

# The Stripe 82 Massive Galaxy Project – II. Stellar mass completeness of spectroscopic galaxy samples from the Baryon Oscillation Spectroscopic Survey

Alexie Leauthaud,<sup>1\*</sup> Kevin Bundy,<sup>1</sup> Shun Saito,<sup>1</sup> Jeremy Tinker,<sup>2</sup> Claudia Maraston,<sup>3</sup> Rita Tojeiro,<sup>4</sup> Song Huang,<sup>1</sup> Joel R. Brownstein,<sup>5</sup> Donald P. Schneider<sup>6,7</sup> and Daniel Thomas<sup>3</sup>

<sup>1</sup>Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan

<sup>2</sup>Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, New York, NY 10003, USA

<sup>3</sup>Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK

<sup>4</sup>School of Physics and Astronomy, University of St Andrews, St Andrews KY16 9SS, UK

<sup>5</sup>Department of Physics and Astronomy, University of Utah, 115 S 1400 E, Salt Lake City, UT 84112, USA

<sup>6</sup>Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA

<sup>7</sup>Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA

Accepted 2016 January 12. Received 2016 January 5; in original form 2015 July 26

## ABSTRACT

The Baryon Oscillation Spectroscopic Survey (BOSS) has collected spectra for over one million galaxies at  $0.15 < z < 0.7$  over a volume of  $15.3 \text{ Gpc}^3$  ( $9376 \text{ deg}^2$ ) – providing us an opportunity to study the most massive galaxy populations with vanishing sample variance. However, BOSS samples are selected via complex colour cuts that are optimized for cosmology studies, not galaxy science. In this paper, we supplement BOSS samples with photometric redshifts from the Stripe 82 Massive Galaxy Catalog and measure the total galaxy stellar mass function (SMF) at  $z \sim 0.3$  and  $z \sim 0.55$ . With the total SMF in hand, we characterize the stellar mass completeness of BOSS samples. The high-redshift CMASS (constant mass) sample is significantly impacted by mass incompleteness and is 80 per cent complete at  $\log_{10}(M_*/M_\odot) > 11.6$  only in the narrow redshift range  $z = [0.51, 0.61]$ . The low-redshift LOWZ sample is 80 per cent complete at  $\log_{10}(M_*/M_\odot) > 11.6$  for  $z = [0.15, 0.43]$ . To construct mass complete samples at lower masses, spectroscopic samples need to be significantly supplemented by photometric redshifts. This work will enable future studies to better utilize the BOSS samples for galaxy-formation science.

**Key words:** galaxies: abundances – galaxies: evolution – galaxies: stellar content – cosmology: observations.

## 1 INTRODUCTION

There is tremendous interest in constraining the size, stellar mass, and halo mass evolution of the most massive galaxies in the universe (Tojeiro et al. 2012; Maraston et al. 2013; Tal et al. 2013; Beifiori et al. 2014; Marchesini et al. 2014; Marsan et al. 2015; van de Sande et al. 2015, to cite a few recent examples). The evolution of these properties places strong constraints on models of galaxy formation which traditionally have difficulty reproducing observed trends such as the amplitude of the stellar mass function (SMF) at the highest masses (e.g. Benson et al. 2003; Maraston et al. 2013;

Knebe et al. 2015), although much progress has been made in the past few years (e.g. Benson 2014; Furlong et al. 2015).

From an observational standpoint, spectroscopic samples of massive galaxies present key advantages over photometric samples. For example, errors in stellar mass estimates are reduced for spectroscopic samples compared to photometric redshift samples. Typical five-band photometric redshifts at  $z \sim 0.5$  have an error<sup>1</sup> of  $\sigma_z = 0.03$  to  $\sigma_z = 0.05$  even for the most massive galaxies (Bundy et al. 2015). This error translates into a stellar mass uncertainty of  $\sim 0.1$  dex which may dominate the total stellar mass error budget. Spectroscopic samples of massive galaxies are also key in order to perform accurate measurements of clustering and/or galaxy–galaxy

\*E-mail: alexie.leauthaud@ipmu.jp

<sup>1</sup> Here  $\sigma_z$  corresponds to the RMS scatter without a  $(1+z)$  normalization.

lensing which place tight constraints on the galaxy–halo connection (e.g. Mandelbaum et al. 2006; Leauthaud et al. 2012; Coupon et al. 2015; Zu & Mandelbaum 2015). Finally, the spectra themselves contain key information and can be used to constrain stellar ages, star formation histories (SFHs), dust extinctions and stellar velocity dispersions (e.g. Chen et al. 2012; Thomas et al. 2013).

For these reasons, from a galaxy-formation perspective, large samples of massive galaxies with spectroscopic redshifts over a wide redshift range are highly desirable. Surveys such as zCOSMOS (Lilly et al. 2007), VVDS (Le Fèvre et al. 2004, 2015), DEEP2 (Newman et al. 2013), PRIMUS (Coil et al. 2011), and VIPERS (Guzzo et al. 2014) provide spectroscopic samples that probe the  $0.2 < z < 1.0$  universe and complement the Sloan Digital Sky Survey (SDSS; York et al. 2000) main sample at  $z = 0.1$  (Strauss et al. 2002). These higher redshift surveys, however, still cover relatively small areas ranging from a few square degrees to a few tens of square degrees (e.g. VIPERS covers 24 deg<sup>2</sup>). The volumes probed by these surveys are insufficient to provide statistically significant samples of the most massive galaxies ( $\log(M_*/M_\odot) > 11.5$ ) which have low number densities ( $\bar{n} \sim 2 \times 10^{-5} \text{ Mpc}^{-3}$ ). For example, the highest mass bin at  $0.48 < z < 0.74$  in the Leauthaud et al. (2012) COSMOS analysis of the stellar-to-halo mass relation only contains 300 galaxies at  $\log(M_*/M_\odot) > 11.29$ , only 71 of which have masses greater than  $\log(M_*/M_\odot) > 11.5$ . The highest mass bin in the Coupon et al. (2015) analysis, which covers a wider area (23.1 deg<sup>2</sup>), contains 6326 galaxies at  $0.5 < z < 1.0$  and  $\log(M_*/M_\odot) > 11.2$ . However, this bin only contains 498 galaxies at  $\log(M_*/M_\odot) > 11.5$ , of which only 234 have a secure spectroscopic redshift from the VIPERS, VVDS, or PRIMUS surveys. Because of these small sample sizes, studies of the stellar-to-halo mass relation, for example, remain poorly constrained at the very high mass end.

An exciting opportunity is the Baryon Oscillation Spectroscopic Survey (BOSS; Eisenstein et al. 2011; Dawson et al. 2013), which, with the final DR12 data release (Alam et al. 2015), has collected spectra for more than one million massive galaxies ( $\log(M_*/M_\odot) > 11.0$ ) at  $0.15 < z < 0.7$  over a volume of 15.3 Gpc<sup>3</sup> (9376 deg<sup>2</sup>), providing the potential to study the most massive galaxy populations with vanishing sample variance. However, the sample selections of Baryon Acoustic Oscillations (BAO) surveys involve complex colour cuts that are optimized for cosmology studies, not galaxy science. While a number of BOSS galaxy studies have been published (Tojeiro et al. 2012; Guo et al. 2013; Maraston et al. 2013; Beifiori et al. 2014; Guo et al. 2014; Montero-Dorta et al. 2014, 2016; Reid et al. 2014; Guo et al. 2015, to cite a few), the stellar mass completeness of the BOSS samples remains poorly understood.

In this paper, the second in a series, we use a new compilation of wide-field survey data, the Stripe 82 Massive Galaxy Catalog (s82-MGC), to address this problem. Bundy et al. (2015, hereafter Paper I) describes the construction of s82-MGC, which matches the 2 mag deeper ‘SDSS Co-add’ optical photometry (Annis et al. 2014) in the equatorial Stripe 82 with the Large Area Survey (LAS) near-IR photometry from the United Kingdom Infrared Telescope Infrared Deep Sky Survey (UKIDSS; Lawrence et al. 2007). Supplementing with a variety of photometric redshifts, s82-MGC enables near-IR based stellar mass estimates for complete samples with  $\log(M_*/M_\odot) > 11.2$  and  $z < 0.7$ .

In this paper, we use the s82-MGC to investigate the stellar mass completeness of the two main BOSS spectroscopic samples, the LOWZ sample at  $0.15 < z < 0.43$  and the CMASS (constant mass) sample at  $0.43 < z < 0.7$ . Our characterization of the completeness

is made with respect to a measurement of the galaxy SMF using the s82-MGC. The evolution of the mass function, a detailed study of potential biases, and implications for galaxy growth are discussed in Bundy et al. (in preparation, hereafter Paper III). Here, we provide convenient fitting formulae which can be used to estimate the completeness of each of the BOSS samples as a function of stellar mass and redshift (Sections 6.1 and 6.2). Convenient files containing our completeness estimates can be found at [www.massivegalaxies.com](http://www.massivegalaxies.com).

This work is complementary to the analysis performed by Montero-Dorta et al. (2014) who studied the magnitude and colour completeness of the high-redshift BOSS CMASS sample. This paper focuses on stellar mass completeness for the full CMASS and LOWZ samples, an aspect that is not addressed in Montero-Dorta et al. (2014).

Upcoming surveys such as the Hyper Suprime Cam survey<sup>2</sup> (HSC) and the Euclid survey (Laureijs et al. 2011) will be able to use photometric redshifts to supplement spectroscopic samples and to construct mass-limited samples over wider redshift and mass ranges than using spectroscopic samples alone. We characterize the level to which spectroscopic samples need to be supplemented by photometric redshifts as a function of mass and redshift (see Section 3).

Finally, a better understanding of the BOSS selection functions will also enable the construction of improved mock catalogues that are critical to investigating the link between galaxies and their host haloes. Our companion paper (Saito et al. 2015) presents improved mock catalogues that account for the stellar mass completeness of the BOSS CMASS sample as a function of redshift.

The layout of this paper is as follows. Section 2 presents the data used in this paper. Section 3 discusses the characteristics of our sample. Section 4 broadly describes the effects of the BOSS sample selection. Section 5 presents our estimate of the total SMF as a function of redshift. Section 6 describes our completeness estimates for CMASS and LOWZ. Section 7 presents mass completeness estimates for several previous studies that used BOSS data. Finally, our summary and conclusions are presented in Section 8. We assume a  $\Lambda$  cold dark matter cosmology with  $\Omega_m = 0.274$ ,  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . Stellar mass is noted  $M_*$  and has been derived using a Chabrier initial mass function (IMF; Chabrier 2003).

## 2 DATA

### 2.1 Boss spectroscopic samples

BOSS is a spectroscopic survey of 1.5 million galaxies over 10 000 deg<sup>2</sup> that was conducted as part of the SDSS-III programme (Eisenstein et al. 2011) on the 2.5 m aperture Sloan Foundation Telescope at Apache Point Observatory (Gunn et al. 1998, 2006). A general overview of the BOSS survey can be found in Dawson et al. (2013), the BOSS spectrographs are described in Smee et al. (2013), and the BOSS pipeline is described in Bolton et al. (2012). BOSS galaxies were selected from Data Release 8 (DR8; Aihara et al. 2011) *ugriz* imaging (Fukugita et al. 1996) using a series of colour–magnitude cuts motivated by the Maraston et al. (2009) Luminous Red Galaxy (LRG) model. This is a passive template mainly dominated by a metal-rich population but also includes a small metal-poor population (3 per cent by mass) to mimic observed metallicity gradients of local massive elliptical galaxies. This template was shown to fit the observed-frame Sloan colours of 2-SLAQ

<sup>2</sup> <http://www.naoj.org/Projects/HSC/HSCProject.html>

galaxies at redshift 0.4–0.6 better than models which include star formation (Wake et al. 2006). This conclusion was also independently confirmed by Montero-Dorta et al. (2016). The choice of a passive template was motivated by the intention of selecting the most massive and passive galaxies for BAO studies.

The BOSS selection uses the following set of colour criteria:

$$c_{\parallel} = 0.7(g_{\text{mod}} - r_{\text{mod}}) + 1.2[(r_{\text{mod}} - i_{\text{mod}}) - 0.18] \quad (1)$$

$$c_{\perp} = (r_{\text{mod}} - i_{\text{mod}}) - (g_{\text{mod}} - r_{\text{mod}})/4 - 0.18 \quad (2)$$

$$d_{\perp} = (r_{\text{mod}} - i_{\text{mod}}) - (g_{\text{mod}} - r_{\text{mod}})/8.0. \quad (3)$$

The subscript ‘mod’ denotes model magnitudes, which are derived by adopting the better fitting luminosity profile between a de Vaucouleurs and an exponential luminosity profile in the  $r$  band (Stoughton et al. 2002). The subscript ‘cmod’ denotes composite model magnitudes, which are calculated from the best-fitting linear combination of a de Vaucouleurs and an exponential luminosity profile (Abazajian et al. 2004). Point spread function (PSF) magnitudes are denoted with the subscript ‘psf’. BOSS colour cuts are computed using model magnitudes, whereas magnitude cuts are computed using cmodel magnitudes. All magnitudes are corrected for Galactic extinction using the dust maps of Schlegel, Finkbeiner & Davis (1998).

BOSS targeted two primary galaxy samples: the LOWZ sample at  $0.15 < z < 0.43$  and the CMASS sample at  $0.43 < z < 0.7$ . The LOWZ sample is an extension of the SDSS I/II LRG sample (Eisenstein et al. 2001) to fainter magnitudes and is defined according to the following selection criteria:

$$|c_{\perp}| < 0.2 \quad (4)$$

$$r_{\text{cmod}} < 13.6 + c_{\parallel}/0.3 \quad (5)$$

$$16 < r_{\text{cmod}} < 19.6 \quad (6)$$

$$r_{\text{psf}} - r_{\text{cmod}} > 0.3. \quad (7)$$

Equation (4) sets the colour boundaries of the sample; equation (5) is a sliding magnitude cut which selects the brightest galaxies at each redshift; equation (6) corresponds to the bright; and faint limits and equation (7) is to separate galaxies from stars. In a similar fashion to the SDSS I/II LRG sample, the LOWZ selection primarily selects red galaxies. Over most of the BOSS footprint, roughly one third of the LOWZ sample has a spectrum from SDSS-II (these objects were not re-observed by BOSS). Galaxies with a spectrum that pre-dates the BOSS survey will be referred to interchangeably either as ‘Legacy objects’ or ‘SDSS KNOWN’.

The CMASS sample targets galaxies at higher redshifts with a surface density of roughly  $120 \text{ deg}^{-2}$ . CMASS targets are selected from SDSS DR8 imaging according to the following cuts:

$$|d_{\perp}| > 0.55 \quad (8)$$

$$i_{\text{cmod}} < 19.86 + 1.6(d_{\perp} - 0.8) \quad (9)$$

$$17.5 < i_{\text{cmod}} < 19.9 \quad (10)$$

$$r_{\text{mod}} - i_{\text{mod}} < 2 \quad (11)$$

$$i_{\text{fib2}} < 21.5, \quad (12)$$

where  $i_{\text{fib2}}$  is the estimated  $i$ -band magnitude in a 2 arcsec aperture diameter assuming 2 arcsec seeing. Star–galaxy separation on CMASS targets is performed via:

$$i_{\text{psf}} - i_{\text{mod}} > 0.2 + 0.2(20.0 - i_{\text{mod}}) \quad (13)$$

$$z_{\text{psf}} - z_{\text{mod}} > 9.125 - 0.46z_{\text{mod}}. \quad (14)$$

In addition to CMASS, BOSS also targeted a smaller ancillary sample known as ‘CMASS SPARSE’ that was designed to test the impact of the colour–magnitude cuts. The SPARSE sample includes fainter and bluer galaxies by extending the sliding cut to:

$$i_{\text{cmod}} < 20.14 + 1.6(d_{\perp} - 0.8). \quad (15)$$

The SPARSE region (contained between the fiducial CMASS sliding cut and the SPARSE sliding cut) is randomly down-sampled to yield about five objects per square degree. Beyond highlighting the SPARSE region in certain figures, we do not utilize this sample in this paper. Instead, our analysis relies on photometric redshifts to probe the full massive galaxy population that extends beyond the BOSS colour boundaries (see Section 2.2).

The CMASS sample was originally designed to loosely follow a constant stellar mass cut at  $0.4 < z < 0.6$  (see fig. 1 in Maraston et al. 2013) and to allow for a wider range of galaxy colours than either the SDSS-II LRG or the LOWZ samples. Using high-resolution HST/ACS imaging, Masters et al. (2011) show that roughly 26 per cent of the CMASS galaxies in the COSMOS survey are morphologically classified as late-types (with an observed colour of  $g - i < 2.35$ ). Using a maximum likelihood approach that accounts for photometric errors as well as the CMASS selection cuts, Montero-Dorta et al. (2014) estimate that 37 per cent of CMASS object may intrinsically belong to the blue cloud.

The CMASS sample is thought to be more complete at higher stellar masses than the SDSS-II LRG and LOWZ samples which are colour selected (Anderson et al. 2014). However, we will demonstrate in Section 6.2 that the opposite is in fact true and that in certain redshift ranges, the LOWZ sample is more complete in terms of stellar mass than CMASS (see Section 6.2).

In this paper, we use the internal DR10 BOSS data release (Ahn et al. 2014). However, our analysis is limited to the Stripe 82 region of the BOSS survey (described in Section 2.2) which was observed in the early phases of the BOSS survey. Because DR11 and DR12 provide no new data in Stripe 82, our analysis and conclusions would be identical using these later data releases. The main samples and catalogues used in this paper are as follows.

(i) The original catalogue that was used to target BOSS galaxies. From this catalogue, we extract the original set of LOWZ and CMASS targets. We will refer to the original target samples as ‘TAR\_LOWZ’ and ‘TAR\_CMASS’. Galaxies that were part of the target catalogue, but did not obtain a fibre due to fibre-collisions are designated as ‘FIBER\_COLLIDED\_LOWZ’ and ‘FIBER\_COLLIDED\_CMASS’.

(ii) The BOSS DR10 large-scale structure (LSS) catalogue created by the BOSS galaxy clustering working group that was used in the Anderson et al. (2014) analysis. We only give a brief description of this catalogue and refer the reader to Anderson et al. (2014) for additional details. This catalogue is constructed by trimming the BOSS redshift file (the SpALL file) by the BOSS mask. Legacy objects are added to the catalogue (Legacy objects are not contained in the SpALL file). Finally, Legacy objects are down-sampled in each sector to match the BOSS completeness. The LSS catalogue contains the flag ‘IMATCH’ which indicates galaxies with

BOSS redshifts (IMATCH = 1) and galaxies with Legacy redshifts (IMATCH = 2). Galaxies from the LSS catalogues will be referred to as ‘LSS\_LOWZ’ and ‘LSS\_CMASS’.<sup>3</sup>

(iii) The BOSS stellar mass catalogue from the Portsmouth group (Maraston et al. 2013).

Further details and the url’s of these catalogues are provided in the appendix.

## 2.2 The stripe 82 massive galaxy catalogue

A key data set for this paper comes from observations of the Stripe 82 region along the celestial equator in the region of the southern galactic sky – a narrow, but deeper subset of the SDSS survey area – for which it is possible to construct a galaxy sample with a well understood completeness function. Stripe 82 is critical for this paper for two reasons. First, it was the subject of repeat imaging campaigns in SDSS, especially by the SDSS Supernova Survey (Frieman et al. 2008). These data have been combined into the ‘SDSS Co-add’ by Annis et al. (2014) and reach roughly 2 mag deeper than the single epoch SDSS imaging with a  $5\sigma$  detection limit of  $r \sim 22.5$ . This added depth is critical for obtaining reliable photometric redshifts (photo- $z$ ’s) for massive galaxies that can be used to supplement the colour-selected BOSS samples out to  $z \sim 0.7$ . Secondly, this region was also imaged by the UKIDSS (Lawrence et al. 2007) which provides near-IR photometry for ensuring robust stellar mass estimates.

The s82-MGCC combines these data sets and delivers matched *ugrizYJHK* photometry using catalogue-level synthetic aperture photometry (Bundy et al. 2012). Galaxies are separated from stars following Baldry et al. (2010) and via a combination of light profile and colour cuts. Redshift information is provided by a number of sources including all SDSS spectroscopic redshifts available in DR10 and several photometric redshift estimates. Among these are neural-network-based photo- $z$ ’s from Reis et al. (2012) and iterative template-fitting estimates from the red-sequence Matched-filter Probabilistic Percolation (redMaPPer; Rykoff et al. 2014). The typical photo- $z$  precision is  $\sigma_z \simeq 0.03$  at  $z \leq 0.7$ .

A redshift estimator,  $z_{\text{best}}$ , is built for s82-MGC that employs spectroscopic redshifts whenever possible and chooses among three photometric-based redshift estimates otherwise. In order of priority,  $z_{\text{best}}$  is set to:

- (i)  $z_{\text{spec}}$ : spectroscopy from SDSS-II or from BOSS.
- (ii)  $z_\lambda$ : for cluster members.
- (iii)  $z_{\text{red}}$ : for field early-type galaxies.
- (iv)  $z_{\text{phot}}$ : photometric redshift from Reis et al. (2012).

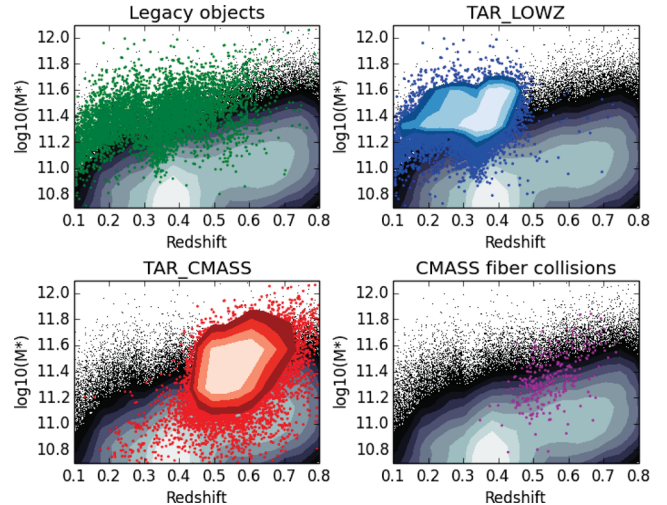
The scatter and bias in the photometric-based redshift estimators are characterized in Paper I.

MANGLE (Swanson et al. 2008) and CUSTOM software is used to apply geometric masking of regions with imaging artefacts in the SDSS Co-adds and UKIDSS imaging and areas around bright stars. The BOSS acceptance mask as well as rejection masks for collisions with the plate post and potentially higher priority BOSS targets are also applied.<sup>4</sup>

Fiducial stellar mass estimates for the s82-MGC are described in Paper I and compared to other publicly available estimates from

<sup>3</sup> Note that we do not apply the large-scale systematic weights that are applied to the BOSS clustering analyses.

<sup>4</sup> For details on BOSS masks, see: [http://www.sdss3.org/dr10/tutorials/lss\\_galaxy](http://www.sdss3.org/dr10/tutorials/lss_galaxy).



**Figure 1.** Mass and redshift distributions of several galaxy samples considered in this paper (indicated by coloured symbols). The underlying black points and grey contours show all galaxies from the s82-MGC as a function of mass and redshift. The s82-MGC catalogue extends to lower mass limits compared to the BOSS spectroscopic samples.

the BOSS survey. A more detailed examination of  $M_*$  estimators and potential biases is presented in Paper III. Briefly, the fiducial  $M_*$  estimates are the result of the Spectral Energy Distribution (SED)-fitting code described in Bundy et al. (2010) applied to the SDSS+UKIDSS PSF-matched photometry and the defined  $z_{\text{best}}$ . Stellar population templates are derived using BC03 models (Bruzual & Charlot 2003) and the STELIB empirical library (Le Borgne et al. 2003) and assuming a Chabrier (2003) IMF. For a prior grid of SED templates spanning a range of ages (restricted to be less than the age of the universe at the redshift of the galaxy), metallicities, dust extinction, and exponential  $\tau$  SFHs, a mass probability distribution is obtained by scaling the model  $M/L$  ratios by the inferred luminosity in the observed  $K$  band. The median of this distribution is taken as the stellar mass estimate.

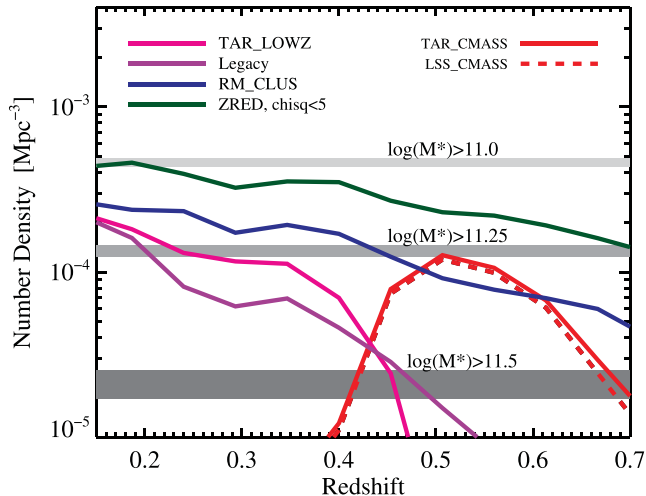
Starting from the s82-MGC parent catalogue, we select the ‘UK-WIDE’ subsample (see Paper I) by rejecting sources in masked regions and requiring all classified galaxies be observed on UKIDSS-LAS frames with  $5\sigma$  detection limits deeper than  $YJHK = [20.3, 20.0, 19.6, 19.5]$  on the AB system of Oke & Gunn (1983). The final UKWIDE sample spans an area of  $139.4 \text{ deg}^2$ .

## 3 SAMPLE CHARACTERISTICS

Fig. 1 displays the stellar mass and redshift distributions for the main galaxy samples considered in this paper compared to the full galaxy population from the s82-MGC. The s82-MGC catalogue extends to lower mass limits compared to the BOSS samples and is mass complete to  $\log_{10}(M_*/M_\odot) = 11.2$  at  $z = 0.7$  (Paper I).

For the LOWZ sample, because of a targeting error<sup>5</sup> related to star–galaxy separation in the early phase of the BOSS survey, a cut of  $\text{TILE} > 10324$  must be applied in order to select a uniform sample of LOWZ galaxies. However, the Stripe 82 region was unaffected by this error (see fig. 2 in Parejko et al. 2013) and this cut is unnecessary for the s82-MGC.

<sup>5</sup> See Parejko et al. (2013) and [http://www.sdss3.org/dr9/algorithms/boss\\_galaxy\\_ts.php](http://www.sdss3.org/dr9/algorithms/boss_galaxy_ts.php).



**Figure 2.** Number density as a function of redshift for several galaxy samples considered in this paper. The magenta line corresponds to LOWZ targets (‘TAR\_LOWZ’). The purple line corresponds to Legacy SDSS-II LRGs (‘Legacy’). The blue line corresponds to redMaPPer cluster members (‘RM\_CLUS’; Rykoff et al. 2014). The green line corresponds to galaxies with  $z_{\text{red}}$  and  $\chi^2_{\text{red}} < 5$  (‘ZRED’). The solid red line corresponds to CMASS targets (‘TAR\_CMASS’) and the dashed red line corresponds to CMASS galaxies from the LSS catalogue (‘LSS\_CMASS’). Grey horizontal lines indicate the number densities of galaxies with  $\log_{10}(M_*/M_\odot) > 11.0$ ,  $\log_{10}(M_*/M_\odot) > 11.25$ , and  $\log_{10}(M_*/M_\odot) > 11.5$ . These numbers were estimated from the s82-MGC assuming no redshift evolution. The redshift evolution of the galaxy SMF will be investigated in detail in Paper III.

Fig. 2 compares the number densities of various samples (e.g. Legacy, LOWZ, CMASS) as a function of redshift. As a reference, Fig. 2 also displays the number densities of galaxies with  $\log_{10}(M_*/M_\odot) > 11.0$ ,  $\log_{10}(M_*/M_\odot) > 11.25$ , and  $\log_{10}(M_*/M_\odot) > 11.5$ .

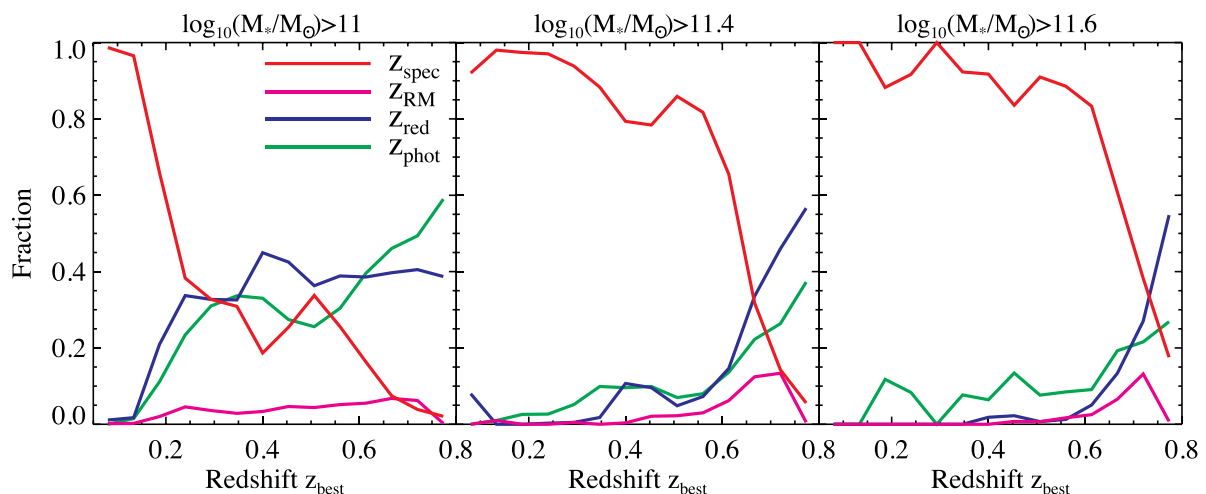
Large imaging surveys that will overlap with the BOSS footprint such as the HSC and Euclid surveys will be able to build mass-limited samples using a combination of spectroscopic redshifts and photometric-based estimators using a strategy similar to

the one we have used here. The ratio  $N_{\text{spec}}/N_{\text{phot}}$  represents a useful quantity when considering trade-offs between the mass limits of such samples and errors introduced by supplementing spectroscopic samples with photometric redshifts. Fig. 3 presents the origin of redshifts that contribute to  $z_{\text{best}}$  for three stellar mass thresholds. At  $z < 0.61$ , it is possible to construct mass-limited samples with  $\log_{10}(M_*/M_\odot) > 11.6$  for which more than 80 per cent of the sample has a spectroscopic redshift. At  $\log_{10}(M_*/M_\odot) > 11.0$ , however, spectroscopic samples must be significantly supplemented by photometric redshifts, except at low redshifts ( $z < 0.1$ ).

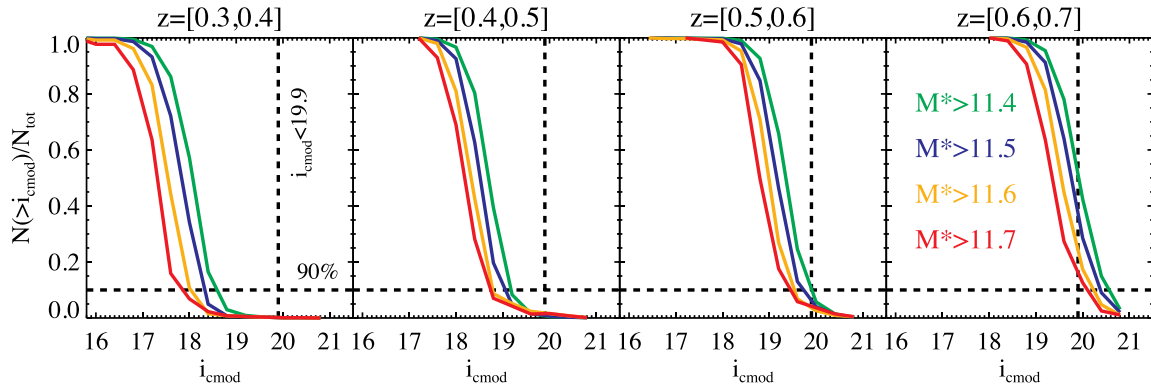
Finally, another question of interest is to determine how faint a spectroscopic survey needs to reach in order to probe a certain mass limit at a given redshift. Fig. 4 presents the cumulative  $i_{\text{cm}}^{\text{mod}}$  magnitude distribution of stellar mass threshold samples as a function of redshift. This figure demonstrates that galaxy samples with  $i_{\text{cm}}^{\text{mod}} < 20$  are roughly 90 per cent complete for  $\log_{10}(M_*/M_\odot) = 11.7$  at  $z = 0.65$ . For comparison, CMASS includes a cut at  $i_{\text{cm}}^{\text{mod}} < 19.9$  which impacts the completeness at higher redshifts. A more in-depth study of the impact of the CMASS cuts is provided in the following section.

#### 4 EFFECTS OF TARGET SELECTION ON BOSS GALAXY SAMPLES

We begin with a broad investigation of the impact of the BOSS target selection on the BOSS galaxy samples as a function of redshift, stellar mass, and colour. To represent galaxy colours, we use an estimate of the birth parameter,  $b_{1000}$ , which is the ratio of the average star formation rate within the previous 1 Gyr to the star formation rate averaged over the galaxy’s history. In this work, we use  $b_{1000}$  estimates from the KCORRECT package (Blanton & Roweis 2007) which are computed from an SED fit of a linear combination of stellar population templates from Bruzual & Charlot (2003, hereafter BC03). As for  $M_*$  estimates,  $b_{1000}$  depends on the assumed models and priors used to fit the observed SEDs. It is a rough estimate of recent star formation that is based on more than a single optical colour (such as  $g - r$ ) and which takes



**Figure 3.** Origin of redshifts that contribute to  $z_{\text{best}}$  as a function of stellar mass. The solid red line indicates the fraction of galaxies with spectroscopic redshifts and with  $\log_{10}(M_*/M_\odot) > 11$  (left-hand panel),  $\log_{10}(M_*/M_\odot) > 11.4$  (middle panel), and  $\log_{10}(M_*/M_\odot) > 11.6$  (right-hand panel) as a function of redshift. At  $z < 0.61$ , it is possible to construct mass-limited samples with  $\log_{10}(M_*/M_\odot) > 11.6$  for which more than 80 per cent of the sample has a spectroscopic redshift. At  $\log_{10}(M_*/M_\odot) > 11.0$ , however, spectroscopic samples must be significantly supplemented by photometric redshifts, except at low redshifts ( $z < 0.1$ ).



**Figure 4.** Cumulative magnitude distribution of stellar mass threshold samples as a function of redshift. Dashed horizontal lines show 90 per cent completeness limits. Dashed vertical lines correspond to  $i_{\text{cmod}} = 20$ . One of the cuts that defines the CMASS samples is a magnitude cut at  $i_{\text{cmod}} < 19.9$ . This figure demonstrates, for example, that galaxy samples with  $i_{\text{cmod}} < 20$  are 90 per cent complete for  $\log_{10}(M_*/M_\odot) = 11.7$  at  $z = 0.65$ .

advantage of the added constraints on dust afforded by the near-IR photometry.

In the s82-MGC, galaxies at  $z \sim 0.55$  and  $\log_{10}(M_*/M_\odot) \sim 11.5$  have  $b_{1000}$  values as high as 0.7, suggesting an occasional high rate of recent star formation, but the vast majority are peaked near  $b_{1000} = 0$  as expected for an old, passively evolving stellar population. Paper I explores in greater detail how  $b_{1000}$  better separates red-sequence galaxies with ongoing dusty star formation from those with truly passive populations.

Fig. 5 shows galaxies from the s82-MGC in the CMASS redshift range as a function of  $i_{\text{cmod}}$  and  $d_\perp$ . Because a number of galaxies in Fig. 5 are not contained in the original BOSS target catalogue,<sup>6</sup>  $i_{\text{cmod}}$  and  $d_\perp$  are taken from the deeper Stripe 82 co-add photometry (Annis et al. 2014). For CMASS galaxies, the typical RMS scatter between the co-add photometry and the shallower photometry of the target catalogue is of the order of  $\sigma_{\text{RMS}} = 0.08$  mag for  $d_\perp$  and  $\sigma_{\text{RMS}} = 0.14$  mag for  $i_{\text{cmod}}$ .

Fig. 5 clearly demonstrates that the CMASS sample is only complete in terms of mass and colour at the highest masses ( $\log_{10}(M_*/M_\odot) > 11.7$ ) and in a narrow redshift window at  $z \sim 0.6$ . The  $d_\perp$  cut mainly affects the sample selection at lower redshifts ( $z < 0.6$ ). The flux limit affects the sample selection at low stellar masses ( $\log_{10}(M_*/M_\odot) < 11.5$ ) and at higher redshifts ( $z > 0.6$ ) but is relatively unimportant for massive galaxies at low redshifts. Galaxies with high values of  $b_{1000}$  (suggesting recent star formation) are excluded from the sample at low redshift due to a combination of all three cuts. At higher redshifts ( $z > 0.6$ ), the sample is mainly flux-limited and includes a larger range of galaxy colours at fixed magnitude.

Fig. 6 presents  $r_{\text{cmod}}$  and  $c_\parallel$  in the LOWZ redshift range where  $r_{\text{cmod}}$  and  $c_\parallel$  are taken from the s82-MGC. For LOWZ galaxies, the typical RMS scatter between the co-add photometry and the shallower photometry of the target catalogue is of the order of  $\sigma_{\text{RMS}} = 0.07$  mag for  $c_\parallel$  and  $\sigma_{\text{RMS}} = 0.11$  mag for  $r_{\text{cmod}}$ . Fig. 6 shows that the LOWZ sample is roughly complete in terms of mass and colour at  $\log_{10}(M_*/M_\odot) > 11.6$  over the redshift range  $0.15 < z < 0.43$ . At lower stellar masses, the mass completeness is limited by the  $c_\parallel$  sliding cut. The flux limit mainly affects the sample selection at low stellar masses ( $\log_{10}(M_*/M_\odot) < 11.7$ ) and at higher redshifts ( $z > 0.3$ ).

<sup>6</sup> bosstarget-1rg-main007-collate.fits. See appendix.

## 5 TOTAL STELLAR MASS FUNCTIONS

To estimate the completeness of the BOSS samples, we first estimate the total SMF in four redshift bins. For a given mass bin and redshift, the stellar mass completeness of any given spectroscopic sample,  $A$ , will be estimated via the ratio  $c = \phi_A/\phi_{\text{tot}}$ . Here,  $\phi_A$  is the number density of sample  $A$  and  $\phi_{\text{tot}}$  is the number density of the total SMF.

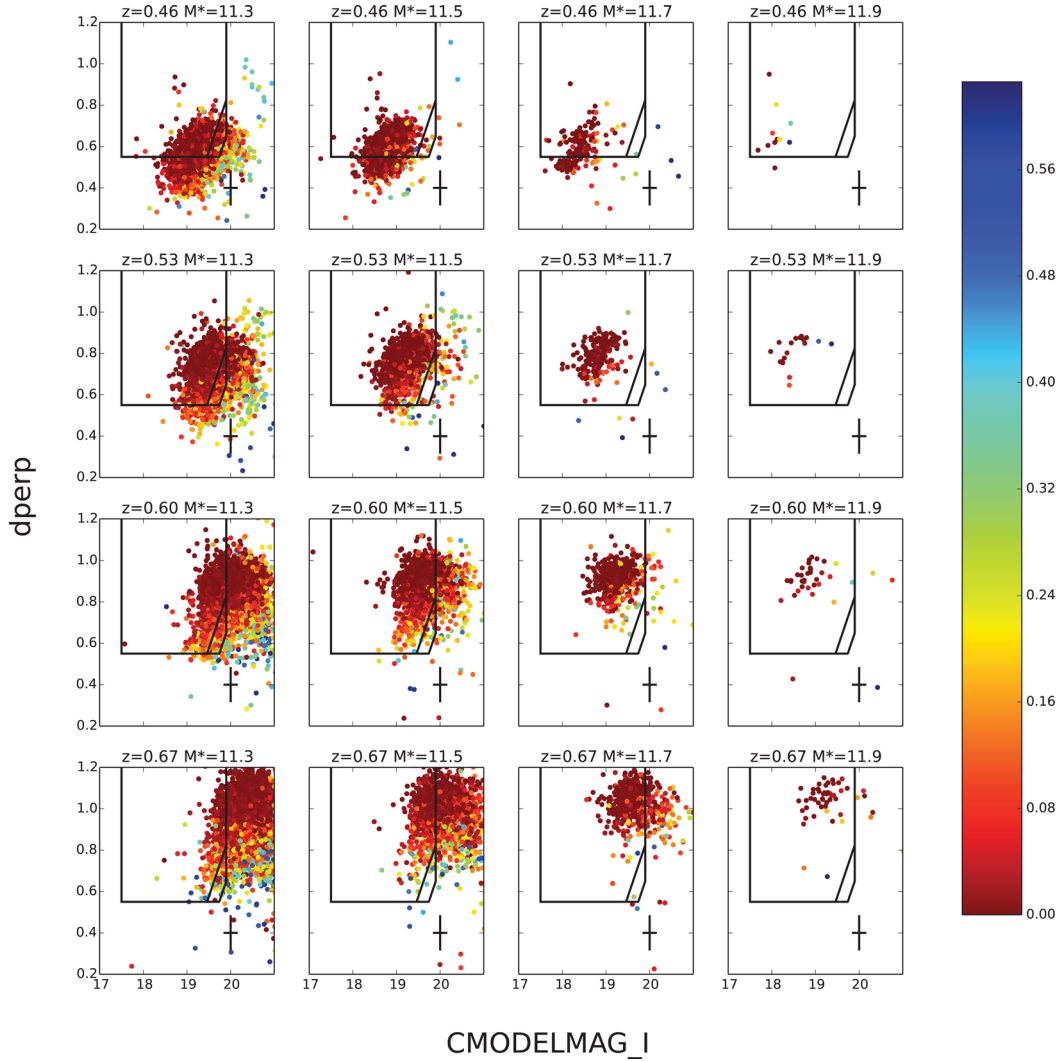
We construct redshift bins that correspond to a volume of 0.05–0.06 Gpc<sup>3</sup>. For the LOWZ sample, this requires a single large redshift bin,  $z_1 = [0.15, 0.43]$ . For CMASS, we divide the sample into three equal volume bins:  $z_2 = [0.43, 0.54]$ ,  $z_3 = [0.54, 0.63]$ , and  $z_4 = [0.63, 0.7]$ . The time span for these redshift bins is 2.6, 0.8, 0.6, and 0.4 Gyr, respectively. We take the redshift-binned mass functions to represent the galaxy distribution sampled at the mid-point of each bin. A more detailed investigation of the redshift evolution of the total SMF is presented and discussed in Paper III.

To construct the total SMF, we use  $z_{\text{best}}$  and  $M_{\text{best}}$  from the s82-MGC to measure the SMF at the high-mass end and data from PRIMUS (Moustakas et al. 2013) to evaluate the low-mass end. Because we use a parent sample that is complete above  $\log M_*/M_\odot = 11.2$  for  $z < 0.7$ ,  $V_{\text{max}}$  corrections are not required. Errors on the Stripe 82 SMFs are derived via bootstrap using 214 roughly equal-area bootstrap regions.

Fig. 7 shows mass functions from PRIMUS and from the s82-MGC over  $0.43 < z < 0.7$ . The figure demonstrates two key points. First, it is clear that mass functions from PRIMUS are insufficient to constrain the SMF at masses above  $\log_{10}(M_*/M_\odot) = 11.5$  – emphasizing the importance of the s82-MGC at these high masses. Secondly, as expected, the SMF of the s82-MGC UKWIDE sample itself begins to turn over due to incompleteness around  $\log_{10}(M_*/M_\odot) \sim 11.2$  at  $z \sim 0.7$ .

To obtain total SMFs for our subsequent analysis, we combine SMF measurements from the PRIMUS survey at  $\log_{10}(M_*/M_\odot) < 11.3$  with those from the s82-MGC at  $\log_{10}(M_*/M_\odot) > 11.3$ . For PRIMUS, we use the mass functions estimated by Moustakas et al. (2013) in the redshift range closest to our bin. The PRIMUS mass functions assume a Chabrier IMF, BC03 stellar population models, and have been adjusted to match our fiducial cosmology. Fig. 8 presents the total SMF in our four redshift bins from  $z = 0.15$  to  $z = 0.7$ .

Stellar mass offsets ( $\sim 0.1$  dex) may exist between our stellar masses and those from PRIMUS. However, PRIMUS did not compute stellar masses for their Stripe 82 field due to a lack



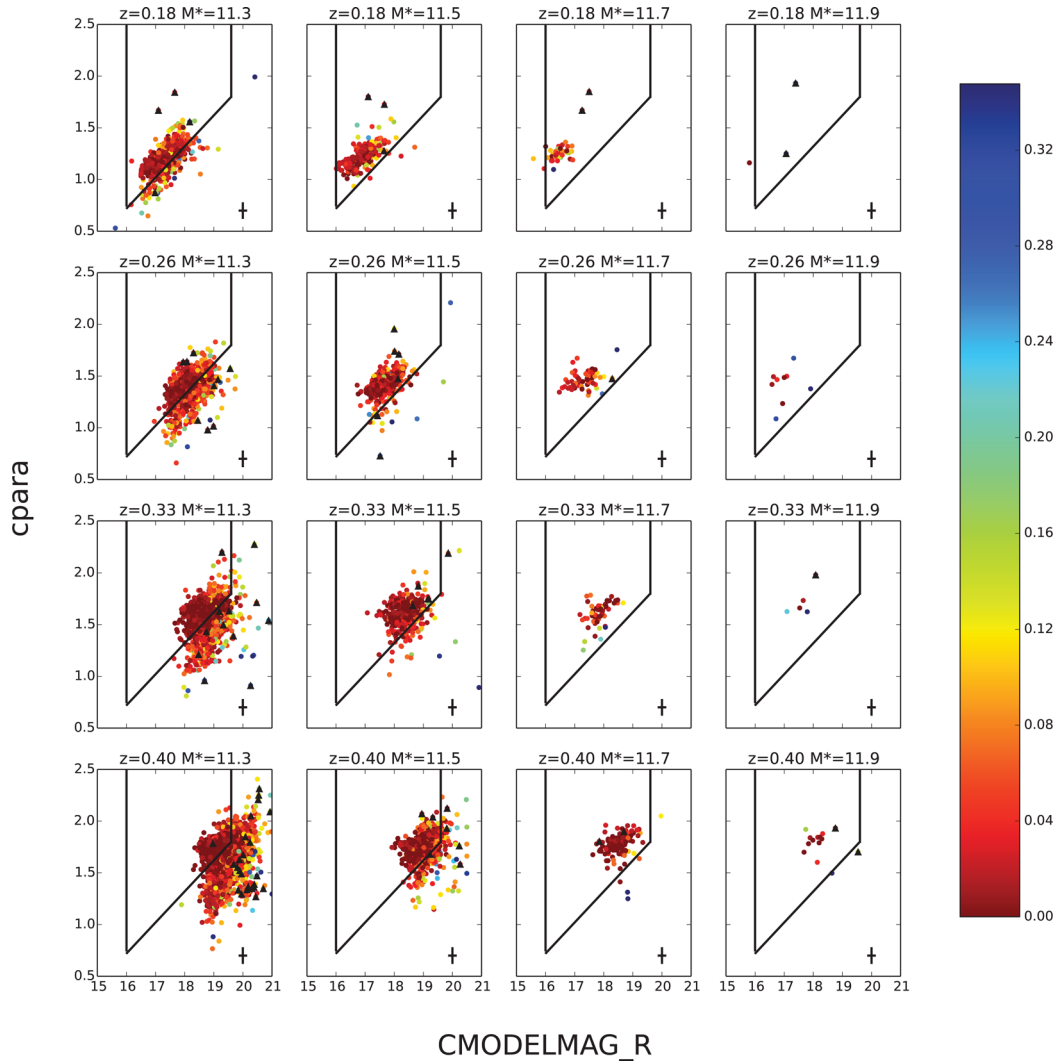
**Figure 5.** Galaxies from the s82-MGC in the CMASS redshift range as a function of  $i_{\text{cmod}}$  and  $d_{\perp}$  also (noted ‘dperp’). Galaxies are colour coded according to their birth parameter  $b_{1000}$ : larger values of  $b_{1000}$  (bluer colours) indicate galaxies that have experienced larger amounts of star formation in the past Gyr. Vertical black lines indicate the CMASS flux limits (equation 10). Horizontal lines indicate the CMASS  $d_{\perp}$  cut (equation 8). Tilted black lines indicate the fiducial sliding cut (equation 9) as well as the sliding cut that defines the sparsely sampled region (equation 15). The vertical error bar indicates the typical scatter between the target photometry and the co-add photometry. The horizontal error bar indicates the typical scatter for  $r_{\text{cmod}}$ . The stellar mass completeness of CMASS is due to the fact that the intrinsic colour distributions of galaxies at fixed mass often extend beyond the BOSS colour boundaries, as well as to scatter across these colour boundaries because of the shallower depth of the target photometry. The later effect is not present in this figure which is based on co-add photometry.

of *Spitzer* data. Hence, we cannot directly compare mass estimates between PRIMUS and the s82-MGC. Instead, to check if our completeness estimates are robust to such effects, we shift the PRIMUS data points by  $\pm 0.1$  dex and recompute the total SMF in each of our four redshift bins. Because we only use the PRIMUS data points at  $\log_{10}(M_*/M_{\odot}) < 11.3$  where the SMF has a relatively shallow slope, we find that 0.1 dex offsets only have a minor effect on our total SMF. Because the total SMF is unchanged, we find that our completeness estimates are robust to  $\pm 0.1$  dex shifts between PRIMUS stellar masses and those from the s82-MGC.

The s82-MGC  $z_{\text{best}}$  parameter uses a combination of spectroscopic redshifts and photometric redshifts. Because redshift errors translate into an error on the luminosity distance, galaxies that do not

have a spectroscopic redshift will have an extra error term for  $M_{\text{best}}$  which we denote as  $\sigma_{M^*-\text{zp}}$ . In the s82-MGC, this error term ranges from about  $\sigma_{M^*-\text{zp}} = 0.05$  dex to  $\sigma_{M^*-\text{zp}} = 0.1$  dex. One concern with our approach is that the inclusion of galaxies with photometric redshifts will cause an increase (due to Eddington bias) in the amplitude of the steep, high-mass end of the total SMF compared to a scenario in which all galaxies in our sample had a spectroscopic redshift. This behaviour would lead our completeness estimates to be underestimated because  $\phi_{\text{tot}}$  would be artificially inflated by this additional scatter term relative to the number densities obtained for a spectroscopic sample.

We account for this effect by forward modelling the SMF by convolving for  $M_*$  scatter induced by photometric redshift uncertainties. We assume an input functional form for the total SMF that



**Figure 6.** Galaxies from the s82-MGC in the LOWZ redshift range as a function of  $r_{\text{cmod}}$  and  $c_{\parallel}$ . Galaxies are colour coded according to their birth parameter  $b_{1000}$ : larger values of  $b_{1000}$  (bluer colours) indicate galaxies that have experienced larger amounts of star formation in the past Gyr. Vertical black lines indicate the LOWZ flux limits (equation 6). Tilted black lines represent the LOWZ sliding cut (equation 5). Black triangles indicate galaxies that are removed by the  $c_{\perp}$  cut (equation 4). The vertical error bar indicates the typical scatter between the target photometry and the co-add photometry. The horizontal error bar indicates the typical scatter for  $r_{\text{cmod}}$ .

follows a double Schechter function (Baldry, Glazebrook & Driver 2008):

$$\phi(M_*) = (\ln 10) \exp \left[ -\frac{M_*}{M_0} \right] \times \left\{ \phi_1 10^{(\alpha_1+1)(\log M_* - \log M_0)} + \phi_2 10^{(\alpha_2+1)(\log M_* - \log M_0)} \right\}, \quad (16)$$

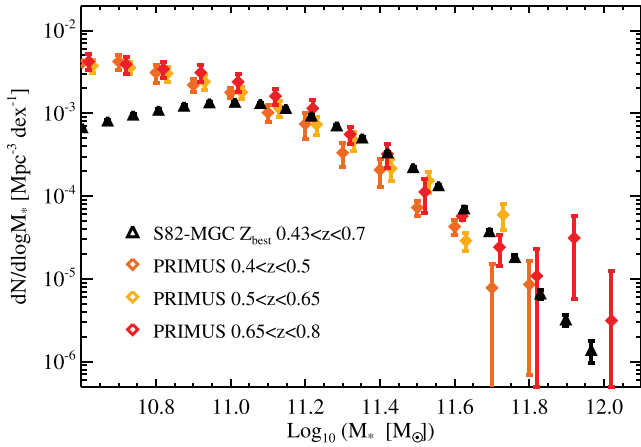
where  $\alpha_2 > \alpha_1$  and the second term dominates at the low-mass end. We generate Monte Carlo realizations of this function that sample various ranges for its parameters. The values of  $\phi_1$ ,  $\phi_2$ , and  $M_0$  are allowed to vary, while we fix  $\alpha_1 = -0.46$  and  $\alpha_2 = -1.58$  (Baldry et al. 2008). The goal of our double Schechter fits to the SMFs is simply to provide a compact and accurate representation of our data – not to provide physically meaningful values to parameters such as  $\phi_1$ ,  $\phi_2$ , and  $M_0$ .

A mock sample is drawn from each realization of the input SMF and additional scatter,  $\sigma_{M^*-z_p}$ , is added to the mock stellar masses of

this sample following the estimated  $\sigma_{M^*-z_p}(M_*, z)$  distribution from the s82-MGC. In our Monte Carlo mock realizations, the fraction of galaxies that have a spectroscopic redshift (i.e.,  $\sigma_{M^*-z_p} = 0$ ) as a function of mass and redshift is identical to the s82-MGC. In this exercise, we have not accounted for additional sources of scatter in  $M_*$ , namely from the mass estimates themselves or from uncertainties in total luminosities. Because these latter sources of errors are also present in the spectroscopic samples, they cancel in our completeness functions which divide spec-z SMFs by those supplemented with photo-zs.

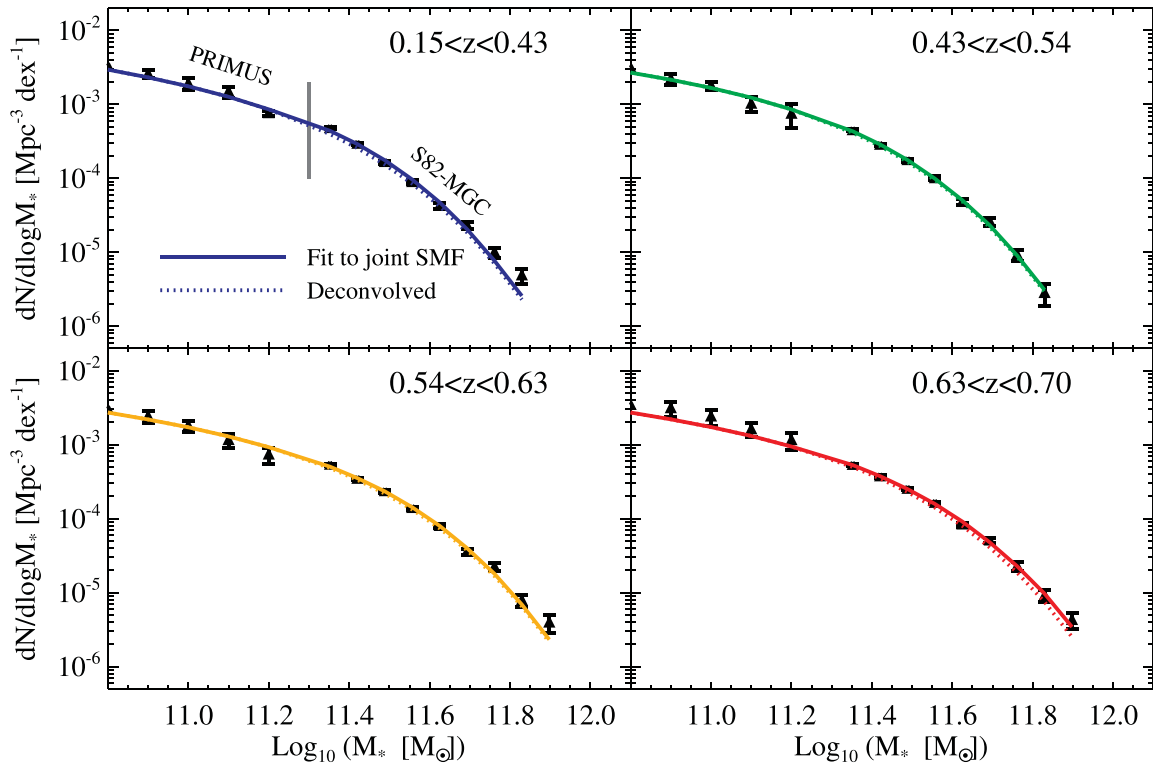
The results of fitting our mock samples to the observed SMF data are presented in Fig. 8. Because a large fraction of galaxies at the high-mass end have a spectroscopic redshift, the inclusion of galaxies with photometric redshifts only has a very minor impact on the SMF. The effects of  $\sigma_{M^*-z_p}$  are hence negligible for the s82-MGC. The best-fitting values for the double Schechter fits are given in Table 1. The statistical errors on our best-fitting Schechter parameters are quite small. However, at this level of precision, the errors on





**Figure 7.** Total SMF derived using  $z_{\text{best}}$  for  $0.43 < z < 0.7$ . Diamonds show the SMFs from PRIMUS in three redshift bins from  $z = 0.4$  to  $z = 0.8$ . Our Stripe 82 sample tightly constrains the high-mass end of the SMF ( $\log_{10}(M_*/M_\odot) > 11.5$ ) while PRIMUS constrains the low-mass end. The Stripe 82 sample is mass complete to  $\log_{10}(M_*/M_\odot) = 11.2$  at  $z = 0.7$ . A qualitative comparison of the SMFs at  $\log_{10}(M_*/M_\odot) > 11.3$  is suggestive of a  $\sim 0.1$  dex offset between PRIMUS and the s82-MGC. We have performed tests to check that our completeness estimates are robust to  $\sim 0.1$  dex offsets between PRIMUS and the s82-MGC.

the SMF are dominated by systematic errors in the determination of stellar mass estimates. The task of quantifying and providing systematic uncertainties is deferred to Paper III.



**Figure 8.** Total SMF in four redshift bins from  $z = 0.15$  to  $z = 0.7$ . To construct the total SMF, we combine data from the PRIMUS survey at  $\log_{10}(M_*/M_\odot) < 11.3$  with data from the s82-MGC at  $\log_{10}(M_*/M_\odot) > 11.3$ . Solid lines indicate our fit to the total SMF and dotted lines show the total SMF deconvolved for the effects of scatter introduced by photometric redshifts. Because a large fraction of galaxies at the high-mass end have a spectroscopic redshift ( $> 80$  per cent at the high-mass end), the inclusion of galaxies with photometric redshifts has a negligible impact on the SMF. Furthermore, we have tested that a shift in the PRIMUS data points of  $\pm 0.1$  dex only has a marginal impact on our determination of the total SMF.

## 6 STELLAR MASS COMPLETENESS OF BOSS SAMPLES

### 6.1 Stellar mass completeness of CMASS

With the total SMF now in hand, we derive the stellar mass completeness for the LSS\_CMASS sample. Completeness is estimated by comparing the total number of LSS\_CMASS galaxies in a given redshift and stellar mass bin to that derived from the total SMF. For the total SMF, we use the three redshift bins  $z_2$ ,  $z_3$ , and  $z_4$ . Fig. 9 presents our total SMF compared to target CMASS galaxies, fibre-collided galaxies, CMASS galaxies from the LSS catalogue, and Legacy objects, in our three fiducial redshift bins.

We further subdivide each of these redshift bins into two roughly equal volume bins to compute the completeness of CMASS in a total of six redshift bins. The CMASS SMFs for each of these six redshift bins are shown in Fig. 10. If the number of LSS\_CMASS galaxies fluctuates above the prediction based on the total SMF, the completeness is simply set to unity. The results are presented in Fig. 11 and the completeness values are given in Table 2.

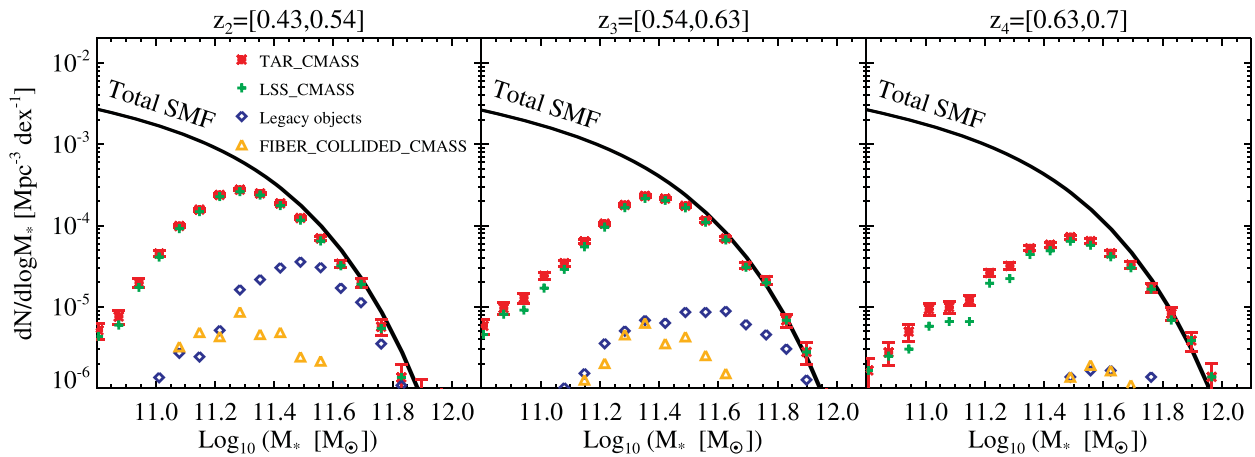
For convenience, we also fit the completeness with the following functional form:

$$c = \frac{f}{2} \left[ 1 + \operatorname{erf} \left( \frac{\log_{10}(M_*/M_1)}{\sigma} \right) \right], \quad (17)$$

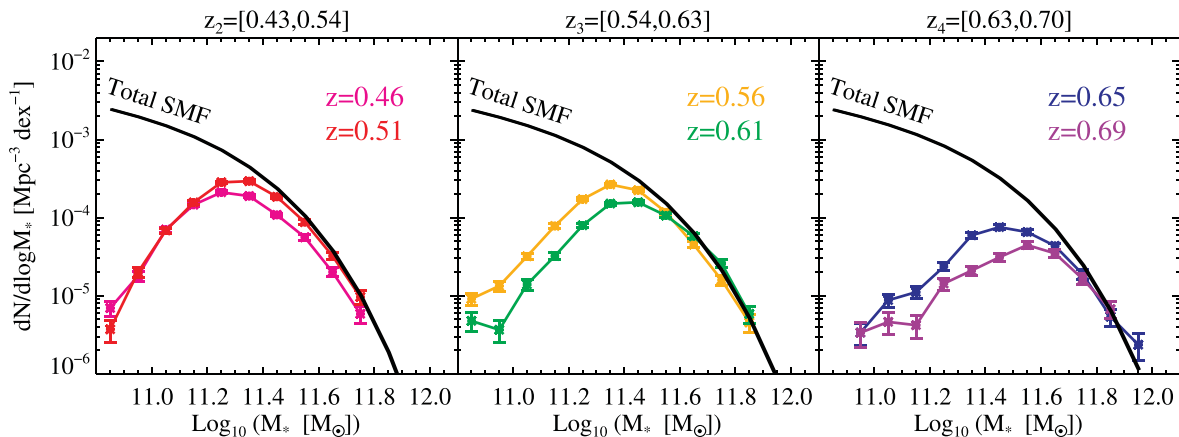
with free parameters  $f$ ,  $\sigma$ , and  $M_1$ . The results are shown on the right-hand side of Fig. 11 and the values of the fitted parameters are given in Table 3. Convenient files containing our completeness estimates can be found at [www.massivegalaxies.com](http://www.massivegalaxies.com).

**Table 1.** Parameters of the double Schechter fit to the total SMF as a function of redshift. Only the first three parameters are varied in the fit. These parameters correspond to the SMF after deconvolving for the effects of scatter due to the inclusion of a subset of galaxies with photometric redshifts. Errors correspond to statistical errors only. Please see Paper III for systematic uncertainty estimates.

Redshift	$\log_{10}(\phi_1/\text{Mpc}^{-3} \text{ dex}^{-1})$	$\log_{10}(\phi_2/\text{Mpc}^{-3} \text{ dex}^{-1})$	$\log_{10}(M_0/M_\odot)$	$\alpha_1$	$\alpha_2$
$z_1 = [0.15, 0.43]$	$-2.97 \pm 0.04$	$-2.79 \pm 0.03$	$10.910 \pm 0.008$	$-0.46$	$-1.58$
$z_2 = [0.43, 0.54]$	$-2.95 \pm 0.04$	$-2.89 \pm 0.04$	$10.922 \pm 0.009$	$-0.46$	$-1.58$
$z_3 = [0.54, 0.63]$	$-3.06 \pm 0.03$	$-2.91 \pm 0.03$	$10.986 \pm 0.007$	$-0.46$	$-1.58$
$z_4 = [0.63, 0.7]$	$-3.06 \pm 0.04$	$-2.92 \pm 0.04$	$10.995 \pm 0.008$	$-0.46$	$-1.58$



**Figure 9.** Total SMF compared to target CMASS galaxies (red stars), fibre-collided galaxies (orange triangles), CMASS galaxies from the LSS catalogue (green plus symbols), and Legacy objects (blue diamonds) in three different redshift bins. We subdivide each of these redshift bins into two roughly equal volume bins to compute the completeness of CMASS in a total of six redshift bins. At low redshifts, Legacy objects make up 50 per cent of galaxies at the high-mass end.

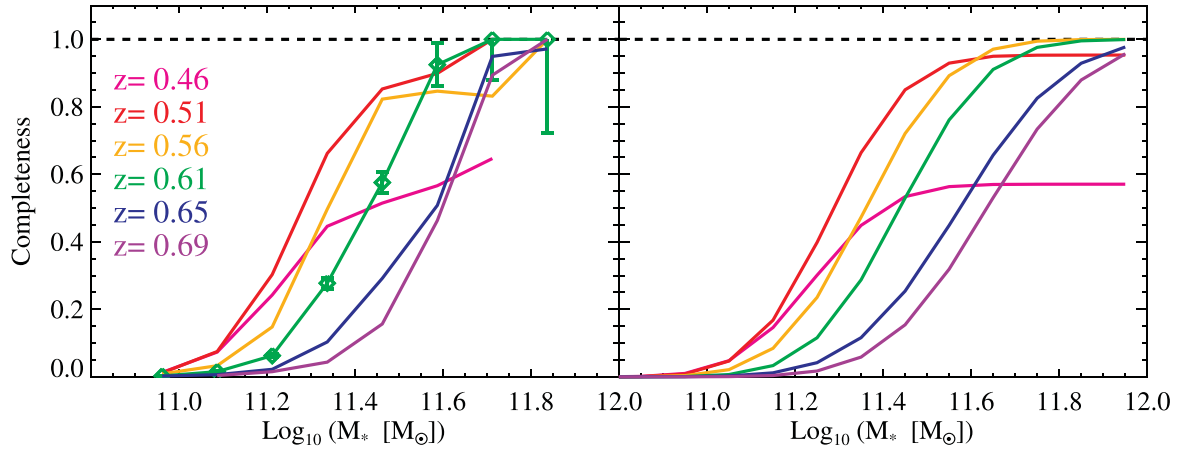


**Figure 10.** Stellar mass functions of the LSS\_CMSS sample in six redshift bins compared to the total SMF. Completeness is estimated via the ratio  $c = \phi_A/\phi_{\text{tot}}$  where  $\phi_A$  is the number density of CMASS in a given redshift bin and  $\phi_{\text{tot}}$  is the number density of the total SMF.

Fig. 11 demonstrates that CMASS is 80 per cent complete at  $\log_{10}(M_*/M_\odot) > 11.6$  in the narrow redshift range  $z = [0.51, 0.61]$ . At the mean redshift of the CMASS sample,  $\bar{z} = 0.55$ , CMASS is roughly 80 per cent complete at  $\log_{10}(M_*/M_\odot) = 11.4$ . For comparison, Maraston et al. (2013) conclude that BOSS is complete above  $\log_{10}(M_*/M_\odot) = 11.3$  at  $z < 0.6$  and at  $\log_{10}(M_*/M_\odot) = 11.6$  at  $z > 0.6$ . This work narrows this statement to  $z = [0.51, 0.61]$  and shows that the completeness decreases at lower and higher redshifts. Based on these considerations, referring to this sample in terms of CMASS should only be considered in loose terms.

## 6.2 Stellar mass completeness of LOWZ

We now proceed in a similar manner for the LOWZ sample in the redshift range  $0.15 < z < 0.43$ . For LOWZ, we use the total SMF estimated in the  $z_1 = [0.15, 0.43]$  redshift bin and shown in Fig. 8. Fig. 12 presents the SMFs of the LOWZ sample compared to the total SMF as a function of redshift. In a similar fashion as in the previous section, completeness is estimated by comparing the total number of LSS\_LOWZ galaxies in a given redshift and stellar mass bin compared to the expectation derived from the total SMF. The results are displayed in Fig. 13 and the completeness



**Figure 11.** Left: stellar mass completeness of the CMASS sample as a function of stellar mass and redshift. For clarity, errors are shown only for the  $z = 0.61$  redshift bin. Right: fits to the completeness using equation (17).

**Table 2.** Measured stellar mass completeness of the CMASS sample. Bins with less than 10 galaxies are marked with an ‘x’ symbol. The bottom line corresponds to the completeness estimates for the combined LOW+CMASS sample at  $0.38 < z < 0.48$ .

$z$	$\log_{10}(M_*/M_\odot)$ =11.09	$\log_{10}(M_*/M_\odot)$ =11.21	$\log_{10}(M_*/M_\odot)$ =11.34	$\log_{10}(M_*/M_\odot)$ =11.46	$\log_{10}(M_*/M_\odot)$ =11.59	$\log_{10}(M_*/M_\odot)$ =11.71	$\log_{10}(M_*/M_\odot)$ =11.84
0.46	$0.07 \pm 0.01$	$0.24 \pm 0.01$	$0.45 \pm 0.03$	$0.51 \pm 0.04$	$0.57 \pm 0.06$	$0.65 \pm 0.11$	x
0.51	$0.08 \pm 0.00$	$0.30 \pm 0.01$	$0.66 \pm 0.03$	$0.85 \pm 0.05$	$0.90 \pm 0.07$	$1.00 \pm 0.17$	x
0.56	$0.03 \pm 0.00$	$0.15 \pm 0.01$	$0.50 \pm 0.02$	$0.82 \pm 0.04$	$0.85 \pm 0.06$	$0.83 \pm 0.10$	$1.00 \pm 0.23$
0.61	$0.02 \pm 0.00$	$0.06 \pm 0.01$	$0.28 \pm 0.02$	$0.58 \pm 0.03$	$0.92 \pm 0.06$	$1.00 \pm 0.12$	$1.00 \pm 0.28$
0.65	$0.01 \pm 0.00$	$0.02 \pm 0.00$	$0.10 \pm 0.01$	$0.29 \pm 0.02$	$0.51 \pm 0.04$	$0.95 \pm 0.09$	$0.97 \pm 0.24$
0.69	$0.00 \pm 0.00$	$0.02 \pm 0.00$	$0.04 \pm 0.01$	$0.16 \pm 0.01$	$0.46 \pm 0.05$	$0.89 \pm 0.11$	$1.00 \pm 0.39$
comb	$0.05 \pm 0.00$	$0.16 \pm 0.01$	$0.42 \pm 0.02$	$0.68 \pm 0.03$	$0.92 \pm 0.06$	$1.00 \pm 0.13$	$1.00 \pm 0.41$

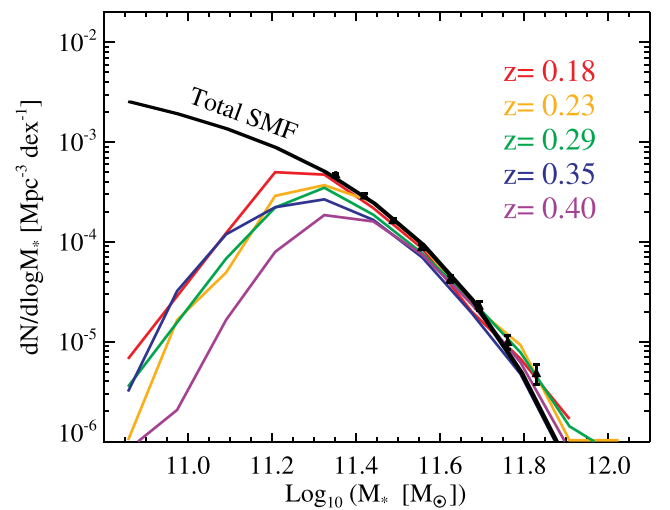
**Table 3.** Functional fits for completeness of the LSS\_CMASS sample. The bottom line corresponds to the combined LOWZ+CMASS sample.

Redshift	$f$	$\sigma$	$\log_{10}(M_1/M_\odot)$
0.46	0.57	0.20	11.24
0.51	0.95	0.20	11.28
0.56	1.00	0.22	11.36
0.61	1.00	0.22	11.44
0.65	1.00	0.27	11.57
0.69	1.00	0.26	11.64
comb	1.00	0.25	11.38

values are listed in Table 4. The values of the functional fits to the completeness are given in Table 5.

The errors on the LOW completeness are large at the high-mass end because the subvolumes used to measure the redshift dependence of the completeness are relatively small. As a result, for example, the best-fitting functional form does not converge to unity at large masses in the  $z = 0.35$  bin but does in the  $z = 0.29$  and  $0.40$  redshift bins. The values provided in this paper should be considered as *estimates* with relatively large errors at the high-mass end. Studies that are sensitive to these completeness estimates must take the reported errors in Table 4 into consideration.

Fig. 13 demonstrates that LOWZ is at least 80 per cent complete at  $\log_{10}(M_*/M_\odot) > 11.6$  over the entire redshift range and at least 90 per cent complete at  $\log_{10}(M_*/M_\odot) > 11.5$  in the redshift range  $z = [0.18, 0.29]$ . At the mean redshift of the LOWZ sample,  $\bar{z} = 0.29$ , LOWZ is roughly 80 per cent complete at  $\log_{10}(M_*/M_\odot) = 11.4$ . Interestingly, these results counter the conventional wisdom



**Figure 12.** Stellar mass functions of the LSS\_LOWZ sample in five redshift bins compared to the total SMF. Completeness is estimated via the ratio  $c = \phi_\Lambda / \phi_{\text{tot}}$ , where  $\phi_\Lambda$  is the number density of LOWZ in a given redshift bin and  $\phi_{\text{tot}}$  is the number density of the total SMF. In this redshift bin, our double Schechter fit slightly underestimates the total SMF at the high-mass end at  $\log_{10}(M_*/M_\odot) \sim 11.8$ .

that CMASS is more complete at higher stellar masses compared to LOWZ (e.g. Anderson et al. 2014).

Finally, although we do find a high stellar mass completeness for LOWZ, Hoshino et al. (2015) reported that the  $r_{\text{cmd}} > 16$  cut

**Table 4.** Measured stellar mass completeness of the LOWZ sample.

$z$	$\log_{10}(M_*/M_\odot)$ =11.09	$\log_{10}(M_*/M_\odot)$ =11.21	$\log_{10}(M_*/M_\odot)$ =11.34	$\log_{10}(M_*/M_\odot)$ =11.46	$\log_{10}(M_*/M_\odot)$ =11.59	$\log_{10}(M_*/M_\odot)$ =11.71
0.18	$0.10 \pm 0.01$	$0.69 \pm 0.04$	$1.00 \pm 0.07$	$1.00 \pm 0.10$	$0.96 \pm 0.15$	x
0.23	$0.04 \pm 0.01$	$0.39 \pm 0.03$	$0.89 \pm 0.07$	$1.00 \pm 0.10$	$1.00 \pm 0.16$	$1.00 \pm 0.35$
0.29	$0.05 \pm 0.01$	$0.29 \pm 0.02$	$0.82 \pm 0.05$	$0.87 \pm 0.07$	$1.00 \pm 0.12$	$1.00 \pm 0.26$
0.35	$0.09 \pm 0.01$	$0.29 \pm 0.02$	$0.62 \pm 0.04$	$0.82 \pm 0.06$	$0.76 \pm 0.09$	$0.94 \pm 0.18$
0.40	$0.01 \pm 0.00$	$0.10 \pm 0.01$	$0.46 \pm 0.03$	$0.79 \pm 0.05$	$0.98 \pm 0.10$	$1.00 \pm 0.19$

**Table 5.** Functional fits for completeness of the LSS\_LOWZ sample.

Redshift	$f$	$\sigma$	$\log_{10}(M_1/M_\odot)$
0.18	1.00	0.11	11.18
0.23	1.00	0.12	11.24
0.29	1.00	0.16	11.27
0.35	0.87	0.20	11.27
0.40	1.00	0.17	11.36

in equation (6) removes a small fraction ( $\sim 5$  per cent) of Brightest Cluster Galaxies from the nominal LOWZ sample.

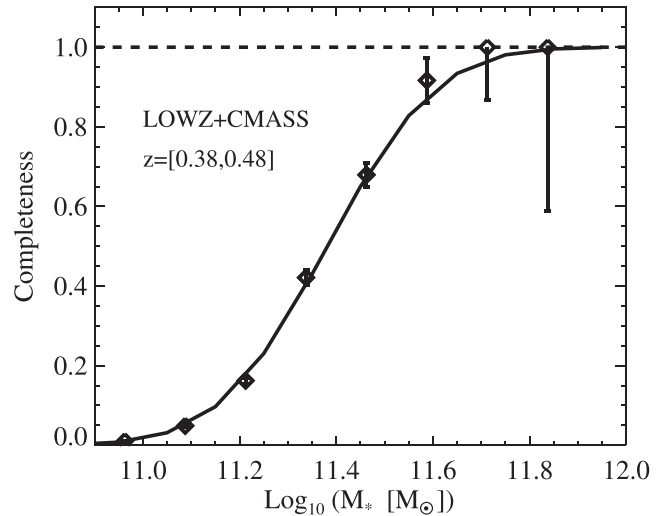
### 6.3 Combined sample

As can be seen from Fig. 2, CMASS and LOWZ overlap in the redshift range  $z \sim 0.38$  to  $z \sim 0.48$ . At these redshifts, it may be useful for certain studies to combine the two samples together. The stellar mass completeness of the combined LOWZ and CMASS samples in the redshift range  $0.38 < z < 0.48$  are presented in Fig. 14. Legacy objects are also included in this combined sample. The combined sample is roughly 80 per cent complete to  $\log_{10}(M_*/M_\odot) = 11.6$ . The completeness values for the combined samples are appended to the bottom of Tables 2 and 3.

In conclusion, the *combination* of LOW and CMASS (and Legacy objects) is 80 per cent complete to  $\log_{10}(M_*/M_\odot) = 11.6$  at  $z < 0.61$ .

## 7 SAMPLES FROM PREVIOUS STUDIES USING BOSS DATA

In this section, we evaluate the potential impact of stellar mass incompleteness for specific samples that have been used in previous

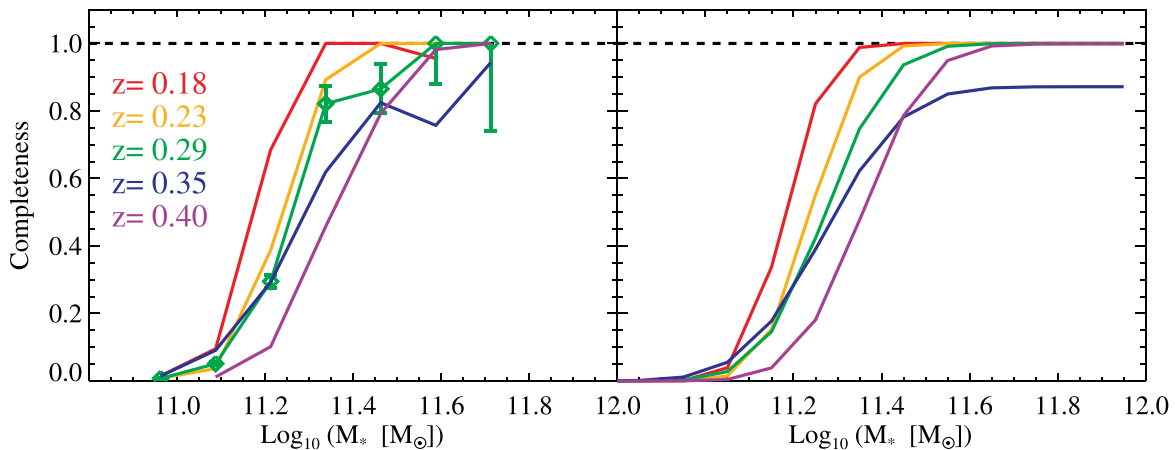


**Figure 14.** Stellar mass completeness of the combined LOWZ+CMASS sample in the redshift range  $0.38 < z < 0.48$ . Legacy objects are included in this combined sample. The solid line corresponds to the best fit using equation (17). The combined sample is roughly 80 per cent complete to  $\log_{10}(M_*/M_\odot) = 11.6$ .

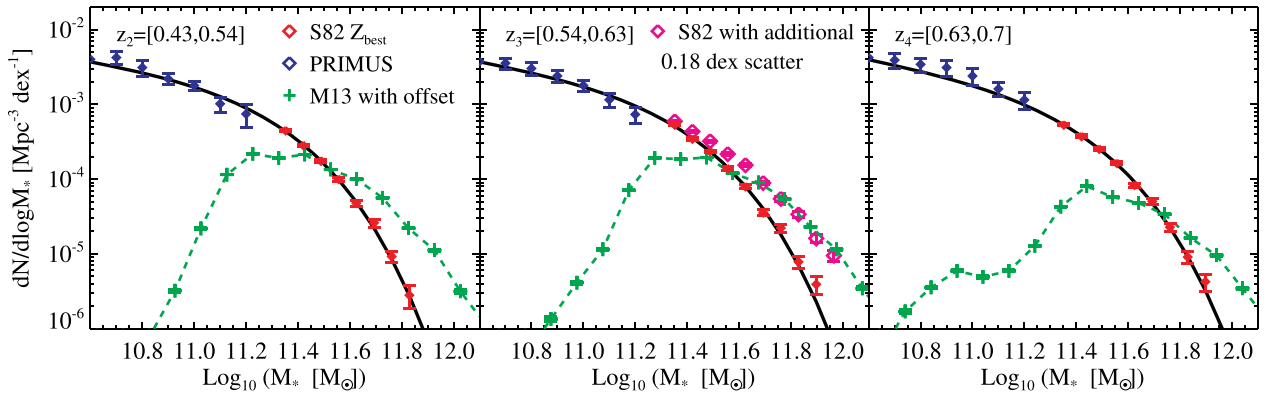
BOSS analyses. We focus on the samples used in Maraston et al. (2013), Shankar et al. (2014), Miyatake et al. (2015), and More et al. (2015).

### 7.1 CMASS mass functions from Maraston et al. (2013)

Maraston et al. (2013) computed stellar masses for BOSS galaxies using SED fits to the SDSS single epoch *ugriz* photometry.



**Figure 13.** Left: stellar mass completeness of the LOWZ sample as a function of stellar mass and redshift. For clarity, errors are shown only for the  $z = 0.29$  redshift bin. Right: fits to the completeness using equation (17).



**Figure 15.** Comparison between the total SMF derived in this work and the CMASS-only SMFs presented in Maraston et al. (2013). Green data points represent the Maraston et al. (2013) SMFs after applying our estimate of the mean offset between the Maraston et al. (2013) masses and s82-MGC masses for each redshift bin (these offsets are of the order of 0.1–0.15 dex). The Maraston et al. (2013) SMFs are computed for CMASS galaxies and do not include any completeness correction factors which explains the downturn at  $\log_{10}(M_*/M_\odot) < 11.4$ . However, there is also a noticeable difference between the Maraston et al. (2013) and the s82-MGC SMFs at the high-mass end. A possible explanation for this difference is that the mass estimates used by Maraston et al. (2013) have a larger scatter compared to the s82-MGC. The magenta diamonds show that adding 0.18 dex of scatter to the s82-MGC mass estimates brings the two SMFs into closer agreement.

SED fits were performed using the `HYERSPECZ` code (Bolzonella, Miralles & Pelló 2000) using the Maraston & Strömbäck (2011) stellar population libraries and assuming a Kroupa IMF (Kroupa 2001). Masses were computed for all BOSS galaxies in two separate runs: one using a passive template and one using a suite of star-forming templates with exponentially declining or truncated SFHs. The stellar masses of star-forming galaxies were adjusted upwards by 0.25 dex to account for a bias when fitting star-forming galaxies in which the best-fitting model may underestimate the total stellar mass because of fitting the brightest population (Maraston et al. 2010). This bias was identified using mock catalogues. For both templates, masses with and without mass-loss were computed in order to allow comparisons with the literature. These stellar masses are a standard output of the BOSS pipeline and are publicly released.

The two separate runs were combined into a single stellar mass catalogue by adopting the star-forming templates for BOSS galaxies with apparent  $g - i$  colours less than 2.35 and the passive template for galaxies with  $g - i > 2.35$  (the so-called Portsmouth Best Estimates). This matched the empirical morphological mix determined from the COSMOS field by Masters et al. (2011).

Maraston et al. (2013) computed the mass function for the CMASS sample in three redshift bins from  $z = 0.45$  to  $z = 0.7$  with the aim at constraining the assembly of the most massive galaxies in relation to galaxy formation models. Because the stellar mass completeness of the CMASS sample was not known at that time, Maraston et al. (2013) did not apply any completeness corrections to these CMASS SMF, opting instead to apply the BOSS selection cuts to semi-analytic models (SAMs) when comparing with the theoretical mass function from galaxy formation models. Maraston et al. (2013) find a deficit of massive galaxies ( $\log(M_*/M_\odot) > 11.3$  for a Kroupa IMF) between the BOSS data and SAMs over the redshift range 0.45–0.6.

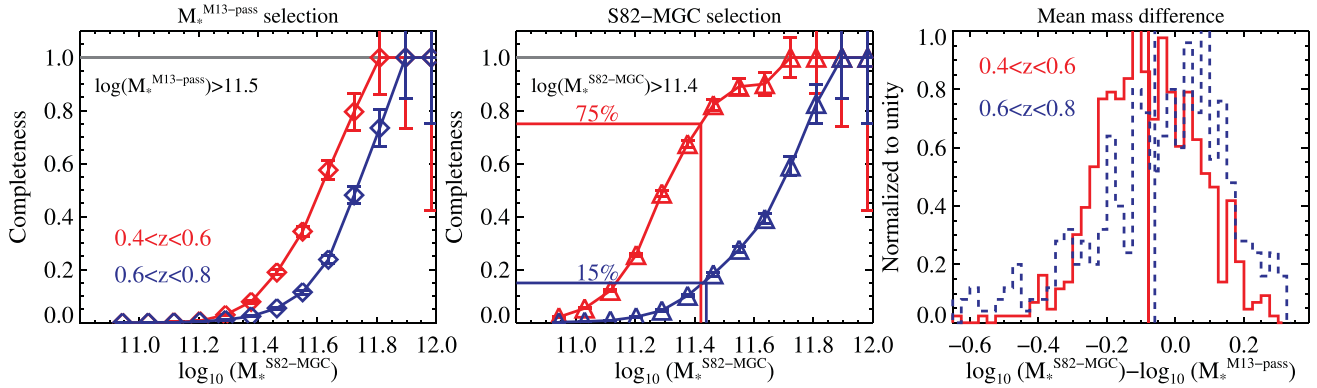
Paper I presents an in-depth comparison between the mass estimates from Maraston et al. (2013) and those from the s82-MGC. For our three fiducial redshift bins, there is a mean offset between the two stellar mass estimates, denoted  $\delta = \log_{10}(M_*^{\text{S82-MGC}}) - \log_{10}(M_*^{\text{M13}})$ . In the range

$11.3 < \log_{10}(M_*^{\text{S82-MGC}}) < 11.6$ ,  $\delta = 0.11$  dex for  $z_2$ ,  $\delta = 0.14$  dex for  $z_3$ ,  $\delta = 0.10$  dex for  $z_4$ .

With a better understanding of the CMASS sample now in hand, we re-investigate the mass completeness of these CMASS SMFs from this early BOSS analysis. Fig. 15 presents a comparison between the total SMF derived in this paper with the CMASS SMFs derived in Maraston et al. (2013, adjusted to our fiducial cosmology). Green data points indicate the Maraston et al. (2013) mass functions after applying the mean offsets for each redshift bin.

We find a lower amplitude in our total SMF at the high-mass end compared to Maraston et al. (2013). A possible explanation for this difference is that the Maraston et al. (2013) mass estimates have a larger scatter compared to the s82-MGC, which are based on deeper optical and NIR photometry. A larger mass error would cause the Maraston et al. (2013) SMFs to be inflated at the high-mass end relative to those computed from the s82-MGC (due to Eddington bias). Fig. 14 in Paper I supports this argument and shows that, compared to other BOSS mass estimates, those employed in Maraston et al. (2013) present a larger scatter (with an average of  $\sigma \sim 0.3$  dex) with respect to the stellar masses of the s82-MGC. This scatter is bimodal and varies with mass and redshift. To check if indeed Eddington bias is a plausible explanation for the differences in the SMFs, we show in Fig. 15 that adding 0.18 dex of scatter to the s82-MGC stellar mass estimates brings the s82-MGC SMF into agreement with Maraston et al. (2013). While this simple test does not demonstrate that the two SMFs are consistent, it does prove that a plausible level of additional scatter can reconcile the difference. A more in-depth analysis would need to take the complex structure of the scatter shown in Paper I into account.

By qualitatively comparing the CMASS mass functions with previous analyses, Maraston et al. (2013) provided a first estimate of the completeness of the CMASS sample, tentatively concluding a rough mass completeness of  $\log_{10}(M_*/M_\odot) = 11.3$  at  $z < 0.6$  and at  $\log_{10}(M_*/M_\odot) = 11.6$  at  $z > 0.6$ . The analysis presented here suggests that these original estimates were reasonable, although somewhat optimistic. A more accurate estimate is presented in Table 2.



**Figure 16.** Stellar mass completeness of the CMASS samples used in the Shankar et al. (2014) analysis. Red diamonds correspond to the low-redshift sample and blue diamonds indicate the high-redshift sample. The  $x$ -axis represents the mass estimate from the s82-MGC. Left-hand panel: mass completeness evaluated using the actual Shankar et al. (2014) sample but relative to s82-MGC masses. This panel shows additional spread compared to a pure threshold sample due to scatter between the Maraston et al. (2013) passive template masses and the s82-MGC masses. Middle panel: mass completeness evaluated using a fixed s82-MGC stellar mass cut. Right-hand panel: distribution of mass differences between the Maraston et al. (2013) passive template masses and the s82-MGC masses. Vertical lines indicate the 50th percentile of the distribution. A mass cut of  $\log_{10}(M_*^{M13-pass}/M_\odot) = 11.5$  corresponds to a mass cut of  $\log_{10}(M_*/M_\odot) \sim 11.43$  in the s82-MGC.

## 7.2 Halo occupation modelling from Shankar et al. (2014)

Shankar et al. (2014) recently studied the high-mass slope and scatter of the stellar-to-halo mass relation to  $z = 1$  using a variety of different data sets. One component of their analysis is based on Halo Occupation Distribution (HOD; Berlind & Weinberg 2002, and references therein) modelling of the projected two-point correlation function of CMASS galaxies. This clustering analysis was performed in two redshift bins:  $0.4 < z < 0.6$  and  $0.6 < z < 0.8$ . A stellar mass cut of  $\log_{10}(M_*/M_\odot) > 11.5$  was applied to both samples. The Maraston et al. (2013) passive template mass catalogue was adopted for this analysis (Guo, private communication).<sup>7</sup>

One underlying assumption of the HOD-type model employed by Shankar et al. (2014) is that stellar mass selected threshold samples should be mass complete. In this case, the central occupation function of central galaxies is generally expected to be well described by a traditional *erf* function that converges to unity at large masses. If the samples under consideration are incomplete in mass, then the amplitude and form of the central occupation function is uncertain.

With this possibility in mind, we evaluate the mass completeness of the two samples employed by Shankar et al. (2014) using the same methodology as in Section 6.1. To estimate the completeness, we recompute the total SMF within each of the two Shankar et al. (2014) redshift bins and show our results in Fig. 16.

There are two aspects of this analysis worth emphasizing. First, in practice, all stellar mass estimates are imperfect and will have scatter relative to the true underlying stellar mass. Hence, any stellar mass ‘threshold’ yields a sample with a more smoothly varying selection function on the true  $M_*$  distribution. We can roughly estimate the magnitude of this effect by assuming that the scatter between the Maraston et al. (2013) passive template masses and the s82-MGC masses is similar to the scatter between either estimate and the true underlying stellar mass. This effect will be captured if we use the actual Shankar et al. (2014) sample (i.e., defined by a selection with respect to the Maraston et al. 2013 masses) but evaluate the completeness using s82-MGC masses. This is shown in the left-hand side of Fig. 16. The samples no longer have a sharp boundary in

mass but instead follow a more smoothly varying completeness function.

Secondly, we can evaluate the completeness of a threshold cut on  $M_*$  defined for one set of  $M_*$  estimates by translating those thresholds into  $M_*$  used in this work. The Shankar et al. (2014) sample, for example, was selected as  $\log_{10}(M_*^{M13-pass}/M_\odot) > 11.5$ . Comparing to the Maraston et al. (2013) passive template masses used in that work, we find a mean offset<sup>8</sup> of  $\delta = -0.06$  dex and  $\delta = -0.06$  for the low- and high-redshift bins, respectively, compared to our masses from s82-MGC. Accounting for this mean offset, the fixed mass cut of  $\log_{10}(M_*^{M13-pass}/M_\odot) > 11.5$  used by Shankar et al. (2014) corresponds to cuts of  $\log_{10}(M_*/M_\odot) > 11.43$  (low- $z$  bin) and  $\log_{10}(M_*/M_\odot) > 11.44$  (high- $z$  bin) using our  $M_*$  estimates. At these limits, CMASS is 75 per cent complete for the low-redshift sample and only 15 per cent in the high-redshift sample (middle panel of Fig. 16).

Finally, the right-hand panel of Fig. 16 displays the difference between the Maraston et al. (2013) passive template masses and the s82-MGC masses at these scales. Both the offset and relative scatter are apparent.

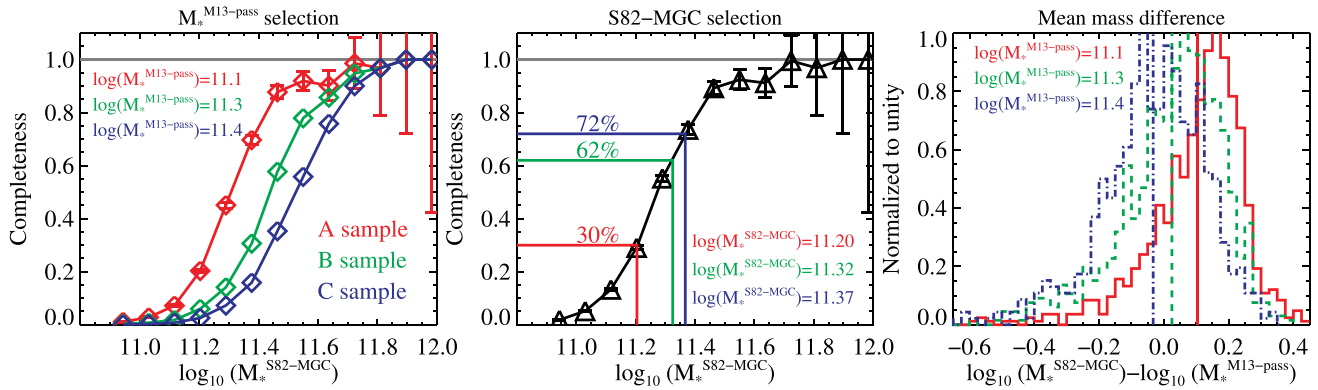
This exercise highlights the stellar mass incompleteness of these samples and demonstrates that that caution needs to be taken when studying CMASS. Understanding to what extent this may or may not have an impact on the original Shankar et al. (2014) analysis is beyond the scope of this paper.

## 7.3 Cosmological Analysis from Miyatake et al. (2015) and More et al. (2015)

In two companion papers, Miyatake et al. (2015) and More et al. (2015) present a joint analysis of the abundance, clustering, and galaxy–galaxy lensing signal measured for CMASS subsamples. Using a HOD framework, they derive constraints on the high-mass end of the stellar-to-halo mass relation and on the cosmological parameters  $\Omega_m$  and  $\sigma_8$ . Miyatake et al. (2015) and More et al. (2015) consider three CMASS samples selected in the range

<sup>7</sup> As opposed to creating a merged catalogue using both the passive and the star-forming templates as in Maraston et al. (2013).

<sup>8</sup> Mass offsets will differ from those quoted in Section 7.1 because in one case a merged (star-forming plus passive) catalogue was created, and in the other the passive template was adopted for all galaxies.



**Figure 17.** Stellar mass completeness of the CMASS samples employed by Miyatake et al. (2015) and More et al. (2015). The x-axis represents the mass estimate from the s82-MGC. Left-hand panel: mass completeness evaluated using the Miyatake et al. (2015) and More et al. (2015) samples but relative to s82-MGC masses. This panel shows additional spread compared to a pure threshold sample due to scatter between the Maraston et al. (2013) passive template masses and the s82-MGC masses. Middle panel: mass completeness evaluated using a fixed s82-MGC stellar mass cut. Right-hand panel: distribution of mass differences between the Maraston et al. (2013) passive template masses and the s82-MGC masses. Vertical lines indicate the 50th percentile of the distribution. A mass cut of  $\log_{10}(M_*^{M13-pass}/M_{\odot}) = 11.1$  corresponds to a mass cut of  $\log_{10}(M_*/M_{\odot}) \sim 11.2$  in the s82-MGC.

$0.47 < z < 0.59$ : ‘Sample A’ with  $\log_{10}(M_*/M_{\odot}) > 11.1$  (‘cut 1’), ‘Sample B’ with  $\log_{10}(M_*/M_{\odot}) > 11.3$  (‘cut 2’), and ‘Sample C’ with  $\log_{10}(M_*/M_{\odot}) > 11.4$  (‘cut 3’). The fiducial sample in the More et al. (2015) cosmological analysis is the A sample. Stellar masses are taken from the Maraston et al. (2013) passive template catalogue and assume a Kroupa IMF.

Using the same methodology as in the previous section, we compute the mass completeness for these three samples. The results are presented in Fig. 17. Sample A is 30 per cent complete at cut 1, sample B is 62 per cent complete at cut 2 and sample C is 72 per cent complete at cut 3. The right-hand panel shows the incompleteness function of these samples evaluated using the original cuts. Again, the incompleteness of the two samples relative to the true underlying mass probably lies between the left-hand and the middle panel of Fig. 17.

The More et al. (2015) HOD analysis accounts for potential incompleteness in the selection of CMASS galaxies at the low-mass end when compared to a true stellar mass threshold sample. However, the difficulty when working with incomplete samples is that galaxies that are excluded from the sample are not a random population. For example, the More et al. (2015) incompleteness model assumes that the CMASS selection corresponds to a random selection at fixed stellar and halo mass, i.e. it assumes that the galaxies that are removed from the sample live in similar halo environments as CMASS galaxies at fixed stellar mass. However, as shown in Fig. 5, CMASS is colour selected and preferentially selects certain regions of colour space at fixed stellar mass. In our companion paper, we demonstrate that a model which accounts for the stellar mass completeness of the CMASS sample, and reproduces the SMF and the two point correlation functions, but assumes that colour is uncorrelated with environment at fixed stellar mass, fails to reproduce the monopole and the quadrupole of the correlation function (Saito et al. 2015). Hence, the modelling of incomplete and colour-selected samples may not be straightforward.

Again, a full evaluation of the impact of incompleteness on the conclusions of Miyatake et al. (2015) and More et al. (2015) is beyond the scope of this paper. The better characterization of the mass and colour completeness of the BOSS samples presented here will make such evaluations possible and improve future attempts to accurately model the BOSS samples.

## 8 SUMMARY AND CONCLUSIONS

The BOSS survey has collected spectra for over one million galaxies at  $0.15 < z < 0.7$  over a volume of  $15.3 \text{ Gpc}^3$  ( $9376 \text{ deg}^2$ ) which provides an opportunity to study the most massive galaxy populations with vanishing sample variance. However, the BOSS sample selections involve complex colour cuts which are not necessarily optimized for galaxy science. As a result, the selection function and stellar mass completeness of these samples are poorly understood. None the less, given the large volumes and consequently large sample sizes at play, these surveys have a tremendous potential to constrain the galaxy–halo connection and to investigate the most massive galaxies in the universe providing that the samples are well understood.

In this paper, we characterize the stellar mass completeness of the BOSS samples with the goal of enabling future studies to better utilize these samples for galaxy-formation science. We use data from Stripe 82, which is roughly 2 mag deeper than the single epoch SDSS imaging, and construct a catalogue of massive galaxies that is complete to  $\log_{10}(M_*/M_{\odot}) > 11.2$  at  $z = 0.7$ . Using this catalogue, we empirically derive the stellar mass completeness of the two main BOSS spectroscopic samples: the LOWZ sample at  $0.15 < z < 0.43$  and the CMASS sample at  $0.43 < z < 0.7$ . We provide convenient fitting formulas which can be used to estimate the completeness of each of these samples as a function of stellar mass and redshift.

We demonstrate that CMASS is significantly impacted by mass incompleteness and is 80 per cent complete at  $\log_{10}(M_*/M_{\odot}) > 11.6$  only in the narrow redshift range  $z = [0.51, 0.61]$ . At the mean redshift of the CMASS sample,  $\bar{z} = 0.55$ , CMASS is roughly 80 per cent complete at  $\log_{10}(M_*/M_{\odot}) = 11.4$ . Based on these considerations, referring to this sample in terms of CMASS should only be considered in loose terms. In contrast, we demonstrate that LOWZ is 80 per cent complete at  $\log_{10}(M_*/M_{\odot}) > 11.6$  over the entire redshift range and 90 per cent complete at  $\log_{10}(M_*/M_{\odot}) > 11.5$  in the redshift range  $z = [0.18, 0.29]$ . At the mean redshift of the LOWZ sample,  $\bar{z} = 0.29$ , LOWZ is 80 per cent complete at  $\log_{10}(M_*/M_{\odot}) = 11.4$ . Hence, our results counter the conventional notion that CMASS is more complete at higher stellar masses compared to LOWZ (Anderson et al. 2014).

Our results suggest an interesting redshift window for studying the evolution of the most massive galaxies. The combination of LOW and CMASS yields a spectroscopic sample that is 80 per cent complete at  $\log_{10}(M_*/M_\odot) > 11.6$  at  $z < 0.61$ .

The values provided in this paper should be considered as *estimates*. As can be seen in Figs 11 and 13, the errors on the completeness at high stellar mass are large due to the limited volume of Stripe 82. Upcoming wide area surveys will be able to repeat this analysis with much higher precision using several hundred to thousands of square degrees. The values presented here are reported with respect to s82-MCG masses so offsets may need to be applied to translate these completeness values to other mass estimates – see Paper I for details.

With upcoming surveys in mind, such as the HSC and Euclid surveys, which will overlap with the BOSS footprint, we also characterize how many supplementary galaxies with photometric redshifts will be needed at any given stellar mass and redshift bin in order to construct mass-limited samples. A sample that is mass limited to  $\log_{10}(M_*/M_\odot) > 11.6$  can be constructed at  $z < 0.61$  by supplementing the BOSS samples with photometric redshifts at the  $\sim 20$  per cent level. At  $\log_{10}(M_*/M_\odot) > 11.0$ , however, spectroscopic samples need to be significantly supplemented by photometric redshifts (at the 80 per cent level).

We use our methodology to evaluate the stellar mass completeness of several specific samples that have been used in past work using BOSS data. We demonstrate that previous work has sometimes overestimated the stellar mass completeness of BOSS samples and suggest that caution needs to be taken when analysing BOSS samples for these types of studies.

The completeness estimates provided by this paper will enable future studies to better utilize the BOSS samples for galaxy-formation and cosmology science. A better understanding of the BOSS selection functions will also enable the construction of improved mock catalogues for the BOSS survey. Our companion paper presents improved mock catalogues that account for the stellar mass completeness of the BOSS CMASS sample as a function of redshift (Saito et al. 2015). These mock catalogues are made publicly available at [www.massivegalaxies.com](http://www.massivegalaxies.com).

## ACKNOWLEDGEMENTS

We thank the anonymous referee for their thoughtful comments which helped improve this manuscript. We thank Jean Coupon, Antonio Montero Dorta, Surhud More, Hironao Miyatake, and Francesco Shankar for useful discussions while preparing this paper. We are grateful to Hong Guo for kindly providing catalogues. This work was supported by World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan. RT acknowledges support from the Science and Technology Facilities Council via an Ernest Rutherford Fellowship (grant number ST/K004719/1). Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the US Department of Energy Office of Science. The SDSS-III web site is <http://www.sdss3.org/>. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max

Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

## REFERENCES

- Abazajian K. et al., 2004, *AJ*, 128, 502  
 Ahn C. P. et al., 2014, *ApJS*, 211, 17  
 Aihara H. et al., 2011, *ApJS*, 193, 29  
 Alam S. et al., 2015, *ApJS*, 219, 12  
 Anderson L. et al., 2014, *MNRAS*, 441, 24  
 Annis J. et al., 2014, *ApJ*, 794, 120  
 Baldry I. K., Glazebrook K., Driver S. P., 2008, *MNRAS*, 388, 945  
 Baldry I. K. et al., 2010, *MNRAS*, 404, 86  
 Beifiori A. et al., 2014, *ApJ*, 789, 92  
 Benson A. J., 2014, *MNRAS*, 444, 2599  
 Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, *ApJ*, 599, 38  
 Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587  
 Blanton M. R., Roweis S., 2007, *AJ*, 133, 734  
 Bolton A. S. et al., 2012, *AJ*, 144, 144  
 Bolzonella M., Miralles J.-M., Pelló R., 2000, *A&A*, 363, 476  
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000 (BC03)  
 Bundy K. et al., 2010, *ApJ*, 719, 1969  
 Bundy K., Hogg D. W., Higgs T. D., Nichol R. C., Yasuda N., Masters K. L., Lang D., Wake D. A., 2012, *AJ*, 144, 188  
 Bundy K. et al., 2015, *ApJS*, 221, 15 (Paper I)  
 Chabrier G., 2003, *PASP*, 115, 763  
 Chen Y.-M. et al., 2012, *MNRAS*, 421, 314  
 Coil A. L. et al., 2011, *ApJ*, 741, 8  
 Coupon J. et al., 2015, *MNRAS*, 449, 1352  
 Dawson K. S. et al., 2013, *AJ*, 145, 10  
 Eisenstein D. J. et al., 2001, *AJ*, 122, 2267  
 Eisenstein D. J. et al., 2011, *AJ*, 142, 72  
 Frieman J. A. et al., 2008, *AJ*, 135, 338  
 Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *AJ*, 111, 1748  
 Furlong M. et al., 2015, *MNRAS*, 450, 4486  
 Gunn J. E. et al., 1998, *AJ*, 116, 3040  
 Gunn J. E. et al., 2006, *AJ*, 131, 2332  
 Guo H. et al., 2013, *ApJ*, 767, 122  
 Guo H. et al., 2014, *MNRAS*, 441, 2398  
 Guo H. et al., 2015, *MNRAS*, 449, L95  
 Guzzo L. et al., 2014, *A&A*, 566, A108  
 Hoshino H. et al., 2015, *MNRAS*, 452, 998  
 Knebe A. et al., 2015, *MNRAS*, 451, 4029  
 Kroupa P., 2001, *MNRAS*, 322, 231  
 Laureijs R. et al., 2011, preprint ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))  
 Lawrence A. et al., 2007, *MNRAS*, 379, 1599  
 Le Borgne J.-F. et al., 2003, *A&A*, 402, 433  
 Le Fèvre O. et al., 2004, *A&A*, 428, 1043  
 Le Fèvre O. et al., 2015, *A&A*, 576, A79  
 Leauthaud A. et al., 2012, *ApJ*, 746, 95  
 Lilly S. J. et al., 2007, *ApJS*, 172, 70  
 Mandelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006, *MNRAS*, 368, 715  
 Maraston C., Strömbäck G., 2011, *MNRAS*, 418, 2785  
 Maraston C., Strömbäck G., Thomas D., Wake D. A., Nichol R. C., 2009, *MNRAS*, 394, L107  
 Maraston C., Pforr J., Renzini A., Daddi E., Dickinson M., Cimatti A., Tonini C., 2010, *MNRAS*, 407, 830  
 Maraston C. et al., 2013, *MNRAS*, 435, 2764  
 Marchesini D. et al., 2014, *ApJ*, 794, 65



- Marsan Z. C. et al., 2015, ApJ, 801, 133  
 Masters K. L. et al., 2011, MNRAS, 418, 1055  
 Miyatake H. et al., 2015, ApJ, 806, 1 preprint, (arXiv:1311.1480)  
 Montero-Dorta A. D. et al., 2014, preprint (arXiv:1410.5854)  
 Montero-Dorta A. D., Shu Y., Bolton A. S., Brownstein J. R., Weiner B. J., 2016, MNRAS, 456, 3265  
 More S., Miyatake H., Mandelbaum R., Takada M., Spergel D., Brownstein J., Schneider D. P., 2015, ApJ, 806, 2  
 Moustakas J. et al., 2013, ApJ, 767, 50  
 Newman J. A. et al., 2013, ApJS, 208, 5  
 Oke J. B., Gunn J. E., 1983, ApJ, 266, 713  
 Parejko J. K. et al., 2013, MNRAS, 429, 98  
 Reid B. A., Seo H.-J., Leauthaud A., Tinker J. L., White M., 2014, MNRAS, 444, 476  
 Reis R. R. R. et al., 2012, ApJ, 747, 59  
 Rykoff E. S. et al., 2014, ApJ, 785, 104  
 Saito S. et al., 2015, preprint (arXiv:1509.00482)  
 Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525  
 Shankar F. et al., 2014, ApJ, 797, L27  
 Smee S. A. et al., 2013, AJ, 146, 32  
 Stoughton C. et al., 2002, AJ, 123, 485  
 Strauss M. A. et al., 2002, AJ, 124, 1810  
 Swanson M. E. C., Tegmark M., Hamilton A. J. S., Hill J. C., 2008, MNRAS, 387, 1391  
 Tal T., van Dokkum P. G., Franx M., Leja J., Wake D. A., Whitaker K. E., 2013, ApJ, 769, 31  
 Thomas D. et al., 2013, MNRAS, 431, 1383  
 Tojeiro R. et al., 2012, MNRAS, 424, 136  
 van de Sande J., Kriek M., Franx M., Bezanson R., van Dokkum P. G., 2015, ApJ, 799, 125  
 Wake D. A. et al., 2006, MNRAS, 372, 537  
 York D. G. et al., 2000, AJ, 120, 1579  
 Zu Y., Mandelbaum R., 2015, MNRAS, 454, 1161

## APPENDIX A: CATALOGUES

Here, we list the various links to the publicly available catalogues used in this paper.

(i) The original catalogue used to target BOSS galaxies. The target catalogue for Stripe 82 is `bosstarget-lrg-main007-collate.fits` and can be found at <http://data.sdss3.org/sas/dr10/boss/target/main007/>. The target selection flags are contained in the bitfield `BOSS_TARGET1`. A description of `BOSS_TARGET1` can be found at [https://www.sdss3.org/dr10/algorithms/bitmask\\_boss\\_target1.php](https://www.sdss3.org/dr10/algorithms/bitmask_boss_target1.php). Various targets can be selected via  $(\text{BOSS\_TARGET1} \ \&\&2^i) > 0$  where  $i$  is a binary digit. In this paper, we are primarily concerned with the samples `GAL_LOZ` ( $i = 0$ ), `GAL_CMASS` which corresponds to the CMASS selection described in Section 2.1 ( $i = 1$ ), `SDSS_KNOWN` which corresponds to objects with Legacy objects ( $i = 6$ ), and `GAL_CMASS_ALL` which includes `GAL_CMASS` and the entire sparsely sampled region ( $i = 7$ ).

(ii) The Portsmouth stellar mass catalogues can be found at [https://www.sdss3.org/dr10/spectro/galaxy\\_portsmouth.php](https://www.sdss3.org/dr10/spectro/galaxy_portsmouth.php).

(iii) The DR10 LSS catalogues (Anderson et al. 2014) that we use are `galaxy_DR10v8_CMASS_South.fits` and `galaxy_DR10v8_LOWZ_South.fits`. These catalogues can be found at <http://data.sdss3.org/sas/dr10/boss/lss/> and the data model can be found at [http://data.sdss3.org/datamodel/files/BOSS\\_LSS\\_REDUX/galaxy\\_DR10v8\\_SAMPLE\\_NS.html](http://data.sdss3.org/datamodel/files/BOSS_LSS_REDUX/galaxy_DR10v8_SAMPLE_NS.html).

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