Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **513,** 3719–3733 (2022)
Advance Access publication 2022 April 15

https://doi.org/10.1093/mnras/stac1042

Hybrid photometric redshifts for sources in the COSMOS and XMM-LSS fields

P. W. Hatfield [1★] M. J. Jarvis [1,2] N. Adams [3] R. A. A. Bowler [1,3] B. Häußler [4] and K. J. Duncan [5]

[1]*Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[2]*Department of Physics, University of the Western Cape, Bellville 7535, South Africa*
[3]*Jodrell Bank Centre for Astrophysics, University of Manchester, Oxford Road, Manchester M13 9PL, UK*
[4]*European Southern Observatory, Alonso de Cordova 3107, Vitacura, Casilla 19001, Santiago, Chile*
[5]*SUPA, Institute for Astronomy, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*

Accepted 2022 April 11. Received 2022 April 11; in original form 2021 November 19

ABSTRACT

In this paper, we present photometric redshifts for 2.7 million galaxies in the XMM-LSS and COSMOS fields, both with rich optical and near-infrared data from VISTA and HyperSuprimeCam. Both template fitting (using galaxy and active galactic nuclei templates within LePhare) and machine learning (using GPz) methods are run on the aperture photometry of sources selected in the K_s-band. The resulting predictions are then combined using a Hierarchical Bayesian model, to produce consensus photometric redshift point estimates and probability distribution functions that outperform each method individually. Our point estimates have a root mean square error of ∼0.08–0.09, and an outlier fraction of ∼3–4 percent when compared to spectroscopic redshifts. We also compare our results to the COSMOS2020 photometric redshifts, which contain fewer sources, but had access to a larger number of bands and greater wavelength coverage, finding that comparable photo-z quality can be achieved (for bright and intermediate luminosity sources where a direct comparison can be made). Our resulting redshifts represent the most accurate set of photometric redshifts (for a catalogue this large) for these deep multisquare degree multiwavelength fields to date.

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

1 INTRODUCTION

Many contemporary astronomical studies in extragalactic astrophysics and cosmology involve estimating the redshifts of large numbers of distant sources (typically galaxies). Galaxy redshift estimates are necessary to probe the time evolution of the Universe, as well as to correctly calculate galaxy properties – estimates of absolute luminosity and related measurements such as galaxy stellar mass rely on the estimated redshift being correct (Hsieh & Yee 2014).

Galaxy, and also active galactic nuclei (AGNs), redshifts can be calculated from their electromagnetic spectrum in two main ways, from spectroscopy or from photometry. *Spectroscopic redshifts* ('spec-z's') are calculated by detecting a known spectral (normally emission) line or feature with a spectrograph, and measuring the 'shift' from the known rest-frame wavelength/frequency. *Photometric redshifts* ('photo-z's') are calculated by measuring the brightness of the source in N broad wavelength ranges, and making a redshift prediction based on this coarse spectral data. Spectroscopic redshifts are far more precise than photo-z's (as long as the spectral feature is correctly identified) but are more costly (in terms of telescope time), so are generally restricted to much smaller samples. Spectroscopic and photometric redshift measurements are thus appropriate for different science goals (Fernandez-Soto et al. 2001).

There are two main methods for photometric redshift calculation (e.g. Salvato, Ilbert & Hoyle 2019); 'template fitting' and 'machine learning'. Template fitting methods typically use a number of galaxy or AGN template spectra (either empirical or synthetic), and use a χ^2-minimization-like method to find the 'best' redshift estimate. Template-fitting based codes in regular use 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[1]).

Machine learning (ML) photo- z methods take a highly empirical approach. The prediction task is treated as a supervised machine learning problem, where predictions must be made based on the photometry, and galaxies with known (usually spectroscopic) redshifts are used as labelled training data. Widely used machine learning 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 Redshifts (METAPHOR; Cavuoti et al. 2017), and many more.

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

[1]https://anaconda.org/astrorama/phosphoros

Template-fitting and machine learning methods typically only make identical predictions in the simplest of cases. In general, the methods make different predictions, with differing claims of levels of precision achieved, and giving different redshift probability distribution functions (PDFs), with corresponding advantages and disadvantages for different science cases (outlined in Salvato et al. 2019). Each method is typically reliable in different parts of colour–magnitude space. This presents an opportunity to achieve redshift prediction performance beyond that of each method individually, seeking 'the best of both worlds'. This has been demonstrated with a number of different approaches, for a number of different data sets (Brodwin et al. 2006; Carrasco Kind & Brunner 2014; Duncan et al. 2018b; Hatfield et al. 2020; Schmidt et al. 2020).

In this work, we present photometric redshift calculations that seek to achieve 'the best of both worlds' for the rich multiwavelength data sets that span the COSMOS and XMM-LSS fields, two of the most well-studied extragalactic fields. These redshifts will be key for a large range of extragalactic studies in these fields.

The structure of this paper is as follows. In Section 2, we describe the data used in this study. In Section 3, we describe the algorithms used, namely GPZ (Almosallam, Jarvis & Roberts 2016a; Almosallam et al. 2016b) and LEPHARE (Arnouts et al. 1999; Ilbert et al. 2006). In Section 4, we discuss our results , and we conclude in Section 5. AB magnitudes are used throughout (Oke & Gunn 1983).

2 DATA

The data in this work cover the COSMOS and *XMM–Newton* Large-Scale Structure (XMM-LSS) fields – see Fig. 1. These fields represent two of the deepest and widest fields used in extragalactic high-redshift survey astronomy, regularly used for a large number of wide-ranging studies e.g. Frayer et al. (2009), Darvish et al. (2017), Ata et al. (2021), Pacaud et al. (2007), Clerc et al. (2014), Chen et al. (2018), and Hale et al. (2018). The catalogues we use are described 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. The data are thus very homogenous across the two fields. Sources were selected in the K_s band (down to a limiting magnitude of $K_s = 24.8$ in COSMOS and $K_s = 23.9$ in XMM-LSS), and forced photometry was performed on all the other bands. 2 arcmin diameter circular apertures were used, which had an aperture correction applied by a model generated with PSFEx (Bertin 2011) for each band. In the COSMOS field 995 049 sources were identified, with 1674 689 sources identified in the XMM-LSS field (2669 738 in total).

The photometry used spans 10 filters; *u* (Canada–France–Hawaii telescope Large Area U-band Deep Survey, CLAUDS, for both COSMOS and the wider area XMM-LSS; Sawicki et al. 2019), *grizy* (Hyper Suprime-Cam Subaru Strategic Program, HSC-SSP, for both COSMOS and XMM-LSS; Aihara et al. 2017, Kawanomoto et al. 2018), and YJHK$_s$ (Visible and Infrared Survey Telescope for Astronomy, VISTA, VISTA Deep Extragalactic Observations, VIDEO for XMM-LSS; Jarvis et al. 2013 and UltraVISTA for COSMOS; McCracken et al. 2012). The photometric depths in the COSMOS (XMM-LSS[2]) field are $u = 27.0$ (=26.9), $g = 27.2$ (=27.0), $r = 26.8$ (=26.5), $i = 26.6$ (=26.4), $z = 25.9$ (=26.3), $y_{\rm HSC} = 25.5$ (=25.6), $Y_{\rm VISTA} = 25.5$ (=25.2), $J = 25.3$ (=24.7), $H = 25.0$ (=24.3), and $K_s = 24.8$ (=23.9).

[2]Depths are not quite identical for the three VISTA tiles in XMM-LSS; here the deepest data value is quoted, see table 1 in Adams et al. (2020) for more details.

The fields have spectroscopic redshifts (used in the training process for the machine learning based photometric redshifts) from a range of sources.[3] The spectroscopic redshifts are taken from the VVDS (Le Fèvre et al. 2013), VANDELS (McLure et al. 2018; Pentericci et al. 2018), Z-COSMOS (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. There were 25 268 spectroscopic redshifts in the COSMOS field, and 14 846 in the XMM-LSS field.

We would note that machine learning based photo-z methods are reliant on the accuracy of the spectroscopic redshifts in the training sample. If the spectroscopic redshifts used in the training process are inaccurate then machine learning methods will simply reproduce the incorrect values (see e.g. Stylianou et al. 2022). For this reason, we only used the most secure spectroscopic redshifts that have flags indicating high quality (confidence of ≥ 95 per cent). Where a source had a secure spectroscopic redshift available from more than one survey, the mean of the secure redshifts was used. Furthermore as discussed in Hatfield et al. (2020), we found that the Primus spectroscopic redshifts could be inconsistent with the higher resolution spectroscopic redshifts at $z > 1$. For this reason, we only use the $z < 1$ Primus spectroscopic redshifts.

[3]This spectroscopic data set is constructed largely similarly to the Catalogue 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/.

3 ALGORITHMS

3.1 Template fitting: LEPHARE

Our template-based photo-z's are calculated in a very similar manner to Adams et al. (2020), using LEPHARE (Arnouts et al. 1999; Ilbert et al. 2006). The only difference is that rather than using AGN templates to find a $\chi^2_{\rm AGN}$ to later use for classification/contamination control, we instead run LEPHARE twice, once with galaxy templates, and once with AGN templates, to obtain two template fitting based photo-z PDFs.

When finding a galaxy template-fitting photo-z PDFs the COSMOS SED template set (Ilbert et al. 2009) was used, where 32 templates were sourced from Polletta et al. (2007) with the GRASIL code (Silva et al. 1998) and from Bruzual & Charlot (2003). The templates cover a range of galaxy morphologies and spectral types (E, S0, Sa, Sb, Sc, Sd, Sdm) and have rest-frame wavelength ranges that cover our optical and near-infrared data set. Within the fitting process, each of these templates is allowed to be modified for the effects of dust attenuation using the Calzetti et al. (2000) attenuation law and an attenuation value in the range $E(B–V) = 0$–1.5. At each redshift, we use the Madau (1995) treatment for absorption by the intergalactic medium. For the AGN template-fitting photo-z PDF calculation, spectra for AGN from Salvato et al. (2009) were instead fit. Zero-point corrections to the photometry were made as in Adams et al. (2020). We do not address the potential impact of AGN variability on photo-z quality (e.g. Simm et al. 2015) in this work. For the LEPHARE calculations broad uniform priors over absolute magnitude ($-28 < M_{\rm abs} < -10$), redshift ($0 < z < 9$), and dust attenuation ($0 < E(B–V) < 1.5$), were used in this work, with an informative prior over redshift being introduced at the Hierarchical Bayesian (HB) combination stage, see Section 3.3. The fiducial

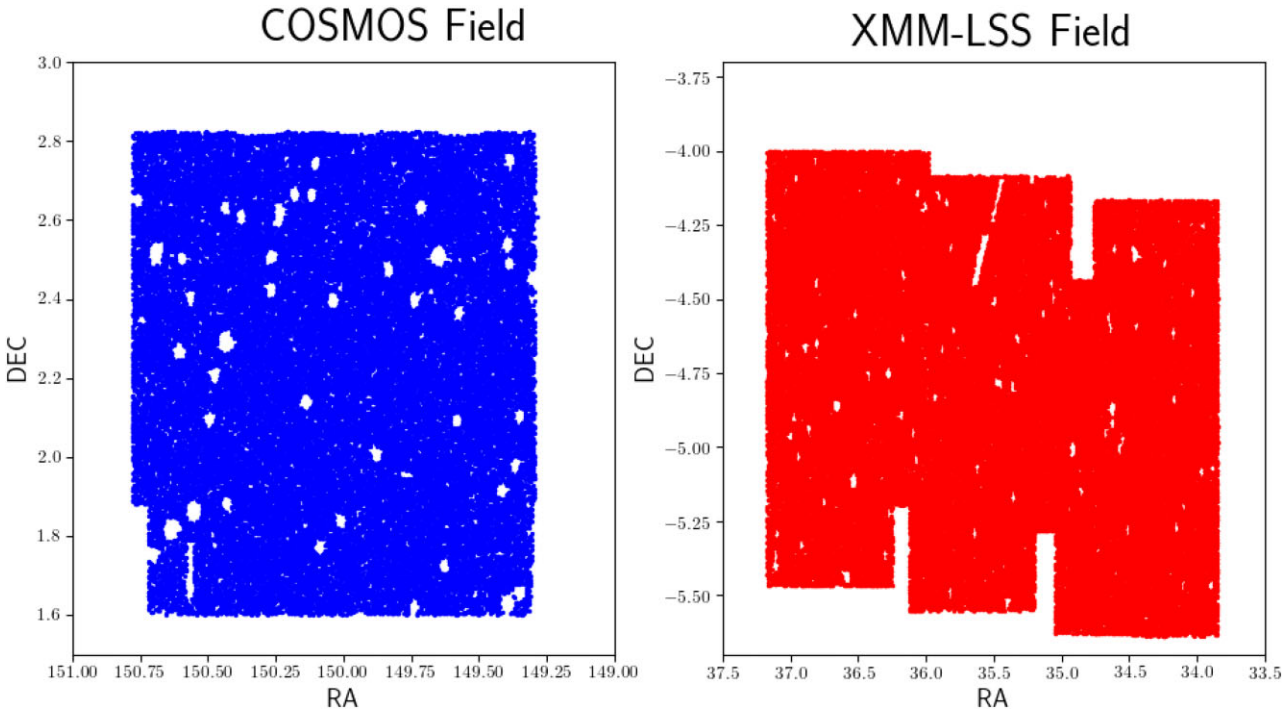

Figure 1. Field geometry of the source populations used in this study, spanning the COSMOS and XMM-LSS fields. Different scales are used for the two sub-plots. Holes and gaps in the fields are from bright nearby stars and artefacts in the imaging.

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

cosmology used was is a standard Flat cosmology of $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70$ km s^{-1} Mpc^{-1}.

3.2 Machine learning: GPz

GPz is a supervised machine learning algorithm developed for the problem of calculating photometric redshifts (Almosallam et al. 2016a,b). The algorithm is 'sparse Gaussian process' (GP) based, see e.g. Rasmussen & Williams (2006). The input data for the algorithm consists of sources with photometry and spectroscopic redshifts (the 'labels'). The algorithm is trained on this data, and then makes predictions for data with no spectroscopic redshift.

GPz has been used on a number of data sets e.g. Gomes et al. (2018), Duncan et al. (2018b), Zuntz et al. (2021), but (as with all algorithms), has some deficiencies, in particular (1) making poor predictions in parts of parameter space underrepresented in the training data and (2) only producing Gaussian PDFs (where in fact non-Gaussian, multimodal PDFs might typically better represent our uncertainty). Hatfield et al. (2020) (see also Duncan 2022) investigated a number of ways in which the bias introduced by these issues could be reduced. They found a combination (referred to as 'GMM-All') of (1) reweighting validation data to be closer to the target data, (2) dividing up the colour–magnitude space into regions and modelling each one separately, and (3) resampling the data many times based on the uncertainties on the photometry improved the resulting predictions. For the data set under consideration here, we calculate GPZ PDFs using GMM-All, rather than using the base GPZ PDFs. Missing bands and uncertainties on photometry were treated by adding the noise variance to the basis functions and constructing a joint distribution of the input parameters as per section 5 of Almosallam (2017).

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

3.3 Hierarchical Bayesian combination

We use an HB model similar to that described in Duncan et al. (2018a) and Duncan et al. (2018b) (which builds on Dahlen et al. 2013). This method seeks to combine the PDFs from *n* different redshift estimates[4] to achieve a consensus PDF that is

[4]The *n* photo-*z* estimates could be different methods within the same 'class' of estimator, or could be different classes of estimate completely. Duncan et al. (2018a) for example combine three different template based photo-*z* estimates ($n = 3$). In Duncan et al. (2018b), an ML based photo-*z* and a template based photo-*z* are combined ($n = 2$). In principle in future other independent approaches to redshift estimation could be incorporated, for example cluster-*z* (e.g. Rahman et al. 2015) and photo-geometric redshifts (e.g. Sonnenfeld 2022).

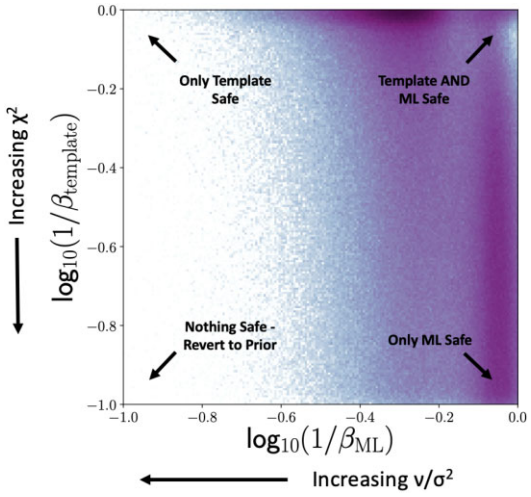

Figure 2. A simple diagram illustrating which photometric redshift methods are more reliable in different parts of colour–magnitude space when the template and the ML PDFs are combined (with the actual β_i values from the data plotted). When the template models fit well and the ML is interpolating, both the template and the ML PDFs are reliable. When the template models fit well, but the ML is extrapolating, only the template PDF is reliable. When the template models become invalid, but the ML still has adequate training data in that part of parameter space, then the ML PDF but not the template PDF is reliable. When the template models do not fit, and the ML is extrapolating, neither approach is reliable, and the best that can be done is to revert to some broad prior.

more accurate than the individual estimators. Note this method is different to that described in section 5.1 of Hatfield et al. (2020), which combined template and ML methods by accepting the ML prediction in the interpolative regime, and the template prediction in the extrapolative regime. The Hatfield et al. (2020) approach permitted use of knowledge of where we would expect the machine learning predictions to be reliable, and where we would expect it to be unreliable, but didn't enable the full information in the individual PDFs to be used. An alternative method of combining PDFs to that described in Duncan et al. (2018a,b) is using a Fréchet mean method, as described in Kodra (2019).

When calculating consensus PDFs, we first used an HB model to combine the galaxy and AGN template PDFs ($n = 2$), to produce a 'best' template-fitting redshift estimate PDF. This 'best' template-fitting redshift estimate PDF is then combined with the machine learning PDF with an HB model (again $n = 2$). See Duncan et al. (2018a) for a full description of the method, but for each source and each redshift estimate i we define:

$$P(z,\, f_{\rm bad})_i = P(z|\text{bad measurement})_i\, f_{\rm bad} + P(z|\text{good measurement})_i(1 - f_{\rm bad}), \quad (1)$$

where $f_{\rm bad}$ is a parameter describing the probability of an estimate being incorrect, $P(z|\text{good measurement})_i$ is the probability distribution assuming that the estimator is correct (i.e. equal to the probability distribution from the estimator), $P(z|\text{bad measurement})_i$ is the probability distribution assuming that the estimator is incorrect (typically chosen as an appropriate prior), and i indexes the n methods (so i indexes over 1 and 2 for the AGN and the galaxy template fits for the first HB model, and then over 1 and 2 again for the template-based and the ML based PDFs for the second HB model). $P(z, f_{\rm bad})_i$ thus describes a PDF from that estimator, allowing for the possibility of the estimate being incorrect.

These n distributions are then combined in the following way:

$$P(z,\, f_{\rm bad}) = \prod_{i=1}^{n} P(z,\, f_{\rm bad})_i^{1/\beta_i}, \quad (2)$$

where the β_i are constants that encode the weights and covariances between the different measurements. Equation (2) represents a small generalization over Duncan et al. (2018a), for which each estimate had the same β values. The product iterates over the n different photo-z PDFs being combined, which are indexed by i.

Finally $f_{\rm bad}$ is marginalized over to get a final PDF:

$$P(z) = \int_{f_{\rm bad}^{\rm min}}^{f_{\rm bad}^{\rm max}} P(z,\, f_{\rm bad}) {\rm d} f_{\rm bad}, \quad (3)$$

where $f_{\rm bad}^{\rm min}$ and $f_{\rm bad}^{\rm max}$ are constants representing the minimum and maximum of the range marginalized over.

The choice of β_i characterizes the weighting and degree of correlation between estimates. When all the β_i are chosen to be equal (to some β), the two extremes are $\beta = 1$ and $\beta = \frac{1}{n}$. $\beta = 1$ corresponds to simply multiplying the PDFs together i.e. treating them as completely independent and multiplying the probabilities. $\beta = \frac{1}{n}$ corresponds to taking the geometric mean of the estimates for the case where they are fully covariant (based completely on the same underlying data). These extremes can to some degree be thought of as corresponding to 'AND' ($\beta = 1$, all the estimates are giving independent information that should be incorporated into the prediction), and 'OR' ($\beta = \frac{1}{n}$, not all the predictions can be independently true) (Duncan et al. 2018a).

3.4 Combining galaxy and AGN template-based PDFs

Our galaxy and AGN template-based PDFs are highly covariant in that they are based on the same data (the photometry) – the only difference is the modelling. If we were agnostic between the galaxy and AGN PDFs, $\beta = \frac{1}{2}$ would be a natural choice (as $n = 2$). However, for individual galaxies, we are not agnostic – we can use the χ^2 to indicate which of the galaxy and AGN templates were better fitting. We hence allow the β_i to vary, requiring that $\sum \frac{1}{\beta_i} = 1$ (essentially generalizing the geometric mean to the weighted geometric mean). We choose

$$\frac{1}{\beta_{\rm galaxy}} = \frac{\exp(-\chi^2_{\rm galaxy}/2)}{\exp(-\chi^2_{\rm galaxy}/2) + \exp(-\chi^2_{\rm AGN}/2)} \quad (4)$$

and

$$\frac{1}{\beta_{\rm AGN}} = \frac{\exp(-\chi^2_{\rm AGN}/2)}{\exp(-\chi^2_{\rm galaxy}/2) + \exp(-\chi^2_{\rm AGN}/2)} \quad (5)$$

to reflect the probabilities implied by the χ^2 values. Note that when $\chi^2_{\rm galaxy} = \chi^2_{\rm AGN}$ we recover $\beta_i = \frac{1}{n} = \frac{1}{2}$, and in the limit of $\chi^2_{\rm galaxy}/\chi^2_{\rm AGN} = 0$ (high confidence that the source is a galaxy not an AGN) we find $\frac{1}{\beta_{\rm galaxy}} = 1$ and $\frac{1}{\beta_{\rm AGN}} = 0$ (and vice versa). We also note that one could treat all individual galaxy and AGN templates in this way and obtain a final PDF. However, for the purposes of this paper, we consider the simpler combination, noting that one could also in principle impose a prior based on the expected number of galaxies and AGN at any given epoch i.e. the luminosity function – however as many luminosity functions are based on photometric redshifts one risks circular arguments.

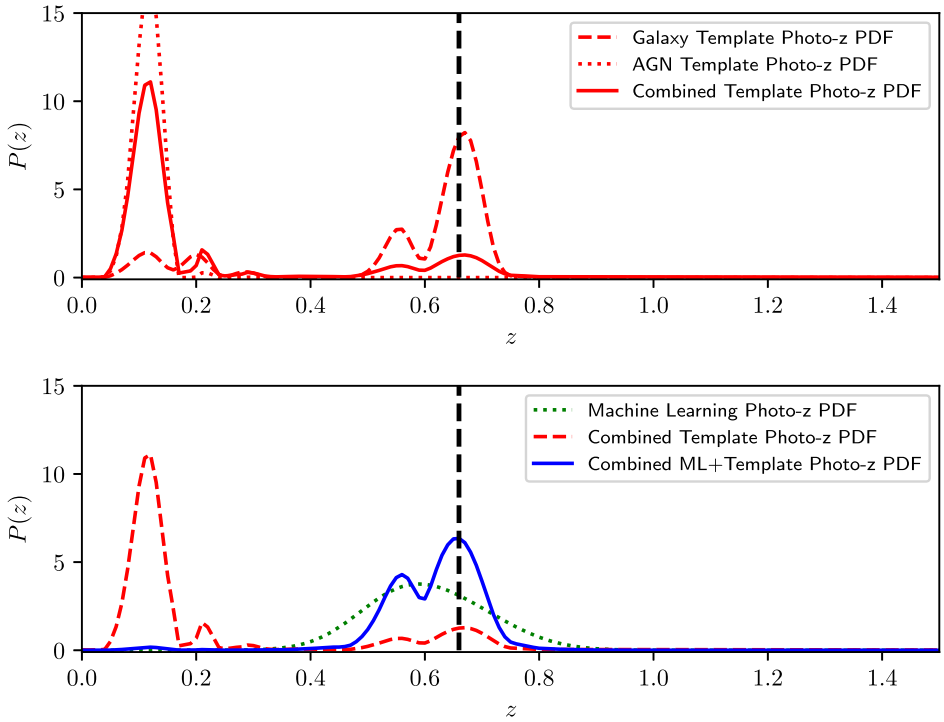

Figure 3. Illustrative example of how the HB model used combines PDFs. The top panel shows the combination of the galaxy and AGN photo-z template PDFs into a consensus template fitting PDF. The lower panel shows the combination of the consensus template fitting PDF with the ML PDF. The vertical line shows the spectroscopic redshift.

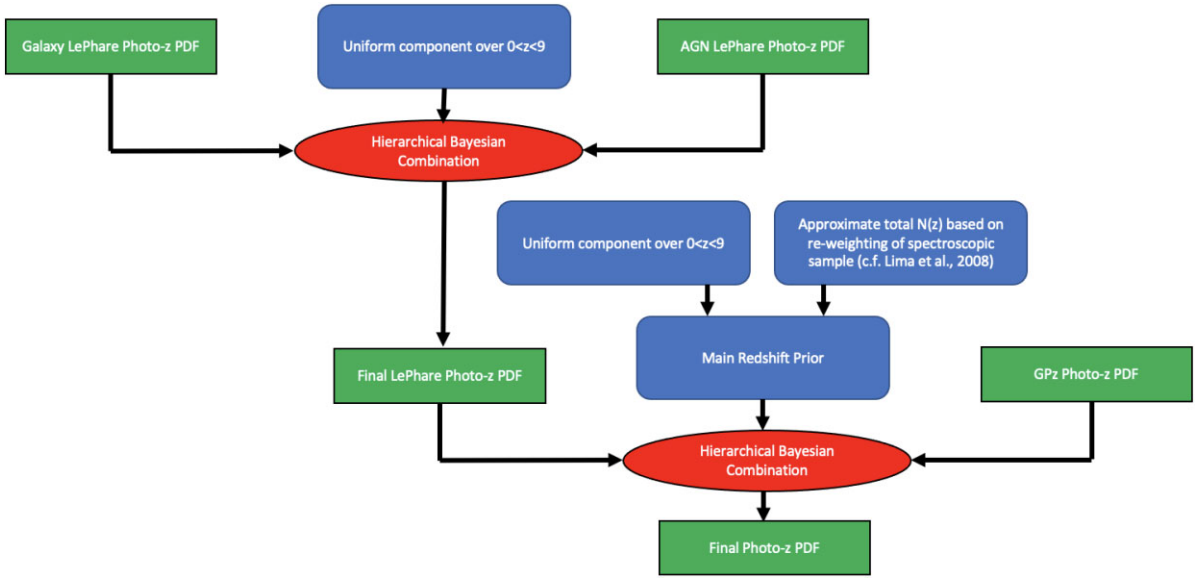

Figure 4. Simple flow chart schematic to illustrate how our final PDFs are constructed. First the galaxy and AGN LePhare template-based PDFs are combined with an HB model. Then the resulting PDF is combined with the GPz ML PDF with a second HB model [using a prior that is a weighted combination of a uniform distribution and the approximate sample $N(z)$].

3.5 Combining template based and machine learning based PDFs

When combining the template-fitting PDF and the ML PDF ($n = 2$), one choice would be to use $\beta_i = 1$ as the estimates are highly independent (the template fitting method does not have access to the information contained within the spectroscopic training data, and the ML method does not have access to our knowledge of the physics implicit within the templates).[5] However, we do have estimates of the 'reliability' of the two methods (as opposed to simply the associated uncertainties). Template fitting predictions, uncertainties,

[5]They are not completely independent, as they both use the same photometry, so scatter on the magnitudes for the two estimates are correlated. However, uncertainty on photometry is a sub-dominant source of photo-z uncertainty.

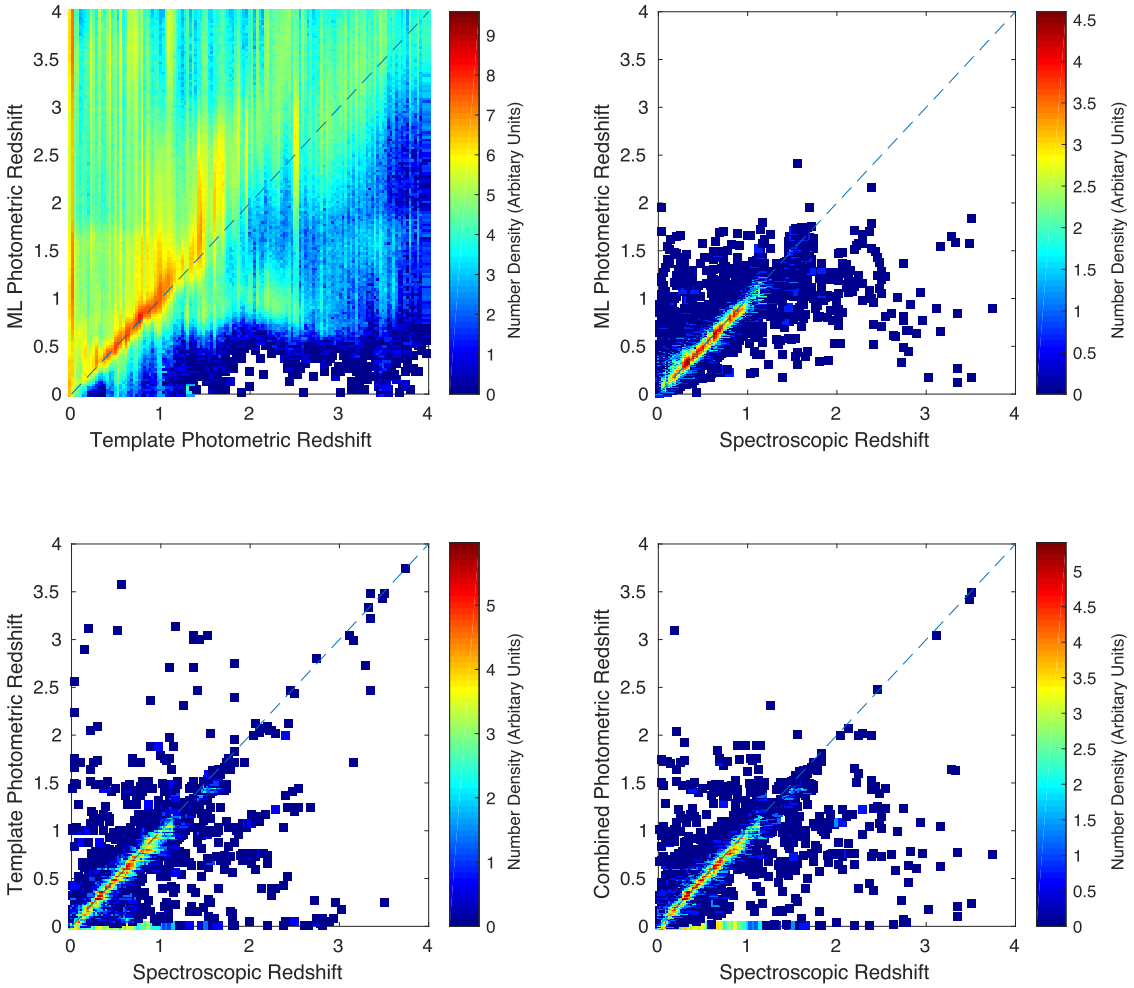

Figure 5. Comparisons of the $z_{\rm phot}$ (from machine learning, template fitting, HB combination) and $z_{\rm spec}$. The top left plot compares the machine learning and template fitting $z_{\rm phot}$, the other three plots compare the $z_{\rm phot}$ predictions to $z_{\rm spec}$. Note that not all sources have a $z_{\rm spec}$, so there are many more points in the top left plot. The diagonal dashed line shows a one-to-one correspondence (if photo-z predictions perfectly agreed with the spectroscopic redshifts).

and PDFs are likely to be unreliable when the χ^2 of the best-fitting template and redshift is high e.g. the photometry is not well fit by any template. Similarly, as discussed in Hatfield et al. (2020), the ML predictions become less reliable in the extrapolative regime, which can be quantified by how much of the uncertainty is due to lack of data in that part of parameter space compared with the total uncertainty (which includes uncertainty from the photometry, as well as intrinsic scatter in output redshift). Thus, we use $1/\beta_{\rm template} = \exp(-\frac{\chi^2}{2})$ and $1/\beta_{\rm ML} = 1 - \frac{\nu}{\sigma^2}$ (where ν is the variance from the lack of data using GPz, σ^2 is the total GPz variance, see Almosallam et al. 2016a and section 5.1 of Hatfield et al. 2020). In the limit of both the ML and template-fitting being reliable the $1/\beta_i \rightarrow 1$, in the limit of templates fitting well but extrapolating far from the spectroscopic training data $1/\beta_{\rm template} \rightarrow 1$, $1/\beta_{\rm ML} \rightarrow 0$, in the limit of no template fitting well, but there being sufficient training data in that part of parameter space $1/\beta_{\rm template} \rightarrow 0$, $1/\beta_{\rm ML} \rightarrow 1$, and finally where no template fits well and there is no nearby training data $1/\beta_{\rm template} \rightarrow 0$, $1/\beta_{\rm ML} \rightarrow 0$ and the resulting PDF reverts to the prior. In Fig. 2, we show a schematic for how the PDFs are combined differently in different parts of parameter space, as well as how our sources actually cover the space. Note that most galaxies have at least one reliable photo-z method. Fig. 3 shows the template and ML based PDFs for a sample galaxy, and the PDFs that result in the HB combination.

We use $f_{\rm bad}^{\rm min} = 0$ and $f_{\rm bad}^{\rm max} = 0.05$ as per Duncan et al. (2018a). In other words, we do not select a value for $f_{\rm bad}$ itself, but instead select the range over which it is marginalized. $f_{\rm bad}^{\rm min}$ and $f_{\rm bad}^{\rm max}$ could be chosen by a tuning process (considered in Duncan et al. 2018a), although this presents difficulties due to the differences between the training and test data, so there is no guarantee that any optimal values found would actually be optimal for the target data. The values from Duncan et al. (2018a) were partially based on a tuning process, but also based on typical outlier fractions for their data (and the surveys considered here have comparable properties and outlier fractions). Both we and Duncan et al. (2018a) found that extremely high values of $f_{\rm bad}^{\rm max}$ (e.g. $f_{\rm bad}^{\rm max} \geq 0.5$) gave worse predictions, but fine tuning beyond choosing a value of approximately 0.05/a value comparable to the expected outlier fraction made negligible difference. Finally, for $P(z|\text{bad measurement})_i$, we use a linear combination of a uniform distribution in redshift over (0,9) (weight 0.001), and the implied approximate sample redshift distribution (weight 0.999). The approximate sample redshift distribution is constructed by taking the redshift distribution of the spectroscopic sample, where each source is weighted by how many sources of similar colour–magnitudes there are in the full sample (based on the Gaussian Mixture Models from

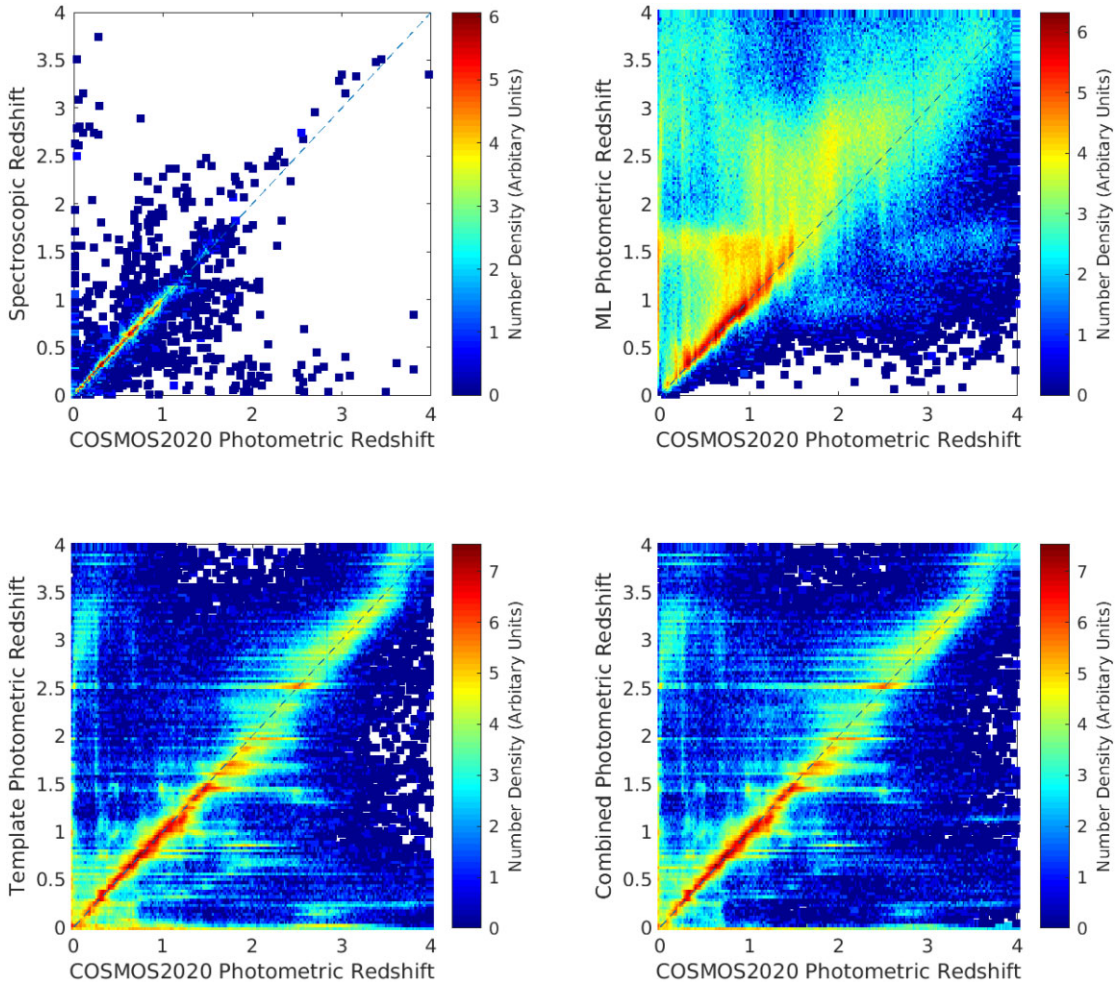

Figure 6. Comparisons of the $z_{\rm phot}$ (from machine learning, template fitting, HB combination) and $z_{\rm COSMOS2020}$. The top left plot compares the spectroscopic redshifts and the COSMOS2020 redshifts, the other three plots compare the $z_{\rm phot}$ predictions to $z_{\rm COSMOS2020}$. Note that not all sources have a $z_{\rm spec}$, so there are many fewer points in the top left plot. The diagonal dashed line shows a one-to-one correspondence (if photo-z predictions perfectly agreed with the spectroscopic redshifts).

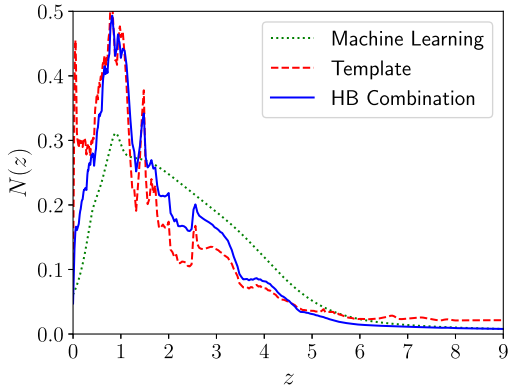

Figure 7. The stacked PDFs from the ML, template, and HB combination photo-z calculations.

the GMM-All calculation), similarly to as in Lima et al. (2008).[6] The addition of the uniform prior is necessary to avoid the distribution being zero for sparse (mainly higher) redshifts. If not included, the prior becomes zero (as opposed to just very small) for redshifts higher than the highest spectroscopic redshift in the sample, which is unphysical (and thus even if the template estimate was highly secure at a higher redshift, the zero-weight prior would dominate and lead to a lower redshift being assigned). Results are relatively insensitive to exact choice of relative weighting of the two components of the prior and were chosen to approximately reflect how many high redshift sources might be expected, to an order of magnitude. Finally, see Fig. 4 for a simple diagram illustrating how the two HB models used here are connected.

[6]Note that this choice of prior is very similar to the 'trainZ' estimator considered in Schmidt et al. (2020) – a simple estimator that assigned each source an identical redshift PDF, that of the population as a whole.

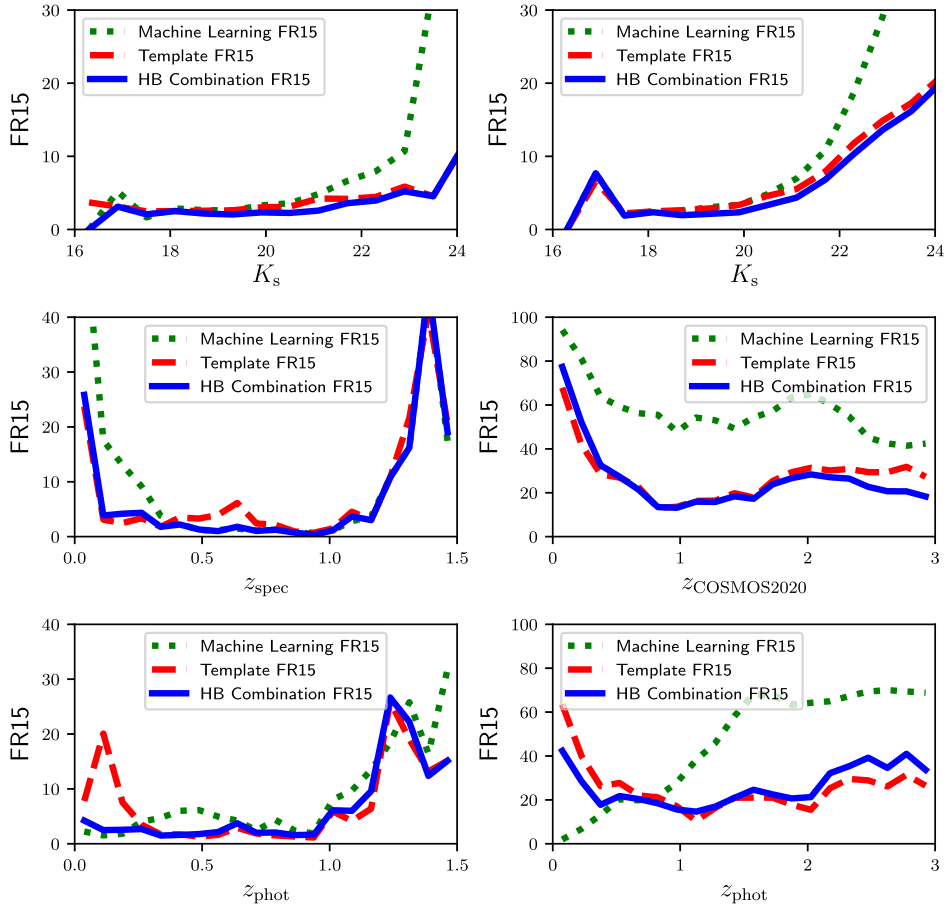

Figure 8. The outlier fraction as a function of 'true' redshift for our three photo-z predictions. The first row shows as a function of K_s-band magnitude, the second row shows as a function of spectroscopic or COSMOS2020 redshift, and the third row shows as a function of photometric redshift. In the left column the 'true' redshifts are the spectroscopic redshift, in the right column the 'true' redshifts are the COSMOS2020 redshifts (the columns are plotted with different redshift ranges, corresponding to the redshift ranges where there were sufficiently large numbers of galaxies). Note prediction performance is likely to be poorer for the sample as a whole as the spectroscopic sample is not representative.

3.6 Selection of 'best' redshift estimate

There is normally no single 'best' choice of point estimate of redshift from a (generally imperfect) redshift PDF,[7] as the statistical properties required depend on science goal e.g. sensitivity to outliers, redshift range of interest etc. Here, we quote the mode of the PDF as the 'best' estimate of redshift, although other point-estimates (e.g. median, mean) can be readily calculated from the PDFs.

4 RESULTS

In this section, we apply the methods discussed in Section 3 to the data described in Section 2. To test the quality of our calculations, one approach is to compare predictions to the spectroscopic sample. However, because not all galaxies have a spectroscopic redshift, the comparisons that include z_{spec} represent a biased sub-set of the whole data set. Furthermore, the ML based predictions were trained on this sample, so we would expect these predictions to be much better than for unseen data (even of the same colour–magnitude distribution). Good performance on the spectroscopic sample is thus necessary but not sufficient.

In this section for the calculation of metrics we remove from the sample sources in the stellar locus (that are likely to be stars) as defined in Jarvis et al. (2013) (which follows the approach of Baldry et al. 2010). We also remove sources with $\chi^2_{Star} < \min(\chi^2_{QSO}, \chi^2_{Galaxy})$. This reduces the COSMOS and XMM-LSS samples to 815 673 and 1557 392, respectively. The sources in the stellar locus are still assigned photometric redshifts for the released catalogue e.g. in case the stellar classification is incorrect due to scatter.

In addition to the spectroscopic sample, our photo-z calculations can also be compared to the COSMOS2020 photo-z catalogue of Weaver et al. (2022) (an update to the COSMOS2015 redshifts of Laigle et al. 2016). We use the LePhare redshifts based on the 'Classic' catalogue. This data set includes optical and NIR data of similar bands to those used in this data set over the COSMOS field. However, in addition to broad-band photometry, it also used a number of medium- and narrow-band filters for the calculation of template-based photometric redshifts (36 bands used in total). Thus, the COSMOS2020 photo-z calculations represent an intermediate category between spectroscopic redshifts and the photo-z we have calculated, in the sense that they are likely (a) more accurate than our template photo-z, but less accurate than spectroscopic redshifts, and (b) less numerous/deep than our sample, but more numerous/deep than the spectroscopic redshift sample. However, they are not used in the training process, and have a different colour–magnitude

[7] See e.g. discussion in Duncan et al. (2019).

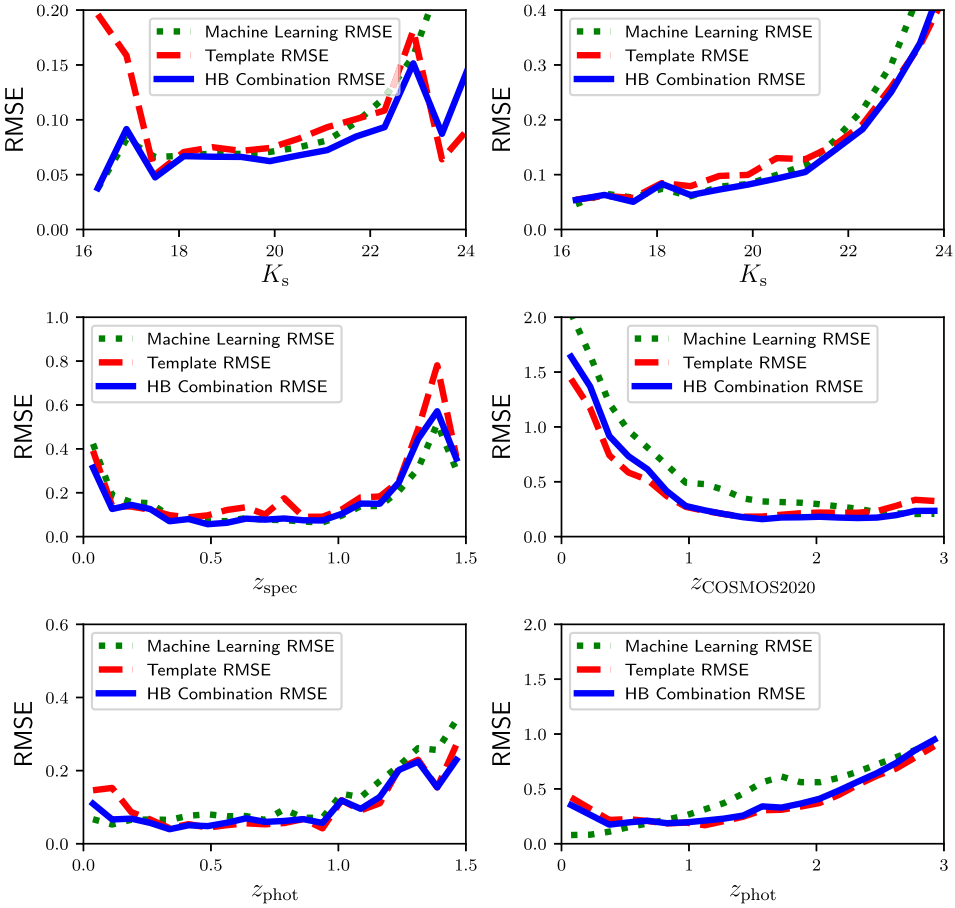

Figure 9. The RMSE scatter as a function of 'true' redshift for our three photo-z predictions. The first row shows as a function of K_s-band magnitude, the second row shows as a function of spectroscopic or COSMOS2020 redshift, and the third row shows as a function of photometric redshift. In the left column the 'true' redshifts are the spectroscopic redshift, in the right column the 'true' redshifts are the COSMOS2020 redshifts (the columns are plotted with different redshift ranges, corresponding to the redshift ranges where there were sufficiently large numbers of galaxies). Note prediction performance is likely to be poorer for the sample as a whole as the spectroscopic sample is not representative.

distribution to the training spectroscopic data, so represent a more realistic test of the photo-z quality. To extract the COSMOS2020 photo-z, we cross-match to our COSMOS data (1 arcsec max error). COSMOS2020 photo-z's were found for 664 322 of our 995 049 COSMOS sources (∼65 per cent).

Fig. 5 shows the point estimates from the three photo-z predictions (ML, template, HB), and the spectroscopic redshifts. Fig. 6 shows the point estimates from the three photo-z predictions (ML, template, HB), and the COSMOS2020 redshifts (and the spectroscopic redshifts to the COSMOS2020 redshifts). In Fig. 5, in the comparison of the ML and template fitting predictions, it can be seen the predictions agree for many sources (the data on the diagonal over $0 < z < 1$), but that there are many objects for which the predictions disagree (predominantly the fainter sources). For the photo-z to spectroscopic redshift comparisons, it can be seen that the photo-z predictions are generally accurate for the spectroscopic sample for all three methods. The HB predictions look qualitatively similar to the template fitting predictions, but with some outliers corrected (e.g. in Fig. 5 the population at $z_{spec} \sim 0.5$ and $z_{phot} \sim 0.2$ for the template fitting is corrected for the HB predictions). Similarly in Fig. 6, it can be seen that our photo-z are in general agreement with the COSMOS2020 photo-z out to $z \sim 4$, although still with a moderate number of outliers (some of which may be due to inaccuracies in our redshifts, and others of which may be due to inaccuracies in COSMOS2020).

Fig. 7 shows the normalized stacked PDF distributions, indicating the implied redshift distribution of our sample. We note that estimates of the population redshift distribution can also be derived with other HB models Leistedt et al. (2016), Malz & Hogg (2020), and Malz (2021). All three distributions are relatively similar and all peak at $z \sim 1$, although there are some differences, in particular the sharp $z = 0$ peak in the template-based distribution is not present in the HB combination distribution, and the high redshift tails have different thicknesses.

4.1 Metrics

Fig. 8 shows the outlier fraction FR15 [the fraction of sources with $|\Delta z| > 0.15(1 + z_{spec})$] as compared to both spectroscopic redshift and COSMOS2020 redshift (for which $z_{COSMOS2020}$ is used instead of z_{spec}). The ML achieves a very low FR15 over $0.2 < z < 1.2$ when compared with the spectroscopic sample, but then degrades at higher and lower redshifts where there is less training data. The template FR15 is relatively flat up to $z \sim 1.2$ and then also rises. The HB photo-z has similar but slightly better FR15 values to the template fitting predictions.

Fig. 9 shows the root mean square error (RMSE; $\sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(\frac{z_{spec}-z_{phot}}{1+z_{spec}})^2}$) scatter again as compared to both spectroscopic

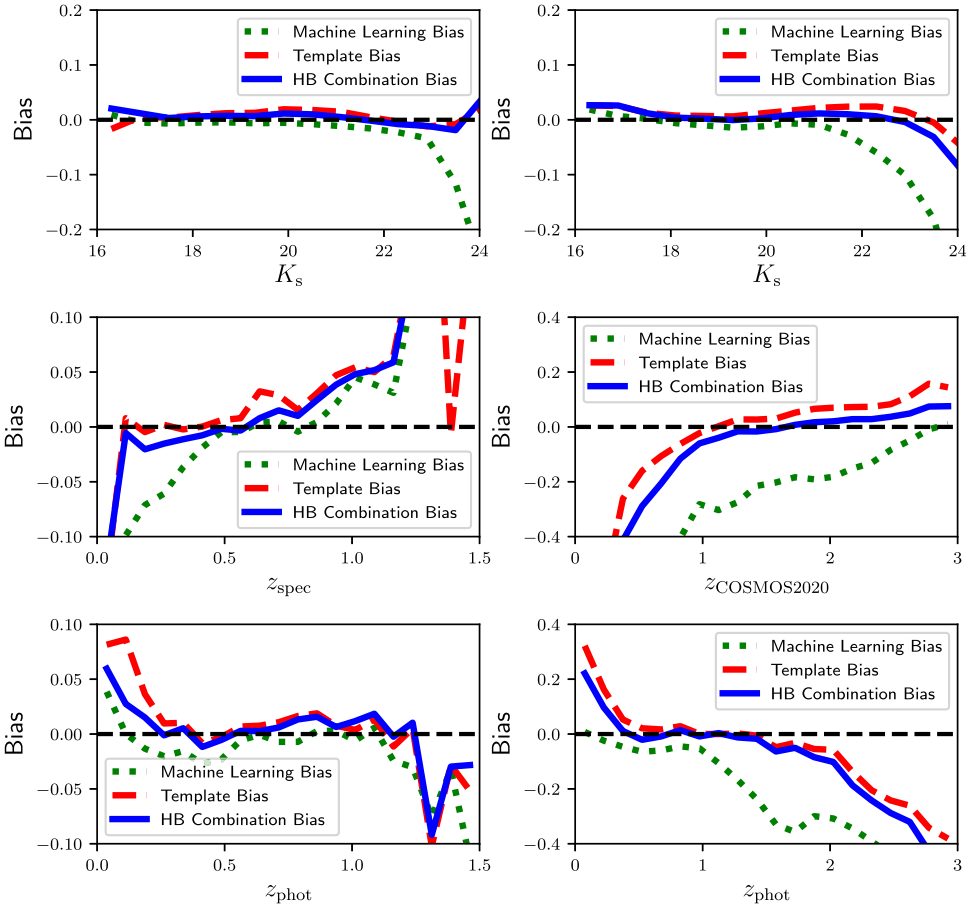

Figure 10. The bias as a function of 'true' redshift for our three photo-z predictions. The first row shows as a function of K_s-band magnitude, the second row shows as a function of spectroscopic or COSMOS2020 redshift, and the third row shows as a function of photometric redshift. In the left column the 'true' redshifts are the spectroscopic redshift, in the right column the 'true' redshifts are the COSMOS2020 redshifts (the columns are plotted with different redshift ranges, corresponding to the redshift ranges where there were sufficiently large numbers of galaxies). Note prediction performance is likely to be poorer for the sample as a whole as the spectroscopic sample is not representative.

Table 2. Summary statistics of the three estimators (see also Figs 8, 9, and 10). The best result for each metric is in bold. Uncertainties are calculated with a bootstrapping method.

Redshift estimate	Metric		
	FR15	*RMSE*	*Bias*
Compared against spectroscopic redshift			
Template fitting	3.6 ± 0.1	0.093 ± 0.005	0.0110 ± 0.0006
Machine learning	4.8 ± 0.2	0.089 ± 0.002	−0.0117 ± 0.0007
Hierarchical combination	**2.8 ± 0.1**	**0.077 ± 0.003**	**0.0044 ± 0.0005**
Compared against COSMOS2020 redshift			
Template fitting	24.7 ± 0.05	**0.527 ± 0.002**	**-0.0773 ± 0.0006**
Machine learning	56.2 ± 0.07	0.788 ± 0.001	−0.388 ± 0.001
Hierarchical combination	**24.6 ± 0.05**	0.602 ± 0.001	−0.151 ± 0.001

redshift and COSMOS2020 redshift (for which $z_{COSMOS2020}$ is used instead of z_{spec}). All methods have lowest RMSE over $0.2 < z < 1.2$ with the HB model giving the lowest scatter for most redshifts. Both the FR15 and RMSE are qualitatively similar to fig. 7 of Duncan et al. (2018b).

Fig. 10 shows the Bias ($\frac{z_{spec}-z_{phot}}{1+z_{spec}}$ when comparing to the spectroscopy, and $\frac{z_{COSMOS2020}-z_{phot}}{1+z_{COSMOS2020}}$ when comparing to the COSMOS2020 redshifts). In most bins, the Hierarchical combination has a value intermediate to the other two estimates.

In terms of K_s-band magnitude dependence, it can be seen that for all three metrics, and all three estimates, performance is broadly better for brighter sources, with relatively flat quality out to $K_s \approx 23$ for the ML predictions, and $K_s \approx 24$ for the other two predictions – before rapid deterioration. The HB combination is the highest performing of the three estimates for most magnitudes, both when compared to the spectroscopic sample, and to the COSMOS2020 sample.

Figs 8, 9, and 10 collectively show the key metrics (outlier fraction, root mean squared error, and bias, respectively) for the performance of the three sets of photo-z predictions, when compared

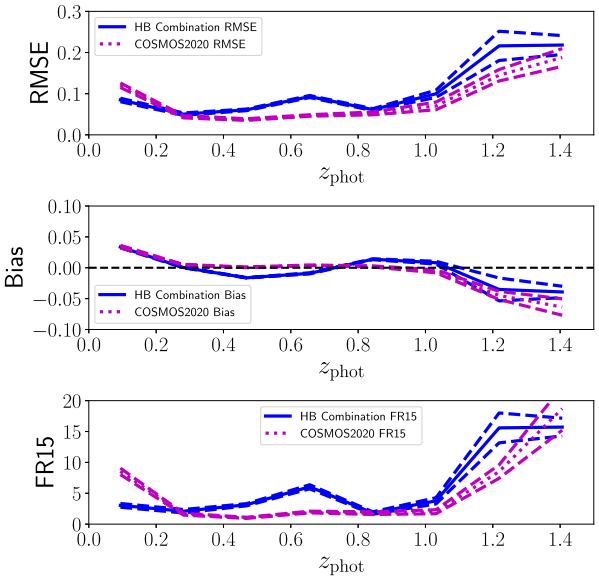

Figure 11. A comparison of the performance of COSMOS2020 redshifts and our HB redshifts (when only the XMM-LSS spectroscopic redshifts are used in the training process), tested using the spectroscopic redshifts in the COSMOS field as the true redshifts. The top plot shows the RMSE scatter, the central plot shows the bias, and the bottom plot shows the FR15. The dashed lines indicate the 1σ uncertainty on the measurements from a bootstrapping resampling analysis.

to the spectroscopic and COSMOS2020 samples. Note that these data sets are not representative of the whole sample, so in general performance is likely to be poorer over the entire source population.[8] Furthermore, the machine learning prediction is trained on the selfsame spectroscopic data, and thus should be expected to perform particularly well. In general, the performance on the spectroscopic sample is very high on all three metrics. For the most part, the HB combination predictions outperform the individual machine learning and template based results (although not quite for every single bin). Performance, in terms of consistency with COSMOS2020, for the entire sample was poorer than for the spectroscopic sample across the three metrics, although still high considering the depth of the sample. The HB combination prediction was still broadly the best performing for most bins, but not nearly as consistently (although this is difficult to definitively draw conclusions about, as COSMOS2020 was also template-based, and thus might be expected to be methodologically correlated with our template fitting results). The performance of the three estimates is summarized in Table 2, where it can be seen that the HB combination performs best for all the comparisons with the spectroscopic data, and the template fitting best for two of the three metrics for the COSMOS2020 comparison, and HB combination best for the third (although only with marginal significance).

If we wish to compare our final redshifts to the COSMOS2020 redshifts, we could make the comparison using the spectroscopic sample. However, this would be slightly preferential to our redshifts as the spectroscopic redshifts in the COSMOS field were actually used in our training process. Thus, we also calculate 'XMM-LSS Trained' redshifts, where we only use the spectroscopic redshifts from the XMM-LSS field when training the ML model. We can then compare these results to the COSMOS2020 results.[9] Fig. 11 shows the RMSE, FR15 and bias for the COSMOS2020 redshifts and our 'XMM-LSS Trained' Hierarchical Combination for the COSMOS spectroscopic sample. These results are summarized in Table 3. In particular, note that (a) the Template Fitting metrics are very similar to the corresponding values in Table 2 (albeit not exactly the same because the test data in Table 3 is just the COSMOS galaxies with spectroscopic redshifts, whereas Table 2 is for spectroscopic redshifts from both COSMOS and XMM-LSS), and (b) the 'XMM-LSS Trained' Machine Learning metrics are poorer than our main Machine Learning estimate metrics, mainly because they have access to less training data. It can be seen in Fig. 11 that the two redshift predictions are relatively comparable – the COSMOS2020 redshift predictions have access to more photometric bands, but our redshift predictions have access to spectroscopic redshifts via the ML predictions. Our HB combinations are a little better at low-z ($z < 0.3$), and the COSMOS2020 predictions a little better at intermediate-z ($0.4 < z < 0.8$). Table 3 shows our calculations actually had very slightly better RMSE (due to higher performance at lower redshift where most of the galaxies are), and only slightly poorer bias and FR15. Thus, this work represents redshift estimations of comparable quality to COSMOS2020 (for bright and intermediate luminosity sources where evaluation is possible), now extended and homogeneous across both the COSMOS and XMM-LSS fields, calculated to fainter luminosities, and using fewer filters.[10]

Fig. 12 shows a Probability Integral Transform (PIT) plot for the PDFs of the spectroscopic sample (used extensively in the literature to assess photo-z quality e.g. Bordoloi, Lilly & Amara 2010, see also Q-Q plots). PIT plots characterize the quality of the PDFs of the predictions, as opposed to the quality of the point estimates. To form a PIT plot, first for each prediction the probability mass of the PDF less than the true value ($z_{\rm spec}$) is calculated. The PIT plot is a (often normalized) histogram of these values e.g. what is the distribution of probability mass in the PDFs less than the true value. A uniform distribution over (0,1) would correspond to a perfectly calibrated PDF e.g. 10 per cent of PDFs have 10 per cent of their probability mass less than the true value. In Fig. 12, the HB model predictions are closest to the horizontal, indicating they are the most realistic PDFs. However, deviations from the horizontal indicate that the PDFs are not calibrated quite perfectly.[11] Note that PIT plots only quantify the 'realism' of PDFs, *not* whether or not the predictions are useful or have any information content. This is most clearly illustrated by the trainZ estimator discussed in Schmidt et al. (2020). This 'algorithm' simply assigned every galaxy a PDF of the redshift distribution of the whole population. This assignment achieves a perfect PIT, but is terrible on almost all other metrics (e.g. point estimates like RMSE) as it contains no information content. As discussed earlier, our prior distribution used in the HB combination mimics the population redshift distribution, so this contribution to the final PDFs will have high quality PIT. In particular for sources where both the template and the ML PDFs are deemed unreliable, the HB combination will revert to this prior, and thus score highly on the PIT, but in general be a poor predictor. This is not necessarily either a positive or a negative property of the predictions, but it is

[8] Also the COSMOS2020 redshifts are not perfectly accurate themselves.

[9] This of course slightly reduces the accuracy of our redshifts as they now have a smaller training set than our 'main' redshift calculations.

[10] Although as discussed, the quality for the parts of colour–magnitude space without spectroscopy is harder to validate.

[11] See for comparison the calibration process presented in Gomes et al. (2018).

Table 3. Summary statistics of the COSMOS2020 estimates, the Template Fitting, the XMM-LSS-trained Machine Learning, and the XMM-LSS-trained Hierarchical Combination estimates (see also Fig. 11). The estimates are compared against the spectroscopic redshifts in the COSMOS field only. Uncertainties are calculated with a bootstrapping method.

Redshift estimate	Metric		
	FR15	*RMSE*	*Bias*
Compared against spectroscopic redshift			
COSMOS2020	**2.9 ± 0.1**	0.109 ± 0.006	**-0.0019 ± 0.0007**
Template fitting	3.6 ± 0.1	0.091 ± 0.005	0.0101 ± 0.0007
XMM-LSS-trained machine learning	8.5 ± 0.2	0.113 ± 0.002	− 0.0295 ± 0.0006
XMM-LSS-trained hierarchical combination	3.9 ± 0.1	**0.090 ± 0.003**	− 0.0029 ± 0.0006

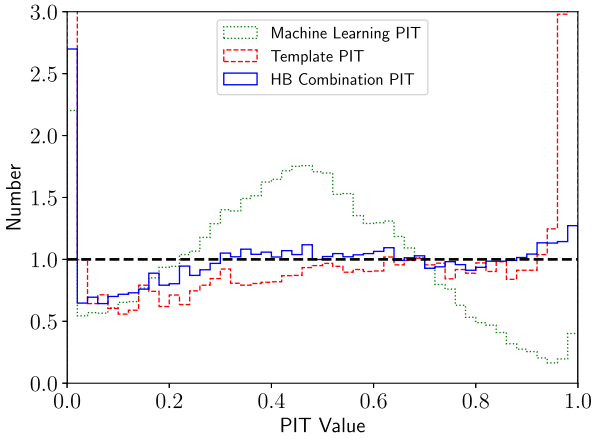

Figure 12. A PIT plot for the GPz machine learning, template fitting, and HB combination photo-z PDFs.

important to emphasize here that PIT score alone is not an indication of quality of prediction. In any case, the majority of sources have at least one reliable photo-z estimate (see Fig. 2); only 19 sources had both $1/\beta_{\rm template}$ and $1/\beta_{\rm ML}$ less than 0.1. The prior still impacts the final PDF even if both estimates are reliable (e.g. if the estimates are 'reliable', but with very large uncertainties), but the generally medium to high $1/\beta_i$ values, combined with the high performance on the point estimates, collectively implies that the PDFs are generally not being dominated by the prior.

4.2 A possible $z \sim 6.8$ galaxy?

In this section, we show how the combined HB method could be used to check the validity of candidate high-z galaxies, where although there are very few high-z training spec-z, the higher accuracy of GPz at lower redshift may give more accurate weighting to low-z solutions.

Endsley et al. (2022) report the detection of a possible $z \sim 6.8$ massive star-forming galaxy. The source, COS-87259, was identified and assigned a redshift using the a Lyman-break narrow-band dropout technique – and is among the COSMOS sources we have analysed. The authors have many more bands (including narrow bands) than used in this work, but do note the possibility of a lower redshift interpretation. Our analysis assigned $z_{\rm ML} = 1.04$, $z_{\rm template} = 7.01$, and $z_{\rm HB} = 0.78$ (PDFs shown in Fig. 13). Our template fitting photo-z thus was consistent with the dropout method, and the ML was consistent with a lower-z solution. In agreement with Endsley et al. (2022), we found low χ^2 values ($\chi^2_{\rm galaxy} = 0.60$ and $\chi^2_{\rm AGN} = 3.9$) indicating good fits, and that the template fitting solution should be reliable.[12] However, for the ML $v/\sigma^2 = 0.013$, indicating that it was not excessively extrapolating, so that the ML prediction ought also be reliable. Furthermore, ML predictions do not rely on the set of templates used, which if incomplete may miss low-z solutions. Both methods have similar β_i weights for the combination of the PDFs. However, the redshift prior distribution favours the low-z solution to a large degree, meaning the $z_{\rm HB}$ is dominated by the $z_{\rm ML}$ value.

Which redshift value ought be believed? First, it should be noted that there are no training spectroscopic redshifts at $z \sim 7$, so the ML will never predict a galaxy to be at $z \sim 7$. The template fitting in this case gave a redshift estimation PDF with a $z \sim 7$ peak, and a much smaller (by a factor of ∼100 in probability) broad $z \sim 1$ peak. Thus, the question is whether the combination with the ML prediction and the prior via the HB model are correctly modifying the relative sizes of the peaks in the template fitting pdf. The final PDF is thus dependent on choice of weighting system (choice of $f_{\rm bad}$, etc.) and choice of prior. As Endsley et al. (2022) note, true confirmation of the redshift of the source will require spectroscopic follow-up. However, the authors do present compelling evidence beyond simply the photometry (e.g. colocation on the sky with an overdensity of other sources at $z \sim 7$) that the source really is at high redshift. Assuming the source really is at high redshift, is our HB model wrong to favour the lower redshift solution? We believe not necessarily, because COS-87259 was *selected* based on the narrow-band photometry. It is perfectly consistent for the majority of sources with this broad-band colour–magnitude to be at lower redshift (causing the ML prediction to take the lower-z value), but a small fraction to be higher redshift. Which value to use depends on science goal and if false-positive or false-negative high-z predictions are more costly. For context, taken at face value our $z_{\rm template}$ values would indicate 6 per cent of our sources are at $z > 6$, whereas our $z_{\rm HB}$ values would indicate closer to 1 per cent. Finally, we would also note Endsley et al. (2022) identify radio continuum emission associated with COS-87259, which may further alter the appropriate redshift distribution prior.

4.3 Choice of point estimate

As we mentioned in Section 3.6, choice of point estimate depends on science goal. Fig. 14 shows FR15 as a function of percentage of data for median and mode point estimates from the HB estimates. It can be seen that for majority of sources the predictions are very similar, but that the mode slightly outperforms the median, mainly for the sources with the highest uncertainty. Everywhere else in this work the mode is used when a point estimate is used.

[12]Note, however, that we did not use any of the narrow bands used in Endsley et al. (2022), so might expect to have a broader PDF than their results.

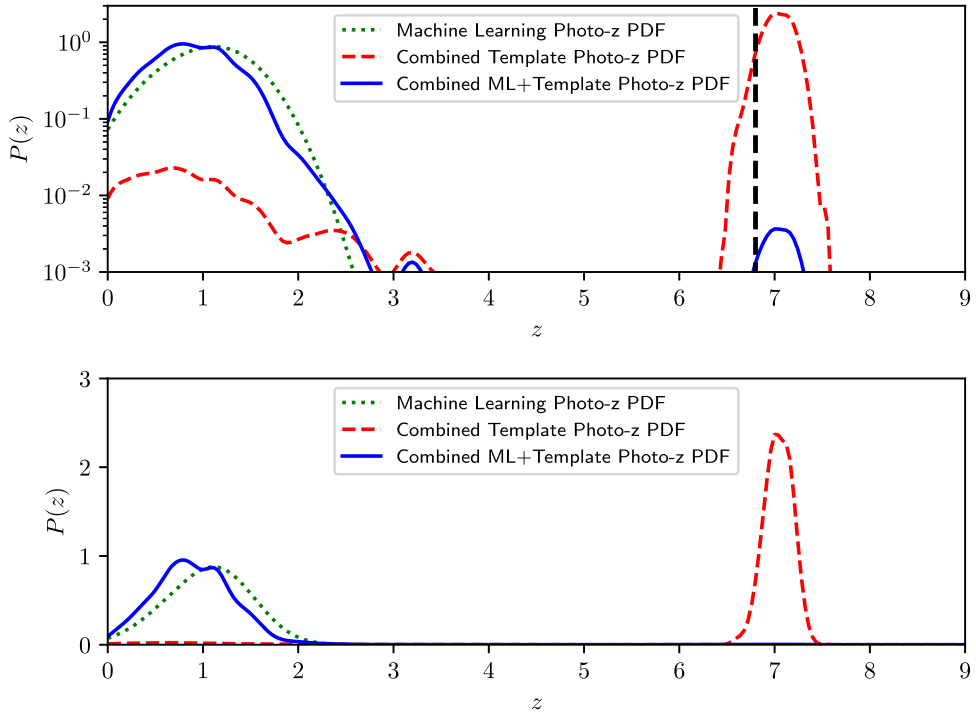

Figure 13. The ML, template, and HB PDFs for COS-87259 (top panel shows with a log-scale, the lower panel with a linear-scale). The vertical line shows the Endsley et al. (2022) redshift.

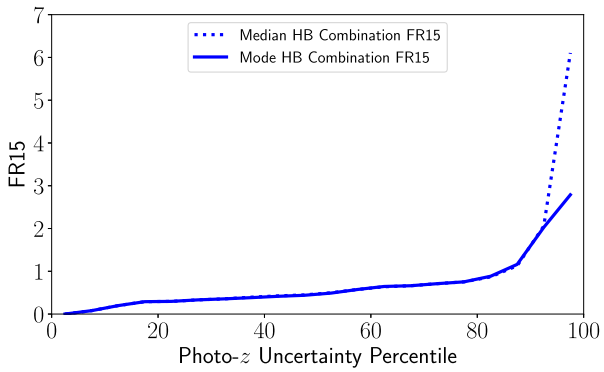

Figure 14. The outlier fraction as a function of 'true' redshift for the median and mode of our HB photo-z PDF predictions, compared to the spectroscopic redshifts. The data are plotted as a function of 'Photo-z Uncertainty Percentile' (as a function of percentile, ranked by uncertainty on estimate).

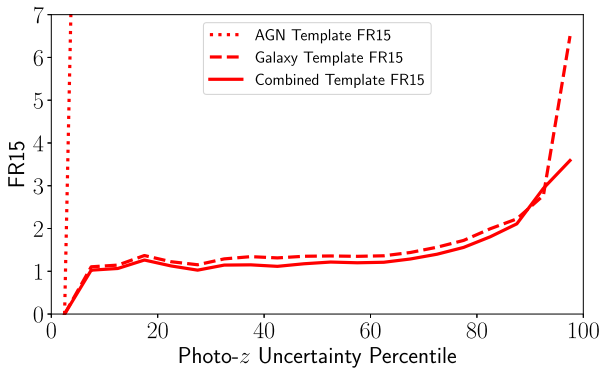

Figure 15. The outlier fraction as a function of 'true' redshift for our three template fitting based photo-z predictions, compared to the spectroscopic redshifts. The data are plotted as a function of 'Photo-z Uncertainty Percentile' (as a function of percentile, ranked by uncertainty on estimate).

4.4 Evaluation of composite template fitting estimates

We also test how great an improvement is achieved by including AGN template PDFs with the galaxy template PDFs, rather than just using galaxy templates (Section 3.4). Fig. 15 shows FR15 as a function of percentage of data for the point estimates for the galaxy-only, AGN-only, and HB combination of the two (i.e. before the machine learning based predictions are added). It can be seen that the AGN-only prediction is much poorer than the other two estimators.[13] The combined estimator is a little bit better than the galaxy-only prediction for most of the data, and much better for the final ∼10 per cent of the data where the uncertainties are greatest (likely the AGN in the sample).

[13]Assuming one is interested in the quality of prediction across the whole sample. If ones science goal specifically concentrated on AGN then of course these predictions might be more helpful.

5 CONCLUSIONS

We have presented the calculation of photometric redshift estimates for sources with deep optical and near-infrared data over the COSMOS and XMM-LSS fields. We calculate template fitting based photometric redshifts for both galaxy and AGN templates using LePhare (building on Adams et al. 2020), as well as machine learning based photometric redshifts using GPz (building on Hatfield et al. 2020), and then used an HB model to combine them (using the method of Duncan et al. 2018b). By combining template fitting and machine learning, we achieve predictions that take the best aspects

of both approaches to photo-z calculation. These redshifts were then tested by comparison to the Weaver et al. (2022) COSMOS2020 photometric redshifts, which are still photometric redshifts, but had access to a higher number of bands. Our redshifts are of comparable quality to the COSMOS2020 redshifts – the information from the spectroscopic training set can make up some of the loss of information from having fewer bands.

The redshifts calculated in this work thus represent the most accurate set of redshifts for a catalogue this large of deep multi-wavelength photometry over multisquare degree surveys. Calculating photometric redshifts is a key challenge in extragalactic astronomy – this work and the resulting data set represents an important large set of reliable high-quality photo-z for future science use over these key extragalactic fields.

ACKNOWLEDGEMENTS

PH acknowledges generous support from the Hintze Family Charitable Foundation through the Oxford Hintze 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. RB acknowledges support from an STFC Ernest Rutherford Fellowship [grant number ST/T003596/1]. 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 CEA (Commissariat à l'énergie atomique)/IRFU (Institut de recherche sur les lois fondamentales de l'Univers), 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.

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

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 photo-z in this study will be made available online at time of publication. Other derived data generated in this research will be shared on reasonable request to the corresponding author.

REFERENCES

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., 2017, 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
Ata M., Kitaura F. S., Lee K. G., Lemaux B. C., Kashino D., Cucciati O., Hernandez-Sanchez M., Le Fevre O., 2021, MNRAS, 500, 3194
Baldry I. K. et al., 2010, MNRAS, 404, 86
Benítez 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., Rot A. H., eds, ASP Conf. Ser. Vol. 442, Automated Morphometry with SExtractor and PSFEx. Astron. Soc. Pac., San Francisco, p. 435
Bordoloi R., Lilly S. J., Amara A., 2010, MNRAS, 406, 881
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
Brodwin M. et al., 2006, ApJ, 651, 791
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483
Carrasco Kind M., Brunner R. J., 2014, MNRAS, 442, 3380
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
Clerc N. et al., 2014, MNRAS, 444, 2723
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
Dahlen T. et al., 2013, ApJ, 775, 93
Darvish B., Mobasher B., Martin D. C., Sobral D., Scoville N., Stroe A., Hemmati S., Kartaltepe J., 2017, ApJ, 837, 16
Duncan K. J., 2022, MNRAS, 512, 3662
Duncan K. J. et al., 2018a, MNRAS, 473, 2655
Duncan K. J., Jarvis M. J., Brown M. J. I., Röttgering H. J. A., 2018b, MNRAS, 477, 5177
Duncan K. J. et al., 2019, A&A, 622, A3
Endsley R. et al., 2022, MNRAS, 512, 4248
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
Frayer D. T. et al., 2009, AJ, 138, 1261
Gomes Z., Jarvis M. J., Almosallam I. A., Roberts S. J., 2018, MNRAS, 475, 331
Hale C., Jarvis M., Delvecchio I., Hatfield P., Novak M., Smolčić V., Zamorani G., 2018, MNRAS, 474, 4133
Hasinger G. et al., 2018, ApJ, 858, 77
Hatfield P. W., Almosallam I. A., Jarvis M. J., Adams N., Bowler R. A. A., Gomes Z., Roberts S. J., Schreiber C., 2020, MNRAS, 498, 5498
Hsieh B. C., Yee H. K., 2014, ApJ, 792, 102

Ilbert O. et al., 2006, A&A, 457, 841
Ilbert O. et al., 2009, ApJ, 690, 1236
Jarvis M. J. et al., 2013, MNRAS, 428, 1281
Kawanomoto S. et al., 2018, PASJ, 70, 66
Kodra D., 2019, in ProQuest Dissertations And Theses, Dissertation Abstracts International, Volume: 80-08(E). University of Pittsburgh, Pittsburgh, p. 164
Laigle C. et al., 2016, ApJS, 224, 24
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
Madau P., 1995, ApJ, 441, 18
Malz A. I., 2021, Phys. Rev. D, 103, 083502
Malz A. I., Hogg D. W., 2020, preprint (arXiv:2007.12178)
McCracken H. J. et al., 2012, A&A, 544, A156
McLure R. J. et al., 2018, MNRAS, 479, 25
Momcheva I. G. et al., 2016, ApJS, 225, 27
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Pacaud F. et al., 2007, MNRAS, 382, 1289
Pentericci L. et al., 2018, A&A, 616, A174
Polletta M. et al., 2007, ApJ, 663, 81
Rahman M., Ménard B., Scranton R., Schmidt S. J., Morrison C. B., 2015, MNRAS, 447, 3500
Rasmussen C. E., Williams C. K. I., 2006, Gaussian Processes for Machine Learning. MIT Press, Cambridge, Massachusetts, p. 248
Sadeh I., Abdalla F. B., Lahav O., 2016, PASP, 128, 104502
Salvato M. et al., 2009, ApJ, 690, 1250
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, 499, 1587
Silva L., Granato G. L., Bressan A., Danese L., 1998, ApJ, 509, 103
Silverman J. D. et al., 2015, ApJS, 220, 12
Simm T. et al., 2015, A&A, 584, 106
Skelton R. E. et al., 2014, ApJS, 214, 24
Sonnenfeld A., 2022, A&A, 659, A133
Stylianou N., Malz A. I., Hatfield P., Crenshaw J. F., Gschwend J., 2022, PASP, 134, 044501
Weaver J. R. et al., 2022, ApJS, 258, 11
Zuntz J. et al., 2021, Open J. Astrophys., 4

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