Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **503,** 4118–4135 (2021) doi:10.1093/mnras/stab711
Advance Access publication 2021 March 10

The PAU Survey: narrow-band photometric redshifts using Gaussian processes

John Y. H. Soo [1,2]★ Benjamin Joachimi,[2] Martin Eriksen,[3] Małgorzata Siudek [3,4] Alex Alarcon [5] Laura Cabayol,[3] Jorge Carretero,[3,6] Ricard Casas,[7,8] Francisco J. Castander,[7,8] Enrique Fernández,[3] Juan García-Bellido [9] Enrique Gaztanaga [7,8] Hendrik Hildebrandt,[10] Henk Hoekstra [11] Ramon Miquel,[3,12] Cristobal Padilla,[3] Eusebio Sánchez,[13] Santiago Serrano[7,8] and Pau Tallada-Crespí[6,13]

[1]*School of Physics, Universiti Sains Malaysia (USM), 11800 USM, Pulau Pinang, Malaysia*
[2]*Department of Physics and Astronomy, University College London (UCL), Gower Street, London WC1E 6BT, UK*
[3]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, E-08193 Bellaterra (Barcelona), Spain*
[4]*National Centre for Nuclear Research, 7 Pasteura Str, PL-02-093 Warsaw, Poland*
[5]*High Energy Physics Division, Argonne National Laboratory, Lemont, IL 60439, USA*
[6]*Port d'Informació Científica (PIC), Universitat Autònoma de Barcelona, Carrer Albareda S/N, E-08193 Bellaterra (Barcelona), Spain*
[7]*Institute of Space Sciences (ICE/CSIC), Universitat Autònoma de Barcelona, Carrer de Can Magrans S/N, E-08193 Cerdanyola del Vallès (Barcelona), Spain*
[8]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[9]*Instituto de Física Teórica (IFT-UAM/CSIC), Universidad Autónoma de Madrid, E-28049 Madrid, Spain*
[10]*German Centre for Cosmological Lensing, Astronomisches Institut, Ruhr-Universität Bochum (AIRUB), Universitätsstr 150, D-44801 Bochum, Germany*
[11]*Leiden Observatory, Leiden University, Niels Bohrweg 2, NL-2333 CA Leiden, the Netherlands*
[12]*Institució Catalana de Recerca i Estudis Avançats (ICREA), E-08010 Barcelona, Spain*
[13]*Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas (CIEMAT), Avenida Complutense 40, E-28040 Madrid, Spain*

Accepted 2021 March 6. Received 2021 March 6; in original form 2021 January 11

ABSTRACT

We study the performance of the hybrid template machine learning photometric redshift (photo-z) algorithm DELIGHT, which uses Gaussian processes, on a subset of the early data release of the Physics of the Accelerating Universe Survey (PAUS). We calibrate the fluxes of the 40 PAUS narrow bands with six broad-band fluxes (*uBVriz*) in the Cosmic Evolution Survey (COSMOS) field using three different methods, including a new method that utilizes the correlation between the apparent size and overall flux of the galaxy. We use a rich set of empirically derived galaxy spectral templates as guides to train the Gaussian process, and we show that our results are competitive with other standard photometric redshift algorithms. DELIGHT achieves a photo-z 68th percentile error of $\sigma_{68} = 0.0081(1 + z)$ without any quality cut for galaxies with $i_{\rm auto} < 22.5$ as compared to $0.0089(1 + z)$ and $0.0202(1 + z)$ for the BPZ and ANNZ2 codes, respectively. DELIGHT is also shown to produce more accurate probability distribution functions for individual redshift estimates than BPZ and ANNZ2. Common photo-z outliers of DELIGHT and BCNZ2 (previously applied to PAUS) are found to be primarily caused by outliers in the narrow-band fluxes, with a small number of cases potentially indicating spectroscopic redshift failures in the reference sample. In the process, we introduce performance metrics derived from the results of BCNZ2 and DELIGHT, allowing us to achieve a photo-z quality of $\sigma_{68} < 0.0035(1 + z)$ at a magnitude of $i_{\rm auto} < 22.5$ while keeping 50 per cent objects of the galaxy sample.

Key words: methods: numerical – methods: statistical – galaxies: distances and redshifts.

1 INTRODUCTION

Photometric redshift (photo-z) estimation continues to be an active research area as it plays a major role in solving the big questions in cosmology. Redshifts provide radial information (distance) to the traditional two-dimensional sky maps of galaxies. They are traditionally determined through spectroscopic methods (spectroscopic redshifts, or spec-zs). Yet since the process requires long telescope time for high completeness, photo-zs are instrumental for the analysis of large surveys containing of order $10^{8–9}$ galaxies. Photo-z methodology has been evolving and improving a lot over the past couple of decades (e.g. Brescia et al. 2018; Salvato, Ilbert & Hoyle 2019), such that it had been sufficiently useful for most recent cosmological researches.

Photo-z, as its name suggests, is often determined through the use of a handful of broad-band photometric filters obtained from large sky surveys. Photo-z estimation methods are generally categorized into two different types: the template-based method, which relies on accurate models of spectral energy distribution (SED) templates of different types of galaxies; and the data-driven empirical method, which relies on training sets of galaxies and machine learning algorithms. Each method however has its own limitations: template-based methods may produce photo-zs with large scatter and catastrophic rates without representative templates; while machine

★ E-mail: johnsooyh@usm.my

learning methods may perform poorly outside the regions of the parameters covered by the training sample (D'Isanto et al. 2018). As a result, hybrid methods have been implemented to utilize the best of both worlds (Cavuoti et al. 2017; Duncan et al. 2018, 2019).

Many current and upcoming surveys such as the Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005), Legacy Survey of Space and Time (LSST; Ivezić et al. 2019), *Euclid* (Laureijs et al. 2011), Kilo-Degree Survey (KiDS; De Jong et al. 2013), *Wide-Field Infrared Survey Telescope* (*WFIRST*; Spergel et al. 2013), and Hyper Suprime-Cam (HSC; Aihara et al. 2018) have set stringent photo-z requirements to ensure that they meet their science goals, forcing the quality of photo-z methodology to constantly improve. For example, LSST's photo-z requirement is to reach a root-mean-square error of $\sigma_{\rm rms} < 0.02(1+z)$, while the *Euclid* requirement is $\sigma_{\rm rms} < 0.05(1+z)$. High-quality photo-$z$s are required for a reliable estimation of e.g. weak lensing (Benjamin et al. 2013), angular clustering (Crocce et al. 2016), intrinsic alignment (Johnston et al. 2021), structure formation, galaxy classification, and galaxy properties (Jouvel et al. 2017; Laigle et al. 2018; Siudek et al. 2018).

The aforementioned surveys are predominantly broad-band surveys that use between four and nine broad-band filters ranging from infrared (IR) to ultraviolet (UV). This work, however, explores the estimation of photo-zs in narrow-band surveys, focusing on the Physics of the Accelerating Universe Survey (PAUS; Padilla et al. 2019), which observes the sky using 40 narrow bands (see Section 2.1). Producing high-quality photo-zs for such a survey requires careful optimization between narrow and broad bands, since machine-learning-based methods have to be optimized for a larger number of inputs (Eriksen et al. 2020), while template-based methods require more attention towards the narrow emission line features.

Martí et al. (2014) used simulations to predict that by using PAUS narrow-band photometry, the photo-z quality could reach an unprecedentedly low 68th percentile error of $\sigma_{68} = 0.0035(1+z)$ at a quality cut of 50 per cent at $i < 22.5$. This has been verified by Eriksen et al. (2019), where they combined the 40 PAUS narrow bands (early data release) with broad bands *uBVriz* from the Cosmic Evolution Survey (COSMOS; Laigle et al. 2016), and using their template-based photo-z code BCNZ2, they showed that this result is achievable when a 50 per cent photometric quality cut was imposed on the final testing set. In a more recent work, Eriksen et al. (2020) used DEEPZ, a deep learning algorithm on the same data set and showed that it outperformed BCNZ2 by reaching 50 per cent lower in σ_{68}. Furthermore, Alarcon et al. (2020) showed that an ever greater precision can be achieved when using additional photometric bands available in the COSMOS field (a total of 66 bands).

We are motivated by the work of Eriksen et al. (2019), but instead of using purely template-based methods, we attempt to achieve this PAUS photo-z precision by utilizing Gaussian processes (GPs, see Section 3.1) to make empirical adjustments to templates, working on the same data set and conditions. We seek to produce an independent method that is competitive, as that will allow us to exploit synergies with BCNZ2 by Eriksen et al. (2019) as shown in this work, DEEPZ (Eriksen et al. 2020), and photo-zs by Alarcon et al. (2020) in the future. Therefore the contents of this paper reflect our findings, putting special emphasis on the performance and application of DELIGHT (Leistedt & Hogg 2017), a hybrid template machine learning photo-z code. When carefully calibrated and combined with COSMOS broad-band fluxes, DELIGHT should achieve equally good results as that of BCNZ2. The main aims of this paper are threefold:

(i) to optimize and test the performance of the hybrid template machine learning photo-z code DELIGHT on a narrow-band survey;

(ii) to develop an optimal method to calibrate the fluxes between the COSMOS broad bands and the PAUS narrow bands;

(iii) to provide an independent photo-z solution for PAUS, enabling the study of photometric and spectroscopic redshift outliers.

This paper is structured as follows. In Section 2, we first introduce PAUS and the sources of photometry and spectroscopic redshifts used in this work. Section 3 describes the algorithms (DELIGHT, ANNZ2, and BPZ) used in this work, together with their optimization settings and SED templates used. Section 4 describes the full details of how the photometry and spectroscopy from PAUS, COSMOS, and zCOSMOS are cross-matched, how the galaxy fluxes are selected, the three methods to calibrate the broad-band and narrow-band fluxes, and the performance metrics used in this work to compare the results between runs and codes. Section 5 shows the photo-z results obtained by DELIGHT, and a thorough analysis is conducted to compare its performance with ANNZ2, BPZ, and BCNZ2. Finally, in Section 6, we study the photo-z outliers of DELIGHT and BCNZ2, and derive new metrics with improved photo-z outlier identifications. Our work is concluded in Section 7.

2 PHOTOMETRY AND SPECTROSCOPY

In this work, photometric data were obtained from PAUS (Section 2.1) and COSMOS (Section 2.2), while spectroscopic redshifts were obtained from zCOSMOS (Section 2.3). In this section, these surveys will be introduced, together with the selection cuts used to obtain our training and testing sets.

2.1 PAUS

Physics of the Accelerating Universe Survey (PAUS) is a narrow-band photometric galaxy survey aimed at mapping the large-scale structure of the Universe up to $i \sim 23.0$. Using 40 narrow bands spaced by 100 Å in the range between 4500 and 8500 Å (filter responses visualized in Eriksen et al. 2019, and Fig. 4), PAUS aims to achieve redshifts with a precision of $\sigma_{\rm rms} < 0.0035(1+z)$ for galaxies with $i_{\rm auto} < 22.5$. PAUS uses the PAUCam instrument (Padilla et al. 2019) on the 4-m William Herschel Telescope (WHT) at Observatorio del Roque de los Muchachos (ORM) in La Palma. It has observed more than 50 deg^2 of sky since the beginning of 2016, and observations to full depth in all narrow bands for 100 deg^2 are planned.

The PAUS forced-aperture co-added photometry has its aperture defined by using the 50 per cent light radius (r_{50}), the point spread function (PSF), ellipticity, and Sérsic index of COSMOS morphology, such that the fluxes measure a fixed fraction of light. The reader is referred to Eriksen et al. (2019) for detailed information on how the PAUS fluxes are measured. In this work, we used the early data release from PAUS (objects are observed at least five times, using an elliptical aperture with 62.5 per cent light radius), and select objects with $i_{\rm auto} \leq 22.5$, entries with no missing measurement, and the COSMOS flag `TYPE=0` (extended objects).

2.2 COSMOS

The Cosmic Evolution Survey (COSMOS; Scoville et al. 2007) covers a sky area of 2 deg^2 ($149.^\circ47 \leq \alpha \leq 150.^\circ7$, $1.^\circ62 \leq \delta \leq 2.^\circ83$) and is known for its high sensitivity, depth, and an exceptionally low and uniform Galactic extinction ($E(B-V) \sim 0.02$).

In this work, we used photometry from the COSMOS2015 catalogue (Laigle et al. 2016); it is a highly complete mass-selected sample to very high redshifts, highly optimized for the study of galaxy evolution and environments in the early Universe. The COSMOS2015 catalogue provides 30-band photometry ranging from near-UV to near-IR wavelengths, all these have been observed through multiple facilities, two of which are the Canada–Hawaii–France Telescope (CFHT) and Subaru Telescope (Miyazaki et al. 2002). From this catalogue we only use the CFHT u^*-band (Boulade et al. 2003) and Subaru B, V, r, i^+, and z^{++} bands (Miyazaki et al. 2002), in conjunction with the narrow-band photometry of PAUS. For simplicity, these bands will be referred to collectively as the *uBVriz* bands; the superscripts are dropped for easier reading.

2.3 zCOSMOS

The zCOSMOS Survey (Lilly et al. 2007) targets galaxies in the COSMOS field using the Visible Multi-Object Spectrograph (VIMOS; Le Fèvre et al. 2003). zCOSMOS-Bright observed 20 689 galaxies in a sky area of 1.7 deg^2, these galaxies have magnitudes $15 < i_{\rm auto} < 22.5$ and redshifts in the range of $0.1 < z < 1.2$, its spectral range is in the red (rest-frame wavelength 5550–9650 Å) to follow strong spectral features around the 4000 Å break to as high redshifts as possible.

In this work, we use data from zCOSMOS-Bright Data Release 3 (DR3).[1] Galaxies with redshift confidence class 3 and 4 (spectroscopic verification rate of 99 per cent and 99.8 per cent, respectively) are selected and cross-matched with PAUS objects.

2.4 Our data set

Using the aforementioned selection cuts, we cross-matched within 1 acsec the 40-narrow-band photometry from PAUS, six-broad-band photometry (*uBVriz*) from COSMOS, and highly reliable redshifts from zCOSMOS to obtain a data sample of 8406 galaxies, which is divided randomly into half for training and testing, respectively. This sample uses a total of 46 bands, and flux calibration between the broad and narrow bands is required as they are obtained from different surveys with different flux measurements. The calibration between these fluxes will be discussed in Section 4.

The colour–magnitude diagram of this sample is shown in Fig. 1, in comparison with the COSMOS2015 sample (all objects with `TYPE=0` and detected in r and i). The slight incompleteness in i magnitude is due to the selection effects in brightness of the spectroscopic redshifts available.

The sample size may seem small, but is sufficient for the GP to work, since the GP essentially creates 4000 + flux-redshift 'templates' to produce photo-zs for objects in the testing set. However, we note that such a small training size has a major effect on the results of ANNZ2 as this training size is close to the lower limit threshold suggested by Bonfield et al. (2010). We also note that the sample we have chosen is very similar to that of Eriksen et al. (2019), the only difference being that they have a more relaxed cut in the number of bands (`N_BANDS`), being `35<N_BANDS<40` (workable for a template code like BCNZ2), while we used `N_BANDS=40`.[2] When comparing results between DELIGHT and BCNZ2, we will only compare photo-zs of the exact same objects. Note that we have used the same broad bands as used by Eriksen et al. (2019).

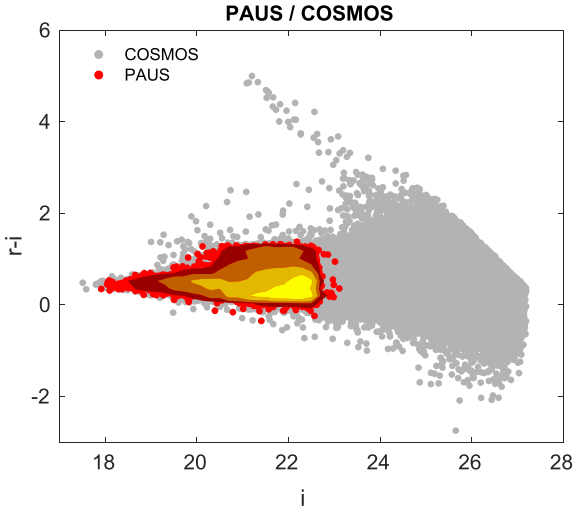

Figure 1. Colour–magnitude diagram for the PAUS data (red) used in this work in comparison with the COSMOS2015 sample (all objects with `TYPE=0` and detected in r and i). The contours represent the density of objects.

3 ALGORITHMS AND TEMPLATES

3.1 DELIGHT and Gaussian processes

DELIGHT[3] (Leistedt & Hogg 2017) is a hybrid template-based and machine learning photo-z algorithm, which was constructed to combine the advantages, and minimize the disadvantages, of both types of algorithms. DELIGHT constructs a large collection of latent SED templates (or physical flux-redshift models) from training data, with a template SED library as a guide to the learning of the model. This conceptually novel approach uses Gaussian processes (GPs) operating in flux-redshift space. DELIGHT was featured in the results of the LSST Photo-z Data Challenge 1 (Schmidt et al. 2020), where it was found to have a low photo-z bias but slightly broader probability density functions (PDFs).

A GP is a supervised learning method, which finds a distribution over the possible functions $f(x)$ that are consistent with the observed data x. Consider Fig. 2: suppose we have a set of observed variables $y = f(x)$, we can fit it using a GP, denoted as $f \sim \mathcal{GP}(\mu, k)$, which assumes that the probability of all $f(x)$ is jointly Gaussian and representable by a mean function $\mu(x)$ and a covariance matrix $\Sigma(x) = k(x_i, x_j)$. $k(x_i, x_j)$ is the kernel function, which relates one variable x_i to another x_j. An example case would be $\mu \equiv 0$ and a kernel function that takes the form of a squared exponential,

$$k(x_i, x_j) = \sigma_f^2 \exp\left[\frac{-(x_i - x_j)^2}{2l^2}\right], \quad (1)$$

where σ_f^2 is the maximum allowable covariance between data (set by the errors on the observation), and l is the tunable correlation length that determines the smoothness of the GP. In this simplistic case, the GP will try to find a marginalization of all possible functions, but μ and k can be modified if an underlying model of the data we want to

[1] http://www.eso.org/qi/catalog/show/65

[2] The relaxed cut resulted in Eriksen et al. (2019) having a larger sample size of 10 801 objects.

[3] https://github.com/ixkael/Delight

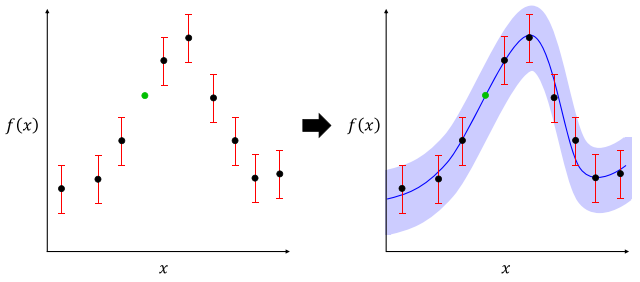

Figure 2. Illustration of a Gaussian process (GP). The left-hand panel shows data points (black dots), with a single datum to be predicted (green dot). The GP trains on the given data points to provide a best-fitting function (blue line) as shown on the right. It also provides a Gaussian confidence interval (blue shaded area) for the prediction.

fit is known. The covariance function is defined such that a smooth function is to be predicted.

Assuming that we have a set of training data $\{x_i, f(x_i)\}$ and would like to find the prediction $\{x_*, f_*(x_*)\}$, the GP models f and f_* as jointly Gaussian, $\mathcal{N}(\mu, \Sigma)$, and therefore

$$\begin{pmatrix} f(x) \\ f_*(x) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \mu_* \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma_* \\ \Sigma_*^{\mathrm{T}} & \Sigma_{**} \end{pmatrix} \right), \tag{2}$$

where $\Sigma = k(x_i, x_j)$ is the covariance between the training data, $\Sigma_* = k(x_*, x_i)$ the covariance between training and the predicted data (superscript T denotes the transpose of the matrix), while $\Sigma_{**} = k(x_*, x_*)$ is the variance of the predicted data.

It follows from the above that the posterior $p(f_*|x_*, x_i, f_i)$ is also Gaussian, therefore a predicted point $f_*(x_*)$ is plotted (green dot in Fig. 2) is modelled by a Gaussian function (smooth blue line) that runs across all points, with its 95 per cent confidence interval ($\pm 1.96\sigma_{f_*}$) represented by the navy shaded area.

In the context of DELIGHT, GPs are used to calculate the predicted fluxes $\hat{F}$ at a certain redshift z for a training object i with fluxes F_i and redshift z_i. This could be better understood by first defining the posterior photo-z distribution $p(z|\hat{F})$ of an object in the testing set. For machine learning methods, it has the form

$$p(z|\hat{F}) \approx \sum_i p(\hat{F}|z, z_i, F_i)\, p(z|z_i, F_i) p(z_i, F_i), \tag{3}$$

where $p(\hat{F}|z, z_i, F_i)$ is the prediction for fluxes of the training galaxy at a different redshift z, while $p(z|z_i, F_i)$ and $p(z_i, F_i)$ are the priors that provide the redshift distributions and abundances, generated from the training data, which are multiplied to give the combined probability $p(z, z_i, F_i)$ for a given redshift z and training object with redshift z_i and fluxes F_i. This is analogous to the one derived from template-based methods,

$$p(z|\hat{F}) \approx \sum_i p(\hat{F}|z, t_i)\, p(z|t_i) p(t_i), \tag{4}$$

where t_i is the template, $p(z|t_i)p(t_i) = p(z, t_i)$ is the prior, and $p(\hat{F}|z, t_i)$ is the probability of the predicted flux $\hat{F}$ at redshift z and for template t_i. Both equations are easily differentiated by the fact that for template-based methods, $p(z|\hat{F})$ is derived using a list of templates t_i, while for machine learning methods it is derived using the individual training set objects with fluxes F_i and spectroscopic redshift z_i.

DELIGHT differs a little from the usual machine learning method in the sense that instead of finding a direct empirical relationship between the fluxes and redshifts of the training objects, it uses a GP to model the predicted fluxes of a training galaxy at different redshifts with the help of SED templates. This creates a latent flux-redshift template for each training object, where for a given set of fluxes in the testing set, it could be compared to several training templates to find the best predicted redshift.

The algorithm first fits a best-fitting SED template to a particular training object i with redshift z_i and fluxes F_i (multiple bands); the best-fitting SED template is then used to formulate the mean function and kernel of a GP to build a flux-redshift template that could predict the expected fluxes of certain band filters when this object is redshifted to a different z. With each training object now becoming a flux-redshift template, the final photo-z posterior distribution of a testing set object is determined by making a pairwise comparison of every training–testing pair, and a weighted solution is obtained based on the best fits of each pair.

In other words, we are computing the probability that the target galaxy has the same SED as the training galaxy but at a different redshift. DELIGHT is thus a hybrid template machine learning photo-z algorithm in the sense that SED templates are used to 'guide' the creation of flux-redshift templates based on the training objects, or, if seen from another perspective, the GP 'corrects' the SED templates by using training data. We refer the reader to Leistedt & Hogg (2017) for more on GPs, and also for the full expressions of the μ and k in relation to the filter responses, flux normalizations, linear mixtures of physical SED templates, and the manually configurable SED residual function of emission lines.

DELIGHT is advantageous over many other photo-z algorithms as its output is less dependent on representative training data, and it does not strictly require the training set to use the same photometric bands. However, it still requires accurate spectroscopic redshifts, high-quality training fluxes, and representative templates to produce high-quality photo-z PDFs, or $p(z)$. As such, given a few photometric bands, DELIGHT is able to predict missing bands or fluxes in an entirely different set of photometric bands, and this function is utilized in Section 4.1 to predict and calibrate the flux values between two surveys.

3.2 DELIGHT optimization

The optimization settings of DELIGHT used in this work are as follows. For the GP set-up, the number of Gaussians to fit the filter curves (`numGpCoeff`) was set to seven instead of the default 20, appropriately selected to accommodate the smaller full width at half-maximum (FWHM) of the narrow-band filters. Other than that, we have mainly used the default hyperparameter settings for DELIGHT with the exception of the widths of the luminosity and redshift priors σ_ℓ and σ_z (`ellPriorSigma` and `zPriorSigma`; see Leistedt & Hogg 2017), which have been lowered to 0.2 and 0.1, respectively, as they produced better results.

As mentioned earlier, the mean function and the kernel of the GP are modelled after the choice of emission lines and SED template sets. We replaced the three default emission lines in DELIGHT with the list provided by Eriksen et al. (2019), although we note that the change in result for this is insignificant. As for the templates, we used the Brown et al. (2014) high-quality templates, which consist of 129 SEDs derived from real nearby galaxies. These templates have wavelengths covering the UV to mid-IR, and encompass a broad range of galaxy types including ellipticals, spirals, merging galaxies, blue compact dwarfs, and luminous IR galaxies. In this work we have also tested the performance of various other template sets (Coleman, Wu & Weedman 1980; Kinney et al. 1996; Bruzual & Charlot 2003; Ilbert et al. 2006; Polletta et al. 2007); however, they do not perform

as well as those of Brown et al. (2014): the root-mean-square photo-z errors could range between 21 and 112 per cent higher when these templates are used. Therefore, the results from these tests are not shown in this work.

We note that DELIGHT requires all magnitudes m_i and magnitude errors to be converted into fluxes F_i and flux variances, with a zero-point adjustment of 26.4 in magnitude (i.e. $F_i = 10^{-0.4(m_i-26.4)}$). We have also added a 3 and 6 per cent flux error in quadrature to the flux variances for the narrow and broad bands, respectively, to account for other flux errors from both the data and the model (values estimated via trial and error). It is also worth mentioning that while DELIGHT is capable of processing negative fluxes (non-detections), the reference band (`referenceBand`) used for flux normalization only handles fluxes with positive values. In this work, we have selected the narrow-band *nb*625 as the reference band, or the COSMOS r band in cases where narrow bands were not used.

Throughout this work, we use $z_{\rm map}$ (the maximum a posteriori of the PDF) to represent the best point estimate photo-z produced by DELIGHT. The output photo-z PDF bins were set to be linear instead of logarithmic, with a step size of 0.001, and a range of $0.02 < z < 1.65$, keeping close to the limits of the spectroscopic redshifts.

3.3 Other algorithms

We are also interested in how DELIGHT compares to other common-template-based or machine-learning-based methods besides BCNZ2 and DEEPZ. Therefore two other photo-z algorithms, ANNZ2 and BPZ, are also used in this work, using the same training and template sets, to be compared with the performance of DELIGHT. In the following paragraphs, we briefly introduce the two algorithms and their optimization settings.

ANNZ2[4] (Sadeh, Abdalla & Lahav 2016) is a machine-learning-based photo-z algorithm that has been widely used in recent works (Bonnett et al. 2016; Jouvel et al. 2017; Bilicki et al. 2018; Soo et al. 2018; Schmidt et al. 2020) due to its high customizability and its ability to produce PDFs. It uses the Toolkit for Multivariate Data Analysis (TMVA; Hoecker et al. 2007) with ROOT (Brun & Rademakers 1997), which allows it to run multiple different machine learning algorithms for training, and outputs photo-zs based on a weighted average of their performance. In this work, we ran ANNZ2 with a mixture of three machine learning methods, namely artificial neural networks (ANNs), boosted decision trees (BDTs), and k-nearest neighbours (KNNs); see Hoecker et al. (2007) for detailed descriptions of these machine learning algorithms. An architecture of $N{:}\frac{2N+1}{3}{:}\frac{N+2}{3}{:}1$ was used for the ANN; the bagging method was used to boost the decision trees; a polynomial kernel was used for the KNN; while the other hyperparameters for each method were individually optimized for best performance. ANNZ2 version 2.3.1 was used in this work, and the mean value of the PDF, $z_{\rm pdf}$, was chosen to represent the photo-z point estimate.

BPZ[5] (Benítez 2000), on the other hand, is one of the long-standing template-based photo-z algorithms, and still widely used today (Martí et al. 2014; Bundy et al. 2015; Cavuoti et al. 2017; Tanaka et al. 2018; Joudaki et al. 2020; Raihan et al. 2020). Other than sharing the usual attributes of a template-based code, BPZ uses Bayesian inference, prior information of redshift distributions, and template interpolation to improve photo-z results. BPZ version 1.99.3 was used in this work, and similar to DELIGHT the Brown templates were used, with the interpolation parameter set to 2. We assumed the same functional form for the Bayesian priors as those used by COSMOS (Laigle et al. 2016). The peak of the PDF, $z_{\rm b}$, was used as the best photo-z point estimate.

Other than ANNZ2 and BPZ, the results of DELIGHT are also compared to the results of BCNZ2, which was developed specifically for the PAUS data (Eriksen et al. 2019). BCNZ2 is able to compute a linear combination of SED templates and is designed to deal with emission lines, extinction, and adjust zero-points between narrow and broad bands, all of which are crucial in the context of PAUS. The introduction of the code BCNZ2 and its early demonstration of PAUS photo-z can be found in Eriksen et al. (2019).

[4]https://github.com/IftachSadeh/ANNZ
[5]http://www.stsci.edu/ dcoe/BPZ/

4 FLUX CALIBRATION

This work utilizes fluxes obtained from two different surveys: the PAUS narrow-band fluxes are measured using an aperture that covers 62.5 per cent of light from the galaxy, while COSMOS broad-band fluxes are measured using a fixed 3 arcsec aperture. Therefore, calibration is required to ensure that the flux values are consistent with one another. We only calibrate the broad-band fluxes, leaving the narrow-band fluxes untouched following Eriksen et al. (2019). The calibration process is done in two steps: first we derive empirical corrections to account for differences in the aperture photometry (calibration for each galaxy), then placing all bands at the same flux zero-point (calibration for each band). For the correction for differences in flux aperture, we note that ideally this could have been easily done if spec-zs are available; however, since the evaluation set would not have spec-zs available, we present three alternatives in the following sections to calibrate the fluxes photometrically.

4.1 Correction for differences in flux aperture

In the first step, we define a parameter $R_{\rm g}$, a correction factor estimated for each galaxy to be multiplied with all of its six *uBVriz* broad-band fluxes. Ideally, this factor is estimated by first finding the best-fitting Brown template for each galaxy using only 40 narrow-band fluxes from PAUS and its true redshift. The best-fitting template is then used to generate the predicted *uBVriz* fluxes, and a weighted mean of the ratios between the predicted flux and the original COSMOS flux $R_{{\rm g},\,b}$ is calculated for each band b, given by

$$R_{\rm g} = \frac{\sum_b R_{{\rm g},b}/\sigma^2_{R_{{\rm g},b}}}{\sum_b 1/\sigma^2_{R_{{\rm g},b}}}, \tag{5}$$

where the sum is over the six COSMOS broad bands, and $\sigma^2_{R_{{\rm g},\,b}}$ is the variance of $R_{{\rm g},\,b}$. Here we have assumed that the Brown templates are sufficiently representative, and therefore the predicted flux derived from it is the true flux of the broad bands. We have also assumed that $R_{{\rm g},\,b}$ should be almost the same across each band for each galaxy. This calibration is motivated by the fact that each galaxy requires a calibration between fixed size and adaptive aperture photometry dependent on its apparent size.

We now explore three different methods to determine $R_{\rm g}$ from the photometric data only.

4.1.1 The photo-z calibration method

The first method, which we call the *photo-z calibration method*, is very similar to the method above except that we replace the spectroscopic redshifts used to determine the predicted *uBVriz* flux for the testing set with photometric redshifts. We first use DELIGHT

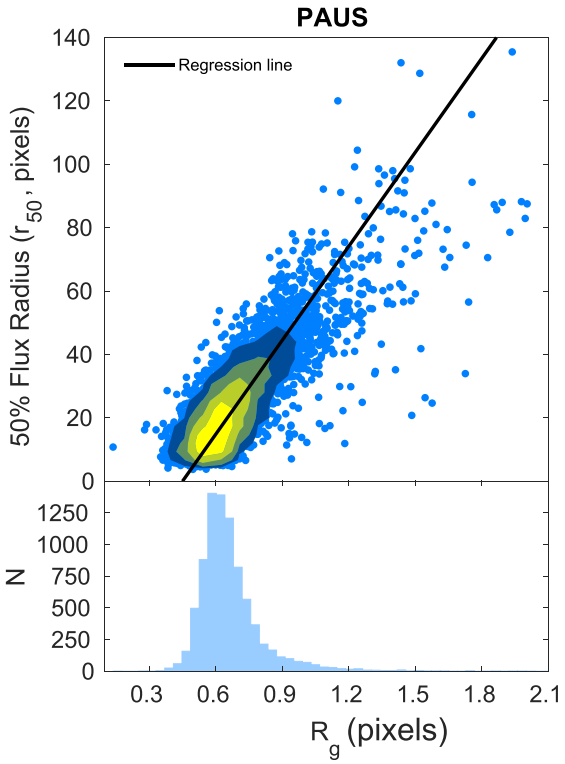

Figure 3. Top: correlation between r_{50} and R_g for the training set, where R_g is a calibration correction factor estimated for each galaxy to be multiplied with all of its six *uBVriz* broad-band fluxes. Bottom: the distribution of R_g of the training set, estimated using the size calibration method. N is the number of galaxies.

and only the 40 narrow bands to produce photo-zs for each object, and then we use these photo-zs to estimate the predicted fluxes, and then later R_g for each galaxy. This implies that the better the quality of the photo-zs produced by only the 40 narrow bands, the better the calibrated broad-band fluxes will be.

4.1.2 The size calibration method

The second method, hereafter the *size calibration method*, does not require the production of predicted fluxes for the testing set. Instead, this method uses the correlation between the sizes of galaxies with their values of R_g in the training set, to predict the values of R_g for objects in the testing set. With the predicted fluxes of the training set known, we plot R_g against the 50 per cent light radius r_{50} (measured in pixels) for each object, and obtain a best-fitting linear-least-squares regression line in the process,

$$R_g = m\, r_{50} + c, \tag{6}$$

where the slope and y-intercept are found to be $m = 0.0101$ and $c = 0.4504$, respectively, with a correlation coefficient of $r = 0.8349$, implying a strong positive correlation between R_g and r_{50}.

With this relationship derived, the values of R_g for each object in the testing set can be estimated. This method is motivated by the fact that the size of galaxies is a defining factor for the difference in their flux values when measured using a fixed aperture or when measured using a fixed light radius. Fig. 3 shows a scatter plot of r_{50} versus R_g for the training set, where the correlation equation is determined. The distribution of R_g is also tabulated in the figure, it is shown to have a median value of 0.6349, implying that on average COSMOS measures more flux for each galaxy than PAUS. We note that in the case when galaxies have undefined values of r_{50}, we substitute them with the mean value of $r_{50} = 22.4934$ pixels.

4.1.3 The flux calibration method

The third and final method is the *flux calibration method*, which is similar to the method used by Eriksen et al. (2019), but simpler in that the GP has a larger capacity to accommodate uncertainties. This method makes use of the fact that there are overlaps in wavelength between the COSMOS broad bands and PAUS narrow bands: the V band overlaps with the narrow bands *nb*505–*nb*585 (nine bands); the r band overlaps with *nb*565–*nb*685 (13 bands); and the i band overlaps with *nb*705–*nb*835 (14 bands). This overlap is illustrated in Fig. 4.

Similar to the previous method, no redshift information is required for flux prediction, the R_g in this case is estimated by first averaging the narrow-band fluxes within the range of the broad-band of interest (V, r, or i), and then taking the ratio between the broad-band flux and the averaged narrow-band fluxes. This will give us three values of $R_{g,\,b}$ for the three *Vri* bands, and finally R_g for each galaxy is taken as the weighted average of the three values.

This method is simple yet effective: it does not involve the spectroscopic redshift, the photo-z derived by 40 narrow bands, or even the size of the galaxy. Here we assume that the R_g estimated using *Vri* is applicable for the *uBz* bands as well. We will compare the overall photo-z quality produced by the three methods above in Section 5.2.

4.2 Correction to flux zero-points

After calibrating the COSMOS broad-band fluxes for each galaxy, we proceed to calibrate the broad-band magnitude offsets within each band. We perform a weighted least-squares fit between the predicted broad-band fluxes (produced by DELIGHT using 40 PAUS narrow-band fluxes, the respective best-fitting Brown templates, and zCOSMOS spec-zs) and the original COSMOS *uBVriz* fluxes in the training set, by using a simple linear equation,

$$\ln(F_{p,b}) = a_b\, \ln(F_{g,b}) + c_b, \tag{7}$$

where $F_{p,\,b}$ is the predicted flux for band b, $F_{g,\,b}$ the COSMOS broad-band flux after undergoing the per-galaxy calibration, and a_b and c_b are constants to be optimized. The values of a_b and c_b estimated for each band using the training set are now used to calibrate the fluxes in the testing set, and these values are tabulated in Table 1. A weighted fit was implemented, with the inverse variances of the fluxes used as the weights, since we expect that objects that are brighter to have relatively lower variances, and by accounting for the variances of objects the fainter objects would be upweighted.

As expected from the table, the values of a_b and c_b are very close to 1 and 0, respectively, since the calibrated flux for aperture correction $F_{g,\,b}$ is already very close to the predicted flux $F_{p,\,b}$. Essentially, this process 'straightens' the correlation line, providing minor yet essential improvements to the overall calibration.

4.3 Overall calibration performance

Fig. 5 shows the correlation between the broad-band fluxes predicted by DELIGHT (using spectroscopic redshifts, PAUS 40 narrow bands,

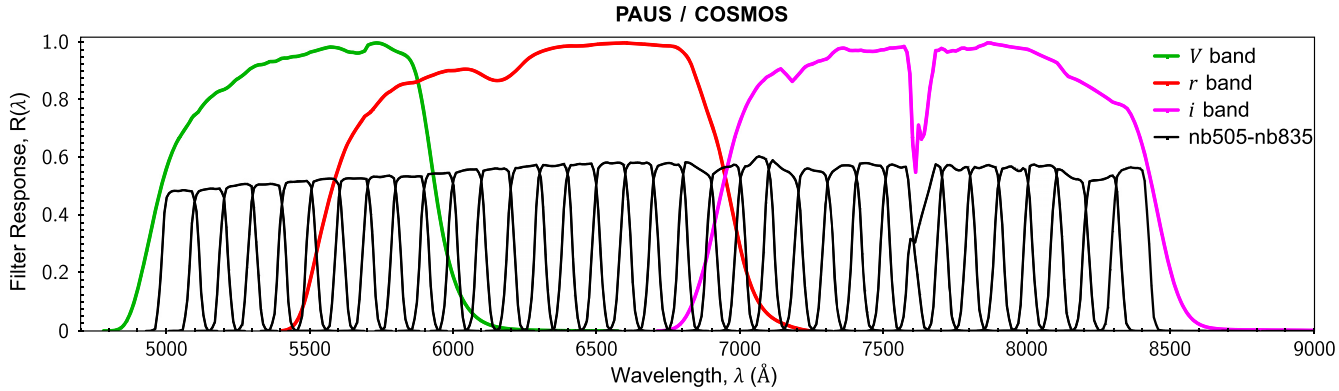

Figure 4. The overlapping wavelengths between 34 PAUS narrow-band filters and three COSMOS broad-band filters: *V* band overlaps with *nb*505–*nb*585 (nine bands); *r* band overlaps with *nb*565–*nb*685 (13 bands); and *i* band overlaps with *nb*705–*nb*835 (13 bands). Note that the filter responses from PAUS and COSMOS are normalized at different values, respectively.

Table 1. List of the best-fitting parameters a_b and c_b for each band b when the predicted and original COSMOS fluxes from the training set were fitted with a weighted least-squares fit, using equation (7).

Bands	a_b	c_b
u	1.0007 ± 0.0001	0.0354 ± 0.0008
B	0.9906 ± 0.0002	0.2163 ± 0.0009
V	0.9988 ± 0.0002	-0.0830 ± 0.0009
r	1.0006 ± 0.0002	0.0015 ± 0.0009
i	1.0202 ± 0.0001	-0.0875 ± 0.0008
z	0.9791 ± 0.0001	0.0424 ± 0.0007

and Brown templates) and the COSMOS broad-band fluxes for our training set, both before and after calibration (red and blue, respectively). The figure only shows the result of the flux calibration method, as the other two methods look very similar graphically (which translates to a small difference in photo-z results shown later in Section 5).

The rms values displayed in Fig. 5 show that for all bands, the scatter between the original fluxes with respect to the predicted fluxes has reduced by 63–88 per cent after the two-step calibration was done. The scatter at low fluxes for the *u* and *B* bands remains evident, which originated from the high uncertainty in flux measurements. Despite the large decrease in scatter, we note that the rms value here is not a metric of improvement for calibration as we do not have the true values of the broad-band fluxes in the matched apertures. However, the calibration of the broad-band fluxes did translate into an improvement in photo-z scatter and 68th percentile error by about 70–80 per cent, as shown in Section 5.

5 RESULTS AND DISCUSSION

Table 2 summarizes the results of this work, it shows all the photo-z metrics we produced, using different algorithms (DELIGHT, ANNZ2, and BPZ), different calibration methods (*flux*, *photo-z*, and *size*), and different number of input fluxes (six broad bands, 40 narrow bands, or both). We divide the analysis of the results into two sections: Section 5.2 studies the performance between the three calibration methods used in DELIGHT, while Section 5.3 compares the best performance of DELIGHT with ANNZ2, BPZ, and BCNZ2. In the following section, we briefly introduce the performance metrics we used in this work.

5.1 Performance metrics

In this work, we use three metrics to quantify the performance of the photo-z point estimates: the root-mean-square error ($\sigma_{\rm rms}$), the 68th percentile error (σ_{68}), and the outlier fraction rate ($\eta_{\rm out}$). With $\Delta z \equiv \frac{z_{\rm phot}-z_{\rm spec}}{1+z_{\rm spec}}$, the above metrics are defined as follows:

$$\sigma_{\rm rms} \equiv \sqrt{\frac{1}{N}\sum_{i}^{N}|\Delta z_i|^2}, \tag{8}$$

$$\sigma_{68} \equiv \frac{Q_{84.1\ \text{per cent}}(\Delta z_i) - Q_{15.9\ \text{per cent}}(\Delta z_i)}{2}, \tag{9}$$

$$\eta_{\rm out} \equiv \text{ per cent objects, where } |\Delta z_i| \geq 0.15. \tag{10}$$

Here N is the total number of galaxies, while Q is a per-centile of the distribution. Since $\sigma_{\rm rms}$ is calculated without the outliers removed, it measures the overall scatter of the sample, whereas σ_{68} measures the scatter with reduced sensitivity to outliers.

With similar motivations as Martí et al. (2014) and Eriksen et al. (2019), we hope to achieve an overall photo-z error of $\sigma_{68} \leq 0.0035(1 + z_{\rm spec})$ for at least 50 per cent of the testing sample after applying an appropriately chosen quality cut. We use the Bayesian odds (Θ) parameter (Benítez 2000) in DELIGHT, similar to its implementation in ANNZ2 by Soo et al. (2018). Θ can be estimated from the photo-z PDF, $p(z)$, using the equation

$$\Theta = \int_{z_{\rm p}-k(1+z_{\rm p})}^{z_{\rm p}+k(1+z_{\rm p})} p(z)\,{\rm d}z, \tag{11}$$

where $z_{\rm p}$ is the peak of $p(z)$ and $k = 0.01$. Θ ranges between 0 and 1, the higher the value, the lower the $p(z)$ width, which implies a more precisely predicted photo-z (though not necessarily accurate). The value of k is arbitrary, appropriately selected such that not too many objects end up having $\Theta = 1$. Therefore, an x per cent quality cut on the sample keeps the top x per cent of objects with the highest values of Θ.

To assess the quality of the $p(z)$, we use probability integral transform (PIT) plots and the continuous ranked probability score (CRPS). The PIT is the cumulative distribution function (CDF) at

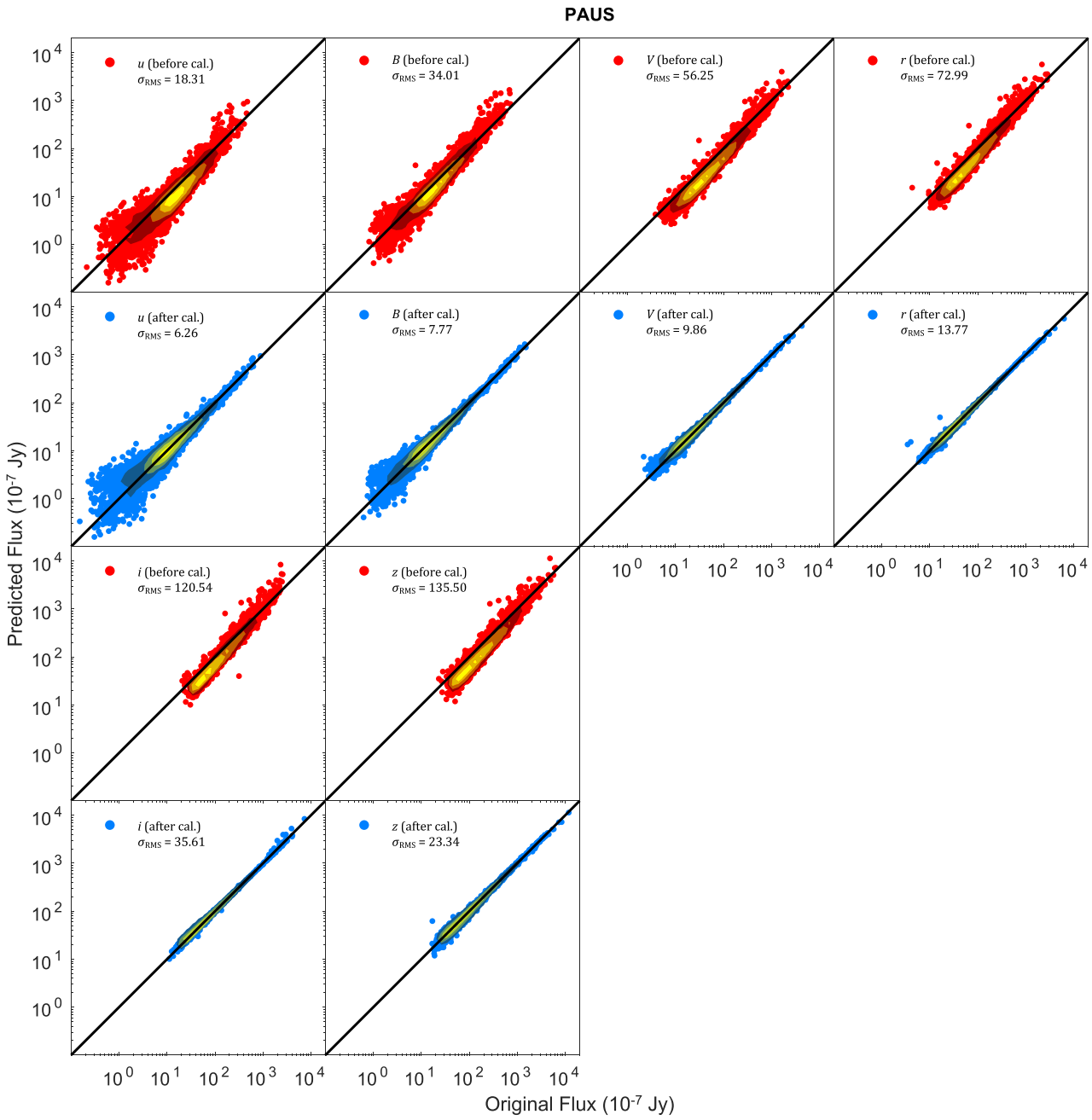

Figure 5. The *uBVriz* broad-band fluxes predicted by DELIGHT plotted against their original COSMOS fluxes, both before and after the two-step calibration process (red and blue, respectively) for our training set, using the flux calibration method as an example. Based on the root-mean-square errors ($\sigma_{\rm rms}$) shown in each panel, the broad-band fluxes match their prediction much better after calibration.

$z_{\rm spec}$ while asserting the $p(z)$ to have an area of unity. Since the photo-z CDF is $C(z) = \int_0^z p(z')\,{\rm d}z'$, PIT is defined to be

$$\mathrm{PIT} = C(z_{\rm spec}) = \int_0^{z_{\rm spec}} p(z)\,{\rm d}z. \tag{12}$$

A PIT distribution tells us on average if the $p(z)$ produced are 'adequately shaped': the shape of the PIT distribution can tell us if the $p(z)$ produced are generally too wide/narrow, or if the $p(z)$ are over-/underpredicting the true redshift.

The CRPS on the other hand tells us how well the $p(z)$ encapsulates or predicts the true redshift ($z_{\rm spec}$). The CRPS of a $p(z)$ can be expressed as

$$\mathrm{CRPS} = \int_{-\infty}^{\infty} \left|C(z) - \mathcal{H}(z - z_{\rm spec})\right|^2 {\rm d}z, \tag{13}$$

where $\mathcal{H}(z - z_{\rm spec})$ is the Heaviside step function with

$$\mathcal{H}(z - z_{\rm spec}) = \begin{cases} 1, & z = z_{\rm spec}, \\ 0, & \text{otherwise}. \end{cases} \tag{14}$$

In this work, we use the symbol $\rho_{\rm CRPS}$ to represent the average CRPS value of all galaxies in the testing sample, in which the smaller the value, the better the $p(z)$ are at predicting their true redshifts.

Table 2. The root-mean-square error ($\sigma_{\rm rms}$), 68th percentile error (σ_{68}), outlier fraction ($\eta_{\rm out}$), mean continuous ranked probability score ($\rho_{\rm CRPS}$), and the root-mean-square error in redshift distribution ($n_{\rm rms}$) for the photo-zs produced in this work, using different algorithms, methods, and number of bands. All results are produced using six broad bands (BB) and 40 narrow bands (NB) unless stated otherwise.

Photo-z methods	$\sigma_{\rm rms}$	σ_{68}	$\eta_{\rm out}$ (per cent)	$\rho_{\rm CRPS}$	$n_{\rm rms}$
DELIGHT (six BB only)	0.0514	0.0441	0.93	0.0388	0.885
DELIGHT (40 NB only)	0.0684	0.0119	4.02	0.0298	0.637
DELIGHT (no calibration)	0.1555	0.0566	9.06	0.0887	0.895
DELIGHT (photo-z calibration method)	0.0335	0.0083	0.71	0.0158	0.634
DELIGHT (size calibration method)	0.0341	0.0095	0.76	0.0165	0.646
DELIGHT (flux calibration method)	0.0331	0.0081	0.86	0.0155	0.636
DELIGHT (flux calibration method, no GP)	0.0442	0.0089	0.98	0.0179	0.639
ANNZ2	0.0556	0.0396	2.66	0.0719	0.465
ANNZ2 (six BB only)	0.0371	0.0202	1.14	0.0522	0.432
BPZ	0.0368	0.0089	0.86	0.0184	0.740
BCNZ2	0.0403	0.0085	1.14	–	–

We refer the reader to Polsterer, D'Isanto & Gieseke (2016) for a detailed description of both PIT and CRPS.

Finally, we also assess the quality of the redshift distribution $n(z)$. We can find how similar the spec-z distribution $n_{\rm spec}(z)$ is compared to the photo-z distribution $n_{\rm phot}(z)$ by estimating $n_{\rm rms}$, the root-mean-square difference between the distributions:

$$n_{\rm rms} = \sqrt{\int \left[n_{\rm phot}(z) - n_{\rm spec}(z)\right]^2 \, {\rm d}z}. \tag{15}$$

$n_{\rm rms}$ provides us a quantitative measure to compare the performances of photo-z with distributions produced by different codes.

5.2 Performance of DELIGHT

Rows 1 and 2 from Table 2 show the photo-zs produced when only trained using the broad and narrow bands individually, and we find that by combining both broad and narrow bands (rows 4–6), we have achieved at least 34 per cent and 20 per cent improvement in the photo-z scatter and σ_{68}, respectively (visualized in Fig. 6).

Rows 3–7 proceed to show the metrics for each calibration method, and on average, the performance of each method is quite similar, all within 4–16 per cent difference in $\sigma_{\rm rms}$ and σ_{68}, respectively. Statistically, the flux calibration method seems to perform slightly better compared to the remaining ones, with the exception of the photo-z calibration method having better values of $\eta_{\rm out}$ and $n_{\rm rms}$. This suggests that while the photo-zs produced by training with only 40 narrow bands are not as competitive as when trained with all 46 bands and calibrated broad bands (see Table 2 and Fig. 6), it is however sufficient to guide the calibration process. Note that we have also included the results of DELIGHT run as a pure template code when calibrated using the flux calibration method for comparison, and we see that without the help of the GP, the photo-z results are similar for most metrics except a degradation in scatter of up to −33.5 per cent. Therefore the good results of DELIGHT shown here are mainly due to the use of the Brown templates, the flux calibration, the combination of broad and narrow bands, and also the work of the GP.

As the three calibration methods presented in Section 4 all result in very similar photo-z performance, we will only show results for the flux calibration method in the following. It is notable however that in all cases, the photo-z requirement of $\sigma_{68} < 0.0035(1+z)$ is achievable for all objects at $i_{\rm auto} < 20.0$, or objects with a 40 per cent Θ cut at $i_{\rm auto} < 22.5$. All three methods also show that despite such high percentage Θ cuts being implemented, a significant number of high photo-z objects still remain in the sample.

5.3 Comparison with other algorithms

Since the DELIGHT results for each of the three calibration methods are very similar to each other, we decided to select only the flux calibration method to be compared to the results obtained by the two other algorithms used in this work, ANNZ2 and BPZ. We also include the point estimates from Eriksen et al. (2019). The values of $\sigma_{\rm rms}$, σ_{68}, and other relevant metrics obtained from these algorithms are shown in rows 8–11 of Table 2, and visualized in Fig. 7.

From the figure, it is found that ANNZ2, being a purely machine-learning-based algorithm, is underperforming compared to the other algorithms. This machine learning method is unable to make full use of the extra information provided by the 40 narrow bands, and is shown to perform better without them. This is partially due to the problem of the curse of dimensionality (Bellman 1957), sharply diluting the pattern recognition power of the algorithm as the number of inputs increases. Besides, the very small training sample size may have heavily affected the potential of ANNZ2. Here we note however that the deep learning code DEEPZ is shown to work well on a similar sample (Eriksen et al. 2020), therefore we hope to do follow-up evaluations of ANNZ2 on PAUS data in the future when a larger training set is available.

In terms of the quality of the point estimate photo-zs, DELIGHT is shown to fare well against BCNZ2 and BPZ (Fig. 7), both of which are purely template-based methods. As both DELIGHT and BPZ used the same template sets in this case (i.e. the Brown templates), we find that the GP contributed to 25 per cent and 9 per cent improvement in the scatter and σ_{68}, respectively, as compared to the pure template fit of BPZ.

Despite the similarities in the point estimates for the entire sample (Table 2), when we cut the sample in percentages of Θ (Figs 8 and 9), we see two major differences. First, the cut in Θ for BPZ does not systematically remove objects with high uncertainties (especially for objects brighter than $i_{\rm auto} = 21$); and secondly, the cut in Θ for BPZ selectively removes objects with lower photo-z. In both cases, DELIGHT is shown to not only perform better in this regard as compared to BPZ, but also better than all other algorithms shown.

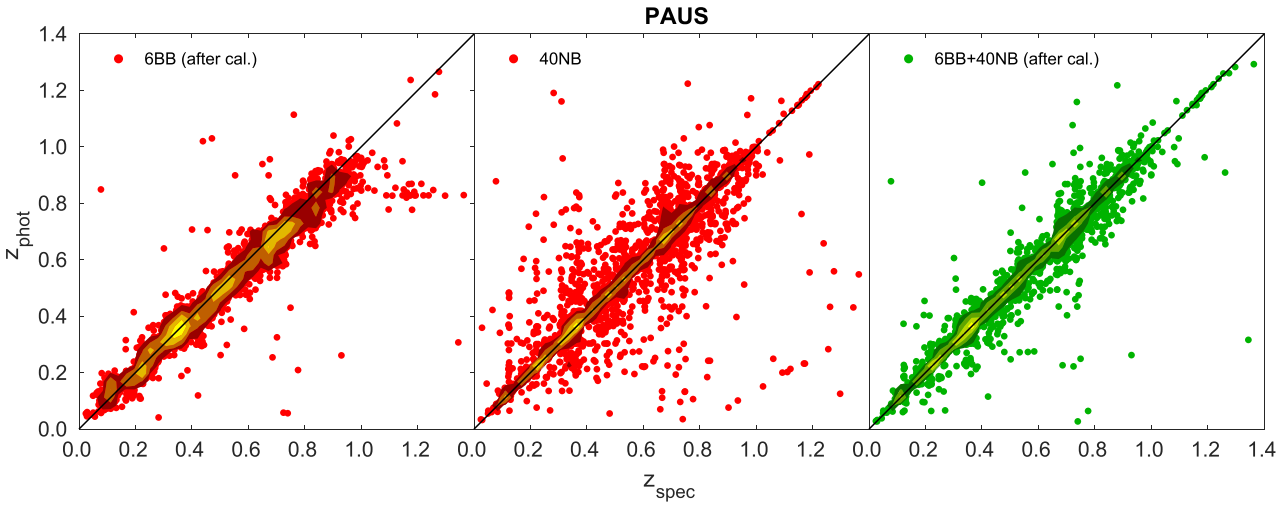

Figure 6. Plot of photo-z versus spec-z, comparing the photo-zs when trained and tested using only six *uBVriz* broad bands (BB, left), only 40 narrow bands (NB, middle), and all 46 bands combined (right). The flux calibration method was used for this plot.

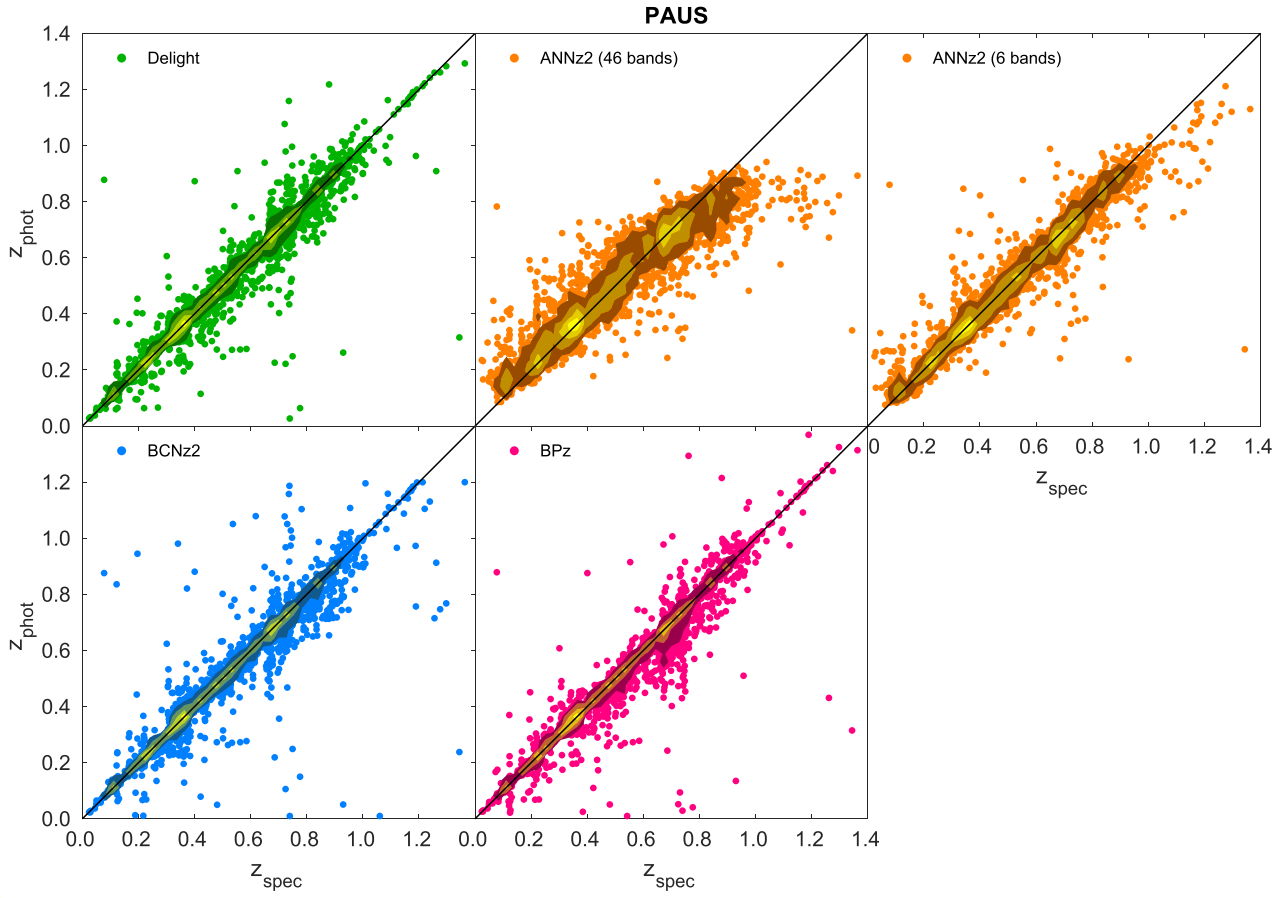

Figure 7. Plots of photo-z versus spec-z, comparing the results of the flux calibration method of DELIGHT (green), ANNZ2 trained with 46 bands/six broad bands (orange), BCNZ2 (blue), and BPZ (magenta). The same colouring scheme will be used to represent the respective methods in the following plots.

A selection of sample $p(z)$ produced by each algorithm is shown in Fig. 10, while the overall quality of the $p(z)$ produced is visualized in the PIT plots as shown in Fig. 11. Once again we see DELIGHT on average producing superior $p(z)$ compared to ANNZ2 and BPZ: it is obvious from the PIT plots that the $p(z)$ produced by ANNZ2 are too narrow (a U-shaped distribution), while those by BPZ are too wide (a significant central peak). In terms of $\rho_{\rm CRPS}$ (see Table 2), DELIGHT once again performs better than both BPZ and ANNZ2, where the adequate shapes and accurately positioned peaks of the $p(z)$ provide good predictions of the true redshift.

We note that the $p(z)$ produced by ANNZ2 are ragged compared to BPZ and DELIGHT, this is due to the limited training sample

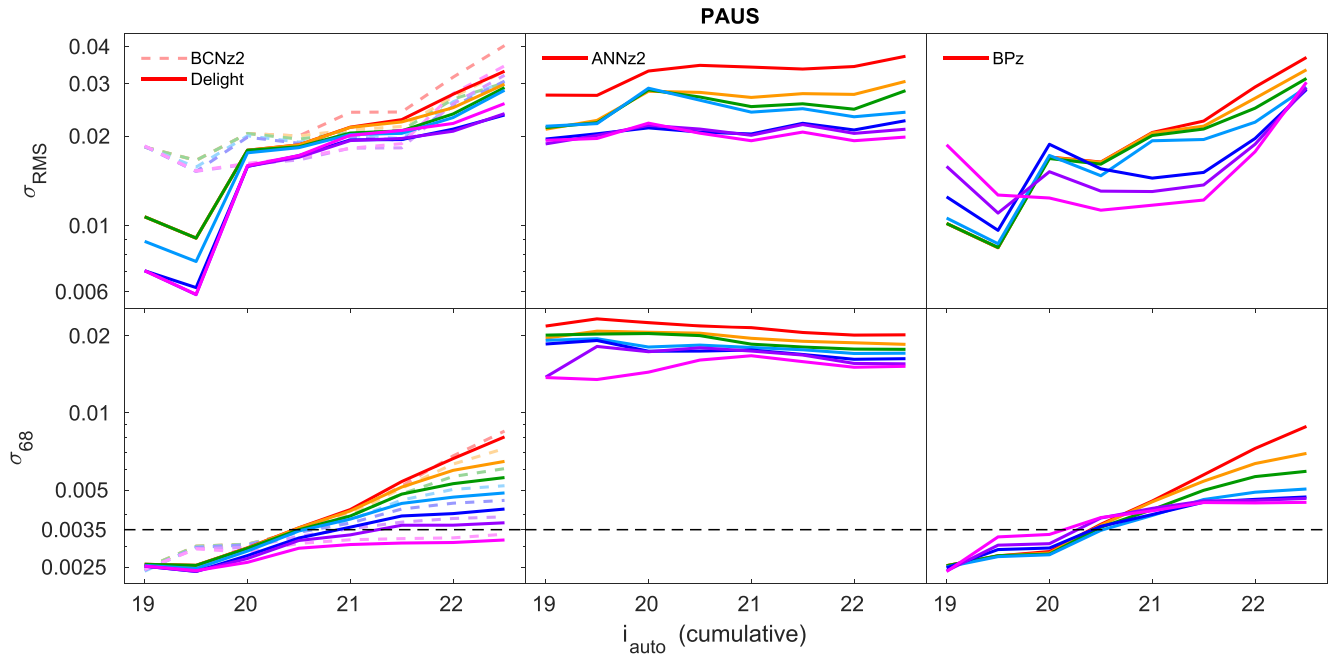

Figure 8. Plots of σ_{rms} (top) and σ_{68} (bottom) with respect to i_{auto} (cumulatively), comparing the performance of DELIGHT (left) with ANNZ2 (middle), BPZ (right), and BCNZ2 (left, dashed lines). The coloured lines represent the sample when cut systematically in the Bayesian odds (Θ), keeping only objects with the best 100 per cent (red), 90 per cent (orange), 80 per cent (green), 70 per cent (blue), 60 per cent (navy), 50 per cent (purple), and 40 per cent (magenta) values. The black horizontal dashed line with $\sigma_{68} = 0.0035(1 + z)$ represents the photo-z quality target of PAUS for 50 per cent of the objects at $i \sim 22.5$.

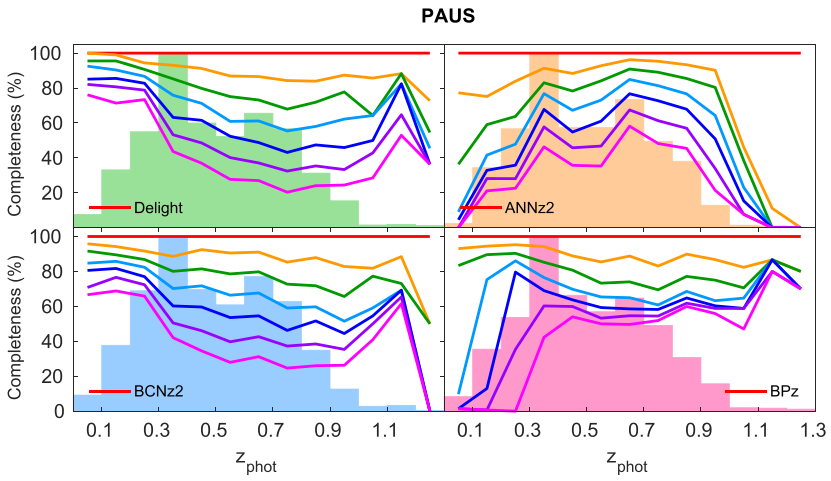

Figure 9. Plot of percentage of objects within each photo-z bin with respect to the cut in Θ value for the results of DELIGHT (top left), ANNZ2 (top right), BCNZ2 (bottom left), and BPZ (bottom right). The lines use the same colour scheme as those in Fig. 8, while the histograms in the background show the photo-z distribution for each method (relative number of objects in each photo-z bin).

size and the low number of network committees used. We intend to look into several methodologies to smoothen machine-learning-based $p(z)$ that are limited by such conditions; this is left for future work. The limited testing size has also produced an $n_{spec}(z)$ distribution that is not smooth, thus despite ANNZ2 producing an $n(z)$ closest to the spectroscopic distribution (lowest n_{rms}), it may have experienced overfitting. Having said that, for the different DELIGHT runs shown in Table 2, the values of n_{rms} are consistent with the other metrics. Therefore, we leave the analysis of $n(z)$ to future work when a large enough testing sample is available.

6 APPLICATION: IDENTIFYING PHOTO-z OUTLIERS

6.1 Analysing the photo-z outliers of DELIGHT and BCNZ2

As we compared the photo-z results, we discovered that there are some galaxies that have similar DELIGHT and BCNZ2 photo-z values; however, these redshift values are far from their respective zCOSMOS spectroscopic redshifts or broad-band photo-zs. Since both BCNZ2 and DELIGHT utilize the PAUS narrow bands, we expect that the photo-zs they produce are more sensitive to emission lines as

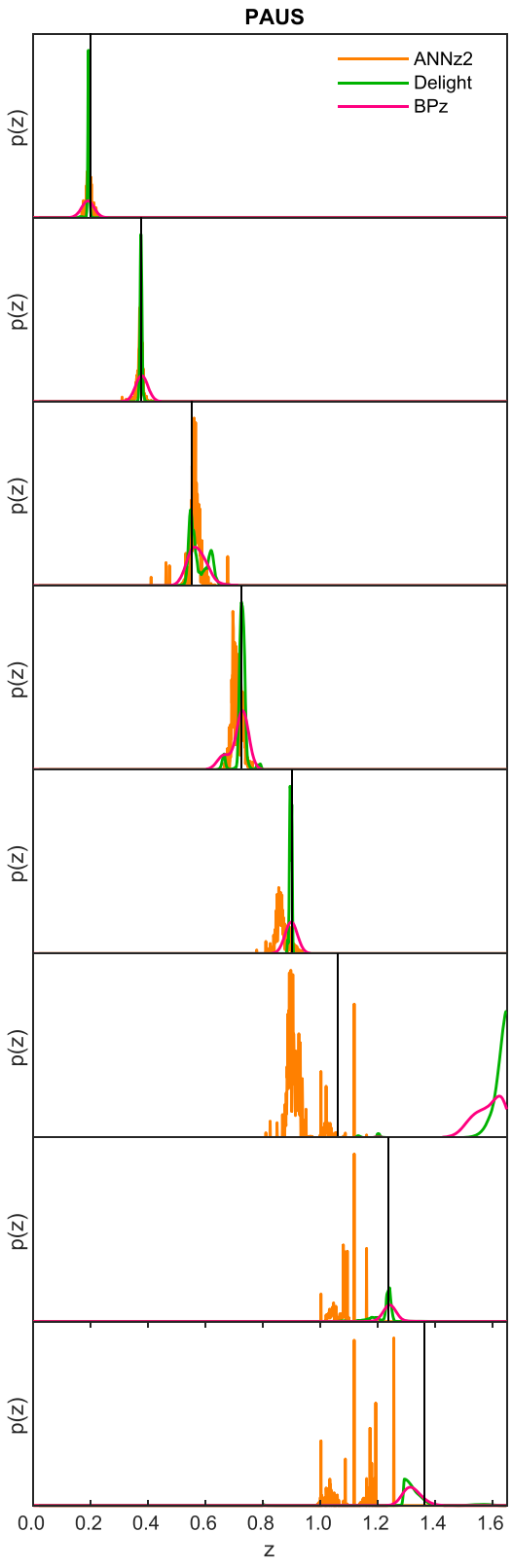

Figure 10. Sample redshift PDF $p(z)$ for the ANNZ2 (orange), DELIGHT (green), and BPZ (magenta). The black vertical lines show the positions of the spectroscopic redshifts.

compared to photo-zs produced using only broad bands. Therefore, we suspect that objects that have similar photo-z values for DELIGHT and BCNZ2 but have disagreeing spec-z values to be an indication of either having (1) a catastrophic zCOSMOS spectroscopic redshift,[6] (2) outlier broad-band or narrow-band fluxes, or (3) misidentification of close neighbours.

For the purpose of this inquiry, we have selected 30 objects from the sample that are photo-z outliers in $z_{\rm DELIGHT}$ versus $z_{\rm spec}$ or $z_{\rm BCNZ}$ versus $z_{\rm spec}$, yet are not outliers in $z_{\rm DELIGHT}$ versus $z_{\rm BCNZ}$. Mathematically, they satisfy the following conditions:

(i) $\frac{|z_{\rm DELIGHT}-z_{\rm spec}|}{1+z_{\rm spec}} \geq 0.15$ or $\frac{|z_{\rm BCNZ}-z_{\rm spec}|}{1+z_{\rm spec}} \geq 0.15$, and
(ii) $\frac{|z_{\rm DELIGHT}-z_{\rm BCNZ}|}{1+\frac{z_{\rm DELIGHT}+z_{\rm BCNZ}}{2}} < 0.15$.

Note that the $z_{\rm DELIGHT}$ used here refers to the photo-z produced using the flux calibration method, trained using 46 bands guided by the Brown templates.

These 30 objects are visualized in the redshift–redshift plots in Fig. 12. Note that in the following paragraphs, we will define a photo-z to be *catastrophic* if it is found to be an outlier with respect to its spec-z, as defined mathematically above. These objects are found to have faint magnitudes ($i_{\rm auto} > 19.75$) and small angular sizes ($r_{50} < 60$ ACS pixels, or 1.8 arcsec), which describe most galaxies of interest for PAUS. We study several different attributes of these objects, namely their respective photo-zs by DELIGHT, BCNZ2, and LEPHARE, photo-z PDFs, best-fitting templates (Brown and GP), spectra, and images. We summarize important observations according to their respective attributes below.

Photo-zs. While these 30 objects have been identified as outliers when trained using 46 bands, we find that two-thirds of these objects have non-catastrophic photo-zs when trained with either only the broad or narrow bands, respectively. In other words, only one-third of these objects have catastrophic photo-zs regardless of which bands were used in the training or fitting process. This suggests that most of the time, outlier fluxes in the broad or narrow bands may have caused a degradation in photo-z quality when trained together (more on this in the *templates* paragraph below). We have also made a comparison between DELIGHT photo-zs with those produced by LEPHARE for the COSMOS2015 catalogue (Laigle et al. 2016), and found that in fact half of the 30 objects have non-catastrophic LEPHARE photo-zs. This suggests that the IR *yJHK* bands could have played a role in improving the PAUS photo-zs, and could be incorporated in future trainings in case the PAUS photometry is problematic.[7]

Photo-z PDFs. We inspected the secondary/tertiary peaks of the PDFs for all DELIGHT runs (trained with six broad bands, 40 narrow bands, or both), and find that less than 20 per cent of these secondary/tertiary peaks coincide with their respective spec-zs. We deduce that despite the importance of secondary PDF peaks in redshift distributions, they do not significantly influence the photo-z quality of these 30 objects.

[6]While we have already selected to use only secure spectroscopic redshifts in this work, we still deem this as a possibility, since a 1 per cent outlier rate in 4000 + spec-z measurements may still yield 40 objects, which is within the same order of number of objects being investigated in this section. Our results later in this section however have verified that most of the outliers are not caused by catastrophic spectroscopic redshifts.

[7]We note that these additional bands will not be available over most of PAUS, which targets Canada–France–Hawaii Telescope Legacy Survey (CFHTLS) wide fields W1 to W4. There is however some IR data on these fields provided by the Wide-field InfraRed Camera (WIRCam) and the VISTA Kilo-degree Infrared Galaxy (VIKING) survey.

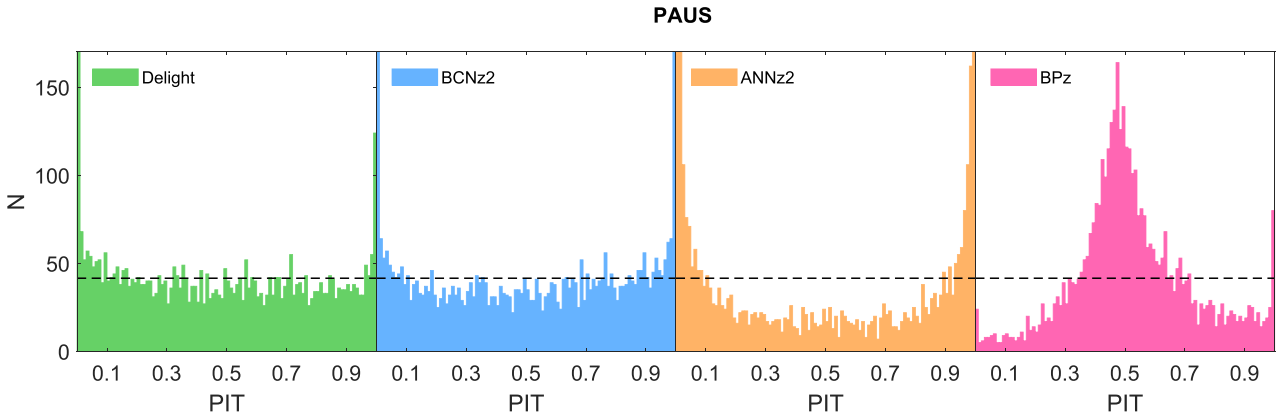

Figure 11. Probability integral transform (PIT) distributions for the $p(z)$ produced by the four different algorithms: DELIGHT (green), BCNZ2 (blue), ANNZ2 (orange), and BPZ (magenta). The dashed horizontal line indicates the mean of the distribution, and a flat distribution is ideal. A U-shaped distribution indicates that the $p(z)$ produced are too narrow, while a mountain-peak shaped distribution indicates that the $p(z)$ produced are too wide.

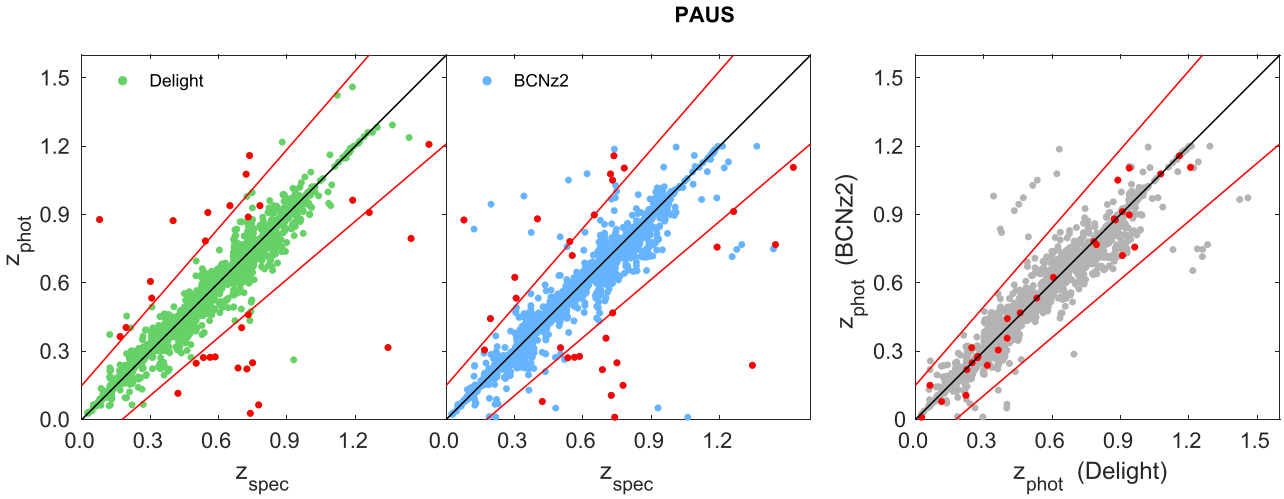

Figure 12. The selected 30 objects (red dots) marked for this outlier analysis. These objects are photo-z outliers of either DELIGHT or BCNZ2 with respect to $z_{\rm spec}$, but are not outliers with respect to each other.

Templates. DELIGHT utilizes the 129 Brown et al. (2014) templates and the 4203 training objects to guide the GP to produce the same number of new flux–redshift templates, which are used to produce photo-zs for the objects. In the training process, DELIGHT would always choose one best-fitting Brown template for each training galaxy to be trained by the GP. Here we inspected two different kinds of best-fitting Brown templates to these 30 outliers: one fixed at the spec-z, and the other with the redshift as a free parameter. In both cases, we examined

(i) if the objects fit to the same templates when trained with only broad bands, only narrow bands, or both, respectively;
(ii) if there are any trends in galaxy morphological types, based on the galaxy type classification indicated by the template;
(iii) if there is any correlation between the χ^2 value of the best-fitting templates and the quality of photo-zs; and
(iv) if any outlier narrow-band fluxes can be identified as the cause of the degradation of photo-z.

As expected, we find that 70 per cent of the outlier objects have different best-fitting Brown templates between the fits at fixed photo-z and spec-z, which contrasts with the case for non-outliers at only 35 per cent. We also find that only slightly more than a third of both the outlier and non-outlier objects were fitted to the same templates when trained using broad bands as compared to trained with all 46 bands. The high percentage of objects with different template fits at different reference redshifts (photo-z or spec-z) and flux combinations (broad bands, narrow bands, or both) also resulted in no trend in galaxy morphological types among the outliers.

However, it was found that up to 60 per cent of the objects have their best-fitting template χ^2 value correlating with the quality in photo-z, which further affirms the usage of this as a metric to remove unreliable photo-zs (see Section 6.2), as also attempted by Eriksen et al. (2019, 2020).

Perhaps a more significant finding from the study of the best-fitting templates is the ability to identify outlier narrow-band fluxes. Fig. 13 shows an example that highlights the importance of identifying outlier narrow-band fluxes, which is shown to significantly affect the photo-z results. It was found that a third of the 30 objects contained outlier narrow-band fluxes, which results in entirely

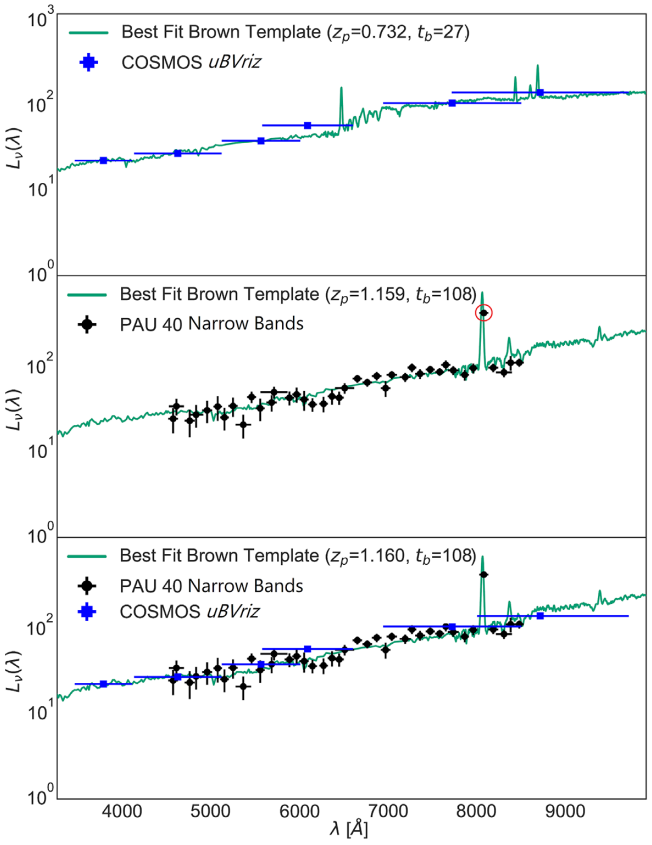

Figure 13. A sample of best-fitting Brown templates (unfixed redshift) when fit to only broad-band fluxes (top), only narrow-band fluxes (middle), and both fluxes (bottom) for the galaxy with zCOSMOS ID 805216. $L_\upsilon(\lambda)$ is the rest-frame luminosity density (or SED) of the galaxy. This galaxy has $z_{\rm spec} = 0.736$, $z_{\rm p}$ and $t_{\rm b}$ in the figure refer to its photo-z and best-fitting Brown template number, respectively. The outlier narrow-band flux shown in the middle panel (red circle) has caused a misfit in template type, resulting in erroneous photo-zs for both cases.

different template fits and photo-zs when trained with narrow bands, as compared to when trained with broad bands only. Among these 10 objects, eight of them are shown to have worse photo-z as compared to training without the narrow bands. We find indications for a significant fraction of narrow-band flux outliers also for galaxies without catastrophic redshift failures. Forthcoming PAUS data reductions will therefore implement methods to identify and correct flux outliers.

Images. We inspect the individual object images compiled by zCOSMOS DR3, these are 5 × 5 arcsec^2 images observed by the *Hubble Space Telescope*/Advanced Camera for Surveys (*HST*/ACS) in the F814W filter (Koekemoer et al. 2007). Among the 30 outlier objects, we find 63.3 per cent and 26.7 per cent of them having bright neighbours within 5 and 3 arcsec of the primary source, respectively. Having said that, we have not found any correlation between the presence of bright neighbours to the other attributes that we have studied thus far. In fact the opposite is true: we find that 60 per cent of the objects with outlier narrow-band fluxes actually have primary sources without any bright neighbours in vicinity.

Spectra. So far we have assumed that the zCOSMOS spectra obtained are reliable, as only entries with high-confidence quality flags have been selected for training (see Section 2.3). In order to probe further, we examined the one-dimensional spectra obtained by the VIMOS spectrograph, which is processed by the VIMOS Interactive Pipeline and Graphical Interface (VIPGI; Scodeggio et al. 2005) to produce the zCOSMOS spec-zs used in this work. The spectra have a range between 5500 and 9450 Å, measured with a resolution of $R \sim 600$ at 2.5 Å pixel^{-1} (Lilly et al. 2009).

We used the redshift measurement tool EZ (Garilli et al. 2010) to inspect the spectra of the 30 outlier objects, and compared our best fits to the spectroscopic redshift produced by zCOSMOS, and also the photo-zs produced by DELIGHT, BCNZ2, DEEPZ, LEPHARE (COSMOS2015), and those of Alarcon et al. (2020).

Upon inspection, we find that up to 10 of these objects (33 per cent) have disputable zCOSMOS spec-z (e.g. two possible redshift values, different best-fitting redshift values, line confusion, and low signal-to-noise ratio). However, most of these potential spec-z failures could be force fitted to the zCOSMOS spec-z and still look satisfactory, which leaves only two (6.7 per cent) of these objects having truly catastrophic spec-zs. Both these objects are found to have better EZ fits at redshift values within 10 per cent uncertainty from the photo-zs produced by DELIGHT and other algorithms. The spectrum of one of these objects is shown in Fig. 14. We have also found one isolated case where the spectra belonged to a bright neighbour and have been mismatched to the PAUS photometry.

Generally, the higher redshift objects are identified by clear O II (3727.1 Å) emission lines, while the lower redshift objects are identified by clear Hα (6564.6 Å) emission lines. We therefore conclude that although catastrophic spec-zs played a role in this situation, our results did not provide enough evidence to say that it is a major cause for catastrophic photo-zs produced by BCNZ2 and DELIGHT. This is not surprising since we have only selected secure spectroscopic redshifts from COSMOS to be used in this work. However this highlights the usefulness of multiple PAUS photo-zs being used to determine failure rates in insecure spectroscopic redshifts.

To summarize this part, we believe that the potentially important source for catastrophic photo-zs in the context of PAUS is the outlier narrow-band fluxes, with weak evidence for the existence of a small number of spec-z failures. We leave the tackling of outlier narrow-band fluxes to future work, but in the following section, we attempt to improve our process to identify and remove these outlier photo-zs.

6.2 New metrics to remove photo-z outliers

In Figs 8 and 9, we have used the Bayesian odds (Θ) to cut the sample, and the aim of this was to keep as many objects as possible while achieving the goal of $\sigma_{68} \leq 0.0035(1+z)$. Here, we extend our previous results further towards that goal by introducing several new metrics to better separate the photo-z outliers from the sample. These metrics are motivated by the inspection of the 30 outliers in Section 6.1, and they are defined as follows.

(i) The *DELIGHT–BCNZ2 metric* ($\Delta_{\rm DB}$),

$$\Delta_{\rm DB} \equiv \frac{|z_{\rm DELIGHT} - z_{\rm BCNZ}|}{1 + \frac{z_{\rm DELIGHT} + z_{\rm BCNZ}}{2}}, \tag{16}$$

a metric used to identify the similarity between DELIGHT and BCNZ2 photo-zs. It is plausible that, in general, the closer the photo-zs between the two algorithms, the more reliable they are.

(ii) The *DELIGHT photo-z standard deviation* ($\sigma_{\rm D}$), which is the standard deviation between all DELIGHT photo-z runs regardless of calibration method and number of bands. Smaller deviations could indicate more reliable photo-zs.

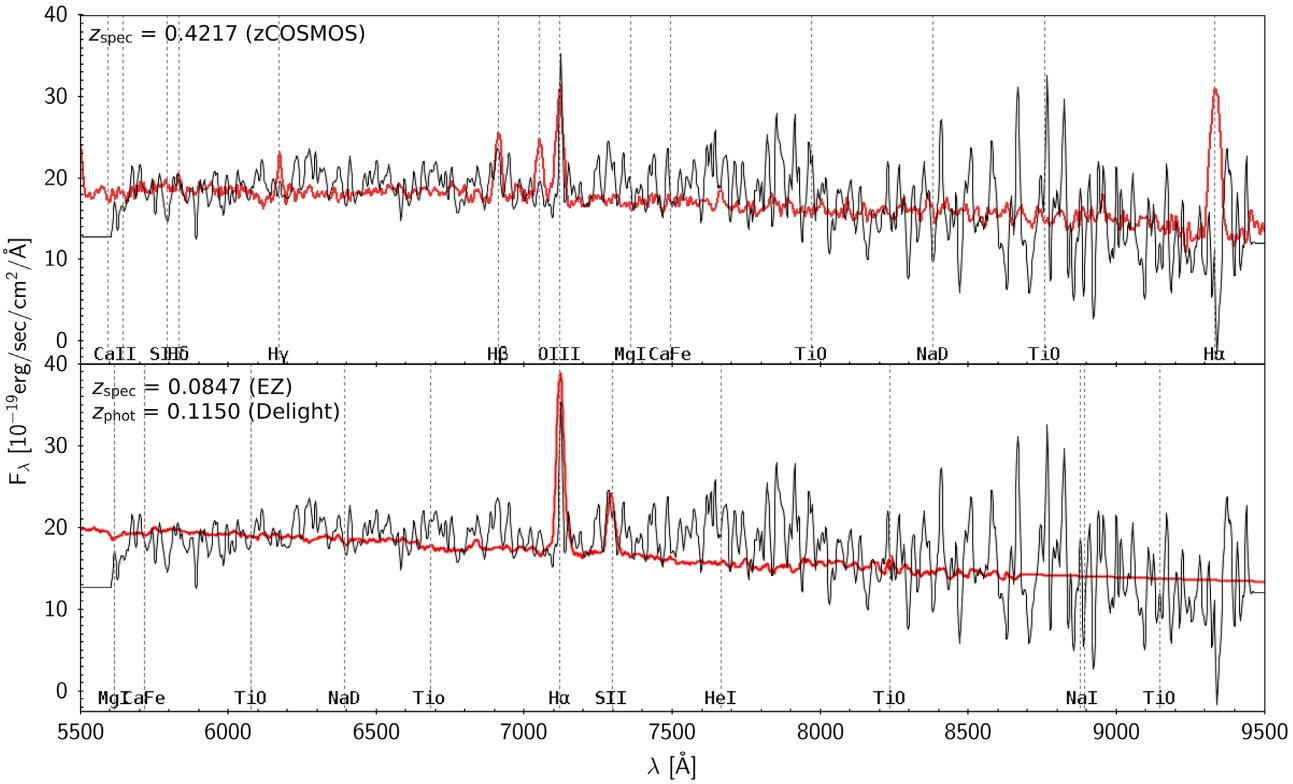

Figure 14. Spectral line fitting (red) for the original spectra (black) of the galaxy with zCOSMOS ID 804179. The spec-z given by zCOSMOS is 0.4217 (top), while the best fit using EZ (Garilli et al. 2010) gives a spec-z of 0.0847 (bottom), which is closer to the photo-z value of 0.1150 estimated by DELIGHT.

(iii) The *chi-squared value of the best-fitting Brown template* (χ^2_t), where we identified a trend that the better the fit, the more reliable the photo-z.

(iv) The *broad-band–narrow-band complementary metric* (ρ^2),

$$\rho^2 \equiv \int p_{BB}(z) p_{NB}(z)\, dz, \tag{17}$$

where $p_{BB}(z)$ and $p_{NB}(z)$ are the $p(z)$ produced by DELIGHT when trained with only broad bands and only narrow bands, respectively. By multiplying these two $p(z)$ and summing over the distribution at each step i, we can identify the consistency between the broad-band and narrow-band $p(z)$. A higher value of ρ^2 means a larger overlap, which indicates more reliable photo-zs.

Together with Θ and the DELIGHT photo-z error (δz), we yield a total of six metrics to experiment with. Using the results from the flux calibration method, we generate and test the individual performance for each of these metrics. For each metric, we measure the σ_{rms} and σ_{68} after systematically removing objects with the worst metric values, 10 per cent of the total sample size each time, until we reach a sample size of only 40 per cent.

We also repeat the exercise by using combined cuts on several metrics, testing all 57 combinations of the six metrics. We note that we do not combine the metrics by averaging or multiplying them, as it would have diluted the impact of the individual metrics. Instead, we rank the values for each metric individually (from best to worst), and remove objects rank by rank, starting with metric values lying in the worst rank. For example, for the combination of metrics $\Theta + \Delta_{DB}$, we first remove all objects that share the worst values of Θ and Δ_{DB}, then remove all objects sharing the second worst values of them, and so on, until we reach a required sample size percentile (90, 80, etc.), where we output the values of σ_{rms} and σ_{68}. We visualize the performance of these metric cuts at several percentiles for σ_{68} with respect to i_{auto} (cumulative) in Fig. 15.

We find that each performance metric cuts the sample differently: while metric cuts of σ_D and ρ^2 reduce the scatter (σ_{rms}) significantly, metric cuts of Θ and Δ_{DB} reduce the σ_{68} instead. The metric χ^2_t, however, does not seem to bring any significant improvement to the results. We have also plotted a cut in $\Delta z = \frac{|z_{phot} - z_{spec}|}{1+z_{spec}}$ (bottom left-hand panel in Fig. 15), which is the theoretical 'best metric', providing an upper limit to be compared with the performance of each of the metrics. Here we noticed that even with the theoretical best metric, a cut of slightly lesser than 70 per cent (blue line) on the sample is still necessary to fulfil the PAUS target of $\sigma_{68} < 0.0035(1 + z)$ (dotted line) for DELIGHT.

Therefore, we select the 60 per cent cut (navy line, retaining 60 per cent of galaxies) as a benchmark to assess the performance of these metrics, we do so by locating where this line cuts the dotted line (i.e. finding the maximum value of i_{auto} where the photo-zs achieves the PAUS target at 60 per cent cut). From Fig. 15, it is clear that cutting in all six metrics does not necessarily outperform the performance when cutting with only Θ, so we searched for the best combination of metrics for σ_{rms} and σ_{68} separately.

For σ_{rms}, the best combination of metrics is $\Delta_{DB} + \sigma_D + \rho^2$, and this combination achieves $\sigma_{rmS} < 0.0035(1 + z)$ at $i_{auto} < 19.27$ at 60 per cent cut, a significant improvement to the case when only Θ was used, where it did not cut the line at all. For σ_{68}, the best combination of metrics is $\Theta + \Delta_{DB}$ where it reached $\sigma_{68} < 0.0035(1 + z)$ at $i_{auto} < 21.25$ at 60 per cent cut, which is also a significant improvement as compared to Θ at $i_{auto} < 20.88$. Here we note that in fact using Δ_{DB} alone, the target can be reached at a higher limit

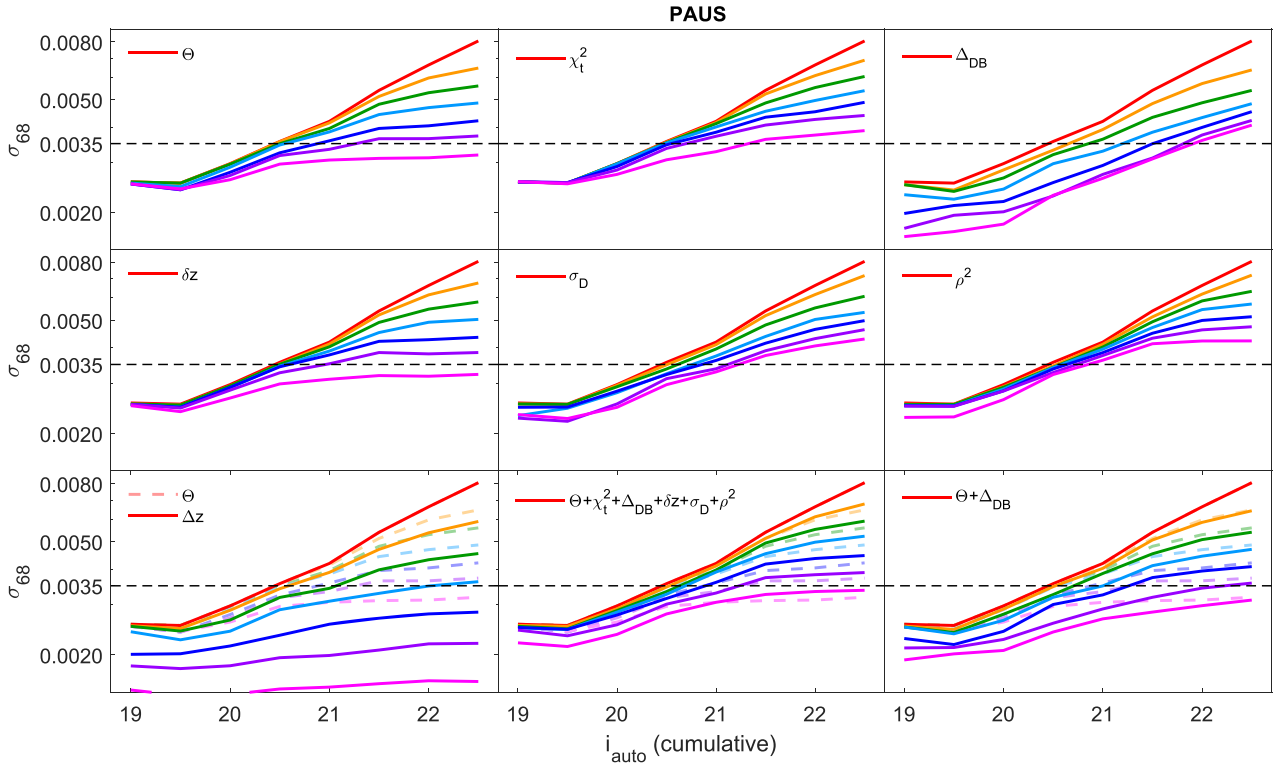

Figure 15. Plot of 68th percentile error (σ_{68}) versus i_{auto} (cumulative) when cut using the following metrics: the Bayesian odds (Θ), best-fitting Brown template χ_t^2 value, DELIGHT–BCNZ2 metric (Δ_{DB}), DELIGHT photo-z error (δz) and standard deviation (σ_D), and the broad-band–narrow-band complementary metric (ρ^2). The coloured lines follow the same percentile cuts as shown in Fig. 8, with the dotted-coloured lines in the background of the bottom panels depicting the results of Θ for easier comparison. The bottom left-hand panel shows the cut in Δz (defined in Section 5.1), the unsurpassable theoretical best used for reference. The bottom-middle panel shows the cuts when all the above metrics were combined, while the bottom right-hand panel shows the combination of metrics that yield the best results.

of $i_{auto} < 21.50$, which highlights the significance of a synergy between DELIGHT and BCNZ2 in selecting a high-quality photo-z sample.

Finally, we also show the performance of the metrics in terms of the completeness with respect to the photo-z (using DELIGHT's flux calibration method), visualized in Fig. 16. We find that metrics like σ_D and ρ^2 tend to selectively remove high photo-z objects, while Θ, χ_t^2, and Δ_{DB} tend to remove mid-ranged photo-z objects. In general, a cut using all six performance metrics at 60 per cent cut shows a balanced result in the completeness, keeping a sufficient number of high-redshift objects in the sample.

To summarize the performance of the individual metrics,

(i) χ_t^2 is the least-performing metric here; it does not bring significant positive impact to the results;

(ii) cuts in σ_D and ρ^2 help to improve the scatter, however, they tend to selectively remove higher photo-z objects from the sample;

(iii) Θ and δz show very similar results, however, Θ tends to keep more high photo-z objects in the sample; and

(iv) Δ_{DB} is the best-performing metric here, and we recommend the use of such a metric to remove outlier photo-zs from a sample.

7 CONCLUSION AND FUTURE WORK

In this work, we have optimized DELIGHT, a hybrid template machine learning algorithm such that it could be used to obtain photo-zs for PAUS, by utilizing its 40 narrow-band fluxes combined with six *uBVriz* COSMOS broad-band fluxes. We have shown three distinct methods to calibrate the broad-band and narrow-band fluxes, and found that all three methods yield comparable results, although the most stable and the one which produces the lowest value of σ_{68} is what we defined as the *flux calibration method*: a method where we calibrate the broad-band fluxes with respect to the narrow-band fluxes by finding the flux ratio of the filter combinations that overlap. This calibration method is entirely photometric, and it was able to produce photo-zs with a scatter reaching as low as $\sigma_{rms} = 0.0331(1 + z)$ and $\sigma_{68} = 0.0081(1 + z)$ for the full PAUS galaxy sample at $i_{auto} < 22.5$.

We have also compared the results of DELIGHT with a machine learning algorithm (ANNZ2) and a template-based algorithm (BPZ and BCNZ2). We find that ANNZ2 underperforms significantly, indicating that ANNZ2 in its basic form is not suitable for narrow-band surveys with large number of bands and small number of training objects.

Despite the photo-z performance of BPZ being within 9 per cent difference of that of DELIGHT, the latter still stood out in terms of the quality of the photo-z PDF $p(z)$ (16 per cent better in ρ_{CRPS}) and the effectiveness of its Bayesian odds (Θ) cut in retaining objects with higher quality photo-z without losing too many high-redshift objects. DELIGHT is also shown to produce competitive results as compared to BCNZ2 (5 per cent lower in σ_{68}), the default photo-z produced for the PAUS.

Further investigation on the common photo-z outliers of DELIGHT and BCNZ2 led to the conclusion that outlier narrow-band fluxes are the main cause for erroneous photo-zs, an insight that will inform

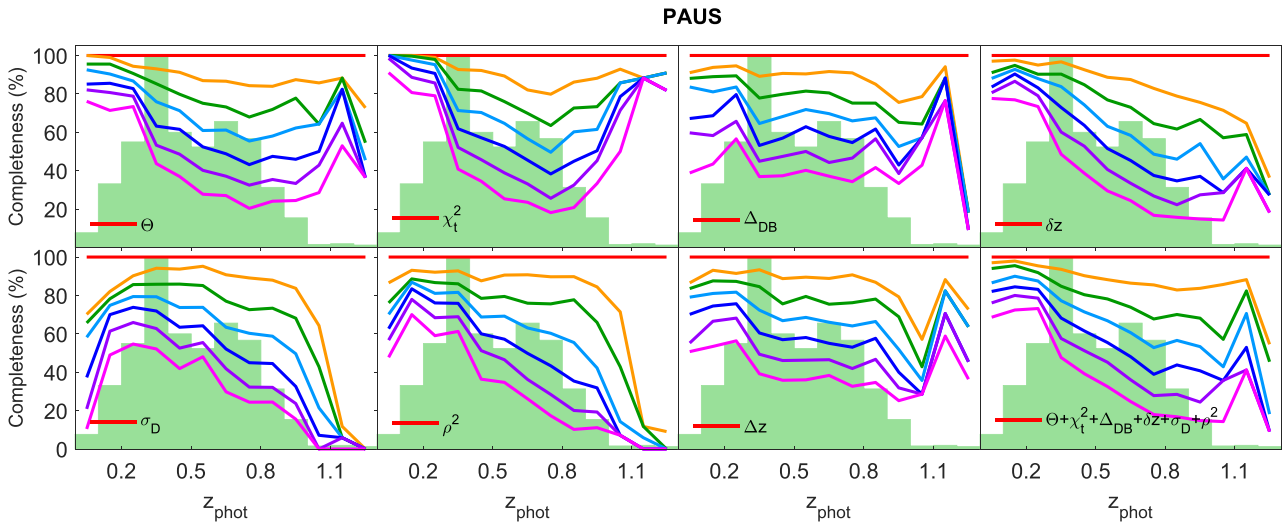

Figure 16. Plot of percentage of objects within each photo-z bin with respect to the cut in performance metric values listed in Fig. 15. The lines show the percentiles of the same colour scheme as in Fig. 8, while the histograms in the background show the relative number of objects in each photo-z bin. The bottom right-hand plot shows when the combination of all six metrics are used to cut the sample.

improvements in forthcoming PAUS data reductions. We have also inspected the spectra and identified catastrophic spec-zs, however, the effects are shown to be insignificant in this work. Motivated by the study of 30 outliers shared between DELIGHT and BCNZ2, we introduced several new metrics to help improve the identification of photo-z outliers and remove them from the sample to achieve better results. From the six metrics compared, our newly introduced DELIGHT–BCNZ2 metric (Δ_{DB}) is shown to significantly improve our photo-z quality, allowing it to reach the PAUS target of $\sigma_{68} < 0.0035(1 + z)$ at $i_{auto} < 21.5$ while retaining 60 per cent of the sample objects. These new metrics could be utilized to return more accurate uncertainties in redshift, which are vital in many cosmological studies.

This opens the door to future studies in finding synergies between different photo-z algorithms and between broad-band and narrow-band photometry. Together with the promising developments of deep learning approaches to deal with narrow-band data (Eriksen et al. 2020), these insights will pave the way towards unprecedentedly precise and accurate photometric redshifts for the full PAUS survey and beyond, like the Javalambre-Physics of the Accelerating Universe Astrophysical Survey (J-PAS; Benítez et al. 2014).

ACKNOWLEDGEMENTS

The authors wish to thank the referee for the helpful and constructive comments. JYHS would like to thank Boris Leistedt for fruitful discussions and the set-up of DELIGHT earlier in this work. JYHS also would like to thank Hwee San Lim and Tiem Leong Yoon for assisting in the set-up of equipment in Universiti Sains Malaysia where most of the computational work of this paper was completed. JYHS acknowledges financial support from the MyBrainSc Scholarship by the Ministry of Education, Malaysia, a studentship provided by Ofer Lahav, and the Short Term Research Grant by Universiti Sains Malaysia (304/PFIZIK/6315395). JYHS and BJ acknowledge support by the University College London Cosmoparticle Initiative. MS acknowledges funding from the National Science Centre, Poland (UMO-2016/23/N/ST9/02963) and the Spanish Ministry of Science and Innovation through the Juan de la Cierva Formación programme (FJC2018-038792-I). HHi acknowledges support by a Heisenberg grant of the Deutsche Forschungsgemeinschaft (Hi 1495/5-1) and an ERC Consolidator Grant (no. 770935). HHo acknowledges support from the Netherlands Organisation for Scientific Research (NWO) through grant 639.043.512. IEEC and IFAE are partially funded by the Institució Centres de Recerca de Catalunya (CERCA) and Beatriu de Pinós Programme of Generalitat de Catalunya. Work at Argonne National Laboratory is supported by UChicago Argonne LLC, Operator of Argonne National Laboratory (Argonne). Argonne, a U.S. Department of Energy Office of Science Laboratory, is operated under contract no. DE-AC02-06CH11357.

This project has received funding from the European Union's Horizon 2020 Framework Programme under the Marie Skłodowska-Curie Actions, through the following projects: Latin American Chinese European Galaxy (LACEGAL) Formation Network (no. 734374), the Enabling Weak Lensing Cosmology (EWC) Programme (no. 776247), and Barcelona Institute of Science and Technology (PROBIST) Post-doctoral Programme (no. 754510).

PAUS is partially supported by the Ministry of Economy and Competitiveness (MINECO, grants CSD2007-00060, AYA2015-71825, ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597 and MDM-2015-0509). Funding for PAUS has also been provided by Durham University (ERC StG DEGAS-259586), ETH Zurich, and Leiden University (ERC StG ADULT-279396).

The PAU data centre is hosted by the Port d'Informació Científica (PIC), maintained through a collaboration of CIEMAT and IFAE, with additional support from Universitat Autònoma de Barcelona and the European Research Development Fund (ERDF).

DATA AVAILABILITY

The data from PAUS (photometry and photo-zs) are currently not yet publicly available. The data from COSMOS were accessed from the ESO Catalogue Facility (https://www.eso.org/qi/), while the data from zCOSMOS (spectra and spec-zs) were accessed from the zCOSMOS data base (http://cesam.lam.fr/zCosmos/). The derived data generated in this research will be shared on reasonable request to the corresponding author.

REFERENCES

Aihara H. et al., 2018, PASJ, 70, S4
Alarcon A. et al., 2021, MNRAS, 501, 6103
Bellman R., 1957, J. Phys. Soc. Jpn., 12, 1049
Benítez N., 2000, ApJ, 536, 571
Benítez N. et al., 2014, preprint (arXiv:1403.5237)
Benjamin J. et al., 2013, MNRAS, 431, 1547
Bilicki M. et al., 2018, A&A, 616, A69
Bonfield D. G., Sun Y., Davey N., Jarvis M. J., Abdalla F. B., Banerji M., Adams R. G., 2010, MNRAS, 405, 987
Bonnett C. et al., 2016, Phys. Rev. D, 94, 042005
Boulade O. et al., 2003, in Iye M., Moorwood A. F. M., eds, Proc. SPIE Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-Based Telescopes. SPIE, Bellingham, p. 72
Brescia M., Cavuoti S., Amaro V., Riccio G., Angora G., Vellucci C., Longo G., 2018, in Kalinichenko L., Manolopoulos Y., Malkov O., Skvortsov N., Stupnikov S., Sukhomlin V., eds, Communications in Computer and Information Science, Vol. 822, Data Analytics and Management in Data Intensive Domains. Springer, Cham, Switzerland, p. 61
Brown M. J. I. et al., 2014, ApJS, 212, 18
Brun R., Rademakers F., 1997, Nucl. Instrum. Methods Phys. Res. Sec. A, 389, 81
Bruzual A. G., Charlot S., 2003, MNRAS, 344, 1000
Bundy K. et al., 2015, ApJS, 221, 15
Cavuoti S. et al., 2017, MNRAS, 466, 2039
Coleman G. D., Wu C.-C., Weedman D. W., 1980, ApJS, 43, 393
Crocce M. et al., 2016, MNRAS, 455, 4301
De Jong J. T. A., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., 2013, Exp. Astron., 35, 25
D'Isanto A., Cavuoti S., Gieseke F., Polsterer K. L., 2018, A&A, 616, A97
Duncan K. J. et al., 2018, MNRAS, 473, 2655
Duncan K. et al., 2019, ApJ, 876, 110
Eriksen M. et al., 2019, MNRAS, 484, 4200
Eriksen M. et al., 2020, MNRAS, 497, 4565
Garilli B., Fumana M., Franzetti P., Paioro L., Scodeggio M., Le Fèvre O., Paltani S., Scaramella R., 2010, PASP, 122, 827
Hoecker A. et al., 2007, preprint (arXiv:physics/0703039)
Ilbert O. et al., 2006, A&A, 457, 16
Ivezić Ž. et al., 2019, ApJ, 873, 111
Johnston H. et al., 2021, A&A, 646, A147
Joudaki S. et al., 2020, A&A, 638, L1
Jouvel S. et al., 2017, MNRAS, 469, 2771
Kinney A. L., Calzetti D., Bohlin R. C., McQuade K., Storchi-Bergmann T., Schmitt H. R., 1996, ApJ, 467, 38
Koekemoer A. M. et al., 2007, ApJS, 172, 196
Laigle C. et al., 2016, ApJS, 224, 24
Laigle C. et al., 2018, MNRAS, 474, 5437
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Le Fèvre O. et al., 2003, in Iye M., Moorwood A. F. M., eds, Proc. SPIE Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-Based Telescopes. SPIE, Bellingham, p. 1670
Leistedt B., Hogg D. W., 2017, ApJ, 838, 5
Lilly S. J. et al., 2007, ApJS, 172, 70
Lilly S. J. et al., 2009, ApJS, 184, 218
Martí P., Miquel R., Castander F. J., Gaztanaga E., Eriksen M., Sanchez C., 2014, MNRAS, 442, 92
Miyazaki S. et al., 2002, PASJ, 54, 833
Padilla C. et al., 2019, AJ, 157, 246
Polletta M. et al., 2007, ApJ, 663, 81
Polsterer K. L., D'Isanto A., Gieseke F., 2016, preprint (arXiv:1608.08016)
Raihan S. F., Schrabback T., Hildebrandt H., Applegate D., Mahler G., 2020, MNRAS, 497, 1404
Sadeh I., Abdalla F. B., Lahav O., 2016, PASP, 128, 104502
Salvato M., Ilbert O., Hoyle B., 2019, Nat. Astron., 3, 212
Schmidt S. J. et al., 2020, MNRAS, 499, 1587
Scodeggio M. et al., 2005, PASP, 117, 1284
Scoville N. et al., 2007, ApJS, 172, 1
Siudek M. et al., 2018, preprint (arXiv:1805.09905)
Soo J. Y. H. et al., 2018, MNRAS, 475, 3613
Spergel D. et al., 2013, preprint (arXiv:1305.5422)
Tanaka M. et al., 2018, PASJ, 70, S9
The Dark Energy Survey Collaboration, 2005, preprint (arXv:astro-ph/0510346)

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