

The *Spitzer* Extragalactic Representative Volume Survey and DeepDrill extension: clustering of near-infrared galaxies

Eelco van Kampen,¹★ Mark Lacy,² Duncan Farrah,^{3,4} Claudia del P. Lagos,^{5,6,7} Matt Jarvis⁸,
 Claudia Maraston,⁹ Kristina Nyland¹⁰, Seb Oliver,¹¹ Jason Surace¹² and Jessica Thorne⁵

¹European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching bei München, Germany

²NRAO, 520 Edgemont Road, Charlottesville, VA 22903, USA

³Department of Physics and Astronomy, Watanabe 416, 2505 Correa Road, Honolulu, HI 96822, USA

⁴Institute for Astronomy, 2680 Woodlawn Drive, Honolulu, HI 96822-1897, USA

⁵International Centre for Radio Astronomy Research (ICRAR), M468, University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia

⁶ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

⁷Cosmic Dawn Center (DAWN), Niels Bohr Institute, University of Copenhagen, Jagtvej 128, DK-2200 Copenhagen N, Denmark

⁸Oxford Astrophysics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

⁹Institute of Cosmology and Gravitation, Dennis Sciama Building, Burnaby Road, Portsmouth PO1 3FX, UK

¹⁰U.S. Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC 20375, USA

¹¹Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Falmer, Brighton BN1 9QH, UK

¹²IPAC, California Institute of Technology, Pasadena, CA 91125, USA

Accepted 2023 May 11. Received 2023 April 28; in original form 2022 April 29

ABSTRACT

We have measured the angular autocorrelation function of near-infrared galaxies in SERVS + DeepDrill, the *Spitzer* Extragalactic Representative Volume Survey and its follow-up survey of the Deep Drilling Fields, in three large fields totalling over 20 deg² on the sky, observed in two bands centred on 3.6 and 4.5 μm . We performed this analysis on the full sample as well as on sources selected by [3.6]–[4.5] colour in order to probe clustering for different redshift regimes. We estimated the spatial correlation strength as well, using the redshift distribution from S-COSMOS with the same source selection. The strongest clustering was found for our bluest subsample, with $\langle z \rangle \sim 0.7$, which has the narrowest redshift distribution of all our subsamples. We compare these estimates to previous results from the literature, but also to estimates derived from mock samples, selected in the same way as the observational data, using deep light-cones generated from the SHARK semi-analytical model of galaxy formation. For all simulated (sub)samples, we find a slightly steeper slope than for the corresponding observed ones, but the spatial clustering length is comparable in most cases.

Key words: surveys – galaxies: evolution – galaxies: formation – galaxies: statistics – cosmology: observations – infrared: galaxies.

1 INTRODUCTION

Studying galaxy environments and their effects on galaxy evolution is one of the main motivations for large, wide-area galaxy surveys. Galaxies are not spread evenly throughout the Universe, but cluster in pairs, groups, and clusters: their distribution is clustered. One consequence of that is that they interact, where the amount of clustering, which is usually quantified using the 2-point correlation function for galaxies (Peebles 1980), and its characteristic length-scale, the clustering strength r_0 , will determine the frequency and severeness of galaxy–galaxy or galaxy–group interaction, and thus the evolutionary history of the galaxy population.

We employ two recent ‘warm *Spitzer*’ surveys to estimate the 2-point galaxy correlation function (usually just called the ‘galaxy correlation function’ for short) for catalogues extracted from the *Spitzer* Extragalactic Representative Volume Survey (SERVS), which covers

18 deg² to 5 σ depths of 1 μJy at 3.6 μm and 2 μJy at 4.5 μm . SERVS, a *Spitzer* ‘Exploration Science’ program (for full details this survey, see Mauduit et al. 2012), and its extension, DeepDrill, performed in a similar fashion (see Lacy et al. 2021 for details). Both are deep enough and wide enough to contain a truly representative volume of the Universe, sampling 0.8 Gpc³ over $1 < z < 6$ in the case of SERVS, and a similar volume for DeepDrill, which was aimed at doubling the original survey size of SERVS. The three SERVS + DeepDrill fields studied here are large enough to reliably estimate the (rest-frame) NIR galaxy correlation function for each field individually, focusing on the *Spitzer* IRAC bands centred on 3.6 and 4.5 μm . In addition, the three fields taken together provide us a large total area for a combined clustering analysis, in order to suppress any remaining sample variance (often called ‘cosmic variance’ in a cosmological context).

SERVS and DeepDrill cover a fairly wide redshift range, peaking at $z \approx 1$. Photometric redshifts are now available (Pforr 2019) for SERVS, employing 12–15 bands (depending on the field) of supplementary data, but their coverage and depth are unfortunately still inhomogeneous and we lack redshifts for the DeepDrill extension.

* E-mail: evkampen@eso.org

We also have few (less than 2 percent) spectroscopic redshifts. So instead of using these, we use colour cuts to select redshift distributions, allowing us to estimate spatial clustering lengths through the Limber equation inversion technique.

Galaxies in the visual wavebands have clustering lengths of the order of 2–10 Mpc, depending on colour, luminosity, and redshift interval, e.g. Coil et al. (2008) for the DEEP2 sample, Zehavi et al. (2011) for the SDSS DR7 sample, and Chistodoulou et al. (2012) for the GAMA sample.

In the far-IR wavebands, for sources with $S_{250\ \mu\text{m}} > 33\ \mu\text{Jy}$, $r_0 \approx 3\text{--}4\ h^{-1}$ Mpc below $z < 0.3$ (van Kampen et al. 2012; Amvrosiadis et al. 2019), while at high redshifts, sub-mm galaxies are more strongly clustered with correlation lengths r_0 of 8.1 ± 0.5 , 8.8 ± 0.8 , and $13.9 \pm 3.9\ h^{-1}$ Mpc at $z = 1\text{--}2$, $2\text{--}3$, and $3\text{--}5$, respectively (Amvrosiadis et al. 2019). In the mid-IR, a clustering estimate at $24\ \mu\text{m}$ is given by Gilli et al. (2007): $r_0 = 4.0 \pm 0.4\ h^{-1}$ Mpc for $f_{24} > 20\ \mu\text{Jy}$, with no redshift selection. Later, Starikova et al. (2012) found $r_0 = 4.98 \pm 0.28\ h^{-1}$ Mpc and $r_0 = 8.04 \pm 0.69\ h^{-1}$ Mpc for $\langle z \rangle = 0.7$ and $\langle z \rangle = 1.7$ populations, respectively, for sources with $S_{24\ \mu\text{m}} > 310\ \mu\text{Jy}$.

The first clustering estimate in the near-infrared was obtained in the *K*-band (i.e. around $2.2\ \mu\text{m}$) for a galaxy population at relatively low redshifts by Baugh et al. (1996). Estimates from deeper and wider surveys followed, e.g. Waddington et al. (2007), de la Torre et al. (2007), Furusawa et al. (2011), Hatfield et al. (2016), Cochrane et al. (2018), and others. These were performed for a variety of NIR bands, survey sizes, and depths. Our current SERVS + DeepDrill sample is larger than these previous efforts, with good survey homogeneity and completeness down to AB magnitude 22 in both the 3.6 and $4.5\ \mu\text{m}$ bands.

The purpose of this paper is to measure the spatial clustering length of flux-selected NIR galaxies in SERVS + DeepDrill to the best accuracy possible, for the full survey as well as for specific subsamples, but also to compare to recent modelling efforts for this population. A very suitable set of simulations for this purpose is provided by the SHARK semi-analytical model (Lagos et al. 2018, 2019), which make use of the SURFS *N*-body simulations (Elahi et al. 2018), the SED modelling software ProSpect (Robotham et al. 2020) and the radiative transfer results of the EAGLE hydrodynamical simulations to model attenuation (Trayford et al. 2020).

These multiwavelength mock data sets include *Spitzer* bands to a depth exceeding our observational data. In fact, the full simulated volume comprises a (wide) light-cone with an area of $107.9\ \text{deg}^2$, a flux selection at the $[3.6]$ band $> 0.575\ \mu\text{Jy}$ (equivalent to an AB magnitude of 24.5), and a redshift range $0 \leq z \leq 6$.

In a first comparison of these simulations to the DeepDrill data, Lacy et al. (2021) found that for number counts the agreement between the model and observations is good overall, particularly at faint magnitudes ($AB > 22$). At intermediate magnitudes, $\sim 18\text{--}21$, the model light-cone predicts a lower number of galaxies compared to both our counts and those in the literature of up to a factor of 1.8 in the $[3.6]$ band and 2.2 in the $[4.5]$ band. We find the models somewhat too faint (~ 0.6 mag), with a small colour difference of 0.05 mag. The main reason for this is probably that the SHARK simulations contain only the emission of galaxies, without the inclusion of active galactic nuclei (AGNs), which are expected to make a larger contribution at intermediate magnitudes. Other aspects of the model, like the star formation history, dust treatment, or the adopted population synthesis model, could also contribute to the number count mismatch.

This paper is organized as follows: in Section 2, we describe in more detail the observational data and simulations used, in Section 3

Table 1. As the SERVS and DeepDrill data were taken at different epochs during the *Spitzer* warm mission, there is a small difference in their fluxes, which we list here as ratios for each band and each field (for which the central coordinates are provided here too). These ratios were set by equalising the overall galaxy numbers density of the two fields in each band, while retaining a colour distribution matching that of S-COSMOS. The ratios thus found are consistent with those listed in the survey paper (Lacy et al. 2021).

Field	RA	Dec.	DeepDrill/SERVS flux ratios	
			$3.6\ \mu\text{m}$	$4.5\ \mu\text{m}$
ES1	00:37:56.2	−44:01:52	0.952	0.959
XMM-LSS	02:22:18.5	−04:48:10	0.957	0.983
CDFS	03:31:50.1	−28:07:06	0.951	0.969

Table 2. Completeness as derived from tables 8 and 9 of Lacy et al. (2021). The completeness fraction for the full sample (denoted as ‘all galaxies’) considered for our clustering analysis, i.e. those in the magnitude interval $18 < AB < 22$, is the weighted mean of the fractions per bin.

AB Mag bin	Completeness fraction	
	$3.6\ \mu\text{m}$	$4.5\ \mu\text{m}$
18–19	1.0	1.0
19–20	0.97	0.97
20–21	0.945	0.94
21–22	0.89	0.915
18–22	0.93	0.94

we elaborate on the methods used, and present in Section 4 the results of applying these methods to both the observed and mock data sets, including a range of subsamples constructed by selecting on colour and/or luminosity. In Section 5, we summarize and discuss our findings. We adopt the same cosmological parameters as used for the SHARK simulations (which we compare to), who used the Planck Collaboration XIII (2016) Lambda cold dark matter cosmology, so we set $H_0 = 67.51\ \text{km s}^{-1}\ \text{Mpc}^{-1}$, $\Omega_\Lambda = 0.6751$, and $\Omega_m = 0.3121$. Clustering length are quoted in h^{-1} Mpc, to facilitate comparison of our estimates to various previous ones found in the literature.

2 OBSERVATIONAL AND MODEL DATA USED

2.1 *Spitzer* imaging and catalogues

The *Spitzer* data, taken in the 3.6 and $4.5\ \mu\text{m}$ bands, were collected in two major efforts, both in the post-cryogenic (‘warm’) part of the *Spitzer* mission. In the early months of that, SERVS (Mauduit et al. 2012) imaged five highly observed astronomical fields (ELAIS-N1, ELAIS-S1, Lockman Hole, Chandra Deep Field South, and XMM-LSS), requiring 1400 h of observing time. Three of these fields (listed in Table 1) form the central areas of the three DeepDrill fields used for this paper: the DeepDrill Survey (Program ID 11086) was observed between 2015-05-04 and 2016-12-26 and specifically designed to extend three SERVS fields, again requiring 1400 h, and doubling the original SERVS survey area. Due to scheduling constraints, the coverage is non-uniform, with areas around the edges of the SERVS fields in particular receiving additional coverage, and some outlying regions not receiving the full coverage. We take this into account in our clustering analysis by excluding these outlying regions.

DeepDrill data processing was similar to that what was done for the SERVS data set (Mauduit et al. 2012), using an existing data cleaning pipeline derived from processing SWIRE and COSMOS

data (Lonsdale et al. 2003; Sanders et al. 2007). In order to achieve homogeneous coverage across each full DeepDrill field, all SERVS image data contained therein were recalibrated to the final warm mission calibration. Full details on the final imaging, source extraction, and matching of the two catalogues (the [3.6] and [4.5] bands) are found in Lacy et al. (2021), and references therein (notably Vaccari 2016).

The total area of the three fields covered by SERVS + DeepDrill is 28 sq. degrees, of which just over 20 sq. degrees is used for the clustering estimates in this paper. The area not included consists of border regions where the survey coverage is inhomogeneous or incomplete. Circular regions around bright stars, in which the source extraction was very incomplete, have also been excluded. This was done by placing circular masks around stars brighter than $K = 8$ in the 2-Micron All Sky Survey (2MASS; Skrutskie et al. 2006) whose size was determined empirically by measuring the radius out to which the wings of the stellar PSF prevented reliable object detection. A regression line was then fit to this radius as a function of K -band magnitude, and objects within a radius $318\text{--}35.2K$ arcseconds (where K is the 2MASS K -band magnitude) of the 2MASS position were masked out. The final sample area consists of three round areas with a radius of 1.5 deg, with the star masks excluded.

Because the SERVS and DeepDrill data sets were taken at different epochs of the ‘warm’ part of the *Spitzer* mission, there is a small difference in photometric calibration between the two data sets. This is discussed in more detail in the survey paper (Lacy et al. 2021), where the ratio between the fluxes in each survey is given for each band in each field. Due to the survey geometry it is important to take this into account as well as we can: the SERVS fields are rectangular, which were extended all around by DeepDrill to yield a circular field of view with the SERVS field in the middle. Any mismatch in flux calibration would result in a different mean source density for the SERVS rectangle as compared to the circular surrounding of the DeepDrill pointings. Such a large-scale mismatch produces an obvious artificial bump in the large-scale part of the angular correlation function, as we don’t expect any ring-like large-scale structure in the galaxy density distribution. We therefore equalise the mean number density of sources for the two data sets to derive a flux ratio that removes this artificial bump (which would arise from small differences in photometric calibration between the SERVS and DeepDrill datasets).

One subtlety here is that we need to do this for both *Spitzer* bands, where different flux ratios for each band changes the colour distribution of the population somewhat. We therefore constrain the colour distribution to be as close as possible to that derived for the S-COSMOS survey (Sanders et al. 2007), which was observed in the same *Spitzer* bands as SERVS + DeepDrill, while at the same time matching the mean source number density for both bands, which is done using number density maps for each *Spitzer* band. This exercise yields the flux ratios listed in Table 1, which are somewhat more precise but still consistent with the flux ratios reported in (Lacy et al. 2021), who derived these from the sources in the overlap regions of the SERVS and DeepDrill, i.e. sources found in both surveys.

Once this flux ratio correction is applied, we cut our source sample to account for completeness (in both magnitude and colour), as defined in the survey definition paper (Mauduit et al. 2012): the full sample is complete down to an AB magnitude of just over 23 in both bands. However, we cut at $[3.6] < 22$ and $[4.5] < 22$ to allow for colour cuts with a range of about a magnitude, so that both band will always remain sufficiently complete. We also cut at $[3.6] > 18$ and $[4.5] > 18$, as the contribution of stars increases significantly at brighter magnitudes (Sanders et al. 2007).

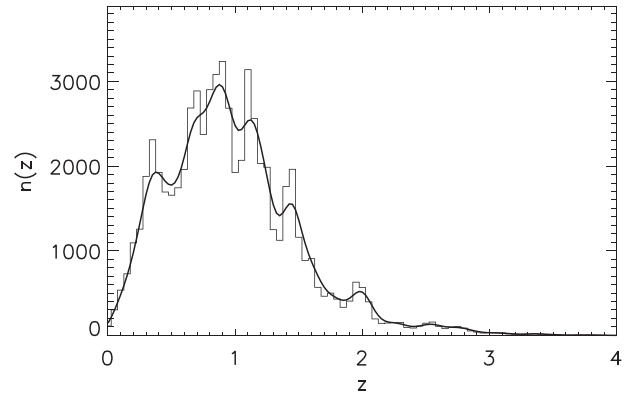


Figure 1. Histogram of the photometric redshift distribution for S-COSMOS galaxies with $18 < [3.6] < 22$ and $18 < [4.5] < 22$, which is the overall flux cut used for our data set for all clustering measures. The smooth function is a Gaussian smoothed (with a width $\sigma_z = 0.1$) version of the histogram, and is subsequently used for estimating the spatial clustering strength (see main text for details).

2.2 Redshift distributions from colour cuts

Following Papovich et al. (2008), we employ the [3.6]–[4.5] colour to select subsamples with a more restricted redshift distribution than the full sample. For angular clustering measures, a narrow redshift range results in a much stronger clustering signal as it is not averaged out as much along the line of sight. The cleanest subsample is derived from the bluest galaxies: Papovich et al. (2008) found that cutting at $[3.6]\text{--}[4.5] < -0.6$ selects (blue) galaxies in a relatively narrow redshift range around $z \sim 0.7$. They also found that cutting at $[3.6]\text{--}[4.5] > -0.1$ selects for predominantly high redshift galaxies ($z > 1.3$).

In order to establish which redshift distribution results from which cuts, both in colour and magnitude, we use the 2 sq. degrees S-COSMOS field (Sanders et al. 2007) observed in the same *Spitzer* bands as SERVS + DeepDrill, and with a wealth of supplementary data (30 bands and homogeneous coverage), which yielded a large number of spectroscopic and photometric redshifts (Ilbert et al. 2009). For our purposes the (Sanders et al. 2007) catalogue is sufficient, as it has a similar depth and completeness as our sample, but we would like to mention that quite recently a deeper COSMOS catalogue became available (Weaver et al. 2022). The photometric redshift distribution for the S-COSMOS sample down to our flux limit in the [3.6] and [4.5] bands is shown in Fig. 1. We currently lack sufficient photometric redshifts for our own samples: we only have these for SERVS (see Pforr et al. 2019), and even for these the coverage and depth are unfortunately fairly inhomogeneous because some data lack the minimum of five bands needed for finding a reliable photometric redshift.

The choice of which colour cuts to use is somewhat arbitrary. The aim is to produce subsamples with sufficient numbers of galaxies each and a redshift distribution that is as narrow as possible and without too many outliers (both at lower and higher redshifts). Using the S-COSMOS sample, we obtain redshift distributions for a range of [3.6]–[4.5] colour cuts, with fixed magnitude cuts in both bands as for our own samples: $18 < [3.6] < 22$ and $18 < [4.5] < 22$. Two sets, one for negative (blue) cuts and one for positive (red) cuts, of such (normalized and somewhat smoothed) redshift distributions are shown in Fig. 2. The actual number of galaxies left in each subsample is decreasing top-down for each of the panels.

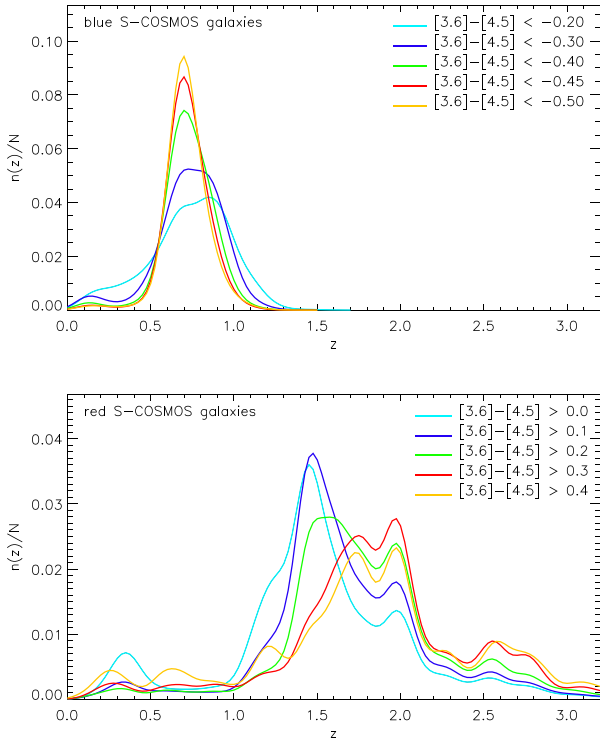


Figure 2. Photometric redshift distribution for S-COSMOS galaxies for different $[3.6]-[4.5]$ colour cuts and down to our flux cuts. These distributions are taken as representative for subsamples taken from our own survey (which lacks sufficient numbers of photometric redshifts) with the same colour cuts. The top panel is for blue galaxies, the bottom panel red ones. The aim is to find those colour cuts that yield the narrowest redshift distributions (see text for details). As the area of the S-COSMOS field is 2 deg^2 , these distributions are not that smooth (especially for the red subsamples).

Based on Fig. 2, we choose two blue and two red cuts for our clustering analysis. To fully exploit the large sample size while still restricting the redshift range as compared to the full sample, we first cut moderately at $[3.6]-[4.5] < -0.3$ and $[3.6]-[4.5] > 0.1$ to form a blue and red subsample, respectively (the blue lines in Fig. 2), each containing roughly a quarter of all galaxies. The redshift distribution of the blue subsample is relatively clean and still somewhat broad, but not as broad as the $[3.6]-[4.5] < -0.2$ subsample (cyan line in Fig. 2). The redshift distribution of the red subsample has a clear peak near $z \approx 1.5$, but with a high redshift tail and a second, smaller peak near $z \approx 2$. It does lack the significant fraction of low-redshift source that a $[3.6]-[4.5] > 0$ cut includes, which partly motivates our choice for a cut at $[3.6]-[4.5] > 0.1$ for the red sample.

For the blue part of the sample a narrower redshift distribution can be achieved by cutting more severely. A cut at $[3.6]-[4.5] < -0.5$ (yellow line in the top panel of Fig. 2) produces the highest peak and the lowest fraction of low and high redshift galaxies, but yields too small a sample for a robust clustering measurement. We therefore cut at $[3.6]-[4.5] < -0.45$ instead (red line in the top panel of Fig. 2), thus retaining sufficient numbers of galaxies for a clean clustering analysis. Note that beyond $[3.6]-[4.5] < -0.5$ the peak decreases again, and yielding very small subsample sizes too. The peak of these ‘very blue’ galaxies is not that different from the ‘blue’ ones. Such a narrow redshift distribution cannot be achieved in the red part of the sample, and what is noteworthy here is that the peak of the redshift distribution moves with the choice of the colour cut. For a ‘very red’ subsample we choose to cut at $[3.6]-[4.5] > 0.3$

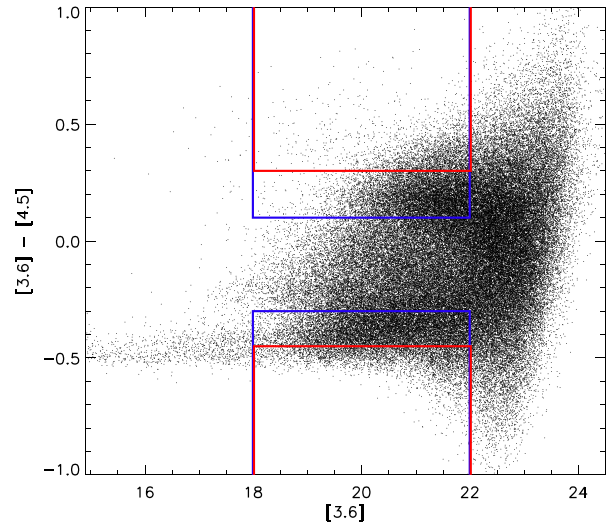


Figure 3. Colour selection and flux cut for the XMM field shown in the colour–magnitude relation of all observed galaxies, including those beyond the flux cut (in both bands). The coloured lines correspond to the coloured redshift distributions shown in Fig. 2, where the top lines, selecting for red galaxies, correspond to those in the bottom panel of Fig. 2, and vice-versa for the bottom lines (selecting for blue galaxies).

(red line in the bottom panel of Fig. 2), as this gives the highest peak at around $z \approx 2$, and less foreground contamination as compared to an even more severe cut. However, there is a significant fraction of high redshift galaxies beyond $z > 2.5$ that cannot be avoided by a $[3.6]-[4.5]$ colour cut alone.

The chosen cuts are shown in Fig. 3, which is the colour–magnitude diagram for the XMM field. The red and blue galaxy distributions resulting from these cuts are shown in Figs 4 and 5. Holes appear where bright star masks have been applied in the observed distribution. The same masks are also applied to the random catalogues used to estimate the clustering strength. The resulting subsamples for each field analysed in this paper are listed in the first three parts of Table 4, where the second column displays the number of galaxies remaining after each selection. The fourth part of the table displays the numbers for the three fields taken together, which is what we use for the combined clustering estimates, helping to overcome field-to-field variations. The bottom part of the table concerns the simulated galaxy populations, discussed next.

2.3 Simulated data used

We employ the SHARK semi-analytical model of galaxy formation (Lagos et al. 2018, 2019), in which the intrinsic SEDs are modelled based on the galaxy’s star formation and gas metallicity history, to then be attenuated based on the galaxy’s dust surface density and employing the radiative transfer results of EAGLE from Trayford et al. (2020). Re-emission in the IR is done using the templates of Dale et al. (2014). We refer the reader to Lagos et al. (2019) and Robotham et al. (2020) for more details on how SEDs are produced in SHARK. The outcome of this process is a set of mock data sets which are at least as complete as the observational data, in both *Spitzer* bands. We cut three simulated light-cones from the complete 107.9 deg^2 SHARK volume with the same field of view and depth as for the three observed SERVS + DeepDrill fields.

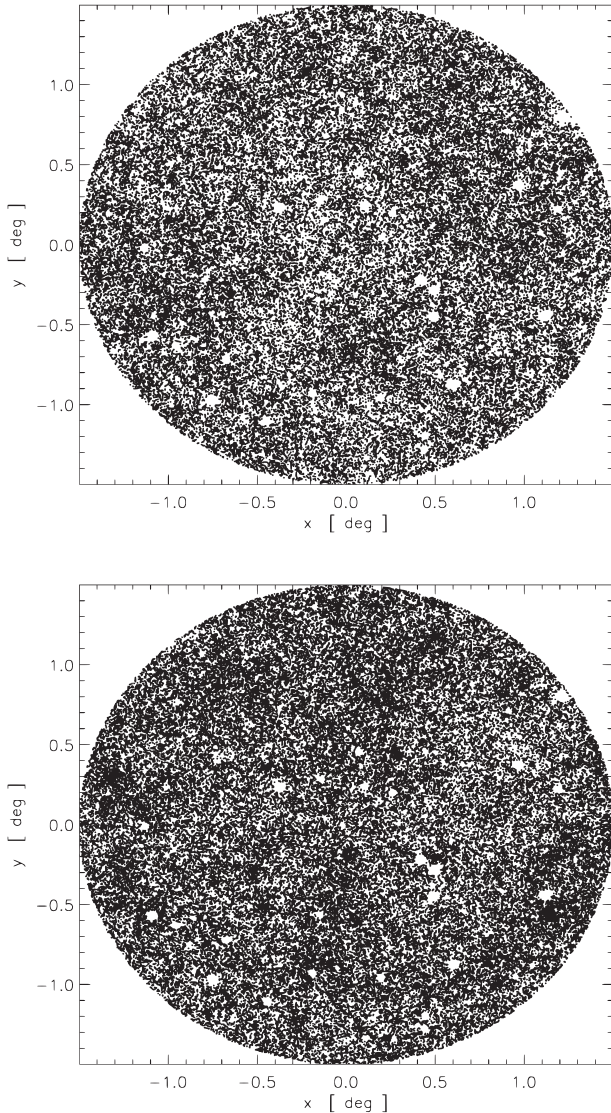


Figure 4. Red (top) and blue (bottom) galaxies selected for the CDFS field. These colour selections, $[3.6]-[4.5] > 0.1$ and $[3.6]-[4.5] < -0.3$, respectively, are aimed at retaining a relatively large number of galaxies for each colour, while still significantly narrowing the redshift distribution as compared to the full sample. The star masks are clearly visible as white circular areas.

2.3.1 Mimicking photometric accuracy

For the flux range we consider, the photometry of the SERVS + DeepDrill data has a fractional uncertainty $\Delta f_{[3.6]}/f_{[3.6]}$ that varies with flux roughly as $f_{[3.6]}^{-0.5}$, as derived from the photometric catalogue of Lacy et al. (2021). This uncertainty with respect to the true flux is important for the position of a galaxy in the colour–magnitude diagram. There is a density gradient in this diagram on the lines representing our flux and colour cuts (the coloured line in Fig. 3). As the photometric uncertainty diffuses positions in the colour–magnitude diagram, this gradient causes more galaxies to move to lower density parts of the diagram (i.e. more extreme colours) than vice-versa, and affects the subsamples selected by the flux and/or colour cuts.

The mock galaxies have no photometric uncertainty, so this diffusion does not take place, and fewer mock galaxies will scatter

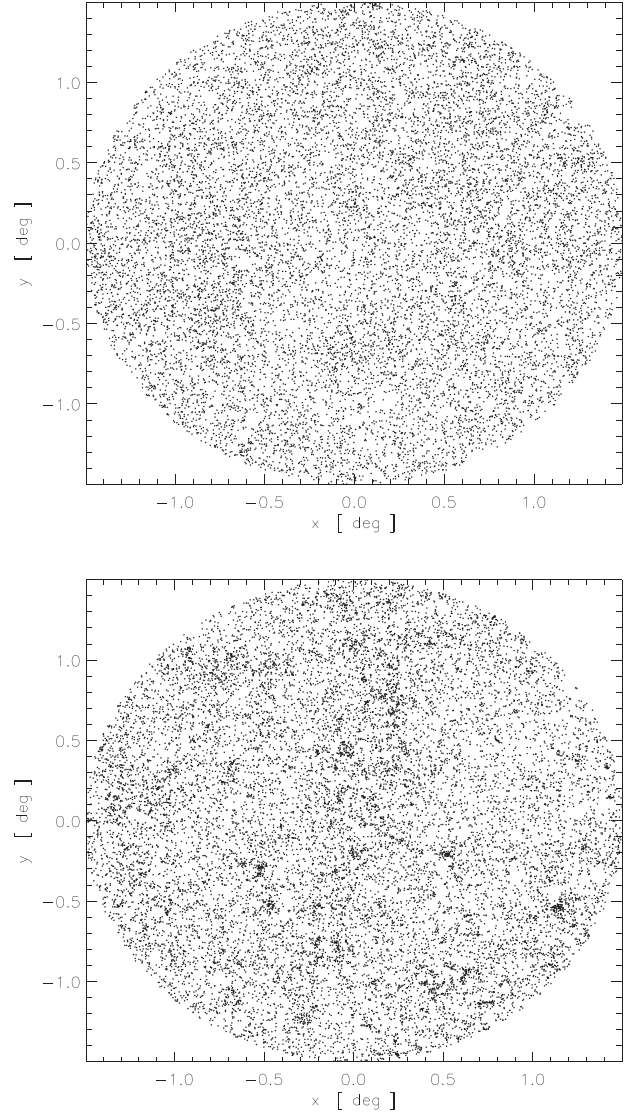


Figure 5. ‘Very red’ (top) and ‘very blue’ (bottom) galaxies selected for the CDFS field. The aim of ‘very red’ selection ($[3.6]-[4.5] > 0.3$) is to select the highest redshift distribution, with a similar dispersion as the blue and red population shown in Fig. 4, whereas the ‘very blue’ cut ($[3.6]-[4.5] < -0.45$) yields the narrowest redshift range with a clear peak $z \sim 0.7$. The clumpiness of the latter is easy to spot by eye.

Table 3. Typical properties for the galaxy population in our five subsamples, as derived from the (brightened) mock catalogues. The table lists median values for redshift, star formation rate, and stellar mass. The latter two are expected to be somewhat larger for the corresponding observed samples, as in the step of brightening the mocks these properties were not changed.

Subsample	Colour cut [3.6]–[4.5]	Median z	Median sfr $M_{\odot} \text{ yr}^{-1}$	Median M_{*} $10^{10} M_{\odot}$
All	–	1.08	4.7	1.79
Very blue	< -0.45	0.68	1.2	4.57
Blue	< -0.3	0.72	2.5	1.23
Red	> 0.1	1.77	12.3	2.63
Very red	> 0.3	1.89	28.2	5.62

Table 4. Clustering measures for all samples considered, observational as well as simulated (see main text for details) for the two-parameter fits. This table also lists the actual sizes of all samples used, as well as the derived spatial clustering strength using the Limber equation inversion technique.

Sample	N	A (arcsec)	δ	r_0 (h^{-1} Mpc)
ES1, all galaxies	289 256	0.37 ± 0.08	0.79 ± 0.04	5.08 ± 1.13
ES1, [3.6]–[4.5] < -0.45	16 352	8.32 ± 1.13	1.04 ± 0.08	9.38 ± 1.44
ES1, [3.6]–[4.5] < -0.3	81 600	2.28 ± 0.26	0.91 ± 0.03	6.40 ± 0.76
ES1, [3.6]–[4.5] > 0.1	69 507	1.59 ± 0.28	0.85 ± 0.04	7.79 ± 1.41
ES1, [3.6]–[4.5] > 0.3	13 301	2.74 ± 1.03	0.90 ± 0.10	10.64 ± 4.16
CDFS, all galaxies	291 734	0.22 ± 0.06	0.67 ± 0.04	4.87 ± 1.37
CDFS, [3.6]–[4.5] < -0.45	17 501	12.77 ± 0.87	0.92 ± 0.05	11.71 ± 1.02
CDFS, [3.6]–[4.5] < -0.3	84 930	2.15 ± 0.27	0.78 ± 0.04	6.75 ± 0.91
CDFS, [3.6]–[4.5] > 0.1	73 846	1.32 ± 0.26	0.72 ± 0.04	8.01 ± 1.61
CDFS, [3.6]–[4.5] > 0.3	14 668	0.88 ± 0.68	0.65 ± 0.12	8.08 ± 6.37
<i>XMM</i> , all galaxies	288 597	0.24 ± 0.09	0.67 ± 0.06	5.14 ± 2.04
<i>XMM</i> , [3.6]–[4.5] < -0.45	16 619	8.63 ± 1.27	0.86 ± 0.06	9.75 ± 1.60
<i>XMM</i> , [3.6]–[4.5] < -0.3	81 216	1.94 ± 0.23	0.78 ± 0.03	6.44 ± 0.80
<i>XMM</i> , [3.6]–[4.5] > 0.1	73 495	1.12 ± 0.24	0.68 ± 0.04	7.73 ± 1.72
<i>XMM</i> , [3.6]–[4.5] > 0.3	14 524	1.51 ± 0.77	0.68 ± 0.10	9.73 ± 5.19
All fields, all galaxies	869 587	0.26 ± 0.05	0.71 ± 0.03	4.96 ± 0.99
All fields, [3.6]–[4.5] < -0.45	50 472	10.22 ± 0.55	0.93 ± 0.03	10.52 ± 0.67
All fields, [3.6]–[4.5] < -0.3	247 746	2.07 ± 0.15	0.81 ± 0.02	6.50 ± 0.51
All fields, [3.6]–[4.5] > 0.1	216 848	1.32 ± 0.15	0.75 ± 0.03	7.83 ± 0.95
All fields, [3.6]–[4.5] > 0.3	42 493	1.72 ± 0.48	0.75 ± 0.06	9.68 ± 2.77
All mock fields, all galaxies	650 496	1.23 ± 0.08	0.91 ± 0.02	7.65 ± 0.52
All mock fields, [3.6]–[4.5] < -0.45	66 333	8.66 ± 0.35	1.01 ± 0.02	9.62 ± 0.45
All mock fields, [3.6]–[4.5] < -0.3	216 273	3.60 ± 0.15	0.95 ± 0.02	7.82 ± 0.36
All mock fields, [3.6]–[4.5] > 0.1	97 562	4.40 ± 0.32	0.99 ± 0.03	11.64 ± 0.91
All mock fields, [3.6]–[4.5] > 0.3	9614	4.32 ± 2.28	0.90 ± 0.16	13.19 ± 7.33
Brightened mocks, all galaxies	946 217	0.74 ± 0.06	0.85 ± 0.02	6.40 ± 0.57
Brightened mocks, [3.6]–[4.5] < -0.45	52 441	7.22 ± 0.44	0.98 ± 0.03	8.82 ± 0.59
Brightened mocks, [3.6]–[4.5] < -0.3	221 476	2.92 ± 0.17	0.88 ± 0.02	7.30 ± 0.46
Brightened mocks, [3.6]–[4.5] > 0.1	220 590	2.39 ± 0.16	0.93 ± 0.02	8.91 ± 0.61
Brightened mocks, [3.6]–[4.5] > 0.3	42 589	2.59 ± 0.56	0.87 ± 0.06	10.55 ± 2.39

towards the more extreme colours as compared to the observed SERVS + DeepDrill galaxies. In order to mimic the photometric uncertainty that is present in the observational data, and its consequences for subsample selection, a random offset for each mock galaxy is drawn from a normal distribution with the same width as the uncertainty in the data as derived from the source finder used by Lacy et al. (2021), ranging from (on average) less than a per cent at the bright end, to more than 5 per cent at our flux limit of $[3.6] = 22$.

2.3.2 Incorporating source blending

The mock light-cones are built from large cosmological volumes, which when projected (in a random direction) will have sources overlapping on the sky. In the [3.6] band where we estimate clustering, the point spread function (PSF) is 1.8 arcsec, so sources separated by less than that in projection on the sky will be blended. These are mostly pairs, but also triples and some quadruples. Those sources that are blended in this way are turned into single sources, with their fluxes summed for both the [3.6] and [4.5] band. This is not immediately visible in the autocorrelation function at the smallest scales, as we only estimate this function from around 5 arcsec onwards. The blending does increase the flux for about 10 per cent of the sources, although this is mostly by a relatively small amount in that many of these blend with fairly faint sources, including those

beyond our faint-end flux cuts (at $AB = 22$ for both bands). The sample size decreases with sources above the flux cut being blended, but increases with sources getting brighter through blending that then end up above the flux cut. The blending somewhat changes the subsamples derived from selection in luminosity bins. Sources that blend also change the colour of the resulting blended source, affecting any subsamples derived from colour cuts.

2.3.3 Matching number counts and completeness

Comparing the number counts for the SERVS + DeepDrill sample and the SHARK light-cones, Lacy et al. (2021) found that these agree relatively well at faint magnitudes ($AB > 22$), but not at brighter magnitudes: this may in part be driven by the lack of modelling of the AGN contribution to the galaxy emission in SHARK. At intermediate magnitudes, ~ 18 – 21 , the model light-cone predicts a lower number of galaxies: the models are somewhat too faint (~ 0.6 mag), with a small offset of 0.05 mag in colour. As this is the range of magnitudes for which we estimate the autocorrelation function, we now consider two sets of mocks: one set in which only the photometric uncertainty and source blending are mimicked, as described above, leaving the physics of the models as generated by SHARK untouched, and a second set in which we additionally match for sample size through flux brightening (as the mock counts are too low), which can be seen

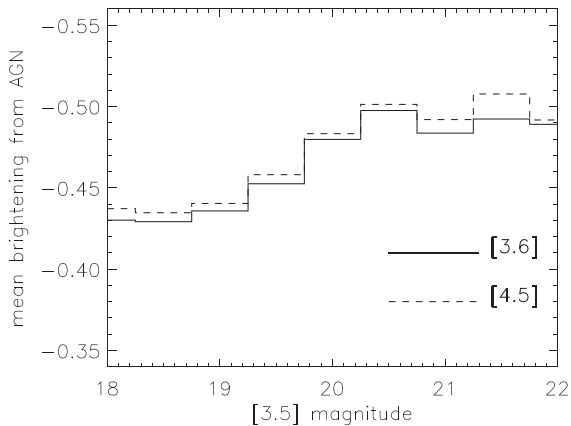


Figure 6. Mean brightening (in magnitude) as a function of [3.6] when an AGN component is included in the SED fitting for the DEVILS sample (see main text and Thorne et al. 2022 for details), for each of the two *Spitzer* bands considered in this paper. At the faint end the offset is half a magnitude, decreasing somewhat for the brighter galaxies.

as compensating for the lack of AGN modelling in SHARK, amongst others.

The contribution of AGN to the overall galaxy SEDs, including the near-infrared bands, was recently estimated for two large samples by Thorne et al. (2022). This was achieved by comparing SED fits (using ProSpect: Robotham et al. 2020) with and without an AGN component. The *Spitzer* fluxes of the galaxies derived from the global best fit SED can then be compared to those where the best-fitting AGN template has been removed, yielding a flux ratio between these for each of the [3.6] and [4.5] bands. In Fig. 6, we show these flux ratios as a magnitude difference (brightening) as a function of the [3.6] magnitude for each of these two bands, for the deepest (and largest) of the two samples studied by Thorne et al. (2022): DEVILS (Davies et al. 2018). For $[3.6] > 20$ the mean offset for each band is about half a magnitude. For the brighter galaxies this goes down a bit to a brightening of -0.43 in magnitude.

For our second set of mocks, we brighten the predicted fluxes with a single offset for each band to get sample sizes that match the observed ones for *all* samples and subsamples considered. Given the uncertainties, it is not realistic to get all simulated and observed sample sizes to match, so the aim is to get a best match for most (sub)samples. We also need to deal with completeness, as the observed sample will not be complete down to the flux limit, but the mock sample is (as it is much deeper). The completeness fractions of SERVS + DeepDrill for [3.6] and [4.5] magnitude intervals are listed in tables 8 and 9 in Lacy et al. (2021). We map these fractions to a bin size of one magnitude, which we use later on, and work out the overall completeness for our flux limits $18 < [3.6] < 22$ and $18 < [4.5] < 22$ by taking the weighted mean of these bins (weighted by number counts). These are listed in Table 2 for each [3.6] and [4.5] catalogue individually. As we select in both bands, completeness will be somewhat worse than either fraction listed (0.93 and 0.94 for [3.6] and [4.5], respectively) because of the variation in colour: a source missing in one band might not be missed in the other, and vice versa, but it could be missing in both too. This means that the completeness for our joint [3.6] and [4.5] cut will be below 0.93, but most likely not below ~ 0.90 . As the simulated light-cones are several magnitudes deeper than the observed data, any mock survey drawn from SHARK down to our shallower flux limit should be practically complete. This implies that for the nearly 870 000 SERVS + DeepDrill sources found for our flux limits $18 < [3.6]$

< 22 and $18 < [4.5] < 22$ there should be between around 935 000 and 967 000 mock galaxies from SHARK (roughly corresponding to completeness fractions of 0.93 and 0.90, respectively).

For these completeness fractions we find that brightening the mock galaxies by -0.565 mag at $3.6 \mu\text{m}$ and -0.61 at $4.5 \mu\text{m}$ gives the best matching sample sizes for the three fields: the combination of mimicking photometric accuracy and source blending, and applying these completeness corrections, the three fields we selected from the full SHARK volume comprise a total of 952 970 galaxies, corresponding to a joint [3.6] and [4.5] completeness fraction for the observational data of 0.916. Besides yielding a reasonable sample size, this brightening of the mock galaxies is consistent with the offsets found by Lacy et al. (2021), including the small colour offset they found. The latter is required to get the ratios of simulated red and blue galaxies to match the ratios for the corresponding observed subsamples. This then constitutes the second set of mocks, besides the first set of mocks for which the offsets for the two fluxes (-0.565 mag at $3.6 \mu\text{m}$ and -0.61 mag at $4.5 \mu\text{m}$) are *not* applied, and therefore has number counts in both bands that are too low, and colours that are somewhat off.

2.3.4 Resulting mock samples

After mimicking photometric accuracy and source blending, we generated two sets of three mock fields, one without any flux corrections, and one where each band is brightened by a fixed offset (see previous subsection). We then treat these mocks in exactly the same way as the observed fields. For the first set of mock fields the counts and colours do not match well to the observed ones, as discussed above, and as can be seen in Table 4 (the rows denoted ‘all mock fields’). For the second (brightened) set of mocks, applying the same colour cuts to the mock data as for the simulated data, for the three fields we retain 225 683 blue galaxies ($[3.6]-[4.5] < -0.3$), and 218 231 red ($[3.6]-[4.5] > 0.1$) ones (shown in the bottom part of Table 4). For the more extreme cuts we get 42 242 ‘very blue’ ($[3.6]-[4.5] < -0.45$) galaxies, and 46 077 ‘very red’ ($[3.6]-[4.5] > 0.3$) ones. There is some variation from the observed counts for each of the colour cuts, but this cannot be changed by a colour offset, as there are too many ‘very blue’ mock galaxies, too few blue ones, and more or less the right amount of red and ‘very red’ mock galaxies. As the choice of the colour cuts remains somewhat arbitrary, this is good enough for our purposes.

In the following we will mostly look at the brightened mock samples (the second set), but also look at estimates for the first set where fluxes (and colours) are not changed at all. Using the actual redshifts of the mock galaxies from the SHARK lightcones we derive the redshift distribution of the colour-selected (brightened) mock subsamples, as we have done for the S-COSMOS field (Fig. 2), with the same colour and flux cuts. The distributions are shown in Fig. 7. What is immediately clear is that the distributions for the blue S-COSMOS and SHARK subsamples are very similar, for all cuts, whereas the distributions for the red ones are not: the SHARK redshift distribution for the red galaxies is quite smooth, whereas the corresponding S-COSMOS one is not, and contains more high-redshift (> 2) galaxies. The main reason for this is field size: the area covered by S-COSMOS is 2 deg^2 , whereas the three SHARK mocks cover just over 20 deg^2 , by construction (to mimic the area of the SERVS + DeepDrill field), to the same depth. This much larger volume clearly smoothens the redshift distribution of the red subsamples.

To get an idea what type of galaxies are selected by the colour cuts, we obtained from the (brightened) mock samples the median values of the redshift, stellar mass, and star formation rate (‘sfr’

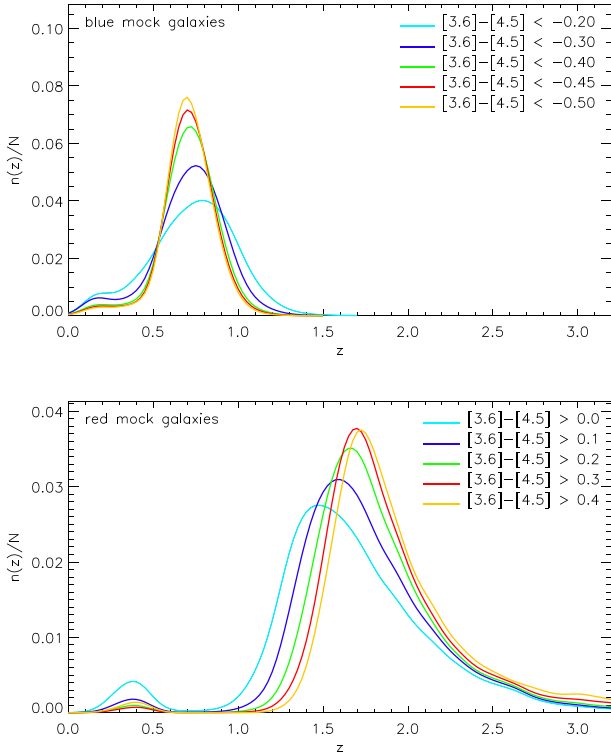


Figure 7. Redshift distribution of the SHARK mock galaxies for the same $[3.6]-[4.5]$ colour cuts as for observed galaxies in the S-COSMOS field, as shown in Fig. 2. The top panel is for blue mock galaxies, the bottom panel red ones. Please note that the area of the S-COSMOS field is 2 deg^2 , whereas the total area of the mocks is the same as for SERVS + DeepDrill, i.e. just over 20 deg^2 . This much larger area yields smoother redshift distributions for the SHARK mock galaxies as compared to the S-COSMOS ones, especially for the red subsamples.

for short) for each of the populations selected. These are listed in Table 3. As the mocks are too faint overall, the stellar mass and star formation rates, which were not changed in the step of brightening the mocks, are to be viewed as lower limits (either or both are likely to be somewhat larger in value for the observed data). However, the relative differences between the (sub)samples should remain similar for the mocks and the observed data sets, allowing us to look for trends. The colour cut correlates with both the median star formation rate and median redshift of the subsample, but as these populations vary significantly in size this is hard to interpret physically: at best the colour cut can be seen as a proxy for these median values. The sample sizes for the red and blue subsamples as well as the two more severely cut samples are similar, so the ‘red’ populations have a higher median stellar mass than the corresponding ‘blue’ ones.

3 METHODS

We combine an estimate for the angular clustering of our galaxies with a redshift distribution obtained for the S-COSMOS sample (Sanders et al. 2007) using the same selection criteria used for our samples. This provides, through inversion of the Limber equation, a measure for spatial clustering. We perform this for various $[3.6]-[4.5]$ colour cuts in each of our three fields, and for the three fields combined. We also estimate clustering for the full sample, which is less informative as the redshift distribution is much broader than for the colour-selected samples, but this is a useful measure for reference

to previous work and to estimates for the various subsamples we consider.

3.1 Estimating correlation functions

We employ the standard estimator (Landy & Szalay 1993) for measuring angular correlations $w(\theta)$: $w_{LS} = (DD - 2DR + RR)/RR$, where DD , DR , and RR are the (normalized) galaxy–galaxy, galaxy–random, and random–random pair counts at separation θ . The source catalogue extracted from the observed maps provide the galaxy positions, and we employ a ten times more abundant random catalogue that Poisson samples the same survey region. For the estimate of $w(\theta)$ and its errors we use the Jackknife technique (e.g. Wall & Jenkins 2003; Norberg et al. 2009), employing 16 pie slices of the circular fields and estimating errors from the Jackknife sampling variance.

The estimator is fitted by its expected value

$$1 + \langle w_{LS} \rangle = [1 + w(\theta)] / (1 + w_{\Omega}), \quad (1)$$

where the ‘integral constraint’ w_{Ω} is the integral of the model function for the two-point correlation function over the survey area.

We consider a two-parameter fit for the generic power-law $w(\theta) = (\theta/A)^{-\delta}$. Our samples and subsamples are large enough to not have to restrict ourselves to a one-parameter fixed-slope power-law function. The fitting technique is the one used by van Kampen et al. (2005), but adapted for the survey geometry of DeepDrill and multiple fields. We employ non-linear χ^2 -fitting using the Levenberg–Marquardt method (Press et al. 1988), which allows us to easily take into account the multiplicative integral constraint, but also produces the covariance matrix of the fitted parameters which provides a good estimate for their uncertainties. The integral constraint is not a free fitting parameter, but a function of the clustering amplitude A and power-law slope δ , and fitted that way. For each fit the χ^2 probability Q is calculated using the incomplete gamma function, and any fits with $Q < 0.1$ are discarded. With the large number of sources available here, this does not actually happen even for our smallest subsample.

3.2 Spatial clustering

We lack spectroscopic redshifts for most (around 98 per cent) of our sources, so we cannot measure spatial clustering directly. Also, we only have photometric redshifts for around half the sources (the original SERVS fields), with inhomogeneous coverage and depth (Pforr et al. 2019). However, we can employ a similar sample from the literature for which sufficient numbers of photometric (and spectroscopic) redshifts, with homogeneous in coverage and depth, are available and to which we can apply the same magnitude and colour cuts as for our own samples. The S-COSMOS data provides such a sample (Ilbert et al. 2009), even though the total area of 2 deg^2 is 10 times smaller than that of our SERVS + DeepDrill sample, resulting in redshift distributions that are not very smooth (which they would be for 10 times the S-COSMOS area).

Assuming such redshift distributions are representative for our own data set, Limber’s equation Limber (1953) provides an estimate for the spatial clustering length r_0 from the angular clustering function (e.g. Phillips et al. 1978). For the Limber equation inversion, we employ the code used by Farrah et al. (2006), using the (Gaussian smoothed with width $\sigma_v = 0.1$) measured redshift distribution from S-COSMOS, corresponding to the same selection (cuts in magnitude and colour) as for our own (sub)samples.

The redshift distribution for the full S-COSMOS sample down to our flux limits of $18 < [3.6] < 22$ and $18 < [4.5] < 22$ is shown in Fig. 1, whereas the corresponding redshift distributions for the colour cuts are shown in Fig. 2, where we only consider two colour cuts in red and blue each.

4 RESULTS

We now employ the methods described above to yield clustering estimates down to our selected flux limit and for various redshift distributions selected through the $[3.6]$ – $[4.5]$ colour cuts. Because the three fields have the same depth, with no systematic differences between the fields, we also combine the estimates for each of the fields to a single estimate for the survey as a whole. We do not use the full coverage of SERVS + DeepDrill: we leave out areas near the border of each field and around bright stars (see Section 2.1), resulting in a clean set of three round fields, totalling just over 20 sq. degrees. Corresponding mock fields of the same field size, depth, and geometry, including bright star removal, are analysed in the same fashion as the observed fields.

4.1 Estimates for all galaxies

For the three SERVS + DeepDrill fields, which each measure 3 degrees in diameter, we first estimate the angular correlation function for all galaxies in the magnitude intervals $18 < [3.6] < 22$ and $18 < [4.5] < 22$. With no colour selection, we probe the full depth of the sample from the local Universe out to $z \approx 2$ – 3 , as can be seen in Fig. 1. This shows the redshift distribution for the full S-COSMOS sample (with the same magnitude selection as for the observed samples). So we are clearly not estimating the local autocorrelation function, but the angular correlation over a wide redshift range. This results in relatively low amplitudes for the estimated angular correlation functions, as can be seen in Fig. 8, with fairly large uncertainties (derived from the Jackknife technique – see Section 3.1). The uncertainty is noticeably smaller for the joint estimate for the three fields taken together. Power-law functions fit well to each of the estimates (the reduced χ^2 values for the fits are within the range expected for the number of bins used), with fitting parameters as listed in Table 4. Despite the low amplitudes for the angular correlation functions, the well-defined redshift distribution allows us to estimate the spatial correlation length r_0 fairly well, especially for the estimate for all three fields together: $r_0 = 4.96 \pm 0.99 h^{-1} \text{Mpc}$, i.e. an uncertainty of around 20 per cent. This estimate mostly serves as a reference for the other estimates which are derived below, for more specific subsamples which sample a more restricted redshift range.

Also using *Spitzer* data, Waddington et al. (2007) employed the SWIRE survey Lonsdale et al. (2003) to estimate clustering at $3.6 \mu\text{m}$ as well, with a survey size of 8 deg^2 . This survey is not as wide and as deep SERVS + DeepDrill, and less complete near our $[3.6] = 22$ limit. Our magnitude limit corresponds to $5.6 \mu\text{Jy}$, which is close to the one for a sample down to $6.3 \mu\text{Jy}$ that they consider. For this sample they find a mean redshift of 0.9, which is very close to our mean of 0.89, and a spatial clustering length $r_0 = 3.18 \pm 0.94 h^{-1} \text{Mpc}$. This is smaller than what we find, but within the uncertainties still, and therefore consistent. The reason for their smaller clustering strength could be related to their larger incompleteness near the faint magnitude limit.

4.2 Estimates for colour selected subsamples

Colour cuts allow us to measure the angular correlation function for narrower redshift distributions, yielding larger amplitudes (as there is

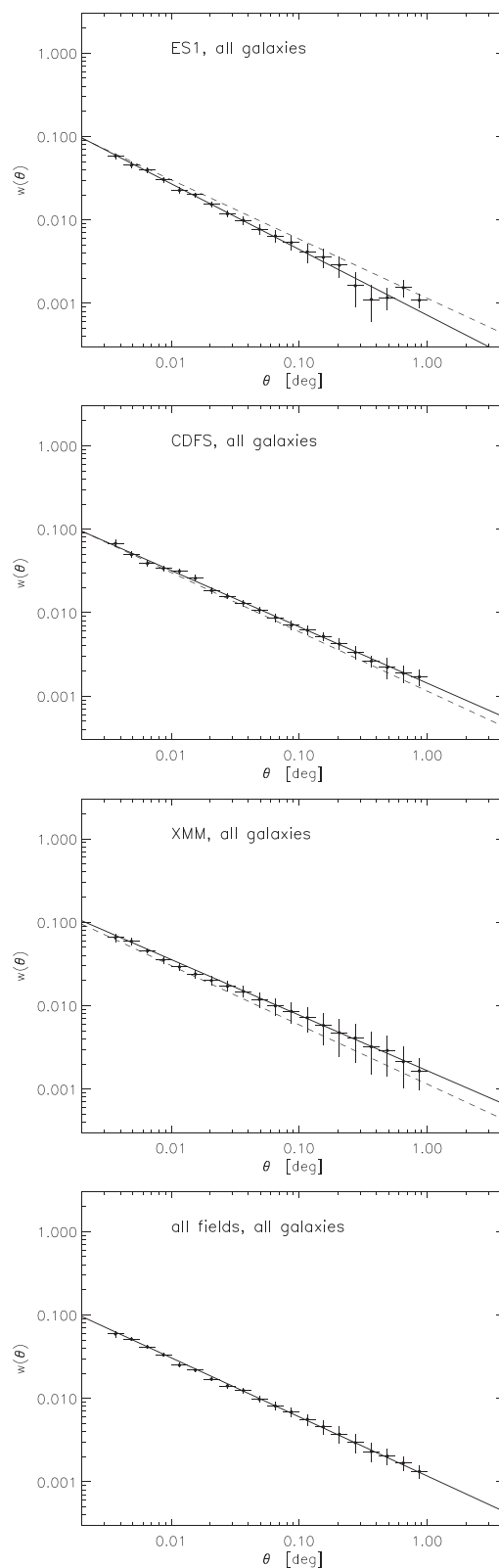


Figure 8. Angular clustering for each of the fields separately and for all fields together (bottom panel), for all galaxies below our flux cut $18 < [3.6] < 22$ and $18 < [4.5] < 22$. Solid lines indicate the power-law functions fitted to these data. To aid visual comparison, the dashed line in the top three panels corresponds to the solid line in the bottom panel (for all three fields).

less averaging out along the line of sight), and cleaner measurements. The colour cuts also select for more specific populations of galaxies, in terms of type, star formation rate, and other properties. For each of the three fields, and for each of the four colour cuts, we estimate the galaxy correlation function and fitted a two-parameter power-law function, as for the full sample. Again we use the S-COSMOS sample to derive the redshift distribution correspond to a given colour cut, which is then used in the Limber equation inversion technique (see Section 3.2).

Looking at the moderate colour cuts first, we select a blue and red subsample cutting at $[3.6]-[4.5] < -0.3$ and $[3.6]-[4.5] > 0.1$, respectively. Their redshift distributions are much narrower than for the total sample, and should provide stronger clustering estimates which are also more informative.

We select two more redshift distributions through more severe colour cuts, yielding much smaller but also much bluer and redder subsamples. As described and argued in Section 2.2, we cut at $[3.6]-[4.5] < -0.45$ and $[3.6]-[4.5] > 0.3$ to achieve this. The same selection for the S-COSMOS sample gives us to the two corresponding redshift distributions. The ‘very blue’ cut produces the narrowest redshift distribution of all four cuts, with a clear peak around $z \approx 0.7$, which has less than half the width of the blue one. In comparison, the ‘very red’ cut mostly moves the red redshift distribution to higher redshifts, and produces a less pronounced peak and a slightly broader distribution as well.

For each of the four subsamples we show our estimates for the galaxy correlation function in Fig. 9 for the blue and red subsamples, and in Fig. 10 for the more extreme cuts (where we use less bins for the fitting). For each cut, the top three panels in each figure (left- and right-hand side) show the galaxy correlation functions obtained for each SERVS + DeepDrill field individually, whereas the bottom panels show the combine estimate using all three fields for each subsample. The latter clearly produce the most accurate estimates (smallest error bars), and also the cleanest fits. The variance between the fields can also be seen: the ‘very blue’ subsamples for each field probably show the most consistency. In all panels, the dotted line represents the measured galaxy correlation function for the full sample, as already shown in the bottom panel of Fig. 8 (which has amplitude $A = 0.18 \pm 0.04$ arcsec and slope $\delta = 0.67 \pm 0.03$), but repeated here for reference.

A power-law shape is observed for each correlation function, with small deviations for some (the largest of these are for the ES1 field). A power-law is robustly fitted in each case, using the methods described in Section 3.1, with reduced χ^2 values for the fits within the expected range for the number of bins used for the fit. The resulting fitted parameters are listed in Table 4. Fractional uncertainties for these are smaller than for the full sample, as can be expected. However, the differences are larger for the fitted amplitudes than for the fitted slopes. The reason for this is the uncertainty in the integral constraint for the individual fields, which is still significant because of the limited size of the fields. Here, again the advantage of using three independent fields is apparent, as this alleviates this uncertainty somewhat.

The uncertainties in the estimates for the clustering strengths using the three fields combined are also smaller than for the individual fields. It is worth checking whether the spread in r_0 found for the individual fields is consistent with the uncertainty in r_0 for the combined field: inspecting the obtained angular galaxy clustering functions shown in Fig. 8, 9, and 10 and the corresponding fitting parameters and derived r_0 in Table 4, this is obvious for the full sample, but also for the blue and red samples. For the ‘very red’ sample, the variance over the three individual fields is $1.25 h^{-1}$ Mpc,

which is still less than the formal error on r_0 for the combined fields. However, for the ‘very blue’, the same variance in r_0 is $1.3 h^{-1}$ Mpc, which is twice the formal error on r_0 for the combined estimate. The reason for this is likely that even though the ‘very blue’ population has the most restricted redshift range, this is also a fixed redshift range around $z \approx 0.7$, and therefore sensitive to specific structures like galaxy clusters, filaments and voids in each of the fields, which are not likely to be synchronous over all fields: galaxy clusters, for example, have a range of formation and evolution histories. This can only be alleviated by a survey over an even larger area. However, for our other subsamples we find that we do not suffer from cosmic variance, and our estimates for the clustering strength r_0 can be considered robust.

Waddington et al. (2007) find the same increasing trend of spatial clustering length with luminosity, going up to values around $r_0 = 6 h^{-1}$ Mpc for their brightest flux cuts (see their table 1), which are again somewhat below what we find for our brightest magnitude bins (see Section 4.4 and Table 5), noting that they considered flux cuts, not flux bins.

4.3 Clustering predictions for the SHARK models

The clustering analysis for the two sets of SHARK mock samples and subsamples was performed in exactly the same way as for the SERVS and DeepDrill data, with similar magnitude and colour cuts, to allow for a meaningful comparison. We therefore selected three random fields from the full SHARK volume that are not adjacent to each other (as is the case for the observational data), with the same area as the three observed data sets from SERVS + DeepDrill. Star masks were applied too, employing the same three sets of masks that were used for the observations, in order to mimic any (small) effect these might have.

For both sets of mocks, the resulting estimates for the angular correlation functions for all galaxies down to our flux limit are shown in Fig. 11. The left-hand side panel shows the estimates for the first set of mocks which only mimic photometric accuracy and source blending, but no changes to the underlying galaxy formation model, whereas the right-hand side panel shows the estimates for the brightened mock samples of set two (see Section 2.3.3). The variation between the three mock fields is similar to that between the three observed fields. The power-law fit for all observed galaxies in all fields (as shown before, in the bottom panel of Fig. 8) is plotted for reference as a single dashed line.

The angular clustering estimates for the same four colour cuts that we employed for the observational data are shown in Fig. 12 for the first set (without brightening) on the left-hand side and for the second set (with brightening) on the right-hand side. In these figures only the joint estimates for the three mock fields are shown, and the dashed lines denote the corresponding observational estimates for the SERVS + DeepDrill subsamples derived from the same colour cut. What is apparent is that the slope is somewhat steeper for all mock (sub)samples. However, the difference in slope between models and data is more pronounced for the first set of mocks (no brightening), especially for the red subsamples. The flux brightening (the second set of mocks) does bring the mock samples towards a better match to the estimates for the corresponding observed samples.

All fitted parameters for the models are listed in the bottom part of Table 4. The spatial clustering length is comparable but mostly somewhat larger than for the observed samples, except for the ‘very blue’ brightened mock subsample. This is also shown graphically in Fig. 15. Overall the clustering lengths for the brightened mock samples are not significantly different from the observed ones:

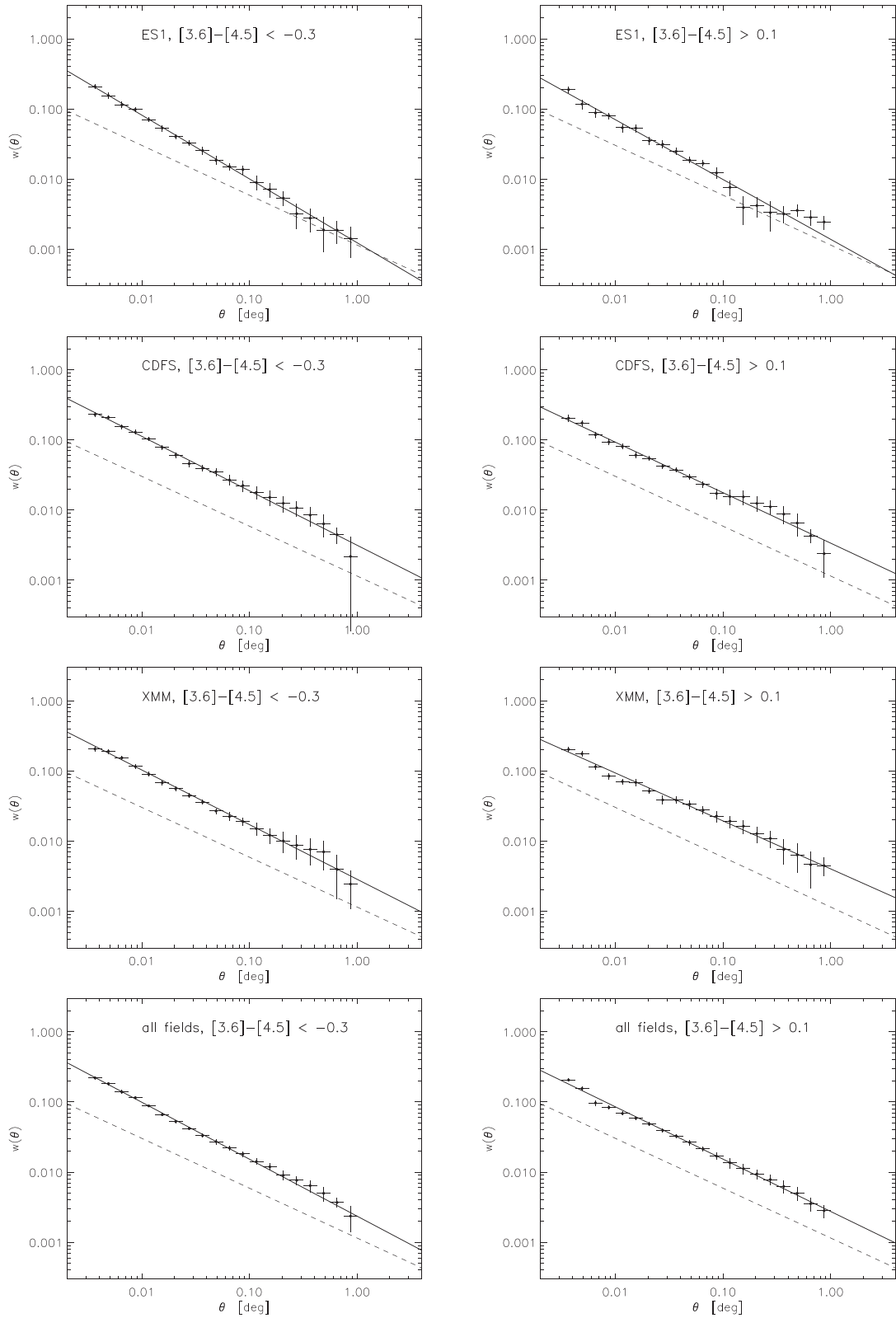


Figure 9. Angular clustering for ES1, CDFS, *XMM*, and for all fields together (bottom panels), for blue (left-hand panels) and red galaxies (right-hand panels), i.e. $[3.6]-[4.5] > 0.1$ and $[3.6]-[4.5] < -0.3$, respectively. Solid lines indicate the power-law functions fitted to these data. To aid visual comparison, the dashed line in each panel corresponds to the solid line in the bottom panel of Fig. 8, which is for all galaxies in all fields.

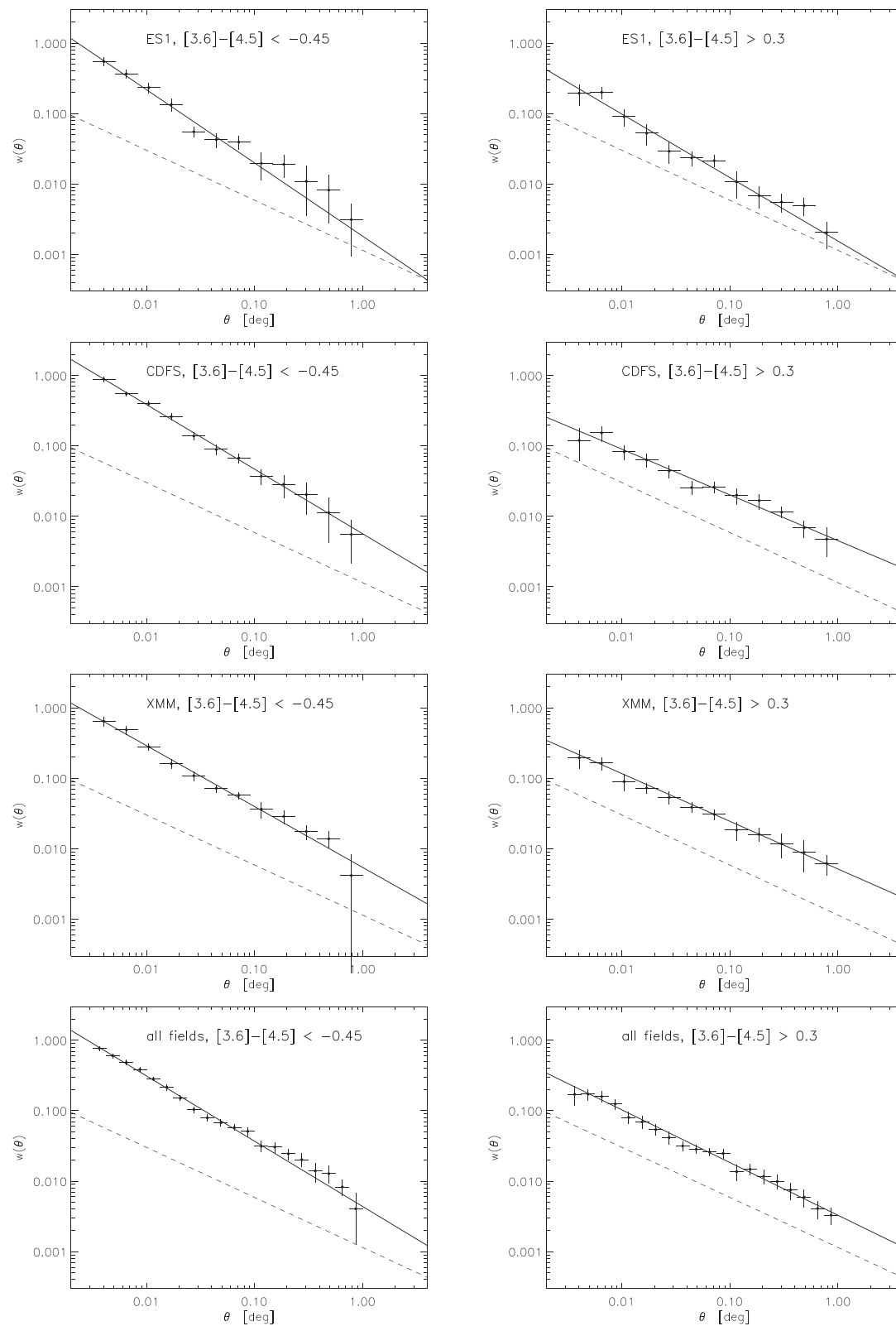


Figure 10. Angular clustering for ES1, CDFS, *XMM*, and for all fields together (bottom panels), for the ‘very blue’ and ‘very red’ galaxy populations, i.e. $[3.6]-[4.5] > 0.3$ and $[3.6]-[4.5] < -0.45$, respectively. Solid lines indicate the power-law functions fitted to these data. To aid visual comparison, the dashed line in each panels corresponds to the solid line in the bottom panel of Fig. 8 (which is for all galaxies in all fields).

Table 5. Clustering measures for subsamples in luminosity bins, observational as well as simulated (see main text for details) for the two-parameter fits. This table also lists the actual sizes of these subsamples, as well as the derived spatial clustering strength using Limber equation inversion technique.

Colour and sample selection	Luminosity bin	N	A (arcsec)	δ	r_0 (h^{-1} Mpc)
All, observed	22 > [3.6] > 18	869 587	0.26 ± 0.05	0.71 ± 0.03	4.96 ± 0.99
All, observed	22 > [3.6] > 21	368 889	0.17 ± 0.05	0.70 ± 0.04	4.50 ± 1.49
All, observed	21 > [3.6] > 20	303 735	0.32 ± 0.08	0.70 ± 0.04	5.17 ± 1.35
All, observed	20 > [3.6] > 19	153 987	1.29 ± 0.19	0.80 ± 0.03	7.20 ± 1.12
All, observed	19 > [3.6] > 18	42 976	3.49 ± 0.56	0.94 ± 0.05	7.76 ± 1.31
All, brightened mocks	22 > [3.6] > 18	946 217	0.74 ± 0.06	0.85 ± 0.02	6.40 ± 0.57
All, brightened mocks	22 > [3.6] > 21	421 926	0.32 ± 0.07	0.76 ± 0.03	5.47 ± 1.22
All, brightened mocks	21 > [3.6] > 20	317 094	0.73 ± 0.09	0.81 ± 0.03	6.32 ± 0.84
All, brightened mocks	20 > [3.6] > 19	155 286	2.10 ± 0.20	0.94 ± 0.03	7.97 ± 0.79
All, brightened mocks	19 > [3.6] > 18	51 911	5.39 ± 0.46	1.01 ± 0.04	9.22 ± 0.85
Blue, observed	22 > [3.6] > 18	247 746	2.07 ± 0.15	0.81 ± 0.02	6.50 ± 0.51
Blue, observed	22 > [3.6] > 21	62 096	1.13 ± 0.31	0.77 ± 0.05	5.52 ± 1.56
Blue, observed	21 > [3.6] > 20	94 439	1.70 ± 0.25	0.76 ± 0.03	5.93 ± 0.92
Blue, observed	20 > [3.6] > 19	68 051	3.24 ± 0.37	0.80 ± 0.03	7.53 ± 0.92
Blue, observed	19 > [3.6] > 18	23 160	3.33 ± 0.98	0.99 ± 0.09	7.21 ± 2.23
Blue, brightened mocks	22 > [3.6] > 18	221 476	2.92 ± 0.17	0.88 ± 0.02	7.30 ± 0.46
Blue, brightened mocks	22 > [3.6] > 21	55 389	0.65 ± 0.24	0.70 ± 0.05	4.65 ± 1.75
Blue, brightened mocks	21 > [3.6] > 20	85 838	1.98 ± 0.28	0.81 ± 0.03	6.16 ± 0.91
Blue, brightened mocks	20 > [3.6] > 19	59 967	4.60 ± 0.38	0.93 ± 0.03	8.42 ± 0.74
Blue, brightened mocks	19 > [3.6] > 18	20 282	12.87 ± 0.84	1.14 ± 0.05	13.33 ± 1.02
Red, observed	22 > [3.6] > 18	216 848	1.32 ± 0.15	0.75 ± 0.03	7.83 ± 0.95
Red, observed	22 > [3.6] > 21	136 986	0.87 ± 0.17	0.72 ± 0.03	7.10 ± 1.38
Red, observed	21 > [3.6] > 20	66 538	3.32 ± 0.40	0.84 ± 0.04	9.88 ± 1.25
Red, observed	20 > [3.6] > 19	11 970	6.76 ± 1.59	0.88 ± 0.08	13.55 ± 3.45
Red, brightened mocks	22 > [3.6] > 18	220 590	2.39 ± 0.16	0.93 ± 0.02	8.91 ± 0.61
Red, brightened mocks	22 > [3.6] > 21	146 756	1.25 ± 0.18	0.86 ± 0.03	7.32 ± 1.09
Red, brightened mocks	21 > [3.6] > 20	60 181	4.42 ± 0.41	1.00 ± 0.04	10.54 ± 1.05
Red, brightened mocks	20 > [3.6] > 19	12 189	14.01 ± 1.41	1.08 ± 0.06	18.46 ± 2.15

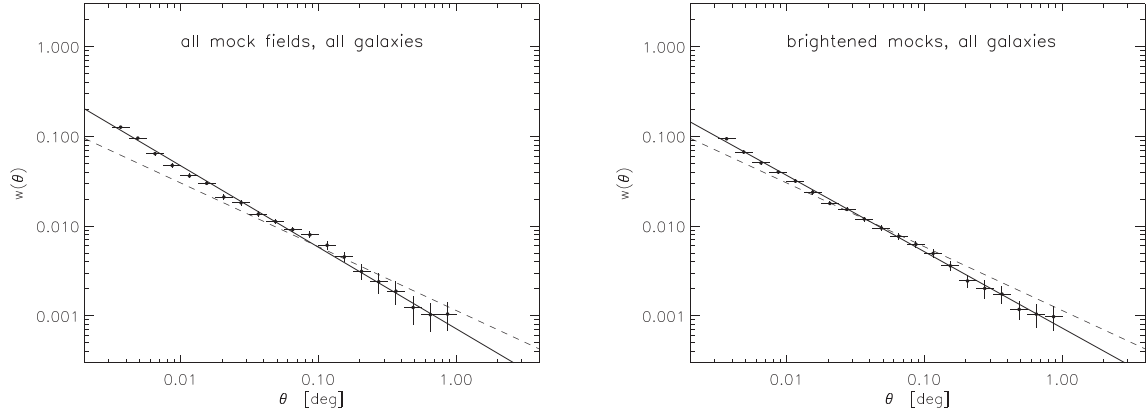


Figure 11. Angular clustering for the two sets of SHARK samples, for the three mock fields combined, for ‘all’ mock galaxies (that is, for $18 < [3.6]_{AB} < 22$ and $18 < [4.5]_{AB} < 22$). The left-hand panel is for the mocks without flux corrections to match sample sizes, the right-hand panel is for the mocks that were corrected, where the two bands have been brightened by somewhat different amounts (see main text). Both sets of mocks do mimic photometric accuracy and source blending: the only difference between the two sets are the flux offsets. The dashed line denotes the power-law fit to the observed angular correlation function for the same flux cut. The simulated galaxies are more strongly clustered, with a steeper slope than the observed population, most significantly for the mocks that have not been corrected (brightened) to match sample size.

only for the ‘very blue’ subsample we find that the difference between the spatial clustering lengths of the brightened mock and observed subsamples is larger than the formal uncertainty of each, but not by much. The clustering lengths for the first set of mock

samples (with no brightening) are typically too large, except for the ‘very blue’ subsample, noting that both the counts and the colours for the first set of mocks do not match the observed ones well.

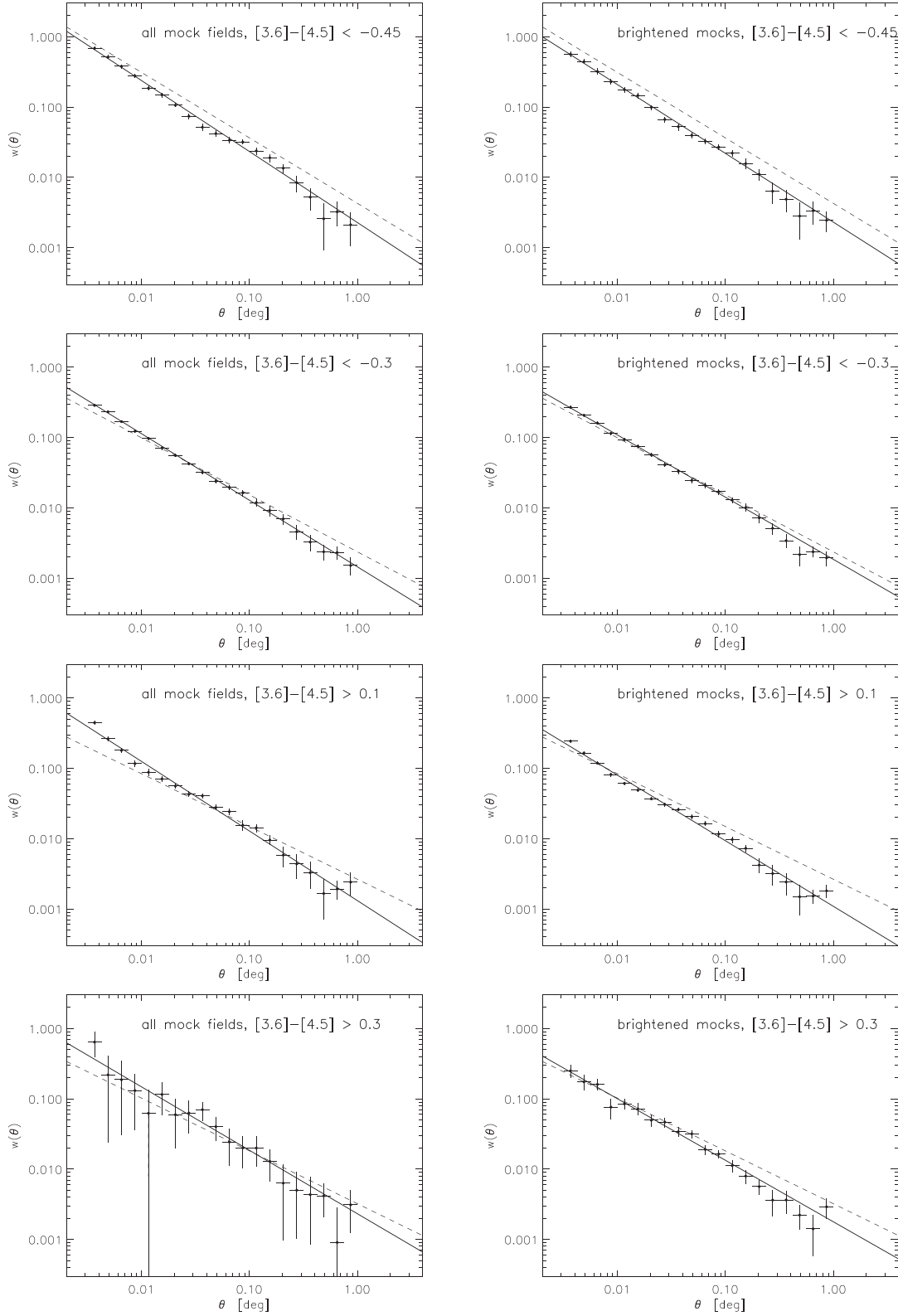


Figure 12. The left-hand side panels show the angular clustering estimates for the first set of SHARK mock samples with photometric accuracy and source blending (1.8 arcsec radius) taken into account, for three mock fields combined. Estimates are presented for the same colour cuts used for the observed samples. The dashed line in each panel shows the power-law fit to the corresponding observed angular correlation function with the same colour cut. As for the full sample shown in Fig. 11, the simulated galaxy population is more strongly clustered than the observed one, with a steeper slope, except for the ‘very blue’ population. The right-hand side panels are for the second set of mocks, with brightened fluxes for each of the two bands (a fixed offset) to match the number counts and completeness of the corresponding observed subsamples. The simulated galaxy population is still more strongly clustered than the observed one (except for the ‘very blue’ subsample), but the clustering amplitudes and slopes are smaller than for the first set of mocks. The spatial clustering lengths derived for these estimated correlation functions are smaller too and closer to the observed values, as seen in Table 4.

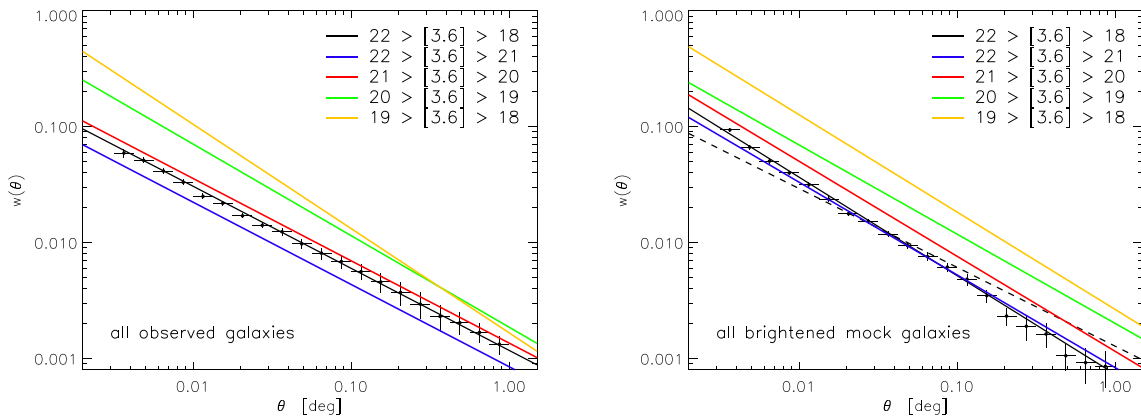


Figure 13. Left-hand side: angular clustering in magnitude bins for the SERVS + DeepDrill sample, for the three fields combined, without colour selection. Right-hand side: angular clustering in magnitude bins for the flux-corrected (brightened) SHARK models, for the three mock fields combined, without colour selection. The dashed line denotes the power-law fit to the observed angular correlation function for the same flux cut (the black solid line in the top panel), to aid visual comparison.

4.4 Estimates for magnitude selected subsamples

In the following analysis, we restrict ourselves to the second set of (brightened) mock samples, as the first set does not match the observations well. For the full sample as well as for the red and blue samples, for both SERVS + DeepDrill and SHARK, there are sufficient numbers of sources to also allow a selection in bins of IRAC [3.6] magnitude. We use a bin size of one magnitude, starting at our faint flux limit of $[3.6] = 22$, and up to where the number of sources becomes too small for a robust clustering estimate, which is the bin $19 > [3.6] > 18$ for the full sample and the blue subsample, and the bin $20 > [3.6] > 19$ for the red subsample. The $19 > [3.6] > 18$ bin for the red subsample only contains of order 1500 galaxies (of the order of 500 per field), for which we would need to use a different estimate for clustering, such as the sky-averaged autocorrelation function (e.g. van Kampen et al. 2005), which makes comparison to the other subsamples much harder. We do not select in the $[4.5] \mu\text{m}$ band: just within the range $18 < [4.5] < 22$, as for the full sample.

The results for the observed and simulated (sub)samples are presented in Table 5, and in more visual form in Figs 13 and 14, where the data points and the black solid line are the same as for the corresponding panels at the bottom of Figs 8 and 9, i.e. for the full luminosity range in both bands. The dashed line in the right-hand side panel is just the fit to the observed data for all galaxies, in order to help visual comparison.

Looking at the fits to the clustering estimates for full sample divided in magnitude bins (Fig. 13), we see a similar trend for both observed and mock galaxies: the estimate for the faintest bin ($22 > [3.6] > 21$, blue line) shows a clustering amplitude that is lower than the one for all galaxies (data points and black line). For each consecutive brighter magnitude bin the angular clustering amplitude and the spatial clustering length get larger. For the brightest magnitude bin ($19 > [3.6] > 18$, orange line) the observed subsample has a fitting function that is steeper than the mock one, but that is the only clear difference between the data and the simulations.

For the red and blue subsamples we see something similar (Fig. 14): in all four panels the estimate for the faintest bin ($22 > [3.6] > 21$, blue line) is below the estimate for the red/blue subsample over the whole magnitude range (data points and black line). Also, again for consecutive brighter magnitude bins the angular clustering amplitude and the spatial clustering length get larger. For the red

observed and red mock subsamples the steepening of the correlation function already starts with the $20 > [3.6] > 19$ bin (green line), but for the blue subsamples again at $19 > [3.6] > 18$ (orange line), this time for both the observed and mock data, with a clear difference in amplitude.

Overall the differences in clustering estimates between observed and (brightened) mock samples are surprisingly small, except for the smallest sample sizes for the brightest magnitude bins. As the uncertainties are larger for these, shallower surveys over a larger area would be required before drawing definite conclusions from such differences. However, it is worth stating that it is also in this magnitude range that the number counts of SHARK differ most from the observations before magnitude corrections were made to the mocks, most likely due to the model ignoring the AGN contribution to the galaxy SED, as shown in Fig. 6.

5 DISCUSSION AND CONCLUSIONS

We have measured the clustering strength for a large sample of near-infrared galaxies in all fields of SERVS + DeepDrill, with uncertainties, and for four colour-selected subsamples as well. The full observational data set is made up of three round fields, which combine to a total survey area of just over 20 deg^2 . Angular clustering measures were obtained down to a flux limit as well as for subsamples selected on the $[3.6]$ – $[4.5]$ colour, which effectively selects for redshift. In order to obtain a spatial clustering estimate from the angular correlation function we used the redshift distribution from S-COSMOS, with the same selection criteria as used for our samples, and the Limber equation inversion technique.

We find that our angular correlation functions are robustly estimated for all (sub)samples, which is not surprising given our sample size, but we also find that these are well fitted by a power-law function over most of its range. We find a correlation length $r_0 = 4.96 \pm 0.99 h^{-1} \text{ Mpc}$ for the full sample down to an AB magnitude limit of 22 in both bands ($[3.6]$ and $[4.5]$), which is consistent with what has traditionally been found in the optical as well for large samples with no redshift selection, and in various other wavebands, such as $24 \mu\text{m}$: Gilli et al. (2007) found a correlation length $r_0 = 4.0 \pm 0.4 h^{-1} \text{ Mpc}$ for all galaxies with $f_{24} > 20 \mu\text{Jy}$ (no redshift selection). In the following, we also compare some of our results to other measurements found in the literature, mostly those listed in Table 6.

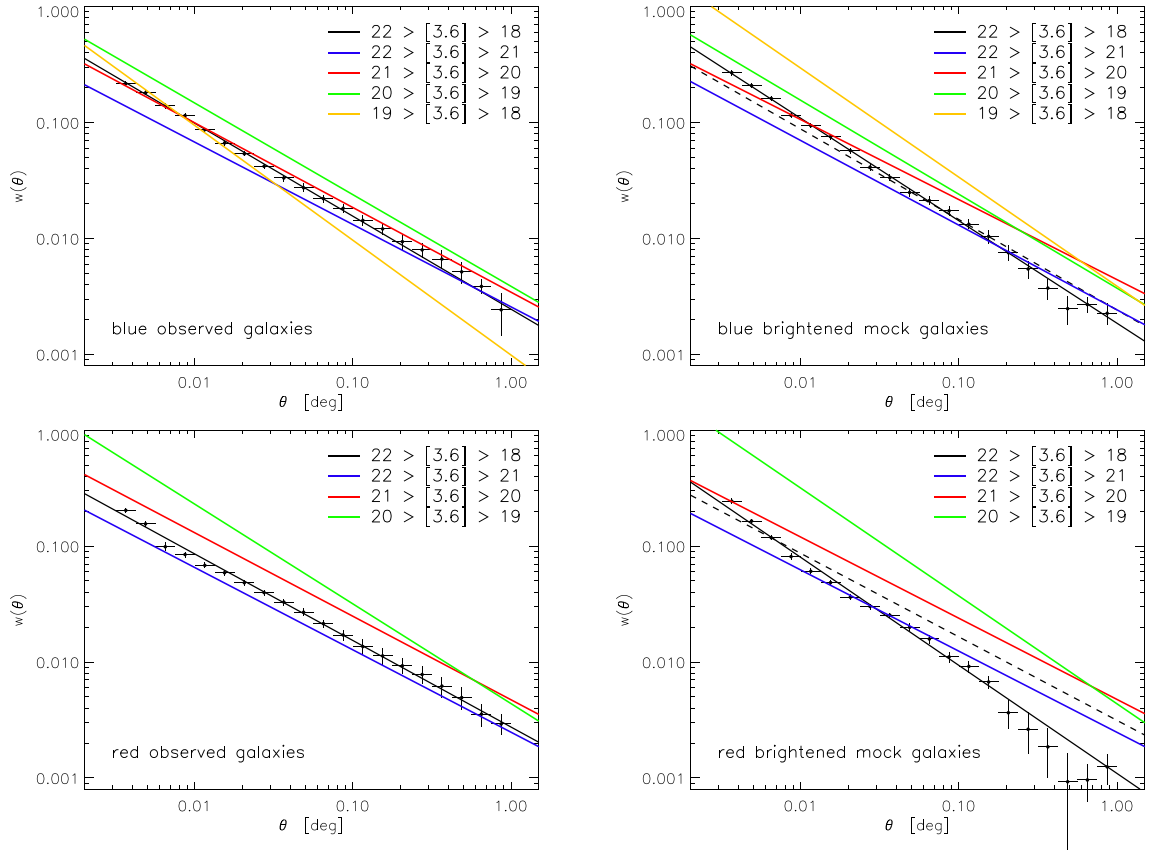


Figure 14. Same as Fig. 13, but for the blue and red subsamples. There are not enough sources in the brightest bin for the red subsample to robustly estimate and fit an autocorrelation function (see main text).

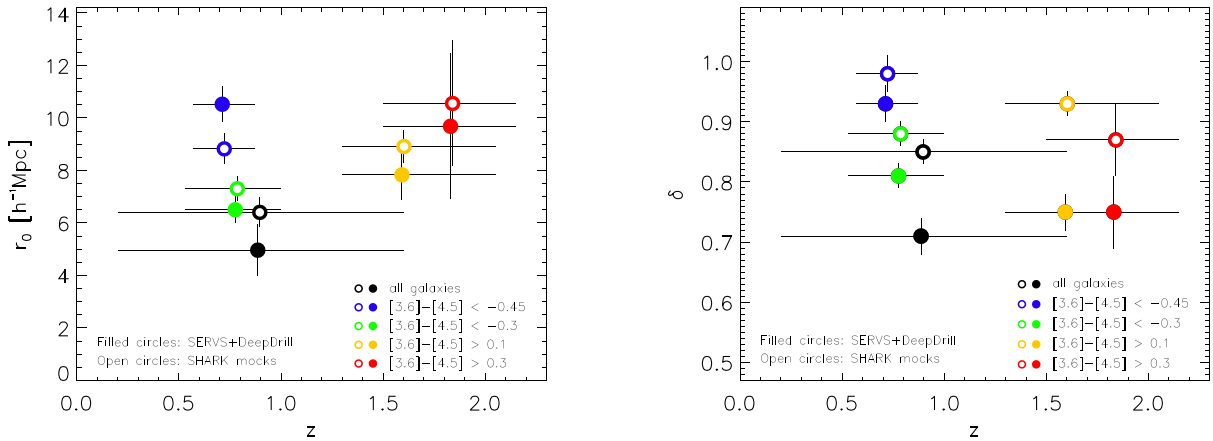


Figure 15. Spatial clustering length and (fitted) clustering slope as a function of redshift (through colour cuts), for the SERVS + DeepDrill datasets (filled circles), and the brightened SHARK mocks (open circles). The vertical bars indicate the uncertainty in the clustering estimate, but the horizontal bar indicates the redshift range covered by the colour cut, as derived from the S-COSMOS sample (see Figs 1 and 2).

de la Torre et al. (2007) studied two samples based on data for the VVDS-SWIRE field: one for the full field (0.82 deg^2) with photometric redshifts, and one for a sub-area (0.42 deg^2) for which spectroscopic redshifts are available (Le Fèvre et al. 2005). Using these redshifts, they divided both samples in redshift bins and estimated the spatial clustering strength for each bin. Averaging over the full redshift range, they find a mean value of $r_0 = 3.9 \pm 0.5 h^{-1} \text{ Mpc}$ for the galaxies selected down to $[3.6] < 21.5$ and a mean value of $r_0 = 4.4 \pm 0.5 h^{-1} \text{ Mpc}$

for the ones selected $[4.5] < 21$. Even though they did not go quite as deep, these values are again consistent with what we found.

In order to study more specific populations or redshift ranges, we use subsamples selected by colour, as we do not have sufficient numbers of photometric or spectroscopic redshift, and those that we have Pforr et al. (2019) are not homogeneous over the whole survey. The colour selection works best for blue colours, in the sense that this yields the most restricted redshift distributions, with a clear

Table 6. Compilation of clustering measurements from the literature, as discussed in the main text. The top part of the table covers magnitude- or flux-selected samples, the bottom part lists clustering strengths for samples also selected in redshift.

Reference	Sample	Selection	r_0 (h^{-1} Mpc)
Waddington et al. (2007)	SWIRE	$S_{3.6} < 6.3 \mu\text{Jy}$	3.18 ± 0.94
Gilli et al. (2007)	GOODS	$f_{24 \mu\text{m}} > 20 \mu\text{Jy}$	4.0 ± 0.4
de la Torre et al. (2007)	VVDS-SWIRE	$[3.6] < 21.5$ $[4.5] < 21$	3.9 ± 0.5 4.4 ± 0.5
Hatfield et al. (2016)	VIDEO, $K_s < 23.5$	$0.75 < z < 0.9$ $1.25 < z < 1.7$	$5.68^{+0.13}_{-0.076}$ $6.47^{+0.11}_{-0.14}$
McCracken et al. (2015)	UltraVISTA-DR1	$0.8 < z < 1.1$	3.5 ± 0.4
Coil et al. (2008)	DEEP2, optical, red subsample	$0.7 < z < 0.925$	5.25 ± 0.26
Amvrosiadis et al. (2019)	H-ATLAS, $S_{250 \mu\text{m}} > 33 \mu\text{Jy}$	$1 < z < 2$	8.1 ± 0.5
Starikova et al. (2012)	SWIRE, $S_{24 \mu\text{m}} < 310 \mu\text{Jy}$	$< z > = 0.7$ $< z > = 1.7$	4.98 ± 0.28 8.04 ± 0.69

peak at $z \approx 0.7$ for the most extreme cut ($[3.6]-[4.5] < -0.45$). The most extreme red cut ($[3.6]-[4.5] > 0.3$) shows a peak at $z \approx 1.5$, but with a tail towards higher redshifts. Not surprisingly, the ‘very blue’ subsample (which has the narrowest redshift distribution) shows the strongest clustering, with a clustering strength r_0 of $10.52 \pm 0.67 h^{-1}$ Mpc. The other three cuts, yield our blue, red, and ‘very red’ subsamples, with a broader redshift distribution. For the ‘very red’ sample, corresponding to our highest redshift distribution (with a median redshift of 1.89), we find $9.68 \pm 2.77 h^{-1}$ Mpc. The less severe colour cuts (our red and blue samples) contain more galaxies, thus providing more accurate estimates. We found $r_0 = 7.83 \pm 0.95 h^{-1}$ Mpc for the red ($[3.6]-[4.5] > 0.1$) cut (median redshift of 1.77), and $r_0 = 6.50 \pm 0.51 h^{-1}$ Mpc for the blue ($[3.6]-[4.5] < -0.3$) cut (median redshift of 0.72).

An interesting trend is that the clustering strength increases somewhat with redshift for the blue, red, and ‘very red’ subsamples (which have median redshifts of 0.72, 1.77, and 1.89, respectively), although this remains hard to interpret as the colour selections are not selecting galaxy populations that are exactly comparable, either differing in redshift (red versus blue), or subsample size (red versus ‘very red’). Furusawa et al. (2011) looked at clustering as a function of stellar mass in the SXDS/UDS field, which measures 0.83 deg^2 and includes galaxies down to $K_{AB} = 23.5$, with photometric redshifts. As their subsample sizes are relatively small, they fit a power law with a constant slope of $\delta = 0.8$ to their correlation functions. They only present correlation lengths in bins of stellar mass, so we cannot directly compare to what we find, but in their fig. 11 they plot the correlation length as a function of redshift, for different stellar mass bins. They find an increase with redshift for fixed stellar mass, for all stellar masses. We find an increase with median redshift for our blue, red, and ‘very red’ subsamples, raising the question whether these three subsample might contain galaxies of similar median stellar mass. Looking at Table 3, this is not the case. The median stellar mass depends mostly on subsample size, not redshift. The subsample with the most severe colour cut (and highest median stellar masses) have the largest clustering strengths. Furusawa et al. (2011) also find that clustering strength increases with stellar mass.

Hatfield et al. (2016) also see this trend, who studied clustering as a function of stellar mass, amongst others, for the VIDEO survey. The full galaxy sample used by Hatfield et al. (2016) for their clustering

analysis comprises 97 052 sources, for $K_s < 23.5$, and various stellar mass cuts. As we do not select for stellar mass, but just for redshift, we consider their estimates for the lowest stellar mass threshold they applied (they do not bin their data in stellar mass). Two redshift ranges overlap with ours: $0.75 < z < 0.9$, where they have 9791 galaxies for their clustering analysis, and $1.25 < z < 1.7$, with 10 800 galaxies remaining. The corresponding correlation strength r_0 they estimated (see their table 1) are $5.68^{+0.13}_{-0.076} h^{-1}$ Mpc and $6.47^{+0.11}_{-0.14} h^{-1}$ Mpc, respectively, which compare fairly well to what we find for our blue and red subsamples, respectively, given that the redshift distributions are different despite the overlap.

A comprehensive study of clustering of K -band ($2.2 \mu\text{m}$) selected galaxies in the UltraVISTA-DR1 survey was performed by McCracken et al. (2015), who measured their clustering strength as a function of stellar mass and star formation activity, amongst others. They found that at fixed redshift and scale, clustering amplitude depends monotonically on the sample stellar mass threshold. Looking at more specific results that allow for a comparison to our findings, their fig. 7 is most useful, which shows the correlation length as a function of sample median stellar mass. One of their redshift ranges is $0.8 < z < 1.1$, which has a similar width as our ‘blue’ subsample. We used the SHARK mocks (which have stellar masses listed) to obtain the median of $\log M_*/M_\odot$, as used by McCracken et al. (2015), for the full ‘blue’ subsample. We find a value of 10.2, which corresponds to a clustering length of $3.5 \pm 0.4 h^{-1}$ Mpc according to fig. 7 of McCracken et al. (2015), where the uncertainty is larger than the error bar in their figure to take into account the uncertainty in the value of 10.2 for the median of $\log M_*/M_\odot$. This is smaller (by almost a factor of 2) than the clustering length of $6.5 \pm 0.51 h^{-1}$ Mpc that we found for our ‘blue’ subsample, which is at somewhat lower redshifts, and has a more pronounced peak in its redshift distribution as compared to the top-hat shape of their $0.8 < z < 1.1$ redshift selection. A larger spread over redshift lowers the clustering strength, and their method used to estimate the clustering length (using a halo model fit) is different from ours. However, it is unclear whether all this sufficiently explains the difference, and merits further study.

An interesting approach is that by Cochrane et al. (2018), who use HiZELS to estimate clustering at very specific redshifts using the $H\alpha$ line. This yields four subsamples, the largest being over 2300 galaxies in 7.6 deg^2 at $z = 0.81 \pm 0.011$. They split this in

stellar mass bins, where only galaxies with stellar masses in the range $4 \times 10^{10} M_{\star}$ to $10^{11} M_{\star}$ produce a clustering strength close to our blue subsample (their fig. 4). What we can also derive from their work is that our more extreme blue cut, with its higher clustering strength, again seems to select higher stellar masses on average, as we found above. Also, their $z = 1.47$ redshift sample (451 galaxies over an area of 2.3 deg^2), best corresponding to our red subsample, also has a larger clustering strength for all stellar mass bins, as in Furusawa et al. (2011).

In the optical, Coil et al. (2008) measured clustering in the DEEP2 survey around $z \sim 1$, for subsamples which all start at $z > 0.7$, and go down to $z \approx 0.8 - 1.2$ depending on the selection for $U - B$ colour in the optical. Their ‘main’ red and blue subsamples have a redshift range of $0.7 - 0.925$ and $0.7 - 1.05$, respectively, which in both cases is somewhat beyond our blue selected galaxies ($[3.6] - [4.5] < -0.3$), which peak at $z \approx 0.7$, and have a range that best matches their ‘main’ red (in the optical) subsample. For this subsample they find a spatial clustering length $r_0 = 5.25 \pm 0.26 h^{-1} \text{ Mpc}$, which is comparable to what we find for our blue (in the NIR) subsample.

In the far-IR wavebands, Amvrosiadis et al. (2019) quote a correlation length $r_0 = 8.1 \pm 0.5 h^{-1} \text{ Mpc}$ for a redshift range $1 < z < 2$. For our redshift ranges we find $r_0 = 7.83 \pm 0.95 h^{-1} \text{ Mpc}$ and $r_0 = 9.68 \pm 2.77 h^{-1} \text{ Mpc}$ for our red and ‘very red’ subsamples, respectively. As the redshift ranges are somewhat different (as are the observed bands), these results are remarkably similar, keeping in mind that the NIR flux is a proxy for the stellar mass while the FIR flux is a proxy for the star formation rate.

Starikova et al. (2012) provides clustering estimates in two different redshift ranges that are not too dissimilar to ours: they found $r_0 = 4.98 \pm 0.28 h^{-1} \text{ Mpc}$ for a population with a mean redshift $\langle z \rangle = 0.7$, similar to our blue subsamples, and $r_0 = 8.04 \pm 0.69 h^{-1} \text{ Mpc}$ for a population with $\langle z \rangle = 1.7$, which is similar to our red subsamples. Their first estimate quote above compares well to our broader $[3.6] - [4.5] < -0.3$ subsample, the latter is consistent with our ‘very red’ $[3.6] - [4.5] > 0.3$ subsample, and with the $1 < z < 2$ clustering length found by Amvrosiadis et al. (2019).

From these comparisons to existing estimates in the literature for a range of samples and wavelength ranges, we find that our findings are consistent, although it remains difficult to reach conclusions from such comparisons as the selected galaxy populations can be fairly different, despite overlapping redshift ranges. There will be overlap in galaxy properties as well, like stellar mass as we saw above, but there is likely to be significant spread, making the overlap far from complete.

This is not an issue for simulated galaxy populations, like the ones derived from the SHARK semi-analytical galaxy formation model (Lagos et al. 2019), as these are complete by definition and the same selections as for the observations can easily be applied. We had to offset the fluxes for the simulated galaxies somewhat to match overall number densities found for our flux cuts, brightening each galaxy by -0.565 mag at $3.6 \mu\text{m}$, and -0.61 mag at $4.5 \mu\text{m}$ (thus also including a minor colour correction to get the number densities right for the colour cuts). The reason this is necessary is likely to be the absence of AGN modelling in the current SHARK semi-analytical models, as can be seen from the counts (Lacy et al. 2021), and from Fig. 6, which is based on the work of Thorne et al. (2022). There are other model components that can act in this direction as well, such as changes in the star formation history, dust treatment, and adopted stellar population synthesis models. We also applied a small completeness correction, as the models are complete down to our flux limit, but the observed data set is ‘only’ ~ 90 per cent complete. We investigated mock samples with and without brightening, and found that the

mocks with brightening best match the observed clustering estimates (and also match number counts and colours, by construction). This is encouraging, and we take this as a motivation to improve the models, through the inclusion of AGN modelling, as well as different population synthesis and dust models. The amount of brightening that we needed to apply will need to be matched by future mocks.

The current mock data sets were treated in the same way as the observed ones, applying the same flux and colour cuts, and the correlation functions obtained in a robust way. These were again fitted to a single power-law (parameters are listed in Table 4). As different colour cuts select for different redshift ranges, we can compare the estimated clustering parameters for these redshift ranges, which is shown in Fig. 15 for the spatial clustering length r_0 (left-hand side panel) and the fitted slope δ (right-hand side panel), for the SERVS + DeepDrill sample (filled circles) as well as the brightened SHARK mocks (open circles). The vertical bars indicate the uncertainty in the fitting to the clustering estimates, but do not include systematic uncertainties like cosmic variance, although these should be fairly small owing to the large angular size of our survey. The horizontal bar denotes the redshift range for each (sub)sample, as derived from the S-COSMOS sample. For the full sample, this is the one shown in Fig. 1. For the subsamples selected on colour, the redshift distributions from Fig. 2 are shown. We find that for the full (brightened) mock sample (down to our flux limit) and the four colour selected subsamples, the fitted slopes are all somewhat larger than for the SERVS + DeepDrill sample (see right-hand panel of Fig. 15), which translates to the spatial clustering strength r_0 being somewhat larger in most cases (except for the ‘very blue’ subsample), as can be seen in the left-hand panel of Fig. 15. The slopes for the full sample and the red sample show the most significant mismatch between the data and the models, with a less significant mismatch for the ‘very red’ subsample. However, below $z < 1$ the slopes match well. This might indicate that the brightening of the mocks should be redshift dependent, or there is another redshift dependence in the treatment of the dust or the modelling of the stellar populations.

In all cases the power-law fits for the mock and real data cross over each other at a given length-scale, which is on the order of a few arcminutes for the full sample and the blue subsample, about an arcminute for both red samples, and below an arcminute for the ‘very blue’ subsample, which is also the most clustered one. So there seem to be a systematic offset in clustering strength, the origin of which is not readily apparent. This is not a large offset, and might be due to the absence of an AGN contribution to the SHARK mock galaxy fluxes: galaxy clustering depends fairly strongly on luminosity and colour (e.g. Zehavi et al. 2011), so if the addition of AGN emission to the mock galaxy fluxes affects a specific environment more than others, adding AGN emission to the models might well remove the small difference between the brightened model and the data we currently find. Alternatively, other changes might need to be made to the modelling, like aspects related to dust treatment and stellar population modelling. There is also an indication that these changes should be redshift dependent. All this will be explored in a follow-up paper for newer versions of SHARK, but also for other models for galaxy formation and evolution that are able to produce large mock catalogues of near-infrared galaxies.

ACKNOWLEDGEMENTS

This work and S-COSMOS are based on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory (JPL), California Institute of Technology, under a

contract with NASA. Support for SERVS was provided by NASA through an award issued by JPL/ Caltech. DeepDrill was funded by a *Spitzer* grant associated with program ID 11086. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. This publication also makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation. Basic research in Radio Astronomy at the U.S. Naval Research Laboratory is supported by 6.1 Base Funding. CL has received funding from the ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. SHARK has been supported by resources provided by The Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. We are grateful to J.C. Mauduit and J. Pforr for their work on the SERVS data processing and catalogues.

DATA AVAILABILITY

The data products from both SERVS and DeepDrill used for this paper (comprising the dual-band catalogues and bright star masks) are available for download from IRSA (irsa.ipac.caltech.edu/data/SPITZER/DeepDrill). This resource makes available mosaic images, coverage maps, uncertainty images, and bright star masks. Each field has two single-band catalogues cut at 5σ and a dual-band catalogue requiring a detection at $>3\sigma$ at both 3.6 and 4.5 μm . The simulated light-cone catalogue from SHARK is also included in the release, its columns are described in table 11 of Lacy et al. (2021).

REFERENCES

- Amvrosiadis A. et al., 2019, *MNRAS*, 483, 4649
- Baugh C. M., Gardner J. P., Frenk C. S., Sharples R. M., 1996, *MNRAS*, 283, L15
- Christodoulou L. et al., 2012, *MNRAS*, 425, 1527
- Cochrane R. K., Best P. N., Sobral D., Smail I., Geach J. E., Stott J. P., Wake D. A., 2018, *MNRAS*, 475, 3730
- Coil A. L. et al., 2008, *ApJ*, 672, 153
- Dale D. A., Helou G., Magdis G. E., Armus L., Díaz-Santos T., Shi Y., 2014, *ApJ*, 784, 11
- Davies L. J. M. et al., 2018, *MNRAS*, 480, 768
- de la Torre S. et al., 2007, *A&A*, 475, 443
- Elahi P. J., Welker C., Power C., Lagos C. del P., Robotham A. S. G., Cañas R., Poulton R., 2018, *MNRAS*, 475, 5338
- Farrah D. et al., 2006, *ApJ*, 641, L17
- Furusawa J., Sekiguchi K., Takata T., Furusawa H., Shimasaku K., Simpson C., Akiyama M., 2011, *ApJ*, 727, 111
- Gilli R. et al., 2007, *A&A*, 475, 83
- Hatfield P. W., Lindsay S. N., Jarvis M. J., Hauessler B., Vaccari M., Verma A., 2016, *MNRAS*, 459, 2618
- Ilbert O. et al., 2009, *ApJ*, 690, 1236
- Lacy M. et al., 2021, *MNRAS*, 501, 892
- Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, *MNRAS*, 481, 3573
- Lagos C. d. P. et al., 2019, *MNRAS*, 489, 4196
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Le Fèvre O. et al., 2005, *A&A*, 439, 877
- Limber D. N., 1953, *ApJ*, 117, 134
- Lonsdale C. J. et al., 2003, *PASP*, 115, 897
- Mauduit J.-C. et al., 2012, *PASP*, 124, 714
- McCracken H. J. et al., 2015, *MNRAS*, 449, 901
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19
- Papovich C., 2008, *ApJ*, 676, 206
- Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton Univ. Press, Princeton
- Pforr J., Vaccari M., Lacy M., Maraston C., Nyland K., Marchetti L., Thomas D., 2019, *MNRAS*, 483, 3168
- Phillips S., Fong R., Ellis R. S., Fall S. M., MacGillivray H. T., 1978, *MNRAS*, 182, 673
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Press W. H., Flannery B. P., Teukolsky S. A., Vetterling W. T., 1988, *Numerical Recipes: The Art of Scientific Computing*. Cambridge Univ. Press, Cambridge
- Robotham A. S. G., Bellstedt S., Lagos C. d. P., Thorne J. E., Davies L. J., Driver S. P., Bravo M., 2020, *MNRAS*, 495, 905
- Sanders D. et al., 2007, *ApJS*, 172, 86
- Skrutskie M. F. et al., 2006, *AJ*, 131, 1163
- Starikova S., Berta S., Franceschini A., Marchetti L., Rodighiero G., Vaccari M., Vikhlinin A., 2012, *ApJ*, 751, 126
- Thorne J. E. et al., 2022, *MNRAS*, 509, 4940
- Trayford J. W., Lagos C. d. P., Robotham A. S. G., Obreschkow D., 2020, *MNRAS*, 491, 3937
- Vaccari M., 2016, in Proceedings of "The many facets of extragalactic radio surveys: towards new scientific challenges" (EXTRA-RADSUR2015). Bologna, Italy, p. 27
- van Kampen E. et al., 2005, *MNRAS*, 359, 469
- van Kampen E. et al., 2012, *MNRAS*, 426, 3455
- Waddington I. et al., 2007, *MNRAS*, 381, 1437
- Wall J. V., Jenkins C. R., 2003, *Practical Statistics for Astronomers*, Cambridge Univ. Press, Cambridge
- Weaver J. R. et al., 2022, *ApJS*, 258, 11
- Zehavi I. et al., 2011, *ApJ*, 736, 59

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