Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **425,** 1527–1548 (2012) doi:10.1111/j.1365-2966.2012.21434.x

Galaxy And Mass Assembly (GAMA): colour- and luminosity-dependent clustering from calibrated photometric redshifts

L. Christodoulou,[1⋆] C. Eminian,[1] J. Loveday,[1] P. Norberg,[2] I. K. Baldry,[3] P. D. Hurley,[1] S. P. Driver,[4,5] S. P. Bamford,[6] A. M. Hopkins,[7] J. Liske,[8] J. A. Peacock,[9] J. Bland-Hawthorn,[10] S. Brough,[7] E. Cameron,[11] C. J. Conselice,[6] S. M. Croom,[10] C. S. Frenk,[2] M. Gunawardhana,[10] D. H. Jones,[12] L. S. Kelvin,[4,5] K. Kuijken,[13] R. C. Nichol,[14] H. Parkinson,[9] K. A. Pimbblet,[12] C. C. Popescu,[15] M. Prescott,[3] A. S. G. Robotham,[4,5] R. G. Sharp,[16] W. J. Sutherland,[17] E. N. Taylor,[18] D. Thomas,[14] R. J. Tuffs,[19] E. van Kampen[8] and D. Wijesinghe[10]

[1]*Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH*
[2]*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE*
[3]*Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Egerton Wharf, Birkenhead CH41 1LD*
[4]*ICRAR (International Centre for Radio Astronomy Research), University of Western Australia, Crawley, WA 6009, Australia*
[5]*SUPA (Scottish Universities Physics Alliance), School of Physics & Astronomy, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS*
[6]*Centre for Astronomy and Particle Theory, University of Nottingham, University Park, Nottingham NG7 2RD*
[7]*Australian Astronomical Observatory, PO Box 296, Epping, NSW 1710, Australia*
[8]*European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany*
[9]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ*
[10]*Sydney Institute for Astronomy, School of Physics, University of Sydney, NSW 2006, Australia*
[11]*Department of Physics, Swiss Federal Institute of Technology (ETH-Zürich), 8093 Zürich, Switzerland*
[12]*School of Physics, Monash University, Clayton, Victoria 3800, Australia*
[13]*Leiden University, PO Box 9500, 2300 RA Leiden, the Netherlands*
[14]*Institute of Cosmology and Gravitation (ICG), University of Portsmouth, Dennis Sciama Building, Burnaby Road, Portsmouth PO1 3FX*
[15]*Jeremiah Horrocks Institute, University of Central Lancashire, Preston PR1 2HE*
[16]*Research School of Astronomy & Astrophysics, Mount Stromlo Observatory, Weston Creek, ACT 2611, Australia*
[17]*Astronomy Unit, Queen Mary University London, Mile End Rd, London E1 4NS*
[18]*School of Physics, University of Melbourne, Victoria 3010, Australia*
[19]*Max Planck Institute for Nuclear Physics (MPIK), Saupfercheckweg 1, 69117 Heidelberg, Germany*

Accepted 2012 May 31. Received 2012 May 30; in original form 2011 October 4

ABSTRACT

We measure the two-point angular correlation function of a sample of 4289 223 galaxies with $r < 19.4$ mag from the Sloan Digital Sky Survey (SDSS) as a function of photometric redshift, absolute magnitude and colour down to $M_r - 5\log h = -14$ mag. Photometric redshifts are estimated from *ugriz* model magnitudes and two Petrosian radii using the artificial neural network package ANNz, taking advantage of the Galaxy And Mass Assembly (GAMA) spectroscopic sample as our training set. These photometric redshifts are then used to determine absolute magnitudes and colours. For all our samples, we estimate the underlying redshift and absolute magnitude distributions using Monte Carlo resampling. These redshift distributions are used in Limber's equation to obtain spatial correlation function parameters from power-law fits to the angular correlation function. We confirm an increase in clustering strength for sub-L^* red galaxies compared with $\sim L^*$ red galaxies at small scales in all redshift bins, whereas for the blue population the correlation length is almost independent of luminosity for $\sim L^*$ galaxies and fainter. A linear relation between relative bias and log luminosity is found

⋆E-mail: L.Christodoulou@sussex.ac.uk

to hold down to luminosities $L \sim 0.03L^*$. We find that the redshift dependence of the bias of the L^* population can be described by the passive evolution model of Tegmark & Peebles. A visual inspection of a random sample from our $r < 19.4$ sample of SDSS galaxies reveals that about 10 per cent are spurious, with a higher contamination rate towards very faint absolute magnitudes due to over-deblended nearby galaxies. We correct for this contamination in our clustering analysis.

Key words: techniques: photometric – surveys – galaxies: distances and redshifts – galaxies: statistics – large-scale structure of Universe.

1 INTRODUCTION

Measurement of galaxy clustering is an important cosmological tool to aid our understanding of the formation and evolution of galaxies at different epochs. The dependence of galaxy clustering on properties such as morphology, colour, luminosity or spectral type has been established over many decades. Elliptical galaxies or galaxies with red colours, which both trace an old stellar population, are known to be more clustered than spiral galaxies (e.g. Davis & Geller 1976; Dressler 1980; Postman & Geller 1984; Loveday et al. 1995; Guzzo et al. 1997; Goto et al. 2003). Recent large galaxy surveys have allowed the investigation of galaxy clustering as a function of both colour and luminosity (Norberg et al. 2002; Budavári et al. 2003; Zehavi et al. 2005; Wang et al. 2007; McCracken et al. 2008; Zehavi et al. 2011). Among the red population, a strong luminosity dependence has been observed whereby luminous galaxies are more clustered because they reside in denser environments.

The galaxy luminosity function shows an increasing faint-end density to at least as faint as $M_r - 5\log h = -12$ mag (Blanton et al. 2005a; Loveday et al. 2012); thus intrinsically faint galaxies represent the majority of the galaxies in the Universe. These galaxies with luminosity $L \ll L^*$ have low stellar mass and are mostly dwarf galaxies with ongoing star formation. However, because most wide-field spectroscopic surveys can only probe luminous galaxies over large volumes, this population is often under-represented. Previous clustering analyses have revealed that intrinsically faint galaxies have different properties from luminous ones. A striking difference appears between galaxy colours in this regime: while faint blue galaxies seem to cluster on a scale almost independent of luminosity, the faint red population is shown to be very sensitive to luminosity (Norberg et al. 2001, 2002; Zehavi et al. 2002; Hogg et al. 2003; Zehavi et al. 2005; Swanson et al. 2008a; Zehavi et al. 2011; Ross, Tojeiro & Percival 2011b). As found by Zehavi et al. (2005), this trend is naturally explained by the halo occupation distribution framework. In this picture, the faint red population corresponds to red satellite galaxies, which are located in high-mass haloes with red central galaxies and are therefore strongly clustered. Recently, Ross et al. (2011b) compiled from the literature bias measurements for red galaxies over a wide range of luminosities for both spectroscopic and photometric data. They showed that the bias measurements of the faint red population are strongly affected by non-linear effects and thus on the physical scales over which they are measured. They conclude that red galaxies with $M_r > -19$ mag are biased similarly to or less than red galaxies of intermediate luminosity.

In this work, we make use of photometric redshifts to probe the regime of intrinsically faint galaxies. Our sample is composed of Sloan Digital Sky Survey (SDSS) galaxies with r-band Petrosian magnitude $r_{\rm petro} < 19.4$. As we have an ideal training set for this sample, thanks to the Galaxy And Mass Assembly (GAMA) survey (Driver et al. 2011), we use the artificial neural network package ANNZ (Collister & Lahav 2004) to predict photometric redshifts. We then calculate the angular two-point correlation function as a function of absolute magnitude and colour. The correlation length of each sample is computed through the inversion of Limber's equation, using Monte Carlo resampling for modelling the underlying redshift distribution. Recently, Zehavi et al. (2011) presented the clustering properties of the DR7 spectroscopic sample from SDSS. They extracted a sample of ∼700 000 galaxies with redshifts to $r \leq 17.6$ mag, covering an area of 8000 deg^2. Their study of the luminosity and colour dependence uses power-law fits to the projected correlation function. Our study is complementary to theirs, since we are using calibrated photo-z values of fainter galaxies from the same SDSS imaging catalogue. We use similar luminosity bins to Zehavi et al., with the addition of a fainter luminosity bin $-17 < M_r - 5\log h < -14$.

Small-scale ($r < 0.1\,h^{-1}$ Mpc) galaxy clustering provides additional tests of the fundamental problem of how galaxies trace dark matter. Previous studies have used SDSS data and the projected correlation function to study the clustering of galaxies at the smallest scales possible (Masjedi et al. 2006), using extensive modelling to account for the fibre constraint in SDSS spectroscopic data. The interpretation of these results offers unique tests of how galaxies trace dark matter and the inner structure of dark matter haloes (Watson et al. 2012). Motivated by these studies, we present measurements of the angular correlation function down to scales of $\theta \approx 0.^{\circ}005$. We work solely with the angular correlation function and pay particular attention to systematics errors and the quality of the data.

On the other hand, on sufficiently large scales ($r > 60\,h^{-1}$ Mpc), it is expected that the galaxy density field evolves linearly following the evolution of the dark matter density field (Tegmark et al. 2006). However, it is less clear whether this assumption holds on smaller scales, where complicated physics of galaxy formation and evolution dominate. In the absence of sufficient spectroscopic data to study the evolution of clustering comprehensively, Ross, Percival & Brunner (2010) used SDSS photometric redshifts to extract a volume-limited sample with $M_r < -21.2$ and $z_{\rm phot} < 0.4$. Their analysis revealed significant deviations from the passive evolution model of Tegmark & Peebles (1998). Here we perform a similar analysis, again using photometric redshifts, for the L^* population.

This paper is organized as follows. In Section 2, we introduce the statistical quantities used to calculate the clustering of galaxies, with an emphasis on the angular correlation function. In Section 3 we present our data for this study and the method for estimating the clustering errors. In Section 4 we describe the procedure that we followed in order to obtain the photometric redshifts. We then investigate the clustering of our photometric sample, containing a large number of intrinsically faint galaxies, in Section 5. In Section 6 we present bias measurements as functions of colour, luminosity and redshift. Our findings are summarized in Section 7. In Appendix A we show how we extracted our initial catalogue from the SDSS DR7

data base and finally in Appendix B we describe in some detail the tests performed to assess systematic errors.

Throughout we assume a standard flat ΛCDM cosmology, with $\Omega_{\rm m} = 0.30$, $\Omega_\Lambda = 0.70$ and $H_0 = 100\,h\,{\rm km\,s^{-1}\,Mpc^{-1}}$.

2 THE TWO-POINT ANGULAR CORRELATION FUNCTION

2.1 Definition

The simplest way to measure galaxy clustering on the sky is via the two-point correlation function, $w(\theta)$, which gives the excess probability of finding two galaxies at an angular separation θ compared with a random Poisson distribution (Peebles 1980, Section 31):

$$dP = \bar{n}^2[1 + w(\theta)]\,d\Omega_1\,d\Omega_2, \quad (1)$$

where dP is the joint probability of finding galaxies in solid angles dΩ_1 and dΩ_2 separated by θ, and $\bar{n}$ is the mean number of objects per solid angle. If $w(\theta) = 0$, then the galaxies are unclustered and randomly distributed at this separation. We consider various estimators for $w(\theta)$ in Section 2.3.

2.2 Power-law approximation

Over small angular separations, the two-point correlation function can be approximated by a power law:

$$w(\theta) = A_w\theta^{1-\gamma}, \quad (2)$$

where A_w is the amplitude. The amplitude of the correlation function of a galaxy population is reduced as we go to higher redshifts, because equal angular separations trace larger spatial separations for more distant objects. By contrast, the slope $1 - \gamma$ of the correlation function is observed to vary little from sample to sample, with $\gamma \approx 1.8$. It is mostly sensitive to galaxy colours (see Section 5).

2.3 Estimator

In practice, the calculation of $w(\theta)$ is done through the normalized counts of galaxy–galaxy pairs $DD(\theta)$ from the data, random–random pairs $RR(\theta)$ from an unclustered random catalogue that follows the survey angular selection function and galaxy–random pairs $DR(\theta)$. Various expressions have been used to calculate $w(\theta)$. In this work we adopt the estimator introduced by Landy & Szalay (1993), which is widely used in the literature:

$$w(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}. \quad (3)$$

Landy & Szalay (1993) showed that this estimator has a small variance, close to Poisson, and allows one to measure correlation functions with minimal uncertainty and bias. The counts $DD(\theta)$, $DR(\theta)$ and $RR(\theta)$ have to be normalized to allow for different total numbers of galaxies $n_{\rm g}$ and random points $n_{\rm r}$:

$$DD(\theta) = \frac{N_{\rm gg}(\theta)}{n_{\rm g}(n_{\rm g} - 1)/2},$$
$$DR(\theta) = \frac{N_{\rm gr}(\theta)}{n_{\rm g}n_{\rm r}},$$
$$RR(\theta) = \frac{N_{\rm rr}(\theta)}{n_{\rm r}(n_{\rm r} - 1)/2}.$$

We use approximately ten times as many random points as galaxies in order that the results do not depend on a particular realization of random distribution. We also tried an alternative estimator proposed by Hamilton (1993), which revealed no significant changes in the correlation function measurements.

Estimates of the angular correlation function are affected by an integral constraint of the form

$$\frac{1}{\Omega^2}\int\int w(\theta_{12})\,d\Omega_1\,d\Omega_2 = 0, \quad (4)$$

where the integral is over all pairs of elements of solid angle Ω within the survey area. The constraint requires that $w(\theta)$ goes negative at large separations to balance the positive clustering signal at smaller separations. However, for wide-field surveys like SDSS the integral constraint has a negligible effect on $w(\theta)$, even on large scales. We find that the additive correction for the integral constraint is at least two orders of magnitude smaller than the value of $w(\theta)$ at $\theta = 9.4^\circ$. Thus the integral constraint does not bias our clustering measurements.

2.4 Spatial correlation function

We are interested in the spatial clustering and the physical separations at which galaxies are clustered, in order to compare data against theory. To this end, we need to calculate the spatial correlation function from our angular correlation function, which is simply its projection on the sky. The spatial correlation function, $\xi(r)$, can be also expressed as a power law:

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \quad (5)$$

where r_0 is the correlation length. It corresponds to the proper separation at which the probability of finding two galaxies is twice that of a random distribution, $\xi(r_0) = 1$. Limber (1953) demonstrated that the power-law approximation for $\xi(r)$ in equation (5) leads to the power law defined in equation (2), with the index γ being the same in both cases. Phillipps et al. (1978) expressed the amplitude of the correlation function, A_w, as a function of the proper correlation length r_0 and the selection function of the survey, whereas later studies propose similar equations where the selection function is implicitly included in the redshift distribution.

Now, writing the angular correlation function as $w(\theta) = A_w\theta^{1-\gamma}$, Limber's equation becomes (Peebles 1980, Section 52, 56)

$$A_w = C\,\frac{\int_{z_{\rm min}}^{z_{\rm max}} r_0^\gamma\, g(z)({\rm d}N/{\rm d}z)^2\,{\rm d}z}{\left[\int_{z_{\rm min}}^{z_{\rm max}}({\rm d}N/{\rm d}z)\,{\rm d}z\right]^2}, \quad (6)$$

where dN/dz is the redshift distribution,[1] which is zero everywhere outside the limits $z_{\rm min}$ and $z_{\rm max}$, and

$$C = \pi^{1/2}\frac{\Gamma[(\gamma - 1)/2]}{\Gamma(\gamma/2)},$$

with Γ the gamma function. The quantity $g(z)$ is defined as

$$g(z) = \left(\frac{{\rm d}z}{{\rm d}x}\right)x^{1-\gamma}F(x),$$

where $F(x)$ is related to the curvature factor k in the Robertson–Walker metric by

$$F(x) = 1 - kx^2.$$

We assume zero curvature, and so $F(x) \equiv 1$.

[1] We use the expressions dN/dz and $N(z)$ interchangeably for the redshift distribution.

When using equation (6), we need to determine the redshift distribution of the sample with precision. We address this issue in Section 4.3. Another subtle complication that arises from the use of equation (6) is that galaxy clustering is assumed to be independent of galaxy properties such as colour and luminosity (Peebles 1980, section 51). Therefore it is particularly important to use samples with fixed colour and luminosity, rather than mixed populations, to study galaxy clustering using Limber's approximation. We address this issue in Section 4.2, where we define the colour and luminosity bins for the clustering analysis.

3 DATA

To carry out this analysis, we take advantage of the Galaxy And Mass Assembly (GAMA) survey (Driver et al. 2011). This spectroscopic sample, at low to intermediate redshifts, forms an ideal training set for predicting photometric redshifts of faint galaxies. The galaxies considered for the calculation of the correlation functions are drawn from the seventh data release of the Sloan Digital Sky Survey photometric sample (SDSS DR7: Abazajian et al. 2009). We briefly outline the properties of these samples below.

3.1 SDSS DR7 photometric sample

At the time of writing, the Sloan Digital Sky Survey (SDSS) is the largest local galaxy survey ever undertaken. The completed SDSS maps almost one quarter of the sky, with optical photometry in u, g, r, i and z bands and spectra for $\sim 10^6$ galaxies. The main goal of the survey is to provide data for large-scale structure studies of the local Universe. A series of papers describes the survey: technical information about the data products and the pipeline can be found in York et al. (2000) and in Stoughton et al. (2002). Details about the photometric system can be found in Fukugita et al. (1996).

The SDSS imaging survey was completed with the seventh data release (Abazajian et al. 2009), which we use in this paper. The main programme of SDSS is concentrated in the Northern Galactic cap with three 2.5° stripes in the Southern Galactic cap. SDSS DR7 contains about 5.5×10^6 galaxies with $r_{\rm petro} < 19.4$ over 7646 deg^2 of sky.

The images are obtained with a 2.5-m telescope located at Apache Point Observatory, New Mexico. Various flux measures are available for galaxies in the SDSS data base (Stoughton et al. 2002), including Petrosian fluxes, model fluxes (corresponding to whichever of a de Vaucouleurs or exponential profile provides a better fit to the observed galaxy profile), and aperture fluxes. In this paper we use model magnitudes to calculate galaxy colours and Petrosian magnitudes to split galaxies into absolute magnitude ranges. After Schlegel, Finkbeiner & Davis (1998), we correct the magnitudes with dust attenuation corrections provided for each object and each filter in the SDSS data base.

The star–galaxy classification adopted by the SDSS photometric pipeline is based on the difference between an object's point-spread function (PSF) magnitude (calculated assuming a PSF profile, as for a stellar source) and its model magnitude. An object is then classified as a galaxy if it satisfies the criterion (Stoughton et al. 2002)

$$m_{\rm psf,tot} - m_{\rm model,tot} > 0.145, \quad (7)$$

where $m_{\rm psf,tot}$ and $m_{\rm model,tot}$ magnitudes are obtained from the sum of the fluxes over *ugriz* photometric bands. This cut works at the 95 per cent confidence level for galaxies with $r < 21$. In Section 3.2 we discuss a different star–galaxy classification, following the GAMA survey, which is the one we adopt for this work (see also Appendix A).

A photometric redshift study can be vulnerable to contamination due not only to stars misclassified as galaxies but also to contamination arising from over-deblended sources (Scranton et al. 2002), usually coming from local spiral galaxies. This imposes limits on the angular scale over which we can probe the correlation function. In order to test for this systematic in our sample, in Appendix B4 we visually inspect random samples of data and then model the contamination as a function of angular separation.

3.2 GAMA sample

The Galaxy And Mass Assembly (GAMA) project[2] is a combination of several ground- and space-based surveys with the aim of improving our understanding of galaxy formation and evolution (Driver et al. 2011). GAMA uses the AAOmega spectrograph of the Anglo-Australian Telescope (AAT) for spectroscopy (Saunders et al. 2004; Sharp et al. 2006). Its targets are selected from the SDSS photometric sample. Target selection is described in detail by Baldry et al. (2010). The main restriction is that the source is detected as an extended object: $r_{\rm psf} - r_{\rm model} > 0.25$. As shown in Appendix A, this criterion is also adopted for our sample extraction from SDSS. This criterion is more restrictive, in the sense that fewer stars will be misclassified as galaxies, than the star–galaxy classification adopted by the SDSS photometric pipeline (previous section), but similar to that used for the SDSS main galaxy spectroscopic sample (Strauss et al. 2002).

The GAMA survey is almost 99 per cent spectroscopically complete over its 144 deg^2 area to $r_{\rm petro} = 19.4$ mag (Driver et al. 2011). GAMA phase 1 (comprising 3 years of observations) includes 95 592 reliable spectroscopic galaxy redshifts to this magnitude limit, extending to redshift $z \approx 0.5$. Of these redshifts, 76 360 have been newly acquired by the GAMA team. The rest come from previous surveys: SDSS (Abazajian et al. 2009), 2dFGRS (Colless et al. 2001; Cole et al. 2005), 6dFGS (Jones et al. 2004), MGC (Driver et al. 2005) and 2SLAQ (Cannon et al. 2006). The overall GAMA redshift distribution is shown in fig. 13 of Driver et al. (2011).

For a consistent training of ANNz it is necessary to match all the GAMA objects with SDSS DR7 übercal photometry (Padmanabhan et al. 2008) and perform identical colour cuts. Once we apply the colour cuts (Section 3.3) necessary for the optimization of ANNz performance and low- and high-redshift cuts ($0.002 < z < 0.5$), 93 584 redshifts remain. They are used to train our photometric redshift neural net algorithm, as described in Section 4.

3.3 Colour cuts

Before we build our final sample from ANNz, we remove galaxies with outlier $u - g$, $g - r$, $r - i$, $i - z$ colours in both the SDSS imaging sample and the training set, because photometric redshift estimates are based primarily on these colours. The complete colour and magnitude cuts are given in Table 1. Fewer than 1 per cent of galaxies are affected by the colour cuts. These colour cuts could in principle affect the mask that we use for correlation-function calculations. To estimate the extent of this effect, we study the distribution on the sky of the colour outliers as well as their angular correlation function. This exercise reveals that colour outliers have a spurious correlation an order of magnitude larger on all angular

[2] http://www.gama-survey.org

Table 1. Colour and apparent magnitude cuts for the optimization of ANNZ. All magnitudes are SDSS model magnitudes.

$12.0 < r_{\rm petro} < 19.4$
$-2 < u - g < 7$
$-2 < g - r < 5$
$-2 < r - i < 5$
$-2 < i - z < 5$

scales than the correlation function of our final sample. However, since the number of these objects is almost three orders of magnitude lower than the total, they would have a negligible effect on $w(\theta)$ measurements if included.

3.4 Final sample

Our aim is to obtain a galaxy sample with photometric properties as close as possible to our training set. To this end, we have selected galaxies from the SDSS DR7 photometric sample with the query used to select GAMA targets (Appendix A). We select galaxies that have 'clean' photometry according to the instructions given on the SDSS website.[3] Our sample is hence limited by $r_{\rm petro} < 19.4$ and satisfies the criterion for star–galaxy separation $r_{\rm psf} - r_{\rm model} > 0.25$. In our analysis, we choose to calculate the correlation function for galaxies located in the SDSS northern cap, corresponding to 92 per cent of SDSS DR7 galaxies. As such, the geometry of the survey is simplified to a contiguous area. Our final sample, after the colour cuts given in Table 1, comprises 4890 965 galaxies.

To evaluate the number of data–random and random–random pairs in equation (3), we need to build a mask for our sample. The mask precisely defines the sky coverage of the sample. We use the file `lss_combmask.dr72.ply` in the NYU Value Added Catalogue[4] (Blanton et al. 2005b), mapping SDSS stripes, as our mask. This file contains the coordinates of the fields observed by SDSS expressed in spherical polygons, excluding areas around bright stars because galaxies in these regions can be affected by photometric errors. It is also suitably formatted for use with the MANGLE software (Hamilton 1993; Hamilton & Tegmark 2004; Swanson et al. 2008b), a tool for manipulating survey masks and obtaining random points with the exact geometry of the mask. Once masking is applied, 4511 011 galaxies remain in our sample.

The upper panel of Fig. 1 shows the boundaries of the final mask for SDSS DR7 that we use for creating random catalogues. Our random catalogues consist of $\sim 10^7$ objects, approximately ten times larger than the number of galaxies in each luminosity and colour bin. Consistency checks have shown that our clustering results are not sensitive to any particular realization of the random catalogue. In Appendix B1 we check the accuracy of the survey mask, as well as the photometric uniformity of the sample, by studying the angular clustering of our sample as a function of r-band apparent magnitude.

3.5 Pixelization scheme and jack-knife resampling

In order to speed up the computation of the correlation function, we pixelize our data according to the SDSSPix[5] scheme. The basic

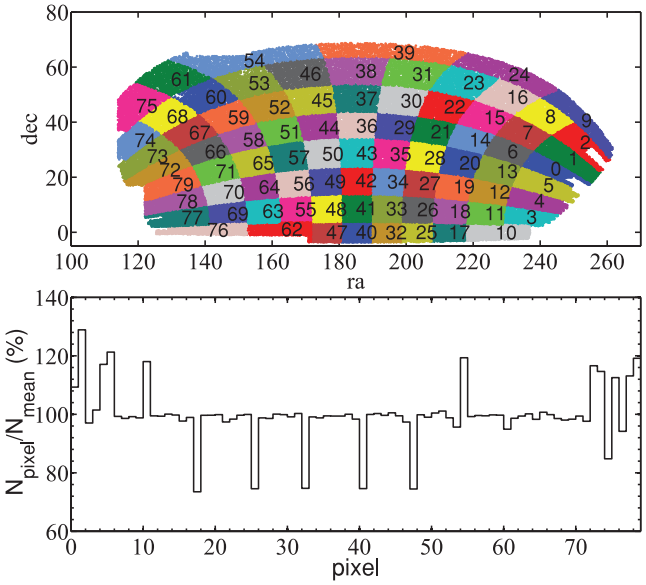

Figure 1. The upper panel shows the jack-knife regions used for the error estimation of our correlation-function measurements. After modifying the SDSSPix scheme there are 80 jack-knife regions, which contain approximately equal numbers of random points. The lower panel reports the normalized area of each pixel, based on a random catalogue. The deviations from uniformity show that differences in the areas of the JK regions are limited to ±30 per cent at most.

concept consists of assigning galaxies located in a portion of the sky to a pixel. After this step, we only need to take into account galaxies in the same pixel and in the neighbouring pixels to calculate the correlation function up to the scale of a pixel. SDSSPix divides the sky along SDSS η and λ spherical coordinates (as defined in section 3.2.2 of Stoughton et al. 2002) into equal spherical areas. Different resolutions are available according to the angular scale of interest. We choose the resolution called basic resolution (resolution = 1). This divides the sky into 468 pixels of size $\sim 9.4 \times 9.4\,{\rm deg}^2$. Then, for galaxies in a given pixel, that pixel and its 8 direct neighbouring pixels include all neighbouring galaxies with separations up to $9.^\circ4$, the largest angular separation we consider (see Section 5).

We also use this pixelization scheme to define the jack-knife (JK) regions for the error analysis. In order to minimize the variation in the number of objects in each JK region, some neighbouring pixels that contain the survey boundary are merged in order that they contain a more nearly equal number of random points. This modification of the SDSSPix pixelization yields 80 JK regions, as shown in the upper panel of Fig. 1. The lower panel of Fig. 1 presents the relative variation in area of each region, as measured by the relative number of randoms each one contains. Hereafter, errors in $w(\theta)$ are determined from 80 JK resamplings, by calculating $w(\theta)$, omitting each region in turn. We have checked that our results are not significantly affected by using either 104 or 40 jack-knife regions. The elements of the covariance matrix, $\mathbf{C}$, are given by

$$C_{ij} = \frac{N-1}{N} \sum_{k=1}^{N} \left[\log\left(w_i^k\right) - \log(\bar{w}_i)\right] \left[\log\left(w_j^k\right) - \log(\bar{w}_j)\right], \quad (8)$$

where w_i^k is the angular correlation function of the kth JK resampling on scale θ_i, $\bar{w}_i$ the mean angular correlation function and N the total number of JK resamplings. In practice, $\bar{w}_i$ is identical to the angular correlation function measurement from the whole survey

[3] http://www.sdss.org/dr7/products/catalogs/flags.html
[4] http://sdss.physics.nyu.edu/vagc/
[5] http://dls.physics.ucdavis.edu/~scranton/SDSSPix/

area. The $N - 1$ factor in the numerator of equation (8) accounts for correlations inherent in the JK procedure (Miller 1974).

The jack-knife procedure is a method of calculating uncertainties in a quantity that that we measure from the data itself. In wide-field galaxy surveys, more often than not large superstructures appear to influence clustering measurements significantly. The best-known example is the SDSS Great Wall (Gott et al. 2005). The presence of such structures makes it tempting to present the results with and without the JK region that encloses them, as done in the clustering studies of Zehavi et al. (2005, 2011). Better still, Norberg et al. (2011) devise a more objective method to remove outlier JK regions consistently from the distribution of all JK measurements that one has at hand. We follow that method in the present analysis, and find that, for all samples considered, the number of JK regions that are outliers and therefore removed is mostly two or three and no more than five.

4 PHOTOMETRIC REDSHIFTS

For the clustering measurements presented in this paper, all distance information comes from photometric redshifts (photo-z). Photo-z values are the basis for estimating the redshift distributions to be used in equation (6) and in estimating distance moduli to calculate absolute magnitudes and colours. For this study we have a truly representative subset of SDSS galaxies down to $r < 19.4$ and we therefore use the artificial neural network package ANNZ developed by Collister & Lahav (2004) to obtain photo-z estimates.

It is important that the training set and the final galaxy sample from SDSS are built using the same selection criteria. The input parameters are the following: übercalibrated, extinction-corrected model magnitudes in *ugriz* bands, the radii enclosing 50 per cent and 90 per cent of the Petrosian r-band flux of the galaxy, and their respective uncertainties. The architecture of the network is 7:11:11:1, with seven input parameters described above, two hidden layers with 11 nodes each and a single output, the photo-z. We use a committee of 5 networks to predict the photo-z values and their uncertainties (see Section 4.1).

4.1 Photometric redshift errors

Before we proceed with the photo-z derived quantities that we use in this study, we investigate the possible biases and errors that ANNZ introduces, using the known redshifts from GAMA. Following standard practice we split our data into three distinct sets: the training set, the validation set and the test set. Half of the objects constitute the test set and the other half the training and validation sets. This investigation is insensitive to the exact numbers in these three sets. The training and validation sets are used for training the network, whereas the test set is treated as unknown. Given predicted photo-z values $z_{\rm phot}$, we can quantify the redshift error for each galaxy in the test set as

$$\delta z \equiv z_{\rm spec} - z_{\rm phot}, \tag{9}$$

the primary quantity of interest as far as true redshift errors are concerned. It can depend on apparent magnitude, colour, the output $z_{\rm phot}$ and the intrinsic scatter $z_{\rm err}$ of ANNZ committees, as well as the position of an object on the sky if the survey suffers from any photometric non-uniformity. We investigate some of these potential sources of error below. The dispersion σ_z of δz is given by the equation

$$\sigma_z^2 = \left\langle (\delta z)^2 \right\rangle - \langle (\delta z) \rangle^2 , \tag{10}$$

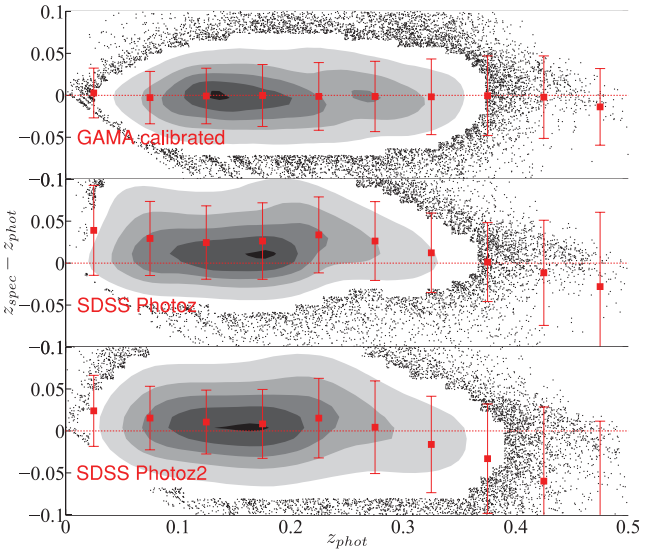

Figure 2. Density/scatter plot of redshift error (spectroscopic minus photometric redshift) against predicted photo-z from this work (top panel) and SDSS (middle and bottom panels). The colour coding is such that the densest area (black contour) is five times denser than the white contour. Points are drawn whenever the density of points is less than 10 per cent of the maximum (black contour). The red squares and error bars represent the mean redshift errors and their standard deviations in photo-z bins of width $\Delta z_{\rm phot} = 0.05$. Horizontal red lines show the zero-error benchmark. The improvement in photometric redshift estimates in this work, due primarily to use of the representative GAMA training set, is clear.

and is found to be $\sigma_z = 0.039$. The standard deviation for the redshift range $0 < z_{\rm phot} < 0.4$, within which we choose to work, is $\sigma_z = 0.035$.

In Fig. 2 we compare our photo-z estimates with the publicly available photo-z from the SDSS website (Oyaizu et al. 2008, tables `photoz1` and `photoz2`). For this comparison we plot the redshift error as a function of photo-z. We then calculate the mean and standard deviation of δz for photo-z bins of width $\Delta z_{\rm phot} = 0.05$. The number of catastrophic outliers (galaxies with $|z_{\rm phot} - z_{\rm spec}| > 3\sigma_z$) for the GAMA calibrated photo-z is 1 per cent or lower for all photo-z bins. We work in fixed photo-z bins, because all our derived quantities are based on the photo-z estimates. This way, any biases with estimated photo-z are readily apparent. Our results based on the GAMA training set outperform the SDSS results: for the redshift range $0.01 < z_{\rm phot} < 0.4$, we obtain essentially unbiased redshift estimates, given the observed scatter. The scatter, in turn, increases with redshift. We note, however, that the `photoz2` catalogue from SDSS DR7 has been improved with the addition of $p(z)$ estimates, which are designed to perform much better in recovering the total redshift probability distribution function of all galaxies (Cunha et al. 2009). Since it is still not clear how to relate a redshift pdf directly to absolute magnitude and colour for a given galaxy, our approach for the study of luminosity- and colour-dependent clustering is easier to interpret.

In Appendix B2, we quantify the photo-z error and possible contamination between redshift bins by cross-correlating photo-z bins that are more than $2\sigma_z$ apart. We find, as expected, that the residual cross-correlation of the different photo-z bins is negligible compared with their auto-correlation.

The distribution of photo-z errors is in general non-Gaussian, albeit less pronounced in the case of a complete training set. Photo-z errors also propagate asymmetrically in absolute magnitude: for a

Table 2. The change in the total number of galaxies as a result of the cuts applied in various stages of the analysis.

Cut description	Number of galaxies left
None	4914 434
Colour cuts (Table 1)	4890 965
Masking	4511 011
$z_{\rm err}^{\rm (ANNz)} < 0.05$ & $0.002 < z_{\rm phot} < 0.4$	4289 223

given redshift error, the error induced in absolute magnitude is larger at low z and smaller at high z, and thus a photo-z analysis is more tolerant to redshift errors for objects at high z. For that reason, it is common practice to scale the redshift error by the quantity $1/(1 + z_{\rm phot})$. Taking into account this redshift stretch, σ_0 can be defined as

$$\sigma_0^2 = \left\langle \left(\frac{\delta z}{1 + z_{\rm phot}} \right)^2 \right\rangle - \left\langle \left(\frac{\delta z}{1 + z_{\rm phot}} \right) \right\rangle^2, \quad (11)$$

giving $\sigma_0 = 0.032$.

We exclude from our analysis galaxies with $z_{\rm phot} < 0.002$ or $z_{\rm phot} > 0.4$. ANNz provides a photo-z error calculated from the photometric errors. Using our test set, we find that this error underestimates the true photo-z error (given from equation 9). We therefore apply a cut on the output parameter $z_{\rm err}$ of ANNz at $z_{\rm err} < 0.05$. These cuts eliminate ∼4 per cent of the galaxies. Cross-checks show that the correlation function measurements do not change if we use a less strict cut, but the chosen cut does improve the $N(z)$ estimates. The final number of galaxies after this cut is 4289 223. We summarize the changes in the number of galaxies in our sample in Table 2. We use Petrosian magnitudes to divide galaxies by luminosity and model magnitudes to calculate galaxy colours.

The photo-z work presented here is similar, but not identical, to that of Parkinson (2012). The latter is appropriate for even fainter SDSS magnitudes as it uses, in its training and validation, all GAMA galaxies with $r_{\rm petro} < 19.8$ and fainter zCOSMOS galaxies (Lilly et al. 2007) matched to SDSS DR7 imaging. Minor differences in the two photo-z pipelines, such as the inclusion of different light-profile measurements, do not significantly affect the estimated photo-z, which presents a similar scatter around the underlying spectroscopic distribution. Our photo-z values agree with those of Parkinson (2012) within the estimated errors.

4.2 Division by redshift, absolute magnitude and colour

Galaxy magnitudes are $k + e$ corrected to $z_{\rm phot} = 0.1$, using KCORRECT version 4.1.4 (Blanton & Roweis 2007) and the passive evolution parameter $Q = 1.62$ of Blanton et al. (2003). In this simple model, the evolution-corrected absolute magnitude is given by $M_{\rm corr} = M - Q(z - z_0)$, where $z_0 = 0.1$ is the reference redshift. We note that Loveday et al. (2012) using GAMA found $Q = 0.7$, which would change evolution-corrected magnitudes by ≈0.3 mag at $z = 0.4$. Approximately equal deviations in absolute magnitude will be induced in our high-z blue galaxy samples, if we use a colour-dependent Q (e.g. Loveday et al. 2012). Assuming a global value for Q, however, allows for a more direct comparison with the SDSS-based clustering studies of Zehavi et al. (2005, 2011). Galaxy colours, derived from SDSS model magnitudes, are referred to as ${}^{0.1}(g - r)$, while absolute magnitudes are derived using the r-band Petrosian magnitude (to match the GAMA redshift survey selection). Fig. 3 shows that the r-band absolute magnitude extends to

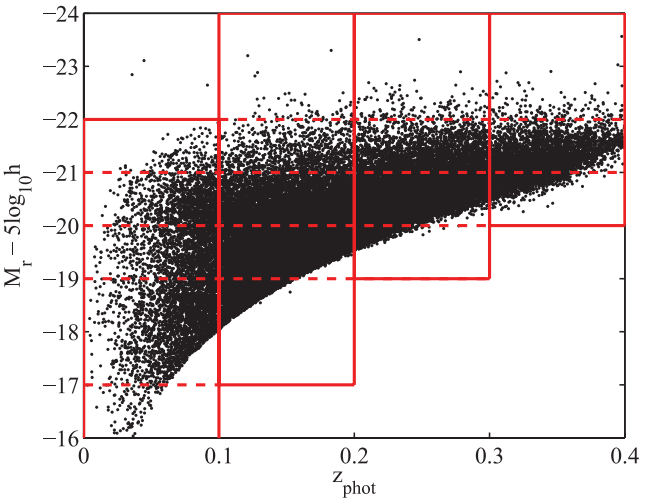

Figure 3. r-band absolute magnitude against photo-z for our photometric sample. Solid red lines show the boundaries of our samples in photo-z and absolute magnitude and dashed lines the further split in absolute magnitude bins. Only 1 per cent of galaxies are shown.

$M_r - 5\log h = -16$ mag with a few galaxies reaching as faint as $M_r - 5\log h = -14$ mag.

We split our galaxy sample into photo-z as well as luminosity bins. Our samples are shown in Fig. 3. Initially we define four photo-z bins in the redshift range $0 < z_{\rm phot} < 0.4$ and then we further split each photo-z-defined sample into six absolute magnitude bins in the range $-24 < M_r - 5\log h < -14$. Thus our photo-$z$ catalogue offers the opportunity for a clustering analysis over the luminosity range $0.03L^* \lesssim L \lesssim 8L^*$, spanning almost three orders of magnitude in L/L^*.

In Fig. 3 some of these redshift–magnitude bins extending beyond the survey flux limit are only partially occupied by galaxies in terms of photometric redshifts and photo-z-derived absolute magnitudes. The true redshift and absolute magnitude distributions for each bin are recovered by Monte Carlo resampling, as discussed in Section 4.3.

Fig. 4 shows colour–magnitude diagrams for our sample split into photo-z bins. The colour bimodality is evident at ${}^{0.1}(g - r) \simeq 0.8$ for all photo-z bins. We have adopted the tilted colour cuts defined by Loveday et al. (2012):

$$M_r - 5\log h = 5 - 33.3 \times {}^{0.1}(g - r)_{\rm model}, \quad (12)$$

which is a slightly modified version of the colour cut used by Zehavi et al. (2011), also shown in Fig. 4.

In Fig. 5 we plot the photo-z error against photo-z for galaxies subdivided into subsamples, where we again have used *photometric* redshifts to estimate galaxy luminosities and colours. There are no obvious systematic biases of $z_{\rm spec} - z_{\rm phot}$ for any of the subsamples, although we do note that the most luminous (faintest) bin contains very few blue (red) galaxies.

The relatively good photo-z notwithstanding, our analysis does not eliminate completely the main systematic error of neural-network-derived photo-z values, which is the overestimation of low redshifts and the underestimation of high redshifts (see e.g. fig. 7 of Collister et al. 2007). As a result, a number of faint galaxies have their redshift overestimated and hence appear brighter in our sample. We note that there is a discrepancy between the fraction of faint red objects in the luminosity bin $-19 < M_r - 5\log h < -17$ between this work and that of Zehavi et al. (2011), which is most probably caused by this systematic shift (see

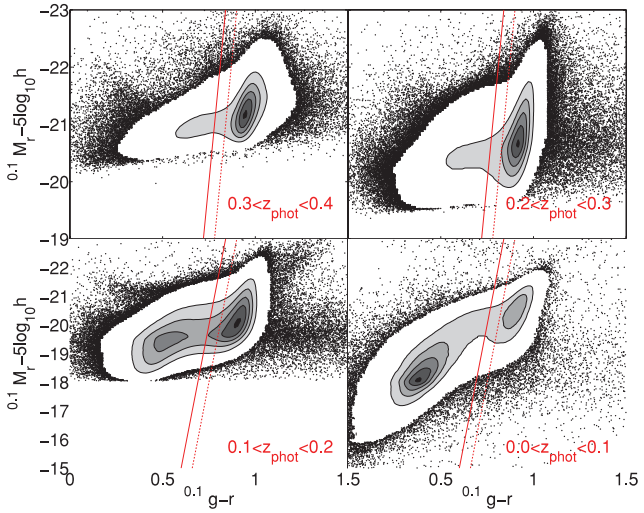

Figure 4. r-band absolute magnitude against $^{0.1}(g-r)$ colour (both k-corrected and passively evolved to $z = 0.1$) for galaxies split into photo-z bins. Solid red lines show the colour cut for red and blue populations suggested by Loveday et al. (2012) and used in this work, dashed red lines the colour cut used by Zehavi et al. (2011).

Table 3). It is possible to cure this by Monte Carlo resampling the photo-z values with their respective errors and then re-derive the absolute magnitudes and colours, but we do not pursue this here.

4.3 Photometric redshift distribution(s)

Despite the fact that ANNz gives fairly accurate and unbiased photo-z values for calculations in broad absolute magnitude bins or photo-z bins, in order to translate the two-dimensional clustering signal to the three-dimensional one using equation (6) the underlying true dN/dz is needed. In this work we loosely follow the approach given in Parkinson (2012) (see also Driver et al. 2011). The GAMA spectroscopic sample is highly representative and allows us to calculate the true redshift errors as a function of photo-z for all objects in GAMA with $r_{\rm petro} < 19.4$. Then, under the assumption of a Gaussian photometric error distribution in each photo-z bin, we perform a Monte Carlo resampling of the ANNz predictions for photo-z values. This is equivalent to replacing each photo-z derived from ANNz with the quantity $z_{\rm MC}$ drawn from a Gaussian distribution, using a photo-z-dependent standard deviation, $\sigma(z_{\rm phot}^{\rm (bin)}) = \delta z_{\rm phot}^{\rm (bin)}$:

$$z_{\rm MC} = G[\mu = z_{\rm phot}, \sigma = \sigma_{\rm phot}(1 + z_{\rm phot})]. \quad (13)$$

Note that *convolving* the imprecise photo-z with additional scatter improves the $N(z)$ redshift distribution: in other words the photo-z process *deconvolves* the $N(z)$ and makes it artificially narrow.

All our sample selections in Fig. 6 have been made using the photo-z derived absolute magnitude $M_r - 5\log h$. We then use the accurate spectroscopic information from GAMA to assess how well Monte Carlo resampling compares with the underlying true dN/dz. Since the GAMA area is much smaller than the SDSS area, we do not wish to recover the exact spectroscopic redshift distribution, merely to match a smoothed version thereof. Our test shows that MC resampling performs rather well in recovering the true dN/dz. This method performs even better with a larger number of objects, which indicates that results are still dominated by statistical errors and therefore there is room for improvement in future when larger

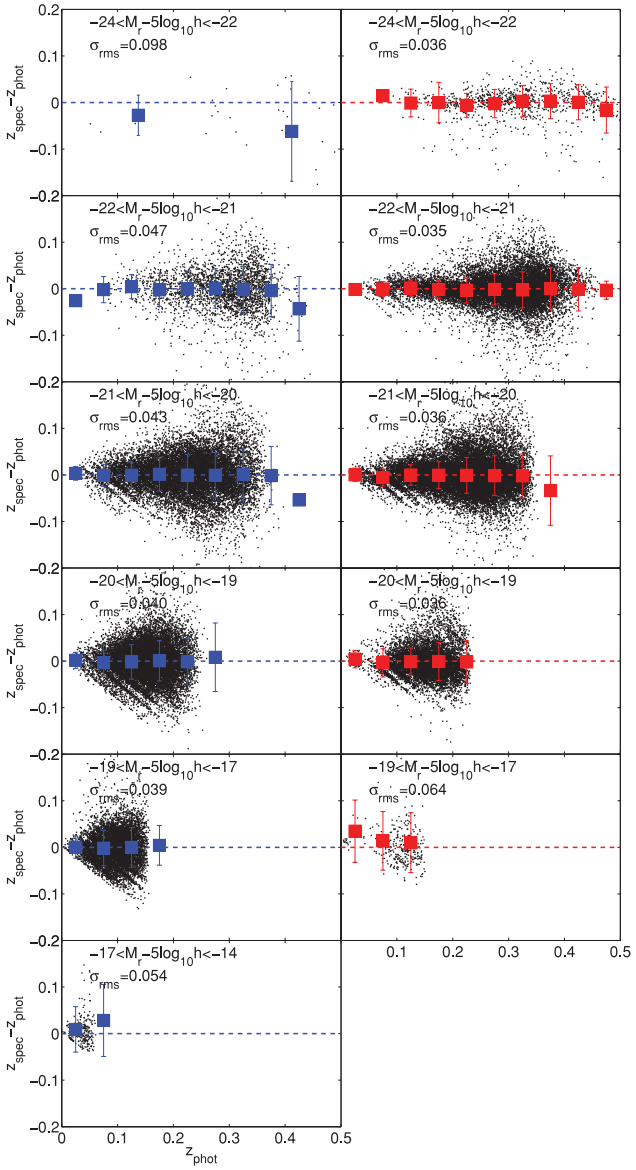

Figure 5. Redshift error against photo-z for our luminosity- and colour-selected GAMA subsamples. The mean redshift error and standard deviation in bins of photo-z are shown by the coloured squares and error bars, while the root-mean-square standard deviation, $\sigma_{\rm rms}$, is listed in each panel. The faint red sample has been omitted due to the small number of galaxies that it contains.

spectroscopic training sets become available. Nevertheless, as an incorrect redshift distribution can cause a systematic error in r_0, in Appendix B3 we test the sensitivity of our results to the assumed dN/dz and compare results using the Monte Carlo recovered dN/dz with those from the weighting method proposed by Cunha et al. (2009).

Fig. 7 shows, for all samples split by photo-z and photo-z-derived absolute magnitude, the photo-z-derived, true underlying and Monte Carlo inferred absolute magnitude distributions (as dashed, thin and thick solid lines respectively). We note that the photo-z-derived absolute magnitude estimates in Fig. 7 are obtained from the resampled redshifts and not by resampling the absolute magnitudes per se.

Table 3. Clustering properties of luminosity-selected samples. Column 1 lists the photo-z-based absolute magnitude ranges, column 2 the median absolute magnitude and the associated 16th and 84th percentiles from the Monte Carlo resampling (Fig. 7) and column 3 the number of galaxies in each sample. Columns 4, 5 and 6 list respectively the slope γ, the correlation length r_0 and the reduced χ^2 χ^2_ν of the power-law fit as defined in Section 2.4. Columns 7, 8 and 9 show the same information but for power-law fits using only the diagonal elements of the covariance matrix. All power-law fits are approximately over the comoving scales $0.1 < r < 20\,h^{-1}$ Mpc. Finally, column 10 presents the relative bias at $5\,h^{-1}$ Mpc measured using equation (14).

Sample $M_r - 5\log h$	Magnitude$^{(MC)}$ $M_r - 5\log h$	N_{gal}	γ	r_0 [h^{-1} Mpc]	χ^2_ν	$\gamma^{(d)}$	$r_0^{(d)}$ [h^{-1} Mpc]	$\chi^{(d)2}_\nu$	b/b^* [h^{-1} Mpc]
All colours $0.3 < z_{phot} < 0.4$									
$[-24, -22)$	$-22.0^{-0.2}_{+0.2}$	13257	2.01 ± 0.15	14.08 ± 2.09	3.41	2.02 ± 0.09	13.68 ± 1.22	2.6	2.13 ± 0.30
$[-22, -21)$	$-21.2^{-0.3}_{+0.3}$	339834	1.94 ± 0.11	8.23 ± 1.54	28.08	1.91 ± 0.09	8.46 ± 1.06	13.0	1.22 ± 0.22
$[-21, -20)$	$-20.8^{-0.2}_{+0.2}$	158860	1.75 ± 0.06	6.96 ± 0.56	3.76	1.78 ± 0.05	6.80 ± 0.33	1.8	1.00 ± 0.01
All colours $0.2 < z_{phot} < 0.3$									
$[-24, -22)$	$-22.0^{-0.3}_{+0.3}$	12294	2.02 ± 0.11	13.29 ± 2.01	2.37	2.01 ± 0.07	13.17 ± 1.13	1.7	2.02 ± 0.32
$[-22, -21)$	$-21.2^{-0.4}_{+0.3}$	284969	1.92 ± 0.09	7.92 ± 1.13	10.91	1.90 ± 0.06	8.12 ± 0.70	5.5	1.17 ± 0.17
$[-21, -20)$	$-20.4^{-0.3}_{+0.4}$	930539	1.75 ± 0.05	6.94 ± 0.76	7.96	1.77 ± 0.05	6.74 ± 0.36	3.3	1.00 ± 0.03
$[-20, -19)$	$-19.8^{-0.3}_{+0.3}$	122870	1.75 ± 0.08	5.84 ± 0.57	2.44	1.76 ± 0.06	5.84 ± 0.29	1.5	0.86 ± 0.10
All colours $0.1 < z_{phot} < 0.2$									
$[-24, -22)$	$-22.0^{-0.4}_{+0.3}$	4311	1.96 ± 0.09	12.58 ± 1.35	0.59	1.95 ± 0.08	12.57 ± 1.13	0.4	2.10 ± 0.35
$[-22, -21)$	$-21.2^{-0.4}_{+0.5}$	106728	1.92 ± 0.05	7.31 ± 0.60	3.56	1.92 ± 0.04	7.40 ± 0.32	1.7	1.22 ± 0.18
$[-21, -20)$	$-20.3^{-0.5}_{+0.5}$	604181	1.75 ± 0.05	6.03 ± 0.77	7.16	1.78 ± 0.06	5.85 ± 0.43	3.9	1.00 ± 0.05
$[-20, -19)$	$-19.5^{-0.4}_{+0.5}$	916563	1.63 ± 0.11	6.36 ± 2.42	42.40	1.71 ± 0.10	5.81 ± 0.75	11.7	1.03 ± 0.30
$[-19, -17)$	$-18.6^{-0.4}_{+0.6}$	211336	1.55 ± 0.08	5.17 ± 0.83	4.41	1.58 ± 0.07	4.89 ± 0.34	1.6	0.87 ± 0.16
All colours $0.0 < z_{phot} < 0.1$									
$[-22, -21)$	$-21.1^{-0.7}_{+0.8}$	19218	1.89 ± 0.13	8.21 ± 2.32	6.36	1.88 ± 0.07	8.09 ± 0.80	1.6	1.15 ± 0.43
$[-21, -20)$	$-20.3^{-0.7}_{+0.9}$	122787	1.68 ± 0.09	7.31 ± 1.40	9.00	1.75 ± 0.05	6.84 ± 0.50	2.1	0.99 ± 0.23
$[-20, -19)$	$-19.4^{-0.6}_{+0.8}$	155147	1.60 ± 0.08	6.23 ± 1.06	9.08	1.65 ± 0.08	6.10 ± 0.64	4.5	0.86 ± 0.20
$[-19, -17)$	$-18.1^{-0.8}_{+1.0}$	271389	1.54 ± 0.06	4.33 ± 0.58	6.20	1.58 ± 0.09	3.97 ± 0.24	2.9	0.65 ± 0.18
$[-17, -14)$	$-16.6^{-0.9}_{+1.4}$	14659	2.03 ± 0.25	4.28 ± 1.56	5.82	2.00 ± 0.28	4.41 ± 1.03	2.1	0.62 ± 0.25

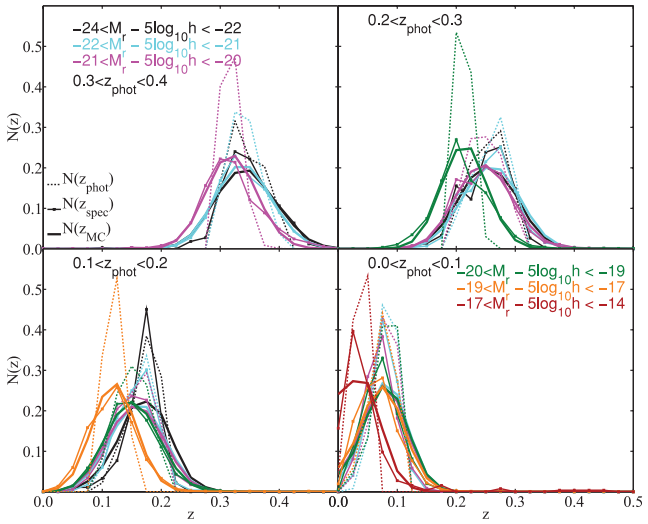

Figure 6. Estimates of the underlying redshift distribution for the luminosity samples used in the clustering analysis. Thin solid lines show the photo-z distribution, which is the basis for the selection, dotted lines the true spectroscopic redshift distribution from GAMA and thick solid lines the average distribution inferred from 100 Monte Carlo resamplings of the photo-z distribution using equation (13).

We then $k + e$ correct every Monte Carlo absolute magnitude realization using the procedure described in Section 4.2. As expected, the true underlying distribution extends well beyond the photo-z inferred luminosity bins, but is yet again rather well described by the Monte Carlo inferred distribution.

It is crucial that we have a good understanding of the true underlying absolute magnitude for all our samples. For galaxy clustering studies with spectroscopic redshifts it is desirable to work with volume-limited samples. Using photometric redshifts, however, one can form only approximately volume-limited samples, since photo-z uncertainties will propagate into absolute magnitude estimates. Essentially, any top-hat absolute magnitude distribution, as selected using photo-z, corresponds to a wider true absolute magnitude distribution, as shown in Fig. 7. This is rather similar to selecting galaxies from a photometric redshift bin and then convolving the initial top-hat distribution with the photo-z error distribution in order to obtain the true $N(z)$. However, using the $w(\theta)$ statistic and an accurate dN/dz for that particular galaxy sample, we can extract its respective spatial clustering signal, which would then correspond to the z_{MC} derived absolute magnitude. Direct comparisons with other studies can then be made, modulo the extent of the overlap between the two absolute magnitude distributions.

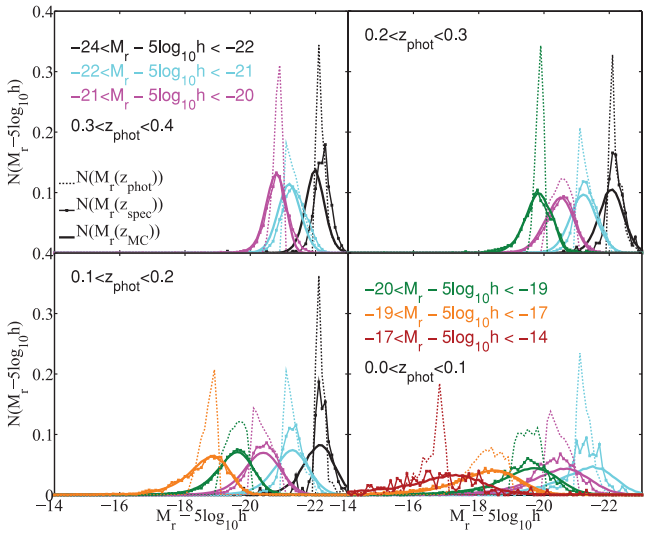

Figure 7. The r-band absolute magnitude distribution for GAMA galaxies with $r_{\rm petro} < 19.4$ split into photo-z and photo-z-derived absolute magnitude slices. Magnitude distributions shown by dashed lines are derived from the raw photo-z, those shown by thin lines from the underlying spectroscopic redshifts and those shown by thick lines from the Monte Carlo derived magnitudes. The latter reproduce the true underlying spec-z inferred magnitude distribution rather well; however for a few samples there is a discrepancy between the spec-z-derived and Monte Carlo derived distributions. All Monte Carlo absolute magnitude estimates are K-corrected and passively evolved following the procedure described in Section 4.2.

5 RESULTS FOR THE TWO-POINT CORRELATION FUNCTION

5.1 Luminosity and redshift dependence

We first calculate the angular correlation function $w(\theta)$ for our samples selected on absolute magnitude and photometric redshift over angular scales from 0.005–9.4° in 15 equally spaced bins in $\log(\theta)$.[6] In a flux-limited survey like SDSS, intrinsically bright galaxies dominate at high redshifts and intrinsically faint objects dominate at low redshifts (see Fig. 4). For that reason, we calculate $w(\theta)$ for the 17 well-populated samples given in Table 3. Errors are estimated using the jack-knife technique, with the covariance matrix given by equation (8). Even if the validity of a given error method based on data alone is still widely debated, it is commonly accepted that the jack-knife method is adequate for angular clustering studies (see e.g. Cabré et al. 2007), while for three-dimensional clustering measurements Norberg et al. (2009) have shown that the jack-knife method suffers from some limitations, in particular on small scales.

Our angular correlation function measurements are broad and probe both highly non-linear and quasi-linear scales. Fig. 8 presents galaxy angular correlation functions for six photo-z-selected absolute magnitude bins. We show the angular scale (lower x-axis) used for the correlation function estimation and the corresponding comoving scale estimated at the mean redshift of the sample (upper x-axis).

Over the range of angular scales fitted, chosen to correspond to approximately 0.1–20 h^{-1} Mpc comoving separation according to the mean redshift of each sample, the angular correlation function can be reasonably well approximated by a power law (equation 2). We perform power-law fits with both the full covariance matrix and the diagonal elements only. The power-law fits for our L^* sample are shown in Fig. 8. Dotted lines in Fig. 8 show the extension of the power laws beyond the scales over which they were fitted. The resulting correlation lengths r_0, slopes γ and quality of the fits as given by the reduced χ^2, χ^2_ν, for all samples are listed in Table 3.

The luminosity dependence of galaxy clustering is present in all photo-z shells: the shape and amplitude of the angular correlation function differ for galaxies with different luminosity. The amplitude of the angular correlation function decreases as we go from bright to faint galaxies for all photo-z bins. The slope of the correlation function also decreases with decreasing luminosity, very much in line with the change in the fraction of red and blue galaxies. As observed in Section 5.2, red (blue) galaxies dominate the brightest (faintest) luminosity bins, with red galaxies preferentially having a steeper correlation function slope than blue galaxies.

For each sample, we estimate the correlation length r_0 via equation (6) using the Monte Carlo inferred redshift distribution described in Section 4.3. The redshift distribution dN/dz is calculated separately for each sample, as shown in Fig. 6. In Appendix B3 we investigate the effects of the assumed dN/dz on the recovered correlation length r_0 and show that the adopted dN/dz recovery method compares favourably with the true underlying dN/dz, as obtained from the smoothed dN/d$z_{\rm spec}$.

For our luminosity bins in the redshift range $0 < z < 0.1$, the correlation length is found to decrease as we go to fainter absolute magnitudes, from $8.21 \pm 2.32\,h^{-1}$ Mpc ($-22 < M_r - 5\log h < -21$) to $4.28 \pm 1.56\,h^{-1}$ Mpc ($-19 < M_r - 5\log h < -17$). This is very much in line with the recent results of Zehavi et al. (2011). Moreover, we do not observe strong evolution with redshift for samples of fixed luminosity. All r_0 and γ measurements are shown in Fig. 9.

There are two main sources of error in the r_0 estimates: (i) the correlated uncertainties in the power-law parameters γ and A_w, which propagate through equation (6) to r_0; (ii) statistical and systematic uncertainties in the modelling of the underlying redshift distribution. The $w(\theta)$ uncertainties and the induced error in r_0 and γ are obtained using the standard deviation from the distribution of JK resampling estimates (Section 3.5). As in the case of the covariance matrix, these uncertainties are multiplied by a factor of $N - 1$ (Norberg et al. 2009). The dN/dz uncertainties are investigated in great detail in Appendix B3, where we show that the Monte Carlo inferred dN/dz performs best while still returning a residual systematic uncertainty of $\pm 0.2\,h^{-1}$ Mpc in r_0 that depends on the sample considered. We find that both sources of uncertainty have a comparable contribution to the errors. In Table 3 we quote the total error in the correlation length after adding the two (independent) errors in quadrature.

5.2 Luminosity, redshift and colour dependence

We repeat the clustering analysis, splitting the samples into red and blue colours using equation (12). For each new sample we re-estimate the underlying redshift distribution used in the inversion of Limber's equation. The corresponding 50th, 16th and 84th percentiles of the underlying absolute magnitude distributions are given in Tables 4 and 5. We also repeat the procedure outlined in Section 5.2 for the error estimation.

In Fig. 10 we present the angular correlation functions in each luminosity and photo-z bin, for red and blue galaxies. The power-law fits over approximately fixed comoving scales and their

[6] Initially our analysis was performed down to $\theta = 0.001°$. However, as shown in Section 5.3 and Appendix B4, the data are not reliable enough on such small scales.

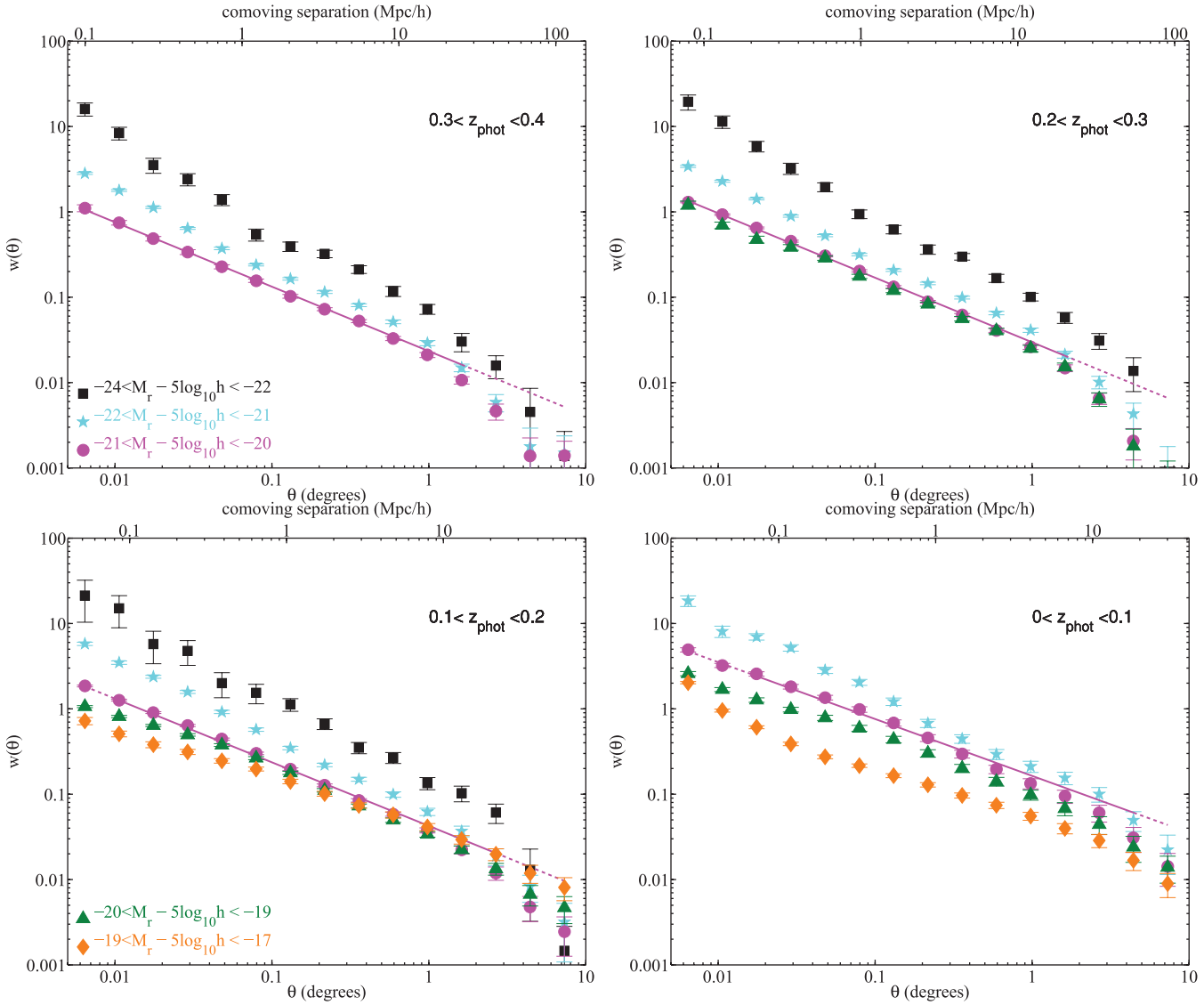

Figure 8. Two-point angular correlation functions $w(\theta)$ of our samples split into photo-z bins and six photo-z-inferred absolute magnitude bins, as indicated in each panel, with jack-knife errors. The solid lines show power-law fits estimated using the full covariance matrix for the L^* sample. Dotted lines show the extension of the power-law fits on scales $<0.1\,h^{-1}$ Mpc and $>20\,h^{-1}$ Mpc.

corresponding errors as well as the quality of the fits and the correlation length are estimated as in Section 5.1 and summarized in Tables 4 and 5. As noted earlier, the power-law fits describe the clustering measurements quite well in a qualitative sense, although certainly not well enough in a quantitative sense, with most samples typically presenting too large a reduced χ^2 (see Tables 4 and 5).

For all absolute magnitude ranges, the red population displays a steeper correlation function slope than the blue one. Blue galaxies have a much shallower slope, which gradually decreases with luminosity until a sudden increase in the slope for the faintest luminosity range probed (Table 5).

The correlation length of red galaxies for all redshift bins presents a minimum value around M^*, with increasing values both faintwards and brightwards (Table 4). We note, however, that this result comes with large uncertainties. For red galaxies, the correlation lengths of the brightest and faintest bin are comparable and faint red objects are more strongly clustered than red objects with intermediate luminosities. For the blue population, r_0 behaves more regularly (like the overall population), gradually decreasing with luminosity and redshift. Blue galaxies generally have smaller uncertainties as well. Our measurement of the correlation length for the faintest luminosity bin ($r_0 = 4.17 \pm 1.41\,h^{-1}$ Mpc) indicates that these galaxies are clustered similarly to blue galaxies of intermediate luminosity. The robustness of this result and some caveats are discussed in Section 5.3.

Due to the complicated way that the slope and the correlation length, as well as their respective uncertainties, change between colour-selected samples, we chose to study the clustering of these samples more quantitatively using the relative bias, i.e. their clustering with respect to the L^* sample. Our relative bias results for all samples, selected by photometric redshift, absolute luminosity and colour, are presented in Section 6.1.

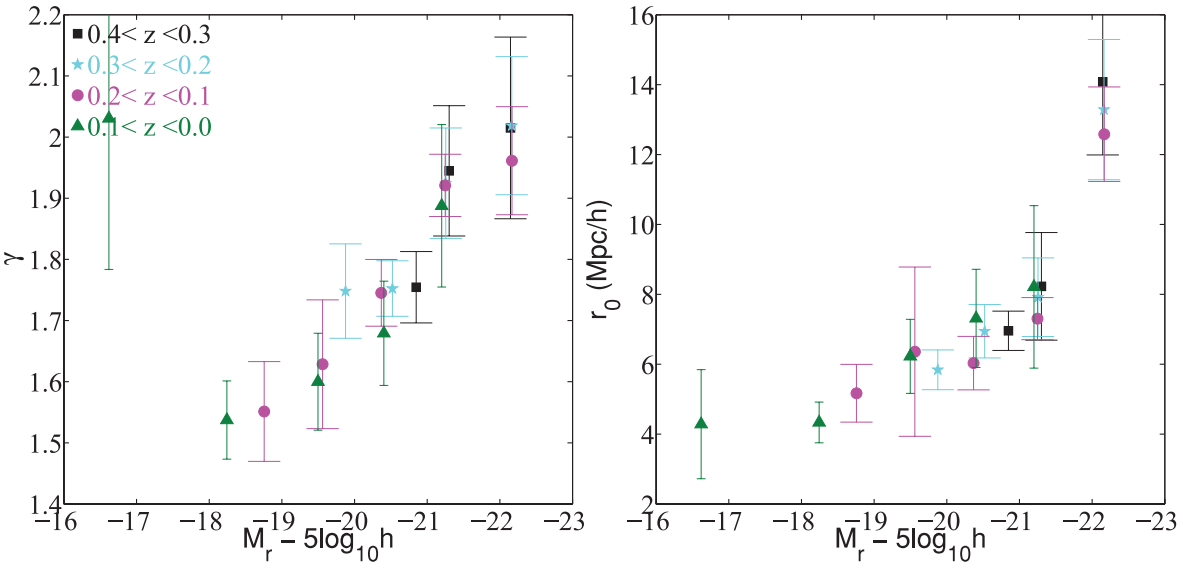

Figure 9. Left: power-law slope, γ, as a function of absolute magnitude and redshift. Right: real-space correlation length, r_0, as a function of absolute magnitude and redshift. Absolute magnitude ranges for which r_0 and γ measurements are valid are given in Table 3.

Table 4. Clustering properties of luminosity-selected red galaxies. Columns are the same as in Table 3.

Sample $M_r - 5\log h$	Magnitude$^{(MC)}$ $M_r - 5\log h$	N_{gal}	γ	r_0 [h^{-1} Mpc]	χ^2_ν	$\gamma^{(d)}$	$r_0^{(d)}$ [h^{-1} Mpc]	$\chi^{(d)2}_\nu$	b/b^* [h^{-1} Mpc]
			Red $0.3 < z_{phot} < 0.4$						
$[-24, -22)$	$-22.0^{-0.2}_{+0.2}$	13095	2.02 ± 0.15	13.91 ± 2.22	3.01	2.03 ± 0.11	13.65 ± 1.86	2.4	1.78 ± 0.26
$[-22, -21)$	$-21.2^{-0.3}_{+0.3}$	287622	1.98 ± 0.10	8.40 ± 1.64	24.60	1.94 ± 0.10	8.71 ± 1.17	13.7	1.06 ± 0.20
$[-21, -20)$	$-20.7^{-0.2}_{+0.2}$	79073	1.86 ± 0.05	8.19 ± 0.54	1.33	1.88 ± 0.05	8.08 ± 0.40	1.2	1.00 ± 0.01
			Red $0.2 < z_{phot} < 0.3$						
$[-24, -22)$	$-22.0^{-0.3}_{+0.3}$	12200	2.02 ± 0.11	13.33 ± 1.95	1.89	2.01 ± 0.07	13.24 ± 1.11	1.8	1.73 ± 0.41
$[-22, -21)$	$-21.2^{-0.4}_{+0.3}$	242452	1.95 ± 0.10	8.26 ± 1.31	11.23	1.92 ± 0.06	8.41 ± 0.72	6.0	1.05 ± 0.25
$[-21, -20)$	$-20.5^{-0.3}_{+0.4}$	597678	1.81 ± 0.06	8.01 ± 1.20	17.10	1.84 ± 0.06	7.69 ± 0.52	6.5	0.98 ± 0.04
$[-20, -19)$	$-19.8^{-0.3}_{+0.3}$	44588	1.95 ± 0.09	8.53 ± 1.30	5.59	1.91 ± 0.08	8.57 ± 0.43	2.8	1.07 ± 0.21
			Red $0.1 < z_{phot} < 0.2$						
$[-24, -22)$	$-22.0^{-0.4}_{+0.3}$	4271	1.96 ± 0.08	12.61 ± 1.26	0.47	1.95 ± 0.08	12.57 ± 1.13	0.4	1.87 ± 0.48
$[-22, -21)$	$-21.2^{-0.4}_{+0.5}$	93975	1.94 ± 0.05	7.56 ± 0.71	2.52	1.93 ± 0.04	7.65 ± 0.36	1.6	1.13 ± 0.28
$[-21, -20)$	$-20.3^{-0.5}_{+0.5}$	393344	1.78 ± 0.11	7.07 ± 1.81	17.30	1.84 ± 0.08	6.68 ± 0.64	6.3	1.03 ± 0.10
$[-20, -19)$	$-19.5^{-0.4}_{+0.5}$	344815	1.71 ± 0.20	9.69 ± 5.98	82.81	1.85 ± 0.12	8.19 ± 1.26	16.9	1.33 ± 0.66
$[-19, -17)$	$-18.7^{-0.4}_{+0.5}$	12942	1.86 ± 0.18	17.86 ± 4.26	9.69	1.84 ± 0.14	17.72 ± 2.88	4.6	2.46 ± 0.83
			Red $0.0 < z_{phot} < 0.1$						
$[-22, -21)$	$-21.1^{-0.7}_{+0.9}$	18631	1.90 ± 0.14	8.20 ± 2.62	5.97	1.88 ± 0.07	8.14 ± 0.78	1.7	0.96 ± 0.47
$[-21, -20)$	$-20.4^{-0.7}_{+0.9}$	83541	1.71 ± 0.11	8.82 ± 2.34	10.98	1.79 ± 0.07	7.90 ± 0.76	3.2	0.97 ± 0.29
$[-20, -19)$	$-19.5^{-0.6}_{+0.8}$	45541	1.77 ± 0.16	10.41 ± 3.89	19.29	1.85 ± 0.14	10.39 ± 1.66	8.1	1.15 ± 0.46
$[-19, -17)$	$-18.7^{-0.5}_{+0.7}$	6690	1.88 ± 0.13	11.59 ± 2.82	2.65	1.90 ± 0.09	11.77 ± 1.32	1.0	1.43 ± 0.51

5.3 Clustering of faint blue galaxies

One of the aims of this paper is to study the clustering of intrinsically faint galaxies for which only photometric redshifts are available in sufficient numbers to calculate $w(\theta)$ reliably. The GAMA depth and the extensive SDSS sky coverage allow us to measure the auto-correlation function of the faintest optically selected galaxies, i.e. those with photo-z estimated absolute magnitudes in the $-17 < M_r - 5\log h < -14$ range and $z_{phot} < 0.08$. This faint sample contains a total of 14 659 galaxies, which are mostly star-forming (as evident by their colours). From the subset with spectroscopic redshifts, the 68th-central percentile of the actual absolute magnitude distribution covers the range $-18 < M_r - 5\log h < -12.7$. However, as shown in Appendix B4, this sample suffers from an overall 50 per cent contamination, with most spurious objects arising from local, over-deblended spiral galaxies.

The upper panel of Fig. 11 shows the correlation functions of all galaxies in our sample with $z_{phot} < 0.08$ split into finer

Table 5. Clustering properties of luminosity-selected blue galaxies. Columns are the same as in Table 3.

Sample $M_r - 5\log h$	Magnitude$^{(MC)}$ $M_r - 5\log h$	N_{gal}	γ	r_0 [h^{-1} Mpc]	χ^2_ν	$\gamma^{(d)}$	$r_0^{(d)}$ [h^{-1} Mpc]	$\chi^{(d)2}_\nu$	b/b^* [h^{-1} Mpc]
			Blue $0.3 < z_{phot} < 0.4$						
$[-22, -21)$	$-21.2^{-0.3}_{+0.3}$	52212	1.71 ± 0.07	6.88 ± 0.47	0.78	1.72 ± 0.07	6.87 ± 0.38	0.6	1.14 ± 0.12
$[-21, -20)$	$-20.8^{-0.2}_{+0.3}$	79787	1.75 ± 0.06	5.86 ± 0.49	1.52	1.75 ± 0.10	5.83 ± 0.44	1.3	1.00 ± 0.01
			Blue $0.2 < z_{phot} < 0.3$						
$[-22, -21)$	$-21.2^{-0.3}_{+0.3}$	42517	1.74 ± 0.11	6.42 ± 0.81	3.05	1.75 ± 0.12	6.46 ± 0.57	1.5	1.17 ± 0.14
$[-21, -20)$	$-20.4^{-0.4}_{+0.4}$	332861	1.63 ± 0.06	5.35 ± 0.48	4.08	1.66 ± 0.05	5.23 ± 0.23	2.6	0.99 ± 0.01
$[-20, -19)$	$-19.8^{-0.3}_{+0.3}$	78282	1.72 ± 0.09	5.08 ± 0.47	1.69	1.72 ± 0.09	4.88 ± 0.34	1.2	0.95 ± 0.11
			Blue $0.1 < z_{phot} < 0.2$						
$[-22, -21)$	$-21.1^{-0.4}_{+0.4}$	12753	1.85 ± 0.13	5.70 ± 0.83	0.86	1.85 ± 0.16	5.67 ± 0.64	0.6	1.22 ± 0.17
$[-21, -20)$	$-20.3^{-0.5}_{+0.5}$	210837	1.67 ± 0.07	4.43 ± 0.32	3.54	1.70 ± 0.06	4.44 ± 0.25	2.6	0.98 ± 0.35
$[-20, -19)$	$-19.4^{-0.5}_{+0.5}$	571748	1.57 ± 0.08	4.75 ± 0.73	11.72	1.62 ± 0.09	4.45 ± 0.42	6.9	1.04 ± 0.14
$[-19, -17)$	$-18.6^{-0.4}_{+0.6}$	198394	1.53 ± 0.06	4.50 ± 0.49	2.26	1.56 ± 0.06	4.31 ± 0.23	1.2	1.00 ± 0.10
			Blue $0.0 < z_{phot} < 0.1$						
$[-21, -20)$	$-20.3^{-0.7}_{+0.9}$	39246	1.61 ± 0.14	4.84 ± 0.82	6.52	1.65 ± 0.13	4.66 ± 0.31	3.2	0.97 ± 0.10
$[-20, -19)$	$-19.3^{-0.7}_{+0.9}$	109606	1.53 ± 0.06	4.63 ± 0.45	2.42	1.57 ± 0.07	4.45 ± 0.40	2.4	0.94 ± 0.21
$[-19, -17)$	$-18.1^{-0.8}_{+1.0}$	264699	1.54 ± 0.08	4.16 ± 0.63	7.29	1.58 ± 0.11	3.85 ± 0.30	4.4	0.86 ± 0.22
$[-17, -14)$	$-16.6^{-0.9}_{+1.3}$	14305	2.02 ± 0.23	4.17 ± 1.41	5.05	1.99 ± 0.28	4.34 ± 1.00	2.1	0.82 ± 0.33

luminosity bins than used previously. There exists a seemingly artificial steepening of $w(\theta)$ on scales $\theta < 0.^\circ1$ for galaxies with $M_r - 5\log h > -17$. In the bottom panel of Fig. 11, we further split the $-17.9 < M_r - 5\log h < -14$ range into two finer luminosity bins and again we find that for fainter samples source contamination affects larger angular scales. We study this contamination and quantify it as a function of scale in Appendix B4.

Having established the angular scales over which we trust our $w(\theta)$ measurements, we proceed to the clustering analysis. Using only the diagonal elements of the covariance matrix,[7] we note that a power law describes the clustering signal rather well, even though there is a hint of an increase in the clustering strength at $\sim 1\,h^{-1}$ Mpc. It is possible that this increase is due to blue galaxies that are satellites in small dark matter haloes. These haloes should not be dense enough to stop star formation and thus we observe only blue galaxies in this luminosity range (Eminian 2008). A recent detailed study of the star-formation history of Hα-selected faint blue galaxies in GAMA can be found in Brough et al. (2011).

In conclusion, the angular clustering for the faintest sample has a spurious amplitude at small angular scales, unless one takes into account the sample contamination. We do this in Appendix B4, where we visually inspect $\sim$10 per cent of the objects in this sample and find that a significant fraction of them are spurious, mainly due to poorly deblended sources. We quantify the effect of this contamination in Appendix B4 for all luminosity bins. This investigation reveals that the angular clustering results on scales $\lesssim 0.^\circ1$ are not trustworthy enough to be considered reliable. We note that the power-law fits are performed on larger scales, which we show are unaffected by this contamination. However, much more detailed investigation of the data is required to confirm robustly the observed increase in the slope of the correlation function. Finally, we note that we have repeated the analysis presented in this section for objects selected from the most recent SDSS release, DR8 (Aihara et al. 2011), and we observe no differences in the results. The contamination from over-deblended spiral galaxies is still present in DR8 for the low-luminosity bin.

[7] Use of diagonal covariance elements only is appropriate for this faint sample, as it covers a rather small volume for which JK resampling is unable to provide an accurate description of the full covariance matrix.

5.4 Quality of fits and the halo occupation distribution formalism

The power-law fits presented in Table 3 are not all satisfactory in a quantitative sense. The angular correlation function is only well-described to first order by a power law. The rather high reduced χ^2 values for some samples are due either to underestimated errors or to the power-law model being inadequate in describing the angular correlation function over a large range of scales. From the test of Section 3.5, we conclude that the JK method gives consistent errors irrespective of the way we define the jack-knife regions and therefore it is most likely that the large reduced χ^2 values are due to a limitation in the power-law model rather than in the error estimates themselves.

A more sophisticated model, like the halo occupation distribution (HOD) model (for a review see Cooray & Sheth 2002), would provide a more physically motivated description of the full correlation function shape, as a function of both colour and luminosity (Zehavi et al. 2004, 2005, 2011; Zheng et al. 2005). The HOD framework, as shown by Zehavi et al. (2005), explains the increase of clustering in the faint red population. Bright red galaxies are central galaxies in massive haloes, whereas faint red galaxies are satellite galaxies in massive haloes. Our measurements suggest that both bright and faint red galaxies are more strongly clustered than red galaxies with intermediate luminosity. We also observe a bump in the angular correlation function of red galaxies at separations $\sim 1\,h^{-1}$ Mpc, which signals the transition (change in slope) between the one-halo and two-halo terms in the correlation function. In contrast, such a change in slope is not evident for the blue population, hence they have a

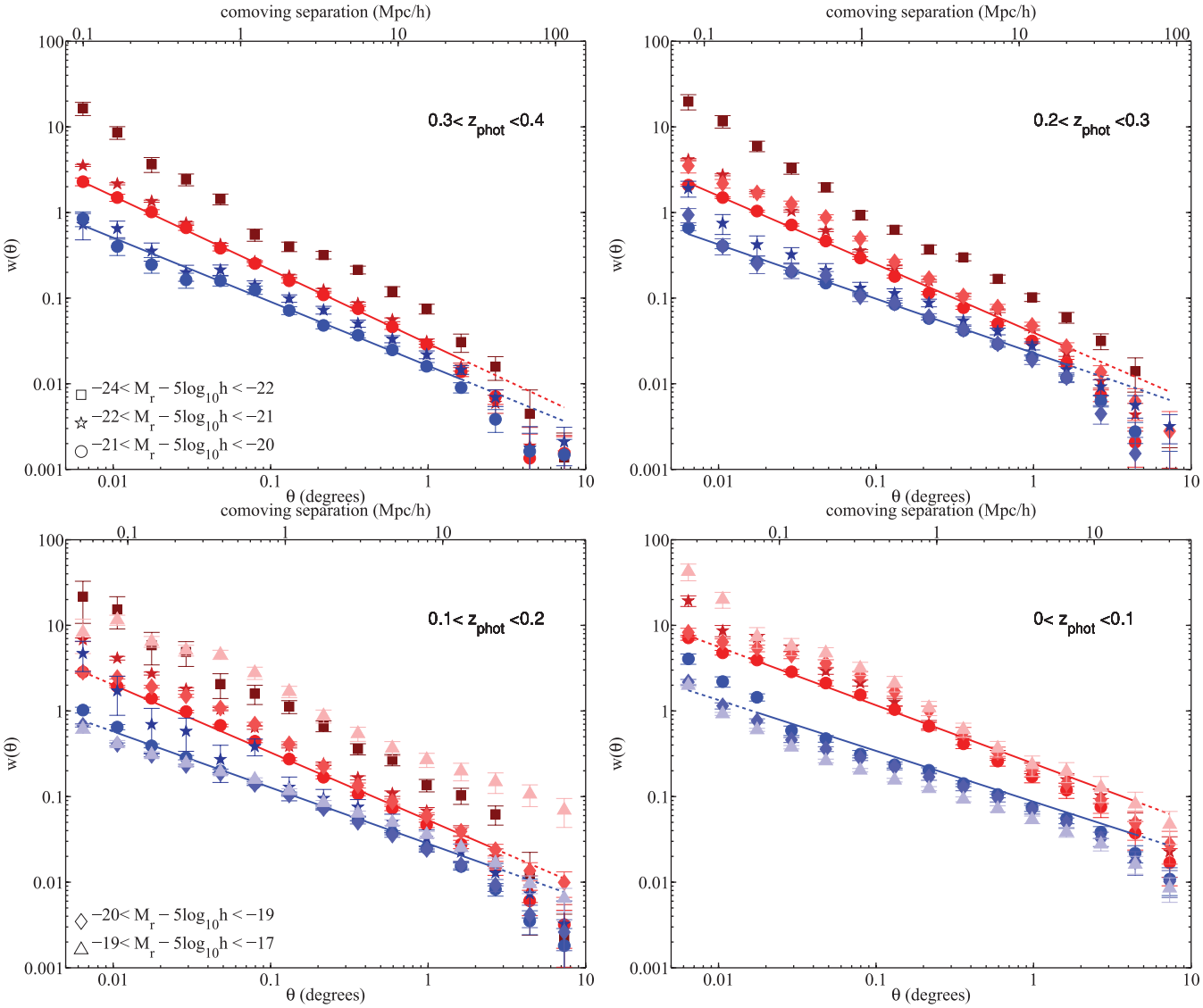

Figure 10. Two-point angular correlation functions $w(\theta)$ split by absolute magnitude and colour, with red circles (blue squares) showing the red (blue) sample. Colour gradients indicate the transition from bright (darker shade) to faint (lighter shade) luminosities. Lines are as in Fig. 8. The faintest (brightest) sample does not contain enough red (blue) galaxies to estimate $w(\theta)$ robustly.

smaller χ^2_ν. This is also in agreement with HOD predictions, which predict a simple power law for blue galaxies with luminosities $M_r - 5\log h < -21$ (Zehavi et al. 2005). A complete HOD modelling of these angular clustering results with photometric redshifts is beyond the scope of the present work, as this would require photo-z dedicated HOD tools to be developed as the standard threshold samples cannot be defined.

6 BIAS MEASUREMENTS

6.1 Relative bias and comparison with previous studies

In this paper we parametrize the real-space correlation function with a power law and infer $\xi(r)$ from angular clustering measurements via a Limber inversion. To ease comparison with samples using a similar, but not identical, selection, we follow Norberg et al. (2002) and define the relative bias of a class of galaxies i with respect to our L^* ($-21 < M_r - 5\log h < -20$) sample as

$$\frac{b_i}{b^*}(r) = \sqrt{\frac{(r_0^i)^{\gamma_i}}{r_0^{\gamma}} r^{\gamma-\gamma_i}}. \quad (14)$$

Equation (14) preserves any scale dependence for samples with different slopes and we choose here to estimate the relative bias at $r = 5\,h^{-1}$ Mpc. The advantage of using this definition of relative bias instead of the raw correlation length for comparison with other studies is twofold. First, the former uses the slope as well as the correlation length, which as we know from equation (6) is strongly correlated. Secondly, if the sample selections are just slightly different, the relative bias is a much more robust way of comparing them as it measures deviations from a series of appropriate reference samples. In this study this is particularly important, as photo-z-inferred properties are not straightforwardly related to the underlying ones, as shown in Section 4.3. Our results are shown in Fig. 12.

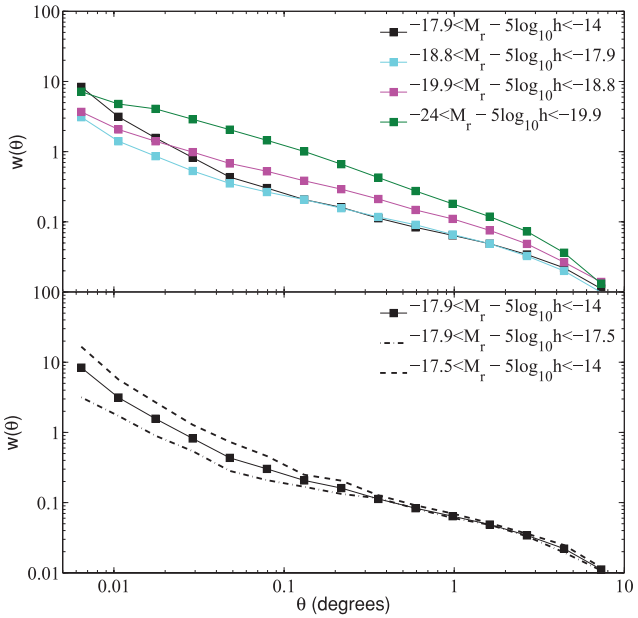

Figure 11. Angular correlation functions for the low-redshift galaxies in our sample split into luminosity bins. The finer luminosity binning allows one to track the scales on which contamination effects (studied and quantified in Appendix B4) are significant. Error bars have been omitted for clarity.

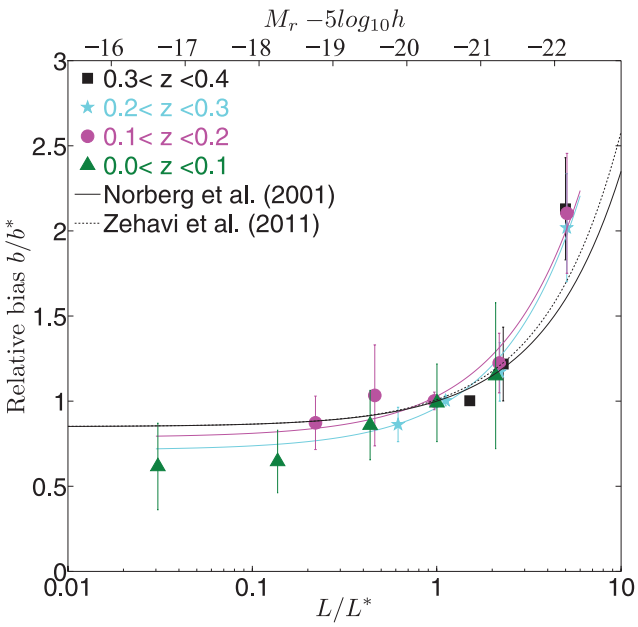

Figure 12. The relative bias, defined in equation (14), at separations $r = 5\,h^{-1}$ Mpc, of all the absolute-magnitude-selected samples used in this study. Data points show the mean and errors of b/b^* obtained from the distribution of 80 jack-knife measurements (Section 3.5) appropriately scaled to account for jack-knife correlations. Cyan and magenta lines show our fits over the redshift ranges $0.2 < z_{\rm phot} < 0.3$ and $0.1 < z_{\rm phot} < 0.2$ respectively. The solid black line shows the fit of Norberg et al. (2001) and the dotted line the fit of Zehavi et al. (2011).

Previous studies from both 2dFGRS (Norberg et al. 2001, 2002) and SDSS (Zehavi et al. 2002, 2005, 2011) have established that the relative bias, b/b^*, as a function of relative luminosity, L/L^*, is well described by an affine relation. We compare our results with these studies in Fig. 12. For all luminosity bins given in Table 3 we

Table 6. Fitted values of a_0 and a_1 in the bias–luminosity relation (equation 15) in three photo-z ranges. Column 1 lists the redshift-bin limits, columns 2, 3 and 4 the fitted values and the quality of fit (reduced χ^2) and column 5 lists $\Delta\chi^2$ between our best-fitting values and the fit by Norberg et al. (2001).

Redshift range	a_0	a_1	χ^2_ν	$\Delta\chi^2$
	All colours			
$0.2 < z_{\rm phot} < 0.3$	0.71 ± 0.04	0.25 ± 0.02	1.10	2.32
$0.1 < z_{\rm phot} < 0.2$	0.82 ± 0.06	0.24 ± 0.03	0.14	1.79
$0.0 < z_{\rm phot} < 0.1$	0.65 ± 0.05	0.27 ± 0.06	0.12	1.18
	Red			
$0.2 < z_{\rm phot} < 0.3$	0.92 ± 0.17	0.12 ± 0.07	0.36	0.29
$0.1 < z_{\rm phot} < 0.2$	1.28 ± 0.43	0.03 ± 0.17	2.33	1.76
	Blue			
$0.2 < z_{\rm phot} < 0.3$	0.84 ± 0.08	0.15 ± 0.06	0.29	0.77
$0.1 < z_{\rm phot} < 0.2$	0.98 ± 0.07	0.08 ± 0.06	0.23	4.22
$0.0 < z_{\rm phot} < 0.1$	0.86 ± 0.02	0.08 ± 0.02	0.07	0.02

fit the equation

$$b/b^* = a_0 + a_1 L/L^*, \quad (15)$$

where a_0 and a_1 are free parameters. Our best-fitting values for samples selected on luminosity, colour and photo-z, using the corresponding L^* for each sample, are given in Table 6. The high-redshift bin only provides three data points and thus we do not include it in this exercise (black squares in Fig. 12). In this table, we also compare with the bias relation of Norberg et al. (2001), who found $(a_0, a_1) = (0.85, 0.15)$. The $\Delta\chi^2$ between our best fit and that of Norberg et al. is 1.2–2.3, which makes the fits statistically compatible, as the 68 per cent confidence interval for 2 degrees of freedom corresponds to $\Delta\chi^2 = 2.31$ (Press et al. 1992). Zehavi et al. (2011) measured the bias relative to dark matter, and in Fig. 12 we rescale their relation with respect to L^*. They also observed a steeper rise in relative bias at high luminosities. Including a power of (L/L^*) in our fit, we also obtain a steeper slope whilst χ^2 remains unchanged, despite the additional degree of freedom.

For samples selected by colour as well as luminosity, it is more difficult to fit equation (15) into each redshift bin. For most photo-z bins we have four or fewer data points. Moreover, using finer luminosity bins would worsen the statistical errors in $N(z)$ and $N(M_r)$ and thus make any fit more difficult to interpret. Fig. 13 shows that the blue population follows a similar trend to the full sample but the relative bias changes more smoothly as a function of luminosity. Table 6 gives the values of a_0 and a_1 for the colour-selected samples. We fit the same linear relation for red galaxies as well, despite the fact that a quadratic function would seem more appropriate. χ^2 values for the linear fit are also shown in Table 6 and, from a purely statistical point of view, a linear relation between b/b^* and L/L^* is still acceptable. Fig. 13 shows that the statistical uncertainty for the two faint red samples is quite large. This is due to the small number of objects in the $-19 < M_r - 5\log h < -17$ sample and the poor quality of fit for the $-20 < M_r - 5\log h < -19$ sample.

6.2 The evolution of absolute bias for L^* galaxies

In Section 6.1 we calculated the relative galaxy bias using the L^* sample ($-21 < M_r - 5\log h < -20$) as our reference sample. In this section we calculate the absolute bias of the L^* population defined as the mean ratio of the observed galaxy correlation function,

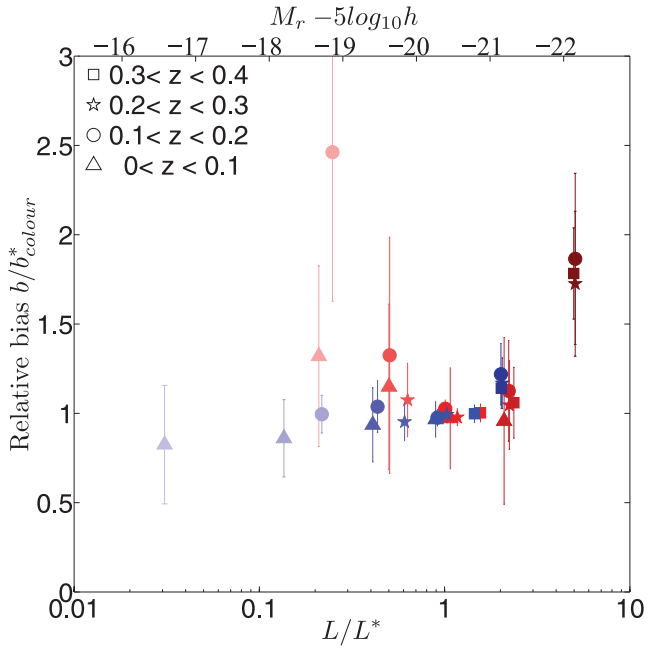

Figure 13. The relative bias, defined in equation (14), at separations $r = 5\,h^{-1}$ Mpc, of all the samples used in this study split by colour (equation 12). Data points show the mean and errors of b/b^* obtained from the distribution of 80 jack-knife measurements (Section 3.5) appropriately scaled to account for jack-knife correlations. Colour coding is as in Fig. 10.

parametrized with a *power law*, over the non-linear dark matter theoretical correlation function

$$b^*(r) = \sqrt{\frac{\xi_{\rm GG}(r)}{\xi_{\rm DM}(r)}} = \sqrt{\frac{(r_0^*)^{\gamma^*}}{r^{\gamma^*}\xi_{\rm DM}(r)}}, \quad (16)$$

where $5\,h^{-1}$ Mpc $< r < 20\,h^{-1}$ Mpc. The theoretical power spectrum $P(k)$ was obtained using `CAMB` (Lewis, Challinor & Lasenby 2000) and the halo-correction recipe of Smith et al. (2003). We then Fourier-transform the non-linear $P(k)$ to obtain the real-space $\xi_{\rm DM}(r)$ using the `FFTLog` package provided by Hamilton (2000).

Since we have correlation-function measurements of the L^* population for a range of redshifts, we can answer the question of whether the evolution of the bias can be described by the passive evolution model introduced by Tegmark & Peebles (1998):

$$[b(z_1) - 1]D(z_1) = [b(z_2) - 1]D(z_2), \quad (17)$$

where D is the growth of structure (Peebles 1980), which we calculate accurately using the `growl` package by Hamilton (2001), which includes corrections to $D(z)$ due to the presence of the cosmological constant. The model described by equation (17) assumes that the galaxy density field linearly traces the dark matter density field and all clustering evolution comes from the growth of structure in the linear regime, i.e. no merging. It is believed that L^* galaxies have undergone very little merging since $z \approx 1$ (Conselice, Yang & Bluck 2009; Lotz et al. 2011).

In the upper panel of Fig. 14, we plot the correlation length as a function of redshift. r_0 is observed to change very little since $z \approx 0.32$. The lowest redshift point has larger errors due to the limited volume sampled. For comparisons with theory, it is more lucid to use the bias instead of the correlation length. In the lower panel of Fig. 14 we plot the evolution of the absolute bias, as defined in equation (16), along with the theoretical prediction of Tegmark & Peebles (1998) for passive clustering evolution (dashed line). In

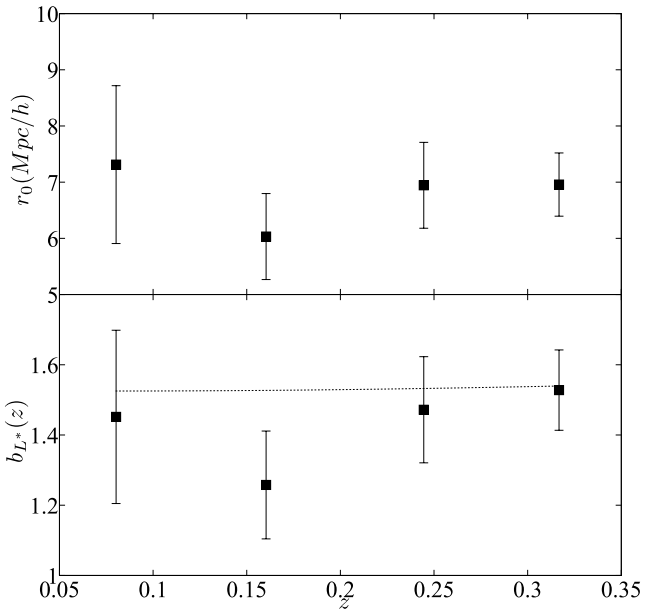

Figure 14. The evolution of clustering of L^* galaxies in the local Universe: the upper panel shows the correlation length r_0, while the lower panel shows the bias $b_{L^*}(z)$ as a function of redshift. The dashed line in the lower panel shows the linear theory prediction from equation (17). Across the redshift range $0.07 < z < 0.32$, the bias of L^* galaxies agrees rather well with the linear theory model.

practice, we fix the high-z value of $b(z)$ and then solve equation (17) over the redshift range $0.07 < z < 0.32$. We find that the evolution of clustering of L^* galaxies is consistent with the model of Tegmark & Peebles (1998).

This agreement between the clustering of L^* galaxies and the passive evolution model was not observed by Ross et al. (2010), who used SDSS photo-z values. The sample selection, modelling of $w(\theta)$ and bias between this study and that of Ross et al. (2010) are very different, as we use GAMA-calibrated photo-z and model the correlation function with a power law, whereas they used SDSS-calibrated photo-z down to $r < 21$ and use halo modelling for the correlation function. Ideally one would expect that the two studies should give consistent results, but it might be that the aforementioned differences in the theoretical modelling and the sample selection influence the results significantly.

7 DISCUSSION AND CONCLUSIONS

Despite their inherent limitations, photometric redshifts offer the opportunity to study the clustering of various galaxy populations using large numbers of objects over a wide range of angular scales with improved statistics, with the caveat that their systematic uncertainties are significantly more complex to deal with. In this section we summarize and discuss the main implications of our results.

Using GAMA spectroscopic redshifts as a training set, we have compiled a photometric redshift catalogue for the SDSS DR7 imaging catalogue with $r_{\rm petro} < 19.4$. We carried out extensive tests to check the robustness of the photo-z estimates and use them for calculating r-band absolute luminosities. We split our sample of 4289 223 galaxies into samples selected on photometric redshift, colour and luminosity and estimate their two-point angular correlation functions. Redshift distributions for the Limber inversion are calculated using Monte Carlo resampling, which we show to be very reliable.

Our clustering results are in agreement with other clustering studies such as those of Norberg et al. (2002) and Zehavi et al. (2011), who used spectroscopic redshifts. We extend the analysis to faint galaxies where photo-z values allow us to obtain representative numbers for clustering statistics. We find that the correlation length decreases almost monotonically towards fainter absolute magnitudes and that the linear relation between b/b^* and L/L^* holds down to luminosities $L \sim 0.03L^*$. For the L^* population, we observe a bias evolution consistent with the passive evolution model proposed by Tegmark & Peebles (1998).

As shown by others (Norberg et al. 2002; Hogg et al. 2003; Zehavi et al. 2005; Swanson et al. 2008a; Zehavi et al. 2011) and confirmed here, the colour dependence is more intriguing because faint red galaxies exhibit a larger correlation length than red galaxies at intermediate luminosities. This trend is explained by HOD models, as shown by Zehavi et al. (2005). Clustering for blue galaxies depends much more weakly on luminosity. We find that at faint magnitudes the SDSS imaging catalogue is badly contaminated by shreds of over-deblended spiral galaxies, which makes interpretation of the clustering measurements difficult. We determine an angular scale beyond which our results are not affected by this contamination, and test this by modelling the scale dependence of the contamination as well as studying its luminosity dependence.

The use of photometric redshifts is likely to dominate galaxy clustering studies in the future. A number of assumptions made in this work might need to be reviewed when we have even better imaging data and training sets. In particular, for cosmology, the non-Gaussianity of photo-z and robust reconstruction of redshift distributions will become a very pressing issue. For galaxy evolution studies, it is essential to study the mapping between a photo-z-derived luminosity range and the true underlying one, as HOD modelling of the galaxy two-point correlation function relies heavily on the luminosity range considered. In this paper, we report only qualitative agreement and leave any HOD study using these photometric-redshift-inferred clustering results for future work.

ACKNOWLEDGMENTS

We thank the anonymous referee for comments and suggestions that significantly improved the paper. LC is financially supported by the Greek State Scholarship Foundation, trustee of the Nik. D. Chrysovergis legacy. CE is grateful to Gavin Dalton and Andrew Liddle for constructive comments about this work during CE's thesis defence. CE was partly supported by the Swiss Sunburst Fund. JL acknowledges support from the Science and Technology Facilities Council, grant numbers ST/F002858/1 and ST/I000976/1. SMC acknowledges the support of an Australian Research Council QEII Fellowship and an J G Russell Award from the Australian Academy of Science. PN acknowledges a Royal Society URF and ERC StG grant (DEGAS-259586).

GAMA is a joint European–Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programmes including GALEX MIS, VST KIDS, VISTA VIKING, WISE, Herschel-ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO and the participating institutions. The GAMA website is http://www.gama-survey.org/.

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

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Aihara H. et al., 2011, ApJS, 193, 29
Baldry I. K. et al., 2010, MNRAS, 404, 86
Benjamin J., van Waerbeke L., Ménard B., Kilbinger M., 2010, MNRAS, 408, 1168
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Blanton M. R. et al., 2003, ApJ, 592, 819
Blanton M. R., Lupton R. H., Schlegel D. J., Strauss M. A., Brinkmann J., Fukugita M., Loveday J., 2005a, ApJ, 631, 208
Blanton M. R. et al., 2005b, AJ, 129, 2562
Brough S. et al., 2011, MNRAS, 413, 1236
Budavári T. et al., 2003, ApJ, 595, 59
Cabré A., Fosalba P., Gaztañaga E., Manera M., 2007, MNRAS, 381, 1347
Cannon R. et al., 2006, MNRAS, 372, 425
Cole S. et al., 2005, MNRAS, 362, 505
Colless M. et al., 2001, MNRAS, 328, 1039
Collister A. A., Lahav O., 2004, PASP, 116, 345
Collister A. et al., 2007, MNRAS, 375, 68
Conselice C. J., Yang C., Bluck A. F. L., 2009, MNRAS, 394, 1956
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Corwin H. G., Jr, Buta R. J., de Vaucouleurs G., 1994, AJ, 108, 2128
Cunha C. E., Lima M., Oyaizu H., Frieman J., Lin H., 2009, MNRAS, 396, 2379
Davis M., Geller M. J., 1976, ApJ, 208, 13
de Vaucouleurs G., de Vaucouleurs A., Corwin H. G., Jr, Buta R. J., Paturel G., Fouque P., 1991, Third Reference Catalogue of Bright Galaxies. Springer-Verlag, New York
Dressler A., 1980, ApJ, 236, 351
Driver S. P., Liske J., Cross N. J. G., De Propris R., Allen P. D., 2005, MNRAS, 360, 81
Driver S. P. et al., 2011, MNRAS, 413, 971
Eminian C., 2008, PhD thesis, University of Sussex
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Goto T., Yamauchi C., Fujita Y., Okamura S., Sekiguchi M., Smail I., Bernardi M., Gomez P. L., 2003, MNRAS, 346, 601
Gott J. R., III, Jurić M., Schlegel D., Hoyle F., Vogeley M., Tegmark M., Bahcall N., Brinkmann J., 2005, ApJ, 624, 463
Guzzo L., Strauss M. A., Fisher K. B., Giovanelli R., Haynes M. P., 1997, ApJ, 489, 37
Hamilton A. J. S., 1993, ApJ, 417, 19
Hamilton A. J. S., 2000, MNRAS, 312, 257
Hamilton A. J. S., 2001, MNRAS, 322, 419
Hamilton A. J. S., Tegmark M., 2004, MNRAS, 349, 115
Hogg D. W. et al., 2003, ApJ, 585, L5

Jones D. H. et al., 2004, MNRAS, 355, 747
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Lilly S. J. et al., 2007, ApJS, 172, 70
Limber D. N., 1953, ApJ, 117, 134
Lotz J. M., Jonsson P., Cox T. J., Croton D., Primack J. R., Somerville R. S., Stewart K., 2011, ApJ, 742, 103
Loveday J., Maddox S. J., Efstathiou G., Peterson B. A., 1995, ApJ, 442, 457
Loveday J. et al., 2012, MNRAS, 420, 1239
Masjedi M. et al., 2006, ApJ, 644, 54
McCracken H. J., Ilbert O., Mellier Y., Bertin E., Guzzo L., Arnouts S., Le Fèvre O., Zamorani G., 2008, A&A, 479, 321
Miller C. J., 1974, Biometrica, 61, 1
Norberg P. et al., 2001, MNRAS, 328, 64
Norberg P. et al., 2002, MNRAS, 332, 827
Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19
Norberg P., Gaztañaga E., Baugh C. M., Croton D. J., 2011, MNRAS, 418, 2435
Oyaizu H., Lima M., Cunha C. E., Lin H., Frieman J., Sheldon E. S., 2008, ApJ, 674, 768
Padmanabhan N. et al., 2008, ApJ, 674, 1217
Parkinson H., 2012, PhD thesis, University of Edinburgh
Peebles P. J. E., 1980, The Large Scale Structure of the Universe. Princeton University Press, Princeton, NJ
Phillipps S., Fong R., Fall R. S. E. S. M., MacGillivray H. T., 1978, MNRAS, 182, 673
Postman M., Geller M. J., 1984, ApJ, 281, 95
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in FORTRAN. The Art of Scientific Computing. Cambridge Univ. Press, Cambridge
Ross A. J., Percival W. J., Brunner R. J., 2010, MNRAS, 407, 420
Ross A. J. et al., 2011a, MNRAS, 1393
Ross A. J., Tojeiro R., Percival W. J., 2011b, MNRAS, 413, 2078
Saunders W. et al., 2004, in Moorwood A. F. M., Iye M., eds, Proc. SPIE Vol. 5492, Ground Based Instrumentation for Astronomy. SPIE, Bellingham, p. 389
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Scranton R. et al., 2002, ApJ, 579, 48
Sharp R. et al., 2006, in McLean I. S., Iye M., eds, Proc. SPIE Vol. 6269, Ground Based and Airborne Instrumentation for Astronomy. SPIE, Bellingham, p. 62690G.
Smith R. E. et al., 2003, MNRAS, 341, 1311
Stoughton C. et al., 2002, AJ, 123, 485
Strauss M. A. et al., 2002, AJ, 124, 1810
Swanson M. E. C., Tegmark M., Blanton M., Zehavi I., 2008a, MNRAS, 385, 1635
Swanson M. E. C., Tegmark M., Hamilton A. J. S., Hill J. C., 2008b, MNRAS, 387, 1391
Tegmark M., Peebles P. J. E., 1998, ApJ, 500, L79
Tegmark M. et al., 2006, Phys. Rev. D, 74, 123507
Wang Y., Yang X., Mo H. J., van den Bosch F. C., 2007, ApJ, 664, 608
Watson D. F., Berlind A. A., McBride C. K., Hogg D. W., Jiang T., 2012, ApJ, 749, 83
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
Zehavi I. et al., 2011, ApJ, 736, 59
Zheng Z. et al., 2005, ApJ, 633, 791

APPENDIX A: SDSS SQL QUERY

The SQL query used to extract our sample from the SDSS DR7 database.

```
SELECT
  objid, g.ra, g.dec, flags, petror50_r,
  petror50Err_r, petror90_r, petror90Err_r,
  petroMag_r - extinction_r as petroMagCor_r,
  petroMagErr_r,
  modelMag_u - extinction_u as modelMagCor_u,
  modelMag_g - extinction_g as modelMagCor_g,
  modelMag_r - extinction_r as modelMagCor_r,
  modelMag_i - extinction_i as modelMagCor_i,
  modelMag_z - extinction_z as modelMagCor_z,
  modelMagErr_u, modelMagErr_g, modelMagErr_r,
  modelMagErr_i,
  modelMagErr_z
FROM galaxy g
JOIN Frame f on g.fieldID = f.fieldID
WHERE
  zoom = 0 and stripe between 9 and 44
  and psfmag_r - modelmag_r > 0.25 and
  petromag_r - extinction_r < 19.4
     AND ((flags_r & 0x10000000) != 0)
     AND ((flags_r & 0x8100000c00a0) = 0)
     PSF_FLUX_INTERP, SATURATED,
     AND (((flags_r & 0x400000000000) = 0) or
     (psfmagerr_r <= 0.2))
     AND (((flags_r & 0x100000000000) = 0) or
     (flags_r & 0x1000) = 0)
```

APPENDIX B: TESTS FOR SYSTEMATICS

Clustering studies using photometric redshifts are subject to systematic errors, which become more pressing as the statistical errors are significantly decreased. In this Appendix we study the most relevant sources of systematic errors that might affect our results. A similar study, for a brighter sample of galaxies at higher redshifts ($0.4 < z < 0.7$), was recently presented by Ross et al. (2011a).

Here we present tests that we believe are more likely to affect the results shown in this paper. We start in Appendix B1 with a scaling test, which mostly tests the reliability of the whole sample for clustering studies. In Appendix B2 we quantify the possible systematics in the clustering signal due to spurious cross-correlations of different photometric redshift bins. In Appendix B3 we test for possible systematics in the spatial correlation function introduced by the redshift distributions used in Limber's equation. Lastly, in Appendix B4 we examine the robustness of the correlation function of the faintest luminosity bin.

B1 Scaling test

With a photometric sample of this size it is prudent to perform a scaling test in order to uncover any dependence of clustering on apparent magnitude. In order to do this we split our sample into apparent magnitude bins and then calculate the angular correlation function. The apparent magnitude ranges are given in Table B1. The angular correlation functions are shown in Fig. B1. For all apparent magnitude bins the slope is approximately equal but the amplitude varies as expected, shifting from high to low values as we go fainter. We then use equation (6) to calculate the correlation length for each magnitude range. We fit over scales of $0.01 < \theta < 2^\circ$ ($0.02 < \theta < 1.2^\circ$ for the $12 < r < 16$ sample). The correlation length for each magnitude bin is found to be equal within the error bars and in

Table B1. Clustering properties in apparent magnitude bins defined by r-band Petrosian magnitude. Column 1 lists the magnitude range, column 2 the number of galaxies, columns 3 and 4 give the values of γ and r_0, defined in equation (5). Column 5 lists the quality of the power-law fits. Errors were calculated using the full covariance matrix, but we do not include the $N(z)$ uncertainty.

r bin (mag)	N_g	γ	r_0	χ^2_ν
$12.0 < r < 16.0$	79543	1.81 ± 0.03	5.01 ± 0.48	1.01
$16.0 < r < 17.0$	201805	1.72 ± 0.02	5.76 ± 0.31	3.1
$17.0 < r < 18.0$	671315	1.73 ± 0.01	5.62 ± 0.20	3.38
$18.0 < r < 18.5$	768620	1.74 ± 0.01	5.58 ± 0.17	2.28
$18.5 < r < 19.0$	1336411	1.73 ± 0.01	5.50 ± 0.12	2.55
$19.0 < r < 19.4$	1720930	1.71 ± 0.01	5.20 ± 0.12	3.48

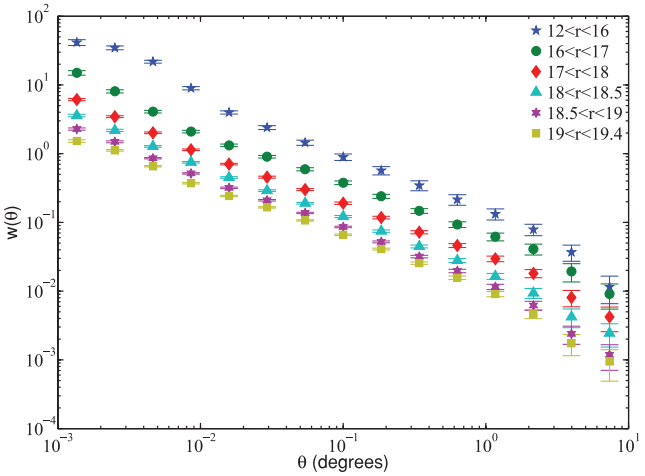

Figure B1. Angular correlation functions of the r-band apparent magnitude bins defined in Table B1.

agreement with the earlier study of Budavári et al. (2003). Thus, for all well-populated apparent magnitude bins we recover the fiducial power law (Peebles 1980)

$$\xi(r) \simeq \left(\frac{r}{5\,h^{-1}\,\mathrm{Mpc}}\right)^{-1.7}. \tag{B1}$$

B2 Cross-correlation of photometric redshift cells

A crucial consistency check, necessary for the validation of our results, is the study of the induced cross-correlations between redshift shells defined by photo-z values from our sample. Since we have established that $\sigma_z \approx 0.04$, we start from $z_{\mathrm{photo}} = 0$ and use five continuous slices with $\Delta z = 0.08$ in order to allow all galaxies with photo-z error of $\lesssim 2\sigma$ to be included in the correct redshift bin. We then cross-correlate slices that are more than one Δz apart.

If a Gaussian with $\sigma = 0.04$ provides a good approximation of the error σ_z, then we can estimate what fraction of galaxies should lie outside the width of each photo-z slice. A galaxy that is outside its redshift slice with width $\Delta z = 0.08$ will have an error greater than 2σ. For a Gaussian distribution, $\sim$5 per cent of all galaxies should lie outside their redshift boundaries. Therefore their residual contribution to the cross-correlation should be $\sim$10 per cent of their auto-correlation.[8] In Fig. B2 we present three auto-correlation functions and their respective cross-correlations. The cross-correlation functions from Fig. B2 are not entirely consistent with zero, but on all scales the residual signal is of the expected order of magnitude. Fig. B2 demonstrates that ANNZ does not produce spurious correlations between physically disjoint galaxies.

B3 Testing dN/dz

Here we test the accuracy of our recovered dN/dz distribution by studying angular clustering in the GAMA area. Since we have precise knowledge of the spectroscopic redshift distributions in the GAMA area, we use these angular clustering measurements to test the robustness of our spatial clustering results using different methods of recovering dN/dz. The methods that we test against the given GAMA spectroscopic redshift distributions are (i) Monte Carlo resampling of the photo-z distributions assuming Gaussian errors (equation 13), which has been used for all the results in this paper, and (ii) the weighting method of Cunha et al. (2009) (also known as nearest neighbour method).

The latter method can be summed up in three distinct steps. First, one estimates the distance in apparent magnitude space to the 200th nearest neighbour of each object in the spectroscopic set, using a Euclidean metric. The exact ordinal number of the neighbouring object should not change the result significantly. For the GAMA number density, $N = 200$ is the best trade-off between smoothing out the large-scale structure and at the same time preserving the locality of the photometric information. Secondly, one calculates the number of objects in the photometric set that are within the hypervolume defined by this distance and then one calculates the weight of each object in the spectroscopic set at point m_i according to the equation

$$w_i = \frac{1}{N_{\mathrm{phot,tot}}} \frac{N(m_i)_{\mathrm{phot}}}{N(m_i)_{\mathrm{spec}}}, \tag{B2}$$

where $N(m_i)_{\mathrm{spec}} = 200$. In the third step, the already known spectroscopic distribution is weighted to match the distribution of the photometric sample. The weighting is done by summing the weights w_i of each object in the spectroscopic sample for all redshift ranges:

$$N(z)_{\mathrm{wei}} = \sum_{i=1}^{N_{\mathrm{spec,tot}}} w_i N(z_1 < z_i < z_2)_{\mathrm{spec}}. \tag{B3}$$

Cunha et al. (2009) show that this method is superior to other methods using photo-z in recovering the true dN/dz, but they do not include Monte Carlo resampling in their comparisons.

The comparison of the different methods is depicted in Fig. B3, where all the clustering measurements are confined to the GAMA area. The errors for the angular clustering measurements are assumed to be Poisson, which is just a lower bound, and the errors in the redshift distributions are obtained from the scatter of Monte Carlo simulations. This test is performed for the same luminosity bins as in Section 4.2, apart from the brightest and faintest bins, which have a very small number of galaxies and hence large statistical errors in $w(\theta)$.

[8] We assume that the two auto-correlations are equal and the number of galaxies in each sample is equal as well. For a detailed treatment of these effects see Benjamin et al. (2010).

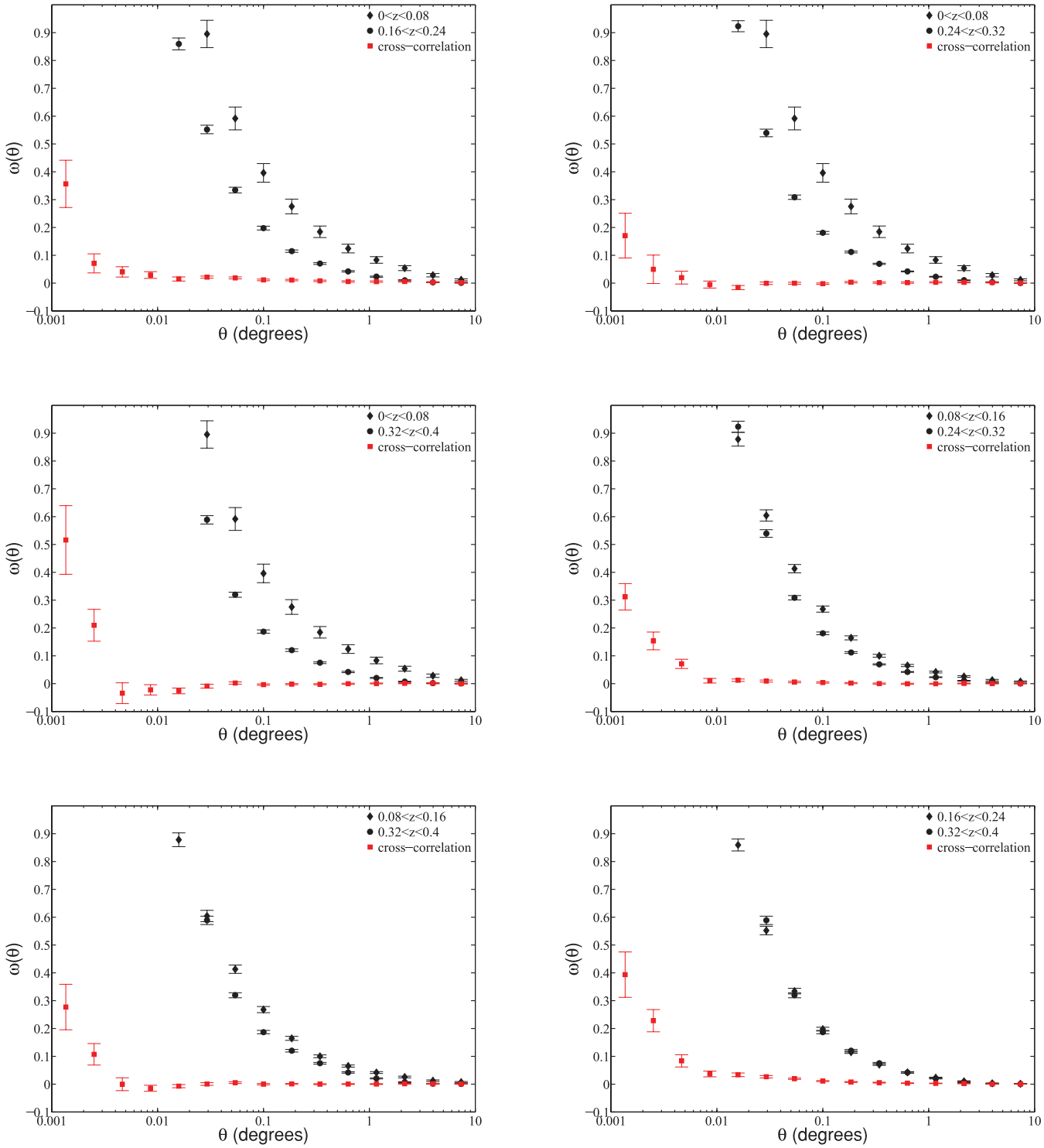

Figure B2. Auto-correlation (diamonds and circles) and cross-correlation (squares) functions for photo-z bins. The cross-correlation signals have negligible magnitude compared with the auto-correlations, and for angular separations $\geq 0.01^\circ$ are consistent with zero. The errors are calculated using JK resampling as explained in Section 3.5.

The (a priori required) agreement between the r_0 measurements from the different methods of recovering dN/dz is not perfect. The r_0 measurements are not significantly affected by the differences between the redshift distributions of Fig. 6. In conclusion, Fig. B3, for the three intermediate and well-populated luminosity bins, implies that the reconstruction of the underlying redshift distribution does not introduce any systematic errors in the r_0 measurements.

This comparison does have its limitations. Samples with small numbers of objects are sensitive to number variations due to the different selections of the two surveys (mainly the more conservative star–galaxy separation that we use in this paper). Moreover, it is

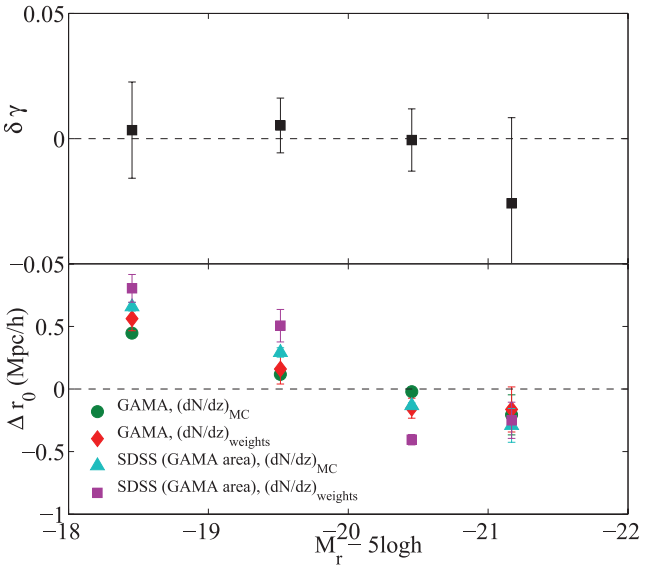

Figure B3. Upper panel: slope residual of correlation-function measurements in the GAMA area, using the measurement of the GAMA sample with spectroscopic redshifts as a reference ($\Delta\gamma = \gamma(\mathrm{SDSS}) - \gamma(\mathrm{GAMA})$). Lower panel: comparison of the effect of the various redshift distributions (as shown in Fig. 6) on r_0 measurements, again using the GAMA sample as a reference ($\Delta r_0 = r_0(\mathrm{i}) - r_0(\mathrm{GAMA})$). Following the discussion in Section 5.1, the error bars show the combined effect of the power-law fit uncertainties (assumed to be Poisson), which are independent of the underlying dN/dz, and the scatter in r_0 due to 100 Monte Carlo resamplings of each dN/dz (only $(\mathrm{d}N/\mathrm{d}z)_{\mathrm{spec}}$ is known precisely).

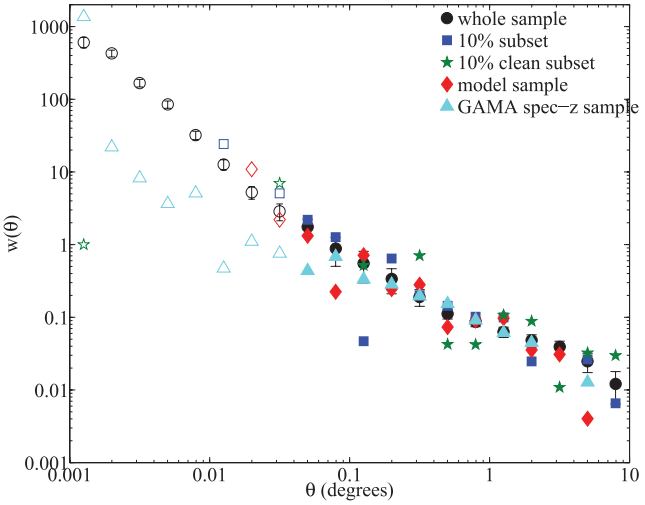

Figure B4. Two-point correlation function of the faintest luminosity bin ($-17 < M_r - 5\log h < -14$). Black circles show the total correlation function, blue squares show the correlation function of the ∼10 per cent subset of objects visually inspected, green stars show the correlation function of the 'clean' part of the inspected subset, red diamonds show the total correlation function corrected to account for spurious pairs on scales $\gtrsim 0.^{\circ}1$ and, finally, cyan triangles show the $w(\theta)$ measurement using only GAMA spectroscopic data. Errors bars for the total sample are calculated using the JK method. Open symbols represent angular scales at which the signal is significantly contaminated and so cannot be trusted.

very difficult to obtain realistic error bars for samples with a small number of galaxies and for which the survey's angular extent is comparable with the angular scales used for the $w(\theta)$ measurements. The difficulty in obtaining the exact angular clustering signal is seen in the upper panel of Fig. B3, which shows the residuals of the measured slopes for the GAMA and SDSS samples. In spite of these, Monte Carlo resampling seems to recover the true r_0 slightly better than the weighting method.

B4 Correlation function for faint galaxies

The correlation function of the faintest sample $[-17, -14)$ exhibits an infeasibly large clustering amplitude at small scales (Fig. B4). This increase in the clustering signal is not hinted at in the $-19 < M_r - 5\log h < -17$ luminosity bin, and so here we investigate whether there is some sort of contamination in the faintest sample.

We randomly select ∼10 per cent of objects in the faintest luminosity bin and visually inspect them to see whether they are genuine galaxies. The fraction of spurious objects is shown in the left panel of Fig. B5 and we observe that it is significant at the very faint end, where the actual number of galaxies is low (red line in the same figure), and ∼40 per cent at the bright end of that luminosity bin. From our visual inspection, most spurious objects are local, over-deblended spiral galaxies; the remainder are merging systems or just sky noise. Evidently, as we go fainter the contamination level is increasing and this presents a serious drawback for clustering studies and a serious limitation for large surveys.

The right panel of Fig. B5 shows the fraction of spurious objects in the other five absolute magnitude bins. We visually inspected ∼100 objects from each of those bins and found that the contamination level is much lower, with a slight increase toward the bright and faint

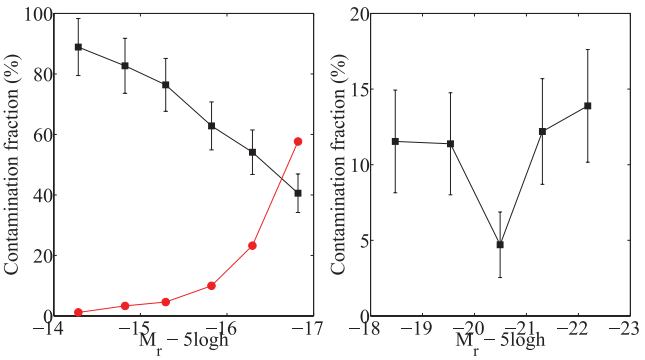

Figure B5. Left panel: black symbols show the fraction of spurious objects for the faintest luminosity bin as a function of absolute luminosity. These fractions are estimated by visually inspecting ∼10 per cent of the total number of objects in that bin. Red symbols show the overall distribution of objects as a function of absolute magnitude. Right panel: fraction of spurious objects as a function of absolute luminosity, obtained by visually inspecting a small subset (∼100) of all objects in each luminosity bin. In both panels the error bars are obtained assuming Poisson statistics.

ends. Our detailed study of the correlation function of the faintest bin shows that it is not affected by contamination on the scales of primary interest ($\theta \gtrsim 0.^{\circ}1$), something that we expect to hold true for all other luminosity bins, as they have a significantly smaller fraction of spurious objects.

The contamination in the $-17 < M_r - 5\log h < -14$ luminosity bin affects the two-point correlation function differently at different angular scales. We address this issue by counting the number of pairs of genuine galaxies in the visually inspected subset. The results are shown in Fig. B4, where we also include the angular correlation

function from the corresponding sample from GAMA.[9] Due to the fact that the subset has a weakened signal at very small scales, we can only draw conclusions for angular scales $>0.^{\circ}1$. From Fig. B4 we see that at these scales the contamination does not significantly affect the correlation function and its fit parameters γ and r_0. For this reason, we present our results limited to angular scales $\theta \gtrsim 0.1^{\circ}$.

[9] GAMA objects have been visually inspected and are therefore more reliable than objects in the SDSS imaging catalogue. On the other hand, GAMA has a smaller area, which increases the statistical errors. For this sample, considering Poisson errors only, the statistical errors in $w(\theta)$ would be at least three times larger than the ones obtained from the SDSS sample.

We also repeated our analysis after masking out areas of sky covered by RC3 galaxies (de Vaucouleurs et al. 1991; Corwin, Buta & de Vaucouleurs 1994) to test whether we could decrease the contamination level. We did not observe any qualitative differences in the power-law parameters estimated and, more importantly, the amplitude of $w(\theta)$ at small scales did not reduce, indicating that the RC3 catalogue does not capture all over-deblended galaxies in the SDSS galaxy catalogue.

Finally, it is important to note (and caution) that source contamination due to over-deblending only became apparent when interpreting the bottom right panels of Figs 8 and 10). Had we completely trusted the results of the scaling test (Appendix B1) or used only the data point near L^* in Fig. B5 (since that population dominates), we would have significantly underestimated the number of spurious objects.

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