

Mass Functions, Luminosity Functions, and Completeness Measurements from Clustering Redshifts

Dominic J. Bates,^{1*} Rita Tojeiro,¹ Jeffrey A. Newman,² Violeta Gonzalez-Perez,³ Johan Comparat,⁴ Donald P. Schneider,^{5,6} Marcos Lima,⁷ Alina Streblyanska⁸

¹*School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK*

²*Department of Physics and Astronomy, University of Pittsburgh, 3941 O'Hara Street, Pittsburgh, PA 15260*

³*University of Portsmouth Institute of Cosmology & Gravitation, Dennis Sciama Building, Portsmouth, PO1 3FX, UK*

⁴*Max-Planck-Institut für Extraterrestrische Physik Giessenbachstraße, 85748 Garching, Germany*

⁵*Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802*

⁶*Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802*

⁷*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, CEP 05314-970, São Paulo, SP, Brazil*

⁸*Instituto de Astrofísica de Canarias 38205 La Laguna, Tenerife, Spain*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

This paper presents stellar mass functions and i-band luminosity functions for Sloan Digital Sky Survey (SDSS) galaxies at $i < 21$ using clustering redshifts, from which we also compute targeting completeness measurements for the Baryon Oscillation Spectroscopic Survey (BOSS). Clustering redshifts is a method of obtaining the redshift distribution of a sample of galaxies with only photometric information by measuring the angular crosscorrelation with a spectroscopic sample in different redshift bins. We construct a spectroscopic sample containing data from the BOSS + eBOSS surveys, allowing us to recover redshift distributions from photometric data out to $z \approx 2.5$. We produce k-corrected i-band luminosity functions and stellar mass functions by applying clustering redshifts to SDSS DR8 galaxies in small bins of colour and magnitude. There is little evolution in the mass function between $0.2 < z < 0.8$, implying the most massive galaxies form most of their mass before $z = 0.8$. These mass functions are used to produce stellar mass completeness estimates for the Baryon Oscillation Spectroscopic Survey (BOSS), giving a stellar mass completeness of 80% above $M_{\star} > 10^{11.4}$ between $0.2 < z < 0.7$, with completeness falling significantly at redshifts higher than 0.7, and at lower masses. Large photometric datasets will be available in the near future (DECaLS, DES, Euclid), so this, and similar techniques will become increasingly useful in order to fully utilise this data.

Key words: galaxies: luminosity function, mass function – surveys – methods: data analysis – galaxies: distances and redshifts

1 INTRODUCTION

Large spectroscopic galaxy surveys are extremely useful tools for studying galaxy evolution. They allow us to determine stellar masses, star formation histories, and dynamics for large numbers of galaxies. In particular, deep, small area surveys such as PRIMUS (Coil et al. 2011), DEEP2 (Newman et al. 2013), and VIPERS (Guzzo et al. 2014) contain data for galaxies over a broad range of masses, colours, morphologies, and redshifts, allowing tests of galaxy evolution on very different objects.

These surveys, however, are not ideal for investigating galaxy evolution at the highest masses, since the number density of galaxies above, for example, $M_{\star} > 10^{11.5} M_{\odot}$, is extremely low. Due to their small area, these pencil-beam surveys typically only target tens or hundreds of galaxies above this mass, and are strongly affected by sample variance.

An ideal approach to study galaxy evolution at these masses is using large-volume cosmological redshift surveys, which typically target the highest mass galaxies over very large regions of the sky. The Baryon Oscillation Spectroscopic Survey (BOSS) (Eisenstein et al. 2011; Dawson et al. 2013) is the most extensive of these to date, measuring spec-

* E-mail: db217@st-andrews.ac.uk (KTS)

tra for roughly 1.5 million luminous red galaxies (LRGs) over $10,000 \text{ deg}^2$ of sky at $z < 0.7$. BOSS contains over 100,000 galaxies with stellar masses $M_\star > 10^{11.5} M_\odot$ (Maraston et al. 2013), so it is able to study this end of the mass function with very little shot noise. Ongoing and future surveys such as eBOSS (Blanton et al. 2017; Dawson et al. 2016) and DESI (DESI Collaboration et al. 2016) will extend this study to higher redshifts and larger numbers of galaxies, providing additional data to better probe these masses.

One limitation of these surveys, however, is that they are optimised for cosmology, not galaxy science. Their target selection therefore involves a number of complex colour cuts, leading to samples of galaxies that are incomplete in both stellar mass and colour. To study galaxy evolution at these masses, we must quantify this incompleteness.

One method of determining incompleteness is by comparing the distribution of galaxies as a function of mass in one sample, to that of another sample which is complete in stellar mass. In Leauthaud et al. (2016), they characterise the stellar mass completeness of BOSS using Stripe 82, a narrow region of the SDSS with deeper *ugriz* photometry, as well as near-IR photometry from the UKIRT Infrared Deep Sky Survey (UKIDSS) (Lawrence et al. 2007), allowing for more accurate photometric redshifts and stellar masses.

Large-area broad-band photometric surveys such as the Sloan Digital Sky Survey (SDSS) (York et al. 2000) are complete, provide data over a large area ($\approx 14000 \text{ deg}^2$) for galaxies over a range of magnitudes and colours, so would be ideal for this purpose. One disadvantage, however, is that for SDSS-like data, photometric redshifts can be unreliable (Rahman et al. 2016b). In this paper, we outline a method of computing luminosity and mass functions (and hence completeness) from broad-band surveys using a technique known as clustering redshifts.

Clustering redshifts is a method of obtaining the redshift distribution of set of galaxies via crosscorrelation with a spectroscopic sample. A number of slightly different techniques have been proposed; however, the main idea is the same: the two-point angular crosscorrelation is measured between a photometric sample of galaxies and different redshift bins of a spectroscopic sample. If the photometric sample overlaps in redshift with a particular bin of the spectroscopic sample, then the measured crosscorrelation will have a positive amplitude. Combining this crosscorrelation with bias information of the two samples, it is possible to accurately measure the redshift distribution of a photometric sample of galaxies.

In this paper we use clustering redshifts to recover the redshift distributions of samples of galaxies from the SDSS photometric survey in small bins of magnitude and colour, isolating galaxies of similar type. After recovering redshift distributions of bins, we use these colours and redshifts to compute stellar masses and luminosities by examining simulated galaxies in the same bins of colour-redshift space.

The layout of this paper is as follows: Section 2 describes both the real and mock data used in this study. Section 3 presents the clustering redshifts method and bias correction, and our method of computing stellar masses. In section 4 we test our clustering redshifts method on mock data, and determine how accurately mass and luminosity functions can be recovered using this technique. In section 5, this technique is applied to real SDSS photometry to produce real

mass and luminosity functions. Section 6 presents completeness measurements for BOSS using these computed mass and luminosity functions. Finally, in section 7, we discuss these results, and outline possible extensions of this work.

2 DATA

This study uses data from two main sources. We seek to compute mass functions from photometric data. In order to perform this task, we apply the clustering redshifts technique, which requires both an “unknown sample” (i.e. a photometric sample of unknown redshifts) and a “reference sample” (a sample over the same region of sky but with spectroscopic redshifts). The unknown sample is crosscorrelated with the reference sample to recover the redshift distribution.

2.1 BOSS & eBOSS Reference Sample

As a reference sample, we use data from both the SDSS-III: BOSS (Eisenstein et al. 2011; Dawson et al. 2013; Gunn et al. 2006; Smee et al. 2013) and SDSS-IV: eBOSS (Blanton et al. 2017; Dawson et al. 2016; Gunn et al. 2006; Smee et al. 2013) surveys. BOSS is a cosmological redshift survey that has measured spectra for ~ 1.5 million luminous red galaxies (LRGs) out to $z = 0.7$, over roughly $10,000 \text{ deg}^2$ of sky. Its primary aim is to map the spatial distribution of the highest mass galaxies over large volumes in order to measure the scale of baryon acoustic oscillations (BAO) in the clustering of galaxies. eBOSS is the extension of this program, ongoing at the moment, targeting $\sim 375,000$ LRGs (Prakash et al. 2016) at $0.7 < z < 0.8$, and $\sim 740,000$ quasars (Myers et al. 2015) over the range $0.9 < z < 3.5$ both over $7,500 \text{ deg}^2$ of sky.

These samples are ideal for a reference sample, as they cover a large area and provide continuous, large numbers of redshifts over the range $0 < z < 3$. The reference sample is therefore a combination of BOSS DR12 (Alam et al. 2015) LRGs in the North Galactic Cap (NGC) covering 6851 deg^2 and eBOSS DR14 data (Abolfathi et al. 2018), containing both the LRG and quasar samples, covering 1011 and 1214 deg^2 respectively, within the BOSS NGC area. The total sample is then 1.1 million galaxies. The number density of both the individual and combined samples are shown in figure 1. When computing correlation functions in later sections, we use large scale structure catalogues from Reid et al. (2012) for the BOSS LRG sample, Bautista et al. (2017) for eBOSS LRGs, and Ata et al. (2018) for eBOSS quasars, using 10x randoms for all samples.

2.2 SDSS Photometric Survey

Our photometric survey (i.e., the sample for which we wish to compute redshift distributions, along with masses and luminosities) consists of data from the SDSS photometric survey (York et al. 2000; Gunn et al. 1998). We use photometry from DR8 (Aihara et al. 2011) which contains *u*, *g*, *r*, *i* and *z* band information. We select only objects morphologically classified as galaxies, and only use data from the primary survey (i.e., the best observation for each object). To create our catalogues, we use *g*, *r* and *i* band modelMag magnitudes (see Stoughton et al. (2002)). We also constrain the sample

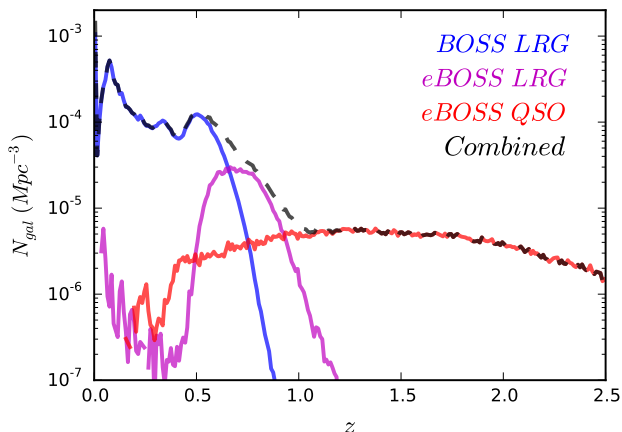


Figure 1. The comoving number density of three different spectroscopic galaxy samples described in section 2.1: At low redshift, BOSS DR12 LRGs (blue), intermediate redshifts, eBOSS DR14 LRGs (magenta), and higher redshifts, eBOSS DR14 quasars (red). The number density for the combination of these three samples is shown as the black dotted line.

to $i < 21$ to avoid incompleteness, and only include galaxies in the same region as the BOSS NGC DR12 footprint using the following masks detailed in Anderson et al. (2012): The survey geometry mask, and veto masks for bright stars mask, unphotometric seeing and bright objects. Finally, we remove all galaxies that are also in our reference sample, leaving 53 million galaxies over $\sim 7000 \text{ deg}^2$. We create random catalogues for this sample using the MANGLE software (Swanson et al. 2008), using 10x randoms.

2.3 Mock Surveys

In Section 4, we assess the reliability of our method on mock data, which requires both a mock reference sample and photometric survey. In later sections, we also use semi-analytic models to compute masses from colours and redshifts. Both these purposes require mock samples, and hence lightcones from two semi-analytic models (SAMs). We firstly take data from LGalaxies (Henriques et al. 2015). This lightcone covers 1/8th of the sky and is run on the Millennium simulation (Springel et al. 2005) rescaled to Planck cosmology (Planck Collaboration et al. 2014). Magnitudes are computed using Maraston & Strömbäck (2011) stellar population models (SSPs), using a Kroupa et al. (2001) initial mass function (IMF). We also use a smaller lightcone from SAGE (Croton et al. 2016) covering (100 deg^2 Area), run on the MultiDark MDPL2 simulation (Klypin et al. 2016; Knebe et al. 2018), with SEDs and magnitudes also computed using Maraston & Strömbäck (2011) SSPs and a Kroupa et al. (2001) IMF. Both catalogues have angular positions, redshifts, SDSS magnitudes (apparent and absolute) with reddening applied (Calzetti et al. 2000), and present day stellar masses.

We add photometric errors to the magnitudes of both SAMs by looking at how the error on a fitted magnitude in the SDSS varies as a function of that magnitude (i.e. g vs σ_g , r vs σ_r , i vs σ_i). For every mock galaxy, we use its

magnitude to compute the mean error at this magnitude in SDSS, then draw a Gaussian random error using this value as the standard deviation. We compute errors for all mock galaxies in the g , r and i bands, and add these errors to our mock galaxy magnitudes.

From these simulated galaxy catalogues, we define a mock reference sample and photometric survey. We define our reference sample by applying the colour and magnitude cuts of the BOSS survey described in Dawson et al. (2013) to both our LGalaxies and SAGE catalogues. This procedure produces samples with comparable redshift distribution to the BOSS survey and is further discussed in section 4.3. We refer to these samples as *BOSS_{LGalaxies}* and *BOSS_{SAGE}*. To create mock SDSS photometric samples, we cut both catalogues to $i < 21$ as in section 2.2, and also remove all galaxies present in our mock reference sample. We refer to these mock photometric surveys as *SDSS_{LGalaxies}* and *SDSS_{SAGE}*.

3 METHOD

Crosscorrelations have long been used to test for physical association (Seldner & Peebles 1976); however, the idea of using crosscorrelations to produce accurate redshift distributions has only become common over the last decade, partly due to the increase in data from large volume spectroscopic and photometric surveys.

Phillipps et al. (1985) investigated determining correlation functions from samples with only partial redshift information; later, in Phillipps & Shanks (1987), luminosity functions are computed given the assumption that galaxies close in the sky are likely at the same redshift. Schneider et al. (2006) more generally investigate this technique by measuring crosscorrelations with galaxies binned by photometric redshift. This approach is built on more formally in Newman (2008), and later Matthews & Newman (2010) and Matthews & A. (2012), where a method is outlined for computing redshift distributions by measuring the angular crosscorrelation between a photometric sample and different redshift bins of a spectroscopic sample. The amplitude of the crosscorrelation is fitted by an analytical form; since the redshift distribution inferred also depends on the evolution in bias of both samples, an iterative technique is employed to correct for this, assuming that clustering amplitude is proportional in both the spectroscopic and photometric sample.

Some variants on this method have subsequently appeared. For example, Schmidt et al. (2013) and Ménard et al. (2013), propose a similar technique, measuring angular crosscorrelations with a spectroscopic sample, but over constant physical scale. Furthermore, bias evolution is corrected for by assuming a bias evolution law, and the effect of this assumption is tested, down to non-linear scales. More recent studies applying these methods include Rahman et al. (2015), Rahman et al. (2016a), Rahman et al. (2016b), Scottez et al. (2016), and Scottez et al. (2018). van Daalen & White (2018) present a model for computing luminosity functions using clustering information and apparent magnitudes.

In Gatti et al. (2018) the performance of three of these methods are investigated: Newman (2008), Schmidt et al. (2013) and Ménard et al. (2013). They by apply all methods to simulated Dark Energy Survey (DES) data, finding

that Newman (2008) method produces slightly noisier redshift distributions due to having two extra degrees of freedom when fitting the crosscorrelation amplitude; furthermore, they report that the proportional bias assumption is not always accurate.

In our preliminary tests, the noise of all techniques was largely due to noise in the crosscorrelation functions; the choice of method made only small differences to the noisiness of the recovered $n(z)$ s. The main difference between methods is how the bias evolution correction is applied.

Since, firstly, Gatti et al. (2018) find that Menard method appears to produce slightly less noisy distributions, and secondly, we will be applying our own bias correction computed from semi-analytic models, we choose to adopt a method based on Ménard et al. (2013).

3.1 Clustering Redshifts Methodology

The method is detailed in Ménard et al. (2013) (hereon M13); we summarise the important points here, along with our alterations. The method is centered around computing the crosscorrelation between a photometric or “unknown” sample, and a number of redshift bins of a spectroscopic or “reference” sample.

Using the simplest Peebles & Hauser (1974) estimator, the angular crosscorrelation between two samples, 1 and 2, can be defined as, $\omega_{12}(\theta) = D_1 D_2(\theta) / R_1 R_2(\theta) - 1$, where $D_1 D_2(\theta)$ is the number of galaxies in sample 1 separated by an angular distance θ from galaxies in sample 2. $R_1 R_2(\theta)$ is the same statistic, but instead for two purely randomly distributed sets of points. The crosscorrelation function therefore describes, as a function of angle, the excess probability that galaxies in one sample will be situated a particular distance from galaxies in another. If the two samples considered overlap in redshift, they will occupy the same density field, and their positions will be correlated, hence this crosscorrelation will have a positive amplitude.

To produce an $n(z)$ measurement, we therefore need to measure the angular crosscorrelation, $\omega_{ur}(\theta, z)$, between an unknown sample, and different redshift bins of a reference sample. Since we are interested in how the amplitude of this quantity evolves with redshift, we integrate over θ to produce

$$\bar{\omega}_{ur}(z) = \int_{\theta_{min}}^{\theta_{max}} d\theta W(\theta) \omega_{ur}(\theta, z) \quad (1)$$

where $W(\theta)$ is the weight function, $W(\theta) = \theta^{-1}$, designed to optimise the signal-to-noise ratio. In order to probe the same physical scale at all redshifts, θ_{min} and θ_{max} are computed differently for each redshift, such that they correspond the same physical scales $r_{p,min}$ and $r_{p,max}$.

From M13, the integrated crosscorrelation is,

$$\bar{\omega}_{ur}(z) \propto \frac{dN_u}{dz}(z) \bar{b}_u(z) \bar{b}_r(z) \bar{\omega}_{DM}(z) \quad (2)$$

where $\frac{dN_u}{dz}(z)$ is the redshift distribution of the unknown sample, $\bar{b}_u(z)$ and $\bar{b}_r(z)$ are the evolution in bias of the unknown and reference samples, respectively, over the same scales, and $\bar{\omega}_{DM}(z)$ is the equivalent evolution in the integrated dark matter correlation function.

3.2 Correcting for Bias Evolution

In order to compute a redshift distribution, we need an estimate of $\bar{b}_u(z)$, $\bar{b}_r(z)$, and $\bar{\omega}_{DM}(z)$. Assuming linear biasing, the integrated autocorrelations of the unknown and reference samples as a function of redshift can be written as $\bar{\omega}_{uu}(z) = \bar{b}_u^2(z) \bar{\omega}_{DM}(z)$ and $\bar{\omega}_{rr}(z) = \bar{b}_r^2(z) \bar{\omega}_{DM}(z)$ respectively. We were able to measure both $\omega_{uu}(z)$ and $\omega_{rr}(z)$, so we can substitute these in to equation 2, producing,

$$\frac{dN_u}{dz}(z) \propto \frac{\bar{\omega}_{ur}(z)}{\sqrt{\bar{\omega}_{uu}(z) \bar{\omega}_{rr}(z)}} \quad (3)$$

We can measure $\bar{\omega}_{ur}(z)$, the integrated crosscorrelation between the unknown sample and each bin in redshift of the reference sample, and also $\bar{\omega}_{rr}(z)$, the integrated autocorrelation of the reference sample over the same redshift bins and physical scale. We can also remove the constant of proportionality by normalising $\frac{dN_u}{dz}(z)$ to the number of galaxies in the unknown sample. The only parameter we cannot compute is $\bar{\omega}_{uu}(z)$, since we have no redshift information for the unknown sample.

M13 show in their figure 1 that for a range of assumed bias evolutions of the unknown sample, if the redshift distribution is narrow, $\sigma_z < 0.2$, the effects of bias evolution on the recovered distribution are small, and therefore the distribution can be estimated as $\frac{dN_u}{dz}(z) \propto \bar{\omega}_{ur}(z)$. Some papers choose to assume this relation, e.g. M13, Schmidt et al. (2013), or factor any deviation from this into their error budgets Gatti et al. (2018).

We will be eventually applying this approach to many different bins of galaxies, and as shown in section 4.1, while distributions are generally narrow, they can often be wider than $\sigma_z < 0.2$. For this reason, we choose instead to correct for this evolution by examining how the bias (or more accurately $\bar{\omega}_{uu}(z)$) evolves in semi-analytic models (SAMs). Note that for all our correlation function measurements, we use the Landy & Szalay (1993) estimator. Furthermore, where cosmology is needed we assume Planck Collaboration et al. (2014) best fit cosmological parameters.

3.3 Computing a Bias Correction

To correct for the bias evolution of the unknown sample, we investigate how this evolves for different samples of galaxies in the LGalaxies SAM described in section 2.3. We must first measure the integrated autocorrelation, $\bar{\omega}_{uu}(z)$ as a function of redshift. We choose to measure the autocorrelation for several galaxy samples in different bins of i -band magnitude, so we can apply a correction as a function of their brightness, as the clustering of galaxies of different luminosities will likely be very different.

We measure $\bar{\omega}_{uu}(z)$ for 10 bins of i -band magnitude of width $\Delta i = 0.25$ between $17 < i < 20$ and $\Delta i = 0.125$ between $20 < i < 21$ (since we have significantly more galaxies at fainter magnitudes). For each magnitude bin, we measure $\bar{\omega}_{uu}(z)$ in redshift bins of width $\Delta z = 0.1/3$, as this is the binning we will apply to test our data in section 4.1. This integrated autocorrelation is shown for three different magnitude bins, and for both small scales ($0.5 < r_p < 1.5$ Mpc) and large scales ($5 < r_p < 15$ Mpc) in figure 2, along with the redshift distributions of the three magnitude bins.

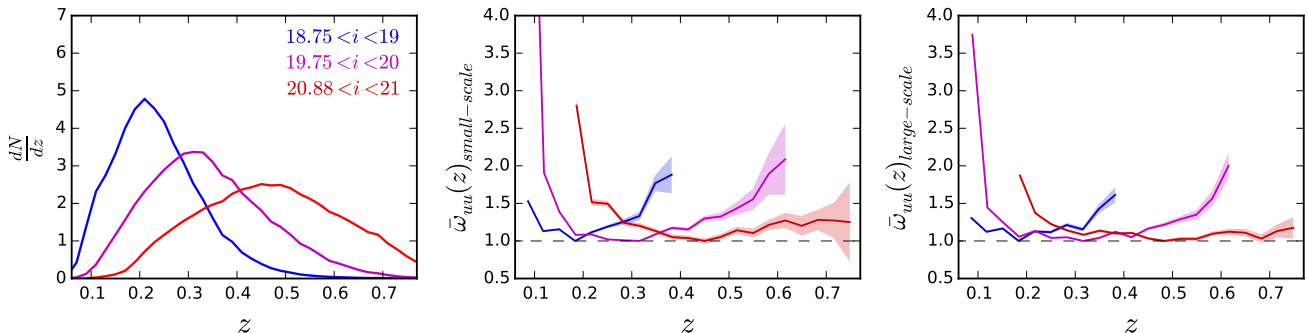


Figure 2. (left) The normalised redshift distributions of three different magnitude bins of LGalaxies data: $18.75 < i < 19$ (blue), $19.75 < i < 20$ (magenta), and $20.88 < i < 21$ (red). The middle and right hand side plots show the integrated angular autocorrelation function, $\bar{\omega}_{uu}(z)$, as a function of redshift for these three bins. The distributions are shown for small scales, $0.5 < r_p < 1.5$ Mpc (middle) and large scales, $5 < r_p < 15$ Mpc, (right). Since the clustering redshifts method only depends on the evolution of $\bar{\omega}_{uu}(z)$, not the overall amplitude, integrated correlation functions for are normalised such that the minimum value is 1.

Looking at both small scale and large scale clustering, in all bins of magnitude, there is a significant increase in the clustering amplitude towards lower redshifts. This behavior is particularly noticeable in the two faintest magnitude bins. There is also, in all magnitude bins, an increase in clustering amplitude towards higher redshifts, although in general this evolution is smaller at larger scales. Since this evolution is very different for different magnitude bins, we suggest that evolution in $\bar{b}_u^2(z)$ is driving this result, rather than $\bar{\omega}_{DM}(z)$.

The reason that the clustering amplitude increases at low redshift is due to the fact that we have magnitude-limited samples. For a given magnitude, galaxies at high redshift will be intrinsically luminous, and likely to be centrals, where as low redshift galaxies will be dim and likely satellites. This change in satellite fraction would lead to a substantial difference in the clustering amplitude, particularly at small scales, as seen in figure 2. This evolution is more significant in faint magnitude bins where galaxies span a larger range of redshifts, and therefore luminosities. The increase in amplitude towards higher redshift is likely due to the fact that for a given magnitude bin, the highest redshift galaxies are likely very luminous and massive, hence strongly biased.

Although large-scale clustering is less dependent on the bias evolution, the recovered $\phi(z)$ is significantly more noisy, mostly due to the angular crosscorrelation signal being diluted by foreground/background galaxies. Furthermore, this signal is more susceptible to spurious correlations due to large-scale structure (e.g., chance alignment of structure at different redshifts). We discuss this issue in appendix A. Because of this effect, when applying to real data we choose to measure crosscorrelations over the scales 1.5 to 5 Mpc, as 1.5 Mpc is the smallest scale we can measure while being safely above the SDSS fiber collision radius of 62 arcseconds.

When applying the clustering redshifts method to data in subsequent sections, we use the measured evolution of $\bar{\omega}_{uu}(z)$ in *SDSSLGalaxies* as an estimate of the bias evolution of the unknown sample, following equation 3. This correction is computed in the same magnitude bin and over the same physical scales as the crosscorrelation is measured. We test this correction in section 4.1.

3.4 Computing Masses and Luminosities

We seek to bin galaxies in small bins of colour and magnitude (i.e. binned in three dimensions by i , $r - i$ and $g - r$). This binning is useful because firstly, since galaxy colours are strongly correlated with redshift, it limits the width of redshift distributions for each bin, which will in turn reduce the importance of the bias evolution correction (Newman 2008; Ménard et al. 2013). Secondly, since we intend to compute masses and luminosities, we require both colour and redshift information, so binning by colour is important.

After recovering the redshift distributions of galaxies in all these colour bins, for each bin, we know the value of i , $r - i$ and $g - r$, along with the number of galaxies at each redshift. At each redshift, we therefore have a measure of the rest-frame spectral energy distribution (SED) of galaxies in this bin. We can compute parameters from this SED (e.g. mass, luminosity) and allocate these to photometric galaxies in the correct quantities.

To compute masses and luminosities for our bins of colour, we choose to use semi-analytic models. After recovering redshift distributions of all bins of colour of our photometric survey, we compute the distribution of mass or luminosity within the same colour-redshift bin of the SAM, and apply this probability density function (PDF) to the real data (i.e. multiply this PDF by the number of galaxies in the same bin of the real data). After recovering masses and luminosities at each redshift, for all bins of colour, these distributions are summed to produce mass and luminosity functions.

Although SAMs do not predict the correct number density of galaxies of given colours, for a given set of colours and redshift, the type of galaxy (i.e. star formation history (SFH), mass, luminosity) should be representative of those in the real universe. This method allows us to account for photometric errors by adding these to our SAM, and to produce a PDF of mass and luminosity rather than just a best fit, ensuring that the correct distribution of mass is allocated in each bin. This technique is, in essence, similar to Pacifici et al. (2012), where a library of physically motivated SFHs is

computed from SAMs, and then used to fit individual galaxy SEDs.

4 TESTING THE METHODS

Before computing real mass and luminosity functions, we test our method using real and mock data. We create mock reference samples and photometric surveys as described section 2.3. This allows us to select galaxies from our photometric survey as function of magnitude and colour, and recover redshift distributions by crosscorrelating with the reference sample. We can then compare the recovered redshift distributions to the true distribution.

4.1 Clustering Redshifts on Mock Data

In order to test the clustering redshifts method, we bin mock photometric survey, *SDSS_LGalaxies* by *i*-band magnitude in bins of width $\Delta i = 0.25$ between $17 < i < 20$ and $\Delta i = 0.125$ between $20 < i < 21$ where the large number of galaxies allows us to bin more finely. Within each of these magnitude bins, we then bin by $r - i$, and then by $g - r$. We choose a number of bins such that each contains $> 100,000$ galaxies, as we found this to be roughly the minimum number of galaxies required to recover a noise-free $n(z)$. At fainter magnitudes, the size of bins is comparable to the photometric error in the SDSS, so smaller bins would not provide significantly more information as galaxies are already scattered between bins. Binning by i , $r - i$ and $g - r$ produces 492 bins.

We then recover the redshift distributions of all bins of *SDSS_LGalaxies* by crosscorrelating with a reference sample, *BOSS_LGalaxies*, as described in sections 3.1 and 3.2. We correct for bias evolution using the computed evolution in *LGalaxies* as in section 3.3. Correlation functions errors are computed using a jackknife method, which in turn is used to compute errors on the final dN/dz following equations 1 and 3. Figures 3 and 4 show the recovered and true redshift distributions of a selection of these colour bins, in bright and faint magnitude bins, respectively.

Figure 3 shows four bins of $r - i$ and $g - r$ covering the extent of the colour space. It can be seen that the redshift distribution, $\phi(z)s$, is recovered well for a range of different values of $g - r$ and $r - i$. Adding a bias correction does not significantly affect the recovered distribution, likely because distributions are narrow, and because the correction, computed in section 3.3, is fairly small at bright magnitudes. Examining the faintest magnitude bin in figure 4, redshift distributions are again recovered well for a range of different values of $g - r$ and $r - i$, however the bias evolution correction becomes more important. This effect appears to be particularly true for wider distributions, where the fraction of satellites is likely changing between low and high redshift, and hence boosting the small-scale clustering amplitude.

If using a photometric survey with smaller photometric error, for example DECaLS or DES, redshift distributions for a given colour bin would be much narrower since galaxies will be less scattered between neighboring bins. The correction therefore becomes less significant, particularly at faint magnitudes, where SDSS errors are large. Errors are visibly larger at higher redshift ($z > 0.65$), where the number density of objects in the reference sample is low, which

can sometimes cause an error in normalisation. This effect should average out over many bins, however, and will be less of a problem when using real data since the true BOSS sample has a larger area, and there are additional eBOSS galaxies and quasars above this redshift.

As a further check of our bias correction, we compute the true median redshift, $z_{med,true}$, for all colour bins, along with the median redshift using clustering redshifts, $z_{med,cz}$, both with and without a bias correction. We compute the error in the median redshift, $z_{med,true} - z_{med,cz}$, for all bins, and present the distribution of errors in figure 5.

Without a bias correction, median redshifts are almost always slightly below the true value. The bias correction shifts the median to higher redshifts, although there remains a similar amount of scatter around the correct value. These errors in the median redshift are fairly small however, relative to the size of our redshift bins ($\Delta z = 0.033$). The scatter is partly due to noise in the recovered redshift distribution, but also may arise because we compute a bias correction for an entire magnitude bin, and this approach may not necessarily describe the bias evolution of all bins of colour within this.

4.2 Clustering Redshifts on Real Data

Since we will be eventually applying this method to real SDSS photometry, we show tests of the method on real data. We use a photometric survey defined from the SDSS, again cut to $17 < i < 21$, described in section 2.2. As in section 4.1, we split the sample into bins of *i*-band magnitude of width $\Delta i = 0.25$ between $17 < i < 20$ and $\Delta i = 0.125$ between $20 < i < 21$, and then bin by $r - i$ and $g - r$ within each of these such that each bin contains $> 100,000$ galaxies. We crosscorrelate each of these bins with our reference sample described in section 2.1, consisting of BOSS and eBOSS LRGs and quasars, in order to recover redshift distributions.

An example of some recovered distributions is presented in figure 6. We also, for reference, show the redshift distribution from the GAMA survey (Baldry et al. 2018) in the same bins of magnitude and colour. GAMA is a spectroscopic survey magnitude limited to $r < 19.8$, targeted over $\sim 286 \text{ deg}^2$ of sky, its primary objective being to study structure on scales of 1 kpc to 1 Mpc. Below $r < 19.8$, GAMA is highly complete ($> 95\%$), although completeness drops for fainter magnitudes. Therefore, for $r \lesssim 19.8$, GAMA redshift distributions should be roughly comparable to the SDSS. We choose an intermediate magnitude bin, $19 < i < 19.25$, in order that we have galaxies over a range of redshifts. In order to lessen the effect of the *r*-band magnitude cut in GAMA, we only show the bluest bins such that bins have $r \lesssim 19.8$, where incompleteness is not significant.

Recoveries of SDSS redshift distributions generally match the corresponding GAMA colour bin well. Some small differences are visible; however, this was also true for the simulated data in figure 4, for which the mass function is recovered well. If we take only bins below $i < 19.25$, we can use GAMA to compute the error in the median redshift of each colour bin, $z_{med,GAMA} - z_{med,cz}$, as in section 4.1. After computing this for all bins, the average error is $\delta z_{med} = -0.01$, indicating no significant offset in the bias evolution.

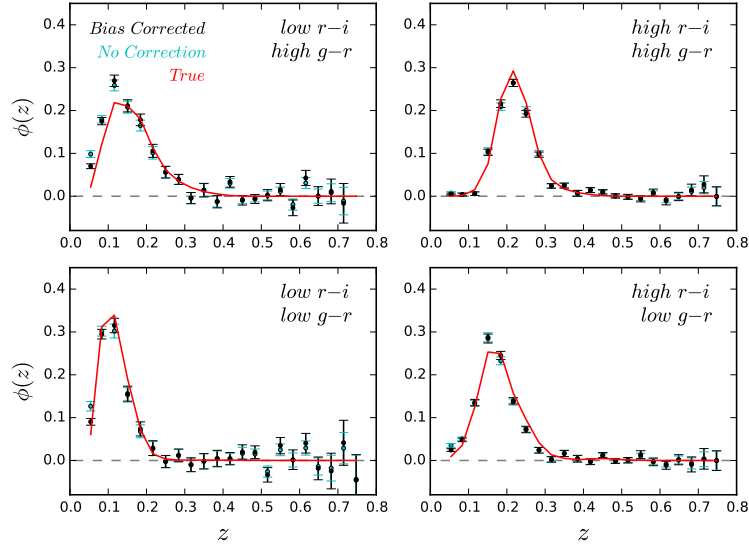


Figure 3. The recovered redshift distributions of different bins of colour of SDSS $LGalaxies$ data, with both no bias correction (cyan), and the bias correction computed in section 3.3 (black). The true distribution is given by the red line. We choose galaxies from a bright magnitude bin ($18 < i < 18.25$), and show four colour bins covering the extent of the colour space.

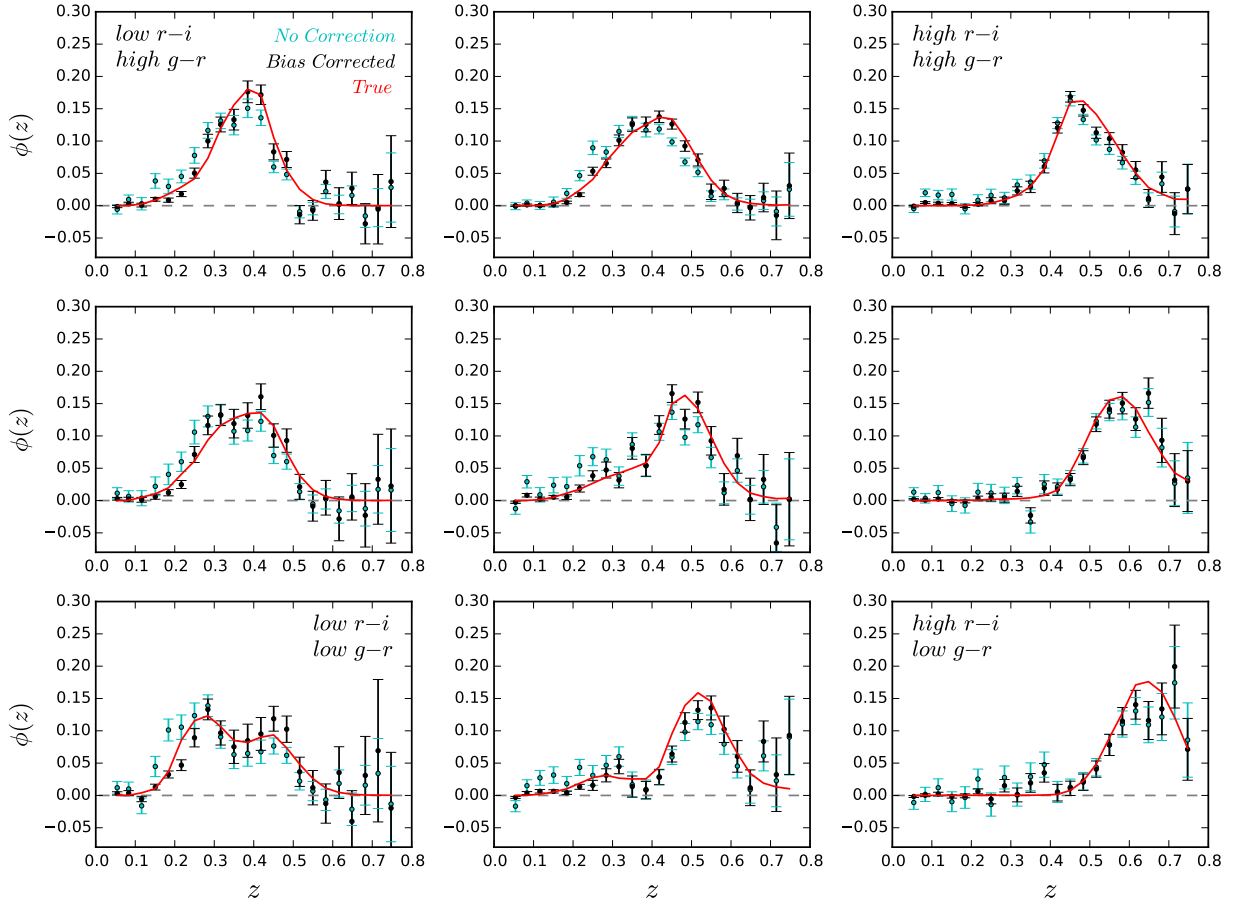


Figure 4. The recovered redshift distributions of different bins of colour of SDSS $LGalaxies$ data, with no bias correction (cyan), and the bias correction computed in section 3.3 (black). The true distribution is given by the red line. Here we choose galaxies from the faintest magnitude bin ($20.875 < i < 21$), containing 7×7 bins in $r-i$ and $g-r$. A selection of bins is presented throughout the colour space (bins 2, 4 and 6 in both dimensions). When computing redshift distributions, the crosscorrelation is integrated over small scales ($0.5 < r_p < 5$ Mpc).

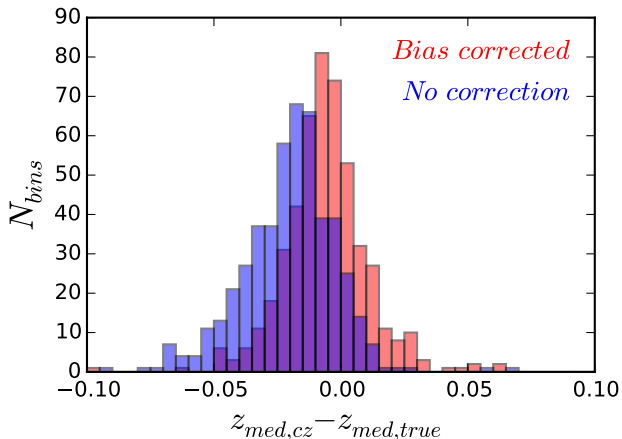


Figure 5. The distribution of errors in the median redshift, $z_{med,true} - z_{med,cz}$, for all bins of colour and magnitude, both with and without the bias correction.

4.3 Mass & Luminosity Functions of Mock Data

We now test our method of computing masses and luminosities, described in section 3.4, on mock data. As in section 4.1, we bin our mock photometric survey, $SDSS_{LGalaxies}$, into bins of colour and magnitude, recovering redshift distribution in each. We then take each of these bins at a given redshift and allocate masses and luminosities by looking in both LGalaxies (the same model, but with different photometric noise applied), and smaller lightcones from SAGE (a different model, also with photometric noise applied). This approach tests how much the choice of model affects the estimated stellar masses and luminosities. After summing the mass and luminosity distributions for all bins of colour and redshift, we produce mass and luminosity functions between $17 < i < 21$. Errors are computed using the error in $\phi(z)$ from the clustering redshifts method. The recovered luminosity and mass function are shown in figures 7 and 8.

Figure 7 displays the recovery of luminosity functions of our $SDSS_{LGalaxies}$ survey, in different redshift bins. Since the luminosities allocated to our galaxies are in the rest-frame, the recovered luminosity functions are by definition k-corrected. We use both LGalaxies and SAGE to compute Luminosities. The true luminosity function is recovered well at all redshifts, independent of whether LGalaxies or SAGE is used to compute an absolute magnitude. This result makes sense, since an absolute magnitude depends only on the redshift, cosmological model, and galaxy SED. Since we have accurate recovered redshifts and i , $r - i$ and $g - r$, we have effectively a rest frame SED, so the computed magnitude from this should not be particularly dependent on the SAM chosen.

Figure 8 shows mass functions, again recovered at different redshifts for the two different models. Using LGalaxies to recover masses works very well (i.e., the same model to convert colours and redshifts to masses), with the recovered mass functions almost exactly matching the true values at all redshifts and masses. Examining the SAGE results, at $M_{\star} < 10^{11.25} M_{\odot}$, mass functions are recovered well; how-

ever, above these masses, the number of high mass galaxies is under-predicted.

In order to understand this difference, we compare the distribution of colours in both models as a function of mass in figure 9. In the two lowest mass bins ($M_{\star} = 10^{9.25} M_{\odot}$ and $10^{10.25} M_{\odot}$), both SAGE and LGalaxies cover roughly the same colour space at both redshifts ($z = 0.25$ and 0.5). This result implies that colours of low mass galaxies ($M_{\star} \lesssim 10^{11} M_{\odot}$) are fairly independent of the semi-analytic model chosen, and explains why the mass function is recovered well at lower masses. In the high mass bin ($M_{\star} = 10^{11.25} M_{\odot}$), colours are visibly different in the two models. This behavior implies that high mass galaxies likely have different formation processes in the two models, and explains why mass functions are not recovered as well.

Since we do not know exactly which model best describes the real universe at high masses, we investigate how well both can reproduce the BOSS survey (containing large numbers of massive galaxies). We apply the colour cuts of BOSS to both samples as described in section 2.3, and compare the redshift distributions of these samples and the real BOSS survey in figure 10. It can be seen that LGalaxies reproduces both samples within the BOSS survey: the LOWZ sample at $0 < z < 0.4$, and CMASS sample at $0.4 < z < 0.8$. These are recovered with broadly the same number density, and although there is a slight offset in the peak of the sample, the overall shape of both samples is recovered well. SAGE manages to select some galaxies with a distribution similar to CMASS; however, at low redshift most galaxies are missing, and the overall shape is significantly different.

For this reason we trust that the colours of galaxies are significantly closer to those of the real universe in LGalaxies. We therefore opt to use LGalaxies when computing masses of real data in section 5.

5 MASS AND LUMINOSITY FUNCTIONS OF SDSS DATA

We now apply the technique to real SDSS data to produce mass and luminosity functions. As described in section 4.2, we recover redshift distributions for many bins of colour of SDSS galaxies between $17 < i < 21$. We compute mass and luminosity distributions for each bin using the colour-mass/luminosity relations of LGalaxies following section 4.3. Since our reference sample was originally removed from the SDSS sample, we compute masses and luminosities for these galaxies in the same way (i.e. for a given colour-redshift bin of our reference sample, we compute the mass distribution within the same bin of LGalaxies, and apply these masses). Note that we use spectroscopic redshift distributions instead of clustering redshifts for our reference sample. After adding these distributions together, we produce global mass and luminosity functions, shown in figures 11 and 12.

5.1 SDSS Mass Functions

The computed mass functions are presented in figure 11. The 95% completeness limits are shown in blue, computed as regions where LGalaxies becomes less than 95% complete due to the magnitude cuts. The bright magnitude cut ($i > 17$) is significant in the two lowest redshift bins; however, the

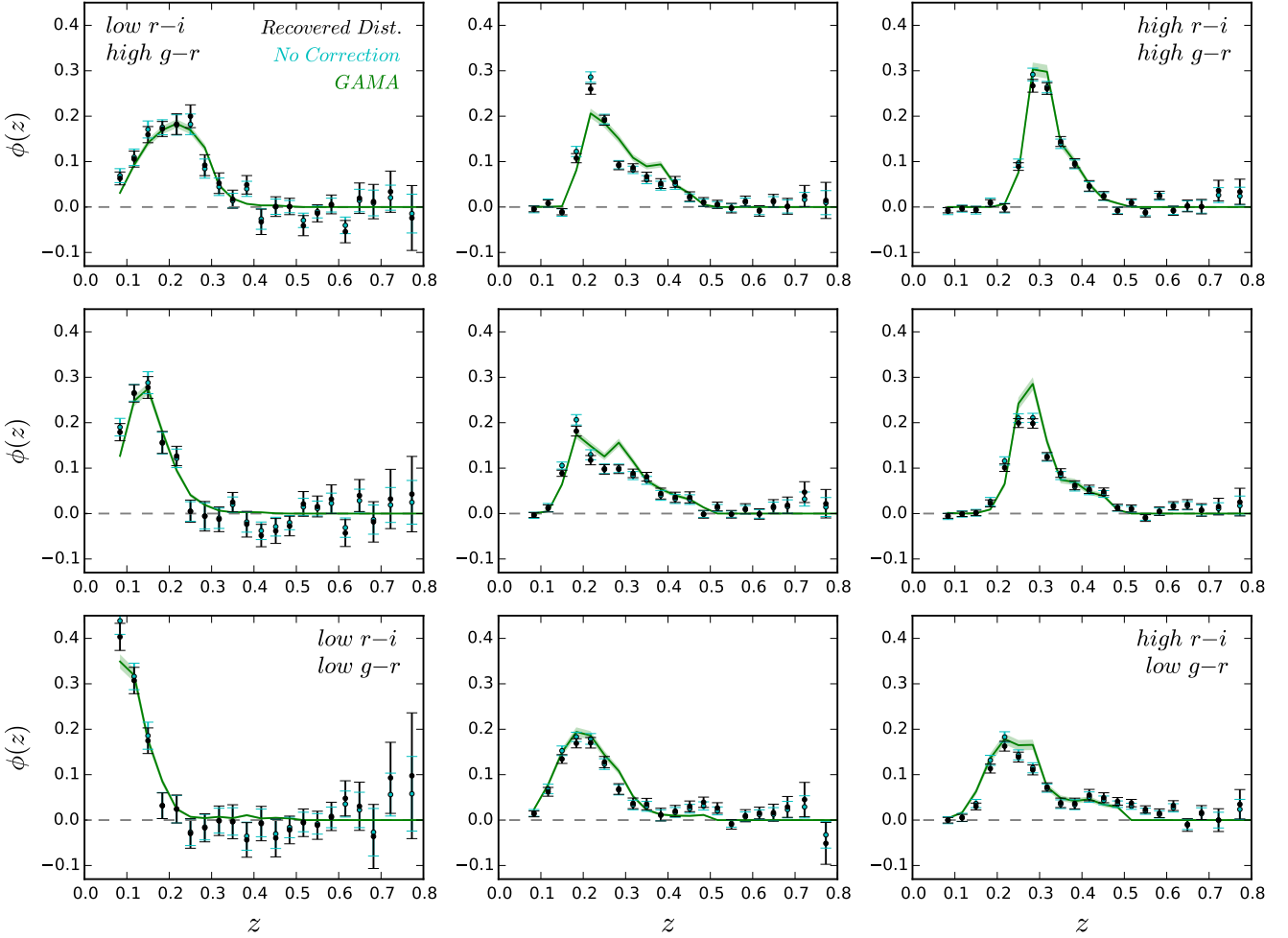


Figure 6. The recovered redshift distributions of different bins of colour of real data both with (black points), and without (cyan points) a correction. The spectroscopic redshift distribution of GAMA galaxies is indicated in the same colour bin (green line). We choose the magnitude bin ($19 < i < 19.25$), and only show bins with small values of $r - i$ in order to avoid the $r < 19.8$ magnitude cut in GAMA.

impact of this becomes less significant at higher redshifts. The faint magnitude cut becomes more significant at higher redshifts; however, we are still mostly complete at the very high mass end ($\gtrsim 10^{11}$) across the range ($0.4 < z < 0.8$). Tabulated versions of these mass functions are presented in appendix C.

Mass functions for the lowest redshift bins match closely with GAMA mass functions ($z < 0.06$) over the complete regions, indicating no significant offsets between our masses and GAMA. At the high mass end ($M_{\star} > 10^{11} M_{\odot}$), little evolution is evident over the redshift range ($0.4 < z < 0.8$), and the mass function is broadly consistent with GAMA ($z < 0.06$), implying there is no significant enhancement of the high mass end of the mass function after $z = 0.8$.

5.2 Constructing a Luminosity Function

Our computed luminosity functions are shown in figure 12. Magnitudes shown are absolute, dust-corrected magnitudes. Incompleteness is again visible for bright galaxies at low redshifts (due to the $i > 17$ cut); however, beyond redshift

0.4 we are complete for $M_i \lesssim -23.5$, allowing us to compare the evolution of the brightest galaxies across multiple bins.

There appears to be a significant amount of evolution over the range ($0.3 < z < 0.8$), with significantly more luminous galaxies present at higher redshifts. If these luminous galaxies are evolving passively, with little ongoing star formation, we would expect their stellar populations to decrease in brightness as young stars die out. Analysis of these luminosity functions to test for passive evolution may be investigated in a future paper.

6 COMPLETENESS

Since we have mass functions out to $z = 0.8$, we can measure the stellar mass completeness of BOSS. We first take both the SDSS and BOSS masses computed in section 5.1. The completeness at a particular redshift is therefore just the mass function of BOSS at that redshift divided by the SDSS mass function. The resulting completeness is displayed in figure 13 for 6 bins of redshift between $0.2 < z < 0.8$.

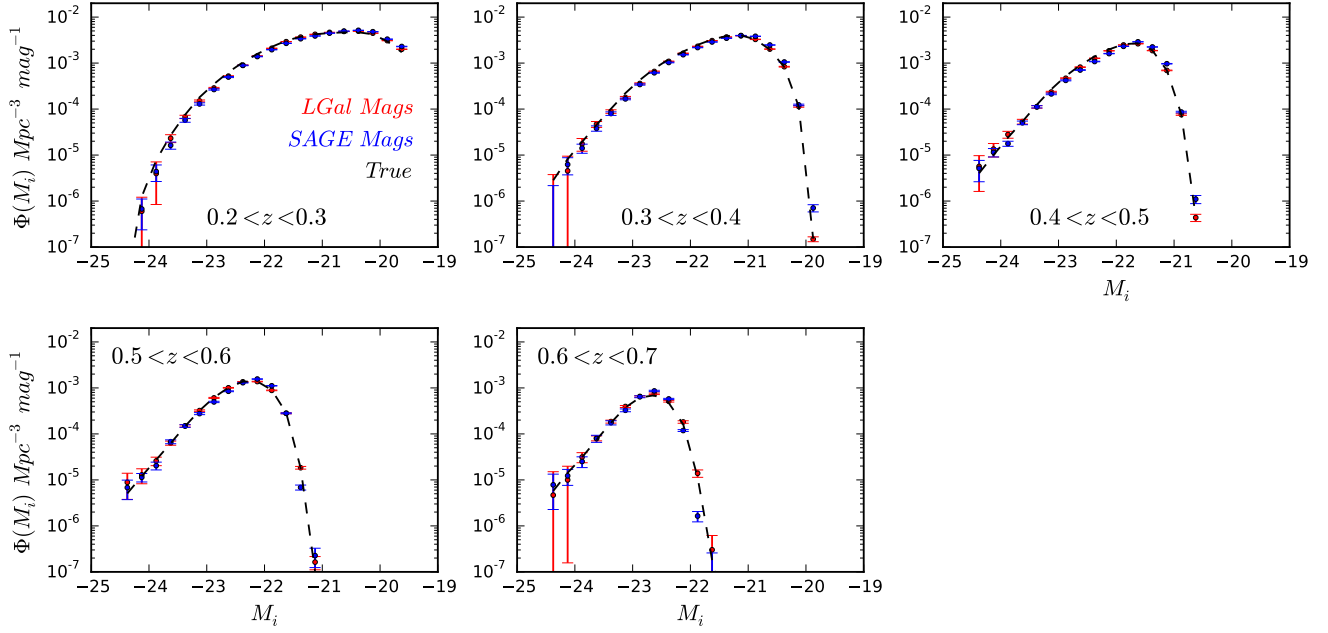


Figure 7. The recovered i-band k-corrected luminosity function of SDSS *LGalaxies* data for a number of different redshift bins between 0.2 and 0.7. Absolute magnitudes are recovered using colour-luminosity relations from both *LGalaxies* (red) and *SAGE* (blue) as described in section 4.3, and the true luminosity function is shown as the black dotted line. The fall-off in the luminosity functions towards faint magnitudes is due to the $i < 21$ cut in our sample.

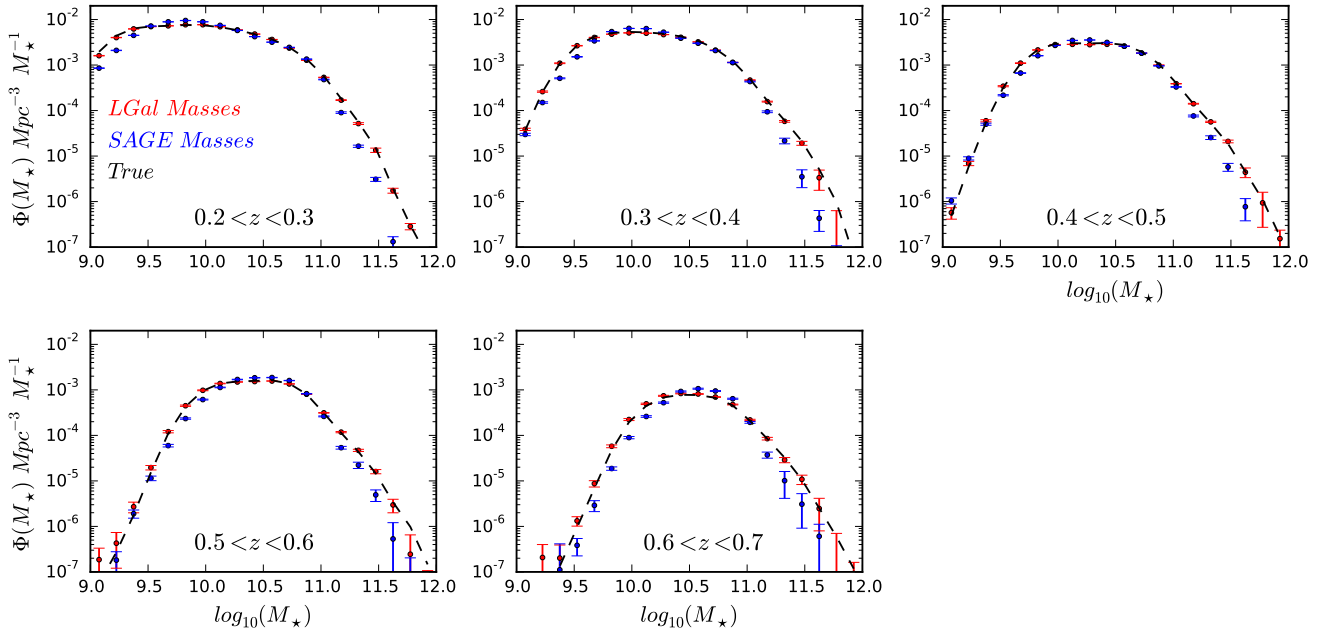


Figure 8. The recovered stellar mass function of SDSS *LGalaxies* data for a number of different redshift bins between 0.2 and 0.7. Masses are recovered using colour-mass relations from both *LGalaxies* (red) and *SAGE* (blue) as described in section 4.3, and the true stellar mass function is shown as the black dotted line. The fall-off in the mass functions towards lower masses is due to the $i < 21$ cut in our sample.

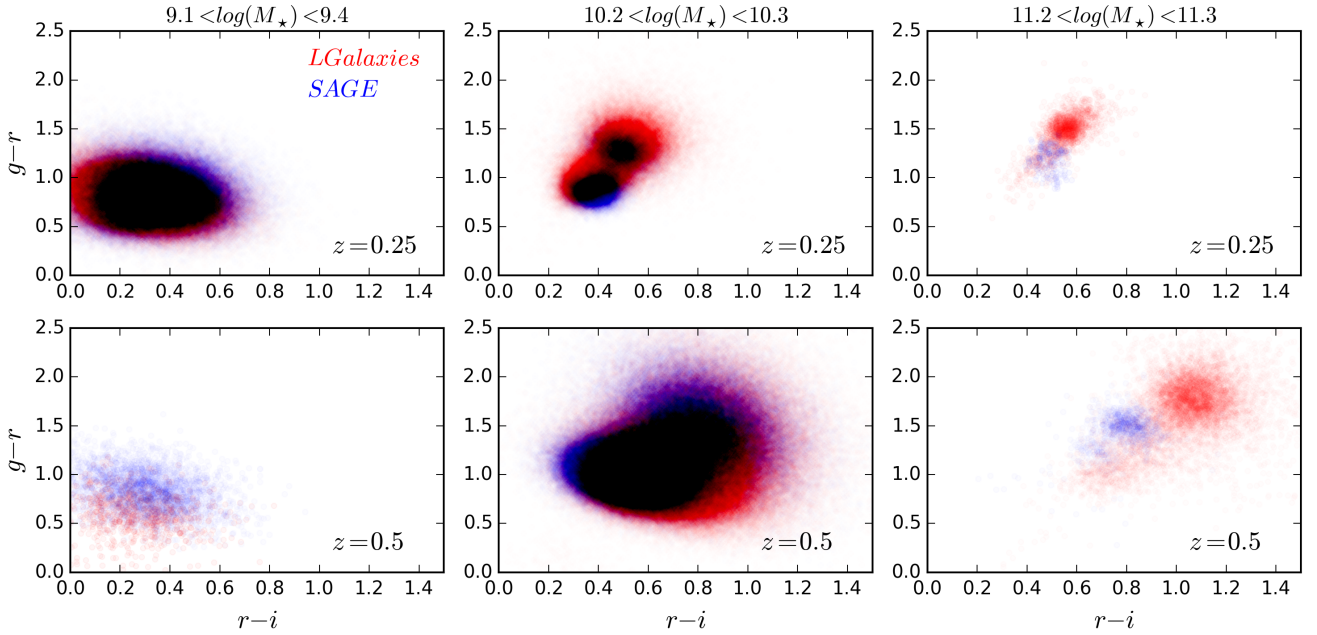


Figure 9. The $r - i$ and $g - r$ colours of galaxies in LGalaxies (red) and SAGE (blue). This distribution is shown for two different redshifts, 0.25 and 0.5, and for three bins of mass centered around 9.25, 10.25 and 11.25 $\log(M_{\odot})$.

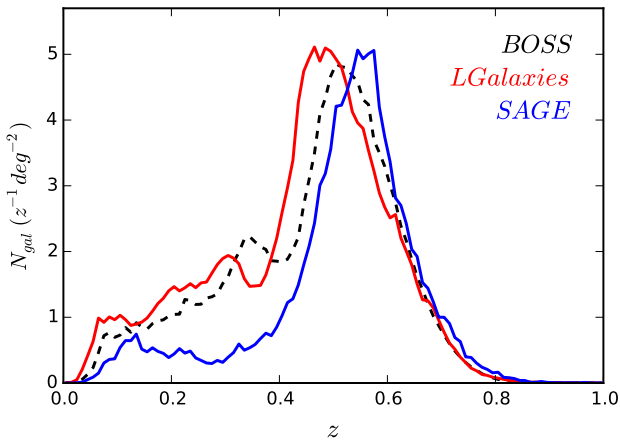


Figure 10. The redshift distribution (number of galaxies $\text{deg}^{-2} z^{-1}$) for the BOSS survey (black dotted line) compared with redshift distributions of LGalaxies (red) and SAGE (blue) with the BOSS colour cuts.

At low redshift $z < 0.4$, our SDSS mass functions are not complete at higher masses due to the bright magnitude cut ($i > 17$). This effect is also true for low masses at higher redshifts due to the faint ($i < 21$) cut. Completeness in these regions may not be fully representative and is shown in blue in figure 13. Between $0.4 < z < 0.8$, however, we are not affected by these cuts over the mass range of BOSS galaxies.

Between $0.2 < z < 0.7$, the stellar mass completeness of BOSS appears similar across all redshifts. Over this redshift range, above $M_{\star} \approx 10^{11.4} M_{\odot}$, BOSS is roughly 80% complete, with completeness falling to roughly zero at masses

lower than $M_{\star} \approx 10^{11} M_{\odot}$. In the $0.6 < z < 0.7$ bin, incompleteness appears at slightly higher masses than in the lower redshift bins. This decrease in completeness mirrors the decrease in number density of the sample shown in figure 1, which peaks just above $z = 0.5$ and falls off at higher redshifts. Looking in the highest redshift bin, BOSS is around 30% complete, only at the highest masses ($M_{\star} \gtrsim 10^{11.6}$).

Stellar masses are dependent on the the method used to obtain them. When mass functions or completeness measurements are compared between methods, any offsets should be taken in to account. We investigate the difference between our method and different BOSS stellar mass estimates in appendix B.

7 DISCUSSION AND CONCLUSIONS

In this study, we have demonstrated that clustering redshift can be used to successfully recover redshift distributions of galaxies in small bins of colour and magnitude of the SDSS by crosscorrelating with galaxies in the BOSS and eBOSS surveys. The importance of the bias correction becomes significant for fainter galaxies, where photometric errors are large, and galaxies are scattered between colour bins.

We have shown that mass and luminosity functions of mock data can be recovered using these recovered redshift distributions by computing masses using simulations in small bins of colour and redshift. We have also recovered mass functions of real data, and find little evolution at high masses between $0.2 < z < 0.8$, suggesting that the most massive galaxies form most of their mass before this time, and do not evolve significantly in mass afterwards. The lack of evolution over these redshifts agrees well with other studies, for example, Pérez-González et al. (2008); Moustakas et al.

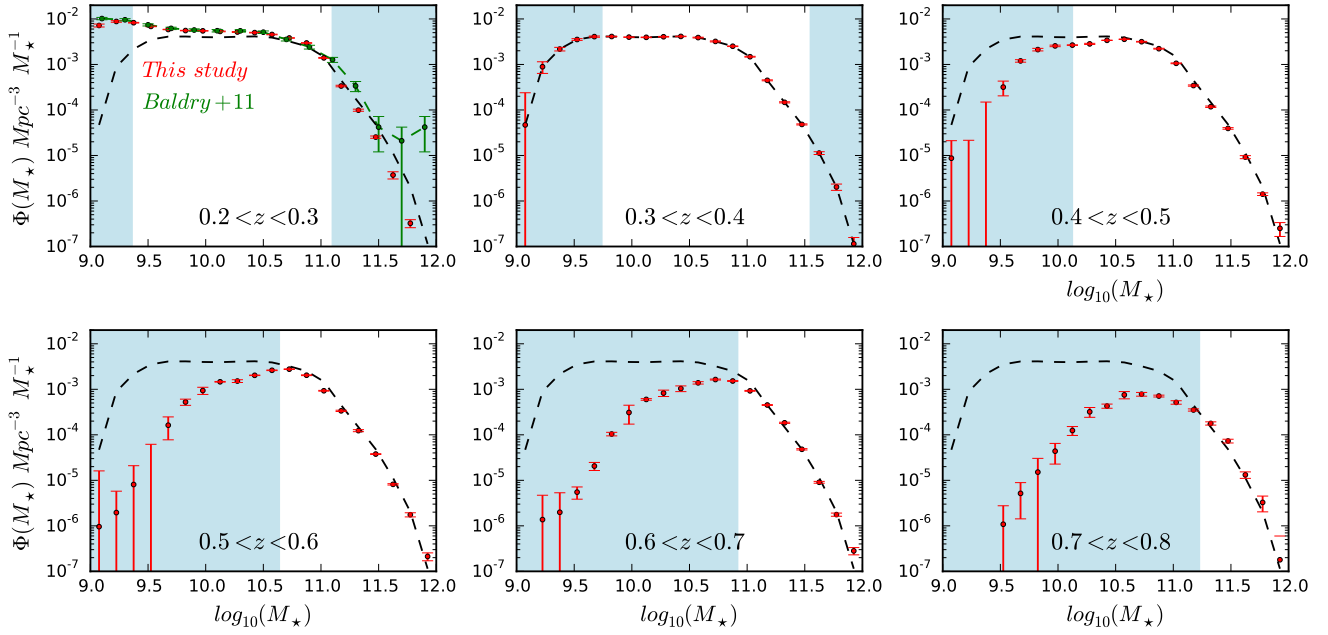


Figure 11. Recovered mass functions for real SDSS data in a number of different redshift bins (red). The green points in the redshift bin $0.2 < z < 0.3$ are the GAMA mass function ($z < 0.06$) (Baldry et al. 2012). For reference, the mass function computed using our method in the $0.3 < z < 0.4$ bin is shown in all bins as the black dotted line. Regions where our mass functions are more than 95% incomplete are shown in blue.

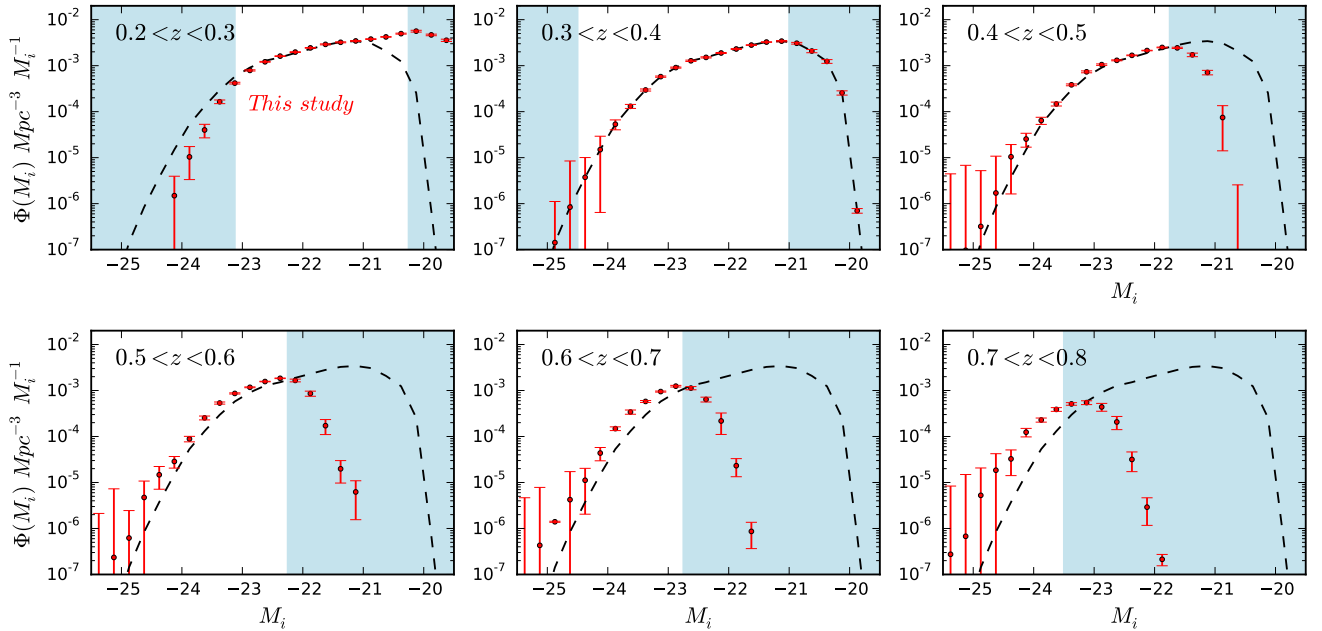


Figure 12. Recovered luminosity functions for real SDSS data in a number of different redshift bins (red). The luminosity function for ($0.3 < z < 0.4$) is shown as the black dotted line in all redshift bins for reference. Regions where our mass functions are more than 95% incomplete are shown in blue.

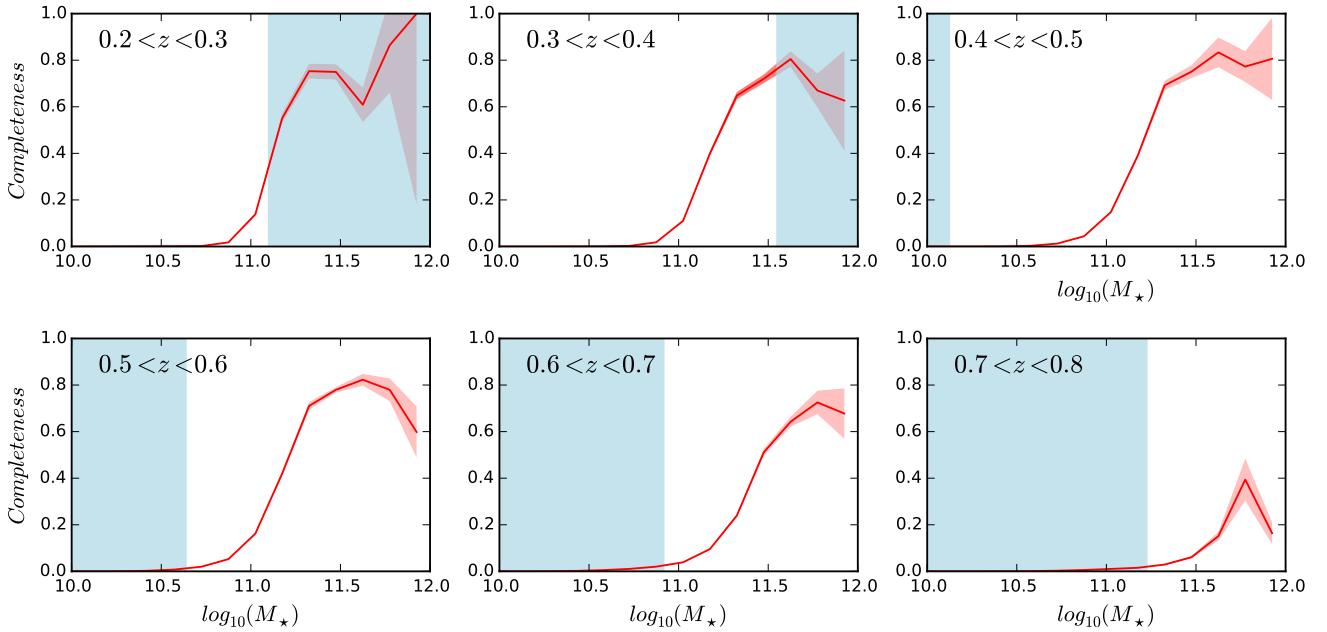


Figure 13. Stellar mass completeness estimates for BOSS between $0.2 < z < 0.8$, computed using the SDSS mass functions recovered in section 5.1. Completeness estimates, along with errors, are shown in red. Regions where the mass functions are more than 95% incomplete are the blue regions.

(2013); Leauthaud et al. (2016); Guo et al. (2018). Our luminosity functions show some evolution with redshift, possibly due to passive evolution; however, this will be investigated in a future paper.

We also produce targeting completeness measurements for BOSS using these mass functions, suggesting that over the redshift range $0.2 < z < 0.7$, BOSS is around 80% complete at high masses ($M_{\star} > 10^{11.4} M_{\odot}$), and falling to almost zero below $M_{\star} < 10^{11} M_{\odot}$. In our highest redshift bin ($0.7 < z < 0.8$) BOSS is strongly affected by incompleteness, and is only about 30% complete at the highest masses $M_{\star} \gtrsim 10^{11.6} M_{\odot}$. We also demonstrate that when comparing mass functions or completeness estimates between methods, significant offsets can be present, which require correction.

Guo et al. (2018) incorporate an missing fraction (incompleteness) component into their conditional stellar mass function model, and analyse the clustering of BOSS galaxies to produce completeness estimates for BOSS. They find that BOSS is around 80% complete above $M_{\star} \gtrsim 10^{11.3} M_{\odot}$ between $0.2 < z < 0.6$, with completeness falling off significantly at higher redshifts. This analysis is in good agreement with our results, showing very similar evolution with redshift and mass, although some offsets may be present due to using different mass estimates. Leauthaud et al. (2016), discussed in section 1, report similar completeness estimates at most redshifts and masses, however predict close to 100% completeness at the highest masses, which is not shown in Guo et al. (2018) or our estimates.

Ongoing and future large-volume spectroscopic surveys, for example eBOSS, DESI and EUCLID (Laureijs et al. 2011), will produce large number of spectra out to higher redshifts. This will firstly allow for better clustering redshifts estimates due to having a larger reference sample, but

also produce large spectroscopic galaxy samples, for which incompleteness must be understood. Combining these data with ongoing and future photometric surveys, for example, The Dark Energy Camera Legacy Survey (DECaLS) (Dey et al. 2018), and The Dark Energy Survey (DES) (DES Collaboration et al. 2017), will allow for redshift distributions to be computed out to higher redshifts, and in much smaller bins of colour, due to these new surveys reaching much deeper and having much smaller photometric error.

The methods used in this study, and similar techniques, will therefore be important tools for the next generation of galaxy surveys in order to utilise these large databases, and to understand the galaxy populations present.

ACKNOWLEDGEMENTS

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

REFERENCES

- Abolfathi B., et al., 2018, ApJS, 235, 42A
- Aihara H., et al., 2011, ApJS, 193, 29A
- Alam S., et al., 2015, ApJS, 219, 12A
- Anderson L., et al., 2012, MNRAS, 427, 3435A
- Ata M., et al., 2018, MNRAS, 473, 4773A
- Baldry I. K., et al., 2012, MNRAS, 421, 621B
- Baldry I. K., et al., 2018, MNRAS, 474, 3875B

Bautista J. E., et al., 2017, preprint (arXiv:1712.08064)
 Blanton M. R., et al., 2017, AJ, 154, 28B
 Calzetti D., et al., 2000, ApJ, 533, 682C
 Chen et al., 2012, MNRAS, 421, 314C
 Coil A. L., et al., 2011, ApJ, 741, 8
 Comparat J., et al., 2017, preprint (arXiv:1711.06575)
 Croton D. J., et al., 2016, ApJS, 222, 22C
 DES Collaboration et al., 2017, preprint (arXiv:1708.01530)
 DESI Collaboration et al., 2016, preprint (arXiv:1611.00036)
 Dawson K. S., et al., 2013, AJ, 145, 10D
 Dawson K. S., et al., 2016, AJ, 151, 44D
 Dey A., et al., 2018, preprint (arXiv:1804.08657)
 Eisenstein D. J., et al., 2011, AJ, 142, 72E
 Gatti M., et al., 2018, MNRAS, 477, 1664G
 Gunn J. E., et al., 1998, AJ, 116, 3040G
 Gunn J. E., et al., 2006, AJ, 131, 2332G
 Guo H., et al., 2018, ApJ, 858, 30G
 Guzzo L., et al., 2014, A&A, 43, 108G
 Henriques B. M. B., et al., 2015, MNRAS, 451, 2663H
 Klypin A., et al., 2016, MNRAS, 457, 4340K
 Knebe A., et al., 2018, MNRAS, 474, 5206K
 Kroupa P., et al., 2001, MNRAS, 322, 231K
 Landy S. D., Szalay A. S., 1993, ApJ, 412, 64L
 Laureijs R., et al., 2011, preprint (arXiv:1110.3193)
 Lawrence A., et al., 2007, MNRAS, 379, 1599
 Leauthaud A., et al., 2016, MNRAS, 457, 4021L
 Maraston C., Strömbäck G., 2011, MNRAS, 218, 2785M
 Maraston C., et al., 2013, MNRAS, 435, 2764M
 Matthews D. J., A. N. J., 2012, ApJ, 745, 180M
 Matthews D. J., Newman J. A., 2010, ApJ, 684, 88N
 Ménard B., et al., 2013, preprint (arXiv:1303.4722)
 Moustakas J., et al., 2013, ApJ, 767, 50M
 Myers A. D., et al., 2015, ApJS, 221, 27M
 Newman J. A., 2008, ApJ, 684, 88N
 Newman J. A., et al., 2013, ApJ, 208, 1
 Pacifici C., et al., 2012, MNRAS, 421, 2002P
 Peebles P. J. E., Hauser M. G., 1974, ApJS, 28, 19P
 Pérez-González P. G., et al., 2008, ApJ, 675, 234P
 Phillipps S., Shanks T., 1987, MNRAS, 227, 115P
 Phillipps S., et al., 1985, MNRAS, 212, 657P
 Planck Collaboration et al., 2014, A&A, 571A, 16P
 Prakash A., et al., 2016, ApJS, 224, 34P
 Rahman M., et al., 2015, MNRAS, 447, 3500R
 Rahman M., et al., 2016a, MNRAS, 457, 3912R
 Rahman M., et al., 2016b, MNRAS, 460, 163R
 Reid B. A., et al., 2012, MNRAS, 426, 2719R
 Schmidt S. J., et al., 2013, MNRAS, 431, 3307S
 Schneider M., et al., 2006, ApJ, 651, 14S
 Scottez V., et al., 2016, MNRAS, 462, 1683S
 Scottez V., et al., 2018, MNRAS, 474, 3921S
 Seldner M., Peebles P. J. E., 1976, ApJ, 227, 30S
 Smee S. A., et al., 2013, AJ, 146, 32S
 Springel V., et al., 2005, Nature, 435, 629S
 Stoughton C., et al., 2002, AJ, 123, 485S
 Swanson M. E. C., et al., 2008, MNRAS, 387, 1391S
 York D. G., et al., 2000, AJ, 120, 1579Y
 van Daalen M. P., White Martin ., 2018, MNRAS, 476, 4649V

APPENDIX A: TESTING THE FITTING SCALE

Here we show how the choice of fitting scale affects the recovered $\phi(z)$. As in section 4.1, we compute redshift distributions of mock data in small bins of magnitude and colour. Here we show the recovery of several colour bins within the faintest magnitude bin, but rather than integrating the crosscorrelation over small scales, as in figure 4, we integrate

over large scales ($15 < r_p < 50$ Mpc). The results are shown in figure A1

While redshift distributions are generally recovered successfully, there is a significant amount of extra noise when compared with the small scale recovery (figure 4). We compute the average error in $\phi(z)$ (i.e. the error due to errors in the correlation functions) for both small scales and large scales. We average this error across all colour bins, and all redshifts; when recovering redshift distributions over large scales, the error is on average 2.4x larger. When noise becomes large, a significant error in normalisation can appear, as seen in, for example, the bin of lowest $r-i$ and highest $g-r$ of figure A1. For this reason, we use only small scale clustering when applying to real data.

APPENDIX B: COMPARING BOSS MASS FUNCTIONS

Here we compare, for BOSS galaxies, our mass functions to mass functions computed using three other methods: 1) [Chen et al. \(2012\)](#), hereon C12, where galaxy parameters are modeled based on a library of model spectra for which principal components have been identified. 2) [Maraston et al. \(2013\)](#), hereon M13, where stellar population models are fit to the observed *ugriz* magnitudes, as well as the spectroscopic redshift of each galaxy. 3) [Comparat et al. \(2017\)](#), hereon C17, which for a given spectra finds the best-fit combination of single-burst SSPs. All three methods use [Maraston & Strömbäck \(2011\)](#) SSPs and a [Kroupa et al. \(2001\)](#) IMF. The four mass functions are presented in figure B1.

Although all methods generally agree on the shape of the mass function, there is a clear offset between methods. In particular, C12 predicts the highest masses. Both M13 and our method predict broadly the same shape as C12 at all redshifts, but this is offset towards slightly lower masses. This result may be related to the fact that in our method and M13, masses are computed from photometry rather than spectra. The shape of the C17 mass function appears slightly different. It predicts a larger number of low mass ($M_\star < 10^{10} M_\odot$) galaxies; however, the number of high mass galaxies is similar to our method. When comparing mass functions or completeness estimates across methods, this offset between different methods must be taken in to account.

APPENDIX C: TABULATED RESULTS

We present tabulated versions of our stellar mass functions and i-band luminosity functions in tables C and C2 respectively, and our completeness estimates in table C

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

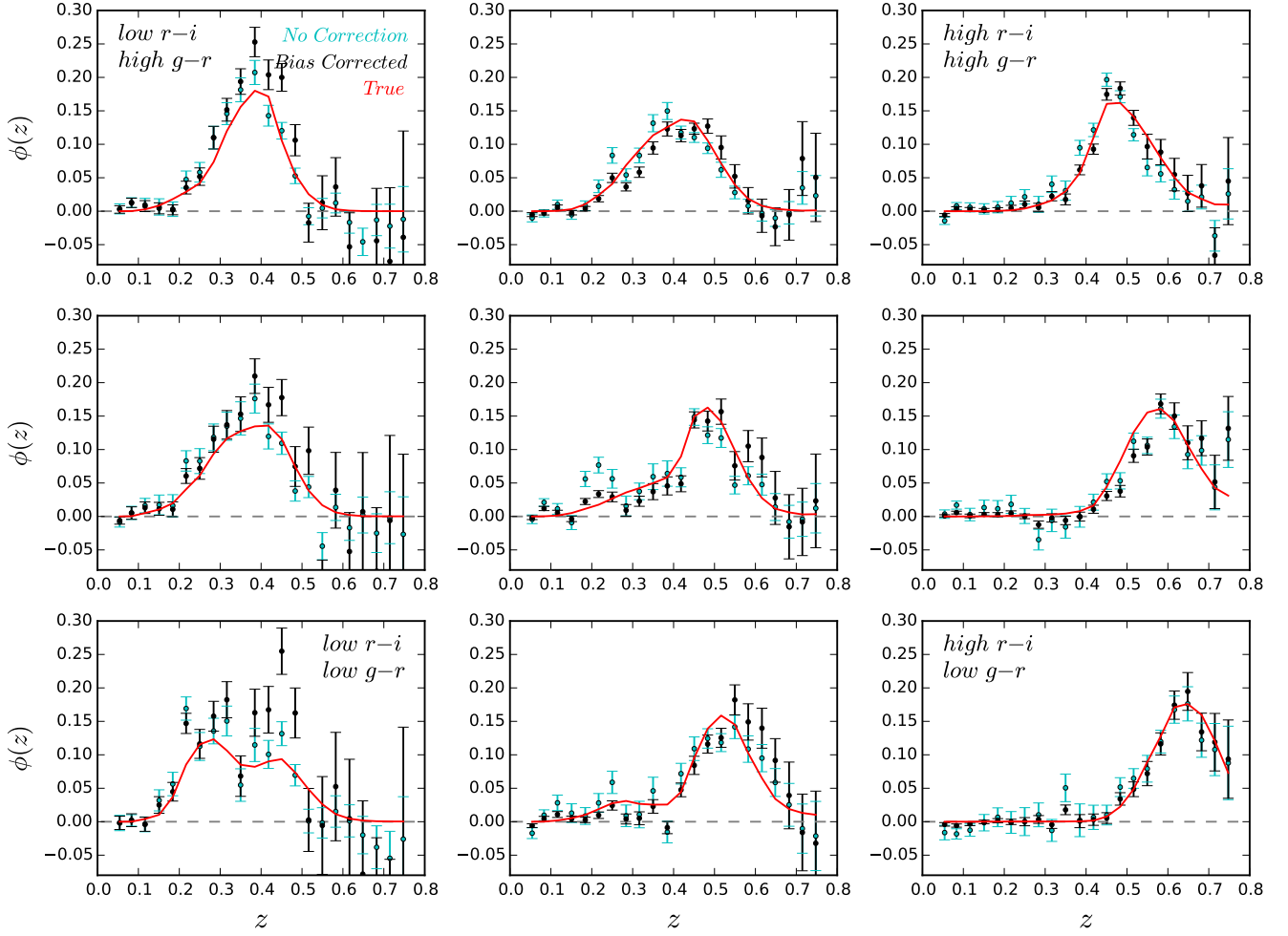


Figure A1. Same as in figure 4, except the crosscorrelation is integrated over large scales ($15 < r_p < 50$ Mpc), rather than smaller scales.

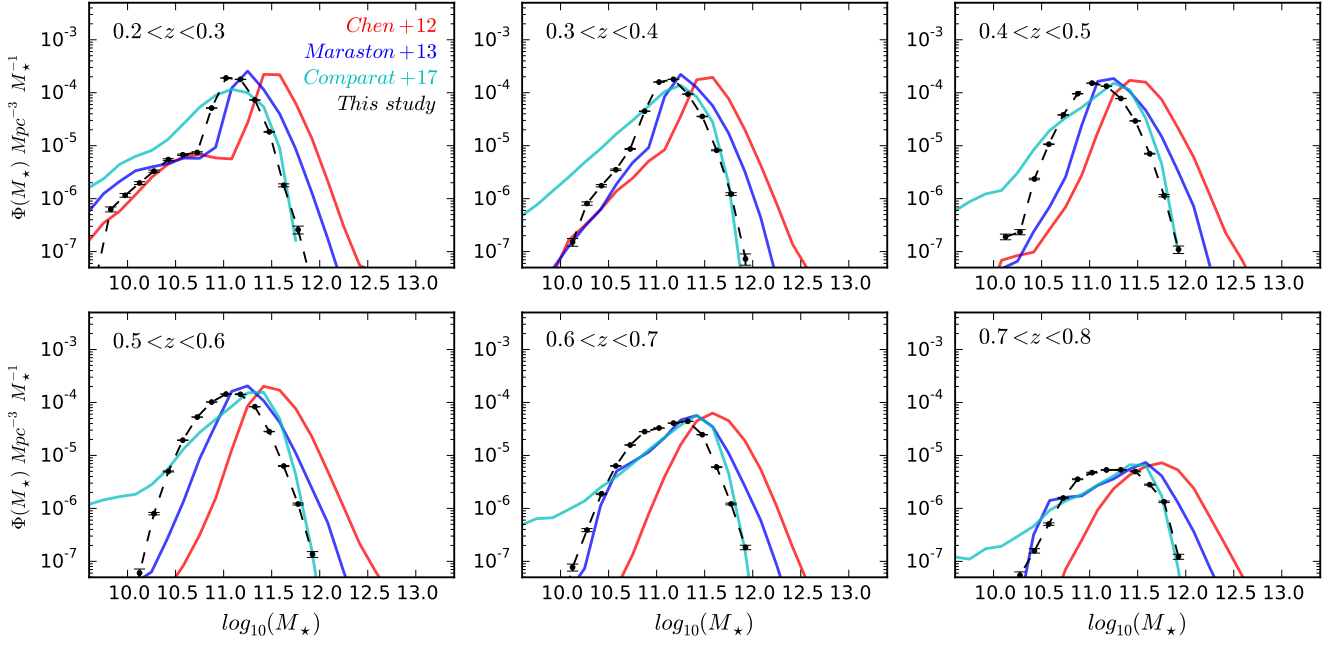


Figure B1. Mass functions of BOSS galaxies using four different methods, shown for six different bins of redshift. Our method described in section 4.3 is shown in black, along with C12 in red, M13 in blue, and C17 in cyan.

Table C1. Tabulated stellar mass functions computed as in section 5.1

0.2 < z < 0.3			0.3 < z < 0.4		
log(M_\star) (M_\odot)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)	log(M_\star) (M_\odot)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)
9.375	8.270	0.143
9.525	6.849	0.092
9.675	5.866	0.068
9.825	5.557	0.067	9.825	4.113	0.048
9.975	5.476	0.062	9.975	3.985	0.048
10.125	5.321	0.099	10.125	3.946	0.055
10.275	5.167	0.108	10.275	4.061	0.068
10.425	5.001	0.081	10.425	4.132	0.051
10.575	4.555	0.074	10.575	3.896	0.074
10.725	3.817	0.053	10.725	3.209	0.080
10.875	2.965	0.045	10.875	2.513	0.039
11.025	1.395	0.030	11.025	1.479	0.030
...	11.175	0.448	0.011
...	11.325	0.147	0.004
...	11.475	0.048	0.001
0.4 < z < 0.5			0.5 < z < 0.6		
log(M_\star) (M_\odot)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)	log(M_\star) (M_\odot)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)
10.275	2.832	0.079
10.425	3.391	0.035
10.575	3.577	0.046
10.725	3.162	0.056	10.725	2.768	0.045
10.875	2.221	0.046	10.875	2.038	0.038
11.025	1.059	0.021	11.025	0.929	0.018
11.175	0.344	0.011	11.175	0.336	0.011
11.325	0.118	0.004	11.325	0.123	0.006
11.475	0.039	0.002	11.475	0.038	0.001
11.625	0.0092	0.0007	11.625	0.0082	0.0003
11.775	0.0014	0.0001	11.775	0.0018	0.0002
11.925	0.00025	0.00008	11.925	0.00021	0.00005
0.6 < z < 0.7			0.7 < z < 0.8		
log(M_\star) (M_\odot)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)	log(M_\star) (M_\odot)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)
11.025	0.924	0.020
11.175	0.452	0.009
11.325	0.184	0.005	11.325	0.178	0.015
11.475	0.048	0.002	11.475	0.073	0.007
11.625	0.0091	0.0004	11.625	0.013	0.002
11.775	0.0018	0.0001	11.775	0.0033	0.0012
11.925	0.00028	0.00005	11.925	0.00018	0.00042

Table C2. Tabulated i-band luminosity functions computed as in section 5.2

0.2 < z < 0.3			0.3 < z < 0.4		
M_i (mag)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)	M_i (mag)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)
...	-24.375	0.0037	0.0063
...	-24.125	0.015	0.014
...	-23.875	0.053	0.013
...	-23.625	0.131	0.013
...	-23.375	0.296	0.015
-23.125	0.417	0.018	-23.125	0.577	0.018
-22.875	0.789	0.024	-22.875	0.907	0.024
-22.625	1.213	0.025	-22.625	1.276	0.043
-22.375	1.612	0.029	-22.375	1.513	0.043
-22.125	1.972	0.037	-22.125	1.886	0.048
-21.875	2.435	0.046	-21.875	2.302	0.049
-21.625	2.905	0.071	-21.625	2.804	0.049
-21.375	3.221	0.078	-21.375	3.26	0.053
-21.125	3.445	0.073	-21.125	3.417	0.107
-20.875	3.778	0.077
-20.625	4.216	0.071
-20.375	4.968	0.121
0.4 < z < 0.5			0.5 < z < 0.6		
M_i (mag)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)	M_i (mag)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)
-25.125	0.000	0.007	-25.125	0.000	0.007
-24.875	0.000	0.005	-24.875	0.001	0.002
-24.625	0.002	0.009	-24.625	0.005	0.006
-24.375	0.010	0.009	-24.375	0.015	0.008
-24.125	0.025	0.008	-24.125	0.029	0.008
-23.875	0.064	0.011	-23.875	0.089	0.012
-23.625	0.147	0.014	-23.625	0.254	0.024
-23.375	0.383	0.018	-23.375	0.534	0.029
-23.125	0.735	0.035	-23.125	0.864	0.025
-22.875	1.053	0.040	-22.875	1.174	0.025
-22.625	1.306	0.031	-22.625	1.567	0.022
-22.375	1.677	0.030	-22.375	1.842	0.028
-22.125	2.132	0.029
-21.875	2.490	0.045
-21.625	2.437	0.078
0.6 < z < 0.7			0.7 < z < 0.8		
M_i (mag)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)	M_i (mag)	Φ ($10^{-3} Mpc^{-3}$)	Φ_{err} ($10^{-3} Mpc^{-3}$)
-25.125	0.000	0.007	-25.125	0.001	0.014
-24.875	0.001	0.001	-24.875	0.005	0.015
-24.625	0.004	0.013	-24.625	0.018	0.024
-24.375	0.011	0.009	-24.375	0.033	0.018
-24.125	0.044	0.014	-24.125	0.124	0.026
-23.875	0.148	0.014	-23.875	0.229	0.023
-23.625	0.341	0.036	-23.625	0.388	0.036
-23.375	0.579	0.028
-23.125	0.949	0.025
-22.875	1.243	0.050

Table C3. Tabulated stellar mass completeness for BOSS computed as in section 6

0.2 < z < 0.3			0.3 < z < 0.4		
log(M_\star) (M_\odot)	Completeness	σ_{comp}	log(M_\star) (M_\odot)	Completeness	σ_{comp}
10.575	0.0015	0.0001	10.575	0.0009	0.0001
10.725	0.0020	0.0001	10.725	0.0027	0.0001
10.875	0.0178	0.0003	10.875	0.0180	0.0003
11.025	0.1380	0.0028	11.025	0.1104	0.0016
...	11.175	0.3997	0.0085
...	11.325	0.6478	0.0158
...	11.475	0.7198	0.0183
0.4 < z < 0.5			0.5 < z < 0.6		
log(M_\star) (M_\odot)	Completeness	σ_{comp}	log(M_\star) (M_\odot)	Completeness	σ_{comp}
10.575	0.0031	0.0001
10.725	0.0124	0.0002	10.725	0.0190	0.0004
10.875	0.0438	0.0009	10.875	0.0501	0.0010
11.025	0.1473	0.0025	11.025	0.1586	0.0024
11.175	0.3893	0.0095	11.175	0.4146	0.0095
11.325	0.6904	0.0204	11.325	0.7030	0.0162
11.475	0.7469	0.0281	11.475	0.7699	0.0111
11.625	0.8322	0.0637	11.625	0.8172	0.0251
11.775	0.7719	0.0660	11.775	0.7703	0.0484
11.925	0.8036	0.1761	11.925	0.5874	0.1085
0.6 < z < 0.7			0.7 < z < 0.8		
log(M_\star) (M_\odot)	Completeness	σ_{comp}	log(M_\star) (M_\odot)	Completeness	σ_{comp}
11.025	0.0357	0.0007
11.175	0.0909	0.0018
11.325	0.2318	0.0061	11.325	0.0270	0.0019
11.475	0.4971	0.0137	11.475	0.0576	0.0046
11.625	0.6266	0.0212	11.625	0.1451	0.0177
11.775	0.7042	0.0499	11.775	0.3818	0.0888
11.925	0.6521	0.1070	11.925	0.1583	0.0458