Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **498,** 5498–5510 (2020) doi:10.1093/mnras/staa2741
Advance Access publication 2020 September 11

Augmenting machine learning photometric redshifts with Gaussian mixture models

P. W. Hatfield [1]★ I. A. Almosallam,[2] M. J. Jarvis [1,3] N. Adams,[1] R. A. A. Bowler [1] Z. Gomes [1] S. J. Roberts[4] and C. Schreiber[1]

[1]*Sub-department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK*
[2]*Saudi Information Technology Company, Riyadh 12382, Saudi Arabia*
[3]*Department of Physics, University of the Western Cape, Bellville 7535, South Africa*
[4]*Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK*

Accepted 2020 August 29. Received 2020 August 20; in original form 2020 May 26

ABSTRACT

Wide-area imaging surveys are one of the key ways of advancing our understanding of cosmology, galaxy formation physics, and the large-scale structure of the Universe in the coming years. These surveys typically require calculating redshifts for huge numbers (hundreds of millions to billions) of galaxies – almost all of which must be derived from photometry rather than spectroscopy. In this paper, we investigate how using statistical models to understand the populations that make up the colour–magnitude distribution of galaxies can be combined with machine learning photometric redshift codes to improve redshift estimates. In particular, we combine the use of Gaussian mixture models with the high-performing machine-learning photo-z algorithm GPz and show that modelling and accounting for the different colour–magnitude distributions of training and test data separately can give improved redshift estimates, reduce the bias on estimates by up to a half, and speed up the run-time of the algorithm. These methods are illustrated using data from deep optical and near-infrared data in two separate deep fields, where training and test data of different colour–magnitude distributions are constructed from the galaxies with known spectroscopic redshifts, derived from several heterogeneous surveys.

Key words: techniques: photometric – surveys – galaxies: distances and redshifts.

1 INTRODUCTION

Many of the current key open questions in cosmology and extragalactic astronomy require extremely wide (probing large areas of the sky/volumes of the Universe) and deep (probing to extremely faint and/or distant sources) galaxy surveys to answer e.g. observations with Euclid (Laureijs et al. 2011) and the Vera C. Rubin Observatory (hereafter Rubin; formerly the Large Synoptic Survey Telescope, LSST) (LSST Science Collaboration 2009). To probe the time evolution/third dimension of the Universe, estimates of the redshift of each galaxy are typically required.

Galaxy, and active galactic nuclei (AGN), redshifts can be calculated from their spectrum in two main ways, from spectroscopy or from photometry. Spectroscopic redshifts ('spec-z's') are calculated by measuring the wavelength of a known spectral (normally emission) line or feature, and comparing it to the known rest-frame wavelength of the line or feature. Photometric redshifts ('photo-z's') are calculated by measuring the brightness of the galaxy in N broad wavelength ranges, and mapping these brightnesses on to a redshift. Typically, spec-z's are far more precise than photo-z's, but can only be measured for much smaller populations of galaxies as spectroscopic observations are more costly, and generally reach shallower depths, see Fernandez-Soto et al. (2001) for a comparison of the strengths and weaknesses of both classes of measurement.

Photometric redshifts themselves can be calculated in two main ways: 'template fitting' methods and 'machine learning' (ML) methods. Template fitting methods are essentially 'theory' based methods – we attempt to use our understanding of the physics behind galaxy spectral energy distributions to map photometry to a spectrum. This in practice can take a number of forms, but typically consists of using a number of model template spectra (either using synthetic spectra or spectra extracted at low redshift from galaxies with observations at many wavelengths), and using a χ^2-minimization-like method to find the 'best' redshift. Notable template-fitting based codes include Photometric Analysis for Redshift Estimate (LEPHARE; Arnouts et al. 1999; Ilbert et al. 2006), Bayesian photometric redshifts (BPZ; Benitez 2000; Benitez et al. 2004; Coe et al. 2006), the Zurich Extragalactic Bayesian Redshift Analyzer (ZEBRA; Feldmann et al. 2006), EAZY (Brammer, van Dokkum & Coppi 2008), and PHOSPHOROS (Paltani et al., in preparation). ML photo-z methods are typically entirely data based; an ML algorithm is given a set of galaxies with photometry and known (usually spectroscopic) redshifts, and is then tasked with predicting the redshift of galaxies without known spec-z's. Widely used ML photo-z codes include Artificial Neural Network Redshifts (ANNZ2; Collister & Lahav 2004; Sadeh, Abdalla & Lahav 2016), Trees for Photo-Z (TPZ; Carrasco Kind & Brunner 2013), self-organizing map redshifts (SOMZ; Carrasco Kind & Brunner 2013), Machine-learning Estimation Tool for Accurate PHOtometric

★ E-mail: peter.hatfield@physics.ox.ac.uk

Redshifts (METAPHOR; Cavuoti et al. 2017), FRANKEN-Z (Speagle et al., in preparation)[1], and many more. The resulting photometric redshifts are required for a variety of science goals (see Desprez et al., in preparation, for a recent photo-z code comparison in the context of Euclid objectives, and Schmidt et al., 2020, in the context of Rubin), but one of the most important, with the most stringent requirements, is weak lensing, where the matter power spectrum is measured by the shear on galaxy shapes, but accurate unbiased redshift estimates are needed for unbiased cosmological inferences, e.g. Banerji et al. (2008), Abdalla et al. (2011), Hearin et al. (2010), Hildebrandt et al. (2017), and Hoyle et al. (2018). Finally, Salvato, Ilbert & Hoyle (2019) present a comprehensive review of contemporary photometric redshift methods, applications and challenges, and a discussion of what advanced approaches must be developed for surveys to best meet their scientific goals in the future.

In this work, we consider how using Gaussian mixture models (GMMs), a Bayesian approach to dividing a set of objects into subpopulations, can support the ML photo-z code GPz (Almosallam, Jarvis & Roberts 2016a; Almosallam et al. 2016b) – although the approach could be employed with other algorithms. In particular, we investigate using GMMs to (i) help account for the different colour space distributions of the training and test data, and (ii) exploit the fact that galaxies and AGN naturally fall into different populations. The approach has some similarities to that described in Fotopoulou & Paltani (2018), who divide their galaxies into separate galaxy populations as part of the photometric redshift calculation process (but for a template fitting method), and also Lima et al. (2008), who use weighting schemes to estimate the redshift distribution, accounting for differences in colour distributions to a reference sample.

The structure of this paper is as follows. In Section 2, we describe the algorithms used in this study, namely GPz and a GMM algorithm, and the data used. In Section 3, we discuss the methods developed. In Section 4, we present our results, discuss the consequences in Section 5, and conclude in Section 6.

2 PRELIMINARIES

2.1 GPz

GPz is an ML regression algorithm originally developed for the problem of calculating photometric redshifts; the details of the algorithm and the key developments in ML theory are described in Almosallam et al. (2016b), Almosallam et al. (2016a). The algorithm is 'sparse Gaussian process' (GP) based, e.g. see Rasmussen & Williams (2006). A GP is a stochastic process with a random variable defined at each point in a space of interest, such that any linear combinations of random variables from different points has a Gaussian distribution; essentially an unparametrized continuous function defined everywhere with Gaussian uncertainties. GPs are very flexible class of supervised non-linear regression algorithms that make very few explicit parametric assumptions about the nature of the function. For this reason, they are well-suited for modelling complex non-linear mappings like photometric redshifts[2] (Rasmussen & Williams 2006; Bonfield et al. 2010; Almosallam et al. 2016b). A GP-based ML algorithm will typically take some set of data over the parameter space of interest and in some sense try and find the GP that was most likely to have produced the data – and then make predictions for other parts of parameter space based on that. The key features introduced by GPz include (i) implementation of a sparse GP framework, allowing the algorithm to run in $O(nm^2)$ instead of $O(n^3)$, where n is the number of samples in the data and m is the number of basis functions, (ii) a 'cost sensitive learning' framework, where the algorithm can be tailored for the precise science goal,[3] and (iii) properly accounting for uncertainty contributions from both variance in the data as well as uncertainty from lack of data in a given part of parameter space (by marginalizing over all the GPs that could have produced the data). GPz was further tested and developed in Gomes et al. (2018), who measured the improvement that could be achieved by also including near-infrared bands and angular sizes, as well as introducing a post-processing calibration that reduced the bias (the difference between the true/spectroscopic redshift, and the photo-z estimate, $z_{\rm spec} - z_{\rm phot}$). Duncan et al. (2018) introduced combining GPz with template-based photometric redshifts using a hierarchical Bayesian model that gave better photometric redshifts than ML or template fitting alone could have produced (a hybrid approach was also considered in Desprez et al., in preparation). GPz is also beginning to be used in other astronomy and physics applications e.g. building surrogate models for and quantifying the uncertainty on inertial confinement fusion experiments (Hatfield et al. 2020) and orbital dynamics (Peng & Bai 2019).

Two key deficiencies of GPz as applied to photo-z's are: (i) GPs ordinarily only produce Gaussian uncertainties, whereas the true probability distribution of a galaxy's redshift based on its photometry is typically not Gaussian,[4] and (ii) typically the target galaxy population has a different colour distribution than the colour distribution of the training set, which introduces biases (a problem common to all ML-based approaches). In this paper, we attempt to account for and mitigate against these difficulties.

Unless otherwise stated, we use the settings in Table 1 (see Almosallam et al. 2016a, b for precise definitions and interpretations).

2.2 GMMs

Mixture Models are probabilistic models for modelling data with subpopulations, where the observed data does not identify which population a datum is from. An everyday example of a mixture model could be length of publication measured in number of words; the distribution would have separate populations of letters, journal articles, and books with very different word length, but typically it would not be possible to separate out short articles from long letters etc. Common astronomical examples include identifying star clusters, identifying populations in surveys, deciding how many classes of a source exist etc. (see Kuhn & Feigelson 2017), and they have also been used for more complex tasks like the identification of strong gravitational lenses (e.g. Cheng et al. 2020). GMMs are a specific type of mixture where each mixture has a Gaussian distribution. GMMs can be viewed as an example of unsupervised ML, in that the algorithm is not told in advance what or how many populations there should be or given any examples of members of

[1] https://github.com/joshspeagle/frankenz

[2] Photometric redshift mappings are likely not perfectly represented by a Gaussian process, but this is likely true of any ML algorithms applied to any real-world problem.

[3] For example, if one is only interested in a science case that requires getting accurate redshifts for $z < 1$ galaxies, and such science is insensitive to poor predictions at higher redshifts. In this case, the learning cost function would include no penalty for getting $z > 1$ galaxy redshifts wrong.

[4] GPs can produce more complex pdfs, but this requires use of a 'warped' GP, a mixture of GPs, or a 'mixture density' GP.

Table 1. Parameter setting of GPz.

Parameter	Value	Description
m	500	Number of basis functions; complexity of GP, in general higher m is more accurate but longer run time
maxIter	500	Maximum number of iterations
maxAttempts	50	Maximum iterations to attempt if there is no progress on the validation set
Method	GPVC	Bespoke covariances on each basis function
Normalize	True	Pre-process the input by subtracting the means and dividing by the standard deviations
Joint	True	Jointly learn a prior linear mean-function

populations. See also D'Isanto & Polsterer (2018), who use a mixture density network to make a GMM of the galaxy redshift posterior.

More formally, we would like to maximize the likelihood distribution:

$$p(\mathbf{X}) = \Pi_{i=1}^{n} p(\mathbf{x}_i), \tag{1}$$

where $\mathbf{X} = \{\mathbf{x}_i\}_i^n$ is the set of samples in the data set, $\mathbf{x}_i \in \mathbb{R}^d$ is the i-th sample in the data set, d is the dimensionality of the input, and $p(\mathbf{x})$ is some multivariate probability density function. Assuming a multivariate normal Gaussian distribution for $p(\mathbf{x})$, the parameters of the distribution that would maximize the likelihood of observing the data, the mean and covariance, can be computed analytically. However, as discussed previously most distributions found in the real world are more complex than a simple unimodal normal distribution. We can model a more complex distribution by assuming that $p(\mathbf{x})$ is the marginal distribution over some latent variable j as follows:

$$p(\mathbf{x}_i) = \sum_{j=1}^{k} p(\mathbf{x}_i|j)p(j). \tag{2}$$

In effect, we have modelled the probability density function $p(\mathbf{x})$ as a weightinged sum of Gaussian distributions, i.e. a GMM. An important property of GMMs is that they can, for some number of mixtures k, model any non-standard probability distribution. The goal now is to find the k means, covariances and mixture weightings that would maximize the probability of observing the data. This typically cannot be solved analytically, and normally requires optimization techniques. The expectation-maximization (EM) algorithm is an iterative method that can be used to search for such parameters. However, the EM algorithm is prone to overfitting, especially as the number of mixtures is increased (at the limit when $k = n$, the means will correspond to the data locations and covariances will be close to zero). To overcome this, we use a Variational Bayes (VB, see Jordan et al. 1999; Jaakkola & Jordan 2000) approach that puts priors on the means, covariances, and mixture weightings to always find the optimal set of mixtures; even if k is set too high, it will automatically prune the extra mixtures by setting their mixture weightings to zeros.

2.3 Data

In a realistic scenario, all galaxies with spectroscopic redshifts would be the training set, and all sources without would be the target test set. Coping with the training and test data sets having different colour–magnitude distributions is a major challenge for ML-based photo-z calculations, e.g. see Beck et al. (2017). Unfortunately, however, the nature of the problem is that it is difficult/impossible to measure the performance on the galaxies without spec-zs. To overcome this problem, and to test how our method performs when the test and training data have different distributions, we construct training and test data sets for which both have spectroscopic redshifts.[5]

In order to ensure a rigorous test of the methods, we consider two separate deep field data sets that have similar photometric coverage. Our training data is from the COSMOS field, and the test data from the *XMM–Newton* large-scale structure (*XMM*-LSS) field (see Fig. 1). We use the catalogues constructed in Bowler et al. (2020) and Adams et al. (2020), which, in order to ensure consistency, used identical procedures to extract the photometry across the two fields. Sources were selected in the K_s band, and forced photometry was performed on all the other bands. We use 2 arcsec diameter circular apertures, which had an aperture correction applied by a model generated with PSFEX (Bertin 2011) for each band.

For this paper, we use the photometry in 10 filters; u (CLAUDS, CFHT, for both COSMOS and *XMM*-LSS) (Sawicki et al. 2019), GRIZY (HSC–SSP, for both COSMOS and *XMM*-LSS) (Aihara et al. 2018), and YJHK$_s$ (VIDEO–VISTA for *XMM*-LSS) (Jarvis et al. 2013), and (UltraVISTA for COSMOS) (McCracken et al. 2012; Laigle et al. 2016), but to a range of different depths, meaning both the colour space probed and the uncertainty on the photometry are quite inhomogeneous – even though the photometry extraction was done in a very homogeneous manner. The end result is two catalogues that span colour–magnitude space very differently, but with photometry very consistent for comparisons between individual galaxies in the two fields.

Similarly, the spectroscopic redshifts come from a range of sources, curated in the same way as in Adams et al. (2020).[6] The spec-zs are taken from the VVDS (Le Fèvre et al. 2013), VANDELS (McLure et al. 2018; Pentericci et al. 2018), zCOSMOS (Lilly et al. 2009), SDSS DR12 (Alam et al. 2015), 3D-*HST* (Skelton et al. 2014; Momcheva et al. 2016), Primus (Coil et al. 2011; Cool et al. 2013), DEIMOS 10K (Hasinger et al. 2018), and FMOS (Silverman et al. 2015) surveys. We would note that ML-based photo-z methods are reliant on the accuracy of the spectroscopic redshifts in the training sample. If the spec-zs used in the training process are inaccurate then ML methods will simply reproduce the incorrect spec-z values. For this reason, we only used the most secure spec-zs that have flags indicating high quality (confidence of ≥ 95 per cent). Where a source had a secure spec-z available from more than one survey, the mean of the secure redshifts was used. Furthermore, we found that the Primus spec-zs were often inconsistent with the higher resolution spec-zs and template-based photo-z's at $z > 1$. For this reason, we only use the $z < 1$ Primus spec-zs. The resulting combination of spectral and

[5]This training data is then split 50–50 into what is described as training and validation in Almosallam et al. (2016b), Almosallam et al. (2016a), but we shall refer to all the data used in the training process as the training set here.
[6]Which itself was constructed largely similarly to the Catalog of Spectroscopic Redshifts from the Hyper Suprime-Cam Subaru Strategic Program Public Data Release, https://hsc-release.mtk.nao.ac.jp/doc/index.php/dr1_specz/.

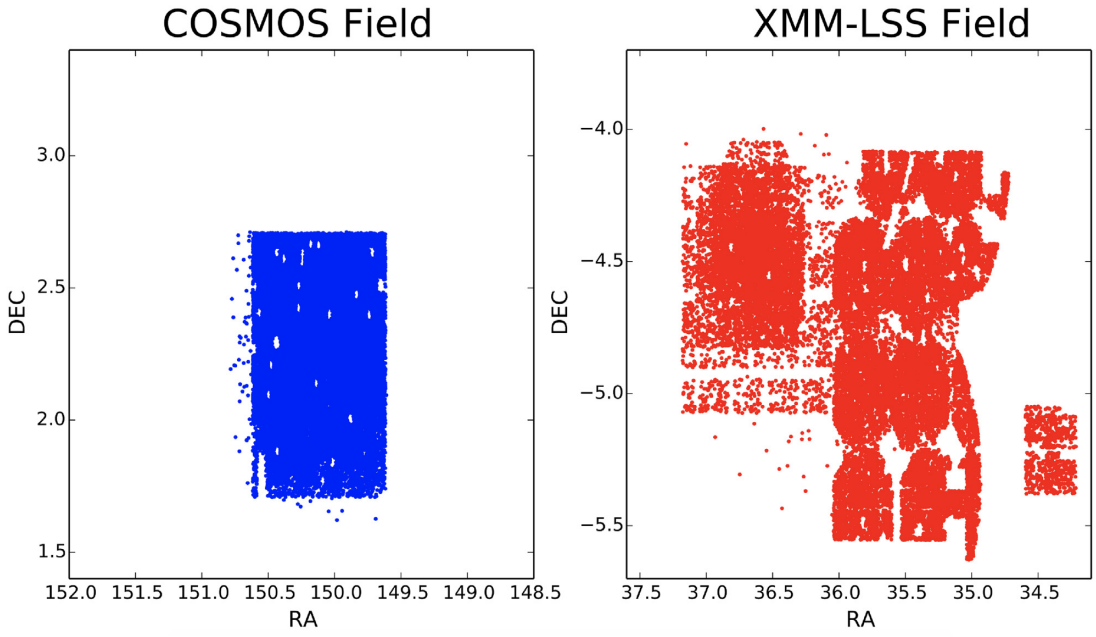

Figure 1. Field geometry of galaxies used in this analysis (each of which has a spectroscopic redshift). The 'COSMOS' galaxies are used for training, and the '*XMM*-LSS' galaxies for testing. The unusual field geometries result from the complex ways in which the various photometric and spectroscopic surveys have overlapped and intersected.

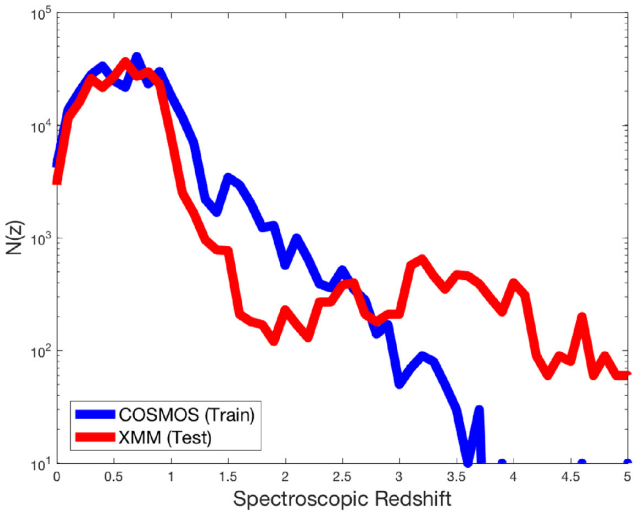

Figure 2. The spectroscopic redshift distributions of the galaxies in the two samples used in this study, the 'COSMOS'/training data, and the '*XMM*-LSS'/testing data.

photometric data used here is thus similar to that presented in Adams et al. (2020) and Bowler et al. (2020).

The *XMM*-LSS data set constructed includes data from the VANDELS survey, with redshifts in the $z = 1 - 4$ range, which are underrepresented in the COSMOS data for $z > 2.5$. This means (i) our training and test data have different colour distributions, and (ii) our test data has a high-redshift tail not present in the training data (see Fig. 2). There are 29 663 galaxies in the COSMOS training set and 24 534 galaxies in the *XMM*-LSS testing set.

In terms of stellar contamination, as we have tried to ensure that we are only using 'secure' redshifts, the vast majority of our sources should be extragalactic. More generally for photo-z's, stars can typically be removed based on a morphological cut (e.g. remove point sources, which will also remove quasars) or with a colour-cut. Conversely, our sample likely has a moderate number of AGN – both in terms of sources that are dominated by AGN light as well as Seyfert-like galaxies, whose photometry has large contributions from both galaxy starlight and a central nucleus. X-ray data is available in both the COSMOS field (Marchesi et al. 2016) and the *XMM*-LSS field (Chen et al. 2018), and could in principle be used to identify AGN (considered for some of the sources in this sample in Adams et al. 2020). We chose however not to use the X-ray data to remove AGN in this work on the rationale that ML methods should in principle be agnostic with regards to whether the source is a galaxy or an AGN – as long as there are similar sources with secure redshifts in the training set, it should be possible to give sources accurate ML-based photo-z's.[7] This conversely is not the case for template-based methods; if templates with sufficient AGN contribution are not included in the fitting process then in general, Seyfert-like galaxies can receive inaccurate redshifts (a point discussed in greater detail in Salvato et al. 2019). The converse strategy of separating a sample into AGN and non-AGN populations in advance is studied in Norris et al. (2019), who measure photo-z performance when X-ray detected sources are included or not included in their training data. We chose not to study AGN versus non-AGN performance separately in this work in order to focus on how to improve global performance, although making such divisions may be a useful method for some science goals.

The way the data sets are constructed from a range of photometric and spectroscopic sources means that they do not have single well-defined depths, and in general have different colour and redshift distributions. For approximate reference however, the 95th percentile faintest training (testing) sources have AB magnitudes $u = 26.9$

[7]There are in practice a few further complications with AGN e.g. temporal variability.

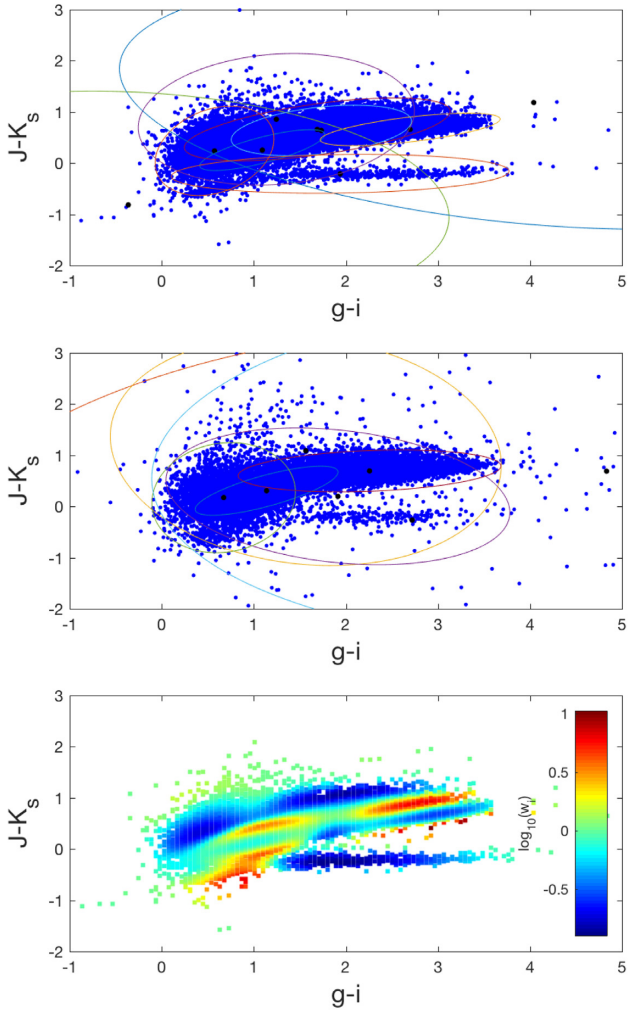

Figure 3. The (g-i)-(J-K$_s$) plane for the training (top subplot) and test (middle subplot) data. Ellipses show the identified mixtures (projected into 2D), with black marks showing the mixture centres. The lower plot shows the w_i values for the training data (essentially the ratio of the data density of the top and middle plots). A stellar locus is visible (cf. fig. 6a in Baldry et al. 2010).

(=26.9), $G = 25.6$ (=25.5), $R = 24.6$ (=24.4), $I = 24.1$ (=23.9), $Z = 23.8$ (=23.6), $Y_{\rm HSC} = 23.6$ (=23.6), $Y_{\rm VIDEO} = 23.6$ (=23.6), $J = 23.3$ (=23.5), $H = 23.2$ (=23.4), and $K_{\rm s} = 23.1$ (=23.3). Note that these are representative depths of spectroscopic data, not the imaging depths. No cosmology needs to be assumed in this work for any of the photometric or redshift calculations.

3 METHODS

3.1 Generalised cost-sensitive learning

A key property of GPz is the ability to tailor the weighting for different science applications. Almosallam et al. (2016b) presented three different weightings, 'Normal', 'Normalized', and 'Balanced'. Normal weighting weightings each galaxy the same (the 'null' option), Normalized weightings each galaxy in the training set by $1/(1 + z_{\rm spec})$, and Balanced weightings each galaxy inversely proportional to the number of galaxies with that redshift in the training set (e.g. galaxies with underrepresented redshifts are upweighted).

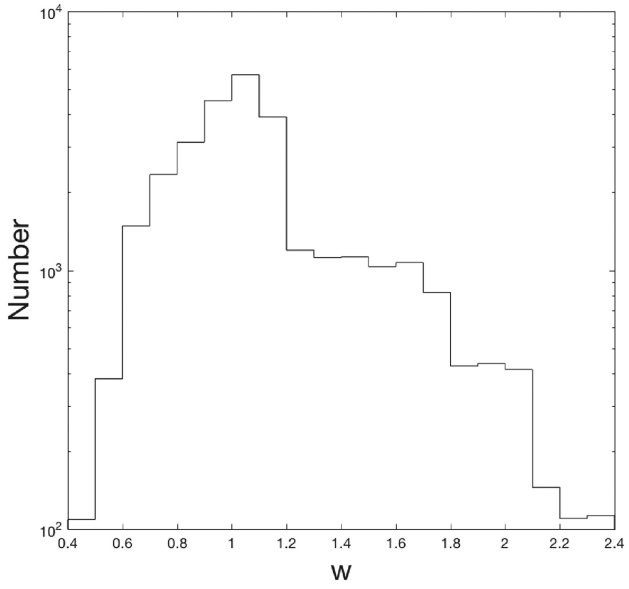

Figure 4. Distribution of w_i values found (note that w_i is capped at 20).

Here, we add to this generalised cost-sensitive learning (GCSL) weighting. We model the training set colour–magnitude space[8] with a GMM to find $p_{\rm train}(\mathbf{x})$, and do the same for the colour–magnitude space of the test set of data to find $p_{\rm test}(\mathbf{x})$, where $\mathbf{x}$ is the colour–magnitude photometry space of the data. This is essentially using the GMM to model the probability distribution in an unbiased way. We then set the weighting for GPz as:

$$w_i = \frac{p_{\rm test}(\mathbf{x}_i)}{p_{\rm train}(\mathbf{x}_i)}. \tag{3}$$

This potentially improves over balanced weighting (i) because it weightings galaxies in colour space rather than redshift space (different parts of colour space correspond to the same redshift), and (ii) it accounts for the colour–magnitude distribution of both the training set *and* the test set, rather than just the training set (it essentially generalises the approach used in Lima et al. 2008 and Duncan et al. 2018). A small number of extreme outlier galaxies end up with extremely high weightings, which were found to distort the entire training process, so we cap the maximum weighting at 20, and instead of equation 3, we in practice use $w_i = \frac{p_{\rm test}(\mathbf{x}_i)+\epsilon}{p_{\rm train}(\mathbf{x}_i)+\epsilon}$ with $\epsilon = 0.01$ to avoid extreme ratios at parts of parameter space with low data densities. Fig. 3 shows (a 2D projection of) the different colour spaces of the training and test data sets, the mixtures that were identified, and the corresponding w_i values. Fig. 4 shows a histogram of resulting w_i values.

If the test set and the training set have the same colour distribution, GCSL reduces to 'Normal' weighting as $p_{\rm train}$ and $p_{\rm test}$ will be identical and the w_i will be equal to one. In this scenario, for rare parts of colour–magnitude space, the performance will be low, but that is unimportant because a proportionately small number of galaxies in the test set will have that colour. 'Balanced' is near-equivalent[9]

[8] 10D, one magnitude and 9 colours.

[9] Only near-equivalent, as in Almosallam et al. (2016b), the weighting is based on redshift rather than colour.

to when the test set has a uniform distribution in colour space, or equivalently desiring a homogeneous performance in colour space. 'Normalised' makes the science judgement that we value percentage error on $1 + z$ rather than absolute error on $1 + z$, as opposed to weighting on sample distributions. One could in principle weighting by 'Normalised-GCSL' with $w_i^{\rm Normalised-GCSL} = w_i^{\rm GCSL}/(1 + z_i)$, although we found in practice this made relatively little difference.

From Fig. 3 it can be seen that there is some stellar contamination in the sample (the population at $J - K_s \approx -0.2$, see fig. 6 a in Baldry et al. 2010), despite them nominally being spectroscopically confirmed galaxies (as is likely to be the case at least to some degree in any putative Euclid and Rubin data set). However, it can be seen that the GMMs do identify the stars as being a separate component in colour–colour space (and down-weighting them, as there do not appear to be as many in the test spec-z sample), so future work could include a probability of each mixture to correspond to a stellar component.

3.2 Weighted validation

ML algorithms generally divide the labelled data into training and validation data sets when training the algorithm. The algorithm typically fits parameters to the training data, measures how well it does on the validation data, and then updates the complexity of the model (if over/under fitting) appropriately based on this. Since we are ultimately trying to predict the test data, making the validation data look more like the test data might help the model better tune itself. To do this, we take each galaxy with a spectroscopic redshift in the COSMOS data set and probabilistically select it to be either training or validation according to:

$$P_{\rm train} = \frac{1}{1 + w_i} \tag{4}$$

and

$$P_{\rm valid} = \frac{w_i}{1 + w_i}, \tag{5}$$

where $P_{\rm train}$ is the probability of the galaxy being assigned to the training data set, $P_{\rm valid}$ is the probability of the galaxy being assigned to the validation data set, and w_i is the weighting from equation 3. Weighted validation is similar in some regards to GCSL but may potentially cope slightly better with parts of parameter space with low data density.

3.3 Resampling

As already mentioned, GP-based ML methods, including GPz, by definition only give Gaussian uncertainties. This is typically, not realistic for photometric redshifts. To generate non-Gaussian ML posteriors with GPz, we trial a resampling method, where GPz is run a large number of times on slightly perturbed copies of the data to produce a large number of Gaussians, which are then summed to produce a non-Gaussian pdf. This is similar to the photometry perturbations of METAPHOR (Cavuoti et al. 2017), and the Monte Carlo approach of FRANKEN-Z (Speagle et al., in preparation).

The resampling method consists of the following steps:

(i) for each galaxy in the training and test sets, based on the uncertainty on each magnitude, resample a new magnitude value for each band;
(ii) train GPz on this resampled set of training data;
(iii) produce Gaussian pdfs for the resampled test data;
(iv) repeat steps 1–3 s times; and,
(v) average the s pdfs produced to obtain the final pdf.

This procedure lets us probe all the variations in the data. The hope is that for galaxies near 'cliffs', that are assigned one redshift with a very small uncertainty, but are very close in colour–magnitude space to galaxies that assigned very different redshifts also with small uncertainty, that the 'cliff' is repositioned slightly each time and the galaxies get more realistic pdfs. This approach can be thought of as a numerical method for GPz with noisy input. GPz conventionally accounts for noisy input, but still produces a single Gaussian. This approach approximates, at the limit of large 's', a multimodal Gaussian for a 'noisy' input of variance equal to the perturbation variance. Of the methods suggested in this text, this is by far the most computationally expensive, as it increases the runtime by a factor of s. However, because of the sparse framework used, GPz runs very fast and it remains practical to use $s \sim 100$ on a laptop (as used in this analysis), and would be viable to use a much higher s on a cluster as each 'run' of GPz can be parallelized.

3.4 Exploiting the population structure

Galaxies naturally fall into different populations. We can use the GMM to find natural galaxy populations, and assign probabilities of being in each population. Given a GMM, and for each galaxy a discrete probability density function for being in each of the populations, there are two natural ways to couple this to an ML algorithm, which we discuss below.

In the basic application of GPz, training the algorithm is $O(nm^2)$ in number of samples n and in number of basis functions m used to represent the mapping from photometry to redshift (in general a higher m will give better results, but take longer to train). Here, we propose using the GMM to split colour space into k regions, and then running GPz with m/k basis functions in each region (we call this 'GMM-Divide'). This typically will reduce the run time by a factor of $\sim k$ – each region will run $\sim k^2$ times faster, but k regions must be run. Regions are defined by which Mixture has the highest probability for that point in colour–magnitude space (a galaxy is in the region of the population it is most likely to be in). It is technically possible for the region for some Mixtures to be the empty set, e.g. tiny Gaussian within a broader one with a much larger amplitude. Regions are thus completely algorithmically determined by the GMM, with no human intervention; see Fig. 3. We define regions of the colour–magnitude distribution of the test set, and separate both the training and test set based on that region division.[10] There are many slight variations on exactly how the GMM algorithm can find populations; we found that letting it select in a magnitude and colour space (see Fig. 3) gave the best results, although precise implementation does not appear to make a large difference. The approach of splitting up parameter space into multiple regions has some similarities to the method of Masters et al. (2015), who use self-organizing maps (SOMs) rather than GMM as the unsupervised learning model used to divide up colour–magnitude space. There is also some overlap to the approach of separating AGN and non-AGN in advance, described in Norris et al. (2019); if X-ray data is available for the training process, then separating based on whether or not a source is X-ray detected is essentially dividing the sample into two populations based on a flux cut. The GMM here attempts to make such divisions in an unsupervised manner.

[10]Similar results are achieved when defining the regions based on the training set.

The other natural approach, which we investigated, would be to let galaxies be in multiple populations probabilistically, e.g. for each population, we train GPz on the entire training set with weightings proportional to how likely they were to have been drawn from that population i.e. $w_i = P(\text{galaxy } i \text{ in population } h)$. We then calculate a redshift pdf for each galaxy in the training set for each population. These are then summed in proportion to the probability that each galaxy in the test set was in each population (e.g. if there was a 75 per cent probability, the galaxy was drawn from population 1, and a 25 per cent chance it was drawn from population 2, the Gaussians from GPz run for population 1 and population 2 are summed with a 3:1 weighting). However, we found that the vast majority of galaxies were in a given population with probability near 1, and that this approach did not typically give better predictions.

3.5 Predicting log(z)

With the exception of Andromeda and other bodies in the Local Group, all galaxies have a positive redshift. ML-based redshift predictions can sometimes have non-trivial fractions of the resulting pdf be negative. One way of coping with this is to effectively set a strong prior that redshift has to be non-negative (essentially, cut off the negative part of the pdf). This unfortunately has the result of introducing a bias in the predictions for low-redshift galaxies (because this approach means you can only overestimate, never underestimate). We test the alternative of trying to predict log(z), which now takes all values, rather than z (considered in Section 3.3 of Almosallam 2017). Note that if $\Theta = \log(z)$, $\mathbb{E}(z) \neq \exp(\mathbb{E}(\Theta))$; Almosallam (2017) shows $\mathbb{E}(z) = \exp(\mathbb{E}(\Theta) + \frac{1}{2}\mathbb{V}(\Theta)))$ and $\mathbb{V}(z) = (\exp(\mathbb{V}(\Theta)) - 1) \times (\mathbb{E}(z))^2$.

This problem is to some degree similar to the issue of treatment of negative fluxes; fluctuations in the noise can lead to galaxies being assigned negative fluxes through the data reduction process (although hopefully, with uncertainties that make the measurements consistent with zero flux). When we train our photometric redshift model, we use log-fluxes, i.e. implicitly assuming uncertainties on fluxes are Gaussian in log-space. For negative flux values, we resample from a Gaussian centred on the negative value, with standard deviation the associated error, until a positive value is found. This obviously gives a slight bias for the faintest galaxies, however, for these sources the uncertainty on the flux is high and as the uncertainty is used within GPz, it has negligible effect. An alternative way to deal with negative fluxes is to use luptitudes (Lupton et al. 1999), as used with GPz in Desprez et al. (in preparation). Fluxes typically have Gaussian uncertainty in linear space near the detection threshold, but Gaussian uncertainty in log-space for brighter sources. The luptitude transformation essentially transforms the flux in such a way as to smoothly transition between these two regimes. Unfortunately, however, this requires consistency of detection threshold (which is used in the transformation), and the detection thresholds are not uniform between the different surveys we use to make up our two samples, so a luptitude in one would not necessarily correspond to the same luptitude in another. Both the methods employed here and luptitudes are not completely accurate, however the effect on our results is minimal.

4 RESULTS

In this section, we trial the methods discussed in Section 3 on the data described in Section 2.3. We calculate photometric redshifts using the following methods:

(i) Base performance, 'Normal' weighting
(ii) 'Generalised CSL' weighting (Section 3.1)
(iii) Weighted validation (Section 3.2)
(iv) Resampled base performance as per Section 3.3 with $s = 100$
(v) GMM-Divide (Section 3.4)
(vi) Modelling log(z) (Section 3.5)
(vii) 'All' - Weighted Validation', GMM-Divide and Resampling done simultaneously [11]

Fig. 5 shows spec-z versus photo-z for these methods, with varying performance. We also show for reference results if only the 70 per cent of data with lowest predicted uncertainty from 'All' is used ('Best').

4.1 Metrics

Figs 6, 7, and 8 compare the performance of our methods[12] as measured by the root mean squared error (RMSE), bias ($z_{\text{spec}} - z_{\text{phot}}$), and fraction of sources within 15 per cent of the true value (FR15), see table 1 in Gomes et al. (2018). These quantities are expressed as a function of 'fraction of the data' (the data is divided into bins of 'error bar size', so ten per cent corresponds to a bin of the best tenth of galaxies in terms of uncertainty size etc.). For RMSE, we also show the data as a function of spectroscopic redshift.

RMSE is approximately 0.05 for the data with the smallest uncertainties, and increases to about 0.25 for the data with the largest uncertainties. As a function of (spectroscopic) redshift, we find the RMSE $\approx$0.2 at $z \sim 0.75$, and $\approx$3 at high redshift. As a function of K_s-band magnitude, the RMSE is between 0 and 0.5 for most magnitudes, but rises rapidly for $K_s > 23$ and $K_s < 17$ galaxies, where there is less training data.[13] FR15 varies from essentially 100 per cent for the data with the smallest uncertainties, to around 85 per cent for the data with the largest uncertainties, dropping off sharply for the final 30 per cent of the data, as one would expect. Bias is less than 0.02 for most of the data, apart from the 20 per cent with the largest uncertainties. The different methods gave moderate variability; in particular 'Resampling' improved RMSE and FR15, but increased the bias. 'Weighted validation' was the only method to largely improve the bias.

Fig. 9 shows (i) the bias on the photo-z's as a function of both spec-z, and photo-z, and (ii) the relative improvement in bias, compared with 'Normal' weighting. It can be seen that the bias is between -0.2 and $+0.2$ for $0 < z < 1$, but steadily increases for higher redshifts. The redshifts are essentially biased towards where most of the data is, which gives rise to the redshift dependence (it is also essentially impossible to underestimate the redshifts of very low-redshift sources because $z > 0$). Whether bias as a function of spec-z or photo-z is more important depends on science goal. For a real science problem, only the photo-z's of the test data will be available, but bias as a function of spec-z can also be relevant depending on whether false-positives or false-negatives are more relevant for a high-redshift science goal etc.

It can be seen in Fig. 9 that 'Divide' reduces the bias at higher redshifts (which is inherited by 'All'). 'Log' slightly improved the bias at the highest redshifts, and 'Resample' impaired the bias (as

[11] There are a large number of ways that all different approaches could be combined; this method we found was the most logical and highest performing.
[12] Comparing the mode of the photo-z pdf to the true spectroscopic redshift where appropriate.
[13] We considered RMSE as a function of K_s because that was the band that detections were made in.

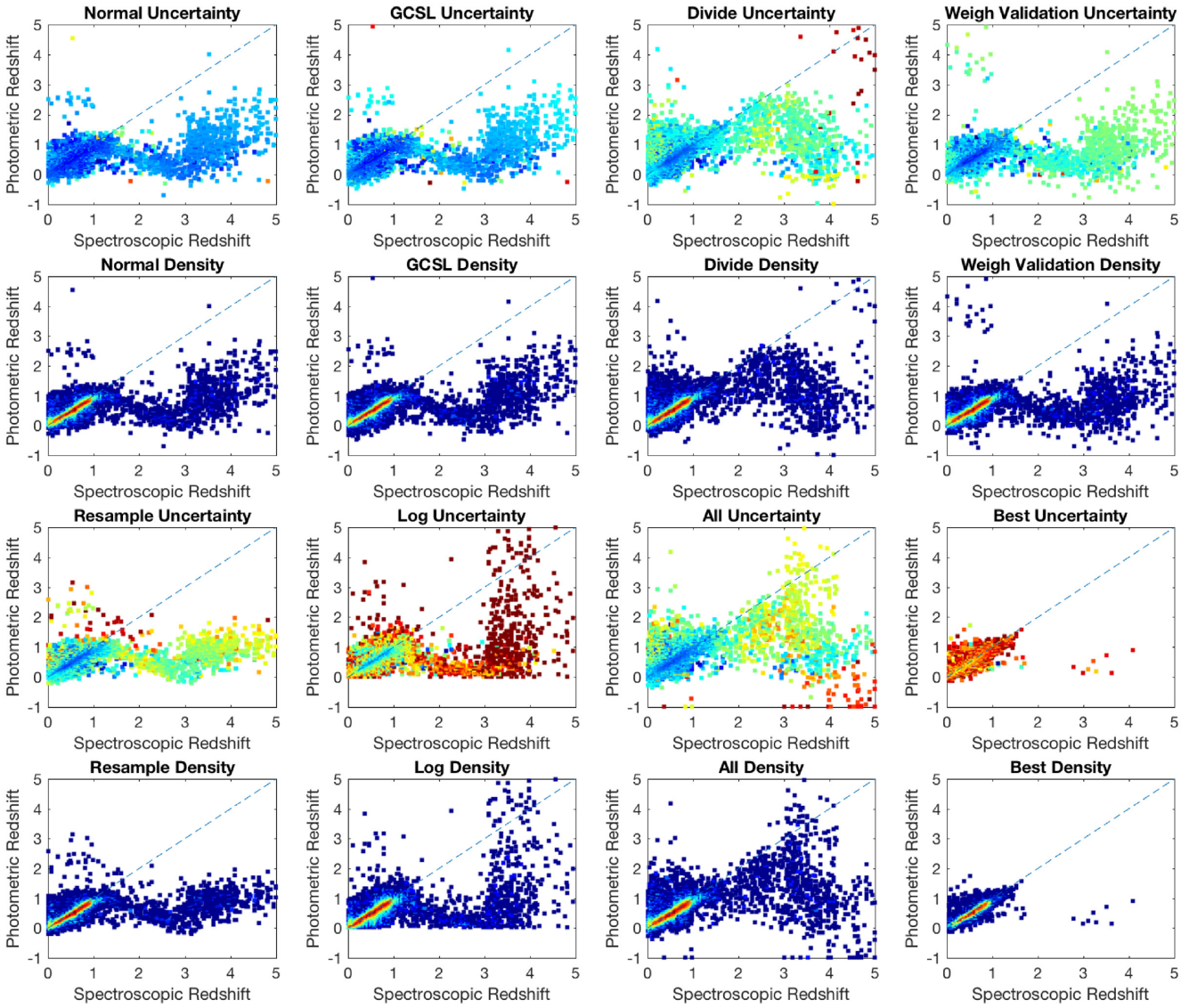

Figure 5. Spectroscopic redshift versus photometric redshift for the Normal implementation, GCSL, GMM-Divide, and Weighing Validation methods (top two rows, going left to right), and the resampling, log, 'All' and 'Best' methods (bottom two rows, going left to right). The first and third rows are coloured by the predicted uncertainty, the second and fourth by the data density.

a function of photo-z), but largely the methods apart from 'Divide' and 'All' didn't give any major improvements.

4.2 Quality of probability distributions

Probability integral transform (PIT) plots can be used to compare how well calibrated these pdfs are (D'Isanto & Polsterer 2018). The plot essentially shows a histogram of cumulative distribution function values at the spectroscopic redshift (e.g. for each galaxy, calculate what fraction of the pdf is less than the true value and plot a histogram of these values). Fig. 10 shows the PIT plot for all $z > 1$ test data, for both 'Normal' and 'All'. The asymmetric 'U' shapes show that the pdfs are slightly overnarrow and biased – but the fact that the 'All' curve is closer to a flat line (which would correspond an unbiased pdf) shows that the quality of the pdfs has indeed been improved at high redshift. For $z < 1$, the PIT plots for 'Normal' and 'All' are essentially identical.

4.3 Population redshift distribution

Finally, we plot the summed probability distributions[14] for both 'Normal' and 'All' to reconstruct estimates of redshift distributions for the test data in Fig. 11. Both 'Normal' and 'All' get the lower redshift 'hump' correct, but both struggle to identify the higher redshift hump. This is not surprising, given the training data, but it does show that using 'All' does manage to ∼triple the estimated number of high redshift galaxies, even if this estimate is still below the true number.

5 DISCUSSION

Our results show that it is possible to obtain significant improvements to the performance of GPz with comparatively little input (and no

[14]More sophisticated methods for estimating the sample redshift distribution do exist, but we do not consider here, e.g. Leistedt et al. (2016).

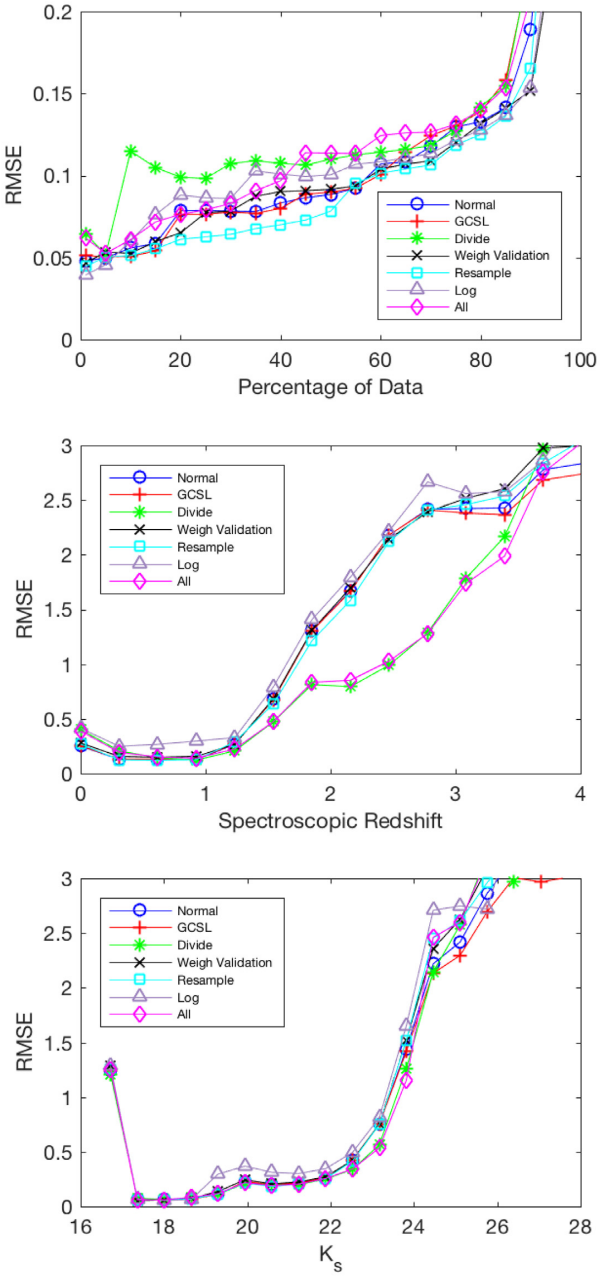

Figure 6. RMSE on on the photometric redshifts, as a function of (i) percentile of the data (top plot, zero with the smallest uncertainties and 100 the largest), (ii) spectroscopic redshift (centre plot), and (iii) K_s band magnitude (bottom plot).

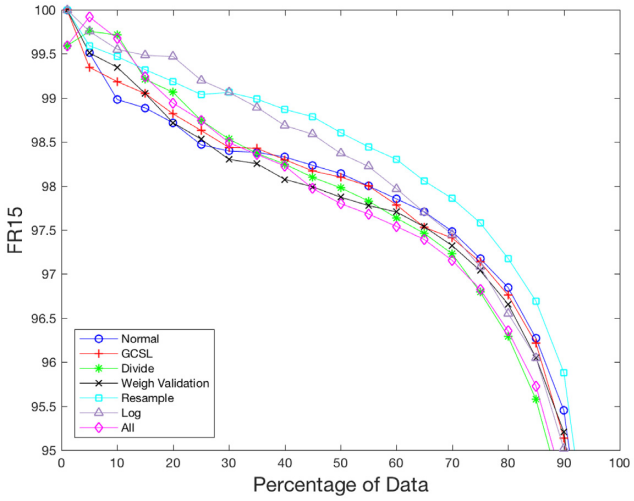

Figure 7. FR15 on on the photometric redshifts, as a function of percentile of the data (zero with the smallest uncertainties, 100 the largest).

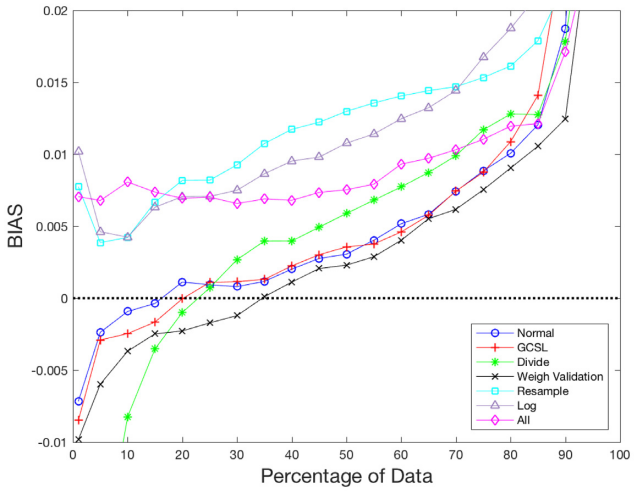

Figure 8. Bias on on the photometric redshifts, as a function of percentile of the data (zero with the smallest uncertainties, 100 the largest).

extra data), taking into account the differences in colour–magnitude distribution of the training and test data. Although we have discussed in the context of GPz, these methods could be easily extended to any other ML photo-z code. In particular, GMM-Divide is easily implemented and gives large improvements, and may be particularly valuable for Euclid and Rubin, where there will be billions of sources, and colour–magnitude space could be profitably split up into hundreds of sections, each still containing $\sim 10^5$–10^7 galaxies. The methods described here could also be combined with the pdf post-processing method described in Gomes et al. (2018).

Improvements at higher redshifts seem to largely come from 'Divide' (modelling different parts of colour–magnitude space separately based on a GMM). This is likely because the 'Divide' method identifies a population that corresponds to higher redshift galaxies. Because each population gets the same share of basis functions, this population gets more basis functions than it normally would in the straightforward implementation of GPz, and can be modelled more accurately. The lower-redshift sources get fewer basis functions than they otherwise would, but still get high performance, as they were only getting a large number of basis functions as it was where the bulk of the data was. An interesting comparison for future analyses where X-ray and/or radio data was available might be to compare the merits and demerits of unsupervised divisions like that discussed in this work, versus cuts explicitly designed to separate AGN and non-AGN.

Resample most improved the RMSE and FR15, but adversely affects the bias. Weight Validation was the best at improving bias. Using 'Divide', 'Resample', and 'Weight Validation' together ('All'), thus seemed the best way to improve both bias, scatter, and behaviour across all redshifts, and this is borne out by the 'All' results for Fig. 9.

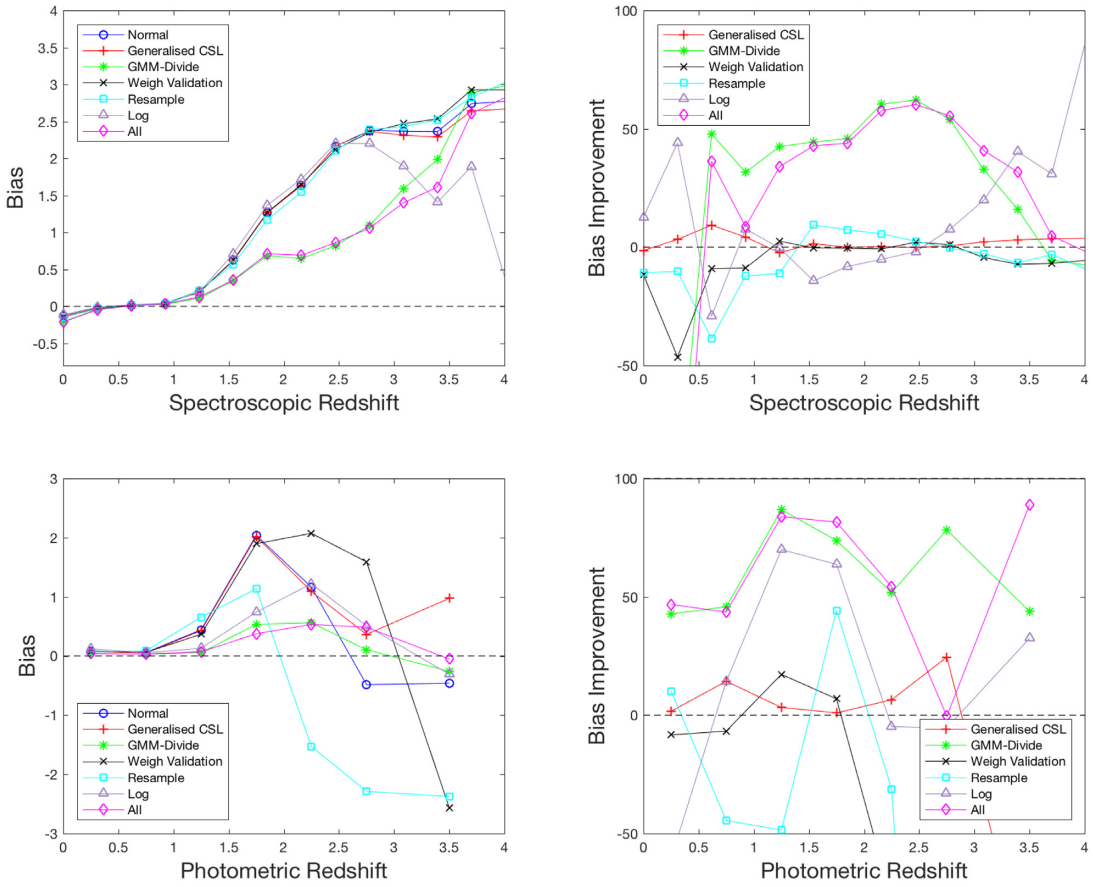

Figure 9. Bias on photometric redshift estimation (left column), and improvement in bias (right column) compared with 'Normal'; 0 per cent is no change in bias and 100 per cent is complete removal of bias. Top row shows the results as a function of spectroscopic redshift, bottom row shows as a function of predicted photometric redshift.

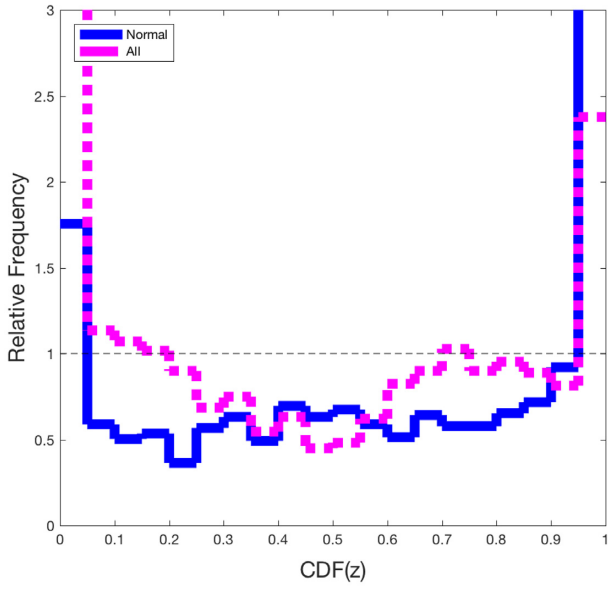

Figure 10. PIT plot for all the test data for 'Normal' and 'All' methods for photometric redshifts $z > 1$. Probability distributions are perfectly calibrated if they lie on the horizontal dashed line.

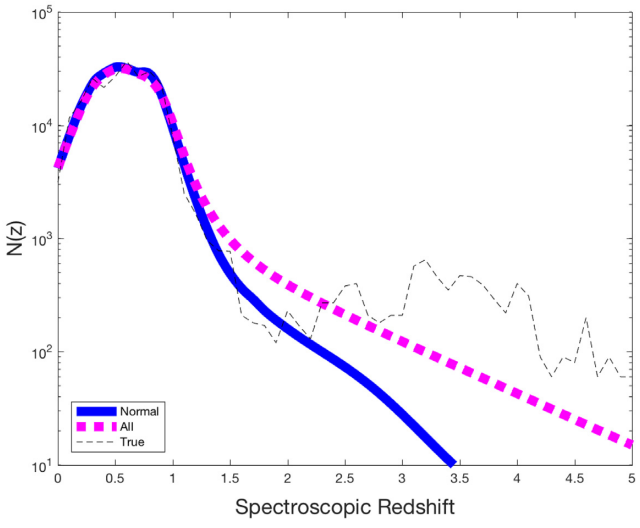

Figure 11. The true underlying redshift distribution (black dashed line), and the stacked pdfs from the (i) 'Normal' (full blue line) and (ii) 'All' (dotted magenta line) methods. It can be seen that 'All' better captures the high redshift distribution.

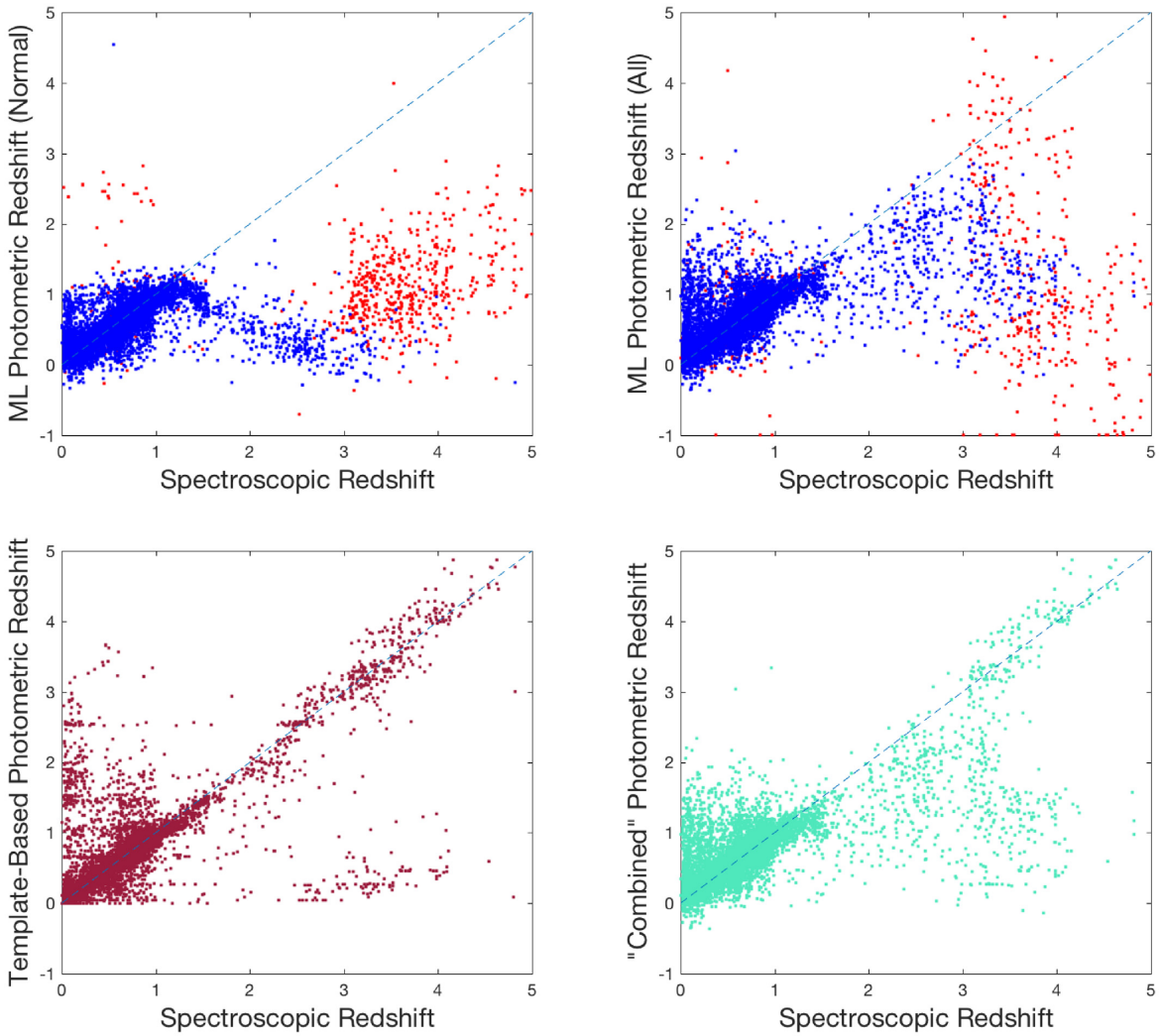

Figure 12. Comparison of ML and template methods. Top-left panel: Our 'Normal' ML predictions. Top-right panel: Our 'All' ML predictions. Bottom-left panel: The template-based predictions of Adams et al. (2020). Bottom-right panel: The 'Combined' predictions that incorporate both the ML and the template-based methods. For the plots in the top row, blue indicates galaxies in the interpolative regions of colour–magnitude space, and red those in the extrapolative regions.

Modelling $\log(z)$ successfully avoids negative redshifts, but typically led to poorer results. It also to some degree pushes the issues at $z = 0$ to $z = \infty$; galaxies up-scattered to high $\log(z)$ end up with predicted redshifts in the hundreds. If choosing to keep with modelling z rather than $\log(z)$ it, as discussed, can be tempting to simply 'cut off' the negative probability (e.g. declare the final pdf for the galaxy's redshift to be a Gaussian multiplied by a step-function). This can be an acceptable solution depending on science goals, but as discussed, it does have the unfortunate feature of biasing results near $z = 0$, e.g. making that cut forces the redshifts to be overestimated. This can be overcome if one requires the whole redshift distribution; some galaxies being assigned negative redshifts can be accepted, and then accounted for when finding the redshift probability distribution for the whole population e.g. with a hierarchical Bayesian model as per FRANKEN-Z (Speagle et al., in preparation).

Fig. 10 suggests that 'All' improves the calibration of the redshift pdfs for $z > 1$. As noted, however the asymmetric 'U' shapes show that the pdfs are generally slightly too narrow. Having accurate pdfs (as opposed to simply point estimates) is essential for many applications, including luminosity functions e.g. López-Sanjuan et al. (2017). More generally, at high redshift it is sometimes found that template-fitting methods can underestimate uncertainty (Dahlen et al. 2013; Salvato et al. 2019),[15] with ML methods usually producing more realistic pdfs (e.g. Brescia et al. 2019).

5.1 Comparison to template-based methods

The main focus of this paper is to identify how best to augment ML-based photo-z's, but it is none the less instructive to see how our ML-based predictions compare to template-based methods. In Fig. 12, we show how our 'Normal' and 'All' photo-z's compare to the template-based photo-z's calculated in Adams et al. (2020) using LEPHARE (Arnouts et al. 1999; Ilbert et al. 2006); it can be seen that the ML and the template-based methods each out-perform the other for different galaxies. In particular, it can be seen that typically the template-based method is better when GPz is in the extrapolative regime (i.e. where there is little or no training data).

[15] Although many works now mitigate against this, e.g. Buchner et al. (2015).

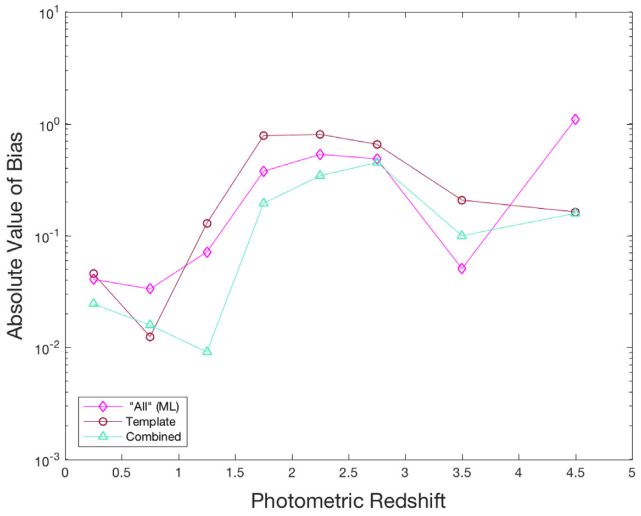

Figure 13. Absolute bias as a function of photometric redshift for (i) our 'All' ML photo-z's, (ii) the Adams et al. (2020) template-based photo-z's, and (iii) the 'Combined' values.

The variance in GPz has three components; ν (uncertainty from lack of data), $\beta_\star^{-1}$ (uncertainty from output noise, e.g. galaxies with different redshifts for the same magnitudes), and γ (uncertainty from input noise, e.g. uncertainty on the magnitudes). The extrapolative uncertainty ν is typically underestimated (furthermore the transition is smooth with no clear boundary), so we define predictions to be extrapolative if $\nu > 0.1 \times (\beta_\star^{-1} + \gamma)$, see Hatfield et al. (2020). This classification can be used to construct a 'Combined' photo-z estimate by using the GPz 'All' prediction if GPz is in the interpolative regime, and the LEPHARE photo-z if GPz is in the extrapolative regime. In Fig. 13, we show the bias of the 'Combined' method compared with the ML and template-based methods, demonstrating that the 'Combined' approach as expected outperforms the individual approaches. For the data sets used here, 4 per cent of the test data is in the extrapolative regime, although this fraction typically will vary depending on training and test data used.

Euclid and Rubin photometry will cover similar wavelengths to this study, over much larger areas (Rhodes et al. 2017). Future work will test these methods on the Euclid and Rubin data challenges, combining the pdfs constructed here with template-based pdfs using both the interpolative/extrapolative method described here, and the Hierarchical Bayesian Model approach of Duncan et al. (2018).

6 CONCLUSIONS

In this work, we began by constructing mock training and test data sets from CFHT, VISTA, and HSC data, both with spectroscopic redshifts, but with different colour and redshift distributions. This mimics the real issue of spectroscopic training sets having a different colour–magnitude distribution to the target distribution. We then discuss and illustrate several ways of using a combination of GMMs and the ML photometric redshift code GPz to obtain improved results over the baseline performance: (i) weighting the data appropriately for the colour–magnitude distribution differences, (ii) modelling different populations separately, (iii) using resampling methods, and (iv) making the validation data closer to the test data. We compare various metrics from the different methods, finding that respectable improvements in bias at higher redshifts ($z \gtrsim 1.5$) can be achieved with these relatively simple methods (and with no additional training data). In particular, modelling different parts of colour–magnitude space separately ('Divide'), Resampling, and weighting the validation data to look more like the test data seem to be the most effective and practical methods. These methods worked well even without removing AGN from the samples.

The key conclusions of this work are:

(i) Weighting schemes that take into account the different colour–magnitude distributions of galaxies in the training and test sets can reduce some of the bias in redshift estimation, particularly at high redshift (without any additional data).

(ii) Using GMMs can help speed up photometric redshift calculation and give improved precision for ML-based photometric redshift calculation.

ACKNOWLEDGEMENTS

PWH acknowledges funding from the Engineering and Physical Sciences Research Council (grant code EP/P01027X/1), generous support from the Hintze Family Charitable Foundation through the Oxford Centre for Astrophysical Surveys, and acknowledges travel support provided by the Science and Technology Facilities Council (STFC) for UK participation in Rubin through grant ST/N002512/1. IAA would like to acknowledge the support of King Abdulaziz City for Science and Technology. ZG is supported by a Rhodes Scholarship granted by the Rhodes Trust. NA acknowledges funding from STFC Grant code ST/R505006/1. RAAB acknowledges support from the Glasstone Foundation. This publication arises from research funded by the John Fell Oxford University Press Research Fund.

Based on data products from observations made with European Southern Observatory (ESO) Telescopes at the La Silla or Paranal Observatories under ESO programme ID 179.A- 2006. Based on observations obtained with MegaPrime/MegaCam, a joint project of Canada-France-Hawaii Telescope (CFHT) and Commissariat à l'énergie atomique (CEA)/Institut de Recherche sur les lois Fondamentales de l'Univers (IRFU), at the 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 work is 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.

This paper makes use of software developed for the Large Synoptic Survey Telescope (now known as Rubin). We thank the LSST Project for making their code available as free software at http://dm.lsst.org.

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper is based, in part, on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which

is operated by Subaru Telescope and Astronomy Data Center at National Astronomical Observatory of Japan. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics, National Astronomical Observatory of Japan.

DATA AVAILABILITY

The derived data generated in this research will be shared on reasonable request to the corresponding author.

REFERENCES

Abdalla F. B., Banerji M., Lahav O., Rashkov V., 2011, MNRAS, 417, 1891
Adams N. J., Bowler R. A. A., Jarvis M. J., Häußler B., McLure R. J., Bunker A., Dunlop J. S., Verma A., 2020, MNRAS, 494, 1771
Aihara H. et al., 2018, PASJ, 70, S8
Alam S. et al., 2015, ApJS, 219, 12
Almosallam I., 2017, PhD thesis, University of Oxford
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., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540
Baldry I. K. et al., 2010, MNRAS, 404, 86
Banerji M., Abdalla F. B., Lahav O., Lin H., 2008, MNRAS, 386, 1219
Beck R., Lin C. A., Ishida E. E. O., Gieseke F., de Souza R. S., Costa-Duarte M. V., Hattab M. W., Krone-Martins A., 2017, MNRAS, 468, 4323
Benitez N., 2000, ApJ, 536, 571
Benitez N. et al., 2004, ApJS, 150, 1
Bertin E., 2011, in Evans I. N., Accomazzi A., Mink D. J., Rots A. H., eds, ASP Conf. Ser. Vol. 442, Astronomical Data Analysis Software and Systems XX. Astron. Soc. Pac., San Francisco, p. 435
Bonfield D. G., Sun Y., Davey N., Jarvis M. J., Abdalla F. B., Banerji M., Adams R. G., 2010, MNRAS, 405, 987
Bowler R. A. A., Jarvis M. J., Dunlop J. S., McLure R. J., McLeod D. J., Adams N. J., Milvang-Jensen B., McCracken H. J., 2020, MNRAS, 493, 2059
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Brescia M., Salvato M., Cavuoti S., Ananna T. T., Riccio G., LaMassa S. M., Urry C. M., Longo G., 2019, MNRAS, 489, 663
Buchner J. et al., 2015, ApJ, 802, 89
Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483
Cavuoti S., Amaro V., Brescia M., Vellucci C., Tortora C., Longo G., 2017, MNRAS, 465, 1959
Chen C. T. J. et al., 2018, MNRAS, 478, 2132
Cheng T. Y., Li N., Conselice C. J., Aragón-Salamanca A., Dye S., Metcalf R. B., 2020, MNRAS, 494, 3750
Coe D., Benitez N., Sanchez S. F., Jee M., Bouwens R., Ford H., 2006, AJ, 132, 926
Coil A. L. et al., 2011, ApJ, 741, 8
Collister A. A., Lahav O., 2004, PASP, 116, 345
Cool R. J. et al., 2013, ApJ, 767, 118
D'Isanto A., Polsterer K. L., 2018, A&A, 609, A111
Dahlen T. et al., 2013, ApJ, 775, 93
Duncan K. J. et al., 2018, MNRAS, 473, 2655
Feldmann R. et al., 2006, MNRAS, 372, 565
Fernandez-Soto A., Lanzetta K. M., Chen H. W., Pascarelle S. M., Yahata N., 2001, ApJS, 135, 41
Fotopoulou S., Paltani S., 2018, A&A, 619, A14
Gomes Z., Jarvis M. J., Almosallam I. A., Roberts S. J., 2018, MNRAS, 475, 331
Hasinger G. et al., 2018, ApJ, 858, 77
Hatfield P., Rose S., Scott R., Almosallam I., Roberts S., Jarvis M., 2020, IEEE Trans. Plasma Sci., 48, 14
Hearin A. P., Zentner A. R., Ma Z., Huterer D., 2010, ApJ, 720, 1351
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Hoyle B. et al., 2018, MNRAS, 478, 592
Ilbert O. et al., 2006, A&A, 457, 841
Jaakkola T. S., Jordan M. I., 2000, Stat. Comput., 10, 25
Jarvis M. J. et al., 2013, MNRAS, 428, 1281
Jordan M. I., Ghahramani Z., Jaakkola T. S., Saul L. K., 1999, Mach. Learn., 37, 183
Kuhn M. A., Feigelson E. D., 2017, preprint (arXiv:1711.11101)
Laigle C. et al., 2016, ApJS, 224, 24
Laureijs R. et al., 2011, ESA report ESA/SRE(2011)12, Euclid Definition Study Report (Red Book). Available at: https://sci.esa.int/web/euclid/-/48983-euclid-definition-study-report-esa-sre-2011-12
Le Fèvre O. et al., 2013, A&A, 559, A14
Leistedt B., Mortlock D. J., Peiris H. V., Leistedt B., Mortlock D. J., Peiris H. V., 2016, MNRAS, 460, 4258
Lilly S. J. et al., 2009, ApJS, 184, 218
Lima M. et al., 2008, MNRAS, 390, 118
López-Sanjuan C. et al., 2017, A&A, 599, A62
LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
Lupton R., Gunn J., Szalay A., Lupton R. H., Gunn J. E., Szalay A. S., 1999, AJ, 118, 1406
McCracken H. J. et al., 2012, A&A, 544, A156
McLure R. J. et al., 2018, MNRAS, 479, 25
Marchesi S. et al., 2016, ApJ, 817, 34
Masters D. et al., 2015, ApJ, 813, 53
Momcheva I. G. et al., 2016, ApJS, 225, 27
Norris R. P. et al., 2019, PASP, 131, 108004
Peng H., Bai X., 2019, Astrodynamics, 3, 325
Pentericci L. et al., 2018, A&A, 616, A174
Rasmussen C. E., Williams C. K. I., 2006, Gaussian processes for machine learning. MIT Press, Cambridge, p. 248
Rhodes J. et al., 2017, ApJS, 233, 21
Sadeh I., Abdalla F. B., Lahav O., 2016, PASP, 128, 104502
Salvato M., Ilbert O., Hoyle B., 2019, Nat. Astron., 3, 212
Sawicki M. et al., 2019, MNRAS, 489, 5202
Schmidt S. J. et al., 2020, MNRAS, preprint (arXiv:2001.03621)
Silverman J. D. et al., 2015, ApJS, 220, 12
Skelton R. E. et al., 2014, ApJS, 214, 24

This paper has been typeset from a TEX/LATEX file prepared by the author.