

# Producing a BOSS CMASS sample with DES imaging

S. Lee<sup>1,2</sup>★ E. M. Huff,<sup>3</sup> A. J. Ross<sup>1</sup> A. Choi,<sup>1</sup> C. Hirata,<sup>1,2</sup> K. Honscheid,<sup>1,2</sup> N. MacCrann,<sup>1,2</sup> M. A. Troxel,<sup>4</sup> C. Davis<sup>1,5</sup> T. F. Eifler,<sup>3,6</sup> R. Cawthon,<sup>7</sup> J. Elvin-Poole,<sup>1,2</sup> J. Annis,<sup>8</sup> S. Avila<sup>1,9</sup> E. Bertin,<sup>10,11</sup> D. Brooks,<sup>12</sup> A. Carnero Rosell<sup>1,14</sup> M. Carrasco Kind,<sup>15,16</sup> J. Carretero,<sup>17</sup> L. N. da Costa,<sup>14,18</sup> J. De Vicente,<sup>13</sup> S. Desai,<sup>19</sup> B. Flaugher,<sup>8</sup> P. Fosalba,<sup>20,21</sup> J. García-Bellido,<sup>9</sup> E. Gaztanaga,<sup>20,21</sup> D. W. Gerdes,<sup>22,23</sup> D. Gruen<sup>1,24,25</sup> R. A. Gruendl,<sup>15,16</sup> J. Gschwend,<sup>14,18</sup> G. Gutierrez,<sup>8</sup> D. L. Hollowood,<sup>26</sup> D. J. James,<sup>27</sup> T. Jeltema,<sup>26</sup> K. Kuehn,<sup>28</sup> M. Lima,<sup>14,29</sup> M. A. G. Maia,<sup>14,18</sup> J. L. Marshall,<sup>30</sup> P. Martini,<sup>1,31</sup> P. Melchior<sup>1,32</sup> F. Menanteau,<sup>15,16</sup> C. J. Miller,<sup>22,23</sup> R. Miquel,<sup>17,33</sup> R. L. C. Ogando,<sup>14,18</sup> A. Palmese<sup>1,8</sup> A. A. Plazas<sup>1,32</sup> E. Sanchez,<sup>13</sup> V. Scarpine,<sup>8</sup> M. Schubnell,<sup>23</sup> S. Serrano,<sup>20,21</sup> I. Sevilla-Noarbe,<sup>13</sup> M. Smith,<sup>34</sup> M. Soares-Santos<sup>1,35</sup> F. Sobreira,<sup>14,36</sup> E. Suchyta<sup>1,37</sup> M. E. C. Swanson,<sup>16</sup> G. Tarle,<sup>23</sup> D. Thomas<sup>1,38</sup> J. Weller,<sup>39,40,41</sup> and J. Zuntz<sup>1,42</sup> (DES Collaboration)

*Affiliations are listed at the end of the paper*

Accepted 2019 June 30. Received 2019 June 3; in original form 2019 June 3

## ABSTRACT

We present a sample of galaxies with the Dark Energy Survey (DES) photometry that replicates the properties of the BOSS CMASS sample. The CMASS galaxy sample has been well characterized by the Sloan Digital Sky Survey (SDSS) collaboration and was used to obtain the most powerful redshift-space galaxy clustering measurements to date. A joint analysis of redshift-space distortions (such as those probed by CMASS from SDSS) and a galaxy–galaxy lensing measurement for an equivalent sample from DES can provide powerful cosmological constraints. Unfortunately, the DES and SDSS-BOSS footprints have only minimal overlap, primarily on the celestial equator near the SDSS Stripe 82 region. Using this overlap, we build a robust Bayesian model to select CMASS-like galaxies in the remainder of the DES footprint. The newly defined DES-CMASS (DMASS) sample consists of 117 293 effective galaxies covering 1244 deg<sup>2</sup>. Through various validation tests, we show that the DMAS sample selected by this model matches well with the BOSS CMASS sample, specifically in the South Galactic cap (SGC) region that includes Stripe 82. Combining measurements of the angular correlation function and the clustering- $z$  distribution of DMAS, we constrain the difference in mean galaxy bias and mean redshift between the BOSS CMASS and DMAS samples to be  $\Delta b = 0.010^{+0.045}_{-0.052}$  and  $\Delta z = (3.46^{+5.48}_{-5.55}) \times 10^{-3}$  for the SGC portion of CMASS, and  $\Delta b = 0.044^{+0.044}_{-0.043}$  and  $\Delta z = (3.51^{+4.93}_{-5.91}) \times 10^{-3}$  for the full CMASS sample. These values indicate that the mean bias of galaxies and mean redshift in the DMAS sample are consistent with both CMASS samples within  $1\sigma$ .

**Key words:** methods: data analysis – techniques: photometric – galaxies: general.

## 1 INTRODUCTION

Since the discovery of the accelerating expansion of the Universe two decades ago (Riess, Filippenko & Challis 1998; Perlmutter et al. 1999), observational and theoretical work has led to a

\* E-mail: lee.5922@osu.edu

concordance cosmological model dominated by 70 per cent dark energy, 25 per cent dark matter, and 5 per cent baryons. Despite the fact that dark energy occupies the majority of the energy density in the Universe, little is understood about its physical nature due to the apparent lack of visible properties. Compelling evidence for the presence of dark energy comes from observations of the underlying matter distribution in the Universe using supernovae, baryon acoustic oscillations (BAOs), and measurements of large-scale structure growth (Frieman, Turner & Huterer 2008; Weinberg et al. 2013; Huterer & Shafer 2018).

To trace out the underlying structure in matter, cosmologists traditionally use galaxies by measuring galaxy clustering as a function of spatial separation. However, using galaxies as tracers results in a biased view of the matter distribution because galaxies form at the peaks of the matter density field where gas reaches high enough density to cool and form stars (Kaiser 1984). The relation between the spatial distributions of galaxies and the underlying dark matter density field is known as galaxy bias. Galaxy bias varies for different scales and galaxy properties such as luminosity or type, and those quantities are degenerate with each other. In the absence of additional information, galaxy bias is indistinguishable from the overall amplitude of matter fluctuations, which makes galaxy bias a major systematic uncertainty in cosmological analyses (Seljak et al. 2005).

Fortunately, weak gravitational lensing provides a direct way to measure the matter distribution, avoiding the issue of galaxy bias. Cosmic shear is the subtle shape distortions of background (source) galaxies by the foreground (lens) matter distribution. It is thus directly connected to the matter distribution and thereby lets us measure the matter distribution without any galaxy bias (see the review in Weinberg et al. (2013) and references therein). However, cosmic shear is technically challenging to measure due to many sources of systematic errors. Because of the small size of the effect compared to the intrinsic random variation in galaxy orientations and ellipticities, weak lensing measurements require a substantial number of source galaxies to achieve small statistical errors. This results in including small and faint galaxies whose systematic errors are challenging to control (Mandelbaum 2018).

Galaxy–galaxy lensing has been shown to be a powerful tool (Baldauf et al. 2010; Choi et al. 2012; Yoo & Seljak 2012; Mandelbaum et al. 2013; van den Bosch et al. 2013; Miyatake et al. 2015; More et al. 2015; Park et al. 2016; Alam et al. 2017; Amon et al. 2018; Singh et al. 2018, 2019; Jullo et al. 2019) that is insensitive to some of the systematic errors that affect cosmic shear (Hirata et al. 2004). It is the cross-correlation function between foreground galaxies and background shear, which represents a direct measurement of the galaxy–matter correlation function. In combination with accurate galaxy clustering information, lensing observables can fully exert their constraining power. In galaxy–galaxy lensing, the galaxy bias is tied to the matter clustering in a different way from galaxy clustering. Combining the two probes breaks the degeneracy between the two constraints.

Some of the sets of galaxies most frequently used as gravitational lenses in cosmological analyses are the BOSS spectroscopic galaxy samples (Reid et al. 2016) from the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013), which is part of the Sloan Digital Sky Survey-III (SDSS-III; Eisenstein et al. 2011). The large sample size and availability of spectroscopic redshifts for all BOSS galaxies allowed the BOSS collaboration to measure the BAO signature with an uncertainty of only 1 per cent for the case of the BOSS CMASS sample, which is the most constraining BAO measurement to date (Reid et al. 2016). This led to several follow-

up studies that combined the BOSS galaxy clustering results with galaxy–galaxy lensing measurements using the BOSS galaxies as lenses.

Mandelbaum et al. (2013) constrained the amplitude of the matter fluctuations at  $z < 0.4$  using data from the SDSS Data Release 7. They utilized two spectroscopic samples – BOSS Main and Luminous Red samples – as lenses and combined galaxy–galaxy lensing between those samples and SDSS source galaxies with galaxy clustering from the same samples. Singh et al. (2018) adopted a similar approach. They combined galaxy clustering from BOSS with galaxy–galaxy lensing and galaxy–CMB (cosmic microwave background) lensing signals, by utilizing the BOSS LOWZ ( $0.15 < z < 0.43$ ) and CMASS ( $0.43 < z < 0.7$ ) samples as lenses. However, due to the shallow depth of SDSS imaging, their measurement of galaxy–galaxy lensing was obtained only with BOSS LOWZ.

Miyatake et al. (2015), More et al. (2015), and Alam et al. (2017) extended this kind of joint analysis to galaxies at a higher redshift  $z \sim 0.5$  by using BOSS CMASS as lenses with the deeper and better quality imaging data from CFHTLenS (Heymans et al. 2012). Jullo et al. (2019) performed a similar analysis with BOSS CMASS galaxies and two weak lensing data sets – CFHTLenS and CFHT-Stripe 82 (Moraes et al. 2014). Amon et al. (2018) utilized three spectroscopic galaxy samples including BOSS LOWZ and CMASS with deep imaging data from KiDS (de Jong et al. 2013) to do a joint analysis of galaxy clustering and galaxy–galaxy lensing. However, the lensing measurements of these analyses are limited to the small overlapping area – a few hundreds of  $\text{deg}^2$ . For instance, the overlapping region between BOSS and CFHTLenS is only  $\sim 105 \text{ deg}^2$ , which is about one hundredth of the BOSS area.

The Dark Energy Survey (DES) is a large photometric survey that images over  $5000 \text{ deg}^2$  of the southern sky to a  $5\sigma$  limiting magnitude of  $\sim 24$  in the  $i$ -band. It observes in the  $grizY$  filter bands. Precise photometry and the largest survey area among the current generation of experiments make DES data an excellent source of imaging data for a joint analysis of galaxy clustering and galaxy–galaxy lensing. However, as with previous measurements combining lensing and clustering, the overlapping region between the DES Year 1 footprint ( $\sim 1800 \text{ deg}^2$ ) and the BOSS footprint is fairly small, consisting of only  $\sim 150 \text{ deg}^2$  near the celestial equator called Stripe 82 (Dark Energy Survey Collaboration 2016; Drlica-Wagner et al. 2018). Simply combining BOSS galaxy clustering with galaxy–galaxy lensing from DES would be limited to the small overlapping area and fail to utilize the full statistical power of DES.

Inspired by the potential power of combining all the available SFDSS and DES measurements, we present in this paper a way of defining a catalogue of DES galaxies from the full footprint of DES, whose properties match with the BOSS CMASS galaxy sample. The resulting DES-CMASS (hereafter Dmass) sample will be the best available for a cosmological analysis combining galaxy–galaxy lensing and galaxy clustering measurements. Thus, we will produce a sample that effectively increases the area available for such studies by a factor of 10 (from 123 to  $1244 \text{ deg}^2$ ).

We start by using the subset of BOSS CMASS galaxies in Stripe 82 where the BOSS footprint overlaps with DES. Using galaxies measured by both DES and BOSS, we train a galaxy selection model using the DES photometric information. Rather than classifying individual galaxies, the model assigns a membership probability to each galaxy and down-weights galaxies that are less likely to be CMASS. To account for spatial dependence of photometric errors, we use the extreme deconvolution (XD) algorithm (Bovy, Hogg & Roweis 2011a) and obtain underlying colour distributions of galaxies from the training sample. The underlying colour distributions

are convolved with photometric errors of the target region, and thereby the model correctly accounts for the photometric errors in the different regions.

This paper is organized as follows. In the following section, we will introduce the BOSS CMASS sample and the DES Y1 Gold catalogue we use for this work and present the selection criteria that were used for the BOSS CMASS sample in detail. We will address the difference between the SDSS and DES photometric systems and explain how it will be accounted for in our probabilistic model. Our model construction can be found in Section 3. The systematic uncertainties of the DMASS sample will be presented in Section 4 and the basic properties of the resulting DMASS catalogue and validation tests will be discussed in Section 5. We will summarize and conclude in Section 6.

The fiducial cosmological model used throughout this paper is a flat  $\Lambda$ CDM model with the following parameters: matter density  $\Omega_m = 0.307$ , baryon density  $\Omega_b = 0.048$ , amplitude of matter clustering  $\sigma_8 = 0.8288$ , spectral index  $n_s = 0.96$ , and Hubble constant  $h \equiv H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1} = 0.677$ .

## 2 DATA

### 2.1 BOSS DR12 CMASS sample

The BOSS (Eisenstein et al. 2011; Bolton et al. 2012; Dawson et al. 2013) was designed to measure the scale of BAOs in the clustering of matter over a larger volume than the combined efforts of all previous spectroscopic surveys of large-scale structure. BOSS uses the same wide-field, dedicated telescope as was employed by SDSS-I and SDSS-II (York et al. 2000), the 2.5 m aperture Sloan Foundation Telescope (Gunn et al. 2006) at Apache Point Observatory in New Mexico. Those surveys imaged over  $10\,000 \text{ deg}^2$  of high galactic latitude sky in the *ugriz* bands, using a mosaic CCD camera (Gunn et al. 1998) with a field of view spanning  $3^\circ$ . BOSS consists primarily of two interleaved spectroscopic surveys observed simultaneously: a redshift survey of 1.5 million luminous galaxies extending to  $z = 0.7$  and a survey of the Lyman alpha forest towards 150 000 quasars in the redshift range  $2.15 < z < 3.5$ . Description of survey design, target selection, and their implications for cosmological analysis are available in Dawson et al. (2013) and Reid et al. (2016).

The BOSS DR12 galaxy survey targeted two distinct samples known as LOWZ and CMASS (Reid et al. 2016). The higher redshift sample CMASS covers redshifts  $0.43 < z < 0.75$  and is designed to select a stellar mass-limited sample of objects of all intrinsic colours, with a colour cut that selects almost exclusively on redshift. The CMASS galaxy sample is selected by the combination of the seven different colour and magnitude cuts. Every source satisfying the selection cuts was targeted by the BOSS spectrograph to obtain their redshifts, except for 5.8 per cent of targets in a fibre collision group and 1.8 per cent of targets for which the spectroscopic pipeline fails to obtain a robust redshift (Reid et al. 2016).

The following three cuts simply limit colours or magnitudes to exclude redshift failures or outliers with problematic photometry:

$$17.5 < i_{\text{cmod}} < 19.9, \quad (1)$$

$$i_{\text{fib2}} < 21.5, \quad (2)$$

$$r_{\text{mod}} - i_{\text{mod}} < 2, \quad (3)$$

where the subscript ‘mod’ denotes model magnitudes, ‘cmod’ denotes cmodel magnitudes, and ‘fib2’ stands for fibre magnitude estimated in a 2 arcsec aperture diameter assuming 2 arcsec seeing. For further details of SDSS magnitudes, we refer readers to the SDSS survey website.<sup>1</sup> The following two cuts are applied to reject stars:

$$i_{\text{psf}} - i_{\text{mod}} > 0.2 + 0.2(20 - i_{\text{mod}}), \quad (4)$$

$$z_{\text{psf}} - z_{\text{mod}} > 9.125 - 0.46z_{\text{mod}}, \quad (5)$$

where ‘psf’ stands for magnitudes computed from the point spread function model.

To exclusively select galaxies on redshift, the BOSS target selection utilizes the quantity  $d_{\perp}$  defined as

$$d_{\perp} \equiv (r_{\text{mod}} - i_{\text{mod}}) - (g_{\text{mod}} - r_{\text{mod}})/8.0. \quad (6)$$

This quantity is designed to approximately follow the colour locus of the passively evolving LRG model in Maraston et al. (2009) at  $z > 0.4$ . Since redshift gradually increases along the colour locus,  $d_{\perp}$  is a good indication of redshift for CMASS-type galaxies.

The following two cuts use  $d_{\perp}$  to select objects with respect to redshift:

$$i_{\text{cmod}} < 19.86 + 1.6(d_{\perp} - 0.8), \quad (7)$$

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

Equation (7) selects the brightest objects at each redshift to keep an approximately constant stellar mass limit over the redshift range of CMASS. Equation (8), the so-called  $d_{\perp}$  cut, is the most restrictive cut among all selections described earlier. This cut isolates intrinsically red galaxies at high redshift. Considering the colour/magnitude space occupied by all SDSS objects, this cut slices the densest region of the sample in the *gri* colour plane and determines the sample’s redshift distribution. This is in contrast to the other cuts, which apply mainly to the edges of the colour/magnitude distributions. Therefore, our work is mainly focused on characterizing the same cut in the DES photometry. More details about the  $d_{\perp}$  cut can be found in Eisenstein et al. (2001) and Padmanabhan et al. (2007), and our derived  $d_{\perp}$  cut in the DES system will be discussed in Section 3.

The colours and magnitudes used in the selection criteria are corrected for Milky Way extinction by the galactic extinction map (Schlegel, Finkbeiner & Davis 1998).

### 2.2 DES Y1 Gold catalogue

The DES (Dark Energy Survey Collaboration 2005; Abbott et al. 2018a) is an imaging survey covering  $5000 \text{ deg}^2$  of the southern sky. This photometric data set has been obtained in five broad-band filters, *grizY*, ranging from  $\sim 400$  to  $\sim 1060 \text{ nm}$  (Li et al. 2016; Burke et al. 2018), using the Dark Energy Camera (Flaugher et al. 2015) mounted on the Blanco 4 m telescope at Cerro Tololo Inter-American Observatory in Chile. The main goal of DES is to improve our understanding of cosmic acceleration and the nature of dark energy using four key probes: weak lensing, large-scale structure, galaxy clusters, and Type Ia supernovae.

<sup>1</sup><https://www.sdss.org/dr12/algorithms/magnitudes>

The Y1A1 Gold wide-area object catalogue<sup>2</sup> (Drlica-Wagner et al. 2018) we use in this work consists of  $\sim 137$  million objects detected in coadd images covering two disjoint areas: one overlapping with the South Pole Telescope (SPT; Carlstrom et al. 2011), and a much smaller area near the celestial equator called Stripe 82 (Annis et al. 2014).

For this work, we refine the DES Y1 Gold catalogue selection by removing imaging artefacts and areas around bright foreground objects such as bright stars and globular clusters. We only keep clean sources with flag  $bit < 1$  in table 4 and flag  $bit < 2$  in table 5 in Drlica-Wagner et al. (2018). We also select sources classified as galaxies by the flag MODEST==1. Furthermore, we remove regions tagged by the DES Y1 BAO study (Crocce et al. 2019) using veto masks. These additional masks select only the wide-area parts of the surveys, namely, those overlapping SPT, and remove a patch of  $18\text{ deg}^2$  where the airmass computation is highly corrupted. The DES Y1 BAO study additionally removes a few  $\text{deg}^2$  sized regions where multi-object fitting (MOF; Drlica-Wagner et al. 2018) photometry is unreliable. However, we do not exclude these regions since we do not use MOF measurements. Further details about the Y1 BAO masks can be found in Crocce et al. (2019). The resulting footprint after applying all aforementioned masks occupies  $1244\text{ deg}^2$  in SPT and  $123\text{ deg}^2$  in Stripe 82.

All magnitudes in the DES Y1 Gold catalogue are shifted by stellar locus regression (SLR), which corrects for Galactic dust reddening (Drlica-Wagner et al. 2018). For consistency with the original CMASS selection, we have removed this SLR correction and instead applied reddening corrections based on the SFD map (Schlegel et al. 1998) as done in SDSS. The correction to the DES magnitude for a band  $b$  is  $A_b = R_b \times E(B - V)_{\text{SFD}}$  with interstellar extinction coefficients for *griz* bands,  $R_b = [3.186, 2.140, 1.569, 1.196]$ , computed in Abbott et al. (2018b).

We applied additional magnitude cuts to the DES Y1 Gold catalogue to exclude outliers in colour space as follows:

$$17 < G_{\text{DET}} < 24, \quad (9)$$

$$17 < R_{\text{DET}} < 24, \quad (10)$$

$$17 < I_{\text{DET}} < 24, \quad (11)$$

$$0 < G_{\text{MOD}} - R_{\text{MOD}} < 2.5, \quad (12)$$

$$0 < R_{\text{MOD}} - I_{\text{MOD}} < 1.5, \quad (13)$$

$$I_{\text{AUTO}} < 21. \quad (14)$$

Sources satisfying the magnitude cuts are kept. Subscripts DET and MOD stand for DES MAG\_DETMODEL<sup>3</sup> magnitude and MAG\_MODEL<sup>4</sup> magnitude, respectively, and AUTO stands for DES

<sup>2</sup><https://des.ncsa.illinois.edu/releases/y1a1>

<sup>3</sup>This magnitude is computed by fitting a galaxy model to the object in a reference detection image taken in one band or a combination of two bands. Then, this fitted model is applied to all measurement images, by fitting only the amplitude.

<sup>4</sup>This magnitude is measured by fitting a galaxy model to the object in each band.

MAG\_AUTO<sup>5</sup> magnitudes. These three magnitudes are computed by an image-processing software called SExtractor.<sup>6</sup> We refer interested readers to the documentation of SExtractor (Bertin & Arnouts 1996) for further details. Note that all DES quantities are written in upper case to avoid confusion with corresponding SDSS quantities. These cuts effectively remove galaxies that are not likely to be CMASS galaxies. Further, these cuts reduce compute time by decreasing the sample size to 10 per cent of the full Y1 Gold sample, while keeping 99.5 per cent of CMASS galaxies in the overlapping region, Stripe 82.

### 2.3 Differences between the SDSS and DES photometry

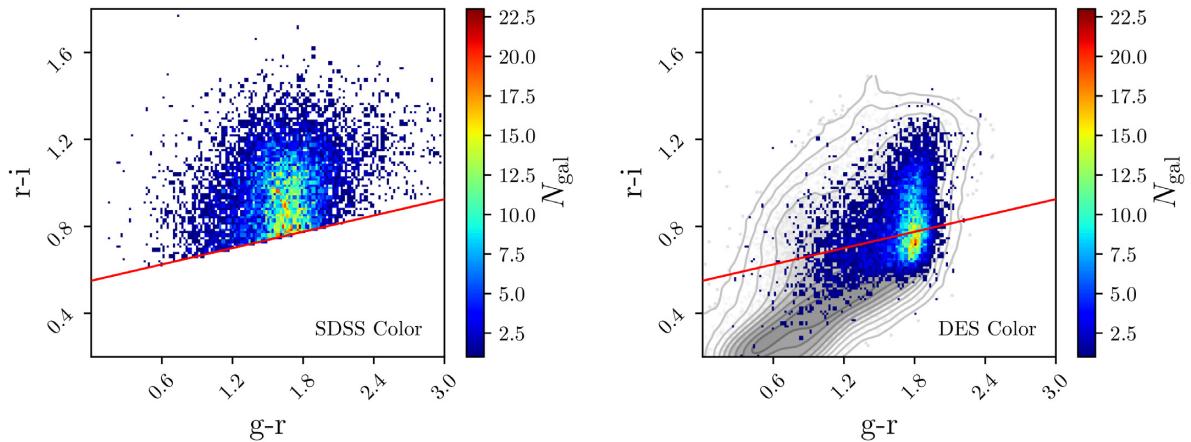
In the DES imaging pipeline, magnitudes for extended sources are derived from different models of luminosity profiles and bands optimized for each source (Drlica-Wagner et al. 2018). This complicated procedure makes magnitudes in one band highly correlated with other bands, as well as the shape or size of galaxies and instruments for each system, and results in magnitudes for the same object being very different in one system from another in a way that is challenging to predict.

Fig. 1 shows the difference in the  $r-i$  versus  $g-r$  colour space of the two different imaging systems, using only tagged CMASS galaxies in the overlap region. The DES colours of CMASS galaxies are obtained by cross-matching the DES Y1 Gold catalogue with the CMASS photometric sample in Stripe 82 by position with a  $2''$  tolerance. For the DES data, MAG\_DETMODEL magnitudes are used. The grey contours in the right-hand panel show all sources from the DES Y1 Gold catalogue. The red solid line in both panels is the  $d_{\perp}$  cut given by equation (8). By noting the large fraction of DES objects below this line, one can clearly see how different the two systems are. In the DES data, the  $d_{\perp}$  cut crosses the most dense part of galaxy sample. Notably, this is a dense region for the full gold sample as well. If we were to blindly apply the  $d_{\perp}$  to the DES data, we would remove almost half of the true CMASS sample. Applying a simple transformation that moves the  $d_{\perp}$  cut to lower  $r-i$  values recovers most of the CMASS galaxies, but at the cost of introducing many non-CMASS galaxies into the sample. Also noticeable in Fig. 1 is the larger scatter in the SDSS distribution, especially in  $g-r$ .

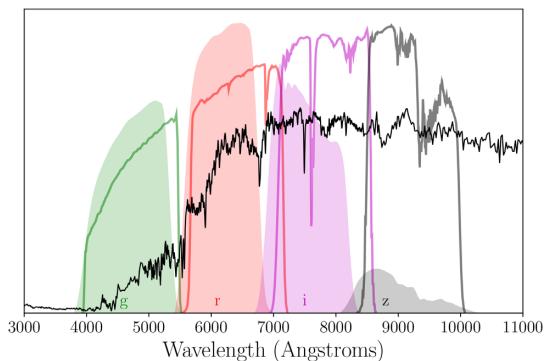
There are several reasons for the discrepancy in the colour space shown in Fig. 1. One is that despite both surveys using *griz* filters, these filters are not identical. Fig. 2 illustrates this fact. The response functions for the five SDSS (shaded) and DES (solid line) filters with the spectral energy density distribution of an elliptical galaxy are shown. The break in the model spectrum at  $4000\text{ \AA}$ , a primary feature of galaxy spectra, migrates through the  $g$ ,  $r$ , and  $i$  filters as the redshift increases (Eisenstein et al. 2001; Padmanabhan et al. 2007). For elliptical galaxies near  $z \sim 0.4$ , the  $4000\text{ \AA}$  break is located at wavelengths where the  $g-r$  transition happens. Near the  $4000\text{ \AA}$  break, the SDSS  $r$  filter (shaded) covers slightly lower wavelengths than the DES  $r$  filter (solid lines) does. That implies galaxies near  $z \sim 0.4$  look redder in SDSS than they do in the DES photometric system. Since the redshift  $z = 0.4$  is also where the  $d_{\perp}$  cut is defined, this discrepancy of the filter transition exacerbates the colour mismatch.

<sup>5</sup>Magnitude is measured in an elliptical aperture, shaped by the second moments of the object and scaled using the Kron radius.

<sup>6</sup><https://sextractor.readthedocs.io>



**Figure 1.** The two-dimensional histograms of CMASS galaxies from Stripe 82 in the  $g-r$  versus  $r-i$  colour plane. The left-hand panel shows SDSS, and the right-hand panel shows DES colours of the same galaxies. The colour bar shows the number of galaxies binned in each histogram bin. The red line is the  $d_{\perp}$  cut. CMASS galaxies look bluer in the DES photometry and the  $d_{\perp}$  cut discards almost half of the CMASS galaxies by crossing the most dense region. The grey contours in the right-hand panel show the full distribution of DES Y1 Gold galaxies in the colour plane. The grey contours show that blindly lowering the  $d_{\perp}$  cut results in accepting more non-CMASS galaxies.



**Figure 2.** The response functions for the  $griz$  SDSS (shaded) and DES (solid lines) filters as a function of wavelength ( $\text{\AA}$ ) with the spectral energy density distribution of an elliptical galaxy at  $z = 0.4$  (black solid line). Near the  $4000 \text{\AA}$  break where the  $g-r$  transition happens, the SDSS  $r$  filter (shaded) covers slightly lower wavelength than the DES  $r$  filter (solid lines) does. This implies the same galaxy near  $z = 0.4$  looks redder in the SDSS photometry than in the DES photometry.

A second cause for the discrepancy in colour space arises from differences in the SDSS and DES imaging pipelines. Magnitudes for extended sources are derived from the flux of a galaxy fitted with a best matched galaxy profile. Widely used galaxy profiles are exponential and de Vaucouleurs profiles (de Vaucouleurs 1948), which perform better for disc and bulge galaxies, respectively. The SDSS imaging pipeline uses either one of these profiles to model magnitudes depending on the shape of a galaxy and uses a linear combination of two profiles for SDSS cmodel magnitudes. The DES imaging pipeline uses only the exponential profile consistently for all magnitudes. The fitting procedure is different as well. For instance, the SDSS pipeline fits galaxies only in the  $r$ -band to obtain model magnitudes.<sup>7</sup> On the other hand, the DES pipeline fits galaxies in a reference image that can be taken from one band

or a combination of more than two bands to obtain the closest analogue, `MAG_DETMODEL` (Abbott et al. 2018b).

The last and most significant reason for the mismatch in colour distributions is the fact that SDSS has significantly larger photometric errors compared to DES. The typical photometric error of the CMASS galaxies is  $\sim 0.2$  along the  $g-r$  axis and  $\sim 0.07$  for the  $r-i$  axis which is  $\sim 5$  times larger than the typical error of DES.<sup>8</sup> The CMASS selection cuts in equations (1)–(8) are simple cuts that do not take into account photometric errors. Ignoring photometric errors does not cause a notable problem for the cuts designed to limit faint magnitudes or to exclude outliers but must be considered thoroughly when it comes to the  $d_{\perp}$  cut. This is due to the location of the  $d_{\perp}$  cut in the densest region of the colour space. Many galaxies with true colours outside of the  $d_{\perp}$  cut have scattered into the sample, while a similar amount of galaxies with true colours within the  $d_{\perp}$  cut could have scattered out of the SDSS selection. From this discussion, we infer that the  $d_{\perp}$  cut used to obtain the BOSS CMASS sample, in terms of true properties, is not a sharp cut shown in Fig. 1, but should instead be a form of likelihood function that accepts or rejects galaxies in a probabilistic way based on galaxy colours and photometric errors.

Based on the three reasons we listed earlier, we constructed a model that can handle the colour mismatch and probabilistic selection near the  $d_{\perp}$  cut all together.

### 3 CONSTRUCTING THE MODEL

While BOSS and DES operate in different hemispheres, the survey footprints overlap in an equatorial area of the sky known as Stripe 82. DES Y1 imaged  $123 \text{ deg}^2$  of this region, thereby providing a region where data from the two surveys can be matched.

By using the photometric information in the overlapping region, we build an algorithm for probabilistic target selection that uses density estimation in colour and magnitude spaces. The general concept of the algorithm is described in Section 3.1. The algorithm is trained in half of the overlapping region and validated in the other half. We discuss the training and validation data sets in Section 3.2.

<sup>7</sup>The term ‘model’ magnitudes here indicates ‘modelMag’ magnitudes used in the BOSS selection criteria.

<sup>8</sup>Based on the information available at catalogues in Section 2.

The tools and detailed fitting procedures for training are presented in Section 3.3. The results of validation and application of the algorithm to the target galaxies can be found in Sections 3.4 and 3.5.

### 3.1 Overview of the algorithm

The probability of being part of the CMASS sample for a source having a property  $\theta$  can be written as the combination of the likelihood and the prior according to Bayes' theorem:

$$P(C|\theta) = \frac{P(\theta|C) P(C)}{P(\theta)}, \quad (15)$$

where

$$P(\theta) = P(\theta|C)P(C) + P(\theta|N)P(N). \quad (16)$$

The notation  $C$  is the class of CMASS, and  $N$  is the class of non-CMASS galaxies.  $P(C)$  is the prior probability that a selected source is part of the CMASS sample, which can be interpreted as the fraction of CMASS in the total galaxy sample.  $P(\theta|C)$  is the likelihood of the source under the probability density function (pdf) of the property  $\theta$  of CMASS. The pdf of the property  $\theta$  of CMASS can be constructed from a histogram of CMASS as a function of  $\theta$ . However, since we use noisy quantities such as observed colours and magnitudes, the resulting pdf might be biased by photometric errors that vary by observing conditions. For example, if the training region has a uniquely different observing condition from other regions, the pdf model drawn from the training galaxies will not accurately represent CMASS. Therefore, the pdf should take into account measurement errors.

To ensure a uniform selection across the survey, we use the XD technique first proposed in Bovy et al. (2011a). The XD algorithm models the observed distribution of data as a mixture of Gaussians, convolved with a multivariate Gaussian model for the measurement errors on each point. It iterates through expectation and maximization steps to solve for the maximum-likelihood estimates of the parameters specifying the underlying mixture model.

The underlying distribution  $P(\theta_{\text{true}}|C)$  derived from XD is an unbiased pdf free from regional measurement errors in the training set. By convolving the underlying distribution back with the measurement uncertainties of the validation sample, we can infer the observed distribution that the same kind of galaxies would have in the validation region. For a given observation  $\theta = \{\theta_{\text{obs}}, \epsilon\}$ , the observed quantity  $\theta_{\text{obs}}$  with a corresponding measurement uncertainty  $\epsilon$ , the likelihood of CMASS is written as

$$P(\{\theta_{\text{obs}}, \epsilon\}|C) = \int d\theta_{\text{true}} p(\{\theta_{\text{obs}}, \epsilon\}|\theta_{\text{true}}) P(\theta_{\text{true}}|C). \quad (17)$$

The first factor  $p(\{\theta_{\text{obs}}, \epsilon\}|\theta_{\text{true}})$  on the right-hand side stands for the distribution function of measurement uncertainty of  $\theta$  in the presence of known measurement uncertainty  $\epsilon$ . We assume that the measurement uncertainty distribution of bright galaxies such as CMASS is nearly a Gaussian with an rms width  $\epsilon$ . The same procedure is repeated for non-CMASS galaxies.

Considering all factors, the resulting posterior probability that will be assigned to a target source having a property  $\theta_{\text{obs}}$  with a measurement uncertainty  $\epsilon$  is given as

$$P(C|\{\theta_{\text{obs}}, \epsilon\}) = \frac{\int d\theta_{\text{true}} p(\{\theta_{\text{obs}}, \epsilon\}|\theta_{\text{true}}) P(\theta_{\text{true}}|C)P(C)}{\int d\theta_{\text{true}} p(\{\theta_{\text{obs}}, \epsilon\}|\theta_{\text{true}})P(\theta_{\text{true}})}, \quad (18)$$

where

$$P(\theta_{\text{true}}) = P(\theta_{\text{true}}|C)P(C) + P(\theta_{\text{true}}|N)P(N). \quad (19)$$

### 3.2 The training and validation sets

We use the overlapping area between BOSS and DES to train and validate the algorithm. To label DES galaxies as CMASS or non-CMASS galaxies, we cross-match the refined DES Y1 Gold catalogue (described in Section 2) to the BOSS CMASS photometric sample<sup>9</sup> using a 2" tolerance. The total number of galaxies labelled as CMASS is 12 639 over the area of 123 deg<sup>2</sup>.

The labelled DES galaxies are split into the training and validation sets. In the overlapping region, the number density of galaxies varies along latitude. Since our probabilistic model assumes that the galaxies are homogeneously distributed in the full sky, we divided the overlapping area into HEALPix<sup>10</sup> (Górski et al. 2005) pixels of resolution  $N_{\text{side}} = 64$  in NEST ordering and took only even values of HEALPix pixels as the training sets to populate the training regions uniformly. The total training set contains 6325 CMASS galaxies and 340 202 non-CMASS galaxies in 62.5 deg<sup>2</sup>. The two samples are used separately to train the algorithm to construct the likelihoods for CMASS and non-CMASS galaxies. Note that this division is used only to test the algorithm and we will later switch to the full Stripe 82 region as the training set for the DES SPT region.

### 3.3 Obtaining true distributions with the XD algorithm

The XD algorithm developed by Bovy et al. (2011a) is a generalized Gaussian mixture model approach to density estimation and is designed to reconstruct the error-deconvolved true distribution function common to all samples, even when noise is significant or there are missing data. Starting from the random initial guess of Gaussian mixtures, the algorithm iteratively calculates the likelihood by varying means and widths of Gaussian components until it finds the best fit of Gaussian mixtures.

We use the Python version of the XD algorithm in the `AstroML`<sup>11</sup> package (VanderPlas et al. 2012). The following four DES properties are fitted by the XD algorithm: ( $G_{\text{DET}} - R_{\text{DET}}$ ), ( $R_{\text{DET}} - I_{\text{DET}}$ ),  $R_{\text{MOD}}$ , and  $I_{\text{MOD}}$ . The two DES colours are selected as they mirror the SDSS information used for the  $d_{\perp}$  cut. The apparent magnitude  $I_{\text{MOD}}$  is included to extract information induced by the cut given in equation (1). There is no  $r$ -band magnitude cut in the CMASS selection criteria, but we include  $R_{\text{MOD}}$  in order to provide extra information to capture the differences between the SDSS and DES filter bands. Star-galaxy separation was performed on DES photometry with the flag `MODEST==1` (Drlica-Wagner et al. 2018); therefore, we do not apply any further cuts to replace cuts (4) and (5).

The `AstroML` XD algorithm leaves the initial number of Gaussian mixture components as a user's choice. One of the well-known methods for choosing the correct number of components is to use the Bayesian information criterion (BIC; Schwarz 1978). We use the Gaussian mixture module in the `scikit-learn`<sup>12</sup> package (Pedregosa et al. 2012) to compute the BIC scores for a different number of components. The optimal number of components found

<sup>9</sup>We do not use the BOSS spectroscopic sample for training. The spectroscopic sample of BOSS CMASS has about 5.8 and 1.8 per cent of missing targets lost by ‘fibre collision’ and ‘redshift failure’, respectively (Reid et al. 2016). Since our probabilistic model is colour based, we utilize the BOSS photometric sample for training in order to include the photometry information from those missing galaxies.

<sup>10</sup><http://healpix.sourceforge.net>

<sup>11</sup><http://www.astroml.org>

<sup>12</sup><https://scikit-learn.org/>

by this exercise is 8 for the CMASS training set and 26 for the non-CMASS training set.

The XD algorithm fits the multidimensional histogram of the four aforementioned DES properties with the optimal number of Gaussian mixtures and returns the values of amplitudes, means, and widths of the best-fitting Gaussian mixture model. The resulting best-fitting model is used as a true distribution  $P(\theta_{\text{true}}|C)$  in equations (18) and (19).

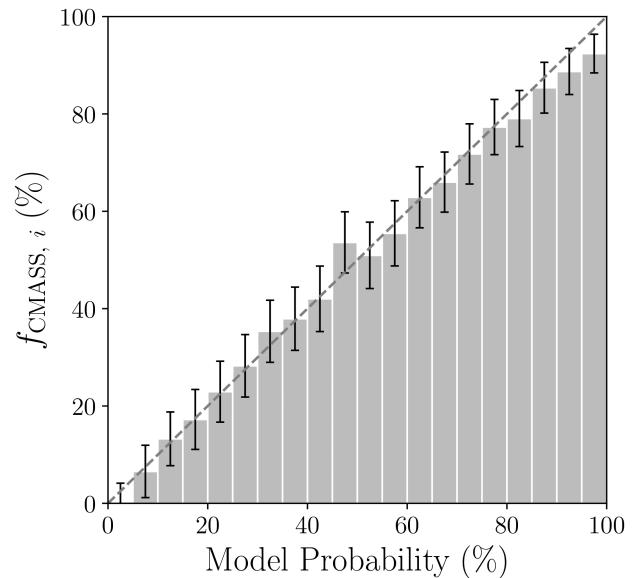
Throughout this work, we assume that there is no correlation between different bands.

### 3.4 Application to the target galaxies

In this section, we apply our probabilistic model to the validation galaxies in order to validate the algorithm. The underlying distributions  $P(\theta_{\text{true}}|C)$  and  $P(\theta_{\text{true}}|N)$  are obtained from the XD algorithm as described in the previous section. The Bayesian priors are given as  $P(C) = 0.018$  and  $P(N) = 1 - P(C)$ . This is based on the fraction of CMASS galaxies in the training set.<sup>13</sup>

The probability of being part of CMASS for a given property  $\theta$  is analogous to the probability of finding a CMASS galaxy in a group of galaxies having the same property  $\theta$ . This implies that in a group of galaxies assigned the same model probability, the assigned probability should be identical to a fraction of galaxies labelled as CMASS. To confirm this argument, we bin validation galaxies in 20 bins based on their assigned probability. In Fig. 3, the x-axis shows the 20 bins of the assigned probability, and the y-axis shows the fraction of true CMASS galaxies in each bin. The grey bars are the fractions of CMASS-labelled galaxies in the validation set with Poisson error bars. The diagonal dashed line represents the ideal case that the probabilistic model would yield if the model successfully recovers the observed distribution of CMASS in the target region. The computed fractions of CMASS show good agreement with the ideal case within error bars. Once all galaxies in the target region are assigned a model probability, the widely accepted next step for classification is dividing target sources into two categories with a threshold probability  $> 50$  per cent. A similar probabilistic approach was used in Bovy et al. (2011b) to distinguish quasars from stars. However, we take a different approach since we are not interested in classifying individual galaxies accurately, but instead we focus on matching the statistical properties of groups of galaxies. In order to produce a statistical match, the membership probability we determine must faithfully reflect the probability that an object would be selected into the BOSS CMASS sample based on SDSS imaging.

Fig. 4 presents the histograms of the  $d_{\perp}$  colour of true CMASS in the training set (blue shaded histogram) binned in the different ranges of the membership probability bins. Training galaxies in low-probability bins tend to have low  $d_{\perp}$  values because of their proximity to the  $d_{\perp}$  cut.  $f_{\text{train}}$  in the top-right corner of each panel is the fraction of training galaxies binned in each probability bin, defined as  $f_{\text{train},i} = N_{\text{train},i}/N_{\text{train},\text{total}}$  for the  $i$ th probability bin. Over the 10 probability bins, training galaxies are distributed uniformly, with a relatively high fraction in the lowest and the highest probability bins. This indicates that galaxies



**Figure 3.** Accuracy of model membership probabilities assigned to the validation set. Galaxies in the validation set are binned based on their probability assigned by the probabilistic model. The x-axis shows 20 bins of the assigned probability, and the y-axis shows the fraction of true CMASS galaxies in each bin. If the model successfully recovers the observed distribution of CMASS in the validation region, the fraction of true CMASS galaxies in each bin should be identical to the assigned model probability. The dashed diagonal line in the figure stands for this ideal case, and the grey bars are given by Poisson errors.

having low membership probabilities contribute to the CMASS sample as significantly as galaxies having high membership probability.

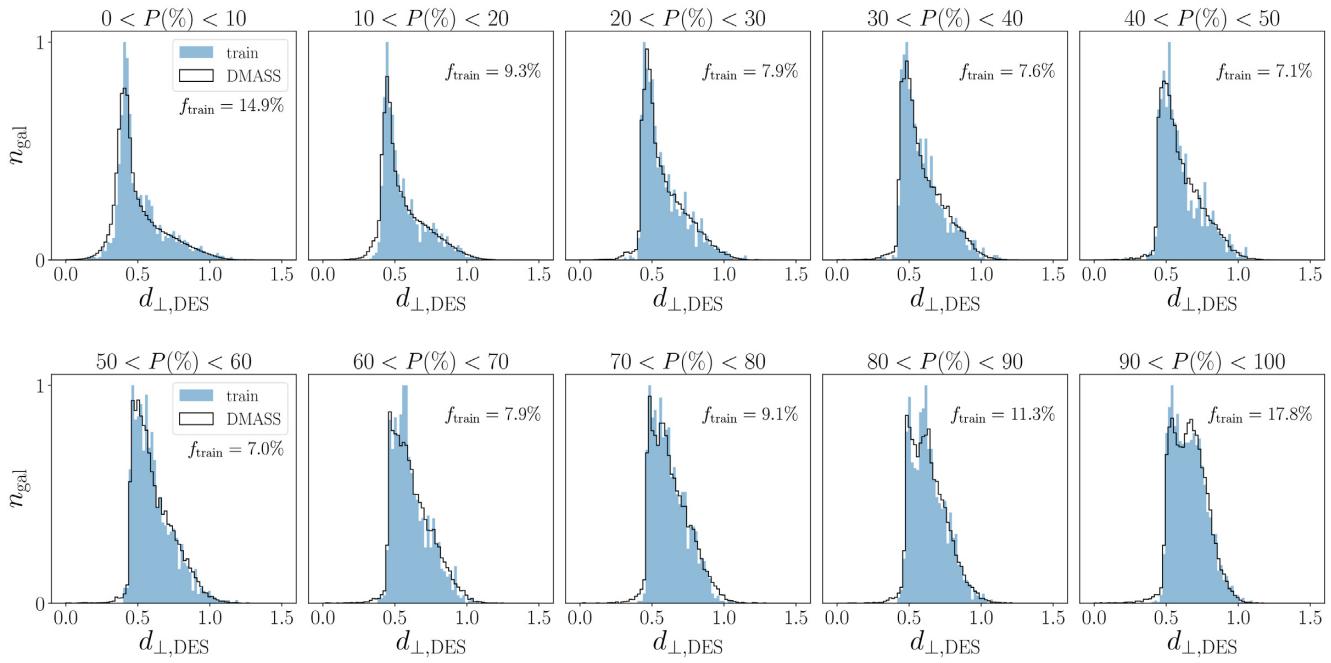
From this, we can infer that in order to generate the same noise level that the original CMASS sample intrinsically has, galaxies should be populated based on their membership probability in the same way that ones in the CMASS sample are. In this sense, the model probability suggests a natural way of how we should make use of the assigned probabilities. We can either sample or weight a galaxy by its assigned probability in order to produce a sample that is a statistical match to the BOSS CMASS sample. Throughout the rest of the work, we use the membership probability as weights.

From now on, we apply the validated algorithm to the DES galaxies outside the training area. If not specified otherwise, the DMASS sample only refers to the DES Y1 Gold galaxies in the SPT region weighted by the assigned membership probability. The black solid lines in Figs 4 and 5 show the various property distributions of the DMASS sample. Fig. 4 shows that the weighting scheme successfully reproduces the noisy quantity  $d_{\perp}$  by populating each probability bin with the DES galaxies (black solid) as the CMASS galaxies (blue) are distributed. Fig. 5 shows the distributions of colours and magnitudes that are used to train the algorithm. The resulting DMASS distributions (black solid) of the colours and magnitudes are in good agreement with the distributions of the training sample (blue).

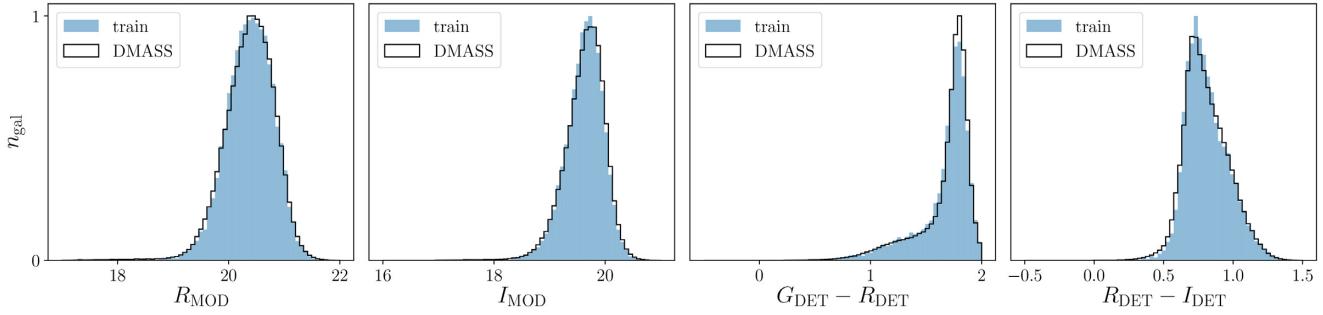
<sup>13</sup>The fraction of CMASS may vary depending on the observing condition of the selected area, but we take the value in the training sample as a global prior for simplicity, assuming CMASS galaxies are homogeneously distributed in the Universe. We will show that this approximation can be justified through validation tests later in this paper.

### 3.5 Excluding low-probability galaxies

The DES Y1 Gold catalogue contains a lot of galaxies that are much fainter than CMASS. This implies the majority of the DES



**Figure 4.** Histograms of the  $d_{\perp}$  colour distributions of CMASS (blue) in the training set and DMASS in SPT (solid black line) in 10 membership probability bins.  $d_{\perp}$  on the  $x$ -axis consists of only DES quantities ( $d_{\perp, \text{DES}} = (R_{\text{DET}} - I_{\text{DET}}) - (G_{\text{DET}} - R_{\text{DET}})/8.0$ , where the subscript ‘DET’ denotes DES MAG\_DETMODEL magnitude).  $f_{\text{train}}$  in the top-right corner of each panel denotes the fraction of training galaxies binned in each probability bin, defined as  $f_{\text{train},i} = N_{\text{train},i} / N_{\text{train},\text{total}}$  for the  $i$ th probability bin.



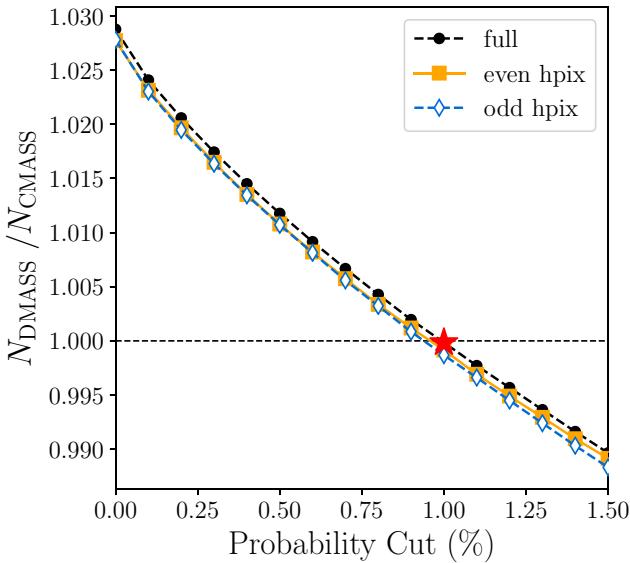
**Figure 5.** Histograms of colour and magnitude distributions of CMASS in the training set (blue) and DMASS in SPT (black solid line). The colours and magnitudes on the  $x$ -axis are DES quantities.

galaxies have extremely low CMASS probabilities. These galaxies are likely to only add noise to the sample and potentially bias our measurements and therefore need to be removed. We carefully test how the low-probability portion of the training sample (even HEALPix pixels) affects the number density. We remove all galaxies lower than a given probability threshold and compare the number density of each sample with those of CMASS in the training sample. Including all sources results in the number density of the resulting sample being about 3 per cent higher than CMASS in the same region, but near a threshold cut  $P > 1$  per cent, the sample yields a similar number density as CMASS. To validate the threshold cut, we construct a model in the same way but by using only the validation sample (odd HEALPix pixels). Fig. 6 shows that the model from the validation sample produces low-probability galaxies that affects the number density of the sample in a very similar way as the training sample. The similarity of the curves from different samples implies that the model tends to boost the number density of the sample in a predictable way, and this

tendency can be remedied by cutting out low-probability galaxies below a certain threshold. The same procedure is performed for the full Stripe 82 region and yields the same number density as CMASS for a threshold cut  $P > 1$  per cent (black points in Fig. 6). Throughout this work, we use a threshold cut  $P > 1$  per cent. This cut excludes  $\sim 90$  per cent of sources in the DMAPS catalogue, but when considering membership probabilities as weights, the effective portion of galaxies eliminated is 2.96 per cent. After applying the probability cut, we determine the effective sample size of the complete DMAPS catalogue by summing the weights. We find the sample size of the DMAPS sample is 117 293.

#### 4 SYSTEMATIC ERROR CHARACTERIZATION

Astrophysical foregrounds, observing conditions, and spatially varying depth are potential sources of systematic uncertainty in galaxy survey analyses. They affect the probability of detecting



**Figure 6.** Number density of DMASS with respect to the probability cut computed from three different regions – the training region (even HEALPix pixels; square), the validation region (odd HEALPix pixels; circle), and the full region (even + odd HEALPix pixels; black diamond). Galaxies below a given probability cut are excluded. Number densities are divided by the number density of CMASS in the corresponding regions. The extremely similar shape of curves from different samples implies that the model tends to boost the number density of the sample in a predictable way and this tendency can be remedied by cutting out low-probability galaxies below a certain threshold. The red star at  $P = 1$  per cent on the black diamond curve is our choice of the probability cut.

sources and also their reconstructed properties, and can thereby result in systematic biases in cosmological analyses (Crocce et al. 2016; Leistedt et al. 2016).

We follow the procedures described in Elvin-Poole et al. (2018) to identify and correct for these kind of systematic biases on the DMASS sample in the SPT region.<sup>14</sup> To search for potential systematic uncertainties that affect galaxy clustering, we study the correlations between the galaxy number density and survey properties. If the galaxy density is independent of a survey property, we do not consider this property to have an impact on our DMASS sample. We use HEALPix maps ( $N_{\text{side}} = 4096$ ) of four observing conditions (airmass, seeing full width at half-maximum (FWHM), sky brightness, exposure time),  $10\sigma$  limiting depth in  $griz$  bands, and the two astrophysical foregrounds of galactic reddening ( $N_{\text{side}} = 1024$ ) and stellar density ( $N_{\text{side}} = 512$ ). A detailed description about constructing HEALPix survey property maps can be found in Leistedt et al. (2016). The construction of stellar density maps is described in Elvin-Poole et al. (2018). The SFD galactic dust map is available at the LAMBDA website<sup>15</sup> (Schlegel et al. 1998).

We mask HEALPix pixels where the galaxy number density deviates by more than 20 per cent from the mean value (1.0) or changes sharply after some threshold value. We mask HEALPix pixels where seeing FWHM in  $r$ -band  $>4.5$  pixels, which removes 2 per cent of the total area.

<sup>14</sup>We do not find any systematic biases from CMASS in the training region. Therefore, systematics addressed in this section were not considered for modelling the probabilistic model.

<sup>15</sup><https://lambda.gsfc.nasa.gov/product/foreground>

Prior to correcting systematics, we rank survey properties from the most to least significance. The survey properties are ranked in order of the value given by

$$\Delta\chi^2 = \chi^2_{\text{null}} - \chi^2_{\text{model}}, \quad (20)$$

where  $\chi^2_{\text{model}}$  is the difference in  $\chi^2$  between the best-fitting model of the number density and data points, and  $\chi^2_{\text{null}}$  is  $\chi^2$  against a null line  $n_{\text{gal}}/\langle n_{\text{gal}} \rangle = 1$ . We minimize  $\chi^2_{\text{model}}$  by fitting a linear model  $N_{\text{gal}} \propto As + B$  against the calculated number density with Poisson errors of each data point.

After ranking properties, we correct for them starting from the highest ranked one using the inverse of the best-fitting model as a weight. Since survey properties are correlated with each other, correcting one survey property can introduce new systematic trends from another survey property. Therefore, the relationship between the galaxy number density and survey properties is recalculated after applying a weight. Then, one moves to the next top-ranked survey property and iterates the procedure.

The weighting scheme we use assumes that the effects of each survey property are separable. However, there is some correlation between systematic maps that may result in overcorrecting the galaxy density field for a large number of systematic maps (Elsner, Leistedt & Peiris 2016). To avoid this, we calculate the impact of adding a systematic weight in every iteration to choose the minimum possible number of survey properties to be corrected.

To investigate the impact of including additional systematic corrections, we utilize the angular correlation function. The angular correlation function  $w^{\delta_g \delta_g}(\theta)$  is computed with systematic weights using the Landy–Szalay estimator (Landy & Szalay 1993) as given by

$$w^{\delta_g \delta_g}(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}, \quad (21)$$

where  $DD$ ,  $DR$ , and  $RR$  are the number of galaxy pairs, galaxy–random pairs, and random pairs separated by a distance  $\theta$ . Systematic weights are applied to individual galaxies as

$$DD(\theta) = \frac{1}{N_{DD}(\theta)} \sum_i^{N_{\text{gal}}} \sum_j^{N_{\text{gal}}} w_i w_j \Theta(\theta_i - \theta_j), \quad (22)$$

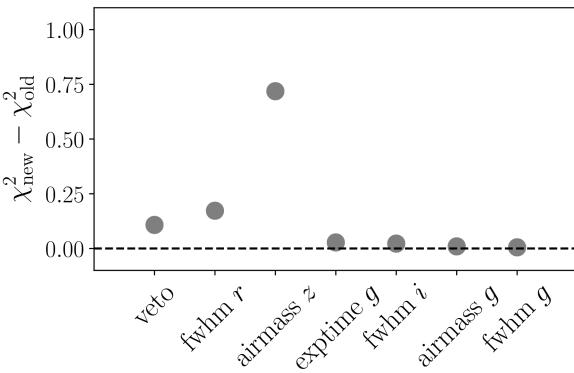
$$DR(\theta) = \frac{1}{N_{DR}(\theta)} \sum_i^{N_{\text{gal}}} \sum_j^{N_{\text{rand}}} w_i w_j \Theta(\theta_i - \theta_j), \quad (23)$$

$$RR(\theta) = \frac{1}{N_{RR}(\theta)} \sum_i^{N_{\text{rand}}} \sum_j^{N_{\text{rand}}} w_i w_j \Theta(\theta_i - \theta_j), \quad (24)$$

where  $w_i$  denotes systematic weight ( $w_i = 1$  for randoms),  $N(\theta)$  is the total number of pairs in a given data set in a given angular bin  $\theta$ , and  $\Theta(\theta_i - \theta_j)$  is 1 if a pair lies at an angular distance  $\theta$ , otherwise zero. The correlation function is measured in 10 logarithmically spaced angular bins over the range  $2.5 \text{ arcmin} < \theta < 250 \text{ arcmin}$ . We adopt the same scales for cross-correlation functions with other surveys later in the paper. All two-point calculations are done with the public code TreeCorr<sup>16</sup> (Jarvis 2015).

Randoms for DMASS are uniformly generated on the surface of a sphere and masked by the same masks described in Section 2. The number density of randoms is chosen to be 50 times larger

<sup>16</sup><https://github.com/rmjarvis/TreeCorr>



**Figure 7.** The impact of systematic weights. Starting from the left, the names of the survey properties are listed on the  $x$ -axis in the order that they are corrected. The weight for the particular property is applied on top of the other weights applied earlier. The  $y$ -axis shows the  $\chi^2$  measured between the correlation function with the new and old weights. ‘veto’ denotes a veto mask applied to remove regions where FWHM in  $r$ -band  $> 4.5$ .

than DMASS, minimizing the impact of any noise from the finite number of randoms and matching the relative number of CMASS randoms.

To construct a covariance matrix for DMASS, we first compute a covariance matrix for CMASS from the 1000 QPM CMASS mock catalogues used in the BOSS-III analyses (Alam et al. 2015):

$$\mathbf{C}(\omega_i, \omega_j) = \frac{1}{N_{\text{mock}} - 1} \sum_{k=0}^{N_{\text{mock}}} (w_i^k - \bar{w}_i)(w_j^k - \bar{w}_j), \quad (25)$$

where  $N_{\text{mock}}$  is the total number of mocks,  $w_i$  represents the  $i$ th bin of the angular correlation function,  $w_i^k$  denotes the  $i$ th bin of the angular correlation function from the  $k$ th mock, and  $\bar{w}$  is the average value of  $w$  over all mocks.

From the resulting CMASS mock covariance matrix, we derive a covariance matrix for DMASS by using the analytic form of the covariance between the angular correlation functions as follows:

$$\mathbf{C}(\theta, \theta') = \frac{(2l + 1)^2}{f_{\text{sky}}(4\pi)^2} \sum_{l=0} P_l(\cos \theta) P_l(\cos \theta') \sigma^2(C_l) + \frac{\delta_{\theta, \theta'}}{n_{\text{pairs}}}, \quad (26)$$

where  $\sigma^2(C_l)$  is the variance of the angular power spectrum  $C_l$ ,  $f_{\text{sky}}$  is the fraction of the sky, and  $n_{\text{pairs}}$  is the total number of galaxy pairs. Assuming DMASS and CMASS have the same galaxy bias and redshift distribution, the first term can be easily adjusted for DMASS by altering the survey area factor. The second term, the shot noise term, can be directly calculated from the data. We obtain the first term of the CMASS covariance matrix by subtracting the inverse of pair counts of the CMASS galaxies from the mock covariance, and rescaling it by the ratio of the survey areas. The derived form of the covariance matrix for DMASS is

$$\mathbf{C}_D = \frac{A_D}{A_C} \left( \mathbf{C}_{\text{mock}, C} - \frac{\delta_{\theta, \theta'}}{n_{\text{pairs}, C}} \right) + \frac{\delta_{\theta, \theta'}}{n_{\text{pairs}, D}}, \quad (27)$$

where  $A$  is the survey area,  $\mathbf{C}_{\text{mock}}$  is the mock covariance, and the subscripts  $C$  and  $D$  stand for CMASS and DMASS, respectively.

Fig. 7 shows the impact of including additional systematic corrections through the value of  $\chi^2$ , computed from the rescaled covariance matrix and residuals between the two measurements – before and after correction. The systematic weights are listed on the  $x$ -axis in the order that they are applied to on top of all the previous weights applied. Survey properties that show notable impacts are

FWHM  $r$ -band and airmass  $z$ -band. Applying corrections for the rest of the survey properties barely affects the angular correlation function. Therefore, we apply systematic weights only for the top two properties. Fig. 8 shows the galaxy number density versus survey property plots before applying the weights (blue) and after (black). We additionally find that correcting the top two systematics removes any trend with stellar number density. Our interpretation is that any trend with the stellar number density is not from pure stellar contamination but from strong correlations between the FWHM and airmass maps.<sup>17</sup>

In the later sections, we will apply the systematic weights and veto mask computed in this section to the DMASS sample and report results along with the no systematics case.

## 5 COMPARISON WITH THE BOSS CMASS SAMPLE

In this section, we compare the properties of the DMASS sample to those of the BOSS CMASS sample. We evaluate the consistency of the overall number density, the amplitude of the auto- and cross-correlation functions, and redshift distribution.

As described in Ross et al. (2011, 2012), the selection functions for BOSS galaxy data in the North Galactic cap (NGC) and the SGC are slightly different due to measurable offsets in the DR8 (Aihara et al. 2011) photometry between the two regions. DMASS tends to mimic SGC CMASS as the training set taken is a subsample of CMASS in the SGC. Therefore, we will specifically compare DMASS with SGC CMASS in addition to comparisons with the full CMASS sample.

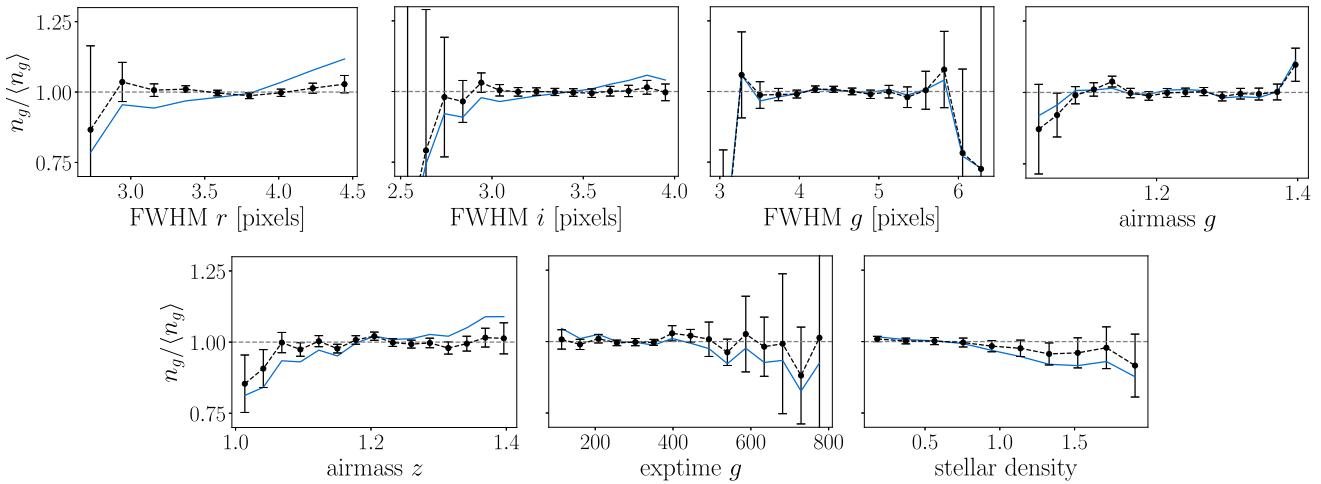
### 5.1 Number density

In this section, we will compare the number density of CMASS in the training data, which is from  $123 \text{ deg}^2$  of Stripe 82 area, to the mean density in three distinct footprints: (1) BOSS CMASS data in the NGC area; (2) BOSS CMASS data in the SGC area; and (3) the DMASS data in the SPT area. We divide each of the three regions into many smaller patches that are the size of Stripe 82. This allows us to determine the expected variance between the number density in the training area and the full region.

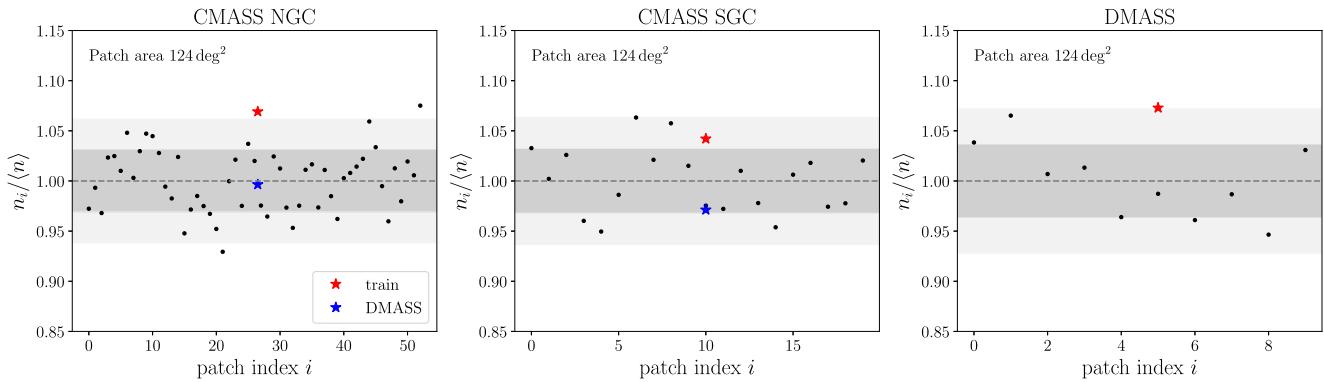
The three large footprints are divided as follows: each region is split into HEALPix pixels at resolution  $N_{\text{side}} = 4096$  where the size of each pixel is  $0.72 \text{ arcmin}^2$ . Then, contiguous sets of  $\sim 606\,000$  pixels are combined to make each patch comparable to the size of Stripe 82. We adopt a slightly larger size for one patch,  $124 \text{ deg}^2$ , in order to include all of the HEALPix pixels in the SPT region while keeping the size of all patches the same. The same patch size is applied for the SGC and NGC regions and the remaining HEALPix pixels that cannot be a complete patch are discarded. The number of patches used for this calculation is 10 for the DES SPT region, 53 for the NGC region, and 20 for the SGC region.

Fig. 9 shows the number density deviation for CMASS in the NGC (left), CMASS in the SGC (middle), and DMASS in the SPT region (right). The black dots are the number density values determined in each of the small patches. All values are divided by the mean number density of each panel. The grey shaded region is the standard deviation of the distribution of the black dots. This represents an estimate for the  $1\sigma$  uncertainty in the number density

<sup>17</sup>The lack of correlation with stellar density is consistent with the results for the DES redMaGiC galaxies at similar redshifts (Elvin-Poole et al. 2018).



**Figure 8.** Galaxy number density with respect to survey properties having the top six  $\Delta\chi^2$  and stellar density. The solid blue lines are calculated without correction weights. The dashed black lines are calculated with weights. The error bars on the black lines are calculated assuming Poissonian statistics.



**Figure 9.** Number density and its deviation in the NGC (left), CMASS in the SGC (middle), and D MASS (right). Each region is divided into Stripe 82-sized (train region) patches. Red stars represent the number density of BOSS CMASS in the training region and blue stars are the total number density of D MASS. All values are divided by the mean number density of each CMASS sample. The dark grey shaded region is  $1\sigma$ , and the light grey shaded region is the  $2\sigma$  level deviation of the black points in each panel.

of a Stripe 82-sized patch. All three cases show a similar level of deviations. The red star in the first and second panels is the number density of the training (BOSS CMASS) galaxies in Stripe 82 and the blue star is the total number density of D MASS. Note that the location of the stars in each panel shows the relative number density in each region. In all panels, the red star is consistently  $\sim 5\text{--}8$  per cent away from the total mean value. One can see that the number density of D MASS is considerably lower than the number density in the training region (red star), but that it is matched to within  $1\sigma$  of the overall CMASS number densities. That is, despite the data in the training region having a significantly greater number density than the overall CMASS sample, our model obtains the number density of D MASS that is a fairly good match to both CMASS SGC and CMASS NGC number densities.

## 5.2 Angular correlation function

We use the angular correlation function as a test to validate that D MASS matches the CMASS sample. Assuming that the number density and redshift distributions are matched, we should expect consistent amplitudes of the correlation functions if we have indeed matched the samples. We can directly compare the

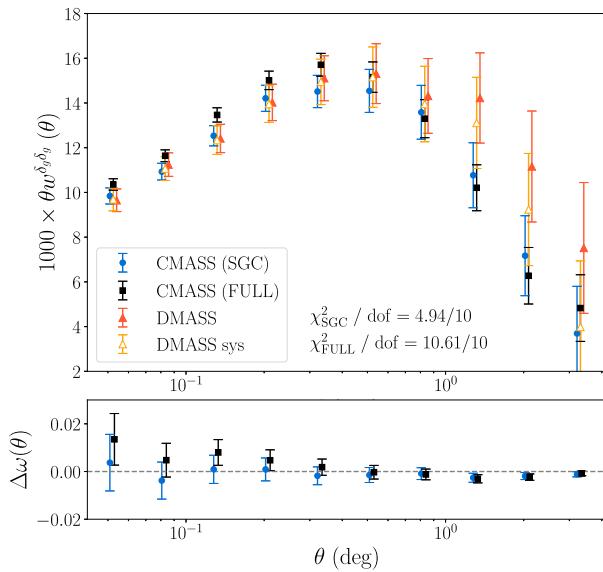
amplitude of the correlation functions of D MASS and CMASS and thus test for consistency without any cosmological assumptions. Three different probes were chosen for this comparison: the galaxy angular autocorrelation and the angular cross-correlation with two full sky surveys – the Wide-field Infrared Survey Explorer all-sky survey (WISE; Wright et al. 2010) and CMB lensing from Planck Collaboration (2016).

### 5.2.1 Auto-angular correlation function

We measure the correlation function  $w_{\delta_g \delta_g}(\theta)$  of CMASS and D MASS galaxies in the same manner as described in Section 4. Each galaxy in the CMASS sample is weighted by systematic (systot), close pair (cp), and redshift failure (zp) weights as given by Reid et al. (2016):

$$w_{\text{total}} = w_{\text{systot}} (w_{\text{cp}} + w_{\text{zp}} - 1). \quad (28)$$

Note that we do not apply these weights in the CMASS training sample because we utilize the BOSS photometric sample for training. The BOSS photometric sample includes all missing galaxies that are dropped from a spectroscopic sample due to fibre collisions and redshift failures.  $w_{\text{systot}}$  is not considered either as we do



**Figure 10.** The top panel shows the angular correlation function calculated with DMASS (red), DMASS corrected by the systematic weights (orange), CMASS SGC (blue), and full CMASS (black). The bottom panel shows residuals between DMAPS and CMASS SGC (blue) or full CMASS (black).  $\chi^2_{\text{SGC}}$  ( $\chi^2_{\text{FULL}}$ ) is the  $\chi^2$  of the observed difference of two-point functions of DMAPS and CMASS SGC (FULL).

not detect any systematic biases from the DES photometry of the CMASS training sample. As done in the previous BOSS analyses (Chuang et al. 2017; Pellejero-Ibanez et al. 2017), we apply the explicit redshift cut  $0.43 < z < 0.75$  to the BOSS CMASS sample. This redshift cut is not considered for training because we utilized only matched photometric information in the training sample. In Appendix A, we show that the redshift cut negligibly affects the 3D two-point functions of BOSS CMASS, which justifies our choice of the CMASS photometric sample as the training set. For DMAPS, galaxies in the SPT region are weighted by the CMASS membership probabilities and systematic weights calculated in Section 4.

The result is shown in Fig. 10. The blue and black data points are the angular correlations of CMASS in the SGC and full CMASS, respectively, and the red data points show the DMAPS angular correlations. Error bars are obtained from the aforementioned mock covariance matrices in Section 4. We find that the angular correlation function of DMAPS has a better agreement with CMASS in SGC than full CMASS. The angular correlation function of full CMASS is slightly higher than the other samples on small scales, as expected from the intrinsic difference between CMASS in the SGC and the NGC. On large scales, DMAPS tends to deviate from the two CMASS samples, but adding systematic weights mitigates the difference by suppressing the correlation function of DMAPS on large scales.

To quantify consistency between CMASS and DMAPS, we use a  $\chi^2$  statistic and its associated probability-to-exceed (PTE) as our primary metric. We take the observed difference of binned two-point functions  $\Delta \mathbf{d} = w_C - w_D$  (shown in the bottom panel in Fig. 10) and its associated covariance as  $\mathbf{C}_{\text{tot}} = \mathbf{C}_C + \mathbf{C}_D$ . Cross-covariance between the CMASS and DMAPS measurements is not considered since the two sets of measurements are carried out on different areas

on the sky. Then, we calculate the  $\chi^2$  of the difference defined by

$$\chi^2 = \sum_{i,j}^{N_{\text{bins}}} \Delta \mathbf{d}_i^T (\mathbf{C}_{\text{tot}}^{-1})_{i,j} \Delta \mathbf{d}_j, \quad (29)$$

and its associated PTE with the degrees of freedom (the number of bins). A probability of  $(100 - \text{PTE})$  per cent = 68 per cent (95 per cent) corresponds to  $1\sigma$  ( $2\sigma$ ) difference.

The  $\chi^2/\text{dof}$  obtained between DMAPS and CMASS SGC is  $4.94/10$  (PTE = 90 per cent) in the range  $2.5 \text{ arcmin} < \theta < 250 \text{ arcmin}$ . For the comparison with the full CMASS sample, we obtain a  $\chi^2/\text{dof}$  of  $10.67/10$  with PTE = 53 per cent. With the systematic weights, we obtain  $2.58/10$  (PTE = 99 per cent) for CMASS in the SGC and  $8.60/10$  (PTE = 47 per cent) for full CMASS.

### 5.2.2 Cross-correlation with WISE galaxies

The WISE satellite surveys 99.86 per cent of the entire sky at wavelengths of  $3.4, 4.6, 12$ , and  $22 \mu\text{m}$  ( $W_1$  through  $W_4$ ). To have a uniform galaxy data set, we select sources to a flux limit of  $W_1 < 15.2$  and remove stars with the cuts  $W_1 - W_2 < 0.2$  and  $W_2 - W_3 < 2.9$ , following Goto, Szapudi & Granett (2012). Regions contaminated by scattered moonlight are excluded by the ‘moonlev’ flag if at least one of the bands has a value higher than 3 (Kovacs et al. 2013). We also remove regions having the extreme level of galactic extinction,  $0.367 \times E(B - V)_{\text{SFD}} > 0.05$ .

The resulting WISE galaxy sample approximately spans the redshift range from 0 to 0.4 with median redshift  $z \sim 0.15$  (Goto et al. 2012; Kovacs et al. 2013). CMASS in the SGC is known to have 5.24 per cent of galaxies and CMASS in the NGC has 3.73 per cent in the low-redshift tail  $z < 0.43$ . If the probabilistic model effectively reproduces the  $d_{\perp}$  cut in the DES photometry, the DMAPS sample would have a similar fraction of galaxies in the low-redshift tail and this would result in the same cross-correlation signal.

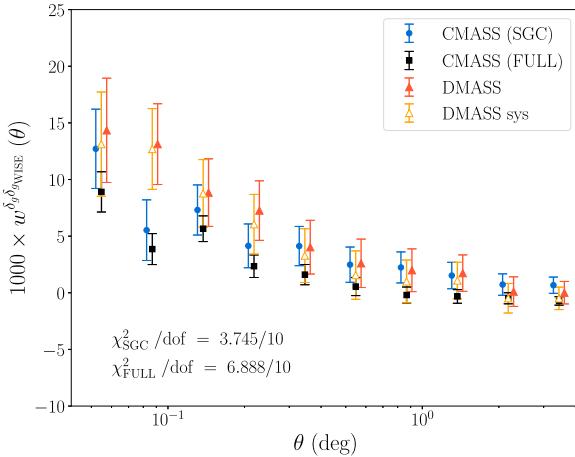
We adopt the Landy–Szalay estimator for the cross-correlation given as

$$w^{\delta_g \delta_{\text{WISE}}}(\theta) = \frac{DD_W - DR_W - D_W R + RR_W}{RR_W}, \quad (30)$$

where  $D_W$  and  $R_W$  stand for WISE galaxies and WISE randoms, respectively. WISE randoms are uniformly generated on the surface of a sphere within the masked region, with a size 50 times larger than the WISE galaxy sample.

Errors are derived from analytic covariance matrices. We calculate the covariance matrices of the cross-correlation as the sum of the Gaussian covariance and non-Gaussian covariance, and the supersample covariance as detailed in Krause & Eifler (2017). We adopt the measurement of galaxy bias  $b_{\text{WISE}} = 1.06$  and the spectroscopic redshift distribution shown in fig. 3 in Goto et al. (2012).

With the same angular binning choice as the autocorrelation function, we measure the cross-correlation function between WISE galaxies and SGC CMASS, full CMASS, and DMAPS as shown in Fig. 11. The cross-correlation of full CMASS shows a slightly lower amplitude than CMASS in the SGC and DMAPS on all scales, which is expected because CMASS in the NGC has a smaller number density than CMASS in the SGC at the low-redshift end. We find that the  $\chi^2/\text{dof}$  of DMAPS computed with respect to CMASS in the SGC is  $9.04/10$  (PTE = 53 per cent), and the one computed with respect to full CMASS is  $12.12/10$  (PTE = 28 per cent). With the systematic weights, the value is  $9.70/10$  (PTE = 47 per cent) with



**Figure 11.** Cross-correlation measurements of the DMass (red), DMass with systematic weights (orange), CMASS SGC (blue), and full CMASS (black) samples with WISE galaxies (left) and CMB convergence map (right). We dropped the first three data points of the cross-correlation with CMB lensing (in grey shaded region) from the measurements of  $\chi^2$  to include only reliable scales where the analytic covariance matrices are valid.

SGC CMASS and 11.42/10 (PTE = 33 per cent) with full CMASS. From the results, we do not find strong discrepancies between any of the CMASS samples and DMass sample. The result also implies that the probabilistic model successfully reproduces the  $d_{\perp}$  cut in the DES system that excludes low-redshift objects.

### 5.2.3 Cross-correlation with the CMB lensing map

The CMB photons released from the time of last scattering ( $z \sim 1100$ ) are gravitationally deflected by the foreground mass distribution as they travel through the large-scale structure. The imprint on CMB anisotropies by this deflection of photons is called CMB lensing (Planck Collaboration 2016). The cross-correlation of galaxy positions and CMB lensing has two advantages for this work. First, CMB lensing is extremely homogeneous compared to galaxy catalogues. All information from the CMB departs from the same redshift  $z = 1100$  (considered as a very thin redshift bin) and travels the same distance until today regardless of the northern or southern part of the sky. Any difference found between the cross-correlation signals between different galaxy samples and the CMB would originate from differences between the galaxy samples themselves. Second, the galaxy bias is tied to the matter-matter correlation function in a different way that might give us complementary information.

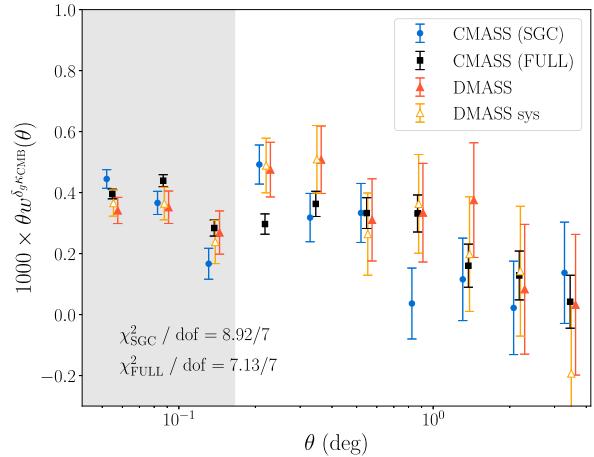
This analysis uses the 2015 CMB convergence map provided by the Planck Collaboration (2016). We use the lensing multipole range of  $8 < l < 2048$  and apply a Gaussian smoothing of  $\theta_{\text{FWHM}} = 1.71$  arcmin to the map. The cross-correlation function is calculated in 10 logarithmically spaced bins in the range  $2.5 \text{ arcmin} < \theta < 250 \text{ arcmin}$  using the estimator (Omori et al. 2019)

$$w^{\delta g \kappa_{\text{CMB}}}(\theta) = D\kappa(\theta) - R\kappa(\theta), \quad (31)$$

with

$$D\kappa(\theta) = \frac{1}{N^{D\kappa}} \sum_{i=1}^{N_{\text{gal}}} \sum_{j=1}^{N_{\text{pix}}} w_i^D w_j^{\kappa} \kappa_{i,j} \Theta(\theta_i - \theta_j), \quad (32)$$

$$R\kappa(\theta) = \frac{1}{N^{R\kappa}} \sum_{i=1}^{N_{\text{gal}}} \sum_{j=1}^{N_{\text{pix}}} w_i^R w_j^{\kappa} \kappa_{i,j} \Theta(\theta_i - \theta_j), \quad (33)$$



where  $D$  and  $R$  stand for galaxies and randoms, respectively,  $w^D$  and  $w^R$  are weights for galaxies and galaxy randoms, respectively,  $N$  in the denominator is the total number of pairs, and  $\kappa_j$  represents the value of convergence at the  $j$ th pixel.

The measurements are shown in Fig. 11. Error bars are from Gaussian covariance matrices computed by cosmoSIS (Zuntz et al. 2015). With the Gaussian covariance matrices and measured cross-correlation functions, we estimate the values of  $\chi^2$  between CMASS and DMass. We find that the  $\chi^2/\text{dof}$  value between CMASS SGC and DMass is 24.38/10 (PTE < 1 per cent) and the value between full CMASS and DMass is 21.56/10 (PTE = 2 per cent). The value between CMASS SGC and NGC is 101.48/10 (PTE < 1 per cent), which is even more extreme than the former two cases. This implies that the large  $\chi^2$  values between CMASS and DMass are not from the difference between CMASS and DMass. Since our analytic covariance matrix is Gaussian, we believe that these large values of  $\chi^2$  are due to the lack of the non-linear contributions on small scales. Therefore, we exclude data points on the scales  $\theta < 10$  arcmin and recalculated  $\chi^2$ . The minimum angular cut is motivated by the measurement of the angular correlation function in Section 5.2.1. We compare the mock covariance for the autocorrelation function with the analytic calculation and find that the analytic calculation underestimates uncertainties by more than 20 per cent at  $\theta < 10$  arcmin. We simply utilize this scale to cut out unreliable information, expecting the non-linear contribution to be dominant on a similar scale in this case. The  $\chi^2/\text{dof}$  values with the minimum scale cut are improved to 8.92/7 (PTE = 26 per cent) between CMASS SGC and DMass, and 7.13/7 (PTE = 42 per cent) between full CMASS and DMass. These values of  $\chi^2/\text{dof}$  are smaller than the  $\chi^2/\text{dof}$  between CMASS NGC and CMASS SGC (shown in Appendix B).

### 5.3 Redshift distribution

In this section, we evaluate the redshift distribution of the DMass sample by cross-correlating DMass galaxies with the DES red-MaGiC sample (Rozo et al. 2016; Elvin-Poole et al. 2018). The concept of this technique called ‘clustering-z’ is to recover redshift distributions of an unknown sample by cross-correlating it with a galaxy sample whose redshift distribution is known and accurate.

The technique was first demonstrated in Newman (2008), and has been developed and applied to various cosmological analyses, including DES (Choi et al. 2016; Rahman et al. 2016; Johnson et al. 2017; Morrison et al. 2017; Scott et al. 2018). Gatti et al. (2018) and Davis et al. (2017) calibrated redshift distributions of the DES Y1 source samples by using the DES redMaGiC sample (Elvin-Poole et al. 2018) as a reference sample. Cawthon et al. (2018) calibrated the DES redMaGiC sample by cross-correlating with the BOSS spectroscopic galaxy samples. For further details about the clustering- $z$  method, we refer interested readers to references in Cawthon et al. (2018).

We utilize the redMaGiC sample as a reference sample and follow the general procedures described in Davis et al. (2017). The redMaGiC galaxies are red luminous galaxies selected by the redMaPPer algorithm (Rykoff et al. 2014) above three different luminosity threshold cuts ( $L/L_* > 0.5$ ,  $L/L_* > 1.0$ , and  $L/L_* > 1.5$ ). These galaxies have excellent photometric redshifts with an approximately Gaussian scatter of  $\sigma_z/(1+z) < 0.02$  and cover the entire redshift range of DMASS within the full DES Y1 footprint. This makes them suitable as a reference sample to evaluate the redshift distribution of the DMASS sample. We opt for the higher luminosity redMaGiC sample selected above a luminosity threshold of  $L > 1.5L_*$  because the sample's redshifts reach up to 0.9.

To obtain the redshift distribution of the unknown sample, we split the reference sample in narrow redshift bins,  $\Delta z = 0.02$ , and measure cross-correlations between the galaxies in each redshift bin and the unknown sample. The cross-correlation for the  $i$ th redshift bin measures the quantity

$$w_{ur}(z_i) = \int_{R_{\min}}^{R_{\max}} dR \int dz' b_u(R, z') b_r(R, z_i) n_u(z') n_r(z_i) \xi_m(R, z_i, z'), \quad (34)$$

where  $b$  denotes galaxy bias of the unknown ('u') and reference ('r') samples,  $n(z)$  stands for a normalized redshift distribution,  $\xi_m$  stands for the matter–matter correlation function,  $z_i$  is the  $i$ th redshift bin of the reference sample, and  $R$  is the co-moving distance,  $R = (1+z)D_A(z)\theta$ . We adopt  $R_{\min} = 500$  kpc and  $R_{\max} = 1500$  kpc based on Gatti et al. (2018) and Davis et al. (2017). We also assume that the galaxy biases of both the reference and unknown sample are scale independent on these scales. Schmidt et al. (2013) demonstrated that implementing a linear bias model in the clustering- $z$  does not significantly affect the methodology, even if these scales are non-linear. The same point was made in Gatti et al. (2018) and Cawthon et al. (2018) as well. For narrow redshift bins  $n_r(z_i) = \delta(z - z_i)$ , equation (34) is simplified as

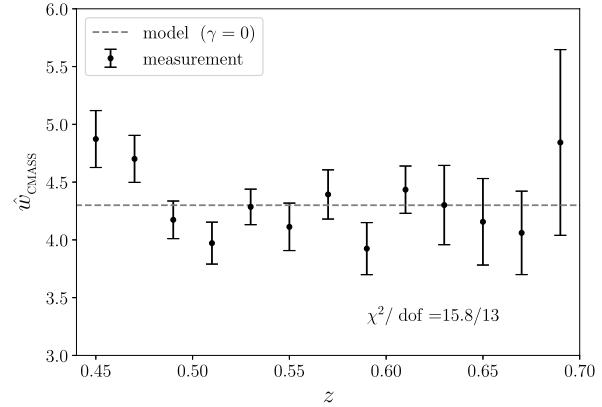
$$w_{ur}(z_i) = n_u(z_i) b_u(z_i) b_r(z_i) w_m(z_i), \quad (35)$$

where  $w_m$  is the integrated matter–matter correlation function between  $R_{\min}$  and  $R_{\max}$ . The cross-correlation was measured with the estimator from Davis & Peebles (1983) as follows:

$$\hat{w}_{ur}(z) = \frac{N_{R_r} \int_{R_{\min}}^{R_{\max}} dR W(R) [D_u D_r(R, z)]}{N_{D_r} \int_{R_{\min}}^{R_{\max}} dR W(R) [D_u R_r(R, z)]} - 1, \quad (36)$$

where  $W(R)$  is a weighting function,  $D_u D_r$  and  $D_u R_r$  stand for the number of galaxy–galaxy and galaxy–random pairs, respectively, and  $N_{R_r}$  and  $N_{D_r}$  stand for the total number of randoms and galaxies of the reference sample, respectively. With the measured cross-correlation, the redshift distribution of the unknown sample is given as

$$n_u(z) \propto \frac{\hat{w}_{ur}(z)}{b_u(z) b_r(z) \hat{w}_m(z)}. \quad (37)$$



**Figure 12.** Integrated autocorrelations (equation 36) of the CMASS SGC sample (black points). The grey dashed line is the redshift evolution model  $(1+z)^\gamma$  with  $\gamma = 0$  (equation 41). The value of  $\chi^2$  between the model and the measurement is 15.8 for 13 data points, which indicates that the measurement is well consistent with the model.

If the redshift bins are sufficiently narrow so that the biases and matter–matter correlations can be considered to be constant in each bin, the autocorrelations of the reference and unknown samples are given as

$$\hat{w}_{rr}(z) = b_r(z)^2 \hat{w}_m(z), \quad (38)$$

$$\hat{w}_{uu}(z) = b_u(z)^2 \hat{w}_m(z). \quad (39)$$

Then, equation (37) is rewritten as

$$n_u(z) \propto \frac{\hat{w}_{ur}(z)}{\sqrt{\hat{w}_{uu}(z) \hat{w}_{rr}(z)}}. \quad (40)$$

For the redshift evolution of  $\hat{w}_{uu}$ , we adopt a power-law parametrization (Cawthon et al. 2018)

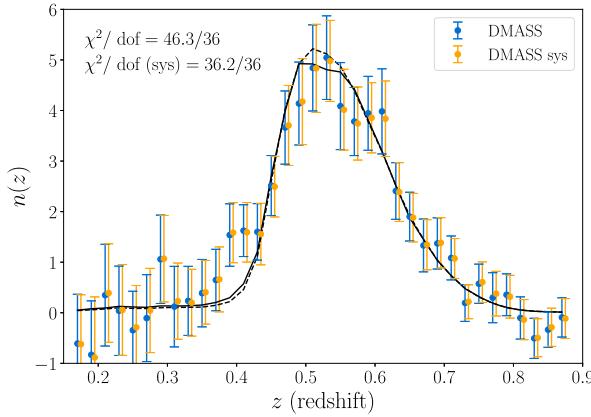
$$\sqrt{\hat{w}_{uu}(z)} \propto (1+z)^\gamma. \quad (41)$$

Since we do not have access to the true redshifts of the DMASS galaxies, we infer the redshift evolution of  $\hat{w}_{uu}$  from the autocorrelations of the CMASS galaxies using their spectroscopic redshifts (see Fig. 12). Based on the nearly constant  $\hat{w}$  of CMASS, we adopt  $\gamma = 0$  for DMASS.

Fig. 13 shows the result obtained from clustering- $z$ . We find an excellent agreement between the clustering- $z$  distribution of DMASS (blue points with error bars) and the spectroscopic redshift of CMASS SGC (solid black curve). The  $\chi^2$  obtained when comparing the two is 46.3 for 36 data points. With the systematic weights, the value is 36.2 for 36 data points. We conclude that the clustering- $z$  method returns  $n(z)$  for DMASS that is consistent with the BOSS SGC  $n(z)$ .

#### 5.4 Difference in galaxy bias

In this section, we present the constraint on the mean galaxy bias difference between DMASS and CMASS derived from the combination of different probes mentioned earlier. Due to the weaker constraining power of the cross-correlation functions compared to the autocorrelation function (Section 5.2.1) and the clustering- $z$  distribution (Section 5.3) in this section.



**Figure 13.** Redshift distribution of DMASS (blue) and DMASS with systematic weights (orange) recovered by the clustering- $z$  method with redMaGiC. The solid black and dashed lines show the spectroscopic redshift distribution of CMASS SGC and full CMASS.

To constrain the shifts in galaxy bias and redshift distribution compared to CMASS, we model the angular correlation function as follows:

$$w^{\delta_g \delta_g}(\theta, b, \Delta b, \Delta z) = \int dz f(z, b, \Delta b) \int dz' f(z', b, \Delta b) \xi_m(R, z, z'), \quad (42)$$

with

$$f(z, b, \Delta b) = (b + \Delta b) n(z + \Delta z), \quad (43)$$

where  $\xi_m$  is the matter angular correlation function,  $R$  is the co-moving distance defined as  $R = (1 + z)D_A(z)\theta$ ,  $n(z)$  is the normalized redshift distribution, and  $b$  is galaxy bias. The galaxy bias of CMASS is known to be nearly a constant within the redshift  $0.43 < z < 0.7$ , so we do not consider redshift evolution of the galaxy bias (Salazar-Albornoz et al. 2017).  $\Delta b$  and  $\Delta z$  are shifts in the galaxy bias and the redshift distribution from fiducial quantities. For CMASS,  $\Delta b$  and  $\Delta z$  are set to zero. Then, the residuals of the angular correlations of CMASS and DMASS are defined as

$$\Delta w^{\delta_g \delta_g}(\theta, b, \Delta b, \Delta z) = w^{\delta_g \delta_g}(\theta, b, \Delta b, \Delta z) - w^{\delta_g \delta_g}(\theta, b, 0, 0). \quad (44)$$

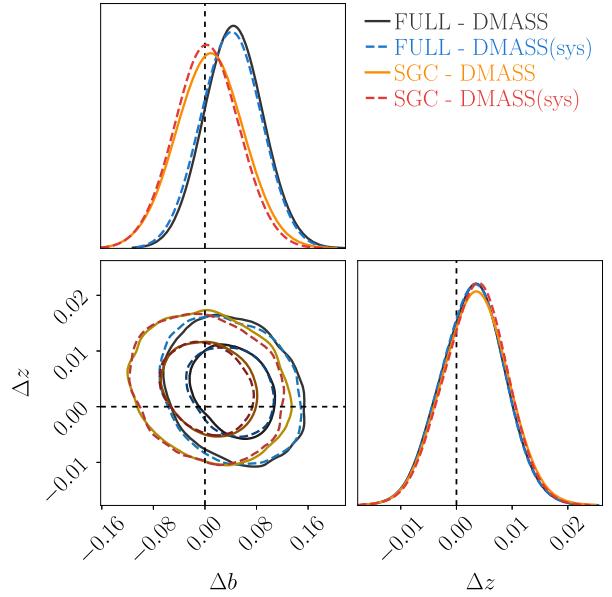
We also model the residuals of the redshift distributions to constrain the redshift shift  $\Delta z$  independently with the clustering- $z$  measurement in the previous section. The residual model is given as

$$\Delta n(z, \Delta z) = n(z + \Delta z) - n(z), \quad (45)$$

where  $\Delta z$  is the same parameter shown in equation (42). We use the spectroscopic redshift distribution of CMASS as the true distribution.

Using a combination of the residuals of the angular correlation (Section 5.2) and clustering- $z$  (Section 5.3) measurements, we perform Markov chain Monte Carlo likelihood analyses to constrain the parameter set of  $\{b, \Delta b, \Delta z\}$ . The likelihood of the combined cosmological probe is given by the sum of individual log likelihoods given as

$$\ln \mathcal{L}(\mathbf{p}) = -\frac{1}{2} (\chi^2_{w^{\delta_g \delta_g}}(\mathbf{p}) + \chi^2_{n(z)}(\mathbf{p})), \quad (46)$$



**Figure 14.** Constraints on the galaxy bias shift  $\Delta b$  and redshift distribution shift  $\Delta z$  from combination of the auto-angular correlation function and clustering- $z$ . The dashed vertical and horizontal lines show the ideal case where DMASS is perfectly matched with CMASS. Orange solid and red dashed contours show shifts from the values of CMASS SGC. Black solid and blue dashed contours show shifts from the values of full CMASS. The DMASS systematic weights are added for dashed contours. Adding systematic weights has very little impact on galaxy bias and redshift distributions.

where  $\mathbf{p}$  is the set of varied parameters. We estimate  $\chi^2$  defined in equation (29). The data vector  $\Delta \mathbf{d}$  is the difference between the measurement and theoretical prediction given as  $\Delta \mathbf{d} = \mathbf{d}_{\text{true}} - \mathbf{d}$ . Equations (44) and (45) are adopted as the true data vector  $\mathbf{d}_{\text{true}}$  for  $\chi^2_{w^{\delta_g \delta_g}}$  and  $\chi^2_{n(z)}$ , respectively. Residuals of the measurements between CMASS and DMASS are used as an input data vector  $\mathbf{d}$  for a corresponding probe as well. The covariance matrix for the angular correlation probe is given as the sum of the CMASS and DMASS covariance matrices:

$$\mathbf{C}_{w^{\delta_g \delta_g}} = \mathbf{C}_{\text{DMASS}} + \mathbf{C}_{\text{CMASS}}, \quad (47)$$

and the covariance matrix for the clustering- $z$  probe  $\mathbf{C}_{n(z)}$  is obtained from the clustering- $z$  calculation in Section 5.3. To evaluate the likelihood values and matter power spectrum for a given cosmology, we use the DES analysis pipeline in CosmoSIS (Zuntz et al. 2015). Further details of the likelihood framework are described in Krause et al. (2017).

Fig. 14 shows the constraints of the bias shift  $\Delta b$  and redshift shift  $\Delta z$ . The orange solid and red dashed contours show shifts  $\Delta b$  and  $\Delta z$  of DMASS when the values of CMASS SGC are fiducial. The black solid and blue dashed contours present shifts  $\Delta b$  and  $\Delta z$  of DMASS when full CMASS is used as fiducial. Dashed contours of both cases are obtained with the systematic weights of DMASS. The resulting numbers are  $\Delta b = 0.01^{+0.045}_{-0.052}$  and  $\Delta z = (3.46^{+5.48}_{-5.55}) \times 10^{-3}$  between CMASS SGC and DMASS, and  $\Delta b = 0.044^{+0.044}_{-0.043}$  and  $\Delta z = (3.51^{+4.93}_{-5.91}) \times 10^{-3}$  between full CMASS and DMASS. Since adding systematic weights has a negligible effect on numbers as shown in Fig. 14, we do not report the results separately. As expected, DMASS has a better agreement with CMASS in SGC. The resulting constraints of

$\Delta b$  show that the mean galaxy bias of DMASS is consistent with both CMASS samples within  $1\sigma$ . Moreover,  $\Delta b$  between DMASS and full CMASS is comparable to 2.6 per cent of the intrinsic difference in CMASS between the SGC and NGC shown in Appendix B.

In this work, we do not consider the redshift bin biases and their uncertainties of the redMaGiC samples, which are known to be  $\Delta z = (0.010, -0.004, -0.004)$  and  $\sigma_{\Delta z} = (0.011, 0.010, 0.008)$  for three redshift bins from  $z = 0.15$  to  $z = 0.6$  (Cawthon et al. 2018). Including the redshift uncertainties as priors would widen the final contours, but still keep the final constraints consistent with CMASS as all the biases are within  $1\sigma$ . Future analyses using DMASS will likely need to use a similar prior on the redshift bias for DMASS as used for DES redMaGiC in DES Y1 (Elvin-Poole et al. 2018).

## 6 CONCLUSION

In this paper, we constructed a catalogue of DES galaxies from the full footprint of DES, whose statistical properties match those of the BOSS DR12 CMASS galaxy sample. We developed an algorithm for probabilistic target selection that uses density estimation in colour and magnitude spaces. The algorithm was trained and validated by the DES photometry from the overlapping area between the DES and BOSS footprints. From the distribution of the input DES galaxies in the overlapping region, the algorithm predicts an observed distribution that the same kind of galaxies would have in the target region. A membership probability calculated based on the predicted observation was assigned to each source in the DES Y1 GOLD catalogue. By weighting galaxies by their assigned probability, the resulting DMAPS sample mimics the noise level of the original CMASS sample.

We showed that the resulting DMAPS catalogue matches well with both the SGC subset of CMASS and the full CMASS sample in various aspects: the number density, auto-angular correlation function, cross-angular correlation function with other full sky surveys, and redshift distribution. We determined differences in galaxy bias and shifts in the redshift distribution between DMAPS and other CMASS samples by combining the angular correlation function and redshift distribution from the clustering- $z$  method. The resulting constraints of  $\Delta b$  show that the galaxy bias of the DMAPS sample is consistent with both CMASS samples within  $1\sigma$ . Furthermore,  $\Delta b$  between DMAPS and full CMASS is comparable to the 2.6 per cent intrinsic difference of CMASS between the SGC and NGC regions.

The resulting DMAPS sample can be used in cosmological analyses in various ways. The most promising application is using DMAPS as a lens sample for galaxy–galaxy lensing. Beyond the sample used in our work, the probabilistic technique used for this work can be easily applied to other image-based and spectroscopic surveys to identify another CMASS-like sample or other specific types of samples. Future surveys such as LSST (LSST Science Collaboration 2009) can be a great application for this novel approach as the survey footprint of LSST occupies the entire southern sky but has only a small overlapping area with spectroscopic surveys such as eBOSS or DESI (DESI Collaboration 2016) that view the northern sky. Producing a spectroscopic galaxy sample with the LSST imaging will enable us to utilize almost the entire sky and yield a wealth of information on the accelerated expansion of the Universe.

## ACKNOWLEDGEMENTS

The figures in this work are produced with plotting routines from matplotlib (Hunter 2007) and ChainConsumer (Hinton 2016). Some of the results in this paper have been derived using the healpy and HEALPix package (Zonca et al. 2019).

The authors thank D. H. Weinberg and C. Chuang for useful conversations in the course of preparing this work.

AC acknowledges support from NASA grant 15-WFIRST15-0008. During the preparation of this paper, CH was supported by the Simons Foundation, NASA, and the U.S. Department of Energy.

Funding for the DES projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians-Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under grant numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2015-71825, ESP2015-66861, FPA2015-68048, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Program (FP7/2007-2013), including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian

Instituto Nacional de Ciéncia e Tecnologia (INCT) e-Universo (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under contract no. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is <http://www.sdss3.org/>.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

This work used resources at the Owens Cluster at the Ohio Supercomputer Center (Ohio Supercomputer Center 1987).

## REFERENCES

Abbott T. M. C. et al., 2018a, *Phys. Rev. D*, 98, 043526  
 Abbott T. M. C. et al., 2018b, *ApJS*, 239, 18  
 Aihara H. et al., 2011, *ApJS*, 193, 29  
 Alam S., Miyatake H., More S., Ho S., Mandelbaum R., 2017, *MNRAS*, 465, 4853  
 Alam S. et al., 2015, *ApJS*, 219, 12  
 Amon A. et al., 2018, *MNRAS*, 479, 3422  
 Annis J. et al., 2014, *ApJ*, 794, 120  
 Baldauf T., Smith R. E., Seljak U., Mandelbaum R., 2010, *Phys. Rev. D*, 81, 063531  
 Bertin E., Arnouts S., 1996, *A&A*, 117, 393  
 Bolton A. S. et al., 2012, *AJ*, 144, 144  
 Bovy J., Hogg D. W., Roweis S. T., 2011a, *Ann. Appl. Stat.*, 5, 1657  
 Bovy J. et al., 2011b, *ApJ*, 729, 141  
 Burke D. L. et al., 2018, *AJ*, 155, 41  
 Carlstrom J. E. et al., 2011, *PASP*, 123, 568  
 Cawthon R. et al., 2018, *MNRAS*, 481, 2427  
 Choi A., Tyson J. A., Morrison C. B., Jee M. J., Schmidt S. J., Margoniner V. E., Wittman D. M., 2012, *ApJ*, 759, 101  
 Choi A. et al., 2016, *MNRAS*, 463, 3737  
 Chuang C.-H. et al., 2017, *MNRAS*, 471, 2370  
 Crocce M. et al., 2016, *MNRAS*, 455, 4301  
 Crocce M. et al., 2019, *MNRAS*, 482, 2807  
 Cuesta A. J. et al., 2016, *MNRAS*, 457, 1770  
 Dark Energy Survey Collaboration, 2005, preprint ([arXiv:astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346))  
 Dark Energy Survey Collaboration, 2016, *MNRAS*, 460, 1270  
 Davis C. et al., 2017, preprint ([arXiv:1710.02517](https://arxiv.org/abs/1710.02517))  
 Davis M., Peebles P. J. E., 1983, *ApJ*, 267, 465  
 Dawson K. S. et al., 2013, *AJ*, 145, 10  
 de Jong J. T. A., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., 2013, *Exp. Astron.*, 35, 25  
 DESI Collaboration, 2016, preprint ([arXiv:1611.00036](https://arxiv.org/abs/1611.00036))  
 de Vaucouleurs G., 1948, *Ann. Astrophys.*, 11, 247  
 Drlica-Wagner A. et al., 2018, *ApJS*, 235, 33  
 Eisenstein D. J. et al., 2001, *AJ*, 122, 2267  
 Eisenstein D. J. et al., 2011, *AJ*, 142, 72  
 Elsner F., Leistedt B., Peiris H. V., 2016, *MNRAS*, 456, 2095  
 Elvin-Poole J. et al., 2018, *Phys. Rev. D*, 98, 042006  
 Flaugher B. et al., 2015, *AJ*, 150, 150  
 Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306  
 Frieman J. A., Turner M. S., Huterer D., 2008, *ARA&A*, 46, 385  
 Gatti M. et al., 2018, *MNRAS*, 477, 1664  
 Gil-Marín H. et al., 2016a, *MNRAS*, 460, 4188  
 Gil-Marín H. et al., 2016b, *MNRAS*, 460, 4210  
 Goto T., Szapudi I., Granett B. R., 2012, *MNRAS*, 422, L77  
 Gunn J. E. et al., 1998, *AJ*, 116, 3040  
 Gunn J. E. et al., 2006, *AJ*, 131, 2332  
 Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759  
 Heymans C. et al., 2012, *MNRAS*, 427, 146  
 Hinton S. R., 2016, *J. Open Source Softw.*, 1, 00045  
 Hirata C. M. et al., 2004, *MNRAS*, 353, 529  
 Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90  
 Huterer D., Shafer D. L., 2018, *Rep. Prog. Phys.*, 81, 016901  
 Jarvis M., 2015, Astrophysics Source Code Library, record ascl:1508.007  
 Johnson A. et al., 2017, *MNRAS*, 465, 4118  
 Jullo E. et al., 2019, *A&A*, 627, A137  
 Kaiser N., 1984, *ApJ*, 284, L9  
 Kitaura F.-S. et al., 2016, *MNRAS*, 456, 4156  
 Kovacs A., Szapudi I., Granett B. R., Frei Z., 2013, *MNRAS*, 431, L28  
 Krause E., Eifler T., 2017, *MNRAS*, 470, 2100  
 Krause E. et al., 2017, preprint ([arXiv:1706.09359](https://arxiv.org/abs/1706.09359))  
 Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64  
 Leistedt B. et al., 2016, *ApJS*, 226, 24  
 Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473  
 Li T. S. et al., 2016, *AJ*, 151, 157  
 LSST Science Collaborations and LSST project, 2009, LSST Science Book, version 2.0, preprint ([arXiv:0912.0201](https://arxiv.org/abs/0912.0201))  
 Mandelbaum R., 2018, *ARA&A*, 56, 393  
 Mandelbaum R., Slosar A., Baldauf T., Seljak U., Hirata C. M., Nakajima R., Reyes R., Smith R. E., 2013, *MNRAS*, 432, 1544  
 Maraston C., Strömbäck G., Thomas D., Wake D. A., Nichol R. C., 2009, *MNRAS*, 394, L107  
 Miyatake H. et al., 2015, *ApJ*, 806, 1  
 Moraes B. et al., 2014, *Rev. Mex. Astron. Astrofis.*, 44, 202  
 More S., Miyatake H., Mandelbaum R., Takada M., Spergel D. N., Brownstein J. R., Schneider D. P., 2015, *ApJ*, 806, 2  
 Morrison C. B., Hildebrandt H., Schmidt S. J., Baldry I. K., Bilicki M., Choi A., Erben T., Schneider P., 2017, *MNRAS*, 467, 3576  
 Newman J. A., 2008, *ApJ*, 684, 88  
 Ohio Supercomputer Center, 1987, Ohio Supercomputer Center, Columbus, Ohio, <http://osc.edu/ark:/19495/f5s1ph73>  
 Omori Y. et al., 2019, *Phys. Rev. D*, 100, 043501  
 Padmanabhan N. et al., 2007, *MNRAS*, 378, 852  
 Park Y. et al., 2016, *Phys. Rev. D*, 94, 063533  
 Pedregosa F. et al., 2012, *Journal of Machine Learning Research*, 12, 2825  
 Pellejero-Ibanez M. et al., 2017, *MNRAS*, 468, 4116  
 Perlmutter S., Aldering G., Goldhaber G., Knop R. A., 1999, *ApJ*, 517, 565  
 Planck Collaboration XV, 2016, *A&A*, 594, A15  
 Rahman M., Mendez A. J., Ménard B., Scranton R., Schmidt S. J., Morrison C. B., Budavári T., 2016, *MNRAS*, 460, 163  
 Reid B. et al., 2016, *MNRAS*, 455, 1553  
 Riess A. G., Filippenko A. V., Challis P., 1998, *AJ*, 116, 1009  
 Ross A. J. et al., 2011, *MNRAS*, 417, 1350

Ross A. J. et al., 2012, *MNRAS*, 424, 564  
 Rozo E. et al., 2016, *MNRAS*, 461, 1431  
 Rykoff E. S. et al., 2014, *ApJ*, 785, 104  
 Salazar-Albornoz S. et al., 2017, *MNRAS*, 468, 2938  
 Schlafly E. F., Finkbeiner D. P., 2011, *ApJ*, 737, 103  
 Schlafly E. F., Finkbeiner D. P., Schlegel D. J., Jurić M., Ivezić Ž., Gibson R. R., Knapp G. R., Weaver B. A., 2010, *ApJ*, 725, 1175  
 Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525  
 Schmidt S. J., Ménard B., Scranton R., Morrison C., McBride C. K., 2013, *MNRAS*, 431, 3307  
 Schwarz G., 1978, *Ann. Stat.*, 6, 461  
 Scottová V., Benoît-Lévy A., Coupon J., Ilbert O., Mellier Y., 2018, *MNRAS*, 474, 3921  
 Seljak U. et al., 2005, *Phys. Rev. D*, 71, 043511  
 Singh S., Alam S., Mandelbaum R., Seljak U., Rodriguez-Torres S., Ho S., 2019, *MNRAS*, 482, 785  
 Singh S., Mandelbaum R., Seljak U., Rodríguez-Torres S., Slosar A., 2018, preprint ([arXiv:1811.06499](https://arxiv.org/abs/1811.06499))  
 van den Bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, *MNRAS*, 430, 725  
 VanderPlas J., Connolly A. J., Ivezić Z., Gray A., 2012, in Proc. Conf. Intelligent Data Understanding (CIDU), IEEE, Boulder, p. 47  
 Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, *Phys. Rep.*, 530, 87  
 Wright E. L. et al., 2010, *AJ*, 140, 1868  
 Yoo J., Seljak U., 2012, *Phys. Rev. D*, 86, 083504  
 York D. G. et al., 2000, *AJ*, 120, 1579  
 Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, *J. Open Source Softw.*, 4, 1298  
 Zuntz J. et al., 2015, *Astron. Comput.*, 12, 45

## APPENDIX A: THE IMPACT OF REDSHIFT TAILS IN BOSS CMASS ON GALAXY BIAS

The BOSS analyses use the CMASS galaxies only within the redshift range ( $0.43 < z < 0.75$ ), by applying the spectroscopic redshift cuts on the CMASS targets selected by photometric cuts in equations (1)–(8) (Chuang et al. 2017; Pellejero-Ibanez et al. 2017). Through the redshift cuts, nearly 10 per cent of sources are discarded from the photometric targets. As the DMass algorithm only utilizes the photometric information of galaxies, the resulting DMass sample includes sources at the low end ( $z < 0.43$ ) or high end ( $z > 0.75$ ), and they cannot be excluded as done in the BOSS CMASS sample. To combine the BOSS measurements with the weak lensing measurements of DMass, the effect of the redshift tails on galaxy clustering should be examined. Here, we test the impact of the redshift tails on the galaxy clustering of BOSS CMASS, specifically on galaxy bias, by computing the correlation function monopole  $\xi_0(s)$  and quadrupole  $\xi_2(s)$ .

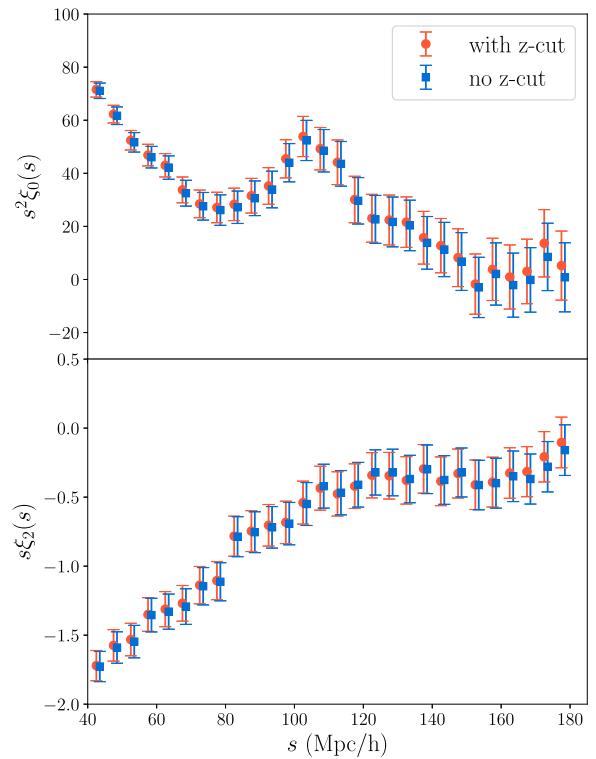
We use the three-dimensional, two-point correlation function estimator given by Landy & Szalay (1993):

$$\xi(s, \mu) = \frac{DD(s, \mu) - 2DR(s, \mu) + RR(s, \mu)}{RR(s, \mu)}, \quad (A1)$$

where  $s$  is the separation of a pair of objects and  $\mu$  is the cosine of the angle between the directions between the line of sight and the line connecting the pair of objects.  $DD$ ,  $DR$ , and  $RR$  represent the normalized galaxy-galaxy, galaxy-random, and random-random pair counts, for a given separation  $s$  and  $\mu$ . The weights described in equation (28) are applied.

To derive the monopole and quadrupole, the two-point correlation function  $\xi(s, \mu)$  is integrated over a spherical shell with radius  $s$ :

$$\xi_l(s) = \frac{1}{N_\mu} \sum_{i=0}^{N_\mu} (2l+1) \xi(s, \mu_i) P_l(\mu_i), \quad (A2)$$



**Figure A1.** Monopole (top) and quadrupole (bottom) correlation function of the CMASS sample before (red) and after (blue) applying redshift cuts at  $40 \text{ h}^{-1} \text{ Mpc} < s < 180 \text{ h}^{-1} \text{ Mpc}$ .

where  $N_\mu$  is the number of  $\mu$  bins and  $P_l(\mu)$  is the Legendre polynomial.

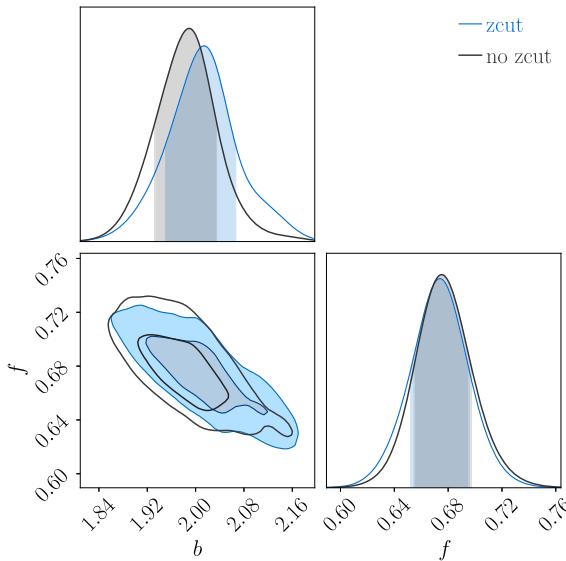
Fig. A1 shows the monopole and quadrupole of CMASS with the redshift cuts (red) and without the redshift cuts (blue). These multipoles are computed in the scale range  $40 \text{ h}^{-1} \text{ Mpc} < s < 180 \text{ h}^{-1} \text{ Mpc}$  with the bin size  $5 \text{ h}^{-1} \text{ Mpc}$ , the same scales and bin size adopted in the previous BOSS analyses (Chuang et al. 2017; Pellejero-Ibanez et al. 2017). Error bars are computed from the MultiDark-PATCHY BOSS DR12 mock catalogues (Kitaura et al. 2016). The amplitudes of the multipoles are overall higher on large scales  $s > 120 \text{ h}^{-1} \text{ Mpc}$  with the redshift tails. This may indicate systematics associated with sources at high redshift, but their impact should be negligible as the offset between the two correlation functions is way smaller than the statistical errors.

From the measured monopole and quadrupole of the correlation function, we constrain galaxy bias  $b$  and the structure growth rate  $f \equiv d \ln D/d \ln a$  at a median redshift  $z = 0.59$  for each case and compare the results. In linear theory, the relative amplitudes of multipoles depend only on the combination of  $b$  and  $f$  as follows:

$$\xi_0(s) = \left( b^2 + \frac{2}{3}bf + \frac{1}{5}f^2 \right) \xi_m(s), \quad (A3)$$

$$\xi_2(s) = \left( \frac{4}{3}bf + \frac{4}{7}f^2 \right) \xi_m(s). \quad (A4)$$

As the linear theory is applicable on the scales we are using, we simply adopt the above equations and expect the potential difference due to the redshift cuts to appear on the constraints of  $b$  and  $f$ . The matter correlation function  $\xi_m(s)$  at  $z = 0.59$  is estimated by CAMB (Lewis, Challinor & Lasenby 2000) with the fiducial cosmology.



**Figure A2.** Comparison of galaxy bias constraints from the CMASS clustering before (blue) and after (red) applying  $z$ -cuts.

We perform a Markov chain Monte Carlo likelihood analysis using `emcee` (Foreman-Mackey et al. 2013). The two parameters  $b$  and  $f$  are varied in the range of  $b = [0.5, 3]$  and  $f = [0.2, 1.0]$ . The likelihood is taken from  $\chi^2$  defined as

$$\chi^2 = \sum_{i,j}^{N_X} [X_{\text{obs},i} - X_{\text{th},i}] \mathbf{C}_{ij}^{-1} [X_{\text{obs},j} - X_{\text{th},j}], \quad (\text{A5})$$

where  $N_X$  is the number of points in the data vector,  $X_{\text{th}}$  is the vector from the theoretical model, and  $X_{\text{obs}}$  is the vector from the measurement. The data points from the multipoles are combined to form a vector  $X$  as

$$X = \{\xi_0(s_0), \xi_0(s_1), \dots, \xi_0(s_N); \xi_2(s_0), \xi_2(s_1), \dots, \xi_2(s_N)\}, \quad (\text{A6})$$

where  $N$  is the number of bins in each multipole.

Fig. A2 shows the constraints of  $b$  and  $f$  at  $z = 0.59$  on the two-dimensional plane. The resulting numbers are  $b = 2.02^{+0.04}_{-0.07}$  and  $f = 0.67 \pm 0.02$  with the redshift cut, and  $b = 2.00^{+0.04}_{-0.06}$  and  $f = 0.68 \pm 0.02$  without the redshift cut. From these results, we do not find a big discrepancy between the two cases.

The negligible impact of high-redshift galaxies has also been studied by the original BOSS analyses. Cuesta et al. (2016) and Gil-Marín et al. (2016a, 2016b) use CMASS within  $0.43 < z < 0.7$  with the effective redshift  $z = 0.57$ . Alam et al. (2015) compared the measurements from these analyses with the ones including higher redshift galaxies up to  $z < 0.75$ , with the effective redshift  $z = 0.59$  (Chuang et al. 2017; Pellejero-Ibanez et al. 2017). To compare the measurements directly at the same redshift, Alam et al. (2015) extrapolated the measurements of Chuang et al. (2017) and Pellejero-Ibanez et al. (2017) at  $z = 0.57$ . A summary of their work can be found in figs 13 and 15 in Alam et al. (2015). All of the BOSS measurements compared in this work show consistency within the  $1\sigma$  level or better.

## APPENDIX B: DIFFERENCES BETWEEN CMASS SGC AND NGC

The BOSS CMASS target selection function is applied differently in the SGC and NGC due to the colour offsets in the DR8 pho-

**Table B1.**  $\chi^2/\text{dof}$  of three probes calculated between two different samples.  $\chi^2_{\text{sys}}/\text{dof}$  in the third column are calculated with systematic weights of DMASS obtained in Section 4. Values in the parentheses are corresponding PTE values. SGC and FULL in bold stand for CMASS in SGC and full CMASS.

Probe	$\chi^2/\text{dof}$ (PTE)	$\chi^2_{\text{sys}}/\text{dof}$ (PTE)
<b>SGC – DMASS</b>		
$w^{\delta_g \delta_g}$	4.94/10 (90%)	2.58/10 (99%)
$w^{\delta_g \delta_{\text{WISE}}}$	9.04/10 (53%)	9.70/10 (47%)
$w^{\delta_g \text{CMB}}$	8.92/7 (26%)	13.25/7 (6%)
<b>FULL – DMASS</b>		
$w^{\delta_g \delta_g}$	10.61/10 (39%)	8.60/10 (57%)
$w^{\delta_g \delta_{\text{WISE}}}$	12.12/10 (28%)	11.42/10 (33%)
$w^{\delta_g \text{CMB}}$	7.13/7 (42%)	7.68/7 (36%)
<b>NGC – SGC</b>		
$w^{\delta_g \delta_g}$	14.53/10 (15%)	–
$w^{\delta_g \delta_{\text{WISE}}}$	11.76/10 (30%)	–
$w^{\delta_g \text{CMB}}$	23.95/7 (0.1%)	–

tometry between two regions. Schlafly et al. (2010) and Schlafly & Finkbeiner (2011) have estimated the level of colour offsets and found that these differences are due to either calibration errors or errors in the galactic extinction corrections (or combination of both). This offset shifts the values of  $d_{\perp}$  (the combination of  $g-r$  and  $r-i$  colours) by 0.0064 magnitudes between the North and South caps, resulting in a few per cent difference in the number density and the amplitude of the angular correlation function. Ross et al. (2011, 2012) have shown that the difference in the number density and the angular correlation function can be mitigated by applying the new cut with the offset,  $d_{\perp} > 0.5564$ , in the SGC. However, the final analyses of BOSS-III were completed with the same  $d_{\perp}$  cut in both regions. Therefore, we do not consider the colour offset either.

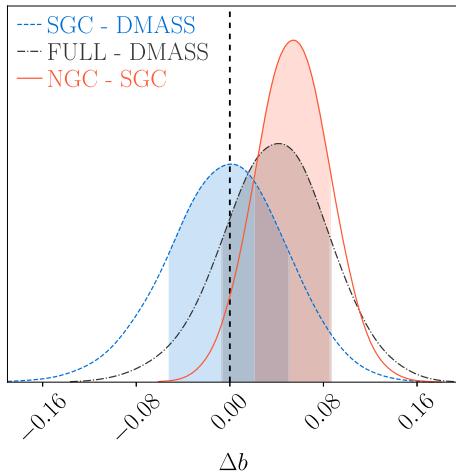
The resulting DMASS is designed to be closer to CMASS in the SGC than NGC since the extreme deconvolution model is trained with the  $d_{\perp}$  colour in the SGC. Therefore, we report the measurements of the angular correlation functions and galaxy biases of CMASS NGC and SGC here in order to show that discrepancy between DMASS and full CMASS originates from the intrinsic difference within CMASS.

Table B1 shows the values of  $\chi^2/\text{dof}$  and its corresponding PTE of all three probes computed in Section 5. The last column includes  $\chi^2/\text{dof}$  between CMASS in NGC and CMASS in SGC. For all three probes,  $\chi^2/\text{dof}$  of ‘NGC – SGC’ is either larger than any of the other two cases or comparable to the case of ‘FULL – DMASS’.

Galaxy biases are derived from the model of the angular correlation function given as

$$w^{\delta_g \delta_g}(\theta) = \int dz b(z) n(z) \int dz' b(z') n(z') \xi_m(R, z, z'), \quad (\text{B1})$$

where  $b$  is galaxy bias,  $n(z)$  is normalized spectroscopic redshift distribution,  $\xi_m$  is matter clustering, and  $R$  is the co-moving distance defined as  $R = (1 + z) D_A(z)\theta$ . With the covariance matrices calculated from the QPM mock catalogues in Section 4, we have estimated best-fitting values of galaxy bias that minimize  $\chi^2$  defined in equation (29). The data vector  $\Delta \mathbf{d}$  in the equation is defined as the residual of the measurement and theoretical prediction given as  $\Delta \mathbf{d} = \mathbf{d}_{\text{true}} - \mathbf{d}$ . The vector  $\mathbf{d}$  corresponds to the measurements of the angular correlation function of CMASS SGC and CMASS NGC computed in Section 5, and  $\mathbf{d}_{\text{true}}$  is the theoretical data vector from equation (B1).



**Figure B1.** Difference in galaxy biases constrained by the angular correlation function of CMASS SGC and CMASS NGC (red).  $\Delta b_{\text{NGC}-\text{SGC}} = 0.056^{+0.031}_{-0.033}$ . The blue dashed and black dot-dashed histograms display  $\Delta b_{\text{SGC}-\text{DMASS}}$  and  $\Delta b_{\text{FULL}-\text{DMASS}}$  obtained in Section 5, respectively. The redshift bin bias  $\Delta z$  of DMASS is marginalized for the latter two cases.

The final constraints of galaxy biases are  $b_{\text{SGC}} = 2.035 \pm 0.026$  and  $b_{\text{NGC}} = 2.088 \pm 0.017$  from CMASS SGC and CMASS NGC, respectively. The derived galaxy bias difference between NGC and SGC is  $\Delta b_{\text{NGC}-\text{SGC}} = 0.056^{+0.031}_{-0.033}$ . Fig. B1 shows the constraint of  $\Delta b_{\text{NGC}-\text{SGC}}$  (red-solid) plotted with  $\Delta b_{\text{SGC}-\text{DMASS}}$  (blue dashed) and  $\Delta b_{\text{FULL}-\text{DMASS}}$  (black dot-dashed) obtained in Section 5. The redshift bin bias  $\Delta z$  of DMASS is marginalized for the latter two cases. The resulting  $\Delta b_{\text{NGC}-\text{SGC}}$  implies that the colour offset between the SGC and NGC naturally yields  $\sim 2.6$  per cent of the difference in galaxy bias, and the constraints of  $\Delta b$  between DMASS and CMASS are safely within this intrinsic difference.

<sup>1</sup>Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA

<sup>2</sup>Department of Physics, The Ohio State University, Columbus, OH 43210, USA

<sup>3</sup>Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

<sup>4</sup>Department of Physics, Duke University, Durham, NC 27708, USA

<sup>5</sup>Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, PO Box 2450, Stanford, CA 94305, USA

<sup>6</sup>Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA

<sup>7</sup>Physics Department, University of Wisconsin-Madison, 2320 Chamberlin Hall, 1150 University Avenue, Madison, WI 53706-1390, USA

<sup>8</sup>Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA

<sup>9</sup>Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain

<sup>10</sup>CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France

<sup>11</sup>Institut d'Astrophysique de Paris, Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, F-75014, Paris, France

<sup>12</sup>Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

<sup>13</sup>Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), E-28040 Madrid, Spain

<sup>14</sup>Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil

<sup>15</sup>Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA

<sup>16</sup>National Center for Supercomputing Applications, 1205 West Clark St, Urbana, IL 61801, USA

<sup>17</sup>Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain

<sup>18</sup>Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil

<sup>19</sup>Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India

<sup>20</sup>Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain

<sup>21</sup>Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain

<sup>22</sup>Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA

<sup>23</sup>Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

<sup>24</sup>Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA

<sup>25</sup>SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

<sup>26</sup>Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA

<sup>27</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

<sup>28</sup>Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia

<sup>29</sup>Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP 05314-970, Brazil

<sup>30</sup>George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA

<sup>31</sup>Department of Astronomy, The Ohio State University, Columbus, OH 43210, USA

<sup>32</sup>Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA

<sup>33</sup>Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain

<sup>34</sup>School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK

<sup>35</sup>Physics Department, Brandeis University, 415 South Street, Waltham, MA 02453, USA

<sup>36</sup>Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, Campinas, SP 13083-859, Brazil

<sup>37</sup>Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

<sup>38</sup>Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK

<sup>39</sup>Excellence Cluster Origins, Boltzmannstr 2, D-85748 Garching, Germany

<sup>40</sup>Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany

<sup>41</sup>Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr 1, D-81679 München, Germany

<sup>42</sup>Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK