

Photometric redshift analysis in the Dark Energy Survey Science Verification data

C. Sánchez,^{1★} M. Carrasco Kind,^{2,3} H. Lin,⁴ R. Miquel,^{1,5} F. B. Abdalla,⁶
 A. Amara,⁷ M. Banerji,⁶ C. Bonnett,¹ R. Brunner,² D. Capozzi,⁸ A. Carnero,^{9,10}
 F. J. Castander,¹¹ L. A. N. da Costa,^{9,10} C. Cunha,¹² A. Fausti,¹⁰ D. Gerdes,¹³
 N. Greisel,^{14,15} J. Gschwend,^{9,10} W. Hartley,^{7,16} S. Jouvel,⁶ O. Lahav,⁶ M. Lima,^{10,17}
 M. A. G. Maia,^{9,10} P. Martí,¹ R. L. C. Ogando,^{9,10} F. Ostrovski,^{9,10} P. Pellegrini,⁹
 M. M. Rau,^{14,15} I. Sadeh,⁶ S. Seitz,^{14,15} I. Sevilla-Noarbe,¹⁸ A. Sypniewski,¹³
 J. de Vicente,¹⁸ T. Abbot,¹⁹ S. S. Allam,^{4,20} D. Atlee,²¹ G. Bernstein,²²
 J. P. Bernstein,²³ E. Buckley-Geer,⁴ D. Burke,^{12,24} M. J. Childress,^{25,26} T. Davis,^{26,27}
 D. L. DePoy,^{28,29} A. Dey,^{21,30} S. Desai,^{31,32} H. T. Diehl,⁴ P. Doel,⁶ J. Estrada,⁴
 A. Evrard,^{13,33,34} E. Fernández,¹ D. Finley,⁴ B. Flaugher,⁴ J. Frieman,⁴
 E. Gaztanaga,¹¹ K. Glazebrook,³⁵ K. Honscheid,³⁶ A. Kim,³⁷ K. Kuehn,³⁸
 N. Kuropatkin,⁴ C. Lidman,³⁸ M. Makler,³⁹ J. L. Marshall,^{28,29} R. C. Nichol,⁸
 A. Roodman,^{12,24} E. Sánchez,¹⁸ B. X. Santiago,^{10,40} M. Sako,²² R. Scalzo,²⁵
 R. C. Smith,¹⁹ M. E. C. Swanson,³ G. Tarle,¹³ D. Thomas,^{8,41} D. L. Tucker,⁴
 S. A. Uddin,^{26,35} F. Valdés,²¹ A. Walker,¹⁹ F. Yuan^{25,26} and J. Zuntz⁴²

Affiliations are listed at the end of the paper

Accepted 2014 September 4. Received 2014 September 3; in original form 2014 June 13

ABSTRACT

We present results from a study of the photometric redshift performance of the Dark Energy Survey (DES), using the early data from a Science Verification period of observations in late 2012 and early 2013 that provided science-quality images for almost 200 sq. deg. at the nominal depth of the survey. We assess the photometric redshift (photo- z) performance using about 15 000 galaxies with spectroscopic redshifts available from other surveys. These galaxies are used, in different configurations, as a calibration sample, and photo- z 's are obtained and studied using most of the existing photo- z codes. A weighting method in a multidimensional colour–magnitude space is applied to the spectroscopic sample in order to evaluate the photo- z performance with sets that mimic the full DES photometric sample, which is on average significantly deeper than the calibration sample due to the limited depth of spectroscopic surveys. Empirical photo- z methods using, for instance, artificial neural networks or random forests, yield the best performance in the tests, achieving core photo- z resolutions $\sigma_{68} \sim 0.08$. Moreover, the results from most of the codes, including template-fitting methods, comfortably meet the DES requirements on photo- z performance, therefore, providing an excellent precedent for future DES data sets.

Key words: Astronomical data bases: surveys – galaxies: distances and redshifts – galaxies: statistics – large-scale structure of Universe.

*E-mail: csanchez@ifae.es

1 INTRODUCTION

Large galaxy surveys provide detailed information on the large-scale structure of the Universe, which, in turn, helps understand its geometry, composition, evolution and fate. On one hand, spectroscopic surveys like 2dF (Colless et al. 2001), VIMOS VLT Deep Survey (VVDS; Le Fèvre et al. 2005), WiggleZ (Drinkwater et al. 2010) or BOSS (Dawson et al. 2013) provide a 3D picture of the galaxy distribution, but they are costly in time and resources, and may suffer from limited depth, incompleteness and selection effects. On the other hand, photometric surveys such as Sloan Digital Sky Survey (SDSS; York et al. 2000), PanSTARRS (Kaiser, Tonry & Luppino 2000), KiDS (de Jong et al. 2013), HSC¹ or LSST (Tyson et al. 2003) are more efficient and usually deeper, more complete and nearly unbiased, but do not provide a complete 3D view of the Universe, due to their limited resolution in the galaxy positions along the line of sight, which are computed by measuring the photometric redshift (photo-*z*) of each galaxy from the fluxes measured through a set of broad-band filters. Even with their limited resolution along the line of sight, photometric surveys, because of their larger volume, are extremely useful for cosmology and, furthermore, uniquely provide some of the most stringent probes of dark energy, such as weak lensing.

There are two main approaches for measuring photometric redshifts: template-fitting methods (e.g. HYPERZ; Bolzonella, Miralles & Pell 2000, BPZ; Benítez 2000; Coe et al. 2006, LEPHARE; Arnouts et al. 2002; Ilbert et al. 2006, EAZY; Brammer, van Dokkum & Coppi 2008), in which the measured broad-band galaxy spectral energy distribution (SED) obtained from the fluxes is compared to a set of redshifted galaxy templates until a best match is found, thereby determining both the galaxy spectral type and its redshift; and training methods (e.g. ANNZ; Collister & Lahav 2004, ARBORZ; Gerdes et al. 2010, TPZ; Carrasco Kind & Brunner 2013), in which a set of galaxies with known spectroscopic redshifts is used to train a machine-learning algorithm (e.g. an artificial neural network, ANN), which is then applied over the galaxy set of interest. Each technique has its own advantages and disadvantages, as we will discuss in this paper, and a combination of them can fully exploit this fact (Carrasco Kind & Brunner 2014).

In order for photo-*z*'s to be useful for cosmological studies, it is necessary to calibrate them, by understanding the statistical properties of the distribution of the differences between the true galaxy redshifts and their photo-*z* estimates: its mean value (for the bias), its width (for the resolution), and its tails (for the fraction of outliers, with grossly misestimated photo-*z*'s). To accomplish this, a sample of galaxies with spectroscopic redshifts is required, ideally with a galaxy population that reproduces the population in the photometric survey.

The Dark Energy Survey (DES; Flaugher 2005) is one such photometric redshift survey, and will cover about one-eighth of the sky (5000 sq. deg.) to an unprecedented depth ($i_{AB} < 24$), imaging about 300 million galaxies in five broad-band filters (*grizY*) up to redshift $z = 1.4$. The DES camera (DECam; Diehl et al. 2012; Flaugher et al. 2012) was installed and commissioned in the second semester of 2012, and a Science Verification (SV) period of observations followed, lasting from 2012 November to 2013 February. The survey officially started in late 2013 August.

The SV observations provided science-quality data for almost 200 sq. deg. at close to the nominal depth of the survey. The SV

footprint was chosen to contain areas already covered by several deep spectroscopic galaxy surveys, including VVDS (Le Fèvre et al. 2005), ACES (Cooper et al. 2012), and zCOSMOS (Lilly et al. 2007), which together provide a suitable calibration sample for the DES photometric redshifts. This paper presents a study of the photo-*z* precision achieved by DES during the SV period, by taking advantage of the available spectroscopic data in its footprint, and by using a large number of photo-*z* algorithms of different nature.

It has been pointed out (Cunha et al. 2012a) that cosmic variance in the spectroscopic samples used for photo-*z* calibration may bias the results of an analysis such as the one we present here, which uses spectra in four relatively small (1 sq. deg. each) patches of sky. A robust photo-*z* calibration requires galaxy spectra distributed all over the photometric survey's footprint, calling for as many as O(50–100) patches (Cunha et al. 2012a). While the plan for the ultimate photo-*z* calibration of the whole DES data will need such a spectroscopic calibration sample, and steps are being taken towards the acquisition of the relevant data, the currently available spectroscopic data set is good enough for a first analysis of the photo-*z* precision that can be achieved with the early DES data. Analogously, the ultimate DES photo-*z* calibration will have to worry about the effects of the possible incompleteness of the spectroscopic calibration samples, effects that we can safely ignore here, given the scope of this first study.

Many studies have been performed in the past comparing in detail several photo-*z* codes over the same, real or simulated, data set (Hogg et al. 1998; Abdalla et al. 2011; Hildebrandt, Wolf & Benítez 2008; Hildebrandt et al. 2010; Dahlen et al. 2013). Particularly comprehensive is the work by Hildebrandt et al. (2010), which compares the performance of 19 photo-*z* codes both over simulated and real (including *Hubble Space Telescope*, *HST*) observations taken in 18 optical and near-infrared bands. Similarly, Dahlen et al. (2013) compare 11 codes over real data in 14 bands, including also some *HST* data. On the other hand, Abdalla et al. (2011) analysed the performance of six photo-*z* algorithms on the MegaZ luminous red galaxy (LRG) sample extracted from the SDSS Data Release 7 five-band photometry, with a magnitude limit around $i_{AB} = 20$. This paper differs from these previous studies in that, on the one hand, it uses solely DECam five-band photometry (*grizY*), and on the other, it studies all kinds of galaxies up to the DES nominal limiting magnitude $i_{AB} = 24$. Furthermore, in the present study, rather than trying to carry out a thorough comparison of all the photo-*z* codes available in the literature, we concentrate on assessing the performance of the early DES data with respect to the photometric redshift determination, and, in order to do so, we try the codes in which members of the DES collaboration have a certain degree of expertise, without attempting to be complete or even necessarily fair in the comparison. Beyond providing a snapshot of the quality of the DES-SV data regarding photo-*z* estimation and accuracy, a secondary goal of this work is to tune these photo-*z* codes to the particular characteristics of the DES data: filter set, depth, etc., in preparation for the upcoming larger data sets.

Since even the deep spectroscopic samples mentioned above fail to reproduce exactly the depth and colours of the DES-SV photometric galaxy sample, a multidimensional weighting technique (Lima et al. 2008; Cunha et al. 2009) was used in order to bring the spectroscopic and photometric samples into better agreement. Matching the galaxies in the spectroscopic samples with those in the DES-SV photometric sample and comparing their spectroscopic redshifts with the DES photo-*z*'s, we will show that, even at this early stage, the DES-SV data fulfil the set of photo-*z* requirements

¹<http://www.naoj.org/Projects/HSC/index.html>

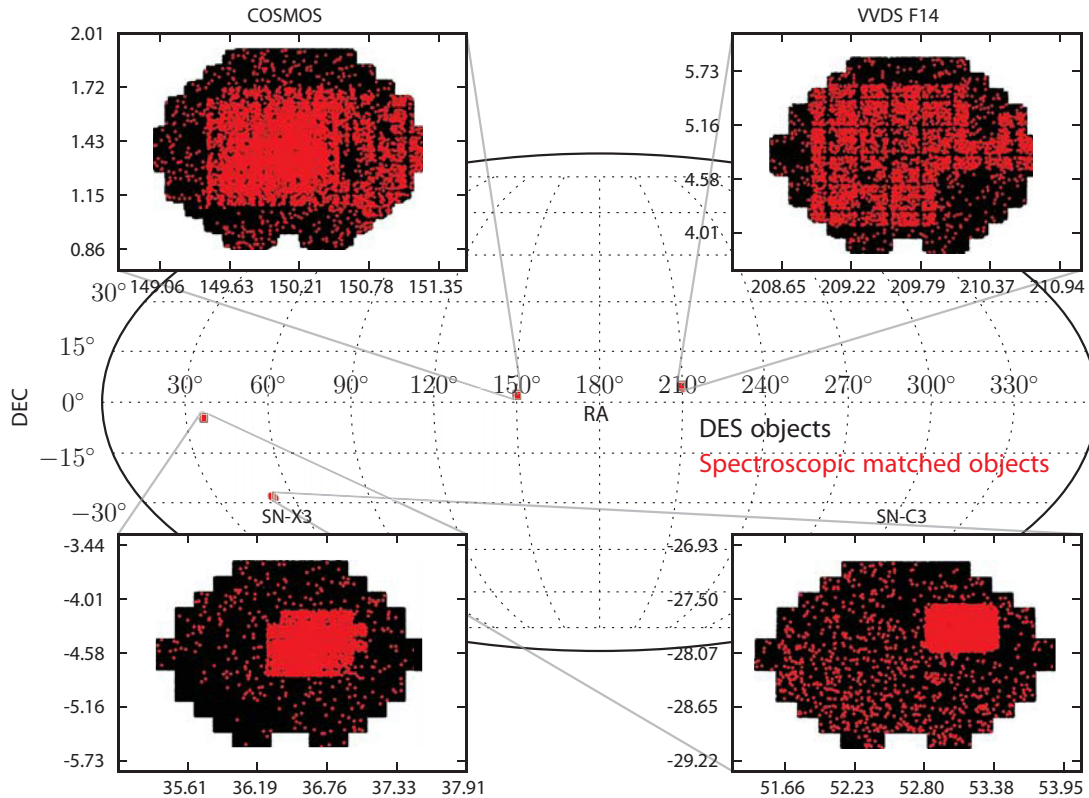


Figure 1. Positions in the sky of the four calibration fields. In the zoomed-in inset panels, it is possible to observe the spectroscopic matched galaxies, in red, in front of all the DES galaxies detected in the fields, in black.

on bias, resolution and outlier fraction that were defined prior to the start of the survey.

The outline of the paper is as follows. Section 2 describes the DES-SV photometric galaxy sample, whereas the spectroscopic galaxy samples are presented in Section 3, together with the weighting technique that has been used to match their depth and colours to those of the DES-SV sample. Section 4 describes briefly the conditions in which the 13 different photo- z codes studied were run, and contains the bulk of the results of the paper, including the comparison between the results obtained with the different photo- z codes, the dependence of the results on both the depth of the DES-SV data and the specific spectroscopic calibration samples used, and an in-depth presentation of the results obtained with four representative photo- z codes, in particular with respect to the set of requirements of the DES survey, which we set up at the beginning of Section 4. A discussion of the main results in the paper can be found in Section 5. Finally, we present our conclusions in Section 6, while we confine to an appendix the detailed description of the metrics used to characterize the photo- z distributions.

2 DES-SV PHOTOMETRIC SAMPLE

DECam imaging on fields overlapping those from deep spectroscopic redshift surveys were obtained for the following four DES fields: SN-X3, SN-C3, VVDS F14, and COSMOS, whose positions in the sky are shown in Fig. 1. SN-X3 and SN-C3 are the two deep fields in the DES supernova (SN) survey, and dithered observations of these fields were obtained routinely during the DES-SV period. The SN-X3 field includes the VVDS-02hr field of the VVDS Deep survey (Le Fèvre et al. 2005, 2013), while SN-C3 overlaps with

the Chandra Deep Field South area of the ACES survey (Cooper et al. 2012). The VVDS F14 field was centred on the VVDS-Wide redshift survey 14hr field (Garilli et al. 2008), and dithered imaging to DES main survey depth of this field was likewise obtained during DES SV. Deep dithered imaging data for the COSMOS field, centred on the Cosmological Evolution Survey (COSMOS) area (Lilly et al. 2007, 2009) were obtained during 2013 February by a DECam community programme.² Each one of the four fields covers about the area of a single DECam pointing, or about 3 deg^2 . See Section 3 for a detailed description of the spectroscopic data matched in each of the fields.

All fields include imaging in the five DES filters *grizY*, and additionally in the *u* band, which is part of DECam but not used by the DES survey. The data have been processed to two imaging depths: Main, corresponding to approximately DES main survey exposure times, and Deep, corresponding to about three times the exposure of a single visit to a DES SN deep field (for SN-X3 and SN-C3) or deeper (for COSMOS). Differences in signal-to-noise ratio (S/N) between the Main and Deep samples can be appreciated in Fig. 2; details of the data, the exposure times used and the magnitude depths are given in Table 1. Similar to DES science requirements convention, the 10σ magnitude limit is defined to be the MAGAUTO value (see definition below in this section) at which the flux in a 2 arcsec diameter aperture is measured at 10σ . Note that for the SN-X3 and SN-C3 fields, we selected those SV observations that approximately met DES main survey sky background and seeing criteria in constructing the processed data used for this paper.

²Proposal 2013A-0351: made available for DES photo- z calibration use by arrangement with PI: Arjun Dey.

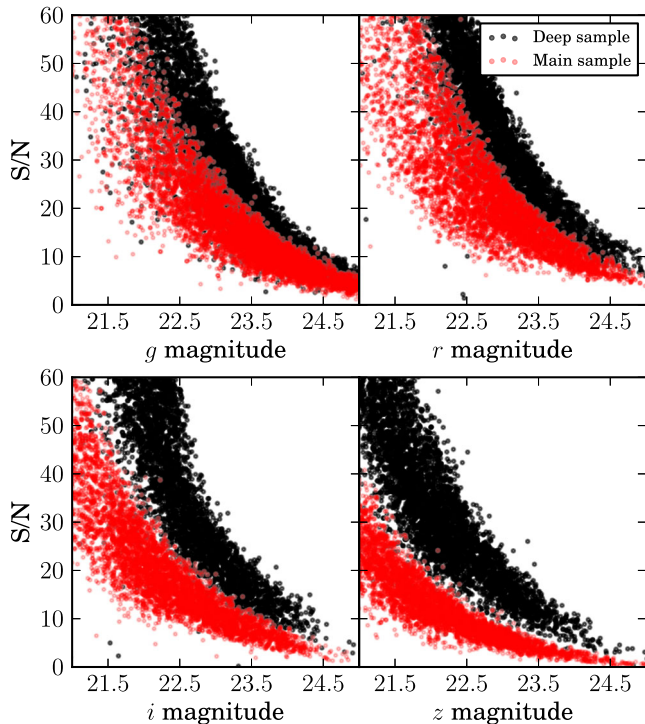


Figure 2. S/N versus magnitude for g , r , i and z DES bands, and for Main (red dots) and Deep (black dots) samples.

The data were processed using the same routines used by DES Data Management (DESDM) in their processing pipeline (Desai et al. 2012; Mohr et al. 2012), in particular for image detrending, astrometric calibration (SCAMP; Bertin 2006), image remapping and co-addition (SWarp; Bertin et al. 2002), point spread function modelling (PSFEX; Bertin 2011), and object detection and photometry (SEXTRACTOR; Bertin & Arnouts 1996). The data were processed by running these codes in standalone mode at Fermilab, rather than by running them within the DESDM processing framework at National Center for Supercomputing Applications (NCSA). Running standalone was needed as the DESDM framework was not yet fully setup at the time (Spring 2013) to process and calibrate the data for these isolated fields all the way through to image co-addition.

Though we basically used the DESDM codes, there were some detailed differences in processing and photometric calibration that we highlight here. For image detrending, we did not include corrections for CCD non-linearity, pupil ghost, and illumination that are now used by DESDM, as these corrections were not available at the time. Image co-addition was done using a median co-add rather than by using a weighted mean as in DESDM. Photometric calibration in the $ugriz$ filters for the SN-X3, VVDS F14, and COSMOS fields was done by matching against overlapping bright stars from the SDSS Data Release 9 data base (Ahn et al. 2012). This was done to calibrate each individual CCD on each separate DECam exposure, before image co-addition. In the Y band for all fields, and in all filters for the SN-C3 field (which did not overlap SDSS), we picked a fiducial exposure for each field, adopted the typical DECam CCD-by-CCD photometric zero-points as determined from DES-SV data, and then tied the photometry for subsequent exposures/CCDs to the fiducial exposure by matching overlapping bright objects. In addition, we also applied a further relative photometric calibration step, by selecting bright $r = 18$ – 22 galaxies in each field and offsetting the zero-points in the other five filters so that the median galaxy

Table 1. Imaging exposure times and depths for photo- z calibration fields.

Field	Filter	Tot. exp. time (s)	10σ Depth
SN-X3 Main	u	900	22.6
SN-X3 Main	g	800	24.1
SN-X3 Main	r	1200	24.3
SN-X3 Main	i	1080	23.6
SN-X3 Main	z	990	22.7
SN-X3 Main	Y	500	21.0
SN-C3 Main	u	900	22.9
SN-C3 Main	g	800	24.3
SN-C3 Main	r	1200	24.3
SN-C3 Main	i	1080	23.6
SN-C3 Main	z	990	22.9
SN-C3 Main	Y	500	20.9
VVDS F14 Main	u	900	22.5
VVDS F14 Main	g	900	24.0
VVDS F14 Main	r	900	23.6
VVDS F14 Main	i	900	23.1
VVDS F14 Main	z	900	22.4
VVDS F14 Main	Y	750	21.6
SN-X3 Deep	u	900	22.6
SN-X3 Deep	g	2000	24.5
SN-X3 Deep	r	3600	24.9
SN-X3 Deep	i	5400	24.5
SN-X3 Deep	z	10890	24.0
SN-X3 Deep	Y	1000	21.0
SN-C3 Deep	u	900	22.8
SN-C3 Deep	g	1800	24.6
SN-C3 Deep	r	3600	24.9
SN-C3 Deep	i	5400	24.5
SN-C3 Deep	z	10890	24.3
SN-C3 Deep	Y	500	20.8
COSMOS Deep	u	33600	25.2
COSMOS Deep	g	4500	24.8
COSMOS Deep	r	4800	24.9
COSMOS Deep	i	12000	24.8
COSMOS Deep	z	7000	23.5
COSMOS Deep	Y	2400	21.6

colours relative to r (i.e. $g - r$, $r - i$, etc.) would match those for fiducial DECam data of the VVDS-Deep 02hr field (part of SN-X3). This additional step was intended to match up the median galaxy colours among the different fields by using photometry of galaxies directly, as the earlier calibration steps use photometry of stars, and there can be small (per cent level) systematic differences between the stellar and galaxy photometric zero-points, in particular due to seeing. We also applied corrections for Milky Way extinction based on the Schlegel, Finkbeiner & Davis (1998) dust maps evaluated at the centre of each field.

As for the use of imaging data by photo- z algorithms, either MAGAUTO or MAGDETMODEL magnitudes (or both) were employed by the different photo- z codes. MAGAUTO magnitudes come from the flux (counts) measured in an elliptical aperture defined as in Bertin & Arnouts (1996). It provides an estimation of the total magnitude of the object. MAGDETMODEL magnitudes are measured from the shape (a Sérsic profile model; Sérsic 1963) fit to the object in the SEXTRACTOR detection image (either the r band or the i band for our data), and the flux is then measured separately in each band using that same model shape. Also available are MAGMODEL magnitudes, which fit the shape of the object independently in each of the bands.

However, MAGDETMODEL magnitudes, which result from one unique best-fitting shape for the object, are in general better suited for colour measurement and hence more appropriate to use for photo- z estimation.

We want to emphasize here that because of the differences mentioned above between the reductions of SV data used in this paper and the improved DESDM reductions of SV data (to be released and described elsewhere), the results of this paper are meant to reflect the photo- z quality achievable from early DES data, rather than from final DES data or even from SV data. We expect that final DESDM reductions of the calibration field data will be better in terms of photometric quality and consequently of photo- z quality, so the results in this paper will serve as a lower bound on the photo- z quality that may be achieved by final DES data. Nonetheless, as we will show later in this paper, the photo- z quality achieved in these early DES data is good and already sufficient to meet the basic DES science requirements on photo- z scatter and outlier fractions.

3 DES-SV SPECTROSCOPIC SAMPLE

In general, to exploit a galaxy photometric survey to its maximum scientific potential, it is necessary to be able to calibrate or control the performance of the photo- z estimation by using data from a spectroscopic survey. To accomplish this, it is necessary to have, for a subset of galaxies, both the spectroscopic redshifts and the estimated photo- z 's. With this information in hand, the characterization of the behaviour of the photometric redshifts is possible, and it becomes a crucial step for cosmological probes such as galaxy clustering or weak lensing. In particular, among other quantities, it is very important to characterize the true redshift distribution of a set of galaxies after a selection in photo- z space.

A photometric survey like DES will therefore need to observe one or several regions of the sky that have been previously covered by a spectroscopic survey, and then match the galaxies in the catalogue of the spectroscopic survey to galaxies observed photometrically by DES. In this paper, the matching between DES and spectroscopic galaxies is performed by using the positions of the galaxies in the sky plane, with a matching radius of 1 arcsec.

Four regions of the sky included in the DES-SV footprint have been used for photo- z calibration in this study (Fig. 1) as given below.

(i) *SN-X3 field*. This area, centred at RA $\sim 36^\circ$, Dec. $\sim -5^\circ$, overlaps with the VIMOS (Le Fèvre et al. 2003) VLT Deep Survey 02hr field. DES photometry has been matched in this field with spectroscopic redshift data from VVDS Deep (Le Fèvre et al. 2005, 2013).

(ii) *VVDS F14 field*. This area, centred at RA $\sim 209^\circ$, Dec. $\sim 5^\circ$, overlaps with the VVDS 14hr field. DES photometry has been matched in this field with spectroscopic redshift data from VVDS Wide (Garilli et al. 2008).

(iii) *SN-C3 field*. This area, centred at RA $\sim 52^\circ$, Dec. $\sim -28^\circ$, overlaps with the Chandra Deep Field South. DES photometry has been matched in this field with spectroscopic redshift data from both VVDS Deep and ACES (Cooper et al. 2012).

(iv) *COSMOS field*. This area, centred at RA $\sim 150^\circ$, Dec. $\sim -1.4^\circ$, overlaps with the Cosmic Evolution Survey field. DES photometry has been matched in this field with spectroscopic redshift data from both VVDS Wide and z COSMOS (Lilly et al. 2007, 2009).

We have also used data from brighter, shallower spectroscopic surveys available in these four regions: OzDES (Kuehn et al. 2014);

Yuan et al., in preparation); SDSS-I/II (Eisenstein et al. 2001; Strauss et al. 2002); SDSS-III BOSS Data Release 9 (Ahn et al. 2012) and 2dF Galaxy Redshift Survey (Colless et al. 2001). Galaxies matched to these surveys help to increase the statistics on the brighter tail of the distribution for galaxies used for this study.

Next we build a set of training and testing samples, using DES photometry from the Main and Deep samples and spectroscopy from different surveys. Note that only high-confidence spectroscopic redshifts, i.e. having redshift flags between 3 and 5, corresponding to secure and very secure (>95 per cent accuracy) redshift determinations, have been selected to construct these samples. Spectroscopic failures can bias cosmological results, as studied in Cunha et al. (2012b), where they showed how for a final DES analysis such failures need to be reduced to the per cent level. While a complete study concerning spectroscopic failures will have to take place for the analysis of the final DES data set, here we rely on the high-confidence redshift flags for the photo- z calibration of the early DES data. Below we describe how the data are distributed in each of the four calibration samples (training and testing for both Main and Deep catalogues).

(i) *Main training sample*. 5859 galaxies, photometry from Main catalogues, spectroscopic redshifts from the following data sets.

–One randomly selected half of VVDS Deep, SDSS/BOSS in the SN-X3 field, ACES, 2dFGRS, OzDES in the SN-C3 field, and VVDS Wide, SDSS/BOSS in the VVDS F14 field.

(ii) *Main testing sample*. 6381 galaxies, photometry from Main catalogues, spectroscopic redshifts from the following data sets.

–The other half left out from the samples in the Main training set.
–All VVDS Deep in the SN-C3 field.

(iii) *Deep training sample*. 7249 galaxies, photometry from Deep catalogues, spectroscopic redshifts from the following data sets.

–One randomly selected half of VVDS Deep, SDSS/BOSS in the SN-X3 field, ACES, 2dFGRS, OzDES in the SN-C3 field and z COSMOS, SDSS/BOSS, 2dFGRS in the COSMOS field.

(iv) *Deep testing sample*. 8358 galaxies, photometry from Deep catalogues, spectroscopic redshifts from the following data sets.

–The other half left out from the samples in the Deep training sample.
–All VVDS Deep in the SN-C3 field.
–All VVDS Wide in the VVDS F14 field.

The spectroscopic redshift distributions of the Main and Deep calibration samples defined above, spanning all the redshift range of interest for DES ($0 < z < 1.4$), are shown in Fig. 3.

3.1 The weighting procedure

In order to assess the photo- z performance of the DES-SV data we would ideally need a calibration sample being representative of the DES-SV full sample, i.e. having exactly the same photometric properties (magnitude and colour distributions). However, spectroscopic galaxy samples are shallower, and suffer from selection effects. A weighting procedure, which assigns a weight to each of the galaxies in the calibration sample so that the distributions of their photometric observables reproduce the distributions of the same observables in the full sample, can be used provided there is enough overlap between the photometric spaces of the calibration and full samples (Lima et al. 2008; Cunha et al. 2009).

Different algorithms can be used to compute the weights, but basically all compare local densities in the photometric spaces of the two samples (calibration and full) and assign a weight to each photometric region of the calibration sample equal to the ratio

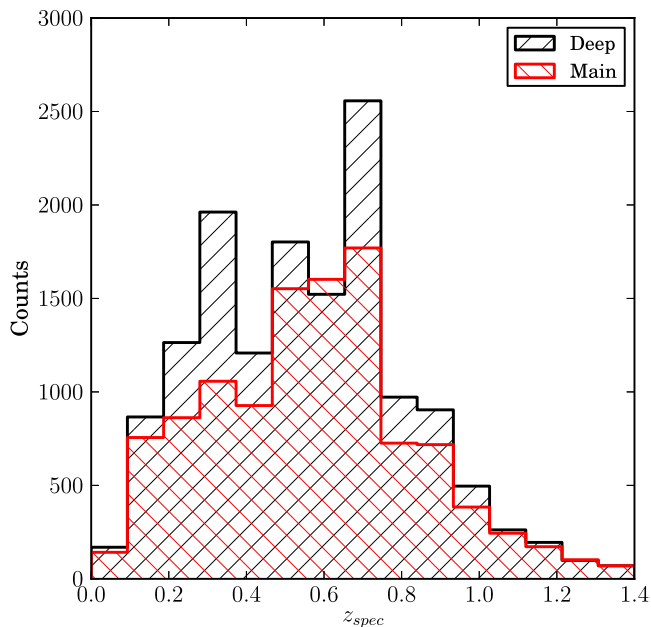


Figure 3. Redshift distributions for the spectroscopic matched galaxies in the Main and Deep calibration samples.

between the densities of galaxies in the full sample and the calibration sample in a given region. In this study, we use a nearest-neighbour algorithm to compute the weights that we use extensively throughout the paper. A detailed description of the method can be found in Lima et al. (2008).

We apply the weighting technique within a region in the multi-dimensional space defined by $18 < i_{AB} < 24$; $0 < g - r < 2$; $0 < r - i < 2$. In Fig. 4, one can check how the weighting procedure is efficiently applied for the sample used in this study. The figure shows, for two DES bands, and the Main and Deep samples, the magnitude distributions for the full sample, the calibration sample and the weighted calibration sample, whose distributions agree very well with those of the full sample.

4 PHOTOMETRIC REDSHIFTS IN THE DES-SV CALIBRATION SAMPLES

In this section, we include all the photo- z analyses using the calibration data defined in Section 3. The analysis is carried out employing an extensive set of statistics. To construct most of the metrics used in this paper, we first define the bias to be $\Delta z = z_{\text{phot}} - z_{\text{spec}}$ and the normalized bias by its reported error as $\Delta z' = (z_{\text{phot}} - z_{\text{spec}})/\epsilon_{\text{phot}}$, where ϵ_{phot} is the error in the estimation of the photo- z 's. We present the standard metrics used to compare the accuracy of the different codes in Table 2, together with the DES science requirements for photo- z 's, set before the start of the survey. The DES science requirements are driven by the Dark Energy science that DES plans to carry out, in particular by weak lensing and large-scale structure tomographic measurements.

The photo- z metrics we consider are intended to measure the quality of the photometric redshifts in terms of their bias, scatter, and outlier fraction statistics, and also in terms of the fidelity of the photo- z errors and of the agreement between the photo- z and true redshift distributions. Detailed definitions of these metrics are given in Appendix , while here we briefly summarize and motivate our choice of metrics.

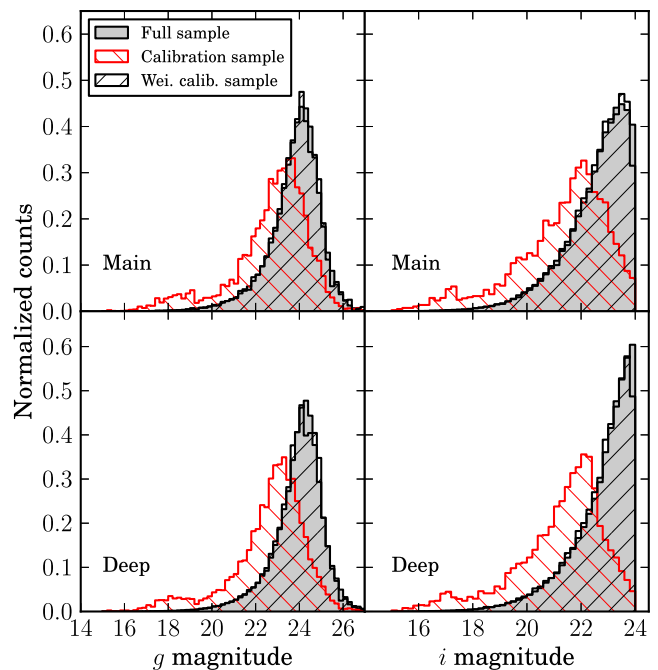


Figure 4. g - and i -magnitude distributions for the full, calibration and weighted calibration sample. The difference between the full and the calibration samples is apparent, the latter being significantly brighter. After applying the weighting procedure described in Lima et al. (2008), the weighted calibration distributions agree very well with the corresponding DES-SV distributions.

(i) *Bias.* To quantify the overall photo- z bias, we use the mean bias $\overline{\Delta z}$ and the median bias Δz_{50} .

(ii) *Scatter.* To measure the photo- z scatter, we use both the standard deviation $\sigma_{\Delta z}$ of Δz and the 68 percentile width σ_{68} of Δz about the median (i.e. $\Delta z_{50} \pm \sigma_{68}$ covers 68 per cent of the Δz distribution). For a Gaussian distribution of Δz , we would have $\sigma_{\Delta z} = \sigma_{68}$. However, in general Δz is not distributed as a Gaussian, so that σ_{68} measures the width of the core of the Δz distribution, whereas $\sigma_{\Delta z}$ is sensitive to the tails. The DES science requirements specify $\sigma_{68} < 0.12$ for a 90 per cent of the selectable sample of galaxies.

(iii) *Outlier fractions.* To quantify the fraction of objects with large $|\Delta z|$, we measure the $2\sigma_{\Delta z}$ and $3\sigma_{\Delta z}$ outlier fractions $\text{out}_{2\sigma}$ and $\text{out}_{3\sigma}$, respectively, both defined relative to the mean photo- z bias $\overline{\Delta z}$. The DES science requirements limit these outlier fractions to be $\text{out}_{2\sigma} < 0.1$ and $\text{out}_{3\sigma} < 0.015$.

(iv) *Fidelity of photo- z errors.* To assess the fidelity of the reported photo- z error ϵ_{phot} , we normalize Δz by ϵ_{phot} and calculate the resulting error-normalized mean bias $\overline{\Delta z'}$ and standard deviation $\sigma_{\Delta z'}$. Ideally, we would obtain $\overline{\Delta z'} = 0$ and $\sigma_{\Delta z'} = 1$, as for a Gaussian distribution of zero mean and unit variance. Deviations from these values would indicate inaccuracies in the photo- z errors.

(v) *Redshift distributions.* Finally, to compare the photometric redshift distribution $N(z_{\text{phot}})$ to the true redshift distribution $N(z_{\text{spec}})$, we use two metrics. The first is N_{Poisson} , which measures the rms difference between $N(z_{\text{phot}})$ and $N(z_{\text{spec}})$, normalized by Poisson fluctuations $\sqrt{N(z_{\text{spec}})}$. The second is the Kolmogorov–Smirnov (KS) metric that tests whether $N(z_{\text{phot}})$ and $N(z_{\text{spec}})$ are consistent with being drawn from the same parent distribution.

Some of the key DES science analyses, such as the galaxy angular correlation functions for large-scale structure studies, or cosmic

Table 2. Definition of the metrics used in the text to present the main results. These are computed in the whole redshift range as well as in bins of width 0.1 in photometric redshift. Detailed definitions can be found in the appendix.

Metric	Description	Requirement
$\overline{\Delta z}$	Mean of the Δz distribution	–
$\sigma_{\Delta z}$	Standard deviation of the Δz distribution	–
Δz_{50}	Median of the Δz distribution	–
σ_{68}	Half width of the interval around Δz_{50} containing 68 per cent of the galaxies	<0.12
$\text{out}_{2\sigma}$	Fraction of galaxies with: $ \Delta z - \overline{\Delta z} > 2\sigma_{\Delta z}$	<0.1
$\text{out}_{3\sigma}$	Fraction of galaxies with: $ \Delta z - \overline{\Delta z} > 3\sigma_{\Delta z}$	<0.015
$\overline{\Delta z'}$	Mean of the $\Delta z' = \Delta z/\epsilon_{\text{phot}}$ distribution	–
$\sigma_{\Delta z'}$	Standard deviation of the $\Delta z'$ distribution	–
N_{Poisson}	Difference between $N(z)^{\text{phot}}$ and $N(z)^{\text{spec}}$ normalized by Poisson fluctuations	–
KS	Kolmogorov–Smirnov statistic for $N(z)^{\text{phot}}$, $N(z)^{\text{spec}}$	–

shear tomography measurements for weak lensing and dark energy constraints, will use DES galaxies divided into separate photo- z bins. For these photo- z bins, the DES science requirements also specify stringent limits on the differences in bias, scatter, and outlier fractions between the DES photometric sample and the spectroscopic calibration sample. For example, there is a requirement that the mean bias $|\overline{\Delta z}| < 0.001(1+z)$ in bins of 0.1 in redshift. Accurate characterization of the full $P(z_{\text{spec}}|z_{\text{phot}})$ distribution, i.e. the distribution of true redshift in bins of photo- z , will also be needed for these DES science analyses. However, consideration of these more stringent DES photo- z science requirements is premature for this paper, because of the limited number and sky distribution of the SV spectroscopic calibration fields. These fields are subject to sample variance effects, i.e. fluctuations in galaxy densities and redshift distributions due to clustering and large-scale structure, and in fact, as detailed in Cunha et al. (2012a), meeting the requirements will necessitate a much larger number of widely distributed spectroscopic calibration fields (e.g. ~ 100) than are currently available from completed redshift surveys. We will thus postpone discussion of these issues and tests of these more stringent requirements for future DES photo- z papers. Nonetheless, we do present in a later section some example $P(z_{\text{spec}}|z_{\text{phot}})$ distribution for several selected photo- z codes.

All the results shown in this paper have been weighted according to the technique presented in Section 3.1, and they include a cut on the 10 per cent of the galaxies having larger estimated photo- z error, as given from each particular code (this 10 per cent cut on photo- z error is allowed by the DES science requirements). This cut introduces small differences among the testing samples used by each photo- z code in the comparison. With this, we test the ability of each particular code to select the most problematic objects in the sample. In addition, clustering measurements can be affected by severe photo- z quality cuts as presented in Marti et al. (2014), where they also show a procedure to correct for these effects.

All the errors for the statistics presented in the paper come from bootstrap resampling using 100 samples, unless otherwise stated.

4.1 Methods

Before going in detail into the photo- z analysis, we present here a brief description of the different methods we have used to estimate photo- z 's, which include most of the relevant photo- z codes available. We have emphasized in the details at the time of running these codes; for an exhaustive description of them, see Table 3. For template-based methods, a standardized set of filter throughput curves has been used. Most of the codes have been run in standalone

mode, while a fair fraction of them has been run within the DES Science Portal, with compatible results. Due to the large number of codes used, the paper, other than showing the DES-SV photo- z capabilities, also serves as a helpful reference to compare different photo- z codes using real data from a deep galaxy survey.

4.1.1 DESDM

The DESDM (default) photo- z 's were computed using the same ANN method that was applied to the SDSS Data Release 6 sample, as described in detail by Oyaizu et al. (2008a). In brief, we used a neural network configuration with 10 input nodes, consisting of the five *grizY* MAGAUTO magnitudes and the five *grizY* MAGDETMODEL magnitudes, followed by three hidden layers with 15 nodes per layer. The formal minimization to determine the neural network weights was done on the training set data, while choosing the set of network weights that gave the lowest photo- z scatter on the testing set, after 300 iterations of the weight minimization. Moreover, to reduce sensitivity to initial conditions in the minimization procedure, we repeated the procedure 10 times, starting each time at a different initial position in the space of weights. The final photo- z for a galaxy was taken to be the average of the photo- z 's computed from the optimal weights for each of the 10 network minimizations.

We also computed accompanying photo- z errors using the empirical *nearest-neighbour error* (NNE) technique, described in detail by Oyaizu et al. (2008b). The NNE method estimates the photo- z error for each galaxy empirically, based on the photo- z 's and true redshifts of the galaxy's 100 nearest neighbours in the spectroscopic testing set, where neighbour distance is defined using a simple flat metric in the space consisting of the 10 input magnitudes noted above. Specifically, the NNE photo- z error σ is defined so that it corresponds to the width of 68 per cent of the $|z_{\text{phot}} - z_{\text{spec}}|$ distribution of the nearest neighbours.

4.1.2 ANNZ

ANNZ (Collister & Lahav 2004) is a training-based method that uses a neural network scheme to find a functional relationship between a given set of input parameters (e.g. magnitudes, colours, etc.) and outputs a desired quantity (e.g. redshift). The results shown in this paper have been obtained by using a neural network architecture of 2 layers with 10 nodes each, and using as inputs the set of 5 MAGAUTO and 5 MAGDETMODEL magnitudes. Attempts to use a larger number of nodes as well as colours as inputs resulted in larger photo- z errors. The uncertainties in the photo- z estimation are computed

Table 3. List of methods used to estimate photo-*z*'s. Code type and main references are given.

Code	Type	Reference
DESDM, artificial neural network	Training based	Oyaizu et al. (2008a)
ANNZ, artificial neural network	Training based	Collister & Lahav (2004)
TPZ, prediction trees and random forest	Training based	Carrasco Kind & Brunner (2013, 2014)
RVMZ, relevance vector machine	Training based	Tipping (2001)
NIP-KNNZ, normalized inner product nearest neighbor	Training based	de Vicente et al., in preparation
ANNZ2, machine-learning methods	Training based	Sadeh et al., in preparation
ARBORZ, boosted decision trees	Training based	Gerdes et al. (2010)
SKYNET, classification artificial neural network	Training based	Bonnett (2013) and Graff et al. (2013)
BPZ, Bayesian photometric redshifts	Template based	Benitez (2000) and Coe et al. (2006)
EAZY, easy and accurate redshifts from Yale	Template based	Brammer et al. (2008)
LEPHARE	Template based	Arnouts et al. (2002) and Ilbert et al. (2006)
ZEBRA, Zurich extragalactic Bayesian redshift analyzer	Template based	Feldmann et al. (2006)
PHOTOZ	Template based	Bender et al. (2001)

using standard propagation of the errors in the input magnitudes to the error in photo-*z*, by using the functional relationship between these input parameters and the output photo-*z*.

4.1.3 Trees for PHOTOZ

TPZ³ (Carrasco Kind & Brunner 2013, 2014) is a machine learning, parallel algorithm that uses prediction trees and random forest techniques to produce both robust photometric redshift pdfs and ancillary information for a galaxy sample. A prediction tree is built by asking a sequence of questions that recursively split the input data taken from the spectroscopic sample, frequently into two branches, until a terminal leaf is created that meets a stopping criterion (e.g. a minimum leaf size or a variance threshold). The dimension in which the data are divided is chosen to be the one with highest information gain among the random subsample of dimensions obtained at every point. This process produces less correlated trees and allows the exploration of several configurations within the data. The small region bounding the data in the terminal leaf node represents a specific subsample of the entire data with similar properties. Within this leaf, a model is applied that provides a fairly comprehensible prediction, especially in situations where many variables may exist that interact in a non-linear manner as is often the case with photo-*z* estimation.

By perturbing the data using their magnitude errors and by taking bootstrapping samples, many (600 in this application) uncorrelated trees can be created whose results are aggregated to construct each individual pdf. For the application to DES-SV data, we have used both MAGAUTO and MAGDETMODEL magnitudes in the five DES bands, together with all the corresponding colours as well as their associated errors.

4.1.4 RVMZ

RVMZ is an empirical photo-*z* code based on the relevance vector machine (RVM) algorithm from Tipping (2001), a Bayesian sparse kernel method for regression. The RVM has characteristics similar to the support vector machine, but includes a Bayesian treatment for the determination of the model weights. This has the advantage that the parameters governing model complexity and noise variance are found in the training run itself, and therefore the RVM does not require cross-validation to optimize these parameters. We use the

RVM implementation in the R-package `kernlab` from Karatzoglou et al. (2004). To obtain photo-*z* estimates, we used MAGDETMODEL magnitudes (*grizY*) and colours ($g - r$, $r - i$, $i - z$, $z - Y$) as input. We reconstruct the pdf by combining the uncertainties in the data sets and the model. In the training set, we use the *k*-fold cross-validation technique, which consists of partitioning the data in *k* groups, then *k* - 1 of these groups are used to train the model which is then evaluated on the hold-out group. This process is then repeated for all possible choices of the hold-out group and the resulting mean-squared error for the redshift prediction is evaluated. At this stage, we obtain the model error as the rms of the predicted pdf. Details of the method will be described in Rau et al. (in preparation).

4.1.5 NIP-KNNZ

NIP-KNNZ (Juan De Vicente et al., in preparation) is a novel technique that computes the photo-*z* from a nearest-neighbour approach based on the Normalized Inner Product (NIP). While Euclidean magnitude–distance ensures that close galaxies in magnitude space are assigned the same redshift, it does not consider as neighbours galaxies with the same colour but separated in overall magnitude. NIP metrics corrects this by considering two galaxies as neighbours, and hence with close redshift, when they have similar colours, rather than magnitudes. The metric is based on the inner product definition:

$$\text{NIP} = \cos \alpha = \frac{\mathbf{M}_t \cdot \mathbf{M}_p}{M_t M_p}, \quad (1)$$

where \mathbf{M}_t and \mathbf{M}_p are the multimagnitude vectors of training and photometric galaxies, respectively. For this particular application, the five MAGDETMODEL magnitudes were used and turned into fluxes. The NIP is related to the angle that the two multimagnitude vectors form. Maximizing NIP is equivalent to minimize the angle between the two vectors. Regarding the photo-*z* error, an empirical formula has been derived to account for three different contributions. The first term is the floor error related to the finite spectroscopic redshift precision, which is all that remains in the best case scenario, when the magnitude vectors of the photometric and the spectroscopic galaxies point in the same direction. The second contribution comes from the characterization of the photo-*z* errors in the spectroscopic sample. NIP-kNN is run over all galaxies in the spectroscopic sample to obtain their photo-*z*s. One half of the difference between the spectroscopic-*z* and the photo-*z* of the spectroscopic galaxy is taken as its photo-*z* error. When NIP-kNN is applied to a galaxy in the photometric sample, it inherits not only the *z* of the closest

³<http://lcdm.astro.illinois.edu/research/TPZ.html>

spectroscopic galaxy but also its photo- z error. The third term is the metric distance $\sin(\alpha)$ that accounts for neighbourhood, multiplied by a constant determined empirically. Assuming the spectroscopic sample spans the range of redshifts of the photometric sample, NIP-KNNZ achieves, by construction, an accurate reconstruction of the redshift distribution $N(z)$

4.1.6 ANN2

ANN2 is a new major version of the public photo- z estimation software, ANNZ (Collister & Lahav 2004), which will be made public in 2015. The new code incorporates several machine-learning methods, such as ANNs, boosted decision trees (BDTs, Freund & Schapire (1997), described below in Section 4.1.7) and k -nearest neighbours (KNNs). The different algorithms are used in concert in order to optimize the photo- z reconstruction performance, and to estimate the uncertainties of the photometric solutions. This is done by generating a wide selection of machine-learning solutions with e.g. different ANN architectures, initialized by different random seeds. The optimization is performed by ranking the different solution according to their performance, which is determined by the respective photo- z scatter of each solution.

The single solution with the best performance is chosen as the nominal photo- z estimator of ANN2. In addition, the entire collection of solutions is used in order to derive a photo- z probability density function (pdf), constructed in two steps. First, each solution is folded with an error distribution, which is derived using the KNN error estimation method of Oyaizu et al. (2008a). The ensemble of solutions is then combined. This is done by weighing the different estimators, in such a way as to produce a pdf which describes the underlying photometric errors. The inputs used in this study were the five MAGAUTO and the five MAGDETMODEL magnitudes.

4.1.7 ARBORZ

The ARBORZ algorithm (Gerdes et al. 2010; Sypniewski 2014) is a training-set-based, publicly available (Sypniewski & Gerdes 2013) photo- z estimator that makes use of BDTs. BDTs were developed to classify objects characterized by a vector of observables \mathbf{x} into two categories. Decision trees are trained iteratively, with initially misclassified objects given higher weight, or ‘boosted’, in the next training cycle. An individual decision tree is a relatively weak classifier. But the ‘forest’ of trees generated during the training process, when their outputs are combined in a way that assigns higher weight to trees with lower misclassification rates, collectively constitutes a strong classifier. To adapt a binary classifier to the problem of determining a continuous quantity like redshift, we divide the redshift range of interest into N discrete bins with a width roughly 25–50 per cent of the expected photo- z resolution, and train a separate BDT classifier for each redshift bin, using a forest size of 50 trees. Each classifier is trained to identify galaxies with a redshift falling in its particular bin as ‘signal,’ and to reject galaxies falling more than $3\sigma_{z,\text{phot}}$ away from its bin as ‘background’. The $3\sigma_{z,\text{phot}}$ exclusion region between signal and background objects is introduced in order to avoid the overtraining that could result from treating a galaxy with, e.g. a redshift of 0.999 as signal and one with 1.001 as background. Each BDT classifier in this ensemble, when presented with a new galaxy, produces a probability that the given galaxy falls within its redshift bin. This collection of probabilities constitutes the ARBORZ pdf. ARBORZ also provides a single best-estimate photo- z (from the median of the pdf) and its error, σ_{68} (from its width);

however, the full pdf provides the best characterization of a galaxy’s photo- z .

In the DES-SV sample, we train ARBORZ using the MAGAUTO and MAGDETMODEL magnitudes in *grizY*, with 50 fixed-width redshift bins out to $z = 1.5$. The results are fairly robust with respect to reasonable variations in the forest size, number of bins, the choice of fixed- or variable-width bins, and the use of either or both sets of magnitudes.

4.1.8 SKYNET, classification neural network

This method, first used in Bonnett (2013), consists of using a neural network to classify galaxies in classes, in this case redshift bins. A neural network with a *softmax* transformed output (Pyle 1999) is able to estimate the probability that an object belongs to a certain class. Given the fact that a galaxy cannot live in more than one redshift bin at the same time, a neural network with a *softmax* transformation is ideally suited to return a pdf for each galaxy. Before training the neural network, we bin our data in n redshift bins, the classes. The neural network is fed by the MAGAUTO magnitudes, the MAGDETMODEL magnitudes and the correct classes. The neural network outputs n values between [0, 1] for each galaxy. These n values sum up to 1 and hence can be interpreted as the probability that a galaxy resides in a redshift bin. In this particular run we used $n = 40$, resulting in a $\delta z = 0.035$ granularity in the pdf. The publicly available neural network software SKYNET (Graff et al. 2013) was used for this work. The neural net is trained using 3 hidden layers with, respectively, 20, 30 and 40 nodes per layer.

4.1.9 BPZ

The Bayesian Photometric Redshifts (BPZ) photo- z code from Benitez (2000) and Coe et al. (2006) is a template-based method that returns the whole probability density distribution $p(z|m_i)$ that the galaxy is at redshift z when its magnitudes in each band are m_i , and also a single photo- z value computed as the maximum of $p(z|m_i)$. Following Bayes Theorem, $p(z|m_i)$ is the product of a likelihood and a prior probability function that represents our previous knowledge of the redshift and spectral type t distributions of the sample in the analysis. In the likelihood, we use the five MAGDETMODEL magnitudes.

(i) *Templates*. We use the eight spectral templates that BPZ carries by default based on Coleman, Wu & Weedman (1980) and Kinney et al. (1996), and add two more interpolated templates between each pair of them by setting the input parameter INTERP=2 (option by default).

(ii) *Prior*. We explicitly calibrate the prior in each test by fitting the empirical function $\Pi(z, t|m_0)$ proposed in Benitez (2000) to the corresponding DES-SV training sample. With this, we are able to remove most of the catastrophic outliers which for template-based methods tend to constitute a sizable fraction of all galaxies.

(iii) *Training*. No other training or calibration has been attempted.

4.1.10 EAZY

The EAZY photo- z code (Brammer et al. 2008) is a template-based maximum-likelihood method that has been specifically optimized for use when representative spectroscopic redshifts are not available for training set based estimators. In this implementation, we use the five MAGDETMODEL magnitudes.

(i) *Templates*. The code uses a novel non-negative matrix factorization algorithm to construct a minimal set of templates which are linear combinations of templates derived from semi-analytic models, based on the Bruzual & Charlot (2003) models but with the star formation histories computed from the semi-analytic models of De Lucia & Blaizot (2007). These model templates are likely to be more representative of galaxies, particularly at high redshifts, compared to the commonly used Coleman et al. (1980) or Bruzual & Charlot (1993) templates. The code also makes use of a template error function to account for uncertainties in templates over specific wavelength ranges.

(ii) *Priors*. No priors are used.

(iii) *Training*. No calibration of the templates is performed using spectroscopic training data. Equally, photometric offsets cannot be derived using spectroscopic training sets. In instances where significant offsets need to be applied to the photometric catalogues, the code is therefore unlikely to perform well relative to template-based codes where these offsets can be directly estimated from the training data.

4.1.11 LEPHARE

LEPHARE (Arnouts et al. 2002; Ilbert et al. 2006) is a public template-fitting code that uses a χ^2 minimization of differences between observed and theoretical magnitudes to find the best template (galaxy type) and redshift. The code also provides for each object upper and lower 1σ limits for this estimate, a maximum-likelihood estimate for the redshift, K -corrections and absolute magnitudes for each band and a probability distribution function. An adaptive method can be used to improve theoretical magnitudes, as well as an $N(z)$ prior to minimize catastrophic errors. The effect of emission lines on the theoretical magnitudes can be estimated and taken into account. Several sets of SEDs and extinction laws are available in the code to be used. For this study, we use the five MAGDETMODEL magnitudes.

(i) *Templates*. A set of 66 SEDs were used in the analysis of the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) data by Ilbert et al. (2006) obtained from interpolation of the largely used Coleman et al. (1980) templates for different Hubble types and Kinney et al. (1996) for starburst galaxies. Since template-fitting codes are time consuming, we searched for a reduced group of templates from this large set, yielding essentially the same overall statistics as far as dispersion and outlier fraction are concerned. We performed several tests using the VVDS-02hr sample with available spectroscopic redshifts and found a reduced set of 21 templates encompassing SEDs for 12 Ellipticals, 6 Spirals, 1 Im and 2 starburst with satisfactory results. Several tests removing the u band from VVDS and CFHTLS data indicated that the discrepancies of photo- z from the true (spectroscopic) value increase due to galaxy type-extinction degeneracy. From these tests, we concluded that in order to minimize this problem we should keep the extinction values $E(B - V)$ less or equal to 0.25 (for types Scd and later) and use only three very late type SEDs (1 Im and 2 starbursts).

(ii) *Prior*. Although the $N(z)$ prior available in LEPHARE refers to the $(B - I)$ colour from the VVDS survey, we verified that applying the procedure to the $g - i$ DES colour we achieved a significant minimization of the outlier fraction, and therefore we used this prior.

(iii) *Training*. The adaptive training method available in LEPHARE was used to obtain a re-calibration (zero-points offsets) in each band. This procedure was first applied to the training sets and the shifts obtained were used when the code was run on the testing samples.

4.1.12 ZEBRA

The Zurich Extragalactic Bayesian Redshift Analyser (Feldmann et al. 2006) is a flexible and powerful photometric redshift code, based around template fitting. The code produces a posterior distribution for each galaxy in redshift and template space, $P(z, T)$, as well as marginalized distributions for $P(z)$ and $p(T)$. For redshift computation, the filters were smoothed over a scale of 100 Å, and the templates smoothed over 30 Å. The tests were conducted with a redshift resolution in linear steps of $\Delta z = 0.01$. The five MAGDETMODEL *grizY* magnitudes were used in this analysis.

(i) *Templates*. 81 templates were used in ZEBRA's Bayesian mode. These templates were selected from a superset of SEDs, and consist of the most frequent best-fitting templates at $z = z_{\text{spec}}$ for galaxies in the training sample. The superset of template SEDs were produced by ZEBRA's template correction module from log interpolations between the Coleman et al. (1980) and Kinney et al. (1996) templates. The module allows the user to define redshift intervals within which the templates are modified to better fit the input photometry. For this stage, a photometric sample from the COSMOS field was used. A fraction of the testing set galaxies have counterparts within this photometric sample, but the photometry is independent from the DES-SV data and the spectroscopic redshifts were not used.

(ii) *Prior*. In Bayesian mode, ZEBRA constructs a self-consistent iterative prior from the galaxy likelihood functions, $L(z, T)$. The approach is similar to that taken by Brodwin et al. (2006), but operates in 2D redshift and template space. The prior constructed from the training sample was used for both samples.

(iii) *Training*. Bordoloi, Lilly & Amara (2010) describe a method of using known (spectroscopic) redshifts to correct the individual marginalized redshift probability distributions, $P(z)$. The method demands that the spectroscopic redshifts sample their respective $P(z)$ fairly, i.e. the distribution of cumulative probabilities between zero and z_{spec} should be flat. We apply a simple first pass of their approach in bins of redshift, with width $\Delta z_{\text{phot}} = 0.1$. Galaxies were assigned to these bins based on their maximum posterior redshift. After correction of the individual $P(z)$, a new z_{phot} was computed as the maximum of the corrected $P(z)$.

4.1.13 PHOTZ

PHOTZ (Bender et al. 2001) is a Bayesian template-fitting photometric redshift code. The redshift probability of an object is obtained by multiplying the probability of a χ^2 fit of template SEDs by prior probabilities for redshift and luminosity. The total probability of a model then reads

$$P(\boldsymbol{\mu}|m) \propto \mathcal{L}(m|\boldsymbol{\mu}) \cdot P(\boldsymbol{\mu}), \quad (2)$$

where m denotes the photometric data (in magnitudes or fluxes), and $\boldsymbol{\mu}$ are the model parameters, i.e. redshift z and luminosity M . In this analysis, we used the five MAGDETMODEL *grizY* magnitudes.

(i) *Templates*. The template set we use contains templates ranging from star forming (blue) to passively evolving (red) galaxies. It includes model SEDs from Bender et al. (2001), which were created from spectroscopically observed objects from the *Hubble Deep Field* North. Another three templates (an S0, Sac, and an Sbc galaxy) are from Mannucci et al. (2001), and two empirical SEDs (of an Scd and an Sbc galaxy) are from Coleman et al. (1980). Our model set additionally includes 13 SEDs from Ilbert et al. (2006) which are based on spectra from Coleman et al. (1980) and were optimized to match local star-forming galaxies. This is a combination of template

sets already used in the past for photometric redshift estimation (e.g. Bender et al. 2001; Brimiouille et al. 2013). Furthermore, we incorporate a set of red SEDs in our model set which were created in order to match the colours of LRGs from SDSS-II (Greisel et al. 2013).

(ii) *Priors.* The redshift and luminosity priors have the form $P(x) \propto \exp(-((x - \hat{x})/\sigma)^p)$, where \hat{x} , σ , and p are defined individually for each model SED. Setting \hat{x} , σ , and p accordingly, we can, for instance, decrease the probability of observing red models at higher redshifts ($z \gtrsim 0.9$), or of galaxies that are too bright in absolute magnitude to exist. In addition to that, we adapt the z and M (absolute magnitude) priors for every model SED in such a way that photometric redshift outliers with $|z_{\text{phot}} - z_{\text{spec}}|/(1+z) > 0.15$ (Ilbert et al. 2006) in the Main and Deep DES-SV training sets are less likely. Therefore, we identify their location in the z versus M space and modify the priors in such a way that they assign smaller probabilities to those regions. This is done solely if the outliers of a template are isolated from good photometric redshift estimates of the same template in the z versus M space.

(iii) *Training.* We iteratively adapt the zero-points for the training catalogues using the median magnitude offsets between the data and the model predictions while optimizing the photometric redshift performance.

4.2 Results of the photo- z analyses

After the description of the codes, we turn to the study of their performance in a number of tests using different configurations of the data samples. While most of these tests focus on estimating how the photo- z determination will perform for the standard DES data, others look for improvements by using deeper photometry or additional bands. We also check the differences in the results under variations in the calibration data and the weights used. Note that the results presented in this subsection are those considering all the galaxies (with quality cuts), which are represented by one single statistic, later in the paper we analyse some of these results in more detail.

4.2.1 Test 1: Main–Main

This test is the most representative of the results shown in this paper, the default case. We use here the Main training sample to train and calibrate the photo- z algorithms and the Main testing sample to validate them, therefore, the test represents the real situation for most of the data collected in the DES survey.

In order to display the performance of all codes, and only for this test in particular, in Fig. 5 we show the z_{phot} versus z_{spec} scatter plot for all the codes listed in Table 3. Furthermore, we compute all the metrics presented in Table 2 and described in Appendix A1. The results, using all the objects in the testing sample except for the 10 per cent quality cut mentioned above, are shown in Table A1 and Figs 6–9. The legend is only shown in Fig. 6, but applies to subsequent figures corresponding to this test.

Fig. 6 shows σ_{68} , related to the precision of the photometric redshifts (and defined in Appendix A1), versus the mean bias of the photo- z 's. The black dashed line sets the DES science requirement on σ_{68} , and one can check how most of the codes presented in this work are below this line, thus fulfilling this important requirement on precision. Also, among the codes satisfying the σ_{68} requirement, there is a subgroup having very low bias as well. In Fig. 6, we show a zoomed-in of this region of interest, where we can see how training-based codes, either producing a single photo- z estimate or a pdf,

$P(z)$, are the ones showing best performance (all the codes in the zoomed-in region belong to the training-based category).

Fig. 7 shows the 3σ versus 2σ outlier fractions for Test 1. The requirement on the 2σ outlier fraction (0.1) is beyond the range of the plot, meaning that all codes fulfil this. However, the 3σ outlier fraction requirement, shown as the black dashed line, is only met by a few codes. Among these codes, there are cases from the two types of photo- z codes, training and template based. Also, there is more homogeneity in this plot: many codes agree with others within error bars. Both of these requirements are set based on the spread of the Δz distribution and no with respect to a fixed distance from the mean, therefore these values quantify how sharp the Δz distribution is with respect to its centre.

Fig. 8 shows rms versus bias for the Δz distribution normalized by its error ($\Delta z'$) of each of the codes analysed in Test 1 ($\sigma_{\Delta z'}$ versus $\overline{\Delta z'}$ in our notation). A large fraction of the codes yield very high values of $\sigma_{\Delta z'}$ (expected to be close to 1), meaning that all these codes underestimate their photo- z errors, however, there is a group of codes with normalized Δz distributions approaching a Gaussian with mean equal to 0 and variance equal to 1. Also, we do not see a particular type of photo- z code being problematic here, both training and template-based codes populate good and bad regions of the plot.

One crucial aspect of photo- z studies, which will be discussed in more detail in Section 4.3, is the estimation and calibration of the true galaxy redshift distributions. In this paper, we use two metrics to compare the reconstruction of the true redshift distribution by the different photo- z algorithms: the N_{Poisson} and KS statistics, defined in Appendix A. In both cases, the smaller the value, the closer are the true redshift distribution and its reconstruction through photo- z 's. Fig. 9 shows these values for all the codes analysed in Test 1. As expected, the two metrics are strongly correlated. It can also be seen how having a redshift pdf for each galaxy, instead of a single-estimate photo- z , helps a given code to have a better redshift reconstruction. This can be inferred looking at the cases where both the pdf and the single estimate are displayed (TPZ, ANNZ2, BPZ): in all these cases, the pdf version of the code obtains better results in terms of these two metrics. As for the results, TPZ and the nearest-neighbour code, NIP-KNNZ, show the best performance in this regard.

To summarize the results from Test 1, we note that most of the codes presented in this work fulfil the requirements for σ_{68} and 2σ outlier fraction, while only a few fulfil the 3σ outlier fraction requirement. Also, training-based codes seem to yield better photo- z precision on average and better $\mathcal{N}(z)$ reconstruction, but, when evaluating other quantities like outlier fraction or the estimation of photo- z errors, there is no clear indication as of which class of photo- z approach show more accurate metrics. As pointed out in Carrasco Kind & Brunner (2014) these results might vary for different regions on the multidimensional photometric space or within the redshift range. Usually, training-based algorithms perform better on areas well populated with training galaxies and poorly on those less dense regions (as in high-redshift bins), fact that we can observe from Fig. 5, where training-based methods tend to have tighter distributions at the centre while some template-based methods can compute photo- z 's for galaxies at higher redshift more efficiently.

4.2.2 Test 2: Deep–Deep

In this test, we train or calibrate the algorithms using the Deep training sample and we apply them on the Deep testing sample (see results in Table A2). The goal of this test is to check the differences

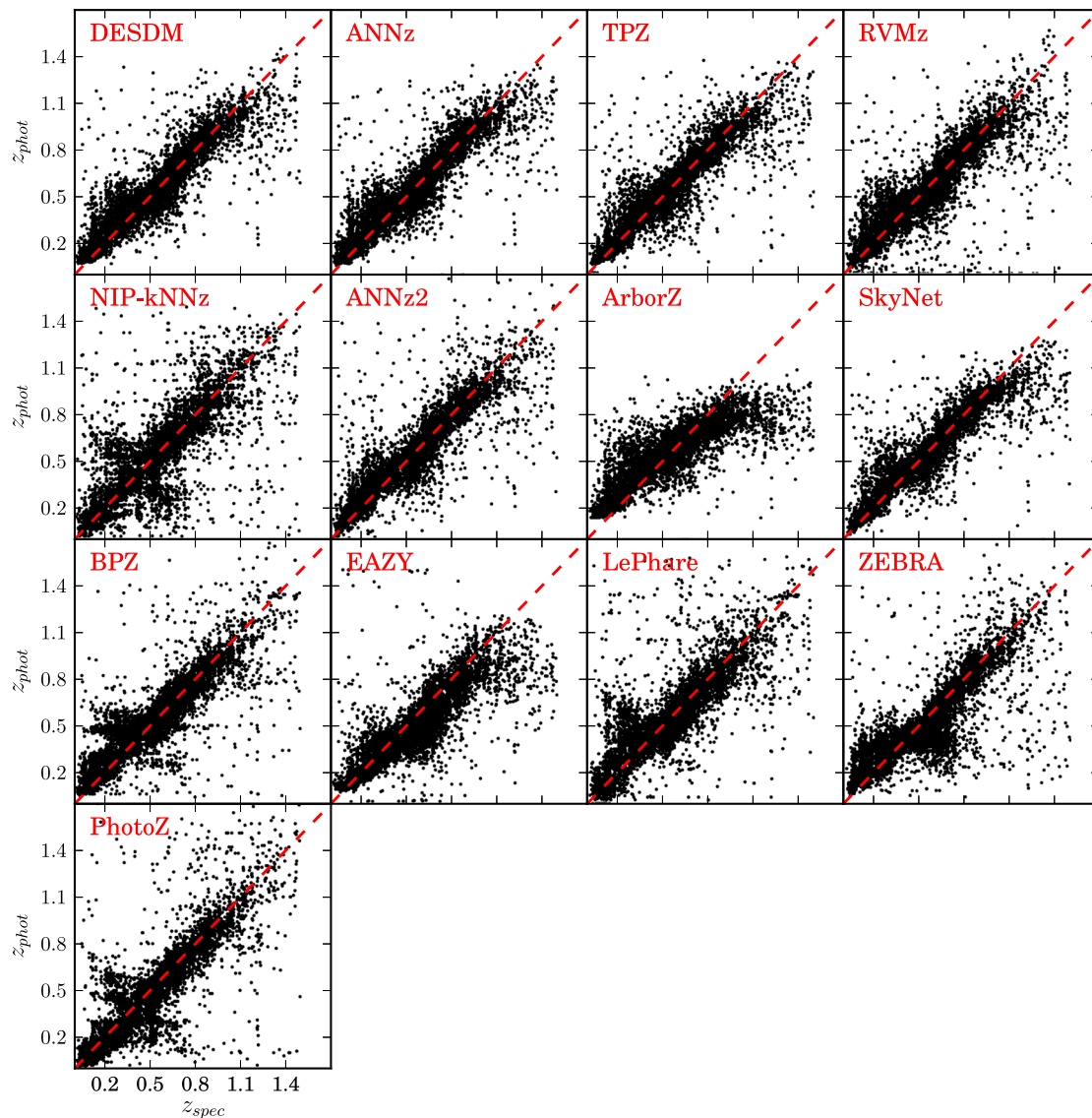


Figure 5. z_{phot} versus z_{spec} scatter plot for all the codes analysed in Test 1 and listed in Table 3.

in photo- z performance when we use higher S/N data, with the caveat that this sample also has a slightly different redshift range. In order to enable an easier comparison with Test 1, we include in the plots for this Test 2 the data from the analogous plots in Test 1, and we do it by including the points for Test 1 (without error bars, for simplicity) connected by a straight line to the corresponding point for Test 2 (with error bars). Fig. 10 shows σ_{68} versus the mean bias of the computed photo- z 's for Test 2. This is analogous to Fig. 6. There we can see that there is general improvement in σ_{68} : most of the codes move to lower values in the plot with respect to Test 1. So we conclude that using higher S/N photometry observations increases the photometric redshift precision of a sample. However, in the case of the bias there is no general trend: some codes improve and some others do not.

Figs 11 and 12 are analogous to Figs 7 and 8, respectively. In the plot showing the outlier fraction (Fig. 11), Test 2 seems to introduce an additional scattering, i.e. codes that did better in Test 1 seem to improve more in Test 2 than codes that did not do so well in Test 1, which now seem to worsen on average. This fact somehow removes the homogeneity we saw in Fig. 7, where many

codes produced results compatible within errors. Fig. 12 shows the rms and bias for the normalized Δz distribution, which, similarly to Test 1, shows that many codes underestimate the photo- z errors. The behaviour here is similar to the one regarding outlier fractions: Test 2 introduces even more differences between codes doing well and codes that are less accurate.

4.2.3 Test 3: Deep+Main

Test 3 uses the Deep training sample for training or calibration of the algorithms and uses the Main testing sample to compute photo- z 's (see results in Table A3 in the appendix). In contrast to Test 2, this is a realistic case since a Deep training sample already exists and is available to use for photo- z calibration of DES data, with the same depth as the Main testing sample. Therefore, this test explores the possibility of improving the photo- z performance of Test 1 by using higher S/N data for training or calibration.

Fig. 13 shows the results on precision and bias for Test 3, and, as before, compares with Test 1. Some template-based codes are

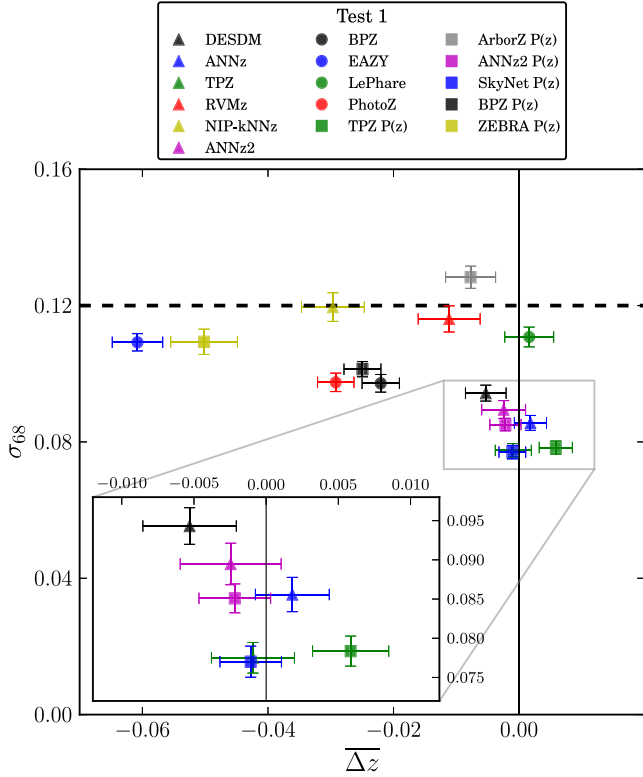


Figure 6. σ_{68} versus bias for all the codes analysed in Test 1. The black dashed lines represent the DES science requirements in this and subsequent figures. Training-based codes have triangles as markers, template-based have circles, and codes producing a pdf for the redshift are marked with a square. This will also be the convention for the next figures. Training-based codes, either producing a single photo- z estimate or a pdf, are the only ones present in the region of best performance (zoom-in).

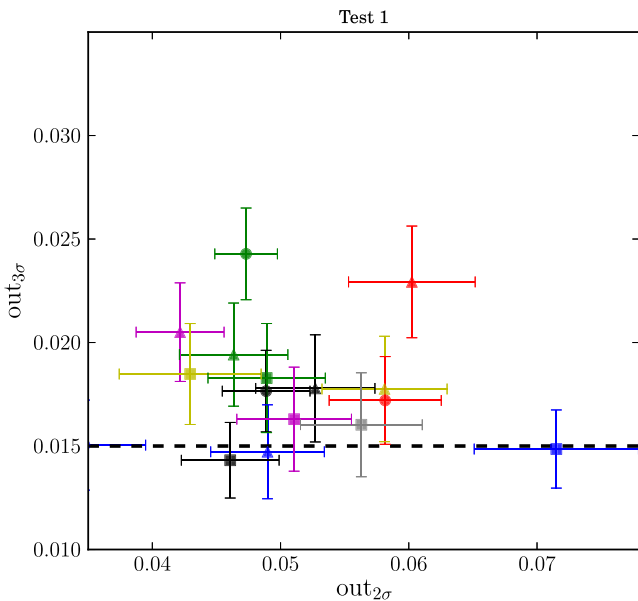


Figure 7. 3σ versus 2σ outlier fraction for all the codes analysed in Test 1. The results are more homogeneous than those for σ_{68} versus bias (Fig. 6).

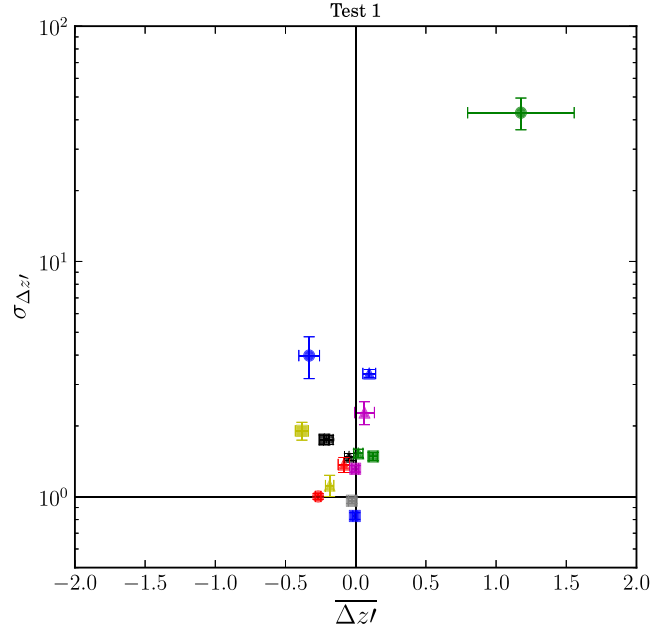


Figure 8. rms versus bias of the normalized Δz distribution $\Delta z' = \Delta z / \epsilon_{\text{phot}}$ for all the codes analysed in Test 1. Most codes underestimate ϵ_{phot} , leading to large values of $\sigma_{\Delta z'}$.

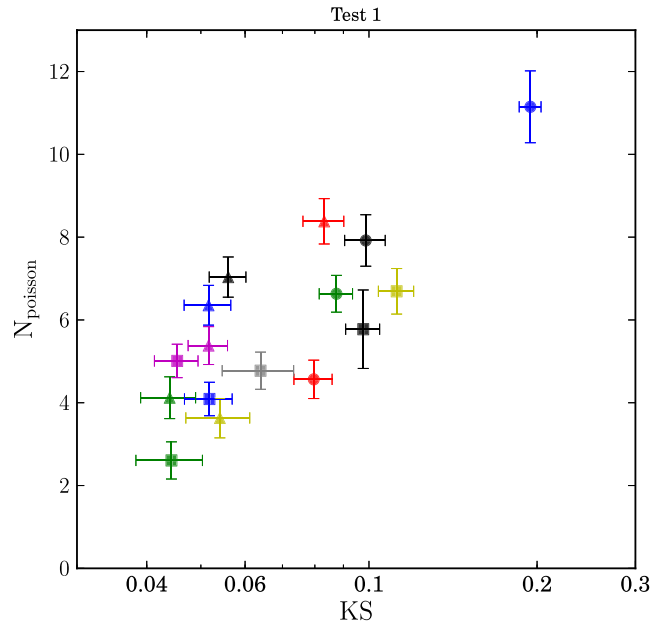


Figure 9. N_{Poisson} versus KS statistics for all the codes analysed in Test 1. Both metrics show how the true galaxy redshift distribution is reconstructed through photo- z 's, for each code. The smaller the value of the metric, the better the reconstruction. A strong correlation between the two metrics is observed, as expected.

not included in this test so less points are shown with respect to Fig. 6. In this test, basically all codes fulfil the σ_{68} requirement. In addition, some codes show important improvement when using this higher S/N data for training, such as ARBORZ. However, there is no general σ_{68} improvement as there was for Test 2, and the results are generally comparable with those from Test 1. Figs 14 and 15 show also a high degree of compatibility with Test 1, contrary to what we saw in Test 2, where larger differences were appreciated.

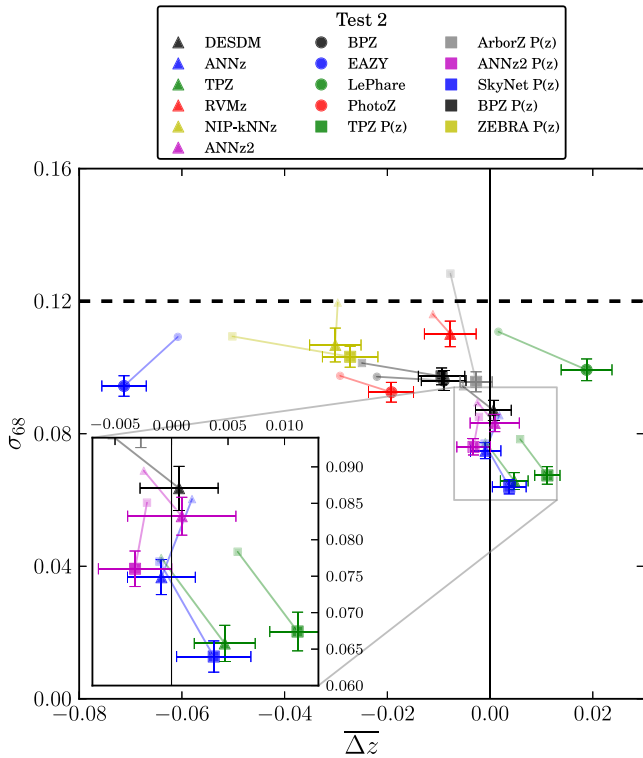


Figure 10. σ_{68} versus bias for all the codes analysed in Test 2. Results of Test 1 for each code are shown as a point without error bars (for simplicity) and connected to the result in this test through a solid line. This convention will also be used in subsequent plots for easier comparison against Test 1, which is the default case. All codes improve in photo- z precision (σ_{68}) due to the deeper photometry.

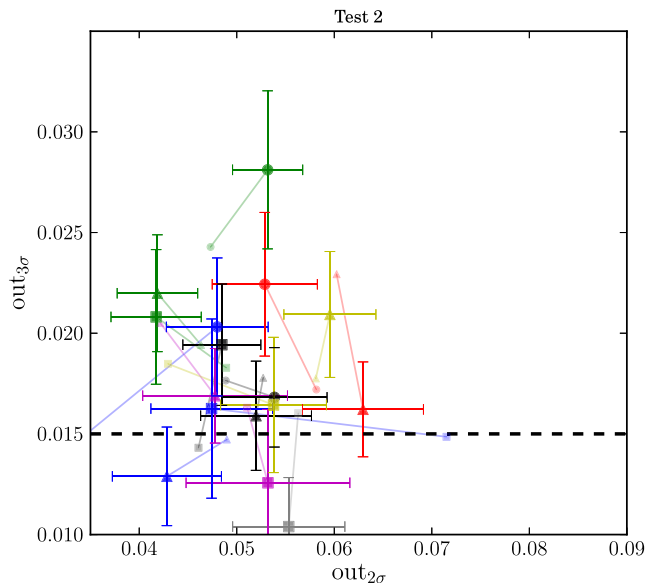


Figure 11. 3σ versus 2σ outlier fraction for all the codes analysed in Test 2. This test shows a larger scatter in the points with respect to the results in Test 1: in general, codes doing well for Test 1 improve in Test 2 while codes doing not so well for Test 1 show a worsening in Test 2.

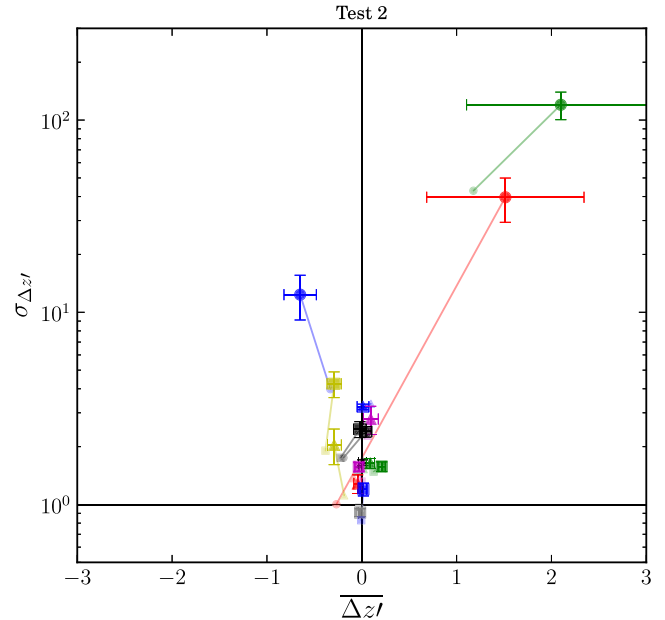


Figure 12. rms versus bias of the normalized photo- z distribution for all the codes analysed in Test 2. Here the results are, in general, close to those observed in Test 1.

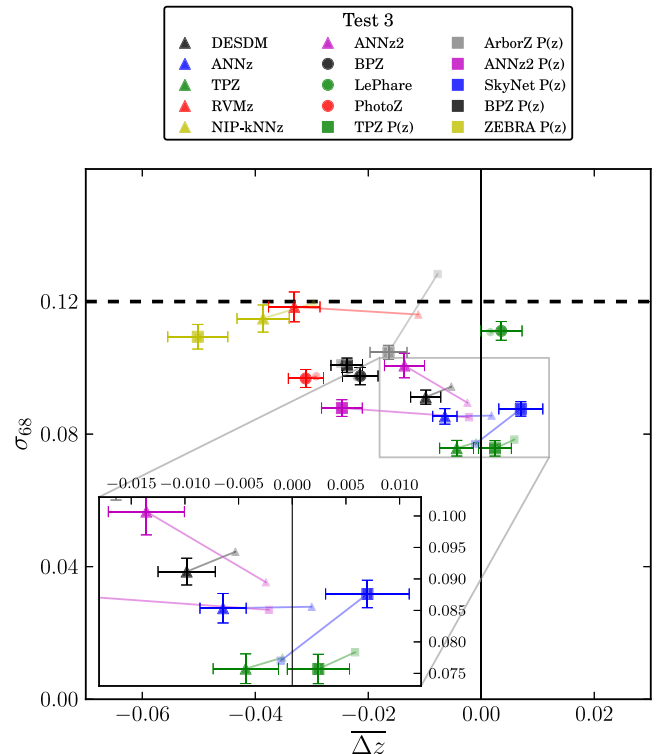


Figure 13. σ_{68} versus bias for all the codes analysed in Test 3. Results are overall compatible with Test 1 although some particular codes, such as ARBORZ, show a significant improvement by using deeper data for training.

4.2.4 Importance of the u band

In this part of the paper, we want to address the effect of adding u -band photometry to the photo- z performance. It is important to stress that a u band is available in DECam, although it is not planned to be used in the DES survey.

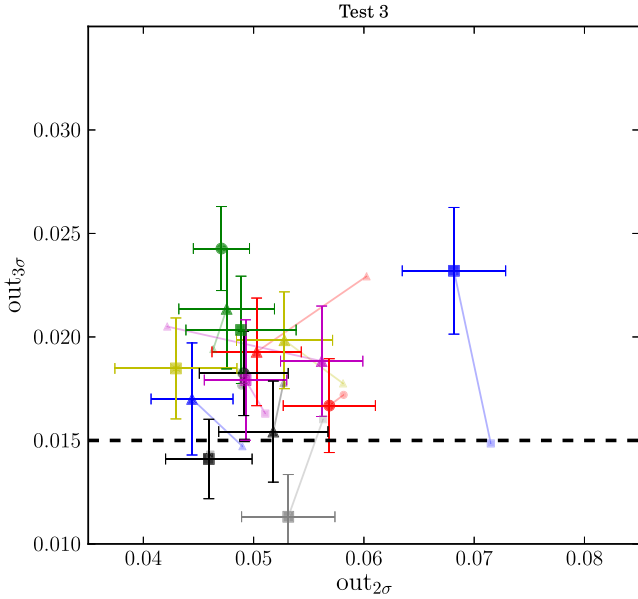


Figure 14. 3σ versus 2σ outlier fraction for all the codes analysed in Test 3. There is a general agreement with the results from Test 1.

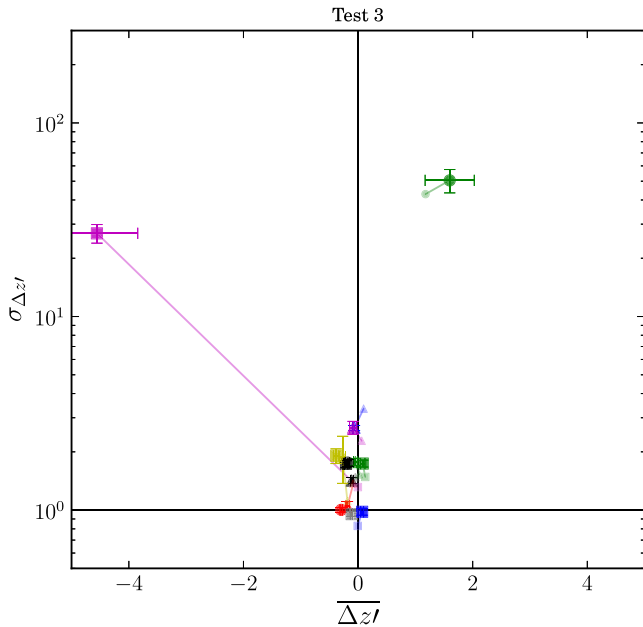


Figure 15. rms versus bias of the normalized photo- z distribution for all the codes analysed in Test 3. Again, results generally agree with Test 1.

We show the effect of incorporating the u band to Test 1. On one hand, in Fig. 16 we can appreciate this effect in the overall photo- z precision for the whole redshift range, and we clearly see an improvement on σ_{68} with respect to Test 1 (shown again as the points without error bars connected through a solid line) for most of the codes present in the plot.

On the other hand, since u is an ultraviolet band one expects it to be more relevant for low-redshift galaxies. Therefore, in Fig. 17, we show the effect of adding the u band to Test 1 in σ_{68} as a function of photometric redshift for four selected photo- z codes (the same ones we select for further analysis in Section 4.3). There we can clearly observe a significant improvement in σ_{68} at low redshift ($z_{\text{phot}} < 0.5$), while the precision at higher redshift is compatible

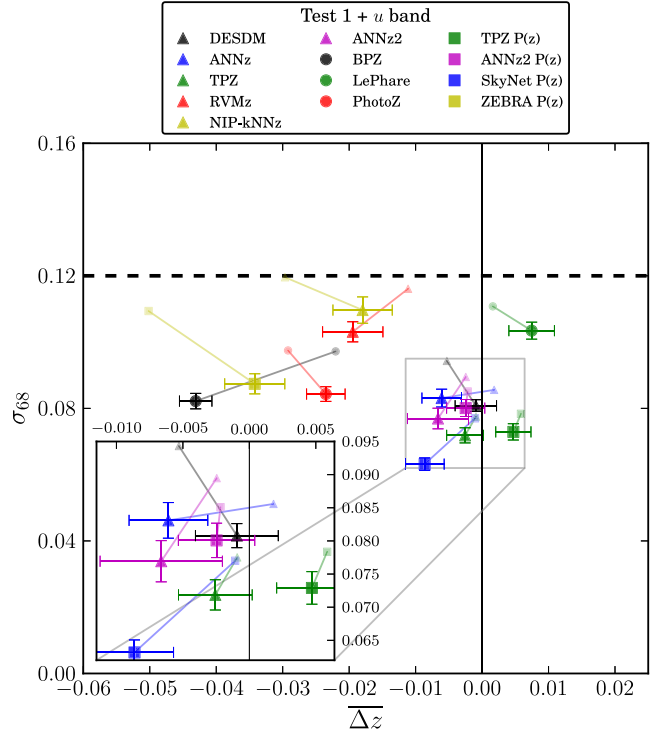


Figure 16. σ_{68} versus bias for all the codes analysed in Test 1 after the addition of the u band. We see how the incorporation of this band significantly improves the photo- z precision.

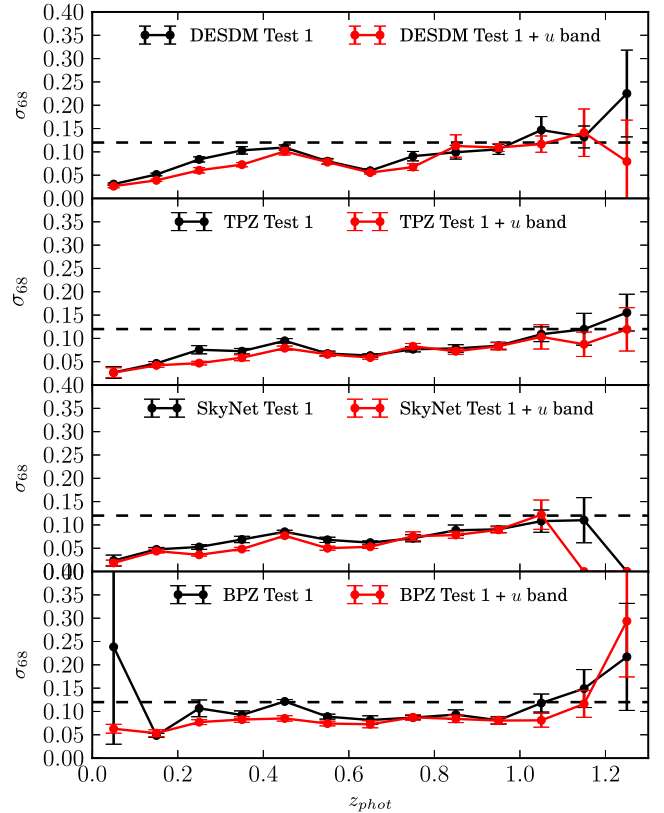


Figure 17. σ_{68} as a function of photometric redshift for four selected codes and for Test 1 + u band. As expected, the improvement in photo- z precision due to the u band addition is more important at low redshift (< 0.5).

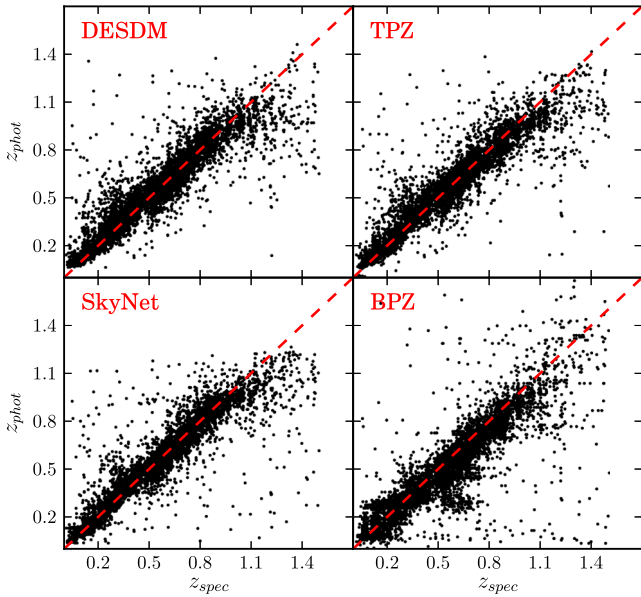


Figure 18. z_{phot} versus z_{spec} scatter plot for four selected codes for Test 1 + u band. When comparing with the corresponding plots in Fig. 5, a clear improvement at low redshift, with an important reduction in the number of outliers, can be appreciated.

within error bars in the two cases. Furthermore, in order to visually appreciate the improvement at low redshift, in Fig. 18 we show the z_{phot} versus z_{spec} scatter plot after adding the u band to Test 1 for the four same codes shown before, to be compared directly with Fig. 5.

4.2.5 Test 4: importance of different spectroscopic sets

In all previous tests, we have considered a single training sample (in two versions: Main and Deep) constructed with joint data from different spectroscopic surveys with different properties, as described in Section 3. Now we want to study the importance of the different major spectroscopic data sets used in the photo- z performance using different photo- z codes, both training and template based. In particular, this test, that we call Test 4, consists on selecting from each of the four calibration fields depicted in Fig. 1 the major spectroscopic set used in the field, then training or calibrating the algorithms using only this subset of spectroscopic objects in the training sample, and then apply the algorithms on the full testing sample, including all the spectroscopy available. In this way, we have selected spectroscopy from VVDS Deep in the SN-X3 field,

from ACES in the SN-C3 field, from VVDS Wide in the VVDS F14 field and from z COSMOS in the COSMOS field. For simplicity, we have considered only the Main training and testing samples (i.e. main survey depth photometry), so the results should be compared with Test 1.

Table 4 shows the photo- z precision (σ_{68}) for the different cases in Test 4 and for Test 1, for comparison. As a first conclusion from this table, we observe how template-based codes, such as LEPHARE, BPZ or PHOTOZ, are not very dependent on the spectroscopic data used for calibration and they show very consistent results for Test 1 and the four cases in Test 4. This is expected since these codes get the photometric information of the galaxies from a given set of pre-defined templates, either empirical or theoretical, and only use the galaxies in the training sample for calibration of the priors (see Section 4.1 for more information about calibration of the priors for each particular template-based code). On the contrary, and as a second conclusion, we observe in the table how training-based codes are dependent on the data used for their training. For this class of codes (e.g. DESDM, ANNZ, TPZ, etc.) we can observe how the Test 1 result is generally better than any of the Test 4 results, given that, in this case, all the photometric information of the galaxies comes from the training set and thus having a more complete set helps in the photo- z performance.

Additionally, there are substantial differences in the photo- z performance depending on the spectroscopic data used for training. In order to understand this, we show in Fig. 19 the i magnitude distribution for each of the four training sets used in this test. On one hand, the VVDS Deep and the ACES cases of Test 4 give the most similar results to Test 1 since VVDS Deep and ACES are the deepest spectroscopic samples, and the only ones reaching $i_{AB} = 24$ as can be appreciated in Fig. 19. On the other hand, Test 4 VVDS Wide gives the poorest results compared to Test 1 due to the fact that the VVDS-Wide spectroscopic data are much shallower ($i_{AB} < 22.5$) than DES, hence the training sample is not complete. Finally, using z COSMOS data as a training sample seems to work better than VVDS Wide but not as well as VVDS Deep and ACES spectroscopy.

4.2.6 Stability after removing the highest weight galaxies

In every application of a weighting method there exists the danger of the analysis being dependent on a few, highly weighted objects in the sample. Here we demonstrate that the conclusions of this analysis do not change after removing the highest weight galaxies. We test it only on the photo- z precision of Test 1, for simplicity.

Table 4. σ_{68} for the four cases in Test 4, corresponding to training on each of the four major spectroscopic samples and testing on the full Main testing sample. The results for Test 1 are also shown, for comparison.

Codes, σ_{68}	Test 1	Test 4 VVDS Deep	Test 4 VVDS Wide	Test 4 ACES	Test 4 z COSMOS
DESDM	0.094 ± 0.002	0.106 ± 0.002	0.139 ± 0.005	0.103 ± 0.003	0.148 ± 0.008
ANNZ	0.086 ± 0.002	0.101 ± 0.003	0.138 ± 0.004	0.091 ± 0.003	0.104 ± 0.003
TPZ	0.078 ± 0.002	0.090 ± 0.002	0.110 ± 0.005	0.093 ± 0.003	0.097 ± 0.003
NIP-KNNZ	0.120 ± 0.004	0.146 ± 0.006	0.156 ± 0.007	0.127 ± 0.005	0.148 ± 0.007
ANNZ2	0.089 ± 0.003	0.099 ± 0.003	0.143 ± 0.008	0.104 ± 0.004	0.137 ± 0.006
BPZ	0.097 ± 0.003	0.096 ± 0.002	0.095 ± 0.002	0.095 ± 0.003	0.095 ± 0.002
LEPHARE	0.111 ± 0.003	0.110 ± 0.003	0.110 ± 0.003	0.111 ± 0.003	0.112 ± 0.003
PHOTOZ	0.097 ± 0.003	0.101 ± 0.003	0.096 ± 0.002	0.096 ± 0.003	0.097 ± 0.003
TPZ P(z)	0.078 ± 0.002	0.091 ± 0.001	0.108 ± 0.004	0.093 ± 0.003	0.094 ± 0.002
BPZ P(z)	0.101 ± 0.002	0.096 ± 0.002	0.097 ± 0.002	0.097 ± 0.002	0.100 ± 0.002
ANNZ2 P(z)	0.085 ± 0.002	0.103 ± 0.003	0.217 ± 0.009	0.103 ± 0.005	0.140 ± 0.009

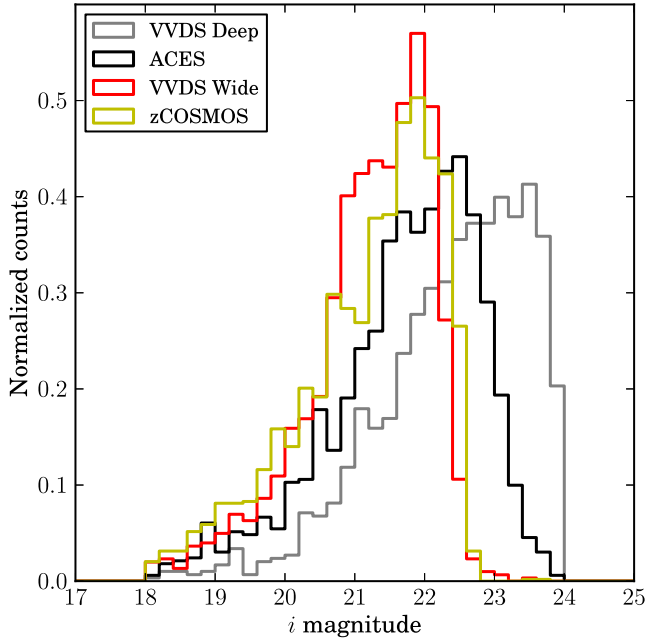


Figure 19. *i*-band magnitude distributions for the four training samples used in Test 4, each corresponding only to one of the four major spectroscopic samples used, one from each of the calibration fields.

Table 5. σ_{68} results after removing the 5 per cent of the galaxies with highest weights in Test 1. Also showing the default Test 1 for comparison.

Codes	Test 1	Test 1 cut
DESDM	0.094 ± 0.002	0.088 ± 0.002
ANNZ	0.086 ± 0.002	0.085 ± 0.002
TPZ	0.078 ± 0.002	0.073 ± 0.002
RVMZ	0.116 ± 0.004	0.103 ± 0.004
NIP-KNNZ	0.120 ± 0.004	0.110 ± 0.004
ANNZ2	0.089 ± 0.003	0.084 ± 0.003
BPZ	0.097 ± 0.003	0.095 ± 0.002
EAZY	0.109 ± 0.003	0.123 ± 0.002
LEPHARE	0.111 ± 0.003	0.109 ± 0.003
PHOTOZ	0.097 ± 0.003	0.089 ± 0.003
TPZ P(z)	0.078 ± 0.002	0.073 ± 0.002
ARBORZ P(z)	0.128 ± 0.003	0.117 ± 0.003
ANNZ2 P(z)	0.085 ± 0.002	0.082 ± 0.002
SKYNET P(z)	0.077 ± 0.002	0.067 ± 0.002
BPZ P(z)	0.101 ± 0.002	0.097 ± 0.002
ZEBRA P(z)	0.109 ± 0.004	0.100 ± 0.004

In Table 5, we show the σ_{68} results for Test 1 together with the same test after removing the 5 per cent of the galaxies with the highest weights (accounting for almost 30 per cent of the total weight) in the Main testing sample. Comparing the two results for each of the photo- z codes we clearly see an improvement in photo- z precision when the 5 per cent cut is applied, but this is expected since removing the highest weight galaxies also means removing the faintest galaxies in the catalogue, which are the most difficult galaxies to get photo- z 's of. However, this predicted improvement is both small and uniform among all codes so that the main conclusions reached in the paper remain valid.

4.3 Results for DESDM, TPZ, SKYNET and BPZ photo- z codes

So far, we have compared a large number of photo- z codes in a variety of situations and configurations. Next, we look in greater detail at four photo- z codes: DESDM, TPZ, SKYNET and BPZ. The DESDM photo- z code, a regression ANN, is integrated within the DESDM service, so its results will be made available together with all the DES data products, making it a clear choice to be studied in detail here. TPZ and SKYNET are state-of-the-art training-based methods using, respectively, random forests and ANNs to compute photo- z 's, and yielding the best performance among all the codes utilized in this analysis. Finally, BPZ is the template-based photo- z code showing best performance in the tests previously shown, and it has been widely used by other galaxy surveys such as CFHTLenS (Heymans et al. 2012; Hildebrandt et al. 2012). All these four codes are public.

In Section 4.2.1, we have studied the default configuration for this analysis, showing the most relevant quantities for all the codes and testing them against the DES requirements. However, we have not looked at the redshift dependence of these quantities, mainly due to difficulties showing that much information for a large number of codes. Fig. 20 shows the photo- z bias, precision and outlier fractions as a function of photometric redshift for the four selected codes. We can see how the four codes behave similarly for all the metrics displayed there. The σ_{68} requirement is fulfilled by the codes in most of the redshift range, except at high redshift where error bars are large due to the small number of objects. As we previously observed, the 3σ outliers fraction is the most difficult requirement to meet, although the results are close to this limit within error bars, while the 2σ outliers fraction required is met in the whole redshift range for all the codes. As mentioned at the beginning of Section 4, the requirement on the mean bias in photo- z bins of width 0.1, $|\overline{\Delta z}| < 0.001(1+z)$, is not currently being analysed, since it necessitates a larger spectroscopic sample in order to be able to

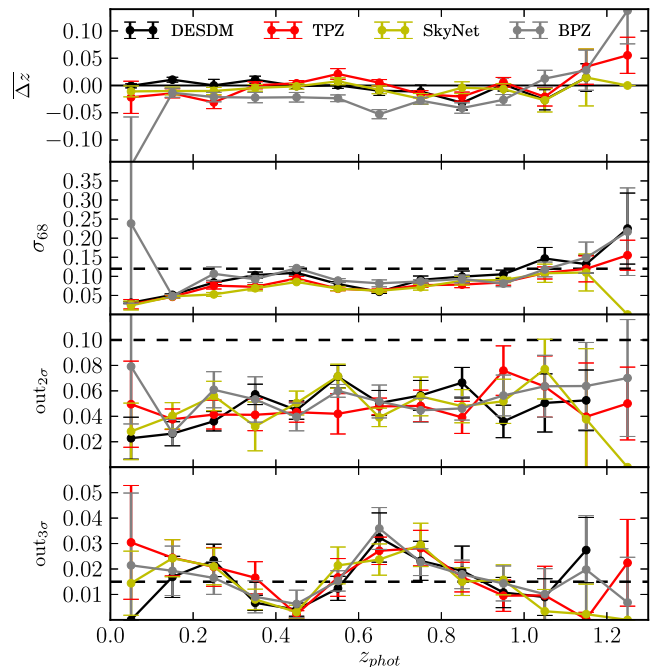


Figure 20. Results of Test 1 for the four selected photo- z codes. From top to bottom, the photo- z bias, σ_{68} and the 2σ and 3σ outlier fractions, in bins of z_{phot} . DES requirements, displayed as a black dashed line, are shown for the latter three metrics.

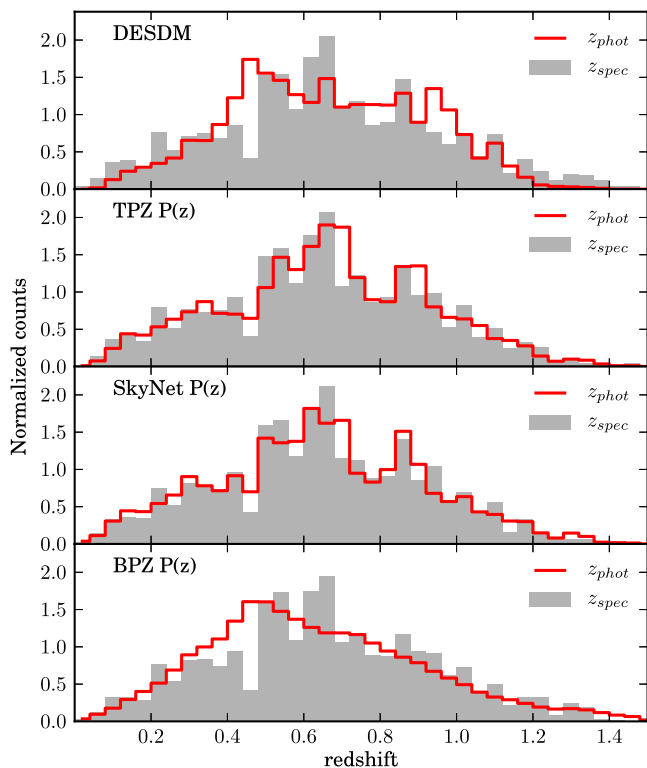


Figure 21. Full weighted spectroscopic redshift distribution and its photo- z reconstruction using the four selected codes for Test 1. TPZ, SKYNET and BPZ produce redshift pdfs for each galaxy, thus yielding smoother photo- z distributions.

calibrate the mean bias away. However, the top plot in Fig. 20 shows that for the training-based codes the mean bias in each photo- z bin is compatible with zero within errors, although the current errors are too large to assess whether the requirement is met. In some cases, the overall photo- z bias is already at the 0.001 level, as can be seen in Tables A1–A3, although again the errors are large.

A very important issue, which is actually the most important result needed from photo- z studies in order to perform many cosmological analyses, is the estimation of true redshift distributions $N(z)$. In Fig. 21, we observe how the full redshift distribution reconstructed from the four photo- z codes compares to the spectroscopic distribution. The DESDM code produces one single value for the photo- z of each galaxy in the testing sample while the other three are $P(z)$ codes, so that they return a pdf for each galaxy to be at a given redshift. This is the reason why the $N(z)$ reconstruction looks smoother for TPZ, SKYNET and BPZ, since these are computed from stacking all individual photo- z pdfs. Quantitatively, one can measure how good an $N(z)$ reconstruction is by looking at the N_{Poisson} and KS metrics in Table A1: the lower these values are, the better is the agreement between the true $N(z)$ and the photo- z -reconstructed one. As for the advantage of using $P(z)$ codes, one can observe in Table A1 how the N_{Poisson} values for TPZ and BPZ are significantly smaller in their $P(z)$ versions than in their single-estimate photo- z versions.

On the other hand, although this full redshift distribution is interesting for photo- z analyses, most of the cosmological studies split the galaxy sample into multiple photo- z bins, therefore there is a need to know the true redshift distribution inside each of those photo- z bins. Fig. 22 shows the redshift distributions, both spectroscopic and photometric, for six photo- z bins of width 0.2 from

$z = 0.1$ to 1.3, and for the four photo- z codes selected. The limited number of spectroscopic galaxies available makes the distributions shown in the figure somewhat noisy, especially in the last photo- z bin, where a very small number of galaxies is available. The third and fourth bins in photo- z are the ones presenting the narrowest spectroscopic redshift distributions, which agrees with the fact that the photo- z precision is the highest in this redshift range as can be appreciated in Fig. 20.

In Fig. 22, we observe how single-estimate photo- z codes produce a top-hat photo- z distribution for each (photo- z selected) redshift bin. In this case, depicted in the left-hand column of Fig. 22, the photometric and spectroscopic redshift distributions of each bin are very different and therefore a spectroscopic sample is needed to calibrate the broadening of the redshift bin due to photo- z errors. On the other hand, when using $P(z)$ codes to bin a sample in photometric redshift, one selects a galaxy to be inside a given redshift bin by looking at the position of the median of the pdf (other choices are also possible, e.g. the mode), checking whether it is within the boundaries of the bin and summing the full pdf of the galaxies inside, including probabilities beyond the bin limits. That makes the photo- z distribution broader than the bin limits and closer to the spectroscopic redshift distribution of the bin, as can be seen in the three rightmost columns in Fig. 22. We can see in those panels how the tails of the spectroscopic distributions are well represented by the photo- z distributions. This is an important point in favour of $P(z)$ codes since their ability to reproduce the spectroscopic redshift distribution of a photo- z selected bin by stacking their redshift pdfs makes them less dependent on a spectroscopic calibration sample.

In summary, we have characterized for each code the true redshift distribution inside each photo- z bin. This is the most important quantity to be extracted from any photo- z analysis, since it is the starting point for many cosmological studies such as galaxy clustering and weak lensing. Regarding the performance in such task, the four codes studied in this section show similar spectroscopic redshift distributions for each photo- z bin, but $P(z)$ codes are able to yield a better reconstruction of these distributions by adding up the redshift pdfs for each galaxy which makes them somewhat less reliant in the precise photo- z calibration.

5 DISCUSSION

We now discuss some of the results and implications of the analyses presented above. First, we consider the results from different types of photo- z codes, and afterwards we compare the main outcomes from this study with previous results in the literature.

The photo- z codes showing the best performance in the analysis are all training-based methods. Among them, there are various codes using ANNs in different ways and configurations (see Section 4.1), and the similarities and differences between them go beyond the network architecture. DESDM uses hyperbolic tangent activation functions in the hidden layer while ANN2 and SKYNET use a sigmoid activation function. ANN2 results are a mean of several runs where both activations functions are used in the different runs. Vanzella et al. (2004) came to the conclusion that the choice of activation between sigmoid and hyperbolic tangent functions has little effect on the photo- z performance. SKYNET was also run with rectified linear units as activation function, which have been shown to outperform more traditional activation functions in object classification (Glorot, Bordes & Bengio 2011), but no improvement in the photo- z problem was observed. SKYNET, ANN2 and ANN2 use weight regularization to avoid overfitting while all the four methods monitor the performance on a validation set to prevent overtraining.

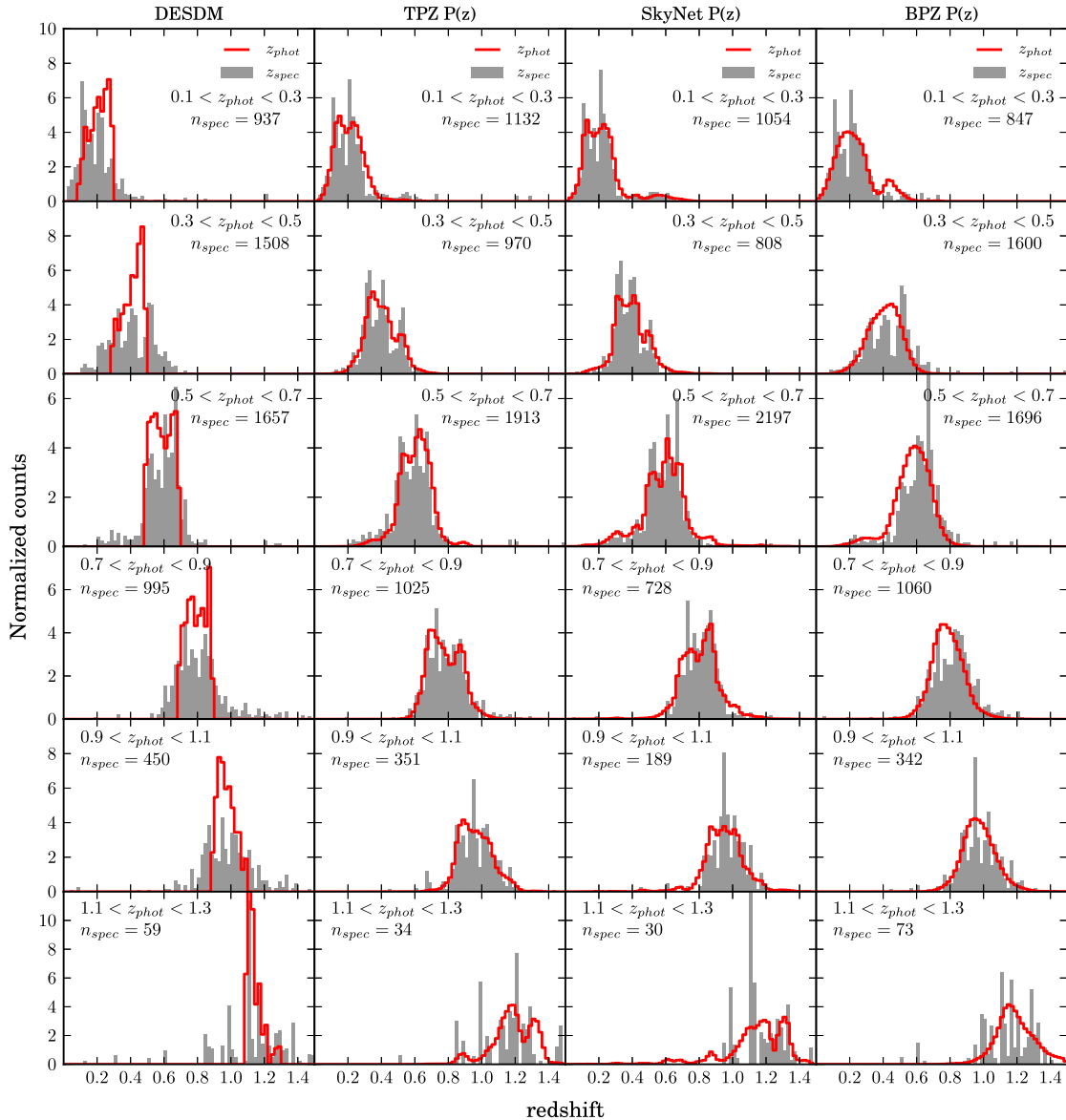


Figure 22. Weighted spectroscopic redshift distributions and their photo- z reconstruction using the four selected codes, for photo- z bins of width 0.2. The number of spectroscopic galaxies inside each photo- z bin is shown. The DESDM is a single-estimate photo- z code, while TPZ, SKYNET and BPZ are $P(z)$ codes. This is the reason why the photo- z distributions returned by the latter codes can reconstruct the tails of the spectroscopic distributions beyond the photo- z bins. The photo- z bins are defined using the best-estimate z_{phot} for each code, while, for TPZ, SKYNET and BPZ the reconstructed redshift distributions are obtained by stacking the pdfs for each galaxy.

DESDM and ANNZ use first-order gradient information while SKYNET and ANNZ2 also use second-order derivative information to train the network. SKYNET and ANNZ2 are the only networks adding a constraint to the outputs: ANNZ2 does this on a redshift bin per redshift bin basis while SKYNET uses a *softmax* transformation in the final layer, adding a constraint on all redshift bins simultaneously. In conclusion, the neural networks with regularization perform better than the unregularized DESDM network while the fact that SKYNET uses a *softmax* constrained output in combination with a back-propagation algorithm that uses second-order derivative information seems to give it the edge over the rest of ANNs.

Aside from ANNs, TPZ, which is a state-of-the-art photo- z code using prediction trees and random forests, performs remarkably well in all the tests in this work. The prediction trees and random forest techniques used by TPZ have the advantage that they

have fewer hyper parameters to be chosen compared to neural networks. Neural networks have, amongst others, to choose the amount of hidden layers, the amount of nodes per hidden layer, the learning rate and at least one regularization parameter if present. Random forests used in TPZ have only two hyper parameters to choose: the amount of trees used and the size of the subsample set of features used at each split. This leaves out the choice of activation function in neural networks and the choice of the measure of information gain at each split in random forests, maximizing its performance.

Furthermore, training-based photo- z codes show lower bias compared to that of template-based codes, which indicates possible systematic inaccuracies in the template sets. This can be solved by using adaptive recalibration procedures, which adjust the zero-point offsets in each band using the training sample. Such technique has

been successfully applied by LEPHARE in this work, as was also the case in Hildebrandt et al. (2010).

The photo-*z* precision values obtained by template-based methods in this study are all compatible with each other within 10 per cent. BPZ and PHOTOZ yield the highest photo-*z* precision among the methods of this class. The reason for this is probably not the template sets they use, which for BPZ is a combination of Coleman et al. (1980) and Kinney et al. (1996), and a more complex combination of templates for PHOTOZ (see Section 4.1), since other codes include similar libraries. The fact that they both use Bayesian priors calibrated on DES data, and not only on previous data sets, can be relevant here, although this makes the result more dependent on how representative the training sample is of the full DES data sample.

In order to set up a context for DES photo-*z* results, it is worth checking the performance obtained by previous similar surveys. Of particular interest is the comparison to the CFHTLS and CFHTLenS (which are two different reductions of the same survey), photo-*z* results (Coupon et al. 2009; Hildebrandt et al. 2012), due to the similarities between CFHTLS and DES in terms of survey specifications. However, it is important to stress the differences between DES and CFHTLS. For instance, CFHTLS uses deep *u*-band photometry by default, although they also test the importance of this band by computing their photo-*z*'s without it (Ilbert et al. 2006). They find a clear degradation in their photo-*z* precision at $z_{\text{spec}} < 0.4$, compatible with our findings. Another difference comes from the fact that they do not apply any weighting technique to their calibration sample, thus leaving room for discrepancies between their calibration and full samples. In addition, we should point out that their spectroscopic samples for training/calibration are much larger than the ones used here by about a factor of 5. Finally, their photometry ($S/N \simeq 11$ at $i_{AB} = 24$) is deeper than the Main reduction in this paper, being close to our Deep sample (see Fig. 2). In Coupon et al. (2009), CFHTLS photo-*z*'s show a precision of about $\sigma_{68} \simeq 0.085$ (roughly translating their result to the metric used here) for *ugriz* photometry in galaxies up to $i_{AB} < 24$. So, despite the differences stated above, their value of the photo-*z* precision is at the level of the DES photo-*z* precision shown in this paper. Moreover, this level of precision is obtained already in Test 1, which uses slightly shallower photometry and lacks the *u* band. Even so, the differences in photo-*z* precision obtained by different algorithms might be relevant here. In particular, the results reported in Coupon et al. (2009) used a template-based photo-*z* code, while the most precise methods reported in this study come from training-based codes. If we do the comparison between template-based codes, the results in this study are close but do not exceed the precision reported by CFHTLS.

In terms of outlier fraction, Coupon et al. (2009) report a value for the fraction of galaxies with $|\Delta z / (1 + z)| > 0.15$ of $\eta = 10.1$ per cent. In order to enable the comparison, since the metric used in this paper differs from their approach, we have computed η for two cases in this study. On the one hand, we have estimated η for a template-based photo-*z* code (BPZ), trying to make a fair comparison since this is the class of algorithm used in Coupon et al. (2009). In this case, we obtain $\eta = 10.0$ per cent, in perfect agreement with their result. On the other hand, computing this outlier rate for a template-based code (DESDM) yields $\eta = 6.1$ per cent, significantly improving the template-based code result.

We conclude that the overall photo-*z* results shown in this paper are within the expectations for DES if we take into account the performance reported by previous comparable studies for a survey of similar characteristics.

6 SUMMARY AND CONCLUSIONS

The DES successfully installed DECam during the second semester of 2012, starting its operations in 2012 November with a SV period lasting until 2013 February. Among the 150 sq. deg. covered by SV observations, four different fields, of about 3 sq. deg. each, overlap with areas with substantial spectroscopic coverage such as VVDS 02hr or COSMOS. Using *grizY* photometry for galaxies matched to the existing spectroscopic data in these four calibration fields, this paper presents the photometric redshift performance of the DES survey in the SV period. Most of the relevant photo-*z* codes have been used in the analysis.

Since spectroscopic galaxy samples are generally shallower, a weighting technique is used to make the calibration sample of galaxies to mimic the DES full sample in magnitude and colour space in order to properly estimate the photo-*z* performance in the DES galaxy sample.

Calibration and testing samples have been produced with two different depths: Main is the default depth in the DES survey, and Deep corresponds to the depth in SNe fields. Test 1, which uses the Main training and testing samples, represents the default case for photo-*z* estimation in DES. Results from 13 different codes are analysed in this case, showing fluctuations in photo-*z* performance but a general agreement in codes of the same type (machine-learning or template-fitting algorithms). In particular, most of the codes analysed comfortably meet the DES science requirements in terms of photo-*z* precision and several also meet the requirements on the fractions of outliers.

In Test 2, we explore the impact of deeper and higher S/N photometry in photo-*z* calculations, showing that all the codes used improve their results significantly, as expected. In Test 3, we explore the possibility of using deeper photometry only for training/calibration of the algorithms. In this case, we see no general improvement, although there is a significant enhancement using specific codes.

In an additional test, we consider the incorporation of the *u* band, which is available in DECam but not used in the DES survey, demonstrating a general improvement in photo-*z* precision, particularly at low redshift (< 0.5). This is expected since the *u* band is crucial for the filter set to bracket the 4000 Å break in low-redshift galaxies. However, due to the high mean redshift of the full DES sample, the impact of not using that band in the overall photo-*z* precision is less important than in previous, shallower surveys such as SDSS. Moreover, we study the importance of the different spectroscopic data sets used, showing how the sets spanning the whole photometric space are crucial for training-based methods, and demonstrated that the results are stable under the removal of the galaxies with highest weights from the analysis.

Generally, training-based photo-*z* codes show the best performance in the tests in terms of photo-*z* precision and bias. Among them, TPZ, using prediction trees and random forest, and SKYNET, a state-of-the-art ANN application, seem to yield the most accurate results, achieving a core photometric redshift resolution below $\sigma_{68} = 0.08$. The fact that these two new codes perform better than others extensively used in the literature shows how there is room for improvement in the photo-*z* industry. On the other hand, all template-fitting methods employed show consistent results between them, although the use of Bayesian priors specifically calibrated on DES data and adaptive template-recalibration procedures appears to help significantly.

Finally, in the last part of the paper, we choose four photo-*z* codes, representing different techniques and types, and we present a more detailed analysis of their results. We show some of the most

relevant metrics as a function of redshift and, most importantly, we study the estimation of the true redshift distributions $N(z)$ computed using photo- z 's. For these four codes, we obtain the true redshift distributions in six photo- z bins. Fig. 21 shows the DES photo- z capabilities in such crucial task, and demonstrates the ability to split the DES full sample in tomographic redshift bins of width 0.2 already with these early data. Furthermore, the calibration of the true redshift distribution of a photo- z selection is the most important ingredient for cosmological studies involving galaxy clustering or weak lensing, and it is an important outcome from this paper, enabling further science analyses.

The photo- z analyses carried out in this work using these early stage DES data will serve as a benchmark for future data releases, and as the survey area grows during the observation period, more spectroscopic data will be available allowing a better calibration and a better sampling for training algorithms. Therefore, these promising early results will do nothing but improve in the near future, which will allow putting tighter constraints on several cosmological parameters. Furthermore, the five band optical and near-infrared photometry of DES can be combined with the infrared J and K_s photometry provided by the VHS survey (McMahon et al. 2013) in 90 per cent of the DES footprint. This should result in improved photometric redshift estimations, particularly at high redshift (Banerji et al. 2008).

ACKNOWLEDGEMENTS

This paper has gone through internal review by the DES collaboration.

We are grateful for the extraordinary contributions of our CTIO colleagues and the DES Camera, Commissioning and Science Verification teams in achieving the excellent instrument and telescope conditions that have made this work possible. The success of this project also relies critically on the expertise and dedication of the DES Data Management organization.

Funding for the DES Projects has been provided by the US Department of Energy, the US National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência e Tecnologia, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratories, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energeticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Edinburgh, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, the Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National

Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University.

The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2009-13936, AYA2012-39559, AYA2012-39620, and FPA2012-39684, which include FEDER funds from the European Union. J.G acknowledges a fellowship from DS-CAPES Programme. This work was partially supported by STFC grant ST/K00090X/1. Contact the author(s) to request access to research materials discussed in this paper.

This research uses data from the VIMOS VLT Deep Survey, obtained from the VVDS data base operated by Cesam, Laboratoire d'Astrophysique de Marseille, France.

Data for the OzDES spectroscopic survey were obtained with the Anglo-Australian Telescope (programme numbers 12B/11 and 13B/12). Parts of this research were conducted by the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the US Department of Energy Office of Science. The SDSS-III website 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 Astrofísica 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.

REFERENCES

- Abdalla F. B., Banerji M., Lahav O., Rashkov V., 2011, *MNRAS*, 417, 1891
 Ahn C. P. et al., 2012, *ApJS*, 203, 21
 Arnouts S. et al., 2002, *MNRAS*, 329, 355
 Banerji M., Abdalla F. B., Lahav O., Lin H., 2008, *MNRAS*, 386, 1219
 Bender R. et al., 2001, in Cris S., Renzini A., Williams R. E., eds, *Deep Fields*. Springer-Verlag, Berlin, p. 96
 Benitez N., 2000, *ApJ*, 536, 571
 Bertin E., 2006, in Gabriel C., Arviset C., Ponz D., Solano E., eds, *ASP Conf. Ser. Vol. 351, Astronomical Data Analysis Software and Systems XV*. Astron. Soc. Pac., San Francisco, p. 112
 Bertin E., 2011, in Evans I. N., Accoma A., eds, *ASP Conf. Ser. 442, Astronomical Data Analysis Software and Systems XX*. Astron. Soc. Pac., San Francisco, p. 435
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
 Bertin E., Mellier Y., Radovich M., Missonnier G., Didelon P., Morin B., 2002, in Bohlender D. A., Durand D., Handley T. H., eds, *ASP Conf. Ser. Vol. 281, Astronomical Data Analysis Software and Systems XI*. Astron. Soc. Pac., San Francisco, p. 228
 Bolzonella M., Miralles J., Pell R., 2000, *A&A*, 492, 476
 Bonnett C., 2013, preprint ([arXiv:1312.1287](https://arxiv.org/abs/1312.1287))
 Bordoloi R., Lilly S. J., Amara A., 2010, *MNRAS*, 406, 881
 Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, 686, 1503
 Brimiouille F., Seitz S., Lerchster M., Bender R., Snigula J., 2013, *MNRAS*, 432, 1046

Brodrin M., Lilly S. J., Porciani C., McCracken H. J., Le Fèvre O., Foucaud S., Crampton D., Mellier Y., 2006, *ApJS*, 162, 20

Bruzual A. G., Charlot S., 1993, *ApJ*, 405, 538

Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000

Carrasco Kind M., Brunner R. J., 2013, *MNRAS*, 432, 1483

Carrasco Kind M., Brunner R. J., 2014, *MNRAS*, 442, 3380

Coe D., Benítez N., Sánchez S. F., Jee M., Bouwens R., Ford H., 2006, *AJ*, 132, 926

Coleman G. D., Wu C.-C., Weedman D. W., 1980, *ApJS*, 43, 393

Colless M. et al., 2001, *MNRAS*, 328, 1039

Collister A. A., Lahav O., 2004, *PASP*, 116, 345

Cooper M. C. et al., 2012, *MNRAS*, 425, 2116

Coupon J. et al., 2009, *A&A*, 500, 981

Cunha C. E., Lima M., Oyaizu H., Frieman J., Lin H., 2009, *MNRAS*, 396, 2379

Cunha C. E., Huterer D., Busha M. T., Wechsler R. H., 2012a, *MNRAS*, 423, 909

Cunha C. E., Huterer D., Lin H., Busha M. T., Wechsler R. H., 2012b, preprint (arXiv:1207.3347v1)

Dahlen T. et al., 2013, *ApJ*, 775, 93

Dawson K. S. et al., 2013, *AJ*, 145, 10

de Jong J. T. A. et al., 2013, *ESO Messenger*, 154, 44

De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2

Desai S. et al., 2012, *ApJ*, 757, 83

Diehl H. T. et al., 2012, *Phys. Procedia*, 37, 1332

Drinkwater M. J. et al., 2010, *MNRAS*, 401, 1429

Eisenstein D. J. et al., 2001, *AJ*, 122, 2267

Feldmann R. et al., 2006, *MNRAS*, 372, 565

Flaugher B., 2005, *Int. J. Mod. Phys.*, A20, 3121

Flaugher B. L. et al., 2012, *Proc. SPIE*, 8446, 844611

Freund Y., Schapire R. E., 1997, *J. Comput. Syst. Sci.*, 55, 119

Garilli B. et al., 2008, *A&A*, 486, 683

Gerdes D. W., Sypniewski A. J., McKay T. A., Hao J., Weis M. R., Wechsler R. H., Busha M. T., 2010, *ApJ*, 715, 823

Glorot X., Bordes A., Bengio Y., 2011, in *JMLR W&CP Proc. Fourteenth Int. Conf. Artif. Intell. Stat. (AISTATS 2011)*, Fort Lauderdale, FL, USA

Graff P., Feroz F., Hobson M. P., Lasenby A. N., 2013, preprint (arXiv:1309.0790)

Greisel N., Seitz S., Drory N., Bender R., Saglia R. P., Snigula J., 2013, *ApJ*, 768, 117

Heymans C. et al., 2012, *MNRAS*, 427, 146

Hildebrandt H., Wolf C., Benítez N., 2008, *A&A*, 480, 703

Hildebrandt H. et al., 2010, *A&A*, 523, A31

Hildebrandt H. et al., 2012, *MNRAS*, 421, 2355

Hogg D. W. et al., 1998, *AJ*, 115, 1418

Ilbert O. et al., 2006, *A&A*, 457, 841

Kaiser N., Tonry J. L., Luppino G. A., 2000, *PASP*, 112, 768

Karatzoglou A., Smola A., Hornik K., Zeileis A., 2004, *J. Stat. Softw.*, 11, 1

Kinney A. L., Calzetti D., Bohlin R. C., McQuade K., Storchi-Bergmann T., Schmitt H. R., 1996, *ApJ*, 467, 38

Kuehn K. et al., 2014, *AAO Observer*, 125, 4

Le Fèvre O. et al., 2003, in Iye M., Moorwood A. F. M., eds, *SPIE Conf. Ser. Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*. SPIE, Bellingham, p. 1670

Le Fèvre O. et al., 2005, *A&A*, 439, 845

Le Fèvre O. et al., 2013, *A&A*, 559, A14

Lilly S. J. et al., 2007, *ApJS*, 172, 70

Lilly S. J. et al., 2009, *ApJS*, 184, 218

Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, *MNRAS*, 390, 118

McMahon R. G., Banerji M., Gonzalez E., Kuposov S. E., Bejar V. J., Lodieu N., Rebolo R., 2013, *The Messenger*, 154, 35

Mannucci F., Basile F., Poggianti B., Cimatti A., Daddi E., Pozzetti L., Vanzani L., 2001, *MNRAS*, 326, 745

Marti P., Miquel R., Bauer A., Gaztanaga E., 2014, *MNRAS*, 437, 3490

Mohr J. J. et al., 2012, *Proc. SPIE*, 8451, 8451

Oyaizu H., Lima M., Cunha C. E., Lin H., Frieman J., Sheldon E. S., 2008a, *ApJ*, 674, 768

Oyaizu H., Lima M., Cunha C. E., Lin H., Frieman J., 2008b, *ApJ*, 689, 709

Pyle D., 1999, *Data Preparation for Data Mining*. Morgan Kaufmann Publishers Inc., San Francisco, California

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525

Sérsic J. L., 1963, *Bol. Asoc. Argentina Astron.*, 6, 41

Strauss M. A. et al., 2002, *AJ*, 124, 1810

Sypniewski A., 2014, PhD thesis, Univ. Michigan

Sypniewski A. J., Gerdes D. W., 2013, *ArborZ home page* (<https://cdcv.fnal.gov/redmine/projects/arborz>)

Tipping M. E., 2001, *J. Mach. Learn. Res.*, 1, 211

Tyson J. A., Wittman D. M., Hennawi J. F., Spergel D. N., 2003, *Nucl. Phys. B - Proc. Suppl.*, 124, 21

Vanzella E. et al., 2004, *A&A*, 423, 761

York D. G. et al., 2000, *AJ*, 120, 1579

APPENDIX A

A1 Description of the metrics

Here we describe the metrics used throughout this work. For each photo-*z* code and each galaxy, we have either the photo-*z* estimation and its associated error or a pdf $P(z)$. As described in the text, a vector of weights were computed in order to match the spectroscopic and photometric samples in multicolour and magnitude space. In each sample, we have a vector ω of weights corresponding to the N galaxies on each test set, where $\sum_{i=1}^N \omega_i = 1$. If no weights are used, the default value $\omega_i = \frac{1}{N}$ is assigned to each galaxy. We define the individual bias as $\Delta z_i = z_{\text{phot}, i} - z_{\text{spec}, i}$ and the statistics used in this work as follows:

(i) mean bias ($\overline{\Delta z}$):

$$\overline{\Delta z} = \frac{\sum \omega_i \Delta z_i}{\sum \omega_i}, \quad (\text{A1})$$

(ii) $\sigma_{\Delta z}$:

$$\sigma_{\Delta z} = \left(\frac{\sum \omega_i (\Delta z_i - \overline{\Delta z})^2}{\sum \omega_i} \right)^{1/2}, \quad (\text{A2})$$

(iii) median (Δz_{50}), the median of the Δz distribution, fulfilling

$$P_{50} = P(\Delta z \leq \Delta z_{50}) = \int_0^{\Delta z_{50}} \omega(\Delta z) d(\Delta z) = \frac{1}{2}, \quad (\text{A3})$$

(iv) σ_{68} , half of the width of the distribution, measured with respect to the median, where 68 per cent of the data are enclosed. This is computed as

$$\sigma_{68} = \frac{1}{2} (P_{84} - P_{16}) \quad (\text{A4})$$

(v) $\text{out}_{2\sigma}$, the fraction of outliers above the $2\sigma_{\Delta z}$ level:

$$\text{out}_{2\sigma} = \frac{\sum W_i}{\sum \omega_i}, \quad (\text{A5})$$

where

$$W_i = \begin{cases} \omega_i, & \text{if } |\Delta z_i - \overline{\Delta z}| > 2\sigma_{\Delta z} \\ 0, & \text{if } |\Delta z_i - \overline{\Delta z}| \leq 2\sigma_{\Delta z}, \end{cases}$$

(vi) $\text{out}_{3\sigma}$, the fraction of outliers above the $3\sigma_{\Delta z}$ level:

$$\text{out}_{3\sigma} = \frac{\sum W_i}{\sum \omega_i}, \quad (\text{A6})$$

where

$$W_i = \begin{cases} \omega_i, & \text{if } |\Delta z_i - \overline{\Delta z}| > 3\sigma_{\Delta z} \\ 0, & \text{if } |\Delta z_i - \overline{\Delta z}| \leq 3\sigma_{\Delta z}, \end{cases}$$

Table A1. Results of all the photo- z metrics listed in Appendix A for all the codes analysed in Test 1. The errors are computed from bootstrap resampling with 100 samples. The weighting procedure has been applied, together with a cut on the 10 per cent of the galaxies with the highest estimated photo- z errors for each code.

Test 1	Δz	Δz_{50}	σ_{68}	$\sigma_{\Delta z}$	out $_{2\sigma}$	out $_{3\sigma}$	$\Delta z'$	$\sigma_{\Delta z'}$	N_{Poisson}	KS
DESDM	-0.005 ± 0.003	-0.003 ± 0.002	0.094 ± 0.002	0.135 ± 0.005	0.053 ± 0.005	0.018 ± 0.003	-0.047 ± 0.032	1.479 ± 0.052	7.035 ± 0.486	0.056 ± 0.004
ANNZ	0.002 ± 0.003	-0.001 ± 0.003	0.086 ± 0.002	0.118 ± 0.004	0.049 ± 0.004	0.015 ± 0.002	0.096 ± 0.046	3.341 ± 0.134	6.355 ± 0.480	0.052 ± 0.005
TPZ	-0.001 ± 0.003	0.004 ± 0.002	0.078 ± 0.002	0.122 ± 0.006	0.046 ± 0.004	0.019 ± 0.002	0.019 ± 0.032	1.529 ± 0.063	4.122 ± 0.505	0.044 ± 0.005
RVMZ	-0.011 ± 0.005	-0.004 ± 0.002	0.116 ± 0.004	0.180 ± 0.008	0.060 ± 0.005	0.023 ± 0.003	-0.084 ± 0.041	1.371 ± 0.098	8.382 ± 0.548	0.083 ± 0.007
NIP-KNNZ	-0.030 ± 0.005	-0.011 ± 0.002	0.120 ± 0.004	0.197 ± 0.009	0.058 ± 0.005	0.018 ± 0.003	-0.186 ± 0.030	1.116 ± 0.117	3.633 ± 0.482	0.054 ± 0.007
ANNZ2	-0.002 ± 0.003	-0.003 ± 0.002	0.089 ± 0.003	0.151 ± 0.009	0.042 ± 0.003	0.021 ± 0.002	0.063 ± 0.071	2.280 ± 0.255	5.381 ± 0.458	0.052 ± 0.004
BFZ	-0.022 ± 0.003	-0.021 ± 0.002	0.097 ± 0.003	0.137 ± 0.006	0.049 ± 0.003	0.018 ± 0.002	-0.194 ± 0.032	1.750 ± 0.075	7.919 ± 0.622	0.099 ± 0.008
EAZY	-0.061 ± 0.004	-0.063 ± 0.003	0.109 ± 0.003	0.153 ± 0.010	0.035 ± 0.005	0.015 ± 0.002	-0.331 ± 0.074	3.982 ± 0.804	11.148 ± 0.868	0.195 ± 0.009
LEPHARE	0.002 ± 0.004	-0.007 ± 0.003	0.111 ± 0.003	0.171 ± 0.008	0.047 ± 0.002	0.024 ± 0.002	1.177 ± 0.379	42.883 ± 6.603	6.632 ± 0.444	0.087 ± 0.006
PHOTOZ	-0.029 ± 0.003	-0.029 ± 0.002	0.097 ± 0.003	0.142 ± 0.006	0.058 ± 0.004	0.017 ± 0.002	-0.268 ± 0.020	1.003 ± 0.030	4.565 ± 0.464	0.080 ± 0.006
TPZ P(z)	0.006 ± 0.003	0.011 ± 0.002	0.078 ± 0.002	0.119 ± 0.006	0.049 ± 0.005	0.018 ± 0.003	0.125 ± 0.030	1.484 ± 0.059	2.607 ± 0.449	0.044 ± 0.006
ARBORZ P(z)	-0.008 ± 0.004	0.001 ± 0.004	0.128 ± 0.003	0.153 ± 0.005	0.056 ± 0.005	0.016 ± 0.003	-0.028 ± 0.024	0.962 ± 0.025	4.774 ± 0.449	0.064 ± 0.009
ANNZ2 P(z)	-0.002 ± 0.002	-0.002 ± 0.002	0.085 ± 0.002	0.118 ± 0.004	0.051 ± 0.004	0.016 ± 0.003	-0.004 ± 0.024	1.315 ± 0.042	5.010 ± 0.404	0.045 ± 0.004
SKYNET P(z)	-0.001 ± 0.002	0.001 ± 0.002	0.077 ± 0.002	0.104 ± 0.003	0.072 ± 0.006	0.015 ± 0.002	-0.006 ± 0.014	0.829 ± 0.027	4.091 ± 0.404	0.052 ± 0.005
BFZ P(z)	-0.025 ± 0.003	-0.025 ± 0.003	0.101 ± 0.002	0.132 ± 0.006	0.046 ± 0.004	0.014 ± 0.002	-0.224 ± 0.033	1.750 ± 0.080	5.776 ± 0.948	0.098 ± 0.007
ZEBRA P(z)	-0.050 ± 0.005	-0.030 ± 0.002	0.109 ± 0.004	0.177 ± 0.012	0.043 ± 0.006	0.018 ± 0.002	-0.383 ± 0.047	1.906 ± 0.166	6.692 ± 0.550	0.112 ± 0.008

Table A2. Results of all the photo- z metrics for all the codes analysed in Test 2. The errors are computed from bootstrap resampling with 100 samples. The weighting procedure has been applied, together with a cut on the 10 per cent of the galaxies with the highest estimated photo- z errors for each code.

Test 2	Δz	Δz_{50}	σ_{68}	$\sigma_{\Delta z}$	out $_{2\sigma}$	out $_{3\sigma}$	$\Delta z'$	$\sigma_{\Delta z'}$	N_{Poisson}	KS
DESDM	0.001 ± 0.003	-0.002 ± 0.002	0.087 ± 0.003	0.123 ± 0.005	0.052 ± 0.006	0.016 ± 0.003	0.008 ± 0.041	1.648 ± 0.059	8.822 ± 0.716	0.058 ± 0.008
ANNZ	-0.001 ± 0.003	-0.003 ± 0.002	0.075 ± 0.002	0.105 ± 0.007	0.043 ± 0.006	0.013 ± 0.002	0.013 ± 0.062	3.228 ± 0.099	7.692 ± 0.679	0.057 ± 0.006
TPZ	0.005 ± 0.003	0.004 ± 0.002	0.066 ± 0.002	0.101 ± 0.005	0.042 ± 0.004	0.022 ± 0.003	0.091 ± 0.042	1.637 ± 0.099	4.123 ± 0.483	0.045 ± 0.007
RVMZ	-0.008 ± 0.005	-0.005 ± 0.003	0.110 ± 0.004	0.168 ± 0.008	0.063 ± 0.006	0.016 ± 0.002	-0.043 ± 0.041	1.283 ± 0.144	10.166 ± 0.750	0.079 ± 0.008
NIP-KNNZ	-0.030 ± 0.005	-0.013 ± 0.003	0.107 ± 0.005	0.169 ± 0.009	0.060 ± 0.005	0.021 ± 0.003	-0.291 ± 0.076	2.040 ± 0.429	4.463 ± 0.652	0.054 ± 0.007
ANNZ2	0.001 ± 0.005	-0.005 ± 0.002	0.083 ± 0.003	0.140 ± 0.010	0.048 ± 0.007	0.017 ± 0.002	0.094 ± 0.080	2.784 ± 0.459	7.737 ± 0.701	0.051 ± 0.006
BFZ	-0.009 ± 0.004	-0.015 ± 0.002	0.096 ± 0.003	0.143 ± 0.007	0.054 ± 0.005	0.017 ± 0.002	0.044 ± 0.062	2.400 ± 0.160	8.205 ± 0.750	0.053 ± 0.005
EAZY	-0.071 ± 0.004	-0.069 ± 0.003	0.094 ± 0.003	0.155 ± 0.010	0.048 ± 0.005	0.020 ± 0.003	-0.650 ± 0.171	12.351 ± 3.233	11.076 ± 0.839	0.210 ± 0.010
LEPHARE	0.019 ± 0.005	-0.008 ± 0.002	0.099 ± 0.003	0.183 ± 0.010	0.053 ± 0.004	0.028 ± 0.004	2.099 ± 0.994	120.100 ± 19.658	7.509 ± 1.169	0.071 ± 0.007
PHOTOZ	-0.019 ± 0.004	-0.029 ± 0.003	0.093 ± 0.003	0.147 ± 0.007	0.053 ± 0.005	0.022 ± 0.004	1.514 ± 0.829	39.656 ± 10.272	4.932 ± 0.601	0.054 ± 0.007
TPZ P(z)	0.011 ± 0.002	0.011 ± 0.002	0.067 ± 0.003	0.099 ± 0.004	0.042 ± 0.005	0.021 ± 0.003	0.210 ± 0.038	1.576 ± 0.090	3.702 ± 0.451	0.051 ± 0.006
ARBORZ P(z)	-0.003 ± 0.003	0.001 ± 0.002	0.096 ± 0.003	0.119 ± 0.006	0.055 ± 0.006	0.010 ± 0.002	-0.017 ± 0.023	0.907 ± 0.043	5.470 ± 0.499	0.055 ± 0.004
ANNZ2 P(z)	-0.003 ± 0.003	-0.003 ± 0.002	0.076 ± 0.002	0.110 ± 0.007	0.053 ± 0.008	0.013 ± 0.004	-0.032 ± 0.041	1.570 ± 0.081	6.682 ± 0.464	0.050 ± 0.005
SKYNET P(z)	0.004 ± 0.003	-0.002 ± 0.002	0.064 ± 0.002	0.110 ± 0.008	0.047 ± 0.006	0.016 ± 0.004	0.017 ± 0.035	1.200 ± 0.092	9.807 ± 0.709	0.079 ± 0.006
BFZ P(z)	-0.009 ± 0.005	-0.019 ± 0.002	0.097 ± 0.003	0.146 ± 0.008	0.048 ± 0.004	0.019 ± 0.003	-0.020 ± 0.067	2.463 ± 0.234	6.169 ± 0.888	0.058 ± 0.007
ZEBRA P(z)	-0.027 ± 0.005	-0.019 ± 0.003	0.103 ± 0.003	0.163 ± 0.008	0.054 ± 0.005	0.016 ± 0.003	-0.292 ± 0.075	4.247 ± 0.649	8.400 ± 1.427	0.071 ± 0.007

Table A3. Results of all the photo- z metrics for all the codes analysed in Test 3. The errors are computed from bootstrap resampling with 100 samples. The weighting procedure has been applied, together with a cut on the 10 per cent of the galaxies with the largest estimated photo- z errors for each code.

Test 3	$\overline{\Delta z}$	Δz_{50}	σ_{68}	$\sigma_{\Delta z}$	out $_{2\sigma}$	out $_{3\sigma}$	$\overline{\Delta z'}$	$\sigma_{\Delta z'}$	N_{Poisson}	KS
DESDM	-0.010 ± 0.003	-0.010 ± 0.002	0.091 ± 0.002	0.127 ± 0.004	0.052 ± 0.005	0.015 ± 0.002	-0.111 ± 0.027	1.421 ± 0.047	6.819 ± 0.523	0.076 ± 0.006
ANNZ	-0.006 ± 0.002	-0.009 ± 0.002	0.085 ± 0.002	0.114 ± 0.003	0.044 ± 0.004	0.017 ± 0.003	-0.063 ± 0.036	2.664 ± 0.078	7.867 ± 0.556	0.096 ± 0.007
TPZ	-0.004 ± 0.003	-0.001 ± 0.002	0.076 ± 0.002	0.124 ± 0.006	0.048 ± 0.004	0.021 ± 0.003	-0.047 ± 0.040	1.770 ± 0.067	3.812 ± 0.837	0.038 ± 0.005
RVMZ	-0.033 ± 0.005	-0.016 ± 0.003	0.118 ± 0.005	0.177 ± 0.009	0.050 ± 0.004	0.019 ± 0.003	-0.186 ± 0.029	1.044 ± 0.059	8.100 ± 0.594	0.088 ± 0.008
NIP-KNNZ	-0.039 ± 0.005	-0.011 ± 0.002	0.115 ± 0.004	0.191 ± 0.009	0.053 ± 0.004	0.020 ± 0.002	-0.270 ± 0.047	1.889 ± 0.516	3.921 ± 0.372	0.068 ± 0.007
ANNZ2	-0.014 ± 0.004	-0.012 ± 0.002	0.101 ± 0.004	0.144 ± 0.005	0.056 ± 0.004	0.019 ± 0.003	-0.087 ± 0.063	2.668 ± 0.203	6.517 ± 0.514	0.071 ± 0.007
BFZ	-0.021 ± 0.003	-0.021 ± 0.002	0.098 ± 0.003	0.138 ± 0.006	0.049 ± 0.004	0.018 ± 0.002	-0.182 ± 0.032	1.759 ± 0.073	7.895 ± 0.626	0.096 ± 0.008
LEPHARE	0.004 ± 0.004	-0.008 ± 0.003	0.111 ± 0.003	0.170 ± 0.007	0.047 ± 0.003	0.024 ± 0.002	1.599 ± 0.429	50.489 ± 6.968	6.371 ± 0.433	0.089 ± 0.006
PHOTOZ	-0.031 ± 0.003	-0.032 ± 0.002	0.097 ± 0.003	0.140 ± 0.006	0.057 ± 0.004	0.017 ± 0.002	-0.294 ± 0.020	1.003 ± 0.029	4.992 ± 0.514	0.088 ± 0.007
TPZ P(z)	0.002 ± 0.003	0.006 ± 0.002	0.076 ± 0.002	0.119 ± 0.006	0.049 ± 0.005	0.020 ± 0.003	0.082 ± 0.038	1.740 ± 0.071	3.321 ± 0.366	0.030 ± 0.004
ARBORZ P(z)	-0.016 ± 0.003	-0.011 ± 0.003	0.105 ± 0.002	0.128 ± 0.005	0.053 ± 0.004	0.011 ± 0.002	-0.112 ± 0.023	0.953 ± 0.030	4.958 ± 0.380	0.081 ± 0.006
ANNZ2 P(z)	-0.025 ± 0.004	-0.015 ± 0.002	0.088 ± 0.003	0.128 ± 0.007	0.049 ± 0.004	0.018 ± 0.003	-4.557 ± 0.712	26.898 ± 2.972	7.402 ± 0.541	0.098 ± 0.008
SKYNET P(z)	0.007 ± 0.004	0.007 ± 0.003	0.088 ± 0.002	0.129 ± 0.006	0.068 ± 0.005	0.023 ± 0.003	0.071 ± 0.025	0.980 ± 0.039	8.149 ± 0.865	0.093 ± 0.006
BFZ P(z)	-0.024 ± 0.003	-0.024 ± 0.003	0.101 ± 0.002	0.130 ± 0.006	0.046 ± 0.004	0.014 ± 0.002	-0.211 ± 0.033	1.734 ± 0.079	5.325 ± 0.503	0.096 ± 0.007
ZEBRA P(z)	-0.050 ± 0.005	-0.030 ± 0.002	0.109 ± 0.004	0.177 ± 0.012	0.043 ± 0.006	0.018 ± 0.002	-0.383 ± 0.047	1.906 ± 0.166	6.692 ± 0.550	0.112 ± 0.008

(vii) $\overline{\Delta z'}$, the mean of the distribution of Δz is normalized by their estimated errors. Ideally, this distribution should resemble a normal distribution with zero mean and unit variance. We define $\Delta z'_i = \Delta z_i / \epsilon_{\text{phot},i}$, where $\epsilon_{\text{phot},i}$ is the computed error of the photometric redshift for galaxy i . Then

$$\overline{\Delta z'} = \frac{\sum \omega_i \Delta z'_i}{\sum \omega_i}, \quad (\text{A7})$$

(viii) $\sigma_{\Delta z'}$:

$$\sigma_{\Delta z'} = \left(\frac{\sum \omega_i (\Delta z'_i - \overline{\Delta z'})^2}{\sum \omega_i} \right)^{1/2} \quad (\text{A8})$$

(ix) N_{Poisson} , a metric that quantifies how close the distribution of photometric redshifts $N(z_{\text{phot}})$ is to the distribution of spectroscopic redshifts $N(z_{\text{spec}})$. For each photometric redshift bin j of width 0.1, we compute the difference of $N(z_{\text{phot}}) - N(z_{\text{spec}})$ normalized by the Poisson fluctuations on $N(z_{\text{spec}})$:

$$n_{\text{Poisson},j} = \frac{\left(\sum_{z_{\text{phot},i} \in \text{bin}_j} \omega_i N - \sum_{z_{\text{spec},i} \in \text{bin}_j} \omega_i N \right)}{\sqrt{\sum_{z_{\text{spec},i} \in \text{bin}_j} \omega_i N}}.$$

Then N_{Poisson} is computed as the rms of the previous quantity:

$$N_{\text{Poisson}} = \left(\frac{1}{n_{\text{bins}}} \sum_{j=1}^{n_{\text{bins}}} n_{\text{Poisson},j}^2 \right)^{1/2} \quad (\text{A9})$$

(x) KS is the Kolmogorov–Smirnov test that quantifies whether the two redshift distributions ($N(z_{\text{phot}})$ and $N(z_{\text{spec}})$) are compatible with being drawn from the same parent distribution, independently of binning. It is defined as the maximum distance between both empirical cumulative distributions. The lower this value, the closer are both distributions. The empirical cumulative distribution function is calculated as

$$F_{\text{spec}}(z) = \frac{\sum_{i=1}^N \Omega_{z_{\text{spec},i} < z}}{\sum \omega_i}$$

where

$$\Omega_{z_{\text{spec},i} < z} = \begin{cases} \omega_i, & \text{if } z_{\text{spec},i} < z \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, the empirical cumulative distribution function $F_{\text{phot}}(z)$ is computed for $N(z_{\text{phot}})$. Then the KS statistic is computed as

$$\text{KS} = \max_z (|F_{\text{phot}}(z) - F_{\text{spec}}(z)|). \quad (\text{A10})$$

For the submissions with a $P(z)$ for each galaxy, these cumulative distributions are computed taking into account the $p(z)$ distribution for each galaxy.

¹*Institut de Física d'Altes Energies, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Barcelona), Spain*

²*Department of Astronomy, University of Illinois, Urbana, IL 61820, USA*

³*National Center for Supercomputing Applications, 1205 West Clark St, Urbana, IL 61801, USA*

⁴*Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA*

⁵*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*

⁶*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK*

⁷*ETH Zurich, Institut für Astronomie, HIT J 11.3, Wolfgang-Pauli-Str. 27, CH-8093 Zurich, Switzerland*

⁸*Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Burnaby Road, Portsmouth PO1 3FX, UK*

⁹*Observatório Nacional (ON/MCT), Rua General José Cristino, 77, Rio de Janeiro - RJ 20921-400, Brazil*

¹⁰*Laboratório Nacional de e-Astronomia, Rua General José Cristino, 77, Rio de Janeiro, RJ 20921-400, São Cristóvão, Brazil*

¹¹*Institut de Ciències de l'Espai (ICE, IEEC/CSIC), E-08193 Bellaterra (Barcelona), Spain*

¹²*Kavli Institute for Particle Astrophysics and Cosmology, 452 Lomita Mall, Stanford University, Stanford, CA 94305, USA*

¹³*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*

¹⁴*University Observatory Munich, Scheinerstrasse 1, D-81679 Munich, Germany*

¹⁵*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*

¹⁶*School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK*

¹⁷*Departamento de Física Matemática, Instituto de Física, Universidade de Sao Paulo, Sao Paulo, SP CP 66318, CEP 05314-970, Brazil*

¹⁸*Centro de Investigaciones Energéticas Medioambientales y Tecnológicas, Av Complutense 40, E-28040 Madrid, Spain*

¹⁹*Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile*

²⁰*Space Telescope Science Institute (STScI), 3700 San Martin Drive, Baltimore, MD 21218, USA*

²¹*National Optical Astronomy Observatory, 950 N Cherry Ave, Tucson, AZ 85719, USA*

²²*Department of Physics & Astronomy, University of Pennsylvania, Philadelphia PA 19104, USA*

²³*Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA*

²⁴*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*

²⁵*Research School of Astronomy and Astrophysics, The Australian National University, Canberra, ACT 2611, Australia*

²⁶*ARC Centre of Excellence for All-Sky Astrophysics (CAASTRO)*

²⁷*School of Mathematics and Physics, University of Queensland, QLD 4072, Australia*

²⁸*George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*

²⁹*Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*

³⁰*Radcliffe Institute for Advanced Study, Harvard University, Byerly Hall, 8 Garden Street, Cambridge, MA 02138, USA*

³¹*Department of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 München, Germany*

³²*Excellence Cluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany*

³³*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*

³⁴*Institut d'Astrophysique de Paris, Univ. Pierre et Marie Curie & CNRS UMR7095, F-75014 Paris, France*

³⁵*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, VIC 3122, Australia*

³⁶*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*

³⁷*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*

³⁸*Australian Astronomical Observatory, North Ryde, NSW 2113, Australia*

³⁹*ICRA, Centro Brasileiro de Pesquisas Físicas, Rua Dr Xavier Sigaud 150, Rio de Janeiro - RJ, CEP 22290-180, Brazil*

⁴⁰*Instituto de Física, UFRGS, Caixa Postal 15051, Porto Alegre, RS - 91501-970, Brazil*

⁴¹*SEPnet, South East Physics Network (www.sepnet.ac.uk)*

⁴²*Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, University of Manchester; Oxford Road, Manchester M13 9PL, UK*

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.