Draft version December 12, 2020
Typeset using LaTeX default style in AASTeX63

DESI Speculator: Stellar Population Synthesis emulator for PROVA-BGS

K. J. Kwon[1,2] and ChangHoon Hahn[3]

[1] *Department of Astronomy, University of California, Berkeley, CA 94720-3411, USA*
[2] *Department of Physics, University of California, Berkeley, CA 94720-3411, USA*
[3] *Department of Astrophysical Sciences, Princeton University, 4 Ivy Ln, Princeton, NJ 08544, USA*

Submitted to

ABSTRACT

We train a neural net emulation model for stellar population synthesis, Speculator (Alsing et al. 2020) to predict DESI galaxy spectra and photometry. We confirm that DESI Speculator reproduces the same SEDs generated by Fsps given a set of galaxy properties with percent level accuracy. We further demonstrate that DESI Speculator can be used to construct fully probabilistic value-added catalogs of bright galaxies from upcoming DESI Bright Galaxy Survey.

Keywords: keywords

1. INTRODUCTION

2. SPS MODELS

2.1. Fsps

Stellar population synthesis (SPS) technique parameterizes the spectral energy distributions of stellar population in terms of their stellar and dust properties. Through SPS evaluation, the properties of galaxies can be estimated from their SEDs, and vice versa. There has been continuous research in SPS techniques, which resulted in multiple publicly available models, such as Bruzual & Charlot (2003) and Maraston (2005). However, galaxy evolution extensively pertains to complicatedly interweaved factors that we do not have clear understanding on. The investigation by Conroy et al. (2009) and Conroy & Gunn (2010) demonstrated that neglecting such factors can lead to systematic errors in derived properties of galaxies. Hence, providing and propagating adequate uncertainties of SPS parameters are critical components in SPS evaluation.

Motivated by the capability of Flexible Stellar Population Synthesis (Fsps) developed by Conroy et al. (2009) and Conroy & Gunn (2010) to propagate uncertainties, we utilize `python-fsps`[1] (Foreman-Mackey et al. 2014), the python implementation of Fsps, to simulate SEDs from the SPS model parameters. We initialize the Fsps Stellar Population object with following choices of the parameters.

```
import fsps
ssp = fsps.StellarPopulation(zcontinuous=1,          # interpolate metallicities
                             sfh=4,                  # sfh type
                             dust_type=4,            # Kriek & Conroy attenuation curve
                             imf_type=1)             # chabrier
```

The initialized stellar population object is configured such that the metallicites are interpolated to the specified values of $\log(Z/Z_{\odot})$, the star formation history is defined by the delayed tau-model, the dust attenuation is modeled with Kriek & Conroy (2013) attenuation curve, and the initial mass function is given by Chabrier (2003). For the rest of the defaulted parameters, we defer reader to `python-fsps` documentation. The initialized SPS model allows for the

kgb0255@berkeley.edu

[1] https://github.com/dfm/python-fsps

choice of model parameters given in Table 1. We sample from the priors to generate a set of galaxy properties and simulate SEDs. Then, we use `emcee` (Foreman-Mackey et al. 2013), an open-source python package for Monte-Carlo Markov Chain (MCMC) methods to infer the posteriors from the SEDs.

Table 1. Fsps model parameters

Parameters	Description	Priors
τ	e-folding time in the delayed tau-model for the star formation history	$\mathcal{U}(0.1, 10)$
dust_1	Dust parameter describing the attenuation of young stellar light	$\mathcal{U}(0, 4)$
dust_2	Dust parameter describing the attenuation of old stellar light	$\mathcal{U}(0, 4)$
$\mathrm{dust}_{\mathrm{index}}$	The strength of the UV Bump in Kriek & Conroy (2013) attenuation curve	$\mathcal{U}(-2.2, 0.4)$
$\log M^*$	log Total mass in $M_\odot$	$\mathcal{U}(8, 13)$

Note—The details for dust_1 and dust_2 parameters can be found in Conroy et al. (2009). The threshold age for the *old* and *young* stellar population is defaulted to 10^7 years.

2.2. Speculator

DESI Bright Galaxy Survey (BGS) will observe ~ 10 million spectra of magnitude-limited samples of galaxies (Ruiz-Macias et al. 2020). With the huge number samples, the SPS evaluation is highly likely to entail significantly high computational cost, which necessitates a fast and accurate SPS model.

Speculator by Alsing et al. (2020), is a fully-connected neural net framework that emulates SPS models. To predict spectra, Speculator performs principal component analysis (PCA) to decompose the SEDs into basis functions, and parameterizes the basis coefficients in terms of the SPS model parameters. We target to emulate Fsps due to its aforementioned advantage in propagating uncertainties. The model parameters that we used is given in Table 2. Note that the model parameters include β_i^{SFH} and γ_i^{ZH}, which are the basis coefficients of respective basis functions. Speculator assumes that the star formation and chemical enrichment histories can be represented by linear combination of their basis functions. The basis functions are pre-established functions of lookback time, obtained by applying a non-negative matrix factorization to the simulated star formation and chemical enrichment histories of galaxies. For the details about the basis functions for star formation and chemical enrichment histories, we defer reader to Alsing et al. (2020), particularly to Figure 3.

Speed-up
racy

Table 2. The training parameters for DESI Speculator

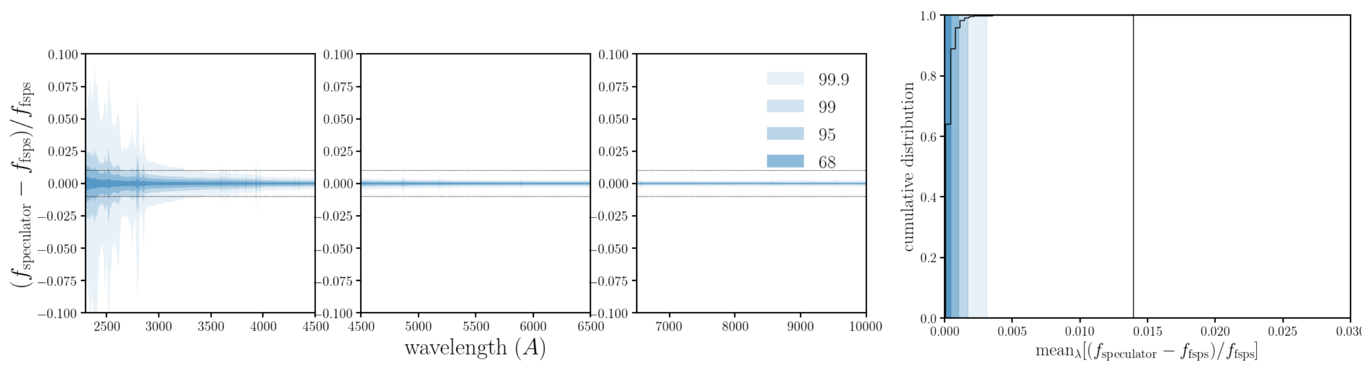
Parameters	Description	Priors
β_1^{SFH}, β_2^{SFH}, β_3^{SFH}, β_4^{SFH}	Star formation history basis function coefficients	$\mathcal{U}(0, 1)$
γ_1^{ZH}, γ_2^{ZH}	Metallicty enrichment history basis function coefficients	$\mathcal{U}(6.5\mathrm{e}{-3}, 7.5\mathrm{e}{-3})$
dust_1	young stellar light attenuation	$\mathcal{U}(0, 3.5)$
dust_2	old stellar light attenuation	$\mathcal{U}(0, 4)$
$\mathrm{dust}_{\mathrm{index}}$	The strength of the UV Bump in Kriek & Conroy (2013) attenuation curve	$\mathcal{U}(-2.5, 0.5)$
$\log M^*$	log Total mass in $M_\odot$	$\mathcal{U}(8.6, 13.8)$

2.2.1. DESI Speculator: *Training & Validation*

To construct the training dataset for DESI Speculator, we first sample large amount of training parameters from the priors listed in Table 2. We generate 5 million combinations of training parameters, which we feed into Fsps to obtain 5 million galaxy SEDs. Each SED is shifted and scaled for normalization. To commence the training, we bin SEDs by three wavelength ranges: $2300\text{Å} \le \lambda < 4500\text{Å}$, $4500\text{Å} \le \lambda < 6500\text{Å}$, and $6500\text{Å} \le \lambda < 11030\text{Å}$. Each of the binned SEDs is decomposed with N_{pca} principal component basis vectors, where N_{pca} varies between the wavelength bins. We emprically found that $N_{pca} = 50$, 30, 30 for each of the wavelength bins to be optimal. Further, we choose to parameterize the basis coefficients with four layers of 256 hidden units, whereas Alsing et al. (2020) used three hidden layers.

We generate the validation dataset in the same way as we generated the training dataset. We sample from priors of the training parameters, and the sampled parameters are run through FSPS to obtain SEDs. Then, the SEDs are binned and decomposed into principal component basis vectors, where the number of the basis vectors for each wavelength bin matches N_{pca} of the trained DESI SPECULATOR.

2.2.2. DESI SPECULATOR: *Validation Result*



(a) (b)

Figure 1. (a) shows the validation results plotted separately for each wavelength bin in terms of the fractional error – $(f_{\rm speculator} - f_{\rm fsps})/f_{\rm fsps}$. The Y axis denotes fractional error of the predicted SEDs, and the X axis denotes the wavelength (Angstrom). The fractional errors for 68%, 95%, 99%, and 99.9% confidence intervals are shown. (b) shows the cumulative distribution of the mean absolute fractional error. For example, an imaginary point $(0.002, 0.5)$ on the graph would imply that 50% of the validation results has fractional error of 0.002 or lower when averaged over the entire wavelength range.

The training and validation procedures were repeated with different training configurations – N_{pca} and number of training samples – until the best validation result was acquired. As shown in Figure 1, the trained model achieves $< 1\%$ accuracy over the most of the wavelength range.

3. RESULTS

3.1. *SED Generation*

Simulate galaxy properties using L-GAL library from igal 0-96. Run jobs to infer galaxy properties (MCMC). Do this with both SPECULATOR and FSPS.

3.2. *Comparison with* FSPS

Similar plots as DESI UG Research Forum, but with FSPS $\hat{\theta}$, not True θ. Explain quantitatively about accuracy.

4. CONCLUSION

DESI SPECULATOR achieves $< 1\%$ accuracy. Can be used for PROVA-BGS because it's much faster and sufficiently accurate.

5. ACKNOWLEDGEMENT

REFERENCES

Alsing, J., Peiris, H., Leja, J., et al. 2020, ApJS, 249, 5, doi: 10.3847/1538-4365/ab917f

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000, doi: 10.1046/j.1365-8711.2003.06897.x

Chabrier, G. 2003, PASP, 115, 763, doi: 10.1086/376392

Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833, doi: 10.1088/0004-637X/712/2/833

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486, doi: 10.1088/0004-637X/699/1/486

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, doi: 10.1086/670067

Foreman-Mackey, D., Sick, J., & Johnson, B. 2014, python-fsps: Python bindings to FSPS (v0.1.1), v0.1.1, Zenodo, doi: 10.5281/zenodo.12157

Kriek, M., & Conroy, C. 2013, ApJL, 775, L16, doi: 10.1088/2041-8205/775/1/L16

Maraston, C. 2005, MNRAS, 362, 799, doi: 10.1111/j.1365-2966.2005.09270.x

Ruiz-Macias, O., Zarrouk, P., Cole, S., et al. 2020, Research Notes of the American Astronomical Society, 4, 187, doi: 10.3847/2515-5172/abc25a