

Angular clustering of $z \sim 2$ star-forming and passive galaxies in 2.5 square degrees of deep CFHT imaging

Taro Sato,^{*} Marcin Sawicki[†] and Liz Arcila-Osejo

Department of Astronomy & Physics and the Institute for Computational Astrophysics, Saint Mary's University, 923 Robie Street, Halifax, Nova Scotia, B3H 3C3, Canada

Accepted 2014 July 4. Received 2014 June 20; in original form 2013 August 15

ABSTRACT

We study the angular clustering of $z \sim 2$ galaxies using $\sim 40\,000$ star-forming (SF) and ~ 5000 passively evolving (PE) galaxies selected from ~ 2.5 deg² of deep ($K_{\text{lim}} = 23\text{--}24$ AB) Canada–France–Hawaii Telescope imaging. For both populations, the clustering is stronger for galaxies brighter in rest-frame optical and the trend is particularly strong for PE galaxies, indicating that passive galaxies with larger stellar masses reside in more massive haloes. In contrast, at rest-frame ultraviolet we find that while the clustering of SF galaxies increases with increasing luminosity, it decreases for PE galaxies; a possible explanation lies in quenching of star formation in the most massive haloes. Furthermore, we find two components in the correlation functions for both SF and PE galaxies, attributable to one- and two-halo terms. The presence of one-halo terms for both PE and SF galaxies suggests that environmental effects were producing passive galaxies in virtualized environments already by $z \sim 2$. Finally, we find notable clustering differences between the four widely separated fields in our study; the popular COSMOS field is the most discrepant (as is also the case for number counts and luminosity functions), highlighting the need for very large areas and multiple sightlines in galaxy evolution statistical studies.

Key words: galaxies: formation – galaxies: haloes – galaxies: statistics – cosmology: observations – dark matter – large-scale structure of Universe.

1 INTRODUCTION

A wide array of evidence has led to a general agreement that the lambda cold dark matter (DM) model successfully describes the structure evolution driven by matter–energy distribution, much of which is dominated by observationally inaccessible dark components (e.g. Komatsu et al. 2011). What has become vastly more accessible through observation during the last couple of decades is the luminous part of the Universe, i.e. galaxies. Through large-scale redshift surveys such as SDSS (York et al. 2000), 2dFGRS (Colless et al. 2001), DEEP2 (Newman et al. 2012), and zCOSMOS (Lilly et al. 2007), there is now a wide agreement that star-forming (SF) activity in the Universe peaked at $z \sim 2$ (see e.g. Hopkins & Beacom 2006 for a compilation) – by which time roughly a fifth of the present-day stellar mass was already in place (e.g. Sawicki 2012b) – and the SF activity in the Universe has been steadily declining ever since (e.g. Lilly et al. 1996; Noeske et al. 2007). This overall trend in star formation history (SFH) of galaxies has been made most clear in the form of the Madau plot (e.g. Lilly et al. 1996; Madau et al. 1996; Sawicki, Lin & Yee 1997), showing the luminosity density

as a function of redshift. Constraining the physical mechanism that gives rise to this trend in SFH has been one of the most pressing issues in extragalactic astrophysics.

The relative importance of internal versus external processes on the evolution of galaxies has been extensively debated. While the importance of baryonic physics has been realized early in the canonical models of galaxy formation (e.g. Rees & Ostriker 1977; Silk 1977), the significance of supernovae/active galactic nuclei feedback in the self-regulation of SF activities has been suggested relatively recently (e.g. Dekel & Silk 1986; Bower et al. 2006; Croton et al. 2006). The environmental effects on the properties of galaxies were already remarked in much earlier times (e.g. Hubble 1936) and have been well established ever since the seminal work by Dressler (1980), quantifying the famous morphology–density relation in clusters of galaxies. In the local Universe, quiescent galaxies are found preferentially in denser environments (e.g. Balogh et al. 2004). In the past, the nearest-neighbour approach was popular in defining the environment surrounding galaxies. At high redshifts, however, the method suffers from lack of sufficiently precise redshift measurements, where spectroscopy is expensive and photometric redshifts require a wide and well-sampled multiwavelength baseline for constraining the spectral energy distributions (SEDs).

Another oft-used approach to quantifying galaxy environment is to construct a two-point correlation function. From theories and

^{*} Currently at LinkedIn.

[†] E-mail: sawicki@ap.smu.ca

simulations, the clustering properties of dark haloes have been well quantified (e.g. Mo & White 1996). Observationally, the clustering measurement of luminous galaxies is relatively straightforward through imaging data. This makes possible the comparisons between the clustering properties of luminous and dark components of galaxies, often in the form of galaxy bias, the ratio of galaxy to dark halo distributions. Of particular interest is the connection between luminous galaxies and the mass of their dark halo hosts, which broadly defines the host environment of galaxies. For the most massive dark haloes (i.e. galaxy clusters), mass estimates may be obtained via X-ray emission, gravitational lensing, cluster-member kinematics, and the Sunyaev–Zeldovich effect. For less massive haloes, however, similar methods are less accessible due to observational costs. On the other hand, clustering measurement offers an efficient way of quantifying the relation between galaxies and dark haloes from ever-increasing catalogues of photometric observations. In fact, the method has been so effective and led to the development of the halo occupation distribution framework (e.g. Jing 1998; Ma & Fry 2000; Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002; Cooray & Sheth 2002; Yang, Mo & van den Bosch 2003; Kravtsov et al. 2004; Zheng et al. 2005) to interpret observations, which describes the probability that a dark halo of virial mass M_h hosts N galaxies of a given set of properties.

In the local Universe, the vast amount of observations from large redshift surveys have steadily improved the measurements of galaxy clustering in terms of their intrinsic properties including luminosity, colour, morphology, and SF properties (e.g. Norberg et al. 2001, 2002; Zehavi et al. 2002, 2005, 2011; Budavári et al. 2003; Madgwick et al. 2003; Li et al. 2006; Swanson et al. 2008; Ross & Brunner 2009; Loh et al. 2010; Ross, Percival & Brunner 2010). Their findings are generally consistent with other approaches on galaxy environments, such as nearest-neighbour measurements. The types of galaxies that tend to cluster more strongly (i.e. denser environments) are luminous/massive, bulge-dominated, and redder/quiescent galaxy populations. Such environmental dependence persists to intermediate redshift $z \sim 1$ (e.g. Coil et al. 2004, 2006, 2008; Le Fèvre et al. 2005; Phleps et al. 2006; Pollo et al. 2006; McCracken et al. 2008; Meneux et al. 2008, 2009; Simon et al. 2009; Abbas et al. 2010).

At $z > 1$, where extensive spectroscopy is difficult, a wide array of photometric colour selection techniques have been devised. These rely on prominent breaks in the SED, such as hydrogen Lyman/Balmer breaks to redshift in between broad-band filters, causing substantial colour changes. The Lyman-break dropout technique (Guhathakurta, Tyson & Majewski 1990; Steidel et al. 1996), for example, made it feasible to cull a large number of SF $z \sim 3$ galaxies purely through photometry. Due to the popularity of the technique, the clustering of Lyman-break galaxies has been well studied (e.g. Adelberger et al. 2005; Ouchi et al. 2005; Lee et al. 2006; Hildebrandt et al. 2009; Savoy et al. 2011; Bielby et al. 2012), showing that clustering strength tends to increase with the ultraviolet (UV) luminosity of galaxies. However, the galaxies selected in rest-frame UV are biased towards SF systems with little dust, and the characterization cannot be extended to wider populations. Various colour selection methods in similar spirit have been devised to target $z > 1$ galaxies: EROs (e.g. Elston, Rieke & Rieke 1988; McCarthy, Persson & West 1992; Hu & Ridgway 1994; Thompson et al. 1999; McCarthy 2004), DRGs (Franx et al. 2003), and BzK (Daddi et al. 2004). These colour selection techniques are efficient, requiring only a few photometric bandpasses, but in general select biased populations, making difficult the comparisons between samples se-

lected differently (e.g. Reddy et al. 2005; Grazian et al. 2007; Lane et al. 2007).

The BzK selection technique (Daddi et al. 2004) has been well tested and employed in several studies of $z \sim 2$ galaxy clustering (e.g. Kong et al. 2006; Hayashi et al. 2007; Blanc et al. 2008; Hartley et al. 2008; McCracken et al. 2010; Ly et al. 2011; Fang et al. 2012; Lin et al. 2012). The popularity of the method comes in part from its ability to construct relatively complete samples of SF and passively evolving (PE) galaxies at $z \sim 2$ without regard for dust reddening. Typically, a magnitude-limited source catalogue is constructed from the K -band image. While sampling redder light in near-infrared is desirable for tracing stellar mass reliably, the lack of deep, wide-field K -band imaging has been a bottleneck for BzK clustering surveys in the past. Furthermore, a clustering measurement from a small field of view suffers from the variance in observation inherent in choosing a particular sightline in the universe (i.e. cosmic variance). A common approach to combat this is to observe a few independent sightlines and/or bootstrap a set of regions within an image to simulate a different set of observations. The wide-field near-infrared imagers [e.g. Kitt Peak National Observatory (KPNO)/NOAO Extremely Wide-Field Infrared Imager (NEW-FIRM) and Canada–France–Hawaii Telescope (CFHT)/Wide-field InfraRed Camera (WIRCam)] of the current generation, however, have been making data accessible lately to lessen these issues.

The visible/near-infrared imaging data from the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS) offer desirable features for clustering measurements. In visible, the CFHT/MegaPrime imaging consists of four independent fields, each being a 1 square degree image, widely separated from each other on the sky. The near-infrared imaging data from the CFHT/WIRCam are mosaics of 21.5 arcmin square images, which make the effective survey areas smaller, but are comparably deep as the current generation of K -limited surveys. These data allow us to study the clustering properties of $z \sim 2$ galaxies in unprecedented scale in terms of number of objects and the survey area. The original BzK technique by Daddi et al. (2004) employs the $(B - z) - (z - K)$ colour–colour plane to separate $z \sim 2$ galaxies from low- z interlopers. We use a revised colour–colour criteria designed to select similar populations at $z \sim 2$ based on $(g - z) - (z - K_s)$ and $(z - H) - (H - K_s)$ colour–colour planes to accommodate the bandpasses differences and the relatively shallow depth of the CFHT g -band data (Arcila-Osejo & Sawicki 2013). In this paper, we describe the data set and simulation method, and present the measurements of angular correlation functions for $z \sim 2$ galaxies sampled by a colour–colour selection method that closely resembles the BzK technique. We defer the measurement of spatial correlation functions, as the de-projection of angular correlation functions requires robust estimates of the redshift distributions of the objects in question. We intend to incorporate photometric redshifts in a future paper. A standard concordance cosmology of $(\Omega_m, \Omega_\Lambda) = (0.3, 0.7)$ and $H_0 = 70 \text{ km s}^{-1}$ is assumed. All magnitude photometries are on the AB scale unless otherwise noted.

2 DATA AND CATALOGUE

This study relies on visible/near-infrared imaging obtained at the CFHT. The visible $u^*g'r'i'z'$ data (which we abbreviate to *ugriz* hereafter) come from the sixth release of the CFHTLS Deep, covering four independent 1 deg² fields (labelled D1, D2, D3, and D4) observed with MegaCam on CFHT. Most of these fields overlap with the regions of the sky that have been studied extensively in other surveys, including the COSMOS (field D2) and the Groth

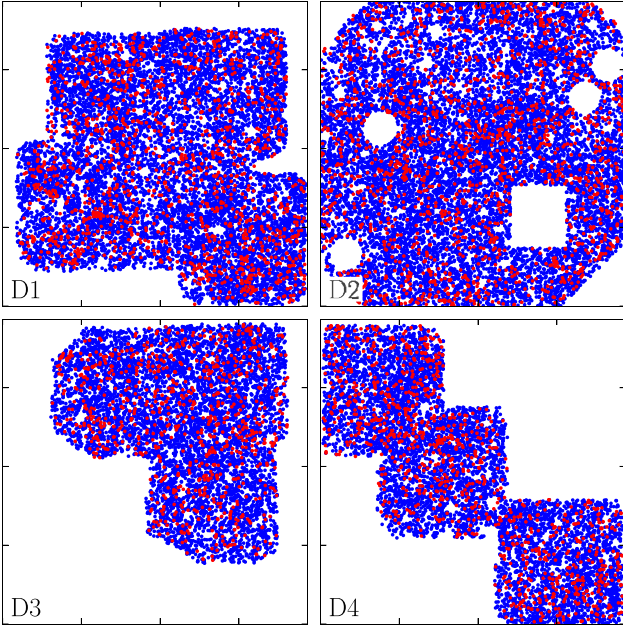


Figure 1. The survey areas showing the distributions of SF (blue) and passive (red) $gzHK_s$ galaxies. Each image extends over 19 354 pixels on the side, which is about 1° (i.e. 0.186 arcsec per pixel). The regions with very poor signal-to-noise ratio (S/N) as well as those close to very bright stars and their diffraction patterns are masked and excluded from our analysis. The peculiar survey boundaries in D1, D3, and D4 are due to mosaicking patterns of the near-infrared pointings.

strip (D3); see e.g. Gwyn (2011) and references therein. The image stacks with the 25 per cent best seeing are used. The near-infrared JHK_s imaging is from the T0002 release of the WIRcam Deep Survey (WIRDS; Bielby et al. 2012). The target fields are subsets of the CFHTLS fields, taken with the CFHT/WIRcam with the exception of the J -band image in D2, for which United Kingdom InfraRed Telescope (UKIRT)/WFCAM data were used. The near-infrared camera has a smaller field of view than MegaCam, and the image for each field generally consists of mosaics. All images were processed by Terapix, astrometrically registered and resampled to a common pixel scale of 0.186 arcsec. The internal astrometric accuracy is ~ 0.05 arcsec between the optical images and ~ 0.1 arcsec with respect to the infrared.

Since the source catalogues are constructed from K_s band images, the survey geometries are largely limited by the exposed regions in K_s , which are generally smaller (except for D2) and not square (Fig. 1). Image masks are generated to flag pixels occupied by stars, diffraction patterns, and other blemishes. Unexposed pixels (based on weight images supplied by Terapix) are also masked. In addition, the regions in which suitable point sources for point spread function (PSF) matching (Appendix A) cannot be found are masked.

For the most part, the data products offered by Terapix are used throughout. To ensure that our fixed-aperture photometry samples a similar physical region for an object in all passbands, however, the PSFs are matched across all the passbands within each field by smoothing (Appendix A). The PSF-matched images are used for colour photometry, yet the source detection is done on unsmoothed images for maximal detectability of faint sources. Object detection, photometry, and the selection of $z \sim 2$ galaxies is described in detail in Arcila-Osejo & Sawicki (2013); below we summarize the main points.

2.1 Photometric measurements

SEXTRACTOR (v2.8.6; Bertin & Arnouts 1996) is used to generate object catalogues. For each field, source detection is carried out on the K_s -band image and then a set of fixed circular aperture photometry is done in the dual mode on sources identified in K_s . A source is considered detected when it extends over the minimum of 5 pixels at the 1.2σ above the sky level. In addition to total magnitude estimates (via `MAG_AUTO`), the colour photometry is done through a 10-pixel diameter circular aperture, which corresponds to 1.86 arcsec projected on the sky. Objects with SEXTRACTOR internal flags of ≤ 4 are considered problem free; objects flagged otherwise are treated as unobserved.

2.1.1 The $gzHK$ colour selection

The objects of our interest are $z \sim 2$ galaxies which are colour selected by a method devised to replicate the BzK technique with the CFHTLS+WIRDS data (Arcila-Osejo & Sawicki 2013), which we summarize here. Such a replication is done by tracing the population synthesis model tracks in the BzK_s and gzK_s colour–colour planes and redefining the demarcation lines in $gzHK_s$, so that the cuts select out similar galaxy populations as in the BzK_s . The distinction between SF and passive galaxies is made in the region where $(z - K_s) > 2.55$, where all objects are considered to be at $z \gtrsim 1.4$. Due to the limited depth in the g photometry (our substitute for B), however, we use supplemental information in near-infrared colours to better constrain their SFH.

For galaxies with $(z - K_s) \leq 2.55$, the ones that satisfy

$$(z - K_s) - 1.27(g - z) \geq -0.022$$

are classified as SF- $gzHK$ galaxies. For objects with $(z - K_s) > 2.55$, SF- $gzHK$ galaxies satisfy the condition

$$(z - H) \leq 2.4(H - K_s) + 1$$

and the rest are considered to be passive (PE- $gzHK$).

We stress that the $gzHK$ technique is tuned to select the same galaxy populations as those selected by the classic Daddi et al. (2004) BzK method. This work is thus fairly similar to most other BzK studies that select high-redshift galaxies after transforming from their filter systems to the classic BzK set. Consequently, the results of our $gzHK$ selection are directly comparable to those from BzK studies by design.

2.1.2 Star–galaxy separation

In the gzK_s plane, stars and galaxies are very clearly separated; this is in fact one of the advantages of the BzK technique. The objects which satisfy the criterion

$$(z - K_s) - 0.45(g - z) \leq -0.57$$

are considered stars. The rest are all considered galaxies. The term all galaxies is used in this paper to refer to the objects which do not satisfy the above criterion, which include SF- $gzHK$ and PE- $gzHK$ galaxies as well as lower- z galaxies.

3 RANDOM OBJECT SIMULATION

A critical component of measuring angular correlation function is to compare the distribution of observed objects to that of the objects whose locations on the sky are uniformly randomized. Since this cannot be done with the real Universe, one runs a set of simulations

to create realizations of the universe from some model, with all underlying parameters tuned to approximate reality. For the purpose of clustering measurements, the object coordinates are uniformly randomized to see if the distribution of observed objects over the same survey area is more clustered compared to that of simulated universe with all objects uniformly distributed. A standard way to achieve this is to draw simulated objects from some priors that are based on theoretical models and/or observations. While the model dependence of this approach makes tractable the analysis and interpretation of results, it has a limitation in that it is a set of those underlying physical properties of the Universe that one often seeks to constrain.

In our simulations, we take a very empirical approach in which simulated objects are constructed to have the same properties as those in the catalogue of observed objects, but with coordinate positions that are uniformly randomized. The method is still model dependent in that the simulated objects are parametrically modelled (Section 3.2) but is adequate in that roughly similar proportions of galaxy types should be simulated to the extent that they are observationally distinguishable. For its simplicity, this approach does not account at all for incompleteness in the observed catalogue, in a sense that a luminosity function calls for.¹ The issue is mitigated by the fact that there is no simple way for carrying out incompleteness-corrected clustering measurements, since data–data pair separations can only be computed from what is observed. So long as the analysis remains on reasonably complete catalogues, the lack of such correction should not render our results useless. The incompleteness in the observed object catalogue and appropriate corrections necessary for simulations are discussed in Section 3.3.

To account for systematics inherent in real observations, the random object catalogue must be subject to the same detection procedure as the one used in the real observations. Briefly, the procedure is as follows: (1) detect sources in an observed image to construct the data catalogue; (2) draw objects randomly from the data catalogue and implant them on to a background image to construct a simulated image; (3) run the same detection process on the simulated image; and finally (4) match the simulated and recovered objects. The (master) random object catalogue is generated from a large number of realizations (400), each constructed from the procedure just outlined. On the computations of pair counts for angular clustering, the random objects are bootstrap-resampled from this catalogue. The details will be discussed in the following sections.

3.1 Definition of photometric recovery

An essential step in photometric recovery of simulated objects (Section 3.2) is to subject them to the same observational/instrumental systematics, to the same detection process as is used in generating the catalogue of observed objects. Such a recovery procedure naturally takes care of common systematics due to survey geometry and varying background noise characteristics over the survey field. When implanted on to an observed image, there may be an additional complication in the recovery process due to overlapping/crowding with existing sources. To better isolate the effect of the background noise, a set of simulated blank sky images were created from observed images. Object mask images generated

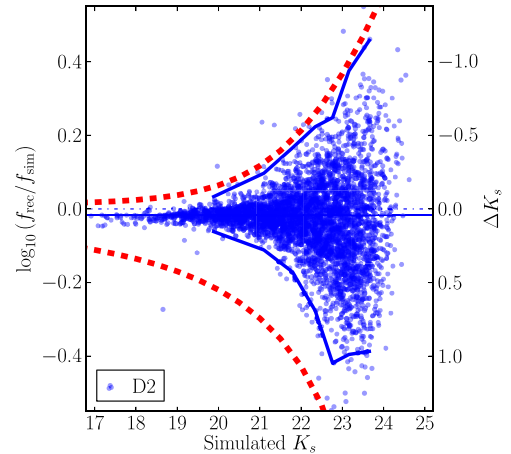


Figure 2. The logarithmic ratios of recovered (f_{rec}) and simulated (f_{sim}) fluxes as a function of simulated K_s magnitudes in D2. The objects are implanted on to a simulated background image. The solid blue curves indicate the bounds used for photometric match (Section 3.1). The solid line just below $\log_{10}(f_{\text{rec}}/f_{\text{sim}}) = 0$ roughly indicates the median of the measurements (for bright sources), implying there is some loss in recovered flux via SExtractor measurements. Repeating this exercise with the observed image as the background will produce a number of recovered fluxes well above the upper bound for photometric matching, indicating overlapping between simulated objects can significantly affect recovery rate. To reduce the file size, the number of objects shown is thinned out to one quarter of the original.

by SExtractor were used to identify source regions, which were then filled with artificial pixel values with variances scaled to local backgrounds. Such images are useful for inspecting recovery rates without the complications caused by overlapping with objects in observed images.

To ensure that recovered objects are indeed the ones implanted, we rely on the discrepancies between the simulated and recovered positions as well as K_s magnitudes. First, matching between simulated and recovered catalogues by positions is done through k -d tree nearest-neighbour lookup with the maximum distance of 10 pixels, large enough for drifting centroids of most objects at the detection threshold. The ID number of the closest object detected is associated with the simulated object. Secondly, the simulated and recovered fluxes are compared, and if they are within the magnitude-dependent bounds (Fig. 2),² the object is tagged as photometrically matched. The photometric cut is fairly conservative and does not aggressively eliminate misidentifications, especially at faint magnitudes. It is done nonetheless to reduce gross misidentifications due to overlapping and errant measurements. Fig. 2 is also useful to assess the number of recovered objects leaking in or out to nearby magnitude bins due to photometric scatter, especially important near the magnitude limit. At and below $K_s \approx 23$, a significant fraction of objects is recovered at more than $\Delta K_s = \pm 0.5$ away from simulated magnitudes, which introduces Eddington bias. The significance of this effect varies depending on binning and

² When presenting figures, we opt to do so with the results in the D2 field. The choice is made to facilitate comparisons with literature where relevant, mainly guided by the fact that the field overlaps with the extensively studied COSMOS survey (e.g. McCracken et al. 2010). Note, however, that the D2 field is the shallowest (in K_s) of all, and the PSFs were matched to the worst seeing in J for colour photometry; consequently, in the other three fields various effects will set in at ~ 1 mag fainter than in D2.

¹ In our LF analysis (Arcila-Osejo & Sawicki 2013) we gauge and correct for incompleteness by using the same simulations described here but selecting simulated objects that match BzK galaxy properties in deep *Hubble Space Telescope* studies.

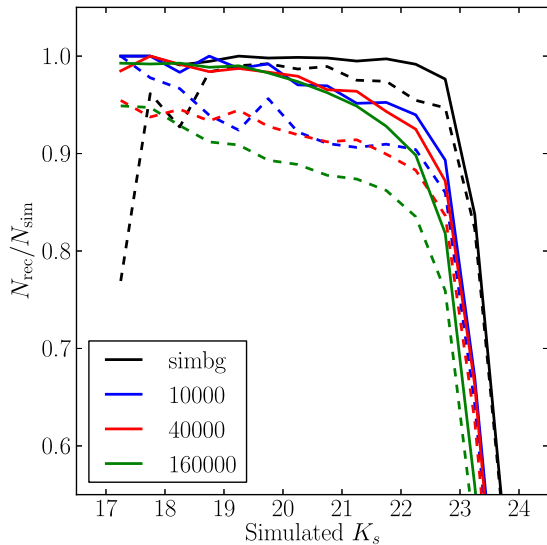


Figure 3. The object recovery fraction as a function of simulated K_s magnitude in D2. The solid lines are completeness functions when object recoveries are based only on their coordinates, whereas dashed lines are for recovery based both on coordinates and photometry; see Section 3.1. In the black lines, we see recovery function when simulated objects are implanted on to a simulated background image. Different numbers of simulated objects are implanted on to the D2 field image: 10 000 (blue), 40 000 (red), and 160 000 (green) to see where overlapping between simulated objects starts affecting object recovery.

photometric errors (Teerikorpi 2004), and should be taken into account in correcting for statistically missing faint populations as in number counting. Some studies of faint galaxy number counts – including ours presented in Arcila-Osejo & Sawicki 2013 – which do not take this effect into account may slightly undercorrect for fainter populations. Also noticeable in the figure is the flux loss of about a few per cent between simulated and recovered objects, which we correct for subsequent simulations.

In real simulation runs, the objects are implanted on to the observed images. This is done to reflect the survey geometries and noise properties, but also to take into account the overlapping of the objects of interest at $z \sim 2$ with foreground/background sources. Crowding the field too much with simulated objects, however, will introduce unwanted systematics due to overlapping among simulated objects themselves.³ The optimal number of simulated objects to be added to an image at once is found by inspecting the drop in the recovery rate as a function of the number of simulated objects (Fig. 3). Compared to the recovery rate for the case when simulated objects are added on to a simulated background image, all recovery rates of the objects on observed images suffer from the loss of objects due to overlapping. In addition, the recovery rate generally drops (~ 10 per cent) when an additional check on photometric match between simulated and recovered magnitude is carried out in addition to position match. Overlapping with existing objects (and other simulated objects) causes both/either the centroids to move and/or the fluxes to be bumped up for the simulated objects upon recovery. Hence the reductions in recovery rates as seen in Fig. 3 are expected. Nonetheless, little change in recovery function is

³ Ideally, a better background image can be constructed by removing only the objects of interest from the field. Implanting simulated objects of interest into such an image would better take into account the systematics caused by object collisions between foreground/background sources.

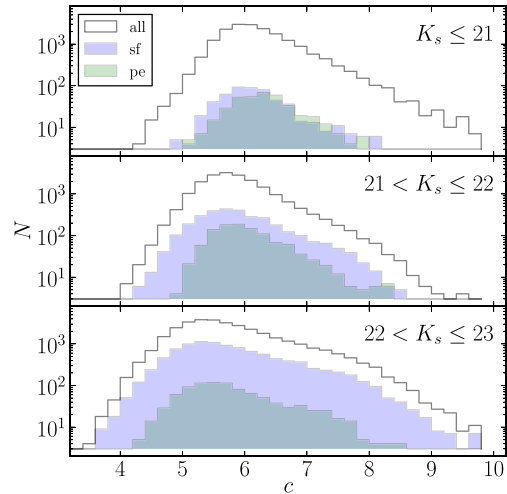


Figure 4. The distributions of K_s -band concentration indices by K_s magnitudes for SF- $gzHK$ (blue histogram), PE- $gzHK$ (green), and all galaxies (white). This figure only shows the observed objects in the D2 field, but similar trends are observed in the other fields.

observed up to 40 000 simulated objects per simulation run per image. To be conservative, no more than 20 000 objects are implanted on to a science image at a time in our simulations.

3.2 Simulated galaxy model

We model a galaxy as a combination of disc and de Vaucouleur (i.e. Sérsic with $n = 4$) spheroid, a parametric model similar to GIM2D (Simard et al. 2002) with the following constraints. Same position angles are assumed for disc and spheroid. The effective radii (i.e. half-light radii) of disc and spheroid are assumed to be similar. Bulge ellipticity and the disc inclination are coupled. In practice, the objects of interest are barely resolved, so these technicalities do not affect our science. While the image quality does not allow us to carry out detailed morphological analysis for $z \sim 2$ objects, we attempt to incorporate some morphological information into our simulation via the concentration index c , as defined by Kent (1985):

$$c = 5 \log (r_{80}/r_{20}),$$

where r_{80} and r_{20} are the radii enclosing 80 and 20 per cent of total flux, respectively. Fig. 4 shows the distributions of concentration indices for galaxies grouped by their K_s magnitudes and types. For K_s -bright objects, PE- $gzHK$ galaxies are slightly more concentrated than SF- $gzHK$ galaxies in general, but the distinction weakens as they become fainter and the morphological information gets lost. As shown in Fig. 5, however, input models do affect the measured concentration indices; bulge-dominated objects tend to get recovered as highly concentrated objects. In our parametrization of simulated object, the bulge-to-total light fraction (B/T) is the primary parameter controlling the light concentration. Using Fig. 5, we empirically map the measured concentration index and half-light radius on to B/T .

While the distributions of concentration indices are not distinguishable for all but the brightest $gzHK$ galaxies, their size distributions are clearly different (Fig. 6). The PE- $gzHK$ galaxies are generally smaller than SF- $gzHK$ galaxies in terms of half-light radius (r_{50}). Since the objects are broadened due to seeing and instrumental systematics, the measured r_{50} are mapped to the model effective radius r_e empirically using Fig. 7, recovering simulated objects with varying model parameters. This is certainly simplistic

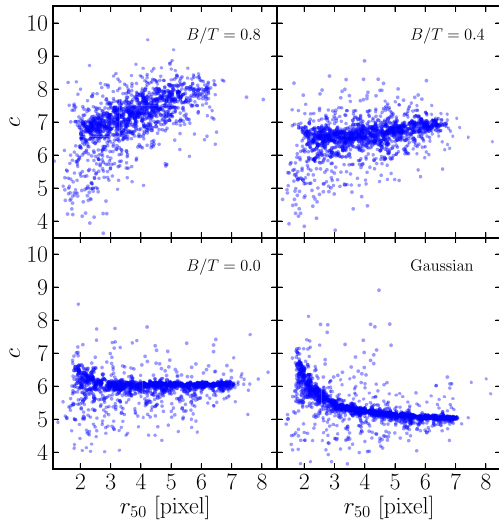


Figure 5. The distributions of measured K_s -band concentration indices c and half-light radii r_{50} for input models with different B/T in D2. The parametric model used for simulated galaxies is described in Section 3.2. The distribution for Gaussian model for point sources (bottom right) is also shown for reference. For clarity, the number of objects shown is thinned out to one tenth of the original.

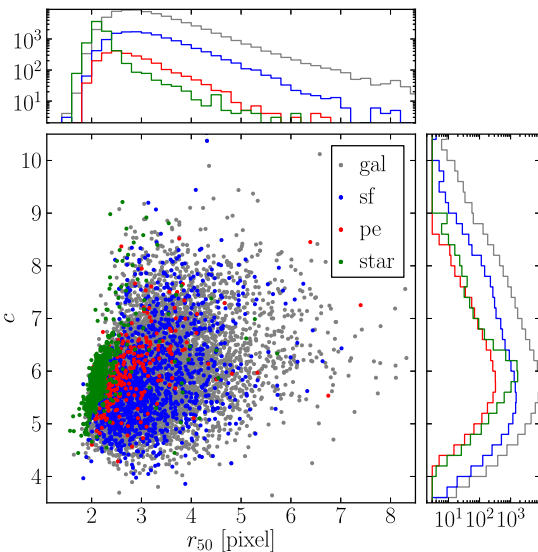


Figure 6. The distributions of K_s -band concentration indices c and half-light radii r_{50} for $K_s \leq 23$ objects in D2. In the scatter plot and histograms, SF- $gzHK$ galaxies are coloured blue, PE- $gzHK$ red, all other galaxies are grey, while stars are green. For clarity, the number of objects is thinned out to one tenth of the original.

and does not take into account such obvious complications as inclination effects, but we find it does not significantly affect completeness (Section 3.3).

3.3 Completeness correction

The simulated objects are randomly drawn directly from the observed object catalogue in an attempt to reproduce the same mix of objects. Corrections still need to be applied to the populations which suffer significant incompleteness; otherwise, the random object catalogue will suffer twice the incompleteness – once when

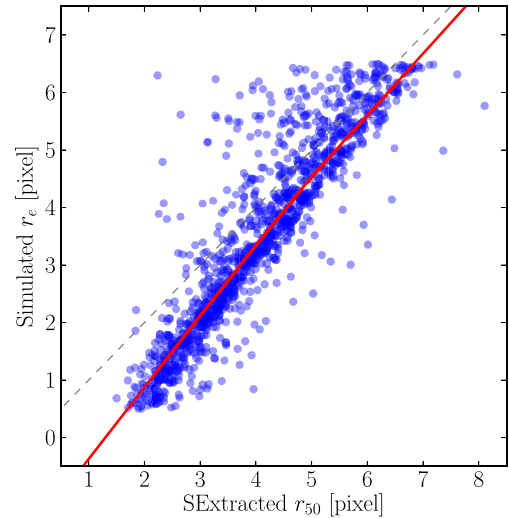


Figure 7. The input model effective radius r_e as a function of the measured half-light radius r_{50} in D2. The dotted line denotes unity where $r_e = r_{50}$. The red solid curve is a fiducial fit to map the measured half-light radius to input model effective radius in our simulations. The curve is chosen mostly to go through the dense parts of distribution, but manual adjustments are made to better match the distribution of r_{50} towards the smaller end in real simulation runs (Fig. 8), i.e. it has not been determined via a rigorous regression analysis. For clarity, the number of objects plotted is thinned out to one tenth of the original.

an observed object catalogue is constructed and yet another when the objects are drawn from the object catalogue, implanted on to a background image, and (un)recovered in simulation.

The primary factors that affect the recovery rate of an object are the brightness and the light profile of that object. The completeness functions were inspected for simulation model parameters that affect these aspects of photometry. The ellipticity e of simulated object does not contribute significantly to incompleteness; the objects with ellipticity $e \gtrsim 0.8$ cannot be recovered by SExtractor, and the completeness at $e \lesssim 0.8$ does not vary (plots not shown). Aside from the total magnitude, it is the effective radius that contributes most significantly to the detectability of objects; see Fig. 8. The same figure also indicates that the model B/T does have a significant effect on the recovery rate towards faint magnitudes. However, morphological information gets lost at faint magnitudes (Fig. 4) and therefore we may not be able to correct for the effect.

Finally, we compare the random object catalogue to the observed catalogue to see how well the former reproduces the latter. As seen in Fig. 9, the observed and completeness-corrected catalogues compare well in terms of K_s magnitude and r_{50} distributions. The figure shows that without corrections we would be missing a large number of objects with small half-light radii r_{50} , the class to which a large number of $gzHK$ galaxies belong (Fig. 6).

4 CLUSTERING PROPERTIES OF $z \sim 2$ GALAXIES

4.1 Two-point angular correlation function

The two-point angular correlation function $\omega(\theta)$ is computed using the Landy and Szalay estimator (Landy & Szalay 1993):

$$\omega_{LS}(\theta) = \frac{r(r-1)DD(\theta)}{n(n-1)RR(\theta)} - \frac{r-1}{n} \frac{DR(\theta)}{RR(\theta)} + 1, \quad (1)$$

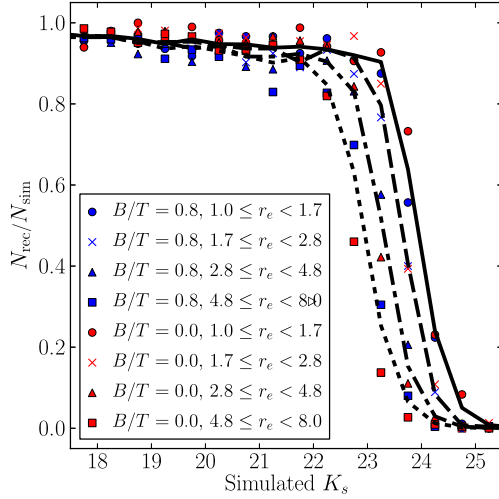


Figure 8. The object recovery fractions as functions of simulated K_s magnitudes, separately computed for different model effective radii in D2. The black lines indicate recovery functions for $1 \leq r_e < 1.7$ (solid), $1.7 \leq r_e < 2.8$ (dashed), $2.8 \leq r_e < 4.8$ (dash-dotted), and $4.8 \leq r_e < 8.0$ (dotted) with r_e in units of pixels. The points indicate recovery fractions computed separately for model bulge fractions $B/T = 0.8$ (blue) and 0 (red), and for $1 \leq r_e < 1.7$ (circle), $1.7 \leq r_e < 2.8$ (cross), $2.8 \leq r_e < 4.8$ (triangle), and $4.8 \leq r_e < 8.0$ (square).

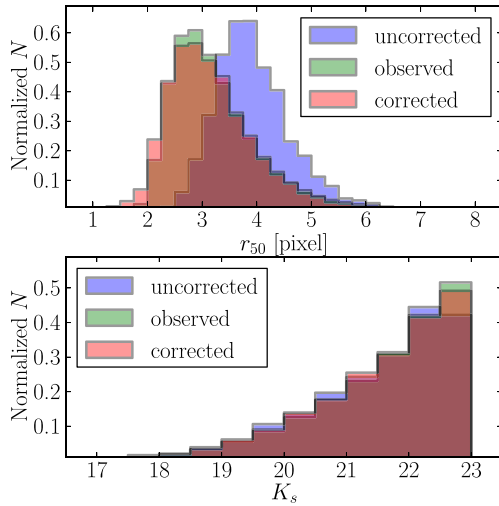


Figure 9. The normalized histograms for measured half-light radii r_{50} (top) and K_s magnitudes for galaxies in D2. The distributions are shown for the observed catalogue (green), the random object catalogue uncorrected for incompleteness (blue), and the random catalogue corrected for incompleteness (red) as described in Section 3.3.

where $DD(\theta)$, $DR(\theta)$, and $RR(\theta)$ are the numbers of data–data, data–random, and random–random pairs with angular separations between $[\theta - \delta\theta/2, \theta + \delta\theta/2]$. The total numbers of objects in the data (n) and random (r) catalogues are used to normalize the pair counts in the above expression. The estimator has become the de facto standard for computing angular correlation functions in galaxy surveys.

For a known angular correlation function $\omega(\theta)$, the number of pairs with separations in $[\theta - \delta\theta/2, \theta + \delta\theta/2]$ is given by

$$n \left(\frac{\delta\Omega_1}{\Omega} \frac{\delta\Omega_2}{\Omega} \right) [1 + \omega(\theta)].$$

Doubly integrating this over the solid angles Ω_1 and Ω_2 for the entire survey area should recover n , as the total number of unique data–data pairs is a fixed quantity; a strong clustering signal at small angular separations must be balanced by a weak clustering signal at large angular separations. However, the normalization of $\omega(\theta)$ depends on the survey geometry as well as how well sampled the clustering signal is. For example, if the survey area is so small that a measurement only captures a clustering signal of a high variance over the limited region, $\omega(\theta)$ measurement tends to underestimate the true angular correlation. Suppose $\omega_{\text{est}}(\theta)$ is computed from some estimator (e.g. ω_{LS}), and $\omega(\theta)$ is the ‘true’ angular correlation function for the sample, such that

$$1 + \omega_{\text{est}}(\theta) = f (1 + \omega(\theta)), \quad (2)$$

where f is a factor that puts $\omega(\theta)$ on a similar scale with $\omega_{\text{est}}(\theta)$. The following constraint

$$n = \iint n \left(\frac{\delta\Omega_1}{\Omega} \frac{\delta\Omega_2}{\Omega} \right) f [1 + \omega(\theta)]$$

leads to

$$f = \left[1 + \iint \omega(\theta) \frac{\delta\Omega_1}{\Omega} \frac{\delta\Omega_2}{\Omega} \right]^{-1} = \frac{1}{1 + \mathcal{I}},$$

where \mathcal{I} is an integral constraint:

$$\mathcal{I} \equiv \frac{1}{\Omega^2} \iint \omega(\theta) d\Omega_1 d\Omega_2. \quad (3)$$

For $\mathcal{I} \ll 1$, equation (2) suggests

$$\omega(\theta) \approx \omega_{\text{est}}(\theta) + \mathcal{I},$$

which leads to the earlier point on how ω_{est} tends to underestimate the true angular correlation function.⁴ Since the true form of $\omega(\theta)$ is not known, a power law of the form

$$\omega(\theta) = A_\omega \theta^{1-\gamma} \quad (4)$$

is often assumed. Using this expression in equation (3) leads to an estimate of integral constraint

$$C \equiv \mathcal{I}/A_\omega = \frac{\sum RR(\theta)\theta^{1-\gamma}}{\sum RR(\theta)} \quad (5)$$

(Roche & Eales 1999) such that

$$\omega(\theta) = A_\omega (\theta^{1-\gamma} - C). \quad (6)$$

The correlation functions computed via equation (1) from the data and random catalogues are often fitted by equation (6). It should be noted that the term integral constraint appears to be used for both \mathcal{I} and C interchangeably in the literature. Henceforth, we assume the form of C when integral constraint is discussed in this paper.

Uncertainties in pair counts over different angular separations are obtained by bootstrap resampling of objects in the data (i.e. observed) and random catalogues (Section 3). At each bootstrap realization, a new data catalogue is generated by randomly resampling (with replacement) the same number of objects from the original data catalogue. The random catalogue is constructed similarly. The angular separation histograms DD , RR , and DR are computed between all unique pairs and are binned up in the range

⁴ This discussion followed the presentation by Wall & Jenkins (2003). For another perspective, see Adelberger et al. (2005), for example.

$-4.5 \leq \log_{10}\theta[^\circ] \leq +1.0$, at a logarithmic interval of $\delta\theta = 0.2$ dex. The correlation function is computed via equation (1). The bootstrap simulation is repeated about 100 times. Since the value of $\omega(\theta)$ within each angular separation bin is found to be normally distributed for each field, the mean and standard deviation are computed in the standard manner to obtain the best estimate and uncertainty for each bin. The average of RR at each angular separation bin is also recorded for the purpose of estimating the integral constraint by means of equation (5). By nature of clustering measurement, $\omega(\theta)$ in different bins are in fact correlated, but we do not take this into account in our uncertainty estimates. For an approach to estimate full covariance, see Wake et al. (2011), for example.

4.2 Fit parameters and their systematic biasing

The parametric function of the form in equation (6) is fitted to the angular correlation functions as outlined in Section 4.1, each bin weighted by its inverse variance. The clustering amplitude A_ω , slope γ , and integral constraint C are estimated via the standard Markov chain Monte Carlo (MCMC) sampling technique as implemented by PyMC (Patil, Huard & Fofonbeck 2010). Due to the form of equation (5) and the fixed set of random–random pairs RR (Section 4.1), there is a one-to-one relation between C and γ when these two parameters are both free.

In using the fitting equation of the form in equation (6), it is often mentioned that the iterative approach required to estimate an integral constraint from equation (5) leads to an unstable solution (e.g. Adelberger et al. 2005; Blanc et al. 2008). Most studies therefore opt to compute the integral constraint independently either by assuming a fiducial slope γ (often $\gamma = 1.8$) or from a theoretical halo model and keeping it fixed for a subsample and given survey field. We argue that the lack of convergence via equations (5) and (6) may at least in part be caused by the particular numerical method that is employed. In Fig. 10, we show the posterior distributions of

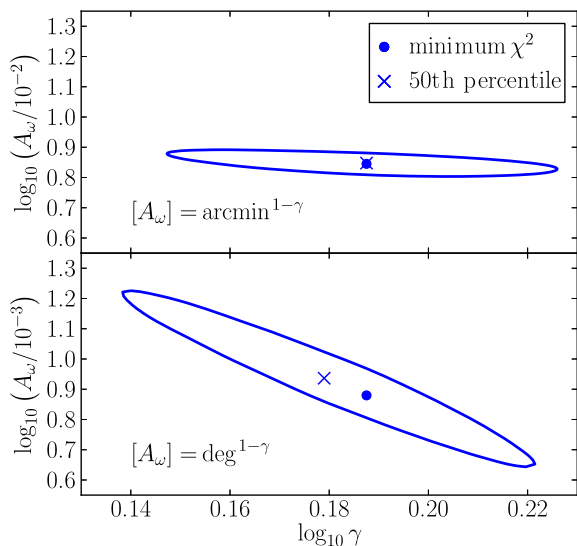


Figure 10. The 68 per cent posterior probability contours of model parameters A_ω and γ of the correlation function for SF-*gzHK* galaxies in D1, as estimated by MCMC sampling technique (Section 4.2). In equation (6), the angular separations are in units of arcminutes (upper) and degrees (lower). Uniform probability density priors are assumed for both parameters. The circular points indicate the parameters at the minimum χ^2 . The crosses indicate the parameters at the 50th percentiles of MCMC posterior distributions.

model parameters A_ω and γ for two similar yet technically different cases of MCMC sampling on the same data. The amplitude A_ω in equation (6) is defined at where the term $(\theta^{1-\gamma} - C)$ is unity. In the case, where θ is in units of arcminutes, only a very weak correlation is observed between A_ω and γ . When θ is in degrees, on the other hand, a strong correlation becomes apparent between the two model parameters; the best estimates for A_ω and γ from the minimum- χ^2 and at the 50th percentiles of posterior distributions also differ quite significantly. The correlation becomes even stronger with θ in radians (not shown). Effectively, we confirm the non-converging tendency and find that estimating A_ω and γ (and C coupled to γ via equation 5) simultaneously as free parameters may lead to systematically lower estimates for γ when the clustering amplitude A_ω is defined in certain angular units. Yet, all these cases supposedly are mathematically similar, so the differences may be attributed to how a particular numerical method spans the parameter space to attain convergence. In other studies, it is often unspecified how A_ω is normalized; however, their clustering data must not be very well sampled over degree scales, so it is possible that they suffer from similar artificial numerical issues when A_ω is defined at those angular scales.

Since our clustering analyses are most relevant on arcminute scales, the angular separations in units of arcminutes are used in equations (5) and (6); hence A_ω technically is in units of $\text{arcmin}^{1-\gamma}$ (so that ω remains dimensionless). The priors for A_ω and γ are uniform probability densities with the range wide enough not to truncate the posterior distributions at extreme values. In all cases, the posterior distributions are observed to be roughly lognormal, and their 50th percentiles mostly match the best estimated parameters at minimum χ^2 .

In this paper, we generally quote the parameters at minimum χ^2 as the best estimates. The 16th and 84th percentile bounds in the marginal distributions of parameters are quoted as the estimates for uncertainties in those parameters. The percentiles are chosen to roughly match commonly cited 1σ uncertainty in literature. We reiterate that the posterior distributions are roughly lognormal.

4.3 Combined angular correlation functions

The distribution of ω_{LS} in each angular separation bin is roughly normal (Section 4.1) and we combine the measurements from our four independent, widely separated fields, to arrive at our best estimates of angular clustering as follows. First, before combining ω from our four fields, we correct them for integral constraints (equation 6). Next, we compute the weighted mean $\bar{\omega}$ of the angular correlation functions, where inverse variance $w = 1/\sigma_\omega^2$ is used as the weight. The variance of the weighted mean is computed from

$$\sigma_{\bar{\omega}}^2 = \frac{1}{\sum_{i=1}^n w_i} \times \frac{1}{n-1} \sum_{i=1}^n w_i (\omega_i - \bar{\omega})^2,$$

where $n = 4$ for the four fields. We then have the combined-field angular correlation function $\bar{\omega}$ and the standard deviation $\sigma_{\bar{\omega}}$ for each angular separation bin. The clustering amplitudes and slopes in Table 1 are estimated from fitting the power law equation (6) (but with $C = 0$) to the combined-field correlation function.

In recent years, it has been realized that angular correlation functions are better modelled by a two-component power law that includes a two-halo term from large-scale clustering of galaxies hosted in separate dark haloes and a one-halo term from objects occupying the same halo. These components are apparent in sufficiently high-quality data (e.g. Zehavi et al. 2004; Wake et al. 2011). We also see

Table 1. The estimated parameters from fitting the angular correlation function ω_{LS} for the combined fields, with varying angular separation ranges. The parameter estimation method and the meanings of the quoted numbers are discussed in Section 4.2. The results vary depending on the fitting domain used; boldfaced values indicate our preferred fitting ranges and fit results.

Sample	$\log_{10}(\theta/1^\circ)$	$A_\omega/10^{-2}$	γ
SF-gzHK	[-2.0, -0.1]	$5.09^{+0.37}_{-0.39}$	$1.73^{+0.07}_{-0.05}$
SF-gzHK	[-2.0, -0.5]	$5.07^{+0.37}_{-0.40}$	$1.70^{+0.08}_{-0.07}$
SF-gzHK	[-2.0, -0.9]	$5.05^{+0.38}_{-0.41}$	$1.61^{+0.09}_{-0.08}$
SF-gzHK	[-3.0, -0.1]	$4.81^{+0.24}_{-0.27}$	$1.76^{+0.03}_{-0.02}$
SF-gzHK	[-3.0, -0.5]	$4.80^{+0.27}_{-0.32}$	$1.76^{+0.03}_{-0.02}$
SF-gzHK	[-3.0, -0.9]	$4.88^{+0.29}_{-0.29}$	$1.76^{+0.03}_{-0.03}$
PE-gzHK	[-2.0, -0.1]	$12.74^{+0.53}_{-0.54}$	$1.90^{+0.06}_{-0.05}$
PE-gzHK	[-2.0, -0.5]	$12.76^{+0.52}_{-0.54}$	$1.92^{+0.06}_{-0.05}$
PE-gzHK	[-2.0, -0.9]	$12.73^{+0.52}_{-0.56}$	$1.87^{+0.07}_{-0.06}$
PE-gzHK	[-3.0, -0.1]	$12.98^{+0.42}_{-0.46}$	$1.95^{+0.03}_{-0.03}$
PE-gzHK	[-3.0, -0.5]	$12.92^{+0.42}_{-0.46}$	$1.96^{+0.03}_{-0.03}$
PE-gzHK	[-3.0, -0.9]	$13.01^{+0.44}_{-0.47}$	$1.95^{+0.03}_{-0.04}$
All galaxies	[-2.0, -0.1]	$2.83^{+0.18}_{-0.20}$	$1.60^{+0.06}_{-0.05}$
All galaxies	[-2.0, -0.5]	$2.83^{+0.17}_{-0.20}$	$1.63^{+0.09}_{-0.07}$
All galaxies	[-2.0, -0.9]	$2.84^{+0.17}_{-0.21}$	$1.60^{+0.11}_{-0.09}$
All galaxies	[-3.0, -0.1]	$2.43^{+0.13}_{-0.14}$	$1.81^{+0.03}_{-0.02}$
All galaxies	[-3.0, -0.5]	$2.33^{+0.13}_{-0.13}$	$1.83^{+0.02}_{-0.02}$
All galaxies	[-3.0, -0.9]	$2.29^{+0.14}_{-0.12}$	$1.84^{+0.02}_{-0.03}$

evidence for the break caused by the two components in our data (Fig. 11). For both SF-gzHK and PE-gzHK the breaks appear around $\log_{10}(\theta/1^\circ) \approx -2.7$, which corresponds to ~ 60 (physical) kpc at $z = 2$ in our adopted cosmology. The projected angular separation at the break is well below the virial radius (r_{200}) of a $\sim 10^{13} M_\odot$ halo, roughly the most massive at $z \sim 2$, so the enhancement of clustering at the small angular scale is expected from one-halo contributions.

In order to measure the large-scale clustering, fitting should be done over angular scales in which the contribution from the one-halo term is not significant, i.e. on scales larger than the break between the two slopes. From practical concerns, studies in the past have exercised varying degrees of care in this regard and consequently we fit over various ranges of angular separations to understand systematics, if any. Table 1 lists the parameters fitted over various angular separation ranges. In all fields, ω_{LS} are sampled well up to $\log_{10}\theta = -0.1$ (θ in degrees). While we have the luxury of having four independent wide fields, the survey geometries vary significantly (Fig. 1), and the angular correlation functions appear to reflect them at larger angular separations (Section 4.4).

In Fig. 12, we see that the estimated fitting parameters are indeed sensitive to the choice of fitting domain. This is particularly notable when the lower limit gets extended to smaller angular separation. However, as long as the lower limit remains above the one-halo/two-halo slope transition, the results remain consistent at $\sim 1\sigma$ level, except for the narrowest interval of $-2.0 \leq \log_{10}\theta < -0.9$, for which the slopes γ appear systematically underestimated. As the fitting domain gets extended to smaller angular separations, the power-law slope γ becomes steeper for both PE-gzHK and SF-gzHK galaxies. This is likely a consequence of greater contribution to the correlation function from the one-halo term (Fig. 11).

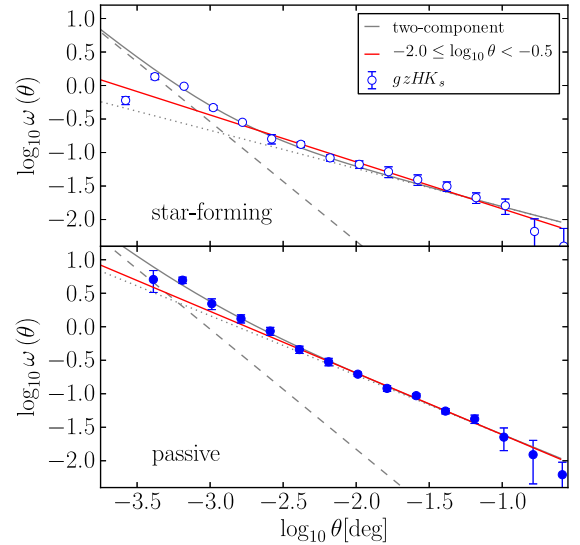


Figure 11. The angular correlation functions as a function of angular separation for SF-gzHK (upper panel) and PE-gzHK (lower panel) galaxies. Each correlation function is fitted via χ^2 minimization with a two-component power law of the form equation (6), ensuring one of the slopes γ is always larger than the other. The components with a flatter slope (dotted curve) and with a steeper slope (dashed curve) are linearly added, which gives an overall better fit (solid curve) to the observed angular correlation function. The fitting is done over the interval of $-3.5 < \log_{10}\theta < -0.5$. The best-fitting curve from a single-component fitting over $-2.0 \leq \log_{10}\theta < -0.5$ is also shown as reference (red curve).

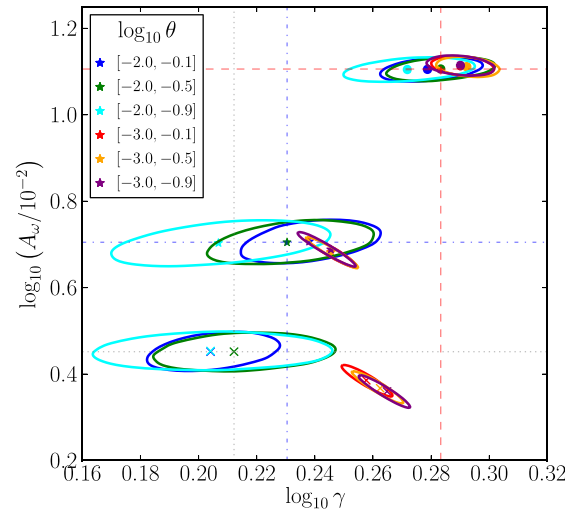


Figure 12. The variation of the 68 per cent confidence intervals of the clustering amplitude A_ω and the power-law slope γ for PE-gzHK (upper contours), SF-gzHK (middle contours), and all galaxies (lower contours) when the fitting angular separation intervals are varied; see Table 1. The colours of contours indicate the fitting interval as indicated in the legend, except the grey contours which are from fitting without integral constraints. The best parameter estimates from the minimum- χ^2 are indicated for PE-gzHK (points), SF-gzHK (stars), and all galaxies (crosses). The red dashed lines indicate the best estimates of A_ω and γ from fitting over $-2.0 \leq \log_{10}\theta < -0.5$ using the combined field data for PE-gzHK (red dashed), SF-gzHK (blue dash-dotted), and all galaxies (grey dotted).

Passive galaxies are known to preferentially reside in dense environments in the local Universe (e.g. Dressler 1980; Balogh et al. 2004), and it is likely that such an environmental relation was already present at $z \sim 1.5$ (e.g. Chuter et al. 2011; Quadri et al. 2012). A stronger one-halo term contribution, in addition to the generally stronger clustering of PE-*gzHK* galaxies over the angular scales probed, could mean that a star formation rate (SFR)–density relation similar to the one observed in the local Universe already existed at $z \sim 2$. While it is difficult to assess exactly how the clustering results relate to nearest-neighbour density estimates for local environment, the observation here suggests that the virialization of galaxies within massive haloes has already progressed significantly by $z \sim 2$ for both SF and passive galaxy populations, leading to situations in which environmental effects can be triggered to produce passive galaxies in such environments. In other words, the presence of the one-halo term for PE galaxies (in addition to that for SF ones) can be taken as evidence for environmental quenching at $z \sim 2$, though it could also indicate other processes at play, such as mass quenching combined with a mass–density relation.

For the purpose of quantifying large-scale clustering, we wish to focus on scales where galaxy bias should be small (i.e. to isolate the two-halo term). In practice, the power-law fitting needs to be limited to above the break in the slopes of the angular correlation function and so we use the range of $-2.0 \leq \log_{10}\theta < -0.5$ for our clustering measurements unless otherwise noted. Strictly speaking, fitting above a certain angular range is not sufficient to isolate the clustering on the larger scales, as clearly seen in the difference between the single power-law fitting over $-2.0 \leq \log_{10}\theta < -0.5$ and the two-component power-law fitting over a larger angular scale in Fig. 11. In order to fit clustering observations better, it is best to cast the measurement in view of theoretical halo models, which we defer to a future opportunity. Nevertheless, the one-component analysis over a limited angular scale in this paper would still be useful to make comparisons to existing studies.

In Fig. 13, the combined angular correlation functions for all fields are compared to those found in other studies (listed in

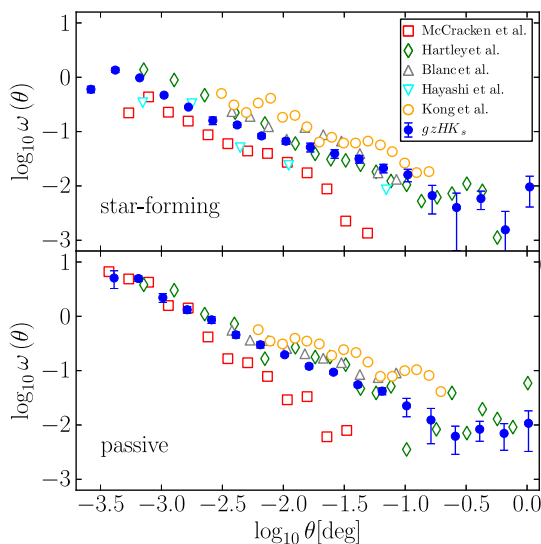


Figure 13. The angular correlation functions as a function of angular separation for SF-*gzHK* (upper panel) and PE-*gzHK* (lower panel) galaxies. The measurements are compared to other surveys employing similar colour selection techniques (*BzK*) that target $z \sim 2$ galaxies. For clarity the uncertainties for these other surveys’ results are not plotted, but they are generally larger than those in our data.

Table 2). All these measurements use adaptations of the classic *BzK* technique to select their samples, and therefore are directly comparable. For both SF-*gzHK* and PE-*gzHK* galaxies, the other correlation functions are roughly consistent, with ours being in the middle ground, although systematic differences obviously exist between different surveys. The correlation functions of Kong et al. (2006) are based on a shallower catalogue ($K_{\text{lim}} \sim 21.8$) and sample a systematically more strongly clustered population than we do (Section 4.5). It is therefore not surprising that their values are systematically higher than ours. In contrast, while deep, the correlation functions of Hartley et al. (2008) are based on samples that do not accurately reproduce the classic *BzK* selection (see Blanc et al. 2008; McCracken et al. 2010). Specifically, some of the galaxies that would be classified as PE by classic *BzK* selection (and in our PE-*gzHK* sample) are absent from the *BzK* sample of Hartley et al. as they are lost predominantly out of the high- z population into the low- z galaxy selection window. While such loss affects the number counts and luminosity functions (see Arcila-Osejo & Sawicki 2013) it should not strongly affect clustering, unless very strong gradients are present in the clustering signal as a function of galaxy colour. The fact that our clustering measurements agree with those of Hartley et al. (2008) confirms this view. Finally, especially curious are the correlation functions computed by McCracken et al. (2010) in the COSMOS field, which exhibit much steeper power-law behaviours than the rest of surveys shown in Fig. 13. Our D2 field is a subset of the COSMOS field, which is larger (2 deg² for COSMOS versus 1 deg² for D2) so their pair counts must be very well sampled over the common angular separation intervals. The correlation functions of *gzHKs* galaxies in the field computed in our study are not consistent with their steep power-law functions (Fig. 14). The origin of this large discrepancy is unclear: although the COSMOS field is the most discrepant from the mean in terms of $z \sim 2$ galaxy number counts and luminosity functions (Arcila-Osejo & Sawicki 2013), field-to-field cosmic variance (Section 4.4) is not sufficient to account for the differences in clustering between the McCracken et al. (2010) and our (and other) studies.

4.4 Field-to-field variations

In Fig. 14, the angular correlation functions for SF-*gzHK* and PE-*gzHK* galaxies from all four CFHTLS deep fields are compared. The four fields are widely separated on the sky, so the sample variance (i.e. cosmic variance) is a significant contributor to the observed differences. The peculiar shapes of the survey geometry (Fig. 1) also lead to varying integral constraints. Such a sample variance is often conveniently cited for explaining apparent disparities among the results from various studies, yet even two closely related fields can yield quite different results based on separate analyses, as in the case for comparing ω_{LS} measurements for our D2 field to that of COSMOS (Section 4.3). Since each field in our study is as large as or larger than the typical survey area used in most of the past studies (Table 2), it would be beneficial to examine the effects of field-to-field variance when the same analysis is performed consistently on data from independent fields.

Table 3 shows the estimated parameters from fitting to ω_{LS} over $-2.0 \leq \log_{10}\theta < -0.5$ in the manner described in Section 4.2, allowing both the clustering amplitude A_ω and slope γ free to vary (the integral constraint C explicitly depends on γ). Compared to the generic values from the combined fields (Table 1), there exist significant field-to-field differences in the slope γ of $\delta\gamma = 0.17, 0.12,$ and 0.21 for SF-*gzHK*, PE-*gzHK*, and all galaxies, respectively. The

Table 2. Summary of selected recent BzK -like clustering study surveys. The K_{lim} are the magnitude cuts used for analysis. Note that the McCracken et al. (2010) field overlaps our D2 field.

Study	Area (deg ²)	Subfields	K_{lim} [AB]
Kong et al. (2006)	0.09 (Deep 3a-F)	2 (one of them is Deep 3a-F)	21.8
Hayashi et al. (2007)	0.05	1	23.2
Blanc et al. (2008)	0.71	2	21.8
Hartley et al. (2008)	0.63	1	23
McCracken et al. (2010)	1.9	1	23
This work	2.59	4	23

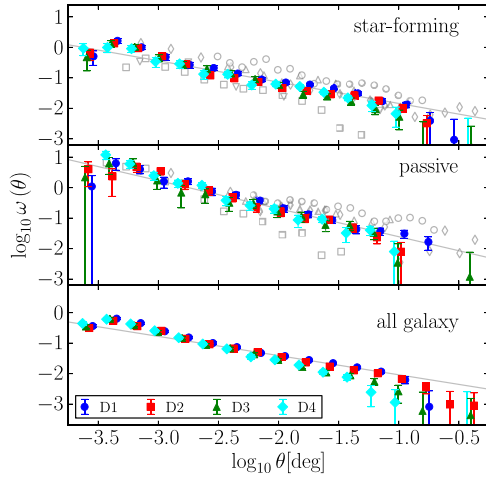


Figure 14. The angular correlation function ω_{LS} as a function of angular separation for SF- $gzHK$ (upper panel) and PE- $gzHK$ galaxies (middle panel), as well as all galaxies (lower panel) in each field. The error bars indicate standard deviations computed in the standard manner within each angular separation bin; see Section 4.2 for details. The solid grey lines are the best-fitting power-law functions for the combined fields over the angular separation range of $-2.0 \leq \log_{10}\theta < -0.5$ (Section 4.3). The various grey points in the upper and middle panels are the same results from the other surveys (Fig. 13) reproduced here for comparison.

Table 3. The angular correlation functions ω_{LS} fitted separately for individual fields. The estimated parameters are over the angular separation range between $-2.0 \leq \log_{10}\theta < -0.5$, with all parameters allowed to vary. The parameter estimation method and the meanings of the quoted numbers are discussed in Section 4.2.

Sample	Field	$A_{\omega}/10^{-2}$	γ	$C/10^{-1}$
SF- $gzHK$	D1	$6.99^{+0.53}_{-0.38}$	$1.54^{+0.08}_{-0.10}$	$1.94^{+0.62}_{-0.41}$
SF- $gzHK$	D2	$5.25^{+1.09}_{-0.41}$	$1.41^{+0.10}_{-0.15}$	$0.27^{+1.56}_{-0.71}$
SF- $gzHK$	D3	$4.00^{+0.52}_{-0.42}$	$1.78^{+0.20}_{-0.22}$	$1.17^{+0.87}_{-0.44}$
SF- $gzHK$	D4	$5.23^{+0.46}_{-0.45}$	$1.81^{+0.14}_{-0.14}$	$1.00^{+0.44}_{-0.29}$
PE- $gzHK$	D1	$14.20^{+1.73}_{-1.36}$	$1.73^{+0.18}_{-0.20}$	$1.16^{+0.88}_{-0.44}$
PE- $gzHK$	D2	$11.66^{+1.46}_{-1.73}$	$2.04^{+0.21}_{-0.18}$	$0.45^{+0.26}_{-0.17}$
PE- $gzHK$	D3	$11.92^{+2.17}_{-2.17}$	$1.93^{+0.30}_{-0.30}$	$0.84^{+0.90}_{-0.39}$
PE- $gzHK$	D4	$12.59^{+1.65}_{-2.03}$	$2.01^{+0.26}_{-0.22}$	$0.63^{+0.42}_{-0.25}$
All galaxies	D1	$3.77^{+0.30}_{-0.19}$	$1.46^{+0.06}_{-0.08}$	$2.46^{+0.61}_{-0.40}$
All galaxies	D2	$3.00^{+0.18}_{-0.13}$	$1.52^{+0.07}_{-0.07}$	$1.91^{+0.47}_{-0.34}$
All galaxies	D3	$2.03^{+0.17}_{-0.13}$	$1.73^{+0.13}_{-0.14}$	$1.35^{+0.57}_{-0.36}$
All galaxies	D4	$1.90^{+0.13}_{-0.14}$	$1.99^{+0.14}_{-0.14}$	$0.65^{+0.24}_{-0.17}$

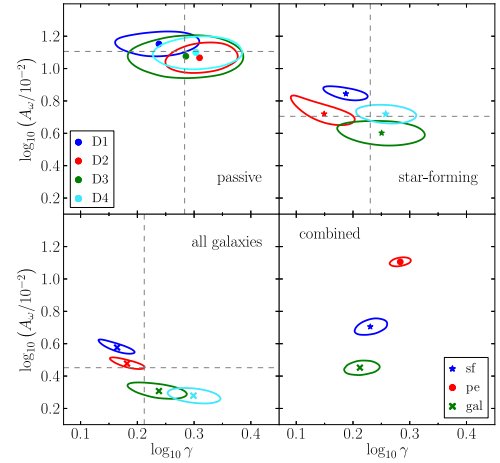


Figure 15. The field-to-field variations of the 68 per cent confidence intervals of the clustering amplitude A_{ω} and the power-law slope γ for PE- $gzHK$ (upper left), SF- $gzHK$ (upper right), all galaxies (lower left). The colours of contours indicate the fields D1 (blue), D2 (red), D3 (green), and D4 (orange). The results are also tabulated in Table 3. The black dashed lines indicate the minimum- χ^2 parameters for each type of galaxies from the combined fields (Table 1), which are also shown in the lower-right panel. The markers inside the contours indicate the parameter set for the minimum- χ^2 fit. All results are from fitting over $-2.0 \leq \log_{10}\theta < -0.5$.

same information is graphically presented in Fig. 15, showing the field-to-field variation is at the 1σ - 2σ level. Since the clustering amplitude A_{ω} and slope γ can be correlated (Section 4.2), they contain partially degenerate information about clustering strength. This makes it somewhat less intuitive when comparing measurements, especially given the amplitude A_{ω} is normalized in units of $[\theta]^1 - \gamma$ and the authors vary on their choice of θ at which A_{ω} is defined as well as the value of γ when fixed. It is best that equation (6) be explicitly evaluated at a specific angular separation for comparison. Our clustering amplitudes A_{ω} are in units of arcminute^{1- γ} with the standard deviations of $\sigma_{A_{\omega}}/10^{-2} = 1.06, 0.99, \text{ and } 0.76$ for SF- $gzHK$, PE- $gzHK$ and all galaxies, respectively.

When observed angular correlation functions are noisy and of lower quality, a common practice is to fix the slope γ at a canonical value such as $\gamma = 1.8$ which is based on low- z measurements of galaxy clustering (e.g. Norberg et al. 2001; Zehavi et al. 2002), and simply let the clustering amplitude A_{ω} carry the clustering information. To facilitate comparisons, we present measurements with $\gamma = 1.8$ in Table 4. However, since the clustering amplitude A_{ω} can be sensitive to the slope γ (Fig. 12) caution must be exercised when comparing clustering amplitudes measured with different values of γ . There is evidence that the power-law slope varies for different galaxy populations. For example, Adelberger et al. (2005) report a

Table 4. The estimated parameters over the angular separation range $-2.0 \leq \log_{10}\theta < -0.5$, with the power-law slope fixed at $\gamma = 1.8$. The amplitude A_ω is allowed to vary, and the integral constraint C is computed via equation (5), i.e. effectively fixed for each field.

Sample	Field	$A_\omega/10^{-2}$	$C/10^{-1}$	r_0 (Mpc)
SF-gzHK	D1	$6.53^{+0.33}_{-0.33}$	0.96	7.87
SF-gzHK	D2	$4.62^{+0.32}_{-0.32}$	0.86	6.49
SF-gzHK	D3	$3.99^{+0.43}_{-0.42}$	1.12	5.98
SF-gzHK	D4	$5.24^{+0.44}_{-0.44}$	1.03	6.96
PE-gzHK	D1	$14.31^{+1.47}_{-1.46}$	0.96	8.63
PE-gzHK	D2	$10.90^{+1.47}_{-1.46}$	0.86	7.42
PE-gzHK	D3	$11.81^{+2.06}_{-2.06}$	1.13	7.76
PE-gzHK	D4	$12.18^{+1.79}_{-1.79}$	1.03	7.90
All galaxies	D1	$3.23^{+0.10}_{-0.10}$	0.96	
All galaxies	D2	$2.68^{+0.08}_{-0.09}$	0.86	
All galaxies	D3	$1.99^{+0.12}_{-0.12}$	1.12	
All galaxies	D4	$1.97^{+0.13}_{-0.13}$	1.03	

Table 5. The estimated parameters over the angular separation range $-2.0 \leq \log_{10}\theta < -0.5$, with γ fixed at the values obtained from the combined fields.

Sample	Field	$A_\omega/10^{-2}$	γ	$C/10^{-1}$	r_0 (Mpc)
SF-gzHK	D1	$6.67^{+0.33}_{-0.33}$	1.70	1.25	8.84
SF-gzHK	D2	$4.71^{+0.32}_{-0.33}$	1.70	1.14	7.20
SF-gzHK	D3	$4.03^{+0.43}_{-0.43}$	1.70	1.44	6.57
SF-gzHK	D4	$5.31^{+0.45}_{-0.45}$	1.70	1.33	7.73
PE-gzHK	D1	$14.49^{+1.49}_{-1.51}$	1.92	0.70	7.67
PE-gzHK	D2	$11.32^{+1.51}_{-1.52}$	1.92	0.62	6.74
PE-gzHK	D3	$11.92^{+2.05}_{-2.04}$	1.92	0.85	6.93
PE-gzHK	D4	$12.44^{+1.80}_{-1.82}$	1.92	0.77	7.08
All galaxies	D1	$3.43^{+0.11}_{-0.11}$	1.63	1.52	
All galaxies	D2	$2.84^{+0.09}_{-0.09}$	1.63	1.40	
All galaxies	D3	$2.10^{+0.13}_{-0.13}$	1.63	1.72	
All galaxies	D4	$2.04^{+0.14}_{-0.14}$	1.63	1.59	

flatter slope (≈ 1.6) for Lyman-break galaxies (i.e. high- z SF galaxies), while McCracken et al. (2010) report a much steeper slope (≈ 2.5) for passive BzK galaxies. Our measurements do suggest a systematically steeper slope for PE-gzHK but the difference seems less pronounced than what these studies present (the McCracken et al. slopes are steeper than other studies for both passive and SF BzK s). Few high- z studies of this nature so far have allowed their correlation function slopes to vary, primarily due to data quality. Our work presents a significant improvement in the accuracy of the measurements over the previous studies.

In Table 5, we present the clustering amplitudes A_ω when the slopes are fixed at the values obtained from the combined fields: $\gamma = 1.70$ for SF-gzHK, 1.92 for PE-gzHK, and 1.63 for all galaxies. The same information is graphically presented in Fig. 16, in which we see that fixing the correlation function slope γ does not lead to a confidence interval similar to the one obtained from marginal posterior distribution and therefore tends to underestimate the parameter uncertainty compared to the case when γ is allowed to vary. On the

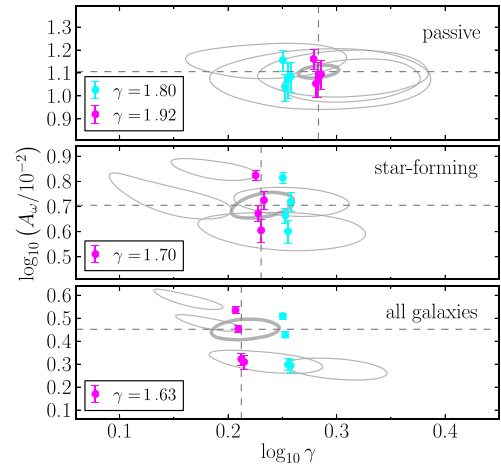


Figure 16. The 68 per cent confidence contours of the clustering amplitude A_ω and the power-law slope γ for PE-gzHK (upper), SF-gzHK (middle), and all galaxies (lower). The thin grey contours are field-to-field results as in Fig. 15, shown for reference. The thicker grey contours are from the combined fields (Fig. 12). The cyan markers indicate the results when the power-law slope is fixed at $\gamma = 1.8$; the magenta markers are also for fixed slopes, $\gamma = 1.92$ for PE-gzHK, 1.70 for SF-gzHK, and 1.63 for all galaxies. The plotted points are slightly offset horizontally for clarity. The error bars for the fixed fitting do not exactly coincide with the 68 per cent confidence contours, since they are derived from independent MCMC sampling runs with fixed γ . All results are from fitting over $-2.0 \leq \log_{10}\theta < -0.5$.

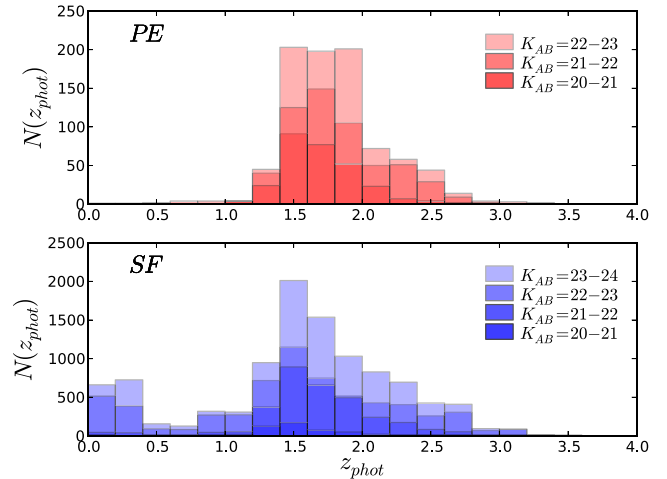


Figure 17. Photometric redshift distributions of our PE and SF galaxies in the Muzzin et al. (2013) catalogue.

other hand, so long as the amplitude A_ω and slope γ are not strongly correlated (which appears to be the case for us), the estimated values for A_ω suffer little systematic effect by fixing γ . However, that is only true under a very fortuitous circumstance in which these parameters are not strongly correlated. We discussed in Section 4.2 that such a correlation apparently can become an issue depending on particular ways in which parameter estimation is carried out. To assess the potential for systematics introduced by fixed γ , it is suggested the χ^2 space be inspected for correlation between A_ω and γ , even if it is not explicitly used to estimate parameter uncertainty.

It should be kept in mind that, for comparison purposes, the clustering amplitude A_ω is not a very meaningful indicator of clustering strength away from where $(\theta^{1-\gamma} - C)$ is unity and when the power-law slope γ is allowed to vary. These are often degenerate, i.e. a

large γ tends to be compensated by a small A_ω and vice versa. The same can be said for spatial correlation length r_0 , which is a scale parameter in a power law of the form

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

and is estimated in a way that depends both on A_ω and γ via the Limber transformation (e.g. Limber 1953; Brainerd, Smail & Mould 1995). In Tables 4 and 5, we list the values of r_0 computed via the same transformation, assuming Gaussian redshift distributions for SF and passive BzK galaxies presented in Blanc et al. (2008).⁵

The correlation length r_0 is a measure of clustering scale, i.e. how galaxies' spatial distributions are biased, and compared among different surveys. In general, larger correlation lengths for the passive BzK galaxies, compared to those for SF BzK s, have been reported (Blanc et al. 2008; McCracken et al. 2010; Lin et al. 2012), which imply stronger clustering of passive galaxies. We confirm this tendency of PE- $gzHK$ galaxies having generally larger correlation lengths than SF- $gzHK$ galaxies, when the same slope γ is used for the two populations (Table 4). When the correlation function slopes γ are not fixed at the same value, however, such a trend is actually not observed (Table 5) and in fact the correlation lengths for SF- $gzHK$ are larger. Table 6 of McCracken et al (2010) shows a similar effect, although these authors do not explicitly comment on this issue. This exercise suggests that direct comparisons of correlation lengths may be relevant only under certain circumstances, such as when the slope γ is fixed for all populations being compared. As noted earlier, it has become apparent that the correlation function slope γ does vary between different galaxy populations at $z \sim 2$. Clearly, the full picture requires reliable measurements of both A_ω and γ , and caution should be exercised when interpreting correlation lengths computed with artificially fixed γ .

4.5 Dependence on galaxy brightness

The K band, which at $z \sim 2$ samples the red part of the rest-frame optical/near-infrared SED, provides an often-used proxy for stellar mass. While a better mass estimate of a high-redshift galaxy can be obtained from full SED analysis (e.g. Sawicki & Yee 1998), single-band estimates remain useful due to their simplicity. Similarly, the rest-frame UV, at $z \sim 2$ sampled in the observed optical wavelengths, is a useful proxy for star formation activity in relatively unobscured galaxies (e.g. Kennicutt 1998; Sawicki 2012b). Here we investigate the clustering of our $gzHK$ galaxies as a function of both observed K - and r -band magnitude.

We note that artificial differences in clustering as a function of galaxy brightness can be introduced if, for example, there are differences between redshift selection windows for galaxies of different magnitude. With that in mind, we checked for such $N(z)$ differences within our samples by cross-matching the $gzHK$ galaxies in the D2 (COSMOS) field with objects in the Muzzin et al. (2013) photo-

⁵ The values are from table 6 of Blanc et al. (2008): $(\bar{z}, \sigma_z) = (1.58, 0.16)$ and $(1.78, 0.31)$ for passive and SF BzK galaxies, respectively, where the σ_z values reflect the intrinsic widths of the distributions after correction for photometric redshift scatter. We verified that the Blanc et al. redshift distributions are consistent with redshift distributions we computed for our PE and SF samples using the photometric redshift catalogues of Muzzin et al. (2013) in the D2/COSMOS field. We also find no evidence for strong differences in $N(z_{\text{phot}})$ as a function of magnitude within our samples (see Fig. 17).

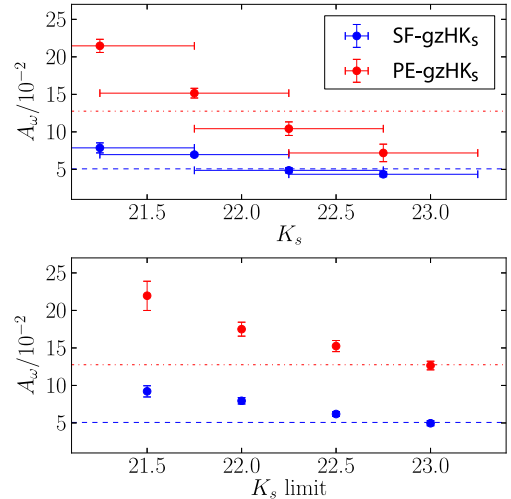


Figure 18. The clustering amplitude for SF- $gzHK$ (blue) and PE- $gzHK$ (red) galaxies as a function of K_s magnitude. In the upper panel, the horizontal bars indicate the K_s magnitude intervals defining subsamples. In the lower panel, the points are plotted at the upper limits for K_s magnitude. All measurements are over $-2.0 < \log_{10}\theta < -0.5$, with γ fixed at the values obtained for the combined field. The red dash-dotted and blue dashed lines indicate the values for the full samples.

metric redshift catalogue and we find that the $N(z)$ distributions do not appear to vary strongly with object apparent magnitude.

4.5.1 Rest-frame optical

The relationship between K -band brightness and clustering strength is of interest because it relates to how galaxies of a given stellar mass cluster as a function of their DM halo mass.

At $z = 0$, brighter galaxies exhibit stronger clustering (e.g. Zehavi et al. 2011). A qualitatively similar trend is observed for K_s -selected SF BzK galaxies at $z \sim 2$ (Kong et al. 2006; Hayashi et al. 2007; McCracken et al. 2010; Lin et al. 2012), which we confirm in Fig. 18. Additionally, we have a large enough sample to study the clustering of PE- $gzHK$ galaxies as well, and their clustering trend is qualitatively similar to that of SF- $gzHK$ s, though with a much stronger dependence on magnitude. This effect is not unexpected if PE- $gzHK$ galaxies are indeed not forming stars: if this is the case then they are dust-free and their observed K -band magnitudes correlate relatively straightforwardly with stellar mass. The newly observed trend then indicates that more massive passive galaxies reside in more massive DM haloes.

4.5.2 Rest-frame UV

In SF galaxies (but not necessarily in passive ones), rest-frame UV luminosity traces the light from young stellar populations acting as a rough proxy for SF activity. In UV-selected samples at $z \sim 2$ and above, stronger clustering is observed for UV-bright galaxies, suggesting that the more vigorously SF galaxies are more clustered at those redshifts (e.g. Ouchi et al. 2004; Adelberger et al. 2005; Lee et al. 2006). At low redshift, $z \leq 1$, this trend with luminosity disappears or perhaps even reverses (Heinis et al. 2007). At intermediate redshifts Savoy et al. (2011) report that the clustering as a function of UV luminosity (their observed R band) turns over at $z \sim 2$ and reverses below that redshift; that is, at $z \gtrsim 2$, UV-bright galaxies cluster more strongly compared to UV-faint galaxies, whereas

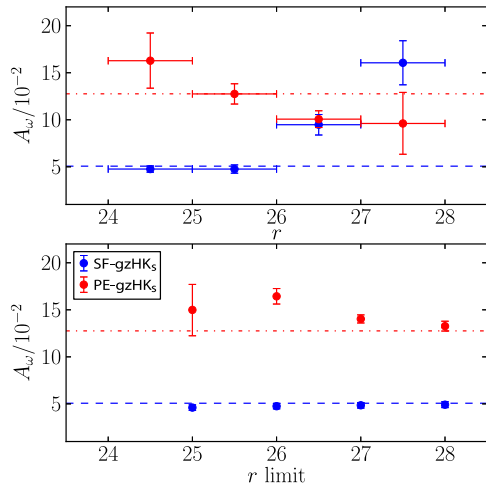


Figure 19. The clustering amplitude for SF-gzHK (blue) and PE-gzHK (red) galaxies as a function of r magnitude. In the upper panel, the horizontal bars indicate the r magnitude intervals defining subsamples, not uncertainties. In the lower panel, the points are plotted at the upper limits for r magnitude. All measurements are over $-2.0 < \log_{10}\theta < -0.5$, with γ fixed at the values obtained for the combined field.

at $z \lesssim 2$ UV-faint galaxies cluster more strongly than UV-bright galaxies. If real, such a transition is very interesting, since the progression of the clustering peak presumably tracks the decrease of star formation in the most massive galaxies across the epochs at which cosmic SF activity peaks and starts to decline.

Our SF-gzHK galaxies are at $z \sim 2$ or below (e.g. Blanc et al. 2008), so comparable to the BM ($z \sim 1.7$) galaxies in Savoy et al. (2011). In Fig. 19, we show the clustering amplitude as a function of r magnitude for both SF-gzHK and PE-gzHK galaxies. Our photometry is not as deep as that of Savoy et al. (2011), so it is difficult to assess how the results compare at the faintest magnitude bin, but even at brighter magnitudes interesting trends emerge.

A remarkable difference is observed in how SF-gzHK and PE-gzHK galaxies cluster in Fig. 19 (top panel), where the clustering properties of gzHK galaxies within different r magnitude slices are plotted. For SF-gzHK, r -faint galaxies cluster more strongly – in agreement with the BM (i.e. $z \sim 1.7$ star formers) of Savoy et al. (2011) – whereas the trend is opposite for PE-gzHK. Given that r -band samples the rest-frame UV, the behaviour of SF-gzHK galaxies may indicate that those SF-gzHKs with less SF activity are found more clustered at that epoch already, consistent with lower- z trends. A possible explanation of this trend is the vigorous activity seen in the most massive haloes at earlier epochs is starting to shut down at $z \sim 2$ (Savoy et al. 2011). The case of PE-gzHK is more straightforward. They are selected for the lack of active star formation and their rest-frame UV light is produced by long-lived low-mass stars rather than recently formed massive ones. Consequently, their luminosity is correlated with their stellar masses, and thus the clustering trend observed for PE-gzHK galaxies likely simply reflects the fact that the more massive (hence rest-UV brighter) PE galaxies reside in the more massive DM haloes, as we found in our analysis using the K band (Section 4.5.1).

4.6 Dependence on colour and sSFR

There has been a recent suggestion by Lin et al. (2012) that the correlation length r_0 for SF BzK galaxies exhibits a turnover as a

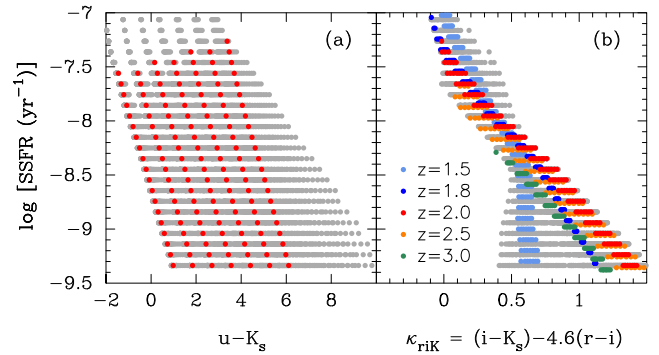


Figure 20. The sSFR as a function of colour combinations for SED models of SF-gzHK galaxies. The grey points show all the models considered, with a wide range of ages, redshift spanning $z = 1.5 - 3.0$, and $E(B - V) = 0.0 - 0.6$. Coloured points highlight selected redshifts. The left-hand panel shows $z = 2.0$ models only but with different amounts of extinction in order to highlight the fact that the use of a single colour results in strong degeneracies in sSFR estimation. The right-hand panel shows a colour combination that better correlates with sSFR. With the exception of objects at the very low end of the plausible redshift range, the colour combination shown in the right-hand panel identifies sSFR almost unambiguously, irrespective of reddening and across the likely redshift range spanned by SF-gzHK galaxies.

function of specific star formation rate (sSFR). That is, SF BzK galaxies with low sSFRs are as strongly clustered as passive BzK galaxies, but SF BzK galaxies with high sSFRs are also more clustered than galaxies with more typical sSFRs; the minimum in the clustering strength appears to be at $sSFR \sim 2 \times 10^{-9} \text{ yr}^{-1}$ (see fig. 5 in their paper).

While we have not measured sSFR directly for our galaxies, a simple proxy measurement of sSFR can be obtained from carefully chosen combinations of broad-band colours. To do this, we determine the suitable colours using Bruzual & Charlot (2003) spectral synthesis models adjusted for Calzetti et al. (2000) dust, cosmological effects, and integrated through filter transmission curves using SEDfit (Sawicki 2012a). Naively, a reasonable choice for sSFR is $u - K_s$ colour, since u band traces the UV light from star formation while K_s band is sensitive to the light from old stellar populations which constitutes the bulk of stellar mass in galaxies. However, dust reddening, which is also most severe in the rest-UV, dilutes the correlation between sSFR and $u - K_s$ colour (Fig. 20a). Such a degeneracy can be alleviated to some extent by tracing $z \sim 2$ galaxy model tracks in a colour–colour plane in which dust degeneracy is projected out, effectively integrating over reddening variations. Such a transformation can be defined in the riK_s colour–colour space: in Fig. 20(b), we see that the correlation between sSFRs and the lines of constant colour combination are not as degenerate in terms of dust reddening as in $u - K_s$. Hence we define

$$\kappa_{riK} \equiv (i - K_s) - 4.6(r - i)$$

as our proxy for sSFR. In general, a higher value of κ maps to a lower sSFR. We note that κ_{riK} is a good proxy for sSFR at higher redshifts, $z \gtrsim 1.7$, but at $z \sim 1.5$ starts to lose its ability to track sSFR at lower sSFR values (light-blue curve in Fig. 20) and completely loses its ability to discriminate sSFR at lower redshifts, $z \lesssim 1.3$. Consequently, if there is a large number of galaxies with $z \lesssim 1.5$ in our sample then the strength of the κ_{riK} sSFR proxy will be diluted (but not erased) by mixing galaxies where κ_{riK} has discriminating power with ‘noise’ from those without it. With these caveats in mind, the clustering as a function of sSFR is summarized in Fig. 21.

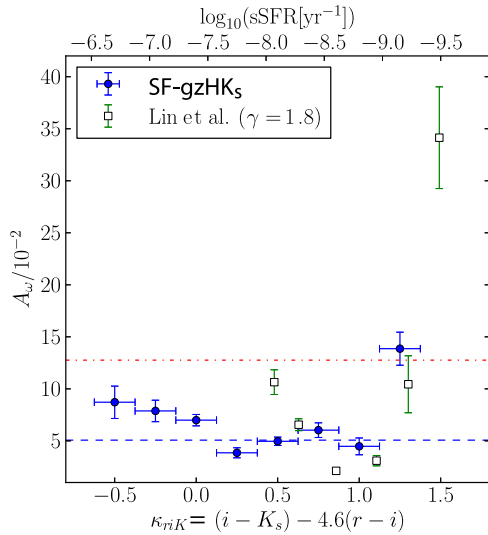


Figure 21. The clustering amplitude for SF-*gzHK* galaxies as a function of κ_{rHK} , a proxy for sSFR (blue points with 1σ error bars); see Section 4.6 for the definition. The corresponding value of sSFR for $z \approx 2$ SF-*gzHK* galaxies is indicated on the upper axis. The colour intervals are $\kappa_{rHK} = \pm 0.125$ around the points shown. The blue and red dashed lines indicate the clustering amplitude A_ω of SF-*gzHK* and PE-*gzHK* galaxies for the combined fields, respectively. The open points are the measurements from Lin et al. (2012).

In both Lin et al. (2012) and our SF-*gzHK* sample, it is clearly evident that the most quiescent $z \sim 2$ SF galaxies are strongly clustered compared to the rest. This is another confirmation that quiescent galaxies are generally more strongly clustered than active galaxies. Lin et al. (2012) additionally reported that their SF *BzK* galaxies with high sSFR (hence low κ in our analysis) were more clustered than those with moderate sSFR, the correlation length r_0 effectively exhibiting a turnover as a function of sSFR. In Fig. 21, we do not see the clustering amplitude exhibiting such a drastic turnover in our sample, although there may be a slight, gradual increase in clustering amplitude towards higher sSFR. However, the comparison is complicated by the fact that the sSFRs are derived by quite different methods. The crude nature of transformation between κ and sSFR here should be remembered, especially away from $z = 2$ where the degeneracy among different model tracks can become severe; as mentioned in the previous section, the bulk of SF-*gzHK* galaxies can be expected to be at $z < 2$ if the redshift distribution of the SF-*gzHK* galaxies were to follow those presented in previous studies. The issue of clustering dependence on sSFR is worth revisiting in the future with more precise sSFRs measurements.

5 CONCLUSIONS

We presented the angular clustering analysis of $z \sim 2$ galaxies selected via the *gzHK* colour–colour method, a substitute for the popular *BzK* technique that we adapted for the CFHT observations of the four CFHTLS deep fields. The angular correlation functions were computed via the Landy–Szalay estimator, both for each field independently and for the combined fields. With four independent fields covering in total $\sim 2.5 \text{ deg}^2$, we have one of the largest area deep surveys and could constrain $z \sim 2$ galaxy clustering through the unprecedented amount of data. In summary, our main results are as follows.

(i) Over the angular scale of $-2.0 \leq \log_{10}\theta < -0.5$, the fitted power-law slopes γ are $1.74^{+0.21}_{-0.21}$ and $1.91^{+0.34}_{-0.30}$ for SF-*gzHK*, PE-*gzHK* galaxies, respectively. The corresponding clustering amplitudes (in units of $10^{-2} \text{ arcmin}^{-(1-\gamma)}$) are $5.15^{+0.81}_{-0.80}$ and $12.73^{+2.45}_{-2.77}$ for the same galaxy types. We thus robustly confirm previous results that passive galaxies are more clustered (i.e. steeper slope γ as well as larger angular clustering amplitude A_ω) than SF galaxies at $z \sim 2$.

(ii) We find evidence for the existence of two components in angular correlation functions not just for SF galaxies (as was previously known) but also for PE galaxies. We attribute these to the longer scale clustering of underlying DM haloes (the two-halo term) and multiple galaxies associated with individual haloes (the one-halo term). The existence of a one-halo term for both PE and SF galaxies suggests that the virialization of galaxies within massive haloes has already progressed significantly by $z \sim 2$ for both SF and passive galaxy populations, leading to situations in which environmental effects can be triggered to produce passive galaxies in such environments. In other words, the presence of the one-halo term for PE galaxies (in combination with that for SF ones) can be taken as evidence for environmental quenching at $z \sim 2$. The break between the one- and two-halo terms appears to be at the angular scale $\log_{10}\theta \approx -2.7$ for both populations, corresponding to 60 kpc at $z \sim 2$.

(iii) We find that the clustering strength of PE galaxies increases as a function of rest-frame UV luminosity, indicating that the more massive among these passive galaxies reside in more massive DM haloes. In dramatic contrast, the clustering of SF galaxies decreases with increasing UV luminosity. A possible explanation of this trend is that the vigorous SF activity seen at earlier redshifts ($z \gtrsim 3$) in the most massive haloes is starting to shut down around $z \sim 2$ leaving the most massive (and thus most clustered) haloes with only low levels of star formation and, thus, low UV luminosities (see Savoy et al. 2011).

(iv) There are noteworthy differences (at the 1σ – 2σ level) in the best-fitting clustering parameters between our four large, widely separated fields. Notably, the difference from the combined-fields value is most pronounced for the D2 field, which is overlapped by the popular COSMOS field. The number counts of $z \sim 2$ galaxies in D2/COSMOS are also anomalous compared to the other fields (Arcila-Osejo & Sawicki 2013), and together these results highlight the need to study several spatially independent areas as is done in this work.

A natural continuation of the two-dimensional clustering analysis presented in this paper is to extend the analysis to three-dimensional clustering. Doing so would allow us to better link the observed clustering to the masses of the underlying DM haloes, and – in conjunction with stellar mass estimates – to gauge the connection between stars that are forming, stars that already formed, and the DM haloes that their host galaxies reside in. We hope to do so at the next opportunity.

ACKNOWLEDGEMENTS

We thank ACEnet and its staff for providing us with a wonderful high-performance computing environment without which the research presented here would not have been possible. We thank Sergiy Khan for his professional assistances throughout the project and for his desire to provide users with the best computing experiences. We thank Andrew Becker for making `HOTPANTS` publicly available, as well as for numerous assistances in getting it to run properly and understanding the underlying principle. We thank

Jerzy Sawicki for a careful reading of the manuscript and the referee for comments that helped improve the quality of this paper. This work benefited tremendously from the PYTHON programming language, its tools, and the community. Computational facilities for this work were provided by ACENet, the regional high-performance computing consortium for universities in Atlantic Canada. ACENet is funded by the Canada Foundation for Innovation (CFI), the Atlantic Canada Opportunities Agency (ACOA), and the provinces of Newfoundland and Labrador, Nova Scotia, and New Brunswick. This research was financially supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) and by an ACENet Fellowship.

This work is based on observations obtained with MegaPrime/MegaCam and WIRCam. The former is a joint project of CFHT and CEA/DAPNIA, at the Canada–France–Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. The latter is a joint project of CFHT, Taiwan, Korea, Canada, France, at the CFHT which is operated by the NRC of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii. This work is based in part on data products produced at TERAPIX and the Canadian Astronomy Data Centre as part of the Canada–France–Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS. This work is also based in part on data products produced at TERAPIX, the WIRDS (WIRCam Deep Survey) consortium, and the Canadian Astronomy Data Centre. This research was supported by a grant from the Agence Nationale de la Recherche ANR-07-BLAN-0228.

REFERENCES

- Abbas U. et al., 2010, *MNRAS*, 406, 1306
- Adelberger K. L., Steidel C. C., Pettini M., Shapley A. E., Reddy N. A., Erb D. K., 2005, *ApJ*, 619, 697
- Alard C., 2000, *A&AS*, 144, 363
- Arcila-Osejo L., Sawicki M., 2013, *MNRAS*, 436, 845
- Balogh M. L., Baldry I. K., Nichol R., Miller C., Bower R., Glazebrook K., 2004, *ApJ*, 615, L101
- Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Bielby R. M. et al., 2012, *MNRAS*, 414, 2
- Blanc G. A. et al., 2008, *ApJ*, 681, 1099
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
- Brainerd T. G., Smail I., Mould J., 1995, *MNRAS*, 275, 781
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Budavári T. et al., 2003, *ApJ*, 595, 59
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682
- Chuter R. W. et al., 2011, *MNRAS*, 413, 1678
- Coil A. L. et al., 2004, *ApJ*, 609, 525
- Coil A. L., Newman J. A., Cooper M. C., Davis M., Faber S. M., Koo D. C., Willmer C. N. A., 2006, *ApJ*, 644, 671
- Coil A. L. et al., 2008, *ApJ*, 672, 153
- Colless M. et al., 2001, *MNRAS*, 328, 1039
- Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- Daddi E., Cimatti A., Renzini A., Fontana A., Mignoli M., Pozzetti L., Tozzi P., Zamorani G., 2004, *ApJ*, 617, 746
- Dekel A., Silk J., 1986, *ApJ*, 303, 39
- Dressler A., 1980, *ApJ*, 236, 351
- Elston R., Rieke G. H., Rieke M. J., 1988, *ApJ*, 331, L77
- Fang G., Kong X., Chen Y., Lin X., 2012, *ApJ*, 751, 109
- Franx M. et al., 2003, *ApJ*, 587, L79
- Grazian A. et al., 2007, *A&A*, 465, 393
- Guhathakurta P., Tyson J. A., Majewski S. R., 1990, *ApJ*, 357, L9
- Gwyn S. D. J., 2011, *AJ*, 143, 38
- Hartley W. G. et al., 2008, *MNRAS*, 391, 1301
- Hayashi M., Shimasaku K., Motohara K., Yoshida M., Okamura S., Kashikawa N., 2007, *ApJ*, 660, 72
- Heinis S. et al., 2007, *ApJ*, 173, 503
- Hildebrandt H., Pielorz J., Erben T., van Waerbeke L., Simon P., Capak P., 2009, *A&A*, 498, 725
- Hopkins A. M., Beacom J. F., 2006, *ApJ*, 651, 142
- Hu E. M., Ridgway S. E., 1994, *AJ*, 107, 1303
- Hubble E. P., 1936, in Hubble E. P., ed., *Realm of the Nebulae*. Yale Univ. Press, New Haven
- Jing Y. P., 1998, *ApJ*, 503, L9
- Kennicutt R. C., 1998, *ARA&A*, 36, 189
- Kent S. M., 1985, *ApJS*, 59, 115
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Kong X. et al., 2006, *ApJ*, 638, 72
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Lane K. P. et al., 2007, *MNRAS*, 379, L25
- Le Fèvre O. et al., 2005, *A&A*, 439, 877
- Lee K.-S., Giavalisco M., Gnedin O. Y., Somerville R. S., Ferguson H. C., Dickinson M., Ouchi M., 2006, *ApJ*, 642, 63
- Li C., Kauffmann G., Jing Y. P., White S. D. M., Börner G., Cheng F. Z., 2006, *MNRAS*, 368, 21
- Lilly S. J., Le Fèvre O., Hammer F., Crampton D., 1996, *ApJ*, 460, L1
- Lilly S. J. et al., 2007, *ApJS*, 172, 70
- Limber D. N., 1953, *ApJ*, 117, 134
- Lin L. et al., 2012, *ApJ*, 756, 71
- Loh Y.-S. et al., 2010, *MNRAS*, 407, 55
- Ly C., Malkan M. A., Hayashi M., Motohara K., Kashikawa N., Shimasaku K., Nagao T., Grady C., 2011, *ApJ*, 735, 91
- Ma C.-P., Fry J. N., 2000, *ApJ*, 543, 503
- McCarthy P. J., 2004, *ARA&A*, 42, 477
- McCarthy P. J., Persson S. E., West S. C., 1992, *ApJ*, 386, 52
- McCracken H. J., Ilbert O., Mellier Y., Bertin E., Guzzo L., Arnouts S., Le Fèvre O., Zamorani G., 2008, *A&A*, 479, 321
- McCracken H. J. et al., 2010, *ApJ*, 708, 202
- Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, *MNRAS*, 283, 1388
- Madgwick D. S. et al., 2003, *MNRAS*, 344, 847
- Meneux B. et al., 2008, *A&A*, 478, 299
- Meneux B. et al., 2009, *A&A*, 505, 463
- Mo H. J., White S. D. M., 1996, *MNRAS*, 282, 347
- Muzzin A. et al., 2013, *ApJS*, 206, 8
- Newman J. A. et al., 2012, *ApJS*, 208, 5
- Noeske K. G. et al., 2007, *ApJ*, 660, L43
- Norberg P. et al., 2001, *MNRAS*, 328, 64
- Norberg P. et al., 2002, *MNRAS*, 332, 827
- Ouchi M. et al., 2004, *ApJ*, 611, 685
- Ouchi M. et al., 2005, *ApJ*, 635, L117
- Patil A., Huard D., Fonnesbeck C. J., 2010, *J. Stat. Softw.*, 35, 1
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Phleps S., Peacock J. A., Meisenheimer K., Wolf C., 2006, *A&A*, 457, 145
- Pollo A. et al., 2006, *A&A*, 451, 409
- Quadri R. F., Williams R. J., Franx M., Hildebrandt H., 2012, *ApJ*, 744, 88
- Reddy N. A., Erb D. K., Steidel C. C., Shapley A. E., Adelberger K. L., Pettini M., 2005, *ApJ*, 633, 748
- Rees M. J., Ostriker J. P., 1977, *MNRAS*, 179, 541
- Roche N., Eales S. A., 1999, *MNRAS*, 307, 703
- Ross A. J., Brunner R. J., 2009, *MNRAS*, 399, 878
- Ross A. J., Percival W. J., Brunner R. J., 2010, *MNRAS*, 407, 420
- Savoy J., Sawicki M., Thompson D., Sato T., 2011, *ApJ*, 737, 92
- Sawicki M., 2012a, *PASP*, 124, 1208

- Sawicki M., 2012b, MNRAS, 421, 2187
 Sawicki M., Yee H. K. C., 1998, AJ, 115, 1329
 Sawicki M. J., Lin H., Yee H. K. C., 1997, AJ, 113, 1
 Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
 Seljak U., 2000, MNRAS, 318, 203
 Silk J., 1977, ApJ, 211, 638
 Simard L. et al., 2002, ApJS, 142, 1
 Simon P., Hetterscheidt M., Wolf C., Meisenheimer K., Hildebrandt H., Schneider P., Schirmer M., Erben T., 2009, MNRAS, 398, 807
 Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJ, 462, L17
 Swanson M. E. C., Tegmark M., Blanton M., Zehavi I., 2008, MNRAS, 385, 1635
 Teerikorpi P., 2004, A&A, 424, 73
 Thompson D. et al., 1999, ApJ, 523, 100
 Wake D. A. et al., 2011, ApJ, 728, 46
 Wall J. V., Jenkins C. R., 2003, Practical Statistics for Astronomers. Cambridge Univ. Press, Cambridge
 Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057
 York D. G. et al., 2000, AJ, 120, 1579
 Zehavi I. et al., 2002, ApJ, 571, 172
 Zehavi I. et al., 2004, ApJ, 608, 16
 Zehavi I. et al., 2005, ApJ, 630, 1
 Zehavi I. et al., 2011, ApJ, 736, 59
 Zheng Z. et al., 2005, ApJ, 633, 791

APPENDIX A: PRODUCING PSF-MATCHED IMAGES

Substantial PSF variations exist across the individual stacked images. In visible bands, such PSF variations are typically smooth over the entire image and can be modelled by low-order polynomials. In near-infrared images, variations are much more complicated due to dithering and mosaicking the limited field of view (Fig. 1). In order to better sample the light from the same parts of objects, images are generally degraded to the worst seeing before carrying out fixed-aperture photometry.

A customized version of `HOTPANTS`⁶ by Andrew Becker is used to find convolution kernels that match PSFs between two images. The program implements the algorithm devised by Alard (2000), constructing a spatially varying kernel K that minimizes the differences between point source stamps in reference and given images $R(x, y)$ and $I(x, y)$, respectively:

$$\sum_i ([R \otimes K](x_i, y_i) - I(x_i, y_i))^2. \quad (\text{A1})$$

The kernel K is decomposed into some basis functions

$$K(u, v) = \sum_n a_n K_n(u, v),$$

⁶ <http://www.astro.washington.edu/users/becker/hotpants.html>

which should be any reasonable orthogonal functions but typically Gaussians with varying widths are used:

$$K_n(u, v) = e^{-(u^2+v^2)/2\sigma_k^2} u^i v^j,$$

where $n = (i, j, k)$. The solutions found over the image are connected via

$$K(u, v) = \sum_n a_n(x, y) K_n(u, v),$$

where

$$a_n(x, y) = \sum_{i,j} b_{i,j} x^i y^j.$$

This effectively finds a polynomial solution that models the smoothly varying convolution kernel over the entire (or parts of an) image. The primary advantage of the algorithm is that it makes computation of spatially varying kernel efficient; see Alard (2000) for details.

Originally designed for detecting time-varying signals by subtracting one PSF-matched image from another, `HOTPANTS` implementation assumes that images R and I are taken through the same passband, with the only variable being their PSFs (due to different seeing, for example). Our modifications are made to match PSFs on images taken through different passbands. This involves an introduction of a factor r_i in equation (A1) which takes into account the flux ratios of point sources in R and I that are used to find the kernel K :

$$\sum_i ([R \otimes K](x_i, y_i) - r_i I(x_i, y_i))^2. \quad (\text{A2})$$

`SEXTRACTOR` is run on all images individually to construct point source catalogues (via `CLASS_STAR`), which are candidate point sources to be used in PSF matching (Fig. A1). Their full width at half-maximums (FWHMs) and light profiles are visually inspected and those sources with abnormal values are removed from the subsequent processing. The modified `HOTPANTS` is then used to match PSFs using the point sources and their flux ratios between two images (Fig. A1).

Since the kernel basis functions that we use are fairly generic superpositions of three or four Gaussians with varying σ_k , our ability to accurately match PSFs between two images is limited by the rather arbitrary choice of kernel basis. In practice, however, most existing studies use a simpler convolution kernel for PSF matching on a smaller field, if at all. The degree of desired accuracy really depends on science goals, and given our wide field of view and the significant differences in seeing/image quality through different passbands, the PSF correction at the level we carry out is necessary and sufficient. For fields D1, D3, and D4, the u -band image has the largest PSF (≈ 0.75 arcsec), to which the rest of the bands are convolved. In D2, the J band has the largest PSF (≈ 0.9 arcsec) and the other images were convolved to that band.

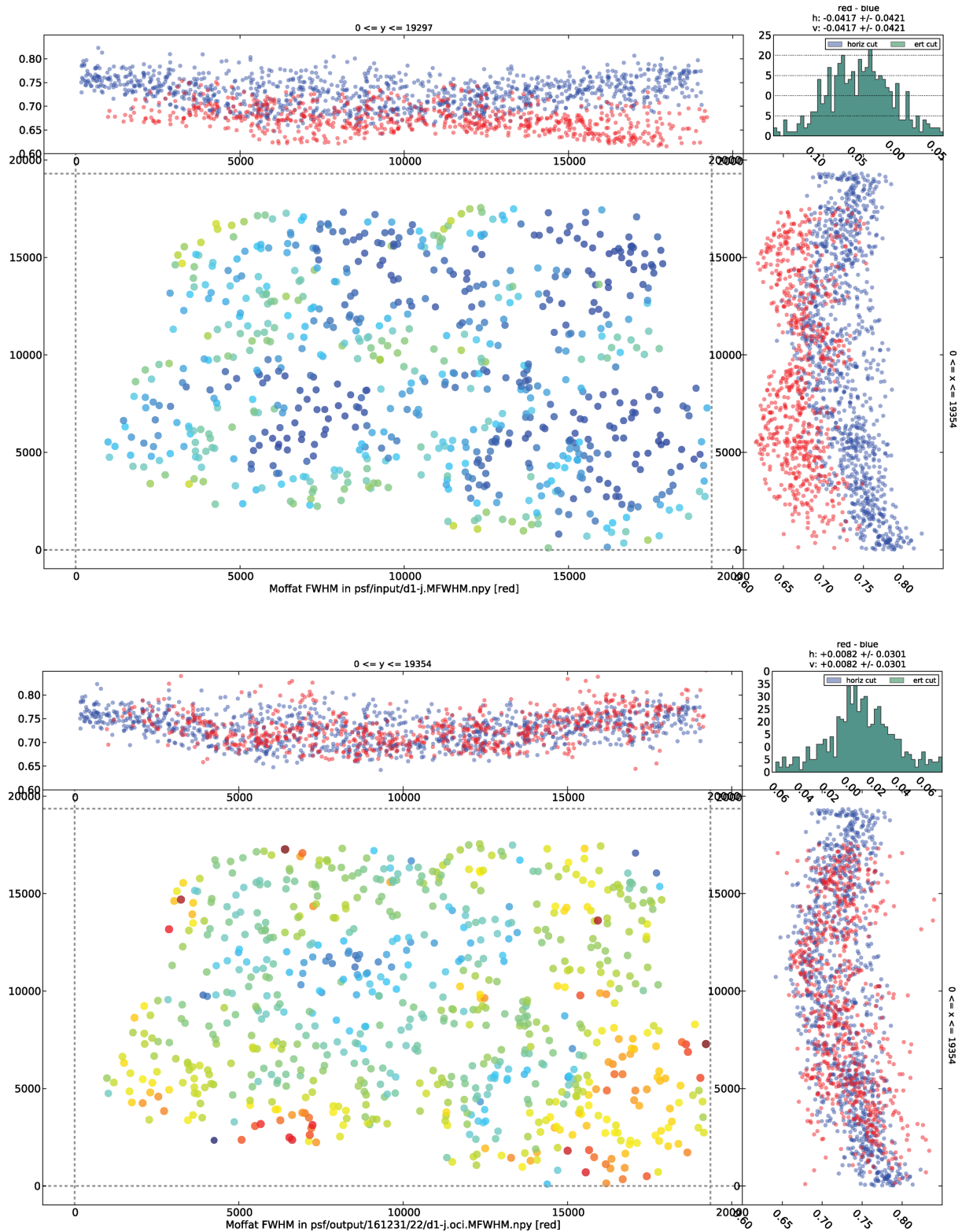


Figure A1. The output from the graphical user interface for the PSF matching utility (between u and J bands in the D1 field in this specific case). The top panel is for before PSF matching and the bottom panel after. The main plot shows the spatial distribution of point sources used for PSF matching, with the colours mapped to their FWHMs. The side plots indicate the spatial variation of the point source size in u band (blue) and J band (red). The spatial variations look worse than they actually are, since the data points are collapsed both vertically and horizontally over the whole image. In practice, smaller slices are made for inspection which make the scatters after the PSF matching more reasonable.

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