

# Learning High Resolution Galaxy Spectra from Broadband Photometry

JOHN FRANKLIN CRENSHAW <sup>1</sup> AND ANDREW J. CONNOLLY <sup>2</sup>

<sup>1</sup>*Department of Physics, University of Washington*

<sup>2</sup>*DIRAC Institute, Department of Astronomy, University of Washington* *Ask Andy what to put here (e.g. eScience)*

## ABSTRACT

**Rough sketch:** We can learn SED's directly from photometric data, without any a priori knowledge of galaxy spectra. This is done by starting with naive templates, matching photometry to these templates, then applying a iterative learning algorithm. This algorithm can also be used to calibrate a set of standard templates to a data set. We apply this method to a data set that I need to describe. The learned SED's have a similar fraction of outliers compared to CWW+SB4, but have lower bias and scatter. The best results come from training the CWW+SB4 templates themselves. Add a link to the Github repo that contains all of the code. **Need to really write this.**

## 1. INTRODUCTION

Redshift is one of the most important observables in extragalactic astronomy and cosmology. Studies of galaxy evolution, galaxy clusters, large-scale structure, weak lensing, etc 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 multi-object spectrographs, spectroscopic measurements are expensive and time-consuming and we can only collect spectra for a small fraction of the galaxies imaged in 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](#)), will 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 must increasingly rely on photometric redshifts (photo-z's; see [Salvato et al. 2019](#) for a review).

Photo-z's are estimates of galaxy redshifts that rely on the changing colors of galaxies as their spectral energy distributions (SED's) redshift through a series of broad-

band 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. [TPZ Kind & Brunner 2013](#), [FlexZBoost Izbicki & Lee 2017](#), [CMNN Graham et al. 2018](#)). Once trained, they can predict galaxy redshifts given photometry only. 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 accounted for in the derived mapping from colors to redshift. The downside of ML methods is that they are not physically motivated. Their success relies on how representative and well-controlled the training set is, and they are unable to extrapolate beyond that set.

Template fitting photo-z estimators (e.g. [LePhare Arnouts et al. 1999](#), [BPZ Benitez 2000](#), [EAZY Brammer et al. 2008](#)) work on the assumption that galaxy photometry are 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 SED's from spectral synthesis models, e.g. [Bruzual A. & Charlot \(1993\)](#), and deriving templates from the observed spectra of local galaxies, e.g. [Benitez et al. \(2004\)](#).

The primary advantage of these methods is that it is not limited to the interpolation of a training set. A key limitation of these methods is that they do not guarantee that the SED templates will span the full distribution of galaxy spectra in a given data set, nor that they will properly account for the effects of dust, the evolution of galaxy spectra with redshift, etc. 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 high- $z$  galaxy spectra is incredibly expensive.

Several previous works have attempted to combine the advantages of these two approaches by deriving SED templates from a photometry training set, and then using the derived templates for photo- $z$  estimation (Budavári et al. 2000; Csabai et al. 2000; Kalmbach & Connolly 2017). This 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, learning templates from a large training set 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 these 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, using a large set of top hat bins as our parameters. We also 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. We demonstrate how leveraging a large set of galaxy photometry, which amount to low resolution spectra, allows us to essentially observe a smaller set of SED templates at an ensemble of redshifts. This effectively over-samples the template SED's, allowing us to reconstruct spectral features at a resolution much higher than that of the broadband filters used to measure the photometry. This method 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 SED's using differential chromatic refraction (DCR; Lee et al. 2019).

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 use templates learned from scratch and pre-existing templates trained on the data to estimate photo- $z$ 's using BPZ (Benítez 2000). We use these learned templates to estimate photo- $z$ 's with BPZ and find that the train-

ing reduces the bias scatter of the redshift estimates, with little impact on the fraction of catastrophic outliers. **Something about the number of templates learned.**

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 we are using for the template training and the redshift estimation. In Section 4, we apply the template training algorithm to sets of naive templates and a pre-existing set of templates derived from galaxy observations and spectral synthesis models. We discuss how the training was performed and discuss the results, including reconstructing and analyzing the emerging spectral features. In Section 5, we use our templates to estimate photo- $z$ 's for a training set of galaxies and analyze the results. We conclude in Section 6 with a discussion of our results and future goals.

## 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 over-sampling 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 can 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 true set of SED templates from which the galaxy spectra are sampled. 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

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

$S(\lambda)$  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,

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

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

The observed fluxes are assigned to the template whose shape is 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 assemble matched photometry sets containing many galaxies for each template.

Examining how the complete photometry set is divided into the smaller matched sets is helpful in assembling the initial set of templates. The initial templates should be selected so that the corresponding matched sets divide the galaxies by the different spectral shapes visible in the photometry. It is also important that each training set contains a sufficient number of fluxes distributed across the wavelengths of interest, as the perturbation algorithm derived in the next section relies on over-sampling 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:

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

We wish to perturb the template so that the template fluxes,  $\hat{f}_n$ , better match the observed fluxes,  $f_n$ . Replacing  $S(\lambda)$  with the discrete representation  $s_k$ , where  $k$  indexes wavelength bins, we can define the cost function

$$\chi^2 = \sum_n \frac{1}{\sigma_n^2} (\hat{f}_n(\{\hat{s}_k\}) - f_n)^2 + \sum_k \frac{1}{\Delta_k^2} (\hat{s}_k - s_k)^2, \quad (4)$$

with respect to the perturbed template,  $\hat{s}_k$ . The optimum perturbation to  $s_k$  can then be found via a multidimensional minimization of the cost function. The first term in Equation 4 penalizes differences between the observed fluxes and the perturbed template fluxes, weighted according to  $\sigma_n$ , the fractional error of the measured flux. The perturbed template fluxes can be calculated with a discretized version of Equation 3:

$$\hat{f}_n(\{\hat{s}_k\}) = \sum_k \hat{s}_k r_{k'}^n \Delta \lambda_{k'} \quad (5)$$

where  $r_{k'}^n$  is the discrete representation of  $R^n(\lambda)$ ,  $k'$  is the wavelength bin corresponding to  $\lambda_{k'} = (1+z_n)\lambda_k$  and  $\Delta \lambda_{k'} = (1+z_n)\Delta \lambda_k$ , where  $\Delta \lambda_k$  is the width of wavelength bin  $k$ . The second term in Equation 4 penalizes large perturbations, weighted by the hyperparameters  $\Delta_k$ . This parameter controls 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:

$$\begin{aligned} \xi_k &= \hat{s}_k - s_k \\ g_n &= f_n - \sum_k s_k r_{k'}^n \Delta \lambda_{k'}. \end{aligned} \quad (6)$$

Then, we have:

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

which can be analytically minimized:

$$\frac{\partial \chi^2}{\partial \xi_l} = 0 \implies \sum_k M_{lk} \xi_k = \nu_l. \quad (8)$$

The matrix  $M$  and vector  $\nu$  are defined

$$M_{lk} = \sum_n \frac{1}{\sigma_n^2} (r_{l'}^n \Delta \lambda_{l'}) (r_{k'}^n \Delta \lambda_{k'}) + \frac{\delta_{lk}}{\Delta_k^2}, \quad (9)$$

$$\nu_l = \sum_n \frac{g_n}{\sigma_n^2} (r_{l'}^n \Delta \lambda_{l'}),$$

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

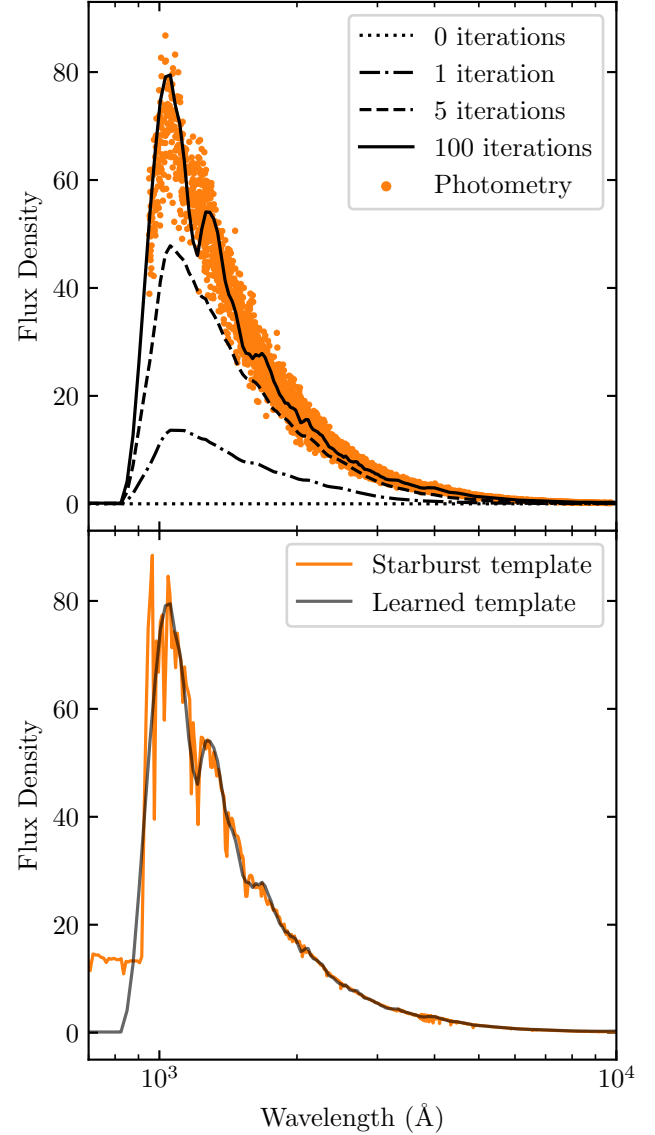
Iterating the perturbation changes the shape of the template SED to better match the measured photometry, as shown in Budavári et al. (2000). An example of this process can be seen in Figure 1. Fluxes in the *ugrizY* filters listed in Table 2 were calculated for a starburst galaxy template at 1000 random 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, higher  $\Delta_k$  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 SED's so that re-running the photometry matching will now results in different sets than previously obtained. As better approximations of the true underlying SED templates are generated by the perturbation algorithm, the photometry matching algorithm will do a better and better job at dividing the photometry data into sets that characterize distinct spectral shapes. 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 roughly separate out the different spectral shapes displayed in the photometry. For further discussion of how the number of templates effects photo- $z$  results, see Section 5.2.

The rest of the hyperparameters consist of the set of  $\Delta_k$ . These parameters, which set the relative weighting of the regularization term in Equation 4, determine the stability and speed of the training algorithm. If the  $\Delta_k$  are too large, training will be very slow and a large number of perturbation and training rounds will be required.



**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 5Myr starburst template from Coe et al. (2006) at 1,000 random redshifts in the range  $z=0$  to  $z=3$  using the *ugrizY* filters listed in Table 2. The simulated photometry has 10% Gaussian error. The template is shown after various stages of the training. Bottom: the learned template is plotted with the original starburst template.

If the  $\Delta_k$  are too low, the training becomes unstable and the final templates will be over-fit. 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 that  $\Delta \equiv \Delta_k$  has a single value for each training set that is independent of wavelength. In choosing the appropriate value of  $\Delta$  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 4. The necessary value of  $\Delta$  will vary by training set, as the number of terms in the sum over fluxes (i.e. the sum over  $n$  in Equation 4) will vary by training set. To this end, we make the following approximation:

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

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

$$w = \frac{\sum_k \frac{1}{\Delta^2} (\hat{s}_k - s_k)^2}{\sum_n \frac{1}{\sigma_n^2} (\hat{f}_n - f_n)^2} \sim \frac{N_k / \Delta^2}{N_n / \bar{\sigma}^2}, \quad (11)$$

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

$$\Delta \simeq \bar{\sigma} \sqrt{\frac{N_k}{w N_n}}. \quad (12)$$

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 2 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<sup>1</sup> 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  $\text{SNR} > 20$ . 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  $1.7 \text{ deg}^2$  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 description<sup>2</sup>, 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 11 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. We use only galaxies detected in all of the optical bands. 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  $8.7 \text{ deg}^2$ , with approximately 25,000 galaxies in the range  $17.5 < i < 22.5$ ; the *Deep* survey covers  $0.74 \text{ deg}^2$ , with approximately 11,000 galaxies in the range  $17.5 < i < 24$ ; the *Ultra-Deep* survey covers  $512 \text{ arcmin}^2$ , 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 (Hudlot 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 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 (Hudlot et al. 2012). The final set contains 71,951 galaxies with redshifts  $z < 2.15$  and magnitudes  $17.7 < i < 23.3$ .

#### 3.4. *DEEP2* and *DEEP3*

*DEEP2* and *DEEP3* are redshift surveys conducted with the DEIMOS spectrograph on the Keck 2 telescope. *DEEP2* (Newman et al. 2013) consists of four

<sup>1</sup> 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.

<sup>2</sup> [https://www.eso.org/sci/observing/phase3/data\\_releases/zcosmos.dr3\\_b2.pdf](https://www.eso.org/sci/observing/phase3/data_releases/zcosmos.dr3_b2.pdf)



**Table 1.** Summary of the spec-z and photometry data sets.  $N_{\text{gal}}$  is the total number of galaxies in the set,  $f_{\text{gal}}$  is the fraction of galaxies in the set, and  $\bar{\sigma}_i$  is the mean fractional flux error for the  $i$ -band photometry.

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

**Table 2.** The 19 filters used to measure the galaxy photometry in the data set. Mean wavelength,  $\lambda_0 = \int \lambda R(\lambda) d\lambda$ , and effective width,  $W_{\text{eff}} = \text{Max}[R(\lambda)]^{-1}$ , are given in angstroms. Filters are listed in order of increasing  $\lambda_0$ . The  $i_2$  band is the replacement to the Megacam  $i$ -band installed in 2007. This filter is named  $y$  in the CFHTLS catalogues (Hudlot 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. The system response functions for each filter were obtained from the Spanish Virtual Observatory (SVO) Filter Profile Service.

Filter	Telescope	Instrument	$\lambda_0$	$W_{\text{eff}}$
<i>NUV</i>	GALEX		2343.1	767.3
<i>u</i>	CFHT	Megacam	3817.7	525.4
<i>B</i>	CFHT	CFH12k	4342.5	873.6
<i>B<sub>J</sub></i>	Subaru	Suprime	4478.4	763.9
<i>g<sup>+</sup></i>	Subaru	Suprime	4808.5	1043.1
<i>g</i>	CHFT	Megacam	4899.9	1293.8
<i>V</i>	CFHT	CFH12k	5393.7	882.7
<i>V<sub>J</sub></i>	Subaru	Suprime	5493.0	862.4
<i>r</i>	CHFT	Megacam	6278.2	1120.2
<i>r<sup>+</sup></i>	Subaru	Suprime	6314.8	1211.4
<i>R</i>	CFHT	CFH12k	6603.5	1138.5
<i>i<sub>2</sub></i>	CHFT	Megacam	7584.5	1409.4
<i>i</i>	CHFT	Megacam	7676.6	1307.6
<i>i<sup>+</sup></i>	Subaru	Suprime	7709.1	1361.7
<i>I</i>	CFHT	CFH12k	8277.3	1816.7
<i>z</i>	CHFT	Megacam	8857.6	1040.1
<i>z<sup>+</sup></i>	Subaru	Suprime	9054.5	1012.3
<i>Y</i>	Subaru	Suprime	10216.0	996.2
<i>J</i>	UKIRT	WFCAM	12508.5	1476.8

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  $\geq 3$ . Photometry was measured in  $u, g, r, i_2, i, z$  on CFHT-Megacam and  $Y$  on Subaru (Miyazaki et al. 2002). The final set contains 10,695 galaxies with redshifts  $z < 1.91$  and magnitudes  $15.3 < i < 25.74$ .

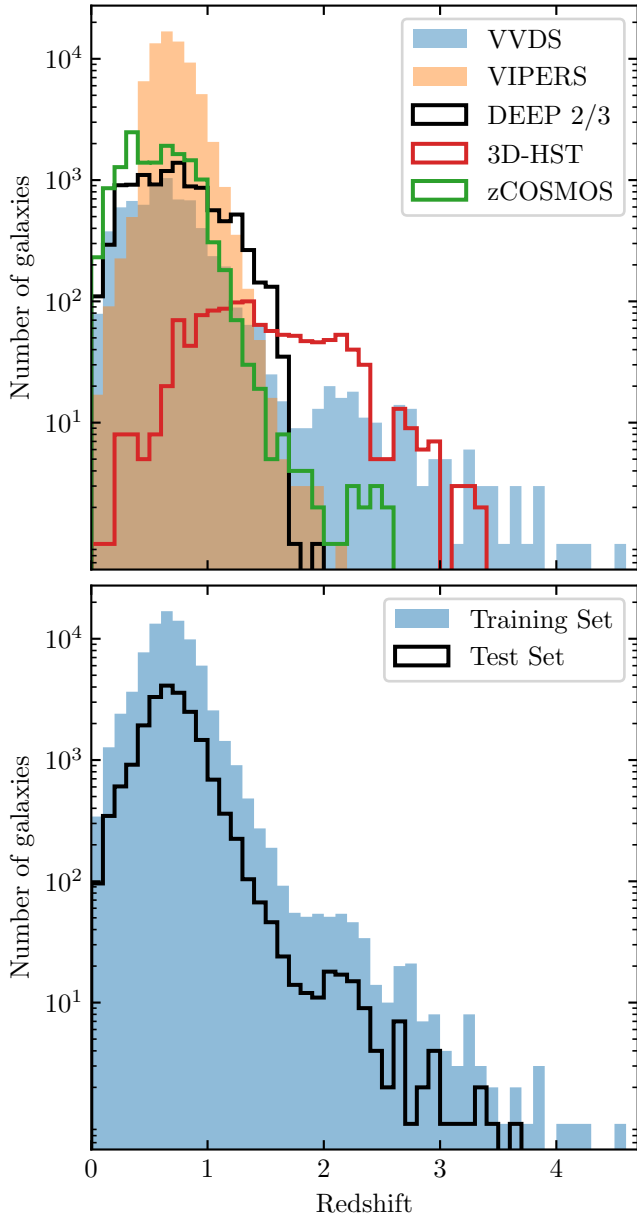
### 3.5. 3D-HST

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 1,153 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 test set, consisting of random 80% and 20% samples respectively of the entire data set. The training set will be used to train the SED templates, while the test set will be used to test the learned templates via photo- $z$  estimation (see Section 5). The training set consists of 81,980 galaxies, with mean redshift  $z_{\text{mean}} = 0.69$ , max redshift  $z_{\text{max}} = 4.54$ , 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. They are “naive” because they are simply chosen by eye



**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.

to roughly divide the photometry into groups by spectral shape, but otherwise are not based on any theoretical models or observed SED's. Each of the naive templates is a log-normal function,

$$S(\lambda) \propto \frac{1}{\lambda} \exp \left[ -\frac{1}{2\sigma^2} \left( \ln \frac{\lambda}{\text{mode}(\lambda)} - \sigma^2 \right)^2 \right], \quad (13)$$

normalized at  $\lambda = 5000 \text{ \AA}$ , with  $\text{mode}(\lambda)$  in the range 1000 to 5500  $\text{\AA}$  and  $\sigma$  in the range 0.35 to 0.9. The tem-

plates extend to 15000  $\text{\AA}$  with 100  $\text{\AA}$  resolution. These eight templates (hereafter N8) can be seen together with their original training sets in Figure 3.

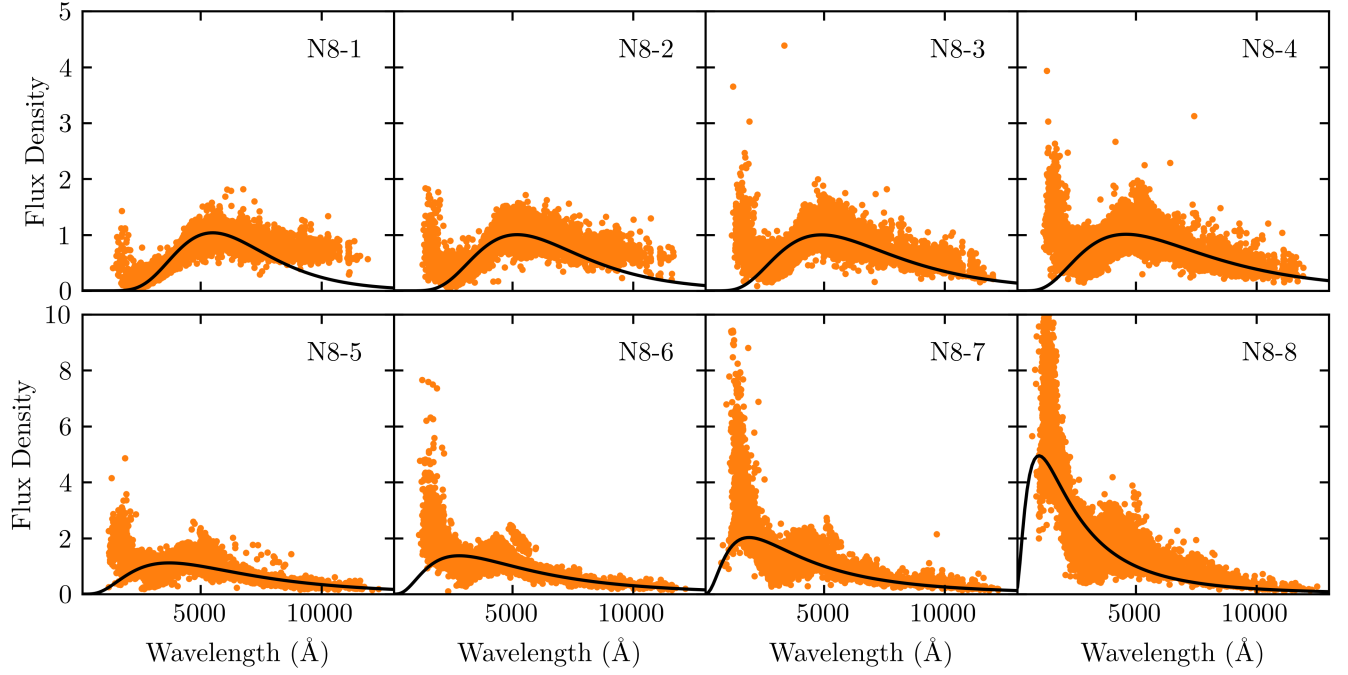
The training algorithm with  $w = 0.5$  is applied to the N8 templates. The convergence of the templates is evaluated via the weighted mean square error,

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

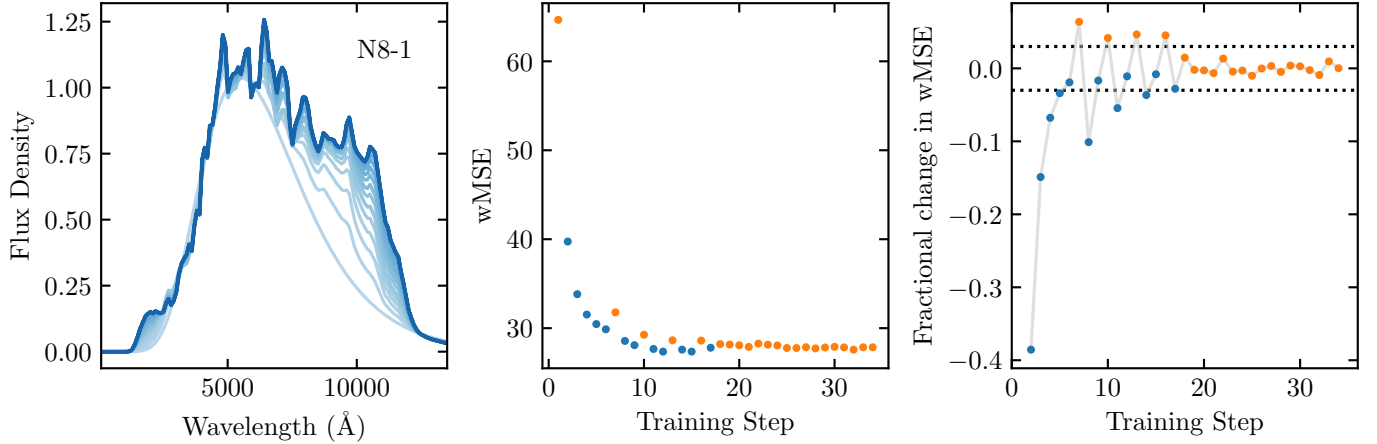
Each template is perturbed until the change in wMSE is less than 3%. When every template has converged to its current photometry set, new photometry sets are generated. Only those templates whose new photometry sets result in a greater than 3% change in wMSE resume perturbation with their new sets. This 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 most accurately 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 22 iterations, and takes approximately 75 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  $\text{\AA}$ , although this was essentially already present in the initial templates. In addition, there are now at a much higher resolution than the broadband filters used for photometry, some of which are labeled with gray lines in Figure 5. Template



**Figure 3.** The 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 N8-1 template is iteratively perturbed to better represent the colors of its photometry set. The final 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  $\pm 0.03$  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.



N8-1 displays Mg and Na absorption lines and template N8-4 contains the beginnings of  $H\alpha$  and  $H\beta$  emission lines. Templates N8-6, N8-7, and N8-8 contain what appear to be  $H\alpha$ ,  $H\beta$ ,  $H\gamma$ ,  $H\delta$ , OII, and OIII 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 the one of the defining features of our method.

In addition to these eight templates, we also train a set of 16 templates from the same range of parameters for the log-normal distribution, creating a more gradual transition of the templates from red to blue. Training for this template set (hereafter N16) took 150 minutes over 24 iterations. The results of the training can be seen in Figure 6. These results closely resemble the N8 results, with the same spectral features emerging. However, the N16 set shows a more gradual transition from red to blue.

In addition to starting from naive templates, one can start with templates derived from spectral synthesis models or observations of local galaxy spectra. 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 210 minutes over 62 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 structure that was originally present in the Im, SB3, and SB2 templates have been decreased in magnitude, while additional structure has been added to the simulated 25Myr and 5Myr templates what 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

**Table 3.** The emission lines reconstructed in the N8-6, N8-7, and N8-8 templates. For each line, we list the amplitude relative to  $H\beta$ ,  $r$ , and the effective width,  $W_\lambda$ .

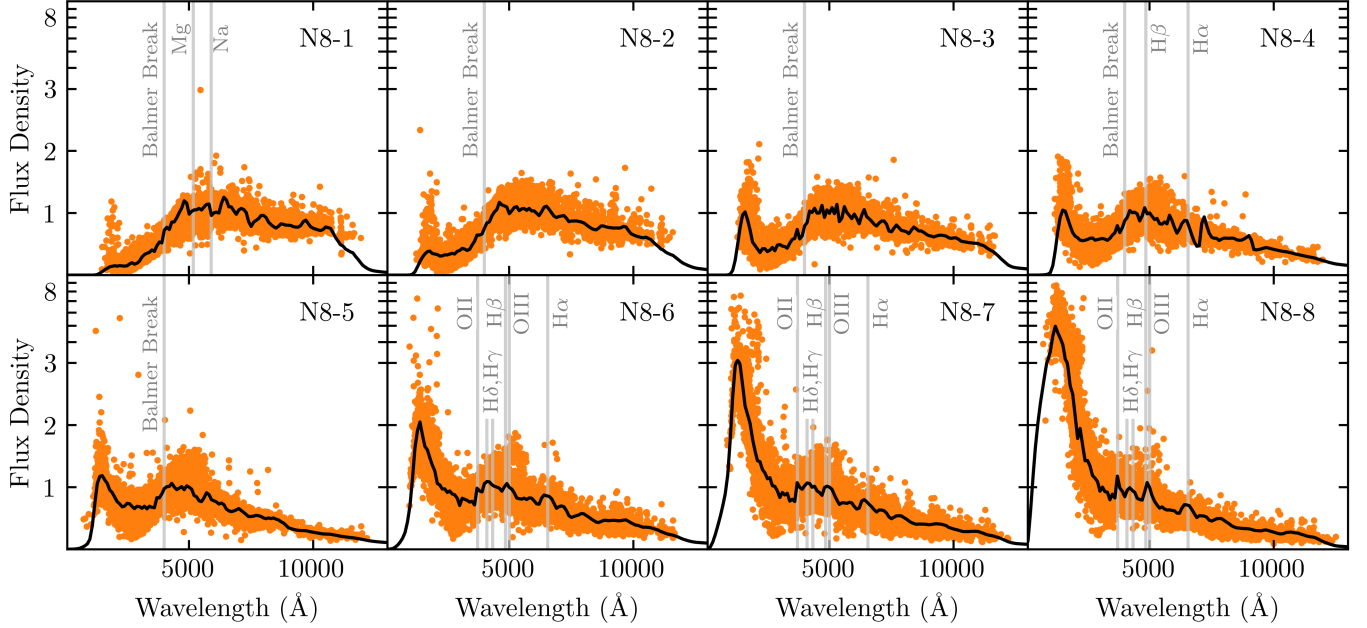
		N8-6		N8-7		N8-8	
	$\lambda$	$r$	$W_\lambda$	$r$	$W_\lambda$	$r$	$W_\lambda$
$H\alpha$	6563	2.86	132.7	2.86	103.3	2.86	115.2
$H\beta$	4861	1.00	32.9	1.00	26.4	1.00	30.3
$H\gamma$	4340	1.18	36.5	1.31	31.6	1.28	37.1
$H\delta$	4102	0.65	19.6	0.72	16.7	0.71	20.7
OII	3727	2.04	58.1	1.27	32.0	0.74	24.4
OIII	5007	2.08	68.0	2.42	66.1	0.86	27.3

resolution photometry due to the oversampling of the underlying SED templates. This includes the emergence of spectral lines in many of the templates (c.f. Figures 5, 6, and 7). Knowledge of these lines allows us to perform post-processing of the learned templates to deconvolve the lines from the broadband filters. Here we perform a simple post-processing 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  $H\beta$  line is impossible to distinguish from the OIII line in our templates because they are so close to one another. The same is true for the  $H\gamma$  and  $H\delta$  lines. To overcome this difficulty, we use the Balmer decrements of  $10^4\text{K}$  SDSS galaxies from Groves et al. (2012):  $H\alpha/H\beta = 2.86$  and  $H\gamma/H\delta = 1.81$ . We calculate the  $H\beta$  flux from  $H\alpha$ , and subtract this from the combined  $H\beta$ -OIII flux, and we calculate  $H\gamma$  and  $H\delta$  from the combined  $H\gamma$ - $H\delta$  flux.

After calculating the flux of the emission lines, the final templates are built by adding Gaussians of equivalent amplitude and  $\text{FWHM} = 20 \text{ Å}$  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  $H\beta$ , and the effective width,  $W_\lambda = \int (1 - F_\lambda/F_0) d\lambda$ , where  $F_\lambda$  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  $H\gamma$  and  $H\delta$  lines relative to  $H\beta$  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 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



**Figure 5.** The 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 H $\alpha$ , H $\beta$ , H $\gamma$ , H $\delta$ , OII, and OIII emission lines are labeled in gray.

20,496 galaxies (20% of the total set), with mean redshift  $z_{\text{mean}} = 0.69$ , max redshift  $z_{\text{max}} = 3.61$ , 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; Benítez 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. Each set of observed fluxes is then matched to a specific template and redshift determined to be the most likely to have produced the observed colors.

For each template, BPZ evaluates a  $\chi^2$  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  $\sigma_n$  denote the calculated flux, the observed flux, and the fractional error as in Equation 4. 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 set of

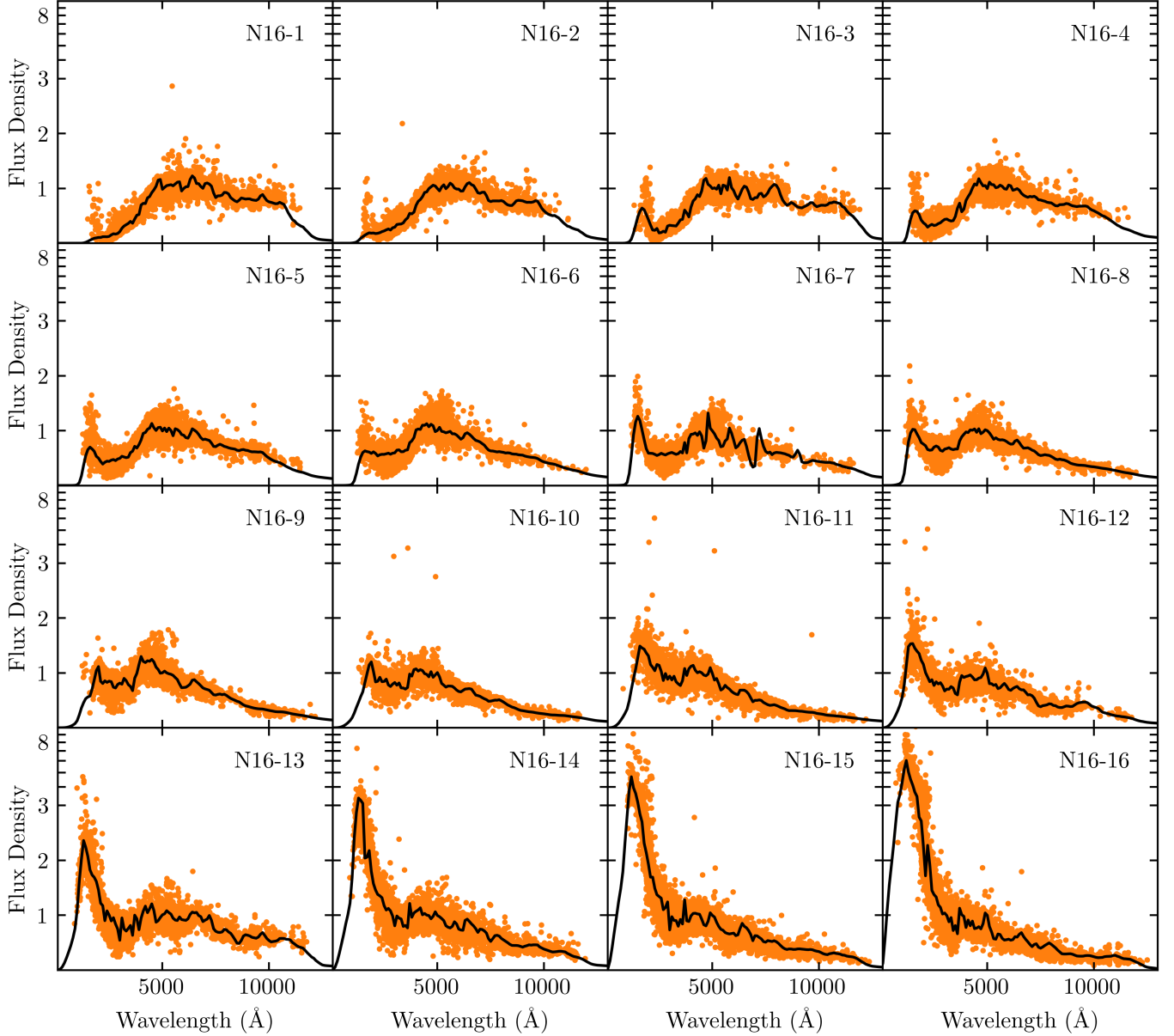
templates:

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

where  $p(z, T|m_0) = p(T|m_0)p(z|T, m_0)$  is a Bayesian 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 peak of the posterior distribution is used to estimate the photo- $z$ .

We use BPZ-v1.99.3<sup>3</sup> (Benítez 2000) to estimate photo- $z$ 's, providing the various sets of templates described in Section 4. We linearly interpolate two templates between each basis template, sorted by rest  $u - g$  color, by setting INTERP=2. We use the default BPZ prior, which requires each SED template be broadly classified as either elliptical, spiral, or irregular/starburst. The SED classifications for each template set are discussed in the next section. We use the one of the  $i$  bands for the magnitude prior, in the following order of priority:  $i$ ,  $i_2$ ,  $I$ ,  $i^+$ . For simplicity, we treat non-detections as non-observations. All other settings were left as default.

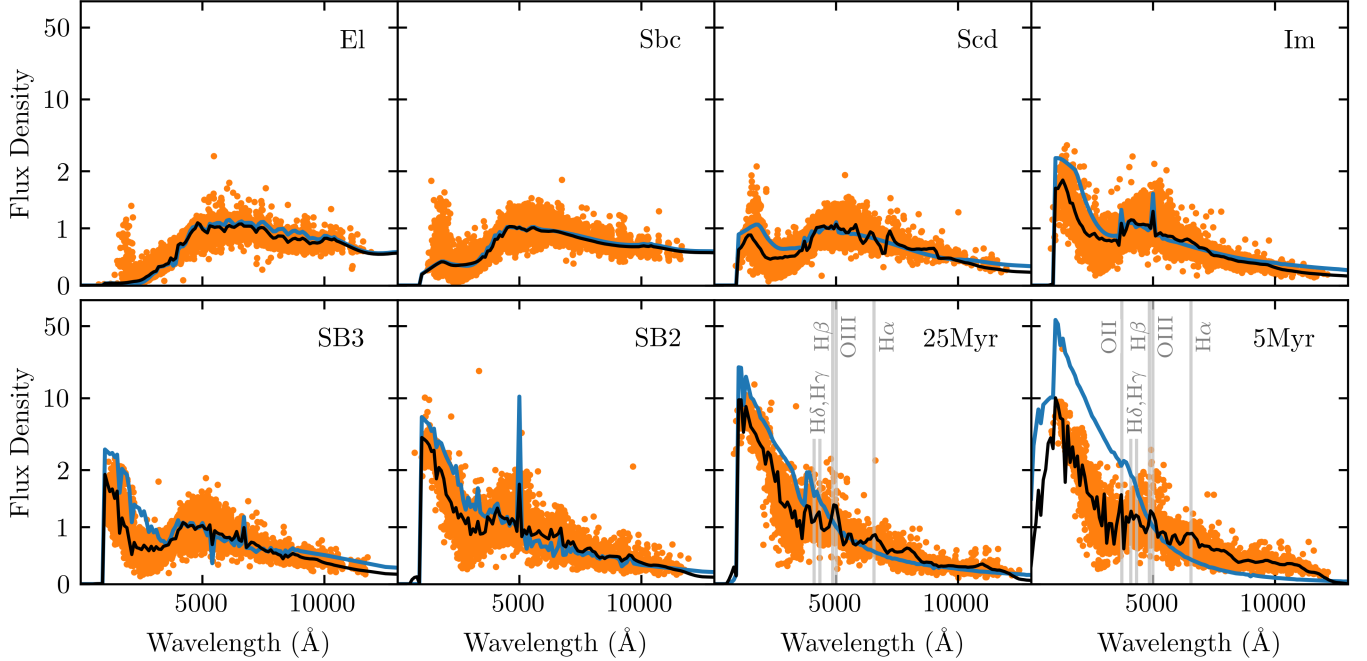
<sup>3</sup> <http://www.stsci.edu/~dcoe/BPZ/>



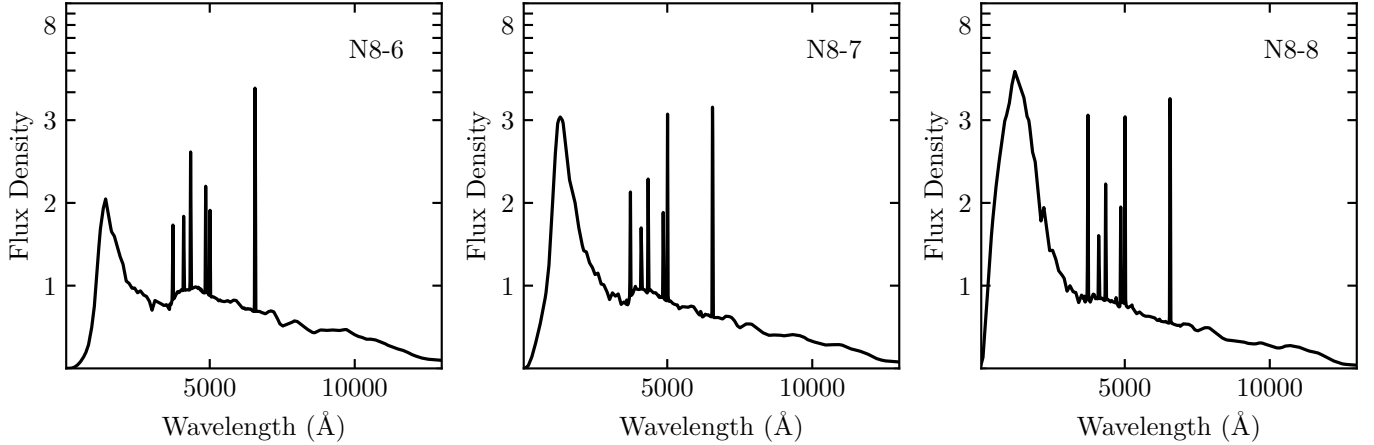
**Figure 6.** The trained N16 templates (black lines) with their final photometry sets (orange points). N16-1 is the reddest template, with each successive template getting bluer. These templates closely resemble the N8 templates and have show the same emerging spectral features (c.f. Figure 5), but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

BPZ provides two metrics for the photo- $z$  estimates: ODDS and  $\chi^2_{\text{mod}}$ . 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.  $\chi^2_{\text{mod}}$  measures how well the best fit template at the predicted redshift matches the observed fluxes. For more about these metrics, see Section 4 of Benítez (2000) and Section 4.3 of Coe et al. (2006). In this work, photo- $z$  estimates with ODDS < 0.95 or  $\chi^2_{\text{mod}} > 1$  are excluded from the analysis, and the fraction excluded on this bases is reported as  $f_{\text{cut}}$ .

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 1000Å Lyman break in a high redshift galaxy spectrum is shifted to the position of the 4000Å 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.



**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 25Myr and 5Myr 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 (cf. Figure 5). The emission lines, left to right, are OII, H $\delta$ , H $\gamma$ , H $\beta$ , OIII, and H $\alpha$ . The wavelengths, relative amplitudes, and effective widths of these lines are in Table 3.

To address this issue, we evaluate the statistics of the interquartile range (IQR) of the data, as these measures are known to be robust to the presence of outliers. We follow [Graham et al. \(2018\)](#) in introducing the quantity  $\Delta z_{1+z} = (z_{\text{spec}} - z_{\text{phot}})/(1 + z_{\text{phot}})$ . 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,  $\sigma_{\text{IQR}}$ , as the width of the IQR in  $\Delta z_{1+z}$ , divided by 1.349 to convert to the equivalent of a Gaussian standard devi-

ation. We define the bias of the photo- $z$  estimates as the mean value of  $\Delta z_{1+z}$  for galaxies within the IQR. The uncertainties of these two values are bootstrapped by calculating the values on 1000 random samples with replacement. Outliers are identified as photo- $z$ 's with  $\Delta z_{1+z} > 3\sigma_{\text{IQR}}$ , and the fraction of outliers is reported as  $f_{\text{out}}$ .

## 5.2. Photo- $z$ Results

Photo- $z$ 's were estimated for the test set using BPZ with the original CWW+SB4 templates, the trained CWW+SB4 templates, and the trained N8 and N16 templates. The Bayesian prior used by BPZ requires each of the SED templates be broadly classified as either elliptical, spiral, or irregular/starburst. Each of our new templates are assigned the type 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, three spiral, and four irregular/starburst galaxies. The N16 templates are determined to have two elliptical, five spiral, and nine irregular/starburst galaxies.

The photo- $z$  results can be seen in Figure 9. The photo- $z$  estimates that passed the cuts on ODDS and  $\chi^2_{\text{mod}}$  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 towards higher photo- $z$ 's. Reduced photo- $z$  accuracy is generally expected for spec- $z$ 's greater than 1.5, as the Balmer break leaves most band sets at around  $z = 1.4$  and the Lyman break does not enter most band sets until  $z = 2.5$ .

By comparing the top two panels, one can see that the training algorithm decreases the fraction cut, the bias, and scatter of the photo- $z$  results for the CWW+SB4 templates, but does not improve the outlier fraction. The bottom two panels show that similar results are obtained with the N8 and N16 template sets, demonstrating that this method can be used to generate photo- $z$  templates without any a priori information about galaxy spectra. The trained N8 and N16 templates have a better cut fraction, better bias, and better scatter than the original CWW+SB4, but have worse outlier fractions. It is clear that our training method improves photo- $z$  estimates for the galaxies with successful photo- $z$  estimates, but does not help resolve the degeneracies that lead to catastrophic outliers.

The value of the metrics as a function of photo- $z$  can be seen in Figure 10. 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  $< 0.02(1 + z_{\text{phot}})$ ;  $f_{\text{out}} < 10\%$ ; average bias  $< 0.003(1 + z_{\text{phot}})$ . The SRD lists these requirements for an  $i < 25$ , magnitude-limited sample of four billion galaxies from  $0.3 < z < 3.0$ . 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 10, you can see that for redshifts  $0.3 < z < 3.0$  we are able to achieve an appropriate outlier fraction, and that our training algorithm makes great progress on the bias and modest progress on scatter.

**Section about the impact of template number.**

## 6. CONCLUSIONS

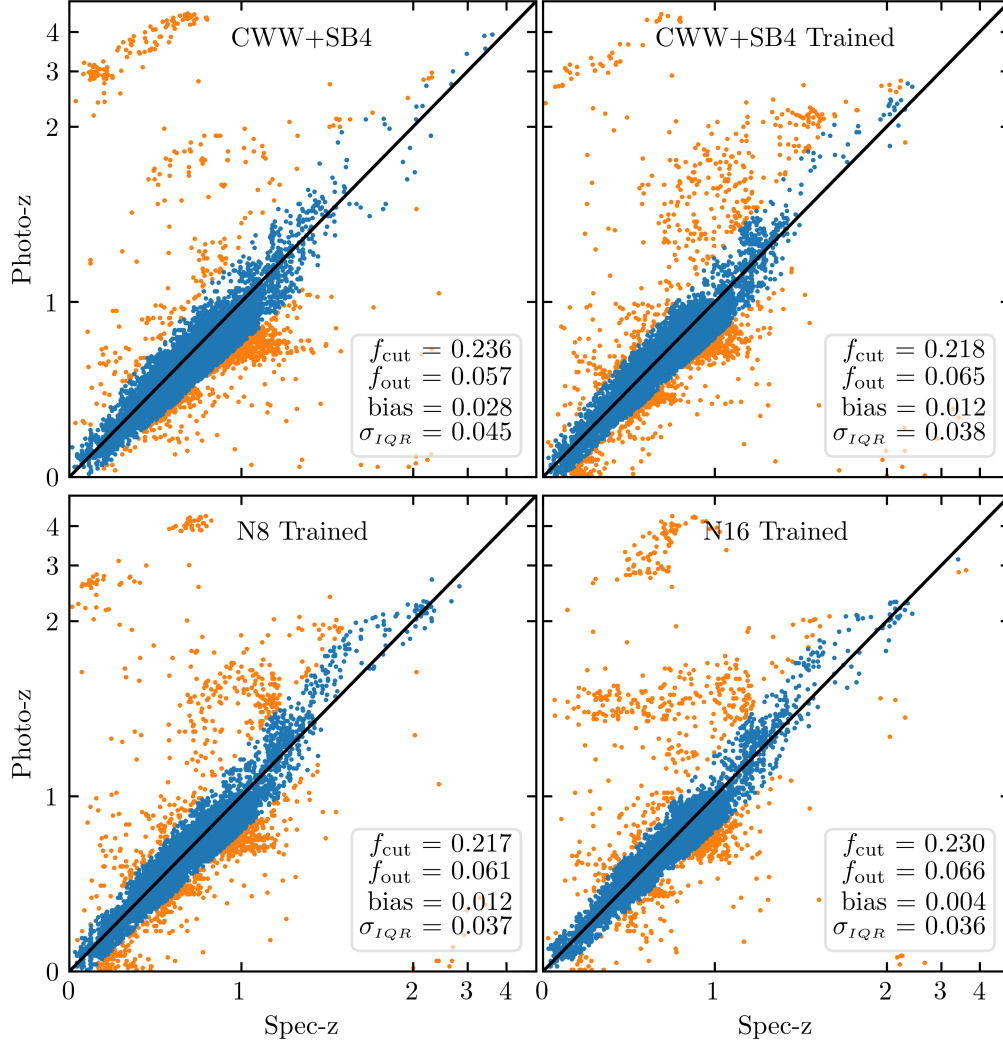
## ACKNOWLEDGMENTS

The authors would like to thank Bryce Kalmbach for providing advice in early stages of this work, Sam Schmidt for his help with BPZ and Melissa Graham for her code to calculate photo- $z$  statistics. This work was supported by the U.S. Department of Energy, Office of Science, under Award Number

This research is based on observations made with ESO Telescopes at the La Silla or Paranal Observatories under programme ID(s) 175.A-0839(B), 175.A-0839(D), 175.A-0839(I), 175.A-0839(J), 175.A-0839(H), 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 from 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).

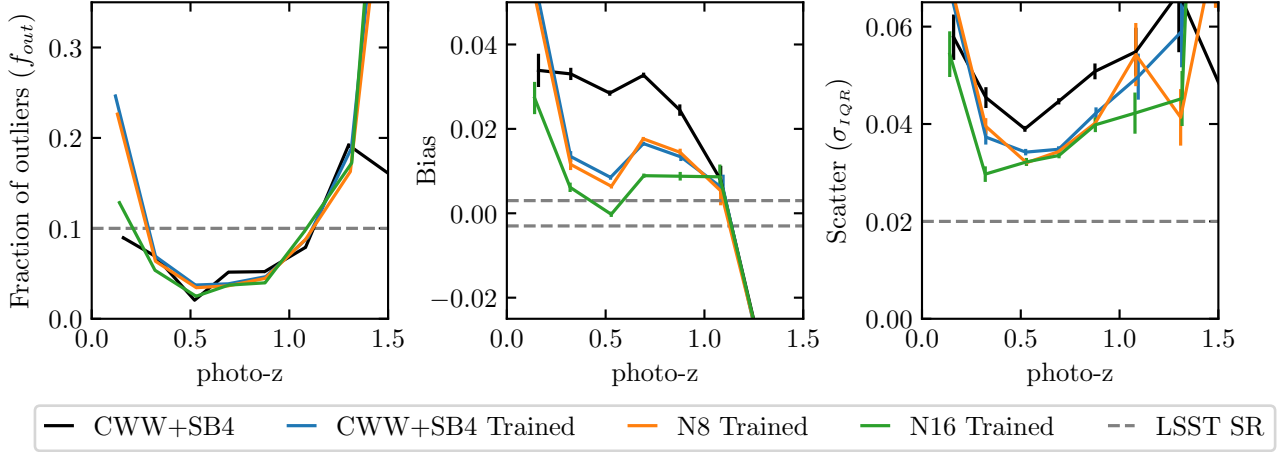




**Figure 9.** Results of photo- $z$  estimation with BPZ, using the four different templates sets. Photo- $z$  estimates are displayed as points: inliers are blue and outliers are orange. The black line represents perfect estimation (i.e. photo- $z$  = spec- $z$ ). The statistics printed in each panel are for the entire data set.

## REFERENCES

- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, Measuring and modelling the redshift evolution of clustering: The Hubble Deep Field North, Tech. Rep. 2, doi: [10.1046/j.1365-8711.1999.02978.x](https://doi.org/10.1046/j.1365-8711.1999.02978.x)
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *Astronomy and Astrophysics*, 558, 33, doi: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068)
- Benítez, N. 2000, *The Astrophysical Journal*, 536, 571, doi: [10.1086/308947](https://doi.org/10.1086/308947)
- Benítez, N., Ford, H., Bouwens, R., et al. 2004, *The Astrophysical Journal Supplement Series*, 150, 1, doi: [10.1086/380120](https://doi.org/10.1086/380120)
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *The Astrophysical Journal*, 686, 1503, doi: [10.1086/591786](https://doi.org/10.1086/591786)
- Bruzual, G., & Charlot, S. 2003, *Monthly Notices of the Royal Astronomical Society*, 344, 1000, doi: [10.1046/j.1365-8711.2003.06897.x](https://doi.org/10.1046/j.1365-8711.2003.06897.x)
- Bruzual A., G., & Charlot, S. 1993, *The Astrophysical Journal*, 405, 538, doi: [10.1086/172385](https://doi.org/10.1086/172385)
- Budavári, T., Szalay, A. S., Connolly, A. J., Csabai, I., & Dickinson, M. 2000, *The Astronomical Journal*, 120, 1588, doi: [10.1086/301514](https://doi.org/10.1086/301514)
- Coe, D., Benítez, N., Sánchez, S. F., et al. 2006, *The Astronomical Journal*, 132, 926, doi: [10.1086/505530](https://doi.org/10.1086/505530)



**Figure 10.** Photo- $z$  metrics as a function of redshift bin. LSST science requirements are displayed as dashed gray lines.

- Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, The Astrophysical Journal Supplement Series, 43, 393, doi: [10.1086/190674](https://doi.org/10.1086/190674)
- Cooper, M. C., Aird, J. A., Coil, A. L., et al. 2011, Astrophysical Journal, Supplement Series, 193, 14, doi: [10.1088/0067-0049/193/1/14](https://doi.org/10.1088/0067-0049/193/1/14)
- Csabai, I., Connolly, A. J., Szalay, A. S., & Budavári, T. 2000, Reconstructing Galaxy Spectral Energy Distributions from Broadband Photometry, Tech. Rep. 1, doi: [10.1086/301159](https://doi.org/10.1086/301159)
- de Jong, J. T., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, Experimental Astronomy, 35, 25, doi: [10.1007/s10686-012-9306-1](https://doi.org/10.1007/s10686-012-9306-1)
- Fruchter, A., & Hook, R. 2002, Drizzle: A Method for the Linear Reconstruction of Undersampled Images, Tech. Rep. 792, doi: [10.1086/338393](https://doi.org/10.1086/338393)
- Graham, M. L., Connolly, A. J., Ivezić, Ž., et al. 2018, The Astronomical Journal, 155, 1, doi: [10.3847/1538-3881/aa99d4](https://doi.org/10.3847/1538-3881/aa99d4)
- Green, J., Schechter, P., Baltay, C., et al. 2012, arXiv e-prints. <https://arxiv.org/abs/1208.4012>
- Groves, B., Brinchmann, J., & Walcher, C. J. 2012, Monthly Notices of the Royal Astronomical Society, 419, 1402, doi: [10.1111/j.1365-2966.2011.19796.x](https://doi.org/10.1111/j.1365-2966.2011.19796.x)
- Hudelot, P., Cuillandre, J.-C., Withington, K., et al. 2012, VizieR Online Data Catalog, II/317. <http://adsabs.harvard.edu/abs/2012yCat.2317...0H>
- Hunter, J. D. 2007, Computing in Science and Engineering, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Ilbert, O., Capak, P., Salvato, M., et al. 2009, Astrophysical Journal, 690, 1236, doi: [10.1088/0004-637X/690/2/1236](https://doi.org/10.1088/0004-637X/690/2/1236)
- Ivezić, Ž., & LSST Science Collaboration. 2018, LSST Project Management LPM-17. <http://ls.st/srd>
- Izbicki, R., & Lee, A. B. 2017, Electronic Journal of Statistics, 11, 2800, doi: [10.1214/17-EJS1302](https://doi.org/10.1214/17-EJS1302)
- Kalmbach, J. B., & Connolly, A. J. 2017, The Astronomical Journal, 154, 277, doi: [10.3847/1538-3881/aa9933](https://doi.org/10.3847/1538-3881/aa9933)
- Kind, M. C., & Brunner, R. J. 2013, Monthly Notices of the Royal Astronomical Society, 432, 1483, doi: [10.1093/mnras/stt574](https://doi.org/10.1093/mnras/stt574)
- Kinney, A. L., Calzetti, D., Bohlin, R. C., et al. 1996, The Astrophysical Journal, 467, 38, doi: [10.1086/177583](https://doi.org/10.1086/177583)
- 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 (IOS Press), 87–90, doi: [10.3233/978-1-61499-649-1-87](https://doi.org/10.3233/978-1-61499-649-1-87)
- Le Fèvre, O., Mellier, Y., McCracken, H. J., et al. 2004, Astronomy and Astrophysics, 417, 839, doi: [10.1051/0004-6361:20031767](https://doi.org/10.1051/0004-6361:20031767)
- Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, Astronomy and Astrophysics, 559, 14, doi: [10.1051/0004-6361/201322179](https://doi.org/10.1051/0004-6361/201322179)
- Lee, M. A., Budavári, T., Sullivan, I. S., & Connolly, A. J. 2019, The Astronomical Journal, 157, 182, doi: [10.3847/1538-3881/ab139f](https://doi.org/10.3847/1538-3881/ab139f)
- Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, Astrophysical Journal, Supplement Series, 184, 218, doi: [10.1088/0067-0049/184/2/218](https://doi.org/10.1088/0067-0049/184/2/218)
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009. <https://arxiv.org/abs/0912.0201>
- Martin, D. C., Fanston, J., Schiminovich, D., et al. 2005, The Galaxy Evolution Explorer : A Space Ultraviolet Survey Mission , Tech. Rep. 1, doi: [10.1086/426387](https://doi.org/10.1086/426387)
- Miyazaki, S., Komiyama, Y., Sekiguchi, M., et al. 2002, Subaru prime focus camera - Suprime-Cam, Tech. Rep. 6, doi: [10.1093/pasj/54.6.833](https://doi.org/10.1093/pasj/54.6.833)

- Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, *The Astrophysical Journal Supplement Series*, 225, 27, doi: [10.3847/0067-0049/225/2/27](https://doi.org/10.3847/0067-0049/225/2/27)
- Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *Astrophysical Journal, Supplement Series*, 208, 57, doi: [10.1088/0067-0049/208/1/5](https://doi.org/10.1088/0067-0049/208/1/5)
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825. <https://arxiv.org/abs/1201.0490>
- Salvato, M., Ilbert, O., & Hoyle, B. 2019, *Nature Astronomy*, 3, 212, doi: [10.1038/s41550-018-0478-0](https://doi.org/10.1038/s41550-018-0478-0)
- Schmidt, S. J., Malz, A. I., Soo, J. Y. H., et al. 2020, Evaluation of probabilistic photometric redshift estimation approaches for LSST, Tech. rep. <https://arxiv.org/abs/2001.03621v1>
- Scodreggio, M., Guzzo, L., Garilli, B., et al. 2018, *Astronomy and Astrophysics*, 609, doi: [10.1051/0004-6361/201630114](https://doi.org/10.1051/0004-6361/201630114)
- The Dark Energy Survey Collaboration. 2005, arXiv e-prints, astro. <https://arxiv.org/abs/0510346>
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in Science and Engineering*, 13, 22, doi: [10.1109/MCSE.2011.37](https://doi.org/10.1109/MCSE.2011.37)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- Zhou, R., Cooper, M. C., Newman, J. A., et al. 2019, *Monthly Notices of the Royal Astronomical Society*, 488, 4565, doi: [10.1093/mnras/stz1866](https://doi.org/10.1093/mnras/stz1866)