Draft version April 21, 2020
Typeset using LaTeX **preprint** style in AASTeX62

DESI: The PRObabilistic Value-Added Bright Galaxy Survey (PROVA-BGS) Mock Challenge

ChangHoon Hahn,[1,2,*] Gyubin Kwon, Malgorzata Siudek, and GQP WG

[1] *Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley CA 94720, USA*
[2] *Berkeley Center for Cosmological Physics, University of California, Berkeley CA 94720, USA*

(Dated: `DRAFT --- 4fe47c0 --- 2020-04-20 --- NOT READY FOR DISTRIBUTION`)

ABSTRACT

In this project, we present the methodology for inferring physical properties of galaxies from joint SED fitting of the DESI optical photometry and spectroscopy. We construct realistic forward modeled DESI data from star formation and chemical enrichment histories of galaxies in the L-Galaxies semi-analytic model and the Illustris TNG hydrodynamic simulations. Then, using these mock observations, we demonstrate that the stellar mass and SFR posteriors from our SED fitting are consistent with constraints from other SED fitting methods in the literature and, more importantly, the input stellar masses and SFRs from the simulations. The SED fitting we present and validate in this project will be used to construct the PRObabilistic Value-Added Bright Galaxy Survey (PROVABGS) from DESI observations.

Keywords: keyword1 – keyword2 – keyword3

1. INTRODUCTION

What is DESI? Provide an overview of DESI specifics, numbers, and science goals, which will mostly be cosmology (BAO, RSD, etc). DESI will be great TODO

What is DESI for galaxy science? On top of all that DESI goodness, DESI will be great for galaxy science! Brief list of some exciting galaxy and quasar physics you can do with the DEIS BGS, QSOs, etc samples. Current status of DESI Comissioning about to start and SV data coming in a year. This is why a mock challenge timely. TODO TODO

Why do we need a mock challenge? We want to test and cement our methodology specifically for our GQP analysis before SV data comes out. As part of the survey preparation, we have all the tools to accurately forward model observations. details on some of the specific tools and what we're able to simulate: realistic spectroscopy. realistic photometry. realistic spectro-photometry All of this gives us a rare opportunity to test our methodology on bespoke simulations. TODO TODO

A mock challenge is also great for testing new methodology. BGS is a bright time survey and will push the boundaries of low SNR spectra. But if we can find a way to infer robust galaxy properties the

* hahn.changhoon@gmail.com

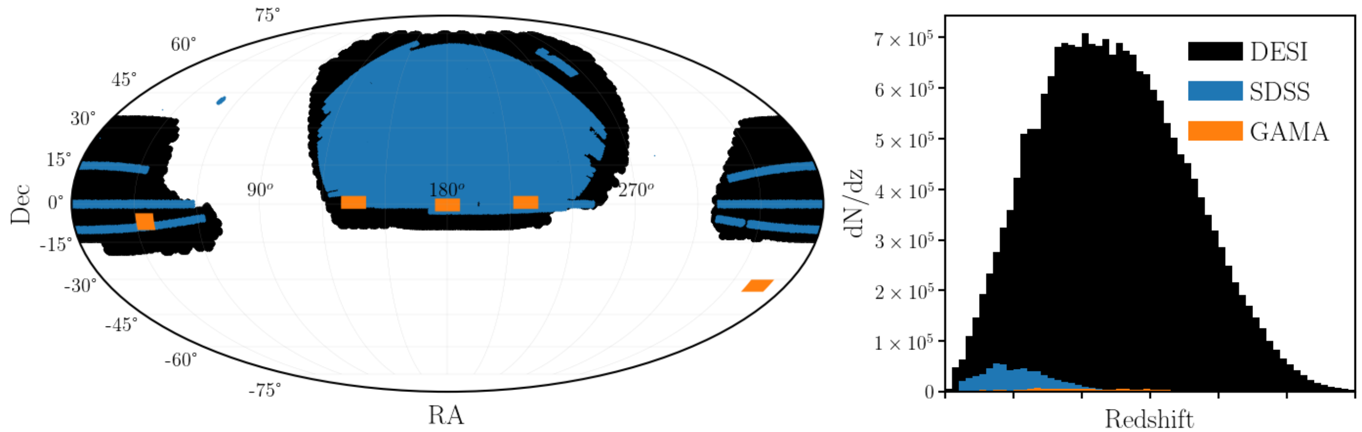

Figure 1. DESI will conduct the largest spectroscopic survey to date covering ∼14, 000 deg^2. During dark time, DESI will measure >20 million spectra of luminous red galaxies, emission line galaxies, and quasars out to $z > 3$. During bright time, DESI will measure the spectra of ∼10 million galaxies out to z∼0.4 with the Bright Galaxy Survey (BGS). *Left*: With its ∼14, 000 deg^2 footprint (black), DESI will cover ∼2× the SDSS footprint (blue; ∼7000 deg^2) and ∼45× the GAMA footprint (orange; ∼300 deg^2). *Right*: Over this footprint, the BGS will provide spectra for a magnitude limited sample of ∼10 million galaxies down to $r < 20$, two orders of magnitude deeper than the SDSS main galaxy sample and 0.2 mag deeper than GAMA.

statistical payout is awesome. Something also about LRGs We're also trying to robustly fit spectra and photometry simultaneously. This has been done before (citations) but not extensively tested on simulations. TODO TODO

broader impact of mock challenge highlight some gqp science cases with a focus on mock challenge papers TODO

2. SIMULATIONS

paragraph describing the DESI data before we talk about how we simulate it. DESI will collect spectroscopy description: wavelength range, spectral resolution, TODO TODO

In addition to galaxy spectra, DESI will also have optical and infrared imaging data from the DESI Legacy Imaging Surveys (hereafter Legacy Surveys; Dey et al. (2019)). The Legacy Surveys are a combination of three public projects (Dark Energy Camera Legacy Survey, Beijing-Arizona Sky Survey, and Mayall z-band Legacy Survey) that jointly imaged the ∼14, 000 deg^2 DESI footprint in three optical bands (g, r, and z). Furthermore, DR8 of the Legacy Survey also includes photometry in the WISE $W1$, $W2$, $W3$, and $W4$ infrared bands. The infrared photometry is from all imaging through year 4 of NEOWISE-Reactivation force-photometered in the unWISE maps at the locations of Legacy Surveys optical sources (cite). TODO

Below we describe how we simulate realistic spectroscopy and photometry for galaxies from state-of-the-art simulations.

2.1. *LGal*

@rita

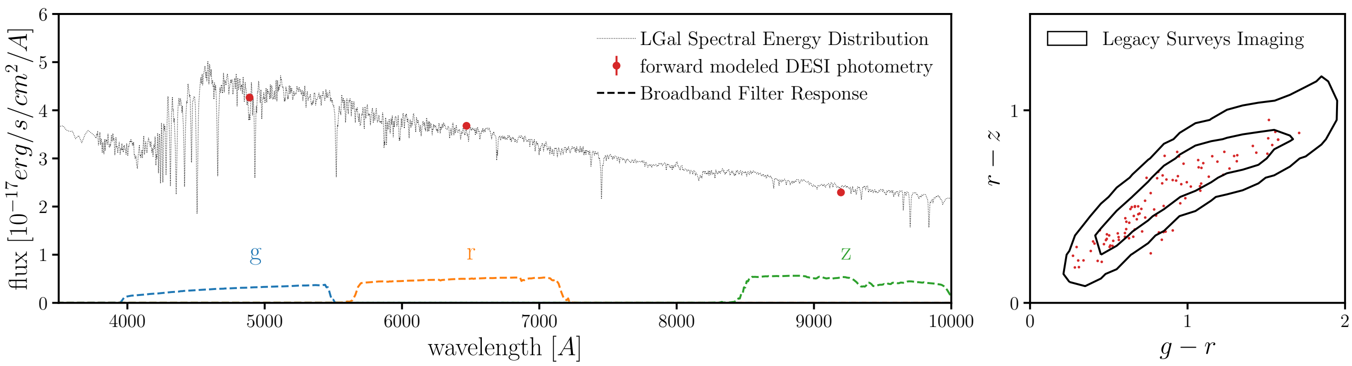

Figure 2. *Left*: We forward model DESI photometry (red) for our simulated galaxies (Section 2.1) by convolving their SEDs with the broadband filters (dashed). We construct the SEDs **CH:** using the star formation and metallicity histories (Section 2.2). We assign uncertainties to the forward modeled photometry by **CH:** matching the colors to the Legacy Survey imaging for BGS (Section 2.3). *Right*: The $g-r$ and $r-z$ color distribution of the forward modeled LGal photometry is consistent with the color distribution of BGS targets from Legacy Survey imaging (black contours).

overview of Lgal. small volume semi-analytic and cosmo-hydro simulations
describe what galaxy properties (SFH, ZH, etc) are available TODO

2.2. *Spectral Energy Distributions*

describe how the SED is generated using the SFH and ZHs TODO

2.3. *Forward Modeled DESI Photometry*

From the noiseless source spetra generated for the galaxies in our simulations (Section 2.1), we generate noiseless photometry by convolving the spectra with the broadband filters of DESI imaging:

$$f_X = \int f(\lambda) R_X(\lambda) \mathrm{d}\lambda \tag{1}$$

$f(\lambda)$, R_X is the transmission.

Fig. 2

Next, we assign realistic photometric uncertainties to the simulated photometric fluxes using the

Afterwards, we impose the target selection criteria of BGS: $r < 20$. This leaves us with X LGal galaxies. TODO

2.4. *Forward Modeled DESI Spectroscopy*

fiber flux assigned when we assigned photometric uncertainty

- 'true' fiber aperture SED generated by scaling down the SED by the assigned r-band fiber flux.
- Meanwhile we add some noise to the fiber flux to have 'measured' fiber flux.
- Furthermore, we later jointly fit photometry and spectra we marginalize over $f_{\rm fiber}$.

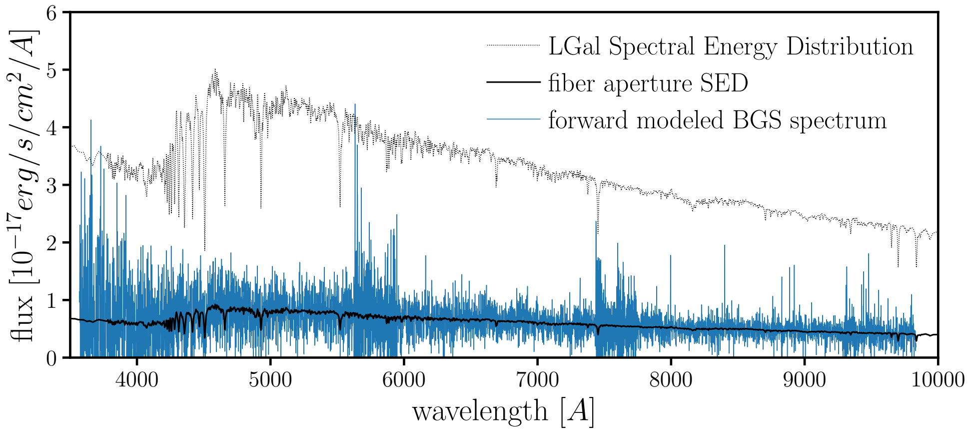

Figure 3. We construct forward model DESI spectra (blue) for our simulated galaxies by first applying a fiber aperture correction to the SED (black solid), then applying a DESI noise model. We apply a fiber aperture correction by scaling down the SED based on the r-band aperture flux we assign from the Legacy Surveys imaging. The noise model includes a DESI instrument model, which accounts for the DESI spectrograph response, and an atmosphere model, which accounts for the bright time observing conditions. We construct simulated DESI spectra for 8 different bright time observing conditions. For more details, we refer readers to Section 2.4.

- run true fiber aperture SED through the DESI noise model to produce DESI-like spectra.
- describe the noise model in detail
- [1]

A paragraph on how we combine these to get the spectrophotometry with figure **figure** showing the spectra and photometry on top of each other

3. INFERRING GALAXY PROPERTIES FROM PHOTOMETRY AND SPECTRA

- overview of SED fitting
- refer to appendix about different set ups
- be explicit about what physical properties are inferred from our SED fitting: stellar masses, ages, metallicities, star-formation histories. Expalin how we derive SFRs from the SFHs (i.e. our choice of 100Myr)
- brief description of other standard SED fitting methods we include for comparison: firefly, VESPA, CIGALE

description of our speculator SED model (Alsing et al. 2019), which is based on FSPS. We use Chabrier IMF **CH:** do we need to justify htis?.

We don't use tau model for SFH because it biases the physical parameters inferred. citation. TODO

[1] https://specsim.readthedocs.io

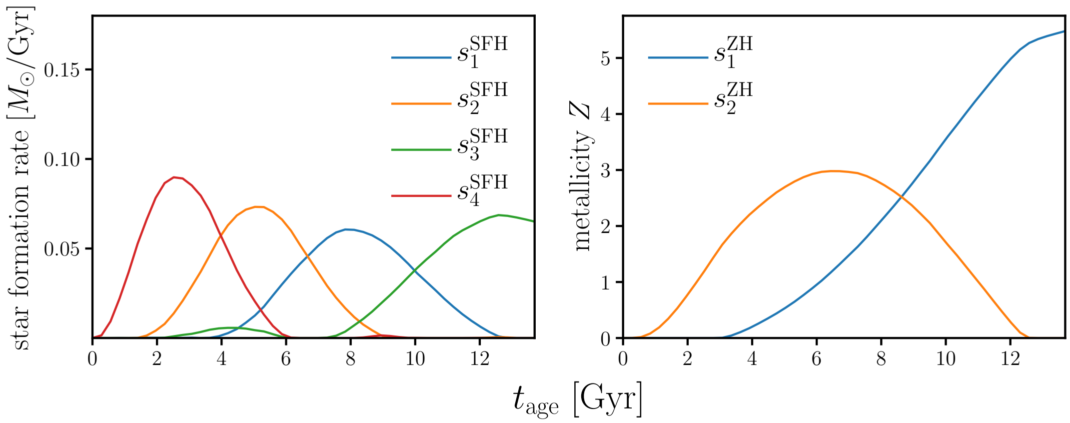

Figure 4. speculator bases

Instead we use the following SFH bases from Rita in prep.

$$\mathrm{SFH}(t, t_{\rm age}) = \sum_{i=1}^{4} \beta_i^{\rm SFH} \frac{s_i^{\rm SFH}(t)}{\int\limits_0^{t_{\rm age}} s_i^{\rm SFH}(t)\,{\rm d}t} \tag{2}$$

Similarly for ZH, we use the following ZH bases

$$\mathrm{ZH}(t) = \sum_{i=1}^{2} \gamma_i^{\rm ZH} s_i^{\rm ZH}(t) \tag{3}$$

$$\mathrm{SFR}_{\rm 100Myr} = \frac{\int\limits_{t_{\rm age}-100{\rm Myr}}^{t_{\rm age}} \mathrm{SFH}(t)\,{\rm d}t}{100\,{\rm Myr}} \tag{4}$$

see Figure 4.

Calzetti dust attenuation curve

Rather than evaluating the SPS directly using FSPS we use Speculator, an emulator for FSPS. Brief explanation of the PCA neural net. motivate why we're using speculator by reiterating Alsing et al. (2019)s speed profiling.

describe our MCMC sampling. James's work in convergence here. talk about how in principle speculator can easily be used with HMC because you can get derivatives with backpropagation. provide detailed profiling of SED fitting and projections for full 10million galaxy BGS sample.

In Figure XXXX we present the posterior as well as the best-fit photometry and spectroscopy using speculator + emcee.

4. RESULTS

Figuree 5

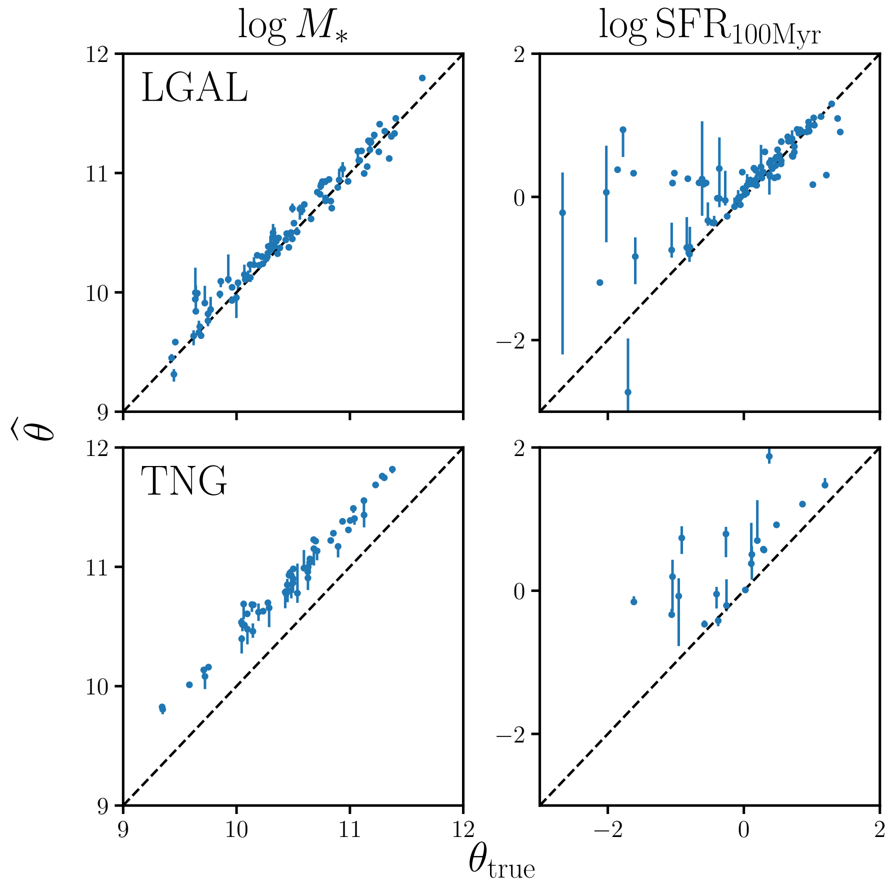

Figure 5. The properties inferred from ifsps spetrophotometry fit as a function of true properties.

In order to quantify the precision and accuracy of the inferred physical properties for our simulated galaxy population, we begin by assuming that the discrepancy between the inferred and true parameters for each galaxy $\Delta_{\theta,i}$)

$$\theta_i^{\rm inf} = \theta_i^{\rm true} + \Delta_{\theta,i} \tag{5}$$

where $\Delta_{\theta,i}$ is sampled from a Gaussian distribution

$$\Delta_{\theta,i} \sim \mathcal{N}(\mu_{\Delta_\theta}, \sigma_{\Delta_\theta}). \tag{6}$$

This Gaussian distribution is described by population hyperparameters μ_{Δ_θ} and σ_{Δ_θ}, the mean and standard deviation, which quantify the accuracy and precision of the inferred physical properties for the population.

Given the photomety and spectrum of our galaxies, $\{\boldsymbol{D}_i\}$, we can get the posteriors for these population hyperparameters $\eta_\Delta = \{\mu_{\Delta_\theta}, \sigma_{\Delta_\theta}\}$ using a hierarchical Bayesian framework (Hogg et al.

2010):

$$p(\eta_\Delta \,|\, \{\boldsymbol{D}_i\}) = \frac{p(\eta_\Delta)\; p(\{\boldsymbol{D}_i\} \,|\, \theta_\Delta)}{p(\{\boldsymbol{D}_i\})} \tag{7}$$

$$= \frac{p(\eta_\Delta)}{p(\{\boldsymbol{D}_i\})} \int p(\{\boldsymbol{D}_i\} \,|\, \{\theta_i\})\; p(\{\theta_i\} \,|\, \eta_\Delta)\; \mathrm{d}\{\theta_i\}. \tag{8}$$

Naively the posteriors for each of the galaxies are not correlated, so we can factorize the expression above

$$p(\eta_\Delta \,|\, \{\boldsymbol{D}_i\}) = \frac{p(\eta_\Delta)}{p(\{\boldsymbol{D}_i\})} \prod_{i=1}^{N} \int p(\boldsymbol{D}_i \,|\, \theta_i)\; p(\theta_i \,|\, \theta_\Delta)\; \mathrm{d}\theta_i \tag{9}$$

$$= \frac{p(\eta_\Delta)}{p(\{\boldsymbol{D}_i\})} \prod_{i=1}^{N} \int \frac{p(\theta_i \,|\, \boldsymbol{D}_i)\; p(\boldsymbol{D}_i)}{p(\theta_i)}\; p(\theta_i \,|\, \eta_\Delta)\; \mathrm{d}\theta_i \tag{10}$$

$$= p(\eta_\Delta) \prod_{i=1}^{N} \int \frac{p(\theta_i \,|\, \boldsymbol{D}_i)\; p(\theta_i \,|\, \eta_\Delta)}{p(\theta_i)}\; \mathrm{d}\theta_i. \tag{11}$$

$p(\theta_i \,|\, \boldsymbol{D}_i)$ is the posterior for galaxy i. Hence, the integral can be which means the integral can be estimated using the MCMC sample from the posterior

$$p(\eta_\Delta \,|\, \{\boldsymbol{D}_i\}) = p(\eta_\Delta) \prod_{i=1}^{N} \frac{1}{S_i} \sum_{j=1}^{S_i} \frac{p(\theta_{i,j} \,|\, \eta_\Delta)}{p(\theta_{i,j})}. \tag{12}$$

S_i is the number of MCMC samples and $\theta_{i,j}$ is the $j^{\rm th}$ sample of galaxy i. We present the maximum a posteriori (MAP) estimates of η_Δ for $\log\,M_*$ and $\log\,$SFR in Figure 6.

η_Δ as a function of SNR/mag/colour.

discussion:

- sfh basis improves SFH accuracy?
- advantage of mcmc over MAP
- non-Gaussian posteriors
- science applicatoins
- comparison to other methods
- different observing condition doesn't impact our results?
- we use super simple dust mmodel. **CH:** more sophistiicated dust models doesn't matter?
- the simulated spectra and the fitting are generated using the same SED. **CH:** We try with different libraries.
- **CH:** we try fitting with a different IMF

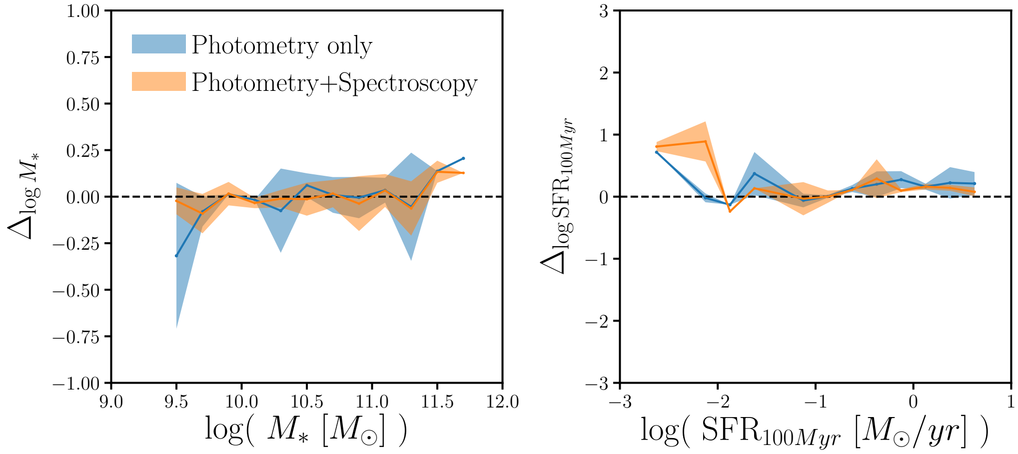

Figure 6. The discrepancies between the inferred and input/"true" M_*s (left) and SFRs (right) for our LGal galaxies. In blue, we infer M_*s and SFRs using only photometry; in orange, we infer M_*s and SFRs by jointly fitting both photometry and spectroscopy. *Jointly fitting spectroscopy and photometry improves constraints on galaxy properties.*

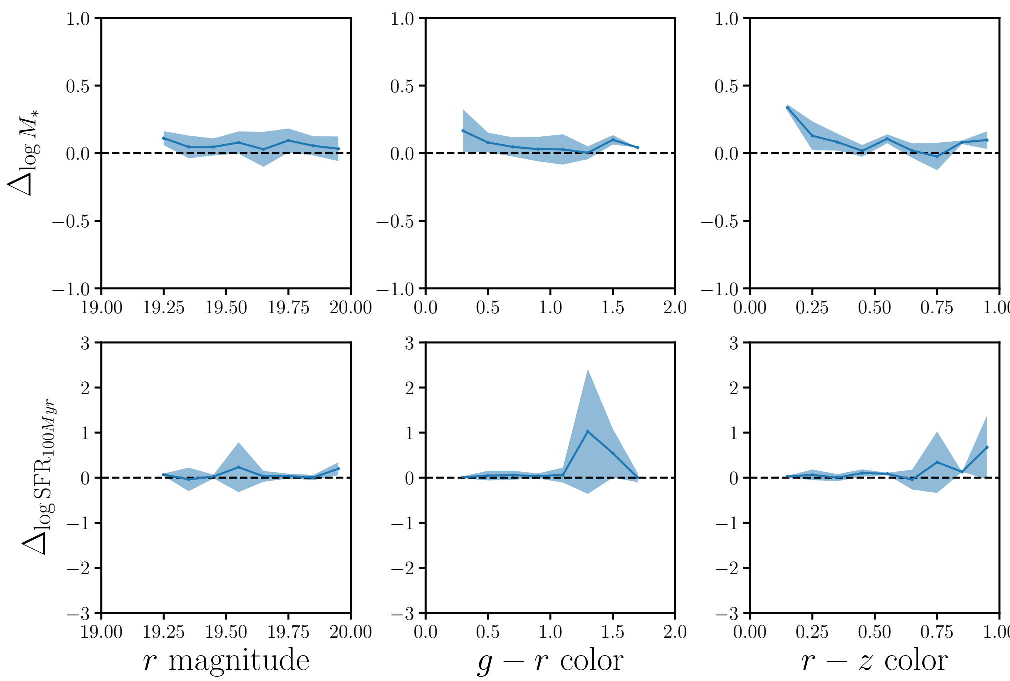

Figure 7. do we want to test bias against other properties? (e.g. obs condition?)

5. SUMMARY

DESI is great.

Also advertise science papers with specific focus on mock challenge science papers.

ACKNOWLEDGEMENTS

Table 1. Table of $\theta_{\rm inf}$ - $\theta_{\rm true}$ and uncertainties for different CIGALE setups

set	CIGALE A	CIGALE B	CIGALE C	CIGALE D
ΔM_{tot}	0.07	0.09	0.27	0.05
M_{err}	0.13	0.20	0.16	0.16
Δ Age	1.78	1.59	1.97	1.92
Age_{err}	2.36	2.63	2.47	2.47
Δ Z	0.0037	0.0026	0.0027	0.0027
Z_{err}	0.0091	0.0085	0.0089	0.0089

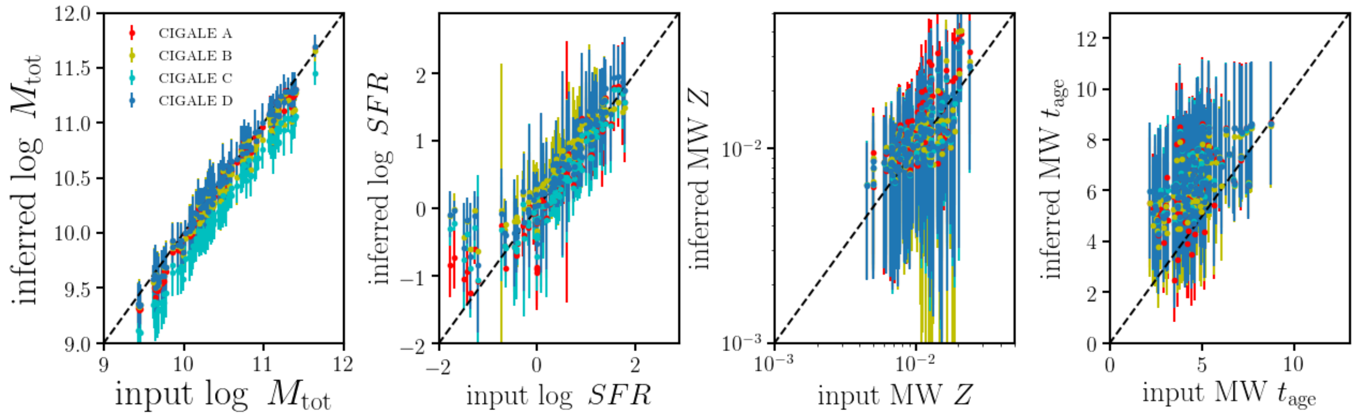

Figure 8. The properties inferred from CIGALE photometry fit as a function of true properties. Configuration CIGALE A, B, C, and D on one plot.

It's a pleasure to thank Joel Leja ... This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under contract No. DE-AC02-05CH11231. This project used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

APPENDIX

A. COMPARISON TO OTHER METHODS

REFERENCES

Alsing J., et al., 2019, arXiv:1911.11778 [astro-ph]

Dey A., et al., 2019, AJ, 157, 168

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