Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **406,** 881–895 (2010) doi:10.1111/j.1365-2966.2010.16765.x

Photo-z performance for precision cosmology

R. Bordoloi,⋆ S. J. Lilly and A. Amara

Institute for Astronomy, ETH Zürich, Wolfgang-Pauli-Strasse 27, CH-8093, Zürich, Switzerland

Accepted 2010 March 30. Received 2010 January 7; in original form 2009 October 29

ABSTRACT

Current and future weak-lensing surveys will rely on photometrically estimated redshifts of very large numbers of galaxies. In this paper, we address several different aspects of the demanding photo-z performance that will be required for future experiments, such as the proposed ESA Euclid mission. It is first shown that the proposed all-sky near-infrared photometry from Euclid, in combination with anticipated ground-based photometry (e.g. PanStarrs-2 or DES) should yield the required precision in individual photo-z of $\sigma_z(z) \leq 0.05(1+z)$ at $I_{\rm AB} \leq 24.5$. Simple a priori rejection schemes based on the photometry alone can be tuned to recognize objects with wildly discrepant photo-z and to reduce the outlier fraction to $\leq$0.25 per cent with only modest loss of otherwise usable objects. Turning to the more challenging problem of determining the mean redshift $\langle z \rangle$ of a set of galaxies to a precision of $|\Delta_{\langle z \rangle}| \leq 0.002(1+z)$ we argue that, for many different reasons, this may be best accomplished by relying on the photo-z themselves rather than on the direct measurement of $\langle z \rangle$ from spectroscopic redshifts of a representative subset of the galaxies, as has usually been envisaged. We present in Appendix A an analysis of the substantial difficulties in the latter approach that arise from the presence of large-scale structure in spectroscopic survey fields. A simple adaptive scheme based on the statistical properties of the photo-z likelihood functions is shown to meet this stringent systematic requirement, although further tests on real data will be required to verify this. We also examine the effect of an imprecise correction for Galactic extinction on the photo-z and the precision with which the Galactic extinction can be determined from the photometric data itself, for galaxies with or without spectroscopic redshifts. We also explore the effects of contamination by fainter overlapping objects in photo-z determination. The overall conclusion of this paper is that the acquisition of photometrically estimated redshifts with the precision required for Euclid, or other similar experiments, will be challenging but possible.

Key words: methods: statistical – galaxies: distances and redshifts – cosmology: observations.

1 INTRODUCTION

Large-scale mapping of the weak-lensing shear field in three dimensions is emerging as a potentially very powerful cosmological probe (Albrecht et al. 2006; Peacock et al. 2006). Weak lensing has the advantage of directly tracing the mass distribution, thereby bypassing much of the complex astrophysics of the baryon component that underpin most of the other probes and which may well dominate the systematic uncertainties in them. In contrast, the underlying physics of weak lensing is extremely simple, and the challenges are primarily on the observational side, particularly the accurate measurement of the weak-lensing distortion and the estimation of distances to very large numbers of faint galaxies.

A weak-lensing survey of half the sky (20 000 deg^2) to a depth of $I_{\rm AB} \sim 24.5$ and with a point spread function (PSF) of $\sim$0.2 arcsec, forms a major part of the proposed ESA Euclid mission.[1] Euclid had its origins in two proposals submitted for the first round of the ESA Cosmic Visions 2015-2025 competition, the DUNE imaging survey (Réfrégier et al. 2006) and the SPACE spectroscopic survey (Cimatti et al. 2009). The combination of the two surveys, plus the anticipated improved information on the cosmic microwave background from *Planck*[2] offers dramatic improvements in our knowledge of the entire dark sector, including the definition of the dark matter power spectrum, the dark energy equation of state parameter w, as well as much else.

Application of weak lensing for cosmology requires at least a statistical knowledge of the distances, i.e. redshifts, of large numbers of individual galaxies. At $I_{\rm AB} < 24.5$, there are about 2.5 billion

⋆E-mail: rongmonb@phys.ethz.ch

[1] http://www.euclid-imaging.net;
http://sci.esa.int/science-e/www/area/index.cfm?fareaid=102
[2] www.rssd.esa.int/Planck

galaxies in the Euclid 2π sr survey area and so, realistically, reliance must be made on photometrically estimated redshifts (hereafter photo-z).

1.1 Required redshift precision for precision cosmology with weak lensing

Several papers have discussed the redshift precision that is needed for weak-lensing analyses to enable the full potential of this approach to be exploited (Ma, Hu & Huterer 2006; Amara & Réfrégier 2007; Abdalla et al. 2008; Ma & Bernstein 2008).

In the lensing tomography approach (Hu 1999), individual galaxies are binned into a number of redshift bins. The shear signal is extracted from the cross-correlation of the shape measurements of individual galaxies in different redshift bins. These correlated alignments then give information (with some distance-weighting function) on the mass distribution between the observer and the nearer of the two redshift bins. Redshift information for the galaxies is required at two conceptually distinct steps: first, the construction of the redshift bins used for the cross-correlation analysis to extract the weak-lensing signal and, secondly, the estimation of the mean redshift of the galaxies in a given bin, which is required to map the results on to cosmological distance and thereby extract the cosmological parameters. It is, of course, possible to do a similar correlation analysis with unbinned data (Castro, Heavens & Kitching 2005; Kitching, Taylor & Heavens 2008), but for our purposes this distinction is unimportant.

The cross-correlation between different redshift bins is undertaken to exclude any galaxy pairs that may be physically associated, i.e. have the same distance. This is to avoid the possibility that physical processes operating around individual galaxies may produce an intrinsic alignment of the galaxies that may be mistaken for the coherent alignment produced by the weak lensing of the foreground mass distribution. The required accuracy of the individual photo-z for the bin-construction task is set by the need to exclude overlaps in the $N(z)$ of individual bins (or the probability distribution for individual galaxies) and thereby remove physically close pairs (King & Schneider 2002). This typically sets a requirement on the precision of individual photo-z of about $\sigma_z = 0.05(1 + z)$ (Bridle & King 2007).

There is a second type of intrinsic alignment effect (Hirata & Seljak 2004), whereby the shape of the further of a given galaxy pair may be affected, through lensing effects, by the shape of the matter distribution around the nearer galaxy, which is likely to be correlated with the visible shape of that galaxy, thereby again producing a correlated alignment of the two galaxies that is unfortunately nothing to do with the lensing signal from the common foreground. Joachimi & Schneider (2008, 2009) have shown that it is possible to implement a nulling approach to eliminate this second intrinsic alignment signal, which again requires a priori knowledge of individual redshifts.

Once the weak-lensing signal is extracted, accurate knowledge of the redshift of the galaxies, as with any cosmological probe gives, amongst other parameters, information on the angular diameter distance $D_\theta(z)$. The sensitivity of $D_\theta(z)$ to the relevant cosmological parameters (Ω_m, Ω_Λ, w, etc.) gives the required accuracy in the mean redshifts that are required to achieve a given precision on the parameters. As an example, Peacock et al. (2006) have shown that a precision of 1 per cent in w requires a typical precision in the mean redshift of about 0.2 per cent in $\langle z \rangle$. The Euclid goal is a 2 per cent precision in w (independent of priors), requiring a precision of order $0.002(1 + z)$ in $\langle z \rangle$. This simple approach is confirmed by extensive analysis of the Fisher matrices (Ma et al. 2006; Amara & Réfrégier 2007). It is generally the case that if the mean redshift of a bin is defined accurately enough, then the higher moments of $N(z)$ within the bin will also have been sufficiently determined. Of course, systematic biases in $\langle z \rangle$ that vary smoothly with redshift are particularly troublesome as they will mimic the effect of changing the cosmological parameters.

In summary, in order to reach the Euclid performance, we require a statistical (random) rms precision of order $0.05(1 + z)$ per galaxy (for the correlation analysis), and a systematic precision in the mean z in each bin of order $0.002(1 + z)$. These are both quite demanding requirements, and together with the shape measurement itself ($\delta\gamma \sim 3 \times 10^{-4}$; Amara & Réfrégier 2008; Bridle et al. 2010), they represent one of the observational challenges that lie along the path to enabling precision cosmology with weak lensing.

Fortunately, there are some mitigating features of weak-lensing analysis. For instance, the analysis is robust (aside from root-n statistics) to the exclusion of individual galaxies, provided only that the exclusion is unrelated to their shapes. One is free therefore to reject galaxies that are likely to have poor photo-z provided that they can be recognized a priori, i.e. from the photometric data alone.

1.2 Challenges for the spectroscopic calibration of $N(z)$

Given the stringent requirements on the systematic error in the mean redshift $\langle z \rangle$ of a particular bin, one approach (Abdalla et al. 2008) is to define the $N(z)$ and mean $\langle z \rangle$ through the acquisition of spectroscopic redshifts for a representative subset of the galaxies. This direct sampling approach is certainly the most conservative, but will be very challenging, in practice, for the following reasons.

First, one clearly requires very large numbers of spectroscopic redshifts. If we have a total redshift interval of Δz, split into m bins, then the number of spectroscopic redshifts N will be of order

$$N \sim m^{-1}(\Delta z/\sigma_{\langle z \rangle})^2. \quad (1)$$

This assumes that the photo-z are perfect, and that there are no outliers. This leads trivially (Amara & Réfrégier 2007) to a requirement for 10^5–10^6 spectroscopic redshifts. This number may be somewhat reduced if galaxies are chosen with a given optimized redshift distribution (Ma & Bernstein 2008).

Secondly, these spectroscopic redshifts must be fully representative of the underlying sample. Any biases in the sampling of the bin or, even harder to reliably quantify, the almost inevitable biases in the ability to secure a reliable spectroscopic redshift, must be dealt with via a weighting scheme (Lima et al. 2008). It should be noted that current routine spectroscopic surveys of typical faint galaxies do not even approach 100 per cent completeness, even at brighter levels. One of the best to date is the zCOSMOS survey (Lilly et al. 2007) on relatively bright $I_\mathrm{AB} < 22.5$ galaxies which yields, at present, a 99 per cent secure redshift for 95 per cent of galaxies at its optimum $0.5 < z < 0.8$ (Lilly et al. 2009). Most other surveys are significantly less complete.

Even more invidious are the effects of large-scale structure in the spectroscopic survey fields, often called cosmic variance. Our own semi-empirical analysis (see Appendix A) of the COSMOS mock catalogues (Kitzbichler & White 2007) shows that, in a given patch of sky, the $N(z)$ at $I_\mathrm{AB} \sim 24$ becomes dominated by cosmic variance as soon as a rather small number of galaxies have been observed. The precise number depends on the field of view of the spectrograph, but is typically about 20–100 for spectroscopic fields of the order of 0.02–1 deg^2, i.e. a sampling rate of only a few per cent. This means that the spectroscopic survey must be split up

over a very large number of independent fields and that to get 10^6 redshifts that are Poisson variance dominated one must effectively cover the whole sky in a sparse sampled way. This is unlikely to be efficient with the large telescopes needed for such faint object spectroscopy. A similar concern comes from the effects of Galactic extinction and reddening, which are likely, even when corrected for, to require spectroscopic sampling across the full range of Galactic $[b, l]$.

These difficulties prompt consideration of other approaches, and in particular, that of placing greater reliance on the photo-z themselves, not only to construct the bins, but also to define their $\langle z \rangle$ with small systematic error.

1.3 Using photo-z for construction of $N(z)$

The performance of photometric redshifts is continually improving. For example, in the COSMOS field (Scoville et al. 2007), where we have very deep 30-band photometry from the ultraviolet (*GALEX*) to 5 μm, several photo-z schemes now achieve a precision of $\sigma_z \sim 0.01(1+z)$, with an outlier fraction [in non-masked areas and defined as a redshift difference greater than $0.15(1+z_{\rm spec})$ of <1 per cent Ilbert et al. 2009], at $I_{\rm AB} < 22.5$ and $0.05 < z < 1.4$, where the photo-z can be checked with about over 10 000 spectroscopic redshifts from zCOSMOS. It should be noted in passing that these 30 bands represent a very inhomogeneous data set in terms of PSF, etc., and so this impressive photo-z performance also demonstrates the feasibility of combining disparate data into homogeneous photometric catalogues. Of course, this outstanding performance in the COSMOS field is unlikely to be achieved over the whole sky for the foreseeable future because of the expense of the required multiband photometry. Nevertheless, the demonstration of this performance in COSMOS suggests that we have not yet reached any fundamental limit to photo-z performance.

There are a number of different approaches to photo-z estimation that can be broadly distinguished between template-matching and more purely empirical approaches, such as artificial neural networks (Collister & Lahav 2004). These have complementary strengths. Template fitting is based on the observed limited dimensionality of galaxy spectral energy distributions (SEDs) plus an astrophysical knowledge of the effects that can modify them, e.g. the redshift itself, and the effects of extinction in our own Galaxy or in the distant galaxy. The more empirical approaches in essence avoid any such assumptions, which is both a strength and a limitation. Although both approaches have passionate adherents, and both methods have their long list of pros and cons, our own view is that both approaches can normally perform equally well and that both are normally limited by the quality of the available data. In practice both use elements of the other, e.g. in template fitting, the actual data can be used to adjust the templates and the photometric zero-points, somewhat blurring the distinction. Finally, it should be noted that both can produce a likelihood distribution in redshift space and subsequently a probability density distribution through the application of priors (Benítez 2000; Bolzonella, Miralles & Pelló 2000; Collister & Lahav 2004; Brammer, van Dokkum & Coppi 2008). In this paper we will base our analysis on a template-fitting algorithm (ZEBRA; Feldmann et al. 2008) since we believe its strengths are well suited to the problem in hand. We also note that the impressive real-life performance in COSMOS described above was achieved with two independent template-fitting codes [LE PHARE (Ilbert et al. 2009) and ZEBRA (Feldmann et al. 2006)].

This improving photo-z performance described above suggests that it may be possible to use the photo-z themselves to construct the $N(z)$, and thus $\langle z \rangle$ for each bin, providing that the systematic uncertainties can be kept below the required level of $0.002(1+z)$. Some spectroscopic calibration would of course still be required, but the focus of this would be on constructing and characterizing the photo-z algorithm, rather than on constructing the $N(z)$ directly.

This approach would have a number of potential advantages over that discussed by Abdalla et al. (2008) and others, and summarized above. At the very least, the number of spectra needed may be substantially reduced, although it is unlikely that one would wish to rely entirely on the photo-z and eliminate the spectroscopic conformation completely. However, to establish that the uncertainties in the photo-z are at the required level ($\sigma_{\langle z \rangle}$) we would need of order

$$N \sim (\sigma_z/\sigma_{\langle z \rangle})^2 \quad (2)$$

spectroscopic redshifts, which may be orders of magnitude or more smaller than that implied by equation (1) since $\sigma_z \sim 0.05\Delta z$.

More importantly, the requirements on completeness and sampling are substantially relaxed since the photo-z characterization is done on individual objects. As one example, it is relatively easy to simulate the degradation in photo-z performance with noisier photometric data, so the calibrating spectroscopic objects need not necessarily extend all the way down to the photometric limit. The cosmic variance problem in spectroscopic calibration is eliminated, and the uncertainties arising from Galactic reddening can also be substantially reduced.

1.4 Subject of this paper

The aim of this paper is to explore the performance of a template-fitting photo-z code as applied to simulated photometric data of the approximate quality that we may realistically expect for a 2π sr *all-sky* ground and space survey within the next decade. Our emphasis is on both the per object performance and on the potential for recognizing and correcting systematic biases, which must be done to a high level if the increased reliance on photo-z is to be possible.

As described in more detail in Section 3, we will assume for definiteness a photometric data set that includes the three-band near-infrared (NIR) photometry that is planned for Euclid itself, plus 5-band *grizy* photometry similar to that which should be produced by the PanStarrs-1, PanStarrs-2 and PanStarrs-4 projects (hereafter PS-1, PS-2 and PS-4).[3] Examining these three generations of ground-based survey probes a range of depths that can be compared against other future surveys such as DES[4] and LSST.[5]

We then explore the following four topics that potentially may limit the photo-z performance and their usefulness to construct $N(z)$ and $\langle z \rangle$.

(i) The photo-z performance on individual objects in terms of the rms scatter (and bias) between the true redshift and the maximum likelihood photo-z, with particular emphasis on how to a priori identify and reject the outliers (*catastrophic failures*) from their individual photo-z $L(z)$ likelihood distributions.

(ii) The construction of $N(z)$ for a given set of photo-z selected galaxies, using their photo-z alone, with an emphasis on how to derive $N(z)$ from summing the probability density distributions that are derived from their individual likelihood functions $L(z)$ to yield the least biased estimate of $N(z)$ and $\langle z \rangle$ for the ensemble.

[3] http://pan-starrs.ifa.hawaii.edu
[4] http://www.darkenergysurvey.org
[5] http://www.lsst.org

(iii) The systematic biases that can enter into the photo-z from an incorrect assumption about the level of foreground Galactic reddening, and how well the photometric data themselves can be used to determine the foreground reddening, both for a set of galaxies at known redshifts, and for those without known redshifts.

(iv) The effects of the photometric superposition of two galaxies at different redshifts, leading to a mixed SED that may perturb the photo-z, with an emphasis on seeing what happens to the redshift likelihood distribution. An interesting question is whether such composite objects can be recognized photometrically as well as 'morphologically' from the images.

Our approach is to try to isolate these problems and to explore each in turn with the aim of providing an *existence proof* that provides a plausible route to achieve the very high photo-z performance that is required for Euclid. In particular, we decided to construct the input photometric catalogues using exactly the same set of approximately 10 000 templates as we subsequently used in the ZEBRA photo-z code. This may strike some readers as being somewhat circular. However, this approach allows us to eliminate the choice of templates as a variable, or uncertainty, in our analysis. This is motivated by the exceptional performance (discussed above) that has already been achieved with the same templates coupled with the exquisite observation data in COSMOS. This strongly suggests to us that the choice of templates is unlikely to be the limiting factor with the degraded photometry that we can realistically expect to have over the whole sky within the time-scale of a decade or so. If however there are some spectra which are unrepresented by templates then the redshift estimation for those galaxies will fail systematically. Moreover, templates are constructed with the assumption that galaxy population is in broad terms the same across the sky.

Although focused on the Euclid cosmology mission, the ideas and results from this paper may be of interest in many other applications that involve photo-z.

2 GENERATION OF THE PHOTOMETRIC CATALOGUES

In order to simulate the catalogues for this paper, we use the COSMOS mock catalogues produced by Kitzbichler & White (2007). The mocks are generated from semi-analytic galaxy formation models using galaxy merger trees derived from the Millennium N-body Simulation. The corresponding cosmological parameters are $\Omega_{\rm m} = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_{\rm b} = 0.045$, $h = 0.73$, $n = 1$ and $w_\Lambda = -1$. The mocks have an area of $1.^\circ4 \times 1.^\circ4$ and are magnitude limited at $I_{\rm AB} \leq 26$. For each galaxy the catalogues give the right ascension (RA), declination (Dec.), the redshift z and the magnitudes for the filters $B_{\rm j}$, $g+$, $r+$, $i+$ and $K_{\rm s}$. The RA and Dec. are in the range $[-0.^\circ7, 0.^\circ7]$. There are a total of 24 mocks, each produced from a different wrapped cone that passes through the cubical simulation so that no object appears twice in a given cone. A given galaxy will however appear at different redshifts in the different mocks. Each mock catalogue contains approximately 600 000 objects, the majority of which are at $z \leq 1$.

In order to produce a photometric catalogue with our own set of filters, we first identified an SED template, from the 10 000 available templates, that well matched the given $B_{\rm j}$, $g+$, $r+$, $i+$ and $K_{\rm s}$ photometry for each galaxy in the Kitzbichler & White (2007) catalogue, at its known redshift. This operation made use of a program kindly provided by Thomas Bschorr. We use templates representing synthetic stellar populations generated from the Bruzual & Charlot (2003) evolutionary models and the parameters varied to compute the whole set of templates are age, star formation history (continuous underlying SFR superimposed with single or multiple bursts), dust obscuration in the galaxy and stellar metallicity. These templates include a range of internal reddening, which ranges from $0 < A_V < 2$ mag (see Fig. 1). This chosen template was then used to compute ideal photometry (i.e. without any observational noise) for this galaxy in any other passband of interest. Intergalactic absorption in each template is compensated for using the Madau law (Madau 1995). In order to match to the proposed Euclid weak-lensing experiment, we consider only objects with $I_{\rm AB} \leq 24.5$. Except for the cosmic variance analysis in Appendix A, we combine all 24 mocks and use a random subsample to mimic a survey over a large area in the sky. Each of our simulations contains at least 100 000 objects.

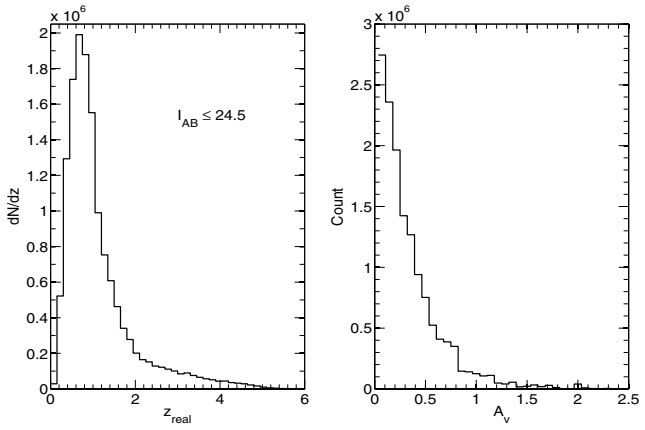

Figure 1. The left-hand panel shows the $N(z)$ distribution of the artificial catalogue with $I_{\rm AB} \leq 24.5$ used in this paper. The right-hand panel shows the distribution of internal A_V in the chosen SED templates for these galaxies.

To generate a photometric catalogue as observed, this ideal photometric catalogue is degraded by adding Gaussian noise (in flux) to the photometry according to the three different sets of survey sensitivities that are listed in Table 1. All three sets contain the nominal NIR photometry expected from Euclid but have different choices for the depth in *grizy*, and therefore explore the requirements for the ground-based complement of the IR survey. Survey-A, which we generally find to be inadequate, uses the point-source sensitivities from Cai et al. (2009). This is therefore a possibly optimistic representation of a PS-1-like survey. Survey-C, which we find exceeds

Table 1. The filter sensitivities for different survey configurations considered. The values quoted here are 5σ errors in AB magnitude.

Assumed sensitivities for the representative surveys			
Band	Survey-A	Survey-B	Survey-C
g	24.66	25.53	26.10
r	24.11	24.96	25.80
i	24.00	24.80	25.60
z	22.98	23.54	24.10
y	21.52	22.01	22.50
	Euclid	NIR	
Y	24.00	24.00	24.00
J	24.00	24.00	24.00
H/K	24.00	24.00	24.00

the requirements, uses the PS-4 extended source sensitivities which is calculated by taking nominal PS-4 point-source sensitivities from Abdalla et al. (2008), degraded by 0.6 mag to account for extended sources. The proposed LSST large area survey would be expected to be slightly deeper than this. Finally, survey-B is intermediate between these two, and approximates what could be expected from PS-2 or the DES. The precise choice of these sensitivities is of course somewhat arbitrary, and they are adopted here for the sake of illustration. Unless stated otherwise, all magnitudes stated in this paper are in the AB system.

2.1 Estimating photo-z values using ZEBRA

In this paper the template-fitting photo-z code ZEBRA (Feldmann et al. 2006) is used to produce photo-z for the galaxies in each of the observationally degraded catalogues. ZEBRA gives a single best-fitting redshift, which we call the 'maximum likelihood redshift' and template type, together with their confidence limits estimated from constant χ^2 boundaries. ZEBRA also outputs the normalized likelihood functions $L(z)$ for individual galaxies in various formats, which we also use in this paper. $L(z)$ can be modified by a Bayesian prior, as desired, but is in any case normalized so that the integral over all redshifts is unity. Further information is available in the ZEBRA user manual.[6]

3 PERFORMANCE ON INDIVIDUAL OBJECTS

In this section, we compare the basic performance of the photo-z estimation by comparing the maximum likelihood photo-z with the known redshifts of the galaxies, for different choice of survey depths for the different simulations presented in Section 2. The aim of this section is to assess what sort of ground-based data are required to complement the Euclid IR photometry and to develop techniques for the automated recognition and elimination of outliers with wildly discrepant photo-z.

3.1 Depth of ground-based photometry

Using the photometric catalogues that were described in the previous section, and as degraded to simulate different survey configurations, we compare their photo-z performance. We first bin the galaxies into narrow redshift bins on the basis of their 'observed' photo-z. We then use the bias (b) and the dispersion $\sigma_z(z)$ to parametrize the performance, defining these as follows. The error per object (δz_i) is

$$\delta z_i = (z_{\mathrm{real},i} - z_{\mathrm{phot},i}), \tag{3}$$

where $z_{\mathrm{real},i}$ and $z_{\mathrm{phot},i}$ are the real and photometric redshifts of the ith galaxy. The mean bias in each photo-z bin $\Delta_z(z)$ is then

$$\Delta_z(z) = \langle \delta z \rangle. \tag{4}$$

The rms deviation in the photo-z estimation within the bin $\sigma_z(z)$ is

$$\sigma_z^2(z) = \langle (\delta z_i - \Delta z)^2 \rangle \tag{5}$$

and the total mean-squared error (MSE) is given as

$$\mathrm{MSE}(z) = \sigma_z^2(z) + \Delta_z^2(z). \tag{6}$$

In Fig. 2 we show the $\sigma_z(z)$ and $\Delta_z(z)$ for the different survey configurations. The blue curves in all the panels give the initial performance of the photo-z code, without any attempt at removing outliers. As expected, increased depth in the optical ground photometry increases the reliability of the photo-z estimates. However, none of the configurations matches, without cleaning, the requirement of $\sigma_z(z)/(1+z) \leq 0.05$, especially at the lower redshifts $z \sim 0.5$ where many of the galaxies in fact lie. The green curves show the effect of removing outliers, recognized purely photometrically (see Section 3.2), and the red curve shows the effect of modifying the individual $L(z)$ as described in Section 3.3.

3.2 A priori identification of outliers

In template fitting, a likelihood function $L(z)$ is derived for each galaxy from which maximum likelihood photo-z is estimated. In an ideal case, with a well-defined photo-z estimate, the $L(z)$ has a single tight peak. Empirically, it is found that many galaxies at the survey limit with poor photo-z estimates have a bimodal likelihood distribution (Brodwin et al. 2006). We therefore developed an algorithm that searches for bimodality in the likelihood curves of each galaxy. If a likelihood function contains more than one peak separated by a certain pre-defined redshift difference and if the ratio between primary and secondary peaks is above a threshold value, then the galaxy is flagged as a likely outlier and can be rejected from the lensing analysis. This pre-defined threshold value can be tuned from simulations of the kind described here, or from spectroscopic measurements of actual redshifts. Of course, this procedure will undoubtedly remove some objects whose photo-z values are actually quite good, but the lensing analysis is stable to this kind of exclusion. There are a number of other ways that the quality of the photometric redshift estimate can be quantified using the likelihoods. For instance Benítez et al. (2009) have adopted the ODDS parameters, which uses the area of the likelihood within a given redshift range around the peak redshift. The approach that we have adopted focuses specifically on trying to identify galaxies with bimodal peaks in their likelihood. We have not performed a detailed comparison of the various methods for detecting outlier since our objective here was to identify at least one method that works. We leave a more thorough study in finding the optimal method to later work.

After removal of doubtful photo-z, the errors in $\sigma_z(z)$ and mean bias $\Delta_z(z)$ are dramatically reduced, as shown by the green lines in Fig. 2. The major improvement in $\sigma_z(z)$ and $\Delta_z(z)$ come from rejection of catastrophic failures rather than a tightening of the 'good' photo-z. As the depth of the photometry increases, it is found that fewer objects need to be rejected to improve the photo-z estimates. In case of survey-A, we find that 23 per cent must be rejected to get below $\sigma_z(z) \leq 0.05(1+z)$, for survey-B it is 12 per cent and for survey-C, only 9 per cent. The trade-off between beneficial cleaning and the wasteful loss of objects determines the robustness of the cleaning. After the above cleaning has been performed, the fraction of 5σ outliers (catastrophic failures) is reduced below 0.25 per cent in all the three surveys (see Table 2). It should be noted that we have not taken into account any additional priors such as the size or luminosity of the galaxies, which might further improve the performance.

3.3 Construction of the probability functions by modifying $L(z)$

ZEBRA, like most photo-z programs, produces a likelihood function $L(z)$ that is a reflection of the probability that the photometric

[6] http://www.exp-astro.phys.ethz.ch/ZEBRA/

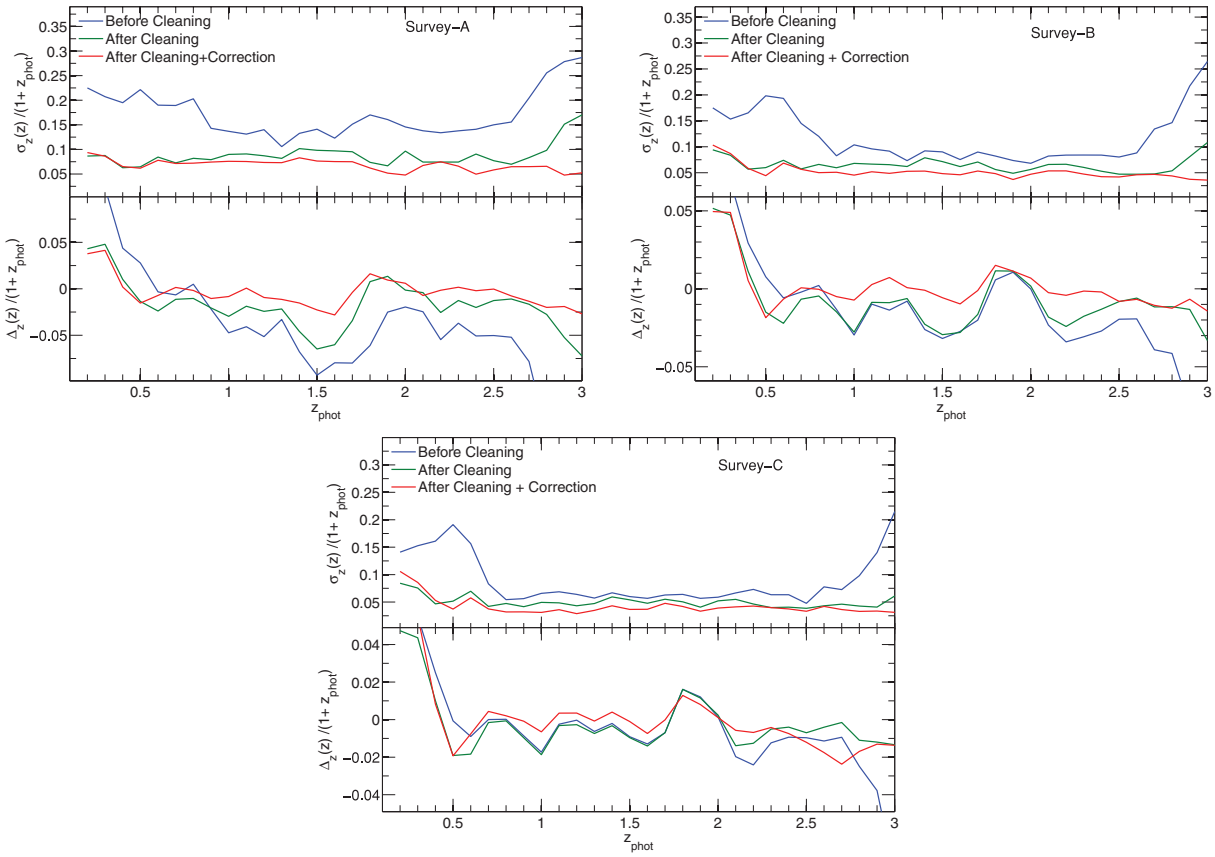

Figure 2. The overall performance of survey-A, survey-B and survey-C whose depths are as quoted in Table 1. The blue lines give the performance without cleaning and the green lines give the performance after cleaning and the red line gives performance after cleaning and applying correction. It is seen that with cleaning and correction $\sigma_z(z)/(1+z) \leq 0.05$ is almost reached in all the cases and systematic bias is also reduced considerably. Survey-B after cleaning and correction reaches $\sigma_z(z)/(1+z) \leq 0.05$ easily. For survey-A 23 per cent, for survey-B 13 per cent and for survey-C 9 per cent rejections were made after cleaning.

Table 2. The percentage of 5σ outliers in various surveys studied. f_{cat} reduces significantly once cleaning of the catalogue is performed, which identifies most of the outliers effectively.

Survey	Percentage of 5σ outliers (f_{cat}) in Before cleaning	After cleaning
Survey-A	1.18	0.2300
Survey-B	0.8820	0.2138
Survey-C	0.8221	0.1776

data would be observed if the galaxy in reality lies at some redshift z. What is of course desired is, given the photometric data, the probability density distribution of the redshift of the galaxy. In the Bayesian approach to statistics, the desired probability distribution for a particular galaxy is obtained by multiplying the likelihood function $L(z)$ by some 'prior' which reflects some assumption(s) about the relative probabilities of different redshifts in the absence of the photometric data in question. Sometimes a 'flat prior' is applied, i.e. assuming that the probability density is simply given by $L(z)\mathrm{d}z$, but this is as much an assumption as more complicated priors. As noted above, the Bayesian prior is a function in z that is applied uniformly to a set of galaxies, regardless of the particular $L(z)$ of each galaxy. As an example, if spectroscopic redshifts are available for a representative subset of galaxies, then the prior could be derived from the $N(z)$ distribution of those galaxies. Other approaches are also possible, e.g. Brodwin et al. (2006), Benítez (2000), etc.

In this paper we develop an alternative empirical approach to constructing the individual probability density distributions from the individual likelihood function $L(z)$ that utilizes information on the actual redshifts, but on an object-by-object basis. The motivation for this is that this approach should be less sensitive to bias or non-uniformities in the acquisition of spectroscopic redshifts.

If the redshift probability function for a given galaxy is meaningful, then the location of the actual redshift within that probability function should be random. Therefore, we first define a variable $P(z)$ for each galaxy that is obtained by integrating the $L(z)$,

$$P(z) = \int_0^z L(z')\mathrm{d}z'. \tag{7}$$

For galaxies where we reliably know the real redshift z_{real} (e.g. from a spectrum), we can then compute $P_{real} = P(z_{real})$. P_{real} is simply a measure of how much of the probability density for a given galaxy lies below the actual (measured) redshift. If $L(z)$ is already a good representation of the probability function, then the distribution of P_{real} for all the galaxies for which one has spectra, i.e. $N(P_{real})$, should be uniform between the extreme P_{real} values of 0 and 1, i.e.

$$N(P_{real})\mathrm{d}P_{real} = \mathrm{d}P_{real}. \tag{8}$$

If this is found not to be the case, as is likely, then a modification or correction to the probability density distributions is required. We

approach this by constructing a global mapping between P and P' that is determined from all objects with reliably known redshifts, such that the distribution of $P'_{\rm real}$ will be flat between 0 and 1. We can write

$$\frac{{\rm d}P}{{\rm d}z} = \frac{{\rm d}P}{{\rm d}P'}\frac{{\rm d}P'}{{\rm d}z}. \quad (9)$$

Note that

$$\frac{{\rm d}P'}{{\rm d}P} \sim N(P_{\rm real}) \quad (10)$$

and $N(P') = 1$ when $P = P'$.

We then produce a new probability distribution $\mathscr{L}'(z)$ from the original likelihood function $L(z)$ for each galaxy, with or without a known redshift, to produce $\mathscr{L}'(z)$ such that, for all z,

$$P'(z) = \int_0^z \mathscr{L}'(z'){\rm d}z'. \quad (11)$$

It is easy to see that this is given by the following simple multiplication to $L(z)$:

$$\mathscr{L}'(z) = L(z)\,N(P(z)), \quad (12)$$

where $N(P(z))$ is the 'observed' distribution of $N(P_{\rm real})$ with applied mapping of z to P for each object using equation (11).

This procedure is clearly related to the application of a conventional Bayesian prior in redshift space to produce the probability density distribution, but is now applied in P (probability) space and is based on the absolute requirement of having a flat $N(P_{\rm real})$ for meaningful probability distributions. The application of the global mapping of P to P' to all objects, independent of their nature, is of course arbitrary and cannot be rigorously justified. However, we find this approach works well, both here and later in the paper.

This procedure is illustrated in Fig. 3, in which the red line gives the $N(P_{\rm real})$ for all the galaxies with survey-C-like sensitivities. We therefore compute an empirical correction so as to make the $N(P_{\rm real})$ curve to be flat. The green line in Fig. 3 is produced by using a subsample of 1000 galaxies and the dashed blue line is produced by inferring the $N(P(z))$ function from a blind mock catalogue. Note that due to discrete binning in P space there is noise introduced in the $N(P(z))$ function and hence the corrected green line is noisy. This noise does not translate into noise in z space.

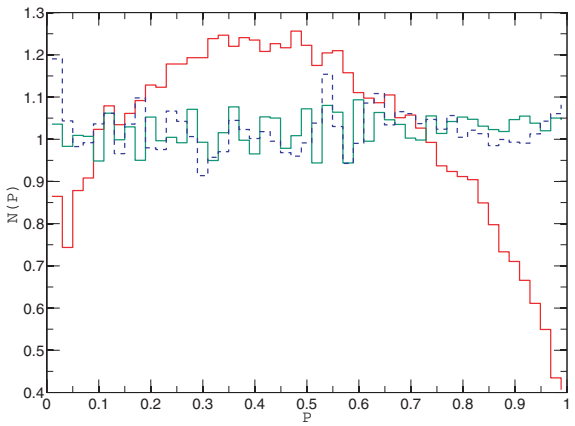

Figure 3. Normalized $N(P)$ before (red) and after (blue dash) applying the correction scheme using a blind sample to estimate the correction function and after (green) correction using a subsample of the same set of objects. In the ideal case this distribution should be flat in P.

From the peaks of the new probability density distribution $\mathscr{L}'(z)$, we can also define a more accurate individual photo-z estimate as given by the red line in Fig. 2. We stress that it is not necessary to know the $N(P)$ very accurately and an average correction of $N(P)$ using a relatively modest number of spectra yields significant improvement to photo-z accuracy.

4 CHARACTERIZATION OF $N(z)$ FROM THE LIKELIHOOD FUNCTIONS

In weak-lensing tomography the photo-z values are used to construct redshift bins which are then used to calculate the lensing power spectrum. The actual $N(z)$ of each bin must then be known for quantitative interpretation of the lensing signal. The mean of the distribution is the most important parameter (Amara & Réfrégier 2007) and we therefore focus on this. Generally a single redshift estimator from the photo-z code (i.e. the maximum likelihood photo-z) is used to construct these bins. However, if using these single redshifts, the $\Delta_{\langle z\rangle}$ requirement cannot be reached, as clearly shown in Figs 2 and 4. This is because the maximum likelihood redshifts cannot, by construction, trace the wings of the $N(z)$ that lie outside of the nominal bins, or trace the remaining catastrophic failures associated with some of the photo-z. Therefore a more sophisticated approach is required.

As noted in Section 1, one approach is to undertake a major spectroscopic survey of large numbers of representative objects in the bin and define the actual $N(z)$ empirically in this way. As discussed there, there are a number of practical difficulties of doing this.

In this paper we explore a different approach, which is to characterize $N(z)$ as the sum of the probability density distributions for each redshift bin. We define the mean redshift inferred from summing the probability density functions as

$$\bar{z} = \left\langle \sum L(z) \right\rangle = \int_0^\infty z \sum L(z){\rm d}z \quad (13)$$

and the bias in estimating $z_{\rm real}$ as

$$\Delta_{\langle z\rangle} = z_{\rm real} - \bar{z}. \quad (14)$$

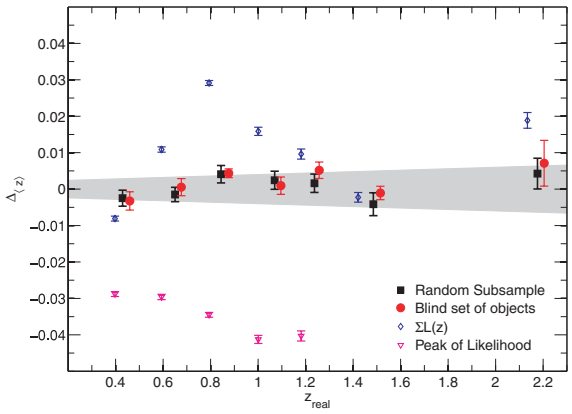

Figure 4. The bias in the mean of the tomographic bins estimates for survey-B. The filled squares are obtained if a random subsample is used to correct the likelihood functions and the filled circles are obtained if correction is done from a blind set of objects. The small random differences presumably reflect the effects of large-scale structure in the 'calibrating' mock. The open blue circles are obtained if only the sum of the likelihood functions is used (before correction) and the open triangles are obtained whilst using maximum likelihood redshifts to estimate the mean.

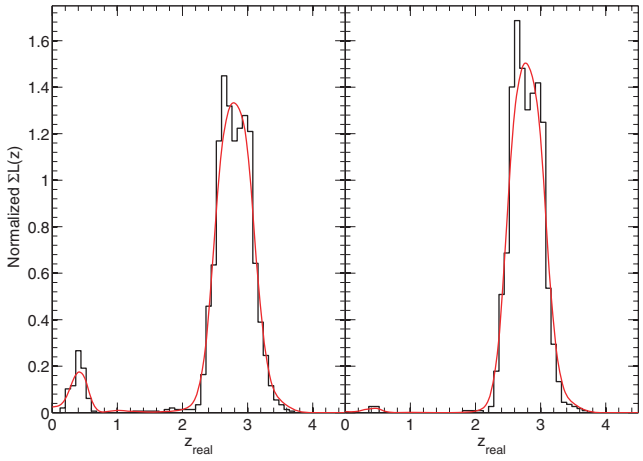

Figure 5. $N(z)$ constructed from the $\sum L(z)$ function before and after cleaning. Here the normalized histogram gives the real redshift distribution in the bin and line is the $N(z)$ constructed from the $\sum \mathscr{L}'(z)$ function. The left-hand panel gives the redshift bin before cleaning and right-hand panel gives the redshift bin after cleaning. The constructed $N(z)$ clearly traces the catastrophic failures.

We apply this approach using the same modification techniques described in Section 3.3. The straight sum of the original likelihood functions $L(z)$ is able to characterize the redshift distribution well, as seen in Fig. 5, which shows that for survey-C the summed $L(z)$ follows (visually) both the catastrophic failures and the wings of the redshift bins well. If we apply the cleaning algorithm described above, the catastrophic failures are removed and wings are constrained more tightly. However, this approach alone is not in fact good enough to characterize the $N(z)$ of the bins to the required precision of $|\Delta_{\langle z\rangle}| \leq 0.002(1+z)$.

To characterize the bins more accurately, the $\mathscr{L}'(z)$ scheme as described in Section 3.3 was developed. We compute $N(P)$ for each redshift bin separately, using a spectroscopically observed subsample of 800–1000 galaxies per bin. After correction, the new likelihood function $\mathscr{L}'(z)$ for each galaxy, and therefore occasionally a new maximum likelihood redshift, is obtained. These are used to rebin the galaxies and the sum of the new $\mathscr{L}'(z)$ is used to construct $N(z)$ for the bins. In Fig. 6 the bias on the mean of the $N(z)$ is given

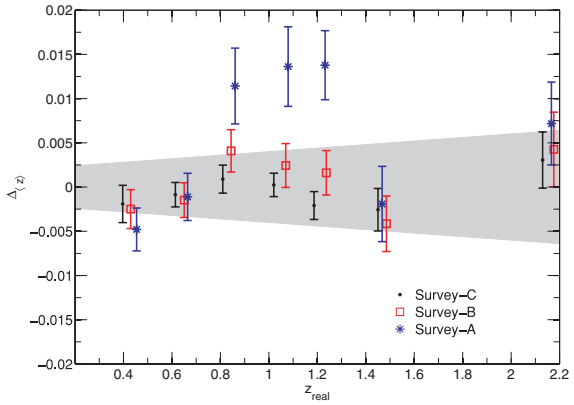

Figure 6. The bias in the mean of the tomographic bins estimates from the normalized $\sum \mathscr{L}'(z)$ functions for survey-C, survey-A and survey-B. For survey-C, with cleaning for catastrophic failures and after applying correction gives $|\Delta_{\langle z\rangle}/(1+z)| \leq 0.002$. Here the shaded region is $|\Delta_{\langle z\rangle}| = 0.002(1+z)$. We have introduced a small offset in x-axis values of survey-B and survey-C for legibility.

for different redshift bins and survey parameters. The error bars on each point shows the effect of randomly picking different subsets for the spectroscopic calibration repeatedly. In Fig. 6 the shaded region gives the Euclid requirement of $|\Delta_{\langle z\rangle}/(1+z)| \leq 0.002$ on the mean redshift of the redshift bins. The black dots are for survey-C, which easily reaches the Euclid requirements. The red open boxes are for survey-B and it just meets the Euclid requirement. The blue stars are for survey-A which do not meet the specifications as given by the shaded region. From this analysis we conclude that for a Euclid-like survey, using a survey-B-like ground-based complement we can characterize the $N(z)$ of the tomographic bins to a precision of $|\Delta_{\langle z\rangle}/(1+z)| \leq 0.002$ and we need around 800–1000 random spectroscopic subsamples per redshift bin to characterize them.

The improvement in accuracy in this method is illustrated in Fig. 4. Here $\Delta_{\langle z\rangle}$ for the tomographic bins are plotted for a survey-B-like scenario. The filled black squares are obtained from a random subsample correction as described above. If we take a small part of the sky to calibrate the whole sample, then the filled red circles are obtained. The small random differences presumably reflect the effect of large-scale structure in the calibration mock. The open blue circles are obtained when only the sum of $L(z)$ is used without any correction and the open triangles are obtained when only the maximum likelihood redshifts are used to estimate the mean of the tomographic bins. Clearly the correction scheme improves the mean estimation over other methods and correction can be applied either from a random subsample of the whole ensemble or from only one patch of the sky with more or less same results.

The great advantage of this approach is that it sidesteps completely the problems associated with the presence of large-scale structure in the spectroscopic survey fields since the spectro-z are used to characterize, and globally modify, the photo-z estimates of individual galaxies, and not to characterize directly the $N(z)$, which will clearly be affected by such structure. It is also potentially less susceptible to incompleteness in the measurement of spectroscopic redshifts, either in selection for spectroscopic observation or in success in measuring a redshift. That said, it is based on an assumption that, for a given spectroscopic target at some maximum likelihood photo-z, the ability to measure a redshift will not systematically depend on the location of the real redshift in P space.

5 INTERNAL CALIBRATION OF GALACTIC FOREGROUND EXTINCTION

In this section we explore the effect that errors or uncertainties in foreground Galactic extinction can have on photo-z, and examine whether the photometry of large numbers of galaxies, with or without known spectroscopic redshifts, can be used to determine an improved extinction map and locally correct the extinction. This latter aspect is an extension of the iterative adjustment of photometric zero-points that is now standard in many template-fitting photo-z codes.

Extragalactic photometry is routinely corrected for the effects of foreground Galactic extinction using reddening maps and an assumed extinction curve. In practical terms, the effect that we should therefore worry about is an error or uncertainty in the A_V, i.e. a ΔA_V, which may be positive or negative. This will cause galaxies to be either too red, or too blue, in the photometric input catalogue.

In this section we construct a catalogue containing 10^4 objects down to $I_{\rm AB} \sim 24.5$. This mimics a roughly 0.1-deg^2 region of the Euclid survey. We consider photometry with the accuracy expected from both survey-A and survey-C, and then perturb these catalogues

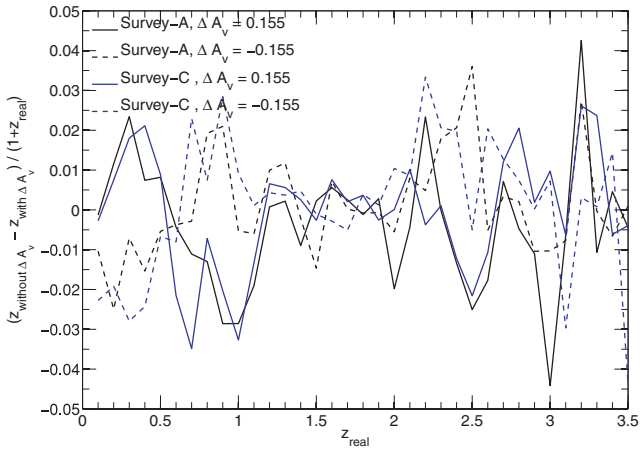

Figure 7. The bias introduced in photo-z estimation due to small offset in photometry after an average correction for effects of reddening. For survey-A and survey-C simulations two opposite ΔA_V offsets were investigated: 0.115 and −0.115. Changing the sign of ΔA_V leads to a more or less mirror inversion in the bias.

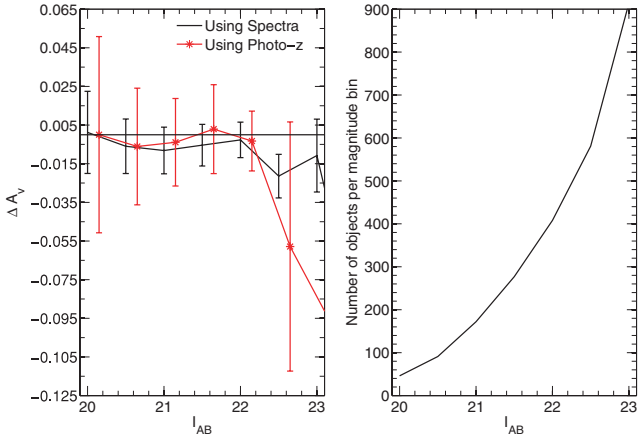

Figure 8. The left-hand panel gives estimates of ΔA_V for different magnitude bins (in the configuration of survey-C). The red line is obtained using internally computed photo-z, without knowledge of the redshifts of the galaxies, and the black line is using spectroscopic information. Here we have introduced a small offset in x-axis on the red line for legibility. The right-hand panel is the number of objects per magnitude bin. With spectra, at $I_{\rm AB} \sim 22$ mag bin around 350 spectra are sufficient to estimate ΔA_V accurately.

by applying a standard mean reddening law (Cardelli, Clayton & Mathis 1989) with a relatively large $\Delta A_V \sim 0.155$, in both the positive and negative directions. We then compare the photo-z for the galaxies with and without these ΔA_V offsets. Fig. 7 shows the bias between these photo-z estimates as functions of redshift. The bias fluctuates in redshift in a somewhat random way, with large systematic excursions at low redshifts. Interestingly, but perhaps not surprisingly, changing the sign of the ΔA_V leads to a largely mirror effect on the redshifts, suggesting that the response of the photo-z scheme to errors in A_V is linear.

Although we chose quite a large 'worst case' error in A_V, the biases with redshift seen in Fig. 7 are almost 10 times worse than the $0.002(1+z)$ photo-z bias that can be tolerated by Euclid's precision cosmology. Large-scale redshift-dependent biases in the photo-z are particularly worrisome as they mimic the effect of cosmological parameters. We therefore explore the possibility of iteratively identifying the residual ΔA_V error as follows. We assume that we will know the wavelength dependence of the reddening in a given field and that the problem is therefore in determining the A_V. At high galactic latitude, we expect that the wavelength dependence could be determined from very large areas of sky, but that $A_V(b, l)$ may vary on small scales.

To estimate the ability of the photometry to determine ΔA_V, we take the input photometric catalogue and 'correct' it for a wide range of assumed ΔA_V around zero. We then run ZEBRA on each of these corrected catalogues and take the $\sum \chi^2_{\rm min}$ of all the individual galaxies (i.e. the sum of the 'best-fitting' chi-squares). The value of ΔA_V that produces the minimum $\sum \chi^2_{\rm min}$ is taken as the best estimate of ΔA_V in that region. The exercise can be undertaken with the redshifts of the galaxies as a free parameter, or by assuming that the galaxies have known redshifts, and looking at the $\sum \chi^2_{\rm min}$ amongst the templates at the known redshifts for each galaxy. The sample used here is magnitude limited to $I_{\rm AB} \leq 24.5$.

To obtain an error bar on ΔA_V we compute a reduced chi squared ($\chi^2_{\rm r}$). For this we need to know the total degrees of freedom (d.o.f.) available in the template-fitting approach, which is non-trivial since it is unclear how many d.o.f. are associated with the 10 000 templates. We assess this by requiring that the $\chi^2_{\rm r}$ be unity (using survey-C, although this should not be important) and find that this gives a dimensionality close to 3, which sounds reasonable given what is known about galaxy spectra (see Connolly et al. 1995). A rule of thumb for estimating the uncertainty in 'one parameter of interest' gives (Avni 1976)

$$\chi^2_{1\sigma\ \text{confidence level}} = \chi^2_{\rm min} + 1. \quad (15)$$

Hence the 1σ uncertainty in ΔA_V estimate is given by the values of ΔA_V which are below $\chi^2_{\rm r} + 1/\text{d.o.f.}$ values.

To estimate the effect of photometric noise in estimating ΔA_V in this internal way, we consider objects in magnitude bins in $I_{\rm AB}$. The results are shown in Fig. 8. We find that it is worth using only galaxies with relatively high signal-to-noise ratio (S/N) photometry, i.e. with the adopted survey parameters, down to $I_{\rm AB} \sim 22$. Below this level, the estimate degrades appreciably. The addition of spectroscopic redshift reduces the error bar significantly, but the method is still practicable down to the same magnitude limit and yields an error on ΔA_V of the order of 0.01 (with known redshifts) or 0.02 (without). To close the loop, we show in Fig. 9 the bias in photo-z introduced by applying 0.01 error to ΔA_V which shows improvement in photo-z errors by a factor of 10.

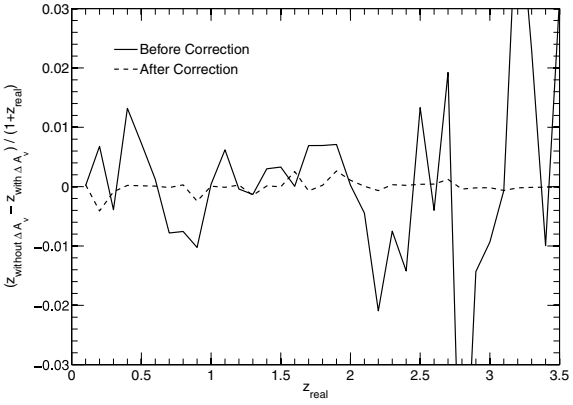

Figure 9. Performance of the photo-z estimates before and after estimated ΔA_V correction have been applied. Using this correction scheme the systematic bias introduced due to ΔA_V is significantly reduced below $0.002(1+z)$ level.

6 IMPACT ON PHOTO-z FROM BLENDED OBJECTS AT DIFFERENT REDSHIFTS

Sometimes multiple galaxies will overlap on the sky, and the photometry will be a composite of the two SEDs. Even with spectroscopy, such objects have composite spectra rendering their spectroscopic redshift estimation non-trivial. In this final section we explore the effects of this blending on photo-z estimates. We simulate many such blended objects by constructing composite SEDs constructed from galaxies at different redshifts, different colours and a wide range of relative brightnesses, from dominance of one through to dominance of the other. For definiteness we look at the photo-z behaviour for a survey-C-like survey.

We select several objects from the main COSMOS mock catalogues at different redshifts and having different colours and normalize their fluxes to have the same $I_{\rm AB}$ brightness. We then adjust the brightnesses of the two objects by ± 6 mag relative to each other, produce a co-added spectrum by averaging the fluxes, and then renormalize the resulting composite back to have $I_{\rm AB} = 23.5$, i.e. 1 mag above the survey limit. Gaussian noise is then added in the usual way to the composite SED to represent the survey-C sensitivities.

In Fig. 10 we compute the total number of such blended objects in the mock catalogues, i.e. fraction of objects which have $\Delta I_{\rm AB} \leq 2$ and are within 2 arcsec angular separation from each other, then they are considered as blended objects. We find that at Euclid survey limit almost 15 per cent of the objects are blended. Out of these objects we identify the composite objects for which the redshift difference between the two individual galaxies is greater than the nominal photo-z error, i.e. $\geq 0.05(1+z)$. These objects will be 'dangerous' as they may affect the photo-z estimation. Physically associated pairs at the same redshift are presumably more benign. We find that almost 12 per cent of the objects are such dangerous pairs.

ZEBRA is then run on these set of blended objects and the resulting likelihood curves of each composite object, along with a single maximum likelihood redshift, are obtained. In Fig. 11 we show one such merged pair. Their individual redshifts are given in brackets and we plot $L(z)$ as a function of redshift and magnitude difference between the pairs. We see clearly that within $\Delta I_{\rm AB} \leq 2$ the photo-z estimates become unreliable and the likelihood curves behave unpredictably.

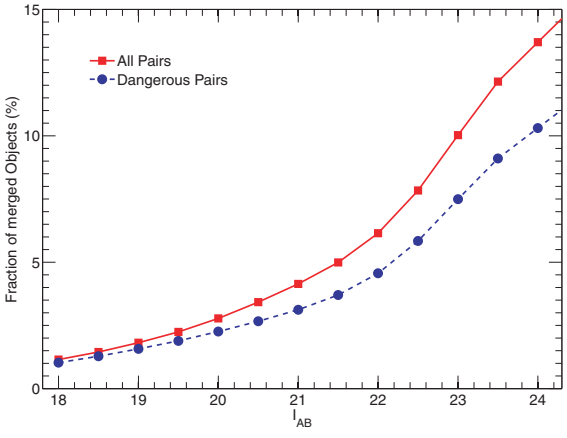

Figure 10. Fraction of objects which will be blended computed from the mock catalogues. Here the solid line gives the total fraction of blended objects with magnitude difference of 2 mag or less and the dashed line gives the fraction of objects where the pairs are at different redshifts and will cause an error in the photo-z estimates.

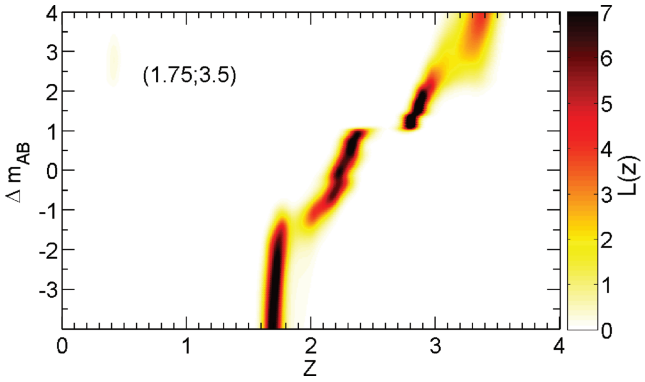

Figure 11. $L(z)$ function of a pair of blended objects. The $L(z)$ function is well behaved in regions other than $\Delta m \leq \pm 2$. In bracket the individual redshifts of the blended pair is given; in this case the low-redshift object is at a redshift of 1.75 and the high-redshift object is at a redshift of 3.5. When Δm approaches zero the likelihood function gets unreliable. Here the objects are taken for survey-C-like depth. The objects are chosen such that one object is a red galaxy and the other is a blue galaxy. The colour axis (vertical) gives the value of the likelihood function.

Bimodality in the likelihood curves can be seen as well as local islands are formed in likelihood redshift space at some intermediate redshifts. In Fig. 12, each box represents a single composite object consisting of a red galaxy and a blue galaxy (individual redshifts are indicated inside the panel within brackets). In each panel, the likelihood curve is plotted as a function (vertically) of the adopted magnitude difference.

We see from Fig. 12 that if a pair of objects consists of two galaxies at similar redshifts then there is a smooth transition from one redshift to the other. When the magnitude difference is 2 mag, or more, the redshift of the brighter galaxy is returned. When the contrast is lower, the maximum likelihood redshift transitions smoothly between the two, but will not accurately represent either component. When the pairs are at different redshifts, i.e. $\Delta z \geq 0.75$ the returned redshifts still trace the redshift of the brighter of the pair for $\Delta I_{\rm AB} > 2$ mag. However, for the region $\Delta I_{\rm AB} < 2$, the behaviour is more varied. Bimodality in the $L(z)$ is often seen, and sometimes a local maximum at a third intermediate redshift. There are often sharp transitions. For the classically degenerate redshift pairs i.e. [0.25,0.5, 2.5,3.5], the photo-z generally make a sharp transition between the two redshifts.

Our general conclusion from this analysis is that the photo-z of blended objects are trustworthy only if the second component is at least 2 mag fainter than the primary (in a waveband close to the middle of the spectral range of the photometry). At smaller magnitude differences the photo-z can be corrupted in a way that is not always recognizable, and we suggest therefore that these blended objects be recognized morphologically from the images, and excluded from the analysis. Fortunately, one would probably want to do this anyway for a lensing analysis because their shape measurements would be hard to use quantitatively.

7 CONCLUSIONS

In this paper we have investigated a number of issues that could potentially limit the photo-z performance of deep all-sky surveys, and thereby impede the ambitious precision cosmology goals of survey programmes such as the proposed ESA Euclid mission. In each case, we find that, while standard techniques do not get to the required accuracy, simple new approaches can be developed

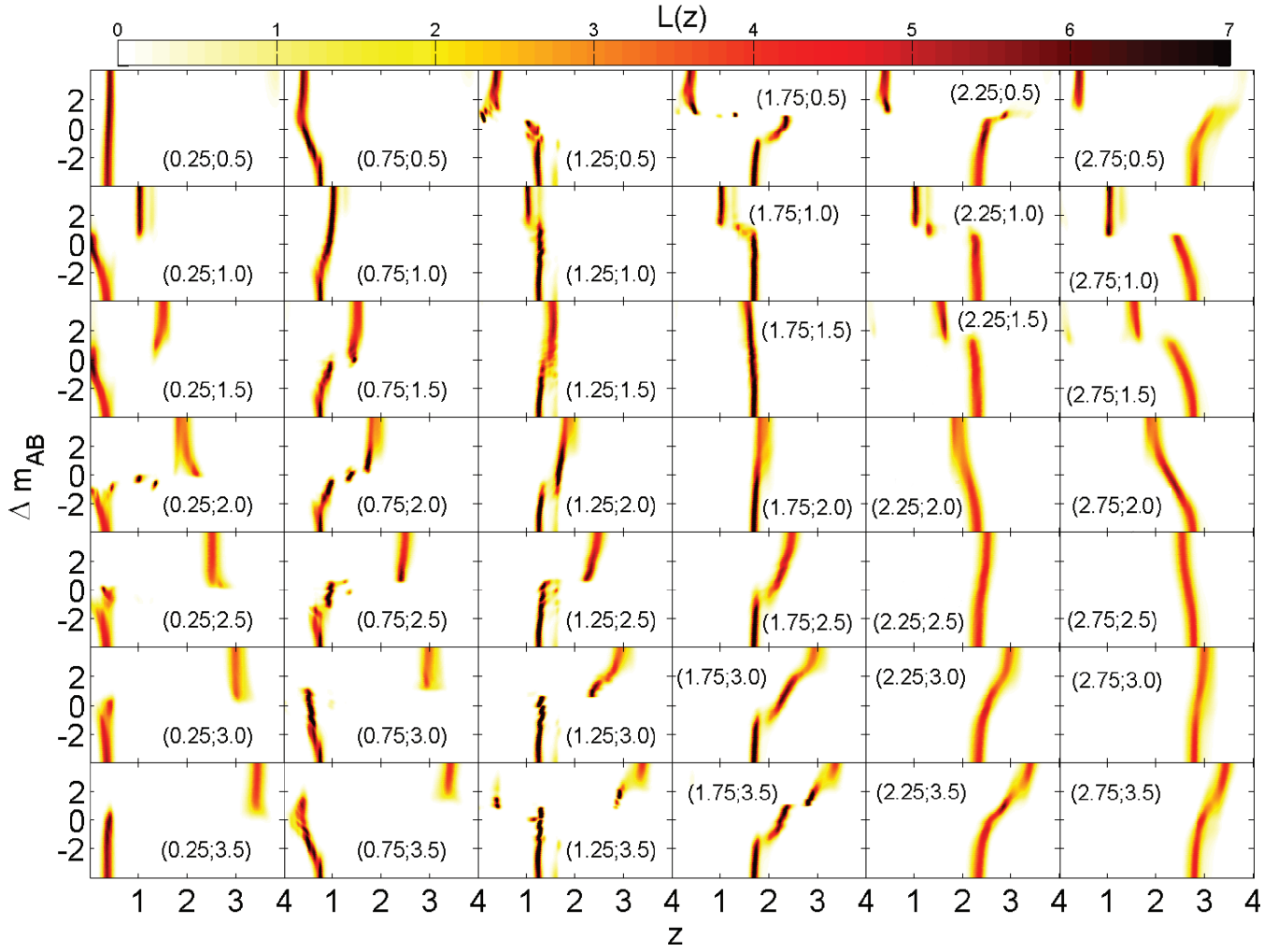

Figure 12. $L(z)$ functions of 42 pairs of blended objects. The objects were boosted to $[-6\ 6]$ mag differences in the I band and variation of likelihood functions are observed. Here, when the low-redshift object is brighter than the high-redshift one, the likelihood function traces the low-redshift object. When both the objects are equally bright we see bimodality in the likelihood function and as magnitude difference increases the likelihood function jumps to trace the high-redshift object. The degeneracy of the likelihood functions at $\Delta m \sim 0$ means that the redshift estimation gets completely unreliable at that region. Here the objects are taken for survey-C-like depth. The objects are chosen such that in each pair one object is a red galaxy and the other is a blue galaxy.

that appear, at least in simple simulations, to get to the required performance.

Knowledge of the redshifts enters into weak-lensing analyses at two distinct steps: first, the selection of objects for shape cross-correlation, and secondly, the accurate knowledge of the mean redshifts of a given set of galaxies. For practical reasons, the first step will likely require the use of photo-z for the foreseeable future. A major motivation for this paper has been to develop techniques that rely on photo-z also for the second step, bypassing the need for very large and highly statistically complete spectroscopic surveys (e.g. Abdalla et al. 2008) and hopefully avoiding the substantial practical difficulties that will be encountered in such spectroscopic surveys from incompleteness and the effects of large-scale structure. Appendix A explores the latter effects in some detail.

The work is based entirely on simulated photometric catalogues that have been constructed to match the expected performance of three generic ground-based surveys, combined with the expected near-IR imaging photometry from Euclid. To construct these catalogues, we use the same set of 10 000 templates as used for the template-fitting photo-z program (ZEBRA). This possibly circular approach allows us to remove the choice of templates as a variable. We believe that it is justified at the current time by the very impressive performance of template-fitting photo-z codes applied to the deep multiband COSMOS photometry, which strongly suggest that the choice of templates will not be a limiting factor at the required level, although further refinement will be desirable.

The analysis is conveniently summarized in terms of the two main requirements on photo-z for precision weak lensing, namely to obtain an rms precision per object of $\leq 0.05(1+z)$ and a systematic bias on the mean redshift of a given set of galaxies of $\leq 0.002(1+z)$. Our main conclusions may be summarized as follows.

(i) To achieve an rms photo-z accuracy of $\sigma_z(z)/(1+z) \sim 0.05$ down to $I_{\rm AB} \leq 24.5$, we need the combination of ground-based photometry with the characteristics of what we call survey-B (similar to PS-2 or DES) and the deep all-sky NIR survey from Euclid itself. This performance also requires the implementation of an 'a priori' rejection scheme (i.e. based on the photometry alone, without knowledge of the actual redshifts of any galaxies) that rejects 13 per cent of the galaxies and reduces the fraction of 5σ outliers to below $f_{\rm cat} < 0.25$ per cent. There is a trade-off between the rejection of outliers and the loss of 'innocent' galaxies with usable photo-z. Deeper photometry improves both the statistical accuracy

Table 3. The $\langle \sigma_z(z)/(1+z) \rangle$ for the three surveys studied. After cleaning and correction have been performed survey-B just about reaches $\sigma_z(z)/(1+z) \sim 0.05$ Euclid requirements.

	$\langle \frac{\sigma_z(z)}{1+z} \rangle$ for different surveys in the range $0.3 \leq z \leq 3.0$		
Survey	Before cleaning	After cleaning	After cleaning + correction
Survey-A	0.1703	0.0884	0.0675
Survey-B	0.1164	0.0640	0.0497
Survey-C	0.0876	0.0492	0.0398

of the photo-z and reduces the wastage in eliminating the catastrophic failures, and the combination of our survey-C (similar to LSST) with Euclid NIR photometry achieves $\sigma_z(z)/(1+z) \sim 0.04$ after 9 per cent rejection (see Table 3).

(ii) A good way to determine the actual $N(z)$ of a set of galaxies in a given photo-z selected redshift bin is to sum their individual probability density functions. We find that the sum of the likelihood functions $\sum L(z)$ already represents quite well both the wings of the $N(z)$ and the remaining catastrophic failures. However, to reach the required performance on the mean of the redshifts $|\Delta_{\langle z \rangle}| \leq 0.002(1+z)$ with the survey-B, or deeper survey-C (together with Euclid IR photometry), we implement a modification scheme to construct individual probability density functions $\mathscr{L}'(z)$ that are based on the individual likelihood functions $L(z)$. This is based on the spectroscopic measurement of redshifts for a rather small number of galaxies (less than 1000) with relaxed requirements on statistical completeness (and no dependence on large-scale structure in the spectroscopic survey fields). We then require that the distribution of the actual redshifts within the probability space that is defined by the individual $\mathscr{L}'(z)$ should be flat across the sample as a whole. This should be true of any set of galaxies, leading to relaxed requirements on sampling of the redshift survey. This approach is similar to the application of a Bayesian prior on the redshifts, but is performed in probability space. Although it cannot be rigorously justified, it is found to work well in practice in our idealized simulations, but should now be tested against more realistic observational data.

(iii) We find that uncertainties in foreground Galactic reddening can have a serious effect in perturbing the photo-z, with a net bias that varies erratically with redshift. However, we also find that such errors in A_V can be identified internally from the photometric data of galaxies, either with or without spectroscopic redshifts. This procedure works best for galaxies with relatively high-S/N photometry $I_{\rm AB} \leq 22$. The required number of galaxies suggests that a reddening map on the scales of 0.1 deg^2 can be internally constructed from the data on galaxies without known redshifts, or from a few hundred galaxies with spectroscopic redshifts.

(iv) We also explored the effect of blended objects, which may constitute 15 per cent of all galaxies at $I_{\rm AB} = 24.5$. The photo-z of the composite SED is a good representation of the redshift of the brighter object as long as the magnitude difference is large, i.e. $\Delta I_{\rm AB} > 2$. When the galaxies are more similar in brightness, $\Delta I_{\rm AB} < 2$, there is a wide range of behaviour. In some cases, multimodal likelihood functions appear, while in others there is a sharp transition from one redshift to another, sometimes with a local maximum at a third, completely spurious redshift. In still others, the likelihood function smoothly transitions between the two redshifts with a single peak at an intermediate redshift. Our conclusion is that composite objects with $\Delta I_{\rm AB} < 2$ should be recognized morphologically from imaging data and removed from the photo-z analysis.

The general conclusion of this paper is that while reaching the photo-z performance required for weak-lensing surveys such as Euclid will not be trivial, the implementation of promising new techniques, coupled with internal calibration of e.g. foreground reddening from the photometric data itself, should allow the required performance to be attained. If more reliance is placed on the photo-z themselves, then this may lead to a significant simplification of the otherwise challenging requirement for spectroscopic calibration of large-scale photometric surveys.

ACKNOWLEDGMENTS

We thank Thomas Bschorr for the use of his template-matching program used in the construction of the mock catalogues, and Pascal Oesch and Robert Feldmann for help in running ZEBRA. We have benefited during the course of this work from useful discussions with many members of the Euclid Science Study Team (ESST) and the Euclid Imaging Consortium (EIC).

REFERENCES

Abdalla F. B., Amara A., Capak P., Cypriano E. S., Lahav O., Rhodes J., 2008, MNRAS, 387, 969
Albrecht A. et al., 2006, preprint (astro-ph/0609591)
Amara A., Réfrégier A., 2007, MNRAS, 381, 1018
Amara A., Réfrégier A., 2008, MNRAS, 391, 228
Avni Y., 1976, ApJ, 210, 642
Benítez N., 2000, ApJ, 536, 571
Benítez N. et al., 2009, ApJ, 692, L5
Bolzonella M., Miralles J.-M., Pelló R., 2000, A&A, 363, 476
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Bridle S., King L., 2007, New J. Phys., 9, 444
Bridle S. et al., 2010, MNRAS, in press (arXiv:0908.0945)
Brodwin M., Lilly S. J., Porciani C., McCracken H. J., Le Fèvre O., Foucaud S., Crampton D., Mellier Y., 2006, ApJS, 162, 20
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Cai Y.-C., Angulo R. E., Baugh C. M., Cole S., Frenk C. S., Jenkins A., 2009, MNRAS, 395, 1185
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Castro P. G., Heavens A. F., Kitching T. D., 2005, Phys. Rev. D, 72, 023516
Cimatti A. et al., 2009, Exp. Astron., 23, 39
Collister A. A., Lahav O., 2004, PASP, 116, 345
Connolly A. J., Szalay A. S., Bershady M. A., Kinney A. L., Calzetti D., 1995, AJ, 110, 1071
Feldmann R. et al., 2006, MNRAS, 372, 565
Feldmann R., Carollo C. M., Porciani C., Lilly S. J., Oesch P., 2008, preprint (arXiv:0801.3275)
Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526
Hu W., 1999, ApJ, 522, L21
Ilbert O. et al., 2009, ApJ, 690, 1236
Joachimi B., Schneider P., 2008, A&A, 488, 829
Joachimi B., Schneider P., 2009, A&A, 507, 105
King L., Schneider P., 2002, A&A, 396, 411
Kitching T. D., Taylor A. N., Heavens A. F., 2008, MNRAS, 389, 173

Kitzbichler M. G., White S. D. M., 2007, MNRAS, 376, 2
Lilly S. J. et al., 2007, ApJS, 172, 70
Lilly S. J. et al., 2009, VizieR Online Data Catalog, 217, 20070
Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118
Ma Z., Bernstein G., 2008, ApJ, 682, 39
Ma Z., Hu W., Huterer D., 2006, ApJ, 636, 21
Madau P., 1995, ApJ, 441, 18
Peacock J. A., Schneider P., Efstathiou G., Ellis J. R., Leibundgut B., Lilly S. J., Mellier Y., 2006, Technical report, ESA–ESO Working Group on 'Fundamental Cosmology'. ESA, Noordwijk
Réfrégier A. et al., 2006, in Mather J. C., MacEwen H. A., de Graauw M. W. M., eds, SPIE Conf. Ser. Vol. 6265, Space Telescopes and Instrumentation I: Optical, Infrared and Millimeter. SPIE, Bellingham, p. 62651Y
Scoville N. et al., 2007, ApJS, 172, 1
Somerville R. S., Lee K., Ferguson H. C., Gardner J. P., Moustakas L. A., Giavalisco M., 2004, ApJ, 600, L171
Springel V. et al., 2005, Nat, 435, 629
Trenti M., Stiavelli M., 2008, ApJ, 676, 767

APPENDIX A: SAMPLING REQUIREMENTS FOR DIRECT SPECTROSCOPIC CALIBRATION

In this appendix, we analyse the number of independent fields that are required if one takes the approach of determining $N(z)$ for a given redshift bin from spectroscopic observations of a representative set of galaxies from that bin. The requirement derives from desiring that the effects of large-scale structure in the galaxy distribution, also known as cosmic variance, are at most equal to the Poisson noise in determining the error in the mean redshift $\langle z\rangle$.

The effect of large-scale structure on the *numbers* of objects in a given survey field is a well-studied problem which can be studied analytically by integrating the two-point correlation function (see e.g. Somerville et al. 2004; Trenti & Stiavelli 2008). We however want to know the effect of structure on the *mean redshift* of the galaxies (within some redshift range) in a given patch of sky, which is a less straightforward problem since it concerns structure within the redshift range rather than the simple overall number of galaxies. For this reason, we take a semi-empirical approach based on the same set of 24 COSMOS mock catalogues (Kitzbichler & White 2007) as was used for the main paper. These 24 mocks are all derived from the same numerical dark matter simulation (the 'Millennium Run') but the 24 light cones sample this in such a way that a given part of the simulation appears in the different mocks at random redshifts. Therefore the large-scale structure at any given redshift is independent from one mock to the next, and comparison of the 24 different mocks will therefore give a good idea of the effect of large-scale structure that would be observed 'on sky'. The galaxy catalogues are of course derived from a particular implementation of a semi-analytic model, and are thus somewhat arbitrary. On the other hand, the clustering of galaxies at $I_{\rm AB} < 24.5$ is not very well constrained empirically on the scales and at the redshifts that are of most interest here, so we believe that our straightforward approach is fully justified. It should be noted that the mock catalogues 'automatically' contain all of the information of the two-point and higher correlation functions, on all scales [up to the Millennium Run box scale of $500\,h^{-1}$ MPc on a side (Springel et al. 2005)]. They also include the effect of peculiar motions in redshift space.

We compute the effects of this large-scale structure on $\langle z\rangle$, the average redshift of a set of galaxies in some defined redshift bin, as follows. We first look at all the galaxies in this redshift bin, across the whole 2-deg^2 field and across all of the 24 mocks. The variance of their individual redshifts is given by the following expression:

$$\sigma^2 = \frac{1}{n}\sum(z-\overline{z})^2 = \frac{1}{n}\sum(z)^2 - (\overline{z})^2. \tag{A1}$$

For a top-hat distribution of redshifts within a redshift bin of width Δz, this quantity would be given by

$$\sigma^2 \approx \frac{(\Delta z)^2}{12}. \tag{A2}$$

We then imagine carrying out a redshift survey using a spectrograph with a particular field of view, which is assumed to be square. We presume that galaxies (within this redshift bin) are randomly selected across this field of view. We then compute for each mock i, the average redshift of all the $n_{{\rm s},i}$ galaxies that fall within the spectrograph field of view, which we denote by ζ_i. The variance of this quantity across the m mocks is then computed as

$$\sigma_\zeta^2 = \frac{1}{m}\sum(\zeta-\overline{\zeta})^2 = \frac{1}{m}\sum(\zeta)^2 - (\overline{\zeta})^2. \tag{A3}$$

In the Poisson case, this variance would be equal to the variance coming from the $n_{\rm s}$ galaxies in the field, which will be approximately given by $\sigma^2/n_{\rm s}$, or more precisely by

$$\sigma_P^2 = \frac{\sigma^2}{m}\sum_{i=1}^{m}\frac{1}{n_{{\rm s},i}} = \sigma^2\left\langle\frac{1}{n_{{\rm s},i}}\right\rangle. \tag{A4}$$

The difference between the 'observed' σ_ζ^2 and that from the Poisson variance σ_P^2 is the contribution of the large-scale structure, or cosmic variance, $\sigma_{\rm CV}^2$:

$$\sigma_{\rm CV}^2 = \sigma_\zeta^2 - \sigma^2\left\langle\frac{1}{n_{{\rm s},i}}\right\rangle. \tag{A5}$$

Given these estimates, we can establish, for each redshift bin and for each spectrograph field of view, a critical number of spectroscopic redshift measurements at which the cosmic variance will be equal to the Poisson variance:

$$N_{\rm crit} = \frac{\sigma^2}{\sigma_{\rm CV}^2}. \tag{A6}$$

As the number of spectroscopic redshifts reaches this level, the standard deviation in the average redshift estimate ζ is already $\sqrt{2}$ times higher than would be the case for Poisson noise alone, and obtaining more spectra in the same field will lead to little further gain in accuracy in the estimate of the global mean redshift $\langle z\rangle$ because the mean redshift in this particular survey field will be dominated by structure in that field.

As a detail, for spectrograph fields of view that are much smaller than the mock, we can repeat the exercise for different locations of the spectrograph field within the 2-deg^2 field of the mocks, averaging the calculations of σ_ζ^2 to improve the definition of the variance.

We consider spectrograph fields of view given by $(1.4/N)^\circ$, with $N = 1,\ 6$, yielding in each case N^2 different locations with the 2-deg^2 field. Fig. A1 shows the derived $N_{\rm crit}$ for galaxies with

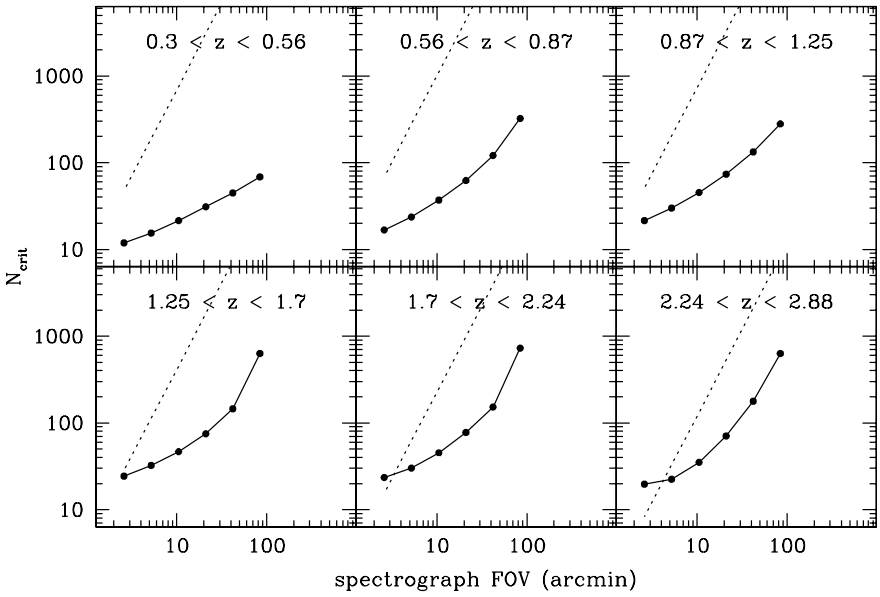

Figure A1. The critical number of redshifts at which the variance in the mean redshift becomes dominated by the effects of large-scale structure or cosmic variance. The solid lines give the derived $N_{\rm crit}$ for $I_{\rm AB} < 24.5$ in six representative redshift bins for spectrograph fields of view from a few arcmin up to 1°.4. The dotted line gives the average number of galaxies within the field of view of the spectrograph. The difference between these lines indicates the maximum permitted sampling rate. A fully sampled spectroscopic survey is likely to be severely cosmic variance limited and much lower sampling rates are required to keep the effects of large-scale structure comparable to the Poisson term.

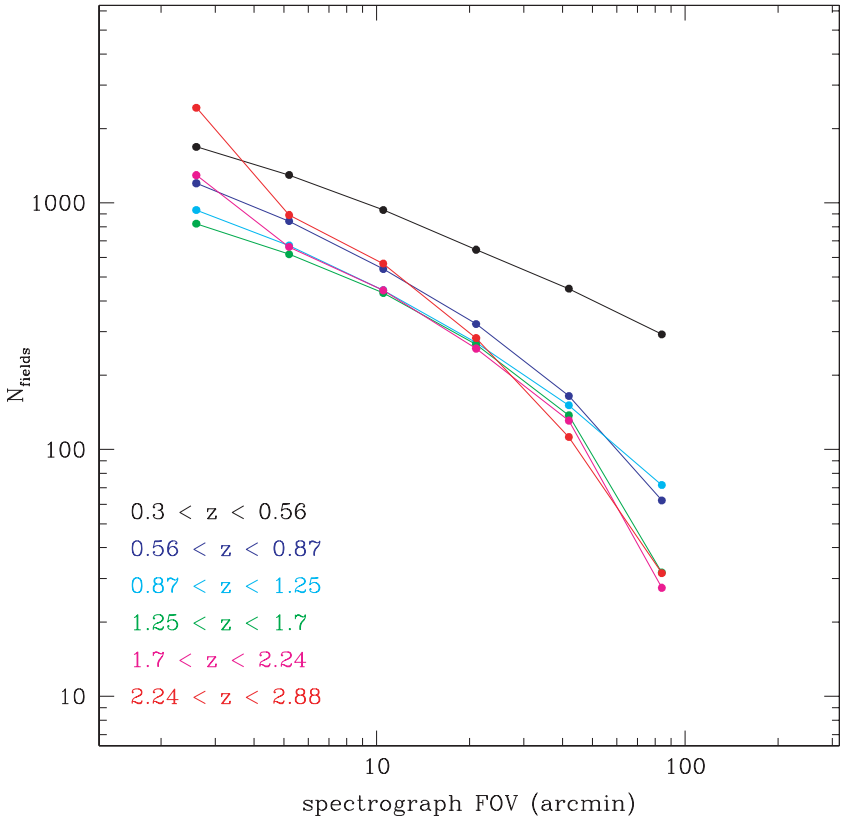

Figure A2. The minimum number of independent fields required in order to ensure that the uncertainty in the mean redshift of a given redshift bin is not dominated by large-scale structure or cosmic variance, and is equal to the Poisson variance from 10^4 galaxies. Clearly for an all-sky survey, the number of required fields will be set by lower redshift range.

$I_{AB} < 24.5$ in six representative redshift bins for spectrograph fields of view ranging from a few arcmin (e.g. NIRSpec on *JWST*) up to fields of 1.4°, which is approaching the largest that is likely practical on an 8-m class telescope. For reference, the current VIMOS spectrograph on the Very Large Telescope has a field of about 15 arcmin^2.

The rather low values of N_{crit} produced by this analysis reflect the familiar observational experience that the large-scale structure in a given survey field starts to become apparent in the $N(z)$ distribution after only a few galaxies have been observed. In particular, it can be seen that the value of N_{crit} is usually very much smaller than the average number of galaxies within the field of view of the spectrograph, which is shown as the dotted line in each panel. The difference between these curves indicates the maximum permitted sampling rate. This emphasizes that the redshift distribution from a fully sampled spectroscopic survey is likely to be severely cosmic variance limited and that very much lower sampling rates are required to keep the effects of large-scale structure comparable to the Poisson term. For example, a survey with the VIMOS spectrograph with a field of view of about 15 arcmin^2 would require a sampling rate at low redshift of 2 per cent (i.e. one galaxy in 50) or less. Only at the highest redshifts and with the smallest field sizes cosmic variance is unimportant. The low sampling rates demanded by this analysis would pose quite severe inefficiencies on the utilization of slit-mask spectrographs for such a program of spectroscopic calibration of photo-z bins. These would be mitigated for fibre-fed spectrographs, although the performance of these at $I_{AB} \sim 24.5$ has not yet been proven. Regardless, it is clear that the survey fields for such a programme would have to be distributed over a significant portion of the sky.

The point at which N_{crit} starts to increase as the square of the field of view (i.e. where the solid line becomes parallel to the dotted lines) shows the point at which it is safe to mosaic adjacent survey fields to build up survey area and galaxy number. This occurs at about degree-scales for $z > 1.7$.

If we take the minimum of N_{crit} and the available number of galaxies in the field, which is almost always given by the former, we can then compute the minimum number of independent fields that will be required in order to attain an uncertainty in the mean redshift of this particular redshift bin $\langle z \rangle$ that is equivalent to the Poisson variance from 10^4 galaxies. This is shown in Fig. A2.

If one imagines doing a single survey to cover the entire redshift range, then the number of fields will generally be set by the lower redshifts, where the effects of large-scale structure are most severe. Only at very small field sizes does the low number density of very high-redshift galaxies become the limiting factor. It can be seen that about 400 widely spaced degree-scale survey fields, 700 VIMOS fields or about 2000 NIRSpec (3-arcmin) survey fields would be required. This requirement clearly approximates an all-sky sparse sampled survey.

The difficulty of implementing such a scheme was a major motivation for considering the alternative approach to constructing $N(z)$ which is developed in the main paper. The effects of large-scale structure are irrelevant in that approach which considers photo-z calibration on an object-by-object level.

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