

# Imaging systematics and clustering of DESI main targets

Ellie Kitanidis,<sup>1★</sup> Martin White<sup>1</sup>,<sup>1</sup> Yu Feng,<sup>1</sup> David Schlegel,<sup>2</sup> Julien Guy,<sup>3</sup>  
Arjun Dey,<sup>4</sup> Martin Landriau,<sup>2</sup> David Brooks,<sup>5</sup> Michael Levi,<sup>2</sup> John Moustakas,<sup>6</sup>  
Francisco Prada,<sup>7</sup> Gregory Tarle<sup>8</sup> and Benjamin Alan Weaver<sup>4</sup>

<sup>1</sup>Department of Physics, University of California, Berkeley, 366 LeConte Hall, Berkeley, CA 94720, USA

<sup>2</sup>Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 93720, USA

<sup>3</sup>Sorbonne Universités, UPMC Université Paris 06, Université Paris-Diderot, CNRS-IN2P3 LPNHE 4 Place Jussieu, F-75252, Paris Cedex 05, France

<sup>4</sup>NSF's National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Avenue, Tucson, AZ 85719, USA

<sup>5</sup>Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK

<sup>6</sup>Department of Physics and Astronomy, Siena College, 515 Loudon Road, Loudonville, NY 12211, USA

<sup>7</sup>Instituto de Astrofísica de Andalucía, Glorieta de la Astronomía, s/n, E-18008 Granada, Spain

<sup>8</sup>Physics Department, University of Michigan Ann Arbor, MI 48109, USA

Accepted 2020 June 3. Received 2020 June 3; in original form 2019 November 13

## ABSTRACT

We evaluate the impact of imaging systematics on the clustering of luminous red galaxies (LRG), emission-line galaxies (ELG), and quasars (QSO) targeted for the upcoming Dark Energy Spectroscopic Instrument (DESI) survey. Using Data Release 7 of the DECam Legacy Survey, we study the effects of astrophysical foregrounds, stellar contamination, differences between north galactic cap and south galactic cap measurements, and variations in imaging depth, stellar density, galactic extinction, seeing, airmass, sky brightness, and exposure time before presenting survey masks and weights to mitigate these effects. With our sanitized samples in hand, we conduct a preliminary analysis of the clustering amplitude and evolution of the DESI main targets. From measurements of the angular correlation functions, we determine power law fits  $r_0 = 7.78 \pm 0.26 h^{-1} \text{Mpc}$ ,  $\gamma = 1.98 \pm 0.02$  for LRGs and  $r_0 = 5.45 \pm 0.1 h^{-1} \text{Mpc}$ ,  $\gamma = 1.54 \pm 0.01$  for ELGs. Additionally, from the angular power spectra, we measure the linear biases and model the scale-dependent biases in the weakly non-linear regime. Both sets of clustering measurements show good agreement with survey requirements for LRGs and ELGs, attesting that these samples will enable DESI to achieve precise cosmological constraints. We also present clustering as a function of magnitude, use cross-correlations with external spectroscopy to infer  $dN/dz$  and measure clustering as a function of luminosity, and probe higher order clustering statistics through counts-in-cells moments.

**Key words:** large-scale structure of Universe.

## 1 INTRODUCTION

The Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration 2016) is a ground-based dark energy experiment whose mission is to produce the largest three-dimensional (3D) map of the universe to date. This map will enable unprecedented constraints on dark energy (for a comprehensive review, refer to Weinberg et al. 2013 or Amendola et al. 2013) by charting the expansion history of the universe through studies of baryon acoustic oscillations (BAO; see Eisenstein 2005 or Bassett & Hlozek 2010 reviews) and constraining the growth of structure through redshift-space distortion measurements (RSD; see e.g. Ruggeri et al. 2019 for

a recent study). In addition, it will provide a means to precisely measure the sum of neutrino masses (Font-Ribera et al. 2014) and to investigate theories of inflation (Gariazzo, Lopez-Honorez & Mena 2015; Tellarini et al. 2016) and modified gravity (Jain & Khoury 2010; Joyce et al. 2015; Casas et al. 2017; Amendola et al. 2019). Installed on the Mayall 4-m telescope at Kitt Peak, DESI is a stage IV dark energy project<sup>1</sup> consisting of a highly multiplexed fibre-fed spectrograph that can measure as many as 5000 spectra in parallel using robot fibre positioners. DESI will obtain spectra for four main target classes selected from imaging, including approximately 6 million luminous red galaxies (LRG) up to  $z = 1.0$ , 17 million [OII] emission-line galaxies (ELG) up to

\* E-mail: [ellick@berkeley.edu](mailto:ellick@berkeley.edu)

<sup>1</sup> As defined in the Dark Energy Task Force report (Albrecht et al. 2006).



$z = 1.6$ , and 2.5 million quasars (QSO). QSOs with  $z < 2.1$  will serve as tracers of the underlying dark matter distribution, while a high redshift sample of QSOs ( $2.1 < z < 3.5$ ) will be used for their Lyman- $\alpha$  absorption features to probe the distribution of neutral hydrogen in the intergalactic medium. During ‘bright time’, when the position of the moon above the horizon impacts the observation of faint, high redshift targets, DESI will conduct the Bright Galaxy Survey, observing over 10 million galaxies up to  $z \sim 0.4$ , and also the Milky Way Survey of local stars.

DESI is currently in its commissioning phase, with survey validation scheduled for the spring of 2020. At this stage of the project, it is vital to ensure that the targets selected from imaging will satisfy the science requirements of the collaboration, which demand meticulous control over all possible systematics. Since it is not obvious how systematics in the distribution of targets selected from imaging will translate to systematics in the 3D clustering of the spectroscopic samples, it is prudent to identify and mitigate them to the greatest extent possible. Furthermore, while acquiring spectra will allow us to remove any low redshift contaminants and deproject any purely angular systematics, the targeting efficiency of the survey will be adversely affected. In this paper, we analyse the impact of potential systematics from imaging and target selection on the observed clustering of the main DESI samples and develop methods to ameliorate those effects. Using the resulting value-added large-scale structure catalogues, we begin to characterize the properties of these tracer samples, both as a first step for analysis in cosmological studies and to aid in generating accurate mock catalogues.

The paper is organized as follows: Section 2 outlines our techniques for measuring and modeling clustering. Section 3 describes the imaging data and explains how we build our large-scale structure catalogues. In Section 4, we implement and test some preliminary survey masks; Section 4.1 deals with imaging completeness and its sensitivity to depth variations and colour cuts, while Section 4.2 covers masking around bright foregrounds. In Section 5, we explore spatial variations in the clustering of the targets and identify additional problematic or anomalous regions. In Section 6, we investigate the effects of varying survey properties and instrument characteristics on the densities of targets, obtaining photometric weights for the most dominant systematics, with Section 6.3 focusing on the impact of stellar contamination in the QSO sample. Section 7 presents our preliminary characterization of the angular distribution of the tracer samples, including mean surface densities, angular correlation functions, angular power spectra, linear biases, counts-in-cells moments, and clustering as a function of magnitude. In Section 8, we use external spectroscopic catalogues to measure real-space projected cross-correlations, clustering  $dN/dz$ , and clustering as a function of luminosity. Section 9 summarizes our results and conclusions.

Throughout, we will work in co-moving coordinates and assume a flat  $\Lambda$ CDM cosmology with  $h = 0.676$ ,  $\Omega_m h^2 = 0.142$ ,  $\Omega_b h^2 = 0.022$ ,  $n_s = 0.962$ , and  $\sigma_8 = 0.848$  (the default parameters in CLASS, see e.g. Blas, Lesgourgues & Tram 2011). Additionally, all magnitudes are quoted as  $AB$  magnitudes, unless otherwise specified.

## 2 CLUSTERING MEASUREMENT AND THEORY

To characterize the properties of DESI main samples and understand how systematics impact their observed distributions, we must be able to accurately quantify clustering as well as compare our results to theoretical predictions and survey expectations. To convert

between angular clustering measurements and 3D clustering theory in Section 7, we assume the fiducial redshift distributions in the DESI Science Final Design Report (DESI Collaboration 2016, henceforth FDR), calculated from cross-matching and photometric methods (private communications: Rongpu Zhou, Anand Raichoor, and Nathalie Palanque-Delabrouille for the DR7, LRG, ELG, and QSO  $dN/dz$ , respectively). In Section 8, we use cross-correlations with external spectroscopy to obtain clustering  $dN/dz$ .

### 2.1 Angular correlation functions

One of the simplest and most powerful measurements of clustering is the two-point correlation function  $\xi(r)$ , which measures the excess probability, compared to a random Poisson distribution, that a pair of objects lie at a given separation (see e.g. Peebles 1980; Peacock 1999). The two-point correlation function and its Fourier transform, the power spectrum, fully characterize a Gaussian random field. For samples that lack redshifts, the two-dimensional (2D) angular correlation function,  $w(\theta)$ , representing the probability in excess of random of finding two objects separated by a given angle, may be used instead.

#### 2.1.1 Pair-count estimators

We measure  $w(\theta)$  of the targets with direct pair-count estimators, namely the Landy–Szalay estimator (Landy & Szalay 1993),<sup>2</sup>

$$\hat{w}_{LS}(\theta) = \frac{D_1 D_2 - D_1 R_2 - D_2 R_1 + R_1 R_2}{R_1 R_2}, \quad (1)$$

where  $DD$ ,  $DR$ , and  $RR$ , respectively, refer to counts of data–data, data–random, and random–random pairs at average separation  $\theta$  (within annular bins  $\theta \pm \delta\theta$ ). For auto-correlations, this simplifies to  $\hat{w}_{LS}(\theta) = (DD - 2DR + RR)/RR$ . When cross-correlating with external data sets that do not have a corresponding random catalogue readily available, we instead use the Davis–Peebles estimator (Davis & Peebles 1983),

$$\hat{w}_{DP}(\theta) = \frac{D_1 D_2 - D_2 R_1}{D_2 R_1}, \quad (2)$$

which has a slightly larger variance (see Landy & Szalay 1993 for a comparison of pair-count estimators) but requires only one set of randoms. We count pairs within 16 logarithmically spaced angular bins between  $\theta = 0.001^\circ$  and  $\theta = 1^\circ$ .

#### 2.1.2 Limber approximation

We wish to generate a prediction for the observed angular clustering of objects in the sky,  $w(\theta)$ , given an assumed model for the full 3D clustering,  $\xi(r)$ , and redshift distribution,  $dN/dz$ . An approximation first introduced by Limber (1953) and Rubin (1954) is frequently employed for this purpose. Briefly, the Limber approximation assumes redshift distributions that do not vary appreciably over the coherence length of the structures defined by  $\xi(r)$ . Though not a requirement, it is often further assumed that the sky may be treated as flat. In this section, we state the general result, as well as the simplified expression for the case of a power law  $\xi(r)$ . A more detailed derivation can be found in e.g. Simon (2007) or Loverde & Afshordi (2008).

<sup>2</sup>Normalization factors are not included here.



Written in centre-of-mass and relative coordinates,  $\bar{r} = (r_1 + r_2)/2$  and  $\Delta r = r_2 - r_1$ , the Limber approximation straightforwardly relates the angular and spatial correlations between two samples,

$$w_{1,2}(\theta) = \int_0^\infty d\bar{r} f_1(\bar{r}) f_2(\bar{r}) \int_{-\infty}^\infty d\Delta r \xi_{1,2}(R, \bar{r}), \quad (3)$$

where  $f_1$  and  $f_2$  are the normalized radial distributions, and  $R = \sqrt{\bar{r}^2 \theta^2 + \Delta r^2}$ . If we further presume a power law for the correlation function, with correlation length  $r_0$ ,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \quad (4)$$

then we can evaluate the  $\Delta r$  integral directly, and Limber's approximation gives a particularly simple result,

$$w_{1,2}(\theta) = \theta^{1-\gamma} r_0^\gamma \sqrt{\pi} \frac{\Gamma(\gamma/2 - 1/2)}{\Gamma(\gamma/2)} \int_0^\infty d\bar{r} f_1(\bar{r}) f_2(\bar{r}) \bar{r}^{1-\gamma} \equiv A_w \theta^{1-\gamma}. \quad (5)$$

Using this equation with tabulated  $dN/dz$ 's, we can compare observation to theory by determining the clustering length  $r_0$  and slope  $\gamma$  that best fit the observed  $w(\theta)$ .

## 2.2 Projected real-space cross-correlations

We can extract additional clustering information by cross-correlating the samples with other samples of known redshift. We begin by deriving a relation between the angular correlation function  $w(\theta)$  and the projected real-space correlation function  $w_p(r_p)$ , the latter being defined as the integral of the 2D spatial correlation function  $\xi(r_\pi, r_p)$  over the line of sight  $r_\pi$  (Davis & Peebles 1983).

Starting with the simple case in which the spectroscopic sample lies at  $\chi = \chi_0$ , the flat-sky approximation yields

$$w(\theta) = \int d\chi f(\chi) \xi\left(\sqrt{\chi_0^2 \theta^2 + (\chi - \chi_0)^2}\right), \quad (6)$$

where  $f(\chi)$  is the normalized radial distribution of the photometric sample, and the second integral over the radial distribution of the spectroscopic sample, a delta function, has been performed. Applying the Limber approximation (Section 2.1.2), this simplifies to

$$w(\theta) \simeq f(\chi_0) \int dr_\pi \xi(r_p, r_\pi) = f(\chi_0) w_p(r_p), \quad (7)$$

where  $w_p(r_p)$  is a real-space measurement, since redshift space distortions affect only  $r_\pi$ .

Generalizing to a narrow spectroscopic redshift slice, such that the clustering can still be treated as constant over the slice, we adopt the approach of Padmanabhan et al. (2009): for each pair in a given bin, we assume that the photometric object lies at the same redshift as the spectroscopic object it is being correlated with, allowing us to re-bin the pair counts in transverse separation,  $w_\theta(r_p)$ , such that equation (7) becomes

$$w_\theta(r_p) = \langle f(\chi) \rangle w_p(r_p), \quad (8)$$

where  $\langle f(\chi) \rangle$  is averaged over the spectroscopic redshift bin in question and  $w_\theta(r_p)$  is the angular correlation function but binned in physical distance instead of angle using the redshifts of the spectroscopic objects for conversion.

To compare with theory, we also note the form  $w_p(r_p)$  takes for a power-law correlation function model  $\xi(r) = (\frac{r}{r_0})^{-\gamma}$ :

$$w_p(r_p) = r_p^{1-\gamma} r_0^\gamma \sqrt{\pi} \frac{\Gamma(\gamma/2 - 1/2)}{\Gamma(\gamma/2)}. \quad (9)$$

## 2.3 Bootstrap errors

As an internal error estimate, we use the bootstrap technique of Efron (1979), splitting the sample into multiple subsamples and then randomly selecting with replacement to obtain many different realizations of the underlying distribution. Since resampling on individual objects has been shown to lead to unreliable errors (Mo, Jing & Boerner 1992; Fisher et al. 1994), with variance underestimated in underdense regions and overestimated in overdense regions, we instead partition the sky into equal area pixels and resample these.

The choice of the bootstrap over similar methods such as the jackknife is motivated by comparative studies (e.g. Norberg et al. 2009) suggesting that, though the bootstrap tends to overestimate variance on all scales, it recovers the principal eigenvectors of the true covariance matrix in an unbiased fashion. As such, we caution that our bootstrap error bars are likely overestimated in some cases.

In detail, we use the HEALPix package<sup>3</sup> (Górski et al. 2005) with  $N_{\text{SIDE}} = 4$  to divide the surface of a sphere into 192 equal area pixels of approximate size  $\sim 215$  square degrees, then throw away any pixels that do not overlap with the footprint, leaving 83 pixels. We then randomly select pixels with replacement until the number of randoms in each bootstrap realization is similar to the number of randoms in the footprint. We use 500 bootstrap realizations to obtain an estimate of the variance. Our results are robust to variations in the  $N_{\text{SIDE}}$  resolution and the number of bootstrap realizations.

## 2.4 Angular power spectra

The angular power spectrum,  $C_\ell$ , is another powerful tool for quantifying clustering, allowing us to study the Fourier modes of the angular distribution of galaxies. It complements the statistical information derived from the angular correlation function to which it is related via a Legendre transform:

$$w(\theta) = \sum_\ell \frac{2\ell + 1}{4\pi} P_\ell(\cos \theta) C_\ell. \quad (10)$$

Large-scale systematics are more clearly visible in the power spectrum than in the correlation function, which potentially has long wavelength modes affecting all angular scales. On the other hand, the correlation function is more sensitive to small scales, where non-linear evolution dominates and introduces correlations between different  $C_\ell$ 's at large  $\ell$ . Additionally, it is faster to compute small-scale clustering in configuration space and large-scale clustering in Fourier space. Thus, we focus our analysis of the angular power spectrum on large scales,  $\ell \leq 500$ , corresponding to angular scales greater than  $\theta \sim 180^\circ/\ell \approx 0.4^\circ$ , or spatial scales greater than a few  $h^{-1}$  Mpc at the characteristic survey depth of  $1 h^{-1}$  Gpc.

### 2.4.1 Measurement

We use HEALPix with  $N_{\text{SIDE}} = 512$  and estimate the angular power spectrum from harmonic analysis of the pixelized map of density

<sup>3</sup><http://healpix.sf.net>



contrast  $\delta_g = n/\bar{n} - 1$ , where  $n$  is the number of galaxies in a given pixel and  $\bar{n}$  is the average density over the entire masked sample multiplied by the given pixel's effective area. We mask out pixels whose effective area is less than 25 per cent of its full area, such that only pixels that are fully (or mostly) inside the survey geometry are considered.<sup>4</sup> Using *anafast*, we obtain the estimated angular power spectrum  $C_\ell$ , which is the sum of the signal and the shot noise. To first-order correct the effects of partial sky, we divide by a factor of  $f_{\text{sky}}$ , the fraction of sky covered by the masked footprint; full deconvolution of the mask is deferred to a future work. We find that the angular power spectrum of the mask has its power concentrated in the large-scale modes, with the mask dropping to half power at  $\ell \sim 10$  and falling below 10 per cent power beyond  $\ell \sim 20$ . We also divide out the pixel window function. The variance of the estimator can be modelled analytically as

$$\sigma_\ell^2 = \frac{1}{f_{\text{sky}}} \frac{2C_\ell^2}{2\ell + 1}. \quad (11)$$

On a small section of sky  $\phi$ , multipole resolution is limited by  $\Delta\ell \approx 180^\circ/\phi$ , with the  $\ell_{\text{min}}$  mode constrained to be the wavelength that fits into the angular patch (Peebles 1980). We bin  $C_\ell$  using 10 linearly spaced bins from  $\ell_{\text{min}} = 30$  to  $\ell_{\text{max}} = 500$  and take the weighted arithmetic mean and variance for each bin.

$$\bar{C}_{\text{bin}} = \frac{\sum_{\ell \text{ in bin}} \frac{C_\ell}{\sigma_\ell^2}}{\sum_{\ell \text{ in bin}} \frac{1}{\sigma_\ell^2}}, \quad \frac{1}{\sigma_{\text{bin}}^2} = \sum_{\ell \text{ in bin}} \frac{1}{\sigma_\ell^2}. \quad (12)$$

#### 2.4.2 Theory

In the first-order correction to the Limber approximation (Loverde & Afshordi 2008), the multipole expansion of the galaxy angular power spectrum is given by

$$C_\ell = \int d\chi f(\chi)^2 \frac{1}{\chi^2} P_g(k = (\ell + 1/2)/\chi, z) \\ = \int d\chi f(\chi)^2 \frac{b(z)^2}{\chi^2} P_m(k = (\ell + 1/2)/\chi, z), \quad (13)$$

where  $f(\chi) \equiv dN/d\chi = dN/dz \frac{H(z)}{c}$  is the normalized radial distribution,  $P_m$  is the linear dark matter power spectrum, and  $b(z)$  is the large-scale bias, which we assume takes the form<sup>5</sup>  $b(z) = b_0/D(z)$  as per the DESI FDR. In Section 7.3, we fit the linear bias  $b_0$  and also explore the scale dependence of the bias in the weakly non-linear regime.

### 2.5 Clustering $dN/dz$

#### 2.5.1 Overview

The idea of using cross-correlations to infer redshift information about objects in the night sky has been circulating for decades

(e.g. Seldner & Peebles 1979; Phillipps & Shanks 1987; Landy, Szalay & Koo 1996; Ho et al. 2008) but has received renewed attention in the context of modern astronomical surveys, which are probing deeper than ever and imaging far more objects than can feasibly be targeted for spectroscopic observation. In recent years, a number of  $dN/dz$  cross-correlation estimators have been proposed and studied (Newman 2008; Matthews & Newman 2010; Schulz 2010; Matthews & Newman 2012; McQuinn & White 2013; Ménard et al. 2013) and applied to real or simulated data (Schmidt et al. 2013; Scottez et al. 2016; Hildebrandt et al. 2017; Chiang, Ménard & Schiminovich 2018; Davis et al. 2018; Gatti et al. 2018; Scottez et al. 2018; Krolewski et al. 2020). Following Ménard et al. (2013), who present a simple and practical method for estimating the clustering  $dN/dz$  of a sample, we probe the redshift distributions of objects targeted for DESI in Section 8.1. Unlike other methods, the Ménard method takes advantage of small-scale clustering information and reduces the impact of systematics by sidestepping autocorrelation functions. We briefly describe the formalism of Ménard method and the details of our implementation below.

#### 2.5.2 Method

Consider two populations. Let the reference (spectroscopic) sample have a redshift distribution  $dN_r/dz$ , a mean surface density  $\bar{n}_r$ , and a total number of objects  $N_r$ . The corresponding properties for the unknown (photometric) sample will be labeled  $dN_u/dz$ ,  $\bar{n}_u$ , and  $N_u$ , respectively. The angular cross-correlation between the reference sample and the unknown sample is estimated by

$$w_{\text{ur}}(\theta, z) = \frac{\langle n_u(\theta, z) \rangle}{\bar{n}_u} - 1, \quad (14)$$

where  $\langle n_u(\theta, z) \rangle$  is the mean surface density of objects from the unknown sample lying within an angular distance  $\theta$  of objects in the reference sample at redshift  $z$ . To calculate this, we bin the reference sample into narrow redshift bins  $\delta z_i$ . Then, for each  $\delta z_i$ , we estimate  $w_{\text{ur}}(\theta, z_i)$  by pair counting with the Davis–Peebles estimator equation (2).

In practice, we actually integrate over an annulus around each reference object, from  $\theta_{\text{min}}$  to  $\theta_{\text{max}}$ , because the sensitivity of the estimator is improved by encoding information from many clustering scales (Ménard et al. 2013). In order to maximize the SNR, we weight each point by  $\theta^{-1}$ , which gives equal amounts of clustering information per logarithmic scale ( $d\theta/\theta = d\log \theta$ ).

$$\bar{w}_{\text{ur}}(z) = \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} d\theta \frac{w_{\text{ur}}(\theta, z)}{\theta}. \quad (15)$$

To avoid excess signal from cross-correlations between duplicate objects that appear in both catalogues, it is necessary to impose a minimum radius,  $\theta_{\text{min}}$ , which is at least as large as the astrometric uncertainties in the survey. Furthermore, as we go to smaller scales  $< 1$  Mpc, clustering becomes increasingly non-linear and bias becomes increasingly scale-dependent, so the assumptions underpinning the estimator break down, potentially affecting the accuracy of the result. Finally, we note that as the scale falls below the mean separation of spectroscopic objects, cross-correlations between redshift bins become more significant. Meanwhile, at larger scales, the advantage of a linear bias<sup>6</sup> must be balanced

<sup>4</sup>Since effective area is calculated using a set of uniformly distributed random points, there is some natural Poisson variance, hence why we do not use a more restrictive threshold.

<sup>5</sup> $D(z)$  is the linear growth function with normalization  $D(z=0) = 1$ . Thus, the approximation  $b \propto D^{-1}(z)$  assumes that the clustering is constant, since the evolution of  $P(k, z)$  is cancelled by the evolution of  $b^2(z)$ .

<sup>6</sup>The bias measured by these angular cross-correlations is dominated by scales of hundreds of kpc to a few Mpc and thus should be distinguished from the large-scale ( $> 10$  Mpc) bias, which may evolve differently.



against the cost of degraded signal-to-noise since the clustering signal decreases with radius and the noise due to systematics increases as more background sources are included in the counts. Thus, for samples that have little to no bias evolution, small scales are ideal for recovering  $dN/dz$ ; for samples with some bias evolution, intermediate scales are recommended; and for samples with extensive bias evolution, restricting to large scales may be the best strategy.

## 2.6 Counts-in-cells

To aid in simulations, we wish to also provide some measurement of the higher order clustering statistics of the DESI samples. In particular, for regions of high density, spectroscopic incompleteness due to the physical limitation of fibre allocation is expected to introduce systematic effects on the observed clustering (Burden et al. 2017; Cahn et al. 2017; Pinol et al. 2017), which must be included in any realistic mock catalogue. Thus, rich clustering information down to the scale of the DESI fibre patrol radius of  $1.4''$  is invaluable for the purpose of mock calibration and validation.

Since a discrete map of galaxies or quasars samples the continuous density field of matter, the number of galaxies within a randomly placed cell (counts-in-cells: Hubble 1934; White 1979; Peebles 1980) provides a window into the higher order correlations. Let  $P(N)$  be the probability that a cell with area  $\Omega$  contains exactly  $N$  galaxies. The factorial moments of this distribution,  $F_p = \langle N(N-1)\dots(N-p+1) \rangle$ , are related to the corresponding moments of the spatially smoothed underlying density field,  $\bar{\mu}_p = \langle (1+\delta)^p \rangle$ , via the simple relation  $\bar{\mu}_p = \frac{F_p}{\langle N \rangle^p}$ , which tidily includes shot-noise corrections arising from the fact that we are dealing with a discrete, locally Poissonian representation of the continuous field (Szapudi & Szalay 1993; Szapudi, Meiksin & Nichol 1996). From these moments, we can extract the correlation functions of corresponding order (also smoothed over the characteristic cell size),  $\bar{w}_p(\Omega) = \frac{1}{\Omega^p} \int_{\Omega} d\Omega_1 \dots d\Omega_p w_p(\theta_1, \dots, \theta_p)$ . The first few are listed below (e.g. Fry 1985; Fry et al. 2011):

$$\begin{aligned}\bar{\mu}_2 &= 1 + \bar{w}_2 \\ \bar{\mu}_3 &= 1 + 3\bar{w}_2 + \bar{w}_3 \\ \bar{\mu}_4 &= 1 + 6\bar{w}_2 + 3\bar{w}_2^2 + 4\bar{w}_3 + \bar{w}_4.\end{aligned}\quad (16)$$

Szapudi & Colombi (1996) classify theoretical errors on counts-in-cells statistics as either cosmic errors (due primarily to shot noise, edge effects, and finite volume) or measurement errors (due to the finite number of sampling cells), the latter scaling as the inverse of the number of sampling cells used. The counts-in-cells distribution and its moments are usually determined by throwing random cells over the region of interest, with massive oversampling required to control the measurement errors. However, Szapudi (1998) describes a method of implementing counts-in-cells that is essentially equivalent to using an infinite number of random cells, thereby eliminating the measurement error entirely. We implement this method using our publicly available code *infic*.<sup>7</sup> Therefore, we present only the uncertainty associated with cosmic errors, which we approximate by calculating counts-in-cells over two large fields, one in each galactic hemisphere, and measuring the mean and dispersion, weighted by effective area.

<sup>7</sup> <https://github.com/ekitanidis/infic>

**Table 1.** Area and sky fraction covered by *exactly* 0,1,2,3,4,5 + exposures in each optical band in DECaLS DR7. The ‘all’ columns are cumulative, such that the  $N_{\text{EXP}} = 1$  row refers to area and sky fraction covered by *at least* one exposure in all bands. Areas are estimated by sampling the footprint with randoms. In this context, the sky fraction is defined relative to the total DECaLS footprint, whereas elsewhere in the paper, it is defined relative to the area where imaging exists ( $N_{\text{EXP}} > 0$ ) in all three optical bands.

$N_{\text{EXP}}$	$f_{\text{sky}}$				Area (deg <sup>2</sup> )			
	$g$	$r$	$z$	All	$g$	$r$	$z$	All
0	0.14	0.11	0.05	1.0	1572.5	1263.8	541.6	11243.6
1	0.17	0.14	0.11	0.82	1865.1	1529.1	1292.5	9273.5
2	0.23	0.21	0.18	0.63	2620.5	2329.9	2035.4	7114.6
3	0.2	0.22	0.24	0.39	2279.5	2430.0	2665.8	4380.5
4	0.09	0.11	0.16	0.19	1032.8	1274.3	1846.3	2185.9
5 +	0.17	0.21	0.25	0.11	1873.2	2416.5	2861.9	1220.6

## 3 IMAGING CATALOGUES

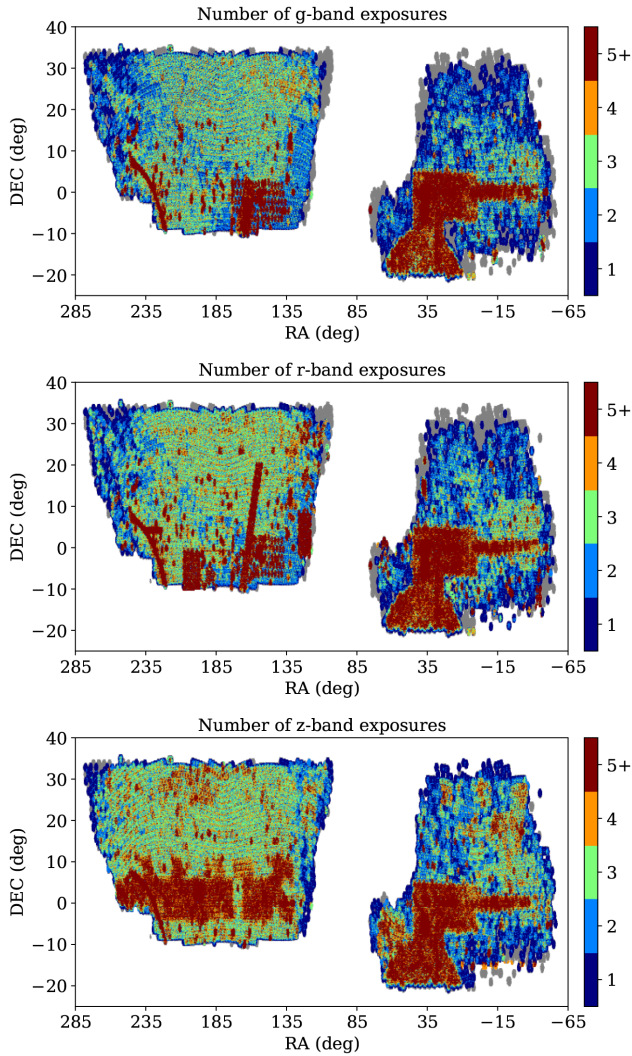
### 3.1 Imaging data

The DECam Legacy Survey (DECaLS) is a wide-field photometric survey amassing deep multicolour imaging within the footprints of ongoing and future spectroscopic surveys. Using the DECam instrument (Flaugher et al. 2015) at the Blanco 4-m telescope, DECaLS observed in three optical/near-infrared bands ( $g$ ,  $r$ ,  $z$ ), complemented by four mid-infrared bands from the Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010). DECaLS aims to obtain images up to  $5\sigma$  point-source depths  $g = 24.7$ ,  $r = 23.9$ , and  $z = 23.0$  AB mag, and is designed to boost the science power of spectroscopic observations by providing publicly available imaging with superior depth (1–2 magnitudes fainter) and enhanced image quality compared to existing photometry from SDSS, ATLAS, and Pan-STARRS. DECaLS covers a large equatorial region (bounded by  $\delta < 32^\circ$  in galactic coordinates), corresponding to roughly two-thirds of the optical imaging used for DESI targeting, and is a key piece of the Legacy Survey project, which has imaged the full 14 000 square degrees of extragalactic sky making up the DESI survey (Dey et al. 2018).

Raw DECam images are processed through the NOAO Community Pipelines, with astrometric calibration and photometric characterization based on Pan-STARRS-1 measurements. The calibrated images are then run through *The Tractor*<sup>8</sup> (Lang, Hogg & Mykytyn 2016), which produces an inference-based catalogue by optimizing the likelihood for source properties, given the data and a noise model. We use Data Release 7 (DR7), the seventh public data release of the Legacy Survey, which is the last DECaLS-only data release, including observations from August 2013 through March 2018 (NOAO survey program 0404). It also uses non-DECaLS observations from DECam conducted between August 2013 and March 2018, including some data from the Dark Energy Survey (DES; DES Collaboration 2005). Together, these cover approximately 9766 square degrees in the  $g$  band, 9853 square degrees in the  $r$  band, and 10610 square degrees in the  $z$  band, with 9298 square degrees observed in all three optical bands (see also Table 1, Fig. 1).

<sup>8</sup> <https://github.com/dstndstn/tractor>





**Figure 1.** Number of exposures in each DECaLS band in DR7, estimated by sampling the footprint with randoms. No map projection is applied here.

**Table 2.** Summary of selection properties for each of the dark-time DESI target classes. W1 and W2 denote WISE bands. ‘Primary’ refers to the band used to define the limiting magnitude, which is relevant for the completeness mask (see Section 4.1).

Target	Redshift range	Selection bands	
		Primary	Other
LRG	0.4–1.0	$z$	$g, r, W1$
ELG	0.6–1.6	$g$	$r, z$
QSO (tracers)	$<2.1$	$r$	$g, z, W1, W2$
QSO (Ly- $\alpha$ )	2.1–3.5	$r$	$g, z, W1, W2$

### 3.2 Target selection

Table 2 summarizes the primary target types evaluated in this paper. These targets are defined in great detail in the FDR and their selection algorithms are briefly outlined below.

#### 3.2.1 LRG

LRG targets are selected from the  $g, r, z$ , and  $W1$  bands by applying a series of colour cuts using extinction-corrected magnitudes. No

morphology cut is applied.

$$\begin{aligned}
 &18.01 < z < 20.41 \\
 &0.75 < r - z < 2.45 \\
 &-0.6 < (z - W1) - 0.8(r - z) \\
 &(z - 17.18) / 2 < r - z < (z - 15.11) / 2 \\
 &(r - z > 1.15) \parallel (g - r > 1.65).
 \end{aligned} \tag{17}$$

#### 3.2.2 ELG

ELG targets are selected from the  $g, r$ , and  $z$  bands by applying a series of colour cuts using extinction-corrected magnitudes. No morphology cut is applied.

$$\begin{aligned}
 &21.00 < g < 23.45 \\
 &0.3 < (r - z) < 1.6 \\
 &(g - r) < 1.15(r - z) - 0.15 \\
 &(g - r) < 1.6 - 1.2(r - z).
 \end{aligned} \tag{18}$$

#### 3.2.3 QSO

QSO targets are selected by applying a machine-learning method based on Random Forests (RF), which relies only on extinction-corrected object colours in the  $g, r, z, W1$ , and  $W2$  bands (Christophe Yèche, private communication). The algorithm is trained using all known QSOs in the footprint with an initial cut of  $r < 23$  against a sample of unresolved objects from Stripe 82 without known QSOs and objects exhibiting QSO-like variations in their light curves. In the target selection itself, a tighter initial cut of  $r < 22.7$  is applied.

### 3.3 Catalogues and randoms

DESI target catalogues and uniform random catalogues are created with our public code ImagingLSS.<sup>9</sup> ImagingLSS processes the outputs of the DECaLS pipeline and selects DESI targets from it, as well as providing the option for auxiliary, user-defined targets. Uniform unclustered randoms are sampled from the imaging survey footprint. Geometric survey masks (for example, vetoing by proximity to bright objects) can be applied to both catalogues and randoms in a consistent manner. In addition to DECaLS data, ImagingLSS uses the SFD98 dust map (Schlegel, Finkbeiner & Davis 1998) to correct for extinction, as well as the Tycho-2 star catalogue (Høg et al. 2000) and the AllWISE catalogue (Cutri et al. 2015) to mask out bright stars.

### 4 IMAGING MASKS

We develop two types of initial masks, completeness and bright star, to reject possibly problematic regions of the imaging data. The effects of these masks on the survey efficiency and effective area are summarized in Table 3. Our baseline sample is selected from regions where imaging exists in all three optical bands used for targeting (‘no mask’), and all sky fractions are quoted relative to this sample. In subsequent sections, we will review the purpose and implementation of each mask, with Section 4.1 focused on imaging completeness, and Section 4.2 describing our bright star mask. We also investigate whether there is need for a mask around extended sources such as large galaxies.

<sup>9</sup><https://github.com/desihub/imaginglss>



**Table 3.** Summary of masks and how each affects the number of targets and the effective area and sky fraction. Here, sky fractions are quoted relative to the ‘no mask’ case, which is simply the joint regions of the footprint where imaging is available in all three optical bands.

Target	Mask	Number	Area (deg <sup>2</sup> )	$f_{\text{sky}}$
LRG	No mask	4882206	9298.91	1.0
	Complete only	4872537	9281.0	1.0
	Bright star only	4304375	8584.68	0.92
	All masks	4296486	8568.65	0.92
ELG	No mask	23224353	9298.91	1.0
	Complete only	23032874	9227.78	0.99
	Bright star only	20329143	8638.84	0.93
	All masks	20166887	8575.15	0.92
QSO	No mask	3125148	9298.91	1.0
	Complete only	3116976	9288.35	1.0
	Bright star only	1814359	6655.11	0.72
	All masks	1810801	6648.68	0.71

#### 4.1 Survey depth and completeness

Tractor catalogues contain an estimation of the imaging depth at each observed pixel in the footprint. This depth is affected by the number of exposures, exposure times, observational conditions, and instrument effects.<sup>10</sup> Due to the multipass nature of the imaging survey and the fact that it is ongoing, variations in depth across the footprint are substantial. In DR7, some of the sky has been covered just once, while the deepest regions have received five or more passes (see Table 1 and Fig. 1).

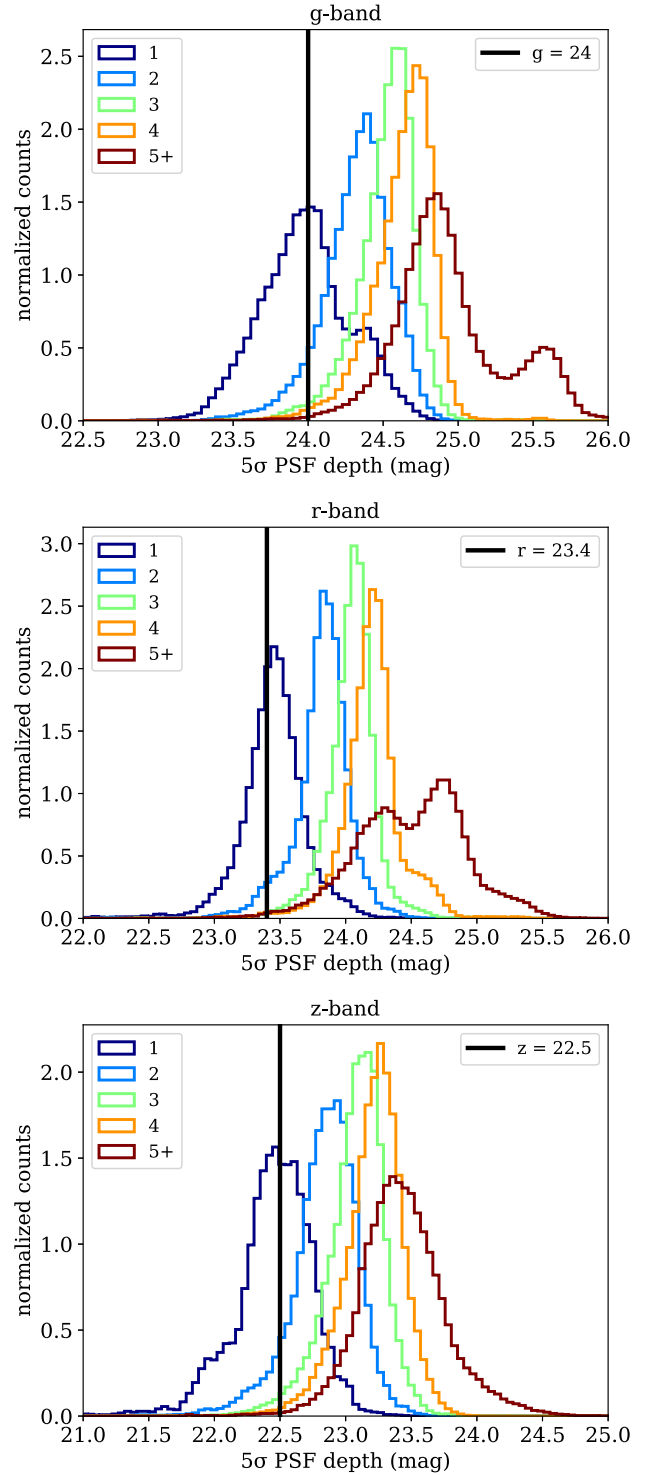
To select a uniform and complete sample, we implement a ‘completeness’ mask. Since two exposures, at minimum, are needed to meet the nominal depth requirements of DESI over most of the footprint (Dey et al. 2018), a reasonable approach might be to mask out areas with fewer than two exposures in each band. However, not every pixel will exceed target depth for a given number of overlapping exposures (see Fig. 2), and thus the result will still be biased towards regions with more passes. Even with perfectly uniform coverage, variations in observing conditions affect depths and therefore the homogeneity of the resulting catalogue.

Instead, using the  $5\sigma$  point-source depths as limiting magnitudes, combined with DESI target definitions, we construct a ‘binary completeness mask,’ in which a particular observed pixel is in the ‘complete’ area for a target type if and only if it meets the following conditions:

- (i) the limiting magnitudes in the bands used for magnitude cuts are sufficient to observe even the faintest targets with  $5\sigma$  confidence;
- (ii) imaging exists for all bands used in the target definitions.

This ensures that only the ‘deep enough’ regions of the sky are used to generate DESI catalogues and randoms for analysis. Fig. 3 shows the sky fraction as a function of depth for each band, with the magnitude cuts for the three targets plotted as vertical lines, to visualize how shifting a magnitude cut up or down would affect usable sky area. As Fig. 3 shows, all depths are sufficient to detect the full target samples.

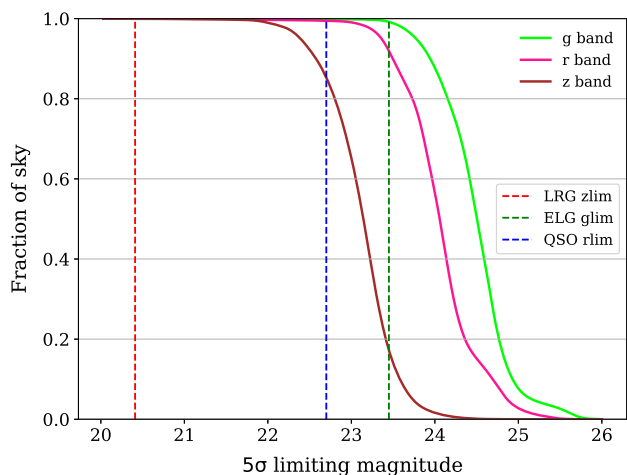
<sup>10</sup>Note that we do not explicitly apply the ALLMASK flag, which uses the NOAO Community Pipeline’s data quality map to mask out bad pixels on the CCD and pixels affected by bleed trail, transients such as cosmic rays, and saturation. These effects are accounted for in the estimation of the depths (Dey et al. 2018) and thus are perforce included in our completeness mask.



**Figure 2.** Histograms of  $5\sigma$  point-source depths of randoms in each band, normalized as probability densities, with the coloured curves corresponding to different numbers of exposures. The solid vertical lines are the DESI nominal  $5\sigma$  depth requirements  $g = 24.0$ ,  $r = 23.4$ , and  $z = 22.5$  for an ELG galaxy with half-light radius of 0.45 arcsec.

By only requiring the existence of imaging in the bands used for colour cuts, we are implicitly assuming that uncertainties in the colours do not affect the reliability of the target selection. Alternatively, we could require a  $5\sigma$  detection in all bands, not





**Figure 3.** Cumulative sky fraction versus  $5\sigma$  limiting magnitudes, with target selection cuts shown as vertical lines to demonstrate the effect that shifting a magnitude limit up or down would have on the completeness of the corresponding target.

just the primary bands used to apply magnitude cuts. However, it is not clear how to define a detection limit in a band that is used only for colour cuts, as the targets are not necessarily bounded in these bands. The other advantage of defining the completeness mask on only the bands used for magnitude cuts is that it makes it easier to apply the definition to target selections of finer granularity in the colour space, for example, represented by a forest of decision trees.

## 4.2 Bright foregrounds

As discussed in Section 3, we correct for galactic extinction by adjusting the colour of objects using the SFD98 dust map. In addition to extinction, bright objects in the foreground (point sources such as stars, and extended sources such as nearby galaxies) can also affect the detection of targets in their angular vicinity, due to charge-coupled device (CCD) saturation and diffraction spikes that contaminate the surrounding pixels. The systematics due to these bright objects in the target catalogues are spatially localized and uncorrelated with the underlying true density of the targets. Thus, we can quantify these systematic effects by measuring the densities of targets as a function of their proximity to the bright foreground objects and mitigate them via masks.

### 4.2.1 Bright stars

We use two bright star catalogues, Tycho-2 and WISE. Tycho-2 is a reference catalogue of the 2.5 million brightest stars in the sky, with photometry in two optical bands,  $\lambda_{BT} = 435$  nm and  $\lambda_{VT} = 505$  nm. It has highly accurate astrometric positions and is 99 per cent complete out to  $VT \sim 11$  magnitude, making it well suited to our analysis. However, some stars not included in the Tycho-2 catalogue may still be bright enough in the near-infrared bands to affect the detection of LRGs and QSOs, both of which use the WISE band W1 in their target selection (QSOs also use W2). To be safe, we create separate veto masks for each star catalogue and apply both to our data. Table 4 compares the effective areas and sky fractions of the two star masks.

For both catalogues, we follow a similar procedure to construct and test a radial mask for each target class. We begin by splitting the sample of stars into magnitude bins, since the masking radius

**Table 4.** Summary of star masks and how each affects the number of targets and the effective area and sky fraction. Here, sky fractions are quoted relative to the ‘no mask’ case, which is simply the joint regions of the footprint where imaging is available in all three optical bands.

Target	Mask	Number	Area (deg <sup>2</sup> )	$f_{\text{sky}}$
LRG	No mask	4882206	9298.91	1.0
	Tycho mask only	4385463	8752.25	0.94
	WISE mask only	4529935	8970.73	0.96
	Both star masks	4304375	8584.68	0.92
ELG	No mask	23224353	9298.91	1.0
	Tycho mask only	21378695	8811.44	0.95
	WISE mask only	21065984	8970.73	0.96
	Both star masks	20329143	8638.84	0.93
QSO	No mask	3125148	9298.91	1.0
	Tycho mask only	2192932	7541.98	0.81
	WISE mask only	2030414	7269.13	0.78
	Both star masks	1814359	6655.11	0.72

will be magnitude-dependent. The bin widths are chosen such that each bin contains a similar number of stars and therefore has comparable Poisson errors. For a given magnitude bin, we use the pair-enumeration algorithm in KDcount<sup>11</sup> to efficiently locate all star-galaxy pairs within a distance of  $\theta = 0.05$  rad  $\approx 10^4$  arcsec of one another. We calculate the density of targets around each star in logarithmically spaced annular bins up to this maximum separation and then average across all stars to determine the mean density of targets in each annular bin. By assigning a conservative cutoff radius to each magnitude bin by eye and fitting the results, we obtain the following magnitude–radius relations:

For Tycho-2:

$$R = \begin{cases} -2.5 \times VT^3 + 77.4 \times VT^2 \\ -813.6 \times VT + 2969 \text{ arcsec}, & \text{LRG, ELG} \\ 2.8VT^2 - 143.4VT + 1387.1 \text{ arcsec}, & \text{QSO} \end{cases} \quad (19)$$

For WISE:

$$R = \begin{cases} 10^{3.29 - 0.18 \times W1} \text{ arcsec}, & \text{LRG, ELG} \\ 10^{3.29 - 0.12 \times W1} \text{ arcsec}, & \text{QSO} \end{cases} \quad (20)$$

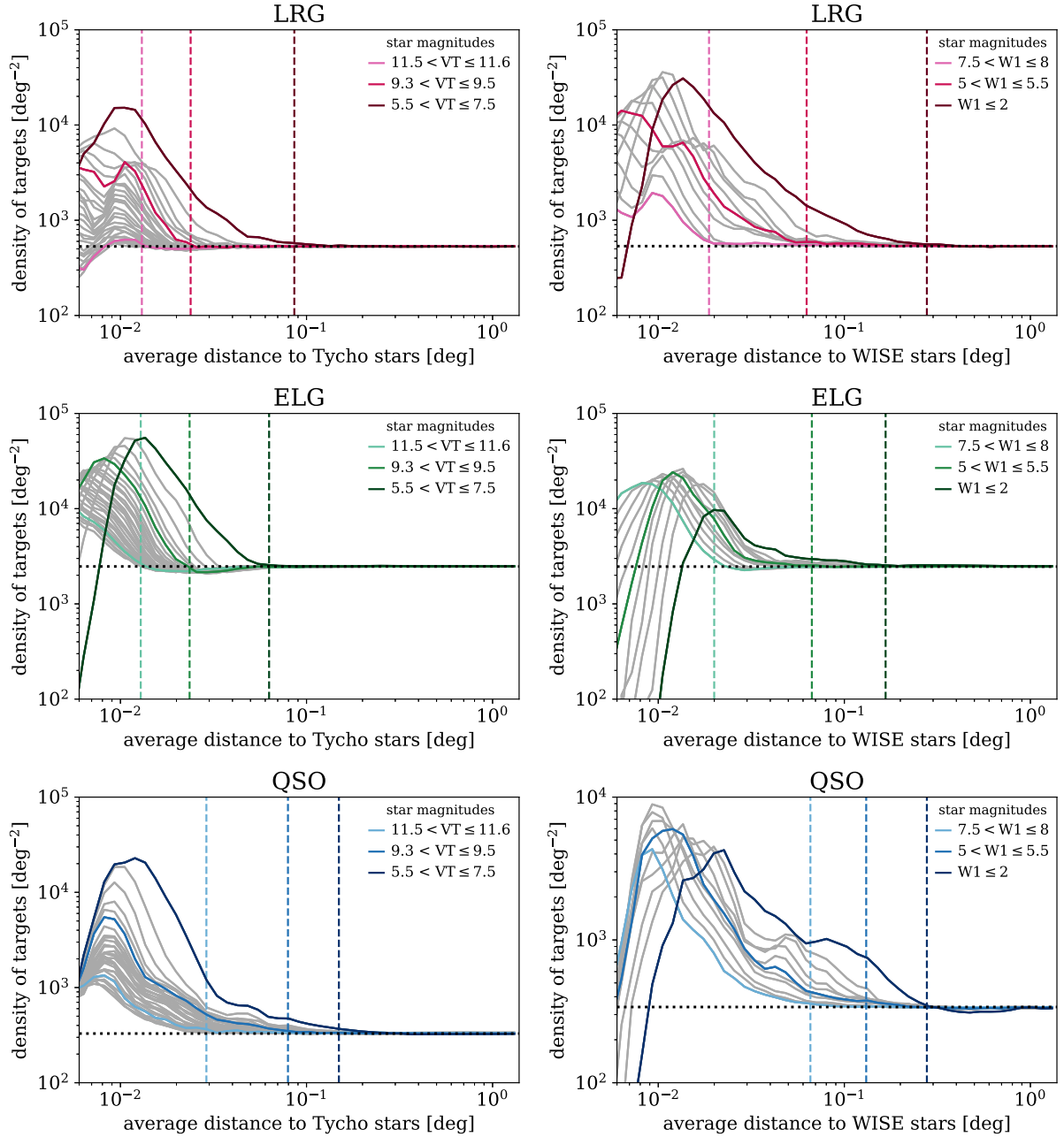
For both types of star mask, the LRGs and the ELGs can be fit to the same magnitude–radius relation, while the QSOs require their own, more conservative mask. We found that relaxing the QSO stellar masks (by, for instance, applying the LRG/ELG masks instead) led to a measurable increase in stellar contamination, evidenced by inflated QSO autocorrelation and QSO–star cross-correlation measurements.

In our analysis, we implement equations (19) and (20) instead of using the MASKBITS column provided by DECaLS to mask out stars. We have found that our geometric masks are more aggressive than the combination of available stellar bitmasks; in particular, they are necessary for removing contaminated areas at larger radii.

Fig. 4 shows the resulting target densities versus distance to bright stars, sectioned by target type and star catalogue, with each magnitude bin plotted separately. Three bins are highlighted for illustration, with the corresponding dashed vertical lines representing the relevant masking radii calculated with the average magnitude of that bin. These plots show how the masks eliminate spurious clustering due to bright stars. In Figs 5 and 6, we also present 2D histograms of target densities around stacks of bright stars, plotted

<sup>11</sup><https://github.com/rainwoodman/kdcount>





**Figure 4.** Average density of DESI targets as a function of distance to bright stars from Tycho-2 (left) and WISE (right). Both star catalogues are divided into magnitude bins, which are spaced such that each contains a roughly comparable number of stars. In each plot, three bins are highlighted for illustration, with the dashed vertical lines representing the corresponding masking radius (equations 19 and 20) calculated using the average magnitude of that bin.

in equatorial coordinates<sup>12</sup> with star masks drawn on as circles, again showing well-fitting mask radii.

#### 4.2.2 Bright extended sources

Similarly to bright stars, we examine the density of targets near bright extended sources such as nearby galaxies. We use the 2MASS

<sup>12</sup>We also performed this analysis in ecliptic coordinates to see if additional structure could be identified. In some of the fainter magnitude bins, resolution and contrast could be manipulated to resolve an x-shaped feature in the 2D stacked plots, but this feature was very fine, and we found that removing it beforehand had no perceptible impact on the 1D density plots.

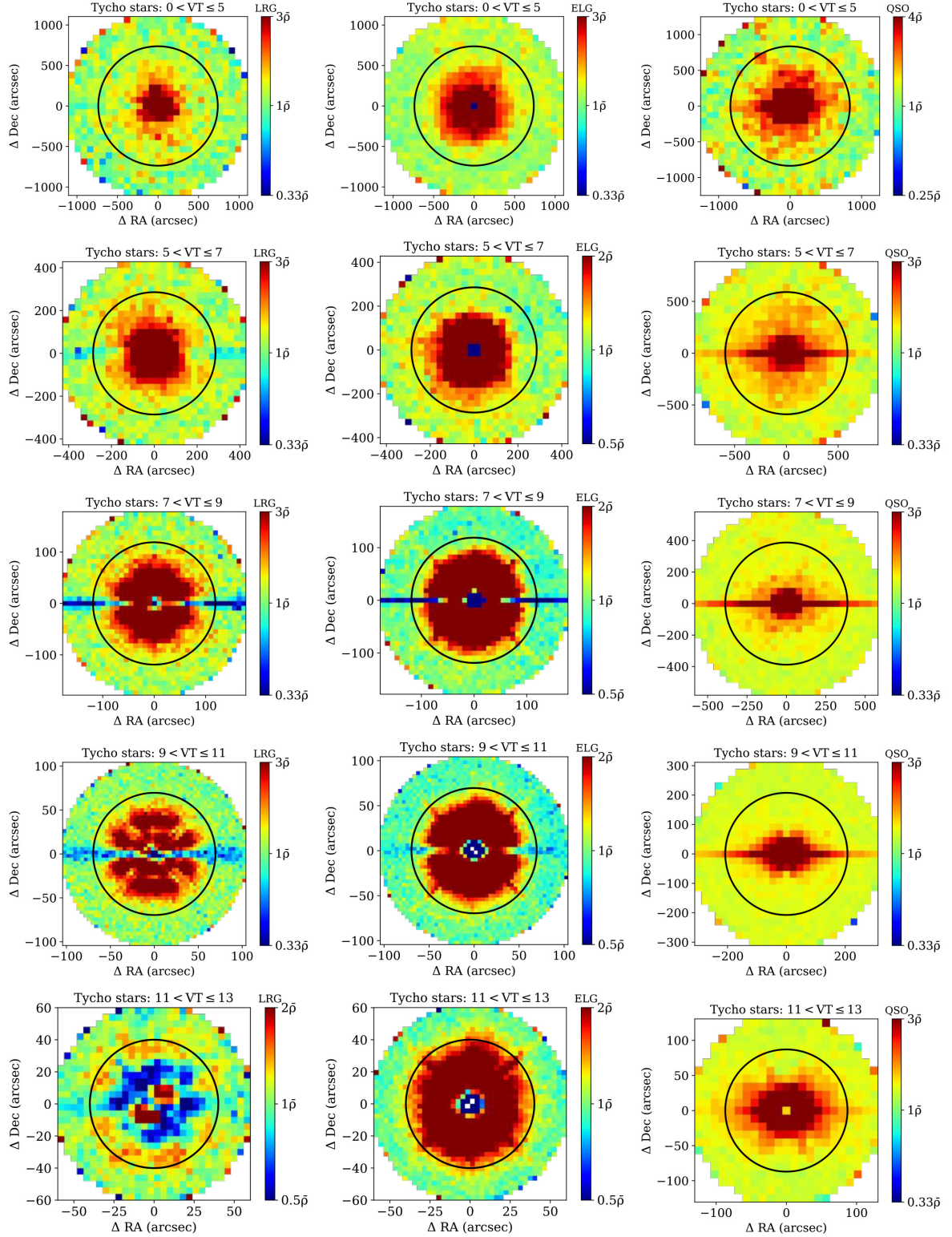
Extended Source Catalog (Jarrett et al. 2000), a catalogue of near-IR extended sources complete for angular sizes greater than  $\sim 10$  arcsec. Restricting to  $10 < J < 15$  total  $J$ -band magnitude, we find no appreciable impact on the density of our DESI dark time targets (Fig. 7), and thus we do not apply a mask.

## 5 SPATIAL VARIATIONS

### 5.1 NGC versus SGC

We calculate the angular correlation functions in the north galactic cap (NGC) and the south galactic cap (SGC) individually, as the two hemispheres may suffer from different systematics, and earlier analyses have found NGC/SGC variations in BOSS data



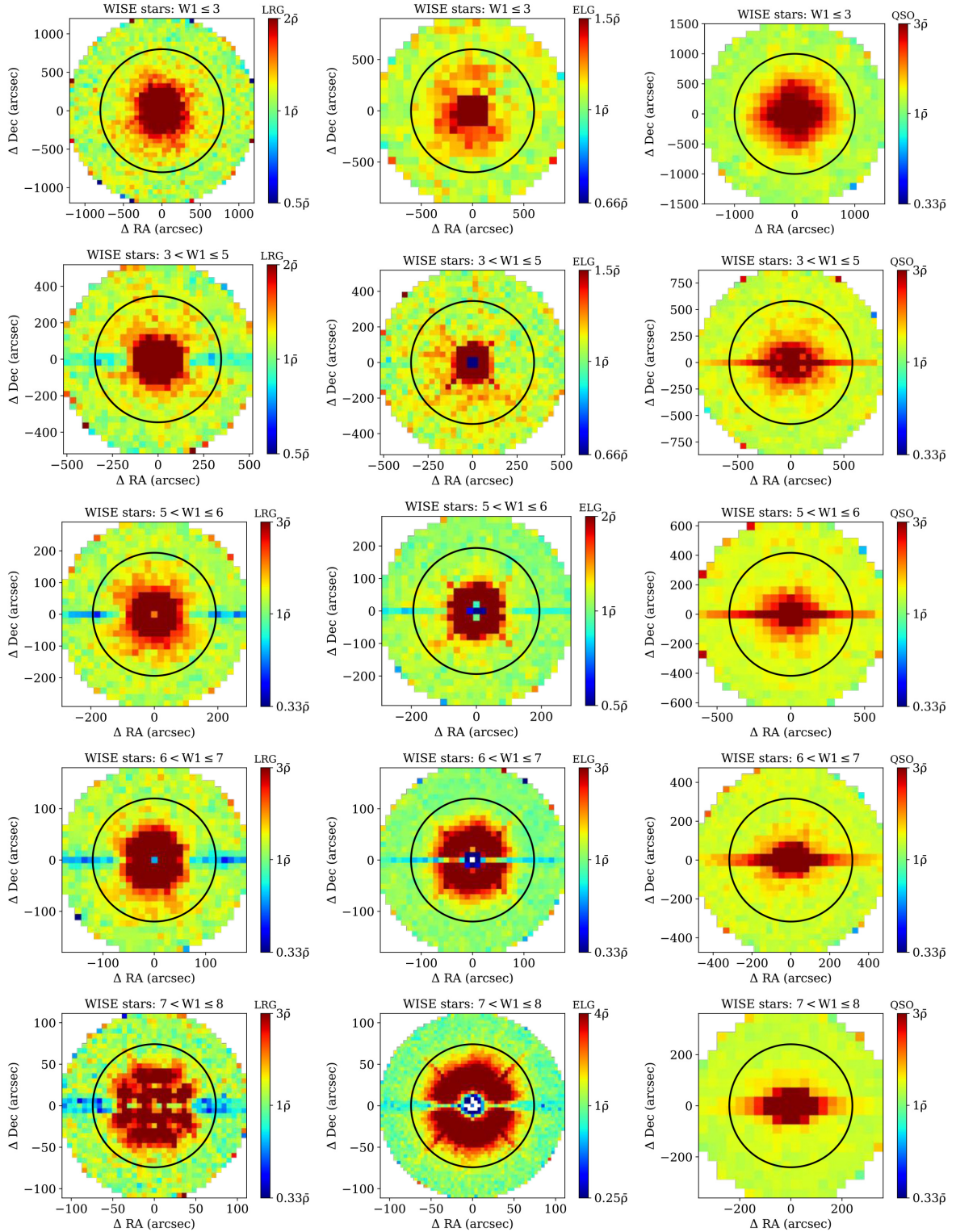


**Figure 5.** 2D histograms of the density of DESI targets around stacks of bright stars from Tycho-2. The solid black circles represent our star masks (equation 19). The horizontal features appearing in some maps, which are due to insufficient masking of charge bleed trails in the CCDs, are only a few arcsec in width, and we find that removing them with a separate rectangular mask has no perceptible impact on the densities around stars.

(see Ross et al. 2012, Section 4.1, for an explanation of the origin of this difference in number density between the NGC and the SGC in BOSS). The results are shown in Fig. 8. For LRGs, the autocorrelations are virtually identical. For ELGs, there is a

slight divergence, most noticeably in the  $0.02^\circ < \theta < 0.09^\circ$  range. For QSOs, the difference between NGC and SGC is significantly more pronounced. The NGC results appear more impacted by systematics, as indicated by a bulge in the correlation function





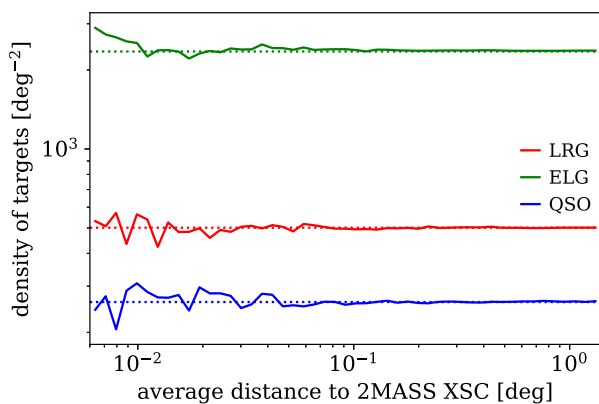
**Figure 6.** 2D histograms of the density of DESI targets around stacks of bright stars from WISE. The solid black circles represent our star masks (equation 20). The horizontal features appearing in some maps, which are due to insufficient masking of charge bleed trails in the CCDs, are only a few arcsec in width, and we find that removing them with a separate rectangular mask has no perceptible impact on the densities around stars.

with extremely large bootstrap errors. This is likely due to the fact that parts of the SGC, where there is DES imaging, are very deep (see also Fig. 9 in the following section to visually observe how the NGC appears more impacted by systematics for QSO).

## 5.2 Visual inspection

As an initial sanity check, we create maps of the density contrast  $\delta = n/\bar{n} - 1$  averaged over HEALPix pixels of  $N_{\text{SIDE}} = 256$ . The results are shown in Fig. 9, with the unmasked catalogues





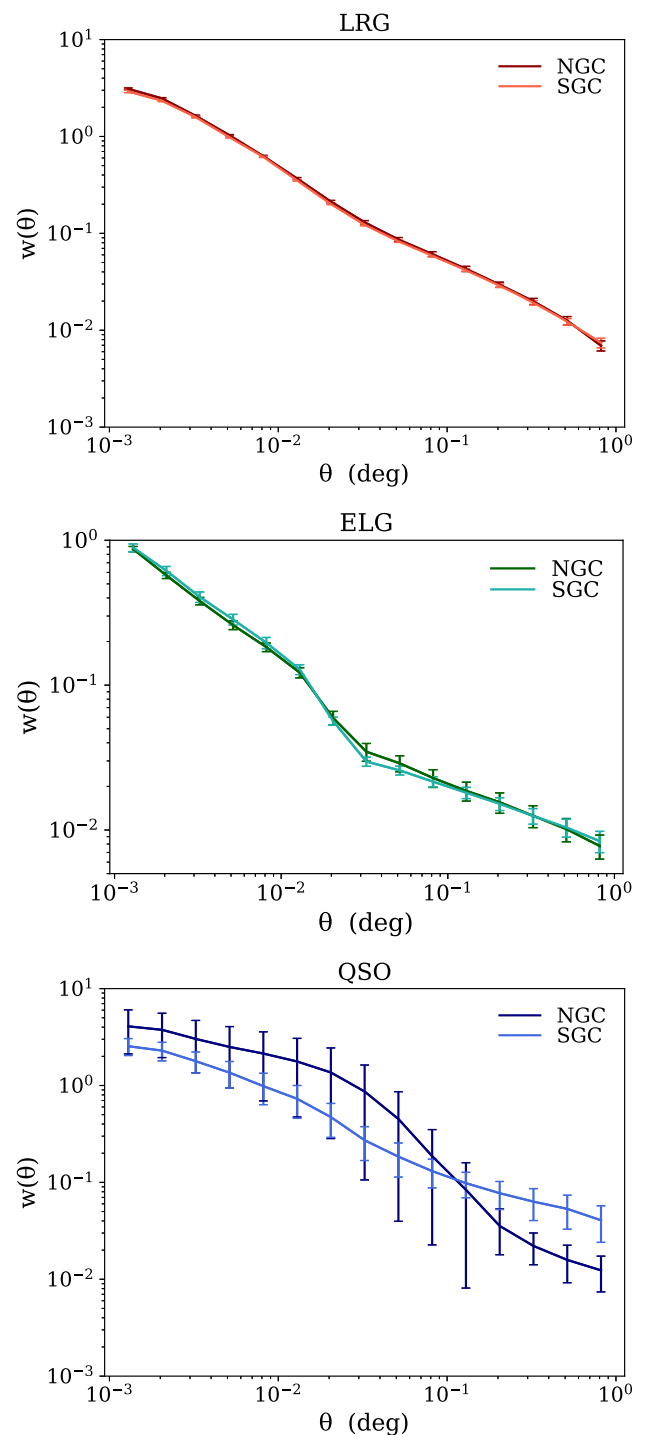
**Figure 7.** Average density of DESI dark time targets as a function of distance to extended sources from 2MASS-XSC.

mapped on the left and their masked counterparts on the right. The masked catalogues are visually cleaner, with the star masks reducing stellar contamination and the completeness masks cancelling the imprint of imaging depth on target density. Several features remaining in the masked maps are highlighted and discussed below.

While LRG clustering appears relatively uniform, ELG clustering shows some troubling large-scale trends. For example, the shape of the DES region in the south is detectable, appearing underdense despite its superior depth. This suggests contamination in non-DES regions, likely due to the effect of low redshift ( $z < 0.8$ ) objects preferentially scattering into the ELG target selection across the low- $z$  colour cut  $(g-r) < 1.15(r-z) - 0.15$  in regions of worse depth. This was tested (Ashley Ross, private communication) by injecting artificial noise into regions of very deep imaging and examining the photometric redshifts of the resulting scattered objects. Additionally, a few suspicious ‘hotspots’ appear in the ELG density. When examined closely, most of these occur around a small set of very bright stars (such as Arcturus) that have been successfully masked but which cause dramatic and complex artefacts in the image beyond the expected masking radius due to reflections of pupil ghosts. Finally, looking at the density of QSOs, there remains noticeable stellar contamination along the galactic plane even after applying conservative masks, as well as the Sagittarius Stream in the north, and there are also some effects at the edges of the footprint.

### 5.3 Outlier analysis

As another test of spatial variation, we perform a jackknife-inspired outlier analysis on the data. Again using the HEALPix scheme, we divide the footprint into large pixels and re-calculate  $w(\theta)$  with each non-empty pixel excluded in turn, still performing the full bootstrap error analysis on the remaining pixels for each iteration. We begin with the coarsest pixels, corresponding to  $N_{\text{SIDE}} = 1$ , and increase the resolution as needed to resolve any anomalies that are detected. For LRGs and ELGs, the results are indistinguishable even at this minimum resolution. However, for QSOs, we find that two pixels at resolution  $N_{\text{SIDE}} = 4$  which, when either is excluded, lead to a significant change in the correlation function (see Fig. 10). Likely, culprits are the Coma Cluster (Abell 1656), which contains over 1000 galaxies, and M3/NGC 5272, one of the largest and brightest globular clusters in the sky.

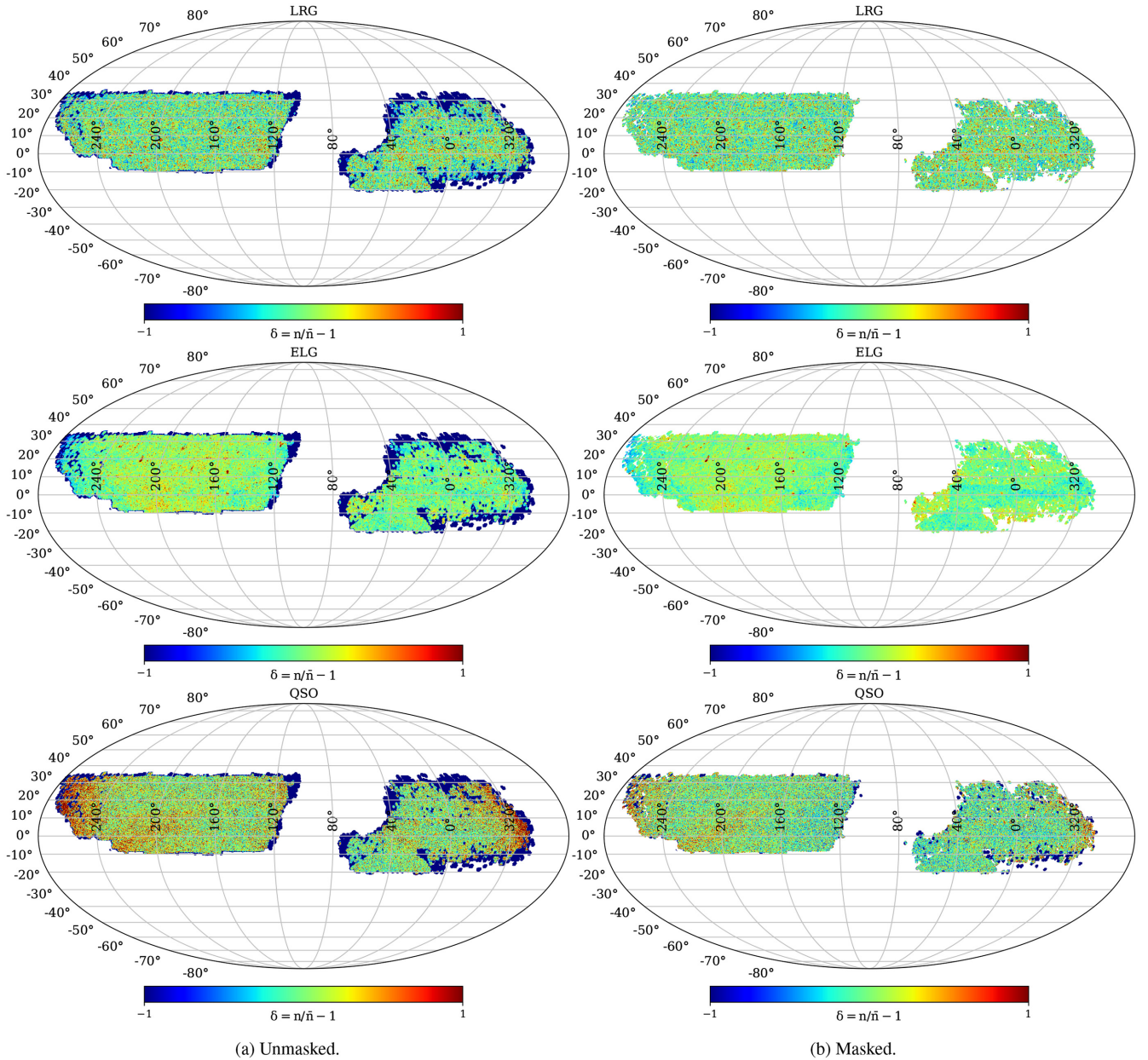


**Figure 8.**  $w(\theta)$  for LRGs, ELGs, and QSOs calculated in NGC and SGC separately. Error bars are from bootstrapping.

## 6 POTENTIAL SYSTEMATICS

Potential systematics include astrophysical foregrounds, variations in observing conditions, and uncertainties in data calibration, processing, and reduction (for similar studies in the context of SDSS, see e.g. Myers et al. 2006; Crocce et al. 2011; and Ross et al. 2011; for similar analyses using DES Verification Data, see e.g. Crocce et al. 2016; Leistedt et al. 2016; Suchyta et al. 2016; and Elvin-Poole et al. 2018). We introduce maps of spatially varying potential





**Figure 9.** Maps of the density contrast  $\delta = n/\bar{n} - 1$  calculated with HEALPix resolution  $N_{\text{SIDE}} = 256$ . Mollweide projection in equatorial coordinates with right ascension centred at  $\text{RA} = 100^\circ$ . Masked data (right) are less impacted by stellar contamination and variations in imaging depth than raw data (left).

systematics in Section 6.1 and examine their impact on the densities of DESI targets, before and after applying photometric weights, in Section 6.2. The purpose of these weights is to mitigate the density trends by upweighting (or downweighting) regions where target density is diminished (or enhanced) due to systematics. Finally, in Section 6.3, we cross-correlate the targets with stars and attempt to quantify stellar contamination in the QSO sample.

### 6.1 Maps of potential systematics

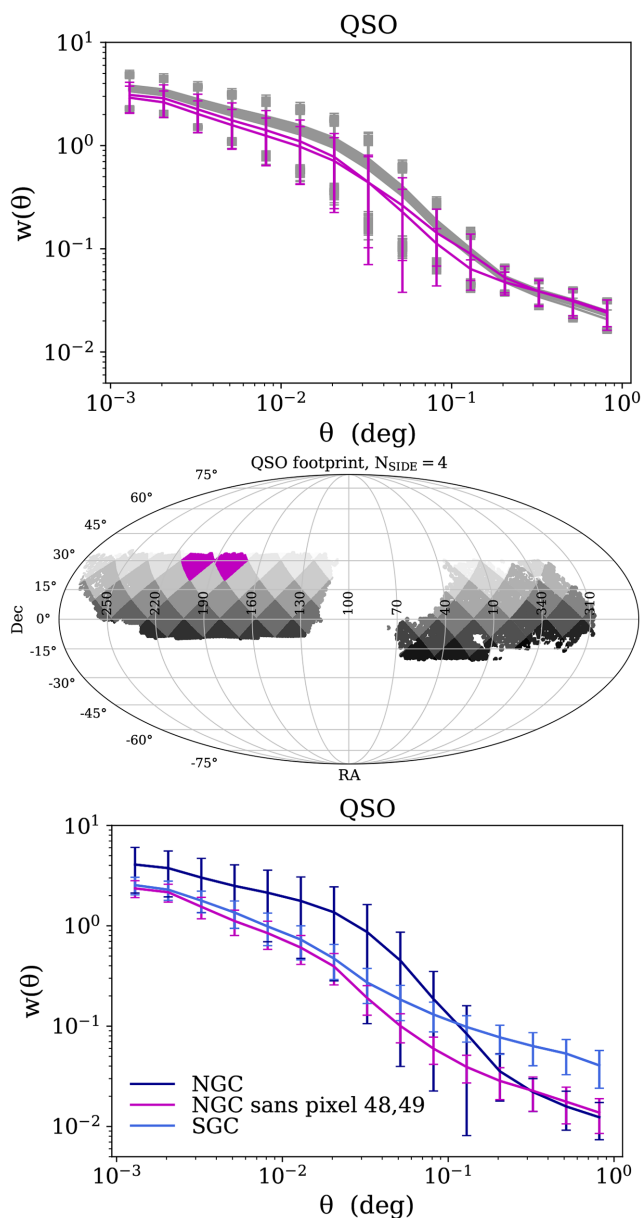
We begin by using the HEALPix scheme with  $N_{\text{SIDE}} = 256$  to divide the data into equal-area pixels of approximately 0.05 square degrees each. This resolution was chosen to avoid the shot-noise limit in which most pixels contain zero or one targets. For our LRG, ELG, and QSO samples with approximate mean densities (per square degree) 500, 2400, and 260, respectively, it produces an average of

10–20 LRGs/QSOs and  $\sim 100$  ELGs per pixel.<sup>13</sup> In every pixel, an average value for the potential systematic is calculated. For survey properties measured per CCD, unless otherwise noted, we first average over overlapping exposures to obtain a mean value for each random and then pixelize using the randoms. The resultant maps are shown below, along with brief descriptions of how they are determined and why they are included in the analysis.

(i) *Stellar density*: In addition to the detection issues near bright foreground stars discussed in Section 4.2.1, the presence of stars impacts the measured density and clustering of galaxies in other ways (e.g. Crocce et al. 2011; Ross et al. 2011). Stars with

<sup>13</sup>Note that most of the systematics studied here (with the exception of  $E_B - V$  and stellar density) are not available at higher resolution than this pixelization scheme in any case, as they are measured per CCD.

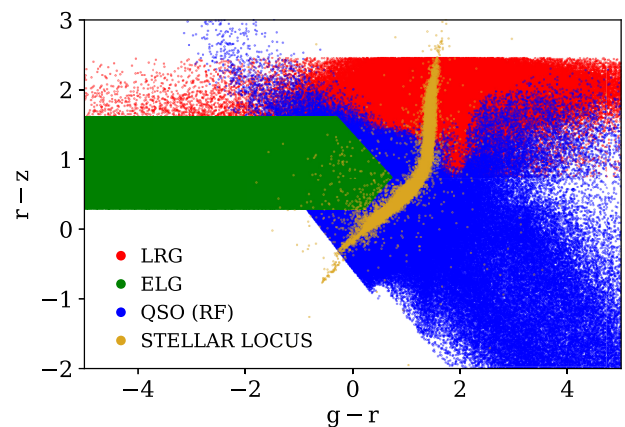




**Figure 10.** Upper: The angular correlation function is re-calculated with each pixel excluded in turn, still performing a complete bootstrap analysis on the remaining pixels. Two neighbouring pixels in the NGC affect the result when either is excluded (violet lines). Middle: Map highlighting the location of the two pixels, which were found to affect clustering. Lower: When both these pixels are excluded, the disparity between the NGC and the SGC is dramatically reduced on small scales.

similar colours (see Fig. 11) can contaminate the samples, and the inclusion of a separate population manifests as an enhanced clustering signal. The observed clustering may also be imprinted with the density gradient of stars, which increases towards the galactic plane. Furthermore, residual point spread function (PSF) tails of some fainter stars may pollute the pixels used to calculate sky background and therefore affect target photometry.

We create a catalogue of stars from DECaLS by selecting objects lying in the stellar locus (using the colour cut  $17 < r < 18$ ) with PSF morphology. The density of this class of objects, shown in the top left-hand panel of Fig. 12, indicates that it is a reliable stellar template.



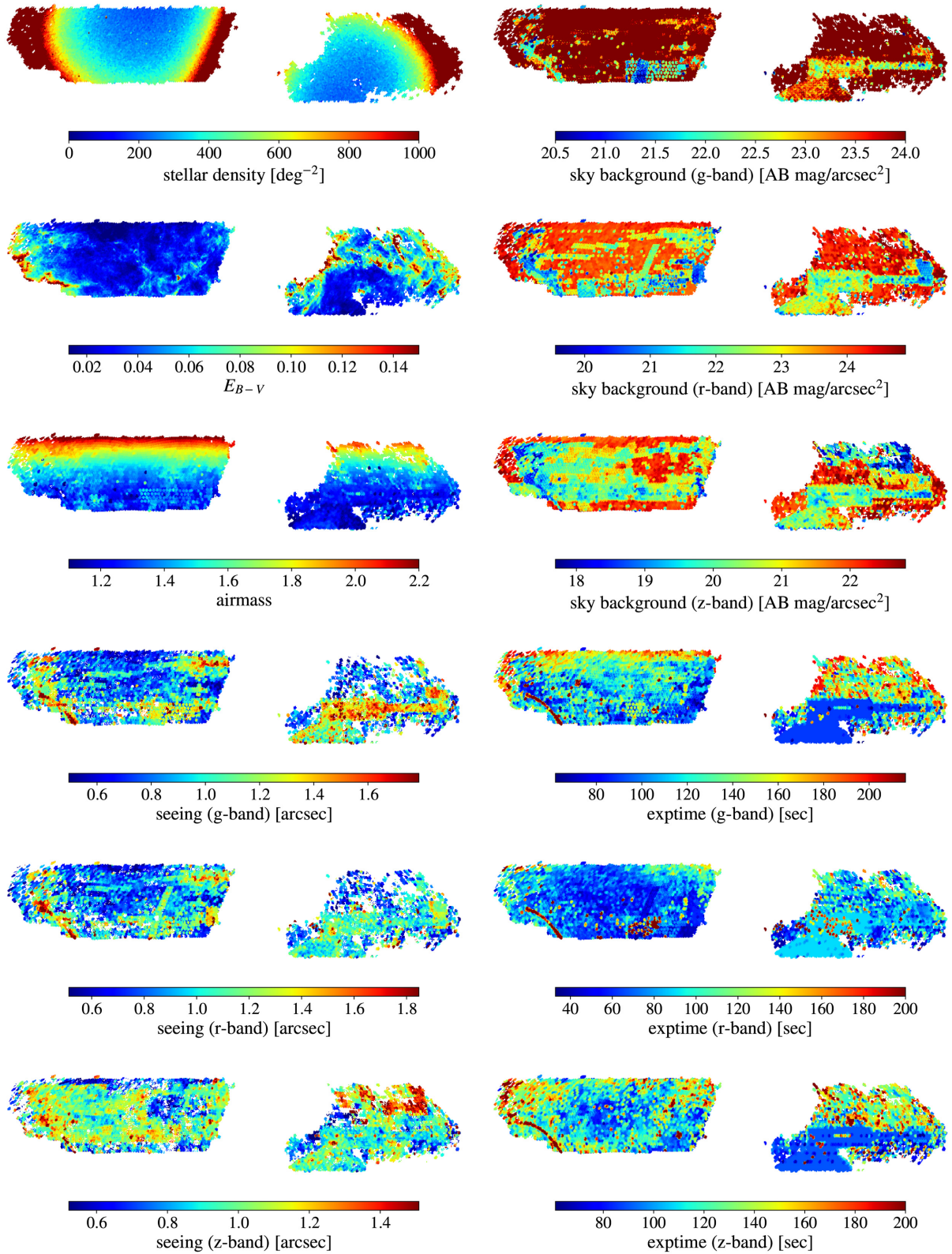
**Figure 11.** Colour-colour plot of LRG, ELG, QSO target selection, with the shape of the stellar locus selected from DECaLS as PSF-type objects with  $17 < r < 18$ .

(ii) *Galactic extinction*: Galactic extinction is the wavelength-dependent absorption and scattering of light by interstellar dust in the Milky Way, causing sources to appear redder. DESI and its imaging surveys deliberately avoid regions of high extinction along the galactic plane. In addition, we use extinction-corrected fluxes in our analysis. The total extinction in each band is provided by the DECaLS pipeline, calculated from the SFD98 dust map combined with a set of extinction coefficients  $A_\lambda/E_{B-V}$  for each DECam and WISE filter. The extinction coefficients were determined at airmass = 1.3 from the values recommended by Schlafly & Finkbeiner (2011) using the Fitzpatrick (1999) extinction curve at  $R_V = 3.1$ . These values are 3.214, 2.165, and 1.211 for  $g$ ,  $r$ , and  $z$ . However, it has been shown in other surveys that residual errors in this correction may cause spurious clustering (see e.g. Scranton et al. 2002; Myers et al. 2006; Ross, Brunner & Myers 2006). In addition, erroneously correcting the photometry of stars in the foreground of the dust could potentially bias their colour and cause some of them to scatter into target selection. Hence, we treat  $E_{B-V}$  as a potential systematic and test its effect on the density field.

(iii) *Seeing*: Astronomical seeing is the blurring of an image due to turbulence in the Earth's atmosphere. The distortions fill out a PSF whose full width at half maximum quantifies the quality of the seeing conditions. In DECaLS, seeing is determined by fitting the median PSF of stars on the CCD to a 2D Gaussian. Since seeing varies between nights and even exposures, a mean value is reported, calculated by averaging the inverse of the effective number of pixels in the PSF (such that images with better seeing dominate the mean, as they contain more information). We use this mean PSF size to determine the impact of seeing conditions on the density field, since bad seeing causes larger magnitude errors as well as more cross-contamination with stars due to poor morphological fits.

(iv) *Airmass*: Airmass is the optical path-length of light through the Earth's atmosphere. When photons from a celestial source travel to a terrestrial observer, they are absorbed and scattered by the atmosphere. Light that must traverse more atmosphere will be attenuated more, so sources appear dimmer at the horizon than at the zenith. For zenith angles  $\lesssim 60^\circ$ , we can approximate the atmosphere as plane-parallel and also assume that its density is more or less constant. In this limit, the airmass is simply the secant of the angle from the zenith to the source location on the sky. We treat the mean airmass as a potential systematic, as it contributes



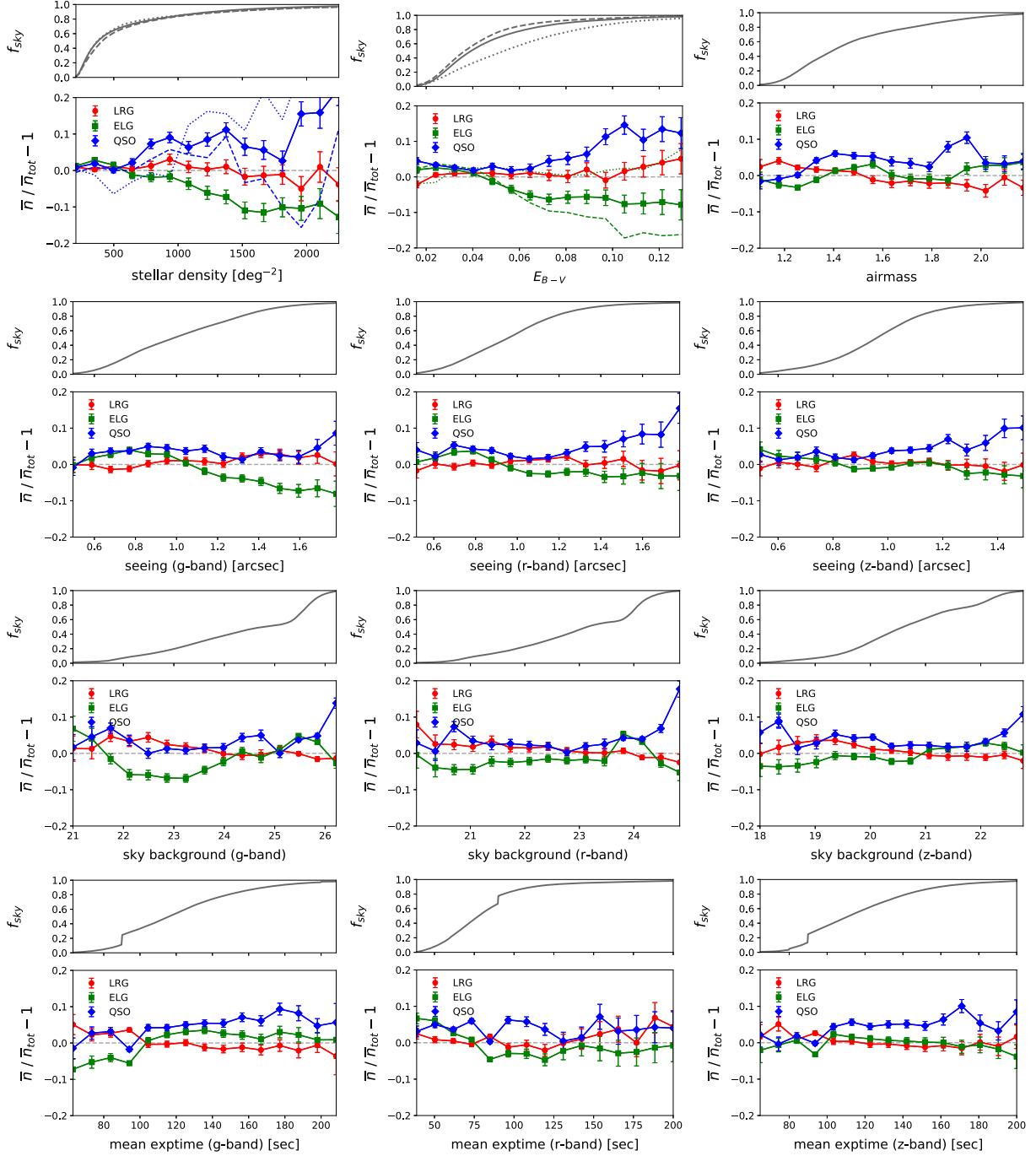


**Figure 12.** Maps of spatially varying potential systematics in equatorial coordinates with Mollweide projection and astronomy convention (east towards left).

to magnitude errors (for instance, airmass induces atmospheric differential refraction as a function of colour, which can affect the photometry) and is entangled with the extinction and seeing corrections.

(v) *Sky brightness*: Variations in the background brightness of the sky (due to various sources such as airglow, scattered starlight, Moon phase, and light pollution) can affect the measured flux errors and therefore the density of targets by scattering objects in or out of





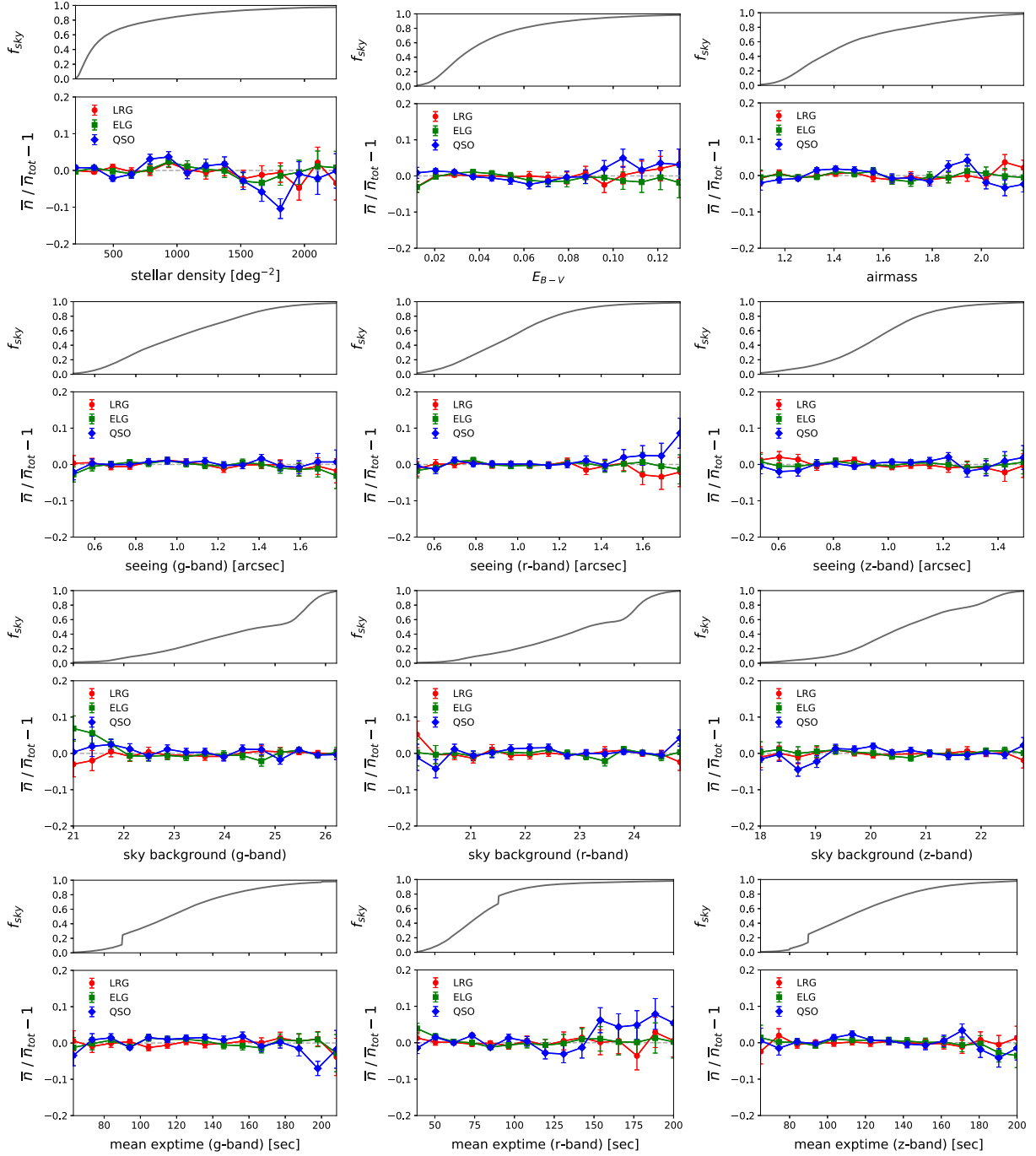
**Figure 13.** Systematic dependences before applying photometric weights: Mean density fluctuations for LRGs (red circles), ELGs (green squares), and QSOs (blue diamonds), as a function of (from top left to bottom right) mean stellar density, colour excess  $E_B - V$ , airmass, seeing, sky background, and exposure time in each band, with Poisson errors. The top panel in each figure is the cumulative sky fraction for each systematic. We generally find that splitting between NGC and SGC has negligible effects on the 1D density trends, with the exception of  $E_B - V$  for ELGs and stellar density for QSOs; for these two cases, we have added the NGC-only (dashed) and SGC-only (dotted) trend lines.

colour cuts. Sky brightness also has a strong dependence on airmass, which increases the brightness of airglow. We include the mean sky background in each band measured on the individual CCDs as a potential systematic.

(vi) Exposure time DECam can attain the depths required for DESI targeting in total exposure times of 166, 134, and 200 s

for  $g$ ,  $r$ , and  $z$  bands, given median observing conditions. As part of the imaging strategy, dynamic exposure times are increased to compensate for poor observing conditions in order to obtain a more uniform sample. We look at variations with mean exposure time in each band, which affects depth and is correlated with other potential systematics, to see how it modulates the observed density.





**Figure 14.** Systematic dependences after applying photometric weights: Mean density fluctuations for LRGs (red circles), ELGs (green squares), and QSOs (blue diamonds), as a function of (from top left to bottom right) mean stellar density, colour excess  $E_B - V$ , airmass, seeing, sky background, and exposure time in each band, with Poisson errors. The top panel in each figure is the cumulative sky fraction for each systematic.

## 6.2 Target densities versus potential systematics

We can determine the post-masking target density per pixel using the random catalogues. Since the randoms are uniformly distributed, counting the number of post-masking randoms in a pixel is equivalent to measuring its effective area, up to a proportionality factor:

$$\delta_i = n_i^{\text{gal}} / \bar{n}^{\text{gal}} - 1 = N_i^{\text{gal}} / N_i^{\text{ran}} \times N_{\text{masked}}^{\text{ran}} / N_{\text{masked}}^{\text{gal}} - 1. \quad (21)$$

For each potential systematic, we bin the pixels by systematic value and then plot the average density versus the average systematic

value of the bins. The results are shown in Fig. 13, with LRGs, ELGs, and QSOs plotted together in each subplot and cumulative sky fraction displayed in the upper panels. The errors bars represent the Poisson noise in each bin; using standard error of the mean gives minuscule error bars, as the variance within each bin is very small, regardless of the exact bin size or pixel resolution used. In general, LRGs show very little dependence on survey properties, while ELGs and QSOs appear more impacted, with QSOs often displaying a non-linear dependence likely due to the more complex selection function. We find that the NGC and SGC density trends



are similar and thus do not need to be plotted separately, with the exception of  $E_B - V$  for ELGs and stellar density for QSOs. For these two special cases, we include the NGC-only (dashed) and SGC-only (dotted) trend lines in Fig. 13.

The most significant systematic effects indicated are from stellar density and extinction. ELG density decreases significantly with increasing stellar density and extinction, with 10 percent level effects in some areas, while QSO density presents the opposite trend. The observed correlation between stellar density and QSO density, and anticorrelation between stellar density and ELG density, is also present in the angular cross-correlation results (Section 6.3). One possible explanation for the extinction dependence is the issue of infrared emission from background galaxies contaminating the SFD dust maps used to correct DECaLS magnitudes. This has been shown to lead to underestimation of the reddening in low extinction ( $E_B - V < 0.15$ ) regions (Yahata et al. 2007; Kashiwagi, Yahata & Suto 2013; Kashiwagi et al. 2015), attributed to the fact that the extragalactic contamination dominates the dust signal in such regions. Looking at the shape of the targets selection functions in colour space (Fig. 11), underestimation of reddening could preferentially scatter objects out of the ELG selection or into the QSO selection. While the resultant underestimation of extinction in these regions is small, it may be highly correlated with the targets.

Some minor dependence on seeing and sky background are also observable, particularly for QSOs, as it is more difficult to distinguish between QSOs and stars in regions with bad seeing and bright sky backgrounds. The relationship between target densities and mean exposure times is more complex. Due to the use of dynamic exposure times, there is entanglement with other systematics; for example, exposures are scaled longer for higher airmass or regions of higher galactic extinction. CCDs with exposure times less than 30 s are automatically removed in the image reduction pipeline, and the ‘jumps’ or discontinuities in sky fraction at various other times are artefacts of the observing strategy. The fact that we are averaging the exposure times over multiple overlapping CCDs slightly muddies the interpretation as well. Nevertheless, it is clear that the attempt to obtain more uniform depths through dynamic exposure times is not perfectly successful for QSOs.

Based on these findings, we create photometric weights to reduce the variance in target densities due to systematics. Working directly with the HEALPix pixels defined in the previous section, we use principal component analysis (PCA) to transform the list of potential systematics into a minimum set of linearly uncorrelated variables. PCA using a full SVD solver reduces the dimensionality from 12 scaled features to 11 components. The first component explains  $\sim 22$  percent of the variance and the last component explains  $\sim 3$  percent of the variance, with  $\sim 50$  percent of the variance explained by the first three components and  $\sim 75$  percent of the variance explained by the first six components.

Unsurprisingly, we find that exposure times contain a great deal of information, as they are correlated with all of the other systematics by design. However, since exposure times are difficult to interpret, we also perform a version of the component analysis with the exposure times removed from consideration, in order to show more clearly how the other features contribute and in what combinations. We find that nearly equal contributions from stellar density and galactic extinction tend to strongly dominate a few components, while more complex mixtures of sky background, seeing, and airmass features dominate the others, as physical intuition might lead us to expect.

We discretize the feature-space to reduce the impact of noisy pixels and outliers and then apply multilinear regression. The

resulting model of density as a function of potential systematics is used to generate weights.<sup>14</sup> We apply our photometric weights to the randoms, modulating the effective area (and therefore target density) of the survey based on the local values of systematics. Weights are normalized in the sense that the proportionality factor in equation (21) changes from the number of randoms in the masked footprint to the sum of their weights. Plots of the density versus systematics after applying weights are shown in Fig. 14, with the linear parts of the trends improved. Fig. 15 demonstrates the effect of applying these weights on the angular correlation functions. All clustering results presented in Section 7 and Section 8 are computed using the weighted values.

### 6.3 Correlation with stars and stellar contamination fraction

We also measure the angular cross-correlation between the targets (after masking but before weighting) and our stellar catalogue, with the results shown in Fig. 16. Consistent with the density trends observed in Fig. 13, LRGs appear uncorrelated with stars, while ELGs demonstrate a small constant anticorrelation, and QSOs show a more significant constant correlation.

Using the angular cross-correlation, we can estimate the fraction of stellar contamination in the QSO sample. Let us assume that the observed number of QSOs at any given location includes some non-trivial number of contaminants, as seems strongly indicated. Let  $N_{\text{star}}$  be the total number of stars that modulate the QSO density in some way:

$$N_{\text{obs}} = N_{\text{true}} + \bar{\epsilon} N_{\text{star}}, \quad (22)$$

where  $\bar{\epsilon}$  is the average number of impacted sources associated with each star. For stars that are simply misclassified as QSOs,  $\epsilon = 1$ . For spurious QSO detections in the immediate vicinity of stars,  $\epsilon > 0$ . For occulted QSOs in the immediate vicinity of stars,  $\epsilon < 0$ . Note that for the latter two effects (spurious or occulted sources near stars), we are considering fainter stars that were not masked out in Section 4.2.1. Thus, any cross-correlations between these sources and their own associated star are negligible beyond very small scales, so the cross-correlation between contaminants and stars is dominated by the autocorrelation of stars times the multiplicative factor  $\epsilon$ .

The fraction of the total sample that is made up of these problematic stars is approximately given by

$$f_{\text{star}} = \frac{\langle N_{\text{star}} \rangle}{\langle N_{\text{obs}} \rangle}, \quad (23)$$

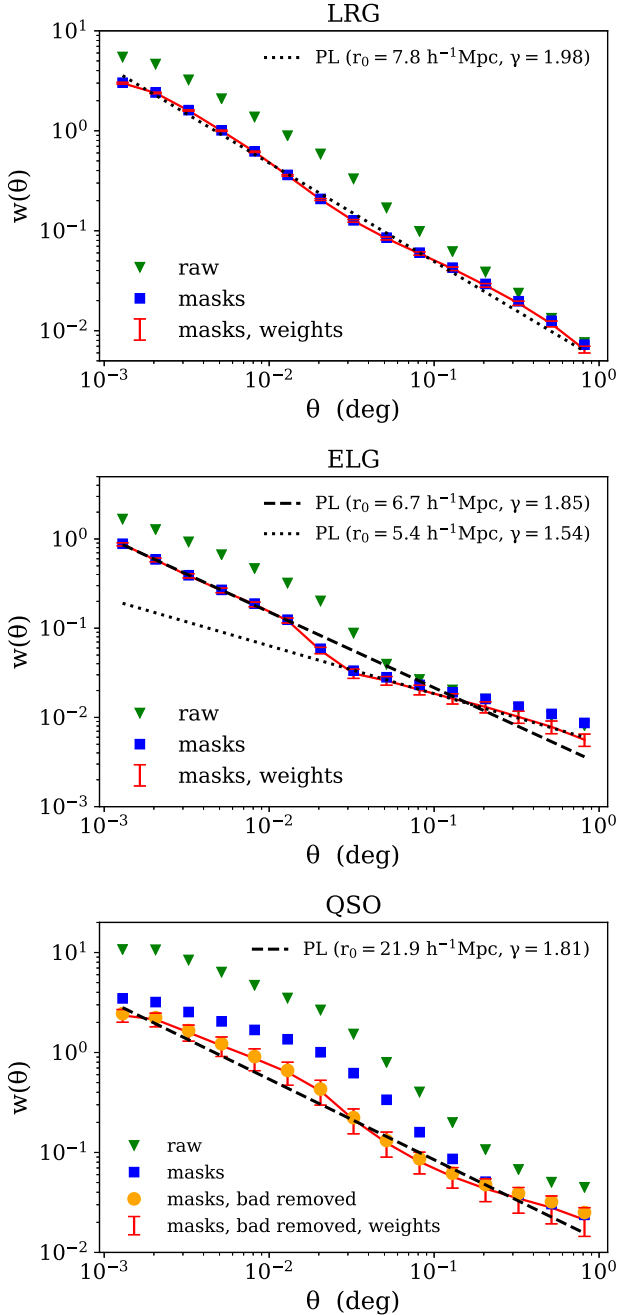
where the brackets signify a spatial average. Similarly, the fraction of true objects is

$$f_{\text{true}} = \frac{\langle N_{\text{true}} \rangle}{\langle N_{\text{obs}} \rangle}. \quad (24)$$

We can rewrite this in terms of the density contrasts  $\delta = N/\langle N \rangle - 1$  by exploiting the fact that  $\langle N_{\text{obs}} \rangle = \langle N_{\text{star}} \rangle / f_{\text{star}} = \langle N_{\text{true}} \rangle / f_{\text{true}}$ , such

<sup>14</sup>For randoms where any of the potential systematics were undefined due to lack of exposures in one or more bands, the weights were manually set to one. The randoms used in our clustering analysis have had the completeness mask of Section 4.1 applied and thus are not affected by this.

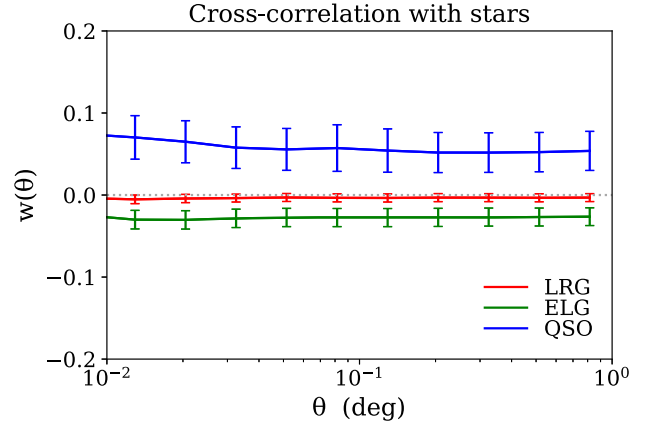




**Figure 15.** Two-point angular correlation functions for LRGs, ELGs, and QSOs at several levels of systematics analysis: without any corrections (green triangles), after applying masks (blue squares), and after applying both masks and photometric weights (red error bars). For QSOs, we perform the additional intermediate step of removing the anomalous regions discussed in Section 5 (orange circles). The red lines are fit to a power-law model for the three-dimensional clustering  $\xi(r) = (r/r_0)^{-\gamma}$ . The values of these fits are listed in more detail in Table 6.

that the observed density of objects is

$$\begin{aligned}
 \delta_{\text{obs}} &= \frac{N_{\text{obs}}}{\langle N_{\text{obs}} \rangle} - 1 = \frac{N_{\text{true}}}{\langle N_{\text{obs}} \rangle} + \bar{\epsilon} \frac{N_{\text{star}}}{\langle N_{\text{obs}} \rangle} - 1 \\
 &= f_{\text{true}} \frac{N_{\text{true}}}{\langle N_{\text{true}} \rangle} + \bar{\epsilon} f_{\text{star}} \frac{N_{\text{star}}}{\langle N_{\text{star}} \rangle} - 1 \\
 &= f_{\text{true}}(\delta_{\text{true}} + 1) + \bar{\epsilon} f_{\text{star}}(\delta_{\text{star}} + 1) - 1 \\
 &= f_{\text{true}}\delta_{\text{true}} + \bar{\epsilon} f_{\text{star}}\delta_{\text{star}}.
 \end{aligned} \tag{25}$$



**Figure 16.** Angular cross-correlation between DESI targets and stars, with errors from bootstrapping on the area.

**Table 5.** Average densities for each target type, calculated over the available footprint. The first column is the uncorrected densities, the second column has had only the masks of Section 4 applied, and the third column has additionally had the photometric weights of Section 6.2 applied. The projected target densities from the FDR are also included for reference.

Target	Average density (deg <sup>-2</sup> )			FDR
	Raw	Masked	Masked and weighted	
LRG	525.0	501.4	498.9	480
ELG	2497.5	2351.8	2352.7	2400
QSO	336.1	261.2	256.2	260

Thus, we can extract the contamination fraction by dividing the QSO-star cross-correlation by the stellar autocorrelation function,  $w_{\text{cross}}(\theta)/w_{\text{star}}(\theta) = \langle \delta_{\text{obs}} \delta_{\text{star}} \rangle / \langle \delta_{\text{star}} \delta_{\text{star}} \rangle = \bar{\epsilon} f_{\text{star}} \equiv f_{\text{contam}}$ , since the cross-correlation between true QSOs and stars vanishes.<sup>15</sup>

As the stellar density varies significantly across the sky (see Fig. 12), with a strong gradient towards the galactic plane, we first divide the sky into three bins:  $|b| < 40$ ,  $40 < |b| < 60$ , and  $|b| > 60$ . For each galactic latitude bin, we calculate  $w_{\text{cross}}(\theta)$  and  $w_{\text{star}}(\theta)$ , averaged across all angular scales (as both correlation functions are flat, within error bars, for all bins), and then bootstrap upon these averaged values to obtain error bars. The resulting stellar contamination fractions are  $f_{\text{contam}} = 7$  per cent  $\pm 4.9$  per cent for  $|b| < 40$ ,  $f_{\text{contam}} = 4.9$  per cent  $\pm 2.7$  per cent for  $40 < |b| < 60$ , and  $f_{\text{contam}} = 4.1$  per cent  $\pm 2.3$  per cent for  $|b| > 60$ .

## 7 ANGULAR CLUSTERING MEASUREMENTS

### 7.1 Mean densities

The average target densities for DR7 are given in Table 5. We present the raw densities as well as the densities after observational effects have been accounted for using the masks and weights described in this paper. Here, densities are calculated by taking the ratio of the total number of objects and the total area, with the latter being estimated using counts of uniform randoms with the masks and weights applied to them. For an independent calculation of raw

<sup>15</sup>We have assumed an ideal stellar template; in reality, there may be a small fraction of true QSOs in the star sample, or a fraction of galaxies in both the QSO and the star samples, which correlate with each other, but these fractions should be much smaller than the fraction of stellar contaminants in the QSO sample, and hence we can ignore them to first order.



**Table 6.** Results from fitting the correlation functions to a power-law  $\xi(r) = (r/r_0)^{-\gamma}$ , where we have used the expected redshift distributions of DESI targets in the Limber approximation to convert between angular clustering and real-space clustering. We list the angular scales fitted over, the expected  $\bar{z}$  of each sample, and the best-fitting parameter values and errors.

Target	$\theta_{\min}, \theta_{\max}$	$r_0$ ( $h^{-1}\text{Mpc}$ )	$\gamma$	$\log_{10} A_w$
LRG	$0.001^\circ, 1^\circ$	$7.78 \pm 0.26$	$1.98 \pm 0.02$	$-4.01 \pm 0.05$
ELG	$0.001^\circ, 1^\circ$	$5.98 \pm 0.30$	$1.90 \pm 0.03$	$-4.25 \pm 0.06$
	$0.001^\circ, 0.01^\circ$	$6.70 \pm 0.10$	$1.85 \pm 0.01$	$-4.01 \pm 0.02$
QSO	$0.05^\circ, 1^\circ$	$5.45 \pm 0.10$	$1.54 \pm 0.01$	$-3.22 \pm 0.02$
	$0.001^\circ, 1^\circ$	$21.9 \pm 1.01$	$1.81 \pm 0.03$	$-3.31 \pm 0.07$

target densities, column 2 of the counts-in-cells tables in Section 7.4 gives the average number of objects  $\bar{N}$  within a square cell of some width and thus can be divided by the corresponding cell area to give the mean (raw) density smoothed over that scale.

## 7.2 Angular correlation functions with $r_0, \gamma$ fits

The measured angular correlation functions for the three target classes are shown in Fig. 15. We present the correlation functions at various stages of analysis to demonstrate the effects of applying masks, weights, and so on. The final, ‘cleanest’ version is fit to theory, as described below.

According to the current paradigm of galaxy formation, galaxies form within collapsed overdensities of dark matter called ‘haloes’ (for a recent review on the galaxy-halo connection, see Wechsler & Tinker 2018). Under this model, the correlation function of galaxies is the sum of two contributions: a 2-halo term corresponding to pairs of galaxies within different haloes, and a 1-halo term corresponding to pairs of galaxies within the same halo. On small scales, where the 1-halo term dominates, the correlation function depends on the complex baryonic physics of galaxy formation and evolution, while on larger scales, where the 2-halo term dominates, the correlation function is characterized by the halo bias describing how dark matter haloes trace the dark matter distribution. When combined, these two terms result in an approximate power law, with a feature corresponding to the 1-halo to 2-halo transition occurring around  $1\text{--}2 h^{-1} \text{Mpc}$ , the typical virial radius of a halo. Motivated by this, we assume the real-space correlation function is a simple power law of the form  $\xi(r) = (r/r_0)^{-\gamma}$ . Using tabulated  $dN/dz$  for each target and applying the Limber approximation (equation 5), we obtain constraints on  $r_0$  and  $\gamma$ , listed in Table 6.

For LRGs, we determine  $r_0 = 7.78 \pm 0.26 h^{-1}\text{Mpc}$  and  $\gamma = 1.98 \pm 0.02$ , which agrees well with previous results for similar samples from the literature; for example, Sawangwit et al. (2011) (Table 2, row 4) finds  $r_0 = 7.56 \pm 0.03 h^{-1}\text{Mpc}$  and  $\gamma = 1.96 \pm 0.01$  for a photometric subsample of LRGs from SDSS imaging with  $\bar{z} = 0.68$  and a similar redshift distribution, over approximately the same range of angular scales. While the LRG correlation function shows some additional structure that is not fit perfectly by a power-law model, no strong features are observed on these scales, which is generally consistent with earlier findings from eBOSS and SDSS LRG studies (see e.g. Zehavi et al. 2005).

We find that the ELG correlation function has a broken form. When fitting from  $\theta = 0.001^\circ$  to  $\theta = 0.01^\circ$ , the correlation function is well fit by  $r_0 = 6.70 \pm 0.10 h^{-1}\text{Mpc}$  and  $\gamma = 1.85 \pm 0.01$ . However, for scales  $\theta > 0.05^\circ$ , the slope becomes shallower, and the correlation function is better fit by  $r_0 = 5.45 \pm 0.10 h^{-1}\text{Mpc}$  and  $\gamma = 1.54 \pm 0.01$ . At the mean effective redshift of the DESI ELG sample,  $z \sim 0.85$ , the co-moving scale of this break

is approximately  $1 h^{-1}\text{Mpc}$ , consistent with a 1-halo to 2-halo transition. The second slope matches with the findings of Favole et al. (2016), who modelled a sample of ELGs selected from the Canada-France Hawaii Telescope Legacy Survey (CFHTLS), cross-matched with BOSS ELG and VIPERS redshifts, at mean redshift  $\bar{z} \approx 0.8$ , to obtain  $s_0 = 5.3 \pm 0.2 h^{-1}\text{Mpc}$  and  $\gamma = 1.6 \pm 0.1$ . Furthermore, when calculating the angular correlation function over the full CFHTLS footprint, they also observed a change of slope occurring at  $\theta \approx 0.01^\circ\text{--}0.05^\circ$  and found that this clustering was consistent with an HOD model having halo masses on the order of  $10^{12} M_\odot$  and satellite fraction  $f_{\text{sat}} \sim 22$  per cent. Similarly, Jouvel et al. (2015) found  $s_0 = 4.2 \pm 0.26 h^{-1}\text{Mpc}$  and  $\gamma = 1.48 \pm 0.04$  for a bright sample of eBOSS ELGs selected from DES photometry at  $\bar{z} = 0.86$ . The real-space clustering amplitudes and slopes for both LRGs and ELGs are also consistent with the results from Mostek et al. (2013) for red and blue galaxy populations in DEEP2.

The QSO correlation function still contains a significantly enhanced clustering signal due to systematics and contamination, with  $r_0 = 21.9 \pm 0.10 h^{-1}\text{Mpc}$  and  $\gamma = 1.81 \pm 0.02$ . By comparison, some fiducial values of QSO clustering amplitude at  $z \sim 2$  from the literature are:  $r_0 = 5.84 \pm 0.33$  and  $\gamma = 1.65 \pm 0.05$  (Croom et al. 2005), or  $r_0 = 4.56 \pm 0.48$  at fixed  $\gamma = 1.5$  (Myers, White & Ball 2009).

## 7.3 Angular power spectra with $b_0, b(k)$ fits

We measure the angular power spectra of the three main DESI target samples using the methods described in Section 2.4. We reiterate that the results presented here have already had the masks and weights of the previous sections applied. Similar to their effect on the angular correlation functions, the impact of the photometric weights derived in Section 6.2 on the angular power spectra is to reduce power on large scales. For LRGs, not using the weights would increase the amplitude by  $\sim 15$  per cent at  $\ell \sim 20$  down to  $\sim 1$  per cent at  $\ell \sim 75$ . For ELGs, it would increase by  $\sim 43$  per cent at  $\ell \sim 20$  down to  $\sim 1$  per cent at  $\ell \sim 150$ . For QSOs, it would increase by  $\sim 12$  per cent at  $\ell \sim 20$  down to  $< 1$  per cent at  $\ell \sim 75$ .

From the angular power spectra, we fit the linear bias. First, we restrict to very large scales where the bias is approximately constant and then relax this restriction as we probe the scale dependence of the bias using the ‘ $P$  model’ (Smith, Scoccimarro & Sheth 2007; Hamann et al. 2008; Cresswell & Percival 2009), which treats the non-linear correction to the bias as an extra non-Poissonian shot noise term arising from the assumption that galaxies populate haloes (Seljak 2001; Schulz & White 2006; Guzik, Bernstein & Smith 2007):

$$P_g \rightarrow P_g + P \Rightarrow b(z)^2 = \frac{b_0^2}{D(z)^2} \rightarrow b(k, z)^2 = \frac{b_0^2}{D(z)^2} \left( 1 + \frac{P}{b_0^2 P_m(k, z)} \right). \quad (26)$$

In terms of the angular power spectra, which involve convolution with the radial distributions (see equation 13), we have

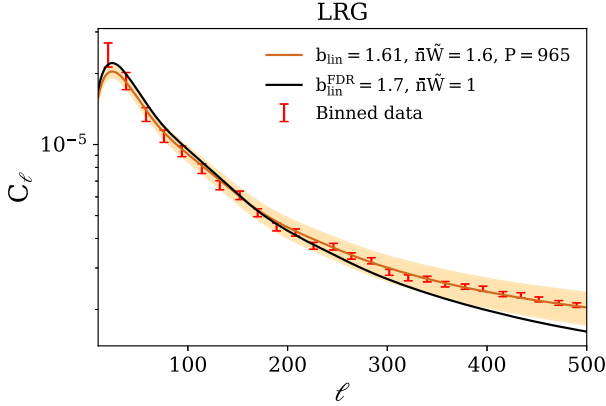
$$C_\ell \rightarrow C_\ell + C \\ C = P \int d\chi f(\chi)^2 \frac{1}{\chi^2} \frac{1}{D(z)^2}. \quad (27)$$

Using CLASS, we compare  $C_\ell$  derived from linear and HALOFIT (Smith et al. 2003) predictions of the matter power spectrum to estimate  $\ell_{\text{max}}$  where they begin to diverge, taking the common assumption that the non-linear correction to the matter power spectrum becomes significant at approximately the same scale as



**Table 7.** Fitting the LRG large-scale bias from the angular power spectra. We initially limit to  $\ell \leq 200$  and fit the linear bias to a constant: first, by fixing the Poisson shot noise term as  $1/\bar{n}$ , and second, by fitting the bias and noise simultaneously. Then we extend to  $\ell = 500$  and fit an additional parameter  $P$  for the scale-dependent bias model: first, by holding the previously found values fixed and fitting only the non-Poisson shot noise term  $P$  at larger  $\ell$ , and second, by fitting the linear bias and  $P$  simultaneously, with  $P$  now absorbing both shot noise terms.

Target	$b_0^{\text{FDR}}$	$\ell_{\text{max}} = 200$ Scale-independent bias			$\ell_{\text{max}} = 500$ $P$ model			
		$b_0$	$\bar{n}\tilde{W}$	$\chi^2_{\text{d.o.f}}$	$b_0$	$\bar{n}\tilde{W}$	$P$	$\chi^2_{\text{d.o.f}}$
LRG	1.7	$1.570 \pm 0.014$	1 (fixed)	5.0 / 9	–	–	$3539 \pm 99$	12.6 / 25
		$1.607 \pm 0.040$	$1.61 \pm 0.43$	3.7 / 8	–	–	$965 \pm 99$	18 / 25
					$1.569 \pm 0.017$	0 (fixed)	$7120 \pm 198$	12.6 / 24



**Figure 17.** Angular power spectrum  $C_\ell$  for LRGs. The red error bars are the binned errors from equation (12). The solid black line is the theoretical curve using the FDR value for the linear bias. The orange envelope is our best-fitting  $P$ -model result, with the model errors dominated by uncertainty in the shot noise terms.

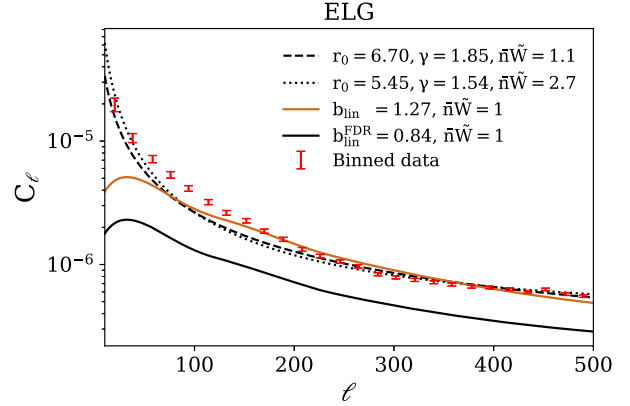
non-linear effects in the galaxy bias (see e.g. Fry & Gaztanaga 1993; Modi, White & Vlah 2017; Desjacques, Jeong & Schmidt 2018; Wilson & White 2019). We find that  $\ell_{\text{max}} \approx 200$  is appropriate for LRGs and ELGs, and slightly conservative for QSOs, which are at higher mean redshift.

### 7.3.1 LRGs

With  $\ell_{\text{max}} = 200$ , we use the linear theory matter power spectrum and fit the scale-independent bias in two ways: first, by fixing the Poisson shot noise term  $\tilde{W} = 1/\bar{n}$  and only fitting  $b_0$ , and second, by simultaneously fitting  $b_0$  and  $\tilde{W}$ . Then, we extend out to  $\ell_{\text{max}} = 500$  and add the additional  $P$  parameter to our model. We fit  $P$  in several ways: both using the previously found values of  $b_0$  and  $\tilde{W}$  from the  $\ell_{\text{max}} = 200$  case and also doing a simultaneous fit to  $b_0$  and  $P$ , with  $\tilde{W}$  absorbed into  $P$ . The results of all fits are shown in Table 7 for LRGs, with all models giving similar results and showing agreement with expectation. In Fig. 17, we plot the binned data, the best-fitting model, and the FDR expectation curve.

### 7.3.2 ELGs

For ELGs, we find that attempting to co-fit the bias and shot noise terms simultaneously returns unphysical negative values for the latter, due to enhanced power at scales  $\ell < 150$  even after applying masks and weights (fortunately, these scales should not directly impact BAO and RSD measurements). However, when fixing the



**Figure 18.** Angular power spectrum  $C_\ell$  for ELGs. The red error bars are the binned errors from equation (12). The solid black line is the theoretical curve using the FDR value for the linear bias. The orange envelope<sup>17</sup> is our best fit for the linear bias, using fixed shot noise  $\tilde{W} = 1/\bar{n}$ . The dashed and dotted lines are the angular power spectra corresponding to the power-law fits to the angular clustering determined in Section 7.2, with the shot noise terms fit as extra free parameters.

shot noise as  $\tilde{W} = 1/\bar{n}$ , we obtain  $b_0 = 1.273 \pm 0.005$ , which agrees well with, e.g. Comparat et al. (2013) and Delubac et al. (2017).

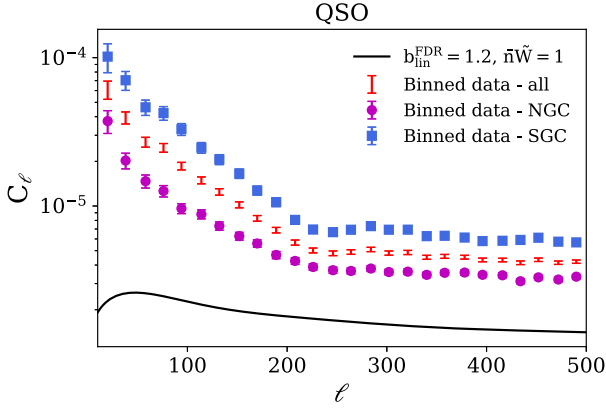
We also calculate the corresponding  $C_\ell$  for each of the two power-law model  $w(\theta)$  fits in Fig. 15 and plot these as well in Fig. 18, with shot-noise contributions fitted as additional free parameters. The results show consistency between our  $w(\theta)$  and  $C_\ell$  results, both of which give clustering parameters falling within the range of fiducial values found in previous studies.

The DESI FDR assumes a conservative lower limit of  $b_0 = 0.84$ , also plotted in Fig. 18, and we confirm that the ELG clustering bias is higher than this value. This is significant as it has the effect of improving the statistical errors on BAO, while also somewhat degrading the RSD forecasts, since more strongly biased tracers exhibit weaker anisotropy. We note that allowing the shot noise term to float in the FDR curve in order to raise its amplitude still results in a very poor fit, as it flattens the curve such that it can achieve only artificial agreement with observation at very large  $\ell$ .

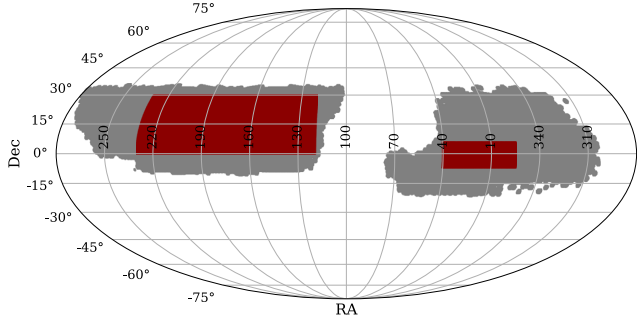
### 7.3.3 QSOs

For QSOs, the angular power spectrum, like the angular correlation function, is significantly inflated with non-cosmological signals. As such, we do not report fitted values but merely plot the results in Fig. 19 alongside FDR expectation to demonstrate the discrepancy. Furthermore, whereas the LRGs and ELGs show no difference when comparing NGC and SGC measurements, QSOs





**Figure 19.** Angular power spectrum  $C_\ell$  for QSOs. The red error bars are the binned errors from equation (12). The magenta circles and blue squares are for measurements restricted to the NGC and the SGC, respectively. The solid black line is the theoretical curve using the FDR value for the linear bias and fixed shot noise  $\tilde{W} = 1/\bar{n}$ .



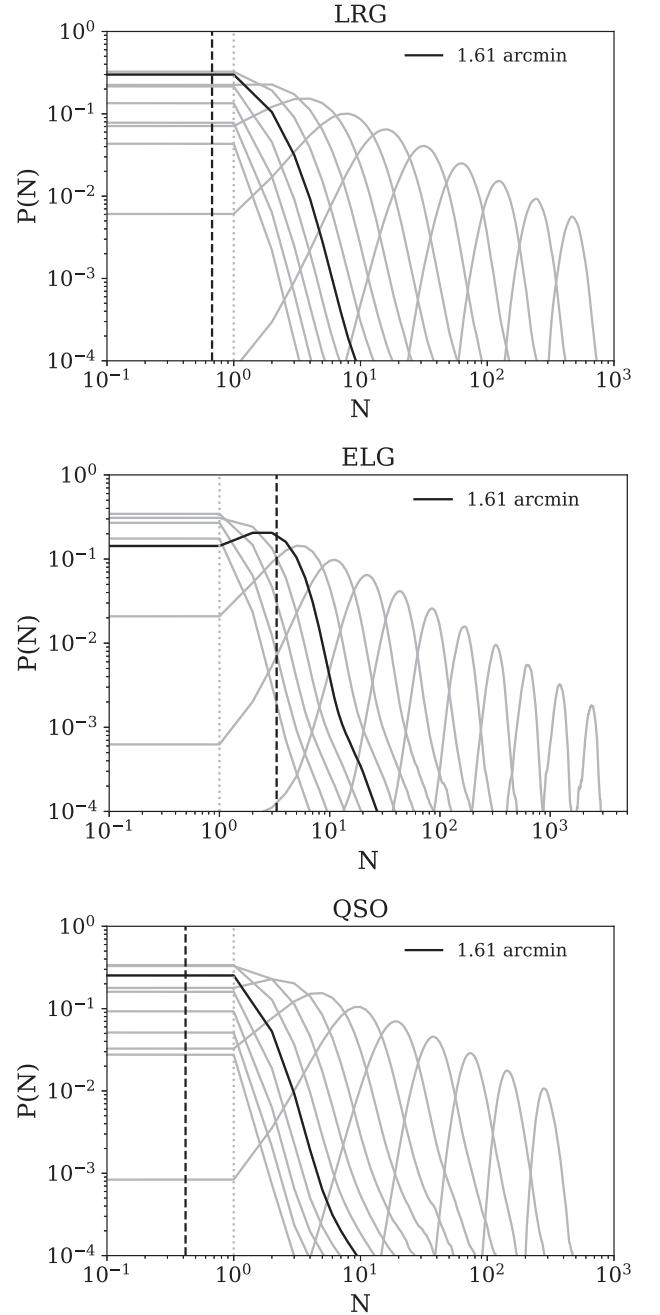
**Figure 20.** Two patches over which we calculate the counts-in-cells distributions and moments. These were chosen by eye as survey regions with no tattered edges and relatively few holes.

once again show a mismatch between galactic hemispheres; after removing the problematic pixels found in Section 5, which caused disproportionately strong small-scale clustering in the NGC, the shapes of the angular power spectra in the NGC and the SGC become identical, but the SGC is enhanced on all scales compared to the NGC. This is consistent with our other findings, namely that the SGC has more stellar contamination than the NGC and that the clustering of these stellar contaminants is relatively flat across scales (Section 6.3).

#### 7.4 Counts-in-cells moments

As discussed in Section 2.6, detailed small-scale clustering information is invaluable for accurate modelling and mock calibration.

We calculate the counts-in-cells statistics over two large fields, one in each galactic hemisphere, with effective areas  $S_{\text{eff}, N} = 3300 \text{ deg}^2$  and  $S_{\text{eff}, S} = 562.5 \text{ deg}^2$  (Fig. 20). We select these fields to be regular in shape and relatively smooth, avoiding areas that are tattered or full of holes. We calculate the probability distribution  $P(N)$  for each field and then measure the weighted average and



**Figure 21.**  $P(N)$  versus  $N$  for 15 logarithmically spaced cell widths from  $\theta = 0.01^\circ$  to  $\theta = 1^\circ$ . The highlighted cell has width  $0.268^\circ \approx 1.61 \text{ arcmin}$ , close to the fibre patrol radius of  $1.4 \text{ arcmin}$ , and the dashed vertical lines correspond to the average target density times the cell area. The dotted vertical line at  $N = 1$  marks the limit where shot noise dominates ( $\tilde{N} < 1$ ).

standard error of the factorial moments  $F_p$  (Wolk et al. 2013):

$$\bar{F}_p = \frac{\sum_{i=N,S} S_{\text{eff},i} F_{p,i}}{\sum_{i=N,S} S_{\text{eff},i}}$$

$$(\Delta F_p)^2 = \frac{\sum_{i=N,S} S_{\text{eff},i} (F_{p,i} - \bar{F}_p)^2}{\sum_{i=N,S} S_{\text{eff},i}}.$$



**Table 8.** Mean, variance, skewness, kurtosis, and cell-averaged  $n$ -point angular correlation functions for each of the three main DESI target classes measured in square cells.

	Width	$\bar{N}$	$\sigma^2$	$\frac{\langle(N-\bar{N})^3\rangle}{\sigma^3}$	$\frac{\langle(N-\bar{N})^4\rangle}{\sigma^4}$	$\bar{w}_2$	$\bar{w}_3$	$\bar{w}_4$
LRG	0.010°	0.0490 ± 0.0019	0.0550 ± 0.0023	7.50 ± 0.23	260 ± 35	2.410 ± 0.014	250 ± 18	88000 ± 8100
	0.014°	0.0950 ± 0.0036	0.1120 ± 0.0047	7.30 ± 0.35	340 ± 49	1.870 ± 0.022	146.0 ± 9.4	40000 ± 3400
	0.019°	0.1840 ± 0.0069	0.2320 ± 0.0097	7.50 ± 0.45	410 ± 58	1.420 ± 0.025	81.0 ± 4.3	16000 ± 1200
	0.027°	0.350 ± 0.013	0.490 ± 0.020	8.10 ± 0.48	470 ± 55	1.060 ± 0.025	45.0 ± 1.5	6200 ± 300
	0.037°	0.680 ± 0.026	1.050 ± 0.042	9.10 ± 0.45	550 ± 44	0.780 ± 0.024	24.90 ± 0.21	2520 ± 21
	0.052°	1.320 ± 0.050	2.290 ± 0.086	10.00 ± 0.38	620 ± 29	0.560 ± 0.020	13.10 ± 0.15	990 ± 26
	0.072°	2.550 ± 0.096	5.10 ± 0.18	10.10 ± 0.22	617.0 ± 4.3	0.390 ± 0.015	6.30 ± 0.21	360 ± 28
	0.100°	4.90 ± 0.19	11.40 ± 0.36	9.200 ± 0.090	500 ± 22	0.270 ± 0.012	2.80 ± 0.13	110 ± 12
	0.139°	9.50 ± 0.36	26.00 ± 0.75	7.550 ± 0.016	330 ± 20	0.1820 ± 0.0086	1.090 ± 0.066	27.0 ± 3.5
	0.193°	18.40 ± 0.70	61.0 ± 1.6	5.710 ± 0.016	190 ± 12	0.1260 ± 0.0063	0.410 ± 0.028	5.80 ± 0.83
	0.268°	35.0 ± 1.3	145.0 ± 3.2	4.070 ± 0.049	95.0 ± 6.9	0.0880 ± 0.0047	0.150 ± 0.012	1.20 ± 0.19
	0.373°	69.0 ± 2.6	359.0 ± 6.0	2.790 ± 0.093	45.0 ± 4.2	0.0620 ± 0.0036	0.0560 ± 0.0060	0.250 ± 0.044
	0.518°	132.0 ± 5.1	906.0 ± 8.7	1.90 ± 0.13	21.0 ± 2.6	0.0440 ± 0.0029	0.0210 ± 0.0031	0.050 ± 0.011
	0.720°	256.0 ± 9.8	2338.0 ± 5.3	1.30 ± 0.14	11.0 ± 1.6	0.0320 ± 0.0024	0.0080 ± 0.0016	0.0100 ± 0.0028
	1.000°	490 ± 19	6110 ± 99	0.90 ± 0.13	6.0 ± 1.1	0.0230 ± 0.0020	0.00360 ± 0.00083	0.00230 ± 0.00080
ELG	0.010°	0.2350 ± 0.0030	0.290 ± 0.010	5.0 ± 1.2	90 ± 47	1.10 ± 0.27	30 ± 15	2000 ± 1100
	0.014°	0.4530 ± 0.0058	0.630 ± 0.034	5.0 ± 1.5	110 ± 56	0.90 ± 0.21	19.0 ± 9.4	800 ± 520
	0.019°	0.870 ± 0.011	1.40 ± 0.10	6.0 ± 1.8	120 ± 61	0.70 ± 0.16	12.0 ± 5.7	400 ± 230
	0.027°	1.690 ± 0.022	3.20 ± 0.30	7.0 ± 2.0	140 ± 67	0.50 ± 0.12	7.0 ± 3.4	200 ± 100
	0.037°	3.260 ± 0.042	7.30 ± 0.79	7.0 ± 2.2	160 ± 75	0.380 ± 0.088	4.0 ± 1.9	80 ± 49
	0.052°	6.290 ± 0.080	17.0 ± 1.9	7.0 ± 2.2	150 ± 72	0.260 ± 0.055	1.90 ± 0.92	30 ± 18
	0.072°	12.10 ± 0.15	37.0 ± 3.7	6.0 ± 1.8	110 ± 54	0.170 ± 0.030	0.80 ± 0.35	7.0 ± 4.6
	0.100°	23.40 ± 0.30	83.0 ± 5.6	5.0 ± 1.2	70 ± 30	0.110 ± 0.013	0.30 ± 0.11	1.60 ± 0.92
	0.139°	45.30 ± 0.57	192.0 ± 3.3	4.00 ± 0.50	50 ± 11	0.0720 ± 0.0036	0.110 ± 0.021	0.40 ± 0.12
	0.193°	87.0 ± 1.1	460 ± 24	3.40 ± 0.25	34.0 ± 3.4	0.0490 ± 0.0018	0.0480 ± 0.0058	0.110 ± 0.019
	0.268°	169.0 ± 2.1	1200 ± 170	3.0 ± 1.0	30 ± 15	0.0350 ± 0.0049	0.030 ± 0.013	0.050 ± 0.038
	0.373°	326.0 ± 3.9	3100 ± 760	3.0 ± 1.7	30 ± 24	0.0260 ± 0.0065	0.020 ± 0.014	0.030 ± 0.034
	0.518°	628.0 ± 7.3	9000 ± 3000	3.0 ± 2.1	30 ± 28	0.0200 ± 0.0071	0.010 ± 0.013	0.020 ± 0.027
	0.720°	1210 ± 14	30000 ± 11000	3.0 ± 2.3	30 ± 29	0.0170 ± 0.0071	0.010 ± 0.010	0.020 ± 0.020
	1.000°	2340 ± 25	80000 ± 39000	2.0 ± 2.2	20 ± 25	0.0140 ± 0.0068	0.0070 ± 0.0079	0.010 ± 0.012
QSO	0.010°	0.02990 ± 0.00011	0.03700 ± 0.00054	18.0 ± 1.6	1500 ± 290	7.90 ± 0.42	2900 ± 440	1900000 ± 420000
	0.014°	0.05780 ± 0.00022	0.0800 ± 0.0018	25.0 ± 3.2	3000 ± 650	6.70 ± 0.43	2300 ± 410	1500000 ± 360000
	0.019°	0.11160 ± 0.00042	0.1820 ± 0.0064	36.0 ± 5.7	5000 ± 1300	5.70 ± 0.44	1800 ± 370	1100000 ± 300000
	0.027°	0.21550 ± 0.00081	0.440 ± 0.022	49.0 ± 8.9	9000 ± 2300	4.70 ± 0.44	1300 ± 310	800000 ± 230000
	0.037°	0.4160 ± 0.0016	1.080 ± 0.077	60 ± 12	12000 ± 3400	3.80 ± 0.41	900 ± 240	500000 ± 150000
	0.052°	0.8030 ± 0.0030	2.70 ± 0.25	70 ± 15	13000 ± 4200	2.90 ± 0.37	600 ± 170	240000 ± 85000
	0.072°	1.5510 ± 0.0058	6.70 ± 0.76	70 ± 17	12000 ± 4200	2.10 ± 0.30	300 ± 100	100000 ± 38000
	0.100°	2.990 ± 0.011	16.0 ± 2.1	70 ± 16	10000 ± 3300	1.50 ± 0.22	170 ± 54	40000 ± 13000
	0.139°	5.780 ± 0.021	39.0 ± 5.2	60 ± 13	7000 ± 2300	1.00 ± 0.15	80 ± 24	11000 ± 4000
	0.193°	11.160 ± 0.040	90 ± 12	50 ± 10	5000 ± 1500	0.640 ± 0.094	34.0 ± 9.9	3000 ± 1100
	0.268°	21.550 ± 0.077	210 ± 27	42.0 ± 7.5	3400 ± 930	0.400 ± 0.055	13.0 ± 3.6	700 ± 250
	0.373°	41.60 ± 0.15	450 ± 59	32.0 ± 5.4	2000 ± 510	0.240 ± 0.033	4.0 ± 1.2	140 ± 49
	0.518°	80.30 ± 0.29	1000 ± 130	23.0 ± 3.7	1000 ± 260	0.140 ± 0.019	1.40 ± 0.38	25.0 ± 8.5
	0.720°	155.20 ± 0.56	2100 ± 270	16.0 ± 2.5	500 ± 130	0.080 ± 0.011	0.40 ± 0.11	4.0 ± 1.4
	1.000°	300.0 ± 1.1	4500 ± 570	10.0 ± 1.5	230 ± 54	0.0460 ± 0.0061	0.120 ± 0.031	0.60 ± 0.20

Following the reasoning of Wolk et al. (2013), we do not perform the complex error propagation from factorial moments to correlation functions, as the error estimate is already a crude approximation limited by the small number of fields.

Fig. 21 shows the probability distributions  $P(N)$  evaluated in square cells with different widths from  $\theta = 0.01^\circ$  to  $\theta = 1^\circ$  for each target class, with a cell width close to the fibre patrol radius ( $1.4' \approx 0.023^\circ$ ) highlighted. The dashed vertical line drawn for this highlighted cell represents the expected number of targets calculated from multiplying the mean target density with the cell area. Note that even for ELGs, the first few angular bins are shot-noise dominated ( $\bar{N} < 1$ , so most cells contain one or zero targets).

Table 8 presents the following quantities for each of the three main target classes:

- (i) Mean  $\bar{N} \equiv \langle N \rangle$
- (ii) Variance  $\sigma^2 \equiv \langle (N - \bar{N})^2 \rangle$
- (iii) Skewness  $\langle (N - \bar{N})^3 \rangle / \sigma^3$
- (iv) Kurtosis  $\langle (N - \bar{N})^4 \rangle / \sigma^4$

- (v) Cell-averaged angular correlation functions  $\bar{w}_2$ ,  $\bar{w}_3$ , and  $\bar{w}_4$

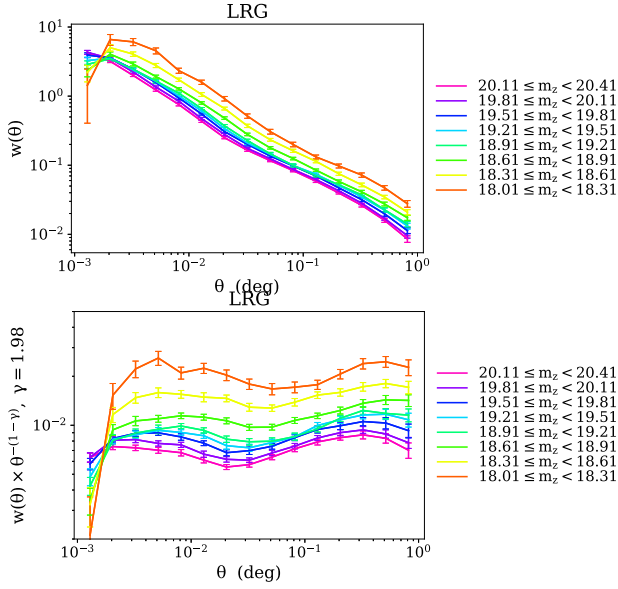
From the above quantities, other quantities of interest can be determined, such as the hierarchical moments  $S_p = \bar{w}_p / \bar{w}_2^{p-1}$  (Szapudi & Szalay 1993; Colombi & Szapudi 2001), fitted power-law parameters for  $\bar{w}_2$  (see e.g. Blake & Wall 2002), etc. More directly, these quantities can be used for the training and testing of mock catalogues.

## 7.5 Clustering as a function of magnitude

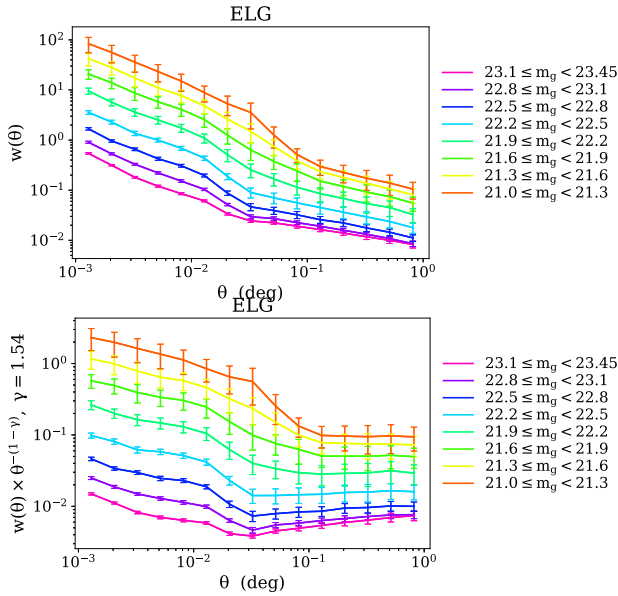
Angular clustering is expected to scale with sample depth (Peebles 1980), so analysing  $w(\theta)$  as a function of magnitude provides another test for the presence of systematics.

We divide the LRG and ELG samples into eight disjoint, equally wide magnitude slices from the bright limit ( $m_z = 18.01$  for LRGs,  $m_g = 21$  for ELGs) to the faint limit ( $m_z = 20.41$  for LRGs,  $m_g = 23.4$  for ELGs) of the target selection. For each bin, we evaluate





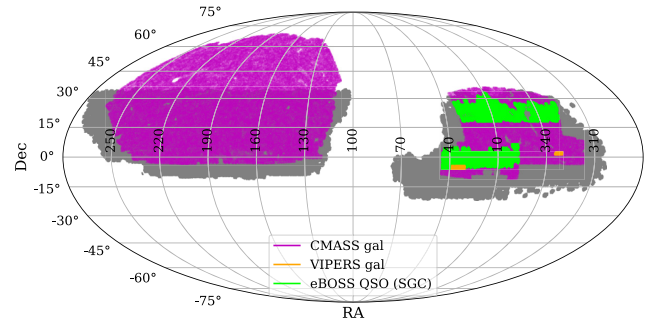
**Figure 22.** Angular correlation functions for LRGs in eight magnitude bins (upper plot), with the angular dependence scaled out for a fixed slope of  $\gamma = 1.98$  (lower plot), which is the slope determined from fitting over the full LRG sample in Table 6.



**Figure 23.** Angular correlation functions for ELGs in eight magnitude bins (upper plot), with the angular dependence scaled out for a fixed slope of  $\gamma = 1.54$  (lower plot), which is the slope determined from fitting over the full ELG sample in Table 6.

$w(\theta)$ .<sup>16</sup> The results are shown in Figs 22 and 23, with the upper plots showing  $w(\theta)$  for each slice and the lower plots showing the same functions but with the angular dependence divided out using a representative value of  $\gamma$  determined from fitting the full sample (Table 6).

<sup>16</sup>Note, we do not re-evaluate the photometric weights for each magnitude bin but instead apply the same weights to all bins.



**Figure 24.** Visualizing the overlap between the DECaLS DR7 footprint (grey) and the footprints of the external catalogues used for cross-correlations. Positions are mapped using Mollweide projection.

Qualitatively, the results match expectation: the brighter subsamples have larger clustering amplitudes and break from the power law form at larger scales. We also note that for both LRGs and ELGs, the minimum inflection slides to smaller scales at fainter magnitudes, again consistent with a 1-halo to 2-halo transition; fainter bins are at higher redshift (so the characteristic scale of the transition will shift to smaller angles due to the larger angular diameter distance) and/or lower luminosity (thus, the 1-halo term will be weaker, as less luminous galaxies reside in less massive haloes).

The best-fitting values for the clustering amplitudes  $A_w$  and slopes  $\gamma$  are reported in Tables 9 and 10. We fit  $A_w$  with the fixed representative value of  $\gamma$ , and also fit  $A_w$  and  $\gamma$  simultaneously, finding similar results in either case. We also perform the fits over different sets of angular scales, starting with a minimum cutoff of  $\theta = 0.005$  for LRGs to avoid scales where the power-law model appears to break down and obtaining fits for  $\theta < 0.05^\circ$  and  $0.05^\circ < \theta < 1^\circ$  separately as well as for the full range.

## 8 SPECTROSCOPIC CROSS-CORRELATIONS

### 8.1 Clustering as a function of redshift

#### 8.1.1 External catalogues

To probe the clustering as a function of redshift through cross-correlations, as described in Sections 2.2 and 2.5, we make use of several external spectroscopic catalogues. We use the CMASS galaxy sample from DR12 of the Baryon Oscillation Spectroscopic Survey (BOSS; Eisenstein et al. 2011; Dawson et al. 2013), which selects higher redshift galaxies at  $0.4 < z < 0.8$  and has significant angular overlap with the DECaLS footprint. We also use galaxies from the final data release of the VIMOS Public Extragalactic Redshift Survey<sup>17</sup> (VIPERS; Scodeggio et al. 2018). VIPERS extends over two narrow CFHTLS fields, W1 and W4, with a combined area of approximately  $23.5 \text{ deg}^2$ , and has nearly 90 000 redshifts out to  $z \sim 1$ . Finally, we use the main sample of QSOs from eBOSS DR14 (Dawson et al. 2016), which overlaps with the DECaLS footprint in the SGC. Fig. 24 shows where the footprints of these surveys intersect with DECaLS DR7, and Fig. 25 demonstrates how their redshift distributions span the expected redshift ranges of the DESI targets.

<sup>17</sup><http://vipers.inaf.it/>



**Table 9.** Best-fitting parameters from modelling the angular clustering of LRGs in  $z$ -band magnitude bins using equation (5). The clustering amplitude  $A_w$  is reported for a fixed slope of  $\gamma = 1.98$  (taken from the fit over the full LRG sample; see Table 6), as well as the results of fitting amplitude and slope simultaneously.  $\theta_{\min}$  and  $\theta_{\max}$  are the angular scales fit over, and  $m_z$  is in AB magnitudes.

$\theta_{\min}, \theta_{\max}$	$m_z$ bin	Med $m_z$	# Objects	LRG		$\gamma$
				$\log_{10} A_w$ ( $\gamma = 1.98$ )	$\log_{10} A_w$	
0.005°, 0.05°	20.11, 20.41	20.26	868849	-3.94 <sup>-3.92</sup> <sub>-3.96</sub>	-4.01 <sup>-3.88</sup> <sub>-4.15</sub>	2.00 <sup>2.07</sup> <sub>1.93</sub>
	19.81, 20.11	19.96	790503	-3.90 <sup>-3.88</sup> <sub>-3.92</sub>	-3.97 <sup>-3.82</sup> <sub>-4.12</sub>	2.00 <sup>2.08</sup> <sub>1.92</sub>
	19.51, 19.81	19.66	692920	-3.83 <sup>-3.81</sup> <sub>-3.86</sub>	-3.91 <sup>-3.72</sup> <sub>-4.10</sub>	2.00 <sup>2.09</sup> <sub>1.91</sub>
	19.21, 19.51	19.37	596349	-3.80 <sup>-3.78</sup> <sub>-3.82</sub>	-3.87 <sup>-3.71</sup> <sub>-4.04</sub>	2.00 <sup>2.08</sup> <sub>1.92</sub>
	18.91, 19.21	19.06	518696	-3.77 <sup>-3.75</sup> <sub>-3.79</sub>	-3.84 <sup>-3.68</sup> <sub>-4.01</sub>	2.00 <sup>2.08</sup> <sub>1.92</sub>
	18.61, 18.91	18.78	434355	-3.69 <sup>-3.68</sup> <sub>-3.70</sub>	-3.76 <sup>-3.66</sup> <sub>-3.86</sub>	2.00 <sup>2.04</sup> <sub>1.96</sub>
	18.31, 18.61	18.48	257434	-3.56 <sup>-3.55</sup> <sub>-3.58</sub>	-3.63 <sup>-3.50</sup> <sub>-3.76</sub>	2.00 <sup>2.06</sup> <sub>1.94</sub>
	18.01, 18.31	18.19	137380	-3.41 <sup>-3.39</sup> <sub>-3.43</sub>	-3.48 <sup>-3.25</sup> <sub>-3.72</sub>	2.00 <sup>2.10</sup> <sub>1.90</sub>
0.05°, 1°	20.11, 20.41	20.26	868849	-3.85 <sup>-3.83</sup> <sub>-3.87</sub>	-3.56 <sup>-3.50</sup> <sub>-3.61</sub>	1.87 <sup>1.91</sup> <sub>1.83</sub>
	19.81, 20.11	19.96	790503	-3.82 <sup>-3.80</sup> <sub>-3.85</sub>	-3.56 <sup>-3.50</sup> <sub>-3.61</sub>	1.88 <sup>1.92</sup> <sub>1.84</sub>
	19.51, 19.81	19.66	692920	-3.78 <sup>-3.76</sup> <sub>-3.81</sub>	-3.43 <sup>-3.39</sup> <sub>-3.48</sub>	1.85 <sup>1.88</sup> <sub>1.82</sub>
	19.21, 19.51	19.37	596349	-3.75 <sup>-3.72</sup> <sub>-3.78</sub>	-3.31 <sup>-3.26</sup> <sub>-3.35</sub>	1.81 <sup>1.85</sup> <sub>1.78</sub>
	18.91, 19.21	19.06	518696	-3.74 <sup>-3.71</sup> <sub>-3.77</sub>	-3.27 <sup>-3.22</sup> <sub>-3.31</sub>	1.80 <sup>1.83</sup> <sub>1.77</sub>
	18.61, 18.91	18.78	434355	-3.67 <sup>-3.65</sup> <sub>-3.69</sub>	-3.25 <sup>-3.23</sup> <sub>-3.27</sub>	1.82 <sup>1.83</sup> <sub>1.81</sub>
	18.31, 18.61	18.48	257434	-3.55 <sup>-3.53</sup> <sub>-3.57</sub>	-3.20 <sup>-3.17</sup> <sub>-3.23</sub>	1.85 <sup>1.86</sup> <sub>1.83</sub>
	18.01, 18.31	18.19	137380	-3.43 <sup>-3.41</sup> <sub>-3.46</sub>	-3.01 <sup>-2.96</sup> <sub>-3.06</sub>	1.82 <sup>1.85</sup> <sub>1.78</sub>
0.005°, 1°	20.11, 20.41	20.26	868849	-3.90 <sup>-3.88</sup> <sub>-3.92</sub>	-3.65 <sup>-3.60</sup> <sub>-3.70</sub>	1.90 <sup>1.93</sup> <sub>1.87</sub>
	19.81, 20.11	19.96	790503	-3.87 <sup>-3.85</sup> <sub>-3.88</sub>	-3.68 <sup>-3.63</sup> <sub>-3.74</sub>	1.92 <sup>1.95</sup> <sub>1.89</sub>
	19.51, 19.81	19.66	692920	-3.81 <sup>-3.80</sup> <sub>-3.83</sub>	-3.67 <sup>-3.62</sup> <sub>-3.73</sub>	1.94 <sup>1.97</sup> <sub>1.91</sub>
	19.21, 19.51	19.37	596349	-3.78 <sup>-3.76</sup> <sub>-3.80</sub>	-3.61 <sup>-3.56</sup> <sub>-3.67</sub>	1.93 <sup>1.96</sup> <sub>1.90</sub>
	18.91, 19.21	19.06	518696	-3.76 <sup>-3.74</sup> <sub>-3.77</sub>	-3.62 <sup>-3.56</sup> <sub>-3.67</sub>	1.94 <sup>1.96</sup> <sub>1.91</sub>
	18.61, 18.91	18.78	434355	-3.68 <sup>-3.67</sup> <sub>-3.69</sub>	-3.55 <sup>-3.51</sup> <sub>-3.59</sub>	1.94 <sup>1.96</sup> <sub>1.92</sub>
	18.31, 18.61	18.48	257434	-3.55 <sup>-3.54</sup> <sub>-3.57</sub>	-3.45 <sup>-3.40</sup> <sub>-3.49</sub>	1.94 <sup>1.97</sup> <sub>1.92</sub>
	18.01, 18.31	18.19	137380	-3.43 <sup>-3.41</sup> <sub>-3.44</sub>	-3.40 <sup>-3.33</sup> <sub>-3.47</sub>	1.97 <sup>2.00</sup> <sub>1.94</sub>

### 8.1.2 Projected real-space cross-correlation functions

We present the real-space projected cross-correlation functions (derived in Section 2.2) for LRGs in Fig. 26, using CMASS galaxies, VIPERS galaxies, and eBOSS QSOs in bins of width  $\delta z = 0.1$ . Some noisy redshift bins are omitted from the plots. The error bars are from bootstrapping on the area and therefore are likely overestimated for VIPERS, which has very small fields.

For ELGs, we initially find that the cross-correlations with CMASS galaxies flatten above  $\theta \sim 0.01^\circ$  for all redshift bins. However, using the CMASS systematics weights in concert with our own photometric weights eliminates this effect, indicating a correlation between systematics in the two catalogues, likely the anticorrelation with stars found in both DESI ELGs and CMASS samples. The projected real-space cross-correlations are plotted in Fig. 27, along with the power-law predictions from the ELG  $w(\theta)$  fits in Table 6, which we translate into  $w_p(r_p)$  using equation (7). We thus have consistency between ELG  $w(\theta)$  in Fig. 15,  $C_\ell$  in Fig. 18, and  $w_p(r_p)$  in Fig. 27. We also note that the break at small scales becomes less pronounced at higher redshift, as the 2-halo term becomes more dominant.

More puzzlingly, ELGs appear to show no correlation with eBOSS QSOs over the overlapping redshift range. This null signal is consistent within error bars across all redshift bins and remains null even when switching to brighter ELG subsamples. At present,

we wish to avoid speculating on why there is no cross-correlation between ELGs and eBOSS QSOs, as a full investigation with survey validation data and spectra is expected to paint a much clearer picture. Given the reasonable ELG autocorrelation and ELG  $\times$  CMASS cross-correlation, we do not believe that this is indicative of catastrophic failure in the ELG sample. Finally, we note that the QSO cross-correlations are too noise-dominated to obtain a meaningful signal.

### 8.1.3 Clustering $dN/dz$

For ELGs and QSOs, issues with cross-correlation measurements discussed in the previous section prevent us from obtaining meaningful  $dN/dz$  over the full redshift ranges of the targets. We therefore focus on LRGs and defer further investigation of ELGs and QSOs to a future work.

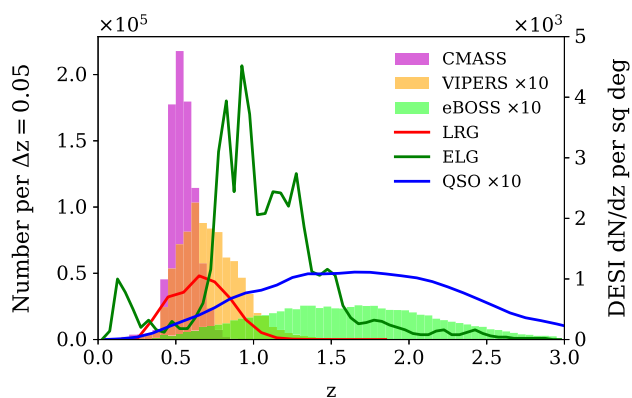
Using the method outlined in Section 2.5, we integrate over each set of cross-correlations in the overlapping redshift ranges to piece together the shape of the LRG  $dN/dz$ . We choose the minimum and maximum physical scales of integration in such a way as to reduce the propagated errors on  $dN/dz$ ; for LRG  $\times$  CMASS and LRG  $\times$  VIPERS, we use  $s_{\min} = 0.05 h^{-1}$  Mpc,  $s_{\max} = 5 h^{-1}$  Mpc, whereas for LRG  $\times$  eBOSS, we use  $s_{\min} = 0.2 h^{-1}$  Mpc,  $s_{\max} = 10 h^{-1}$  Mpc.

To minimize the potential impact of bias evolution in the photometric sample, we first divide it by colour before cross-correlating



**Table 10.** Best-fitting parameters from modelling the angular clustering of ELGs in  $g$ -band magnitude bins using equation (5). The clustering amplitude  $A_w$  is reported for a fixed slope of  $\gamma = 1.54$  (taken from the fit over the full ELG sample; see Table 6), as well as the results of fitting amplitude and slope simultaneously.  $\theta_{\min}$  and  $\theta_{\max}$  are the angular scales fit over, and  $m_g$  is in AB magnitudes.

$\theta_{\min}, \theta_{\max}$	$m_z$ bin	Med $m_z$	# Objects	ELG		$\log_{10} A_w$	$\gamma$
				$\log_{10} A_w (\gamma = 1.54)$			
$0^\circ, 0.05^\circ$	23.1, 23.4	23.27	8530522	-3.15	$^{+0.08}_{-0.22}$	-4.10	$^{+1.87}_{-1.73}$
	22.8, 23.1	22.97	4768510	-2.98	$^{+0.21}_{-0.07}$	-4.66	$^{+2.02}_{-1.95}$
	22.5, 22.8	22.68	2462312	-2.67	$^{+0.29}_{-0.76}$	-4.43	$^{+2.05}_{-1.95}$
	22.2, 22.5	22.38	1224671	-2.36	$^{+0.27}_{-0.47}$	-4.09	$^{+2.06}_{-1.94}$
	21.9, 22.2	22.07	614394	-1.97	$^{+0.88}_{-0.09}$	-3.68	$^{+2.05}_{-1.95}$
	21.6, 21.9	21.77	334669	-1.74	$^{+0.61}_{-0.90}$	-3.31	$^{+2.05}_{-1.95}$
	21.3, 21.6	21.47	205434	-1.74	$^{+0.45}_{-0.28}$	-3.00	$^{+2.03}_{-1.97}$
	21.0, 21.3	21.16	138431	-1.74	$^{+0.17}_{-0.31}$	-2.68	$^{+2.01}_{-1.98}$
$0.05^\circ, 1^\circ$	23.1, 23.4	23.27	8530522	-3.23	$^{+0.21}_{-0.26}$	-2.93	$^{+1.43}_{-1.39}$
	22.8, 23.1	22.97	4768510	-3.15	$^{+0.13}_{-0.17}$	-2.80	$^{+1.41}_{-1.40}$
	22.5, 22.8	22.68	2462312	-2.99	$^{+0.98}_{-0.01}$	-2.76	$^{+1.45}_{-1.44}$
	22.2, 22.5	22.38	1224671	-2.77	$^{+0.76}_{-0.78}$	-2.64	$^{+1.49}_{-1.48}$
	21.9, 22.2	22.07	614394	-2.47	$^{+0.46}_{-0.48}$	-2.53	$^{+1.56}_{-1.54}$
	21.6, 21.9	21.77	334669	-2.22	$^{+0.19}_{-0.25}$	-2.51	$^{+1.66}_{-1.61}$
	21.3, 21.6	21.47	205434	-2.04	$^{+0.01}_{-0.08}$	-2.47	$^{+1.72}_{-1.65}$
	21.0, 21.3	21.16	138431	-1.93	$^{+0.89}_{-0.97}$	-2.37	$^{+1.72}_{-1.62}$
$0^\circ, 1^\circ$	23.1, 23.4	23.27	8530522	-3.17	$^{+0.12}_{-0.21}$	-4.08	$^{+1.79}_{-1.74}$
	22.8, 23.1	22.97	4768510	-3.05	$^{+0.99}_{-0.11}$	-4.16	$^{+1.87}_{-1.81}$
	22.5, 22.8	22.68	2462312	-2.80	$^{+0.73}_{-0.87}$	-4.09	$^{+1.92}_{-1.87}$
	22.2, 22.5	22.38	1224671	-2.53	$^{+0.45}_{-0.63}$	-4.00	$^{+1.98}_{-1.94}$
	21.9, 22.2	22.07	614394	-2.32	$^{+0.22}_{-0.43}$	-3.69	$^{+2.00}_{-1.96}$
	21.6, 21.9	21.77	334669	-2.13	$^{+0.04}_{-0.24}$	-3.35	$^{+2.00}_{-1.97}$
	21.3, 21.6	21.47	205434	-1.98	$^{+0.88}_{-0.09}$	-3.11	$^{+2.00}_{-1.96}$
	21.0, 21.3	21.16	138431	-1.89	$^{+0.80}_{-0.01}$	-2.99	$^{+2.00}_{-1.91}$



**Figure 25.** Visualizing the redshift ranges for DESI targets compared to catalogues from spectroscopic surveys that overlap the DECaLS footprint. The solid lines correspond to the expected  $dN/dz$  per square degree of the DESI target classes, while the histograms are from the external catalogues.

each subsample separately, as discussed by Ménard et al. (2013), Schmidt et al. (2013), Rahman et al. (2015), and Gatti et al. (2018). In fig. 1 of Prakash et al. (2016), the photometric redshifts of eBOSS LRGs are plotted in colour space, with a transition from mostly  $z < 0.6$  objects to mostly  $0.6 < z < 1.0$  objects occurring when  $r - W1$  is in the range between 2 and 3. Motivated by this, we select a roughly median value of  $r - W1 = 2.6$  to create two similarly sized LRG

subsamples. We cross-correlated these two subsamples separately with the three external catalogues, with the combined results shown in Fig. 28 and compared to the results derived without binning the sample.

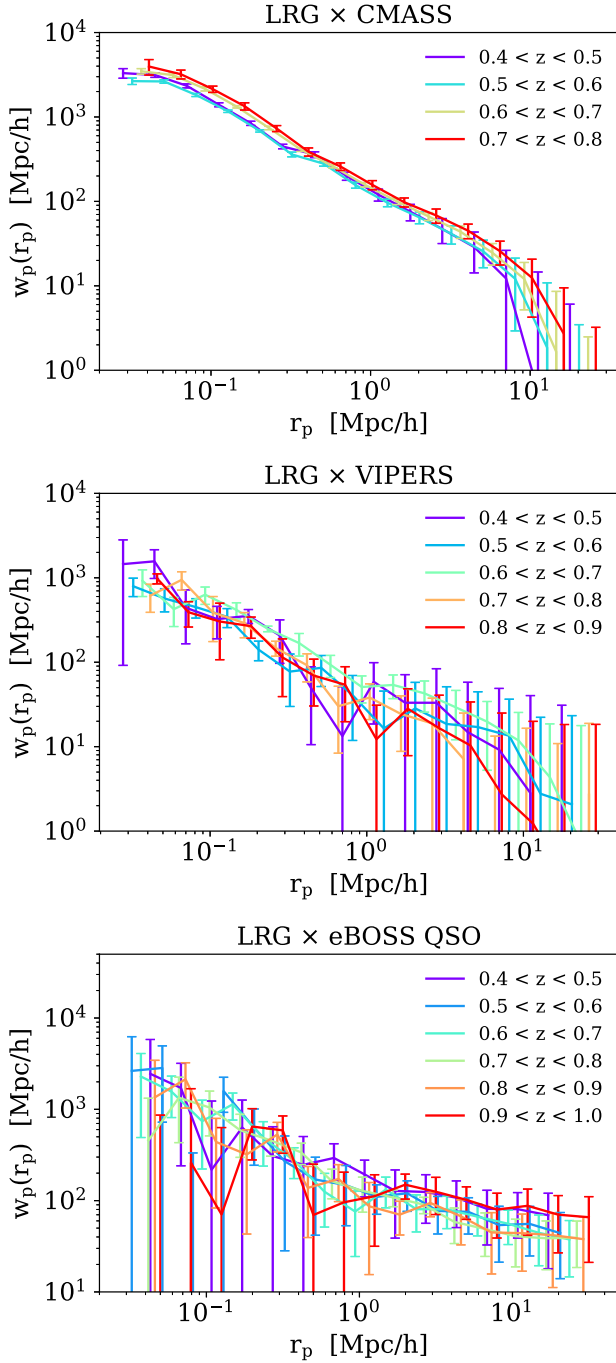
We find that the clustering  $dN/dz$  from all three cross-correlations match very well with the fiducial FDR  $dN/dz$ .<sup>18</sup> Along with the excellent agreement between measured and fiducial bias found in Section 7.3, this suggests that the LRG sample will be able to fully meet the cosmology goals of the collaboration. Additionally, the upper panel of Fig. 28 confirms that the colour cut at  $r - W1 = 2.6$  effectively splits the LRG sample into high and low redshift subsamples with an approximate boundary at  $z \sim 0.65$ .

## 8.2 Clustering as a function of luminosity

By cross-correlating magnitude binned LRGs with redshift binned spectroscopic catalogues, we can also probe the luminosity dependence of the sample. We begin by dividing the LRGs into three broader magnitude bins from  $m_z = 18.01$  to  $m_z = 20.41$ , the bright and faint limits, respectively, of the target selection. To improve signal-to-noise, we also double the widths of the redshift bins to  $\delta z = 0.2$  and focus on the cross-correlations with CMASS

<sup>18</sup>We note that the fiducial redshift distributions in Fig. 25 are of the targets selected from imaging, including contaminants.

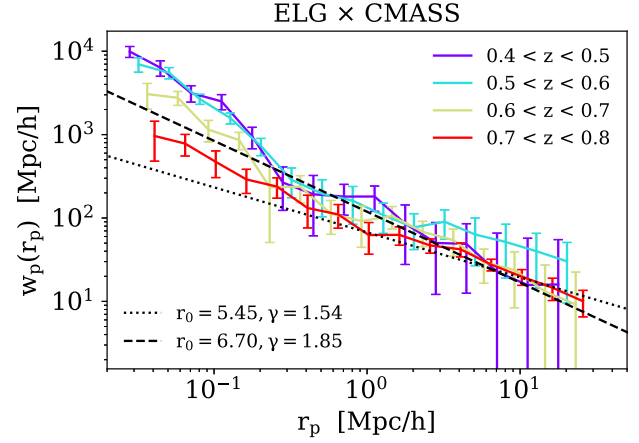




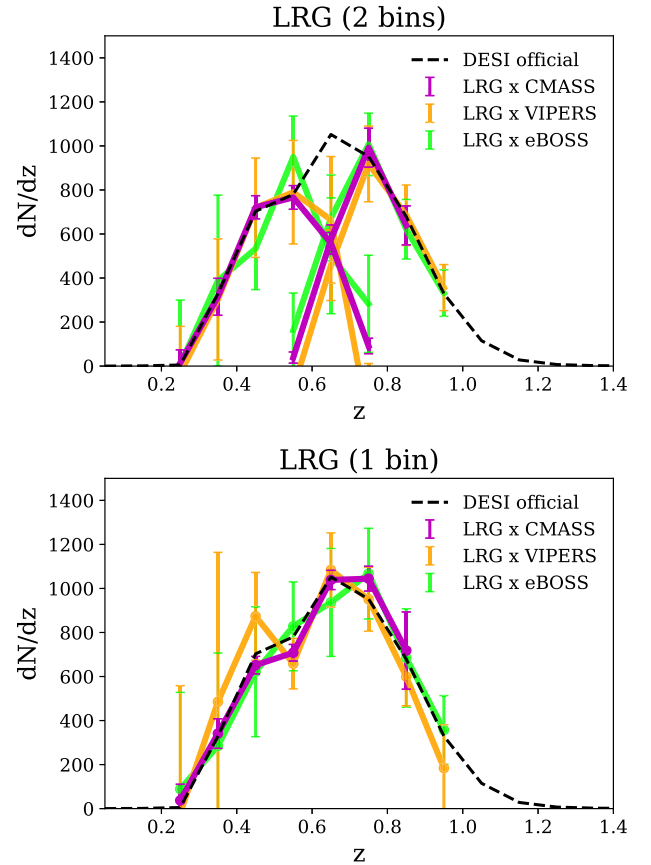
**Figure 26.** Projected real-space cross-correlations between LRGs and three external samples with spectroscopic redshifts: CMASS galaxies, VIPERS galaxies, and eBOSS QSOs. Error bars are from bootstrapping.

galaxies, which involve the smallest error bars. Through these cross-correlations, we can crudely reconstruct  $dN/dz$  for each of the magnitude bins, shown in Fig. 29. The behaviour is as expected; brighter objects are at lower mean redshift, with the redshift distributions generally appearing as deeper and deeper copies of each other.

This result allows us to convert angular cross-correlations into projected real-space cross-correlations, as detailed in Section 2.2, giving us  $w_p(r_p)$  in three broad luminosity bins for a given redshift bin. We select a bin near the middle of the CMASS redshift

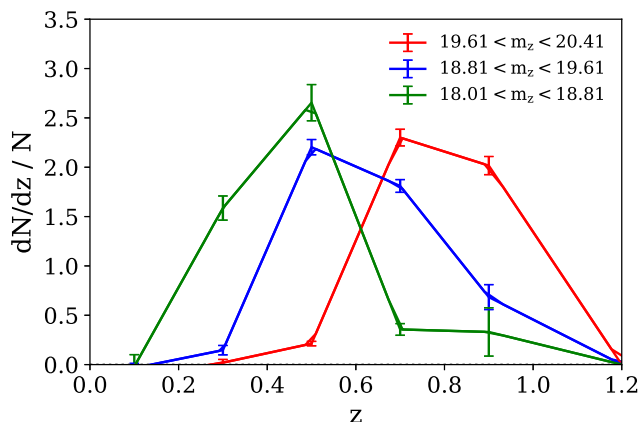


**Figure 27.** Projected real-space cross-correlations between ELGs and CMASS galaxies. Error bars are from bootstrapping. Dashed and dotted lines are the power-law fits of the ELG autocorrelation from Table 6.



**Figure 28.** The clustering-based  $dN/dz$  for LRGs derived from cross-correlations with CMASS galaxies (magenta), VIPERS galaxies (orange), and eBOSS QSOs (lime), with the expected  $dN/dz$  plotted as a dashed line. The upper plot shows  $dN/dz$  calculated using two  $r - W1$  colour bins, a proxy for a  $z \sim 0.6$  cut in order to reduce the impact of bias evolution, while the lower plot is determined using the full sample. Error bars are from propagating bootstrap errors from the cross-correlations through the  $dN/dz$  calculation.





**Figure 29.** Normalized  $dN/dz$  for each of three broad magnitude bins, derived from cross-correlations with CMASS galaxies in broad redshift bins of  $\delta z = 0.2$ . Error bars are from propagating bootstrap errors from the cross-correlations through the  $dN/dz$  calculation.

**Table 11.** Fits to the luminosity binned LRG-CMASS cross-correlations assuming  $w_p(r_p) = Ar_p^{1-\gamma}$ . The first column specifies the LRG apparent magnitude bins, while the second column calculates the corresponding absolute magnitude bin at the midpoint of the CMASS redshift bin  $0.4 < z < 0.6$ .

$m_z$ bin	LRG $\times$ CMASS, $0.4 < z < 0.6$	
	$M_z(\bar{z} = 0.5)$	$\log_{10} A$ ( $\gamma = 1.98$ )
19.61, 20.41	−21.87, −21.07	−3.40 <sup>+3.36</sup> <sub>−3.44</sub>
18.81, 19.61	−22.67, −21.87	−2.51 <sup>+2.49</sup> <sub>−2.53</sub>
18.01, 18.81	−23.47, −22.67	−2.56 <sup>+2.54</sup> <sub>−2.59</sub>

range<sup>19</sup>,  $0.4 < z < 0.6$ , and fit the clustering for each corresponding luminosity bin to a power law,  $w_p(r_p) = Ar_p^{1-\gamma}$ . The results are given in Table 11.

## 9 SUMMARY AND CONCLUSIONS

In order to fully realize the statistical power of the DESI experiment, it is vital to assess the quality of the imaging data and target definitions and to account for any non-cosmological sources of spatial fluctuations in the galaxy catalogues that could bias the cosmological analyses. In the first part of this paper, we diagnose causes of systematic errors in the clustering of DESI main targets selected from imaging and present masks and photometric weights aimed at reducing these effects. The masks and weights will be used in construction of cosmological clustering samples. Our key results are summarized below:

(i) We find that obscuration due to bright stars in the foreground creates significant variations in density, particularly for QSOs. Implementing aggressive masks around Tycho-2 and WISE stars, which remove 7–8 per cent of the usable sky area for galaxies and 28 per cent of the usable sky area for QSOs, dramatically improves the agreement of the angular correlation functions with cosmological predictions.

(ii) We determine that a general mask around large galaxies and other extended sources is not indicated; however, by visual inspection, we discover that failing to mask around the Coma cluster and M3 will create a significantly overestimated clustering signal for QSOs in the NGC. Additionally, the images of a small number

of very bright stars are plagued by complex patterns of reflected pupil ghosts, which create structures in the ELG density beyond the scope of even our highly conservative bright star masks.

(iii) We find that masked LRGs exhibit only minor density variations as a function of potential systematics such as stellar density, extinction, airmass, seeing, sky brightness, and exposure time. By contrast, masked ELG and QSO densities still fluctuate significantly, with the most dominant systematics, stellar density, and extinction, affecting densities by as much as 10 per cent. We find that ELGs are anticorrelated with stars and extinction, while QSOs are positively correlated. Photometric weights calculated by performing multilinear regression on these trends significantly ameliorate them.

(iv) We perform angular cross-correlations between the targets and stars and again find that LRGs are uncorrelated, ELGs are anticorrelated, and QSOs are positively correlated. Dividing the stars into three galactic latitude bins, we use QSO–star cross-correlations and star autocorrelations to estimate the stellar contamination fraction in the QSO sample as a function of galactic latitude.

We stress that this process has been highly iterative, with our efforts continuously informing the evolution of DESI’s imaging data reduction pipelines and target selection algorithms. The Legacy Survey Data Release 9, which is currently being processed and will be used for DESI target selection, has made several algorithmic upgrades motivated in part by the feedback in this study. These include changes to the pixel-level flat-field response functions and an improved modeling of sky subtraction. The latter improvements are most pronounced for ELG targets that are faint relative to the sky and therefore have their target densities modulated by errors in the sky modeling. The new sky model also helps with issues of scattered light around bright stars that affect the selection of QSO targets. In addition, the imaging team has added more aggressive foreground masking and has flagged other bad data from the list of problematic regions identified in this study (such as the ghost pupils around certain visibly bright stars and the M3/NGC contaminants in the north). Similarly, the target selection algorithms have been iteratively updated many times in response to our findings.

In addition to being a crucial first step towards constraining cosmology with DESI clustering measurements, our findings have important implications for other ongoing and future imaging surveys. As multi-epoch surveys become deeper and more sensitive, they will be increasingly limited by systematic uncertainties from instrument calibration, survey characteristics, and observing conditions. Our framework for identifying and mitigating the effects of such systematics, such as our new approach to quantifying contamination due to stars, is therefore highly relevant and widely applicable to future imaging surveys.

After applying masks and weights, we devote the second part of this paper to modeling the properties of the samples, providing the first large-scale clustering analysis of DESI targets. Modeling the samples is an important first step for doing cosmology with DESI, and our clustering results will also aid in the creation and validation of accurate mock catalogues. Additionally, we present several new methodologies, including the technique of probing the luminosity-dependent clustering by cross-correlating magnitude-binned photometric samples with redshift-binned spectroscopic samples. These methods can be applied to other clustering studies with deep photometric data, for instance in future studies with data from the Large Synoptic Survey Telescope



(LSST Science Collaboration 2009). Our main results are outlined below:

(i) We present the average densities before and after the corrections have been applied, finding that all three target densities are in reasonably good agreement with expectation after masking and weighting.

(ii) We model the angular correlation functions of the samples, assuming power-law spatial correlation functions. For LRGs, we recover values that agree very well with earlier studies. For ELGs, we see a broken power law, with different slopes for  $\theta < 0.01^\circ$  and  $\theta > 0.05^\circ$ , which agree reasonably well with similar studies. For QSOs, we obtain a highly inflated value for the clustering amplitude, indicating that substantial contamination remains.

(iii) We compare the observed angular power spectra to theory to determine the linear large-scale bias and also probe the scale dependence of the bias in the weakly non-linear regime. For LRGs, we find a value of  $b_0$  that agrees very well with the DESI FDR prediction. For ELGs, we find a value of  $b_0$  that is higher than the conservative lower limit given by the FDR but is similar to values from the literature and is self-consistent with our angular and real-space clustering measurements. By contrast, the observed angular power spectrum for QSOs is a poor fit to theory, with all scales seemingly affected by non-cosmological signals.

(iv) We use cross-correlations with external spectroscopy to determine real-space projected cross-correlation functions in redshift bins, through which we also derive clustering  $dN/dz$ . For LRGs, the clustering as a function of redshift behaves as expected, and we see an excellent match with the expected  $dN/dz$  from target selection. For ELGs, the redshift-binned cross-correlations with CMASS are consistent with expectation, but cross-correlations with eBOSS QSOs show no significant correlation. For QSOs, the cross-correlations are not currently clean enough to meaningfully model  $dN/dz$ .

(v) The clustering of LRGs and ELGs as a function of magnitude also behaves as predicted, with clustering amplitude scaling with depth. We provide fits to the angular correlation functions in magnitude bins. We also cross-correlate magnitude binned LRGs with redshift binned CMASS galaxies to probe luminosity-dependent clustering.

(vi) We present counts-in-cells moments and cell-averaged higher order correlation functions to further facilitate mock calibration and validation.

Overall, our results suggest that the quality of the imaging and the selection of targets are suitable for achieving the ambitious scientific objectives of the DESI collaboration. With imaging surveys completed and spectroscopic first light announced in October 2019, the commissioning phase is on track for completion in January 2020. After a survey validation period in the spring, the 5-year survey is expected to begin in the summer of 2020. We look forward to the exciting and impactful new science that DESI will enable in the coming decade.

## ACKNOWLEDGEMENTS

The authors thank Stephen Bailey, Daniel Eisenstein, Shirley Ho, Dustin Lang, Jeffrey Newman, Anand Raichoor, Ashley Ross, Hee-Jong Seo, Mike Wilson, Christophe Yèche, and Pauline Zarrouk for many useful discussions. E. K. and M. W. are supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Award No. DE-SC0017860. D. S., J. G., M. L., and J. M. are supported by the Director, Office of Science, Office

of High Energy Physics of the U.S. Department of Energy under contract no. DE-AC02-05CH11231, and by the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility under the same contract. This work also made extensive use of the NASA Astrophysics Data System and of the astro-ph preprint archive at arXiv.org. Additional support for DESI is provided by the U.S. National Science Foundation, Division of Astronomical Sciences under contract no. AST-0950945 to the National Optical Astronomy Observatory; the Science and Technologies Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the National Council of Science and Technology of Mexico; and by the DESI Member Institutions. The authors are honored to be permitted to conduct astronomical research on Iolkam Duag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

## DATA AVAILABILITY STATEMENT

The data underlying this article were accessed from The Legacy Survey (<http://legacysurvey.org/>). The derived data generated in this research will be shared upon reasonable request to the corresponding author.

## REFERENCES

- Albrecht A. et al., 2006, preprint (arXiv:0609591)
- Amendola L. et al., 2013, *Living Rev. Relativ.*, 16, 6
- Amendola L., Bettoni D., Marta Pinho A., Casas S., 2019, *Universe*, 6, 20
- Bassett B., Hlozek R., 2010, *Dark Energy: Observational and Theoretical Approaches*. Cambridge Univ. Press, Cambridge, UK, p. 246
- Blake C., Wall J., 2002, *MNRAS*, 337, 993
- Blas D., Lesgourgues J., Tram T., 2011, *J. Cosmology Astropart. Phys.*, 2011, 034
- Burden A., Padmanabhan N., Cahn R. N., White M. J., Samushia L., 2017, *J. Cosmology Astropart. Phys.*, 3, 001
- Cahn R. N., Pinol L., Hand N., McDonald P., Seljak U., 2017, APS April Meeting Abstracts, p. S5.004
- Casas S., Kunz M., Martinelli M., Pettorino V., 2017, *Phys. Dark Universe*, 18, 73
- Chiang Y.-K., Ménard B., Schiminovich D., 2018, *ApJ*, 877, 150
- Colombi S., Szapudi I., 2001, in Banday A. J., Zaroubi S., Bartelmann M., eds, *Mining the Sky*. Springer-Verlag, Berlin, Heidelberg, p. 262
- Comparat J. et al., 2013, *MNRAS*, 433, 1146
- Cresswell J. G., Percival W. J., 2009, *MNRAS*, 392, 682
- Crocce M., Gaztañaga E., Cabré A., Carnero A., Sánchez E., 2011, *MNRAS*, 417, 2577
- Crocce M. et al., 2016, *MNRAS*, 455, 4301
- Croom S. M. et al., 2005, *MNRAS*, 356, 415
- Cutri R. M. et al., 2015, Technical Report, Explanatory Supplement to the NEOWISE Data Release Products
- Davis C. et al., 2018, *MNRAS*, 477, 2196
- Davis M., Peebles P. J. E., 1983, *ApJ*, 267, 465
- Dawson K. S. et al., 2013, *AJ*, 145, 10
- Dawson K. S. et al., 2016, *AJ*, 151, 44
- Delubac T. et al., 2017, *MNRAS*, 465, 1831
- DES Collaboration, 2005, preprint (arXiv:0510346)
- DESI Collaboration, 2016, preprint (arXiv:0912.0201)
- Desjacques V., Jeong D., Schmidt F., 2018, *Phys. Rep.*, 733, 1
- Dey A. et al., 2018, *AJ*, 157, 168
- Efron B., 1979, *Ann. Statist.*, 7, 1
- Eisenstein D. J., 2005, *New A Rev.*, 49, 360
- Eisenstein D. J. et al., 2011, *AJ*, 142, 72
- Elvin-Poole J. et al., 2018, *Phys. Rev. D*, 98, 042006
- Favole G. et al., 2016, *MNRAS*, 461, 3421



- Fisher K. B., Davis M., Strauss M. A., Yahil A., Huchra J., 1994, *MNRAS*, 266, 50
- Fitzpatrick E. L., 1999, *PASP*, 111, 63
- Flaugher B. et al., 2015, *AJ*, 150, 150
- Font-Ribera A., McDonald P., Mostek N., Reid B. A., Seo H.-J., Slosar A., 2014, *J. Cosmology Astropart. Phys.*, 5, 023
- Fry J. N., 1985, *ApJ*, 289, 10
- Fry J. N., Gaztanaga E., 1993, *ApJ*, 413, 447
- Fry J. N., Colombi S., Fosalba P., Balaraman A., Szapudi I., Teyssier R., 2011, *MNRAS*, 415, 153
- Gariazzo S., Lopez-Honorez L., Mena O., 2015, *Phys. Rev. D*, 92, 063510
- Gatti M. et al., 2018, *MNRAS*, 477, 1664
- Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
- Guzik J., Bernstein G., Smith R. E., 2007, *MNRAS*, 375, 1329
- Hamann J., Hannestad S., Melchiorri A., Wong Y. Y. Y., 2008, *J. Astropart. Phys.*, 2008, 017
- Hildebrandt H. et al., 2017, *MNRAS*, 465, 1454
- Ho S., Hirata C., Padmanabhan N., Seljak U., Bahcall N., 2008, *Phys. Rev. D*, 78, 043519
- Hubble E., 1934, *ApJ*, 79, 8
- Høg E. et al., 2000, *A&A*, 355, L27
- Jain B., Khoury J., 2010, *Ann. Phys.*, 325, 1479
- Jarrett T. H., Chester T., Cutri R., Schneider S., Skrutskie M., Huchra J. P., 2000, *AJ*, 119, 2498
- Jouvel S. et al., 2017, *MNRAS*, 469, 2771
- Joyce A., Jain B., Khoury J., Trodden M., 2015, *Phys. Rep.*, 568, 1
- Kashiwagi T., Yahata K., Suto Y., 2013, *PASJ*, 65, 43
- Kashiwagi T., Suto Y., Taruya A., Kayo I., Nishimichi T., Yahata K., 2015, *ApJ*, 799, 132
- Krolewski A., Ferraro S., Schlafly E. F., White M., 2020, *J. Cosmol. Astropart. Phys.*, 2020, 047
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Landy S. D., Szalay A. S., Koo D. C., 1996, *ApJ*, 460, 94
- Lang D., Hogg D. W., Mykytyn D., 2016, *Astrophysics Source Code Library*, record ascl:1604.008
- Leistedt B. et al., 2016, *ApJS*, 226, 24
- Limber D. N., 1953, *ApJ*, 117, 134
- Loverde M., Afshordi N., 2008, *Phys. Rev. D*, 78, 123506
- LSST Science Collaboration, 2009, preprint ([arXiv:0912.0201](https://arxiv.org/abs/0912.0201))
- Matthews D. J., Newman J. A., 2010, *ApJ*, 721, 456
- Matthews D. J., Newman J. A., 2012, *ApJ*, 745, 180
- McQuinn M., White M., 2013, *MNRAS*, 433, 2857
- Ménard B., Scranton R., Schmidt S., Morrison C., Jeong D., Budavari T., Rahman M., 2013, preprint ([arXiv:1303.4722](https://arxiv.org/abs/1303.4722))
- Mo H. J., Jing Y. P., Boerner G., 1992, *ApJ*, 392, 452
- Modi C., White M., Vlah Z., 2017, *J. Cosmology Astropart. Phys.*, 2017, 009
- Mostek N., Coil A. L., Cooper M., Davis M., Newman J. A., Weiner B. J., 2013, *ApJ*, 767, 89
- Myers A. D. et al., 2006, *ApJ*, 638, 622
- Myers A. D., White M., Ball N. M., 2009, *MNRAS*, 399, 2279
- Newman J. A., 2008, *ApJ*, 684, 88
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19
- Padmanabhan N., White M., Norberg P., Porciani C., 2009, *MNRAS*, 397, 1862
- Peacock J. A., 1999, *Cosmological Physics*. Cambridge Univ. Press, Cambridge, UK
- Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton Univ. Press, Princeton, New Jersey
- Phillipps S., Shanks T., 1987, *MNRAS*, 227, 115
- Pinol L., Cahn R. N., Hand N., Seljak U., White M., 2017, *J. Cosmology Astropart. Phys.*, 4, 008
- Prakash A. et al., 2016, *ApJS*, 224, 34
- Rahman M., Ménard B., Scranton R., Schmidt S. J., Morrison C. B., 2015, *MNRAS*, 447, 3500
- Ross A. J., Brunner R. J., Myers A. D., 2006, *ApJ*, 649, 48
- Ross A. J. et al., 2011, *MNRAS*, 417, 1350
- Ross A. J. et al., 2012, *MNRAS*, 424, 564
- Rubin V. C., 1954, *PNAS*, 40, 541
- Ruggeri R. et al., 2019, *MNRAS*, 483, 3878
- Sawangwit U., Shanks T., Abdalla F. B., Cannon R. D., Croom S. M., Edge A. C., Ross N. P., Wake D. A., 2011, *MNRAS*, 416, 3033
- Schlafly E. F., Finkbeiner D. P., 2011, *ApJ*, 737, 103
- 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
- Schulz A. E., 2010, *ApJ*, 724, 1305
- Schulz A. E., White M., 2006, *Astropart. Phys.*, 25, 172
- Scodegg M. et al., 2018, *A&A*, 609, A84
- Scott V. et al., 2016, *MNRAS*, 462, 1683
- Scott V., Benoit-Lévy A., Coupon J., Ilbert O., Mellier Y., 2018, *MNRAS*, 474, 3921
- Scranton R. et al., 2002, *ApJ*, 579, 48
- Seldner M., Peebles P. J. E., 1979, *ApJ*, 227, 30
- Seljak U., 2001, *MNRAS*, 325, 1359
- Simon P., 2007, *A&A*, 473, 711
- Smith R. E. et al., 2003, *MNRAS*, 341, 1311
- Smith R. E., Scoccamarro R., Sheth R. K., 2007, *Phys. Rev. D*, 75, 063512
- Suchyta E. et al., 2016, *MNRAS*, 457, 786
- Szapudi I., 1998, *ApJ*, 497, 16
- Szapudi I., Colombi S., 1996, *ApJ*, 470, 131
- Szapudi I., Szalay A. S., 1993, *ApJ*, 408, 43
- Szapudi I., Meiksin A., Nichol R. C., 1996, *ApJ*, 473, 15
- Tellarini M., Ross A. J., Tasinato G., Wands D., 2016, *J. Cosmology Astropart. Phys.*, 6, 014
- Wechsler R. H., Tinker J. L., 2018, *ARA&A*, 56, 435
- Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, *Phys. Rep.*, 530, 87
- White S. D. M., 1979, *MNRAS*, 186, 145
- Wilson M. J., White M., 2019, *J. Cosmology Astropart. Phys.*, 2019, 015
- Wolk M., McCracken H. J., Colombi S., Fry J. N., Kilbinger M., Hudelot P., Mellier Y., Ilbert O., 2013, *MNRAS*, 435, 2
- Wright E. L. et al., 2010, *AJ*, 140, 1868
- Yahata K., Yonehara A., Suto Y., Turner E. L., Broadhurst T., Finkbeiner D. P., 2007, *PASJ*, 59, 205
- Zehavi I. et al., 2005, *ApJ*, 621, 22

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