Mon. Not. R. Astron. Soc. **396,** 2379–2398 (2009) doi:10.1111/j.1365-2966.2009.14908.x

Estimating the redshift distribution of photometric galaxy samples – II. Applications and tests of a new method

Carlos E. Cunha,[1,2,3⋆] Marcos Lima,[2,4,5] Hiroaki Oyaizu,[1,2] Joshua Frieman[1,2,6] and Huan Lin[6]

[1]*Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA*
[2]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[3]*Department of Physics, University of Michigan, 450 Church St., Ann Arbor, MI 48109, USA*
[4]*Department of Physics, University of Chicago, Chicago, IL 60637, USA*
[5]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[6]*Center for Particle Astrophysics, Fermi National Accelerator Laboratory, Batavia, IL 60510, USA*

Accepted 2009 April 14. Received 2009 April 14; in original form 2008 October 16

ABSTRACT

In Lima et al. we presented a new method for estimating the redshift distribution, $N(z)$, of a photometric galaxy sample, using photometric observables and weighted sampling from a spectroscopic subsample of the data. In this paper, we extend this method and explore various applications of it, using both simulations and real data from the Sloan Digital Sky Survey (SDSS). In addition to estimating the redshift distribution for an entire sample, the weighting method enables accurate estimates of the redshift probability distribution, $p(z)$, for *each* galaxy in a photometric sample. Use of $p(z)$ in cosmological analyses can substantially reduce biases associated with traditional photometric redshifts, in which a single redshift estimate is associated with each galaxy. The weighting procedure also naturally indicates which galaxies in the photometric sample are expected to have accurate redshift estimates, namely those that lie in regions of photometric-observable space that are well sampled by the spectroscopic subsample. In addition to providing a method that has some advantages over standard photo-z estimates, the weights method can also be used in conjunction *with* photo-z estimates e.g. by providing improved estimation of $N(z)$ via deconvolution of $N(z_{\rm phot})$ and improved estimates of photo-z scatter and bias. We present a publicly available $p(z)$ catalogue for ∼78 million SDSS DR7 galaxies.

Key words: galaxies: distances and redshifts – galaxies: statistics – distance scale – large-scale structure of Universe.

1 INTRODUCTION

Optical and near-infrared (NIR) wide-area surveys planned for the next decade will increase the size of photometric galaxy samples by an order of magnitude, delivering measurements of billions of galaxies. Much of the utility of these samples for astronomical and cosmological studies will rest on knowledge of the redshift distributions of the galaxies they contain. For example, surveys aimed at probing dark energy via clusters, weak lensing and baryon acoustic oscillations (BAO) will rely on the ability to coarsely bin galaxies by redshift, enabling approximate distance–redshift measurements as well as study of the growth of density perturbations. The power of these surveys to constrain cosmological parameters will be limited in part by the accuracy with which the galaxy redshift distributions can be determined (Huterer et al. 2004, 2006; Ma, Hu & Huterer 2006; Zhan 2006; Zhan & Knox 2006; Lima & Hu 2007).

Photometric redshifts – approximate estimates of galaxy redshifts based on their broad-band photometric observables e.g. magnitudes or colours – offer one set of techniques for approaching this problem. However, photo-z estimators are typically biased to some degree, and they can suffer from catastrophic failures in certain regimes. These problems motivate the development of potentially more robust methods.

In Lima et al. (2008) we presented a new, empirical technique aimed not at estimating individual galaxy redshifts but instead at estimating the redshift distribution, $N(z)$, for an entire photometric galaxy sample or suitably selected subsample. The method is based upon matching the distributions of photometric observables (e.g. magnitudes, colours etc.) of a spectroscopic subsample to those of the photometric sample. The method assigns weights to galaxies in the spectroscopic subsample (hereafter denoted the training

⋆E-mail: cunha@uchicago.edu

set, in analogy with machine-learning methods of photo-z estimation), such that the weighted distributions of observables for these galaxies match those of the photometric sample. The weight for each training-set galaxy is computed by comparing the local 'density' of training-set galaxies in the multidimensional space of photometric observables to the density of the photometric sample in the same region. We estimate the densities using a nearest neighbour approach that ensures that the density estimates are both local and stable in sparsely occupied regions of the space. The use of the nearest neighbours ensures optimal binning of the data, which minimizes the requisite size of the spectroscopic subsample. After the training-set galaxy weights are derived, we sum them in redshift bins to estimate the redshift distribution for the photometric sample.

As Lima et al. (2008) show, this weighting method provides a precise and nearly unbiased estimate of the underlying redshift distribution for a photometric sample without recourse to photo-z estimates for individual galaxies. Moreover, the spectroscopic training set does *not* have to be representative of the photometric sample, in its distributions of magnitudes, colours or redshift, for the method to work. (By contrast, the performance of training-set-based photo-z estimators generally degrades as the training set becomes less representative of the photometric sample.) The only requirement is that the spectroscopic training set *covers*, even sparsely, the range of photometric observables spanned by the photometric sample. The weighting method can be applied to different combinations of photometric observables that correlate with redshift – here, we confine our analysis to magnitudes and colours.

In this paper we present additional applications of the weighting method, test its performance on simulated data sets and show results of those applications using data from the Sloan Digital Sky Survey (SDSS). The applications of the weighting method naturally fall into two categories, those that enhance photo-z estimators and those that (potentially) replace photo-z estimation. In the first category, we show that the weighting method can be used to improve estimates of the scatter and bias of training-set-based photo-z estimates as functions of (true) spectroscopic redshift. Knowledge of such errors are very important, since uncertainties in photo-z bias and scatter are nuisance parameters that significantly degrade the power of cosmological probes (e.g. Huterer et al. 2004; Ma et al. 2006; Lima & Hu 2007). We also show that the weights can be used to obtain improved estimates of the error distribution of the photo-zs, $P(z_{\rm phot}|z_{\rm spec})$, and thereby improve the deconvolution procedure used to infer the underlying redshift distribution, $N(z)$, from the distribution of photo-zs (Padmanabhan et al. 2005).

In the second category of applications, we consider the weighting technique on its own, independently of 'traditional' photo-z estimates. The accuracy of the weighting method in directly reconstructing $N(z)$ is affected by photometric errors and by sparse or incomplete coverage by the training set of the space of photometric observables spanned by the photometric data. We develop and test a bootstrap technique to estimate random errors in the weighted $N(z)$ estimate and present a technique for detecting systematic errors in it as well. We also discuss the effects of training-set non-representativeness on the $N(z)$ estimate. Perhaps most importantly, we show that the weighting procedure can be used to estimate not only the redshift distribution for the (entire) photometric sample, $N(z)$, but also a redshift probability distribution, $p(z)$, for each galaxy in the photometric sample. Such a distribution contains much more information than a discrete photo-z estimate, $z_{\rm phot}$. Use of $p(z)$ instead of $z_{\rm phot}$ in cosmological analyses can potentially greatly reduce the biases arising from photo-zs.

The paper is organized as follows. In Section 2 we review and extend the weighting method for estimating the redshift distribution and the redshift probability distribution, focusing in particular on sources and estimates of errors in the method. In Section 3 we describe the actual and simulated SDSS galaxy catalogues that we use to test the weighting method and its alternatives. We demonstrate how the weighting method improves upon photometric-redshift estimates in the mock catalogue in Section 4, and we demonstrate its effectiveness in estimating $N(z)$, in comparison with photo-z-based methods, in Section 5. We apply the new methods to the real SDSS Data Release (DR6) in Section 6. We present our conclusions in Section 7 and include some technical details of the analysis in the appendices.

2 THE WEIGHTING METHOD

In this section, we briefly review and extend the weighting method introduced in Lima et al. (2008). We define the weight, w, of a galaxy in the spectroscopic training set as the normalized ratio of the density of galaxies in the photometric sample to the density of training-set galaxies around the given galaxy. These densities are calculated in a local neighbourhood in the space of photometric observables e.g. multiband magnitudes. More formally, given a training-set galaxy, we define its weight by

$$w \equiv \frac{1}{N_{\rm P,tot}} \frac{\rho_{\rm P}}{\rho_{\rm T}}, \tag{1}$$

where $N_{\rm P,tot}$ is the total number of galaxies in the photometric sample, and $\rho_{\rm P}$ and $\rho_{\rm T}$ are the local number densities in the space of observables for the photometric and training sets,

$$\rho_{\rm P,T} \equiv \frac{N_{\rm P,T}}{V_{\rm P,T}}, \tag{2}$$

where $N_{\rm P(T)}$ is the number of photometric (training) set galaxies within volume $V_{\rm P(T)}$.

We adopt a nearest neighbour approach to estimating the density of galaxies in magnitude space, because it enables control of statistical errors (shot noise) while also ensuring adequate 'locality' of the volume in magnitude space. We define the distance $d_{\alpha\beta}$ in magnitude space between the αth and βth galaxies in a (photometric or spectroscopic) sample using a Euclidean metric,

$$(d_{\alpha\beta})^2 \equiv (\boldsymbol{m}_\alpha - \boldsymbol{m}_\beta)^2 = \sum_{a=1}^{N_m} \left(m_\beta^a - m_\alpha^a\right)^2, \tag{3}$$

where N_m denotes the number of magnitudes (i.e. different passbands) measured for each galaxy. We use this distance to find the set of *nearest neighbours* to the αth object, i.e. the set of galaxies with the smallest $d_{\alpha\beta}$. For a fixed number of nearest neighbours $N_{\rm nei}$, if we order the neighbours by their distance from the αth galaxy, then we can define the hypervolume in terms of the distance from galaxy α to the $N_{\rm nei}^{\rm th}$ nearest neighbour, indexed by γ, i.e. $V_m = (d_{\alpha\gamma})^{N_m}$.

Estimating the local density in the spectroscopic training set using a fixed value for $N(\boldsymbol{m}_\alpha)_{\rm T} = N_{\rm nei}$ ensures that the density estimate is positive-definite and that the resulting weight is well defined. To estimate the corresponding density in the photometric sample, we simply count the number of galaxies in the photometric sample, $N(\boldsymbol{m}_\alpha)_{\rm P}$, that occupy the *same* hypervolume V_m around the point $\boldsymbol{m}_\alpha$. Since the densities are estimated in the spectroscopic and photometric sets using the same hypervolume, the ratio of the densities in equation (1) is simply the ratio of the corresponding numbers of objects within the volume, and the weight for the αth training-set

galaxy is therefore given by

$$w_\alpha = \frac{1}{N_{\rm P,tot}} \frac{N(\boldsymbol{m}_\alpha)_{\rm P}}{N(\boldsymbol{m}_\alpha)_{\rm T}}. \quad (4)$$

$N_{\rm nei}$ can be chosen to balance locality, which favours small V_m, against statistical errors, which favour large $N_{\rm nei}$.

2.1 Weights and the redshift distribution $N(z)_P$

As shown in Lima et al. (2008), by construction the *weighted* spectroscopic training set has essentially identical distributions of multiband magnitudes and colours as the photometric sample from which it is drawn, even though the spectroscopic set is in general not representative of the photometric sample. The weighting procedure in effect corrects for that non-representativeness, provided the training set adequately spans the range of the photometric-observable space covered by the photometric sample. Since the weighted training set has identical distributions of photometric observables as the photometric sample, it is reasonable to assume that the former also provides an accurate estimate of the binned redshift distribution of the photometric sample,

$$N(z)_{\rm wei} \equiv \hat{N}(z_1 < z < z_2)_{\rm P} = \sum_{\beta=1}^{N_{\rm T,tot}} w_\beta N(z_1 < z_\beta < z_2)_{\rm T}, \quad (5)$$

where the weighted sum is over all galaxies in the training set. Lima et al. (2008) show that this indeed provides a nearly unbiased estimate of the redshift distribution of the photometric sample, $N(z)_{\rm P}$, under suitable conditions. Examples of this application will be discussed in Section 5.1.

2.2 Weights and the redshift probability distribution $p(z)$

Although knowledge of the redshift distribution for a photometric sample, $N(z)_{\rm P}$, is sufficient for many applications, there are of course instances in which one would like redshift information about individual galaxies in the sample. As noted in the Introduction, photo-z estimators provide one approach to this problem. However, photo-z estimates are limited by the fundamental assumption that there is a functional relationship between the photometric observables and redshift. In fact, galaxies occupying a small cell in the space of photometric observables will have a range of redshifts. One can therefore associate that cell with a redshift probability distribution function (PDF), $p(z|\text{observables})$. The shape of the PDF is determined by the choice of observables, the size of the cell, the photometric errors and the range of spectral energy distributions (SEDs) of the galaxies. If the PDF is narrowly peaked, photo-z estimates can be both precise (small scatter) and accurate (small bias). However, if the distribution is broad, skewed or multiply peaked, then photo-z estimates will suffer large scatter, bias and potentially catastrophic failures. The ubiquitous positive bias of photo-z estimates for low-redshift galaxies and negative bias for high-redshift galaxies are consequences of this fundamental assumption. Low- and high-redshift objects can in some cases occupy the same cell of magnitude space, but photo-z estimators will assign them all essentially the same redshift.

To overcome these problems and avoid the biases intrinsic to photo-z estimates, it is preferable to use the full redshift PDF for the galaxies in a small cell in the space of photometric observables, $p(z) \equiv p(z|\text{observables})$. This PDF encodes all the information available about the redshift of an individual galaxy in a photometric sample. One can choose to extract a single redshift estimate from the PDF, e.g. its mean, median or mode, but often that is not necessary in applications.

The weighting method described above can be straightforwardly applied to estimate $p(z)$ using a spectroscopic training set. The estimator $\hat{p}(z)$ for a galaxy in the photometric sample is given by the weighted redshift distribution of its $N_{\rm nei}$ nearest neighbours in the training set, using the metric of equation (3),

$$\hat{p}(z) = \sum_{\beta=1}^{N_{\rm nei}} w_\beta \delta(z - z_\beta), \quad (6)$$

where, as before, $N_{\rm nei}$ can be determined from simulations by minimizing the sum of the shot-noise and 'non-locality' errors. In practice, we estimate $p(z)$ in redshift bins. This estimate for $p(z)$ was used in a study of galaxy–galaxy lensing by Mandelbaum et al. (2008) and was shown to yield significantly smaller lensing calibration bias than use of photo-z estimates.

We can also construct a new estimator for $N(z)_{\rm P}$ by summing the $\hat{p}(z)$ distributions for all galaxies in the photometric sample,

$$\hat{N}(z)_{\rm P} = \sum_{i=1}^{N_{\rm P,tot}} \hat{p}_i(z). \quad (7)$$

This estimator is similar but not identical to that of equation (5). We will see in Section 5.1.3 that these two are comparable in recovering the true redshift distribution of a photometric sample.

2.3 Sources of errors in the weighting method

The errors arising in the weights method can be considered the errors in estimating $p(z|$ observables) for a galaxy in the photometric sample from the information in the training set. Any differential selection effect between the spectroscopic and photometric samples will lead to errors in $\hat{p}(z)$. There are several kinds of selection effects: (1) statistical effects, (2) large-scale structure (LSS), (3) spectroscopic failures in the training set, (4) survey selection in the photometric observables, (5) survey selection in non-photometric observables and (6) non-locality of the weights.

Statistical errors arise because the training set is just a subsample of the photometric survey and is subject to statistical fluctuations. These fluctuations can be significant in regions of magnitude space where the training set is very sparse. In such regions, the shot-noise errors in $\hat{p}(z)$ will either be large or else the nearest neighbour volume must be made large, leading to increased non-locality (see below). Statistical errors can be estimated by bootstrap resampling the training and photometric sets. If the magnitude errors are well known, one can further Monte Carlo resample the magnitudes. We present results of bootstrap error estimation in Section 5.1.4.

Errors due to LSS can be significant if certain regions of the space of photometric observables are only represented in the training set by a spectroscopic survey that covers a small solid angle, in which one or a few large structures dominate. In this case, $p(z|\text{observables})$ for the training set will comprise one or a few redshift spikes rather than a smooth distribution. If these effects occur in regions of magnitude space where the true redshift PDF is broad or multiply peaked, they can potentially cause systematic errors in the estimates of $p(z)$ or $N(z)$ for the photometric sample. The resulting errors may be large if the linear size of the training-set volume is not large compared to the galaxy clustering correlation length. The errors from LSS can in principle be estimated by constructing mock training-set volumes using N-body simulations of structure formation.

Spectroscopic failures, i.e. targeted objects in the training set for which redshifts could not be obtained, can also lead to systematic errors in $\hat{p}(z)$ if the failures happen systematically, for instance, if they occur preferentially for a particular galaxy spectral type and if that type has a different redshift PDF from other galaxy types in the same region of magnitude space. Since such spectroscopic failures will tend to occur in specific and identifiable regions of magnitude space, however, one can at minimum excise or down-weight those regions in estimating quantities for the photometric sample (see Section 2.4), at the cost of incompleteness.

The severity of these systematic errors is regulated by the width of the redshift PDF. In the limit of a large number of photometric observables with very small measurement errors and a large spectroscopic training set, the redshift PDF in a small cell in magnitude space approaches a δ function. In this regime, the effects of LSS and of spectroscopic failures would be simply to increase the statistical errors in certain regions of observable space, an effect accounted for in the bootstrap error estimate. As one moves away from this ideal limit, the systematic errors grow, in the sense that one can no longer reliably estimate $p(z|$ observables) for a galaxy in the photometric sample from its training-set neighbours. That effect is not captured by the bootstrap and must be estimated by other means e.g. using simulations. The mock SDSS DR6 catalogue we have constructed for this paper (see Section 3.2) does not simulate LSS or spectroscopic failures; we plan to study such effects in the future. Some of the surveys that comprise the training set for the real DR6 data are individually affected by LSS effects. Having a combination of them helps to alleviate the problem, though more testing is required to quantify the possible systematics.

LSS and spectroscopic failures lead to unavoidable differences in the selection functions for the photometric and spectroscopic samples. In addition, there are differential selection effects that are built in by those designing the spectroscopic survey. For example, one typically makes magnitude and colour cuts in selecting spectroscopic targets from a photometric sample. In this case, where the selection is made explicitly in the photometric observables, there will be regions of observable space where the weights cannot be used to reliably estimate redshift distributions. Again, such regions are known from the target selection cuts and can be safely excised from the photometric sample (see Section 2.4). If, on the other hand, there are differences in spectroscopic and photometric selection based on non-photometric observables, then systematic errors in $\hat{p}(z)$ can occur.

A variant of this problem arises when the training set is selected using photometric observables that are different from the ones measured in the photometric sample. For example, for the SDSS DR6 photometric catalogue, the spectroscopic target selection for the Deep Extragalactic Evolutionary Probe 2 (DEEP2) sample in the training set used a different magnitude system (coming from different photometric samples) from the SDSS. Similarly, the selection of the 2dF-SDSS Luminous Red Galaxy (LRG) and QSO (2SLAQ) spectroscopic catalogue made use of photometric observables that were not used in the photo-z estimation. Whether such cuts will cause systematic errors depends on how well the selection in those systems can be approximated using the SDSS *ugriz* filters.

Finally, the non-locality of the weights solution is a source of systematic error. Here, non-locality refers to the fact that, in the nearest neighbour approach, we are using information from a finite volume to estimate the density at a point in observable space, and the density varies over the space. This procedure corresponds to applying a smoothing kernel to the density field. Non-locality becomes a problem if the volume occupied by the neighbours (or the scale of the smoothing kernel) becomes comparable to or larger than the scale over which the density changes appreciably. In this limit, the shape of the volume used to select the nearest neighbours may be important. Non-locality errors are reduced by choosing a smaller neighbour volume for the density estimate, but at the cost of increasing the shot-noise errors. Ultimately, the combined errors can be reduced by increasing the density of the training set in a particular region of observable space, i.e. by measuring more spectra.

2.4 Selecting the 'recoverable' part of a photometric sample

One of the necessary conditions for the weights procedure to work is that the spectroscopic training set covers the same region of photometric observables as the photometric sample. That is, the weights can only recover the redshift PDF of a galaxy in the photometric sample if it lies in the region of intersection of the redshift–observables hypersurfaces of the training and photometric sets. Defining this region of intersection is not always trivial, especially given the high number of dimensions that may be involved. To do so, we count how many times a galaxy in the photometric sample is used in the weights calculation for all members of the training set. By definition, photometric galaxies that are never counted in the weights procedure are not in the region of intersection, hence the redshift distribution of those galaxies will not be accurately recovered by the weighting procedure. We make use of this criterion below. If one does not require the photometric sample to be complete, one can choose to excise such galaxies from consideration. Using several real and mock catalogues, we have found empirically that using $\sim$five nearest neighbours in the weights calculation is optimal for determining the intersection region for the mock catalogue.

As examples, consider the mock and real SDSS DR6 catalogues of Section 3. From Figs 1(b), 2(b) and 3(b), one might expect that the combined training set covers the same region of observables as the photometric sample. However, using the definition of the previous paragraph, more than $\sim$43 per cent of the mock photometric-sample galaxies are not used in the weights calculation, i.e. they are not well represented in the training set. Fortunately, in the real SDSS DR6 catalogue, by the same criterion we find that $\sim$98 per cent of photometric-sample galaxies with $r < 22$ are well represented in the training set. It is important to apply such a recoverability test whenever a training-set method is used.

3 CATALOGUES

To test the performance of the weighting method and compare it with standard photo-z estimates, we employ two kinds of catalogues. The first is drawn from the SDSS DR6 (Adelman-McCarthy et al. 2008) photometric sample and various spectroscopic subsamples of it and allows us to display results of the weighting method on real data. The second is a mock catalogue constructed to have properties similar to the SDSS DR6 photometric and spectroscopic samples. The goal of the mock catalogue is not to precisely reproduce all features of the SDSS catalogue but to have a sample with realistic spectroscopic and photometric features and for which we have ground truth (i.e. redshifts and galaxy types) for all galaxies. In this section, we describe the relevant features of the real and mock catalogues, relegating the details to Appendix A.

3.1 SDSS DR6 data

The SDSS DR6 photometric and spectroscopic data samples are drawn from those used by Oyaizu et al. (2008a) to produce a neural network photo-z catalogue.

3.1.1 Photometric sample

We use a random 1 per cent subset of the galaxies in the SDSS DR6 Photoz2 catalogue described in Oyaizu et al. (2008a) as our photometric sample. This subset contains approximately 769 582 galaxies with $r < 22$. The catalogue is approximately flux limited at this magnitude limit. For details of the parent sample, see Appendix A and Oyaizu et al. (2008a). The r magnitude, $g - r$ and $r - i$ colour distributions are shown in the bottom right-hand panel of Fig. 1(a) and the bottom panels of Fig. 2(a).

3.1.2 Spectroscopic training set

The spectroscopic training sample we use for SDSS DR6 is drawn from a number of spectroscopic galaxy catalogues that overlap with SDSS DR6 imaging. We impose a magnitude limit of $r < 23$ on the spectroscopic samples as well as additional cuts based on the quality of the spectroscopic redshifts reported by the different surveys (see Appendix A). The SDSS spectroscopic sample provides 531 594 redshifts, principally from the MAIN and LRG samples. The remaining redshifts are 20 381 from the Canadian Network for Observational Cosmology (CNOC) Field Galaxy Survey (CNOC2; Yee et al. 2000), 1531 from the Canada–France Redshift Survey (CFRS; Lilly et al. 1995), 11 040 from the DEEP (Davis et al. 2001) and DEEP2 (Weiner et al. 2005), 654 from the Team Keck Redshift Survey (TKRS; Wirth et al. 2004) and 52 762 LRGs from the 2SLAQ Survey (Cannon et al. 2006).

The r magnitude and colour ($g - r$ and $r - i$) distributions for the spectroscopic samples are shown in Figs 1(a) and 2(a). Although the magnitude and colour distributions of the combined spectroscopic sample are not identical to those of the photometric sample, the spectroscopic sample does span the ranges of apparent magnitude and colours of the photometric sample. Fig. 3(a) gives the spectroscopic redshift distribution for the combined spectroscopic sample.

3.2 SDSS DR6: mock catalogue

Using spectral template libraries and observational data on the redshift-dependent luminosity functions of galaxies of different types, we have constructed mock photometric and spectroscopic samples that reproduce the main features of the real SDSS DR6 samples. We describe these briefly below.

3.2.1 Mock photometric sample

The simulated SDSS catalogue contains 10^7 galaxies with redshift $z < 2.0$ and magnitude $14 < r < 22$. We use the LF_MOCK_SCHECHTER code from the KCORRECT package (Blanton et al. 2003) to generate redshift, type and i–magnitude relations. The inputs to the code are the redshift range, Schechter luminosity function parameters and the ranges of absolute and apparent r magnitudes. The code outputs a list of redshifts and apparent r magnitudes. We set the range of absolute i-band magnitudes to $(-24, -14)$. Using data from the VIMOS VLT Deep Survey (VVDS), Zucca et al. (2006) estimated galaxy luminosity functions and Schechter-function fits thereto for different galaxy types in redshift bins of size $\Delta z = 0.2$ from $z_{\rm min} = 0.2$ to $z_{\rm max} = 1.5$. We fit simple polynomial functions to the Schechter parameters of Zucca et al. (2006) to derive a continuous relationship between the Schechter parameters M^*, α, ϕ^*, redshift z and galaxy type T, using the centroid of each redshift bin for the fit. To regularize the fits, we visually extrapolated the results of Zucca et al. (2006) to the $z = (0, 0.2)$ bin and, where needed (for certain galaxy types), for the (1.2, 1.5) bin. The detailed fits are given in Appendix B.

Galaxy colours are generated using the four Coleman, Wu, & Weedman spectral templates (Coleman, Wu & Weedman 1980) – E, Sbc, Scd, Im – extended to ultraviolet (UV) and NIR

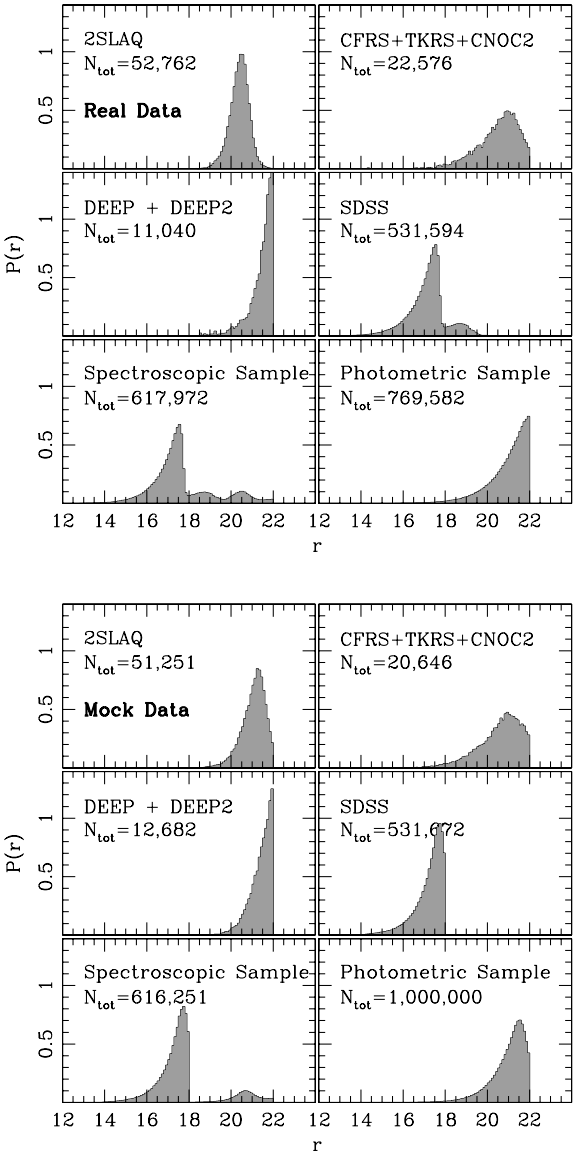

Figure 1. Normalized r magnitude distributions for the catalogues comprising the real SDSS DR6 (top: figure a) and the mock SDSS DR6 (bottom: figure b) catalogues. In each figure, the top four panels indicate the distributions for the different spectroscopic subsamples (see text), bottom left-hand panels indicate flux distributions for the combined spectroscopic samples and bottom right-hand panels indicate distributions for the photometric samples. In each panel, $N_{\rm tot}$ denotes the total number of galaxy measurements used in each sample.

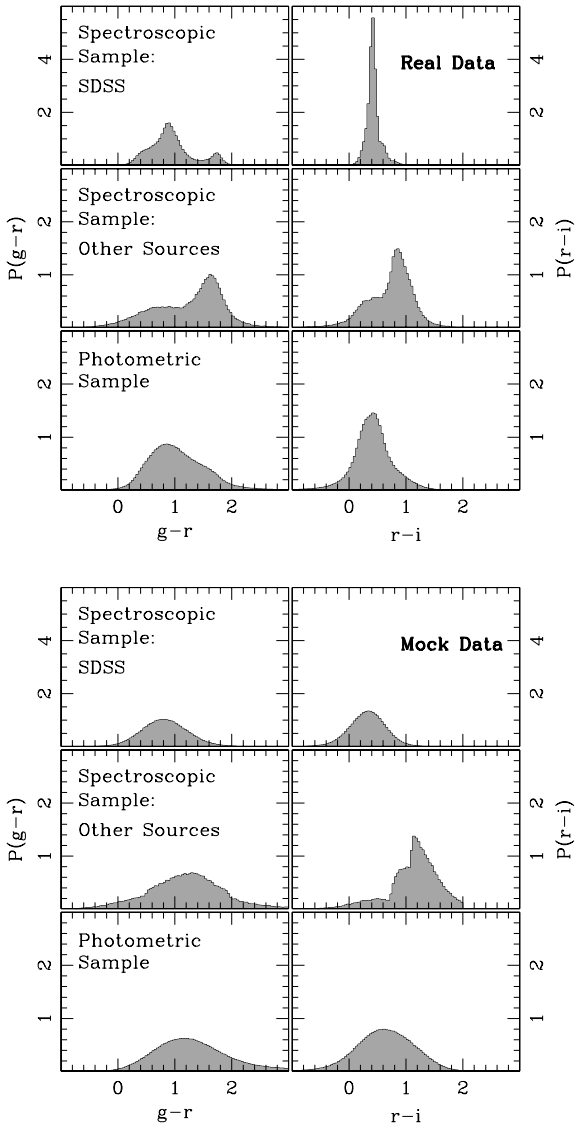

Figure 2. Distributions of $g - r$ and $r - i$ colours for the catalogues comprising spectroscopic training and photometric sets for the real SDSS DR6 (top: figure a) and the mock SDSS DR6 (bottom: figure b). Top rows give distributions for the SDSS spectroscopic sample, middle rows the distributions for the other spectroscopic samples and bottom rows the distributions for the photometric samples. The real and mock SDSS spectroscopic colour distributions differ primarily because the latter does not include LRGs.

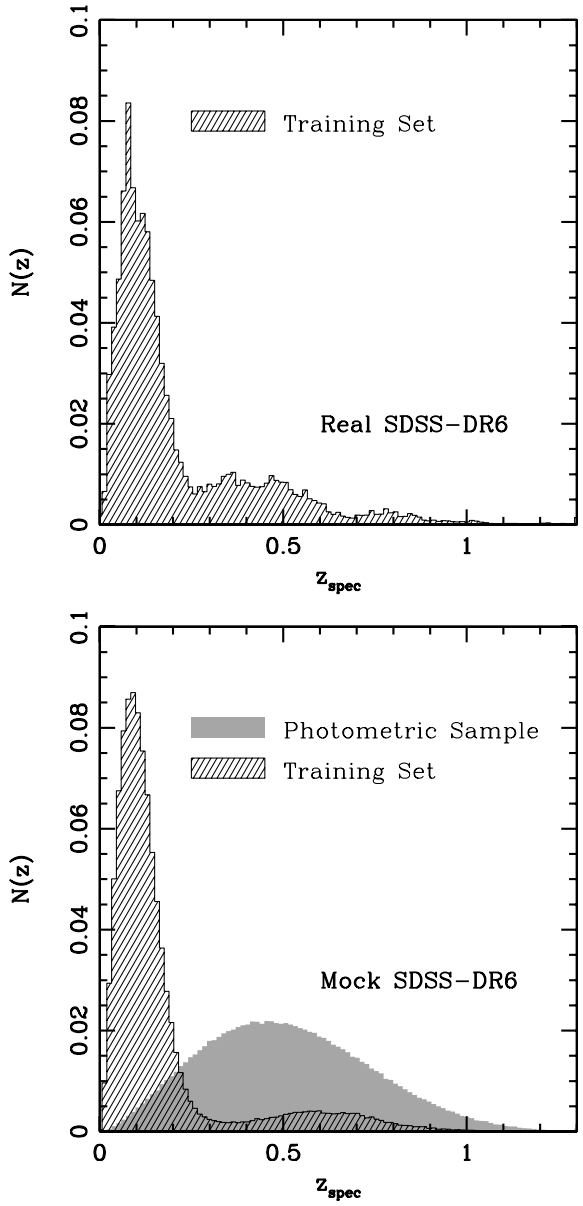

Figure 3. Top, figure a: spectroscopic redshift distribution for the combined SDSS DR6 spectroscopic training set. Bottom, figure b: spectroscopic redshift distributions for the mock SDSS DR6 training and photometric sets.

wavelengths using synthetic templates from Bruzual & Charlot (1993). These templates are mapped to galaxy SED type T (used by Zucca et al. 2006) as (E, Sbc, Scd, Im) $\rightarrow$ T = (1, 2, 3, 4). To improve the sampling and coverage of colour space, we have created additional templates by interpolating between adjacent templates. The redshift, r magnitude and type relations are first generated without photometric errors; errors are then added to produce observed magnitudes. Magnitude errors are modelled as sky-background-dominated errors approximated as Gaussians that are uncorrelated between the different SDSS filters.

The resulting magnitude and colour distributions for the mock photometric sample are shown in the lower right-hand panel of Fig. 1(b) and the bottom panels of Fig. 2(b). The redshift distribution for the sample is shown as the dark grey region in Fig. 3(b). The r magnitude distribution of the mock photometric sample peaks at slightly brighter magnitude than for the actual DR6 photometric sample, and the $r - i$ distribution is slightly less peaked than that of the real data, but overall the real and mock distributions are quite similar in their photometric properties. As noted above, the goal of the mocks is *not* to exactly reproduce the real data distributions.

3.2.2 *Mock spectroscopic training set*

We construct the mock spectroscopic training set by piecing together a variety of different catalogues with different selection functions, each meant to qualitatively represent one of the spectroscopic training samples described above in Section 3.1.2. We obtain each component catalogue of the training set by generating an independent realization of the mock photometric sample and applying the selection cuts of the spectroscopic catalogue to the realization. The selection cuts we use for each component

Table 1. Mock spectroscopic training set properties: number of galaxies and photometric selection cuts applied.

Catalogue	Unique objects	All objects	Selection cuts
mockSDSS	531 672	531 672	$r \leq 18.0$
mockDEEP+DEEP2	2 419	31 716	$g - r < 2.35(r - i) - 0.45$, $g - r < 1.95$, $1.1 < r - i < 2$, $r < 22$
mockTKRS+CFRS+CNOC2	1 827	23 681	$u < 23$, $g < 23$, $r < 22$, $i < 22$
mock2SLAQ	11 082	51 251	$((r - i) - (g - r))/8 \geq 0.55$, $0.7(g - r) + 1.2(r - i - 0.18) \geq 1.6$, $17.5 \leq i \leq 19.8$, $0.5 < g - r < 3$, $r - i < 2$

spectroscopic catalogue are given in Table 1. As discussed in Appendix A2, many of the real training set galaxies are located in the southern celestial stripe, which was imaged repeatedly by the SDSS. In the real training set, multiple photometric measurements of the same galaxy were treated as independent. We have simulated this effect in the mock training set by regenerating the magnitudes of each galaxy in the mock training sets as needed. The number of unique mock galaxies and total number of galaxies (counting all realizations of the same galaxy as different objects) are shown in the second and third columns of Table 1. For comparison, we have also generated spectroscopic catalogues with the same total number of objects but using only unique objects. We found no discernible differences in the resulting photo-zs or weights.

The r magnitude, colour ($g - r$ and $r - i$) and spectroscopic redshift distributions of the spectroscopic samples for the mock SDSS DR6 data are shown in Figs 1, 2 and 3. As is evident from comparison of the *a* and *b* components of Figs 1 and 2, there are some noteworthy differences between the selection cuts used for the mock training set and the actual target selection cuts applied in constructing the spectroscopic surveys described in Section 3.1.2 and Appendix A2. For example, for the SDSS spectroscopic catalogue, the mock sample is flux limited at $r = 18$, while the actual spectroscopic catalogue comprises the MAIN sample, with a flux limit of $r = 17.7$, and the LRG sample, with red colours and a flux distribution that peaks around $r \approx 19$. For the other spectroscopic surveys, the actual photometric selection cuts were typically made in non-SDSS passbands, while our mock data and selection cuts were generated using the SDSS *ugriz* bands. Therefore, the mock photometric cuts do not exactly match the actual cuts used. As a result of this mismatch, e.g. the peak of the r magnitude distribution of the mock 2SLAQ sample is about one magnitude fainter than the corresponding peak in the real data, as shown in the upper left-hand panels of Figs 1(a) and (b).

4 APPLICATIONS OF THE WEIGHTING METHOD I: IMPROVING PHOTOMETRIC REDSHIFT MEASURES

With the mock and real galaxy catalogues in hand, we can now test the performance of the weighting method in different applications. In this section, we describe the utility of the weighting method in improving the performance of traditional photo-z estimates. In the next section, we use the weighting method to directly estimate $N(z)$ and compare the results with photo-z-based estimates.

4.1 Estimating photo-z bias and scatter

We have applied an artificial neural network (ANN) photo-z estimator, described in Appendix C and in more detail in Oyaizu et al. (2008a), to the SDSS DR6 mock catalogue of Section 3.2. Despite the fancy name, an ANN is simply a function which relates redshifts to photometric observables. The training set is used to determine the best-fitting value for the free parameters of the ANN. The best-fitting parameters are found by minimizing the overall scatter (see definition below) of the photo-zs determined for the training set galaxies. The ANN configurations are not unique in the sense that different sets of parameters can result in the same overall scatter. The best-fitting parameters found after minimizing the scatter depend on where in parameter space the optimization run begins. Hereafter we refer to an ANN function using a given set of best-fitting parameters as a neural network solution. The network is trained on the mock spectroscopic training set described in Section 3.2.2 and used to estimate redshifts for the mock photometric sample of Section 3.2.1. We have also trained and applied the network using the real DR6 data described in Section 3.1.

The results of the ANN photo-z estimator are displayed in Fig. 4, which shows the inferred redshift $z_{\rm phot}$ versus true redshift $z_{\rm spec}$. Panel (b) shows the results for the mock spectroscopic training set, while panel (c) shows the results for the mock photometric sample. For comparison, panel (a) shows results for the real SDSS DR6 training set data. As was seen in Fig. 3(b), the redshift distribution of the mock photometric sample is considerably deeper than that of the mock training set. Not surprisingly, the photo-z errors as a function of redshift for the mock photometric sample are somewhat larger than one would estimate based on the training set (compare the 68 and 95 per cent contours in panels b and c). This is a problem since, for real (as opposed to mock) galaxy catalogues, one does not have the information necessary to make panel (c), i.e. one can only estimate photo-z performance using the training set. Since the training set is, as in this mock example, usually not representative of the photometric sample, the statistics of photo-z quality for the training set are not accurate indicators of photo-z quality for the photometric sample.

To make this point more quantitative, we consider two standard statistical measures of photo-z quality, the scatter and bias as functions of spectroscopic redshift:

$$\sigma^2(z_j) \equiv (1/N_j) \sum_{i=1}^{N_j} |z_{{\rm phot},i} - z_{{\rm spec},i}|^2, \quad (8)$$

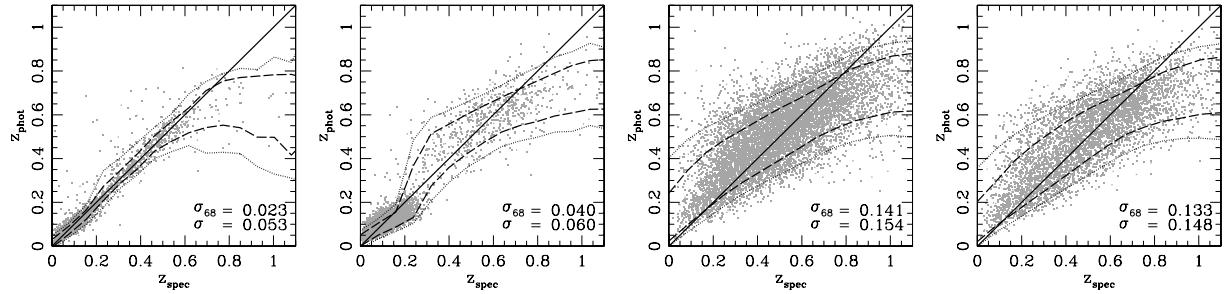

Figure 4. $z_{\rm phot}$ versus $z_{\rm spec}$ for (from left to right): (a) the real SDSS DR6 training set, (b) the mock SDSS training set, (c) the full mock photometric set and (d) the recoverable mock photometric set, i.e. the part of the mock photometric set that is well represented in the training set. The dashed and dotted curves enclose 68 and 95 per cent of the points in each $z_{\rm spec}$ bin. In the lower right of each panel, σ is the rms photo-z scatter averaged over all objects in the catalogue, and σ_{68} is the range containing 68 per cent of the objects in the distribution of $z_{\rm phot}-z_{\rm spec}$.

$$b(z_j) \equiv (1/N_j)\sum_{i=1}^{N_j}(z_{{\rm phot},i} - z_{{\rm spec},i}), \tag{9}$$

where N_j is the number of objects in the jth $z_{\rm spec}$ bin, i.e. with true redshifts in the interval $z_j \pm \Delta z$. Fig. 5 shows these measures for the mock training sample (left-hand panels) and photometric sample (right-hand panels) for five different neural network solutions. These five solutions come from networks with the same structure (same number of layers and nodes per layer, see Appendix C) but with different initial values for the network weights $w_{i\alpha\beta}$. The left-hand panels of Fig. 5 show that the different solutions yield essentially identical results for the scatter and bias for the training set, but the right-hand panels show a dispersion of quality measures for the photometric sample. We can address this issue by working with the average of the five photo-z solutions for each galaxy. The solid (black) curve in the top right-hand panel of Fig. 5 shows that the average photo-z solution results in a $b(z)$ that is the average of the biases of the individual neural net solutions, as may

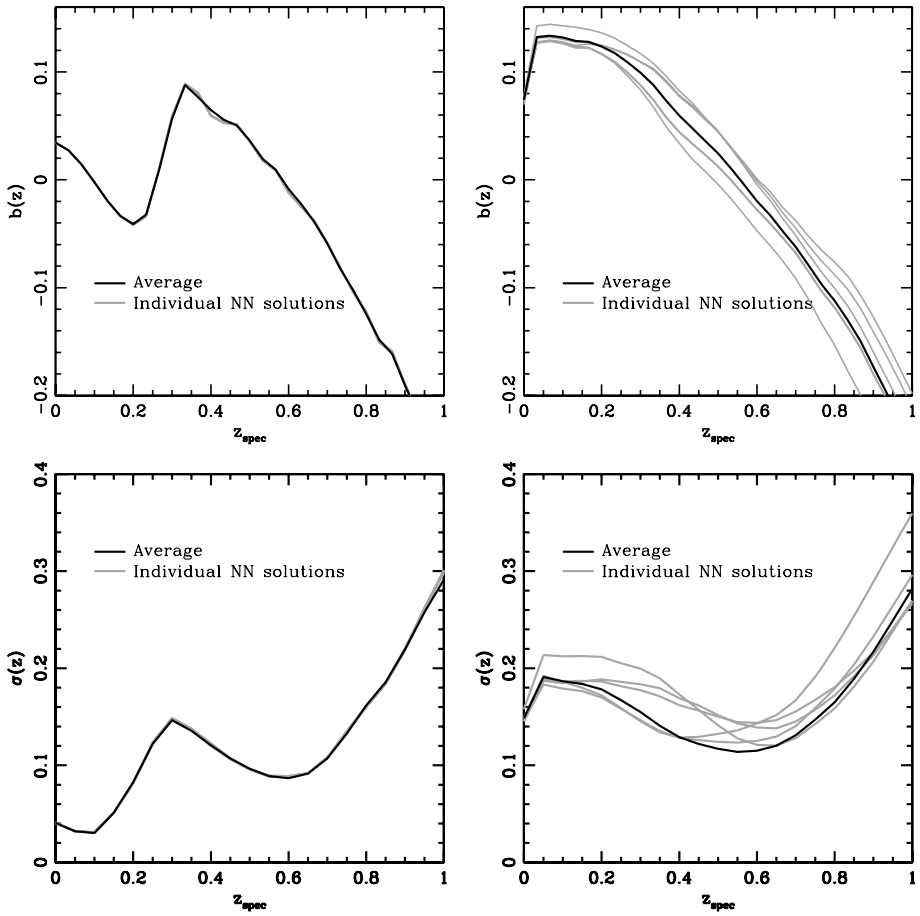

Figure 5. Upper panels: photo-z bias b versus $z_{\rm spec}$ for the five neural network photo-z solutions for the mock SDSS sample: (top left) training set (unweighted) and (top right) photometric set. Lower panels: photo-z scatter σ versus $z_{\rm spec}$ for the five neural network photo-z solutions of the (bottom left) training set and (bottom right) photometric set.

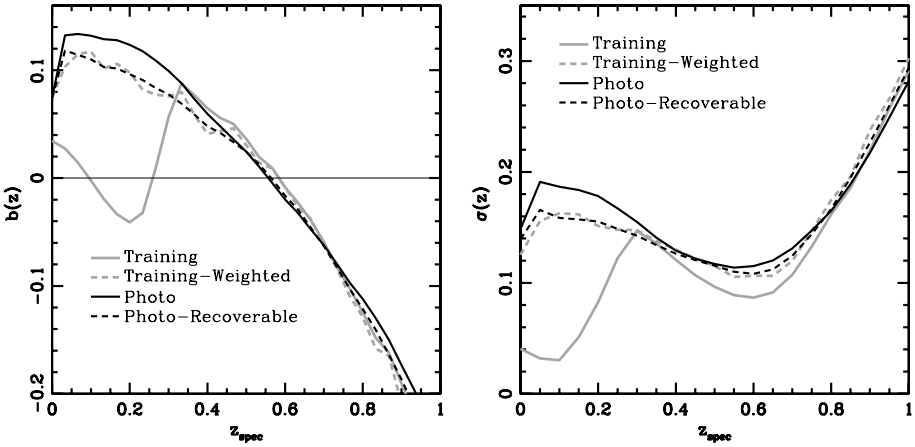

Figure 6. (Left) photo-z bias versus $z_{\rm spec}$ and (right) scatter versus $z_{\rm spec}$ for the weighted and unweighted mock SDSS training set as well as for the mock photometric set and the recoverable photometric set. The weighted training set results more accurately match those for the photometric set and very accurately match those for the recoverable photometric set.

be expected. The bottom right-hand panel of Fig. 5 shows a more interesting result, that the scatter of the average photo-z solution is considerably smaller than the average scatter of the individual neural net solutions.

Even if one uses the average photo-z solution, comparison of the left- and right-hand panels of Fig. 5 demonstrates the qualitative point made above, that the scatter and bias versus redshift for the training set are not accurate estimators of the scatter and bias over the full redshift range for the photometric sample. As shown more explicitly in Fig. 6, the training set scatter and bias tend to underestimate those measures for the photometric sample, particularly at redshifts $z_{\rm spec} < 0.3$. This is simply because the training-set objects are generally brighter than those in the photometric set at similar redshift, which implies that the training-set galaxies have smaller photometric errors and consequently smaller photo-z errors.

The weighting procedure provides a straightforward avenue for addressing this problem of estimating the photo-z scatter and bias for the photometric sample. Since the *weighted* training set has, by construction, magnitude distributions similar to those of the photometric set, we can instead use weighted versions of $\sigma(z)$ and $b(z)$ for the training set as estimates of the scatter and bias for the photometric set, i.e.

$$\sigma_{\rm w}^2(z_j) \equiv (1/N_j)\sum_{i=1}^{N_j} w_i |z_{{\rm phot},i} - z_{{\rm spec},i}|^2, \tag{10}$$

$$b_{\rm w}(z_j) \equiv (1/N_j)\sum_{i=1}^{N_j} w_i (z_{{\rm phot},i} - z_{{\rm spec},i}), \tag{11}$$

where the weights w_i are given by equation (1) and the sums are over all objects in the training set. Fig. 6 shows the scatter and bias for the training set, the weighted training set and the full photometric set, where the average photo-z of the five neural network solutions has been used. We see that the weighted training set yields estimates of scatter and bias that are much closer to those of the photometric set over the entire redshift range. Moreover, as noted in Section 2.4, we expect the weighting method to work best for the *recoverable* portion of the photometric sample. Fig. 6 also shows the scatter and bias versus redshift for the recoverable photometric sample, showing that the weighted training-set estimates are very accurate in this case.

Since the weights can be used to improve the estimates of photo-z scatter and bias for the photometric set, one might hope that the weights could also be used to improve the photo-z solution itself. However, because of the large number of degrees of freedom of the ANN, most of the information for the photo-z solution comes from small regions in the space of photometric observables around each training-set object. The weights do not vary strongly over those small regions, and therefore the photo-z solution does not change significantly between the unweighted and weighted cases.

4.2 Estimating photo-z errors

As demonstrated above, the weighting procedure improves the estimates of photo-z scatter and bias for a photometric sample but does not improve the photo-z accuracy itself. Another issue, which we now discuss, is the accuracy of photo-z error estimates.

We estimate photo-z errors for objects in the photometric catalogue using the nearest neighbour error (NNE) estimator (Oyaizu et al. 2008b). The NNE method is training-set based and associates photo-z errors to photometric objects by considering the errors for objects with similar multiband magnitudes in a spectroscopic sample, hereafter termed the 'validation set'. The validation set is chosen to be independent of the training set in order to avoid the issue of overfitting, i.e. so that the ANN is not trained to fit the statistical fluctuations of the training set, which would result in NNE underestimating the photo-z errors.

The NNE procedure to estimate the redshift error $\sigma_{\rm NNE}$ for a galaxy in the photometric sample is as follows. Using the distance measure of equation (3), we find the validation-set nearest neighbours in magnitude space to the galaxy of interest. Since the selected nearest neighbours are in the spectroscopic sample, we know their photo-z errors, $\delta z = z_{\rm phot} - z_{\rm spec}$, where $z_{\rm phot}$ has been estimated using the neural network method. We calculate the 68 per cent width of the δz distribution for the neighbours and assign that number as the photo-z error estimate for the photometric galaxy. Here we select the nearest 100 neighbours of each object to estimate its photo-z error. In studies of photo-z error estimators applied to mock and real galaxy catalogues, we found that NNE accurately predicts the photo-z error when the training set is representative of the photometric sample (Oyaizu et al. 2008b). Here we investigate what happens when the training set is *not* representative, and we also consider the impact of weighting the neighbours using equation (4) in computing the NNE estimate.

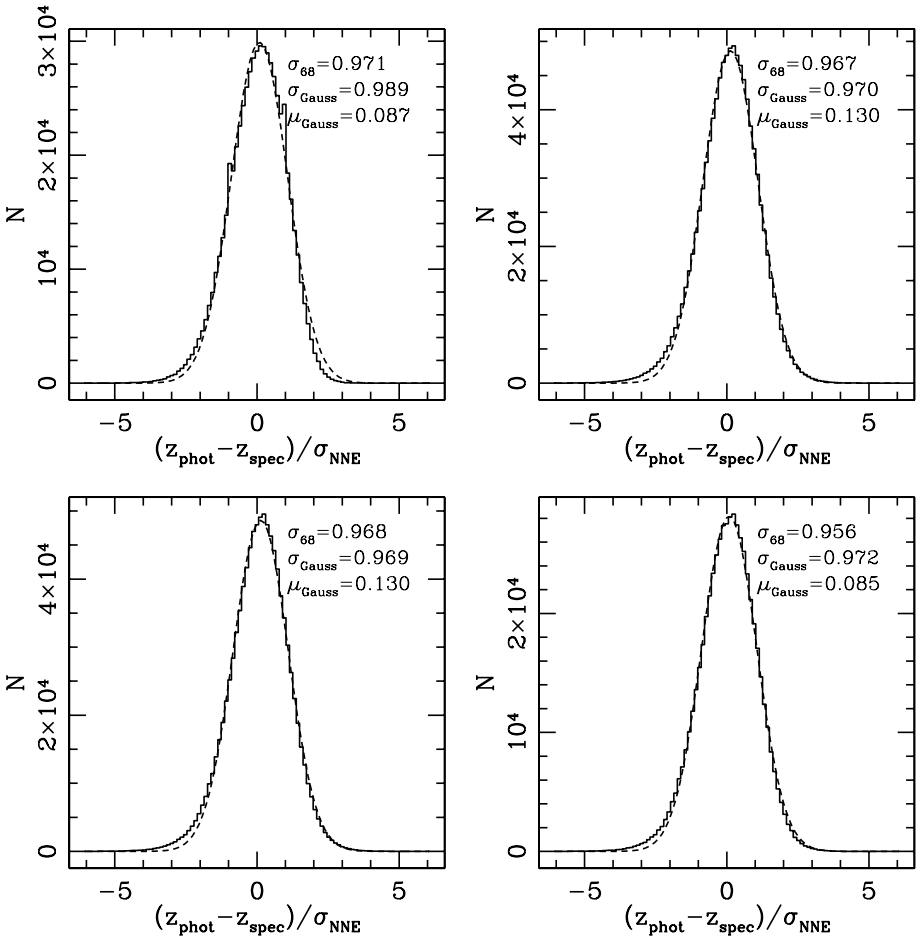

Figure 7. Distributions of $(z_{\rm phot}-z_{\rm spec})/\sigma_{\rm NNE}$ for the (top left) training set, (top right) photometric set (using unweighted validation set), (bottom left) photometric set (using weighted validation set) and (bottom right) recoverable photometric set (using weighted validation set).

Fig. 7 shows the distributions of $(z_{\rm phot} - z_{\rm spec})/\sigma_{\rm NNE}$, i.e. the photo-$z$ error distribution normalized by the NNE error estimate $\sigma_{\rm NNE}$, for the training set (upper left), for the photometric set using unweighted (upper right) and weighted (lower left) validation-set objects and for the recoverable photometric set (lower right) using the weighted validation set. The dashed curves in these panels show Gaussian fits to the error distributions; we also indicate the best-fitting Gaussian means ($\mu_{\rm Gauss}$) and standard deviations ($\sigma_{\rm Gauss}$), as well as the σ_{68} widths (about zero) of the distributions (not of the fits). The Gaussian fits give equal weight to each bin of the distributions and ignore objects for which $\sigma_{\rm NNE} = 0$. We see that the overall normalized error distributions are close to Gaussian for all the catalogues and that there is little difference among the four cases. We conclude that the NNE error estimate is robust even when the training set is not representative and that the weights do not significantly affect the NNE estimator. In retrospect the latter is not too surprising since the NNE estimate is derived from a typically small nearest neighbour region, over which the weights do not vary strongly.

4.3 Deconvolving the photo-z distribution

The photometric redshift distribution is the convolution of the true redshift distribution $N(z_{\rm spec})$ with the distribution of photometric redshift errors. For discrete distributions we can express this as

$$N(z_{\rm phot})_i = \sum_j P(z_{\rm phot}|z_{\rm spec})_{ij} N(z_{\rm spec})_j, \tag{12}$$

where the indices i and j refer to bins of $z_{\rm phot}$ and $z_{\rm spec}$, respectively, and $P(z_{\rm phot}\,|z_{\rm spec})_{ij}$ is the probability that a galaxy has photo-z in bin i given that its spectroscopic redshift is in bin j.

As noted in Padmanabhan et al. (2005), we can solve equation (12) for $N(z_{\rm spec})$ by inverting $P(z_{\rm phot}|z_{\rm spec})_{ij}$. However, the inversion problem is ill-conditioned for two reasons. First, the convolution is a smoothing operation, and some of the information in $N(z_{\rm spec})_j$ is irretrievably lost in that process. Second, small errors in $P(z_{\rm phot}\,|z_{\rm spec})_{ij}$ are magnified by the matrix inversion.

Both problems can be alleviated by using prior information to regularize the inversion and restore some of the lost information. Following Padmanabhan et al. (2005), we use a forward difference operator, defined as

$$S = \sum_{j=0}^{N_{\rm bin}-1} \{[N(z)]_{j+1} - [N(z)]_j\}, \tag{13}$$

as a prior on the smoothness of the reconstruction. To incorporate the prior information into the deconvolution procedure, we must represent the deconvolution as a minimization problem. If we define

$$E_0 \equiv \sum_i \left| P^{-1}(z_{\rm phot}|z_{\rm spec})_{ij}[N(z_{\rm spec})]_j - [N(z_{\rm phot})]_i \right|^2, \tag{14}$$

then the deconvolution can be stated as the problem of minimizing E_0 with respect to $N(z)$. To incorporate the prior, we define

$$E = E_0 + \lambda S, \tag{15}$$

and the regularized deconvolution is achieved by minimizing E. The parameter λ sets how much importance is given to the smoothing and is often chosen ad hoc. Here, following Press et al. (1992), we set

$$\lambda = \frac{\mathrm{Tr}\left[P^{\mathrm{T}}(z_{\mathrm{phot}}|z_{\mathrm{spec}})\,P(z_{\mathrm{phot}}|z_{\mathrm{spec}})\right]}{\mathrm{Tr}\left(B^{\mathrm{T}}\,B\right)}, \tag{16}$$

where B is the $(N_{\mathrm{bin}} - 1)$ (N_{bin}) first difference matrix given by $B = \delta_{(i+1)j} - \delta_{ij}$. This choice of λ gives comparable weight to both parts of the minimization.

The preceding discussion summarizes the 'standard' photo-z deconvolution method for estimating the redshift distribution. The weighting method can provide a better estimate of $P(z_{\mathrm{phot}}|z_{\mathrm{spec}})_{ij}$ for the photometric sample, reducing the need for regularization and thereby improving the deconvolution estimate of $N(z_{\mathrm{spec}})$. We can incorporate the weights into the estimation of $P(z_{\mathrm{phot}}|z_{\mathrm{spec}})_{ij}$ by calculating, for each z_{spec} bin, the z_{phot} distribution for the weighted training-set galaxies.

We postpone discussion of the performance of the deconvolution and weighted deconvolution methods to the next section, where we compare them with direct application of the weighting method to estimation of $N(z_{\mathrm{spec}})$.

5 APPLICATIONS OF THE WEIGHTING METHOD II: ESTIMATES OF $N(z)$ AND $p(z)$ IN MOCK PHOTOMETRIC SAMPLES

5.1 The redshift distribution $N(z)$

We now have at hand a number of methods for estimating the true redshift distribution $N(z)$ for a photometric galaxy sample. Using photo-zs, one can simply use the photo-z distribution itself, $N(z_{\mathrm{phot}})$, as an estimator, or the deconvolved photo-z distribution described in Section 4.3 or the weighted, deconvolved photo-z distribution mentioned at the end of Section 4.3. Alternatively, one can use the weighted spectroscopic redshift distribution of the training-set galaxies to directly estimate $N(z)$, i.e. equation (5), without recourse to photo-zs. Finally, we can sum the redshift probability distributions $p(z)$ for each galaxy in the photometric sample (again estimated from the weighted training set) to estimate $N(z)$, using equation (6). In this section, we compare results of these different estimates of $N(z)$ using the mock SDSS DR6 sample. The results are summarized in Tables 2 and 3 and the best results for each method are shown in Figs 8–10.

5.1.1 Measures of reconstruction quality

To compare the different methods, we need a statistical measure of the quality of the reconstruction of the estimated redshift distribution. We use two. The first is a χ^2 statistic (per degree of freedom and per galaxy), defined here as

$$(\chi^2)^{\mathrm{X}} \equiv \frac{1}{N_{\mathrm{bin}} - 1} \sum_{i=1}^{N_{\mathrm{bin}}} \frac{\left[N(z^i)^{\mathrm{X}} - N(z^i_{\mathrm{spec}})^{\mathrm{P}}\right]^2}{N(z^i_{\mathrm{spec}})^{\mathrm{P}}\Delta z}. \tag{17}$$

Here N_{bin} is the number of redshift bins used, Δz is the width of the bins and $N(z^i)^{\mathrm{X}}$ is equal to $N(z^i_{\mathrm{spec}})^{\mathrm{T}}_{\mathrm{wei}}$ if the weighting procedure is used or to $N(z^i_{\mathrm{phot}})^{\mathrm{P}}$ if the redshift distribution is instead estimated using photo-zs. The usual definition of χ^2 uses the numbers of objects in given bins instead of the normalized probability $N(z^i)$; multiplying our χ^2 by $N_{\mathrm{P,tot}}\,\Delta z$ gives the usual definition. We chose the above statistic so that it is independent of the number of galaxies and the number of redshift bins, allowing us to more fairly compare reconstruction quality across different data sets. Since the probabilities are normalized, the number of degrees of freedom is $N_{\mathrm{bin}} - 1$.

The second measure we employ is a binned version of the Kolmogorov–Smirnov (KS) statistic, defined as the maximum difference between the two cumulative redshift distributions being compared, for example, the cumulative distributions corresponding to $N(z^i_{\mathrm{spec}})^{\mathrm{T}}_{\mathrm{wei}}$ and $N(z^i_{\mathrm{spec}})^{\mathrm{P}}$. The KS statistic is more sensitive to differences in the medians of the two distributions being compared, whereas the χ^2 statistic tends to stress the regions of the distribution that are least well sampled, i.e. regions where $N(z^i)$ is small. In our implementation, we use binned cumulative distributions instead of unbinned cumulative distributions, so this statistic is not strictly the KS statistic.

Note that we do not use the absolute values of these statistics as formal goodness-of-fit measures. Rather, we use their relative values for the different estimators to compare the quality of the different reconstructions – see Table 2.

Table 2. Redshift distribution reconstruction statistics – 30 bins.

Full photometric set	χ^2	KS parameter
Photo-z	0.107	0.0848
Photo-z deconvolution (no weights)	0.577	0.124
Photo-z deconvolution (100 nb)	0.521	0.0989
Weights (100 nb)	0.0341	0.0456
Recoverable photometric set		
Photo-z	0.105	0.0674
Photo-z deconvolution (no weights)	0.499	0.140
Photo-z deconvolution (2 nb)	0.0682	0.0295
Photo-z deconvolution (5 nb)	0.0648	0.0266
Photo-z deconvolution (100 nb)	0.102	0.0351
Weights (2 nb)	0.006 24	0.0129
Weights (5 nb)	0.005 71	0.0145
Weights (100 nb)	0.006 43	0.0246
$p(z)$ (2 nb)	0.005 40	0.0219
$p(z)$ (5 nb)	0.004 93	0.0201
$p(z)$ (100 nb)	0.005 34	0.0241

Note. nb – neighbours.

Table 3. Redshift distribution reconstruction statistics – 20 bins.

Recoverable photometric set	χ^2	KS parameter
Photo-z	0.105	0.0674
Photo-z deconvolution (no weights)	0.404	0.125
Photo-z deconvolution (2 nb)	0.0509	0.0235
Photo-z deconvolution (5 nb)	0.0566	0.0232
Photo-z deconvolution (100 nb)	0.0971	0.0290
Weights (2 nb)	0.004 84	0.0129
Weights (5 nb)	0.004 67	0.013 27
Weights (100 nb)	0.005 47	0.0232

Note. nb – neighbours.

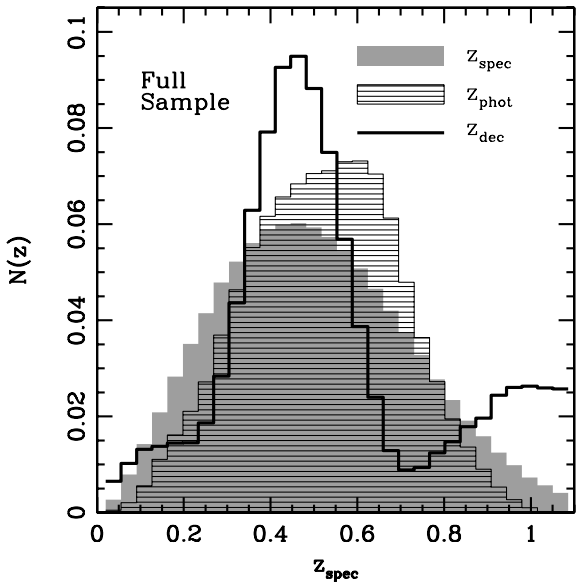

Figure 8. True spectroscopic redshift distribution (solid grey) of the mock SDSS photometric sample, and estimates of the redshift distribution using the photo-z distribution (hatched) and deconvolved photo-z distribution (black line).

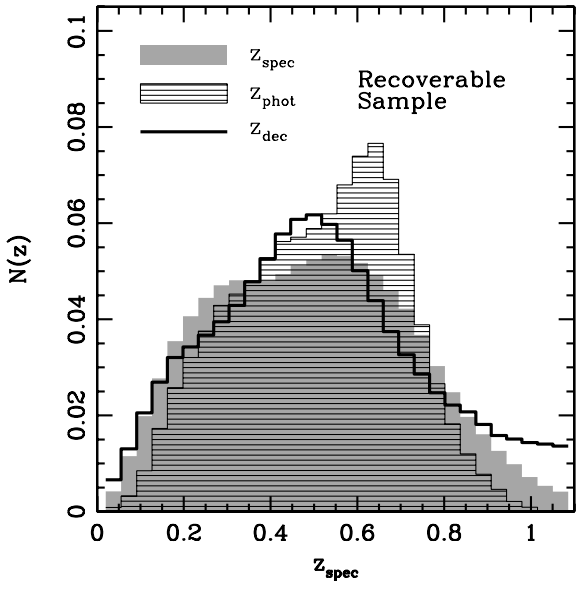

Figure 9. True spectroscopic redshift distribution (solid) of the recoverable mock photometric sample, and estimates of the redshift distribution using the photo-z distribution (hatched) and deconvolved photo-z distribution (black line).

5.1.2 Photo-z estimates of N(z)

The photo-z estimate $N(z_{\rm phot})$ of the true redshift distribution for the mock SDSS photometric sample is shown in Fig. 8 (hatched histogram). We can see that $N(z_{\rm phot})$ underestimates the true distribution, $N(z_{\rm spec})$ (grey histogram), at both low and high redshifts and overestimates it at intermediate redshifts, $0.4 < z_{\rm spec} < 0.8$. In addition, the peak of $N(z_{\rm phot})$ is biased with respect to the peak of $N(z_{\rm spec})$. Comparing the two distributions, we find that $\chi^2 = 0.107$ and KS $= 0.0848$. The photo-z and true redshift distributions for the recoverable photometric sample are shown in Fig. 9 (hatched and

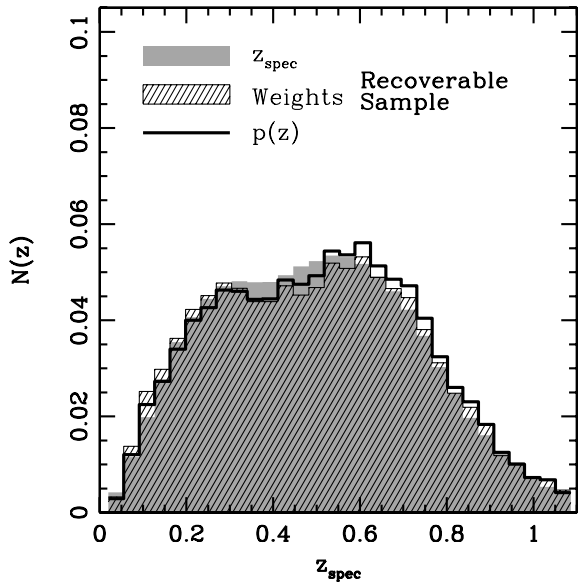

Figure 10. True spectroscopic redshift distribution (solid grey) of the recoverable mock photometric sample, and estimates of the redshift distribution using the weights (hatched) and $p(z)$ (line) methods.

grey histograms). Again, $N(z_{\rm phot})$ underestimates $N(z_{\rm spec})$ at low and high redshifts and overestimates it in between. The reconstruction statistics are similar to those for the full photometric sample, $\chi^2 = 0.105$ and KS $= 0.0674$. This indicates that the faithfulness of $N(z_{\rm phot})$ as an estimate of the true redshift distribution is not very sensitive to whether the training set is representative of the photometric sample: the errors in the recovered redshift distribution are dominated by a systematic effect. The fact that $N(z_{\rm phot})$ is more sharply peaked than $N(z_{\rm spec})$ is a common feature of training-set-based photo-z estimates and results from the breakdown of the fundamental photo-z assumption that a single $z_{\rm phot}$ can represent a full redshift distribution.[1] For the full photometric sample the peak in $N(z_{\rm phot})$ is not as pronounced as it is for the recoverable photometric sample, because the larger photo-z scatter in regions not covered by the training set smoothes out the peak.

We have also tested the photo-z deconvolution method of Section 4.3 as an estimate of the redshift distribution. The standard (unweighted) deconvolution was not successful at recovering $N(z)$, with $\chi^2 = 0.577$, KS $= 0.124$ for the full photometric sample and $\chi^2 = 0.499$, KS $= 0.140$ for the recoverable photometric sample. The result for the *weighted* deconvolution method, where the weights have been estimated using the five nearest neighbours, is shown by the black line in Fig. 8; it is also not very effective for the full photometric sample, with $\chi^2 = 0.521$ and KS $= 0.989$. Although the peak of the deconvolved redshift distribution is at the correct redshift, the distribution shows an oscillatory behaviour with redshift. However, as shown in Fig. 9 (black line versus grey histogram), the weighted deconvolution performs much better for the recoverable photometric sample, with $\chi^2 = 0.0648$ and KS $=$ 0.0266.

The deconvolution estimate of the redshift distribution oscillates about the true distribution. This kind of behaviour is typical of the

[1] Maximum-likelihood template-fitting photo-z methods suffer from a similar problem but with opposite consequences. Because of the different way in which $p(z|\text{observables})$ is estimated in those cases, $N(z_{\rm phot})$ tends to be flatter than the true redshift distribution (Brodwin et al. 2006).

inversion techniques used to perform the deconvolution. It can be alleviated by either increasing the training-set size, decreasing the number of redshift bins, or using prior knowledge to improve the estimate of $P(z_{\rm phot}|z_{\rm spec})$. We briefly investigate the second of these possibilities. As Table 3 shows, using only 20 as opposed to 30 redshift bins improves the deconvolution estimate, $\chi^2 = 0.0509$ and KS $= 0.0235$ (here with weights calculated using the two nearest neighbours). However, fewer bins means coarser redshift information, so it would be preferable to find a method that can accommodate a large number of redshift bins. Table 3 also shows that the other methods are not as sensitive to the number of bins. The deconvolution can also be improved by Monte Carlo resampling the training set (Padmanabhan et al. 2005). Ideally, the resampling should be done in the space of observables used to calculate the photo-zs. However, this approach is prohibitively time consuming for large data sets, and it requires accurate knowledge of the magnitude errors – which may be hard to obtain.

5.1.3 *Weighting method estimates of N(z)*

The direct estimate of the redshift distribution for the photometric sample using the weighting method of equation (5), $N(z)_{\rm wei}$, is shown by the hatched region in Fig. 10. By construction, this estimate is the same for both the full and the recoverable photometric samples, that is, the weighting method in practice provides an estimate of the redshift distribution for the recoverable photometric sample. Comparison with the true redshift distribution of the recoverable sample (solid grey histogram in Fig. 10) shows that the weighting method provides the best redshift distribution estimate of any of the methods under consideration here. For the full photometric sample, $\chi^2 = 0.0341$ and KS $= 0.0456$ (using 100 nearest neighbours), and for the recoverable sample, $\chi^2 = 0.00571$ and KS $= 0.0145$ (with five nearest neighbours). As shown in Table 2, $N(z)_{\rm wei}$ is relatively insensitive to the number of neighbours used in the calculation.

Finally, using the sum of the $p(z)$ estimates for each galaxy in the photometric sample is almost identical to using the weights to estimate $N(z)$. The estimate $N[\sum p(z)]$ of the redshift distribution is shown by the solid black line in Fig. 10, using five nearest neighbours to estimate $p(z)$. For this case, from Table 2 we have $\chi^2 = 0.004\,93$, KS $= 0.0241$ for the recoverable photometric set, quite close to the values for the $N(z)_{\rm wei}$ estimate. Table 2 also shows that using fewer nearest neighbours slightly improves the KS statistic but not the χ^2 statistic. Moreover, by using fewer neighbours, one is unable to accurately characterize $p(z)$, so we caution against using fewer than 100 neighbours in the weighted estimate of $p(z)$.

5.1.4 *Error estimate for N(z$_{\rm wei}$)*

From equations (1) and (2), the errors in the weights depend upon the uncertainties in determining the volumes of the training-set and photometric-set regions around an object and upon the uncertainties in the number of nearest neighbours for both the training and photometric sets. All of these quantities are correlated, making error estimation for the weighting method a challenge. Instead, we apply a bootstrap resampling procedure to directly estimate the errors in the quantity of interest, in this case the weighted estimate of the redshift distribution, $N(z)_{\rm wei}$. We sample with replacement from the training and photometric sets to generate resampled training and photometric sets of the same sizes as the originals. Then, for each pair of resampled training and photometric sets, we calculate the weights using equation (4) and generate $N(z)_{\rm wei}$ using equation (5). We repeat this procedure 10 000 times and estimate the covariance matrix by

$$C(z_\alpha, z_\beta) = \frac{1}{n_{\rm s}-1} \times \sum_{i=1}^{n_{\rm s}} [\hat{N}_i(z_\alpha) - \langle \hat{N}(z_\alpha)\rangle][\hat{N}_i(z_\beta) - \langle \hat{N}(z_\beta)\rangle], \quad (18)$$

where $n_{\rm s}$ is the number of bootstrap samples, $\hat{N}_i(z)$ is the weighted estimate of the redshift distribution in the ith bootstrap sample and $\langle \hat{N}(z)\rangle$ is the mean of the bootstrap estimates. The correlation matrix is defined in the usual way by $\rho(z_\alpha, z_\beta) = C(z_\alpha, z_\beta)/\sigma(z_\alpha)\sigma(z_\beta)$.

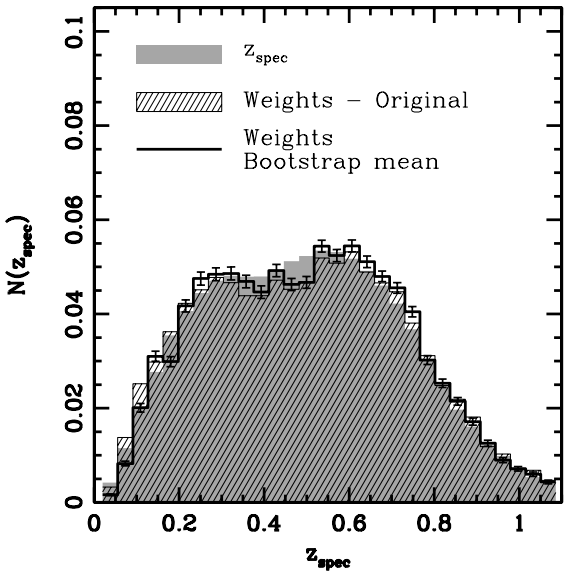

Figure 11. True spectroscopic redshift distribution (solid grey) of the recoverable photometric set, the estimated redshift distribution using the weighting method (hatched region) and the mean of the bootstrap samples for the weighting method (black line). The error bars are given by the square root of the diagonal terms of the covariance matrix calculated from the bootstrap samples.

Fig. 11 shows $N(z)_{\rm wei}$ (hatched), the mean of the bootstrap estimates (solid black) and error bars given by the square root of the diagonal elements of the covariance matrix. There are small anticorrelations between nearby redshift bins, of at most −0.2. Correlations between non-adjacent bins are smaller by at least an order of magnitude.

5.1.5 *Correcting systematic errors in the N(z) estimate*

From Fig. 10, we note that the $N(z_{\rm wei})$ distribution is slightly flatter than $N(z_{\rm spec})$, a feature that also shows up in other catalogues (see e.g. Lima et al. 2008). This smoothing of the redshift distribution is a consequence of using non-negligibly small regions in magnitude space around the training-set galaxies to estimate the weights. This is especially problematic for regions where the training set is sparse, for then the 'neighbour volume' used to calculate the weights may be large compared to the typical scale of change of the redshift/observable hypersurface. The problem is compounded when photometry errors are large, because large errors broaden the redshift distribution in a bin of observables. Broader distributions require a larger number of training-set objects in order to be well characterized, but increasing the number of training-set nearest neighbours in the weights calculation increases the non-locality of

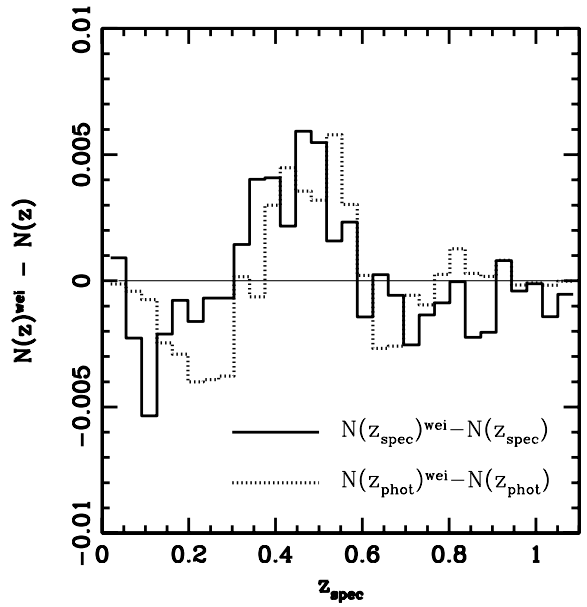

Figure 12. Bias in the weighting method estimate of the redshift distribution for the recoverable photometric set. Solid line shows the bias in the true redshift distribution. Dotted line shows the bias in the weighted photo-z distribution, also for the recoverable photometric set. Since they approximately match, we can use the bias in the weighted photo-z distribution, which is an observable, to estimate the bias in the weighted true redshift distribution.

the estimate. The ideal solution would be to increase the total number of training-set objects in the sample, or at least the number in sparsely covered regions, but that is not always an option. The poor man's alternative is to develop ways to characterize and correct for the systematic errors.

An empirical approach we have developed makes use of the photometric redshifts in the following way. Starting with the training set, compute the photo-z distribution of the *weighted* training set, $N(z_{\rm phot})_{\rm wei}$, i.e. use equation (5) but with z replaced by $z_{\rm phot}$ everywhere. The difference between $N(z_{\rm phot})_{\rm wei}$ and the photo-z distribution for the photometric sample, $N(z_{\rm phot})$, is shown by the dotted line in Fig. 12. The bias we are actually interested in is $N(z_{\rm spec})_{\rm wei} - N(z_{\rm spec})$, shown by the solid line in Fig. 12. We see that these two differences have similar behaviour with redshift, presumably due to similar non-locality of the weight solution in regions where the training set is sparse. We can therefore use $N(z_{\rm phot})_{\rm wei} - N(z_{\rm phot})$, the bias in the weighted photo-z distribution and which is an observable for the photometric sample, to estimate $N(z_{\rm spec})_{\rm wei} - N(z_{\rm spec})$, the systematic error in the weighted estimate of the true redshift distribution. The redshift distribution estimate can then be approximately corrected for this bias.

To reduce the effect of random errors in the estimation of the bias, we smooth $N(z_{\rm phot})_{\rm wei} - N(z_{\rm phot})$ using a 'moving window' method. Each redshift window has width greater than half of the separation between window centroids. The smoothing factor is the ratio of the window size to the redshift bin size when no smoothing is used. We have used smoothing factors of 1, 2, 3 and 5 to calculate $N(z_{\rm phot})_{\rm wei} - N(z_{\rm phot})$. A smoothing factor of 1 corresponds to a window size of 0.0367 in redshift. We picked the other smoothing factors based on the natural scales set by the σ and σ_{68} of the photo-zs in the training and photometric sets.

Table 4 shows the recovery statistics for the distributions corrected for systematics in this way, and Fig. 13 shows the improvement in the $N(z)$ estimate when the correction with smoothing factor of 2 is applied. While these results are suggestive, more testing should be done before adopting this method as a correction for systematic errors in practice.

Table 4. Redshift distribution reconstruction statistics – correction of systematics.

Recoverable photometric set – 5 neighbours – $0 < z < 1.1$		
Smoothing factor	χ^2	KS parameter
No correction	0.005 71	0.0145
Unsmoothed	0.004 87	0.0151
2	0.003 51	0.0127
3	0.003 49	0.0134
5	0.003 55	0.0131
Bootstrap mean (no correction)	0.006 00	0.0189

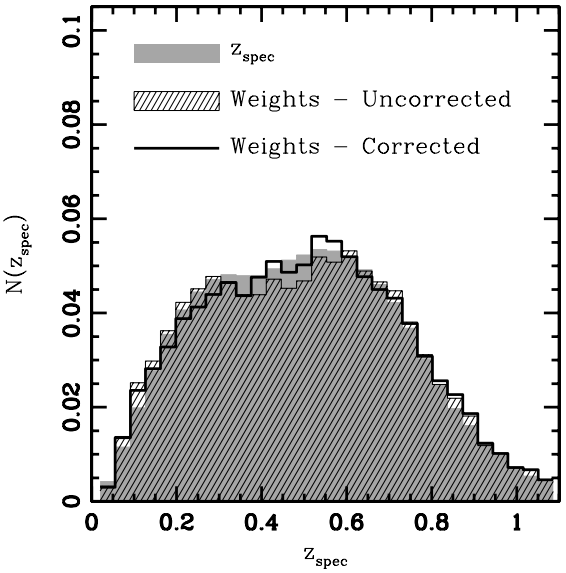

Figure 13. True spectroscopic redshift distribution (solid grey) of the recoverable photometric set, and the estimated redshift distribution using the weighting method, showing both the uncorrected results (hatched) and results corrected for systematic errors (black line) as described in the text.

5.2 The probability distribution $p(z)$

In this section we examine the effectiveness of the weighted training set in estimating the redshift probability distribution $p(z)$ for individual galaxies, and the relation between $p(z)$, $z_{\rm phot}$ and $z_{\rm spec}$. For this study, we have increased the size of the mock photometric set to 9000 000 galaxies in order to improve the statistics. As before, we calculate the training-set estimate of $p(z)$, hereafter $p(z_{\rm train})$, for a training-set galaxy by selecting its 100 nearest neighbours in the training set. The spectroscopic redshift distribution of these objects is $p(z_{\rm train})$. We then select all the galaxies in the photometric sample that are closer to the given galaxy in magnitude space than its 100th-nearest training-set neighbour. The spectroscopic redshift distribution of the selected photometric galaxies is, barring statistical fluctuations and non-locality, the true redshift distribution, hereafter $p(z_{\rm true})$, of the region of observable space centred about the selected galaxy.

In Fig. 14 we show the redshift distributions for three galaxies. In each panel, $p(z_{\rm true})$ is shown as a grey histogram with 60 bins, and

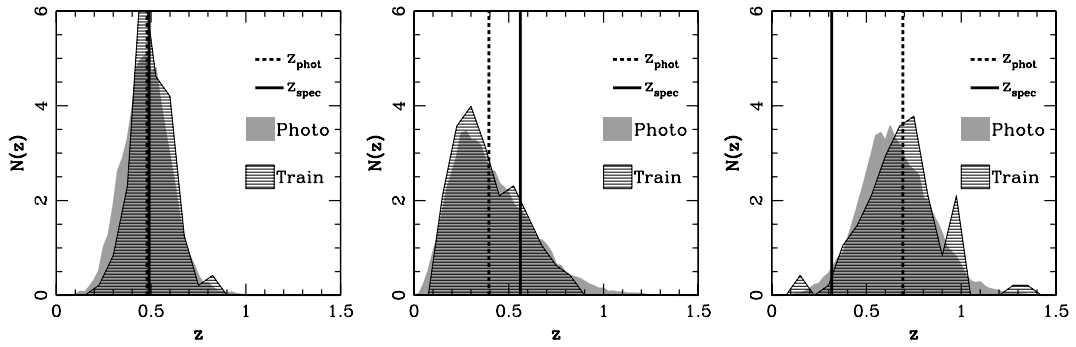

Figure 14. Distributions of $p(z_{\rm true})$ (solid grey histograms) and $p(z_{\rm train})$ (hatched histograms) for three training-set galaxies in the mock SDSS sample. The vertical solid (dashed) lines indicate $z_{\rm spec}$ ($z_{\rm phot}$) for each galaxy. Left: an early-type galaxy at $z = 0.48$; middle: a late-type galaxy at $z = 0.56$; right: a faint, early-type galaxy at $z = 0.31$.

$p(z_{\rm train})$ is shown as the hatched histogram with 20 bins. We have rescaled the histograms by multiplying each by the width of the histogram bin for easier comparison of the distributions. The solid vertical line indicates the true redshift of the galaxy and the dashed vertical line indicates its ANN $z_{\rm phot}$ estimate. The left-hand panel of the figure is for an early-type ($T = 1.5$) galaxy with r magnitude of 20.67 and $z_{\rm spec} = 0.48$. This galaxy has 4006 neighbours in the photometric sample, i.e. that many photometric objects are as close to it in magnitude space as its 100 nearest training-set neighbours. In this example, the true redshift distribution of this region of observable space is narrow, $p(z_{\rm train})$ is a quite accurate estimate of $p(z_{\rm true})$, $z_{\rm phot}$ is very near $z_{\rm spec}$ and both are at the peak of the $p(z)$ distributions.

The middle panel shows the distributions for a late-type ($T = 3.1$) galaxy with $r = 21.1$ and $z_{\rm spec} = 0.56$. There were 14 606 neighbours to this galaxy in the photometric sample. With the exception of the extreme tails of the distribution, $p(z_{\rm train})$ provides an accurate estimate of $p(z_{\rm true})$. The redshift PDF $p(z_{\rm true})$ for this galaxy is much broader than that for the galaxy in the left panel, in part because the magnitudes of late-type galaxies do not correlate with redshift as well as those of early types. The neural network photo-z is 0.39 for this object, higher than the peak of $p(z_{\rm train})$ at $z = 0.3$ or its median at $z = 0.34$. The true redshift of this object, $z_{\rm spec} = 0.56$, is far removed from the peak of its redshift distribution. However, the photo-z error, $z_{\rm phot} - z_{\rm spec} = 0.16$, is comparable to the photo-z scatter at this redshift, $\sigma(z_{\rm spec} = 0.56) \sim 0.13$ (see bottom right-hand plot of Fig. 5), which shows that this example is not atypical. The broader $p(z_{\rm true})$ is, the more likely it is that $z_{\rm spec}$ will be far from the peak of the distribution. In that case, the photo-z estimator cannot zero in on the correct redshift, and a single-point $z_{\rm phot}$ estimate will be a poor redshift estimate for a large fraction of the objects in this region of observable space.

The right-hand panel of Fig. 14 shows the distributions for another early-type ($T = 1.4$) galaxy with $r = 21.8$ and $z_{\rm spec} = 0.31$, with 18 366 neighbours in the photometric set. This is the most pathological of the three examples. The large width of $p(z_{\rm true})$ for this galaxy is due to its faintness, which results in large magnitude errors. The peaks of $p(z_{\rm train})$ and $p(z_{\rm true})$ are offset by ~0.1–0.2, and $p(z_{\rm train})$ shows a spurious second peak at $z \sim 1$. Such fluctuations are not uncommon when one uses 100 galaxies to estimate $p(z)$. The true redshift of this galaxy is at the low-redshift tail of $p(z_{\rm true})$, and $z_{\rm phot}$ for this object is catastrophically wrong even though it is near the peak of $p(z_{\rm true})$. The catastrophic error results from using a single number to represent a very broad distribution, and in this case the galaxy in question is quite different from most of its neighbours in magnitude space. For a photometric survey, the redshift distribution is typically broad near the photometric limit of the survey. To avoid catastrophic errors and biases, one should work with the full redshift probability distribution per object.

6 APPLICATION TO SDSS DR6 DATA

Now that we have tested the weighting method on mock SDSS photometric samples, we apply it to the actual SDSS DR6 photometric sample.

6.1 Bias and scatter in SDSS photo-*z*s

Oyaizu et al. (2008a) estimated photo-zs for the SDSS DR6 photometric sample using an ANN (see Appendix C) and several different combinations of photometric observables. One version, denoted there by D1, used as input observables the five magnitudes *ugriz* and five concentration indices, also splitting the training set and the photometric sample into five bins of r magnitude and performing separate ANN fits in each bin. Version CC2 used as inputs the four colours $u - g$, $g - r$, $r - i$, $i - z$, plus the concentration indices in g, r and i. Here, as in Section 4.1, we use the weighting method to obtain improved estimates of the bias and scatter of these photo-z estimates. Fig. 15 shows the weighted and unweighted $b(z)$ and $\sigma(z)$ estimates derived from the training set, along with third-order polynomial fits to the weighted estimates. The polynomial fit coefficients are given in Table 5. The differences between the weighted and unweighted $b(z)$ and $\sigma(z)$ curves are qualitatively consistent with the results on the mock sample (Fig. 6), but the real data have larger scatter and bias than the mocks.

6.2 The SDSS redshift distribution

Fig. 16 shows the weighting method estimate, $N(z)_{\rm wei}$, for the redshift distribution of the SDSS DR6 photometric sample with $r < 22$. The error bars on $N(z)_{\rm wei}$ are given by the square root of the diagonal elements of the covariance matrix obtained by the bootstrap resampling procedure described in Section 5.1.4.

The coarse-grained structure of the redshift distribution is similar to that of the mock SDSS sample (Fig. 3). However, the fine-grained structure shows peaks and dips that the study of Section 5.1.5 suggests are indications of systematic error. As noted there, large photometric errors, combined with sparseness of the training set, can lead to distortions of the inferred redshift distribution. This effect is likely present in the weighted estimate of the SDSS DR6 redshift distribution for galaxies with $r < 22$. The bump in $N(z)_{\rm wei}$ around $z = 0.75$ is the result of the magnification of the sampling errors in

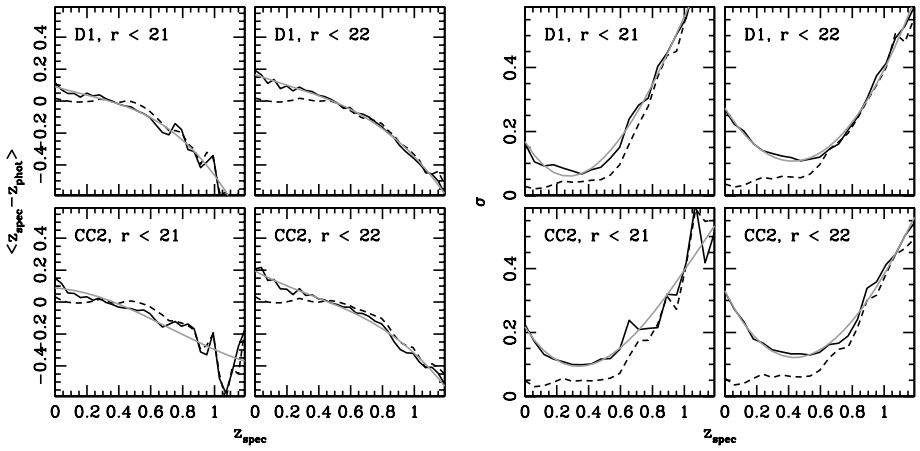

Figure 15. Left panels: estimated photo-z bias versus redshift for the weighted and unweighted training set of the SDSS DR6 catalogue for four cases: (top left) D1 photo-zs with $r < 21$, (top right) D1 photo-zs with $r < 22$, (bottom left) CC2 photo-zs with $r < 21$ and (bottom right) CC2 photo-zs with $r < 22$. Right-hand panels: estimated photo-z scatter versus redshift for the weighted and unweighted training set of the SDSS DR6 catalogue for the same cases depicted in the left-hand panels. In each plot the dashed line corresponds to the unweighted result, the solid dark line to the weighted result and the solid red line is a third-order polynomial fit to the weighted result. The fit coefficients are given in Table 5.

Table 5. Fit coefficients to the weighted estimates of photo-z bias and scatter versus redshift for SDSS DR6 catalogue.

	$r < 21$	$r < 22$
	D1 photo-zs	
$b(z)$	[0.090 0269, −0.293 255, 0.262 842, −0.523 857]	[0.165 74, −0.350 82, 0.192 806, −0.355 683]
$\sigma(z)$	[0.167 949, −0.823 95, 1.698 19, −0.484 006]	[0.273 305, −0.788 055, 0.951 591, −0.042 6683]
	CC2 photo-zs	
$b(z)$	[0.088 4344, −0.057 4277, −0.607 687, 0.279 678]	[0.193 711, −0.527 042, 0.421 479, −0.408 717]
$\sigma(z)$	[0.217 213, −0.776 92, 1.360 55, −0.406 967]	[0.329 747, −1.000 9, 1.316 67, −0.262 655]

Note. All fits are third-order polynomials of the form $a_1 + a_2 z + a_3 z^2 + a_4 z^3$.

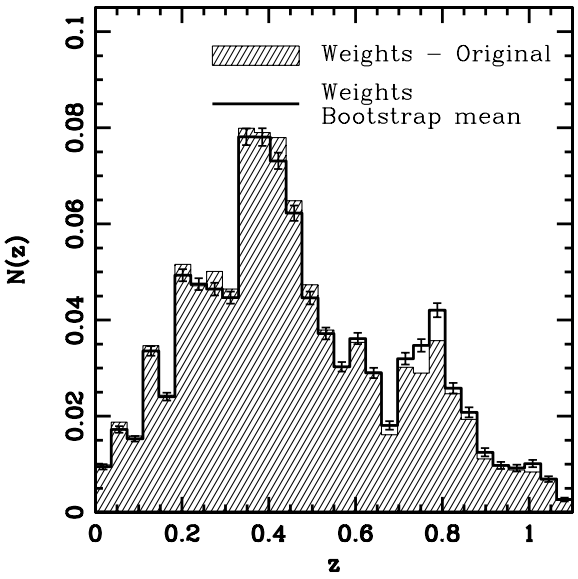

Figure 16. Estimated redshift distribution for the SDSS DR6 sample (with $r < 22$), computed using the weighting method (hatched) and the mean of the bootstrap samples (solid line). The error bars are the diagonal bootstrap errors.

the training set caused by the lack of redshift information in the photometry of faint galaxies, combined with the lack of training-set coverage in that redshift range.

When we impose more stringent r-magnitude cuts, Fig. 17 (left) shows that the feature disappears. In Fig. 17 (right) we show the $N(z)$ distribution estimated using $p(z)$ using two different training sets. In one case we use the full training set to estimate the $p(z)$s while in the other we remove all galaxies from DEEP/DEEP2 and 2SLAQ (totalling 84 568 galaxies) from the training set and we add 6069 from two approximately flux-limited samples, DEEP2-EGS (Davis et al. 2007) and zCOSMOS (Lilly et al. 2007), which we describe in further detail in Section 6.3. The bump at $z = 0.75$ disappears when DEEP/DEEP2 is not included, showing that the selection in DEEP/DEEP2, which was done to target $z \sim 0.7$ galaxies and in a different photometric system from SDSS is responsible for the bump. The effects of 2SLAQ are much less pronounced, and consist in a small overall shift of the distribution. 2SLAQ has morphology cuts (in addition to the SDSS *ugriz* magnitude cuts) which could have yielded some systematic biases. As mentioned previously, the selection effects are amplified by the photometry errors, so that the systematics are reduced if one imposes more stringent magnitude cuts. If one is primarily interested in the overall redshift distribution, the $N(z)$ estimate using the training set without DEEP/DEEP2 or 2SLAQ is more reliable. However, if one requires redshift

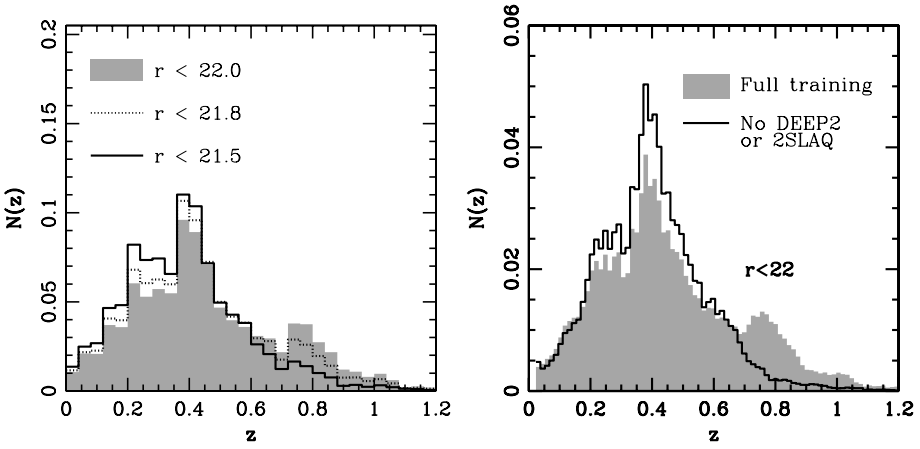

Figure 17. Left: weighted estimates of the redshift distribution for the SDSS DR6 photometric sample, with $r < 21.5$, $r < 21.8$ and $r < 22$. Right: $p(z)$ estimates of redshift distribution for SDSS DR7 photometric sample using the full training set as well as a training set without the DEEP2 (the non-EGS part) and without 2SLAQ.

information for individual galaxies, the estimate using the full training set is still preferable. Without DEEP/DEEP2 and 2SLAQ, the training set is too sparse at faint magnitudes. As a result, the individual $p(z)$ estimates are derived using training-set objects spread out over a large region of observable space, which makes the $p(z)$s poor representations of the *local* redshift distributions around the corresponding galaxies. Given the training sets available, the best way to reduce the effects of selection issues while having reliable $p(z)$ estimates is to perform magnitude cuts.

We see another feature in $N(z)_{\rm wei}$ in the range $0.2 < z < 0.4$ that does not go away with tighter r-magnitude cuts (see Fig. 17). Similar features can be seen in the zCOSMOS+DEEP2/EGS redshift distribution used by Mandelbaum et al. (2008) (see the bottom right-hand panel of fig. 4 of Mandelbaum et al. 2008), in the CNOC2 distribution used in our training set (see fig. 2 in Lima et al. 2008) and in the full CNOC2 sample shown in Lin et al. (1999). The feature in the DEEP2 data appears to be caused, at least partially, by spectroscopic failures affecting both early- and late-type galaxies in that redshift range (J. Newman, private communication), and it is possible that this is affecting the weighted estimate. In general, one should not expect that the redshift distribution of a sample flux limited in one filter will be smooth, due to k-correction-like effects. The complex shape of the SEDs of galaxies implies that a flux limit based on a single filter will preferentially select certain galaxy types at certain redshifts. We do not see such a feature in the mock SDSS catalogue, because the mock was *created* with a smooth r magnitude distribution and redshift distribution, and we only applied a cut in the r band.

6.3 A $p(z)$ catalogue for SDSS DR7

We calculated $p(z)$s for the full SDSS DR7 sample satisfying the selection cuts of the Photoz2 photometric redshift table described in Appendix A1 and in Oyaizu et al. (2008a) – a total of 78 135 961 galaxies. We added a sample of 4241 galaxies with spectra from zCOSMOS with quality flags 2.5, 3.4, 3.5, 4.4, 4.5, 9.3, 9.4, 9.5 (Lilly et al. 2007) and 1828 galaxies from DEEP2-EGS (Davis et al. 2007) with $z_{\rm quality} \geq 3$ to the training set. We do not use the ubercalibrated magnitudes (Padmanabhan et al. 2008) available for DR7 because these were not available for most of our training set galaxies. The catalogue is available from the SDSS DR7 value-added catalogues website.[2] There are 240 files, ordered by RA, one for every 0.1 h of RA from 0 to 23 h Thus, the file named pofz.ra12h3.dat has photo-zs and $p(z)$s for objects with $12.3 \leq$ RA < 12.4 h, and so forth. The $p(z)$ values are tabulated for 100 redshift bins, centred at $z = 0.03$ to 1.47, with redshift spacing $dz = 1.44/99$. To reduce effects of Poisson noise we adopt a 'moving window' smoothing technique. Each entry for a given $p(z)$ is calculated based on a bin of width $4\,dz$. As discussed in Section 6.2, the quality of the estimates degrades rapidly for $r > 21.5$. We therefore recommend a cut in brightness of at least $r < 21.8$.

7 DISCUSSION AND FUTURE WORK

We have extended and applied the weighting technique of estimating redshift distributions (Lima et al. 2008). The weighting procedure allows one to use a spectroscopic training set to accurately estimate the bias and scatter of photo-zs as a function of redshift. In addition, the weighting method provides a natural, robust way to select galaxies in the photometric sample that are well represented in the training set. Moreover, we have shown that the weighting technique provides a precise estimate of the redshift distribution of a photometric sample in the region of observable space where the training set and the photometric sample intersect. The estimate $N(z)_{\rm wei}$ more accurately estimates the redshift distribution for a photometric sample than methods based on photo-zs. We have also extended the weighting method to estimate the redshift PDF for individual galaxies, $p(z)$. Use of this PDF can substantially reduce biases associated with the use of single-point photo-zs, and we recommend its use in the analysis of future photometric galaxy surveys.

We have outlined the potential different sources of error of the weights technique and we have demonstrated how to use information from the photo-z distribution to reduce systematic errors in the weights. We have shown that for the SDSS DR7, selection effects in the training set are the dominant source of error in the estimation of $N(z)$, and that this systematic increases sharply with r magnitude. In particular, we have found that the selection of the DEEP2 survey, which uses a different set of filters from the SDSS, is the dominant source of systematic errors.

[2] http://www.sdss.org/dr7/products/value_added/index.html

We have made public a catalogue of $p(z)$ for ∼78 million SDSS DR7 galaxies. We have also provided fitting functions for the weights-based estimates of the bias and scatter of photo-zs as a function of redshift for the D1 and CC2 photo-zs of the SDSS DR6.

For the future, investigations of the weighting method should include study of optimizing the weights estimation, e.g. with a variable number of nearest neighbours in different regions of observable space, and inclusion of systematic effects, e.g. associated with LSS and spectroscopic failures, in the mock catalogues.

ACKNOWLEDGMENTS

We acknowledge useful conversations with Jeff Newman. CEC would like to thank Rachel Mandelbaum and Reiko Nakajima for extensive testing of the SDSS DR7 p(z) catalogue. This work was supported by the KICP under NSF Nos PHY-0114422 and PHY-0551142, by NSF grants AST-0239759, AST-0507666 and AST-0708154 at the University of Chicago, by the DOE at the University of Chicago and Fermilab, and by DOE contract number DE-AC02-07CH11359.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The SDSS website is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington.

REFERENCES

Adelman-McCarthy J. K. et al., 2008, ApJS, 175, 297
Blake C., Bridle S., 2005, MNRAS, 363, 1329
Blanton M. R. et al., 2003, AJ, 125, 2348
Brodwin 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
Cannon R. et al., 2006, MNRAS, 372, 425
Coleman G. D., Wu C. C., Weedman D. W., 1980, ApJS, 43, 393
Collister A. A., Lahav O., 2004, PASP, 116, 345
Davis M., Newman J. A., Faber S. M., Phillips A. C., 2001, in Cristiani S., Renzini A., Williams R. E., eds, Proc. ESO Workshop, The DEEP2 Redshift Survey. Springer-Verlag, Berlin, p. 241
Davis M. et al., 2007, ApJ, 660, L1
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Gunn J. E. et al., 2006, AJ, 131, 2332
Hogg D. W., Finkbeiner D. P., Schlegel D. J., Gunn J. E., 2001, AJ, 122, 2129
Huterer D., Kim A., Krauss L. M., Broderick T., 2004, ApJ, 615, 595
Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101
Lilly S. J., Le Fevre O., Crampton D., Hammer F., Tresse L., 1995, ApJ, 455, 50
Lilly S. J. et al., 2007, ApJS, 172, 70
Lima M., Hu W., 2007, Phys. Rev. D, 76, 123013
Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118
Lin H., Yee H. K. C., Carlberg R. G., Morris S. L., Sawicki M., Patton D. R., Wirth G., Shepherd C. W., 1999, ApJ, 518, 533
Lupton R., Gunn J. E., Ivezić Z., Knapp G. R., Kent S., 2001, in Harnden F. R., Jr, Primini F. A., Payne H. E., eds, ASP Conf. Ser. Vol. 238, Astronomical Data Analysis Software and Systems X. Astron. Soc. Pac., San Francisco, p. 269
Ma Z., Hu W., Huterer D., 2006, ApJ, 636, 21
Mandelbaum R. et al., 2008, MNRAS, 386, 781
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
Padmanabhan N. et al., 2005, MNRAS, 359, 237
Padmanabhan N. et al., 2008, ApJ, 674, 1217
Pier J. R., Munn J. A., Hindsley R. B., Hennessy G. S., Kent S. M., Lupton R. H., Ivezić Ž., 2003, AJ, 125, 1559
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in C: The Art of Scientific Computing. Cambridge Univ. Press, Cambridge
Smith J. A. et al., 2002, AJ, 123, 2121
Stoughton C. et al., 2002, AJ, 123, 485
Tucker D. L. et al., 2006, Astron. Nachr., 327, 821
Weiner B. J. et al., 2005, ApJ, 620, 595
Wirth G. D. et al., 2004, AJ, 127, 3121
Yee H. K. C. et al., 2000, ApJS, 129, 475
York D. G. et al., 2000, AJ, 120, 1579
Zhan H., 2006, J. Cosmol. Astropart. Phys., 8, 8
Zhan H., Knox L., 2006, ApJ, 644, 663
Zucca E. et al., 2006, A&A, 455, 879

APPENDIX A: SDSS DR6 DATA SAMPLE

A1 Photometric set

The SDSS comprises a large-area imaging survey of the north Galactic cap, a multi-epoch imaging survey of an equatorial stripe in the south Galactic cap and a spectroscopic survey of roughly 10^6 galaxies and 10^5 quasars (York et al. 2000). The survey used a dedicated, wide-field, 2.5-m telescope (Gunn et al. 2006) at Apache Point Observatory, New Mexico. Imaging was carried out in drift-scan mode using a 142 mega-pixel camera (Gunn et al. 2006) that gathers data in five broad-bands, $ugriz$, spanning the range from 3000 to 10 000 Å (Fukugita et al. 1996), with an effective exposure time of 54.1 s band^{-1}. The images were processed using specialized software (Lupton et al. 2001; Stoughton et al. 2002) and were astrometrically (Pier et al. 2003) and photometrically (Hogg et al. 2001; Tucker et al. 2006) calibrated using observations of a set of primary standard stars (Smith et al. 2002) observed on a neighbouring 20-inch telescope.

The imaging in the SDSS DR6 (Adelman-McCarthy et al. 2008) covers a nearly contiguous region of the north Galactic cap. In any region where imaging runs overlap, one run was declared primary[3] and was used for spectroscopic target selection; other runs were declared secondary. The area covered by the DR6 primary imaging survey, including the southern stripes, is 8520 deg^2, but DR6

[3] For the precise definition of primary objects see http://cas.sdss.org/dr6/en/help/docs/glossary.asp

includes both the primary and secondary observations of each area and source (Adelman-McCarthy et al. 2008).

In this paper, we use a random 1 per cent subset of the SDSS DR6 Photoz2 catalogue described in Oyaizu et al. (2008a) as our photometric sample. The Photoz2 catalogue contains all primary objects from DR6 (drawn from the SDSS CasJobs website[4]) that have the TYPE flag equal to 3 (the type for galaxy) and that do not have any of the flags BRIGHT, SATURATED, SATUR_CENTER or NOPETRO_BIG set. For the definitions of these flags we refer the reader to the PHOTO flags entry at the SDSS website.[5] The full Photoz2 photometric sample comprises 77 418 767 galaxies. The r magnitude, $g - r$ and $r - i$ colour distributions are shown in the bottom panels of Figs 1(a) and 2(a).

A2 Spectroscopic training samples

As noted in the text, the spectroscopic training sample we use for SDSS DR6 is drawn from a number of spectroscopic galaxy catalogues that overlap with SDSS imaging. Each survey providing spectroscopic redshifts defines a redshift quality indicator; we refer the reader to the respective publications listed below for their precise definitions. For each survey, we chose a redshift quality cut roughly corresponding to 90 per cent redshift confidence or greater. The SDSS spectroscopic sample provides 531 672 redshifts, principally from the MAIN and LRG samples, with confidence level $z_{\rm conf} > 0.9$. The remaining redshifts are 21 123 from the CNOC2 (Yee et al. 2000), 1830 from the CFRS (Lilly et al. 1995) with class >131 716 from the DEEP (Davis et al. 2001) with q_z = A or B and from DEEP2 (Weiner et al. 2005)[6] with $z_{\rm quality} \geq 3728$ from the TKRS (Wirth et al. 2004) with $z_{\rm quality} > -1$, and 52 842 LRGs from the 2SLAQ Survey (Cannon et al. 2006)[7] with $z_{\rm op} \geq 3$.

We positionally matched the galaxies with spectroscopic redshifts against photometric data in the SDSS *BestRuns* CAS data base, which allowed us to match with photometric measurements in different SDSS imaging runs. The above numbers for galaxies with redshifts count independent photometric measurements of the same objects due to multiple SDSS imaging of the same region; in particular SDSS Stripe 82 has been imaged a number of times. The numbers of *unique* galaxies used from these surveys are 1435 from CNOC2, 272 from CFRS, 6049 from DEEP and DEEP2, 389 from TKRS and 11 426 from 2SLAQ. The SDSS spectroscopic samples were drawn from the SDSS primary galaxy sample and therefore are all unique.

APPENDIX B: SDSS DR6 MOCK CATALOGUE

Using spectral template libraries and observational data on the redshift-dependent luminosity functions of galaxies of different types, we have constructed mock photometric and spectroscopic samples that reproduce the main features of the real SDSS DR6 samples. In particular, we fit simple polynomial functions to the Schechter parameters of Zucca et al. (2006) to derive a continuous relationship between the Schechter parameters M^*, α, ϕ^*, redshift z and galaxy type T, using the centroid of each redshift bin for the fit. To regularize the fits, we visually extrapolate the results of Zucca et al. (2006) to the $z = (0, 0.2)$ bin and, where needed, for the (1.2, 1.5) bin.

Table B1. Schechter luminosity function parameters (Zucca et al. 2006) used to derive polynomial fits to the relationships between the Schechter luminosity function parameters, redshift and galaxy spectral type. The parameters in Zucca et al. (2006) were derived using the B band of the VVDS; here we use them to generate the r-band magnitude distributions, using the appropriate k-corrections by galaxy type (Blake & Bridle 2005).

Type	z bin	α	$M^*_{\rm AB} - 5\log(h)$	ϕ^* ($10^{-3}\,h^3\,{\rm Mpc}^{-3}$)
1	0.0–0.2	$-0.15^{+0.30}_{-0.30}$	$-20.00^{+0.30}_{-0.30}$	$6.15^{+0.70}_{-0.70}$
1	0.2–0.4	$-0.04^{+0.28}_{-0.27}$	$-20.27^{+0.27}_{-0.31}$	$5.15^{+0.64}_{-0.64}$
1	0.4–0.6	$-0.40^{+0.20}_{-0.20}$	$-20.49^{+0.17}_{-0.18}$	$3.12^{+0.30}_{-0.30}$
1	0.6–0.8	$-0.22^{+0.17}_{-0.17}$	$-20.22^{+0.09}_{-0.10}$	$3.53^{+0.25}_{-0.25}$
1	0.8–1.0	$-0.01^{+0.25}_{-0.24}$	$-20.73^{+0.11}_{-0.12}$	$2.36^{+0.18}_{-0.18}$
1	1.0–1.2	$-1.23^{+0.34}_{-0.34}$	$-20.53^{+0.11}_{-0.12}$	$2.39^{+0.22}_{-0.22}$
1	1.2–1.5	$-1.30^{+0.40}_{-0.40}$	$-20.50^{+0.30}_{-0.30}$	$2.3^{+0.30}_{-0.30}$
2	0.0–0.2	$-0.60^{+0.20}_{-0.20}$	$-20.00^{+0.20}_{-0.20}$	$7.60^{+0.90}_{-0.90}$
2	0.2–0.4	$-0.67^{+0.13}_{-0.13}$	$-20.13^{+0.19}_{-0.21}$	$6.50^{+0.56}_{-0.56}$
2	0.4–0.6	$-0.50^{+0.15}_{-0.14}$	$-19.97^{+0.12}_{-0.12}$	$4.35^{+0.31}_{-0.31}$
2	0.6–0.8	$-0.57^{+0.13}_{-0.13}$	$-20.39^{+0.09}_{-0.10}$	$4.58^{+0.26}_{-0.26}$
2	0.8–1.0	$-0.60^{+0.20}_{-0.20}$	$-20.55^{+0.10}_{-0.11}$	$3.54^{+0.22}_{-0.22}$
2	1.0–1.2	$-0.76^{+0.34}_{-0.33}$	$-20.77^{+0.12}_{-0.13}$	$3.01^{+0.23}_{-0.23}$
2	1.2–1.5	$-1.57^{+0.61}_{-0.62}$	$-20.82^{+0.13}_{-0.14}$	$2.19^{+0.22}_{-0.22}$
3	0.0–0.2	$-0.80^{+0.30}_{-0.30}$	$-19.00^{+0.60}_{-0.60}$	$10.0^{+0.60}_{-0.60}$
3	0.2–0.4	$-0.84^{+0.10}_{-0.10}$	$-19.14^{+0.12}_{-0.13}$	$9.82^{+0.54}_{-0.54}$
3	0.4–0.6	$-1.07^{+0.10}_{-0.10}$	$-20.04^{+0.11}_{-0.11}$	$6.31^{+0.30}_{-0.30}$
3	0.6–0.8	$-0.79^{+0.13}_{-0.13}$	$-20.10^{+0.09}_{-0.09}$	$7.11^{+0.29}_{-0.29}$
3	0.8–1.0	$-0.87^{+0.15}_{-0.15}$	$-20.33^{+0.08}_{-0.08}$	$6.27^{+0.27}_{-0.27}$
3	1.0–1.2	$-1.39^{+0.26}_{-0.26}$	$-20.38^{+0.10}_{-0.10}$	$5.57^{+0.33}_{-0.33}$
3	1.2–1.5	$-1.86^{+0.55}_{-0.59}$	$-20.81^{+0.12}_{-0.13}$	$3.67^{+0.27}_{-0.27}$
4	0.0–0.2	$-1.55^{+0.20}_{-0.20}$	$-19.60^{+0.40}_{-0.40}$	$2.60^{+0.40}_{-0.40}$
4	0.2–0.4	$-1.59^{+0.11}_{-0.12}$	$-19.73^{+0.29}_{-0.33}$	$2.59^{+0.13}_{-0.13}$
4	0.4–0.6	$-1.53^{+0.18}_{-0.19}$	$-19.38^{+0.17}_{-0.18}$	$4.10^{+0.19}_{-0.19}$
4	0.6–0.8	$-1.35^{+0.15}_{-0.15}$	$-19.95^{+0.12}_{-0.12}$	$4.07^{+0.16}_{-0.16}$
4	0.8–1.0	$-1.68^{+0.20}_{-0.21}$	$-20.10^{+0.12}_{-0.12}$	$4.72^{+0.20}_{-0.20}$
4	1.0–1.2	$-1.99^{+0.33}_{-0.34}$	$-20.19^{+0.12}_{-0.12}$	$6.95^{+0.36}_{-0.36}$
4	1.2–1.5	$-2.50^{+0.52}_{-0.91}$	$-20.53^{+0.12}_{-0.12}$	$4.34^{+0.32}_{-0.32}$

The Schecter luminosity function is defined as

$$\phi(M)\,{\rm d}M = \frac{2}{5}\phi^*(\ln 10)\left[10^{(2/5)(M^*-M)}\right]^{\alpha+1} \times \exp\left[-10^{(2/5)(M^*-M)}\right]\,{\rm d}M, \quad \text{(B1)}$$

where $\phi(M)\,{\rm d}M$ is the number of galaxies with absolute magnitudes between M and $M + {\rm d}M$.

The Schechter parameters we use are shown in Table B1. The polynomials we derive are

$$\alpha = b_1 T^2 + b_2 T z + b_3 z + b_4 z^2 + b_5, \quad \text{(B2)}$$

$$M^* = c_1 T^2 + c_2 T z + c_3 z + c_4 z^2 + c_5, \quad \text{(B3)}$$

[4] http://casjobs.sdss.org/casjobs/
[5] http://cas.sdss.org/dr6/en/help/browser/browser.asp
[6] http://deep.berkeley.edu/DR2/
[7] http://lrg.physics.uq.edu.au/New_dataset2/

$$\phi^* = d_1 T^2 + d_2 T z + d_3 z + d_4 z^2 + d_5 + d_6 T^2 z + d_7 T^3. \quad \text{(B4)}$$

We find the best-fitting coefficients to be

$\boldsymbol{b} = [-0.087,\ 0.050,\ 0.998,\ -1.143, -0.383]$,

$\boldsymbol{c} = [0.068,\ -0.202,\ -0.806,\ 0.227,\ -19.86]$,

$\boldsymbol{d} = [2.04,\ -5.20,\ -0.636,\ 0.910,\ 4.181,\ 1.417,\ -0.536]$.

APPENDIX C: ARTIFICIAL NEURAL NETWORK PHOTO-*z*s

For comparison with the weighting method, we use an ANN method to estimate photometric redshifts (Collister & Lahav 2004; Oyaizu et al. 2008a). We use a particular type of ANN called a feed forward multilayer perceptron (FFMP), which consists of several nodes arranged in layers through which signals propagate sequentially. The first layer, called the input layer, receives the input photometric observables (magnitudes, colours etc.). The next layers, denoted hidden layers, propagate signals until the output layer, whose outputs are the desired quantities, in this case the photo-z estimate. Following the notation of Collister & Lahav (2004), we denote a network with k layers and N_i nodes in the ith layer as $N_1 : N_2 : \cdots : N_k$.

A given node can be specified by the layer it belongs to and the position it occupies in the layer. Consider a node in layer i and position α with $\alpha = 1, 2, \ldots, N_i$. This node, denoted $P_{i\alpha}$, receives a total input $I_{i\alpha}$ and fires an output $O_{i\alpha}$ given by

$$O_{i\alpha} = F(I_{i\alpha}), \quad \text{(C1)}$$

where $F(x)$ is the activation function. The photometric observables are the inputs $I_{1\alpha}$ to the first layer nodes, which produce outputs $O_{1\alpha}$. The outputs $O_{i\alpha}$ in layer i are propagated to nodes in the next layer $(i + 1)$, denoted $P_{(i+1)\beta}$, with $\beta = 1, 2, \ldots, N_{i+1}$. The total input $I_{(i+1)\beta}$ is a weighted sum of the outputs $O_{i\alpha}$:

$$I_{(i+1)\beta} = \sum_{\alpha=1}^{N_i} w_{i\alpha\beta} O_{i\alpha}, \quad \text{(C2)}$$

where $w_{i\alpha\beta}$ is the weight that connects nodes $P_{i\alpha}$ and $P_{(i+1)\beta}$. Iterating the process in layer $i + 1$, signals propagate from hidden layer to hidden layer until the output layer. There are various choices for the activation function $F(x)$ such as a sigmoid, a hyperbolic tangent, a step function, a linear function etc. The choice of the activation function typically has no important effect on the final photo-zs, and different activation functions can be used in different layers. In our implementation, we use a network configuration N_m:15:15:15:1, which receives N_m magnitudes and outputs a photo-z. We use hyperbolic tangent activation functions in the hidden layers and a linear activation function for the output layer.

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