The Astrophysical Journal, 888:83 (9pp), 2020 January 10 https://doi.org/10.3847/1538-4357/ab5a79

Photometric Redshift Estimation with Galaxy Morphology Using Self-organizing Maps

Derek Wilson[1], Hooshang Nayyeri[1], Asantha Cooray[1], and Boris Häußler[2]
[1] Department of Physics & Astronomy, University of California, Irvine, CA 92697, USA
[2] European Southern Observatory, Alonso de Cordova 3107, Vitacura, Casilla 19001, Santiago, Chile
Received 2019 June 11; revised 2019 November 18; accepted 2019 November 20; published 2020 January 10

Abstract

We use multiband optical and near-infrared photometric observations of galaxies in the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey to predict photometric redshifts using artificial neural networks. The multiband observations span from 0.39 to 8.0 μm for a sample of $\sim$1000 galaxies in the GOODS-S field for which robust size measurements are available from *Hubble Space Telescope* Wide Field Camera 3 observations. We use self-organizing maps (SOMs) to map the multidimensional photometric and galaxy size observations while taking advantage of existing spectroscopic redshifts at $0 < z < 2$ for independent training and testing sets. We show that use of photometric and morphological data led to redshift estimates comparable to redshift measurements from modeling of spectral energy distributions and from SOMs without morphological measurements.

Unified Astronomy Thesaurus concepts: Galaxy radii (617); Galaxies (573); Multi-color photometry (1077); Galaxy properties (615); Redshifted (1379)

1. Introduction

Photometric redshift (photo-z) estimation is crucial for astrophysical applications because obtaining spectroscopic redshifts for large samples of distant galaxies is often infeasible. Physical properties of extragalactic sources further depend on accurate redshift measurements. The photometric redshift can also be used as a good proxy for distance for mapping the large-scale structure and performing weak lensing studies (Munshi et al. 2008).

Unfortunately, due to selective sampling of the galaxy spectral energy distribution (SED), photometric redshifts suffer from much higher uncertainties than spectroscopic redshifts. Errors in photometric redshifts can significantly affect measurements of cosmological parameter in, for example, studies of weak lensing (e.g., Huterer et al. 2006; Ma et al. 2006; Bernstein & Huterer 2010) and baryon acoustic oscillation (e.g., Zhan & Knox 2006; Chaves-Montero et al. 2018).

The observable quantity available for photo-z estimation is galaxy photometry in multiple wavelength bands, and a large number of techniques have been developed to estimate redshift while trying to minimize $z_{\rm phot} - z_{\rm spec}$. Photometric redshift estimation is primarily done via template-fitting (e.g., Lanzetta et al. 1996; Fernández-Soto et al. 1999) and/or statistical (e.g., Connolly et al. 1995) and machine learning techniques. As surveys grow ever larger, machine learning techniques that can process enormous amounts of data with minimal human input are becoming increasingly important.

Some techniques for photo-z estimation involve using artificial neural networks with photometry and/or morphology data (e.g., Firth et al. 2003; Ball et al. 2004; Collister & Lahav 2004; Vanzella et al. 2004; Bonfield et al. 2010; Soo et al. 2018), support vector machines (e.g., Wadadekar 2005; Jones & Singal 2017), the Multi-Layer Perceptron with Quasi-Newton Algorithm (MLPQNA, Brescia et al. 2013), and the conditional density estimator FLEXCODE (Izbicki & Lee 2017). Statistical models have also been developed, such as the surface brightness and photometry model of Kurtz et al. (2007), the algorithm based on surface brightness, Sèrsic index, and photometry developed in Wray & Gunn (2008), and the Gaussian process regression model (Way & Srivastava 2006; Bonfield et al. 2010; Way 2011; Almosallam et al. 2016a, 2016b), which also appears in Gomes et al. (2018) when applied to infrared- and visible-band photometry in conjunction with angular size. Wadadekar (2005) uses support vector machines to estimate redshifts from photometric data as well as the 50% and 90% Petrosian radii for their sources. They observe 15% greater accuracy when they use the two Petrosian radii with photometry than when they use photometry alone. The empirical techniques in Vince & Csabai (2007) use photometry and morphological data from the Sloan Digital Sky Survey (SDSS), and they find that the weak correlation between morphology and redshift leads to only negligible gains in accuracy of photo-z estimation. Singal et al. (2011) use a principal component analysis including morphological parameters to estimate photometric redshifts for the All-wavelength Extended Groth Strip International Survey (AEGIS; Davis et al. 2007). They conclude that the additional noise added to the data set by including morphological parameters will offset any of the gains coming from correlations between redshift and morphology. Jones & Singal (2017) use a support vector machine to estimate photometric redshifts. Their work includes principal components of eight morphological parameters; however, they observe no significant decrease in the rms error or in the number of outliers (i.e., the number of galaxies with $(z_{\rm phot} - z_{\rm spec})/(1 + z_{\rm spec})$ greater than some value, such as the value of 0.15 in Hildebrandt et al. 2010) when using morphological data. Machine learning models are trained on photometric and/or morphological features that have been derived from the galaxy images. Hoyle (2016) develops a deep neural network that is trained directly on galaxy images, so the network itself decides which parts of the image are important. The paper does not note a significant improvement in redshift accuracy. A similar approach is found in Menou (2019), which uses a multilayer perceptron/convolutional neural network (MLP-convnet) architecture that analyzes galaxy-integrated features such as fluxes and colors using the MLP framework while adding in morphological information found by analyzing images directly with the

convnet framework. They find that the MLP-convnet architecture does lead to a significant improvement in accuracy but has no effect on the number of outliers.

We now focus on the use of a machine learning technique known as a self-organizing map (SOM; Kohonen 1982, 1990), which has increased in the last decade. An SOM is an artificial neural network whose main advantage is its ability to reduce the dimensionality of input data while preserving the relationships between data points, thus making those relationships easier to visualize. We use the SOM to characterize the multidimensional space of observed galaxy SEDs. In the literature, Tagliaferri et al. (2003) combine multilayer perceptrons with SOMs to analyze photometric data from SDSS. There is also MLZ (Machine Learning and photo-z, Carrasco Kind & Brunner 2013, 2014), which performs two regression algorithms for computing photo-zs: TPZ, which uses prediction trees and random forests, and SOMZ, which uses SOMs. SOMs are also used by Masters et al. (2015) to estimate redshifts and identify regions in galaxy color space where spectroscopic redshifts have not been obtained in past surveys. If these gaps could be filled in by future surveys, such a complete training set would be a powerful tool for photo-z estimation using machine learning. Recent work by Speagle & Eisenstein (2017a) develops a photo-z technique that combines template-fitting methods with SOMs. Concerning the predictive power of SOMs, Geach (2012) finds that SOMs can reach accuracies in photometric redshifts that are competitive with template-fitting and other empirical methods. When trained on mock data from the Large Synoptic Survey Telescope and *Euclid*, they find that their technique can predict redshifts to the accuracy required for *Euclid* weak lensing measurements (Speagle & Eisenstein 2017a, 2017b).

In this paper, we explore the effect that the addition of galaxy morphology to SOM training data has on the accuracy of redshift estimation. This paper is organized as follows: Section 2 describes the catalog data from GOODS-S used in our study. In Section 3, we summarize the SOM algorithm. Sections 4 and 5 discuss the performance of the SOMs when photometry alone and photometry plus morphology, respectively, are used for training. The AB magnitude system is used, and a flat-ΛCDM cosmology of $\Omega_{m_0} = 0.27$, $\Omega_{\Lambda_0} = 0.73$, and $H_0 = 70$ km s^{-1} Mpc^{-1} is assumed. The code developed herein will be made publicly available at https://github.com/derkwilson/PhotSOM.

2. Data

We use publicly available data from the GOODS-S field (centered at R.A. = $03^h32^m30^s$, decl. = $-27^d48^m20^s$), which covers an area of approximately 150 arcmin^2. Our training and testing catalogs are pulled from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011).[3] The full CANDELS GOODS-S catalog (Guo et al. 2013) includes optical, near-, and mid-infrared photometry from the *Hubble Space Telescope* (*HST*), the Very Large Telescope (VLT), and the *Spitzer* Infrared Array Camera (IRAC). Our primary training and testing catalogs each consist of 506 galaxies in the GOODS-S field with colors computed from the 15 bands listed in Table 1, comparable to the training and testing sets of Dahlen et al. (2013). We have an additional training set with about 1360 sources, and the results using this training set do not differ significantly from those with the training set of 506 sources, so we will focus on the results from the latter. We note that Bonfield et al. (2010) find that photo-z estimates deteriorate with fewer than 2000 training objects when using artificial neural networks and Gaussian process regression, but that the size and architecture of the network may permit reasonable results with fewer training objects. All sources in the training and testing sets have $z_{\rm spec} < 2$, and the distribution of redshifts is shown in Figure 1. Dahlen et al. (2013) previously released a training/testing catalog set with

[3] https://archive.stsci.edu/prepds/candels/

Table 1
The 19 Features Used in the Training and Testing of the SOMs

Feature	Wavelength (μm)	References.
VLT/VIMOS U	∼0.36	N09, G13
HST/ACS F435W	0.4320	G04, K11, G13
HST/ACS F606W	0.5956	G04, K11, G13
HST/ACS F775W	0.7760	G04, K11, G13
HST/ACS F814W	0.8353	G04, K11, G13
HST/ACS F850LP	0.8320	G04, K11, G13
HST/WFC3 F098M	0.985	W11, G13
HST/WFC3 F105W	1.045	K11, G13
HST/WFC3 F125W	1.250	K11, G13
HST/WFC3 F160W	1.545	K11, G13
VLT/ISAAC Ks	2.16	R10, G13
Spitzer/IRAC 3.6	3.6	A13, G13
Spitzer/IRAC 4.5	4.5	A13, G13
Spitzer/IRAC 5.8	5.8	G13
Spitzer/IRAC 8.0	8.0	G13
R_{50}	0.4320	H13
Concentration (C)	1.250	P16
Asymmetry (A)	1.250	P16
Smoothness (S)	1.250	P16

Note. The first 15 lines of the table are the photometry, showing the instrument and filter used as well as the central wavelength of the filter. The bottom four lines of the table show the morphological quantities used and the corresponding wavelengths. References: G04: Giavalisco et al. (2004), N09: Nonino et al. (2009), R10: Retzlaff et al. (2010), K11: Koekemoer et al. (2011), W11: Windhorst et al. (2011), A13: Ashby et al. (2013), G13: Guo et al. (2013), H13: Häußler et al. (2013), P16: Peth et al. (2016).

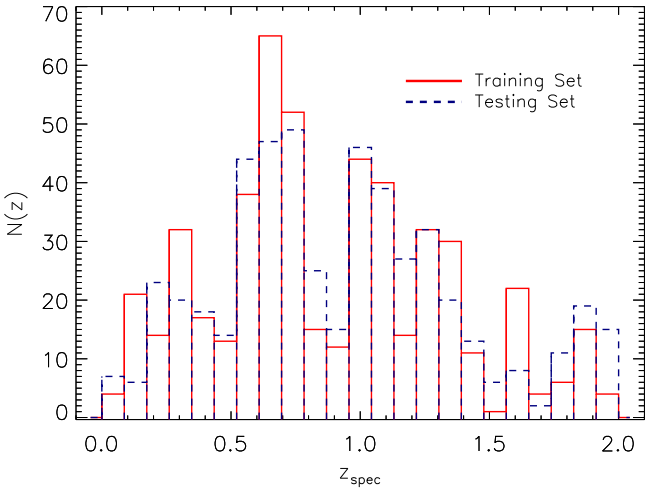

Figure 1. Histograms of the galaxy spectroscopic redshifts comprising the training set (red) and testing set (blue dashed). The training and testing sets each contain 506 individual galaxies up to a redshift of 2.

photometry in the same bands (except ACS F814W) extending up to $z \sim 5$ in redshift, so we also test our SOMs with these catalogs for comparison.

In addition to the photometry, we use half-light radii (Häußler et al. 2013) and concentration, asymmetry, and smoothness data from Peth et al. (2016) (see Table 1). In total, we use 15 photometric features and four morphological features when training and testing our SOMs. Half-light radii come from a single-Sèrsic fit to sources extracted from *H*-band images. Peth et al. (2016) extract morphological quantities from the WFC3 F125W and F160W images obtained by CANDELS. We use the *H*-band morphologies from the catalog of Peth et al. (2016). Training data consist of the colors (Guo et al. 2013) and sizes/morphologies (Häußler et al. 2013; Peth et al. 2016) for ∼500 galaxies with known spectroscopic redshifts. We match the size/morphology data to the photometry for each of the sources in these catalogs based on sky coordinates.

Galaxy morphologies are captured by a number of quantities; for example, radius, concentration, asymmetry, smoothness, Sèrsic index, axis ratio, Gini coefficient, and second-order moment (e.g., Conselice et al. 2000; Conselice 2003; Lotz et al. 2004; Peth et al. 2016). A galaxy's spatial extent can be characterized through measurements of half-light radius (hereafter R_{50}), which is the radius within which 50% of the galaxy's total flux falls. Concentration (Kent 1985; Bershady et al. 2000; Conselice 2003) describes the extent to which a galaxy's light is concentrated toward the center. The concentration is taken to be the ratio between the radii containing 80% and 20% of the galaxy's light within 1.5 Petrosian (Petrosian 1976) radii (e.g., Peth et al. 2016). Large-scale asymmetries in the light distribution of the source are described by the asymmetry statistic (Conselice et al. 2000). High asymmetry is typical for blue, star-forming galaxies and can be indicative of systems that have undergone mergers (Conselice et al. 2000; Conselice 2003). Smoothness (Conselice 2003), also known as clumpiness, traces structures with high spatial frequencies, such as star-forming regions. In contrast, objects such as elliptical galaxies consist primarily of low spatial frequencies, due to their smooth light distributions. Conselice (2003) define clumpiness as the ratio between the flux in high-frequency spatial structures and the total flux of the galaxy. There are alternative methods for identifying clumps, such as resolved rest-frame (U–V) color selections (Hemmati et al. 2014, see also Wuyts et al. 2012; Guo et al. 2015), which yield comparable results.

Together, concentration, asymmetry, and smoothness form the CAS structural parameter system (Conselice 2003). The CAS parameters form a three-dimensional volume that can be used to classify galaxies into elliptical, spiral, dwarf irregular, dwarf elliptical, and merger classes. We include the CAS system in our analysis to see whether the evolution of morphological parameters correlates strongly enough with redshift to improve photo-z estimates.

We provide a brief summary of other interesting morphological quantities that could also potentially be used in training the SOMs, though they were not used in this study. The Gini coefficient (Lorenz 1905; Abraham et al. 2003; Lotz et al. 2004) is a quantity used to measure how equally light is distributed among pixels in a galaxy image. The Gini coefficient is also correlated with concentration (Abraham et al. 2003). The second-order moment (Lotz et al. 2004) measures the flux in pixels weighted by their squared distance from the galaxy center. This statistic is sensitive to bright features such as galactic nuclei, bars, spiral arms, and star clusters (Lotz et al. 2004).

3. Redshift Measurement Algorithm

We use the SOM to identify correlations between redshift and observed galaxy colors as measured from the multiband optical and near-infrared data. Morphological information on a galaxy is included in the SOM algorithm in a later section. When the SOM is given the color/morphology data of a test galaxy, it searches for the node that is closest in color–morphology space to that test galaxy and makes an approximation of its redshift based on the location of the node within the map. In theory, we could supply the SOM with any observable quantity (photometric or morphological; such as color, half-light radius, Sèrsic index, asymmetry, concentration, Gini coefficient, etc.), and the SOM would cluster the input data according to the correlations that it locates in the data. For studies of galaxy SEDs, this means that we can explore any of the mapped properties and associate them with a measured value given the clustered information.

The construction of the SOM is similar to the self-organizing map association network (SOMA) from Yamakawa et al. (2001), though our method of association differs. A SOMA infers a set of perfect (complete) information from a set of incomplete information. For the case presented here, we take the perfect information to be a vector of data points consisting of galaxy photometry, morphology, and spectroscopic redshift, and the incomplete information would be a vector of photometric and morphological data points without a redshift. The SOMs are constructed and organized from a set of training samples consisting of perfect information; subsequently, samples composed of incomplete information and unknown spectroscopic redshift can be presented to the map for redshift classification. Note that *perfect* in this sense does not mean without error, but rather that the data *exist*.

The SOM is initialized to an $m \times n$ array of nodes. Each node contains a weight vector that covers the attribute (e.g., color, size, spectroscopic redshift) space of the input data. This weight vector is initialized to random values, and as the map is trained these weight vectors will update themselves to be more representative of the data. This training process is repeated for each galaxy in the training sample. The map as a whole has a topology that we take to be toroidal. Various works in the literature (e.g., Yamakawa et al. 2001; Masters et al. 2015) describe the training process in detail. We summarize the same process here and borrow their notation. One training iteration begins with the selection of a random training sample with feature vector $\boldsymbol{x}$ containing photometric and morphological data as well as a spectroscopic redshift. Next is the identification of the best-matching unit (BMU), the node that is closest in attribute space to the training sample according to the reduced χ^2 distance given by

$$d_k^2(\boldsymbol{x}, \boldsymbol{w}_k) = \frac{1}{m} \sum_{i=1}^{m} \frac{(x_i - w_{k,i})^2}{\sigma_{x_i}^2} \quad (1)$$

where d_k is the reduced χ^2 distance, m is the length of the feature vector $\boldsymbol{x}$, x_i is the ith component of $\boldsymbol{x}$, σ_{x_i} is the uncertainty associated with x_i, and $\boldsymbol{w}_k$ is the kth weight vector in the SOM. In the cases in which a training object or testing

object was missing a data feature (i.e., a value of −99 for flux in some band), the reduced χ^2 distances for each node were computed by taking the missing feature to be exactly equal to the node weight that corresponded to the missing feature; i.e., setting x_i equal to $w_{k,i}$ for that feature. This means that only the non-missing data will contribute to the sum in Equation (1). In this way, the incomplete training/testing vector can still exist in the m-dimensional feature space, but its reduced χ^2 distance will only depend on the features that are not missing. This technique also works if more than one feature is missing.

The goal is to have nodes with similar weights located near each other in the map. The nodes in the "neighborhood" of the BMU are determined by the neighborhood function H_k, which we take to be Gaussian:

$$H_k(t) = e^{-d_k^2/\sigma^2(t)} \tag{2}$$

where the standard deviation $\sigma^2(t)$ of the neighborhood function is

$$\sigma(t) = \sigma_0\left(\frac{1}{\sigma_0}\right)^{t/N_{\rm iters}} \tag{3}$$

where σ_0 is an arbitrary initial value, and t is an integer ranging from 1 to the total number of training iterations, $N_{\rm iters}$.

The BMU and surrounding nodes are then rewarded for being nearest to the training sample and are allowed to update their weights according to the relation

$$\boldsymbol{w}_k(t+1) = \boldsymbol{w}_k(t) + a(t)H_k(t)[\boldsymbol{x}(t) - \boldsymbol{w}_k(t)] \tag{4}$$

where we adopt the learning function $a(t)$:

$$a(t) = e^{-t/N_{\rm iters}}. \tag{5}$$

While other learning functions exist in the literature (e.g., Masters et al. 2015), we selected this one because it gave the lowest outlier fraction (OLF). The learning function decreases monotonically and is intended to desensitize the SOM to new training data as time progresses, allowing it to converge to a stable solution.

The multitude of SOM parameters (e.g., number of nodes, number of training iterations, learning rate, neighborhood function) affect the performance of the SOM as a whole. The number of nodes and training iterations used will depend on the total number of training samples available. A larger training set will require more training iterations to fully capture the data; however, it is possible to overtrain a map with too many training iterations, where the SOM learns the training data well but does not generalize to data it has not seen before. The number of nodes affects the number and size of clusters that form in the trained map. If the number of nodes is too small, the map may not capture the full set of relations present in the data. Increasing the number of nodes and training iterations comes at a cost in computing time as well. We determined by cross-validation that a map size of 150 pixels by 150 pixels had optimal predictive ability. Cross-validation involves removing a subset of samples (the validation set) from the training set, training the map on the remaining samples, and then using the validation set as testing samples. The grid size of the map is varied, and the optimal value of the grid size hyperparameter is selected based on performance on the validation set.

To extract a redshift prediction from the SOM, it is presented with a test vector that contains the same photometric and morphological attributes as the training vectors, but without the spectroscopic redshift. While ignoring the redshift attribute of the SOM nodes, the reduced χ^2 distance is computed between the test vector and each node in the map, identifying the BMU (node). The redshift of the BMU becomes the redshift associated with the test vector and represents the best prediction of the redshift of the test source.

Table 2
Summary of Performance when using the Median of Multiple SOM Predictions after Training

	σ_F	σ_o	OLF
Photometry Only	∼0.14	∼0.05	10%–1%
With R_{50}	∼0.14	∼0.06	12%–4%
With CAS	∼0.13	∼0.05	10%–2%

Note. Training was done with photometry alone, photometry plus half-light radius, and photometry plus concentration, asymmetry, and smoothness. The addition of morphological parameters had an insignificant effect on photometric redshift estimation. OLF denotes the outlier fraction, the fraction of sources with $\sigma > 0.15$.

4. SOMs on Galaxies

In order to test the SOM, the known spectroscopic redshifts of galaxies must be compared with the predictions of the map. However, the galaxies used to test the map must not be sources that the map has seen before; that is, they cannot appear in the training set. A study of several photometric redshift codes was performed by Dahlen et al. (2013), and they have released the training and control catalogs based on GOODS-S data that were used in the study. As a first test, our SOMs were trained and tested using this training/control set, which contained only photometric data. For each source, the quantity $\sigma = \Delta z/(1 + z_{\rm spec})$, where $\Delta z = z_{\rm BMU} - z_{\rm spec}$, is determined. There are several measures of performance (e.g., Dahlen et al. 2013), denoted by σ_F (=rms[$\Delta z/(1 + z_{\rm spec})$]), σ_O (the same as σ_F but it has sources with $\sigma > 0.15$ removed), and the OLF specifying the fraction of sources with $\sigma > 0.15$. Individual SOMs were trained using the training/testing set from Dahlen et al. (2013), and the performance of individual maps was found to be $\sigma_F \sim 0.17$, $\sigma_O \sim 0.042$–0.044, and OLF ∼9%–10%. To obtain a slight improvement in accuracy, the median of the results of 500 SOMs was found (since each SOM will be slightly different because the initial node weights are random and the training samples may be presented in different orders), giving $\sigma_F \sim 0.15$, $\sigma_O \sim 0.036$–0.038, and OLF ∼6%–8%.

Next, we trained and tested the SOMs using three training/testing set pairs each composed of ∼500 sources with $z < 2$. The first training/testing set contains only 13 colors (computed from 14 photometric bands), the second set contains R_{50} from a single-Sèrsic fit in addition to the colors, and the third set contains the colors as well as CAS data. We select sources with $z < 2$ because morphological measurements for sources at higher redshift will be inherently less precise. A single SOM trained and tested with our training set of $z < 2$ sources produced a typical σ_F in the range 0.14–0.16 and σ_o in the range 0.048–0.052 with OLFs of ∼10%–12%. By computing the median of multiple SOMs, we produced slightly lower values of σ_o. By averaging the SOM outputs in this way, we obtained the results in Table 2 when using photometry alone, and photometry with either R_{50} or CAS. An example of typical results is shown in Figure 2.

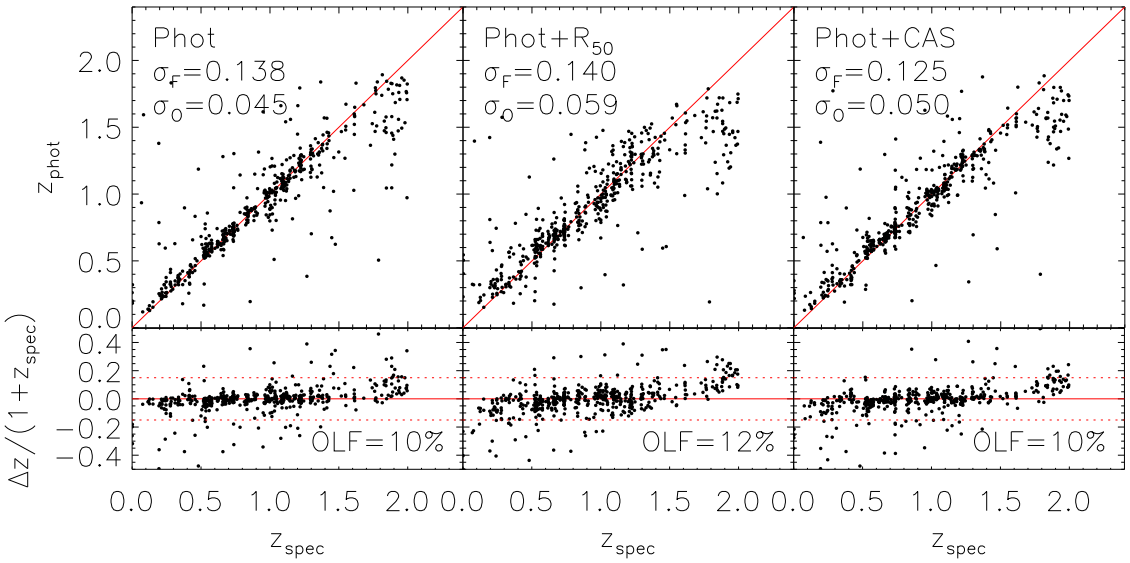

Figure 2. Top row shows a comparison of photo-z to spec-z for GOODS-S field using different subsets of data features. The bottom shows the normalized residuals given by $(z_{\rm phot} - z_{\rm spec})/(1 + z_{\rm spec})$. Left: SOM predictions using only photometric data. Middle: using photometry and half-light radius. Right: using photometry and CAS data.

For comparison, we run several public photo-z codes on the three training/testing set pairs. The photo-z codes used were PHOTORAPTOR using MLPQNA (Brescia et al. 2013), FLEXCODE (Izbicki & Lee 2017), and TPZ and SOMZ from the MLZ package (Carrasco Kind & Brunner 2013, 2014). Here we will only give a brief summary of these algorithms. MLPQNA uses a supervised learning technique involving multilayer perceptrons, a network of neurons that is trained by minimizing a loss function. The loss function is minimized by iteratively updating the weights in the neural network. The quasi-Newton algorithm is used to compute the Hessian of second derivatives, which is necessary for computing the amount by which the network weights are updated. We use a three-layer network with 15, 16, or 18 neurons in the first layer (if the training set contains just photometry, phot + R_{50}, or phot + CAS, respectively), 64 neurons in the second layer, and one neuron in the final layer. We set a decay rate of 0.001 and use a maximum of 10,000 iterations.

FLEXCODE employs a conditional density estimator method that seeks to improve photo-zs by constructing a full conditional density distribution from the data. This is done using an orthogonal series formulation, with the series coefficients determined by regression. The result is a conditional probability distribution that is useful for handling the multimodality in a photo-z prediction. When running FLEXCODE, we use the XGBoost regression method with a cosine basis system.

MLZ can perform regression using two different methods: a prediction tree and random forest algorithm and a self-organizing map algorithm. Prediction trees work by splitting the data into multiple branches based on some attribute. This process is repeated recursively until a stopping criterion is met, at which point a photo-z prediction can be made. A random forest is a collection of prediction trees whose predictions can be combined to produce more accurate results. The SOM component of MLZ works similarly to the SOM algorithm described in this work. The main difference between the SOM algorithms is the way in which spectroscopic redshift is used to train the SOM. In the MLZSOMZ, the spectroscopic redshift does not enter in the training of the SOM. Only after the map has been trained are the spectroscopic redshifts from the training sample associated with the nodes in the map, with the mean redshift of the sources associated with each node becoming the final redshift of that node. For our study with TPZ, we set the MinLeaf parameter to 10. For SOMZ, we use a periodic grid with a size of 64 nodes and 3000 training iterations.

Our implementation of the SOM algorithm uses a supervised approach. The spectroscopic redshift is included during the training process, and the final trained map will contain weights corresponding to the final redshift associated with each node. Overall, the performances of our SOM algorithm and the other photo-z codes were comparable, though missing data negatively affected the performance of some of the codes. As almost every source was missing photometry in one band or another, the replacement of the missing value with -99 may not allow the codes to perform optimally, while at the same time, removal of all data points with a missing value was not possible. The results from the photo-z codes are shown in Figure 3, and the corresponding metrics are listed in Table 3. FLEXCODE returned similar results for all three testing sets. The TPZ algorithm from MLZ was generally less accurate for the testing sets that included morphological data. We note that it is possible that there may exist hyperparameters for the FLEXCODE and TPZ algorithms that may improve their predictions but which we may have missed while tuning these models, despite our best efforts to find the optimal hyperparameters. MLPQNA and the SOMZ algorithm had large OLFs, with the number of outliers increasing when morphological data were used in training. It is likely that the large OLFs may be caused by missing data.

5. Probability Distributions

Many photo-z methods return a probability distribution in redshift space (e.g., LEPHARE: Arnouts et al. 1999; Ilbert et al. 2006, PROBWTS: Cunha et al. 2009) because methods that only give point estimates can miss important information; e.g., a probability distribution may be double-peaked, but a point

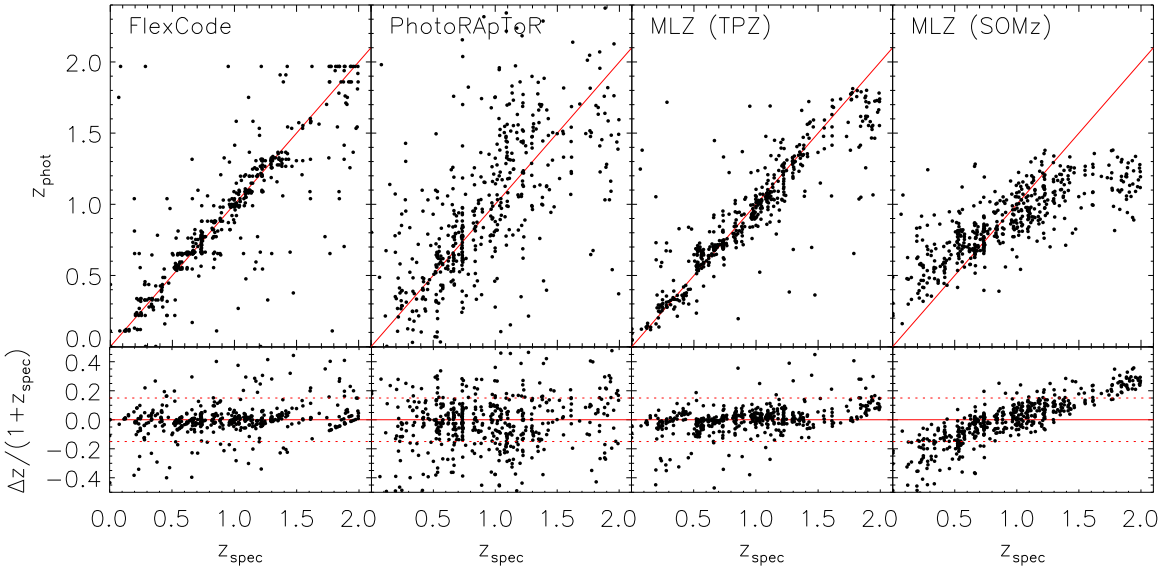

Figure 3. Example of the results from the literature photo-z codes when applied to our training/testing set containing photometry and R_{50}. See text for references and Table 3 for quantitative metrics of the results. We find that our SOM implementation produces results that are similar in dispersion and outlier fraction. PhotoRApToR and SOMz had unusually large outlier fractions, which we attribute to the effects of missing data in the training/testing sets. It is possible that a more extensive search over hyperparameter space may yield better results.

Table 3
Typical Results Obtained by Running Photo-z Codes from the Literature on our Training/testing Sets Including Photometry and Morphologies

	σ_F	σ_o	OLF
FlexCode	∼0.15	∼0.05	11%–13%
PhotoRApToR (MLPQNA)	∼0.44	∼0.07	21%–27%
MLZ (TPZ)	∼0.12	∼0.05	9%–10%
MLZ (SOM)	∼0.16	∼0.07	24%–28%

Note. The results from our SOM implementation are about the same as the results from these other software products.

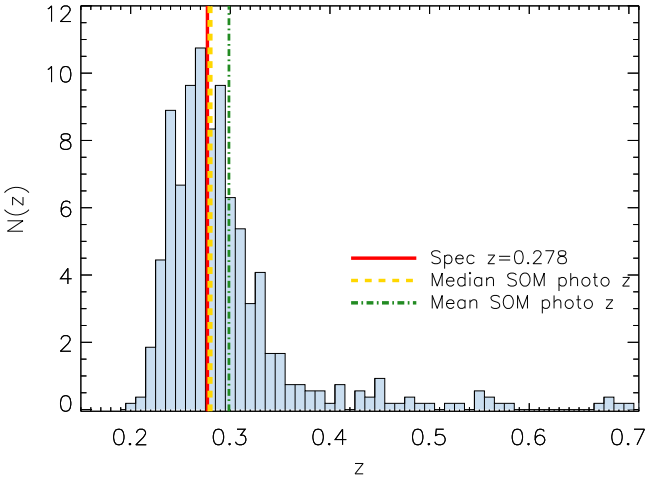

Figure 4. Example of a redshift probability distribution generated using 500 different SOMs. The spectroscopic redshift for this source is $z = 0.278$. Since each of the 500 SOMs is initialized with a different random set of parameters, each will converge to its own estimate of the redshift. The median of multiple SOMs provided measurements that were more closely aligned with the spectroscopic redshifts, due to its insensitivity to outliers.

estimate may only see the larger peak and miss the information in the secondary peak (Mandelbaum et al. 2008; Cunha et al. 2009; Wittman 2009; Bordoloi et al. 2010; Abrahamse et al. 2011; Sheldon et al. 2012). By using an ensemble of SOMs, the algorithm that we employ can be extended to return a probability distribution. Each individual SOM in the ensemble is initialized randomly, with no two SOMs having the same starting parameters. The different initializations will lead each map to converge to different weights after the training process is completed, and thus each map will predict a different photometric redshift for a test source. The results from the ensemble of SOMs are histogrammed with a bin size of $\Delta z = 0.01$ to form the final probability distribution function (PDF) (see Figure 4), and the median of the distribution is taken to be the final point estimate of the redshift.

The quality of the PDFs is tested using the probability integral transform (PIT) described in Polsterer et al. (2016) and the confidence test from Wittman et al. (2016). The PIT (Dawid 1984) is given by the histogram of the cumulative probabilities of each redshift PDF computed at the value of the spectroscopic redshift. The PIT histogram serves as a visual guide for how well calibrated the probability distribution is (Polsterer et al. 2016). Figure 5 shows an example derived from the SOM distribution functions. Ideally, the PIT should be nearly uniform if the PDFs are well calibrated. The U-shape of the histogram in Figure 5 indicates that our PDFs are underdispersed, i.e., that the dispersion in the redshift PDFs predicted by the SOMs is too small and the spectroscopic redshifts are too often ending up in the tails of the PDFs. As such, it appears that there is an overabundance of PDFs in which the statistical likelihood is very low for the spectroscopic redshift associated with the galaxy for that PDF. This means that the PDFs do not adequately represent the spectroscopic redshifts, and more work is required to make them more accurate.

The second metric used to test the SOM PDFs is the test developed by Wittman et al. (2016) to determine whether the widths of PDFs are over- or underconfident. We refer readers to the original paper for a more in-depth explanation of the test but provide a brief summary here. This confidence test is based on the principle that, ideally, a sample of galaxies should have

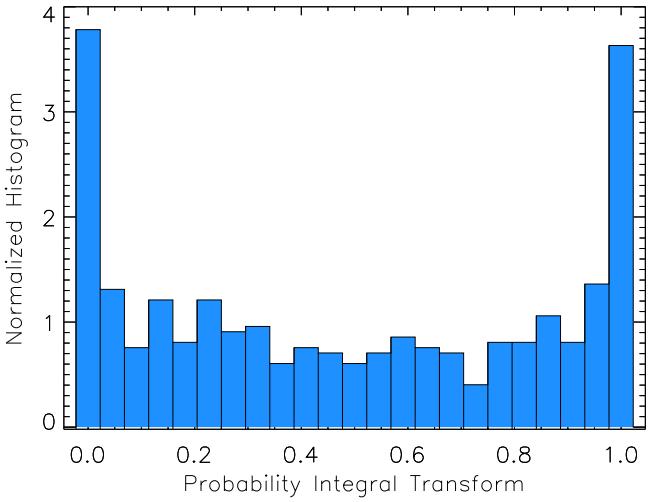

Figure 5. The PIT (e.g., Polsterer et al. 2016) from a set of redshift PDFs. A set of well-calibrated PDFs will have a near-uniform PIT. The U-shape of our PIT indicates that our redshift PDFs are underdispersed..

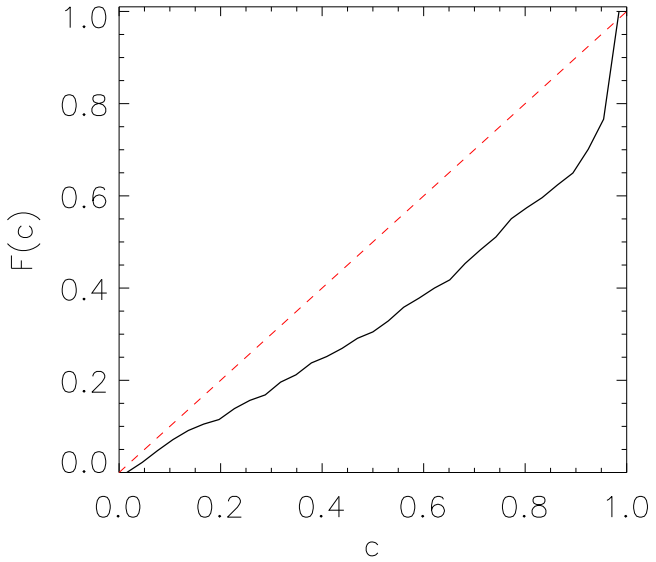

Figure 6. Confidence test from Wittman et al. (2016). Shown in black is the cumulative distribution function, $F(c)$, of the binned threshold credibilities, c. The red dashed line represents the case in which the redshift PDFs have a well-calibrated width. The plot indicates that at least some of our redshift PDFs are overconfident, i.e., that their widths are too narrow.

1% of its spectroscopic redshifts fall in the 1% credibility intervals (CI) of the corresponding PDFs, 2% of spectroscopic redshifts fall in the 2% CI, 50% of spectroscopic redshifts fall in the 50% CI, and so on. To perform the test, the threshold credibility, c_i, is computed for each galaxy in the testing set. The cumulative probability function $F(c)$ is then found from the distribution of the c_i. This cumulative distribution function is plotted in Figure 6. Ideally, the curve should lie on the red dashed line, if 1% of $z_{\rm spec}$ fall in the 1% CI, etc. In our case, the black curve lies below the ideal case, indicating that our redshift PDFs are overconfident, i.e., that the confidence intervals are too narrow and the uncertainties are underestimated. Again, more work is needed to improve the PDFs.

6. Discussion

Figure 7 shows the difference between the SOM photo-z using photometry alone and the SOM photo-z using photometry in conjunction with R_{50}. For each galaxy in the test sample, we calculate its redshift with and without R_{50} as input data and then determine the absolute difference between the two photo-zs ($|\Delta z_{\rm phot}|$ and $|\Delta z_{\rm phot+size}|$) and the spectroscopic redshift. If R_{50} had no effect on the redshift determination, then $|\Delta z_{\rm phot}| - |\Delta z_{\rm phot+size}|$ should be zero. If, however, R_{50} led to some improvement, then $|\Delta z_{\rm phot}| - |\Delta z_{\rm phot+size}|$ would be positive, since the deviation of $z_{\rm phot}$ from $z_{\rm spec}$ would be larger than the deviation of $z_{\rm phot+size}$ from $z_{\rm spec}$. Negative values would indicate that R_{50} had a detrimental effect. In Figure 7, 67% of data points lie below zero, indicating that half-light radius did not improve photo-z estimation.

We find that the addition of galaxy morphological data does not significantly improve the redshift estimation from the SOMs. The scatter introduced by the morphological data most likely dominates any benefit coming from the correlation between redshift and morphology. These results appear to be in line with the results from Soo et al. (2018), who find that adding morphological quantities such as galaxy size, Sèrsic index, surface brightness, and ellipticity does not significantly improve photo-z estimates when combined with a complete set of good photometry (in their case, full *ugriz* photometry). Soo et al. (2018) conclude that including a full set of photometric bands may saturate the amount of redshift information

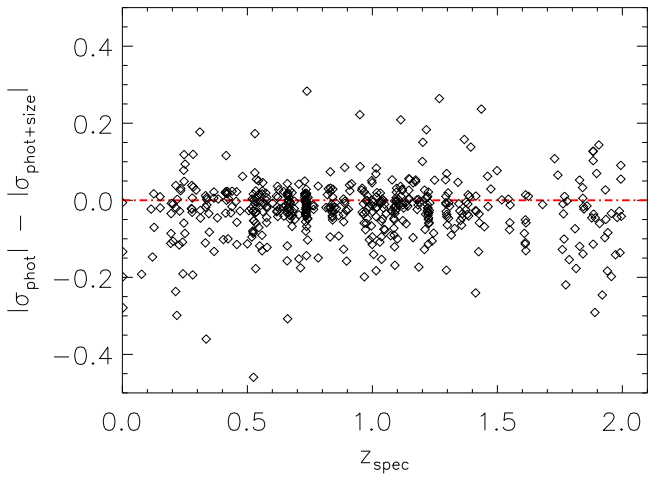

Figure 7. Comparison of SOM predictions for each test source with and without half-light radius. The quantity Δz is given by $z_{\rm spec} - z_{\rm phot}$. We show Δz calculated with photometry alone ($\Delta z_{\rm phot}$) minus Δz calculated with photometry and size ($\Delta z_{\rm phot\ +\ size}$). Positive values indicate that the use of half-light radius increased the accuracy of the SOM, while negative values indicate a decrease in accuracy.

available, which is reasonable given that they find improvement in photo-z estimates when morphology is used in conjunction with sub-optimal photometry or photometry in fewer than all five *ugriz* bands. Similarly, we conclude that our use of morphology, at its present precision, may not be providing any new information that is not already contained in our 15 bands of photometry. Soo et al. (2018) also compare the effects of low-quality versus high-quality morphology by studying galaxy radii measured by the SDSS Stripe 82 survey and by the Canada–France–Hawaii Telescope (CFHT) in Stripe 82 (CS82), the latter of which they assume to be of higher quality due to its 0.″6 seeing. However, they do not find any improvement in photo-zs when using the CS82 data over the SDSS data. In comparison, we find that improvement might be

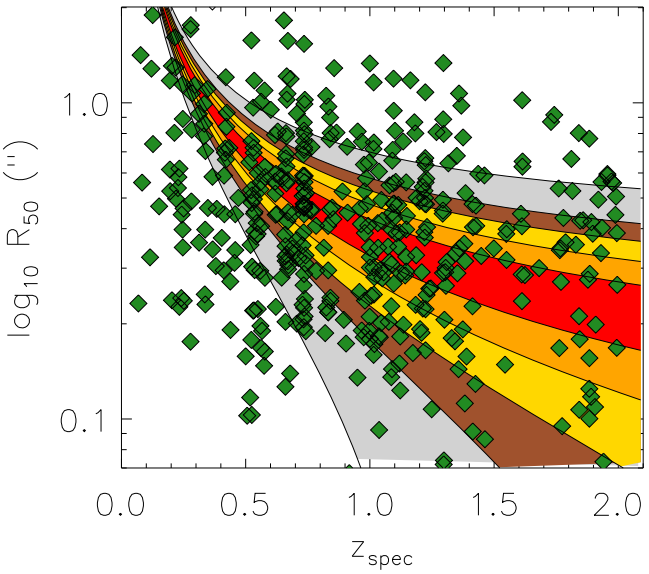

Figure 8. Comparison of simulated R_{50} with real R_{50} data (green diamonds). The regions correspond to simulated R_{50} with different Gaussian spreads around a presumed average trend; red: $\sigma = 0\farcs05$, orange: $\sigma = 0\farcs1$, yellow: $\sigma = 0\farcs15$, brown: $\sigma = 0\farcs2$, and gray: $\sigma = 0\farcs32$. The scatter of the real R_{50} is $\sim 0\farcs32$, with approximately 68% of data points falling within the gray region. We find improvement in photo-z estimates that include R_{50} only when the spread in R_{50} is smaller than $0\farcs05$. Such a spread in real data may be impossible to achieve due to the intrinsic variation in R_{50}, even with increased telescopic precision.

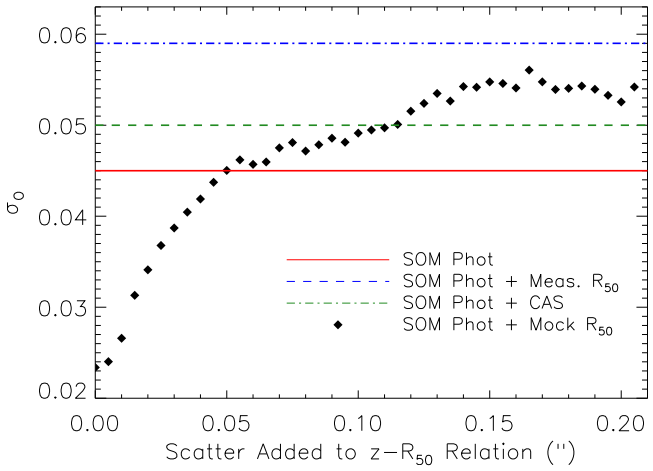

Figure 9. Redshift uncertainty as a function of the scatter added to the theoretical size relation for the GOODS-S field (black dots). The training data for the SOM results given by the black dots consist of photometry and size (half-light radius, computed according to the relation in Mosleh et al. (2012) (see also van der Wel et al. 2014). For comparison, we show the performance of the SOM when using photometry alone (red line), photometry and half-light radius from GALFIT (blue line), and the existing precision of photo-zs in the CANDELS catalog. The SOMs with photometry+size would perform better than with photometry alone if the variation in size at a particular redshift was less than about $0\farcs02$. If future surveys with higher precision instruments could measure half-light radii to this precision, the SOM networks presented here may offer improvement in photo-z estimates.

possible if the scatter in radii is less than $0\farcs05$ (Figure 9), which is well below the CS82 seeing.

While morphological parameters did not lead to significant increases in accuracy, we would like to see whether future morphological measurements with increased precision may lead to better SOM predictions. To do this, we pass simulated R_{50} data to the SOMs during training and testing. The mock size data are generated by taking the power-law fits for $\log(r_e)$ as a function of redshift for Lyman-break galaxies in Mosleh et al. (2012) to be the true relation between size and redshift (see also van der Wel et al. 2014). The simulated R_{50} are drawn from a Gaussian distribution with a variable standard deviation (scatter) and mean equal to the half-light radius at each redshift from the "true relation." Figure 8 shows a comparison of the simulated R_{50} with the actual R_{50} from the data. In Figure 9, we examine the effect that increased precision in R_{50} has on σ_o for a sample of galaxies. As the amount of scatter (black points) is lowered, improvement in photo-z estimation is achieved when the deviation in half-light radius from the theoretical relation is less than $0\farcs05$. Even with next-generation space telescopes such as the *James Webb Space Telescope* (*JWST*) and the Wide Field Infrared Survey Telescope (*WFIRST*) with diameters of 6.5 m and 2.4 m, respectively, the best angular resolution possible would be $0\farcs05$ and $0\farcs15$ for the H band at 1.65 μm. Improvement in photo-z estimation using half-light radius may not be viable in the near future. It may also be the case that the intrinsic scatter in radii at the same redshift may be too large (i.e., greater than $0\farcs03$) for any correlation to improve redshift estimates.

7. Summary

We apply the SOM algorithm to photometric and morphological data in the GOODS-S field to study the effect that morphological parameters have on estimating photometric redshifts. The SOMs are trained on photometry in 15 wavelength bands and on half-light radius, concentration, asymmetry, and smoothness for about 500 galaxies with known spectroscopic redshifts up to $z \sim 2$. The SOMs make predictions for the redshifts of about 500 galaxies in a separate testing set and are compared to the spectroscopic redshifts of those sources. The results indicate no significant improvement in the accuracy of the SOM redshift predictions when using morphology plus photometry, in comparison to photometry alone. Similar results are obtained after cursory studies using our training and testing data on other photo-z codes, leading to typical results of $\sigma_F \sim 0.13$–0.16, $\sigma_O \sim 0.05$–0.07, and OLF $\sim$ 10%–14% in the best cases. We attribute this result to the large scatter in the morphological data and the possibility that morphology is not introducing any new information that is not already contained in the photometry.

Redshift PDFs are produced by the SOMs in addition to point estimates. PDFs are more sensitive to multimodality in the results of SOM predictions. At the present, tests of our redshift pdfs show that they are underdispersed as well as overconfident (or too narrow in width), and more work is required to improve their accuracy.

Lastly, we explore the effect that a strong radius–redshift relation would have on the SOM predictions. The goal was to identify how tight a radius–redshift relation would have to be in order to give improvement in photo-z estimation. This was done by simulating half-light radii with varying levels of scatter around a theoretical radius–redshift relation. The simulated radii were used along with photometry to train and test a group of SOMs. Improvement was found only for very small scatter less than $\sim 0\farcs05$ around a theoretical radius–redshift relation.

This material is based upon work supported by the National Science Foundation under award number 1633631. Additional support for this paper was provided in part by GAANN

P200A150121, NSF grant AST-1313319, NASA grant NNX16AF38G, HST-GO-13718, HST-GO-14083. This work is based on observations taken by the CANDELS Multi-Cycle Treasury Program with the NASA/ESA *HST*, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

ORCID iDs

Derek Wilson https://orcid.org/0000-0001-6002-423X
Hooshang Nayyeri https://orcid.org/0000-0001-8242-9983

References

Abraham, R. G., van den Bergh, S., & Nair, P. 2003, ApJ, 588, 218
Abrahamse, A., Knox, L., Schmidt, S., et al. 2011, ApJ, 734, 36
Almosallam, I. A., Jarvis, M. J., & Roberts, S. J. 2016a, MNRAS, 462, 726
Almosallam, I. A., Lindsay, S. N., Jarvis, M. J., & Roberts, S. J. 2016b, MNRAS, 455, 2387
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, ApJ, 769, 80
Ball, N. M., Loveday, J., Fukugita, M., et al. 2004, MNRAS, 348, 1038
Bernstein, G., & Huterer, D. 2010, MNRAS, 401, 1399
Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, AJ, 119, 2645
Bonfield, D. G., Sun, Y., Davey, N., et al. 2010, MNRAS, 405, 987
Bordoloi, R., Lilly, S. J., & Amara, A. 2010, MNRAS, 406, 881
Brescia, M., Cavuoti, S., D'Abrusco, R., Longo, G., & Mercurio, A. 2013, ApJ, 772, 140
Carrasco Kind, M., & Brunner, R. J. 2013, MNRAS, 432, 1483
Carrasco Kind, M., & Brunner, R. J. 2014, MNRAS, 438, 3409
Chaves-Montero, J., Angulo, R. E., & Hernández-Monteagudo, C. 2018, MNRAS, 477, 3892
Collister, A. A., & Lahav, O. 2004, PASP, 116, 345
Connolly, A. J., Csabai, I., Szalay, A. S., et al. 1995, AJ, 110, 2655
Conselice, C. J. 2003, ApJS, 147, 1
Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, ApJ, 529, 886
Cunha, C. E., Lima, M., Oyaizu, H., Frieman, J., & Lin, H. 2009, MNRAS, 396, 2379
Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, ApJ, 775, 93
Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, ApJL, 660, L1
Dawid, A. P. 1984, Journal of the Royal Statistical Society, 147, 278
Fernández-Soto, A., Lanzetta, K. M., & Yahil, A. 1999, ApJ, 513, 34
Firth, A. E., Lahav, O., & Somerville, R. S. 2003, MNRAS, 339, 1195
Geach, J. E. 2012, MNRAS, 419, 2633
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJL, 600, L93
Gomes, Z., Jarvis, M. J., Almosallam, I. A., & Roberts, S. J. 2018, MNRAS, 475, 331
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, ApJ, 800, 39
Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24
Häußler, B., Bamford, S. P., Vika, M., et al. 2013, MNRAS, 430, 330
Hemmati, S., Miller, S. H., Mobasher, B., et al. 2014, ApJ, 797, 108
Hildebrandt, H., Arnouts, S., Capak, P., et al. 2010, A&A, 523, A31
Hoyle, B. 2016, A&C, 16, 34
Huterer, D., Takada, M., Bernstein, G., & Jain, B. 2006, MNRAS, 366, 101
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Izbicki, R., & Lee, A. B. 2017, EJSta, 11, 2800
Jones, E., & Singal, J. 2017, A&A, 600, A113
Kent, S. M. 1985, ApJS, 59, 115
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kohonen, T. 1982, Biological Cybernetics, 43, 59
Kohonen, T. 1990, IEEEP, 78, 1464
Kurtz, M. J., Geller, M. J., Fabricant, D. G., Wyatt, W. F., & Dell'Antonio, I. P. 2007, AJ, 134, 1360
Lanzetta, K. M., Yahil, A., & Fernández-Soto, A. 1996, Natur, 381, 759
Lorenz, M. O. 1905, Publications of the American Statistical Association, 9, 209
Lotz, J. M., Primack, J., & Madau, P. 2004, AJ, 128, 163
Ma, Z., Hu, W., & Huterer, D. 2006, ApJ, 636, 21
Mandelbaum, R., Seljak, U., Hirata, C. M., et al. 2008, MNRAS, 386, 781
Masters, D., Capak, P., Stern, D., et al. 2015, ApJ, 813, 53
Menou, K. 2019, MNRAS, 489, 4802
Mosleh, M., Williams, R. J., Franx, M., et al. 2012, ApJL, 756, L12
Munshi, D., Valageas, P., van Waerbeke, L., & Heavens, A. 2008, PhR, 462, 67
Nonino, M., Dickinson, M., Rosati, P., et al. 2009, ApJS, 183, 244
Peth, M. A., Lotz, J. M., Freeman, P. E., et al. 2016, MNRAS, 458, 963
Petrosian, V. 1976, ApJL, 209, L1
Polsterer, K. L., D'Isanto, A., & Gieseke, F. 2016, arXiv:1608.08016
Retzlaff, J., Rosati, P., Dickinson, M., et al. 2010, A&A, 511, A50
Sheldon, E. S., Cunha, C. E., Mandelbaum, R., Brinkmann, J., & Weaver, B. A. 2012, ApJS, 201, 32
Singal, J., Shmakova, M., Gerke, B., Griffith, R. L., & Lotz, J. 2011, PASP, 123, 615
Soo, J. Y. H., Moraes, B., Joachimi, B., et al. 2018, MNRAS, 475, 3613
Speagle, J. S., & Eisenstein, D. J. 2017a, MNRAS, 469, 1186
Speagle, J. S., & Eisenstein, D. J. 2017b, MNRAS, 469, 1205
Tagliaferri, R., Longo, G., Andreon, S., et al. 2003, LNCS, 2859, 226
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28
Vanzella, E., Cristiani, S., Fontana, A., et al. 2004, A&A, 423, 761
Vince, O., & Csabai, I. 2007, in IAU Symp. 241, Stellar Populations as Building Blocks of Galaxies, ed. A. Vazdekis & R. Peletier (Cambridge: Cambridge Univ. Press), 573
Wadadekar, Y. 2005, PASP, 117, 79
Way, M. J. 2011, ApJL, 734, L9
Way, M. J., & Srivastava, A. N. 2006, ApJ, 647, 102
Windhorst, R. A., Cohen, S. H., Hathi, N. P., et al. 2011, ApJS, 193, 27
Wittman, D. 2009, ApJL, 700, L174
Wittman, D., Bhaskar, R., & Tobin, R. 2016, MNRAS, 457, 4005
Wray, J. J., & Gunn, J. E. 2008, ApJ, 678, 144
Wuyts, S., Förster Schreiber, N. M., Genzel, R., et al. 2012, ApJ, 753, 114
Yamakawa, T., Horio, K., & Kubota, R. 2001, Advances in Self-Organising Maps (London: Springer), 15
Zhan, H., & Knox, L. 2006, ApJ, 644, 663