Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **455,** 1553–1573 (2016) doi:10.1093/mnras/stv2382

SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues

Beth Reid,[1,2] Shirley Ho,[3] Nikhil Padmanabhan,[4] Will J. Percival,[5]★ Jeremy Tinker,[6] Rita Tojeiro,[7] Martin White,[2,8] Daniel J. Eisenstein,[9] Claudia Maraston,[5] Ashley J. Ross,[10] Ariel G. Sánchez,[11] David Schlegel,[1,2] Erin Sheldon,[12] Michael A. Strauss,[13] Daniel Thomas,[5] David Wake,[14,15] Florian Beutler,[2] Dmitry Bizyaev,[16,17] Adam S. Bolton,[18] Joel R. Brownstein,[18] Chia-Hsun Chuang,[19] Kyle Dawson,[18] Paul Harding,[20] Francisco-Shu Kitaura,[21] Alexie Leauthaud,[22] Karen Masters,[5] Cameron K. McBride,[9] Surhud More,[22] Matthew D. Olmstead,[23,18] Daniel Oravetz,[16] Sebastián E. Nuza,[24] Kaike Pan,[16] John Parejko,[4] Janine Pforr,[5,25] Francisco Prada,[19,26,27] Sergio Rodríguez-Torres,[19,26,28] Salvador Salazar-Albornoz,[11,29] Lado Samushia,[30,31,5] Donald P. Schneider,[32,33] Claudia G. Scóccola,[34,35,36] Audrey Simmons[16] and Mariana Vargas-Magana[37]

Affiliations are listed at the end of the paper

Accepted 2015 October 12. Received 2015 October 12; in original form 2015 July 22

ABSTRACT

The Baryon Oscillation Spectroscopic Survey (BOSS), part of the Sloan Digital Sky Survey (SDSS) III project, has provided the largest survey of galaxy redshifts available to date, in terms of both the number of galaxy redshifts measured by a single survey, and the effective cosmological volume covered. Key to analysing the clustering of these data to provide cosmological measurements is understanding the detailed properties of this sample. Potential issues include variations in the target catalogue caused by changes either in the targeting algorithm or properties of the data used, the pattern of spectroscopic observations, the spatial distribution of targets for which redshifts were not obtained, and variations in the target sky density due to observational systematics. We document here the target selection algorithms used to create the galaxy samples that comprise BOSS. We also present the algorithms used to create large-scale structure catalogues for the final Data Release (DR12) samples and the associated random catalogues that quantify the survey mask. The algorithms are an evolution of those used by the BOSS team to construct catalogues from earlier data, and have been designed to accurately quantify the galaxy sample. The code used, designated MKSAMPLE, is released with this paper.

Key words: cosmology: observations – (cosmology:) large-scale structure of Universe.

1 INTRODUCTION

The size of galaxy redshift surveys has grown exponentially over the last decade and will continue do so into the next, thanks to the continuing development of instrumentation to undertake multi-object spectroscopy on dedicated telescopes. The scientific driver for this dramatic increase is that galaxy redshift surveys provide a wealth of cosmological and extragalactic information. The most easily accessible cosmological information is encoded in two-point clustering statistics of the overdensity field, which contain both the Baryon Acoustic Oscillation (BAO) and Redshift Space Distortion (RSD) signals. The BAO scale is a comoving large-scale enhancement in pairs of galaxies separated by ∼150 Mpc, which can be used to track cosmological expansion. It arises from the propagation of sound waves in the early Universe (Peebles & Yu 1970; Sunyaev & Zel'dovich 1970; Doroshkevich, Zel'dovich & Sunyaev 1978), and is quite insensitive to astrophysical processing that occurs on smaller scales; thus BAO experiments are affected by a low level of systematics (see review by Weinberg et al. 2013 for a comparison of

*E-mail: will.percival@port.ac.uk

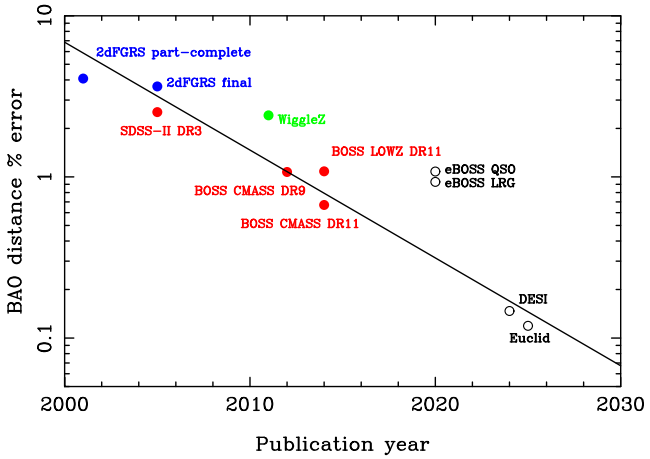

Figure 1. BAO measurement errors predicted for various surveys as a function of the year of publication. In order to calculate these with a consistent methodology we plot 'predictions' using the code of Seo & Eisenstein (2007) based on a single number of galaxies, and volume for each survey. The surveys plotted are 2dFGRS early (Percival et al. 2001) and final (Cole et al. 2005); SDSS-II LRGs (Eisenstein et al. 2005); WiggleZ (Blake et al. 2011); BOSS DR9 CMASS (Anderson et al. 2012); BOSS DR11 LOWZ (Tojeiro et al. 2014) and CMASS (Anderson et al. 2014b). In terms of survey volume, BOSS DR12 is very close to DR11 and we do not show it here. We also present approximate predictions for the eBOSS, DESI and Euclid future surveys (see the text for details).

different methods). RSD arise from the peculiar velocities of galaxies within a comoving frame, which produce coherent distortions in the measured redshifts compared to those produced by the Hubble expansion (Kaiser 1987). As these velocities are gravitational in origin, the amplitude depends on the rate of structure growth, and hence RSD allow tests of General Relativity (GR) on large scales.

The BAO signature has now been detected in many different galaxy surveys and analysed using a variety of methods. To show the exponential growth in BAO measurements, Fig. 1 presents the predicted error on the BAO scale expected for different surveys, calculated as if the clustering signal from different directions was optimally combined to provide the best possible single BAO position measurement. We include results from various stages of the 2-degree-Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001, 2003), Sloan Digital Sky Survey (SDSS; York et al. 2000) and WiggleZ (Drinkwater et al 2010), and predictions for the continuation of the SDSS project with eBOSS (Dawson et al. 2015). For consistency, all calculations used the code of Seo & Eisenstein (2007), approximating each survey as a single volume, limited in redshift and area, and sampled by a constant density of galaxies, with numbers approximately matching those of the actual surveys. Thus the results themselves are not precise and are designed to simply demonstrate the evolution rather than provide a quantitative comparison between experiments. The best-fitting line shows the growth in the impact of past surveys, following the development of multi-object spectrographs (MOS) on the Anglo-Australian telescope (Lewis, Glazebrook & Taylor 1998) and the Sloan telescope (Gunn et al. 2006), which continues to the next generation with a new MOS being developed for the Hobby–Eberly telescope (HETDEX; Hill et al. 2008), the Mayall telescope (DESI; Levi et al. 2013), the VISTA telescope (4MOST; de Jong et al. 2014), the William Herschel Telescope (WEAVE; Dalton et al. 2014), the Subaru telescope (PFS; Takada et al. 2014) and the satellite experiments Euclid (Laureijs et al. 2011) and WFIRST (Spergel et al. 2015). For clarity, we only plot approximate DESI and Euclid predictions in Fig. 1 to show the general expected trend from these new instruments, as our simplified approach is insufficient to provide a careful differential analysis of these future projects. Also, there is significant uncertainty in the predictions for Euclid, as a consequence of our lack of knowledge about the galaxy population targeted: the prediction here uses the predicted volume and galaxy density of Laureijs et al. (2011). The higher redshift surveys of eBOSS and WiggleZ are inherently more difficult and consequently they lie above the line: they push into new redshift ranges, rather than to larger volumes.

In this paper, we present the target selection and catalogue generation of the Data Release 12 (DR12; Alam et al. 2015) samples of galaxies selected from the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2012), which is part of SDSS-III (Eisenstein et al. 2011). The spectroscopic sample has two primary catalogues: LOWZ at $z \lesssim 0.4$, and CMASS covering $0.4 \lesssim z \lesssim 0.7$ (see Section 3 for details). An overview of the BOSS observations is provided in Section 2; see Dawson et al. (2012) for a full description of the survey.

The work presented here follows on from the analysis of previous data releases: DR12 is the third public SDSS data release containing BOSS spectroscopic results. The first was DR9 (Ahn et al. 2012), when the survey was approximately one-third complete. The creation of the large-scale structure (LSS) catalogues from these data was outlined in Anderson et al. (2012), alongside the isotropic BAO results, with the anisotropic results following in Anderson et al. (2014a). The development of a method to remove potential systematic errors in the clustering measurements caused by fluctuations in the target catalogue was presented in Ross et al. (2012). The DR9 catalogues were used extensively for science, which also tested the catalogues themselves. The clustering was compared with simulations in Nuza et al. (2013), and with full model fits in Sanchez et al. (2012). RSD were measured by Reid et al. (2012) and enhanced by knowledge of passive galaxy evolution in Tojeiro et al. (2012); the resulting GR tests were presented in Samushia et al. (2013). Primordial non-Gaussianity was constrained by Ross et al. (2013), while Zhao et al. (2013) reported neutrino masses, and Scoccola et al. (2013) examined the time variation of physical constants. This work led to further refinements of the catalogue creation algorithm for the analysis of the second public BOSS data release DR10 (Ahn et al. 2014), which coincided with an internal release (called DR11). In particular, the code was rewritten into a modular version, called MKSAMPLE, new weights were used to correct for fluctuations in the expected target density, and new masks were used for 'bad' areas. These refinements were presented alongside the BAO results for the CMASS sample in Anderson et al. (2014b) and the LOWZ sample in Tojeiro et al. (2014), and were confirmed to be robust to colour (Ross et al. 2014) and against possible systematics in the fit (Vargas-Magana et al. 2014). As for DR9, the results were extensively used, further testing the catalogues: RSD measurements have been made in a number of different ways (Beutler et al. 2014a; Samushia et al. 2014; Sanchez et al. 2014; Chuang et al. 2013), the bispectrum calculated and analysed (Gil-Marin et al. 2015a,b), and neutrino mass constraints presented (Beutler et al. 2014b). Saito et al. (2015) account for redshift dependent selection effects and compare clustering and RSD with predictions from abundance matching.

We have now analysed the final BOSS DR12 galaxy sample using an algorithm that builds on the work described above. This paper on the targeting algorithm and catalogue creation method is complemented by a series of papers measuring and analysing

clustering, splitting the BOSS galaxies into subsamples delineated by the primary targeting algorithms LOWZ and CMASS samples (see Section 3 for details). BAO measurements are presented in configuration-space (Cuesta et al. 2015) and Fourier-space (Gil-Marin et al. 2015c), and RSD measurements made in Fourier-space are presented in Gil-Marin et al. (2015d). Two further support papers are provided in this set: Ross et al. (in preparation) considers the BOSS selection function in more detail, presenting the observational foot-print, masks for image quality and Galactic extinction, and weights to account for density relationships intrinsic to the imaging and spectroscopic portions of the survey. Vargas-Magana et al. (2015) presents systematic tests on the reconstruction algorithm used for anisotropic BAO analyses. A subsequent set of analyses to be released soon, will consider jointly analysing the full BOSS sample, without splitting by target selection.

Because the key cosmological measurements depend on the density field, galaxy properties (except how they trace this field, commonly quantified by a linear deterministic bias b), are unimportant once redshifts have been measured, and cosmological surveys are free to choose which galaxies to observe to optimize survey efficiency and the optimal bias b. BOSS targets luminous galaxies for spectroscopic observations as they have a large bias, are relatively easy to target, and have strong spectral features that ease redshift determination. The target selection adopted by BOSS is an extension of the targeting algorithms for the SDSS-II (Eisenstein et al. 2001) and 2SLAQ (Cannon et al. 2006) luminous red galaxies (LRGs), targeting fainter and bluer galaxies in order to achieve the desired number density of $\sim 3 \times 10^{-4}\, h^3\, \mathrm{Mpc}^{-3}$. The majority of the galaxies are old stellar systems whose prominent 4000 Å break makes them relatively easy to target using multicolour data. The data from which the samples are targeted is described in Section 2, and the LOWZ and CMASS target selection algorithms are discussed in detail in Section 3.

In order to do LSS analyses with the sample of spectroscopically observed galaxies, we have put together catalogues including information on the detailed angular and radial mask of the sample including the redshift completeness, the observing conditions when the imaging and spectroscopic observations were made, and the appropriate weights to give each object, as well as random (i.e. unclustered) catalogues with the same selection function. These collectively make up the LSS catalogues, whose contents are detailed in this paper.

Key to creating these catalogues for the BOSS galaxy surveys is the ability to predict where we could have observed galaxies, as well as where galaxies exist, thus defining the survey or sample mask. This mask is intricately linked with the selection of galaxies: in general, corrections for selection effects can be applied to either the mask or the galaxy sample to produce a match between the two. In order to understand the mask, we need to understand both the target sample and the subsequent spectroscopy and redshift measurement, which we briefly summarize in Section 4. The BOSS galaxy mask is quantified using a 'random catalogue', a Poisson sampling of the volume covered by the selected galaxies, including any variations in density other than the cosmological clustering signal we wish to measure. The '3D mask' does not have to be quantified by a Poisson sampling, but this is a straightforward approach to this – in effect providing a Monte Carlo sampling of the volume covered. This weighted random sample and the weighted galaxy sample form the starting point for the key BOSS galaxy clustering analyses. Section 5 presents the method adopted by the BOSS team to prepare catalogues of galaxies and randoms, using routines made publicly available in a code called MKSAMPLE. This is a further extension of the code used for the early DR9 analyses, which is described in Anderson et al. (2012).

Although the targeting algorithm adopted for BOSS is isotropic and the catalogue of target objects covers an angular area larger than that of spectroscopic observations, the mask is complicated by various anisotropic effects including variations in imaging depth due to recalibration of the SDSS photometric scale and rereduction of the imaging as the spectrosocpic survey progressed, variation of seeing, variation with stellar density caused by occultation by stars, the inability to measure spectra for close to another target observed at the same time, and the failure to measure spectra as a function of signal-to-noise ratio in the spectrum. These effects are often corrected by applying a weight to the galaxies (e.g. Ross et al. 2012), but could instead be incorporated into the mask. The quality of the DR12 data is such that we can now observe systematic effects that couple radial and angular fluctuations, and we introduce 3D corrections for these. The manner adopted to deal with these effects for BOSS is described in Section 6.

BOSS includes a number of galaxy catalogues with different selection functions, some of which spatially overlap. The combination of these to optimally quantify the underlying matter overdensity field is non-trivial, and we present the method adopted by the BOSS team in Section 7.

The MKSAMPLE code will be released upon publication of this paper, and we will also publish the resulting LSS catalogues, with a full datamodel describing each. These will all be linked from the main SDSS web site http://www.sdss.org/surveys/boss.

2 DATA

2.1 Imaging data

The SDSS-I/II (York et al. 2000) imaged approximately 7606 deg^2 of the Northern Galactic Hemisphere and 600 deg^2 of the Southern Galactic Hemisphere in the *ugriz* bands (Fukugita et al. 1996; Smith et al. 2002; Doi et al. 2010), using a specially designed camera (Gunn et al. 1998) on the 2.5 m Sloan telescope (Gunn et al. 2006) at the Apache Point Observatory in New Mexico. The SDSS-III project (Eisenstein et al. 2011) obtained additional imaging to make the region of the Southern Galactic Hemisphere contiguous, covering 3172 deg^2. As part of this effort, the original SDSS-I/II data and the SDSS-III data were reduced with the latest versions of the SDSS image processing and calibration pipelines (Lupton et al. 2001; Pier et al. 2003; Padmanabhan et al. 2008). These data were released as part of Data Release 8 (Aihara et al. 2011), and form the parent imaging catalogue for the BOSS galaxy target selection. There are a number of differences between the processing performed for DR8 (see the DR8 paper Aihara et al. 2011 for a detailed discussion) and earlier reductions; reproducing BOSS galaxy samples derived from the imaging data requires using the appropriate algorithms.

BOSS obtained spectra and redshifts for 1372 737 galaxies over 9376 deg^2. The targets are assigned to tiles of diameter 3°, using a tiling algorithm that is adaptive to the density of targets on the sky (Blanton et al. 2003). Spectra are then obtained using the BOSS spectrographs (Smee et al. 2013). Each observation is performed in a series of 900 s exposures, integrating until a minimum signal-to-noise ratio is achieved for the faint galaxy targets. Redshifts are then measured using the methods described in Bolton et al. (2012). The spectroscopic observations were split into distinct areas of sky, which we call chunks, targeted separately and sequentially in time, each defined by a subset of the total footprint. Later chunks can overlap earlier chunks and recover unobserved targets. The angular

Table 1. Basic definitions for the geometric description of SDSS-III BOSS observations and the LSS MANGLE masks.

Spherical polygon	The base unit of a MANGLE mask. Spherical polygons are used to represent the boundaries of the imaging survey from which the targets are drawn, the circular fields defined by spectroscopic tiles, as well as regions to be removed from the survey footprint (e.g. the centerpost of each spectroscopic tile; see Section 5.1.1 for a full list).
Spectroscopic tile	Output of the tiling algorithm providing a central location on the sky and list of targets to be observed for spectroscopic observations. Each tile has a circular field of view of radius 1.°49 , and can be observed by multiple plates.
Plate	Physical plate with a hole drilled for each target, based on the anticipated airmass of observation. Spectroscopic tiles may be observed using multiple plates.
Chunk	Basic unit of sky input to the tiling algorithm. It consists of a set of rectangles in a spherical coordinate system. The SDSS-III BOSS survey is composed of 38 chunks.
Sector	The union of spherical polygons defined by a unique intersection of spectroscopic tiles. The survey completeness is treated as uniform within a sector.

distribution of chunks 2–11, which are special as they reflect early versions of the target selection (see Appendix A), but also serve to show how the survey is built up from chunks are shown in Fig. A1, and basic definitions for geometrical descriptors used in this paper are provided in Table 1.

The start of spectroscopic observations preceded the finalization of the DR8 imaging reductions, so the imaging data used by BOSS are based on the photometric measurements *available at the time of tiling* (see Section 4.2), which may differ from the quantities available for an object in the DR8 catalogue. BOSS targeting was performed using three different versions of the reduction software that resolves the catalogues from overlapping imaging data (RESOLVE; see Aihara et al. 2011). Chunks 1–4 used a version of the RESOLVE software tagged on 14-06-2009, chunks 5–11 used a version tagged on 16-11-2009, and chunks 12 onwards used the same version as that used to produce DR8. In total, 17 per cent of targets were targeted with pre-DR8 RESOLVE versions. Because these different versions of the software selected different imaging data to be designated as 'primary' (i.e. either the only or the best observation of this object; see the DR8 documentation for more details), approximately 9 per cent of the imaging data used for targeting CMASS galaxies is now designated as secondary[1] in the DR8 data base.

2.2 Parent catalogue

The selection of galaxy targets for spectroscopic observation is based on a parent catalogue of photometrically identified objects within the imaging data. The parent catalogue was based on objects chosen from 3172 deg^2 in the Southern Galactic Cap (SGC) and 7606 deg^2 in the Northern Galactic Cap (NGC), as described in this section. The SDSS imaging pipeline returns a number of different measurements of the photometry of galaxies. Full descriptions may be found on the SDSS website[2] and in Stoughton et al. (2002). For galaxy target selection, we use three photometric measurements, which have all been corrected for Galactic extinction using the Schlegel, Finkbeiner & Davis (1998) dust maps.

The colours of galaxies are based on SDSS model magnitudes (denoted by the subscript *mod*). These are determined by using the best-fitting [point spread function (PSF)-convolved] deVaucouleurs or exponential profile fit in the r band to determine the fluxes in the other bands (full details are provided in Abazajian et al. 2004). Cuts in apparent magnitude are made with 'cmodel' magnitudes (denoted by a subscript cmod). These are a linear combination of the flux from the best-fitting exponential and deVaucouleurs profile fit in each band separately[3]

$$f_{\rm cmod} = (1 - P) f_{\rm exp} + P f_{\rm deV}, \quad (1)$$

where P is the best-fitting coefficient obtained from a fit of the linear combination of the deVaucouleurs and exponential profile fits to the image, and weights the different contributions (reported as $\rm frac_{deV}$ by the SDSS pipelines), and f represents the flux (*not magnitude*) assuming an exponential or deVaucouleur profile. Star–galaxy separation compares the PSF magnitudes of galaxies (denoted by a subscript PSF) with model or cmodel magnitudes; PSF magnitudes underestimate the flux from extended sources compared with the model fits (see Section 3.3.2 for details). Finally, we use 'fibre2' (denoted by subscript fib2) magnitudes to estimate the expected flux through the SDSS-III 2 arcsec fibres.

The parent sample for the BOSS galaxy target selection is constructed by selecting all detected objects that the photometric pipeline classifies as galaxies, and that are chosen by RESOLVE to be 'primary'. The targeting software uses the photometry of the primary objects to select targets for spectroscopic follow-up. The variation in selected targets from different imaging data is consistent with that expected given the photometric uncertainties, and so we treat the regions targeted with pre- and final DR8 imaging as statistically identical. We do not make any cuts on photometricity at this stage; unphotometric data is discarded at the catalogue creation stage (see Section 5.1.1). Users constructing their own samples for science analyses are advised to use the CALIB_STATUS flag to cut on photometricity (restricting to photometric observations corresponds to `CALIB_STATUS==1`). We cull objects with suspect photometry as reported in the flags set by the imaging pipeline. In particular, we require objects that are detected in the r and i bands. In the Image Processing pipeline, this is indicated by having one of the BINNED1, BINNED2 or BINNED4 flags set in both the r and i bands. We also require that the OBJC_FLAG flag, which is a combination of the per-filter flags appropriate for the whole object (the full definition is provided in Stoughton et al. 2002) has

(i) objects not to be saturated: `(NOT SATUR) OR (SATUR AND (NOT SATUR_CENTER))`,
(ii) blended objects: `(NOT BLENDED) OR (NOT NODEBLEND)`,
(iii) other photometric quality flags: `(NOT BRIGHT) AND (NOT TOO_MANY_PEAKS) AND (NOT PEAKCENTER) AND (NOT NOTCHECKED) AND (NOT NOPROFILE)`.

[1] I.e. there is an overlapping observation with higher quality photometry.
[2] http://www.sdss3.org/dr8
[3] Contrasted with model magnitudes where the fit in the r band is used.

3 TARGET SELECTION

We now turn to the specifics of the target selection algorithms used to define the BOSS spectroscopic galaxy samples. We first summarize the criteria that we wish our algorithm to satisfy (Section 3.1), with the aim of defining a uniformly selected sample over a broad redshift range. The galaxy sample is targeted using two different algorithms, which we term 'LOWZ' (detailed in Section 3.2) and 'CMASS' (for 'Constant (stellar) Mass', Section 3.3), respectively. Star–galaxy separation is treated differently in the CMASS sample than elsewhere in SDSS, as we describe in Section 3.3.2. A variant of the CMASS algorithm was used to explore the colour boundaries of the sample (Section 3.4).

3.1 Requirements and criteria

The BOSS sample was designed to measure the BAO signature in the two-point galaxy clustering signal, and in particular to meet error requirements on the measurement of the angular diameter distance d_A and Hubble parameter H at $z = 0.35$ and 0.6. These requirements can be met by a survey covering an area of approximately 10 000 deg^2 with a comoving number density of galaxies of $3 \times 10^{-4}\ h^3\ \mathrm{Mpc}^{-3}$ for $0.1 < z < 0.6$. This density is close to optimal for large-scale cosmological studies (e.g. Kaiser 1986). To efficiently undertake such a survey using the Sloan telescope and spectrographs, we need to select a subsample of the parent catalogue of photometrically identified objects that fulfil the following criteria simultaneously:

(i) galaxies that lie in the desired redshift range $0.1 < z < 0.6$,
(ii) sufficient galaxies to meet the desired density over the full redshift range,
(iii) well-defined limits in stellar populations, to isolate a strongly clustered subsample of galaxies,
(iv) redshifts that can be measured in a relatively short exposure with our telescope,
(v) few contaminating objects that are not part of the desired sample,
(vi) selectable uniformly across the desired area,
(vii) selection is not sensitive to systematic errors in the data used.

The challenge of target selection is to provide an algorithm for selecting the subsample of the parent imaging catalogue that optimally meets these goals. Selection based solely on an apparent magnitude cut, as used for the SDSS-I and -II Main Galaxy Sample (Strauss et al. 2002) in general selects too many low-redshift and low-luminosity galaxies. Rather, in BOSS we follow a similar philosophy to the selection of LRGs in SDSS-I and -II (Eisenstein et al. 2001) and the 2SLAQ survey (Cannon et al. 2006) using colour–magnitude and colour–colour cuts, selecting luminous galaxies with strong spectral features (item (iv) above).

At redshifts $z < 0.4$, we can select such a sample by extending to fainter LRGs than observed in SDSS-I and -II. At higher redshifts, we do not restrict ourselves to red galaxies, and instead select an approximately stellar mass-limited sample of objects of all intrinsic colours. As in Eisenstein et al. (2001), two sets of colours are necessary to describe the colour locus: one when the 4000 Å break lies in the SDSS g-band, and the other when it redshifts into the r band at $z \sim 0.4$. Selecting these two subsamples requires defining fiducial colours that track the locus of a passively evolving population of galaxies in gri colour space. Following Eisenstein et al. (2001) and Cannon et al. (2006), we define

$$c_{||} = 0.7(g_{\rm mod} - r_{\rm mod}) + 1.2(r_{\rm mod} - i_{\rm mod} - 0.18) \quad (2)$$

$$c_{\perp} = (r_{\rm mod} - i_{\rm mod}) - (g_{\rm mod} - r_{\rm mod})/4.0 - 0.18 \quad (3)$$

to describe the low-redshift locus and

$$d_{\perp} = (r_{\rm mod} - i_{\rm mod}) - (g_{\rm mod} - r_{\rm mod})/8, \quad (4)$$

to describe the high-redshift locus. As discussed above, the colours are defined using SDSS model magnitudes, and are corrected for Milky Way extinction. The efficiency of these selections to select luminous galaxies as a function of redshift is demonstrated in Fig. 2, which shows how $c_{||}$ and $r_{\rm mod} - i_{\rm mod}$ versus redshift for observed BOSS galaxies.

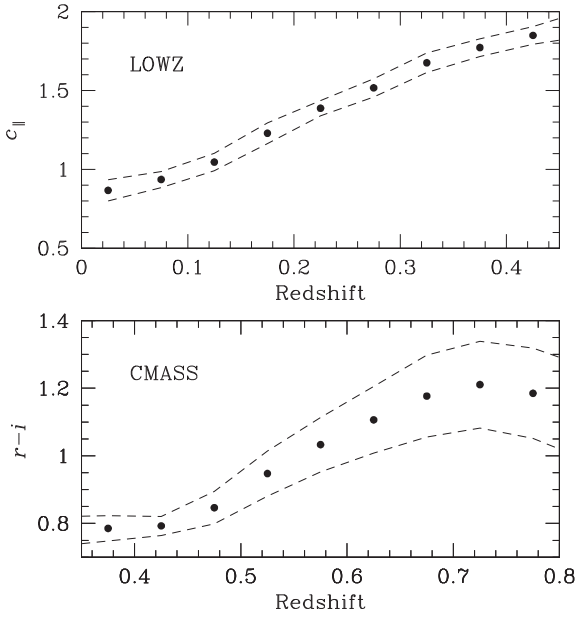

Figure 2. Top panel: black dots show median $c_{||}$ for LOWZ spectroscopically confirmed galaxies as a function of measured redshift, with the dashed lines showing the interquartile range. The efficiency of using this quantity to track redshift is clear. Bottom panel: median $r_{\rm mod} - i_{\rm mod}$ as a function of redshift for confirmed CMASS galaxies, with interquartile range (dashed lines). The way in which we can track the high-redshift locus of galaxies using this colour, and select as a function of redshift, is clear.

Where the targeting algorithms use colour selection, they are built on model magnitudes, which are based on the flux measured through equivalent apertures in all band and thus provide unbiased colours of galaxies. Brightness limits are instead based on cmodel magnitudes, which provide better estimates of the total light observed.

3.2 The LOWZ sample

The LOWZ sample is designed to extend the SDSS-I/II Cut I LRG sample (Eisenstein et al. 2001) to $z \approx 0.4$ to fainter luminosities, in order to increase the number density of the sample by roughly a factor of 3. Fig. 3 shows how the colours $c_{||}$ and $c_{\perp}$ describe the evolution of a passively evolving stellar population with redshift. Redshift increases from the bottom left to upper right. The black line shows the passively evolving LRG model of Maraston et al. (2009). The Maraston et al. (2009) 'LRG' template is a model of a metal-rich population in passive evolution containing a small fraction of a metal-poor coeval population. This model was found to be

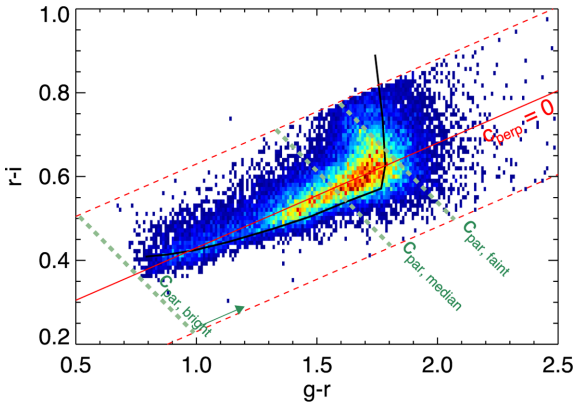

Figure 3. Density plot of LOWZ galaxies in the $(g - r, r - i)$ colour plane; red corresponds to higher density and dark blue to lower density, in an arbitrary normalization and linear scale. Redshift increases rightwards and upwards along the galaxies locus, starting at $z \simeq 0.1$ on the bottom-left corner. The knee on the galaxy locus is caused by the 4000 Å break transitioning between the *g*- and *r*-band filters, and happens at $z \approx 0.4$. The colours $c_\perp$ and $c_{||}$ are simple rotations of this colour plane, and trace the position of a target in parallel and perpendicular, respectively, to the data locus. The black thick line represents the passively evolving LRG model of Maraston et al. (2009). The green and red dashed lines are the colour and magnitude targeting cuts – see the main text for details. The few targets seen outside of the selection cut are due to differences in the targeting and final photometry, see Section 2.

a good fit to the *g*, *r*, *i* colours of LRGs from the 2SLAQ survey (Cannon et al. 2006) as a function of redshift, over models containing star formation in various amount. The same model also better fit the overall luminosity evolution of BOSS galaxies (Montero-Dorta et al. 2015). The knee seen in the galaxy locus corresponds to the transition of the 4000 Å break from the *g*- to the *r*- band. The parameter $c_{||}$ quantifies the position of a galaxy along the main locus, and $c_\perp$ characterizes the departure of a galaxy from the centre of the locus; $c_\perp = 0$ lies approximately at the centre of the galaxy distribution.

We select targets at low redshift ($z < 0.4$) around the predicted colour locus using

$$|c_\perp| < 0.2, \tag{5}$$

(red dashed lines in Fig. 3) and we select the brightest and reddest objects at each redshift using a sliding colour–magnitude cut with $c_\perp$ (an effective proxy for a photometric redshift):

$$r_{\rm cmod} < 13.5 + c_{||}/0.3. \tag{6}$$

The dashed green lines in Fig. 3 show the effective cuts in $c_{||}$ for three different *r*-band magnitudes: $r = 16$, 18.73 and 19.6 mag corresponding to the faint boundary, the median magnitude and the bright boundary of the sample, respectively. Thus fainter objects must be redder to pass the cut. This cut is the most important criterion in the selection of LOWZ galaxies – it drives the number density of the sample by effectively setting the magnitude limit as a function of redshift, and aims to produce a constant number density over the desired redshift range. The number of galaxies in the sample is therefore highly sensitive to this cut (see Ross et al. 2012; Tojeiro et al. 2014). The resulting space density of the sample is shown in Fig. 11; the sample is close to volume-limited (constant space density at $\sim 3 \times 10^{-4}\, h^3\, {\rm Mpc}^{-3}$) over the redshift range $0.2 < z < 0.4$.

We impose brightness limits on the targets, such that

$$16 < r_{\rm cmod} < 19.6. \tag{7}$$

The faint limit ensures a high-redshift success rate. The bright limit excludes a significant number of low-redshift blue galaxies that would otherwise pass the colour cut, but also excludes a fraction of brightest cluster galaxies in low-redshift massive clusters (Hoshino et al. 2015). A bright cut was not needed in SDSS-I/II as such galaxies were already targeted by the SDSS-I/II Main Galaxy Sample (Strauss et al. 2002), but a significant fraction of the BOSS footprint lies outside that of SDSS-I and -II.

The star–galaxy separation follows the same procedure introduced in Eisenstein et al. (2001) for the LRGs,

$$r_{\rm psf} - r_{\rm cmod} > 0.3. \tag{8}$$

The cmodel magnitude is a proxy for a 'total' magnitude for a galaxy, while the PSF magnitude fits the unresolved component of the object. The difference between the two is therefore a measure of the extendedness of the galaxy.

In summary, the LOWZ selection algorithm, as implemented after commissioning, is as follows:

$$r_{\rm cmod} < 13.5 + c_{||}/0.3 \tag{9}$$

$$|c_\perp| < 0.2 \tag{10}$$

$$16 < r_{\rm cmod} < 19.6 \tag{11}$$

$$r_{\rm psf} - r_{\rm cmod} > 0.3. \tag{12}$$

The galaxies in the LOWZ sample may be selected from the DR12 data base using the following flags, whose definitions can be found on the SDSS website:[4]

(i) BOSS_TARGET1 && 2^0 Objects targeted by the LOWZ algorithm.
(ii) SPECPRIMARY == 1 Objects with spectra, removing duplicate observations.
(iii) ZWARNING_NOQSO == 0 Objects whose spectroscopic redshifts are cleanly measured.
(iv) CLASS_NOQSO == 'GALAXY' Objects whose spectra are those of a galaxy (as opposed to a quasar or star).

The basic properties of the LOWZ sample are presented in Parejko et al. (2013), who fitted the small-scale clustering of the galaxies using halo occupation distribution (HOD) modelling. They demonstrated that these galaxies lie in massive haloes, with a mean halo mass of $5.2 \times 10^{13}\, h^{-1}\, {\rm M}_\odot$, a large-scale bias of ∼2.0 and a satellite fraction of 12 ± 2 per cent. These galaxies occupy haloes with average masses between those of the CMASS sample and the original SDSS I/II LRG sample.

3.2.1 *Exceptions to the LOWZ targeting*

During the first nine months of BOSS observations, the incorrect star–galaxy separation criterion was used to identify LOWZ targets, removing a significant fraction of galaxies (see Appendix A). To select a uniformly targeted sample from all LOWZ redshifts, with the selection criteria described in this section, the simplest procedure is to avoid those data with the use of an additional cut

[4] http://www.sdss.org

(i) TILEID≥10324,

where TILEID identifies spectroscopic tiles, and this cut corresponds to chunk numbers larger than 6.

Further details on this issue, and other slight changes in the targeting of LOWZ galaxies in early chunks, can be found in Appendix A. Briefly, LOWZ targets in chunk 2, and LOWZ targets in chunks 3–6, were selected with different algorithms from those of subsequent data. For the purposes of a LSS catalogue, in previous data releases we simply removed chunks 2–6 from the LOWZ sample and the corresponding mask. In Section 7, we construct separate samples using the chunk 2 ('LOWZE2') and chunk 3–6 ('LOWZE3') selections, and combine all three LOWZ catalogues with the CMASS samples to construct a single unified sample appropriate for analyses restricted to large scales, such as BAO fitting. The effects of these changes on the density of galaxies measured as a function of redshift can be seen in Fig. 11.

3.3 The CMASS sample

The CMASS sample uses similar selection cuts to those utilized by the Cut-II LRGs from SDSS-I/II and the LRGs in 2SLAQ, but extends them both bluer and fainter in order to increase the number density of targets in the redshift range $0.4 < z < 0.7$ and get closer to a mass limited sample.

The quantity $d_\perp$ (Fig. 4) effectively discards low-redshift galaxies by choosing

$$d_\perp > 0.55. \quad (13)$$

We do not apply any further colour cuts, with the exception of a sliding colour–magnitude cut that selects the brightest objects at each redshift, in such a way as to keep an approximately constant stellar mass limit over the redshift range of CMASS according to the passively evolving model of Maraston et al. (2009):

$$i_{\rm cmod} < \min(19.86 + 1.6(d_\perp - 0.8), 19.9). \quad (14)$$

This approach is a significant departure from SDSS-I/II Cut-II and 2SLAQ LRGs – which consisted of essentially a flux-limited sample with a colour cut to isolate the reddest galaxies.

We impose model and magnitude limits as follows:

$$17.5 < i_{\rm cmod} < 19.9, \quad (15)$$

$$i_{\rm fib2} < 21.5. \quad (16)$$

The faint magnitude limits are set to ensure a high-redshift success rate, whereas the bright limit protects against some low-redshift interlopers. In the first 14 tiling chunks, CMASS objects were targeted with $i_{\rm fib2} < 21.7$, but the redshift failure rate at the faint end of this range was quite poor, so we revised this limit to the final value of $i_{\rm fib2} < 21.5$.

To exclude outliers with problematic deblending, we further impose the following cuts on colour and $r_{{\rm dev},i}$ (the effective radius in the fit to the deVaucouleurs profile for the i-band magnitude, measured in pixels):

$$r_{\rm mod} - i_{\rm mod} < 2 \quad (17)$$

$$r_{{\rm dev},i} < 20.0\,{\rm pix}. \quad (18)$$

These cuts remove a very small fraction of targets. The CMASS star–galaxy separation is described in detail in the next section.

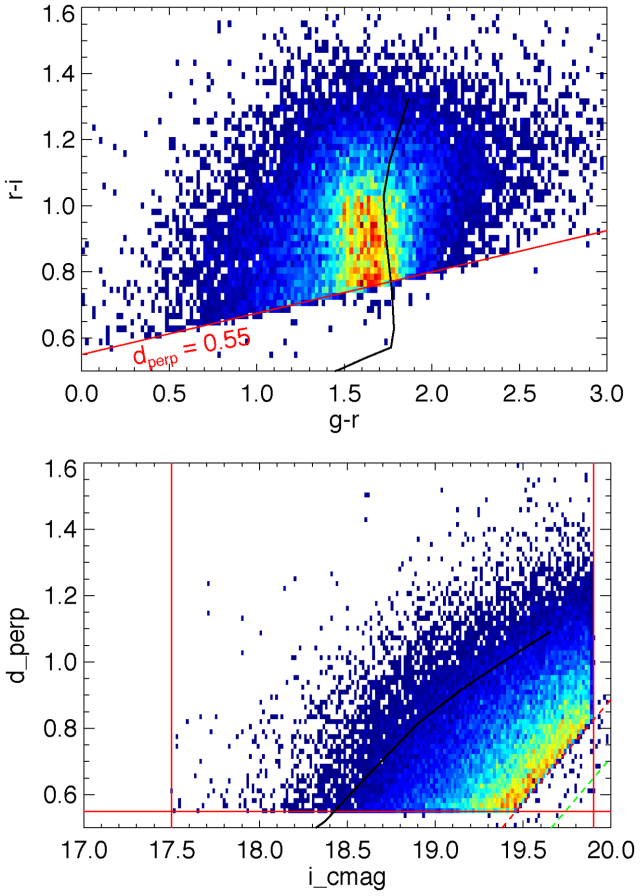

Figure 4. Both panels show density plots of CMASS galaxies; red corresponds to higher density and dark blue to lower density in an arbitrary normalization and linear scale. The black thick line shows the passively evolving LRG model of Maraston et al. (2009). Top: redshift increases upwards, starting at $z \simeq 0.4$ at $d_\perp = 0.55$. Bottom: the sliding cut in $d_\perp$ with i-band magnitude, designed to select an approximately stellar-mass complete sample. Stellar mass increases with the perpendicular distance to the sliding cut, represented here by the red dashed line – see Maraston et al. (2013) for details. The green dashed line shows the sliding cut adopted for the CMASS SPARSE sample (see Section 3.4). Vertical solid lines show the magnitude limits. On both panels, the small fraction of targets that lie outside of the selection cut are due to differences in the targeting and final photometry, see Section 2. Only chunks greater than 6 are shown.

CMASS galaxies can be selected from the DR12 data base using the following flags:

(i) BOSS_TARGET1 && 2^1
(ii) SPECPRIMARY == 1
(iii) ZWARNING_NOQSO == 0
(iv) CLASS_NOQSO == 'GALAXY'.

The basic clustering properties of the CMASS sample are presented in White et al. (2011), which fitted the small-scale clustering of the galaxies using HOD modelling. They showed that these galaxies lie in massive haloes, with a mean halo mass of $2.6 \times 10^{13}\ h^{-1}\ {\rm M}_\odot$, a large-scale bias of ∼2.0 and a satellite fraction of 10 per cent. These galaxies occupy haloes with lower masses than those of the LOWZ sample, although the bias is similar, a consequence of them being at higher redshift.

CMASS galaxies are massive, with $M_* > 10^{11}\,{\rm M}_\odot$ (e.g. Chen et al. 2012; Maraston et al. 2013), and the majority are dominated by

old stellar populations with low star formation rates (e.g. Chen et al. 2012; Tojeiro et al. 2012; Thomas et al. 2013). Maraston et al. (2013) argues that the CMASS sample becomes significantly incomplete at stellar masses $M_* < 10^{11.3}\,M_\odot$ and $z > 0.6$ for a Kroupa initial mass function, and is roughly consistent with a volume-limited sample at higher masses and lower redshift. Thomas et al. (2013) presented similar results, showing that stellar velocity dispersions of BOSS galaxies peak at ∼240 km s^1 with a narrow distribution virtually independent of redshift. Most recently, Leauthaud et al. (2015) quantified the stellar mass completeness of CMASS and LOWZ using data from the Stripe 82 region of sky along the celestial equator – a narrow, but deeper subset of the SDSS imaging survey region, that is 2 magnitudes deeper than the single epoch SDSS imaging (Annis et al. 2014). Using the Stripe 82 Massive Galaxy Catalog (Bundy et al. 2015), they estimate that CMASS is 80 per cent complete at $\log_{10}(M_*/M_\odot) \geq 11.6$ in the redshift range $z = [0.51, 0.61]$. The stellar mass completeness of CMASS decreases at lower and higher redshifts and the denomination 'constant mass' should be considered only as a loose approximation outside of the redshift window $z = [0.51, 0.61]$. However, the combination of LOWZ and CMASS yields a spectroscopic sample that is 80 per cent complete at $\log_{10}(M_*/M_\odot) \geq 11.6$ at $z < 0.61$. Compared to cut-II LRGs, CMASS galaxies have a larger range of properties including morphology (Masters et al. 2011), star formation rates (Chen et al. 2012; Thomas et al. 2013) and star-formation histories (Tojeiro et al. 2012), partly because no red cut has been imposed on the $g-r$ observed-frame colour. It should be noted however that, for example, galaxies with detectable emission-lines (hence hosting very young stellar populations) still represent only 4 per cent of the sample (see Thomas et al. 2013).

3.3.1 *Exceptions in CMASS targeting flag*

The meaning of BOSS_TARGET1 && 2^1 (the CMASS targeting flag) evolved during the first 14 chunks of the survey. Therefore BOSS_TARGET1 && 2^1 will *not* select CMASS galaxies (as defined by the equations in the previous sections) in these regions, and further subsampling is required based on galaxy colours and magnitudes to recover the final selection in these regions. Alternatively, these chunks can be explicitly excluded. For the first 14 chunks the following exceptions should be noted:

(i) Chunks 1 & 2: the data taken in the commissioning phase (chunks 1 & 2) used a significantly broader selection criteria (see Section 3.3.3), and therefore must be dealt with carefully.

(ii) Chunks 3–6: the data taken in chunks 3–6 used a slightly looser $i_{\rm cmod}$ cut, selecting instead on $i_{\rm cmod} < 19.92 + 1.6(d_\perp - 0.8)$.

(iii) Chunks 1–14: as mentioned above, the cut in $i_{\rm fib2}$ changed during the survey. In chunks 1–14, the targeting required $i_{\rm fib2} < 21.7$.

With the exception of chunk 1, all of these chunks are included in the LSS catalogue after applying the required subsampling based on colours and/or magnitudes.

3.3.2 *Star–Galaxy Separation in the CMASS sample*

The difference between PSF and model magnitudes is a measure of the extendedness of a source, thus making it useful to separate stars from galaxies. For the commissioning phase of the survey, we applied a star–galaxy separation criterion identical to that used in the 2SLAQ survey (Cannon et al. 2006), a sloping cut in $i_{\rm psf} - i_{\rm mod}$:

$$i_{\rm psf} - i_{\rm mod} > 0.2 + 0.2(20 - i_{\rm mod}). \quad (19)$$

Whilst this cut is effective at removing the bulk of the stars, roughly 6.9 per cent of ∼7000 CMASS targets from the commissioning runs had stellar spectra, mainly cool M-dwarfs.

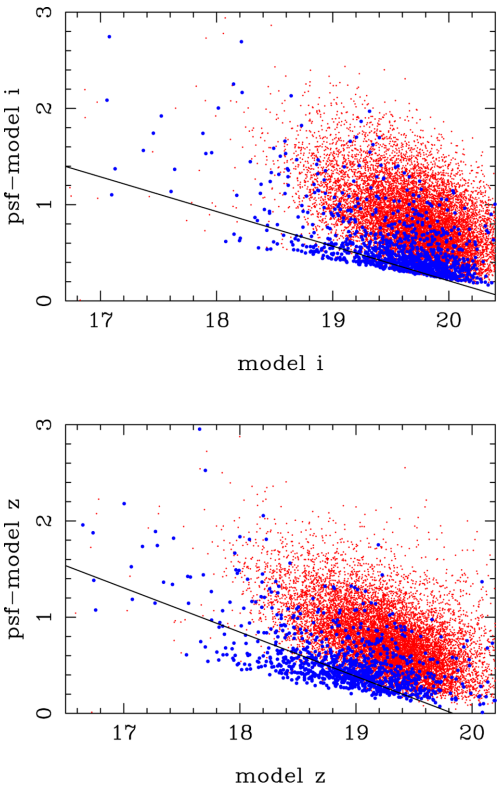

Figure 5. The distribution of spectroscopically confirmed stars (large blue points) and galaxies (small red points) in the PSF-model versus model i-band (top) and z-band (bottom) planes selected in the CMASS sample of the commissioning data. The black lines are the linear cuts that remove the most spectroscopically confirmed stars whilst removing less than 1 per cent of the galaxies. The z-band cut was added to the original i-band cut targeting from chunk 3 onwards.

Fig. 5 displays the distribution of the spectroscopically classified stars (blue) and galaxies (red) in the $i_{\rm psf} - i_{\rm mod}$ versus $i_{\rm mod}$ and $z_{\rm psf} - z_{\rm mod}$ versus $z_{\rm mod}$ planes for these commissioning targets. As expected, the stars preferentially occupy lower values of PSF-model than the galaxies, and so applying a more restrictive cut would remove more stars but at the expense of removing some galaxies. For a maximum loss of just 1 per cent of galaxies, we found the linear cuts that would remove the largest numbers of stars. These cuts, shown as the black lines in Fig. 5, remove 31 and 52 per cent of the stars that remained in the commissioning data for the i- and z-band cuts, respectively. Since the z-band cut,

$$z_{\rm psf} - z_{\rm mod} > 9.125 - 0.46 z_{\rm mod}, \quad (20)$$

performed significantly better, it was added to the original i-band cut (equation 19) for all data from Chunk 3 onwards (i.e. after the commissioning runs), such that targets have to pass both cuts to be selected. Even though the z-band cut alone removes the vast majority of the stars excluded by the i-band cut, we kept the i-band cut in place to ensure that we could apply a consistent star–galaxy separation throughout the survey. This is achieved

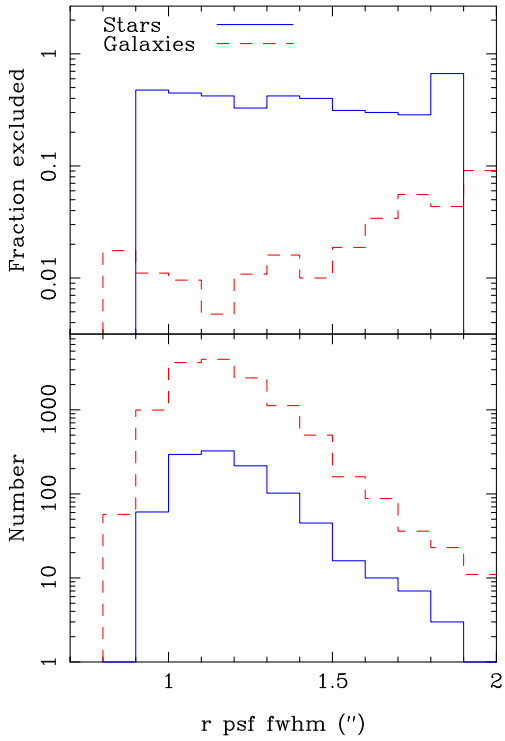

Figure 6. The dependence of the star galaxy separation on the FWHM of the imaging PSF. The top panel shows the fraction, and the bottom panel the number, of spectroscopically classified stars and galaxies in the commissioning data that are excluded by the additional z −band star-galaxy separation as functions of r−band FWHM.

simply by retroactively applying the z-band cut to the commissioning data.

Since these star–galaxy separation criteria measure the compactness of the objects in the SDSS imaging, their effectiveness will depend on the imaging PSF. Based on the commissioning data, Fig. 6 shows how the fraction of stars and galaxies removed by the new z-band criteria depends on the r-band PSF. The fraction of galaxies removed is fairly flat at ∼1 per cent for PSF FWHM <1.5 arcsec and then rapidly increases at higher FWHMs. The fraction of stars removed displays the opposite trend. Fig. 6 also presents the numbers of stars and galaxies as a function of r-band FWHM, demonstrating that the vast majority of the sample is selected from imaging with FWHM <1.5 arcsec. This slight seeing-dependent star–galaxy separation will result in the imprint of a spatial dependence in the density of galaxies across the survey, which can be corrected using seeing-dependent weights (see 6.4 for details).

Whilst the above analysis addresses the fraction of galaxies lost due to the addition of the z-band star galaxy separation criteria, it provides no indication of how many compact galaxies were removed by the original i-band cut. To investigate this issue, we combined the deep co-added SDSS Stripe82 imaging (Abazajian et al. 2009; Annis et al. 2014) with near-infrared J- and K-band imaging from the UKIDSS Large Area Survey Data Release 4 (Hewett et al. 2006; Casali et al. 2007; Lawrence et al. 2007; Hambly et al. 2008) in order to define a robust set of stars and galaxies over an area of 150 deg^2. A $J-K<1.1$ colour cut provides an excellent separation between stars and galaxies in the colour–magnitude region that the CMASS galaxies occupy. When this information is combined with the higher S/N measurement of $z_{\rm psf}-z_{\rm mod}$ from the co-added imaging we can confidently separate stars and galaxies. Using these data, we estimate that the final star–galaxy separation cuts removes 2.3 per cent of the full sample of galaxies selected by the CMASS colour cuts.

3.3.3 Summary of CMASS target selection

In summary, the CMASS target selection for the bulk of the survey is as follows:

$$i_{\rm cmod} < 19.86 + 1.6(d_\perp - 0.8) \quad (21)$$

$$17.5 < i_{\rm cmod} < 19.9 \quad (22)$$

$$d_\perp > 0.55 \quad (23)$$

$$i_{\rm psf} - i_{\rm mod} > 0.2 + 0.2(20 - i_{\rm mod}) \quad (24)$$

$$z_{\rm psf} - z_{\rm mod} > 9.125 - 0.46 z_{\rm mod} \quad (25)$$

$$r_{\rm mod} - i_{\rm mod} < 2 \quad (26)$$

$$i_{\rm fib2} < 21.5 \quad (27)$$

$$r_{{\rm dev},i} < 20.0\,{\rm pix}. \quad (28)$$

During commissioning (chunks 1 and 2), we used significantly looser criteria; the CMASS_COMM sample (BOSS_TARGET&&2^2), just under 25 000 galaxies, was selected as follows:

$$i_{\rm cmod} < 20.14 + 1.6(d_\perp - 0.8) \quad (29)$$

$$17.5 < i_{\rm cmod} < 20.0 \quad (30)$$

$$d_\perp > 0.55 \quad (31)$$

$$i_{\rm psf} - i_{\rm mod} > 0.2 + 0.2(20 - i_{\rm mod}) \quad (32)$$

$$r_{\rm mod} - i_{\rm mod} < 2 \quad (33)$$

$$i_{\rm fib2} < 22. \quad (34)$$

See other exceptions to these criteria in Section 3.3.1.

3.4 Sparse sampling cuts

Motivated by the wish to study objects of slightly lower stellar mass and bluer intrinsic colour, we designed the CMASS_SPARSE sample. It extends the CMASS selection by altering the $i_{\rm cmod}$–$d_\perp$ sliding colour–magnitude cut to

$$i_{\rm cmod} \geq 19.86 + 1.6(d_\perp - 0.8) \quad (35)$$

$$i_{\rm cmod} < 20.14 + 1.6(d_\perp - 0.8), \quad (36)$$

with the other cuts unchanged (i.e. the area between the red and green dashed lines in the bottom panel of Fig. 4). These galaxies were randomly subsampled down to a number density on the sky of 5 deg^{-2}, corresponding to approximately 1 in 10 targets. This sample was selected across the full BOSS footprint.

CMASS_SPARSE galaxies may be selected with

(i) BOSS_TARGET1 && 2^3
(ii) SPECPRIMARY == 1

(iii) ZWARNING_NOQSO == 0
(iv) CLASS_NOQSO == 'GALAXY'

after excluding the commissioning chunks.

Altering the CMASS target selection in this way produces a sample of galaxies at somewhat lower redshift and stellar mass. The median redshift of CMASS_SPARSE is $z = 0.51$, with a stellar mass distribution that peaks at $10^{11.2}$ $M_\odot$ (using the stellar masses of Chen et al. 2012), relative to the peak CMASS mass of $10^{11.4}$ $M_\odot$.

4 SPECTROSCOPIC OBSERVATIONS

4.1 Previously known redshifts

Fractions of the LOWZ and CMASS targets have a previous robust object classification and redshift determined from the SDSS-II survey (York et al. 2000; Abazajian et al. 2009). We therefore matched our target sample to a sample of 'known objects' with predetermined secure classifications and redshifts and did not spectroscopically reobserve these galaxies within BOSS. This subsample of targets has a complicated angular distribution on the sky: the majority of the NGC was covered by SDSS-II, but only a few stripes in the SGC were observed. These pre-observed targets account for 43 per cent (9 per cent) of the LOWZ targets in the north (south). A much smaller fraction of CMASS targets were pre-observed: 1.7 per cent (0.7 per cent) in the N (S).

4.2 Target collation and spectroscopic tiling

We start with the list of targets provided by the target selection algorithms detailed above, and remove targets with known redshifts as defined above. The *tiling algorithm* assigns the remaining targets to spectroscopic tiles. The sky was tiled in a piecemeal fashion as the survey progressed; each of these regions is called a 'chunk'; see Section 2.1 and Dawson et al. (2012) for further details. DR12 contains observations from 38 chunks. The survey mask and collated target catalogue both indicate the chunk to which a region or specific object was assigned.

The tiling algorithm (Blanton et al. 2003) determines the location of the 3° diameter spectroscopic tiles and allocates the available fibres among the targets, including targets from other programmes within BOSS. Because of the size of the cladding on the fibres, fibres may not lie within 62 arcsec of one another on a given spectroscopic tile. The algorithm therefore divides target galaxies into friends-of-friends groups with a linking length of 62 arcsec, and then assigns fibres to the groups in a way that maximizes the number of targets with fibres. The choice of which galaxies are assigned fibres is otherwise random. The algorithm adapts to the density of targets on the sky, such that regions with a larger than average number density tend to be covered by more than one tile. For the DR12 sample, 42 per cent (55 per cent) of the area in the north (south) is covered by multiple tiles, and the number density of CMASS targets is larger by 4.7 per cent (3.4 per cent) in those regions. The tile overlap – target density correlation is less pronounced for the LOWZ sample (1.6 and 2.4 per cent enhancement in north and south, respectively). The LOWZ sample constitutes only 35 per cent of the galaxy targets, and particularly in the north many galaxies in dense regions already have spectra from the SDSS-II and thus were not targeted for SDSS-III BOSS spectroscopy (see Section 4.1).

Fibre collisions are partially resolved only in the multiple tile regions, and therefore may not be representative of the unresolved fibre collisions in lower target density regions. Fibre-collided galaxies cannot simply be accounted for by reducing the completeness of their sector, since they are a non-random subset of targets (conditioned to have another target within 62 arcsec). As discussed further in Section 6.1, we provide a set of weights that treat these objects as if they were observed, and assign their weight to the nearest object of the same target class. Finally, since quasar targets are given higher priority by the tiling algorithm, we account for their presence by simply including a 62 arcsec veto mask (see Section 5.1.1) around each high priority quasar target.

4.3 Spectroscopic reductions

Each 'tile' output from the tiling algorithm specifies a central location on the sky and the list of targets to be observed. Physical plates are drilled at the University of Washington based on the anticipated airmass of observation. Multiple plates can cover the same tile, and plates may be observed on multiple nights until the desired signal-to-noise ratio is reached (Dawson et al. 2012).

The BOSS spectroscopic reduction pipeline is detailed in Bolton et al. (2012), with minor updates given in Alam et al. (2015). The final DR12 catalogues used the v5_7_0 tag of the IDLSPEC2D software package[5] for spectroscopic calibration, extraction, classification, and redshift analysis. We restrict the LSS catalogues to only include data from plates with PLATEQUALITY set to 'good.' The criteria for this designation are a minimum of three exposures, the number of spectroscopic pixels flagged as bad must be less than 10 per cent, and a minimum signal-to-noise ratio requirement for both the blue and red arms of the spectrograph must be met (Dawson et al. 2012).

The classification and redshift of each object are determined by a Maximum Likelihood fit of the co-added spectra to a linear combination of redshifted 'eigenspectra' in combination with a low-order polynomial. The polynomial (quadratic for galaxies, quasars, and cataclysmic variable stars; cubic for all other stars) allows for residual extinction effects or broad-band continua not otherwise described by the templates. The templates are derived from a rest-frame principal-component analysis (PCA) of training samples of galaxies, quasars and stars using stellar population templates at the BOSS resolution (from Maraston et al. 2013). The reduced χ^2 versus redshift is measured in redshift steps corresponding to the logarithmic pixel scale of the spectra, where $\Delta\log_{10}(\lambda) = 0.0001$. Galaxy templates are fit from $z = -0.01$ to 1.00, quasar templates from $z = 0.0033$ to 7.00, and star templates from $z = -0.004$ to 0.004 (± 1200 km s^{-1}). The template fit with the best reduced χ^2 is selected as the classification and redshift, with warning flags set for poor wavelength coverage, broken/dropped and sky-target fibres, and best fits which are within $\Delta\chi^2/\mathrm{dof} = 0.01$ of the next best fit (comparing only to fits with a velocity difference of more than 1000 km s^{-1}). This method is a development of that used for the SDSS DR8 (Aihara et al. 2011), and is explained in further detail in Bolton et al. (2012), and in Ahn et al. (2012, 2014).

For galaxy targets, a dominant source of false identifications is due to quasar templates with unphysical fit parameters, e.g. large negative amplitudes causing a quasar template emission feature to fit a galaxy absorption feature. Thus, for galaxy targets, the best classification and redshift are selected only from the fits to galaxy and star templates, and we restrict the sample to fits the pipeline classifies as robust. The results of these fits are tabulated in the '*_NOQSO' versions of various quantities in the LSS catalogues.

[5] http://www.sdss3.org/svn/repo/idlspec2d/tags/v5_7_0/

Table 2. Basic parameters of the DR12 CMASS,LOWZ, LOWZE2, and LOWZE3 samples. We track these classifications on a sector-by-sector basis in order to compute the BOSS fibre completeness in each sector of the survey. In this table, we report $\bar{N}_X = \sum_{\rm sectors} N_X$, the sum over all sectors retained in the final BOSS mask. Target classification counts and areas for the LOWZE2 and LOWZE3 samples are reported for chunk 2 and chunk 3–6, respectively. To estimate the target density for those samples, we use the full NGC footprint to reduce cosmic variance.

Property sample	NGC CMASS	SGC CMASS	Total	NGC LOWZ	SGC LOWZ	Total	NGC LOWZE2	NGC LOWZE3
$\bar{N}_{\rm gal}$	607 357	228 990	836 347	177 336	132 191	309 527	2985	11 195
$\bar{N}_{\rm known}$	11 449	1841	13 290	140 444	13 073	153 517	2730	6371
$\bar{N}_{\rm star}$	14 556	8262	22 818	1043	976	2019	24	61
$\bar{N}_{\rm fail}$	10 188	5157	15 345	868	602	1470	21	55
$\bar{N}_{\rm cp}$	34 151	11 163	45 314	4459	4422	8881	16	167
$\bar{N}_{\rm missed}$	7997	3488	11 485	10 295	3499	13 794	114	609
$\bar{N}_{\rm used}$	568 776	208 426	777 202	248 237	113 525	361 762	4336	15 380
$\bar{N}_{\rm obs}$	632 101	242 409	874 510	179 247	133 769	313 016	3030	11 311
$\bar{N}_{\rm targ}$	685 698	258 901	944 599	334 445	154 763	489 208	5890	18 458
Total area (deg^2)	7429	2823	10 252	6451	2823	9274	144	834
Veto area (deg^2)	495	263	759	431	264	695	10	55
Used area (deg^2)	6934	2560	9493	6020	2559	8579	134	779
Effective area (deg^2)	6851	2525	9376	5836	2501	8337	131	755
Targets/deg^2	98.9	101.1	99.5	55.6	60.5	57.0	43.4	23.5

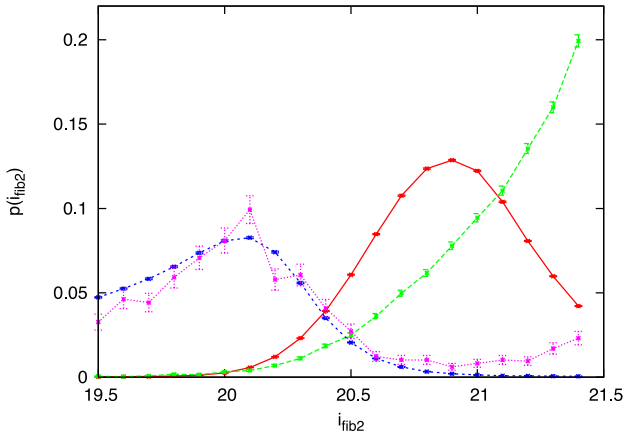

Figure 7. Normalised $i_{\rm fib2}$ distributions of redshift failures (green, dashed) and redshift successes (red, solid) for the CMASS sample. Redshift failures constitute 1.8 per cent of the CMASS targets observed by SDSS-III BOSS. These are contrasted against normalized distributions for the LOWZ sample of redshift failures (pink, dotted) and redshift successes (blue, dashed). Error bars were calculated assuming Poisson statistics. Note that some LOWZ galaxies have $i_{\rm fib2} < 19.5$, which is why the normalization for LOWZ curves looks lower than for CMASS.

Table 2 lists the total number of CMASS and LOWZ targets that were assigned a fibre within the survey footprint ($\bar{N}_{\rm obs}$) as well as the breakdown for each of the three possible outcomes: the number of CMASS and LOWZ targets robustly classified as stars ($\bar{N}_{\rm star}$) or galaxies ($\bar{N}_{\rm gal}$), and the number of targets for which the pipeline failed to find a robust classification and redshift ($\bar{N}_{\rm fail}$). A total of 2.3 per cent (3.4 per cent) of CMASS targets are stars and 1.6 per cent (2.1 per cent) are redshift failures in the north (south). Only 0.6 per cent of LOWZ targets are stars and 0.5 per cent are redshift failures.

Fig. 7 demonstrates that the pipeline is less likely to obtain a successful redshift for CMASS targets with fainter $i_{\rm fib2}$ magnitudes. Section 6.3 discusses how we account for this strong dependence in the redshift failure weights.

5 LSS CATALOGUE CREATION

The creation of the BOSS LSS catalogues involves a number of steps. We start with a list of targets based on the target selection procedure described above, with the previously known redshifts and outcome of the spectral analysis for each object for which we have a spectrum, matched to this list. Next we construct the survey mask, which specifies the regions of the sky that will be included in the LSS catalogues and the completeness in each included region. Finally, we use the mask and observed redshifts to generate a set of 'random' galaxies, Poisson sampling the sky coverage specified by the mask with the same expected density distribution as the galaxies. The random galaxies are assigned redshifts to match the distribution of the target sample. Together, the data and random catalogues can be used for statistical analyses such as N-point functions. These steps and some of the subtleties involved are now described in detail.

5.1 Mask

We use the MANGLE software (Swanson et al. 2008) to track the areas covered by the BOSS survey and the angular completeness of each distinct region; our terminology is summarized in Table 1. The mask is constructed of spherical polygons, which form the base unit for the geometrical decomposition of the sky. The angular mask of the survey is formed from the intersection of the imaging boundaries (expressed as a set of polygons) and the spectroscopic tiles. We define each unique intersection of spectroscopic tiles to be a sector (see Blanton et al. 2003; Tegmark et al. 2004; Aihara et al. 2011).

We compute sector completeness based on the distribution of targets across various outcomes of the tiling pipeline and spectroscopic reductions. In each sector (indexed by i) included in the LSS catalogue, we distinguish the following outcomes (separately for each target class):

(i) galaxies with redshifts from good BOSS spectra (we denote the number in each sector by $N_{\rm gal,i}$),
(ii) galaxies with redshifts from pre-BOSS spectra ($N_{\rm known,i}$),
(iii) spectroscopically confirmed stars ($N_{\rm star,i}$),
(iv) objects with BOSS spectra from which stellar classification or redshift determination failed ($N_{\rm fail,i}$),

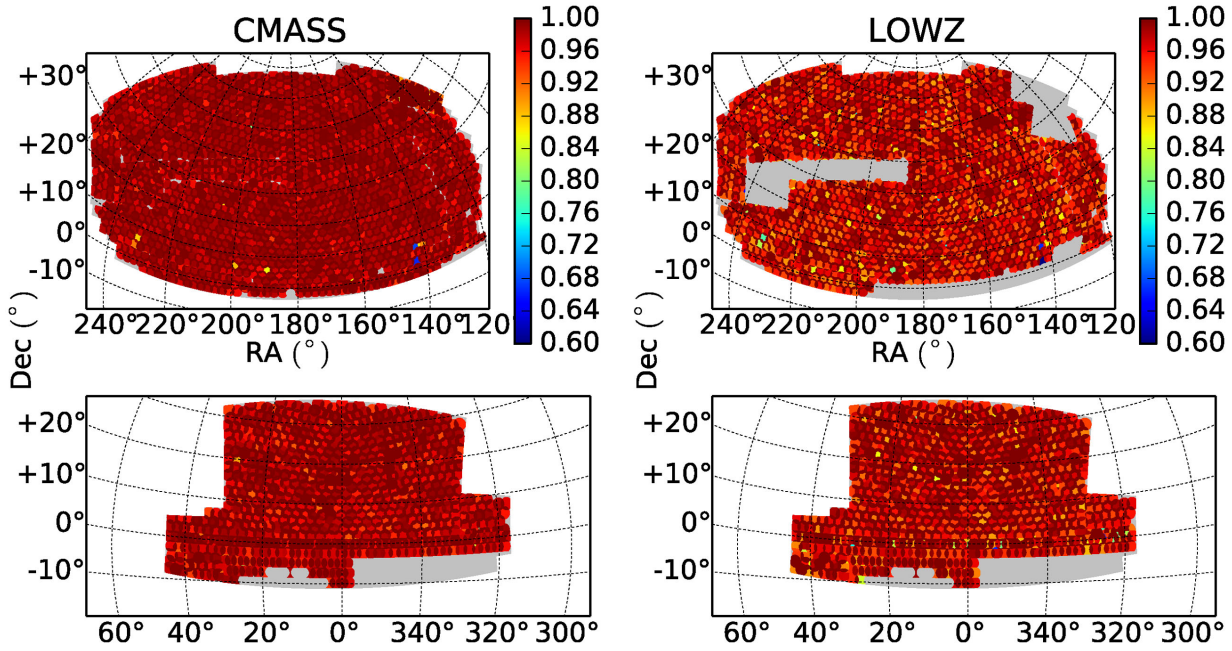

Figure 8. Completeness maps for both the LOWZ and CMASS samples in the north and south Galactic caps. The mean completeness is 98.8 per cent for the CMASS sample shown in the left-hand panels, and 97.2 per cent for the LOWZ sample in the right-hand panels. Gaps correspond to early chunks as shown in Fig. A1. Each patch of different colour corresponds to a plate, with the colour determined by the completeness of that plate. This is surrounded by the higher completeness regions that overlap that plate with other plates. This leaves a pattern that looks like a darker, higher completeness 'mesh', covering the survey.

(v) objects with no spectra, in a fibre collision group with at least one object of the same target class ($N_{\rm cp,i}$),[6]

(vi) objects with no spectra, if in a fibre collision group then with no other objects from the same target class ($N_{\rm missed,i}$).

These quantities, summed over all sectors included in the LSS catalogues, are given in Table 2. As each target is classed by one of these descriptors, we have that the total number of targets in sector i is

$$N_{\rm targ,i} = N_{\rm star,i} + N_{\rm gal,i} + N_{\rm fail,i} + N_{\rm cp,i} + N_{\rm missed,i} + N_{\rm known,i}, \quad (37)$$

and we define the number of targets observed by BOSS as

$$N_{\rm obs,i} = N_{\rm star,i} + N_{\rm gal,i} + N_{\rm fail,i}. \quad (38)$$

Matching our analyses for DR9, DR10 and DR11, the LOWZ catalogue is then cut to $0.15 < z < 0.43$, and the CMASS catalogue is cut to $0.43 < z < 0.7$ to avoid overlap, and to make the samples independent. The number of galaxies used in the final catalogue $N_{\rm used}$ is the subset of $N_{\rm gal,i} + N_{\rm known,i}$ that pass these redshift cuts.

From these descriptions, we define a BOSS fibre completeness in sector i

$$C_{\rm BOSS,i} = \frac{N_{\rm obs,i} + N_{\rm cp,i}}{N_{\rm star,i} + N_{\rm gal,i} + N_{\rm fail,i} + N_{\rm cp,i} + N_{\rm missed,i}}. \quad (39)$$

This completeness definition excludes the 'known' objects observed by SDSS-II. $C_{\rm BOSS,i}$, shown in Fig. 8, is recorded in the mangle mask files released with the LSS catalogues and is used in the random catalogue generation (see Section 5.2). By this definition, the area-weighted average completeness is 99 per cent (97 per cent) for the CMASS (LOWZ) samples. We compute the effective mask area in Table 2 by weighting the used area of each sector by its completeness.

The boundaries of the spectroscopic tiles can be seen by eye in Fig. 8 as discontinuities in the value of completeness; the unique intersection of those tiles define individual sectors, in which we treat the BOSS fibre completeness as uniform. On average, the completeness is larger in regions covered by more than one spectroscopic tile. The raw sky area covered by spectroscopic tiles is 10 338 deg^2, of which 10 252 deg^2 remain (7429 deg^2 in the NGC and 2823 deg^2 in the SGC) after restricting the mask to sectors for which every planned tile has been observed with 'good' PLATEQUALITY.

We also define a galaxy redshift completeness, assuming that stars are always correctly classified spectroscopically

$$C_{\rm red,i} = \frac{N_{\rm gal,i}}{N_{\rm obs,i} - N_{\rm star,i}}, \quad (40)$$

and define a target completeness

$$C_{\rm targ,i} = \frac{N_{\rm gal,i} + N_{\rm known,i}}{N_{\rm targ,i}}, \quad (41)$$

which gives the number of good galaxies spectroscopically observed in BOSS combined with previously known redshifts divided by the number of targets calculated in each sector. Fig. 9 shows the fraction of the total BOSS area that has target completeness greater than a specified value, and how this would change if we could ignore various effects. This shows the relative importance of different categories of targets to the target completeness of BOSS, from the least important, which is redshift failures, to fibre collisions, which is the most important.

Previous LSS catalogues (DR9, DR10, DR11) had to deal with sizeable regions where BOSS spectra were not complete, and we made a number of cuts on sectors to include in the LSS catalogues to minimize the impact of this effect. In particular, sectors meeting

[6] *cp* is used because each galaxy exists in a 'close-pair' with one another.

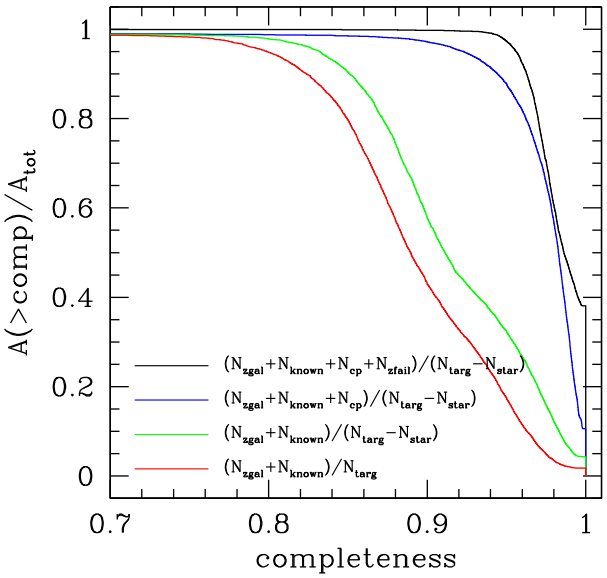

Figure 9. The fraction of the total survey area that has a target completeness greater than the value shown, where target completeness is defined as the number of good galaxies spectroscopically observed in BOSS and those with previously known redshifts divided by the number of targets calculated in each sector as $C_{\rm targ,i} = (N_{\rm gal,i} + N_{\rm known,i})/N_{\rm targ,i}$, as in equation (41). We compare this completeness with those we would have obtained had we not had to include various classes of targets. If there had been no stars in our target list, the completeness would have been $(N_{\rm gal,i} + N_{\rm known,i})/(N_{\rm targ,i} - N_{\rm star,i})$ (green line). If additionally we had not had to deal with fibre collisions, we would have observed a completeness $(N_{\rm gal,i} + N_{\rm known,i} + N_{\rm cp,i})/(N_{\rm targ,i} - N_{\rm star,i})$ (blue line), and if additionally there were no redshift failures $(N_{\rm gal,i} + N_{\rm known,i} + N_{\rm cp,i} + N_{\rm fail,i})/(N_{\rm targ,i} - N_{\rm star,i})$ (black line). From the definition of $N_{\rm targ,i}$ in equation (37), we see that the remaining decrement of the black line from $C_{\rm targ,i} = 1$ is due to missed galaxies $N_{\rm missed,i}$.

any of the following criteria were removed from the LSS mask:

(i) $C_{\rm BOSS,i} < 0.7$ (equation 39); removing part-complete sectors on the edges of the survey missing a significant fraction of redshifts.

(ii) $C_{\rm red,i} < 0.8$ (equation 40) and $N_{\rm gal,i} > 10$; removing regions with bad spectroscopic observations.

(iii) $N_{\rm obs,i} = 0$ and there is not another sector within 2° in the $\pm$ right ascension or declination directions; removing isolated regions without galaxies.

But this was not done for the DR12 sample. If we had additionally applied the fibre completeness cut (first criterion above), for DR12 we would have rejected an additional 30 (56) deg^2 from the CMASS (LOWZ) mask; if instead we had applied the redshift success cut in DR12 (second criterion above), we would have rejected an additional 1.7 (1.4) deg^2 from the CMASS (LOWZ) mask. The difference between the earlier mask selection and the algorithm described above applied to DR12 constitute negligible changes on the survey mask. The two algorithms agree to within 0.3 per cent of the total mask area for both the CMASS and LOWZ samples. Finally, the classification of $N_{\rm cp,i}$ and $N_{\rm missed,i}$ has slightly changed in DR12 relative to DR9–DR11; see Section 6.1.

5.1.1 *Veto masks*

While the basic geometry of the survey is encapsulated in the survey mask described in the previous sections, there remain many small regions within it where we could not have observed galaxies. Although they are individually small, they are not randomly distributed across the sky, and sum to a significant area, and so we exclude them from any analysis. We represent those regions by a set of veto masks, and remove 'randoms' that fall within these masks. The masks are as follows.

(i) Centerpost mask: each Sloan plate is secured to the focal plane by a central bolt: no targets coinciding with the centerpost of a spectroscopic tile can be observed. This mask reduces the survey area by 0.04 per cent.

(ii) Collision priority mask: Ly–α quasar targets receive higher priority than BOSS galaxy targets in the tiling algorithm; in regions of only a single spectroscopic tile, BOSS galaxy targets are unobservable within a fibre collision radius (62 arcsec) of those targets. Treating the high-priority quasar target locations as uncorrelated with the galaxy density field and neglecting any recovered galaxy targets in tile overlap regions, we can simply account for the high-priority quasars by masking a 62 arcsec radius around each. This mask reduces the survey area by 1.5 per cent.

(iii) Bright stars mask: we mask an area around stars in the Tycho catalogue (Høg et al. 2000) with Tycho $B_{\rm T}$ magnitude within [6,11.5] with magnitude-dependent radius

$$R = (0.0802 B_{\rm T}^2 - 1.860 B_{\rm T} + 11.625)\,\text{arcmin}. \quad (42)$$

This mask reduces the area by 1.9 per cent.

(iv) Bright objects mask: the standard bright star mask occasionally misses some bright stars that impact the SDSS imaging data quality. Additionally, a small number of bright local galaxies saturate the imaging as well, affecting target selection in their outskirts. These objects were identified by visual inspection, and the mask radii for each object were also determined in this manner, ranging from $0.^\circ1$ to $1.^\circ5$. The number of objects in this mask is $\sim$125, subtending a total area of 43.8 deg^2. The list of objects is described in section 2.1 of (Rykoff et al. 2014). This mask covers 0.4 per cent of the BOSS area.

(v) Non-photometric conditions mask: we mask regions where the imaging was not photometric in g, r, or i bands, the PSF modelling failed, the imaging reduction pipeline timed out (usually due to too many blended objects in a single field, caused by a high stellar density), or the image was identified as having any other critical problems. This mask reduces the area by 3.4 per cent.

(vi) Seeing cut: we discard regions where the point spread function full width half-maximum (labelled 'PSF_FHWM' in the catalogues) is greater than 2.3, 2.1, 2.0 in the g, r, and i band, respectively. The rationale for this cut is to decrease the variation of target density and properties with seeing due to the star galaxy separation (equations 12, 19 and 20) and $i_{\rm fib2}$ cuts. This cut removes an additional 0.5 per cent (1.7 per cent) of the NGC (SGC) footprint.

(vii) Extinction cut: for similar reasons, we also discard areas where the $E(B - V)$ extinction (labelled 'EB_MINUS_V' in the catalogues, from Schlegel et al. 1998) exceeds 0.15. This cut removes an additional 0.06 per cent (2.2 per cent) of the NGC (SGC) footprint.

In the catalogue creation pipeline, the list of targets is immediately passed through these veto masks, so that targets in vetoed regions do not contribute to the sector completeness calculation. All random galaxies within the veto regions must also be removed. Table 2 shows that in total, 6.6 per cent (9.3 per cent) of the area within the north (south) galactic cap footprint was removed by the veto masks.

5.2 Random catalogue generation

All of our clustering analyses make use of random catalogues with the same angular and redshift selection functions as the data. To produce these catalogues, we first use the MANGLE ransack command to generate one ≈10× and two ≈50× catalogues, where the angular density of the random galaxies is proportional to the completeness value in the mask for each sector.[7] As the random catalogue follows the redshift completeness per sector, it automatically corrects for any systematic effects caused by the decrease in fiducial exposure times starting roughly half-way through the BOSS survey. Next, we remove random galaxies using the set of veto masks described in Section 5.1.1. Only the angular coordinates of the 10× random catalogue are used to fit for angular systematic weights; see Section 6.4. Since the true underlying redshift distribution of our targets is unknown and can only be estimated from the empirical redshift distribution, we assign redshifts to the galaxies in the two 50× random catalogues by randomly drawing from the measured galaxy redshifts, but with a weight for each galaxy given by $w_{\rm tot,i}$, defined in equation (48). This procedure ensures that the (weighted) galaxy and random catalogues have exactly the same redshift distribution, apart from (small) stochasticity from the random redshift assignment. Ross et al. (2012) compare this random redshift assignment scheme with approaches that fit a spline of varying knot number to the measured galaxy redshift distribution, and then sample from the resulting spline directly. Based on analysis of mock catalogues, their fig. 19 demonstrates that the former method provides the smallest bias in fits to the monopole and quadrupole correlation function.

6 ACCOUNTING FOR OBSERVATIONAL ARTEFACTS IN LSS CATALOGUES

In this section, we describe in detail how we weight the targeted galaxies when computing LSS statistics, in order to minimize the impact of observational artefacts on our estimate of the true galaxy overdensity field. We identify various effects that affect the completeness of the sample, which we quantify with weights applied per sector. These weights are a development of those presented in Anderson et al. (2012, 2014b). In particular, we discuss treatment of 'known' redshifts from SDSS-II that were not re-observed in SDSS-III BOSS, galaxies not observed due to fibre collisions, observed galaxies for which a robust redshift was not obtained, and a weighting scheme to null non-cosmological fluctuations imprinted on the catalogue by the target selection step. The weights described below are available for each galaxy in the LSS catalogues. In this section, we also summarize weights we apply to minimize our statistical error on the observed power spectrum.

6.1 Fibre collision corrections

Galaxies that were not assigned a spectroscopic fibre due to fibre collisions are not a random subsample of the full target sample since they are within a fibre collision radius (62 arcsec) of another target. This is potentially a large effect; in the SGC, where the coverage of known targets from SDSS-II is lowest, approximately 20 per cent of galaxy targets are in a collision group containing other CMASS or LOWZ galaxy targets. As a result, 5.8 per cent of CMASS targets and 3.3 per cent of LOWZ targets were not assigned a spectroscopic fibre.

These objects preferentially occupy denser environments and therefore have higher than average large-scale bias. They are also more likely than average to occupy the same dark matter halo as a neighbouring galaxy target. Accurate fibre collision corrections are therefore particularly important for applications relying on the absolute value of galaxy bias (i.e. in a comparison of the lensing and clustering amplitude) or those that use small-scale clustering to deduce halo occupation statistics and satellite fractions.

In the default LSS catalogue that focuses on obtaining unbiased galaxy density fields on large scales, we simply upweight the nearest galaxy from the same target class that was assigned a fibre to account for collided galaxies that were not assigned fibres. This information is tracked by incrementing a weight $w_{\rm cp}$, labelled WEIGHT_CP in the DR12 LSS catalogues. The upweighted nearest neighbour could be classified by the spectroscopic pipeline as a good galaxy redshift, a star, or a redshift failure. Upweighting the neighbour without reference to its classification is the appropriate thing to do as the missed object could be in any of these classes.

We correct 34151 (11163) CMASS targets and 4459 (4422) LOWZ targets by nearest neighbour upweighting in the NGC (SGC). This amounts to 5.0 per cent (4.3 per cent) of CMASS targets in the NGC (SGC), and 1.3 per cent (2.9 per cent) of the LOWZ targets. The difference between the hemispheres is due to both higher tile density in the SGC (so more fibre collisions fall in overlap regions where they can be partially resolved) and to most of the previously known SDSS-II redshifts falling in the NGC.

The algorithm used to generate the DR12 catalogues differs slightly from the one used for the DR9–DR11 catalogues. The new algorithm uses the output from the tiling algorithm to determine membership in fibre collision groups. Targets with the same 'FINALN' and 'INGROUP' field flags output from the tiling code share a collision group. We choose the nearest object of the same target class and collision group to carry the weight of the unobserved target. We also allow 'known' galaxies to carry the weight if they are closer than all BOSS-observed targets. In DR9–DR11 catalogues, we did not refer to the fibre collision group indices, but simply identified collision pairs in the same target class if they were separated by less than 62 arcsec. None the less, the two algorithms select the same nearest neighbour ∼94 per cent of the time.

Our adopted fibre collision correction scheme neglects a few subtle cases.

(i) No corrections are applied for objects that are the only members of their target class in their fibre collision group, and did not receive a fibre. For CMASS, this class represents 4 per cent of all targets in fibre collision groups, and 0.7 per cent of all CMASS targets overall. Since there are more CMASS targets per unit area, this effect is larger for LOWZ targets: 12 per cent of all collided LOWZ targets and 1.4 per cent of the full sample. Treating such collision pairs as unassociated is still a good approximation. To verify this assumption, we examined all collision groups consisting of a single LOWZ target and single CMASS target, and for which we obtained both redshifts. Only 11 per cent of such pairs had line-of-sight separations smaller than 50 h^{-1}Mpc.

(ii) No corrections are applied when none of the multiple objects of the same target class in a fibre collision group were assigned a fibre. These galaxies are treated as random incompleteness in the survey coverage and comprise 0.14 per cent of the total galaxy sample.

[7] To exactly reproduce the officially released random catalogues, one must use the ransack version included in the SDSS idlutils product with version v5_4_25 or higher (Surhud More, private communication). Random seeds input to ransack are provided in the catalogue generation scripts accompanying MKSAMPLE.

Table 3. Distribution of galaxies across fibre collision group sizes. The largest collisions group (not listed) contains 17 galaxy targets. The first column provides the fraction of CMASS targets in groups with $n_{\rm group}$ CMASS targets, restricted to groups with at least two CMASS targets. The second column shows the same calculation for LOWZ targets. The final column lists the fibre collision group size distribution, where $n_{\rm group}$ includes both CMASS and LOWZ targets. For consistency across the mask these results were computed from the LOWZ sample footprint (chunks ≥ 7). For reference, the fraction of galaxies that are not in any collision group is 77 per cent.

$n_{\rm group}$	$f_{\rm CMASS}$	$f_{\rm LOWZ}$	$f_{\rm C+L}$
2	0.7631	0.8456	0.7566
3	0.1687	0.1182	0.1726
4	0.0440	0.0270	0.0461
5	0.0146	0.0070	0.0150
6	0.0059	0.0010	0.0057
7	0.0023	0.0005	0.0022
8	0.0007	0.0003	0.0008
9	0.0003	0.0006	0.0003

Table 4. Fibre collision statistics for targets in regions covered by $n_{\rm tiles}$ spectroscopic tiles. The second column shows the fraction of the total mask area covered by $n_{\rm tiles}$ tiles. The third column gives $f_{\rm fibre}$, the fraction of all collided galaxies that were assigned a fibre. The remaining columns specify the fraction of pairs of galaxy targets (CMASS + LOWZ) in the same collision group for which both targets received a fibre, both globally ($c_{\rm pair}$), and as a function of $n_{\rm group}$. We use the global fraction $c_{\rm pair}$ to remove collided pairs and approximate the fibre collision effect in our mock galaxy catalogues. We track $c_{\rm pair}$ separately for the NGC and SGC and for CMASS, LOWZ or combined catalogues, but in practice the values are similar in each case to those reported here.

$n_{\rm tiles}$	$f_{\rm area}$	$f_{\rm fibre}$	$c_{\rm pair}$	$c_{\rm pair}(2)$	$c_{\rm pair}(3)$	$c_{\rm pair}(4)$
1	0.54	0.561	0.092	0.042	0.159	0.142
2	0.41	0.945	0.820	0.971	0.685	0.589
3	0.05	0.992	0.966	0.992	0.985	0.915
4	0.0005	1.000	1.000	1.000	1.000	–

(iii) Finally, 0.3 per cent of targets did not receive a fibre due to collisions with targets other than CMASS and LOWZ but of the same priority. Again, we treat these missing redshifts as random.

Tables 3 and 4 provide statistics about the distribution of CMASS and LOWZ galaxies in fibre collision groups and how the probability of assigning a fibre to a pair of collided galaxies in the same fibre collision group depends on the size of the collision group. Approximately 75 per cent of collided galaxies are in a group of only two, and group sizes above four are quite rare. In Table 4, $f_{\rm fibre}$ reports the fraction of galaxies in a collision group that received a spectroscopic fibre, as a function of $n_{\rm tiles}$, the number of spectroscopic tiles covering their sector. In the remaining columns, we report the fraction of pairs of CMASS+LOWZ targets in the same collision group for which both targets received a fibre, both globally ($c_{\rm pair}$), and as a function of $n_{\rm group}$. In regions covered by a single spectroscopic tile, only a small fraction of pairs with $n_{\rm group} = 2$ both receive a spectroscopic fibre (4 per cent). Such pairs must be sourced from collision groups containing at least one target of another class, oriented such that the two CMASS/LOWZ targets in the group are separated by more than 62 arcsec. As expected, for $n_{\rm tiles} > 1$ pairs in smaller collision groups are more likely to be resolved, and the majority of fibre collisions are removed.

Finally, to understand the impact of fibre collision corrections on our estimate of the true galaxy density field, we examine the apparent separation for pairs of galaxies in the same fibre collision group for which good redshifts were obtained for both. Fig. 10 shows the distribution near $\Delta z \approx 0$, although the tails extend to much larger separations. We have converted redshift separations to apparent distance separations using the fiducial cosmological model. The observed distribution (coloured lines) can be fitted by a flat background and an exponential distribution centred on $\Delta z = 0$ (black lines). The fraction of resolved fibre collision pairs that are 'correlated' (i.e. contribute to the exponential component in the fit to the pairwise separation histogram) is 52 per cent for pairs of CMASS targets and 62 per cent for pairs of LOWZ targets, i.e. nearly half of fibre collision pairs are unassociated projections. Interestingly, the width of the distribution is consistent with $\sigma_{\rm exp} = 5.4\ {\rm h^{-1}Mpc}$ for both target classes and is generally consistent with halo modelling expectations.

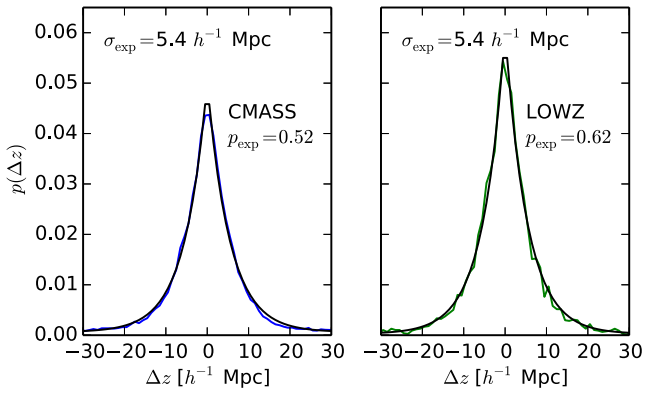

Figure 10. The probability distribution of apparent line-of-sight separations for pairs of galaxies in the same fibre collision group and for which both have good redshifts. The left-hand panel uses pairs of CMASS targets and the right-hand panel uses pairs of LOWZ targets. Both distributions can be fitted with the sum of a background term and an exponential: $ae^{-|\Delta z|/\sigma} + b$ in the range $|\Delta z| < 50\ {\rm h^{-1}Mpc}$. A total of 52 per cent (62 per cent) of the CMASS (LOWZ) pairs contribute to the exponential term. The best-fitting width σ of the exponential component is $5.4\ {\rm h^{-1}Mpc}$ for both CMASS and LOWZ targets.

Since the choice of which galaxies are assigned fibres in a collision group is completely random (apart from maximizing the number of targets receiving a fibre), the object not assigned a fibre is statistically equivalent to the one we upweight, and so once upweighted correlations at transverse separations larger than the fibre collision scale should be unbiased. However, correlations at transverse separations below the collision scale will be biased, since we are removing these small-scale pairs. Additionally, these small-scale variations will be anisotropic, and therefore likely to have a stronger effect on the quadrupole, rather than monopole moments of two-point clustering statistics, for example. We therefore advocate constructing statistics that do not apply these weights in situations where these effects are important; see Reid et al. (2014) for an example configuration space statistic.

6.2 Treatment of 'known' targets

As the pre-observed 'known' sample is complete (no failures are kept), it does not match the angular distribution induced by variations in completeness of the galaxies spectroscopically observed by BOSS. Rather than try to model the distribution of known galaxies, we instead subsampled these data to match BOSS completeness in each sector, thus imposing the BOSS mask on the known galaxies.

In this way, we make the sample indistinguishable from BOSS-observed targets. In earlier data releases (DR9-11), we also marked a fraction of the galaxies in a 62 arcsec close pair containing at least one object from the 'known' sample as fibre-collided; we did not apply this step in our DR12 analysis and describe the difference in more detail in the next section.

In DR9-DR11 catalogues, we additionally marked a fraction of the galaxies in a 62 arcsec close pair containing at least one object from the 'known' sample as fibre-collided, and assigned its weight to its nearest neighbour. This fraction was determined by measuring the fraction of 62 arcsec BOSS targets that were fibre collision corrected in each sector. In sectors covered only by a single spectroscopic tile all 62 arcsec pairs were collided. The original motivation of this correction was to impose the same fibre collision completeness on the 'known' targets as the BOSS targets. In DR12, we did not apply this correction. The rationale was that on sufficiently large scales the nearest neighbour upweighting scheme restores the correct clustering statistics, and so should therefore be equivalent to using the measured redshifts. However, we expect the effective shot noise to be larger when using the former procedure. Correlation function and power spectrum analyses that marginalize over a shot noise term should be unaffected by this choice; analyses of smaller scale clustering should examine this issue further. This change is particularly important for clustering of the LOWZ sample because of the large overlap with the 'known' galaxy sample.

6.3 Redshift failures

For 1.8 per cent (0.5 per cent) of CMASS (LOWZ) targets, the spectroscopic pipeline fails to obtain a robust redshift. We do not necessarily expect these to be distributed randomly with respect to e.g. plate centre or redshift, and so we again adopt a nearest neighbour upweighting scheme to account for these objects. Redshift failure galaxies were permitted to be upweighted because of a nearest neighbour fibre collision. We therefore transfer the total weight to the nearest neighbour of the redshift failure, incrementing a weight $w_{\rm noz}$, labelled WEIGHT_NOZ in the DR12 LSS catalogues. The upweighted object must be classified either as a good galaxy or star redshift.

In DR9–DR11 LSS catalogues, we removed sectors with redshift success rates below 80 per cent and at least 10 good redshifts; in our DR12 catalogue, we exclude troublesome observations by restricting mask regions with PLATEQUALITY of 'good', and do not remove the handful of sectors that would have been excluded using the DR9–11 criteria. Upon closer examination, we found that sectors failing the DR9–11 cut contained a small number of targets and therefore subject to small number statistics; we checked that targets in those sectors were drawn from plates with high-redshift success rates.

In DR9–11, we searched for redshift failure neighbours to upweight only in the same sector; in the DR12 catalogue, we only consider neighbours observed on the same plate (which spans multiple sectors) and same date, which restricts the neighbour search to galaxies observed under approximately the same conditions, and means the weighted number of classified objects in each sector matches the number of targets. The majority of close neighbours restricted to the same sector versus restricted to the same plate and date are the same neighbour. The median angular separation between galaxies without a good redshift and their closest neighbour using the updated algorithm is 3.7 arcmin (3.9 arcmin) in the north (south), compared with 2.9 arcmin using the sector-based algorithm. Total counts of redshift failures for CMASS and LOWZ galaxies are listed in Table 2.

In CMASS, redshift failures are more likely to occur on faint targets – see Fig. 7. In the weighting scheme described above the neighbouring, up-weighted, galaxies are drawn from the distribution of observed galaxies, which in turn are brighter on average than the galaxies that failed to yield a good redshift. Given the slight correlation of $i_{\rm fib2}$ with redshift, this introduces a small redshift-dependent bias on the LSS catalogues. To ameliorate this effect, we modify the redshift-failure weights such that the *weighted* distribution of $i_{\rm fib2}$ of the corrective weight matches the $i_{\rm fib2}$ of the targets with failed redshifts. In practice, acknowledging that an up-weighted galaxy might be a neighbour to more than one redshift failure, we compute $w_{\rm noz,new} = 1 + (w_{\rm noz,old} - 1)w_{\rm ifib2}$, where $w_{\rm ifib2} = n(i_{\rm fib2,noz})/n(i_{\rm fib2,cp})$ with $n(i_{\rm fib2,noz})$ and $n(i_{\rm fib2,cp})$ corresponding to the green and red lines of Fig. 7, respectively. To avoid $w_{\rm ifib2}$ being dominated by Poisson noise in any given bin of $i_{\rm fib2}$, we set $w_{\rm ifib2} = 1$ for any bin where $n(i_{\rm fib2,noz})$ or $n(i_{\rm fib2,cp})$ are less than 10. The weights are normalized such that $\sum w_{\rm noz,new} = \sum w_{\rm noz,old}$. This scheme effectively transfers weight from bright to faint neighbours of redshift-failure weights. We only apply this extra correction to CMASS for two reasons: first, the LOWZ redshift-failure rate is very small (0.5 per cent) and, secondly, we find no significant dependence of redshift failure with $i_{\rm fib2}$ for LOWZ targets.

6.4 Angular systematic weights

For the DR12 data, we follow the same approach as described in Ross et al. (2012) and updated in Anderson et al. (2014b) to remove non-cosmological fluctuations in CMASS target density with stellar density and seeing. The LOWZ targets are brighter and do not show significant variations with these quantities, so LOWZ targets do not require these weights.

In DR12, we update the $n_{\rm side} = 128$ HEALPIX[8] (Górski et al. 2005) stellar density map to include all stars with i-band magnitudes between 17.5 and 19.9; the map used in DR10/DR11 did not impose the 17.5 bright cut. The two maps also differ by a factor of the pixel area, 0.210 deg^{-2}. The functional form for $w_{\rm star}$ was also updated in DR12 to be the inverse of a linear relation:

$$w_{\rm star}(n_{\rm s}, i_{\rm fib2}) = (A_{i_{\rm fib2}} + B_{i_{\rm fib2}} n_{\rm s})^{-1}, \quad (43)$$

while in DR10/DR11 $w_{\rm star}$ was linearly dependent on $n_{\rm s}$; see Ross et al. (in preparation) for details. These two differences explain the changes to the values of the $A_{i_{\rm fib2}}$ and $B_{i_{\rm fib2}}$ parameters between DR10/DR11 and DR12. The DR12 parameter values for $w_{\rm star}$, determined using all galaxies in the CMASS catalogue with $0.35 < z < 1.0$, are $A_{i_{\rm fib2}} = [0.959, 0.994, 1.038, 1.087, 1.120]$ and $B_{i_{\rm fib2}} = [0.826, 0.149, -0.782, -1.83, -2.52] \times 10^{-4}$, computed in 0.3 mag width $i_{\rm fib2}$ bins centred at [20.45, 20.75, 21.05, 21.35], as in Anderson et al. (2014b). The parameter $w_{\rm star}$ is determined for each galaxy by first linearly interpolating the $A_{i_{\rm fib2}}$ and $B_{i_{\rm fib2}}$ fits to derive a value at each galaxy's $i_{\rm fib2}$, and then using equation (43). The distribution of weight values is similar in the NGC and SGC and, overall, 93 per cent of CMASS galaxies have $0.95 < w_{\rm star} < 1.1$.

For DR10/DR11 analyses, a map of the DR8 i-band seeing, S_i, was created by taking the mean seeing value within HEALPIX pixels with $n_{\rm side} = 1024$ over the primary SDSS galaxies in the DR8 Catalogue Archive Server. For DR12, we instead directly query

[8] http://healpix.jpl.nasa.gov/

the imaging data to determine the conditions estimated for each galaxy's parent imaging field. Per-object and per-field seeing estimates are calculated differently. Empirically, these two methods for determining S_i differ by a factor of ∼0.9. There is also scatter between per-field and per-object estimates of sky flux and airmass. The DR12 galaxy and random catalogues contain fields for 'PSF_FWHM', 'AIRMASS', 'SKYFLUX', 'EB_MINUS_V', and 'IMAGE_DEPTH' if users want to further explore systematics relationships. In what follows, the i-band seeing S_i = PSF_FWHM[3]. For DR12, we adopted a slightly different parameter convention from that of earlier catalogues:[9]

$$w_{\rm see}(S) = A_{\rm see}^{-1}\left[1 - {\rm erf}\left(\frac{S_i - B_{\rm see}}{\sigma_{\rm see}}\right)\right]^{-1}. \quad (44)$$

In addition, we fit the systematic relationship separately for the NGC and SGC, again restricting the fits to objects in the CMASS LSS catalogues with $0.35 < z < 1.0$. The DR12 parameter values are $A_{\rm see} = 0.5205$ (0.5344), $B_{\rm see} = 2.844$ (2.267), and $\sigma_{\rm see} = 1.236$ (0.906) for the NGC and SGC, respectively. In DR10/DR11, we also set $w_{\rm see}(S_i > 2.5\,{\rm arcsec}) = w_{\rm see}(S_i = 2.5)$ arcsec; this action is no longer necessary since the DR12 veto masks remove all area with $S_i > 2.0$ arcsec.

Finally, the application of the CMASS z-band star/galaxy separation cut in the LOWZE3 sample induced a significant dependence on the sample number density with S_i that varies with the i-band model magnitude; see Ross et al. (in preparation) for details. The systematic weight for this sample is

$$\ell = \max\left(-2, b + m(i_{\rm mod} - 16)^{-0.5}\right) \quad (45)$$

$$w_{\rm see,LOWZE3} = \min\left(5, (1. + (S_i - 1.25)\ell)^{-1}\right) \quad (46)$$

with parameters $b = 0.875$ and $m = -2.226$, fit using all objects in the LOWZE3 catalogue with $0.2 < z < 0.5$, including objects in chunks ≥6 in addition to the LOWZE3 targeted region, chunks 3–6.

The total angular systematic weights are simply the product of $w_{\rm star}$ and $w_{\rm see}$ for each object with index i:

$$w_{\rm systot,i} = w_{\rm star,i} w_{\rm see,i}. \quad (47)$$

6.5 Total galaxy weights

Finally, we combine the angular systematics weight $w_{\rm systot,i}$ with the fibre collision and redshift failure nearest neighbour weights to produce a final weight for each object i in the final catalogue:

$$w_{\rm tot,i} = w_{\rm systot,i}(w_{\rm cp,i} + w_{\rm noz,i} - 1). \quad (48)$$

Since the default values of both $w_{\rm cp,i}$ and $w_{\rm noz,i}$ are 1, the term in parentheses conserves the total number of galaxy targets. This is the galaxy weighting consistent with the construction of the LSS catalogues provided, and must be used to obtain unbiased estimates of the galaxy density field, since this weight is used when assigning the random galaxy redshifts; see Section 5.2.

6.6 Angular density and redshift distribution

We estimate the angular density of galaxy targets as the total number of targets within the final LSS mask divided by the total non-vetoed area within the sample LSS mask. The values for each target class are listed in the final line of Table 2.[10] We convert this angular target density into a three-dimensional space density through a properly normalized redshift probability distribution:

$$p(z_j, z_j + {\rm d}z){\rm d}z \propto \sum_{z_i \in [z_j, z_j + {\rm d}z]} w_{\rm tot,i} / \sum w_{\rm tot,i}, \quad (49)$$

where we sum over all objects in the catalogue with good spectroscopic redshifts, and $w_{\rm tot,i}$ is the total weight assigned to target i to account for various observational artefacts (equation 48). The inclusion of $w_{\rm systot,i}$ in the estimate for $p(z)$ accounts for any impact of the angular systematics on the (normalized) redshift distribution, through e.g. the $i_{\rm fib2}$ dependence of the stellar weights. However, our estimator for the angular target density does not recover the true target density in the absence of stars and imperfect seeing, but an average target density over the survey footprint. Finally, we use the fiducial cosmology to determine the number of targets per $h^{-3}{\rm Mpc}^3$. The result is shown in Fig. 11 for all four target classes, as well as the sum of the CMASS and LOWZ sample number densities (with duplicate CMASS and LOWZ targets counted only once). The CMASS+LOWZ number density reaches a local minimum in the overlap region of $\bar{n}(z = 0.41) = 2.2 \times 10^{-4}\ h^{-3}{\rm Mpc}^3$. As reported in the previous sections, survey incompleteness, fibre collisions, redshift failures, and stars in the target sample all reduce the average angular density of good galaxy redshifts compared to the angular target density; their aggregate impact is a 10 per cent (4.4 per cent) reduction for CMASS (LOWZ). Finally, we compute

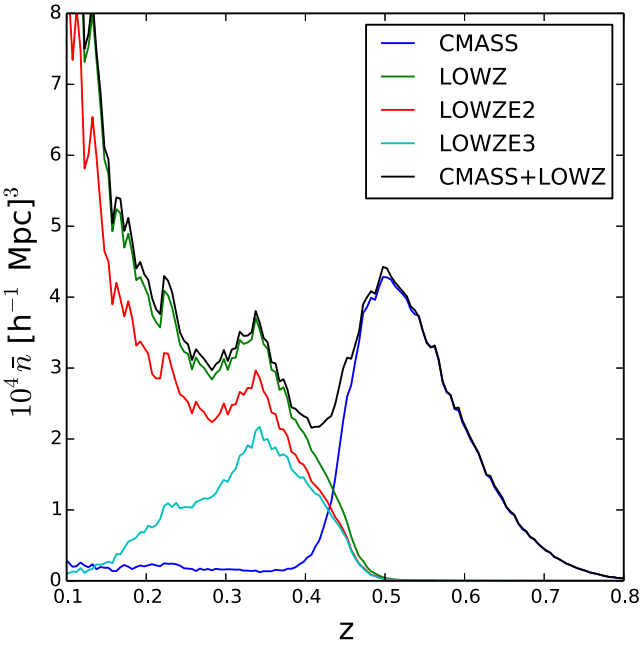

Figure 11. Number density of all four target classes assuming our fiducial cosmology with $\Omega_{\rm m} = 0.31$, along with the sum of the CMASS and LOWZ number densities (black).

[9] Equation (19) of Anderson et al. (2014b) should state $w_{\rm see}(S) = 2A_{\rm see}^{-1}[1 - {\rm erf}(\frac{S_i - B_{\rm see}}{\sigma_{\rm see}})]^{-1}$ for the parameter values listed in that text.

[10] Our calculation of the 'NBAR' field in the galaxy and random catalogues estimates the angular density of the sample as $A_{\rm eff}^{-1}\sum_i(w_{\rm cp} + w_{\rm noz} - 1)$. Here, $A_{\rm eff}$ is the completeness weighted area inside the mask and the sum is over all galaxies in the catalogue with good redshifts. This method is slightly noisier since the completeness in each region is estimated from a finite number of galaxies. We verified that the two methods agree to within 0.02 per cent.

the effective volume $V_{\rm eff}$, which quantifies the reach of a sample for making cosmological measurements, for the CMASS and LOWZ samples following the same algorithm outlined in Anderson et al. (2014b), summing over 200 redshift shells

$$V_{\rm eff} = \sum_i \left(\frac{\bar{n}(z_i)P_0}{1+\bar{n}(z_i)P_0} \right)^2 \Delta V(z_i), \quad (50)$$

where $\Delta V(z_i)$ is the volume of the shell at z_i, and we assume that $P_0 = 10\,000\,h^{-3}{\rm Mpc}^3$, which we have changed since DR11, so the numbers are not directly comparable to Anderson et al. (2014b). We find $V_{\rm eff} = 5.1$ Gpc^3 for CMASS and 2.3 Gpc^3 for LOWZ.

6.7 FKP weights

Feldman, Kaiser & Peacock (1994), hereafter FKP, showed that the optimal weighting of galaxies as a function of redshift depends on the number density of galaxy tracers. The optimal weight $w_{\rm FKP}$ depends on the amplitude of the power spectrum in the power spectrum bin of interest. In practice, we use the same value $P_0 = 10\,000$ h^{-3}Mpc^3 to estimate both the power spectrum and correlation function on all scales. This value of P_0 corresponds to the observed power spectrum at $k \approx 0.15\ h$ Mpc^{-1}. The field 'WEIGHT_FKP' in the DR12 galaxy and random catalogues is given by

$$w_{\rm FKP,i} = \frac{1}{1+\bar{n}(z_i)P_0} \quad (51)$$

for an object with redshift z_i, where $\bar{n}(z_i)$ is computed by linear interpolation over bins with $\Delta z = 0.005$ starting at $z = 0$. The $w_{\rm FKP}$ weight is optional in LSS analyses. To utilize these weights in a LSS analysis, one must weight both data and random objects; the final weight of galaxy i is therefore $w_{\rm tot,i}w_{\rm FKP,i}$ and the final weight of random object j is $w_{\rm FKP,j}$. If one does not use the FKP weights (i.e. as in Reid et al. 2014), consistent weightings of the galaxy and random catalogues are $w_{\rm tot,i}$ and $w_j = 1$, respectively.

Earlier data releases adopted a different fiducial cosmology and assumed $P_0 = 20\,000\,h^{-3}{\rm Mpc}^3$ to compute $w_{\rm FKP}$. Percival, Verde & Peacock (2004) updated the analysis of Feldman et al. (1994) to a weighting scheme that accounts for luminosity-dependent clustering; such weights will be presented for the BOSS galaxy samples in a forthcoming BOSS team paper. However, because our target selection algorithm is so efficient at selecting massive galaxies, the gain provided by luminosity-dependent weights is modest for our sample.

7 COMBINED CATALOGUE CREATION

For the purpose of providing a maximally contiguous three-dimensional density field estimate, in DR12 we provide a new catalogue that combines the CMASS sample with the three lower redshift samples: LOWZE2 (chunk 2), LOWZE3 (chunks 3–6), and LOWZ (chunks $\geq$7). See Appendix A for details of the LOWZE2 and LOWZE3 samples. A precise geometric description of the sky area covered by each sample is provided in mangle mask format, constructed such that every sector included in the CMASS mask is included in exactly one of the LOWZE2, LOWZE3, or LOWZ footprints. We also construct two additional masks, one including the LOWZE2 + LOWZ sky coverage and another including the LOWZE3 + LOWZ sky coverage.

Using those masks, we first generate a LOWZE2 catalogue including chunk 2 and chunks $\geq$7 and a LOWZE3 catalogue including chunks $\geq$3 using the target selection algorithms detailed in Appendix A. This is possible since all the galaxies passing LOWZE2 and LOWZE3 cuts will also pass the LOWZ cuts. Producing a catalogue across a larger fraction of the sky allows a more accurate estimate of $\bar{n}(z)$ for the LOWZE2 and LOWZE3 samples (and therefore a better means of assigning redshifts to the random galaxy sample). Without this step, the average density in chunk2 and chunks 3–6 would be poorly determined and could lead to erroneous reconstruction flows towards or away from those regions in the final combined catalogues. As discussed in Section 6.4 and Ross et al. (in preparation), there is a significant correlation between i-band seeing and LOWZE3 target density which we remove using a systematic weight given by equation (46); LOWZE2 and LOWZ samples require no systematic weight corrections. We follow this same procedure with some minor but important differences when combining CMASS and LOWZ catalogues. After full footprint data and random catalogues are produced, we trim each catalogue back to its original targeted region (i.e. LOWZE2 in chunk 2, LOWZE3 in chunks 3–6, and LOWZ in chunks $\geq$7) using the mutually exclusive masks discussed above.

Our algorithm to generate the combined catalogue from the four different samples (CMASS, LOWZ, LOWZE2, LOWZE3) is as follows.

(i) Renormalize the CMASS galaxy systematic weights $\tilde{w}_{\rm systot,i} \propto w_{\rm systot,i}$ such that

$$1 = \frac{\sum_i \tilde{w}_{\rm systot,i}(w_{\rm cp,i} + w_{\rm noz,i} - 1)}{\sum_i (w_{\rm cp,i} + w_{\rm noz,i} - 1)}. \quad (52)$$

This ensures that in the combined catalogue, a CMASS target and a LOWZ target on average have equal weight in each of the three distinct regions. The functional form chosen for $w_{\rm see}$ and $w_{\rm star}$ does not guarantee this normalization. Fibre collision and redshift failure weights are left the same as in the original CMASS-only catalogue and the parameters for the systematic weights are identical to the ones in the CMASS-only catalogue (apart from the renormalization).

(ii) For each of LOWZ, LOWZE2, LOWZE3 samples ('LOWZX'), read in the targets (including those in chunks $\geq$7), and remove objects already in the CMASS catalogue. Duplicate targets are 2.6, 2.4, and 4.4 per cent of the LOWZ, LOWZE2, and LOWZE3 samples, respectively. Fibre collision and redshift failure weights are then recomputed on each duplicate excluded LOWZX sample. As in the previous catalogues, fibre collision and redshift failure weights are only assigned to other LOWZX targets (not CMASS targets). For the LOWZE3 sample, systematic weights are assigned using the same parameters as the LOWZE3-only sample, but renormalized as in equation (52).

(iii) Concatenate the CMASS and LOWZX samples and compute the completeness of the combined sample in each sector. The rest of the catalogue creation steps, i.e. random catalogue generation and $\bar{n}$ estimation, are identical to the algorithms used for the CMASS and LOWZ catalogues described previously.

When analysing the combined catalogue, as well as allowing for any evolution in the bias across the sample, one also has to consider the differential bias between LOWZ and CMASS samples. Although this is expected to be small due to the relatively benign transition from LOWZ to CMASS (Ross et al., in preparation), a full exploration of this issue is left for a forthcoming BOSS team paper.

8 DISCUSSION

The small statistical errors achievable on cosmological measurements from BOSS data require removal of potential systematic issues to an unprecedented level. Spectroscopic target selection and mask creation are key areas where systematic problems can be introduced if care is not taken to fully understand both. In this paper, we have presented the target selection for the three primary spectroscopic galaxy catalogues within BOSS: LOWZ, CMASS and Sparse, and for variations on these used for some early data. Each sample has different sky coverage and expected redshift distribution.

We have also presented the methods used to turn the target catalogue and redshift measurement data into galaxy and random catalogues, which enable clustering measurements to be quickly made, as well as methods to mitigate potential systematics. It may be that some analyses are best done without the corrections provided – for example, it may be cleaner for small-scale clustering analyses not to apply the close-pair weights, but to correct in some other manner.

In addition to a number of improvements over the catalogue creation method used for DR9, DR10 and DR11 samples we have described how we have created a single BOSS catalogue, combining CMASS and LOWZ samples. This allows us to include some extra galaxies, and maximize the effective volume covered by galaxies within BOSS. It also allows us to use a binning scheme in redshift different from those of CMASS and LOWZ, optimizing our cosmological measurements.

The resulting galaxy and random catalogues, the largest in the world, are hosted at http://www.sdss.org/dr12/ as well as supplemental catalogue and target information. In this final release, we also provide copies of our source code, MKSAMPLE, to reproduce the DR10, DR11, and DR12 catalogues. The reader should consult the source code directly to resolve any ambiguities in our description here.

Next generation spectroscopic experiments, such as eBOSS (Dawson et al. 2015), DESI (Levi et al. 2013), HETDEX (Hill et al. 2008), 4MOST (de Jong et al. 2014), WEAVE (Dalton et al. 2014) PFS (Takada et al. 2014), Euclid (Laureijs et al. 2011) and WFIRST (Spergel et al. 2015), are expected to make cosmological measurements with precision either comparable or higher by up to an order of magnitude compared to that of BOSS, requiring a thorough understanding and extremely careful treatment of potential systematic effects. Although each of these future experiments have different observing strategies, they will encounter challenges in the process of catalogue creation similar to those of BOSS (e.g. variations in the galaxy surface density due to galactic extinction is an effect inherent to our observable Universe). The lessons learned from the catalogue creation method applied within BOSS, and described in this paper, will be of strong benefit for these future surveys.

ACKNOWLEDGEMENTS

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 Astrofisica 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.

We thank Eli Rykoff for providing the Bright Objects Mask described in Section 5.1.1.

The author order reflects the following: BR led the development of the MKSAMPLE software package. She is followed in the author list by two alphabetical lists of scientists who provided major contributions to galaxy targeting and/or catalogue creation. Out of these, WJP is corresponding author as he took-over the creation of this paper when BR left the field of astronomy. This list is followed by an alphabetical list of scientists who provided moderate contributions to these topics, and who contributed to the BOSS project as a whole.

REFERENCES

Abazajian K. N. et al., 2004, AJ, 128, 502
Abazajian K. N. et al., 2009, ApJS, 182, 543
Ahn C. P. et al., 2012, ApJS, 203, 21
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., 2012, MNRAS, 427, 3435
Anderson L. et al., 2014a, MNRAS, 439, 83
Anderson L. et al., 2014b, MNRAS, 441, 24
Annis J. et al., 2014, ApJ, 794, 120
Beutler F. et al., 2014a, MNRAS, 443, 1065
Beutler F. et al., 2014b, MNRAS, 444, 3501
Blake C. et al., 2011, MNRAS, 418, 1707
Blanton M., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003, AJ, 125, 2276
Bolton A. et al., 2012, AJ, 144, 144
Bundy K. et al., 2015, preprint (arXiv:1509.01276)
Cannon R. et al., 2006, MNRAS, 372, 425
Casali M. et al., 2007, A&A, 467, 777
Chen Y. M. et al., 2012, MNRAS, 421, 314
Chuang C.-H., 2013, MNRAS, 433, 3559
Cole S. et al., 2005, MNRAS, 362, 505
Colless M. et al., 2001, MNRAS, 328, 1039
Colless M. et al., 2003, preprint (astro-ph/0306581)
Cuesta A. et al., 2015, preprint (arXiv:1509.06371)
Dalton G. et al., 2014, in Ramsay S. K., McLean I. S., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 9147, Ground-Based and Airborne Instrumentation for Astronomy V. SPIE, Bellingham, p. 91470L
Dawson K. et al., 2012, AJ, 145, 10
Dawson K. et al., 2015, preprint (arXiv:1508.04473)
de Jong R. S. et al., 2014, in Ramsay S. K., McLean I. S., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V. SPIE, Bellingham, p. 91470M
Doi M. et al., 2010, AJ, 139, 1628
Doroshkevich A. G., Zel'dovich Ya. B., Sunyaev R. A., 1978, Sov. Astron., 22, 523
Drinkwater M. et al., 2010, MNRAS, 401, 1429
Eisenstein D. J. et al., 2001, AJ, 122, 2267
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Eisenstein D. J. et al., 2011, AJ, 142, 72
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23

Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Gil-Marin H., Noreña J., Verde L., Percival W. J., Wagner C., Manera M., Schneider D. P., 2015a, MNRAS, 451, 539
Gil-Marin H. et al., 2015b, MNRAS, 452, 1914
Gil-Marin H. et al., 2015c, preprint (arXiv:1509.06373)
Gil-Marin H. et al., 2015d, preprint (arXiv:1509.06386)
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Gunn J. E. et al., 1998, AJ, 116, 3040
Gunn J. E. et al., 2006, AJ, 131, 2332
Hambly N. C. et al., 2008, MNRAS, 384, 637
Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T., 2006, MNRAS, 367, 454
Hill G. J. et al., 2008, in Kodama T., Yamada T., Aoki K., eds, ASP Conf. Ser. Vol. 346, Panoramic Views of Galaxy Formation and Evolution. Astron. Soc. Pac., San Francisco, p. 115
Hoshino H. et al., 2015, MNRAS, 452, 998
Høg E. et al., 2000, A&A, 355, 27
Kaiser N., 1986, MNRAS, 219, 785
Kaiser N., 1987, MNRAS, 227, 1
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Lawrence A. et al., 2007, MNRAS, 379, 1599
Leauthaud A. et al., 2015, preprint (arXiv:1507.04752)
Levi M. et al., 2013, preprint (arXiv:1308.0847)
Lewis I. J., Glazebrook K., Taylor K., 1998, Proc. SPIE, 3355, 828
Lupton R., Gunn J. E., Ivezić Z., Knapp G., Kent S., 2001, in Harnden F. R., Jr, Primini F. A., Payne H. E., eds, ASP Conf. Ser. Vol. 238, Astronomical Data Analysis Software and Systems X. Astron. Soc. Pac., San Francisco, p. 269
Maraston C., Strömbäck G., Thomas D., Wake D. A., Nichol R. C., 2009, MNRAS, 394, 107
Maraston C. et al., 2013, MNRAS, 435, 2764
Masters K. L. et al., 2011, MNRAS, 418, 1055
Montero-Dorta A. D. et al., 2015, preprint (arXiv:1410.5854)
Nuza S. E. et al., 2013, MNRAS, 432, 743
Padmanabhan N. et al., 2008, ApJ, 674, 1217
Parejko J. K. et al., 2013, MNRAS, 429, 98
Peebles P. J. E., Yu J. T., 1970, ApJ, 162, 815
Percival W. J. et al., 2001, MNRAS, 327, 1297
Percival W., Verde L., Peacock J., 2004, MNRAS, 347, 645
Pier J. R., Munn J. A., Hindsley R. B., Hennessy G. S., Kent S. M., Lupton R. H., Ivezi' Ž., 2003, AJ, 125, 1559
Reid B. et al., 2012, MNRAS, 426, 2719
Reid B., Seo H.-J., Leauthaud A., Tinker J., White M., 2014, MNRAS, 444, 476
Ross A. et al., 2012, MNRAS, 428, 1116
Ross A. et al., 2013, MNRAS, 424, 564
Ross A. et al., 2014, MNRAS, 437, 1109
Rykoff E. S. et al., 2014, ApJ, 785, 104
Saito S. et al., 2015, preprint (arXiv:1509.00482)
Samushia L. et al., 2013, MNRAS, 429, 1514
Samushia L. et al., 2014, MNRAS, 439, 3504
Sanchez et al., 2012, MNRAS, 425, 415
Sanchez et al., 2014, MNRAS, 440, 2692
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Scoccola C. G. et al., 2013, MNRAS, 434, 1792
Seo H.-J., Eisenstein D. J., 2007, ApJ, 665, 14
Smee S. et al., 2013, AJ, 146, 32
Smith J. A. et al., 2002, AJ, 123, 2121
Spergel D. et al., 2015, preprint (arXiv:1503.03757)
Stoughton C. et al., 2002, AJ, 123, 3487
Strauss M. A. et al., 2002, AJ, 124, 1810
Sunyaev R. A., Zel'dovich Ya. B., 1970, ApSS, 7, 3
Swanson M. E. C., Tegmark M., Hamilton A. J. S., Hill J. C., 2008, MNRAS, 387, 1391
Takada M. et al., 2014, PASP, 66, 1
Tegmark M. et al., 2004, ApJ, 606, 702
Thomas D. et al., 2013, MNRAS, 431, 1383
Tojeiro R. et al., 2012, MNRAS, 424, 136
Tojeiro R. et al., 2014, MNRAS, 440, 2222
Vargas-Magana M. et al., 2014, MNRAS, 445, 2
Vargas-Magaña M., Ho S., Fromenteau S., Cuesta A. J., 2015, preprint (arXiv:1509.06384)
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87
White M. et al., 2011, ApJ, 728, 126
York D. G. et al., 2000, AJ, 120, 1579
Zhao G.-B. et al., 2013, MNRAS, 436, 2038

APPENDIX A: LOWZ EARLY SELECTION ALGORITHMS

As the survey progressed, these were slight changes to the targeting pipeline. In some instances, the newer algorithm was stricter than the one used in the past, so we simply apply the same cuts to the objects targeted earlier as well. One special case is the LOWZ targets in chunks 2–6. The star–galaxy separation algorithm for CMASS was erroneously applied to those galaxies as well, resulting in a drastic reduction in the target density. There are other differences, and so we define two algorithms, LOWZE2 as that applied to chunk 2, and LOWZE3 as that applied to chunks 3–6. In analyses thus far we have simply eliminated these early regions from our LOWZ catalogue, but we are actively pursuing a sufficient description of that population to robustly recover clustering measurements in those regions. Removing this area results in a 10 per cent reduction in the LOWZ survey mask area. The distributions of the early chunks on the sky are shown in Fig. A1. Chunk 2 was commissioning data, and used the LOWZE2 version; Chunks 3–6 used LOWZE3, and chunks 7–11 used older photometric reductions, and a different version of RESOLVE (see Section 2.1). Chunk 1 was used for very early commissioning runs and is not of sufficient uniformity to be used to create LSS catalogues. Chunk 1 is located at Dec.=0° in the SGC footprint (commonly referred to as 'Stripe 82'). This area was later reobserved with updated target selection as Chunk 11.

(i) Chunk 2: the LOWZE2 sample had slightly different $r_{\rm cmod}$ cuts and the CMASS i-band star–galaxy separation cut was erroneously applied. The catalogue was later trimmed to $16 < r_{\rm cmod}$ as well.

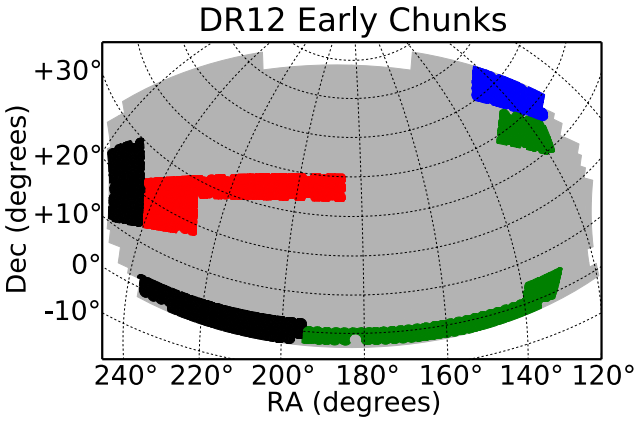

Figure A1. The location of early chunks, where targeting and/or photometric reduction versions differed from the later chunks. Chunk 2 (blue), chunks 3–4 (green), chunks 5–6 (red), and chunks 7–11 (black) are shown. Chunks 7–11 used an early version of the imaging data reduction software (see Section 2).

This selection yields a target density ∼15 per cent lower than the nominal LOWZ target sample,

$$r_{\rm cmod} < 13.4 + c_{||}/0.3 \tag{A1}$$

$$|c_{\perp}| < 0.2 \tag{A2}$$

$$r_{\rm cmod} < 19.5 \tag{A3}$$

$$r_{\rm psf} - r_{\rm cmod} > 0.3 \tag{A4}$$

$$i_{\rm psf} - i_{\rm mod} > 0.2 + 0.2(20 - i_{\rm mod}). \tag{A5}$$

(ii) Chunks 3–6: the LOWZE3 sample is the same as chunk 2 but with a stricter $17 < r_{\rm cmod}$ bound and both star–galaxy separation cuts. This selection yields a target density ∼45 per cent lower than the nominal LOWZ target sample,

$$r_{\rm cmod} < 13.4 + c_{||}/0.3 \tag{A6}$$

$$|c_{\perp}| < 0.2 \tag{A7}$$

$$17 < r_{\rm cmod} < 19.5 \tag{A8}$$

$$r_{\rm psf} - r_{\rm cmod} > 0.3 \tag{A9}$$

$$i_{\rm psf} - i_{\rm mod} > 0.2 + 0.2(20 - i_{\rm mod}) \tag{A10}$$

$$z_{\rm psf} - z_{\rm mod} > 9.125 - 0.46 z_{\rm mod}. \tag{A11}$$

(iii) The $i_{\rm fib2} < 21.5$ CMASS cut was applied to chunks 15 and above. Our CMASS LSS catalogue applies this cut to all chunks.

[1]*Department of Physics, Berkeley Center for Cosmological Physics, University of California, Berkeley, CA, 94720, USA*
[2]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[3]*Department of Physics, Bruce and Astrid McWilliams Center for Cosmology, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA*
[4]*Yale Center for Astronomy and Astrophysics, Yale University, New Haven, CT 06520, USA*
[5]*Institute of Cosmology & Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth PO1 3FX, UK*
[6]*Department of Physics, Center for Cosmology and Particle Physics, New York University, 4 Washington Place, New York, NY 10003, USA*
[7]*School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK*
[8]*Departments of Physics and Astronomy, University of California, Berkeley, CA 94720, USA*
[9]*Harvard–Smithsonian Center for Astrophysics, 60 Garden St, Cambridge, MA 02138, USA*
[10]*Center for Cosmology and AstroParticle Physics, The Ohio State University, Columbus, OH 43210, USA*
[11]*Max-Planck-Institut für extraterrestrische Physik, ch 1312, Giessenbachstr., D-85741 Garching, Germany*
[12]*Brookhaven National Laboratory, Bldg 510, Upton, New York, NY 11973, USA*
[13]*Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA*
[14]*Department of Astronomy, University of Winsconsin-Madison, 475 N. Charter Street, Madison, WI 53706-1582, USA*
[15]*Department of Physical Sciences, The Open University, Milton Keynes MK7 6AA, UK*
[16]*Apache Point Observatory and New Mexico State University, PO Box 59, Sunspot, NM 88349-0059, USA*
[17]*Sternberg Astronomical Institute, Moscow State University, 119992 Moscow, Russia*
[18]*Department Physics and Astronomy, University of Utah, 115 S 1400 E, Salt Lake City, UT 84112, USA*
[19]*Instituto de Física Teórica, (UAM/CSIC), Universidad Autonoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*
[20]*Department of Astronomy, Case Western Reserve University, 10900 Euclid Ave, Cleveland, OH 44106, USA*
[21]*Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany*
[22]*Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[23]*Department of Chemistry and Physics, King's College, Wilkes Barre, PA 18711, USA*
[24]*Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany*
[25]*Aix Marseille Universit, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, F-13388 Marseille, France*
[26]*Campus of International Excellence UAM+CSIC, Cantoblanco, E-28049 Madrid, Spain*
[27]*Instituto de Astrofisica de Andaluca (CSIC), Glorieta de la Astronoma, E-18080 Granada, Spain*
[28]*Departamento de Física Teórica, Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*
[29]*Universitäts-Sternwarte München, Scheinerstrasse 1, D-81679 Munich, Germany*
[30]*Kansas State University, Manhattan, KS 66506, USA*
[31]*National Abastumani Astrophysical Observatory, Ilia State University, 2A Kazbegi Ave, GE-1060 Tbilisi, Georgia*
[32]*Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA*
[33]*Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA*
[34]*Instituto de Astrofísica de Canarias (IAC), C/Vía Láctea, s/n, E-38200 La Laguna, Spain*
[35]*Facultad de Ciencias Astronómicas y Geofísicas -Universidad Nacional de La Plata, Paseo del Bosque, S/N (1900) La Plata, Argentina*
[36]*CONICET, Rivadavia 1917, 1033 Buenos Aires, Argentina*
[37]*Instituto de Fisica, Universidad Nacional Autónoma de México, Apdo. Postal 20-364, Mexico City, México*

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