

RECEIVED: January 17, 2024

REVISED: March 21, 2024

ACCEPTED: April 2, 2024

PUBLISHED: May 9, 2024

Simulation-based inference of deep fields: galaxy population model and redshift distributions

Beatrice Moser^{id},^{a,*} Tomasz Kacprzak^{id},^{a,b} Silvan Fischbacher^{id},^a
Alexandre Refregier^{id},^a Dominic Grimm^a and Luca Tortorelli^{id}^c

^a*Institute for Particle Physics and Astrophysics, ETH Zurich,
Wolfgang-Pauli-Strasse 27, CH-8093 Zurich, Switzerland*

^b*Swiss Data Science Center, Paul Scherrer Institute,
Forschungsstrasse 111, 5232 Villigen, Switzerland*

^c*University Observatory, Faculty of Physics, Ludwig-Maximilian-Universität München,
Scheinerstrasse 1, 81679 Munich, Germany*

E-mail: moserb@phys.ethz.ch, tomaszk@phys.ethz.ch, silvanf@phys.ethz.ch,
alexandre.refregier@phys.ethz.ch, dominic.grimm@phys.ethz.ch,
luca.tortorelli@physik.lmu.de

ABSTRACT: Accurate redshift calibration is required to obtain unbiased cosmological information from large-scale galaxy surveys. In a forward modelling approach, the redshift distribution $n(z)$ of a galaxy sample is measured using a parametric galaxy population model constrained by observations. We use a model that captures the redshift evolution of the galaxy luminosity functions, colours, and morphology, for red and blue samples. We constrain this model via simulation-based inference, using factorized Approximate Bayesian Computation (ABC) at the image level. We apply this framework to HSC deep field images, complemented with photometric redshifts from COSMOS2020. The simulated telescope images include realistic observational and instrumental effects. By applying the same processing and selection to real data and simulations, we obtain a sample of $n(z)$ distributions from the ABC posterior. The photometric properties of the simulated galaxies are in good agreement with those from the real data, including magnitude, colour and redshift joint distributions. We compare the posterior $n(z)$ from our simulations to the COSMOS2020 redshift distributions obtained via template fitting photometric data spanning the wavelength range from UV to IR. We mitigate sample variance in COSMOS by applying a reweighting technique. We thus obtain a good agreement between the simulated and observed redshift distributions, with a difference in the mean at the 1σ level up to a magnitude of 24 in the

*Corresponding author.



i band. We discuss how our forward model can be applied to current and future surveys and be further extended. The ABC posterior and further material will be made publicly available at <https://cosmology.ethz.ch/research/software-lab/ufig.html>.

KEYWORDS: galaxy surveys, galaxy evolution, Bayesian reasoning, cosmological simulations

ARXIV EPRINT: [2401.06846](https://arxiv.org/abs/2401.06846)

Contents

1	Introduction	1
2	Data	3
2.1	Deep/UltraDeep data from HSC PDR3	3
2.2	COSMOS2020 catalog	4
2.3	Validation sample	4
3	Method	5
3.1	Galaxy population model	6
3.2	Image simulations of HSC DUD fields	9
3.3	Source extraction and matching	11
3.4	COSMOS2020 redshift assignment to HSC deep fields	12
3.5	Factorised ABC inference	12
4	Results	15
4.1	ABC posterior	15
4.2	Comparison of simulations and real data	19
4.3	Redshift distributions	22
5	Conclusions	23
A	Galaxy population model priors	26
A.1	Template coefficient priors from COSMOS2015 catalog	26
B	Details of the ABC runs	28
C	SExtractor settings	31

1 Introduction

Cosmological probes allow us to investigate the structure and components of our Universe, by posing constraints on a cosmological model. The standard model of cosmology, known as Λ CDM, comprises of three main components: dark energy, dark matter and baryons. These components can be traced in a large-scale galaxy survey, by measuring the positions and shapes of galaxies and their correlations. In recent years, state-of-the-art experiments such as the Dark Energy Survey¹ (DES; [1]), the Kilo-Degree Survey² (KiDS; [2]) and the Hyper Suprime-Cam Subaru Strategic Program³ (HSC; [3]) have reported their constraints resulting from galaxy clustering, cosmic shear and galaxy-galaxy lensing and their combination, known as 3×2 point analysis [4–6]. Precise determination of the redshift distribution $n(z)$ of samples of galaxies is critical for obtaining cosmological constraints from galaxy surveys.

¹<http://www.darkenergysurvey.org/>.

²<http://kids.strw.leidenuniv.nl/>.

³<https://hsc.mtk.nao.ac.jp/ssp/survey/>.

Redshift information allows the separation of source and lens sample, and the computation of cosmological observables. Spectroscopy is prohibitively time-consuming as a technique to provide accurate redshifts of all galaxies in a wide survey, and is furthermore subject to selection biases, especially for faint samples. Surveys thus rely on integrated measurements in a limited number of broad-bands in order to determine the redshift distribution of the sample of interest (for review, see [7, 8]). This has proven to be a challenging task, especially since the relationship between redshift and colour in a limited wavelength range is subject to degeneracies [8]. The characterization of photometric redshift (photo- z) distributions is one of the key systematics affecting cosmic shear measurements since errors in the calibration of redshift distributions and their uncertainties can lead to biases in the retrieved cosmological parameters [8–15]. Traditional photo- z approaches include template fitting (for example LePhare [16, 17], BPZ [18], ZEBRA [19] and EAZY [20]) and machine learning methods (for example ANNz [21], ANNz2 [22, 23] and DNF [24]).

In a cosmological survey it is common to employ methods to constrain the overall redshift distribution of the sample of interest rather than the redshifts of single objects, either by an empirical reweighting of a well measured redshift sample [25–27], by using spatial cross-correlations [28], or by a combination of the two [29, 30]. Another approach, which has been introduced in recent years, is simulation-based inference (SBI) [31–33]. SBI relies on forward modelling the survey of interest: the redshift distribution of a sample of galaxies is the result of the statistical properties of the observed galaxy population, the observing conditions and limitations of the detector, and the selection applied to define the target sample. Accurate modelling of magnitude, colour and (at second-order) size distributions of galaxies as a function of redshift, taking into account the observational and instrumental effects, enables robust determination of the redshift distribution of the sample, as well as a straightforward estimation of its uncertainty. Furthermore, the constraints on the galaxy population model provide some insights on the statistical nature of galaxies and allow a robust treatment of complicated selection functions. The methodology used in this paper has been developed in [31] and further extended and applied in [32] and [34, 35]. We refer to this forward modelling framework as Monte Carlo Control Loops (MCCL) [36]. The method relies on an empirical galaxy population model to describe the intrinsic properties of galaxies and stars. We render the objects with photon shooting methods using an image simulator called Ultra Fast Image Generator (UFIG [37]) in a set of broad-bands described by the filter throughputs of the telescope used. Moreover, we simulate observational and instrumental effects such as sky and detector noise, point spread functions (PSF), reddening and saturation. In this way, we can post-process the simulations and the real images in the same way. We run SExtractor [38] on both to obtain catalogs of objects and apply the same selection functions to both the simulated and the real catalog, which simplifies the treatment of selection biases. The model parameters are constrained using the observed data via Approximate Bayesian Computation (ABC), using distance measures that ensure that the photometric properties of the objects in simulations statistically agree with real data. The method has been used to simulate and perform a cosmic shear measurement of the Dark Energy Survey Year 1 [32, 39, 40], for redshift calibration on Subaru data [31], and to obtain the luminosity functions of blue and red galaxies at different redshifts with Canada-France-Hawaii Telescope Legacy Survey (CFHTLS [41]) data [34]. Furthermore, it has been applied to simulate the

narrow band imaging of Physics of the Accelerating Universe survey (PAUS [42]) [35, 43] and galaxy spectra from the Sloan Digital Sky Survey (SDSS [44]) CMASS sample [45, 46].

In this work, we use HSC Deep/UltraDeep (DUD) data [47] and accurate many-band photometric redshifts from COSMOS2020 [48] in order to obtain tight constraints on the model parameters at high redshift. Previous constraints to the model parameters extended to a redshift of $z \sim 1$, whereas in this work we explore the regime of Stage IV surveys. [49] suggests that HSC data is the most powerful for constraining the Schechter parameters of the luminosity function and thus the redshift distribution of galaxies, because of its exquisite depth. As done in previous work, we use an ABC framework, with several important practical improvements. After tuning the model, we use the obtained posterior parameters to simulate HSC DUD data in the COSMOS field and validate the $n(z)$ for different magnitude cuts. While the HSC DUD fields are deeper than current and upcoming wide-field surveys, the magnitude cuts that we consider in this work are consistent with the weak lensing source samples of current (i band magnitude cut between 23 and 24) and upcoming surveys (expected i band magnitude cut between 24.5 and 25.5). Using our ABC posterior directly to calibrate the redshift distributions of these survey would entail running image simulations of the survey of interest using the parameters from the ABC posterior obtained in this work. The resulting $n(z)$ distribution would include errors propagated from the ABC posterior constraints and due to selection effects. Calibrating the model with deeper data has the advantage that objects above the magnitude cut are constrained, which is necessary to include noise and source confusion, thus making the sample selection realistic.

The paper is structured as follows. Section 2 describes the HSC DUD data and the COSMOS2020 catalog used both to tune the model and validate our results. In section 3, we describe the methodology and introduce the changes compared to previous work. Section 4 reports the results of our analysis. We conclude the paper in section 5. We assume a standard Λ CDM cosmology with $h=0.7$, $\Omega_m=0.3$, and $\Omega_\Lambda=0.7$ throughout the paper.

2 Data

In this section, we present the data used to constrain our model of the galaxy population and to validate the obtained redshift distributions. We rely on data from the Deep and UltraDeep layers of the third data release (PDR3) of the Hyper Suprime-Cam Subaru Strategic Program (HSC) [47]. In order to provide our model with additional redshift information and validate the $n(z)$, we complement the HSC data with accurate photo- z estimates from the COSMOS2020 panchromatic photometric catalog [48].

2.1 Deep/UltraDeep data from HSC PDR3

HSC is a large multi-band imaging survey conducted with the 8.2-metre Subaru telescope. It comprises of three layers: Wide, Deep and UltraDeep. The Wide layer covers 1470 deg^2 , considering partially observed areas in five broad-band filters (g, r, i, z, y). The Deep/UltraDeep (DUD) layers cover $\sim 36 \text{ deg}^2$ in the five broad-band filters and four additional narrow-band filters. In this work, we use the publicly available coadded broad-band DUD images with

local sky subtraction from PDR3.⁴ There are four different fields: COSMOS, DEEP2-3, SXDS+XMM-LSS and ELAIS-N1. Each field is separated in tracts which are equi-area rectangular regions on the sky, divided in 9×9 patches comprising of 12 arcmin per side corresponding to 4200 pixels and overlapping by 100 pixels. We use all the available DUD patches for tuning the model. The dataset consists initially of roughly 1500 patches. We blacklist patches where:

- more than 30% of the image area is flagged as `NO_DATA`,
- more than 50% of the image area is covered by the `BRIGHT_OBJECT` mask,
- the image overlaps for more than 30% of the area with another patch.

`NO_DATA` and `BRIGHT_OBJECT` masks correspond to flags 8 and 9 in the mask layer of the data. The overlap, on the other hand, is computed using the footprint of the images and by masking the pixels on the top or upper edge of each coadd that are also covered by another patch. After blacklisting, we retain a total of 746 patches.

2.2 COSMOS2020 catalog

The COSMOS2020 catalog [48] consists of nearly 1 million high quality photometric redshifts derived via template fitting of many broad and narrow band observations ranging from UV to IR wavelengths. There are four different publicly available catalogs, which differ in the method used for extracting photometry (`SEXTRACTOR`, used in `CLASSIC`, and `THE FARMER`) and for the photometric redshift template fitting code (`LePhare` [16, 17] and `EAZY` [20]). Since we use `SEXTRACTOR` for the photometric measurement on the HSC data in our pipeline, we also work with the COSMOS2020 `CLASSIC` catalog. We use both `LePhare` and `EAZY` photo-*zs*. We remove areas where the photometry is unreliable or with partial coverage by means of the `FLAG_COMBINED` parameter thus reducing the area to 1.27 deg^2 . We select objects that have `MAG_AUTO` < 99, `LePhare` (`lp_zBEST`) or `EAZY` (`ez_z_phot`) photo-*z* between 0 and 8 (removing `Nan` values), `LePhare` object type galaxy (`lp_type=0`) and `SEXTRACTOR` `FLAGS` < 4. The COSMOS2020 catalog is used both for providing redshift information while constraining the model and for validation. The validation sample is explained in the following section, while the reweighting procedure used during the ABC analysis is detailed in section 3.4.

2.3 Validation sample

In order to build our validation sample, we compare the COSMOS2020 `COMBINED` footprint with the HSC DUD data in the COSMOS field and find 63 overlapping patches, out of which 56 are almost fully covered. We perform `SEXTRACTOR` forced photometry on these coadds using the *i* band for detection and match the obtained catalog with objects in the COSMOS2020 catalog by position and magnitude (using the magnitude `MAG_APER` measured in a circular 3'' diameter aperture). The `BRIGHT_OBJECT` masks from HSC PDR3 are very conservative and cause a loss of roughly one third of the COSMOS2020 objects, since COSMOS2020 uses the less conservative HSC PDR2 masks. We compare the simulated final redshift distributions

⁴https://hsc-release.mtk.nao.ac.jp/doc/index.php/available-data___pdr3/.

Value	<i>i</i> band mag cut 23	<i>i</i> band mag cut 24	<i>i</i> band mag cut 25
Δz sample variance LePhare	0.014 ± 0.012	0.001 ± 0.019	0.002 ± 0.019
Δz sample variance EAZY	0.015 ± 0.012	0.001 ± 0.019	0.003 ± 0.018
$\Delta z = \langle z_{\text{EAZY}} \rangle - \langle z_{\text{LePhare}} \rangle$	0.014	0.018	0.0006

Table 1. Shifts in photometric redshift due to sample variance and difference between the two photo-*z* codes used in COSMOS2020.

in the COSMOS field to both **LePhare** and **EAZY** photo-*z*s at three different magnitude cuts and with the selection defined in appendix B. We describe our procedure to account for sample variance in COSMOS in the next subsection.

2.3.1 Sample variance in COSMOS

The COSMOS field only spans 2 deg^2 of the sky. This means that, while the volume spanned by COSMOS observations is large due to the considerable depth [50], there are notable sample variance effects at low redshifts. In order to estimate the impact of sample variance, we assign COSMOS2020 photo-*z*s to all galaxies in the other deep fields as described in section 3.4 and look at the offset in mean redshift. This ensures that we span a larger area and the difference in depth is negligible when only considering galaxies with *i* band magnitude below 25.⁵ In order to also measure the scatter due to sample variance, we produce 10 sub-fields the size of COSMOS (56 images) and measure the standard deviation of the 10 mean redshifts. We prefer this approach to a standard Bootstrap in order to preserve the locality of the effect, which is due to the inhomogeneity of large-scale structure on small scales. The redshift offsets are reported in table 1. We observe a mean redshift offset between the COSMOS field and the other deep fields of $\Delta z = \langle z_{\text{COSMOS}} \rangle - \langle z_{\text{deep}} \rangle \approx 0.015$ when we cut at *i* band magnitude of 23, meaning that sample variance causes a bias for the brightest sample.

2.3.2 Offset between EAZY and LePhare

We notice that, when applying a simple magnitude cut in the *i* band and the selection described in appendix B, there is a systematic offset between **EAZY** and **LePhare** photometric redshifts from COSMOS2020. Similarly to sample variance, this effect has a stronger impact on the brightest sample where **EAZY** predicts systematically higher redshifts than **LePhare**. The systematic offset is $\Delta z = \langle z_{\text{EAZY}} \rangle - \langle z_{\text{LePhare}} \rangle = 0.014$ for a magnitude cut at **MAG_APER3_i=23** and 0.018 for a magnitude cut at **MAG_APER3_i=24**. The offset is negligible when cutting at **MAG_APER3_i=25**. We report these offsets in table 1.

3 Method

The backbone of our forward modelling framework is an empirical parametric model of the galaxy population, used to generate distributions of intrinsic properties of galaxies. Once a galaxy catalog is generated given a set of model parameters, we simulate an image of the survey of interest, in our case the HSC DUD fields. In order to obtain a realistic simulation,

⁵We use the *i* band aperture magnitude in a $3''$ circular aperture **MAG_APER3_i**, as described in section 3.3.

we include the effects of the instrument and the known observational systematics that impact the photometric measurement. The end-to-end process from a set of model parameters to a realistic telescope image is implemented in the Ultra Fast Image Generator (UFIG [37]). UFIG has been developed as a simulator for MCCL, with speed as one of the primary features. Computational speed is critical for this task, since a large number of simulations is required to tune the parameters of the model. The inference is performed by running an Approximate Bayesian Computation (ABC), where the realism of our simulated images is increased by minimizing a set of distance measures. Despite the optimized simulator, the computational cost of performing a full ABC inference amounts to roughly a million CPU hours on the Euler HPC cluster of ETH Zurich. In the following, we describe the galaxy population model, how we extend UFIG to reproduce realistic HSC DUD images, how we include redshift information from COSMOS2020 reducing the impact of cosmic variance and the details of our ABC scheme. We focus on the novelties introduced in this work.

3.1 Galaxy population model

We include in our model two different populations of galaxies: red and blue, often referred to as quiescent and star-forming galaxies. The separation between red and blue galaxies is intrinsic: we have separate sets of parameters in our model for the two galaxy populations. We sample absolute magnitudes M and redshifts z from Schechter luminosity functions

$$\phi(z, M) = \frac{2}{5} \ln 10 \phi^*(z) 10^{\frac{2}{5}(M^*(z)-M)(\alpha+1)} \exp\left(-10^{\frac{2}{5}(M^*(z)-M)}\right), \quad (3.1)$$

where $M^*(z)$ and $\phi^*(z)$ are functions of redshift and have separate parameters for blue and red galaxies. We then assign a spectral energy distribution (SED) to each galaxy as a linear combination of 5 spectral templates from KCORRECT [44]

$$\text{SED}(\lambda) = \sum_{i=0}^4 c_i T_i(\lambda).$$

The coefficients of the templates are also different for blue and red galaxies and evolve with redshift. We show the five KCORRECT template spectra for reference in figure 1. We assign sizes to galaxies using a log-normal distribution for the half light radius and a Sersic light profile. The ellipticities are sampled from a Beta distribution. As explained in the following subsections, in this work all galaxy population parameters are constrained separately for red and blue galaxies. Furthermore, we add stars to our simulations using the Besançon model of the Milky Way [51]. The magnitudes from the catalog of pseudo-stars are sampled with replacement, and the positions are assigned randomly within the HEALPIX pixel (`nside=8`, see [31, 32] for more details). The positions of the bright end of the star distribution is taken from the Gaia DR3 catalog [52, 53] and abundance-matched to the Besançon model. For an extensive description of the galaxy population model, see [31, 34].

In the following, we highlight the modifications to the galaxy population model compared to [31, 32, 34]: (i) modification of the luminosity function parametrization, (ii) addition of new parameters in the morphology sector to allow different characteristics for blue and red galaxies, (iii) small changes to the parametrization of ellipticities and Sersic indices, (iv) modification of parametrization for template coefficients of the SED.

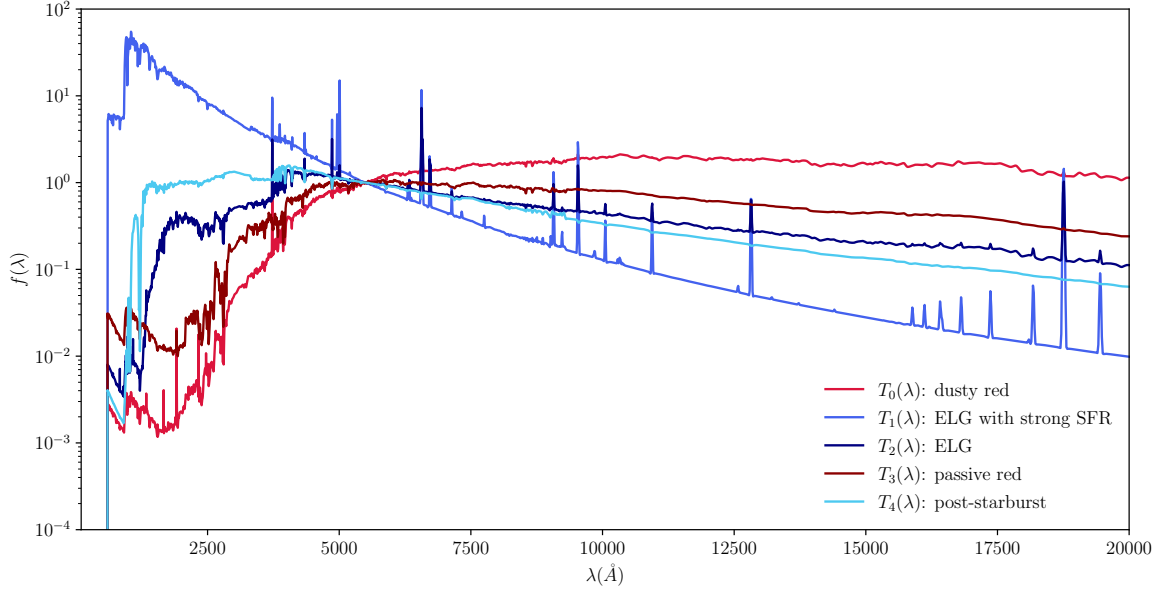


Figure 1. The five KCORRECT templates [54] which are combined linearly to obtain the spectral energy distributions of galaxies. They are renormalized so that $f(\lambda) = 1$ at $\lambda = 5500(\text{\AA})$.

3.1.1 Luminosity function parametrization

We modify the redshift evolution of the luminosity function parameters, in accordance with galaxy evolution models [55]. In the Pure Luminosity Evolution (PLE) scenario, massive galaxies assemble and form most of their stars at high redshifts. They then evolve without merging. This results in the functional form

$$M^*(z) = M_{\text{intcpt}}^* + M_{\text{slope}}^* \log(1 + z) \quad (3.2)$$

for the evolution of the characteristic absolute magnitude with different parameters M_{intcpt}^* and M_{slope}^* for blue and red galaxies. In the Pure Density Evolution (PDE) scenario, galaxies undergo mergers so that they are more massive but less numerous at lower redshifts. This scenario can be modelled through the evolution of the normalization of the luminosity function as

$$\phi^*(z) = \phi_{\text{ampl}}^* (1 + z)^{\phi_{\text{exp}}^*} \quad (3.3)$$

where ϕ_{ampl}^* and ϕ_{exp}^* also depend on the galaxy population. We vary these 8 parameters during the ABC. Furthermore, α_{blue} and α_{red} , describing the steepness of the faint-end slope of the luminosity function, are also varied in this analysis, differently from previous work.

3.1.2 Updated galaxy morphology

We added new parameters in the morphology section. The relation between galaxy half light radius r_{50} and absolute magnitude is described by three parameters: $\log r_{50}^{\text{intcpt}}$, $\log r_{50}^{\text{slope}}$ and $\log r_{50}^{\text{std}}$ [31]. We sample $\log r_{50}$ from a normal distribution with mean

$$\log r_{50}^{\text{mean}} = \log r_{50}^{\text{slope}} M + \log r_{50}^{\text{intcpt}}$$

and standard deviation $\log r_{50}^{\text{std}}$, where M is the absolute magnitude of the galaxy. The distribution of r_{50} is thus log-normal (as found in e.g. [56, 57]). In our updated model we have a separate set of these parameters for red and blue galaxies. We also constrain the $\log r_{50}^{\text{std}}$ parameter during the ABC inference, which was fixed in previous work.

3.1.3 Ellipticity and Sersic indices

The parameterization of ellipticity $p(e)$ using a Beta distribution is slightly modified; parameters e_{mode} , e_{spread} correspond to the mode and concentration of the Beta distribution respectively. They are related to the parameters of the Beta distribution α and β through $e_{\text{spread}} = \alpha + \beta$ and $e_{\text{mode}} = \frac{\alpha-1}{e_{\text{spread}}-2}$. This change makes the parameters easier to interpret and allows for designing simpler priors. We also modified the prescription for modelling the distribution of Sersic indices. We use a Betaprime distribution with free parameter n_s , which is the mode of the distribution. It is related to parameters α and β of the Betaprime distribution as $n_s = \frac{\alpha-1}{\beta+1}$. Parameter β is responsible for the scatter, and fixed throughout the analysis, to the following values: $\beta_{\text{blue}} = 5$ and $\beta_{\text{red}} = 50$. These values were chosen so that the distributions of Sersic indices match that of [58].

3.1.4 Spectral templates parametrization

In our model, the coefficients of the SED templates are drawn from a five-dimensional Dirichlet distribution separately for blue and red galaxies, similarly to [31, 32, 34]. The Dirichlet distribution is used because the samples drawn from it sum to 1 and the spectrum can then be rescaled to match the absolute magnitude of the galaxy. The parameters of the Dirichlet distribution evolve with redshift: we use two separate sets of parameters for $z=0$ and $z=3$, with parameters for other redshifts being an interpolation between them

$$\alpha_i(z) = (\alpha_{i,0})^{1-\frac{z}{3}} \times (\alpha_{i,3})^{\frac{z}{3}}.$$

In previous work the α_i were constrained at redshifts $z = 0$ and $z = 1$; we now use $z = 3$ since the functional form is fixed (so that we do not need to have a large sample of galaxies at $z = 3$ to pose limits on the parameters' values) and this allows us to enforce prior bounds at higher redshifts.

Previously, the prior on this distribution was also a Dirichlet variable with unity weights, multiplied by a uniform number between [5,15], which accounted for the variance. This way, the α_i parameters were affecting both the mean and variance of the Dirichlet variable. We change the model to capture the mode and variance in separate parameters. Furthermore, [31] derived weights to apply to each template using the New York University Value-Added Galaxy Catalog [54] thus effectively using different template spectra for blue and red galaxies. We removed the weights and reparametrized the template spectra to be normalized to 1 at wavelength 5500 Å. We use a redundant parametrization with modes of the Dirichlet distributions $\bar{\alpha}_i$, $i = 0, \dots, 4$ and two new $\alpha_{\text{std},0/3}$ parameters. The parameters $\alpha_{\text{std},0/3}$ correspond to the standard deviation of the 5-dimensional Dirichlet coefficients with equal concentrations at redshifts $z = 0$ and $z = 3$ and evolve in redshift the same way as the template coefficients α_i . We enforce the normalization $\sum_i \bar{\alpha}_i = 1$, and the final Dirichlet

coefficients are calculated as

$$\alpha_i = 1 + \bar{\alpha}_i \cdot \left[\frac{1}{N} \left(1 - \frac{1}{N} \right) \alpha_{\text{std}}^{-2} - N - 1 \right]$$

with $N=5$. This new parametrization reduces the number of local minima in the problem and makes the $\bar{\alpha}_i$ variables more interpretable. Finally, the template spectra are the same for red and blue galaxies and we encapsulate the information about the different galaxy populations in the ABC prior on the template coefficients. The ABC prior on template coefficients was obtained in a preprocessing step where we performed an ABC on catalog level using the COSMOS2015 catalog [59], as described in appendix A. The final model has 46 parameters, out of which 4 are redundant (see table 3 in appendix A).

3.2 Image simulations of HSC DUD fields

The catalogs of intrinsic galaxy properties are used to create simulated HSC DUD images. The image generation procedure, including realistic observational and instrumental effects, is as follows. We input the metadata provided by the HSC database⁶ about size of the image in pixels, pixel scale (0.168"/pixel) and sky coordinates of the images. We perform our simulations using the g , r , i , z and y broad-bands. HSC replaced the r and i filters with more uniform filters $r2$ and $i2$ which have been coadded together with r and i . In our simulations, we use the filter throughputs from $r2$ and $i2$, after checking that the magnitude shifts are small. In order to compute the apparent magnitude of a galaxy in a specific broad-band, we integrate over its SED and the filter throughput taking into account k-corrections and reddening due to galactic extinction. The computation of arbitrary magnitudes in the AB system is described in section 3.2.3 of [31] and the wavelength dependent extinction to account for reddening in appendix D of the same paper. The magnitude zeropoint is set to 27 mag/ADU for the HSC coadds.

We simulate PDR3 coadded images directly, as introduced in section 2.1. In order to simulate the coaddition process, we use systematic maps derived from the metadata. We create a map of the exposure times and number of exposures per pixel for each patch. The CCD gain of a single exposure multiplied by the number of exposures per pixel gives us a rough estimate of the effective gain to convert between ADUs and number of photons. Galaxies are randomly distributed on the image and rendered by sampling individual photons according to the galaxy's Sersic profile. This procedure naturally includes Poisson noise [37, 39]. The Point Spread Function (PSF) is rendered as a distortion to the light profile of the galaxy. In order to estimate the impact of the PSF in the real images, we use a Convolutional Neural Network (CNN) as presented in [60] and updated in [32]. The PSF is estimated at the position of stars matched with Gaia DR3 [52, 53] with magnitudes included between 18 and 22 in the i band, which have SExtractor FLAGS 0 or 16 and are not at the image boundary. We perform this selection because stars with apparent magnitudes lower than 18 in the i band are included in the bright objects masks of HSC. The matching is done with a Balltree with a maximum distance of 1.5 pixels. We reserve 15% of the selected stars for validation. Each PSF parameter is then interpolated across the coadd using a Chebyshev polynomial basis of maximum order 4 (see appendix C of [32] for details).

⁶<https://hsc-release.mtk.nao.ac.jp/datasearch/>.

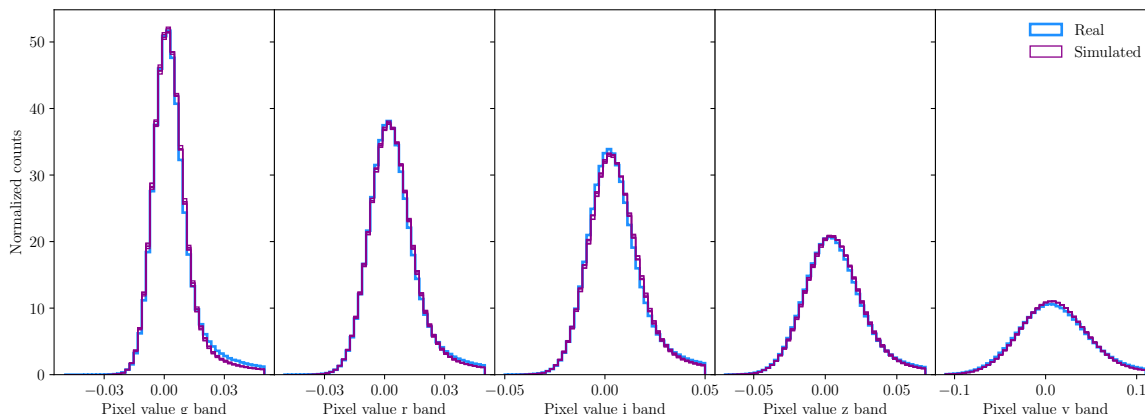


Figure 2. Example of pixel histogram around the background values from `tract` 9813 patch 702. The simulated background is shown from 5 different ABC configurations.

In order to simulate the background noise in an image, we first derive a map of the root-mean-square of the noise from the real data using the `Background2D` function with `SExtractorBackground` estimator and 3σ sigma-clipping from `photutils` [61]. This map is derived from the data individually for each patch. We then add Gaussian noise to the simulation with mean read off from the image header and standard deviation taken from the map (different for each pixel). Since the standard deviation of the noise that we apply is already different in each pixel and is estimated from background subtracted images, we do not need to perform any background subtraction, including local background subtraction which has the most impact in the surroundings of bright objects, which are anyway masked. The resulting simulated background is in good agreement with the background in the real data. We show an example of the pixel histogram of an image for real data and simulations for low pixel values in figure 2. We observe that there is a slight overestimation of the background level due to the lack of background subtraction.

An alternative approach to background estimation would amount to adding a Gaussian background using the parameters in the image headers (both mean and standard deviation) and then applying global and local background subtraction. In our tests this procedure worsened the agreement between data and simulations. We create masks of the areas with no data and surrounding bright stars using the bit flags 8 and 9 from the second layer of the fits files.

The final steps of the simulation process convert photons to ADUs by dividing out the effective gain and saturate pixels that are above the maximum value of the real data. This is a simplistic estimate of the saturation limit, which is good enough in practice since the saturated areas are always masked. We show an example of a simulated image compared to real data in figure 3. The most noticeable difference between the real image and the simulation is the presence of some large galaxies in the simulation. These are not ruled out by our distance measures and need further investigation. The lack of local background subtraction in the simulated image is noticeable around bright objects.

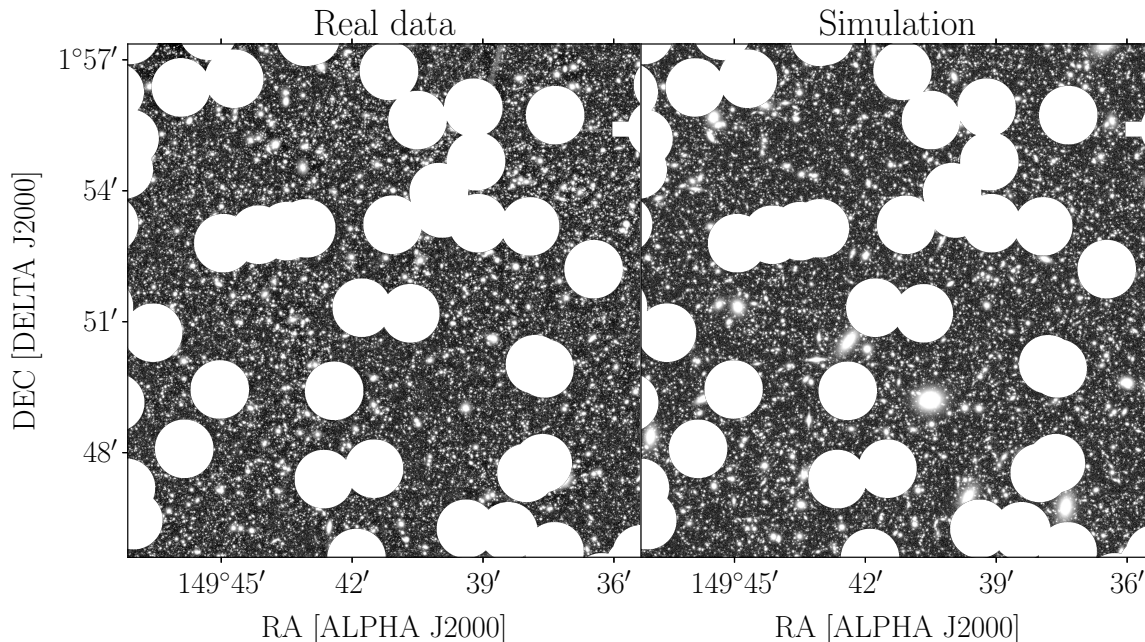


Figure 3. A comparison between a real and a simulated coadd for tract 9813, patch 702 in the HSC COSMOS field is shown. The bright star mask, as derived from the data, is applied to both the real data and simulation.

3.3 Source extraction and matching

In order to proceed in our analysis, we run SExtractor in dual-image mode with the same settings (reported in appendix C) on real images and simulations. We use the i band image as detection image. In the simulations, the SExtractor detections are matched by position and magnitude to the true properties of the injected galaxies. This procedure can have an impact on the resulting photometric properties of galaxies, on the ABC posterior and on the redshift distribution since not all SExtractor detections are matched to an injected galaxy or matched correctly. We use the segmentation map produced by SExtractor and find, for each detection, the overlapping simulated object that minimizes the sum of the differences Δmag between MAG_AUTO and true magnitude in all bands

$$\Delta\text{mag} = \sum_{b \in g, r, i, z, y} |\text{mag}_b - \text{MAG_AUTO}_b|. \quad (3.4)$$

This matching procedure improves on our previous technique of matching each detected object to the closest detected object below a predefined magnitude difference, especially in the case of a crowded field. Nevertheless, the procedure to match true and detected objects and the impact of blending will need further investigation in the future. While in the matching procedure we use MAG_AUTO (which is closest to the UFIG true magnitude), in the following we always use MAG_APER in a $3''$ aperture (referred to as MAG_APER3 from here on, with related quantities MAGERR_APER3, FLUX_APER3 and FLUXERR_APER3), unless otherwise specified. This induces photometric biases for bright large objects (which are larger than the fixed $3''$ aperture) but provides more reliable colours and reduces photometric biases for faint

objects. `FLUX_AUTO` is the sum of the pixel values assigned to the object and thus depends on the adaptive determination of the object’s size. By injecting the same catalog in a deep and ultradeep image, we observed a dependence of `MAG_AUTO` on exposure time, since more pixels of an object rise above the background. A selection in `MAG_AUTO` is undesirable in our case, since we calibrate the model on deep images and then extrapolate it to ultradeep images.

3.4 COSMOS2020 redshift assignment to HSC deep fields

Before describing our ABC scheme, we show how we incorporate redshift information from COSMOS2020 in the SExtractor catalogs obtained from other deep fields. We apply the reweighting technique described in section 4.2 of [13]. We start from the validation sample introduced in section 2.3 where we have galaxy photometry from our own SExtractor run in the HSC COSMOS field overlapping with COSMOS2020 and LePhare and EAZY photo- z s derived from position matching the COSMOS2020 catalog. In order to assign a redshift estimate to a target galaxy in another deep field, we first add Gaussian noise to the COSMOS galaxies’ fluxes until the noise level is equal to that of the target galaxy (the images in the COSMOS field are UltraDeep and thus less noisy than in the other fields). We discard COSMOS galaxies that have larger flux errors than the target galaxy. We then match a COSMOS2020 galaxy to the target galaxy by minimizing the flux χ^2

$$\chi^2 \equiv \sum_b \left(\frac{f_b - f_b^{\text{COSMOS}}}{\sigma_b} \right)^2 \quad (3.5)$$

where $b \in g, r, i, z, y$ and f_b is the `FLUX_APER3` of an object in band b and σ_b its `FLUXERR_APER3`. In this way, we reweight the COSMOS2020 $n(z)$ to match the colour distribution of galaxies in the image we are considering. We verify that a COSMOS2020 galaxy is never matched more than 5 times in the same image (multiple matches only happen for very bright galaxies). In the following subsections, we present two contributions to the uncertainties of the redshift distributions from COSMOS2020, beyond the photo- z errors on individual objects. These are taken into account when validating the $n(z)$ derived from our forward modelling approach against COSMOS2020 photo- z s.

3.5 Factorised ABC inference

We constrain the 46 parameters of our galaxy population model using the HSC deep data and the COSMOS2020 catalog described in section 2. This data combination constitutes a unique sample to precisely constrain our galaxy population model at high redshift, given its completeness up to high magnitudes. We perform simulation-based inference (SBI) to derive a posterior distribution of the parameters of the model, since the likelihood of the observables is unknown, but we have the ability to sample from it through simulations. Our ABC scheme is similar to the one used in [34], and involves prior-to-posterior iterations. The base idea behind ABC is that the model posterior $p(\theta|x)$, where x is the observed data and θ the parameters of the model, can be approximated by $p(\theta|\rho(x, y) < \epsilon)$, where ρ is a distance metric, y is the simulated data and ϵ is a threshold.

The unique property of our problem is that the dataset comprises of a large number of images, which can be considered as semi-independent. We divide our full dataset \mathbf{d} into N

smaller parts $\mathbf{d} = d_1, \dots, d_N$. This way we can factorize the posterior on the full dataset into posterior from its parts

$$p(\theta|\mathbf{d}) \sim \prod_{i=1}^N p(\theta|d_i). \quad (3.6)$$

Then we use the posterior of one part of the data as prior for another part

$$p(\theta|d_1) \sim p(d_1|\theta)p(\theta), \quad (3.7)$$

$$p(\theta|d_{i+1}) \sim p(d_{i+1}|\theta)p(\theta|d_i), \quad (3.8)$$

where $p(\theta)$ is the prior on the model parameters θ . This factorization allows for efficient application of the simple rejection ABC algorithm, which allows for very low complexity of our high performance computing implementation. In practice, we begin by sampling 10000 points from the model's prior $p(\theta)$ and using each model parameter configuration to simulate a part of the data d_1 . We accept the parameters θ where the combined distance metric computed from the simulations falls in the 20th percentile. We then resample the obtained distribution and iterate the procedure. The details of the factorized ABC are discussed at the end of this section and in appendix B.

We modify the ABC inference engine compared to [31, 32, 34] as follows: (i) updated distance metrics, (ii) modifications to the ABC iteration engine and posterior modelling, (iii) modification of model's priors.

3.5.1 Distance metrics

We use the Maximum Mean Discrepancy (MMD), a kernel two-sample test for high dimensional probability distributions [62], as our primary distance measure

$$d_{\text{MMD}} = \frac{1}{N(N-1)} \sum_{i,j} k(x_i, x_j) + k(y_i, y_j) - k(x_i, y_j) - k(y_i, x_j) \quad (3.9)$$

where x_i is a property of the i th object in the real data and y_j is a property of the j th object in the simulated data. The kernel we use is Gaussian

$$k(x_i, y_j) = \exp\left(-\frac{\|x_i - y_j\|^2}{2\sigma}\right) \quad (3.10)$$

with free parameter σ . We describe how we choose the value of σ in appendix B. We extend the input vector of the MMD compared to previous work to include magnitudes `MAG_APER3`, sizes `FLUX_RADIUS`, the two photo- z estimates assigned as described in section 3.4, ellipticities calculated from windowed moments `**_WIN_IMAGE`, and a new variable called the *flux fraction* f_b , calculated as:

$$f_b = \frac{\text{FLUX_APER3_b}}{\sum_j \text{FLUX_APER3_j}} \quad (3.11)$$

where `FLUX_APER_b` is the `SExtractor` flux in band b in a circular $3''$ aperture. Flux fractions capture similar information as colours. This information is technically also present

in the magnitudes, but these mostly impact constraints on the luminosity function parameters. We found that the addition of the flux fractions improved our capacity to constrain the $\bar{\alpha}_i$ parameters of the SED template coefficients distribution. Since our main goal is redshift calibration, we decide to mostly focus on colours, magnitudes and redshifts. For this reason, we include `MAG_APER3` and f_b in all 5 bands, but only include the ellipticity and `FLUX_RADIUS` in the reference i band. Our model does not account for colour gradients in the galaxy size, so it is a reasonable approximation to only constrain the size model in the reference band. The addition of information about the galaxy profile, for instance the concentration index of galaxies (the ratio between the semi-major axes containing 50% and 90% of the elliptical Petrosian flux, see e.g. [63, 64]), in the MMD distance could help to constrain the size model. In our experiments, we found little correlation between the observed concentration indices and morphological parameters from our model, probably due to the faintness of the selected galaxy sample, in which shapes are dominated by the point spread function.

The MMD input vector is 14-dimensional. Before calculating the MMD distance, each column is scaled, so that its mean is close to zero and its standard deviation to 1. The scaling is obtained from the real data and used throughout the analysis for both real and simulated data. Since the fraction of outliers in the COSMOS2020 catalog for galaxies above magnitude 25 is extremely large ($\simeq 25\%$), we only select objects with `MAG_APER3` < 25 in the i band. The galaxy sample is dominated by faint galaxies. In order to increase the weight assigned to galaxies that are the target of current wide-field surveys, we include an MMD distance with magnitude cut `MAG_APER3_i=23`. This choice ensures that the photometric properties of the bright end of the galaxy population are well reproduced by our ABC posterior. Omitting this distance worsens the fit slightly. Other selection cuts to remove stars and objects with bad measurements are described in appendix B.

As the MMD does not capture the differences in number counts, which has an important impact on the normalization of the luminosity function, we combine it with a fractional difference in number of objects

$$d_{ng} = \frac{|N_{\text{SIM}} - N_{\text{HSC}}|}{N_{\text{HSC}}} \quad (3.12)$$

where N_{SIM} is the number of objects in the simulation and N_{HSC} in the real data. This distance is also included for both magnitude cuts at `MAG_APER3_i` < 23 and `MAG_APER3_i` < 25.

The distance metrics in an iteration n are computed for all patches that are included in that iteration ($|d_n|$). We aggregate each distance using the median, which is robust to outliers, so that we have four distances ($d_{ng,25}, d_{ng,23}, d_{\text{MMD},25}$ and $d_{\text{MMD},23}$) for each parameter configuration. We then rescale the distribution of each distance across ABC points so that it has minimum equal to zero and median equal to 1 and finally add the distances with weights:

$$d_{\text{comb}} = 0.1 \cdot d_{ng,25} + 0.1 \cdot d_{ng,23} + 0.6 \cdot d_{\text{MMD},25} + 0.2 \cdot d_{\text{MMD},23}. \quad (3.13)$$

The weights are chosen to rebalance the sample and upweight bright galaxies (below i magnitude of 23), which are the target of Stage III cosmological surveys, and would otherwise only account for 20% of the sample.

3.5.2 Iteration engine and posterior modelling

We use sets of HSC patches randomly selected without replacement and increase their number in each iteration: the first set has $|d_1| = 10$ patches. Using fewer images at the beginning of the ABC allows us to eliminate very unlikely areas of parameter space without wasting computing time. We then add 1 patch at every subsequent iteration. If the sets of patches are representative of the full sample, then our factorized ABC procedure corresponds to a standard rejection ABC, with reduced computational costs. If the sub-sample of images in one iteration is biased or has local properties, then the resulting ABC posterior could be biased. We consider the probability that more than 10 randomly selected images would present a systematic bias very unlikely and aggregate the distance measures computed from the different images using a median, which is robust to outliers. This way, even if some of the patches have local effects, we still obtain reliable ABC posteriors at each iteration.

In each iteration, we simulate the $|d_n|$ patches for 10000 parameter samples and keep the 2000 samples with the lowest combined distance as the posterior. Then, we perform a density estimation for the posterior using a Gaussian Mixture Model (GMM) with 20 components. We draw the new 10000 samples from this GMM to create a resampled posterior, and pass it as a prior to the next iteration. The GMM fitting is performed in a transformed space to make it easier for the GMM to capture non-Gaussian distributions. We check that the GMM fitting accurately resamples the posterior and monitor the evolution of the distance measures at each iteration. When there is no more improvement in any of the distance metrics, we stop iterating the algorithm. This stopping condition is similar to that of [34], where we look at the evolution of each distance separately because the combined distance is rescaled differently at each iteration. We ran 23 iterations of the algorithm. The details of these iterations are summarised in table 6 in appendix B. Appendix A introduces the prior that we used for the ABC run, gives an overview of all model parameters and reports the resulting mean and standard deviation of each parameter in the posterior.

4 Results

In this section, we present the results obtained from tuning our galaxy population model. We iteratively performed ABC inference on randomly selected batches of images taken from the HSC deep fields and complemented with reweighted COSMOS2020 many-band photometric redshifts as described in sections 3.4, 3.5 and appendix B. We show the resulting posterior distribution of the model parameters in the following section. Then, we use samples from the posterior to run simulations in the COSMOS field and compare the photometric properties obtained by running SExtractor with the same settings on simulations and real data. We choose to use these patches as validation set, since we have redshift information for individual objects from the photo-z codes LePhare and EAZY. We conclude with a comparison of the obtained redshift distributions with the COSMOS2020 catalog at different magnitude cuts.

4.1 ABC posterior

Figures 4, 5, and 6 show the posterior distribution obtained after 23 iterations of the ABC algorithm, where we fulfill our stopping condition (see section 3.5). The model has 46

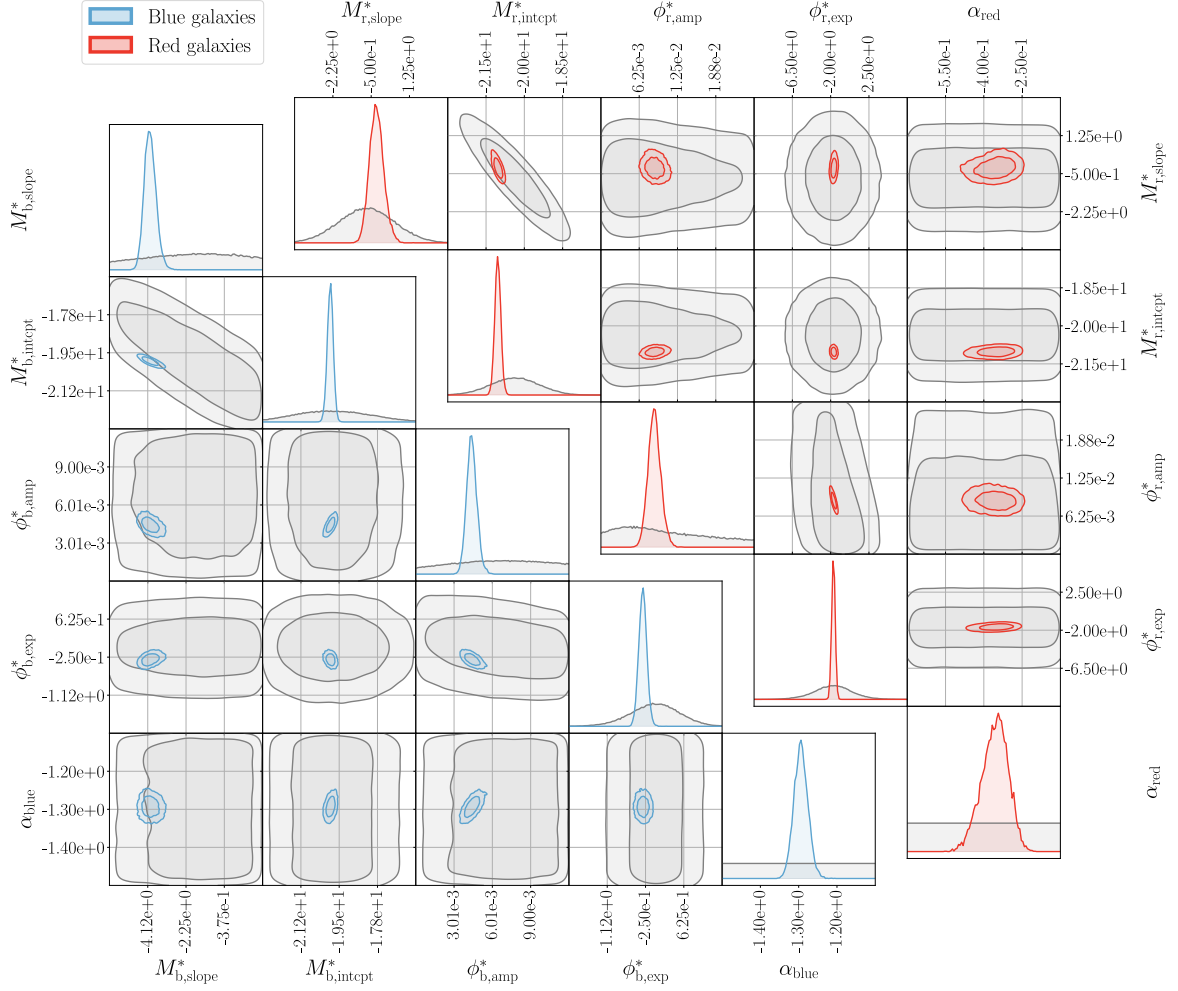


Figure 4. ABC posterior for the luminosity function parameters. The parameters for blue and red galaxies are shown in light blue and red, respectively. The ABC prior is shown in grey.

parameters, that we show divided into three categories for clarity: luminosity function parameters, SED template coefficients and galaxy morphology parameters.

The luminosity function parameters are the most relevant for determining the redshift distribution. In figure 4 we show the parameters for blue galaxies in the lower left triangle and for red galaxies in the upper right triangle. We also plot the prior distribution in grey. We notice that the parameters for blue galaxies are better constrained than those for red galaxies, possibly because blue galaxies are more abundant. The parameters M_{intcpt}^* are generally very well constrained by the ABC, and we notice several strong correlations between the luminosity function parameters, most notably between M_{intcpt}^* and M_{slope}^* and between ϕ_{amp}^* and ϕ_{exp}^* . The parameter α_{blue} we obtain is very close to the fiducial value of -1.3 whereas $\alpha_{\text{red}} \approx -0.35$ is slightly higher than the fiducial -0.5 .

In figure 5 we show the constraints on the SED templates coefficients. We show the prior, obtained from the catalog level ABC run described in appendix A.1, in grey. The posterior distribution of the parameters controlling the blue galaxy population is shown in

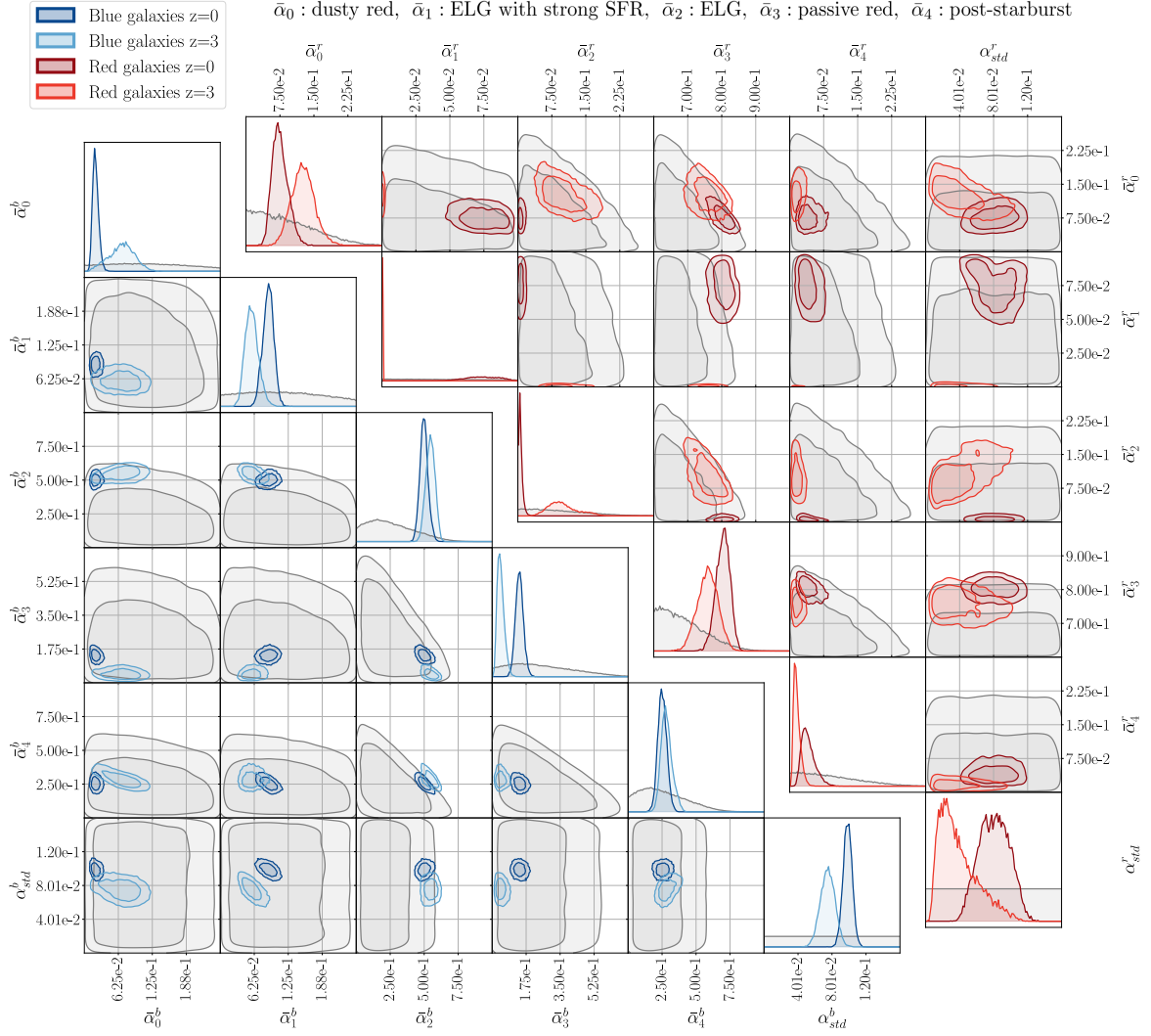


Figure 5. ABC posterior for the parameters controlling the spectral energy distributions. The parameters for blue galaxies at redshift 0 are displayed in light blue and in dark blue at redshift 3. Similarly, the parameters controlling red galaxies at redshift 0 are in red and at redshift 3 in dark red. The modes of the Dirichlet coefficients are encoded in the first five parameters and one final parameter controls the standard deviation of the Dirichlet distribution. The ABC prior is shown in grey.

the lower triangle of figure 5 (in dark blue at redshift $z=0$ and light blue at redshift $z=3$) and that controlling red galaxies in the upper triangle (in dark red for redshift $z=0$ and light red for redshift $z=3$). The $\bar{\alpha}_i$ parameters are non-negative, since a template cannot contribute less than zero to the galaxy SED. The fact that the posterior of some coefficients is very close to the lower bound of the prior is thus a sign of that coefficient not contributing to the SED of a galaxy type at redshift 0 or 3. We notice that the red galaxies are less constrained, especially at high redshift. The passive galaxy template (T_3) is dominating the SED of red galaxies, as imposed by the prior and there is a significant contribution of the dusty red template (T_0), more prominent at $z = 3$. The contribution of the ELG template (T_2) increases with redshift, whereas that of the ELG with strong star formation template

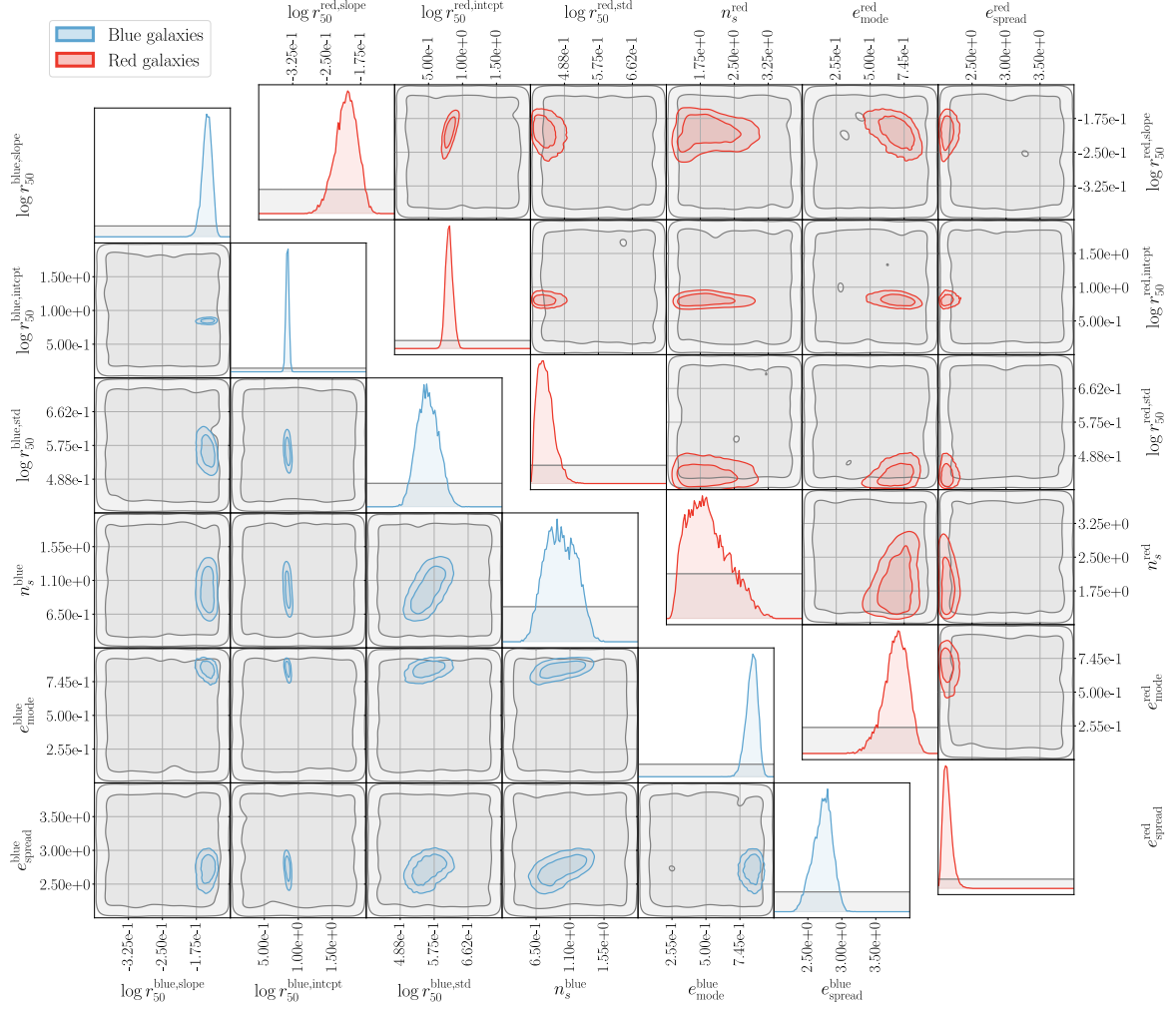


Figure 6. ABC posterior for the parameters controlling galaxy morphology. In light blue we show the parameters for blue galaxies and in red for red galaxies. The ABC prior is shown in grey.

(T_1) decreases with redshift. This contamination indicates that our galaxy population coming from the red luminosity function is not completely passive (probably includes galaxies from the green valley). The post-starburst template (T_4) contributes little and decreases with redshift. The blue galaxies are better constrained and dominated by the ELG template (T_2). The post-starburst template (T_4) is also present in the blue galaxy population at all redshifts, whereas the contribution of the ELG template with strong star formation (T_1) slightly decreases with redshift. The contribution of the dusty red template (T_0) increases with redshift, whereas that of the passive red template (T_3) decreases with redshift. We should not overinterpret the mixture of modes of the template coefficients, since our model allows large freedom due to the scatter parameters and the mixture of templates with redshift evolving coefficients. The KCORRECT templates (shown in figure 1) are derived from SDSS data [65] and do not provide an accurate characterization of galaxies at high redshifts. We also need to consider that there is a smooth transition between red and blue galaxies and that these two categories might not provide a good description of high redshift galaxies.

In figure 6, we show the parameters describing galaxy morphology. These are considerably more difficult to constrain for our distance metrics, since the effects of the parameters are often degenerate and describe the full intensity profile of the galaxies, which is not sufficiently captured by ellipticity and radius alone. We again show the prior in grey and the posterior distributions for blue and red galaxies in red and blue respectively. We note that the best constrained parameter is $\log r_{50}^{\text{intcpt}}$ for both blue and red galaxies, whereas other parameters are less constrained. The posterior on $e_{\text{spread}}^{\text{red}}$ is close to the prior boundary. This is probably due to a prior volume effect: the Beta distribution becomes uniform when $e_{\text{spread}}^{\text{red}} = 2$ thus making $e_{\text{mode}}^{\text{red}}$ irrelevant. We do not extend the prior below $e_{\text{spread}}^{\text{red}} = 2$, since this would correspond to U-shaped distributions of the intrinsic ellipticities which are unphysical. The posterior on $\log r_{50}^{\text{red, std}}$ is also close to the prior boundary. We plan to extend the prior in future work. We test the impact on mean redshift of changing this parameter outside the prior range and obtain a negligible shift. The distribution of Sersic indices for red galaxies remains very broad, encompassing values between 1.5 and 3, whereas the distribution for blue Sersic indices is centered at ≈ 1 . Sersic indices lack redshift evolution in our model, which is observationally measured (high redshift galaxies have smaller Sersic indices [66]).

4.2 Comparison of simulations and real data

We sample 30 parameter configurations from the ABC posterior at random and simulate the 56 HSC images overlapping with the COSMOS2020 COMBINED footprint. We run SExtractor consistently in dual-image mode using the i band image for detection on the 30 simulations and the real data and compare the obtained photometric properties of galaxies. We show 2D contours and 1D histograms of selected photometric properties of simulated galaxy samples and HSC real data up to MAG_APER3_i of 25 in figure 7. In the lower triangle we show MAG_AUTO in the r, i, z bands and FLUX_RADIUS in the i band. The magnitudes show excellent agreement both in the 1D projections and in the 2D contours. The sizes, on the other hand, are more discrepant: there is a tail of large galaxies in the simulations and also a population of galaxies smaller than the smallest galaxies in the data. This is an indication of limitations in the modelling of galaxy morphology. Our model does not include size evolution with redshift at fixed absolute magnitude, which can cause a model bias in the size distribution. We do not expect the lack of size evolution with redshift to impact the redshift distributions, since our sample selection in size is very broad. We plan to explore extensions of the size model in future work. The upper triangle of figure 7 shows the colours of galaxies and their correlations. Colours are very important, since they strongly correlate with redshift. We observe a rather good agreement of the colour distributions between the simulations and the real data, with some differences in the tails of the distributions.

In order to highlight the agreement and the discrepancies between colours, sizes and magnitudes, and also display their evolution with redshift, we include a scatter plot in figure 8. We show the relation between i band MAG_AUTO and FLUX_RADIUS, $r - i$ and $i - z$ colours and redshift (from the LePhare photo- z code for the real data). We observe that many trends are present both in the data and simulations: in particular the colour-redshift degeneracies are well reproduced by the simulations up to $z \approx 4$. Since simulating realistic galaxy colours is generally a challenging task, this highlights the effectiveness and precision of this method.

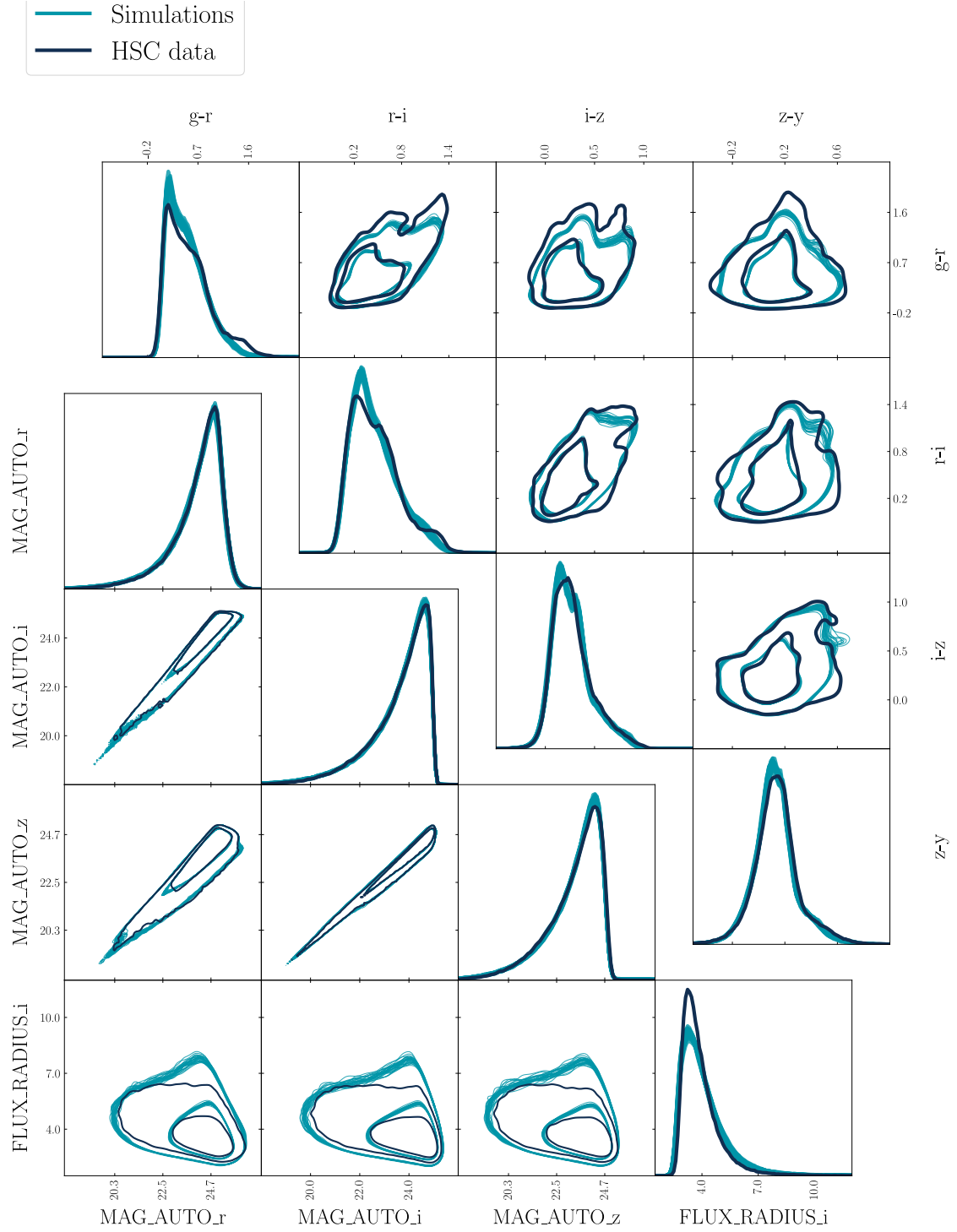
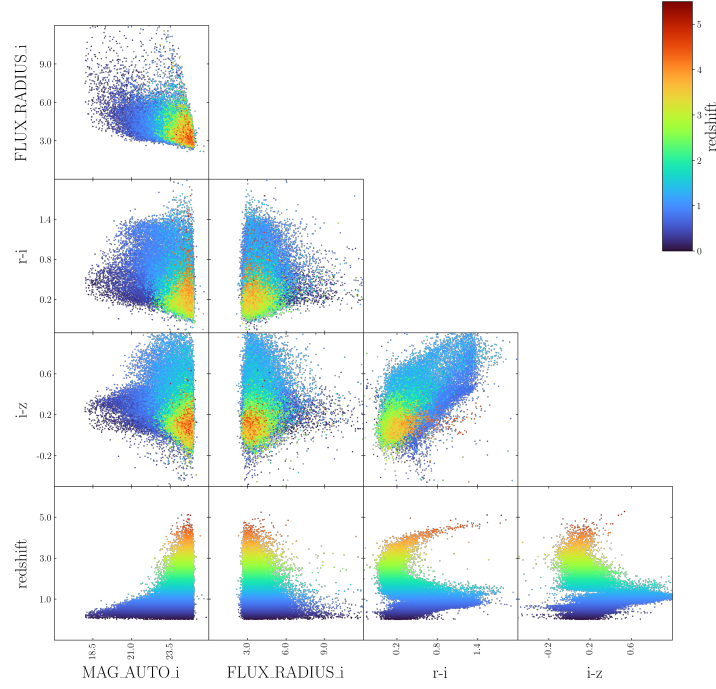
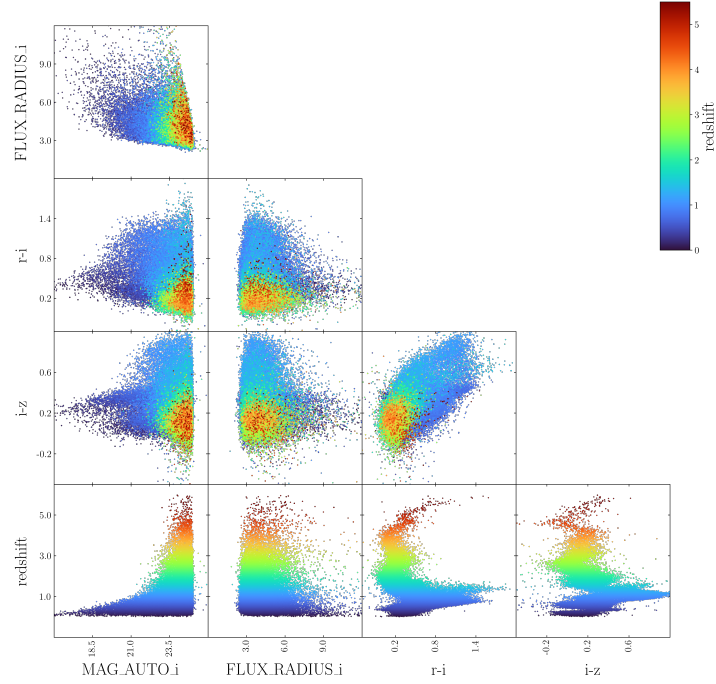


Figure 7. Comparison of selected photometric properties (MAG_AUTO in r , i and z band, $FLUX_RADIUS$ in the i band and colours) of HSC real data (in dark blue) and 30 simulations (in teal) in the COSMOS field.



(a) Scatter plot of selected photometric properties of the data in the COSMOS field. Redshift refers to LePhare photo-zs.



(b) Scatter plot of selected photometric properties of one simulation in the COSMOS field.

Figure 8. Scatter plots of real data and one of the simulations in the COSMOS field. We include `MAG_AUTO` and `FLUX_RADIUS` in the i band, $r - i$ and $i - z$ colours and redshift (from COSMOS2020 CLASSIC LePhare in the case of the real data). Each point corresponds to a randomly selected galaxy from the catalog.

We note that the simulations include an excess of galaxies at very high redshifts ($z > 4.5$) that are not observed in the data. Due to the relation between absolute magnitude and size (and the evolution of angular diameter distances with redshift), these objects are large enough to be detected. This heavy high- z tail is likely to bias the redshift distribution. The current parametrization of the luminosity function implies that the characteristic absolute magnitude $M^*(z)$ becomes brighter at higher redshifts (with a $\log(1+z)$ redshift dependence). In a Universe that grows hierarchically, this is not the case at high enough redshifts. One possibility would be to truncate the growth of $M^*(z)$ at a redshift z_{max} so that $M^*(z) = \text{const}$ for $z > z_{\text{max}}$ but constraining this parameter would be difficult since it would affect only a small fraction of the objects. We also observe an abundance of large objects in the simulations at all redshifts (as already observed in the 1D FLUX_RADIUS histogram and in the comparison between real and simulated images). Our model does not account for the evolution of galaxy sizes with redshift at fixed absolute magnitude, which is seen in observations where high redshift galaxies are up to five times smaller than local galaxies (see [66] and references therein). We leave high redshift model refinements for future work.

4.3 Redshift distributions

Since the photometric properties of the simulations are in statistical agreement with the data, we derive the posterior redshift distributions. In figure 9 we show the $n(z)$ s from 30 simulations in the COSMOS field together with those from the COSMOS2020 photometric

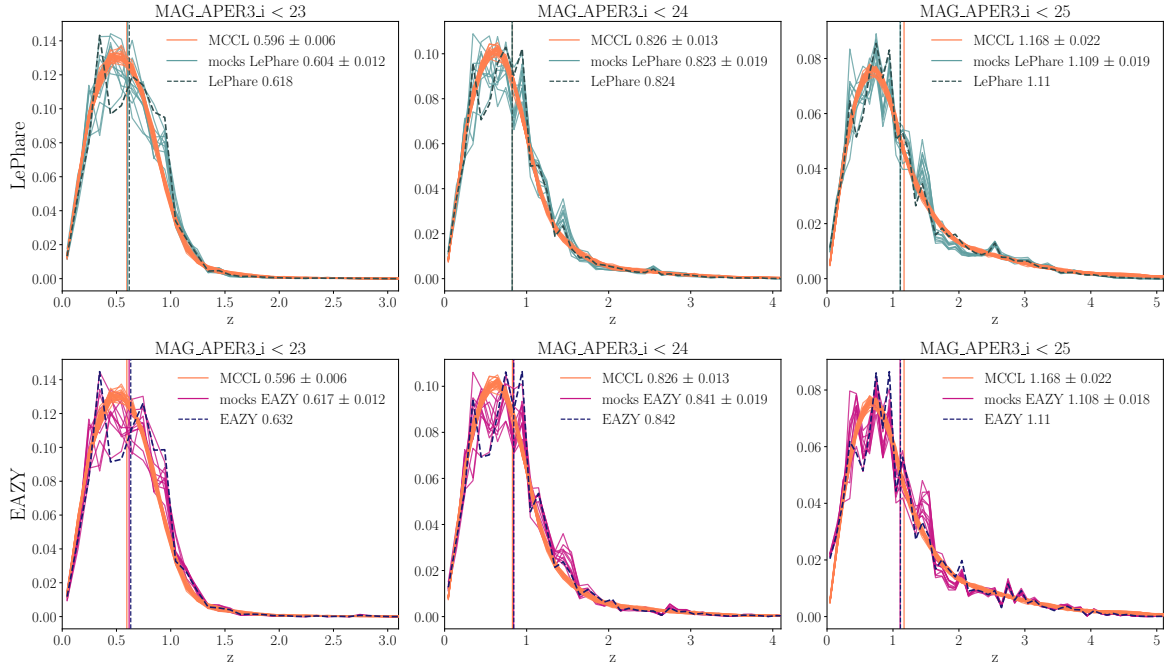


Figure 9. Redshift distributions in the COSMOS field, from 30 simulations from the ABC posterior (in orange), from COSMOS2020 photo- z s and from the COSMOS mocks generated by reweighting the COSMOS2020 photo- z s based on HSC deep fields photometry (as explained in section 3.4) in 10 sets of 56 patches. The top row shows LePhare photo- z s (in dark grey for COSMOS and sea green for the mocks) and the bottom row EAZY photo- z s (in dark blue for COSMOS and purple for the mocks).

redshift catalog in the same field. We also include the redshift distributions obtained by assigning COSMOS2020 photo- z s to 10 sets of 56 continuous patches selected at random in the HSC deep fields, using the reweighting procedure presented in 3.4 to reduce the impact of sample variance. We show three different magnitude cuts in the i band MAG_APER3 in the three columns ($\text{MAG_APER3_i} < 23, 24, 25$) and the two different photo- z codes used in COSMOS2020 (LePhare and EAZY) in the two rows. We notice that the redshift distributions obtained from MCCL are smooth due to the absence of clustering in the simulations.

We report mean redshift errors for our simulations, corresponding to the standard deviation of the means of the 30 simulations. We estimate an error on the mean redshift of 0.002 at all magnitude cuts from COSMOS2020 using Bootstrap. Considering sample variance and the systematic offsets between the two photo- z estimates in COSMOS2020 reported in section 2.3.1 and 2.3.2, we expect a minimum error of 0.02 in the mean redshift from COSMOS2020. By summing this in quadrature with the MCCL errors, we obtain a rough estimation of the combined errors $\sigma_{23} \approx 0.02$, $\sigma_{24} \approx 0.025$ and $\sigma_{25} \approx 0.03$ for the three magnitude cuts at $\text{MAG_APER3_i} = 23, 24, 25$. We report the mean redshifts and errors in table 2. Our estimated redshift distribution for i band magnitude cut of 23 has low mean and is more concentrated around the mean than the $n(z)$ of COSMOS2020, especially when compared to EAZY photo- z s. This can be partly explained by the fact that the simulations are only affected by shot noise due to the limited number of objects, but do not include sample variance, since the objects are randomly distributed in space without accounting for clustering. When considering the $n(z)$ distributions reweighted according to deep field photometry in figure 9, we observe the shift of ≈ 0.015 towards lower redshift reported in table 1. To make sure that the systematic offset is not due to the reweighting methodology, we also create mocks from the simulations by assigning the redshifts from a simulated COSMOS field to equal area sets of patches from other simulated HSC deep fields. The obtained redshift shifts are negligible ($\Delta z \approx 0.001$ at all magnitude cuts). The sample variance reduction procedure leads to a 1σ agreement between the redshift distribution of the simulations and the COSMOS2020 data at magnitude cut of 23 in the i band. The redshift distribution for objects below i band magnitude of 24 is in excellent agreement with COSMOS2020. The presence of heavier high redshift tails in the simulations, originating from the extrapolated redshift growth of $M^*(z)$ in the luminosity function parametrization, as explained in section 4.2, reduces the mean redshift agreement for the $\text{MAG_APER3_i} < 25$ sample to 2σ . Removing all simulated objects at $z > 4.5$ reduces the MCCL mean redshift estimate of this sample to $\bar{z}_{\text{MCCL},25} = 1.131 \pm 0.014$, in better agreement with COSMOS2020.

5 Conclusions

Redshift calibration is one of the key systematics affecting cosmic shear measurements. Shifts in the mean of the $n(z)$ lead to biased cosmological constraints from large-scale structure surveys [8–15]. It is therefore important to explore and combine a wide range of different methodologies to infer accurate photometric redshift distributions.

In this work, we presented a simulation-based inference approach to obtain redshift distributions from coadded telescope images, extending the result from [31] to deep HSC data and increasing the accuracy of the method. This choice of dataset enables us to test

	MAG_APER3_i < 23	MAG_APER3_i < 24	MAG_APER3_i < 25
\bar{z}_{MCCL}	0.596	0.826	1.168
\bar{z}_{LePhare}	0.618	0.824	1.110
$\bar{z}_{\text{LePhare, mocks}}$	0.604	0.823	1.109
\bar{z}_{EAZY}	0.632	0.842	1.110
$\bar{z}_{\text{EAZY, mocks}}$	0.617	0.841	1.108
σ_{MCCL}	0.006	0.013	0.022
σ_{combined}	0.02	0.025	0.03

Table 2. We report the mean redshifts obtained from MCCL in the COSMOS field, together with the LePhare and EAZY mean redshifts from COSMOS2020 at three different magnitude cuts (MAG_APER3_i < 23, 24, 25). We also report the LePhare and EAZY mean redshifts obtained when reweighting according to HSC deep photometry 10 sets of 56 images (mocks), the MCCL errors and a rough estimation of the combined errors.

our method in the regime of Stage IV surveys. We developed several extensions of the methodology, both in terms of modelling and inference. We calibrated the parameters of our galaxy population model using photometric properties from galaxies in the HSC deep fields and accurate photometric redshifts from COSMOS2020, and obtained realistic simulations. We report the resulting parameters of the model in table 3. We compared our results with photometric properties and photo- z s from the COSMOS2020 catalog, simulating the same area with the Subaru telescope in five broad bands, and found good agreement. We showed how sample variance in COSMOS has a strong impact on bright magnitude limited samples. We found a systematic redshift offset in the COSMOS field for objects below magnitude 23 in the HSC i band, common to both photometric redshift methods (LePhare and EAZY). Previous work [14, 29] found that the use of the COSMOS field high quality photo- z s alone for redshift calibration could bias low the mean retrieved redshift. This is not in contrast with our results, since redshift calibration strongly depends on the selection function and we are only estimating the sample variance in the COSMOS field itself, and not assessing the impact of using COSMOS2020 as a calibration sample within a methodology. Once this effect is taken into account, our simulations achieve 1σ agreement with the mean of the redshift distribution of COSMOS2020 up to MAG_APER3_i=24, and 2σ agreement up to MAG_APER3_i=25. The overall shape of the $n(z)$ agrees well. The presence of a high redshift tail at $z > 4.5$ requires further investigation and is an indication of model bias in the luminosity function.

Forward modelling has several advantages that can benefit cosmological large-scale structure surveys in different ways. On one hand, realistic simulations can be used to optimize the survey strategies and model the selection function to the needed level of accuracy. On the other hand, as done in this work, simulations can be used for calibration and to study effects that are difficult to model otherwise, for example the impact of blending and how unrecognized blends can affect the shear measurement. The MCCL method has good prospects of applicability to data from upcoming Stage IV surveys such as the Legacy Survey of Space and Time (LSST) [67] and Euclid [68, 69], which will have depths comparable to the HSC deep fields. The error on the mean redshift per tomographic bin required by these surveys is $\Delta z < 0.001(1+z)$, about an order of magnitude tighter than the current work. In

order to make this possible, a number of extensions and improvements are desirable. First of all, imaging of wider deep fields with many band photometry in order to reduce sample variance or deeper spectroscopy with clean selection cuts would greatly benefit photometric redshift calibration in general and forward modelling methods in particular. Secondly, it will be necessary to investigate the evolution of the luminosity function at high redshifts and find a suitable parametrization in order to avoid an excess of high redshift galaxies. This also includes designing good distance metrics to constrain the tail of the distribution using simulation-based inference. It would be beneficial to extend our galaxy population model to include effects that are well understood but not currently modelled, for example the size evolution of galaxies with redshift at fixed absolute magnitude and a relation between the absolute magnitude and colour of red galaxies (more massive galaxies are redder [70]). Another possible improvement of the model of morphologies is the inclusion of bulges and disks instead of a single Sersic profile, as well as the redshift evolution of the Sersic index. A very promising path to a more physically motivated modelling is the use of stellar population synthesis (SPS) models instead of spectral templates to model the galaxies' SEDs. This entails sampling a stellar mass function rather than two luminosity functions and constructing SEDs directly from physical properties of galaxies (such as star formation rates, metallicities and gas properties). This has become feasible in terms of computing time through the emulation of SPS models [33, 71–73]. Emulators can also be used to speed up the UFIG simulations, by mapping the transfer function between catalogs obtained from the galaxy population model and realistic detections. This requires a more detailed understanding of selection effects and the impact of blending. Faster simulations would allow us to test extensions of the model more extensively. Finally, we have discussed how sample variance can impact redshift distributions, when the area considered is limited. In order to obtain realistic sample variance in our simulations, we need to distribute galaxies following the underlying large-scale structure. This can be achieved with the required computational speed by using Subhalo Abundance Matching (SHAM) and approximate simulations as described in [74].

Acknowledgments

We thank Daniel Grün, John Weaver, Will Hartley, Pascale Berner and Jamie McCullough for useful discussions. This project was supported in part by grant 200021_192243 from the Swiss National Science Foundation. This paper is based on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Centre (ADC) at NAOJ.

COSMOS2020 is based on observations collected at the European Southern Observatory under ESO programme ID 179.A-2005 and on data products produced by CALET and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

We thank the High Performance Computing group at ETH Zurich for support with the Euler cluster, which was heavily used throughout this work. We submitted jobarrays to the cluster using `esub-epipe` [75, 76]. We used functionalities from several Python packages: `numpy` [77], `scipy` [78], and `matplotlib` [79]. Corner plots were created with `trianglechain` [15, 80].

A Galaxy population model priors

We use a similar model as [31, 34, 36], with a number of modifications, described in section 3. We summarize the model with a description of model parameters, prior distributions and allowed ranges in table 3. The model has 46 parameters, but 4 of them are redundant: modes of the template coefficients $\bar{\alpha}_i$ are always forced to sum to $\sum \bar{\alpha}_i = 1$. The luminosity function parameters use the same prior as [34], with standard deviation scaled by a factor of $\times 3$. The prior column in table 3 shows the distribution (Normal, Uniform, or Dirichlet) of the prior, as well as the additional bounds applied. For all variables using the Uniform distribution, a joint Sobol sequence was used to generate the prior. A suitable prior for the template coefficients $\bar{\alpha}_i$ is obtained through a catalog level ABC using the COSMOS2015 catalog [59]. We describe this procedure in the next subsection. Table 3 also lists the mean and standard deviation of each parameter’s 1D posterior distribution.

A.1 Template coefficient priors from COSMOS2015 catalog

As mentioned in section 3.1, we do not rely on the weights derived in [31] using the New York University Value-Added Galaxy Catalog to differentiate the SED between red and blue galaxies, but impose different priors on the Dirichlet $\bar{\alpha}_i$ parameters for the two galaxy types. This is motivated by the changes to the spectral energy distribution modelling described in section 3.1, that make the model more interpretable. To capture redshift-colour dependencies of higher redshift galaxies, we derive these priors from a comparison with the COSMOS2015 catalog [59].

First, we select galaxies from this catalog using the following cuts: $z \in [0.3, 4]$, $\text{mag}_{\text{ip}} \in [10, 24.5]$, $\text{TYPE} == 0$. The comparison between the simulated and observed galaxies is performed using their redshift and colours, defined with respect to the reference band. We use colour and redshift to avoid constraining the luminosity function, and exploit just the colour-redshift information. We use the reference band ip and compute the colours as a difference with bands NUV, u, B, V, r, zpp, Y, J, H, Ks. For the simulated galaxies, these magnitudes are calculated using UFIG up to catalog generation, with the use of filters provided in the COSMOS2015 dataset. The comparison is performed for red and blue galaxies separately. In the real data we use the **CLASS** provided by COSMOS2015 to separate between star-forming and quiescent galaxies (classified using NUV-r / r - J diagram). We use the nearest-neighbour estimator of Universal Divergence [81] as a distance metric between simulations and real data.

We do not intend to create a posterior on the template coefficient values, but rather find the upper and lower limits on the Dirichlet modes of template coefficient values $\bar{\alpha}_i$. We then perform an iterative procedure of progressively narrowing down the ranges for all coefficients. Starting with a uniform range for $\bar{\alpha}_i \in [0, 1]$, we generate 10000 samples from the luminosity

	Parameter	Meaning	Prior	Posterior
Luminosity functions (10)	$M_{b,\text{slope}}^*$	Slope of the redshift evolution of the parameter M^* in the Schechter LF for blue galaxies, see equation (3.2)	Prior from [34] $\times 3$, $\in [-6, 1.5]$	-4.0 ± 0.3
	$M_{b,\text{intcpt}}^*$	Intercept of the redshift evolution of the parameter M^* in the Schechter LF for blue galaxies, see equation (3.2)	Prior from [34] $\times 3$, $\in [-23, -16]$	-19.9 ± 0.1
	$M_{r,\text{slope}}^*$	Slope of the redshift evolution of the parameter M^* in the Schechter LF for red galaxies, see equation (3.2)	Prior from [34] $\times 3$, $\in [-4, 3]$	-0.3 ± 0.3
	$M_{r,\text{intcpt}}^*$	Intercept of the redshift evolution of the parameter M^* in the Schechter LF for red galaxies, see equation (3.2)	Prior from [34] $\times 3$, $\in [-23, -17]$	-21.0 ± 0.1
	$\phi_{b,\text{exp}}^*$	Exponent of the redshift evolution of the parameter ϕ^* in the Schechter LF for blue galaxies, see equation (3.3)	Prior from [34] $\times 3$, $\in [-2, 1.5]$	-0.31 ± 0.09
	$\phi_{b,\text{amp}}^*$	Amplitude of the redshift evolution of the parameter ϕ^* in the Schechter LF for blue galaxies, see equation (3.3)	Prior from [34] $\times 3$, $\in [1.1\text{e}-5, 1.2\text{e}-2]$	0.0044 ± 0.0004
	$\phi_{r,\text{exp}}^*$	Exponent of the redshift evolution of the parameter ϕ^* in the Schechter LF for red galaxies, see equation (3.3)	Prior from [34] $\times 3$, $\in [-11, 7]$	-1.7 ± 0.2
	$\phi_{r,\text{amp}}^*$	Amplitude of the redshift evolution of the parameter ϕ^* in the Schechter LF for red galaxies, see equation (3.3)	Prior from [34] $\times 3$, $\in [2\text{e}-8, 2.5\text{e}-2]$	0.009 ± 0.001
	α_{blue}	Steepness of the faint-end slope in the Schechter LF for blue galaxies, see equation (3.1)	$\mathcal{U}[-1.5, -1.1]$	-1.29 ± 0.02
Galaxy morphology (12)	$\log r_{50}^{\text{blue/red slope}}$	Slope of the evolution of the average intrinsic physical size of galaxies with absolute magnitude	$\mathcal{U}[-0.4, -0.1]$	b: -0.15 ± 0.01 r: -0.21 ± 0.02
	$\log r_{50}^{\text{blue/red intcpt}}$	Intercept of the evolution of the average intrinsic physical size of galaxies with absolute magnitude	$\mathcal{U}[0, 2]$	b: 0.84 ± 0.02 r: 0.81 ± 0.05
	$\log r_{50}^{\text{blue/red std}}$	Standard deviation of the normal distribution we use to sample intrinsic physical galaxy sizes	$\mathcal{U}[0.4, 0.75]$	b: 0.56 ± 0.03 r: 0.44 ± 0.02
	n_s^{blue}	Mode of the Sersic index distribution of blue galaxies	$\mathcal{U}[0.2, 2]$	1.0 ± 0.2
	n_s^{red}	Mode of the Sersic index distribution of red galaxies	$\mathcal{U}[1, 4]$	2.0 ± 0.4
	$e_{\text{mode}}^{\text{blue/red}}$	Ellipticity distribution mode for blue/red galaxies	$\mathcal{U}[0.01, 0.99]$	b: 0.83 ± 0.04 r: 0.69 ± 0.08
	$e_{\text{spread}}^{\text{blue/red}}$	Ellipticity distribution spread for blue/red galaxies	$\mathcal{U}[2, 4]$	b: 2.7 ± 0.1 r: 2.14 ± 0.06
	$\bar{\alpha}_{i,0/3}^{\text{blue/red}}$	Normalized Dirichlet concentration parameters at $z=0/3$ from which the template coefficients for blue/red galaxies are sampled, $i=0, \dots, 4$, $\sum_i \bar{\alpha}_i=1$	5-dimensional Dirichlet, see A.1 for details	See table 5
SED coeff. (24)	$\alpha_{\text{std},0/3}^{\text{blue/red}}$	Standard deviation of the normalized Dirichlet concentration parameters at $z=0/3$ from which the template coefficients for blue/red galaxies are sampled	$\mathcal{U}[1\text{e}-4, 0.16]$	See table 5

Table 3. Table with galaxy population model parameters, priors and resulting 1D posteriors. Luminosity function is shortened as LF. The details of the prior ranges of the Dirichlet distributions for the template coefficients are explained in subsection A.1. We briefly describe the model parameters and report the type of prior and the its boundaries. The last column shows the resulting mean value and standard deviation from the ABC posterior.

parameter	$\bar{\alpha}_{0,z}^{\text{blue}}$	$\bar{\alpha}_{1,z}^{\text{blue}}$	$\bar{\alpha}_{2,z}^{\text{blue}}$	$\bar{\alpha}_{3,z}^{\text{blue}}$	$\bar{\alpha}_{4,z}^{\text{blue}}$	$\bar{\alpha}_{0,z}^{\text{red}}$	$\bar{\alpha}_{1,z}^{\text{red}}$	$\bar{\alpha}_{2,z}^{\text{red}}$	$\bar{\alpha}_{3,z}^{\text{red}}$	$\bar{\alpha}_{4,z}^{\text{red}}$
lower limit	0	0	0	0	0	0	0	0	0.6	0
upper limit	0.25	0.25	1	0.7	1	0.3	0.1	0.3	1	0.3

Table 4. Upper and lower limits on the modes of the Dirichlet coefficients $\bar{\alpha}_i$ for blue and red galaxies derived from the catalog level ABC. The same boundaries are imposed at redshifts $z = 0$ and $z = 3$. These are the upper and lower limits of the Dirichlet priors of the subsequent ABC run on HSC DUD data.

type, z	$\bar{\alpha}_0$	$\bar{\alpha}_1$	$\bar{\alpha}_2$	$\bar{\alpha}_3$	$\bar{\alpha}_4$	α_{std}
blue, $z = 0$	0.021 ± 0.005	0.09 ± 0.01	0.50 ± 0.03	0.14 ± 0.02	0.25 ± 0.03	0.099 ± 0.005
blue, $z = 3$	0.07 ± 0.02	0.06 ± 0.01	0.55 ± 0.03	0.04 ± 0.02	0.28 ± 0.04	0.075 ± 0.009
red, $z = 0$	0.08 ± 0.02	0.07 ± 0.01	0.006 ± 0.004	0.80 ± 0.02	0.04 ± 0.02	0.08 ± 0.02
red, $z = 3$	0.13 ± 0.02	$1e - 5 \pm 2e - 5$	0.10 ± 0.03	0.76 ± 0.03	0.016 ± 0.008	0.04 ± 0.02

Table 5. ABC posterior means and standard deviations of the SED coefficients for red and blue galaxies at redshift $z = 0$ and $z = 3$.

function prior described in section 3.1. We then calculate Universal Divergence between the redshifts and colours from the simulated and COSMOS galaxies. The columns are scaled before the comparison, using a robust scaler from the `scikit-learn` package. We select 2000 best points and calculate the lower and upper limits on $\bar{\alpha}_i$. We then input these new limits and generate another 10000 samples with them. We repeat this process 20 times. The ranges for the coefficients obtained from this procedure define the prior in the main ABC run and are shown in table 4, with coefficients rounded roughly to 0.1. The obtained ranges agree with expectations for both blue and red galaxies, with $\bar{\alpha}_2$ and $\bar{\alpha}_3$ dominating for the blue and red galaxies, respectively. We report the means and standard deviations of the obtained posterior of the SED template coefficients in table 5.

B Details of the ABC runs

In this appendix, we describe the details of our ABC analysis that were omitted in section 3.5. The ABC iteration engine is similar to the one presented in [34] and depends on a sequence of prior-to-posterior iterations.

Iterations. We start by sampling 10000 parameter configurations from the prior defined in table 3, with limits in the Dirichlet coefficients $\bar{\alpha}_i$ from table 4. We discard samples when the simulation fulfills one of the rejection criteria: (i) having more than 1 million blue or red galaxies (ii) having less than 300 or more than 20000 objects below magnitude 24 in the i band. These are considered extreme conditions, that no simulation that is similar to the real data would fulfil and help us restrict to more likely parts of parameter space, without an excessive use of computing time. For each of the 10000 configurations, we simulate 10 HSC patches in the first iteration, as described in section 3.5. In the following iterations, we increase the number of simulated patches by 1, whereas the number of parameter configurations

iteration	$N_{p,\text{sim}}$	sky area
1	10	0.38 deg ²
2	11	0.42 deg ²
3	12	0.46 deg ²
4	13	0.5 deg ²
5	14	0.54 deg ²
6	15	0.58 deg ²
7	16	0.61 deg ²
8	17	0.65 deg ²
9	18	0.69 deg ²
10	19	0.73 deg ²
11	20	0.77 deg ²
12	21	0.8 deg ²
13	22	0.85 deg ²
14	23	0.88 deg ²
15	24	0.92 deg ²
16	25	0.96 deg ²
17	26	1 deg ²
18	27	1.04 deg ²
19	28	1.08 deg ²
20	29	1.11 deg ²
21	30	1.15 deg ²
22	31	1.19 deg ²
23	32	1.23 deg ²

Table 6. Number of patches used in each iteration and corresponding sky area covered by the HSC DUD patches.

is fixed to 10000. Table 6 shows the number of patches simulated per iteration and the corresponding sky area in deg².

Sample selection. The distance measures described in section 3.5.1 are computed using the SExtractor catalogs, created in all *grizy* bands, based on the detection in the *i* band. We perform the PSF estimation using a Convolutional Neural Network [60], in the same way as in [32]. We run SExtractor on the HSC data first, and then on the simulated images during the ABC iterations. From the catalogs, we select galaxies with the following set of cuts applied in all bands with strict **and** conditions:

$$\begin{aligned}
& \text{FLAGS} < 4, \quad 14 < \text{MAG_APER3} < 30, \quad \text{MAG_AUTO} < 99, \quad 0.1 < \text{FLUX_RADIUS} < 100, \\
& -3 < \log_{10}(\text{SNR}) \equiv \log_{10}\left(\frac{\text{FLUX_AUTO}}{\text{FLUXERR_AUTO}}\right) < 4, \quad 0 < \text{ELL} < 1, \\
& \text{N_EXPOSURES} > 0, \quad 0.5 < r_{50}/\text{PSF_FWHM}, \quad \text{CLASS_STAR} < 0.95
\end{aligned}$$

where `ELL` is the absolute ellipticity calculated from windowed moments `**_WIN_IMAGE`, `N_EXPOSURES` is the number of exposures in the coadd at the position of the object, r_{50} is the object size defined as $r_{50} = 2 \cdot \ln(2) \cdot (X2_WIN_IMAGE + Y2_WIN_IMAGE)^{1/2}$, as in [32]. Both the PSF size cut and the `CLASS_STAR` cut are applied to create a pure galaxy sample. In addition, we impose `MAG_APER3` < 25 in the i band. Note that the `SEXTRACTOR` detections are matched in the simulations to the true properties of the injected galaxies, as explained in section 3.2, so that a further criterion for simulated objects to be selected for the MMD distances is that the detection has been matched to a true simulated object. We additionally require that the objects do not lie on the image mask. When computing $d_{\text{MMD},23}$ we additionally impose `MAG_APER3_i` < 23 .

Optimization of the kernel radius parameter. To obtain the most sensitive MMD distance, it is common to optimize the parameters of the kernel used to compute it [62]. We use a Gaussian kernel with a single parameter σ , corresponding to the correlation scale. We compute σ for the different MMD distances as the median distance between samples drawn from the same probability distribution (the real data) [31, 62].

Modelling of posterior distributions. We create the posterior distribution at each iteration by setting the 20th percentile as a threshold, thus selecting the 2000 out of 10000 samples with the smallest combined distance. We then create a model of this posterior using a Gaussian Mixture Model (GMM), from the `scikit-learn` implementation.⁷ We use 20 Gaussians to fit the distribution. Before fitting, we transform the model parameter samples to a gaussianized space. This is done to make the GMM more suited for fitting the distribution, especially for parameters with uniform priors. First we rescale the parameters to lie between $\in [1e-8, 1 - 1e-8]$, and then apply a Gaussian inverse-CDF transform. We draw 10000 new samples in the gaussianized space using Sobol sampling and invert them back to the original space. We verify that the GMM model in the transformed space is a good representation of the posterior by comparing the 2D marginal projections of all parameter combinations for the original 2000 samples and 10000 new GMM samples. The GMM samples from the model posterior are passed as priors to the next iteration of the ABC algorithm.

⁷scikit-learn.org/stable/modules/generated/sklearn.mixture.GaussianMixture.html.

C SExtractor settings

We report in table 7 the SExtractor configuration.

SExtractor parameter	Value
CATALOG_TYPE	FITS_1.0
DETECT_TYPE	CCD
DETECT_MINAREA	5
THRESH_TYPE	RELATIVE
DETECT_THRESH	1.5
ANALYSIS_THRESH	1.5
FILTER	Y
FILTER_NAME	gauss_3.0_5x5.conv
DEBLEND_NTHRESH	32
DEBLEND_MINCONT	0.00001
CLEAN	Y
CLEAN_PARAM	1.0
MASK_TYPE	CORRECT
MAG_ZEROPOINT	27
PIXEL_SCALE	0.168
STARNNW_NAME	default.nnw
BACK_SIZE	128
BACK_FILTERSIZE	3
BACKPHOTO_TYPE	LOCAL
BACKPHOTO_THICK	24
WEIGHT_TYPE	NONE

Table 7. SExtractor configuration used in this work both on real images and simulations. The missing parameters change per patch and band and are described in the text.

References

- [1] DES collaboration, *The Dark Energy Survey: more than dark energy — an overview*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 1270 [[arXiv:1601.00329](#)] [[INSPIRE](#)].
- [2] ASTRO-WISE and KiDS collaborations, *The Kilo-Degree Survey*, *Exper. Astron.* **35** (2013) 25 [[arXiv:1206.1254](#)] [[INSPIRE](#)].
- [3] H. Aihara et al., *The Hyper Suprime-Cam SSP Survey: Overview and Survey Design*, *Publ. Astron. Soc. Jap.* **70** (2018) S4 [[arXiv:1704.05858](#)] [[INSPIRE](#)].
- [4] DES collaboration, *Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing*, *Phys. Rev. D* **105** (2022) 023520 [[arXiv:2105.13549](#)] [[INSPIRE](#)].
- [5] C. Heymans et al., *KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints*, *Astron. Astrophys.* **646** (2021) A140 [[arXiv:2007.15632](#)] [[INSPIRE](#)].

- [6] S. More et al., *Hyper Suprime-Cam Year 3 results: Measurements of clustering of SDSS-BOSS galaxies, galaxy-galaxy lensing, and cosmic shear*, *Phys. Rev. D* **108** (2023) 123520 [[arXiv:2304.00703](#)] [[INSPIRE](#)].
- [7] J.A. Newman and D. Gruen, *Photometric Redshifts for Next-Generation Surveys*, *Ann. Rev. Astron. Astrophys.* **60** (2022) 363 [[arXiv:2206.13633](#)] [[INSPIRE](#)].
- [8] M. Salvato, O. Ilbert and B. Hoyle, *The many flavours of photometric redshifts*, *Nature Astron.* **3** (2018) 212 [[arXiv:1805.12574](#)].
- [9] D. Huterer, M. Takada, G. Bernstein and B. Jain, *Systematic errors in future weak lensing surveys: Requirements and prospects for self-calibration*, *Mon. Not. Roy. Astron. Soc.* **366** (2006) 101 [[astro-ph/0506030](#)] [[INSPIRE](#)].
- [10] C.E. Cunha, D. Huterer, M.T. Busha and R.H. Wechsler, *Sample variance in photometric redshift calibration: cosmological biases and survey requirements*, *Mon. Not. Roy. Astron. Soc.* **423** (2012) 909 [[arXiv:1109.5691](#)] [[INSPIRE](#)].
- [11] D. Huterer, C.E. Cunha and W. Fang, *Calibration errors unleashed: effects on cosmological parameters and requirements for large-scale structure surveys*, *Mon. Not. Roy. Astron. Soc.* **432** (2013) 2945 [[arXiv:1211.1015](#)] [[INSPIRE](#)].
- [12] S. Joudaki et al., *CFHTLenS revisited: assessing concordance with Planck including astrophysical systematics*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 2033 [[arXiv:1601.05786](#)] [[INSPIRE](#)].
- [13] DES collaboration, *Dark Energy Survey Year 1 Results: Redshift distributions of the weak lensing source galaxies*, *Mon. Not. Roy. Astron. Soc.* **478** (2018) 592 [[arXiv:1708.01532](#)] [[INSPIRE](#)].
- [14] S. Joudaki et al., *KiDS+VIKING-450 and DES-Y1 combined: Cosmology with cosmic shear*, *Astron. Astrophys.* **638** (2020) L1 [[arXiv:1906.09262](#)] [[INSPIRE](#)].
- [15] S. Fischbacher, T. Kacprzak, J. Blazek and A. Réfrégier, *Redshift requirements for cosmic shear with intrinsic alignment*, *JCAP* **01** (2023) 033 [[arXiv:2207.01627](#)] [[INSPIRE](#)].
- [16] S. Arnouts et al., *Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North*, *Mon. Not. Roy. Astron. Soc.* **310** (1999) 540 [[astro-ph/9902290](#)] [[INSPIRE](#)].
- [17] O. Ilbert et al., *Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey*, *Astron. Astrophys.* **457** (2006) 841 [[astro-ph/0603217](#)] [[INSPIRE](#)].
- [18] N. Benítez, *Bayesian photometric redshift estimation*, *Astrophys. J.* **536** (2000) 571 [[astro-ph/9811189](#)] [[INSPIRE](#)].
- [19] R. Feldmann et al., *The Zurich Extragalactic Bayesian Redshift Analyzer (ZEBRA) and its first application: COSMOS*, *Mon. Not. Roy. Astron. Soc.* **372** (2006) 565 [[astro-ph/0609044](#)] [[INSPIRE](#)].
- [20] G.B. Brammer, P.G. van Dokkum and P. Coppi, *EAZY: A Fast, Public Photometric Redshift Code*, *Astrophys. J.* **686** (2008) 1503 [[arXiv:0807.1533](#)] [[INSPIRE](#)].
- [21] A.A. Collister and O. Lahav, *ANNz: Estimating photometric redshifts using artificial neural networks*, *Publ. Astron. Soc. Pac.* **116** (2004) 345 [[astro-ph/0311058](#)] [[INSPIRE](#)].
- [22] M. Carrasco Kind and R.J. Brunner, *Exhausting the Information: Novel Bayesian Combination of Photometric Redshift PDFs*, *Mon. Not. Roy. Astron. Soc.* **442** (2014) 3380 [[arXiv:1403.0044](#)] [[INSPIRE](#)].
- [23] I. Sadeh, F.B. Abdalla and O. Lahav, *ANNz2: Photometric Redshift and Probability Distribution Function Estimation using Machine Learning*, *Publ. Astron. Soc. Pac.* **128** (2016) 104502 [[arXiv:1507.00490](#)] [[INSPIRE](#)].

- [24] J. De Vicente, E. Sánchez and I. Sevilla-Noarbe, *DNF — Galaxy photometric redshift by Directional Neighbourhood Fitting*, *Mon. Not. Roy. Astron. Soc.* **459** (2016) 3078 [[arXiv:1511.07623](#)] [[INSPIRE](#)].
- [25] M. Lima et al., *Estimating the redshift distribution of photometric galaxy samples*, *Mon. Not. Roy. Astron. Soc.* **390** (2008) 118 [[arXiv:0801.3822](#)] [[INSPIRE](#)].
- [26] H. Hildebrandt et al., *KiDS+VIKING-450: Cosmic shear tomography with optical and infrared data*, *Astron. Astrophys.* **633** (2020) A69 [[arXiv:1812.06076](#)] [[INSPIRE](#)].
- [27] A.H. Wright, H. Hildebrandt, J.L. van den Busch and C. Heymans, *Photometric redshift calibration with self-organising maps*, *Astron. Astrophys.* **637** (2020) A100 [[arXiv:1909.09632](#)] [[INSPIRE](#)].
- [28] DES and eBOSS collaborations, *Dark Energy Survey Year 3 Results: clustering redshifts — calibration of the weak lensing source redshift distributions with redMaGiC and BOSS/eBOSS*, *Mon. Not. Roy. Astron. Soc.* **510** (2022) 1223 [[arXiv:2012.08569](#)] [[INSPIRE](#)].
- [29] DES collaboration, *Dark Energy Survey Year 3 results: redshift calibration of the weak lensing source galaxies*, *Mon. Not. Roy. Astron. Soc.* **505** (2021) 4249 [[arXiv:2012.08566](#)] [[INSPIRE](#)].
- [30] M.M. Rau et al., *Weak Lensing Tomographic Redshift Distribution Inference for the Hyper Suprime-Cam Subaru Strategic Program three-year shape catalogue*, *Mon. Not. Roy. Astron. Soc.* **524** (2023) 5109 [[arXiv:2211.16516](#)] [[INSPIRE](#)].
- [31] J. Herbel et al., *The redshift distribution of cosmological samples: a forward modeling approach*, *JCAP* **08** (2017) 035 [[arXiv:1705.05386](#)] [[INSPIRE](#)].
- [32] DES collaboration, *Monte Carlo control loops for cosmic shear cosmology with DES Year 1 data*, *Phys. Rev. D* **101** (2020) 082003 [[arXiv:1906.01018](#)] [[INSPIRE](#)].
- [33] J. Alsing et al., *Forward Modeling of Galaxy Populations for Cosmological Redshift Distribution Inference*, *Astrophys. J. Suppl.* **264** (2023) 29 [[arXiv:2207.05819](#)] [[INSPIRE](#)].
- [34] L. Tortorelli et al., *Measurement of the B-band galaxy Luminosity Function with Approximate Bayesian Computation*, *JCAP* **09** (2020) 048 [[arXiv:2001.07727](#)] [[INSPIRE](#)].
- [35] L. Tortorelli et al., *The PAU survey: measurement of narrow-band galaxy properties with approximate bayesian computation*, *JCAP* **12** (2021) 013 [[arXiv:2106.02651](#)] [[INSPIRE](#)].
- [36] A. Réfrégier and A. Amara, *A way forward for Cosmic Shear: Monte-Carlo Control Loops*, *Phys. Dark Univ.* **3** (2014) 1 [[arXiv:1303.4739](#)] [[INSPIRE](#)].
- [37] J. Bergé, L. Gamper, A. Réfrégier and A. Amara, *An Ultra Fast Image Generator (UFig) for wide-field astronomy*, *Astron. Comput.* **1** (2013) 23.
- [38] E. Bertin and S. Arnouts, *SExtractor: Software for source extraction*, *Astron. Astrophys. Suppl. Ser.* **117** (1996) 393 [[INSPIRE](#)].
- [39] C. Bruderer et al., *Calibrated Ultra Fast Image Simulations for the Dark Energy Survey*, *Astrophys. J.* **817** (2016) 25 [[arXiv:1504.02778](#)] [[INSPIRE](#)].
- [40] C. Bruderer et al., *Cosmic shear calibration with forward modeling*, *JCAP* **08** (2018) 007 [[arXiv:1707.06233](#)] [[INSPIRE](#)].
- [41] J.-C.J. Cuillandre et al., *Introduction to the CFHT Legacy Survey final release (CFHTLS T0007)*, in the proceedings of the *Observatory Operations: Strategies, Processes, and Systems IV* (2012) [[DOI:10.1117/12.925584](#)].

- [42] P. Martí et al., *Precise photometric redshifts with a narrow-band filter set: The PAU Survey at the William Herschel Telescope*, *Mon. Not. Roy. Astron. Soc.* **442** (2014) 92 [[arXiv:1402.3220](#)] [[INSPIRE](#)].
- [43] L. Tortorelli et al., *The PAU Survey: A Forward Modeling Approach for Narrow-band Imaging*, *JCAP* **11** (2018) 035 [[arXiv:1805.05340](#)] [[INSPIRE](#)].
- [44] EBOSS collaboration, *Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies and the Distant Universe*, *Astron. J.* **154** (2017) 28 [[arXiv:1703.00052](#)] [[INSPIRE](#)].
- [45] M. Fagioli et al., *Forward Modeling of Spectroscopic Galaxy Surveys: Application to SDSS*, *JCAP* **11** (2018) 015 [[arXiv:1803.06343](#)] [[INSPIRE](#)].
- [46] M. Fagioli et al., *Spectro-Imaging Forward Model of Red and Blue Galaxies*, *JCAP* **06** (2020) 050 [[arXiv:2002.04039](#)] [[INSPIRE](#)].
- [47] H. Aihara et al., *Third data release of the Hyper Suprime-Cam Subaru Strategic Program*, *Publ. Astron. Soc. Jap.* **74** (2022) 247 [[arXiv:2108.13045](#)] [[INSPIRE](#)].
- [48] J.R. Weaver et al., *COSMOS2020: A Panchromatic View of the Universe to $z \sim 10$ from Two Complementary Catalogs*, *Astrophys. J. Suppl.* **258** (2022) 11 [[arXiv:2110.13923](#)] [[INSPIRE](#)].
- [49] P. Sudek, L.F. de la Bella, A. Amara and W.G. Hartley, *The sensitivity of the redshift distribution to galaxy demographics*, *Mon. Not. Roy. Astron. Soc.* **516** (2022) 1670 [[arXiv:2112.11345](#)] [[INSPIRE](#)].
- [50] N. Scoville et al., *The Cosmic Evolution Survey (COSMOS): Overview*, *Astrophys. J. Suppl.* **172** (2007) 1 [[astro-ph/0612305](#)] [[INSPIRE](#)].
- [51] A.C. Robin, C. Reylé, S. Derrière and S. Picaud, *A synthetic view on structure and evolution of the Milky Way*, *Astron. Astrophys.* **409** (2003) 523 [[astro-ph/0401052](#)] [[INSPIRE](#)].
- [52] GAIA collaboration, *The Gaia Mission*, *Astron. Astrophys.* **595** (2016) A1 [[arXiv:1609.04153](#)] [[INSPIRE](#)].
- [53] A. Vallenari et al., *Gaia Data Release 3. Summary of the content and survey properties*, *Astron. Astrophys.* **674** (2023) A1 [[arXiv:2208.00211](#)].
- [54] SDSS collaboration, *New york university value-added galaxy catalog: A galaxy catalog based on new public surveys*, *Astron. J.* **129** (2005) 2562 [[astro-ph/0410166](#)] [[INSPIRE](#)].
- [55] R. Johnston, *Shedding Light on the Galaxy Luminosity Function*, *Astron. Astrophys. Rev.* **19** (2011) 41 [[arXiv:1106.2039](#)] [[INSPIRE](#)].
- [56] R.S. de Jong and C.G. Lacey, *The local space density of sb-sdm galaxies as function of their scale size, surface brightness, and luminosity*, *Astrophys. J.* **545** (2000) 781 [[astro-ph/0008071](#)] [[INSPIRE](#)].
- [57] S. Shen et al., *The size distribution of galaxies in the Sloan Digital Sky Survey*, *Mon. Not. Roy. Astron. Soc.* **343** (2003) 978 [[astro-ph/0301527](#)] [[INSPIRE](#)].
- [58] DES collaboration, *A catalogue of structural and morphological measurements for DES Y1*, *Mon. Not. Roy. Astron. Soc.* **481** (2018) 2018 [[arXiv:1807.10767](#)] [[INSPIRE](#)].
- [59] C. Laigle et al., *The COSMOS2015 Catalog: Exploring the $1 < z < 6$ Universe with half a million galaxies*, *Astrophys. J. Suppl.* **224** (2016) 24 [[arXiv:1604.02350](#)] [[INSPIRE](#)].
- [60] J. Herbel et al., *Fast point spread function modeling with deep learning*, *JCAP* **07** (2018) 054 [[arXiv:1801.07615](#)] [[INSPIRE](#)].
- [61] L. Bradley et al., *astropy/photutils: 1.5.0*, (2022), [[DOI:10.5281/zenodo.6825092](#)].

- [62] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf and A. Smola, *A kernel two-sample test*, *J. Mach. Learn. Res.* **13** (2012) 723.
- [63] SDSS collaboration, *Statistical Properties of Bright Galaxies in the Sloan Digital Sky Survey Photometric System*, *Astron. J.* **122** (2001) 1238 [[astro-ph/0105401](#)] [[INSPIRE](#)].
- [64] C. Yamauchi et al., *Morphological classification of galaxies using photometric parameters: The concentration index versus the coarseness parameter*, *Astron. J.* **130** (2005) 1545 [[astro-ph/0507060](#)] [[INSPIRE](#)].
- [65] M.R. Blanton and S. Roweis, *K-corrections and filter transformations in the ultraviolet, optical, and near infrared*, *Astron. J.* **133** (2007) 734 [[astro-ph/0606170](#)] [[INSPIRE](#)].
- [66] C.J. Conselice, *The Evolution of Galaxy Structure over Cosmic Time*, *Ann. Rev. Astron. Astrophys.* **52** (2014) 291 [[arXiv:1403.2783](#)] [[INSPIRE](#)].
- [67] LSST collaboration, *LSST: from Science Drivers to Reference Design and Anticipated Data Products*, *Astrophys. J.* **873** (2019) 111 [[arXiv:0805.2366](#)] [[INSPIRE](#)].
- [68] EUCLID collaboration, *Euclid Definition Study Report*, [arXiv:1110.3193](#) [[INSPIRE](#)].
- [69] EUCLID collaboration, *Euclid preparation. I. The Euclid Wide Survey*, *Astron. Astrophys.* **662** (2022) A112 [[arXiv:2108.01201](#)] [[INSPIRE](#)].
- [70] I.K. Baldry et al., *Quantifying the Bimodal Color-Magnitude Distribution of Galaxies*, *Astrophys. J.* **600** (2004) 681 [[astro-ph/0309710](#)] [[INSPIRE](#)].
- [71] J. Alsing et al., *SPECULATOR: Emulating Stellar Population Synthesis for Fast and Accurate Galaxy Spectra and Photometry*, *Astrophys. J. Suppl.* **249** (2020) 5.
- [72] A.P. Hearin et al., *DSPS: Differentiable Stellar Population Synthesis*, *Mon. Not. Roy. Astron. Soc.* **521** (2023) 1741 [[arXiv:2112.06830](#)] [[INSPIRE](#)].
- [73] K.J. Kwon, C.H. Hahn and J. Alsing, *Neural Stellar Population Synthesis Emulator for the DESI PROVABGS*, *Astrophys. J. Suppl.* **265** (2023) 23.
- [74] P. Berner et al., *Fast forward modelling of galaxy spatial and statistical distributions*, *JCAP* **04** (2024) 023 [[arXiv:2310.15223](#)] [[INSPIRE](#)].
- [75] D. Zürcher et al., *Cosmological Forecast for non-Gaussian Statistics in large-scale weak Lensing Surveys*, *JCAP* **01** (2021) 028 [[arXiv:2006.12506](#)] [[INSPIRE](#)].
- [76] D. Zürcher et al., *Towards a full w CDM map-based analysis for weak lensing surveys*, *Mon. Not. Roy. Astron. Soc.* **525** (2023) 761 [[arXiv:2206.01450](#)] [[INSPIRE](#)].
- [77] C.R. Harris et al., *Array programming with NumPy*, *Nature* **585** (2020) 357 [[arXiv:2006.10256](#)] [[INSPIRE](#)].
- [78] P. Virtanen et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, *Nature Meth.* **17** (2020) 261 [[arXiv:1907.10121](#)] [[INSPIRE](#)].
- [79] J.D. Hunter, *Matplotlib: A 2D Graphics Environment*, *Comput. Sci. Eng.* **9** (2007) 90 [[INSPIRE](#)].
- [80] T. Kacprzak and J. Fluri, *DeepLSS: Breaking Parameter Degeneracies in Large-Scale Structure with Deep-Learning Analysis of Combined Probes*, *Phys. Rev. X* **12** (2022) 031029 [[arXiv:2203.09616](#)] [[INSPIRE](#)].
- [81] Q. Wang, S.R. Kulkarni and S. Verdu, *Divergence Estimation for Multidimensional Densities Via k -Nearest-Neighbor Distances*, *IEEE Trans. Inform. Theory* **55** (2009) 2392.