PROVABGS Probabilistic Stellar Mass Function of the BGS One-Percent Survey

ChangHoon Hahn[1],[*]

[1]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton NJ 08544, USA*

(Dated: `DRAFT -- 6d00025 -- 2021-01-14 -- NOT READY FOR DISTRIBUTION`)

ABSTRACT

Keywords: keyword1 – keyword2 – keyword3

1. INTRODUCTION

2. THE DESI BRIGHT GALAXY SURVEY: EARLY DATA RELEASE

DESI began its five years of operations in May 14, 2021. something about EDR Before its start, TODO DESI conducted the Survey Validation (SV) campagin to verify that the survey will meets its scientific and performance requirements. The SV campaign was divided into two main programs: the first, SV1, characterized the survey's performance for different observing conditions and was used to optimize sample selection. The second, the One-Percent Survey (or SV3), observed a dataset that can be used for representative clustering measurements and deliver a 'truth' sample with high completeness over an area at least 1% of the expected main survey footprint. We refer readers to **?** for details on the DESI SV programs. In this work, we focus on BGS galaxies observed during the One-Percent Survey.

The One-Percent Survey observed on 38 nights from April 2021 to the end of May 2021. During this time DESI observed 288 bright time exposures that cover 214 BGS 'tiles', planned DESI pointings. A set of 11 overlapping tiles so that their centers are arranged around a 0.12 deg circle, forming a 'rosette' completeness pattern. In total, the One-Percent Survey observed 20 rosettes covering 180 deg^2 spanning the northern galactic cap (see Figure 1 in Hahn et al. 2022).

All BGS spectra observed during the One-Percent Survey are reduced using the 'Fuji' version of the DESI spectroscopic data reduction pipeline (**?**). First, spectra are extracted from the spectrograph CCDs using the *Spectro-Perfectionsim* algorithm of **?**. Then, fiber-to-fiber variations are corrected by flat-fielding and a sky model, empirically derived from sky fibers, is subtracted from each spectrum. Afterwards, the fluxes in the spectra are calibrated using stellar model fits to standard stars. The final processed spectra is then derived by co-adding the calibrated spectra across expoures of the same tile. In total, DESI observed spectra of 155,022 BGS Bright and 109,418 BGS Faint targets during the One-Percent Survey.

For each target spectrum, we measure its redshift using Redrock[1], the redshift fitting algorithm for DESI (**?**). Redrock uses χ^2 minimization computed from a linear combination of Principal

---

[*] changhoon.hahn@princeton.edu
[1] https://redrock.readthedocs.io

Component Analysis (PCA) basis spectral templates in three template classes ("stellar", "galaxy", and "quasar") to find the best-fit redshift. It also provides a measure of redshift confidence, $\Delta\chi^2$, which corresponds to the difference between the χ^2 values of the best-fit REDROCK model and the next best-fit model.

Among the One-Percent Survey BGS targets, we restrict our samples to targets with reliable redshift measurements. We, therefore, exclude targets observed using malfunctioning fiber positioners. We also restrict our sample to targets with: spectra classified as galaxy spectra by REDROCK, no REDROCK warning flags, $\Delta\chi^2 > 40$, and REDROCK redshift uncertainty ZERR $< 0.0005(1+z)$. Lastly, we only include galaxies within $0 < z < 0.6$. After these cuts, our One-Percent Survey BGS sample includes 143,074 BGS Bright galaxies and 96,771 BGS Faint galaxies.

3. PROVABGS SED MODELING

For each BGS galaxy, we derive its M_* and other properties, $\overline{\mathrm{SFR}}$, Z_{MW}, and $t_{\mathrm{age,MW}}$ from DESI photometry and spectroscopy using the PROVABGS SED modeling framework (Hahn et al. 2022). PROVABGS models galaxy SEDs using stellar population synthesis with non-parametric star-formation history (SFH) with a starburst, a non-parametric metallicity history (ZH) that varies with time, and a flexible dust attenuation prescription. The non-parameteric SFH and ZH prescriptions are derived from SFHs and ZHs of simulated galaxies in the Illustris hydrodynamic simulation (Vo-gelsberger et al. 2014; Genel et al. 2014; Nelson et al. 2015) and provide compact and flexibly representations of SFHs and ZHs. For the stellar population synthesis, PROVABGS uses the Flexible Stellar Population Synthesis (FSPS; Conroy et al. 2009, 2010) model with MIST isochrones (Paxton et al. 2011, 2013, 2015; Choi et al. 2016; Dotter 2016), Chabrier (2003) initial mass function (IMF), and a combination of MILES (Sánchez-Blázquez et al. 2006) and BaSeL (Lejeune et al. 1997, 1998; Westera et al. 2002) spectral libraries.

Furthermore, PROVABGS provides a Bayesian inference framework that infers full posterior probability distributions of the SED model parameter: $p(\theta\,|\,\mathbf{X}^{\mathrm{photo}},\mathbf{X}^{\mathrm{spec}})$, where $\mathbf{X}^{\mathrm{photo}}$ represents the photometry and $\mathbf{X}^{\mathrm{spec}}$ represents the spectroscopy. In total, θ has 13 parameters: M_*, 6 parameters specifying the SFH ($\beta_1, \beta_2, \beta_3, \beta_4, f_{\mathrm{burst}}, t_{\mathrm{burst}}$), 2 parameters specifying ZH (γ_1, γ_2), 3 parameters specifying dust attenuation ($\tau_{\mathrm{BC}}, \tau_{\mathrm{ISM}}, n_{\mathrm{dust}}$), and a nuisance parameter for the fiber aperture effect. Posteriors have distinct advantages over point estimates because they accurately estimate uncertainties and degeneracies among galaxy properties. Furthermore, as we later demonstrate, they are essential for principled population inference: *e.g.* SMF.

In practice, accurately estimating a 13 dimensional posterior requires a large number ($\gtrsim$100,000) SED model evaluations, which requires prohibitive computational resources — ~ 10 CPU hours per galaxy. To address this challenge, PROVABGS samples the posterior using the Karamanis & Beutler (2020) ensemble slice Markov Chain Monte Carlo (MCMC) sampling with the ZEUS Python package[2]. PROVABGS further accelerates the inference by using neural emulators for the SED models. The emulators are accurate to subpercent level and $> 100\times$ faster than the original SED model based on FSPS (Kwon et al. 2022). With ZEUS and neural emulation, deriving a posterior takes $\sim$5 min
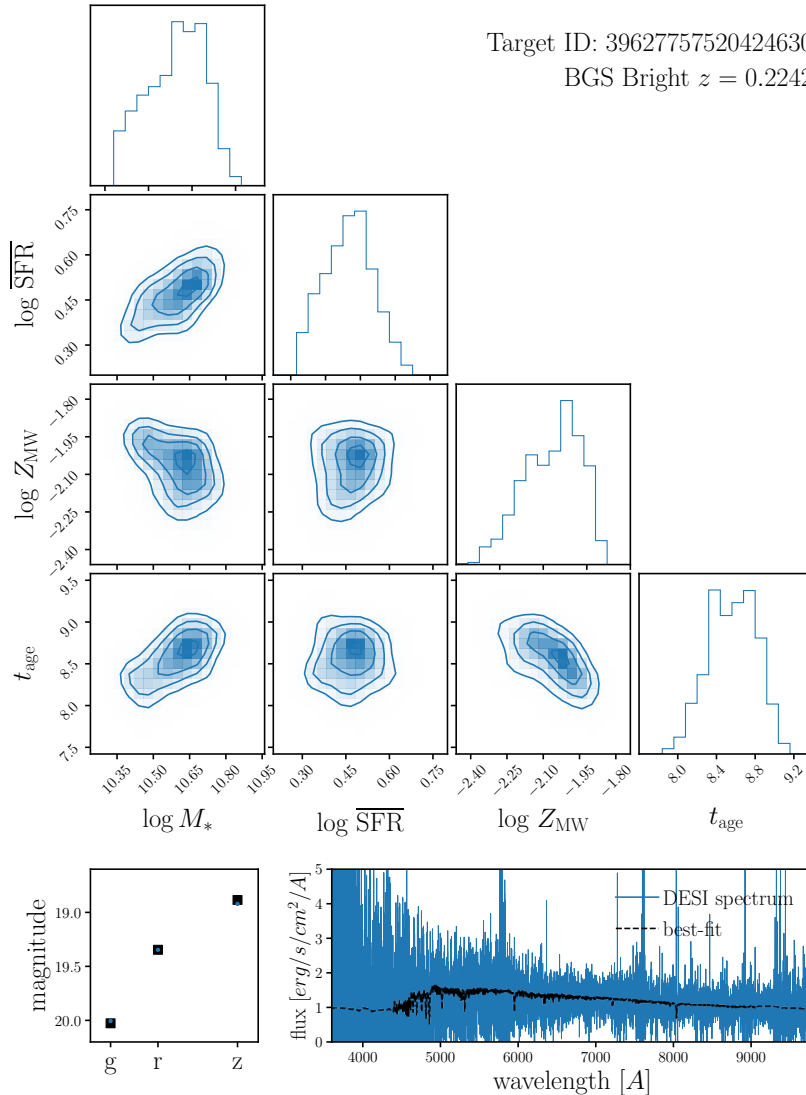

Figure 1. *Top panels*: Posteriors of galaxy properties, M_*, $\overline{\mathrm{SFR}}$, Z_{MW}, and $t_{\mathrm{age,MW}}$, for a randomly selected BGS Bright galaxy with $z = 0.2242$ (target ID: 39627757520424630) inferred using the PROVABGS SED modeling framework from DESI photometry and spectroscopy. The contours mark the X, X, and X percentiles of posterior. With the PROVABGS posteriors, we accurately estimate the galaxy properties, their uncertainties, and any degeneracies among them. *Bottom panels*: Comparison of the best-fit PROVABGS SED model prediction (black) to observations (blue). We compare the g, r, and z band photometry in the left panel and spectra in the right panel. We infer the posterior of galaxy properties for every BGS galaxies in the DESI One-Percent Survey.

per galaxy with PROVABGS. Moreover, Hahn et al. (2022) demonstrated PROVABGS can accurately infer M_* overall the full expected M_* range of BGS, using forward modeled synthetic DESI observations.

In Figure 1, we demonstrate the PROVABGS SED modeling framework for a randomly selected BGS Bright galaxy with $z = 0.2242$ (target ID: 39627757520424630). In the top panels, we present
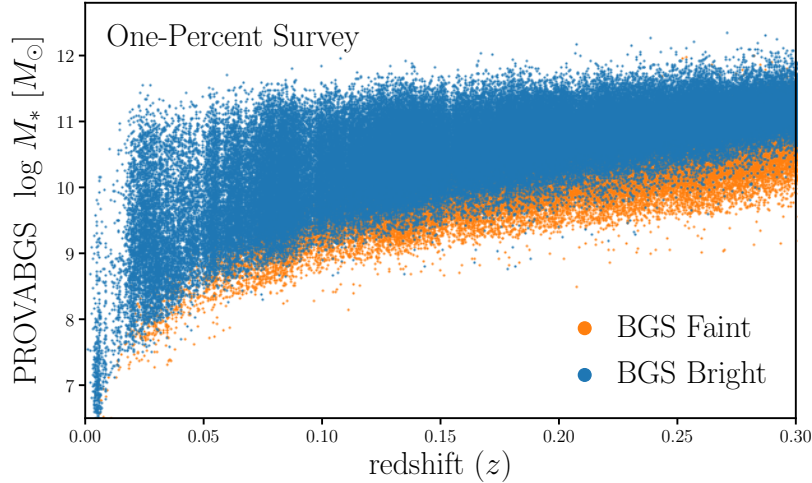
Figure 2. M_* as a function of z of BGS Bright (blue) and Faint (orange) galaxies in the DESI One-Percent Survey. For M_*, we use the best-fit values derived using PROVABGS. BGS Bright is a magnitude-limited sample to $r < 19.5$ while BGS Faint includes fainter galaxies $19.5 < r < 20.175$ selected using $r_{\rm fib}$ and color (**?**). In total, we infer the posteriors of 143,017 BGS Bright and 95,499 BGS Faint galaxies in the DESI One-Percent Survey spanning $0 < z < 0.6$.

the posteriors of galaxy properties, M_*, $\overline{\rm SFR}$, $Z_{\rm MW}$, and $t_{\rm age,MW}$, inferred from DESI photometry and spectroscopy. We mark the X, X, and X percentiles of posterior with the contours. The posteriors illustrate that we can precisely measure the properties of BGS galaxies from DESI photometry and spectroscopy. Furthermore, with the full posterior, we accurately estimate the uncertainties on the galaxy properties and the degeneracies among them (*e.g.* M_* and $\overline{\rm SFR}$). In the bottom panels, we compare the PROVABGS SED model prediction using the best-fit parameter values (black) to DESI observations (blue). The left panel compares the optical g, r, and z band photometry while the right panel compares the spectra. The comparison shows good agreement between the best-fit model and the observations.

We derive a PROVABGS posterior (*e.g.* Figure 1) for every galaxy in the DESI One-Percent Survey. In Figure 2, we present the best-fit M_* measurements as a function of z for the BGS galaxies in DESI One-Percent Survey. We mark the galaxies in the BGS Bright sample in blue and the ones in the BGS Faint sample in orange. We infer the posteriors of 143,017 BGS Bright and 95,499 BGS Faint galaxies.

4. RESULTS

From the posteriors of galaxy properties inferred using PROVABGS (Section 3), we derive the marginalized posteriors: $p(M_* \,|\, \boldsymbol{X}_i)$, the marginalized 1D posterior of M_* for galaxy i. Using these posteriors, we can estimate the SMF of BGS galaxies using population inference in a hierarchical Bayesian framework (*e.g.* Hogg et al. 2010; Foreman-Mackey et al. 2014; Baronchelli et al. 2020). In other words, we infer $p(\phi \,|\, \{\boldsymbol{X}_i\})$, the probability distribution of population hyperparameters ϕ that describe the SMF, $\Phi(M_*; \phi)$, given the DESI observations, $\{\boldsymbol{X}_i\}$. For the SMF, we use a Gaussian Mixture Model (GMM; Press et al. 1992; McLachlan & Peel 2000), which provides a highly flexible

parameterization for describing population distributions:

$$\Phi(M_*; \phi) = \sum_{j=1}^{k} \mathcal{N}(M_*; \phi_j).$$ (1)

k represents the number of Gaussian components. ϕ_j represent the mean and standard deviation of the jth Gaussian component of the GMM.

We take this approach over the standard approach that use point estimates of M_* because we can correctly propagate the uncertainties in our M_* measurements and more robustly estimate the M_* distribution — *ie.* SMF. Malz & Hogg (2020) demonstrated in the context of inferring redshift distributions from individual photometric redshift measurements that using point estimates is statistically incorrect and can lead to biased redshift distributions. We emphasize that Malz & Hogg (2020) is an analogous analysis with a similar goal of measuring a 1D galaxy property distribution.

To infer $p(\phi \,|\, \{\boldsymbol{X}_i\})$, we follow the same approach described in Hahn et al. (2022):

$$p(\phi \,|\, \{\boldsymbol{X}_i\}) = \frac{p(\phi) \, p(\{\boldsymbol{X}_i\} \,|\, \phi)}{p(\{\boldsymbol{X}_i\})}$$ (2)

$$= \frac{p(\phi)}{p(\{\boldsymbol{X}_i\})} \int p(\{\boldsymbol{X}_i\} \,|\, \{\theta_i\}) \, p(\{\theta_i\} \,|\, \phi) \, \mathrm{d}\{\theta_i\}.$$ (3)

$$= \frac{p(\phi)}{p(\{\boldsymbol{X}_i\})} \prod_{i=1}^{N} \int p(\boldsymbol{X}_i \,|\, \theta_i) \, p(\theta_i \,|\, \phi) \, \mathrm{d}\theta_i$$ (4)

$$= \frac{p(\phi)}{p(\{\boldsymbol{X}_i\})} \prod_{i=1}^{N} \int \frac{p(\theta_i \,|\, \boldsymbol{X}_i) \, p(\boldsymbol{X}_i)}{p(\theta_i)} \, p(\theta_i \,|\, \phi) \, \mathrm{d}\theta_i$$ (5)

$$= p(\phi) \prod_{i=1}^{N} \int \frac{p(\theta_i \,|\, \boldsymbol{X}_i) \, p(\theta_i \,|\, \phi)}{p(\theta_i)} \, \mathrm{d}\theta_i.$$ (6)

We estimate the integral using S_i Monte Carlo samples from the individual posteriors $p(\theta_i \,|\, \boldsymbol{X}_i)$:

$$\approx p(\phi) \prod_{i=1}^{N} \frac{1}{S_i} \sum_{j=1}^{S_i} \frac{p(\theta_{i,j} \,|\, \phi)}{p(\theta_{i,j})}.$$ (7)

Since the sample of BGS galaxies is not volume-limited and complete as a function of M_*, we must account for the selection effect and incompleteness when estimating the SMF. To account for the selection effects of the BGS samples, we include weights derived from z^{max}, the maximum redshift that galaxy i could be placed and still be included in the BGS samples. We derive z_i^{max} for every galaxy using by redshifting the SED predicted by the best-fit parameters. We then derive V_i^{max}, the comoving volume out to z_i^{max}, and include a factor of $1/V_i^{\text{max}}$ in the galaxy weight w_i.

Next, we include correction weights for spectroscopic incompleteness driven by fiber assignment and redshift failures. Incompletenss from fiber assignment is due to the fact that DESI is not able to assign fibers to all galaxies included in the BGS target selection. Furthermore, due to galaxy clustering there is significant variation in the assignment probability. Meanwhile, incompleteness from redshift

failure is caused by the fact that we do not successfully measure the redshift for every spectra and the redshift failure rate depends on the surface brightnesses of the galaxies and the signal-to-noise ratio of the spectra. We describe how we derive $w_{i,\text{FA}}$ and $w_{i,\text{ZF}}$, the incompleteness correction weights for fiber assignment and redshift failures in Appendix A. Each BGS galaxy is assigned a weight of $w_i = (w_{i,\text{FA}} \times w_{i,\text{ZF}})/V_i^{\max}$.

We modify Eq. 2 to include w_i:

$$p(\phi \,|\, \{\boldsymbol{X}_i\}) \approx \frac{p(\phi)}{\prod\limits_{i=1}^{N} p(\boldsymbol{X}_i)^{w_i}} \prod_{i=1}^{N} \left(\int p(\boldsymbol{X}_i \,|\, \theta_i) \, p(\theta_i \,|\, \phi) \, \mathrm{d}\theta_i \right)^{w_i} \tag{8}$$

$$\approx \frac{p(\phi)}{\prod\limits_{i=1}^{N} p(\boldsymbol{X}_i)^{w_i}} \prod_{i=1}^{N} \left(\sum_{j=1}^{S_i} \frac{p(\theta_{i,j} \,|\, \phi)}{p(\theta_{i,j})} \right)^{w_i} \tag{9}$$

$$\approx \frac{p(\phi)}{\prod\limits_{i=1}^{N} p(\boldsymbol{X}_i)^{w_i}} \prod_{i=1}^{N} \left(\sum_{j=1}^{S_i} \frac{q_\phi(\theta_{i,j})}{p(\theta_{i,j})} \right)^{w_i}. \tag{10}$$

In practice, we do not derive the full posterior $p(\phi \,|\, \{\boldsymbol{X}_i\})$. Instead we derive the maximum a posteriori (MAP) hyperparameter ϕ_{MAP} that maximizes $p(\phi \,|\, \{\boldsymbol{X}_i\})$ or $\log p(\phi \,|\, \{\boldsymbol{X}_i\})$. We expand,

$$\log p(\phi \,|\, \{\boldsymbol{X}_i\}) \approx \log p(\phi) + \sum_{i=1}^{N} w_i \log \left(\sum_{j=1}^{S_i} \frac{q_\phi(\theta_{i,j})}{p(\theta_{i,j})} \right). \tag{11}$$

Since the first two terms are constant, we derive ϕ_{MAP} by maximizing

$$\max_{\phi} \quad \sum_{i=1}^{N} w_i \log \left(\sum_{j=1}^{S_i} \frac{q_\phi(\theta_{i,j})}{p(\theta_{i,j})} \right). \tag{12}$$

using the ADAM optimizer (Kingma & Ba 2017). We derive ϕ_{MAP} for BGS galaxies in redshift bins of width $\Delta z = 0.04$ starting from $z = 0.01$. This enables us to examine the redshift evolution of the SMF within BGS.

4.1. *The Probabilistic Stellar Mass Function*

We present the probabilistic SMF of $0.01 < z < 0.05$ BGS galaxies in the One-Percent Survey in Figure 3 (blue). We represent the uncertainties on the SMF from sample variance in the shaded region. The uncertainties are estimated using a standard jackknife technique, which we describe in Appendix B (Eq. B4) and provide conservative uncertainty estimates (Norberg et al. 2009). We also include SMF measurements from previous spectroscopic surveys PRIMUS (?) (black circle) and GAMA (Driver et al. 2022) (black triangle). Since ? and Driver et al. (2022) derive M_* using different SED models we include a M_* offset based on overlapping galaxies.
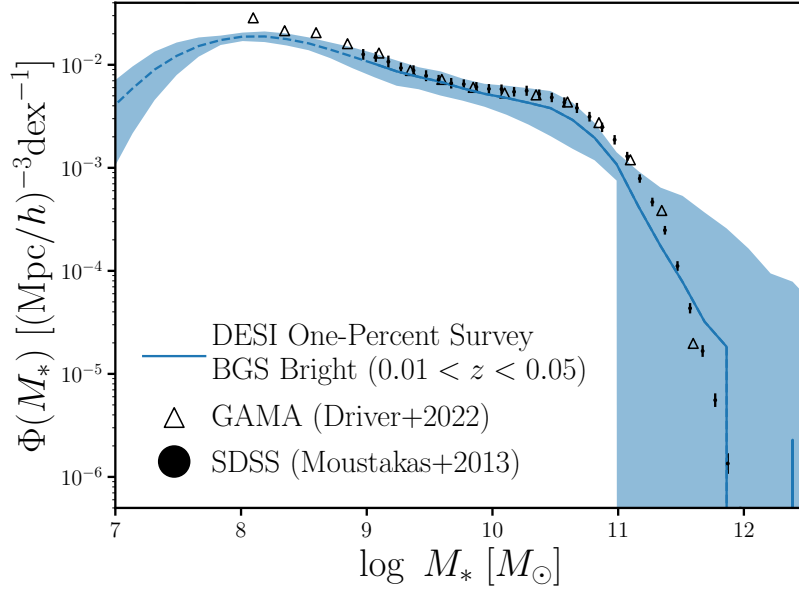
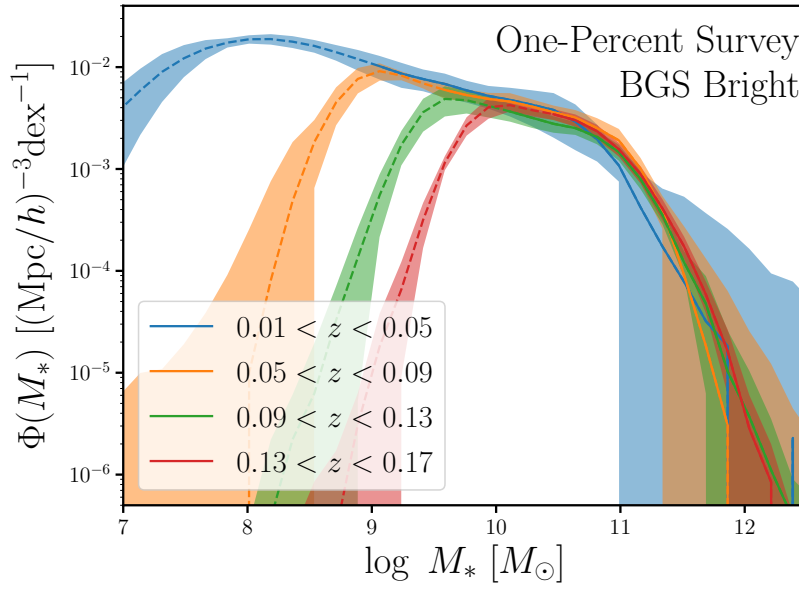**Figure 3.** The probabilistic SMF of $0.01 < z < 0.05$ BGS galaxies in the One-Percent Survey.

Figure 4. The probabilistic SMF of $0.01 < z < 0.05$ BGS galaxies in the One-Percent Survey.

ACKNOWLEDGEMENTS

It's a pleasure to thank

APPENDIX

A. SPECTROSCOPIC COMPLETENESS

Spectroscopic galaxy surveys, such as BGS, do not successfully measure the redshift for all of the galaxies they target. As a result, this spectroscopic incompleteness must be accounted for when measuring galaxy population statistics such as the SMF. In this appendix, we present how we estimate the spectroscopic incompleteness for BGS and derive the weights we use to correct for its impact on the SMF.

For BGS, spectroscopic incompleteness is primarily driven by fiber assignment and redshift failures. DESI uses 10 fiber-fed spectrographs with 5000 fibers but targets more galaxies than available fibers. For instance, the BGS Bright and Faint samples have ~ 860 and $530\,\mathrm{targets/deg}^2$, respectively. For the 8 deg^2 field-of-view of DESI, this roughly correspond to 11,000 targets, significantly more than the 5000 available fibers. DESI only measures the spectra of targets that are assigned fibers. In fact, of the 5000, a minimum of 400 'sky' fibers are dedicated to measuring the sky background for accurate sky subtraction and an additional 100 fibers are assigned to standard stars for flux calibration **?**.

Furthermore, each fiber is controlled by a robotic fiber positioner on the focal plane. These positioners can rotate on two arms and be positioned within a circular patrol region of radius 1.48 arcmin (**????**). Although the patrol regions of adjacent positioners slightly overlap, the geometry of the positioners cause higher incompleteness in regions with high target density (**?**). To mitigate the incompleteness from the fiber assignment, BGS will observe its footprint with four passes. With this strategy, BGS achieves $\sim 80\%$ fiber assignment completeness (**?**).

To estimate fiber assignment completeness, we run the fiber assignment algorithm (**?**) on BGS targets 128 separate times. For each BGS galaxy, i, we count the total number of times out of 128 that the galaxy is assigned a fiber: $N_{i,\mathrm{FA}}$. Then to correct for the fiber assignment incompleteness, we assign correction weights

$$w_{i,\mathrm{FA}} = \frac{128}{N_{i,\mathrm{FA}}} \tag{A1}$$

to each BGS galaxy. explain what this means

Although we measure a spectrum for each galaxy assigned a fiber, we do not accurately measure redshifts for every spectra. This redshift measurement failure significantly contributes to spectroscopic incompleteness. For BGS, redshift failure of an observed galaxy spectrum depends mainly on fiber magnitude and a statistic, TSNR2. Fiber magnitude is the predicted flux of the BGS object within a $1.5''$ diameter fiber; we use r-band fiber magnitude, r_{fiber}. TSNR2 roughly corresponds to the signal-to-noise ratio of the spectrum and is the statistic used to calibrate the effective exposure times in DESI observations (CITE).

TODO

In Figure 5, we present the redshift, z, success rate of BGS Bright galaxies as a function of r_{fiber} and TSNR2. In each hexbin, the color map represents the mean z-success rate. We include all hexbins with more than 2 galaxies. Overall, the z-success rate depends significantly on r_{fiber}: galaxies with fainter r_{fiber} have lower z-success rates. However, the r_{fiber} dependence itself varies in bins of TSNR2. We mark the edges of the bins in black dashed: $\log \mathrm{TSNR2} = 2.0, 2.5, 3.0, 3.5, 3.85$. Within each of the TSNR2 bins, the r_{fiber} dependence of the z-success rate does not vary significantly. In Figure 6,
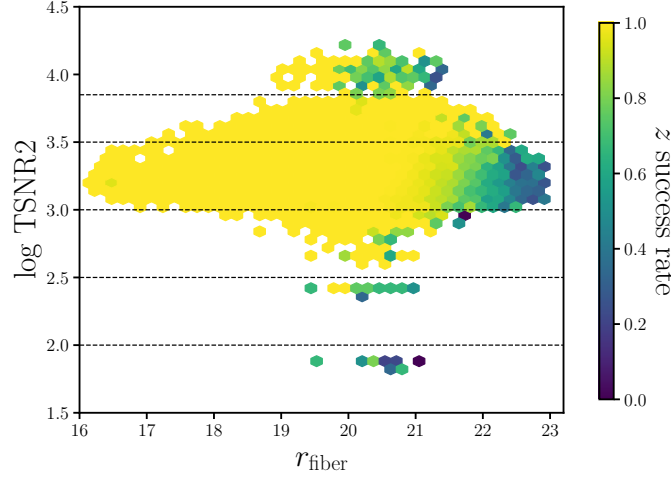
Figure 5. Redsift success rate of BGS Bright galaxies as a function of $r_{\rm fiber}$ and TSNR2. TSNR2 is a statistic that quantifies the signal-to-noise ratio of the observed spectrum. The color map represents the mean redshift success rate in each hexbin. We mark the TSNR2 bins (black dashed) that we use to separately fit the redshift success rate as a function of $r_{\rm fiber}$ using Eq. A3. In each TSNR2 bin, redshift success decreases as $r_{\rm fiber}$ increases.

we present the z-success rate of BGS Bright galaxies as a function of $r_{\rm fiber}$ for each of the 6 TSNR2 bins. We mark the range of TSNR2 in the bottom left of each panel. The errorbars represent the Poisson uncertainties of the z-success rate.

To correct for the effect of redshift failures, we include an additional correction weight for each BGS galaxy:

$$w_{i,{\rm ZF}} = \frac{1}{f_{z-{\rm sucess}}(r_{{\rm fiber},i}, {\rm TSNR2}_i)}. \tag{A2}$$

$f_{z-{\rm sucess}}(r_{{\rm fiber},i}, {\rm TSNR2}_i)$ is the z-success rate as a function of $r_{\rm fiber}$ and TSNR2 of the galaxy. Galaxies with $f_{z-{\rm sucess}} = 1$ (100% z-success) will have $w_{i,{\rm ZF}} = 1.0$ while galaxies with $f_{z-{\rm success}} = 0.1$ (10% z-success) will have $w_{i,{\rm ZF}} = 10$. For $f_{z-{\rm sucess}}(r_{{\rm fiber},i}, {\rm TSNR2}_i)$, we fit the following functional form for each TSNR2 bin:

$$f_{z-{\rm success}}(r_{\rm fiber}) = \frac{1}{2}\left(1 - {\rm erf}(c_0(r_{\rm fiber} - c_1))\right). \tag{A3}$$

In Figure 6, we present the best-fit $f_{z-{\rm sucess}}(r_{\rm fiber})$ for each of the TSNR2 bins in dashed. The best-fit coefficients, c_0, c_1, are derived from χ^2 minimization. We repeat this procedure indepedently for BGS Bright galaxies as well as the BGS Faint galaxies with $(z - W1) - 1.2(g - r) + 1.2 \geq 0$, and BGS Faint galaxies with $(z - W1) - 1.2(g - r) + 1.2 < 0$. We list the best-fit values in bins of TSNR2 for each of the samples in Table 1.

B. UNCERTAINTIES ON THE SMF

We estimate the uncertainties of the SMF from sample variance using the standard jackknife technique. This involves splitting our BGS sample into subsamples and then estimating uncertainties
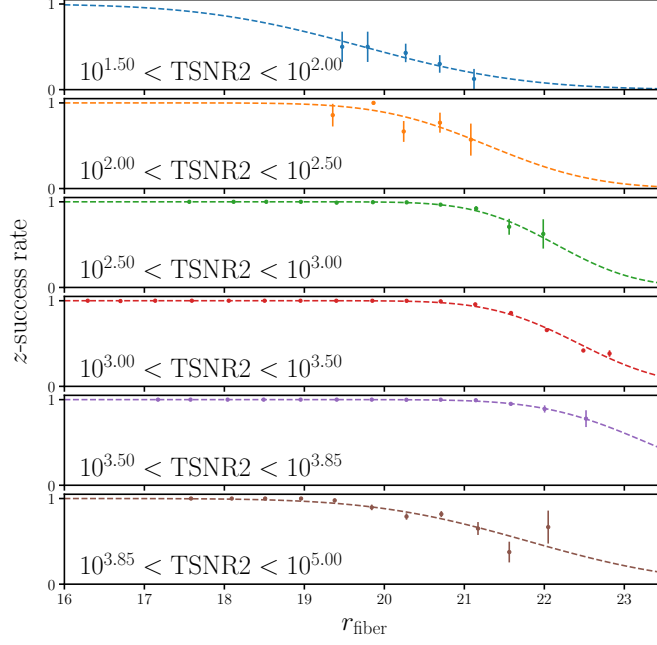
Figure 6. Redshift success rates of BGS Bright galaxies as a function of r_{fiber} in 6 TSNR2 bins. The error bars represent the poisson uncertainties. In each panel, we include the best-fit analytic (Eq. A3) approximation of the redshift success rate (dashed) derived from χ^2 minimization. We use this analytic approximation to calculate the galaxy weights to correct for spectroscopic incompleteness caused by failures to accurately measure redshifts from observed spectra.
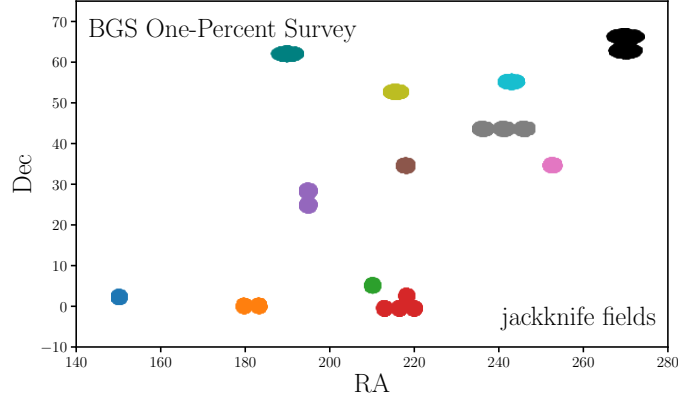

Figure 7. The RA and Dec of the 12 jackknife fields of the BGS One-Percent Survey used to estimate the uncertainties on the SMF from sample variance. We mark each field with a distinct color.

using the subsample-to-subsample variations:

$$\sigma_\Phi = \left(\frac{N_{\text{jack}} - 1}{N_{\text{jack}}} \sum_{k=1}^{N_{\text{jack}}} (\Phi_k - \Phi)^2 \right). \tag{B4}$$

Table 1. Best-fit coefficients of the z-success rate as a function of $r_{\rm fiber}$ for different TSNR2 bins for BGS Bright and Faint samples.

TSNR2 range	c_0	c_1
BGS Bright		
$10^{1.5} - 10^2$	0.443	19.7
$10^2 - 10^{2.5}$	0.668	21.3
$10^{2.5} - 10^3$	0.888	22.1
$10^3 - 10^{3.5}$	0.822	22.4
$10^{3.5} - 10^{3.85}$	0.698	23.3
$10^{3.85} - 10^5$	0.465	21.8
BGS Faint		
$(z - W1) - 1.2(g - r) + 1.2 \geq 0$		
$10^{1.5} - 10^{2.5}$	1.67	21.1
$10^{2.5} - 10^3$	1.65	21.8
$10^3 - 10^{3.1}$	1.49	22.1
$10^{3.1} - 10^{3.2}$	1.32	22.3
$10^{3.2} - 10^{3.3}$	1.33	22.4
$10^{3.3} - 10^{3.5}$	0.907	23.1
$10^{3.5} - 10^{3.85}$	1.03	23.0
$10^{3.85} - 10^5$	0.924	21.6
BGS Faint		
$(z - W1) - 1.2(g - r) + 1.2 < 0$		
$10^{2.5} - 10^3$	1.48	20.9
$10^3 - 10^{3.1}$	2.40	21.2
$10^{3.1} - 10^{3.2}$	1.30	21.8
$10^{3.2} - 10^{3.3}$	1.27	22.0
$10^{3.3} - 10^{3.5}$	1.83	21.6
$10^{3.5} - 10^{3.85}$	0.798	22.9
$10^{3.85} - 10^5$	1.29	20.6

$N_{\rm jack}$ is the number of jackknife subsamples and Φ_k represents the SMF estimated from the BGS galaxies excluding the jackknife subample k. In this work, we split the BGS sample into 12 jackknife fields based on the angular positions of galaxies. We present the jackknife fields in Figure 7 with distinct colors.

C. STELLAR MASS COMPLETENESS

In this appendix, we describe how we derive $M_{\rm lim}$, the M_* limit above which our BGS Bright sample is complete. Although there are various methods for estimating $M_{\rm lim}$ in the literature, *e.g.*
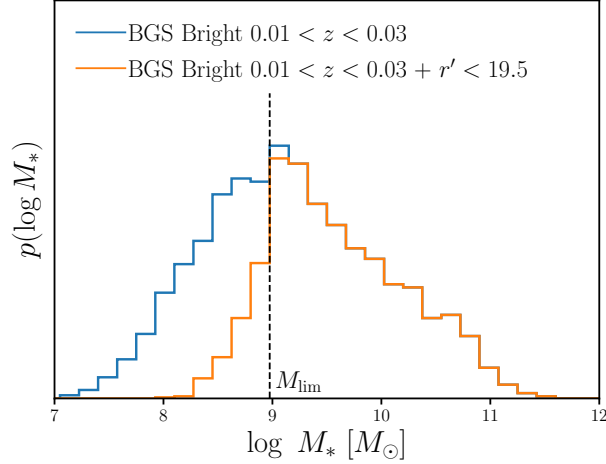
Figure 8. The M_* distribution of BGS Bright galaxies with $0.01 < z < 0.03$ (blue) and the M_* distribution of same set of galaxies that would remain in the BGS Bright magnitude limit if they were redshifted to $z' = z + 0.02$: $r' < 19.5$. We set the stellar mass completeness limit, $M_{\rm lim}$, for $0.01 < z < 0.05$ to the M_* where more than 10% of galaxies are excluded in the latter distribution.

Table 2. Stellar mass completeness limit, $M_{\rm lim}$ for redshift bins of width $\Delta z = 0.04$.

z range	$\log_{10} M_{\rm lim}$
$0.01 - 0.05$	8.975
$0.05 - 0.09$	9.500
$0.09 - 0.13$	10.20
$0.13 - 0.17$	10.38
$0.17 - 0.21$	10.72

based on estimating the mass-to-light ratio (**??**), we adopt a simple approach that takes advantage of the fact that BGS Bright is a magnitude-limited sample.

To derive $M_{\rm lim}$ in redshift bins of width $\Delta z = 0.04$, we first split the galaxy sample into narrower bins of $\Delta z/2$. For each narrower redshift bin, $i\Delta z/2 < z < (i+1)\Delta z/2$, we take all the best-fit PROVABGS SEDs from all galaxies in the bin and artificially redshift it to $z' = z + \Delta z/2$:

$$f'_\lambda = f_\lambda \frac{d_L(z)^2}{d_L(z')^2}. \tag{C5}$$

$d_L(z)$ reprents the luminosity distance at redshift z. Afterward, we calculate the r-band magnitudes, r', for f'_λ and impose the $r' < 19.5$ magnitude limit of the BGS Bright. We then compare the M_* distribution of all the galaxies in $i\Delta z/2 < z < (i+1)\Delta z/2$ to the galaxies in $i\Delta z/2 < z < (i+1)\Delta z/2$ with $r' < 19.5$. For instance, we present the M_* distributions of all BGS Bright galaxies in $0.01 < z < 0.03$ (blue) and the BGS Bright galaxies in $0.01 < z < 0.03$ with $r' < 19.5$ (orange) in Figure 8.
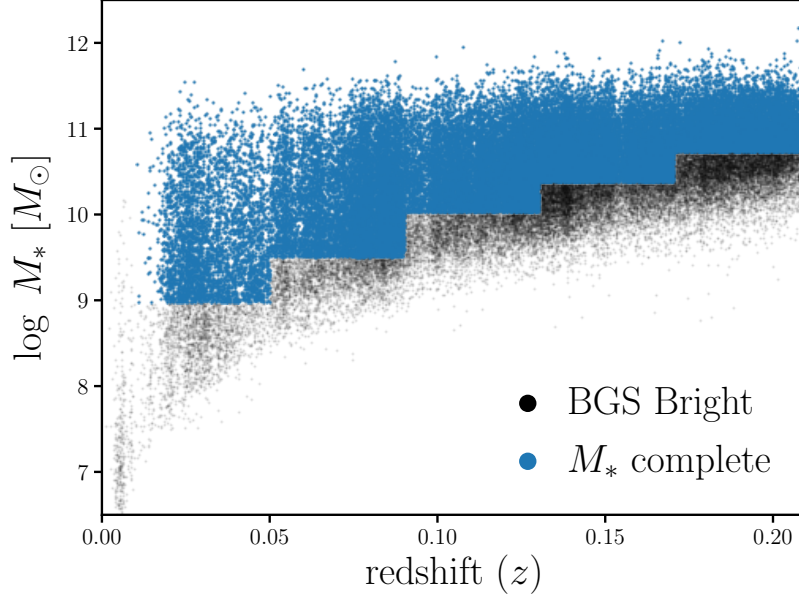
Figure 9. M_* and redshift relation of BGS Bright galaxies in the EDR (black) and the galaxies within the stellar mass completeness limit ($M_* < M_{\rm lim}$; blue). $M_{\rm lim}$ is derived in redshift bins of width $\Delta z = 0.04$. The lowest redshift bin ($0.01 < z < 0.05$) is complete down to $M_* < 10^9 M_\odot$.

Since galaxies become fainter when they are placed at higher redshifts, *i.e.* $r' > r$, the $r' < 19.5$ sample has fewer low M_* galaxies. We determine the M_* at which, more than 10% of galaxies are excluded in the $r' < 19.5$ sample (black dashed) and set this limit as $M_{\rm lim}$ for the redshift bins: $0.01 < z < 0.05$. Our procedure for deriving $M_{\rm lim}$ takes advantage of the fact that galaxy samples at lower redshifts are complete down to lower M_* than at higher redshifts. We repeat this procedure for all the $\Delta z = 0.04$ redshift bins that we use to measure the SMF. In Table 2, we list $M_{\rm lim}$ values for each of the redshift bins. Furthermore, we present the M_* and redshift relation of BGS Bright galaxies (black) and the stellar mass complete sample (blue) in Figure 9.

REFERENCES

Baronchelli L., Nandra K., Buchner J., 2020, Monthly Notices of the Royal Astronomical Society, 498, 5284

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

Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, The Astrophysical Journal, 823, 102

Conroy C., Gunn J. E., White M., 2009, The Astrophysical Journal, 699, 486

Conroy C., White M., Gunn J. E., 2010, The Astrophysical Journal, 708, 58

Dotter A., 2016, The Astrophysical Journal Supplement Series, 222, 8

Driver S. P., et al., 2022, Monthly Notices of the Royal Astronomical Society, p. stac472

Foreman-Mackey D., Hogg D. W., Morton T. D., 2014, The Astrophysical Journal, 795, 64

Genel S., et al., 2014, Monthly Notices of the Royal Astronomical Society, 445, 175

Hahn C., et al., 2022, The DESI PRObabilistic Value-Added Bright Galaxy Survey (PROVABGS) Mock Challenge

Hogg D. W., Myers A. D., Bovy J., 2010, The Astrophysical Journal, 725, 2166

14

Karamanis M., Beutler F., 2020, arXiv e-prints, p. arXiv:2002.06212

Kingma D. P., Ba J., 2017, arXiv:1412.6980 [cs]

Kwon K. J., Hahn C., Alsing J., 2022, Neural Stellar Population Synthesis Emulator for the DESI PROVABGS

Lejeune T., Cuisinier F., Buser R., 1997, A & A Supplement series, Vol. 125, October II 1997, p.229-246., 125, 229

Lejeune T., Cuisinier F., Buser R., 1998, Astronomy and Astrophysics Supplement, v.130, p.65-75, 130, 65

Malz A. I., Hogg D. W., 2020, How to Obtain the Redshift Distribution from Probabilistic Redshift Estimates

McLachlan G., Peel D., 2000, Finite Mixture Models. Wiley-Interscience

Nelson D., et al., 2015, Astronomy and Computing, 13, 12

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

Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, The Astrophysical Journal Supplement Series, 192, 3

Paxton B., et al., 2013, The Astrophysical Journal Supplement Series, 208, 4

Paxton B., et al., 2015, The Astrophysical Journal Supplement Series, 220, 15

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in C (2Nd Ed.): The Art of Scientific Computing. Cambridge University Press, New York, NY, USA

Sánchez-Blázquez P., et al., 2006, Monthly Notices of the Royal Astronomical Society, 371, 703

Vogelsberger M., et al., 2014, Monthly Notices of the Royal Astronomical Society, 444, 1518

Westera P., Lejeune T., Buser R., Cuisinier F., Bruzual G., 2002, Astronomy and Astrophysics, 381, 524