Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **423,** 909–924 (2012) doi:10.1111/j.1365-2966.2012.20927.x

Sample variance in photometric redshift calibration: cosmological biases and survey requirements

Carlos E. Cunha,[1,2⋆] Dragan Huterer,[1] Michael T. Busha[2,3] and Risa H. Wechsler[2,4,5]

[1]*Department of Physics, University of Michigan, 450 Church St, Ann Arbor, MI 48109-1040, USA*
[2]*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA*
[3]*Institute for Theoretical Physics, University of Zurich, 8057 Zurich, Switzerland*
[4]*Department of Physics, Stanford University, Stanford, CA 94305, USA*
[5]*SLAC National Accelerator Laboratory, 2575 Sand Hill Rd., MS 29, Menlo Park, CA 94025, USA*

Accepted 2012 March 14. Received 2012 March 9; in original form 2011 October 9

ABSTRACT

We use *N*-body/photometric galaxy simulations to examine the impact of sample variance of spectroscopic redshift samples on the accuracy of photometric redshift (photo-z) determination and calibration of photo-z errors. We estimate the biases in the cosmological parameter constraints from weak lensing and derive requirements on the spectroscopic follow-up for three different photo-z algorithms chosen to broadly span the range of algorithms available. We find that sample variance is much more relevant for the photo-z error calibration than for photo-z training, implying that follow-up requirements are similar for different algorithms. We demonstrate that the spectroscopic sample can be used for training of photo-zs and error calibration without incurring additional bias in the cosmological parameters. We provide a guide for observing proposals for the spectroscopic follow-up to ensure that redshift calibration biases do not dominate the cosmological parameter error budget. For example, assuming optimistically (pessimistically) that the weak lensing shear measurements from the Dark Energy Survey could obtain 1σ constraints on the dark energy equation of state w of 0.035 (0.055), implies a follow-up requirement of 150 (40) patches of sky with a telescope such as Magellan, assuming a 1/8 deg^2 effective field of view and 400 galaxies per patch. Assuming (optimistically) a VIMOS-VLT Deep Survey-like spectroscopic completeness with purely random failures, this could be accomplished with about 75 (20) nights of observation. For more realistic assumptions regarding spectroscopic completeness, or with the presence of other sources of systematics not considered here, further degradations to dark energy constraints are possible. We test several approaches for making the requirements less stringent. For example, if the redshift distribution of the overall sample can be estimated by some other technique, e.g. cross-correlation, then follow-up requirements could be reduced by an order of magnitude.

Key words: cosmological parameters – cosmology: observations – cosmology: theory – dark energy – large-scale structure of Universe.

1 INTRODUCTION

One of the principal systematic errors affecting surveys that utilize the large-scale structure (LSS) to study dark energy is the quality of the photometric redshifts (hereafter photo-zs). Because of time and throughput constraints it is costly and impractical to obtain spectroscopic redshifts for more than a small fraction of galaxies. Upcoming surveys such as the Dark Energy Survey[1] (DES), PanStarrs,[2] Hyper-Suprime Cam survey[3] (HSC) and the Large Synoptic Survey Telescope[4] (LSST) will have to rely on the photo-zs in order to utilize the three-dimensional information from the large number of galaxies observed in these surveys. Without the redshift information, one loses the ability to perform weak lensing tomography (Hu 1999), and thus degrades the ability to measure the temporal evolution of dark energy in the recent ($z \lesssim 1$) history of the Universe (for reviews, see Bartelmann & Schneider 2001; Hu 2002; Huterer 2002, 2010; Amara & Refregier 2007; Hoekstra & Jain 2008; Munshi et al. 2008).

⋆E-mail: ccunha@stanford.edu
[1] http://darkenergysurvey.org
[2] http://pan-starrs.ifa.hawaii.edu
[3] http://oir.asiaa.sinica.edu.tw/hsc.php
[4] http://lsst.org

Photo-z techniques use broad-band photometry, i.e. the measured flux through a few bands, to estimate approximate galaxy redshifts. Other observable quantities (hereafter 'observables'), such as galaxy shape measures, can also be used, but they typically have limited redshift information. The intrinsic uncertainty of photo-zs can contribute significantly to the error in the inferred cosmological parameters.

There are two broad, overlapping, categories of photo-z estimators. Template-fitting algorithms (e.g. Arnouts et al. 1999; Benitez 2000; Bolzonella, Miralles & Pelló 2000; Budavári et al. 2000; Csabai et al. 2003; Feldmann et al. 2006) assign photo-zs to a galaxy by finding the template and redshift that best reproduce the observed fluxes. Training set methods (e.g. Connolly et al. 1995; Firth, Lahav & Somerville 2003; Wadadekar 2005; Wang et al. 2007; Gerdes et al. 2010), on the other hand, use a spectroscopic sample to characterize a relation between the photometric observables and the redshifts, which is then applied to the full photometric sample. The distinction between the two categories is muddled because template-fitting methods can also use spectroscopic redshifts to improve the fitting. Conversely, training set methods can be based on catalogues simulated using templates. For reviews and comparison of methods see e.g. Hogg et al. (1998), Koo (1999), Hildebrandt et al. (2010) and Abdalla et al. (2011).

Spectroscopic redshifts (hereafter spec-zs) play three important roles in photo-z analysis. First, as described above they improve the accuracy of photo-z estimation. Having accurate photo-zs is highly desirable for cosmology, as photo-z errors inevitably smear the radial information describing galaxy clustering. Secondly, spectroscopic redshifts characterize the photo-z errors (see Ref. Oyaizu et al. 2008b, for a review). With accurate error estimation, one can remove or downweight the less reliable photo-zs, decreasing their impact on the cosmological analysis. Thirdly, spec-zs characterize the uncertainties in the photo-z error distribution ('error in the error'), which is a key quantity that needs to be accurately known. In particular, even if photo-zs are not exceptionally accurate and there are regions of badly misestimated redshift (i.e. the 'catastrophic errors'), one can still recover the cosmological information provided the bias, scatter and ideally the full distribution in the $z_{\rm p}$–$z_{\rm s}$ plane, are accurately calibrated using the subset of galaxies with spectroscopic information.

The requirements on spectroscopic samples due to the three requirements mentioned are not independent, but have been treated as such in the past. For example, the amount of spectroscopic follow-up required for calibration depends on the intrinsic accuracy of the photo-zs (Huterer et al. 2006; Ma, Hu & Huterer 2006; Ma & Bernstein 2008) and on the identification of regions with unreliable photo-zs (Sun et al. 2009; Bernstein & Huterer 2010, hereafter BH10; Hearin et al. 2010), but the ability to do both of these is strongly dependent on the use of spec-zs for training and error estimation. At this point, the careful reader may wonder: can the same spectroscopic sample be used for photo-z training, error estimation and calibration without significantly biasing cosmological results? Yes, it turns out, as we will show in this paper.

Obtaining spectra for thousands of galaxies needed for photo-z studies is a very difficult task, which complicates their use in photo-z studies. Spectroscopic surveys can be far from a representative subsample of the photometric sample for five principal reasons.

(i) *Shot noise.* Spectroscopic samples to the depth required are quite small, hence Poisson fluctuations due to the finite number of galaxies are significant.

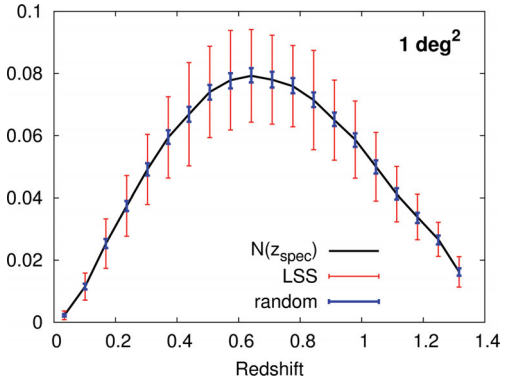

Figure 1. Normalized spectroscopic redshift distribution for the full data. The red (light grey) error bars show the 1σ variability in the redshift distribution for contiguous 1 deg^2 angular patches. The blue (dark grey) error bars show the variability in the redshift distribution assuming random samples of with the same mean number of objects as the 1 deg^2 patches. We assume that only a 25 per cent random subsample of each patch is targeted for spectroscopy, yielding about 1.2×10^4 galaxies per patch on average.

(ii) *Sample variance.* Spectroscopic surveys designed to reach the magnitude limits of the upcoming photometric surveys typically have very small angular apertures, much smaller than fluctuations introduced by large-scale clustering of galaxies. The fluctuations due to sample variance can be an order of magnitude larger than shot-noise fluctuations for samples of around 1 deg^2 (see e.g. van Waerbeke et al. 2006, and Fig. 1).

(iii) *Type incompleteness.* Strength of spectral features vary significantly with galaxy type. In addition, the wavelength coverage of most spectrographs is not sufficient to detect some of the main features through the full redshift range of interest.

(iv) *Incorrect redshifts.* Line misidentification can yield incorrect redshifts. The number of incorrect spectroscopic redshifts can be reduced – by keeping only the most reliable galaxies – at the cost of increasing the incompleteness.

(v) *Sample variance in observing conditions.* Variations in imaging conditions (e.g. seeing and photometric quality) during a survey imprint an angular dependence to the survey depth and completeness.

Past papers on the effects of photometric redshift errors on dark energy constraints (Huterer et al. 2006; Ma et al. 2006; Amara & Refregier 2007; Abdalla et al., 2008; Kitching, Taylor & Heavens 2008; Ma & Bernstein 2008; Bordoloi, Lilly & Amara 2010; Hearin et al. 2010) have studied in detail the distribution of photometric redshifts (more specifically, the full probability density function $P(z_{\rm p}|z_{\rm s})$). Some of these works have extended the analysis to estimate the number of spectra required in order to calibrate the photo-zs. However, in essentially all cases the requirements on the spectroscopic sample have only assumed shot noise, i.e. that the accuracy of the photo-z bias and error in some redshift bin labelled by μ is equal to $\Delta z_{\rm bias}(z_\mu) = \sigma_z(z_\mu)\sqrt{1/N^\mu_{\rm spec}}$ and $\Delta\sigma_z(z_\mu) = \sigma_z(z_\mu)\sqrt{2/N^\mu_{\rm spec}}$, where $N^\mu_{\rm spec}$ is the size of the spectroscopic follow-up sample in that bin (see equation 18 in Ma et al. 2006).

Sample variance was taken into account in spectroscopic follow-up requirements in van Waerbeke et al. (2006) and Ishak & Hirata (2005); however, they only considered the overall redshift distribution of the source sample and did not include photometric redshifts in the simulations (see also Bordoloi et al. 2010, for a related discussion). Requirements on spectrograph design in order to minimize spectroscopic failures were investigated in Jouvel et al. (2009),

with emphasis on designing spectrographs to calibrate redshifts for space-based missions. Finally, the effects of sample variance in observing conditions was investigated by Nakajima et al. (2012) using Sloan Digital Sky Survey (SDSS) imaging and spectroscopic redshifts from several surveys overlapping the SDSS. That paper found that atypical imaging conditions in the spectroscopic fields can lead to biases in galaxy–galaxy lensing analysis, but fortunately concluded that this type of bias can be at least partly corrected (see also Sheldon et al. 2011, for a related discussion).

The main goals of this paper are to study the impact of sample variance in spectroscopic samples to the training of photo-*z*s, error estimation and error calibration, and to assess implications for cosmological constraints from weak lensing tomography analyses. The paper is organized as follows. In Section 2 we describe the photo-*z* algorithms we use in our tests. In Section 3 we describe our construction of the different simulated samples used. We detail the procedure of estimating biases in cosmological constraints from the weak lensing tomography in Section 4. Results are given in Section 5 with a discussion of potential improvements in Section 6. We provide a guide for determining spectroscopic observational requirements in Section 7 and present our conclusions in Section 8. The construction of the simulations is described in Appendix A.

2 PHOTO-*z* ALGORITHMS

We consider three different photo-*z* algorithms that broadly span the space of possibilities. Namely, we use a basic template-fitting code without any priors, a training set fitting method and a training set method that does not perform a fit, but uses the local density in the neighbourhood of an object to derive a redshift probability distribution. We briefly describe each below.

2.1 Template-fitting redshift estimators

Template-fitting estimators derive photometric redshift estimates by comparing the observed colours of galaxies to colours predicted from a library of galaxy spectral energy distributions (SEDs). We use the publicly available LePhare photo-*z* code[5] (Arnouts et al. 1999; Ilbert et al. 2006) as our template-fitting estimator. We chose the extended Coleman–Wu–Weedman (CWW) template library (Coleman, Wu & Weedman 1980) because it yielded the best photo-*z*s for our simulation.

We purposefully ignore all priors for reasons that we now describe. There are essentially two classes of priors, those derived from completely different surveys, and those based on targeted follow-ups of a subsample of the survey for which photo-*z*s are desired. The use of the latter makes template-fitting results quite similar to the training set methods, and would make the template-fitting code subject to a training procedure which would be affected by the sample variance. The use of the former could reduce some outliers, but would also complicate the interpretation of the results; there are several choices of external priors, and if the selection of the sample used to determine the priors is different from that of the survey at hand then redshifts could be biased (see e.g. Abrahamse et al. 2011). As we shall see, the photo-*z* quality is not a dominant factor in our analysis, and a more thorough experimentation of the template-fitting algorithms is not expected to affect conclusions.

[5] http://www.cfht.hawaii.edu/~arnouts/LEPHARE/lephare.html

2.2 Nearest neighbour redshift probability estimators

2.2.1 Weights

In this subsection, we briefly review the weighting method[6] of Lima et al. (2008), which is required for computing redshift probabilities, henceforth $p(z)$. 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. In this case, the DES *griz* magnitudes are our observables. The hypervolume used to estimate the density is set here to be the Euclidean distance of the galaxy to its Nth nearest neighbour in the training set. We set $N = 50$ for the $p(z)$ estimate. Smaller N lead to less broad $p(z)$s and better reconstruction of the overall redshift distribution at the cost of increased shot noise in individual $p(z)$s. If one does not care about individual $p(z)$s then we recommend choosing a smaller N; the optimal choice will depend on the training set size. The bias analysis is not sensitive to the choice of N.

The weights can be used to estimate the redshift distribution of the photometric sample using

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

where the weighted sum is over all galaxies in the training set. Lima et al. (2008) and Cunha et al. (2009) show that this provides a nearly unbiased estimate of the redshift distribution of the photometric sample, $N(z)_{\rm P}$, provided the differences in the selection of the training and photometric samples are solely done in the observable quantities used to calculate the weights.

2.2.2 Probability density $p(z)_{\rm w}$

To estimate the redshift error distribution for each galaxy, $p(z)_{\rm w}$, we adopt the method of Cunha et al. (2009). We use the subscript 'w' to differentiate between our particular estimator and the general concept for redshift probability distributions. The $p(z)_{\rm w}$ for a given object in the photometric sample is simply the redshift distribution of the N (in this case 50) nearest neighbours in the *training* set:

$$p(z)_{\rm w} = \sum_{\beta=1}^{N} w_\beta \delta(z - z_\beta). \quad (2)$$

We estimate $p(z)_{\rm w}$ in 20 redshift bins between $z = 0$ and 1.35.

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

$$N(z)_{\rm p(z)} = \sum_{i=1}^{N_{\rm P,tot}} p_i(z)_{\rm w}. \quad (3)$$

This estimator becomes identical to that of equation (1) in the limit of very large training sets. For training sets smaller than tens of thousands of galaxies, one can improve the $p(z)_{\rm w}$ estimate by multiplying each $p(z)_{\rm w}$ by the ratio of $N(z)_{\rm wei}$ to $N(z)_{\rm p(z)}$.

[6] The weights, $p(z)$ and polynomial codes are available at http://kobayashi.physics.lsa.umich.edu/ ~ccunha/nearest/. The codes can also be obtained in the git repository PROBWTS at http://github.com

We note that several public photo-z codes exist that can output $p(z)$s, e.g. the template-fitting codes LePhare (Arnouts et al. 1999; Ilbert et al. 2006), zebra (Feldmann et al. 2006), bpz (Coe et al. 2006) and the training-set based ArborZ (Gerdes et al. 2010). We do not expect qualitative differences in our conclusions from using the above methods because, as we will show, sample variance affects mostly spectroscopic properties, not photometric.

2.3 Nearest neighbour polynomial fitting redshift estimators

For each galaxy in the photometric sample, the nearest neighbour polynomial (NNP) fitting algorithm uses the N nearest neighbouring galaxies with spectra (i.e. in the training set) to fit a low-order polynomial relation between the redshift and the observable quantities (e.g. colours and magnitudes). It then applies this function to the observables of the galaxy in the photometric sample and assigns it a redshift. We use a second-order polynomial in this study and check that a first-order polynomial does not change results by more than a few per cent. The NNP method was introduced by Oyaizu et al. (2008a) and produces photo-zs that are very similar to the neural networks. We chose the NNP here because it is very fast compared to other codes for photometric samples with up to a few million objects in size. In addition, we can directly compare the results of the NNP photo-zs with the $p(z)_{\rm w}$ since both are based on the same set of training-set galaxies. As with the $p(z)_{\rm w}$ method, the choice of which N nearest neighbours are to be used does not affect results significantly, provided there are enough galaxies to characterize the coefficients of the polynomial fit and avoid overfitting. For a second-order polynomial with four observables, we find that $N = 100$ is a good comprise between retaining locality of colour information and stability of the fit. Results presented here use a slightly more aggressive $N = 80$, but this does not affect the bias results meaningfully.

3 SIMULATED DATA

3.1 Selection

We use a cosmological simulation, populated with galaxies and their photometric properties, fully described in Appendix A. The simulation consists of a 220 deg^2 photometric survey in the *grizY* DES bands with 10σ magnitude limits of [24.6, 24.1, 24.4, 23.8, 21.3]. For this study, we disregard the Y band since we find it does not improve the photo-zs. We select only galaxies with $i < 24$ which are also detected (to 5σ) in the grz bands. The original catalogue contains 13 550 386 galaxies, and after the cuts we are left with $N_{\rm data}$ = 10 780 625 galaxies. To speed up the training and calibration of the photo-zs, we pick a random subsample of about $N_{\rm phot} = 4000\,000$ galaxies to be our photometric sample.

3.2 Training and calibration samples

We construct our spectroscopic training and calibration samples by splitting the simulation output into several sets of $N \times N$ patches of equal area, with each patch being nearly square in shape. When comparing the different photo-z algorithms we use three binning schemes, setting $N = 6$, 15 and 30, which correspond, roughly, to patches of area 6, 1 and 0.25 deg^2, respectively. Because spectroscopic surveys are far from complete, in a sense that they include spectra of only a subset of all photometrically discovered galaxies, we randomly pick a subsample from each patch. Unless stated otherwise, we simulate 25 per cent random completeness, i.e. we use a Monte Carlo approach to downsample by drawing a random number between 0 and 1 for each galaxy and selecting the galaxies for which the number is less than 0.25. The mean number of galaxies per pixel available for training and calibration is about 74 865, 11 978 and 2995 for the 6, 1 and 0.25 deg^2 pixel sets. We refer to the sample created by splitting the data in angular patches as the *LSS samples*.

For each set of LSS samples, we generate what we call the *random-equivalent samples*. The random-equivalent samples are sets of random samples drawn from the full survey but with size similar to the LSS sample patches. For example, the random equivalent patches of the 1 deg^2 LSS patches are generated as follows. There are 225 patches in the 1 deg^2 case. The random-equivalent patches are generated by performing random draws of galaxies from the full data set to generate a new set of 225 patches; each such (random equivalent) patch is generated by including every galaxy from the original catalogue with the probability $N_{\rm patch}/N_{\rm gal}$, where $N_{\rm patch}$ is the average number per patch (e.g. 11 978 in the 1 deg^2 case), while $N_{\rm gal}$ is the total number of galaxies in the simulation. This yields 225 samples that have the same average number of galaxies per patch as the LSS patches.

As discussed in the Introduction, in real spectroscopic surveys the incompleteness is caused not only by random subselection, but also the inability to get spectra for some galaxies. These spectroscopic failures can lead to biases in the training and calibration and we shall explore them in a follow-up paper. Throughout, we use the same set of patches for both training and calibration. In Section 5, we show that this does not add appreciable error to the cosmological constraints.

4 WEAK LENSING BIAS

We wish to quantify how much sample variance due to the LSS contributes to errors in weak lensing shear, and thus errors in the derived cosmological parameter constraints. For simplicity, we only study the shear–shear correlations, and not the related shear–galaxy and galaxy–galaxy power spectra. The observable quantity we consider is the convergence power spectrum:

$$C^{\kappa}_{ij}(\ell) = P^{\kappa}_{ij}(\ell) + \delta_{ij}\frac{\langle\gamma^2_{\rm int}\rangle}{\bar{n}_i}, \quad (4)$$

where $\langle\gamma^2_{\rm int}\rangle^{1/2}$ is the rms intrinsic ellipticity in each component, $\bar{n}_i$ is the average number of galaxies in the ith redshift bin per steradian and ℓ is the multipole that corresponds to structures subtending the angle $\theta = 180^\circ/\ell$. For simplicity, we drop the superscripts κ below. For most of this work we take $\langle\gamma^2_{\rm int}\rangle^{1/2} = 0.16$, which yields very stringent follow-up requirements. We discuss the impact of this choice in Section 5.4.1.

We closely follow the formalism of BH10, where the photometric redshift errors are algebraically propagated into the biases in the shear power spectra. These biases in the shear spectra can then be straightforwardly propagated into the biases in the cosmological parameters. We now review briefly this approach.

Let us assume a survey with the (true) distribution of source galaxies in redshift $n_{\rm S}(z)$, divided into B bins in redshift. Let us define the following terms.

(i) *Leakage* $P(z_{\rm p}|z_{\rm s})$ (or $l_{\rm sp}$ in BH10 terminology): fraction of objects from a given spectroscopic bin that are placed into an incorrect (non-corresponding) photometric bin.

(ii) *Contamination* $P(z_{\rm s}|z_{\rm p})$ (or $c_{\rm sp}$ in BH10 terminology): fraction of galaxies in a given photometric bin that come from a non-corresponding spectroscopic bin.

When specified for each tomographic bin, these two quantities contain the same information. Note in particular that the two quantities satisfy the integrability conditions:

$$\int P(z_{\rm p}|z_{\rm s})\,{\rm d}z_{\rm p} \equiv \sum_{\rm p} l_{\rm sp} = 1, \tag{5}$$

$$\int P(z_{\rm s}|z_{\rm p})\,{\rm d}z_{\rm s} \equiv \sum_{\rm s} c_{\rm sp} = 1. \tag{6}$$

A fraction $l_{\rm sp}$ of galaxies in some spectroscopic-redshift bin $n_{\rm s}$ 'leak' into some photo-z bin $n_{\rm p}$, so that $l_{\rm sp}$ is the fractional perturbation in the spectroscopic bin, while the contamination $c_{\rm sp}$ is the fractional perturbation in the photometric bin. The two quantities can be related via

$$c_{\rm sp} = \frac{N_{\rm s}}{N_{\rm p}}\, l_{\rm sp}, \tag{7}$$

where $N_{\rm s}$ and $N_{\rm p}$ are the absolute galaxy numbers in the spectroscopic and photometric bin, respectively. Then

$$n_{\rm s} \to n_{\rm s}, \tag{8}$$

$$n_{\rm p} \to (1 - c_{\rm sp})\, n_{\rm p} + c_{\rm sp}\, n_{\rm s} \tag{9}$$

and the photometric bin normalized number density is affected (i.e. biased) by photo-z catastrophic errors. The effect on the cross power spectra is then (BH10)

$$\begin{aligned}
C_{pp} &\to (1 - c_{\rm sp})^2 C_{pp} + 2c_{\rm sp}(1 - c_{\rm sp})C_{sp} + c_{\rm sp}^2 C_{ss},\\
C_{mp} &\to (1 - c_{\rm sp})C_{mp} + c_{\rm sp}\, C_{ms} \qquad (m < p),\\
C_{pn} &\to (1 - c_{\rm sp})C_{pn} + c_{\rm sp}\, C_{sn} \qquad (p < n),\\
C_{mn} &\to C_{mn} \qquad\qquad\qquad\quad ({\rm otherwise})
\end{aligned} \tag{10}$$

(since the cross power spectra are symmetrical with respect to the interchange of indices, we only consider the biases in power spectra C_{ij} with $i \le j$). Note that these equations are exact for a fixed contamination coefficient $c_{\rm sp}$.

The bias in the observable power spectra is the right-hand side (rhs)–left-hand side (lhs) difference in the above equations.[7] The cumulative result due to all contaminations in the survey (or, $P(z_{\rm s}|z_{\rm p})$ values for each $z_{\rm s}$ and $z_{\rm p}$ binned value) can be obtained by the appropriate sum

$$\begin{aligned}
\delta C_{pp} &= \sum_{\rm s} (-2c_{\rm sp} + c_{\rm sp}^2)C_{pp} + 2c_{\rm sp}(1 - c_{\rm sp})C_{sp} + c_{\rm sp}^2 C_{ss},\\
\delta C_{mp} &= \sum_{\rm s} \left(-c_{\rm sp} C_{mp} + c_{\rm sp}\, C_{ms}\right),\\
\delta C_{pn} &= \sum_{\rm s} \left(-c_{\rm sp} C_{pn} + c_{\rm sp}\, C_{sn}\right)
\end{aligned} \tag{11}$$

for each pair of indices (m, p), where the second and third line assume $m < p$ and $p < n$, respectively.

The bias in cosmological parameters is given by using the standard linearized formula (Knox, Scoccimarro & Dodelson 1998; Huterer & Turner 2001), summing over each pair of contaminations (s, p):

$$\delta p_i \approx \sum_j (\mathbf{F}^{-1})_{ij} \sum_{\alpha\beta} \frac{\partial \bar{C}_\alpha}{\partial p_j} (\mathbf{Cov}^{-1})_{\alpha\beta}\, \delta C_\beta, \tag{12}$$

[7] We have checked that the quadratic terms in $c_{\rm sp}$ are unimportant, but we include them in any case.

where $\mathbf{F}$ is the Fisher matrix and $\mathbf{Cov}$ is the covariance of shear power spectra (see just below for definitions). This formula is accurate when the biases are 'small', that is, when the biases in the cosmological parameters are much smaller than statistical errors in them, or $\delta p_i \ll (\mathbf{F}^{-1})_{ii}^{1/2}$. Here i and j label cosmological parameters, and α and β each denote a *pair* of tomographic bins, i.e. $\alpha, \beta = 1, 2, \ldots, B(B+1)/2$, where recall B is the number of tomographic redshift bins. To connect to the C_{mn} notation in equation (10), for example, we have $\beta = mB + n$.

We calculate the Fisher matrix $\mathbf{F}$ assuming perfect redshifts, and following the procedure used in many other papers (e.g. Huterer & Linder 2007). The weak lensing Fisher matrix is then given by

$$\mathbf{F}_{ij}^{\rm WL} = \sum_\ell \frac{\partial \boldsymbol{C}}{\partial p_i}\, \mathbf{Cov}^{-1}\, \frac{\partial \boldsymbol{C}}{\partial p_j}, \tag{13}$$

where p_i are the cosmological parameters and $\mathbf{Cov}^{-1}$ is the inverse of the covariance matrix between the observed power spectra whose elements are given by

$$\begin{aligned}
{\rm Cov}\left[C_{ij}(\ell), C_{kl}(\ell')\right] &= \frac{\delta_{\ell\ell'}}{(2\ell + 1)\, f_{\rm sky}\, \Delta\ell}\\
&\quad \times \left[C_{ik}(\ell)C_{jl}(\ell) + C_{il}(\ell)C_{jk}(\ell)\right].
\end{aligned} \tag{14}$$

The fiducial weak lensing survey corresponds to expectations from the DES, and assumes 5000 deg^2 (corresponding to $f_{\rm sky} \simeq 0.12$) with tomographic measurements in $B = 20$ uniformly wide redshift bins extending out to $z_{\rm max} = 1.35$. The effective source galaxy density is 12 galaxies per square arcmin, while the maximum multipole considered in the convergence power spectrum is $\ell_{\rm max} = 1500$. The radial distribution of galaxies, required to determine tomographic normalized number densities n_i in equation (4), is determined from the simulations and shown in Fig. 1.

We consider a standard set of six cosmological parameters with the following fiducial values: matter density relative to critical $\Omega_{\rm M} = 0.25$, equation of state parameter $w = -1$, physical baryon fraction $\Omega_{\rm B}h^2 = 0.023$, physical matter fraction $\Omega_{\rm M}h^2 = 0.1225$ (corresponding to the scaled Hubble constant $h = 0.7$), spectral index $n = 0.96$ and amplitude of the matter power spectrum $\ln A$, where $A = 2.3 \times 10^{-9}$ (corresponding to $\sigma_8 = 0.8$). Finally, we add the information expected from the *Planck* survey given by the *Planck* Fisher matrix (Hu, private communication). The total Fisher matrix we use is thus

$$\mathbf{F} = \mathbf{F}^{\rm WL} + \mathbf{F}^{\rm Planck}. \tag{15}$$

The fiducial constraint on the equation of state of dark energy assuming perfect knowledge of photometric redshifts is $\sigma(w) = 0.035$.

Our goal is to estimate the biases in the cosmological parameters due to imperfect knowledge of the photometric redshifts. In particular, the relevant photo-z error will be the difference between the inferred $P(z_{\rm s}|z_{\rm p})$ distribution for the calibration (or training) set and that for the actual survey. Therefore, we define

$$\delta C_\beta = C_\beta^{\rm train} - C_\beta^{\rm phot} \tag{16}$$

$$= \delta C_\beta^{\rm train} - \delta C_\beta^{\rm phot}, \tag{17}$$

where the second line trivially follows given that the true, underlying power spectra are the same for the training and photometric galaxies. All of the shear power spectra biases δC can straightforwardly be evaluated from equation (11) by using the contamination coefficients for the training and photometric fields, respectively. Therefore, the effective error in the power spectra is equal to the difference in the biases of the training set spectra (our *estimates*

of the biases in the observable quantities) and the photometric set spectra (the actual biases in the observables).

5 RESULTS

We present our results in this section. In Section 5.1 we compare the effects of sample variance on the spectroscopic redshifts and the photometric observables, concluding that the effects on the redshifts are dominant. We then discuss the impact of sample variance on photo-z training in Section 5.2, finding that the effect on the photo-z scatter statistics is negligible, but that it does introduce variability in the estimate of the overall redshift distribution. The effect is much smaller for photo-z methods that use a fitting function, such as the NNP, but pronounced for the density-based estimators such as the $p(z)_{\rm w}$. In Section 5.3, we look at the impact of sample variance in calibration of the photo-z error distributions, finding that it dominates shot noise for the scenarios we simulate. Finally, in Section 5.4 we examine the dependence of our results on our choices of parametrizations.

5.1 Spectroscopic redshift variance versus photo-z variance

LSS not only correlates the spatial distribution of galaxies, but also correlates the distribution of galaxy types, colours and other properties. For example, if there is a big galaxy cluster in some patch on the sky, red galaxies will be over-represented in that patch. Since red galaxies typically have better photo-zs than blue galaxies, an estimate of the redshift error distribution using this patch may not be representative of the error distribution of the full sample. In addition, objects in this region will have a smaller dispersion in the quality of their redshifts than predicted by Poisson statistics. Because this extra systematic is indirectly caused by the existence of LSSs, we refer to it as sample variance of the photo-zs, to differentiate it from sample variance purely in galaxy positions, which we hereafter refer to as the sample variance in the spec-zs.

We use the conditional probabilities $P(z_{\rm p}|z_{\rm s})$ and $P(z_{\rm s}|z_{\rm p})$ to disentangle the two sources of sample variance. The key point is that $P(z_{\rm s}|z_{\rm p})$ is sensitive to changes in the $z_{\rm s}$ distribution, but not in the $z_{\rm p}$ distribution. Conversely, $P(z_{\rm p}|z_{\rm s})$ is only sensitive to changes in the $z_{\rm p}$ distribution, but not in $z_{\rm s}$ (one can be convinced of this point by constructing simple toy examples).

We now estimate the variability of the error distributions across patches of the sky. For $P(z_{\rm p}|z_{\rm s})$ we define the standard deviation about the mean:

$$\sigma(P(z_{\rm p}|z_{\rm s})) = \sqrt{\frac{\sum_{\rm patches}\left(P(z_{\rm p}|z_{\rm s}) - \overline{P(z_{\rm p}|z_{\rm s})}\right)^2}{N_{\rm patches}}}, \qquad (18)$$

where $\overline{P(z_{\rm p}|z_{\rm s})}$ is the mean 'leakage' (between the patches) of galaxies from the spectroscopic bin centred at $z_{\rm s}$ being registered as having the photometric redshifts in the bin centred at $z_{\rm p}$. We also introduce the equivalently defined quantity $\sigma(P(z_{\rm s}|z_{\rm p}))$. We are interested in the increase in variability relative to the case of a random subsample, where effects of clustering due to the LSS have been zeroed out.

In the top panel of Fig. 2 we show the *ratio* of $\sigma(P(z_{\rm p}|z_{\rm s}))$ calculated for the 0.25 deg^2 LSS patches and the corresponding 0.25 deg^2 random-equivalent patches. In the bottom panel of the same figure, we show the corresponding ratio for $\sigma(P(z_{\rm s}|z_{\rm p}))$. We perform this test using the template photo-zs so as to isolate the importance of sample variance on the calibration of the error matrices. Comparing the two plots, we see that sample variance of the photo-zs does not increase appreciably between the random and the LSS patches, i.e. the ratios in each pixel are very close to unity. The sample variance of the spec-zs, on the other hand, shows marked increase, as was already apparent from Fig. 1. In Section 6.1 we show that the insensitivity of $P(z_{\rm p}|z_{\rm s})$ to LSS can be used to reduce spectroscopic follow-up requirements.

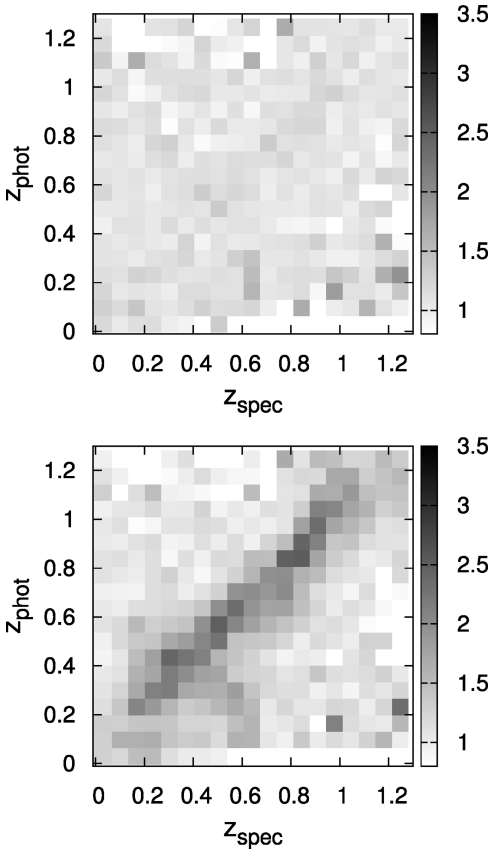

Figure 2. Top panel: ratio of $\sigma(P(z_{\rm p}|z_{\rm s}))$ (see equation 18) calculated for the 0.25 deg^2 LSS patches and the corresponding 0.25 deg^2 random patches using template photo-zs. Bottom panel: same, but for $\sigma(P(z_{\rm s}|z_{\rm p}))$. The ratios are much bigger on the bottom plot than on the top, indicating that sample variance affects the spectroscopic redshifts much more than the photometric redshifts.

5.2 Sample variance in photo-z training

In this section we examine the effects of sample variance in the training of photo-zs. We find that the commonly reported scatter in the photo-z estimation is affected by the shot noise but not by sample variance.

Table 1 shows the average photo-z scatter of the photometric sample for the polynomial method as well as the average width of the $p(z)_{\rm w}$s. The photo-z scatter is defined as the standard deviation (around zero) of the $P(z_{\rm p}–z_{\rm s})$ distribution. The average mean width of the $p(z)_{\rm w}$ is defined as the average, over all training iterations, of the mean 1σ width of the $p(z)_{\rm w}$s of the galaxies in the photometric sample. Comparison of the corresponding 'LSS' and 'Random' columns in the table shows that LSS does not affect the photo-z or $p(z)_{\rm w}$ statistics significantly. The training set size is important, however, as larger training sets have lower shot noise. For the polynomial photo-zs, we see a 12 per cent degradation in the scatter between the 6 and 0.25 deg^2 cases. The $p(z)_{\rm w}$s are much more sensitive, with a degradation of 63 per cent. In Fig. 3 one can

Table 1. 1σ scatter of the polynomial photo-zs (averaged over all training iterations) and mean 1σ width of the $p(z)_{\rm w}$s, (averaged over all training iterations). These mean scatters are shown for different patch areas and training set sizes. For comparison, the mean scatter of the template-fitting photo-zs is 0.157. Note that the LSS does not affect the photometric redshift statistics significantly, but the total number of galaxies in the training set does.

Photo-z scatter and training set size					
		LSS		Random	
Area	Mean $N_{\rm gals}$	$\sigma_{\rm poly}$	$\sigma_{p(z)}$	$\sigma_{\rm poly}$	$\sigma_{p(z)}$
6 deg^2	7.4×10^4	0.099	0.104	0.099	0.104
1 deg^2	1.2×10^4	0.106	0.129	0.105	0.129
0.25 deg^2	3.0×10^3	0.114	0.162	0.113	0.163

see that the decreased scatter of the polynomial method translates into a more diagonal $P(z_{\rm s}|z_{\rm p})$ error matrix.

This demonstrates that one can significantly decrease the variance of the recovered redshifts by fitting the redshift–observable relation (e.g. using the polynomial method) instead of using a pure density estimator (e.g. the $p(z)_{\rm w}$) – however, this comes at the cost of *biasing* the recovered redshift distribution, as seen in Fig. 4. What are the options, then, for improving the latter class of methods? To reduce the width of the $p(z)_{\rm w}$ one can either use repeat observations to decrease the mean neighbour separation in the training set, decrease the number of nearest neighbours used or adopt a fit to the redshift–observable density distribution in the neighbourhood of each galaxy. We leave these explorations for a future work.

The message of this section is that the intrinsic uncertainty of photo-zs is much greater than any systematic introduced by LSS, so that there is no significant degradation of photo-z scatter *itself* by using training sets obtained from pencil beam surveys. However, the commonly reported photo-z scatter is not sufficient to gauge biases on cosmological parameters. Below we will show that sample variance introduced by the LSS does in fact lead to significant biases in cosmological parameter estimates.

5.3 Sample variance in photo-z calibration

In this section, we describe how the sample variance in the spectroscopic parameters biases the calibration of the photo-z error distributions (i.e. the $P(z_{\rm s}|z_{\rm p})$), and how this translates into bias in cosmological parameters. The main metric we use to quantify the cosmological bias is the fractional bias in the equation of state w. We define the fractional bias as the absolute bias in w obtained from equation (12) divided by the fiducial statistical error:

$$\frac{\delta w}{\sigma(w)}, \tag{19}$$

where the marginalized statistical error in the equation of state is, recall, $\sigma(w) = 0.035$ for the DES+*Planck* combination (see Section 4).

We begin by examining a single patch in Section 5.3.1 and then discuss statistics of the biases for all the calibration patches in the simulation.

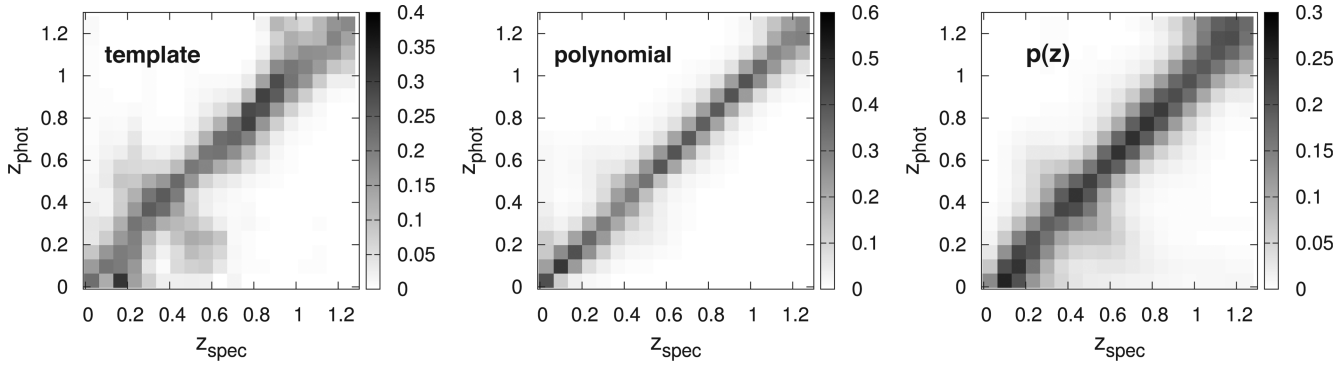

Figure 3. Mean $P(z_{\rm s}|z_{\rm p})$ for the three methods. The template is on the left, polynomial at the centre and $p(z)_{\rm w}$ on the right. For the polynomial and $p(z)_{\rm w}$, the mean $P(z_{\rm s}|z_{\rm p})$ depend on the training size. We show the 6 deg^2 result for both. Note the different scales in the three plots.

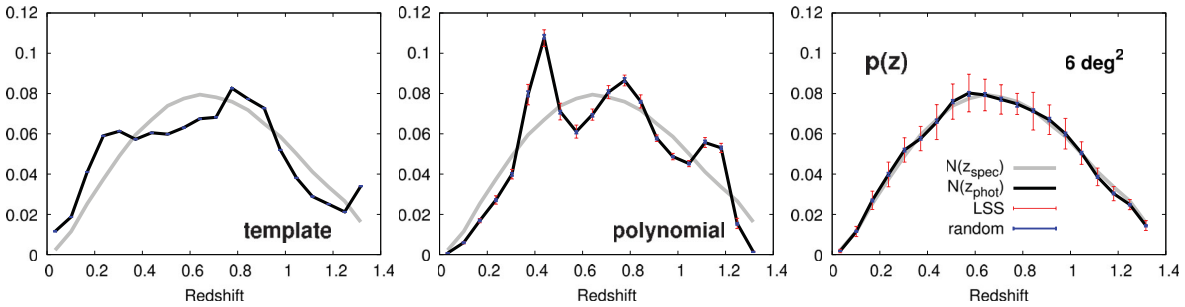

Figure 4. Redshift distribution estimates using the (left) template fitting, (centre) polynomial and (right) $p(z)_{\rm w}$ estimator. The true redshift distribution is shown in grey, and the estimates are in black. The weights estimate is not shown as it is indistinguishable from the true redshift distribution. The red (light grey) error bars shows the 1σ variability of the estimates for the 6 deg^2 patches. The hardly visible blue (dark grey) error bars show the corresponding error bars derived using the random equivalent subsamples. Note that the template fitting and polynomial methods produce very precise but highly biased estimates of the redshift distribution.

5.3.1 Case study: Patch 37

To understand how fluctuations in the redshift distribution of the calibration sample affect the estimation of $P(z_s|z_p)$ and the resulting cosmological biases, we focus on a single 1 deg^2 calibration patch, Patch 37 (out of, recall, 225 total patches). We choose this patch (which happens to be 37th in our ordering) randomly, but check that it is fairly typical, with total fractional bias well within the 1σ limits of the fractional bias distribution for the two methods we investigate.

Before we get to the details we review a result (covered in BH10) which we will utilize. Fig. 5 shows the ratio of biases in the dark energy equation of state w divided by its statistical error induced by each individual photo-z error corresponding to a fixed contamination $P(z_s|z_p)$ of 0.01 in each(z_s,z_p) bin. The points to note are that cosmological biases generally worsen with distance from the $z_p = z_s$ line, i.e. as the photo-z error becomes 'more catastrophic'. Conversely, contamination is relatively harmless at low z_p or at z_p near the survey median.

Now we are ready to examine Patch 37. The examination consists of two steps. In step 1, we look into how the differences between the overall redshift distribution and the redshift distribution of Patch 37 affect the estimation of the error distribution $P(z_s|z_p)$ for the polynomial and template methods. In step 2, we look at how the errors in the estimation of $P(z_s|z_p)$ in any given (z_s, z_p) bin propagate to biases in the dark energy equation of state w.

(i) *Step 1: Patch 37 redshift biases*. Fig. 6 shows the spectroscopic redshift distribution of the whole survey (i.e. of the photometric sample) $N(z_s)^{\rm phot}$ in black colour, as well as that of Patch 37, $N(z_s)^{\rm p37}$, in blue (grey). The deviations of the redshift distribution of Patch 37 from that of the full survey directly affect the estimation of $P(z_s|z_p)$, regardless of photo-z method. The top-row panels of Fig. 7 show the difference of $P(z_s|z_p)$ for the full sample and Patch 37 (the calibration sample) for the polynomial method (top left) and template method (top right). Comparing Fig. 6 to the top-row panels of Fig. 7, we see that each downward fluctuation of $N(z_s)^{\rm p37}$ relative to $N(z_s)^{\rm phot}$ translates into a negative $\Delta P(z_s|z_p)$ for the corresponding z_s column regardless of photo-z method used. The converse is also true: if $N(z_s)^{\rm p37}$ overestimates $N(z_s)^{\rm phot}$ at a given z_s bin, then $\Delta P(z_s|z_p)$ will be biased high in that z_s column as well.

(ii) *Step 2: Patch 37 biases in w*. The bottom-row panels of Fig. 7 show the corresponding fractional biases in the dark energy equation of state w in each (z_s,z_p) bin. For each (z_s, z_p) bin, the fractional bias in w is essentially a product between the sensitivity in fractional w bias to unit redshift errors (shown in Fig. 5) and the actual redshift bias (shown in the left-hand column panels of Fig. 7 for the two photo-z methods). Even though the sensitivities for fixed contamination are smallest near the $z_s \approx z_p$ diagonal, the actual values of $\Delta P(z_s|z_p)$ are largest near the diagonal. Overall, the latter effect wins, as the right-hand panels of Fig. 7 show, and the biases in w are contributed largely – though not exclusively – by $\Delta P(z_s|z_p)$ errors near the diagonal, $z_s \approx z_p$. A noticeable exception is the bin near $z_s = 0.4$, $z_p = 1.3$, in the polynomial results (left-hand column). Overall, the contribution of this bin lowered the overall fractional bias in w, which turns out to be $\delta w/\sigma(w) = 0.27$ for the polynomial method and 0.52 for the template method. Hence, if it was not for the big negative bias in that bin, the polynomial would have lost to the template method in this patch! The conclusion is that the final w bias is the result of several cancellations, which reduce the importance of the choice of photo-z method. However, it is desirable that photo-zs be accurate because it implies that the $P(z_s|z_p)$ will more diagonal, which, for comparably stable methods, implies smaller biases in w. And perhaps most importantly, better photo-zs imply better fiducial constraints, which our analysis is not sensitive to.

5.3.2 Statistics of the biases in w

In this section we examine statistics of the biases in w when different patches are used for training and/or calibration of the photo-zs. Fig. 8 shows the distribution of the fractional biases when using the $p(z)_{\rm w}$ and template-fitting estimators as a function of the biases obtained when the polynomial technique is used. The top panel shows the 1 deg^2 LSS case, and the bottom plot shows the 1 deg^2 random equivalent. Clearly, biases in w introduced by sample variance for the different methods are very correlated while those introduced by Poisson fluctuations alone are not. This suggests that one cannot reduce the effects of sample variance by simply combining estimates based on different photo-z methods.

In Table 2, we show the mean fractional bias in the equation of state w, its σ_{68} statistics and the median total shift in chi-squared (defined below) corresponding to the full-dimensional cosmological parameter space. We define σ_{68} as the range encompassing 68 per cent of the area of the distribution of $|\delta w|/\sigma(w)$, where δw is the bias in the equation of state in any given patch and $\sigma(w)$ is the marginalized statistical error in the equation of state. Moreover, we define the total chi-square as

$$\Delta\chi^2_{\rm tot} = (\delta \boldsymbol{p})^{\rm T} \mathbf{F} \boldsymbol{p}, \tag{20}$$

where $\delta \boldsymbol{p}$ is a six-dimensional vector containing cosmological parameter biases and $\mathbf{F}$ is the (statistical only) Fisher matrix defined in equation (15). We then define $\Delta\chi^2_{\rm med}$ to be the median of the distribution of $\Delta\chi^2_{\rm tot}$.

We find that the distribution of fractional biases are typically reasonably Gaussian, in the sense that our definition of σ_{68} matches the standard deviation of the fractional bias distribution (without the absolute value) to a few per cent, and an equivalent definition of σ_{95} is quite close to twice the standard deviation. In Section 7, we will assume the distribution of fractional biases is Gaussian to estimate follow-up requirements for the DES survey.

Actual spectroscopic calibration samples should be composed of several sets of patches of sky. Ideally, the patches should be separated enough so as to be statistically independent. Because of the small size of our simulation it is not possible to combine many independent patches; recall, our simulation covers only $\sim 15^\circ$ on a side. As a simple alternative, we combine several randomly selected patches to create the spectroscopic training and calibration sample. We consider two scenarios, one composed of patches 120 of 1/8 deg^2 with each galaxy selected with probability of 0.03125 – with average total of 2.4×10^4 galaxies. The other scenario is composed of 180 patches of 1/32 deg^2 with galaxies selected with probability 0.125, and with the average total of 3.4×10^4 galaxies. We repeat the procedure for generating these combined samples several times to generate the statistics shown in Table 3.

The point we want to make is that, in the more realistic scenarios with calibration samples coming from separate patches, all of the photo-z methods we tested yield very similar results. Combining patches randomly is far from ideal, hence the bias statistics presented in Table 3 are pessimistic. We consider the spectroscopic requirements with optimal patch selection in Section 7.

The conclusions of this section are the following.

(i) The LSS and random-equivalent cases lead to very different bias statistics. Conversely, differences between the photo-z methods

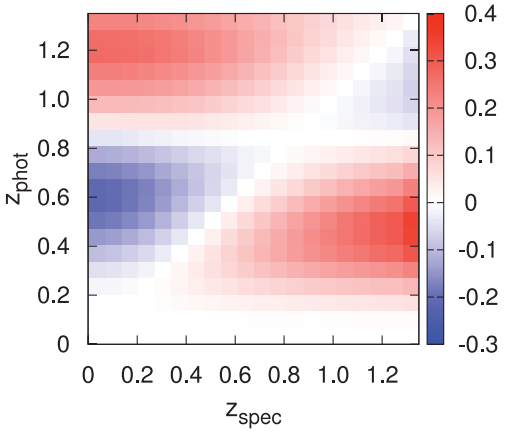

Figure 5. Bias/error ratio in the dark energy equation of state, $\delta w/\sigma(w)$, for a fixed contamination of 0.01 as a function of position in z_p–z_s space.

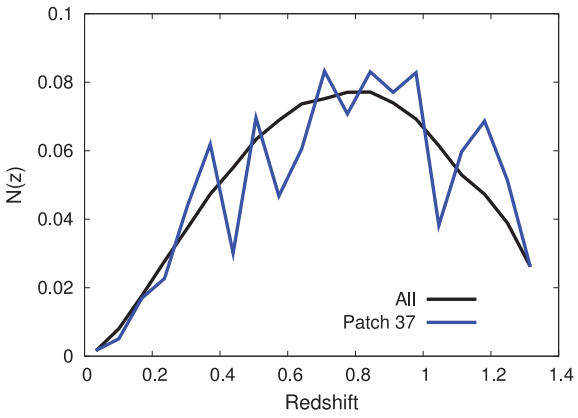

Figure 6. Spectroscopic redshift distribution of the whole survey (i.e. the photometric sample), $N(z_s)^{phot}$, in black, and of Patch 37, $N(z_s)^{p37}$, shown in blue (grey).

do not affect the bias statistics considerably. In particular, when many patches are combined, the photo-z estimators perform nearly identically.

(ii) The $p(z)_w$ method is the most sensitive to sample variance. This is expected because it is a purely density-based estimator, and it degrades the fastest as the area and size of the training set decrease. However, comparing the statistics of the $p(z)_w$ for different areas in the random equivalent cases suggests that the $p(z)_w$ estimator is not as sensitive to shot noise. Moreover, the $p(z)_w$ method is the only method that yields a perfect reconstruction of the overall redshift distribution in the limit of large area of spectroscopic samples.

(iii) The polynomial-fitting method appears to have slightly larger mean fractional bias than the $p(z)_w$ and template fitting in the cases shown in Table 2. However, the mean fractional bias is significantly smaller than the σ_{68} width in all cases. In addition,

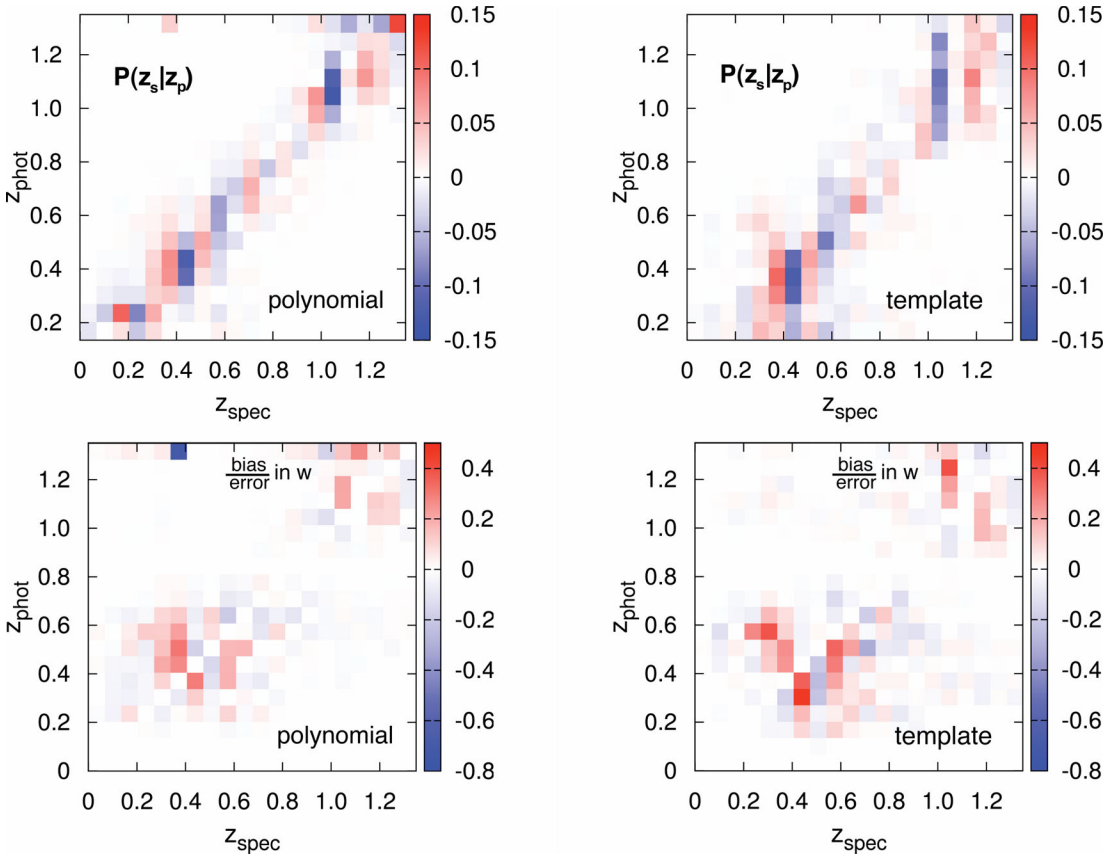

Figure 7. Biases in Patch 37. The top-row panels shows the difference of $P(z_s|z_p)$ for the photometric and calibration samples for the polynomial (top left-hand panel) and template (top right-hand panel) method. The bottom-row panels show the corresponding contribution to bias/error ratio in the dark energy equation of state w due to photometric redshift errors in each z_s, z_p bin. The fractional biases in w shown in the bottom row panels are equal to the product of the photometric redshifts errors (shown in the top row panels) and the sensitivity to a fixed photometric redshift (shown in Fig. 5).

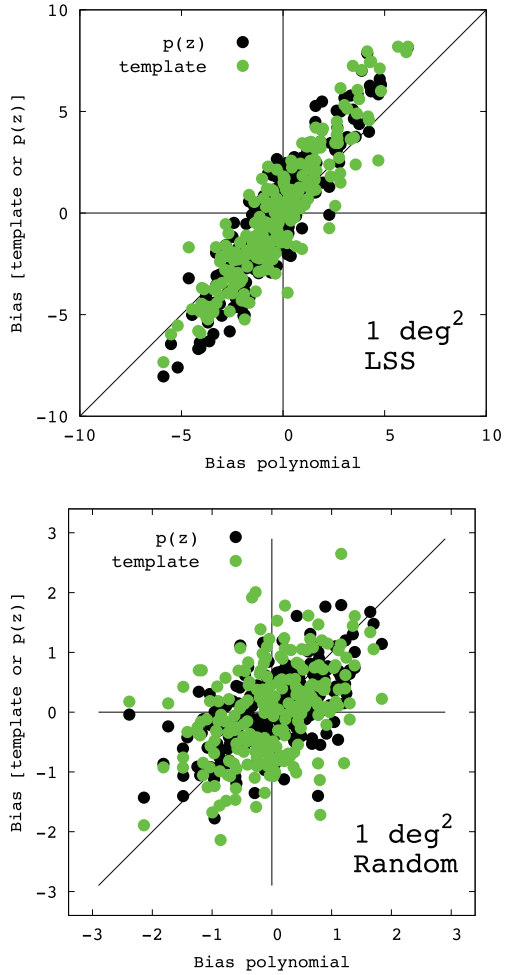

Figure 8. Fractional biases in w (i.e. the bias/error ratios in w) for the different 1 deg^2 patches used to train and/or calibrate the photometric redshifts. The top panel shows that errors in different photo-z methods produce correlated biases in the equation of state w in the presence of the LSS. The x-axis indicates the fractional bias in w for the polynomial estimator, while the y-axis shows the corresponding bias for template estimator (black points) and the $p(z)_w$ estimator, shown in green (grey). The bottom panel shows the random equivalent patches where the correlation is much less pronounced.

the polynomial technique outperforms the other methods in almost all scenarios, suggesting that use of a training set yields improvements superior to any bias introduced by using the same patch to train and calibrate the photo-zs. We believe that the conclusion that one can use the same sample to train and calibrate photo-zs should hold for other training-set-dependent photo-z techniques provided the method has some control for the degrees of freedom it utilizes and thereby avoid biases due to overfitting.

5.4 Dependence on simulations and parametrizations

In this section we discuss some of our choices of survey parameters.

5.4.1 *Dependence on intrinsic ellipticity*

For most of the results shown in this paper, we have assumed the optimistic value of $\langle\gamma_{\rm int}^2\rangle^{1/2} = 0.16$ for the rms intrinsic ellipticity. The effective intrinsic ellipticity is somewhat difficult to estimate before the survey has started taking data, and there is a range of forecasted values in the literature; for example, $\langle\gamma_{\rm int}^2\rangle^{1/2} = 0.23$ (Kirk et al. 2011; Laszlo et al. 2011). We tested using rms ellipticity of 0.26 with the template photo-zs, and found that the change affects primarily the fiducial constraints, degrading e.g. marginalized error in w by a factor of $\sim$1.6 (from 0.035 to 0.055). The overall degradation in the σ_{68} of the distribution of $|\delta w|/\sigma(w)$ degrades by a factor of $\sim$1.9 for the LSS cases and $\sim$1.6 for the random equivalent cases. Since we find that the intrinsic galaxy ellipticity primarily affects the fiducial cosmological parameter errors (i.e. $\sigma(w)$, rather than the systematic bias δw), we use it as a control parameter to vary our

Table 2. Mean fractional bias in w (i.e. mean of $\delta w/\sigma(w)$) and σ_{68} (i.e. width of the $|\delta w|/\sigma(w)$ distribution) for the different techniques, assuming patches of area 6, 1, 1/4 deg^2 for training and calibration or a random subsample with the same number of galaxies. The $\Delta\chi^2_{\rm med}$ column indicates the median value (among all patches) of $\Delta\chi^2_{\rm tot}$ of the fit over all cosmological parameters; see equation (20).

	Bias in w					
	LSS			Random		
Technique	$\overline{\delta w/\sigma(w)}$	σ_{68}	$\Delta\chi^2_{\rm med}$	$\overline{\delta w/\sigma(w)}$	σ_{68}	$\Delta\chi^2_{\rm med}$
6 deg^2						
Template	0.04	2.56	3.14	0.04	0.44	0.14
Polynomial	−0.07	1.53	2.04	−0.04	0.39	0.12
$p(z)_w$	0.05	2.33	2.56	0.07	0.31	0.10
1 deg^2						
Template	−0.04	3.75	7.36	0.01	0.92	0.75
Polynomial	−0.19	2.96	4.74	0.00	0.93	0.64
$p(z)_w$	−0.01	3.99	9.05	0.029	0.78	0.50
1/4 deg^2						
Template	0.03	4.61	16.4	−0.15	1.9	2.9
Polynomial	−0.11	3.99	10.3	−0.17	1.7	2.2
$p(z)_w$	0.07	5.88	32.3	−0.10	2.0	3.0

Table 3. Mean fractional bias in w (i.e. $\delta w/\sigma(w)$) and σ_{68} (i.e. width of the $|\delta w|/\sigma(w)$ distribution) for the different techniques, assuming 120 randomly selected patches of area 1/8 deg^{2}or 180 patches of area 1/32 deg^2 were used for training and calibration. Galaxies selected from the 1/8 deg^2 and the 1/32 deg^2 patches with probabilities 0.125 and 0.03125, respectively. The $\Delta\chi^2_{\rm med}$ column indicates the median $\Delta\chi^2_{\rm tot}$ of the fit over all cosmological parameters.

Bias in w (combined random patches)			
1/8 deg^2 – fraction = 0.03125 120 patches – $\overline{N} = 2.2 \times 10^4$			
Technique	$\overline{\delta w/\sigma(w)}$	σ_{68}	$\Delta\chi^2_{\rm med}$
Template	0.12	0.84	0.59
Polynomial	−0.16	0.76	0.54
$p(z)_w$	0.05	0.84	0.54
1/32 deg^2 – fraction = 0.125 180 patches – $\overline{N} = 3.4 \times 10^4$			
Technique	$\overline{\delta w/\sigma(w)}$	σ_{68}	$\Delta\chi^2_{\rm med}$
Template	0.19	0.76	0.41
Polynomial	−0.10	0.62	0.29
$p(z)_w$	0.12	0.74	0.39

baseline cosmological parameter error assumptions.[8] Henceforth, we adopt $\langle\gamma_{\rm int}^2\rangle^{1/2} = 0.16$ as the optimistic case for the dark energy fiducial errors (which leads to more challenging follow-up requirements), and $\langle\gamma_{\rm int}^2\rangle^{1/2} = 0.26$ as the pessimistic error case (which leads to more relaxed requirements). Unless mentioned otherwise, results assume the former, optimistic case.

5.4.2 *Dependence on redshift range*

After the completion of the paper, we obtained a newer version of the DES simulations that reached $z = 2$. We found that the redshift range 1.35–2 only composed of about 6.5 per cent of the sample and had little impact on the results despite the significantly worse photo-*z*s for galaxies in that range. Fractional biases degrade by 10 per cent, an effect driven primarily by the improvement in fiducial constraints – which assume perfect photo-*z*s.

5.4.3 *Dependence on number of tomographic bins*

We have adopted a rather aggressive redshift slicing as our baseline case, assuming 20 tomographic redshift bins distributed in the $0 < z < 1.35$ range. We expect that with fewer redshift slices, photo-*z* errors will be less pronounced while the statistical errors will increase slightly; and thus that the spectroscopic follow-up requirements derived in this paper will be somewhat relaxed. This expectation is backed up by numerical checks that we now describe.

In addition to $B = 20$, we also consider cases of $B = 5$, 10, 15, 30 and 40 tomographic bins using alternately the template and polynomial photo-*z* methods. We find that the dependence of biases in cosmological constraints on the number of bins is rather weak. As B increases from 5 to 20, the bias in the dark energy equation of state *decreases* by ∼30 per cent and converges at this point, not increasing appreciably for higher B (reflecting the fact that such small-redshift-scale fluctuations are not degenerate with cosmological information). Moreover, as B increases from 5 to 20, the *statistical* errors on w decrease by 10 per cent, and drop a further ∼10 per cent as B is increased to 40. Therefore the bias-to-error ratio decreases by a total of ∼20 per cent up to $B = 20$ but then increases by ∼10 per cent for $B = 20$–40. Given these unremarkable dependencies for such a wide range of B, and the fact that higher B implies more stringent requirements, we conclude that 20 tomographic bins is indeed a good representative choice for the calculations in this paper.

6 DISCUSSION: CAN THINGS BE IMPROVED?

In this section, we discuss possibilities for reducing the impact of sample variance. In Section 6.1, we present tests we have performed and in Section 6.2 we discuss other possibilities that should be explored.

6.1 Performed tests

(i) *Culling.* We used the width of the $p(z)_{\rm w}$ as a criterion to identify catastrophic photo-*z*s. We removed all galaxies for which $\sigma(p(z)) \geq 0.15$, which culled 10 per cent of the galaxies in our simulation. The impact of this selection is summarized in Table 4. The scatter in the photo-*z*s improved by 13 and 15 per cent for the template and polynomial methods, respectively, and the mean $p(z)_{\rm w}$ width improved 10 per cent. The width of the fractional w bias distribution, as described by σ_{68} improved by 6 and 11 per cent for the polynomial and $p(z)_{\rm w}$ techniques, respectively, but only improved the template estimator results by the negligible 3 per cent.

Table 4. Mean and σ_{68} scatter of the fractional bias in w for the different techniques, assuming patches of area 6, 1, 1/4 deg^2 for training and calibration or a random subsample with the same number of galaxies. The $\Delta\chi^2_{\rm med}$ column indicates the median $\Delta\chi^2_{\rm tot}$ of the fit over all cosmological parameters. In this table, 10 per cent of the galaxies were removed based on $p(z)_{\rm w}$ width. The $R(\sigma_z)$ shows the ratio of the photo-*z* scatters (or the $p(z)_{\rm w}$ width) of results on this table to the corresponding value in Table 2. The $R(\sigma_{68})$ shows the ratio of the σ_{68} used in this table, to the corresponding value in Table 2, assuming the same fiducial statistical constraint for both cases. As a result, this ratio compares the change in total bias, not fractional. To get the change in fractional bias one should note that the culling degrades the statistical constraints on w by 6 per cent.

	Bias in w (with culling)					
Technique	$\overline{\delta w/\sigma(w)}$	σ_{68}	$\Delta\chi^2_{\rm med}$	$R(\sigma_z)$	$R(\sigma_{68})$	$R(\Delta\chi^2_{\rm med})$
6 deg^2						
Template	0.01	2.48	3.20	0.87	0.97	1.02
Template*	−0.06	2.90	2.59	0.92	1.13	0.82
Polynomial	−0.17	1.44	1.75	0.85	0.94	0.86
$p(z)_{\rm w}$	0.03	2.08	1.94	0.90	0.89	0.75

We also tried to perform the culling using an error estimate from the template-fitting code.[9] The results are in the entry Template*, in Table 4. We see that the template error estimation was less efficient than the $p(z)_{\rm w}$ width for improving the photo-*z* scatter. With the same fraction of objects removed, the mean scatter improved by only 8 per cent compared to 13 per cent when the $p(z)_{\rm w}$ width was used. In addition, the culling actually resulted in worsening of the bias in w, despite an improvement in the overall cosmological parameter fit measured by the improvement in the median $\Delta\chi^2_{\rm tot}$.

The conclusion is that culling of outliers does not seem to be a very efficient way to improve the bias due to photo-*z* calibration even when it works reasonably well in improving the mean photo-*z* scatter.

(ii) $P(z_{\rm p}|z_{\rm s})$. If the true redshift distribution of the photometric sample is known somehow (e.g. using cross-correlation techniques (Newman 2008), or from theoretical priors), then one can use it to improve results. As discussed in Section 5.1, the quantity $P(z_{\rm p}|z_{\rm s})$ is much less sensitive to sample variance than $P(z_{\rm s}|z_{\rm p})$. If $N(z_{\rm s})$ for the photometric sample is known, we use the fact that

$$P\left(z_{\rm s}^i|z_{\rm p}^j\right) = P\left(z_{\rm p}^j|z_{\rm s}^i\right)\frac{N_{\rm s}^i}{N_{\rm p}^j} \tag{21}$$

to estimate $P(z_{\rm p}|z_{\rm s})$ from $P(z_{\rm s}|z_{\rm p})$. Table 5 shows the improvement in the statistics of the dark energy equation of state bias. For the 6 deg^2 case, we see from the last column that the statistics from template-fitting and $p(z)_{\rm w}$ methods improve by a factor of ∼5 relative to the fiducial results shown in Table 2; this corresponds to 25 times smaller follow-up samples needed to achieve the same calibration! Improvements for the 1 deg^2 are not as pronounced, but

[8] Note, it would not be hard to come up with other ways to improve the fiducial constraints, such as adding other two-point correlations to the analysis, or including magnification. Conversely, one could add intrinsic alignments and other sources of errors to degrade the constraints.

[9] The error estimate we use is the difference between the Z_BEST68_HIGH and Z_BEST68_LOW outputs of the LEPHARE code.

Table 5. Mean and σ_{68} scatter of the fractional bias in w for the different techniques, assuming patches of area 6, 1, 1/4 deg^2 for training and calibration or a random subsample with the same number of galaxies. The $\Delta\chi^2_{\rm med}$ column indicates the median $\Delta\chi^2_{\rm tot}$ of the fit over all cosmological parameters. Results in this table assume the true redshift distribution of the photometric sample was known, allowing us to use $P(z_{\rm p}|z_{\rm s})$ instead of $P(z_{\rm s}|z_{\rm p})$ as described in the text. The $R(\sigma_{68})$ shows the ratio of the σ_{68} used in this table to the corresponding value in Table 2.

Bias in w (with $P(z_{\rm p}\|z_{\rm s})$)					
Technique	Mean	σ_{68}	$\Delta\chi^2_{\rm med}$	$R(\sigma_{68})$	$R(\Delta\chi^2_{\rm med})$
6 deg^2					
Template	−0.06	0.52	0.36	0.20	0.11
Polynomial	−0.13	0.87	0.43	0.57	0.21
$p(z)_{\rm w}$	−0.14	0.52	0.34	0.22	0.13
1 deg^2					
Template	−0.17	1.28	1.39	0.34	0.19
Polynomial	−0.39	1.31	1.69	0.44	0.36
$p(z)_{\rm w}$	−0.29	0.98	1.14	0.25	0.13

are still substantial. These results are idealized, because the redshift distribution is assumed to be perfectly known. How well does $N(z_{\rm s})$ need to be known for this technique to be useful is an open question.

If one uses a $p(z)$ estimator (from any algorithm), the $p(z)$s can be corrected using the improved $P(z_{\rm p}|z_{\rm s})$. The ability to correct the redshift estimates is only possible for $p(z)$ estimators but not for single-value photo-zs.

6.2 Other possible improvements

In this section we briefly describe potentially interesting techniques to reduce the spectroscopic follow-up requirements, but that go beyond the scope of this paper.

(i) *Smoothing, fitting and deconvolution*. With enough theoretical priors, one may use assumptions about smoothness or a functional form of the overall redshift distribution to fit the weights estimate of the redshift distribution. Alternatively, since the redshift sample variance is due to the projection along the line of sight of the linear power spectrum, one can perhaps use Fourier techniques to deconvolve the LSS from the redshift distribution estimates.

(ii) *Repeat observations*. The use of repeat photometric observations would help reduce the shot-noise component of the photo-z training procedure. Unfortunately, the sample variance would not be affected. The reduction of such noise might be relevant to help stabilize deconvolution techniques.

7 GUIDE FOR OBSERVING PROPOSALS

In this section we provide a guide for observers to determine what observing requirements are needed for photo-z calibration given a specific telescope's effective angular aperture, number of spectroscopic fibres and collecting area. Typically, calibration requirements have been represented in terms of total number of galaxies. We argue that calibration requirements should be phrased in terms of variables more closely related to total observing time or cost. With this purpose in mind, we define the number of pointings, $N_{\rm point}$, to be the product of the number of patches times the number of repeat observations of each patch. For constant collecting area, the number of pointings is a direct measure of total observational time required.

The previous sections focused on calibration requirements from a single patch. If independent patches are combined, the requirements decrease with the square-root of the number of independent patches. This square-root scaling only applies exactly to the template-fitting method because it does not use a training procedure. For simplicity, and because the previous results were rather insensitive to photo-zs, we only use the template photo-zs in this section.

As an example, we consider the case of the DES. To reach reasonable spectroscopic completeness at the limiting magnitudes of the DES requires very large telescopes. We thus tune our guide to two of the telescopes that will be available for the calibration: Visible Multi-Object Spectrograph (VIMOS)-Very Large Telescope (VLT) and Inamori-Magellan Areal Camera and Spectrograph (IMACS)-Magellan. VLT is an 8-m class telescope with angular aperture of 250 arcmin^2 (or about 1/16 deg^2). Magellan is a 6.5-m class telescope with collecting area of 0.25 deg^2. We assume that in each observation, VLT and Magellan can observe about 300–500 galaxies if a low-dispersion setting is used. In real observations, the need to disperse the spectra in the focal plane reduces much of the available collecting area. This is not a random reduction, however. Roughly speaking, spectra cannot be at the edges of the focal plane so that there is room left in the focal plane to disperse the spectra. The design of VLT already accounts for this, but for Magellan there is a loss of up to a half of the total area. To roughly cover the possibilities for existing telescopes of large angular aperture we perform our tests assuming 1/4, 1/8 and 1/32 deg^2 fields of view.

Fig. 9 shows the number of independent patches that must be observed as function of the number of galaxies per patch so that the photo-z calibration leads to bias in w that is smaller than the statistical error in w with 95 per cent probability. One can see that, for fixed number of galaxies per patch, the larger the telescope, the smaller the number of independent patches that need to be observed. Hence, assuming equal throughput and same number of available fibres, a telescope such as Magellan is more efficient than VLT for spectroscopic calibration. For reference, we also show the results assuming the full 1/4 deg^2 field of view of Magellan is available for spectroscopy. For the case of the 1/4 deg^2 collecting area, if the telescope can observe 400 galaxies at once, then about 140 independent patches – or a total of 5.6×10^4 galaxies – would be needed to ensure, with 95 probability that the bias in the equation of state is less than the statistical error (i.e. bias/error $\leq$1.0). The requirement increases to about 150 and 180 patches for effective angular apertures of 1/8 and 1/32 deg^2. The requirement for VLT would be about 165 patches (not shown).

The contours in Fig. 9 were constructed by varying the mean fraction of galaxies that are sampled from each patch. The right tip of each contour line corresponds to using 100 per cent of the galaxies in a patch. For a fixed angular aperture, the total number of galaxies required decreases with decreasing sampling fraction. By sampling fewer galaxies per patch one more efficiently beats down the sample variance, up to the point where shot noise dominates. The total number of galaxies required can never be smaller than the requirements from shot noise only estimates. In our case, this is about 4×10^4 galaxies. The upturn in the contours at low sampling fraction indicates the shot-noise domination regime, at which point reducing the number of galaxies per patch yields no benefit.

How does one use Fig. 9 to deduce more stringent requirements on dark energy parameter biases, or implements different survey assumptions? The distribution of fractional bias in w is roughly Gaussian, hence to get $N\sigma$ requirements on the bias, one can

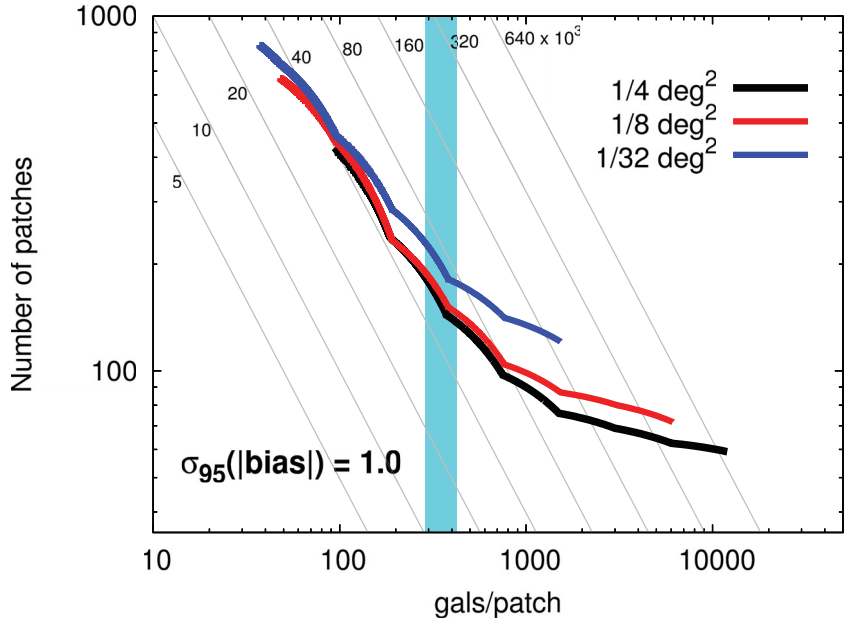

Figure 9. Relation between number of independent patches and galaxies observed per patch so that the calibration bias will yield a bias/error ratio in w that is less than 1.0 with 95 per cent probability. We consider three different telescope apertures based on capabilities of existing telescopes: 1/4 deg^2 (solid black), 1/8 deg^2 (solid red) and 1/32 deg^2 (or 112.5 arcmin^2; blue). The first two scenarios correspond to the optimistic and pessimistic assumptions about the effective observing area of Magellan. The VIMOS-VLT instrument could observe about 1/16 deg^2. The diagonal light grey lines indicate contours of fixed total number of galaxies, while the vertical band indicates typical number of galaxies per observed patch possible with a single pointing of Magellan or VLT. For a fixed number of galaxies per patch, the total number of patches required is higher for a smaller patch area in order to compensate for the increased sample variance per patch. Similarly, if the survey can observe more galaxies in each patch, then the total number of patches obviously decreases since fewer patches will be required to calibrate the shot noise, at the expense of increasing the total number of galaxies required.

simply multiply the 2σ requirement plotted by $N/2$. For example, the requirement of keeping the bias/error less than 1.0 at 2σ roughly implies that the bias is less than 0.5 at 1σ. One can use a square-root scaling to deduce more stringent requirements; for example, if one would like the bias/error in w to be less than 0.25 at 1σ, then the number of independent patches required increases by four. Because the effect of the independent number of patches is only a square-root, it is well worth investigating techniques that decrease the sensitivity to the sample variance. For example, as we saw in Section 6, if the redshift distribution of the photometric sample could be perfectly known, the calibration biases would decrease by factors of up to 5 which would decrease the number of patches required for photo-z calibration by more than a factor of 25!

Finally, recall that usage of a more realistic intrinsic galaxy ellipticity of 0.26 increases the fiducial w error by a factor of 1.6 (from 0.035 to 0.055) and leaves the biases in w largely unaffected, resulting in the decreased follow-up requirements by a factor of $\sim$3.5 in the number of patches required. Nevertheless, we think that usage of the smaller value of the intrinsic rms ellipticity used throughout is preferred, given that the fractional biases $\delta w/\sigma(w)$ could be larger than expected. This could happen in two ways: either the fiducial error $\sigma(w)$ could be improved by other weak lensing techniques (three-point function, other cross-correlations, etc.), or additional systematics might increase the bias δw. We therefore erred on the side of being conservative in terms of the spectroscopic follow-up requirements, and adopted $\langle\gamma_{\rm int}^2\rangle^{1/2} = 0.16$, or $\sigma(w) = 0.035$. Our best current understanding is that only three kinds of systematics would increase spectroscopic follow-up requirements: non-random spectroscopic failures, imperfect star–galaxy separation and variability in observing conditions. Other systematics would likely only cause a degradation in the fiducial cosmological parameter constraints, thereby decreasing follow-up requirements.

The time required for completing observations depends on the requirements on spectroscopic completeness. If we assume that a completeness level comparable to that of the VIMOS-VLT Deep Survey[10] (VVDS) is sufficient,[11] then two patches of sky can be covered per night using VLT or Magellan, if a single pointing is required per patch. In the absence of spectroscopic failures, the ideal strategy is clearly to use a single pointing per patch to beat down sample variance as fast as possible. However, spectroscopic failures typically cannot be ignored, which makes it harder to determine the optimal observing strategy. The key difficulty is that spectroscopic failure rates vary strongly with galaxy type, which implies that different observing times are needed for different types of galaxies to yield reliable redshifts. In addition, for a fixed galaxy type, there is a broad distribution of intrinsic luminosities. An optimized survey would, at the very least, requires a carefully weighted target selection function to ensure the final spectroscopic sample is a representative subsample of the full photometric survey. At best, the ideal survey would combine several telescopes, each optimized for a certain depth and galaxy population. For example, planned surveys such as BigBOSS[12] and DESpec[13] will have very wide fields of view and be able to obtain several thousand spectra per pointing.

[10] http://cesam.oamp.fr/vvdsproject/
[11] The VVDS-DEEPS survey obtained redshifts for about 44 per cent of their sample with confidence above 91–97 per cent, and of these, about 22 per cent had confidence of 99 per cent (Le Fèvre et al. 2005).
[12] http://bigboss.lbl.gov/
[13] http://eag.fnal.gov/DESpec/Home.html

An interesting strategy would be to use these telescopes – perhaps with massive co-addition of images – to obtain a large sample to depths slightly brighter than $i \simeq 24$, for galaxy types with more easily detectable spectra. This way, an 8-m class telescope could concentrate exclusively on the very faintest galaxies.

In a forthcoming follow-up to this paper, we incorporate a simulated spectroscopic pipeline to our analysis to determine the levels of spectroscopic completeness that are required for dark energy studies.

8 CONCLUSIONS

We used cosmological N-body simulations populated with galaxies with DES photometry to investigate the impact of shot noise and sample variance in the spectroscopic observations necessary to train the photo-zs and calibrate their error distributions. Our conclusions are as follows.

(i) For typical spectroscopic surveys, sample variance is much larger than shot noise.

(ii) Sample variance affects the spectroscopic properties more strongly than photometric properties. Consequently, the error distribution $P(z_s|z_p)$ is much more sensitive to sample variance than $P(z_p|z_s)$. Unfortunately, for cosmological analysis $P(z_s|z_p)$ is the error distribution that we have to use, which results in calibration requirements that are quite demanding. If the overall distribution of the photometric sample is known somehow, e.g. using cross-calibration techniques, then one can estimate $P(z_s|z_p)$ from $P(z_p|z_s)$, which can reduce follow-up requirements by more than an order of magnitude. In addition, if one uses $p(z)$s instead of single-number photo-z estimates, the improved $P(z_p|z_s)$ estimate can be used to correct and improve the $p(z)$s.

(iii) The use of the same spectroscopic sample to train photo-zs and calibrate the photo-z error distribution does not introduce additional cosmological biases. In addition, the scatter in the photo-zs is, on average, not degraded by sample variance.

(iv) For small training sets the $p(z)_w$ method is the most affected by sample variance because it is a pure density estimator (cf. Fig. 4). Conversely, the $p(z)_w$ estimate is the only unbiased method in the sense that, for large enough training, it recovers the true redshift distribution of the photometric sample.

(v) Biases in the dark energy equation of state obtained from the different photo-z methods are highly correlated for sample-variance-dominated calibration samples, suggesting that a simple combination of photo-z methods cannot reduce the biases. Conversely, for shot-noise-dominated calibration samples, biases are largely uncorrelated.

(vi) Culling of catastrophic outliers is not very effective at reducing calibration requirements, with the decrease in the bias in w being comparable to degradation of the statistical errors due to the reduction of the sample size.

(vii) We provide a guide to observing proposals of spectroscopic samples directed towards the calibration of photo-zs for the DES. We focus on Magellan and VLT, the two telescopes best suited for DES calibration. To reduce sample variance effects one should spread the observations to as many patches as possible, using as many spectroscopic fibres as possible in each observation. We find that VLT and Magellan would need about 165 and 150 patches, respectively, in order to ensure, with 95 per cent probability, that the photo-z-calibration-induced bias in w does not dominate its statistical error. This estimate assumes that 400 galaxies can be observed per patch. If a VVDS level of completeness is sufficient, these observations would require about 85 and 75 nights of observation for VLT and Magellan, respectively, assuming the optimistic fiducial uncertainty of $\sigma(w) = 0.035$. For a more pessimistic fiducial error $\sigma(w) = 0.055$, the requirements decrease by a factor of about 3.5. Nevertheless, the former number may be more useful as a guideline, since the overall requirements might be increased by including the type incompleteness and spectroscopic redshift failures, something that we will fully investigate in a forthcoming companion paper.

ACKNOWLEDGMENTS

CC would like to thank Joerg Dietrich, Huan Lin, Anja von der Linden and Jeff Newman for discussions about spectroscopic surveys and Stephanie Jouvel for help with the LePhare code. We thank Gary Bernstein, Joanne Cohn, Martin White and an anonymous referee for very useful comments. We would like to thank the Kavli Institute for Theoretical Physics in Santa Barbara where some of this work was carried out. MTB and RHW would like to thank their collaborators of the LasDamas project for use of their simulation data. CC and DH are supported by the DOE OJI grant under contract DE-FG02-95ER40899. CC is also supported by a Kavli Fellowship at Stanford University. DH is additionally supported by NSF under contract AST-0807564, and NASA under contract NNX09AC89G. RHW received support from the US Department of Energy under contract number DE-AC02-76SF00515. MTB was supported by Stanford University and the Swiss National Science Foundation under contract 2000 124835/1. This research was supported in part by the National Science Foundation under Grant No. PHY05-51164.

REFERENCES

Abdalla F. B., Amara A., Capak P., Cypriano E. S., Lahav O., Rhodes J., 2008, MNRAS, 387, 969
Abdalla F. B., Banerji M., Lahav O., Rashkov V., 2011, MNRAS, 417, 1891
Abrahamse A., Knox L., Schmidt S., Thorman P., Tyson J. A., Zhan H., 2011, ApJ, 734, 36
Amara A., Refregier A., 2007, MNRAS, 381, 1018
Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Benitez N., 2000, ApJ, 536, 571
Bernstein G., Huterer D., 2010, MNRAS, 401, 1399 (BH10)
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Blanton M. R. et al., 2003, ApJ, 592, 819
Bolzonella M., Miralles J.-M., Pelló R., 2000, A&A, 363, 476
Bordoloi R., Lilly S. J., Amara A., 2010, MNRAS, 406, 881
Budavári T., Szalay A. S., Connolly A. J., Csabai I., Dickinson M., 2000, AJ, 120, 1588
Busha M. T., Wechsler R. H., Behroozi P. S., Gerke B. F., Klypin A. A., Primack J. R., 2011, ApJ, 743, 117
Coe D., Benítez N., Sánchez S. F., Jee M., Bouwens R., Ford H., 2006, AJ, 132, 926
Coleman G. D., Wu C. C., Weedman D. W., 1980, ApJS, 43, 393
Connolly A. J., Csabai I., Szalay A. S., Koo D. C., Kron R. G., Munn J. A., 1995, AJ, 110, 2655
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Cooper M. C., Tremonti C. A., Newman J. A., Zabludoff A. I., 2008, MNRAS, 390, 245
Csabai I. et al., 2003, AJ, 125, 580
Cunha C. E., Lima M., Oyaizu H., Frieman J., Lin H., 2009, MNRAS, 396, 2379
Feldmann R. et al., 2006, MNRAS, 372, 565
Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339, 1195

Gerdes D. W., Sypniewski A. J., McKay T. A., Hao J., Weis M. R., Wechsler R. H., Busha M. T., 2010, ApJ, 715, 823
Hearin A. P., Zentner A. R., Ma Z., Huterer D., 2010, ApJ, 720, 1351
Hildebrandt H. et al., 2010, A&A, 523, A31
Hoekstra H., Jain B., 2008, Annu. Rev. Nuclear Part. Sci., 58, 99
Hogg D. W. et al., 1998, AJ, 115, 1418
Hu W., 1999, ApJ, 522, L21
Hu W., 2002, Phys. Rev. D, 66, 083515
Huterer D., 2002, Phys. Rev. D, 65, 063001
Huterer D., 2010, Gen. Relativ. Gravity, 42, 2177
Huterer D., Linder E. V., 2007, Phys. Rev. D, 75, 023519
Huterer D., Turner M. S., 2001, Phys. Rev. D, 64, 123527
Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101
Ilbert O. et al., 2006, A&A, 457, 841
Ishak M., Hirata C. M., 2005, Phys. Rev. D, 71, 023002
Jouvel S. et al., 2009, A&A, 504, 359
Kirk D., Laszlo I., Bridle S., Bean R., 2011, preprint (arXiv:1109.4536)
Kitching T. D., Taylor A. N., Heavens A. F., 2008, MNRAS, 389, 173
Knox L., Scoccimarro R., Dodelson S., 1998, Phys. Rev. Lett., 81, 2004
Koo D. C., 1999, in Weymann R., Storrie-Lombardi L., Sawicki M., Brunner R., eds, ASP Conf. Ser. Vol. 191, Photometric Redshifts and the Detection of High Redshift Galaxies. Astron. Soc. Pac., San Francisco, p. 3
Laszlo I., Bean R., Kirk D., Bridle S., 2011, preprint (arXiv:1109.4535)
Le Fèvre O. et al., 2005, A&A, 439, 845
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
Munshi D., Valageas P., Van Waerbeke L., Heavens A., 2008, Phys. Rep., 462, 67
Nakajima R., Mandelbaum R., Seljak U., Cohn J. D., Reyes R., Cool R., 2012, MNRAS, 420, 3240
Newman J. A., 2008, ApJ, 684, 88
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
Sheldon E. S., Cunha C., Mandelbaum R., Brinkmann J., Weaver B. A., 2011, preprint (arXiv:1109.5192)
Sun L., Zu-Hui F., Charling T., Jean-Paul K., Stéphanie J., André T., 2009, ApJ, 699, 958
van Waerbeke L., White M., Hoekstra H., Heymans C., 2006, Astropart. Phys., 26, 91
Wadadekar Y., 2005, PASP, 117, 79
Wang D., Zhang Y. X., Liu C., Zhao Y. H., 2007, MNRAS, 382, 1601
Wetzel A. R., White M., 2010, MNRAS, 403, 1072

APPENDIX A: THE SIMULATIONS

The simulated galaxy catalogue used for the present work was generated using the Adding Density Determined Galaxies to Lightcone Simulations (ADDGALS) algorithm (Busha et al., in preparation; Wechsler et al., in preparation). This algorithm attaches synthetic galaxies to dark matter particles in a lightcone output from a dark matter N-body simulation. The model is designed to match the luminosities, colours and clustering properties of galaxies.

The simulation used here was based on a single 'Carmen' simulation from the LasDamas project (McBride et al., in preparation). This simulation was run with the publicly available GADGET-2 code and modelled a flat Λ cold dark matter (ΛCDM) cosmology with $\Omega_{\rm m} = 0.25$ and $\sigma_8 = 0.8$ in a $1\,{\rm Gpc}\,h^{-1}$ box with 1120^3 particles. The lightcone output necessary for the ADDGALS algorithm was created by pasting together 33 snapshots in the redshift range $z = 0$–1.33. This results in a $220\,{\rm deg}^2$ lightcone whose orientation was selected such that there are no particle replications in the inner $\sim 100\,{\rm deg}^2$ and minimal replications in the outer regions.

The ADDGALS algorithm used to create the galaxy distribution consists of two steps: galaxies based on an input luminosity function are first assigned to particles in the simulated lightcone, after which multiband photometry is added to each galaxy using a training set of observed galaxies. For the first step, we begin by defining the relation $P(\delta_{\rm dm}|M_r, z)$ – the probability that a galaxy with magnitude M_r, a redshift z resides in a region with local density $\delta_{\rm dm}$, defined as the radius of a sphere containing $1.8 \times 10^{13}\,h^{-1}\,{\rm M}_\odot$ of dark matter. This relation can be tuned to reproduce the luminosity-dependent galaxy two-point function by using a much higher resolution simulation combined with the technique known as subhalo abundance matching. This is an algorithm for populating very high resolution dark matter simulations with galaxies based on halo and subhalo properties that accurately reproduces properties of the observed galaxy clustering (Conroy, Wechsler & Kravtsov 2006; Behroozi, Conroy & Wechsler 2010; Wetzel & White 2010; Busha et al. 2011). The relationship $P(\delta_{\rm dm}|M_r, z)$ can be measured directly from the resulting catalogue. Once this probability relation has been defined, galaxies are added to the simulation by integrating a (redshift dependent) r-band luminosity function to generate a list of galaxies with magnitudes and redshifts, selecting a $\delta_{\rm dm}$ for each galaxy by drawing from the $P(\delta_{\rm dm}|M_r, z)$ distribution, and attaching it to a simulated dark matter particle with the appropriate $\delta_{\rm dm}$ and redshift. The advantage of ADDGALS over other commonly used approaches based on the dark matter haloes is the ability to produce significantly deeper catalogues using simulations of only modest size. When applied to the present simulation, we populate galaxies as dim as $M_r \approx -16$, compared with the $M_r \approx -21$ completeness limit for a standard halo occupation (HOD) approach.

While the above algorithm accurately reproduces the distribution of satellite galaxies, central objects require explicit information about the mass of their host haloes. Thus, for haloes larger than $5 \times 10^{12}\,h^{-1}\,{\rm M}_\odot$, we assign central galaxies using the explicit mass–luminosity relation determined from our calibration catalogue. We also measure $\delta_{\rm dm}$ for each haloes, which is used to draw a galaxy from the integrated luminosity function with the appropriate magnitude and density to place at the centre.

For the galaxy assignment algorithm, we choose a luminosity function that is similar to the SDSS luminosity function as measured in Blanton et al. (2003), but evolves in such a way as to reproduce the higher redshift observations of the NOAO Deep Wide-Field Survey (NDWFS) and Deep Extragalactic Evolutionary Probe 2 (DEEP2) observations. We use a Schechter function with $\phi^* = 1/81 \times 10^{-2z/3}$, $M_* = -20.34 + 3.5(a - 0.91)$ and $\alpha = -1.03$, where a is the cosmological expansion factor.

Once the galaxy positions have been assigned, photometric properties are added. We begin with a training set of spectroscopic galaxies and the simulated set of galaxies with r-band magnitudes generated earlier. For each galaxy in both the training set and simulation we measure Δ_5, the distance to the fifth nearest galaxy on the sky in a redshift bin. Each simulated galaxy is then assigned an SED based on drawing a random training-set galaxy with the appropriate magnitude and local density, k-correcting to the appropriate redshift and projecting on to the desired filters. The k-corrections and projections are performed using the KCORRECT code (Blanton et al. 2003). The construction of the SEDs in KCORRECT is described in Blanton & Roweis (2007).

Differences between the training set and simulated galaxy sample complicate the process of colour assignment. In order to compile a sufficiently large training set, we use a magnitude-limited sample of SDSS spectroscopic galaxies brighter than $m_r = 17.77$ with $z < 0.2$. The simulated sample, on the other hand, is a volume-limited

sample, spanning a broader redshift range. When measuring Δ_5 we restrict ourselves to neighbours brighter than $M_r = -19.7$ in the simulation sample, while using all objects in the observational catalogue. To mitigate differences in luminosity and redshift, each galaxy is rank ordered according to its density in its redshift bin, and require that objects be in the same percentile bin in each sample rather than having the same the absolute value of Δ_5. This is similar to the method used in Cooper et al. (2008).

The final step for producing a realistic simulated catalogue is the application of photometric errors. While the photometric errors generated here are particular to DES, the algorithm can be generalized for any survey. For each galaxy, we add a noise term to the intrinsic galaxy flux, where the noise is drawn from a Gaussian of width:

$$\text{noise} = \sqrt{t_e n_p n_s + f_{g,i} t_e}, \tag{A1}$$

where t_e is the exposure time, n_p the number of pixels covered by a galaxy, n_s the flux of the sky in a single detector pixel and $f_{g,i}$ is the intrinsic flux of the galaxy. Application of the above relation to objects from the SDSS catalogue shows that it is able to faithfully reproduce the reported errors of the survey.

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