Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **452,** 3710–3725 (2015) doi:10.1093/mnras/stv1567

Accurate photometric redshift probability density estimation – method comparison and application

Markus Michael Rau,[1,2]★ Stella Seitz,[1,2] Fabrice Brimioulle,[3] Eibe Frank,[4] Oliver Friedrich,[1,2] Daniel Gruen[1,2] and Ben Hoyle[1,5]

[1]*Ludwig-Maximilians-Universität München, Universitäts-Sternwarte, Scheinerstr. 1, D-81679 Munich, Germany*
[2]*Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse 1, D-85748 Garching, Germany*
[3]*Observatório Nacional, Rua General José Cristino, 20921-400 Rio de Janeiro, Brazil*
[4]*Department of Computer Science, University of Waikato, Private Bag 3105, Hamilton, New Zealand*
[5]*Excellence Cluster Universe, Bolzmannstr. 2, D-85748 Garching, Germany*

Accepted 2015 July 10. Received 2015 July 8; in original form 2015 April 6

ABSTRACT

We introduce an ordinal classification algorithm for photometric redshift estimation, which significantly improves the reconstruction of photometric redshift probability density functions (PDFs) for individual galaxies and galaxy samples. As a use case we apply our method to CFHTLS galaxies. The ordinal classification algorithm treats distinct redshift bins as ordered values, which improves the quality of photometric redshift PDFs, compared with non-ordinal classification architectures. We also propose a new single value point estimate of the galaxy redshift, which can be used to estimate the full redshift PDF of a galaxy sample. This method is competitive in terms of accuracy with contemporary algorithms, which stack the full redshift PDFs of all galaxies in the sample, but requires orders of magnitude less storage space. The methods described in this paper greatly improve the log-likelihood of individual object redshift PDFs, when compared with a popular neural network code (ANNZ). In our use case, this improvement reaches 50 per cent for high-redshift objects ($z \geq 0.75$). We show that using these more accurate photometric redshift PDFs will lead to a reduction in the systematic biases by up to a factor of 4, when compared with less accurate PDFs obtained from commonly used methods. The cosmological analyses we examine and find improvement upon are the following: gravitational lensing cluster mass estimates, modelling of angular correlation functions and modelling of cosmic shear correlation functions.

Key words: catalogues – surveys – galaxies: distances and redshifts.

1 INTRODUCTION

The determination of distance, or redshift, estimates to galaxies is a vital requirement before using large-scale photometric galaxy surveys for many cosmological analyses. Large-scale surveys, such as the SDSS (York et al. 2000), PanSTARRS (Tonry et al. 2012), DES (Flaugher 2005) and LSST (Tyson et al. 2003), rely on a combination of photometric and more accurate spectroscopic redshifts when providing distance estimates to photometrically identified galaxies.

Photometric redshifts are used throughout astrophysics and cosmology, for example in large-scale structure analyses (Staniszewski et al. 2009; de Simoni et al. 2013), in galaxy cluster weak lensing analyses (Gruen et al. 2013) and in galaxy–galaxy lensing analyses (Brimioulle et al. 2013). Photometric redshifts are obtained using either machine learning methods or template fitting techniques (see e.g. Benítez 2000; Csabai et al. 2000; Bender et al. 2001; Feldmann et al. 2006; Ilbert et al. 2006; Greisel et al. 2013). Machine learning techniques range from early works employing artificial neural networks (Firth, Lahav & Somerville 2003; Collister & Lahav 2004) as photometric point predictors to recent developments that estimate the full photometric redshift probability density function (PDF) of the galaxy (Lima et al. 2008; Cunha et al. 2009; Carrasco Kind & Brunner 2013; Bonnett 2015). For detailed reviews and comparisons of different photometric redshift techniques, we refer the reader to Sánchez et al. (2014), Hildebrandt et al. (2010) and Dahlen et al. (2013). This work focuses on machine learning methods for photometric redshift PDF estimation for samples of galaxies (hereafter sample PDF) as well as individual galaxies (hereafter individual PDFs). We apply the results to a range of analyses in weak gravitational lensing, cosmic shear and large-scale structure.

★ E-mail: mmrau@usm.lmu.de

In general, machine learning algorithms learn a mapping between the photometry of an object and the spectroscopic redshift. To train the machine learning models to learn this mapping, one typically identifies spectrophotometric data that overlap with the photometric feature space of the final data sample for which one would like to estimate redshifts. However, recent work shows that machine learning can also be performed with spectroscopic reference data that are brighter than the photometric sample (Hoyle et al. 2015b). Many photometric surveys include a dedicated spectroscopic follow-up programme, which allows such a machine learning system to be built, e.g. SDSS-I/II (York et al. 2000), 2dF (Colless et al. 2001), VVDS (Le Fèvre et al. 2005) and WiggleZ (Drinkwater et al. 2010).

The mapping obtained with machine learning is only approximate: the redshift of an object cannot be exactly determined by its corresponding photometry. Moreover, most machine learning methods produce a point estimate, which reduces the individual PDF to one number. The point estimate only predicts the most likely value of the redshift, irrespective of the quality of the photometry, and the shape of the distribution. In order to enter the era of precision cosmology, one must be able to incorporate the uncertainty in the redshift estimate into the cosmological analysis. This means that the use of single point redshift predictions is no longer sufficient. To achieve precision cosmology, we are required to incorporate the full redshift uncertainty using the individual PDFs.

We can obtain a sample PDF by stacking the individual PDFs. This distribution describes the probability that a randomly sampled galaxy has a certain redshift. The accurate estimation of the redshift distribution of the full sample is important for many cosmological analyses, e.g., in large-scale structure, weak gravitational lensing and cosmic shear.

However, effectively estimating and storing the photometric redshift PDF instead of the point estimate, for each object in a large astronomical data set, is a challenging task. This process requires efficient and accurate photometric estimation algorithms, and scalable data storage solutions. These algorithms must be benchmarked using carefully constructed performance metrics to be useful for the next-generation large-scale structure photometric surveys (e.g. Laureijs et al. 2011).

We discuss such metrics to quantify performance of photometric redshift PDF estimation in Section 2. We describe the ordinal class PDF (OCP) algorithm in Section 3.2, which improves the estimation accuracy over commonly used non-ordinal classification architectures. We continue in Section 3.4 by showing how the OCP method can become more storage efficient, by combining it with the Gaussian mixture model. This enables the storage of the PDFs of individual galaxies even within massive data sets without significant demands on disc space.

Many applications in cosmology require an estimation of the sample PDF. We propose a single point estimator for this quantity in Section 3.5, and show how this single floating point number can be computed very efficiently, and achieves good performance when compared with algorithms that stack individual PDFs. The performance of the proposed techniques is demonstrated and analysed in a method comparison in Sections 5.1 and 5.2 using a spectrophotometric data set (Section 4) obtained from the public CFHTLS Wide survey.

Finally, we demonstrate in Section 5.3 that the methods introduced in this work improve the precision of gravitational lensing cluster mass estimates, measurements of angular correlation functions and analyses of cosmic shear correlation functions, when compared with results obtained using a common neural network code. We conclude and summarize in Section 6.

2 FUNDAMENTAL CONCEPTS

The following section gives a brief review of important statistical concepts needed in this work. We start with a short introduction to density estimation, introduce metrics to quantify the performance of density estimators and finally describe a scheme to assess the performance of a machine learning model.

2.1 Kernel density estimation

The goal of kernel density estimation is to find a good estimator[1] $\hat{p}(\boldsymbol{x})$ for the PDF $p(\boldsymbol{x})$ of a random variable $\boldsymbol{X}$ using N samples $\boldsymbol{x}_i$. Consider a small region $\mathcal{R}$ centred on a point $\boldsymbol{x}$. We can then assume that $p(\boldsymbol{x})$ is approximately constant across $\mathcal{R}$. Based on this assumption, we can estimate the density at point $\boldsymbol{x}$ as

$$\hat{p}(\boldsymbol{x}) = \frac{k}{N V_{\mathcal{R}}} \,. \tag{1}$$

The number of objects[2] k in equation (1) can be estimated by considering a D dimensional hyper cube with volume

$$V_{\mathcal{R}} = h^D \tag{2}$$

centred on the point $\boldsymbol{x}$ with side length h. Using equation (1), we obtain k as

$$k = \sum_{i=1}^{N} K\left(\frac{\boldsymbol{x} - \boldsymbol{x}_i}{h}\right) , \tag{3}$$

where

$$K(\boldsymbol{d}) = \begin{cases} 1, & |d_i| \leq 1/2, \quad 1 \leq i \leq D \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

is an example of a kernel function. Note that this kernel has discontinuities at the boundaries. The bandwidth h determines how much the kernel density estimate interpolates (or smoothes) between the given data points. A bandwidth that is too large oversmoothes important structures in the density whereas one that is too small leads to a noisy density estimate. The density estimate $\hat{p}(\boldsymbol{x})$ can then be written as

$$\hat{p}(\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{h^D} K\left(\frac{\boldsymbol{x} - \boldsymbol{x}_i}{h}\right) =: \frac{1}{N}\sum_{i=1}^{N} \tilde{K}\left(\boldsymbol{x}, \boldsymbol{x}_i, h\right) . \tag{5}$$

Instead of using equation (4), which has discontinuities at the boundaries, we can alternatively use smooth and symmetric functions, for example, a Gaussian.

The estimation of photometric redshift PDFs for individual objects (individual PDFs) is an application of conditional PDF estimation, since the individual PDF $p(z|\boldsymbol{f})$ is conditional on the object's photometry $\boldsymbol{f}$. The estimation of conditional PDFs can be formulated in close analogy with equation (5). We can estimate the individual PDF $p(z|\boldsymbol{f})$ as a weighted kernel density estimate in redshift space of the form

$$\hat{p}(z|\boldsymbol{f}) = \sum_{i=1}^{N_{\mathrm{tr}}} w_i(\boldsymbol{f}) K(z, z_i^{\mathrm{spec}}, h) \,, \tag{6}$$

using a data set, the so-called training set, containing N_{tr} objects. $K(z, z_i^{\mathrm{spec}}, h)$ denotes a kernel function with bandwidth h centred

[1] In the following, we will mark the estimator for a quantity with a hat.
[2] Fixing the number of points k that fall into $\mathcal{R}$ and estimating the volume $V_{\mathcal{R}}$ leads to the k nearest-neighbour density estimation technique (see e.g. Scott 1992).

on the spectroscopic redshift values $z_i^{\rm spec}$. The weights $w_i(\boldsymbol{f})$ sum to unity and depend on the photometry $\boldsymbol{f}$ of the object.

The conditional cumulative distribution function $F(z|\boldsymbol{f})$ defined as

$$F(z|\boldsymbol{f}) = \int_{-\infty}^{z} p(z'|\boldsymbol{f}) \mathrm{d}z' \tag{7}$$

can be estimated (Meinshausen 2006) as

$$\hat{F}(z|\boldsymbol{f}) = \sum_{i=1}^{N_{\rm tr}} w_i(\boldsymbol{f}) I(z_i^{\rm spec} \leq z)\,. \tag{8}$$

$I(z_i^{\rm spec} \leq z)$ equates to unity if $z_i^{\rm spec} \leq z$ and to zero otherwise.

The redshift distribution $\hat{p}(z)$ of a sample (sample PDF) containing N objects can be estimated by stacking the individual PDFs

$$\hat{p}(z) = \sum_{i=1}^{N} w_{{\rm stack},i}\, \hat{p}(z|\boldsymbol{f}_i)\,. \tag{9}$$

The normalized weights $w_{{\rm stack},\,i}$ can be set to $1/N$ or chosen to give more weight to certain subpopulations. For example, we can favour certain redshift intervals $z \in [a, b]$ by defining weights as

$$w_{\rm stack} = \int_a^b p(z|\boldsymbol{f}) \mathrm{d}z = \hat{F}(b|\boldsymbol{f}) - \hat{F}(a|\boldsymbol{f})\,, \tag{10}$$

and we show an example of such a weighting in Section 5.2. The above weights are normalized afterwards to sum to unity.

2.2 The Gaussian mixture model

In this paper, we consider kernel density estimators and Gaussian mixture models for density estimation. A Gaussian mixture model (see, for example, Bishop 2006) for the PDF $p(x)$ of a random variable X is a linear combination of K normal densities defined as

$$p(x) = \sum_{i=1}^{K} \alpha_i \mathcal{N}(x, \mu_i, \sigma_i)\,, \tag{11}$$

where α_i is the amplitude, μ_i is the mean and σ_i is the standard deviation of the mixture component i.

We define the weight proportion $\gamma_k(x)$ of component k as

$$\gamma_k(x) = \frac{\alpha_k \mathcal{N}(x, \mu_k, \sigma_k)}{\sum_{j=1}^{K} \alpha_j \mathcal{N}\left(x, \mu_j, \sigma_j\right)}\,, \tag{12}$$

where $\gamma_k(x)$ determines how much a certain component of the Gaussian mixture model contributes to the total density at point x.

2.3 Evaluation metrics

Consider an estimate $\hat{p}(\boldsymbol{x})$ of the true PDF $p(\boldsymbol{x})$ describing the distribution of the random variable $\boldsymbol{X}$. We can measure the quality of the estimate $\hat{p}(\boldsymbol{x})$ by its distance $D(\hat{p}(\boldsymbol{x})||p(\boldsymbol{x}))$ to the true distribution $p(\boldsymbol{x})$, which is generally unknown. The Kullback–Leibler divergence between the true density $p(\boldsymbol{x})$ and the estimate $\hat{p}(\boldsymbol{x})$ is defined using the natural logarithm as

$$D(p||\hat{p}) = \int_{-\infty}^{\infty} p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{\hat{p}(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x}\,. \tag{13}$$

A good estimate $\hat{p}$ for p should minimize $D(p||\hat{p})$. Rewriting the logarithm we obtain

$$D(p||\hat{p}) = \int_{-\infty}^{\infty} p(\boldsymbol{x}) \log\left(p(\boldsymbol{x})\right) \mathrm{d}\boldsymbol{x} - \int_{-\infty}^{\infty} p(\boldsymbol{x}) \log\left(\hat{p}(\boldsymbol{x})\right) \mathrm{d}\boldsymbol{x}, \tag{14}$$

and we note that the first term is a constant that does not depend on the model parameters, for example bandwidth, kernel, or shape of kernel function. Thus, the second term in equation (14) can be used as a relative measure of the accuracy of $\hat{p}(\boldsymbol{x})$. If we use the sample mean to estimate the expectation with respect to $p(\boldsymbol{x})$, we obtain the mean negative log-likelihood loss, hereafter MNLL (Habbema, Hermans & Van den Broek 1974; Duin 1976),

$$\mathrm{MNLL} = -\frac{1}{N} \sum_{i=1}^{N} \log\left(\hat{p}(\boldsymbol{x}_i) + \epsilon\right), \tag{15}$$

where we set $\epsilon = 10^{-6}$ to avoid floating point underflow. The Kullback–Leibler divergence is a distance and thus non-negative and it is smallest if the MNLL is smallest.

A suitable loss function for individual PDFs can be defined analogously (see e.g. Frank & Bouckaert 2009; Takeuchi, Nomura & Kanamori 2009; Sugiyama et al. 2010). We estimate $p(z|\boldsymbol{f}_i)$ for each of the N objects in the sample, in order to establish performance using a sample of objects for which spectroscopic redshift values have been observed. We then evaluate $\hat{p}(z|\boldsymbol{f}_i)$ at the object's observed spectroscopic redshift $\hat{p}(z = z_{{\rm spec},i}|\boldsymbol{f}_i)$. In the rest of the paper, the abbreviation MNLL refers to the mean negative log-likelihood loss evaluated for individual PDFs.

2.4 Model training

We randomly sample three non-overlapping data sets without replacement from the available data: the training set, the validation set and the test set. The model is trained on the training set and the model parameters (Table 1) are chosen by testing the performance of the trained model with different parameter settings on the validation set.

Table 1. Model parameters of the QRF, the classification-based PDF estimation algorithms (OCP/NOCP) and the OCP algorithm used with the Gaussian mixture model OCP GMM. 'nodesize' and 'mtry' are model parameters of the random forest described in Section 3. 'BW mod' is the bandwidth modification factor employed in the Scott's rule (equation 24) and Gauss Comp. denotes the maximum number of components allowed in the Gaussian mixture model. The best parameter configuration for each algorithm picked on the validation set during model tuning (Section 2.4) is marked in bold type.

	QRF/HWE	OCP	NOCP	OCP GMM
nodesize	3,**5**,7,10	1,2,3,**5**,7,9	**1**,2,3,5	**1**,2,3,5,7,9
mtry	1,2,3,4,**5**	1,2,3,**4**,5	1,2,3,4,**5**	1,2,3,4,**5**
BW mod	0.5,0.6,...,**1.8**,...,3.0	0.5,0.6,...,**2.5**,...,3.0	0.5,0.6,...,**2.0**,...,3.0	–
Gauss Comp.	–	–	–	**1**,2,3

The validation set is used during model tuning and therefore it does not provide a good estimate of the performance on unseen data. We measure this generalization performance on a test set that is not used during training and tuning.

To evaluate the machine learning algorithms, we construct a training set containing 9000 objects, a validation set containing 3000 objects and a test set containing 22 072 objects. After the validation set has been used to determine the best combination of model parameters, we merge the training set and the validation set, and train the respective model again with this best setup. In this way, we make optimal use of the available data to build the final model. All results described in Section 5 are obtained on the test set, which, we reiterate, was not used in all prior steps of model training and tuning.

In this work, we choose to use the aperture magnitudes of the CFHTLS Wide five-band photometry as input attributes. Other photometric features may be used, for example see Hoyle et al. (2015a) for a feature importance analysis.

3 ALGORITHMS

We have introduced the estimator for the photometric redshift PDF of individual objects (individual PDF) in equation (6) as a weighted kernel density estimate that depends on the weights $w(\boldsymbol{f})$. The following section discusses two algorithms that can be used to estimate these weights.

3.1 Quantile regression forest (QRF)

The quantile regression forest[3] (Meinshausen 2006) is a generalization of the random forest (Breiman 2001) that can be used to reconstruct individual PDFs, an algorithm known as TPZreg (Carrasco Kind & Brunner 2013) in astrophysics.

A regression/classification tree partitions the input space and returns the mean/majority vote of the response values (i.e. the redshift values) of the training set objects in each partition as the final prediction for new objects falling into that partition. The tree partitions the input data such that the training set objects in each partition are most similar with respect to their response values. In regression, we measure similarity using the sum of squares loss function SSE, defined as

$$\mathrm{SSE} = \sum_{\tau=1}^{l} \sum_{\boldsymbol{f}_i \in \mathcal{R}_\tau} \left(z_{\mathrm{spec},i} - \langle z_{\mathrm{spec},\tau} \rangle\right)^2 . \tag{16}$$

The sum runs over all l leaf nodes of the tree $1 \leq \tau \leq l$, each of which represents a certain partition $\mathcal{R}_\tau$ in input space, and over all objects in the training set $(\boldsymbol{f}_i, z_{\mathrm{spec},i})$ with attribute values $\boldsymbol{f}_i$ that fall into $\mathcal{R}_\tau$. The term $\langle z_{\mathrm{spec},\tau} \rangle$ denotes the mean spectroscopic redshift of all training set objects that fall into $\mathcal{R}_\tau$.

The binary tree is recursively grown by choosing a splitting attribute and split point for each region using brute-force search such that the SSE is minimized.

The random forest algorithm combines several trees by bootstrap aggregation which is described as follows. New training sets are drawn from the original training set with replacement, which is also known as bootstrapping. We train a tree model on each of these bootstrapped training sets, to obtain an ensemble of trees. Combining the estimates from all trees in the ensemble reduces variance. In addition, the random forest algorithm makes the resulting models even more diverse by modifying the way each tree is grown. Before each split selection, the routine randomly selects a certain number of attributes, as specified by the 'mtry' parameter, on which the algorithm can perform the split.

The complexity of the tree model is governed by the size of the leaves of the tree. We stop the recursive tree building process when a specified minimum number of objects in each leaf, denoted as 'nodesize', is reached. If the nodesize is small, very complex trees are grown and the tree might overadapt to the training set. This is an example of overfitting. The prediction from the random forest is the mean, in regression, or the majority vote, in classification, of the predictions from the ensemble of trees.

A single tree in the random forest splits the space spanned by the input attributes derived from the photometry of the objects into partitions which are represented by the tree leaves. Each leaf defined in this manner is associated with the mean spectroscopic redshift value of the training set objects in this leaf. The tree therefore approximates the underlying smooth function by a step function. If a new object is queried, it will be placed in a leaf containing objects with similar photometry. Following the formulation by Meinshausen (2006), we can write the photometric redshift prediction

$$z_{\mathrm{phot}}(\boldsymbol{f}) = \sum_{i=1}^{N_{\mathrm{tr}}} w_i(\boldsymbol{f}) z_{\mathrm{spec},i} \tag{17}$$

as a weighted sum over the spectroscopic redshift values $z_{\mathrm{spec},i}$ of the N_{tr} training set objects. In order to distinguish the different trees in the ensemble, which are characterized by different split selections, we introduce a parameter θ, which characterizes each tree. All training set objects with photometry $\boldsymbol{f}_i^{\mathrm{tr}}$ that are located in the same region $\mathcal{R}_{l(\boldsymbol{f},\theta)}$ (defined by the leaf $l(\boldsymbol{f},\theta)$) as the newly queried object with photometry $\boldsymbol{f}$ get a constant weight, and all other training set objects get zero weight. This can be written as

$$w_i(\boldsymbol{f},\theta) = \frac{I\left(\boldsymbol{f}_i^{\mathrm{tr}} \in \mathcal{R}_{l(\boldsymbol{f},\theta)}\right)}{\sum_{j=1}^{N_{\mathrm{tr}}} I\left(\boldsymbol{f}_j^{\mathrm{tr}} \in \mathcal{R}_{l(\boldsymbol{f},\theta)}\right)} , \tag{18}$$

where the weights are normalized such that they sum to unity.

The same concept holds for the random forest prediction, in which the weights associated with each training set object are averaged over k trees, each grown on different bootstrapped data sets, and therefore each described by a different parameter θ_b:

$$w_i(\boldsymbol{f}) = \frac{1}{k} \sum_{b=1}^{k} w_i(\boldsymbol{f},\theta_b) . \tag{19}$$

The weights can be used to estimate the individual PDF and corresponding statistics like the conditional mean, the conditional cumulative distribution function, or the conditional standard deviation defined as

$$\hat{\sigma}^2(z|\boldsymbol{f}) = \sum_{i=1}^{N_{\mathrm{tr}}} w(\boldsymbol{f}_i) \left(z_{\mathrm{spec},i} - z_{\mathrm{phot}}(\boldsymbol{f}_i)\right)^2 . \tag{20}$$

The following section introduces an alternative way of estimating the weights in equation (6), using a classification scheme.

3.2 OCP estimation

The basic idea of classification-based PDF estimation is to bin the spectroscopic data by redshift and use a classification algorithm that outputs probabilities for bin membership to reconstruct the PDF. Bin membership is viewed as an ordinal variable. Ordinal scale

[3] The method was originally developed to estimate conditional quantiles, and hence the name quantile regression forest.

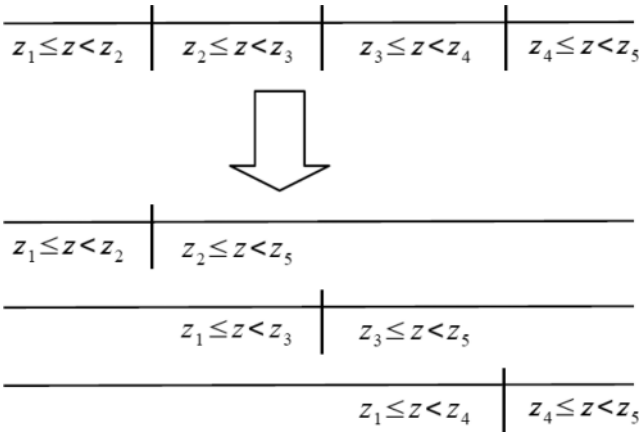

Figure 1. An illustrative example of a nominal classification problem with four redshift bins. These bins can be reformulated into three binary classification problems by merging neighbouring bins. The class probabilities from the binary classification problems can be recombined to incorporate the ordering between the redshift bins (see the text) into the final classification.

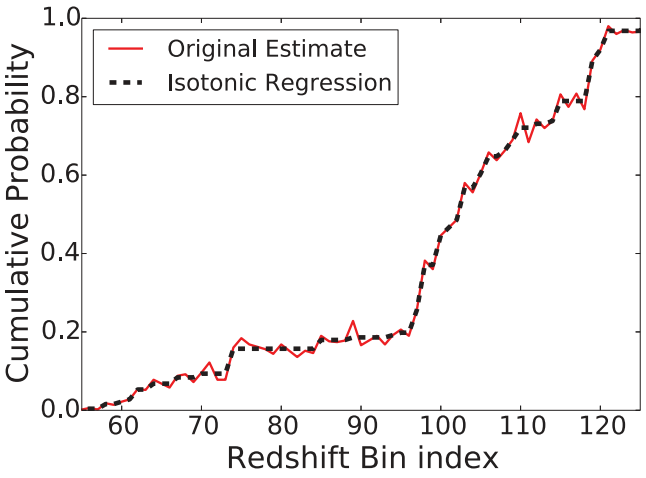

Figure 2. Ordinal classification can result in non-monotonic cumulative distribution functions. We calibrate them using isotonic regression. Isotonic regression (black) approximates the original estimate (red) as a monotonically increasing step function.

variables, in contrast to nominal ones, exhibit an intrinsic order. If the classes in a classification problem are ordinal, we can use this information to improve the classification (Frank & Hall 2001).

Current classification-based PDF estimation methods in the astrophysics literature (e.g. Carrasco Kind & Brunner 2013; Bonnett 2015) treat redshift bins as nominal classes. In the following, we will refer to the latter as the non-ordinal class PDF (NOCP) algorithm.

The OCP algorithm trains a separate classifier that estimates the probability $p(z \geq z_i)$ that a new object has redshift z above a certain threshold z_i given by the edge of the respective redshift bin. This scheme is illustrated in Fig. 1. The probability that the redshift of an object resides in the original bins is then calculated from these separate classification models as (Frank & Hall 2001)

(1) $p(z \in [z_1, z_2[) = 1 - p(z \geq z_2)$
(2) $p(z \in [z_{i-1}, z_i[) = p(z \geq z_{i-1}) - p(z \geq z_i)$
(2) $p(z \in [z_{k-1}, z_k[) = p(z \geq z_{k-1})\,, 1 < i < k\,.$

The reconstruction of the class probabilities $p(z_i)$ has the idealistic assumption that each of the classifiers used to estimate the probability $p(z \geq z_i)$ outputs perfect probabilities. In practice, this will not be the case and the recovered cumulative distribution function, which is a monotonically increasing function, has to be calibrated. Schapire et al. (2002) and Frank & Bouckaert (2009) use a heuristic approach to ensure this monotonicity requirement. Alternatively, we use the 'isotonic' regression technique to calibrate the class probabilities. Isotonic regression is synonymous for monotonically increasing regression and is a technique for which efficient implementations are available (de Leeuw, Hornik & Mair 2009).

For increasing bin index, isotonic regression optimizes the mean squared error between the original function values and the isotonic fit such that the fit is a monotonic increasing step function as shown in Fig. 2.

We use bins of fixed size $\Delta z = 0.01$ in the range between the minimum and the maximum spectroscopic redshift values in the training set, since we found that equal-frequency binning degrades photometric redshift accuracy for catalogues with long-tailed sample PDF. The weights sum to unity and are calculated using

$$w_i(\boldsymbol{f}) = \frac{\hat{p}(b_i|\boldsymbol{f})}{n_{b_i}}\,, \tag{21}$$

where n_{b_i} is the number of training objects with a redshift value in bin b_i. The quantity $\hat{p}(b_i|\boldsymbol{f})$ is an estimate for the class probability that a newly queried object with photometry $\boldsymbol{f}$ has a spectroscopic redshift inside the bin b_i. The method used to obtain the class probability estimates $\hat{p}(b_i|\boldsymbol{f})$ is interchangeable [e.g. using neural networks (Bonnett 2015) or the random forest (Frank & Bouckaert 2009; Carrasco Kind & Brunner 2013)]. We use the random forest algorithm for consistency with the QRF and implemented the OCP algorithm using the 'RANDOMFOREST' (Liaw & Wiener 2002) package for the R programming language (R Core Team 2013).

The original paper by Schapire et al. (2002) used the histogram estimator defined in Frank & Bouckaert (2009) as

$$\hat{p}(z|\boldsymbol{f}) = \sum_{i=1}^{N_{\rm tr}} w_i(\boldsymbol{f}) \frac{I(b_{z_i} = b_z)}{r_{b_z}}\,. \tag{22}$$

Here b_z is defined as an index denoting the bin in which z is located and r_{b_z} denotes the corresponding bin width. We can interpret this histogram as a weighted kernel density estimate with value $r_{b_z}^{-1}$ for all training set objects in a bin specified by b_z and zero outside. Frank & Bouckaert (2009) improved the algorithm using a Gaussian kernel function and demonstrated its superiority over the histogram kernel in numerical experiments on machine learning benchmark data sets that are unrelated to the photometric redshift problem.

3.3 Bandwidth selection

The algorithms we use to obtain PDFs for individual objects require the selection of an appropriate bandwidth for the weighted kernel density estimator (equation 6). This section proposes a bandwidth selection scheme that selects the bandwidth for the Gaussian kernel during model tuning using the MNLL.

The choice of a proper bandwidth depends on the shape of the underlying distribution and the number of objects available to construct the estimator. Assuming a normal distribution and a Gaussian kernel function, one can obtain the optimal bandwidth as

$$\sigma_{\rm scott} = 1.06 \frac{\hat{\sigma}}{N^{1/5}}\,, \tag{23}$$

where $\hat{\sigma}$ is the sample standard deviation and N denotes the number of objects. This so-called Scott's rule is commonly used in the

machine learning and statistics literature (e.g. Wang & Wang 2007; Takeuchi et al. 2009). To apply this bandwidth selection rule to weighted data, we need to calculate the weighted standard deviation from the weighted training set using equation (20). Scott's rule gives a good first estimate of a suitable bandwidth for distributions which are approximately normal.

Photometric redshift PDFs are in general not normal distributions and equation (23) can pick a non-optimal bandwidth. Thus, we modify equation (23) as

$$\sigma_{\rm scott} = a\,\frac{\hat{\sigma}}{N^{1/5}}\,, \tag{24}$$

with a pre-factor a that is chosen to minimize the MNLL on the validation set. We can stack the $N_{\rm te}$ individual PDFs in the test set using an individual bandwidth σ_a for each object

$$\hat{p}(z) = \sum_{a=1}^{N_{\rm te}} w_{{\rm stack},a} \sum_{i=1}^{N_{\rm tr}} w_i(\boldsymbol{f}_a)\mathcal{N}(z, z_i, \sigma_a)\,, \tag{25}$$

or we can use a global bandwidth $\sigma_a = \sigma$.

3.4 The Gaussian mixture model estimator

Storing the individual PDFs obtained by weighted kernel density estimation for every element in the test set requires a large amount of storage. Carrasco Kind & Brunner (2014) proposed several different methods, including a Gaussian mixture model, to more efficiently store a previously obtained estimate. The authors store individual PDFs using 10–20 numbers compared with 200 used previously. Instead of giving a previously estimated individual PDF a sparse representation, we fit the Gaussian mixture model directly to the weighted spectroscopic redshift values in the training set and ensure model sparsity by penalizing the model likelihood dependent on the number of components in the mixture model.

More specifically, we fit the Gaussian mixture to the weighted spectroscopic data with the expectation maximization algorithm (for an introduction see Chen & Gupta 2010) as implemented in the RMIXMOD package (Biernacki et al. 2006; Auder et al. 2014). In Sections 5.1 and 5.2 during the analysis using CFHTLS, we select the number of Gaussian components for each object in the test set using the normalized entropy criterion (Celeux & Soromenho 1996; Biernacki, Celeux & Govaert 1999), abbreviated as NEC in the following. The maximum number of Gaussian components that can be included in the mixture model is a parameter that is selected during model tuning as described in Section 2.4.

For a K component Gaussian mixture model fitted on the weighted training data, the NEC criterion reads

$${\rm NEC}(K) = \frac{E(K)}{L(K) - L(1)}, \tag{26}$$

where $L(K)$ denotes the maximum weighted log-likelihood

$$L(K) = \sum_{i=1}^{N} w(\boldsymbol{f}_i) \log\left(\sum_{k=1}^{K} \alpha_k \mathcal{N}\left(z_{{\rm spec},i}, \mu_i, \sigma_i\right)\right) \tag{27}$$

for the K component Gaussian mixture model. The entropy $E(K)$ is defined as

$$E(K) = -\sum_{k=1}^{K}\sum_{i=1}^{N} w(\boldsymbol{f}_i)\gamma_k\left(z_{{\rm spec},i}\right) \log\left(\gamma_k\left(z_{{\rm spec},i}\right)\right) \;\leq 0\,, \tag{28}$$

where the definition of the component weight proportions, following equation (12), is used. We pick the number of components K such that the NEC criterion is minimized, where ${\rm NEC}(1) = 1$ (Biernacki et al. 1999).

The NEC criterion normalizes the entropy by the maximum weighted log-likelihood, in which the offset for a one-component mixture is subtracted. There are two reasons (Celeux & Soromenho 1996) why we cannot use the entropy $E(K)$ directly. The entropy for $K = 1$ provides a lower bound

$$E(K) \geq E(1) \quad \forall K > 1 \tag{29}$$

and the maximum weighted log-likelihood function is an increasing function of K, which makes $E(K)$ unequal for different values of K. The entropy term $E(K)$ measures how much overlap there is between the different components of the Gaussian mixture model. In the case where the components in the model fit completely separated data clusters, the entropy term approaches zero. If we select too many components, the quantity $E(K)$ will increase because the components will overlap strongly. This can be compensated by the higher likelihood of the more complex model. In this way, we can efficiently determine a suitable number of components to include into the mixture.

3.5 Highest weight element

A common application for individual PDFs is the estimation of the sample PDF. Storing and processing individual PDFs is computationally expensive. We propose the highest weight element (hereafter HWE), a single point estimate for each object from which we can accurately reconstruct the sample PDF. We first run the QRF algorithm to determine weights as for individual PDF estimation. Instead of using the individual PDF, we select the spectroscopic redshift value that is associated with the maximum weight. If more than one spectroscopic redshift value has the same maximum weight, we randomly select one of those values.

4 DATA SET

We use photometric imaging data from the CFHTLS Wide survey using the u^*, g', r', i' and z' bands as obtained from the public CFHTLenS data release (Erben & CFHTLenS Collaboration 2012).[4] We obtain the photometry analogously to Brimioulle et al. (2013), i.e. we degrade all images to match the band with the worst seeing, and use the unconvolved i' band as the detection band and the convolved frames as the extraction band. Then we correct for the remaining zero-point calibration uncertainties and varying galactic extinction by comparing the measured star colours from the catalogues with predictions of the Pickles star library (Pickles 1998). In this way, we eliminate possibly remaining field-to-field variations in the photometric calibration.

We then match our photometric catalogues to public spectroscopic redshift samples. These samples are the Visible Multiobject Spectrograph (VIMOS) VLT Deep Survey (VVDS; Le Fèvre et al. 2004; Garilli et al. 2008), VVDS-F22, the Deep Extragalactic Evolutionary Probe-2 (DEEP-2) programme (Vogt et al. 2005; Weiner et al. 2005; Davis et al. 2007) and the VIMOS Public Extragalactic Redshift Survey (Garilli et al. 2014; Guzzo et al. 2014). We only make use of spectroscopic redshifts with confidence values of at least 95 per cent and only use pointings where the i' data are available and where the i' band serves as detection band.

[4] http://www.cfhtlens.org

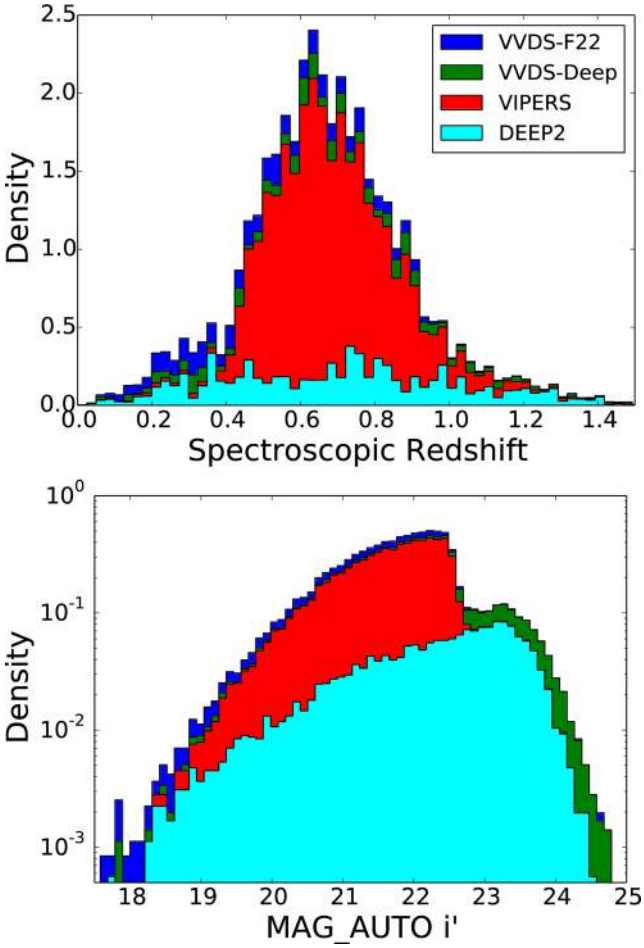

Figure 3. Spectroscopic redshift and MAG_AUTO i' distributions of the compiled data set described in Section 4. Objects matched from different spectroscopic surveys are indicated by different colours. We limit the spectroscopic redshift range to $z_{\rm spec} < 1.5$ in the plots excluding 34 objects with higher redshift.

This produces a total sample of 28 159 objects with $i' \leq 22.5$ and additional 5893 objects with $22.5 < i' \leq 24.5$ with spectroscopic redshifts and five-band photometry. We illustrate the spectroscopic redshift and MAG_AUTO i' distributions of the compiled dataset in Fig. 3.

5 RESULTS

Future large-area photometric surveys will produce large amounts of photometric data for which we need to obtain redshift information. Efficiency in terms of runtime and disc space will be important in order to use algorithms for photometric redshift estimation effectively on these large data sets. Additionally, we are required to produce high-quality photometric redshift PDFs in order to obtain accurate constraints on, for example, cosmological parameters or cluster masses.

We use the public CFHTLS data described in Section 4, to compare the accuracy of photometric redshift PDFs estimated by the algorithms described in Section 3. We show that these methods improve the modelling of angular correlation functions, cluster mass estimates and the modelling of shear correlation functions compared to results obtained with the neural network code ANNZ (Collister & Lahav 2004) commonly used in the literature (e.g. Sheldon et al. 2009; Williamson et al. 2011; Smith et al. 2012; Planck Collaboration 2015).

Table 2. Point prediction performance of the neural network code ANNZ and the template fitting code PHOTOZ quantified by the metrics described in Section 5.1.

	η	$\sigma(\Delta z)$	$\langle\Delta z\rangle$	σ_{68}
ANNZ	1.23 per cent	0.092	−0.001	0.044
PHOTOZ	2.27 per cent	0.129	−0.008	0.050

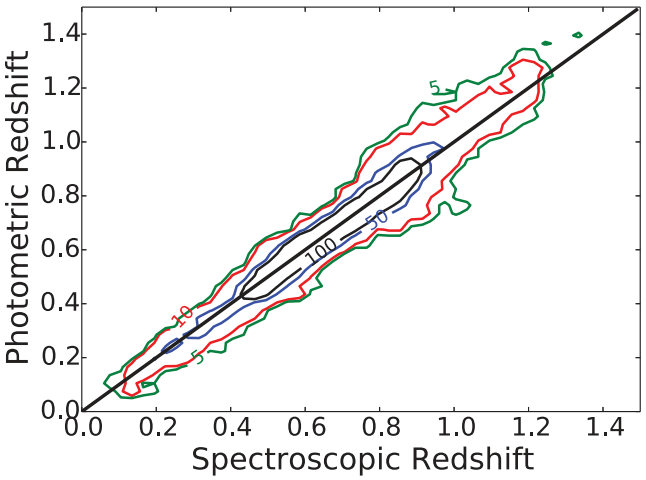

Figure 4. Density contours of photometric redshift estimates from ANNZ against the spectroscopic redshift.

5.1 Comparison with ANNZ

We train an ensemble of 20 neural networks with two hidden layers, each consisting of 12 nodes, following the methodology described in Section 2.4.

The photometric redshift estimates obtained from ANNZ are competitive compared to those obtained with the template fitting code PHOTOZ (Bender et al. 2001; Brimioulle et al. 2008; Greisel et al. 2013) in terms of common photometric redshift performance metrics. As shown in Table 2, ANNZ improves upon the photometric redshift performance obtained with PHOTOZ by 46, 29, 88 and 12 per cent in terms of outlier rate, scatter, bias and spread of the residuals. The outlier rate η is defined as the fraction of objects with $|z_{\rm spec} - z_{\rm phot}| > 0.15$. The bias $\langle\Delta z\rangle$ and scatter $\sigma(\Delta z)$ are the mean and standard deviation of the distribution of the residuals $\Delta z = z_{\rm phot} - z_{\rm spec}$. The spread of the residual distribution is measured by the σ_{68} metric which is defined as half the difference between the 16 and 84 per cent quantile.

The quality of the photometric redshifts obtained with ANNZ is illustrated in Fig. 4. It shows a tightly aligned correlation between photometric and spectroscopic redshift. We estimate sample PDFs from the ANNZ point predictions and the stacked normal densities constructed from the ANNZ error estimates, in the following referred to as 'ANNZ-stack'. While showing excellent point prediction performance, ANNZ and ANNZ-stack do not accurately estimate the sample PDF as shown in Fig. 5. The sample PDF constructed from ANNZ-stack deviates from the true spectroscopic redshift distribution in the central redshift range [0.45, 0.85]. We will show in the following sections that these deviations from the true spectroscopic redshift PDF introduce a systematic bias in several important analyses in

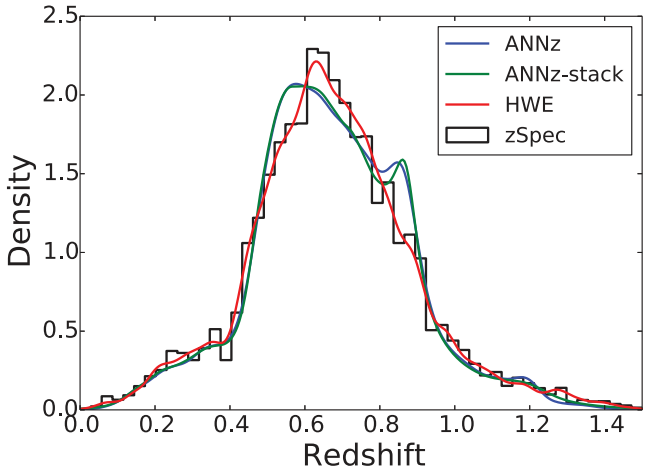

Figure 5. Sample PDF estimated using ANNZ and the HWE. The histogram shows the true spectroscopic redshift distribution.

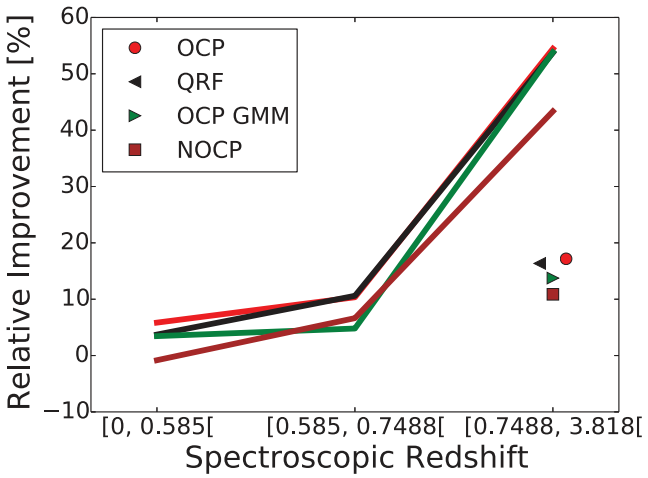

Figure 6. Relative improvement in MNLL over the performance of ANNZ-stack. We compare the classification-based PDF estimators (OCP, NOCP), the ordinal classification PDF estimator combined with a Gaussian mixture model (OCP GMM) and the QRF in three spectroscopic redshift bins. The plotted points show the average improvement over the full spectroscopic redshift range.

Table 3. MNLL of the QRF, the classification-based PDF estimation algorithms (OCP/NOCP) and the OCP algorithm used with the Gaussian mixture model (GMM). The values are evaluated over the full spectroscopic redshift range and in three bins. The result is illustrated in Fig. 6.

	Total	[0, 0.585[	[0.585, 0.7488[	[0.7488, 3.818[
OCP	−1.3577	−1.3905	−1.6432	−1.0395
NOCP	−1.2847	−1.3029	−1.5880	−0.9648
QRF	−1.3483	−1.3627	−1.6470	−1.0347
GMM	−1.3181	−1.3591	−1.5606	−1.0354
ANNZ	−1.1588	−1.3138	−1.4891	−0.6731

cosmology. To compare the quality of photometric redshift PDFs of individual objects (individual PDFs), we evaluate the MNLL (equation 15) of the four discussed algorithms (QRF, NOCP, OCP and OCP GMM) on the full range of redshift values and in three redshift bins ([0, 0.585[, [0.585, 0.7488[and [0.7488, 3.818[). The results are shown in Table 3 and illustrated in Fig. 6. QRF, NOCP and OCP employ the weighted kernel density estimate. OCP GMM denotes the Gaussian mixture model applied in combination with weights determined using the ordinal classification method described in Section 3.2.

We illustrate the relative improvement $\mathrm{MNLL}_{\mathrm{rel}}$ gained by applying these algorithms compared with ANNZ-stack

$$\mathrm{MNLL}_{\mathrm{rel}} = \left(\frac{\mathrm{MNLL}_{\mathrm{ANNz}} - \mathrm{MNLL}_{\mathrm{alg.}}}{|\mathrm{MNLL}_{\mathrm{ANNz}}|} \right) \tag{30}$$

in Fig. 6. A high value in terms of $\mathrm{MNLL}_{\mathrm{rel}}$ translates into an improvement in the log-likelihood of the individual PDFs over those obtained with ANNZ-stack. The boundaries of the redshift intervals are picked such that they contain approximately the same number of test set objects. All discussed methods improve over ANNZ-stack. For the highest redshift objects, our methods show improvement of up to 50 per cent. The OCP routine performs the best and improves the NOCP routine. This verifies the superiority of the ordinal classification technique. The QRF performs on par with OCP. OCP GMM shows mediocre results, but provides the most efficient parametrization using a single normal density per object.

5.2 Stacked photometric redshift distribution

Applications like shear tomography require the photometric selection of objects in a certain redshift range. We stack the individual PDFs compared in Section 5.1 using weights that quantify their overlap with a certain redshift interval using equation (10). These estimates are compared with the weighted kernel density estimate obtained from the spectroscopic redshift values using the same weights. The weights are determined using each of the OCP, NOCP, OCP GMM and QRF methods individually. The HWE uses the weighted kernel density estimate with weights determined using the QRF algorithm. We use Scott's rule to choose the bandwidth for the weighted kernel density estimates of the HWE and the spectroscopic redshift values. The sample PDFs obtained with the OCP, NOCP and QRF algorithms are very similar. We therefore restrict the following discussion to the OCP method.

The results shown in Fig. 7 compare the weighted sample PDFs obtained with the HWE, OCP and OCP GMM methods in the redshift intervals [0, 0.585[, [0.585, 0.7488[and [0.7488, 3.818[. They differ mainly in the amount of smoothing present in the estimate. Notably the OCP GMM method oversmoothes features in the density estimate. This is because a single Gaussian was selected during model tuning based on the performance of individual object PDFs. Allowing more components reduces the amount of smoothing. The HWE is competitive with methods that estimate the individual PDFs, with the advantage that the HWE is extremely efficient to calculate and, being a point estimate, requires storing only a single floating point number per object.

The weighted distributions of all methods have tails that extend outside the desired redshift range. We can reduce these tails by stacking only the objects with the highest weight in the respective redshift bin as demonstrated in the lower-right panel of Fig. 7. We estimate the sample PDF from the HWE predictions of the objects with the 5000 highest weights in the respective redshift bin. The estimated weighted sample PDF of these objects has less overlap with neighbouring redshift bins, compared with the estimate that incorporates all objects. Furthermore, it agrees well with the equally weighted spectroscopic redshift distribution of the corresponding objects.

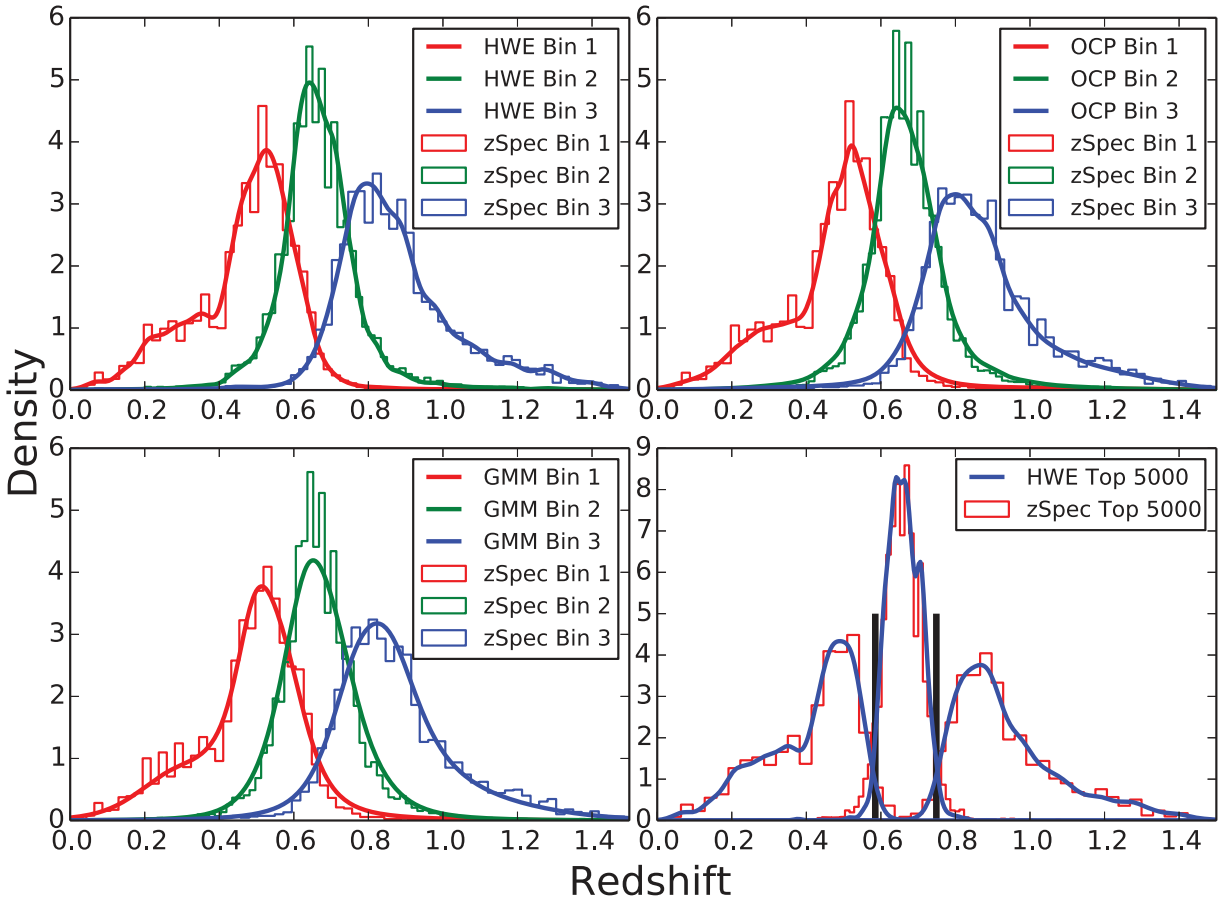

Figure 7. Sample PDFs weighted in three redshift intervals [0, 0.585[, [0.585, 0.7488[and [0.7488, 3.818[. The PDFs are obtained using the HWE (upper left), the ordinal classification PDF estimator (upper right) and the ordinal classification PDF estimator combined with a Gaussian mixture model (lower left). The histograms show the weighted spectroscopic redshift distribution using weights determined using the respective algorithms. The lower-right panel shows the weighted distribution of the HWE predictions for the objects with the 5000 highest weights in the three intervals (blue) and the corresponding weighted histogram of spectroscopic redshifts (red).

Instead of weighting the objects in the respective redshift range, we can select objects based on a photometric redshift point estimate in analogy with Benjamin et al. (2013). We perform the same cut in `MAG_AUTO` $i' < 23.0$ and estimate the sample PDF in the same photometric redshift intervals selected after our ANNZ estimate. The results for the HWE are shown in Fig. 8 and agree well with the spectroscopic redshift distribution. The agreement is better in the central bins, which contain more objects, because the histogram approximates the underlying distribution better.

5.3 Applications to cosmology

We now investigate how the previously discussed methods can be used to improve analyses that use photometric redshifts. We estimate the sample PDF using the HWE and ANNZ. We use kernel density estimates with bandwidths selected using Scott's rule.

Where required, we impose a flat Λ cold dark matter (ΛCDM) cosmology with $\Omega_\mathrm{m} = 0.3, \Omega_\Lambda = 0.7, n_\mathrm{s} = 0.96, H = 0.7, \sigma_8 = 0.79$.

5.3.1 The angular power spectrum

The angular power spectrum measures the clustering of galaxies and is an important tool to constrain cosmological models.

In the following, we adopt the notation of Thomas, Abdalla & Lahav (2010). Consider the line-of-sight projection of the 3D mass distribution in the universe, $\delta_{2\mathrm{D}}$. The harmonic modes of $\delta_{2\mathrm{D}}$ are given by

$$\delta_\ell = i^\ell \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \delta(\boldsymbol{k}) W_\ell(k)\,, \tag{31}$$

where the window function $W_\ell(k)$ is sensitive to the sample PDF of light sources, $p(z)$, and can be computed by the integral

$$W_\ell(k) = \int p(z) D(z) \left(\frac{\mathrm{d}z}{\mathrm{d}x}\right) j_\ell(kz)\mathrm{d}z\,. \tag{32}$$

Here $D(z)$ is the linear growth factor, $j_\ell(kz)$ are the Bessel functions and $\left(\frac{\mathrm{d}z}{\mathrm{d}x}\right)$ relates the redshift to the radial comoving coordinate x.

The angular power spectrum C_ℓ is the variance of the modes δ_ℓ,[5]

$$C_\ell = \left\langle \delta_\ell \delta_\ell^* \right\rangle = 4\pi \int \Delta^2(k) W_\ell^2(k) \frac{\mathrm{d}k}{k}\,, \tag{33}$$

where the dimensionless 3D power spectrum $\Delta^2(k)$ is given in terms of the usual 3D matter power spectrum $P_\delta(k)$ as

$$\Delta^2(k) = \frac{4\pi k^3 P_\delta(k)}{(2\pi)^3}\,. \tag{34}$$

[5] In our analysis, we are assuming a galaxy-dark matter bias equal to one.

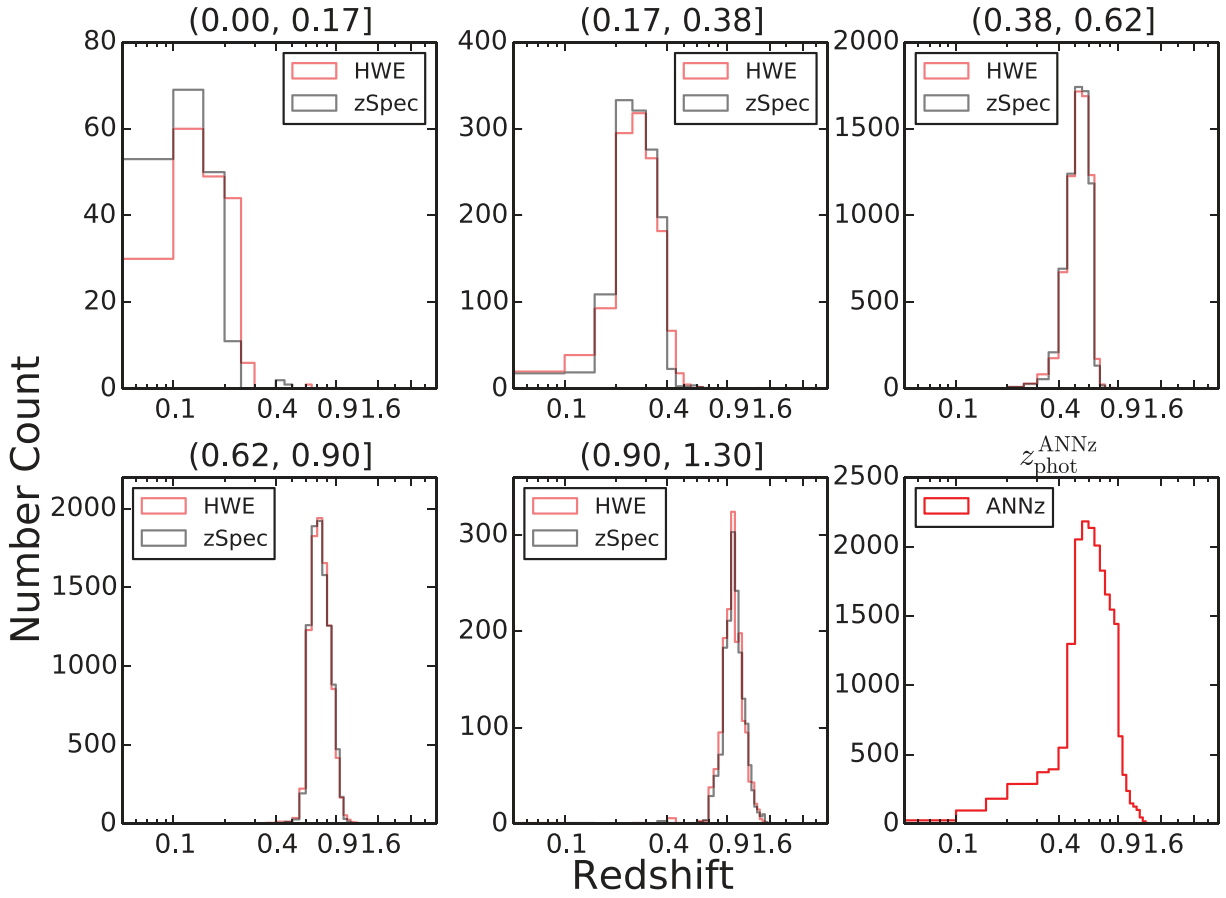

Figure 8. Sample PDFs estimated using the HWE for subsamples selected in analogy with Benjamin et al. (2013, fig. 1) using a cut at MAG_AUTO $i' < 23.0$. The subsamples are selected using the photometric redshift estimates from ANNZ in intervals shown in the subfigure titles.

From equation (32) it can be seen that the modelling of C_ℓ depends highly on the assumed sample PDF of the data. We use the distributions shown in Fig. 5 to model the angular correlation power spectrum with the CLASS software package (Blas, Lesgourgues & Tram 2011). We define the bias introduced by the $C_\ell^{\rm phot}$ of the angular correlation function estimated using photometric redshifts, as the relative difference to the results based on the PDF of spectroscopic redshifts $C_\ell^{\rm spec}$:

$$\mathrm{Bias}_{C_\ell} = \left(\frac{C_\ell^{\rm phot} - C_\ell^{\rm spec}}{C_\ell^{\rm spec}} \right) . \tag{35}$$

The resulting biases are shown in Fig. 9. We find that the results obtained with the HWE have a lower systematic bias in C_ℓ by a factor of 4 compared to the ANNZ results and that the improvement is almost independent of ℓ.

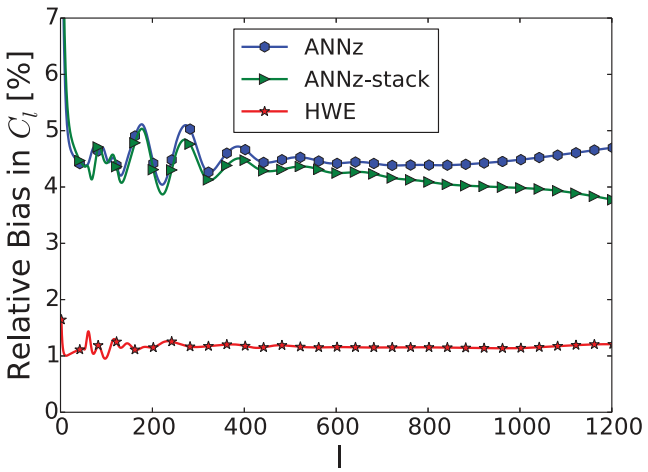

Figure 9. Bias in the angular correlation power spectrum obtained for different estimates for the sample PDF. We restrict the comparison to $\ell < 1200$.

5.3.2 *Gravitational lensing*

We investigate two important applications in gravitational lensing: quantifying cluster masses by the light deflection from background sources and obtaining cosmic shear correlation functions. In contrast to the previously considered analysis of the angular correlation function, applications in gravitational lensing require careful selection of sources with successfully measured shapes. Since the spectrophotometric data set used previously is not representative for data sets generally used in gravitational lensing analyses, we first weight our catalogue such that it mimics a CFHTLS shape catalogue. To do this, we obtain a photometric shape catalogue from public CFHTLS data, which is then used as the reference to weight the spectrophotometric data set.

5.3.3 *Catalogue creation and weighting*

Whether the shape of an object can be measured depends primarily on its intrinsic size and magnitude in the respective band. We therefore re-weight our spectrophotometric catalogue such that it resembles a CFHTLS shape catalogue in terms of these properties. We obtain the shape catalogue in analogy with Brimioulle et al. (2013) for the full CFHTLS survey region. Intrinsic sizes $s_{\rm intr}$ are calculated for each object from the measured $\mathrm{FWHM}_{\rm image}$ and corrected for seeing as follows:

$$s_{\rm intr} = \sqrt{\mathrm{FWHM}_{\rm image}{}^2 - \left\langle \mathrm{FWHM}_{\rm psf} \right\rangle^2}\,, \tag{36}$$

where $\left\langle \mathrm{FWHM}_{\rm psf} \right\rangle$ is the average size of the point spread function for the respective chip.[6]

In this way, we obtain $s_{\rm intr}$ and MAG_AUTO i' entries for each object in the shape and spectrophotometric catalogue. We now determine weights for the spectrophotometric catalogue such that, after weighting, it matches the size and magnitude distribution of the shape catalogue. Furthermore, the results obtained with the re-weighted spectrophotometric catalogue have to be robust against the removal of the objects with the highest weights (Sánchez et al. 2014). Since we do not have enough spectroscopically observed objects to mimic the shape catalogue at the faint end, we have to employ a magnitude cut in order to fulfil both requirements. For the analyses presented in Sections 5.3.4 and 5.3.5, we employ a magnitude cut at MAG_AUTO $i' < 23.5$ and MAG_AUTO $i' < 23.0$, respectively. We give a detailed discussion of these cuts in Appendix A.

We combine bootstrap re-sampling with the k nearest-neighbour estimator to determine weights for the elements in the spectrophotometric catalogue, such that the weighted catalogue mimics the distribution of the shape catalogue in the two-dimensional space spanned by the intrinsic size of the objects and their magnitude MAG_AUTO i'. To this end, we draw bootstrap samples from the shear catalogue and find the k nearest neighbours in the spectrophotometric catalogue. The nearest neighbour of an object in the spectrophotometric catalogue is the object in the shear catalogue with the lowest Euclidean distance to this object. Accordingly, the k nearest-neighbour algorithm selects the k nearest objects with respect to the Euclidean distance. The number of times an object in the spectrophotometric catalogue is selected as one of the k nearest neighbours corresponds to its weight. This process is similar to previous work done by Lima et al. (2008), which employs a nearest-neighbour-based approach to determine weights for objects in a spectroscopic sample to estimate the sample PDF of the photometric data. In contrast to our method, which is based on bootstrap re-sampling, they calculate the density ratio between the distributions characterizing the two catalogues using a nearest-neighbour approach. For the data at hand, we draw 10^6 bootstrap samples and consider three nearest neighbours $k = 3$. This method accurately weights the spectrophotometric data to mimic the size and i-band magnitude distributions of the shape catalogue, as shown in Figs 10 and 11. The following analysis uses the estimated weights to weight the sample PDF of ANNZ, ANNZ-stack, the HWE and the spectroscopic data as shown in Fig. 12.

[6] We work on image stacks, but (as in Brimioulle et al. 2013) only consider objects, for which all images contribute to the stack from the same CCD-chip.

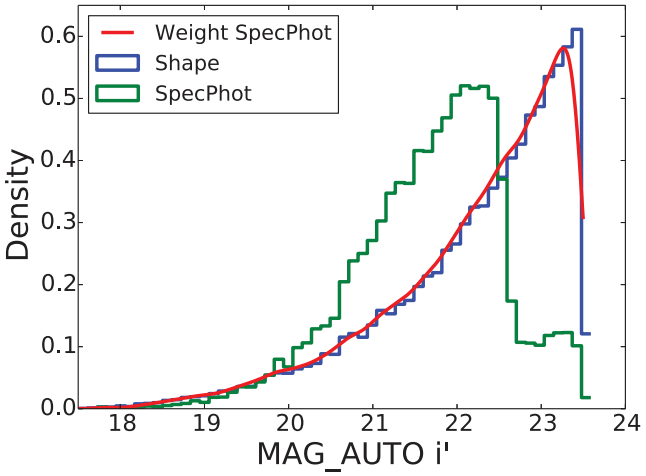

Figure 10. Distributions in MAG_AUTO i' band for the original spectrophotometric data set, the re-weighted spectrophotometric data set and the shape catalogue for MAG_AUTO $i' < 23.5$.

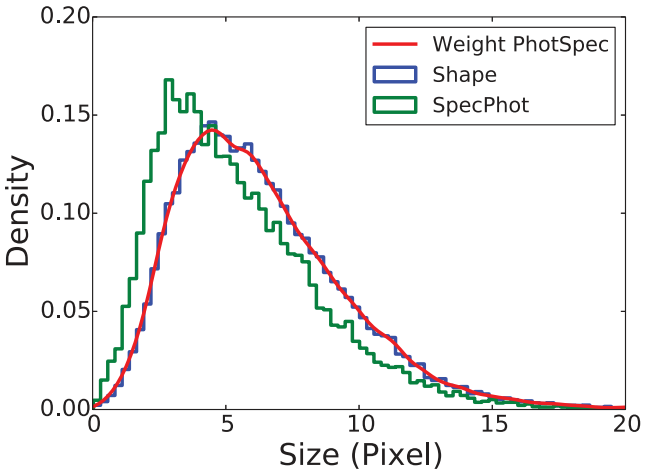

Figure 11. Distributions in intrinsic size (equation 36) for the original spectrophotometric data set, the re-weighted spectrophotometric data set and the shape catalogue for MAG_AUTO $i' < 23.5$.

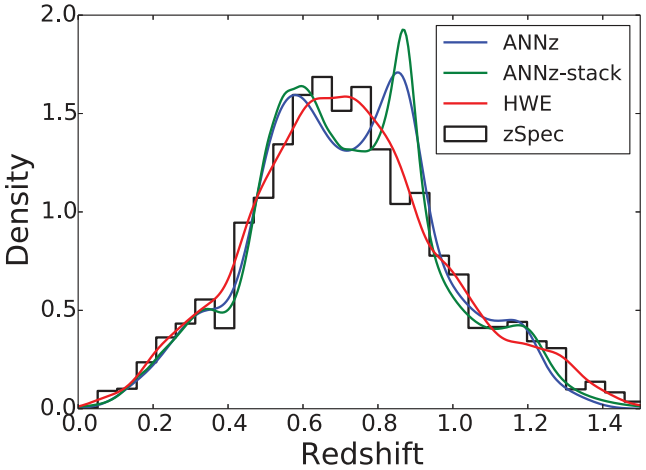

Figure 12. Weighted stacked sample PDF estimated using ANNZ, ANNZ-stack and the HWE. The histogram shows the weighted spectroscopic redshift distribution. We use a cut on MAG_AUTO $i' < 23.5$. The used weights and cuts are described in Section 5.3.3.

5.3.4 Cluster mass measurement

Galaxy clusters are one of the primary tools to probe the ΛCDM model (for a review, see e.g. Allen, Evrard & Mantz 2011). Cluster masses can be determined by measuring the tangential alignment of gravitationally lensed galaxies[7] located behind the clusters. The accuracy of these weak lensing mass estimates suffers from uncertainties in the photometric redshift of the lensed sources. In combination with other effects such as cluster mass profile variances, they can introduce systematics at the 5–10 per cent level (see e.g. Applegate et al. 2014). In the following, we will only consider uncertainties due to errors in photometric redshift estimates (Seitz & Schneider 1997; Mandelbaum et al. 2008; Dawson et al. 2012; Gruen et al. 2013, 2014; Applegate et al. 2014). The excess surface density inside radius R

$$\langle\Sigma(r)\rangle_{r<R} - \overline{\Sigma}(R) = \Sigma_{\rm crit}\,\gamma_{\rm tan}(R) \tag{37}$$

is proportional to the critical surface density

$$\langle\Sigma_{\rm cr}\rangle \propto \int_{z_{\rm Lens}}^{\infty} {\rm d}z\; p(z) \left(\frac{D_{\rm d}(z_{\rm Lens}) D_{\rm ds}(z_{\rm Lens}, z)}{D_{\rm s}(z)}\right) \tag{38}$$

of the lens at redshift $z_{\rm Lens}$. Here $D_{\rm d}$, $D_{\rm s}$ and $D_{\rm ds}$ denote the angular diameter distance to the lens, the source and between the lens and the source, respectively. Uncertainties in the sample PDF of background sources $p(z)$ will propagate into systematic errors in the determination of the critical surface density. This introduces systematic errors in the excess surface density and therefore in the cluster mass estimate.

We quantify the systematic bias of the critical surface density as

$${\rm Bias}_{\langle\Sigma_{\rm cr}\rangle} = \left(\frac{\langle\Sigma_{\rm cr}\rangle_{\rm photo} - \langle\Sigma_{\rm cr}\rangle_{\rm spec}}{\langle\Sigma_{\rm cr}\rangle_{\rm spec}}\right), \tag{39}$$

where $\langle\Sigma_{\rm cr}\rangle_{\rm photo}$ is estimated from the photometry of the objects (e.g. using machine learning) and $\langle\Sigma_{\rm cr}\rangle_{\rm spec}$ from the spectroscopic redshifts.

We estimate the error σ on this bias with respect to our test set containing N objects as

$$\sigma^2 = \left(\frac{\sigma_{\rm photo}(\Sigma_{\rm cr})}{\sqrt{N}\,\langle\Sigma_{\rm cr}\rangle_{\rm spec}}\right)^2. \tag{40}$$

The mean and standard deviation of the distribution of $\Sigma_{\rm cr}$ are estimated using the PDF estimates obtained from ANNZ and the HWE and we present the results in Fig. 13.

The HWE estimate for the sample PDF reduces the systematic bias in the critical surface density compared with ANNZ by a factor of 4 for medium lens redshifts $z \in [0.45, 0.6]$. The systematic bias in $\langle\Sigma_{\rm cr}\rangle$ obtained from the HWE is consistent with zero for lens redshifts $z < 0.7$ and, in general, outperforms the results obtained with ANNZ. Higher lens redshifts are however unrealistic for current survey depths.

5.3.5 Cosmic shear

Cosmic shear is the weak lensing effect generated by the inhomogeneous matter distribution of the universe and has become an important tool to constrain cosmological parameters (see, e.g., Kilbinger et al. 2013, and references therein). Similar to our discussion of the angular correlation function, we derive a power spectrum

[7] For an introduction into gravitational lensing, we refer to Bartelmann & Schneider (2001).

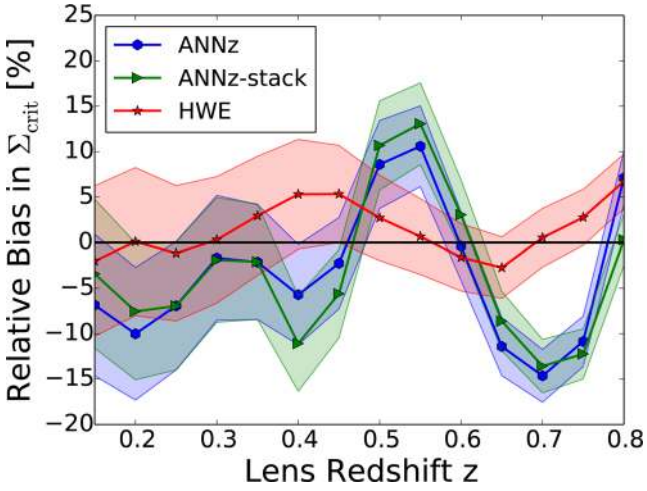

Figure 13. Relative bias in the mean critical surface density (equation 39) for different lens redshifts obtained using different estimates for the sample PDF. The filled area shows the 1σ error interval.

$P_\kappa(\ell)$ of the lensing convergence κ, which is the source of the lensing potential, defined with respect to the radial comoving coordinate x

$$P_\kappa(\ell) = \int_0^\infty {\rm d}x \left(\frac{q^2(x)}{x^2}\right) P_\delta\left(\frac{\ell}{x}, x\right). \tag{41}$$

We calculate the power spectrum $P_\delta\left(\frac{\ell}{x}, x\right)$ using the halofit formula from Smith et al. (2003). The lensing efficiency $q(x)$ quantifies how strongly the objects in an infinitesimal shell of radial comoving coordinates deflect the light coming from background sources. Since the radial comoving coordinates of the objects are related to their redshifts, the lensing efficiency $q(x)$ depends on the sample PDF $p(z)$. From the lensing convergence power spectrum, we can obtain the two shear correlation functions (Kaiser 1992) as

$$\xi_\pm(\theta) = \frac{1}{2\pi}\int_0^\infty {\rm d}\ell\;\ell\;P_\kappa(\ell)\;J_{0,4}(\ell\theta), \tag{42}$$

where the Bessel function J_0 (J_4) corresponds to the ξ_+ (ξ_-) correlation function. In analogy with the previous sections, we quantify the bias in the shear correlation functions obtained from photometric data $\xi_\pm^{\rm phot}$ by their relative error with respect to the results obtained from the spectroscopic data $\xi_\pm^{\rm spec}$,

$${\rm Bias}_{\xi_\pm} = \left(\frac{\xi_\pm^{\rm photo} - \xi_\pm^{\rm spec}}{\xi_\pm^{\rm spec}}\right). \tag{43}$$

The results are presented in Figs 14 and 15. We reduce the bias in the shear correlation function estimates, using the HWE estimate instead of the photometric redshift estimates from ANNZ, by a factor of 12 for ξ_- and a factor of 6 for ξ_+.

6 SUMMARY AND CONCLUSIONS

The next-generation photometric surveys will measure the positions on the sky of thousands of millions of galaxies. We must be able to reliably estimate the distance to, or the redshift of, each photometrically identified galaxy before we can use these galaxies in analyses to derive the values of cosmological parameters. Furthermore to maximize the precision and accuracy of any derived parameters, we require a complete understanding of the full shape of the photometric redshift PDF for both each individual object and the entire galaxy sample.

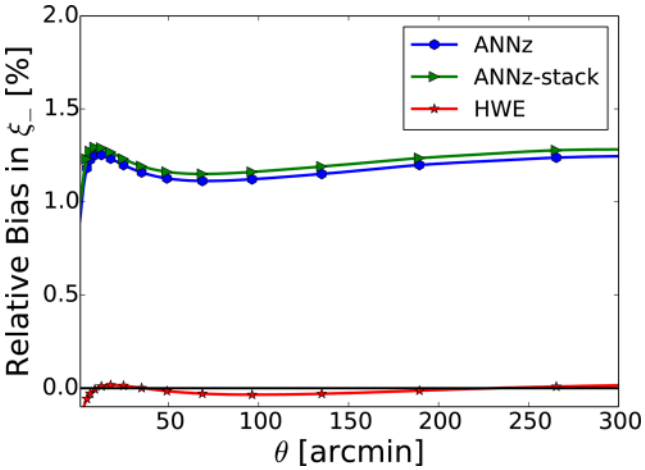

Figure 14. Relative bias in the shear correlation function estimate for ξ_- (equation 43) obtained using different estimates for the sample PDF.

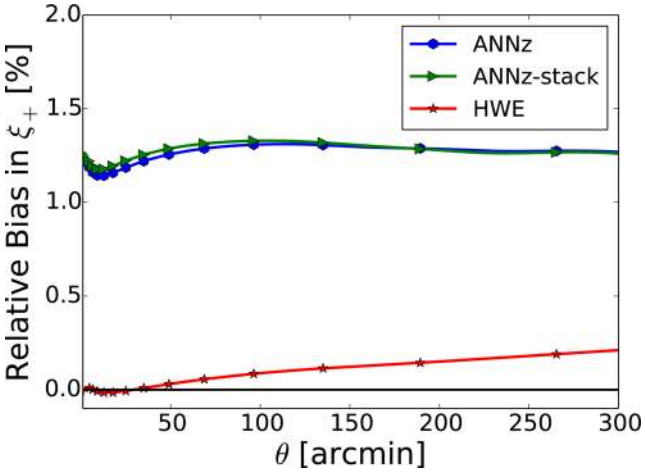

Figure 15. Relative bias in the shear correlation function estimate for ξ_+ (equation 43) obtained using different estimates for the sample PDF.

In this work, we develop and discuss methods drawn from machine learning, to accurately estimate photometric redshift PDFs, which will meet both the future storage demands of large surveys and the precision demands for cosmological parameter estimation.

As a working example, we apply these algorithms to a sample of galaxies selected from the CFHTLS survey for a set of cosmological analyses. We demonstrate that these methods reduce the biases in all of the analyses examined. We also show that these biases result from the mishandling of the full shape of the photometric redshift PDFs.

This advancement is quantified by comparing several accurate methods to estimate photometric redshift PDFs for individual objects. We estimate individual PDFs using a classification scheme that classifies objects into redshift bins and thereby constructs the PDF using the probabilities for bin membership. In contrast to the classification-based PDF estimation methodology commonly used in the astrophysics literature, we incorporate the order of consecutive redshift bins into the classification framework. This produces more accurate individual PDFs. We quantify the performance of the methods by measuring the average log-likelihood of all PDF estimates in a test sample. Our method outperforms other non-ordinal classification and regression schemes, for example classification trees and neural networks. Specifically, for high-redshift objects, our method reaches performance gains of over 50 per cent in average log-likelihood when compared with the results obtained using the common neural network code ANNZ. We construct the individual PDFs using kernel density estimation which inherently requires the selection of a suitable bandwidth to govern the smoothing scale. We propose an efficient method to choose the smoothing scale on an object-by-object basis. We further discuss a Gaussian mixture model, whose complexity is adaptively selected for each individual object, using a criterion that penalizes model complexity. This method shows solid performance compared with kernel density estimates, while providing a more efficient parametrization of individual PDFs.

Many cosmological analyses require accurate knowledge of the full shape of the galaxy sample PDF, instead of estimates for the individual PDFs of each galaxy. Sample PDFs are typically obtained by stacking the PDFs of individual galaxies, and so their estimation and storage are required. This reconstruction of the individual PDF typically requires the storage of several hundred floating point numbers. Complex post processing algorithms can reduce this number to 10–20 floating point numbers per object at the expense of additional computation time. However, in this work, we propose a new single point estimator for each galaxy, called highest weight element (HWE), which can be used to accurately reconstruct the full sample PDF. This leads to a significant reduction in the storage requirements of future photometric surveys. Furthermore, we note that reconstructing the full sample PDF using the point estimator method described in this paper requires orders of magnitude less computation time than using other common redshift codes.

Applications such as shear tomography require the accurate photometric selection of objects in redshift bins. We weight photometrically observed galaxies such that their sample PDF lies within the predefined redshift range. The weights are estimated from the overlap between the individual redshift PDFs and the redshift selection interval. We further use these weights to improve the selection of a sample of galaxies, such that their sample redshift PDF is more accurately confined to be within the predefined redshift bin.

We now return our attention to the specific use case highlighted above using CFHTLS galaxies. In particular, we examine the following cosmological analyses: the estimation of cluster masses using weak gravitational lensing, the modelling of galaxy angular correlation functions and the modelling of cosmic shear correlation functions. In each case, we compare the results, and estimate biases, using results obtained with ANNZ.

For lensing clusters within the redshift interval $0.45 < z < 0.6$, we show that our methods reduce the relative bias in the cluster mass reconstruction by up to a factor of 4. Furthermore, our methods improve the relative biases in the modelling of the explored large-scale structure and cosmic shear correlation functions by similar values.

In this paper, we have shown that the usual point estimate of a photometric redshift is a poor estimator when used to reconstruct the full sample redshift PDF. We note that these point estimates are still used in many recent analyses, and we have shown that their continued use can lead to large biases in cosmological analysis. By using the new HWE point estimator method, highlighted in this paper, we show that the full shape of the sample PDF can be estimated more accurately and that this reduces the biases incurred by misestimating the sample PDF.

The results discussed in this paper have been obtained under the idealized assumption that the data used to train the models are completely representative of the test data. In applications where this is

not the case, data augmentation techniques (Hoyle et al. 2015b) can be used to artificially populate regions of colour–magnitude space, which are not fully covered by spectroscopy. These techniques assume a model for the data distribution and can be seen as a form of extrapolation. Weighting methods (Section 5.3.3) are in some cases an alternative to data augmentation. If all relevant attributes are included, these algorithms can be used to determine weights, such that the weighted data set resembles a reference data set.

To aid the common adoption of these tools and techniques, we will make the source code of all algorithms publicly available on the homepage of the first author.

ACKNOWLEDGEMENTS

SS thanks Ofer Lahav for drawing her attention towards machine learning techniques several years ago, and for inspiring discussions on machine learning versus template fitting during mutual visits.

MMR especially thanks Jolanta Krzyszkowska and Natascha Greisel for useful discussions.

This work was supported by SFB-Transregio 33 'The Dark Universe' by the Deutsche Forschungsgemeinschaft (DFG) and the DFG cluster of excellence 'Origin and Structure of the Universe'.

REFERENCES

Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409
Applegate D. E. et al., 2014, MNRAS, 439, 48
Auder B., Lebret R., Iovleff S., Langrognet F., 2014, Rmixmod: An Interface for MIXMOD. R Package Version 2.0.2
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bender R. et al., 2001, in Cristiani S., Renzini A., Williams R. E., eds, ESO Proc. Deep Fields. Springer, Berlin, p. 96
Benítez N., 2000, ApJ, 536, 571
Benjamin J. et al., 2013, MNRAS, 431, 1547
Biernacki C., Celeux G., Govaert G., 1999, Pattern Recognit. Lett., 20, 267
Biernacki C., Celeux G., Govaert G., Langrognet F., 2006, Comput. Stat. Data Anal., 51, 587
Bishop C. M., 2006, Pattern Recognition and Machine Learning. Springer, Berlin
Blas D., Lesgourgues J., Tram T., 2011, J. Cosmol. Astropart. Phys., 7, 34
Bonnett C., 2015, MNRAS, 449, 1043
Breiman L., 2001, Mach. Learn., 45, 5
Brimioulle F., Lerchster M., Seitz S., Bender R., Snigula J., 2008, preprint (arXiv:0811.3211)
Brimioulle F., Seitz S., Lerchster M., Bender R., Snigula J., 2013, MNRAS, 432, 1046
Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483
Carrasco Kind M., Brunner R. J., 2014, MNRAS, 441, 3550
Celeux G., Soromenho G., 1996, J. Classif., 13, 195
Chen Y., Gupta M. R., 2010, Technical Report No. UWEETR-2010-0002, Em Demystified: An Expectation-Maximization Tutorial, available at: https://www.ee.washington.edu/techsite/papers/
Colless M. et al., 2001, MNRAS, 328, 1039
Collister A. A., Lahav O., 2004, PASP, 116, 345
Csabai I., Connolly A. J., Szalay A. S., Budavári T., 2000, AJ, 119, 69
Cunha C. E., Lima M., Oyaizu H., Frieman J., Lin H., 2009, MNRAS, 396, 2379
Dahlen T. et al., 2013, ApJ, 775, 93
Davis M. et al., 2007, ApJ, 660, L1
Dawson W. A. et al., 2012, ApJ, 747, L42
de Leeuw J., Hornik K., Mair P., 2009, J. Stat. Softw., 32, 1
de Simoni F. et al., 2013, MNRAS, 435, 3017
Drinkwater M. J. et al., 2010, MNRAS, 401, 1429
Duin R., 1976, IEEE Trans. Comput., 25, 1175
Erben T., CFHTLenS Collaboration 2012, in American Astronomical Society Meeting Abstracts Vol. 219, p. 130.09
Feldmann R. et al., 2006, MNRAS, 372, 565
Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339, 1195
Flaugher B., 2005, Int. J. Mod. Phys. A, 20, 3121
Frank E., Bouckaert R., 2009, in Zhou E.-H., Washio T., eds, Springer Proc. First Asian Conference on Machine Learning, ACML 2009, Conditional Density Estimation with Class Probability Estimators. Springer, Berlin, p. 65
Frank E., Hall M., 2001, in De Raedt L., Flach P., eds, Lecture Notes in Computer Science, Vol. 2167, Machine Learning: ECML 2001. Springer, Berlin, p. 145
Garilli B. et al., 2008, A&A, 486, 683
Garilli B. et al., 2014, A&A, 562, A23
Greisel N., Seitz S., Drory N., Bender R., Saglia R. P., Snigula J., 2013, ApJ, 768, 117
Gruen D. et al., 2013, MNRAS, 432, 1455
Gruen D. et al., 2014, MNRAS, 442, 1507
Guzzo L. et al., 2014, A&A, 566, A108
Habbema J. D. F., Hermans J., Van den Broek K., 1974, in Bruckman G., ed., Proc. Computational Statistics, Compstat 1974, A Stepwise Discrimination Analysis Program Using Density Estimation. Physica Verlag, Vienna, p. 101
Hildebrandt H. et al., 2010, A&A, 523, A31
Hoyle B., Rau M. M., Zitlau R., Seitz S., Weller J., 2015a, MNRAS, 449, 1275
Hoyle B., Rau M. M., Bonnett C., Seitz S., Weller J., 2015b, MNRAS, 450, 305
Ilbert O. et al., 2006, A&A, 457, 841
Kaiser N., 1992, ApJ, 388, 272
Kilbinger M. et al., 2013, MNRAS, 430, 2200
Laureijs R. et al., 2011, Euclid definition study report (Red Book), preprint (arXiv:1110.3193)
Le Fèvre O. et al., 2004, A&A, 417, 839
Le Fèvre O. et al., 2005, A&A, 439, 845
Liaw A., Wiener M., 2002, R News, 2, 18
Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118
Mandelbaum R. et al., 2008, MNRAS, 386, 781
Meinshausen N., 2006, J. Mach. Learn. Res., 7, 983
Pickles A. J., 1998, PASP, 110, 863
Planck Collaboration 2015, preprint (arXiv:1502.01595)
R Core Team 2013, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna
Sánchez C. et al., 2014, MNRAS, 445, 1482
Schapire R. E., Stone P., McAllester D. A., Littman M. L., Csirik J. A., 2002, in Sammut C., Hoffmann A. G., eds, Proc. ICML '01, Modeling Auction Price Uncertainty Using Boosting-Based Conditional Density Estimation. Morgan Kaufmann, San Francisco, p. 546
Scott D. W., 1992, Multivariate Density Estimation: Theory, Practice, and Visualization, 1st edn. Wiley, New York
Seitz C., Schneider P., 1997, A&A, 318, 687
Sheldon E. S. et al., 2009, ApJ, 703, 2217
Smith R. E. et al., 2003, MNRAS, 341, 1311
Smith D. J. B. et al., 2012, MNRAS, 427, 703
Staniszewski Z. et al., 2009, ApJ, 701, 32
Sugiyama M., Takeuchi I., Suzuki T., Kanamori T., Hachiya H., Okanohara D., 2010, IEICE Trans. Inf. Syst., E93-D, 583
Takeuchi I., Nomura K., Kanamori T., 2009, Neural Comput., 21, 533
Thomas S. A., Abdalla F. B., Lahav O., 2010, preprint (arXiv:1011.2448)
Tonry J. L. et al., 2012, ApJ, 750, 99
Tyson J. A., Wittman D. M., Hennawi J. F., Spergel D. N., 2003, Nucl. Phys. B, 124, 21
Vogt N. P. et al., 2005, ApJS, 159, 41
Wang B., Wang X., 2007, preprint (arXiv:0709.1616)
Weiner B. J. et al., 2005, ApJ, 620, 595
Williamson R. et al., 2011, ApJ, 738, 139
York D. G. et al., 2000, AJ, 120, 1579

APPENDIX A: TESTS OF WEIGHTING SCHEME

The analyses in Sections 5.3.5 and 5.3.4 have been carried out by weighting the photospectroscopic data set such that it resembles a shape catalogue. If only a few objects in the re-weighted catalogue are given high weights, the analyses can strongly depend on these objects. We lack spectrocopically observed objects at the faint end of the shape catalogue and therefore employ a magnitude cut to avoid giving large weight to the faint, unrepresentative part of the spectrophotometric catalogue. In analogy with Sánchez et al. (2014), we test the robustness of our weighting scheme with respect to the considered applications by excluding the top 5 per cent of the objects that are given the highest weights.

The bias in the critical surface density is robust against the exclusion of the highest weighted objects for a magnitude cut at MAG_AUTO $i' < 23.5$ as shown in Fig. A1. The results improve

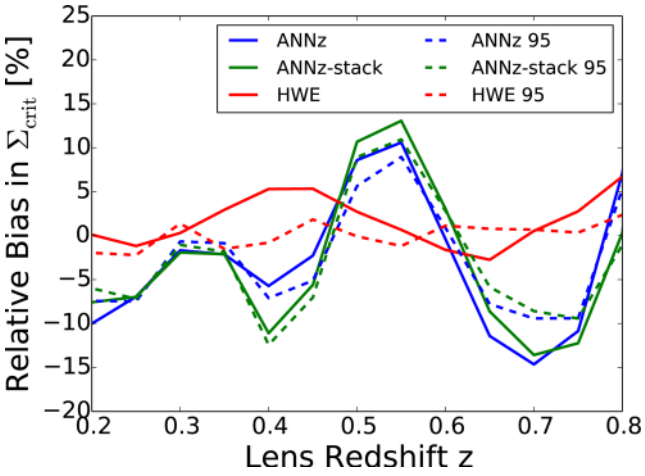

Figure A1. Relative bias in the mean critical surface density (equation 39) for different lens redshifts obtained using different estimates for the sample PDF. We show the relative biases obtained for the weighted data set cut at MAG_AUTO $i' < 23.5$ in solid lines, and the corresponding results with the 5 per cent highest weighted objects removed in dashed lines.

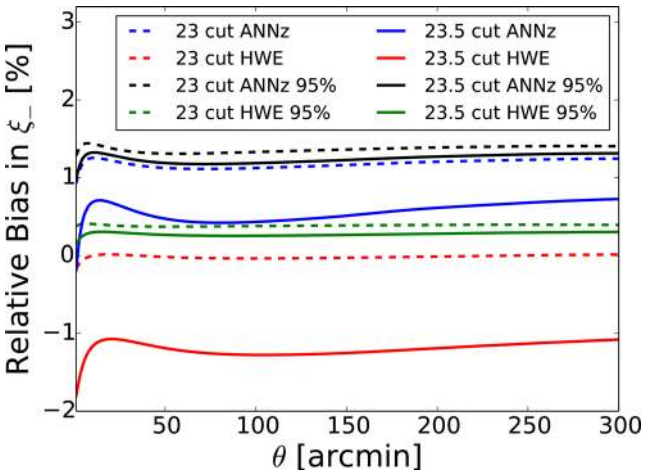

Figure A2. Relative bias in the shear correlation function estimate for ξ_- (equation 43) obtained using different estimates for the sample PDF. We show the relative biases obtained for the weighted data set cut at MAG_AUTO $i' < 23.5$ in solid lines and MAG_AUTO $i' < 23.0$ in dashed lines and the corresponding results with the 5 per cent highest weighted objects removed.

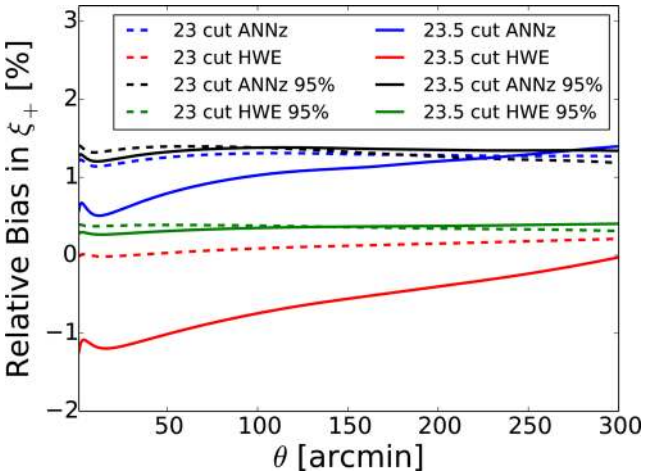

Figure A3. Relative bias in the shear correlation function estimate for ξ_+ (equation 43) obtained using different estimates for the sample PDF. We show the relative biases obtained for the weighted data set cut at MAG_AUTO $i' < 23.5$ in solid lines and MAG_AUTO $i' < 23.0$ in dashed lines and the corresponding results with the 5 per cent highest weighted objects removed.

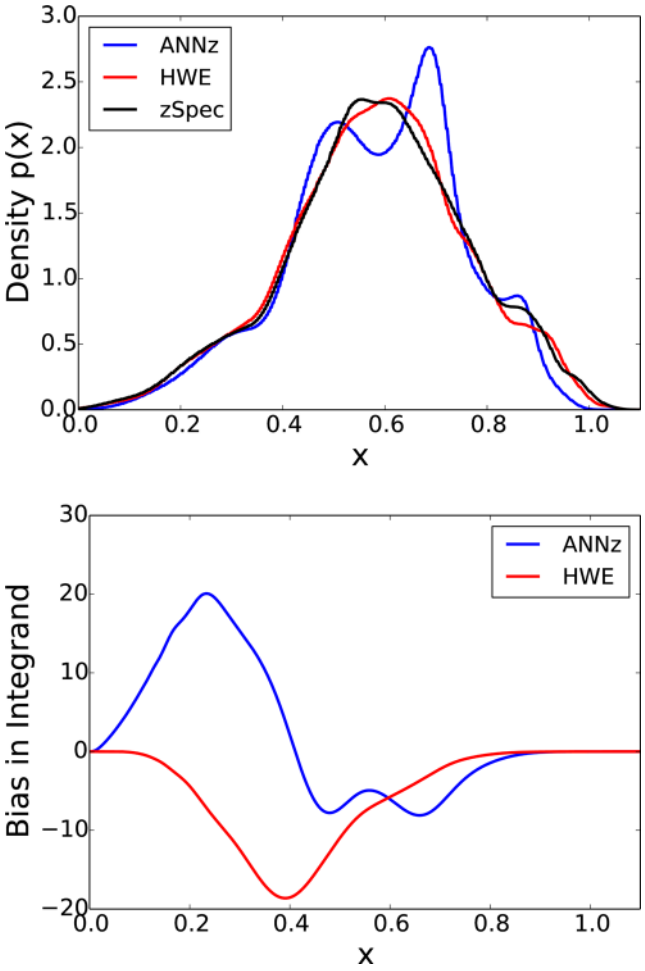

Figure A4. The sample PDFs for a cut at MAG_AUTO $i' < 23.5$ expressed in radial comoving coordinates x for the spectroscopic, ANNZ and HWE reconstructions (top panel). The bias (equation A1) obtained with the sample PDFs from ANNZ and the HWE for the example $l = 100$ (bottom panel).

for all algorithms if these objects are removed. The conclusions of the analysis, i.e. that the HWE leads to a lower bias compared with ANNZ, remain valid.

The analysis of the biases incurred in estimates of the cosmic shear correlation functions requires a more conservative cut at MAG_AUTO $i' < 23.0$, to be robust against the removal of a small number of highly weighted objects, as can be seen in Figs A2 and A3. For a magnitude cut at MAG_AUTO $i' < 23.5$, ANNZ[8] gives a better overall result compared with the HWE, while the opposite is true if the 5 per cent objects with the highest weight are left out.

Note that this is *not* because the $p(z)$ reconstruction of ANNZ is superior at faint magnitudes. Instead, this can be explained by considering the bias in the integrand in equation (42) with respect to the spectroscopic result given as

$$\text{Bias} = \frac{\ell}{2\pi} J_{0,4}(l\theta) \left(P_\kappa^{\text{phot}}(\ell) - P_\kappa^{\text{spec}}(\ell) \right) . \tag{A1}$$

As shown in Fig. A4, ANNZ both partly underestimates and overestimates the true spectroscopic integrand at different redshift values such that these two effects compensate each other. Since the lensing efficiency is dominated by the high-redshift tail of the stacked PDF, the peculiar shape of the ANNZ reconstruction in this range happens to outperform the otherwise superior HWE method. The shape of the high-redshift tail strongly depends on a small number of faint objects, which are given a high weight. Accordingly, this artefact is no longer present if the top 5 per cent of the objects with the highest weights are left out. For a more conservative cut at MAG_AUTO $i' < 23.0$, the analysis is no longer dominated by a few highly weighted objects at the faint end of our spectrophotometric catalogue, the ANNZ analysis does not outperform the HWE and the interpretation does not depend on the removal of the objects with the highest weights.

[8] The results for ANNZ-stack are very similar. Therefore, we do not show them here.

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