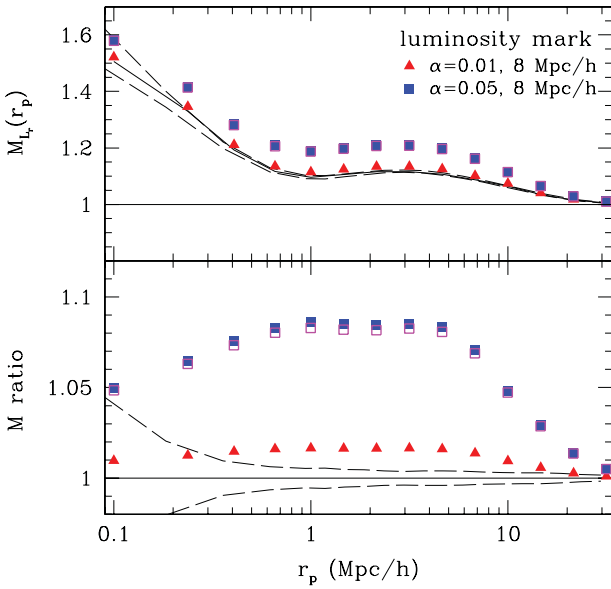
We show analogous mark CFs using NN overdensities in Fig. 8. These overdensities are clearly more sensitive to small-scale environmental correlations than large-scale ones. In addition, the new mark signal for  $\alpha = 0.01$  is larger than for the FA overdensities (triangle points in Fig. 7), which is due to the long tail of the NN overdensities at large  $1+\delta$ . In Section 4, we account for this effect and find that these overdensities are particularly sensitive to environmental correlations on scales smaller than  $600 \text{ kpc } h^{-1}$ . An advantage of the mark correlation approach is that it quantifies the scale dependence of environmental trends, and that it exploits the entire data set (rather than splitting a sample with overdensity cuts, for example).

## 4 RANK-ORDERED MARK CORRELATIONS

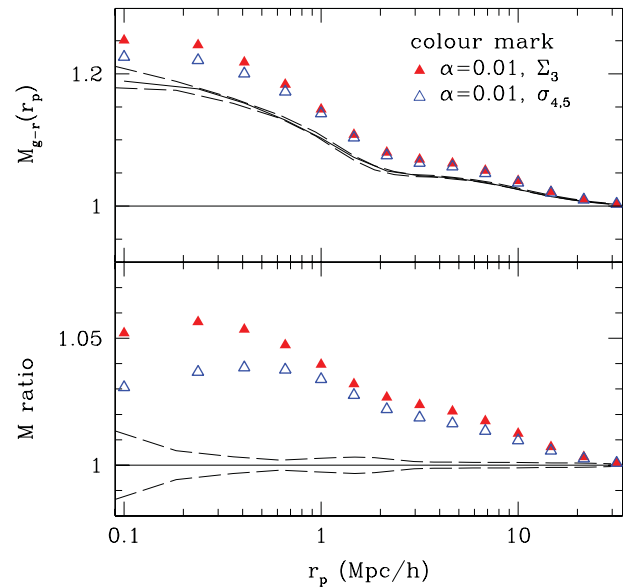
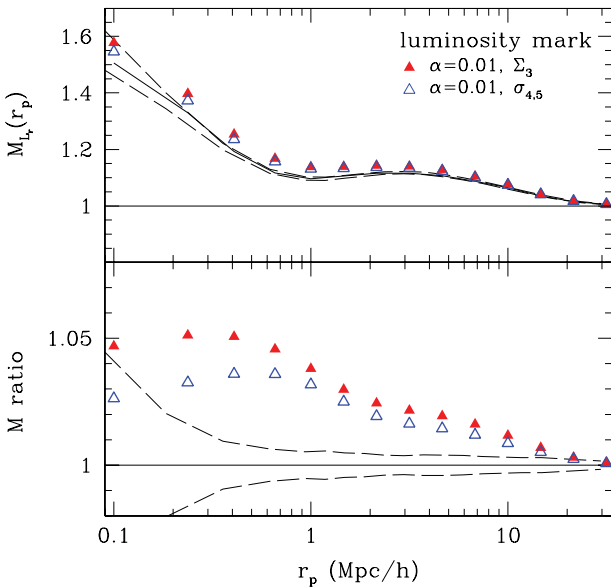
### 4.1 The traditional WW/DD measurement

We now demonstrate how mark CFs quantify the scale dependence of environmental correlations by using the FA and NN estimates of the local density around galaxies,  $1+\delta$ , as the weight (or mark). As these quantities are intended to be direct probes of galaxy environments, rather than indirect ones, such as luminosity or colour, one would expect stronger mark correlation signals than those obtained in the previous section.

<sup>2</sup> Statistical errors are estimated with ‘jack-knife’ resampling, using 27 subsamples of the full mock cube. The variance of the clustering measurements yields the error estimates. [For details, see Zehavi et al. (2005) and Norberg et al. (2009).]



**Figure 7.**  $r$ -band luminosity (left) and  $g - r$  colour (right) mark CFs (upper panels), with additional environmental correlations using FA overdensities:  $(1 + \delta)^\alpha$  with  $\alpha = 0.01$  (red triangles) and  $0.05$  (blue squares). The lower panels show the ratio  $M(\alpha \neq 0)/M(\alpha = 0)$  to more clearly indicate the effect of the additional environmental correlations. The  $\alpha$  dependence of the effect on the mark signal can be easily estimated (open squares; see Appendix A). The dashed lines show the uncertainty of the measurements.



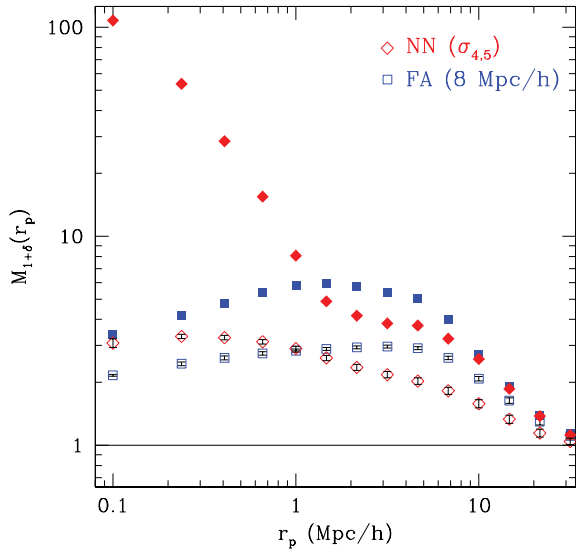
**Figure 8.**  $r$ -band luminosity (left) and  $g - r$  colour (right) mark CFs (upper panels), like Fig. 7, but with additional environmental correlations using NN overdensities: third nearest neighbour (open triangles) and fourth and fifth nearest neighbours (solid triangles). The results are shown for  $\alpha = 0.01$ , and the effect is relatively strong because of the long tail in the NN overdensity distributions (Fig. 3).

The filled symbols in Fig. 9 show the result. The NN weights, using  $\sigma_{4,5}$  overdensities, produce a much stronger signal, especially on small scales. This is consistent with previous work, in which an NN local density was used as a weight (White & Padmanabhan 2009). In Appendix B1, we also illustrate the effect of using the small-scale environment as a weight.

The obvious jump in amplitude at  $r \leq 2h^{-1}$  Mpc for the NN weight is consistent with the expectation that it probes scales within haloes. In addition, the fact that the FA signal reaches a maximum at  $1h^{-1}$  Mpc, which is roughly the scale of a group or cluster, suggests that these are the pairs which are in the densest larger

scale environments. The decrease on smaller scales indicates that an increasing fraction of the closest pairs, which may be low-mass interacting galaxies, are not in particularly dramatic larger scale ( $\sim 8h^{-1}$  Mpc) overdensities.

On the other hand, comparing the FA signal to the NN signal is less straightforward. For example, it is not clear what to make of the fact that the two weights have the same amplitude at scale  $r \geq 10h^{-1}$  Mpc (other than that pairs separated by  $10h^{-1}$  Mpc have weights which are above average by the same factor). This is because the two sets of weights have rather different distributions (Fig. 3). We explore how to remove this in the next section.



**Figure 9.** Mark CFs where FA and NN estimates of the local density ( $8 \text{ Mpc } h^{-1}$  spheres and  $\sigma_{4,5}$ , respectively) were used as marks. The filled symbols show the original measurement, and the open symbols show the result of rank ordering and rescaling to  $[0,1]$  before making the measurement.

#### 4.2 Rank-ordered marks

As stated previously, the strength of mark correlations is affected by the shape of the marks' distributions, which makes it difficult to fairly compare the mark correlations of different marks. To remove the effect of the distributions, we first rank order the marks. We could then have scaled one of the distributions to the other, but this would not allow us to compare either of these with a third mark, for example. Instead, as a more general solution, we perform the rank ordering and then *use the rank itself as the mark*. The open symbols in Fig. 9 show the result of doing this and then remeasuring WW/DD. (In practice, we rank order and then match to a uniform distribution on  $[1, N]$ . In this way, all marks are scaled to the same distribution, so the mark correlation signal can be compared between marks. However, the matching to a uniform distribution is not really necessary.)

The rescaling changes the ratio of the small- to large-scale signal dramatically, particularly for the NN weights, for which the required rescaling is much larger. Evidently, the large  $\delta$  tail in Fig. 3 contributes significantly to the small  $r$  signal; while not unexpected, it is nice to see this confirmation that the NN weights really do correspond to small-scale environments. (We will return to the flatness of the signal shortly.)

On the other hand, note that now the feature at  $1 h^{-1}$  Mpc in the FA signal has gone away. This shows that rescaling comes with a cost, since there may be information in the shape of the distribution, which rank ordering removes.

To illustrate the effect of rank ordering, a model based on cluster and field populations is described in Appendix B2.

#### 4.3 Effect of redshift-space distortions

Throughout this paper we have used environment measures based on redshift-space distances. It is important to consider the effect of redshift-space distortions on these environments, in the context of rank-ordered mark correlations. The non-linear virial motions of galaxies in haloes spread out objects in groups and clusters along the line of sight to produce 'fingers of god' (FOG). These small-

scale distortions can affect how galaxy environments are assessed, as we see using the Park et al. (2007) NN environment measures<sup>3</sup> in Fig. 10. In the left-hand panel, the real-space and  $z$ -space overdensities clearly have considerable scatter between them, and in  $z$ -space there is a deficit of very large overdensities, due to the FOG spreading them out.

This is also seen in the mark CFs (right-hand panel, analogous to the rescaled mark correlations in Fig. 9). The FOG distortions can result in underestimated densities for small-separation pairs  $r_p < 1 \text{ Mpc } h^{-1}$ , but overestimated densities for more widely separated pairs ( $r_p > 1 \text{ Mpc } h^{-1}$ ) where the FOG reach into underdense regions (see also Abbas & Sheth 2007). We have tested this using smaller scale overdensities ( $\Sigma_3$ , used in Fig. 8), which have a similar result but the transition between underestimated and overestimated overdensities occurs at smaller separations. Note that the small-scale downturn at  $r_p < 400 \text{ kpc } h^{-1}$  in the mark correlations (in the right-hand panel of Fig. 10 and in Fig. 9 with the FA overdensity marks) is *not* due to FOG, since it occurs with the real-space overdensities as well; it occurs because these environment measures best probe larger scale environments, as opposed to environment measures with smaller apertures or fewer neighbours.

Note too that even though the  $\delta_r - \delta_z$  mark correlations depart from unity at both small and large scales, rescaling to a uniform distribution nonetheless results in  $1 + \delta$  mark correlations that are approximately consistent with the real-space ones especially at larger separations, demonstrating the utility of the rank-ordered mark correlations.

#### 4.4 Colour and luminosity

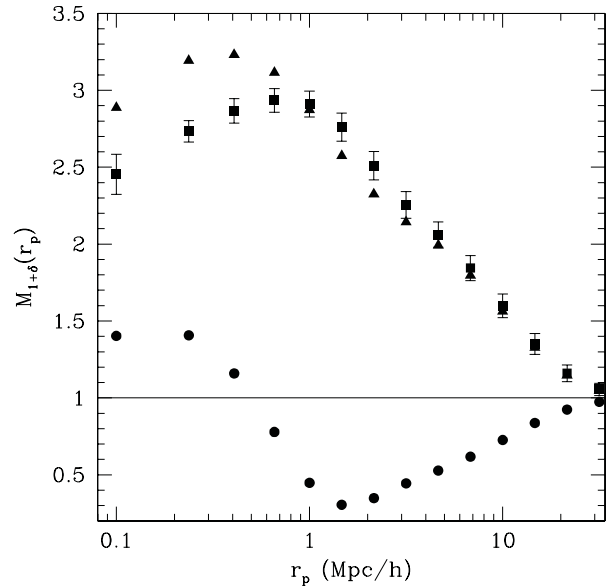
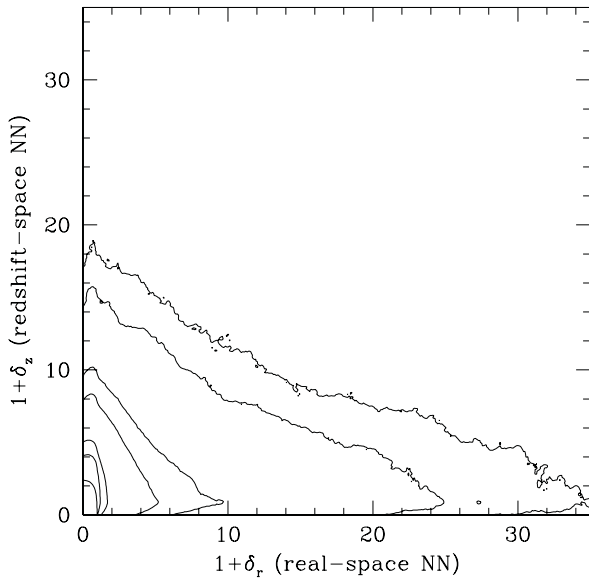
We remarked in the introduction that it is difficult to compare the usual measurements of the colour and luminosity mark correlations with one another. These measurements are shown in Fig. 11. Note that, in contrast to the previous section, here we also compare measurements in the mocks with similar measurements in the SDSS. The agreement between the triangles and crosses shows that the mock catalogues faithfully reproduce the luminosity and colour dependence of clustering; since these WW/DD signals were not used to construct the mocks, they represent non-trivial tests of the mock-making algorithm.

While this is reassuring, a puzzle lies in the fact that we naively expect colour to correlate more strongly with environment than does luminosity (e.g. Butcher & Oemler 1984; Bower et al. 1998; Diaferio et al. 1999; Blanton et al. 2005), but the WW/DD signals do not show this: the amplitude of the luminosity mark correlations is stronger than that of colour. This is primarily because the two weights have very different distributions (Figs 4 and 5); the distribution of luminosity is much broader, such that bright galaxies have  $L \gg \bar{L}$ . Fig. 12 shows the result of rank ordering and using the rank as a mark instead. Unlike the previous figure, now colour clearly produces the stronger signal. It is also worth noting that the error bars are much larger for the  $L$ -weighted signal, showing that a large range of  $L$  ranks contributes at each  $r$ ; this range is much narrower for the  $g - r$  colour.

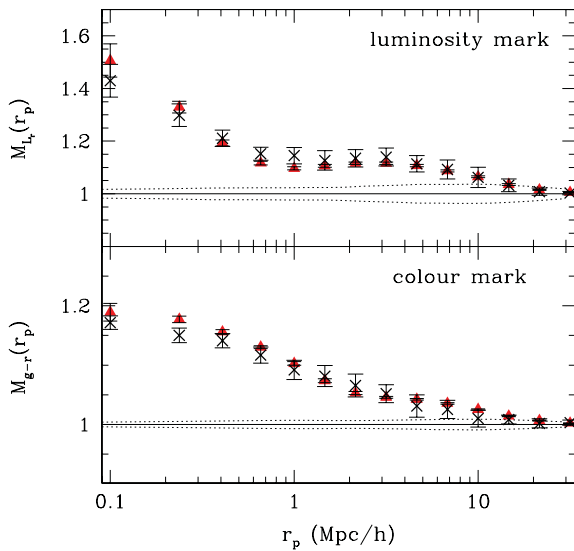
Finally, our rank-ordering procedure allows us to compare these measurements with those in Section 4.2. Comparing the luminosity

<sup>3</sup> Park et al. (2007) local densities are estimated by using 20 nearest-neighbour galaxies, with the galaxies centrally weighted by a spherical adaptive smoothing kernel (and hence not equivalent to  $\Sigma_{20}$  as defined in equation 5).





**Figure 10.** Left: contour plot of real-space versus  $z$ -space Park et al. (2007) NN overdensities. Right: rank-ordered and rescaled mark CFs of these real-space and  $z$ -space overdensities (triangle and square points, respectively), analogous to the mark correlations in Fig. 9. The  $\delta_r - \delta_z$  mark correlations (circles) are also shown.



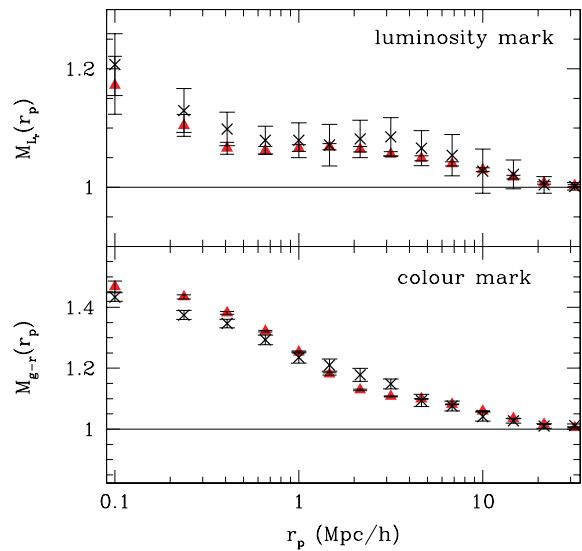
**Figure 11.**  $r$ -band luminosity and  $g - r$  colour mark CFs for galaxies with  $M_r < -19$  in the SDSS DR7 (crosses) and in the mock catalogue (triangles). The error bars show jack-knife errors on the measurements, and the dotted lines show the scatter from randomly scrambling the marks (see the text for details).

and colour mark CFs (Fig. 11) to the local density mark correlations (Fig. 9) shows that both luminosity and colour produce significantly weaker signals.

## 5 CONCLUSIONS AND DISCUSSION

Our key results can be summarized as follows.

(i) Mark CFs are particularly sensitive to environmental correlations, specifically when using enhanced weights of  $(1 + \delta)^\alpha$  with  $\alpha > 0.01$ , though this sensitivity depends on the environment measure, scale and the mark's uncertainty.



**Figure 12.** Same as the previous figure, but now the weights have all been rank ordered and then scaled to a uniform distribution. Although the qualitative trends are the same as in the previous figure, now the signal is stronger in the bottom panel (i.e. when weighting by colour), indicating that the colour–density correlation is stronger than the luminosity–density one.

(ii) Small environmental correlations are difficult to detect with more traditional methods, as highlighted by the (lack of) variation of most of the mark distributions at fixed overdensity, and in the colour–density relation.

(iii) Rank ordering the marks and then using the rank as the weight provides a simple way to compare results for different marks, because it removes any dependence on the marks' distributions.

The analysis in this paper highlights the advantages (and disadvantages) of mark clustering statistics. The fact that they are sensitive to weak environmental correlations that traditional methods find difficult to detect demonstrates their utility. Mark statistics

are particularly useful for identifying and quantifying environmental trends. This owes to the fact that the statistics of entire samples can be folded together, producing clearer correlations than simply binning environments or splitting galaxies into ‘field’ and ‘cluster’ subsamples, for example. Nonetheless, one cannot determine from these trends alone which galaxies occupy which environments; more information is needed (e.g. from halo models of galaxy clustering) in order to associate particular galaxies with environments of particular halo mass or overdensity (Skibba et al. 2006; Skibba & Sheth 2009).

In contrast, methods that characterize individual local galaxy environments can associate galaxies with overdensities, though they too have strengths and weaknesses. FA and nearest-neighbour overdensities are sensitive to inter- and intrahalo environments, respectively, consistent with the findings of M12 and Haas, Schaye & Jeason-Daniel (2012). We showed this with the scale-dependent mark clustering measurements, which overlapped at  $r_p \sim 600 \text{ kpc } h^{-1}$ , within the ‘one-halo term’; FAs, if sufficiently large, can encompass entire haloes as well as some of the surrounding regions. Nonetheless, the interpretation of environmental trends can be difficult and depends crucially on how the overdensities are measured and on the DDP.

One can also interpret our results in terms of central and satellite galaxies in haloes. Since satellite luminosities, colours and stellar masses depend only weakly on halo mass, and hence only weakly on the environment (e.g. Skibba, Sheth & Martino 2007; van den Bosch et al. 2008; Skibba 2009; see also Neistein et al. 2011; De Lucia et al. 2012), the majority of the environmental correlations that we detect are due to the central galaxies. For example, the colours of central galaxies are strongly halo mass dependent, and this is clearly shown by the colour mark correlations, which are especially sensitive to the dependence on overdensity.

Finally, we note that rank-ordered mark CFs are applicable to any comparative analysis of environmental trends involving large catalogues of objects in surveys or simulations with sufficiently accurate distances and marks, and are useful for testing or constraining models. Rank-ordered mark correlations could be useful for quantifying and comparing measures of ‘halo assembly bias’ (e.g. Sheth & Tormen 2004; Harker et al. 2006; Wechsler et al. 2006; Croton, Gao & White 2007; Croft et al. 2012), such that halo formation time, concentration or occupation is weakly correlated with the environment at fixed mass. These statistics could also be applied to tests of ‘halo abundance matching’ (Conroy, Wechsler & Kravtsov 2006; Vale & Ostriker 2006; Neistein et al. 2011; Trujillo-Gomez et al. 2011; Kang et al. 2012) methods, in which central/satellite galaxies and dark matter haloes/subhaloes are rank ordered by their luminosities, masses or circular velocities, and their cumulative number densities are matched.

## ACKNOWLEDGMENTS

RAS is supported in part by the NASA *Herschel* Science Center, JPL contract #1350371, and in part by the NSF grant AST-1055081. RKS is supported in part by NSF 0908241 and NASA NNX11A125G, and is grateful to the GEPI and LUTH groups at Meudon Observatory for their hospitality during the summers of 2011 and 2012. DC acknowledges receipt of a QEII Fellowship by the Australian Research Council. We thank Idit Zehavi for valuable discussions about the SDSS clustering measurements. We also thank the anonymous referee for insightful comments that helped to improve the paper.

## REFERENCES

- Abazajian K. et al., 2009, *ApJS*, 182, 543  
 Abbas U., Sheth R. K., 2005, *MNRAS*, 364, 1327  
 Abbas U., Sheth R. K., 2007, *MNRAS*, 378, 641  
 Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, *MNRAS*, 373, 469  
 Balogh M. L., Baldry I. K., Nichol R., Miller C., Bower R., Glazebrook K., 2004, *ApJ*, 615, L101  
 Beisbart C., Kerscher M., 2000, *ApJ*, 545, 6  
 Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2002, *ApJ*, 629, 143  
 Bower R. G., Kodama T., Terlevich A., 1998, *MNRAS*, 299, 1193  
 Butcher H., Oemler Jr. A., 1984, *ApJ*, 285, 426  
 Coil A. et al., 2008, *ApJ*, 672, 153  
 Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201  
 Cooper M. C. et al., 2006, *MNRAS*, 370, 198  
 Croft R., Di Matteo T., Khandai N., Springel V., Jana A., Gardner J., 2012, *MNRAS*, 425, 2766  
 Croton D. J. et al., 2005, *MNRAS*, 356, 1155  
 Croton D. J., Gao L., White S. D. M., 2007, *MNRAS*, 374, 1303  
 de la Torre S. et al., 2010, *MNRAS*, 409, 867  
 De Lucia G., Weinmann S., Poggianti B. M., Aragón-Salamanca A., Zaritsky D., 2012, *MNRAS*, 423, 1277  
 Diaferio A., Kauffmann G., Colberg J., White S. D. M., 1999, *MNRAS*, 307, 537  
 Gallazzi A. et al., 2009, *ApJ*, 690, 1883  
 Gardner J. P., Connolly A., McBride C., 2007, preprint (astro-ph/0709.1967)  
 Haas M. R., Schaye J., Jeason-Daniel A., 2012, *MNRAS*, 419, 2133  
 Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, *MNRAS*, 367, 1039  
 Hogg D. W. et al., 2003, *ApJ*, 585, L5  
 Kang X., Li M., Lin W. P., Elahi P. J., 2012, *MNRAS*, 422, 804  
 Li C., Kauffmann G., Jing Y. P., White S. D. M., Börner G., Cheng F. Z., 2006, *MNRAS*, 368, 21  
 Martínez V. J., Arnalte-Mur P., Stoyan D., 2010, *A&A*, 513, 22  
 More S., van den Bosch F. C., Cacciato M., Mo H. J., Yang X., Li R., 2009, *MNRAS*, 392, 801  
 Muldrew S. I. et al., 2012, *MNRAS*, 419, 2670 (M12)  
 Neistein E., Li C., Khochfar S., Weinmann S. M., Shankar F., Boylan-Kolchin M., 2011, *MNRAS*, 416, 1486  
 Norberg P. et al., 2002, *MNRAS*, 332, 827  
 Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19  
 Paranjape A., Sheth R. K., 2012, *MNRAS*, 423, 1845  
 Park C., Choi Y.-Y., Vogeley M. S., Gott J. R., III, Blanton M. R., 2007, *ApJ*, 658, 898  
 Sheth R. K., 2005, *MNRAS*, 364, 796  
 Sheth R. K., Tormen G., 2004, *MNRAS*, 350, 1385  
 Sheth R. K., Connolly A. J., Skibba R., 2005, preprint (astro-ph/0511773)  
 Sheth R. K., Jimenez R., Panter B., Heavens A. F., 2006, *ApJ*, 650, L25  
 Skibba R. A., 2009, *MNRAS*, 392, 1467  
 Skibba R. A., Sheth R. K., 2009, *MNRAS*, 392, 1080  
 Skibba R. A., Sheth R. K., Connolly A. J., Scranton R., 2006, *MNRAS*, 369, 68  
 Skibba R. A., Sheth R. K., Martino M. C., 2007, *MNRAS*, 382, 1940  
 Springel V. et al., 2005, *Nat*, 435, 629  
 Stoyan D., Stoyan H., 1994, *Fractals, Random Shapes, and Point Fields*. Wiley, Chichester  
 Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, *ApJ*, 742, 16  
 Vale A., Ostriker J. P., 2006, *MNRAS*, 371, 1173  
 van den Bosch F. C., Pasquali A., Yang X., Mo H. J., Weinmann S., McIntosh D. H., Aquino D., 2008, preprint (arXiv:0805.0002)  
 Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, *ApJ*, 652, 71  
 Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, *MNRAS*, 366, 2

White M., Padmanabhan N., 2009, MNRAS, 395, 2381  
 Wilman D. J., Zibetti S., Budavári T., 2010, MNRAS, 406, 1701  
 York D. G. et al., 2000, AJ, 120, 1579  
 Zehavi I. et al., 2005, ApJ, 630, 1  
 Zehavi I. et al., 2011, ApJ, 736, 59

## APPENDIX A: EFFECT OF THE ADDITIONAL ENVIRONMENTAL CORRELATION

We provide an estimate of the effect of the added environmental correlation (equation 2) on the marked CF.

As described in Section 4, the statistic  $M(r)$  can be approximated by the simple pair count ratio WW/DD, where WW is the sum over all pairs with separation  $r$ , weighting each member of the pair by its mark, and DD is the total number of such pairs. To use a specific example in the paper (see Section 3), suppose that rather than weighting galaxies by their luminosity  $L$ , we modify the weight by adding a small dependence on the  $8 \text{ Mpc } h^{-1}$  overdensity, which we will denote as  $1 + \delta_8$ . In this case, the modified weight can be expressed as

$$w = L(1 + \delta_8)^\alpha \sim L(1 + \alpha\delta_8), \quad (\text{A1})$$

where  $\alpha$  is small. For the mark CF, we will normalize by the mean mark, and the mean of the above expression is simply

$$\langle w \rangle \sim \langle L(1 + \alpha\delta_8) \rangle \sim \langle L \rangle. \quad (\text{A2})$$

Therefore, for a pair of galaxies  $i$  and  $j$  at separation  $r$ , we have

$$\text{WW}(r) = \frac{\langle L_i(1 + \delta_i)(1 + \alpha\delta_{8,i}) L_j(1 + \delta_j)(1 + \alpha\delta_{8,j}) \rangle}{\langle w \rangle^2}. \quad (\text{A3})$$

If  $L$  is not significantly correlated with density (which is not quite true, because  $L$  is correlated with  $M_{\text{halo}}$  and hence  $\delta$ ), then this becomes

$$\text{WW}(r) = \langle [(1 + \delta)(1 + \alpha\delta_8)]_i [(1 + \delta)(1 + \alpha\delta_8)]_j \rangle. \quad (\text{A4})$$

Keeping the lowest order in  $\alpha$ , this can be expanded as follows:

$$\begin{aligned} \text{WW}(r) &\sim \langle (1 + \delta_i)(1 + \delta_j)(1 + \alpha\delta_{8,i} + \alpha\delta_{8,j}) \rangle \\ &\sim \langle 1 + \delta_i + \delta_j + \delta_i\delta_j + \alpha\delta_{8,i}(1 + \delta_i + \delta_j \\ &\quad + \delta_i\delta_j) + \alpha\delta_{8,j}(1 + \delta_i + \delta_j + \delta_i\delta_j) \rangle \\ &\sim 1 + \langle \delta_i\delta_j \rangle + 2\alpha(\langle \delta_{8,i}\delta_i \rangle + \langle \delta_{8,i}\delta_j \rangle + \langle \delta_{8,i}\delta_i\delta_j \rangle). \end{aligned} \quad (\text{A5})$$

Since  $\text{DD} = 1 + \langle \delta_i\delta_j \rangle$ , we have

$$\frac{\text{WW}}{\text{DD}} \sim 1 + 2\alpha \frac{\langle \delta_{8,i}\delta_i \rangle + \langle \delta_{8,i}\delta_j \rangle + \langle \delta_{8,i}\delta_i\delta_j \rangle}{\text{DD}}, \quad (\text{A6})$$

which shows that the new environmental correlation that we introduced in the text produces a new signal proportional to  $2\alpha$ . The approximations here appear to be accurate, as the mark CFs with  $\alpha = 0.05$  can be predicted from those with  $\alpha = 0.01$ , and vice versa, at all separations  $r$  (see e.g. Fig. 7).

## APPENDIX B: SMALL-SCALE ENVIRONMENT AS A WEIGHT

### B1 A toy model

We will use a simple toy model to illustrate the effect of using the environment as a weight.

Suppose that all the mass is in haloes which all have the same mass  $m$  distributed around each halo centre according to

$$\begin{aligned} \frac{\rho(r)}{\bar{\rho}} &= \frac{m}{\bar{\rho}} \frac{\exp(-r^2/2R_v^2)}{(2\pi R_v^2)^{3/2}} \\ &= \Delta_v e^{-r^2/2R_v^2}, \end{aligned} \quad (\text{B1})$$

where the final expression defines  $\Delta_v$ , the central density [because  $(2\pi)^{3/2} R_v^3$  is the volume of the profile].

Then the unweighted CF is

$$\begin{aligned} \xi(r) &= \frac{\bar{\rho}}{m} \frac{m^2}{\bar{\rho}^2} \frac{\exp(-r^2/4R_v^2)}{(2\pi 2R_v^2)^{3/2}} \\ &= \frac{\Delta_v}{2^{3/2}} e^{-r^2/4R_v^2}, \end{aligned} \quad (\text{B2})$$

where the first factor of  $\bar{\rho}/m$  is the number density of haloes (i.e. all the mass is in haloes of mass  $m$ ).

If we model the weight as the local value of the density smoothed with an FA of scale  $s$ , then the weight associated with a distance  $r$  from the halo centre is

$$w(r) = e^{-r^2/2(R_v^2 + s^2)}. \quad (\text{B3})$$

If we define

$$R_s^2 \equiv R_v^2(R_v^2 + s^2)/(2R_v^2 + s^2), \quad (\text{B4})$$

then the mean weight is

$$\bar{w} = 4\pi \int dr r^2 \rho(r) w(r)/m = (R_s/R_v)^3. \quad (\text{B5})$$

Therefore, the normalized weighted CF is

$$\xi_w(r) = \frac{\bar{\rho}}{m} \frac{m^2}{\bar{\rho}^2} \frac{e^{-r^2/4R_s^2}}{(2\pi 2R_s^2)^{3/2}}. \quad (\text{B6})$$

To see what this implies for WW/DD, suppose that  $s < R_v$ . Then  $R_s^2 \rightarrow R_v^2/2$ . On small scales,  $\text{WW/DD} = (1 + \xi_w)/(1 + \xi) \approx \xi_w/\xi$  because we are interested in the case in which  $\Delta_v \gg 1$ . In this limit,  $\text{WW/DD} \approx 2^{3/2} \exp(-r^2/2R_v^2)/\exp(-r^2/4R_v^2)$ . This shows that WW/DD has the same shape as  $\xi$  itself, and the small-scale amplitude is  $2^{3/2}$  times the unweighted one. The amplitude is  $2^{3/2}$  because the assumed Gaussian profile (equation B1) is relatively flat; a centrally cusped profile (such as a Navarro, Frenk & White (1996) one) will produce a stronger signal. The small-scale shape of WW/DD should not come as a surprise: when  $s \ll R_v$ , then  $\xi_w$  is like the convolution of  $\rho^2$  with itself, making  $\xi_w \propto \xi^2$ . More generally,  $\text{WW/DD} \propto \xi^{1/(1+s^2/R_v^2)}$ .

### B2 A model in terms of cluster and field populations

To gain intuition about the effect of rank ordering, first note that for a list of length  $N$  marks, the mean mark is  $[N(N+1)/2]/N = (N+1)/2$ , so normalizing is particularly simple. Now, suppose that the distribution of environments is bimodal, with one population associated with ‘close’ pairs (separations less than some  $R_c$ ) in dense regions and another with underdense ones. Suppose that ‘cluster’ galaxies have close neighbours but ‘field’ galaxies do not (i.e. they are like hard spheres). Then the total clustering signal is  $n_i^2(1 + \xi_{ii}) = n_c^2(1 + \xi_{cc}) + n_f^2(1 + \xi_{ff}) + 2n_c n_f(1 + \xi_{cf})$ , where

$n_t \equiv n_c + n_f$  and the mean mark is  $\bar{w} \equiv (n_c w_c + n_f w_f)/n_t$ . Therefore,

$$\frac{WW}{DD} = \frac{(n_c w_c)^2 (1 + \xi_{cc}) + (n_f w_f)^2 (1 + \xi_{ff}) + \frac{2n_c w_c n_f w_f (1 + \xi_{cf})}{\bar{w}^2 n_t^2 (1 + \xi_{tt})}}{(B7)}$$

If we have rank ordered the marks, then  $n_t \bar{w} = N(N + 1)/2$ ,  $n_f w_f = N_f(N_f + 1)/2$  and  $n_c w_c = n_t \bar{w} - n_f w_f$ . On scales smaller than  $R_c$  we know that both  $\xi_{ff}$  and  $\xi_{cf}$  equal  $-1$ , making

$$\frac{WW}{DD} = \frac{n_c^2 w_c^2 (1 + \xi_{cc})}{n_t^2 \bar{w}^2 (1 + \xi_{cc})} = \left( \frac{w_c}{\bar{w}} \right)^2 = [1 + N_f/(N + 1)]^2. \quad (B8)$$

Thus, the small-scale signal is a measure of the field fraction  $N_f/N$ , but note that it cannot exceed 4.

If we were to interpret our measured value of  $WW/DD = 3$  in Fig. 9 in these terms, we would infer a field fraction of about 70 per

cent; it is interesting that this implies a cluster fraction (30 per cent) that is close to the satellite fraction usually quoted in halo model analyses of galaxy clustering (e.g. Zehavi et al. 2005; van den Bosch et al. 2007) and the satellite fraction of the mock catalogue used in this paper (Section 2.1). If we assume that on intermediate scales  $\xi_{ff}$  and  $\xi_{cf}$  both are approximately equal to zero, then

$$\frac{WW}{DD} = \frac{1 + (n_c/n_t)^2 (w_c/\bar{w})^2 \xi_{cc}}{1 + (n_c/n_t)^2 \xi_{cc}}. \quad (B9)$$

In this approximation, the scale dependence of  $WW/DD$  codes information about the cluster or field fraction and the CF of the cluster population. In the  $\xi(r) \gg 1$  limit, it smoothly asymptotes to the previous expression.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.