Population-Level Inference for Galaxy Properties from Broadband Photometry

Jiaxuan Li [1] **Peter Melchior** [1 2] **ChangHoon Hahn** [1] **Song Huang** [3]

Abstract

We present a method to infer galaxy properties and redshifts at the population level from photometric data using normalizing flows. Our method POPSED can reliably recover the redshift and stellar mass distribution of 10^5 galaxies using SDSS $ugriz$ photometry with < 1 GPU-hour, being 10^6 times faster than the traditional SED modeling method. The approach can also be applied to spectroscopic data including DESI and Gaia XP spectra. Our method provides an efficient and self-consistent way to learn the population posterior without deriving the posteriors for every individual object and then combining them.

1. Introduction

Galaxies are the building blocks of the Universe. Billions of galaxies will be characterized by the upcoming surveys such as LSST (Ivezić et al., 2019), Euclid (Racca et al., 2016), and Roman (Spergel et al., 2015), opening a huge discovery space for the evolution of galaxies as a population. Decoding the physical properties of galaxies, including the star formation history (SFH) and chemical enrichment history, from the observed spectral energy distributions (SEDs) requires modeling their SEDs with stellar population synthesis (SPS) models (e.g., Conroy, 2013; Carnall et al., 2018; Johnson et al., 2021). After modeling individual SEDs, the posteriors of each galaxy can be combined to study the galaxy population including the stellar mass function (Wright et al., 2017) and star-forming main sequence (Speagle et al., 2014). However, SED modeling is a high-dimensional problem (typically > 10 dimensions) that requires evaluating the SPS model and sampling the posterior several million times for one galaxy. Bayesian SED fitting with traditional SPS models (e.g., FSPS, Conroy et al., 2009) takes ~ 20 CPU-hours per galaxy (Leja et al., 2019), making it computationally infeasible to analyze billions of galaxies. Although recent advances have accelerated SED fitting using neural networks and simulation-based inference (e.g., Alsing et al., 2020; Hearin et al., 2021; Hahn & Melchior, 2022; Wang et al., 2023), some of these methods need sophisticated training and still face the problem of rigorously combining individual posteriors (see §4).

In this paper, we present POPSED, an efficient and self-consistent method to infer the properties of galaxy populations from photometric data without fitting individual galaxies. Our method could reliably recover the population statistics of 10^5 galaxies within 1 GPU-hour. We describe our method in §2, validate it using mock and real data in §3, and discuss the implication and outlook in §4.

2. Method

We recover the population-level distribution of galaxy properties from photometry data as follows. We denote the observed data by $\{\boldsymbol{X}_i\}$ and the physical properties of any galaxy by $\boldsymbol{\theta}$. We seek to approximate the population posterior $p(\boldsymbol{\theta}|\{\boldsymbol{X}_i\})$ by a normalizing flow $p_{\boldsymbol{\phi}}(\boldsymbol{\theta})$, which is a highly flexible model for probability density distributions. By repeatedly sampling from the flow $\boldsymbol{\theta}_j \sim p_{\boldsymbol{\phi}}(\boldsymbol{\theta})$, we predict the corresponding photometry $\hat{\boldsymbol{X}}_j^{\boldsymbol{\phi}} = F(\boldsymbol{\theta}_j)$ using the forward model F, which is a neural network-based emulator for galaxy SEDs. We compare the distributions of the observed photometry and the predicted photometry by calculating the Wasserstein distance between the two distributions $D_W[p(\{\hat{\boldsymbol{X}}_j^{\boldsymbol{\phi}}\}), p(\{\boldsymbol{X}_i\})]$. This distance is then used as a loss to train the normalizing flow $\boldsymbol{\phi}$ using the gradient descent method. In the end, the flow will be able to approximate the population posterior such that the predicted photometry agrees well with the observed data. We train an ensemble of flows and combine their results. Figure 1 shows a schematic diagram describing our method.

2.1. SPS Model and SED emulator

The forward model F translates a set of physical parameters $\boldsymbol{\theta}_j$ to the corresponding photometry $\hat{\boldsymbol{X}}_j^{\boldsymbol{\phi}}$. This mapping is prescribed by an SPS model and is accelerated by an SED emulator. The SPS model predicts a galaxy spectrum by combing SFH, chemical enrichment history, initial mass function, dust model, and spectra templates for simple stellar populations. In this work, we take the `PROVABGS` model described in Alsing et al. (2020); Hahn et al. (2022a), where

[1]Department of Astrophysical Sciences, 4 Ivy Lane, Princeton University, Princeton, NJ 08544, USA [2]Center for Statistics & Machine Learning, Princeton University, Princeton, NJ 08544, USA [3]Department of Astronomy and Tsinghua Center for Astrophysics, Tsinghua University, Beijing 100084, China. Correspondence to: Jiaxuan Li <jiaxuanl@princeton.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

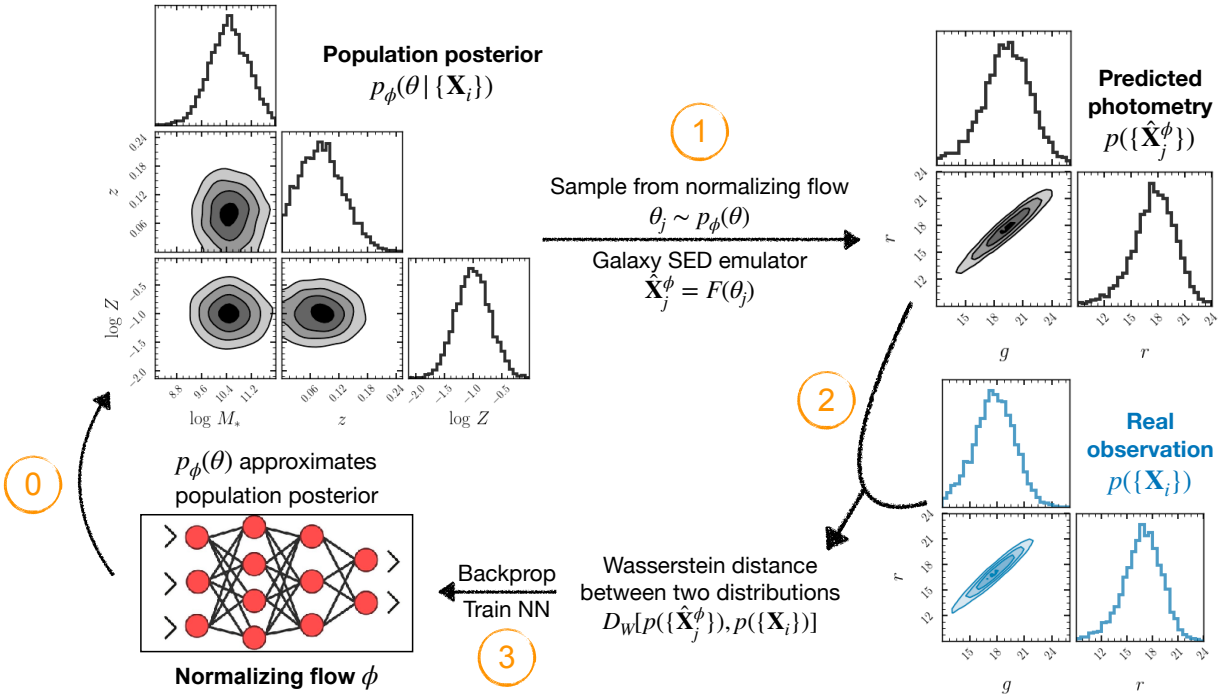

Figure 1. The schematic plot showing how POPSED works as described in §2. The population posterior $p_{\boldsymbol{\phi}}(\boldsymbol{\theta}|\{\boldsymbol{X}_i\})$ is approximated by a normalizing flow $p_{\boldsymbol{\phi}}(\boldsymbol{\theta})$. We sample from the normalizing flow and forward model the photometry $\{\hat{\boldsymbol{X}}_j^{\boldsymbol{\phi}}\}$ using the galaxy SED emulator $F(\boldsymbol{\theta}_j)$. Then we compare the distributions of observed photometry and the predicted photometry by calculating the Wasserstein distance $D_W[p(\{\hat{\boldsymbol{X}}_j^{\boldsymbol{\phi}}\}), p(\{\boldsymbol{X}_i\})]$, which is used as a loss to train the normalizing flow. After training, the population posterior describes a galaxy population whose photometry distribution agrees with observation.

the SFH is described by a linear combination of four bases and one burst component. The SFH bases are generated from galaxies in the Illustris simulation (Vogelsberger et al., 2014) using the non-negative matrix factorization. We simplify the metallicity history to be a constant metallicity Z over time. We use `FSPS` (Conroy et al., 2009; Johnson et al., 2021) to generate galaxy spectrum for given redshift z, total formed stellar mass $M_\star$, SFH, and metallicity Z. Then we add dust attenuation to the galaxy spectra following the Charlot & Fall (2000) recipe. As shown in Hahn et al. (2022a), this SPS prescription is flexible enough to model real galaxies. In total, our SPS model contains 12 parameters as summarized in Table 1 in the appendix. We emphasize that any SPS model can be used to perform population-level inference just by retraining the SED emulator, as described below.

We train a differentiable emulator for galaxy SED following Alsing et al. (2020) where a neural network is trained to predict the PCA coefficients of a galaxy spectrum. The emulator takes the physical parameters listed in Table 1 and predicts the corresponding restframe spectra from 1,000Å to 60,000Å that is later shifted to the observed frame for a given redshift. In order to generate representative training data, we sample the SPS parameter space according to the prior distributions listed in Table 1. We use uninformative priors to avoid introducing any bias in training the emulator. Trained with 3×10^6 spectra, the emulator achieves an accuracy of ~ 0.01 mag in SDSS $ugriz$ filters (Doi et al., 2010). Noise needs to be added in order to meaningfully forward model the observed data. We apply Gaussian noise with uncertainties sampled from an SNR distribution, which we describe in further detail in §2.4.

2.2. Normalizing flow

We use normalizing flows to approximate the population posterior $p_{\boldsymbol{\phi}}(\boldsymbol{\theta}|\{\boldsymbol{X}_i\})$. Normalizing flows (Tabak & Vanden-Eijnden, 2010; Kobyzev et al., 2019) map a complex distribution $p(\boldsymbol{\theta})$ to a simple base distribution $\pi(\boldsymbol{z})$ using an invertible bijective transformation parameterized by a neural network. From many different flow models, we use the Neural Spline Flow model (NSF, Durkan et al., 2019) where the base function is a multivariate Gaussian distribution and the transformations are described by monotonic rational-quadratic splines. The flexibility of the NSF model is well-suited to describe the SPS parameter distribution. We use the NSF implementation in the `sbi` package (Greenberg et al., 2019; Tejero-Cantero et al., 2020).

The original NSF model is initialized such that $p_{\boldsymbol{\phi}}(\boldsymbol{\theta})$ is a standard multivariate Gaussian distribution. However, this initialization presents challenges during training. Firstly, nonphysical parameters (e.g., negative redshift) will be drawn from this initial distribution. Secondly, the strong

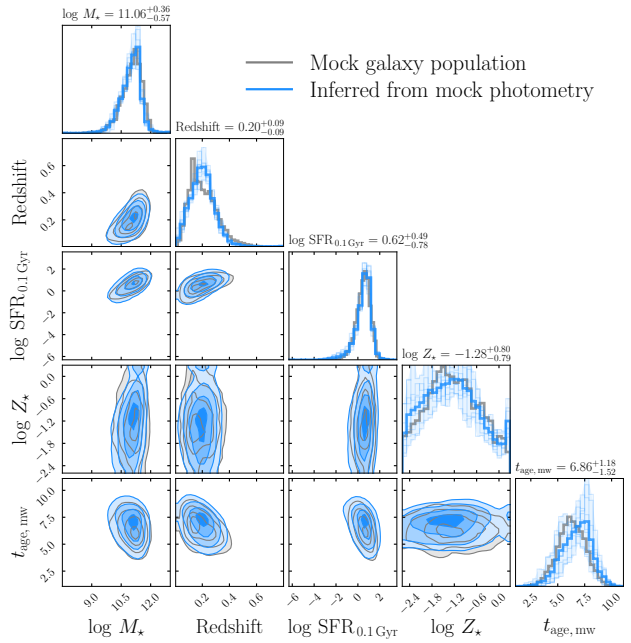

Figure 2. Mock galaxy population (gray contours) and the inferred galaxy population (blue contours) using our method. The mock observations are in SDSS $ugriz$-bands and imitate the noise property of GAMA. The lighter blue histograms show the individual normalizing flow and the dark blue histogram is the result after combining 20 flows.

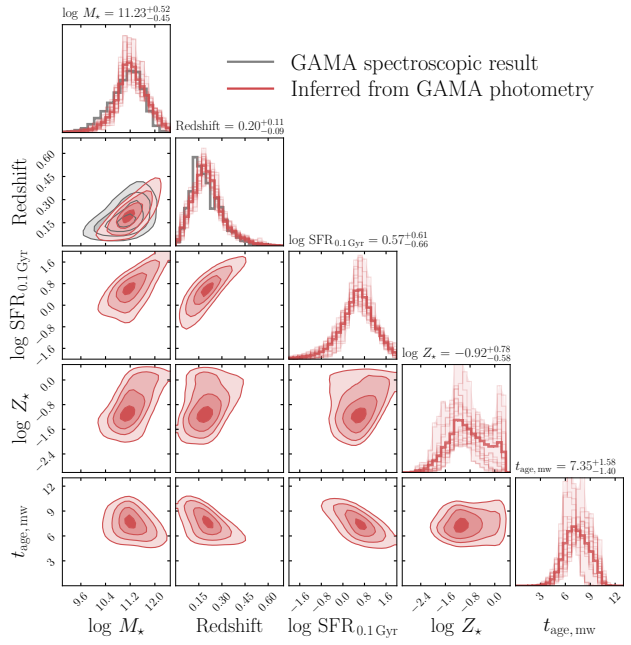

Figure 3. The inferred galaxy population from GAMA photometry data is shown as red contours. For comparison, the gray histograms show the spectroscopic redshift and the stellar mass from the GAMA catalog. The redshift and stellar mass distribution inferred from photometry are in excellent agreement with spectroscopic redshifts.

prior imposed by this initial distribution becomes problematic when the available data is not informative enough. To address these concerns, we append a cumulative distribution function (CDF) transformation layer to the flow which converts a Gaussian distribution into a uniform distribution within a predefined range (Table 1), ensuring that all parameters drawn from $p_{\boldsymbol{\phi}}(\boldsymbol{\theta})$ are physically valid.

2.3. Training

Once the normalizing flow is initialized, we sample from $p_{\boldsymbol{\phi}}(\boldsymbol{\theta})$ and generate the corresponding photometry $\{\hat{\boldsymbol{X}}_j^{\boldsymbol{\phi}}\}$ using the emulator F. The difference between the observed and predicted photometry is considered as the loss for training the flow. The commonly-used Kullback–Leibler (KL) divergence is hard to apply here because the photometry data is high-dimensional and another density estimator for $p(\{\boldsymbol{X}_i\})$ is needed to evaluate the KL divergence. Therefore, we turn to the Wasserstein distance (also known as the earth mover's distance) which describes the minimum "cost" of moving a distribution to the target distribution. The Wasserstein distance is symmetric about the two distributions and has been applied in Generative Adversarial Networks (GAN, Arjovsky et al., 2017)) and approximate Bayesian inference (e.g., Ambrogioni et al., 2018; Bernton et al., 2019). We use the Python package `GeomLoss` (Charlier et al., 2021; Feydy et al., 2019) to compute the Wasserstein distance $D_W[p(\{\hat{\boldsymbol{X}}_j^{\boldsymbol{\phi}}\}), p(\{\boldsymbol{X}_i\})]$, whose gradient is back-propagated to train the flow $\boldsymbol{\phi}$. The whole network is implemented in `PyTorch` (Paszke et al., 2019) and trained with the `Adam` optimizer (Kingma & Ba, 2014).

The realistic noise added to the predicted photometry $\hat{\boldsymbol{X}}_j^{\boldsymbol{\phi}}$ can make the training much harder at the beginning when the normalizing flow has not yet reached a plausible solution. Therefore, we take an "annealing" strategy where the noise (described by an effective SNR) is added gradually as the training goes on. We set the SNR of the noise in the forward model to follow an exponential decay $\mathrm{SNR}(t) = 1 + \mathrm{SNR}_0 \cdot \exp(-\tau \cdot t/T)$, where τ is the decay rate and T is the total epochs of training. In this way, the normalizing flow will be guided by the bulk of the data in the beginning without paying much attention to the noise. The flow is exposed to more realistic noises later on and adjusts its shape to match the details in the observed data. We set $\mathrm{SNR}_0 = 30$, $\tau = 12$, $T = 800$ for the case studies in Section 3. The time for training one flow is about 40 mins using one NVIDIA A100 for 10^5 galaxies.

Since the photometry data is not informative enough to put stringent constraints on the population posterior, a single normalizing flow might be trapped into a local optimum. We take the ensemble learning approach and train a number of flows with the same architecture but different random seeds to map out the full distribution of physical parameters. We combine these models by sampling from each of them

and combining the samples.

2.4. Data

To test the performance of POPSED on inferring the galaxy properties and redshifts, we use the photometry data from the Galaxy And Mass Assembly (GAMA, Driver et al., 2011) survey Data Release 3 (DR3, Baldry et al. 2018). The GAMA survey is a spectroscopic survey targeting galaxies selected from the photometric surveys down to $r < 19.8$ mag. It is therefore an ideal data set to test how well can we recover the redshift distribution by comparing it with the spectroscopic redshifts. The photometry in $ugriz$-bands comes from SDSS DR7 (Abazajian et al., 2009). We construct the noise model using the aperture-matched photometry (Driver et al., 2016) by calculating the SNR for each source in the catalog and deriving the mean (μ_X) and the standard deviation (σ_X) of the SNR as a function of magnitude in each band. Despite its simplicity, this model describes the noise property of GAMA data well. In the end, we apply an additional cut $\mathrm{SNR}_X > 1$ in all bands to remove marginally-detected objects with poor photometry quality.

3. Results

3.1. Mock Observation

We design a mock observation to validate POPSED. The parameter distribution $p(\boldsymbol{\theta})_{\mathrm{mock}}$ for the mock test is set to mimic a real galaxy population. The stellar mass and redshift follow their joint distribution in GAMA DR3, whereas the other parameters follow truncated Gaussian distributions as described in the appendix. The true distributions of characteristic quantities such as redshift, total-formed stellar mass, recent SFR, metallicity, and mass-weighted age are shown in Figure 2 as gray contours. Then we generate mock data $\{\boldsymbol{X}_i\}$ for 10^5 galaxies in SDSS $ugriz$-bands using our forward model.

We run POPSED on these mock data by training 20 flows with different random seeds. The recovered parameter distribution is shown as blue contours in Figure 2. The light blue shades show the variation among the flows. The inferred galaxy population is in excellent agreement with the ground truth in all dimensions. Moreover, the ensemble of flows (blue shades) covers the true distribution (gray lines). We find that photometric data can tightly constrain the distributions of stellar mass, redshift, and SFR. Nevertheless, the metallicity and mass-weighted age have relatively large variations among different flows, meaning that the data is less informative to constrain them.

3.2. GAMA Data

Motivated by the success of the mock test, we apply POPSED to real GAMA photometry data. After the quality cuts (see §2.4) and an additional signal-to-noise ratio cut $\mathrm{SNR} > 1$ for all SDSS $ugriz$-bands, there are 83,692 objects in our GAMA sample. We use the `AUTO` photometry of SDSS $ugriz$-bands in the GAMA DR3 catalog.

Figure 3 shows the inferred galaxy population as red contours. We overlay the spectroscopic redshifts and the stellar masses from GAMA as black histograms without considering the uncertainties in GAMA measurement. Our inferred redshift distribution is in excellent agreement with the spectroscopic redshift distribution, despite the fact that photometry is much less informative than spectroscopy on constraining redshifts. Our recovered stellar mass distribution also agrees with the GAMA result but is skewed towards the high-mass end. This level of discrepancy is expected, since the Gama stellar mass is derived using an SPS model with different SFH and spectral templates (Taylor et al., 2011). POPSED also reveals the correlation among the physical parameters including the correlation between stellar mass and recent SFR (i.e., star-forming main sequence) shown in Figure 3.

4. Discussion

Population analyses of galaxies have provided valuable insights into astrophysics, but traditional SED modeling on individual galaxies is very expensive. Despite recent advancements in accelerating SED fitting (e.g., Hahn & Melchior, 2022; Wang et al., 2023), analyzing a large sample of 10^8 galaxies still requires $\sim 10^5$ GPU-hours. One drawback of SBI-based methods is the cost of training and generating training data, making it difficult to change the SPS and noise prescription once the SBI model is trained. Our method recovers the population posterior 10^{2-3} times faster than the SBI-based methods and does not require demanding training. It is also straightforward to change the SPS model and noise level in POPSED, making it more flexible to be applied to different surveys.

Furthermore, it is non-trivial to construct the population posterior from individual posteriors or point estimates in a statistically rigorous way (e.g., Leistedt et al., 2016; Malz & Hogg, 2020; Wong et al., 2020; Alsing et al., 2022). The population posterior is often reconstructed using hierarchical Bayesian methods where hyper-parameters $\boldsymbol{\psi}$ describe the population posterior by $p(\boldsymbol{\psi}|\{\boldsymbol{X}_i\}) = p(\boldsymbol{\psi}) \cdot \Pi_{i=1}^{N} \int \frac{p(\boldsymbol{\theta}_i|\boldsymbol{X}_i)p(\boldsymbol{\theta}_i|\boldsymbol{\psi})}{p(\boldsymbol{\theta}_i)} \mathrm{d}\boldsymbol{\theta}_i$. Evaluating this posterior requires the calculation of N Monte Carlo integrals using samples from individual posteriors. Furthermore, MCMC is required to sample the population posterior, which will also be computationally intensive. POPSED skips obtaining individual posteriors but directly constrains the population posterior, which is of scientific interest for most studies.

Many science cases would benefit from having the population posterior in hand. It can be marginalized to learn

about the key relations such as stellar mass function and star-forming main sequence. The photo-z distribution presented in Fig. 3 is helpful for weak lensing studies (Mandelbaum, 2018). When applied to two samples that are selected differently (e.g., by optical colors), POPSED will tell how the two samples are different in their physical properties. To design future photometric surveys, one can sample from a population posterior to generate mock observations in different filters with different depths. The target selection for spectroscopic surveys can now be done by selecting interesting regions in the population posterior rather than in data space. Additionally, the idea presented in this paper could be extended to stellar spectra (e.g., Gaia XP spectra, Zhang et al., 2023) and galaxy spectra (e.g., DESI, Hahn et al., 2022b).

References

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Allende Prieto, C., An, D., Anderson, K. S. J., Anderson, S. F., Annis, J., Bahcall, N. A., Bailer-Jones, C. A. L., Barentine, J. C., Bassett, B. A., Becker, A. C., Beers, T. C., Bell, E. F., Belokurov, V., Berlind, A. A., Berman, E. F., Bernardi, M., Bickerton, S. J., Bizyaev, D., Blakeslee, J. P., Blanton, M. R., Bochanski, J. J., Boroski, W. N., Brewington, H. J., Brinchmann, J., Brinkmann, J., Brunner, R. J., Budavári, T., Carey, L. N., Carliles, S., Carr, M. A., Castander, F. J., Cinabro, D., Connolly, A. J., Csabai, I., Cunha, C. E., Czarapata, P. C., Davenport, J. R. A., de Haas, E., Dilday, B., Doi, M., Eisenstein, D. J., Evans, M. L., Evans, N. W., Fan, X., Friedman, S. D., Frieman, J. A., Fukugita, M., Gänsicke, B. T., Gates, E., Gillespie, B., Gilmore, G., Gonzalez, B., Gonzalez, C. F., Grebel, E. K., Gunn, J. E., György, Z., Hall, P. B., Harding, P., Harris, F. H., Harvanek, M., Hawley, S. L., Hayes, J. J. E., Heckman, T. M., Hendry, J. S., Hennessy, G. S., Hindsley, R. B., Hoblitt, J., Hogan, C. J., Hogg, D. W., Holtzman, J. A., Hyde, J. B., Ichikawa, S.-i., Ichikawa, T., Im, M., Ivezić, Ž., Jester, S., Jiang, L., Johnson, J. A., Jorgensen, A. M., Jurić, M., Kent, S. M., Kessler, R., Kleinman, S. J., Knapp, G. R., Konishi, K., Kron, R. G., Krzesinski, J., Kuropatkin, N., Lampeitl, H., Lebedeva, S., Lee, M. G., Lee, Y. S., French Leger, R., Lépine, S., Li, N., Lima, M., Lin, H., Long, D. C., Loomis, C. P., Loveday, J., Lupton, R. H., Magnier, E., Malanushenko, O., Malanushenko, V., Mandelbaum, R., Margon, B., Marriner, J. P., Martínez-Delgado, D., Matsubara, T., McGehee, P. M., McKay, T. A., Meiksin, A., Morrison, H. L., Mullally, F., Munn, J. A., Murphy, T., Nash, T., Nebot, A., Neilsen, Eric H., J., Newberg, H. J., Newman, P. R., Nichol, R. C., Nicinski, T., Nieto-Santisteban, M., Nitta, A., Okamura, S., Oravetz, D. J., Ostriker, J. P., Owen, R., Padmanabhan, N., Pan, K., Park, C., Pauls, G., Peoples, John, J., Percival, W. J., Pier, J. R., Pope, A. C., Pourbaix, D., Price, P. A., Purger, N., Quinn, T., Raddick, M. J., Re Fiorentin, P., Richards, G. T., Richmond, M. W., Riess, A. G., Rix, H.-W., Rockosi, C. M., Sako, M., Schlegel, D. J., Schneider, D. P., Scholz, R.-D., Schreiber, M. R., Schwope, A. D., Seljak, U., Sesar, B., Sheldon, E., Shimasaku, K., Sibley, V. C., Simmons, A. E., Sivarani, T., Allyn Smith, J., Smith, M. C., Smolčić, V., Snedden, S. A., Stebbins, A., Steinmetz, M., Stoughton, C., Strauss, M. A., SubbaRao, M., Suto, Y., Szalay, A. S., Szapudi, I., Szkody, P., Tanaka, M., Tegmark, M., Teodoro, L. F. A., Thakar, A. R., Tremonti, C. A., Tucker, D. L., Uomoto, A., Vanden Berk, D. E., Vandenberg, J., Vidrih, S., Vogeley, M. S., Voges, W., Vogt, N. P., Wadadekar, Y., Watters, S., Weinberg, D. H., West, A. A., White, S. D. M., Wilhite, B. C., Wonders, A. C., Yanny, B., Yocum, D. R., York, D. G., Zehavi, I., Zibetti, S., and Zucker, D. B. The Seventh Data Release

of the Sloan Digital Sky Survey. *ApJS*, 182(2):543–558, June 2009. doi: 10.1088/0067-0049/182/2/543.

Alsing, J., Peiris, H., Leja, J., Hahn, C., Tojeiro, R., Mortlock, D., Leistedt, B., Johnson, B. D., and Conroy, C. SPECULATOR: Emulating Stellar Population Synthesis for Fast and Accurate Galaxy Spectra and Photometry. *ApJS*, 249(1):5, July 2020. doi: 10.3847/1538-4365/ab917f.

Alsing, J., Peiris, H., Mortlock, D., Leja, J., and Leistedt, B. Forward modeling of galaxy populations for cosmological redshift distribution inference. *arXiv e-prints*, art. arXiv:2207.05819, July 2022.

Ambrogioni, L., Güçlü, U., Güçlütürk, Y., Hinne, M., Maris, E., and van Gerven, M. A. J. Wasserstein Variational Inference. *arXiv e-prints*, art. arXiv:1805.11284, May 2018. doi: 10.48550/arXiv.1805.11284.

Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein GAN. *arXiv e-prints*, art. arXiv:1701.07875, January 2017. doi: 10.48550/arXiv.1701.07875.

Baldry, I. K., Liske, J., Brown, M. J. I., Robotham, A. S. G., Driver, S. P., Dunne, L., Alpaslan, M., Brough, S., Cluver, M. E., Eardley, E., Farrow, D. J., Heymans, C., Hildebrandt, H., Hopkins, A. M., Kelvin, L. S., Loveday, J., Moffett, A. J., Norberg, P., Owers, M. S., Taylor, E. N., Wright, A. H., Bamford, S. P., Bland-Hawthorn, J., Bourne, N., Bremer, M. N., Colless, M., Conselice, C. J., Croom, S. M., Davies, L. J. M., Foster, C., Grootes, M. W., Holwerda, B. W., Jones, D. H., Kafle, P. R., Kuijken, K., Lara-Lopez, M. A., López-Sánchez, Á. R., Meyer, M. J., Phillipps, S., Sutherland, W. J., van Kampen, E., and Wilkins, S. M. Galaxy And Mass Assembly: the G02 field, Herschel-ATLAS target selection and data release 3. *MNRAS*, 474(3):3875–3888, March 2018. doi: 10.1093/mnras/stx3042.

Bernton, E., Jacob, P. E., Gerber, M., and Robert, C. P. Approximate Bayesian computation with the Wasserstein distance. *arXiv e-prints*, art. arXiv:1905.03747, May 2019. doi: 10.48550/arXiv.1905.03747.

Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koornneef, J., and Storchi-Bergmann, T. The Dust Content and Opacity of Actively Star-forming Galaxies. *ApJ*, 533(2):682–695, April 2000. doi: 10.1086/308692.

Carnall, A. C., McLure, R. J., Dunlop, J. S., and Davé, R. Inferring the star formation histories of massive quiescent galaxies with BAGPIPES: evidence for multiple quenching mechanisms. *MNRAS*, 480(4):4379–4401, November 2018. doi: 10.1093/mnras/sty2169.

Charlier, B., Feydy, J., Glaunès, J. A., Collin, F.-D., and Durif, G. Kernel operations on the gpu, with autodiff, without memory overflows. *Journal of Machine Learning Research*, 22(74):1–6, 2021. URL http://jmlr.org/papers/v22/20-275.html.

Charlot, S. and Fall, S. M. A Simple Model for the Absorption of Starlight by Dust in Galaxies. *ApJ*, 539(2):718–731, August 2000. doi: 10.1086/309250.

Conroy, C. Modeling the Panchromatic Spectral Energy Distributions of Galaxies. *ARA&A*, 51(1):393–455, August 2013. doi: 10.1146/annurev-astro-082812-141017.

Conroy, C., Gunn, J. E., and White, M. The Propagation of Uncertainties in Stellar Population Synthesis Modeling. I. The Relevance of Uncertain Aspects of Stellar Evolution and the Initial Mass Function to the Derived Physical Properties of Galaxies. *ApJ*, 699(1):486–506, July 2009. doi: 10.1088/0004-637X/699/1/486.

Doi, M., Tanaka, M., Fukugita, M., Gunn, J. E., Yasuda, N., Ivezić, Ž., Brinkmann, J., de Haars, E., Kleinman, S. J., Krzesinski, J., and French Leger, R. Photometric Response Functions of the Sloan Digital Sky Survey Imager. *AJ*, 139(4):1628–1648, April 2010. doi: 10.1088/0004-6256/139/4/1628.

Driver, S. P., Hill, D. T., Kelvin, L. S., Robotham, A. S. G., Liske, J., Norberg, P., Baldry, I. K., Bamford, S. P., Hopkins, A. M., Loveday, J., Peacock, J. A., Andrae, E., Bland-Hawthorn, J., Brough, S., Brown, M. J. I., Cameron, E., Ching, J. H. Y., Colless, M., Conselice, C. J., Croom, S. M., Cross, N. J. G., de Propris, R., Dye, S., Drinkwater, M. J., Ellis, S., Graham, A. W., Grootes, M. W., Gunawardhana, M., Jones, D. H., van Kampen, E., Maraston, C., Nichol, R. C., Parkinson, H. R., Phillipps, S., Pimbblet, K., Popescu, C. C., Prescott, M., Roseboom, I. G., Sadler, E. M., Sansom, A. E., Sharp, R. G., Smith, D. J. B., Taylor, E., Thomas, D., Tuffs, R. J., Wijesinghe, D., Dunne, L., Frenk, C. S., Jarvis, M. J., Madore, B. F., Meyer, M. J., Seibert, M., Staveley-Smith, L., Sutherland, W. J., and Warren, S. J. Galaxy and Mass Assembly (GAMA): survey diagnostics and core data release. *MNRAS*, 413(2):971–995, May 2011. doi: 10.1111/j.1365-2966.2010.18188.x.

Driver, S. P., Wright, A. H., Andrews, S. K., Davies, L. J., Kafle, P. R., Lange, R., Moffett, A. J., Mannering, E., Robotham, A. S. G., Vinsen, K., Alpaslan, M., Andrae, E., Baldry, I. K., Bauer, A. E., Bamford, S. P., Bland-Hawthorn, J., Bourne, N., Brough, S., Brown, M. J. I., Cluver, M. E., Croom, S., Colless, M., Conselice, C. J., da Cunha, E., De Propris, R., Drinkwater, M., Dunne, L., Eales, S., Edge, A., Frenk, C., Graham, A. W., Grootes,

M., Holwerda, B. W., Hopkins, A. M., Ibar, E., van Kampen, E., Kelvin, L. S., Jarrett, T., Jones, D. H., Lara-Lopez, M. A., Liske, J., Lopez-Sanchez, A. R., Loveday, J., Maddox, S. J., Madore, B., Mahajan, S., Meyer, M., Norberg, P., Penny, S. J., Phillipps, S., Popescu, C., Tuffs, R. J., Peacock, J. A., Pimbblet, K. A., Prescott, M., Rowlands, K., Sansom, A. E., Seibert, M., Smith, M. W. L., Sutherland, W. J., Taylor, E. N., Valiante, E., Vazquez-Mata, J. A., Wang, L., Wilkins, S. M., and Williams, R. Galaxy And Mass Assembly (GAMA): Panchromatic Data Release (far-UV-far-IR) and the low-z energy budget. *MNRAS*, 455(4):3911–3942, February 2016. doi: 10.1093/mnras/stv2505.

Durkan, C., Bekasov, A., Murray, I., and Papamakarios, G. Neural Spline Flows. *arXiv e-prints*, art. arXiv:1906.04032, June 2019. doi: 10.48550/arXiv.1906.04032.

Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S.-i., Trouve, A., and Peyré, G. Interpolating between optimal transport and mmd using sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2681–2690, 2019.

Greenberg, D. S., Nonnenmacher, M., and Macke, J. H. Automatic Posterior Transformation for Likelihood-Free Inference. *arXiv e-prints*, art. arXiv:1905.07488, May 2019. doi: 10.48550/arXiv.1905.07488.

Hahn, C. and Melchior, P. Accelerated Bayesian SED Modeling using Amortized Neural Posterior Estimation. *arXiv e-prints*, art. arXiv:2203.07391, March 2022.

Hahn, C., Kwon, K. J., Tojeiro, R., Siudek, M., Canning, R. E. A., Mezcua, M., Tinker, J. L., Brooks, D., Doel, P., Fanning, K., Gaztañaga, E., Kehoe, R., Landriau, M., Meisner, A., Moustakas, J., Poppett, C., Tarle, G., Weiner, B., and Zou, H. The DESI PRObabilistic Value-Added Bright Galaxy Survey (PROVABGS) Mock Challenge. *arXiv e-prints*, art. arXiv:2202.01809, February 2022a.

Hahn, C., Wilson, M. J., Ruiz-Macias, O., Cole, S., Weinberg, D. H., Moustakas, J., Kremin, A., Tinker, J. L., Smith, A., Wechsler, R. H., Ahlen, S., Alam, S., Bailey, S., Brooks, D., Cooper, A. P., Davis, T. M., Dawson, K., Dey, A., Dey, B., Eftekharzadeh, S., Eisenstein, D. J., Fanning, K., Forero-Romero, J. E., Frenk, C. S., Gaztañaga, E., Gontcho, S. G. A., Guy, J., Honscheid, K., Ishak, M., Juneau, S., Kehoe, R., Kisner, T., Lan, T.-W., Landriau, M., Le Guillou, L., Levi, M. E., Magneville, C., Martini, P., Meisner, A., Myers, A. D., Nie, J., Norberg, P., Palanque-Delabrouille, N., Percival, W. J., Poppett, C., Prada, F., Raichoor, A., Ross, A. J., Safonova, S., Saulder, C., Schlafly, E., Schlegel, D., Sierra-Porta, D., Tarle, G., Weaver, B. A., Yèche, C., Zarrouk, P., Zhou, R., Zhou, Z., and Zou, H. DESI Bright Galaxy Survey: Final Target Selection, Design, and Validation. *arXiv e-prints*, art. arXiv:2208.08512, August 2022b. doi: 10.48550/arXiv.2208.08512.

Hearin, A. P., Chaves-Montero, J., Alarcon, A., Becker, M. R., and Benson, A. DSPS: Differentiable Stellar Population Synthesis. *arXiv e-prints*, art. arXiv:2112.06830, December 2021.

Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., Angel, J. R. P., Angeli, G. Z., Ansari, R., Antilogus, P., Araujo, C., Armstrong, R., Arndt, K. T., Astier, P., Aubourg, É., Auza, N., Axelrod, T. S., Bard, D. J., Barr, J. D., Barrau, A., Bartlett, J. G., Bauer, A. E., Bauman, B. J., Baumont, S., Bechtol, E., Bechtol, K., Becker, A. C., Becla, J., Beldica, C., Bellavia, S., Bianco, F. B., Biswas, R., Blanc, G., Blazek, J., Blandford, R. D., Bloom, J. S., Bogart, J., Bond, T. W., Booth, M. T., Borgland, A. W., Borne, K., Bosch, J. F., Boutigny, D., Brackett, C. A., Bradshaw, A., Brandt, W. N., Brown, M. E., Bullock, J. S., Burchat, P., Burke, D. L., Cagnoli, G., Calabrese, D., Callahan, S., Callen, A. L., Carlin, J. L., Carlson, E. L., Chandrasekharan, S., Charles-Emerson, G., Chesley, S., Cheu, E. C., Chiang, H.-F., Chiang, J., Chirino, C., Chow, D., Ciardi, D. R., Claver, C. F., Cohen-Tanugi, J., Cockrum, J. J., Coles, R., Connolly, A. J., Cook, K. H., Cooray, A., Covey, K. R., Cribbs, C., Cui, W., Cutri, R., Daly, P. N., Daniel, S. F., Daruich, F., Daubard, G., Daues, G., Dawson, W., Delgado, F., Dellapenna, A., de Peyster, R., de Val-Borro, M., Digel, S. W., Doherty, P., Dubois, R., Dubois-Felsmann, G. P., Durech, J., Economou, F., Eifler, T., Eracleous, M., Emmons, B. L., Fausti Neto, A., Ferguson, H., Figueroa, E., Fisher-Levine, M., Focke, W., Foss, M. D., Frank, J., Freemon, M. D., Gangler, E., Gawiser, E., Geary, J. C., Gee, P., Geha, M., Gessner, C. J. B., Gibson, R. R., Gilmore, D. K., Glanzman, T., Glick, W., Goldina, T., Goldstein, D. A., Goodenow, I., Graham, M. L., Gressler, W. J., Gris, P., Guy, L. P., Guyonnet, A., Haller, G., Harris, R., Hascall, P. A., Haupt, J., Hernandez, F., Herrmann, S., Hileman, E., Hoblitt, J., Hodgson, J. A., Hogan, C., Howard, J. D., Huang, D., Huffer, M. E., Ingraham, P., Innes, W. R., Jacoby, S. H., Jain, B., Jammes, F., Jee, M. J., Jenness, T., Jernigan, G., Jevremović, D., Johns, K., Johnson, A. S., Johnson, M. W. G., Jones, R. L., Juramy-Gilles, C., Jurić, M., Kalirai, J. S., Kallivayalil, N. J., Kalmbach, B., Kantor, J. P., Karst, P., Kasliwal, M. M., Kelly, H., Kessler, R., Kinnison, V., Kirkby, D., Knox, L., Kotov, I. V., Krabbendam, V. L., Krughoff, K. S., Kubánek, P., Kuczewski, J., Kulkarni, S., Ku, J., Kurita, N. R., Lage, C. S., Lambert, R., Lange, T., Langton, J. B., Le Guillou, L., Levine, D., Liang, M., Lim, K.-T., Lintott, C. J., Long, K. E., Lopez, M., Lotz, P. J.,

Lupton, R. H., Lust, N. B., MacArthur, L. A., Mahabal, A., Mandelbaum, R., Markiewicz, T. W., Marsh, D. S., Marshall, P. J., Marshall, S., May, M., McKercher, R., McQueen, M., Meyers, J., Migliore, M., Miller, M., Mills, D. J., Miraval, C., Moeyens, J., Moolekamp, F. E., Monet, D. G., Moniez, M., Monkewitz, S., Montgomery, C., Morrison, C. B., Mueller, F., Muller, G. P., Muñoz Arancibia, F., Neill, D. R., Newbry, S. P., Nief, J.-Y., Nomerotski, A., Nordby, M., O'Connor, P., Oliver, J., Olivier, S. S., Olsen, K., O'Mullane, W., Ortiz, S., Osier, S., Owen, R. E., Pain, R., Palecek, P. E., Parejko, J. K., Parsons, J. B., Pease, N. M., Peterson, J. M., Peterson, J. R., Petravick, D. L., Libby Petrick, M. E., Petry, C. E., Pierfederici, F., Pietrowicz, S., Pike, R., Pinto, P. A., Plante, R., Plate, S., Plutchak, J. P., Price, P. A., Prouza, M., Radeka, V., Rajagopal, J., Rasmussen, A. P., Regnault, N., Reil, K. A., Reiss, D. J., Reuter, M. A., Ridgway, S. T., Riot, V. J., Ritz, S., Robinson, S., Roby, W., Roodman, A., Rosing, W., Roucelle, C., Rumore, M. R., Russo, S., Saha, A., Sassolas, B., Schalk, T. L., Schellart, P., Schindler, R. H., Schmidt, S., Schneider, D. P., Schneider, M. D., Schoening, W., Schumacher, G., Schwamb, M. E., Sebag, J., Selvy, B., Sembroski, G. H., Seppala, L. G., Serio, A., Serrano, E., Shaw, R. A., Shipsey, I., Sick, J., Silvestri, N., Slater, C. T., Smith, J. A., Smith, R. C., Sobhani, S., Soldahl, C., Storrie-Lombardi, L., Stover, E., Strauss, M. A., Street, R. A., Stubbs, C. W., Sullivan, I. S., Sweeney, D., Swinbank, J. D., Szalay, A., Takacs, P., Tether, S. A., Thaler, J. J., Thayer, J. G., Thomas, S., Thornton, A. J., Thukral, V., Tice, J., Trilling, D. E., Turri, M., Van Berg, R., Vanden Berk, D., Vetter, K., Virieux, F., Vucina, T., Wahl, W., Walkowicz, L., Walsh, B., Walter, C. W., Wang, D. L., Wang, S.-Y., Warner, M., Wiecha, O., Willman, B., Winters, S. E., Wittman, D., Wolff, S. C., Wood-Vasey, W. M., Wu, X., Xin, B., Yoachim, P., and Zhan, H. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *ApJ*, 873(2):111, March 2019. doi: 10.3847/1538-4357/ab042c.

Johnson, B., Foreman-Mackey, D., Sick, J., Leja, J., Byler, N., Walmsley, M., Tollerud, E., Leung, H., and Scott, S. dfm/python-fsps: python-fsps v0.4.1rc1, May 2021. URL https://doi.org/10.5281/zenodo.4737461.

Johnson, B. D., Leja, J., Conroy, C., and Speagle, J. S. Stellar Population Inference with Prospector. *ApJS*, 254 (2):22, June 2021. doi: 10.3847/1538-4365/abef67.

Kingma, D. P. and Ba, J. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, art. arXiv:1412.6980, December 2014. doi: 10.48550/arXiv.1412.6980.

Kobyzev, I., Prince, S. J. D., and Brubaker, M. A. Normalizing Flows: An Introduction and Review of Current Methods. *arXiv e-prints*, art. arXiv:1908.09257, August 2019. doi: 10.48550/arXiv.1908.09257.

Kriek, M. and Conroy, C. The Dust Attenuation Law in Distant Galaxies: Evidence for Variation with Spectral Type. *ApJL*, 775(1):L16, September 2013. doi: 10.1088/2041-8205/775/1/L16.

Leistedt, B., Mortlock, D. J., and Peiris, H. V. Hierarchical Bayesian inference of galaxy redshift distributions from photometric surveys. *MNRAS*, 460(4):4258–4267, August 2016. doi: 10.1093/mnras/stw1304.

Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P., Speagle, J. S., Brammer, G., Momcheva, I., Skelton, R., Whitaker, K. E., Franx, M., and Nelson, E. J. An Older, More Quiescent Universe from Panchromatic SED Fitting of the 3D-HST Survey. *ApJ*, 877(2):140, June 2019. doi: 10.3847/1538-4357/ab1d5a.

Malz, A. I. and Hogg, D. W. How to obtain the redshift distribution from probabilistic redshift estimates. *arXiv e-prints*, art. arXiv:2007.12178, July 2020. doi: 10.48550/arXiv.2007.12178.

Mandelbaum, R. Weak Lensing for Precision Cosmology. *ARA&A*, 56:393–433, September 2018. doi: 10.1146/annurev-astro-081817-051928.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative style, high-performance deep learning library. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance.pdf.

Racca, G. D., Laureijs, R., Stagnaro, L., Salvignol, J.-C., Lorenzo Alvarez, J., Saavedra Criado, G., Gaspar Venancio, L., Short, A., Strada, P., Bönke, T., Colombo, C., Calvi, A., Maiorano, E., Piersanti, O., Prezelus, S., Rosato, P., Pinel, J., Rozemeijer, H., Lesna, V., Musi, P., Sias, M., Anselmi, A., Cazaubiel, V., Vaillon, L., Mellier, Y., Amiaux, J., Berthé, M., Sauvage, M., Azzollini, R., Cropper, M., Pottinger, S., Jahnke, K., Ealet, A., Maciaszek, T., Pasian, F., Zacchei, A., Scaramella, R., Hoar, J., Kohley, R., Vavrek, R., Rudolph, A., and Schmidt, M. The Euclid mission design. In MacEwen, H. A., Fazio, G. G., Lystrup, M., Batalha, N., Siegler, N., and Tong, E. C. (eds.), *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave*, volume 9904

of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp. 99040O, July 2016. doi: 10.1117/12.2230762.

Speagle, J. S., Steinhardt, C. L., Capak, P. L., and Silverman, J. D. A Highly Consistent Framework for the Evolution of the Star-Forming "Main Sequence" from z ~0-6. *ApJS*, 214(2):15, October 2014. doi: 10.1088/0067-0049/214/2/15.

Spergel, D., Gehrels, N., Baltay, C., Bennett, D., Breckinridge, J., Donahue, M., Dressler, A., Gaudi, B. S., Greene, T., Guyon, O., Hirata, C., Kalirai, J., Kasdin, N. J., Macintosh, B., Moos, W., Perlmutter, S., Postman, M., Rauscher, B., Rhodes, J., Wang, Y., Weinberg, D., Benford, D., Hudson, M., Jeong, W. S., Mellier, Y., Traub, W., Yamada, T., Capak, P., Colbert, J., Masters, D., Penny, M., Savransky, D., Stern, D., Zimmerman, N., Barry, R., Bartusek, L., Carpenter, K., Cheng, E., Content, D., Dekens, F., Demers, R., Grady, K., Jackson, C., Kuan, G., Kruk, J., Melton, M., Nemati, B., Parvin, B., Poberezhskiy, I., Peddie, C., Ruffa, J., Wallace, J. K., Whipple, A., Wollack, E., and Zhao, F. Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report. *arXiv e-prints*, art. arXiv:1503.03757, March 2015. doi: 10.48550/arXiv.1503.03757.

Tabak, E. and Vanden-Eijnden, E. Density estimation by dual ascent of the log-likelihood. *Communications in Mathematical Sciences*, 8(1):217–233, 2010. ISSN 1539-6746. doi: 10.4310/CMS.2010.v8.n1.a11.

Taylor, E. N., Hopkins, A. M., Baldry, I. K., Brown, M. J. I., Driver, S. P., Kelvin, L. S., Hill, D. T., Robotham, A. S. G., Bland-Hawthorn, J., Jones, D. H., Sharp, R. G., Thomas, D., Liske, J., Loveday, J., Norberg, P., Peacock, J. A., Bamford, S. P., Brough, S., Colless, M., Cameron, E., Conselice, C. J., Croom, S. M., Frenk, C. S., Gunawardhana, M., Kuijken, K., Nichol, R. C., Parkinson, H. R., Phillipps, S., Pimbblet, K. A., Popescu, C. C., Prescott, M., Sutherland, W. J., Tuffs, R. J., van Kampen, E., and Wijesinghe, D. Galaxy And Mass Assembly (GAMA): stellar mass estimates. *MNRAS*, 418(3):1587–1620, December 2011. doi: 10.1111/j.1365-2966.2011.19536.x.

Tejero-Cantero, A., Boelts, J., Deistler, M., Lueckmann, J.-M., Durkan, C., Gonçalves, P. J., Greenberg, D. S., and Macke, J. H. sbi: A toolkit for simulation-based inference. *Journal of Open Source Software*, 5(52):2505, 2020. doi: 10.21105/joss.02505. URL https://doi.org/10.21105/joss.02505.

Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Nelson, D., and Hernquist, L. Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. *MNRAS*, 444(2):1518–1547, October 2014. doi: 10.1093/mnras/stu1536.

Wang, B., Leja, J., Villar, V. A., and Speagle, J. S. SBI++: Flexible, Ultra-fast Likelihood-free Inference Customized for Astronomical Application. *arXiv e-prints*, art. arXiv:2304.05281, April 2023. doi: 10.48550/arXiv.2304.05281.

Wong, K. W. K., Contardo, G., and Ho, S. Gravitational-wave population inference with deep flow-based generative network. *PhRvD*, 101(12):123005, June 2020. doi: 10.1103/PhysRevD.101.123005.

Wright, A. H., Robotham, A. S. G., Driver, S. P., Alpaslan, M., Andrews, S. K., Baldry, I. K., Bland-Hawthorn, J., Brough, S., Brown, M. J. I., Colless, M., da Cunha, E., Davies, L. J. M., Graham, A. W., Holwerda, B. W., Hopkins, A. M., Kafle, P. R., Kelvin, L. S., Loveday, J., Maddox, S. J., Meyer, M. J., Moffett, A. J., Norberg, P., Phillipps, S., Rowlands, K., Taylor, E. N., Wang, L., and Wilkins, S. M. Galaxy And Mass Assembly (GAMA): the galaxy stellar mass function to z = 0.1 from the r-band selected equatorial regions. *MNRAS*, 470(1):283–302, September 2017. doi: 10.1093/mnras/stx1149.

Zhang, X., Green, G. M., and Rix, H.-W. Parameters of 220 million stars from Gaia BP/RP spectra. *arXiv e-prints*, art. arXiv:2303.03420, March 2023. doi: 10.48550/arXiv.2303.03420.

Table 1. SPS parameters in our model and their priors for training the emulator

Parameter	Description	Prior for training the emulator
z	Redshift	Uniform: min $= 0$, max $= 1.5$
$\log(M_\star/M_\odot)$	Total stellar mass formed	Fixed to $M_\star = 1\ M_\odot$ when training the emulator
$\beta_1, \beta_2, \beta_3, \beta_4$	Coefficients of SFH bases	Flat Dirichlet prior with $0 \leqslant \beta_i \leqslant 1$, $\sum_i \beta_i = 1$
$t_{\rm burst}$	The lookback time when star formation burst happens	Uniform: min $= 10^{-2}$ Gyr, max $= 13.27$ Gyr
$f_{\rm burst}$	The fraction of total formed stellar mass in the star formation burst	Uniform: min $= 0$, max $= 1$
$\log(Z_\star/Z_\odot)$	stellar metallicity ($\log Z_\odot = 0.019$)	Uniform: min $= -2.6$, max $= 0.3$
$n_{\rm dust}$	The power-law index of the Calzetti et al. (2000) attenuation curve	Uniform: min $= -3.0$, max $= 1.0$
τ_1	Birth-cloud dust optical depth	Uniform: min $= 0$, max $= 3.0$
τ_2	Diffuse dust optical depth	Uniform: min $= 0$, max $= 3.0$

Table 2. The distribution of SPS parameters of the mock galaxy population

Parameter	Distribution
z and $\log(M_\star/M_\odot)$	Follow the joint distribution from GAMA DR3
$\beta_1, \beta_2, \beta_3, \beta_4$	Flat Dirichlet distribution with $0 \leqslant \beta_i \leqslant 1$, $\sum_i \beta_i = 1$
$t_{\rm burst}$	Truncated normal: min $= 10^{-2}$ Gyr, max $= 13.27$ Gyr, $\mu = 12$ Gyr, $\sigma = 7$ Gyr
$f_{\rm burst}$	Truncated normal: min $= 0$, max $= 1$, $\mu = 0.1$, $\sigma = 0.7$
$\log(Z_\star/Z_\odot)$	Truncated normal: min $= -2.6$, max $= 0.3$, $\mu = -1.2$, $\mu = 0.9$
$n_{\rm dust}$	Truncated normal: min $= -3.0$, max $= 1.0$, $\mu = 2$, $\sigma = 2$
τ_1	Truncated normal: min $= 0$, max $= 3.0$, $\mu = 1$, $\sigma = 0.8$
τ_2	Truncated normal: min $= 0$, max $= 3.0$, $\mu = 0.6$, $\sigma = 0.8$

A. Details on the SPS model and mock test

Here we describe the SPS model and how we generate the mock galaxy population. The SFH in our SPS model is described by a linear combination of four SFH bases $s_i^{\rm SFH}$ and one burst component:

$$\mathrm{SFH}(t, t_{\rm age}) \propto (1 - f_{\rm burst}) \sum_{i=1}^{4} \beta_i s_i^{\rm SFH}(t) + f_{\rm burst}\, \delta(t - t_{\rm burst}). \tag{1}$$

The β_i are the coefficients of each SFH base, $t_{\rm burst}$ is the lookback time when the star formation burst happens, and $f_{\rm burst}$ is the fraction of total stellar mass that is formed in the burst. As shown in Hahn et al. (2022a), four NMF bases are sufficient to capture the SFH of Illustris galaxies. Among the SFH bases, $s_1^{\rm SFH}$ corresponds to the most recent star formation, whereas $s_4^{\rm SFH}$ corresponds to the oldest star formation (see Fig. 5 in Hahn et al. 2022a). Such SFH bases are non-negative by construction and are physically intuitive to understand. We require $\sum_i \beta_i = 1$ and we normalize overall SFH to be $M_{\star,\rm formed} = \int_{t=0}^{t_{\rm age}} \mathrm{SFH}(t, t_{\rm age})$. In this work, we refer to stellar mass as the total formed stellar mass $M_{\star,\rm formed}$.

In order to synthesize the stellar populations, we discretize the lookback time t into time bins. The first time bin corresponds to $\log(t/\mathrm{yr}) < 6.05$, and each time bin has a width of 0.1 dex until the lookback time reaches the age of the galaxy $t_{\rm age}$, which is determined by its redshift z. For the stellar population in each time bin, we evaluate the corresponding SFR and metallicity according to SFH and generate spectra using `FSPS`. In the end, we add spectra together weighted by the total stellar mass formed in each time bin. We do not add nebular emissions in the SPS model.

We add dust attenuation to the galaxy spectra following Charlot & Fall (2000) recipe, which includes the birth-cloud attenuation and the diffuse dust screens. The birth-cloud attenuation only acts for stars younger than 10^7 yrs (Conroy et al., 2009) and also attenuates nebular emissions. The diffuse dust component affects all stars and we adopt the Kriek & Conroy (2013) formulation. Therefore, the optical depth of dust attenuation is:

$$\hat{\tau}_\lambda(t) = \begin{cases} \tau_1 (\lambda/5500\text{Å})^{-1.0} & t \leqslant 10^7 \text{ yr} \\ \frac{\tau_2}{4.05} (\lambda/5500\text{Å})^{n_{\rm dust}} \left[k'(\lambda) + D(\lambda)\right] & \text{for all ages} \end{cases} \tag{2}$$

where τ_1 is the birth-cloud optical depth at 5500Å, τ_2 is the diffuse dust optical depth at 5500Å, $n_{\rm dust}$ is the slope of dust attenuation curve, $k'(\lambda)$ is the Calzetti et al. (2000) attenuation curve, and $D(\lambda)$ is a Lorentzian-like Drude profile describing

the UV dust bump. In total, our SPS model contains 12 parameters as summarized in Table 1. We refer the interested readers to Hahn et al. (2022a) for a more complete description of PROVABGS. Table 1 also includes the uninformative prior from which we draw the spectra for training the emulator (§2.1).

Section 3.1 shows promising results when using POPSED to infer the galaxy population from mock observation. The mock galaxy population is designed to mimic real galaxy populations. In reality, the distributions of redshift and stellar mass are correlated because of the depth of the survey. In order to introduce such a correlation in our mock test, we take the stellar masses and spectroscopic redshifts from the GAMA DR3 catalog and resample their joint distribution. The distributions of other SPS parameters are summarized in Table 2. Most of them follow a truncated normal distribution whose boundary is set by the prior used to train the emulator (Table 1). For simplicity, we do not introduce correlations among these SPS parameters except for stellar mass and redshift. As shown in Figure 2, such a mock galaxy population has appropriate properties in SFR and metallicity and demonstrates some realistic correlations among SFR, stellar mass, and redshift. It is sufficient for the purpose of a mock test.