The Astrophysical Journal Supplement Series, 264:29 (18pp), 2023 February https://doi.org/10.3847/1538-4365/ac9583

Forward Modeling of Galaxy Populations for Cosmological Redshift Distribution Inference

Justin Alsing[1], Hiranya Peiris[1,2], Daniel Mortlock[1,3,4], Joel Leja[5,6,7], and Boris Leistedt[3]
[1] Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Stockholm SE-106 91, Sweden; justin.alsing@fysik.su.se
[2] Department of Physics and Astronomy, University College London, Gower Street, London, WC1E 6BT, UK
[3] Department of Physics, Imperial College London, Blackett Laboratory, Prince Consort Road, London, SW7 2AZ, 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 16802, USA
[7] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
Received 2022 July 12; revised 2022 September 16; accepted 2022 September 18; published 2023 January 18

Abstract

We present a forward-modeling framework for estimating galaxy redshift distributions from photometric surveys. Our forward model is composed of: a detailed population model describing the intrinsic distribution of the physical characteristics of galaxies, encoding galaxy evolution physics; a stellar population synthesis model connecting the physical properties of galaxies to their photometry; a data model characterizing the observation and calibration processes for a given survey; and explicit treatment of selection cuts, both into the main analysis sample and for the subsequent sorting into tomographic redshift bins. This approach has the appeal that it does not rely on spectroscopic calibration data, provides explicit control over modeling assumptions and builds a direct bridge between photo-z inference and galaxy evolution physics. In addition to redshift distributions, forward modeling provides a framework for drawing robust inferences about the statistical properties of the galaxy population more generally. We demonstrate the utility of forward modeling by estimating the redshift distributions for the Galaxy And Mass Assembly (GAMA) survey and the Vimos VLT Deep Survey (VVDS), validating against their spectroscopic redshifts. Our baseline model is able to predict tomographic redshift distributions for GAMA and VVDS with respective biases of $\Delta z \lesssim 0.003$ and $\Delta z \simeq 0.01$ on the mean redshift—comfortably accurate enough for Stage III cosmological surveys—without any hyperparameter tuning (i.e., prior to doing any fitting to those data). We anticipate that with additional hyperparameter fitting and modeling improvements, forward modeling will provide a path to accurate redshift distribution inference for Stage IV surveys.

Unified Astronomy Thesaurus concepts: Redshift surveys (1378); Galaxy photometry (611); Galaxy stellar content (621); Galaxy evolution (594); Cosmological parameters from large-scale structure (340); Gravitational lensing (670); Weak gravitational lensing (1797)

1. Introduction

Accurate inferences of the redshift distributions of ensembles of galaxies from their photometry are of central importance for deriving cosmological constraints from weak-lensing surveys. Ongoing and upcoming surveys—such as the Dark Energy Survey (Flaugher 2005), the Kilo-Degree Survey (De Jong et al. 2015), the Hyper Suprime-Cam (Aihara et al. 2018), the Vera C. Rubin Observatory's Legacy Survey of Space and Time (Abell et al. 2009), and Euclid (Laureijs et al. 2011)—will map large volumes of the universe, measuring the angular positions, images, and photometry for billions of galaxies. The unprecedented statistical power of these surveys will make them increasingly sensitive to systematic biases, with the systematics on the redshift distributions in particular being expected to constitute the single largest contributor to the total systematic error budget (see, e.g., Hildebrandt et al. 2017). In order to reach the full potential of the upcoming Stage IV surveys, the characterizations of the redshift distributions (e.g., their measured means and variances) will need to improve by roughly an order of magnitude, compared to the current state of the art (Newman & Gruen 2022).

Three main approaches exist for estimating cosmological redshift distributions: cross correlation (Schneider et al. 2006; Newman 2008; McQuinn & White 2013; Ménard et al. 2013; Schmidt et al. 2013; Morrison et al. 2017; Davis et al. 2018), direct calibration (Lima et al. 2008; Hildebrandt et al. 2016, 2020; Buchs et al. 2019; Wright et al. 2020), and template-based methods (Benitez 2000; Ilbert et al. 2006; Brammer et al. 2008; Arnouts & Ilbert 2011; Hildebrandt et al. 2012; Leistedt et al. 2016, 2019; Hoyle et al. 2018; Tanaka et al. 2018).

Cross correlation and direct calibration both rely on spectroscopic redshift samples, to compare against the photometric data, in order to calibrate their redshift distributions. These approaches have the appeal that they leverage reliably measured redshifts to calibrate the photo-z distributions, and are relatively insensitive to modeling assumptions about the photometric data. On the other hand, they are limited by the lack of available spectroscopic redshifts at the depths probed by ongoing and upcoming surveys, and are vulnerable to biases due to spectroscopic selection effects that are not well represented by reweighting in the (broadband) colors (Buchs et al. 2019; Hartley et al. 2020), as well as the uncertainties in galaxy bias modeling, in the case of cross-correlation methods (Gatti et al. 2018).

Template-based methods instead rely on building a statistical model for the photometric observations, in order to constrain the redshifts of individual galaxies, and the redshift distributions of ensembles, from their photometry alone. Template approaches assert that galaxies belong to one of a finite set of "types," where each type has an associated (rest-frame) spectral template that defines its colors as a function of redshift. These templates can then be compared to the observed colors, to give the redshift (and type) likelihoods for each galaxy, which can then be combined in a hierarchical model to infer the redshift distributions of ensembles of galaxies (Leistedt et al. 2016, 2019). However, template methods are limited by the inability of template sets (and priors over galaxy types) to characterize the galaxy population at the necessary level of realism. Finite template sets tend to be overly restrictive, while methods that allow for linear combinations of templates result in large swathes of prior volumes being populated by unphysical galaxy spectra. Further, by relying on models for the photometric data, template-based models need to treat selection effects explicitly in order to draw robust inferences at the population level (e.g., redshift distributions): this has so far not been done.

Physically, the redshift distributions of selected samples of galaxies arise from a sequence of three main processes. The statistical properties of the galaxy population (i.e., the intrinsic distributions of physical characteristics and redshifts) define an intrinsic distribution for galaxy colors, from which galaxies in the universe are sampled. The colors (photometry) of those galaxies in some patch of the sky then get observed by a survey, resulting in a catalog of noisy (measured) photometry. Selection cuts are then applied to those measurements, to ensure a clean and high-quality galaxy sample and to sort the galaxies into tomographic redshift bins. The redshift distributions of interest, then, are those of the galaxies that make it past the selection cuts and into a given tomographic bin. Therefore, if one is able to accurately characterize the galaxy population, observational processes, and selection effects, one can predict the redshift distributions of interest.

In this paper, we develop a forward-modeling framework for estimating redshift distributions by explicitly modeling the processes that give rise to them. We construct a population model describing the joint distribution of the physical characteristics (e.g., the stellar, dust, and gas contents) of galaxies, encoding galaxy evolution physics in the relationships between the physical properties of the galaxies. We use a stellar population synthesis (SPS) model to connect those physical parameters to the rest-frame spectra and, hence, the photometry for each galaxy. The observation process is then characterized by a data model that captures the measurement noise, heterogeneous observing conditions and strategies, and photometric calibration. Finally, we have a selection model specifying the selection cuts. This parameterized forward model then forms the basis for the Bayesian inference of cosmological redshift distributions, either by hierarchical inference or by simulation-based inference (SBI).

This forward-modeling approach can be thought of as resolving the current limitations of template-based methods, by replacing template sets with a continuous SPS model, by explicitly treating selection effects, and by inferring population and data model parameters in a self-consistent fashion. In particular, by replacing finite template sets with a continuous model, we are able to better capture the diversity of real galaxy spectra, while having full control over the priors describing the statistical properties of the galaxy population. The use of SPS models in analyzing large samples of galaxies has only recently become feasible, thanks to fast neural emulators (e.g., `speculator`; Alsing et al. 2020). The use of physically motivated priors (e.g., Tanaka 2015 and Ramachandra et al. 2022) and continuous physical models for galaxy spectra (Ramachandra et al. 2022) has already led to promising improvements in photometric redshift inferences for individual galaxies.

The structure of this paper is as follows. In Section 2, we describe the frameworks for forward-modeling photometric surveys and estimating the redshift distributions in forward-modeling contexts. In Section 3, we describe our SPS model, galaxy population model, and data model assumptions. In Sections 4 and 5, we show the ability of our baseline forward model to recover tomographic redshift distributions for the Galaxy And Mass Assembly (GAMA) survey and the Vimos VLT Deep Survey (VVDS), respectively. We outline a roadmap for future forward-modeling efforts and conclude in Section 7.

In a companion paper (B. Leistedt et al. 2023, in preparation) we validate the forward-modeling framework for inferring individual galaxy redshifts, including the hierarchical calibration of hyperparameters.

2. Forward Modeling of Galaxy Surveys for Redshift Inference

In this section, we describe our generative modeling framework for photometric surveys. We frame the forward model as a pipeline for simulating mock catalogs, which can then be compared to the observed catalog in an SBI setting, and either used to estimate the implied tomographic redshift distributions for a given set of modeling and hyperparameter choices or used as a basis for hierarchical inference, via Markov Chain Monte Carlo (MCMC) sampling.

The notation is summarized in Section 2.1 and Table 1. We describe the generative model in Section 2.2, and discuss how to perform the inference of the model parameters in Section 2.4.

2.1. Notation

Each galaxy is described by a set of SPS parameters $\boldsymbol{\varphi}$, which describe the stellar, gas, and dust contents of the galaxy. The rest-frame spectrum $l(\lambda) \equiv l(\lambda; \boldsymbol{\varphi})$ is connected to the parameters $\boldsymbol{\varphi}$ via an SPS model, which computes a composite spectrum from the stars in the population, given their initial mass functions, ages, and metallicities, from their star formation and metallicity histories, plus modifications due to dust as well as nebular emission (see Conroy 2013 for a review).

Combined with the redshifts z, the SPS parameters predict the model photometry for each galaxy, i.e., the fluxes $\{f_b\}$ in the bandpasses $\{W_b(\lambda)\}$, defined by

$$f_b(\boldsymbol{\varphi}, z) = \frac{(1+z)^{-1}}{4\pi d_L^2(z)} \int_0^\infty l(\lambda/(1+z); \boldsymbol{\varphi}) e^{-\tau(z,\lambda)} W_b(\lambda) d\lambda, \quad (1)$$

where $d_L(z)$ is the luminosity distance for the redshift z and $\tau(z, \lambda)$ is the optical depth of the intergalactic medium.

We denote the vector of the measured fluxes for each galaxy with $\boldsymbol{d}$, where the photometric data model specifies the sampling distribution $P(\boldsymbol{d}|\boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta})$ of the measured fluxes,

Table 1
Notation for All the Model Parameters Included in the Forward Model

Parameter	Description
	Population Model Parameters
ψ	Hyperparameters describing the galaxy population model
Φ_0	Present-day comoving volume density of galaxies
$\rho(z;\psi)$	Evolution in the relative comoving number density of galaxies ($\rho(0;\psi)=1$)
	Data Model Parameters
$\boldsymbol{\eta}$	Nuisance parameters determining the data model
$\mathcal{O}$	Parameters governing the distribution of photometric uncertainties
	Latent Parameters
$\boldsymbol{\varphi}_{1:N}$	Stellar population parameters describing the rest-frame spectrum (per galaxy)
$z_{1:N}$	Redshift (per galaxy)
	Derived Quantities
$\bar{N}=\bar{N}(\Phi_0,\psi,\boldsymbol{\eta},\mathcal{O})$	Expected number of selected galaxies, given the population, selection, and data models
$\boldsymbol{f}_{1:N}=\boldsymbol{f}(\boldsymbol{\varphi}_{1:N}, z_{1:N})$	Model fluxes determined by the SPS model (per galaxy)
	Data
$\boldsymbol{d}_{1:N}$	Data vector of measured fluxes (per galaxy)
$\boldsymbol{\sigma}_{1:N}$	Flux measurement uncertainties (per galaxy)
N	Observed number of selected galaxies
	Selection
$S_{1:N}$	Selection into the sample based on photometric cuts (per galaxy)
$S^{(k)}_{1:N}$	Color-based selection into tomographic bin k (per galaxy)

given the model fluxes and measurement uncertainties $\boldsymbol{\sigma}$. The data model is parameterized by the nuisance parameters $\boldsymbol{\eta}$, which characterize the properties of the noise distribution, calibration (e.g., zero-point) parameters, and modeling error terms.

The intrinsic distribution of the galaxy characteristics (the SPS parameters and redshift) is described by a population model $P(\boldsymbol{\varphi}, z|\psi)$, with hyperparameters ψ. This describes the joint distribution of the galaxy characteristics for the background galaxy population, in the absence of any selection effects.

We assume that the selection cuts for the main galaxy sample are made using observed photometry, with a selection probability $P(S|\boldsymbol{d},\boldsymbol{\sigma})$ equal to one or zero for selection or rejection., i.e., selection is deterministic, given the photometric data vector.

The subsequent sorting of galaxies into tomographic bins introduces an additional selection $S^{(k)}$ (for the kth bin), based on the measured galaxy colors (fluxes). Typically, tomographic binning will be based on some estimator for the redshift, which will be a deterministic function of the measured fluxes and their uncertainties:

$$P(S^{(k)}|\hat{z}(\boldsymbol{d},\boldsymbol{\sigma})) = \begin{cases} 1 & z_{\rm l}^{(k)} < \hat{z}(\boldsymbol{d},\boldsymbol{\sigma}) < z_{\rm u}^{(k)} \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where $\hat{z}(\boldsymbol{d},\boldsymbol{\sigma})$ is an estimator for the redshift, with $z_{\rm l}^{(k)}$ and $z_{\rm u}^{(k)}$ being estimators for the lower and upper limits for the kth tomographic bin.

The measurement uncertainties on the photometry will depend on the observing conditions (e.g., seeing) and strategy (e.g., exposure time), which will typically vary to some extent across the survey. The uncertainties will also scale with flux, owing to the Poisson photon count contribution to the errors, and have additional intrinsic scatter, due to the varying difficulty in extracting fluxes from galaxy images with different morphologies. We denote the distribution of the photometric uncertainties across the survey (as a function of flux) with the uncertainty model $P(\boldsymbol{\sigma}|\boldsymbol{f}(\boldsymbol{\varphi}, z), \mathcal{O})$, where $\mathcal{O}$ denotes the model assumptions (and parameters) characterizing that distribution.

2.2. Generative Model

The forward model proceeds as follows:

1. Draw SPS parameters $\boldsymbol{\varphi}$ and redshifts z from the population model $P(\boldsymbol{\varphi}, z|\psi)$;
2. Compute the model photometry, given the SPS parameters, using the SPS model, $\boldsymbol{f}\equiv\boldsymbol{f}_{\rm SPS}(\boldsymbol{\varphi}, z)$;
3. Draw uncertainties from the uncertainty model $P(\boldsymbol{\sigma}|\boldsymbol{f}, \mathcal{O})$;
4. Draw noisy (calibrated) photometry $\boldsymbol{d}$, given the model fluxes and uncertainties, from the data model $P(\boldsymbol{d}|\boldsymbol{f}(\boldsymbol{\varphi}, z), \boldsymbol{\sigma}, \eta)$;
5. Apply selection cuts on the noisy photometry, where $P(S|\boldsymbol{d},\boldsymbol{\sigma})$ equals one (zero) for passing (failing) selection; and
6. Assign a tomographic bin label, based on the photo-z estimator, $\hat{z}(\boldsymbol{d},\boldsymbol{\sigma})$.

The result is a catalog of noisy photometry for selected galaxies, with associated tomographic bin labels.

Repeating the above process until one obtains the same number N of selected objects as in the observed catalog provides a draw from the assumed generative model for the data conditioned on N selected objects.[8] This can therefore be used as a generative model for inferring the hyperparameters via SBI, or as the basis for Bayesian hierarchical inference, as described below in Section 2.4.

Repeating this process in the limit of $N\to\infty$ selected samples, then examining the redshift distributions of the galaxies that make it into each bin, provides the tomographic redshift distributions that are implied for a given set of modeling and hyperparameter assumptions. Note that the target redshift distributions are given by a (typically intractable) integral over the population, data, and selection models. The tomographic redshift distribution $n_k(z)$ for the galaxies passing selection both into the analysis sample, and subsequently into the kth tomographic bin, is given by:

$$\begin{aligned} n_k(z) &\equiv P(z|S^{(k)},\psi,\boldsymbol{\eta},\mathcal{O}) \\ &= \int P(\boldsymbol{\varphi}, z|\psi, S^{(k)},\boldsymbol{\sigma},\boldsymbol{\eta}) \\ &\quad\times P(\boldsymbol{\sigma}|S^{(k)},\psi,\boldsymbol{\eta},\mathcal{O})d\boldsymbol{\varphi}d\boldsymbol{\sigma} \\ &= \int \frac{P(\boldsymbol{\varphi}, z|\psi)P(S^{(k)}|\boldsymbol{\varphi}, z,\boldsymbol{\sigma},\boldsymbol{\eta})}{P(S^{(k)}|\psi,\boldsymbol{\sigma},\boldsymbol{\eta})} \\ &\quad\times P(\boldsymbol{\sigma}|S^{(k)},\psi,\boldsymbol{\eta},\mathcal{O})d\boldsymbol{\varphi}d\boldsymbol{\sigma} \\ &= \int \frac{P(\boldsymbol{\varphi}, z|\psi)P(S^{(k)}|\boldsymbol{d},\boldsymbol{\sigma})}{P(S^{(k)}|\psi,\boldsymbol{\sigma},\boldsymbol{\eta})}P(\boldsymbol{d}|\boldsymbol{\varphi}, z,\boldsymbol{\sigma},\boldsymbol{\eta}) \\ &\quad\times P(\boldsymbol{\sigma}|S^{(k)},\psi,\boldsymbol{\eta},\mathcal{O})d\boldsymbol{d}\,d\boldsymbol{\varphi}d\boldsymbol{\sigma}, \end{aligned} \tag{3}$$

where $S^{(k)}$ denotes selection into both the analysis sample and the kth tomographic bin. Forward simulating from the

[8] Note that this is implicitly marginalized over the expected present-day volume number density of the galaxies; see Appendix A.

generative model described above provides a way of estimating the target redshift distributions for a given set of hyperparameters, without the need for direct integration.

2.3. Emulation of SPS Models

Any applications of this forward model—either simulations of large mocks or MCMC sampling of the associated posterior—will require a vast number of SPS model calls. This is only made tractable by the neural emulation of SPS models (Alsing et al. 2020), which speeds up SPS computations by a factor of 10^4 compared to FSPS (Conroy & Gunn 2010).

2.4. Inference

The inference of redshift distributions under the generative model described in Section 2.2 requires inferring the population and data model parameters ψ and $\boldsymbol{\eta}$ under the forward-model assumptions, which, in turn, provide marginal posteriors for the tomographic redshift distributions, via Equation (3).[9]

In this paper, we are focused on validating a baseline forward model for predicting tomographic redshift distributions, without performing inference (or optimization) of the forward-model parameters. Nonetheless, it is useful to consider how inference works within our forward-modeling framework, and to highlight the advantages of performing redshift distribution inference in this fashion.

The joint posterior for the generative model described in Section 2.2 is given by (see Appendix A for a derivation):

$$P(\psi, \boldsymbol{\eta}, \{\boldsymbol{\varphi}, z\}_{1:N} | \{\boldsymbol{d}, \boldsymbol{\sigma}, S\}_{1:N}, N, \mathcal{O}) = P(\psi)P(\boldsymbol{\eta}) \times \prod_{i=1}^{N} \frac{P(\boldsymbol{\varphi}_i, z_i \mid \psi)P(\boldsymbol{d}_i \mid \boldsymbol{\varphi}_i, z_i, \boldsymbol{\sigma}_i, \boldsymbol{\eta})}{P(S_i \mid \psi, \boldsymbol{\sigma}_i, \boldsymbol{\eta})}, \quad (4)$$

where the selection term in the denominator is given by

$$P(S|\psi, \boldsymbol{\sigma}, \boldsymbol{\eta}) = \int P(S|\boldsymbol{d}, \boldsymbol{\sigma})P(\boldsymbol{d}|\boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta}) \times P(\boldsymbol{\varphi}, z|\psi) d\boldsymbol{d}\, d\boldsymbol{\varphi}\, dz. \quad (5)$$

This joint posterior can then be sampled (using MCMC) to jointly infer the population and data model parameters, as well as the SPS parameters and redshifts for each galaxy. The marginal posterior over the hyperparameters and data model parameters then provides a posterior over the target tomographic redshift distributions, via Equation (3).

Phrasing the redshift distribution inference task as a hierarchical model in this way has a number of advantages. First, note that, so far, this model contains only photometry: the method does not explicitly require spectroscopic calibration data. Where spec-zs (or other spectroscopically derived constraints on SPS parameters) are available for some subsets of the galaxies, they can be straightforwardly included by simply appending additional (sharply peaked) likelihoods for those galaxies. Importantly, the fact that any external spec-z calibration data are not representative of the main sample is unimportant in this approach, provided that selection cuts are only performed with respect to the main survey data.

Similarly, the inclusion of additional data (e.g., additional bands) from external surveys for subsets of galaxies is also straightforwardly achieved, by simply appending additional likelihood terms for those galaxies. Again, the fact that auxiliary data is only available for biased (unrepresentative) subsets of the galaxies is not important, provided that selection is not performed with respect to those auxiliary data.

Note that the population model $P(\boldsymbol{\varphi}, z|\psi)$ that appears in Equation (4) is a "global" quantity: it describes the statistical properties of the background galaxy population (without selection effects), and is therefore the same for all tomographic bins and across all surveys, regardless of their differing selection functions. This opens up strong synergies with the galaxy evolution community, who are concerned with constraining (various aspects of) $P(\boldsymbol{\varphi}, z|\psi)$: as our understanding of the statistical properties of the galaxy population improves, this can be fed directly into improved photometric redshift inferences, via improved priors on the hyperparameters ψ.

Finally, since the population and data model parameters are inferred from the photometric data in a self-consistent fashion, the uncertainties in the population and data model parameters will be fully propagated through to the final $n(z)$ inferences.

However, MCMC sampling of the joint posterior in Equation (4) requires computing the selection integral in Equation (5) for every galaxy in the sample, in every likelihood evaluation. This presents a severe computational bottleneck for sampling-based methods. In practice, for such sampling to be computationally tractable, the selection integral will require replacement with a fast emulator (e.g., Talbot & Thrane 2022).

Alternatively, SBI (aka likelihood-free inference) provides a framework for performing Bayesian inference under complex forward models using only simulations, bypassing the need to compute the likelihood (and selection integral) entirely (e.g., Alsing et al. 2018, 2019; Jeffrey & Wandelt 2020). For a recent application of SBI to a population model with selection effects, see Gerardi et al. (2021).

3. Baseline Forward Model

In this paper, we are focused on demonstrating the ability of a baseline forward model to recover redshift distributions, without performing additional inference or optimization of population-level parameters. We lay out the baseline forward-model assumptions for the SPS model in Section 3.1, for the galaxy population model in Section 3.2, and for the data model in Section 3.3.

3.1. SPS Model

We assume an SPS model with nine free parameters, summarized in Table 2 and described below.

Star formation histories (SFHs; star formation rates, or SFRs, as a function of time) are parameterized by a double power law:

$$\dot{M}(t; \alpha, \beta, \tau, z) \propto \frac{M}{(t/t^*)^{\alpha} + (t/t^*)^{-\beta}}, \quad (6)$$

where the transition time is defined as $t^* \equiv \tau t_{\rm univ}(z)$ for the lookback time $t_{\rm univ}(z)$, and the SFH is normalized such that it integrates to give the total stellar mass $\int_0^{t_{\rm univ}(z)} \dot{M} dt = M$. We define the SFR as the average of $\dot{M}$ over the past 100 Myr.

The gas-phase metallicity $\log_{10} Z_{\rm gas}$ is a free parameter, and the gas ionization parameter u is set so that it tracks the SFR, assuming the Kaasinen et al. (2018) relation between gas ionization and SFR.

[9] In practice, the mapping between the hyperparameters and tomographic redshift distributions will be done via simulating large mocks, following Section 2.2.

Table 2
SPS Model Parameters and Their Prior Ranges

Parameter	Description	Limits
$\log_{10}(M/M_\odot)$	Stellar mass	[7, 13]
$\log_{10} Z_{\rm gas}$	Gas-phase metallicity	[−1.98, 0.5]
$\log_{10} u$	Gas ionization parameter	[−4, −1]
τ_1	Birth cloud (stars younger than 10 Myr) dust attenuation	[0, 2]
τ_2	Diffuse dust attenuation	[0, 2]
δ	Negative offset from the Calzetti dust attenuation index	[0, 0.4]
α, β	Indices of double-power-law SFH	$[10^{-3}, 10^{1}]$
τ	Transition time of double-power-law SFH, as a fraction of lookback time	[0.007, 1]
z	Redshift	[0, 2.5]

The metallicity history of the stellar population is assumed to build up with the stellar mass production, such that the present-day stellar and gas-phase metallicities are identical:

$$Z(t) = (Z_{\rm gas} - Z_{\rm min})\frac{1}{M}\int_0^t \dot{M}(t)dt + Z_{\rm min}, \tag{7}$$

where $Z_{\rm min}$ is the minimum metallicity covered by the stellar templates.

Dust attenuation is modeled with two components describing the birth cloud (stars younger than 10 million years) and diffuse dust screens, respectively, following Charlot & Fall (2000; see Leja et al. 2017 for details). The birth cloud (τ_1) and diffuse (τ_2) attenuation, and the power-law index n of the Calzetti et al. (2000) attenuation curve for the diffuse component, are all free model parameters.

We assume MIST stellar evolution tracks and isochrones (Choi et al. 2016; Dotter 2016; based on MESA—Paxton et al. 2010, 2013, 2015). Nebular line and continuum emissions are generated with CLOUDY (Ferland et al. 2013), using model grids from Byler et al. (2017).

We emulate the photometry (apparent magnitudes) in all relevant bands using `speculator` (Alsing et al. 2020). The apparent magnitude in each band as a function of the SPS parameters and redshift is parameterized by a dense neural network with four hidden layers of 128 units each, and activation functions as described in Alsing et al. (2020). Each emulator is trained on 6.4×10^6 training samples, with SPS parameters and redshifts drawn from the population model (see Section 3.2 below) and model photometry computed using FSPS (Conroy & Gunn 2010). Training is performed following the prescription in Alsing et al. (2020), and we ensure that the 99.9% intervals for the emulator error distributions are better than 2% in all bands.

3.2. Galaxy Population Model

Specifying a population model for the SPS parameters and redshift amounts to specifying the joint (prior) distribution that characterizes the statistical properties of the galaxy population. Galaxy formation and evolution physics result in complex relationships between the stellar population parameters. In an effort to capture as much of this phenomenology as possible, we factorize the population prior into the following (generic) structure:

$$\begin{aligned} P(\text{SPS parameters, redshift}) = {} & P(\text{mass, redshift}) \\ & P(\text{metallicity}|\text{SFR, mass}) \\ & P(\text{SFR}|\text{mass, redshift}) \\ & P(\text{dust}|\text{SFR, mass, metallicity}) \\ & P(\text{age}|\text{SFR, mass}). \end{aligned} \tag{8}$$

Factorized this way, the population model is decomposed into a number of well-studied relations between galaxy characteristics: P(mass, redshift) is given by the redshift-evolving mass function; P(metallicity | SFR, mass) characterizes the fundamental metallicity relation (FMR); P(SFR | mass, redshift) characterizes the star-forming sequence (SFS); P(dust | SFR, mass, metallicity) specifies the relationship between dust and the star formation and chemical enrichment histories; and P(age | SFR, mass) specifies the empirical relationship between ages and SFHs.

For the SPS model setup chosen for this study (summarized in Table 2), the specific population model assumptions and default parameters are taken as follows.

3.2.1. Mass Function

The joint distribution of mass and redshift is defined by

$$P(M, z) \propto \Phi(M, z)dV(z), \tag{9}$$

where $\Phi(M, z)$ is the unnormalized mass function and $dV(z)$ is the differential comoving volume element. For the mass function, we assume a mixture of two Schechter functions, with default parameter values and redshift evolution taken from Leja et al. (2020). We assume a Planck 2015 (Ade et al. 2016) cosmology for the comoving volume element. The assumed mass function and redshift prior is shown in Figure 1.

3.2.2. FMR

Galaxies undergo continuous chemical evolution, as heavier elements are produced in stars and expelled into the interstellar medium, and gas flows regulate the metal content, by either the dilution or expulsion of enriched gas out of the galactic potential. On global scales, this results in a tight relationship between the gas-phase metallicity and the SFH of a galaxy (e.g., mass and SFR)—the so-called FMR (Yates et al. 2012; Andrews & Martini 2013; Nakajima & Ouchi 2014; Salim et al. 2014, 2015; Yabe et al. 2015; Kashino et al. 2016; Cresci et al. 2019; Curti et al. 2020). Qualitatively, galaxies on the FMR tend toward lower metallicities for higher SFRs, and higher metallicities for higher masses, but the overall shape is a nonlinear function of both mass and SFR. The FMR is typically considered to be independent of redshift, with galaxies moving along the relation as they evolve (and preferentially occupying different regions of the relation at different redshifts), but the relation itself remains constant over cosmic history.[10]

We take the FMR parameterization and default parameter values from Curti et al. (2020), where the FMR was measured over the broad stellar mass and SFR ranges covered by the Sloan Digital Sky Survey. The median gas-phase metallicity as

[10] The physics governing chemical enrichment is assumed to be constant over cosmic time.

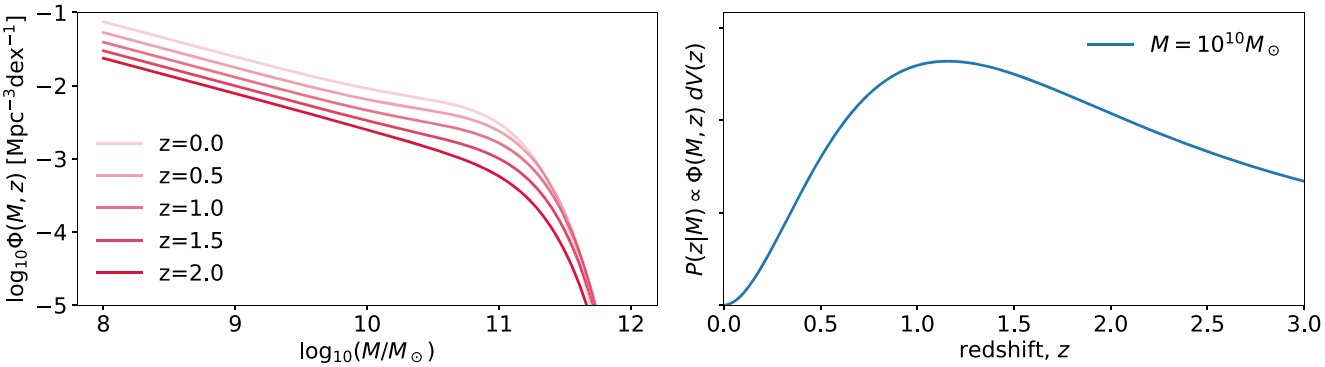

Figure 1. Left: the redshift-evolving mass function from Leja et al. (2020). Right: the implied redshift distribution for the background galaxy population (without selection) at $M = 10^{10}\, M_\odot$.

a function of mass and SFR is parameterized as

$$\langle \log_{10} Z_{\rm gas} \rangle = \Delta Z_0 - \frac{\gamma}{\beta} \log_{10}\left[1 + \left(\frac{M}{M_0(\mathrm{SFR})}\right)^{-\beta}\right], \quad (10)$$

where

$$\log_{10} M_0(\mathrm{SFR}) = m_0 + m_1 \log_{10}(\mathrm{SFR}), \quad (11)$$

with default parameters $\Delta Z_0 = 0.09$, $\gamma = 0.3$, $m_0 = 10.1$, $m_1 = 0.56$, and $\beta = 2.1$. We assume a Student's t-distribution (hereafter, student-t; with 2 degrees of freedom) for $P(\log_{10} Z_{\rm gas}|M, \mathrm{SFR})$, where the mean is given by Equation (10) and the FWHM is equal to 0.05. The FMR is shown in Figure 2.

The measurement of the FMR is sensitive to the way in which the SFRs and metallicities are estimated (see, e.g., Telford et al. 2016 and Cresci et al. 2019), with the SFRs and metallicities that are used to calibrate the FMR from spectra and those arising in a given SPS model being subtly different proxies for those quantities. Therefore, when fitting the FMR to photometric data, as part of the population model, it would be prudent to set reasonably broad priors on the FMR parameters, in order to capture any biases relative to measurements based on spectra.

While the majority of the galaxies are expected to live on the FMR (since the same physics drives chemical enrichment for most galaxies), those processes will be disrupted in merger events. Merged galaxies are therefore not expected to follow the same FMR relation (Bustamante et al. 2020). This can be compensated for by adding heavy tails to the FMR (as we do here, with the student-t distribution) or by deriving a separate model for the metallicities of merged galaxies.

3.2.3. SFS

The SFS characterizes the (redshift-evolving) relationship between the SFR and mass, with the vast majority of galaxies forming most of their mass either on or when passing through the SFS (Leitner 2012; Abramson et al. 2015). Qualitatively, the SFS is characterized by star-forming and quiescent galaxies with ongoing and negligible SFRs, respectively. The clustering of galaxies into these two populations, with different characteristic SFRs, results in a highly non-Gaussian—sometimes bimodal—distribution of SFRs (conditioned on mass and redshift) in the galaxy population as a whole (Daddi et al. 2007; Noeske et al. 2007; Karim et al. 2011; Rodighiero et al. 2011; Whitaker et al. 2012, 2014; Speagle et al. 2014; Renzini & Peng 2015; Schreiber et al. 2015; Tomczak et al. 2016; Leslie et al. 2020; Leja et al. 2022).

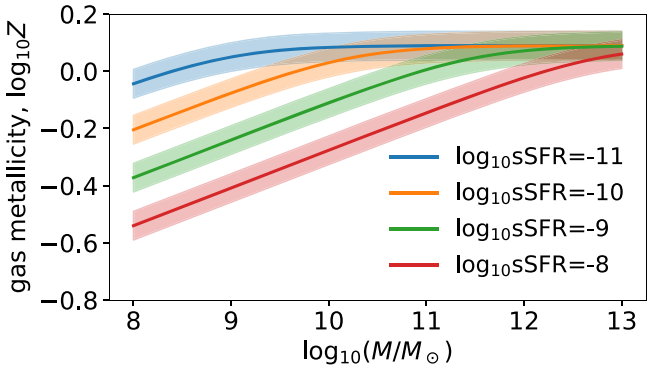

Figure 2. FMR priors on gas-phase metallicity conditioned on mass and specific SFR (in units of Gyr^{-1}). The solid lines represent the mean, and the bands show the FWHMs of the assumed student-t distribution.

In this work, we base our SFS model on the measured relation from Leja et al. (2022), who fit a normalizing flow to learn $P(\mathrm{SFR} \mid M, z)$ from three-dimensional Hubble Space Telescope (HST) data. The galaxy sample used in Leja et al. (2022) is mass-complete, down to around $10^9 M_\odot$, over the redshift range relevant for this study ($z \leqslant 1.5$). We note that $P(\mathrm{SFR} \mid M, z)$ is a very smooth function of mass and redshift, and to ensure sensible extrapolation below the mass limit, we fit a surrogate model (with good extrapolation properties) to the normalizing flow of Leja et al. (2022; see Figure 3, as well as Appendix B for technical details).

Specifying a population model requires that we put a prior on the SFH parameters (in this case, the three parameters of the double-power-law SFH), but the SFS provides only a prior constraint on a derived quantity: the SFR, $\mathrm{SFR}(\alpha, \beta, \tau, z)$. Hence, we need to define our prior over the SFH parameters in such a way that the target prior over the SFR is satisfied. One of the problems with parametric SFH models, such as the double power law, is that simple priors on the SFH parameters lead to strong (and undesirable) implied priors on the derived quantities, such as the SFR (e.g., Carnall et al. 2019). For example, taking a baseline uniform prior in $(\log_{10}\alpha, \log_{10}\beta, \tau)$ over reasonable ranges (see Table 2) leads to the undesirable implicit SFR prior $P_0(\mathrm{SFR} \mid z)$ shown in Figure 4.

To ensure that our SFH priors only encode the SFS, without spurious additional contributions from any baseline prior

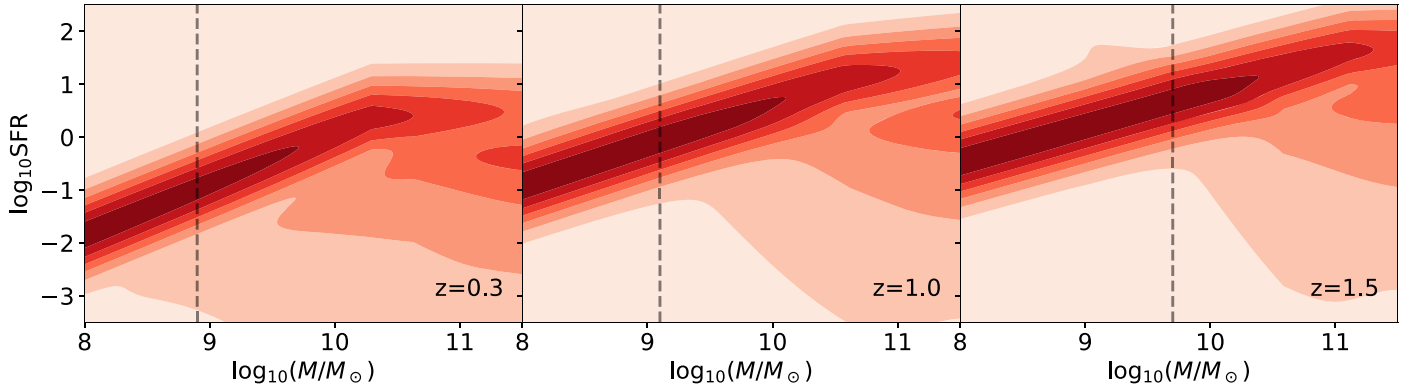

Figure 3. SFS prior on the SFR (in units of $M_\odot \mathrm{yr}^{-1}$) conditioned on mass and redshift, based on the measurement from Leja et al. (2022; see Appendix B for details). The vertical lines represent the mass-complete limit for the three-dimensional HST data on which the Leja et al. (2022) fit was performed; we extrapolate the SFS below the mass limit, as shown in the images.

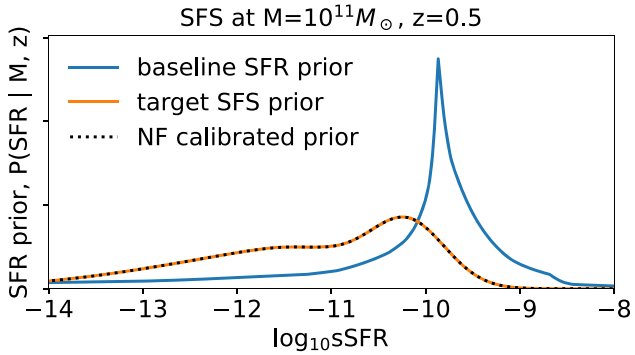

Figure 4. The prior on the SFR implied by taking uniform priors over the double-power-law SFH parameters ($\log_{10}\alpha$, $\log_{10}\beta$, τ) over the ranges specified in Table 2 is shown in blue, while the target prior (specified by the SFS measurement of Leja et al. 2022) is shown in orange. The SFR prior obtained by recalibrating the baseline prior with a normalizing flow (as described in Section 3.2.3) is shown by the black dotted line, giving excellent agreement with the target SFS prior.

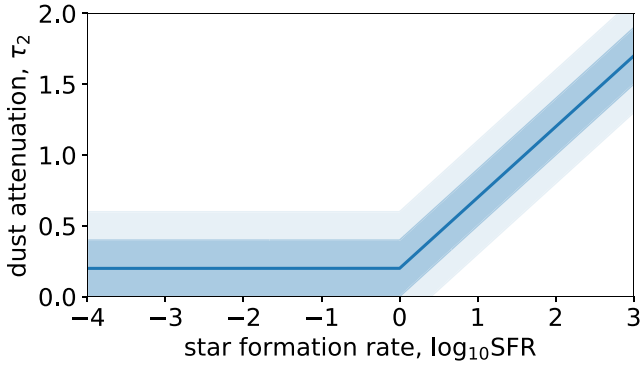

Figure 5. Prior on the diffuse dust attenuation conditioned on SFR. The solid line shows the mean and the bands show the 1σ and 2σ contours.

assumptions, we define the SFH priors as

$$P(\alpha, \beta, \tau|M, z) = \frac{\pi_0(\alpha, \beta, \tau)\, P(\mathrm{SFR}|M, z)}{P_0(\mathrm{SFR}|z)}, \tag{12}$$

where $\pi_0(\alpha, \beta, \tau)$ is the baseline (uniform) prior on the SFH parameters, $P(\mathrm{SFR} \mid M, z)$ is the target SFS prior, and $P_0(\mathrm{SFR} \mid z)$ is the implicit prior on the SFR implied by π_0. The implicit SFR prior is defined by the surface integral

$$P_0(\mathrm{SFR}|z) = \int_{\mathrm{SFR=const.}} \pi_0(\alpha, \beta, \tau) dS, \tag{13}$$

where dS is the surface element in the SFH parameter space (α, β, τ).

Dividing out the implicit SFR prior in this way ensures that the overall prior on the SFR is specified by the SFS only. To avoid having to compute the surface integrals in Equation (13) directly, we train a normalizing flow to learn the conditional density $P_0(\mathrm{SFR} \mid z)$, so that it can be conveniently divided out (the technical details are given in Appendix C).

3.2.4. Dust

The amount of dust attenuation, and the shape of the effective attenuation law, is governed by the total amount of dust, the grain composition, the dust–star–gas geometry in the galaxy, and its inclination relative to the observer. This can be encoded as a relationship between dust attenuation parameters and SFHs (SFR and mass), as well as, potentially, metallicity and redshift (see Salim & Narayanan 2020 for a review).

We take a relatively simple dust prior model, where the dust attenuation in the diffuse component is assumed to scale with SFR according to (following Tanaka 2015)

$$\langle \tau_2 \rangle = 0.2 + 0.5 \log_{10} \mathrm{SFR}\, \Theta(\log_{10} \mathrm{SFR}), \tag{14}$$

where Θ is the Heaviside step function. We assume a Gaussian prior on the diffuse dust attenuation with the mean given above, and a standard deviation of 0.2. The dust attenuation prior is shown in Figure 5.

The index of the dust attenuation law (for the diffuse component) is assumed to vary as a function of the total dust attenuation, with a mean given by

$$\langle \delta \rangle = -0.095 + 0.111\tau_2 - 0.0066\tau_2^2, \tag{15}$$

where δ is the (negative) offset from the index of the Calzetti attenuation curve (Calzetti et al. 2000). We take a Gaussian prior on δ, with the mean given above and a standard deviation of 0.4.

For the birth cloud component, we take a Gaussian prior on the ratio τ_1/τ_2 of the birth cloud and the diffuse dust components, with a mean equal to 1 and a standard deviation of 0.3. This is consistent with previous findings that the dust optical depth in nebular emission lines is roughly twice that of

the stellar component (Calzetti et al. 1994; Price et al. 2014; Reddy et al. 2015).

We leave more sophisticated dust prior modeling, e.g., where dust attenuation properties are conditioned on SFR, mass, metallicity, and redshift (e.g., Nagaraj et al. 2022), to future work.

3.2.5. Age

The double-power-law SFH parameterization and prior implicitly links age and SFR, with older (younger) galaxies having lower (higher) SFRs, qualitatively in line with expectations. However, the assumed priors on the double-power-law SFH parameters allow for a tail down to very low ages; we therefore impose a lower cut of 1 Gyr on the galaxy ages[11] to eliminate spuriously young galaxies. We do not impose any additional priors on age.

3.3. Data Model

The data model characterizes the sampling distribution of the measured photometry, given the true (model) fluxes, encoding both calibration (i.e., zero-points), modeling errors, and measurement noise. We treat the sampling distribution of the observed fluxes as a student-t distribution, with 2 degrees of freedom:

$$P(\boldsymbol{d}|\boldsymbol{f}, \boldsymbol{\sigma}, \boldsymbol{\eta}) \propto \prod_{b=1}^{N_{\rm bands}} \left[1 + \frac{1}{2}\left(\frac{d_b - \alpha_{{\rm zp},b} f_b}{\Sigma_b(f_b)}\right)^2\right]^{-3/2}, \quad (16)$$

where the total uncertainty $\boldsymbol{\Sigma}$ is given by

$$\Sigma_b^2 = \sigma_b^2 + (\beta_b \alpha_{{\rm zp},b} f_b)^2. \quad (17)$$

The data model parameters $\boldsymbol{\eta} = (\boldsymbol{\alpha}_{\rm zp}, \boldsymbol{\beta})$ characterize the zero-points $\boldsymbol{\alpha}_{\rm zp}$ and error floors $\boldsymbol{\beta}$ (encoding modeling errors, emulator errors, and an effective noise floor on the measurement uncertainties).

In the application to GAMA and VVDS data in Sections 4–5, we fix the zero-points to the values published by the respective survey collaborations, and assume a default value of 0.03 for the (fractional flux) error floors in all bands.

We note that this data model is readily extendable to include additional error terms for emission line modeling errors, spectral energy distribution modeling errors as a function of rest-frame wavelength, and the parameters describing the shape (e.g., skewness or tail weight) of the data sampling distribution. A more sophisticated error model (including the hierarchical calibration of hyperparameters) is investigated in our companion paper (B. Leistedt et al. 2023, in preparation).

3.4. Uncertainty Model

The uncertainty model describes the distribution of photometric measurement uncertainties over the survey, which will vary from galaxy to galaxy, due to heterogeneous observing conditions and strategies, as well as the varying difficulty of extracting photometry from galaxy images with different morphologies and geometries, and will also scale with the (true) flux, owing to the Poisson photon count contribution to the overall measurement error.

[11] We take age here to mean the mass-weighted age.

We take a data-driven approach to uncertainty modeling, where we learn the distribution $P(\boldsymbol{\sigma}|\boldsymbol{f}, \mathcal{O})$ directly from the data. This process proceeds in two steps. First, we fit each of the galaxies under the SPS model, population prior, and data model assumptions described above, by MCMC sampling their individual posteriors (see Equation (4) with fixed hyperparameters). For each galaxy, this provides a maximum a posteriori (MAP) estimate of their true fluxes $\boldsymbol{f}$. We then take the catalog of MAP estimated fluxes and associated uncertainties $\{\boldsymbol{f}, \boldsymbol{\sigma}\}_{1:N}$, and train a mixture density network (MDN) to learn $P(\boldsymbol{\sigma}|\boldsymbol{f}, \mathcal{O})$. The MDN parameterizes the uncertainty distribution as a Gaussian mixture:

$$P(\boldsymbol{\sigma}|\boldsymbol{f}, \mathcal{O}) = \sum_{c=1}^{N_{\rm comp.}} r_c(\boldsymbol{f}; \boldsymbol{w}) \mathcal{N}[\boldsymbol{\mu}_c(\boldsymbol{f}; \boldsymbol{w}), \boldsymbol{\Sigma}_c(\boldsymbol{f}; \boldsymbol{w})], \quad (18)$$

where the component weights, means, and variances are functions of flux, parameterized by a dense neural network (whose weights and biases are denoted by $\boldsymbol{w}$). The MDN is trained by minimizing the total (negative) log-likelihood of the data under the model with respect to the network weights:[12]

$$\mathcal{L}(\boldsymbol{w}) = -\sum_{i=1}^{N_{\rm galaxies}} \ln P(\boldsymbol{\sigma}_i|\boldsymbol{f}_i, \boldsymbol{w}). \quad (19)$$

Throughout this paper, we take a default MDN with 12 components and a single hidden layer with 256 units and leaky ReLU activation functions. Examples of trained MDNs for the uncertainty distributions for GAMA and VVDS are shown in Figures 6 and 7.

4. Case Study I: GAMA

The GAMA survey covers 250 square degrees, and has obtained ∼230,000 spectroscopic redshifts over the past decade (Driver et al. 2011). The survey was designed to have simple target selection, based on photometry alone (discussed below), making it an ideal data set for validating our forward model. We take GAMA data release 4 (DR4; Driver et al. 2022), with photometry in the KiDS *ugri* and VIKING *ZYHJKs* bands.

The main selection is performed on the KiDS r band, with spectroscopic redshifts measured for all galaxies with $r < 19.65$. An additional color cut of $(J - Ks) > 0.025$ is made for star–galaxy separation. With these cuts, we are left with a sample of 206,454 galaxies with 9-band photometry (*ugri-ZYHJKs*) and measured spectroscopic redshifts (for validation).

4.1. Data Model

We assume student-t uncertainties on the fluxes, as described in Section 3.3, and take the extinction and zero-point corrections provided with GAMA DR4 (Driver et al. 2022).

We train an MDN to model the distribution of the measurement uncertainties as a function of flux, using the GAMA data, as described in Section 3.4. The uncertainty distributions and corresponding trained models are shown side by side in Figure 6.

[12] Note that this is equivalently a Monte Carlo estimate of the Kullback–Leibler divergence between the model and the true distribution, up to an additive constant.

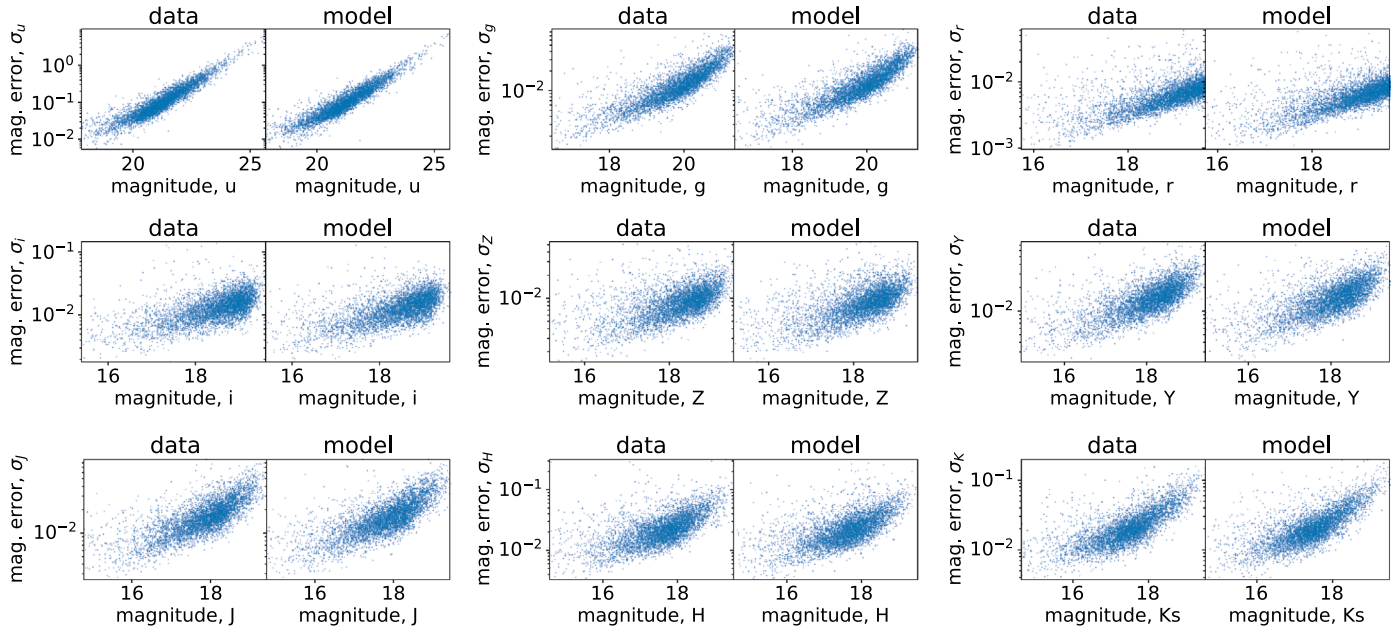

Figure 6. Magnitude uncertainties vs. magnitudes for the GAMA data (left panels) vs. the trained MDN model for the error distribution conditioned on flux (right panels). Note that the magnitudes in both the left and right panels are MAP magnitudes, from an initial fit of the SPS model to the GAMA galaxies, as described in Section 3.4.

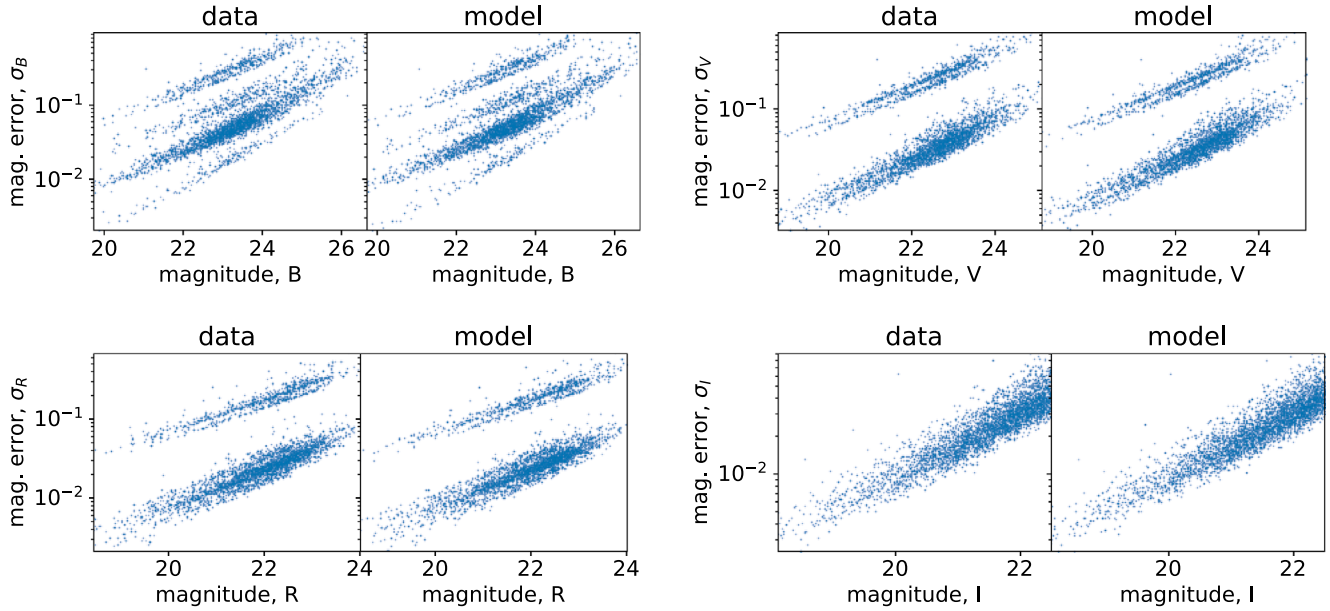

Figure 7. Magnitude errors vs. magnitudes for the VVDS data (left panels) vs. the trained MDN model for the error distribution conditioned on flux (right panels). Note that the magnitudes in both the left and right panels are MAP magnitudes, from an initial fit of the SPS model to the VVDS galaxies, as described in Section 3.4.

4.2. Tomographic Binning

For the purpose of tomographic binning, we train a simple (dense) neural network estimator for the redshift, given the measured KiDS and VIKING photometry. We take a dense network with four hidden layers of 64 units each and leaky ReLU activations, passing the measured magnitudes and magnitude errors in the nine bands as inputs and the estimated redshift as output. The network is trained to minimize the mean square error on the redshift, using the GAMA photometry and redshifts. Training is performed with `Adam`, using a batch size of 1024, a training : validation split of 90 : 10, and by triggering early stopping when the validation loss has ceased to improve after 30 epochs. The resulting redshift estimator has an overall accuracy of around $\sigma_z \simeq 0.06$.

The GAMA galaxies are binned into two tomographic bins, based on their estimated redshift: $0 < \hat{z} < 0.2$ and $\hat{z} > 0.2$ for the two bins, respectively.

4.3. Results

Forward-model predictions are obtained by generating a large mock catalog, following the prescription in Section 2.2. We MCMC sample the population model (imposing a prior limit of $r < 20.65$), draw uncertainties and add noise according to the uncertainty and data models described above, and apply

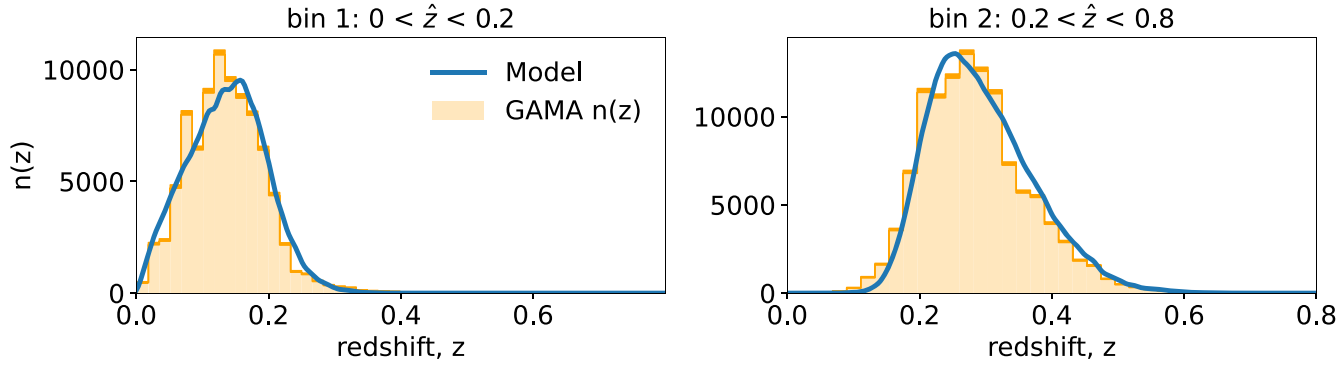

Figure 8. Tomographic redshift distributions obtained by the forward model (blue) compared to histograms of the GAMA spectroscopic redshifts (orange). The model predictions are in excellent agreement with the distributions of the spectroscopic redshifts.

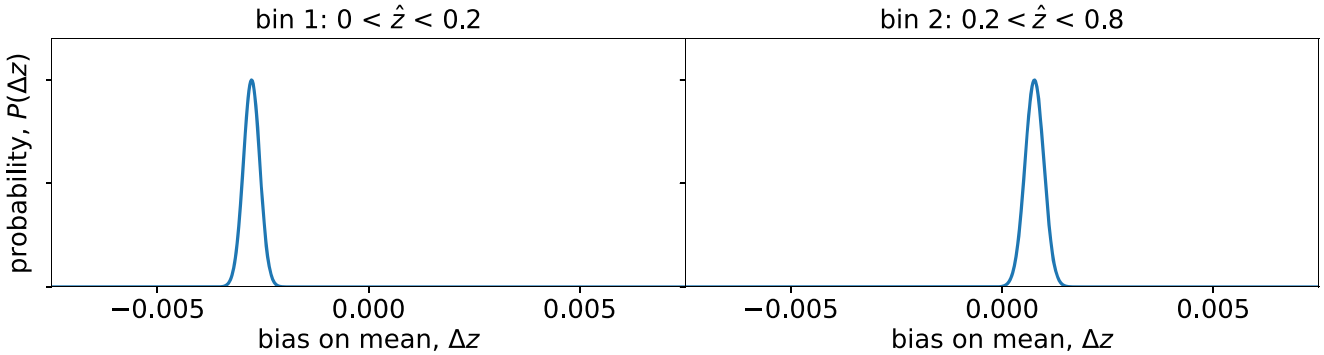

Figure 9. The bias on the mean redshifts of the model redshift distributions vs. the data for GAMA, $\Delta z = \langle z_{\rm model}\rangle - \langle z_{\rm data}\rangle$. The distributions are obtained by bootstrapping samples from the model $n(z)$, taking a kernel density estimate of the bootstrapped sample means, then centering the kernel density estimate on the difference between the sample mean of the spec-zs and the mean of the model $n(z)$.

selection cuts $r < 19.65$ and $(J - Ks) > 0.025$ to the simulated noisy photometry. Tomographic bin labels are assigned based on the redshift estimator described above. We continue sampling until $5 \cdot 10^5$ selected samples are obtained.

The tomographic redshift distributions predicted by the forward model are shown alongside the spec-z histograms in Figure 8, and the corresponding biases on the means of the redshift distributions are shown in Figure 9. The forward model is able to predict the redshift distributions with biases of around 0.003 and 0.001 on the mean, for the two respective tomographic bins. This is comfortably accurate enough for ongoing Stage III surveys (e.g., Asgari et al. 2021), where the statistical error on the mean redshift per tomographic bin is $\mathcal{O}(0.01)$, and cosmological parameter constraints should be insensitive to biases of $\lesssim$0.04 (Hildebrandt et al. 2016). The model predictions are very close to the accuracy requirements for Stage IV surveys, where the bias on the mean should not exceed[13] $\Delta z < 0.002(1+z)$ (Mandelbaum et al. 2018).

We reiterate that these model predictions are obtained by assuming (fixed) default parameters for the population model, as described in Section 3.2; fitting the uncertainty distribution to the GAMA data, to characterize the distribution of the photometric errors (as a function of flux); and assuming the zero-point calibration provided by the GAMA collaboration (Driver et al. 2022).

[13] For the LSST year one analysis, the requirement on the mean bias per tomographic bin is $\Delta z < 0.002(1+z)$, decreasing to $\Delta z < 0.001(1+z)$ by year 10 (Mandelbaum et al. 2018).

5. Case Study II: VVDS

Similar to GAMA, VVDS (Le Fèvre et al. 2013) is a spectroscopic survey designed to have simple photometric target selection. We focus on VVDS-Wide, which covers 8.7 square degrees and has obtained spectra for 25,805 galaxies down to $I < 22.5$.

We take photometric data in the *BVRI* bands from VVDS-Wide, which were obtained with the CFH12K camera at CFHT. The main selection is performed in the I band, with spectra obtained for objects with $17.5 < I < 22.5$. The imaging survey is sufficiently deep (a limiting magnitude of $I = 24.8$) to ensure 100% completeness down to $I = 22.5$ for the spectroscopic sample. Star–galaxy separation was performed on the spectra, so no other photometric cuts were performed. We selected only galaxies with redshift quality flags of 3 or 4 (>95% probability of obtaining a correct redshift, according to Le Fèvre et al. 2013). The relevant information for modeling any correlations between the assessed reliability and redshift in detail is not publicly available for this catalog, hence we make no attempt to model this implicit selection effect. However, Figure 13 in Le Fèvre et al. (2013), for a VVDS-Deep sample in the same magnitude range, suggests that the spectroscopic success rate for this sample is expected to be roughly uniform up to $z \simeq 1$, then drop thereafter. This only impacts ~1%–2% of the sample, in a regime where photometric selection is expected to strongly dominate in any case. Hence, we do not expect a significant impact on our results from the redshift dependence of the spectroscopic success rate.

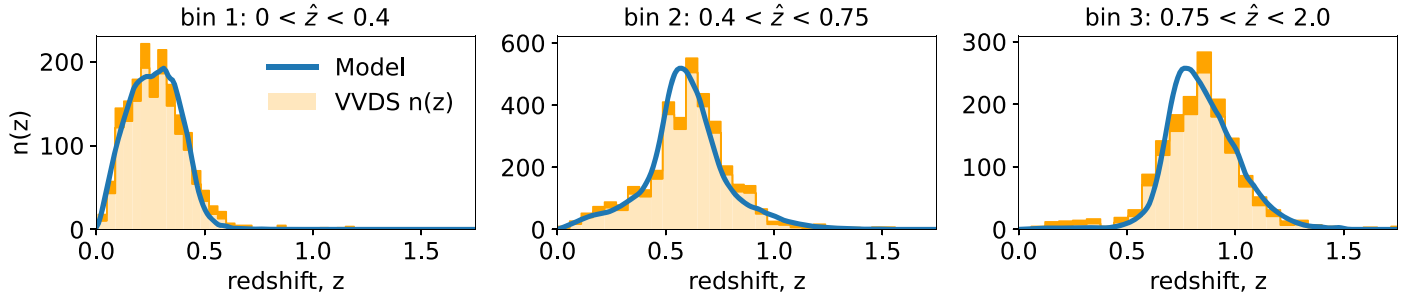

Figure 10. Tomographic redshift distributions obtained by the forward model (blue) compared to histograms of the VVDS spectroscopic redshifts (orange). The widths of the histogram bars indicate the $\pm\sqrt{N}$ Poisson noise. The model predictions are in excellent agreement with the distributions of the spectroscopic redshifts.

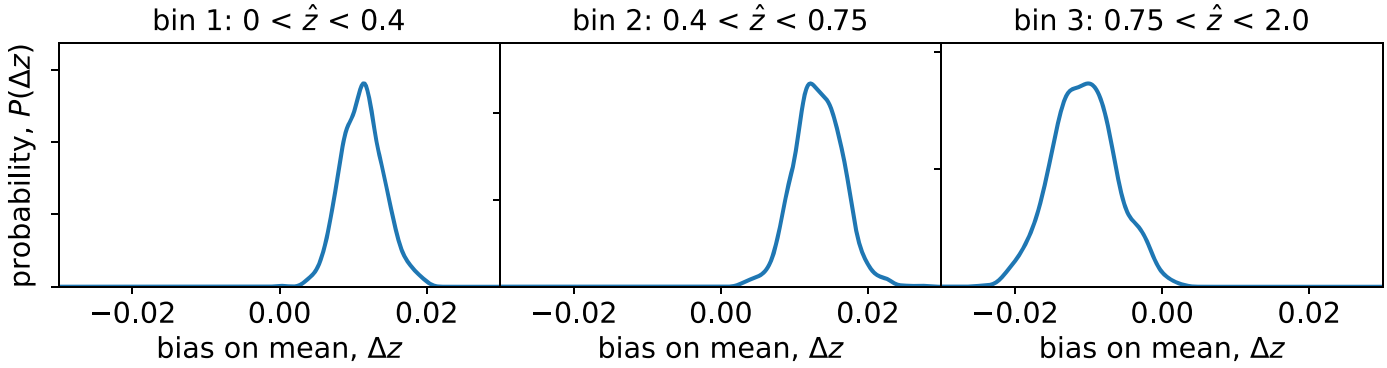

Figure 11. The bias on the mean redshifts of the model redshift distributions vs. the data for VVDS, $\Delta z = \langle z_{\rm model}\rangle - \langle z_{\rm data}\rangle$. The distributions are obtained by bootstrapping samples from the model $n(z)$, taking a kernel density estimate of the bootstrapped sample means, then centering the kernel density estimate on the difference between the sample mean of the spec-zs and the mean of the model $n(z)$.

5.1. Data Model

We again assume student-t uncertainties on the fluxes, as described in Section 3.3, and take the extinction and zero-point corrections provided by the VVDS team (Le Fèvre et al. 2013).

We train an MDN to model the distribution of the measurement uncertainties as a function of flux, using the VVDS data, as described in Section 3.4. Unlike GAMA, the VVDS *BVRI* photometry is patchy, with roughly a quarter of the objects missing measurements in at least one band. We set the fractional uncertainties for the missing values to 1000, and constrain one component in the mixture model to be a delta function at 1000, where the relative weight of that component in the MDN then encodes the relative probability of having a missing value, as a function of flux. The uncertainty distributions (for nonmissing values) and corresponding trained models are shown side by side in Figure 7. Note the multimodal structure in the uncertainty distributions, owing to the varying depth of the photometry in all but the *I* band. This makes for a more challenging test case for the forward-modeling framework.

5.2. Tomographic Binning

For the purpose of tomographic binning, we train a simple (dense) neural network estimator for the redshift, given the measured *BVRI* photometry. We again take a dense network with four hidden layers of 64 units each and leaky ReLU activations, passing the measured magnitudes and magnitude errors in the four bands as inputs and the estimated redshift as output. The network is trained on the VVDS photometry and spec-zs, as described in Section 4.2. The resulting redshift estimator has an overall accuracy of around $\sigma_z \simeq 0.2$. Note that the redshift estimator is considerably less accurate in this case as compared to GAMA, owing to the poorer constraining power of the *BVRI* bands, the prevalence of missing values in the VVDS photometry, and the smaller training set.

The VVDS galaxies are binned into three tomographic bins, based on their estimated redshift: $0 < \hat{z} \leqslant 0.4$, $0.4 < \hat{z} \leqslant 0.75$, and $0.75 < \hat{z} \leqslant 2$.

5.3. Results

As in Section 4.3, model predictions are obtained by MCMC sampling the population model (imposing a prior limit of $I < 23.5$), drawing uncertainties and adding noise according to the uncertainty and data models described above, and applying selection cuts $17.5 < I < 22.5$ to the simulated noisy *I*-band magnitudes. Tomographic bin labels are assigned based on the redshift estimator described above. Sampling is continued until $5 \cdot 10^5$ selected samples are obtained.

The tomographic redshift distributions predicted by the forward model are shown alongside the spec-z histograms for VVDS in Figure 10, and the corresponding biases on the means of the redshift distributions are shown in Figure 11.

The forward model is able to predict the redshift distributions with a bias of $\Delta z \simeq 0.01$ on the mean in all three bins. This is comparable to the statistical error on the mean redshift per tomographic bin for Stage III surveys (e.g., Asgari et al. 2021), and below the threshold where cosmological parameter biases become significant ($\Delta z \lesssim 0.04$; Hildebrandt et al. 2016). The model predictions are within a factor of a few of the requirements for Stage IV surveys ($\Delta z < 0.002(1+z)$; Mandelbaum et al. 2018), and we note that the Stage IV

requirements are contained within the error distribution on the mean bias for VVDS, as shown in Figure 11.

We reiterate that these model predictions are obtained by assuming (fixed) default parameters for the population model, as described in Section 3.2; fitting the uncertainty distribution to the VVDS data, to characterize the distribution of photometric errors (as a function of flux); and assuming the zero-point calibration provided by the VVDS collaboration (Le Fèvre et al. 2013).

We note that our data model for VVDS has residual uncertainties that may be responsible for some of the redshift bias. We have assumed standard Johnson *BVRI* filters, which are approximately correct, but there are some differences in detail (Le Fèvre et al. 2013). Zero-point calibration for VVDS was also reported to be challenging, with large (and sometimes differential) zero-points required in some bands to achieve consistency between photometry and spectra (Vincent Le Brun, private communication). We have not included calibration uncertainties in our results.

6. Discussion

While our baseline model is able to accurately recover the tomographic redshift distributions for GAMA and VVDS, a number of improvements are possible.

The baseline SPS model assumes a simple double-power-law SFH parameterization. Such a simple SFH parameterization is not expected to capture the full diversity of real SFHs, and can lead to overly restrictive correlations between important derived quantities (such as SFR and age), which might not be representative of real galaxies. These limitations can be alleviated by nonparametric (binned) SFH models (e.g., Leja et al. 2019) or more physical SFH parameterizations (e.g., Alarcon et al. 2023).

Regarding the population model, the largest modeling uncertainties are expected to come from the dust attenuation prior, where our baseline model assumed that dust attenuation scales with SFR only. In reality, dust characteristics are expected to be related to the detailed star formation and metallicity enrichment histories, motivating a more sophisticated dust prior model (e.g., Nagaraj et al. 2022).

The data model also has a number of simplifying assumptions. SPS modeling errors are expected to vary as a function of rest-frame wavelength, with emission lines in particular being subject to potentially significant modeling biases (B. Leistedt et al. 2023, in preparation). In the context of inferring SPS parameters for individual galaxies, photometric error floors have often been used to capture both modeling and calibration uncertainties, in order to increase the uncertainties on the inferred SPS model parameters and reduce the biases. While this strategy is fine for individual galaxies, increasing the variances in order to cover the potential biases in this way can lead to overdispersion of the population-level parameters. Therefore, data modeling efforts should instead focus on parameterizing and modeling such biases directly, rather than treating them as extra variance terms. The shape of the noise distribution also merits careful investigation, with (for example) the skewness and tail weights of the photometric measurement errors likely varying between bands, and as a function of flux and background noise levels.

Regarding the selection modeling, we have so far considered a scenario where selection is performed with respect to the measured photometry alone. However, for weak-lensing surveys, some additional selection cuts will typically be made on the images, such as image quality cuts to ensure reliable shear measurements, image-based star–galaxy separation, deblending, surface brightness cuts, etc. Because galaxy image characteristics correlate with SPS parameters and redshift, image-based cuts will induce additional selection effects that could modify the resulting redshift distributions.

In Appendix A, we show that the effects of image-based selection cuts can be addressed by replacing the population model with an effective population prior describing the statistical properties of the galaxy population that passes the image cuts (conditioned on the characteristics of the survey, etc.). Hence, image-based selection can be incorporated into the forward-modeling framework, by parameterizing the effects of those image cuts on the population prior over SPS parameters and redshift, and inferring those additional hyperparameters alongside the other population and data model parameters. Alternatively, if it can be demonstrated (or orchestrated) that the photometric cuts are sufficiently stronger than any image-based cuts, such that the image cuts have a negligible impact on the analysis sample, then those unmodeled selection effects can be safely ignored.

In addition to modeling improvements, the inference (or optimization) of population and data model parameters from the photometric data should lead to additional improvements in accuracy. For robust inferences, the data model parameters should be self-consistently calibrated using the photometric data themselves, with the photometric redshifts expected to be particularly sensitive to zero-point calibrations (but with all data model parameters playing a role). Regarding the population model, we note that different aspects of the model are better constrained than others by external data. In particular, the dust prior and FMRs are expected to be the least well understood and constrained, meriting broader priors on their parameters. The SFS is somewhat better constrained (e.g., Leja et al. 2022), while the mass function is relatively tightly constrained (e.g., Leja et al. 2020).

7. Conclusions

We have presented a forward-modeling framework for photometric surveys, which is capable of accurately predicting the tomographic redshift distributions required for cosmological analyses. Scaling this forward-modeling approach to large surveys is made possible by the neural emulation of SPS models (Alsing et al. 2020).

Forward modeling has a number of advantages over existing methods for estimating cosmological redshift distributions. In contrast to direct calibration methods, forward modeling does not require external spectroscopic data: it is therefore not hampered by the (lack of) availability of spectroscopic redshifts at the depths required for photometric surveys, and it is not vulnerable to biases arising from spectroscopic selection effects that cannot be well described by reweighting in broadband color space. In contrast to cross-correlation-based estimators, it is not sensitive to galaxy bias modeling assumptions. Our forward-modeling framework also resolves a number of the limitations of existing template-based methods, by replacing template sets with a continuous physical model for galaxy spectra (with associated physical priors), carefully treating selection effects, and enabling the self-consistent inference of model parameters describing the galaxy population and data model.

By explicitly modeling the processes that give rise to the target redshift distributions, forward modeling allows for fine control over the relevant modeling assumptions. In particular, it creates synergies between galaxy evolution physics and photometric redshift inference: as our constraints on the statistical properties of the galaxy population improve, these will lead directly to improved priors on the population model parameters, and hence improved photometric redshift inferences.

We have demonstrated the utility of our forward-modeling framework by accurately recovering the redshift distributions for the GAMA and VVDS surveys, validating them against their spectroscopic redshifts. The model is able to predict the tomographic redshifts for those two surveys, with biases of $\Delta z \lesssim 0.003$ for GAMA and $\Delta z \simeq 0.01$ for VVDS, respectively, without performing inference or optimization of the model parameters describing the galaxy population and photometric calibration. This accuracy is sufficient for the ongoing Stage III surveys, and approaches the accuracy requirements for Stage IV surveys. We anticipate that with additional modeling improvements and the optimization of the model hyperparameters, forward modeling can provide a path to accurate cosmological redshift distribution inference for Stage IV surveys.

In a companion paper (B. Leistedt et al. 2023, in preparation), we demonstrate the utility of this forward-modeling framework for inferring individual redshifts, including the hierarchical calibration of data model hyperparameters.

We thank George Efstathiou, Angus Wright, Konrad Kuijken, Hendrik Hildebrandt, Will Hartley, and Jeff Newman for valuable discussions. We also thank Vincent Le Brun and Henry McCracken for helpful communications regarding the VVDS data. This project has received funding from the European Research Council (ERC), under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101018897 CosmicExplorer). This work has also been enabled by support from the research project grant "Understanding the Dynamic Universe," funded by the Knut and Alice Wallenberg Foundation, under Dnr KAW 2018.0067. J.A., H.V.P., and D.J.M. were partially supported by the research project grant "Fundamental Physics from Cosmological Surveys," funded by the Swedish Research Council (VR), under Dnr 2017-04212. The work of H.V.P. was additionally supported by the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine. B.L. is supported by the Royal Society, through a University Research Fellowship. H.V.P. and D.J.M. acknowledge the hospitality of the Aspen Center for Physics, which is supported by National Science Foundation grant No. PHY-1607611. The participation of H.V.P. and D.J.M. at the Aspen Center for Physics was supported by the Simons Foundation.

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

Appendix A
Derivation of the Joint Posterior including Selection Effects

For the purpose of deriving the joint posterior for the forward model described in Section 2.2, it is useful to consider the generation of a selected sample directly, as follows (the notation is summarized in Table 1).

The total number N of selected galaxies is drawn (assuming Poisson statistics), given the expected number of selected objects under the population model, data model, and selection effects:

$$\bar{N}(\Phi_0, \psi, \boldsymbol{\eta}, \mathcal{O}) = A \int \Phi_0 \rho(z; \psi) \frac{dV}{dz} P(\boldsymbol{\varphi}|z, \psi) P(S|\boldsymbol{\varphi}, z, \boldsymbol{\eta}, \mathcal{O}) d\boldsymbol{\varphi}\, dz, \tag{A1}$$

where $P(S|\boldsymbol{\varphi}, z, \boldsymbol{\eta}, \mathcal{O})$ is the selection probability for a given set of galaxy parameters (and data model parameters $\boldsymbol{\eta}$ and $\mathcal{O}$) and A is the survey area. The SPS parameters, redshifts, measurement uncertainties, and data vectors for each (selected) galaxy are then drawn from their respective distributions, conditioned on selection. The generative model is hence given by

$$\begin{aligned} &P(\psi, \Phi_0, \boldsymbol{\eta}, \{\boldsymbol{\varphi}, z\}_{1:N}, \{\boldsymbol{d}, \boldsymbol{\sigma}, S\}_{1:N}, N|\mathcal{O}) \\ &= P(\psi) P(\boldsymbol{\eta}) P(\Phi_0) \frac{\bar{N}(\Phi_0, \psi, \boldsymbol{\eta})^N e^{-\bar{N}(\Phi_0, \psi, \boldsymbol{\eta})}}{N!} \prod_{i=1}^{N} P(\boldsymbol{\varphi}_i, z_i|\psi, \boldsymbol{\sigma}_i, S_i, \boldsymbol{\eta}) P(\boldsymbol{d}_i|\boldsymbol{\varphi}_i, z_i, \boldsymbol{\sigma}_i, S_i, \boldsymbol{\eta}) P(\boldsymbol{\sigma}_i|\boldsymbol{f}_i, \mathcal{O}, \psi, \boldsymbol{\eta}, S_i). \end{aligned} \tag{A2}$$

Taking a log-uniform prior for the present-day volume density, $P(\Phi_0) = 1/\Phi_0$, and noting that $\bar{N} \propto \Phi_0$, we can marginalize out Φ_0 analytically and obtain

$$P(\psi, \boldsymbol{\eta}, \{\boldsymbol{\varphi}, z\}_{1:N}, \{\boldsymbol{d}, \boldsymbol{\sigma}, S\}_{1:N}, N|\mathcal{O}) = N^{-1} P(\psi) P(\boldsymbol{\eta}) \prod_{i=1}^{N} P(\boldsymbol{\varphi}_i, z_i|\psi, \boldsymbol{\sigma}_i, S_i, \boldsymbol{\eta}) P(\boldsymbol{d}_i|\boldsymbol{\varphi}_i, z_i, \boldsymbol{\sigma}_i, S_i, \boldsymbol{\eta}) P(\boldsymbol{\sigma}_i|\boldsymbol{f}_i, \mathcal{O}, \psi, \boldsymbol{\eta}, S_i). \tag{A3}$$

The joint posterior is hence given by

$$P(\psi, \eta, \{\varphi, z\}_{1:N}|\{\boldsymbol{d}, \boldsymbol{\sigma}, S\}_{1:N}, N, \mathcal{O}) = P(\psi)P(\eta)\prod_{i=1}^{N} P(\varphi_i, z_i|\psi, \boldsymbol{\sigma}_i, S_i, \eta)P(\boldsymbol{d}_i|\varphi_i, z_i, \boldsymbol{\sigma}_i, S_i, \eta), \tag{A4}$$

where we have dropped the $P(\boldsymbol{\sigma}_i|\boldsymbol{f}_i, \mathcal{O}, \psi, \eta, S_i)$ term in the posterior, since its sensitivity to the latent parameters and hyperparameters is typically negligible compared to the likelihood and prior terms.

Written this way, the population model and likelihood terms are conditioned on selection. This parameterization has the drawback that a population model conditioned on selection is typically hard to parameterize directly, and becomes a survey-specific quantity. Also, if the selection cuts involve more than simple (independent) flux or signal-to-noise ratio cuts in each band, then a data model conditioned on selection is also hard to compute directly.

It is instead desirable to rewrite the model in terms of the population model for the background galaxy population, and the likelihood without selection. Using the chain rule, the population model and likelihood terms can be rewritten as

$$\begin{aligned} P(\varphi, z|\psi, \boldsymbol{\sigma}, S, \eta) &= \frac{P(\varphi, z|\psi)P(S|\varphi, z, \boldsymbol{\sigma}, \eta)}{P(S|\psi, \boldsymbol{\sigma}, \eta)}, \\ P(\boldsymbol{d}|\varphi, z, \boldsymbol{\sigma}, S, \eta) &= \frac{P(\boldsymbol{d}|\varphi, z, \boldsymbol{\sigma}, \eta)P(S|\boldsymbol{d}, \boldsymbol{\sigma})}{P(S|\varphi, z, \boldsymbol{\sigma}, \eta)}. \end{aligned} \tag{A5}$$

Inserting these into Equation (A4), we obtain (after cancellation and dropping parameter-independent terms)

$$P(\psi, \eta, \{\varphi, z\}_{1:N}|\{\boldsymbol{d}, \boldsymbol{\sigma}, S\}_{1:N}, N, \mathcal{O}) = P(\psi)P(\eta) \times \prod_{i=1}^{N} \frac{P(\varphi_i, z_i|\psi)P(\boldsymbol{d}_i|\varphi_i, z_i, \boldsymbol{\sigma}_i, \eta)}{P(S_i|\psi, \boldsymbol{\sigma}_i, \eta)}, \tag{A6}$$

where the selection term in the denominator is given by

$$P(S|\psi, \boldsymbol{\sigma}, \eta) = \int P(S|\boldsymbol{d}, \boldsymbol{\sigma})P(\boldsymbol{d}|\varphi, z, \boldsymbol{\sigma}, \eta)P(\varphi, z|\psi)d\boldsymbol{d}\, d\varphi\, dz. \tag{A7}$$

Note that the selection term only depends explicitly on the hyperparameters and the data model parameters ψ and η. In a special case where one is only interested in inferring the latent parameters (φ, z) for each galaxy, with the hyperparameters and data model parameters being fixed, selection appears to "drop out" of the problem, so that one should infer the latent parameters under the population prior and data models without selection. Note that in this case (although it might seem counterintuitive), selection effects are still properly included: they only enter implicitly via the ensemble of selected galaxies that have made it into the analysis sample.

When inferring the hyperparameters and data model parameters, though, the selection term is important: the high-dimensional integral over the parameter and data space in Equation (5) usually represents the computational bottleneck for sampling Bayesian hierarchical models under selection effects.

A.1. Selection Effects on Both Photometry and Images

In the model derived above, we have assumed that selection was performed with respect to the photometric data vector only (i.e., based on the measured fluxes and their uncertainties). In reality, for weak-lensing surveys, some selection cuts will occur at the level of images, such as image quality cuts to ensure robust shear measurements, the removal of blended objects, surface brightness cuts, etc. Therefore, it is useful to consider the impact of image-based selection cuts on the generative model. To this end, in this section, we derive the joint generative model for galaxy photometry and images, and subsequently marginalize over the images, to explore a typical case where one wants to perform redshift inference with respect to photometry only, but needs to account for image-based selection cuts.

We will assume that galaxy images are characterized by the parameters $\boldsymbol{\zeta}$ (in addition to the SPS parameters and redshift). We denote galaxy image data vectors and uncertainties with $\boldsymbol{D}$ and $\boldsymbol{\Sigma}$, respectively, where $\boldsymbol{D}$ can be taken to mean either the full pixelized image or some low-dimensional summary statistics derived from the galaxy images on which the selection is performed. For the purpose of this derivation, we distinguish photometric and image-based selection cuts with S_d and S_D, respectively, and use S to denote combined selection.

Conceptually, the forward model proceeds as before. The total number N of selected galaxies is drawn (assuming Poisson statistics), given the expected number of selected objects:

$$\begin{aligned} &\bar{N}(\Phi_0, \psi, \eta, \mathcal{O}) \\ &= A\int \Phi_0 \rho(z; \psi)\frac{dV}{dz}P(\boldsymbol{\zeta}, \varphi|z, \psi) \\ &\times P(S|\boldsymbol{\zeta}, \varphi, z, \eta, \mathcal{O})d\boldsymbol{\zeta}d\varphi dz. \end{aligned} \tag{A8}$$

The parameters, redshifts, measurement uncertainties, and data vectors for each (selected) galaxy are then drawn from their respective sampling distributions, conditioned on selection. The generative model is hence given by

$$\begin{aligned} &P(\psi, \Phi_0, \eta, \{\varphi, \boldsymbol{\zeta}, z\}_{1:N}, \{\boldsymbol{d}, \boldsymbol{D}, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, S\}_{1:N}, N|\mathcal{O}) \\ &= P(\psi)P(\eta)P(\Phi_0)\frac{\bar{N}(\Phi_0, \psi, \eta)^N e^{-\bar{N}(\Phi_0, \psi, \eta)}}{N!} \\ &\times \prod_{i=1}^{N} P(\varphi_i, \boldsymbol{\zeta}_i, z_i|\psi, \boldsymbol{\sigma}_i, \boldsymbol{\Sigma}_i, S_i, \eta) \\ &\times P(\boldsymbol{d}_i|\varphi_i, z_i, \boldsymbol{\sigma}_i, S_{d,i}, \eta)P(\boldsymbol{D}_i|\boldsymbol{\zeta}_i, \boldsymbol{\Sigma}_i, S_{D,i}, \eta) \\ &\times P(\boldsymbol{\sigma}_i|\boldsymbol{f}_i, \mathcal{O}, \psi, \eta, S_i)P(\boldsymbol{\Sigma}_i|\mathcal{O}, \psi, \eta, S_i), \end{aligned} \tag{A9}$$

where we have assumed that the errors on the photometry and images are uncorrelated. We note that while it is true that the fluxes are also summary statistics extracted from the images, the other summary statistics $\boldsymbol{D}$ extracted from the images on which cuts are made (e.g., image quality flags) are likely to characterize very different features of the raw pixelized galaxy image, so the assumption that their errors are uncorrelated is probably reasonable.

Taking a log-uniform prior for the present-day volume density and marginalizing out Φ_0 gives

$$\begin{aligned}
&P(\psi, \boldsymbol{\eta}, \{\varphi, \zeta, z\}_{1:N}, \{\boldsymbol{d}, \boldsymbol{D}, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, S\}_{1:N}, N|\mathcal{O})\\
&= N^{-1} P(\psi)P(\boldsymbol{\eta}) \prod_{i=1}^{N}\\
&\times P(\varphi_i, \zeta_i, z_i|\psi, \boldsymbol{\sigma}_i, \boldsymbol{\Sigma}_i, S_i, \boldsymbol{\eta})P(\boldsymbol{d}_i|\varphi_i, z_i, \boldsymbol{\sigma}_i, S_{d,i}, \boldsymbol{\eta})\\
&\times P(\boldsymbol{D}_i|\zeta_i, \boldsymbol{\Sigma}_i, S_{D,i}, \boldsymbol{\eta})P(\boldsymbol{\sigma}_i|\boldsymbol{f}_i, \mathcal{O}, \psi, \boldsymbol{\eta}, S_i)\\
&\times P(\boldsymbol{\Sigma}_i|\mathcal{O}, \psi, \boldsymbol{\eta}, S_i).
\end{aligned} \tag{A10}$$

Using the chain rule, the population model and likelihood terms can be rewritten as

$$\begin{aligned}
&P(\varphi, \zeta, z|\psi, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, S, \boldsymbol{\eta})\\
&= \frac{P(\varphi, \zeta, z|\psi)P(S|\varphi, \zeta, z, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, \boldsymbol{\eta})}{P(S|\psi, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, \boldsymbol{\eta})},\\
&P(\boldsymbol{d}|\varphi, z, \boldsymbol{\sigma}, S_d, \boldsymbol{\eta})\\
&= \frac{P(\boldsymbol{d}|\varphi, z, \boldsymbol{\sigma}, \boldsymbol{\eta})P(S_d|\boldsymbol{d}, \boldsymbol{\sigma})}{P(S_d|\varphi, z, \boldsymbol{\sigma}, \boldsymbol{\eta})}\\
&P(\boldsymbol{D}|\zeta, \boldsymbol{\Sigma}, S_D, \boldsymbol{\eta})\\
&= \frac{P(\boldsymbol{D}|\zeta, \boldsymbol{\Sigma}, \boldsymbol{\eta})P(S_D|\boldsymbol{D}, \boldsymbol{\Sigma})}{P(S_D|\zeta, \boldsymbol{\Sigma}, \boldsymbol{\eta})}.
\end{aligned} \tag{A11}$$

Inserting these into Equation (A10), we obtain (after cancellation)

$$\begin{aligned}
&P(\psi, \boldsymbol{\eta}, \{\varphi, \zeta, z\}_{1:N}, \{\boldsymbol{d}, \boldsymbol{D}, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, S\}_{1:N}, N|\mathcal{O})\\
&= N^{-1}P(\psi)P(\boldsymbol{\eta}) \prod_{i=1}^{N} P(\varphi_i, \zeta_i, z_i|\psi)\\
&\times P(\boldsymbol{d}_i|\varphi_i, z_i, \boldsymbol{\sigma}_i, \boldsymbol{\eta})P(\boldsymbol{D}_i|\zeta_i, \boldsymbol{\Sigma}_i, \boldsymbol{\eta})\\
&\times \frac{P(S_{d,i}|\boldsymbol{d}_i, \boldsymbol{\sigma}_i)P(S_{D,i}|\boldsymbol{D}_i, \boldsymbol{\Sigma}_i)}{P(S_i|\psi, \boldsymbol{\sigma}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\eta})}\\
&\times P(\boldsymbol{\sigma}_i|\mathcal{O}, \psi, \boldsymbol{\eta}, S_i)P(\boldsymbol{\Sigma}_i|\mathcal{O}, \psi, \boldsymbol{\eta}, S_i).
\end{aligned} \tag{A12}$$

For the joint modeling of both photometry and images, the joint posterior is hence given by

$$\begin{aligned}
&P(\psi, \boldsymbol{\eta}, \{\varphi, \zeta, z\}_{1:N}|\{\boldsymbol{d}, \boldsymbol{D}, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, S\}_{1:N}, N, \mathcal{O})\\
&= P(\psi)P(\boldsymbol{\eta}) \prod_{i=1}^{N} P(\varphi_i, \zeta_i, z_i|\psi)P(\boldsymbol{d}_i|\varphi_i, z_i, \boldsymbol{\sigma}_i, \boldsymbol{\eta})P(\boldsymbol{D}_i|\zeta_i, \boldsymbol{\Sigma}_i, \boldsymbol{\eta}) \times \frac{1}{P(S_i|\psi, \boldsymbol{\sigma}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\eta})},
\end{aligned} \tag{A13}$$

where the selection term in the denominator is given by

$$\begin{aligned}
&P(S|\psi, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, \boldsymbol{\eta})\\
&= \int P(S_D|\boldsymbol{D}, \boldsymbol{\Sigma})P(S_d|\boldsymbol{d}, \boldsymbol{\sigma})P(\boldsymbol{d}|\varphi, z, \boldsymbol{\sigma}, \boldsymbol{\eta})P(\boldsymbol{D}|\zeta, \boldsymbol{\Sigma}, \boldsymbol{\eta})\\
&\quad \times P(\varphi, \zeta, z|\psi)d\boldsymbol{d}\, d\boldsymbol{D}\, d\varphi\, d\zeta\, dz,
\end{aligned} \tag{A14}$$

and, as before, we have dropped the $P(\boldsymbol{\sigma}_i|\mathcal{O}, \psi, \boldsymbol{\eta}, S_i)$ and $P(\boldsymbol{\Sigma}_i|\mathcal{O}, \psi, \boldsymbol{\eta}, S_i)$ terms, based on the notion that their parameter sensitivity will be negligible compared to the likelihood and prior terms.

Now, in order to determine the effect of image-based cuts when performing inference with respect to photometry alone, we need to marginalize over the image parameters, image data vectors, and their uncertainties. Taking the joint generative model in Equation (A12) and marginalizing over the image data vectors $\boldsymbol{D}_{1:N}$, one obtains

$$\begin{aligned}
&P(\psi, \boldsymbol{\eta}, \{\varphi, \zeta, z\}_{1:N}, \{\boldsymbol{d}, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, S\}_{1:N}, N|\mathcal{O})\\
&= N^{-1}P(\psi)P(\boldsymbol{\eta}) \prod_{i=1}^{N} P(\varphi_i, \zeta_i, z_i|\psi)P(\boldsymbol{d}_i|\varphi_i, z_i, \boldsymbol{\sigma}_i, \boldsymbol{\eta})\\
&\times \frac{P(S_{d,i}|\boldsymbol{d}_i, \boldsymbol{\sigma}_i)P(S_D|\zeta_i, \boldsymbol{\Sigma}_i, \boldsymbol{\eta})}{P(S_i|\psi, \boldsymbol{\sigma}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\eta})}\\
&\times P(\boldsymbol{\sigma}_i|\boldsymbol{f}_i, \mathcal{O}, \psi, \boldsymbol{\eta}, S_i)P(\boldsymbol{\Sigma}_i|\mathcal{O}, \psi, \boldsymbol{\eta}, S_i).
\end{aligned} \tag{A15}$$

We can then rewrite the selection term in the denominator as

$$P(S|\psi, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, \boldsymbol{\eta}) = P(S_d|S_D, \psi, \boldsymbol{\sigma}, \boldsymbol{\eta})P(S_D|\psi, \boldsymbol{\Sigma}, \boldsymbol{\eta}), \tag{A16}$$

and absorb $P(S_D|\psi, \boldsymbol{\Sigma}, \boldsymbol{\eta})$ and $P(S_D|\zeta, \boldsymbol{\Sigma}, \boldsymbol{\eta})$ into the population prior term (again using the chain rule), to give

$$\begin{aligned}
&P(\psi, \boldsymbol{\eta}, \{\varphi, \zeta, z\}_{1:N}, \{\boldsymbol{d}, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, S\}_{1:N}, N|\mathcal{O})\\
&= N^{-1}P(\psi)P(\boldsymbol{\eta}) \prod_{i=1}^{N} P(\varphi_i, \zeta_i, z_i|\psi, S_{D,i}, \boldsymbol{\Sigma}_i, \boldsymbol{\eta})\\
&\times P(\boldsymbol{d}_i|\varphi_i, z_i, \boldsymbol{\sigma}_i, \boldsymbol{\eta}) \frac{P(S_{d,i}|\boldsymbol{d}_i, \boldsymbol{\sigma}_i)}{P(S_{d,i}|S_{D,i}, \psi, \boldsymbol{\sigma}_i, \boldsymbol{\eta})}\\
&\times P(\boldsymbol{\sigma}_i|\boldsymbol{f}_i, \mathcal{O}, \psi, \boldsymbol{\eta}, S_{d,i})P(\boldsymbol{\Sigma}_i|\mathcal{O}, \psi, \boldsymbol{\eta}, S_{D,i}).
\end{aligned} \tag{A17}$$

Marginalizing over the image uncertainties $\boldsymbol{\Sigma}_{1:N}$ and the parameters governing the galaxy images $\boldsymbol{\zeta}_{1:N}$ then gives

$$\begin{aligned}
&P(\psi, \boldsymbol{\eta}, \{\varphi, z\}_{1:N}, \{\boldsymbol{d}, \boldsymbol{\sigma}, S\}_{1:N}, N|\mathcal{O})\\
&= N^{-1}P(\psi)P(\boldsymbol{\eta}) \prod_{i=1}^{N} P(\varphi_i, z_i|\psi, S_{D,i}, \boldsymbol{\eta}, \mathcal{O})P(\boldsymbol{d}_i|\varphi_i, z_i, \boldsymbol{\sigma}_i, \boldsymbol{\eta}) \frac{P(S_{d,i}|\boldsymbol{d}_i, \boldsymbol{\sigma}_i)}{P(S_{d,i}|S_{D,i}, \psi, \boldsymbol{\sigma}_i, \boldsymbol{\eta})}P(\boldsymbol{\sigma}_i|\boldsymbol{f}_i, \mathcal{O}, \psi, \boldsymbol{\eta}, S_{d,i}).
\end{aligned} \tag{A18}$$

Dropping parameter-independent terms (and assuming that the parameter sensitivity of $P(\boldsymbol{\sigma}_i|\boldsymbol{f}_i, \mathcal{O}, \boldsymbol{\psi}, \boldsymbol{\eta}, S_{d,i})$ is negligible relative to the prior and likelihood terms) then gives the posterior

$$P(\boldsymbol{\psi}, \boldsymbol{\eta}, \{\boldsymbol{\varphi}, z\}_{1:N}, |\{\boldsymbol{d}, \boldsymbol{\sigma}, S\}_{1:N}, N, \mathcal{O}) \propto P(\boldsymbol{\psi})P(\boldsymbol{\eta}) \prod_{i=1}^{N} P(\boldsymbol{\varphi}_i, z_i|\boldsymbol{\psi}, S_D, \boldsymbol{\eta}, \mathcal{O})P(\boldsymbol{d}_i|\boldsymbol{\varphi}_i, z_i, \boldsymbol{\sigma}_i, \boldsymbol{\eta}) \frac{1}{P(S_d|S_D, \boldsymbol{\psi}, \boldsymbol{\sigma}_i, \boldsymbol{\eta})}, \tag{A19}$$

where the selection term in the denominator is given by

$$\begin{aligned} &P(S_d|S_D, \boldsymbol{\psi}, \boldsymbol{\sigma}, \boldsymbol{\eta}) \\ &= \int P(S_d|\boldsymbol{d}, \boldsymbol{\sigma})P(\boldsymbol{d}|\boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta}) \\ &\times P(\boldsymbol{\varphi}, \boldsymbol{\zeta}, z|\boldsymbol{\psi}, S_D, \boldsymbol{\eta}, \mathcal{O}) d\boldsymbol{d}d\boldsymbol{\varphi}dz. \end{aligned} \tag{A20}$$

Crucially, note how this posterior has exactly the same form as in Equation (A6), but the image-based selection cuts have been completely absorbed into an effective population prior model for the galaxies that pass the image-based selection cuts. Those selection effects will not be well specified without assuming a detailed model for the joint distribution of SPS parameters and galaxy images; therefore, they will typically need parameterizing and inferring.

Appendix B
SFS Model

Leja et al. (2022) use a normalizing flow to model the SFS, i.e., the distribution of the SFR conditioned on mass and redshift. While their flow provides a state-of-the-art measurement of the SFS, their model utilized a dummy variable that needs integrating over in order to compute the log-probability $P(\mathrm{SFR} \mid M, z)$, and it has no explicit constraints to ensure that it extrapolates sensibly below the mass-complete limit for the data that it is fitted to. To address these two issues, we construct a surrogate model to approximate their normalizing flow. The SFS is modeled as a mixture of a Gaussian, characterizing star-forming galaxies, and a SinhArcSinh distribution, characterizing quiescent galaxies:

$$\begin{aligned} &P(\log_{10}\mathrm{SFS}|\mathcal{M}, z) \\ &= r_{\mathrm{q}}(\mathcal{M}, z)\mathcal{S}(\log_{10}\mathrm{SFS}|\mu_{\mathrm{q}}(\mathcal{M}, z), \sigma_{\mathrm{q}}(\mathcal{M}, z), k_{\mathrm{q}}) \\ &+(1 - r_{\mathrm{q}}(\mathcal{M}, z))\,\mathcal{N}(\log_{10}\mathrm{SFS}|\mu_{\mathrm{sf}}(\mathcal{M}, z), \sigma_{\mathrm{sf}}), \end{aligned} \tag{B1}$$

where $\mathcal{M} \equiv \log_{10} M$, $\mathcal{N}$ denotes the normal distribution and $\mathcal{S}$ denotes the SinhArcSinh distribution, defined as a bijection of the unit normal distribution: $x \rightarrow \mu + \sigma \sinh^{-1}(\sinh(x) + k)$.

The location and scale of the SinhArcSinh, the relative weight of the two mixture components, and the location of the Gaussian are functions of mass and redshift, defined by

$$\begin{aligned} \mu_{\mathrm{sf}}(\mathcal{M}, z) &= [a(z)\Theta(\mathcal{M} - \mathcal{M}^*_{\mathrm{sf}}(z)) + b(z)\Theta(\mathcal{M}^*_{\mathrm{sf}}(z) - \mathcal{M})](\mathcal{M} - \mathcal{M}^*_{\mathrm{sf}}(z)) + c(z) \\ \mu_{\mathrm{q}}(\mathcal{M}, z) &= [d(z)\Theta(\mathcal{M} - \mathcal{M}^*_{\mathrm{sf}}(z)) + e(z)\Theta(\mathcal{M}^*_{\mathrm{sf}}(z) - \mathcal{M})](\mathcal{M} - \mathcal{M}^*_{\mathrm{sf}}(z)) + f(z) - 1 \\ \sigma_{\mathrm{q}}(\mathcal{M}, z) &= \sigma_{q0} + (\sigma_{q0} - \sigma_{q1})\mathrm{sigmoid}((\mathcal{M} - \sigma_{q2})/\sigma_{q3}) \\ r_{\mathrm{q}}(\mathcal{M}, z) &= \mathrm{sigmoid}(j) * \mathrm{sigmoid}((\mathcal{M} - (g + h * z))/i), \end{aligned} \tag{B2}$$

Table 3
Fitted Values of the Surrogate SFS Model

Parameter	Fitted Value
(a_0, a_1, a_2)	(−0.15040097, 0.9800668, −0.50802046)
(b_0, b_1, b_2)	(1.0515388, − 0.28611764, 0.02131329)
(c_0, c_1, c_2)	(0.05053138, 1.0766244, − 0.02015052)
(d_0, d_1, d_2)	(−0.13125503, 0.7205097, −0.18212801)
(e_0, e_1, e_2)	(1.5429502, − 1.5872463, −0.04843145)
(f_0, f_1, f_2)	(0.65359867, 0.92735046, −0.17695354)
(g, h, i, j)	(10.442122, 0.56389964, 0.7500511, 2.0604856)
$(\mathcal{M}^*_{q0},\mathcal{M}^*_{q1},\mathcal{M}^*_{q2})$	(10.611108, 0.08009169, −0.06575003)
$(\mathcal{M}^*_{sf0},\mathcal{M}^*_{sf1},\mathcal{M}^*_{sf2})$	(10.335057, −0.3050156, 0.5491848)
$(\sigma_{q0}, \sigma_{q1}, \sigma_{q2}, \sigma_{q3})$	(0.54855245, 0.44964817, 11.159543, 0.11614972)
σ_{sf}	0.3912887
k_q	−1.5658572

where the functions $a(z)$ through $f(z)$, as well as $\mathcal{M}^*_{sf}(z)$ and $\mathcal{M}^*_q(z)$, are quadratic functions in redshift, defined as $a(z) = a_0 + a_1 z + a_2 z^2$, etc.

We fit this parametric form to the normalizing flow of Leja et al. (2022) by generating 10^6 samples from the flow and maximizing the total log-likelihood of those samples under our parametric model. Optimization is performed with `Adam` (Kingma & Ba 2014), with a learning rate of 10^{-4}, no minibatching, and 10^4 epochs. The fitted parameters are given in Table 3.

Appendix C
Setting Priors on Derived Quantities using Normalizing Flows

As described in 3.2, we want to put a prior on the SFH parameters, such that the resulting prior on the SFR is given by our desired assumptions about the SFS. To recap from the main text, we construct a prior over the SFH parameters as

$$P(\alpha, \beta, \tau|M, z) = \frac{\pi_0(\alpha, \beta, \tau)\, P(\mathrm{SFR}|M, z)}{P_0(\mathrm{SFR}|z)}, \tag{C1}$$

where $\pi_0(\alpha, \beta, \tau)$ is the baseline (uniform) prior on the SFH parameters, $P(\mathrm{SFR} \mid M, z)$ is the target SFS prior, and $P_0(\mathrm{SFR} \mid z)$ is the implicit prior on the SFR implied by π_0. The implicit SFR prior is defined by the surface integral

$$P_0(\mathrm{SFR}|z) = \int_{\mathrm{SFR=const.}} \pi_0(\alpha, \beta, \tau) dS, \tag{C2}$$

where dS is the surface element in the SFH parameter space (α, β, τ). In order to circumvent having to compute those surface integrals directly every time that we need to evaluate the prior density, we train a normalizing flow to learn $P_0(\mathrm{SFR} \mid z)$ as follows. We construct a training set by drawing SFH parameters from the baseline prior $\pi_0(\alpha, \beta, \tau)$, which we take as log-uniform in α and β, and uniform in τ, over the prior ranges given in Table 2 and redshifts drawn uniformly between 0 and 2. For each baseline prior (and redshift) sample, we compute the specific SFR for those SFH and redshift parameters, defined as the fractional mass formed in the last 100 Myr. This provides a training set of $\{\mathrm{sSFR}, \alpha, \beta, \tau, z\}$, on which we can train a conditional density estimator to learn $P_0(\mathrm{sSFR} \mid z)$.

We train a neural spline flow (Durkan et al. 2019) to learn $P_0(\log_{10}\mathrm{sSFR}|z)$, with 16 spline knots spaced between −15 and −6 (in $\log_{10}\mathrm{sSFR}$), a single hidden layer of 16 units and leaky ReLU activation functions, and a base Gaussian density with location −11 and scale 1. Training is performed using `Adam`, with a learning rate of $1e-2$ for 1000 epochs, with no batching on a training set of 10^6 samples (generated as described above). The implicit prior $P_0(\log_{10}\mathrm{sSFR}|z)$ and the trained normalizing flow is shown in Figure 12, and the use of

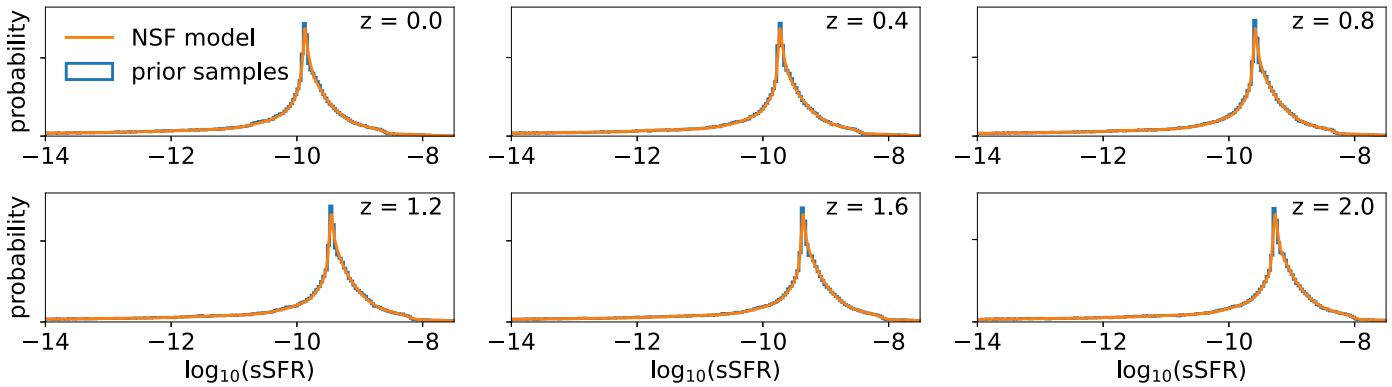

Figure 12. Samples from the prior on the SFR implied by taking uniform priors over the double-power-law SFH parameters ($\log_{10}\alpha$, $\log_{10}\beta$, τ), over the ranges specified in Table 2, are shown as the blue histograms, while the learned normalizing flow model for the implied distribution (as a function of redshift) is shown in orange.

the trained normalizing flow to divide out the implicit SFR prior is shown in Figure 4.

ORCID iDs

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
Boris Leistedt https://orcid.org/0000-0002-3962-9274

References

Abell, P. A., Allison, J., Anderson, S. F., et al. 2009, arXiv:0912.0201
Abramson, L. E., Gladders, M. D., Dressler, A., et al. 2015, ApJL, 801, L12
Ade, P. A., Aghanim, N., Arnaud, M., et al. 2016, A&A, 594, A13
Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S8
Alarcon, A., Hearin, A. P., Becker, M. R., & Chaves-Montero, J. 2023, MNRAS, 518, 562
Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. 2019, MNRAS, 488, 4440
Alsing, J., Peiris, H., Leja, J., et al. 2020, ApJS, 249, 5
Alsing, J., Wandelt, B., & Feeney, S. 2018, MNRAS, 477, 2874
Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140
Arnouts, S., & Ilbert, O. 2011, LePHARE: Photometric Analysis for Redshift Estimate, Astrophysics Source Code Library, ascl:1108.009
Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, A&A, 645, A104
Benitez, N. 2000, ApJ, 536, 571
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Buchs, R., Davis, C., Gruen, D., et al. 2019, MNRAS, 489, 820
Bustamante, S., Ellison, S. L., Patton, D. R., & Sparre, M. 2020, MNRAS, 494, 3469
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
Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, ApJ, 429, 582
Carnall, A. C., Leja, J., Johnson, B. D., et al. 2019, ApJ, 873, 44
Charlot, S., & Fall, S. M. 2000, ApJ, 539, 718
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Conroy, C. 2013, ARA&A, 51, 393
Conroy, C., & Gunn, J. E. 2010, FSPS: Flexible Stellar Population Synthesis, Astrophysics Source Code Library, ascl:1010.043
Cresci, G., Mannucci, F., & Curti, M. 2019, A&A, 627, A42
Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020, MNRAS, 491, 944
Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156
Davis, C., Rozo, E., Roodman, A., et al. 2018, MNRAS, 477, 2196
De Jong, J. T., Kleijn, G. A. V., Boxhoorn, D. R., et al. 2015, A&A, 582, A62
Dotter, A. 2016, ApJS, 222, 8
Driver, S. P., Bellstedt, S., Robotham, A. S. G., et al. 2022, MNRAS, 513, 439
Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, MNRAS, 413, 971
Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. 2019, Advances in Neural Information Processing Systems 32, ed. H. Wallach et al. (Red Hook, NY: Curran Associates), https://proceedings.neurips.cc/paper/2019/hash/7ac71d433f282034e088473244df8c02-Abstract.html
Ferland, G., Porter, R., Van Hoof, P., et al. 2013, RMxAA, 49, 137
Flaugher, B. 2005, IJMPA, 20, 3121
Gatti, M., Vielzeuf, P., Davis, C., et al. 2018, MNRAS, 477, 1664
Gerardi, F., Feeney, S. M., & Alsing, J. 2021, PhRvD, 104, 083531
Hartley, W. G., Chang, C., Samani, S., et al. 2020, MNRAS, 496, 4769
Hildebrandt, H., Choi, A., Heymans, C., et al. 2016, MNRAS, 463, 635
Hildebrandt, H., Erben, T., Kuijken, K., et al. 2012, MNRAS, 421, 2355
Hildebrandt, H., Köhlinger, F., Van den Busch, J., et al. 2020, A&A, 633, A69
Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454
Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, MNRAS, 478, 592
Ilbert, O., Arnouts, S., McCracken, H., et al. 2006, A&A, 457, 841
Jeffrey, N., & Wandelt, B. D. 2020, arXiv:2011.05991
Kaasinen, M., Kewley, L., Bian, F., et al. 2018, MNRAS, 477, 5568
Karim, A., Schinnerer, E., Martínez-Sansigre, A., et al. 2011, ApJ, 730, 61
Kashino, D., Renzini, A., Silverman, J. D., & Daddi, E. 2016, ApJL, 823, L24
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14
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
Leitner, S. N. 2012, ApJ, 745, 149
Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2019, ApJ, 876, 3
Leja, J., Johnson, B. D., Conroy, C., Dokkum, P. G. v., & 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
Leslie, S. K., Schinnerer, E., Liu, D., et al. 2020, ApJ, 899, 58
Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, MNRAS, 390, 118
Mandelbaum, R., Eifler, T., Hložek, R., et al. 2018, arXiv:1809.01669
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
Nagaraj, G., Forbes, J. C., Leja, J., Foreman-Mackey, D., & Hayward, C. C. 2022, ApJ, 932, 54
Nakajima, K., & Ouchi, M. 2014, MNRAS, 442, 900
Newman, J. A. 2008, ApJ, 684, 88
Newman, J. A., & Gruen, D. 2022, ARA&A, 60, 363
Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJL, 660, L43
Paxton, B., Bildsten, L., Dotter, A., et al. 2010, 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
Price, S. H., Kriek, M., Brammer, G. B., et al. 2014, ApJ, 788, 86
Ramachandra, N., Chaves-Montero, J., Alarcon, A., et al. 2022, MNRAS, 515, 1927
Reddy, N. A., Kriek, M., Shapley, A. E., et al. 2015, ApJ, 806, 259
Renzini, A., & Peng, Y.-j. 2015, ApJL, 801, L29
Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, ApJL, 739, L40
Salim, S., Lee, J. C., Dave, R., & Dickinson, M. 2015, ApJ, 808, 25
Salim, S., Lee, J. C., Ly, C., et al. 2014, ApJ, 797, 126
Salim, S., & Narayanan, D. 2020, ARA&A, 58, 529
Schmidt, S. J., Menard, B., Scranton, R., Morrison, C., & McBride, C. K. 2013, MNRAS, 431, 3307
Schneider, M., Knox, L., Zhan, H., & Connolly, A. 2006, ApJ, 651, 14
Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, A&A, 575, A74
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Talbot, C., & Thrane, E. 2022, ApJ, 927, 76
Tanaka, M. 2015, ApJ, 801, 20
Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2018, PASJ, 70, S9
Telford, O. G., Dalcanton, J. J., Skillman, E. D., & Conroy, C. 2016, ApJ, 827, 35
Tomczak, A. R., Quadri, R. F., Tran, K.-V.-H., et al. 2016, ApJ, 817, 118
Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104
Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, ApJL, 754, L29
Wright, A. H., Hildebrandt, H., Van den Busch, J. L., & Heymans, C. 2020, A&A, 637, A100
Yabe, K., Ohta, K., Akiyama, M., et al. 2015, PASJ, 67, 102
Yates, R. M., Kauffmann, G., & Guo, Q. 2012, MNRAS, 422, 215