ASM Sc. J., 19, 2024

https://doi.org/10.32802/asmscj.2023.1767

Improving Photometric Redshifts by Training Complementary Features of Galaxy Templates Using Machine Learning

Ishaq Y.K. Alshuaili, John Y.H. Soo*, M.Z. Mat Jafri and M.R. Omar

School of Physics, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia

This study aims to improve the photometric redshifts (photo-zs) of galaxies by combining the use of template-based and machine learning algorithms, notably the Bayesian Photometric Redshift (BPz) and Artificial Neural Networks for Redshifts 2 (ANNz2), so we can advantage-leverage the complementary aspects of both techniques and achieve improved photo-z predictions. In this work, we introduce a technique where the outputs of the template-based photo-z (the best-fit template type and photo-z) are added as inputs to ANNz2, and we see that there is an improvement in $[\sigma_{\rm RMS}, \sigma_{68}]$ giving values as low as $[0.0474, 0.0471]$, $[0.0368, 0.0253]$ and $[0.0213, 0.0168]$ in the SDSS Stripe-82, CMASS, and LOWZ samples, respectively. This study is considered an extension of our previous work to improve photo-z values, which enhances its use in fainter and deeper sky surveys, opening broader horizons to develop these methods and finding improved methods for measuring galaxy photo-zs.

Keywords: galaxies; distances and redshift methods; photometric methods; data analysis.

I. INTRODUCTION

Modern astronomy relies heavily on the measurement of galaxy redshifts to help us comprehend the nature and development of the Cosmos. Understanding a galaxy's creation, development, and grouping depends on knowing its velocity, distance, and other cosmic characteristics, most of which may be determined by its redshift. Hubble (1929) established what is now known as Hubble's law, which links a galaxy's redshift to its distance, now known as Hubble's law. Also, the Big Bang model of the universe was established in large part because of the observation of galaxy redshifts. The Big Bang model foresaw the cosmic microwave background (CMB) radiation, which was found in 1964 by Penzias and Wilson (Penzias & Wilson, 1965). The redshift of the CMB is consistent with the universe's expansion. They make it possible to examine phenomena as a function of time and distance, as well as to identify structure formations like galaxy clusters, measuring distance-dependent quantities such as luminosities and masses.

Redshifts are also essential for separating large-scale structures and galaxies along the line of sight. The characteristics of the supermassive black holes in the centres of galaxies have also been studied using measurements of galaxy redshifts. For instance, millions of quasars, which are very brilliant objects propelled by supermassive black holes, have been found by the Sloan Digital Sky Survey (SDSS, Adelman-McCarthy *et al.*, 2007). Artificial neural networks (ANNs) are used by machine learning algorithms like ANNz2 to discover intricate correlations between the observed photometric characteristics and the accompanying spectroscopic redshifts (spec-zs). These algorithms are highly suited for dealing with varied and complicated galaxy populations because they can capture detailed patterns and correlations in the data. The observed photometric data of galaxies with known spectroscopic redshifts are compared using template-based methods, like Bayesian Photometric Redshifts (BPz & Benitez, 2000), which draw on a library of spectral templates. The redshift distribution's probabilistic estimates are provided by BPz by fitting the observed data to

*Corresponding author's e-mail: johnsooyh@usm.my

these templates. This approach is especially helpful for populations of well-studied galaxies with a constrained variety of spectral properties.

Various studies have investigated adding morphological information, beyond the usual galaxy-integrated colours, as training input for the machine to improve photometric redshift (photo-z) estimations. For instance, Menou (2019) has developed a multilayer perceptron / convolutional neural network (MLP-convnet) they find that it can significantly improve the accuracy of photo-z estimation by incorporating morphological features of galaxies. The SDSS photometry and morphological data are used in the empirical methodology of Vince and Csabai (2006), and they discovered that the poor association between morphology and redshift results in very minor improvements in photo-z estimate accuracy. Soo *et al.* (2018) found that the impact of galaxy morphology on photo-z quality is more significant for bright SDSS samples than for general samples of galaxies with good 5-band *ugriz* photometry. However, we can note that many estimations are purposefully geared towards the target demographic, and the estimates perform best for populations where large spectroscopic samples are available for training and testing. Although difficult, improving photo-z quality is necessary to further our knowledge of cosmology. As we analyse deeper and wider data sets for more and more stringent tests of cosmological models, requirements on photo-z methods steeply increase.

Numerous attempts exist to estimate and explore the synergy between different photo-z methods. For instance, studies explore how the synergies between narrow-band photometric data and large imaging surveys can be utilised to enhance broadband photo-zs, while Gomes *et al.* (2018) found that incorporating near-infrared YJHK filters and angular size data in the training, validation, and testing of photo-z estimation significantly improved accuracy. they have used two different photo-z algorithms, BPz and ANNz2, to obtain photo-zs for galaxies. They compare the results from these two codes to the KiDS pipeline solution.

In this work, we present how we can improve photo-zs of galaxies via exploring the synergy of two methods: template and machine learning, by using the parameters t_b (the best fit template set in BPz) and z_b (the best fit template-based photo-z) from BPz that we can use to improve the performance of the ANNz2 algorithm when utilising the artificial neural network (ANN) approach to improve photo-zs.

II. PHOTOMETRIC DATA

A. Sloan Digital Sky Survey (SDSS)

The Sloan Digital Sky Survey (SDSS; York *et al.*, 2000), which has been in continuous operation since April 2000, provides a chance to create the biggest and most comprehensive cluster sample. The photometry covers $14\,000\ \mathrm{deg}^2$ and is provided in five wide bands ($ugriz$), along with the subsequent spectroscopic measurements. Millions of astronomical objects such as stars, galaxies, quasars, and other celestial phenomena, have been studied by the survey. The gathered information has been applied to a variety of scientific studies, including mapping galaxy distributions, identifying dark matter characteristics, and discovering new celestial. In this work, we will present different methods to improve the estimation of the redshift of galaxies. We use different data sets from SDSS through three different samples, namely the Stripe-82 (S82) Sample, the Low Redshift (LOWZ) Sample, and the CMASS Sample, each representing galaxy samples of a certain redshift regime and colour selection.

B. Stripe-82 (S82) Sample

Stripe-82 is a 2.5° wide stripe in the Southern Galactic Cap ($-50° < \alpha < 60°, 1.25° < \delta < 1.25°$) that spans all five SDSS bands and covers a total size of $275\ \mathrm{deg}^2$. Stripe-82 has been observed in multiple SDSS data releases, due to its location on the celestial sphere and its scientific significance for many astrophysical investigations. Numerous types of studies have made use of the data from Stripe-82, including investigations into the properties and distribution of galaxies, the universe's large-scale structure, star populations, study transient and variable phenomena.

To test our methodology on the S82 sample, we used 29 541 galaxies with the following cut: redshift range of $0.1 < z < 1.2$, `class=GALAXY`, `type=3` (extended object), magnitude cuts of $16.24 < u < 30.99$, $16.04 < g < 28.85$, $16 < r < 24.49$, $15.57 < i < 24.60$ and $15.17 < z < 24.98$ to remove outliers. We applied the cuts to the entire sample of galaxies, which was later divided into three groups (testing, validation, and training) of 9847 galaxies each.

C. Low Redshift (LOWZ) Sample

The Low Redshift (LOWZ, Tojeiro *et al.*, 2014) sample is a specific selection of galaxies with low redshift values from the SDSS. The LOWZ sample was designed to study galaxies' characteristics and development in the nearby universe. The LOWZ sample was introduced as part of the Baryon Oscillation Spectroscopic Survey (BOSS), a component of SDSS-III (Dawson *et al.*, 2013). The chosen galaxies are among the brightest and reddest in the population of low-redshift galaxies. To nearly triple the sample's number density, the LOWZ sample is intended to extend the SDSS-I/II Cut I in the Luminous Red Galaxy (LRG, Eisenstein et al. 2001) sample to fainter luminosities at $z \approx 0.4$.

In this work, the LOWZ sample focuses on galaxies with low redshifts, where a total of 45 600 galaxies (testing, training, and validation) with redshifts ranging from $0.1 < z < 0.5$ were selected for this study. We ensured that in the sample, `type=3`, `class=GALAXY` and the warning flag `zWarning=0`.

D. The CMASS Galaxy Sample

The BOSS program within SDSS selected the CMASS galaxy sample to study the universe's large-scale structure and investigate the aggregation properties of galaxies (Eisenstein *et al.*, 2011). Colour and magnitude cuts were used as the selection criteria for the CMASS sample, where they used similar selection cuts to those utilised by Cut-II of LRGs from SDSS-I/II but extended them both bluer and fainter to increase the number density of targets in the redshift range $0.4 < z < 0.7$ (Reid *et al.*, 2016).

In this work, we applied the same methodology as above, by selecting 55 140 galaxies, divided into three groups (testing, training, and validation) of 18 380 galaxies each. Galaxies from the CMASS sample with a redshift of $0.1 < z < 0.9$, `type=3`, `class=GALAXY` and `zWarning=0` were selected.

III. PHOTOMETRIC REDSHIFT ALGORITHMS

The photo-z technique described in the literature can be classified into two broad categories: the empirical training set method, and the fitting of spectral energy distributions (SED) by synthetic or empirical template spectra. The first approach is also known as the machine learning method, an empirical relationship between magnitudes and redshifts is derived using a subsample of objects (the training set) in which both the redshifts and photometry are available (Connolly *et al.*, 1995). A slightly modified version of this method was used by (Wang *et al.*, 1998) to derive redshifts in the Hubble Deep Field (HDF–N) by means of a linear function of colours. In the SED-fitting approach, a spectral library is used to compute the colours of various types of sources at any plausible redshift, and a matching technique is applied to obtain the "best-fitting" redshift. This technique has been used extensively in deep cosmological surveys. Although most methods proposed to improve the estimates of galaxy redshifts have shown that machine learning methods have a clear performance advantage, however, both template and machine learning methods have their advantages that we can benefit from.

In this work, we improved and enhanced the redshift of galaxies using the complementary features of both approaches by combining the advantages of ANNz2 with template-based data and redshift estimates from BPz, the best-fitting redshift is referred to as z_b (or `z_b` in the algorithm). Based on each object's observed photometric data and comparison to template spectra or empirical models, the BPz algorithm determines the probability distribution function (PDF) of redshift for each item (Benitez, 2000). Whereas the best-fit template model for a given galaxy is represented by the coefficient t_b (or `t_b` in the algorithm) in BPz, the template index attached to t_b indicates the template applied at the best-fit estimate based on the SED template set used. The parameters t_b and z_b from BPz (by leveraging the complementary aspects of both techniques) can be used to improve the performance of the ANNz2 algorithm to improve photo-zs. In the following sections, we will explain the algorithms ANNz2 and BPz in further detail.

A. ANNz2

ANNz2 is a machine learning algorithm that Sadeh *et al.* (2016) created to overcome the difficulties and restrictions associated with photo-z estimation. This algorithm aims to correctly extract the related uncertainties, create both single-value solutions and PDFs, and optimise the performance of the photo-z estimate.

ANNz2 has been used to estimate photo-zs for BOSS (Dawson *et al.*, 2013), the Dark Energy Survey (DES, Abbott *et al.*, 2005), and the Legacy Survey of Space and Time (LSST; Schmidt *et al.*, 2020). It has been demonstrated to be a very efficient technique for calculating photo-zs. In this work, we chose to use ANNz2 because it is capable of doing regression estimation of single-value photo-z solutions as well as PDFs. We optimise ANNz2 by employing an ANN architecture of $N:2N:2N:1$, where N is the number of inputs used for the photo-z determination, which may range from 5 to 7, depending on the number of input parameters used (*ugriz*, z_b or t_b).

B. BPz

The Bayesian approach of Benitez (2000) is implemented in the Bayesian Photometric Redshift (BPz) algorithm. The redshift-type likelihood $\mathcal{L}(C|z,T)$, where C is the colour of a galaxy, z a certain redshift and T a spectral type, is produced by comparing the observed galaxy magnitudes with the redshifted template library weighted by BPz using a prior probability $p(z,T|C)$. Due to colour/redshift degeneracies, $\mathcal{L}(C|z,T)$ is frequently multimodal. The addition of previous knowledge aids in the elimination of unrealistic solutions and compactification of $p(z,T)$, which increases photo-z accuracy and lowers the incidence of catastrophic outliers. BPz can compute redshifts based on galaxy spectra information, even at high redshifts, by accurately modelling their spectral evolution. This eliminates the requirement for expensive and imprecise spectroscopic source observations for training sets.

In this work, the template-based algorithm BPz settings are set as follows: the template list used was those from Brown *et al.* (2014), the prior used was 'hdfn_gen', the redshift resolution was set to 0.002, and our minimum magnitude uncertainty is 0.001. As in our previous paper (Alshuaili *et al.*, 2022), we chose the Brown templates as it gives better results when compared to other templates, such as the CWW templates.

IV. IMPROVEMENT IN PHOTOMETRIC REDSHIFTS

Each of the samples described above (the Stripe-82, LOWZ and CMASS samples) was tested with BPz independently, and then we extracted t_b and z_b from BPz and included them into the machine learning algorithms (ANNz2) to enhance its performance. As mentioned in Section IIIA, ANNs are used in ANNz2 to calculate photo-zs. The ANN is trained using a training set of objects with known spectroscopic redshifts, and it then learns the intricate relationship between photometric magnitudes and the redshift.

For each sample running on ANNz2, we made four tests by varying the number of inputs: the first one is only $ugriz$, the second is $ugriz + z_b$, third is $ugriz + t_b$ and the last is $ugriz + t_b + z_b$. The goal of this step is to draw conclusions and study the extent of improvement in the performance metrics for each of these processes. The performance metrics used in this work are the root-mean-square error (σ_{RMS}), the 68th percentile error (σ_{68}) and the outlier fraction rate (η_{out}), which were all defined and used in Soo *et al.* (2018).

A. Photo-zs on the Stripe-82 Sample

The inclusion of t_b and z_b samples from BPz enhances the training and validation process of the ANNz2 algorithm, resulting in more accurate and reliable redshift estimates in the Stripe-82 sample, this can be seen from the results in **Table 1** and **Figure 1**, which show an improvement in the value of σ_{68} and σ_{RMS}. This is especially true when adding z_b to ANNz2, as the value of $\sigma_{\mathrm{RMS}} = 0.0475$ is lower as compared to the default $ugriz$ value in ANNz2 (0.0477) and BPz (0.0720).

Table 1. Performance of photo-z for the Stripe-82 Sample, as shown through the root-mean-square error (σ_{RMS}), 68th percentile error (σ_{68}), and outlier fraction (η_{out}) when including z_b and t_b from BPz as inputs in ANNz2. The lowest values are shown in green.

Training Parameters	σ_{RMS}	σ_{68}	η_{out} (%)
u, g, r, i, z, z_b	0.0474	0.0204	1.76
u, g, r, i, z, t_b	0.0486	0.0225	1.92
u, g, r, i, z, z_b, t_b	0.0479	0.0216	1.63
u, g, r, i, z (ANNz2)	0.0477	0.0210	1.75
u, g, r, i, z (BPz)	0.0720	0.0364	4.64

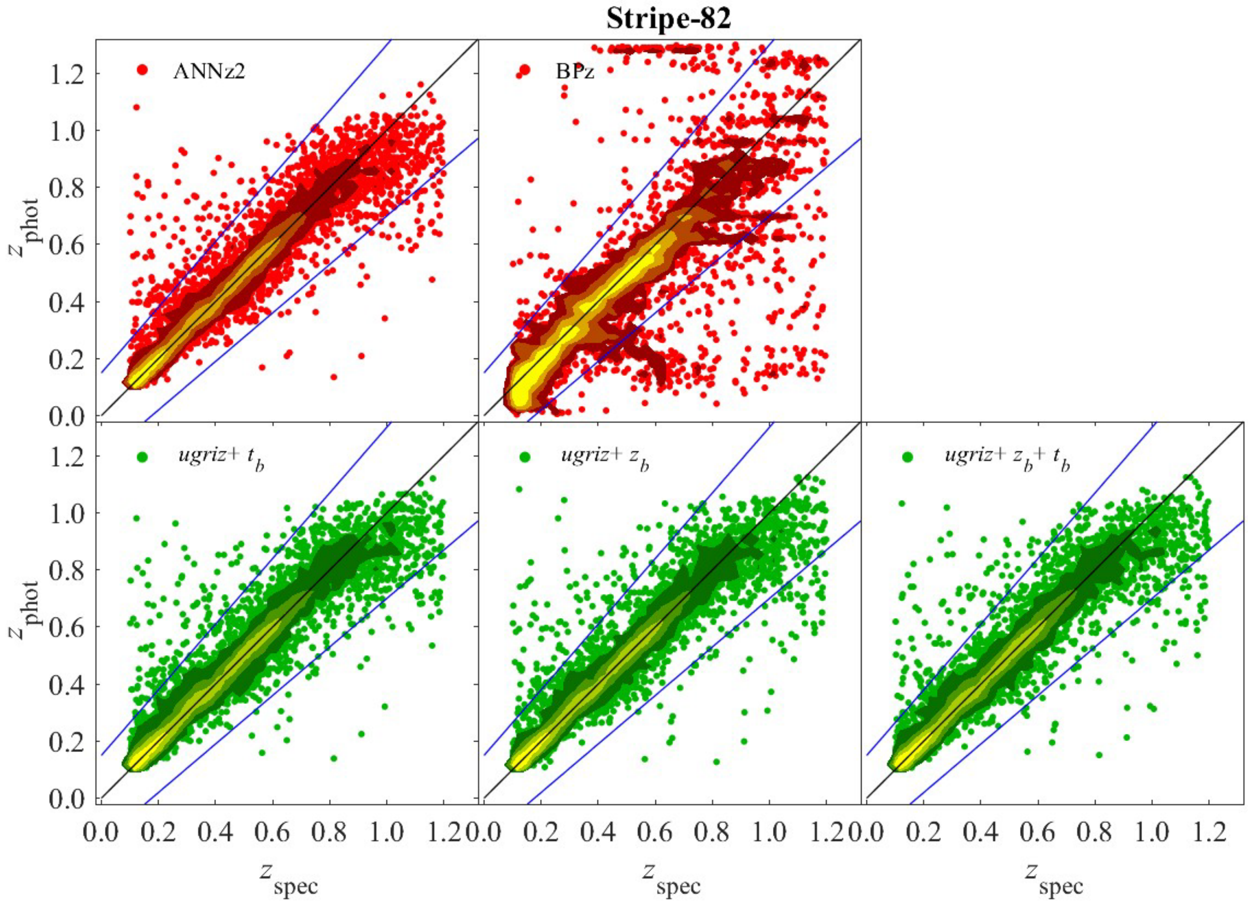

Figure 1. Photo-z vs. spec-z, comparing the performance of ANNz2 (top left), BPz (top right) with the methods we used to improve the photo-zs in the Stripe-82 Sample, i.e. adding t_b (bottom left), z_b (bottom middle) and both z_b and t_b (bottom right) as input for ANNz2. The blue lines define the limits for outliers, defined by η_{out} (see Soo *et al.*, 2018).

The percentage of improvement may be low overall (0.4%), however, we note that the density of points closer to the diagonal lines in **Figure 1** has increased particularly in the higher redshift region, for example, the $ugriz + z_b$ run increases the 1σ contour line redshift upper limit from 0.7 to 0.9. This is helpful as it allows more higher redshift galaxies to be retained in the situation where an error cut is needed to filter out redshifts with higher uncertainties, this parallels the effects of adding galaxy morphology into a similar sample in Soo *et al.* (2018). We note also that the improvement in σ_{RMS} here is close to the improvement we found in our previous work (Alshuaili *et al.*, 2022) on the same sample, which achieves a value of $\sigma_{\mathrm{RMS}} = 0.0471$.

B. Photo-zs on the LOWZ and CMASS Samples

The use of LOWZ and CMASS samples in this work are meant to complement each other in the redshift range, since LOWZ has $0.1 < z < 0.5$, while CMASS has $0.4 < z < 0.7$, allowing us to study the effects of our methodology on both a low and high redshift range. For LOWZ, we observe in **Table 2** that the result is slightly better when using all 7 inputs ($ugriz + z_b + t_b$) as compared to when using only $ugriz + z_b$, while as visualised in **Figure 2**, the performance improvement is very similar between the two methods. We also compared these results with those of Meshcheryakov *et al.* (2015), Brescia *et al.* (2014) and Soo *et al.* (2018) and we found that our value of σ_{RMS} value is shown to be better.

Table 2. Performance of photo-z for the LOWZ Sample, as shown through σ_{RMS}, σ_{68} and η_{out} when including z_b and t_b from BPz as inputs in ANNz2. The lowest values are shown in green, and the results of ANNz2, BPz, and three previous studies are shown as references.

Training Parameters	σ_{RMS}	σ_{68}	η_{out} **(%)**
u, g, r, i, z, z_b	0.0217	0.0178	0.025
u, g, r, i, z, t_b	0.0224	0.0182	0.037
u, g, r, i, z, z_b, t_b	0.0216	0.0178	0.025
u, g, r, i, z (ANNz2)	0.0219	0.0177	0.037
u, g, r, i, z (BPz)	0.0405	0.0249	0.820
Meshcheryakov *et al.* (2015)	0.0252	–	–
Brescia *et al.* (2014)	0.0280	–	–
Soo *et al.* (2018)	0.0228	0.0177	0.100

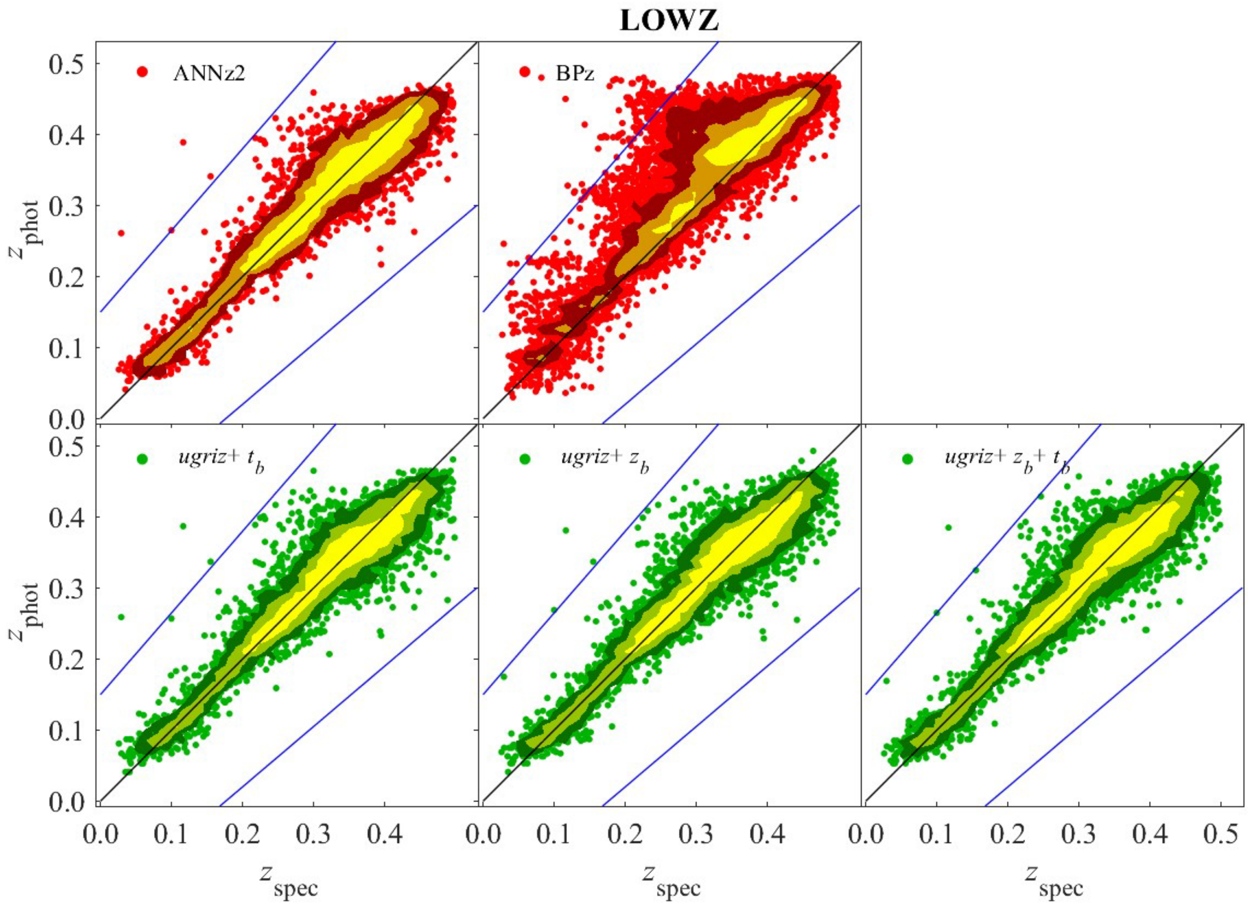

Figure 2. Photo-z vs. spec-z, comparing the performance of ANNz2 (top left), BPz (top right) with the methods we used to improve the photo-zs in the LOWZ Sample (bottom, similar arrangement as in Figure 1).

On the other hand, for the CMASS sample, ANNz2 is still consistently better than BPz (**Table 3**), while an improvement in σ_{RMS} is observed when using $ugriz + z_b$ as inputs to ANNz2, giving $\sigma_{\mathrm{RMS}} = 0.0368$ compared to 0.0374 when training with only the five $ugriz$ magnitudes (visualised in **Figure 3**). It also appears to be an improved result when compared to the results of previous studies (Soo *et al.*, 2018), considering the change in the method used in these two approaches. Here we also note that the improvement shown in the CMASS sample is more than that of the LOWZ sample, which is in line with our previous deduction that z_b brings more improvement to photo-zs of galaxies in the higher redshift regime.

Table 3. Performance of photo-z for the CMASS Sample, as shown through σ_{RMS}, σ_{68} and η_{out} when including z_b and t_b from BPz as inputs in ANNz2. The lowest values are shown in green, and the results of ANNz2 and BPz and Soo *et al.* (2018) are shown as reference.

Training Parameters	$\boldsymbol{\sigma_{\mathrm{RMS}}}$	$\boldsymbol{\sigma_{68}}$	$\boldsymbol{\eta_{\mathrm{out}}}$ **(%)**
u, g, r, i, z, z_b	0.0368	0.0253	0.838
u, g, r, i, z, t_b	0.0375	0.0257	0.903
u, g, r, i, z, z_b, t_b	0.0372	0.0255	0.838
u, g, r, i, z (ANNz2)	0.0374	0.0256	0.854
u, g, r, i, z (BPz)	0.0465	0.0284	1.420

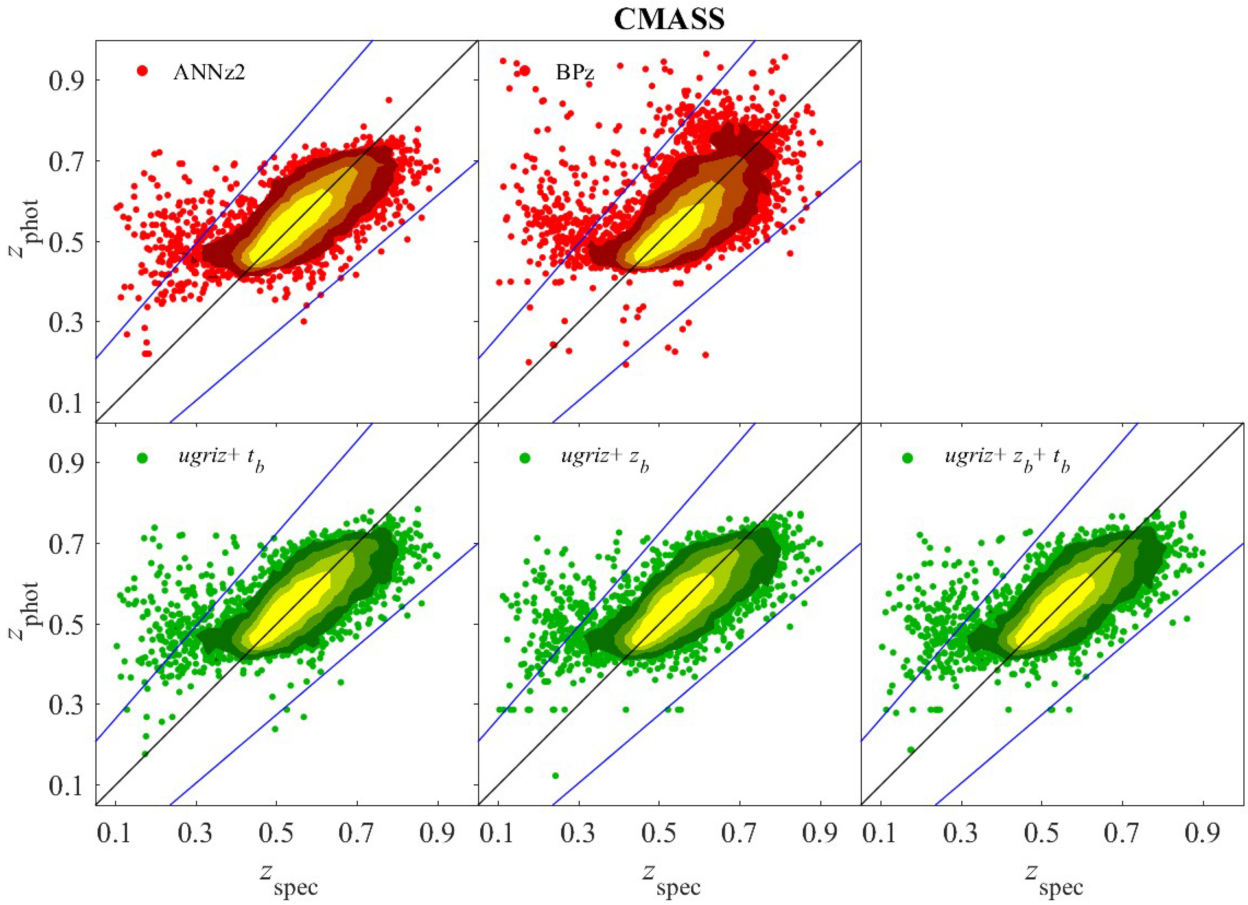

Figure 3. Photo-z vs. spec-z comparing the performance of ANNz2 (top left), BPz (top right) with the methods we used to improve the photo-zs in the CMASS Sample (bottom, similar arrangement as in Figure 1).

The analysis of the four samples in this work also demonstrates that the runs using $ugriz + t_b$ do not seem to provide any noticeable improvement, but instead mainly added noise to the machine learning algorithm. The template type t_b output from BPz is a number running from 1 to 129, representing each of the 129 Brown *et al.* (2014) templates arranged based on their Hubble types (ellipticals, spirals and others). It was initially hoped that the machine learning algorithm would pick up the correlation between the galaxy types to their redshifts through t_b to improve the photo-z estimate, however, it is shown to be not the case. In the future, we attempt to improve the representation of this correlation, e.g. by using k-means clustering, grouping of galaxy types, etc, to see if this correlation plays a role in improving photo-zs at all.

V. CONCLUSION AND FUTURE WORK

In the context of improving photo-z estimates using the ANNz2, it is found that the best-fit template photometric redshift (z_b) from BPz can be utilised as an additional input feature to enhance the performance of the photo-zs produced. The best-fit template type (t_b) from BPz could potentially provide valuable information about the spectral characteristics of the galaxies, however, including this information as a feature in the ANNz2 algorithm does not improve the photo-z estimation. This work demonstrates the capability of adding z_b is shown to improve the photo-z accuracy for various samples of galaxies (Stripe-82, LOWZ and CMASS), improving the root-mean-square error (σ_{RMS}) values by percentages of 2.0%, 1.4%, 1.6% and 0.7%, respectively, for these samples. We believe there is still room for improvement in this research finding, and we intend to explore and combine other related methodologies to synergise different photo-z algorithms. In this study, we concentrated on a particular set of complementary features, but we can potentially explore more feature spaces.

We believe there could be additional characteristics or data sources that might offer helpful details for estimating photo-z. Other feature spaces like multi-wavelength photometry, spectroscopic measurements, galaxy morphological data, or even auxiliary data from other surveys, may be explored and included in future studies. We may significantly advance the accuracy and precision of photo-zs by following these approaches for future research, which will enable us to learn new things about the universe and its development.

VI. ACKNOWLEDGEMENT

JYHS acknowledges financial support via the Fundamental Research Grant Scheme (FRGS) by the Malaysian Ministry of Higher Education with code FRGS/1/2023/STG07/USM/02/14.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

VII. REFERENCES

Abbott, T, Aldering, G, Annis, J, Barlow, M *et al.* 2005, 'The Dark Energy Survey'. arXiv: astro-ph/0510346.

Adelman-McCarthy, JK, Agüeros, MA, Allam, SS *et al.* 2007, 'The fifth data release of the Sloan Digital Sky Survey', *Astrophysical Journal Supplement Series,* vol. 172, no. 1, pp. 634-644.

Alshuaili, IYK., Soo, JYH, Mat Jafri, MZ & Rafid, Y 2022, 'Improving Photometric Redshifts by Merging Probability Density Functions from Template-Based and Machine Learning Algorithms', *Astronomy Letters*, vol. 48, pp. 665-675.

Benitez, N 2000, 'Bayesian Photometric Redshift Estimation', *The Astrophysical Journal*, vol. 536, no. 2, pp. 571–583.

Brown, MJI, Moustakas, J, Smith, JDT *et al.* 2014, 'An atlas of galaxy spectral energy distributions from the ultraviolet to the mid-infrared', *Astrophysical Journal Supplement Series*, vol. 212, no. 2, pp. 10-30.

Brescia, M, Cavuoti, S, Longo, G & De Stefano 2014, 'A catalogue of photometric redshifts for the SDSS-DR9 galaxies', *Astronomy and Astrophysics*, vol. 568, p. A126.

Connolly, AJ, Csabai, A, Szalay, AS *et al.* 1995, 'Slicing through multicolour space: galaxy redshifts from broadband photometry', *The Astronomical Journal*, vol. 110, pp. 2655-2664.

Dawson, KS, Schlegel, DJ, Ahn, CP *et al.* 2013, 'The baryon oscillation spectroscopic survey of SDSS-III', *Astronomical Journal*, vol. 145, pp. 131-132.

Eisenstein, DJ, Weinberg, DH & Agol, E 2011, 'SDSS-III: Massive spectroscopic surveys of the distant Universe, the MILKY WAY, and extra-solar planetary systems', *Astronomical Journal*, vol. 142, pp. 123-145.

Gomes, Z, Jarvis, MJ, Almosallam, IA & Roberts, SJ 2018, 'Improving photometric redshift estimation using GPZ: Size information, post-processing, and improved photometry', *Monthly Notices of the Royal Astronomical Society*, vol. 475, pp. 331-342.

Hubble, EP 1929, 'A relation between distance and radial velocity among extra-galactic nebulae', *National Academy of Science*, vol. 15, pp. 42, 45, 55.

Meshcheryakov, AV, Glazkova, VV, Gerasimov, SV *et al.* 2015, 'High-Accuracy Redshift Measurements for Galaxy Clusters at $z < 0.45$ Based On SDSS-III Photometry', *Astronomy Letters*, vol. 41, pp. 307-316.

Menou, K 2019, 'Morpho-photometric redshifts', *Monthly Notices of the Royal Astronomical Society*, vol. 489, pp. 4802-4808.

Penzias, AA & Wilson, RW 1965, 'A Measurement of Excess Antenna Temperature at 4080 Mc/s', *Astrophysical Journal*, vol. 142, pp. 419-421.

Reid, B, Ho, S, Padmanabhan, N, Percival, WJ, Tinker, J *et al.* 2016, 'SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: Galaxy target selection and large-scale structure catalogues', *Monthly Notices of the Royal Astronomical Society*, vol. 455, pp. 1553-1573.

Sadeh, I, Abdalla, FB & Lahav, O 2016, 'ANNZ2: Photometric Redshift and Probability Distribution Function Estimation Using Machine Learning', *Publications of the Astronomical Society of the Pacific*, vol. 128, pp. 104502-104520.

Schmidt, SJ, Malz, AI, Soo, JYH *et al.* 2020, 'Evaluation of probabilistic photometric redshift estimation approaches for the Rubin Observatory Legacy Survey of Space and Time (LSST)', *Monthly Notices of the Royal Astronomical Society*, vol. 499, pp. 1587-1606.

Soo, JYH, Moraes, B, Joachimi, B *et al.* 2018, 'Morpho-z: Improving photometric redshifts with galaxy morphology', Monthly Notices of the Royal Astronomical Society, vol. 475, pp. 3613-3632.

Tojeiro, A, Ross, AJ, Burden, A, Samushia, L *et al.* 2014, 'The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Galaxy clustering measurements in the low-redshift sample of Data Release 11', *Monthly Notices of the Royal Astronomical Society*, vol. 440, pp. 2222-2235.

Vince, O & Csabai, I 2006, 'Toward more precise photometric redshift estimation', *International Astronomical Union Symposium*, vol. 241, pp. 573-574.

Wang, Y, Bahcall, N & Turner, EL 1998, 'A method for deriving redshifts in the Hubble Deep Field using a linear function of colours', *The Astronomical Journal*, vol. 116, no. 5, pp. 2081-2087.

York, DG, Adelman, J, Anderson, JE *et al.* 2000, 'The Sloan Digital Sky Survey: technical summary', *The Astronomical Journal*, vol. 120, pp. 1579-1587.