# Hierarchical Bayesian Inference of Photometric Redshifts with Stellar Population Synthesis Models

Boris Leistedt[1] , Justin Alsing[2] , Hiranya Peiris[2,3] , Daniel Mortlock[1,2,4] , and Joel Leja[5,6,7]

[1] Department of Physics, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK; b.leistedt@imperial.ac.uk
[2] Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Stockholm SE-106 91, Sweden
[3] Department of Physics and Astronomy, University College London, Gower Street, London, WC1E 6BT, UK
[4] Department of Mathematics, Imperial College London, London SW7 2AZ, UK
[5] Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA
[6] Institute for Computational & Data Sciences, The Pennsylvania State University, University Park, PA, USA
[7] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA

## Abstract

We present a Bayesian hierarchical framework to analyze photometric galaxy survey data with stellar population synthesis (SPS) models. Our method couples robust modeling of spectral energy distributions with a population model and a noise model to characterize the statistical properties of the galaxy populations and real observations, respectively. By self-consistently inferring all model parameters, from high-level hyperparameters to SPS parameters of individual galaxies, one can separate sources of bias and uncertainty in the data. We demonstrate the strengths and flexibility of this approach by deriving accurate photometric redshifts for a sample of spectroscopically confirmed galaxies in the COSMOS field, all with 26-band photometry and spectroscopic redshifts. We achieve a performance competitive with publicly released photometric redshift catalogs based on the same data. Prior to this work, this approach was computationally intractable in practice due to the heavy computational load of SPS model calls; we overcome this challenge by the addition of neural emulators. We find that the largest photometric residuals are associated with poor calibration for emission-line luminosities and thus build a framework to mitigate these effects. This combination of physics-based modeling accelerated with machine learning paves the path toward meeting the stringent requirements on the accuracy of photometric redshift estimation imposed by upcoming cosmological surveys. The approach also has the potential to create new links between cosmology and galaxy evolution through the analysis of photometric data sets.

*Unified Astronomy Thesaurus concepts:* Observational cosmology (1146); Large-scale structure of the universe (902); Redshift surveys (1378)

## 1. Introduction

Deriving accurate redshifts and redshift distributions from photometry alone is of central importance for the scientific exploitation of modern galaxy surveys. In particular, cosmological analyses involving galaxy clustering and weak gravitational lensing require exquisitely accurate estimates of the redshift distributions of the selected galaxy populations (see, e.g., Hildebrandt et al. 2020; Abbott et al. 2022). Photometric redshifts (photo-$z$'s) of individual galaxies are also useful for selecting targets of interest (e.g., at high redshift).

Photo-$z$'s are typically obtained via two types of methods: template fitting, and machine learning.[8] Template fitting (e.g., Benitez 2000; Ilbert et al. 2006, 2006; Brammer et al. 2008; Acquaviva et al. 2015; Tanaka 2015; Carnall et al. 2018; Battisti et al. 2019) relies on lists of grids or templates for the spectral energy distributions (SEDs) of the galaxies considered, combined with a prior on the parameters involved (galaxy type, magnitude, etc.). Machine-learning techniques

(e.g., Collister & Lahav 2004; Kind & Brunner 2013) involve flexible models trained on spectroscopic or synthetic data. Despite their past successes, both these approaches are unable to satisfy the accuracy requirements imposed by upcoming Stage IV surveys, as they are not robust to errors in the data (calibration offsets, underestimated uncertainties), in the SED model, or to the non-representativity or sample variance of the spectroscopic data used for training and validation (e.g., Hartley et al.2020; Gatti et al. 2021; Myles et al. 2021; Newman & Gruen 2022). For example, the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) will require the uncertainty on the mean of redshift distributions to be smaller than $0.001(1 + z)$ by its Year 10 data release (The LSST Dark Energy Science Collaboration et al. 2018), which in turn imposes very stringent requirements on the accuracy of individual photo-$z$'s.

We introduce a framework capable of addressing these needs: a hierarchical model of galaxy photometry, consisting of three components describing a process for generating survey data:

1. galaxy SEDs generated with stellar population synthesis;
2. a population model describing the distributions of the intrinsic parameters of galaxies in the sample analyzed;
3. a noise model for generating photometric observations (i.e., a likelihood function).

---

[8] A third technique, clustering redshifts, can constrain redshift distributions, but cannot deliver individual photo-$z$'s, see, e.g., Schneider et al. (2006), Newman (2008), Ménard et al. (2013), McQuinn & White (2013), Schmidt (2013), Morrison et al. (2017), Gatti et al. (2021).

This provides a powerful combination of empirical modeling with physics-driven components, incorporating known (astro) physics while leaving sufficient room for data-driven corrections, addressing the aforementioned issues of flexibility, interpretability, miscalibration and extrapolation outside of training data. Thus, once applied and calibrated to large survey data sets, it has the potential to yield both precise and accurate inferences on intrinsic galaxy parameters, including redshift. Building on previous work on hierarchical modeling in photometric surveys (Leistedt et al. 2016; Jones & Heavens 2018; Sánchez & Bernstein 2018; Rau et al. 2019; Alarcon et al. 2020; Sánchez et al. 2020), our framework takes advantage of the increasing accuracy and flexibility of SPS, and of methodological innovations in Bayesian inference for large hierarchical models, in particular involving machine-learning-facilitated acceleration of computationally intensive analysis steps.

In this paper, we demonstrate the efficacy of this framework by applying it to the COSMOS2020 data set (Weaver et al. 2022). For the choice of SED model, we consider PROSPECTOR-$\alpha$ (Leja et al. 2017, 2019; Johnson et al. 2021). We address the high computational needs of its model calls with neural emulators (Alsing et al. 2020) applied to both SPS total fluxes and emission-line contributions. We find that the emission-line contributions are responsible for the largest photometric residuals in our analysis and therefore require calibration in order to derive accurate photo-$z$'s. Together, the SED, data, and population models form a complex hierarchical model which would normally be computationally intractable, a challenge we overcome with neural emulator-based acceleration. We obtain redshift estimates competitive with those publicly released with the COSMOS2020 data using this baseline model, providing a benchmark validation test demonstrating the power of the new methodology. The extensive 26-band wavelength coverage of COSMOS2020 simplifies parameter inference and allows us to demonstrate the conceptual advantages of the new methodology without the significant overhead of carrying out a Markov Chain Monte Carlo (MCMC) analysis. Application to broadband-only photometric surveys, such as the LSST, will require a more involved treatment of the uncertainties and selection effects, which we defer to future work. In a companion paper (Alsing et al. 2022), we show how to correctly include selection effects in the forward modeling of such surveys, and in particular for the inference of galaxy redshift distributions.

The outline of this paper is as follows. In Section 2 we describe our framework in the form of a generative model of photometric data. An inference formalism is presented in Section 3. We describe the data in Section 4 and the results in Section 5. We discuss the results in further detail in Section 6 and conclude in Section 7. In what follows, a Planck 2015 (Ade et al. 2016) and a Chabrier (2003) initial mass function are adopted for all relevant calculations. $p(\cdot \,|\, \cdot)$ refers to the probability density of the quantity before the "|," conditioned on quantities behind it. We work with fluxes in AB units.

## 2. Generative Model for Galaxy Photometry

Hierarchical models tackling photometric redshifts and redshift distributions have been developed (Leistedt et al. 2016; Jones & Heavens 2018; Sánchez & Bernstein 2018; Rau et al. 2019; Alarcon et al. 2020; Sánchez et al. 2020), but typically assume a fixed SED model. Leistedt et al. (2019) pioneered the approach of inferring redshifts jointly with the

calibration of the hyperparameters of an SED model. However, like most photo-$z$ methodologies (Ilbert et al. 2006; Brammer et al. 2008; Acquaviva et al. 2015; Tanaka 2015; Carnall et al. 2018; Battisti et al. 2019), the model still relied on a curated set of SED templates, which made it difficult to construct more structured, interpretable corrections to the SEDs or to the population model. Our work introduces the use of full continuous stellar population synthesis (SPS) models for redshift estimation, tailored to the multilevel calibration and inference that will be needed for the upcoming generation of wide-deep photometric surveys (Newman & Gruen 2022).

### 2.1. SED Model

SPS provides a powerful way to model the SED of a galaxy. It exploits fundamental principles to generate and sum the contributions of "simple" stellar populations (as observed in, e.g., star clusters) and to apply the effects of additional complications, such as dust and nebular emission, in order to create galaxy SEDs within a forward modeling framework. This step involves a set of specific modeling choices for these components; for example an explicit star formation history (SFH). We use the Flexible Stellar Population Synthesis (FSPS; Conroy et al. 2009a, 2009b; Conroy & Gunn 2010) code, accessed through the PYTHON-FSPS binding (Foreman-Mackey et al. 2014).

For the SED model we consider a variation of PROSPECTOR-$\alpha$, which was first introduced in Leja et al. (2017), revised in Leja et al. (2019), Johnson et al. (2021), and has been successfully used to constrain Bayesian models for the galaxy population in Leja et al. (2020, 2022), Nagaraj et al. (2022) and Whitler et al. (2022). While these works fixed the redshift to the value measured with spectroscopy (spec-$z$), we relax this assumption and treat it as a free parameter. In our variation of PROSPECTOR-$\alpha$, a galaxy SED is (deterministically) defined by 15 parameters:

1. Eight parameters describing the SFH: one parameter for the total stellar mass formed $\log M_\star$ $[M_\odot]$ (integral of the SFH); one parameter for the stellar metallicity $\log(Z_\star/Z_\odot)$, assumed to be the same for all stars in the galaxy; and six parameters for the relative age SFH bins (ratios $r_i$ of the piece-wise SFR in adjacent temporal bins), where the seven time bins are spaced following Leja et al. (2020).
2. Three parameters for the dust attenuation, following the model of Charlot & Fall (2000) (see Leja et al. 2017 for details) with birth cloud $\tau_1$, and diffuse attenuation $\tau_2$ with a power law (of index $n$) from Calzetti et al. (2000).
3. A gas-phase metallicity parameter $\log(Z_{gas}/Z_\odot)$ for the nebular emission (decoupled from the stellar metallicity) modeled after the grids of Byler et al. (2017).
4. Two parameters, $\log f_{AGN}$ and $\tau_{AGN}$, for the AGN torus emission model of Nenkova et al. (2008).
5. Redshift, $z$.

The SED includes dust heating from stars via energy balance, via a dust SED of fixed shape (Draine & Li 2007).

### 2.2. Emission Lines

PROSPECTOR-$\alpha$ includes a nebular emission model where the gas is ionized by the same stars synthesized in the SED (Byler et al. 2017; Byler 2018). There is a large variability of

the line strengths, as demonstrated by Byler et al. (2017). In practice, previous studies have required the flexible addition of emission lines on top of SED templates in order to obtain accurate photo-$z$'s (Ilbert et al. 2006, 2008; Brammer et al. 2008; Alarcon et al. 2021), so we adopt a similar approach.

We model offsets and uncertainty in the strength of the emission lines by encoding them into parametric bias and variance parameters. For a set of SPS parameters $\varphi$, the total energy density as a function of wavelength, $\ell$, consists of the base FSPS prediction $\ell_{SPS}$, with an extra additive contribution from emission lines. In practice, in FSPS, emission lines are modeled as delta functions, integrated in bandpasses and added to the model photometry. We can therefore write the SED as

$$\ell(\lambda; \varphi, \alpha) = \ell_{SPS}(\lambda; \varphi) + \sum_{j=1}^{N_{lines}} \alpha_j \ell_j(\varphi) \delta^D(\lambda - \lambda_j), \quad (1)$$

where $\ell_j$ is the amplitude of the $j$th line, $\lambda_j$ the rest-frame wavelength of the line and $\delta^D(\cdot)$ is the Dirac delta function. $\alpha_j$ is the hyperparameter for the additional contribution from line $j$. Because line $\ell_j$ is already included in $\ell_{SPS}$ with a default weight, its total contribution in $\ell$ is $1 + \alpha_j$.

### 2.3. Synthetic Photometry

We can obtain total model fluxes in the $b$th band by integrating the SED through the bandpass filter $W_b(\lambda)$, such that

$$F_b(\varphi, z, \alpha)$$
$$= \frac{(1+z)^{-1}}{4\pi d_L^2(z)} \int_0^\infty \ell\left(\frac{\lambda}{1+z}; \varphi, \alpha\right) e^{-\tau(z,\lambda)} W_b(\lambda) d\lambda, \quad (2)$$

where $d_L(z)$ is the luminosity distance and $\tau(z, \lambda)$ is the effective optical depth of the intergalactic medium, calculated with the model of Madau (1995).

### 2.4. Population Model

So far we have a mechanism for using SPS to generate the photometry of a single galaxy given its intrinsic properties, $\varphi$ and redshift, $z$. We now describe a formalism for generating a sample of galaxies with a probability distribution $p(\varphi, z|\kappa)$, where $\kappa$ are the hyperparameters which describe the population. These could in principle be inferred as well, but this typically requires a careful treatment of selection effects (i.e., how galaxies are selected into the sample at hand), which must be included at the population level. We show how to treat selection effects in a companion paper, Alsing et al. (2022). For our demonstration with COSMOS2020 in this paper, we focus on redshift inference for individual galaxies. Due to the constraining power of the data, the population prior will have very little impact on the galaxy posterior distributions, especially the redshifts. Therefore, we are able to bypass an explicit treatment of selection effects and also adopt a fairly uninformative population model.

We adopt minor variations on the fiducial PROSPECTOR-$\alpha$ prior (see Leja et al. 2019; Johnson et al. 2021),

**Table 1**
Summary of Parameters and Priors Describing the SPS Model

| Parameter | Prior Bounds | Prior |
|---|---|---|
| $\log M_\star \, [M_\odot]$ | $[7, 13]$ | Uniform(7, 13) |
| $\log(Z_\star/Z_\odot)$ | $[-1.98, 0.19]$ | Uniform$(-1.98, 0.19)$ |
| $\{\log r_i\}_{i=1,\cdots,6}$ | $[-5, 5]$ | Stu(0, 0.3, 2) |
| $\tau_2$ | $[0, 4]$ | $\mathcal{N}_c(\mu_2, \sigma_2, 0, 4)$ |
| $\tau_1/\tau_2$ | $[0, 2]$ | $\mathcal{N}_c(\mu_1, \sigma_1, 0, 2)$ |
| $n$ | $[-1, 0.4]$ | $\mathcal{N}_c(\mu_n, \sigma_n, -1, 0.4)$ |
| $\log f_{AGN}$ | $[10^{-5}, 150]$ | LogUniform$(10^{-5}, 150)$ |
| $\tau_{AGN}$ | $[-2, 0.5]$ | LogUniform(5, 150) |
| $\log(Z_{gas}/Z_\odot)$ | $[-2, 0.5]$ | $\mathcal{N}_c(\log(Z_\star/Z_\odot), \sigma_Z, -2, 0.5)$ |
| $z$ | $[0, 2.5]$ | Uniform(0, 2.5) |

**Note.** $\mathcal{N}_c(l, s, m, M)$ refers to a truncated normal distribution of location $l$, scale $s$, and in the range $[m, M]$. Our fiducial values for the parameters below are $\mu_n = -0.095 + 0.111\,\tau_2 - 0.0066\,\tau_2^2$, $\sigma_n = 0.4$, $\sigma_Z = 3.0$, $(\mu_1, \sigma_1) = (1.0, 0.3)$, $(\mu_2, \sigma_2) = (0.3, 1.0)$, following Leja et al. (2019).

summarized in Table 1. It factorizes as

$$
\begin{aligned}
p(\varphi, z|\kappa) &= p(\varphi|z, \kappa)\,p(z) \\
&= p(\text{mass}) \times p(\text{SFH}) \times p(\text{dust}|\kappa) \\
&\quad \times p(\text{stellar metallicity}) \times p(\text{AGN}) \times p(z) \\
&\quad \times p(\text{gas-phase metallicity}|\text{stellar metallicity}, \kappa).
\end{aligned} \quad (3)
$$

Most parameters have simple uniform or log-uniform priors, except in the cases discussed below.

For the gas-phase metallicity, we take a normal[9] prior of mean at the stellar metallicity and standard deviation $\sigma_Z = 3.0$. Despite the fact that they should track each other approximately, it is more robust to treat the gas-phase metallicity as a nuisance parameter (Leja et al. 2020). We take a normal prior on the diffuse dust components $\tau_2$ with mean equal to $\mu_2 = 0.3$ and standard deviation $\sigma_2 = 1.0$. For the birth cloud component $\tau_1$, we take a normal prior on the ratio $r = \tau_1/\tau_2$, with mean equal to $\mu_1 = 1$ and standard deviation $\sigma_1 = 0.3$. The index of the dust attenuation law (for the diffuse component) is assumed to vary as a function of the total dust attenuation, with mean given by: $\langle\delta\rangle = \mu_n = -0.095 + 0.111\,\tau_2 - 0.0066\,\tau_2^2$, where $\delta$ is the (negative) offset from the index of the Calzetti attenuation curve (Calzetti et al. 2000). We take a normal prior on $\delta$, with mean $\mu_n$ given above and standard deviation $\sigma_\delta = 0.4$. This is a simple average of the results of Leja et al. (2019), but could be updated with a more complex prior, such as that of Nagaraj et al. (2022).

Finally, we collect the following hyperparameters in a vector $\kappa = (\mu_r, \sigma_r, \sigma_n, \sigma_\delta, \sigma_Z)$. We will infer them jointly with the other components of the model, in order to demonstrate how one could, in principle, learn about the underlying physics with this type of forward modeling framework. However, we expect little sensitivity to these hyperparameters for the data set we use. A more detailed population prior, encoding more of the known relations from galaxy formation and evolution, was developed by Alsing et al. (2022), and successfully reproduced galaxy redshift distributions given broadband photometry.

---

[9] Since we apply normal priors to parameters only defined in finite ranges, a renormalization is necessary, giving rise to truncated normals, as described in Table 1. For simplicity we omit this detail throughout this section.

**Table 2**
Notation for All Model Parameters

| Parameter | Description |
|---|---|
| | *Hyperparameters* |
| $\omega_b$ | Zero-point offset (for $b$th band, relative to the $i$ band) |
| $\gamma_b$ | Additional flux uncertainty contribution for the $b$th band (fraction of the total model flux) |
| $\alpha_j$ | Additional flux for the $j$th emission line (fraction of the line strength in FSPS) |
| $\beta_j$ | Additional flux uncertainty contribution for the $j$th emission line (fraction of the line strength in FSPS) |
| $\kappa$ | Hyperparameters describing the galaxy population model |
| | *Latent parameters* |
| $\varphi$ | Stellar population parameters describing the rest-frame spectrum (per galaxy) |
| $z$ | Redshift (per galaxy) |
| | *Derived parameters* |
| $\ell_{\mathrm{SPS}}(\varphi)$ | Flux energy density predicted by FSPS (per galaxy) |
| $\ell_j(\varphi)$ | Amplitude of emission line (Dirac delta model) in FSPS (per galaxy) |
| $L_{jb}(\varphi, z)$ | Flux of the $j$th emission line in the $b$th band (per galaxy), i.e., $\ell_j$ inserted in Equation (3) |
| $F_b(\varphi, z, \boldsymbol{\alpha})$ | Total model flux in the $b$th band, with all contributions (per galaxy), i.e., Equation (1) inserted into Equation (3) |
| $\Sigma_b(\varphi, z, \boldsymbol{\alpha}, \gamma_b, \boldsymbol{\gamma})$ | Additional flux uncertainty in the $b$th band, with all contributions (per galaxy), see Equation (5) |
| | *Data* |
| $\hat{F}_b$ | Measured flux in the $b$th band (per galaxy) |
| $\sigma_b$ | Flux measurement uncertainty in the $b$th band (per galaxy) |

**Note.** Vectors are indicated with bold symbols. Our total log likelihood and graphical model presented below include an extra index $i$ per galaxy, which we have omitted in this table and all other equations in the paper for simplicity.

### 2.5. Noise Model

The final stage of the generative model is to simulate observed photometry $\hat{F}_b$ from the model flux $F_b$, by adding noise. The noise model can be fully characterized using a likelihood function. We use a scaled and translated Student's-t distribution[10] with two degrees of freedom, which is more robust to outliers than a normal distribution, since it has heavier tails:

$$p(\hat{F}_b | F_b, \sigma_b, \Sigma_b, \omega_b)$$
$$= \mathrm{Stu}(\hat{F}_b; \omega_b F_b, \sigma_b^2 + (\omega_b \Sigma_b)^2, 2). \qquad (4)$$

Aside from the observational uncertainty $\sigma_b$, which is given for each object (typically measured from the images), we introduce two additional terms: a zero-point offset $\omega_b$ and the model uncertainty $\Sigma_b$ added in quadrature. $\omega_b$ appears in front of the uncertainty because the latter will involve the model $F_b$, which must be rescaled by $\omega_b$ (although there is some freedom in choosing how to define these terms self-consistently).

We write the model uncertainty as the sum (in quadrature) of a fraction of the total model flux and a fraction of the total line

---

[10] The density of the scaled and translated Student's-t distribution is

$$\mathrm{Stu}(x; l, s, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

with $t = (x - l)/s$ as the residuals and $\Gamma(\cdot)$ as the Gamma function. $l$ and $s$ are often referred to as location and scale parameters and $\nu$ as the number of degrees of freedom.
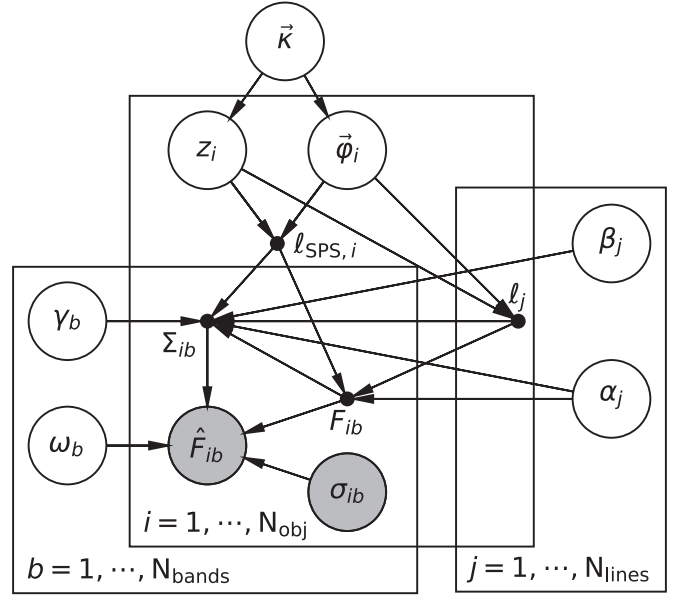


**Figure 1.** Graphical representation of the main parameters involved of our model. Circles are inferred random variables, shaded circles are observed and dots indicate random variables that are deterministic in their inputs.

fluxes,

$$\Sigma_b^2(\varphi, z, \boldsymbol{\alpha}, \gamma_b, \boldsymbol{\gamma})$$
$$= (\gamma_b F_b(\varphi, z))^2 + \sum_{j=1}^{N_{\mathrm{lines}}} (\beta_j(1 + \alpha_j)L_{jb}(\varphi, z))^2, \qquad (5)$$

where $L_{jb}$ is the photometric flux of the $j$th line in the $b$th band, i.e., the result of applying Equation (2) to $\ell_j \delta^{\mathrm{D}}(\lambda - \lambda_j)$. The first term operates at the level of the band (controlled by a hyperparameter $\gamma_b$) and is analogous to the magnitude uncertainty floor often employed in photo-$z$ techniques. The second term is similar, but adds a contribution from each line. As before, the hyperparameter $\beta_j$ can be interpreted in terms of a fraction of the total emission-line flux (which is now $1 + \alpha_j$). In summary, this allows us to separate the extra uncertainty $\Sigma_b$ preferred by the data into contributions from calibration uncertainty (via $\gamma_b$) and SED uncertainty (via $\beta_j$).

## 3. Inference Formalism

We now assemble the components presented in the previous section into a hierarchical model for generating and analyzing photometric data. We describe an inference methodology and also computational acceleration using machine learning.

### 3.1. Hierarchical Model

A summary of the parameters of our model is provided in Table 2. A graphical representation is provided in Figure 1. Each galaxy has 15 intrinsic SPS parameters (including redshift), as described in Section 2.1, and in Sections 2.2–2.5 we have introduced five sets of hyperparameters, aiming to add some targeted sources of flexibility into the model. Furthermore, a hierarchical modeling approach has the ability to fix subsets of parameters, and to observe the effect on the others, offering additional robustness diagnostics.

The full posterior distribution of our model (adding a subscript $i$ for galaxies) is

$$p(\underbrace{\alpha, \beta, \omega, \gamma, \kappa}_{\text{hyper}}, \underbrace{\{\varphi_i, z_i\}}_{\text{latent}}) = \underbrace{p(\alpha, \beta, \omega, \gamma, \kappa)}_{\text{global prior}}$$

$$\prod_{i=1}^{N_{\text{obj}}} \underbrace{p(\varphi_i, z_i | \kappa)}_{\text{population model}} \prod_{b=1}^{N_{\text{bands}}} \underbrace{p(\hat{F}_{ib} | \underbrace{F_{ib}(\varphi_i, z_i, \alpha)}_{\text{SED model}}, \sigma_{ib}, \Sigma_{ib}^2(\varphi_i, z, \alpha, \gamma_b, \gamma), \omega_b)}_{\text{noise model}}. \tag{6}$$

## 3.2. Priors

In our approach it is straightforward to encode prior knowledge or constraints on corrections such as zero-points or emission-line contributions, as the prior on hyperparameters appears explicitly as $p(\alpha, \beta, \omega, \gamma, \kappa)$. The priors we use are given in Table 3.

We find that the Laplace prior for each $\alpha_j$ is preferable to a normal prior, since it promotes sparsity of the coefficients, i.e., the coefficients should be zero unless they significantly improve the model fits. Without this regularization there is a risk that line contributions would mimic other narrow features in the SED, such as absorption lines, which should instead be associated with a separate set of stellar evolution physics.

## 3.3. Optimization Strategy

Our inference strategy is to optimize the posterior distribution with respect to all parameters. This is because the mixture of broad bands (typically deeper) and noisier intermediate or narrow bands makes it possible to pin down SED features accurately and unequivocally. As a result, it is possible to derive estimates of the hyper- and latent parameters through optimization, without the need for more involved inference strategies, such as MCMC. We adopt the ADAM optimizer with a learning rate of $10^{-3}$.

Challenges that need to be overcome at this stage involve the computational needs of SPS calls, and the fact that they are not differentiable analytically (although differentiable SPS models are now being developed, Hearin et al. 2021; Alarcon et al. 2023). We address these issues with neural emulators of the model predictions, which are both $\sim10^4$ faster than native SPS calls and differentiable (Alsing et al. 2020). We return to the details of our neural emulators in the next section.

In principle, one could determine all parameters in a single run. But, for the purpose of this demonstration study, we adopt a two-stage approach (e.g., Laigle et al. 2016; Weaver et al. 2022) that is commonly used and leverages the availability of spectroscopic redshifts. We first determine the hyperparameters with redshifts fixed at the values acquired via spectroscopy (optimizing all SPS parameters other than the redshift). A second run is then performed by fixing the hyperparameters at the inferred values, and only optimizing the latent SPS parameters (now including the redshift). The resulting optimized redshifts are denoted $\hat{z}^{\text{MAP}}$ (omitting the galaxy index).

Finally, we employ bijectors to map the original parameters to new variables which have zero-mean, unit-variance normal priors. These are easier to work with than the original variables, since they do not have boundaries, which facilitates optimization (and sampling). The parameter bounds are shown explicitly in Table 1.

## 3.4. Emulators

As previously mentioned, one drawback associated with SPS models is the computational cost of each model call, which is

**Table 3**
Priors on Hyperparameters

| Hyperparameter | Prior |
|---|---|
| $\{\alpha_j\}_{j=1,\cdots,N_{\text{lines}}}$ | Laplace(0, 10) |
| $\{\beta_j\}_{j=1,\cdots,N_{\text{lines}}}$ | Normal(0, 1) |
| $\{\omega_b\}_{b=1,\cdots,N_{\text{bands}}}$ | Normal(1, 1) |
| $\{\gamma_b\}_{b=1,\cdots,N_{\text{bands}}}$ | Normal(0, 1) |

**Note.** The numbers in parentheses indicate the location and scale of the distribution.

the result of the many intensive numerical steps involved in predicting flux densities and photometry from stellar isochrones, initial mass function and the other ingredients of SPS. While this has been overcome with MCMC in the past and applied to medium-sized ($\sim10^5$ objects) data sets (e.g., Leja et al. 2019), photometric galaxy surveys are approaching the order of a billion objects (The LSST Dark Energy Science Collaboration et al. 2018; Abbott et al. 2022), so brute-force inference is not practical. Hyperparameters or model calibration would make the computational load even heavier. We now discuss how we employ emulators[11] of fluxes and line emission in order to overcome this challenge.

### 3.4.1. SPS Fluxes

For each band, we use the methodology described in Alsing et al. (2020) to train an emulator to predict the photometric flux, i.e., $\ell_{\text{SPS}}$ fed into Equation (3). While, in principle, one could train emulators to predict the $\ell_{\text{SPS}}(\lambda)$, it is advantageous to emulate photometric fluxes directly in order to avoid evaluating Equation (3). The emulator for each band is a fully connected 4-layer neural network (with 128 neurons in each layer). We train with staggered learning rates ($10^{-3}$, $10^{-4}$, $10^{-5}$, $10^{-6}$) and batch sizes (1000, 10,000, 50,000, all objects) to improve convergence. We train using 4 million samples from the prior, setting 10% of the training data aside for validation and stopping the training when the loss does not improve for 20 epochs.

### 3.4.2. Emission-line Fluxes

Our model photometry has additional contributions from emission lines, so for any parameter vector $(\varphi, z)$ we need to calculate them in addition to the base SPS predictions. This extra computational load can again be alleviated with emulators. However, in this case building emulators for individual emission lines and all bands would be prohibitive. Instead, we adopt a different strategy. We train an emulator to model the fluxes $\{\ell_j\}$ on the rest-frame wavelength grid $\{\lambda_j\}$,

---

[11] Note that emulators are not the only approach to accelerating inference. For example, Hahn & Melchior (2022) and Ramachandra et al. (2022) showed how to approximate posterior distributions directly for the estimation of SPS parameters or redshifts from broadband photometry.

following the procedure of Alsing et al. (2020). In our setting the size of the wavelength grid is the number of emission lines in FSPS (all 128 lines in Byler et al. 2017), so the accuracy of the emulator is easier to keep under control.[12] To predict the photometric flux of an emission line, we need to implement Equation (3). Because the bandpass $\{W_b(\lambda)\}$ are in tabular form, it would be preferable to work with an interpolated or continuous approximation. We find that fitting the sum of three generalized normal distributions to each $W_b$ offers a simple solution, since it is straightforward to fit to the tabulated data without any tuning. This approach delivers accurate and fast results for the final emission-line fluxes.

### 3.4.3. Emulator Accuracy

When comparing the initial and the emulator-reconstructed fluxes (computed from fluxes), we find that they are accurate at a level better than 1% (99th percentile of the magnitude differences in the range spanned by the COSMOS2020 data below). This will be comfortably covered by $\Sigma_b$, for example, thanks to the $\gamma_b$ hyperparameter being greater than 0.01. The accuracy of emulation of emission-line fluxes is typically below 0.1% (99th percentile), so the errors in this emulation can safely be ignored.

## 4. Data

We now construct a data set that demonstrates that our framework is able to produce accurate photometric redshift inferences from a benchmark sample. This consists of selecting broad- and narrow-band photometry in the COSMOS field, and cross-matching with a selection of spectroscopic data sets to enable validation of inferred redshifts.

### 4.1. Photometry

Weaver et al. (2022) presented COSMOS2020, the latest compilation of imaging in the COSMOS field. Two sets of photometric measurements were performed on these data: *Classic* and *Farmer*. For our case study we choose *Farmer*, which is based on the model-fitting software The Tractor.[13] The photo-$z$'s published with this data set were derived using EAZY (Brammer et al. 2008), as well as LEPHARE (Ilbert et al. 2006, 2008) using slightly different subsets of bands. While we will compare our results to both of these methods, we adopt the same subset of selected bands as EAZY; its SED templates are also derived using FSPS, thus providing a more direct comparison. This selection of bands exclude the Subaru Suprime-Cam broad bands (shallower than other filters at similar wavelengths) and the GALEX bands (shallow and with broad PSFs) (Weaver et al. 2022). We apply the recommended "combined" mask, which retains the deepest regions with the greatest number of available bands and removes areas corrupted by bright stars and artifacts in all the relevant bands.

This photometric data set is prepared for further analysis using the code released with the COSMOS2020 data, which applies various flux corrections (including Galactic extinction) and unit conversions.

### 4.2. Spectroscopic Redshifts

We compile the positions and spectroscopic redshifts from three large campaigns covering the COSMOS field: zCosmos-bright (Lilly et al. 2007), DEIMOS (Hasinger et al. 2018)], and C3R2 (Masters et al. 2017, 2019; Stanford et al. 2021).

zCosmos-bright (Lilly et al. 2007) is a highly representative sample of 20,000 bright ($i^* \leqslant 22.5$) galaxies. DEIMOS (Hasinger et al. 2018) is a sample of $\sim 4 \times 10^3$ objects selected from a variety of input catalogs based on multiwavelength observations in the field and thus represents a diverse selection function. Finally, the Complete Calibration of the Color-Redshift Relation (C3R2, Masters et al. 2017, 2019; Stanford et al. 2021) is a multi-institution, multi-instrument survey targeting faint galaxies ($i \sim 24.5$) populating regions of color space under-sampled by other surveys, in order to provide validation data for upcoming weak lensing surveys. We use data from the Data Releases 1, 2, and 3 ($\sim 4 \times 10^3$ objects).

For all three samples, we keep objects with quality flags 3 and 4. We perform a spatial cross-match between the positions of these objects in the spectroscopic and photometric catalogs. Specifically, we keep the nearest unique photometric source within a radius of $0''.6$ around each spectroscopic object, as recommended by Weaver et al. (2022). We also remove objects classified as stars by EAZY or LEPHARE, as well as the few sources with spectroscopic redshift >2.5.

We are left with 12,473 objects: 6920, 3168, and 2043 for zCosmos-bright, DEIMOS and C3R2, respectively. This is the data set upon which we carry out our demonstration study.

This data set combines three samples and so has a complicated selection function (i.e., the probability, conditional on its intrinsic properties, that a galaxy appears in the sample). The selection function must be incorporated into any population level analysis, as is treated explicitly in the inference of redshift distributions presented by Alsing et al. (2022). But if the focus is on inference of object-level parameters (i.e., individual galaxy redshifts) then selection effects can be ignored if the data on detected objects is sufficiently constraining; in the limit that the photometry determines the redshift (and other parameters) perfectly, the posterior is a delta function and hence is completely decoupled from the population. While not strictly satisfied here—the high-quality photometry and broad wavelength coverage do yield finite parameter uncertainties—this is an excellent approximation. We hence do not consider selection effects further.

## 5. Results

We now evaluate the performance of our inference framework applied to the crossmatched data set described in the previous section. In what follows, unless explicitly stated, the hyperparameters of the population model ($\boldsymbol{\kappa}$) are not optimized and are set to the fiducial values defined previously.

In practice, to avoid degeneracies with other parameters of the model, we anchor the zero-point of the HSC $i$ band to unity and define the others relative to this band.

### 5.1. Emission Lines

We perform an initial fit to the data with our model with no contribution from lines, i.e., setting $\alpha_j = \beta_j = 0 \,\forall j$. Thus, we optimize the zero-point offsets and uncertainty hyperparameters $\{\omega_b, \gamma_b\}_{b=1,\cdots,N_{\mathrm{bands}}}$, as well as the SPS parameters of each galaxy, with the procedure outlined in Section 3.3. This

---

[12] For reference, a standard FSPS spectrum with the default MILES spectral library has $\sim 6 \times 10^3$ wavelength points.
[13] https://github.com/dstndstn/tractor

consists of a run, with redshifts fixed at the spectroscopic redshifts, to estimate the hyperparameters, and a second run now relaxing redshifts but fixing the hyperparameters. The resulting maximum a posteriori redshift estimates, $\hat{z}^{\rm MAP}$, are poor compared to the released COSMOS2020 photo-$z$'s. We also observe sharp features in the redshift residuals $\hat{z}^{\rm MAP}-$ spec-$z$ and in the flux residuals (data minus model at the best-fit parameters) binned in spec-$z$. Many of these features correspond to strong emission lines, such as [O II], demonstrating that their luminosities are not correctly calibrated for our data set. Re-calibrating them is straightforward within our framework.

We must first determine a suitable subset of emission lines (out of the 128 implemented in FSPS by Byler et al. 2017) to include in the model. To achieve this we take the best-fit SEDs of our first run described above (at $z = $ spec-$z$) and calculate the contribution of each emission line to the total photometric flux, i.e., the flux ratio $L_{jb}/F_b$, for all objects, lines, and bands. We then take the average of these values over objects and bands. This procedure yields the lines which make the strongest contributions to the data at hand. This set includes most of the lines with intrinsic large variance according to Byler et al. (2017). Importantly, this procedure also filters out lines that are outside our wavelength coverage. We are left with 44 lines. Note that nearby emission lines are not necessarily all well-resolved by the data, i.e., each may only unequivocally contribute to a few objects. This could be improved with more informative priors on the lines or on line selection; however we do not expect this to have a significant impact on our redshift results.

In total, following this line-selection procedure, we have $2 \times N_{\rm bands} + 2 \times N_{\rm lines} = 140$ hyperparameters (excluding $\kappa$), and $15 \times N_{\rm objects} \sim 2 \times 10^5$ "latent" SPS parameters. We now optimize all of these parameters and extract maximum a posteriori redshift estimates $\hat{z}^{\rm MAP}$.

### 5.2. Photo-z's

We evaluate the performance of this model using some commonly used metrics. For every object in our catalog, a spectroscopic redshift $z_{\rm spec}$ and a redshift estimate $\hat{z}^{\rm MAP}$ are available. The main quantity of interest is the difference between the two, divided by the standard $1 + z$ factor accounting for the expected scaling of the quality of the estimates due to the effect of redshift on wavelength,

$$\Delta_z = \frac{z_{\rm spec} - \hat{z}^{\rm MAP}}{1 + z_{\rm spec}}. \tag{7}$$

We compute the mean, median, and standard deviation of $\Delta_z$ over our sample of galaxies. For the standard deviation, an estimator more robust to outliers than the sample variance is $\Sigma_{\rm NMAD}$, the median absolute deviation (the median of $|\Delta_z|$) multiplied by 1.48. Finally, the outlier fraction is defined as the percentage of objects with $|\Delta_z| > 0.15$.

Figure 2 shows these metrics, calculated in redshift and magnitude bins (in the reference HSC $i$ band, which is not affected by zero-points), for our framework, as well as the publicly released COSMOS2020 LEPHARE and EAZY redshifts.

Overall, our approach yields comparable levels of bias (as measured by the median of $\Delta_z$), fewer outliers, and a slightly larger level of scatter (as measured by $\sigma_{\rm NMAD}$). The
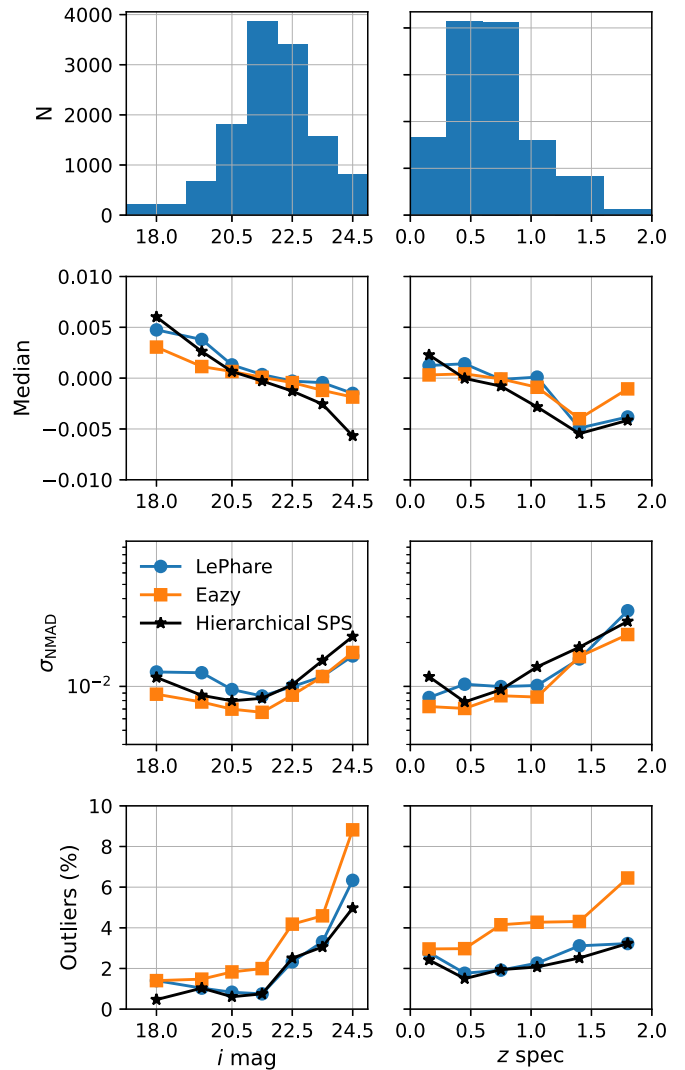


**Figure 2.** Photo-$z$ metrics calculated using our sample for our hierarchical model compared with the COSMOS2020 released LEPHARE and EAZY photo-$z$'s. We compute the median of $\Delta_z = (z_{\rm spec} - \hat{z}^{\rm MAP})/(1 + z_{\rm spec})$, as well as $\sigma_{\rm MAD} = 1.48 \times {\rm median}(|\Delta_z|)$, and the outlier fraction as the fraction of objects with $|\Delta_z| > 0.15$.

comparison with LEPHARE is not straightforward, as it uses a slightly different set of bands. On the other hand, we use the same set of bands as EAZY, also based on templates derived from a grid of parameters using an SPS model. EAZY includes a magnitude-redshift prior as well as a template error function (Brammer et al. 2008). These differences may be responsible for the small improvement in $\sigma_{\rm NMAD}$. Note that we have not attempted to tune or improve our model in order to carry out redshift inferences on these data, beyond the procedure described above.

While these photo-$z$ metrics are informative summaries of the results, it is also interesting to examine the distributions of redshift estimates themselves. The residuals $\Delta_z$ are shown in Figure 3, while Figure 4 also shows a conventional comparison of the estimates and the spectroscopic redshifts. These confirm the results captured by the metrics above. This also highlights that the redshift estimates are noisier for the DEIMOS and C3R2 samples because these surveys targeted populations typically under-represented in other spectroscopic campaigns
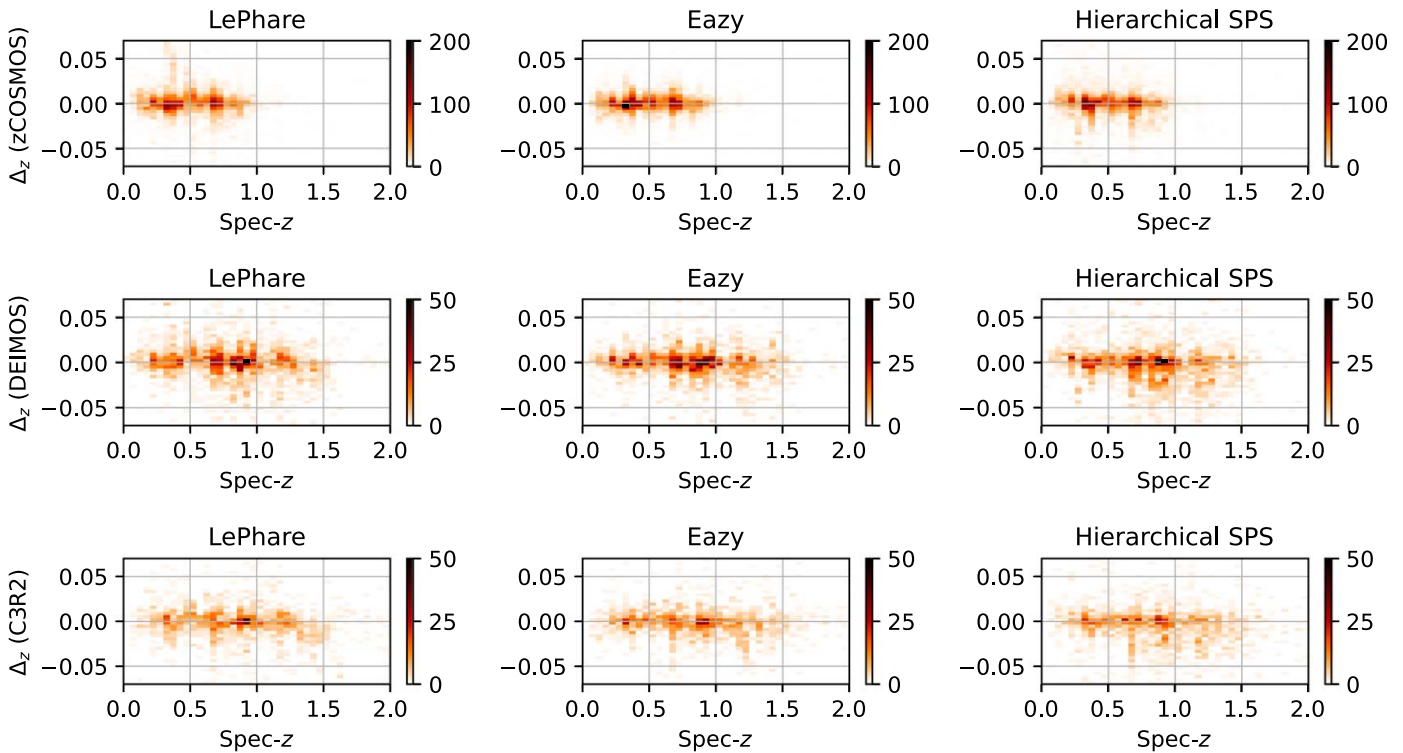
**Figure 3.** Differences between the maximum a posteriori redshift estimates and the spectroscopic redshifts (divided by $1 + z$) for the new model compared with the COSMOS2020 released LEPHARE and EAZY redshifts. This highlights the areas where they achieve similar performance in terms of scatter and outliers. Splitting between the three sources of spectroscopic redshifts also reveals the differences in photo-$z$ performance for samples with different selection, as described in Section 4.
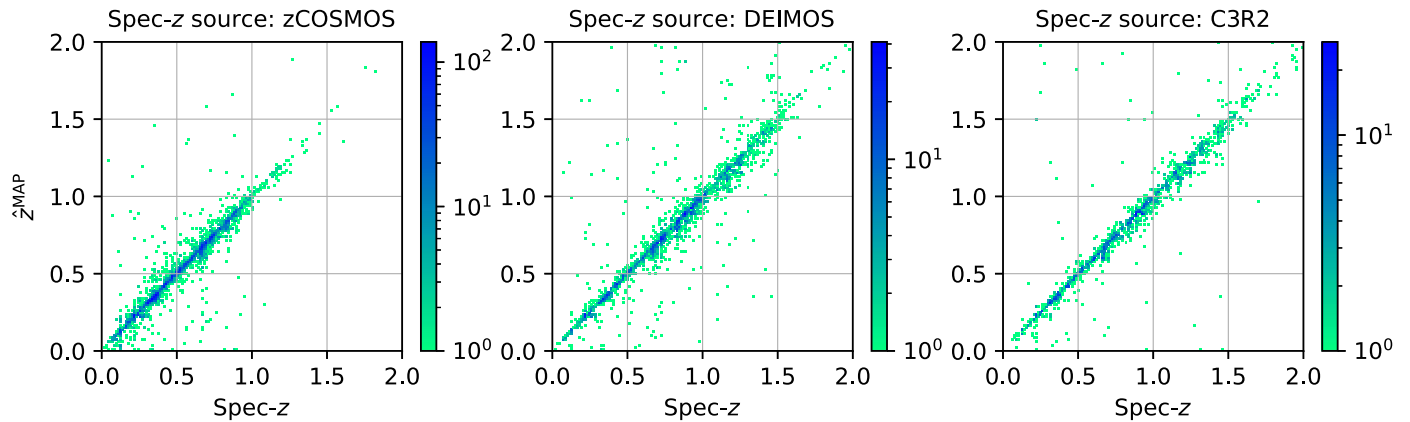


**Figure 4.** Maximum a posteriori redshift estimates versus spectroscopic redshifts for the three spectroscopic sources comprising the data we analyze. Differences are best highlighted in Figure 3, whereas these scatter plots with a logarithmic color bar better highlight the overall distributions and the outliers (which are rare).

(fainter, bluer, higher redshift). These populations also present greater challenges in terms of accurate photo-$z$ estimation.

### 5.3. Hyperparameters

In this section, we investigate the results of our optimization procedure for the model's hyperparameters and their implications for the robustness of the data model. The hyperparameters resulting from the optimization are shown in Figure 5. We also show the result of a run with the noise and emission-line offsets set to fiducial values $\gamma_b = 0.01 \, \forall b$, $\alpha_i = 0 \, \forall i$. Throughout this section we will also comment on robustness checks obtained by running variations of these models.

#### 5.3.1. Zero-point Offsets

The inferred zero-points $\{\omega_b\}$ are at the level of a few percent in most bands. They are well-constrained by the data and robust to changes in the model. The biggest changes are shifts in the intermediate bands caused by emission-line offsets $\alpha_i$, as shown in Figure 5. These results are qualitatively consistent with the EAZY zero-points released with COS-MOS2020. This is not a surprise given that we use the same set of bands and that EAZY also involves a grid of SPS templates. Some possible explanations for the disparities are differences in the SPS modeling, the approximation resulting from using a grid as opposed to a continuous model, or even the fact that EAZY calibrates the zero-points based on the whole
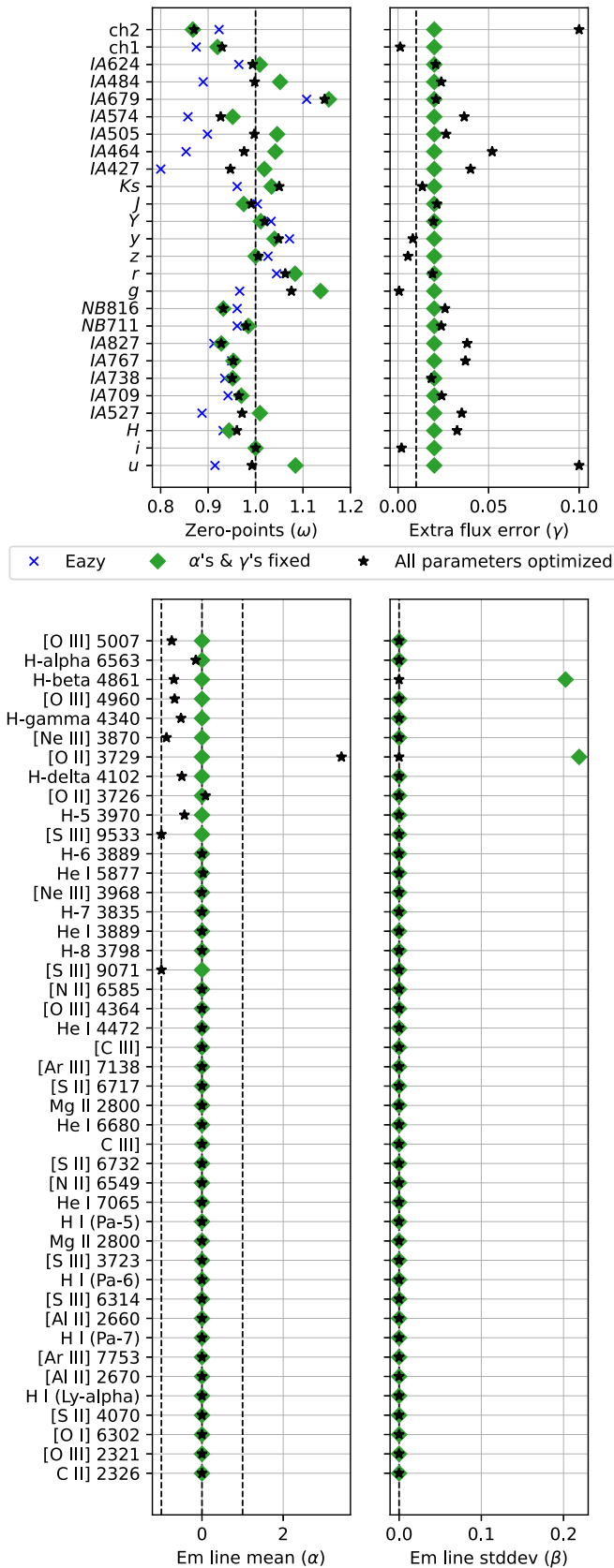
**Figure 5.** Hyperparameters obtained with the optimization procedure of Section 3.3. The black stars correspond to our best model, with all parameters optimized, while the green diamonds correspond to an additional model (used for robustness checks, see text for details) fixing the emission-line offsets $\alpha_j = 0 \ \forall j$ and error floors $\gamma_b = 0.01 \ \forall b$.

COSMOS2020 data set (not only objects with spectroscopic redshifts).

### 5.3.2. Examination of Photometric Uncertainties

Figure 5 shows that the inferred values of flux uncertainties $\gamma_b$ are at the level of a few percent in most bands: smaller for the broad bands (which are typically deeper), and larger for intermediate and narrow bands (noisier and also more sensitive to the details of the SED modeling, such as emission lines). The $u$ and ch2 bands are the most noisy, consistent with previous findings (Laigle et al. 2016; Weaver et al. 2022). Just as for zero-points, the inferred values are mostly stable when switching other hyperparameters on and off. However, there is no uncertainty contributed by emission lines in this model (bottom right panel), indicating that the band-to-band uncertainty is preferred by the data. If we set the band-to-band uncertainty to 1% (green diamonds), then emission lines contribute to the uncertainty budget (via non-zero $\beta$) to compensate, as expected. Thus, we conclude that the noise model performs well, but that emission-line uncertainty is simply not required in our full model for this data set.

### 5.3.3. Emission-line Offsets

We find that including extra contributions from emission lines significantly improves the quality of the derived photo-$z$'s. This is consistent with previous findings by Ilbert et al. (2006, 2008), Brammer et al. (2008) and Alarcon et al. (2021). Note that in Figure 5 emission lines are ordered by their contribution relative to the total broadband flux, following the procedure described above. Offsets are primarily given to the strongest lines, with some of them being almost entirely canceled ($\alpha \approx -1$). The Laplace prior is effective at keeping most line offsets at zero, unless they bring a significant improvement to the fits. We also note that the interpretation of the hyperparameters is conditioned on the sample at hand: changing the subset of COSMOS2020 galaxies analyzed may change the resulting values of zero-points and emission-line offsets, for example.

### 5.4. Population Model

We perform one additional run, also optimizing the hyperparameters of the population model, $\kappa$. We find that the data do not significantly perturb the fiducial values originally adopted; nor do these parameters appear degenerate with other parameters of the model. This demonstrates the robustness of the model to the prior and additionally shows that varying these hyperparameters does not lead to significantly improved fits. However, it should be noted that we have not included selection effects in our model. If we had found that the inferred values had significantly moved from the fiducial values, we would have been unable to determine whether this had been caused by incomplete modeling of the selection process (Alsing et al. 2022). However in this setting, as we have previously argued, the selection effects should be negligible so this concern does not arise.

### 5.5. Residuals

Finally, we check the ability of the presented model to explain the statistical properties of the data. We examine the
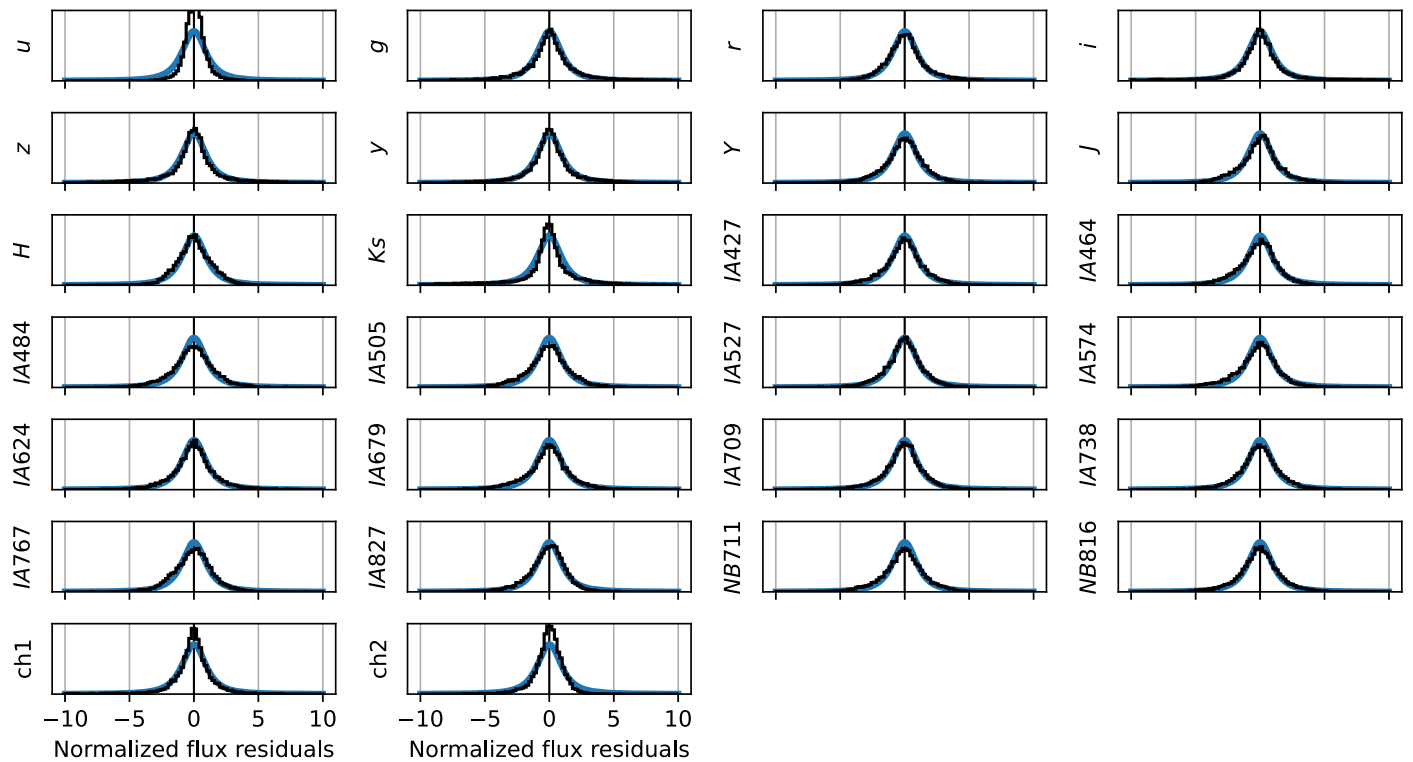
**Figure 6.** Flux residuals (black), as defined in Equation (8), compared with the likelihood function of Equation (4) in blue, showing a satisfactory agreement between the data and the model predictions.

distribution of the flux residuals, defined as

$$r_b = \frac{\hat{F}_b - \omega_b F_b}{\sqrt{\sigma_b^2 + (\omega_b \Sigma_b)^2}}. \tag{8}$$

These are shown in Figure 6. The blue lines show our likelihood function, the Student's-t distribution described in Equation (4). There is good agreement between the chosen likelihood and the residuals. We have checked that this agreement significantly deteriorates when all of the hyperparameters are not optimized (not shown here). The deviations in the $u$ and ch2 bands are likely to be due to residual data systematics which are not well described by our model.

When repeating this analysis with a normal likelihood function, we find that the residuals are worse and provide a poor fit of the likelihood function, despite optimizing the other components of the model. This setting also results in worse redshift estimates. This demonstrates the presence of outliers and shows that accommodating them with a suitable likelihood function can significantly improve the ability of a model to fit complicated data.

## 6. Discussion

Our results demonstrate that the hierarchical SPS inference framework presented here (and in Alsing et al. 2022) delivers accurate redshifts on a benchmark data set, therefore passing a stringent validation test. We now discuss possible extensions which relax some of its assumptions.

The model itself could be made more sophisticated in various ways in order to better suit specific applications and increase the accuracy of the photo-$z$'s further. Such extensions include a stellar model, which could straightforwardly be incorporated to allow star-galaxy separation, akin to what is done in EAZY and LEPHARE with stellar templates.

In the future, it will be critical to include effects which are currently neglected, but which affect the data (and in turn, photometric redshifts) at significant levels for LSST, such as airmass-dependent fluxes (Graham et al. 2018). Furthermore, extending our model to include spatial information (Sánchez & Bernstein 2018; Jasche & Lavaux 2019; Alarcon et al. 2020; Sánchez et al. 2020) could provide a powerful probe of the connection between galaxies and dark matter and exploit more of the information offered by photometric surveys.

Our chosen optimization strategy, paired with the posterior distribution taken as a loss function, minimizes the residuals between the measured and the model fluxes. Other choices may be possible, especially when including spectroscopic redshifts. For example, one could explicitly minimize the residuals between photometric (inferred) and spectroscopic redshifts, although this will render the results more sensitive to selection effects in the data.

We neglected the uncertainties in hyperparameters and SPS parameters. This is because the extensive wavelength coverage of the 26-band COSMOS2020 data, the small flux uncertainties, and the availability of spectroscopic redshifts combine to yield well-constrained galaxy SEDs and in turn hyperparameters. Nevertheless, there are fundamental degeneracies expected between some of the SPS parameters (Leja et al. 2017, 2019), as well as with the redshift when it is not fixed to a spectroscopic value. Selection effects may also be important, and can be accounted for in the posterior distributions as described in Alsing et al. (2022). In analyzing data sets where the uncertainties in hyperparameters or SPS parameters (including redshift) are not negligible (e.g., broadband-only surveys like the LSST), one can resort to the variety of

Bayesian inference techniques (i.e., MCMC, variational inference, Laplace approximation, simulation-based inference) available for hierarchical models. Since the impact of selection cuts and parameter uncertainties is strongly dependent on the specific data and models being considered, we defer further discussion and the development of effective inference strategies to future work.

## 7. Conclusion

We have presented a hierarchical model to infer redshift and other intrinsic galaxy properties from photometric data. This approach makes it possible to self-consistently encode knowledge of (1) galaxy formation and evolution via a population model, (2) astrophysics via SEDs predicted with stellar population synthesis, and (3) observational effects via a noise model and a likelihood function. By formulating the inference within a Bayesian hierarchical model, we are able to parameterize additional sources of bias and uncertainty in each of these components and to solve for them self-consistently. Thus, any manual tuning or inversion methods can be avoided, with the added benefit that it is possible to set informative priors on these additional terms.

Our photo-$z$'s (as measured with traditional metrics on redshift point estimates) are competitive with the publicly released photo-$z$'s from COSMOS2020 (Weaver et al. 2022). While this is a benchmark test on a relatively bright, low-redshift sample of spectroscopically confirmed galaxies, these results demonstrate the power and flexibility of our methodology and motivate its further development toward addressing the challenging setting presented by stage IV cosmological surveys. We made our inference tractable using neural emulators to accelerate model calls (Alsing et al. 2020).

In future applications of our framework, we will relax some of the modeling assumptions made here. In particular, we demonstrate how to incorporate selection in the companion paper Alsing et al. (2022). Comparing inferences from multiple SPS and population models will also provide valuable consistency checks. In fact, refining the SED and population models directly from the upcoming survey data is a promising avenue for directly probing the underlying (astro)physics using large amounts of untapped statistical power. Leveraging hybrid data sets and incorporating additional data available (e.g., spec-$z$'s, or even full spectra) will be instrumental for correctly separating the different physical effects at play. This is straightforwardly included in hierarchical models (see, e.g., Alarcon et al. 2021; Nagaraj et al. 2022). Together, these attributes make the methodology presented here well-suited for addressing the challenges of upcoming photometric surveys and the robust scientific exploitation of their data.

## Author Contributions

B.L.: conceptualization, methodology, software, validation, formal analysis, writing—original draft. J.A.: conceptualization, methodology, software, validation, writing—review and editing. H.V.P.: conceptualization, methodology, validation, writing—review and editing, funding acquisition. D.M.: conceptualization, methodology, validation, writing—review and editing, funding acquisition. J.L.: conceptualization, methodology, validation, writing—review and editing.

## ORCID iDs

Boris Leistedt ⓘ https://orcid.org/0000-0002-3962-9274
Justin Alsing ⓘ https://orcid.org/0000-0003-4618-3546
Hiranya Peiris ⓘ https://orcid.org/0000-0002-2519-584X
Daniel Mortlock ⓘ https://orcid.org/0000-0002-0041-3783
Joel Leja ⓘ https://orcid.org/0000-0001-6755-1315

## References

Abbott, T., Aguena, M., Alarcon, A., et al. 2022, PhRvD, 105, 023520
Acquaviva, V., Raichoor, A., & Gawiser, E. 2015, ApJ, 804, 8
Ade, P. A., Aghanim, N., Arnaud, M., et al. 2016, A&A, 594, A13
Alarcon, A., Gaztanaga, E., Eriksen, M., et al. 2021, MNRAS, 501, 6103
Alarcon, A., Hearin, A. P., Becker, M. R., & Chaves-Montero, J. 2023, MNRAS, 518, 562
Alarcon, A., Sánche, C., Bernstein, G. M., & Gaztañaga, E. 2020, MNRAS, 498, 2614
Alsing, J., Leistedt, B., Leja, J., Peiris, H., & Efstathiou, D. M. G. 2022, ApJS, in press
Alsing, J., Peiris, H., Leja, J., et al. 2020, ApJS, 249, 5
Battisti, A. J., da Cunha, E., Grasha, K., et al. 2019, ApJ, 882, 61
Benitez, N. 2000, ApJ, 536, 571
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Byler, N. 2018, nell-byler/cloudyfsps: Initial release of cloudyFSPS, v1.0.0, Zenodo, doi:10.5281/zenodo.1156412
Byler, N., Dalcanton, J. J., Conroy, C., & Johnson, B. D. 2017, ApJ, 840, 44
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
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
Collister, A. A., & Lahav, O. 2004, PASP, 116, 345
Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833
Conroy, C., Gunn, J. E., & White, M. 2009a, ApJ, 699, 486
Conroy, C., White, M., & Gunn, J. E. 2009b, ApJ, 708, 58
Draine, B. T., & Li, A. 2007, ApJ, 657, 810
Foreman-Mackey, D., Sick, J., & Johnson, B. 2014, python-fsps: Python bindings to FSPS (v0.1.1), Zenodo, doi:10.5281/zenodo.12157
Gatti, M., Giannini, G., Bernstein, G. M., et al. 2021, MNRAS, 510, 1223
Graham, M. L., Connolly, A. J., Ivezić, Ž, et al. 2018, AJ, 155, 1
Hahn, C., & Melchior, P. 2022, ApJ, 938, 11
Hartley, W. G., Chang, C., Samani, S., et al. 2020, MNRAS, 496, 4769
Hasinger, G., Capak, P., Salvato, M., et al. 2018, ApJ, 858, 77
Hearin, A. P., Chaves-Montero, J., Alarcon, A., Becker, M. R., & Benson, A. 2021, arXiv:2112.06830
Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69

Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Ilbert, O., Capak, P., Salvato, M., et al. 2008, ApJ, 690, 1236
Jasche, J., & Lavaux, G. 2019, A&A, 625, A64
Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, ApJS, 254, 22
Jones, D. M., & Heavens, A. F. 2018, MNRAS, 483, 2487
Kind, M., & Brunner, R. J. 2013, MNRAS, 432, 1483
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
Leistedt, B., Hogg, D. W., Wechsler, R. H., & DeRose, J. 2019, ApJ, 881, 80
Leistedt, B., Mortlock, D. J., & Peiris, H. V. 2016, MNRAS, 460, 4258
Leja, J., Johnson, B. D., Conroy, C., et al. 2019, ApJ, 877, 140
Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., & Byler, N. 2017,
    ApJ, 837, 170
Leja, J., Speagle, J. S., Johnson, B. D., et al. 2020, ApJ, 893, 111
Leja, J., Speagle, J. S., Ting, Y.-S., et al. 2022, ApJ, 936, 165
Lilly, S. J., Fevre, O. L., Renzini, A., et al. 2007, ApJS, 172, 70
Madau, P. 1995, ApJ, 441, 18
Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2017, ApJ, 841, 111
Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, ApJ, 877, 81
McQuinn, M., & White, M. 2013, MNRAS, 433, 2857
Ménard, B., Scranton, R., Schmidt, S., et al. 2013, arXiv:1303.4722
Morrison, C. B., Hildebrandt, H., Schmidt, S. J., et al. 2017, MNRAS,
    467, 3576
Myles, J., Alarcon, A., Amon, A., et al. 2021, MNRAS, 505, 4249
Nagaraj, G., Forbes, J. C., Leja, J., Foreman-Mackey, D., & Hayward, C. C.
    2022, ApJ, 932, 54
Nenkova, M., Sirocky, M. M., Ivezić, Ž., & Elitzur, M. 2008, ApJ,
    685, 147
Newman, J. A. 2008, ApJ, 684, 88
Newman, J. A., & Gruen, D. 2022, ARA&A, 60, 363
Ramachandra, N., Chaves-Montero, J., Alarcon, A., et al. 2022, MNRAS,
    515, 1927
Rau, M. M., Wilson, S., & Mandelbaum, R. 2019, MNRAS, 491, 4768
Sánchez, C., & Bernstein, G. M. 2018, MNRAS, 483, 2801
Sánchez, C., Raveri, M., Alarcon, A., & Bernstein, G. M. 2020, MNRAS,
    498, 2984
Schmidt, S. J., Ménard, B., Scranton, R., Morrison, C., & McBride, C. K. 2013,
    MNRAS, 431, 3307
Schneider, M., Knox, L., Zhan, H., & Connolly, A. 2006, ApJ, 651, 14
Stanford, S. A., Masters, D., Darvish, B., et al. 2021, ApJS, 256, 9
Tanaka, M. 2015, ApJ, 801, 20
The LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T.,
    et al. 2018, arXiv:1809.01669
Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, ApJS, 258, 11
Whitler, L., Stark, D. P., Endsley, R., et al. 2022, arXiv:2206.05315