Defining the DESI Bright Galaxy Survey target selection

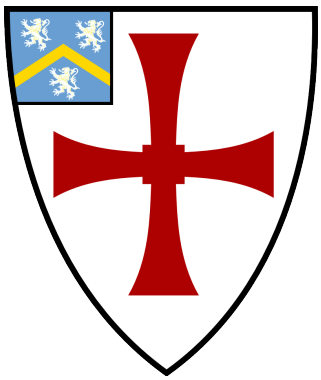
OMAR ANTONIO RUIZ-MACIAS

Defining the DESI Bright Galaxy Survey target selection

Omar Antonio Ruiz Macias

A thesis presented for the degree of

Doctor of Philosophy



Institute for Computational Cosmology

Durham University

United Kingdom

19th July 2021

Defining the DESI Bright Galaxy Survey target selection

Omar Antonio Ruiz Macias

Abstract

This thesis aims to cover the steps taken for the selection of the input galaxy catalogue for the Dark Energy Spectroscopic Instrument (DESI) Bright Galaxy Survey (BGS) using the photometric Legacy Surveys. The BGS is a redshift survey of bright galaxies that will be performed using the DESI 5,000 fibre spectrograph on the 4-m Mayall telescopein at Kitt Peak, Arizona.

Our galaxy selection implements a new way to perform star galaxy separation using the *Gaia* photometry. The purity of our sample is assessed with previous galaxy surveys GAMA* and the SDSS Main Galaxy Sample, and with the MXXL light-cone mock catalogue through clustering measurements. The robustness of the BGS selection criteria are assessed by quantifying the dependence of the target galaxy density on imaging and other properties. Systematic correlations are found with amplitudes of less than 5 per cent. This work also presents the first results from the Survey Validation (SV) stage of DESI. Using the SV data we were able to assess our BGS selection and tune it to achieve a high redshift success rate. The final catalogue includes nearly 30 Million galaxies for a 14,000 $\deg^2$ area that covers the North and South Galactic Caps.

Supervisors: Prof. Shaun M. Cole, Prof. Carlton Baugh and Dr Peder Norberg.

*This and subsequent acronyms are defined in the glossary

Acknowledgements

It is four years now since the beginning of this amazing journey and it is time to thank everyone who made this possible. This little section won't be enough to thank my supervisors, Shaun, Carlton, and Peder. Their guidance and support over my stay at the ICC was superb, I felt welcome all the time and they looked after me in all aspects, academic and personal life. I appreciate the patience they had with me, especially with my poor communications skills and the multiple corrections over all my written documents, including this thesis! To Pauline, who supports me through my job seeking and was my mentor (a great one!) in the last couple of years. To John for the nice talks and guidance. To Alex, Mike, Chang and many more for their DESI guidance.

I would like to acknowledge CONACyT for providing the funding that made my PhD possible, to Durham University and DESI UK for the extra funding and enabling me to travel to the multiple DESI meetings.

I had the pleasure to meet very nice people across these years. To my CDT friends (the Durham data-miners!) for all the good times and all the A.I./machine learning/linear regression (whichever name you prefer) lectures and workshops we attend. To my office mates, with whom I spent great times, Myles, Aidan, Behzad, and many more. To Shufei, Lindsay and Sabine for the nice talks and aid with administrative duties. To my other American family (the real America, the continent), Cesar and his wonderful family who made me feel like home when I needed the most, and Cristian for the nice talks at lunch time.

Finally, I would like to thank my family, my wife Luz and my new born son Demian. It was because of their love and lots of patience, that I managed to get this far. To my parents and brothers who have always supported me, and to my new parents, my parents in law, who have always looked after me like a son.

To Luz and Demian

Contents

Declaration

The work in this thesis is based on research carried out by the author between 2017 and 2021 while the author was a research student under the supervision of Prof. Shaun Cole, Prof. Carlton Baugh, Dr Peder Norberg, and with contributions from Dr. Pauline Zarrouk at the Institute for Computational Cosmology, Department of Physics, Durham University, England.

No part of this thesis has been submitted elsewhere for any other degree or qualification, and it is the sole work of the author unless referenced to the contrary in the text.

Figures 2.16, 2.17, 3.7, 3.8, 3.9, 3.10, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7 were made by Dr. Pauline Zarrouk.

Figure 5.1 was taken from the DESI Survey Validation plan web page.

Some of the work presented in this thesis has been published in journals - the relevant publications are listed below.

Publications

Chapter 2 has been published in the form of a paper:

- Ruiz-Macias, O., Zarrouk, P., Cole, S., Baugh, C. M., Norberg, P., Lucey, J., Dey, A., Eisenstein, D. J., Doel, P., Gaztañaga, E., Hahn, C., Kehoe, R., Kitanidis, E., Landriau, M., Lang, D., Moustakas, J., Myers, A. D., Prada, F., Schubnell, M., Weinberg, D. H., Wilson, M. J., "Characterizing the target selection pipeline for the Dark Energy Spectroscopic Instrument Bright Galaxy Survey", 2021, *MNRAS*, **502**, 4328-4349.

Some parts of Sections 2.3 and 2.4 of Chapter 2 have been summarised and published as part of a paper:

- Ruiz-Macias, O., Zarrouk, P., Cole, S., Norberg, P., Baugh, C. M., Brooks, D., Dey, A., Duan, Y., Eftekharzadeh, S., Eisenstein, D. J., Forero-Romero, J., Gaztañaga, E., Hahn, C., Kehoe, R., Landriau, M., Lang, D., Levi, M. E., Lucey, J., Meisner, A., Moustakas, J., Myers, A. D., Palanque-Delabrouille, N., Poppett, C., Prada, F., Raichoor, A., Schlegel, D. J., Schubnell, M., Tarlé, G., Weinberg, D. H., Wilson, M. J., Yèche, C., "Preliminary Target Selection for the DESI Bright Galaxy Survey (BGS)", 2020, *Res. Notes AAS*, **4**, 187.

Chapters 3 and 4 are to be submitted as part of a paper. Dr. Pauline Z. lead most of Chapter 4 where I contributed with the data analysis. I lead most of Chapter 3 with the exception of the angular correlation functions.

- Zarrouk, P., Ruiz-Macias, O., Cole, S., Baugh, C. M., "Preliminary clustering properties of the DESI BGS bright targets using DR9 Legacy Imaging Surveys", 2021.

Chapter 5 is to be submitted as part of a planned paper:

- Ruiz-Macias, O. and DESI Collaboration, "The DESI Bright Galaxy Survey in Survey Validation", 2021.

List of Figures

List of Tables

Glossary

ΛCDM Λ Cold Dark Matter

2MRS 2MASS Redshift Survey (Huchra et al., 2012)

2SLAQ-LRG 2dF-SDSS LRG and QSO Luminous Red Galaxy Survey (Cannon et al., 2006)

2dFGRS Two Degree Field Galaxy Redshift Survey (Colless et al., 2001, 2003)

6dFGS Six Degree Field Galaxy Redshift Survey (Jones et al., 2009, 2004)

AGES AGN and Galaxy Evolution Survey (Kochanek et al., 2012)

ALLWISE A combination of WISE and NEOWISE (Secrest et al., 2015).

Autofib Autofib Redshift Survey (Ellis et al., 1996)

BAO Baryon Acoustic Oscillations

BASS The Beijing-Arizona Sky Survey.

BGS DESI Bright Galaxy Survey.

BOSS Baryon Oscillation Spectroscopic Survey (Eisenstein et al., 2011; Parejko et al., 2012; Dawson et al., 2013)

BS Bright Stars

BTS Bright Time Survey

CC Colour Cuts

CCD Charge-Couple Device

CFRS Canada-France Redshift Survey (Lilly et al., 1995)

CNOC2 Canadian Network for Observational Cosmology 2 Field Galaxy Redshift Survey (Yee et al., 2000)

COMP Composite of DEV and EXP TRACTOR profiles fit.

COSMOS Cosmic Evolution Survey (Scoville et al. 2007a)

CP Community Pipeline

CfA Center for Astrophysics Redshift Survey (Huchra et al., 1983)

CfA2 Center for Astrophysics 2 Redshift Survey (Geller and Huchra, 1989)

DECaLS The Dark Energy Camera Legacy Surveys.

DECam The Dark Energy Camera (Flaugher et al., 2015)

DEEP2 DEEP Extragalactic Evolutionary Probe 2 Redshift Survey (Davis et al., 2003; Newman et al., 2013)

DES The Dark Energy Survey (Crocce et al., 2017).

DESI The Dark Energy Spectroscopic Instrument (DESI Collaboration et al., 2016)

DEV A de Vaucouleurs TRACTOR profile fit.

DTS Dark Time Survey

DURS Durham-UKST Redshift Survey (Ratcliffe et al., 1996)

ELG DESI Emission Line Galaxy.

ESP Eso Slice Project (Vettolani et al., 1997)

EUCLID The European Space Agency mission with a singular goal of characterizing dark energy via wide-field surveys of the extragalactic sky (Laureijs et al., 2011).

EXP Exponential TRACTOR profile fit.

FMC Fibre Magnitude Cut

GAIA ESO collaboration space-base survey to measure low-resolution spectra of the Milky Stars (Gaia Collaboration et al., 2018).

GAMA Galaxy And Mass Assembly redshift survey (Driver et al., 2012; Liske et al., 2015; Baldry et al., 2017)

GC Globular Cluster

H-AAO Hawaii+AAO K-Band Redshift Survey (Huang et al., 2003)

LBG-z3 Lyman Break Galaxies at $z \sim 3$ (Steidel et al., 2003)

LCRS Las Campanas Redshift Survey (Shectman et al., 1996)

LG Large Galaxies

LRG DESI Luminous Red Galaxy.

LS Legacy Surveys (Dey et al., 2019).

LSST The Large Synoptic Survey Telescope (Ivezić et al., 2019).

MGC Millennium Galaxy Catalogue (Liske et al., 2003)

MWS DESI Milky Way Star.

MXXL The Millennium-XXL simulation (Angulo et al., 2012a).

MzLS The Mayall z-band Legacy Surveys.

NANOMAGIES (nMgy) 1 nMgy corresponds to an AB magnitude of 22.5.

NASA National Aeronautics and Space Administration.

NEOWISE Near-Earth Objects Wide-field Infrared Survey (Mainzer et al., 2011).

NGC North Galactic Cap.

NOAO National Optical Astronomy Observatory.

NOBS Number of Observations in the Imaging Legacy Surveys.

OpenNGC Is a database containing positions and main data of NGC (New General Catalogue) and IC (Index Catalogue).

PRIMUS PRIsm MUlti-Object Survey (Coil et al., 2011; Cool et al., 2013)

PSF Point Spread Function.

QC Quality Cuts.

QSO DESI Quasi Stellar Object.

REDROCK Redshift estimate algorithm for DESI.

REX Round Exponential with variable radius TRACTOR profile fit.

RSD Redshift Space Distortions.

SAPM Stromlo Automated Photographic Measurement (Loveday et al., 1992)

SDSS-LRG Sloan Digital Sky Survey Luminous Red Galaxies (Eisenstein et al., 2001)

SDSS-MGS Sloan Digital Sky Survey Main Galaxy Survey (Strauss et al., 2002)

SDSS-s82 Sloan Digital Sky Survey Stripe 82 galaxy survey (Adelman-McCarthy et al., 2006; Baldry et al., 2005)

SER Sersic TRACTOR profile fit.

SGA Siena Galaxy Atlas (Moustakas in prep.).

SGC South Galactic Cap.

SSRS2 Southern Sky Redshift Survey 2 (da Costa et al., 1998)

SV Survey Validation.

TRACTOR Pixel-level forward-modelling software to extract the source catalogues for the Legacy Surveys (Lang et al., 2016).

TYCHO s an astrometric reference catalogue containing positions, proper motions and two-colour photometry for the 2.5 million brightest stars in the sky (Høg et al., 2000).

VIMOS Visible Multi-Object Spectrograph

VIPERS VIMOS Public Extragalactic Redshift Survey (Garilli et al., 2014; Guzzo et al., 2014)

VLT Very Large Telescope

VVDS VIMOS VLT Deep Survey (Le Fèvre et al., 2004; Garilli et al., 2008)

WCS World Coordinate System.

WISE Wide-field Infrared Survey Explorer mission (Wright et al., 2010).

WiggleZ WiggleZ Dark Energy Survey (Drinkwater et al., 2010; Parkinson et al., 2012)

eBOSS Extended Baryon Oscillation Spectroscopic Survey (Dawson et al., 2016)

zCOSMOS redshift survey in the COSMOS field (Lilly et al., 2007, 2009)

Chapter 1

Introduction to galaxy redshift surveys and DESI

1.1 Overview and impact of galaxy redshift surveys

In the last decades of the 20th century, physics has been obligated to change its conception of the Universe.

To the known issues with the standard model of particle physics – regarding neutrino mass (Bahcall and Davis, 1976; Fukuda et al., 2001), the baryon asymmetry (Farrar and Shaposhnikov, 1993), and a theory for gravity at subatomic scales (Lykken, 2010) – we have to add the recent data from cosmological surveys. In the last decades, cosmology has transitioned from being an almost speculative and theoretical discipline, to be a data-rich precision branch of physics. The measurements of the cosmic microwave background (CMB) from COBE (Smoot et al., 1992), WMAP (Bennett et al., 2003), PLANCK (Planck Collaboration et al., 2020), the measurements of the Hubble parameter by the Hubble Space Telescope (HST) (Kennicutt et al., 1995; Efstathiou, 2021), the discovery of the accelerated expansion of the Universe with observations of Type Ia Supernovae (Riess et al., 1998; Perlmutter et al., 1999; Knop et al., 2003; Tonry et al., 2003), to mention some examples, have set the parameters that best describes the evolution of the

Universe. Although many questions were answered thanks to the new data reported by cosmological surveys, many more questions have arisen, challenging our understanding of physics once again. We faced a very dark path when we discovered that 95% of the contents of the Universe is unknown, and that all matter and radiation studied in laboratories can explain only the remaining 5% which is composed of baryons, leptons, and radiation. Of the unknown 95% of the contents of the Universe, around 70% is thought to be in the form of dark energy (DE) and the remaining 35% is known as dark matter (DM). The DE was introduced as an explanation of the recent accelerated expansion of the Universe. The mass of our Galaxy can be estimated from the distribution of the stellar light, to account for gas and dust mass, the mean mass-to-light ratio of stellar population is used. From this mass estimate, using Newtonian mechanics, one would expect that the rotational velocity as a function of radius ($V(r)$) decreases outside the bulge of the Galaxy. However, what we observe is a flat rotational curve indicating that our Galaxy contains significantly more mass than is visible in the form of stars. This additional mass is called dark matter (DM) (Rubin et al., 1980).

The standard model of cosmology (ΛCDM) considers a Universe that is homogeneous and isotropic, however, at early stages of the Universe there must exists some small perturbations that grow and evolve into the large scale structures – galaxy clusters, voids and filaments – we see today (see Schneider, 2006, chap 1.2.5). Before recombination, photons, electrons, and protons were coupled in the photon-baryon plasma. Compression and enhancement of regions within the plasma caused a series of acoustic oscillations until the Universe was cold enough for photons to decouple from baryons. These photons, now free to travel across the Universe, carried the signal of the acoustic oscillations encoded their temperature distribution and this is what we see in the CMB (Nieves and Volkas, 2003). The angular power spectrum of the CMB give us a precise measurement of the contents and geometry of the Universe (Hu and Dodelson, 2002; Miller et al., 2000; de Bernardis et al., 2000; Halverson et al., 2002; Hanany et al., 2000). The position of the

first peak (at $l \sim 200$), for instance, give us information of the age of the Universe and the horizon radius (R_H) at the decoupling time, and hence, the content of DE and the geometry of the Universe. The amplitude of the first peak contains information on the content of matter

The cosmological theory also predicts that these oscillations will be imprinted onto the late-time matter power spectrum (Peebles and Yu, 1970; Bond and Efstathiou, 1984; Holtzman, 1989; Hu and Sugiyama, 1996; Eisenstein and Hu, 1998). The feature corresponds to the maximum distance that the acoustic oscillation sound wave could travel before matter and radiation decouple. This distance is at ~ 150 Mpc, the size of the sound horizon at recombination (R_H). The acoustic feature is manifested as an enhancement in clustering due R_H as a small single spike in the correlation function at 150 Mpc separation.

The scale R_H is known from CMB observations and thus provides us with a standard ruler. Using the distribution of baryonic matter from galaxy surveys we can use its apparent size to measure the distance to the effective redshift of the survey galaxies. Measuring the distance as a function of redshift gives us information of the expansion rate and geometry of the Universe. Two-point correlation measurements will also detect the anisotropies in galaxy clustering - redshift space distortions (RSD) - due to the peculiar velocities of galaxies generated by density perturbations (Jackson, 1972; Kaiser, 1987; Zarrouk et al., 2018). This probes the content of matter (Ω_m) and the bias of the tracers, and gives a direct measurement of the properties of gravity at each redshift, through its effect on galaxy motions.

In order to constrain the various models that try to explain the underlying mystery behind the dark energy and the dark matter, it is necessary to have a three-dimensional map of the distribution of matter in the Universe.

1.2 Brief historical review of galaxy surveys

It is from spectroscopic surveys (Spec-z) like CfA (Huchra et al., 1983), SAPM (Loveday et al., 1992), 2dF (Colless et al., 2001, 2003), 6dF (Jones et al., 2009, 2004), DEEP2 (Davis et al., 2003; Newman et al., 2013), SDSS (Smee et al., 2013; Strauss et al., 2002), GAMA (Driver et al., 2012; Liske et al., 2015; Baldry et al., 2017), PRIMUS (Coil et al., 2011; Cool et al., 2013), VIPERS (Garilli et al., 2014; Guzzo et al., 2014), VVDS (Le Fèvre et al., 2004; Garilli et al., 2008), WiggleZ (Drinkwater et al., 2010; Parkinson et al., 2012), and zCOSMOS (Lilly et al., 2007, 2009), that we can get detailed information about the large scale structure of the Universe. However, getting this $3D$ picture of the galaxy distribution requires spending much time, resources and money.

The spectroscopic surveys listed above that probed intermediate and higher redshifts tended to have smaller solid angles and sampled part of the overall galaxy distribution.

On the other hand we have the photometric redshifts (photo-z) like in SDSS (York et al., 2000; Abazajian et al., 2009), PanSTARRS (Chambers et al., 2016), KiDS (de Jong et al., 2013), or the HSC-SSP (Aihara et al., 2018). With photometric redshifts it is possible to study a much larger number of objects, and the sampling is more homogenous, covering all of the galaxies down to some flux limit, rather than being biased* towards those for which spectroscopic redshifts can be successfully measured e.g. those with strong emission lines. This approach consists of measuring the brightness of an object through various filters or passbands of colours, and the idea is to isolate features like the 4000 Angstrom break or the Lyman break – if the filters fall either side of such a break, then the corresponding colour is red . There are two main approaches to estimate the redshifts from these observed magnitudes and then estimate the distance with Hubble's law (i) template

*photo-z redshifts can also be biased due systematics such as the depth of the imaging survey where faint sources tend to have high scattering towards true redshifts.

fitting methods (Benitez, 2000; Coe et al., 2006; Arnouts et al., 2002) in which the spectral energy distribution (SED) of the objects is compared with spectroscopic or well known galaxies templates which are redshifted to, hence finding their galaxy spectral type and and their redshifts; (ii) and more recently, the machine learning techniques such as kd-tree nearest neighbour fit (KF) (Csabai et al., 2007) used in SDSS DR12 (Alam et al., 2015), or the random forest like the one performed in SDSS DR10 (Ahn et al., 2014) in which the different fluxes of each galaxy serve as features for the model to be trained with known spectroscopic redshifts. Each of these techniques has its advantages and disadvantages (see Sánchez et al. (2014) and a review from Zheng and Zhang (2012)).

The idea of photo-z was originally developed by Baum in 1962 to find galaxy clusters but it was not until the nineties that astronomers returned to this technique with the development of large and deep field surveys which improved cosmological studies such as galaxy clustering and weak lensing. However, photometric redshifts are susceptible to larger random and systematic errors than spectroscopic redshifts, this is why it is important to understand the uncertainties if we want to derive cosmological results. Random and systematic redshift errors can lead to errors in the luminosity and mass function (Marchesini et al., 2009; Bates et al., 2019), and in galaxy clustering. In order to constrain cosmological parameters it is essential to have good redshift measurements, or at least an understanding of the systematics. For good calibration of the photo-z error distributions we must have a spectroscopic sample that is representative of the target sample in (i) large area, (ii) high completeness and (iii) few wrong redshifts (Cunha et al., 2014).

On the other hand, having high resolution spectra does not mean we have a secure redshifts, in order to obtain the redshifts, the spectra are compared with a set of galaxy templates, then redshift is derived by redshifting the template to improve the match with the observed spectrum; this is often done by cross correlating the template with the observed spectrum. To quantify the quality of the spectra there are four main items to consider (i) Spectral coverage, (ii) integration time, (iii) the

template set and (iv) observing conditions.

Going back to the CfA redshift survey (Center for Astrophysics) in 1982, the very first spectroscopic redshift survey in the modern era, Davis, Huchra and company measured the largest sample of galaxies distribution with 2400 redshifts over 9000 deg^2 area, pointing the telescope at one galaxy at a time, with a B magnitude limited to 14.5, showing that the distribution of galaxies was anything but random. Spectra resolution was $\Delta\lambda = 0.5\,\text{nm}$ over the wavelength range $450 - 710\,\text{nm}$, having a resolving power of $R = \lambda/\Delta\lambda \sim 710/0.5 = 1420$. Integration time was between 15 to 50 minutes. The CfA survey was followed by CfA2 (Geller and Huchra, 1989). CfA2 was released in 1999 with over 18000 redshifts over 17000 deg^2 and with a limited magnitude of $B < 15.5$. Spectral resolution was 0.6 nm covering the wavelength range of $370 - 750\,\text{nm}$. It was with CfA2 that the *Great Wall* was discovered, a galaxy filament structure of over 60 Mpc width and 5 Mpc in thickness running all the way across between 8 hours and 17 hours RA. Figure 1.1 shows the sky coverage of CfA2 in RA and redshift, although the *Great Wall* cannot be appreciated as this is more visible at wider declination slice.

Twenty years after CfA, the 2dF Galaxy Redshift Survey (2dFGRS) measured over 220 000 redshifts in 1500 deg^2 area and with a much fainter magnitude limit of $b_J < 19.45$. 2dFGRS is a multifibre spectrograph with 400 robotic actuated fibres, and hence was able to measure the spectra of 400 objects simultaneously. While CfA2 has a median redshift of $z \sim 0.02$, 2dFGRS has a median redshift of $z \sim 0.11$ and goes as deep as $z \sim 0.3$. The wavelength range runs from $360 - 800$nm. The measurements from 2dFGRS had a much better accuracy than previous spectroscopic surveys that led to a better galaxy distribution on scales up to 600 Mpc. The main cosmological results of 2dFGRS was to provide, when combined with the CMB, confirmation that a dark energy term was needed in the theory (Efstathiou et al., 2002), and measurements in the BAO and power-spectrum ($P(k)$) (Cole et al., 2005; Peacock et al., 2001; Percival et al., 2001). Other contributions include mass and clustering derivations with the luminosity function (Cole et al.,

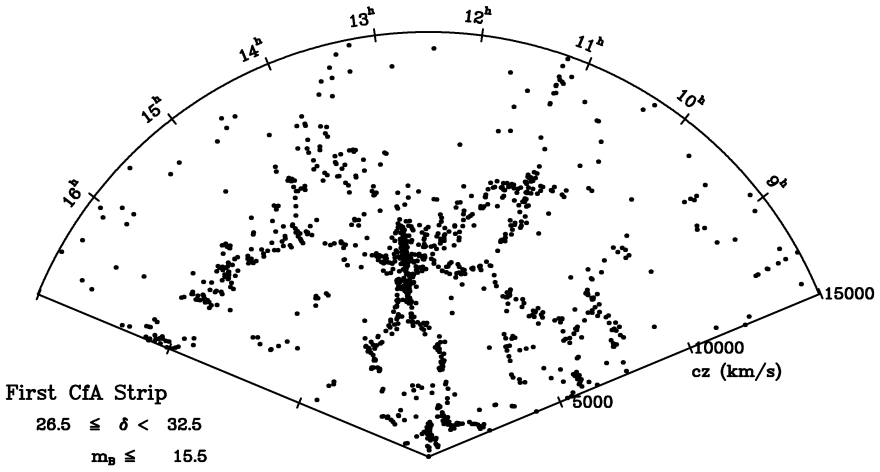


Figure 1.1: CfA2 sky coverage for a 6 degree slice in declination, $26.5 < \delta < 32.5$. Image taken from John Huchra's `https://www.cfa.harvard.edu/~dfabricant/huchra/zcat/`

2001; Norberg et al., 2001).

With instruments becoming more and more precise, many more spectroscopic surveys arose throughout the years. While 6dFGS was the largest in area, covering $17,000$ deg^2 others covered more volume by going to fainter magnitudes and higher redshift. SDSS Main Galaxy Sample (MGS) (Strauss et al., 2002) measured 1 million galaxy redshifts in an area of $10\,000$ deg^2 with a r-band magnitude limit of 17.7. The GAMA survey, on the other hand, covers a smaller area on the sky (286 deg^2), but goes deeper in magnitude, reaching galaxies as faint as r-band magnitude of 19.8 and a median redshift of $\bar{z} = 0.2$.

To map the large-scale structure at high redshifts, surveys should look for specific tracers in order to save time or to match the density of targets to the number of redshifts that can be measured in a single pointing. Previous mentioned surveys covered the magnitude limited samples without colour selection; such galaxies tend to be unbiased measurements of the galaxy to matter distribution and offer a way

to measure the clustering of the underlying dark matter at scales of $0.02 < k < 0.15$ h Mpc^{-1} (Lahav et al., 2002). However, further in redshift we use other tracers such as Luminous Red Galaxies (LRGs), Emission Line Galaxies (ELGs), and quasars (QSOs). Surveys like BOSS (Eisenstein et al., 2011; Parejko et al., 2012; Dawson et al., 2013), and its successor, eBOSS (Dawson et al., 2016), measure redshifts for these tracers up to $z = 3$. Fig. 1.2 illustrates some of the most important surveys of the time in area and density.

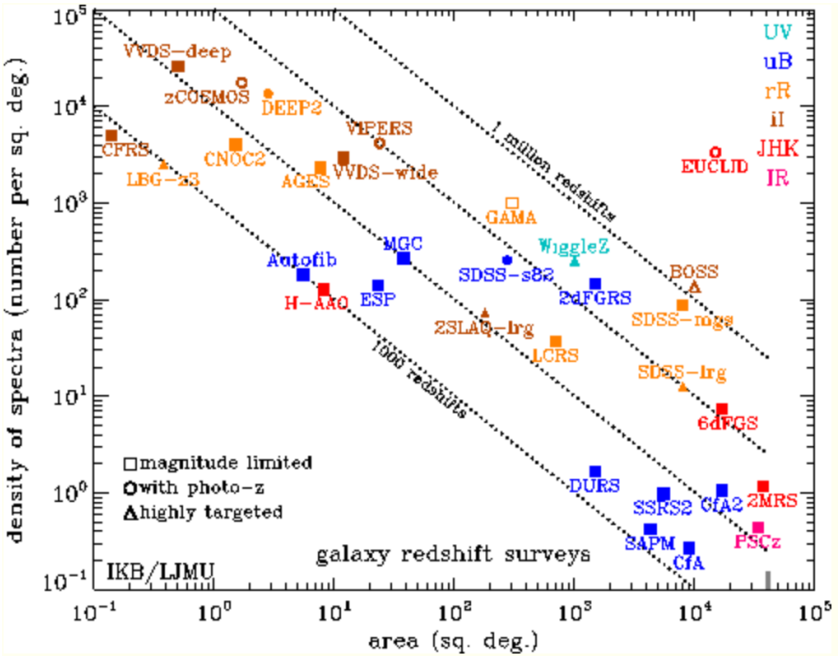


Figure 1.2: Comparison between galaxy surveys in area and density. The definition of the survey acronyms and papers describing the surveys are given in the glossary. Image taken from Baldry's `https://www.astro.ljmu.ac.uk/~ikb/research/galaxy-redshift-surveys.html`.

The upcoming Dark Energy Spectroscopic instrument (DESI) aims to get redshifts of around 30 million galaxies, an order of magnitude greater than its predecessors SDSS, BOSS and eBOSS. In section 1.3 we will introduce DESI on which this work is focused.

1.3 The DESI survey

The Dark Energy Spectroscopic Instrument* (DESI) (DESI Collaboration et al., 2016) is a multi-fibre spectrograph that will be used to carry out a number of wide-field surveys of galaxies and quasars to map the large-scale structure of the Universe. These surveys will probe the form of dark energy by allowing high precision measurements of the baryon acoustic oscillation (BAO) scale and the growth rate of structure using redshift-space distortions (RSD).

DESI is a robotically-actuated, fibre-fed spectrograph that is capable of collecting 5 000 spectra simultaneously. The spectra cover the wavelength range 360 to 980 nm, with a spectral resolution of $\Delta\lambda = 0.18$ nm and a resolving power of $R = \lambda/\Delta\lambda$ between 2 000 and 5 500, depending on the wavelength. DESI will be used to conduct a five-year survey, which already started in 2021 and with the aim of measuring redshifts over a solid angle of 14 000 deg^2. More than 30 million spectroscopic targets will be selected for four different tracer samples drawn from the imaging data. These are (i) luminous red galaxies (LRGs) in the redshift range $z = 0.3$ to $z = 1$, (ii) emission line galaxies (ELGs) to $z = 1.7$, (iii) quasars to higher redshifts ($2.1 < z < 3.5$), and for the $Ly - \alpha$ forest absorption features in their spectra, which will be used as tracers of the large scale structure, and (iv) a magnitude-limited BGS out to $z \approx 0.6$ with a median redshift of $z \approx 0.2$ which is the focus of this thesis.

DESI observations are divided into two main programmes: the Bright Time Survey (BTS) and the Dark Time Survey (DTS). The BGS will be part of the BTS and is conducted when the Moon is above the horizon and the sky is too bright to allow efficient observation of fainter targets. The BTS excludes the few nights closest to full Moon and BGS always targets fields that are at least $40-50$ deg away from the Moon. BGS alone will be ten times larger than the SDSS-I and SDSS-II main galaxy samples (MGS) of 1 million bright galaxies that were observed over

*`http://desi.lbl.gov/`

the time period 1999 – 2008 (Abazajian et al., 2003), while DESI will observe all of its tracers, including BGS in a five year period.

DESI provides at least an order of magnitude improvement over BOSS both in the comoving volume it probes and the number of galaxies it will map. This will significantly advance our understanding of the expansion history of the Universe. Precision measurements of the expansion history of the Universe is a powerful probe of the nature of dark energy. If we want to quantify how better DESI will be compared to previous surveys, we can use the Dark Energy Task Force figure of merit (DETF FoM), which measures the combined precision on the dark energy equation of state today ω_0, and its evolution with redshift ω_a. DESI achieves an DETF FoM of more than a factor of three better than all the Stage-III galaxy BAO measurements combined (DESI Collaboration et al., 2016). This increases even more with the inclusion of Ly-α forest BAO, and more including galaxy broadband power spectrum to scales of $k = 0.1$ h Mpc^{-1}. The BGS will enable the best ever measurement of low redshift BAO and RSD with at least 10 times more precision than SDSS MGS.

1.4 The Imaging Legacy Surveys

The Dark Energy Camera Legacy Survey (DECaLS), the Beijing-Arizona Sky Survey (BASS), and the Mayall z-band Legacy Survey (MzLS) are the combination of public projects that together constitute the DESI Legacy Imaging Surveys (hereafter the Legacy Surveys) (Dey et al., 2019). The imaging Legacy Surveys (LS) were created with the aim of attaining photometry with the necessary target density, coverage and depth required for DESI. The SDSS MGS (Strauss et al., 2002) and Pan-STARRS1 (Chambers et al., 2016) catalogues are both too shallow to be used to reliably select the DESI survey targets. The DES survey (The Dark Energy Survey Collaboration, 2005) does reach the target depth for DESI, but only covers 5000 deg^2, mostly in the South Galactic Cap (SGC), with only $\sim$ 1130 deg^2

observable from Kitt Peak where the Mayall telescope that hosts DESI is sited.

In this work we focused on the last two releases of the LS, the data release 8* (DR8), and the data release 9† (DR9). Main differences within the scope of this work are listed in Sec. 3.1.1.

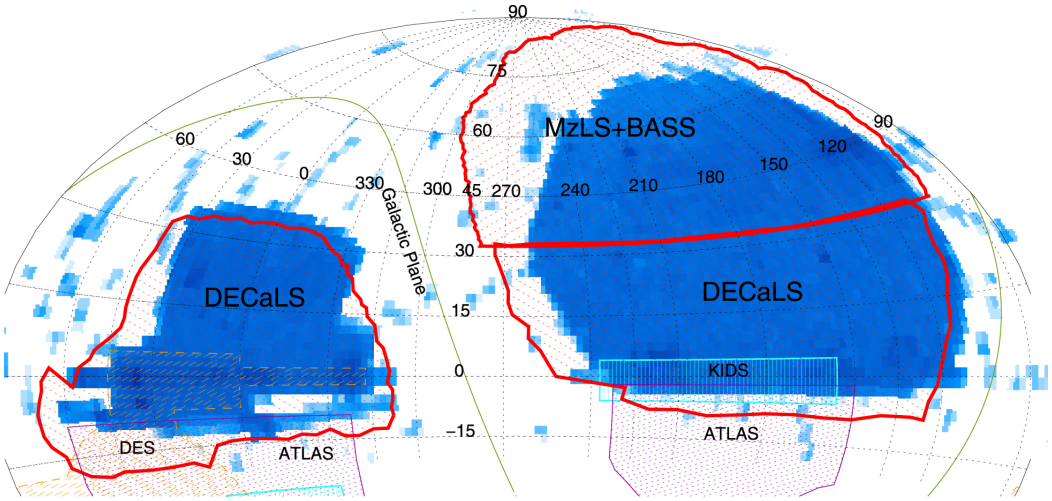


Figure 1.3: Footprints of the optical imaging surveys contributing to DESI imaging, demarcated by the thick red outlines, are shown here in an equal-area Aitoff projection in equatorial coordinates. The region covered by the BASS and MzLS surveys is almost entirely in the North Galactic Cap (NGC) at declinations $\delta \geq +32$ deg, and DECaLS covers the entire South Galactic Cap and the $\delta \leq +34$ deg regions in the NGC. The regions covered by existing wide-area spectroscopic redshift surveys (SDSS, 2dF, and BOSS) are shown in the blue gray scale in the map provided, where the darker colours represent a higher density of spectroscopic redshifts. The Legacy Surveys provide deeper imaging and can leverage the existing spectroscopy in these regions, unlike most other existing or ongoing deep imaging surveys (e.g., DES, ATLAS, KIDS, etc.). Credit: Fig.1 of Dey et al. (2019)

In Fig. 1.3 we compare the Legacy imaging Surveys footprint with existing imaging and spectroscopic surveys in these regions.

1.5 TRACTOR

All data from the Legacy Surveys are first processed at the NSF's National Optical-Infrared Astronomy Research Laboratory in Tucson (NOIRLab) through the NOIR-

*See details of DR8 here: `https://www.legacysurvey.org/dr8/description/`
†See details of DR9 here: `https://www.legacysurvey.org/dr9/description/`

Lab Community Pipeline* (CP). The CP takes raw data as an input and provides detrended and calibrated data products such as instrumental calibration (e.g. bias subtraction and flat fielding), astrometric calibration (e.g. mapping the distortions and providing a world coordinate system, or WCS), photometric characterization (e.g. magnitude zero point calibration) and artifact identification, masking and/or removal (e.g. removal of cross-talk and pupil ghosts, and identification and masking of cosmic rays).

The source catalogues for the Legacy Surveys are constructed using the LEGACYPIPE† software, which uses the TRACTOR‡(Lang et al., 2016) code for pixel-level forward-modelling of astronomical sources. This is a statistically rigorous approach to fitting the differing point spread functions (PSF) and pixel sampling of these data, which is particularly important as the optical data has a PSF width around 1 arcsec, and the WISE data a PSF of 6 arcsec in W1-W3, and $\sim$ 12 arcsec in W4.

The steps in the LEGACYPIPE processing are described in Dey et al. (2019); we briefly summarize relevant parts here.

After initial source detection and defining the contiguous set of pixels associated with each detection (termed a blob), LEGACYPIPE proceeds to fit these pixels with models of the surface brightness, including a point-source and a variety of galaxy models. These fits are performed on the individual optical images (in g, r and z bands), taking into account the different PSF and sensitivity of each image, using TRACTOR.

Besides the PSF model, TRACTOR fits four other light profile models to sources: a round exponential with a variable radius (referred to as REX), an exponential profile (EXP), a de Vaucouleurs profile(DEV), and a SERSIC profile for DR9 or composite of DEV and EXP profiles (COMP) for DR8. The decision as to

*https://www.noao.edu/noao/staff/fvaldes/CPDocPrelim/PL201_3.html
†https://github.com/legacysurvey/legacypipe
‡https://github.com/dstndstn/tractor

whether or not to retain an object in the catalogue and the choice of the model to best describe its light profile is treated as a penalized-χ^2 model selection problem.

This process results in object fluxes and colours that are consistently measured across the wide-area imaging surveys that form the input into the DESI target selection. In general, TRACTOR improves the target selection for all DESI surveys by allowing information from low resolution and low signal-to-noise measurements to be combined with those from high resolution and high signal-to-noise data. The TRACTOR catalogues include source positions, fluxes, shape parameters, and morphological quantities that can be used to discriminate extended sources from point-sources, together with errors on these quantities.

1.6 Introduction to DESI BGS

The characterisation and definition of the target list for each DESI survey is a critical step for efficient survey execution and to allow reliable measurements of galaxy clustering. In this thesis I describe this process for the DESI bright galaxy survey (hereafter BGS), a flux limited sample of around 10 million galaxies, using photometry from the new imaging survey, the Legacy Surveys* (LS).

The target sample for the BGS is intended to be a galaxy sample that is flux-limited in the r-band. The magnitude limit is determined by the total amount of bright observing time and the exposure times required to achieve the desired redshift efficiency. This target selection is, in essence, a deeper version of the target selection for the SDSS MGS (Strauss et al., 2002).

To make predictions for BGS target sample we make use of the mock galaxy catalogue created from the Millennium-XXL (MXXL) N-body simulation of Angulo et al. (2012b) by Smith et al. (2017). This mock is tuned match the luminosity function, colour distribution, and clustering properties of the SDSS MGS at low

*`http://legacysurvey.org/`

redshift, and the evolution of these statistics to redshift $z \approx 0.5$ as measured from the GAMA survey (Driver et al., 2012; Liske et al., 2015; Baldry et al., 2017).

The DESI BGS is expected to have a target density of just over 800 galaxies per square degree in a primary sample defined by a faint r-band magnitude limit of 19.5. Then, in a lower priority sample, a secondary sample of $\sim$ 600 galaxies deg^{-2} defined by the magnitude range $19.5 < r < 20$ (DESI Collaboration et al., 2016). From hereon in we will refer to these BGS samples as BGS BRIGHT and BGS FAINT respectively. A few per cent of galaxies in the DESI BGS will be lost due to deblending errors, superposition with bright stars, and other artifacts that typically affect imaging catalogues. Our aim is to provide a reliable input galaxy catalogue for the DESI BGS and to characterize its properties, such as the surface density of galaxies and their clustering.

1.7 Outline of remaining chapters

In Chap. 2 we introduce the DECaLS imaging survey and made our first attempt to characterise the BGS target selection. Here we define the star-galaxy approach used for this and the subsequent chapters, as well as the photometric and spatial cuts that define the BGS selection. In Chap. 2 we assess the BGS selection for the first time using GAMA, and present the first results of clustering analysis and systematics. In Chap. 3 we present the BGS target selection for the LS DR9 and for the three surveys DECaLS, BASS, and MzLS for the first time. We included an analysis of the difference in the photometry between DECaLS and BASS/MzLS, an assessment of target selection with the help of a visual inspection web tool, the study of systematics, and a possible contamination by stars and spurious around large galaxies using angular cross-correlations. Chap. 4 include a deeper clustering analysis of the BGS target selection from LS DR9. The analysis include two-point angular correlation functions to check over the consistency between the surveys, a comparison with the MXXL light-cone catalogue with the angular clustering as

a function of r-band magnitude, and as a function of colour, and higher-order statistics using the counts-in-cells method. Finally, in Chap. 5, we present the BGS target selection for the DESI survey validation observations, and using this data, we assess and present the final BGS selection for the main survey. In Chap. 5 we define the requirements that the BGS target selection have to meet such as redshift success rate, stellar contamination, and exposure time. Chap. 6 show the conclusions of this thesis work and the future work.

Chapter 2

Characterising the BGS Target Selection with DECaLS DR8

2.1 Introduction

In this Chapter we define and characterize the BGS target selection based on the DECaLS release, DR8, which covers $\sim 2/3$ of the full 14 000 deg^2 of DESI footprint. The resulting catalogue is defined in Ruiz-Macias et al. (2020) and here we present the details of that selection and associated analysis of the catalogue.

This BGS catalogue was used by DESI in the commissioning stage of the early survey validation observations. The final BGS catalogue will be based on the next, DR9, Legacy Survey data release. Details of improvements of DR9 with respect to DR8 are given in Section 3.1.1.

This Chapter is organised as follows: in Section 2.2 we describe the Legacy Surveys imaging data used to select our targets and the secondary datasets used to tune the selection. In Sections 2.3 and 2.4 we define the spatial and photometric cuts used to select BGS targets and to get rid of artifacts that might become problematic for DESI observations plus the removal of poor quality imaging data. In Section 2.4 we define our star-galaxy classification using *Gaia* DR2. In Section 2.5

we compare the BGS catalogue with its overlap of the GAMA DR4* (Driver et al., 2012; Liske et al., 2015; Baldry et al., 2017) to assess the completeness and contamination of the BGS and to quantify its expected redshift distribution. In Section 2.5.2 we look at eight potential systematics that might be affecting our BGS target selection and try to mitigate these effects with linear weights determined using the stellar density. Section 2.5.3 shows the clustering of our BGS selection before and after applying the weights and we compare it with SDSS and the MXXL lightcone catalogue (Smith et al., 2017). Finally, in Section 2.6, we summarize our results and present our conclusions.

2.2 Photometric Data sets

During the BGS target selection process we make use of several catalogues. The main data set used is the Legacy Surveys DR8 (hereafter LS DR8, Dey et al., 2019) imaging catalogue from which we select our targets. We also make use of secondary catalogues for masking purposes, such as the TYCHO2 star catalogue (Høg et al., 2000), the *Gaia* DR2 (Gaia Collaboration et al., 2016b), the Siena Galaxy Atlas - 2020 (SGA-2020) (Moustakas in prep.) and globular clusters from the OpenNGC† catalogue. We also use a combination of *Gaia* DR2 and LS photometry to perform star-galaxy separation.

2.2.1 Legacy Survey DR8 (DECam)

This work in this Chapter is based on the eighth release of the Legacy Survey project (LS DR8) which is the first release to integrate data from all of the individual

*It is not the proper GAMA DR4 but an unreleased version of the GAMA catalogue in between DR3 and DR4 that the GAMA collaboration made available to us. This version is essentially the same as GAMA DR3, but with more redshifts. Meanwhile, the proper GAMA DR4 replaces SDSS photometry with KiDS and has a new magnitude limit.

†OpenNGC, `https://github.com/mattiaverga/OpenNGC`, is a database containing positions and main data of NGC (New General Catalogue) and IC (Index Catalogue) objects constructed by the GAVO data center team by merging data from NED, HyperLEDA, SIMBAD, and several databases available at HEASARC (`https://heasarc.gsfc.nasa.gov/`).

components of the Legacy Surveys (BASS, DECaLS and MzLS). However, this Chapter focuses only on DECaLS data.

The DECaLS data in the LS DR8 data release comprises observations from 9th August 2014 through 7th March 2019. DECam images come from the Dark Energy Camera (DECam Flaugher et al., 2015) at the 4-m Blanco telescope at the Cerro Tololo Inter-American Observatory. DECam has 62 2048×4096 pixel format 250 μm-thick LBNL CCDs arranged in a roughly hexagonal ~ 3.2 deg^2 field of view. The pixel scale is 0.262 arcsec/pix and the camera has high sensitivity across a broad wavelength range of $\sim 400 - 1000$ nm. Since LS DR8 data goes beyond the intended DESI footprint* of $\sim 14\,000$ deg^2, we are going to consider only data within the DESI footprint. This corresponds to $\sim 9\,717$ deg^2 of DECaLS data of which $\sim 1\,114$ deg^2 are covered by DECam data coming from the DES (The Dark Energy Survey Collaboration, 2005). We essentially have two DECam data sets, i) DECam imaging taken for the LS programme which we refer to as DECam LS and ii) the DECam data coming from the DES programme which we refer to as DECam DES. DECam LS and DECam DES combined to form the DECaLS data set. The needs of the DES survey required deeper limiting magnitudes than DESI. Therefore, the main differences between images taken for the DECam DES compared to images taken for DECam LS relies in the magnitude depths and the profile fitted to extract the source brightness, mostly PSF for DECam DES and an exponential profile for DECam LS. Fig. 2.1 shows the sky map coverage of DECaLS imaging indicating the DECaLS imaging that lies within the DESI footprint. DECaLS is the only survey that covers the entire SGC ($4\,394$ deg^2) and the NGC ($5\,323$ deg^2) regions of the DESI survey at declination $\delta \leq +32.375°$.

In order to fulfil the target selection required for the different DESI surveys (BGS, LRGs, ELGs and QSOs), it was concluded that a three-band g, r and z optical imaging programme, complemented by *Wide-field Infrared Survey Explorer*

*Current LS DR8 imaging covers around $\sim 20\,332$ deg^2 of which $15\,174$ deg^2 corresponds to DECaLS.

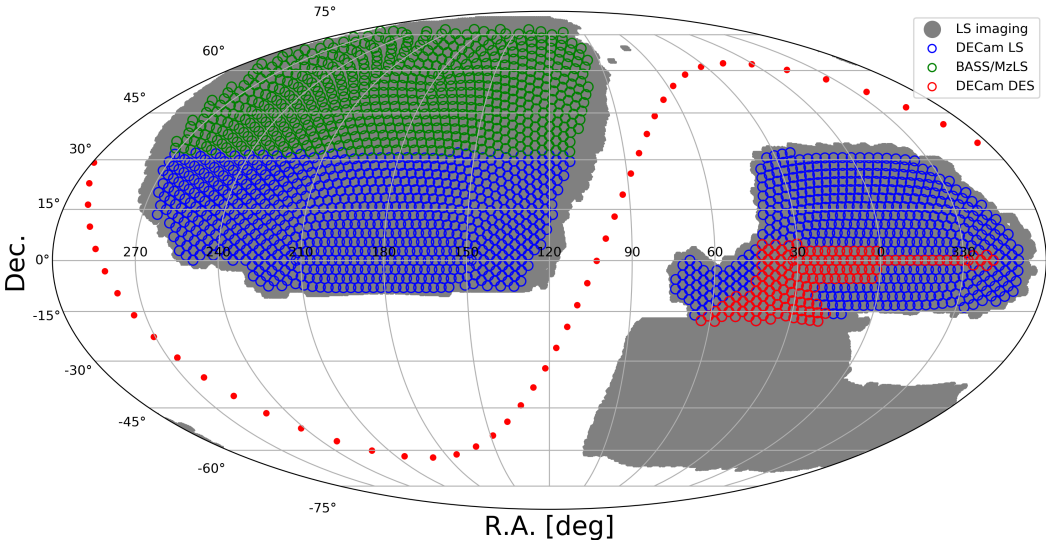


Figure 2.1: The sky map of the footprint of all the LS imaging used in DECaLS and in BASS and MzLS is shown in gray. The red and blue circles show the DESI tiles that define the portion of DESI survey footprint that lies within DECaLS. The blue tiles are those for which the data comes from the DECam LS imaging while the red tiles come from DECam DES imaging. The green tiles show the northern DESI footprint whose imaging data comes from the BASS and MzLS surveys which are not the focus of this Chapter. The red dots show the locus of the Galactic plane.

(*WISE*) W1 and W2 photometry, would be sufficient. The minimal depth* required is $g = 24.0$, $r = 23.4$ and $z = 22.5$. DECam LS reaches these required depths in total exposure times of 140, 100 and 200 sec in g, r, z respectively in nominal† conditions, typically in a minimum of two visits per field.

The BGS is flux limited in the r-band. However, since TRACTOR performs simultaneous fits in g, r and z (see Section 1.5) we also chose to impose quality cuts in the other bands as well as those in the r band when selecting the BGS targets. As described in Section 1.5, TRACTOR perform 5 different fitting models to extract sources, a point spread function (PSF), a round exponential with a variable radius (REX), an exponential profile (EXP), a de Vaucouleurs profile(DEV), and a composite of DEV and EXP profiles (COMP). These profiles are chosen based on

*The depths are defined as the optimal-extraction (forced-photometry) 5σ depths for a galaxy near the limiting depth of DESI, where that galaxy is defined to be an exponential profile with a half-light radius of $r_{\rm half} = 0.45$ arcsec.

†Here 'nominal' is defined as photometric and clear skies with seeing FWHM of 1.3 arcsec, airmass of 1.0, and sky brightness in g, r and z of 22.04, 20.91 and 18.46 AB mag arcsec^{-2}, respectively.

a penalized-χ^2 model selection problem and are widely used in posterior analysis.

The main TRACTOR outputs required for the BGS are the total fluxes* corresponding to the best-fitting source model (i.e., PSF, REX, EXP, DEV or COMP) in all three bands (g, r and z), the number of observations (NOBS) in the three bands, the predicted flux (in the r-band only) within the aperture of a fibre which is around 1.5 arcsec diameter (FIBERFLUX†) in 1 arcsec Gaussian seeing. The Galactic extinction values are derived from the SFD98 maps (Schlegel et al., 1998) and are reported in linear units of transmission (MW_TRANSMISSION) in the g, r and z bands, with a value of unity representing a fully transparent region of the Milky Way and 0 indicating a fully opaque region. The extinction coefficients for the DECam filters were computed through an airmass of 1.3, for a source with a 7 000 K thermal spectrum (Schlafly and Finkbeiner, 2011). The resulting coefficients are $A/E(B-V) = 3.995, 3.214, 2.165, 1.592, 1.211, 1.064$ in $ugrizY$. These are then multiplied by the SFD98 $E(B-V)$ values at the coordinates of each object to derive the g, r and z MW_TRANSMISSION values. Finally, in each band, there is a set of quality measures called FRACMASKED, FRACFLUX and FRACIN that quantify the quality of the data in each profile fit. We describe these in more detail in Section 2.4.4.

The fluxes returned by TRACTOR can be transformed into AB magnitudes as follows:

$$magr = 22.5 - 2.5 \log_{10}(\text{FLUX}), \tag{2.1}$$

$$mag = 22.5 - 2.5 \log_{10}(\text{FLUX/MW\_TRANSMISSION}), \tag{2.2}$$

where Eqn. (2.1) does not include the correction for Galactic extinction, unlike Eqn. (2.2). The r in Eqn. (2.1) stands for raw.

Table 2.1 shows the area covered by photometry in each of the three bands of DECaLS DR8 with 1, 2 or 3 passes. These values are just for the data within the

*The fluxes output by TRACTOR are in units called NANOMAGGIES (nMgy). A flux of 1 NANOMAGGIE corresponds to an AB magnitude of 22.5.

†The FIBERFLUX is in units of NANOMAGGIES

Table 2.1: The area, in square degrees, of DECaLS DR8 covered by at least 1, 2 or 3 passes in each of the three filters (grz) individually (first three rows), and combined (i.e. at least 1, 2 or 3 passes in each of the 3 bands; bottom row). We have restricted our results to observations within the DESI footprint as shown in Fig. 2.1.

Band/Number of Passes	≥ 1	≥ 2	≥ 3
g-band	9 687	9 454	7 769
r-band	9 686	9 422	7 569
z-band	9 686	9 487	8 036
combined	9 669	9 257	6 870

DESI footprint, as shown in Fig. 2.1. This DECaLS footprint covers a total of 9 717 deg^2. Expressed in percentages, 99.5 per cent of this area has at least one pass in all of the three bands g, r and z, 95.3 per cent has at least two passes and 70.7 per cent has at least three passes in all three bands.

2.2.2 Secondary catalogues

Here we list other catalogues that are used either to exclude regions of the sky in which the extraction of galactic sources is compromised by the presence of other objects, or to perform star-galaxy separation.

2.2.2.1 Tycho 2

Bright stars can impinge upon the estimation of the photometric properties of nearby galaxies or may even lead to the generation of spurious sources. Hence, it is prudent to simply exclude or veto regions close to known bright stars to avoid such problems. Regions near bright stars are masked out of the target catalogue using the TYCHO2 catalogue (Høg et al., 2000). The TYCHO2 catalogue contains positions, proper motions, and two-colour photometry for 2 539 913 of the brightest stars in the Milky Way.

2.2.2.2 Gaia DR2

Gaia (Gaia Collaboration et al., 2016b) is a European Space Agency mission that was launched in 2013 with the aim of observing ≈ 1 per cent of all the stars in the Milky Way, measuring accurate positions for them along with their proper motions, radial velocities, and optical spectrophotometry. The wavelength coverage of the astrometric instrument, defined by the white-light photometric G-band magnitude, is 330 - 1050 nm (Carrasco et al., 2016). These photometric data have a high signal-to-noise ratio and are particularly suitable for variability studies.

Since the first release of *Gaia* data (Gaia Collaboration et al., 2016a), this survey has been widely used by the DESI LS (i.e. for astrometric calibrations, proper motions, bright star masking) and is also ideal for constructing a star-galaxy separator for the BGS. There are 1.7 billion stars in the second *Gaia* data release (DR2)*, over the whole sky to $G = 20.7$, which is sufficiently deep to detect all stars that might contaminate the BGS FAINT sample. We describe how we use a combination of *Gaia* and LS photometry to perform star-galaxy separation in Section 2.4.1.

2.2.2.3 Globular clusters and planetary nebulae

Globular clusters and planetary nebulae are bright extended sources that can affect the identification of extragalactic sources in a similar way to bright stars. In the LS, an area of sky around such objects is excluded to minimize their impact on target selection. The OpenNGC catalogue† is used to provide a list of such sources. The extent and impact of masking around globular clusters and planetary nebulae is discussed in Section 2.3.1.1.

*DR2 covers 22 months of observations and was released on 25 April 2018.

†`https://github.com/mattiaverga/OpenNGC`

2.2.2.4 The Siena Galaxy Atlas

Large galaxy images can be broken up by photometric pipelines, which, for example, could mistake H II regions inside the galaxy for individual extended sources. Also, spurious sources could be generated around the boundaries of large galaxies. The Siena Galaxy Atlas - 2020 (SGA-2020)* is an ongoing project to select the largest galaxies in the LS using optical data from the HyperLeda catalogue† (Makarov et al., 2014) and infrared data from the ALLWISE catalogue (Secrest et al., 2015). Currently the catalogue contains 535 292 galaxies that have an angular major axis (at the 25 mag/arcsec^2 isophote) larger than 20 arcsec. The use of the SGA-2020 in the spatial mask of the BGS is described in Section 2.3.1.2.

2.3 Spatial Masking

Our main goal is to produce a reliable BGS input catalogue that fulfils the DESI science requirements. If the target list contains spurious objects, these will mistakenly be allocated fibres leading to a reduction in the efficiency and completeness of the redshift survey. Furthermore, spurious objects could imprint a systematic effect in the measured clustering.

A step towards minimising the number of spurious objects is to mask out regions of the sky around bright stars, since features such as extended halos, ghosts, bleed trails and diffraction spikes around the stars can compromise the measurement of the photometry of neighbouring objects. Similarly we must remove areas around very large galaxies and globular clusters and planetary nebulae; such objects can also affect the photometric measurements of their neighbours, leading to incorrect properties or spurious objects.

Within the same framework, we have to propagate instrumental effects such as saturated pixels, bad pixels, bleed trails, etc. that the NOAO CP tracks and

*https://github.com/moustakas/SGA
†http://leda.univ-lyon1.fr/

TRACTOR reports in the LS catalogue*

One way to avoid contamination of the catalogue with spurious objects is to exclude regions around bright stars and galaxies. This can be done with a simple but effective circular mask for stars and by using elliptical masks for galaxies. In Section 2.3.1 we set out the geometrical masking functions we have applied around bright stars, large galaxies and globular clusters to minimize the number of spurious targets in our BGS catalogue. In Section 2.3.2 we describe the masks applied to reduce the number of spurious targets due to imaging artifacts such as bad pixels resulting from saturation and bleed trails.

For subsequent analysis (e.g. estimating clustering statistics), it is very important to keep a record of the areas of the survey that are removed by these masks. For this purpose we have made use of the randoms catalogue developed by the DESITARGET† team. The randoms catalogue has a total density of 50 000 objects/deg^2 divided into 10 subsets, each with density of 5 000 objects/deg^2. Each random carries with it some of the DECam imaging information computed from the image pixel (in each band and exposure) in which it is located and supplementary information such as the dust extinction extracted from HEALPix‡ maps (Zonca et al., 2019). These imaging attributes include the number of observations (NOBS_G, NOBS_R, NOBS_Z), galactic extinction (EBV), the bitwise mask for optical data (MASKBITS), etc§.

In Fig. 2.2 we show a flow chart which summarizes the spatial masking applied when constructing the BGS catalogue. The spatial masking is broken down into two classes: *geometrical masking* and *pixel masking*. The blue boxes of the flow chart report the survey area (in deg^2) and mean target densities (in objects/deg^2) after successively applying each mask (gray hexagonal boxes). The red boxes record

*In the LS DR8 catalogue information on whether or not the photometric parameters measured for an object have the possibility of being influenced by a bad pixel is flagged by the ALLMASK MASKBITS.

†`https://github.com/desihub/desitarget`

‡`http://healpix.sourceforge.net`

§For more information on the properties of randoms see: `http://legacysurvey.org/dr8/files/#random\protect\discretionary{\char\hyphenchar\font}{}{}catalogs`

the same information for the rejected area and objects. The final BGS catalogue does not depend on the order in which the masks are applied, but as some areas and targets are rejected by more than one mask the information in the red boxes depends on the ordering. For example, the area and number of objects shown as being rejected by the pixel masking excludes what would be rejected by this mask if the geometric masks had not been applied first. Overall, for the DECaLS footprint of 9 717 deg^2, the spatial masking removes 3.25 per cent of the area.

2.3.1 Geometrical masking

2.3.1.1 Bright star mask (BS)

The bright star (BS) mask is based on the locations of stars from *Gaia* DR2 (Gaia Collaboration et al., 2018) and the TYCHO2 (Høg et al., 2000) catalogue after correcting for epoch and proper motions. This mask consists of the union of circular exclusion regions around each star, where the radius of the exclusion region, estimated from an earlier stacking analysis, depends on the magnitude of the star in the following way:

$$\begin{aligned} R_{\rm BS}(m) &= 39.3 \times 2.5^{(11-m)/3} \text{ arcsec}, \quad m > 2.9 \\ &= 471.6 \text{ arcsec}, \qquad\qquad\quad m < 2.9. \end{aligned} \tag{2.3}$$

Here m is either *Gaia* G-mag or TYCHO2 MAG_VT with *Gaia* G-mag being used when both are available. Stars fainter than $m = 13$ have no exclusion zone around them.

The BS masking uses a total of 773 673 *Gaia* DR2 objects (82 objects/deg^2) with *Gaia* G-mag brighter than 13, while from TYCHO2, we have a total of 3 349 objects ($\sim$ 0.36 objects/deg^2) to a TYCHO2 visual magnitude brighter than MAG_VT = 13. In order to avoid overlaps both catalogues have been matched after applying

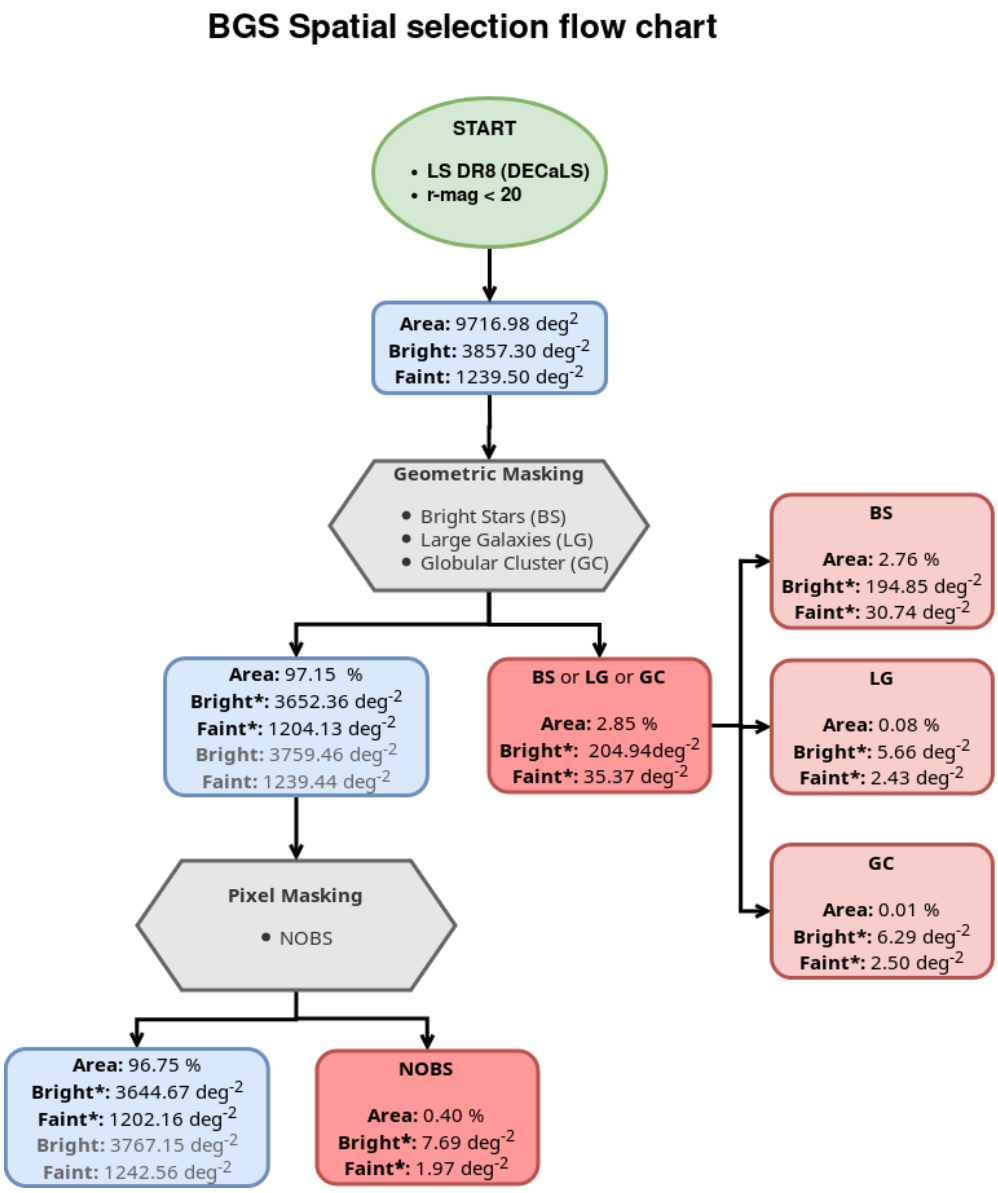


Figure 2.2: The flow chart shows the effects of the spatial masks that are applied as part of BGS target selection for the DECaLS DR8 data. The spatial masking is divided into two classes, one defined by the geometrical cuts which exclude regions around bright sources (bright stars, large galaxies and globular clusters), and the other by pixel-based cuts which use information such as the number of observations (NOBS). The boxes in the flow chart show the survey area (in deg^2) and the target number density (per square deg) split into BGS BRIGHT ($r < 19.5$) and BGS FAINT ($19.5 < r < 20$) after each mask is applied. The blue boxes give this information for the portion of the survey that is retained while the red boxes give this information for the areas removed. If more than one mask is combined at a single stage (as indicated within the gray hexagonal boxes), then the dark-red boxes show the results for the combination of these masks and the light-red boxes shows the results for each individual mask. As some of the masks can overlap the numbers in the light-red boxes do not necessarily add up to those in the dark-red boxes. The target densities with the (*) superscript are computed without correcting for the area removed by the masking while those without the (*) superscript are corrected for the masked area. The gray hexagonal boxes describe the different masks. Note that star-galaxy separation is not yet applied here and this is why we have a high target density in the blue boxes.

proper motions to bring *Gaia* objects to the same epoch as TYCHO2 and keeping only the TYCHO2 objects that are not found in *Gaia*. These TYCHO2 stars represents only a 0.4% of total stars used for the BS masking. For the magnitude, m, used to compute the mask radius in equation (2.3) is the *Gaia* G-band magnitude for the *Gaia* stars and the TYCHO2 visual magnitude, MAG_VT, for the retained TYCHO2 stars. The overall median difference between the TYCHO2 and *Gaia* magnitude is 0.4 with TYCHO2 being fainter. This 0.4 magnitude difference translates into a median decrease in masking radius of 50 arcsecs for *Gaia* stars with magnitude of 3 and a decrease of 2 arcsecs for *Gaia* stars with magnitude of 13 from equation 2.3. Within $R_{\rm BS}(m)$ TRACTOR forces all the sources it detects to be fit with the PSF profile to avoid artificially fitting diffraction spikes and stellar haloes as large extended sources. Thus any galaxies detected within $R_{\rm BS}$ will have their fluxes underestimated. Consequently to define a reliable galaxy catalogue we must veto all sources within $R_{\rm BS}$ of a bright star. In Fig. 2.2 we show that this Bright star mask covers 2.76 per cent of the initial footprint and rejects ∼195 potential BGS BRIGHT objects/deg^2 and ∼31 potential BGS FAINT objects/deg^2 when averaged over the full initial footprint. It should be noted that most of these objects are stars as star-galaxy separation has not been applied at this stage in the flow chart shown in Fig. 2.2. An alternative ordering of the flow chart with star-galaxy separation applied first is shown in Fig. A.1. There we see that for galaxies the corresponding numbers are 13.7 galaxies/deg^2 for BGS BRIGHT and 8.5 galaxies/deg^2 for BGS FAINT.

To determine if the bright star mask is adequate or whether the effects of stellar haloes causes a systematic error in the photometry of neighbouring galaxies that extends to larger radii, we plot in Fig. 2.3 the average density of BGS galaxies in the vicinity of bright stars prior to applying the bright star mask. If the photometry of galaxies has been compromised by any means, this can be seen in the galaxy number density to a fixed magnitude due to the strong dependence of galaxy number density on apparent magnitude. The term BGS galaxy refers to the BGS sample after

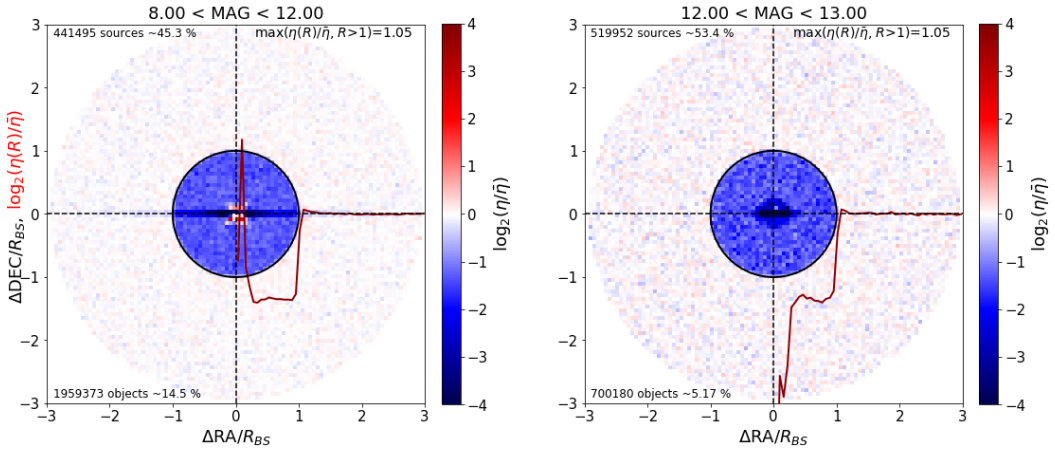


Figure 2.3: 2D histograms of the positions of BGS objects relative to their nearest Bright Star (BS) taken from the *Gaia* and TYCHO2 sources down to G-mag and visual magnitude MAG_VT of 13 respectively. These stacks are performed in magnitude bins in the BS catalogue from magnitude 8 to 12 (left) and 12 to 13 (right). The stacks are made using angular separations rescaled to the masking radius function given in Eqn 2.3, which means that objects within a scaled radius of 0 to 1 will be masked out by the BS veto while objects with $R = r/R_{\rm BS} > 1$ will not (here $r^2 = (\Delta{\rm RA}^2\cos({\rm DEC})^2 + \Delta{\rm DEC}^2)$. The colour scale shows the ratio of the density per pixel (η) to the mean density ($\bar{\eta}$) within the shell $1.1 < r/R_{\rm BS} < 3$. The density ratio is shown on a $\log_2$ scale where red shows overdensities, blue corresponds to underdensities and white shows the mean density. The black solid circle shows extent of the BS exclusion zone. The red solid line shows the radial density profile on the same scale as the colour distribution $\log_2(\eta(R)/\bar{\eta})$ where $\eta(R)$ is the target density within the annulus at radius R of width $\Delta R \sim 0.06$.

applying the star-galaxy separation and the spatial and photometric cuts down to the r-band magnitude of 20, which will be covered in the subsequent subsections of Section 2.3 and in Section 2.4. The stacks are made by expressing the angular separation, r, of the BGS galaxies prior to applying the bright star mask from their nearest bright star in units of the bright star masking radius $R_{\rm BS}$, as given by Eqn. 2.3. In these rescaled coordinates, $R = r/R_{\rm BS}$, galaxies within a radius of unity, shown by the black circle, are within the BS masking zone. We show stacks for two magnitude bins defined by the G-mag and visual magnitude MAG_VT for *Gaia* DR2 and TYCHO2 stars respectively, one with bright stars of magnitude between 8 to 12 and one fainter with magnitude between 12 to 13. The radial profile (red solid line) shows the variation in the target density, defined as $\Delta\rho(R) \equiv$

$\log_2(\eta(R)/\bar{\eta})$ where $\eta(R)$ is the target density in an annulus at radius R of width $\Delta R \sim 0.06$, and $\bar{\eta}$ is the mean target density evaluated over the region $1.1 < R < 3$. This means that $\Delta\rho(R) = 0$ corresponds to the mean density, $\Delta\rho(R) \geq 1$ to an overdensity at least twice the mean density, and $\Delta\rho(R) < 0$ to an underdensity. The large underdensity at radius $R \leq 1$ is due to TRACTOR forcing all objects within this region to be fit by the PSF model. In Section 2.4.1 we will see how stars and galaxies are defined for the BGS target selection, which does not depend on TRACTOR PSF designation, therefore, galaxies in the region $R < 1$ are allowed. In the left panel of Fig. 2.3, we see a spike of spurious galaxies for $R < 0.2$. In contrast the right panel shows a strong deficit of galaxies at $R < 0.2$. For $R > 1$, the stacks show uniform density close to mean, suggesting the star mask is working. There is a small bump just outside the masking radius where a ~ 6 per cent excess is seen in both panels. This may need to be revisited for accurate clustering studies, but is not large enough to be a concern for the efficiency of target selection.

2.3.1.2 Large galaxies mask (LG)

Without special treatment, large galaxies in which spiral arms and other structures such as H II regions are resolved would be artificially fragmented by TRACTOR into multiple sources. To avoid this and to achieve more accurate photometry for large galaxies in the SGA-2020 catalogue (see §2.2.2.4), TRACTOR is seeded with different priors, and within an elliptical mask centred on the large galaxy TRACTOR (in DR8) fits secondary detections using only the PSF model. This reduces the spurious fragmentation of large galaxy images, but also means that genuine neighbouring galaxies within the masked area have compromised photometry. The elliptical mask that is used has the same position, 25 mag/arcsec^2 isophotal major axis angular diameter, D_{25}, semi-minor to semi-major ratio, B/A and position angle, PA as the ones used to define the large galaxies in the SGA-2020 catalogue. Defining an effective masking radius of $r = \sqrt{ab}$, where a and b are the semi-major and semi-minor axes of the elliptical mask, the median masking

radius for the LG galaxies is 10.8 arcsecs.

We apply these same masks to reject objects from the BGS catalogue but then we reinstate the large galaxies provided they are not also masked by the bright star or globular cluster mask. The area covered by the combined LG mask amounts to only 0.08 per cent of the initial area and the number of objects removed amounts to 5.7 objects/deg^2 BGS BRIGHT and 2.4 objects/deg^2 BGS FAINT objects over the full initial area.

2.3.1.3 Globular cluster mask (GC)

The globular cluster (GC) mask works in a similar way to the BS mask, by applying a circular exclusion zone around the GC. The masking radius is defined by the major axis attribute for the object in the OpenNGC catalogue.

The GC mask has the smallest impact of the geometric masks, rejecting only 0.01 per cent of the initial area, accounting for densities of 6.3 objects/deg^2 in BGS BRIGHT and 2.5 objects/deg^2 in BGS FAINT. TRACTOR also force fits as PSFs everything within this mask.

2.3.2 Pixel masking

Some of the effects that compromise the photometry on a pixel basis and the model fitting include bad pixels, saturation, cosmic rays, bleed trails, and transients. The NOAO DECam CP identifies these instrumental effects during its various calibrations* (see Table 5 in Dey et al. (2019) for a list of the calibrations) and these are passed through TRACTOR and compiled in the ALLMASK BITMASK†. ALLMASK denotes a source blob that overlaps with any of the mentioned bad pixels in all of the overlapping images.

*The document that lists all the calibrations and which includes details about the various maskings can be found at: `https://www.noao.edu/noao/staff/fvaldes/CPDocPrelim/PL201_3.html`

†Details of this BITMASK can be found here: `http://www.legacysurvey.org/dr8/bitmasks/#allmask-x-anymask-x`

Besides the bad pixels which arise due to instrumental defects, the BGS requires a complete sample in the three bands (g, r and z). We therefore impose a requirement that there is at least one observation in each of the bands through the NOBS parameter. NOBS stands for Number of Observations, and is defined as the number of images that contribute to the central pixel of each detected source in each of the bands. Both ALLMASK and NOBS are pixel-based and hence this information is also available in the random catalogue. However, we find that virtually all of the area (97 per cent) (and hence virtually all of the randoms) rejected by ALLMASK is also rejected by using NOBS = 0 (in any band). In addition, ALLMASK rejects a significant number of objects (196 objects/deg^2) but with a small associated area (0.01 per cent of the full area). Virtually all the objects rejected by ALLMASK and many others are already rejected by the quality cuts in FRACMASKED, FRACIN and FRACFLUX (in any band); these cuts will be reviewed in Section 2.4.

In conclusion, there is little to be gained from using ALLMASK and we have therefore decided to use only NOBS as our pixel level mask, shrinking the area by 0.4 per cent and reducing the target density by 7.7 objects/deg^2 in BGS BRIGHT and 2 objects/deg^2 in BGS FAINT over the initial area.

2.4 Photometric selection

Following the spatial masking described in the previous section, the next step in the construction of the BGS target list is to incorporate information about photometric measurements into the selection process. According to the science requirements of the BGS and the mock BGS catalogues made by Smith et al. (2017), the survey is expected to have a target density of 800 galaxies deg^{-2} to an r-band limit of 19.5. For the faint sample ($19.5 < r < 20$), which is second priority in BGS, a density of 600 galaxies deg^{-2} is expected.

One of the major challenges for the BGS is the separation of stars and galaxies.

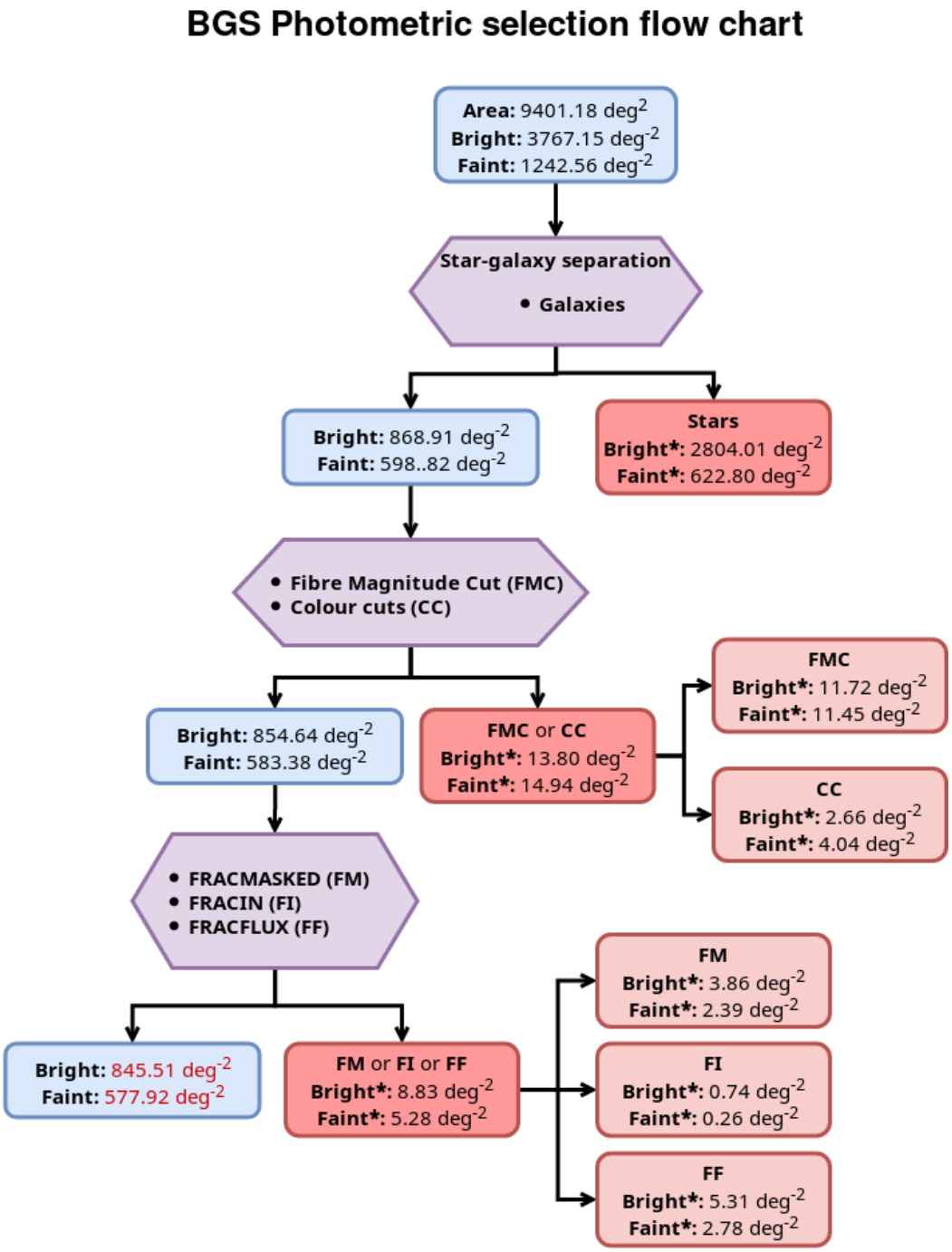


Figure 2.4: Flow chart of the BGS target selection in the Legacy Surveys DR8 based on photometric considerations. The photometric selection of BGS targets is divided into four stages; star-galaxy separation, fibre magnitude cuts (FMC), colour cuts (CC) and quality cuts (QCs). The photometric cut flow chart is a continuation of the spatial cut flow chart (Fig. 2.2) and therefore we start from the area and object densities reported at the end of the spatial cut flow chart. We report densities for the bright and faint samples separately, showing in blue boxes the values for the sources remaining after each of the BGS cuts. The densities of the removed objects are shown in red/pink boxes. The different cuts applied are shown in purple hexagonal boxes.

In Section 2.4.1 we describe how we compare high angular resolution point source magnitudes from *Gaia* DR2 (Gaia Collaboration et al., 2018) with total magnitudes from the best-fitting light profile model selected by TRACTOR to distinguish point sources from extended sources.

In Section 2.4.2 we describe how we reject spurious objects that have incongruous light profiles by comparing their total magnitudes with the fibre magnitude that TRACTOR computes from the fitted profile assuming 1 arcsec Gaussian seeing and 1.5 arcsec fibre diameter. We place a cut in the fibre magnitude versus total magnitude plane that is motivated by the locus of confirmed galaxies from the GAMA DR4 survey.

Further posterior cuts which use photometry include removing colour outliers in $g-r$ and $r-z$ (see § 2.4.3), and applying quality cuts that indicate low accuracy in the flux measurement for an object (see § 2.4.4). The quality cuts make use of the quantities FRACMASKED, FRACFLUX and FRACIN measured by TRACTOR for each object in each of the three bands (grz). These are defined and discussed in § 2.4.4.

In Fig. 2.4 we show the second part of the BGS target selection flow chart. This flow chart focuses on the photometric selection cuts and starts from where the previous flow chart (Fig. 2.2), showing the spatial cuts, left off. The BGS catalogue, in the DECaLS subregion, ends up having a reduced area of 9 401 deg^2 out of the initial 9 717 deg^2, and target densities of 846 objects/deg^2 and 578 objects/deg^2 for BGS BRIGHT and BGS FAINT respectively.

2.4.1 Star-galaxy separation

The classification of images as star or galaxies is an old problem that is of great importance when defining target catalogues for the efficient use of multi-object spectrographs. Sophisticated techniques are employed which include algorithms using machine learning methods applied to both colour and morphological inform-

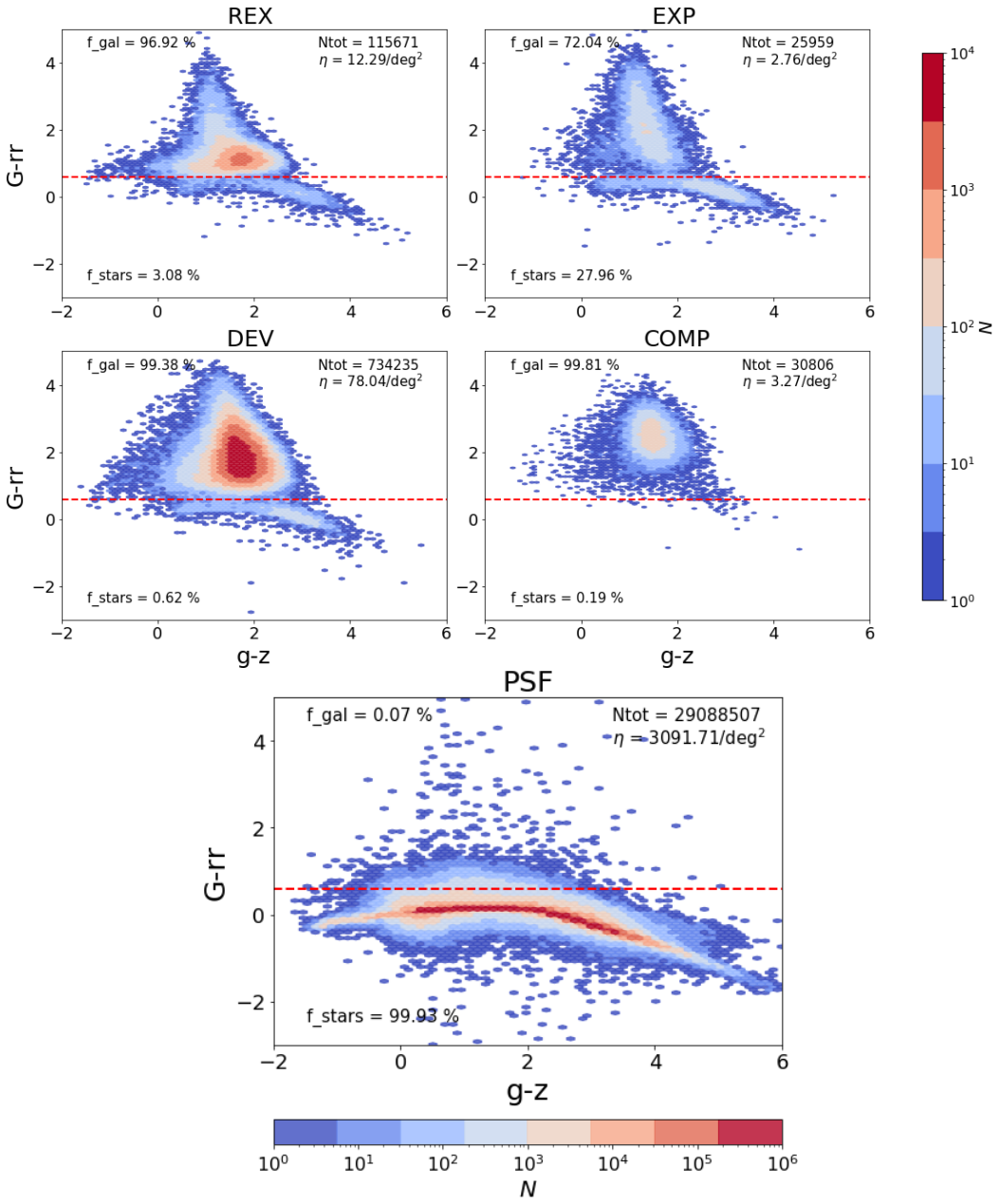


Figure 2.5: Separately for objects classified by TRACTOR as type REX, EXP, DEV COMP and PSF we show the difference between the *Gaia* (PSF) magnitude G and total non-dust corrected r-band model magnitude measured by TRACTOR, rr versus TRACTOR extinction corrected $g-z$ colour. All the objects plotted have passed the geometrical and pixel cuts detailed in Fig. 2.2, and all but the star-galaxy classification cut of the photometric-based cuts detailed in Fig. 2.4. The plots show objects that have been cross-matched between LS DR8 objects and *Gaia* DR2. Each panel shows a different morphological class, as labelled, according to the best-fitting light profile assigned by TRACTOR. The red-dashed line indicates our adopted division at $G-rr = 0.6$ with stars below and galaxies above the line. The colour in the plots shows the number counts of objects in an hexagonal cell, ranging from 1 to 10 000, except for the case of PSF-type objects, in which case the colour scale covers the range from 1 to 1 million as indicated in the colour bars. We display the fraction of galaxies and stars according to this classification at the top-left corner and bottom-left corner respectively. The total number of objects ($N_{\rm tot}$) in each plot and the target density (η) this represents is displayed in the top-right corner.

ation e.g. artificial neural networks (Odewahn et al. 1992; Bertin and Arnouts 1996), support vector machines (Fadely et al. 2012) and decision trees (Weir et al. 1995). TRACTOR uses a rigorous statistical approach to determine the best fitting light profile model to each object. In this way it classifies objects as either point sources (PSF) or extended sources (DEV, EXP, COMP or REX). However, this pipeline is not infallible and it is inevitable with ground based seeing that some compact galaxies will be misclassified as being of PSF type rather than extended. As we want to avoid incompleteness that depends on the variable seeing of the images we have instead made use of the space based high angular resolution *Gaia* photometry to distinguish point sources from extended sources. This is possible for the BGS as virtually* all stars brighter than the BGS magnitude limit of $r < 20$ are bright enough to be detected by *Gaia*.

The *Gaia* DR2 catalogue (Gaia Collaboration et al., 2018) that we use is primarily a catalogue of stars but has some galaxy and quasar contamination as reported by Bailer-Jones et al. (2019). This means we cannot simply classify all of the BGS objects that are in *Gaia* as stars. However, by comparing TRACTOR magnitude measurements with the higher spatial resolution magnitude measurements from *Gaia* we can determine which objects have extended light profiles. The *Gaia* magnitudes are computed assuming all objects are point sources. This results in accurate magnitudes for stars but magnitudes that are systematically fainter than the associated total magnitudes for sources that are extended compared to the ~ 0.4 arcsec PSF achieved by *Gaia*. In contrast, the model magnitudes computed by TRACTOR should capture more fully the total magnitude of the object. Consequently, if *Gaia* and TRACTOR magnitudes were measured in the same band, we would expect them to agree for point sources but for the TRACTOR magnitude to be brighter than the *Gaia* magnitude for extended sources. We would even expect this to be true for extended objects that TRACTOR mis-classifies as PSF since

*Over nearly all the sky *Gaia* DR2 is complete between $12 < G\text{-mag} \lesssim 20.5$, but there are very small stripes where the depth is as shallow as G-mag<17 (see 6.2 of Gaia Collaboration et al., 2018)

the wide, ground-based PSF of TRACTOR would capture more of the total flux than the narrow PSF of *Gaia*. The complication is that the *Gaia* G band is a much wider filter than the DESI r band, but as we shall see, the colour dependence is weak.

Based on these considerations we define TRACTOR objects with $r < 20$ as being galaxies if either of the following two conditions is met:

- The object is not in the *Gaia* catalogue.
- The object is in the *Gaia* catalogue but has $G - rr > 0.6$.

In the above, the G-band is the G photometric *Gaia* magnitude and rr is the raw r-band magnitude from the LS DR8 *without* applying a correction for Galactic extinction. This choice is made because the *Gaia* magnitude is not corrected for Galactic extinction. The discussion above explains that G and rr magnitudes are measured in different effective apertures and so the quantity $G - rr$ should be thought of as a measure of how spatially extended an object is and not its colour. The first criterion above is satisfied by most (93 per cent) of the BGS objects. It leaves very little stellar contamination in the BGS, as essentially any star brighter than $r = 20$ is bright enough to be detected and catalogued by *Gaia*. The second criterion is required to keep the BGS completeness high by not rejecting galaxies that are in the *Gaia* catalogue.

In Fig. 2.5 we show the $G - rr$ versus $g - z$ plane for objects in *Gaia* DR2 that are matched with objects in the LS DR8. The panels show different objects as classified by the TRACTOR model fits (i.e., PSF, COMP, DEV, EXP, REX). The cross-matched objects have been subject to all the BGS cuts (i.e. both spatial and photometric) with the exception of the star-galaxy separation itself. For objects classified by TRACTOR as PSF-type, we can see the stellar locus around $G-rr = 0$ with a weak colour dependence. For the extended sources (i.e., COMP, DEV, EXP,

REX), we see part of the galaxy locus* in the upper part of the plot, just above $G - rr = 0$.

From Fig. 2.5 we can see that the assignment of the best fitting TRACTOR model supports our *Gaia* classification using $G - rr > 0.6$, but we can still see some remnants of the stellar locus for objects that have not been assigned PSF-type by TRACTOR. For the objects classified PSF-type by TRACTOR we see in the right-most panel of Fig. 2.5 that 99.93 per cent fall on the stellar side of our $G - rr$ cut. For the objects classified by TRACTOR as the extended types (REX, DEV and COMP) the stellar contamination (i.e. objects with $G - rr < 0.6$) is at most 3.1 per cent. However, the contamination of the EXP-type objects is approaching 30 per cent.

The BGS target selection has the expected surface density after applying the star-galaxy separation. From the spatial cut flow chart in Fig. 2.4, we find a bright target density of 868.91 objects/deg^2 and a faint target density of 598.82 objects/deg^2. Rejected *Gaia* stars have a target density of 2 804.01 objects/deg^2 bright stars and 622.80 objects/deg^2 faint stars.

2.4.2 Fibre magnitude cut

In order to reduce the number of image artefacts and fragments of 'shredded' galaxies that would otherwise be classified as BGS targets we apply a cut on the fibre magnitude that is defined as a function of r-band magnitude as follows:

$$\text{rfibmag} < \begin{cases} 22.9 + (r - 17.8) & \text{for } r < 17.8 \\ 22.9 & \text{for } 17.8 < r < 20 \end{cases} \tag{2.4}$$

where rfibmag is the magnitude of the predicted r-band fibre flux and r is the total r-band magnitude, both extinction corrected. The location of this cut was guided by inspecting postage stamp images of a selection of the objects with the faintest

*We have to remember that Fig. 2.5 only includes stars and galaxies that are cross-matched between LS DR8 and *Gaia* DR2.

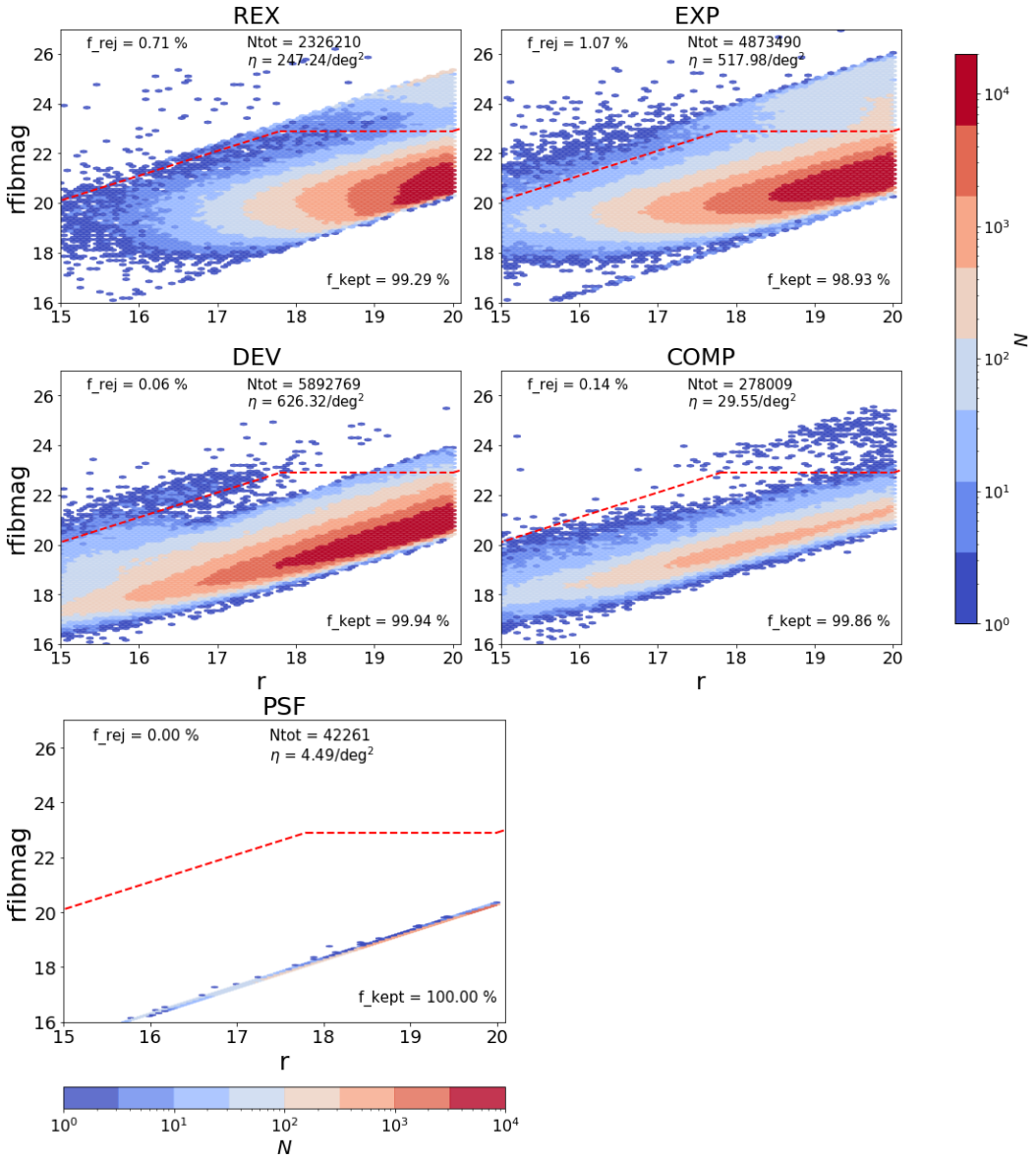


Figure 2.6: BGS galaxies in the r-band total magnitude (x-axis) versus r-band fibre magnitude (y-axis) plane in the LS DR8. The results are divided into the five different TRACTOR best-fitting light profile models, as labelled at the top of each panel. The colour bar shows the number counts of objects in an hexagonal cell covering the range from 1 to 20 000 for four of the light profile models with the exception of PSF-type galaxies, in which case the scale covers 1 to 10 000. The red-dashed line shows the fibre magnitude cut (FMC): we reject every object that is above this threshold. The numbers shown in top-left and bottom-right corners give the fraction of galaxies rejected and kept, respectively, while the number in the top-right corner shows the total number of galaxies (Ntot) and the corresponding target density (η).

fibre magnitudes with the aim of rejecting objects that appear to be artefacts while retaining nearly all of the genuine galaxies. In addition, at the bright end our threshold was guided by the location of spectroscopically confirmed GAMA galaxies, as discussed further in Section 2.5.1. Fig. 2.6 shows the distribution of the BGS objects in the rfibmag vs. rmag plane, with a separate panel for the different TRACTOR classes, and a red-dashed line indicating the location of the *fibre magnitude cut* (hereafter FMC). In the first four panels we can see that the galaxy locus has a tight core and, in general, is well below the FMC. The FMC removes 1.2 per cent of the objects classified as EXP and even smaller fractions of the other light profile classes.

All BGS objects in the PSF class lie on a stellar locus. Whether all these objects are stars or whether this is an artefact of TRACTOR only fitting the PSF model to *Gaia* sources with low astrometric excess noise (AEN) is revisited in Section 2.5.1, where we compare our classification with that of the GAMA DR4 survey. The stellar locus is also visible in the other photometric classes indicating there is some stellar contamination in our sample, but it is at a very low level.

In summary the adopted FMC rejects a further 23.17 objects/deg^2 of which 11.72 are in BGS BRIGHT and 11.45 are in BGS FAINT from the objects that have passed the previous cuts which include the rejection of stars by our star-galaxy classifier.

2.4.3 Colour cuts

An efficient way of rejecting further spurious targets from the BGS is to reject objects with bizarre colours. The limits we impose to reject outliers are:

$$\begin{aligned} -1 &< g-r &< 4 \\ -1 &< r-z &< 4. \end{aligned} \tag{2.5}$$

Fig. 2.7 shows the $g-r$ vs. $r-z$ colour-colour distribution of the objects retained in BGS if all but the colour cut (CC) were applied. The red box indicates the colour

range we keep. We can see from this plot that the locus of normal galaxy colours lies well within the range we retain and the cuts are only removing objects/artefacts with bizarre colours. It is evident that some stellar contamination remains as the stellar locus can be seen as a spur of objects with very red $r - z$ colours. However the density of objects in this spur, and its blueward extension which overlaps the galaxy locus, is no more than a few objects/deg^2 as we shall see in Section 2.5.1. The colour cuts (CC) we apply reject an additional 6.7 objects/deg^2, with 2.66 in BGS BRIGHT and 4.04 in BGS FAINT.

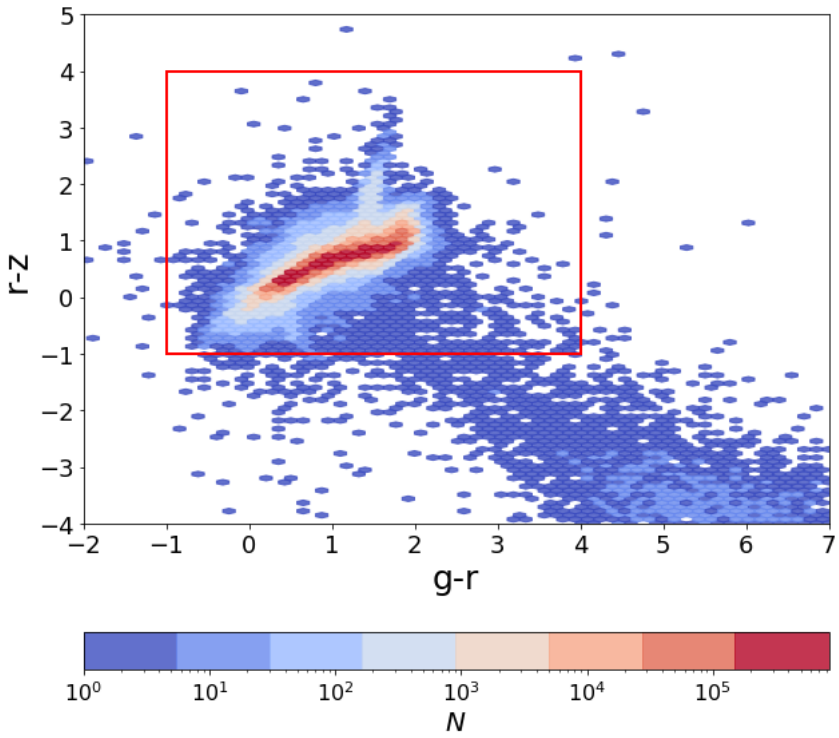


Figure 2.7: Colour-colour distribution showing $g - r$ versus $r - z$ for BGS objects without applying the CC. The colour bar shows the number counts of objects in an hexagonal cell covering the range from 1 to 800 000. The solid red box shows CC defined in Equation 2.5. Sources outside of this box are excluded from the BGS.

2.4.4 Quality cuts

Each object in the TRACTOR catalogue has three measures of the quality of its photometry recorded in each of the three bands (grz). These are:

- FRACKMASK (FM): The profile-weighted fraction of pixels masked in all observations of the object in a particular band. This quantity lies in the range $[0, 1]$. High values indicate that most of the flux of the fitted model lies in pixels for which there is no data due to masking and so the measurement is unreliable.

- FRACIN (FI): The fraction of the model flux that lies within the set of contiguous pixels (termed a 'blob') to which the model was fitted. FRACIN is close to unity for most real sources. Low values indicate that most of the model flux is an extrapolation of the model into regions in which no data was available to constrain it.

- FRACFLUX (FF): The profile-weighted fraction of the flux from other sources divided by the total flux of the object in question. FRACFLUX is zero for isolated objects but can become large for faint objects detected in the wings of brighter objects that are nearby.

Once the other cuts have been applied, in particular, the cut on NOBS and the BS mask, the distribution of each of these quantities is tightly peaked around the favoured values of FRACMASKED ≈ 0, FRACIN ≈ 1 and FRACFLUX ≈ 0. However, each quantity has a distribution with a fairly featureless tail that extends out to less desirable values. There are also clear correlations between the three quantities for a given photometric band and in some cases between photometric bands. The choice of the best set of thresholds to reject outliers is not trivial. We have adopted the following quality cuts (QCs):

$$\begin{aligned} \text{FRACMASKED\_}i &< 0.4, \\ \text{FRACIN\_}i &> 0.3, \\ \text{FRACFLUX\_}i &< 5, \qquad \text{where } i = g,\ r \text{ or } z, \end{aligned} \tag{2.6}$$

based on visual inspection of postage stamp images.

As mentioned in Section 2.3.2, we find that the objects flagged by the TRACTOR quantity ALLMASK are essentially a subset of the objects that are rejected by applying the quality cuts listed in Eqn. 2.6. While cutting on ALLMASK would have the advantage that it could also be applied to the randoms, we find that it is important to apply the QCs to remove spurious objects that are missed by the other cuts. For instance, some spurious objects that are outliers in either the fibermag vs. mag plane or in the colour-colour space that just pass the FMC and CC are removed by considering FRACMASKED or FRACIN.

As shown in the flow chart, Fig. 2.4, the QCs reject an additional 14.11 objects/deg^2 of which $\sim$ 60 per cent are removed by FRACFLUX, $\sim$ 45 per cent by FRACMASKED and $\sim$ 7 per cent due to FRACIN. The overlap between the FRACMASKED, FRACIN and FRACFLUX cuts is minimal, with only 1.05 objects/deg^2 for objects with $r < 19.5$, and in round 0.15 objects/deg^2 for objects with $19.5 < r < 20$ being rejected by more than one of the cuts. Separately for BGS BRIGHT and BGS FAINT, we show the target density of objects rejected by these cuts after applying all the previous cuts. The largest overlap between these cuts is between FRACMASKED and FRACFLUX for BGS BRIGHT, but even here it amounts to less than 1 object/deg^2. For BGS FAINT this overlap is small, 0.11 object/deg^2, and there is no overlap with FRACIN.

In Appendix A.1 we present another version of the selection cut flow chart in which the cuts are applied in a different order. There we give a galaxy view of the target selection by first applying the star-galaxy classification so that all the subsequent cuts apply only to galaxies. The final selected sample which comprises of 845.5 galaxies/deg^2 in BGS BRIGHT and 577.9 galaxies/deg^2 in BGS FAINT, is exactly the same, as the order of the cuts does not matter. The objects rejected by each filter, however, does change as many objects are rejected by more than one filter. To illustrate this point we have also swapped the order of the FMC and QCs cuts so one can see how these influence one another.

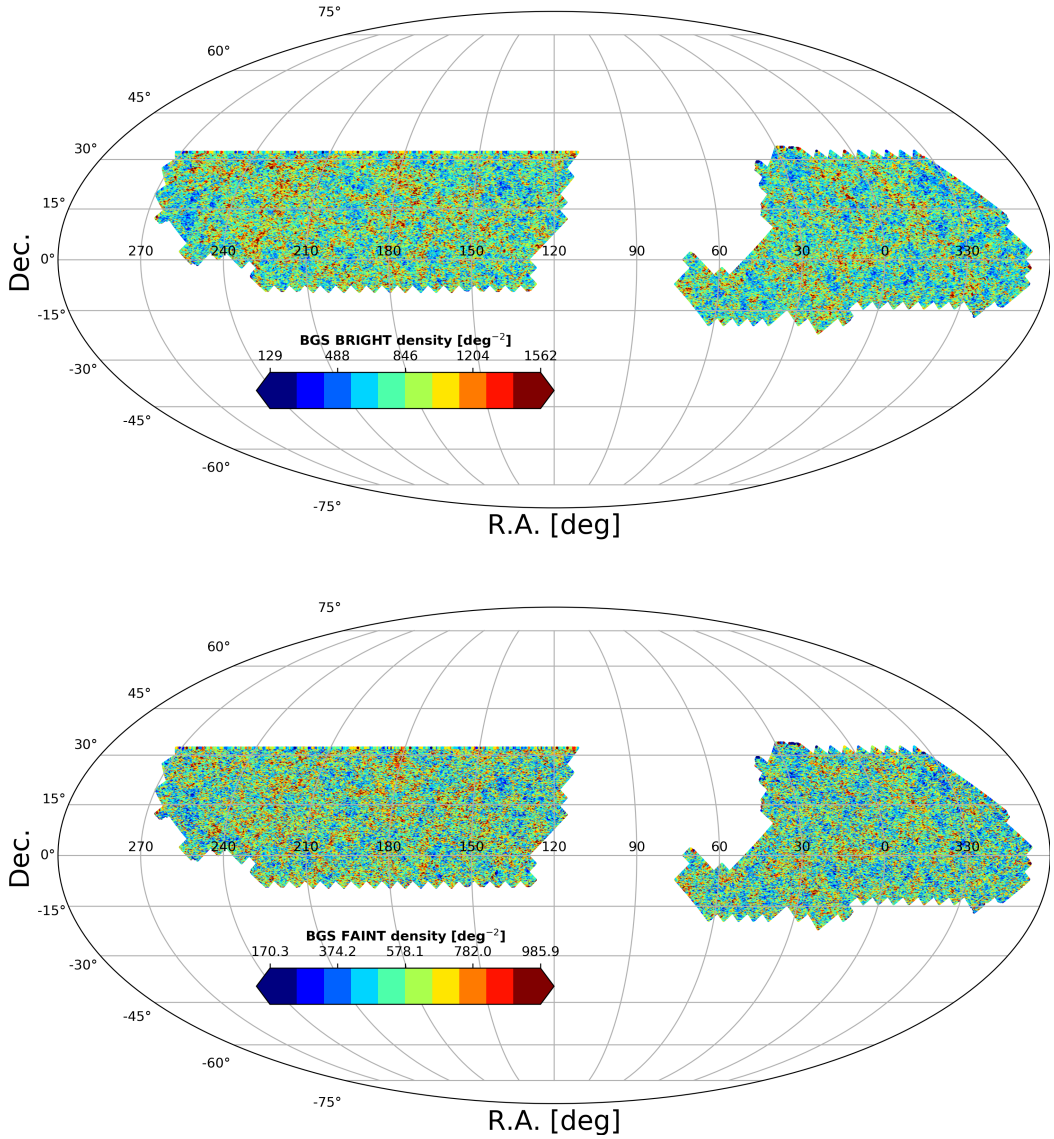


Figure 2.8: The distribution on the sky of the BGS BRIGHT (upper map) and BGS FAINT (bottom map) target density in objects/deg^2, computed on a HEALPix grid with a resolution of $N_{\text{side}} = 256$. The mean densities are 846 and 579 objects/deg^2 for the bright and faint BGS respectively.

Table 2.2: The BGS target densities for each of the TRACTOR best-fitting photometric models. The first column labels the photometric model. The next three columns list the surface density of objects per deg^2 for the BGS BRIGHT and BGS FAINT samples separately and their combined sum. The area covered by the DECaLS portion of the BGS is $9,401$ deg^2.

Model	$\eta_{\rm bright}$ [deg^{-2}]	$\eta_{\rm faint}$ [deg^{-2}]	$\eta_{\rm overall}$ [deg^{-2}]
DEV	427	202	629
EXP	284	230	514
REX	104	141	246
COMP	27	3	31
PSF	3	2	5
Total	846	578	1423

2.5 Catalogue properties

The final BGS catalogue in the DECam region in the South Galactic Cap (SGC) covers the declination range $-17 \lesssim \mathrm{DEC} \lesssim 32$ degrees, and in the North Galactic Cap (NGC) the range $-10 \lesssim \mathrm{DEC} \lesssim 32$ degrees. The BGS has a total of $13,378,062$ galaxies of which $7,944,975$ are in BGS BRIGHT and $5,433,087$ are in BGS FAINT. The total area covered by the BGS in the DECaLS subregion defined by the footprint of the tiles in Fig. 2.1 and after accounting for the spatial cuts described in Section 2.3 is $9\,401$ deg^2. In Table 2.2 we list the target density of the BGS catalogue for each of the best-fitting photometric models used in TRACTOR.

In Fig. 2.8 we show the BGS BRIGHT and BGS FAINT sky map densities computed with the HEALPix scheme using

$$\begin{aligned} \eta_i &= N_i^{\rm BGS}/A_{\rm eff}, \\ A_{\rm eff} &= N_i^{\rm R}/\eta^{\rm R}, \end{aligned} \tag{2.7}$$

where for each pixel $N_i^{\rm BGS}$ is the number of BGS targets, $A_{\rm eff}$ is the effective area computed from the number of randoms, $N_i^{\rm R}$, and the total surface density of the randoms, without any masking, is $\eta^R = 15,000$ objects/deg^2. We use a HEALPix grid of $N_{\rm side} = 256$ giving a pixel area of $A_{\rm pix} = 0.052$ deg^2. The appearance of the density fluctuations is very similar in the two disjoint regions and show no variation

with galactic latitude. We look more closely at systematic variations in the target density in Section 2.5.2.

2.5.1 Cross-comparison with GAMA

The main target sample in GAMA (Baldry et al., 2017) is a complete sample of galaxies with SDSS Petrosian r-band magnitude brighter than $r = 19.8$. The Petrosian magnitude is measured within a circular aperture of twice the Petrosian radius, where the radius is computed using the r-band surface brightness profile (Adelman-McCarthy et al., 2008). The GAMA photometric selection is very similar to that of DESI BGS and so we expect a very similar redshift distribution as GAMA which has median of $z = 0.2$ and a 90 percentile value of $z = 0.5$.

Star-galaxy separation in GAMA was conservative in that it aimed for very high completeness at the expense of some stellar contamination. These properties combined with its very high spectroscopic completeness (high quality redshift have been obtained for more 98.85 per cent of the GAMA targets) make it a nearly ideal "truth table" from which to assess the completeness of the BGS target selection and measure the expected redshift distribution of the BGS BRIGHT sample. Below we make use of GAMA to examine various aspects of our BGS catalogue. In Sec. 2.5.1.1 we compare the r-band phototometry of the matched objects and determine the redshift distribution of the BGS galaxies that match with galaxies in the GAMA survey. Section 2.5.1.2 explores an issue related to TRACTOR only providing PSF photometry for some of the BGS galaxies. In Section 2.5.1.3 we assess incompleteness in BGS relative to GAMA and quantify how much is caused by each of the various geometric and photmetric selections.

2.5.1.1 Magnitude definition and redshift distribution

We match the GAMA Main Survey DR4 galaxy catalogue (Driver et al., 2012; Liske et al., 2015; Baldry et al., 2017), which is defined by a Petrosian magnitude

(R_PETRO) limit of $r = 19.8$, to the BGS target catalogue. We use a maximum linking length of 1 arcsec to match them. The mean separation of the matches we find is 0.093 arcsec with a 1σ dispersion of 0.091 arcsec. We focus on three of the five GAMA fields: G09, G12, G15. We omit G02 as this GAMA field is only partially within the DECaLS footprint, and G23 is far to the south. The redshift completeness of the main GAMA survey is extremely high in the sense that 98.85 per cent of the objects in the catalogue yield redshifts with a quality flag NQ ≥ 3.

The GAMA spectroscopic redshifts can be used to reliably reject stars with a cut at $z = 0.002$. In what follows we restrict our GAMA catalogue to the spectroscopically confirmed galaxies ($\sim$ 98 per cent of the full catalogue). The area of each of the GAMA fields considered is 59.98 deg^2 which means that our matched sample has a total area of ~ 180 deg^2. The overall density of sources that are cross-matched between BGS and GAMA galaxies is ~ 970 objects/deg^2 with a mean redshift of $z = 0.224$.

For this matched catalogue, Fig. 2.9 compares the DR8 r-band total magnitude ($r_{\rm LS}$) with the Petrosian r-band magnitude from GAMA ($r_{\rm GAMA}$) by plotting $r_{\rm LS} - r_{\rm GAMA}$ vs $r_{\rm GAMA}$. To see how this difference depends on galaxy morphology, we divide the LS galaxies into the five photometric classes assigned by TRACTOR. In each panel we show the fraction of matched galaxies in each TRACTOR model fit class; DEV and EXP classes together make up 80 per cent of the sample and the PSF class just 2.5 per cent. We mark on the plot the $r_{\rm LS} < 20$ limit of BGS, but note this has not been applied when defining the LS sample that was matched to GAMA.

Differences in the effective passbands of the r-band filters of the LS and SDSS result in offsets in $r_{\rm LS} - r_{\rm GAMA}$ of around -0.05 and -0.1 for blue and red galaxies respectively (Dey et al., 2019). One also has to consider the difference in magnitude definitions which contributes the more to this magnitude offset. To the extent that the best fit profiles accurately describe the actual light profiles of the objects, LS provides total magnitudes. In contrast, the SDSS Petrosian magnitudes used by

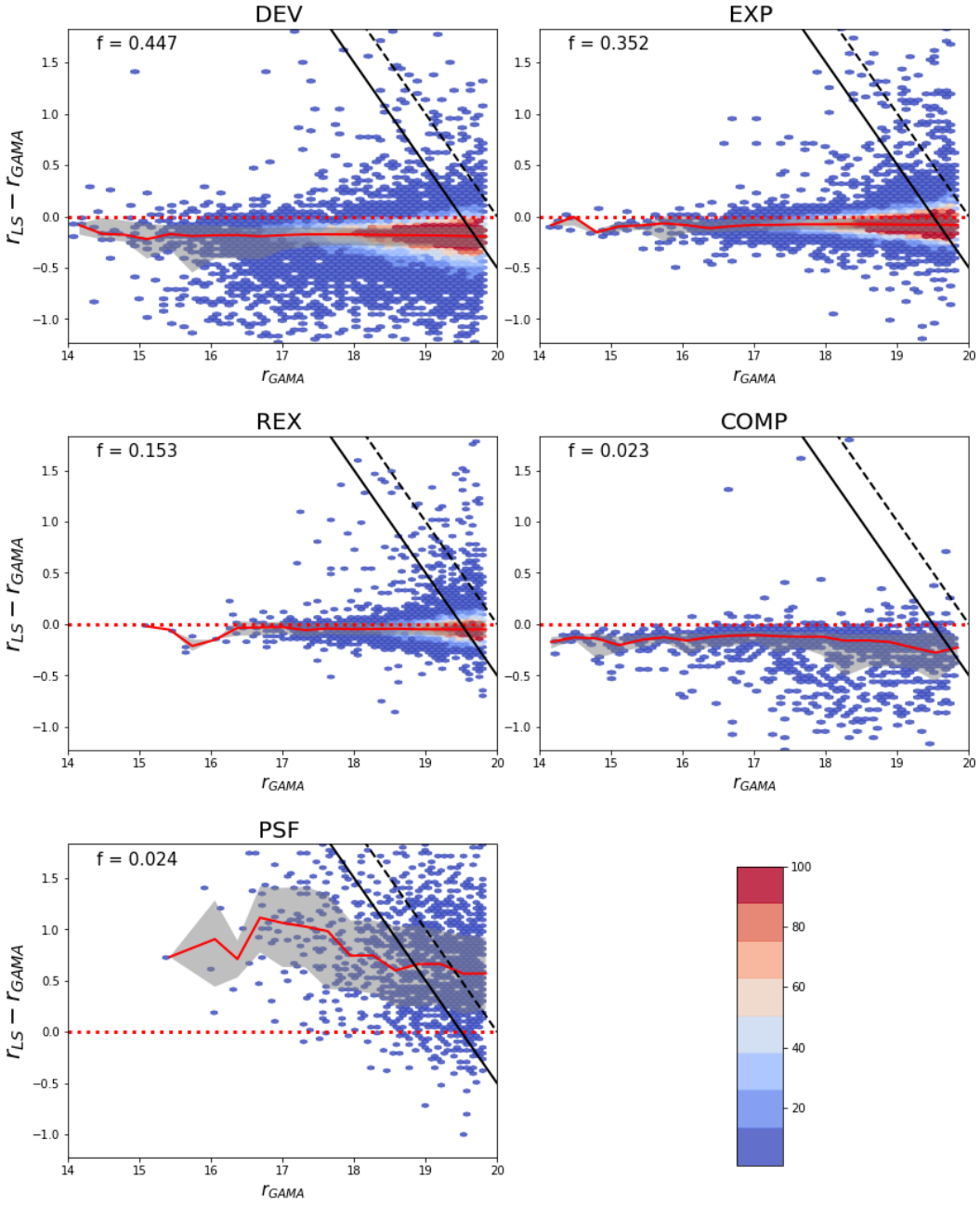


Figure 2.9: The r-band total magnitude in the LS ($r_{\rm LS}$) vs the SDSS r-band Petrosian magnitude in GAMA ($r_{\rm GAMA}$) for LS DR8 objects cross-matched with GAMA. Each plot corresponds to one of the five photometric model fits assigned by TRACTOR. The red solid line shows the median value of $r_{\rm LS} - r_{\rm GAMA}$ as function $r_{\rm LS}$; the gray shading shows the 20 to 80 percentile range; the dashed black line shows the limiting magnitude of $r_{\rm LS} = 20$ for BGS and the solid black line shows limiting magnitude of $r_{\rm LS} = 19.5$ for BGS. The colour bar shows the number counts of objects in an hexagonal cell covering the range from . The fraction of LS DR8 objects plotted out of the total number matched with GAMA is shown in the top-left corner of each panel.

GAMA quantify only the flux within twice the Petrosian radius (Blanton et al., 2001). The fraction of the flux within this aperture depends on the light profile. For EXP profile it captures 99.4 per cent, but for the DEV profile, which is more sharply peaked but with broader wings, only 82 per cent is captured. It is these differences in definition which largely drive the differences in median offsets we see in the DEV, EXP, REX and COMP classes. In all these cases the LS magnitude is brighter (more negative) than the GAMA magnitude with median offsets being -0.085 magnitudes for EXP and -0.188 magnitudes for DEV. In contrast for the PSF case the median $r_{\rm LS} - r_{\rm GAMA}$ is positive, which means that the LS PSF model magnitude captures less flux than the GAMA Petrosian magnitude. For true point sources we would expect these two magnitudes to be almost equal. The positive difference appears to happen because TRACTOR force fits PSF models to sources that are actually extended (deemed extended by our *Gaia* based star-galaxy separation) and consequently underestimates their fluxes. The reason this happens is discussed in Section 2.5.1.2.

If we take account of the scatter between the BGS and GAMA magnitudes we can use GAMA to assess the level of contamination in the BGS catalogue. If we treat GAMA as being a 100 per cent complete galaxy catalogue then any objects in BGS that are not in GAMA would be contamination in the form of stars or image artefacts. This is not true at $r = 20$ as here some BGS objects will not be in GAMA simply because of the $r_{\rm petro} < 19.8$ magnitude limit in GAMA. This can be seen in Fig. 2.9 from the location of the $r_{\rm LS} = 20$ dashed line relative to where the GAMA data truncates at $r_{\rm GAMA} = 19.8$. To avoid this problem if we apply a brighter magnitude limit $r < r_{\rm lim}$ to BGS then for a broad range of $18.5 \lesssim r_{\rm lim} \lesssim 19.3$ we find that ~ 3 percent of BGS objects are not matched with GAMA galaxies. This sets an upper limit (in this magnitude range) of 3 per cent contamination in BGS as GAMA itself may not be 100 per cent complete.

Fig. 2.10 shows the distribution of redshifts for BGS objects that have been cross-matched with GAMA galaxies. The overall distribution is shown along with

those for the BGS FAINT and BGS BRIGHT. We expect this distribution to be representative of the BGS BRIGHT sample as we can see from Fig. 2.9 that incompleteness caused by the GAMA magnitude limit to be very small. However the redshift distribution plotted for BGS FAINT is more strongly affected by the GAMA magnitude limit and its true redshift distribution is expected to be more extended.

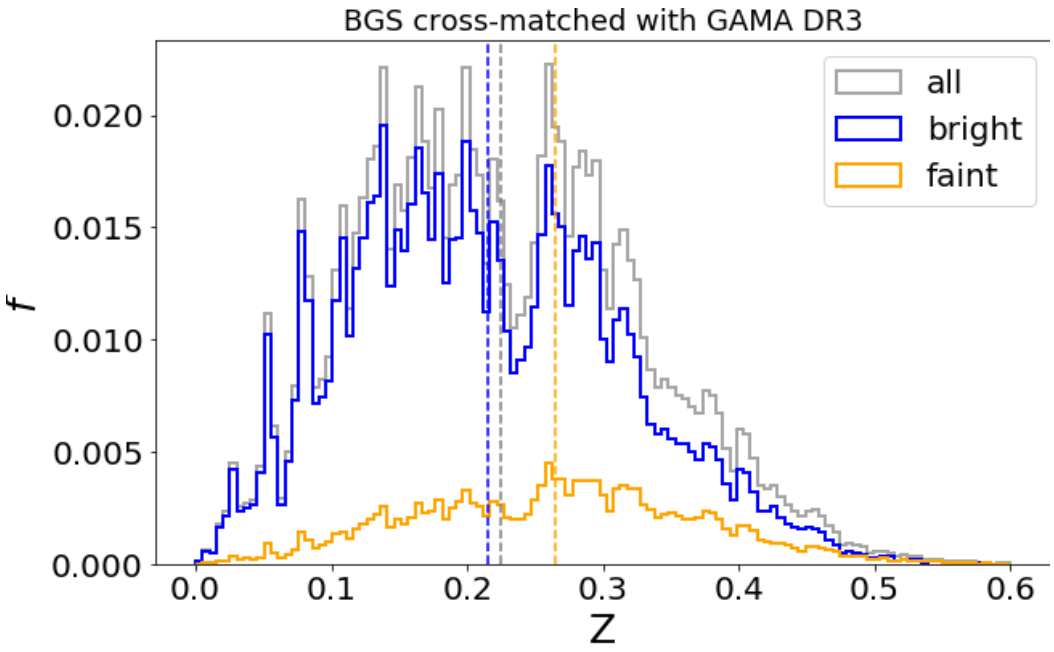


Figure 2.10: The redshift distribution of BGS objects cross-matched with GAMA DR4 broken into bright ($r < 19.5$, blue) and faint ($19.5 < r < 20$, orange) galaxies according to the BGS r-band. The gray histogram shows the overall redshift distribution of BGS galaxies cross-matched with GAMA. The mean redshift values for each distribution are: 0.215 for the bright sample (dashed blue line), 0.265 for the faint sample (dashed orange) and 0.224 for all galaxies (dashed gray).

2.5.1.2 Galaxies with TRACTOR type PSF

To avoid stars being classified as extended sources TRACTOR uses a catalogue of stars from *Gaia* to pre-select a set of objects on which it will only allow PSF fits. The *Gaia* objects for which it does this are based on the following cut on the *Gaia*

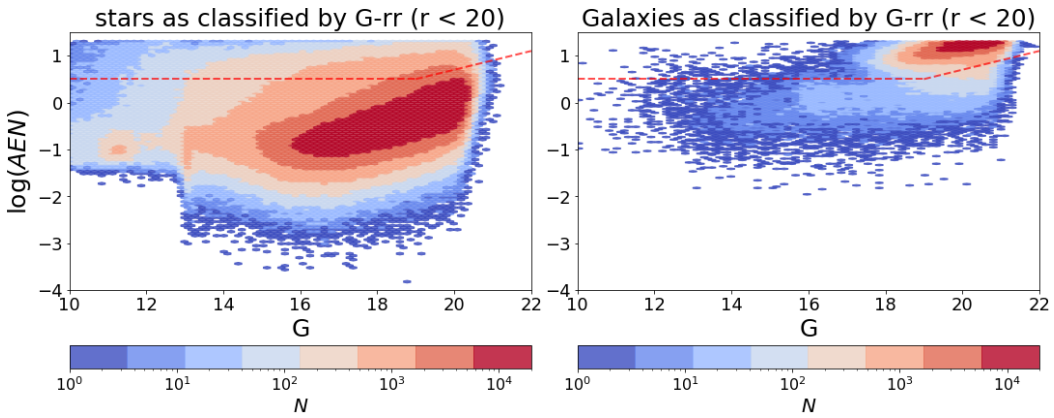


Figure 2.11: The *Gaia* Astrometric Excess Noise parameter (AEN) versus G-band magnitude. The top panel shows *Gaia* objects classified as stars by BGS and the bottom those classified as galaxies. Both plots only show *Gaia* objects with magnitud limit of $r < 20$. The red dashed-line represents the threshold limit for the AEN classification used in TRACTOR, therefore everything below the line is a star and everything above is a galaxy according to the AEN classification. The colour bar shows the number counts of objects in an hexagonal cell covering the range from 1 to 20 000.

astrometric excess noise parameter , AEN,

$$\begin{aligned} \text{AEN} &< 10^{0.5}, \quad G \leq 19 \\ \text{AEN} &< 10^{0.5+0.2(G-19)}, \quad G \geq 19, \end{aligned} \tag{2.8}$$

where G is the *Gaia* photometric G-band. The AEN can be used as measure of whether a source is extended as for extended sources the astrometric measurements are noisier than one would expect for a point source.

In contrast, in the BGS we use the difference between the *Gaia* G-band magnitude and the TRACTOR raw r-band magnitude, rr, (not corrected for extinction) as a measure of how extended the object is (see Section 2.4.1). In Fig. 2.11 we have plotted log(AEN) versus G separately for objects classified as stars and galaxies by our $G-rr$ classifier. The threshold adopted by TRACTOR can be seen to separate the bulk galaxies from the stars. For 96 objects/deg^2 the two criteria agree the object is a galaxy, but the distributions are extended and the agreement is not perfect. There are 36 objects/deg^2 that the AEN criterion classifies as galaxies which $G - rr$ classifies as stars. More problematic are the 5 objects/deg^2 that

Table 2.3: The surface density of PSF-type objects in the BGS in the G09, G12 and G15 GAMA fields combined before ($\eta_{\rm BM}$) and after ($\eta_{\rm AM}$) cross-matching with GAMA (top half of table). The bottom half of the table shows the surface density and percentage of objects in disjoint subsamples of the PSF-type BGS sample, as listed in the first column: objects that are not in *Gaia*, objects that the AEN scheme classifies as stars and those that the AEN scheme classifies as galaxies.

Sample	$\eta_{\rm BM}$ [deg^{-2}]		$\eta_{\rm AM}$ [deg^{-2}]	
PSF-type BGS	4.10		1.76	
Subsample	$\eta_{\rm BM}$ [deg^{-2}]	$\%_{\rm BM}$	$\eta_{\rm AM}$ [deg^{-2}]	$\%_{\rm AM}$
Not in *Gaia*	1.72	42.0	0.04	2.3
Gaia AEN star	2.26	55.2	1.69	96.4
Gaia AEN galaxy	0.11	2.8	0.02	1.3

the AEN criterion classifies as stars which $G - rr$ classifies as galaxies. This is an issue as it means some objects that are classified as galaxies in the BGS are treated by TRACTOR as stars and only have a PSF light profile fitted. Overall in the BGS there are 5 objects/deg^2 with PSF type within the DECaLS footprint (see Table 2.2). These objects have fibre magnitudes that are consistent with the locus of stars in Fig. 2.6 which makes us question if they really are galaxies. We investigate this below by making use of GAMA to determine whether or not they are galaxies.

First, we restrict our attention to the 180 deg^2 of our matched GAMA catalogue. The BGS PSF-type galaxies (main sample) have a density of 4.10 objects/deg^2, somewhat less than the 5 objects/deg^2 which is the average over the full DECaLS area. This reduces further to 1.76 objects/deg^2 after cross-matching with GAMA. We further subdivide these two cases (BGS PSF type and BGS PSF type cross-matched with GAMA) into three disjoint sub samples: i) those that are not in *Gaia*, ii) those that are in *Gaia* and which are classified using the AEN value as stars, and iii) those that are in *Gaia* and which are classified using the AEN value as galaxies.

The subsample sizes are reported in Table 2.3, where we give the surface density of objects before and after the cross-match with GAMA ($\eta_{\rm BM}$ and $\eta_{\rm AM}$) along with

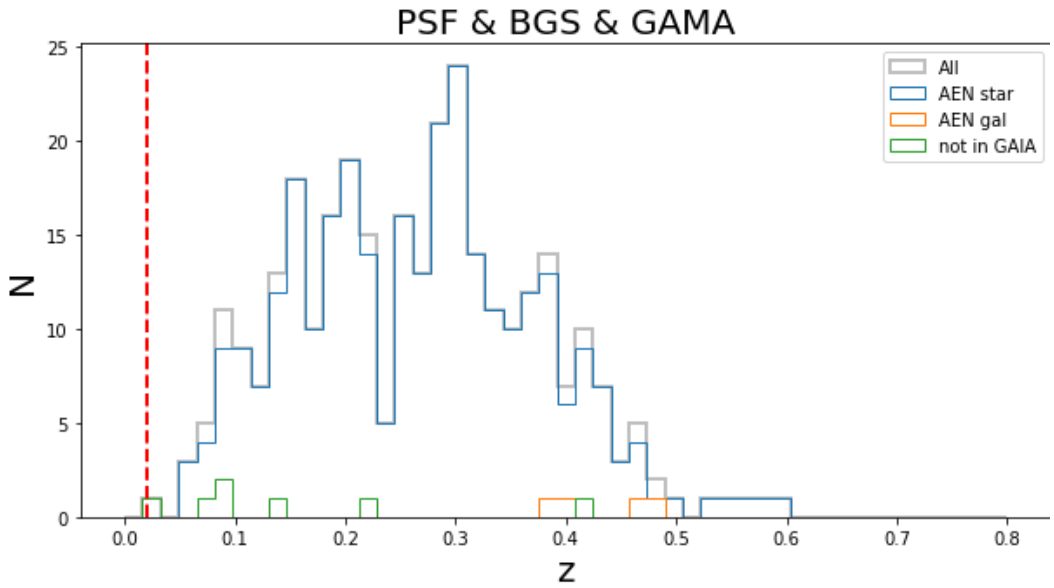


Figure 2.12: Redshift distribution of PSF-type BGS galaxies cross-matched with galaxies from three GAMA fields (G09, G12, G15). Redshifts are taken from GAMA DR4. The four distributions correspond to the matched sample (gray) and the disjoint subsamples comprising galaxies not in *Gaia* (green), and stars (blue) and galaxies (red), as defined by the AEN classification. The red dashed line marks the redshift $z = 0.002$; objects with redshifts smaller than this are stars.

the percentage of the total number of objects represented by each subsample. This shows that $\sim$ 96 per cent of the BGS PSF-type cross-matched with GAMA are *Gaia* AEN stars, which represents the $\sim$ 55 per cent in the non-matched sample. For the remaining 45 per cent in the non-matched sample, GAMA is not reliable to assess this as only 3.6 per cent of those are cross-matched with GAMA. Fig. 2.12 shows the GAMA redshift distribution for the BGS PSF-type cross-matched with GAMA broken into the three clases shown in Table 2.3. These objects shown a redshift distribution very similar to that of the full BGS sample. The reason for this mis-classification lies in the fact that for objects classified by the *Gaia* AEN criterion as stars TRACTOR only fits PSF models. For the galaxies that this *Gaia* AEN criterion falsely classifies as stars TRACTOR underestimates the total flux of the galaxy resulting in the offset with the GAMA photometry we saw in the PSF panel of Fig. 2.9 and putting these galaxies close to the stellar locus in Fig. 2.6.

2.5.1.3 Incompleteness of BGS relative to GAMA

To the depth of GAMA we can assess the completeness of the BGS catalogue by cross-matching the full depth LS DR8 catalogue with GAMA DR4. This cross-match yields a catalogue of 1011 objects/deg^2 which represents of 99.6 per cent of the GAMA catalogue. Visual inspection reveals some of the remaining 0.4 per cent are deblending issues where GAMA fragments a galaxy into two objects while TRACTOR keeps it as a single object. Of the matched objects 970 objects/deg^2 are in BGS while the other 41 objects/deg^2 are excluded from the BGS catalogue by one or other of our selection cuts.

Due to the scatter between SDSS r-band Petrosian magnitude used by GAMA and the TRACTOR model magnitude used by BGS (see Fig. 2.9), the BGS $r_{\rm LS} = 20$ magnitude limit excludes 20 faint GAMA galaxies per square degree. This leaves 20.8 objects/deg^2 in GAMA that are missing from the BGS. Whether this represents potential problematic incompleteness in BGS or just a difference in sample definition depends on which selection cuts remove the objects. We quantify and discuss this below.

The diagonal elements in Fig. 2.13 indicate the number density of spectroscopically confirmed GAMA galaxies missing from the BGS catalogue as result of each of the following spatial and photometric cuts: the bright star mask (BS); the large galaxy mask (LG); the number of observations (NOBS); star-galaxy classification (SG); fibre magnitude cut (FMC); colour cut (CC); the FRACMASKED quality cut (QCs FM); the FRACIN quality cut (QCs FI); the FRACFLUX quality cut (QCs FF). The off-diagonal entries in Fig. 2.13 show the surface density of GAMA galaxies that are removed by both of the two cuts indicated by the labels on the x and y axes.

The objects removed by the spatial BS and NOBS cuts are benign in that they do not affect BGS clustering measurements. These spatial masks are uncorrelated with BGS galaxy positions and so can be fully accounted for in clustering analyses

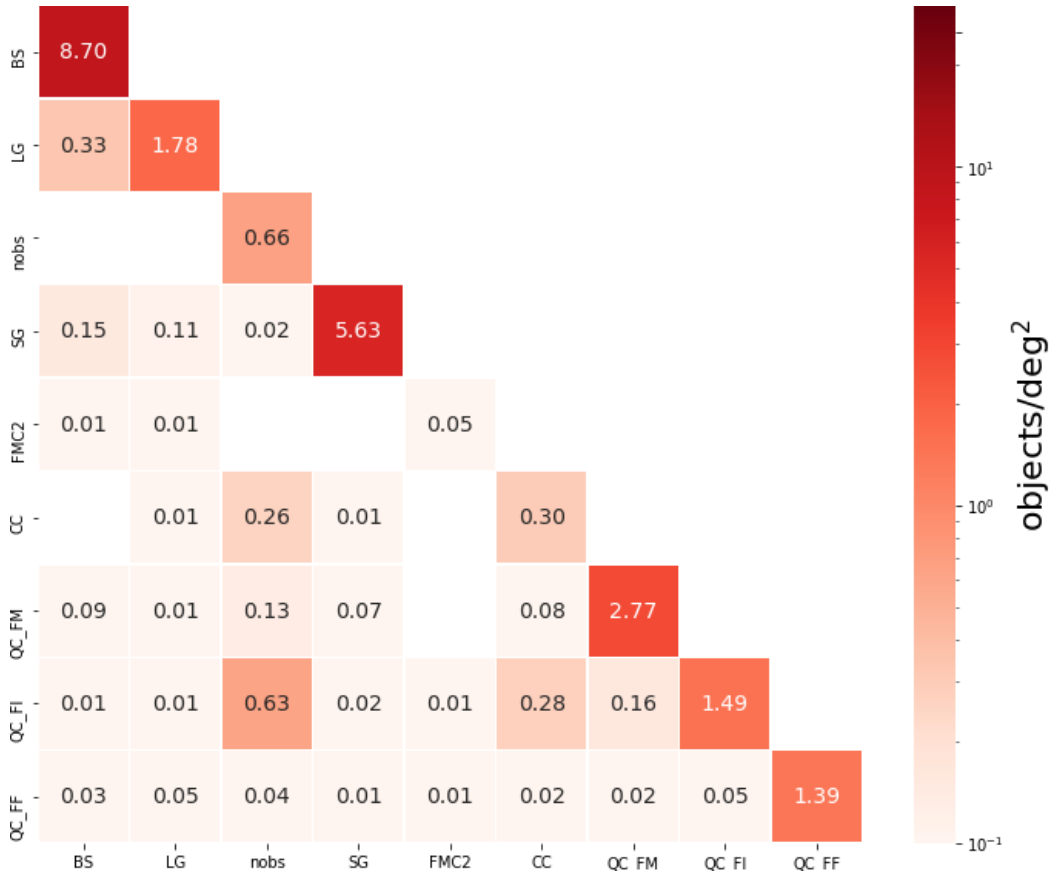


Figure 2.13: Heatmap showing the target density of GAMA galaxies ($z > 0l002$) that are missed in the BGS. The diagonal shows the number of objects per square degree removed by each of the individual spatial and photometric cuts applied in the BGS while the off-diagonal entries show the densities of objects removed by both cuts labelled on the x and y axes.

by applying the same masks to the random catalogue. The values given in Fig. 2.13 show that these two masks have no overlap and together remove 9.36 objects/deg^2.

Applying these two spatial cuts leaves us with 11.43 galaxies/deg^2 that are in GAMA but are missed by BGS. The cuts that remove these objects are almost completely independent. 5.36 objects/deg^2 are removed by the our SG classification. These objects are close to the cut imposed for the *Gaia* star-galaxy separation ($G - rr = 0.6$), but fall on the stellar side. We find that 98 per cent of these missed GAMA galaxies are classified as stars according to the *Gaia* AEN condition, which means that their photometry has been compromised as TRACTOR only fitted PSF models. If these are extended objects, then their flux as reported by TRACTOR

is a fraction of what it should be and hence their rr-magnitude is shifted to fainter values. This results in BGS galaxies shifting to lower values of $G - rr$, moving them out of the galaxy locus and into the stellar one. If the flux from these PSF-fitted galaxies were correct we would expect the residual incompleteness to be 6.07 galaxies/deg^2, equivalent to $6.07/970 = 0.62$ per cent. The proportions of this produced by the LG QCs FM, QCs FI and QCs FF cuts are 23.5, 41.2, 13.8 and 21.4 per cent respectively with a negligible fraction removed by the CC and FMC.

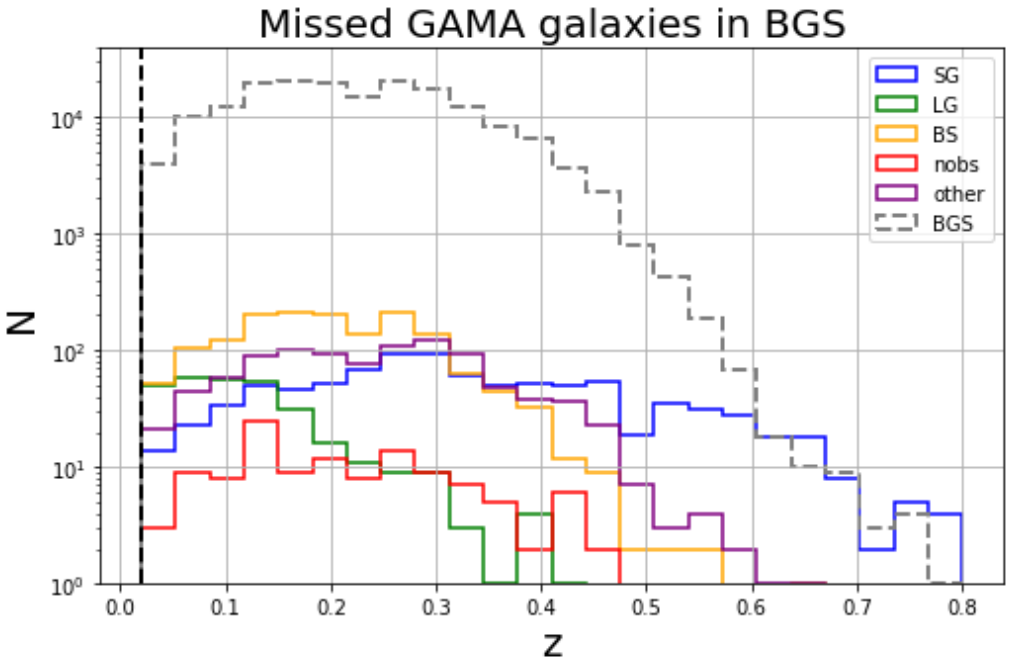


Figure 2.14: Redshift distribution of the GAMA galaxies that are not included in the BGS, with objects rejected by different cuts indicated by different line colours as labelled: blue shows GAMA objects missed due the star-galaxy separation applied (SG), green due to large galaxy masking (LG), yellow – bright star masking (BS), red – number of observations (NOBS) and purple due to the remaining cuts (CC, FMC and all the QCs). The dashed gray line shows the redshift distribution of BGS galaxies cross-matched with GAMA. The vertical black dashed line marks the redshift boundary between stars ($z < 0.002$) and galaxies.

In Fig. 2.14 we show the redshift distribution of the GAMA galaxies that are not present in the BGS. The solid coloured lines show the distribution for GAMA galaxies rejected by different BGS cuts, as labelled in the figure. We also plot the overall redshift distribution of BGS galaxies for comparison. GAMA galaxies

removed by the bright star masking and by the restrictions on the number of observations have a similar redshift distribution to the overall BGS. GAMA galaxies that are removed by the large galaxy mask have a distribution that is shifted to lower redshifts than the overall BGS distribution. GAMA galaxies can be found within the geometric BGS mask as GAMA does not use masking to deal with large galaxies, and so GAMA galaxies can be found in the regions that the BGS rejects around large galaxies. However, GAMA does perform masking around bright stars but this is less aggressive than the LS DR8 bright star masking. This can be seen from the areas rejected: the bright star masking in GAMA removes ~ 1 object/deg^2 (Baldry et al., 2010) whereas LS DR8 removes ~ 5 objects/deg^2.

2.5.2 Potential systematics

Here we look at potential systematic effects that could influence the homogeneity of the BGS catalogue and show how to mitigate these. As in any survey, the density of BGS targets is affected by observational effects which arise for a number of reasons. These include astrophysical foregrounds such as Galactic extinction, variations in the density of stars in the Milky Way, as well as variations in depth for the different imaging surveys and uncertainties in the data calibration.

To study the impact of these systematics on the observed galaxy density, we use a HEALPix map that divides the whole sky into $12N_{\rm side}^2$ equal area pixels, adopting $N_{\rm side} = 256$. Each pixel contains the median value of the systematics values within the pixel and the BGS target density. The corresponding BGS target density in each pixel, $\eta_{\rm i}$, is defined in Equation 2.7.

We study the effect of eight systematics on the BGS target density:

- Stellar density: we use stars from the *Gaia* DR2 catalogue with $12 < G < 17$ to construct the stellar density in each HEALPix pixel.

- Galactic extinction: the extinction values were computed using the SFD98 dust maps as reviewed in Section 2.2.1.

- PSF size (seeing) in the grz bands: the PSF size measures the full width at half maximum (FWHM) of the point spread function (PSF) which determines how much the transmission of light through turbulence in the Earth's atmosphere blurs the observed images. The seeing varies across the multiple observations.

- Photometric depth in the grz bands: the depth of the photometry, as characterised by the 5σ AB magnitude detection limit for a 0.45 arcsec round exponential galaxy profile, varies across the survey due to changes in the observing conditions.

To determine if the BGS target density has a systematic dependence on any of these quantities, we bin the HEALPix pixels according to the value of the quantity and for each bin determine the mean target density, η_i, and the error on the mean, $\sigma_i/\sqrt{N_i}$. In Fig. 2.15 we show how the mean BGS target density, η, varies with respect to each of the quantities listed above. Each panel shows the mean and error on the mean for three samples, BGS BRIGHT, BGS FAINT and the combined BGS sample (labelled simply BGS). The histogram below the curves in each panel shows (on an arbitrary scale) the number of HEALPix pixels contributing to each estimate. In general, the systematic variation in the BGS target density is less than 5 per cent, with the one exception being a ~ 7 per cent decrease in the target density in regions of high stellar density.

Stars could impact the BGS target density in at least five ways: i) Stellar contamination of the BGS selection could lead to increased target density in regions of the sky with high stellar density. ii) While the impact of very bright stars is dealt with by masking (see Section 2.3.1.1), the halos and diffraction spikes around slightly fainter stars could still affect the photometry of neighbouring galaxies. iii) High stellar density could lead to an overestimate of the local sky brightness

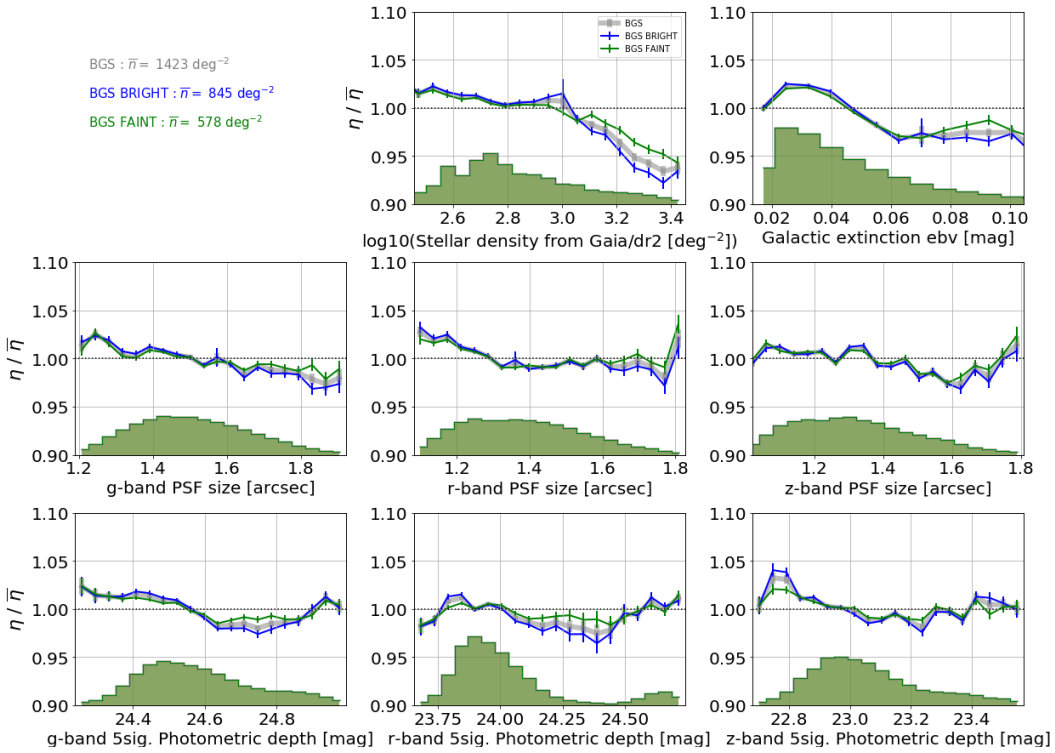


Figure 2.15: The systematic variation of the BGS BRIGHT (blue) and BGS FAINT (green) and combined (bgs_any, gray) target densities with respect to different properties: the logarithm of the stellar density from *Gaia* DR2, Galactic extinction, PSF size in the three bands (grz) and the photometric depth in each of the three bands (grz). The target densities and these eight quantities were computed in pixels on the sky using a HEALPix grid with resolution of $N_{\rm side} = 256$. Histograms shows the distribution for each of the x-axis properties. The error bars show the errors on the mean. Each target density, η is expressed in units of its mean across the whole survey $\bar{\eta}$ as given in the legend.

which, when subtracted, would lead to fainter galaxy fluxes and hence a lower BGS target density. iv) Star/galaxy superposition. v) Binary stars that TRACTORs resolution is not capable of resolving.

Stellar contamination would lead to an increase in target density with increasing stellar density, whereas we see a decrease that sets in above a stellar density of 10^3 deg^{-2}. Hence, stellar contamination cannot be the dominant systematic influence on the target density.

Galaxy photometry directly compromised by nearby stars that were not subject to masking also seems unlikely to be the cause for the variation in target density.

We test this by implementing the medium bright stars mask with a very little impact on target density and clustering. A further masking with 2 and 3 times the masking radius of equation (2.3) was also tested with no improvement on target density at high stellar densities.

The effect of high stellar density on the estimation of the sky levels deserves further investigation, but is deferred to another study. There is some variation of the target density with galactic extinction which could indicate systematic errors in the estimation of the amount of dust extinction. However, as there are spatial correlations between stellar density and dust extinction, these trends could be driven by the variation in stellar density and can be mitigated with several techniques such as linear and non-linear regressions and machine learning techniques such as Artificial Neural Networks (Rezaie et al., 2020).

Due to variations in observing conditions, the PSF size varies across the survey. The explicit modelling of the PSF of each image by TRACTOR should make the photometry robust to this variation. Also, our use of *Gaia* to perform star-galaxy separation should also make this classification independent to variations in the seeing. This appears to be borne out by the results shown in Fig. 2.15 which exhibit only very weak trends with PSF.

In the BGS, while the primary selection is in the r-band, TRACTOR simultaneously fits objects in all 3 bands and so the model parameters are affected by data in all three bands. However, any dependence on the depth of the photometry appears very weak in all three bands. This to be expected as the photometric depth is typically 3 to 4 magnitudes deeper than the $r = 20$ selection limit of the BGS.

2.5.2.1 Mitigation of systematics using linear weights based on stellar density

One way to mitigate the effect of the systematics in our catalogue is to apply a weight that corrects the target density. If we treat the systematic dependence of the

observed target density on a particular quantity, S, as a simple regression problem, we can define the observed target density, $\eta_{\rm i}^{\rm o}$, averaged over HEALpix pixels with a particular value of $S = S_{\rm i}$, as

$$\eta_{\rm i}^{\rm o} = \eta_{\rm i}\, W_{\rm i}(S_{\rm i}). \tag{2.9}$$

Here, $\eta_{\rm i}$ is the true target density and $W_{\rm i}(S_{\rm i})$ is the weight for a given systematic attribute, S. As shown in Fig. 2.15, the most important target density variation is driven by stellar density. Here, we assume that the weight is a simple linear function, $W_{\rm i}(S_{\rm i}) = mS_{\rm i} + c$, where $S_{\rm i}$ is the the stellar density, as we would expect any contamination (or anti-contamination) to be proportional to the stellar density and not to the $\log_{10}$(stellar density). The best fitting coefficients we find when applying this model to the combined BGS BRIGHT and BGS FAINT sample are $c = 1.03$ and $m = -3.96 \times 10^{-5}$. By construction, this weighting removes the general trend with stellar density for the combined sample and most of the trend with stellar density for the individual BGS BRIGHT and BGS FAINT samples. At the same time this weighting also reduces the weak systematic trend of target density with galactic extinction.

2.5.3 Angular correlation function

We measure the angular correlation function, $w(\theta)$, in five apparent magnitude bins from $r_{\rm AB} = 15$ to $r_{\rm AB} = 20$ for the BGS targets in DECaLS South Galactic Cap (SGC) and North Galactic Cap (NGC). Angular correlations were computed using the publicly available code CUTE (Alonso, 2012). We compare these with measurements from the mock BGS lightcone catalogue (Smith et al., 2017). This mock catalogue was built by populating the MXXL N-body simulation with galaxies based on a halo occupation distribution model. By construction, the HOD parameters of this mock reproduces both the luminosity function and 2-point clustering measured in the SDSS at low redshift and the GAMA survey at higher redshift.

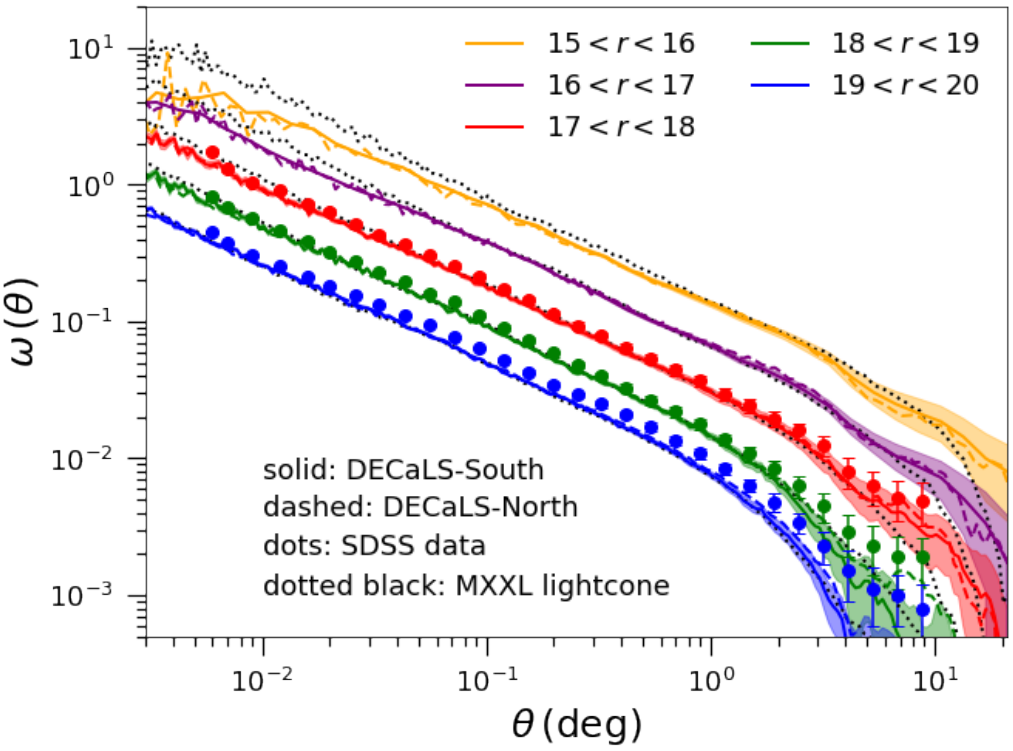


Figure 2.16: The angular correlation function, $w(\theta)$, measured for the BGS targets in bins of apparent magnitude; different colours indicate different magnitude bins as labelled. The shaded area shows the standard deviation obtained from 100 jackknife regions. The solid curves show the results for DECaLS-South, the dashed curves show DECaLS-North and the dotted curves show the angular clustering in the MXXL lightcone catalogue. The symbols with error bars show measurements from the SDSS by Wang et al. (2013).

Fig. 2.16 shows the comparison of angular clustering measured for the BGS targets with error bars corresponding to the standard deviation of 100 jackknife realisations, the MXXL mock and the SDSS observations by Wang et al. (2013). The angular clustering measurements are consistent between the DECaLS North and South regions, which demonstrates the homogeneity between these two parts of DECaLS. The angular clustering of the BGS targets agrees very well with that displayed in the MXXL lightcone. The HOD parameters of the MXXL mock have been tuned to attempt to match the clustering measured from SDSS MGS, however on large scales HOD models can only alter the amplitude and not the shape of the correlation. Moreover the shape of the large scale correlation function of MXXL is

very similar to that of all ΛCDM models that are consistent with CMB observations. Hence it is interesting that for the two faintest bins BGS is more consistent with MXXL (and hence with ΛCDM) than is SDSS MGS – possibly indicating reduced systematic errors.

We also look at the angular clustering of the BGS targets after applying the weights that depend on stellar density, as described in the previous section. Overall, applying stellar density weights has a small impact at angular scales larger than $3 - 4$ deg. Both the clustering with and without the weights are consistent with each other, within the error bar.

A further test of the fidelity of our BGS catalogue is to check for any spatial correlation of the distribution of BGS targets with stars in the Milky Way. Here we focus our attention on the fainter stars, $12 < G < 17$, which, ideally, should be removed from the BGS targets by our star-galaxy separation scheme. We find a significant anticorrelation on very small scales but no correlation on scales larger than 100 arc seconds.

2.5.4 Angular cross-correlation with large galaxies

In order to determine whether we are missing faint BGS targets around large galaxies due to the LG mask defined in Section 2.2.2.4, we measure the angular cross-correlation function between the SGA-2020 and faint BGS targets in $18 < r < 19$ (dash-dotted) as shown in Fig. 2.17. We also measure the angular cross-correlation function between these faint BGS targets and brighter BGS targets in the magnitude range $15 < r < 16$ (solid) where we assume that most of the large galaxies lie, and we do the same using the MXXL lightcone (dashed). The vertical dotted line shows the mean mask radius around large galaxies, which is about 10 arcsec.

The agreement between the results from the BGS catalogue (solid) and from the MXXL lightcone (dashed) suggests that our treatment of large galaxies is sat-

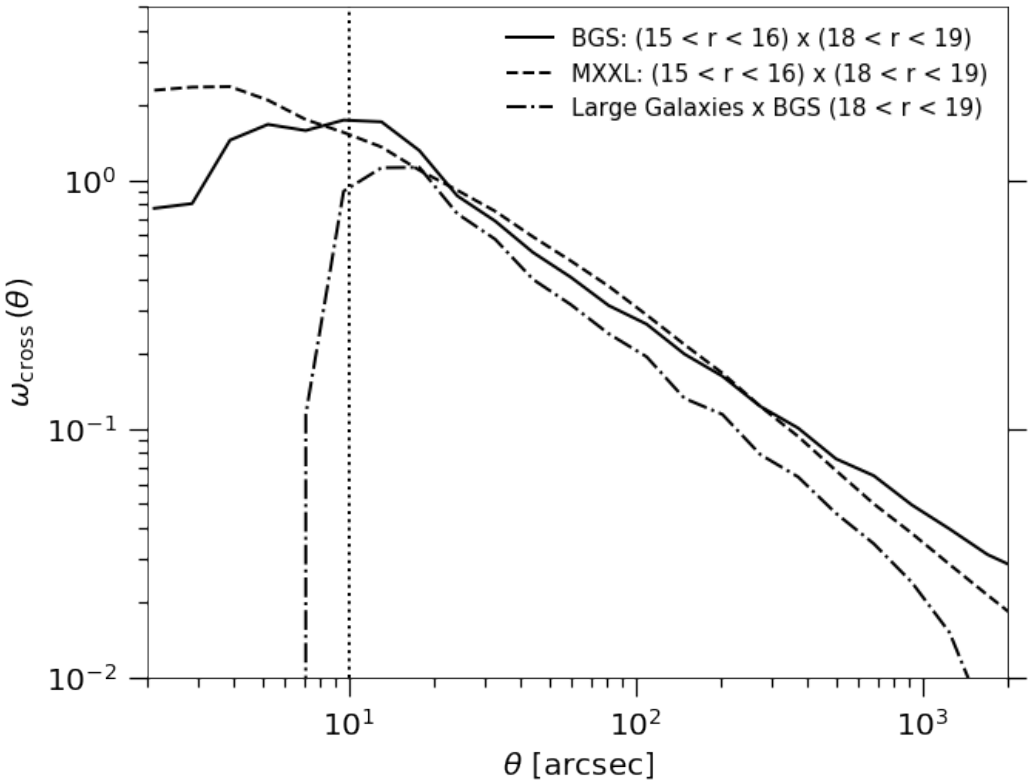


Figure 2.17: The angular cross-correlation function measured between faint BGS targets in $18 < r < 19$ and large galaxies from the SGA-2020 (dash-dotted) and between the same faint BGS targets and brighter BGS targets in $15 < r < 16$ (solid), the magnitude range in which most of the large galaxies reside. We also compare with the angular cross-correlation between these two bins in apparent magnitude measured in the MXXL lightcone (dashed). The vertical dotted line shows the mean LG mask radius which is about 10 arcsec.

isfactory and we are only missing BGS targets on scales below 10 arcsec, which is the median large galaxy masking radius (see Section 2.3.1.2). The difference in amplitude between the solid and dash-dotted curves, with a lower value when cross-correlating with the SGA-2020, suggests that the catalogue of large galaxies contains either more low-z galaxies or more brighter galaxies, or both, compared to the BGS targets in $15 < r < 16$.

2.6 Conclusions

Here we have presented the steps needed to define and select the Bright Galaxy Survey (BGS) targets for the Dark Energy Spectroscopic Instrument (DESI) project. Our galaxy selection uses DECaLS LS imaging data from Data Release 8 (DR8) reduced by the NOAO CP and TRACTOR pipelines. Our BGS target selection has two main components, one which imposes spatial cuts and the other which applies photometric selections. Figs. 2.2 and 2.4 show the flowcharts that set out these two selections. At each step these flowcharts report the remaining survey area and surface density of targets.

The main features of our spatial and photometric cuts are the following:

- The BGS spatial target selection removes area near bright stars (BS mask), large galaxies (LG mask), and globular clusters (GC mask), as well as galaxies with less than a specified minimum number of observations (NOBS mask). The BS mask is a circular aperture that scales with the magnitude of the bright star (see Eqn. 2.3). The exclusion of areas around bright stars removes ~ 270 deg^2, this is 2.76 per cent of initial footprint. Inspection of stacked images around bright stars (i.e. those with *Gaia* $G < 13$ or TYCHO2 $V < 13$) shows that the BS masking radius used in TRACTOR is well-motivated, with no sign of contamination around the bright stars in the BGS target density. There is a modest ~ 6 per cent increase in BGS target density just beyond the edge of the masked region. We find that there is a negligible angular cross-correlation between stars and galaxies at scales > 100 arcsec. Below 100 arcsec we have an anti-correlation possibly caused by the stars masked within the range $12 < G < 13$.

- The LG and GC masks account for a smaller number of contaminants than the BS mask, removing just ~ 9 deg^2 of survey area or 0.09 percent of initial footprint.

- DECaLS DR8 is complete to 99.5 per cent with at least one observation in the three bands grz, as described by the value of NOBS. The selection made on NOBS removes ~ 39 deg^2 of imaging data.

- We use *Gaia* DR2 to separate stars and galaxies as described in Section 2.4.1. This classification exploits the small PSF of the *Gaia* imaging compared with that typically present in ground-based observations. In our classification scheme we compare the measurement of the flux of an object by *Gaia* with that from TRACTOR through the parameter $G - rr$. Objects with a TRACTOR flux that is greater than that reported by *Gaia* are considered to be galaxies because this difference implies that they are extended sources (see Fig. 2.5).

- A small fraction (~ 0.35 per cent) of BGS galaxies are of PSF type according to TRACTOR. About half of these are compact sources for which the PSF model is the best fit, but the other half have only PSF photometry as they were designated stars based on the *Gaia* Astrometric Excess Noise (AEN) parameter before TRACTOR was run. For these objects TRACTOR only performs PSF fits. Matching to GAMA reveals that most (96 per cent) of these BGS PSF-type objects are confirmed to be galaxies by the GAMA spectroscopy. In addition, we find that the ~ 7 GAMA galaxies/deg^2 that are missed in BGS are mostly (~ 98 per cent) of PSF type according to TRACTOR. We conclude that using the AEN classification is i) causing ~ 0.17 per cent of BGS galaxies to be falsely classfied as of PSF type and ii) compromising the photometry of another 7 objects/deg^2 which then due to having their fluxes underestimated are falsely classified as stars by the BGS $G - rr$ star-galaxy classification.

- Possible systematic effects in DECaLS leave a small imprint on surface density of BGS sources. The variation in the target density of BGS sources as a function of the main possible systematic effects, such as the stellar density,

galactic extinction, seeing and imaging depth, is less than 10 per cent in the case of stellar density and under 5 per cent for the remaining systematics. We implement a weighting scheme based on a linear regression model which uses the density of stars to mitigate these effects. Applying the resulting weights, variation in the target density with stellar density is removed by construction, and is greatly reduced when plotted against the other systematic quantities.

- Angular clustering measurements made from our BGS target catalogue are compared with previous measurements from SDSS and the predictions from the MXXL lightcone mock catalogue, which on large scales can be taken as a prediction of ΛCDM models (see 2.5.3). On small scales, the three measurements of the angular correlation function agree well, with the exception of the brightest galaxies considered. At large scales, the angular clustering we find for the BGS targets is closer to that recovered from the MXXL mock catalogue than the SDSS measurements. The agreement between the BGS and the MXXL lightcone is even better after applying the linear weights based on stellar density to the BGS.

Galleries with examples of BGS targets divided in BGS BRIGHT and BGS FAINT can be found at `http://astro.dur.ac.uk/~qmxp55/bgs_ts_paper_gallery.html` along with galleries showing examples of rejected objects by the different spatial and photometric cuts we apply in BGS. We included also examples of discrepancies between our star-galaxy (SG) classification using *Gaia* with TRACTORs divided into 1) TRACTORs extended objects that fail our SG classification, and the TRACTORs point sources objects that pass our SG classification and 2) are *Gaia* and 3) are not *Gaia* sources. Finally, examples of discrepancies between TRACTORs point source classification for *Gaia* objects and our SG classification divided in two samples: 1) are galaxies by our SG classification but stars according to TRACTORs assessment of *Gaia* sources using the Astrometric Excess Noise (AEN) parameter from *Gaia*, and 2) stars by our SG classification but galaxies by

their AEN classification.

In Chapter 3 we will focus on applying this framework to select BGS targets using the additional LS, BASS and MzLS imaging data, and set out what is needed to tune our selection to use the release of the LS, DR9. Chapter 3 summarises the main changes in the imaging between both releases and main changes between both BGS selections, and hence should refer to this Chapter for most of the details of the BGS selection. Chapter 3 will also include a more complete clustering analysis using mock catalogues and colour based clustering measurements.

Chapter 3

BGS selection with Legacy Surveys DR9

3.1 The Legacy Surveys DR9

The DESI Legacy Imaging Surveys produced an inference model catalogue of the sky from a set of optical and infrared imaging data, comprising 20,000 deg^2 of extragalactic sky visible from the northern hemisphere in three optical bands (g, r, z) and four infrared bands. The sky coverage is approximately bounded by -18 $<$ Dec $<$ +84 deg in celestial coordinates and $|b| > 18$ deg in Galactic coordinates. To achieve this goal the DESI Legacy Imaging Surveys conducted 3 imaging projects on different telescopes.

The following surveys constitute the imaging in the optical for DESI targeting:

- DECaLS: The Dark Energy Camera Legacy Survey. The program completed in March, 2019. The DECaLS program made use of other DECam data within the DESI footprint. The most significant of these other data sets is from the Dark Energy Survey (DES The Dark Energy Survey Collaboration, 2005), which includes a 5000 deg^2 contiguous area in the South Galactic Cap. DECaLS explicitly did not re-image that area, instead incorporating the DES imaging itself.

- BASS: The Beijing-Arizona Sky Survey imaged regions at Dec $\geq +32$ deg in the Northern Galactic Cap (NGC), in the g and r optical bands. Although the focus of the survey was Dec $\geq +32$ deg, about 4% of the observations included in the Legacy Surveys are at Dec $< +32$ deg. Notably, 1% of these are in equatorial regions (to facilitate studies of imaging in a region where BASS overlaps with DECaLS).

- MzLS: The Mayall z-band Legacy Survey imaged the Dec ≥ 32 deg region of the Dark Energy Spectroscopic Instrument (DESI) footprint. Although the focus of the survey was Dec ≥ 32 deg, a few per cent of the observations included in the Legacy Surveys are at Dec < 32 deg. Notably, 2% of these are in equatorial regions (to facilitate studies of imaging in a region where MzLS overlaps with DECaLS).

3.1.1 Main differences with DR8

Compared to its predecessor, DR9 incorporates some major and minor changes that affect the photometry of the objects, and hence the target selection. These changes are incorporated in TRACTOR *, the code that generates image modelling of multi-band and multi-epoch data sets. Below we list the most important changes relevant for BGS target selection:

1. **Iterative detection**: After the first round of fitting, TRACTOR conducts a second round of detections over the data-model residuals with the aim of finding additional sources.

2. **Extended PSF model**: An extended PSF model is used to subtract the wings of bright stars from DECam images only.

3. **Sersic fitting model**: The composite (COMP) morphological model has been replaced by a Sersic profile (SER). A source is classified as SER if a

*`https://github.com/dstndstn/tractor`

Sersic profile provides a better fit than other profiles, PSF, EXP and DEV, as quantified by a χ^2 that takes account of additional free parameters of the Sersic fit.

4. **Relaxed *Gaia* PSF criterion**: TRACTOR forces *Gaia* objects to be fitted as PSF sources if they meet the condition ($G \leq 18$ & AEN $< 10^{0.5}$) OR ($G \leq 13$), that was previously set to ($G \leq 19$ & AEN $< 10^{0.5}$) OR ($G \leq 19$ & AEN $< 10^{0.5+0.2(G-19)}$). Here G is GAIA DR2 G-band magnitude, and AEN is the *Gaia* astrometric excess noise parameter.

5. **Pre-fitting for large sources**: Regions around large galaxies and globular clusters have their own local source extraction, which is performed separately from the normal TRACTOR run. The parent catalogues of these objects have improved extensively since DR8.

3.2 Target selection cuts

The target sample for the BGS is intended to be a galaxy sample that is magnitude-limited in the r-band. The magnitude limit is determined by the total amount of bright observing time and the exposure times required to achieve the desired redshift efficiency. This target selection is, in essence, a deeper version of that for the SDSS Main Galaxy Sample (Strauss et al., 2002). The depth of the DESI BGS is comparable to GAMA and the area it covers is 50 times larger, but in dense regions the fraction of targets observed by DESI will be lower (Driver et al., 2012; Liske et al., 2015; Baldry et al., 2017). We describe the BGS target selection below and in Section 3.3 we compare the main changes with the BGS selection from the Legacy Surveys DR8 define in Sec. 2 and published in these papers Ruiz-Macias et al. (2020); Ruiz-Macias et al. (2021).

1. **Star-Galaxy separation**: based on *Gaia* DR2, a galaxy in BGS is defined by $(G - rr > 0.6)$ `or` $(G = 0)$ where G is the *Gaia* G-mag and rr is the LS r-

band magnitude (without any extinction correction). The term $G = 0$ means non-detection of a *Gaia* source. The aim of this selection is to compare LS models with the better PSF model of *Gaia*.

2. **Spatial masking**: this includes geometrical masking around i) bright stars (BS) and ii) globular clusters (GC) and iii) a pixel masking. The geometric masks require that LS MASKBITS 1 and 13 are not set. The bright star mask (defined by `bit`=1) combines stars from *Gaia* DR2 (Gaia Collaboration et al., 2018) and the TYCHO2 (Høg et al., 2000) catalogue, corrected for epoch and proper motions. This mask consists of a circular exclusion region with a radius $R_{\rm BS}(m)$ (Eqn. 3.1), that depends on the magnitude of the star, m. The magnitude is either the TYCHO2 MAG_VT or *Gaia* G-mag with *Gaia* G-mag taking precedence. Stars fainter than $m = 13$ are not masked. The globular cluster (GC) mask (`bit`=13) consists of a circular exclusion zone around known GCs from the OpenNGC catalogue*.

$$R_{\rm BS}(m) = 815 \times 1.396^{-m} \text{ arcsec} \tag{3.1}$$

3. **Photometric cuts**: these are i) colour-colour cuts in $g - r$ and $r - z$ (see equations 3.2 and 3.3), ii) cuts in fiber magnitude to increase the redshift success rate (see equation 3.4), and iii) cuts in objects with low quality photometry (see equation 3.5).

$$(-1 < g - r < 4) \tag{3.2}$$

$$(-1 < r - z < 4) \tag{3.3}$$

$$\text{rfibmag} < \begin{cases} 22.9 + (r - 17.8) & \text{for } r < 17.8 \\ 22.9 & \text{for } 17.8 < r < 20 \end{cases} \tag{3.4}$$

*`https://github.com/mattiaverga/OpenNGC`

Table 3.1: Target density in objects/deg^2 (η) and the effective area ($A_{\rm eff}$) in deg^2 of the BGS target selection we have adopted for the LS DR9. We show results for BGS BRIGHT and BGS FAINT each divided into three regions; BASS/MzLS, DECaLS NGC and DECaLS SGC. $A_{\rm eff}$ is the area after accounting for the spatial masking.

BGS	BASS/MzLS	DECaLS NGC	DECaLS SGC
$\eta_{\rm bright}$	813.4	875.6	839.0
$\eta_{\rm faint}$	569.4	598.8	581.1
$A_{\rm eff}$	4493	5263	4326

$$\begin{aligned} \text{FRACMASKED\_}i &< 0.4, \\ \text{FRACIN\_}i &> 0.3, \\ \text{FRACFLUX\_}i &< 5, \\ \text{where } i &= \{g,r\},\ \{g,z\} \text{ or } \{r,z\}. \end{aligned} \tag{3.5}$$

A detailed view of the cuts implemented in BGS can be found in Appendix A.2. Here, Fig. A.2 and Fig. A.3 shows a flow chart of the BGS selection for DECaLS and BASS/MzLS respectively, starting with the galaxies observed by the DR9 Legacy Imaging Surveys using the star-galaxy separation defined in this section.

3.3 Comparison of the BGS target selection between DR8 and DR9

The BGS selection criteria defined above include three main changes from that defined using DR8 DECaLS in Chapter 2 (also Ruiz-Macias et al., 2020; Ruiz-Macias et al., 2021). In this section, we summarize these changes and the motivation behind them.

The masking radius around bright stars has been reduced by a factor of two in DR9 compared to DR8. Fig. 3.1 shows the stacked average density of BGS close to bright stars. Distances have been rescaled to the masking radius $R_{\rm BS}(m)$ of

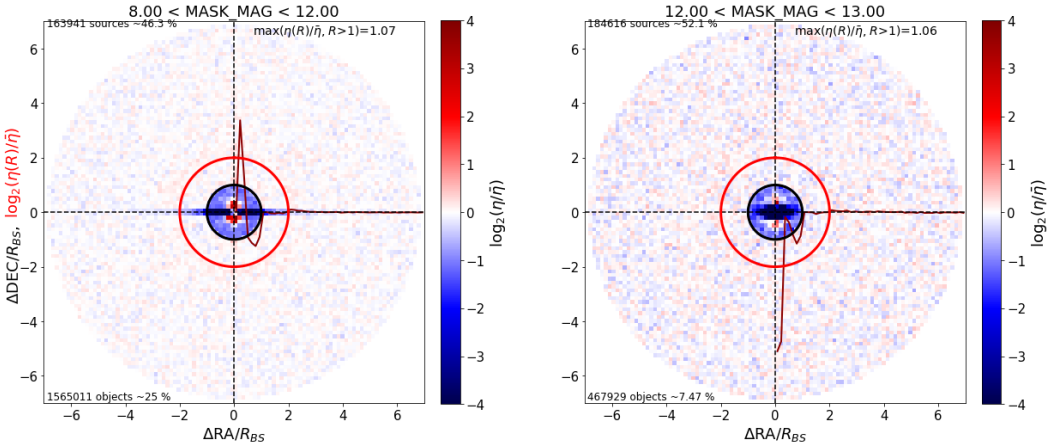


Figure 3.1: 2D histograms of the positions of BGS objects from BASS/MzLS relative to their nearest Bright Star (BS) taken from the *Gaia* and Tycho catalogues down to G-mag and visual magnitude MAG_VT of 13 respectively. These stacks are performed in magnitude bins in the BS catalogue from magnitude 8 to 12 (left) and 12 to 13 (right). The stacks are made using angular separations rescaled to the masking radius function given in Eqn. 3.1 (inner black circle), which means that objects within a scaled radius of 0 to 1 will be masked out by the BS veto while objects with $R = r/R_{\rm BS} > 1$ will not (here $r^2 = \Delta{\rm RA}^2\cos({\rm DEC})^2 + \Delta{\rm DEC}^2$). The colour scale shows the ratio of the density per pixel (η) to the mean density ($\bar{\eta}$) within the shell $1.1 < r/R_{\rm BS} < 7$. The density ratio is shown on a $\log_2$ scale where red shows over-densities, blue corresponds to under-densities and white shows the mean density. The outer red circle shows the masking radius of the LS DR8 data. The red solid line shows the radial density profile on the same scale as the colour distribution $\log_2(\eta(R)/\bar{\eta})$ where $\eta(R)$ is the target density within the annulus at radius R of width $\Delta R \sim 0.06$.

equation 3.1. The DR9 masking radius is represented by the smaller black circle while the radius applied in DR8 is shown by the larger red circle. Contamination by bright stars seems to be higher for the brightest stars ($8 < m < 12$) but the density profile of BGS objects (solid red line in Fig. 3.1) shows almost no sign of contamination for BGS sources that fall outside the masking radius defined by equation 3.1 but inside the masking radius used in DR8. The reduction in the masking radius reduces the area masked around bright stars from ~ 2.8 per cent to ~ 0.9 per cent. Further tests on the effect of stars on the BGS target density are done in Section 3.6.1.

Further changes in the BGS selection include not masking around the Large Galaxies of the Siena Galaxy Atlas (SGA-2020, Moustakas, Lang, in preparation),

and a revisiting of the quality cuts (QCs). The SGA galaxy catalogue has gone through a series of improvements and TRACTOR was run separately within the regions immediately surrounding these galaxies, which has led to a reduction in the number of spurious objects around these galaxies in DR9 compared to DR8. These spurious objects were due to large galaxies not being appropriately fitted by TRACTOR and as a result, these galaxies were fragmented in many fake sources. In addition to that, in DR8, TRACTOR forced PSF fits to all the objects around large galaxies, compromising the photometry of the potential BGS targets in these regions and forcing us to mask them. In DR9, we have visually inspected around 1 per cent of the BGS within the SGA mask and found $\sim$ 50 per cent of them are galaxies and the remaining $\sim$ 50 are either stars or fragmented galaxies. We have decided to target all of these to ensure completeness for clustering studies. We can reject spurious objects at a later stage. Fig. 3.2 shows 2D histograms of BGS targets in the vicinity of the large galaxies for the cases: i) with the LG mask is applied, and ii) without the LG mask being applied. The radial distance was rescaled to the size of the minor-axis of the masking ellipse.

In the left hand panels of Fig. 3.2, we see the BGS target density falling at small radii is a result of applying the LG mask. In contrast, the right hand panels show the BGS target density rising steadily to much smaller scales although eventually turning over when very close to the central galaxy. This increasing target density is partially a result of the galaxy correlation function but also enhanced due to spurious sources. However, using GAMA DR4*, we match BGS with the GAMA Bright Galaxy Sample and for the matched objects within the LG mask, we were able to identify that around 40 per cent are spectroscopically confirmed galaxies that would otherwise be rejected by the LG mask.

Turning to the "quality cuts" (QCs) defined in Chapter 2, `FRACMASKED_`$i < 0.4$, `FRACIN_`$i > 0.3$, `FRACFLUX_`$i < 5$, where $i \equiv g$, r or z. `FRACIN` is used to reject

*This is an unreleased version that the GAMA collaboration made available to us. It is essentially the same as GAMA DR3, but with more redshifts.

Table 3.2: Increase in target density in objects/deg^2 (η) of the current BGS target selection (DR9) compared with the BGS selection defined for DECaLS DR8. We show results for BGS BRIGHT and BGS FAINT and for three regions; BASS/MzLS, DECaLS NGC and DECaLS SGC.

BGS	BASS/MzLS	DECaLS NGC	DECaLS SGC
$\Delta\eta_{\rm bright}$	+12.7	+14.1	+12.5
$\Delta\eta_{\rm faint}$	+ 7.7	+ 8.2	+ 7.3

sources for which a large fraction of the model flux lies outside the contiguous pixels to which the model was fitted, `FRACFLUX` is used to reject objects that are swamped by flux from adjacent sources, and `FRACMASKED` is used to veto objects with a high fraction of masked pixels. The overall improvement in the quality of the photometry assessed in next sections 3.3.1 and 3.4 allows us to adopt a less conservative definition of the QC. Instead of requiring objects to pass these cuts in each of the three bands, we now only require them to pass the cuts in any two of the three bands (see Eq. 3.5). In subsequent analysis, we will refer to the DR8 QCs (see Eqn. 2.6) as old FRACS, and the QCs in this Chapter as new FRACS. The objects rejected by new FRACS are a subset of around 1/3 of those rejected by old FRACS. In Section 3.4, we assess again the completeness with respect to GAMA and we remind the reader that for DR8 we were missing about 70 true galaxies/deg^2 because of old FRACS. In order to assess the current selection given by equation 2.6, we perform a visual inspection (VI) of the imaging with a sample of targets flagged by the old FRACS. The details of the VI set up and results are given in Sec. 3.3.1.

Compared with the BGS selection defined for DECaLS DR8, the current BGS selection increases the target density by 20 objects/deg^2. Table 3.2 shows the gain in target density for BGS BRIGHT and BGS FAINT, and for the three regions BASS/MzLS, DECaLS NGC and DECaLS SGC.

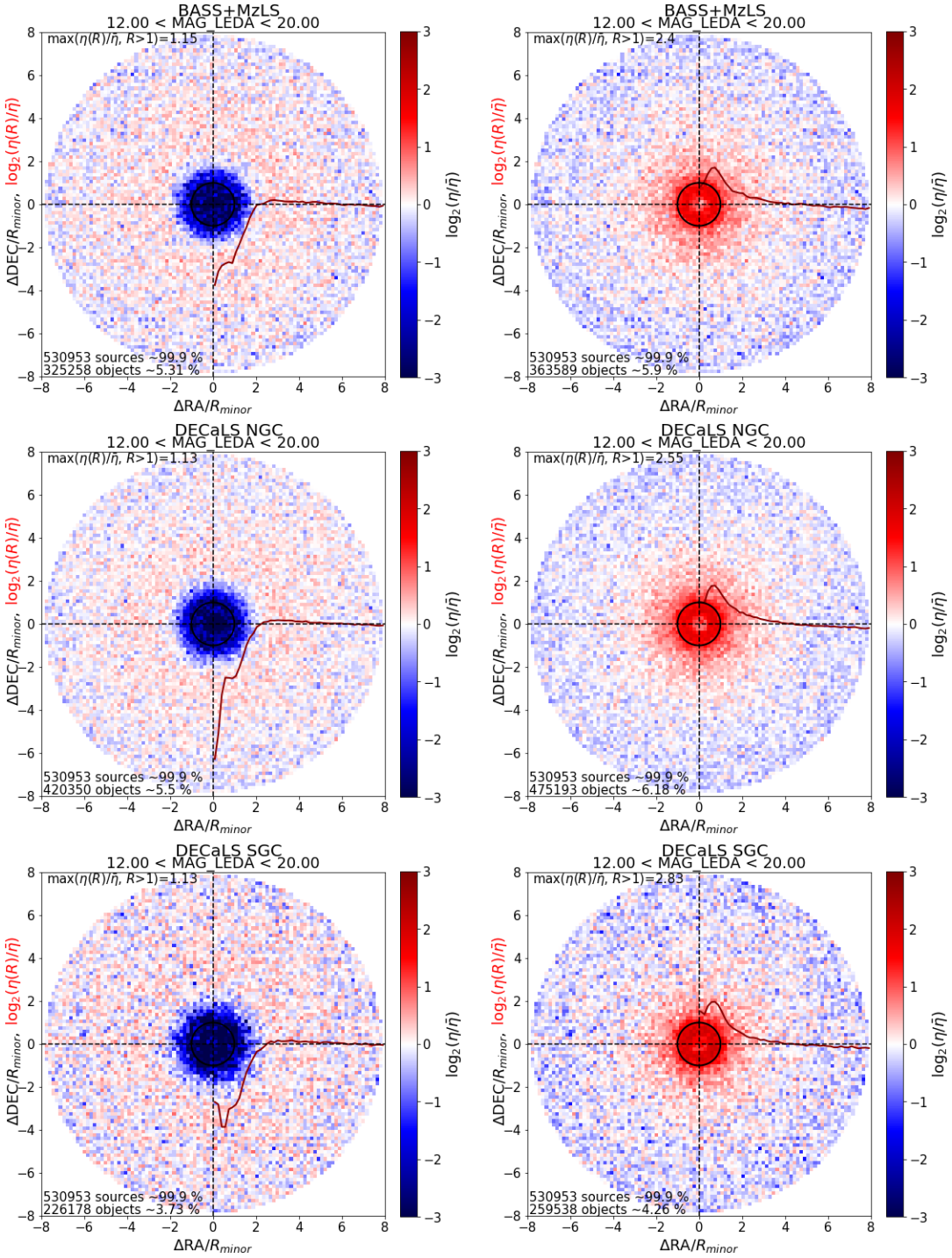


Figure 3.2: 2D histograms of the BGS targets in the vicinity of the large galaxies. Left: When applying the LG mask. Right: Without applying the LG mask. The radial distance was rescaled to the size of the minor-axis of the masking ellipse (R_{minor}) represented by the solid black circle.

3.3.1 Visual inspection of the imaging

The aim of the visual inspection(VI) of the imaging that we perform is to assess our selection while minimizing human uncertainties in the classification by involving as many people as possible. For this purpose, we created the LSVI* (Legacy Surveys Visual Inspection) tool, an interactive web framework that automates this process. LSVI essentially creates postage stamp galleries from the Legacy Surveys Sky Viewer web site viewer† (D. Lang in prep.) with corresponding classification radio buttons.

To assess whether the QCs we adopt for DR9 are correctly only rejecting spurious objects, we took a sparsely selected sample of 2000 BGS targets in a ~ 420 deg^2 area in DECaLS DR9 flagged by the old FRACS. This sample accounts for ~ 35 per cent of the total available in this area. The radio button classification labels are, GAL: the object is a galaxy; STAR the object is a star; BT: the object is contaminated by (or is a fake source from) a bleed trail; DS/H: the object is contaminated by (or is a fake source from) a diffraction spike or a stellar halo; FRAG: the object is spurious from a fragmented large galaxy; JUNK: the object does not fit in any of the previous classifications but it is clearly spurious; UNK: the object does not fit in any of the previous classifications and it is unclear whether it is a galaxy or not.

A total of six people participated in the classification, and the results were classified in three categories: i) confirmed galaxy (CG): two or more people say the object is a galaxy and less than two people say it is anything but a galaxy; ii) confirmed non-galaxy (CNG): two or more people say the object is anything but a galaxy and less than two people say it is a galaxy; and iii) inconsistent classification (IC): is neither of previous categories. Out of the 2000 objects, 35 per cent are CG, 43 per cent are CNG, and 22 per cent are IC. The CNG include stars and other artifacts such as bleed trails, diffraction spikes, stellar halos and

*`https://lsvi-webtool.herokuapp.com/`
†Legacy Surveys / D. Lang (Perimeter Institute) `legacysurvey.org/viewer`

fragmented images. We have looked at various combinations of image parameters but have found no easy way to isolate the CG while still rejecting spurious sources. However, for the subset of above sample that are rejected by the new FRACS cuts defined by equation 3.5, the classification results show that around 10 per cent are CG, around 60 CNG, and around 30 per cent are IC.

Galleries with snapshots of the LSVI webpage for the visual inspection of the objects we describe above can be found in Appendix B.

3.4 Validation with GAMA

In DR8 DECaLS, we assessed the completeness of our BGS catalogue with respect to GAMA, whose main target sample (Baldry et al., 2017) is highly complete for galaxies with $14 < r_{\rm SDSS} < 19.8$, where $r_{\rm SDSS}$ is the SDSS Petrosian r-band magnitude, with 98.85% of the objects in the catalogue having good redshifts with a quality flag NQ ≥ 3, after applying a redshift cut at $z > 0.002$ to remove the remaining stars. We match the BGS targets with the GAMA Main Survey DR4 galaxy catalogue* using a maximum linking length of 1 arcsec and we focus on three of the five GAMA fields: G09, G12, G15.

In order to have a realistic comparison between the catalogues, we have excluded targets in GAMA that matched with LS targets that are vetoed by the BS, GC, and the NOBS masks. We cross-matched the resulting GAMA catalogue with the Legacy Imaging Surveys DR9. This results in a matched catalogue of 1007.5 objects/deg^2 in a 178 deg^2 area, which represents 99.8 per cent of all the resulting GAMA galaxies. We define four different BGS selections to test the completeness with respect to GAMA. The four samples are: i) DR8 cuts (nominal DR8), ii) nominal DR9 with no LG mask and new FRACS (current selection), iii) nominal DR9 with no LG mask and old FRACS (Current old FRACS), iv) nominal DR9

*This is an unreleased version that the GAMA collaboration made available to us. It is essentially the same as GAMA DR3, but with more redshifts.

Table 3.3: Number of BGS targets that matched with GAMA, and the GAMA galaxies that are missing in BGS. Each row is a different BGS sample and numbers are objects/deg^2.

BGS samples	BGS in GAMA	GAMA not in BGS
Nominal DR8	995	13
Current	1,003	4
Current old FRACS	998	9
Current no FRACS	1,004	3

with no LG mask and no FRACS (Current no FRACS). The number of BGS targets for each of these samples that match with GAMA is given in Table 3.3. The table also lists the number of GAMA galaxies that are missing in BGS. The numbers are in objects/deg^2 and show that the BGS sample with no LG mask and no FRACS has the highest completeness with respect to GAMA.

Fig. 3.3 shows the completeness with respect to GAMA of the four samples obtained by computing the ratio of BGS matched with GAMA over the GAMA objects. These results show that the sample with no LG and no FRACS has the highest completeness with respect to GAMA, followed by the nominal DR9 sample; both these samples have more than 99.5 per cent completeness. However, if we had taken the ratio of the BGS matched with GAMA over the total of BGS targets, we would find that the nominal DR9 sample has the highest ratio, which can be interpreted as being the sample with the least contamination. Focusing on our current BGS selection, the nominal DR9 sample, there are 4.5 objects/deg^2 from GAMA that are missing in BGS, of which 3.8 objects/deg^2 are due to the star-galaxy separation, and the remaining 0.7 objects/deg^2 come from the QCs. Based on these results, we can conclude that our current BGS sample has a completeness with respect to GAMA which is above 99.5%. Note that none of the BGS samples in Table 3.3 represents the final BGS selection. However, the final BGS selection is close to the Current no FRACS sample and is defined in Section 5.9.

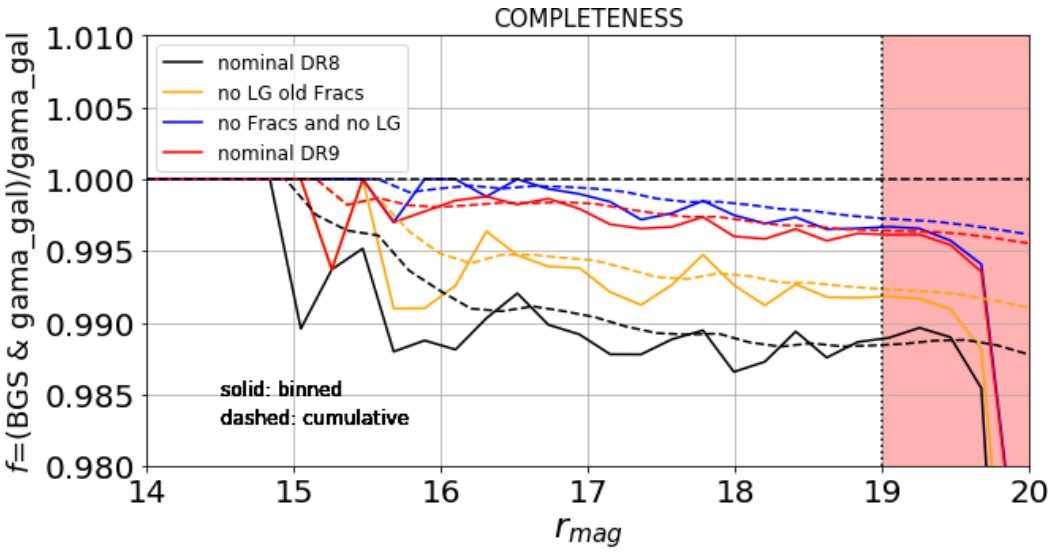


Figure 3.3: Completeness of the BGS targets according to different choices of target selection cuts with respect to GAMA which can be considered as complete for $14 < r_{\mathrm{SDSS}} < 19.8$. The solid lines are binned estimates and the dashed lines cumulative.

3.5 Comparison of DR9 DECaLS and DR9 BASS+MzLS

Looking at the survey depth in each band, characterised by the 5σ AB magnitude detection limit for a 0.45 arcsec REX galaxy profile, the DECaLS g and r bands go deeper than the equivalent bands in BASS (see the bottom row of plots in Fig. 3.6). In the z band, however, DECaLS and MzLS have a similar depth. Whilst, for the purpose of this work these depths are sufficient for the BGS selection, we are interested here to see how the magnitudes measured in the same bands differ between the surveys, which were conducted with different instruments at different telescopes.

DECaLS and BASS/MzLS overlap in the NGC at around Dec = 32 deg within $29 < \mathrm{Dec} < 35\,\mathrm{deg}$. For this analysis we looked at a 76 deg^2 area in the region $200 < \mathrm{RA} < 240\,\mathrm{deg}$ and $29 < \mathrm{Dec} < 35\,\mathrm{deg}$. The area was computed using a random catalogue with density of 15,000 objects per deg^2 and a HEALPix grid of $N_{\mathrm{side}} = 1024$. In order to compare the photometry in DECaLS and BASS/MzLS,

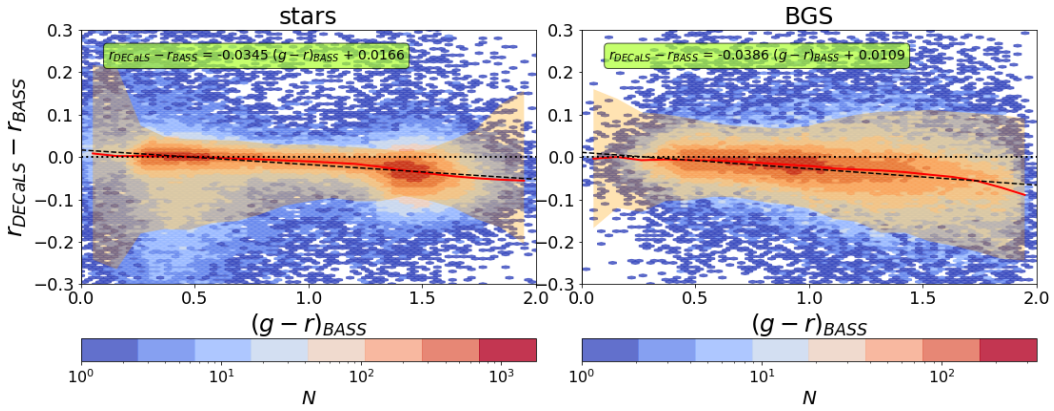


Figure 3.4: DECaLS and BASS/MzLS matched Stars (left) and BGS targets (right) showing $r_{\rm DECaLS} - r_{\rm BASS\text{-}MzLS}$ as a function of $(g-r)_{\rm BASS}$ colour. Stars are as defined by our *Gaia* star-galaxy classification. The colour bar indicates the number of objects within the hexagonal cell. The red solid line shows the median and the orange shaded region shows the the 3 and 97 percentiles. The black dashed lines show the best straight line fits to the median relation as given in the green boxes above.

we find all target matches within a distance of 0.5 arcsec with a mean separation of 0.12 arcsec and a standard deviation of 0.1 arcsec. To avoid incomplete regions, we require $\mathrm{NOBS}_i > 0$ for $i = g, r$ and z for the three surveys. After the match, we define the BGS objects for the three surveys using the BGS target selection we defined in Section 3. We find agreement between $1,328$ BGS objects/deg^2, and disagreement for 66 objects/deg^2 that are in BGS in DECaLS but which are not BGS in BASS/MzLS; conversely there are 28 BGS objects/deg^2 in BASS/MzLS that are not in the DECaLS BGS.

We find that most of the disagreements are due to a shift in the r-band magnitude. Fig. 3.4 shows $r_{\rm DECaLS} - r_{\rm BASS\text{-}MzLS}$ as a function of $(g-r)_{\rm BASS}$ for two samples: one with only stars in both BASS/MzLS and in DECaLS, and the other with only BGS objects. We can fit the magnitude difference as a linear function in $(g-r)_{\rm BASS}$. Equations 3.6 and 3.7 show these r-band photometry transformations of the BASS/MzLS system to the DECaLS system for BGS matches ($r'_{\rm gg}$) and for star matches ($r'_{\rm ss}$), respectively

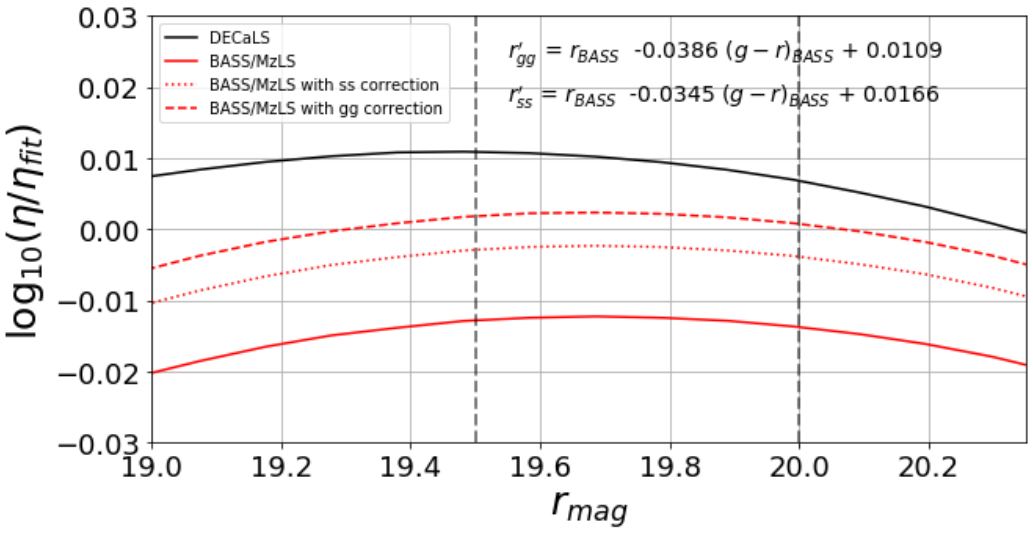


Figure 3.5: The BGS target density (η) divided by the the target density of a fiducial linear fit, $\log_{10}(\eta_{\rm fit}) = 0.46 \times r_{\rm mag} + 6.10$. The solid black and red lines show the DECaLS and BASS/MzLS BGS target densities within their full footprints. The red dashed and dotted lines shows BASS/MzLS BGS target density after applying the r-mag transformation equations 3.6 and 3.7 respectively.

$$r'_{\rm gg} = r_{\rm BASS} - 0.039(g-r)_{\rm BASS} + 0.011 \quad (3.6)$$

$$r'_{ss} = r_{\rm BASS} - 0.035(g-r)_{\rm BASS} + 0.017. \quad (3.7)$$

From equations 3.6 and 3.7 we can see that the r-band magnitude in BASS/MzLS is fainter than in DECaLS. The median r-band magnitude offset is $r_{\rm BASS} - r_{\rm DECaLS} \approx 0.026$ for BGS objects. In Fig. 3.5 we show the number counts as a function of r-mag for BGS objects in the full DECaLS and BASS/MzLS footprints. For the BASS/MzLS region we show these counts both as a function of $r_{\rm BASS\text{-}MzLS}$ and after transforming the BASS/MzLS magnitudes using equations 3.6 and 3.7. These results show that both regions can achieve similar target densities if we apply a linear transformation in BASS/MzLS, increasing the overall target density to 1430 objects/deg^2. Compared to the 1383 objects/deg^2 without the colour correction, this represents a increase of 3.4 per cent.

3.6 Study of potential systematics

In this section we investigate the potential systematics associated with the BGS target selection. Fig. 3.6 shows the variation of the BGS target density with common imaging systematics for BASS/MzLS (in blue), DECaLS-NGC (in red) and DECaLS-SGC (in green). We can see that the strongest dependence is due to the stars with a maximum 7 per cent variation in the target density with stellar density, which was also found with DR8 DECaLS (Ruiz-Macias et al., 2021). Only *Gaia* stars with $12 < G < 17$ are used here, and we remind the reader that only the brightest stars with $G < 13$ are masked (recall that G is the *Gaia* G-band magnitude). In Section 3.6.1, we study the cross-correlation signal of the BGS targets with this stellar catalogue. We also check the behaviour of the BGS targets in the vicinity of the large galaxies (Section 3.6.2) to see whether additional cuts are necessary to remove spurious objects. Both Sections 3.6.1 and 3.6.2 use auto-correlation and cross-correlation functions that are determined using the Landy-Szalay estimator (Landy and Szalay, 1993) defined by equation 4.1 in Section 4.1.

3.6.1 Cross-correlation with stars

First, we measure the angular cross-correlation between the BGS targets after masking and the stellar catalogue for the three imaging surveys. We tested several configurations for the Target Selection cuts: i) with and without applying the masking around large galaxies (LG), ii) considering the three options for the 'FRACS cut' defined in Section 3, not applying the FRACS (no FRACS), applying the conservative definition of DR8 (old FRACS), iii) applying a less conservative definition (new FRACS). The consequences of these different choices for the measured angular cross-correlation function of the BGS targets with *Gaia* stars are shown in Fig. 3.7. As expected, the large galaxy mask has no effect on the stellar contamination and the other configurations show a negligible impact given the error

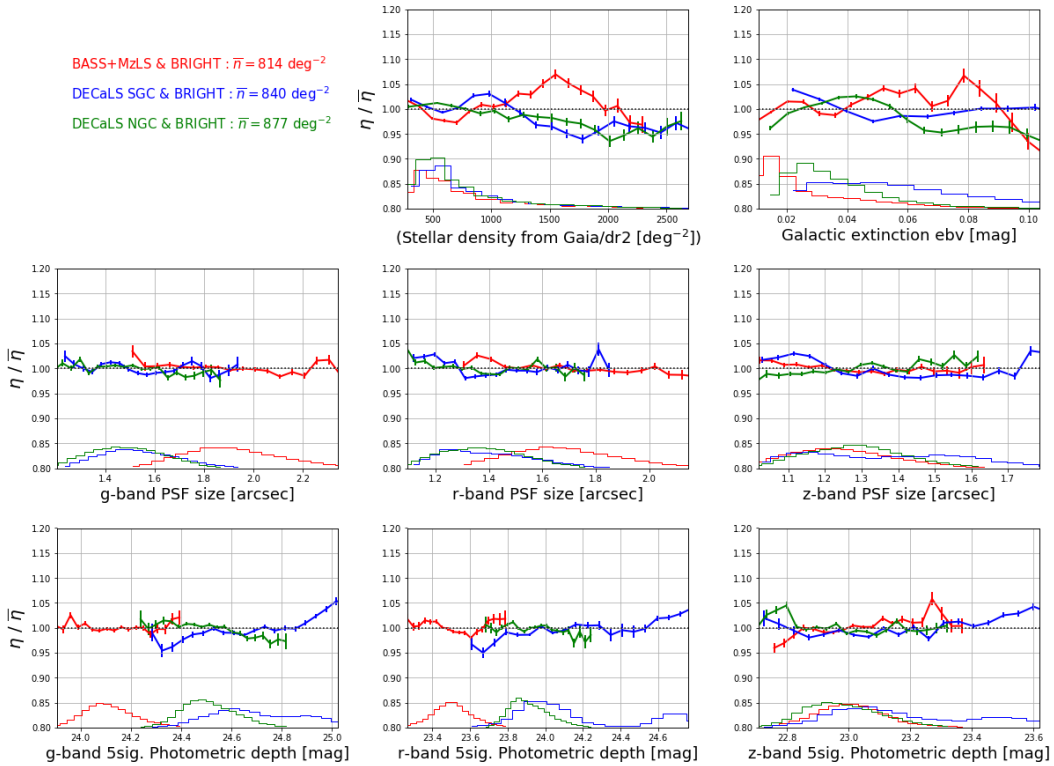


Figure 3.6: BGS BRIGHT target density variation with imaging systematics for BASS/MzLS (red), DECaLS-NGC (green) and DECaLS-SGC (blue). Lower histograms shows the distribution of BGS targets per region.

bar. Therefore, we conclude that none of these options should affect significantly the angular clustering of the BGS targets, as confirmed later in Section 4.1.

Fig. 3.8 shows the ratio between the cross-correlation function of BGS targets and stars, and the auto-correlation function of the stars. At large angular scales, this ratio gives an estimate of the fraction of contaminating stars in the BGS sample. The error bars are estimated using 100 jackknife regions for both the cross- and auto-correlation functions. The BGS targets seem uncorrelated with stars which is consistent with the small target density trends observed in Fig. 3.6. Indeed, although stars represent the main systematic in the BGS selection, we note that the effect remains small compared that seen for other DESI targets, such as the Emission Line Galaxies (ELG), which are fainter, or with Quasars (QSO), that are point-source objects. The correlation of these dark-time DESI targets with stars was shown in Kitanidis et al. (2019). Fig. 3.8 shows that there is no

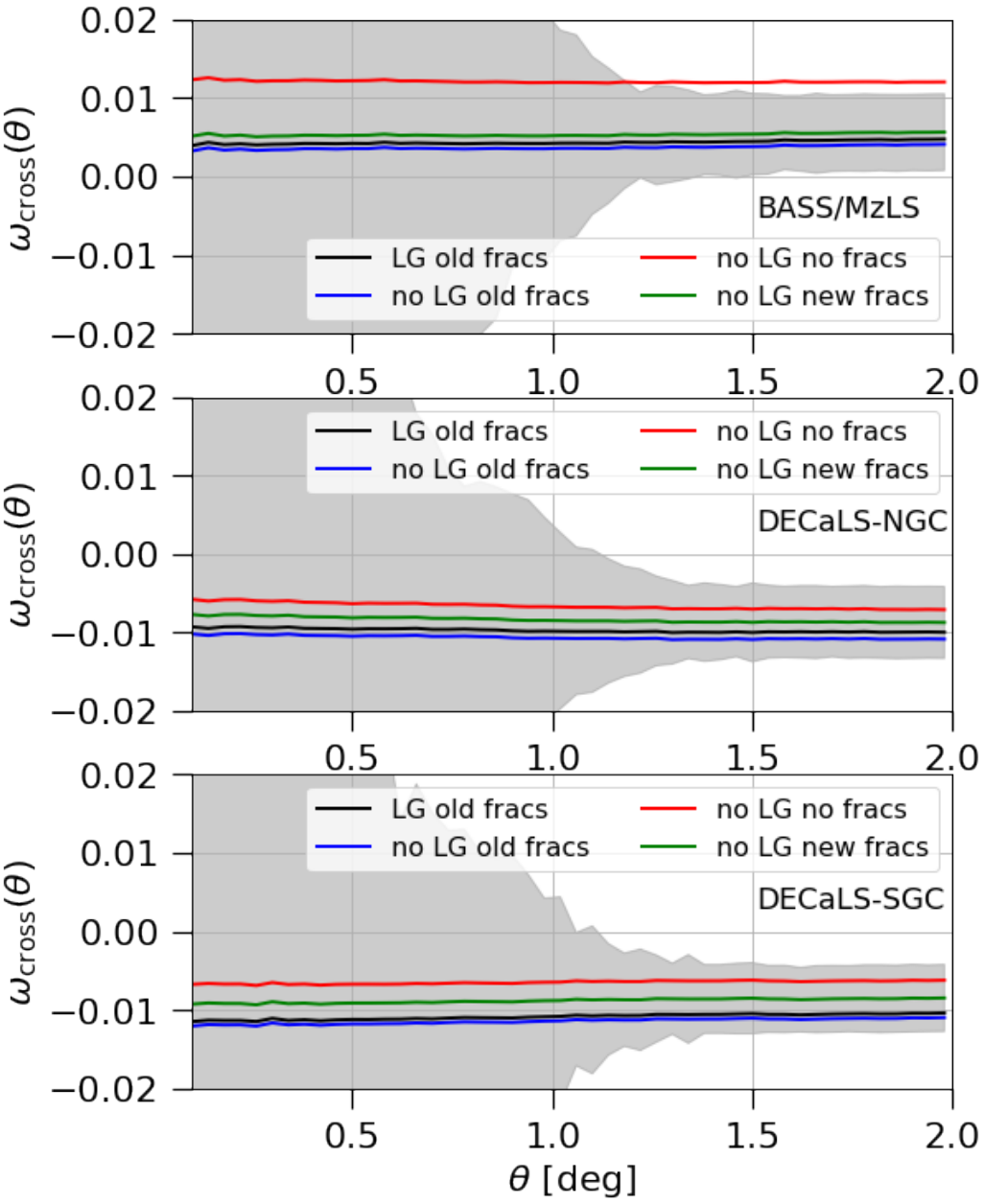


Figure 3.7: Angular cross-correlation function between the BGS Bright targets in each region and *Gaia* stars for different configurations of the BGS target selection. The *Gaia* stars have $12 < G < 17$. The shading shows the 1-σ error estimated using a jackknife resampling of the data for LG old FRACS selection only.

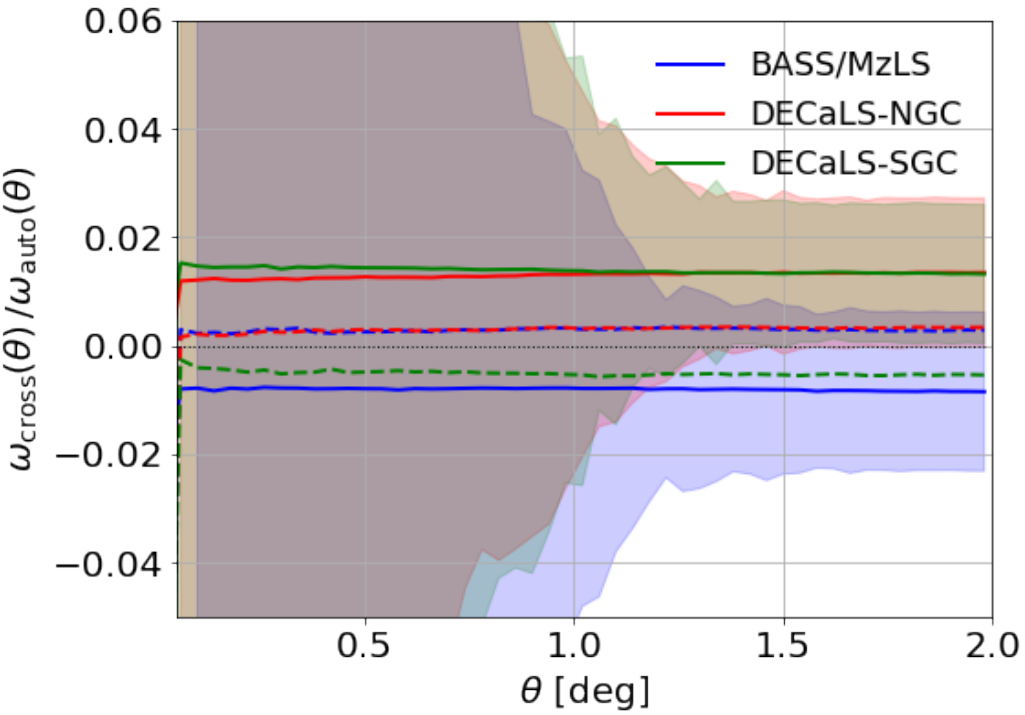


Figure 3.8: Ratio between the angular cross-correlation function between the BGS targets and *Gaia* stars and the auto-correlation function of *Gaia* stars. Dashed line shows this ratio after removing regions of high stellar density (stellar density < 1000 objects/deg^2). The shading shows the 1-σ error estimated using a jackknife resampling of the data. The *Gaia* stars have $12 < G < 17$.

significant stellar contamination in the BGS sample, which is also consistent with the results from the star-galaxy separation when considering *Gaia* objects. In the light of this result, we confirm that we meet the requirement according to which the BGS selection contains less than 2% stellar contamination. This will be further tested and confirmed, in Chapter 5, using spectroscopic star-galaxy classification provided by the DESI Survey Validation (SV) data. For SV we adopted a less conservative choice for the stellar rejection by selecting all the objects classified as non-PSF by Tractor regardless of whether or not they are classified as galaxies by our *Gaia* based classification as well as all objects classified as galaxies by our *Gaia* classification.

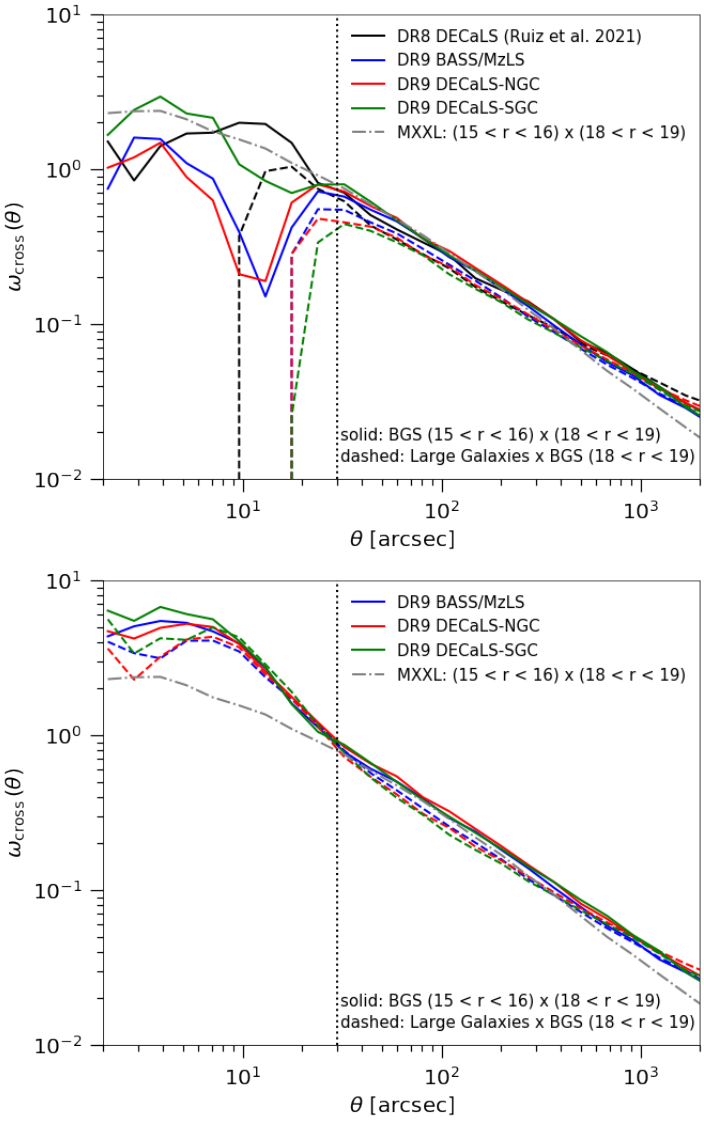


Figure 3.9: The angular cross-correlation function measured between faint BGS targets in $18 < r < 19$ and large galaxies from the SGA-2020 (dashed) and between the same faint BGS targets and brighter BGS targets in $15 < r < 16$ (solid), the magnitude range in which most of the large galaxies reside. We also compare with the angular cross-correlation between these two bins in apparent magnitude measured in the MXXL lightcone (dashed-dot). The vertical dotted line shows the mean LG mask radius which is about 20 arcsec. Top: applying the large galaxy mask when selecting the BGS targets. Bottom: without applying the large galaxy mask.

3.6.2 Cross-correlation with large galaxies

Large galaxies correspond to the brightest BGS galaxies in our sample. By comparing the angular cross-correlation between the large galaxies, which have typical magnitudes of about 15–16, and BGS faint targets (with $18 < r < 19$) and the angular cross-correlation between the BGS bright targets ($15 < r < 16$ for instance) with the same BGS faint galaxies, we can estimate whether we have an excess or deficit of BGS targets in the vicinity of the large galaxies due spurious or misclassified sources. A similar test for DECaLS DR8 was performed in Sec. 2.5.4 and those results are reproduced as the black lines on the top panel of Fig. 3.9. This panel also shows the result of this test when masking around the large galaxies in DR9, which also includes BASS/MzLS. In DR9, the median large galaxy masking radius is about 20 arcsec (shown by the dotted vertical line in Fig. 3.9) which is twice the size of that used in DR8. As expected, and as we found in DECaLS DR8, when masking is applied the dashed curves that correspond to the cross-correlation function between the large galaxies and the BGS faint targets drop dramatically on scales below the masking radius, meaning that we are missing BGS targets on these scales.

The bottom panel of Fig. 3.9 shows a similar study but without applying the large galaxy mask. The solid and dashed curves now agree much better on scales below 20 arcsec, meaning that we recover the BGS targets in the vicinity of the large galaxies. However, the overall amplitude seems larger than what is obtained from the MXXL lightcone for BGS (Smith et al., 2017) when measuring the cross-correlation function between the bright and faint BGS galaxies in the simulation (grey curve). This suggests that the BGS selection contains some spurious objects in the vicinity of the large galaxies that could be removed by additional cuts.

In Fig. 3.10, we show the impact on the cross-correlation signal at scales below the size of the masking radius around large galaxies of different choices for defining a quality cut based on FRACS. In Section 3, we presented what this set of cuts

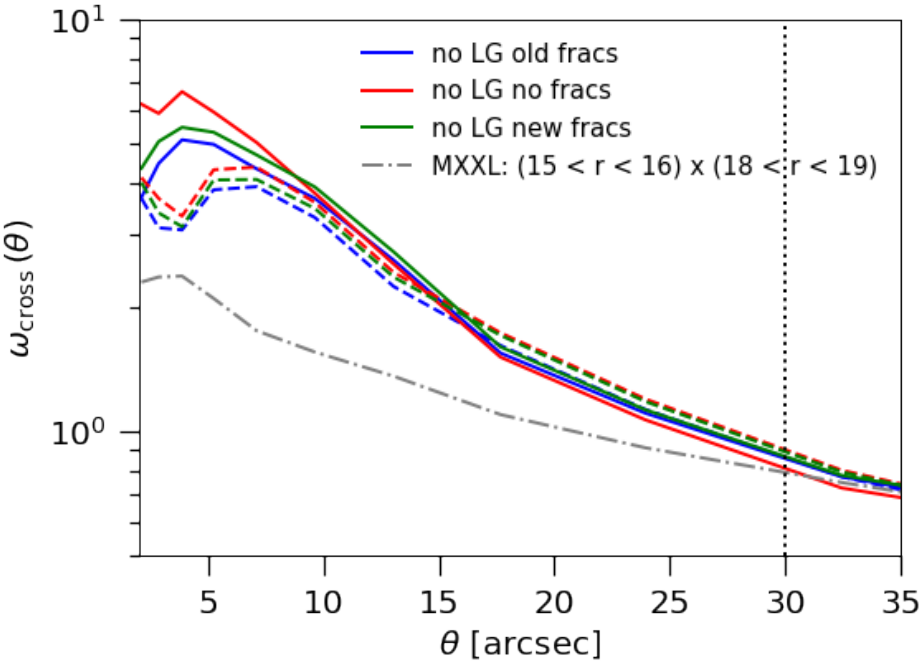


Figure 3.10: Same as in Fig. 3.9, but here we focus on scales below the masking radius around large galaxies and on the use of different quality cuts aimed at removing spurious objects in the vicinity of these large galaxies without removing true BGS targets.

corresponds to and in DR8 we adopted a conservative definition (old FRACS), with DR9 we investigated the effect of adopting a less conservative cut (new FRACS) or no cut at all (no FRACS). Not applying this cut increases the fraction of contamination around large galaxies which translates into a higher amplitude at these scales compared to MXXL and the other cases. The conservative approach adopted in DR8 seems to provide the best agreement with the results from the simulated lightcone. However, in the next section we will see that it also removes true BGS targets. As a consequence, there is a balance to be found between keeping true BGS targets while removing the spurious objects around the large galaxies. For this reason, we are developing a webtool to visually inspect a random fraction of the BGS targets in the vicinity of the large galaxies with the goal of determining the exact fraction of spurious objects and identifying common properties that could define more suitable cuts.

3.7 Conclusions

We have presented the target selection pipeline that selects the BGS targets ($r \leq 19.5$) from the latest release of the Legacy Imaging Surveys (DR9) that uses BASS, MzLS and DECaLS over the DESI footprint of $14,000$ deg^2. This includes several changes with respect to what was first presented using DECaLS DR8 in Ruiz-Macias et al. (2021) that we summarize here:

- Due the major improvements in the iterative fitting around bright stars, the radius of the bright star mask is half the size in DR9 than it was in DR8. We have checked that this change does not introduce any spurious effects. We found that our star-galaxy separation based on *Gaia* yields less than 2 per cent stellar contamination which is also confirmed by the cross-correlation signal of BGS targets and bright stars at large scales.

- Due the major improvements in the photometry in DR9, we no longer need to apply a spatial masking around large galaxies: we were missing true BGS targets in the vicinity of the large galaxies that we are now able to recover. We checked that this change improves the completeness with respect to GAMA which is highly complete in $14 < r < 19$.

- Using the visual inspection webtool, we found that we could make a less conservative choice on some of the quality cuts that involve FRACIN, FRACMASKED and FRACFLUX which increases the completeness slightly while ensuring a negligible fraction of spurious objects in the vicinity of the large galaxies.

- Finally, the DR9 selection cuts yield a completeness with respect to GAMA DR4, which is complete in $14 < r < 19$, that is above 99 per cent. DR9 also results in a BGS bright sample ($r \leq 19.5$) that meets the requirement for target density which is above 800 deg^{-2} for the three imaging surveys.

Chapter 4

Clustering analysis with DR9

While many cosmological studies require knowledge of the three-dimensional distribution of galaxies, the angular clustering also provides valuable information about both cosmology and the galaxy–halo connection. In Section 4.1, we present a detailed study of the angular correlation function that includes a comparison with theory and the MXXL lightcone of Smith et al. (2017) for BGS. In particular, we study the consistency between the BASS/MzLS and DECaLS BGS catalogues in terms of the angular correlation function (Section 4.1.2), then we analyse the clustering as a function of magnitude (Section 4.1.3) and as a function of colour (Section 4.1.4). Finally, in Section 4.2, we investigate the higher-order statistics of the galaxy density field using counts-in-cells.

4.1 Angular correlation function

4.1.1 Methodology

We measure the angular correlation function, $w(\theta)$, of the BGS targets using the estimator of Landy and Szalay (1993):

$$w_{\rm LS}(\theta) = \frac{D_1D_2(\theta) - D_1R_2(\theta) - D_2R_1(\theta) + R_1R_2(\theta)}{R_1R_2(\theta)}, \tag{4.1}$$

where DD, DR and RR are, respectively, the data-data, data-random and random-random pair counts at average separation θ. The random catalogue is provided by the Legacy Imaging Surveys*. The form given in equation 4.1 is for the cross-correlation of two samples. For the auto-correlation function, the labels 1 and 2 are indistinguishable and this simplifies to $w_{\rm LS}(\theta) = (DD - 2DR + RR)/RR$. We used the cross-correlation version of the estimator in Sec. 3.6.1 (BGS targets and stars) and in Sec. 3.6.2 (BGS targets and Large Galaxies). We use the publicly available code TWOPCF† to compute the angular correlation function together with jackknife errors. These jackknife errors are obtained by dividing the footprint into 100 independent regions of similar area such that each region contains the same number of points in the random catalogue.

In order to characterise the clustering of the BGS targets, we compare it to theoretical predictions based on the halo model (e.g. Peacock and Smith, 2000; Seljak, 2000; Cooray and Sheth, 2002). In the current paradigm of galaxy formation, galaxies form within dark matter halos and the overall galaxy clustering can be modelled by two contributions: one contribution due to galaxy pairs within dark matter halos (the 1-halo term) and another contribution due to galaxy pairs in separate halos (the 2-halo term; see, for example, Benson et al., 2000; Zheng et al., 2005). 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 a few h^{-1}Mpc, the typical virial radius of a halo, as first measured in the SDSS Main Galaxy Sample (Zehavi et al., 2004). Then, to obtain a prediction for the observed angular clustering, $w(\theta)$, based on a model for the full three-dimensional clustering, $\xi(r)$, we can use Limber's approximation (Limber, 1953) to project the real-space clustering into angular space, assuming a flat sky and small angular separations (for a discussion of the validity of Limber's approximation, see Simon, 2007):

$$w(\theta) = \frac{2}{c}\int_0^\infty \mathrm{d}z\, H(z)\left(\frac{\mathrm{d}N}{\mathrm{d}z}\right)^2 \int_0^\infty \mathrm{d}u\, \xi\left(r = \sqrt{u^2 + x^2(z)\theta^2}\right), \tag{4.2}$$

*`http://www.legacysurvey.org/dr9/files/random-catalogs`
†`https://github.com/lstothert/two_pcf`

where $\mathrm{d}N/\mathrm{d}z$ is the normalised redshift distribution, $x(z)$ is the comoving distance to redshift z and the integral takes account of the reduction or dilution of clustering due to the chance alignments of uncorrelated galaxies at significantly different redshifts along the line of sight. This dilution effect is larger when the sample covers a wider range of redshift. Re-writing this following the notation in Kitanidis et al. (2019), with the centre-of-mass, $\bar{r} = (r_1 + r_2)/2$, relative coordinates, $\Delta_r = r_2 - r_1$ and where $f(\bar{r})$ is the normalised radial distribution, the equation becomes:

$$w(\theta) = \int_0^\infty \mathrm{d}\bar{r}\, f(\bar{r})^2 \int_{-\infty}^\infty \mathrm{d}\Delta r\, \xi(R, \bar{r}). \tag{4.3}$$

Previous studies showed that the observed correlation function can be modelled as a single power law in r and z up to separations of about $\simeq 10h^{-1}$Mpc (e.g. Davis and Peebles 1983; Gaztanaga 1995; Maddox et al. 1996):

$$\xi(r) = \left(\frac{r_0}{r}\right)^\gamma (1+z)^{-(3+\epsilon)}, \tag{4.4}$$

where r_0 is the clustering length, the scale at which $\xi = 1$, and γ is the power-law slope. When the clustering properties do not evolve with proper coordinates, we have $\epsilon = 0$ (Gaztanaga, 1995). Assuming this power-law form for the correlation function, equation. 4.3 becomes:

$$w(\theta) = \theta^{1-\gamma}\, r_0^\gamma\, \sqrt{\pi}\, \frac{\Gamma(\gamma/2 - 1/2)}{\Gamma(\gamma/2)} \int_0^\infty \mathrm{d}\bar{r}\, f(\bar{r})^2\, (1+z)^{(\gamma-3)}\, \bar{r}^{1-\gamma}, \tag{4.5}$$

This final equation can be considered as $w(\theta) = A_{\gamma,r_0}\theta^{1-\gamma}$, where the integral and Γ functions have been absorbed into a constant, A_{γ,r_0}, whose value is set by the choices for γ and r_0. Plotting $w(\theta) \times \theta^{-(1-\gamma)}$ will result in a constant if the power law model is a good description of the measured angular clustering. As one can see, there is a degeneracy between the inherent clustering amplitude and the redshift distribution of the galaxies in the sample. In what follows, we will fit the observed angular clustering with this theoretical prediction in order to extract the clustering length r_0 and slope γ, using the $\mathrm{d}N/\mathrm{d}z$ from the MXXL lightcone simulation (Smith et al., 2017) which matches the expected BGS redshift distribution. We note that the values of these functions and parameters that describe the BGS clustering properties could be used to create more realistic mock catalogues.

4.1.2 Consistency between BASS/MzLS and DECaLS

First, we test the consistency between clustering in the three imaging surveys when considering BGS Bright. Fig. 4.1 shows the angular correlation function of the BGS targets in BASS/MzLS (blue), in DECaLS-NGC (red) and in DECaLS-SGC (green), together with the angular clustering from the MXXL BGS lightcone (black). Given the choice of quantity plotted on the vertical axis, the plateau we see up to angular scales of $\simeq 2$ deg shows that the power-law form is an excellent description on these scales with $\gamma \simeq 1.8$. Beyond $\simeq 2$ deg there is a rapid reduction in the clustering away from the small-scale power law. We can see a very good agreement overall between the three imaging surveys and the MXXL lightcone, which is further confirmed by the results of the fitting given in Table 4.1. We find a consistent clustering length and slope between the three imaging surveys and the MXXL. Comparison with previous measurements using SDSS EDR (Stoughton et al., 2002), SDSS DR7 (Wang et al., 2013), and the APM (Maddox et al., 1990) indicate that the angular clustering of the DESI BGS sample has a steeper slope (i.e. the clustering strength drops more rapidly with increasing angular separation) which can be explained by the dependency in the redshift distribution number counts (dN/dz) from equation 4.2. The BGS redshift distribution is much larger than previous surveys.

In order to investigate the impact of any potential remaining imaging systematics, we also look at the angular correlation function on large scales. Fig. 4.2 shows the angular clustering up to 20 deg for the three imaging surveys and the MXXL. The solid curves correspond to the nominal configuration, the dashed ones to the case where we remove regions of high stellar density (i.e. we keep stellar density $< 1000/\mathrm{deg}^2$), and the dotted curves to the case where we remove regions of low Galactic latitude (we keep $|\mathrm{b}| > 30$ deg). These two tests have a negligible impact on the clustering given the size of the error bars at these large scales. The overall agreement is reasonably good. At angular scales between 5 and 15 deg,

Table 4.1: Best-fitting values for the clustering length, r_0, and power-law slope, γ, of the BGS targets in BASS/MzLS, DECaLS-NGC and DECaLS-SGC, compared to the results from the MXXL lightcone simulation, when using a power-law approximation over the fitting range $0.001 < \theta < 1$ deg.

dataset	r_0 [h^{-1}Mpc]	γ
BGS Bright		
BASS/MzLS	5.477 ± 0.117	1.792 ± 0.007
DECaLS-NGC	5.653 ± 0.118	1.781 ± 0.007
DECaLS-SGC	5.010 ± 0.079	1.818 ± 0.005
MXXL	4.817 ± 0.106	1.789 ± 0.006
$15 < r_{\rm mag} < 16$		
BASS/MzLS	6.173 ± 0.703	1.642 ± 0.033
DECaLS-NGC	4.413 ± 0.498	1.761 ± 0.036
DECaLS-SGC	5.446 ± 0.558	1.698 ± 0.034
MXXL	5.731 ± 0.628	1.736 ± 0.039
$16 < r_{\rm mag} < 17$		
BASS/MzLS	5.889 ± 0.359	1.744 ± 0.021
DECaLS-NGC	5.309 ± 0.448	1.761 ± 0.027
DECaLS-SGC	5.962 ± 0.368	1.715 ± 0.022
MXXL	6.189 ± 0.181	1.753 ± 0.029
$17 < r_{\rm mag} < 18$		
BASS/MzLS	5.844 ± 0.198	1.776 ± 0.012
DECaLS-NGC	6.226 ± 0.275	1.746 ± 0.015
DECaLS-SGC	5.514 ± 0.225	1.793 ± 0.015
MXXL	5.909 ± 0.206	1.788 ± 0.012
$18 < r_{\rm mag} < 19$		
BASS/MzLS	5.360 ± 0.146	1.750 ± 0.008
DECaLS-NGC	5.444 ± 0.237	1.742 ± 0.013
DECaLS-SGC	5.393 ± 0.122	1.745 ± 0.007
MXXL	4.590 ± 0.140	1.803 ± 0.007
$19 < r_{\rm mag} < 20$		
BASS/MzLS	5.286 ± 0.098	1.725 ± 0.006
DECaLS-NGC	5.336 ± 0.122	1.720 ± 0.007
DECaLS-SGC	5.032 ± 0.100	1.740 ± 0.006
MXXL	4.382 ± 0.107	1.774 ± 0.006

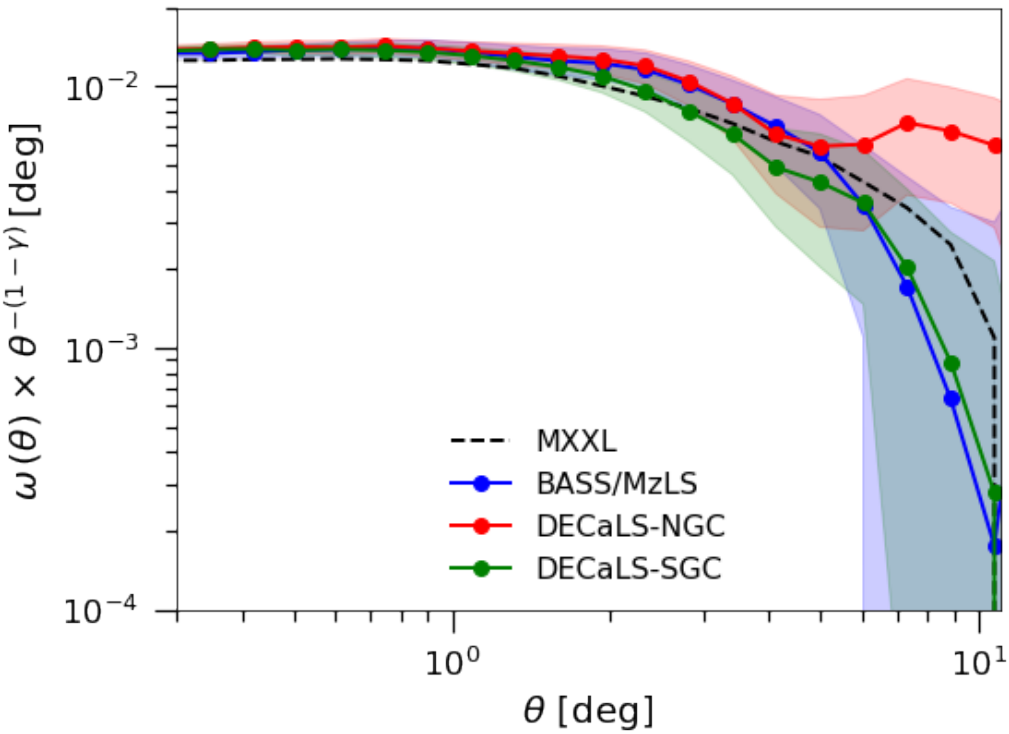


Figure 4.1: Angular clustering of the BGS targets in BASS/MzLS (blue), DECaLS-NGC (red), DECaLS-SGC (green), together with the results from the MXXL BGS lightcone (black). We have scaled the angular correlation function by $\theta^{-(1-\gamma)}$ using $\gamma = 1.8$ to highlight departures from the power law recovered at small angular separations. The shading shows the 1-σ error estimated using a jackknife resampling of the data.

DECaLS-NGC seems to have a higher amplitude but again, the errors bars are important at these very large angular scales. One may question the validity of the jackknife errors at these scales. In order to test this we computed the error bars using 10, 25 and 50 jackknife regions and compared with the errors when using 100 jackknife regions. We notice a slight under-estimation when increasing the size of the jackknife region as expected, but otherwise the effect remains small which validates our interpretation of Fig. 4.2: the difference in clustering amplitude in this regime is consistent with being due to a statistical fluctuation.

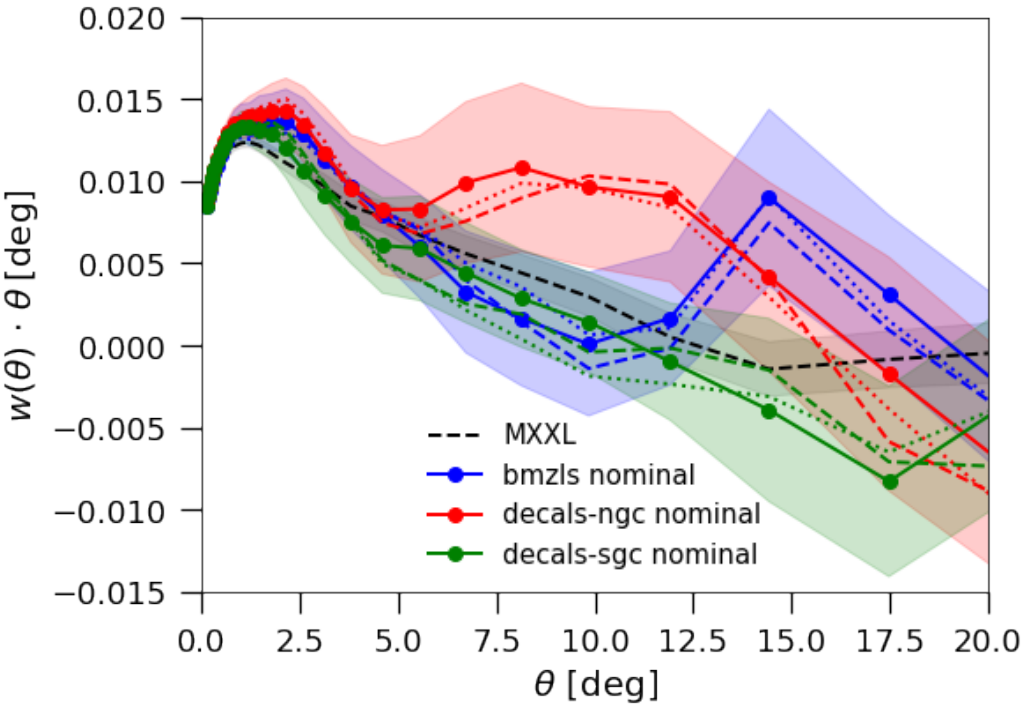


Figure 4.2: Angular clustering of the BGS targets when removing regions of high stellar density (dashed) or low Galactic latitude (dash-dot) compared to the original case (solid); these three estimates are consistent within the 1-σ jackknife errors shown by the shaded regions. The angular clustering at large scales is also shown for the MXXL BGS lightcone (dashed black). Note in this plot the angular correlation function is plotted multiplied by θ.

4.1.3 Clustering as a function of magnitude

As an additional check for systematics, we compute the angular correlation function for different apparent magnitude bins and compare the results of the BGS targets with the MXXL simulation as shown in Fig. 4.3. The quantity plotted on the y-axis, $w(\theta) \times \theta^{-(1-\gamma)}$ with $\gamma = 1.8$, was chosen such that one can see the domain of validity of the power-law form, as for Fig. 4.1. Table 4.1 presents the results of the power-law fitting on both DR9 and MXXL for the five apparent magnitude bins we consider.

In order to help interpret these results we first quantify some properties of the matching MXXL mock catalogue. In Fig. 4.4, we see that the distribution of absolute magnitude has very little dependence on the apparent magnitude range

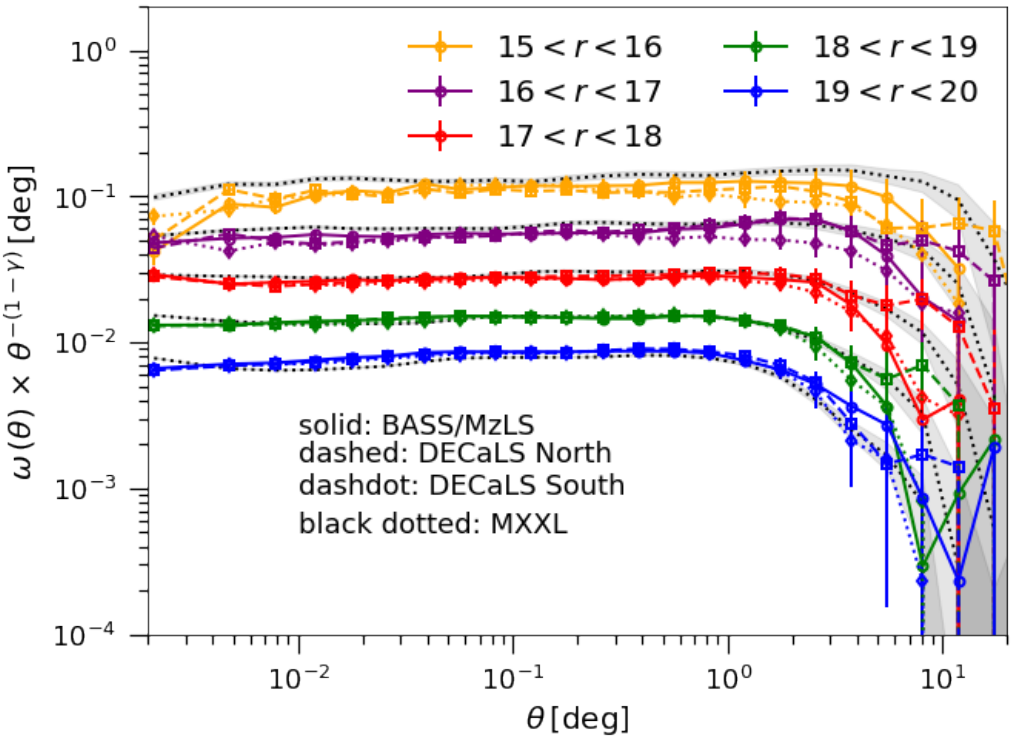


Figure 4.3: Angular clustering as a function of the apparent magnitude in the r-band for BASS/MzLS (solid), DECaLS-NGC (dashed), DECaLS-SGC (dashdot), together with the results from the MXXL BGS lightcone mock (dotted).

of the sample. Hence we would expect each of our apparent magnitude samples to be dominated by galaxies of the same absolute magnitude and hence have similar 3-dimensional clustering, $\xi(r)$. The main way in which the samples differ is in their normalized dN/dz shown in the bottom panel of Fig. 4.4. The shallower more sharply peaked dN/dz of the brighter samples will lead to stronger angular clustering, due to the $(dN/dz)^2$ term in Limber's equation (Eqn. 4.2), and the break away from the small scale power-law will occur on larger angular scales due to a fixed comoving separation subtending a larger angle at low redshift. This is precisely how the observational results shown in Fig. 4.3 behave.

To summarise, we find an overall consistent clustering strength and slope between the three imaging surveys and MXXL. Moreover, compared to the reference SDSS measurements (Wang et al., 2013), the DESI BGS allows us to obtain more precise measurements due to the larger size of the sample and greater reliability on large

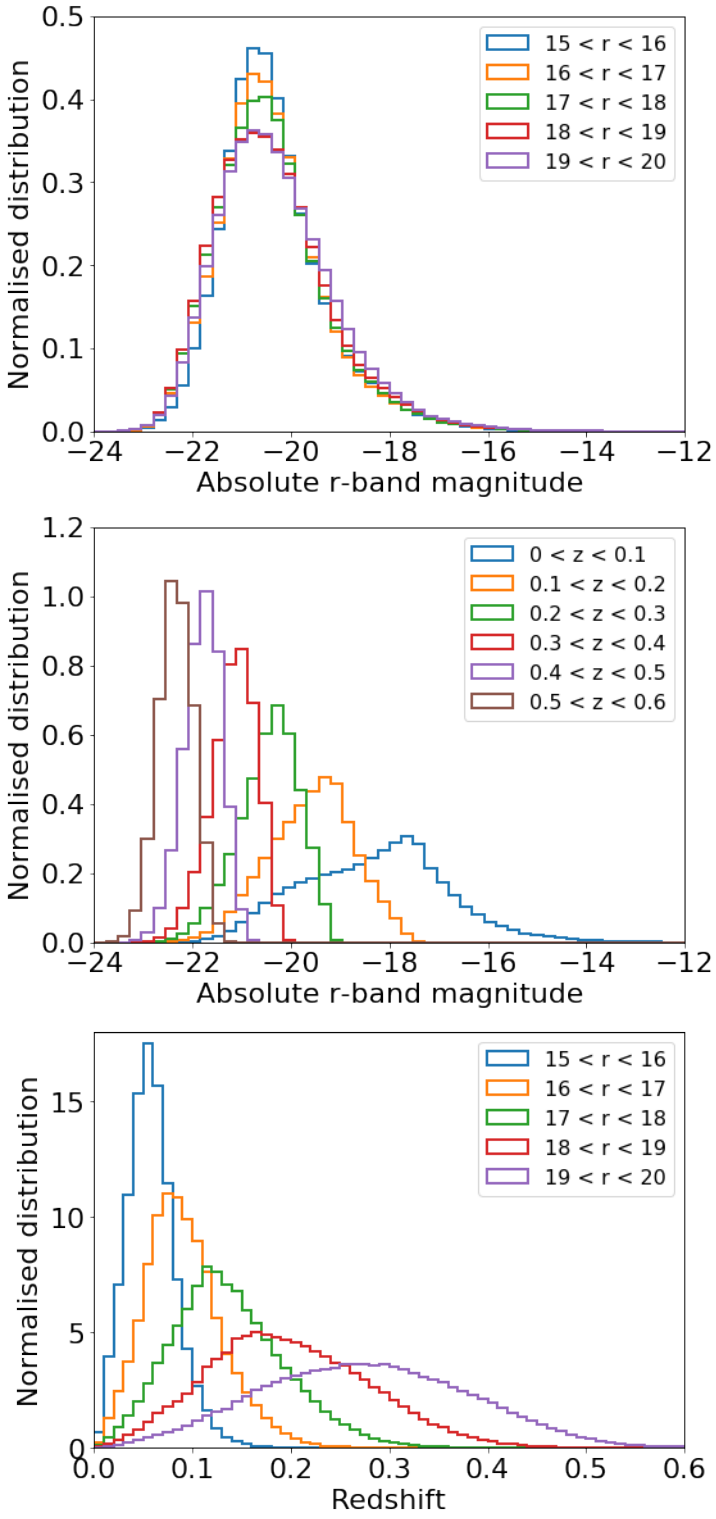


Figure 4.4: Top: Normalised absolute r-band magnitude distribution in MXXL for different apparent r-band magnitude slices. Middle: Normalised absolute r-band magnitude distribution in MXXL for different redshift slices. Bottom: Normalised redshift distribution in MXXL for different apparent r-band magnitude slices.

scales.

4.1.4 Clustering as a function of colour

A galaxy's colour reflects its composite stellar population, which in turn depends on its star formation history, the chemical enrichment history of the star-forming gas, and the attenuation of the starlight by dust; these processes are influenced by the mass of the galaxy's host dark matter halo (for reviews, see Conroy 2013; Somerville and Davé 2015). Therefore, massive galaxies with red colours typically have older stellar populations while galaxies with intermediate masses are bluer and younger with higher star formation rates.

In order to disentangle the colour, luminosity and redshift dependence of the galaxy clustering, we compute the colour-dependent clustering in two apparent magnitude bins for both BGS DR9 and MXXL. For each apparent magnitude bin, we split the sample into the 50 per cent bluest galaxies and 50 per cent reddest galaxies using $g - r$ colour. We found that considering a fixed fraction of blue/red galaxies instead of fixed colour cuts results in a fairer comparison between BGS DR9 and MXXL, as the colour distribution in the MXXL simulation does not match perfectly that of the observations, particularly at fainter magnitudes. The results of this exercise are shown in Fig. 4.5, where the top panel corresponds to galaxies with $17 < r_{\rm mag} < 18$ and the bottom panel to galaxies with $19 < r_{\rm mag} < 20$. For each magnitude bin the angular clustering of blue and red galaxies is shown for DR9 BGS (solid) and the MXXL simulation (dotted). As expected, we can see that red galaxies are more strongly clustered than blue ones at intermediate to small angular separations. The overall agreement with the lightcone is good over a large range of angular scales, which thus validates the colour-assignment procedure in MXXL.

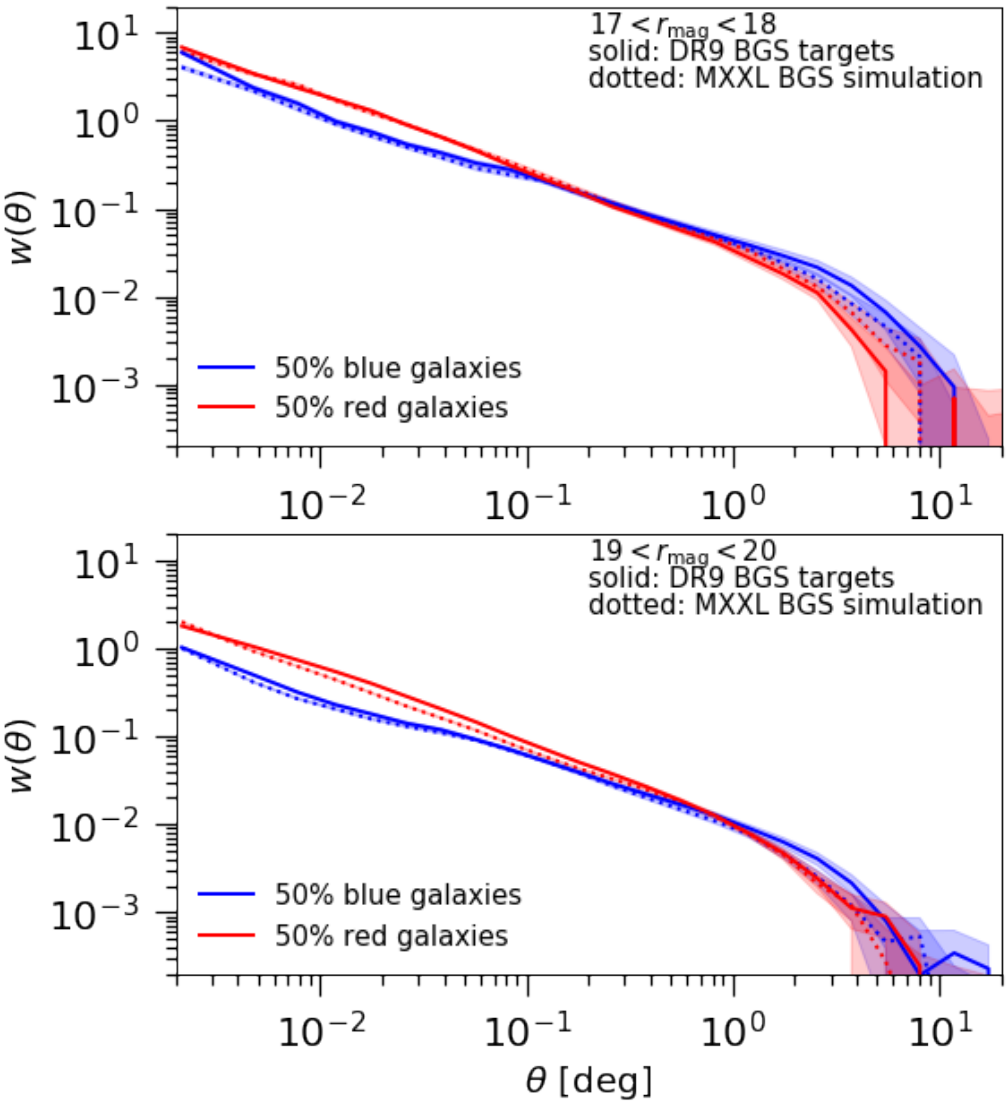


Figure 4.5: Angular clustering of the BGS targets with $17 < r_{\rm mag} < 18$ (top panel) and $19 < r_{\rm mag} < 20$ (bottom panel), for samples divided by colour into red and blue galaxies. The BGS measurements are shown by solid lines. The results using the MXXL BGS lightcone mock are also shown (dotted) for the same configurations as for the data.

4.2 Higher-order statistics using counts-in-cells

If the density field is a purely Gaussian random field, then its probability distribution function can be described by just two numbers: the mean and the variance. A Gaussian primordial density field is well supported by observations of the

CMB (Planck Collaboration et al., 2018). However, one can easily show that, in the gravitational instability scenario, this primordial distribution of density fluctuations will evolve into a distinctly asymmetric density field. Thus, the observed higher-order moments of the local density field contain information besides the two-point statistics, such as the departure from Gaussianity, which can inform us about the growth of cosmic structures, and more specifically on the bias between galaxies and the underlying matter distribution (see the review by Bernardeau et al. 2002). Moreover, in order to produce more realistic mock catalogues, it is essential to reproduce the higher-order clustering statistics of the BGS sample, especially for regions of high density where spectroscopic incompleteness due to the finite size of the fibre allocation* has a significant impact on clustering (Burden et al., 2017; Hahn et al., 2017; Bianchi et al., 2018; Smith et al., 2019; Bianchi and Verde, 2020). These higher-order statistics can be explored using counts-in-cells (CIC, see for example White 1979; Peebles 1980; Fry and Gaztanaga 1993).

The CIC analysis of projected galaxy counts in wide-field galaxy surveys has a long history, stretching back to visually measured counts on photographic plates (Groth and Peebles, 1977). Gaztanaga (1994) measured the distribution of CIC up to ninth order from the Automated Plate Machine survey (Maddox et al., 1990), showing that the galaxies are essentially unbiased tracers of the matter distribution on large scales. Ross et al. (2006, 2007) applied CIC to the third release of SDSS in order to measure the higher-order angular correlation functions of SDSS that can be used for testing the hierarchical clustering model and higher-order bias terms. Salvador et al. (2019) developed the technique to measuring the linear and non-linear galaxy bias of the Dark Energy Survey Science Verification data. More recently, Repp and Szapudi (2020) developed a theoretical prediction of the CIC galaxy probability function as a function of σ_8 and b to measure these parameters from the SDSS Main Galaxy Sample.

With the goal of providing a complete characterisation of the clustering prop-

*The DESI patrol radius is about 1.4 arcmin which corresponds to 0.017 deg.

erties of the BGS sample, in this section we investigate the higher-order statistics of the density field up to fourth order: mean, variance, skewness and kurtosis by making use of the CIC method. We use the HEALPix* package (Górski et al., 2005) which divides the sky such that each pixel covers the same surface area. This method works for the entire DESI footprint, unlike the one used in Kitanidis et al. (2019), which is based on a transformation of the angular coordinates into cartesian coordinates. This is a good approximation for regions close to the Galactic plane, such as the rectangles defined in their paper for DECaLS, but is no longer valid when considering BASS/MzLS for instance. In the HEALPix pixelation, the lowest resolution partition is comprised of 12 base pixels and the resolution increases by dividing each pixel into four new ones such that $N_{\rm side} = 2^{\rm resolution}$ is the number of pixels per side and $N_{\rm pix} = 12 \times N_{\rm side}^2$ is the total number of pixels in the map. In what follows, we consider resolutions above 4 to the maximum 10. The maximum resolution corresponds to a cell size of roughly 0.06 deg across which is larger than the DESI fibre patrol radius. For each resolution of the HEALPix maps, we remove pixels that are not fully within the survey boundaries by determining a threshold based on the expected number density using the random catalogue. The threshold is determined such that these outliers in the HEALPix pixels distribution are removed while decreasing the effective area by less than 10%, as confirmed in Fig. 4.6 which shows the difference in effective area after and before removing the outliers for BASS/MzLS (red), DECaLS-NGC (blue) and DECaLS-SGC (green) based on the random catalogue.

For each resolution of the HEALPix map, we compute the effective mean density per square degree, the standard deviation, the skewness and the kurtosis. Szapudi and Colombi (1996) showed that the CIC statistics are sensitive to sample variance (shot noise, edge effects and finite volume) and to measurements errors due to the finite number of sampling cells. Szapudi (1998) proposed a method of infinite oversampling that enables the noise that is introduced by having only one set of

*https://healpix.sourceforge.io

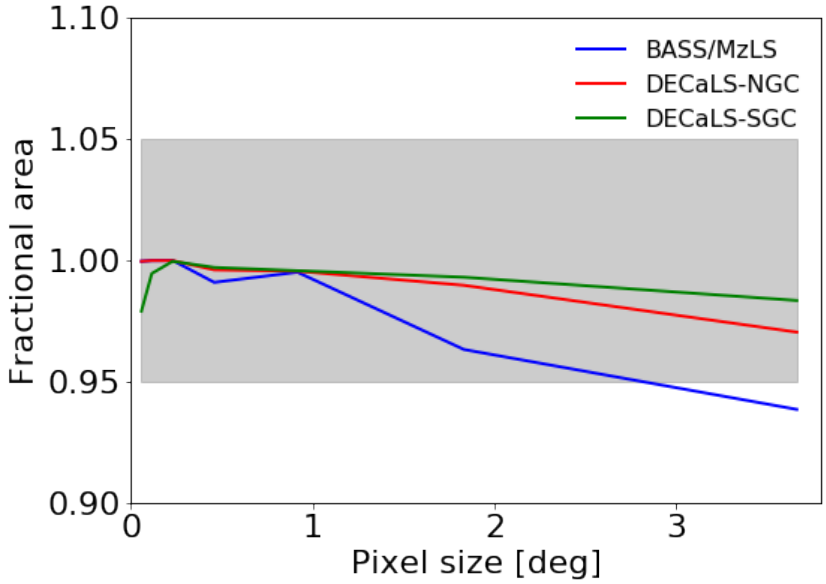


Figure 4.6: Ratio of the effective area after and before removing the outliers in the HEALPix distribution based on the random catalogue for the three imaging surveys: BASS/MzLS (blue), DECaLS-NGC (red), DECaLS-SGC (green). The shaded grey region shows 10% variation of this fractional area.

sampling cells to be beaten down and thus to eliminate the measurement errors. In order to reproduce this oversampling effect, we dither by a fraction of cell size each HEALPix map, compute the CIC statistics for each rotation and take the average. In practice, first we convert the RA, DEC into x, y, z coordinates and then rotate the coordinates by an angle ϕ (in degrees) around an arbitrary rotation axis vector. The angle ϕ is randomly chosen in a Gaussian distribution of width the HEALPix cell size (we also tried twice and five times the HEALPix cell size). Eventually we convert back the shifted x, y, z into new RA, DEC. We do 5 rotations and compute the mean and standard deviation of each quantity above.

We did not find any shift in the mean value of each CIC statistics associated with this shifting of pixels, which confirms that we are carrying out a robust sampling and that the tails of the counts distribution are well measured and not unduly affected by the sampling of extreme voids or overdensities. In order to estimate errors, we define a set of 100 jackknife regions, the same set for every pixel size, and we compute the effective density, standard deviation, skewness and kurtosis

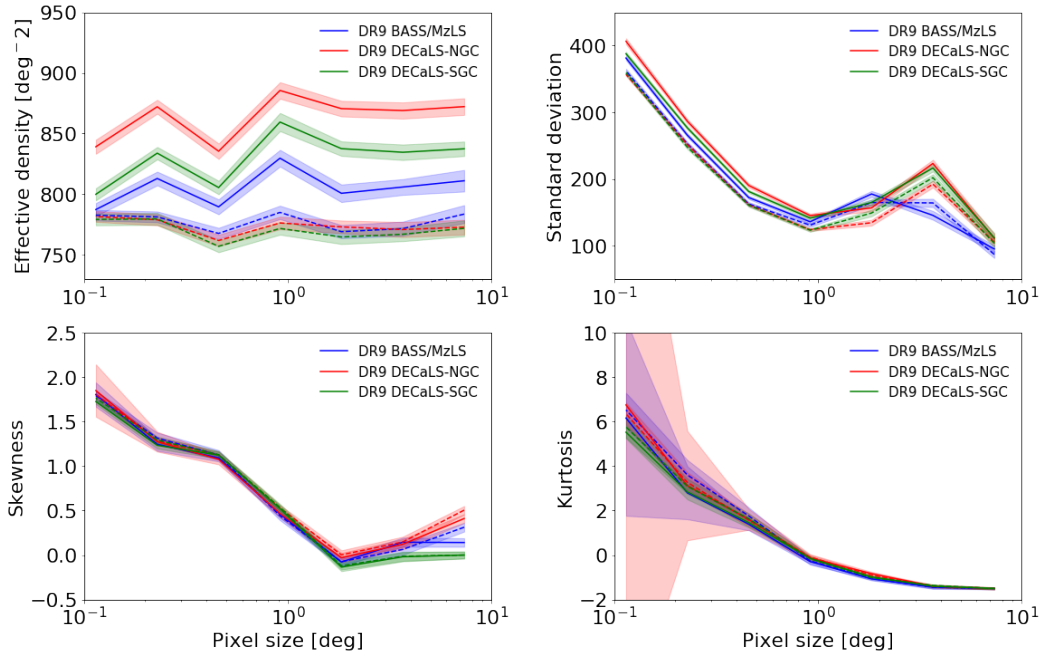


Figure 4.7: Effective mean density, standard deviation, skewness and kurtosis as a function of HEALPix cell size in degrees for the MXXL lightcone (dashed) and for the BGS DR9 targets (solid) where both are restricted to the same imaging region with errors bars from 100 jackknife regions.

in each region, take the mean and the standard deviation. First, we test the procedure using the MXXL lightcone that we split into BASS/MzLS, DECaLS-NGC and DECaLS-SGC regions. Fig. 4.7 shows the results of the MXXL lightcone in dashed for each statistic as a function of HEALPix cell size for BASS/MzLS (blue), DECaLS-NGC (red), DECaLS-SGC (green) with the coloured regions representing the 1-σ errors from the 100 jackknife regions. As expected, the measurements for the different MXXL regions all agree to within the errors. The solid curves show the same results for the BGS DR9 targets with their jackknife errors. The values of the target density for the three imaging surveys are consistent with the ones given in Section 3.3 with 7% difference at maximum when correcting for the magnitude and colour shift between BASS/MzLS and DECaLS.

The other statistics show a better agreement between the three imaging regions of the BGS data and with the MXXL lightcone, even for the third (skewness) and fourth (kurtosis) moments of the galaxy density field although no direct in-

formation about these higher-order statistics was included in the creation of the MXXL lightcone for BGS. We note that both the skewness and kurtosis are non-zero as expected from a primordial Gaussian random density field which evolved under gravitational instability and led to the hierarchy of gravitationally-bound structures that form the cosmic web with filaments, sheets, knots and voids.

4.3 Conclusions

The angular clustering shows very good consistency between the three imaging surveys and with the MXXL BGS lightcone (Smith et al., 2017). It is also consistent with a power-law model on angular scales below 1 deg and it gives comparable clustering strength r_0 and slope γ, both between the three imaging surveys, and with the mock and with previous measurements in the literature.

The two-point angular clustering and the higher-order clustering using the counts-in-cells technique shows a good agreement between DESI and MXXL mock. Based on the construction of the MXXL mock catalogue, we can say that the BGS objects are consistent with a ΛCDM universe populated with galaxies using a standard HOD description to relate the galaxy distribution to the mass distribution.

The two-point angular clustering strength depends strongly on both apparent magnitude and colour. These trends also agree well with those in the MXXL mock indicating that the prescription used in the mock in which the fraction of red galaxies increases with halo mass is realistic.

Chapter 5

Assessment of BGS selection in DESI Survey Validation (SV)

5.1 Introduction to the DESI SV

DESI will conduct surveys of bright galaxies, luminous red galaxies (LRGs), emission line galaxies (ELGs), and quasars (QSOs). These samples, numbering roughly 30 million spectroscopic targets in total, will constrain the cosmic distance scale using the apparent size of the baryonic acoustic oscillations (BAO) and, using the measured power spectrum, extract cosmological information such as constraints on the neutrino masses, the nature of modified gravity, and the physics of inflation. DESI will also obtain spectra of millions of stellar sources to probe the physics associated with stellar evolution and the formation of the Milky Way. Because the surface density and faintness of the planned DESI samples far exceeds the capabilities of current spectroscopic facilities, such catalogues have not been extensively explored. It is therefore essential to explore the quality of the selection algorithms and spectra using the DESI instrument itself before the commencement of the five year program. Can the instrument measure redshifts at the desired success rate in the stated exposure time? If the exposure time is varied, how does the success rate change? These are some of the questions that DESI Survey Validation is trying to

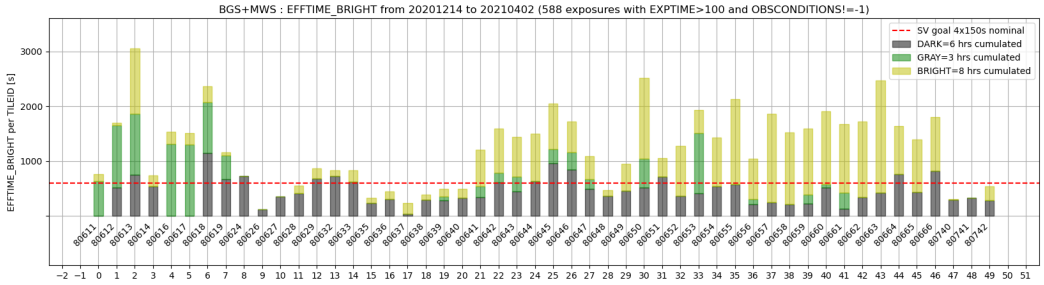


Figure 5.1: Cumulative effective exposure time per BGS+MWS tile for DESI SV. The histogram shows the cumulative exposure time per tile with the colour coding indicating the contributions coming from dark, gray and, bright conditions. The x-axis labels indicate the tile IDs. The red dashed line shows the SV target exposure time which equals four times the nominal BGS exposure time.

answer. Depending on the answers, the target selection for the different surveys might be fine tuned, prior to full survey mode starting.

DESI conducted observations prior to the start of the primary survey (scheduled for May 2021) in a phase of 'Survey Validation' or SV for short. In this chapter we make use of these observations to test the quality of these data against the science requirements for the BGS through the target selection decision tree described in Section 5.2.

The DESI observation programme is made up of three components: i) the dark time programme, ii) the bright time programme, and iii) the backup programme. BGS targets and Milky Way stars (MWS) will be observed mostly during bright time. In Section 5.4 we give more details about how DESI switches between these three programmes each night, specifically for the bright time observations.

SV observations were carried from mid December 2020 to early April 2021, with a total of 1931 exposures over 76 nights. From these exposures, 588 were dedicated to BGS+MWS targets and have an effective exposure time in bright time greater than 100 seconds, considered as the minimum exposure time to get useful data. The effective exposure time for bright time is defined in Eq. 5.2. A summary of these observations in this effective exposure time per tile can be found in Fig. 5.1. In this plot, the x-axis shows the tile IDs, and the corresponding cumulative exposure time for the BGS+MWS survey programme in dark, bright, and backup conditions.

In Section 5.2 we describe the decision tree towards the final BGS target selection. An overview of the SV target selection algorithms is provided in Section 5.3. In Section 5.4 the survey speed and the effective exposure time for the bright time programme is defined. In Section 5.5 we estimate the redshift success rate. In Section 5.6 a new colour-based selection is introduced to BGS FAINT in order to improve the redshift completeness at faint magnitudes. In Section 5.7 we test our star-galaxy separation. In Section 5.8 we assess the quality cuts, and in Section 5.9 we define the BGS target selection for the main survey. We use SV observations to assess the BGS target selection following the instructions set out in Section 5.2. The conclusions of this chapter are given in Section 5.10.

5.2 BGS decision tree for target selection

Using the SV data, we build a target selection algorithm very close to the one that will be used in the final survey and attempt to understand the data and redshift fitting well enough to determine whether the metrics on redshift performance will be met. In preparation for this, we define a decision tree according to how we characterize the BGS sample.

The biggest questions relating to the optimization of the BGS sample pertain to identifying samples within the SV selections that have low contamination, and high redshift efficiencies with the smallest exposure times possible.

- **Star-galaxy separation, spatial and quality cuts:** We test our *Gaia*-based star-galaxy classification in SV by targeting extended objects that fail the $G - rr > 0.6$ cut. We will consider the default star-galaxy classification criteria to be acceptable if the stellar contamination for the combined BGS sample is below 2 per cent. Otherwise we will consider more restrictive criteria balancing a trade off between missing galaxies and removing stellar contaminants. We will also tune our spatial and quality cuts by looking at

the galaxies that fail these cuts. As well as the star-galaxy classification, we will scrutinize any flag or mask that excludes more than 1% of galaxies that would otherwise satisfy the target selection criteria. If the contaminant fraction of these flagged/masked galaxies is below 20%, then we will restore these galaxies (or relax the flag/mask criterion).

- **Assess fiducial exposure times:** The wide range of galaxies being targeted in BGS vary in surface brightness, total magnitude, and emission line strengths. The exposure times required to determine a robust redshift will vary depending on such galaxy properties and the redshift. We will determine the relationship between exposure time, galaxy properties, and photometric criteria. Of particular importance will be the question of whether a subset of BGS targets requires more than one exposure to achieve competitive redshift success rates.

- **Adapt strategy based on BGS bright:** We seek criteria that satisfy the 'Level 2' survey requirements: a surface density > 700 objects/deg^2, including an input catalogue of at least 80% of all galaxies with $r \leq 19.5$. We choose the combination of criteria that satisfy these requirements, along with the 95% redshift completeness target, whilst using the smallest value of the baseline exposure time, ($T_{\rm bright}$), as this will give the largest footprint. If this value of $T_{\rm bright}$ yields a footprint of > 9000 deg^2 with 22% contingency or margin in the schedule, then we adopt these modified criteria and this value of $T_{\rm bright}$. Because reducing the footprint below $14,000$ deg^2 allows us to avoid the areas of sky that have the lowest efficiency in terms of redshift success rate, observing conditions or stellar contamination, we anticipate that we can gain more time than is suggested by the fractional area lost, e.g., a $12,000$ deg^2 survey should be completed in significantly less than 6/7 of the time of a $14,000$ deg^2 survey by avoiding the galactic plane. If this value of $T_{\rm bright}$ does not yield a footprint of > 9000 deg^2 with 22% margin, then we must decide which requirement to relax. In the worst case it seems likely that

we would first give up on the requirement of including at least 80% of the magnitude limited sample (e.g. by having an explicit fibre magnitude limit which would exclude some galaxies that would be in a pure magnitude limited sample), and second relax the requirement of achieving a surface density of > 700 deg^2. We might also decide to accept a margin smaller than 22%.

- **High redshift success rate for BGS FAINT:** There are no explicit requirements for the BGS secondary sample (i.e. BGS FAINT). Our goal is to maximize the overall scientific value of the sample, which will depend heavily but not exclusively on its spectroscopic redshift completeness. Using SV data, we will investigate which galaxies not in the primary BGS sample with $r \leq 20.0$ have the lowest redshift success for our chosen $t_{\rm BGS}$ and which galaxies with $r > 20$ have the highest redshift success. We will then consider swaps of galaxies from the second set for galaxies from the first set. We will accept such swaps if they noticeably increase the expected redshift completeness of BGS secondary sample without significantly degrading its scientific value, e.g. by making clustering analyses difficult because of complex selection criteria or by omitting scientifically important classes of galaxies.

5.3 BGS SV target selection

The BGS is the lowest-redshift sample of DESI extragalactic targets. These galaxies will be observed during the time when the Moon is significantly above the horizon, and the sky is too bright to allow efficient observation of fainter targets. Approximately 10 million of the brightest galaxies within the DESI footprint will be observed over the course of the survey, sampling redshifts $z < 0.5$ at high density. This sample alone will be ten times larger than the SDSS-I and SDSS-II main galaxy sample of one million bright galaxies observed between $1999 - 2008$.

In this Section, we compare the proposed SV BGS target selections with the nominal BGS selections for the main survey that we defined back in Section 3.2.

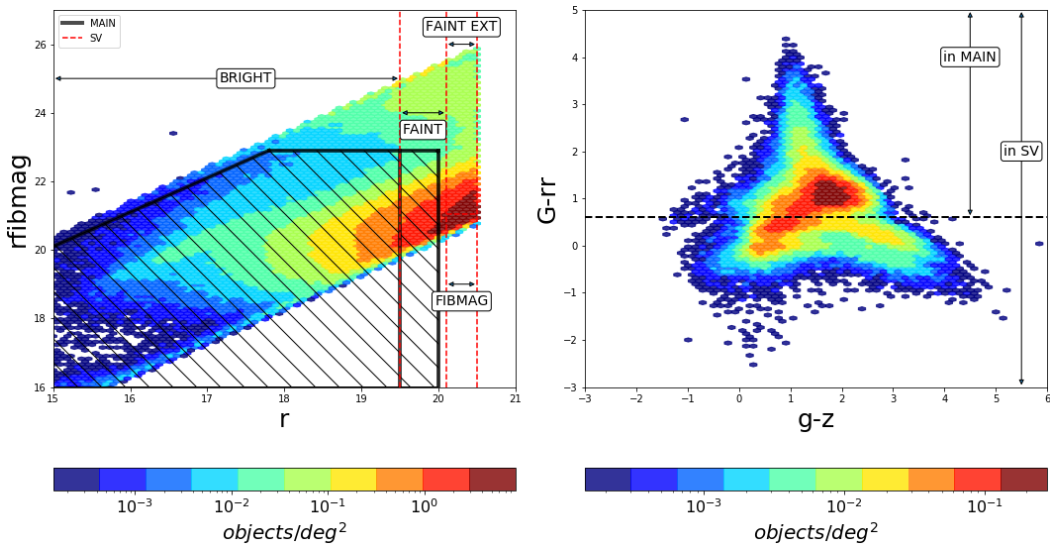


Figure 5.2: Comparison between the extended SV selection and the nominal main survey selection for BGS targets using the LS DR9 imaging catalogues. **Left**: r-band fibre magnitude versus r-band total magnitude, where r-band fibre magnitude is determined using an aperture matched to the fibre radius under nominal seeing conditions. Four of the five extended selection criteria are presented with red dashed lines. The nominal selection for the main survey is shown in the hatched black region. **Right**: *Gaia* $G - rr$ versus $g - z$ colour used to discriminate between stars and galaxies in the nominal BGS selection. Fluxes are corrected for extinction except for those used to measure the $G - rr$, where the raw magnitudes measured with the DECaLS and BASS r filters are denoted rr. Star galaxy separation for the nominal selection is denoted by the dashed-black line at $G - rr = 0.6$. In SV, objects below this threshold are not rejected if the morphological classification in the LS indicates the objects are extended (i.e., non-PSF).

The nominal selections are primarily defined by total r-band magnitude limits of $r < 19.5$ (BGS BRIGHT) and $19.5 < r < 20$ (BGS FAINT) and a magnitude-dependent, faint fibre magnitude limit that removes mainly spurious sources. The region covered by these two samples is shown by the hatched black area in the left panel of Fig. 5.2. The main changes compared with the nominal selection include:

- Stars and galaxies in BGS are separated based on the *Gaia* versus LS $G - rr$ magnitude difference in the main selection, as shown in the right panel of Fig. 5.2. In SV, we extend this selection to include all TRACTOR non-PSF objects that lie below this threshold.

- In SV, we do not apply the photometric fibre magnitude cut described in

equation 2.4.

- In SV, we do not apply the photometric low quality cut described in equation 2.6.

The BGS BRIGHT selection in SV remains the same as in the nominal selection after applying the extensions described above. However, we add another subclass to BGS BRIGHT with the peculiarity that this sample contains some of the targets that do not meet one or more of our quality criteria such as the NOBS, the CC and the QCs described in equation 2.6. The extended BGS selection in SV therefore includes:

- **Bright** ($r < 19.5$) and BGS SV.
- **Low-quality** ($r < 20.1$) and BGS SV that fails to meet one or more of the NOBS, CC and QCs cuts described in equation 2.6.

There are no explicit requirements for the faint selection in BGS regarding L2 science requirements. One goal of SV is to explore the potential of extending the main selection with fainter targets and thus maximize the overall scientific value of the sample. In Section 5.5 we will investigate which galaxies not in the BGS BRIGHT sample have the highest redshift success rates and evaluate the benefits of altering the main selection to include these targets. In the SV extended selection, we explore several alternatives to the simple faint selection proposed for the main survey:

- **Faint** ($19.5 < r < 20.1$) and BGS SV. The sample extends 0.1 magnitudes fainter than BGS FAINT for the nominal BGS selection. The goal of these observations is to explore the relationship between magnitude and redshift success rate.
- **Faint extended** ($20.1 < r < 20.5$ and $r_{\rm fib} > 21.051$) and BGS SV. The extension to even fainter magnitudes will allow us to probe whether there are

subsets within these data (e.g. satisfying colour constraints) and with fainter fibre magnitude limits for which the redshift completeness remains high.

- **Fibre magnitude extended** ($20.1 < r < 20.5$ and $r_{\rm fib} < 21.051$) and BGS SV. The extension to even fainter magnitudes will again allow us to probe whether there are subsets within these data (e.g. satisfying colour constraints) and with brighter fibre magnitude limits for which redshift completeness remains high. (This sample was separated from the previous sample for historical reasons.)

5.4 The DESI BRIGHT programme

DESI will conduct observations in three main programmes, DARK, BRIGHT and BACKUP, with BGS being conducted in BRIGHT time. The classification of whether a given night, or part of a night, is designated as DARK, BRIGHT or BACKUP is based on the concept of survey speed. The survey speed parameter, SPEED_BRIGHT, is an instantaneous estimate of the speed that survey could proceed that ignores read noise, Poisson noise, variations in atmospheric absorption and Galactic extinction. Survey speed is essentially a measurement of the sky conditions. For a given BGS observation the exposure time is varied so as try and obtain spectra of the same signal-to-noise ratio in all observations. To achieve this an effective exposure time ($T_{\rm bright}$ or EFFTIME_BRIGHT) that would be equal to the actual open-shutter exposure time under nominal conditions is defined. This effective exposure time takes account of the weather conditions and also accounts for read noise, Poisson noise, varying atmospheric absorption, and Galactic extinction.

For a single exposure, the survey speed and the effective exposure time for BGS

are given by,

$$\text{SPEED\_\_BRIGHT} = \langle|\text{FIBERFRAC}|\rangle^2 \times \langle\text{AIRFAC}\rangle^2 \times \frac{\text{SKY\_\_NOMINAL}}{\text{SKY}} \quad (5.1)$$

$$\text{EFFTIME\_\_BRIGHT} = \text{EXPTIME} \times \langle|\text{FIBERFRAC}|\rangle^2 \times \frac{\text{SKY\_\_NOMINAL}}{\text{EFFSKY\_\_BRIGHT}} \times \frac{1}{\text{EBVFAC}^2} \quad (5.2)$$

$$|\text{FIBERFRAC}| = \frac{\text{FIBERFRAC\_\_BGS} \times \text{TRANSPARENCY}}{\text{FIBERFRAC\_\_BGS\_\_NOMINAL} \times \text{AIRFAC}} \quad (5.3)$$

$$\text{EFFSKY\_\_BRIGHT} = \frac{\langle\text{SKY}\rangle + \text{SKY\_\_RDNOISE} \times 1000/\text{EXPTIME} + \text{FIBERFLUX\_\_BRIGHT}}{1 + (\text{SKY\_\_RDNOISE} + \text{FIBERFLUX\_\_BRIGHT})/\text{SKY\_\_NOMINAL}} \quad (5.4)$$

where brackets ($\langle\rangle$) denote the mean over an exposure, and the other terms are defined as follows:

- **TRANSPARENCY:** equal to 1.0 when conditions are photometric and the telescope and corrector optics are clean.
- **SKY:** photometric sky flux level in the r-band in units of nMgy/arcsec^2 where nMgy is NANOMAGGIES and 1 nMgy corresponds to an AB magnitude of 22.5.
- **SKY_NOMINAL:** is a reference value that is meant to be close to the median dark-time photometric sky level in the r-band.
- **FIBERFRAC_BGS:** corresponds to the fraction of light of an object that makes it to a DESI fibre. The assumed morphology is a de Vaucouleurs fit with a half-light radius of $r_{\text{half}} = 1.5$ arcsec. Expressed as a polynomial fit in terms of the FWHM: $\log(\text{FIBERFRAC}) = 0.0341 \times \log(\text{FWHM})^3 - 0.3611 \times \log(\text{FWHM})^2 - 0.7175 \times \log(\text{FWHM}) - 1.5643$.
- **FIBERFRAC_BGS_NOMINAL:** similar to the above quantity with a FWHM $= 2\alpha\sqrt{2^{1/\beta} - 1} = 1.1$, where $\alpha = 1.175$, and $\beta = 3.5$. For BGS, this value corresponds to 0.195, and the assumed morphology is a *de Vaucouleur* fit with a half-light radius of $r_{\text{half}} = 1.5$ arcsec.

- **AIRFAC:** airmass factor, AIRFAC = $10^{k(airmass-1)/2.5}$ where $k = 0.114$ in the r-band.
- **EBVFAC:** Galactic extinction factor, computed in the r-band for the median E_{B-V} of the targets on the tile EBVFAC = $10^{2.165(E_{B-V})/2.5}$. This quantity modifies the measured sky values to account for read noise and Poisson noise, while keeping the value unchanged when observing in nominal conditions.
- **SKY_RDNOISE:** 0.932 nMgy/arcsec^2, which is 1/4 of the nominal sky brightness value of 3.73 nMgy/arcsec^2.
- **FIBERFLUX_BRIGHT:** 1.71 nMgy/arcsec^2 in nominal seeing. This represents the mean surface brightness within the fibre aperture in typical conditions for a fiducial BGS galaxy at the BGS BRIGHT magnitude limit of 19.5.

In order for a bright time observation to take place, the survey speed needs to meet one or both of the following criteria:

1 First tile of the night shift and sun angle > -16 deg at the start of the exposure, or after 1000 seconds later.

2 Conditions are such that the survey speed in the range $1/2.5 >$ SPEED_BRIGHT $> 1/6$ (time-averaged over non-twilight time in the last 30 min).

5.5 BGS redshift success rate

For this analysis, we consider SV DEEP and SINGLE exposures. The DEEP exposures are combinations of multiple SINGLE exposures and, for the purposes of SV, can be used to approximate truth tables. We use DEEP exposures with an effective exposure time ($T_{\rm bright}$) defined by Eq. 5.2 longer than 1000 sec.

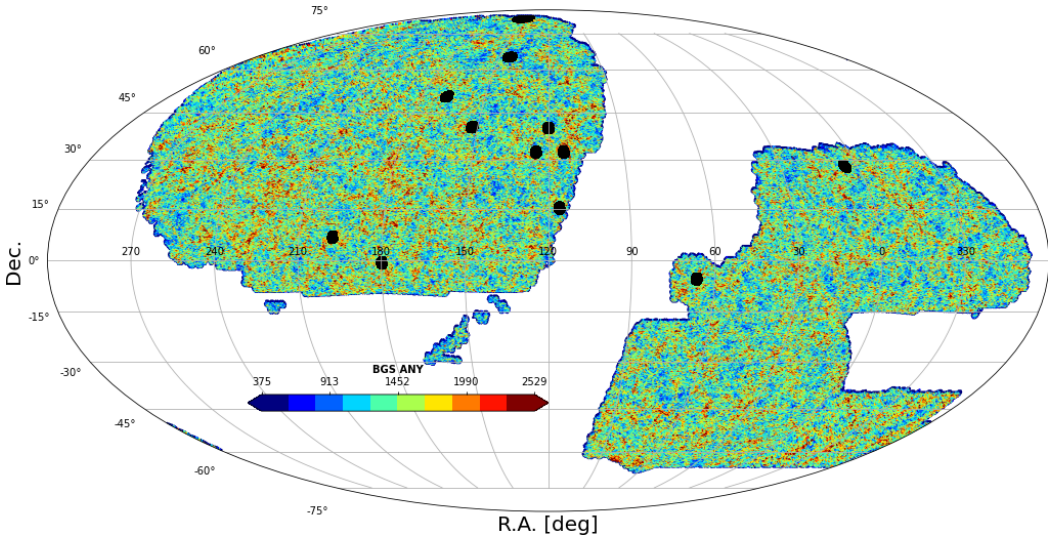


Figure 5.3: BGS SV target density across the sky. The black regions show the BGS+MWS tiles used for the analysis of the redshift success rate.

These single exposures are filtered to extract only the BGS+MWS tiles, and finally, to get a realistic sample, we take only tiles with $T_{\rm bright}$ close to the nominal 180 seconds exposure time for the BGS, however, $T_{\rm bright}$ will be tuned if necessary based on SV results. We use Equation 5.2 to estimate $T_{\rm bright}$, which allows us to correct for the weather and read-out noise, Poisson noise, atmospheric absorption and Galactic extinction.

We end up with 19 single exposure tiles of which 12 are unique tiles. Fig. 5.3 shows the BGS SV target density across the sky, and the black circles show the configuration of the tiles used for this analysis.

DESI uses the REDROCK* redshift algorithm, which uses a combination of the Bolton et al. (2012) template fitting approach and an archetype (Cool et al., 2013) approach similar to that applied in REDMONSTER (Hutchinson et al., 2016). A summary of the general process is as follows: Classification and redshift determination are performed via a fit of a linear combination of spectral templates to each spectrum. Fitting is done over a range of redshifts for three different classes of templates that independently characterize stellar, galaxy, and quasar spectral diversity. The redshift and spectral class that give the lowest value of χ^2 are

*`https://github.com/desihub/redrock`

considered the best description of the spectrum. A fit is only considered reliable, or good, if it can be differentiated from the second best fit by a sufficiently large difference in the χ^2. We denote this parameter as $\Delta\chi^2$.

To quantify the redshift success rate, we use the DEEP exposures as a truth table. For each tile, we consider only the BGS SV targets that meet the following criteria from the DEEP exposure:

- ZWARN = 0: i.e. there are no warnings from the spectral fitting pipeline.
- $\Delta\chi^2 > 40$: difference in χ^2 between fits to the spectrum with the best fitting and the second best fitting redshift (redshift fits are only considered distinct if they differ by at least 1000 km s^{-1}).
- SPECTYPE $\neq$ STAR: the the best-fitting spectroscopic type (i.e. SPECTYPE) is not a star.
- $0 < z < 0.6$: the redshift lies in the range expected for the BGS.
- $\sigma_z < 0.0005(1 + z)$: redshift error less than 150 km s^{-1}.

Any targets that do not meet these criteria are considered as targeting issues, as with these deep exposures genuine targets should easily meet all these criteria. We determine the redshift success for SINGLE exposures if the above criteria are met, along with the condition:

- $z_{\rm crit} = \dfrac{|z - z_{\rm DEEP}|}{(1 + z_{\rm DEEP})} < 0.003,$

where $z_{\rm DEEP}$ is the redshift obtained from the DEEP exposure.

Fig. 5.4 shows the redshift success rate as a function of r-band magnitude in bins, and as a function of r-band fibre magnitude in bins. The goal for BGS BRIGHT, according to BGS decision tree in Section 5.2, is to achieve 95 per cent redshift completeness for the nominal BGS effective exposure time. Our results

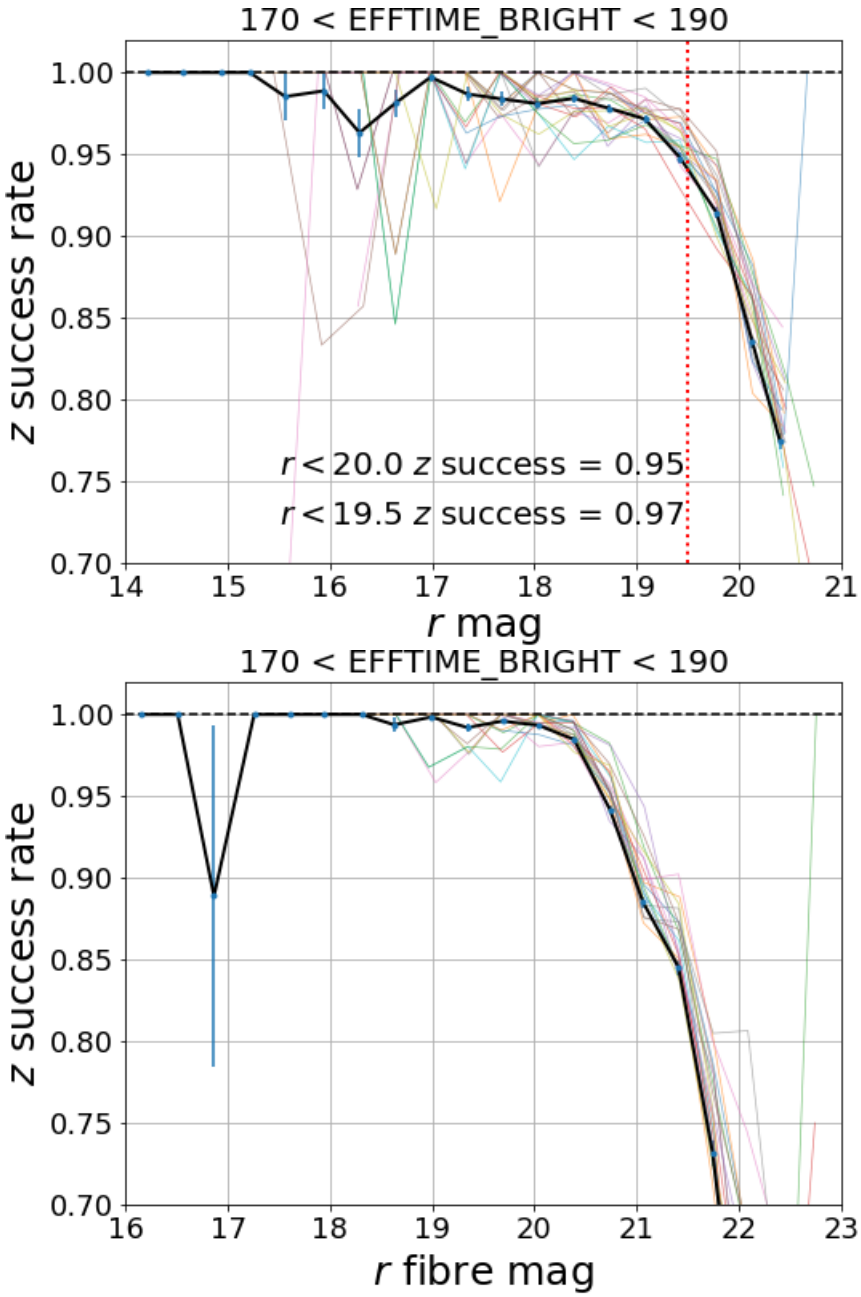


Figure 5.4: The redshift success rate in the BGS SV dataset as a function of r-mag (Top) and r-fibre mag (Bottom) in bins. The solid black curve corresponds to all 19 tiles combined while coloured curves show the results for each individual tile. Top plot shows the cumulative redshift success rates at $r < 19.5$ (97 per cent), and at $r < 20$ (95 per cent).

show that we meet and surpass this requirement by achieving a redshift completeness of 97 percent to a limiting magnitude of $r = 19.5$. While at a limiting magnitude of $r = 20$ the redshift completeness decreases to 95 per cent which is still good. However, we note that there is not a specific requirement on redshift completeness for BGS FAINT. In Section 5.6 we investigate if replacing the simple flux cut used in BGS FAINT by a colour-based selection can improve the redshift success rate.

Tables 5.1 and 5.2 give a summary of each of the 19 tiles for BGS BRIGHT SV targets and the remaining BGS SV targets ignoring the low-quality targets. Each table contains the tiles ID (TILEID), the exposure ID (EXPID), the effective time for Bright time survey ($T_{\rm bright} \equiv$ EFFTIME_BRIGHT), the tile centres (TILERA and TILEDEC), the number of BGS targets (N), the percentage of BGS objects that fail the individual redshift success rate criteria: i) $z_{\rm warn} \equiv$ ZWARN $\neq 0$, ii) $\Delta\chi^2_{\rm low} \equiv \Delta\chi^2 < 40$, iii) SPECTYPE = STARS, iv) $z > 0.6$, v) $\sigma_z^{\rm high} \equiv \sigma z > 150\,{\rm km\ s^{-1}}$, vi) $z_{\rm crit}^{\rm high} \equiv$ not in $z_{\rm crit}$, and the redshift success rate ($z_{\rm success}$) as the percentage of BGS targets that passes all the redshift success rate criteria.

5.6 Improving the redshift success rate in BGS FAINT

In BGS SV we go fainter than the nominal selection, to $r = 20.5$. This extension in magnitude allows us to investigate whether some subset of the targets fainter than the nominal faint selection have a better redshift success rate. Fig. 5.5 shows the redshift success rate in the r-mag vs r-fibre mag plane. While the r-band magnitude does not seem to be correlated with the redshift success rate, the r-fibre magnitude does show that fainter than $r_{\rm fibre} = 21.7$ the redshift success rate is lower than 70 per cent.

Using a colour combination with g, r, z, and $W1$ band magnitudes, one can separate the galaxies with strong emission lines. Fig. 5.6 shows a sample of galaxies

Table 5.1: Summary of BGS BRIGHT SV targets per tile. Information includes the tiles ID (TILEID), the exposure ID (EXPID), the effective time for Bright time survey ($T_{\rm bright} \equiv$ EFFTIME_BRIGHT), the tile centres (TILERA and TILEDEC), the number of BGS targets (N), the percentage of BGS objects that fail the individual redshift success rate criteria, and the redshift success rate ($z_{\rm success}$) as the percentage of BGS targets that pass all the redshift success rate criteria.

TILEID	EXPID	$T_{\rm bright}$	TILERA	TILEDEC	N	$z_{\rm warn}$%	$\Delta\chi^2_{\rm low}$%	STARS%	$z > 0.6$%	$\sigma_z^{\rm high}$%	$z_{\rm crit}^{\rm high}$%	$z_{\rm success}$%
80612	68644	187.26	66.68	-5.24	2426	0.16	0.82	0.00	0.25	0.0	0.33	99.18
80612	68645	184.63	66.68	-5.24	2466	0.32	0.81	0.00	0.24	0.0	0.49	99.15
80619	68673	176.87	144.00	65.00	2076	0.29	1.11	0.00	0.29	0.0	0.43	98.84
80619	68674	173.17	144.00	65.00	2046	0.29	1.27	0.00	0.34	0.0	0.44	98.73
80617	68799	176.35	5.00	28.00	2321	0.26	1.12	0.00	0.26	0.0	0.39	98.88
80617	68801	176.37	5.00	28.00	2319	0.30	0.78	0.00	0.09	0.0	0.30	99.18
80643	70734	177.26	114.00	32.37	2128	0.38	1.27	0.00	0.38	0.0	0.66	98.54
80643	70735	171.87	114.00	32.37	2116	0.28	0.95	0.05	0.33	0.0	0.57	98.82
80643	70736	172.80	114.00	32.37	2105	0.48	0.95	0.00	0.24	0.0	0.57	98.86
80644	70737	176.06	116.00	15.50	2019	0.50	1.24	0.05	0.20	0.0	0.69	98.42
80660	70762	170.58	168.38	50.46	2240	0.54	1.29	0.04	0.40	0.0	0.76	98.62
80660	70763	170.72	168.38	50.46	2253	0.53	1.29	0.00	0.40	0.0	0.80	98.54
80644	71115	183.01	116.00	15.50	2001	0.50	1.45	0.00	0.50	0.0	0.65	98.45
80645	71116	176.09	120.00	40.00	2340	0.56	1.37	0.00	0.51	0.0	0.64	98.59
80646	71273	181.91	125.00	32.37	2297	0.17	0.83	0.00	0.26	0.0	0.39	99.13
80653	72517	180.57	152.54	40.30	2306	0.13	0.69	0.00	0.00	0.0	0.17	99.31
80662	72530	174.67	180.00	-0.50	2331	0.04	0.51	0.00	0.04	0.0	0.09	99.49
80665	72716	173.36	198.00	7.00	2285	0.13	0.53	0.00	0.04	0.0	0.22	99.43
80655	74814	171.57	155.00	83.00	2126	0.42	1.08	0.00	0.28	0.0	0.42	98.87

Table 5.2: Summary of BGS SV targets with $r > 19.5$ and ignoring the low-quality targets per tile. The table contains the same information as the above table for BGS BRIGHT.

TILEID	EXPID	T_{bright}	TILERA	TILEDEC	N	z_{warn}%	$\Delta\chi^2_{\text{low}}$%	STARS%	$z > 0.6$%	σ_z^{high}%	$z_{\text{crit}}^{\text{high}}$%	z_{success}%
80612	68644	187.26	66.68	-5.24	2426	1.90	8.12	0.00	0.78	0.00	1.81	91.71
80612	68645	184.63	66.68	-5.24	2466	1.58	7.58	0.04	0.57	0.00	1.34	92.38
80619	68673	176.87	144.00	65.00	2076	2.46	8.53	0.05	0.67	0.05	1.59	91.38
80619	68674	173.17	144.00	65.00	2046	1.96	10.61	0.00	0.68	0.00	1.61	89.39
80617	68799	176.35	5.00	28.00	2321	2.54	9.69	0.00	1.16	0.04	2.24	90.26
80617	68801	176.36	5.00	28.00	2319	2.67	8.62	0.04	1.08	0.00	2.11	91.25
80643	70734	177.26	114.00	32.37	2128	3.01	9.54	0.00	1.17	0.00	2.68	90.23
80643	70735	171.87	114.00	32.37	2116	2.36	9.97	0.00	1.32	0.05	2.41	89.79
80643	70736	172.80	114.00	32.37	2105	2.80	8.93	0.00	1.43	0.00	2.95	90.78
80644	70737	176.06	116.00	15.50	2019	2.92	9.46	0.05	1.39	0.00	2.53	90.49
80660	70762	170.58	168.38	50.46	2240	2.72	8.93	0.04	1.29	0.04	2.50	90.94
80660	70763	170.72	168.38	50.46	2253	2.93	9.28	0.00	1.46	0.00	2.93	90.55
80644	71115	183.01	116.00	15.50	2001	2.75	9.85	0.00	1.15	0.00	2.20	90.10
80645	71116	176.09	120.00	40.00	2340	2.86	9.57	0.00	1.58	0.00	2.52	90.30
80646	71273	181.91	125.00	32.37	2297	2.22	8.05	0.00	1.00	0.00	1.70	91.82
80653	72517	180.57	152.54	40.30	2306	1.99	8.24	0.00	0.61	0.00	1.78	91.72
80662	72530	174.67	180.00	-0.50	2331	0.60	6.82	0.00	0.21	0.00	0.39	93.18
80665	72716	173.36	198.00	7.00	2285	2.45	8.05	0.04	1.44	0.00	2.71	91.73
80655	74814	171.57	155.00	83.00	2126	2.63	9.92	0.00	1.32	0.00	2.49	89.84

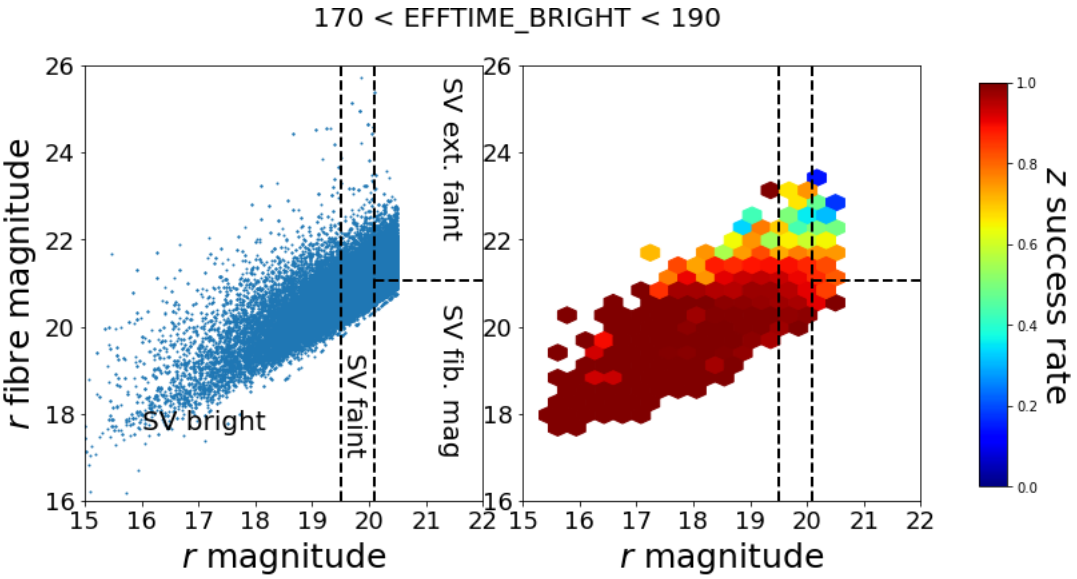


Figure 5.5: BGS SV targets as a function r-band magnitude and r-fibre mag. The colour plot shows the redshift success rate values, as indicated by the key. Black dashed lines indicate the limits for each of the BGS SV subclasses.

from the AGES survey that have $r < 19.5$ from the DECaLS DR7 photometry. The galaxies are plotted in two different colours depending on whether the AGES spectra have strong (red) or weak (black) emission lines defined by whether the Hβ equivalent width is greater or smaller than 2 Angstroms (Moustakas, priv com.). The line in the plot is the relation $(z - w1) - 3/2.5(g - r) + 1.2 = 0$ and seems to do a good job of separating out the emission line objects.

Using the $(z-w1)-3/2.5(g-r)+1.2 = 0$ relation, and the r fibre mag, we can get rid of most of the BGS objects with low redshift success rate. The colour-based selection is given by equation 5.6, and is represented in Fig. 5.7.

$$\Delta(g, r, z, w1) = (z - w1) - 3/2.5(g - r) + 1.2 \tag{5.5}$$

$$Z_{\rm high} = \begin{cases} \text{rfibmag} < 20.75 & \Delta(g, r, z, w1) \leq 0 \\ \text{rfibmag} < 21.5 & \Delta(g, r, z, w1) > 0 \end{cases} \tag{5.6}$$

We compare the two selections, i) the r-fibre mag cut, and ii) the colour-based selection for BGS SV in the faint region (i.e., $r > 19$) with the current selection

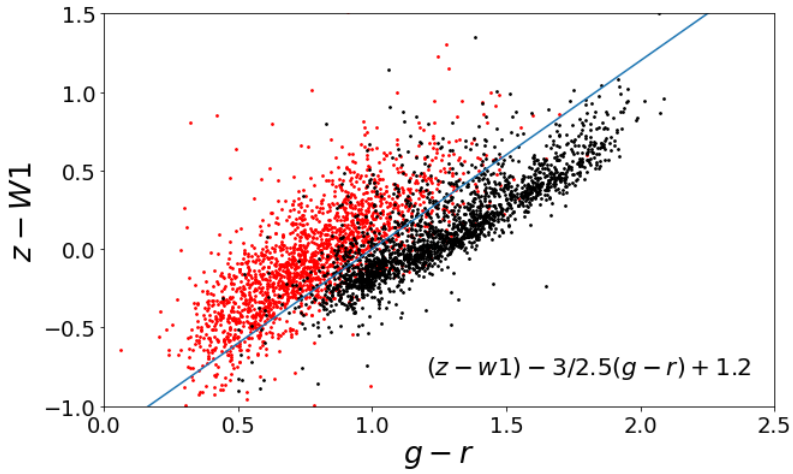


Figure 5.6: AGES galaxies with DECaLS DR7 photometry with $r < 19.5$. Red points are galaxies with strong Hβ emission lines (i.e. equivalent width greater than 2 Angstroms), while black dots are galaxies with weak Hβ emission lines. The solid blue line shows the relation $(z - w1) - 3/2.5(g - r) + 1.2 = 0$.

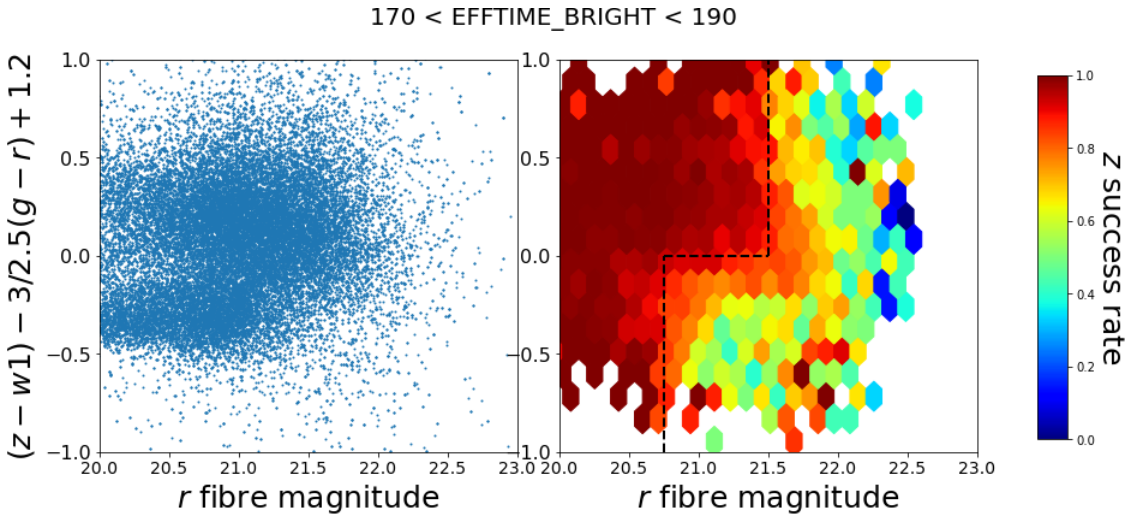


Figure 5.7: BGS SV targets as a function r-fibre mag and the colour from Eq. 5.5. The colour plot shows the redshift success rate values. Black dashed lines show the colour-based cut which is designed to increase the redshift success rate.

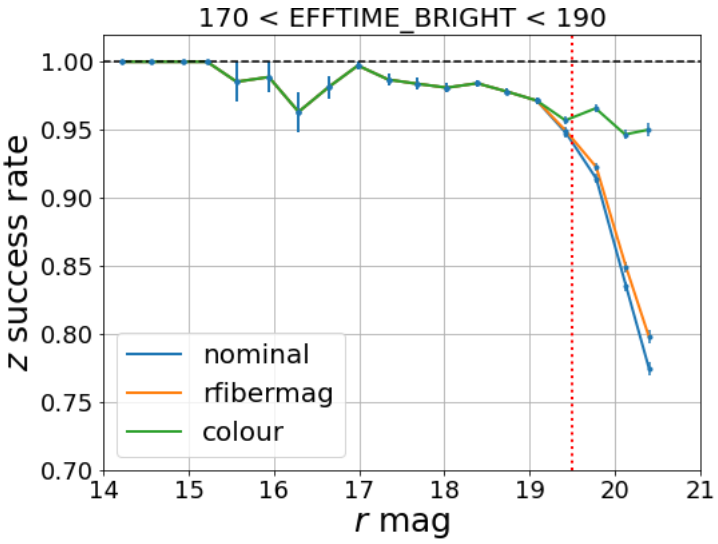


Figure 5.8: Comparison of redshift success rate as a function of r-band magnitude for three scenarios: BGS SV (blue), BGS SV and r-fibre mag cut (orange), and BGS SV and colour-based selection (green). Overall, the redshift success rate at $r = 20.3$ is 92, 93, and 97 per cent respectively, for these three cases.

(nominal). The r-fibre mag cut shows a minimal improvement in redshift success compared to the nominal selection. However, the colour-based selection has a huge impact, improving the redshift success rate to 97 per cent at $r < 20.3$, compared to the 92 per cent for $r < 20.3$ alone, as shown by Fig. 5.8.

5.7 Star-galaxy separation assessment

The aim of this section is to test the *Gaia*-based star-galaxy classification method and whether the extended star-galaxy classification used in the BGS selection for SV works better than the nominal star-galaxy classification used in Sections 2.1 and 3. Since we want to get a true estimate of the stellar contamination in the BGS SV selection, we decided to use the DEEP exposure tiles instead of the SINGLE exposure tiles for this analysis. We have a total of 26 tiles with nearly 67k targets.

We split the BGS SV selection into three disjoint categories: i) not in *Gaia* (NG), ii) in *Gaia* and $G - rr > 0.6$ ($\mathrm{IG_{main}}$), and iii) in *Gaia* and $G - rr < 0.6$ ($\mathrm{IG_{sv}}$). The subset NG contains the majority of targets with 93.4 per cent out

Table 5.3: Percentage of targets in BGS SV that fail each, and all (bottom row), of the redshift success criteria for three disjoint GAMA subsets. Percentages are expressed in terms of the total number of targets per subset.

criteria	NG %	IG_{main} %	IG_{sv} %
z_{warn}	2.0	0.3	0.8
$\Delta\chi^2_{low}$	4.0	0.7	2.0
STARS	1.6	7.1	85.3
$z > 0.6$	5.3	5.8	49.5
σ_z^{high}	0.0	0.0	0.0
$z_{failure}$	8.0	9.5	92.8

of total of the BGS targets for SV. IG_{main} contains 5 per cent, and IG_{sv} only 1.6 per cent. Among these subsets, NG and IG_{main} have less stellar contamination according to REDROCK spectra classification with 1.6 and 7.1 per cent of stars out of the total for each of this subsets respectively. On the other hand, IG_{sv} is highly contaminated by stars with 85 per cent of its targets being stars. Table 5.3 shows the percentage of targets, for each of the above *Gaia* subsets, that fail all the redshift criteria ($z_{failure}$), and individually, that were defined in Section 5.5.

The REDROCK spectral classification* (i.e. GALAXY, STAR, or QSO) for each of the BGS subclass defined in Section 5.3 is shown in Table 5.4. The numbers represent percentages out of total of BGS SV targets ignoring the low-quality subclass. Overall, the stellar contamination, is at 3.2 per cent. For BGS BRIGHT, this is of the order of 4.3 per cent, higher than the tolerated maximum specified of 2 per cent. Table 5.5 shows the same as Table 5.4 but ignoring targets from the IG_{sv} subset. Applying this change, we are able to reduce the amount of stars by a third reaching 1.9 per cent of stars overall. BGS BRIGHT is also reduced to 2.7 per cent but still higher than the 2 per cent margin. Fig. 5.9 shows a comparison per REDROCK spectral classification of the targets in BGS SV with and without the IG_{sv} as a function of the r-band magnitude. Avoiding the IG_{sv} targets offers a clear improvement in our galaxy sample regarding stellar contamination, particularly at the bright end.

*Spectra with GALAXY best spectral fit but with $z < 200$ km s^{-1} are classified as STARS.

Table 5.4: Spectral classification from REDROCK for four out of the five BGS SV subclasses (ignoring the low-quality subclass). The numbers are percentages of the total number of targets across all four of these BGS SV target classes.

class	QSO	GALAXY	STAR
BRIGHT	0.6	38.4	1.7
FAINT	0.4	28.7	0.6
FAINT EXT	0.2	20.1	0.1
FIBMAG EXT	0.2	8.3	0.8
OVERALL	1.3	95.5	3.2

Table 5.5: As Table 5.4, but now excluding the IG_{sv} subset as well as the low-quality subclass.

class	QSO	GALAXY	STAR
BRIGHT	0.5	39.0	1.1
FAINT	0.3	29.1	0.3
FAINT EXT	0.2	20.4	0.1
FIBMAG EXT	0.2	8.4	0.4
OVERALL	1.2	96.9	1.9

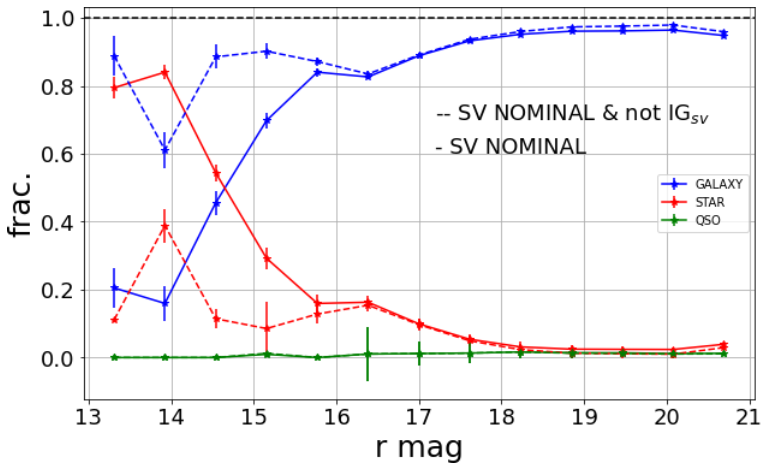


Figure 5.9: Comparison per REDROCK spectral classification (i.e. GALAXY, STAR, or QSO) of the targets in BGS SV with (solid) and without (dashed) the IG_{sv} as a function of the r-band magnitude.

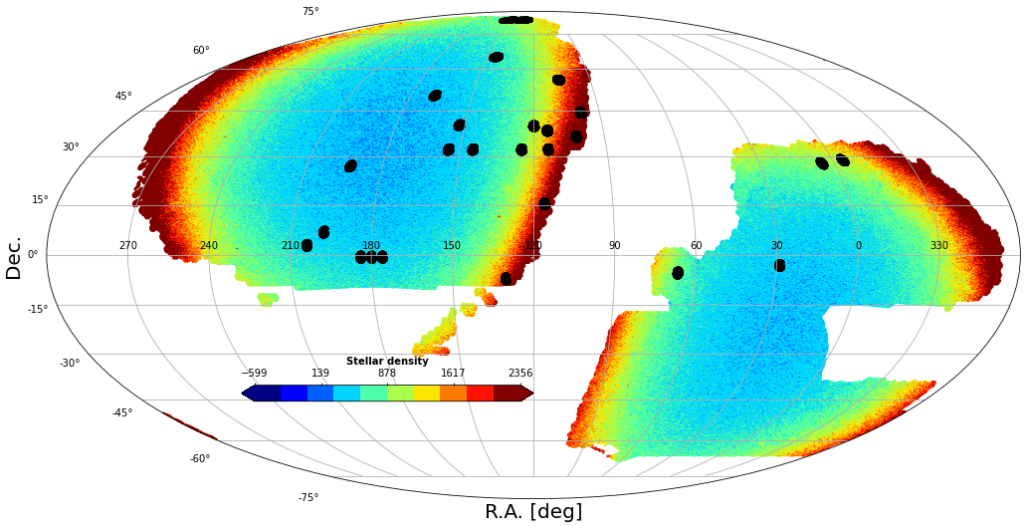


Figure 5.10: Stellar density map in the DESI footprint. The black circles show the position of the DEEP tiles used in this analysis.

To check whether our target footprint is representative to assess the stellar contamination, we compute the stellar density ratio $\bar{\eta}_{\rm DEEP}/\bar{\eta}_{\rm DESI}$ where $\bar{\eta}_{\rm DEEP}$ is the mean stellar density in the DEEP tiles and $\bar{\eta}_{\rm DESI}$ is the mean stellar density in the approximately $20,000$ deg^2 of the LS imaging footprint. Fig. 5.10 shows the stellar density map in the DESI footprint, with the black circles showing the position of the DEEP tiles used in this analysis. The stellar density is defined by *Gaia* stars with $12 < G < 17$ and with AEN $= 0$ OR $\log_{10}(\text{AEN}) < 0.3(G - 5.3)$. Fig. 5.11 shows the distribution of the stellar density in logarithm scale for $\eta_{\rm DEEP}$ and $\eta_{\rm DESI}$ with mean densities of $\bar{\eta}_{\rm DEEP} = 813$ objects/deg^2, and $\bar{\eta}_{\rm DESI} = 824$ objects/deg^2. The density ratio $\bar{\eta}_{\rm DEEP}/\bar{\eta}_{\rm DESI} = 1.12$ shows that the footprint used for this analysis has 12 per cent more stars than if we compare with the mean stellar density of the larger footprint of the imaging LS. If applying this correction, the stellar contamination gets closer to the 2 per cent margin.

5.8 Quality cuts assessment

Currently, we have two different samples of quality cuts: i) the so called old FRACS defined in Eqn. 2.6, and ii) the new FRACS defined in Eqn. 3.5 which is a subset

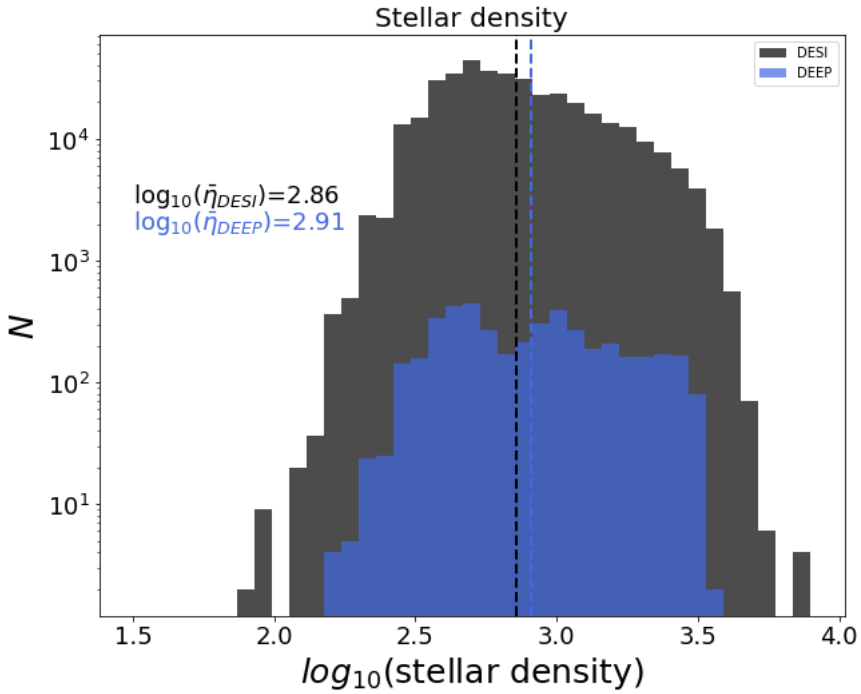


Figure 5.11: The distribution of the stellar density, on a logarithmic scale. The blue histogram shows the stellar density for the footprint within the DEEP exposure tiles while black histogram corresponds to a larger footprint of nearly $20,000$ deg^2. The blue and black dashed lines show the mean density of the same colour histogram.

Table 5.6: Spectral classification from REDROCK split into two samples: old FRACS and new FRACS. The sample new FRACS is a subset of the old FRACS sample. The numbers are percentages out of the total BGS SV targets.

class	QSO	GALAXY	STAR
old FRACS	0.1	3.0	1.6
new FRACS	0.1	1.6	1.4

of the old FRACS cuts. In SV, BGS does not apply any of these cuts. On the contrary, in BGS we have a subclass dedicated to target a fraction of these FRACS objects. Out of all BGS SV targets, old FRACS accounts for 4.7 per cent while new FRACS accounts for 3.1 per cent. Table 5.6 shows the REDROCK spectra classification results for the two sets of cuts. While the number of stars correctly rejected by new FRACS (1.4 per cent) is only slightly less than with old FRACS (1.6 per cent), the number of galaxies wrongly rejected by new FRACS (1.6 per cent) is almost reduced by a factor of two compared to old FRACS (3.0 per cent).

5.9 The BGS target selection for the main survey

The selection presented in this section corresponds to the final BGS selection already being used for the main survey programme. The selection was tuned following the above assessments (redshift success rate, star-galaxy classification, and quality cuts), but also the implementation of a different selection in BASS/MzLS compared to DECaLS following the r-band magnitude offset of $\Delta r_{\rm offset} = 0.04$ between BASS and DECaLS with the aim to match target densities between the surveys* (see Sec. 3). Eqn. 5.8 shows the equivalent colour-based selection given in Eqn 5.6 but tuned for BASS/MzLS $\Delta r_{\rm offset}$.

r-band magnitude limits at the faint end are imposed to restrict our BGS sample to a density of $1,400$ objects/deg^2. At the bright end we apply limits to the r-band magnitude and r-band total fibre magnitude (rfibtotmag), which includes contributions to the fibre magnitude from nearby overlapping sources, to avoid objects close to very bright objects being retained such objects close to saturated bright stars.

The **common cuts** are defined as follow:

- **Star-Galaxy separation**: based on *Gaia* DR2, a galaxy in BGS is defined by $(G - rr > 0.6)$ `or` $(G = 0)$ where G is the *Gaia* G-mag and rr is the LS r-band magnitude (without any extinction correction).
- **Spatial masking**: this includes geometrical masking around i) bright stars (BS) and ii) globular clusters (GC) and iii) a pixel masking of at least one observation in each of the g, r and z bands ($\rm NOBS_i > 0$). The geometric masks require that LS MASKBITS 1 and 13 are not set.
- **Photometric cuts**: these are i) colour-colour cuts in $g - r$ and $r - z$ (see equations 3.2 and 3.3), ii) cuts in fibre magnitude (see equation 3.4) and, iii) bright limit cuts $r > 12$ and rfibtotmag > 15.

*The offset applies also to rfibmag.

$$\Delta(g,r,z,w1)^{\rm BASS} = (z-w1) - 3/2.5(g-(r-\Delta r_{\rm offset})) + 1.2 \quad (5.7)$$

$$Z_{\rm high}^{\rm BASS} = \begin{cases} {\rm rfibmag} < 20.75 + \Delta r_{\rm offset} & \Delta(g,r,z,w1)^{\rm BASS} \leq 0 \\ {\rm rfibmag} < 21.5 + \Delta r_{\rm offset} & \Delta(g,r,z,w1)^{\rm BASS} > 0 \end{cases} \quad (5.8)$$

The BGS BRIGHT and BGS FAINT selection for DECaLS is:

- **BGS BRIGHT**: ($r < 19.5$) & (common cuts),
- **BGS FAINT**: ($19.5 < r < 20.175$) & ($Z_{\rm high}$) & (common cuts),

and for BASS/MzLS:

- **BGS BRIGHT**: ($r < 19.5 + \Delta r_{\rm offset}$) & (common cuts),
- **BGS FAINT**: ($19.5 + \Delta r_{\rm offset} < r < 20.22$) & ($Z_{\rm high}^{\rm BASS}$) & (common cuts).

The target density for BGS BRIGHT is $\sim$ 865 objects/deg^2 for DECaLS as well as for BASS/MzLS, and $\sim$ 535 objects/deg^2 for BGS FAINT for both regions too.

5.10 Preliminary conclusions

From Sec. 5.5 (and in particular from Fig. 5.4) we conclude that the bright sample ($r < 19.5$) in BGS SV already meets the requirements in terms of the redshift success rate by achieving a 97 per cent completeness under the nominal $T_{\rm bright} \sim$ 180 seconds. The faint sample ($r > 19.5$) on the other hand, does not have specific requirements for the redshift success rate. However, BGS SV proves that we can improve the redshift success rate if we adjust the selection implementing the colour-based cut in Eqn. 5.6 and by going deeper in r-band magnitude. The colour-based adjustment improves the redshift success rate from 92 to 97 per cent.

The requirements for stellar contamination state a margin of 2 per cent. Sec. 5.7 shows that we are close to meeting this requirement if we avoid targets that are in *Gaia* and have $G - rr < 0.6$, and even closer (~ 2.3 per cent) if we correct for the stellar density bias presented in the chosen tiles.

In Section 5.8 the assessment of the quality cuts (i.e. FRACS) shows that the stellar contamination in the old FRACS and in new FRACS is 34 and 45 per cent respectively. Our requirements in Sec. 5.2 allow a maximum of 20 per cent of stellar contamination in order to keep these targets. We believe that the old FRACS sample is not that far from the 20 per cent margin, and therefore, for the final BGS selection we have decided to not to apply any of the FRACS cuts. The additional galaxies might be relevant for clustering analysis and we can always get rid of the stars at a later stage.

Chapter 6

Conclusions and future work

6.1 Target selection and clustering analysis

The focus of this thesis is to build the cleanest and most reliable input Bright Galaxy Survey (BGS) catalogue for the Dark Energy Spectroscopic Instrument (DESI DESI Collaboration et al., 2016). Since no other bright galaxy catalogue of the specifications of the BGS has been built up before, our catalogue represents the most complete sample of bright galaxies to date and can be of great use for future spectroscopic surveys, and studies of large-scale structure and galaxy formation.

One notable feature of the BGS catalogue is the new method applied to perform star-galaxy classification, described in Sec. 2.4.1, that uses the *Gaia* DR2 objects, and which has already proven, by SV, to be efficient (see Sec. 5.7). After this classification, the BGS target catalogue is almost fully defined by a pure magnitude limit. In the BASS/MzLS and DECaLS regions only around 5 and 2 per cent respectively of the objects not classified as stars are rejected by other selection cuts (see Sec. 5.9). Of these, the majority (around 60 per cent) are rejected by the spatial masks, which, as they are also applied to the randoms, have no effect on clustering measurements.

The imaging systematics that might affect the reliability of our target catalogue are minimal, as reported in Sec. 3.6, with the exceptions of the stellar density and

galactic extinction, for which our target density varies by somewhat more than 5 per cent when plotted against these quantities. This is below the level seen in other large scale structure catalogues where its effect has been successfully mitigated by the use of compensating weights (Ross et al., 2020; Merz et al., 2021).

The BGS target catalogue has undergone two main quality reviews prior to the main assessment with the DESI survey validation programme. These are listed below:

- **Completeness with GAMA:** GAMA is similar in redshift and depth to BGS but covers a much smaller solid angle (Driver et al., 2012; Liske et al., 2015; Baldry et al., 2017). Our BGS target catalogue has a completeness above 99 per cent with respect to GAMA spectroscopically confirmed galaxies and most of these are BGS BRIGHT galaxies. There are around 400 BGS objects/deg^2 that are not in GAMA that has already been assessed in SV and confirmed spectroscopically as galaxies, showing that our BGS sample goes deeper than GAMA.
- **Cosmology and clustering with MXXL:** by comparing our BGS catalogue with the MXXL light-cone mock catalogue (Smith et al., 2017) through the two-point angular clustering, we were able to check that the BGS objects agree with a ΛCDM universe populated with galaxies that follow a standard HOD model. The clustering measurements show consistency between the DECaLS NGC, DECaLS SGC, and BASS/MzLS regions.

6.2 Final assessment of BGS with Survey Validation

We made use of the DESI Survey Validation data to tune and assess four key aspects of the BGS target catalogue that are essential for the DESI science requirements:

- **redshift success rate:** BGS BRIGHT shows a high redshit success rate of 97 per cent, 2 per cent higher than the requirement. BGS FAINT was tuned

with a colour-based selection cut that increased the redshift success rate to 97 per cent.

- **Exposure time:** The high completeness in redshift success rate was achieved with $T_{\text{eff\_bgs}}$ close to the nominal value of 180 ± 20 seconds.

- **Star-galaxy separation:** the stellar contamination in our BGS catalogue is slightly higher than the stated two per cent margin. We are, however, very close to the margin and have decided to keep the method as it is for the final catalogue.

- **Quality cuts:** usually referred as OLD FRACS for LS DR8 or NEW FRACS for LS DR9 (see equations 2.6 and 3.5 respectively). Either of these samples has more than 20 per cent stellar contamination and so is above our stated tolerance. However, this sample is small, of the order of 1 per cent of the total BGS smaple, and therefore including this sample does not significantly increase the overall stellar contamination. Henceforth, we prefer to not exclude this small sample, as a precise clustering analysis requires a matching mask in the randoms as well which we would not have for a sample with these cuts applied.

6.3 Future Work

There are still some ongoing issues in BGS like the systematic trend in galaxy density at high stellar densities, that could be improved at the bright end by using additional cuts in the r-band magnitude and the r-band fibre magnitude plane. For the faint end, stacking around faint stars reveals systematic density variations that could be addressed with additional masking. Another issue is the selection of spurious sources in the vicinity of large galaxies, here, the imaging visual inspection web tool (see Sec. 3.3.1 and Appendix B) is useful and may enable some combination of image properties to be combined to flag the spurious sources, but

if not we will have to use the spectra to determine whether the observed sources are spurious, fragments of the large galaxy or genuine distinct galaxies.

Despite these ongoing issues, the BGS target selection catalogue for DESI is a powerful dataset in its own right for many studies. Here, we outline some of the ways this work can be used.

- BGS can be used to create a photo-z catalogue at low redshifts. The bands available in the LS span a wide range of wavelengths, covering g, r, z bands in the optical, and four bands in the infrared from the ALLWISE catalogue (Secrest et al., 2015).

- The high density and footprint of BGS allows an unprecedented sample for a deeper clustering analysis, probing galaxy clustering as a function of luminosity, colour and redshift. This can be used to constrain HOD models and physical models of galaxy formation, and to build better mock catalogues.

- BGS can be used to create synergies with future deep and wide area surveys like the LSST. A cross-match with the LSST can only be carried out with DESI as it is deep enough in redshift, and has a large enough overlap ($\approx 6,000$ deg^2). The DESI $z < 1$ galaxy sample (that includes BGS) will probe the halo mass range where the impact of baryonic effects on cosmic shear signals is expected to be the largest ($M_{\rm halo} \sim 10^{14}$ M$_\odot$). DESI will provide accurate spectroscopic redshifts for groups and clusters in this range, thus minimizing the impact of projection effects that impact photometric cluster selections. At $z < 0.3$, the flux-limited BGS survey will ensure that a wide variety of host galaxy types are included. An increase in SN Ia host statistics and a larger solid angle of the combined LSST+DESI survey will also improve tests of gravity and ΛCDM using peculiar-velocity measurements of SNe Ia (e.g., Huterer et al., 2017).

Appendix A

Galaxy view

Flow charts with a galaxy view (i.e., ignoring stars) of the different BGS target selection process for the different LS releases, DR8, and DR9.

A.1 DECaLS DR8

In contrast to the approach taken in Sections 2.3 and 2.4, here we present a 'galaxy' view of the BGS selection by implementing the star-galaxy separation before the other BGS cuts (with the exception of first applying the nominal BGS magnitude limit $r < 20$). The results of this exercise are shown in Fig. A.1. In this view, the geometric masking does not look as aggressive as it did in Fig. 2.2, with the size of the rejected area and number of objects typically reduced at each step by an order of magnitude compared to what was seen in Fig. 2.2. The BS mask step is the stage that is the most affected by this change in order. Next is the application of the selection on NOBS which has half the effect that it did in Fig. 2.2. Note that the area removed by the cuts remains unchanged as this does not depend on the number of targets but is calculated using the randoms.

In addition to changing the order in which the star-galaxy separation is applied compared to the selection criteria presented in Sections 2.3 and 2.4, we swap the order of the FMC and CC with the QCs. When comparing both schemes, (Fig. 2.4

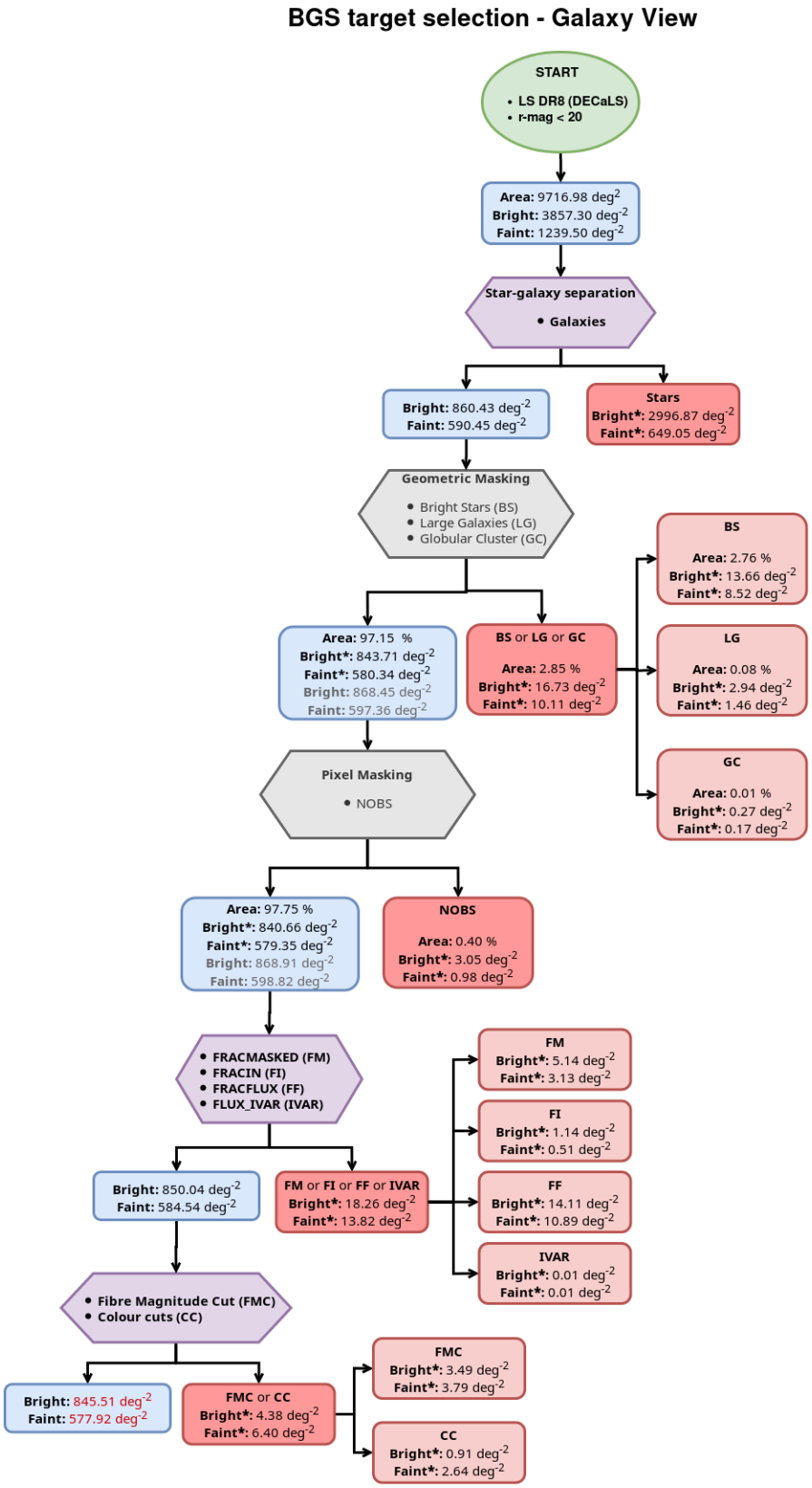


Figure A.1: Flow chart showing the spatial and photometric BGS target selections applied to the Legacy Surveys DR8 (continued on the following page).

Figure A.1: The spatial selections are shown by gray boxes and are divided into two kinds, one defined by geometric cuts around bright sources i.e. bright stars (BS), large galaxies (LG) and globular clusters (GC), and the other which is at the pixel level, such as the number of observations (NOBS). The photometric selection of BGS targets is divided into four types and is shown by purple boxes; star-galaxy separation, fibre magnitude cuts (FMC), colour cuts (CC) and quality cuts (QCs) which include FRACMASKED, FRACIN, FRACFLUX and FLUX_IVAR. The blue boxes show the area (in degrees) and the number density (per square degree) of objects retained after each selection, broken down into the numbers for the bright and faint components of the BGS. The red boxes show the equivalent information for the rejected objects. If more than one cut or selection is applied at a given stage, then the darker red boxes show the information about removed objects for the combination of cuts and the lighter red boxes show the corresponding values for each individual cut. The superscript (*) denotes target densities without correcting for the area removed by cuts up to that point, while densities without a superscript (*) do take into account the reduction in area.

and Fig. A.1), we see a high overlap between the QCs and the FMC of $\sim$ 15 objects/deg^2 which represent $\sim 2/3$ the galaxies rejected by FMC in Section 2.4. CC is also affected by the to the sequence of cuts and the rejections due to this cut are reduced by a factor of 2 in the galaxy view.

A.2 DECaLS and BASS/MzLS DR9

Here we present a 'galaxy' view of the BGS selection presented in Sec. 3.2 where we implement the star-galaxy separation before the other BGS cuts (with the exception of first applying the nominal BGS magnitude limit $r < 20$). The results of this exercise are shown in Fig. A.2 for DECaLS, and in Fig. A.3 for the BASS/MzLS regions.

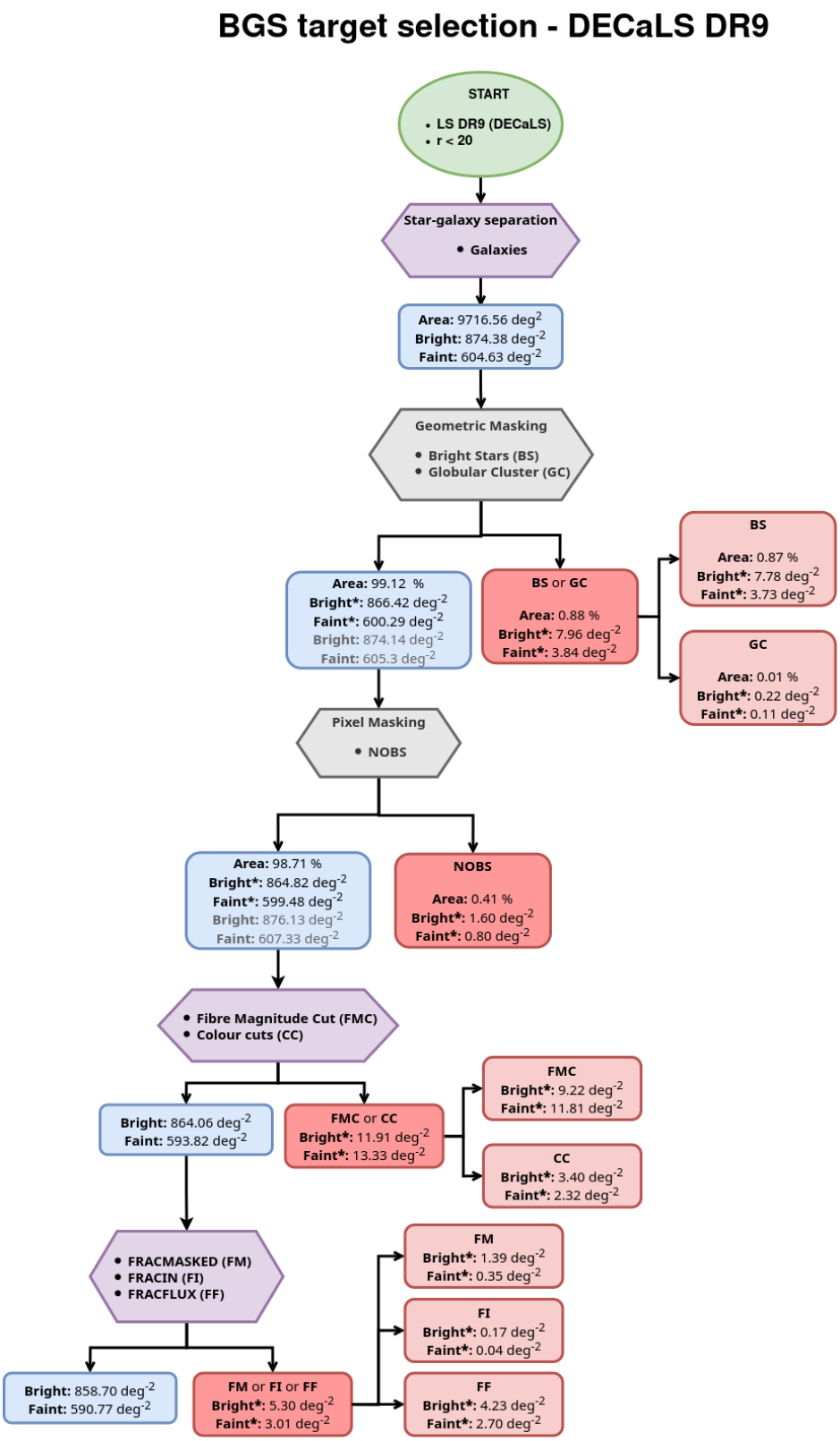


Figure A.2: Flow chart showing the spatial and photometric BGS target selections given by Sec. 3.2 and applied to the DECaLS region from the Legacy Surveys DR9 (continued on the following page).

Figure A.2: The spatial selections are shown by gray boxes and are divided into two kinds, one defined by geometric cuts around bright sources i.e. bright stars (BS) and the other which is at the pixel level, such as the number of observations (NOBS). The photometric selection are shown by purple boxes. The blue boxes show the area (in degrees) and the number density (per square degree) of objects retained after each selection, broken down into the numbers for the bright and faint components of the BGS. The red boxes show the equivalent information for the rejected objects. If more than one cut or selection is applied at a given stage, then the darker red boxes show the information about removed objects for the combination of cuts and the lighter red boxes show the corresponding values for each individual cut. The superscript (*) denotes target densities without correcting for the area removed by cuts up to that point, while densities without a superscript (*) do take into account the reduction in area.

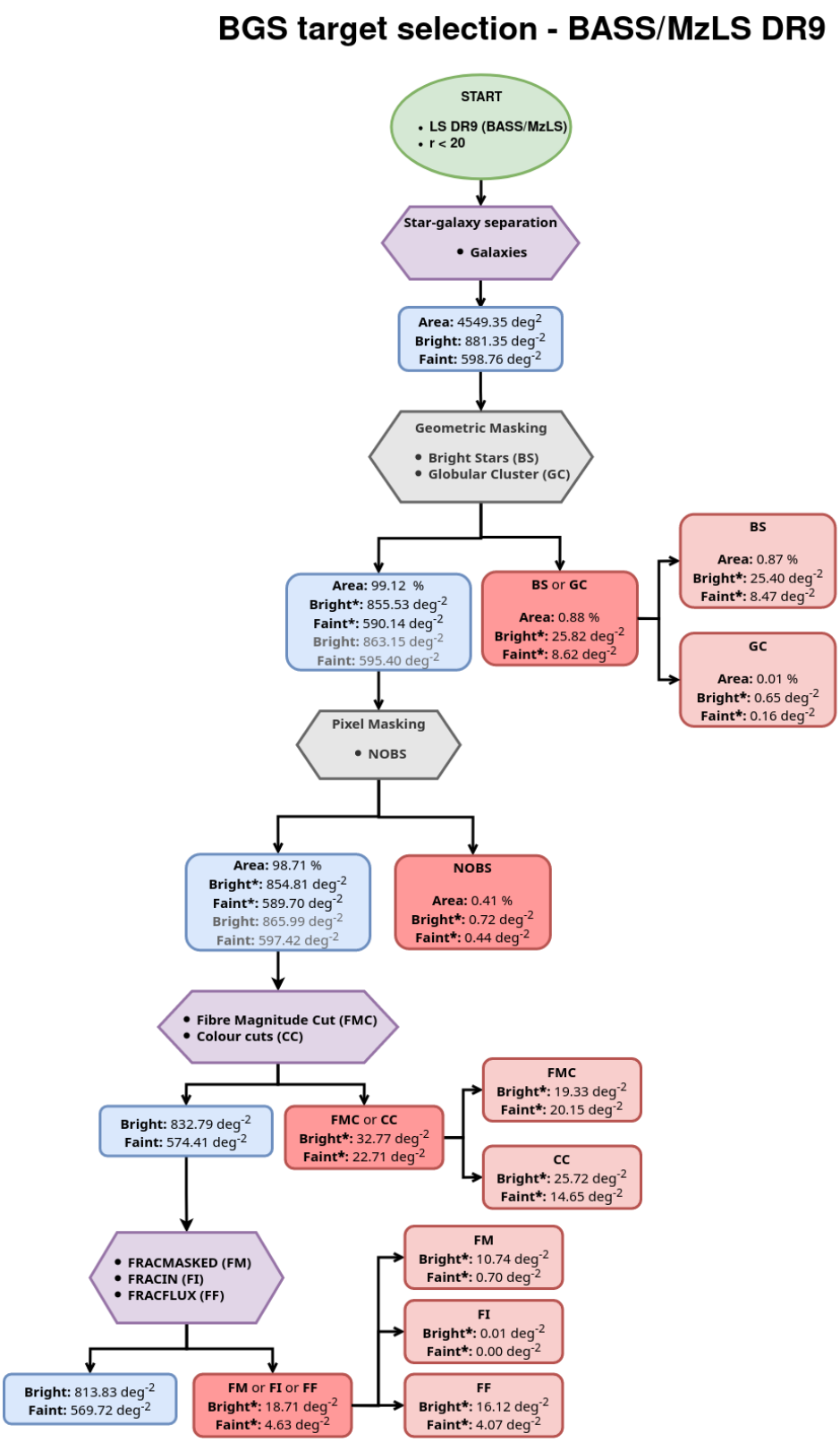


Figure A.3: As for Fig. A.2, but for the BASS/MzLS region of the Legacy Surveys.

Appendix B

LSVI webtool galleries

With the aim of validating the BGS target selection, galleries of postage stamp images were constructed to test the purity and contamination of a variety of selections. Here we present a couple of pages from one such gallery that was set up to for the Visual Inspection described in Sec. 3.3.1.

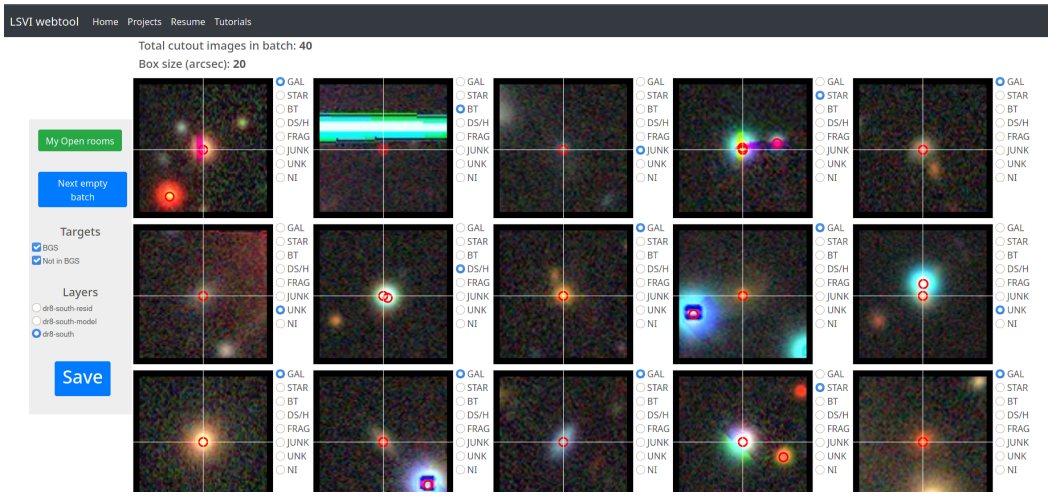


Figure B.1: Example of the LSVI webtool for the visual inspection described in Sec. 3.3.1. Each postage stamp includes radio buttons in the right-hand side for the classification of the object. Page control buttons control the type of image displayed (i.e. image, model, or residuals) and target overlays (i.e. green if it is a BGS target or red if not) are placed in left-hand side panel of the page.

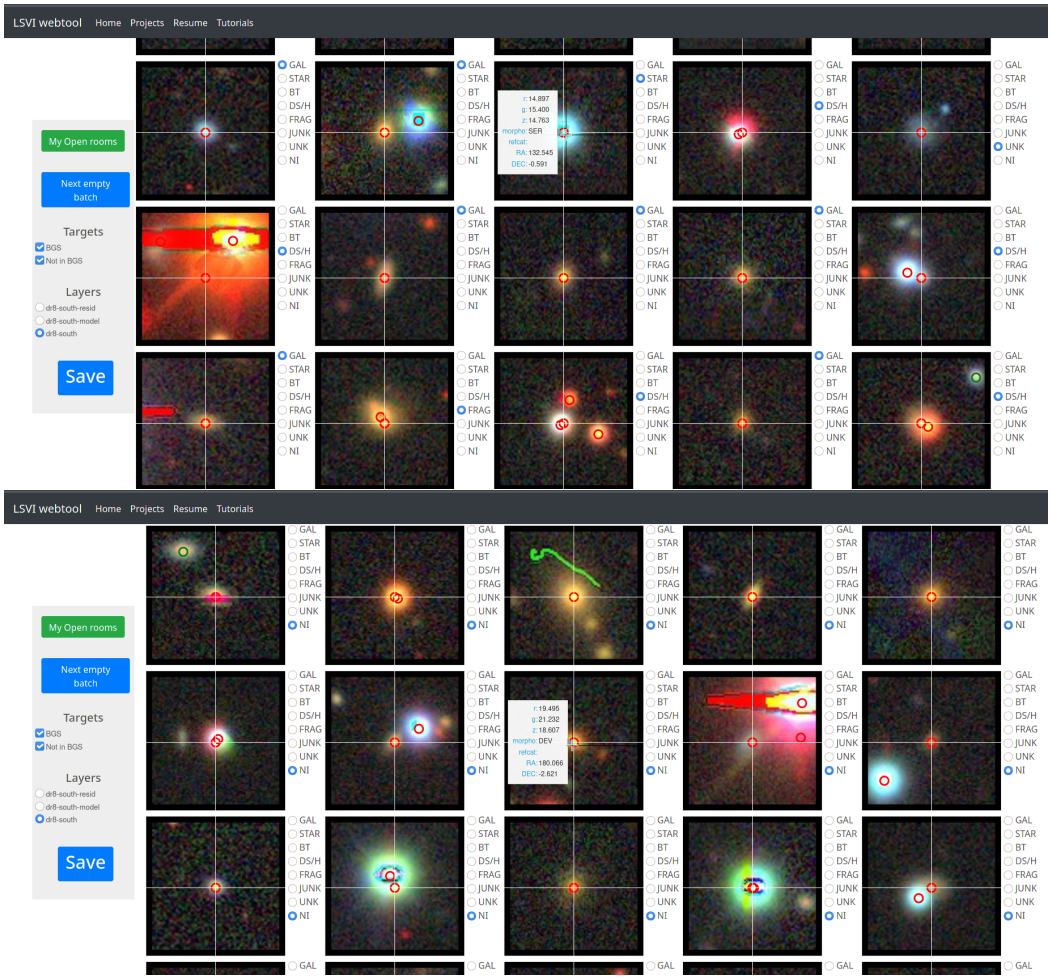


Figure B.2: This shows examples with the kind of description that can be placed for all the objects appearing in the postages images. This example shows the magnitude in the g, r and z bands, the best-fit morphology by TRACTOR, and the positions in RA and DEC.

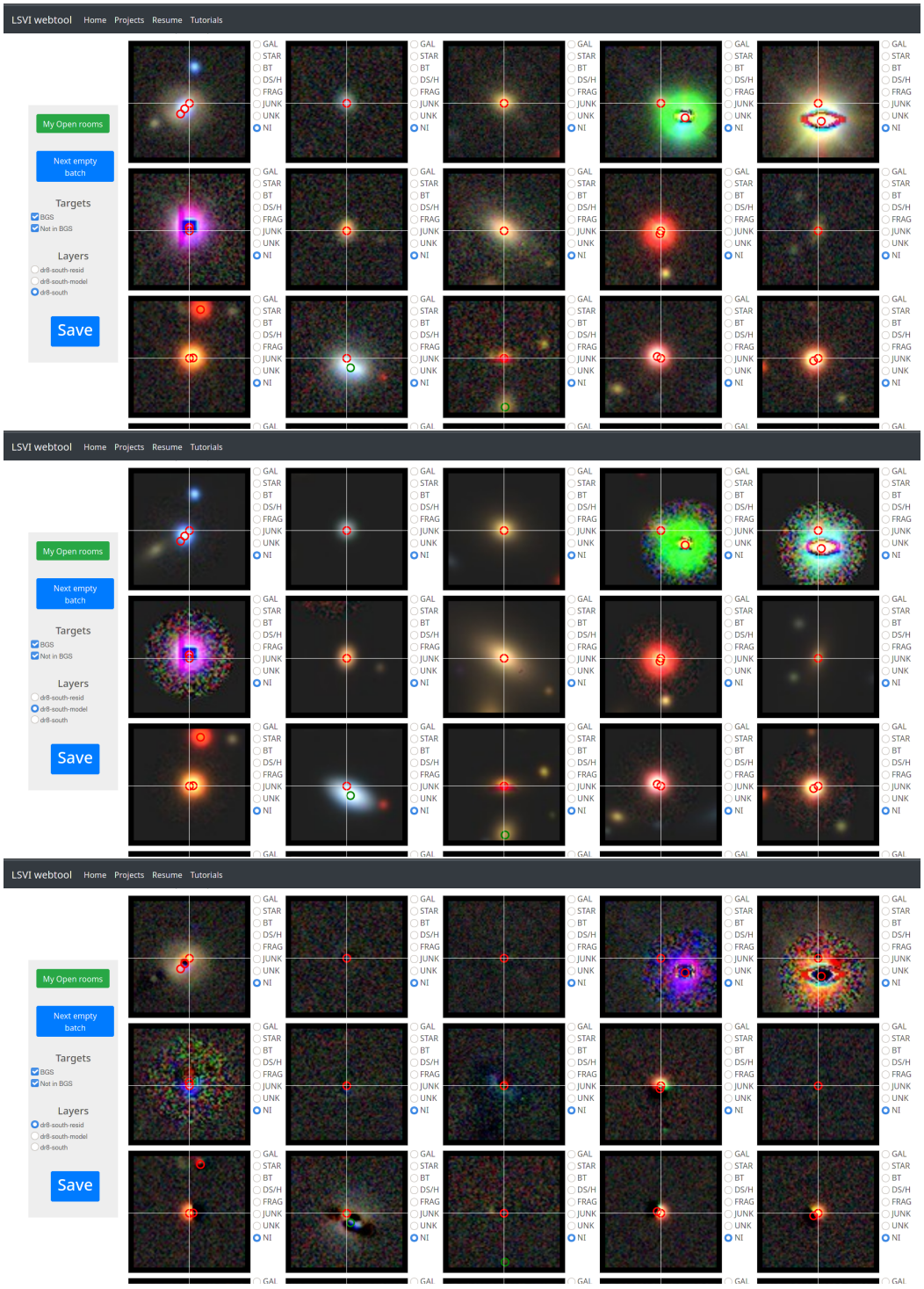


Figure B.3: Examples of the choice of image displayed. **top:** image, **middle:** model, and **bottom:** residual.

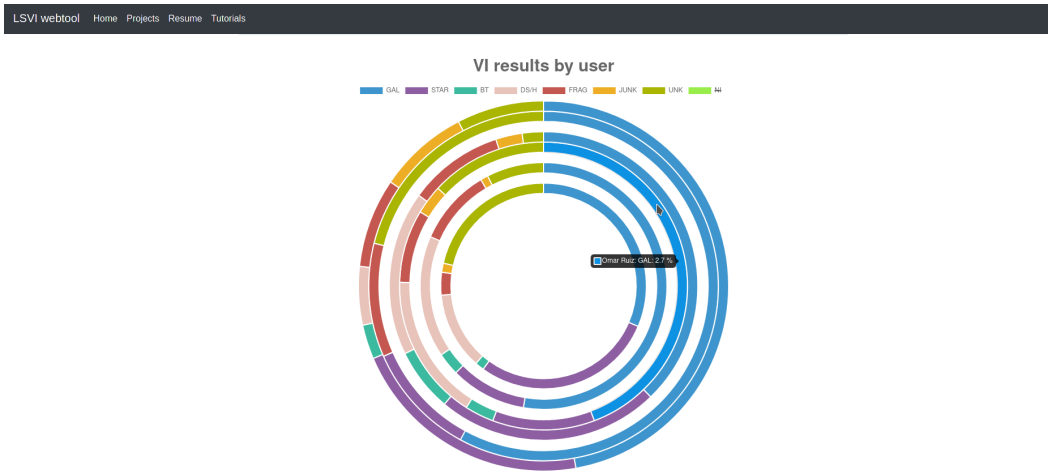


Figure B.4: Example of classification results. Each circle correspond to results from each user involved in the classification where colours represent the classification labels.

Bibliography

K. Abazajian et al. The First data release of the Sloan Digital Sky Survey. *Astron. J.*, 126:2081, 2003. doi: 10.1086/378165.

K. N. Abazajian, J. K. Adelman-McCarthy, M. A. Agüeros, et al. The Seventh Data Release of the Sloan Digital Sky Survey. *Astrophysical Journal, Supplement*, 182 (2):543–558, June 2009. doi: 10.1088/0067-0049/182/2/543.

J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, et al. The Fourth Data Release of the Sloan Digital Sky Survey. *Astrophysical Journal, Supplement*, 162 (1):38–48, Jan. 2006. doi: 10.1086/497917.

J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, et al. The Sixth Data Release of the Sloan Digital Sky Survey. *Astrophysical Journal, Supplement*, 175 (2):297–313, April 2008. doi: 10.1086/524984.

C. P. Ahn, R. Alexandroff, C. Allende Prieto, et al. The Tenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Apache Point Observatory Galactic Evolution Experiment. *Astrophysical Journal, Supplement*, 211(2):17, April 2014. doi: 10.1088/0067-0049/211/2/17.

H. Aihara, R. Armstrong, S. Bickerton, et al. First data release of the Hyper Suprime-Cam Subaru Strategic Program. *Publications of the ASJ*, 70:S8, January 2018. doi: 10.1093/pasj/psx081.

S. Alam, F. D. Albareti, C. Allende Prieto, et al. The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III. *Astrophysical Journal, Supplement*, 219(1):12, July 2015. doi: 10.1088/0067-0049/219/1/12.

D. Alonso. CUTE solutions for two-point correlation functions from large cosmological datasets. *ArXiv e-prints*, October 2012.

R. E. Angulo, V. Springel, S. D. M. White, et al. Scaling relations for galaxy clusters in the Millennium-XXL simulation. *Monthly Notices of the RAS*, 426 (3):2046–2062, Nov. 2012a. doi: 10.1111/j.1365-2966.2012.21830.x.

R. E. Angulo, V. Springel, S. D. M. White, et al. Scaling relations for galaxy clusters in the Millennium-XXL simulation. *Monthly Notices of the RAS*, 426 (3):2046–2062, Nov 2012b. doi: 10.1111/j.1365-2966.2012.21830.x.

S. Arnouts, L. Moscardini, E. Vanzella, et al. Measuring the redshift evolution of clustering: the Hubble Deep Field South. *Monthly Notices of the Royal Astronomical Society*, 329(2):355–366, 01 2002. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2002.04988.x. URL `https://doi.org/10.1046/j.1365-8711.2002.04988.x`.

J. N. Bahcall and R. Davis. Solar neutrinos: A scientific puzzle. *Science*, 191 (4224):264–267, 1976. ISSN 0036-8075. doi: 10.1126/science.191.4224.264. URL `https://science.sciencemag.org/content/191/4224/264`.

C. A. L. Bailer-Jones, M. Fouesneau, and R. Andrae. Quasar and galaxy classification in Gaia Data Release 2. *Monthly Notices of the RAS*, 490(4):5615–5633, December 2019. doi: 10.1093/mnras/stz2947.

I. K. Baldry, K. Glazebrook, T. Budavári, et al. The Sloan Digital Sky Survey u-band Galaxy Survey: luminosity functions and evolution. *Monthly Notices of the RAS*, 358(2):441–456, Apr. 2005. doi: 10.1111/j.1365-2966.2005.08799.x.

I. K. Baldry, A. S. G. Robotham, D. T. Hill, et al. Galaxy And Mass Assembly (GAMA): the input catalogue and star–galaxy separation. *Monthly Notices of the Royal Astronomical Society*, 404(1):86–100, 04 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.16282.x. URL `https://doi.org/10.1111/j.1365-2966.2010.16282.x`.

I. K. Baldry, J. Liske, M. J. I. Brown, et al. Galaxy And Mass Assembly: the G02 field, Herschel–ATLAS target selection and data release 3. *Monthly Notices of the Royal Astronomical Society*, 474(3):3875–3888, 11 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx3042. URL `https://doi.org/10.1093/mnras/stx3042`.

D. J. Bates, R. Tojeiro, J. A. Newman, et al. Mass functions, luminosity functions, and completeness measurements from clustering redshifts. *Monthly Notices of the RAS*, 486(3):3059–3077, July 2019. doi: 10.1093/mnras/stz997.

N. Benitez. Bayesian photometric redshift estimation. *The Astrophysical Journal*, 536(2):571–583, jun 2000. doi: 10.1086/308947. URL `https://doi.org/10.1086/308947`.

C. L. Bennett, M. Halpern, G. Hinshaw, et al. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results. *Astrophysical Journal, Supplement*, 148(1):1–27, September 2003. doi: 10.1086/377253.

A. J. Benson, S. Cole, C. S. Frenk, et al. The nature of galaxy bias and clustering. *Monthly Notices of the RAS*, 311(4):793–808, February 2000. doi: 10.1046/j.1365-8711.2000.03101.x.

F. Bernardeau, S. Colombi, E. Gazta

textasciitilde naga, et al. Large-scale structure of the Universe and cosmological perturbation theory. *Physics Reports*, 367(1-3):1–248, September 2002. doi: 10.1016/S0370-1573(02)00135-7.

E. Bertin and S. Arnouts. SExtractor: Software for source extraction. *Astronomy and Astrophysics, Supplement*, 117:393–404, June 1996. doi: 10.1051/aas:1996164.

D. Bianchi and L. Verde. Confronting missing observations with probability weights: Fourier space and generalized formalism. *Monthly Notices of the RAS*, 495(1):1511–1529, June 2020. doi: 10.1093/mnras/staa1267.

D. Bianchi, A. Burden, W. J. Percival, et al. Unbiased clustering estimates with the DESI fibre assignment. *Monthly Notices of the RAS*, 481(2):2338–2348, December 2018. doi: 10.1093/mnras/sty2377.

M. R. Blanton, J. Dalcanton, D. Eisenstein, et al. The Luminosity Function of Galaxies in SDSS Commissioning Data. *Astronomical Journal*, 121(5):2358–2380, May 2001. doi: 10.1086/320405.

A. S. Bolton, D. J. Schlegel, É. Aubourg, et al. Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey. *Astronomical Journal*, 144(5):144, Nov. 2012. doi: 10.1088/0004-6256/144/5/144.

J. R. Bond and G. Efstathiou. Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter. *Astrophysical Journal, Letters*, 285:L45–L48, October 1984. doi: 10.1086/184362.

A. Burden, N. Padmanabhan, R. N. Cahn, et al. Mitigating the impact of the DESI fiber assignment on galaxy clustering. *Journal of Cosmology and Astroparticle Physics*, 2017(3):001, March 2017. doi: 10.1088/1475-7516/2017/03/001.

R. Cannon, M. Drinkwater, A. Edge, et al. The 2dF-SDSS LRG and QSO (2SLAQ) Luminous Red Galaxy Survey. *Monthly Notices of the RAS*, 372(1):425–442, Oct. 2006. doi: 10.1111/j.1365-2966.2006.10875.x.

J. M. Carrasco, D. W. Evans, P. Montegriffo, et al. Gaia Data Release 1. Principles of the photometric calibration of the G band. *Astronomy and Astrophysics*, 595: A7, Nov 2016. doi: 10.1051/0004-6361/201629235.

K. C. Chambers, E. A. Magnier, N. Metcalfe, et al. The Pan-STARRS1 Surveys. *arXiv e-prints*, art. arXiv:1612.05560, Dec 2016.

D. Coe, N. Benítez, S. F. Sánchez, et al. Galaxies in the hubble ultra deep field. i. detection, multiband photometry, photometric redshifts, and morphology. *The Astronomical Journal*, 132(2):926–959, jul 2006. doi: 10.1086/505530. URL `https://doi.org/10.1086/505530`.

A. L. Coil, M. R. Blanton, S. M. Burles, et al. The PRIsm MUlti-object Survey (PRIMUS). I. Survey Overview and Characteristics. *Astrophysical Journal*, 741 (1):8, November 2011. doi: 10.1088/0004-637X/741/1/8.

S. Cole, P. Norberg, C. M. Baugh, et al. The 2dF galaxy redshift survey: near-infrared galaxy luminosity functions. *Monthly Notices of the RAS*, 326(1):255–273, September 2001. doi: 10.1046/j.1365-8711.2001.04591.x.

S. Cole, W. J. Percival, J. A. Peacock, et al. The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *Monthly Notices of the RAS*, 362(2):505–534, September 2005. doi: 10.1111/j.1365-2966.2005.09318.x.

M. Colless, G. Dalton, S. Maddox, et al. The 2dF Galaxy Redshift Survey: spectra and redshifts. *Monthly Notices of the RAS*, 328(4):1039–1063, December 2001. doi: 10.1046/j.1365-8711.2001.04902.x.

M. Colless, B. A. Peterson, C. Jackson, et al. The 2dF Galaxy Redshift Survey: Final Data Release. *arXiv e-prints*, art. astro-ph/0306581, June 2003.

C. Conroy. Modeling the Panchromatic Spectral Energy Distributions of Galaxies. *Annual Review of Astron and Astrophys*, 51(1):393–455, August 2013. doi: 10.1146/annurev-astro-082812-141017.

R. J. Cool, J. Moustakas, M. R. Blanton, et al. The PRIsm MUlti-object Survey (PRIMUS). II. Data Reduction and Redshift Fitting. *Astrophysical Journal*, 767 (2):118, April 2013. doi: 10.1088/0004-637X/767/2/118.

A. Cooray and R. Sheth. Halo models of large scale structure. *Physics Reports*, 372:1–129, December 2002. doi: 10.1016/S0370-1573(02)00276-4.

M. Crocce et al. Dark Energy Survey Year 1 Results: Galaxy Sample for BAO Measurement. *Submitted to: Mon. Not. Roy. Astron. Soc.*, 2017.

I. Csabai, L. Dobos, M. Trencséni, et al. Multidimensional indexing tools for the virtual observatory. *Astronomische Nachrichten*, 328(8):852, October 2007. doi: 10.1002/asna.200710817.

C. E. Cunha, D. Huterer, H. Lin, et al. Spectroscopic failures in photometric redshift calibration: cosmological biases and survey requirements. *Monthly Notices of the RAS*, 444(1):129–146, October 2014. doi: 10.1093/mnras/stu1424.

L. N. da Costa, C. N. A. Willmer, P. S. Pellegrini, et al. The Southern Sky Redshift Survey. *Astronomical Journal*, 116(1):1–7, July 1998. doi: 10.1086/300410.

M. Davis and P. J. E. Peebles. A survey of galaxy redshifts. V. The two-point position and velocity correlations. *Astrophysical Journal*, 267:465–482, April 1983. doi: 10.1086/160884.

M. Davis, S. M. Faber, J. Newman, et al. Science Objectives and Early Results of the DEEP2 Redshift Survey. In P. Guhathakurta, editor, *Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II*, volume 4834 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 161–172, February 2003. doi: 10.1117/12.457897.

K. S. Dawson, D. J. Schlegel, C. P. Ahn, et al. The Baryon Oscillation Spectroscopic Survey of SDSS-III. *Astronomical Journal*, 145(1):10, January 2013. doi: 10.1088/0004-6256/145/1/10.

K. S. Dawson, J.-P. Kneib, W. J. Percival, et al. The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data. *Astronomical Journal*, 151(2):44, February 2016. doi: 10.3847/0004-6256/151/2/44.

P. de Bernardis, P. A. R. Ade, J. J. Bock, et al. A flat Universe from high-resolution maps of the cosmic microwave background radiation. *Nature*, 404(6781):955–959, April 2000. doi: 10.1038/35010035.

J. T. A. de Jong, G. A. Verdoes Kleijn, K. H. Kuijken, et al. The Kilo-Degree Survey. *Experimental Astronomy*, 35(1-2):25–44, January 2013. doi: 10.1007/s10686-012-9306-1.

DESI Collaboration, A. Aghamousa, J. Aguilar, et al. The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, art. arXiv:1611.00036, October 2016.

A. Dey, D. J. Schlegel, D. Lang, et al. Overview of the DESI legacy imaging surveys. *The Astronomical Journal*, 157(5):168, apr 2019. doi: 10.3847/1538-3881/ab089d. URL `https://doi.org/10.3847%2F1538-3881%2Fab089d`.

M. J. Drinkwater, R. J. Jurek, C. Blake, et al. The WiggleZ Dark Energy Survey: survey design and first data release. *Monthly Notices of the RAS*, 401(3):1429–1452, January 2010. doi: 10.1111/j.1365-2966.2009.15754.x.

S. P. Driver, D. T. Hill, L. S. Kelvin, et al. VizieR Online Data Catalog: Galaxy And Mass Assembly (GAMA) DR1 (Driver+, 2011). *VizieR Online Data Catalog*, art. J/MNRAS/413/971, September 2012.

G. Efstathiou. To H0 or not to H0? *arXiv e-prints*, art. arXiv:2103.08723, March 2021.

G. Efstathiou, S. Moody, J. A. Peacock, et al. Evidence for a non-zero Λ and a low matter density from a combined analysis of the 2dF Galaxy Redshift Survey and cosmic microwave background anisotropies. *Monthly Notices of the RAS*, 330(2):L29–L35, February 2002. doi: 10.1046/j.1365-8711.2002.05215.x.

D. J. Eisenstein and W. Hu. Baryonic Features in the Matter Transfer Function. *Astrophysical Journal*, 496(2):605–614, March 1998. doi: 10.1086/305424.

D. J. Eisenstein, J. Annis, and J. E. o. Gunn. Spectroscopic Target Selection for the Sloan Digital Sky Survey: The Luminous Red Galaxy Sample. *Astronomical Journal*, 122(5):2267–2280, Nov. 2001. doi: 10.1086/323717.

D. J. Eisenstein, D. H. Weinberg, E. Agol, et al. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *Astronomical Journal*, 142(3):72, September 2011. doi: 10.1088/0004-6256/142/3/72.

R. S. Ellis, M. Colless, T. Broadhurst, et al. Autofib Redshift Survey - I. Evolution of the galaxy luminosity function. *Monthly Notices of the RAS*, 280(1):235–251, May 1996. doi: 10.1093/mnras/280.1.235.

R. Fadely, D. W. Hogg, and B. Willman. Star-Galaxy Classification in Multi-band Optical Imaging. *Astrophysical Journal*, 760(1):15, November 2012. doi: 10.1088/0004-637X/760/1/15.

G. R. Farrar and M. E. Shaposhnikov. Baryon asymmetry of the Universe in the minimal standard model. *Physical Review Letters*, 70(19):2833–2836, May 1993. doi: 10.1103/PhysRevLett.70.2833.

B. Flaugher, H. T. Diehl, K. Honscheid, et al. The Dark Energy Camera. *Astronomical Journal*, 150(5):150, November 2015. doi: 10.1088/0004-6256/150/5/150.

J. N. Fry and E. Gaztanaga. Biasing and Hierarchical Statistics in Large-Scale Structure. *Astrophysical Journal*, 413:447, August 1993. doi: 10.1086/173015.

S. Fukuda, Y. Fukuda, M. Ishitsuka, et al. Constraints on neutrino oscillations using 1258 days of super-kamiokande solar neutrino data. *Phys. Rev. Lett.*, 86: 5656–5660, Jun 2001. doi: 10.1103/PhysRevLett.86.5656. URL `https://link.aps.org/doi/10.1103/PhysRevLett.86.5656`.

Gaia Collaboration, A. G. A. Brown, A. Vallenari, et al. Gaia Data Release 1. Summary of the astrometric, photometric, and survey properties. *Astronomy and Astrophysics*, 595:A2, Nov 2016a. doi: 10.1051/0004-6361/201629512.

Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, et al. The Gaia mission. *Astronomy and Astrophysics*, 595:A1, Nov 2016b. doi: 10.1051/0004-6361/201629272.

Gaia Collaboration, A. G. A. Brown, A. Vallenari, et al. Gaia Data Release 2. Summary of the contents and survey properties. *Astronomy and Astrophysics*, 616:A1, Aug 2018. doi: 10.1051/0004-6361/201833051.

B. Garilli, O. Le Fèvre, L. Guzzo, et al. The Vimos VLT deep survey. Global properties of 20,000 galaxies in the I_$AB < 22.5$ WIDE survey. *Astronomy and Astrophysics*, 486(3):683–695, August 2008. doi: 10.1051/0004-6361:20078878.

B. Garilli, L. Guzzo, M. Scodeggio, et al. The VIMOS Public Extragalactic Survey (VIPERS). First Data Release of 57 204 spectroscopic measurements. *Astronomy and Astrophysics*, 562:A23, February 2014. doi: 10.1051/0004-6361/201322790.

E. Gaztanaga. High-Order Galaxy Correlation Functions in the APM Galaxy Survey. *Monthly Notices of the RAS*, 268:913, June 1994. doi: 10.1093/mnras/268.4.913.

E. Gaztanaga. On the Interpretation of Clustering from the Angular APM Galaxy Survey. *Astrophysical Journal*, 454:561, December 1995. doi: 10.1086/176509.

M. J. Geller and J. P. Huchra. Mapping the universe. *Science*, 246(4932):897–903, 1989. ISSN 0036-8075. doi: 10.1126/science.246.4932.897. URL `https://science.sciencemag.org/content/246/4932/897`.

K. M. Górski, E. Hivon, A. J. Banday, et al. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *Astrophysical Journal*, 622(2):759–771, April 2005. doi: 10.1086/427976.

E. J. Groth and P. J. E. Peebles. Statistical analysis of catalogs of extragalactic objects. VII. Two- and three-point correlation functions for the high-resolution Shane-Wirtanen catalog of galaxies. *Astrophysical Journal*, 217:385–405, October 1977. doi: 10.1086/155588.

L. Guzzo, M. Scodeggio, B. Garilli, et al. The VIMOS Public Extragalactic Redshift Survey (VIPERS). An unprecedented view of galaxies and large-scale structure at $0.5 < z < 1.2$. *Astronomy and Astrophysics*, 566:A108, June 2014. doi: 10.1051/0004-6361/201321489.

C. Hahn, R. Scoccimarro, M. R. Blanton, et al. The Effect of Fiber Collisions on the Galaxy Power Spectrum Multipoles. *Monthly Notices of the RAS*, 467(2): 1940–1956, May 2017. doi: 10.1093/mnras/stx185.

N. W. Halverson, E. M. Leitch, C. Pryke, et al. Degree angular scale interferometer first results: A measurement of the cosmic microwave background angular power spectrum. *The Astrophysical Journal*, 568(1):38–45, mar 2002. doi: 10.1086/338879. URL `https://doi.org/10.1086/338879`.

S. Hanany, P. Ade, A. Balbi, et al. MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of 10'-5°. *Astrophysical Journal, Letters*, 545(1):L5–L9, December 2000. doi: 10.1086/317322.

E. Høg, C. Fabricius, V. V. Makarov, et al. The Tycho-2 catalogue of the 2.5 million brightest stars. *Astronomy and Astrophysics*, 355:L27–L30, March 2000.

J. A. Holtzman. Microwave Background Anisotropies and Large-Scale Structure in Universes with Cold Dark Matter, Baryons, Radiation, and Massive and Massless Neutrinos. *Astrophysical Journal, Supplement*, 71:1, September 1989. doi: 10.1086/191362.

W. Hu and S. Dodelson. Cosmic microwave background anisotropies. *Ann.Rev.Astron.Astrophys.*, 40:171–216, 2002. doi: 10.1146/annurev.astro.40.060401.093926.

W. Hu and N. Sugiyama. Small-Scale Cosmological Perturbations: an Analytic Approach. *Astrophysical Journal*, 471:542, November 1996. doi: 10.1086/177989.

J. S. Huang, K. Glazebrook, L. L. Cowie, and C. Tinney. The Hawaii+Anglo-Australian Observatory K-Band Galaxy Redshift Survey. I. The Local K-Band

Luminosity Function. *Astrophysical Journal*, 584(1):203–209, Feb. 2003. doi: 10.1086/345619.

J. Huchra, M. Davis, D. Latham, et al. A survey of galaxy redshifts. IV - The data. *Astrophysical Journal, Supplement*, 52:89–119, June 1983. doi: 10.1086/190860.

J. P. Huchra, L. M. Macri, K. L. Masters, et al. The 2MASS Redshift Survey—Description and Data Release. *Astrophysical Journal, Supplement*, 199 (2):26, Apr. 2012. doi: 10.1088/0067-0049/199/2/26.

T. A. Hutchinson, A. S. Bolton, K. S. Dawson, et al. Redshift Measurement and Spectral Classification for eBOSS Galaxies with the redmonster Software. *Astronomical Journal*, 152(6):205, Dec. 2016. doi: 10.3847/0004-6256/152/6/205.

D. Huterer, D. L. Shafer, D. M. Scolnic, and F. Schmidt. Testing λcdm at the lowest redshifts with SN ia and galaxy velocities. *Journal of Cosmology and Astroparticle Physics*, 2017(05):015–015, may 2017. doi: 10.1088/1475-7516/2017/05/015. URL `https://doi.org/10.1088/1475-7516/2017/05/015`.

Ž. Ivezić, S. M. Kahn, J. A. Tyson, et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *Astrophysical Journal*, 873(2):111, Mar. 2019. doi: 10.3847/1538-4357/ab042c.

J. C. Jackson. A Critique of Rees's Theory of Primordial Gravitational Radiation. *Monthly Notices of the Royal Astronomical Society*, 156(1):1P–5P, 02 1972. ISSN 0035-8711. doi: 10.1093/mnras/156.1.1P. URL `https://doi.org/10.1093/mnras/156.1.1P`.

D. H. Jones, W. Saunders, M. Colless, et al. The 6dF Galaxy Survey: samples, observational techniques and the first data release. *Monthly Notices of the RAS*, 355(3):747–763, December 2004. doi: 10.1111/j.1365-2966.2004.08353.x.

D. H. Jones, M. A. Read, W. Saunders, et al. The 6dF Galaxy Survey: final redshift release (DR3) and southern large-scale structures. *Monthly Notices of the RAS*, 399(2):683–698, October 2009. doi: 10.1111/j.1365-2966.2009.15338.x.

N. Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the RAS*, 227:1–21, July 1987. doi: 10.1093/mnras/227.1.1.

J. Kennicutt, Robert C., W. L. Freedman, and J. R. Mould. Measuring the Hubble Constant with the Hubble Space Telescope. *Astronomical Journal*, 110:1476, October 1995. doi: 10.1086/117621.

E. Kitanidis, M. White, Y. Feng, et al. Imaging Systematics and Clustering of DESI Main Targets. *arXiv e-prints*, art. arXiv:1911.05714, November 2019.

R. A. Knop, G. Aldering, R. Amanullah, et al. New Constraints on Ω_M, Ω_Λ, and w from an Independent Set of 11 High-Redshift Supernovae Observed with the Hubble Space Telescope. *Astrophysical Journal*, 598(1):102–137, November 2003. doi: 10.1086/378560.

C. S. Kochanek, D. J. Eisenstein, R. J. Cool, et al. AGES: The AGN and Galaxy Evolution Survey. *Astrophysical Journal, Supplement*, 200(1):8, May 2012. doi: 10.1088/0067-0049/200/1/8.

O. Lahav, S. L. Bridle, W. J. Percival, et al. The 2dF Galaxy Redshift Survey: the amplitudes of fluctuations in the 2dFGRS and the CMB, and implications for galaxy biasing. *Monthly Notices of the RAS*, 333(4):961–968, July 2002. doi: 10.1046/j.1365-8711.2002.05485.x.

S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *Astrophysical Journal*, 412:64–71, July 1993. doi: 10.1086/172900.

D. Lang, D. W. Hogg, and D. Mykytyn. The Tractor: Probabilistic astronomical source detection and measurement. Astrophysics Source Code Library, April 2016.

R. Laureijs, J. Amiaux, S. Arduini, et al. Euclid Definition Study Report. *arXiv e-prints*, art. arXiv:1110.3193, Oct. 2011.

O. Le Fèvre, Y. Mellier, H. J. McCracken, et al. The VIRMOS deep imaging survey. I. Overview, survey strategy, and CFH12K observations. *Astronomy and Astrophysics*, 417:839–846, April 2004. doi: 10.1051/0004-6361:20031767.

S. J. Lilly, O. Le Fevre, D. Crampton, et al. The Canada-France Redshift Survey. I. Introduction to the Survey, Photometric Catalogs, and Surface Brightness Selection Effects. *Astrophysical Journal*, 455:50, Dec. 1995. doi: 10.1086/176555.

S. J. Lilly, O. Le Fèvre, A. Renzini, et al. zCOSMOS: A Large VLT/VIMOS Redshift Survey Covering $0 < z < 3$ in the COSMOS Field. *Astrophysical Journal, Supplement*, 172(1):70–85, September 2007. doi: 10.1086/516589.

S. J. Lilly, V. Le Brun, C. Maier, et al. The zCOSMOS 10k-Bright Spectroscopic Sample. *Astrophysical Journal, Supplement*, 184(2):218–229, October 2009. doi: 10.1088/0067-0049/184/2/218.

D. N. Limber. The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. *Astrophysical Journal*, 117:134, January 1953. doi: 10.1086/145672.

J. Liske, D. J. Lemon, S. P. Driver, et al. The Millennium Galaxy Catalogue: 16 $<=B_{MGC} < 24$ galaxy counts and the calibration of the local galaxy luminosity function. *Monthly Notices of the RAS*, 344(1):307–324, Sept. 2003. doi: 10.1046/j.1365-8711.2003.06826.x.

J. Liske, I. K. Baldry, S. P. Driver, et al. Galaxy And Mass Assembly (GAMA): end of survey report and data release 2. *Monthly Notices of the RAS*, 452:2087–2126, September 2015. doi: 10.1093/mnras/stv1436.

J. Loveday, B. A. Peterson, G. Efstathiou, and S. J. Maddox. The Stromlo–APM Redshift Survey. I. The Luminosity Function and Space Density of Galaxies. *Astrophysical Journal*, 390:338, May 1992. doi: 10.1086/171284.

J. D. Lykken. Beyond the Standard Model. *arXiv e-prints*, art. arXiv:1005.1676, May 2010.

S. J. Maddox, G. Efstathiou, W. J. Sutherland, et al. Galaxy correlations on large scales. *Monthly Notices of the Royal Astronomical Society*, 242(1):43P–47P, February 1990. ISSN 0035-8711. doi: 10.1093/mnras/242.1.43P. URL `http://mnras.oxfordjournals.org/cgi/doi/10.1093/mnras/242.1.43P`.

S. J. Maddox, G. Efstathiou, and W. J. Sutherland. The APM Galaxy Survey - III. an analysis of systematic errors in the angular correlation function and cosmological implications. *Monthly Notices of the RAS*, 283(4):1227–1263, December 1996. doi: 10.1093/mnras/283.4.1227.

A. Mainzer, J. Bauer, T. Grav, et al. Preliminary Results from NEOWISE: An Enhancement to the Wide-field Infrared Survey Explorer for Solar System Science. *Astrophysical Journal*, 731(1):53, Apr. 2011. doi: 10.1088/0004-637X/731/1/53.

D. Makarov, P. Prugniel, N. Terekhova, et al. HyperLEDA. III. The catalogue of extragalactic distances. *Astronomy and Astrophysics*, 570:A13, October 2014. doi: 10.1051/0004-6361/201423496.

D. Marchesini, P. G. van Dokkum, N. M. F. Schreiber, et al. THE EVOLUTION OF THE STELLAR MASS FUNCTION OF GALAXIES FROMz= 4.0 AND THE FIRST COMPREHENSIVE ANALYSIS OF ITS UNCERTAINTIES: EVIDENCE FOR MASS-DEPENDENT EVOLUTION. *The Astrophysical Journal*, 701(2):1765–1796, aug 2009. doi: 10.1088/0004-637x/701/2/1765. URL `https://doi.org/10.1088/0004-637x/701/2/1765`.

G. Merz, M. Rezaie, H.-J. Seo, et al. The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey quasar sample: Testing observational systematics on the Baryon Acoustic Oscillation measurement. *arXiv e-prints*, art. arXiv:2105.10463, May 2021.

A. D. Miller, R. Caldwell, M. J. Devlin, et al. A Measurement of the Angular Power Spectrum of the CMB from l = 100 to 400. In *American Astronomical Society Meeting Abstracts Vol. 195*, volume 195 of *American Astronomical Society Meeting Abstracts*, page 55.08, May 2000.

J. A. Newman, M. C. Cooper, M. Davis, et al. The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts. *Astrophysical Journal, Supplement*, 208:5, September 2013. doi: 10.1088/0067-0049/208/1/5.

J. F. Nieves and R. R. Volkas. Neutrinos, Flavor Physics, and Precision Cosmology. *American Institute of Physics Conference Series*, 689, October 2003.

P. Norberg, C. M. Baugh, E. Hawkins, et al. The 2dF Galaxy Redshift Survey: luminosity dependence of galaxy clustering. *Monthly Notices of the RAS*, 328 (1):64–70, November 2001. doi: 10.1046/j.1365-8711.2001.04839.x.

S. C. Odewahn, E. B. Stockwell, R. L. Pennington, et al. Automated Star/Galaxy Discrimination With Neural Networks. *Astronomical Journal*, 103:318, January 1992. doi: 10.1086/116063.

J. K. Parejko, T. Sunayama, N. Padmanabhan, et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: the low-redshift sample. *Monthly Notices of the Royal Astronomical Society*, 429(1):98–112, 12 2012. ISSN 0035-8711. doi: 10.1093/mnras/sts314. URL `https://doi.org/10.1093/mnras/sts314`.

D. Parkinson, S. Riemer-Sørensen, C. Blake, et al. The WiggleZ Dark Energy Survey: Final data release and cosmological results. *Physical Review D*, 86(10): 103518, November 2012. doi: 10.1103/PhysRevD.86.103518.

J. A. Peacock and R. E. Smith. Halo occupation numbers and galaxy bias. *Monthly Notices of the RAS*, 318:1144–1156, November 2000. doi: 10.1046/j.1365-8711.2000.03779.x.

J. A. Peacock, S. Cole, P. Norberg, et al. A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey. *Nature*, 410(6825): 169–173, March 2001.

P. J. E. Peebles. *The large-scale structure of the universe*. Princeton University Press, 1980.

P. J. E. Peebles and J. T. Yu. Primeval Adiabatic Perturbation in an Expanding Universe. *Astrophysical Journal*, 162:815, December 1970. doi: 10.1086/150713.

W. J. Percival, C. M. Baugh, J. Bland-Hawthorn, et al. The 2dF Galaxy Redshift Survey: the power spectrum and the matter content of the Universe. *Monthly Notices of the RAS*, 327(4):1297–1306, November 2001. doi: 10.1046/j.1365-8711.2001.04827.x.

S. Perlmutter, G. Aldering, G. Goldhaber, et al. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *Astrophysical Journal*, 517(2):565–586, June 1999. doi: 10.1086/307221.

Planck Collaboration, N. Aghanim, Y. Akrami, et al. Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, art. arXiv:1807.06209, July 2018.

Planck Collaboration, N. Aghanim, Y. Akrami, et al. Planck 2018 results. I. Overview and the cosmological legacy of Planck. *Astronomy and Astrophysics*, 641: A1, September 2020. doi: 10.1051/0004-6361/201833880.

A. Ratcliffe, T. Shanks, A. Broadbent, et al. The Durham/UKST galaxy redshift survey - I. Large-scale structure in the universe. *Monthly Notices of the RAS*, 281:L47–L52, Aug. 1996. doi: 10.1093/mnras/281.3.L47.

A. Repp and I. Szapudi. Galaxy bias and σ_8 from counts in cells from the SDSS main sample. *Monthly Notices of the RAS*, 498(1):L125–L129, November 2020. doi: 10.1093/mnrasl/slaa139.

M. Rezaie, H.-J. Seo, A. J. Ross, and R. C. Bunescu. Improving galaxy clustering measurements with deep learning: analysis of the DECaLS DR7 data. *Monthly Notices of the RAS*, 495(2):1613–1640, June 2020. doi: 10.1093/mnras/staa1231.

A. G. Riess, A. V. Filippenko, P. Challis, et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *Astronomical Journal*, 116(3):1009–1038, September 1998. doi: 10.1086/300499.

A. J. Ross, R. J. Brunner, and A. D. Myers. Precision Measurements of Higher Order Angular Galaxy Correlations Using 11 Million SDSS Galaxies. *Astrophysical Journal*, 649(1):48–62, September 2006. doi: 10.1086/506269.

A. J. Ross, R. J. Brunner, and A. D. Myers. Higher Order Angular Galaxy Correlations in the SDSS: Redshift and Color Dependence of Nonlinear Bias. *Astrophysical Journal*, 665(1):67–84, August 2007. doi: 10.1086/519020.

A. J. Ross, J. Bautista, R. Tojeiro, et al. The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Large-scale structure catalogues for cosmological analysis. *Monthly Notices of the RAS*, 498(2):2354–2371, Oct. 2020. doi: 10.1093/mnras/staa2416.

V. C. Rubin, J. Ford, W. K., and N. Thonnard. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc). *Astrophysical Journal*, 238:471–487, June 1980. doi: 10.1086/158003.

O. Ruiz-Macias, P. Zarrouk, S. Cole, et al. Preliminary Target Selection for the DESI Bright Galaxy Survey (BGS). *Research Notes of the American Astronomical Society*, 4(10):187, October 2020. doi: 10.3847/2515-5172/abc25a.

O. Ruiz-Macias, P. Zarrouk, S. Cole, et al. Preliminary target selection for the DESI bright galaxy survey (BGS). *Research Notes of the AAS*, 4(10):187, oct 2020. doi: 10.3847/2515-5172/abc25a. URL `https://doi.org/10.3847/2515-5172/abc25a`.

O. Ruiz-Macias, P. Zarrouk, S. Cole, et al. Characterizing the target selection pipeline for the Dark Energy Spectroscopic Instrument Bright Galaxy Survey. *Monthly Notices of the Royal Astronomical Society*, 502(3):4328–4349, 02 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab292. URL `https://doi.org/10.1093/mnras/stab292`.

A. I. Salvador, F. J. Sánchez, A. Pagul, et al. Measuring linear and non-linear galaxy bias using counts-in-cells in the Dark Energy Survey Science Verification data. *Monthly Notices of the RAS*, 482(2):1435–1451, November 2019. doi: 10.1093/mnras/sty2802.

C. Sánchez, M. Carrasco Kind, H. Lin, et al. Photometric redshift analysis in the Dark Energy Survey Science Verification data. *Monthly Notices of the RAS*, 445 (2):1482–1506, December 2014. doi: 10.1093/mnras/stu1836.

E. F. Schlafly and D. P. Finkbeiner. Measuring Reddening with Sloan Digital Sky Survey Stellar Spectra and Recalibrating SFD. *Astrophysical Journal*, 737:103, August 2011. doi: 10.1088/0004-637X/737/2/103.

D. J. Schlegel, D. P. Finkbeiner, and M. Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *Astrophysical Journal*, 500:525–553, June 1998. doi: 10.1086/305772.

P. Schneider. *Extragalactic Astronomy and Cosmology*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. ISBN 978-3-540-33174-2. doi: 10.1007/978-3-540-33175-9. URL `http://dx.doi.org/10.1007/978-3-540-33175-9`.

N. J. Secrest, R. P. Dudik, B. N. Dorland, et al. Identification of 1.4 Million Active Galactic Nuclei in the Mid-Infrared using WISE Data. *Astrophysical Journal, Supplement*, 221(1):12, Nov 2015. doi: 10.1088/0067-0049/221/1/12.

U. Seljak. Analytic model for galaxy and dark matter clustering. *Monthly Notices of the RAS*, 318:203–213, October 2000. doi: 10.1046/j.1365-8711.2000.03715.x.

S. A. Shectman, S. D. Landy, A. Oemler, et al. The Las Campanas Redshift Survey. *Astrophysical Journal*, 470:172, Oct. 1996. doi: 10.1086/177858.

P. Simon. How accurate is Limber's equation? *Astronomy and Astrophysics*, 473 (3):711–714, October 2007. doi: 10.1051/0004-6361:20066352.

S. A. Smee, J. E. Gunn, A. Uomoto, et al. THE MULTI-OBJECT, FIBER-FED SPECTROGRAPHS FOR THE SLOAN DIGITAL SKY SURVEY AND THE BARYON OSCILLATION SPECTROSCOPIC SURVEY. *The Astronomical Journal*, 146(2):32, jul 2013. doi: 10.1088/0004-6256/146/2/32. URL `https://doi.org/10.1088/0004-6256/146/2/32`.

A. Smith, S. Cole, C. Baugh, et al. A Lightcone Catalogue from the Millennium-XXL Simulation. *Mon. Not. Roy. Astron. Soc.*, 470(4):4646–4661, 2017. doi: 10.1093/mnras/stx1432.

A. Smith, J.-h. He, S. Cole, et al. Correcting for fibre assignment incompleteness in the DESI Bright Galaxy Survey. *Monthly Notices of the RAS*, 484(1):1285–1300, March 2019. doi: 10.1093/mnras/stz059.

G. F. Smoot, C. L. Bennett, A. Kogut, et al. Structure in the COBE Differential Microwave Radiometer First-Year Maps. *Astrophysical Journal, Letters*, 396:L1, September 1992. doi: 10.1086/186504.

R. S. Somerville and R. Davé. Physical Models of Galaxy Formation in a Cosmological Framework. *Annual Review of Astron and Astrophys*, 53:51–113, August 2015. doi: 10.1146/annurev-astro-082812-140951.

C. C. Steidel, K. L. Adelberger, A. E. Shapley, et al. Lyman Break Galaxies at Redshift z ∼3: Survey Description and Full Data Set. *Astrophysical Journal*, 592 (2):728–754, Aug. 2003. doi: 10.1086/375772.

C. Stoughton, J. Adelman, J. T. Annis, et al. Data Processing Factory for the Sloan Digital Sky Survey. In *Survey and Other Telescope Technologies and Discoveries*, volume 4836, pages 339–349, December 2002. doi: 10.1117/12.457014.

M. A. Strauss, D. H. Weinberg, R. H. Lupton, et al. Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample. *Astronomical Journal*, 124:1810–1824, September 2002. doi: 10.1086/342343.

I. Szapudi. A New Method for Calculating Counts in Cells. *Astrophysical Journal*, 497(1):16–20, April 1998. doi: 10.1086/305439.

I. Szapudi and S. Colombi. Cosmic Error and Statistics of Large-Scale Structure. *Astrophysical Journal*, 470:131, October 1996. doi: 10.1086/177855.

The Dark Energy Survey Collaboration. The Dark Energy Survey. *arXiv e-prints*, art. astro-ph/0510346, October 2005.

J. L. Tonry, B. P. Schmidt, B. Barris, et al. Cosmological Results from High-z Supernovae. *Astrophysical Journal*, 594(1):1–24, September 2003. doi: 10.1086/376865.

G. Vettolani, E. Zucca, G. Zamorani, et al. The ESO Slice Project (ESP) galaxy redshift survey. I. Description and first results. *Astronomy and Astrophysics*, 325: 954–960, Sept. 1997.

Y. Wang, R. J. Brunner, and J. C. Dolence. The SDSS Galaxy Angular Two-Point Correlation Function. *Mon. Not. Roy. Astron. Soc.*, 432:1961, 2013. doi: 10.1093/mnras/stt450.

N. Weir, U. M. Fayyad, and S. Djorgovski. Automated Star/Galaxy Classification for Digitized Poss-II. *Astronomical Journal*, 109:2401, June 1995. doi: 10.1086/117459.

S. D. M. White. Further simulations of merging galaxies. *Monthly Notices of the RAS*, 189:831–852, December 1979. doi: 10.1093/mnras/189.4.831.

E. L. Wright, P. R. M. Eisenhardt, A. K. Mainzer, et al. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance.

Astronomical Journal, 140:1868-1881, December 2010. doi: 10.1088/0004-6256/140/6/1868.

H. K. C. Yee, S. L. Morris, H. Lin, et al. The CNOC2 Field Galaxy Redshift Survey. I. The Survey and the Catalog for the Patch CNOC 0223+00. *Astrophysical Journal, Supplement*, 129(2):475–492, Aug. 2000. doi: 10.1086/313426.

D. G. York, J. Adelman, J. Anderson, John E., et al. The Sloan Digital Sky Survey: Technical Summary. *Astronomical Journal*, 120(3):1579–1587, September 2000. doi: 10.1086/301513.

P. Zarrouk, E. Burtin, H. Gil-Marín, et al. The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: measurement of the growth rate of structure from the anisotropic correlation function between redshift 0.8 and 2.2. *Monthly Notices of the RAS*, 477(2):1639–1663, June 2018. doi: 10.1093/mnras/sty506.

I. Zehavi, D. H. Weinberg, Z. Zheng, et al. On Departures from a Power Law in the Galaxy Correlation Function. *Astrophysical Journal*, 608(1):16–24, June 2004. doi: 10.1086/386535.

H. Zheng and Y. Zhang. Review of techniques for photometric redshift estimation. In N. M. Radziwill and G. Chiozzi, editors, *Software and Cyberinfrastructure for Astronomy II*, volume 8451, pages 1012 – 1023. International Society for Optics and Photonics, SPIE, 2012. doi: 10.1117/12.925314. URL `https://doi.org/10.1117/12.925314`.

Z. Zheng, A. A. Berlind, D. H. Weinberg, et al. Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies. *Astrophysical Journal*, 633(2):791–809, November 2005. doi: 10.1086/466510.

A. Zonca, L. Singer, D. Lenz, et al. healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in python. *Journal of Open Source*

Software, 4(35):1298, March 2019. doi: 10.21105/joss.01298. URL `https://doi.org/10.21105/joss.01298`.

Colophon

This thesis is based on a template developed by Matthew Townson and Andrew Reeves. It was typeset with LaTeX 2_ε. It was created using the *memoir* package, maintained by Lars Madsen, with the *madsen* chapter style. The font used is Latin Modern, derived from fonts designed by Donald E. Knuth.