Forward modeling of galaxy populations for cosmological redshift distribution inference

Justin Alsing,[1] Hiranya Peiris,[2, 1] Daniel Mortlock,[3, 4, 1] 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
[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, USA
[7] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA

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 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 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) and Vimos VLT Deep (VVDS) surveys, validating against their spectroscopic redshifts. Our baseline model is able to predict tomographic redshift distributions for GAMA and VVDS with a bias of $\Delta z \lesssim 0.003$ and $\Delta z \simeq 0.01$ on the mean redshift respectively – comfortably accurate enough for Stage III cosmological surveys – *without any hyper-parameter tuning* (i.e., prior to doing any fitting to those data). We anticipate that with additional hyper-parameter fitting and modeling improvements, forward modeling can provide a path to accurate redshift distribution inference for Stage IV surveys.

Keywords: photometric redshifts - galaxy surveys - cosmological parameters

1. INTRODUCTION

Accurate inference of the redshift distributions of ensembles of galaxies from their photometry is of central importance for deriving cosmological constraints from weak lensing surveys. Ongoing and upcoming surveys such as the Dark Energy Survey (Dark Energy Survey Collaboration: Fermilab & Flaugher 2005), the Kilo-Degree Survey (De Jong et al. 2015), 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 makes them increasingly sensitive to systematic biases, with systematics on the redshift distributions in particular expected to be 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 upcoming Stage IV surveys, the characterization of the redshift

Corresponding author: Justin Alsing
justin.alsing@fysik.su.se

arXiv:2207.05819v2 [astro-ph.CO] 17 Nov 2022

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; Hoyle et al. 2018; Tanaka et al. 2018; Leistedt et al. 2016, 2019).

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 re-weighting in the (broad-band) colors (Buchs et al. 2019; Hartley et al. 2020), and 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 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 incapability 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 volume populated by unphysical galaxy spectra. Further, by relying on a model 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 distribution of physical characteristics and redshifts) defines 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, then 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 physical characteristics (e.g., stellar, dust and gas content) of galaxies, encoding galaxy evolution physics in the relationships between galaxies physical properties. We use a stellar population synthesis (SPS) model to connect those physical parameters to the rest-frame spectra and hence photometry for each galaxy. The observation process is then characterized by a data-model that captures measurement noise, heterogeneous observing conditions and strategy, and photometric calibration. Finally, we have a selection model specifying the selection cuts. This parameterized forward model then forms the basis for Bayesian inference of cosmological redshift distributions, either by hierarchical inference, or 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, explicitly treating selection effects, and inferring population- and data-model parameters in a self-consistent fashion. In particular, in replacing finite template-sets by a continuous model we are able to better capture the diversity of real galaxy spectra whilst having full control over the prior describing the statistical properties of the galaxy population. The use of SPS models for 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 Ra-

machandra et al. 2021) and continuous physical models for galaxy spectra (Ramachandra et al. 2021) has already led to promising improvements in photometric redshift inference for individual galaxies.

The structure of this paper is as follows. In §2 we describe the framework for forward modeling photometric surveys, and estimating redshift distributions in a forward modeling context. In §3 we describe our SPS model, galaxy population model, and data model assumptions. In §4 and 5 we show the ability of our baseline forward model to recover the tomographic redshift distributions for GAMA and VVDS respectively. We outline a roadmap for future forward modeling efforts and conclude in §7.

In a companion paper (Leistedt et al. 2022) we validate the forward modeling framework for inferring individual galaxy redshifts, including hierarchical calibration of hyper-parameters.

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 a simulation-based inference setting, used to estimate the implied tomographic redshift distributions for a given set of modeling and hyper-parameter choices, or can be used as a basis for hierarchical inference via MCMC sampling.

Notation is summarized in §2.1 and Table 1. We describe the generative model in §2.2, and discuss how to perform inference of the model parameters in §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 a stellar population synthesis model, which computes the composite spectrum from the stars in the population given their initial mass function, ages and metallicities from the star formation and metallicity histories, plus modifications due to dust, and nebular emission (see Conroy 2013 for a review).

Combined with redshifts z, the SPS parameters predict the model photometry for each galaxy, i.e., the fluxes $\{f_b\}$ in band-passes $\{W_b(\lambda)\}$, 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 $d_L(z)$ the luminosity distance for redshift z and $\tau(z, \lambda)$ is the optical depth of the inter-galactic medium.

We denote the vector of measured fluxes for each galaxy by $\mathbf{d}$, where the photometric *data-model* specifies the sampling distribution $P(\mathbf{d}|\boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta})$ of measured fluxes given the model fluxes and measurement uncertainties $\boldsymbol{\sigma}$. The data model is parameterized by 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 galaxy characteristics (SPS parameters and redshift) is described by a *population model* $P(\boldsymbol{\varphi}, z|\boldsymbol{\psi})$, with hyper-parameters $\boldsymbol{\psi}$. This describes the joint distribution of galaxy characteristics for the background galaxy population, in the absence of any selection effects.

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

Subsequent sorting of galaxies into tomographic bins introduces an additional selection $S^{(k)}$ (for the k-th 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:

$$P(S^{(k)}|\hat{z}(\mathbf{d}, \boldsymbol{\sigma})) = \begin{cases} 1 & z_l^{(k)} < \hat{z}(\mathbf{d}, \boldsymbol{\sigma}) < z_u^{(k)} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $\hat{z}(\mathbf{d}, \boldsymbol{\sigma})$ is an estimator for the redshift, and $z_l^{(k)}$ and $z_u^{(k)}$ and the lower and upper limits for the k-th tomographic bin.

The measurement uncertainties on the photometry will be 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 photometric uncertainties across the survey (as a function of flux) by the *uncertainty model* $P(\boldsymbol{\sigma}|\mathbf{f}(\boldsymbol{\varphi}, z), \mathcal{O})$, 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* $P(\boldsymbol{\varphi}, z|\boldsymbol{\psi})$;

Parameter	Description
Population-model parameters	
$\boldsymbol{\psi}$	Hyper-parameters describing the galaxy population model
Φ_0	Present-day comoving volume-density of galaxies
$\rho(z; \boldsymbol{\psi})$	Evolution in the relative comoving number-density of galaxies ($\rho(0; \boldsymbol{\psi}) = 1$)
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 rest-frame spectrum (per galaxy)
$z_{1:N}$	Redshift (per galaxy)
Derived quantities	
$\bar{N} = \bar{N}(\Phi_0, \boldsymbol{\psi}, \boldsymbol{\eta}, \mathcal{O})$	Expected number of selected galaxies given the population, selection and data models
$\mathbf{f}_{1:N} = \mathbf{f}(\boldsymbol{\varphi}_{1:N}, z_{1:N})$	Model fluxes determined by SPS model (per galaxy)
Data	
$\mathbf{d}_{1:N}$	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)
$S_{1:N}^{(k)}$	Color-based selection into tomographic bin k (per galaxy)

Table 1. Notation for all model parameters included in the forward model.

2. Compute model photometry given the SPS parameters, using the SPS model, $\mathbf{f} \equiv \mathbf{f}_{\mathrm{SPS}}(\boldsymbol{\varphi}, z)$;

3. Draw uncertainties from the *uncertainty model* $P(\boldsymbol{\sigma}|\mathbf{f}, \mathcal{O})$;

4. Draw noisy (calibrated) photometry $\mathbf{d}$ given the model fluxes and uncertainties, from the *data-model* $P(\mathbf{d}|\mathbf{f}(\boldsymbol{\varphi}, z), \boldsymbol{\sigma}, \eta)$;

5. Apply selection cuts on the noisy photometry, where $P(S|\mathbf{d}, \boldsymbol{\sigma})$ equals one (zero) for passing (failing) selection;

6. Assign tomographic bin label based on the photo-z estimator, $\hat{z}(\mathbf{d}, \boldsymbol{\sigma})$.

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[1]. This can hence be used as a generative model for inferring the hyper-parameters via simulation-based inference, or as the basis for Bayesian hierarchical inference, as described below in §2.4.

Repeating this process in the limit of $N \to \infty$ selected samples, and examining the redshift distributions of the galaxies that make it into each bin, provides the tomographic redshift distributions implied for a given set of modeling and hyper-parameter 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 galaxies passing selection both into the analysis sample, and subsequently into the k-th tomographic

---

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

bin, is given by:

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

where $S^{(k)}$ denotes selection into both the analysis sample and k-th tomographic bin. Forward simulating from the generative model described above provides a way of estimating the target redshift distributions for a given set of hyper-parameters, without the need for direct integration.

2.3. Emulation of SPS models

Any applications of this forward model – either simulation of large mocks, or MCMC sampling the associated posterior – will require a vast number of SPS model calls. This is only made tractable by 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

Inference of redshift distributions under the generative model described in §2.2 requires inferring the population- and data-model parameters $\boldsymbol{\psi}$ and $\boldsymbol{\eta}$ under the forward model assumptions, which in turn provide marginal posteriors for the tomographic redshift distributions via Equation (3)[2].

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 highlight the advantages of performing redshift distribution inference in this fashion.

---

[2] In practice, the mapping between hyper-parameters and tomographic redshift distributions will be done via simulating large mocks, following §2.2

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

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

where the selection term in the denominator is given by:

$$
\begin{aligned}
P(S|\boldsymbol{\psi}, \boldsymbol{\sigma}, \boldsymbol{\eta}) &= \int P(S|\mathbf{d}, \boldsymbol{\sigma})P(\mathbf{d}|\boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta}) \\
&\qquad \times P(\boldsymbol{\varphi}, z|\boldsymbol{\psi})\, d\mathbf{d}\, d\boldsymbol{\varphi}\, dz.
\end{aligned}
\tag{5}
$$

This joint posterior can then be sampled (using MCMC) to jointly infer the population- and data-model parameters, and SPS parameters and redshifts for each galaxy. The marginal posterior over the hyper- 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. Firstly, 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 sub-sets 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 selection cuts are only performed with respect to the main survey data.

Similarly, 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) sub-sets of the galaxies is not important, provided selection is not performed with respect to those auxiliary data.

Note that the population model $P(\boldsymbol{\varphi}, z|\boldsymbol{\psi})$ 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) $P(\boldsymbol{\varphi}, z|\boldsymbol{\psi})$: 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 hyper-parameters ψ.

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

However, MCMC sampling 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 practise, for sampling to be computationally tractable the selection integral will require replacement with a fast emulator (e.g., Talbot & Thrane 2022).

Alternatively, simulation-based inference (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 §3.1, the galaxy population model in §3.2, and the data model in §3.3.

3.1. *Stellar population synthesis (SPS) model*

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

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

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

The metallicity history of the stellar population assumed to build up with stellar mass production, such

that present-day stellar and gas-phase metallicities are identical:

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

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

Dust attenuation is modelled with two components describing birth cloud (stars younger than ten 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 emission is 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 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 §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 results 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:

$$P(\text{SPS parameters, redshift}) =$$
$$P(\text{mass, redshift})$$
$$P(\text{metallicity} | \text{SFR, mass})$$
$$P(\text{SFR} | \text{mass, redshift})$$
$$P(\text{dust} | \text{SFR, mass, metallicity})$$
$$P(\text{age} | \text{SFR, mass}). \quad (8)$$

Parameter	Description	Limits
$\log_{10}(M/M_\odot)$	Stellar mass	$[7, 13]$
$\log_{10}Z_{\mathrm{gas}}$	Gas phase metallicity	$[-1.98, 0.5]$
$\log_{10}u$	Gas ionization parameter	$[-4, -1]$
τ_1	Birth cloud (stars younger than 10Myr) 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 star-formation history	$[10^{-3}, 10^1]$
τ	Transition-time of the double power-law star formation history, as fraction of lookback-time	$[0.007, 1]$
z	Redshift	$[0, 2.5]$

Table 2. SPS model parameters and their prior ranges.

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

For the SPS model set-up 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:

$$P(M, z) \propto \Phi(M, z)dV(z), \qquad (9)$$

where $\Phi(M, z)$ is the un-normalized mass function, and $dV(z)$ 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 show in Figure 1.

3.2.2. *Fundamental metallicity relation (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 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 star formation history of a galaxy (e.g., mass

and SFR) – the so-called fundamental metallicity relation (Yates et al. 2012; Andrews & Martini 2013; Nakajima & Ouchi 2014; Yabe et al. 2015; Salim et al. 2014, 2015; Kashino et al. 2016; Cresci et al. 2019; Curti et al. 2020). Qualitatively, galaxies on the FMR tend towards lower metallicities for higher star formation rates, and higher metallicities for higher masses, but the overall shape is a non-linear 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 being constant over cosmic history[3].

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 SDSS. The median gas-phase metallicity as a function of mass and SFR is parameterized as:

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

where

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

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[4] (with two degrees-of-freedom) for $P(\log_{10}Z_{\mathrm{gas}}|M, \mathrm{SFR})$, where the mean is given by Equation (10), and the FWHM is equal to 0.05. The FMR is shown in Figure 2.

Measurement of the FMR is sensitive to the way in which SFRs and metallicities are estimated (see e.g.,

---

[3] The physics governing chemical enrichment is assumed to be constant over cosmic time.
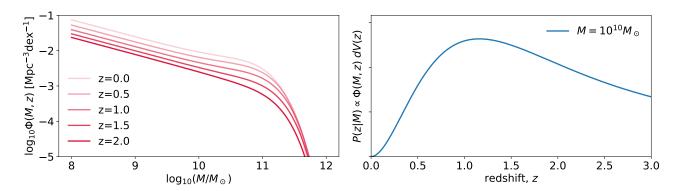
[4] Hereafter, student-t.

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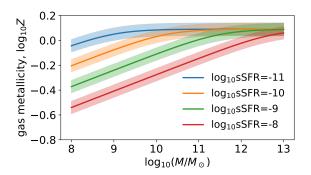

Figure 2. Fundamental metallicity relation prior on gas-phase metallicity conditioned on mass and specific star formation rate (in units of Gyr^{-1}). Solid lines represent the mean, and the bands show the FWHM of the assumed student-t distribution.

Telford et al. 2016 and Cresci et al. 2019), with the SFRs and metallicities 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 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 deriving a separate model for the metallicities of merged galaxies.

3.2.3. *Star forming sequence*

The star-forming sequence (SFS) characterizes the (redshift-evolving) relationship between star formation rate and mass, with the vast majority of galaxies forming most of their mass either on or passing through the star-forming sequence (Leitner 2012; Abramson et al. 2015). Qualitatively, the SFS is characterized by star-forming and quiescent galaxies with ongoing and negligible star formation rates respectively. The clustering of galaxies into those 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. 2021).

In this work, we base our SFS model on the measured relation from Leja et al. (2021), who fit a normalizing flow to learn $P(\mathrm{SFR}|M, z)$ from 3D HST data. The galaxy sample used in Leja et al. (2021) is mass-complete down to around $10^9 M_\odot$ over the redshift range relevant for this study ($z \leq 1.5$). We note that $P(\mathrm{SFR}|M, z)$ 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. (2021) (see Figure 4, and 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 SFH), but the star-forming sequence provides only a prior constraint on a derived quantity: the star formation rate, $\mathrm{SFR}(\alpha, \beta, \tau, z)$. 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 derived quantities
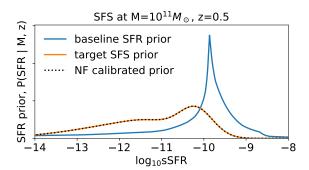
SFS at M=$10^{11} M_\odot$, z=0.5

— baseline SFR prior
— target SFS prior
⋯ NF calibrated prior

Figure 3. The prior on the star formation rate implied by taking uniform priors over the double power-law SFH parameters ($\log_{10}\alpha, \log_{10}\beta, \tau$) over the ranges specified in Table 2 is shown in blue, while the target prior (specified by the SFS measurement of Leja et al. (2021)) is shown in orange. The SFR prior obtained by re-calibrating the baseline prior with a normalizing flow (as described in 3.2.3) is shown by the black dotted line, giving excellent agreement with the target SFS prior.

such as 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 (c.f. Table 2) leads to the undesirable implicit SFR prior $P_0(\text{SFR}|z)$ shown in Figure 3.

To ensure that our SFH priors *only* encode the SFS, without a spurious additional contribution from any baseline prior assumptions, we define the SFH priors as:

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

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

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

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

Dividing out the implicit SFR prior in this way insures that the overall prior on 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 $P_0(\text{SFR}|z)$ so that it can be conveniently divided out (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 star formation histories (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):

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

where Θ is the Heaviside step function. We assume a Gaussian prior on the diffuse dust attenuation with the mean given above, and 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 mean given by:

$$\langle \delta \rangle = -0.095 + 0.111\tau_2 - 0.0066\tau_2^2, \qquad (15)$$

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 mean given above and standard deviation 0.4.

For the birth cloud component, we take a Gaussian prior on the ratio τ_1/τ_2 of the birth cloud and diffuse dust components with mean equal to 1 and standard deviation 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) star-formation rates, 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 one Gyr on the galaxy ages[5] to
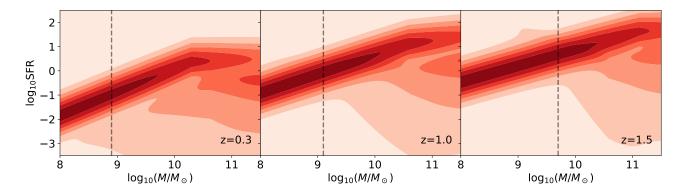
---

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

Figure 4. Star forming sequence prior on star formation rate (in units of $M_\odot \mathrm{yr}^{-1}$) conditioned on mass and redshift, based on the measurement from Leja et al. (2021) (see Appendix B for details). The vertical lines represent the mass-complete limit for the 3D HST data on which the Leja et al. (2021) fit was performed; we extrapolate the SFS below the mass limit as shown in the Figure.
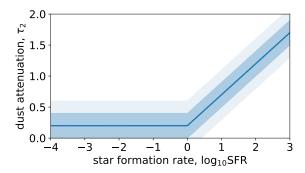

Figure 5. Prior on the diffuse dust attenuation conditioned on SFR. Solid line shows the mean, and the bands show the 1- and 2-σ contours.

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 observed fluxes as a student-t distribution with two degrees-of-freedom,

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

(16)

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

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

(17)

The data-model parameters $\boldsymbol{\eta} = (\boldsymbol{\alpha}_\mathrm{zp}, \boldsymbol{\beta})$ characterize the zero points $\boldsymbol{\alpha}_\mathrm{zp}$, 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 §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, SED modeling-error as a function of rest-frame wavelength, and parameters describing the shape (e.g., skewness or tailweight) of the data sampling distribution. A more sophisticated error-model (including hierarchical calibration of hyper-parameters) is investigated in our companion paper Leistedt et al. (2022).

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 strategy, varying difficulty in extracting photometry from galaxies 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.

We take a data-driven approach to uncertainty modeling, where we learn the distribution $P(\boldsymbol{\sigma}|\mathbf{f}, \mathcal{O})$ 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 (c.f. Equation (4) with fixed hyper-parameters). For each galaxy, this provides a maximum a posteriori (MAP) estimate for their true fluxes $\mathbf{f}$. We then take the catalog of MAP estimated fluxes and associated uncertainties $\{\mathbf{f}, \boldsymbol{\sigma}\}_{1:N}$, and train a Mixture Density Network (MDN) to learn $P(\boldsymbol{\sigma}|\mathbf{f}, \mathcal{O})$. The MDN parameterizes

the uncertainty distribution as a Gaussian mixture:

$$P(\boldsymbol{\sigma}|\mathbf{f}, \mathcal{O}) = \sum_{c=1}^{N_{\text{comp.}}} r_c(\mathbf{f}; \mathbf{w}) \mathcal{N}\left[\boldsymbol{\mu}_c(\mathbf{f}; \mathbf{w}), \boldsymbol{\Sigma}_c(\mathbf{f}; \mathbf{w})\right],$$
(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 $\mathbf{w}$). The MDN is trained by minimizing the total (negative) log-likelihood of the data under the model with respect to the network weights[6]:

$$\mathcal{L}(\mathbf{w}) = -\sum_{i=1}^{N_{\text{galaxies}}} \ln P(\boldsymbol{\sigma}_i|\mathbf{f}_i, \mathbf{w}).$$
(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 9.

4. CASE STUDY I: GAMA

The Galaxy And Mass Assembly (GAMA) survey covers 250 square degrees, and has obtained $\sim 230,000$ 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 dataset for validating our forward model. We take the GAMA data release 4 (Driver et al. 2022), with photometry in KiDS $ugri$ and VIKING $ZYHJK_s$ 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 $(J - K_s) > 0.025$ is made for star-galaxy separation. With these cuts, we are left with a sample of $206,454$ galaxies with 9-band photometry ($ugriZYHJK_s$), and measured spectroscopic redshifts (for validation).

4.1. Data model

We assume student-t uncertainties on the fluxes as described in §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 measurement uncertainties as a function of flux, trained on the GAMA data, as described in §3.4. The uncertainty distributions and corresponding trained models are shown side-by-side in Figure 6.

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 with 64 units each and leaky-ReLU activations, passing the measured magnitudes and magnitude errors in the 9-bands as inputs and the estimated redshift as output. The network is trained to minimize the mean square error on the redshift, trained on the GAMA photometry and redshifts. Training is performed with Adam using a batch size of 1024, a training:validation split of 90:10, and 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 §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 selection cuts $r < 19.65$ and $(J - K_s) > 0.025$ 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 7, and the corresponding bias on the mean of the redshift distribution in shown in Figure 8. The forward model is able to predict the redshift distributions with a bias 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[7] $\Delta z < 0.002(1 + z)$ (Mandelbaum et al. 2018).

---

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

[7] For LSST year 1 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).
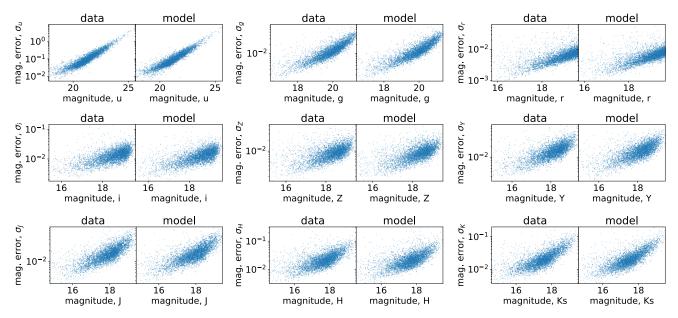
Figure 6. Magnitude uncertainties versus magnitudes for the GAMA data (left panels) versus the trained MDN model for the error distribution conditioned on flux (right panels). Note that the magnitudes in both the left and right panels are maximum a posteriori (MAP) magnitudes, from an initial fit of the SPS model to the GAMA galaxies as described in §3.4.

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

5. CASE STUDY II: VVDS

Similar to GAMA, the VIMOS VLT Deep Survey (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 obtained spectra for 25 805 galaxies down to $I < 22.5$.

We take photometry in the $BVRI$ bands for 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 (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 to obtain 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, and hence we make no attempt to model this implicit selection effect. However, Fig. 13 in Le Fèvre

et al. (2013) for the VVDS-*Deep* sample in the same magnitude range suggests that the spectroscopic success rate for this subsample is expected to be roughly uniform up to $z \simeq 1$ and drop thereafter. This only impacts $\sim 1 - 2\%$ 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.

5.1. *Data model*

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

We train a MDN to model the distribution of measurement uncertainties as a function of flux, trained on the VVDS data, as described in §3.4. Unlike GAMA, the VVDS $BVRI$ photometry is patchy with roughly a quarter of objects missing measurements in at least one band. We set the fractional uncertainties for 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 non-missing values) and corresponding trained models are shown side-by-side in Figure 9. Note the multi-modal structure in the uncertainty distributions, owing
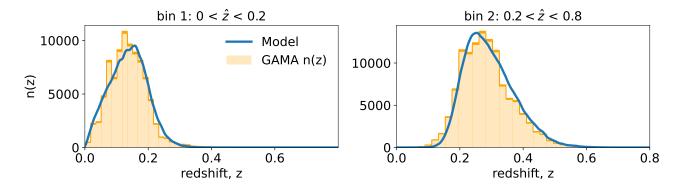
Figure 7. Tomographic redshift distributions obtained by the forward model (blue) compared to the histogram of the GAMA spectroscopic redshifts (orange). The model predictions are in excellent agreement with the distributions of the spectroscopic redshifts.
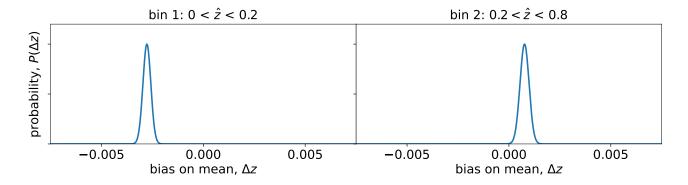

Figure 8. Bias on the mean redshift of the model redshift distributions versus the data for GAMA, $\Delta z = \langle z_{\mathrm{model}} \rangle - \langle z_{\mathrm{data}} \rangle$. The distributions are obtained by bootstrapping samples from the model $n(z)$, taking a kernel density estimate (KDE) of the bootstrapped sample means, and centering the KDE on the difference between the sample mean of the spec-zs and the mean of the model $n(z)$.

to varying depth 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 with 64 units each and leaky-ReLU activations, passing the measured magnitudes and magnitude errors in the 4-bands as inputs and the estimated redshift as output. The network is trained on the VVDS photometry and spec-zs as described in §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 comapred 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: $0 < \hat{z} \leq 0.4$; $0.4 < \hat{z} \leq 0.75$; and $0.75 < \hat{z} \leq 2$.

5.3. *Results*

As in §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 bias on the mean of the redshift distribution in shown in Figure 11.
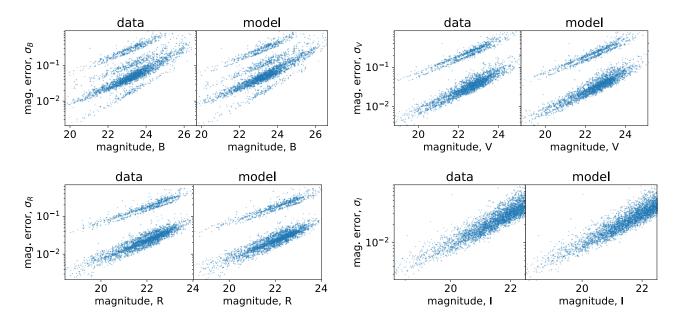
Figure 9. Magnitude errors versus magnitudes for the VVDS data (left panels) versus the trained MDN model for the error distribution conditioned on flux (right panels). Note that the magnitudes in both the left and right panels are maximum a posteriori (MAP) magnitudes, from an initial fit of the SPS model to the VVDS galaxies as described in §3.4.

.

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 re-iterate that these model predictions are obtained by assuming (fixed) default parameters for the population model described in §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 LeBrun, pri-

vate 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 star-formation history parameterization. Such a simple SFH parameterization is not expected to capture the full diversity of real star formation histories, 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 non-parametric (binned) star formation history models (e.g., Leja et al. 2019), or more physical SFH parameterizations (e.g., Alarcon et al. 2022).

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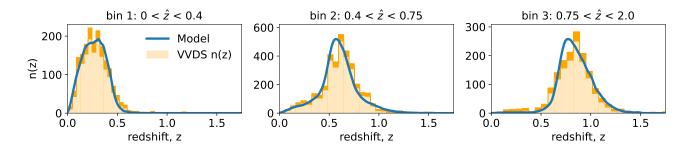
Figure 10. Tomographic redshift distributions obtained by the forward model (blue) compared to the histogram of the VVDS spectroscopic redshifts (orange). The width of the histogram bars indicates $\pm\sqrt{N}$ Poisson noise. The model predictions are in excellent agreement with the distributions of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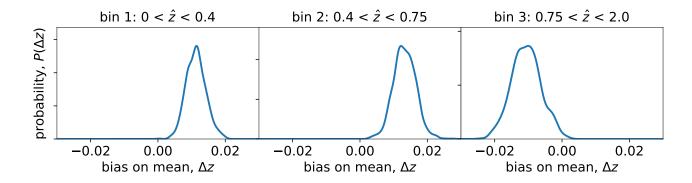

Figure 11. Bias on the mean redshift of the model redshift distributions versus the data for VVDS, $\Delta z = \langle z_{\mathrm{model}}\rangle - \langle z_{\mathrm{data}}\rangle$. The distributions are obtained by bootstrapping samples from the model $n(z)$, taking a kernel density estimate (KDE) of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

lines in particular being subject to potentially significant modeling biases (Leistedt et al. 2022). 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 inferred SPS model parameters and reduce biases. While this strategy is fine for individual galaxies, increasing variances in order to cover potential biases in this way can lead to over-dispersion in population-level parameters. Therefore, data-modeling efforts should instead focus on parameterizing and modeling 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 photometric measurement errors likely varying between bands, and as a function of flux and background noise levels.

Regarding the selection modeling, so far we have considered the 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 effect 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 effect of those image cuts on the population prior over SPS parameters and redshift, and inferring those additional hyper-parameters 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 unmodelled selection effects can be safely ignored.

In addition to modeling improvements, inference (or optimization) of population- and data-model parameters from the photometric data should lead to additional improvements in accuracy. For robust inferences, data-model parameters should be self-consistently calibrated using the photometric data themselves, with 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. The dust prior and fundamental metallicity relations in particular are expected to be the least well understood and constrained, meriting broader priors on their parameters. The star-forming sequence is somewhat better constrained (e.g., Leja et al. 2021), 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 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 depth required for photometric surveys, and is not vulnerable to biases arising from spectroscopic selection effects that cannot be well-described by re-weighting in broad-band colour-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 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 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, those 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 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 ongoing Stage III surveys, and approaching the accuracy requirements of Stage IV surveys. We anticipate that with additional modeling improvements, and optimization of model hyper-parameters, forward modeling can provide a path to accurate cosmological redshift distribution inference for Stage IV surveys.

In a companion paper (Leistedt et al. 2022), we demonstrate the utility of this forward modeling framework for inferring individual redshifts, including hierarchical calibration of data-model hyper-parameters.

APPENDIX

A. DERIVATION OF THE JOINT POSTERIOR INCLUDING SELECTION EFFECTS

For the purpose of deriving the joint posterior for the forward model described in §2.2 it is useful to consider the generation of a selected sample directly, as follows (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:

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

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:

$$P(\boldsymbol{\psi}, \Phi_0, \boldsymbol{\eta}, \{\boldsymbol{\varphi}, z\}_{1:N}, \{\mathbf{d}, \boldsymbol{\sigma}, S\}_{1:N}, N|\mathcal{O}) =$$

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

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:

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

The joint posterior is hence given by:

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

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

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

It is instead desirable to re-write 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 re-written as:

$$P(\boldsymbol{\varphi}, z|\boldsymbol{\psi}, \boldsymbol{\sigma}, S, \boldsymbol{\eta}) = \frac{P(\boldsymbol{\varphi}, z|\boldsymbol{\psi})P(S|\boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta})}{P(S|\boldsymbol{\psi}, \boldsymbol{\sigma}, \boldsymbol{\eta})},$$

$$P(\mathbf{d}|\boldsymbol{\varphi}, z, \boldsymbol{\sigma}, S, \boldsymbol{\eta}) = \frac{P(\mathbf{d}|\boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta})P(S|\mathbf{d}, \boldsymbol{\sigma})}{P(S|\boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta})}. \tag{A5}$$

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

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

where the selection term in the denominator is given by:

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

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

When inferring hyper- and data-model parameters though, the selection term is important: the high-dimensional integral over 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 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 insure robust shear measurements, 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 the 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 parameters ζ (in addition to the SPS parameters and redshift). We denote galaxy image data-vectors and uncertainties by $\mathbf{D}$ and Σ respectively, where $\mathbf{D}$ can be taken to mean either the full pixelized image, or some low-dimensional summary statistics derived from the galaxy images on which selection is performed. For the purpose of this derivation, we distinguish photometric and image-based selection cuts by 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:

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

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:

$$P(\psi, \Phi_0, \eta, \{\varphi, \zeta, z\}_{1:N}, \{\mathbf{d}, \mathbf{D}, \sigma, \Sigma, S\}_{1:N}, N|\mathcal{O}) = P(\psi)P(\eta)P(\Phi_0)\frac{\bar{N}(\Phi_0, \psi, \eta)^N e^{-\bar{N}(\Phi_0, \psi, \eta)}}{N!}$$

$$\times \prod_{i=1}^{N} P(\varphi_i, \zeta_i, z_i|\psi, \sigma_i, \Sigma_i, S_i, \eta)\, P(\mathbf{d}_i|\varphi_i, z_i, \sigma_i, S_{d,i}, \eta)P(\mathbf{D}_i|\zeta_i, \Sigma_i, S_{D,i}, \eta)\, P(\sigma_i|\mathbf{f}_i, \mathcal{O}, \psi, \eta, S_i)P(\Sigma_i|\mathcal{O}, \psi, \eta, S_i),$$
$$\tag{A9}$$

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, other summary statistics $\mathbf{D}$ 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:

$$P(\psi, \eta, \{\varphi, \zeta, z\}_{1:N}, \{\mathbf{d}, \mathbf{D}, \sigma, \Sigma, S\}_{1:N}, N|\mathcal{O}) = N^{-1} P(\psi)P(\eta) \prod_{i=1}^{N} P(\varphi_i, \zeta_i, z_i|\psi, \sigma_i, \Sigma_i, S_i, \eta)P(\mathbf{d}_i|\varphi_i, z_i, \sigma_i, S_{d,i}, \eta)$$

$$\times P(\mathbf{D}_i|\zeta_i, \Sigma_i, S_{D,i}, \eta)P(\sigma_i|\mathbf{f}_i, \mathcal{O}, \psi, \eta, S_i)P(\Sigma_i|\mathcal{O}, \psi, \eta, S_i). \tag{A10}$$

Using the chain rule, the population model and likelihood terms can be re-written as:

$$P(\boldsymbol{\varphi}, \boldsymbol{\zeta}, z | \boldsymbol{\psi}, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, S, \boldsymbol{\eta}) = \frac{P(\boldsymbol{\varphi}, \boldsymbol{\zeta}, z | \boldsymbol{\psi}) P(S | \boldsymbol{\varphi}, \boldsymbol{\zeta}, z, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, \boldsymbol{\eta})}{P(S | \boldsymbol{\psi}, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, \boldsymbol{\eta})},$$

$$P(\mathbf{d} | \boldsymbol{\varphi}, z, \boldsymbol{\sigma}, S_d, \boldsymbol{\eta}) = \frac{P(\mathbf{d} | \boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta}) P(S_d | \mathbf{d}, \boldsymbol{\sigma})}{P(S_d | \boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta})}$$

$$P(\mathbf{D} | \boldsymbol{\zeta}, \boldsymbol{\Sigma}, S_D, \boldsymbol{\eta}) = \frac{P(\mathbf{D} | \boldsymbol{\zeta}, \boldsymbol{\Sigma}, \boldsymbol{\eta}) P(S_D | \mathbf{D}, \boldsymbol{\Sigma})}{P(S_D | \boldsymbol{\zeta}, \boldsymbol{\Sigma}, \boldsymbol{\eta})}. \tag{A11}$$

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

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

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

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

where the selection term in the denominator is given by:

$$P(S | \boldsymbol{\psi}, \boldsymbol{\sigma}, \boldsymbol{\Sigma}, \boldsymbol{\eta}) = \int P(S_D | \mathbf{D}, \boldsymbol{\Sigma}) P(S_d | \mathbf{d}, \boldsymbol{\sigma}) P(\mathbf{d} | \boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta}) P(\mathbf{D} | \boldsymbol{\zeta}, \boldsymbol{\Sigma}, \boldsymbol{\eta}) P(\boldsymbol{\varphi}, \boldsymbol{\zeta}, z | \boldsymbol{\psi}) \, d\mathbf{d} \, d\mathbf{D} \, d\boldsymbol{\varphi} \, d\boldsymbol{\zeta} \, dz, \tag{A14}$$

and as before, we have dropped the $P(\boldsymbol{\sigma}_i | \mathcal{O}, \boldsymbol{\psi}, \boldsymbol{\eta}, S_i)$ and $P(\boldsymbol{\Sigma}_i | \mathcal{O}, \boldsymbol{\psi}, \boldsymbol{\eta}, S_i)$ 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 $\mathbf{D}_{1:N}$, one obtains:

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

We can then re-write the selection term in the denominator as,

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

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

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

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

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

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

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

where the selection term in the denominator is given by:

$$P(S_d|S_D, \boldsymbol{\psi}, \boldsymbol{\sigma}, \boldsymbol{\eta}) = \int P(S_d|\mathbf{d}, \boldsymbol{\sigma})P(\mathbf{d}|\boldsymbol{\varphi}, z, \boldsymbol{\sigma}, \boldsymbol{\eta})P(\boldsymbol{\varphi}, \boldsymbol{\zeta}, z|\boldsymbol{\psi}, S_D, \boldsymbol{\eta}, \mathcal{O})\, d\mathbf{d}\, d\boldsymbol{\varphi}\, dz. \qquad (A20)$$

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 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.

B. STAR FORMING SEQUENCE MODEL

Leja et al. (2021) use a normalizing flow to model the star-forming sequence (SFS), i.e., the distribution of star-formation rate 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 $P(\mathrm{SFR}|M, z)$, and has no explicit constraints to ensure that it extrapolates sensibly below the mass-complete limit for the data that it was 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:

$$P(\log_{10}\mathrm{SFS}|\mathcal{M}, z) = r_{\mathrm{q}}(\mathcal{M}, z)\mathcal{S}(\log_{10}\mathrm{SFS}|\mu_{\mathrm{q}}(\mathcal{M}, z), \sigma_{\mathrm{q}}(\mathcal{M}, z), k_{\mathrm{q}}) + (1 - r_{\mathrm{q}}(\mathcal{M}, z))\mathcal{N}(\log_{10}\mathrm{SFS}|\mu_{\mathrm{sf}}(\mathcal{M}, z), \sigma_{\mathrm{sf}}),$$
$$(B21)$$

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: $x \to \mu + \sigma \sinh^{-1}(\sinh(x) + k)$.

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

$$\mu_{\mathrm{sf}}(\mathcal{M}, z) = [a(z)\Theta(\mathcal{M} - \mathcal{M}_{\mathrm{sf}}^*(z)) + b(z)\Theta(\mathcal{M}_{\mathrm{sf}}^*(z) - \mathcal{M})](\mathcal{M} - \mathcal{M}_{\mathrm{sf}}^*(z)) + c(z)$$
$$\mu_{\mathrm{q}}(\mathcal{M}, z) = [d(z)\Theta(\mathcal{M} - \mathcal{M}_{\mathrm{sf}}^*(z)) + e(z)\Theta(\mathcal{M}_{\mathrm{sf}}^*(z) - \mathcal{M})](\mathcal{M} - \mathcal{M}_{\mathrm{sf}}^*(z)) + f(z) - 1$$
$$\sigma_{\mathrm{q}}(\mathcal{M}, z) = \sigma_{q0} + (\sigma_{q0} - \sigma_{q1})\mathrm{sigmoid}((\mathcal{M} - \sigma_{q2})/\sigma_{q3})$$
$$r_{\mathrm{q}}(\mathcal{M}, z) = \mathrm{sigmoid}(j) * \mathrm{sigmoid}((\mathcal{M} - (g + h * z))/i) \qquad (B22)$$

where the functions $a(z)$ through $f(z)$, as well as $\mathcal{M}_{\mathrm{sf}}^*(z)$ and $\mathcal{M}_{\mathrm{q}}^*(z)$, 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. (2021) 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 mini-batching and 10^4 epochs. Fitted parameters are given in Table 3.

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

Table 3. Fitted values of the surrogate SFS model.

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 star-forming sequence (SFS). To re-cap from the main text, we construct a prior over SFH parameters as:

$$P(\alpha, \beta, \tau | M, z) = \frac{\pi_0(\alpha, \beta, \tau) \, P(\mathrm{SFR}|M, z)}{P_0(\mathrm{SFR}|z)} \tag{C23}$$

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

$$P_0(\mathrm{SFR}|z) = \int_{\mathrm{SFR=const.}} \pi_0(\alpha, \beta, \tau) dS, \tag{C24}$$

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 we need to evaluate the prior density, we train a normalizing flow to learn $P_0(\mathrm{SFR}|z)$ 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 100Myr. This provides a training set of $\{\mathrm{sSFR}, \alpha, \beta, \tau, z\}$, on which we can train a conditional density estimator to learn $P_0(\mathrm{sSFR}|z)$.

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 with 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 use of the trained normalizing flow to divide out the implicit SFR prior is shown in Figure 3.

REFERENCES

Abell, P. A., Allison, J., Anderson, S. F., et al. 2009, arXiv preprint arXiv:0912.0201

Abramson, L. E., Gladders, M. D., Dressler, A., et al. 2015, The Astrophysical Journal Letters, 801, L12

Ade, P. A., Aghanim, N., Arnaud, M., et al. 2016, Astronomy & Astrophysics, 594, A13

Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, Publications of the Astronomical Society of Japan, 70, S8
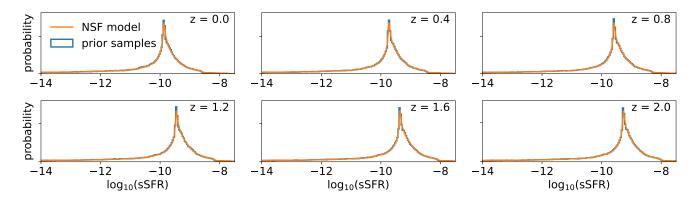
Figure 12. Samples from the prior on the star formation rate implied by taking uniform priors over the double power-law SFH parameters $(\log_{10}\alpha, \log_{10}\beta, \tau)$ over the ranges specified in Table 2 is shown as blue histograms, while the learned normalizing flow model for the implied distribution (as a function of redshift) is shown in orange.

Alarcon, A., Hearin, A. P., Becker, M. R., & Chaves-Montero, J. 2022, arXiv preprint arXiv:2205.04273

Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. 2019, Monthly Notices of the Royal Astronomical Society, 488, 4440

Alsing, J., Wandelt, B., & Feeney, S. 2018, Monthly Notices of the Royal Astronomical Society, 477, 2874

Alsing, J., Peiris, H., Leja, J., et al. 2020, The Astrophysical Journal Supplement Series, 249, 5

Andrews, B. H., & Martini, P. 2013, The Astrophysical Journal, 765, 140

Arnouts, S., & Ilbert, O. 2011, Astrophysics Source Code Library, ascl

Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, Astronomy & Astrophysics, 645, A104

Benitez, N. 2000, The Astrophysical Journal, 536, 571

Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, The Astrophysical Journal, 686, 1503

Buchs, R., Davis, C., Gruen, D., et al. 2019, Monthly Notices of the Royal Astronomical Society, 489, 820

Bustamante, S., Ellison, S. L., Patton, D. R., & Sparre, M. 2020, Monthly Notices of the Royal Astronomical Society, 494, 3469

Byler, N., Dalcanton, J. J., Conroy, C., & Johnson, B. D. 2017, The Astrophysical Journal, 840, 44

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, The Astrophysical Journal, 533, 682

Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, The astrophysical journal. Chicago. Vol. 429, no. 2, pt. 1 (July 1994), p. 582-601

Carnall, A. C., Leja, J., Johnson, B. D., et al. 2019, The Astrophysical Journal, 873, 44

Charlot, S., & Fall, S. M. 2000, The Astrophysical Journal, 539, 718

Choi, J., Dotter, A., Conroy, C., et al. 2016, The Astrophysical Journal, 823, 102

Conroy, C. 2013, Annual Review of Astronomy and Astrophysics, 51, 393

Conroy, C., & Gunn, J. E. 2010, Astrophysics Source Code Library, ascl

Cresci, G., Mannucci, F., & Curti, M. 2019, Astronomy & Astrophysics, 627, A42

Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020, Monthly Notices of the Royal Astronomical Society, 491, 944

Daddi, E., Dickinson, M., Morrison, G., et al. 2007, The Astrophysical Journal, 670, 156

Dark Energy Survey Collaboration: Fermilab, University of Illinois at Urbana-Champaign, U. o. C. L. B. N. L. C.-T. I.-A. O., & Flaugher, B. 2005, International Journal of Modern Physics A, 20, 3121

Davis, C., Rozo, E., Roodman, A., et al. 2018, Monthly Notices of the Royal Astronomical Society, 477, 2196

De Jong, J. T., Kleijn, G. A. V., Boxhoorn, D. R., et al. 2015, Astronomy & Astrophysics, 582, A62

Dotter, A. 2016, The Astrophysical Journal Supplement Series, 222, 8

Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, Monthly Notices of the Royal Astronomical Society, 413, 971

Driver, S. P., Bellstedt, S., Robotham, A. S., et al. 2022, Monthly Notices of the Royal Astronomical Society, 513, 439

Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. 2019, Advances in neural information processing systems, 32

Ferland, G., Porter, R., Van Hoof, P., et al. 2013, Revista mexicana de astronomía y astrofísica, 49, 137

Gatti, M., Vielzeuf, P., Davis, C., et al. 2018, Monthly Notices of the Royal Astronomical Society, 477, 1664

Gerardi, F., Feeney, S. M., & Alsing, J. 2021, Physical Review D, 104, 083531

Hartley, W. G., Chang, C., Samani, S., et al. 2020, Monthly Notices of the Royal Astronomical Society, 496, 4769

Hildebrandt, H., Erben, T., Kuijken, K., et al. 2012, Monthly Notices of the Royal Astronomical Society, 421, 2355

Hildebrandt, H., Choi, A., Heymans, C., et al. 2016, Monthly Notices of the Royal Astronomical Society, 463, 635

Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, Monthly Notices of the Royal Astronomical Society, 465, 1454

Hildebrandt, H., Köhlinger, F., Van den Busch, J., et al. 2020, Astronomy & Astrophysics, 633, A69

Hoyle, B., Gruen, D., Bernstein, G., et al. 2018, Monthly Notices of the Royal Astronomical Society, 478, 592

Ilbert, O., Arnouts, S., McCracken, H., et al. 2006, Astronomy & Astrophysics, 457, 841

Jeffrey, N., & Wandelt, B. D. 2020, arXiv preprint 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, The Astrophysical Journal, 730, 61

Kashino, D., Renzini, A., Silverman, J., & Daddi, E. 2016, The Astrophysical Journal Letters, 823, L24

Kingma, D. P., & Ba, J. 2014, arXiv preprint arXiv:1412.6980

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv preprint arXiv:1110.3193

Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, Astronomy & Astrophysics, 559, A14

Leistedt, B., Alsing, J., Peiris, H., Mortlock, D., & Leja, J. 2022, In preparation

Leistedt, B., Hogg, D. W., Wechsler, R. H., & DeRose, J. 2019, The Astrophysical Journal, 881, 80

Leistedt, B., Mortlock, D. J., & Peiris, H. V. 2016, Monthly Notices of the Royal Astronomical Society, 460, 4258

Leitner, S. N. 2012, The Astrophysical Journal, 745, 149

Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2019, The Astrophysical Journal, 876, 3

Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., & Byler, N. 2017, The Astrophysical Journal, 837, 170

Leja, J., Speagle, J. S., Johnson, B. D., et al. 2020, The Astrophysical Journal, 893, 111

Leja, J., Speagle, J. S., Ting, Y.-S., et al. 2021, arXiv preprint arXiv:2110.04314

Leslie, S. K., Schinnerer, E., Liu, D., et al. 2020, The Astrophysical Journal, 899, 58

Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, Monthly Notices of the Royal Astronomical Society, 390, 118

Mandelbaum, R., Eifler, T., Hložek, R., et al. 2018, arXiv preprint arXiv:1809.01669

McQuinn, M., & White, M. 2013, Monthly Notices of the Royal Astronomical Society, 433, 2857

Ménard, B., Scranton, R., Schmidt, S., et al. 2013, arXiv preprint arXiv:1303.4722

Morrison, C. B., Hildebrandt, H., Schmidt, S. J., et al. 2017, Monthly Notices of the Royal Astronomical Society, 467, 3576

Nagaraj, G., Forbes, J. C., Leja, J., Foreman-Mackey, D., & Hayward, C. C. 2022, arXiv preprint arXiv:2202.05102

Nakajima, K., & Ouchi, M. 2014, Monthly Notices of the Royal Astronomical Society, 442, 900

Newman, J. A. 2008, The Astrophysical Journal, 684, 88

Newman, J. A., & Gruen, D. 2022, arXiv preprint arXiv:2206.13633

Noeske, K., Weiner, B., Faber, S., et al. 2007, The Astrophysical Journal, 660, L43

Paxton, B., Bildsten, L., Dotter, A., et al. 2010, The Astrophysical Journal Supplement Series, 192, 3

Paxton, B., Cantiello, M., Arras, P., et al. 2013, The Astrophysical Journal Supplement Series, 208, 4

Paxton, B., Marchant, P., Schwab, J., et al. 2015, The Astrophysical Journal Supplement Series, 220, 15

Price, S. H., Kriek, M., Brammer, G. B., et al. 2014, The Astrophysical Journal, 788, 86

Ramachandra, N., Chaves-Montero, J., Alarcon, A., et al. 2021, arXiv preprint arXiv:2111.12118

Reddy, N. A., Kriek, M., Shapley, A. E., et al. 2015, The Astrophysical Journal, 806, 259

Renzini, A., & Peng, Y.-j. 2015, The Astrophysical Journal Letters, 801, L29

Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, The Astrophysical journal letters, 739, L40

Salim, S., Lee, J. C., Davé, R., & Dickinson, M. 2015, The Astrophysical Journal, 808, 25

Salim, S., Lee, J. C., Ly, C., et al. 2014, The Astrophysical Journal, 797, 126

Salim, S., & Narayanan, D. 2020, arXiv preprint arXiv:2001.03181

Schmidt, S. J., Ménard, B., Scranton, R., Morrison, C., & McBride, C. K. 2013, Monthly Notices of the Royal Astronomical Society, 431, 3307

Schneider, M., Knox, L., Zhan, H., & Connolly, A. 2006, The Astrophysical Journal, 651, 14

Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, Astronomy & Astrophysics, 575, A74

Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, The Astrophysical Journal Supplement Series, 214, 15

Talbot, C., & Thrane, E. 2022, The Astrophysical Journal, 927, 76

Tanaka, M. 2015, The Astrophysical Journal, 801, 20

Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2018, Publications of the Astronomical Society of Japan, 70, S9

Telford, O. G., Dalcanton, J. J., Skillman, E. D., & Conroy, C. 2016, The Astrophysical Journal, 827, 35

Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2016, The Astrophysical Journal, 817, 118

Whitaker, K. E., Van Dokkum, P. G., Brammer, G., & Franx, M. 2012, The Astrophysical Journal Letters, 754, L29

Whitaker, K. E., Franx, M., Leja, J., et al. 2014, The Astrophysical Journal, 795, 104

Wright, A. H., Hildebrandt, H., Van den Busch, J. L., & Heymans, C. 2020, Astronomy & Astrophysics, 637, A100

Yabe, K., Ohta, K., Akiyama, M., et al. 2015, Publications of the Astronomical Society of Japan, 67

Yates, R. M., Kauffmann, G., & Guo, Q. 2012, Monthly Notices of the Royal Astronomical Society, 422, 215