Draft version May 5, 2020
Typeset using LaTeX **twocolumn** style in AASTeX63

Learning Spectral Templates for Photometric Redshift Estimation from Broadband Photometry

John Franklin Crenshaw[1] and Andrew J. Connolly[2]

[1]*Department of Physics, University of Washington*
[2]*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

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 will 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 will 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, I will discuss a heuristic for selecting hyperparameters for the training.

Work in analogy to Drizzle (Fruchter & Hook 2002) and to this image reconstruction paper (Lee et al. 2019).

Corresponding author: John Franklin Crenshaw
jfc20@uw.edu

2.1. *Generating Training 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

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

$S(\lambda)$ 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.

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 training sets containing many galaxies for each template.

Examining how the photometry set is divided into training sets is helpful in assembling the initial set of templates. The initial templates should be selected so that the different spectral shapes visible in the photometry data are sorted into different training sets. 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:

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

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

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

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 σ_n, the fractional error of the measured flux. The perturbed template fluxes can be calculated with a discretized version of Equation 3:

$$\hat{f}_n(\{\hat{s}_k\}) = \sum_k \hat{s}_k\, r^n_{k'} \Delta\lambda_{k'} \tag{5}$$

where r^n_k 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 Δ_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:

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

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:

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

The matrix M and vector ν are defined

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

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

Iterating the perturbations changes the shape of the template SED to better match the measured photometry. An example of this can be seen in Figure 1. Fluxes in the *ugrizY* filters listed in Table 2 were calculated for a starburst galaxy template at 1,000 random redshifts in the range z=0 to z=3. Starting with an $S(\lambda) = 0$ template SED, the perturbation algorithm is applied iteratively. After 100 iterations, the learned template closely matches the original template in the wavelength range for which photometry exists. While the learned template is a smoothed version of the original, high resolution features have been recovered, despite the relatively low resolution of the filters. Note that in practice, 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 shapes of the template SED's, so that re-running the algorithm to

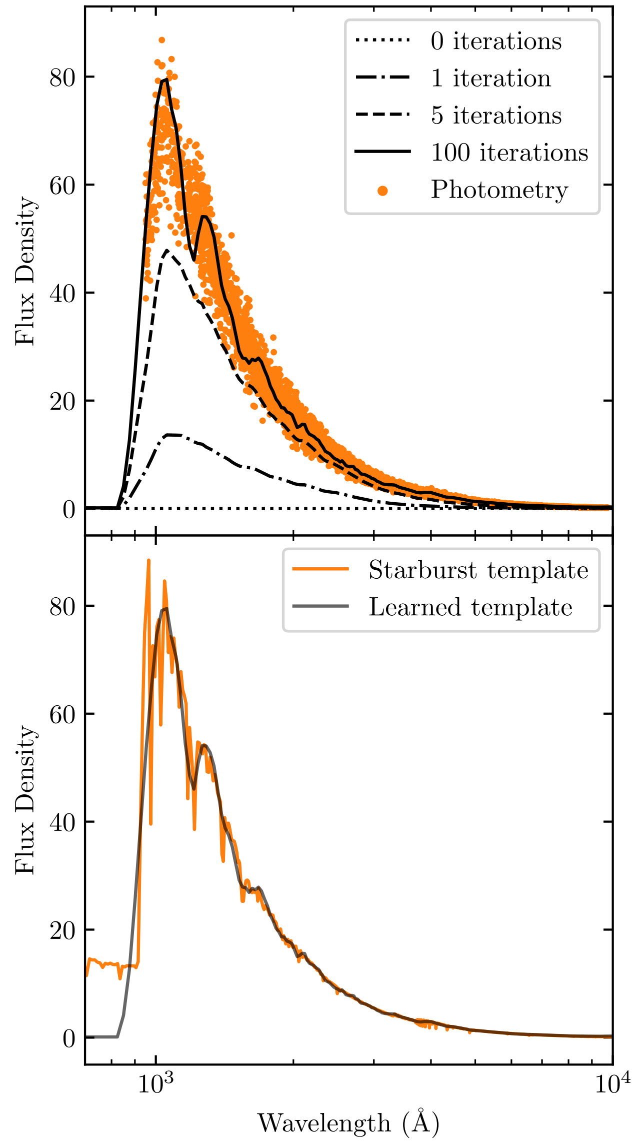

Figure 1. Perturbing a naive template, in this case a flat line, to better match a photometry set. Top: the orange points are simulated observations of the 5Myr starburst template from Coe et al. (2006) at 1,000 random redshifts in the range z=0 to z=3 using the *ugrizY* filters listed in Table 2. The 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.

generate training sets will now result in different sets than previously obtained. As better approximations of the true underlying SED templates are generated by the perturbation algorithm, the training set generation algorithm will do a better and better job at dividing the photometry data into training sets that characterize distinct spectral shapes. We iterate the full training algorithm (generate training sets, perturb the templates, re-generate training sets, perturb the templates again, etc), until the SED templates converge.

2.3. *Selecting Hyperparameters*

The training algorithm relies on the choice of four hyperparameters. The first is the number of templates. As discussed in Section 2.1, this choice can be made by using the training set generation 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 REF TEMPLATE # SECTION.

The final three hyperparameters control the template perturbation algorithm, namely the values of the Δ_k, the number of iterations ($N_{\rm iter}$), and the number of template perturbations per iteration ($N_{\rm pert}$). These three are interconnected and choosing them appropriately affects the stability and speed of the training algorithm. Here we present a heuristic for selecting the Δ_k. Discussion of N_iter and N_pert is deferred to Section 2.

The most important parameter to select is Δ_k, which controls the relative weighting of the regularization term in Equation 4. If this value is too large, training will be very slow and a large number of perturbation and training rounds will be required. If this value is too low, the training becomes unstable and the final templates will be over-fit.

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 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 4. 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 4) will vary by training set. To this end, we make the following approximation:

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

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

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

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

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

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 2 yields roughly the same results.

3. DATA

Need to update numbers and mention how only fluxes with SNR $>$ 20 and galaxies with at least 4 fluxes were kept. Also, I removed FUV because it's not broadband. Our data set consists of 111,797 galaxies with redshifts $z < 4.54$ and i-band magnitudes in the range $13.8 < i < 27.4$. This set is divided into a training set, from which SED templates are learned, and a test set, for which photo-z's are estimated using the learned templates. The redshift surveys in each set are distinct, but there is overlap in the broadband filters used to measure the photometry. The entire data set is summarized in Table 1, the filters used to measure photometry are listed in Table 2, and the redshift distributions are shown in Figure 2. The filter transmission were obtained from the Spanish Virtual Observatory (SVO) Filter Profile Service.

3.1. *Training Set*

The training set is the set of galaxies from which the SED templates are learned, according to the methods described in Section 2. It consists of 95,927 galaxies (86% of all galaxies) with redshifts $z < 4.54$ and i-band magnitudes in the range $13.8 < i < 26.8$. The training set is composed of three spec-z surveys, zCOSMOS-*bright* (Lilly et al. 2009), VVDS (Le Fèvre et al. 2013), and VIPERS (Scodeggio et al. 2018), each conducted with the VIMOS spectrograph mounted on the European Southern Observatory's (ESO) Very Large Telescope (VLT). The spec-z's are matched to broadband photometry from the GALEX, CFHT (Megacam and CFH12k), Subaru, and UKIRT telescopes (see Table 2). Cite Martin when I cite GALEX.

3.1.1. *zCOSMOS-bright*

zCOSMOS (Lilly et al. 2009) is a redshift survey of 1.7 deg^2 of the COSMOS field, 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 only use galaxies in the recommended sample described in the ESO data release description[1], 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 broadband photometry from Ilbert et al. (2009). The photometry is measured from the ultraviolent to the near-infrared in 12 broadband filters: FUV and NUV on GALEX, u and i on CFHT using Megacam, B and V on CFHT using CFH12k, g^+, r^+, i^+, and z^+ on Subaru, and J on UKIRT (details in Table 2). We use only galaxies that were not masked in any of the optical bands. The final set consists of 14,311 galaxies with redshifts $z < 2.52$ and i-band magnitudes in the range $16.9 < i < 24.3$.

3.1.2. *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 $17.5 < i < 24$; 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. Photometry was obtained in the u, g, r, i, z and B, V, R, I bands from CFHT. See Table 2 for details. The final set contains 9,665 galaxies out to redshifts $z < 4.5$, with magnitudes $13.8 < i < 20.9$. Cite CFH12K camera at CFHT photometry from LeFevre 2004 and McCracken.

3.1.3. *VIPERS*

The VIMOS Public Extragalactic Redshift Survey (VIPERS, Scodeggio et al. 2018) is a dense, large-volume redshift survey, focusing on redshifts $0.5 < z < 1.2$. The survey includes 76,552 galaxies with redshifts reliable at a 95% confidence level. We keep only galaxies with `photoMask` and `spectroMask` = 1. The final set contains 71,951 galaxies with redshifts < 2.15 and magnitudes $17.7 < i < 23.3$. VIPERS photometry are measured in FUV and NUV on GALEX, and $u, g, r, i2, i, z$ on CFHT Megacam. Note that i_2 is the replacement i-band filter on Megacam.

3.2. *Test Set*

The test set is the set of galaxies for which photo-z's are estimated with `BPZ`, as described in Section 5.1, using the SED templates learned from the training set. It consists of 15,650 galaxies (14% of all galaxies) with redshifts $z < 3.32$ and i-band magnitudes in the range $15.3 < i < 27.4$ The test set is composed of broadband photometry and spec-z's from the DEEP2/3 and 3D-HST surveys compiled in Zhou et al. (2019).

Paragraph about the Zhou stuff. (Zhou et al. 2019; Newman et al. 2013; Momcheva et al. 2016). Need to write this.

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

Table 1. Summary of the redshift and photometric data sets. $N_{\rm gal}$ is the total number of galaxies in the set, $f_{\rm gal}$ is the fraction of galaxies in the set, and $\bar{\sigma}_i$ is the mean fractional flux error for the i-band photometry. NOTE: Need to update this after imposing SNR $>$ 20.

Data Set	$N_{\rm gal}$	$f_{\rm gal}$	$z_{\rm mean}$	$z_{\rm max}$	i-band range	$i_{\rm mean}$	$\bar{\sigma}_i$	Reference
zCOSMOS	14311	0.13	0.57	2.52	$16.87 \leq i \leq 24.29$	21.19	0.06	Lilly et al. (2009)
VVDS	9665	0.09	0.64	4.54	$13.84 \leq i \leq 26.75$	20.95	0.03	Le Fèvre et al. (2013)
VIPERS	71951	0.64	0.70	2.15	$17.66 \leq i \leq 23.25$	21.41	0.02	Scodeggio et al. (2018)
DEEP2/3	14029	0.13	0.73	1.98	$15.30 \leq i \leq 25.36$	21.57	0.04	Zhou et al. (2019); Newman et al. (2013)
3D-HST	1621	0.01	1.50	3.32	$19.10 \leq i \leq 27.41$	23.83	0.04	Zhou et al. (2019); Momcheva et al. (2016)
Training	89261	0.80	0.70	4.54	$13.84 \leq i \leq 26.75$	27.41	0.03	
Test	22316	0.20	0.69	3.74	$16.46 \leq i \leq 27.41$	26.78	0.03	
Total	111577	1.00	0.70	4.54	$13.84 \leq i \leq 27.41$	21.36	0.03	

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, $W_{\rm eff} = {\rm Max}[R(\lambda)]^{-1}$, are given in angstroms. Filters are listed in order of increasing λ_0.

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

4. APPLICATION TO DATA

Start by describing the training and test sets here. Mention their redshift and magnitude ranges. Calculate the min and max wavelength of the training set, and say that we will try to reconstruct spectra over that wavelength range.

4.1. *The Training and Test Set*

Move discussion of the training and test set here.

4.2. *Training Templates on the Data*

Eight naive templates were chosen to represent the underlying SED shapes of the data set according to the principles described at the end of Section 2.1. They are "naive" because they are just chosen by eye to roughly split the photometry into groups by shape. Each is a log-normal distribution,

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

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

These eight templates were chosen to approximately separate the galaxy photometry into distinct groups based on spectral shape, and to ensure that each template had a sizeable and well distributed training set. While this requires some guess and check with the assembled training sets, the strength of this method is that you do not need a priori information of galaxy spectra.

After the training sets are assembled, outliers are removed by means of an Isolation Forest in the flux-wavelength space. Outliers are determined on a per-flux basis instead of per-galaxy. Removing outliers before training the templates helps stabilize the results of the perturbation algorithm described in the next section.

The training algorithm is applied to the N8 templates, resulting in the final templates seen in Figure 4. The templates were trained for five rounds, with each round consisting of three perturbations with $w = 0.5$. After the training, the templates more closely resemble real galaxy spectra. You can see there are now features in the templates at a much higher resolution than the broadband filters. Add lines to guide the eye for 400nm break and spectral lines, where they are visible. Mention those here as well.

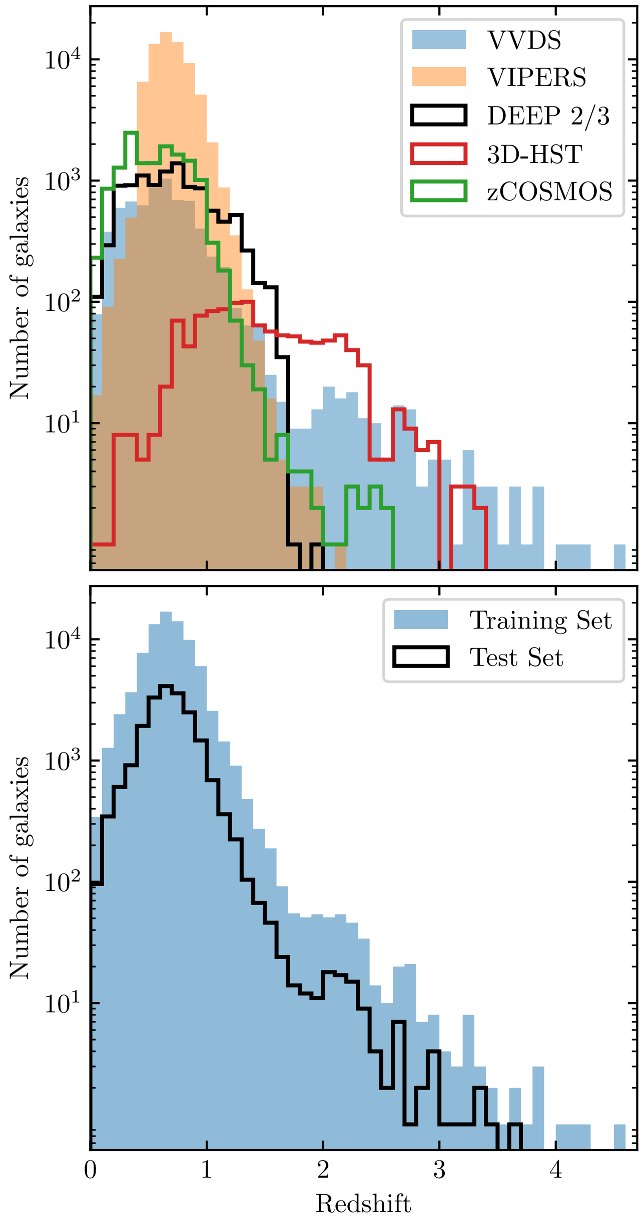

Figure 2. Redshift distribution of the galaxy surveys. The top and bottom panels show the redshifts of the training and test sets respectively, including the constituent surveys.

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. This template set (hereafter N16) can be seen with the final training sets in Figure 5. These were again trained for five rounds, with each round consisting of three perturbations with $w = 0.5$.

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 that comes with `BPZ`. 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.

These templates were trained for five rounds, with each round consisting of one perturbation with $w = 2$. The results of the training can be seen together with the original CWW+SB4 templates and the final training sets in Figure 6. Say something about how the training added additional high resolution structure to the templates and added a red tilt to Im, SB2, 25Myr, and 5Myr.

5. ESTIMATING PHOTO-Z'S

5.1. *Bayesian Photometric Redshift*

Bayesian Photometric Redshift (`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 the 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 observations.

For each template, `BPZ` evaluates a χ^2 function at each redshift on the grid:

$$\chi^2(z, T, A) = \sum_n \frac{1}{\sigma_n^2} \left(A\, \hat{f}_n(z, T) - f_n \right)^2, \tag{14}$$

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 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: $p(\{f_n\}|z, T) \propto \exp\left(-\chi^2/2\right)$. The redshift posterior is then calculated by marginalizing over the set of templates:

$$\begin{aligned} p(z|\{f_n\}, m_0) &= \sum_T p(z, T|\{f_n\}, m_0) \\ &\propto \sum_T p(z, T|m_0)\, p(\{f_n\}|z, T), \end{aligned} \tag{15}$$

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

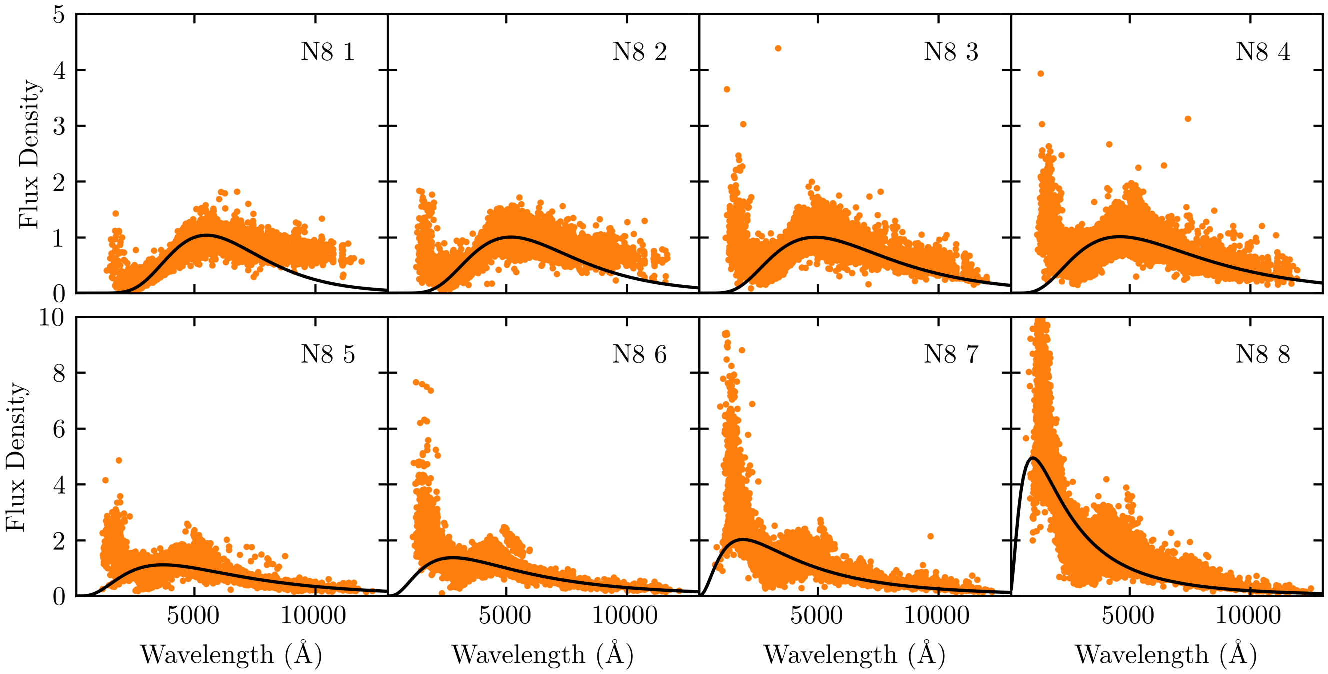

Figure 3. The untrained N8 templates (black lines) with their original training sets (orange points). N8 1 is the reddest template, with each successive template getting bluer.

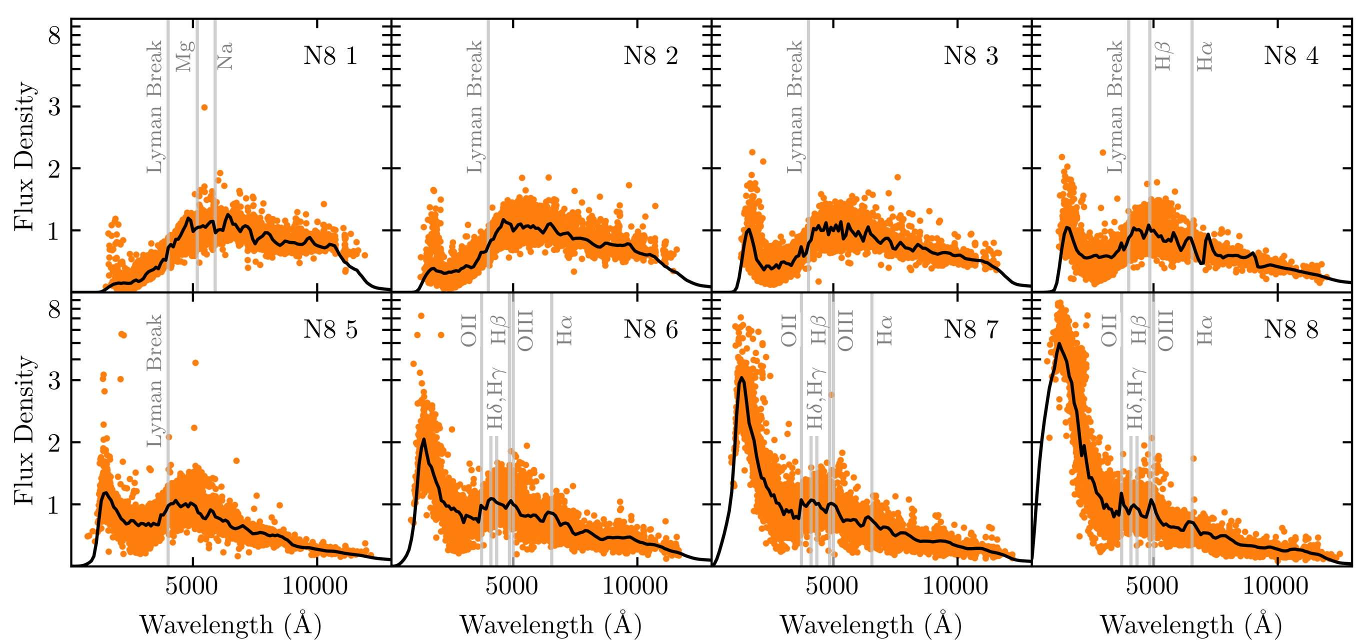

Figure 4. The trained N8 templates (black lines) with their final training sets (orange points). N8 1 is the reddest template, with each successive template getting bluer. Say something about the lines added to guide the eye to spectral features.

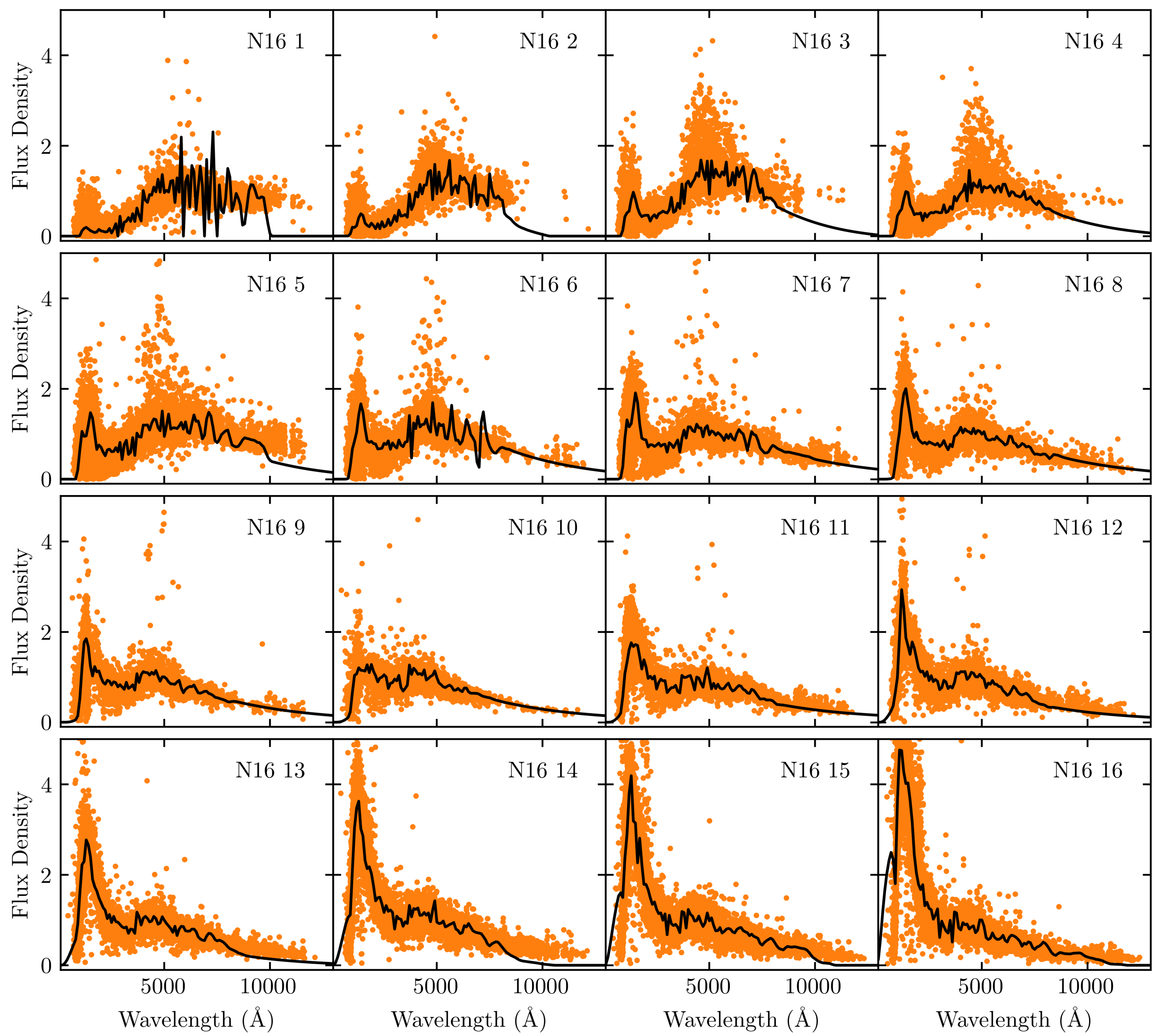

Figure 5. The trained N16 templates (black lines) with their final training sets (orange points). N16 1 is the reddest template, with each successive template getting bluer. Say something about the lines added to guide the eye to spectral features.

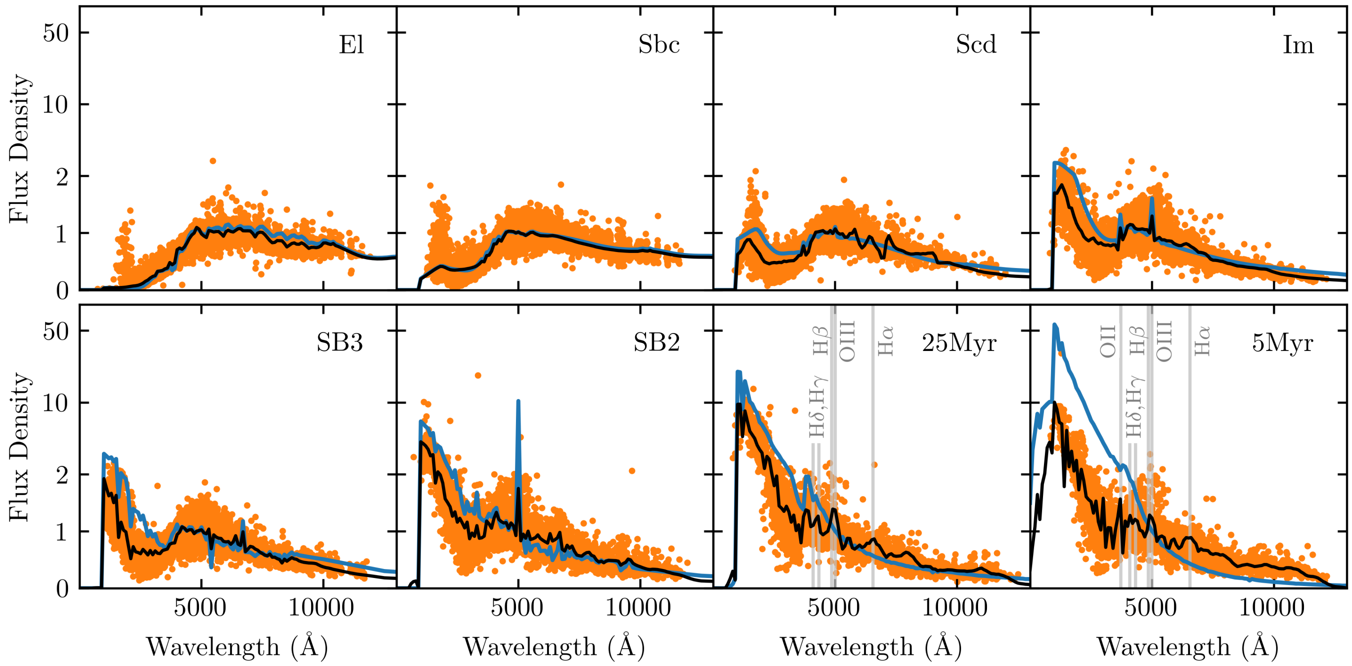

Figure 6. Result of training the CWW+SB4 templates. The original templates are in blue, the trained templates in black, and the final training sets are displayed as orange points.

encoded in the redshift posterior generated by BPZ and other photo-z codes (cite examples of this). In this work, however, only the peak of the posterior distribution is used to estimate the photo-z.

In this work, we use BPZ-v1.99.3[2] (Benitez 2000) to estimate photo-z's. We provide various sets of SED templates, as 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 Section 5.2. We use the CFHTLS i band for the magnitude prior and for simplicity, we treat non-detections as non-observations. All other settings were left as default.

BPZ provides two metrics for the photo-z estimates: ODDS and $\chi^2_{\rm mod}$. 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. $\chi^2_{\rm mod}$ 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 $\chi^2_{\rm mod} > 1$ are excluded from the analysis, and the fraction excluded on this bases is reported as $f_{\rm cut}$.

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 caused by a degeneracy wherein the 100nm Lyman break in a high redshift galaxy spectrum is shifted to the position of the 400nm 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 known to be robust to the presence of outliers. We follow Graham et al. (2018) in introducing the quantity $\Delta z_{1+z} = (z_{spec} - z_{phot})/(1 + z_{phot})$. 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, $\sigma_{\rm IQR}$, 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. Outliers are identified as photo-z's with

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

$\Delta z_{1+z} > 3\sigma_{\rm IQR}$, and the fraction of outliers is reported as $f_{\rm out}$.

5.2. *Template Classification*

The Bayesian prior used by BPZ requires each of the SED templates be broadly classified as either elliptical, spiral, or irregular/starburst. To perform this classification, we compare the colors of the trained templates with the original CWW+SB4 templates, which are already classified. In Figure 7, the colors u-g and NUV-u are plotted for every template. Each template is classified by assigning to it the classification of the nearest CWW+SB4 template in this color space. The regions of the color space corresponding to elliptical, spiral, and irregular/starburst are color-coded red, green, and blue respectively.

In the left panel of Figure 7, you can see the classifications of each of the original CWW+SB4 templates, and how the classifications are changed by the training algorithm. Originally, CWW+Sb4 contains one elliptical galaxy, two spiral, and five irregular/starburst; after training, the irregular template "Im" has been converted to a spiral template. In the middle panel, you can see that N8 contains one elliptical galaxy, four spiral galaxies, and three irregular/starburst galaxies. In the right panel, you can see that N16 contains one elliptical galaxy, eleven spiral galaxies, and four irregular/starburst galaxies.

5.3. *Photo-z Results*

We use our trained templates to estimate photo-z's for the galaxies in the validation set using BPZ (Benitez 2000), and evaluate the results by comparing to the spectroscopic redshifts and photo-z estimates using the original CWW+SB4 templates. BPZ was run with each of the template sets above, using the settings described in Section 5.1, with the SED classifications listed in the preceding section.

The photo-z results can be seen in Figure 8. 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. By comparing the top two panels, one can see that the training algorithm decreases the outlier fraction, bias, and scatter of the photo-z results for the CWW+SB4 templates. 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. Noticeably, the N8 and N16 sets perform considerably worse in the $z = 1.5$ to $z = 2.5$ range. This behavior is generally expected of photo-z estimators, as the Balmer break leaves most band sets at around $z = 1.4$ and the Lyman break does not enter most band sets until $z = 2.5$. However, this flaw is especially pronounced in the N8 and N16 sets as there is far less data for galaxies in this redshift range for the sets to train on (c.f. Figure 2). This, together with the fact that the trained CWW+SB4 set performs the best, indicates that the training algorithm should be combined with spectral synthesis models and observed spectra to yield the best results over the widest redshift range.

The results for N8 and N16 are similar, indicating that the addition of further templates has minimal impact on photo-z estimation. In addition to these four template sets, we considered one additional "augmented" set, which consists of the 16 templates from N16, as well as the El, 25Myr, and 5Myr templates from CWW+SB4. This set consists of 19 templates, and spans the widest range of the color space visible in Figure 7 (i.e. it contains the N16 template set, plus the red template in the top right and the two blue templates in the bottom left in the right panel of Figure 7). These results were virtually indistinguishable from the N16 results, and are thus not displayed in Figure 8. This further supports the case that additional templates have little impact on photo-z's.

The value of the metrics as a function of photo-z can be seen in Figure 9. For comparison, plotted in gray are the LSST science requirements for the metrics as listed in the LSST Science Requirement Document (SRD; Ivezić & the 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 from $0.3 < z < 3.0$. For comparison, our test set consists of 15,650 galaxies with $i < 27.4$, in the range $z < 3.3$, including 13,510 galaxies with $i < 25$, in the range $0.3 < z < 3.0$. In Figure 9, you can see that our training algorithm goes a long way towards reaching the LSST goals for bias and scatter.

Work in a reference to the fact that the templates were tested on a different survey from what they were trained on. This demonstrates its cross-survey robustness... although there is overlap in the filters used between the two... That might be something I need to change. Ask Andy.

6. CONCLUSIONS

ACKNOWLEDGMENTS

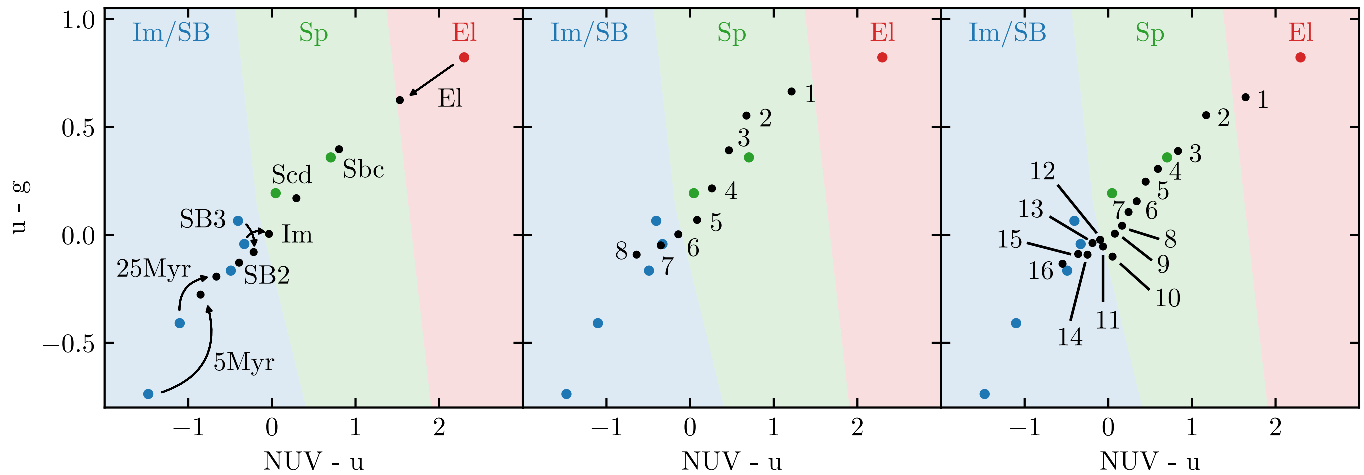

Figure 7. The colors used to classify each SED template. Templates that are in the red, green, and blue regions are classified as elliptical, spiral, and irregular/starburst respectively. The x-axis is the difference of the GALEX *NUV* band and the CFHTLS *u* band. The y-axis is the difference of the CFHTLS *u* and *g* bands. The original CWW+SB4 templates are plotted in each of the three panels, represented by colored points according to their classifications. They can be identified by the labels in the left panel. Left: the original CWW+SB4 template colors, together with their trained colors in black. The arrows point from the original colors to the trained colors, and thus show how the training moves the templates through the color space. Arrows are not shown but can be inferred for templates whose points are very close together. Middle: the colors of the trained N8 templates, labeled by their number. Right: the colors of the N16 templates.

REFERENCES

Benitez, N. 2000, Astrophys. J., 536, 571, doi: 10.1086/308947

Benitez, N., Ford, H., Bouwens, R., et al. 2004, Astrophys. J. Suppl. Ser., 150, 1, doi: 10.1086/380120

Bruzual, G., & Charlot, S. 2003, Mon. Not. R. Astron. Soc., 344, 1000, doi: 10.1046/j.1365-8711.2003.06897.x

Budavári, T., Szalay, A. S., Connolly, A. J., Csabai, I., & Dickinson, M. 2000, Astron. J., 120, 1588, doi: 10.1086/301514

Coe, D., Benítez, N., Sánchez, S. F., et al. 2006, Astron. J., 132, 926, doi: 10.1086/505530

Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, Astrophys. J. Suppl. Ser., 43, 393, doi: 10.1086/190674

Fruchter, A., & Hook, R. 2002, Drizzle: A Method for the Linear Reconstruction of Undersampled Images, Tech. Rep. 792, doi: 10.1086/338393

Graham, M. L., Connolly, A. J., Ivezić, Ž., et al. 2018, Astron. J., 155, 1, doi: 10.3847/1538-3881/aa99d4

Ilbert, O., Capak, P., Salvato, M., et al. 2009, Astrophys. J., 690, 1236, doi: 10.1088/0004-637X/690/2/1236

Ivezić, Ž., & the LSST Science Collaboration. 2018, LSST Proj. Manag. LPM-17. http://ls.st/srd

Kinney, A. L., Calzetti, D., Bohlin, R. C., et al. 1996, Astrophys. J., 467, 38, doi: 10.1086/177583

Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, Astron. Astrophys., 559, 14, doi: 10.1051/0004-6361/201322179

Lee, M. A., Budavári, T., Sullivan, I. S., & Connolly, A. J. 2019, Astron. J., 157, 182, doi: 10.3847/1538-3881/ab139f

Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, Astrophys. Journal, Suppl. Ser., 184, 218, doi: 10.1088/0067-0049/184/2/218

Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, Astrophys. J. Suppl. Ser., 225, 27, doi: 10.3847/0067-0049/225/2/27

Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, Astrophys. Journal, Suppl. Ser., 208, 57, doi: 10.1088/0067-0049/208/1/5

Scodeggio, M., Guzzo, L., Garilli, B., et al. 2018, Astron. Astrophys., 609, doi: 10.1051/0004-6361/201630114

Zhou, R., Cooper, M. C., Newman, J. A., et al. 2019, Mon. Not. R. Astron. Soc., 488, 4565, doi: 10.1093/mnras/stz1866

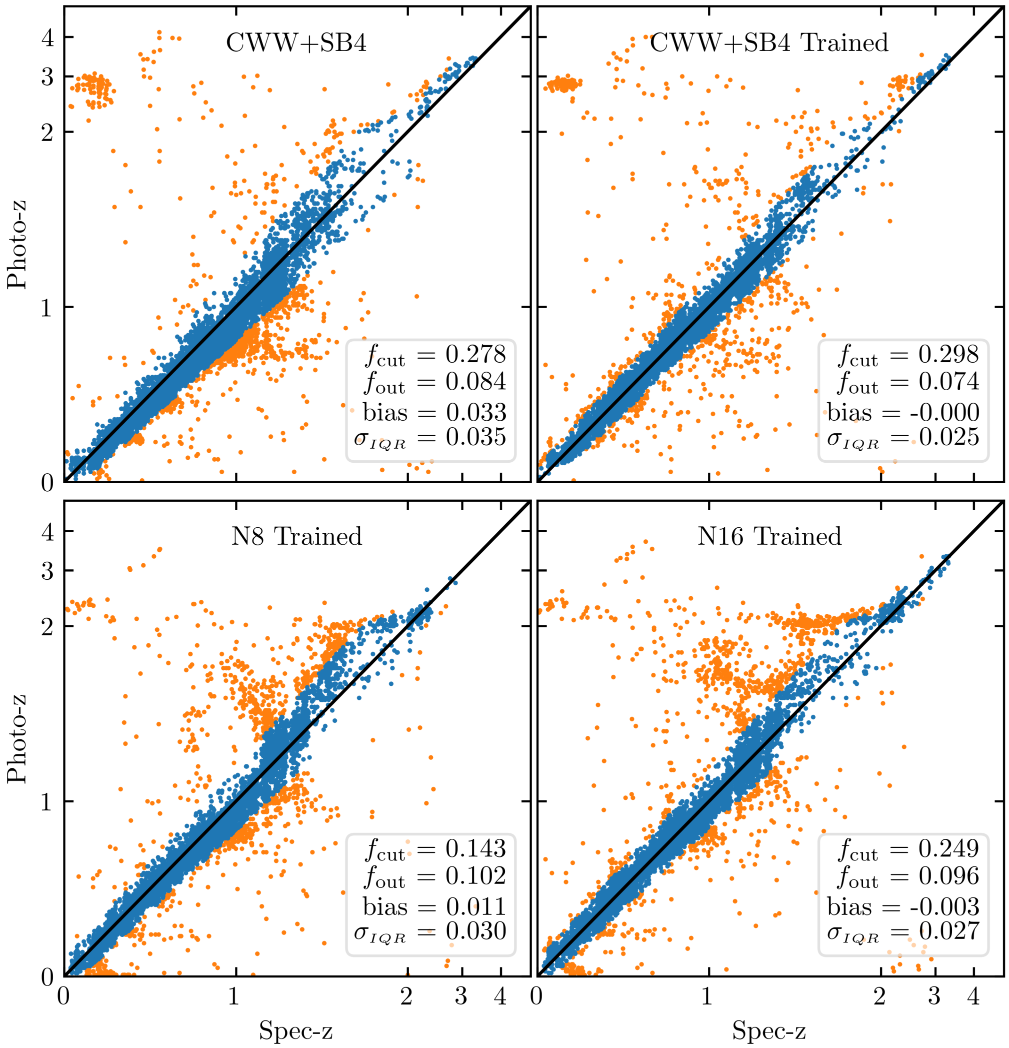

Figure 8. Results of photo-z estimation with BPZ, using the four different templates sets. Photo-z estimates are displayed as points: inliers are blue and outliers are orange. The black line represents perfect estimation (i.e. photo-z = spec-z). The statistics printed in each panel are for the entire data set.

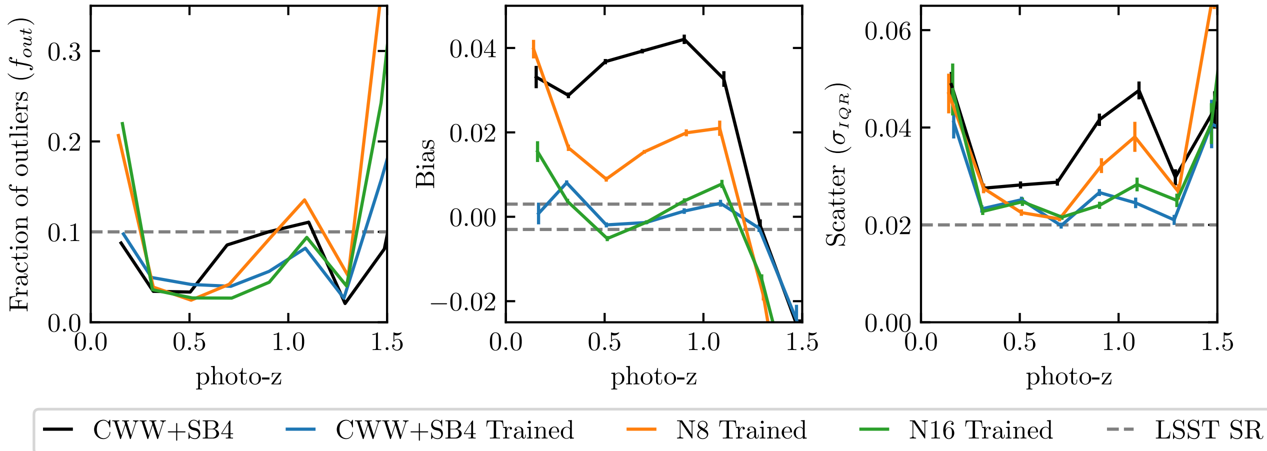

Figure 9. Photo-z metrics as a function of redshift bin. LSST science requirements are displayed as dashed gray lines.