The Astronomical Journal, 160:191 (16pp), 2020 October https://doi.org/10.3847/1538-3881/abb0e2

Learning Spectral Templates for Photometric Redshift Estimation from Broadband Photometry

John Franklin Crenshaw[1] and Andrew J. Connolly[2,3]
[1] Department of Physics, University of Washington, Box 351560, Seattle, WA 98195, USA; jfc20@uw.edu
[2] DIRAC Institute, Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA
[3] eScience Institute, University of Washington, Box 351570, Seattle, WA 98195, USA
Received 2020 July 2; revised 2020 August 10; accepted 2020 August 18; published 2020 September 30

Abstract

Estimating redshifts from broadband photometry is often limited by how accurately we can map the colors of galaxies to an underlying spectral template. Current techniques utilize spectrophotometric samples of galaxies or spectra derived from spectral synthesis models. Both of these approaches have their limitations: either the sample sizes are small and often not representative of the diversity of galaxy colors, or the model colors can be biased (often as a function of wavelength), which introduces systematics in the derived redshifts. In this paper, we learn the underlying spectral energy distributions from an ensemble of ∼100 K galaxies with measured redshifts and colors. We show that we are able to reconstruct emission and absorption lines at a significantly higher resolution than the broadband filters used to measure the photometry for a sample of 20 spectral templates. We find that our training algorithm reduces the fraction of outliers in the derived photometric redshifts by up to 28%, bias up to 91%, and scatter up to 25%, when compared to estimates using a standard set of spectral templates. We discuss the current limitations of this approach and its applicability for recovering the underlying properties of galaxies. Our derived templates and the code used to produce these results are publicly available in a dedicated Github repository: https://github.com/dirac-institute/photoz_template_learning.

Unified Astronomy Thesaurus concepts: Galaxy photometry (611); Photometry (1234); Astronomical techniques (1684); Spectral energy distribution (2129); Redshifted (1379); Cosmology (343); Redshift surveys (1378); Computational methods (1965); Astronomical methods (1043); Astronomy data analysis (1858)

1. Introduction

Studies of galaxy evolution, galaxy clusters, large-scale structure, weak lensing, and so on all rely on the determination of galaxy redshift. Spectroscopic surveys of galaxies can provide very accurate redshifts by measuring the shifted wavelengths of sharp spectral features such as emission and absorption lines. Despite advancements in multiobject spectrographs, spectroscopic measurements are expensive and time-consuming, and we can only collect spectra for a small fraction of the galaxies that can be imaged by modern surveys, such as the Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005) and the Kilo-Degree Survey (KiDS; de Jong et al. 2013). This problem will only increase in magnitude as the next generation of surveys, such as the Vera Rubin Observatory Legacy Survey of Space and Time (LSST; LSST Science Collaboration et al. 2009) and the Wide-Field Infrared Survey Telescope (WFIRST; Green et al. 2012), image orders of magnitude more galaxies at fainter magnitudes than are present in current data sets. As a result, rather than rely on spectroscopic redshifts (spec-z's), modern surveys increasingly rely on photometric redshifts (photo-z's; see Salvato et al. 2019 for a review).

Photo-z's are estimates of galaxy redshifts derived from changes in the colors of galaxies as their spectral energy distributions (SEDs) redshift through a series of broadband filters. This estimation is typically done using one of two approaches: machine learning (ML) or template fitting (see, e.g., Schmidt et al. 2020 for an evaluation of many examples of the two).

Machine learning approaches train on a data set of photometry with spec-z's and attempt to directly learn an empirical relationship between galaxy colors and redshift (e.g., Connolly et al. 1995, `TPZ` Kind & Brunner 2013, `FlexZ-Boost` Izbicki & Lee 2017, `CMNN` Graham et al. 2018). Once trained, they can predict galaxy redshifts given photometry alone. The advantage of ML methods is that the effects of dust, galaxy evolution, and other relevant variables are encoded in the training set, and thus it is possible for ML methods to account for these in the derived mapping from colors to redshift if the data encapsulate these effects. The success of this mapping depends on the choice and complexity of the ML model and the corresponding hyperparameters. The downside of ML methods is that their success relies on how representative and well controlled the training set is, and that they are unable to extrapolate beyond that set.

Template fitting of photo-z estimators (e.g., `LePhare`, Arnouts et al. 1999; `BPZ`, Benitez 2000; `EAZY`, Brammer et al. 2008) works on the assumption that galaxy photometry is sampled from a relatively small set of underlying spectral types, characterized by the eponymous SED templates. These estimators calculate photo-z's by selecting the template and redshift with simulated fluxes most similar to the observed fluxes. In order for this method to work, the underlying SED templates from which the galaxies are sampled must be known. Common methods for generating these templates include simulating galaxy SEDs from spectral synthesis models (e.g., Bruzual & Charlot 1993) and deriving templates from the observed spectra of local galaxies (e.g., Benitez et al. 2004).

The primary advantage of the template fitting method is that it is not limited to the bounds of a training set. A key limitation is that it does not guarantee that the SED templates will span the full distribution of galaxy spectra in a given data set, nor that it will properly account for the effects of dust or spectral

evolution. In addition, spectral synthesis models are only able to generate spectra with a discrete set of physical parameters (e.g., temperature, age, metallicity), and obtaining real galaxy spectra is expensive, especially at the redshifts and magnitudes that will be observed by LSST.

Several previous works have attempted to combine the advantages of these two approaches by deriving SED templates from a photometric training set, and then using the derived templates for photo-z estimation (Budavári et al. 2000; Csabai et al. 2000; Assef et al. 2008). These approaches leverage a large set of galaxy photometry, which amount to low-resolution spectra, to sample a smaller set of SED templates across a broad range of rest wavelengths. This effectively oversamples the template SEDs, allowing us to reconstruct spectral features at a resolution much higher than that of the broadband filters used to measure the photometry. This is analogous to the Drizzle technique used to reconstruct higher resolution images for the Hubble Space Telescope (HST; Fruchter & Hook 2002) and the reconstruction of SEDs using differential chromatic refraction (DCR; Lee et al. 2019).

This template learning approach retains the physical motivation and extensibility of the template fitting method, while taking advantage of learning the systematics and confounding variables implicit in the training set. In addition, it opens the possibility of learning a smooth continuum of galaxy spectra, in contrast to the discrete set offered by the limited galaxy observations and galaxy modeling codes.

While previous works attempt to learn galaxy templates from data using a set of eigenspectra, we adapt the algorithm of Budavári et al. (2000) to directly learn a set of templates from the data. We extend these earlier works by applying our methods to a large data set of 102,476 galaxies with spec-z's and photometry in 19 bands. In this manner, we are able to learn a variable number of SED templates with clear spectral features, and with simple postprocessing, we are able to further reconstruct emission lines in the bluest templates.

We show that templates can be learned from scratch or as perturbations of preexisting templates. We use these learned templates to estimate photo-z's with `BPZ` and find that the training reduces the bias and scatter of the redshift estimates, with little impact on the fraction of catastrophic outliers. In addition, we find that both bias and scatter decrease with the number of SED templates used in the photo-z estimation.

The outline of the paper is as follows. In Section 2 we describe the template training algorithm, including how to match photometry to templates, how to perturb templates to better match the photometry, and how to select the hyperparameters for training. In Section 3, we describe the spec-z and photometric data sets used in the template training and redshift estimation. In Section 4, we apply the template training algorithm to sets of naive templates and to a preexisting set of templates derived from galaxy observations and spectral synthesis models. We discuss the performance of the algorithm, including its convergence and the accuracy of the reconstructions. In Section 5, we use our templates to estimate photo-z's for a training set of galaxies and analyze the results. We discuss our results and future goals in Section 6 and conclude in Section 7.

2. Template Training Algorithm

In this section, we will present an approach for learning SED templates directly from broadband photometry, using a modified version of the algorithm developed in Budavári et al. (2000). If we assume that the galaxies in our data set are sampled from a small set of underlying spectra, the SED templates, and we know the spectroscopic redshift for each galaxy, we can shift the photometry to the rest frame and treat each observation of a redshifted galaxy as a *rest-frame observation* of one of the templates with a different set of effective filters. With a large enough data set, the wavelengths of the effective filters will overlap substantially. This oversampling allows us to recover higher resolution features in the templates, even though the data are low-resolution observations of different galaxies.

Let us assume we have a set of SED templates as a starting point, which can represent rudimentary guesses and need not resemble true galaxy spectra. In the first part of this section, we describe a method by which we create a training set of broadband photometry for each template from a large data set of galaxy photometry. In the second part, we derive the perturbation algorithm that is used to train each SED template on its corresponding photometry set. The full training algorithm is an expectation maximization that consists of iterating these two steps: matching photometry to templates, and perturbing templates to better match the photometry. This process is iterated until the SED templates converge. In the final part, we discuss a heuristic for selecting the training hyperparameters.

2.1. Matching Photometry Sets

Assume we have a set of naive SED templates and a large set of observed fluxes, $\{f_m\}$, with known spectroscopic redshifts, z_m. Our goal is to train each template on an appropriate subset of the $\{f_m\}$ so that the naive templates better represent the colors of the galaxies. To assemble these training sets, we consider subsets $\{f_n\} \subset \{f_m\}$, corresponding to the observed fluxes of a single galaxy at redshift z, where the subscript n denotes different filters. We compare these observed fluxes with the template fluxes $\{\hat{f}_n\}$, where

$$\hat{f}_n = \int S\left(\frac{\lambda}{1+z}\right) R^n(\lambda) d\lambda, \tag{1}$$

$S(\lambda)$ is an SED template, and $R^n(\lambda)$ is the normalized response function of the filter used to measure the flux f_n. For photon counting detectors,

$$R(\lambda) = \frac{\lambda\, T(\lambda)}{\int \lambda\, T(\lambda) d\lambda}, \tag{2}$$

where $T(\lambda)$ is the system response function that captures the transmittance of the atmosphere and the response of the detector (Bessell 2005).

The observed fluxes are assigned to the template whose colors are most similar, which is determined by normalizing the observed and template fluxes in the same band and picking the template that minimizes the squared differences of the fluxes. The normalization band is chosen by selecting the band for which the ratio $\hat{f}_n / f_n$ is the median of the flux ratios for that galaxy. By performing this matching and renormalization for each galaxy in the photometry set, we associate a subset of the galaxies (and the corresponding photometry) to each template.

Examining how the galaxies are assigned to the individual templates is helpful in selecting the initial set of templates. The initial templates should be chosen so that the matching

algorithm roughly divides the galaxies by their colors. It is also important that each set contains a sufficient number of fluxes distributed across the wavelengths of interest, as the perturbation algorithm derived in the next section relies on oversampling to reconstruct higher resolution features of the SED templates.

2.2. The Perturbation Algorithm

Assume we have a set of photometry, $\{f_n\}$, which constitute observations of the same underlying SED template, $S(\lambda)$, at various known redshifts, z_n. These observed fluxes should approximately match the template fluxes calculated via Equation (1). However, we can also calculate the template fluxes by imagining that we are observing the template in its rest frame using a set of effective, blueshifted filters:

$$\hat{f}_n = \int S(\lambda)\, R^n[(1+z_n)\lambda]\, d[(1+z_n)\lambda] \tag{3}$$

$$= \sum_k s_k\, r^n_{k'} \Delta\lambda_{k'}, \tag{4}$$

where in the second line s_k and r^n_k are the discrete representations of $S(\lambda)$ and $R^n(\lambda)$, respectively, parameterized by the wavelength bins $\{\lambda_k\}$ with widths $\{\Delta\lambda_k\}$. Primed indices indicate redshifted wavelengths, that is, $\lambda_{k'} = (1+z_n)\lambda_k$ and $\Delta\lambda_{k'} = (1+z_n)\Delta\lambda_k$.

We wish to perturb the template so that the template fluxes, $\hat{f}_n$, better match the observed fluxes, f_n. Letting $\hat{s}_k$ be a new template resulting from a perturbation of s_k, we define the cost function (Budavári et al. 2000, Equation (7)):

$$\chi^2 = \sum_n \frac{1}{\sigma_n^2}(\hat{f}_n(\{\hat{s}_k\}) - f_n)^2 + \sum_k \frac{1}{\Delta_k^2}(\hat{s}_k - s_k)^2. \tag{5}$$

The optimum perturbation is found via a multidimensional minimization of the cost function. The first term in Equation (5) penalizes differences between the observed fluxes and the perturbed template fluxes, weighted according to σ_n (the fractional error of the measured flux). The second term in Equation (5) penalizes large perturbations, weighted by the hyperparameters Δ_k. This parameter controls the learning rate and also helps stabilize the results. See the next section for more details.

We follow Budavári et al. (2000) by introducing the simplifying perturbation and constant terms:

$$\begin{aligned} \xi_k &= \hat{s}_k - s_k \\ g_n &= f_n - \sum_k s_k\, r^n_{k'} \Delta\lambda_{k'}. \end{aligned} \tag{6}$$

Then, we have

$$\chi^2 = \sum_n \frac{1}{\sigma_n^2}\left(g_n - \sum_k \xi_k\, r^n_{k'} \Delta\lambda_{k'}\right)^2 + \sum_k \frac{\xi_k^2}{\Delta_k^2}, \tag{7}$$

which can be analytically minimized:

$$\frac{\partial\chi^2}{\partial\xi_l} = 0 \Rightarrow \sum_k M_{lk}\tilde{\xi}_k = \nu_l. \tag{8}$$

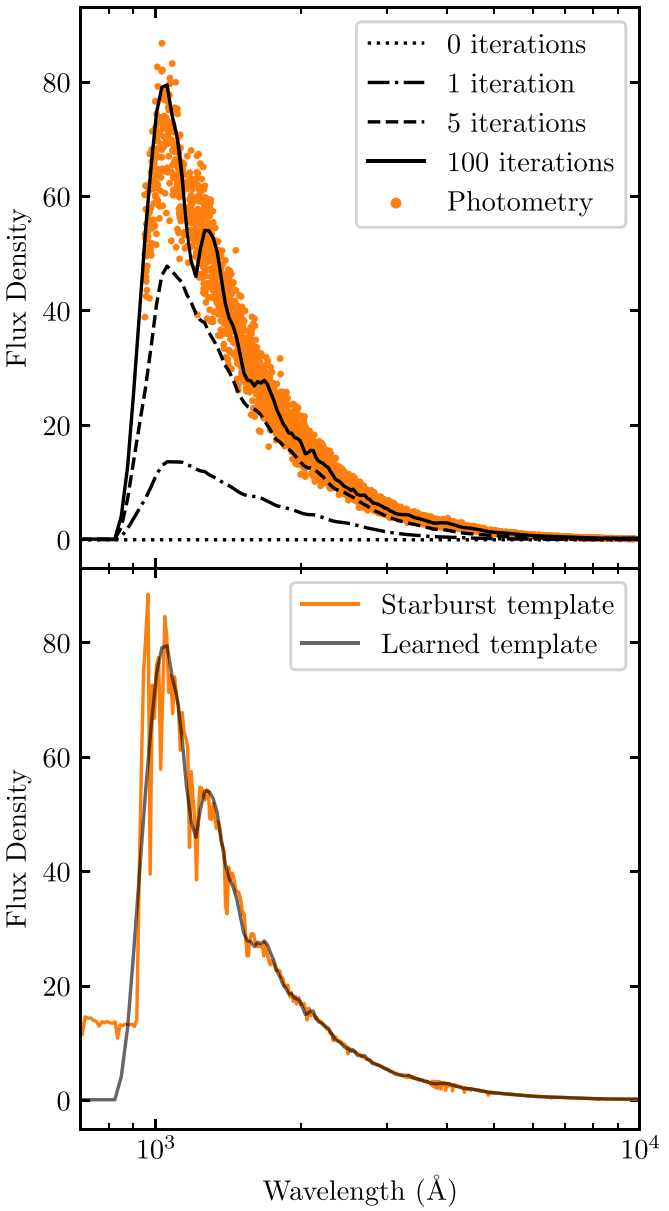

Figure 1. Perturbing a naive template, in this case a flat line, to better match a photometry set. Top: The orange points are simulated observations of the 5 Myr starburst template from Coe et al. (2006) at 1000 random redshifts in the range $z = 0$ to $z = 3$ using the *ugrizY* filters listed in Table 2. The simulated photometry has a 10% Gaussian error. The template is shown after various stages of the training. Bottom: The learned template is plotted with the original starburst template.

The matrix M and vector ν are defined as

$$\begin{aligned} M_{lk} &= \sum_n \frac{1}{\sigma_n^2}(r^n_{l'}\Delta\lambda_{l'})(r^n_{k'}\Delta\lambda_{k'}) + \frac{\delta_{lk}}{\Delta_k^2}, \\ \nu_l &= \sum_n \frac{g_n}{\sigma_n^2}(r^n_{l'}\Delta\lambda_{l'}), \end{aligned} \tag{9}$$

where δ_{lk} is the Kronecker delta. One can then solve for $\tilde{\xi}$. The perturbed spectrum is then $\hat{s}_k = s_k + \tilde{\xi}_k$.

Iterating the perturbation changes the shape of the template SED to better match the measured photometry, as shown in Budavári et al. (2000). An example of this process can be seen in Figure 1. Fluxes in the *ugrizY* filters listed in Table 2 were calculated for a starburst galaxy template at 1000 random

Table 1
Summary of the Spectrophotometric Data Sets

Data Set	N_{gal}	f_{gal}	z_{mean}	z_{max}	i-band Range	i_{mean}	$\bar{\sigma}_i$	Link to Catalog
zCOSMOS	14298	0.14	0.57	2.52	$16.87 \leqslant i \leqslant 24.18$	21.19	0.022	http://cesam.lam.fr/hstcosmos/
VVDS	6915	0.07	0.67	4.54	$13.84 \leqslant i \leqslant 24.97$	20.86	0.014	https://cesam.lam.fr/vvds/index.php
VIPERS	69415	0.68	0.70	2.15	$17.66 \leqslant i \leqslant 23.08$	21.38	0.017	http://vipers.inaf.it:8080/
DEEP2/3	10695	0.10	0.71	1.91	$15.30 \leqslant i \leqslant 25.36$	21.42	0.020	http://d-scholarship.pitt.edu/36064/
3D-HST	1153	0.01	1.46	3.32	$19.10 \leqslant i \leqslant 25.74$	23.56	0.027	http://d-scholarship.pitt.edu/36064/
Training	81980	0.80	0.69	4.54	$13.84 \leqslant i \leqslant 25.74$	21.32	0.018	
Test	20496	0.20	0.69	3.61	$16.46 \leqslant i \leqslant 25.69$	21.34	0.018	
Total	102476	1.00	0.69	4.54	$13.84 \leqslant i \leqslant 25.74$	21.33	0.018	

Note. N_{gal} is the total number of galaxies in the set, f_{gal} is the fraction of all galaxies in the set, and $\bar{\sigma}_i$ is the mean fractional flux error in the i-band.

redshifts $z < 3$. Starting with an $S(\lambda) = 0$ template SED, the perturbation algorithm is applied iteratively. After 100 iterations, the trained template closely matches the original template in the wavelength range for which photometry exists. While the trained template is a smoothed version of the original, high-resolution features have been recovered, despite the relatively low resolution of the filters. In practice, a higher Δ_k can be chosen so that fewer iterations are required in the training; a lower value was chosen here so that the effects of successive iterations can be more clearly seen. See Section 2.3 for further discussion of selecting the hyperparameters.

The perturbation algorithm changes the shape of the template SEDs so that rerunning the photometry matching will now result in different subsets of galaxies assigned to each template. The full training algorithm is iterated until the SED templates converge.

2.3. Selecting Hyperparameters

The success of the training algorithm depends on the chosen hyperparameters. The first is the number of templates. As discussed in Section 2.1, this choice can be made by using the photometry matching algorithm and choosing the appropriate number of templates to approximately separate out the different spectral shapes displayed in the photometry. For further discussion of how the number of templates affects photo-z results, see Section 5.3.

The rest of the hyperparameters consist of the set of Δ_k. These parameters, which set the relative weighting of the regularization term in Equation (5), determine the stability and speed of the training algorithm. If the Δ_k are too large, training will be very slow and a large number of iterations will be required. If the Δ_k are too low, the training becomes unstable and the final templates will be overfit. Here we present a heuristic for selecting an appropriate value to balance these two extremes.

For the work presented below, the index k is dropped, so $\Delta \equiv \Delta_k$ has a single value for each training set that is independent of wavelength. In choosing the appropriate value of Δ for each training set, it is desirable to select a value that corresponds to a constant ratio, w, of the flux and regularization terms in Equation (5). The necessary value of Δ will vary by training set, as the number of terms in the sum over fluxes (i.e., the sum over n in Equation (5)) will vary by training set. To this end, we make the following approximation:

$$\frac{\sum_k (\hat{s}_k - s_k)^2}{\sum_n (\hat{f}_n - f_n)^2} \sim \frac{N_k}{N_n}, \tag{10}$$

where $N_k \equiv \sum_k$ and $N_n \equiv \sum_n$. This permits the following approximation of the ratio w:

$$w = \frac{\sum_k \Delta^{-2} (\hat{s}_k - s_k)^2}{\sum_n \sigma_n^{-2} (\hat{f}_n - f_n)^2} \sim \frac{N_k/\Delta^2}{N_n/\bar{\sigma}^2}, \tag{11}$$

where $\bar{\sigma} = \sum_n \sigma_n / N_n$. Then, for a desired ratio w, the requisite Δ can be approximated:

$$\Delta \simeq \bar{\sigma} \sqrt{\frac{N_k}{w N_n}}. \tag{12}$$

In practice, we have found that $w = \mathcal{O}(1)$ works well. The results of the training are relatively robust to the selection of w, in that changing w by, for example, a factor of two yields similar results.

3. Data

We collect a set of galaxy spectroscopic redshifts, paired with broadband photometry, from various surveys to test our training algorithm. Our set consists of 102,476 galaxies with redshifts $z < 4.54$ and i-band magnitudes[4] in the range $13.8 < i < 25.7$. For all surveys, we use galaxies with highly reliable spec-z's, photometry in one of the i-bands, and photometry in at least three bands with signal-to-noise ratio S/N > 20. The entire data set is summarized in Table 1, the filters used to measure the photometry are listed in Table 2, and the redshift distributions are shown in Figure 2.

3.1. zCOSMOS-bright

zCOSMOS (Lilly et al. 2009) is a redshift survey of 1.7 deg^2 of the COSMOS field, conducted with the VIMOS spectrograph mounted on the European Southern Observatory's (ESO) Very Large Telescope (VLT). The survey is divided into two parts, *bright* and *deep*. We make use of the former, consisting of approximately 20,000 galaxies with redshifts $z < 1.2$. We use galaxies recommended in the ESO data release

[4] The i-band magnitudes quoted in this section denote the magnitude in one of i, i_2, I, or i^+ as listed in Table 2. For galaxies with photometry in multiple i-bands, the magnitude used is the first to appear in that list.

Table 2
List of Broadband Filters

Filter	Telescope	Instrument	λ_0	$W_{\rm eff}$
NUV	GALEX		2343.1	767.3
u	CFHT	Megacam	3817.7	525.4
B	CFHT	CFH12k	4342.5	873.6
B_J	Subaru	Suprime	4478.4	763.9
g^+	Subaru	Suprime	4808.5	1043.1
g	CHFT	Megacam	4899.9	1293.8
V	CFHT	CFH12k	5393.7	882.7
V_J	Subaru	Suprime	5493.0	862.4
r	CHFT	Megacam	6278.2	1120.2
r^+	Subaru	Suprime	6314.8	1211.4
R	CFHT	CFH12k	6603.5	1138.5
i_2	CHFT	Megacam	7584.5	1409.4
i	CHFT	Megacam	7676.6	1307.6
i^+	Subaru	Suprime	7709.1	1361.7
I	CFHT	CFH12k	8277.3	1816.7
z	CHFT	Megacam	8857.6	1040.1
z^+	Subaru	Suprime	9054.5	1012.3
Y	Subaru	Suprime	10216.0	996.2
J	UKIRT	WFCAM	12508.5	1476.8

Notes. Mean wavelength, $\lambda_0 = \int \lambda R(\lambda) d\lambda$, and effective width, $W_{\rm eff} = {\rm Max}[R(\lambda)]^{-1}$, are given in angstroms. Filters are listed in order of increasing λ_0. The response functions for each filter were obtained from the Spanish Virtual Observatory (SVO) Filter Profile Service.

description[5] determined to have 99% spectroscopic verification (i.e., `zflag` = 3.x, 4.x, 2.5, 2.4, 1.5, 9.5, 9.3, 18.5, 18.3).

The zCOSMOS redshifts are matched to photometry from Ilbert et al. (2009). The photometry is measured from the ultraviolet to the near-infrared in 10 broadband filters: *NUV* on GALEX (Martin et al. 2005); *u* and *i* on CFHT-Megacam; *B* and *V* on CFHT-CFH12k; g^+, r^+, i^+, and z^+ on Subaru; and *J* on UKIRT. The final set consists of 14,298 galaxies with redshifts $z < 2.52$ and *i*-band magnitudes in the range $16.9 < i < 24.2$.

3.2. VVDS

The VIMOS VLT Deep Survey (VVDS; Le Fèvre et al. 2013) is a redshift survey consisting of three component surveys: *Wide*, *Deep*, and *Ultra-Deep*. The Wide survey covers 8.7 deg^2, with approximately 25,000 galaxies in the range $17.5 < i < 22.5$; the Deep survey covers 0.74 deg^2, with approximately 11,000 galaxies in the range $17.5 < i < 24$; and the Ultra-Deep survey covers 512 arcmin^2, with approximately 900 galaxies in the range $23 < i < 24.75$. We use redshifts with quality flags 3 and 4, indicating a 98% spec-z confidence. The photometry was measured in nine filters: *u*, *g*, *r*, *i*, *z* on CFHT-Megacam (Hudelot et al. 2012) and *B*, *V*, *R*, *I* on CFHT-CFH12k (Le Fèvre et al. 2004). The final set contains 6,915 galaxies out to redshifts $z < 4.5$, with magnitudes $13.8 < i < 25.0$.

3.3. VIPERS

The VIMOS Public Extragalactic Redshift Survey (VIPERS; Scodeggio et al. 2018) is a dense, large-volume redshift survey focusing on redshifts $0.5 < z < 1.2$. We use VIPERS galaxies

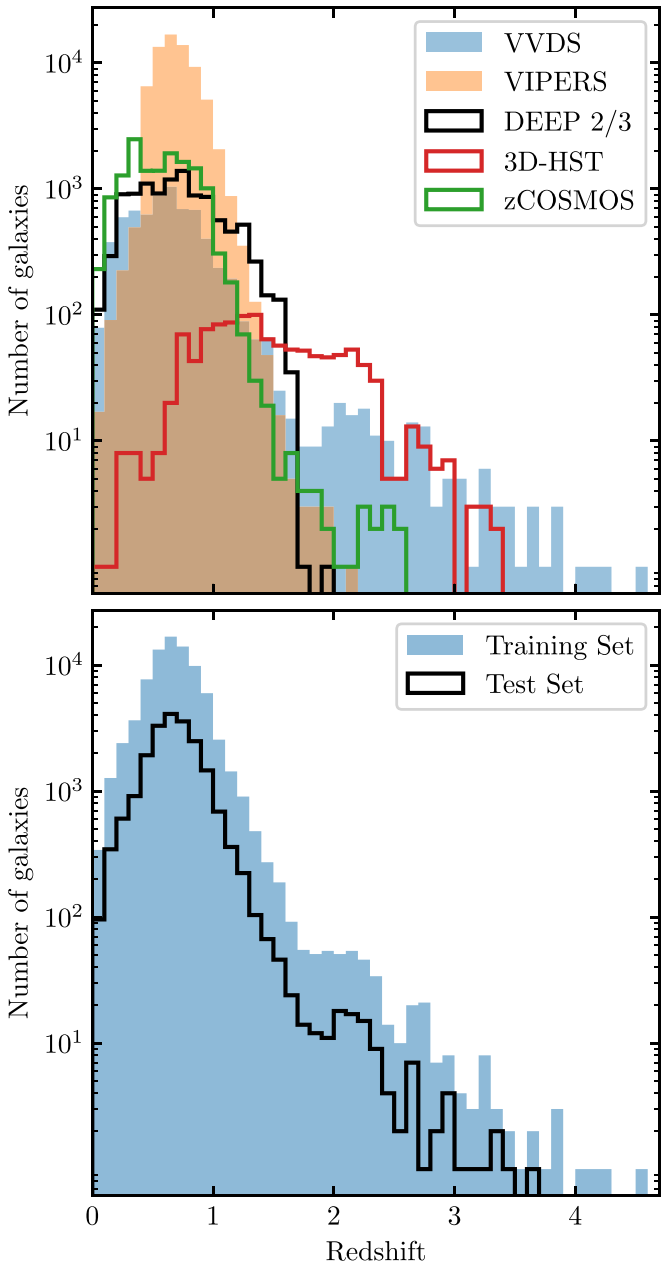

Figure 2. Redshift distribution of the galaxy surveys. The top panel shows the distributions of each of the constituent surveys. The bottom panel shows the redshift distributions of the training and test sets used for template training and photo-z estimation, respectively.

with spec-z's reliable at the 95% confidence level (`zflag` = 2.X, 3.X, 4.X) and with `photoMask` and `spectroMask` = 1. The redshifts are matched to photometry measured in *NUV* on GALEX (Martin et al. 2005) and *u*, *g*, *r*, i_2, *i*, *z* on CFHT-Megacam[6] (Hudelot et al. 2012). The final set contains 71,951 galaxies with redshifts $z < 2.15$ and magnitudes $17.7 < i < 23.3$.

[5] https://www.eso.org/sci/observing/phase3/data_releases/zcosmos_dr3_b2.pdf

[6] The i_2 band is the replacement to the Megacam *i*-band installed in 2007. This filter is named *y* in the CFHTLS catalogs (Hudelot et al. 2012), but we follow Zhou et al. (2019) in naming it i_2 to avoid confusion with the longer *y* bands used in Subaru and LSST.

3.4. DEEP2 and DEEP3

DEEP2 and DEEP3 are redshift surveys conducted with the DEIMOS spectrograph on the Keck II telescope. DEEP2 (Newman et al. 2013) consists of four fields; we use galaxies from the first field in the Extended Groth Strip (EGS), which had no redshift preselection. DEEP3 (Cooper et al. 2011) expanded on the DEEP2 survey of the EGS. Redshifts from these surveys are matched with aperture-corrected photometry provided by Zhou et al. (2019). We use galaxies with CFHTLS flag 0, SExtractor flags less than 4 in every band, and redshift quality flag $\geqslant 3$. Photometry was measured in u, g, r, i_2, i, z on CFHT-Megacam[6] and Y on Subaru (Miyazaki et al. 2002). The final set contains 10,695 galaxies with redshifts $z < 1.91$ and magnitudes $15.3 < i < 25.74$.

3.5. 3D-HST

In addition to the spectroscopic surveys above, we include grism redshifts from the 3D-HST survey (Newman et al. 2013; Momcheva et al. 2016). Redshifts for this survey were analyzed and matched with aperture-corrected photometry by Zhou et al. (2019). We select the galaxies with CFHTLS flag 0, SExtractor flags less than 4 in every band, and the flag `use_zgrism1` $= 1$. For galaxies in both the DEEP2/3 and 3D-HST sets, we use DEEP2/3 redshifts instead. Photometry was measured in u, g, r, i_2, i, z on CFHT-Megacam and Y on Subaru. After these cuts, the 3D-HST set contains 1153 galaxies with redshifts $z < 3.32$ and magnitudes $23.6 < i < 25.7$.

4. Application to Data

Using the training algorithm described in Section 2, we will learn galaxy SED templates directly from the broadband photometry described in Section 3. We divide the data set into a training and a test set, consisting of random 80% and 20% samples of the entire data set. The training set will be used to train the SED templates, while the test set will be used to evaluate the learned templates via photo-z estimation (see Section 5). The training set consists of 81,980 galaxies with mean redshift $z_{\rm mean} = 0.69$, max redshift $z_{\rm max} = 4.54$, and magnitudes $13.8 < i < 25.7$. A full summary of the set can be seen in Table 1, and the redshift distribution can be seen in Figure 2.

Eight naive templates were chosen to represent the underlying SED shapes of the photometry set according to the principles described at the end of Section 2.1. We chose the number eight to allow a direct comparison to the standard template set described below. They are "naive" because they are simply chosen by eye to roughly divide the photometry into groups by spectral shape, but otherwise they are not based on any theoretical models or observed SEDs. Each of the naive templates is a log-normal function,

$$S(\lambda) \propto \frac{1}{\lambda} \exp\left[-\frac{1}{2\eta^2}\left(\ln\frac{\lambda}{\mathrm{mode}(\lambda)} - \eta^2\right)^2\right], \tag{13}$$

normalized at $\lambda = 5000$ Å, with mode (λ) in the range 1000–5500 Å and η in the range 0.35–0.9. The templates extend to 15000 Å with 100 Å resolution. These eight templates (hereafter N8) can be seen together with their original training sets in Figure 3.

The training algorithm with $w = 0.5$ is applied to the N8 templates. The convergence of the templates is evaluated via the weighted mean square error:

$$\mathrm{wMSE} = \sum_n \frac{1}{\sigma_n^2}(\hat{f}_n(\{\hat{s}_k\}) - f_n)^2. \tag{14}$$

Each template is perturbed until the change in wMSE is less than 3%, which was chosen empirically to balance sufficient template reconstruction and the algorithm's run time. When every template has converged to its current photometry set, new photometry sets are generated. Only those templates whose new photometry sets result in a greater than 3% change in wMSE resume perturbation with their new sets. This process is iterated until no template has a new photometry set that results in a greater than 3% change in wMSE. This indicates that the photometry is sorted into distinct sets, and that further perturbation is unlikely to improve the photometry-matching results.

The progress of the training algorithm is shown in Figure 4 for the template N8-1. The left panel shows the progress of the perturbation algorithm as it deforms the originally smooth N8-1 template to better match the colors of the matched photometry sets. In particular, N8-1 becomes redder and acquires higher resolution structure, which will be discussed below. The middle panel shows the wMSE, and the right panel shows the fractional change in the wMSE throughout the training. Orange points indicate values after a photometry-matching stage, and blue points indicate values after a perturbation. You can see that the wMSE drops as the template is perturbed, and perturbation continues until the magnitude of the fractional change in wMSE drops below 0.03, indicated by the dotted black lines in the right panel. Once this occurs, new photometry is matched, resulting in an increase in wMSE. This process is iterated, with fewer and fewer perturbations needed per iteration. Eventually, all of the points are orange, indicating that after each new photometry matching, N8-1 is not perturbed, as it already sufficiently matches its photometry set.

The training continues for 12 rounds and takes approximately 15 minutes. The final results for the N8 templates can be seen in Figure 5. The templates are now a much better match to the photometry and more closely resemble physical galaxy spectra. Most of the templates have a Balmer break at 4000 Å, although this was essentially already present in the initial templates. In addition, there are now emission and absorption lines visible in the spectra at a much higher resolution than the broadband filters used for the photometry (some of which are labeled with gray lines in Figure 5). Template N8-1 displays Mg and Na absorption lines, and template N8-4 contains the beginnings of Hα and Hβ emission lines. Templates N8-6, N8-7, and N8-8 contain what appear to be Hα, Hβ, Hγ, Hδ, O II, and O III emission lines (see Section 4.1 for more analysis). The emergence of these high-resolution features from a large ensemble of low-resolution data is one of the defining features of this method.

In addition to these eight templates, we double the template number and train a set of 16 templates, in order to demonstrate the algorithm's ability to reconstruct templates with a more gradual transition of the colors from red to blue. This set (hereafter N16) was drawn from the same range of parameters for the log-normal function and trained for 50 minutes over 26 rounds. The results of the training can be seen in Figure 6.

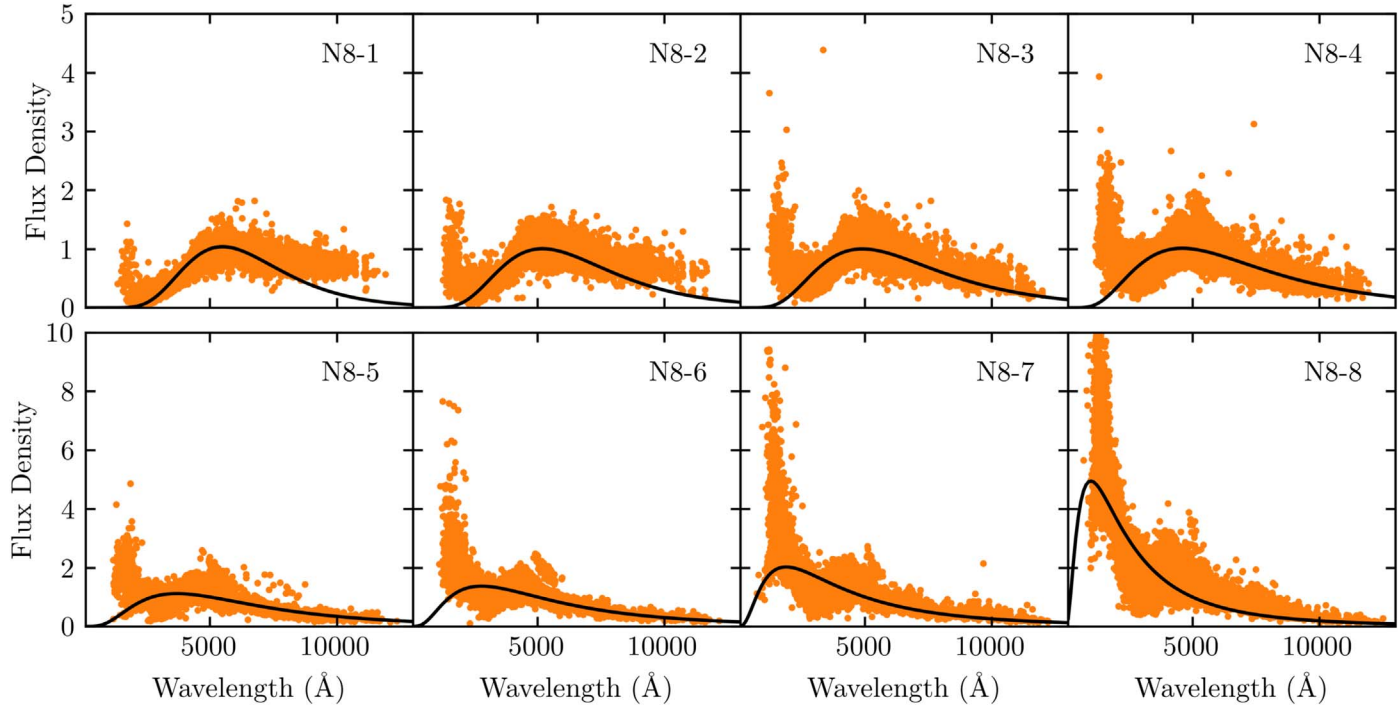

Figure 3. Untrained N8 templates (black lines) with their corresponding photometry sets (orange points), generated with the algorithm described in Section 2.1. N8-1 is the reddest template, with each successive template getting bluer.

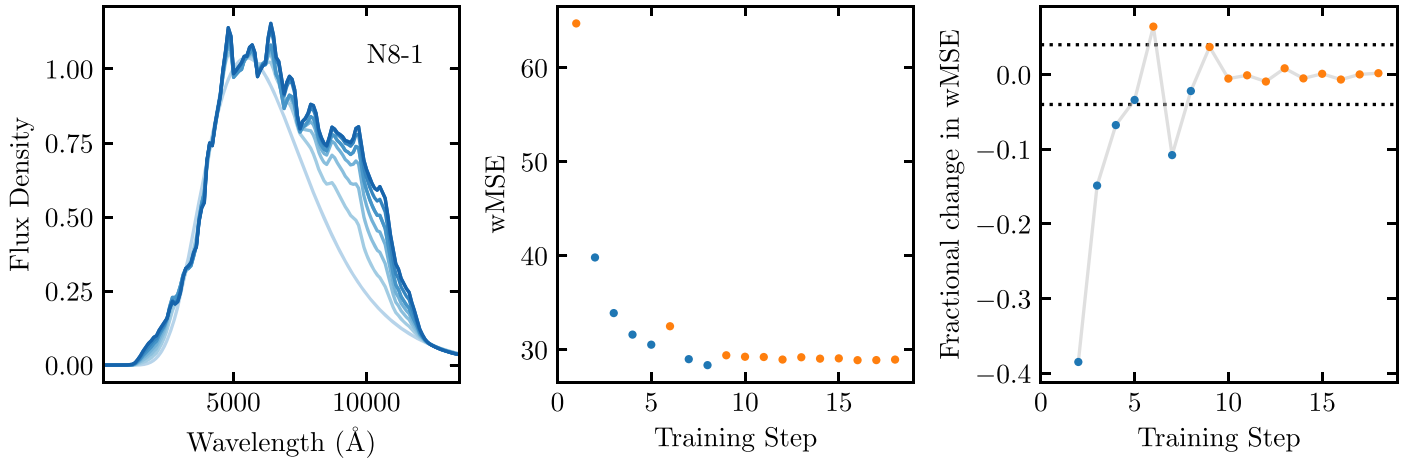

Figure 4. Training of N8-1. Left: The initial (light blue) N8-1 template is iteratively perturbed to better represent the colors of its photometry set. The final (dark blue) template is redder and has more structure. Middle: wMSE of the N8-1 template throughout the training process. Orange points represent the wMSE after a photometry matching stage, while blue points represent the wMSE after a perturbation. Right: fractional change in the wMSE. Orange points represent the fractional change due to a new photometry matching stage, while blue points represent a fractional change due to a perturbation. The dotted black lines show the ±0.03 cutoff. When a perturbation results in a fractional change of magnitude less than 0.03, perturbation is halted and new photometry is matched. After the sixth photometry match, the template is not perturbed because it already sufficiently matches the photometry.

These results closely resemble the N8 results, with the same spectral features emerging. However, the N16 set shows a more gradual transition in color.

In addition to starting from naive templates, one can start with templates derived from spectral synthesis models or observations of local galaxy spectra (Budavári et al. 2000; Csabai et al. 2000). Here we apply the training algorithm to a standard set of SED templates commonly used for photo-z estimation (e.g., BPZ; see Section 5.1). This set (hereafter CWW+SB4) consists of four templates from Coleman et al. (1980) and two starburst templates from Kinney et al. (1996), the latter of which were added to account for faint blue galaxies in the HDF-N. These six templates were recalibrated by Benitez et al. (2004) to correct for systematic differences between the observed and predicted galaxy colors in the HDF-N and other spectroscopic catalogs. In addition to these six, CWW+SB4 contains two synthetic starburst templates from Bruzual & Charlot (2003), added by Coe et al. (2006) to account for even bluer galaxies in the UDF.

The CWW+SB4 templates were trained with $w = 2$ for 46 minutes over 32 iterations. The results of the training can be seen in Figure 7. The original templates are plotted in blue, with the trained templates plotted in black, along with the final photometry sets in orange. You can see that the El and Sbc templates have barely been altered. The remaining templates have all systematically become redder. The high-resolution

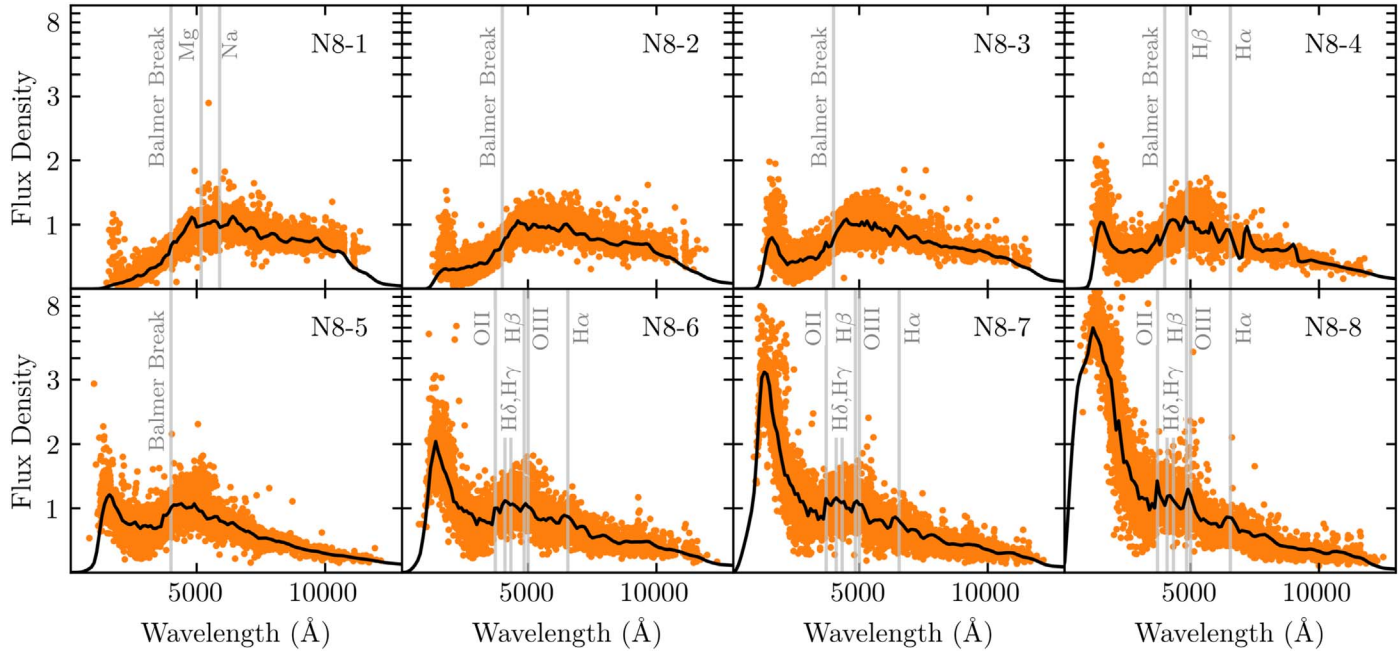

Figure 5. Trained N8 templates (black lines) with their final photometry sets (orange points). N8-1 is the reddest template, with each successive template getting bluer. The templates now more closely resemble physical galaxy spectra and have acquired structure at a higher resolution than the broadband templates. The Balmer break, Mg and Na absorption lines, and Hα, Hβ, Hγ, Hδ, O II, and O III emission lines are labeled in gray.

structure that was originally present in the Im, SB3, and SB2 templates has been decreased in magnitude, while additional structure has been added to the simulated 25 and 5 Myr templates, which were originally smooth. These new features have been labeled in gray.

4.1. Reconstructing Spectral Lines

The template training algorithm allows the reconstruction of high-resolution spectral features from low-resolution photometry, due to the oversampling of the underlying SED templates. This includes the emergence of spectral lines in many of the templates (see Figures 5, 6, and 7). Knowledge of these lines allows us to perform postprocessing of the learned templates to deconvolve the lines from the broadband filters. Here we perform a simple postprocessing of the N8-6, N8-7, and N8-8 templates to reconstruct the emission lines labeled in Figure 5. The templates are up-sampled to 10 Å, and the continuum of each is linearly interpolated around the emission lines. The excess flux is attributed to the corresponding spectral lines. The flux of the Hβ line is impossible to distinguish from the O III line in our templates because they are so close to one another. The same is true for the Hγ and Hδ lines. To overcome this difficulty, we use the Balmer decrements of 10^4 K SDSS galaxies from Groves et al. (2012): Hα/Hβ = 2.86 and Hγ/Hδ = 1.81. We calculate the Hβ flux from Hα and subtract this from the combined Hβ–O III flux, and we calculate Hγ and Hδ from the combined Hγ–Hδ flux.

After calculating the flux of the emission lines, the final templates are built by adding Gaussians of equivalent amplitude and FWHM = 20 Å to the continuum. The templates with the reconstructed spectral lines can be seen in Figure 8. For each line, we calculate the amplitude relative to Hβ and the effective width, $W_\lambda = \int (1 - F_\lambda/F_0) d\lambda$, where F_λ is the total flux and F_0 is the continuum flux. These values can be seen in Table 3. Note that the amplitudes of our reconstructed Hγ and Hδ lines relative to Hβ are approximately three times greater than those listed in Groves et al. (2012).

5. Estimating Photo-z's

We evaluate the results of our template training algorithm by using our learned templates to estimate photo-z's for the test set of galaxies using the software package BPZ (Benitez 2000), and by comparing the results to the spec-z's and the photo-z's estimated using the original CWW+SB4 templates. The test set consists of 20,496 galaxies (20% of the total set) with mean redshift $z_{\rm mean} = 0.69$, max redshift $z_{\rm max} = 3.61$, and magnitudes $13.8 < i < 25.7$. See Table 1 for a full summary and Figure 2 for the redshift distribution.

5.1. Bayesian Photometric Redshifts

Bayesian Photometric Redshifts (BPZ; Benitez 2000) is a template-based photo-z estimator. Template-based estimators take a set of SED templates, assumed to be spanning and exclusive, and calculate observed fluxes over a grid of redshift values. For each template, BPZ evaluates a χ^2 function at each redshift on the grid:

$$\chi^2(z, T, A) = \sum_n \frac{1}{\sigma_n^2} (A \hat{f}_n(z, T) - f_n)^2, \quad (15)$$

where T denotes the template, z denotes the redshift, A is a normalization, and $\hat{f}_n$, f_n, and σ_n denote the calculated flux, the observed flux, and the fractional error as in Equation (5). The sum over n is a sum over the filters for the set of observed fluxes. BPZ then evaluates the likelihood for producing the observed galaxy fluxes: $p(\{f_n\}|z, T) \propto \exp(-\chi^2/2)$. The redshift posterior is then calculated by marginalizing over the

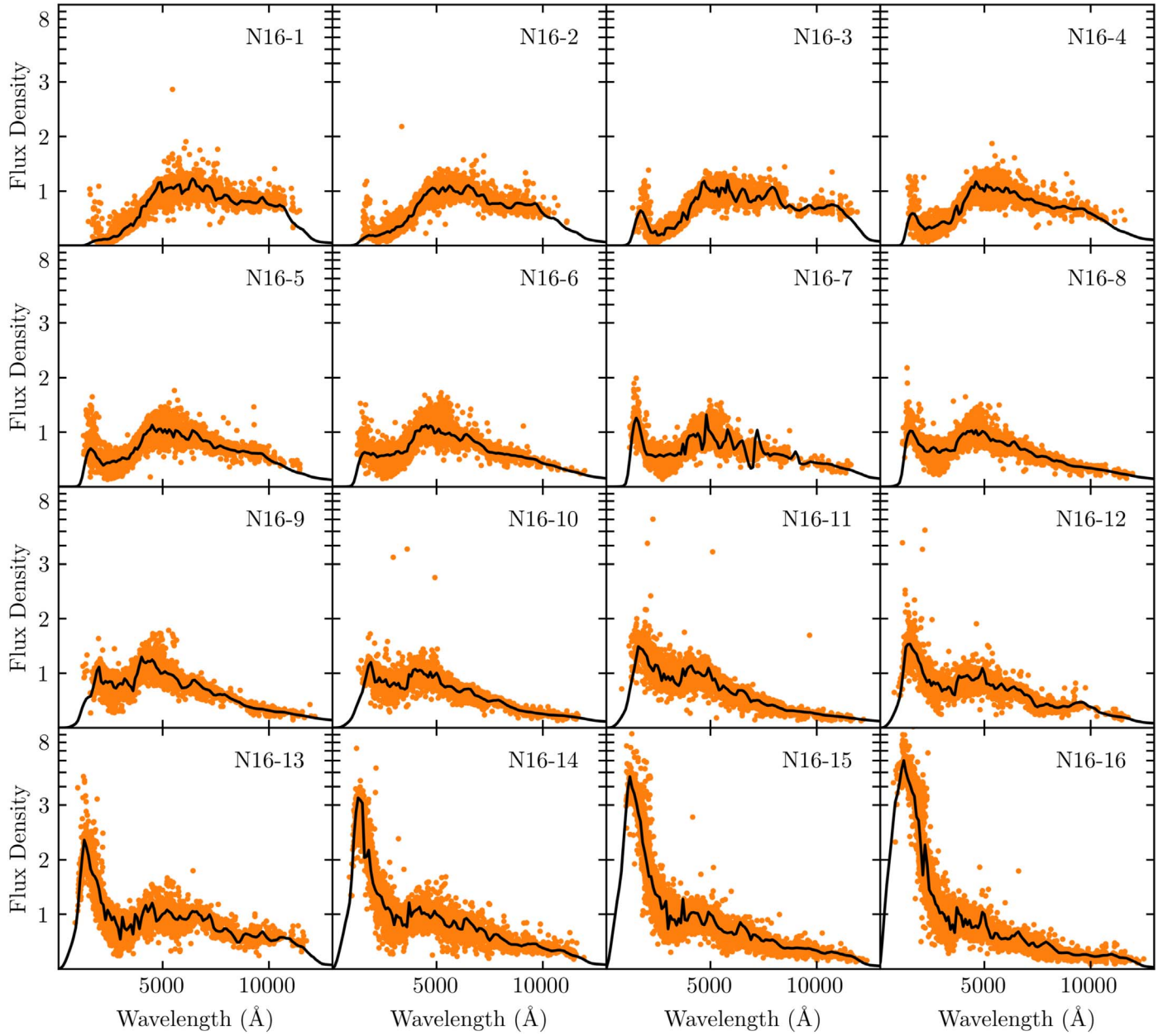

Figure 6. Trained N16 templates (black lines) with their final photometry sets (orange points). N16-1 is the reddest template, with each successive template getting bluer. These templates closely resemble the N8 templates and show the same emerging spectral features (see Figure 5), but consist of a more continuous transition from red to blue spectra.

set of templates:

$$p(z|\{f_n\}, m_0) = \sum_T p(z, T|\{f_n\}, m_0) \propto \sum_T p(z, T|m_0)\, p(\{f_n\}|z, T), \quad (16)$$

where $p(z, T|m_0)$ is a prior over the apparent magnitude m_0. Work is underway to determine how best to use the full information encoded in the redshift posterior generated by BPZ and other photo-z codes (e.g., Schmidt et al. 2020). In this work, however, only the mode of the posterior distribution is used to estimate the photo-z.

We use BPZ -v1.99.3[7] to estimate photo-z's. We turn off template interpolation by setting INTERP = 0. For simplicity, we treat nondetections as nonobservations. We use the various sets of SED templates described in Section 4 and use the prior described in the following section. All other settings were left as default.

5.2. Galaxy Magnitude Priors

Before estimating photo-z's with BPZ, we must first construct the magnitude priors, $p(z, T|m_0)$, calibrated to the galaxies in our training set. We separate the prior into two

[7] http://www.stsci.edu/~dcoe/BPZ/

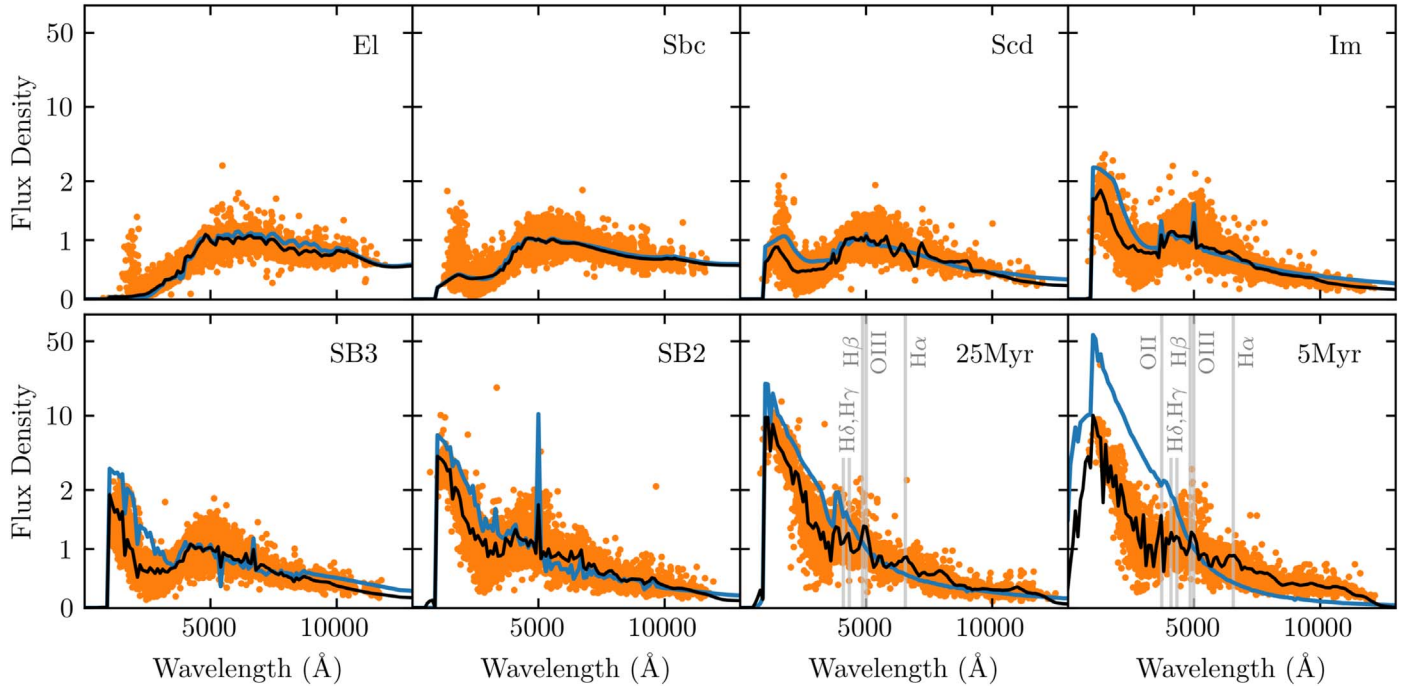

Figure 7. Result of training the CWW+SB4 templates. The original templates are in blue, the trained templates in black, and the final training sets are displayed as orange points. The 25 and 5 Myr templates have acquired emission lines that were not present in the initial templates. These are labeled in gray.

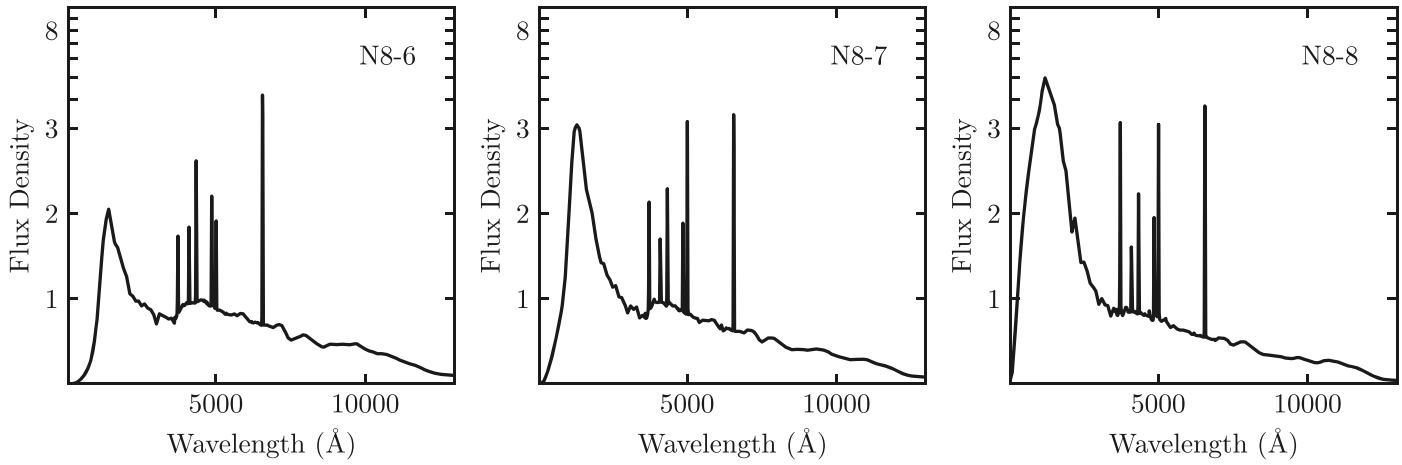

Figure 8. The N8-6, N8-7, and N8-8 templates with reconstructed emission lines (see Figure 5). The emission lines, left to right, are O II, Hδ, Hγ, Hβ, O III, and Hα. The wavelengths, relative amplitudes, and effective widths of these lines are in Table 3.

Table 3
Reconstructed Emission Lines

		N8-6		N8-7		N8-8	
Line	λ	r	W_λ	r	W_λ	r	W_λ
Hα	6563	2.86	132.7	2.86	103.3	2.86	115.2
Hβ	4861	1.00	32.9	1.00	26.4	1.00	30.3
Hγ	4340	1.18	36.5	1.31	31.6	1.28	37.1
Hδ	4102	0.65	19.6	0.72	16.7	0.71	20.7
O II	3727	2.04	58.1	1.27	32.0	0.74	24.4
O III	5007	2.08	68.0	2.42	66.1	0.86	27.3

Note. For each emission line, r is the amplitude relative to Hβ, and W_λ is the effective width in angstroms.

parts:

$$p(z, T|m_0) = p(T|m_0)\, p(z|T, m_0). \quad (17)$$

For the magnitude m_0, we use one of the i-bands in the following order of priority: i, i_2, I, i^+. Instead of constructing a different prior for each template, we follow Benitez (2000) in dividing our templates into three broad classifications: elliptical (El), spiral (Sp), or irregular/starburst (Im/SB). The CWW +SB4 templates are already classified under this scheme. We classify our new templates and each of the galaxies in the training set by assigning the classification of the CWW+SB4 template with the most similar colors, determined by minimizing the mean square error of the fluxes.

The N8 templates are determined to have one elliptical, four spiral, and three irregular/starburst galaxies; the N16 templates are determined to have two elliptical, eight spiral, and six

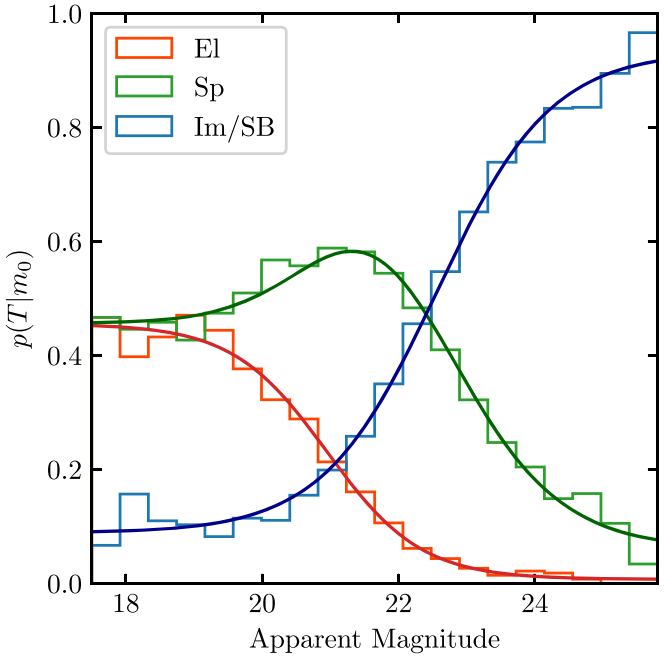

Figure 9. Fraction of each spectral class as a function of apparent magnitude. The histograms represent the fractions in the training set, and the curves are the spectral type priors fit to the data.

irregular/starburst galaxies. The fraction of each classification as a function of magnitude for the training-set galaxies is displayed in Figure 9.

We assume that the El and Im/SB galaxies have spectral priors of the form

$$p(T|m_0) = \frac{L_T}{1 + e^{-\kappa_T(m_0 - m_T)}} + C_T, \tag{18}$$

while $p(\mathrm{Sp}|m_0) = 1 - p(\mathrm{El}|m_0) - p(\mathrm{Im/SB}|m_0)$. The values of $\{L_T, \kappa_T, m_T, C_T\}$ for the El and Im/SB galaxies are found by fitting to the distributions in Figure 9. All three priors are plotted in the same figure, and the parameter values are listed in Table 4.

For the redshift prior, we use Equations (23) and (24) from Benitez (2000):

$$p(z|T, m_0) = \frac{1}{N_T} \exp\left\{-\left(\frac{z}{Z_T}\right)^{\alpha_T}\right\}, \tag{19}$$

where the normalization is

$$N_T = \frac{{Z_T}^{\alpha_T+1}}{\alpha_T} \Gamma\left(\frac{\alpha_T + 1}{\alpha_T}\right), \tag{20}$$

and the "median" redshift Z_T is chosen to have the linear dependence

$$Z_T(m_0) = z_{0T} + k_T(m_0 - 20). \tag{21}$$

Equation (19) reproduces the exponential cutoff at high redshifts present in the training set and can reasonably approximate any unimodal redshift distribution, from very narrow ($\alpha \gg 2$) to very broad ($\alpha \ll 1$). This flexibility reduces the bias introduced by the functional form of the prior (Benitez 2000). The nine parameters $\{\alpha_T, z_{0T}, k_T\}$ are determined by maximizing the likelihood $L = \prod_i p(z_i|T_i, m_{0i})$, where the product is over the galaxies in the training set. The parameters and their bootstrapped uncertainties are listed in Table 4.

5.3. Photo-z Results

We estimate photo-z's for the test-set galaxies using BPZ with the settings and priors described in the previous two sections. We used four template sets: the original CWW+SB4 templates, the trained CWW+SB4 templates, and the trained N8 and N16 templates.

BPZ provides two metrics for the photo-z estimates: ODDS and χ^2_{mod}. ODDS measures how narrowly peaked the posterior distribution $p(z|\{f_n\}, m_0)$ is around the estimated photo-z. Galaxies with low ODDS have either broad redshift posteriors or posteriors with multiple peaks. Here, χ^2_{mod} measures how well the best-fit template at the predicted redshift matches the observed fluxes. For more about these metrics, see Section 4 of Benitez (2000) and Section 4.3 of Coe et al. (2006). In this work, photo-z estimates with ODDS < 0.95 or $\chi^2_{\mathrm{mod}} > 1$ are excluded from the analysis, and the fraction excluded on this basis is reported as f_{cut}.

To further evaluate the results of BPZ, we calculate the scatter, bias, and outlier fraction of the photo-z estimates. Photo-z estimates are known to be contaminated with a significant number of outliers. This is largely driven by a degeneracy wherein the 1000 Å Lyman break in a high-redshift galaxy spectrum has optical colors similar to the 4000 Å Balmer break in a low-redshift galaxy spectrum. BPZ attempts to break this degeneracy with the galaxy magnitude prior (i.e., galaxies with brighter apparent magnitudes are more likely to be at a lower redshift), yet there are still a large number of outliers.

To address this issue, we evaluate the statistics of the interquartile range (IQR) of the data, as these measures are robust to the presence of outliers. We follow Graham et al. (2018) in introducing the quantity $\Delta z_{1+z} = (z_{\mathrm{spec}} - z_{\mathrm{phot}})/(1 + z_{\mathrm{phot}})$. The numerator quantifies the photo-z error, and the denominator compensates for the larger uncertainty at high redshifts. We define the scatter of the photo-z estimates, σ_{IQR}, as the width of the IQR in Δz_{1+z}, divided by 1.349 to convert to the equivalent of a Gaussian standard deviation. We define the bias of the photo-z estimates as the mean value of Δz_{1+z} for galaxies within the IQR. The uncertainties of these two values are bootstrapped by calculating the values on 1000 random samples with replacement.

Table 4
Parameters for the Priors, $p(z, T|m_0)$

Spectral Type	L_T	κ_T	m_T	C_T	α_T	z_{0T}	k_T
El	0.448 ± 0.017	−1.45 ± 0.16	21.0 ± 0.1	0.007 ± 0.009	3.88 ± 0.04	0.484 ± 0.003	0.119 ± 0.002
Sp	…	…	…	…	3.40 ± 0.04	0.493 ± 0.003	0.124 ± 0.002
Im/SB	0.845 ± 0.031	1.20 ± 0.11	22.6 ± 0.1	0.089 ± 0.013	2.22 ± 0.03	0.361 ± 0.009	0.130 ± 0.008

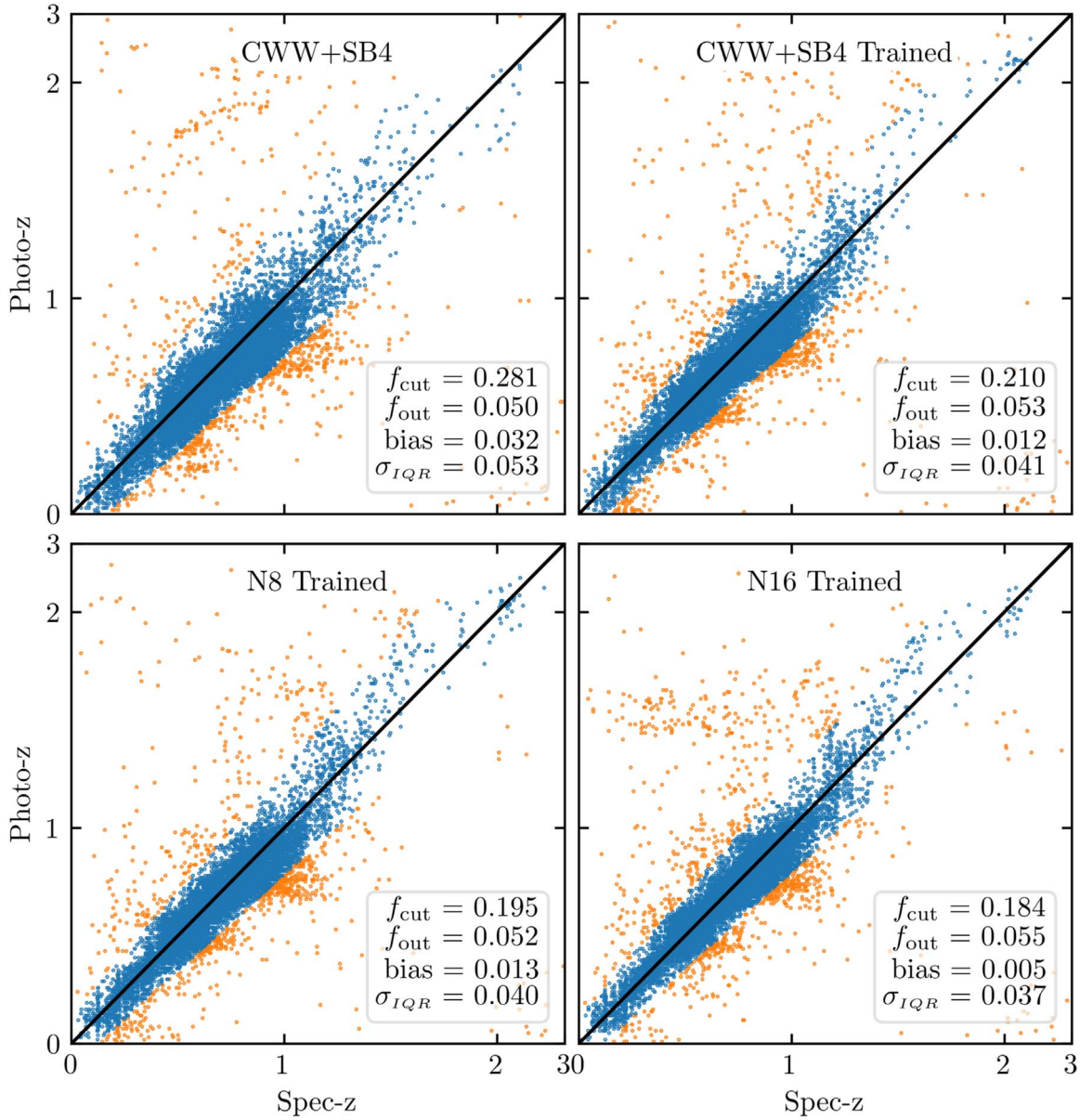

Figure 10. Results of photo-z estimation with BPZ, using the four different template sets. Photo-z estimates are displayed as points: inliers are blue and outliers are orange. The black line represents perfect estimation (i.e., photo-z = spec-z). The statistics printed in each panel are for the entire data set.

Outliers are identified as photo-z's with $\Delta z_{1+z} > 3\sigma_{\rm IQR}$, and the fraction of outliers is reported as $f_{\rm out}$.

The photo-z results can be seen in Figure 10. The photo-z estimates that passed the cuts on ODDS and $\chi^2_{\rm mod}$ are displayed as points: the inliers in blue, the outliers in orange. The values of the photo-z statistics for each template set are printed in each panel. For all four template sets, the photo-z estimation is reasonably accurate for spec-z's $z < 1.5$. For higher redshifts, there appears to be a systematic bias toward higher photo-z's. Reduced photo-z accuracy is generally expected for spec-z's greater than 1.5, as the Balmer break leaves the optical bands at around $z = 1.4$ and the Lyman break does not enter the ultraviolet bands until $z = 2.5$.

For the CWW+SB4 templates, the training algorithm decreased the fraction of photo-z's cut by 25%, the bias by 63%, and the scatter by 23%, but it did not improve the outlier fraction. We were able to achieve similar photo-z results using the trained N8 and N16 template sets, demonstrating that our training algorithm can be used to generate photo-z templates without any a priori information about galaxy spectra. Compared to the CWW+SB4 templates, N8 templates decreased $f_{\rm cut}$ by 31%, bias by 59%, and scatter by 25%. The N16 templates decreased $f_{\rm cut}$ by 35%, bias by 84%, and scatter by 30%. In all cases, the training algorithm decreases the fraction of bad photo-z's ($f_{\rm cut} + f_{\rm out}$), the bias, and the scatter.

Comparing the results for the N8 and N16 template sets indicates that increasing the number of templates can reduce the fraction cut and the bias and scatter of the photo-z estimates. To further investigate this relationship, we calculate the photo-z statistics for a range of template numbers, the results of which are in Figure 11. We find that increasing the number of templates decreases the fraction cut and the bias, as well as slightly decreasing the scatter. The trend for outlier fraction is less clear.

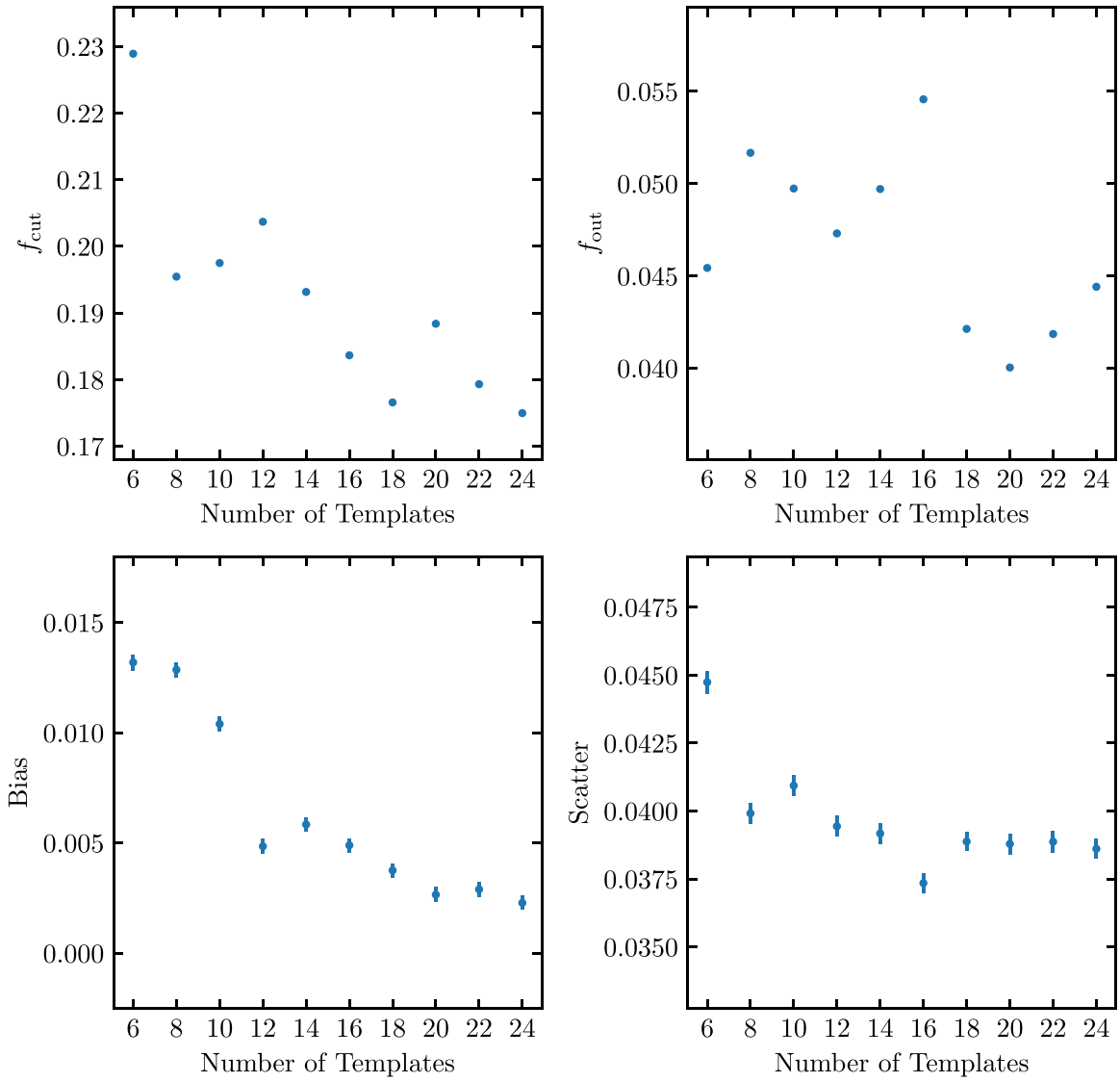

Figure 11. Photo-z statistics as a function of template number. Statistics are for the full redshift range.

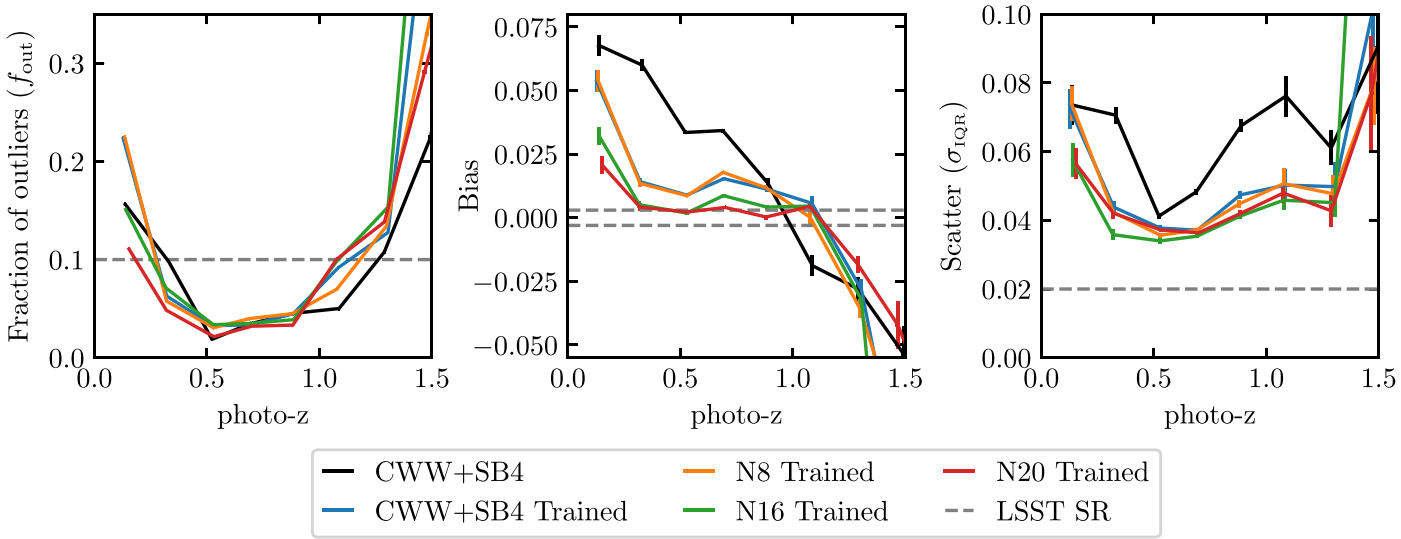

Figure 12. Photo-z metrics for the various template sets as a function of redshift bin. LSST science requirements are shown as dashed gray lines.

The N20 set has $f_{\rm cut} = 0.188$ (a 33% decrease compared to CWW+SB4), $f_{\rm out} = 0.040$ (a 20% decrease), bias = 0.003 (a 91% decrease), and scatter = 0.039 (a 26% decrease).

The value of the metrics as a function of photo-z can be seen in Figure 12. In addition to the template sets plotted above, we add the N20 set. For comparison, plotted in gray are the LSST science requirements for the metrics as listed in the LSST Science Requirement Document (SRD; Ivezić & LSST Science Collaboration 2018). The SRD lists the following minimum requirements to enable the envisioned LSST cosmological studies: root-mean-square error $<0.02(1 + z_{\rm phot})$; $f_{\rm out} < 10\%$; average bias $< 0.003(1 + z_{\rm phot})$. The SRD lists these requirements for an $i < 25$, magnitude-limited sample of four billion galaxies in $0.3 < z < 3.0$. For comparison, our test set consists of 20,496 galaxies with $i < 25.7$ in the range $z < 3.6$, including 19,391 galaxies with $i < 25$ in the range $0.3 < z < 3.0$. In Figure 12, we show that for redshifts $0.3 < z < 1.2$ we are able to achieve an appropriate outlier fraction, and that our training algorithm makes great progress on the bias, almost reaching the threshold required for LSST. We also make modest progress on the scatter, but reduction by another factor of two is still required. Beyond redshift $z = 1.2$, all of our metrics fail the LSST science requirements.

6. Discussion

In Section 2, we demonstrated that our training algorithm could learn galaxy SED templates from photometry at a high resolution relative to the filters used to make the observations. We are able to learn a set of templates over twice the size of the standard CWW+SB4, showing a smooth progression of galaxy colors from red to blue. The spectra contain relatively high-resolution spectral features, and postprocessing can further reconstruct emission and absorption lines. The bluer templates contain more structure as they represent star-forming galaxies and thus have stronger emission lines. In addition, the bluer templates have a larger number of high-redshift galaxies compared to the red templates, which aids the reconstruction of high-resolution features. While the high-redshift galaxies number in the hundreds instead of thousands, our results indicate that high-resolution features can be reliably reconstructed with only a few hundred galaxies.

Our method has a number of limitations. The success of our algorithm relies on the ability to generate a naive set of templates as a starting point that will reliably divide the photometry by the spectral type of the galaxy. This is relatively easy to accomplish for fewer than 20 templates, as was demonstrated by our simple photometry-matching procedure and the log-normal templates we used. This is a strength of the algorithm as it is relatively robust to the starting templates. If, however, you wish to derive more than 20 templates from the photometry, care must be taken in the division of the photometry set to ensure there are sufficient galaxies in each subset to fully sample the entire wavelength range for the templates. In addition, the inherently discretized way in which we divide the photometry set stands in the way of generating a truly continuous set of SED templates. For a more continuous set of templates, one might imagine taking two "adjacent" photometry sets and assembling a photometry set "between" them by taking the bluer half of one set together with the redder half of the other. Equally, we could construct a moving window that progressively subdivides a sample based on color (with galaxies allowed to be present in more than one subset).

Our data consists only of broadband photometry, but our algorithm would work equally well with narrow bands as well. Combining broadband and narrowband photometry would expand the data set and further constrain the templates. In particular, the addition of narrowband photometry should increase the resolution of spectral features recovered, and it may allow one to resolve features such as the Hγ and Hδ emission lines that we had to treat as a single feature. One could also include bands from a wider range of wavelengths to increase the wavelength range over which the templates are constrained. We attempted to include fluxes from the *K*-bands included with the zCOSMOS and VIPERS catalogs to learn infrared wavelengths for the templates, but there appeared to be systematic calibration issues in the data that we could not resolve. There is evidence that the inclusion of near-infrared and near-ultraviolet photometry in photo-z estimation can reduce outliers and scatter by up to 50% each (Graham et al. 2020).

In addition, for the results presented here, we used only galaxy fluxes with S/N greater than 20. One can use galaxies with lower S/N if outlier fluxes are removed from the photometry sets before training (we had success using an Isolation Forest; Liu et al. 2008; Liu et al. 2012). However, lowering the S/N of the photometry generally reduces the resolution of the structure that one can reconstruct.

The training algorithm itself could be made more sophisticated by restoring the wavelength dependence of the hyperparameter Δ_k. We also hope to move beyond an iterative regression approach into deep learning, perhaps using generative adversarial networks (GANs; Goodfellow et al. 2014).

When constructing the `BPZ` prior, we sorted our templates into broad spectral classes. In the N8 set, for example, we determined that one template was elliptical, four were spiral, and three were irregular/starburst. Each of our templates has approximately the same number of galaxies matched to it, and the photometry matched to the elliptical templates does not display more variance than the photometry matched to other templates. These observations indicate that our data set contains a larger number of spiral and irregular/starburst galaxies than elliptical galaxies, rather than suggesting that the space of elliptical galaxy spectra is less finely sampled. For this reason, we do not expect the imbalance of the template number in each class to have a large impact on the photo-z quality, but nevertheless we note that a more sophisticated prior could be constructed without relying on this broad classification scheme, which may provide better redshift estimates.

We found in Section 5.3 that our training algorithm can improve the bias and scatter of photo-z estimates. We found that increasing the number of templates enhances these improvements, with the best results for 20 templates. As mentioned above, with our current method for generating photometry sets, we struggle to reliably reconstruct more than 20 templates, so whether these benefits continue to decrease with template number is unknown.

We can compare our method for generating more SED templates with `BPZ`'s method of linearly interpolating between templates. N8 with `INTERP = 2` generates 22 total templates. Table 5 compares the photo-z results using these templates with the results using 22 templates learned from the photometry with `INTERP = 0`. It is clear that, as far as $f_{\rm out}$ and bias, our method for generating extra templates is superior to the linear interpolation used by `BPZ`.

Table 5
Comparison to BPZ Interpolation

	INTERP	Total N	f_{cut}	f_{out}	Bias	Scatter
N8	0	8	0.228	0.058	0.014	0.040
N8	2	22	0.209	0.060	0.012	0.037
N22	0	22	0.214	0.045	0.004	0.039

Note. Total N is the total number of SED templates in the set, including those interpolated by BPZ. Statistics quoted are for the full redshift range.

The photo-z estimation with our learned template sets outperforms the results of the standard CWW+SB4 templates. However, more work needs to be done to reach the requirements set for LSST, especially for redshifts $z > 1$. Templates can be trained for LSST science using the substantial overlap of LSST photometry with the eBoss (Dawson et al. 2016) and Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016) surveys, which will provide hundreds of thousands of spec-z's for LSST photo-z training and calibration (Schmidt et al. 2014; Newman et al. 2015).

Our training method can be extended to other domains (e.g., stellar spectral reconstruction) where one can take a large set of incomplete data, segment that data into classes, and treat the set of unique observations in each class as an ensemble of observations of some class archetype, and thereby reconstruct more complete information. We plan to adapt the method to reconstruct supernova light curves from supernova photometry.

7. Conclusions

We have shown that galaxy SED templates can be learned directly from a data set of broadband photometry. Large sets of photometry at various redshifts can be leveraged to reconstruct high-resolution features, such as the Hα, Hβ, Hγ, Hδ, O II, and O III emission lines, as well as Na and Mg absorption lines. Simple postprocessing can further improve the resolution of these reconstructed lines. The number of templates learned is variable and can be increased to more continuously sample the space of galaxy spectra and to improve photo-z results.

We used our templates to estimate photo-z's for a test set of galaxies using BPZ. We found that training the standard set of templates that comes with BPZ decreases the fraction of bad photo-z's by 21%, the bias by 63%, and the scatter by 23%. Our own trained naive templates yielded better results. We learned a set of 20 templates from the data that reduced the fraction of bad photo-z's by 31%, the bias by 91%, and the scatter by 26%. These derived templates outperform the interpolated spectra used by BPZ. The improvements in bias are almost sufficient to meet the requirements set for LSST, but another reduction by a factor of two is needed for the scatter.

The templates derived with our training algorithm demonstrate that accurate galaxy spectra can be learned from broadband photometry. Our SEDs could potentially be used for applications other than photo-z's, and our learning algorithm can be extended to other applications, such as learning supernova light curves from photometry.

Our derived templates and the code used to produce these results are publicly available in a dedicated Github repository: https://github.com/dirac-institute/photoz_template_learning.

The authors would like to thank Bryce Kalmbach for providing advice in early stages of this work, Sam Schmidt for his help with BPZ, Melissa Graham for her code to calculate photo-z statistics, and Tamás Budavári for comments on the manuscript. This work was supported by the U.S. Department of Energy, Office of Science, under award DE-SC-0011635. The authors also acknowledge partial support from NSF grants AST-1715122 and OAC-1739419.

This research is based on observations made with ESO Telescopes at the La Silla or Paranal Observatories under program IDs 175.A-0839(B), 175.A-0839(D), 175.A-0839(I), 175.A-0839(J), 175.A-0839(H), and 175.A-0839(F). This research is also based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada–France–Hawaii Telescope (CFHT), which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This research is also based in part on data products produced at Terapix, available at the Canadian Astronomy Data Centre as part of the Canada–France–Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS. We use data from the VIMOS VLT Deep Survey, obtained from the VVDS database operated by Cesam, Laboratoire d'Astrophysique de Marseille, France. We also use data from the VIMOS Public Extragalactic Redshift Survey (VIPERS). VIPERS has been performed using the ESO Very Large Telescope, under the "Large Programme" 182.A-0886. The participating institutions and funding agencies are listed at http://vipers.inaf.it. This research has also made use of the SVO Filter Profile Service (http://svo2.cab.inta-csic.es/theory/fps/) supported by the Spanish MINECO through grant AYA2017-84089.

Software: Astropy (Astropy Collaboration et al. 2013), BPZ (Benitez 2000), Jupyter (Kluyver et al. 2016), Matplotlib (Hunter 2007), Numpy (Van Der Walt et al. 2011), Scikit-learn (Pedregosa et al. 2011), Scipy (Virtanen et al. 2020).

ORCID iDs

John Franklin Crenshaw https://orcid.org/0000-0002-2495-3514
Andrew J. Connolly https://orcid.org/0000-0001-5576-8189

References

Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Assef, R. J., Kochanek, C. S., Brodwin, M., et al. 2008, ApJ, 676, 286
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, 33
Benitez, N. 2000, ApJ, 536, 571
Benitez, N., Ford, H., Bouwens, R., et al. 2004, ApJS, 150, 1
Bessell, M. S. 2005, ARA&A, 43, 293
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Bruzual, A. G., & Charlot, S. 1993, ApJ, 405, 538
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Budavári, T., Szalay, A. S., Connolly, A. J., Csabai, I., & Dickinson, M. 2000, AJ, 120, 1588
Coe, D., Benítez, N., Sánchez, S. F., et al. 2006, AJ, 132, 926
Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, ApJS, 43, 393
Connolly, A. J., Csabai, I., Szalay, A. S., et al. 1995, AJ, 110, 2655
Cooper, M. C., Aird, J. A., Coil, A. L., et al. 2011, ApJS, 193, 14
Csabai, I., Connolly, A. J., Szalay, A. S., & Budavári, T. 2000, AJ, 119, 69
Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44
de Jong, J. T., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, ExA, 35, 25

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
Fruchter, A., & Hook, R. 2002, PASP, 114, 144
Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., et al. 2014, arXiv:1406.2661
Graham, M. L., Connolly, A. J., Ivezić, Ž., et al. 2018, AJ, 155, 1
Graham, M. L., Connolly, A. J., Wang, W., et al. 2020, AJ, 159, 285
Green, J., Schechter, P., Baltay, C., et al. 2012, arXiv:1208.4012
Groves, B., Brinchmann, J., & Walcher, C. J. 2012, MNRAS, 419, 1402
Hudelot, P., Cuillandre, J.-C., Withington, K., et al. 2012, yCat, 2317, 0
Hunter, J. D. 2007, CSE, 9, 90
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Ivezić, Ž. & LSST Science Collaboration 2018, LSST Project Management LPM-17, http://ls.st/srd
Izbicki, R., & Lee, A. B. 2017, EJSta, 11, 2800
Kind, M. C., & Brunner, R. J. 2013, MNRAS, 432, 1483
Kinney, A. L., Calzetti, D., Bohlin, R. C., et al. 1996, ApJ, 467, 38
Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, ed. F. Loizides & B. Schmid (Amsterdam: IOS Press), 87
Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, 14
Le Fèvre, O., Mellier, Y., McCracken, H. J., et al. 2004, A&A, 417, 839
Lee, M. A., Budavári, T., Sullivan, I. S., & Connolly, A. J. 2019, AJ, 157, 182
Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, ApJS, 184, 218
Liu, F. T., Ting, K. M., & Zhou, Z.-H. 2008, in ICDM 2008. Eighth IEEE Int. Conf. on Data Mining (IEEE Computer Society), ed. F. Giannotti et al. (Los Alamitos, CA: IEEE Computer Society), 413
Liu, F. T., Ting, K. M., & Zhou, Z.-H. 2012, ACM Trans. Knowl. Discov. Data, 6, 39
LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201
Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJL, 619, L1
Miyazaki, S., Komiyama, Y., Sekiguchi, M., et al. 2002, PASJ, 54, 833
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Newman, J. A., Abate, A., Abdalla, F. B., et al. 2015, APh, 63, 81
Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 57
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Salvato, M., Ilbert, O., & Hoyle, B. 2019, NatAs, 3, 212
Schmidt, S. J., Malz, A. I., Soo, J. Y. H., et al. 2020, arXiv:2001.03621v1
Schmidt, S. J., Newman, J. A., Abate, A. & t. S. N. W. P. Team 2014, arXiv:1410.4506
Scodeggio, M., Guzzo, L., Garilli, B., et al. 2018, A&A, 609, A84
The Dark Energy Survey Collaboration 2005, arXiv:astro-ph/0510346
Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Zhou, R., Cooper, M. C., Newman, J. A., et al. 2019, MNRAS, 488, 4565