Draft version December 23, 2021
Typeset using LaTeX **preprint** style in AASTeX63

Accelerated Bayesian SED Modeling using Amortized Neural Posterior Estimation

ChangHoon Hahn[1,*] and Peter Melchior[1]

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

ABSTRACT

State-of-the-art spectral energy distribution (SED) analyses use a Bayesian framework to infer the physical properties of galaxies from spectroscopic and photometric observations. These methods, however, require sampling a high dimensional space of SED model parameters and, thus, take $>$ 100 CPU hours per galaxy. They are not computationally scalable and cannot be feasily used to analyze *millions* of galaxy photometry and spectra from the next-generation galaxy surveys (*e.g.* Rubin, James Webb, and Roman). short description of our SBI method. We combine this SBI-based method with the provabgs SED model, which uses non-parameteric star formation and chemical enrichment histories and a flexible dust attenuation model. We then analyze the multi-wavelength photometry of the NASA-Sloan Atlas with our method as well as with the origin provabgs SED model using standard Markov Chain Monte Carlo (MCMC) sampling. Our method accurately recovers the posteriors for the SED model parameters and inferred galaxy properties. More importantly, while the standard MCMC sampling took 100 CPU hours, our method took 0.2 secs to infer the full posterior.

TODO

1. INTRODUCTION

TODO

paragraph on SED modeling

TODO

paragraph on SED Modeling with a full Bayesian approach - marginalize over nuisance parameters - accurately capture degeneracies among model parameters - allows informative priors based on prior observations

TODO

limitation of Bayesian SED modeling - These methods involve sampling a high dimensional space of SED model parameters. - The dimensionality only increases as SED models become more sophisticated both in terms of stellar population synthesis modeling and accounting for observational effects. - State-of-the-art methods use MCMC or HMC to efficiently sample a high dimensional parameter space. Go through all the latest works and detail their dimensionality and setups. - Despite the use of modern sampling techniques, SED modeling take hundreds of hours per galaxy. Even for $\sim$ 4000 galaxies in the LEGA-C ESO Public Spectroscopic Survey , this required 3.5 million CPU hours Bagpipes.

TODO

simulation based inference provide an overview of simulation based inference

2. SIMULATION-BASED INFERENCE

* changhoon.hahn@princeton.edu.com

The ultimate goal of Bayesian SED modeling, and probabilistic inference more broadly, is to infer the posterior probability distributions of galaxy properties, θ, given observations, $\mathbf{x}_{\rm obs}$ — $p(\theta \,|\, \mathbf{x}_{\rm obs})$. We can evaluate the posterior at a specific θ and $\mathbf{x}$ using Bayes' rule, $p(\theta \,|\, \mathbf{x}_{\rm obs}) \propto p(\theta)\ p(\mathbf{x}_{\rm obs} \,|\, \theta)$. $p(\theta)$ is the prior distribution, which we specify. And $p(\mathbf{x}_{\rm obs} \,|\, \theta)$ is the likelihood, which is *typically* evaluated using a surrogate Gaussian functional form:

$$\ln p(\mathbf{x}_{\rm obs} \,|\, \theta) = -\frac{1}{2}(\mathbf{x}_{\rm obs} - m(\theta))^t \mathbf{C}^{-1} (\mathbf{x}_{\rm obs} - m(\theta)). \tag{1}$$

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

In the standard approach, the full posterior distribution is derived by evaluating the posterior with a sampling technique such as Markov Chain Monte Carlo (MCMC) or nested sampling (*e.g.* Carnall et al. 2017; Leja et al. 2019b; Tacchella et al. 2021). These sampling techniques are essential for the efficient exploration of the relatively higher dimensionality of SED model parameter space. Even advanced techniques, however, are subject to major limitations. For instance, MCMC sampling techniques can struggle to accurately estimate multimodal and degenerate posteriors. Many also require significant hand-tuning by the user. More importantly, despite their efficiency, these techniques require on the order of a *million* SED model evaluations to derive a posterior — this can take ∼100 of CPU hours per galaxy. Analyzing the tens of millions of spectra or billions of photometry from upcoming surveys (*e.g.* DESI, Rubin, Roman) with these approaches would thus require *billions of CPU hours*.

Simulation-based inference (SBI; also known as "likelihood-free" inference) offers a more scalable approach to Bayesian SED modeling. At its core, SBI involves any method that uses a forward model of the observed data to directly estimate the posterior ($p(\theta \,|\, \mathbf{x}_{\rm obs})$), the likelihood ($p(\mathbf{x} \,|\, \theta)$, or the joint distribution of the parameters and data ($p(\theta, \mathbf{x})$). SBI methods have 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; Kacprzak et al. 2018; Alsing et al. 2018; Wong et al. 2020; Huppenkothen & Bachetti 2021; Zhang et al. 2021), and more broadly in physics (*e.g.* Brehmer et al. 2019; Cranmer et al. 2020)

One simple and pedagogical example of SBI is Approximate Bayesian Computation (ABC; Rubin 1984; Pritchard et al. 1999; Beaumont et al. 2002), which uses a rejection sampling framework to estimate the posterior. First, parameter values are sampled from the prior: $\theta' \sim p(\theta)$. The forward model, F is then run on the sampled θ' to generate simulated data $F(\theta') = \mathbf{x}'$. If the simulated $\mathbf{x}'$ is 'close' to the observed $\mathbf{x}_{\rm obs}$, usually based on a threshold on some distance metric $\rho(\mathbf{x}', \mathbf{x}_{\rm obs}) < \epsilon$, θ' is kept. Otherwise, θ' is rejected. This process is repeated until there are enough samples to estimate the posterior. The estimated posterior from ABC can be written as $p(\theta \,|\, \rho(F(\theta), \mathbf{x}_{\rm obs}) < \epsilon)$. In the case where $\epsilon \to 0$, the conditional statement is equivalent to the condition $F(\theta) = \mathbf{x}_{\rm obs}$; thus, the estimated ABC posterior is *equivalent* to the true posterior: $p(\theta \,|\, \rho(F(\theta), \mathbf{x}_{\rm obs}) < \epsilon \to 0) \equiv p(\theta \,|\, \mathbf{x}_{\rm obs})$.

ABC produces unbiased estimates of the posterior and only requires a forward model of the observed data. It makes no assumptions on the likelihood and, therefore, relaxes the assumptions

that go into surrogate likelihood methods. Nevertheless, ABC is based on rejection sampling and thus requires comparable number of model evaluations as standard MCMC sampling based techniques. ABC is only the simplest SBI method; new SBI methods can infer posteriors with much fewer model evaluations. Density estimaiton-based SBI methods (*e.g.* Papamakarios et al. 2017; Alsing et al. 2018; Hahn et al. 2019; Greenberg et al. 2019; Tejero-Cantero et al. 2020), for instance, use model evaluations to fit estimates of $p(\theta \,|\, \mathbf{x}_{\rm obs})$, $p(\mathbf{x} \,|\, \theta$, or $p(\theta, \mathbf{x})$ probability distributions. They can exploit recent advances in neural density estimation (NDE) that increasingly enable high-fidelity density estimation with fewer samples of the distribution. For instance, the NDE in Papamakarios et al. (2017) accurately estimates the $28 \times 28 = 784$-dimensional distribution of the MNIST dataset[1] with only tens of thousands of samples.

2.1. *Amortized Neural Posterior Estimation*

Density estimation SBI provides a key advantage over MCMC sampling-based inference methods — it enables *amortized inference*.

sbi (Greenberg et al. 2019; Tejero-Cantero et al. 2020)

Wong et al. (2020) Dax et al. (2021) Zhang et al. (2021)

- example normalizing flows and MAF

3. NASA-SLOAN ATLAS

As a demonstration of its speed and accuracy, we apply SEDflow to optical photometry from the NASA-Sloan Atlas[2] (NSA) with some additional quality cuts. The NSA catalog is a re-reduction of SDSS DR8 (Aihara et al. 2011) that includes an improved background subtraction (Blanton et al. 2011) and near and far UV photometry from GALEX (). For optical photometry, we use SDSS photometry in the u, g, r, i, and z bands, which are corrected for galactic extinction using Schlegel et al. (1998). For UV photometry, we use GALEX photometry in the $W1$ and $W2$ bands based on DR6[3]. details about the GALEX force photometry TODO

We impose a number of additional quality cuts to the NSA photometry. The SDSS photometric pipeline can struggle to accurately define the center of objects near the edge or at low signal-to-noise. In some cases, the centroiding algorithm will report the position of the peak pixel in a given band as the centroid. These cases are often associated with spurious objects, so we exclude them from our sample. We also exclude objects that have pixels, which were not checked for peaks by the deblender. The SDSS pipeline interpolates over pixels classified as bad (*e.g.* cosmic ray). We exclude objects where more than 20% of point-spread function (PSF) flux is interpolated over as well as objects where the interpolation affected many pixels and the PSF flux error is inaccurate. We also exclude objects where the interpolated pixels fall within 3 pixels of their center and they contain a cosmic ray that was interpolated over. Lastly, we exclude any objects that were not detected at $\geq 5\sigma$ in the original frame, that contain saturated pixels, or where their radial profile could not be extracted. By imposing these quality cuts, we avoid complications from artificats in the photometry that we do not

[1] http://yann.lecun.com/exdb/mnist/
[2] http://nsatlas.org/
[3] http://galex.stsci.edu/GR6/

model. In principle, we can relax the cuts if we were to include observational effects in our model. For additional details on the quality flags, we refer readers to the SDSS documentation[4]. After the quality cuts, we have 33,887 galaxies in our NSA sample.

In Figure **??**, we present the distribution of optical and UV magnitudes of the NSA catalog (color).

4. SEDFLOW

In this work we apply ANPE to SED modeling of galaxy spectra. tl;dr of intro

4.1. *SED Modeling: PROVABGS*

To model SEDs, we use the state-of-the-art stellar population synthesis (SPS) model of the PROVABGS (Hahn *et al.* 2022). With SPS modeling, we model the SED of a galaxy a as a composite of stellar populations defined by stellar evolution theory (in the form of isochrones, stellar spectral libraries, and an initial mass function) and its star formation and chemical enrichment histories (SFH and ZH), attenuated by dust (see Conroy 2013, for a review). The PROVABGS model, in particular, utilizes a non-parametric SFH with a starburst, a non-parametric 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, a starburst component. The SFH contribution from the NMF component is a linear combination of four NMF SFH basis functions, that are derived from performing NMF (Lee & Seung 1999; Cichocki & Phan 2009; Févotte & Idier 2011) on smoothed SFHs of simulated galaxiese of the Illustris cosmological hydrodyanmic simulations (Vogelsberger et al. 2014; Genel et al. 2014; Nelson et al. 2015). The NMF SFH prescription provides a compact and flexible representation of the SFH, assuming that the SFHs of Illustris galaxies resemble the SFHs of real galaxies. To add stochasticity to the SFH, we include a second star burst component that consists of a single stellar population (SSP).

The ZH is similar defined using two NMF bases dervied from Illustris ZHs. Most SPS models assume constant metallicity over time (*e.g.* Carnall et al. 2017; Leja et al. 2019a); however, this assumption can significantly bias inferred galaxy properties (Thorne et al. 2021). Instead by using the NMF prescription, we can flexibly model a range of different ZHs with only two extra parameters. 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) initial mass function (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-space time bin and SSPs are evalulated at each time bin using FSPS. The SSPs are summed up to get the unattenuated galaxy SED.

Finally, 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 the diffuse dust using the Kriek & Conroy (2013) attenuation curve. Then, the BC component provides extra dust attenuation on SSPs younger than 100 Myr with

[4] https://www.sdss.org/dr16/algorithms/flags_detail

young stars that are embedded in modecular clouds and HII regions. In total the PROVABGS SED model has 12 free parameters: stellar mass (M_*), six SFH parameters ($\beta_1, \beta_2, \beta_3, \beta_4, t_{\rm burst}, f_{\rm burst}$), two ZH parameters ($\gamma_1, \gamma 2$), and three dust attenuation parameters ($\tau_{\rm BC}, \tau_{\rm ISM}, n_{\rm dust}$). Each PROVABGS model evaluation takes X seconds.

4.2. *Training Data*

In this section, we describe how to we construct the training data using our SED model. First, we sample $N_{\rm train}$ SED model parameters from a prior: $\theta' \sim p(\theta)$. We use the same priors as Hahn *et al.* (2022): uniform priors over $M_*, t_{\rm burst}, f_{\rm burst}, \gamma_1, \gamma 2, \tau_{\rm BC}, \tau_{\rm ISM}, n_{\rm dust}$ and Dirichlet prior over $\beta_1, \beta_2, \beta_3, \beta_4$. The Dirichlet prior is chosen for the normalization of the NMF SFH while the rest are chosen to span an extensive range of galaxy SEDs.

For each of the sampled SED parameters, we construct mock observables. In this work, our observables are optical photometry of the NASA-Sloan Atlas (Section 3). We first input θ' into the SED model to construct galaxy SEDs: $F(\lambda; \theta')$. Afterwards, we convolve the SEDs with optical broadband filters, R_X to generate noiseless photometric fluxes:

$$f_X(\theta') = \int F(\lambda; \theta')\, R_X(\lambda)\, {\rm d}\lambda \tag{2}$$

Next, to generate synthetic photometry, we must apply a realistic noise model.

- noise model

4.3. *Training ANPE*

description of the ANPE training

- architecture
- validation

5. RESULTS

validate the normaling flow SBI posteriors for a single case compare the corner plot of a posterior derived from MCMC with the SBI TODO

validate the derived properties of galaxies for a handful of galaxies TODO

paragraph on the computational advantage TODO

6. SUMMARY

ACKNOWLEDGEMENTS

It's a pleasure to thank Adam Carnall, Kartheik Iyer, Jenny E. Green, Uroš Seljak, Michael A. Strauss, ... for valuable discussions and comments.

APPENDIX

REFERENCES

Aihara, H., Allende Prieto, C., An, D., et al. 2011, The Astrophysical Journal Supplement Series, 193, 29, doi: 10.1088/0067-0049/193/2/29

Alsing, J., Wandelt, B., & Feeney, S. 2018, arXiv:1801.01497 [astro-ph]. https://arxiv.org/abs/1801.01497

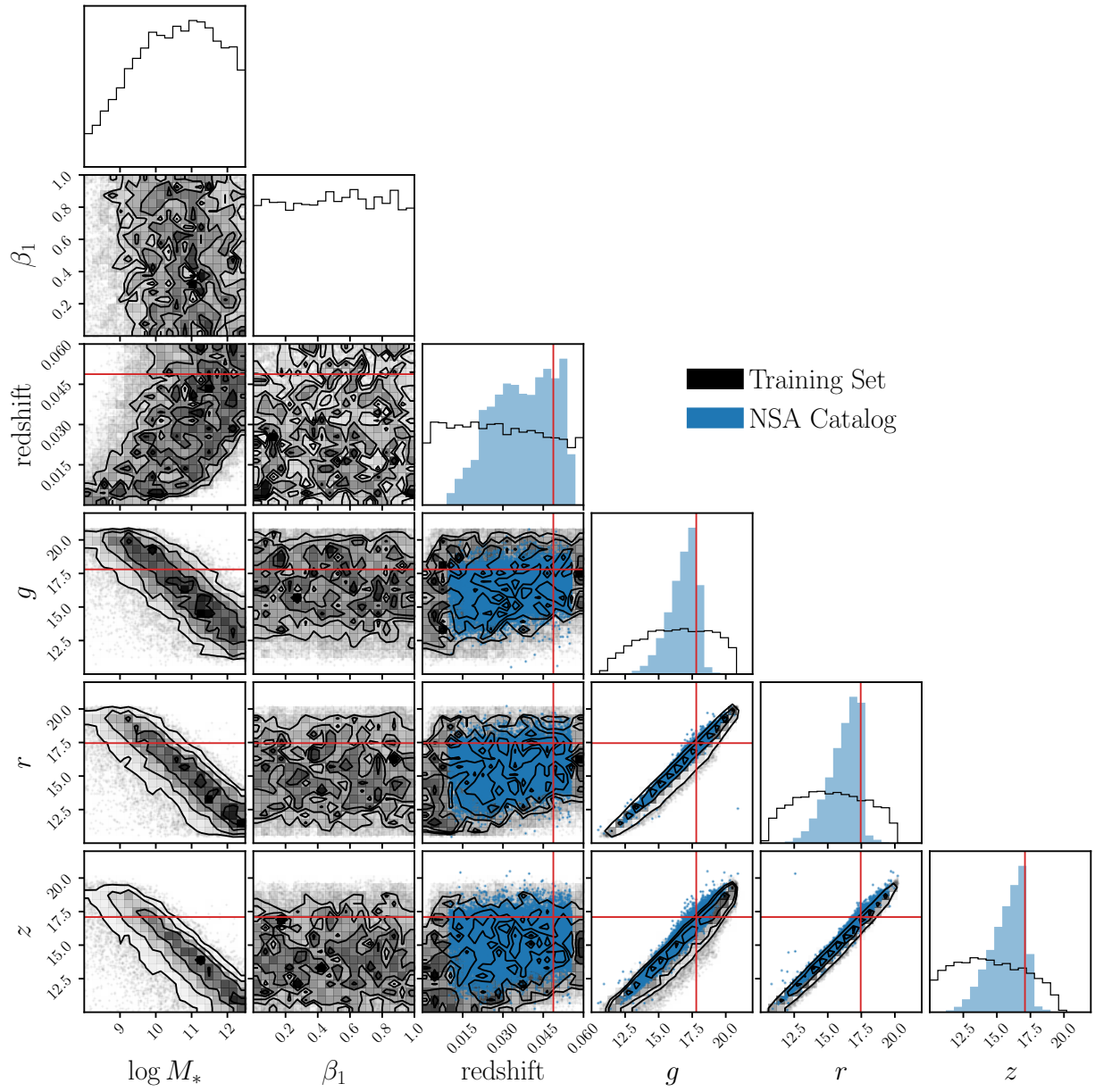

Figure 1. Joint distribution of SED model parameters ($\log M_*$, β_1, redshift) and photometric magnitudes (g, r, z) for our training set. The training set was constructed by sampling parameter values from the prior (Table **??**), constructing SEDs using a theoretical SPS model, and applying our noise model. For details, we refer readers to Section **??**. For comparison, we present the distribution of magnitudes for galaxies in the NSA catalog (blue). *The training set fully encompasses the observations, thus, our* SEDFLOW *method can be used to infer the posterior for all NSA galaxies.*

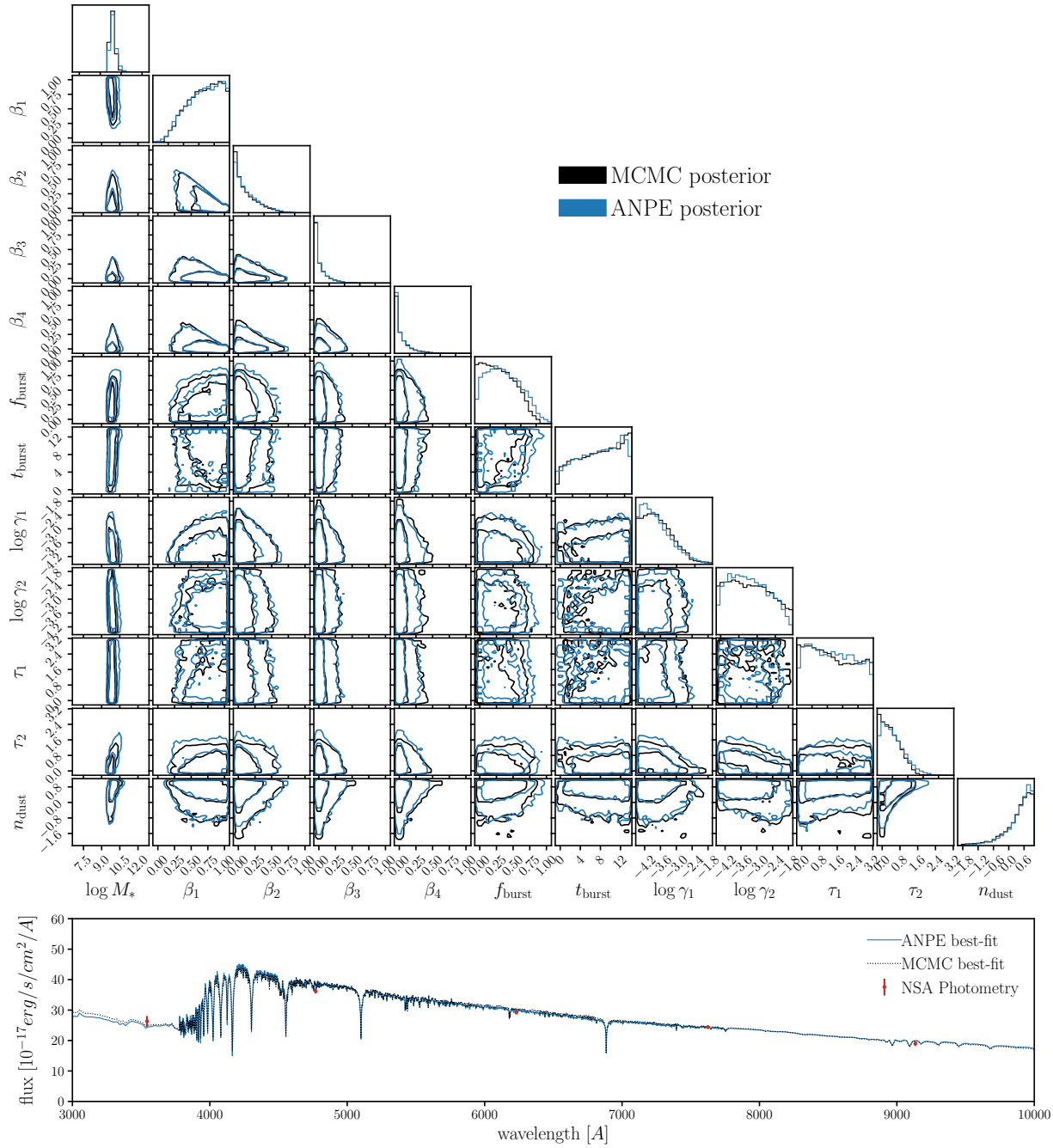

Figure 2. A comparison of the posteriors of the 12 SED model parameters derived from standard MCMC sampling (black) and ANPE (orange) for a randomly selected NSA galaxy. The posteriors are in excellent agreement for all of the parameters. Estimating the posterior using MCMC sampling requires X hours. Even using neural emulators to accelerate likelihood evaluations, MCMC sampling requires Y hours. *With ANPE, inferring the full posterior for a galaxy only requires 1 second.*

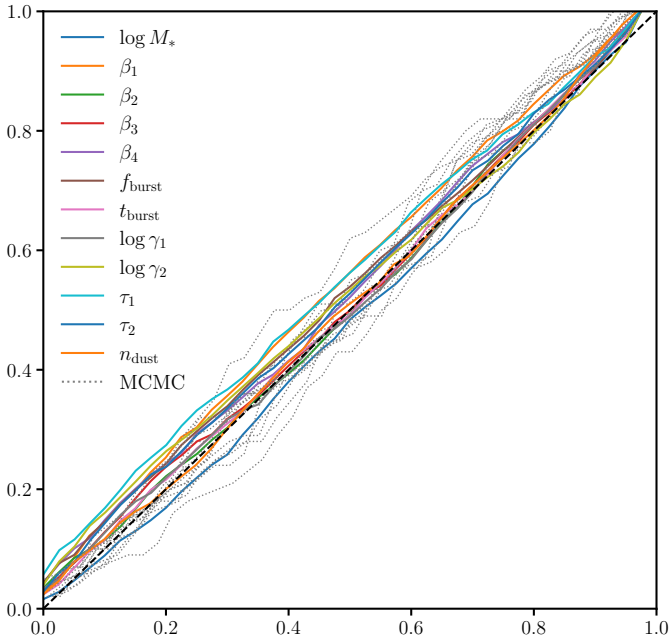

Figure 3. Probability-probability (p-p) plot of the ANPE for 1000 simulated test data. For each SPS parameter, we plot the cumulative distribution function (CDF) of the percentile score of the true value within the ANPE marginalized posterior. For the true posteriors, the percentile score is uniformly distributed so the CDF is diagonal (black dashed). The test data is constructed in the same way as the training data (Section 4.2). For reference, we include the p-p plot of the posterior estimated from MCMC sampling (gray). *The ANPE is in good agreement the true posterior.*

Beaumont, M. A., Zhang, W., & Balding, D. J. 2002, Genetics, 162, 2025

Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011, The Astronomical Journal, 142, 31, doi: 10.1088/0004-6256/142/1/31

Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. 2019, arXiv:1805.12244 [hep-ph, physics:physics, stat]. https://arxiv.org/abs/1805.12244

Cameron, E., & Pettitt, A. N. 2012, Monthly Notices of the Royal Astronomical Society, 425, 44, doi: 10.1111/j.1365-2966.2012.21371.x

Carnall, A. C., McLure, R. J., Dunlop, J. S., & Davé, R. 2017, arXiv:1712.04452 [astro-ph]. https://arxiv.org/abs/1712.04452

Chabrier, G. 2003, Publications of the Astronomical Society of the Pacific, 115, 763, doi: 10.1086/376392

Charlot, S., & Fall, S. M. 2000, The Astrophysical Journal, 539, 718, doi: 10.1086/309250

Choi, J., Dotter, A., Conroy, C., et al. 2016, The Astrophysical Journal, 823, 102, doi: 10.3847/0004-637X/823/2/102

Cichocki, A., & Phan, A.-H. 2009, IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences, 92, 708, doi: 10.1587/transfun.E92.A.708

Conroy, C. 2013, Annual Review of Astronomy and Astrophysics, 51, 393, doi: 10.1146/annurev-astro-082812-141017

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

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

Cranmer, K., Brehmer, J., & Louppe, G. 2020, Proceedings of the National Academy of Sciences, 117, 30055, doi: 10.1073/pnas.1912789117

Dax, M., Green, S. R., Gair, J., et al. 2021, arXiv:2106.12594 [astro-ph, physics:gr-qc]. https://arxiv.org/abs/2106.12594

Dotter, A. 2016, The Astrophysical Journal Supplement Series, 222, 8, doi: 10.3847/0067-0049/222/1/8

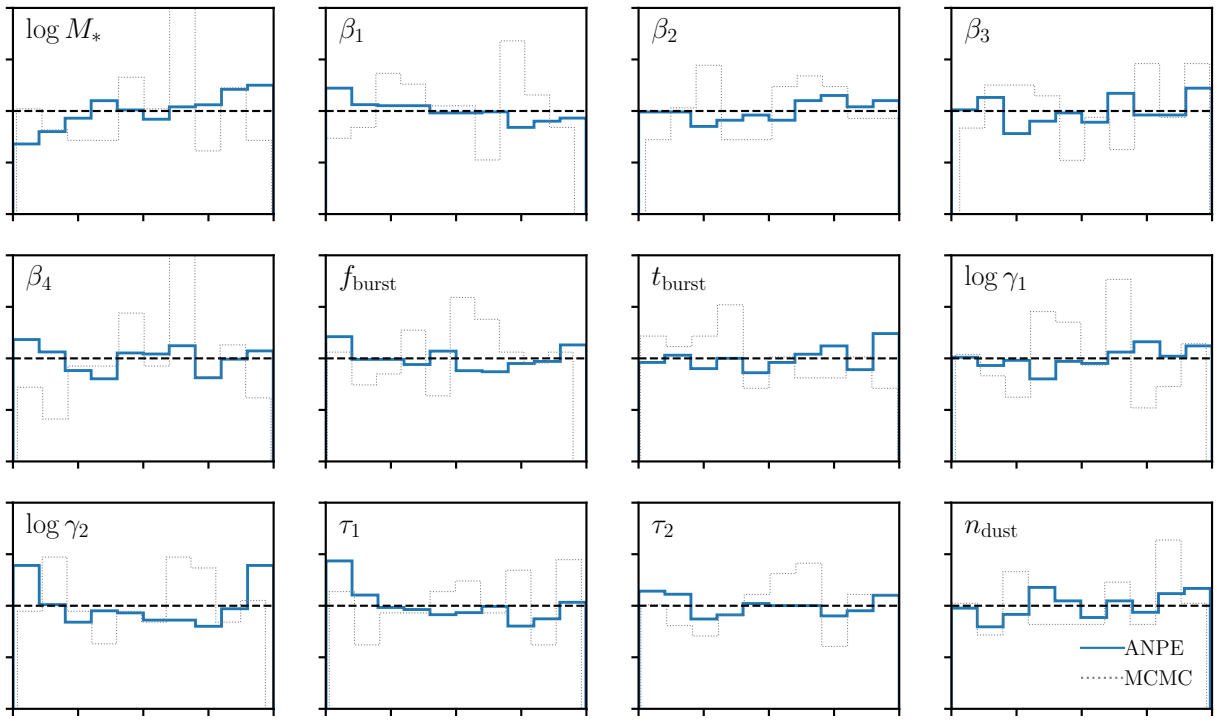

Figure 4. Simulation-based calibration plot of the ANPE for 1000 simulated test data. For each SPS parameter, we plot the cumulative distribution function (CDF) of the percentile score of the true value within the ANPE marginalized posterior. For the true posteriors, the percentile score is uniformly distributed so the CDF is diagonal (black dashed). The test data is constructed in the same way as the training data (Section 4.2). For reference, we include the p-p plot of the posterior estimated from MCMC sampling (gray). *The ANPE is in good agreement the true posterior.*

Févotte, C., & Idier, J. 2011, arXiv:1010.1763 [cs]. https://arxiv.org/abs/1010.1763

Genel, S., Vogelsberger, M., Springel, V., et al. 2014, Monthly Notices of the Royal Astronomical Society, 445, 175, doi: 10.1093/mnras/stu1654

Greenberg, D. S., Nonnenmacher, M., & Macke, J. H. 2019, Automatic Posterior Transformation for Likelihood-Free Inference

Hahn, C., Beutler, F., Sinha, M., et al. 2019, Monthly Notices of the Royal Astronomical Society, 485, 2956, doi: 10.1093/mnras/stz558

Hahn, C., Vakili, M., Walsh, K., et al. 2017, Monthly Notices of the Royal Astronomical Society, 469, 2791, doi: 10.1093/mnras/stx894

Huppenkothen, D., & Bachetti, M. 2021, Accurate X-Ray Timing in the Presence of Systematic Biases With Simulation-Based Inference

Kacprzak, T., Herbel, J., Amara, A., & Réfrégier, A. 2018, Journal of Cosmology and Astro-Particle Physics, 2018, 042, doi: 10.1088/1475-7516/2018/02/042

Kriek, M., & Conroy, C. 2013, The Astrophysical Journal Letters, 775, L16, doi: 10.1088/2041-8205/775/1/L16

Lee, D. D., & Seung, H. S. 1999, Nature, 401, 788, doi: 10.1038/44565

Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2019a, ApJ, 876, 3, doi: 10.3847/1538-4357/ab133c

Leja, J., Speagle, J. S., Johnson, B. D., et al. 2019b, arXiv, arXiv:1910.04168

Lejeune, T., Cuisinier, F., & Buser, R. 1997, A & A Supplement series, Vol. 125, October II 1997, p.229-246., 125, 229, doi: 10.1051/aas:1997373

—. 1998, Astronomy and Astrophysics Supplement, v.130, p.65-75, 130, 65, doi: 10.1051/aas:1998405

Nelson, D., Pillepich, A., Genel, S., et al. 2015, Astronomy and Computing, 13, 12, doi: 10.1016/j.ascom.2015.09.003

Papamakarios, G., Pavlakou, T., & Murray, I. 2017, arXiv e-prints, 1705, arXiv:1705.07057

Paxton, B., Bildsten, L., Dotter, A., et al. 2011, The Astrophysical Journal Supplement Series, 192, 3, doi: 10.1088/0067-0049/192/1/3

Paxton, B., Cantiello, M., Arras, P., et al. 2013, The Astrophysical Journal Supplement Series, 208, 4, doi: 10.1088/0067-0049/208/1/4

Paxton, B., Marchant, P., Schwab, J., et al. 2015, The Astrophysical Journal Supplement Series, 220, 15, doi: 10.1088/0067-0049/220/1/15

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., & Feldman, M. W. 1999, Molecular Biology and Evolution, 16, 1791, doi: 10.1093/oxfordjournals.molbev.a026091

Rubin, D. B. 1984, The Annals of Statistics, 12, 1151

Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, Monthly Notices of the Royal Astronomical Society, 371, 703, doi: 10.1111/j.1365-2966.2006.10699.x

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, The Astrophysical Journal, 500, 525, doi: 10.1086/305772

Tacchella, S., Conroy, C., Faber, S. M., et al. 2021, arXiv e-prints, 2102, arXiv:2102.12494

Tejero-Cantero, A., Boelts, J., Deistler, M., et al. 2020, Journal of Open Source Software, 5, 2505, doi: 10.21105/joss.02505

Thorne, J. E., Robotham, A. S. G., Davies, L. J. M., et al. 2021, Monthly Notices of the Royal Astronomical Society, 505, 540, doi: 10.1093/mnras/stab1294

Vogelsberger, M., Genel, S., Springel, V., et al. 2014, Monthly Notices of the Royal Astronomical Society, 444, 1518, doi: 10.1093/mnras/stu1536

Westera, P., Lejeune, T., Buser, R., Cuisinier, F., & Bruzual, G. 2002, Astronomy and Astrophysics, 381, 524, doi: 10.1051/0004-6361:20011493

Weyant, A., Schafer, C., & Wood-Vasey, W. M. 2013, The Astrophysical Journal, 764, 116, doi: 10.1088/0004-637X/764/2/116

Wong, K. W. K., Contardo, G., & Ho, S. 2020, Physical Review D, 101, 123005, doi: 10.1103/PhysRevD.101.123005

Zhang, K., Bloom, J. S., Gaudi, B. S., et al. 2021, doi: 10.3847/1538-3881/abf42e