Durham E-Theses

Investigating the Impact of Environment on the Galaxy Luminosity Function.

MOORE, SAMUEL,GARETH

Investigating the Impact of Environment on the Galaxy Luminosity Function

Samuel Moore

A Thesis presented for the degree of
Doctor of Philosophy

Institute for Computational Cosmology
Durham University
United Kingdom
September 2024

Abstract

In this thesis, we present the DESI Y1 Bright Galaxy Survey (BGS) Luminosity Functions (LFs) in the r, g, z and w_1-bands from $0.002 < z < 0.6$. To do this, we describe a method for finding k-corrections based on the FastSpecFit (FSF) k-corrections; an evolutionary model for finding e-corrections; and a series of weights to account for target and redshift incompleteness of the BGS Y1 survey. We construct LFs that are extremely faint, reaching magnitudes of $M_r - 5 \log h \sim -10$ in the r-band. Moreover, we observe the existence of an upturn in the LFs at $M_r - 5 \log h > -15$. These LFs are in agreement with those presented in the literature. We further validate our results by using a range of methodologies ($1/V_{\mathrm{max}}$, $1/V_{\mathrm{dc, max}}$, SWML) to construct the LFs, which broadly agree with each other. We note some areas where further investigation is needed. This includes a disparity between the North and South LFs at the bright end, a potential need for a more complex evolutionary model, and the issue of imaging systematics at the faint end of the LF. Nevertheless, the small jackknife errorbars on our global LFs demonstrate that our results are well-constrained, and these errors will only become smaller with the release of the Y3 and Y5 BGS datasets. Moreover, we confirm that our methodology of dealing with the differing photometry in North and South is broadly successful.

We extend this analysis by using the luminosity function to investigate the environmental dependencies of galaxies. In particular, we investigate how luminosity and colour depend on local density, and present our results here. These results agree well with prior results in the literature from GAMA, indicating that we have developed a successful methodology for dealing with boundary corrections and holes in the survey. Although the effective volume of the BGS survey is similar to the GAMA results due to boundary corrections, this will yield promising results on the Y3 and Y5 results which will be more complete and will have a much larger effective volume.

Finally, we then use our methodology to generate $1/V_{\mathrm{max}}$ Stellar Mass Functions (SMFs). With these SMFs, we help to test and validate a new methodology called Photometric Objects Around Cosmic Webs (PAC). PAC was developed to estimate

the excess projected density distribution of a spectroscopic catalogue by utilising the signal in the cross-correlation of faint galaxies from the Legacy photometric surveys with the brighter DESI BGS galaxies with known spectroscopic redshift. In doing so, PAC can estimate the SMF in a novel way, and yields results that agree with our $1/V_{\mathrm{max}}$ SMF above $10^9 M_\odot$ but currently differs at the low-mass end. Both our LF and SMF results serve as useful results that can act to better constrain and distinguish between different galaxy formation models.

Declaration

The work in this thesis is based on research carried out at the Institute for Computational Cosmology, Durham University, United Kingdom. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text. For this thesis, we make use of FastSpecFit (FSF), which is the work of Moustakas et al. (2023), and Photometric Objects Around Cosmic Webs (PAC), which is the work of Xu et al. (2022a). The work presented in this thesis is intended to be published as two papers. Chapters 2 and 3, presenting methodology and global LF results, will be published in the first paper (Moore et al., in prep., a). Chapter 4, which extends this to density-dependent LF results, will be published in the second paper (Moore et al., in prep., b).

Acknowledgements

This thesis could not have come to fruition without the support and guidance of a broad range of people. First, I wish to thank my thesis advisor, Shaun Cole, who has been a source of valuable advice throughout the entire process. I am especially appreciative of his extensive knowledge and expertise which has helped to drive this project forward. Additionally, I wish to thank my other advisor, Peder Norberg, for his level of insight and scrutiny in dealing with some of the challenges of this thesis as well as his attention to detail. Another person to thank in this research is Michael Wilson, whose guidance in coding and general problem solving proved to be extremely useful in getting over some of the early hurdles of this thesis. I have fond memories of walking to the Picnic Basket for lunch and discussing the problems of the day.

More broadly, I extend my thanks to the Durham Quizbowl team, including my former housemate Michael Kohn, Kai Madgwick, Isaiah Silvers and Harry Scully. There are too many high points and fond memories to mention them all - but the accumulation of our efforts to win ACF Regionals must be recognised. I also thank my COQL team, Kai Madgwick and Toby Attala, and my broader OQL team. These thanks go to the wider quizbowl community in general, which in its rigour in the pursuit of knowledge it must be confessed has taught me more about my own subjects than I might care to admit.

Like all large projects, there is much to be said about the minutia of day-to-day life - those small parts that nonetheless are a vital part of the process. There are, by definition, too many to mention, but I will highlight the broader ICC and Durham community. There are also countless other individuals who have aided me over the past four years - you know who you are, and have my eternal thanks.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

The nature of galaxy formation and evolution has been a significant question in the history of astrophysics. In Chapter 1, we outline the background of galaxy formation and evolution, highlighting the question of interest in this thesis: how does the environment affect galaxy evolution?

1.1 Introduction to Cosmology

In order to understand galaxy formation, it is important to outline some of the fundamental principles of cosmology which are important in this process. The universe is observed to be expanding (Bahcall, 2015; Hubble, 1929) with recent observations finding that this rate of expansion is accelerating (e.g: as demonstrated by Type Ia Supernovae observations in Riess et al., 1998). In addition, the universe is considered to be homogeneous and isotropic on sufficiently large scales, typically at length scales of $> 150h^{-1}\mathrm{Mpc}$. This is known as the Cosmological Principle (Barrow, 1989; Marinoni et al., 2012). The assumption that the Cosmological Principle holds is a major underpinning of cosmology, allowing for the use of the Friedmann-Lemaître-Robinson-Walker (FLRW) metric and the later derivation of a number of

fundamental equations in cosmology (Melia, 2022). In order for homogeneity and isotropy to be preserved over time, the universe must expand by some scale factor $a(t)$ that is the same throughout the universe, such that

$$D_{\mathrm{p}}(t) = a(t)D_{\mathrm{p}}(t_0), \qquad (1.1)$$

where $D_{\mathrm{p}}(t)$ is the proper distance at time t, and t_0 represents the present day. The proper distance is the length between two objects defined along a space-like path (Wald, 1984). By definition, $a(t_0) = 1$. The scale factor allows for a useful conversion between proper and co-moving coordinates: $\mathbf{r}(t) = a(t)\mathbf{x}$, where $\mathbf{x}$ is the co-moving coordinate and $\mathbf{r}(t)$ is the proper coordinate. From this definition of the scale factor, the relationship between velocity and proper distance may be derived as

$$v = \frac{dD_{\mathrm{p}}}{dt} = \frac{d}{dt}(aD_{\mathrm{p,0}}) = \dot{a}D_{\mathrm{p,0}} = \frac{\dot{a}}{a}D_{\mathrm{p}} = H(t)D_{\mathrm{p}}, \qquad (1.2)$$

where we define $H(t) = \dot{a}/a$. The velocity of objects caused solely by the expansion of the universe is known as the Hubble flow. Note that $H(t_0) = H_0$ is the Hubble constant, the present-day rate of expansion that is a constant in space, but not time. There is significant debate as to the exact value of the Hubble constant, with low redshift observations giving notably different results from high redshift CMB observations. For example, Planck CMB observations give $H_0 = 67.66 \pm 0.42$ $\mathrm{km\,s^{-1}Mpc^{-1}}$ (Planck Collaboration et al., 2020), while more local observations such as SH0ES find $H_0 = 73.04 \pm 1.04$ $\mathrm{km\,s^{-1}Mpc^{-1}}$ (Riess et al., 2022). This disparity, known as 'Hubble tension', is an unsolved problem in cosmology to date and likely illustrates that we have an incomplete picture of cosmology (see Freedman, 2021, for a comprehensive summary with discussion of potential solutions). As we shall subsequently show in this section, many cosmological quantities such as distances are dependent on the value of H_0. Many papers define $H_0 \equiv 100h$ $\mathrm{km\,s^{-1}Mpc^{-1}}$, where h is a dimensionless constant used in distance units (e.g. $h^{-1}\mathrm{Mpc}$) so that the numerical values of distances do not depend on the uncertainty in the Hubble constant.

The expansion of the universe results in photons being redshifted, where the

wavelength $\lambda \propto a$. A source can also be redshifted due to peculiar velocities, due to velocities relative to the Hubble flow causing a Doppler effect, or due to photons leaving gravitational potentials leading to gravitational redshifting (Weinberg, 2008). However, on sufficiently large scales, cosmological redshift dominates. The redshift, z, of an object may be defined as

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{emit}}}{\lambda_{\text{emit}}}, \tag{1.3}$$

where λ_{obs} is the wavelength of the photon measured by an observer, and λ_{emit} is the emitted wavelength of the photon at its source. Using the fact that $\lambda \propto a$, it follows that

$$a(z) = \frac{1}{1 + z}. \tag{1.4}$$

The expansion of the universe may be characterised by the Friedmann Equations, which are derived from the Einstein Field Equations (Weinberg, 2008). The first and second Friedmann Equations are

$$H^2(t) \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G \rho(t)}{3} - \frac{Kc^2}{a(t)^2} + \frac{\Lambda c^2}{3}, \tag{1.5}$$

and

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \tag{1.6}$$

where $\rho(t)$ is the density of the universe at time t. K is a constant that corresponds to the curvature of the universe. For a flat universe, $K = 0$. For a positively curved universe, $K > 0$, and for a negatively curved universe, $K < 0$. Empirical observations suggest that the universe is flat (Efstathiou & Gratton, 2020). Λ is the cosmological constant, which drives the acceleration of the expansion of the universe (see the discussion of dark energy below). p is the pressure. For a flat universe with $\Lambda = 0$, $H^2 = 8\pi G\rho/3$. From this, we define

$$\rho_{\text{crit}} = \frac{3H_0^2}{8\pi G} \tag{1.7}$$

3

as the critical density of the universe (the mean density in a flat universe with no cosmological constant). This is then used as a normalisation factor to define density parameters for different components of the universe,

$$\Omega_i = \frac{\rho_i}{\rho_{\text{crit}}}. \tag{1.8}$$

By modelling the universe as a perfect fluid, we can use the equation of state relating ρ and p

$$p = w\rho c^2, \tag{1.9}$$

where w is a constant. By combining this with the second Friedmann equation,

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a} \tag{1.10}$$

which has the solution

$$\rho(t) \propto a(t)^{-3(1+w)}. \tag{1.11}$$

In other words, the density of the universe is dependent on $a(t)$ and w. There exist several different components of the universe, which in turn have different values of w. This indicates that different components of the universe have a differing density evolution. The components of particular interest are matter (baryonic and dark) (Ω_{m}), radiation (Ω_{r}), vacuum energy ($\Omega_\Lambda \equiv \Lambda c^2/(3H_0^2)$), and curvature ($\Omega_K \equiv -Kc^2/a^2H_0^2$).

Specifically, we denote matter to refer to material with a non-relativistic thermal velocity, including baryonic matter. From this non-relativistic condition, $p_{\text{m}} \ll \rho_{\text{m}}c^2$, which implies that $w = 0$. From this, $\Omega_{\text{m}} \propto (1+z)^3$. Radiation refers to particles that are relativistic. This includes photons as well as other high-velocity particles. Radiation follows $p_r = (1/3)\rho_r c^2$ with $w = 1/3$, resulting in $\Omega_{\text{r}} \propto (1+z)^4$.

In addition to matter and radiation, we require a negative pressure to account for the observed acceleration of the expansion of the universe. In order for $\ddot{a} > 0$,

one requires $\rho + 3p/c^2 < 0$, which means $p < (-1/3)\,\rho c^2$. In other words, there exists a component with $w < -1/3$. The vacuum energy is a Lorentz invariant non-zero energy that is postulated to exist due to the quantum field of the vacuum. Due to this invariance across frames, the pressure is constant such that $p_\Lambda = -\rho_\Lambda c^2$. The vacuum energy is equivalent to the cosmological constant with $\Lambda = 8\pi G \rho_\Lambda / c^2$, where a value of $\Lambda > 0$ results in acceleration. Additional components may be seen in the literature, such as neutrinos (Ω_ν), but are not used in this thesis. We note that $\Omega = \sum_i \Omega_i = 1$ for a flat ($K = 0$) universe, which follows from Eqn. 1.5. From the above, we can write

$$\left(\frac{H}{H_0}\right)^2 = \Omega_{\mathrm{m}}(1+z)^3 + \Omega_\Lambda + \Omega_{\mathrm{r}}(1+z)^4. \tag{1.12}$$

Given that these densities have different time dependencies, the dominant component of the universe has changed over time. In the early universe, the universe was radiation dominated. Later on, the universe became matter dominated. Today, the universe is dominated by the (dark) energy density associated with the cosmological constant.

The current prevailing consensus amongst cosmologists is that the universe is best described by the ΛCDM model (Blanchard et al., 2024). In this model, there exists non-relativistic ('cold') dark matter which is unobservable except through gravitational interactions. Dark matter existence has been inferred through galaxy rotation curves (Rubin et al., 1980), and through weak lensing studies (Huterer, 2010). The identification of DM remains an ongoing mystery for observers and theorists alike, with major contenders for DM including axions (Adams et al., 2023) and WIMPS (Roszkowski et al., 2018). In addition, there exists dark energy (in the form of the cosmological constant Λ), which acts as a negative pressure driving the accelerated expansion of the universe. The ΛCDM model is not perfect, and a number of problems have emerged over the last few decades (Perivolaropoulos & Skara, 2022). However, the ΛCDM model remains the most widely supported model in cosmology at the present time (Blanchard et al., 2024).

There are various distance definitions that are used in cosmology. One useful

definition is the comoving distance (d_c, sometimes denoted χ in the literature). We define a comoving coordinate system which moves with the Hubble flow - that is, comoving coordinates expand with the universe. From this, the comoving distance relates to the proper distance in a flat universe as

$$d_c = \frac{D_p(t)}{a(t)} = D_p(t)(1+z).\tag{1.13}$$

We note that $D(t_0) = D_p(t)/a(t)$, confirming the time-independence of the comoving distance. Moreover, we can define the luminosity distance, d_L, in a flat universe as

$$d_L = c(1+z)\int_0^z \frac{dz'}{H(z)},\tag{1.14}$$

which is in turn related to the comoving distance as

$$d_L = \frac{d_c}{a(t)} = d_c(1+z).\tag{1.15}$$

We emphasise that these equations are for a flat universe, see e.g: Section 3.4 of Peacock (1999) for the more general case.

1.2 Introduction to Galaxy Formation

Very shortly after the Big Bang, the universe underwent a period of rapid expansion known as cosmic inflation (Guth, 1981). During inflation, quantum fluctuations in the scalar field are stretched, resulting in the initial density perturbations from which structure grows (Linde, 1982). These density perturbations, otherwise known as overdensities, are observed to be scale-invariant (as is predicted in the inflationary model). Over time, this leads to hierarchical structure formation - whereby these density perturbations led to the formation of Large Scale Structure (LSS). Density perturbations are formally defined as

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(\mathbf{x}, t)}{\bar{\rho}(\mathbf{x}, t)},\tag{1.16}$$

where $\rho(\mathbf{x}, t)$ is the density of the point and $\bar{\rho}(\mathbf{x}, t)$ is the mean density of the universe. For an overdensity, $\rho(\mathbf{x}, t) > \bar{\rho}(\mathbf{x}, t)$. Initially, overdensities undergo linear growth when $|\delta| = |\frac{\delta\rho}{\rho}| << 1$. In this linear regime, the perturbations can be calculated exactly using perturbation theory (Peebles, 1980). Assuming a ΛCDM universe, the linear perturbation equation is

$$\frac{\partial^2 \delta}{\partial t^2} + 2H(t)\frac{\partial \delta}{\partial t} - \frac{\nabla_x^2 \delta p}{a^2 \bar{\rho}} - 4\pi G \bar{\rho}(t)\delta = 0. \tag{1.17}$$

This equation assumes that the fluid is Newtonian. In the relativistic case, the final term is modified to $(32\pi G/3)\bar{\rho}(t)\delta$ (Lima et al., 1997). For the case of cold dark matter, which can be treated as a fluid with $p = 0$, the above equation may be solved using separation of variables, $\delta(\mathbf{x}, t) = A(\mathbf{x})D(t)$, to yield

$$\delta(\mathbf{x}, t) = A(\mathbf{x})D_+(t) + B(\mathbf{x})D_-(t), \tag{1.18}$$

where $D_+(t)$ and $D_-(t)$ are the growing and decaying modes respectively. Typically the decaying mode is neglected as the ratio of the growing modes between two time periods t_1 and t_2 will be much greater than the comparable decaying mode ratio for $t_2 >> t_1$. In other words, any decaying mode will decay to a negligible amplitude in the long-run. In the linear regime, all modes evolve as $\delta \propto D_+(a)$ where $D_+(a)$ is called the growth factor. The differential equation that characterises linear growth is evidently dependent on the cosmology of the universe. It can be shown that

$$D_+(a) \propto \begin{cases} a^2, & \text{if } a < a_{\text{eq}} \\ a, & \text{if } a_{\text{eq}} < a < a_{\text{DE}} \\ \text{constant}, & a > a_{\text{DE}}, \end{cases} \tag{1.19}$$

where a_{eq} represents the scale factor at matter-radiation equality, and a_{DE} represents the scale factor at matter-dark energy equality. Eventually, as the overdensity becomes larger over time, it can no longer be modelled by linear theory as it moves into the non-linear growth regime. Whilst structures in the linear regime expand with the Hubble flow of the universe, non-linear structures will separate from the

Hubble flow and start to collapse. It is not possible to analytically solve for clustering in the non-linear regime. As a result, various N-body simulation techniques have been developed to follow perturbations into the non-linear regime, pioneered by Efstathiou & Eastwood (1981) and Aarseth et al. (1979). This results in the formation of gravitationally bound structures. In particular, dark matter (DM) haloes are formed in this process, which can then act as gravitational potential wells for further structure formation. These haloes may undergo two major methods that lead to mass growth: the accretion of dark matter onto the halo, and DM halo mergers. This follows a 'bottom-up' approach of hierarchical structure formation where small structures form first and subsequently merge to create larger ones (White & Frenk, 1991).

Baryonic gas may form an inflow into the DM halo, resulting in regions of high density gas. This gas is typically heated by shocks during accretion. Various processes result in the cooling of the gas. For example, the dominant form of cooling on galactic scales are recombination lines. Another less prominent mechanism for cooling is Bremsstrahlung for high temperature gas ($T \sim 10^7 K$). Additionally, inverse Compton scattering is important in the early universe. This cooling leads to a reduction in the gas pressure that acts against gravity, leading to collapse into discs that are supported by angular momentum. Eventually, once the surface density is high enough, star formation will begin to occur, resulting in a galaxy (Rees & Ostriker, 1977).

Galaxies are typically classified as a given morphology based on the Hubble tuning fork which divides galaxies into ellipticals and spirals, as well as irregulars (Longair, 2008). This system is useful as a classification system for most luminous galaxies, with only a small proportion of irregular galaxies. The origin of the abundances of these different morphologies remains a somewhat open question. In the paradigm of hierarchical structure formation, the inflow of baryonic gas forms spiral galaxies. However, later galactic mergers and tidal forces are proposed as a dominant mechanism for the production of elliptical galaxies (Gott, 1975; Kormendy et al., 2009). In addition, the proportion of ellipticals and S0 galaxies increases with local density, while the proportion of spiral galaxies decreases (Dressler, 1980). This

indicates an important role for the environment in influencing the morphological evolution of galaxies.

Whilst the fundamental details of galaxy formation are outlined above, this is by no means a solved area of astrophysics. In particular, the picture becomes more complicated when baryonic matter must be considered. Various models have sought to simulate galaxy formation with different parameters - including semi-analytic models and hydrodynamic simulations to better understand the role of various processes in galaxy formation (Bower et al., 2006; Cole et al., 2000; Crain et al., 2015; Kauffmann et al., 1993).

Hydrodynamic simulations can directly follow the shock heating and cooling gas dynamics but require sub-grid analytic models for other processes. This includes star formation and the corresponding feedback into the ISM from supernovae and stellar wind, AGN accretion and feedback, and black hole dynamics. Examples of hydrodynamic simulations include EAGLE (Crain et al., 2015) and FLAMINGO (Schaye et al., 2023). Parameters in all of these models are calibrated based on empirical observations. For example, this typically includes selecting star formation parameters based on the Kennicutt-Schmidt law (Kennicutt, 1998). Generally, simulations are calibrated based on the local Stellar Mass Function (SMF) and the galaxy LF which can constrain a combination of the halo mass function feedback parameters.

Semi-analytic models (SAM) use approximate, analytical methods to investigate galaxy formation processes. SAMs make use of DM merger trees which are typically found using N-body simulations with specific cosmological parameters. Alternatively, they may use analytic approximations such as extended Press-Schechter formalism (Bond et al., 1991; Bower, 1991; Press & Schechter, 1974). For SAMs, it is necessary to consider the processes that lead to observable galaxy properties from the DM haloes. This includes gas cooling, star formation, feedback systems, galaxy mergers, chemical/metallicity evolution, and many more factors. Despite significant progress over the past few decades, many of these processes remain poorly understood. As a result, these processes are not directly simulated in SAMs, but rather these are modelled by general scaling/power laws. Prominent examples of

semi-analytic models include models from the Durham group such as GALFORM (Cole et al., 2000), models from the Munich group such as SAGE (Croton et al., 2006; Croton et al., 2016), and models from the Santa Cruz group (Somerville et al., 2008).

There are various reasons why one approach may be favoured over the other. For example, SAMs tend to be less computationally expensive to run compared to hydrodynamic simulations due to their treatment of sub-grid physics. This allows for a much larger number of galaxies to be used. It should be noted that comparisons of the two approaches generally show good agreement in specific cases - for example, in a 'stripped-down' version of GALFORM that only incorporates shock heating and radiative cooling of gas, there is strong agreement between the predictions of the two models for the galaxy mass distribution and the two-point correlation functions. (Helly et al., 2003). However, there are many observed differences between the predictions of many of these models. A comparison of EAGLE and GALFORM finds that there are some substantial differences in their predictions. For example, Guo et al. (2016) finds that GALFORM leads to different metallicity distributions as a function of stellar mass and broadly a low evolution in the star formation rate. Mitchell et al. (2017) extends this analysis by galaxy matching, showing that there exists divergence in the stellar mass distribution at high redshift (and hence different stellar evolution) between the two models. Moreover, GALFORM overestimates the stellar specific angular momentum. In short, there remain significant differences between various hydrodynamic and semi-analytic models. Generally, these various models also still differ from empirical observations, suggesting that there remains physics to be understood (Naab & Ostriker, 2017).

Importantly, in both hydrodynamic simulations and semi-analytic models, there are freedoms as they are not able to directly simulate a number of the processes. These free parameters have to set by comparison to observations, of which the galaxy LF is one of the most fundamental. In particular, the LF has been essential in constraining the SN and AGN feedback parameters. Specifically, SN feedback aids the process of reducing galaxy formation in low mass haloes, resulting in a reduction in the faint-end of the LF (Kim et al., 2012) and AGN feedback is important in

reducing galaxy formation in high mass haloes (Bower et al., 2006). Moreover, different choices of these free parameters will alter such things as the balance between AGN and SN feedback and the haloes and environments in which they occur. This will inevitably give rise to differing environmental dependencies of galaxy properties for different parameter choices. As a result of this, accurate measurements of how galaxy properties depend on environment can distinguish between galaxy formation models.

1.3 The Role of Environment

The role of the environment in influencing the properties of a galaxy in its formation and evolution is a topic of significant interest. In the past, several studies have sought to investigate these environmental dependencies, including the dependence of luminosity and colour on local density (McNaught-Roberts et al., 2014) and cosmic web classification (Kraljic et al., 2017).

These studies have been facilitated by large redshift surveys which have surveyed a large area of the sky to sufficiently high redshifts to high completeness. This has allowed for data analysis on a large catalogue of galaxies to produce statistically significant results. In particular, the galaxy luminosity function (LF) is a method of choice for investigating these environmental dependencies. The LF, sometimes split by colour or morphology, has been used to investigate how the distribution of luminosity depends on environmental measures such as local density (McNaught-Roberts et al., 2014), void/filament/cluster classification (Kraljic et al., 2017; Moorman et al., 2015), group multiplicity (Robotham et al., 2010), virial mass, and other factors.

For example, McNaught-Roberts et al. (2014) uses GAMA DR2 data (Driver et al., 2011) to confirm a significant impact of environment on colour, finding that red galaxies dominate in overdense environments while blue galaxies dominate in underdense environments. Tempel, E. et al. (2011) finds using SDSS DR7 data that the impact of environment is highly dependent on morphology. In particular, the LF of elliptical galaxies is highly dependent on environment, whereas the LF of spiral galaxies is independent of environment. Moreover, density is found to be

more important for influencing the red elliptical LF than the blue elliptical LF.

In addition, the impact of environment on the faint-end of the LF has been investigated in various papers (Barrena, R. et al., 2012; Liu et al., 2008; Trentham & Tully, 2002). For example, Liu et al. (2008) finds that in the COSMOS field, the faint-end slopes of the LF differ for different SED categories, and additionally become less steep as a function of redshift. This has been attributed to either evolution of the LF, or the existence of low surface brightness dwarf galaxies that are undetected at high redshifts. The analysis of Trentham & Tully (2002) presents the LF as faint as $M_r - 5 \log h \sim 10$ for different galaxy densities and morphologies, finding the faint end of the LFs to be lower than predicted.

An empirical measurement of the LF is important to better assess and constrain theoretical models of galaxy formation; theoretical models must agree with the measured LF. Given the large amount of data that DESI will be recording, it will significantly improve LF measurements from its large sample size. Statistical errors in the LF will be small - instead, it will be more important than ever to carefully consider the systematics. The LF has been a useful method for validating various galaxy formation models. In particular, models can investigate the processes that impact the faint-end of the LF. For example, Mathis & White (2002) use simulations of the local universe to investigate voids, and predict that in underdense regions, the slope of the LF will be higher than in overdense regions. McNaught-Roberts et al. (2014) finds empirically that the faint end of the LF steepens with overdensity only in overdense regions for a given galaxy population.

A number of papers have found the faint-end of the LF to be less steep than predicted by numerical simulations (Hoyle et al., 2005; Moretti, A. et al., 2015) suggesting that there is further scope to improve galaxy simulations. A major aim of this thesis is to use the DESI survey to make more accurate measures of the environmental dependence of the galaxy luminosity function.

1.4 Introduction to DESI

The Dark Energy Spectroscopic Instrument (DESI) is a next-generation Stage IV survey[1] based at Kitt Peak, Arizona as part of the 4-m Mayall Telescope. The primary goal of DESI is to measure various aspects of the LSS in order to constrain a range of cosmological parameters to sub-percent precision. This includes Baryon Acoustic Oscillations (BAOs) across a range of redshift ranges, in addition to Redshift Space Distortions (RSDs). In order to accomplish this, DESI will observe galaxies over an approximate area of 14,000 deg^2.

The DESI focal plane consists of 5000 robotically controlled fibre positioners which each hold a unique fibre for object targeting. There exist fibres which are used for other purposes, such as sky brightness observations. This focal plane is divided into 10 separate regions called 'petals', each consisting of 500 fibres (Silber et al., 2022). Each fibre from each petal is in turn connected a three-arm spectrograph containing three spectroscopic cameras with wavelength ranges 3600–5800 Å, 5760–7620 Å and 7520–9824 Å (DESI Collaboration et al., 2024c).

As part of the DESI project, there are five main target classes: Emission Line Galaxies (ELGs), Luminous Red Galaxies (LRGs), quasars (QSOs), the Milky Way Survey (MWS) and the Bright Galaxy Survey (BGS) (DESI Collaboration et al., 2016). There are additional secondary classes not discussed here. This thesis will focus on BGS, further detailed in Section 2.1.

DESI operations began in 2019 and have continued to the present, with some minor disruptions. As a brief outline to DESI operations, the major data releases are SV1, SV3 and Y1. SV1 was a data commissioning release that ran from 2020-dec-14 to 2021-april-02 and was used for initial testing of target selection (Myers et al., 2023). This dataset was spatially highly incomplete with varied target selection cuts (with the aim of investigating the most efficient selection cuts) but had longer

---

[1]Stage IV is a designation of the Dark Energy Task Force (DETF) commissioned by NSF, NASA and DOE to designate larger, long-term projects that can better constrain dark energy measurements. In particular, Stage IV projects aim in combination to minimise the area of the 95% confidence limit in the $w_0 - w_a$ plane by a factor of 10 from prior dark energy studies, such as SDSS-II, PanSTARRS-1, ACT, etc. Other examples of Stage IV projects include LSST/Rubin andSKA. (Albrecht, 2006)

exposures resulting in higher S/N spectra. SV3 (otherwise called the One-Percent Survey) ran from 2021-apr-05 to 2021-jun-10 and was designed to be a highly complete sample with high fibre assignment completeness ($\sim 95\%$ complete), covering an area of 180 deg^2 (comparable to the total size of the GAMA equatorial fields, G9, G12 and G15 (Driver et al., 2011)). As such, this is a useful reference dataset for verifying incompleteness corrections. Moreover, it was designed to be highly overlapping with other surveys, making it a useful dataset for verification of other measurements. The Early Data Release (EDR) is a release of spectra taken from the Survey Validation datasets. In other words, it is a combination of the SV1, SV2 and SV3 datasets (among others) (DESI Collaboration et al., 2024c). This has been followed by the DESI DR1 data, otherwise known as the Year One (Y1) data. The Y1 dataset was taken up to 14-jun-2022 and encompasses the EDR data (DESI Collaboration, prep). Further information on the Y1 BGS data is presented in Chapter 2. Initial work for this thesis was conducted on the SV3 and EDR datasets, however, this thesis primarily presents results from the Y1 dataset.

1.5   Thesis Roadmap

With the background of galaxy formation outlined, we proceed to investigate LFs with the DESI BGS Y1 dataset. In Chapter 2 of this thesis, we outline the selection criteria for DESI BGS Bright and other notable details. In Section 2.4, we describe our methodology for calculating polynomial-derived k-corrections for each of the BGS galaxies. In Section 2.5, we outline incompleteness weights, including a new redshift incompleteness weight that we calculate. In Chapter 3, we outline different LF methodologies in Section 3.1, including an outline of error estimation. We also derive an evolutionary correction factor (e-corrections) for the absolute magnitudes, outlined in Section 3.2. We present the global LFs in the g, r, z and w_1 bands in Section 3.3. We additionally present the colour-dependent LFs in Section 3.5. Additional analysis of the LFs, including potential imaging issues are discussed in Section 3.6. In Chapter 4, we outline a method for calculating the local density of a galaxy, and present density-dependent LFs in Section 4.3. In Chapter 5, we apply

our LF methodology to present Stellar Mass Functions (SMFs). Final conclusions on this thesis are presented in Chapter 6.

DESI BGS

In this Chapter, we outline the DESI Bright Galaxy Survey (BGS). We present a method of estimating for all galaxies absolute magnitudes in the same rest-frame waveband. Additionally, we present a method of calculating redshift incompleteness corrections for the survey, which we use to calculate new incompleteness weights for the LF.

2.1   Introduction to BGS

BGS is the division of DESI that focuses on the mapping of more than 10 million galaxies from $0 < z < 0.6$. BGS is divided into two categories: BGS Bright and BGS Faint. BGS Bright is a $r \lesssim 19.5$ magnitude limited sample. There is an additional faint fibre magnitude (r_{fibre}) limit included that is far away from the locus of galaxies (Hahn et al., 2023). The purpose of this selection cut is to remove spurious objects (Ruiz-Macias et al., 2021). The selection criteria of BGS Bright are shown in Fig. 2.1[1]. BGS Faint covers the fainter range $19.5 < r < 20.175$

---

[1]A straight-line locus of objects with a best fit photometric profile in the Legacy Survey of a PSF can be seen in the left-hand panel of Fig. 2.1. These objects are included in the selection as they are classified as galaxies by the BGS star-galaxy separator that makes use of higher resolution
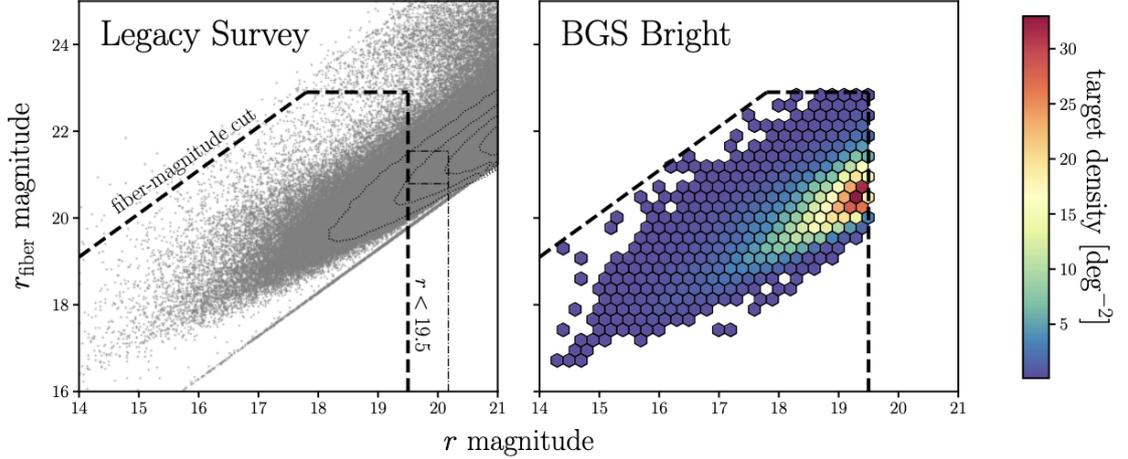
Figure 2.1: Targets for the BGS Bright sample are chosen based on the selection cuts described in Section 3.1 of Hahn et al. (2023) and an $r < 19.5$ magnitude cut. In the left panel, we show these cuts (based on fibre magnitude) and the $r < 19.5$ cut (black dashed) on the distribution of r vs. r_{fibre} magnitude for LS objects that pass our star–galaxy selection (grey). The contours mark the 11.7, 39.3, 67.5, and 86.4 percentiles of the distribution (dotted). We also include the r and r_{fibre} cuts for the BGS Faint sample (dotted–dashed). We impose selection cuts on BGS targets in order to minimise the number of spurious objects and mitigate any systematic effects that can affect galaxy clustering analyses. In the right panel, we present the target density of the BGS Bright targets (colour map). In total, we have 864 targets deg^{-2} for the BGS Bright sample. Taken from Hahn et al. (2023).
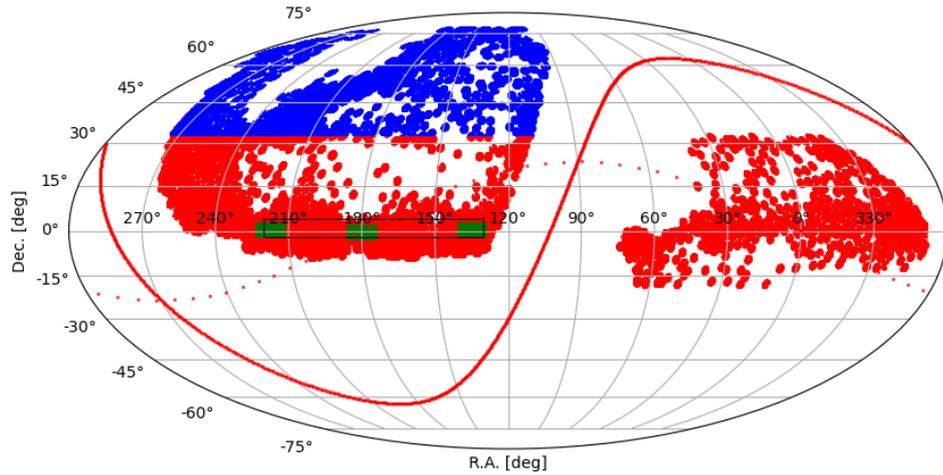
Figure 2.2: Sky plot of DESI Y1 BGS data. Blue shows the North photometry region while Red shows the South. The green region is the comparable GAMA equatorial fields (G9, G12, G15). The solid red line shows the galactic plane while the dotted red line shows the ecliptic plane. The black box shows the RA and DEC of the cone plot in Fig. 2.5.
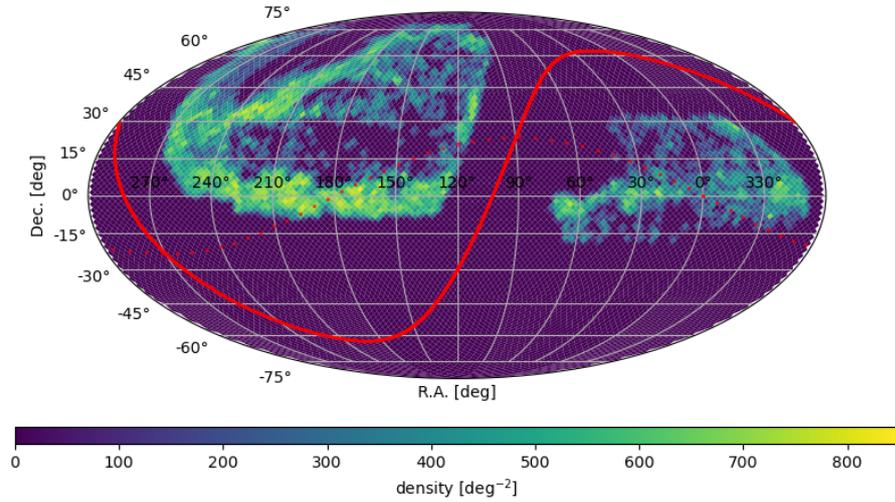
Figure 2.3: Sky plot of DESI Y1 BGS data showing the density of galaxies with good redshifts per degree squared. The differing density is due to certain areas of the sky receiving more passes, making them more complete.
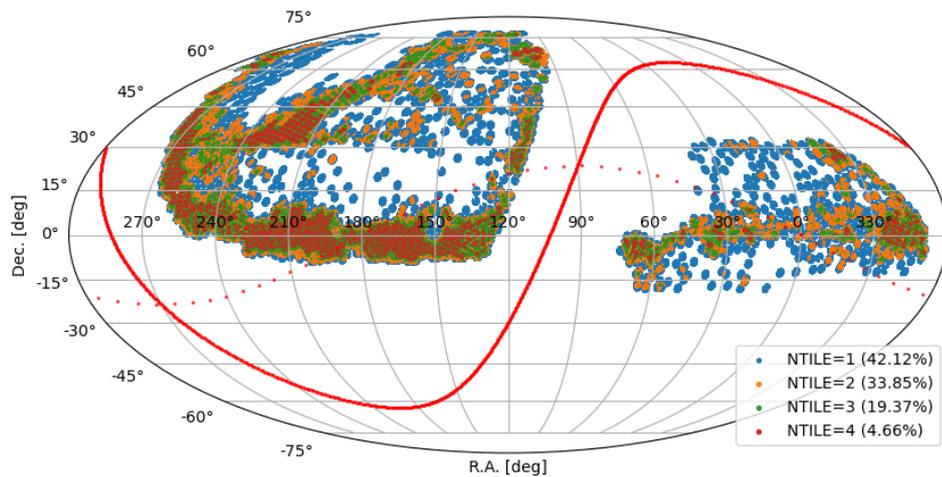
Figure 2.4: Sky plot of DESI Y1 BGS data categorised by NTILE - the number of times a galaxy could be targeted by a fibre. Large areas of Y1 BGS have only received one pass, indicating that there are regions of high incompleteness. Solid and dashed lines are the same as in Fig. 2.2.

but additionally has a colour-dependent fibre-magnitude limit with the purpose of removing objects for which the redshift success rate would be low (Hahn et al., 2023). DESI targets, including those in BGS, are selected based on applying photometric criteria to the photometric data taken with the DESI Legacy Surveys, BASS/MzLS in the North and DECaLS in the South (Dey et al., 2019). The target selection for BGS was developed in Ruiz-Macias et al. (2020, 2021), and the final target selection procedure is outlined in Hahn et al. (2023); Myers et al. (2023). As a result of the differing photometry in North and South, the DESI BGS Bright survey is magnitude-limited to $r < 19.54$ in the North and $r < 19.5$ in the South. These are extinction-corrected apparent magnitudes which are based on the SFD dust map (Schlegel et al., 1998). Both of these limits were chosen to produce the same surface density of targets, $\mathcal{N} \approx 800 \text{ deg}^{-2}$ in North and South. BGS Bright - which will be the focus of this paper - is similar to the depth of the Galaxy And Mass Assembly (GAMA) survey (Driver et al., 2011) and two magnitudes fainter than the SDSS Main Galaxy Survey (Strauss et al., 2002), but covers a much larger area of the sky at 14,000 deg^2. Moreover, BGS Bright has a median redshift of $z \approx 0.2$, double that of SDSS.

In this thesis, we make use of DESI DR1 data (otherwise known as Year One (Y1) data). This is the main DESI survey recorded after approximately one year of operations (as outlined in Section 1.4), with key DESI results released in 2024 (DESI Collaboration et al., 2024a,b). Our catalogue contains a total of 4,108,130 galaxies, shown in Fig. 2.2. Fig. 2.2 shows the large area covered by the North (in blue) and the South (in red). For comparision, the three equatorial fields in GAMA DR4 (G9, G12, G15) are plotted in green. As is typical for most surveys, BGS avoids the galactic plane due to stellar crowding. Fig. 2.3 and Fig. 2.4 show the same footprint with the sky density and NTILE values respectively, where NTILE is defined as the number of times that a target galaxy has been reachable by a fibre. For example, NTILE=2 means that the target has been within the patrol radius of a fibre in two separate passes. These plots indicate that there is a wide amount of variation in

---

GAIA data. Their number density is low as shown in the right hand panel of Fig. 2.1.
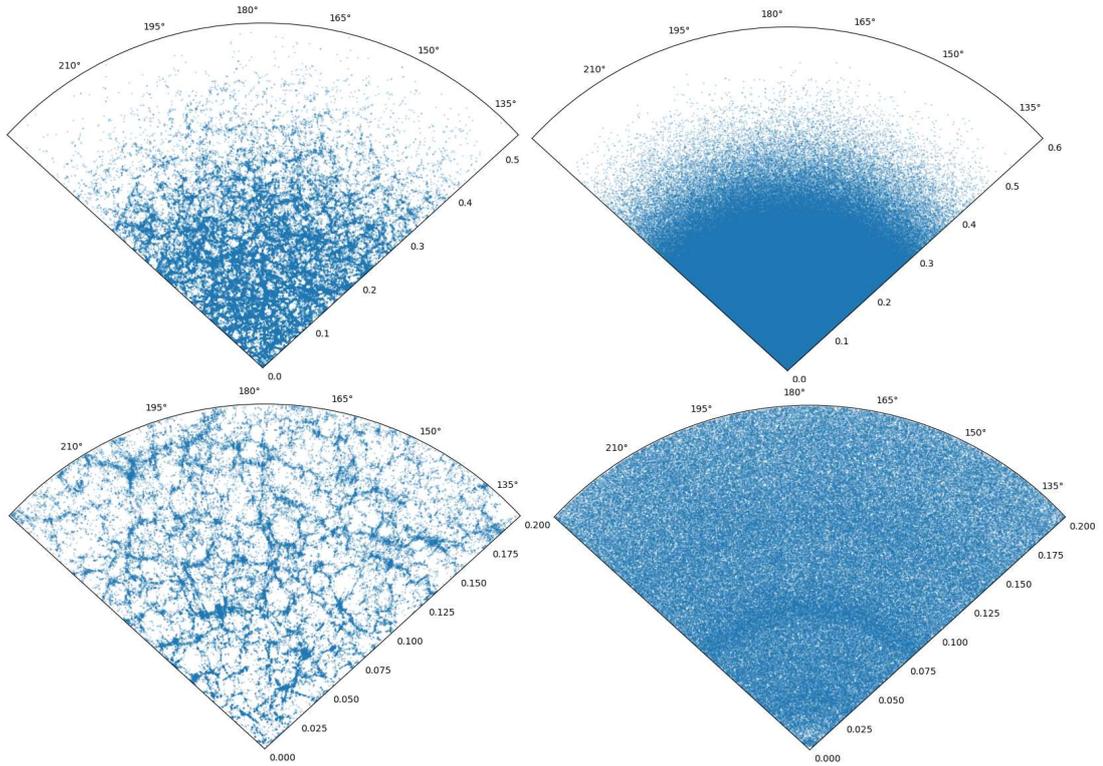
Figure 2.5: Top: Cone plots showing a slice of BGS Y1 for $0 < z < 0.6$ for the data catalogue (left) and the randoms catalogue (right). Bottom: Similar cone plots showing a zoomed-in slice of BGS Y1 for $0 < z < 0.2$, illustrating the LSS seen in DESI BGS. For all plots, the range is $130 < \mathrm{RA} < 220$ and $-2 < \mathrm{DEC} < 4$ (as shown in Fig. 2.2).
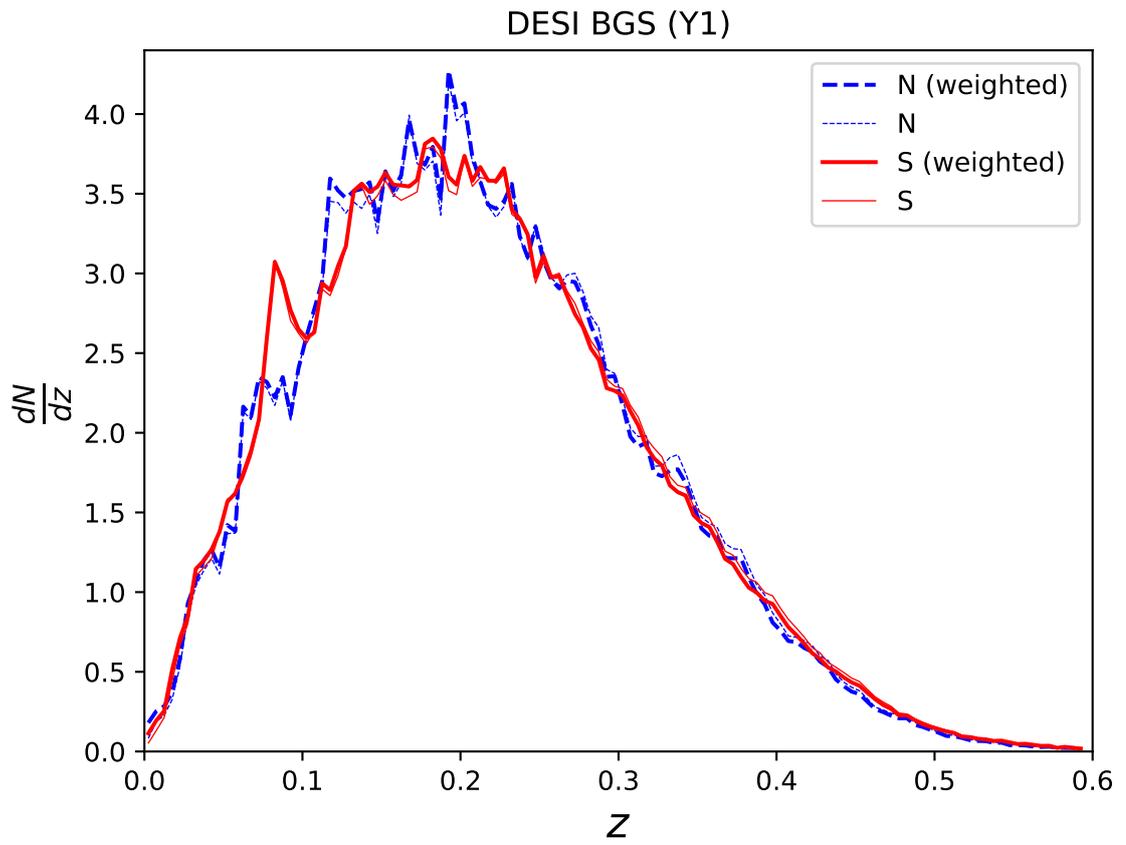
Figure 2.6: The redshift distribution of North and South Y1 BGS objects normalised by survey area. The weighted distributions make use of the weights defined in Eqn. 2.4.

the targeting, with some regions repeatedly targeted by multiple passes, and some regions targeted just once.

The Large Scale Structure apparent in BGS Bright is shown in Fig. 2.5, which plots a cone from $130 < \mathrm{RA} < 220$ degrees, $-4 < \mathrm{DEC} < 2$ degrees. The top-left plot displays the BGS survey over the entire redshift range considered in this thesis (from $0 < z < 0.6$), while the bottom-left plot zooms in (from $0 < z < 0.2$). From these plots, one can see a great deal of structure, with a large number of observable filaments and walls. The right-hand plots show a single realisation of the random catalogues, which covers the entire footprint (see Section 2.3 for further discussion of the randoms). In addition, we observe in Fig. 2.6 that North and South have similar $dN(z)/dz$ distributions as expected, although we observe the presence of some structure as expected.

In addition, the DESI Y1 catalogue also contains data from the Wide-Field Infrared Survey Explorer (WISE). WISE is a space-based infrared survey which mapped the whole sky with multiple passes (Wright et al., 2010). This was continued with the unWISE survey, which obtained deeper imaging through coadds (Schlafly et al., 2019). In particular, we use magnitudes from the w_1 band centred at 3.4μm with an angular resolution of 6.1", in order to later present a w_1-band LF. The observer-frame colours for different bands (including the w_1 band) are presented in Fig. 2.7 for both North and South.

2.2 Redshifts

In order to determine the redshifts of galaxies, DESI makes use of Redrock (Bailey et al., prep), a template-based classifier which classifies spectra as GALAXY, STAR or QSO and assigns a redshift to each spectrum based on χ^2 minimisation of a linear combination of Principal Component Analysis (PCA) basis spectral templates. This method also rules out unphysical PCA combinations in its fitting, such as negative spectra combinations. Redrock templates are constructed using SDSS spectra. Redrock additionally generates a $\Delta\chi^2$ value which is the difference in χ^2 between the best and second-highest likelihood peaks with different redshifts, which acts as a
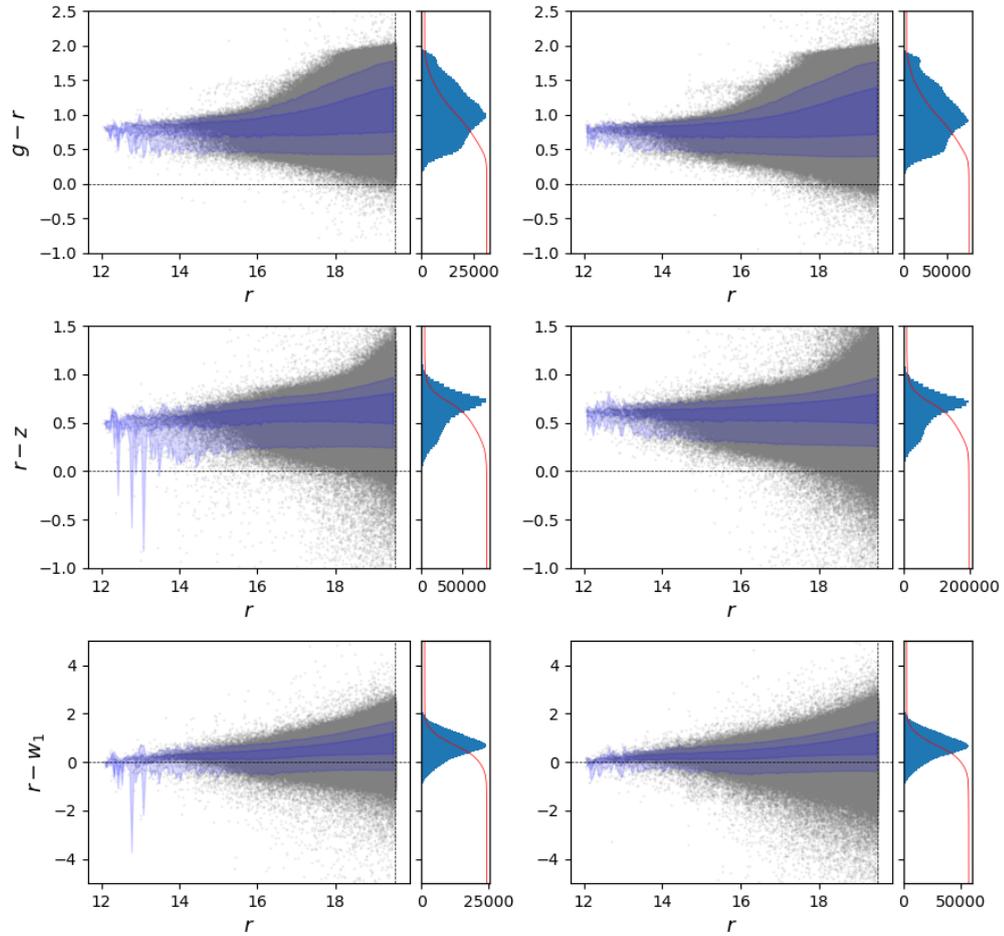
Figure 2.7: Observer-frame colour distributions for Y1 BGS Bright. Graphs on the left represent the North, while graphs on the right represent the South. The dashed vertical line shows the faint r-band limit of the survey ($r = 19.54$ for North, $r = 19.5$ for South). The red line plots the cumulative distribution function. The blue contours represent the 25-75th and 5-95th percentile ranges.

metric of confidence. Examples of these spectra and their Redrock fits are shown for BGS Bright and BGS Faint objects in Fig. 2.8.

A large number of spectra were evaluated by visual inspectors to assess the validity of the Redrock algorithm, as detailed in Lan et al. (2023). For the Main Survey (Y1), BGS Bright had an assessed Visual Inspection Quality $> 2.5^2$ for 99.6% of galaxies based on 1037 sources, indicating that Redrock returns robust redshifts for the vast majority of objects. Moreover, Lan et al. (2023) investigates the use of $\Delta\chi^2$ as a metric for assessing the reliability of the Redrock redshifts. For BGS Bright, the authors find a good redshift purity of 100% for $\Delta\chi^2 > 40$ in the VI sample of 2718 BGS target galaxies, indicating that this is a useful threshold for ensuring that sources have the correct redshifts. 97.2% of the VI BGS spectra have $\Delta\chi^2 > 40$.

2.3 Randoms

To define the selection function of the survey, a catalogue of randomly positioned points is generated over the DESI sky footprint with a number density of 2500 objects per square degree per set of randoms. Then, only the randoms that are reachable by a good fibre of an observed tile are kept in the catalogue. These randoms therefore completely map out the observed region of the survey. In particular, this footprint excludes the regions within the observed fields that were never reachable due to various factors - including but not limited to the gaps between the petals and broken fibre positioners. These randoms are additionally assigned redshifts and other galaxy properties, with the assignment done randomly from the selected galaxy sample. This means that for any subset of the randoms with the same selection as the data (e.g: with the same apparent or absolute magnitude cuts), the corresponding data will have the same $dN(z)/dz$ distribution. It should be noted that North

---

[2]The Visual Inspection Quality is the average score assigned by human assessers of the BGS spectra. The scoring criteria is defined as follows: Q=4: Confident classification with two or more secure features. Q=3: Probable classification with at least one secure spectral feature and continuum or many weak spectral features. Q=2: Possible classification with one strong (but unknown) spectral feature. Q=1: Unlikely classification with some signal but features unidentified. Q=0: No signal. Further details are found in Lan et al. (2023).
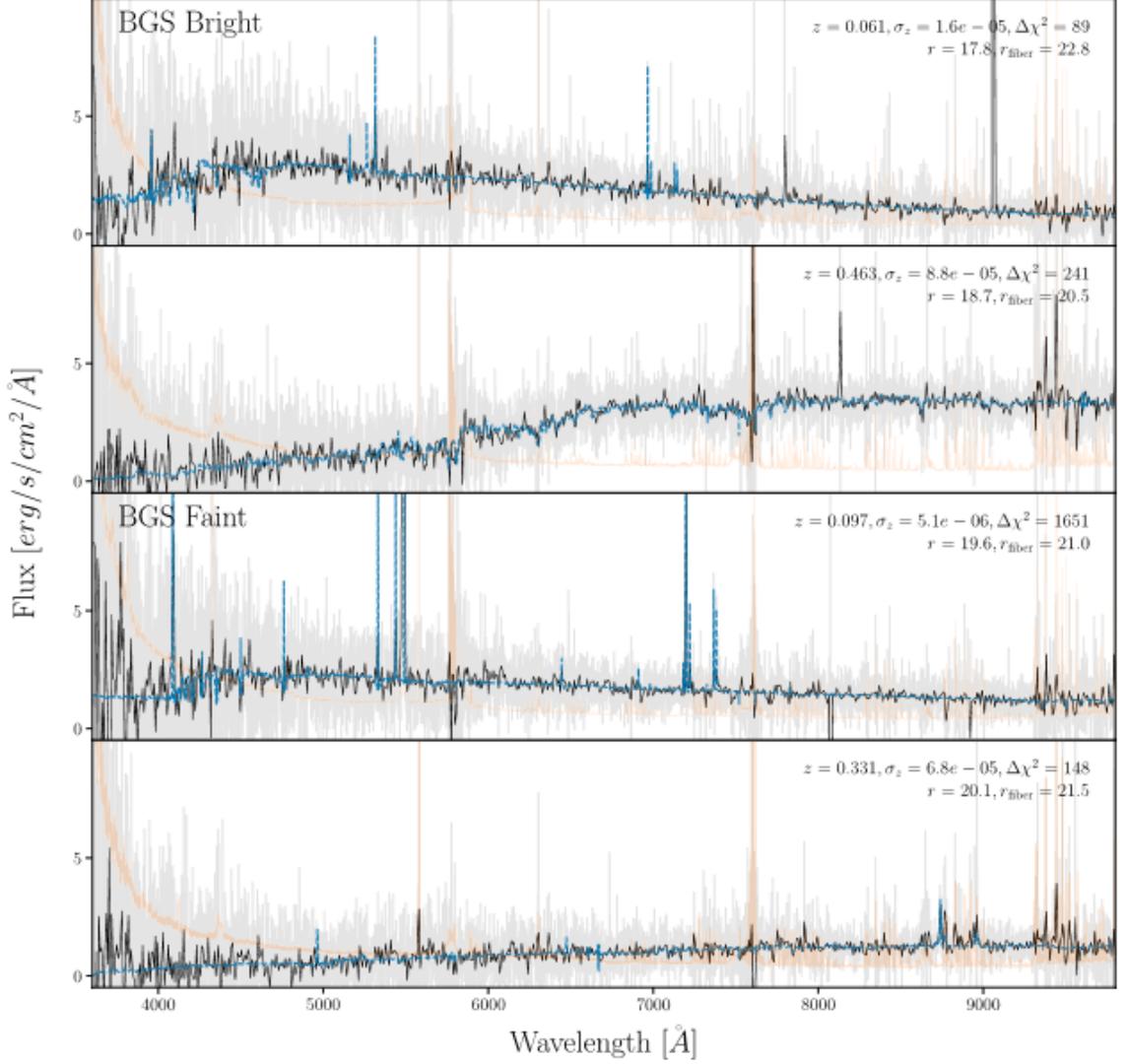
Figure 2.8: BGS galaxy spectra from the One-Percent Survey (grey). We present spectra of a blue and a red galaxy from the BGS Bright sample and a blue and a red galaxy from the BGS Faint samples (top to bottom panels). In each panel, we also plot the spectrum rebinned to a coarser wavelength grid (black) and the measured uncertainties (orange), and we include the best-fit Redrock template used to measure the redshift (blue). The redshift measurement, uncertainty, and $\Delta\chi^2$ from Redrock are included in the upper right corner, along with the r-band magnitude and fibre magnitude of the galaxy. Taken from Hahn et al. (2023)

.

and South property assignment is done separately from each other to allow for the difference in selection functions. As described in Section 2.1, DESI assigns fibres to obtain the spectra of galaxies. This means that a single pass will suffer from target incompleteness as there will be 'holes' in the survey, due to the gaps between petals, holes from where the Guiding, Focusing and Alignment (GFA) cameras are, and malfunctioning fibres. The distribution of the randoms fully takes this into account, as shown in Fig 2.5. We use the total number of randoms to quantify the total area of the sky observed, but also their distribution to map this incompleteness (see Fig 2.5, in addition to figures 2 and 3 in Ross et al. (prep)) so that we can make unbiased estimates of the local density when analysing the dependence of the LF on density.

2.4   k-corrections

Typically, absolute magnitudes are corrected by k-corrections to account for band-shifting effects, as the observed passband maps to different rest-frame passbands for galaxies at different redshifts (see Hogg et al., 2002, for a comprehensive overview of k-corrections). Specifically, in order to compare the photometric properties of galaxies at different redshifts, we need to transform their photometry to that of a fixed combination of reference frame and filter curve. For ease of comparison, we have chosen the SDSS r-band (and g-band) filter curves with a reference redshift of $z_{\text{ref}} = 0.1$ as adopted by Zehavi et al. (2005) and Loveday et al. (2012).

To compute absolute magnitudes and k-corrections, FastSpecFit (FSF) was developed to perform fast spectral synthesis and emission-line fitting of DESI spectra and broadband photometry (Moustakas et al., 2023). In particular, FSF works by simultaneously fitting model SEDs to a combination of the broadband photometry and the aperture-corrected DESI spectral photometry (Moustakas et al., 2023). This is shown in Fig. 2.9, which shows an FSF fit for an example BGS object from the SV3 survey. From these model SEDs, FSF absolute magnitudes and k-corrections have been calculated. However, these are not convenient for our purposes. Instead, we require full redshift-dependent k-correction functions so that we can calculate

V_{max} [3] as defined in Eqn. 3.3. As a result, we use the FSF k-correction catalogue as a basis to which we fit k-correction polynomials as described below.

With the BGS catalogues we provide $^{0.1}M_r$, the absolute magnitude in the SDSS r-band with reference redshift $z_{\mathrm{ref}} = 0.1$, defined by

$$^{0.1}M_r - 5\log_{10} h = m_{r_{LS}} - 5\log_{10}\left(\frac{d_L(z)}{h^{-1}\mathrm{Mpc}}\right) - 25$$
$$- \,^{0.1}k_{r_{LS}\to r}(z, \,^{0.1}(g-r)). \quad (2.1)$$

Here, the subscript r represents the SDSS r-band, r_{LS} represents the Legacy Survey band (BASS for the North; DECaLS for the South), $^{0.1}k_{r_{LS}\to r}(z, \,^{0.1}(g-r))$ represents the derived polynomial k-correction of the galaxy from the Legacy Survey observer frame r_{LS}-band to the rest frame SDSS r-band with reference redshift $z_{\mathrm{ref}} = 0.1$. $d_L(z)$ is the luminosity distance to the redshift z, determined using the same cosmology defined in the previous subsection. Optionally and in addition, an e-correction may be applied in order to account for the intrinsic luminosity evolution of a galaxy over time (not included in Eqn. 2.1; this is discussed in Section 3.2). Our methodology is flexible and can provide results for different reference-frame redshifts. For the sake of simplicity, for subsequent results and plots we use $z_{\mathrm{ref}} = 0.1$.

First, we create a rest-frame colour lookup table, where we generate a 2D histogram of observer-frame $g - r$ colour against redshift and compute the median rest-frame FSF colour in each pixel (Fig. 2.10). Using this table, and Cloud-In-Cell interpolation (Hockney & Eastwood, 1988), each galaxy in the Y1 DESI catalogue is assigned the median rest-frame colour corresponding with its observed colour and redshift. This assigned colour is then used to bin the galaxies into 7 rest-frame colour bins each containing an equal number of objects.

Each colour bin is split into a range of 50 redshift-bins and the median FSF k-correction is found in each bin. A least-squares polynomial fit is then performed

---

[3] V_{max} is the volume in which the galaxy can be re-positioned and still satisfy all the selection criteria to be included in the sample that is being analysed. For instance, there is a maximum redshift to which the galaxy could be relocated before its apparent magnitude is too faint for it to be included include in the sample.
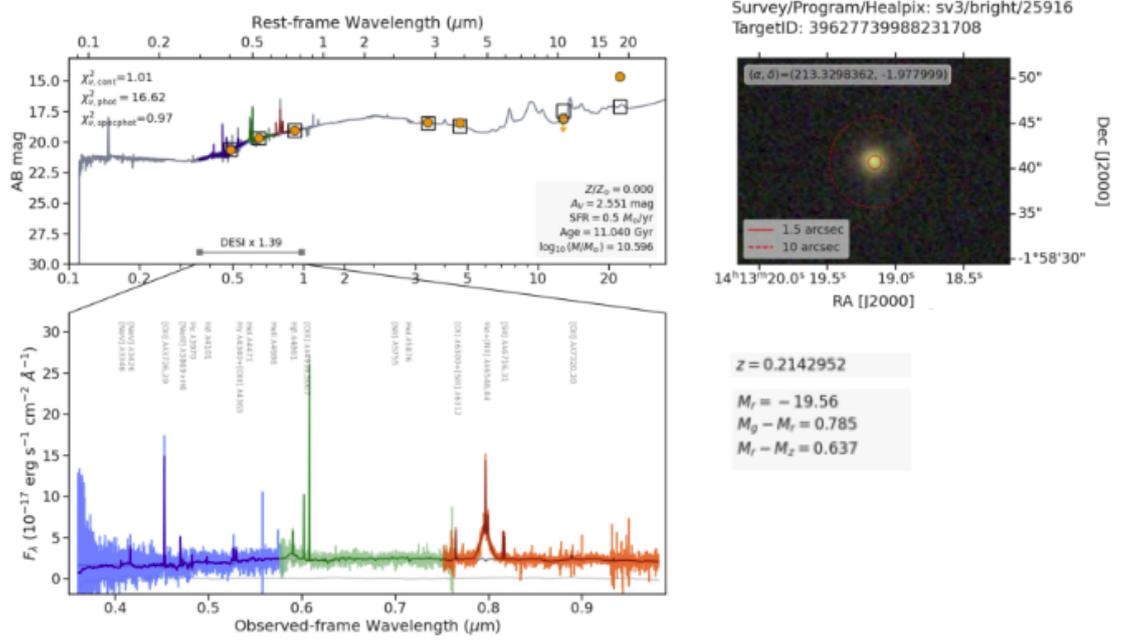
Figure 2.9: An example of the FastSpecFit (FSF) spectral fitting. Left: the derived SED model (grey) against the actual spectra (blue, green, red) where the open squares represent the predicted photometry in the various bandpasses and the circles show the actual measurements. Right: the BGS galaxy with apertures shown and quantities presented which have been derived from the FSF model. The spectrum captures only the light from the inner aperture shown in the top right panel defined by the fibre. This is then rescaled by the flux ratio between the outer and inner apertures. Various quantities are then extracted from this model fit, including the absolute magnitudes and k-corrections. Adapted with permission from John Moustakas.

on these median k-corrections to find a 7th-order polynomial for each colour bin. The choice to fit a 7th-order polynomial was motivated by a number of reasons. We desired a polynomial fit that would fit well to the data over the entire redshift range. Moreover, we required that the polynomials did not cross over at any point to avoid having non-unique k-corrections. Previous studies (McNaught-Roberts et al., 2014) had conducted the fitting using a 4th-order polynomial, however, we determined that this did not adequately fit low-redshift objects and led to a slight disparity between North and South objects. We additionally investigated the use of splines to neatly fit to the median bins separately for low-redshift and high-redshift regions; however, this proved to be too flexible, often causing the polynomials to cross at high-redshift. We settled on the 7th-order polynomial as this fits our data well including at low-redshift, but does not seem to be prone to over-fitting.

We confirm that the median values match very closely to the fitted polynomials across all redshifts. This is seen for the South k-correction polynomials in Fig. 2.11. The North and South are modelled separately, with the polynomial fits compared in Fig. 2.12. From this plot, we see that the North and South curves are slightly different reflecting the differences in the North/South photometry. Furthermore, we note that there is a pinch point at about $z = 0.14$ which corresponds to the redshift at which the central wavelength of the DESI r-band filter best matches with that of the SDSS r-band filter at a reference redshift $z_{\mathrm{ref}} = 0.1$. This corresponds to the 3.73% shift in the effective wavelength between the SDSS r-band filter (6205.83Å) and the BASS r-band filter (6437.79Å), where we note that (1+0.141) / (1+0.1) = 3.73%.

We conducted additional tests to ensure that these k-correction polynomials remain largely invariant to choices in colour and redshift bin size. For example, choosing a larger number of colour-bins has a small but negligible impact on the k-correction polynomials at the colour extremes.

Furthermore, we derive the reference-frame $g - r$ colour and thus also the g-band k-correction using Eqn. 2.2,

$$^{0.1}k_{g_{LS}\to g}(z, {}^{0.1}(g-r)) = {}^{0.1}k_{r_{LS}\to r}(z, {}^{0.1}(g-r))$$
$$+ {}^{0.1}(g-r) - \left({}^{0.1}M_g - {}^{0.1}M_r\right) \quad (2.2)$$

Here, we note that $M_g - M_r$ (the rest/reference frame colour) comes from the lookup table and is therefore a function of $g - r$ and z. This is shown in Fig. 2.12. For these polynomials, we then utilise a cubic interpolation scheme between these colour-bin polynomials when assigning k-corrections.

As validation of our k-correction fitting, we confirm that there is a negligible difference between the M_r given by FSF, and the M_r that we calculate in Eqn. 2.1 using our k-correction polynomials. This is shown in Fig. 2.13, where the median difference is < 0.01 magnitudes for the two colour bins plotted. Moreover, this difference remains well-bounded across the entire redshift range, suggesting that our results should be unaffected by the choice to use polynomial fits rather than the direct FSF k-corrections.

Fig. 2.14 presents the $^{0.1}(g-r)$ rest-frame colour distribution and the r-band absolute magnitude distribution. We observe that there exists a slight offset in the North and South colour distributions. We attribute this to a possible small error (0.01 magnitudes) in the calibration of the filter curves.

In addition, we note that we can calculate k-corrections for $z_{\text{ref}} = 0.0$ from $z_{\text{ref}} = 0.1$. We present additional results for k-correction transformations in Appendix A, including the derivation of the relationship between k-corrections of different reference redshifts.

We additionally note that we conducted a similar method of polynomial fitting for k-corrections for EDR, that we outline in Section 4.2.3 of DESI Collaboration et al. (2023). We emphasise that this EDR methodology of polynomial fitting is different to the final version used here for the Y1 data. In particular, those EDR k-correction polynomials are 4th-order polynomial fits to GAMA DR4 data. Whilst this was valuable at the time (prior to the release of FSF k-correction data), this assumes that there is no difference in photometry - i.e: North and South objects with
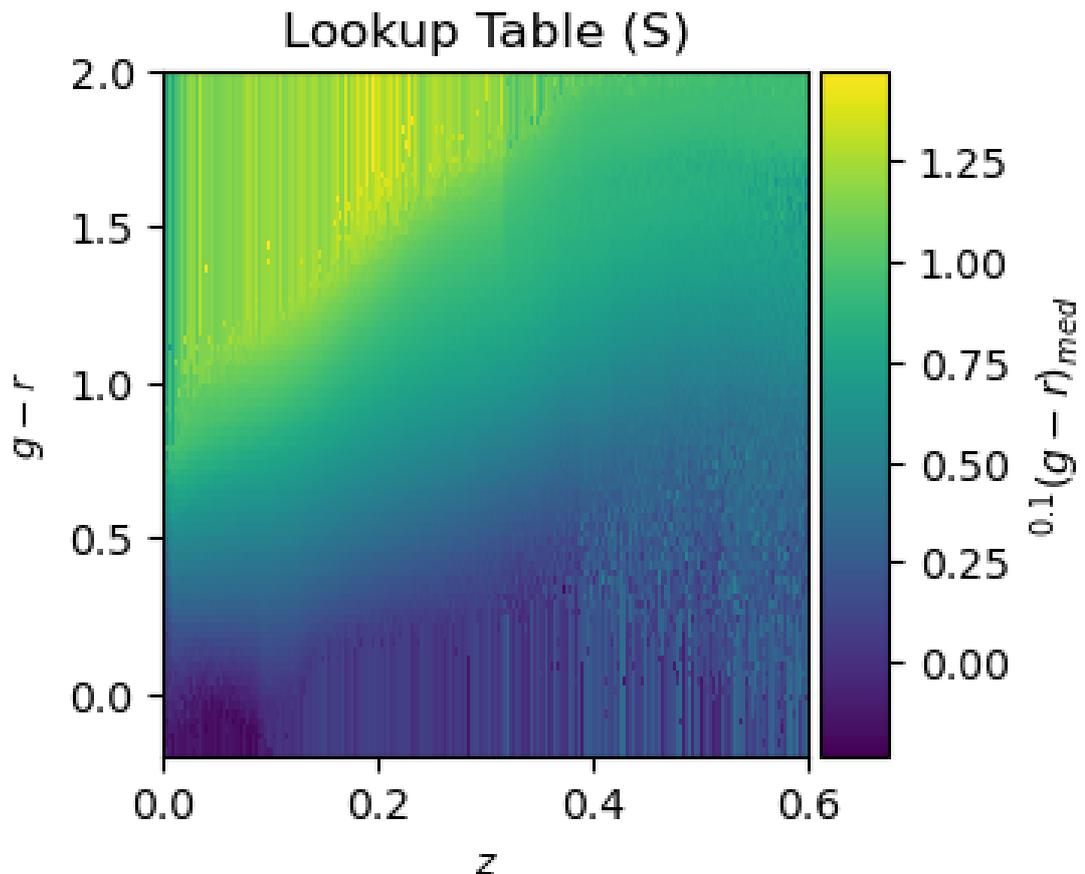
Figure 2.10: The rest-frame lookup table for South as a function of redshift and observer-frame $g - r$ colour. Each pixel value is the median rest-frame colour, $^{0.1}(g-r)_{\mathrm{med}}$ for the galaxies in that pixel bin, based on the FSF rest-frame $^{0.1}(g-r)$ colours. For pixels with no galaxies, that pixel is assigned the value of the closest pixel at that redshift range. A separate lookup table is generated and used for the North.

the same z and $g - r$ values receive the same k-correction. Here, we have updated the methodology to make use of a colour lookup table and to directly use DESI BGS galaxies in order to take account of the DESI photometry and its differences in the North and South regions.

2.5 Incompleteness Corrections

To construct galaxy luminosity functions, it is important to correct for incompleteness within the DR1 dataset, including systematic effects in the input catalogue, target incompleteness and redshift incompleteness.
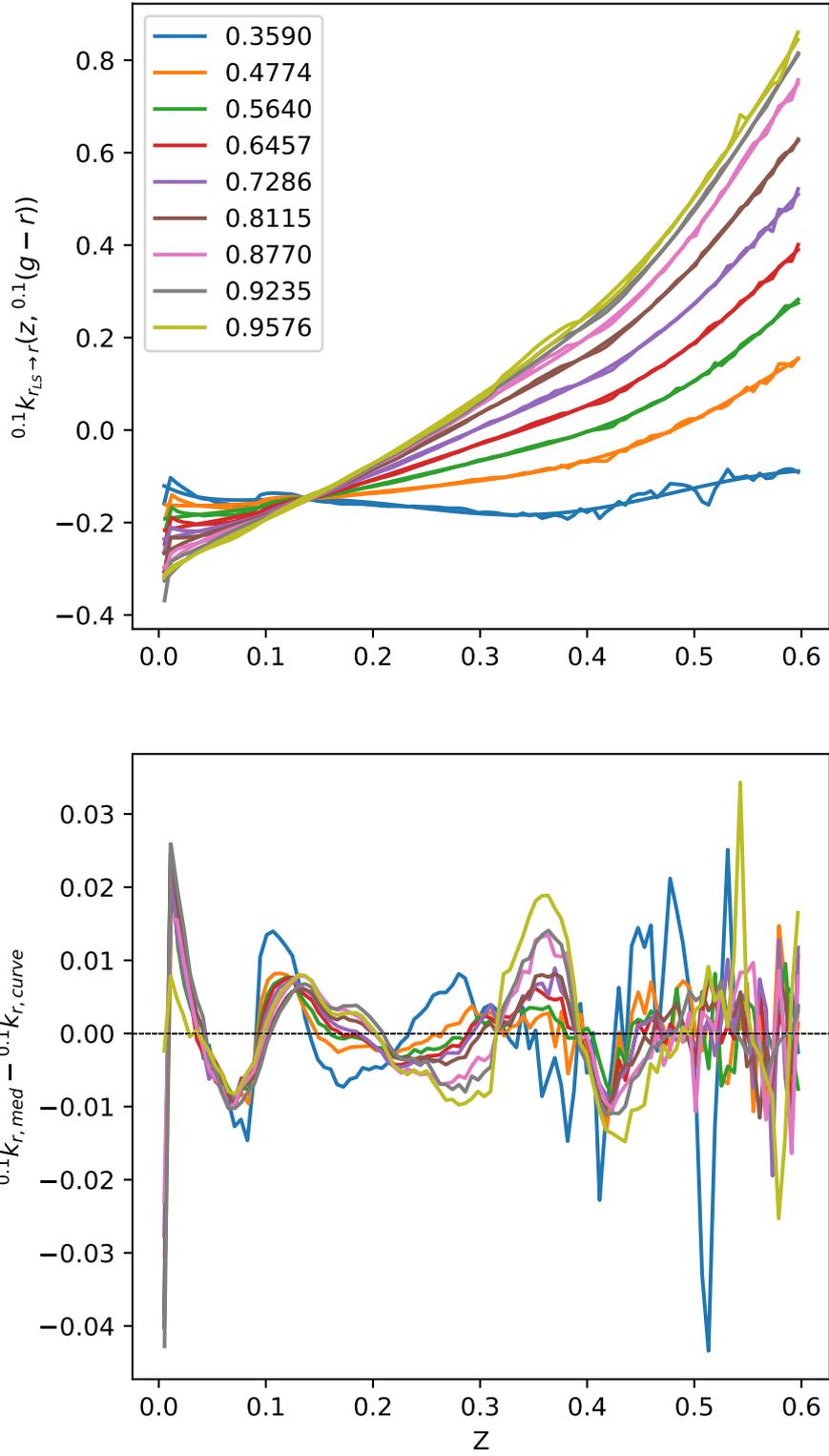
Figure 2.11: Top: Polynomial fits to the medians in each rest-frame colour bin for the South. The legend indicates the median rest-frame colour in that rest-frame colour bin. Bottom: the residual plot, showing that these polynomials are good fits to the medians. Here, $^{0.1}k_{\mathrm{r,med}}$ is shorthand for the $^{0.1}k_{\mathrm{r_{LS}\to r}}(z, ^{0.1}(g-r))$ value for each bin, while $^{0.1}k_{\mathrm{r,curve}}$ is the value derived from the polynomial fit.
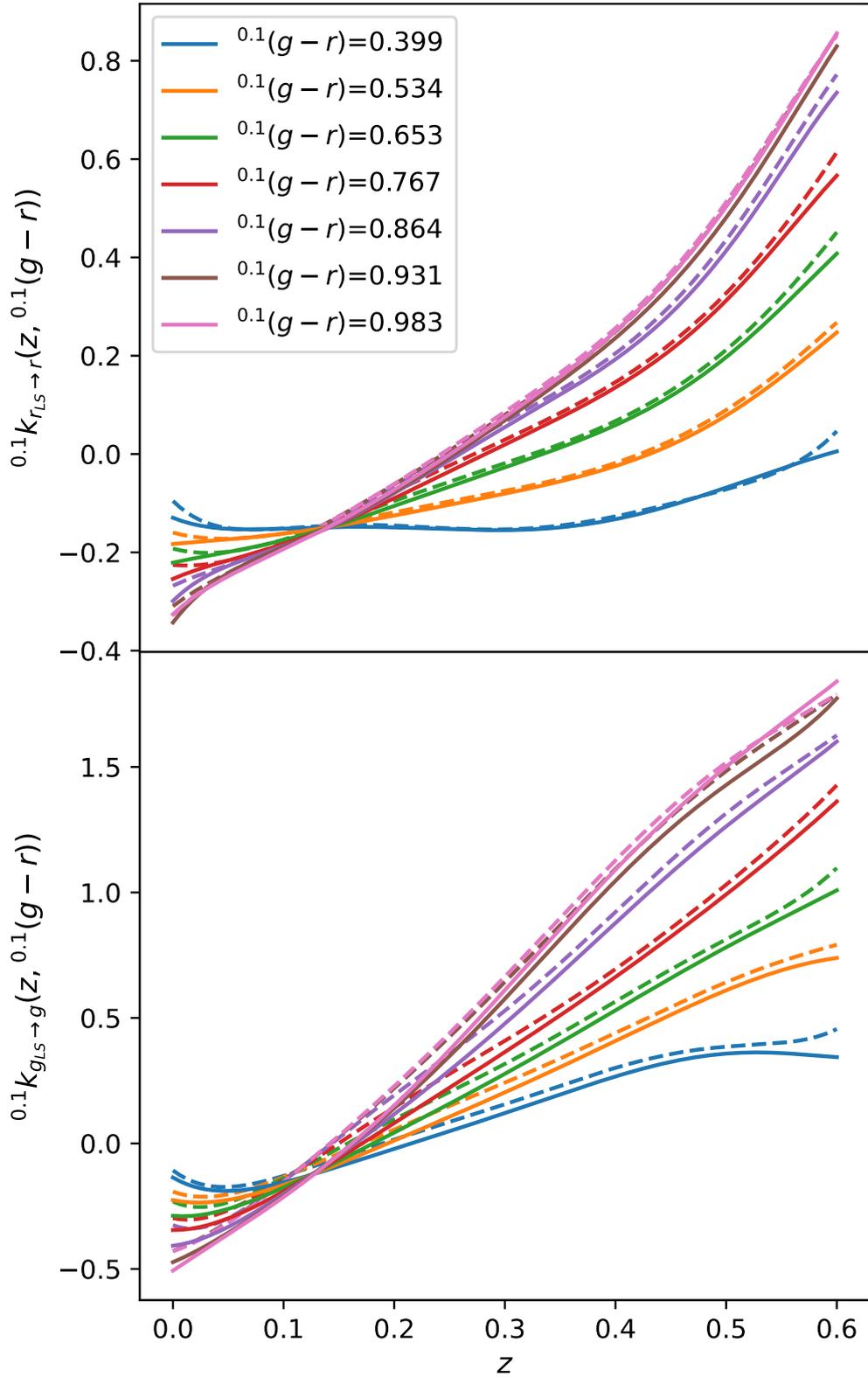
Figure 2.12: Top: k-correction polynomials to the SDSS r-band with $z_{\mathrm{ref}} = 0.1$ to the DECaLS r-band (South, solid line) and the BASS/MzLS r-band (North, dashed line). Bottom: The r-band polynomials transformed to g-band k-correction polynomials using Eqn. 2.2.
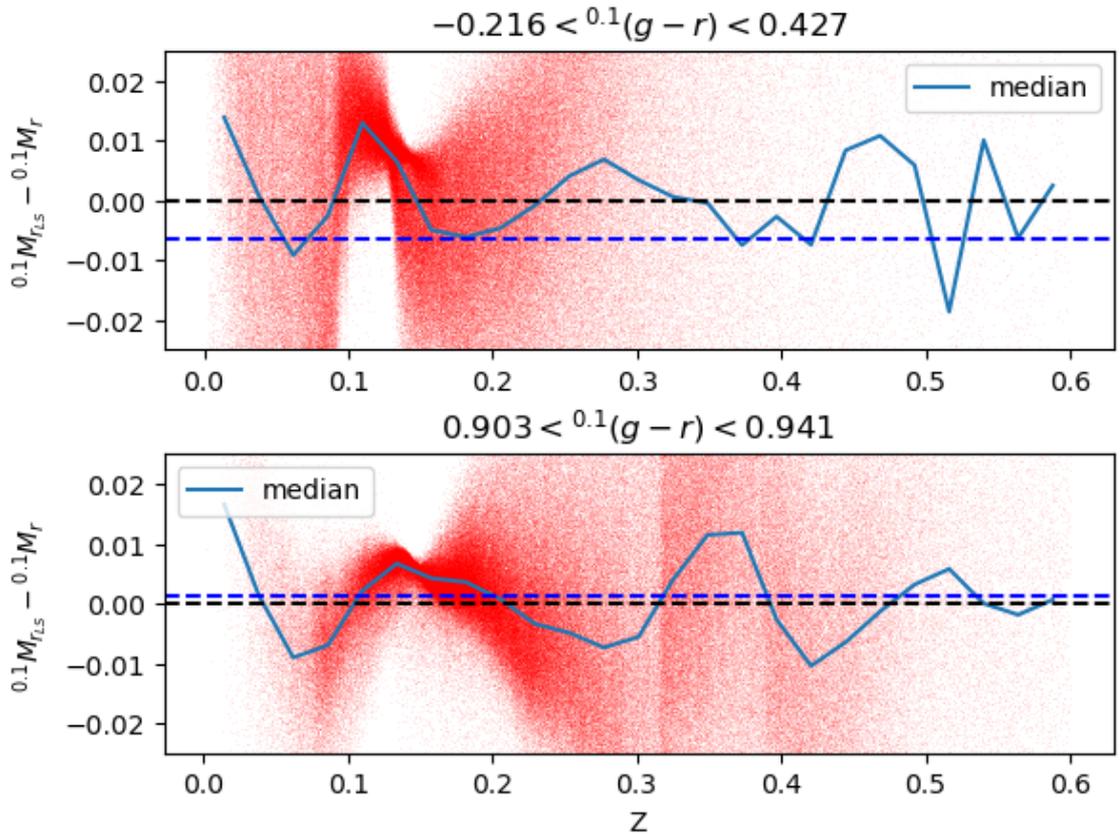
Figure 2.13: The difference of the direct FSF r-band absolute magnitude and our derived r-band absolute magnitude as a function of redshift using our k-correction model for the South. The solid blue line plots the median in different redshift bins, while the dashed blue line is the overall median across all redshifts. The dashed black line is a reference line at zero. Similar plots exist for the North.
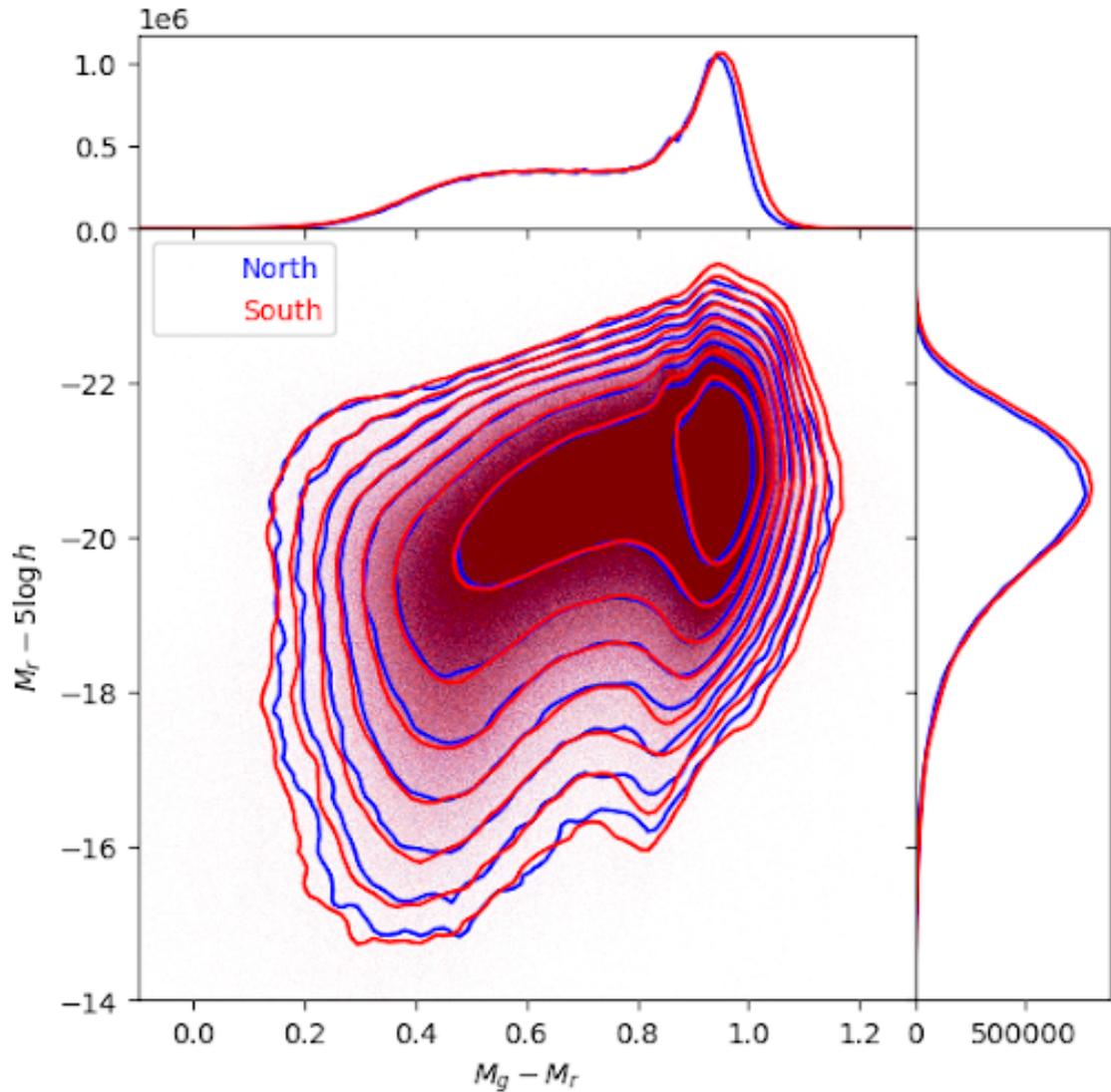
Figure 2.14: Rest-frame colour vs. absolute magnitude distributions in North and South. Contours are plotted representing number density, with each successive contour a factor of 2 larger in number density (starting at 160 objects per bin). The top histogram shows the $^{0.1}(g-r)$ colour distribution for North and South. The right histogram shows the r-band absolute magnitude distribution. Both histograms have been normalised by sky-area to adjust for the fact that South is approximately double the size of North.

35

2.5.1 Systematic Weights

Typically, a systematic weight (w_sys, denoted WEIGHT_SYS in DESI catalogues) is used to account for target density fluctuations due to imaging conditions and foregrounds. Specifically, this weight corrects for unphysical correlations of the target density with dust extinction, stellar density, photometric depth and HI maps (DESI Collaboration, private correspondence). In clustering analyses, these weights are important as they account for variation across the sky which impacts the clustering. However, the LF will only be sensitive to the mean value of the systematic weights. We observe that these weights are close to unity (with a mean weight close to unity) and make a negligible difference to our results (shown later on in Section 3.3, Fig. 3.10). Moreover, these weights have only been calculated for specific sections used in the LSS analysis, and not as a function of apparent magnitude (which would be necessary if we were to use them). As such, we have chosen to ignore this systematic weight in this paper.

2.5.2 Target Completeness Weights

An additional source of incompleteness is target incompleteness, which in this case is a correction factor to account for targets that were not observed. In tiles for which there has only been a single pass, there will be target incompleteness as only one object can be targeted within the unique patrol region of each fibre. This means that not all targets will be assigned to fibres in regions of high target density. Furthermore, fibres cannot be placed arbitrarily close to each other due to the physical and mechanical constraints of the fibres. This limit can also lead to observable targets not being assigned a fibre. The footprint of the survey is determined by the area of sky that is the union of sky reachable by good fibres on observed fields. DESI conducts multiple passes to reduce target incompleteness. However, in regions covered by multiple tiles, the target incompleteness is reduced but is rarely completely removed. The target completeness weight (w_comp, denoted WEIGHT_COMP in DESI catalogues) is defined in DR1 as the inverse of f_TLID, i.e: the inverse of the number of targets sharing the same same tile location ID (TILELOCID) (Ross et al., prep).

w_{comp} is therefore a weight that compensates for the observable targets that were not assigned a fibre by gathering their weights and applying them to a neighbouring object that was assigned a (working) fibre.

2.5.3 Redshift Weights

In addition, not all targeted galaxies may receive redshifts (e.g: due to a failure of Redrock to fit a model spectra with confidence). Whilst the DESI LSS catalogues do calculate a redshift completeness weight (WEIGHT_ZFAIL), this weight is designed to correct to a uniform sample (with uniform incompleteness) rather than to a complete sample. As such, this weight is adequate for clustering but less useful for LFs. We choose instead to make our own direct estimate of the redshift completeness weights ($w_{z,\mathrm{new}}$). For comparison, we denote the old redshift weight (WEIGHT_ZFAIL) as $w_{z,\mathrm{old}}$.

To do this, we take the full LSS catalogue and define a 'zgood' subset that is subject to the following selection cuts:

1. DELTACHI2 ($\Delta\chi^2$) > 40.

2. ZWARN = 0

3. $0.002 < z < 0.6$

These selection cuts ensure that the redshift is reliable and excludes the redshift range contaminated by stars.

The overall redshift completeness of the survey is simply the number of objects in 'zgood' over the number of objects in the full observed catalogue. We calculate the incompleteness in bins based on the r-fibre magnitude and the template signal to noise squared for BGS (TSNR2BGS, defined in Guy et al., 2023). TSNR is defined by

$$\mathrm{TSNR}^2 = \sum_i \frac{T_i^2}{\sigma_i^2} \langle (\delta F)^2 \rangle. \tag{2.3}$$

Here, δF is defined as the difference between a template spectrum and the median filtered version of that same spectrum. The function of $\langle (\delta F)^2 \rangle$ is shown in figure 43

of Guy et al. (2023). T_i represents a calibration coefficient, and σ_i is the flux measurement uncertainty. TSNR2 is defined for all DESI galaxy types; TSNR2BGS specifically uses BGS templates. Importantly, TSNR2BGS is proportional to the effective exposure time, and does not depend on the photometric properties of a specific target but simply depends on the target class template. As such, we select r_{fibre} and TSNR2BGS as variables for this calculation as we expect redshift completeness to decrease for fainter r_{fibre} and smaller TSNR2BGS.

This acts as a lookup table so that galaxies may be assigned a redshift completeness weight. The weight is given by the inverse of the completeness value given by the lookup table, where we make use of Cloud-In-Cell interpolation to smooth over the pixels (Hockney & Eastwood, 1988). This result is shown in Fig. 2.15.

The total weight calculated for each galaxy is given in Eqn. 2.4. We confirm that our total weights are higher on average than the total weight ('WEIGHT') provided in the DESI catalogue.

$$w_i = w_{\text{comp},i} \cdot w_{\text{z, new},i} \tag{2.4}$$

In particular, we notice that the redshift weight ($w_{z,\text{new}}$) that we derive is significantly higher for certain galaxies than the old redshift weight, as seen in Fig. 2.16. We find that $\bar{w}_{z,\text{old}} = 1.0010$, while $\bar{w}_{z,\text{new}} = 1.0112$. However, Fig. 2.16 shows that there are a number of objects that have significantly altered total weights as a result of using this new redshift weight.
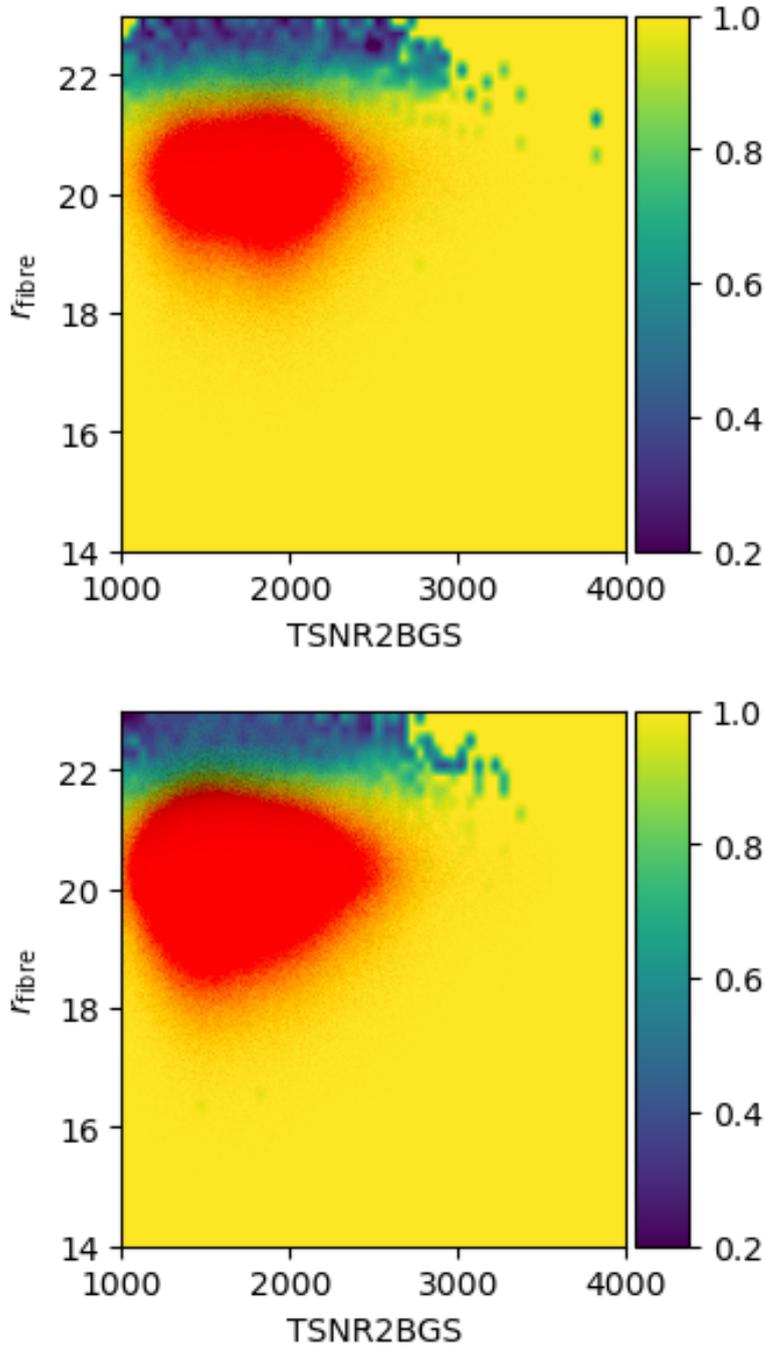
Figure 2.15: The empirical redshift completeness as a function of the fibre magnitude, r_{fibre}, and the expected spectral target signal-to-noise ratio for a fiducial BGS source, TSNR2BGS. The North (top) and South (bottom) completeness is plotted with the galaxies for that region plotted as red points. The majority fall where the redshift completeness is very high. We note the difference in the distribution of the galaxies between North and South. To define a weight, $w_{\text{z, new}}$, to correct for this incompleteness we interpolate the binned incompletness using the Cloud-In-Cell technique and take its inverse.
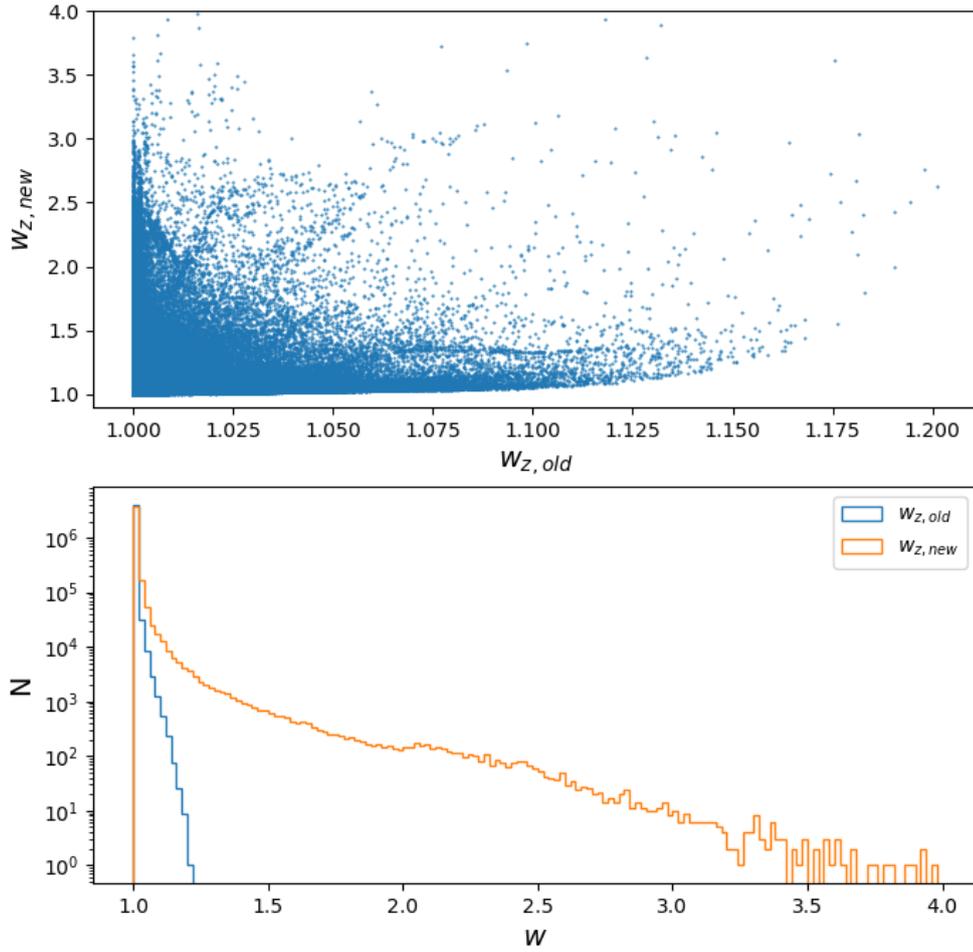
Figure 2.16: Top: The value of DESI WEIGHT_ZFAIL ($w_{z,\mathrm{old}}$) vs. the newly calculated $w_{z,\mathrm{new}}$, demonstrating that there exists a significant number of galaxies that have a much higher redshift completeness correction. Bottom: The distribution of the old and new total weights (w_{old} and w_{new}), where w_{old} incorporates $w_{z,\mathrm{old}}$ and w_{new} incorporates $w_{z,\mathrm{new}}$.

CHAPTER 3

Luminosity Functions

In this Chapter, we introduce the methodology for finding the global luminosity function (LF) from DESI BGS. A method of calculating evolutionary corrections (e-corrections) for each galaxy calibrated using the $V/V_{\max}$ distribution is presented. We present a number of bivariate and univariate LFs in the g, r, z and w_1 bands, and include a thorough analysis of Poisson and jackknife errors, in addition to various LF fits, such as the Schechter function. Whilst we typically make use of the $1/V_{\max}$ LF method in this Chapter, we additionally validate that our results are invariant with respect to method employed by also constructing density-corrected $1/V_{\max}$ LFs ($V_{\mathrm{dc,max}}$) and Stepwise Maximum Likelihood (SWML) LFs. This is an important validation step as we desire our LF results to be largely invariant of the LF estimator used. Furthermore, we present colour-dependent LFs to better understand the global LFs. We investigate potential causes of a difference in the bright-end of the LF for North and South. We also conduct further investigations to better understand an upturn in the LF at faint magnitudes, including an assessment of the DESI imaging.

3.1 Global LF Methods

3.1.1 $1/V_{\mathrm{max}}$ estimator

In this thesis, we primarily make use of the V_{max} estimator as outlined in Schmidt (1968). In the V_{max} method the LF estimator is

$$\phi(L)dL = \sum_{i=1}^{N} \frac{w_i W(L - L_i)}{V_{\mathrm{max}}(L_i)} \tag{3.1}$$

where $V_{\mathrm{max}}(L_i)$ is defined as the maximum volume over which the galaxy can be seen in the survey, w_i represents other included weights. The $W(L - L_i)$ represents a binning function that is defined as

$$W(L - L_i) = \Theta(L_i - L + dL/2) - \Theta(L + dL/2 - L_i) \tag{3.2}$$

where Θ is the Heaviside step function.

Formally, V_{max} may be defined as

$$V_{\mathrm{max},i} = \frac{4}{3}\pi f_{\mathrm{sky}}[d(z_{\mathrm{max},i})^3 - d(z_{\mathrm{min},i})^3] \tag{3.3}$$

where $d(z)$ represents the comoving distance of a galaxy at redshift z, and z_{min} and z_{max} represent respectively the minimum and maximum redshift at which a galaxy could be observed given its r-band apparent magnitude and the redshift limits of the sample being considered. f_{sky} is the fraction of the sky that has been surveyed by DESI. This is computed using the number of objects in the matched random catalogue, given the known surface density of the randoms.

By down-weighting each object by its maximum detection volume, this method corrects for the issue that intrinsically faint objects are only detected in a small volume, leading to the preferential detection of intrinsically bright objects. Additionally, the $1/V_{\mathrm{max}}$ estimator has the advantages of not assuming a functional form for the LF and automatically having the correct normalisation. The disadvantage of this estimator is that it assumes all sources follow a uniform spatial distribution. This can result in distortion in the case of overdense or underdense regions (Efs-

tathiou et al., 1988). We note that for deep wide-area surveys such as DESI, higher redshift slices fairly sample both overdense and underdense regions. As such, this distortion is only a potential problem for very faint luminosities as such galaxies can only be detected at very low redshift, where the volume is very small.

3.1.2 SWML estimator

In addition, to the V_{max} estimator, we make use of the Stepwise Maximum Likelihood (SWML) estimator. Broadly, likelihood methods consider the probabilities of observing a galaxy at redshift z_i and magnitude M_i (or luminosity L_i) within a magnitude-limited survey. This can be used to construct a likelihood function as

$$\mathcal{L} = \prod_i p_i, \tag{3.4}$$

where

$$p_i = \frac{\Phi(L_i)}{\int_{L_{\mathrm{min}}(d_i)}^{L_{\mathrm{max}}(d_i)} \Phi(L)dL}. \tag{3.5}$$

Parametric estimators will typically assume a functional form of $\phi(M)$ (such as a Schechter function) with some set parameters, and then will maximise the likelihood with respect to those parameters. This is the basis for the STY method (Sandage et al., 1979). We instead choose to make use of the non-parametric method set out in Efstathiou et al. (1988), otherwise known as the Stepwise Maximum Likelihood (SWML) method. This method does not require the assumption of a functional form of the LF. Moreover, the SWML estimator is unbiased by density fluctuations if one assumes that the shape of the LF is independent of the density (unlike the V_{max} method, which can be biased by density fluctuations). In the past, the SWML estimator has been preferred for datasets with smaller sample sizes as it acts to smooth the effect of density perturbations on the LF.

$\Phi(M)$ is specified by the values of a uniformly spaced histogram across the magnitude interval, with the histogram values being determined by maximising the

likelihood in equation (3.4). This yields

$$\Phi(M) = \frac{\sum_{i=1}^{N} W(M_i - M_k)}{\sum_{i=1}^{N} \frac{H(M_i - M_k)\Delta M}{\sum_{j=1}^{N_P} \phi_j H(M_j - M_{\min})\Delta M}} \tag{3.6}$$

where

$$W(x) = \begin{cases} 1, & \text{if } -\frac{\Delta M}{2} \leq x \leq \frac{\Delta M}{2}. \\ 0, & \text{else.} \end{cases} \tag{3.7}$$

is the binning function and

$$H(x) = \begin{cases} 0, & \text{if } x \leq -\frac{\Delta M}{2}. \\ \frac{x}{\Delta M} + \frac{1}{2}, & \text{if } -\frac{\Delta M}{2} \leq x \leq \frac{\Delta M}{2}. \\ 1, & \text{if } x \geq \frac{\Delta M}{2}. \end{cases} \tag{3.8}$$

is related to its integral by $H(x) = \int_0^x W(x')dx'/\Delta M$.

3.1.3 $V_{\text{dc,max}}$ estimator

Finally, we make use of a density-corrected $V_{\max}$ method (hereafter called $V_{\text{dc,max}}$), based on a method from Cole (2011). In this method, the effective volume of a galaxy may be calculated with Eq. 3.9

$$V_{\text{dc,max},i} = \int_{z_{\min,i}}^{z_{\max,i}} \Delta(z)\frac{dV}{dz}dz = \sum_j \Delta V \Delta_j G(V_j) \tag{3.9}$$

where the sum is over volume shells, the overdensity parameter in each shell $\Delta_j = 1$ in the first iteration, and G is a binning function that corresponds to:

$$
G(V_j) = \begin{cases} 0, & \text{if } V_j - \Delta V/2 > V_{\text{max},j}. \\[2mm] 1, & \text{if } \frac{\min(V_j+\Delta V/2, V_{\text{max},j}) - \max(V_j - \Delta V/2, V_{\text{min},j})}{\Delta V}. \\[2mm] 0, & \text{if } V_j + \Delta V/2 < V_{\text{min},j}. \end{cases} \tag{3.10}
$$

where we use the volume $V = \int (dV/dz)dz$ rather than redshift as the radial coordinate. For subsequent iterations, we calculate

$$
\Delta_j = \frac{N_j}{N_{\text{exp},j}} \tag{3.11}
$$

where

$$
N_{\text{exp},j} = \sum_i \frac{G(V_{\text{min},i} < V_j < V_{\text{max},i})\Delta V}{\Delta_j V_{\text{max},i}}. \tag{3.12}
$$

Here, $N_{\text{exp},j}$ is the expectation value for the number of galaxies that we would expect in each volume element if they were uniformly distributed in space given their individual $V_{\text{min},i}$ and $V_{\text{max},i}$.

From this, the LF for each iteration may be calculated as:

$$
\phi = \frac{1}{\Delta M} \sum_i \frac{w_i}{V_{\text{dc,max}}} \Theta(M_i; \Delta M) \tag{3.13}
$$

If $\Delta = 1$, this equation reduces to Eqn. 3.1. We additionally enforce for each iteration the constraint that the volume averaged overdensity $\langle \Delta_j \rangle = 1$.

3.1.4 Error Estimation

We calculate the Poisson errors on the $1/V_{\text{max}}$ LF using

$$
\frac{\Delta\phi(L)}{\phi(L)} = \sqrt{\frac{\sum_i (w_i^2) W(L - L_i)}{(\sum_i w_i)^2}}. \tag{3.14}
$$

Here we note that if all weights are unity, then this expression reduces to the standard $1/\sqrt{N}$ fractional error.
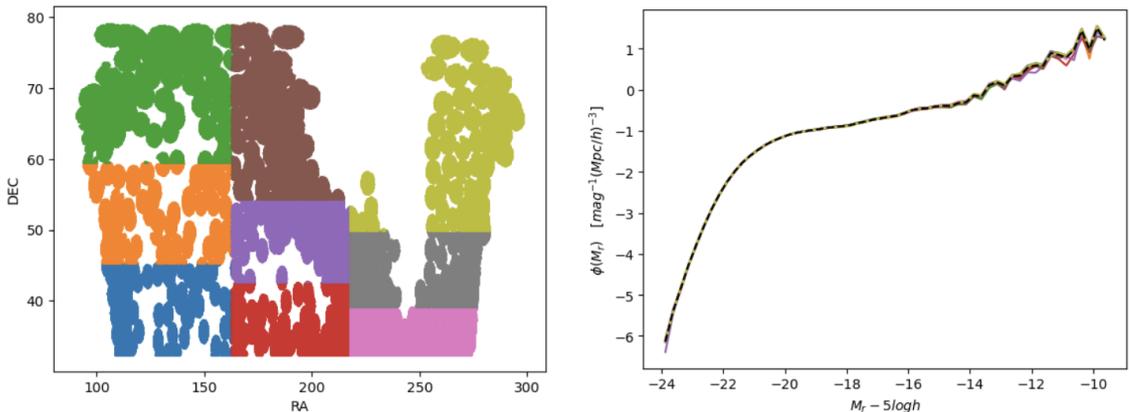
Figure 3.1: Left: Plot showing the jackknife regions for BGS North. These regions are all equal area regions as determined by the random catalogue. Right: The LFs for each jackknife sample. We confirm that the mean of the jackknife sample LFs is identical to the global LF for the entire region.

We additionally calculate jackknife errors for the $1/V_{\max}$ estimator. To begin, we define jackknife areas. To do this, we use the random catalogues and divide the North and South separately into equal object bins which provides us a footprint of equal area regions. To begin, we choose to subdivide the North into 9 areas (3 by 3), and the South into 20 areas (4 by 5), such that each region is approximately the same size for both North and South (North: 265.94 deg^2; South: 267.91 deg^2, see Fig. 3.1). For the North and South separately, we then iteratively remove a jackknife area, calculate the value of ϕ using the remaining areas, and then restore that area, moving on to the next area. The jackknife error is calculated as

$$\text{Var}(x) = \frac{N-1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2, \tag{3.15}$$

where $x_i = \phi(L)_i$, and N is the number of jackknife regions. The pre-factor accounts for the fact that jackknife samples are not independent (Norberg et al., 2009).

We verify that the mean of the jackknife LFs is identical to the global LF, as expected. We conduct a number of additional tests to ensure that the jackknife errors are robust and well-understood. First of all, we 'shift' the jackknife regions such that the regions are defined in different locations in RA. We find that the errors are similar, including in faint magnitude bins.

Second, we test whether the larger errors in the North are quantitatively consis-
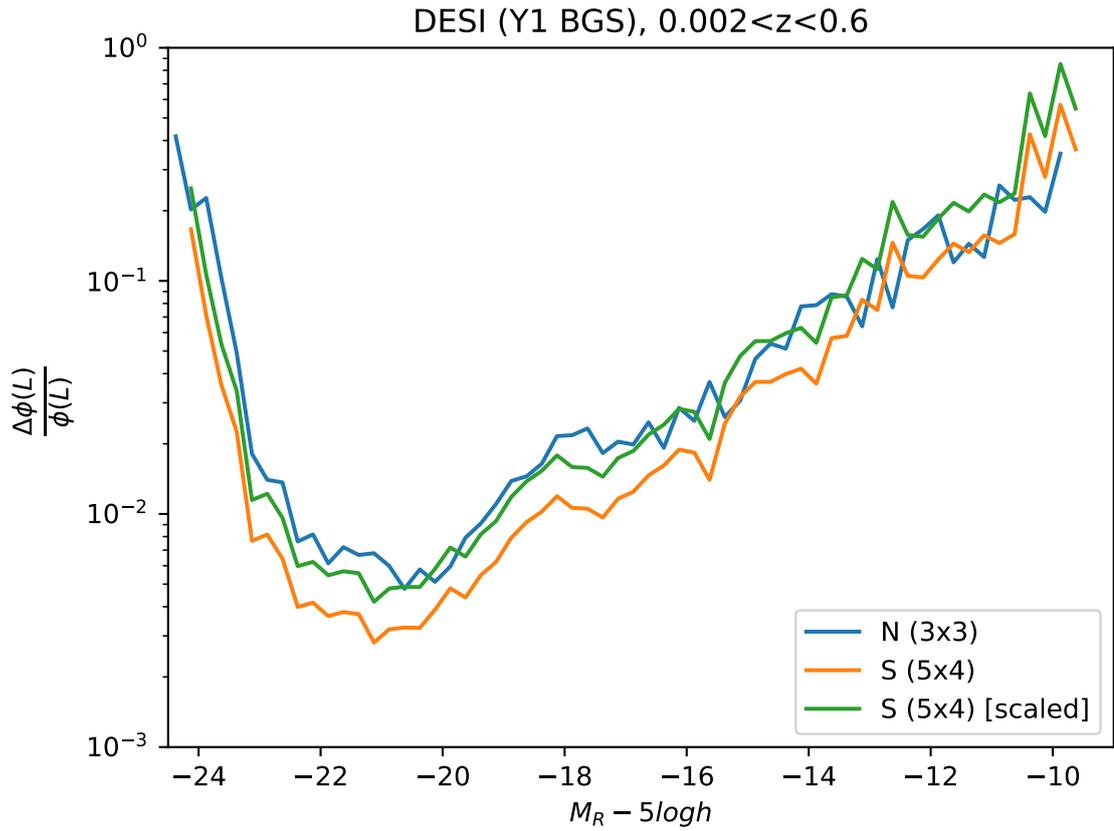
Figure 3.2: Jackknife errors presented for North and South. Before scaling, the North errorbars are larger than the South due to the difference in volume between the North and South area. After scaling the South errorbar by the square root of the ratio of the areas, the errors between North and South broadly agree at all magnitudes.

tent with how much smaller the North survey area is compared to the that of the South. To do this, we scale the South jackknife areas by the square root of the ratio of the areas. Fig. 3.2 shows the jackknife errors for North and South (for the choice 3 by 3 in the North and 5 by 4 in the South). We additionally show the result of scaling the South jackknife errors. We see this brings them into good agreement across almost the entire magnitude range indicating that the jackknife errors scale with survey area in the manner expected.

Finally, it is important to consider the interpretation of jackknife errors at the very faintest absolute magnitudes. We carefully note that objects at the faintest absolute magnitudes are at very low redshifts - as such they occupy a very small volume in the survey. In particular, the faint absolute magnitudes are prone to the issue of super-sample variance - where the full volume of the sample is small compared to the scale of large scale structure. As a result, the jackknife method will underestimate the true error. Moreover, there are very few objects in the faintest absolute magnitude bins - as such, the jackknife regions start to fail to be representative of the full variation across the universe. From a visual inspection of the distribution of points in different magnitude bins, we observe that certain regions start to appear visibly underpopulated around $M_r - 5\log_{10} h > -14$. As a consequence of this, we notice that the estimate of the error becomes noisy and quantised at faint magnitudes due to the small number of data points. In particular, we observe that there were occasions when shifting the jackknife regions did not change the error calculated for some of the faintest absolute magnitude bins. A visual analysis of data points showed that because there are few data points at the faintest absolute magnitudes (of order $N \sim 10$), then there is a reasonable probability that a shift in the jackknife regions will not change the distribution of points across jackknife regions. As such, the error calculated would be the same.

3.2   e-corrections

In order to correct for the evolution of a galaxy with redshift, some authors implement an additional e-correction in their calculation of absolute magnitude. This

changes the absolute magnitude equation from Eqn. 2.1 to the following:

$$^{0.1}M_r - 5\log_{10} h = m_{r_{LS}} - 5\log_{10}\left(\frac{d_L(z)}{h^{-1}\mathrm{Mpc}}\right) - 25 - {}^{0.1}k_{r_{LS}\to r}(z, {}^{0.1}(g-r)) - E(z).$$

(3.1)

For the rest of this thesis, all results will incorporate an e-correction, $E(z)$. We follow the convention of McNaught-Roberts et al. (2014) which uses the functional form

$$E(z) = -Q(z - z_{\mathrm{ref}}).$$

(3.2)

It should be noted that this functional form is determined solely by Q, the evolution of the galaxy luminosity. Other analyses (e.g. Loveday et al., 2012) may also incorporate evolution of the galaxy number density with redshift by introducing a second parameter, P. Loveday et al. (2012) and Loveday et al. (2015) find that for all bands and colour samples, P and Q are strongly anticorrelated due to the estimation of P being dependent on Q. In other words, this means that there is redundancy in P and Q such that different values can yield the same LF. Based on this, we choose to fix $P = 0$ (no number density evolution) and solely estimate Q. One advantage of doing this is that it allows for a more direct comparison to the results in McNaught-Roberts et al. (2014) who also fixed $P = 0$. This was useful the initial stages of our research when validating our methodology on GAMA DR4 data.

We calculate the ratio of V/V_{max} for each galaxy - where V is the survey volume below the redshift at which the galaxy is observed and V_{max} is the maximum volume over which the galaxy could be observed, given the absolute magnitude of the galaxy and the apparent magnitude limits of the survey. If the chosen value of Q correctly models the evolution then in the absence of large scale structure the V/V_{max} distribution should be uniform. In practice, there will be some small variation due to the presence of large-scale structure. In order to determine Q, we assume that the V/V_{max} ratio is uniform. Using this, we vary Q such that χ^2 is minimised between
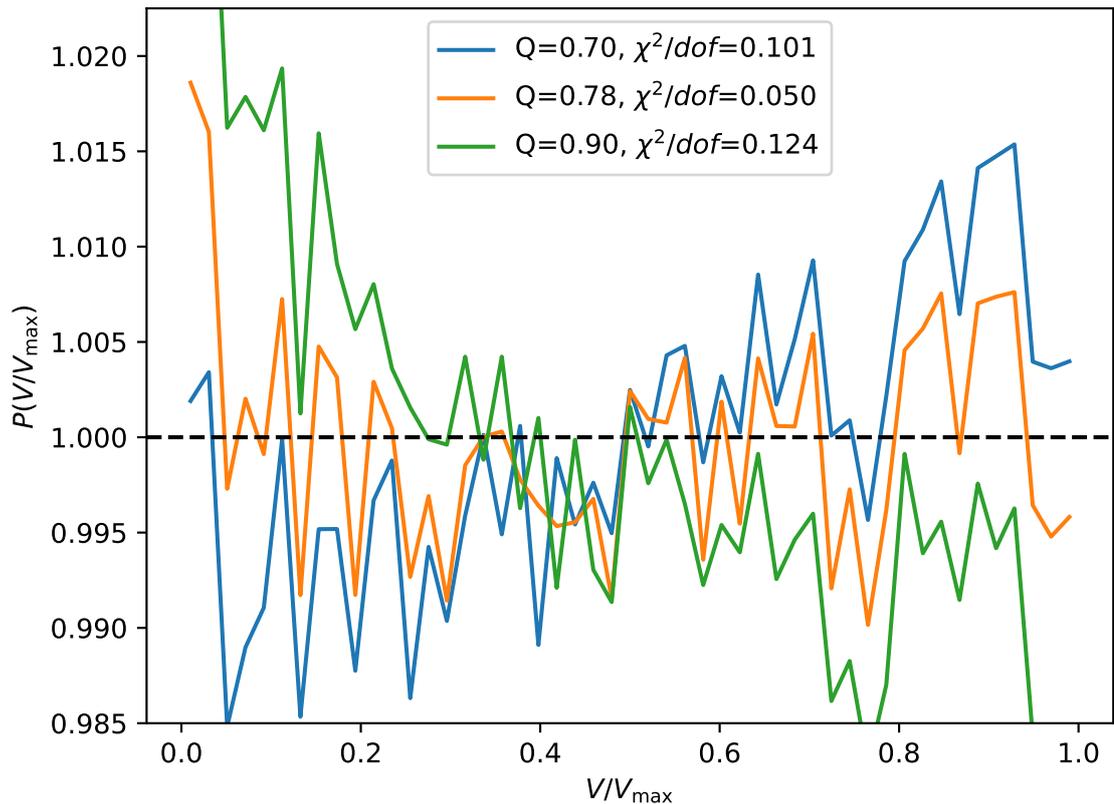
Figure 3.3: The V/V_{max} distribution for $Q = 0.78$ (the found optimal value) and a number of nearby values of Q. The reduced χ^2 value is presented for each Q-value.

the ideal V/V_{max} distribution and the actual V/V_{max} distribution (see Fig. 3.3 and Fig. 3.4). This is conducted over the full apparent magnitude-limited sample from $0.002 < z < 0.6$. From this, we find a global $Q = 0.78 \pm 0.2$, which is compatible with $Q = 0.97 \pm 0.15$ in McNaught-Roberts et al. (2014). In our case, we have derived a jackknife errorbar on Q. We confirm that we generate the same Q-value for the North and the South separately. It is important to note that our e-correction methodology differs from that described in McNaught-Roberts et al. (2014), which calculates the Q-value that preserves the LF in three different redshift bins ($0.01 < z < 0.21$, $0.21 < z < 0.31$, $0.31 < z < 0.51$). If this involved a $1/V_{max}$ LF, we note that our methodology of optimising for a flat V/V_{max} distribution is equivalent. However, McNaught-Roberts et al. (2014) instead made use of an SWML estimator.
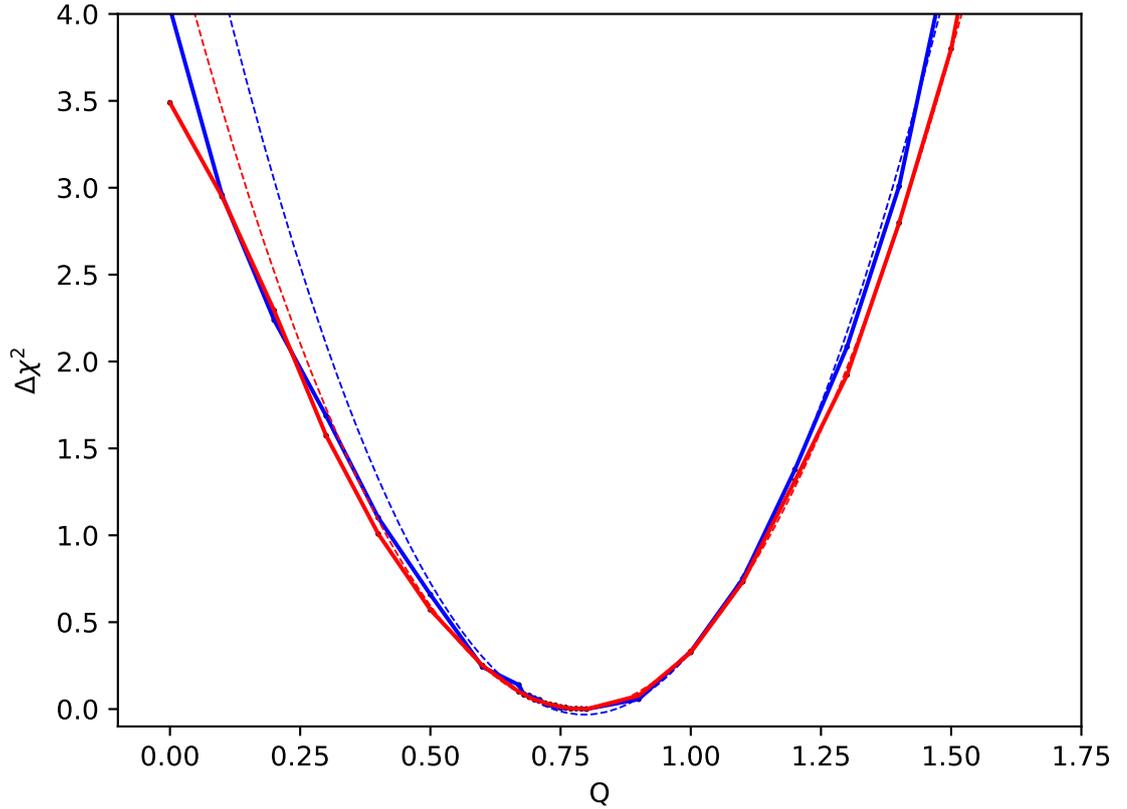
Figure 3.4: The $\Delta\chi^2$ value of the fit to the $V/V_{\max}$ distributions to a uniform distribution for different Q-values when using the BGS Y1 data. This fitting is done over the full apparent magnitude-limited sample from $0.002 < z < 0.6$. Blue represents the North while red represents the South. The dashed lines shows the optimal parabola fits over a narrow range of $\Delta Q = 0.5$ (e.g: 0.6-1.1). From this, we find that $Q = 0.78$ is the optimal value in both the North and South. We separately determine the error from jackknife sampling.

3.3 Global LF Results and Discussion

We generate bivariate LFs as a useful means to visualise and understand where the overall LFs become incomplete. These are presented in Fig. 3.5. For simplicity, the following explanation focuses on the $g - r$ bivariate LF, but the discussion is applicable to the $z - r$ and $w_1 - r$ bivariate LFs as well. The $g - r$ bivariate LF is estimated from objects distributed in bins in the $M_g - M_r$ plane and weighted by $1/V_{\max}$. Bins which contain only one object are indicated by red dots. In addition, the subplots show the projected LFs for the g and r band. The shaded regions in these plots represent the Poisson errors on the LFs.

BGS Bright is an r-band magnitude-limited survey. Whilst there is no explicit limit on the g, z or w_1 bands, this r-band limit will result in a colour-dependent completeness limit in the other bands. In order to better understand our other LFs, we seek to estimate this completeness limit. The black curve represents the completeness limit in the $M_g - M_r$ plane - and is calculated from Eqn. 3.1 using the r-band and redshift limits. Specifically, the limits are $z = 0.002$, and $r = 19.54$ in the North, $r = 19.5$ in the South. We present results with a less strict redshift limit in Appendix B.1.

We use this bivariate plot to estimate the completeness limit of the g-band. To do this, we calculate the 5th and 95th percentiles of the g-band magnitude distribution for each r-band magnitude bin. Then, we conduct an OLS linear regression on these percentiles, weighting for the number of objects in each bin. These linear regressions are shown as dashed pink lines on the bivariate plot. Although not shown here, we note that the linear regression lines are generally good representations of the percentile curves for the $g - r$ and $z - r$ bivariate LFs, however, they trail off at faint magnitudes for the $w_1 - r$ bivariate LF. The intersection of the 5th percentile regression line and the completeness locus allows for a rough approximation of the completeness limit in the g-band, and this is plotted in red on the univariate g-band LF.

We emphasise that this method is an upper-bound on the completeness limit. This is exemplified by the the $w_1 - r$ bivariate LF. As mentioned above, it is unlikely that the methodology discussed above accurately yields a completeness limit for the

w_1-band LF. This can be visually seen in the locus of the bivariate LF – it looks more likely that incompleteness starts to be a factor at around $M_{w_1} - 5 \log_{10} h = -12$ to -14.

More importantly, we note that there are observable 'spikes' in the w_1-band LF that correspond to a surge in both the value of $\phi(M_{w_1})$ and its Poisson error. Moreover, these spikes represent an increase in $\phi(M_{w_1})$ by a factor of 10 or more. Subsequent analysis shows that these spikes are caused by single objects in each bin with a high $1/V_{\max}$ weight and a high total weight w. This combination allows these objects to disproportionately contribute to the LF. Our bivariate LF plot shows that these spikes correspond to single-bin objects that exist far from the locus, and have unrealistic $r - w_1$ rest-frame colours (e.g: the first spike around $M_{w_1} - 5 \log_{10} h = -18$ is caused by an object with $^{0.1}(r - w_1) = 6$).

As such, our solution is to impose an additional $^{0.1}(r - w_1)$ rest-frame colour cut. We select a conservative colour cut of $^{0.1}(r - w_1) < 2.25$ as this value is the minimum value required to remove the spurious peaks brighter than $M_r - 5 \log_{10} h < -14$. This is shown in the $w_1 - r$ bivariate LF plot (Fig. 3.5) as the diagonal red dashed line. Many of the objects that are far from the $r - w_1$ locus (and removed by the colour cut) appear to be single bin objects that appear to be too bright in the w_1 band, assuming that the r-band DESI magnitude is reasonably accurate. We manually observe the SDSS and WISE images of some of these objects to further investigate what is happening. We observe that many of the objects removed by the colour cut appear to be spurious objects - with a number corresponding to mergers, overlapping significantly with closer objects, or being highly fragmented. This supports the idea that these objects have an unrealistically bright w_1-band magnitude, and justifies the use of the $^{0.1}(r - w_1)$ rest-frame colour cut. We conduct further investigations on the faint-end imaging as it pertains to the r-band LF in Section 3.6.

We present the global LFs, after applying the colour cuts, in Fig. 3.6 for the g, r, z, and w_1 bands. We observe that there is broadly good agreement between North and South at both the 'knee' of the LFs and the faint-end of the LFs. There is a breakdown in this agreement at bright magnitudes. This will be discussed further in Section 3.4.
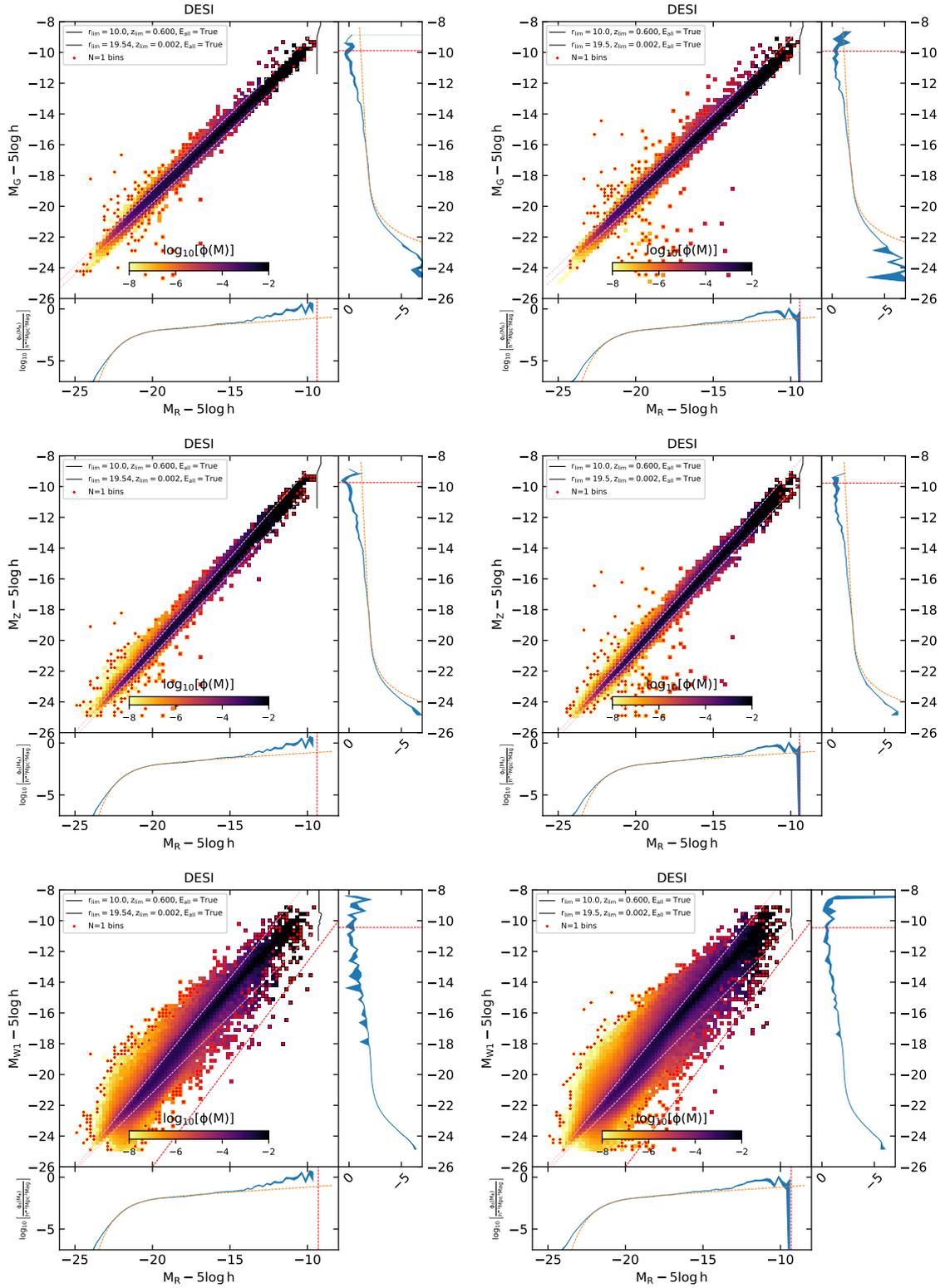
Figure 3.5: Bivariate LFs for North (left column) and South (right column) in the g, z and w_1 bands. The black curves represent the limit of the survey. The dashed red lines give completeness limits for the r-band and g-band luminosity functions based on the intersection of the completeness curves and the 95th percentile contour of the bivariate LF. The $r - w_1$ bivariate LFs have an additional red-dashed line showing a selection cut that is later incorporated to remove spurious objects. The orange dashed lines show the Schechter fits from Loveday et al. (2012).
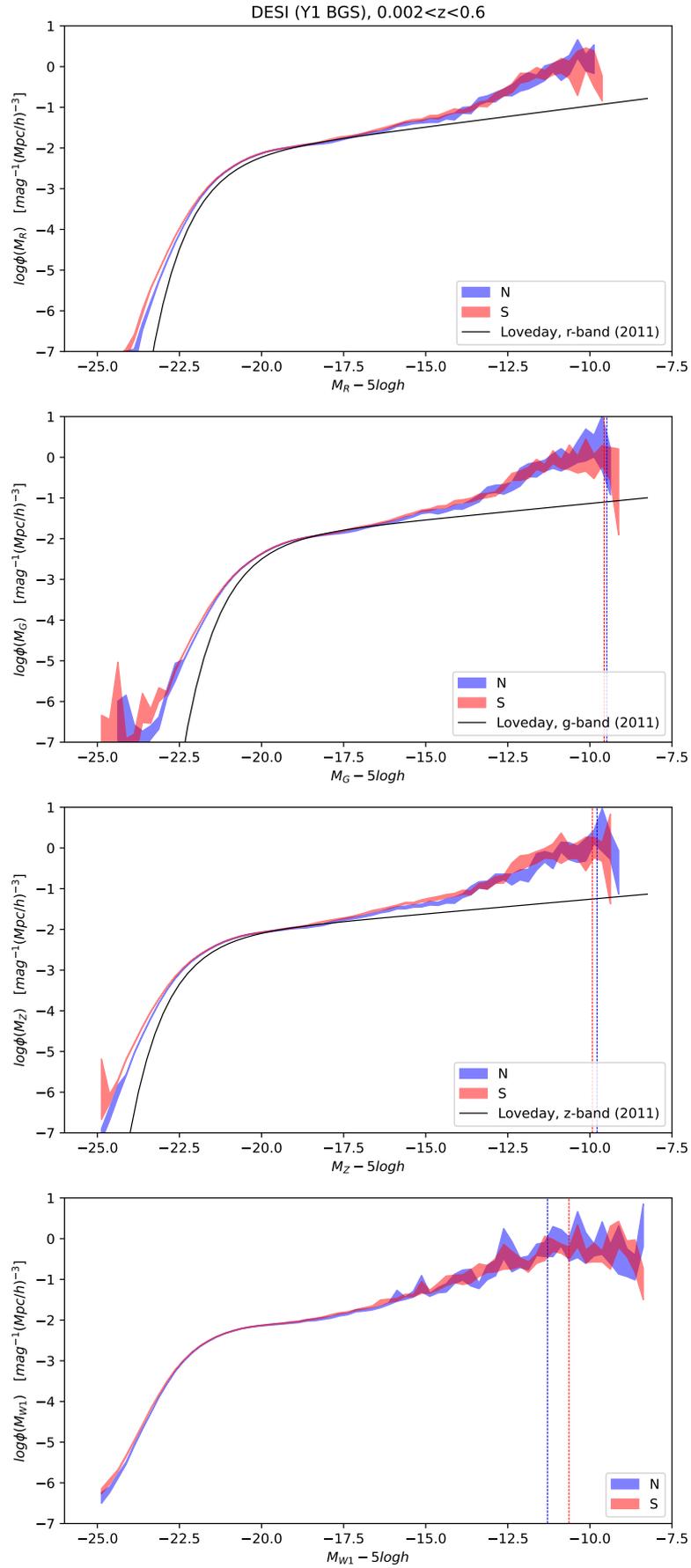
Figure 3.6: The global $1/V_{\max}$ LF for Y1 data for North and South in the r, g, z and w_1 bands. The width of each LF represents the jackknife error. The dashed lines represent the completeness limits derived from the corresponding bivariate LFs. Schechter functions are taken from Table 3 of Loveday et al. (2012).

We observe that the corresponding Loveday Schechter fits are not a good fit to our global LFs from $0.002 < z < 0.6$. We determine that the Loveday Schechter fits are better fits when the upper redshift limit of the sample is $z < 0.2$. This includes the sample $0.002 < z < 0.1$, which is equivalent to the GAMA sample for which the Loveday Schechter functions were fitted (Fig. 3.7). This is clearer when considering the errorbars of the LFs (Fig. B.2). The fact that we get different results for different redshift ranges implies that our evolution model does not fully capture the evolution that occurs in the galaxy sample. In order to resolve this, we investigate different e-correction models to deal with this disparity. One such model is to use different Q-values for red and blue galaxies. We define a red galaxy as $^{0.1}(g-r) > 0.75$ and a blue galaxy as $^{0.1}(g-r) < 0.75$ (see Section 3.5 for evidence for this choice of colour split). We then determine Q-values for each population using the same methodology as described in Section 3.2, labelled Q_{red} and Q_{blue}. We find that $Q_{\mathrm{red}} = 0.23 \pm 0.3$ and $Q_{\mathrm{blue}} = 1.59 \pm 0.2$ (Fig. 3.8). Using these values, we find that the r-band LF from $0.002 < z < 0.6$ is far closer to the Loveday Schechter fit. However, there is still a visible discrepancy at the knee of the LF, suggesting that this is not a perfect fix and that a more complex evolutionary model may be required.

With the global LFs found, we attempt to parameterise them. We make use of a Schechter function to do this. We also investigated other variations, including the double Schechter function. The choice to fit these functions has two major benefits. First, it allows for a closer comparison to the known literature, and we can assess whether a Schechter function is a reasonable approximation of the global LF. Second, even if the Schecher function is a poor fit, it is a useful reference against which we can quantify the significance of the upturn observed in the faint-end of our LFs. The functional form of the Schechter function is given as

$$\Phi(L)dL = \Phi^* \left(\frac{L}{L_*}\right)^\alpha \exp\left(-\frac{L}{L_*}\right)\frac{dL}{L_*} \tag{3.1}$$

where, noting that

$$M - M^* = -2.5\log_{10}\left(\frac{L}{L_*}\right) \tag{3.2}$$
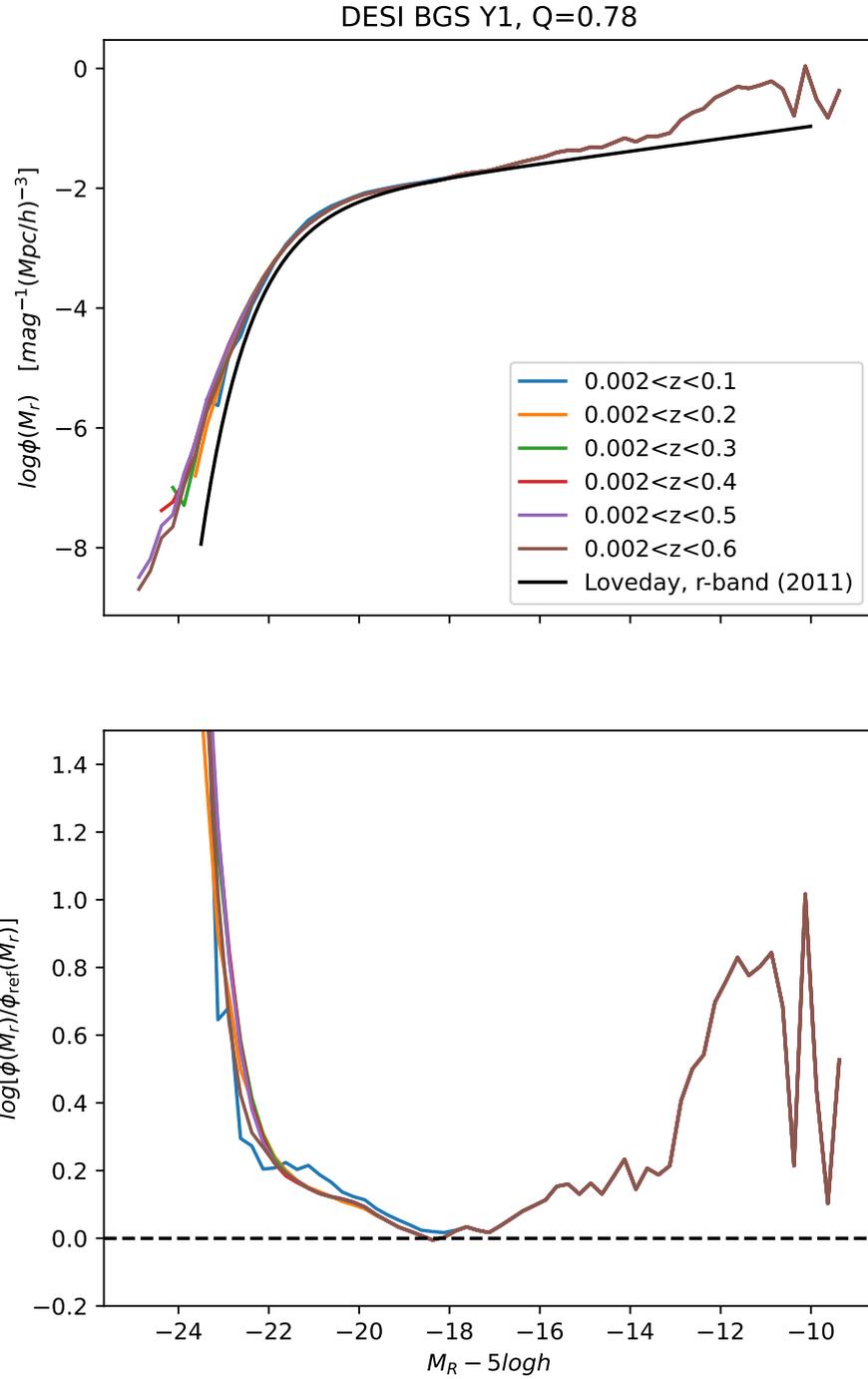
Figure 3.7: Top: The $1/V_{\max}$ LFs for different redshift ranges, using $Q = 0.78$ as our e-correction model. We compare this to the GAMA Schechter function presented in Loveday et al. (2012). Bottom: The residual plot of our $1/V_{\max}$ LFs against the Loveday Schechter fit.
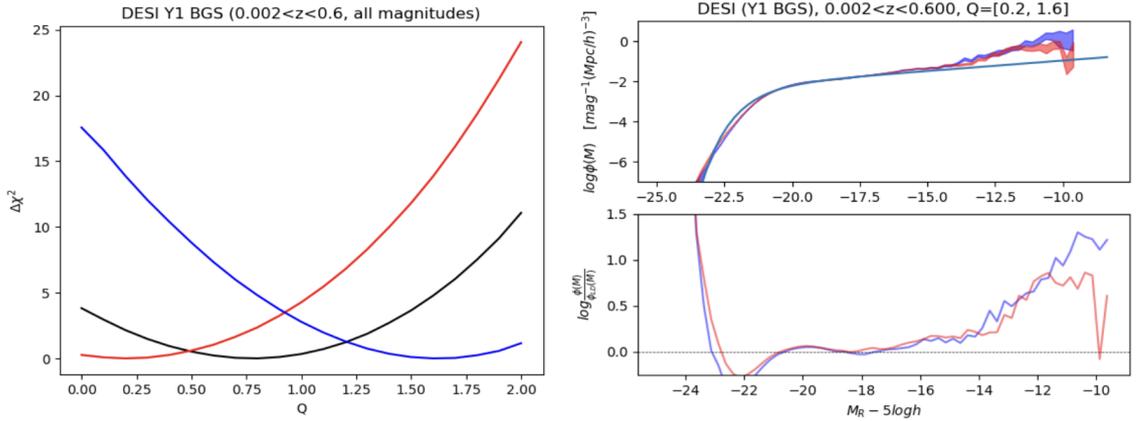
Figure 3.8: Left: The $\Delta\chi^2$ value of the fit to the $V/V_{\max}$ distributions to a uniform distribution for different Q-values when using the BGS Y1 data, split into red and blue samples. This fitting is done over the full apparent magnitude-limited sample from $0.002 < z < 0.6$. Blue represents galaxies with $^{0.1}(g-r) < 0.75$ while red represents galaxies with $^{0.1}(g-r) > 0.75$. From this, we find that $Q_{\mathrm{red}} = 0.23$ and $Q_{\mathrm{blue}} = 1.59$. We separately determine the error from jackknife samples. Right: The $1/V_{\max}$ LF using Q_{red} for red galaxies and Q_{blue} for blue galaxies. In this graph, the blue LF (and corresponding residual line) is the North while the red LF is the South.

and

$$\frac{\Phi(M)}{\Phi(L)} = -\frac{dL}{dM} \tag{3.3}$$

we can define the Schechter function more conveniently in terms of magnitudes as

$$\Phi(M) = \frac{\ln 10}{2.5}\Phi^* 10^{0.4(1+\alpha)(M-M^*)} \exp[-10^{0.4(M-M^*)}] \tag{3.4}$$

Here, M^* parameterises the position of the 'knee' of the Schechter function, α parameterises the faint end slope, and ϕ^* represents a normalisation constant.

It has been observed in prior papers (Loveday et al., 2012) that the Schechter function may not always be the best functional form of the LF (a result that we verify below). As such, there have been various suggested modifications to the Schechter function that add additional parameters to find a better fit. One such example is the double Schechter function, which combines two separate Schechter functions to give a five-parameter model (where there is only one normalisation constant required). The functional form of the double Schechter function (in magnitudes) is given below.

Table 3.1: Schechter parameters for the $0.002 < z < 0.6$ LFs. Here, $M_{\min}$ and $M_{\max}$ refer to the range of $M_x - 5 \log h$ over which the Schechter function was fitted. All other parameters are defined by Eqn. 3.4.

	$M_{\min}$	$M_{\max}$	$\log_{10} \Phi^*$	$M^* - 5 \log h$	α
All (North)					
g	-24	-10	-2.38	-20.68	-1.56
r	-24	-10	-2.08	-20.97	-1.32
z	-24	-10	-2.14	-21.81	-1.26
w_1	-24	-10	-2.20	-21.84	-1.27
All (South)					
g	-24	-10	-2.69	-21.13	-1.76
r	-24	-10	-2.15	-21.12	-1.36
z	-24	-10	-2.16	-21.91	-1.28
w_1	-24	-10	-2.16	-21.81	-1.23
Red (North)					
g	-24	-12	-2.47	-20.55	-1.33
r	-24	-12	-2.15	-20.82	-0.85
z	-24	-12	-2.07	-21.44	-0.66
w_1	-24	-12	-2.08	-21.39	-0.65
Red (South)					
g	-24	-12	-2.58	-20.76	-1.42
r	-24	-12	-2.20	-20.98	-0.93
z	-24	-12	-2.10	-21.57	-0.73
w_1	-24	-12	-2.10	-21.47	-0.71
Blue (North)					
g	-24	-12	-2.70	-20.54	-1.71
r	-24	-12	-2.40	-20.64	-1.52
z	-24	-12	-2.41	-21.26	-1.43
w_1	-24	-12	-2.63	-21.51	-1.49
Blue (South)					
g	-24	-12	-2.79	-20.63	-1.78
r	-24	-12	-2.45	-20.70	-1.56
z	-24	-12	-2.45	-21.34	-1.47
w_1	-24	-12	-2.65	-21.56	-1.49

$$\Phi(M) = \Phi^*(10^{0.4(M_1^*-M)(1+\alpha)} + 10^{0.4(M_2^*-M)(1+\beta)}) \exp[-10^{0.4(M_1^*-M)}] \qquad (3.5)$$

We make use of Markov Chain Monte Carlo (MCMC) methods to fit the various functions to the LFs and calculate appropriate errors on each parameter. In order
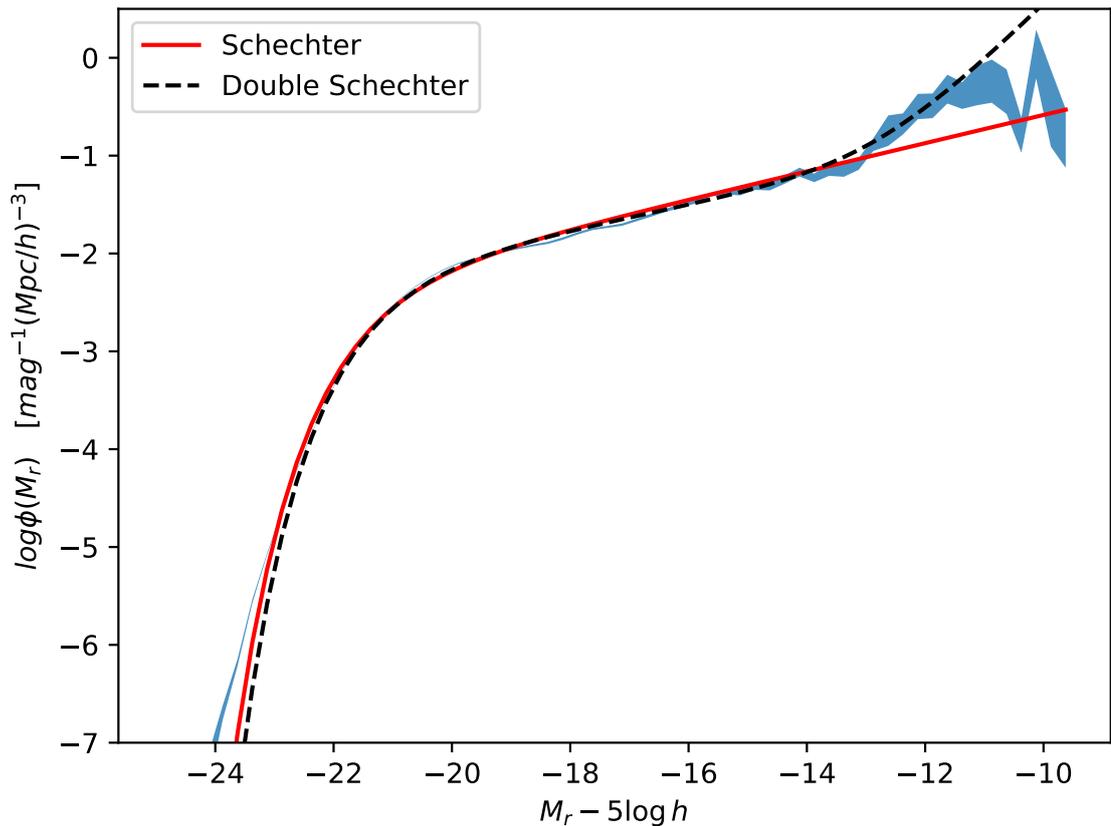
Figure 3.9: Schechter fits to the $1/V_{\max}$ South r-band LF. The parameters for each Schechter fit is found in Table 3.1 and 3.2.

to ensure that the estimated parameters are reasonable, we introduce a number of checks - most notably:

1. We make use of a large number of walkers (> 512) to sample a large proportion of the parameter space. We check that the parameters remain largely invariant to increasing the number of walkers further.

2. We initially make use of burn-ins to ensure that the results are not adversely affected if the initial position of the walkers is too far from the 'true' result.

3. We re-run the MCMC fitting, setting the initial parameters to be the returned parameter results of the previous run. We ensure that the results are the same. This double-checks that the initial conditions of the walkers is not affecting the results.

Table 3.2: Double Schechter parameters for the $0.002 < z < 0.6$ LFs. Here, M_{min} and M_{max} refer to the range of $M_x - 5\log_{10} h$ over which the Double Schechter function was fitted. All other parameters are defined by Eqn. 3.5.

	M_{min}	M_{max}	$\log_{10}\Phi^*$	$M_1^* - 5\log_{10} h$	$M_2^* - 5\log_{10} h$	α	β
All (North)							
g	-24	-10	-2.24	-20.12	-22.24	-0.89	4.79
r	-24	-10	-2.10	-20.95	-14.39	-1.30	-2.50
z	-24	-10	-2.09	-21.66	-16.53	-1.12	-1.79
w_1	-24	-10	-2.06	-21.52	-16.94	-0.98	-1.85
All (South)							
g	-24	-10	-2.73	-21.13	-9.25	-1.76	-1.76
r	-24	-10	-2.27	-21.12	-16.48	-1.36	-1.36
z	-24	-10	-2.15	-21.82	-16.64	-1.19	-1.77
w_1	-24	-10	-2.09	-21.61	-17.12	-1.01	-1.77
Red (North)							
g	-24	-12	-2.24	-20.12	-22.24	-0.89	4.79
r	-24	-12	-2.08	-20.57	-22.57	-0.64	4.36
z	-24	-12	-2.47	-21.44	-20.56	-0.66	-0.66
w_1	-24	-12	-2.09	-21.29	-23.12	-0.61	3.54
Red (South)							
g	-24	-12	-2.31	-20.28	-22.36	-0.99	4.89
r	-24	-12	-2.74	-20.70	-20.28	-1.56	-1.56
z	-24	-12	-2.78	-21.34	-21.23	-1.47	-1.47
w_1	-24	-12	-2.10	-21.37	-23.22	-0.66	3.78
Blue (North)							
g	-24	-12	-2.55	-20.29	-22.47	-1.62	2.50
r	-24	-12	-2.33	-20.47	-22.71	-1.46	2.50
z	-24	-12	-2.60	-21.26	-19.12	-1.43	-1.43
w_1	-24	-12	-2.61	-21.41	-23.36	-1.47	2.50
Blue (South)							
g	-24	-12	-3.56	-20.63	-22.71	-1.78	-1.78
r	-24	-12	-2.49	-20.70	-12.56	-1.56	-1.56
z	-24	-12	-2.51	-21.34	-14.86	-1.47	-1.47
w_1	-24	-12	-2.62	-21.43	-23.29	-1.47	2.50

The results for the Schechter fits are shown in Table 3.1. The results for the double Schechter fits are shown in Table 3.2. In addition, characteristic graphs are shown in Fig. 3.9. From Fig. 3.9, it is visually clear that the standard Schechter function is not a good fit for any of the LFs over the full range of magnitudes. We find that the errors are incredibly small on each Schechter parameter. However,

this is compensated by a very high χ^2 value, and so we do not cite these errors here. However, for each LF, the fit seems reasonable at $M_g - 5\log_{10} h < -16$, $M_r - 5\log_{10} h < -16.5$, $M_z - 5\log_{10} h < -17.5$, $M_{w_1} - 5\log_{10} h < -18.25$. Moreover, we can more carefully quantify the effect of the upturn using the Schechter function. For example, in the r-band LF at $M_r - 5\log_{10} h = -12$, we can see that the our Schechter fit is off by 0.7 dex. The double Schechter function appears to be a better fit, as the additional flexibility from additional parameters allows for the faint end to be better modelled. However, we note that there is still some difficulty in modelling the faint end of the LFs. In particular we warn that some of these double Schechter fits remain poor (especially for the colour-split LFs in the g-band). We find better fits to the faint-end when fitting in the range $-22 < M_g < -12$, but this is at the expense of the bright-end. As such, we recommend that all comparisons are made to the LFs directly. Finally, we note that one alternative to the Schechter fits is to instead fit high-order polynomials to the LFs. We manage to do this with much better success for all bands, although this is prone to overfitting and can be highly variable to the order of the polynomial in modelling the faint end. Because of this, we do not present these polynomial fits here.

In addition to the $1/V_{\mathrm{max}}$ LFs, we also present the r-band SWML and $1/V_{\mathrm{dc,max}}$ estimates. We find that there is good agreement between the different LF methods across all magnitudes (Fig. 3.10). In particular, we observe that for most magnitudes, the $1/V_{\mathrm{max}}$ and SWML LFs agree strongly with each other in both North and South. There is some discrepancy at the faintest magnitudes as expected due to the impact of density fluctuations on the small volume over which faint magnitude objects can be observed. We find that the $1/V_{\mathrm{dc,max}}$ method (which corrects for these fluctuations) agrees more closely with the $1/V_{\mathrm{max}}$ LF. There is a slight systematic offset between the $1/V_{\mathrm{dc,max}}$ LF and the SWML LF across all magnitudes. Nonetheless, this difference is small and does not alter our conclusion that the LF methods broadly agree with each other except at the faintest magnitudes where there is a large degree of variance. As such, we can be confident that our results are independent of the choice of luminosity function estimator.

Finally, now that we have outlined our LF methodology, we confirm (as discussed

in Section 2.5.1) that using the systematic weights has a negligible impact on our LFs. In particular, Fig. 3.11 shows a series of r-band LF for North and South with and without the use of w_{sys}. We observe that the residual remains incredibly small across the knee of the LF. This further justifies our decision to not make use of the systematic weights.

3.4 Bright-end Analysis

In our analysis, we note that there is an offset between North and South at the bright end of the r-band LF. We sought to better quantify the cause of this difference by investigating the FSF k-corrections - noting that these should correct for the differing photometry. In particular, we seek to investigate whether this offset is caused by a problem in our k-correction polynomial pipeline, whether this is an inherent issue caused by the original FSF k-corrections, or whether this is a real observational difference (noting that this is highly unlikely given the large footprint of both DESI North and South).

We attempt to verify whether it is the colour offset that is affecting the LF. This colour offset between North and South is seen in Fig. 2.14. We take the South $^{0.1}(g-r)$ rest-frame colour distribution to be the 'true' distribution, and reassign the North colours according to that distribution. This results in both North and South regions having an identical rest-frame colour distribution. From this, we recompute k-corrections using these new magnitudes. Although this succeeds in getting good agreement between the resulting colour distributions as expected, this does not seem to resolve the offset in the LFs. This suggests that the problem is not corrected for by changing the colour distribution. This result is unsurprising as the colour offset that we observe is on the order of 0.01 magnitudes while the bright-end shift in the LF is closer to the order of 0.1 magnitudes.

We conduct additional tests to try and determine the cause of the North/South discrepancy. There exists an overlap region where objects exist in both BASS and DECaLS, which we match up and then compare their absolute magnitudes. The idea is that this can help reveal if there are differences in the initial photometry
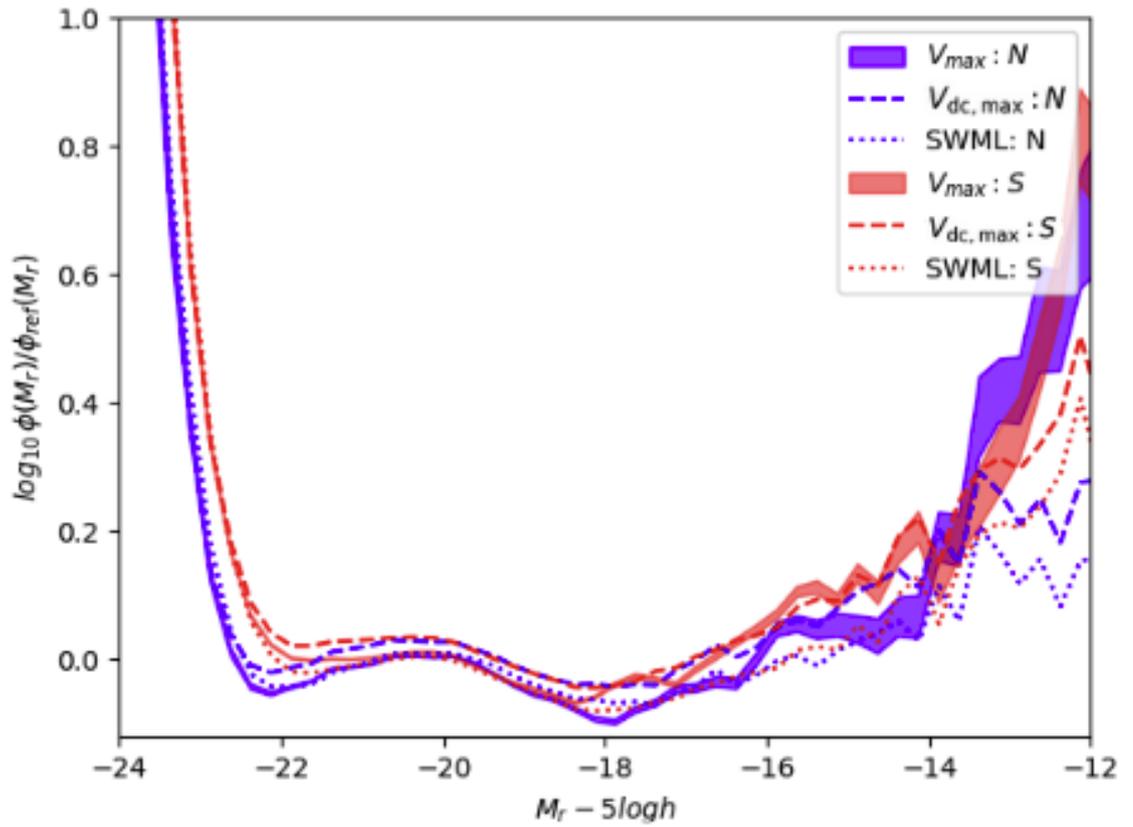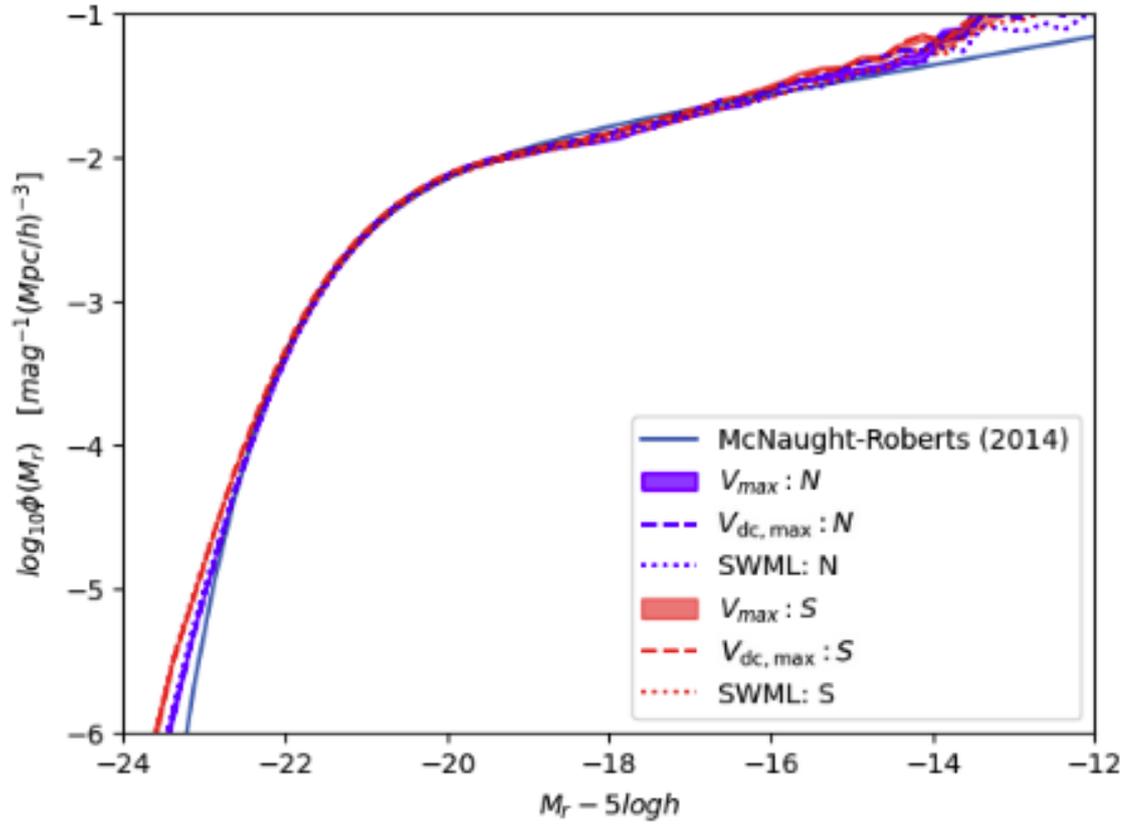
Figure 3.10: A comparison of the different LF estimators ($1/V_{\mathrm{max}}$, $1/V_{\mathrm{dc,max}}$, SWML) in the North and South. Bottom: The ratio plots of the three LFs against the global r-band Schechter function from McNaught-Roberts et al. (2014).
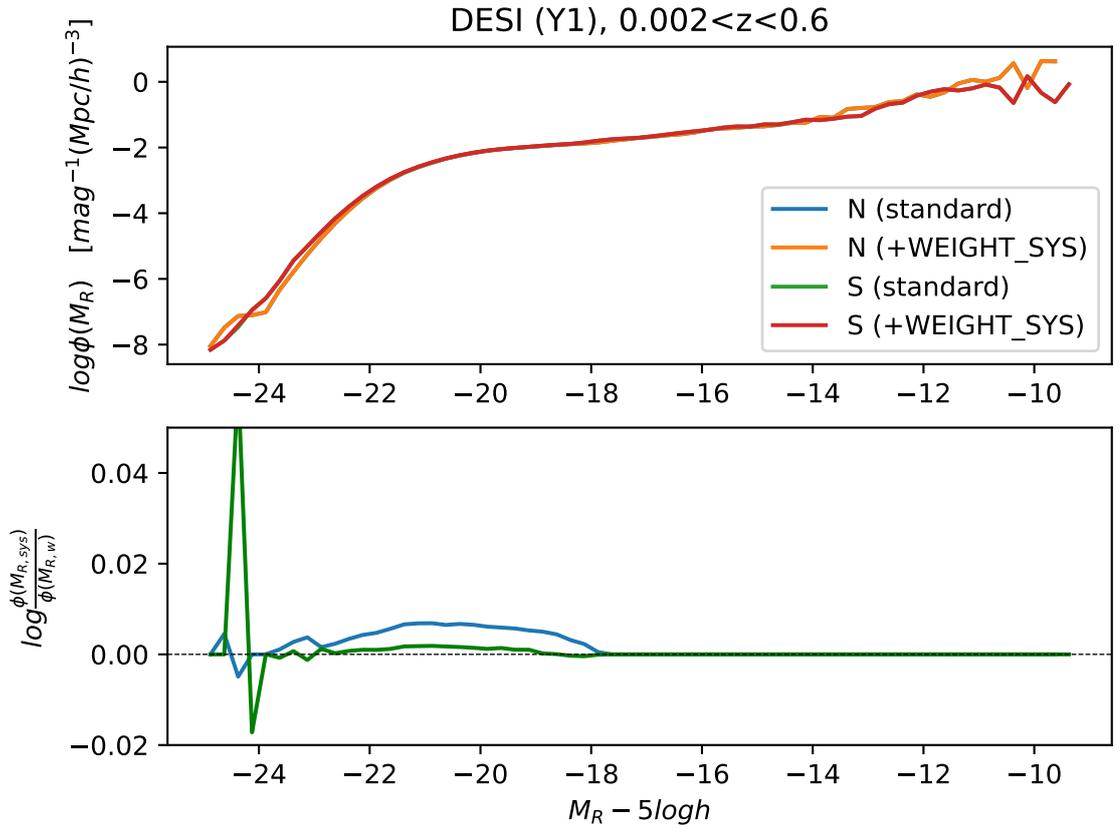
Figure 3.11: Comparison of the $1/V_{\max}$ LF when using the standard weight defined in Eqn. 2.4 against LFs using weights that additionally incorporate w_{sys} (WEIGHT_SYS) as an additional multiplicative factor to Eqn. 2.4.

that could account for the difference in the LFs. First, we perturb the LF by the offset in the BASS/DECaLS magnitude function, noting that as the offset between the two magnitudes diverge as a function of the r-band absolute magnitude, then this could explain an offset at brighter magnitudes. This methodology and the corresponding results are outlined in Appendix C. We find that adjusting for the magnitude residual between BASS and DECaLS yields North and South LFs that are far closer to each other at bright magnitudes. As such, the different photometries may play a role in explaining this discrepancy in the LFs. We note that the bright-end of the LF is in general highly dependent on photometry (see, e.g: fig. 3 from Bernardi et al., 2013) so this is a reasonable hypothesis. One working hypothesis is that BASS and DECaLS may deal with the extended light profiles of the galaxies differently. However, it should be noted that offsetting the LFs does not perfectly fix the issue, so there may be other factors at play.

3.5 Colour-split LF Results and Discussion

Here, we present the LFs as split by colour. For clarity, we refer to LF split by colour as the 'red' or 'blue' LFs. This is not to be confused with an 'r-band' LF, which references the magnitude used.

In order to find colour-dependent LFs, we first split our population into red and blue populations. To define the appropriate colour-cut, we make use of the bimodal nature of the $^{0.1}(g - r)$ rest-frame colour histogram, which has two distinct populations that can each be characterised by a Gaussian curve. By fitting Gaussians to each curve with an Expectation-Maximisation (EM) algorithm, we can split the populations based on the intersection of those two fitted Gaussians. This yields a threshold of $^{0.1}(g - r) = 0.75$. Alternatively, we can visually identify a difference in the two populations by splitting the $^{0.1}(g - r)$ histogram into different M_r absolute magnitude bins as there is a region where the trough of the combined histogram tends to be very similar (Fig. 3.12). This method is consistent with the result found with the EM method. The results of the red and blue $V_{\max}$ LFs are presented for each band (Fig. 3.13). We note that the red and blue populations are consistently
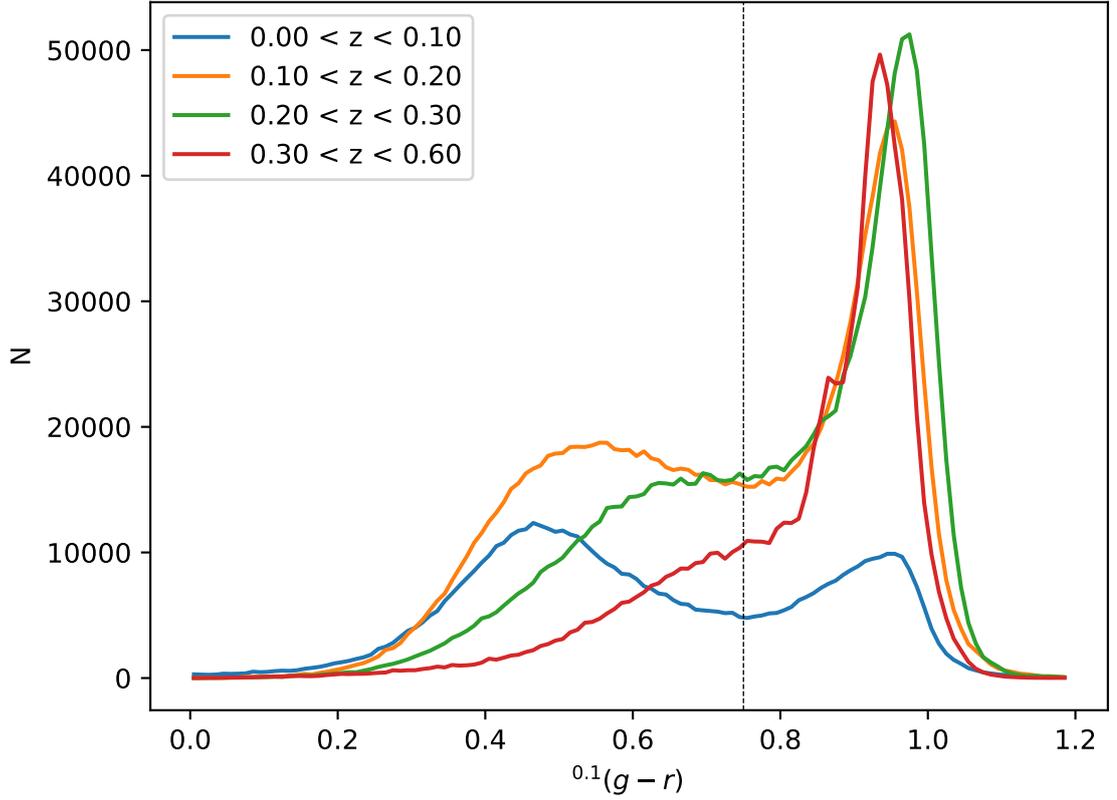
Figure 3.12: The rest-frame $^{0.1}(g-r)$ distributions for different redshift bins. These show a clear bimodal distribution. The dashed vertical line at $^{0.1}(g-r) = 0.75$ is where our Expectation-Maximisation algorithm splits the population into two separate Gaussians. As such, we denote $^{0.1}(g-r) \leq 0.75$ to be blue and $^{0.1}(g-r) \geq 0.75$ to be red.

defined by $^{0.1}(g-r) = 0.75$, even when examining the z and w_1 bands.

The colour LF plots are particularly useful in characterising the upturn observed in the global LFs. In particular, we note that the upturn seems to be driven predominately by the red LF. Moreover, this is observed across the r, g and z bands in both the North and South. We present Schechter and double Schechter fits in Table 3.1 and 3.2. Similarly to the global LFs, the Schechter fits help to quantify the extent of the observed upturn at faint magnitudes. showing it to be significant in the red LF in comparison to the jackknife errors.
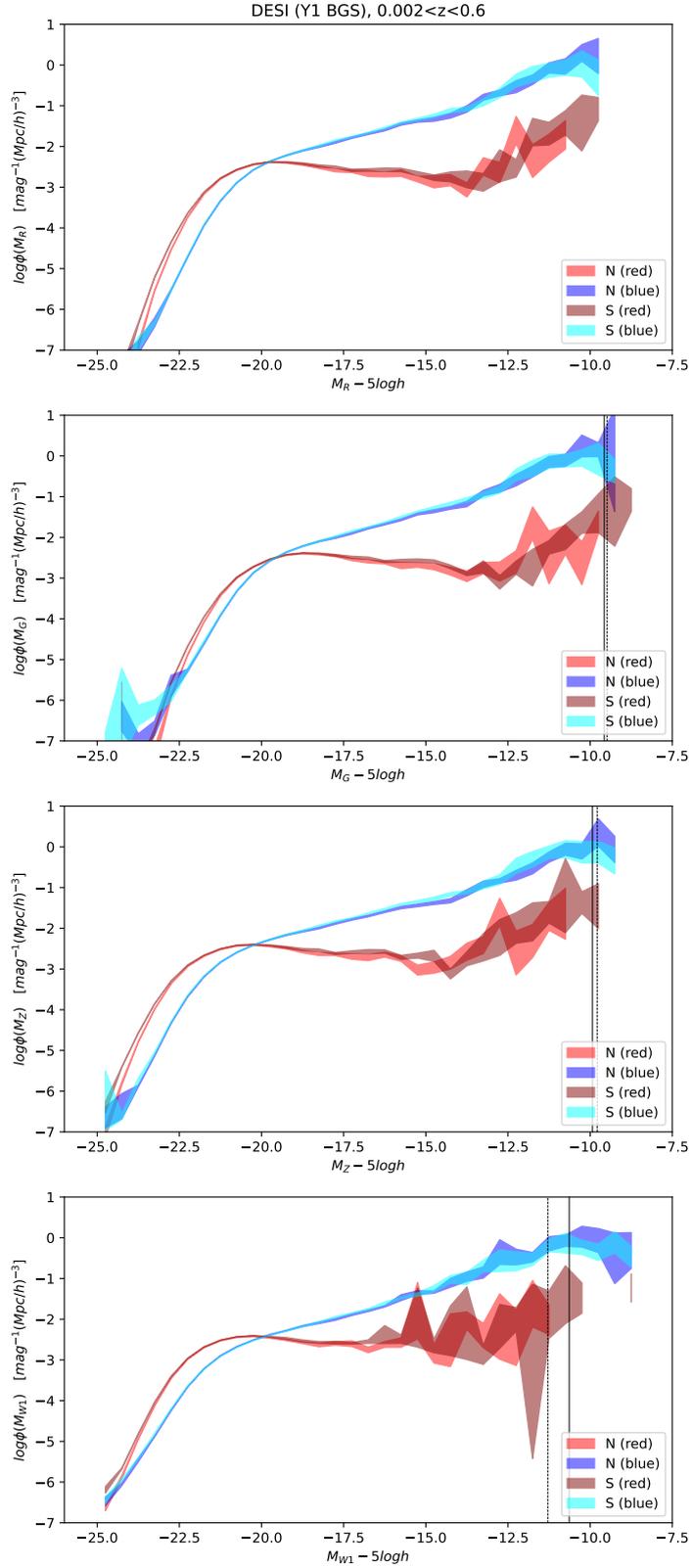
Figure 3.13: The global $V_{\max}$ LF for Y1 data for North and South in the r, g, z and w_1 bands split by colour. The red LFs represent galaxies with $^{0.1}(g-r) > 0.75$, while the blue LFs are galaxies with $^{0.1}(g-r) < 0.75$. The width of each LF represents the Jackknife error. The dashed lines represent the completeness limits derived from the corresponding bivariate LFs.

3.6   Imaging

We observe that our LFs reach extremely faint magnitudes across all bands. Whilst we have outlined a method for estimating the completeness of the g, z, and w_1 bands, we note carefully that this comes with problems. Notably, it is a theoretical completeness limit that does not consider other factors that may introduce error in the faint-end of the LF (including for the r-band LF). In order to better understand the completeness of our data, we conduct a number of visual inspections of the DESI Y1 galaxies using the Legacy Survey Sky Browser. In particular, we are interested as to whether the galaxies at very faint r-band absolute magnitudes are real galaxies with accurate fluxes. Additionally, we are interested in further justifying our $r - w_1$ colour cut, to ensure that we are removing bad galaxies.

The Legacy Survey Sky Browser is a viewing tool that allows us to visually inspect the imaging from various surveys across the whole sky. This viewer usefully contains the Legacy Survey DR9 images in the $ugriz$ bands as well as the unWISE W1/W2 NEO7 images, allowing for quick visual comparisons of different galaxies. In addition, these images can be overlayed with DESI information, including the DESI Bright-time targets and the DESI footprint and fibres. We write a simple code to extract images from the Sky Browser, allowing for easier comparisons of objects. An example image is shown below in Fig. 3.14 for a galaxy found to be bright in the w_1-band but faint in the r-band. Close inspection shows this specific object to be a non-typical object - possibly a galaxy merger. Moreover, viewing the DESI targets suggests that this object may be fragmented, leading to a fainter r-band magnitude than expected.

We created code to conduct quality assurance. This code takes 20 random objects within a magnitude bin and extracts their Sky Browser DR9 image. Details such as the RA, DEC, z, and TARGETID are added to the image. From this, visual inspection may quickly be conducted for each magnitude bin. The categories are delineated as follows:

1. Good: the galaxy is clearly visible and is the dominant object targeted by the fibre. It is unlikely that other features in the image will affect flux extraction.
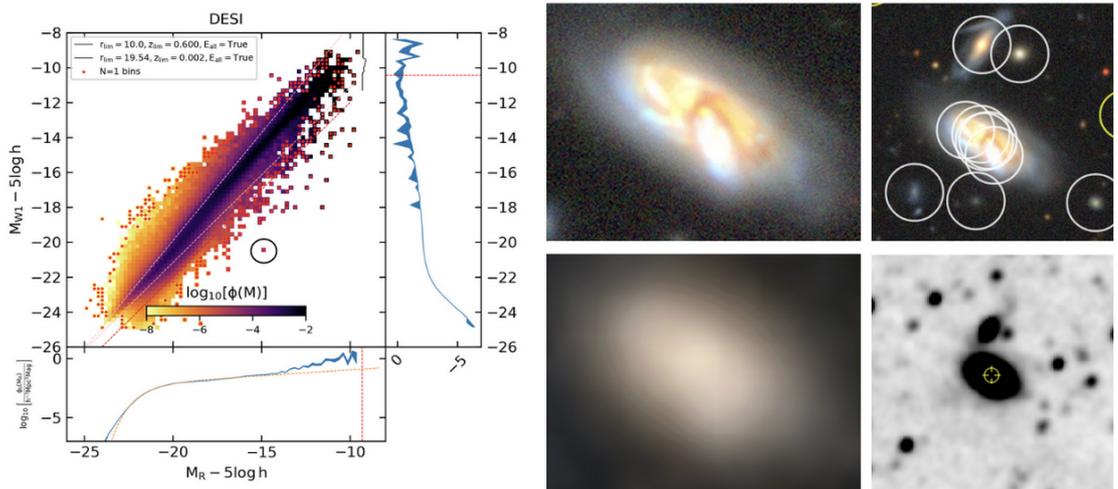
Figure 3.14: An example of a rogue galaxy on the $r - w_1$ bivariate luminosity plot. Further investigation using the Legacy Survey Sky Browser and WISE imaging shows this specific galaxy to be spurious - possibly the result of a galaxy merger or otherwise poor imaging effects. From top left going clockwise: 1) The $r - w_1$ bivariate LF with the target object circled. 2) the grz composite image of the galaxy from Legacy Survey Sky Browser. 3) A zoom out of the grz composite image, with target BGS Bright objects shown as white circles (here multiple circles on the galaxy suggest possible fragmentation). 4) The direct WISE w_1-band image. 5) The w_1-band image presented on Legacy Survey Sky Browser.

2. Possible: the galaxy is visible. There may be other similar objects or features in the image that could conceivably affect the flux extraction. However, we are confident that the object targeted is a galaxy.

3. Problem: it is unlikely that this is a galaxy, or it is a galaxy that is obviously incorrectly imaged.

An example of each type of galaxy is shown in Fig. 3.15. Moreover, we present examples of the 20 galaxies in Appendix D.1 for different absolute magnitude bins.

We observe all 41 galaxies with $M_r - 5\log_{10} h > -10$. Of these, we find 9 of them to be 'good' - that is, are likely to be galaxies (even if the magnitude could be wrong), 29 to be 'problems', and 3 to be 'possible'. In particular, the vast majority of problematic galaxies appear to exist within another more dominant galaxy or dust cloud. As such, we raise the concern that many of the objects at the faintest magnitudes may be artificial and erroneous. For example, there is the issue of fragmentation, where a single extended source is targeted multiple times. As a
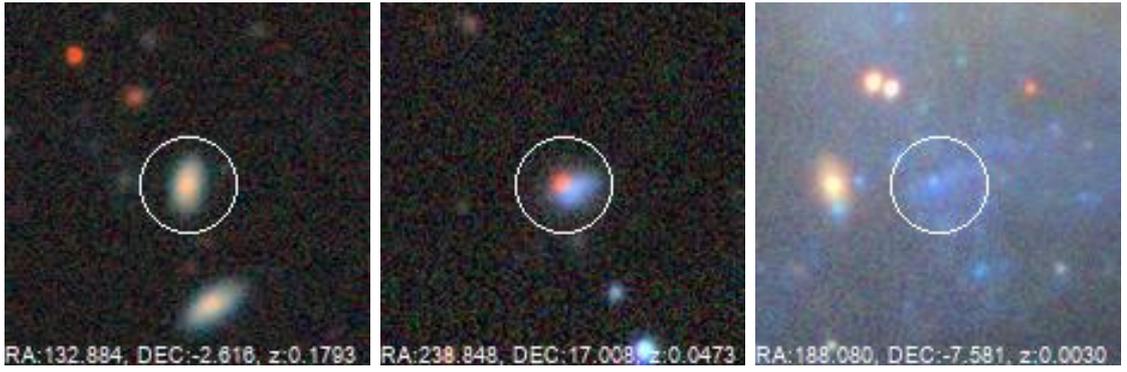
Figure 3.15: Examples of our VI classification metric. The left galaxy is a 'good' object, the middle galaxy is a possible' object, and the right galaxy is a 'bad' object.

result of this, there may exist multiple targets for a single large galaxy where each target gets a lower flux than expected.

We continue with the series of 20 random samples taken from different magnitude bins. The table of our results is shown below (Table 3.3). We present examples of the imaging for different r-band magnitude bins in Appendix D.1. From this work, we suggest that the validity of our r-band LF only extends up to $M_r - 5\log_{10} h \approx -14$. At fainter magnitudes, there exist a large number of bins that are dominated by problematic objects. We carefully note that this leads to the possibility that the upturn in the global LF is a consequence of the imaging analysis artefacts. Specifically, fragmentation leads to a larger number of objects being counted in the fainter magnitude bins. We present further investigations of this in Appendix D.2, where we attempt to use a Neural Network to classify galaxies based on the criteria above. We note that these results do not fully eliminate the upturn in the r-band LF, meaning that it is still possible that the upturn is real, but further investigation is required to better confirm this.

3.7   Conclusions

In summary, in this Chapter we have developed and used a methodology for defining k-corrections for each galaxy in the BGS Y1 dataset. This has the advantage over the FSF k-corrections as our polynomial method allows for V_{max} to be calculated for each galaxy, in addition to consistent rest-frame absolute magnitudes.

Table 3.3: Table showing the VI classification for random samples in several magnitude bins. M_r^{faint} and M_r^{bright} define the $M_r - 5\log h$ range. N_{sample} is the number of objects that we inspect in that magnitude bin, while N_{total} is the total number of objects in that bin.

M_r^{faint}	M_r^{bright}	N_{sample}	N_{total}	$N_{\text{sample}}/N_{\text{total}}$	imaging categorisation (r-band)		
					Good	Possible	Problem
	-10.0	41	41	1.0	9	3	29
-10.00	-10.25	20	28	0.714	3	9	8
-10.25	-10.5	20	28	0.714	3	4	13
-10.5	-10.75	20	40	0.5	2	6	12
-10.75	-11.0	20	64	0.313	5	7	13
-11.0	-11.25	20	102	0.196	4	9	7
-11.25	-11.5	20	124	0.161	7	8	5
-11.5	-11.75	20	143	0.140	4	5	11
-11.75	-12.0	20	176	0.114	12	3	5
-12.0	-12.25	20	189	0.106	8	2	10
-12.25	-12.5	20	228	0.088	10	3	7
-12.5	-12.75	20	282	0.071	13	5	2
-12.75	-13.0	20	333	0.060	9	3	8
-13	-13.25	20	385	0.052	10	2	8
-13.25	-13.5	20	471	0.04246	15	0	5
-13.5	-13.75	20	585	0.03419	13	1	6
-13.75	-14.0	20	776	0.02577	15	0	5
-14.0	-14.25	20	1054	0.01898	11	1	8
-14.25	-14.5	20	1368	0.01462	15	1	4
-14.5	-14.75	20	1723	0.01161	15	0	5
-14.75	-15.0	20	2454	0.00815	13	1	6
-15.0	-15.25	20	3212	0.00623	14	3	3
-15.25	-15.5	20	4427	0.00452	13	1	6
-15.5	-15.75	20	5910	0.00338	17	1	2
-15.75	-16.0	20	7401	0.00270	17	1	2
-16.0	-16.25	20	9246	0.00216	17	1	2
-16.25	-16.5	20	11655	0.00172	20	0	0
-16.5	-16.75	20	14894	0.00134	19	0	1
-16.75	-17.0	20	18891	0.00106	20	0	0
-17.0	-17.25	20	23853	0.00084	19	1	0
-17.25	-17.5	20	31112	0.00064	20	0	0
-17.5	-17.75	20	40719	0.00049	19	0	1
-17.75	-18.0	20	50444	0.00040	20	0	0
-18.0	-18.25	20	62566	0.00032	19	0	1
-18.25	-18.5	20	79371	0.00025	20	0	0

With this, we present $1/V_{\mathrm{max}}$ LFs in the g, r, z and w_1 bands with jackknife errorbars. These LFs are well-determined with small statistical errors over the range $0.002 < z < 0.6$ and $-24 < M_r - 5\log_{10} h < -10$. Moreover, when considering the $1/V_{\mathrm{max}}$ LFs from $0.002 < z < 0.2$, we find good agreement with the GAMA results presented by Loveday et al. (2012). This result is robust to how galaxy evolution is treated, in part due to our choice of reference redshift $z_{\mathrm{ref}} = 0.1$. In addition, we make use of two different LF methodolgies: SWML and $V_{\mathrm{dc,max}}$. Both of these methods broadly agree with our $1/V_{\mathrm{max}}$ estimator, showing that our results are robust to the estimation method.

Although the North and South LFs for BGS Y1 agree very well over a large magnitude range, we observe an offset between the North and South LFs at the brightest absolute magnitudes. This is a highly significant systematic difference. We determine that this is unlikely to be a consequence of our k-correction model. A preliminary investigation of the overlap regions between North and South suggests that the difference is due to the same galaxies being given different magnitudes in the two surveys - in other words, it is an issue in the initial photometry. Moreover, this difference is restricted to the bright red galaxies. Our working hypothesis is that this may be due to a combination of factors. In particular, this may be caused by TRACTOR fitting more complicated photometric models to higher S/N data, which is different in North and South due to the different depths of the two surveys.

In order to find our LFs, we have developed a single parameter e-correction model that estimates Q from the V/V_{max} distribution. We find that $Q = 0.78$ for both the North and the South datasets. However, this model does not yield LFs that are consistent with each other from samples over different redshift ranges. This also leads to an inconsistency between our $0.002 < z < 0.6$ r-band LF and the GAMA LF from Loveday et al. (2012). This means that the single-parameter model does not fully capture the evolution of the LF. We improve the evolutionary model by splitting the galaxies into red and blue samples and finding Q_{red} and Q_{blue} using the same method as before, and this results in a series of LFs that are more consistent between samples over different redshift ranges. However, this is still not perfect and is not consistent to within the small statistical precision provided by the DESI

dataset.

We are able to find the LF up to very faint magnitudes, e.g: the r-band LF extends to $M_r - 5 \log h = -10$. We observe an upturn in the LF at around $M_r - 5 \log h > -14$ which is mainly observed in the red LF. Although this upturn has been observed in some prior studies and is statistically robust given our errorbars, it nonetheless remains unclear in our analysis how much of this upturn may be due to imaging issues. In particular, we visually observe a non-negligible proportion of fragments above $M_r - 5 \log h < -15$, with the proportion of bad objects dominating at the faintest magnitudes. This is seen in Table 3.3. This suggests that this upturn may be artificial, at least in part.

We attempt to correct for the imaging issues using two different methods. The first is to weight the LF in each magnitude bin by the proportion of problem galaxies that we observe in our visual inspection. The second is to utilise a Neural Network to classify a larger subset of objects and remove the problem galaxies before recomputing the LF. This is detailed further in Appendix D. In both cases, we still observe an upturn in the LF, albeit reduced. Given that there remain limitations with these two methods, further work should be conducted to better quantify this issue.

Density-Dependent Luminosity Functions

In this Chapter, we introduce a method for determining the local density around a given galaxy, including corrections for boundaries and 'holes' in the survey. From this, we present density-dependent luminosity functions for BGS. The motivation behind this Chapter is to empirically measure how galaxies are influenced by their environment. By doing so, we can distinguish between different galaxy formation models which make different predictions.

4.1 DDP Definition

In order to estimate the local density of galaxies, we require a tracer population that we describe as a Density Defining Population (DDP). This is a volume-limited sample defined within a range of redshifts and absolute magnitudes. From the selected absolute magnitude limits, we calculate the corresponding redshift limits based on the bright and faint apparent magnitude limits of the survey. This additionally makes use of a conservative global k-correction, in this case making use of a very red value of $^{0.1}(g - r) = 0.9$ to calculate the k-correction, resulting in a stricter redshift range. We assume an e-correction of $Q = 0.78$ when determining the redshift limit.
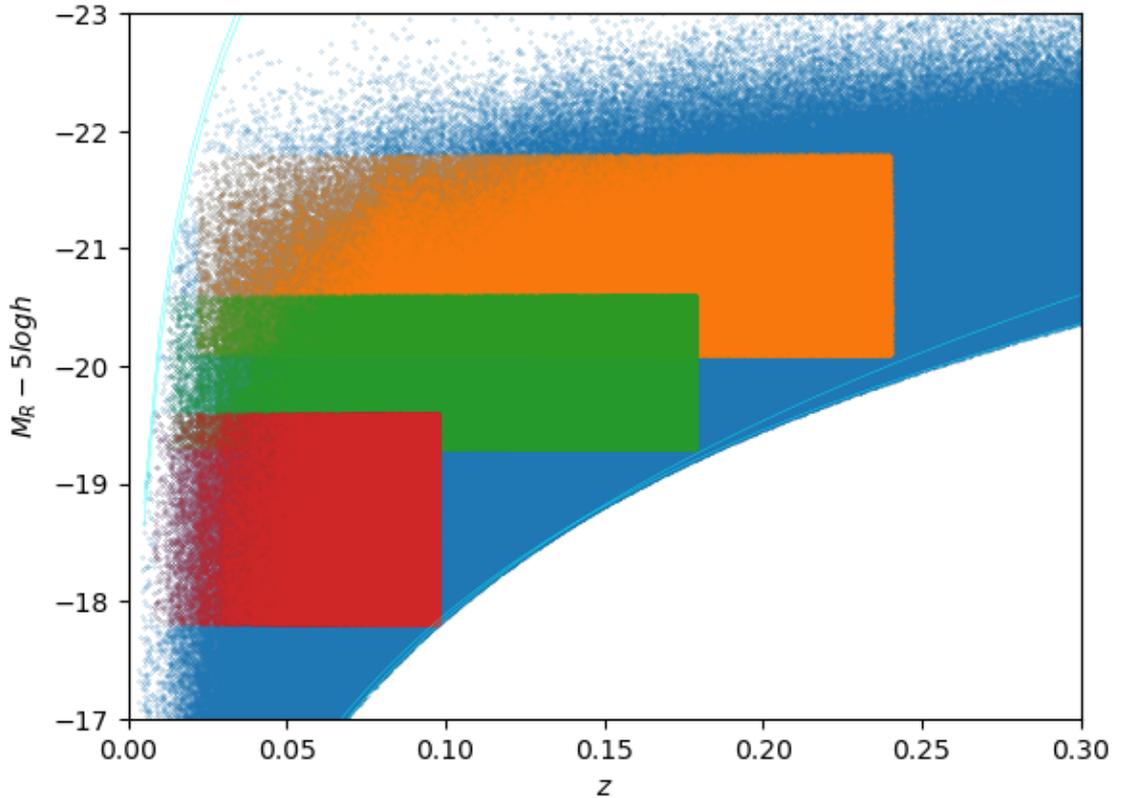
Figure 4.1: Absolute magnitude against redshift for all DESI BGS South galaxies. DDP samples are coloured. The cyan lines correspond to the bright and faint apparent magnitude limits of $r = 12$ and $r = 19.5$ making use of a conservative global k-correction, with $^{0.1}(g - r) = 0.9$ for all redshifts. The North DDPs are defined similarly, with $r = 19.54$ as the faint limit. The absolute magnitude and redshift limits for each DDP sample are given in Table 4.1.

We define all galaxies within these redshift and absolute magnitude limits as DDP galaxies.

We define three different DDP regions based on different absolute magnitude ranges to test that the tracer population is a robust way of defining local density. These can be seen in Fig. 4.1, with the limits and properties of these DDP regions shown in Table 4.1. We mostly focus on using DDP1, as this covers a large volume over which we can calculate local density. Notably, there is some overlap in the galaxies in different DDP regions. As a result, DDP2 and DDP3 are used for validation, as we desire our local density estimates to be largely invariant of our choice of DDP.

Table 4.1: Properties of the three DDP samples, including r-band absolute magnitude and redshift limits. Also shown are the effective comoving volume of the DDP sample, the number density of DDP galaxies and the number of galaxies that fall within the DDP redshift limits. Note that $M_{\min}$ and $M_{\max}$ corresponding to the minimum and maximum values of $M_r - 5\log h$ for the DDP. The units of V_{DDP} are $h^{-3}\mathrm{Mpc}^3$, and the units of ρ_{DDP} are $h^3\mathrm{Mpc}^{-3}$.

DDP	Region	$M_{\min}$	$M_{\max}$	$z_{\min}$	$z_{\max}$	$V_{\mathrm{DDP}}/10^6$	$\rho_{\mathrm{DDP}}/10^{-3}$	$N_{\mathrm{GAL}}/10^3$
1	N	-21.8	-20.1	0.0213	0.2424	77.69	3.52	762.92
1	S	-21.8	-20.1	0.0215	0.2401	169.36	3.85	1 795.93
2	N	-20.6	-19.3	0.0124	0.1804	33.65	5.97	494.78
2	S	-20.6	-19.3	0.0125	0.1787	73.25	6.41	1 170.97
3	N	-19.6	-17.8	0.0079	0.0994	5.99	12.71	162.93
3	S	-19.6	-17.8	0.0080	0.0979	12.80	15.19	413.75

4.2 Local Density

The DDP galaxies as defined above are then used to determine the local density. We define $N_{8,j}$ as follows.

$$N_{8,j} = \sum_{i;\, r_{ij} < r_8}^{N_{\mathrm{ddp}}} w_i \tag{4.1}$$

In other words, $N_{8,j}$ is the weighted sum of DDP galaxies around the jth galaxy in a sphere of radius $r_8 = 8$ Mpc/h. Here, r_{ij} is the distance between DDP galaxy i and galaxy j, and N_{ddp} is the number of DDP galaxies. The choice of an 8 Mpc/h sphere follows from analysis in McNaught-Roberts et al. (2014) who found this value to be a good balance between probing dense environments well and minimising the impact of redshift-space distortions.

Peculiar motions will have some impact on the density estimates through the finger-of-god effect, as peculiar velocities may elongate the galaxy distribution along the line-of-sight. In a non-uniform density field, this may have the effect of reducing the local density estimate within the sphere. Alternative density estimation methods include kernel density estimation (KDE) and various adaptations of this (see Ferdosi et al. (2011)), although we do not made use of them in this thesis.

We then define $\bar{N}_8$ as the expected number of DDP galaxies in an given $r_8 = 8$

Mpc/h sphere given the volume and number density of the DDP sample

$$\bar{N}_8 = \frac{4}{3}\pi r_8^3 \cdot \frac{1}{V_{\text{ddp}}} \sum_i^{N_{\text{ddp}}} w_i. \tag{4.2}$$

Here, V_{ddp} is the volume for the DDP region, defined as:

$$V_{\text{ddp}} = \frac{4\pi}{3}A \cdot [d_c(z_{\text{max}}^{\text{ddp}})^3 - d_c(z_{\text{min}}^{\text{ddp}})^3] \tag{4.3}$$

where A, determined from the number of random galaxies in our matched random catalogue, is the fraction of the total sky covered by the sample. $z_{\text{min}}^{\text{ddp}}$ and $z_{\text{max}}^{\text{ddp}}$ are the minimum and maximum redshift values for the DDP sample (see 4.1). We emphasise that the volumes and number densities are different between North and South. This yields a standard measure for the overdensity:

$$\delta_{8,j} = \frac{N_{8,j}}{\bar{N}_8 \cdot f_j} - 1 \tag{4.4}$$

where f_j represents the fill factor of each galaxy, defined as the ratio of randoms in the $r_8 = 8$ Mpc/h sphere centred on galaxy j to the expected number in a complete sphere

$$f_j = \frac{N_{\text{rand},8,j}}{\bar{N}_{\text{rand},8}}. \tag{4.5}$$

Here, $N_{\text{rand},8,j}$ is defined similarly to $N_{8,j}$: it is the number of DDP randoms around galaxy j in an $r_8 = 8$ Mpc/h sphere, while $\bar{N}_{\text{rand},8}$ is the expected number of DDP randoms in a complete $r_8 = 8$ Mpc/h sphere. The value of the expected local density must be corrected to account for survey holes and boundaries. The fill factor is incorporated to account for the fact that if a galaxy lies on the edge of the survey, then N_8 will be an underestimate as DDP galaxies outside the survey volume will not be detected. For example, if a galaxy lies right on the boundary of the survey, then about 50% of the sphere is within the survey volume, and so has a fill factor of $f_j = 0.5$. We discount all galaxies with a fill factor $f_j < 0.8$ in our analysis in order to avoid introducing large errors on local density estimates for

boundary galaxies.

We present the spatial distribution of the fill factors in Fig. 4.2. This plot shows a slab in the South that has a thickness of 10 Mpc/h with galaxies colour-coded by their fill factors. From this, we can observe two clear features. First, there are visible regions where the fill factor is low at the boundaries of the survey. This includes all of the edges of the survey as expected, including at the z_{min} and z_{max} that are set by the bright and faint apparent magnitude limit. Second, there appear to be radial shells where the fill factor is visibly different from the surrounding area. We ascribe this to the fact that the distribution of randoms follows the $dN(z)/dz$ distribution of the BGS catalogue. As a result, the fill factors will also be correlated with fluctuations in these distributions. Currently, this means that these radial variations in the density of randoms will modulate the assigned value of delta. For example, a radial region of higher $dN(z)/dz$ will have a higher fill factor distribution, which would decrease the estimates of delta in that region. This could be improved by fitting a smooth curve to the random $dN(z)/dz$ distribution and applying additional redshift dependent weights to force the weighted random $dN(z)/dz$ to agree with the smooth fit. However, because the modulation amplitude is small we do not expect this to affect the trends we find with density.

It is important to carefully consider the definition that we use to define local density. In particular, we desire that our DDP tracer population is not biased in any way to skew our overdensity estimates. For example, it is possible that a DDP sample predominately consisting of luminous galaxies may be biased towards overdense regions. As a result, we test that our measurement of overdensity is robust to the choice of DDP used. To do this, we calculate overdensities using DDP2 and DDP3, both of which extend to fainter magnitudes than DDP1 (see Fig. 4.1 and Table 4.1). We then compare the overdensities of galaxies that we can measure using two different DDPs (noting that galaxies outside the redshift range of either of the DDPs cannot be compared). These results are shown for DDP2 and DDP3 (Fig. 4.3). We find that there no significant difference between the overdensities using different DDPs.

79
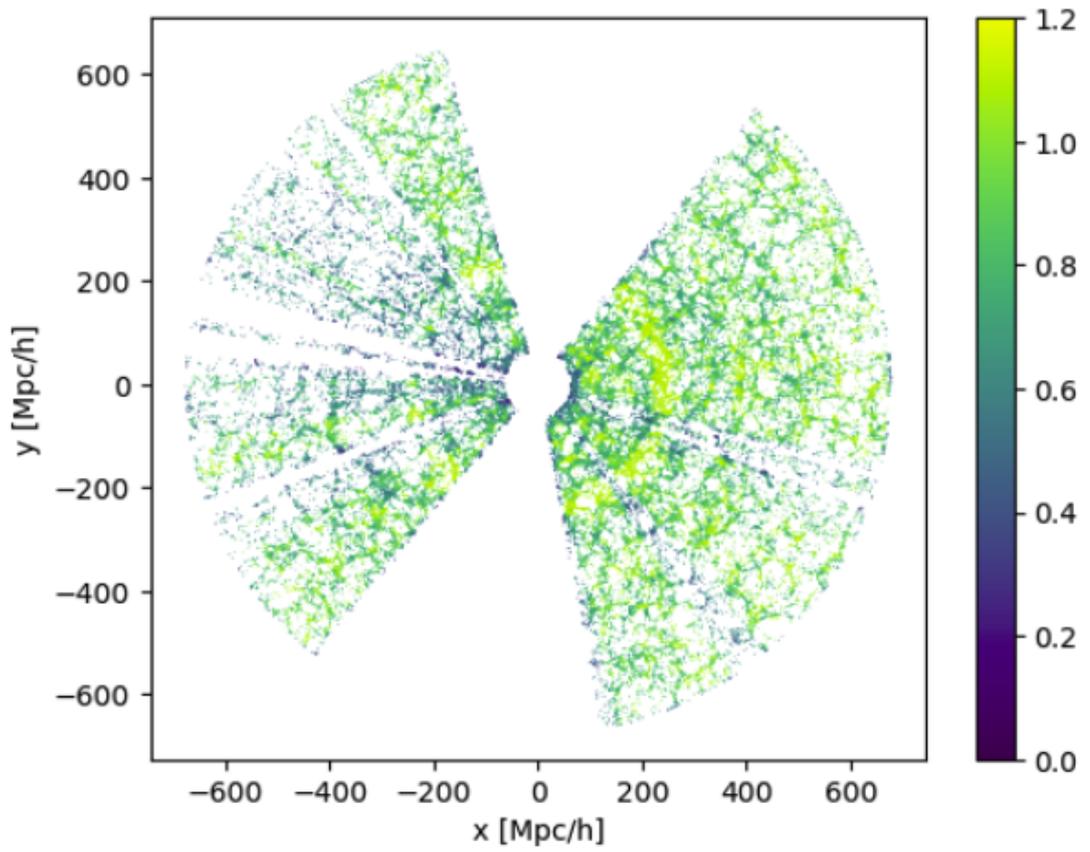
Figure 4.2: The fill factor distribution in space of the Y1 BGS galaxies for a 10 Mpc/h thick slab. This plot includes boundary galaxies with $f < 0.8$, which visibly lie at the edges of the survey.
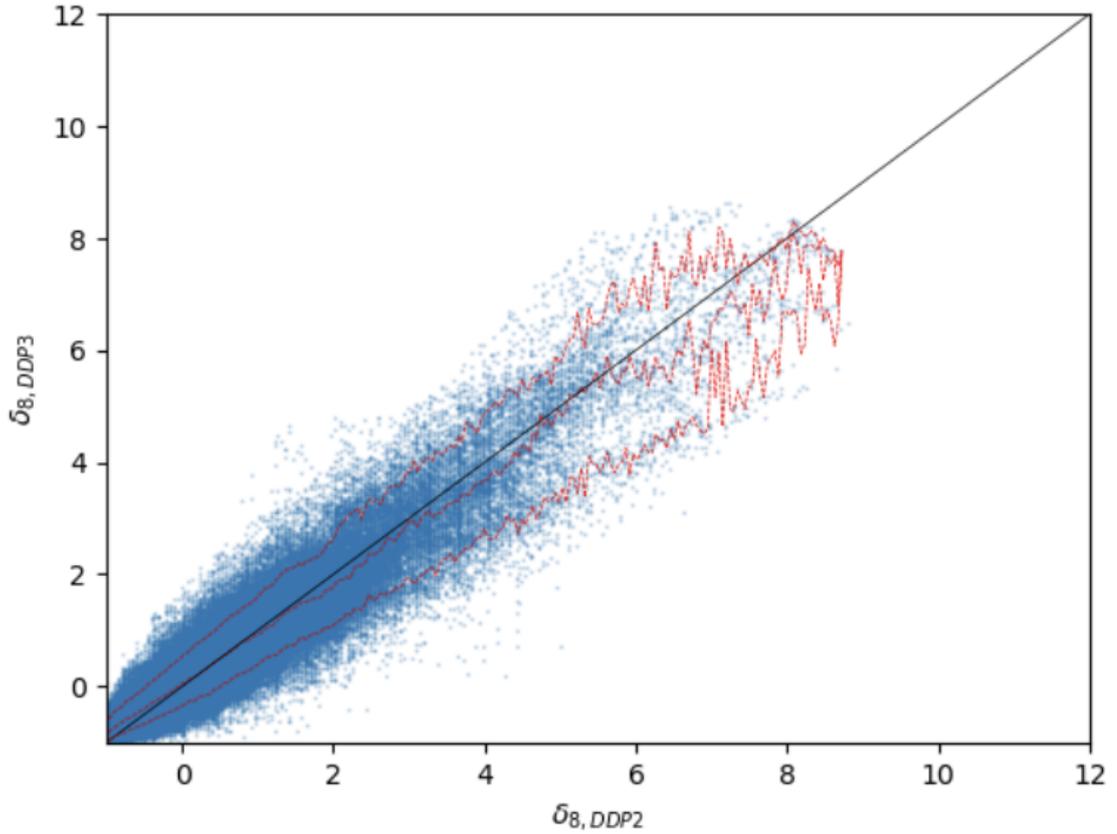
Figure 4.3: Comparisons of the overdensities for different DDP populations. Here, we present the comparison for North objects for DDP2 and DDP3. We only compare objects that can be assigned an overdensity from both DDPs due to an overlap in the redshift range. The red lines present the median, 25th and 75th percentile values of the DDP3 overdensity distribution for each DDP2 overdensity bin. The black dashed line shows the one-to-one correspondence line. Similar plots exist for DDP1/DDP2 and the South.
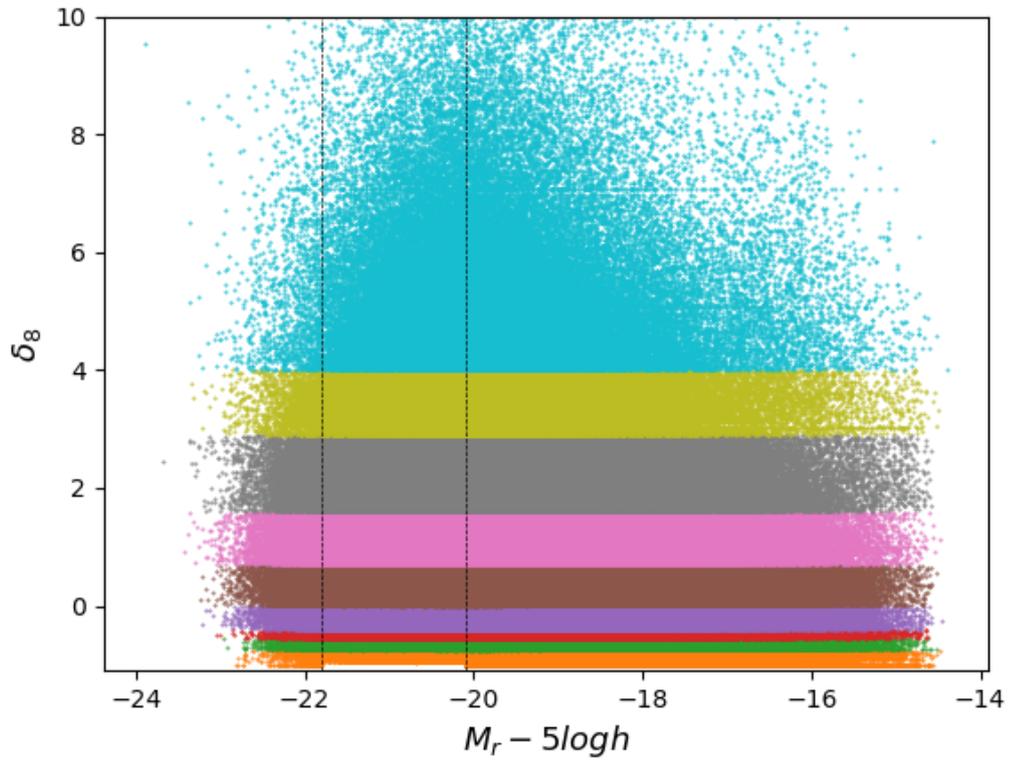
Figure 4.4: The overdensity (δ_8) against absolute magnitude for the South BGS objects. The overdensity is calculated using DDP1. The colours represent the overdensity tier (given in Table 4.2). The vertical lines represent the absolute magnitude limits for the North DDP1 population. It should be noted that because DDPs are self-counting, $\delta_8 \neq -1$ within the DDP1 absolute magnitude range, causing a visible gap.
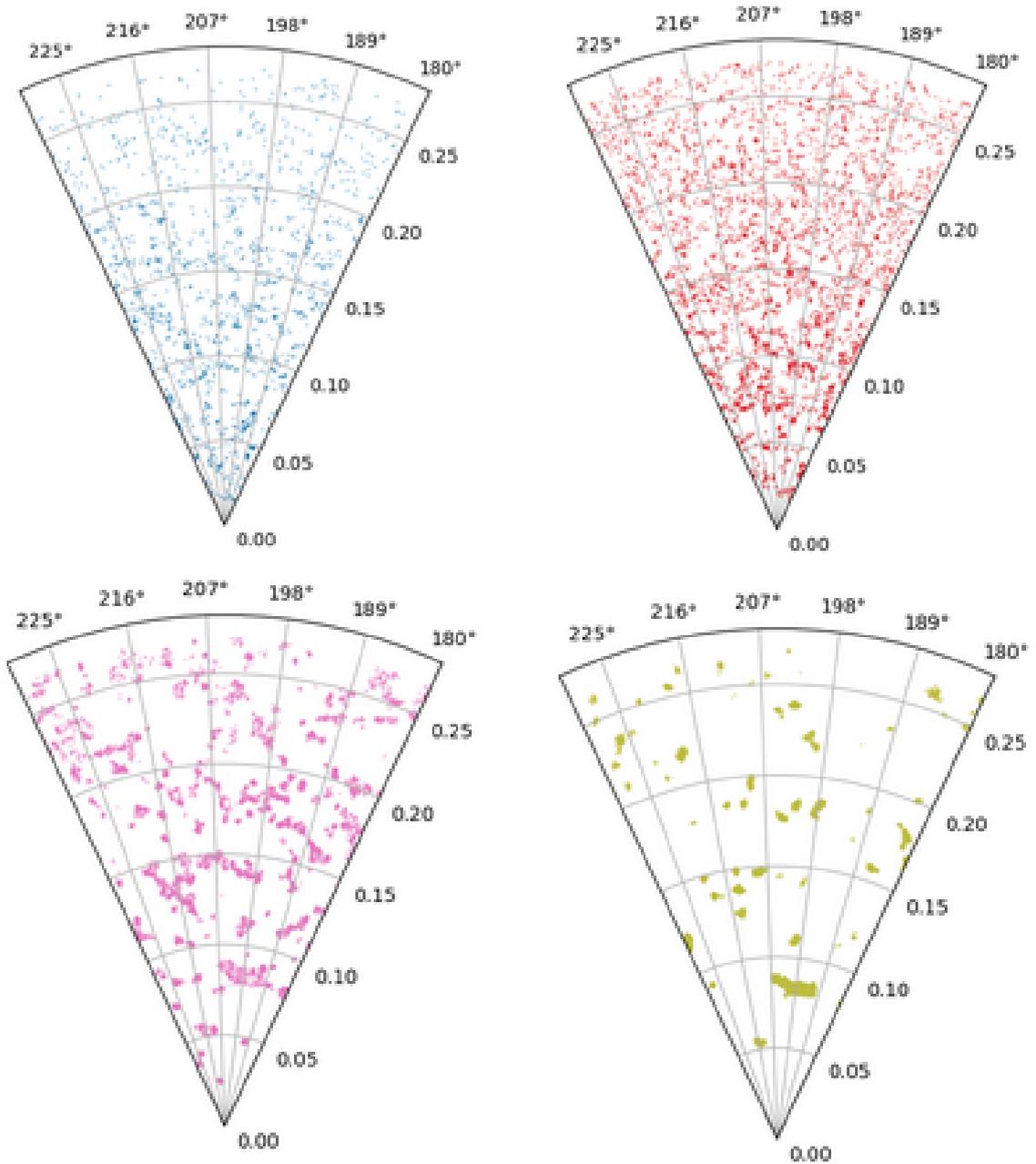
Figure 4.5: Cone plots showing the galaxies for the δ_8 tiers. Top: d1 (left), d4 (right). Bottom: d6 (left), d9 (right).

4.3 DDP Luminosity Function

With the local density defined for each galaxy (as detailed in Section 4.2), we can also find density-dependent LFs based on the local density calculated in an 8 Mpc/h sphere, hereafter denoted as δ_8 LFs.

To begin with, we categorise all galaxies into a "δ_8 tier", that is, a category based on its δ_8 value. Again, for ease of comparison, we use the same δ_8 tiers as McNaught-Roberts et al. (2014). This is shown in Table 4.2 and Fig. 4.4. In addition, we present cone plots of the data split into four δ_8 tiers (d1, d4, d6, d9). This is shown in Fig. 4.5. We observe from these cone plots that higher δ_8 tier galaxies tend to be less uniform over the entire space, instead existing as clumpy structures. Moreover, there are few high δ_8 tier galaxies at very low redshift. This is a result of the volumes being extremely small at low redshift.

We present the overdensity values, δ_8, by tier against absolute magnitude for DDP1 (Fig. 4.4). From this graph, there are a number of noticeable features that are useful for understanding our methodology. First, there is an absence of objects at $\delta_8 = -1$ within the DDP1 absolute magnitude range. This is because DDP galaxies are self-counting, and all galaxies within that absolute magnitude range (and redshift limits of the DDP population) are DDP galaxies by definition. Second, there is some visible quantisation of points, where zooming in on this plot reveals visible striations in δ_8 where points are more likely to lie. To understand this, consider the case where all galaxies are assigned a total completeness weight of unity. Then, $N_{8,j}$ is an integer value. As such, the impact of increasing $N_{8,j}$ by 1 is to increase $\delta_{8,j}$ by $1/(N_{8,j} \cdot f_j)$. Given that f_j is close to unity for most galaxies, this is effectively a discrete step with some small amount of scatter. Introducing the total weight then has two effects on this δ_8 distribution. The target completeness weight is an integer value and so can simply shift galaxies from one striation to a higher one. The redshift completeness weight is then a non-integer value, and so introduces further scatter to each striation. However, as shown in our analysis in Section 2.5.3, the mean of these redshift completeness weights is close to unity and so has a marginal effect.

For the DDP range, we define a new value, $V_{\mathrm{max},i}^{\mathrm{ddp}}$ to account for the fact that the

Table 4.2: Table of DDP1 overdensity bins. Listed are overdensity limits, $\delta_{\min}$ and $\delta_{\max}$, the number of galaxies in a given δ_8 tier, $N_{\delta,\text{DDP1}}$, and the mean δ_8 value in each tier, $\langle \delta_8 \rangle$.

Tier	Region	$\delta_{\min}$	$\delta_{\max}$	$N_{\delta,\text{DDP1}}/10^3$	$\langle \delta_8 \rangle$
d1	N	-1.00	-0.75	59.280	-0.884
d1	S	-1.00	-0.75	124.439	-0.880
d2	N	-0.75	-0.55	57.123	-0.647
d2	S	-0.75	-0.55	130.097	-0.646
d3	N	-0.55	-0.40	46.888	-0.474
d3	S	-0.55	-0.40	105.138	-0.474
d4	N	-0.40	0.00	121.904	-0.203
d4	S	-0.40	0.00	288.934	-0.202
d5	N	0.00	0.70	176.008	0.328
d5	S	0.00	0.70	422.432	0.331
d6	N	0.70	1.60	140.168	1.108
d6	S	0.70	1.60	340.776	1.106
d7	N	1.60	2.90	98.296	2.151
d7	S	1.60	2.90	232.92	2.149
d8	N	2.90	4.00	36.381	3.369
d8	S	2.90	4.00	80.907	3.381
d9	N	4.00	29.49	26.877	5.353
d9	S	4.00	26.36	70.283	5.492

volume over which a galaxy may be observed is now limited by a stricter redshift range.

$$V_{\max,i}^{\text{ddp}} = \frac{4\pi}{3} A [d^3(z_{\max,i}^{\text{ddp}}) - d^3(z_{\min,i}^{\text{ddp}})] \tag{4.6}$$

where

$$z_{\max,i}^{\text{ddp}} = \min(z_{\max,i}, z_{\max}^{\text{ddp}}) \tag{4.7}$$

and

$$z_{\min,i}^{\text{ddp}} = \max(z_{\min,i}, z_{\min}^{\text{ddp}}) \tag{4.8}$$

In other words, $z_{\min,i}^{\text{ddp}}$ and $z_{\max,i}^{\text{ddp}}$ refer to the minimum and maximum redshifts

at which the ith galaxy could be observed within the DDP. The maximum possible value of $z_{\mathrm{max},i}^{\mathrm{ddp}}$ is the upper redshift of the DDP itself, for example, $z_{\mathrm{max}}^{\mathrm{ddp}} = 0.242$ in DDP1 in the North. A similar constraint exists for $z_{\mathrm{min},i}^{\mathrm{ddp}}$. We note that by definition, $V_{\mathrm{max},i}^{\mathrm{ddp}} \leq V_{\mathrm{max},i}$.

When using the $1/V_{\mathrm{max}}$ estimator on different δ_8 tiers, it is important to correct for the fact that each δ_8 tier will occupy a different volume within the total DDP volume. This can be quantified using the random catalogue. We define the fractional volume in Eqn. 4.9 where $N_\delta(< z)$ is the cumulative number of randoms found in δ_8 tier δ at redshift z and $N_{\mathrm{all}}(< z)$ is the cumulative number of total randoms found across all δ_8 tiers found at redshift z. From this, we define a corrective volume fraction for each galaxy.

$$f_{\delta,i} = \frac{N_\delta(< z_{\mathrm{max},i}^{\mathrm{ddp}}) - N_\delta(< z_{\mathrm{min},i}^{\mathrm{ddp}})}{N_{\mathrm{all}}(< z_{\mathrm{max},i}^{\mathrm{ddp}}) - N_{\mathrm{all}}(< z_{\mathrm{min},i}^{\mathrm{ddp}})} \tag{4.9}$$

From this, the V_{max} quantity used in our estimator for each δ_8 tier is

$$V_{\mathrm{max},i}^{\delta} = V_{\mathrm{max},i} \times f_{\delta,i} \tag{4.10}$$

Using $V_{\mathrm{max},i}^{\delta}$ defined above, we present the LFs for four different δ_8 tiers (d1, d4, d6 and d9) in Fig. 4.6. Here we show the zeropoint LFs where DDPs are not self-counting, and thus values of $\delta_8 = -1$ are possible within the DDP range. Although this is different from the overdensities presented in Fig. 4.4 and additionally leads to a discontinuity in the density field (at the centres of the DDP galaxies), this has the advantage of being more directly comparable to McNaught-Roberts et al. (2014). We observe that there is good agreement between the North and South LFs for all tiers except for d9, the highest δ_8 tier, which differ at faint magnitudes. This is explainable by the fact that faint magnitude objects are only observable in a very small volume close to the observer. As such, the sample variance is very large as the result depends on whether or not there is a large overdense region in that small volume.
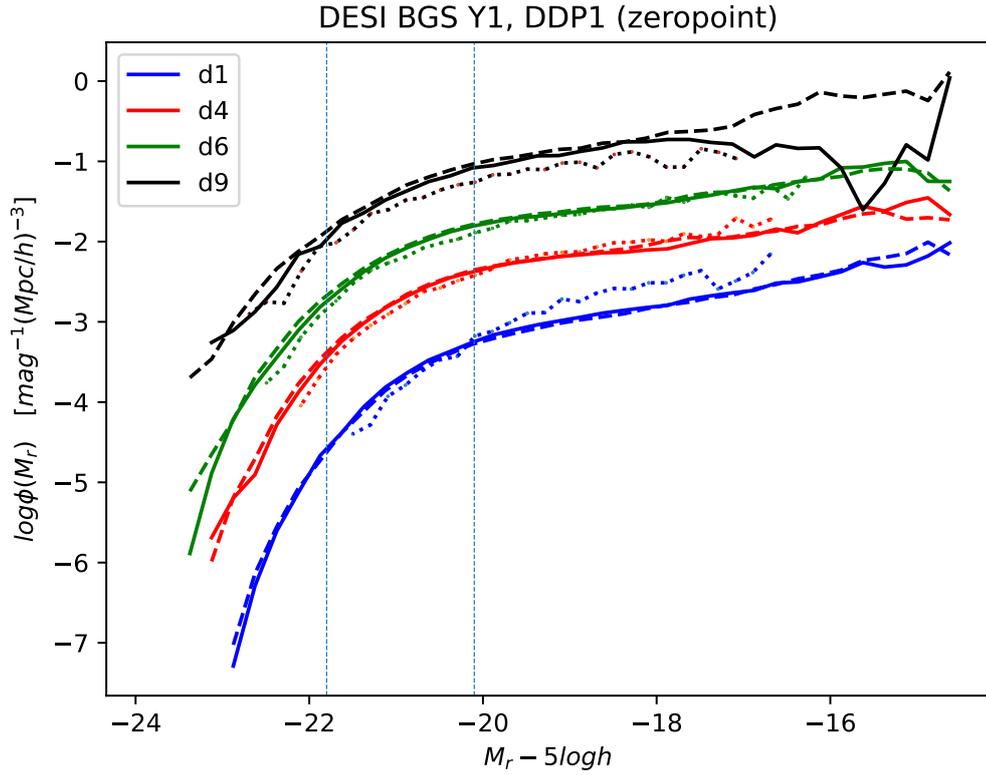
Figure 4.6: The density-dependent $1/V_{\max}$ LFs for North and South. North is represented by the solid line, while South is represented by the dashed line. Each LF represents a different δ_8 tier, as outlined in Table 4.2. Note that here we are presenting the zeropoint LFs (where the DDPs are not self-counting). The dotted line shows the equivalent density-dependent LFs for GAMA as presented in McNaught-Roberts et al. (2014). The vertical dashed lines indicate the absolute magnitude limits of DDP1.

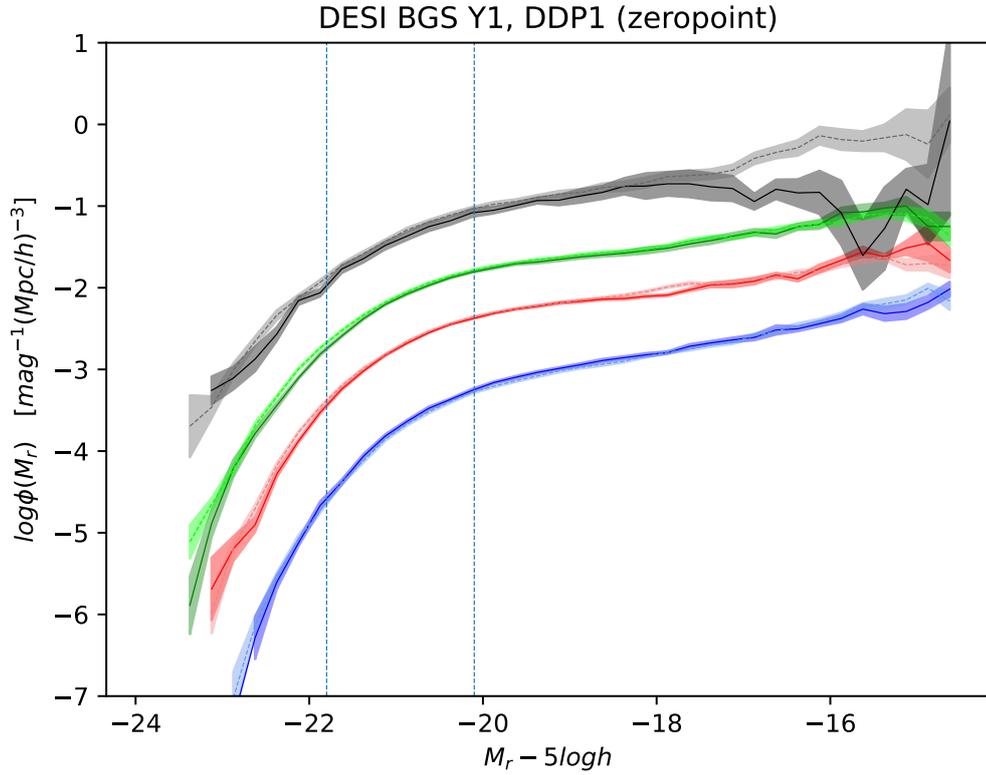Figure 4.7: The density-dependent $1/V_{\max}$ LFs for North and South. Here, we present the jackknife errors for each LF. For clarity, we use darker colour shades to represent the Norths LFs and lighter colour shades to represent the South LFs. Note that here we are presenting the zeropoint LFs (where the DDPs are not self-counting). The vertical dashed lines indicate the absolute magnitude limits of DDP1.
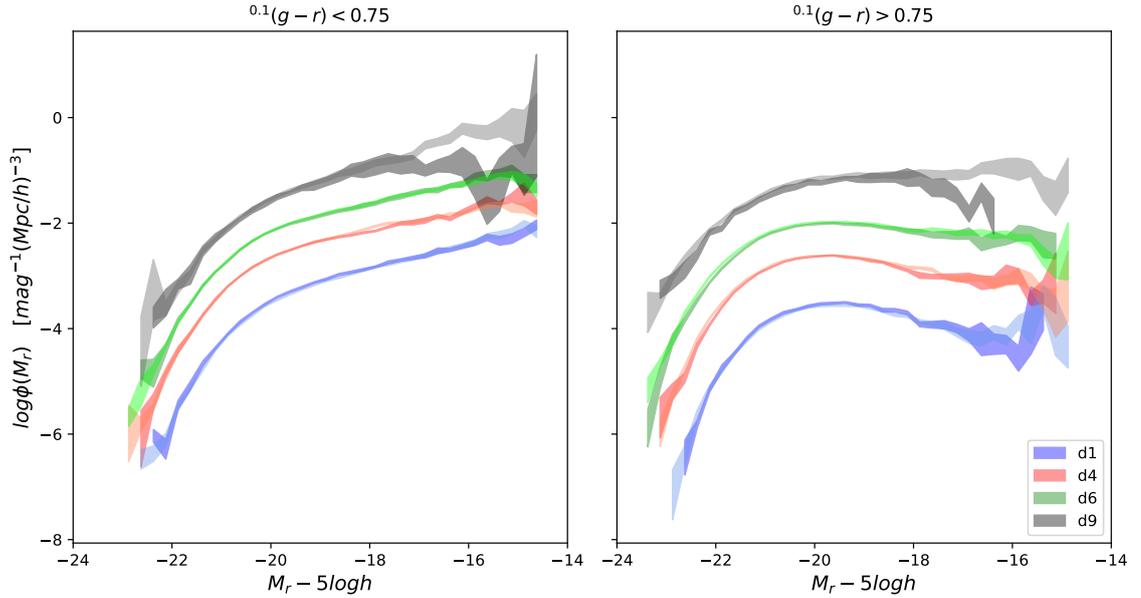
Figure 4.8: The $1/V_{\max}$ LFs split by both overdensity and colour. Red galaxies are defined as $^{0.1}(g-r) > 0.75$, while blue galaxies are defined as $^{0.1}(g-r) < 0.75$. Here, North LFs are presented in darker colour shades while South LFs are presented in lighter colour shades.

Our results are in broad agreement with the density-dependent LFs for GAMA DR2 as presented by McNaught-Roberts et al. (2014). We observe that the d4 and d6 δ_8 tier LFs have a very similar shape and normalisation to the GAMA LFs. There is some difference at d1, where we notice that the faint-end slope of the GAMA LF is steeper than that of our BGS Y1 results. In addition, the highest δ_8 tier seems to have a similar shape and faint-end slope, but is systmatically higher across all magnitudes.

We emphasise that we do not expect perfectly comparable results due to various differences between the datasets and methodology. For example, we observe different galaxies in different regions of the sky. In addition, we note that our methodology differs from the analysis in McNaught-Roberts et al. (2014) who instead used a fractional volume weighting on initial density-dependent SWML LFs to normalise the GAMA LFs.

We also present jackknife errors for our BGS Y1 density-dependent LFs (Fig. 4.7). We observe that the errors for both North and South LFs are generally very small, with the exception of the highest δ_8 tier (which contains fewer objects than the other

δ_8 tiers). Our errors are comparable to those presented for the GAMA results in McNaught-Roberts et al. (2014). We attribute this to the fact that both the GAMA and Y1 BGS samples have a similar number of galaxies once we have applied our fill factor cut.

We present further results of our density-dependent LFs. Fig. 4.8 presents the density-dependent LFs additionally split by colour, using a colour-split of $^{0.1}(g-r) = 0.75$. We continue to see good agreement between North and South LFs across all of our density tiers, with the sole exception again being the faint-end of the highest δ_8 tier. The jackknife errors remain small for these LFs.

4.4 Conclusions

In conclusion, we have developed a robust methodology for finding density-dependent LFs, which involves defining a Density Defining Population (DDP) of galaxies. This method adjusts the $1/V_{\mathrm{max}}$ estimator to account for the fact that different δ_8 tiers occupy different volumes within the DDP volume. Moreover, this methodology incorporates fill factors to account for the issue of boundary galaxies - where galaxies near the edge of the survey may receive lower local density estimates.

The results agree between the North and South galaxies for the BGS Y1 data, except at the highest δ_8 tier - which is expected due to the large sample variance at faint magnitudes for the highest overdensities. These results also generally agree well with prior results from GAMA presented in McNaught-Roberts et al. (2014). In particular, we observe that the middle δ_8 tiers, d4 and d6, are of an extremely similar shape and normalisation to that of GAMA. The major point of difference is the lowest δ_8 tier, d1, where the GAMA result has a different shape to our BGS Y1 result. We broadly observe that M^* decreases as a function of density, while α remains the same. These trends are also observed in GAMA by McNaught-Roberts et al. (2014).

In addition, we find small jackknife errors on most of the density-dependent LFs for both North and South. Although this shows that our LFs are well-constrained, these errors are comparable to those of GAMA. This is because currently our results

are limited by the small effective volume of galaxies that we have. This is a direct result of the Y1 BGS catalogue being highly incomplete which in turn causes there to be a large number of boundary galaxies (with $f < 0.8$) that are ignored in our density-dependent LFs. Although this is a major limitation for the results as presented here, we emphasise that we have developed a robust methodology which will yield more promising results in the future Y3 and Y5 catalogues, which will be far more complete.

Stellar Mass Function

In this Chapter, we present the Stellar Mass Function (SMF) of the BGS Y1 dataset. We compare this result to various SMF estimates, including the novel Photometric objects Around Cosmic webs (PAC) model of Xu et al. (2022a), outlined in Section 5.3.

5.1 Motivation

The Stellar Mass Function (SMF) quantifies the number density of galaxies as a function of stellar mass. The SMF is a useful tool for understanding the growth of the stellar population of galaxies. Most importantly, the SMF can act as a fundamental constraint on galaxy formation models. However, one must be cautious as the SMF is typically more model dependent than the luminosity function (LF) as a result of various assumptions that are made in order to estimate the stellar masses of a galaxy from observed quantities.

The SMF has been a point of investigation for many decades, but initial studies came with very large uncertainties. In recent years, a number of studies have made very precise measurements of the SMF at low redshifts (Baldry et al., 2012; Cole

et al., 2001; Wright et al., 2017). In addition, there have been major extensions to the accuracy of SMFs with larger sample sizes at higher redshifts (Weaver et al., 2023). For example, McLeod et al. (2021) present a number of ground-based estimates of the SMF, in addition to a combined HST and ground-based SMF up to the redshift bin $2.75 < z < 3.75$.

The SMF has become extremely important for both the calibration and validation of a number of galaxy simulations. For example, EAGLE (Crain et al., 2015; Schaye et al., 2015), one of the first cosmological hydrodynamic simulations to produce a realistic galaxy population, assumes the Planck 2013 cosmological parameters (Planck Collaboration et al., 2014) and then is calibrated to reproduce the observed SMF at $z = 0.1$. It is additionally calibrated to reproduce the galaxy mass-BH relation. (Crain et al., 2015). EAGLE has then been able to generate SMFs at higher redshifts, which acts as a validation test of the simulation. This is similarly true of a successor to EAGLE, the FLAMINGO simulation, which calibrates its subgrid modelling with the present SMF (Schaye et al., 2023). Due to the importance of the SMF in the calibration of these galaxy formation models, it is very important to have accurate estimates.

In addition, the SMF is a highly useful tool to understand galaxy formation processes. In particular, the difference between the SMF and the halo mass function offers an insight into physical processes such as SN and AGN feedback. The direct link between DM halo mass and stellar mass remains an open question, but various studies have sought to quantify the role of feedback in shaping the SMF. For example, Bower et al. (2012) finds that implementing a low wind speed expulsion model combined with AGN feedback in GALFORM yields a model SMF that agrees closely with the empirical SMF at high masses.

An additional motivation for investigating the SMF is to better characterise the nature of dark matter. As discussed in Chapter 1, the nature of dark matter is an unresolved problem in astrophysics. However, the nature of dark matter is extremely important in structure formation, with different models having been proposed, including warm dark matter (WDM). Due to the increased kinetic energy of WDM, WDM notably suppresses structure formation at low-mass scales when compared to

CDM. In particular, WDM would predict a decrease in the number density at the low-mass end of the SMF when compared to CDM. This has been quantified by a number of simulation that have investigated the evolution of the WDM and CDM SMFs over time (Dayal & Giri, 2024). As a result, measuring the SMF can act as an important constraint on WDM models.

5.2   SMF Methods

In order to calculate the SMF, it is first necessary to calculate the stellar mass of a galaxy from its photometry. Typically, this will make use of a Stellar Population Synthesis (SPS) model, e.g: Bruzual & Charlot (2003), to determine the SED produced by a given stellar population (defined by its age and metallicity) and then these can be fit to an observed galaxy SED to determine the stellar mass of the galaxy (and its age and metallicty). However, the results of this SED generation is highly model dependent, relying on a number of non-unique choices. SPS modelling typically operates as follows: To begin, an SPS model seeks to track the trajectory of a (main sequence) star as it evolves on the Hertzsprung-Russell diagram. In particular, these stellar tracks are sampled in both mass and time in order to generate stellar isochrones for given ages and metallicities. With some choice of isochrones selected, the SPS model will use a spectral library to convert stellar evolution calculations (e.g: effective temperatures) into SEDs, where the choice of spectral library may be empirical or theoretical.

Another consequential aspect is the choice of the Initial Mass Function (IMF) which empirically describes the mass distribution of a population of stars at formation. Popular examples of IMFs include Salpeter (1955), Kroupa (2002) and Chabrier (2003) which each differ significantly at the low-mass end of the distribution. The choice of IMF can make a significant difference in the normalisation of the mass-to-light ratio (Bernardi et al., 2017) and hence in the inferred stellar masses.

Finally, dust modelling has an important effect on the SED as ISM dust causes attenuation in the optical and UV ranges, but is a contributor to emission in the far-IR range. As such, different dust models will also result in different SEDs and

different inferred stellar masses.

In calculating the stellar masses, Xu et al. (2022b) makes use of the CIGALE SED code using the grz-band measurements. In addition, Xu et al. (2022b) uses the Bruzual & Charlot (2003) stellar population synthesis models, the initial mass function from Chabrier (2003), a delayed star formation history model, and the Calzetti et al. (2000) extinction law for dust reddening.

For illustration, we present in Fig. 5.1 the stellar masses calculated by Xu et al. (2022b) against those calculated by FSF (Moustakas et al., 2023). We see that although the majority of the galaxies have very similar stellar masses, there is a significant amount of scatter. Moreover, we observe visually different results in the SMF between the two models at the high-mass end (Fig. 5.2), demonstrating the impact of these assumptions. We observe that the $1/V_{\mathrm{max}}$ SMFs seems to fall off around $10^7 M_\odot$. We attribute this to the completeness limit of the survey, whereby fainter objects are more likely to have low galactic stellar mass.

5.3 Photometric Objects Around Cosmic Webs

In this Subsection, we outline the PAC method that was developed in Xu et al. (2022a) to calculate the SMF from the stellar masses.

To begin, we consider the quantification of galaxy clustering. Due to the hierarchical structure formation described in Chapter 1, galaxies are not randomly distributed in the universe with a Poisson distribution, but rather form clusters. This is most notably demonstrated with the visual presence of filaments and voids in the universe. In order to quantify this clustering, we define the spatial two-point correlation function as the excess probability that a galaxy is separated from another galaxy at some distance r. More formally, we write this as

$$dP = n[1 + \xi(r)]dV, \tag{5.1}$$

where n is the mean number density of the galaxy sample and dV is the volume element. In practice, this spatial two-point correlation function is found by using
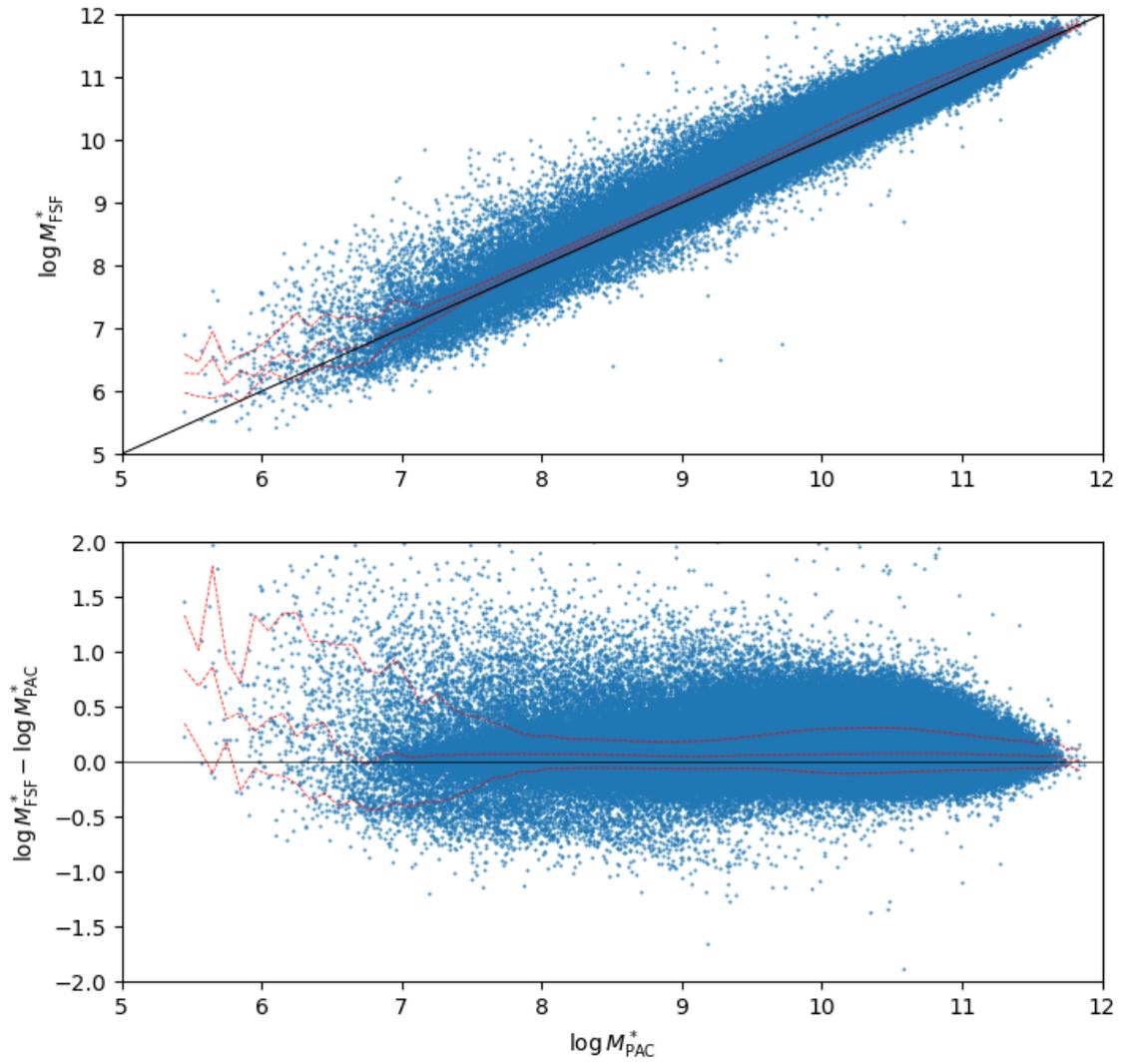
Figure 5.1: Comparison of the Xu et al. (2022b) stellar masses and FSF stellar masses. The solid black line represents the one-to-one line between the two quantites. The dashed red lines represent the 10th and 90th percentiles.
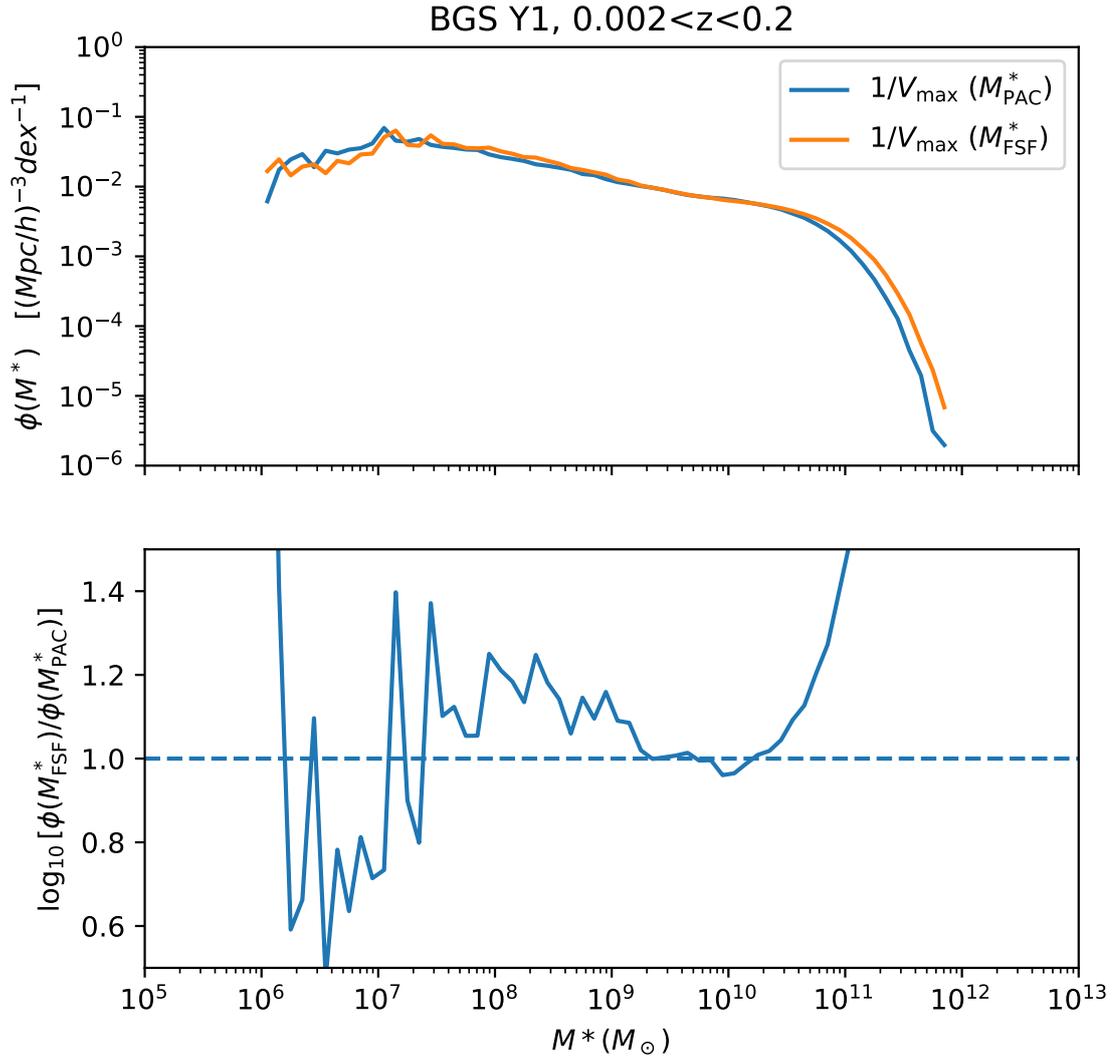
Figure 5.2: A comparison of the $1/V_{\mathrm{max}}$ SMF using the stellar masses detailed in Xu et al. (2022b) and the $1/V_{\mathrm{max}}$ SMF using the FSF stellar masses.

the estimator from Landy & Szalay (1993)

$$\xi(r) = \frac{DD - 2DR + RR}{RR}, \qquad (5.2)$$

where DD is the pairwise count (of pairs of separation r) from the data catalogue, RR is the pairwise count from the random catalogue, and DR is the pairwise count from the data catalogue with the random catalogue. Similarly, we can define the angular two-point correlation function, $\omega(\theta)$, as the excess probability of finding one galaxy within a given angular separation of another galaxy. This is defined in the same manner as Eqn. 5.2, except the pairwise counts occur within a given on-sky angular separation. More broadly, one can also find cross-correlations between two different populations. In this case, the correlation function of one population is defined with respect to another population.

Photometric Objects around Cosmic Webs (PAC) is a method developed by Xu et al. (2022a) that estimates the projected density distribution of photometric objects around spectroscopic objects. The motivation for this is that estimating the SMF at high redshifts has proved difficult due to a lack of spectroscopic data. Furthermore, photometric redshifts may have large uncertainties at high redshifts. This difficulty particularly affects the low-mass end of the SMF, resulting in high uncertainties on the SMF. As such, a method that could better estimate the number density of photometric objects would be particularly useful.

Let us consider two catalogues of galaxies - one is a set of galaxies with spectroscopic redshifts, and one is a set of galaxies with photometric redshifts (in Xu et al. (2022a), these are called pop_1 and pop_2 respectively).

PAC works by estimating the excess projected density distribution, $\bar{n}_2\omega_p(r_p)$, of the spectroscopic catalogue as

$$\bar{n}_2\omega_p(r_p) = \frac{\bar{S}_2}{r_1^2}w_{12,\mathrm{weight}}(\theta) \qquad (5.3)$$

where $\bar{S}_2$ is the mean angular surface density of the photometric survey, r_1 is the comoving distance to pop_1 and $w_{12,\mathrm{weight}}$ is the weighted angular cross-correlation function between the photometric and spectroscopic population. For clarification,

$\bar{n}_2$ is the mean number density of the photometric population, and $w_p(r_p)$ is the projected cross-correlation function of the spectroscopic and photometric catalogues with $r_p = r_1\theta$.

The methodology of PAC is as follows: First, the spectroscopic catalogue is divided into narrow redshift bins (which effectively form a continuum of values). Then, each galaxy in the photometric catalogue is assumed to have the same redshift as the mean redshift in each bin. From this, the stellar mass (and other physical properties) can be calculated using the SED and assuming that mean redshift. Then, in each redshift bin, one can use Eqn. 5.3 to calculate $n_2 w_p(r_p)$ for a given stellar mass. The power of this technique is that only photometric galaxies at the same redshift as the spectroscopic galaxies they are being correlated with contribute to the cross-correlation. Thus photometric galaxies that were assigned the wrong stellar mass because they were assigned the wrong redshift do not systematically bias the measurement of $n_2 w_p(r_p)$.

Armed with the measurement of $n_2 w_p(r_p)$, there are a number of ways of estimating n_2. One such method is to divide through by $w_p(r_p)$, where this makes the assumption that $w_p(r_p)$ estimated from the spectroscopic sample only depends on stellar mass. Alternatively, one can use a catalogue of subhalos in a cosmological N-body simulation with an assumed set of cosmological parameters (e.g. the Planck cosmology) and use the subhalo abundance matching (SHAM) technique (Vale & Ostriker, 2004) to yield a secondary joint constraint between n_2 and $w_p(r_p)$ and solve for n_2. In summary, PAC utilises the photometric catalogue to get very low stellar masses under a number of assumptions.

5.4   SMF Comparisons

The PAC method is very powerful as it enables the full depth of the parent photometric Legacy Survey catalogue to be used to infer stellar masses and not just the brighter subset that is in the spectroscopic BGS catalogue. However, it makes additional assumptions including either that $w_p(r_p)$ only depends on stellar mass or assumes both a particular cosmological model and the SHAM ansatz. It is therefore

useful to test these assumptions using a more direct estimate of the SMF. Hence we make use of stellar masses defined by Xu et al. (2022b) together with the $1/V_{\max}$ method described in Subsection 3.1.1 to find the SMF. We then compare this to the PAC estimates (both with the SHAM ansatz and the asssumption of $w_p(r_p)$ only depending on stellar mass). In addition we compare to the Driver et al. (2022) estimate from GAMA (modified as in Xu et al. (2022b) to account for the systematic difference in the inferred stellar masses due to the different choice of SPS model). These comparisons are shown in Fig 5.3.

The agreement between our direct $1/V_{\max}$ estimate and the published estimate from GAMA (Driver et al., 2022) is in good agreement down to $10^7 M_{\odot}$. Below this the BGS estimate becomes incomplete, while the smaller area of the GAMA estimate makes it susceptible to fluctuations due to cosmic variance. We can see that above $10^9 M_{\odot}$ there is excellent agreement between the two PAC estimates and our direct $1/V_{\max}$ estimate. This lends strong support to the PAC method and its ability to infer the SMF by cross-correlation techniques. However below $10^9 M_{\odot}$ we see the two PAC estimates diverge with our $1/V_{\max}$ estimate. The SHAM-based estimate stays consistent with our direct estimate down to $10^9 M_{\odot}$, but the one based on assuming $w_p(r_p)$ only depends on stellar mass becomes slightly less steep. The reason for this needs more investigation. It could be a cosmic variance issue as at this stellar mass the estimate of $w_p(r_p)$ from the spectroscopic sample is limited to a very small volume. Alternatively, it could be a systematic error induced by $w_p(r_p)$ depending on both stellar mass and galaxy colour combined with the completeness of the BGS sample at fixed stellar mass depending on colour. These issues will have to be understood and modelled before the PAC method can live up to its promise of enabling the SMF to be estimated well below $10^8 M_{\odot}$.

5.5 Conclusions

In this Chapter, we have compared the galaxy stellar masses for the BGS Y1 catalogue as estimated by Xu et al. (2022b) and Moustakas et al. (2023). Here, we highlight that there is a general level of agreement between the two estimates, with
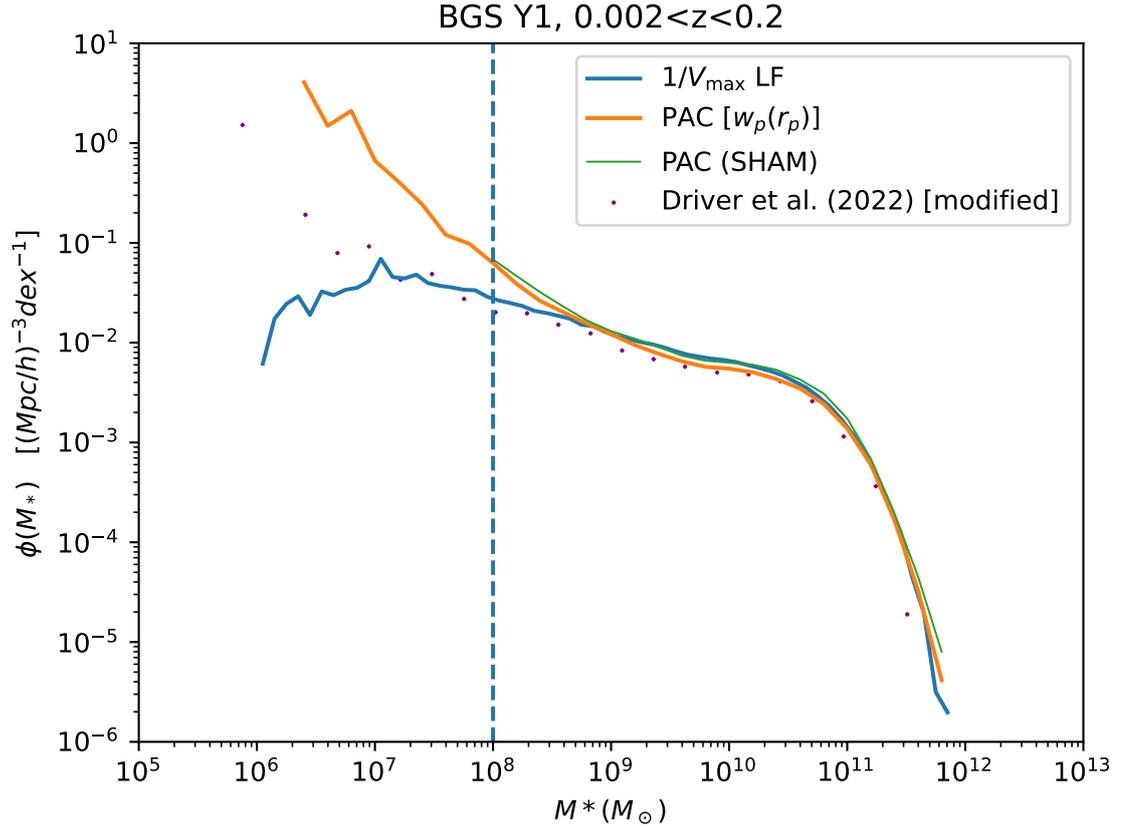
Figure 5.3: The $1/V_{\text{max}}$ SMF compared to the PAC model assuming $w_p(r_p)$ only depends on stellar mass and the PAC model using the SHAM ansatz. A modified empirical SMF from Driver et al. (2022) is presented, where the SMF has been shifted by the median offset between the individual stellar masses in Driver et al. (2022) and Xu et al. (2022b) for the galaxies they have in common. The dashed line at $10^8 M_\odot$ represents the current limit on the validity of the PAC model determined by where the choice of assumptions leads to diverging estimates.

the median residual being close to zero and the interquartile range remaining small over most magnitudes. Nevertheless, we highlight that there are a number of outliers in addition to the existence of non-negligible scatter, especially at the low-mass end.

We proceed to investigate the novel PAC method developed by Xu et al. (2022a). PAC is a powerful method which is able to exploit the signal in the cross-correlation of faint galaxies from the Legacy photometric surveys with the brighter DESI BGS galaxies with known spectroscopic redshift. In doing so, the PAC method is able to estimate the galaxy SMF to very low stellar masses.

As a test of the PAC methodology, we compare the PAC SMF to a direct $1/V_{\mathrm{max}}$ estimate, using the same methodology as detailed in Chapter 3. We find broadly good agreement at the high-mass end. However, below $10^9 M_\odot$ we find that there is significant divergence between the PAC estimate and our $1/V_{\mathrm{max}}$ estimate. As a result, we suggest that there are still some potential issues with PAC at low masses, and more careful modelling of the auto-correlation function of stellar mass selected samples is required. If PAC is able to do this and estimate stellar mass functions to as low as $10^6 M_\odot$ then it will be a useful constraint on DM models.

Conclusions

6.1 Overview

6.1.1 Global LFs

In this thesis, we have developed a methodology to determine the global $1/V_{\max}$ LF for the DESI BGS Y1 data in the r, g, z and w_1 bands. In particular, we have developed a polynomial-based method for defining k-corrections based on the FastSpecFit (FSF) k-corrections that allows for the calculation of $V_{\max}$ for all BGS galaxies. We have also developed a single-parameter e-correction derived from the $V/V_{\max}$ distribution of the galaxies. From this, we have found LFs that have small jackknife errorbars across a wide range of magnitudes. We confirm that we get good agreement with prior literature - in particular, we get comparable results to the complete GAMA data when assuming the same redshift limits and evolution. This means that we successfully account for the target and redshift incompleteness in the Y1 data. With this successful result, we can move on in the future to find the Y3 and Y5 LFs which will be more complete, and thus can yield even stronger constraints on the global LF.

6.1.2 Density-dependent LFs

We additionally present density-dependent LFs for DESI BGS Y1 in the r-band. The value of these LFs is that different galaxy formation models give different predictions as to the impact of environment on galaxy formation. As such, accurate measures of LFs split by environmental measures can act as a stringent test on those models.

We find that the North and South LFs agree very well across all density tiers, with the exception of the most dense tier, which disagrees at faint magnitudes. Given the small volume over which the faint magnitude objects can be detected, it is reasonable to expect that there may be a large discrepancy in finding dense regions between North and South. In addition, we show relatively good agreement between our DESI BGS Y1 LFs and the equivalent LFs from GAMA DR2, which is indicative of the fact that we have developed a robust pipeline which successfully takes into account the impact of boundaries and holes on the calculations of local density. There remains some difference in the shape of the lowest δ_8 tier between BGS Y1 and GAMA DR2. We note that due to the incompleteness of the BGS Y1 data, a large percentage of the data is removed from the analysis due to our condition on removing boundary galaxies. With the removal of boundary galaxies, the effective volume of the BGS survey is close to that of GAMA, meaning that the jackknife errorbars we derive are also comparable with GAMA. Nevertheless, we demonstrate a clear methodology for density-dependent LFs. The Y3 and Y5 results will be far more promising as they will have a far higher completeness, and as a result, a far higher effective volume. We therefore hope to better constrain the density-dependent LF errorbars in the future.

6.1.3 SMFs

In addition, we also present $1/V_{\mathrm{max}}$ Stellar Mass Functions (SMFs) on the BGS Y1 dataset. In doing so, we have used galaxy stellar masses from Xu et al. (2022b) in order to make a direct comparison to PAC, a powerful method of estimating the SMF to low masses that exploits the signal in the cross-correlation of faint galaxies from the Legacy photometric surveys with brighter galaxies with redshift in DESI

BGS.

We find reasonably good agreement at high stellar masses, but there is significant divergence at $M_\star < 10^9 M_\odot$. This suggests that whilst PAC is a promising model for constraining DM models, further work will need to be done to improve its low-mass estimate of the SMF, such as through more careful modelling of the auto-correlation function of stellar mass selected samples.

6.2　Future Directions

6.2.1　Evolutionary Modelling

There are nevertheless some points of interest that need further investigation. We note from our investigation of the LFs that the evolution is non-trivial. Relative to GAMA, the use of a single Q-value from $0.002 < z < 0.2$ gives consistent results. However, we find that a single value of Q is inadequate to fully capture the entire evolution over the redshift range of $0.002 < z < 0.6$, where we start to observe disagreement with the GAMA Schechter function (as well as disagreement with DESI LFs at different redshift ranges). This suggests that a more developed evolution model is required. Prior literature has divided galaxy samples into red and blue populations. We confirm that the DESI Y1 data appears to have two distinct Gaussian populations broadly separable by colour. Dividing the galaxy population into these two distinct colour classes and using two distinct Q-values (Q_{red}, Q_{blue}) yields more consistent results with the LFs diverging from each other to a lesser extent over larger redshift ranges. Nevertheless, there remains some divergence from the Loveday Schechter fit. It must also be considered that this introduces additional assumptions into our model. In particular, the two-parameter model assumes that the shapes of the red and blue LFs stay the same over time, which is not necessarily true. This suggests that while using two separate Q-values seems to improve our LF measurements, there is still ample room for improvement. We consider that calculating Q as a function of luminosity ($Q = Q(L)$) may be a better description of the evolution. The has the added advantage of removing the assumption that the shapes of the red and blue colour LFs stay the same with time. This is currently a

point of ongoing investigation.

6.2.2 Photometry

In addition to this, we find that the agreement between the North and South LFs is largely robust, especially around the knee of the LF. This suggests that we have successfully accounted for the photometry difference in North and South through the use of our k-correction model. However, we do note that there is a North/South discrepancy at the brightest end of the LF. Our preliminary investigations find that this is not a result of any rest-frame colour discrepancy between the North and South populations, as assigning the South rest-frame colour-distribution to the North does not remove the discrepancy. As such, we can discount the possibility of our k-correction model causing this problem. Nevertheless, this problem remains unsolved, but is likely a problem in the initial photometry. One suggestion is that this may be due to BASS and DECaLS assigning different light profiles to galaxies, resulting in galaxies receiving a higher r-band flux measurement in one region compared to the other. To investigate this, we hope to compare the Sersić indices of the overlap galaxies from BASS and DECaLS once this data becomes available to us to see if there is any notable difference.

6.2.3 Imaging

One of the advantages of our LFs is that we manage to extend our LFs to very faint magnitudes, with reasonable jackknife errorbars. Moreover, we observe an upturn in the LF at these faint magnitudes which is predominately caused by the red galaxies. We sought to better probe the accuracy of our LFs by conducting visual inspections of galaxies at these faint magnitudes. Our investigations into the imaging systematics suggest that many of the objects at $M_r - 5 \log h < -14$ suffer from fragmentation and other imaging problems, which we have sought to better quantify. For example, we find in our samples that a majority of objects at $M_r - 5 \log h < -10$ suffer from being problem galaxies. We attempt to adjust our LF using both a crude correction from our visual inspection statistics as well

as a neural network classification scheme. Both of these methods fail to completely remove the upturn suggesting that it is a real trend in the data. However, both of these methodologies have limitations and can be developed and tested further. As such, more work is needed to better assess the veracity of the faint-end of the LF. Further work will need to be conducted to investigate what other properties correlate with imaging problems. In doing so, this can hopefully lead to the development of a selection cut to remove these spurious objects.

6.2.4 DESI Comparison Work

At the moment, a major limitation of the above results is the incompleteness of the Y1 catalogue. Although this will affect the global LF errorbars, the most notable impact is on our density-dependent LFs. This is because we implement a fill factor cut of $f_j < 0.8$ which removes galaxies around the boundaries and holes of the survey. This means that while the DESI survey is highly extensive in terms of its target coverage, the current effective volume is comparable to that of GAMA, meaning that our density-dependent LF errorbars are also comparable. In the future, we hope to conduct this analysis on the Y3 and Y5 catalogues which will be far more complete and therefore will have a much larger effective volume.

We also note the challenge of making direct comparisons between LFs from different surveys as a result of the differing definition in magnitudes. For example, SDSS used Petrosian magnitudes, whereas DESI took the approach of calculating total magnitudes. There are different advantages and disadvantages of each method, such as the fact that total magnitudes are dependent on the S/N of the data. Regardless, this adds a further challenge that our DESI LFs may not be directly comparable with LFs from other surveys like SDSS. To investigate this, one future direction may be to calculate Petrosian magnitudes for DESI and observe how this affects the LF.

There is also the possibility of comparing the magnitudes of different surveys with DESI magnitudes. For example, EUCLID is deep enough to compare BGS galaxies. The utility of this is that we can further verify whether our LFs agree across different surveys. This includes probing some of the features in our LFs, such as the bright-end disparity between North and South. For example, if the

LFs end up agreeing with each other in a different survey with different magnitude definitions, then it constrains the problem to the DESI photometry.

6.2.5 Further Environmental Measures

In addition, there are other measures of environment which can be investigated. For this thesis, we focused on using δ_8 as a measurement for local density due to its comparability to prior GAMA results as well as the fact that δ_8 can be sensibly estimated from the incomplete Y1 dataset. Nevertheless, there are also other estimators that may be more physical. One such example would be a cosmic web classification scheme where galaxies could be classified as residing in filaments, sheets, clusters and voids. From this, LFs could be found for each classification and compared to prior results in the literature. We did not pursue this method as there is substantial difficulty in accurately classifying galaxies this way in the highly incomplete Y1 catalogue. However, as the Y3 and Y5 catalogues will be more complete, such an analysis will be possible in the future.

6.2.6 Simulations

As part of the impact of this thesis, we hope that the DESI LFs can be used to better constrain galaxy formation models. As detailed in Chapter 1, there are still a number of open questions in galaxy formation - including but not limited to the role of feedback. In order to further our analysis, a direct comparison of the density-dependent LFs to simulations such as FLAMINGO should be conducted. FLAMINGO has conducted hundreds of different simulations with varying parameters, including differing feedback parameters that account for different AGN and SN models. By comparing our LF results against the LFs from the set of FLAMINGO simulations, we may be able to place better constraints on different feedback parameters and rule out certain AGN models. Furthermore, even if these simulation runs are comparatively small in box size, they can nonetheless flag areas on interest so that future simulation runs can be larger with specific feedback parameters.

Bibliography

Aarseth S. J., Gott J. R. I., Turner E. L., 1979, The Astrophysical Journal, 228, 664 1.2

Adams C. B., et al., 2023, Axion Dark Matter (`arXiv:2203.14923`), `https://arxiv.org/abs/2203.14923` 1.1

Aggarwal C. C., 2018, Neural Networks and Deep Learning: A Textbook, 1st edn. Springer Publishing Company, Incorporated D.2

Albrecht A., 2006, in APS April Meeting Abstracts. APS Meeting Abstracts. p. G1.002 1

Bahcall N. A., 2015, Proceedings of the National Academy of Sciences, 112, 3173 1.1

Bailey S., et al., 2025, in prep. 2.2

Baldry I. K., et al., 2012, Monthly Notices of the Royal Astronomical Society, 421, 621 5.1

Barrena, R. Girardi, M. Boschin, W. Mardirossian, F. 2012, A&A, 540, A90 1.3

Barrow J. D., 1989, Quarterly Journal of the Royal Astronomical Society, 30, 163 1.1

Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, Monthly Notices of the Royal Astronomical Society, 436, 697 3.4

Bernardi M., et al., 2017, Monthly Notices of the Royal Astronomical Society, 475, 757 5.2

Blanchard A., Héloret J.-Y., Ilić S., Lamine B., Tutusaus I., 2024, The Open Journal of Astrophysics, 7 1.1

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, The Astrophysical Journal, 379, 440 1.2

Bower R. G., 1991, Monthly Notices of the Royal Astronomical Society, 248, 332 1.2

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, Monthly Notices of the Royal Astronomical Society, 370, 645 1.2

Bower R. G., Benson A. J., Crain R. A., 2012, Monthly Notices of the Royal Astronomical Society, 422, 2816 5.1

Bruzual G., Charlot S., 2003, Monthly Notices of the Royal Astronomical Society, 344, 1000 5.2

Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, The Astrophysical Journal, 533, 682 5.2

Chabrier G., 2003, Publications of the Astronomical Society of the Pacific, 115, 763 5.2

Cole S., 2011, Monthly Notices of the Royal Astronomical Society, 416, 739 3.1.3

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, Monthly Notices of the Royal Astronomical Society, 319, 168 1.2

Cole S., et al., 2001, Monthly Notices of the Royal Astronomical Society, 326, 255 5.1

Crain R. A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 450, 1937 1.2, 5.1

Croton D. J., et al., 2006, Monthly Notices of the Royal Astronomical Society, 365, 11 1.2

Croton D. J., et al., 2016, The Astrophysical Journal Supplement Series, 222, 22 1.2

DESI Collaboration in prep., DESI 2024 I: Data Release 1 of the Dark Energy Spectroscopic Instrument 1.4

DESI Collaboration et al., 2016, The DESI Experiment Part I: Science,Targeting, and Survey Design (arXiv:1611.00036) 1.4

DESI Collaboration et al., 2023, arXiv e-prints, p. arXiv:2306.06308 2.4

DESI Collaboration et al., 2024a, arXiv e-prints, p. arXiv:2404.03000 2.1

DESI Collaboration et al., 2024b, arXiv e-prints, p. arXiv:2404.03002 2.1

DESI Collaboration et al., 2024c, The Astronomical Journal, 168, 58 1.4

Dayal P., Giri S. K., 2024, Monthly Notices of the Royal Astronomical Society, 528, 2784 5.1

Dey A., et al., 2019, The Astronomical Journal, 157, 168 2.1

Dressler A., 1980, The Astrophysical Journal, 236, 351 1.2

Driver S. P., et al., 2011, Monthly Notices of the Royal Astronomical Society, 413, 971 1.3, 1.4, 2.1

Driver S. P., et al., 2022, Monthly Notices of the Royal Astronomical Society, 513, 439 5.4, 5.3

Efstathiou G., Eastwood J. W., 1981, Monthly Notices of the Royal Astronomical Society, 194, 503 1.2

Efstathiou G., Gratton S., 2020, Monthly Notices of the Royal Astronomical Society: Letters, 496, L91 1.1

Efstathiou G., Ellis R. S., Peterson B. A., 1988, Monthly Notices of the Royal Astronomical Society, 232, 431 3.1.1, 3.1.2

Ferdosi B. J., Buddelmeijer H., Trager S. C., Wilkinson M. H. F., Roerdink J. B. T. M., 2011, AA, 531, A114 4.2

Freedman W. L., 2021, The Astrophysical Journal, 919, 16 1.1

Gott J. Richard I., 1975, The Astrophysical Journal, 201, 296 1.2

Guo Q., et al., 2016, Monthly Notices of the Royal Astronomical Society, 461, 3457 1.2

Guth A. H., 1981, Phys. Rev. D, 23, 347 1.2

Guy J., et al., 2023, The Astronomical Journal, 165, 144 2.5.3, 2.5.3

Hahn C., et al., 2023, The Astronomical Journal, 165, 253 (document), 2.1, 2.1, 2.8

Helly J. C., Cole S., Frenk C. S., Baugh C. M., Benson A., Lacey C., Pearce F. R., 2003, Monthly Notices of the Royal Astronomical Society, 338, 913 1.2

Hockney R. W., Eastwood J. W., 1988, Computer simulation using particles 2.4, 2.5.3

Hogg D. W., Baldry I. K., Blanton M. R., Eisenstein D. J., 2002, The K correction (`arXiv:astro-ph/0210394`) 2.4, A.1.1, A.1.2

Hoyle F., Rojas R. R., Vogeley M. S., Brinkmann J., 2005, The Astrophysical Journal, 620, 618 1.3

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168 1.1

Huterer D., 2010, General Relativity and Gravitation, 42, 2177–2195 1.1

Kauffmann G., White S. D. M., Guiderdoni B., 1993, Monthly Notices of the Royal Astronomical Society, 264, 201 1.2

Kennicutt Robert C. J., 1998, The Astrophysical Journal, 498, 541 1.2

Kim H.-S., Power C., Baugh C. M., Wyithe J. S. B., Lacey C. G., Lagos C. D. P., Frenk C. S., 2012, Monthly Notices of the Royal Astronomical Society, 428, 3366 1.2

Kormendy J., Fisher D. B., Cornell M. E., Bender R., 2009, The Astrophysical Journal Supplement Series, 182, 216 1.2

Kraljic K., et al., 2017, Monthly Notices of the Royal Astronomical Society, 474, 547 1.3

Kroupa P., 2002, Science, 295, 82 5.2

Lan T.-W., et al., 2023, The Astrophysical Journal, 943, 68 2.2, 2

Landy S. D., Szalay A. S., 1993, The Astrophysical Journal, 412, 64 5.3

Lima J. A. S., Zanchin V., Brandenberger R., 1997, Monthly Notices of the Royal Astronomical Society, 291, L1 1.2

Linde A., 1982, Physics Letters B, 116, 335 1.2

Liu C. T., Capak P., Mobasher B., Paglione T. A. D., Rich R. M., Scoville N. Z., Tribiano S. M., Tyson N. D., 2008, The Astrophysical Journal, 672, 198 1.3

Longair M., 2008, Galaxy Formation. Astronomy and astrophysics library, Springer, https://books.google.co.uk/books?id=fRqMmwEACAAJ 1.2

Loveday J., et al., 2012, Monthly Notices of the Royal Astronomical Society, 420, 1239 2.4, 3.2, 3.5, 3.6, 3.7, 3.3, 3.7, B.2, B.2

Loveday J., et al., 2015, Monthly Notices of the Royal Astronomical Society, 451, 1540 3.2

Marinoni C., Bel J., Buzzi A., 2012, Journal of Cosmology and Astroparticle Physics, 2012, 036 1.1

Mathis H., White S. D. M., 2002, Monthly Notices of the Royal Astronomical Society, 337, 1193 1.3

McLeod D. J., McLure R. J., Dunlop J. S., Cullen F., Carnall A. C., Duncan K., 2021, Monthly Notices of the Royal Astronomical Society, 503, 4413 5.1

McNaught-Roberts T., et al., 2014, Monthly Notices of the Royal Astronomical Society, 445, 2125 1.3, 2.4, 3.2, 3.2, 3.10, 4.2, 4.3, 4.3, 4.6, 4.4

Melia F., 2022, Modern Physics Letters A, 37, 2250016 1.1

Mitchell P. D., et al., 2017, Monthly Notices of the Royal Astronomical Society, 474, 492 1.2

Moorman C. M., Vogeley M. S., Hoyle F., Pan D. C., Haynes M. P., Giovanelli R., 2015, The Astrophysical Journal, 810, 108 1.3

Moretti, A. et al., 2015, A&A, 581, A11 1.3

Moustakas J., Scholte D., Dey B., Khederlarian A., 2023, FastSpecFit: Fast spectral synthesis and emission-line fitting of DESI spectra, Astrophysics Source Code Library, record ascl:2308.005 (ascl:2308.005) (document), 2.4, 5.2, 5.5

Myers A. D., et al., 2023, The Astronomical Journal, 165, 50 1.4, 2.1

Naab T., Ostriker J. P., 2017, Annual Review of Astronomy and Astrophysics, 55, 59 1.2

Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, Monthly Notices of the Royal Astronomical Society, 396, 19 3.1.4

Peacock J. A., 1999, Cosmological Physics. Cambridge University Press 1.1

Peebles P. J. E., 1980, The large-scale structure of the universe 1.2

Perivolaropoulos L., Skara F., 2022, New Astronomy Reviews, 95, 101659 1.1

Planck Collaboration et al., 2014, A&A, 571, A16 5.1

Planck Collaboration et al., 2020, A&A, 641, A6 1.1

Press W. H., Schechter P., 1974, The Astrophysical Journal, 187, 425 1.2

Rees M. J., Ostriker J. P., 1977, Monthly Notices of the Royal Astronomical Society, 179, 541 1.2

Riess A. G., et al., 1998, The Astronomical Journal, 116, 1009 1.1

Riess A. G., et al., 2022, The Astrophysical Journal Letters, 934, L7 1.1

Robotham A., Phillipps S., De Propris R., 2010, Monthly Notices of the Royal Astronomical Society, 403, 1812 1.3

Ross A., et al., 2025, in prep. 2.3, 2.5.2

Roszkowski L., Sessolo E. M., Trojanowski S., 2018, Reports on Progress in Physics, 81, 066201 1.1

Rubin V. C., Ford W. K. J., Thonnard N., 1980, The Astrophysical Journal, 238, 471 1.1

Ruiz-Macias O., et al., 2020, Research Notes of the American Astronomical Society, 4, 187 2.1

Ruiz-Macias O., et al., 2021, Monthly Notices of the Royal Astronomical Society, 502, 4328 2.1

Salpeter E. E., 1955, The Astrophysical Journal, 121, 161 5.2

Sandage A., Tammann G. A., Yahil A., 1979, The Astrophysical Journal, 232, 352 3.1.2

Schaye J., et al., 2015, Monthly Notices of the Royal Astronomical Society, 446, 521 5.1

Schaye J., et al., 2023, Monthly Notices of the Royal Astronomical Society, 526, 4978 1.2, 5.1

Schlafly E. F., Meisner A. M., Green G. M., 2019, The Astrophysical Journal Supplement Series, 240, 30 2.1

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, The Astrophysical Journal, 500, 525 2.1

Schmidt M., 1968, The Astrophysical Journal, 151, 393 3.1.1

Silber J. H., et al., 2022, The Astronomical Journal, 165, 9 1.4

Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, Monthly Notices of the Royal Astronomical Society, 391, 481 1.2

Strauss M. A., et al., 2002, The Astronomical Journal, 124, 1810 2.1

Tempel, E. Saar, E. Liivamägi, L. J. Tamm, A. Einasto, J. Einasto, M. Müller, V. 2011, A&A, 529, A53 1.3

Trentham N., Tully R. B., 2002, Monthly Notices of the Royal Astronomical Society, 335, 712 1.3

Vale A., Ostriker J. P., 2004, Monthly Notices of the Royal Astronomical Society, 353, 189 5.3

Wald R. M., 1984, General Relativity. The University of Chicago Press 1.1

Weaver J. R., et al., 2023, A&A, 677, A184 5.1

Weinberg S., 2008, Cosmology 1.1, 1.1

White S. D. M., Frenk C. S., 1991, The Astrophysical Journal, 379, 52 1.2

Wright E. L., et al., 2010, The Astronomical Journal, 140, 1868 2.1

Wright A. H., et al., 2017, Monthly Notices of the Royal Astronomical Society, 470, 283 5.1

Xu K., Zheng Y., Jing Y., 2022a, The Astrophysical Journal, 925, 31 (document), 5, 5.3, 5.3, 5.5

Xu K., Jing Y. P., Gao H., 2022b, The Astrophysical Journal, 939, 104 (document), 5.2, 5.1, 5.2, 5.4, 5.5, 5.3, 6.1.3

Zehavi I., et al., 2005, The Astrophysical Journal, 630, 1 2.4

k-corrections

A.1 k-corrections

In this appendix, we derive some of the general equations for k-corrections. In particular, we present the result for reference-redshift transformations which allow us to transform $^{0.1}k_r$ to $^{0.0}k_r$ if required.

A.1.1 Reference Redshift

Formally, we note that the definition of a k-correction may be defined following from Hogg et al. (2002) as

$$^{z_{\mathrm{ref}}}k(z) = -2.5 \log_{10} \left[(1+z) \frac{\int f_\nu(\nu/(1+z_{\mathrm{ref}}))\, R(\nu)\, d\nu/\nu}{\int f_\nu(\nu/(1+z))\, R(\nu)\, d\nu/\nu} \right], \tag{A.1.1}$$

where $f_\nu(\nu)$ is the SED of the object and $R(\nu)$ is the filter bandpass. For ease, we can rewrite this equation for a reference redshift z_{ref} as

$$^{z_{\mathrm{ref}}}k(z) = -2.5\log_{10}(1+z) + A(z_{\mathrm{ref}}) - A(z) \qquad (\mathrm{A.1.2})$$

where we define

$$A(z) = -2.5\log_{10}\int f_\nu\Big(\frac{\nu}{1+z}\Big)\,R(\nu)\,\frac{d\nu}{\nu} \qquad (\mathrm{A.1.3})$$

We can then define the following:

$$^{z_2}k(z) - {}^{z_{\mathrm{ref}}}k(z) + {}^{z_{\mathrm{ref}}}k(z_2) = A(z_2) - A(z_{\mathrm{ref}}) - 2.5\log_{10}(1+z_2) - A(z_2) + A(z_{\mathrm{ref}})$$
$$= -2.5\log_{10}(1+z_2),$$

$$(\mathrm{A.1.4})$$

and rearrange to get

$$\boxed{^{z_2}k(z) = {}^{z_{\mathrm{ref}}}k(z) - {}^{z_{\mathrm{ref}}}k(z_2) - 2.5\log_{10}(1+z_2).} \qquad (\mathrm{A.1.5})$$

A.1.2   Band-shifting

Furthermore, we can use Equation 8 of Hogg et al. (2002) for the case of converting from observed (Legacy) filter $L(\nu)$ for an object at redshift z to reference (Sloan) filter $S(\nu)$ at reference redshift z_{ref}. The k-correction is given by

$$^{z_{\mathrm{ref}}}k_{\mathrm{SL}}(z) = -2.5\log_{10}\left[(1+z)\frac{\int f_\nu(\nu/(1+z_{\mathrm{ref}}))\,S(\nu)\,d\nu/\nu}{\int f_\nu(\nu/(1+z))\,L(\nu)\,d\nu/\nu}\right], \qquad (\mathrm{A.1.6})$$

where $f_\nu(\nu)$ is the object's SED. Here, we have made the assumption that both filters are normalised such that $\int L(\nu)d\nu/\nu = \int S(\nu)d\nu/\nu = 1$.

Now let us specialise to the case where the two filters are identical apart from an offset in $\log\nu$, i.e. $S(\nu) = L(\beta\nu)$. From this

$$^{z_{\mathrm{ref}}}k_{\mathrm{SL}}(z) = -2.5\log_{10}\left[(1+z)\frac{\int f_\nu(\nu'/(1+z_{\mathrm{ref}}))\,L(\beta\nu')\,d\nu'/\nu'}{\int f_\nu(\nu/(1+z))\,L(\nu)\,d\nu/\nu}\right] \qquad (\mathrm{A.1.7})$$

By making the substitution that $\nu = \beta\nu'$, this becomes

$$^{z_{\text{ref}}}k_{\text{SL}}(z) = -2.5\log_{10}\left[(1+z)\frac{\int f_\nu(\nu/[\beta(1+z_{\text{ref}})])\,L(\nu)\,d\nu/\nu}{\int f_\nu(\nu/(1+z))\,L(\nu)\,d\nu/\nu}\right]. \qquad (A.1.8)$$

By comparing to equation A.1.1 (in which we substitute $R \to L$)

$$^{z_{\text{ref}}}k_{\text{LL}}(z) = -2.5\log_{10}\left[(1+z)\frac{\int f_\nu(\nu/(1+z_{\text{ref}}))\,L(\nu)\,d\nu/\nu}{\int f_\nu(\nu/(1+z))\,L(\nu)\,d\nu/\nu}\right], \qquad (A.1.9)$$

$$^{z_{\text{ref}}}k_{\text{SL}}(z) = {}^{z'_{\text{ref}}}k_{\text{LL}}(z), \qquad \text{where} \qquad 1 + z'_{\text{ref}} = \beta(1+z_{\text{ref}}). \qquad (A.1.10)$$

But the k-corrections we have actually tabulated from GAMA are

$$^{z_{\text{ref}}}k_{\text{SS}}(z) = -2.5\log_{10}\left[(1+z)\frac{\int f_\nu(\nu/(1+z_{\text{ref}}))\,S(\nu)\,d\nu/\nu}{\int f_\nu(\nu/(1+z))\,S(\nu)\,d\nu/\nu}\right], \qquad (A.1.11)$$

with $z_{\text{ref}} = 0.1$. Substituting using $S(\nu) = L(\beta\nu)$ we have

$$^{z_{\text{ref}}}k_{\text{SS}}(z) = -2.5\log_{10}\left[(1+z)\frac{\int f_\nu(\nu'/(1+z_{\text{ref}}))\,L(\beta\nu')\,d\nu'/\nu'}{\int f_\nu(\nu'/(1+z))\,L(\beta\nu')\,d\nu'/\nu'}\right], \qquad (A.1.12)$$

Substituting $\nu = \beta\nu'$ this becomes

$$^{z_{\text{ref}}}k_{\text{SS}}(z) = -2.5\log_{10}\left[\beta(1+z)\frac{\int f_\nu(\nu/[\beta(1+z_{\text{ref}})])\,L(\nu)\,d\nu/\nu}{\int f_\nu(\nu/[\beta(1+z)])\,L(\nu)\,d\nu/\nu}\right] + 2.5\log_{10}(\beta), \qquad (A.1.13)$$

By comparing to Eqn. A.1.9, we can write this as

$$^{z_{\text{ref}}}k_{\text{SS}}(z) = {}^{z'_{\text{ref}}}k_{\text{LL}}(z') + 2.5\log_{10}(\beta) \qquad (A.1.14)$$

where $1 + z'_{\text{ref}} = \beta(1+z_{\text{ref}})$ and $1 + z' = \beta(1+z)$. This is equivalent to

$$^{z_{\text{ref}}}k_{\text{LL}}(z) = {}^{z''_{\text{ref}}}k_{\text{SS}}(z'') - 2.5\log_{10}(\beta) \qquad (A.1.15)$$

where $1 + z''_{\text{ref}} = (1+z_{\text{ref}})/\beta$ and $1 + z'' = (1+z)/\beta$.

Hence, using Eqn. A.1.10 we have

$$\boxed{^{z_{\mathrm{ref}}}k_{\mathrm{SL}}(z) = {^{z_{\mathrm{ref}}}k_{\mathrm{SS}}(z''')} - 2.5\log_{10}(\beta)} \qquad \text{where} \qquad 1+z''' = (1+z)/\beta. \quad \text{(A.1.16)}$$

Further Luminosity Functions

B.1 $0.02 < z < 0.6$

In this section, we present LFs in the redshift range $0.02 < z < 0.6$ (as compared to $0.002 < z < 0.6$ in Fig. 3.5 and Fig. 3.6). This is because at very low redshifts, there may be a multitude of factors that affect the completeness of the sample. At a slightly higher limit of $z > 0.02$, the completeness limits that we present are better established. These new LFs are presented in Fig. B.1.

B.2 $Q = 0.78$ **LFs**

In this section, we present a set of r-band LFs using $Q = 0.78$ for our e-correction model. We observe good agreement between the GAMA r-band Schechter fit from Loveday et al. (2012) for $0.002 < z < 0.1$, however, there is some divergence around the knee of the LF as we extend to higher redshifts. These LFs are presented in Fig. B.2.
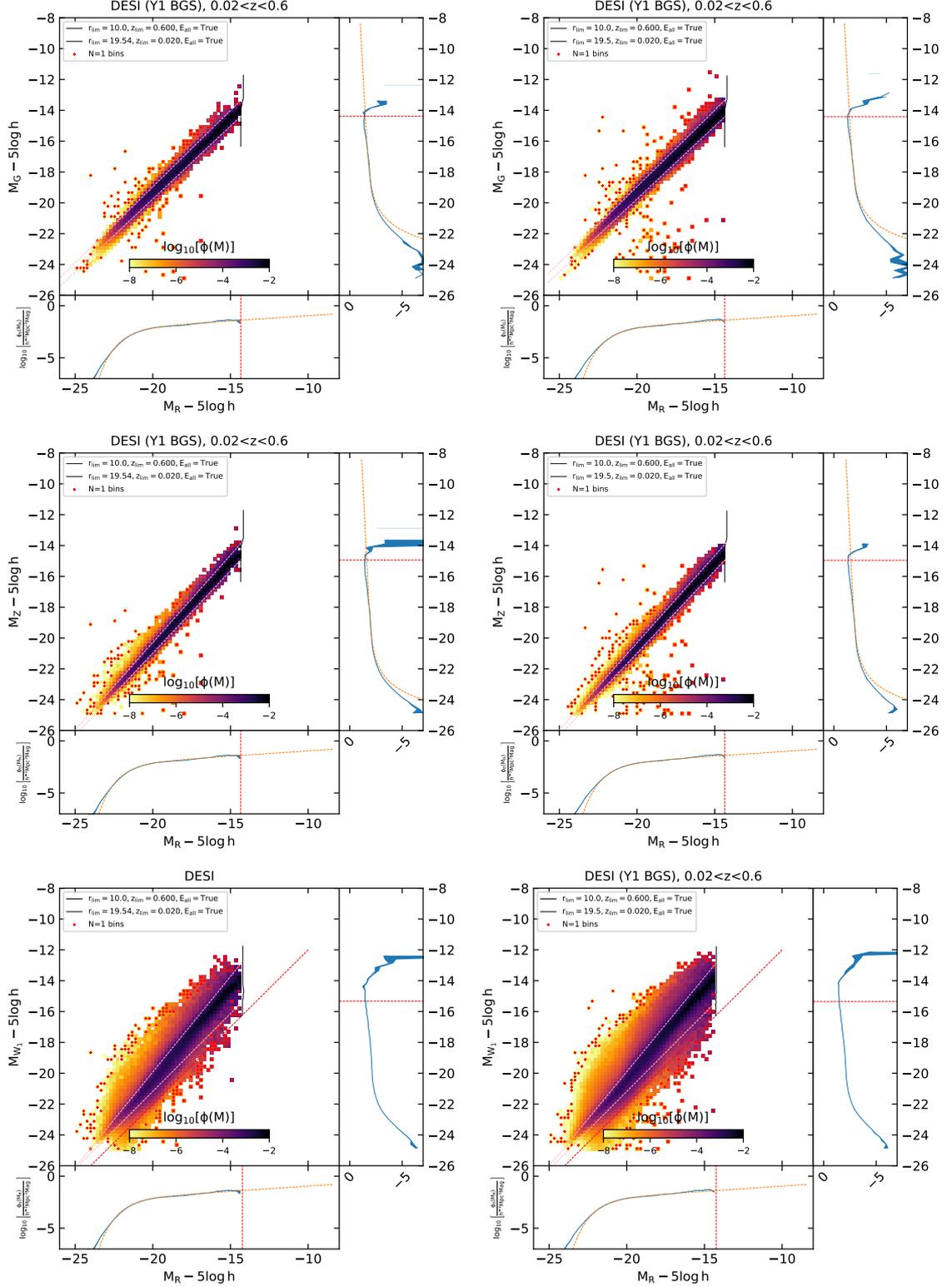
Figure B.1: Bivariate LFs for North (left column) and South (right column) in the g, z and w_1 bands. Here, the redshift limits are $0.02 < z < 0.6$. The black curves represent the limit of the survey. The dashed red lines give completeness limits for the r-band and g-band luminosity functions based on the intersection of the completeness curves and the 95th percentile contour of the bivariate LF. The $r - w_1$ bivariate LFs have an additional red-dashed line showing a selection cut that is later incorporated to remove spurious objects.
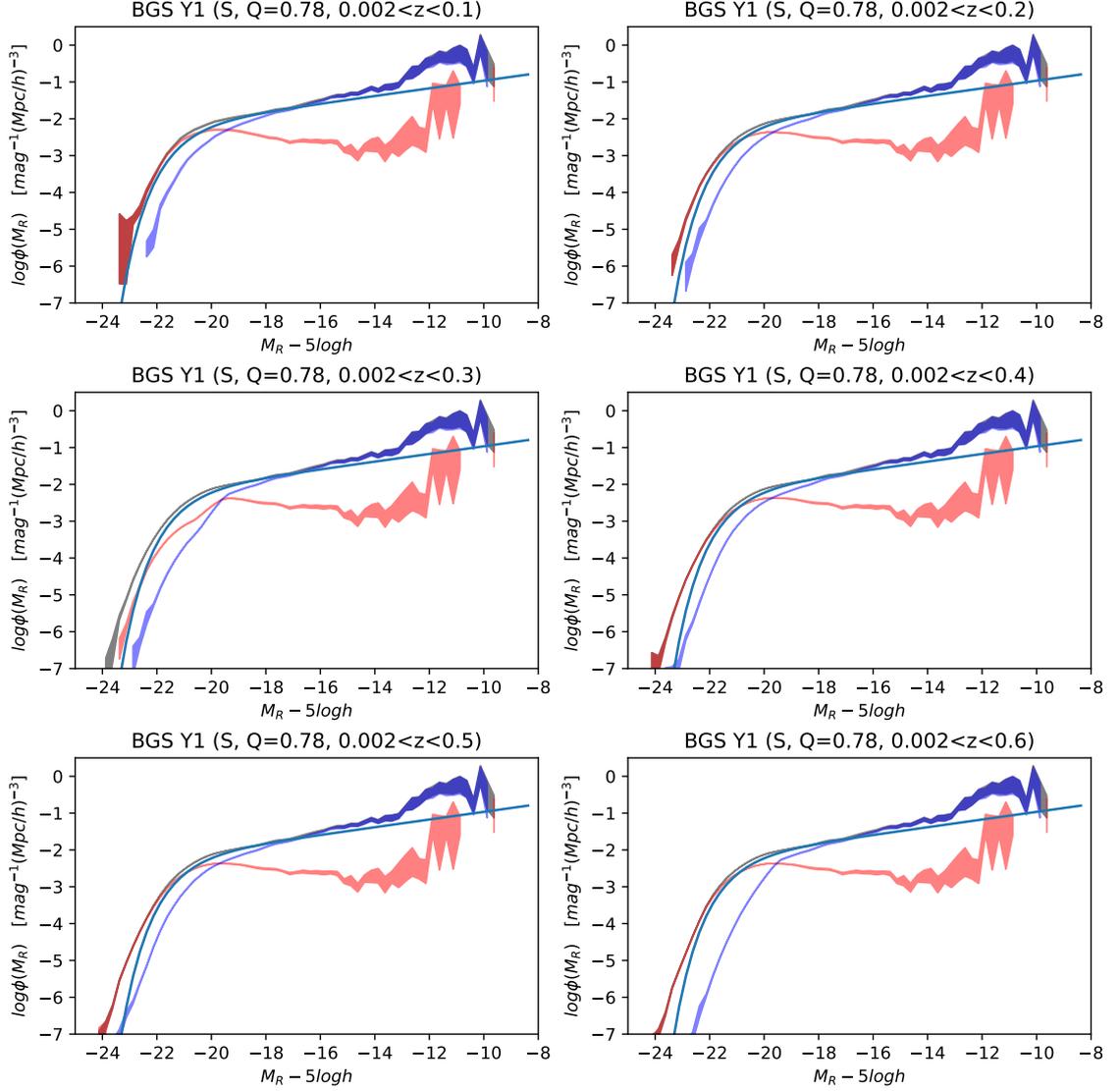
Figure B.2: $1/V_{\max}$ r-band LFs for the South presented across different redshift ranges. Here, we use $Q = 0.78$ as our e-correction model. The total LF is presented in grey. We additionally present the red and blue LFs, as well as the r-band GAMA Schechter fit from Loveday et al. (2012).

Bright-end LF Analysis

C.1 FSF Plots

An additional concern with our LFs is that there is a discrepancy between the North and South r-band LFs at the bright end. We present our investigation on the potential cause of this difference.

First, we consider the possibly that there may be an inherent discrepancy in the FSF k-correction fits. To do this, we make use of the overlap region in the DESI survey - that is, the region where there exists both North and South photometry data for the same objects. This region is shown in Figure C.1.

First, we compare the apparent magnitudes (g and r) in BASS and DECaLS. (Fig. C.2). We note that there is a strong correlation between the g-band apparent magnitudes for BASS and DECaLS using Tractor, with a median residual of -0.0585. The same is true for the r-band apparent magnitudes, with a median residual of -0.0229. There does appear to be a systematic difference between BASS r-band apparent magnitudes and DECaLS r-band apparent magnitudes (and likewise for the g-band), but this is expected given the different filter functions.

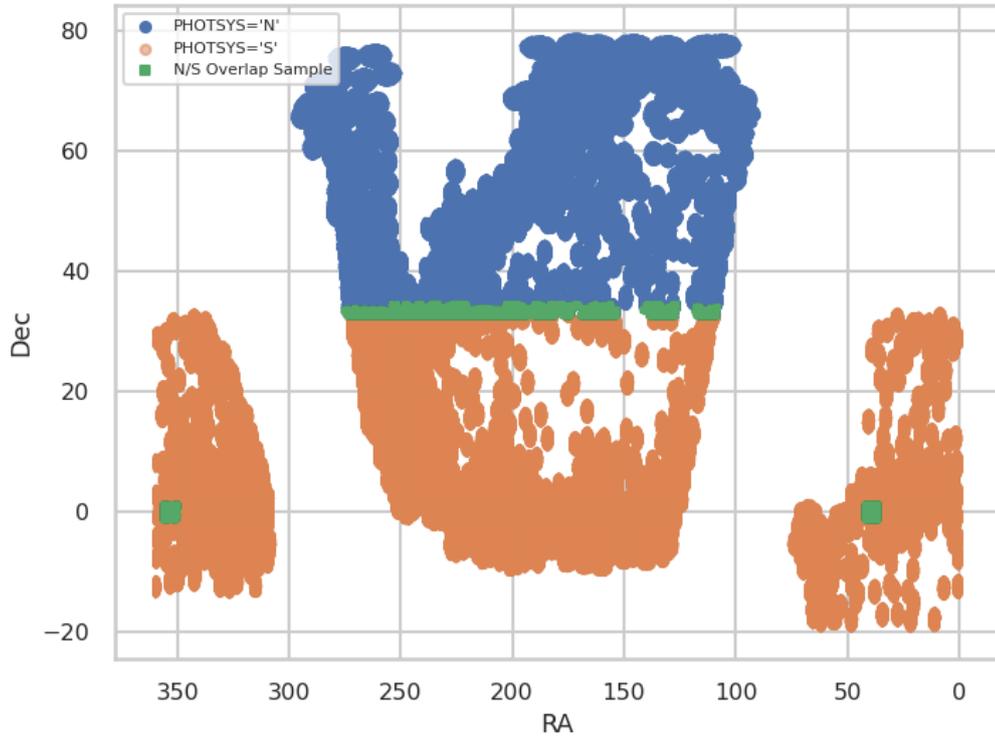We investigate whether there are differences in BASS and DECaLS for the r-

Figure C.1: A plot showing the North and South regions of the DESI survey. The green region represents an overlap region - the area where objects received both North and South photometry fluxes (based on BASS/MzLS and DECaLS).
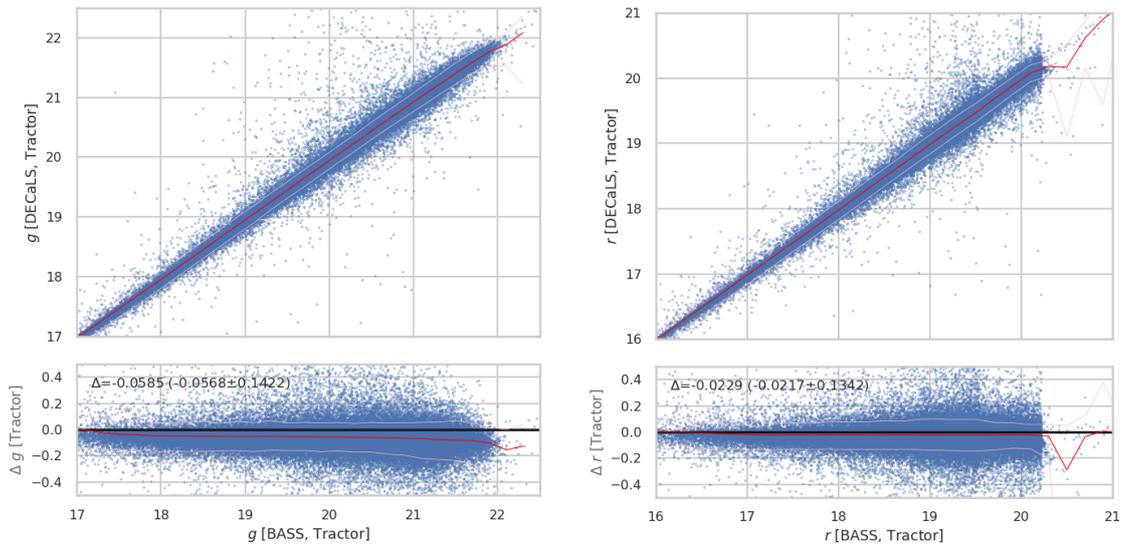

Figure C.2: A comparison of BASS and DECaLS apparent magnitudes, showing the corresponding residual plot. Left: The BASS/DECaLS comparison for the g-band apparent magnitudes. Right: The BASS/DECaLS comparison for the r-band apparent magnitudes.
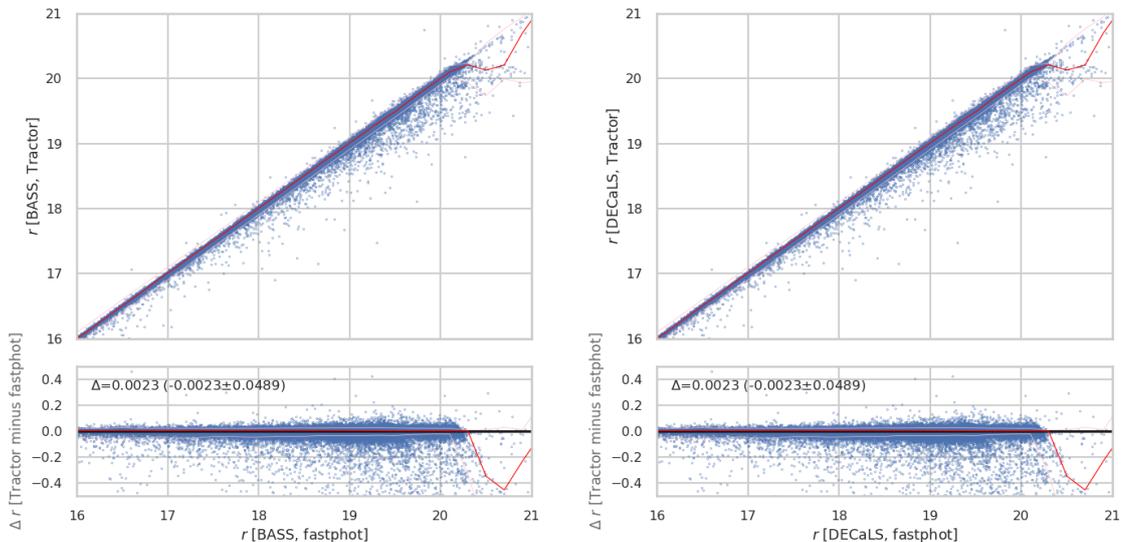
Figure C.3: Left: The r-band absolute magnitudes for BASS vs the fastphot model. Right: The r-band absolute magnitudes for DECaLS vs the fastphot model.

band and g-band apparent magnitudes when using Tractor compared to using the fastphot model. The fastphot model differs from the fastspec model as the fastphot model only fits the broadband photometry. This is in contrast to the fastspec model which also fits the spectrophotometry with an aperture correction.

Moreover, there is also a strong correlation between the Tractor magnitudes and the fastphot magnitudes (for both BASS and DECaLS in both the r and g-band), indicating that the fastphot model has good concurrent validity (Fig. C.3, C.4). In particular, the fastphot model agrees with the Tractor model for both the g and r bands to within 0.01 mag for both BASS and DECaLS. This shows that fastphot is a very good fit to the data.

We then proceed to investigate the absolute magnitudes found using BASS and DECaLS. Although the magnitudes broadly agree at fainter magnitudes, we observe a difference at bright magnitudes in the r-band. Further investigation shows that this difference predominately occurs for red objects of $^{0.1}(g - r) > 0.75$, where we are careful to note that there exist an adequate number of blue galaxies at bright magnitudes to demonstrate that these does not hold for blue galaxies (Fig. C.5).

We investigate the effect that this magnitude difference has on the LF by 'perturbing' all of the galaxies in each magnitude bin by its BASS/DECaLS residual.
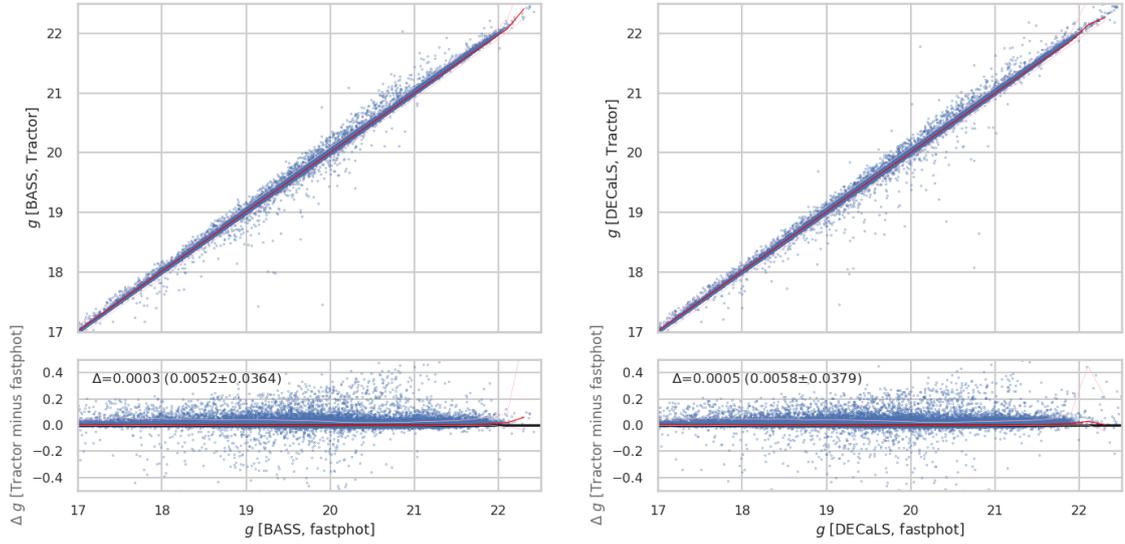
Figure C.4: Left: The g-band absolute magnitudes for BASS vs the fastphot model. Right: The g-band absolute magnitudes for DECaLS vs the fastphot model.
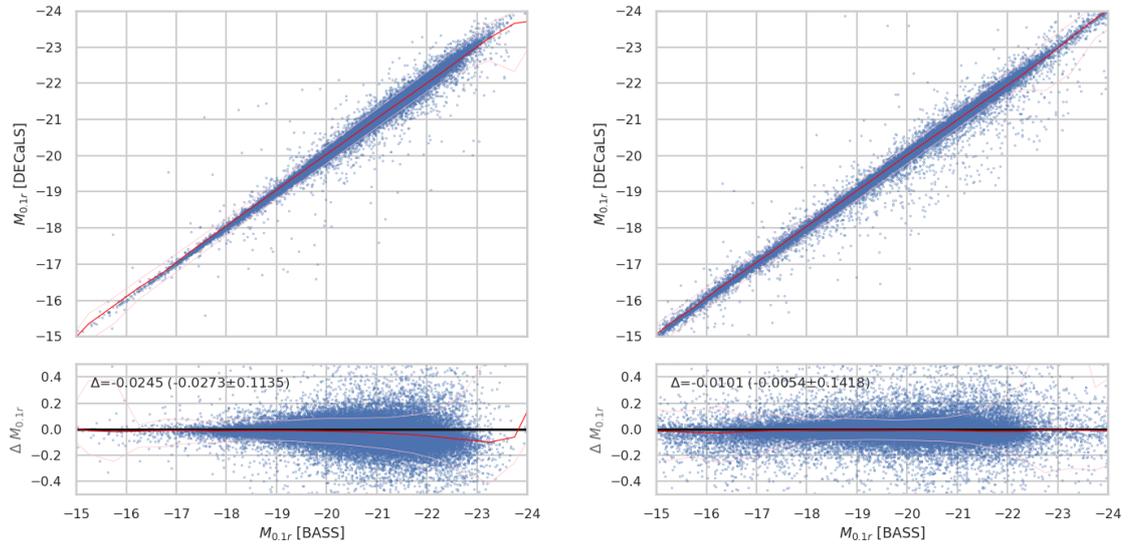

Figure C.5: Left: A scatter plot showing the r-band absolute magnitudes for BASS vs DECaLS for the red galaxies, with a corresponding residual plot. Right: The equivalent plot but for blue galaxies.
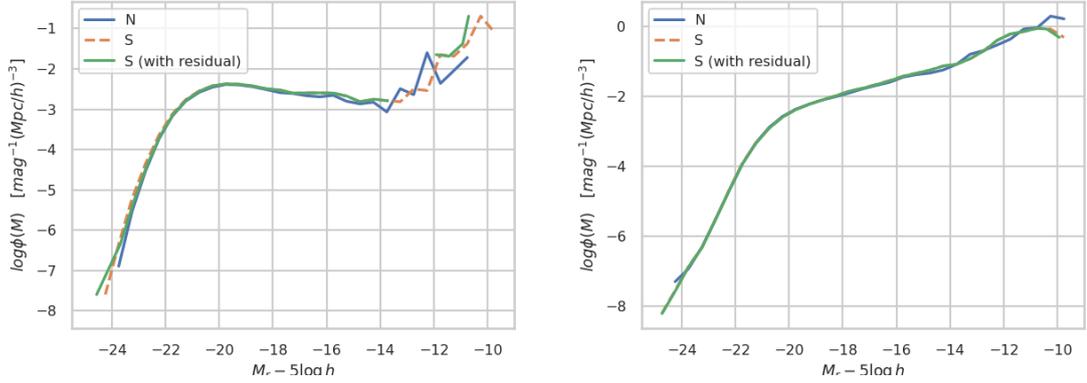
Figure C.6: Left: a plot of the North and South r-band LFs for red galaxies. Here. each galaxy in the South has its M_r changed by an offset according to the residual of the BASS/DECaLS M_r for that magnitude bin. This yields the South LF 'with residual' offset. Right: The same LFs but for the blue galaxies.

That is:

$$M_{r,i} = M_{r,i} + M_{r,\text{BASS},j} - M_{r,\text{DECaLS},j} \tag{C.1.1}$$

where i corresponds to each galaxy and j corresponds to each luminosity bin. As shown in Figure C.5, this adjustment makes a notable difference to the LF at bright magnitudes, bringing the North LF far closer to the South LF.

Imaging Results

D.1 Imaging

As discussed in Section 3.6, we investigated the imaging of some of the faint absolute magnitude galaxies in order to better understand the limitations of our data. In particular, we were interested in better quantifying the completeness of the LF at faint absolute magnitudes. Presented here are additional figures showing comparison plots that were generated to assess the imaging in different magnitude bins (Fig. D.1, D.2, D.3).

D.2 Neural Network Classification

In order to further investigate this problem, we attempt to use a neural network to evaluate a larger proportion of the objects. In doing so, we can further quantify the impact of poor imaging on the bright end of the LF and can gain a better idea of how many galaxies might be affected. In addition, by classifying all of the galaxies of magnitude $M_r - 5\log_{10} h < -15$, we can better determine whether there truly exists an upturn in the faint-end of the LF.
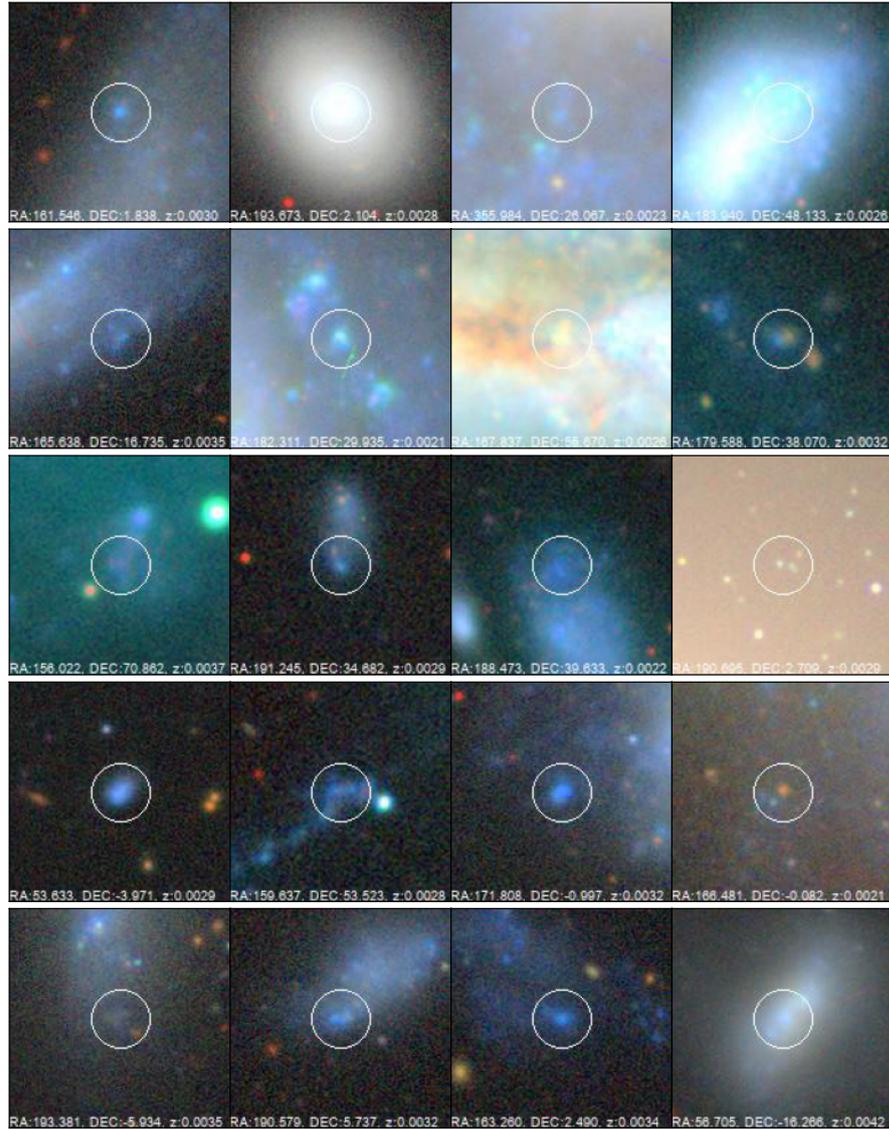
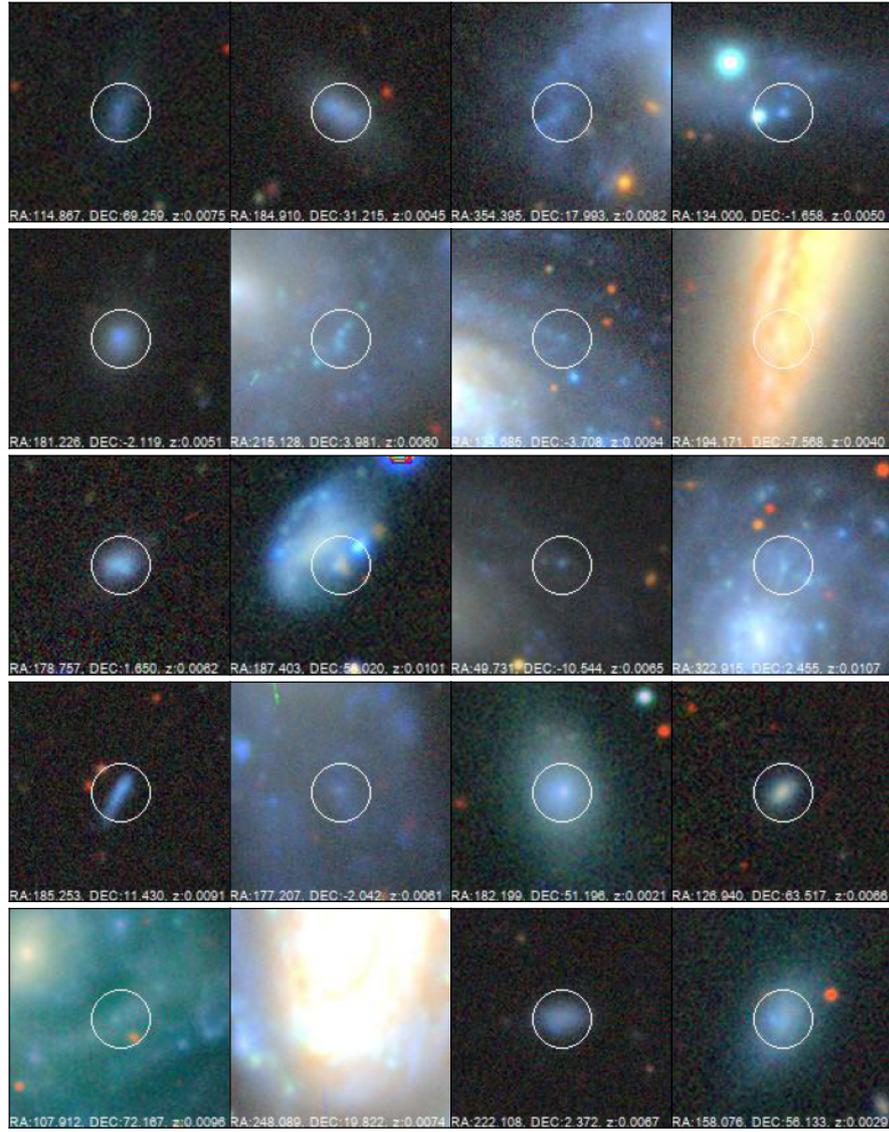Figure D.1: Plot showing a random sample of 20 galaxies in the range $-11 < M_r - 5\log_{10} h < -10.75$.

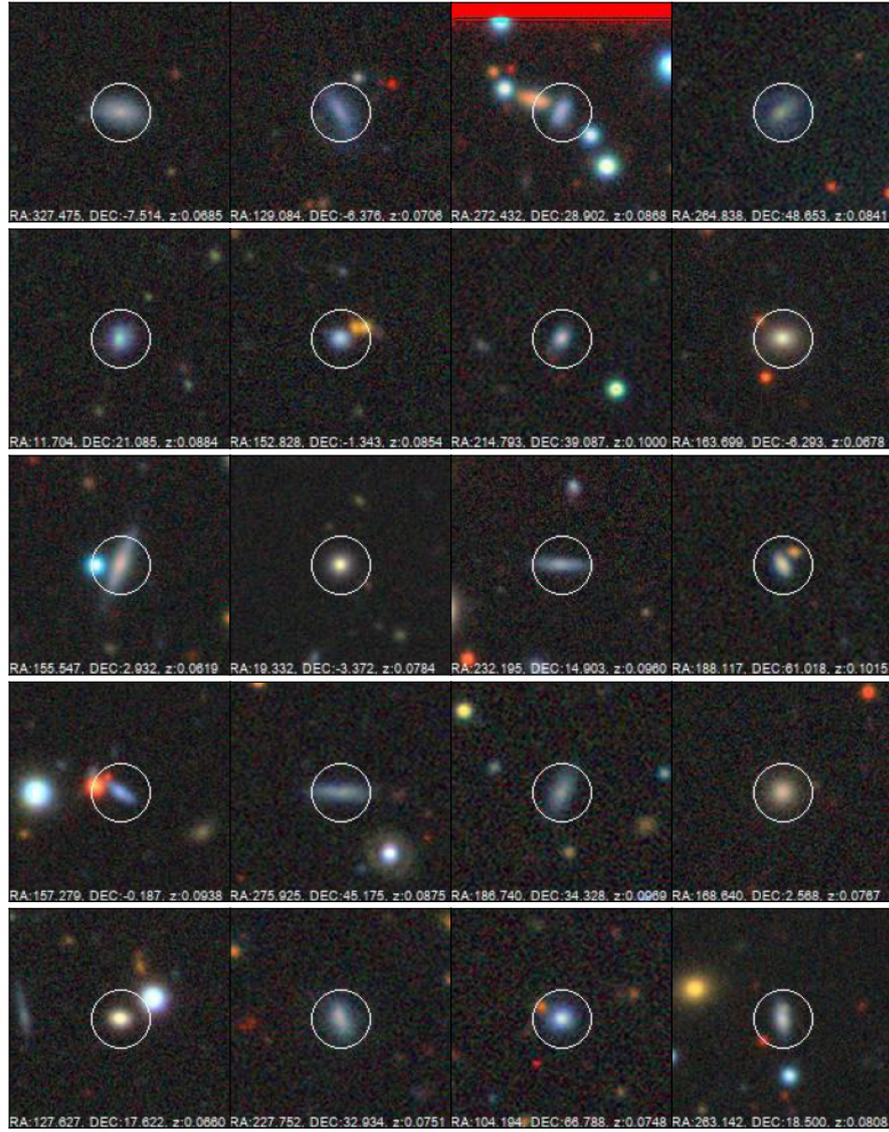Figure D.2: Plot showing a random sample of 20 galaxies in the range $-13 < M_r - 5\log_{10} h < -12.75$.

Figure D.3: Plot showing a random sample of 20 galaxies in the range $-18 < M_r - 5\log_{10} h < -17.75$.

A neural network is a layered model inspired by the brain that makes use of a large number of interconnected nodes (often termed 'neurons') to learn and classify data. These nodes are organised into layers - the input layer, the output layer, and the hidden layer(s). In this case, the input layer consists of the pixels of the image. The output layer is a 1x3 array of probabilities - specifically the probability that a galaxy is 'good', 'possible' or 'problem' (as defined by our visual inspection criteria in Section 3.6). The hidden layers consist of a range of nodes that process the input data, where each node applies a weight to the input and utilises an activation function to generate the output. These weights may be assigned and calibrated on a training set by a number of methods, such as backpropagation. There is no obvious interpretation for each hidden layer, although it is noted that adding more hidden layers is a method for classifying more complex structures. (Aggarwal, 2018)

We test a range of neural networks and select a particular network that performed well on both the training set and the test set (Fig. D.4). Fig. D.5 shows examples of the classification of the neural network on the training set. This shows that the neural network is reasonably good at classifying problem galaxies correctly. However, we observe that the model can be overly conservative - often classifying 'good' galaxies as 'possible' galaxies. As such, we choose to remove galaxies classified as 'bad' but keep galaxies classified as 'possible'. We do not classify every galaxy in DR1 BGS for timing reasons, as it takes a non-negligible amount of time to download the image of each galaxy. However, we do classify a large number of galaxies across all magnitude bins in order to assess the validity of the neural network. In particular, we confirm that the network remains accurate for faint galaxies (and so has not been overtrained on bright galaxies). We find that this neural network yields similar results to Table 3.3, as expected.

We present an r-band LF in the South where each magnitude bin is weighted by the number of good and possible galaxies over the total number of galaxies in that sample, using the samples described in Table 3.3. We find that this correction results in an LF with a reduced upturn at the faint-end. However, there is still a visible upturn in the LF. A similar result holds for the North r-band LF. In addition, removing problem objects with our Neural Network yields a similar result to that

Layer (type)	Output Shape	Param #
flatten (Flatten)	(None, 69312)	0
dense (Dense)	(None, 128)	8,872,064
dense_1 (Dense)	(None, 128)	16,512
dense_2 (Dense)	(None, 128)	16,512
dense_3 (Dense)	(None, 128)	16,512
dense_4 (Dense)	(None, 3)	387

Figure D.4: An example of one of the neural networks that we tested.
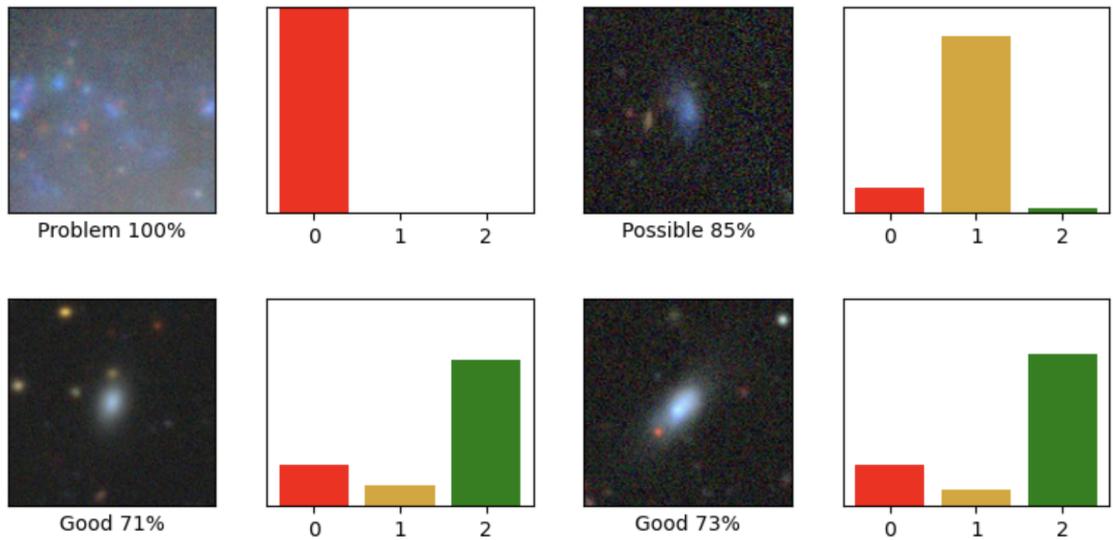

Figure D.5: An example of the neural network classification on the training set for one of our models. An image is shown and the Neural Network assigns a probability that the galaxy is a problem ('0'), possible ('1') or good ('2'). The galaxy is then assigned the class with the highest probability.

of the table-corrected LF, although it is highly variable to the model used and rather variable as we have not classified every galaxy. However, we caution that for the reasons outlined above that we do not classify galaxies across all magnitudes. Moreover, different choices in the neural network (such as the number of hidden layers) can cause significant differences in the corrected LF. As such, although our results seem to show that the upturn is a real feature (albeit not as prominent), we recognise that more work is needed to confirm this.
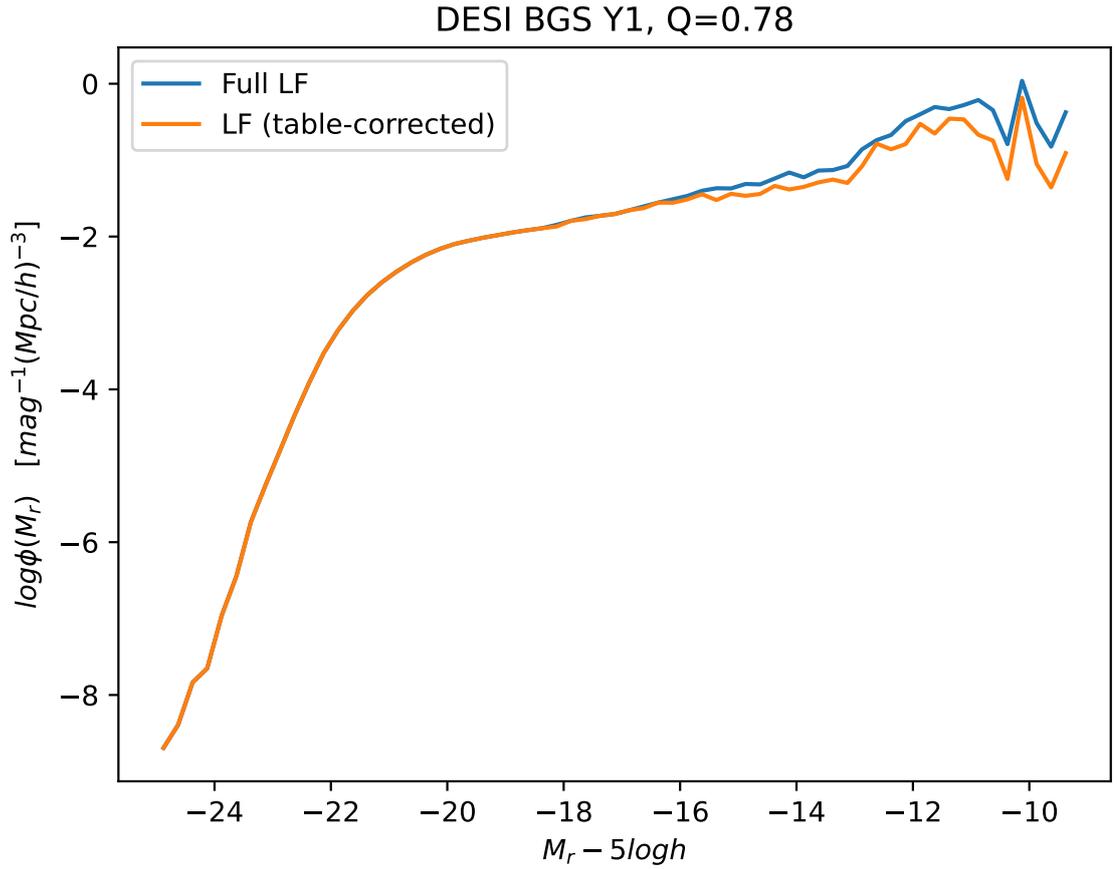
Figure D.6: The South $1/V_{\mathrm{max}}$ r-band LF. Here, we present the full LF with all objects (including potentially spurious objects), identical to that in Fig. 3.6. We then present a table-corrected LF which adjusts each magnitude bin by a factor based on the proportion of problem objects found in our visual inspection for that bin.

Glossary

ACT Atacama Cosmology Telescope. 13

AGN Active Galactic Nuclei. 9, 10, 11, 93, 108

BAO Baryon Acoustic Oscillation. 13

BASS Beijing-Arizona Sky Survey. 27, 29, 33, 63, 66, 106, 123, 124

BGS Bright Galaxy Survey. ii, 13

BH Black Hole. 93

CDM Cold Dark Matter. 94

CMB Cosmic Microwave Background. 2

DDP Density Defining Population. 75, 76

DECaLS Dark Energy Camera Legacy Survey. 27, 33, 63, 66, 106, 123, 124

DM Dark Matter. 5, 8, 9, 93, 105

DOE Department of Energy. 13

EDR Early Data Release. 14, 30

ELG Emission Line Galaxies. 13

EM Expectation-Maximisation. 66

FLRW Friedmann-Lemaître-Robinson-Walker. 1

FSF FastSpecFit. ii, ix, 26, 30, 31, 63, 97, 103

GAMA Galaxy and Mass Assembly. ii, 11, 14, 17, 19, 30, 49, 56, 57, 73, 89, 103, 105

GFA Guiding, Focusing and Alignment. 26

IMF Initial Mass Function. 94

ISM Interstellar Medium. 9, 94

LF Luminosity Function. ii

LRG Luminous Red Galaxies. 13

LSS Large Scale Structure. 6, 37

LSST Large Synoptic Survey Telescope (renamed to the Vera C. Rubin Observatory). 13

MCMC Markov Chain Monte Carlo. 59, 60

MWS Milky Way Survey. 13

MzLS Mayall z-band Legacy Survey. 33, 124

NASA National Aeronautics and Space Administration. 13

NSF National Science Foundation. 13

OLS Ordinary Least Squares. 52

PAC Photometric Objects around Cosmic Webs. 92, 95

PCA Principal Component Analysis. 22

QSO Quasar. 13, 22

RSD Redshift Space Distortion. 13

SAM Semi-Analytic Model. 9, 10

SDSS Sloane Digital Sky Survey. 11, 19, 22, 26, 27, 29, 33, 53

SED Spectral Energy Distribution. 12, 26, 94

SFD Schlegel, Finkbeiner & Davis (1998). 19

SHAM Subhalo Abundance Matching. 99

SKA Square Kilometre Array. 13

SMF Stellar Mass Function. ii, 9, 15, 92, 104

SN Supernova. 10, 11, 93, 108

SPS Stellar Population Synthesis. 94

STY Sandage, Tammann & Yahil (commonly used to name a LF method.). 43

SWML Stepwise Maximum Likelihood. ii, 41, 43, 73, 89

TSNR Target Signal-to-Noise. 37

TSNR2 Target Signal-to-Noise Squared. 38

WDM Warm Dark Matter. 93

WIMPS Weakly Interacting Massive Particles. 5

WISE Wide-Field Infrared Survey Explorer. 22, 53