



# Accelerated Bayesian SED Modeling Using Amortized Neural Posterior Estimation

ChangHoon Hahn<sup>1</sup> and Peter Melchior<sup>1,2</sup>

<sup>1</sup> Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA; [changhoon.hahn@princeton.edu](mailto:changhoon.hahn@princeton.edu)

<sup>2</sup> Center for Statistics and Machine Learning, Princeton University, Princeton, NJ 08544, USA

Received 2022 March 12; revised 2022 June 16; accepted 2022 June 21; published 2022 October 10

## Abstract

State-of-the-art spectral energy distribution (SED) analyses use a Bayesian framework to infer the physical properties of galaxies from observed photometry or spectra. They require sampling from a high-dimensional space of SED model parameters and take  $>10\text{--}100$  CPU hr per galaxy, which renders them practically infeasible for analyzing the *billions* of galaxies that will be observed by upcoming galaxy surveys (e.g., the Dark Energy Spectroscopic Instrument, the Prime Focus Spectrograph, the Vera C. Rubin Observatory, the James Webb Space Telescope, and the Roman Space Telescope). In this work, we present an alternative scalable approach to rigorous Bayesian inference using Amortized Neural Posterior Estimation (ANPE). ANPE is a simulation-based inference method that employs neural networks to estimate posterior probability distributions over the full range of observations. Once trained, it requires no additional model evaluations to estimate the posterior. We present, and publicly release, SEDFLOW, an ANPE method for producing the posteriors of the recent Hahn et al. SED model from optical photometry and redshift. SEDFLOW takes  $\sim 1$  s per galaxy to obtain the posterior distributions of 12 model parameters, all of which are in excellent agreement with traditional Markov Chain Monte Carlo sampling results. We also apply SEDFLOW to 33,884 galaxies in the NASA–Sloan Atlas and publicly release their posteriors.

*Unified Astronomy Thesaurus concepts:* Galaxies (573); Galaxy photometry (611); Algorithms (1883); Astrostatistics (1882); Astrostatistics techniques (1886); Astrostatistics tools (1887); Bayesian statistics (1900); Posterior distribution (1926); Astrostatistics strategies (1885); Redshift surveys (1378); Galaxy properties (615)

## 1. Introduction

The physical properties of galaxies are the building blocks of our understanding of galaxies and their evolution. We can determine properties such as the stellar mass ( $M_*$ ), star formation rate (SFR), metallicity ( $Z$ ), and age ( $t_{\text{age}}$ ) of a galaxy by analyzing its spectral energy distribution (SED). The theoretical modeling of galaxy SEDs is currently based on stellar population synthesis (SPS) and describes the SED as a composite stellar population constructed from isochrones, stellar spectra, an initial mass function (IMF), a star formation and chemical evolution history, and dust attenuation (e.g., Bruzual & Charlot 2003; Maraston 2005; Conroy et al. 2009; see Walcher et al. 2011 and Conroy 2013 for comprehensive reviews). Some models also include dust and nebular emissions, as well as emissions from active galactic nuclei (e.g., Johnson et al. 2021). In state-of-the-art SED modeling, theoretical SPS models are compared to observed SEDs using Bayesian inference, which accurately quantifies the parameter uncertainties and degeneracies among them (Acquaviva et al. 2011; Chevallard & Charlot 2016; Leja et al. 2017; Carnall et al. 2018; Johnson et al. 2021; Hahn et al. 2022a). The Bayesian approach also enables marginalization over nuisance parameters, which are necessary for modeling the effects of observational systematics (e.g., flux calibration).

However, current Bayesian SED modeling methods, which use Markov Chain Monte Carlo (MCMC) sampling techniques, take 10–100 CPU hr per galaxy (e.g., Carnall et al. 2019; Tacchella et al. 2021). While this is merely very expensive for current data sets of hundreds of thousands of galaxy SEDs, as

observed by the Sloan Digital Sky Survey (SDSS; York et al. 2000), DEEP2 (Davis et al. 2003), COSMOS (Scoville et al. 2007), and GAMA (Baldry et al. 2018), it will be prohibitive for the next generation of surveys. Over the next decade, surveys with the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016), the Prime Focus Spectrograph (PFS; Takada et al. 2014), the Vera C. Rubin Observatory (Ivezic et al. 2019), the James Webb Space Telescope (Gardner et al. 2006), and the Roman Space Telescope (Spergel et al. 2015), will observe *billions* of galaxy SEDs. The task of the SED modeling alone for these surveys would amount to tens or hundreds of billions of CPU hours, exceeding, e.g., the entire computing allocation of the Legacy Survey of Space and Time (LSST) data release production<sup>3</sup> by at least two orders of magnitude. Recently, Alsing et al. have (2020) adopted neural emulators to accelerate SED model evaluations by three to four orders of magnitude—posterior inference takes minutes per galaxy. While this renders current data sets within reach, the next-generation data sets will still require tens or hundreds of millions of CPU hours whenever any aspect of the SED model is altered. Furthermore, this still practically precludes rapid analyses of upcoming transient surveys, especially LSST, which will report  $\sim 10,000$  alerts per minute.<sup>4</sup>

But Bayesian inference does not require MCMC sampling. Simulation-based inference (SBI) is a rapidly developing class of inference methods that offers alternatives for many applications (see Cranmer et al. 2020, and references therein). Many SBI methods leverage the latest developments in statistics and machine learning for more efficient posterior estimation (Papamakarios et al. 2017; Alsing et al. 2019; Hahn



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

<sup>3</sup>  $\approx 2$  billion core hr (<https://dmtn-135.lsst.io/>).

<sup>4</sup> <https://dmtn-102.lsst.io/>

et al. 2019; Jeffrey & Wandelt 2020; Dax et al. 2021; Huppenkothen & Bachetti 2021; Zhang et al. 2021). Of particular interest for SED modeling is a technique called Amortized Neural Posterior Estimation (ANPE). Instead of using MCMC to sample the posteriors for every single galaxy separately, ANPE uses neural density estimators (NDEs) to build a model of the posteriors for *all* observable galaxies. Once the NDE has been trained, generating the posterior requires only the observed SED and no additional model evaluations.

In this work, we present SEDFLOW,<sup>5</sup> a method that applies ANPE to Bayesian galaxy SED modeling using the recent Hahn et al. (2022a) SED model. We demonstrate that we can derive accurate posteriors<sup>6</sup> with SEDFLOW and make Bayesian SED modeling fully scalable for the billions of galaxies that will be observed by upcoming surveys. As further demonstration, we apply SEDFLOW to the optical photometry of  $\sim 33,000$  galaxies in the NASA–Sloan Atlas (NSA). We begin in Section 2 by describing SBI using ANPE. We then present how we design and train SEDFLOW in Section 3, and describe the NSA observations in Section 4. We validate the accuracy of the posteriors from SEDFLOW in Section 5, and we discuss the implications of our results in Section 6.

## 2. Simulation-based Inference

The goal of Bayesian SED modeling, and probabilistic inference more broadly, is to infer the posterior probability distributions  $p(\theta | \mathbf{x})$  of the galaxy properties  $\theta$ , given the observations  $\mathbf{x}$ . For a specific  $\theta$  and  $\mathbf{x}$ , we typically evaluate the posterior using Bayes’s rule,  $p(\theta | \mathbf{x}) \propto p(\theta)p(\mathbf{x} | \theta)$ , where  $p(\theta)$  denotes the prior distribution and  $p(\mathbf{x} | \theta)$  the likelihood, which is typically assumed to have a Gaussian functional form:

$$\ln p(\mathbf{x} | \theta) = -\frac{1}{2}(\mathbf{x} - m(\theta))^T \mathbf{C}^{-1} (\mathbf{x} - m(\theta)). \quad (1)$$

$m(\theta)$  is the theoretical model, in our case a galaxy SED model from SPS.  $\mathbf{C}$  is the covariance matrix of the observations. In practice, off-diagonal terms are often ignored and measured uncertainties are used as estimates of the diagonal terms.

SBI (also known as “likelihood-free” inference) offers an alternative that requires no assumptions about the form of the likelihood. Instead, SBI uses a generative model, i.e., a simulation  $F$ , to generate mock data  $\mathbf{x}'$  given the parameters  $\theta'$ :  $F(\theta') = \mathbf{x}'$ . It uses a large number of simulated pairs  $(\theta', \mathbf{x}')$  to directly estimate the posterior  $p(\theta | \mathbf{x})$ , the likelihood  $p(\mathbf{x} | \theta)$ , or the joint distribution of the parameters and data  $p(\theta, \mathbf{x})$ . SBI has already been successfully applied to a number of Bayesian parameter inference problems in astronomy (e.g., Cameron & Pettitt 2012; Weyant et al. 2013; Hahn et al. 2017; Alsing et al. 2018; Kacprzak et al. 2018; Wong et al. 2020; Huppenkothen & Bachetti 2021; Jeffrey et al. 2021; Zhang et al. 2021) and in physics (e.g., Brehmer et al. 2019; Cranmer et al. 2020).

### 2.1. Amortized Neural Posterior Estimation

SBI provides another a critical advantage over MCMC inference methods—it enables *amortized inference*. For SED modeling using MCMC, each galaxy requires  $> 10^5$  model

evaluations to accurately estimate  $p(\theta | \mathbf{x})$  (Hahn et al. 2022a, J. Kwon et al. 2022, in preparation). Moreover, model evaluations for calculating the posterior of one galaxy cannot be used for another. This makes MCMC approaches to the SED modeling of upcoming surveys computationally infeasible.

With density estimation SBI, we require a large number ( $\sim 10^6$ ) of model evaluations only initially, to train an NDE, a neural network with parameters  $\phi$  that is trained to estimate the density  $p_\phi(\theta | \mathbf{x}')$ . If the training covers the entire or the practically relevant portions of the  $\theta$  and  $\mathbf{x}$  spaces, we can evaluate  $p_\phi(\theta | \mathbf{x}_i)$  for each galaxy  $i$  with minimal computational cost. The inference is therefore amortized and no additional model evaluations are needed to generate the posteriors for each galaxy. This technique is ANPE, and it has recently been applied to a broad range of astronomical applications, from analyzing gravitational waves (e.g., Wong et al. 2020; Dax et al. 2021) to binary microlensing lensing (Zhang et al. 2021). For SED modeling, the choice in favor of using ANPE is easy: the entire upfront cost for ANPE model evaluations would only yield posteriors of tens of galaxies with MCMC.

ANPE makes two important assumptions. First, that the simulator  $F$  is capable of generating mock data  $\mathbf{x}'$  that are practically indistinguishable from the observations. In terms of the expected signal,  $m$  in Equation (1), this is the same requirement as for any probabilistic modeling approach. But unlike likelihood-based evaluations, such as conventional MCMC, the data generated for SBI need to include all relevant noise terms as well. We address both aspects in Sections 3.2 and 6.1. Second, ANPE assumes that the NDE is well trained:  $p_\phi(\theta | \mathbf{x}')$  is a good approximation of  $p(\theta | \mathbf{x}')$ , and therefore of  $p(\theta | \mathbf{x})$ . We assess this in Section 5.

ANPE commonly employs so-called “normalizing flows” (Tabak & Vanden-Eijnden 2010; Tabak & Turner 2013) as density estimators. Normalizing flow models use an invertible bijective transformation,  $f$ , to map a complex target distribution to a simple base distribution,  $\pi(z)$ , that is quick to evaluate. For ANPE, the target distribution is  $p(\theta | \mathbf{x})$  and  $\pi(z)$  is typically a simple multivariate Gaussian, or a mixture of Gaussians. The transformation  $f: z \rightarrow \theta$  must be invertible and have a tractable Jacobian. This is so that we can evaluate the target distribution from  $\pi(z)$  by a change of variable:

$$p(\theta | \mathbf{x}) = \pi(z) \left| \det \left( \frac{\partial f^{-1}}{\partial \theta} \right) \right|. \quad (2)$$

Since the base distribution is easy to evaluate, we can also easily evaluate the target distribution. A neural network is trained to obtain  $f$ , and the collection of its parameters forms  $\phi$ . The network typically consists of a series of simple transforms (e.g., shift and scale transforms) that are each invertible and whose Jacobians are easily calculated. By stringing together many such transforms,  $f$  provides an extremely flexible mapping from the base distribution.

Many different normalizing flow models are now available in the literature (e.g., Germain et al. 2015; Durkan et al. 2019). In this work, we use Masked Autoregressive Flow (MAF; Papamakarios et al. 2017). The autoregressive design (Uria et al. 2016) of MAF is particularly well suited to modeling conditional probability distributions such as the posterior. Autoregressive models exploit the chain rule to expand the joint probabilities of a set of random variables as products of

<sup>5</sup> <https://github.com/changhoonhahn/SEDFlow/>

<sup>6</sup> <https://changhoonhahn.github.io/SEDFlow/>

one-dimensional conditional probabilities:  $p(\mathbf{x}) = \prod_i p(x_i | x_{1:i-1})$ . They then use neural networks to describe each conditional probability,  $p(x_i | x_{1:i-1})$ . In this context, we can add a conditional variable  $y$  on both sides of the equation,  $p(\mathbf{x} | \mathbf{y}) = \prod_i p(x_i | x_{1:i-1}, \mathbf{y})$ , so that the autoregressive model describes a conditional probability  $p(\mathbf{x} | \mathbf{y})$ . One drawback of autoregressive models is their sensitivity to the ordering of the variables. Masked Autoencoder for Distribution Estimation (MADE; Germain et al. 2015) models address this limitation by using binary masks to impose the autoregressive dependence and by permuting the order of the conditioning variables. A MAF model is built by stacking multiple MADE models. Hence, it has the autoregressive structure of MADE, but with greater flexibility to describe complex probability distributions. In practice, we use the MAF implementation in the `sbi` Python package<sup>7</sup> (Greenberg et al. 2019; Tejero-Cantero et al. 2020).

### 3. SEDFLOW

In this section, we present SEDFLOW, which applies ANPE to galaxy SED modeling for a scalable and accelerated approach. For our SED model, we use the state-of-the-art PROVABGS model from Hahn et al. (2022a). Although many SED models have been recently used in the literature—e.g., BAGPIPES (Carnall et al. 2018) and PROSPECTOR (Leja et al. 2017; Johnson et al. 2021)—we choose PROVABGS because it will be used to analyze  $>10$  million galaxy spectro-photometry observations measured by the DESI Bright Galaxy Survey (Hahn et al. 2022b). Below, we describe the PROVABGS model, the construction of the SEDFLOW training data using PROVABGS, and the training procedure for SEDFLOW.

#### 3.1. SED Modeling: PROVABGS

We use the state-of-the-art SPS model of PROVABGS (Hahn et al. 2022a). The SED of a galaxy is modeled as a composite of stellar populations defined by stellar evolution theory (in the form of isochrones, stellar spectral libraries, and an IMF) and its star formation and chemical enrichment histories (SFH and ZH, respectively), attenuated by dust (see Walcher et al. 2011 and Conroy 2013 for a review). In particular, the PROVABGS model utilizes a nonparametric SFH with a starburst, a nonparametric ZH that varies with time, and a flexible dust attenuation prescription.

The SFH has two components: one based on non-negative matrix factorization (NMF) bases and the other being a starburst component. The SFH contribution from the NMF component is a linear combination of four NMF SFH basis functions, derived from performing NMF (Lee & Seung 1999; Cichocki & Phan 2009; Févotte & Idier 2011) on the SFHs of galaxies in the Illustris cosmological hydrodynamical simulation (Genel et al. 2014; Vogelsberger et al. 2014; Nelson et al. 2015). The NMF SFH prescription provides a compact and flexible representation of the SFH. The second starburst component consists of a single stellar population (SSP) and adds stochasticity to the SFH.

The ZH is similarly defined using two NMF bases derived from Illustris. This ZH prescription enables us to flexibly model a wide range of ZHs and, unlike most SED models, it does not assume constant metallicity over time, which can

significantly bias the inferred galaxy properties (Thorne et al. 2021). The stellar evolution theory is based on Flexible Stellar Population Synthesis (FSPS; Conroy et al. 2009; Conroy & Gunn 2010) with the MIST isochrones (Paxton et al. 2011, 2013, 2015; Choi et al. 2016; Dotter 2016), the Chabrier (2003) IMF, and a combination of the MILES (Sánchez-Blázquez et al. 2006) and BaSeL (Lejeune et al. 1997, 1998; Westera et al. 2002) libraries. The SFH and ZH are binned into 43 logarithmically-spaced time bins and SSPs are evaluated at each time bin using FSPS. The SSPs are summed up to get the unattenuated rest-frame galaxy SED.

Lastly, PROVABGS attenuates the light from the composite stellar population using the two-component Charlot & Fall (2000) dust attenuation model with diffuse dust (ISM) and birth cloud (BC) components. All SSPs are attenuated by ISM using the Kriek & Conroy (2013) attenuation curve. The BC component then provides extra dust attenuation on SSPs younger than 100 Myr with young stars that are embedded in molecular clouds and HII regions. In total, the PROVABGS SED model has 12 parameters: stellar mass ( $M_*$ ), six SFH parameters ( $\beta_1, \beta_2, \beta_3, \beta_4, t_{\text{burst}}, f_{\text{burst}}$ ), two ZH parameters ( $\gamma_1, \gamma_2$ ), and three dust attenuation parameters ( $\tau_{\text{BC}}, \tau_{\text{ISM}}, n_{\text{dust}}$ ). Each PROVABGS model evaluation takes  $\sim 340$  ms.

#### 3.2. Training Data

In this section, we describe how we construct the training data for SEDFLOW using the PROVABGS SED model. First, we sample  $N_{\text{train}}$  model parameters from a prior:  $\theta' \sim p(\theta)$ . We use the same priors as Hahn et al. (2022a): uniform priors over  $M_*$ ,  $t_{\text{burst}}$ ,  $f_{\text{burst}}$ ,  $\gamma_1$ ,  $\gamma_2$ ,  $\tau_{\text{BC}}$ ,  $\tau_{\text{ISM}}$ , and  $n_{\text{dust}}$ , with broad conservative ranges, and Dirichlet priors over  $\beta_1, \beta_2, \beta_3$ , and  $\beta_4$ . We choose a symmetric Dirichlet prior for the normalization of the NMF SFH so that  $\sum_i \beta_i = 1$ . For each  $\theta'$ , we also uniformly sample a redshift within the range of the NSA:  $z' \sim \mathcal{U}(0, 0.2)$ . Next, we forward model mock observables. We calculate the rest-frame galaxy SED from PROVABGS and redshift it:  $F(\lambda; \theta', z)$ . Afterward, we convolve  $F$  with optical broadband filters,  $R_X$ , to generate noiseless photometric fluxes:

$$f_X(\theta', z') = \int F(\lambda; \theta', z') R_X(\lambda) d\lambda. \quad (3)$$

The next step of the forward model is to apply noise. We assign photometric uncertainties,  $\sigma'_X$ , by sampling an estimate of the observed  $p(\sigma_X|f_X)$  of NSA galaxies. Then, we apply Gaussian noise,

$$\hat{f}_X(\theta', z', \sigma'_X) = f_X(\theta', z') + n_X \quad \text{where } n_X \sim \mathcal{N}(0, \sigma'_X), \quad (4)$$

to derive the forward-modeled photometric flux.

For our estimate of  $p(\sigma_X|f_X)$ , we use an empirical estimate based on NSA photometry and measured uncertainties. For each of the five optical bands, we separately estimate

$$\hat{p}(\sigma_X|f_X) = \mathcal{N}(\mu_{\sigma_X}(f_X), \sigma_{\sigma_X}(f_X)) \quad (5)$$

as a Gaussian in magnitude space.  $\mu_{\sigma_X}$  and  $\sigma_{\sigma_X}$  are the median and standard deviation of  $\sigma_X$  as a function of  $f_X$ , and we estimate them by evaluating them in  $f_X$  bins and interpolating over the bins. Any  $\theta'$  that is assigned a negative  $\sigma'_X$  is removed from our training data. We also remove any training data with  $f_X(\theta')$  outside the range of NSA photometry.

<sup>7</sup> <https://github.com/mackelab/sbi/>

As SBI requires an accurate noise model, one might be concerned about the simplicity of our model. If the noise model is incorrect, estimates of the parameters  $\theta$  would be biased by an amount that is impossible to predict. We therefore add the noise variances  $\sigma_X$  to the conditioning variables of the posterior model, i.e., we train and evaluate  $p_\phi(\theta | \mathbf{x}, \{\sigma_X\})$ . This means that we merely have to ensure that  $\sigma'_X$  spans the observed  $\sigma_X$  values in order to have a posterior that is robust to our choice of noise model. We discuss this further in Section 6.

In total, we construct  $N_{\text{train}} = 1,131,561$  sets of SED parameters, redshift, photometric uncertainties, and mock NSA photometry. In Figure 1, we present the distribution of the training data  $\{(\theta', z, \sigma'_X, \hat{f}_X)\}$  (black). We include select SED model parameters ( $\log M_*, \beta_1$ ), redshift, photometry in the  $g$  and  $r$  bands, and photometric uncertainty in the  $r$  band. We also include the  $(z, \sigma_X, f_X)$  distribution of NSA galaxies (blue), for comparison. The photometry and uncertainties are in magnitude space. The distribution of the training data spans the distribution of NSA galaxies.

### 3.3. Training SEDFLOW

For SEDFLOW, we use a MAF normalizing flow model (Section 2.1) with 15 MADE blocks, each with two hidden layers and 500 hidden units. In total, the model has 7,890,330 parameters,  $\phi$ . We determine this architecture through experimentation. Our goal is to determine  $\phi$  of the MAF model,  $p_\phi(\theta | \mathbf{x})$ , so that it accurately estimates the posterior probability distribution  $p(\theta | \mathbf{x})$ .  $\theta$  represents the SED parameters and  $\mathbf{x} = (f_X, \sigma_X, z)$ . We do this by minimizing the Kullback–Leibler divergence between  $p_\phi(\theta | \mathbf{x})$  and  $p(\theta | \mathbf{x})$ :  $D_{\text{KL}}(p || p_\phi)$ .

In practice, we split the training data into a training set and a validation set, with a 90/10 split. Afterward, we maximize the total log likelihood  $\sum_i \log p_\phi(\theta_i | \mathbf{x}_i)$  over the training set, which is equivalent to minimizing  $D_{\text{KL}}(p || p_\phi)$ . We use the ADAM optimizer (Kingma & Ba 2017) with a learning rate of  $5 \times 10^{-4}$ . To prevent overfitting, we evaluate the total log likelihood on the validation data at every training epoch and stop the training when the validation log likelihood fails to increase after 20 epochs. Training our model with a batch size of 50 takes roughly a day on a single 2.6 GHz Intel Skylake CPU. Given our small batch size, we find similar training times when using CPUs or GPUs.

## 4. NASA–Sloan Atlas

As a demonstration of its speed and accuracy, we apply SEDFLOW to optical photometry from the NSA.<sup>8</sup> The NSA catalog is a rerduction of SDSS Data Release 8 (Aihara et al. 2011) that includes improved background subtraction (Blanton et al. 2011). We use SDSS photometry in the  $u$ ,  $g$ ,  $r$ ,  $i$ , and  $z$  bands, which are corrected for galactic extinction using Schlegel et al. (1998). To ensure that the galaxy sample is not contaminated, we impose a number of additional quality cuts by excluding:

1. Objects where the centroiding algorithm reports the position of the peak pixel in a given band as the centroid. The SDSS photometric pipeline can struggle to accurately define the center of objects near the edge or at low signal-to-noise ratios, so these cases are often spurious objects.

2. Objects that have pixels that were not checked for peaks by the deblender.
3. Objects where more than 20% of the point-spread function (PSF) flux is interpolated over, as well as objects where the interpolation has affected many pixels and the PSF flux error is inaccurate. The SDSS pipeline interpolates over pixels classified as bad (e.g., cosmic rays).
4. Objects where the interpolated pixels fall within three pixels of the center and they contain a cosmic ray that was interpolated over.
5. Objects that were not detected at  $\geq 5\sigma$  in the original frame, objects that contain saturated pixels, or objects whose their radial profiles could not be extracted.

By excluding these objects, we avoid complications from artifacts in the photometry that we do not model. For additional details of the quality flags, we refer readers to the SDSS documentation.<sup>9</sup> After the quality cuts, we have a total of 33,884 NSA galaxies in our sample.

## 5. Results

Now that we have trained SEDFLOW, we can estimate the posterior,  $p(\theta | \mathbf{x}_i)$ , for any  $\mathbf{x}_i = \{f_{X,i}, \sigma_{X,i}, z_i\}$ . In practice, we do this by drawing samples from the SEDFLOW NDE model. Since we use a normalizing flow, this is trivial: we generate samples from the target distribution of the normalizing flow, a multivariate Gaussian distribution in our case, then we transform the samples using the bijective transformation in Equation (2) that we trained. The transformed samples follow  $p_\phi(\theta | \mathbf{x}_i)$  and estimate the posterior,  $p(\theta | \mathbf{x}_i)$ .

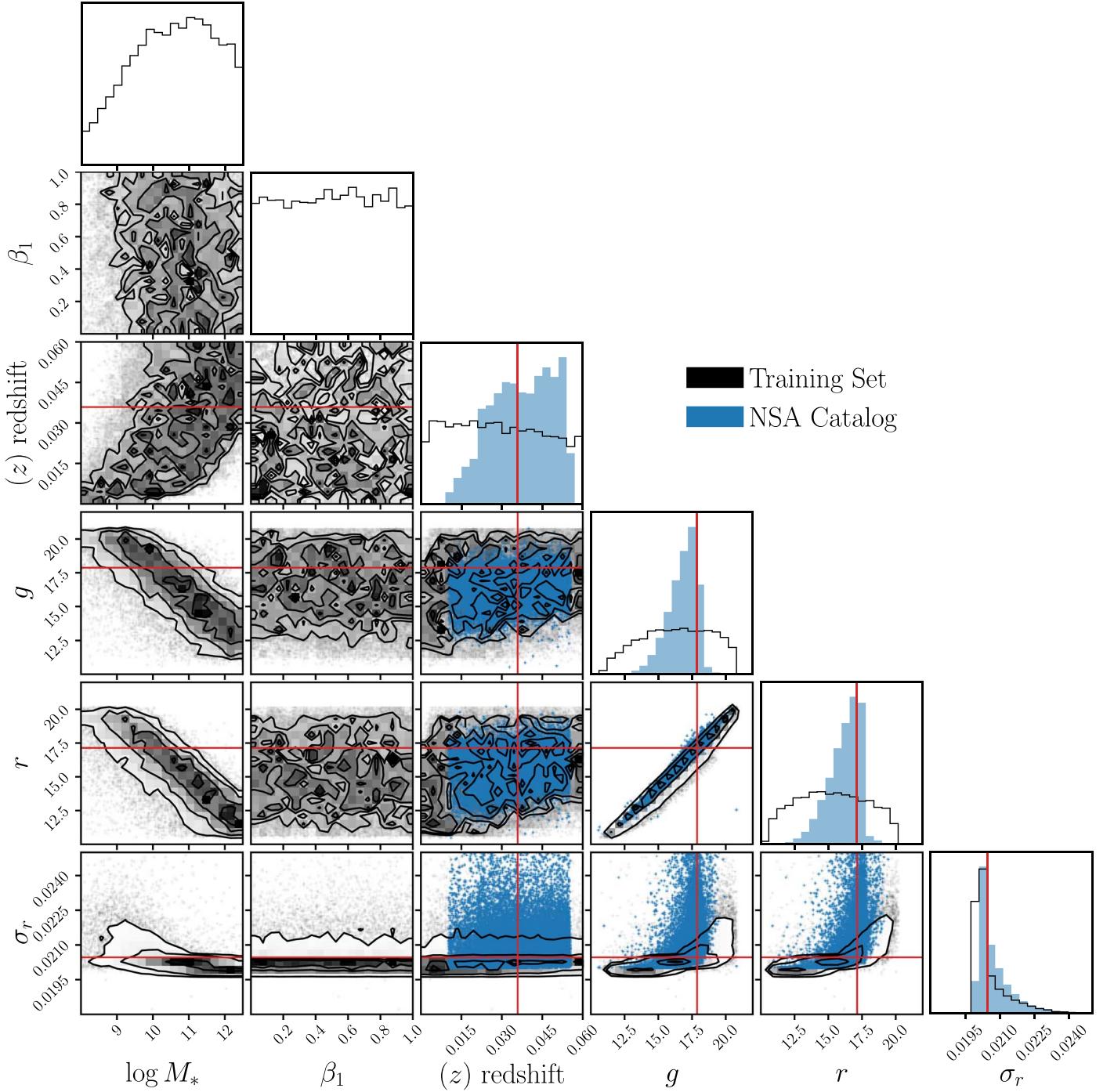
Next, we validate the accuracy of the SEDFLOW posterior estimates. As a first test, we compare the posterior from SEDFLOW to the posterior derived from MCMC sampling for a single arbitrarily chosen NSA galaxy in Figure 2 (NSAID = 26). At the top, we present the posterior distributions of the 12 SED model parameters for the SEDFLOW posteriors (blue) and the MCMC posteriors (black). The SEDFLOW posteriors are in excellent agreement with the MCMC posteriors for all of the SED parameters.

At the bottom of Figure 2, we compare the SEDs of the best-fit parameter values from SEDFLOW (blue lines) and the MCMC posteriors (black dotted lines). We also include the NSA photometric flux of the selected galaxy (red dots). The best-fit SED from the SEDFLOW posterior is in good agreement with both the MCMC best-fit SED and the NSA photometry.

The key advantage of ANPE is that it enables accurate Bayesian inference orders of magnitude faster than conventional methods. We derive the MCMC posterior using the ZEUS ensemble slice sampler (Karamanis & Beutler 2020), with 30 walkers and 10,000 iterations. 2000 of the iterations are discarded for burn-in. In total, the MCMC posterior requires  $>100,000$  SED model evaluations. Since each evaluation takes  $\sim 340$  ms, it takes  $\sim 10$  CPU hr to calculate a single MCMC posterior. Recently, SED modeling has adopted neural emulators to accelerate SED model evaluations (Alsing et al. 2020). In Hahn et al. (2022a), for instance, the PROVABGS emulator takes only  $\sim 2.9$  ms per evaluation,  $>100$  times faster than the original model. Yet, even with emulators, due to the number of evaluations necessary for convergence, an MCMC

<sup>8</sup> <http://nsatlas.org/>

<sup>9</sup> [https://www.sdss.org/dr16/algorithms/flags\\_detail](https://www.sdss.org/dr16/algorithms/flags_detail)



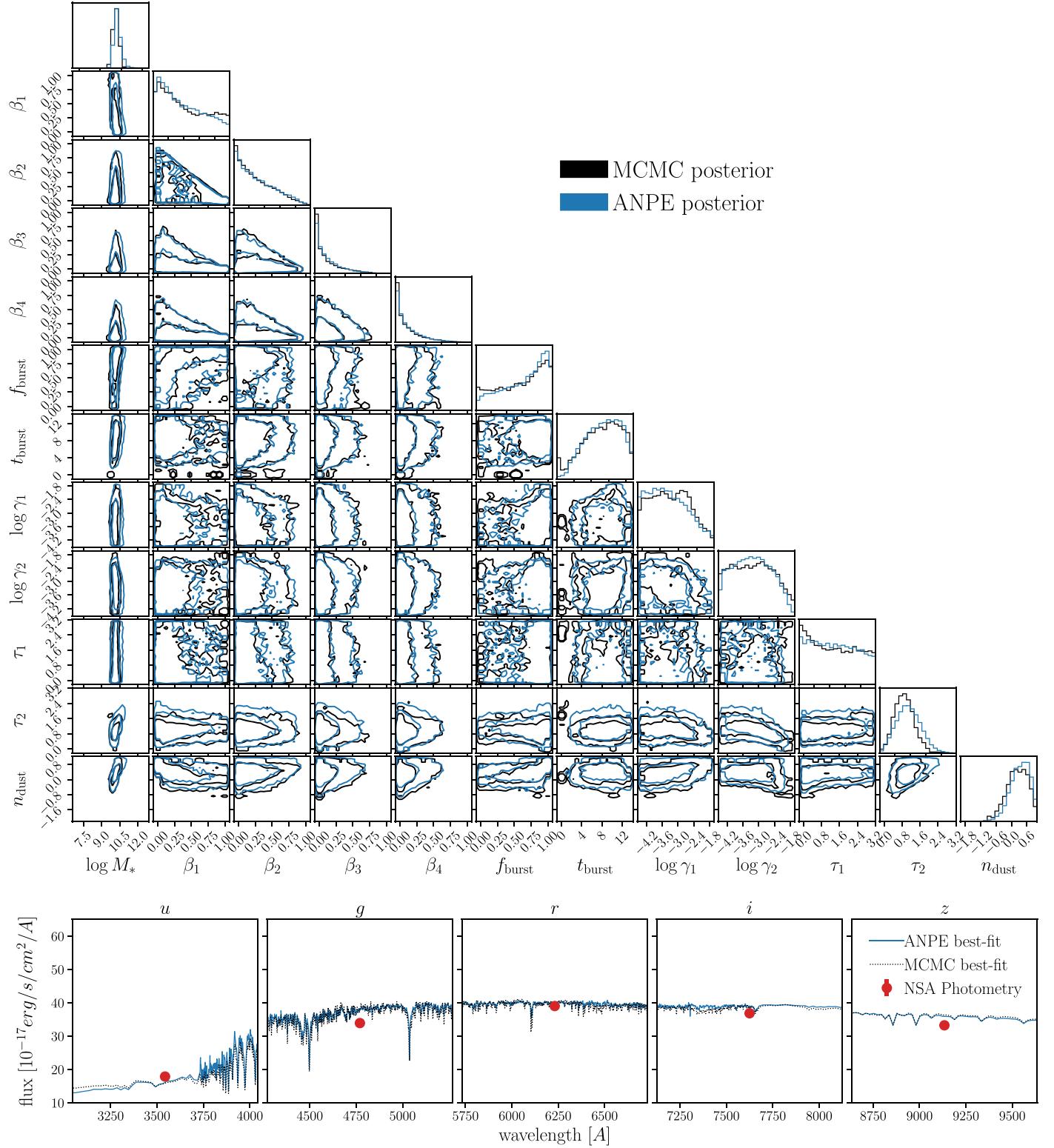
**Figure 1.** The distribution of the SED model parameters, redshift, photometric bands, and uncertainties of the data used to train SEDFLOW. We plot a subset of the parameters ( $\log M_*$ ,  $\beta_1$ ) and photometric bands for clarity. The training data was constructed by sampling SED model parameters from the prior and forward modeling optical photometry using the PROVABGS and noise models (Section 3). For comparison, we present the distribution of the redshift, magnitudes, and uncertainties for galaxies in the NSA catalog (blue). The training set encompasses the observations; thus, SEDFLOW can be used to infer the posteriors for the NSA galaxies.

posterior takes  $\sim 10$  CPU minutes. Meanwhile, after training, the SEDFLOW posterior takes  $1 \text{ s} < 10^4$  times faster than MCMC.

The posteriors from SEDFLOW and MCMC are overall in excellent agreement for the NSA galaxies, besides the one in Figure 2. However, we do not know the true SED parameters for these galaxies, so to further validate SEDFLOW we use test synthetic photometry, where we know the truth. We sample 1000 SED parameters from the prior,  $\{\theta_i^{\text{test}}\} \sim p(\theta)$ , and

forward model synthetic NSA observations,  $\{x_i^{\text{test}}\}$ , for them in the same way as for the training data (Section 3.2). Afterward, we generate posteriors for each  $x_i^{\text{test}}$  using SEDFLOW:  $\{p(\theta | x_i^{\text{test}})\}$ .

In Figure 3, we present the probability–probability (P–P) plot of the SEDFLOW posteriors for the test data. The P–P plot presents the cumulative distribution function (CDF) of the percentile score of the true value within the marginalized posterior for each parameter. For true posteriors, the percentiles

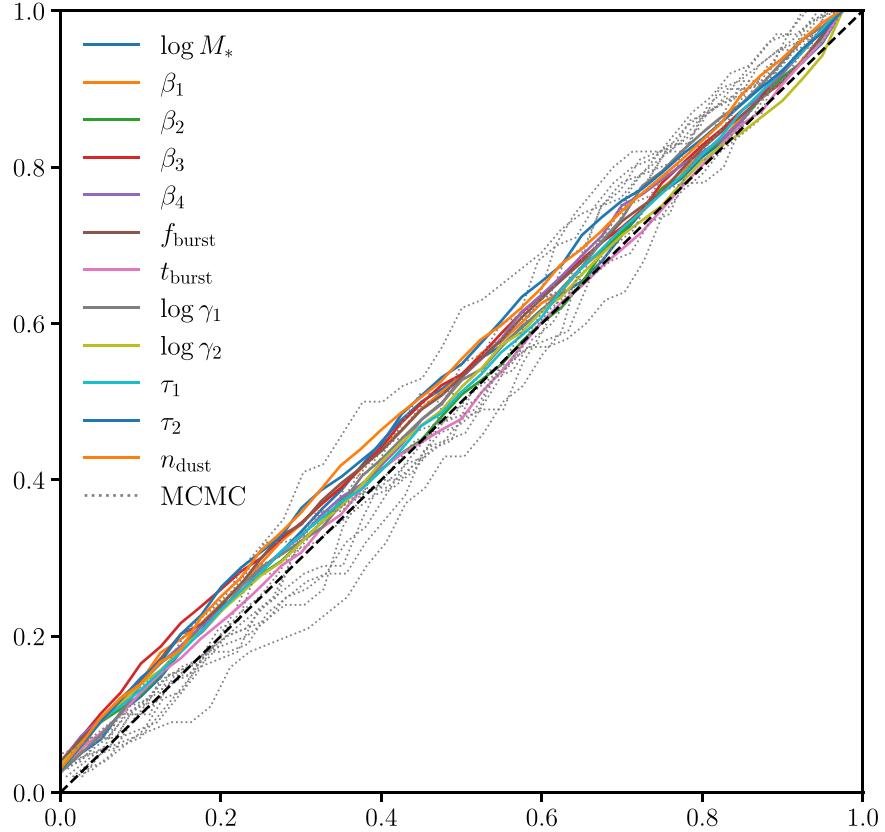


**Figure 2.** A comparison of the posteriors of the 12 SED model parameters derived from standard MCMC sampling (black) and our SEDFLOW (blue) for an arbitrarily selected NSA galaxy (NSAID = 26; red in Figure 1). The posteriors are in excellent agreement for all of the SED parameters. In the bottom panel, we present the SEDs of the best-fit parameter values from SEDFLOW (blue lines) and the MCMC posteriors (black dotted lines), which are both in good agreement with the observed NSA photometric flux (red dots). Estimating the posterior using MCMC sampling requires  $\sim 10$  CPU hr. Even using neural emulators to accelerate the likelihood evaluations, MCMC sampling requires  $\sim 10$  CPU minutes. With SEDFLOW, inferring the full posterior takes 1 s per galaxy.

are uniformly distributed, so the CDF is a diagonal (black dashed line). Overall, the CDFs for SEDFLOW lie close to the diagonal for each of the SED parameters. Hence, the

SEDFLOW posteriors are in excellent agreement with the true posteriors.

In Figure 3, we also include the CDFs of the SED parameters for the MCMC posteriors derived for a subset of 100 test



**Figure 3.** P–P plot of the SEDFLOW posteriors for 1000 synthetic test observations, with the known true parameter values. We plot the CDFs of the percentile scores of the true values within the marginalized SEDFLOW posteriors for each SED parameter. For the true posteriors, the percentile scores are uniformly distributed, so the CDF is diagonal (black dashed line). For reference, we include the P–P plot of the posteriors estimated from the MCMC sampling (gray dotted lines). The SEDFLOW posteriors are in excellent agreement with the true posteriors.

observations (gray dotted lines). Comparing the CDFs from the MCMC posteriors to those of SEDFLOW, we find that the SEDFLOW posteriors are actually in better agreement with the true posteriors. This is due to the fact that the MCMC posteriors are only estimates of the true posterior and are also subject to limitations in initialization, sampling, and convergence. The posteriors from SEDFLOW, which has been properly trained, are not impacted by these limitations, so the comparison highlights the additional advantages of an ANPE approach besides the  $>10^4$  times performance improvement.

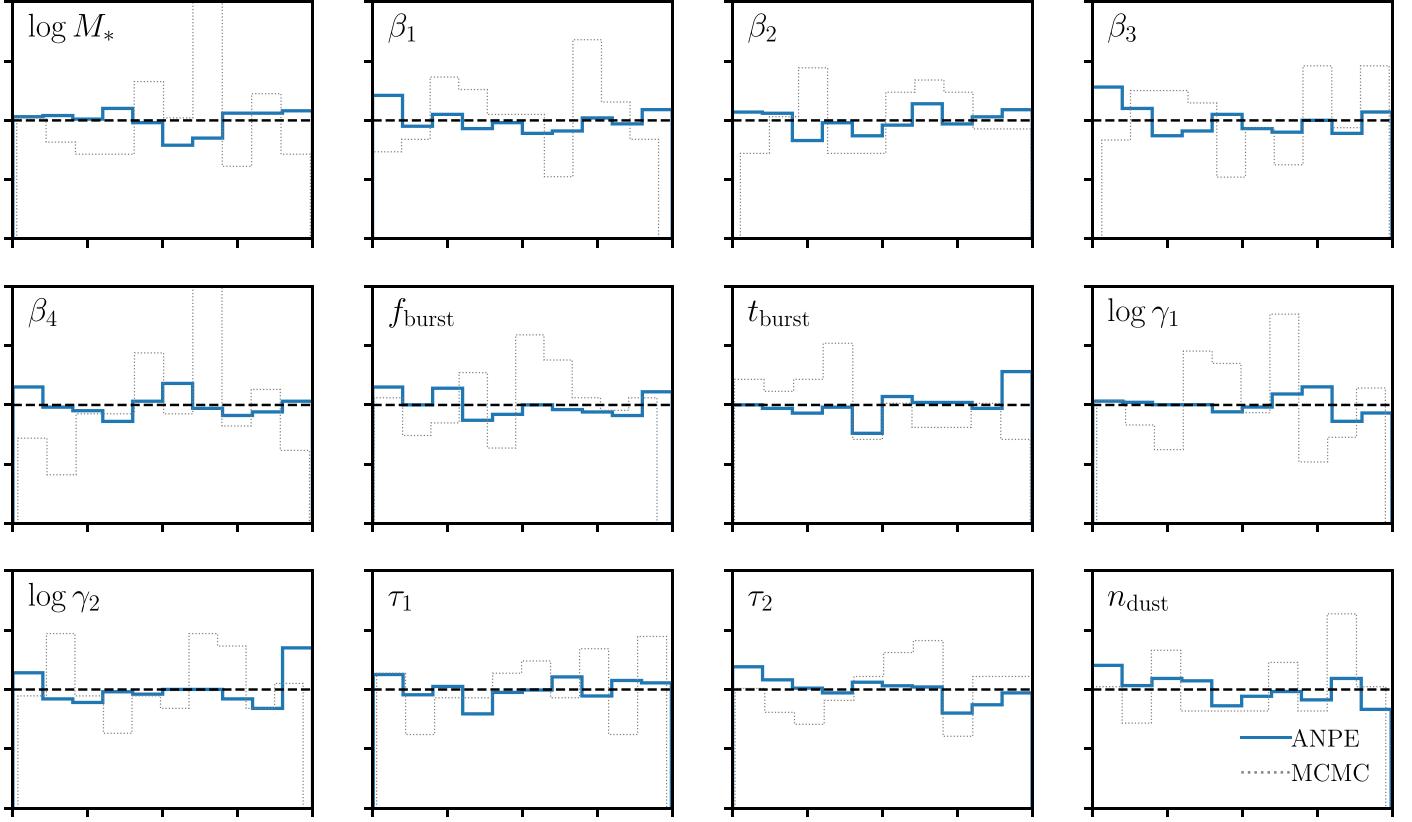
We examine another validation of the SEDFLOW posteriors using simulation-based calibration (SBC; Talts et al. 2020). Rather than using percentile scores, SBC examines the distribution of the rank statistics of the true parameter values within the marginalized posteriors. It addresses the limitations that the CDFs only asymptotically approach the true values and the discrete sampling of the posteriors can cause artifacts in the CDFs. In Figure 4, we present the SBCs of each SED parameter for the SEDFLOW posteriors (blue lines) using the 1000 test observations. For comparison, we include the SBCs for the MCMC posteriors (gray dotted lines). Similar to the percentile scores, the distributions of the rank statistics are uniform for the true posterior (black dashed lines). The rank statistic distributions for the SEDFLOW posteriors are nearly uniform for all SED parameters. Hence, we confirm that the SEDFLOW posteriors are in excellent agreement with the true posteriors.

An advantage of SBC is that by examining the deviation of the rank statistics distribution from uniformity, we can

determine how the posterior estimates deviate from the true posteriors. For instance, if the distribution has a U shape, where the true parameter values are more often at the lowest and highest ranks, then the posterior estimates are narrower than the true posteriors. If the distribution has a  $\cap$  shape, then the posterior estimates are broader than the true posteriors. Any asymmetry in the distribution implies that the posterior estimates are biased. For the SEDFLOW posteriors, we find none of these features for any of the SED parameters. Hence, SEDFLOW provides unbiased and accurate estimates of the true posteriors for all SED parameters.

With the accuracy of SEDFLOW validated, we apply it to derive the posteriors for all of our NSA galaxies (Section 4). Analyzing all 33,884 NSA galaxies takes  $\sim$ 12 CPU hr. For each galaxy, we generate 10,000 samples of the 12-dimensional posterior,  $p(\theta | \mathbf{x})$ . These posteriors on SED parameters represent posteriors on the stellar mass, SFH, ZH, and dust content of the NSA galaxies. To maximize the utility of the posteriors further, we use them to derive posteriors on the following additional galaxy properties: SFR averaged over 1 Gyr, mass-weighted metallicity, and mass-weighted stellar age (see Equation (17) in Hahn et al. 2022a for the exact calculation). We publicly release<sup>10</sup> all of the posteriors for our NSA sample, as well as all of the software and data used to train and validate SEDFLOW (Hahn & Melchior 2022).

<sup>10</sup> <https://changhoonhahn.github.io/SEDflow/>



**Figure 4.** SBC plot of the SEDFLOW posteriors for 1000 synthetic test observations. The histograms in each panel represent the distributions of the rank statistics of the true values within the marginalized SEDFLOW posteriors (blue lines) for each SED parameter. For the true posteriors, the rank statistics will have a uniform distribution (black dashed lines). For reference, we include the rank distributions of the MCMC posteriors for a subset of 100 test data (gray dotted lines). The rank statistic distributions of SEDFLOW are nearly uniform for all of the SED parameters. Therefore, SEDFLOW provides unbiased and accurate estimates of the true posteriors.

## 6. Discussion

### 6.1. Forward Model

In the previous section, we demonstrated the accuracy of the SEDFLOW posteriors. Nevertheless, a primary determining factor for the fidelity of SEDFLOW, or any machine-learning model, is the quality of the training data set and, thus, the forward model used to construct it. Below, we discuss the caveats and limitations of our forward model, which has two components: the PROVABGS SPS model and the noise model (Section 3.2).

First, for our noise model, we assign uncertainties to noiseless photometric fluxes, based on an empirical estimate of  $p(\sigma_x | f_x)$ , for each band independently. This is a simplistic prescription and, as the bottom right panels of Figure 1 ( $g - \sigma_r$  and  $r - \sigma_r$ ) reveal, there are discrepancies between the magnitude—uncertainty distributions of the training data and the observations. Despite these discrepancies, SEDFLOW provides excellent estimates of the true posterior. This is because we design our ANPE to include  $\sigma_x$  as a conditional variable (Section 3.3). The  $f_x - \sigma_x$  distribution of our training data does not impact the accuracy of the posteriors, as long as there are sufficient training data near  $x$  to train the NDE in that region.

A more accurate noise model will, in theory, improve the performance of SEDFLOW, because the  $x$ -space of the training data will more efficiently span the observations. Fewer training data would be expended in regions of  $x$ -space that are devoid of observations. However, for our application, we do not find

significant improvement in how well we estimate the true posterior when we alter the noise model. For example, we used a noise model that assigns uncertainties to fluxes based on  $p(\sigma_{u,g,r,i,z} | f_{u,g,r,i,z})$ , estimated using a normalizing flow. This suggests that even with our simplistic noise model, the  $x$ -space of the observations is covered sufficiently well by the training data. We note that when we decrease  $N_{\text{train}}$  to below 500,000, the SEDFLOW posteriors are significantly less accurate. A more realistic forward model may reduce this  $N_{\text{train}}$  threshold for accurate posteriors. However, generating  $N_{\text{train}} \sim 1,000,000$  training SEDs has a negligible computational cost compared to MCMC SED modeling, so we do not consider it necessary to explore this further.

Next, we consider the limitations in the PROVABGS SPS model used in our forward model. Our SPS model uses compact and flexible prescriptions for SFH and ZH that can describe a broad range of SFHs and ZHs. However, the prescriptions are derived from simulated Illustris galaxies, whose SFHs and ZHs may not reflect the full range of SFHs and ZHs of real galaxies. If certain subpopulations of observed galaxies have SFHs and ZHs that cannot be well described by the PROVABGS prescription, they cannot be accurately modeled. Even if the PROVABGS SFH and ZH prescriptions are sufficient, there are limitations in our understanding of stellar evolution.

There is currently no consensus on stellar evolution, stellar spectral libraries, or the IMFs of galaxies (e.g., Treu et al. 2010; van Dokkum & Conroy 2010; Rosani et al. 2018; Ge et al. 2019; Sonnenfeld et al. 2019). The PROVABGS model uses

MIST isochrones, the Chabrier (2003) IMF, and MILES + BaSeL spectral libraries. These choices limit the range of SEDs that can be produced by the training data. For instance, if galaxies have significant variations in their IMFs, assuming a fixed IMF would falsely limit the range of our training data. A more flexible SED model that includes uncertainties in SPS would broaden the range of galaxy SEDs that can be modeled. Data-driven approaches may also enable SED models to be more descriptive (e.g., Hogg et al. 2016; Portillo et al. 2020). Improving SED models, however, is beyond the scope of this work. Our focus is on improving the Bayesian inference framework. In this regard, the limitations of the SED model equally impact conventional approaches with MCMC.

We encounter the caveats above when we apply SEDFLOW to the NSA catalog. For a small fraction of NSA galaxies (588 out of 33,884), SEDFLOW generates posteriors that are outside the prior volume. This is because the photometry or the uncertainties of these galaxies lie outside the support of the training data, where SEDFLOW is well trained. They either have higher photometric uncertainties, for a given magnitude, or bluer photometric colors than the training data. Some of these may be observational artifacts or problematic or biased photometry. Nevertheless, SEDFLOW fails in these cases because we cannot construct training data near these galaxies with our limited noise and SPS models. Since this only affects a small fraction of the NSA galaxies, we flag them in our catalog and, for completeness, infer their galaxy properties by applying PROVABGS with MCMC sampling. For more details, we refer readers to Appendix.

To test the limitations of the forward model, we could construct additional tests for the posteriors derived from ANPE. For instance, the  $\chi^2$  of the best-fit parameter value from the estimated posterior could be used to assess whether the best-fit model accurately reproduces the observations. This would only require one additional model evaluation per galaxy. One could also construct an Amortized Neural Likelihood Estimator (ANLE) using the same training data. Unlike the ANPE, which estimates  $p(\theta | f_X, \sigma_X, z)$ , the ANLE would estimate  $p(f_X | \theta, \sigma_X, z)$ . We could then further validate the posteriors by assessing whether the observed photometry lies within the ANLE distribution. Based on the overall high level of accuracy of the SEDFLOW posteriors, we do not explore these additional tests; however, they could be used to further validate any ANPE posteriors.

## 6.2. Advantages of SEDFLOW

The primary advantage of SEDFLOW is its computational speed. This becomes even more pertinent if we want to add additional parameters to address concerns about the choices in current SPS models, described above. To relax these assumptions, SPS models would need to introduce additional parameters that flexibly model these uncertainties (Conroy et al. 2009; Conroy & Gunn 2010). While the dimensionality of the current SPS models already makes MCMC methods computationally infeasible, ANPE has been applied to higher-dimensional applications. For instance, Dax et al. (2021) constructed an accurate ANPE for a 15-dimensional model parameter and 128-dimensional conditional variable spaces. NDE is an actively developing field in machine learning and new methods are constantly emerging (e.g., Wu et al. 2020; Dhariwal & Nichol 2021). Since ANPE can handle higher dimensionality, in the future we will be able to include

additional parameters that model the uncertainties in SPS. This will not only improve our SED modeling, but also improve our understanding of stellar evolution and the IMF.

In addition to enabling scalable SED modeling for the next generation of galaxy surveys, SEDFLOW will also enable us to tackle other key challenges in SED modeling. For example, recent works have demonstrated that the priors of SED models can significantly impact the inferred galaxy properties (Carnall et al. 2018; Leja et al. 2019; Hahn et al. 2022a). Even “uninformative” uniform priors on SED model parameters can impose undesirable priors on derived galaxy properties, such as  $M_*$ , SFR, SFH, and ZH. To avoid significant biases, galaxy studies must carefully select priors and validate their results using multiple different choices. With an MCMC approach, selecting a different prior means reevaluating every posterior and repeating all the SED model evaluations in the MCMC sampling. For an ANPE approach, the prior is set by the distribution of the parameters in the training data. For a new prior, instead of reconstructing the training data, we can resample it in such a way that the parameters follow the new prior. The ANPE model can then be retrained, revalidated on the test data, and redeployed on the observations. Each of these steps will require substantially less computational resources than generating a new set of training data or using MCMC methods. Hence, the ANPE approach provides a way of efficiently varying the prior without multiplying the computational costs.

## 7. Summary and Outlook

By analyzing the SED of a galaxy, we can infer detailed physical properties, such as its stellar mass, SFR, metallicity, and dust content. These properties serve as the building blocks of our understanding of how galaxies form and evolve. State-of-the-art SED modeling methods use MCMC sampling to perform Bayesian statistical inference. They derive posterior probability distributions of galaxy properties, given observations that accurately estimate the uncertainties and parameter degeneracies, to enable more rigorous statistical analyses. For the dimensionalities of current SED models, deriving a posterior requires  $\gtrsim 100,000$  model evaluations and takes  $\gtrsim 10\text{--}100$  CPU hr per galaxy. Upcoming galaxy surveys, however, will observe *billions* of galaxies, e.g., DESI, PFS, the Vera C. Rubin Observatory, the James Webb Space Telescope, and the Roman Space Telescope. Analyzing all of these galaxies with the current Bayesian SED models would be infeasible and would require hundreds of billions of CPU hours. Even with the recently proposed emulators, which accelerate model evaluations by three to four orders of magnitude, the computation costs of the SED modeling would remain a major bottleneck for galaxy studies.

We demonstrate in this work that ANPE provides an alternative *scalable* approach for Bayesian inference in SED modeling. ANPE is an SBI method that formulates Bayesian inference as a density estimation problem and uses NDEs to approximate the posteriors over the full space of the observations. The NDE is trained using parameter values drawn from the prior and mock observations simulated with these parameters. Once trained, a posterior can be obtained from the NDE by providing the observations as the conditional variables, without any additional model evaluations.

In this work, we present SEDFLOW, a galaxy SED modeling method using ANPE and PROVABGS, a flexible SED model

that uses compact nonparametric SFH and ZH prescriptions and was recently validated in Hahn et al. (2022a). Furthermore, we apply SEDFLOW to optical photometry from the NSA as demonstration and validation of our ANPE approach. We present the following key results from our analysis.

1. We train SEDFLOW using a data set of  $\sim 1$  million SED model parameters and forward-modeled synthetic SEDs. The parameters are drawn from a prior and the forward model is based on the PROVABGS and noise models. We design the ANPE to estimate  $p(\theta | f_X, \sigma_X, z)$ , where  $f_X$ ,  $\sigma_X$ , and  $z$  are the photometry, photometric uncertainty, and redshift, respectively. For its architecture, we use a MAF normalizing flow with 15 MADE blocks, each with two hidden layers and 500 hidden units. Training SEDFLOW requires roughly 1 day on a single CPU. Once trained, deriving the posteriors of galaxy properties for a galaxy takes  $\sim 1$  s,  $10^5$  times faster than traditional MCMC sampling.
2. The posteriors derived using SEDFLOW show excellent agreement with the posteriors derived from MCMC sampling. We further validate the accuracy of the posteriors by applying SEDFLOW to synthetic observations with known true parameter values. Based on statistical metrics used in the literature (a P-P plot and an SBC plot), we find excellent agreement between the SEDFLOW posteriors and the true posteriors.
3. Lastly, we demonstrate the advantages of SEDFLOW by applying it to the NSA. Estimating the posterior of  $\sim 33,000$  galaxies takes  $\sim 12$  CPU hr. We make the catalog of posteriors publicly available online. For each galaxy, the catalog contains the posteriors of all 12 PROVABGS SED model parameters, as well as the galaxy properties:  $M_*$ , average SFR over 1 Gyr, mass-weighted metallicity, and mass-weighted stellar age.

This work highlights the advantages of using an ANPE approach to Bayesian SED modeling. Our approach can easily be extended beyond this application. For instance, we can include multiwavelength photometry at ultraviolet or infrared wavelengths. We can also modify SEDFLOW to infer redshift from photometry. In SEDFLOW, we include redshift as a conditional variable, since NSA provides spectroscopic redshifts. However, redshift can be included as an inferred variable, rather than a conditional one. We can then apply SEDFLOW to infer galaxy properties from photometric data sets without redshift measurements, while marginalizing over the redshift prior. If we do not require spectroscopic redshifts, SEDFLOW can be extended to much larger data sets that span fainter and broader galaxy samples. Conversely, we can use SEDFLOW to infer more physically motivated photometric redshifts, where we marginalize over our understanding of galaxies rather than using templates.

The ANPE approach to SED modeling can also be extended to galaxy spectra. Constructing an ANPE for the full data space of the spectra would require estimating a dramatically higher-dimensional probability distribution. SDSS spectra, for instance, have  $\sim 3600$  spectral elements. In our approach, we include the uncertainties of observables as conditional variables, which doubles the curse of dimensionality. Recent works, however, have demonstrated that galaxy spectra can be represented in a compact low-dimensional space using autoencoders (Portillo et al. 2020, P. Melchior & C. Hahn

2022, in preparation). In Portillo et al. (2020), they demonstrate that SDSS galaxy spectra can be compressed into seven-dimensional latent variable space, with little loss of information. Such spectral compression dramatically reduces the dimensionality of the conditional variable space to dimensions that can be tackled by current ANPE methods. We will explore the SED modeling of galaxy spectrophotometry using ANPE and spectral compression in a following work.

It is a pleasure to thank Adam Carnall, Miles Cranmer, Kartheik Iyer, Andy Goulding, Jenny E. Green, Joel Leja, Jiaxuan Li, Uroš Seljak, and Michael A. Strauss for valuable discussions and comments. This work was supported by the AI Accelerator program of the Schmidt Futures Foundation.

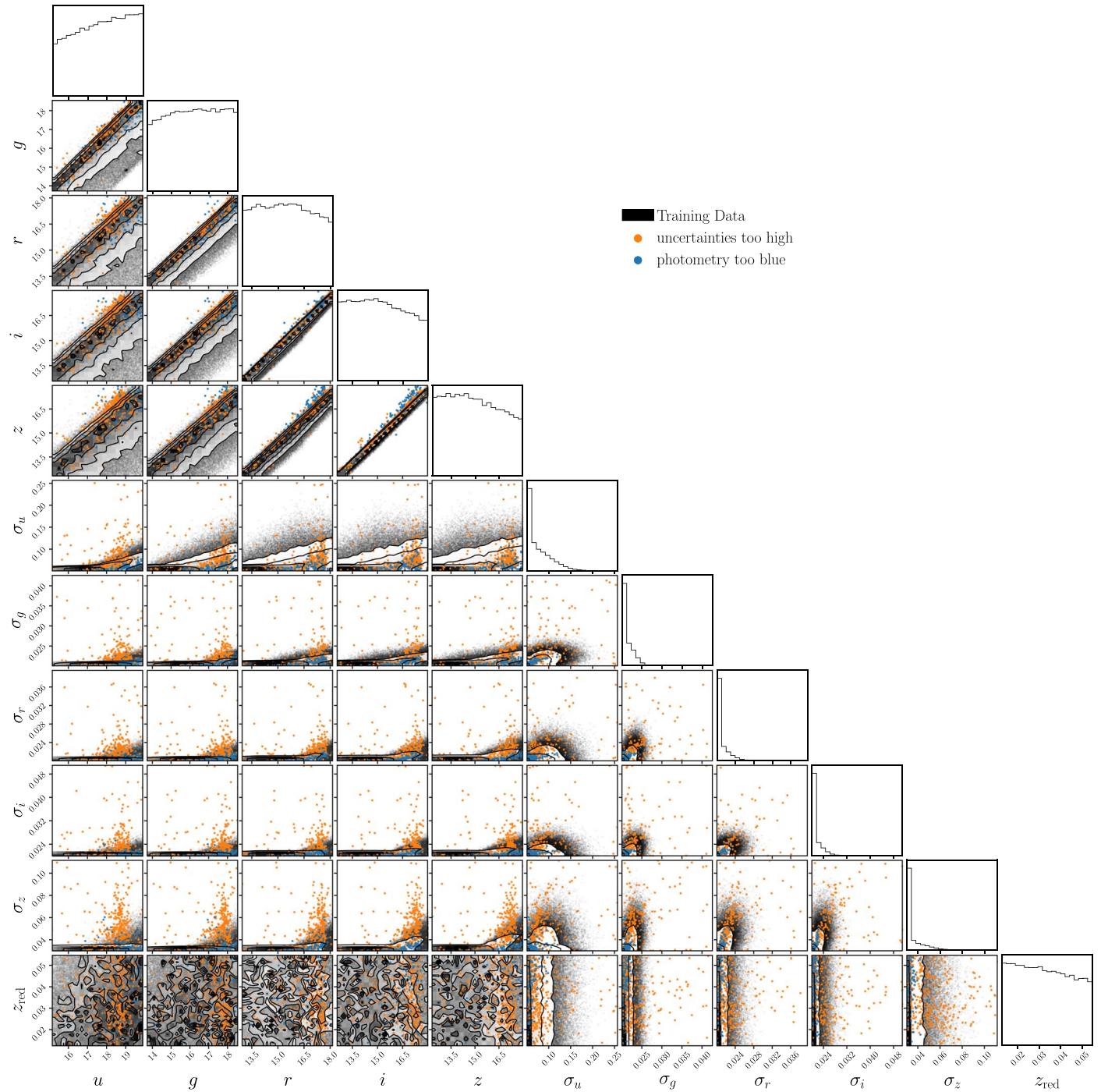
## Appendix

### Testing outside the SEDFLOW Training Range

For a small number of NSA galaxies, 588 of 33,884, SEDFLOW does not produce valid posteriors. The normalizing flow of SEDFLOW generates posteriors that are entirely outside the prior volume, because the photometry or uncertainties of the galaxies lie outside of the support of the SEDFLOW training data (Section 3.2). These galaxies either (1) have unusually high photometric uncertainties that are not accounted for in our noise model, or (2) they have photometric colors that cannot be modeled by our SED model. In Figure 5, we present the distributions of the photometric magnitudes, uncertainties, and redshifts ( $\mathbf{x} = \{f_X, \sigma_X, z\}$ ) of these NSA galaxies. We mark the galaxies that are outside the SEDFLOW training data support (black) due to (1) in orange and (2) in blue.

We classify the galaxies as (1) if they have  $\sigma_X$  that is unusually high for a given  $f_X$  in at least one photometric band:  $\sigma_X > \mu_{\sigma_X}(f_X) + 3\sigma_{\sigma_X}(f_X)$  (see Equation (4)). There are 490 galaxies without valid SEDFLOW posteriors due to (1). Many of them lie well beyond the locus of the training data points. The SEDFLOW estimate of  $p(\theta | f_X, \sigma_X, z)$  is only accurate in regions of  $\mathbf{x}$ -space where there are sufficient training data. This requirement is not met for these galaxies. In principle, if we use a more conservative noise model than Equation (4), and construct noisier training data, we can expand the support of SEDFLOW. SEDFLOW would then produce sensible posteriors for more NSA galaxies. However, there is an inherent trade-off. For a training data set of fixed size, a more conservative noise model would reduce the amount of training data in the  $\mathbf{x}$ -space where the vast majority of NSA galaxies lie, so it could reduce the accuracy of the posteriors in these regions. Since SEDFLOW fails for only a small fraction of NSA galaxies, we do not explore more conservative noise models in this work.

Next, we examine the galaxies that lie outside the SEDFLOW support because they have colors that cannot be modeled by our SED model. In Figure 5, we classify NSA galaxies as (2) if any of their colors (e.g.,  $u - r$ ,  $u - g$ , or  $r - z$ ) are bluer than the 99.9% percentile of the training data color. There are 98 galaxies without valid SEDFLOW posteriors due to (2). In particular, the  $i$  versus  $z$  magnitude panel highlights how a significant number of galaxies are bluer than the training data. The fact that the training data do not span these colors suggests that the PROVABGS SED model may not fully describe all types of galaxies in observations. As we discuss in Section 6, this may be due to the limitations of the SFH and ZH prescriptions in the PROVABGS model or our understanding of stellar evolution. The limitations of the SED model equally



**Figure 5.** The distributions of the photometric magnitudes, uncertainties, and redshifts of the NSA galaxies for which SEDFLOW does not produce valid posteriors (blue or orange). These galaxies lie outside the support of the SEDFLOW training data (black), so SEDFLOW cannot accurately estimate their posteriors. We mark the NSA galaxies that have unusually high photometric uncertainties that are not accounted for in our noise model in orange and the galaxies that have photometric colors that cannot be modeled by our SED model in blue.

impact conventional MCMC sampling approaches, and are beyond the scope of this work. Therefore, we do not examine this issue further. For completeness, we derive the posteriors for the NSA galaxies on which SEDFLOW fails, using PROVABGS with MCMC sampling with the same configuration as Hahn et al. (2022a).

#### ORCID iDs

ChangHoon Hahn <https://orcid.org/0000-0003-1197-0902>

Peter Melchior <https://orcid.org/0000-0002-8873-5065>

#### References

- Acquaviva, V., Gawiser, E., & Guaita, L. 2011, *ApJ*, **737**, 47
- Aihara, H., Allende Prieto, C., An, D., et al. 2011, *ApJS*, **193**, 29
- Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. 2019, *MNRAS*, **488**, 4440
- Alsing, J., Wandelt, B., & Feeney, S. 2018, arXiv:1808.06040
- Alsing, J., Peiris, H., Leja, J., et al. 2020, *ApJS*, **249**, 5
- Baldry, I. K., Liske, J., Brown, M. J. I., et al. 2018, *MNRAS*, **474**, 3875

- Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011, *AJ*, **142**, 31
- Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. 2019, *PNAS*, **117**, 5242
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000
- Cameron, E., & Pettitt, A. N. 2012, *MNRAS*, **425**, 44
- Carnall, A. C., Leja, J., Johnson, B. D., et al. 2019, *ApJ*, **873**, 44
- Carnall, A. C., McLure, R. J., Dunlop, J. S., & Davé, R. 2018, *MNRAS*, **480**, 4379
- Chabrier, G. 2003, *PASP*, **115**, 763
- Charlot, S., & Fall, S. M. 2000, *ApJ*, **539**, 718
- Chevallard, J., & Charlot, S. 2016, *MNRAS*, **462**, 1415
- Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, **823**, 102
- Cichocki, A., & Phan, A.-H. 2009, *IEETF*, **92**, 708
- Conroy, C. 2013, *ARA&A*, **51**, 393
- Conroy, C., & Gunn, J. E. 2010, *ApJ*, **712**, 833
- Conroy, C., Gunn, J. E., & White, M. 2009, *ApJ*, **699**, 486
- Cranmer, K., Brehmer, J., & Louppe, G. 2020, *PNAS*, **117**, 30055
- Davis, M., Faber, S. M., Newman, J., et al. 2003, *Proc. SPIE*, **4834**, 161
- Dax, M., Green, S. R., Gair, J., et al. 2021, *PhRvL*, **127**, 241103
- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
- Dhariwal, P., & Nichol, A. 2021, arXiv:2105.05233
- Dotter, A. 2016, *ApJS*, **222**, 8
- Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. 2019, arXiv:1906.04032
- Férotte, C., & Idier, J. 2011, arXiv:1010.1763
- Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, *SSRv*, **123**, 485
- Ge, J., Mao, S., Lu, Y., Cappellari, M., & Yan, R. 2019, *MNRAS*, **485**, 1675
- Genel, S., Vogelsberger, M., Springel, V., et al. 2014, *MNRAS*, **445**, 175
- Germain, M., Gregor, K., Murray, I., & Laroche, H. 2015, Proc. of the 32nd Int. Conf. on Machine Learning (Lille, France: PMLR), 881, arXiv:1502.03509
- Greenberg, D. S., Nonnenmacher, M., & Macke, J. H. 2019, arXiv:1905.07488
- Hahn, C., Beutler, F., Sinha, M., et al. 2019, *MNRAS*, **485**, 2956
- Hahn, C., & Melchior, P. 2022, SEDflow: Accelerated Bayesian SED Modeling using Amortized Neural Posterior Estimation, v0.1, Zenodo, doi:10.5281/zenodo.6345467
- Hahn, C., Vakili, M., Walsh, K., et al. 2017, *MNRAS*, **469**, 2791
- Hahn, C., Kwon, K. J., Tojeiro, R., et al. 2022a, arXiv:2202.01809
- Hahn, C., Wilson, M. J., & Ruiz-Macias, O. 2022b, arXiv:2208.08512
- Hogg, D. W., Casey, A. R., Ness, M., et al. 2016, *ApJ*, **833**, 13
- Huppenkothen, D., & Bachetti, M. 2021, *MNRAS*, **511**, 5689
- Ivezic, Z., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, **873**, 111
- Jeffrey, N., Alsing, J., & Lanusse, F. 2021, *MNRAS*, **501**, 954
- Jeffrey, N., & Wandelt, B. D. 2020, arXiv:2011.05991
- Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, *ApJS*, **254**, 22
- Kacprzak, T., Herbel, J., Amara, A., & Réfrégier, A. 2018, *JCAP*, **2018**, 042
- Karamanis, M., & Beutler, F. 2020, arXiv:2002.06212
- Kingma, D. P., & Ba, J. 2017, arXiv:1412.6980
- Kriek, M., & Conroy, C. 2013, *ApJL*, **775**, L16
- Lee, D. D., & Seung, H. S. 1999, *Natur*, **401**, 788
- Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2019, *ApJ*, **876**, 3
- Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., & Byler, N. 2017, *ApJ*, **837**, 170
- Lejeune, T., Cuisinier, F., & Buser, R. 1997, *A&AS*, **125**, 229
- Lejeune, T., Cuisinier, F., & Buser, R. 1998, *A&AS*, **130**, 65
- Maraston, C. 2005, *MNRAS*, **362**, 799
- Nelson, D., Pillepich, A., Genel, S., et al. 2015, *A&C*, **13**, 12
- Papamakarios, G., Pavlakou, T., & Murray, I. 2017, arXiv:1705.07057
- Paxton, B., Bildsten, L., Dotter, A., et al. 2011, *ApJS*, **192**, 3
- Paxton, B., Cantiello, M., Arras, P., et al. 2013, *ApJS*, **208**, 4
- Paxton, B., Marchant, P., Schwab, J., et al. 2015, *ApJS*, **220**, 15
- Portillo, S. K. N., Parejko, J. K., Vergara, J. R., & Connolly, A. J. 2020, *AJ*, **160**, 45
- Rosani, G., Pasquali, A., La Barbera, F., Ferreras, I., & Vazdekis, A. 2018, *MNRAS*, **476**, 5233
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, **500**, 525
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, **172**, 1
- Sonnenfeld, A., Jaelani, A. T., Chan, J., et al. 2019, *A&A*, **630**, A71
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757
- Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, *MNRAS*, **371**, 703
- Tabak, E. G., & Turner, C. V. 2013, *Commun. Pure Appl. Math.*, **66**, 145
- Tabak, E. G., & Vanden-Eijnden, E. 2010, *Commun. Math. Sci.*, **8**, 217
- Tacchella, S., Conroy, C., Faber, S. M., et al. 2021, *ApJ*, **926**, 36
- Takada, M., Ellis, R. S., Chiba, M., et al. 2014, *PASJ*, **66**, R1
- Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. 2020, arXiv:1804.06788
- Tejero-Cantero, A., Boeltz, J., Deistler, M., et al. 2020, *JOSS*, **5**, 2505
- Thorne, J. E., Robotham, A. S. G., Davies, L. J. M., et al. 2021, *MNRAS*, **505**, 540
- Treu, T., Auger, M. W., Koopmans, L. V. E., et al. 2010, *ApJ*, **709**, 1195
- Uria, B., Côté, M.-A., Gregor, K., Murray, I., & Laroche, H. 2016, arXiv:1605.02226
- van Dokkum, P. G., & Conroy, C. 2010, *Natur*, **468**, 940
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, *MNRAS*, **444**, 1518
- Walcher, J., Groves, B., Budavári, T., & Dale, D. 2011, *Ap&SS*, **331**, 1
- Westera, P., Lejeune, T., Buser, R., Cuisinier, F., & Bruzual, G. 2002, *A&A*, **381**, 524
- Weyant, A., Schafer, C., & Wood-Vasey, W. M. 2013, *ApJ*, **764**, 116
- Wong, K. W. K., Contardo, G., & Ho, S. 2020, *PhRvD*, **101**, 123005
- Wu, H., Köhler, J., & Noé, F. 2020, arXiv:2002.06707
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, **120**, 1579
- Zhang, K., Bloom, J. S., Gaudi, B. S., et al. 2021, *AJ*, **161**, 11