

KiDS + VIKING-450 cosmology with Bayesian hierarchical model redshift distributions

George T. Kyriacou,¹ Arrykrishna Mootoovaloo,^{1,2} Alan F. Heavens,^{1*} Andrew H. Jaffe,¹
Florent Leclercq³ and Konrad Kuijken⁴

¹Imperial Centre for Inference and Cosmology (ICIC), Imperial Astrophysics, Imperial College, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK

²Oxford Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

³CNRS & Sorbonne Université, UMR 7095, Institut d'Astrophysique de Paris, 98 Bis Boulevard Arago, F-75014 Paris, France

⁴Leiden Observatory, Leiden University, P.O. Box 9513, NL-2300 RA Leiden, the Netherlands

Accepted 2026 May 13. Received 2026 May 13; in original form 2024 September 10

ABSTRACT

Tomographic redshift distributions from photometric data are crucial ingredients in cosmic shear analysis, since they are required for the theoretical calculation of the signal based on the redshift distribution of the galaxies where the shear field is sampled. In this paper, we develop as a proof of concept Leistedt et al.'s template-based Bayesian Hierarchical Model framework into an application to weak lensing data by sampling the redshift distributions of the galaxies in the KiDS + VIKING-450 survey. We also use a principal component analysis to provide a set of representative templates drawn from a large superset. For computational tractability, subsets of 10^5 galaxies are chosen to determine the redshift distributions, and we test the sensitivity of the cosmological inference to the subset chosen, finding it to be subdominant compared to the statistical error. We marginalize over the inferred redshift distributions and find that the Bayesian method increases the clustering parameter compared with previous studies, alleviating the S_8 tension with Planck, where $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3} = 0.756 \pm 0.039$, assuming flat lambda cold dark matter. The tension with Planck for this survey is reduced from 2.3σ to 1.9σ . We also infer a value for the matter density, $\Omega_m = 0.31 \pm 0.10$.

Key words: gravitational lensing; weak – methods: data analysis – methods: statistical – techniques: photometric – cosmological parameters.

1 INTRODUCTION

Testing cosmological models through the inference of cosmological parameters is one of the main goals of large galaxy surveys. One observable is cosmic shear, the distortion of images of galaxies due to weak gravitational lensing along the line-of-sight by cosmic structure (for reviews, see e.g. M. Bartelmann & P. Schneider 2001; D. Munshi et al. 2008; M. Kilbinger 2015). Due to the large volumes required to obtain a measurable signal, this is usually accomplished with broad-band photometric surveys rather than spectroscopy, leading to some challenges in determining the redshift distribution of the sources, which is required to compute the cosmic shear or convergence power spectrum P_κ . This is sensitive particularly to the median redshift z_m (F. Bernardeau, L. van Waerbeke & Y. Mellier 1997; B. Jain & U. Seljak 1997; D. Huterer et al. 2006), but in detail the full $n(z)$ is required. Since cosmic shear increases with both source redshift and power spectrum amplitude, any bias in the median redshift results in an opposite bias in the inferred power spectrum amplitude. The demand for accurate $n(z)$ is exacerbated through

the use of tomographic bins, which are needed to increase the information content. Methods have been developed to estimate photometric redshifts from low spectral resolution data, such as template fitting (J. Walcher et al. 2011) and more specifically Bayesian Photometric Redshifts (BPZ; N. Benítez 2000). Template fitting is the matching of photometric data to that of redshifted spectra that have been multiplied by the appropriate filter response and a free amplitude, and integrated.

As accurate as these approaches may be in giving estimates and redshift posteriors for *individual* galaxies, they do not translate to developing usable and accurate $n(z)$ for the galaxies in the survey, which is what is used for cosmology inference. Previous surveys such as the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) and the Dark Energy Survey (DES) used the stacking of individual redshift posterior distributions (H. Hildebrandt et al. 2012; C. Bonnett et al. 2016). However, this approach magnifies the scatter of the underlying $n(z)$ and may not be the most suitable way to infer the distribution (N. Padmanabhan et al. 2005; C. Hikage et al. 2019). This may lead to biases at a level intolerable for current and future surveys (J. A. Newman et al. 2015; A. Choi et al. 2016; A. I. Malz & D. W. Hogg 2022; Euclid Collaboration 2024). Some methods do not give uncertainties in

* E-mail: a.heavens@imperial.ac.uk

the values of $n(z)$, which are then often treated by introducing wholesale shifts in the tomographic redshift distributions, which cannot reflect the full variability of the redshift distributions, so marginalization over the redshift distribution uncertainty is only approximate.

Current surveys have attempted to alleviate this problem through the use of a complementary spectroscopic survey. One such method is the weighted direct calibration, referred to as the direct redshift calibration method (DIR) (H. Hildebrandt et al. 2020; B. Stölzner et al. 2021; J. Ruiz-Zapatero et al. 2023). First proposed by M. Lima et al. (2008), the method takes accurate spectroscopic redshifts and calibrates the final $n(z)$ by re-weighting the spectroscopic redshift distribution so that the weighted spectroscopic survey matches the magnitude and colour distributions of the photometric survey. This method has been shown to be both accurate and gives uncertainties, through the use of bootstrap samples of said surveys, in their estimations. Unfortunately, this accuracy relies on a spectroscopic survey that covers the full range of magnitudes, redshifts and types of galaxies in the photometric sample, a luxury one will not have as photometric surveys become deeper in redshift and broader in scope (LSST Science Collaboration 2009; Euclid Collaboration 2019) as noted by S. Joudaki et al. (2020). Furthermore, uncertainties are produced but samples cannot be taken in a strictly Bayesian sense.

Cosmic shear analysis from the first year data of the Hyper Suprime-Cam (HSC; C. Hikage et al. 2019) used a similar approach to DIR for obtaining redshift distributions, the main difference being that overlapping spectroscopic surveys (O. Le Fèvre et al. 2013; J. Liske et al. 2015) do not have galaxy numbers large enough for the reweighting process to represent accurately properties of the photometric sources. As an alternative, the HSC survey uses the COSMOS 30-band photo- z catalogue (O. Ilbert et al. 2009) – a significantly larger number of bands than KV450, at the expense of an extra source of variability in the errors in the COSMOS photo- z s.

An alternative technique for inferring the $n(z)$ redshift distributions is the cross-correlation (CC) method developed by J. A. Newman (2008), with which we will also make comparisons. This method uses the CC functions between photometric and spectroscopic objects to finding the photometric redshift distributions.

An increasingly widespread technique is self-organizing maps (SOMs; R. Buchs et al. 2019; A. H. Wright et al. 2020), where colours are mapped to cells, and spectroscopic data used to assign redshifts to cells. Recently, M. M. Rau et al. (2023) developed a method to draw posterior samples of redshift distributions for the HSC Y3 analysis. In the DES Y3 analysis, the SOM method was used to generate samples of the redshift distributions and these were then used to obtain cosmological constraints from galaxy clustering and weak lensing (J. Myles et al. 2021; T. M. C. Abbott et al. 2022). A similar approach using SOM was used in the KiDS-1000 analysis (M. Asgari et al. 2021). A challenge of methods that rely on spectroscopic data is covariate shift – mismatched spectroscopic and photometric samples, for which care needs to be taken, such as in the weighting scheme of StratLearn (M. Autenrieth et al. 2024). This challenge motivates a method for obtaining Bayesian $n(z)$ distributions and uncertainties from photometric data alone.

In this paper, we adapt the method presented by B. Leistedt, D. J. Mortlock & H. V. Peiris (2016), who presented a Bayesian Hierarchical Method (BHM) with the main goal of inferring the redshift distribution of the population from which they are drawn

(see also J. Alsing et al. 2023; B. Leistedt et al. 2023). For weak lensing, current surveys do not account for source clustering, so we instead infer the redshifts of the background galaxies where the shear field is sampled. While individual galaxy samples of template type, magnitude and redshift are generated as an intermediate step in the Gibbs scheme of B. Leistedt et al. (2016), they are not retained or used as a direct inferential output in that framework. The innovation here is to retain these samples and use them to construct tomographic redshift histograms for each observed galaxy subsample, shifting the inferential target from the population distribution to the distribution of the specific galaxies used in the shear analysis. Hence in this work we develop a practical algorithm to sample the individual redshifts of the galaxies in the survey, rather than the population from which they are drawn. We also refine the choice of templates by developing a principal component analysis (PCA)-based method for grouping similar templates. These are empirical, largely based on observed galaxy spectra, in contrast to modelling the spectra with stellar population models; see e.g. B. Leistedt et al. (2023); J. Alsing et al. (2024).

This paper is organized as follows. In Section 2, we develop the method and test on simulated data in Section 3. The KV450 data are described in Section 4, and the redshift distributions derived. In Section 5, we present the cosmological results and discuss various challenges, and finally we present our conclusions in Section 6. Appendices present the simulation details and the (standard) background theory for completeness.

2 THE BAYESIAN HIERARCHICAL MODEL

The Bayesian approach used here is similar to that presented by B. Leistedt et al. (2016). We parametrize the galaxies in terms of a template t , redshift z , and (r -band) apparent magnitude m , with the ultimate goal being to sample the population distribution $p(t, z, m)$, finally marginalizing over the templates and magnitudes to produce samples of the redshift distributions for each tomographic bin. To do this we alternately sample (a) the individual galaxy t, z, m values, and (b), the parameters describing the population distribution of t, z, m , all conditioned on the measured galaxy fluxes $\{\hat{F}_b\}$.

For each galaxy, the measured flux \hat{F}_b in a photometric band labelled by b is assumed to have a Gaussian error distribution,

$$\hat{F}_b \sim \mathcal{N}(aF_b, \sigma_b^2), \quad (1)$$

where σ_b^2 is the flux variance of the galaxy in band b , and a is an amplitude. F_b is the noise-free flux, given by integrating the redshifted rest-frame template spectrum $L_t(\nu)$ multiplied by the filter response $W_b(\nu)$ over frequency,

$$F_b(t, z) = \frac{1+z}{4\pi D_L^2(z)} \int_0^\infty \frac{d\nu}{\nu} L_t(\nu) [\nu(1+z)] W_b(\nu), \quad (2)$$

where $D_L(z)$ is the luminosity distance. Note that the templates have an arbitrary normalization, so an amplitude a is introduced, and set by dividing the unnormalized flux (equation (2)) by the unnormalized r -band flux, and then multiplying F_b by the desired r -band flux, determined by m .

We discretize the three-dimensional (3D) distribution $p(t, z, m)$ in a 3D histogram of template, redshift and magnitude, labelled by indices i, j, k , with f_{ijk} representing the fraction of the population residing in each bin. Note that i is a genuinely discrete label, whereas j and k label continuous parameters which are binned (note that j represents a much finer binning

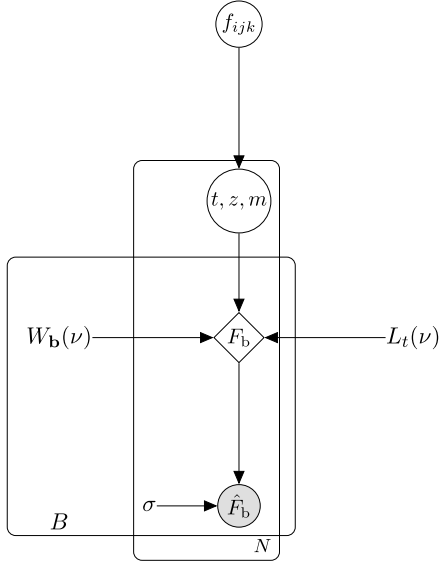


Figure 1. Hierarchical forward model. f_{ijk} is the fraction of galaxies in bins of template t , redshift z and magnitude m , labelled by i, j, k respectively. t, z, m values are drawn for all N galaxies, from the population specified by f_{ijk} , and noise-free fluxes F_b computed for the $B = 9$ photometric bands $b = 1 \dots B$. The bands are defined by transmissions $W_b(\nu)$, and the templates by $L_t(\nu)$. The observed fluxes \hat{F}_b include flux errors given by σ . Empty circles indicate quantities drawn from probability distributions, the diamond shows a deterministic relation, the filled circle represents (fixed) observed data, and quantities on their own are considered known precisely. The plates labelled by B and N indicate multiple instances. This can be applied to the whole sample or a subsample assigned (e.g. by BPZ) to a tomographic bin.

than that of the tomographic redshift distribution of interest; for reference, the binning in m is from $m = 19$ to $m = 25$ in steps of 0.1, and in z , steps of 0.05 up to $z = 3$). The survey population 3D histogram then approximates the joint distribution $p(t, z, m)$, integrated over the bins of redshift and magnitude:

$$f_{ijk} = \int_{z_{j,\min}}^{z_{j,\max}} \int_{m_{k,\min}}^{m_{k,\max}} p(t_i, z, m) dz dm. \quad (3)$$

Similarly, the conditional sampling distribution of the fluxes of galaxies in a bin is

$$p(\hat{F}_b | ijk) = \int_{z_{j,\min}}^{z_{j,\max}} \int_{m_{k,\min}}^{m_{k,\max}} p(\hat{F}_b | t_i, z, m) dz dm, \quad (4)$$

where the probability distribution for \hat{F}_b in the integrand is given by equation (1). The probability distribution for the set of measured fluxes of a galaxy, $\{\hat{F}_b\}$, is a product of the individual band likelihoods, assuming independence.

Each coefficient f_{ijk} is a population parameter to be inferred and the sampling captures any inter-bin correlations, which may be important for parameter inference. We assume a uniform prior for f_{ijk} , subject to constraints $0 \leq f_{ijk} \leq 1$ and $\sum_{ijk} f_{ijk} = 1$ (the Dirichlet distribution). The uniform prior in template index i is a prior on the galaxy types in the template set used, which we discuss in Section 2.2. The relevant directed acyclic graph is shown in Fig. 1.

We employ a two-step Gibbs sampler, alternating between sampling the set of individual galaxy (g) templates, redshifts and mag-

nitudes, $\{(t, z, m)_g\}$, and the population parameters $\{f_{ijk}\}$. The sampling differs slightly from the efficient implementation of B. Leistedt et al. (2016), which describes a Gibbs sampling implementation that marginalizes over the individual galaxy properties (t, z, m) , and samples the underlying population $\{f_{ijk}\}$ from which the galaxies were drawn. However, for use in tomographic weak lensing analysis, we require the properties not of the population from which the observed galaxies are drawn, but the redshifts of the galaxies that are actually in the sample, since these are the points where the shear field is sampled. Sampling both the coefficients $\{f_{ijk}\}$ and the individual properties $(t, z, m)_g$ for all galaxies allows us to use the latter samples, with each galaxy labelled by its tomographic bin, to construct the histograms for each tomographic bin separately.

We note that the redshift distributions inferred by this method differ conceptually from the population-level distributions that appear in the standard derivations of the lensing power spectra (e.g. M. Bartelmann & P. Schneider 2001), which are fixed properties of the survey selection function and binning, whereas our method samples the histograms of redshifts of specific galaxies in the survey. These two quantities differ in principle since our method includes the effect of source clustering on the average redshift distribution. However, we argue that for the purposes of a wide-field survey such as KV450, this distinction is negligible in practice for two reasons. First, these deviations are small, as the volumes in which we sample the redshift distributions are large, averaged over the several hundred square degrees of the survey, in bins that are of order 50 Mpc thick, and this is a completely subdominant effect. Secondly, source clustering is a small effect even for much higher precision surveys such as Euclid (L. Linke et al. 2025). The distinction between the two inferential targets is therefore not of practical significance for the present analysis, though we acknowledge it as an interesting conceptual question and an important direction for future work as survey volumes and statistical precision increase.

2.1 Tomographic bin sampling

For each sample (which is a full histogram and set of properties for all observed galaxies), we finally allocate each galaxy to its predetermined (by BPZ) tomographic bin and counting these gives a sample of the numbers, $n_{ijk,\text{bin}}$, from which we marginalize over template i and magnitude k to create a histogram of the redshifts. This produces samples of $n(z)_{\text{bin}}$ which can be marginalized over in the cosmological analysis; we also have access to the posterior distribution of the properties of each source which may separately be of interest.

Note that there are three independent uses of ‘binning’ in our model. The initially BPZ tomographic binning procedure is deterministic, and performed outside of and essentially separately from this hierarchical model. In practice, BPZ may use some of the same templates as in our set $\{t_i\}$, but it is important to understand that BPZ is used *only* as a mechanism to divide the sample into tomographic bins. Inevitable errors in BPZ photometric redshifts mean that the actual distribution of redshifts will extend beyond the nominal redshift limits of each bin, but this does not matter. The crucial requirement is that we have sufficient samples to represent the actual redshift distributions.

To set up the Gibbs sampling, where we alternately sample from (a) the i, j, k for each galaxy and (b) the population parameters f_{ijk} , we start by using Bayes theorem: assuming independent

galaxy observations, the joint distribution (see B. Leistedt et al. (2016) for full details) is

$$p(n_{ijk}, \{f_{ijk}\} | \{\hat{F}_b\}_g) \propto p(\{f_{ijk}\}) \prod_{g=1}^{N_{\text{gal}}} p(\{\hat{F}_b\}_g | ijk) p(ijk | \{f_{ijk}\}). \quad (5)$$

For given f_{ijk} , the Gibbs sampling step then samples i, j, k for each galaxy, with probability

$$p(ijk | \{f_{ijk}\}, \{\hat{F}_b\}_g) = f_{ijk} \times p(\{\hat{F}_b\}_g | ijk). \quad (6)$$

Note that no tomographic binning is done at this stage. Counting galaxies in each ijk then gives a sample n_{ijk} for the whole population of sources. Using this n_{ijk} one draws samples for f_{ijk} following a multinomial distribution:

$$p(\{f_{ijk}\} | \{n_{ijk}\}) = \frac{(N_{\text{gal}} + N_t N_z N_m - 1)! \delta_D (1 - \sum_{ijk} f_{ijk})}{\prod_{i=1}^{N_t} \prod_{j=1}^{N_z} \prod_{k=1}^{N_m} n_{ijk}!} \Theta(f_{ijk}) f_{ijk}^{n_{ijk}}. \quad (7)$$

So far, the approach described above follows B. Leistedt et al. (2016), whose main goal was to sample the underlying distribution f_{ijk} from which survey galaxies are drawn, given the selection effects. However, for a weak lensing analysis, we require the redshift distribution of the individual galaxies that occupy each tomographic bin of that survey, not the population from which they are drawn, which requires a different approach. Each galaxy has a label determined by BPZ, so the algorithm is simply to add up the number of galaxies in each bin ijk for each tomographic bin label for each realization. We then marginalize over template i and magnitude k to obtain a sample of each tomographic redshift distribution.

2.2 Template set

One of the key assumptions of this model is that all galaxies within the survey are represented closely enough by one of the templates in the set used in the analysis. Thus, one of the issues is that of template incompleteness; one requires a large enough set of templates that they cover the spectral range of the photometric data. This would suggest using a very large template set. Our prior is that each template is equally probable, so the template set determines the prior. For example, if a set contained six elliptical templates for each spiral, the prior would favour elliptical galaxies. This imbalance can propagate into the template posterior, if there is not enough data to override the prior.

We have constructed a template superset of around 300 galaxies from the BPZ source at <https://www.stsci.edu/~dcoe/BPZ/>. However, using so many templates makes the code slow. Therefore, we used smaller sets of templates which still capture the colours of the survey galaxies. In addition to choosing random sets of templates, we also created representative sets by using PCA (using SKLEARN) and k -means clustering in the 4D space of the largest four principal components, to divide the templates into five groups. We do not use templates constructed from the principal components themselves, but use PCA only to divide the templates into similar groups, from each of which we choose 10 templates at random, thus avoiding over-representation of similar templates which may be over-abundant in the complete template sample. This turned out to be an effective way to reduce the number of templates required to obtain stable results. The four components are shown in the upper panel of Fig. 2 in which

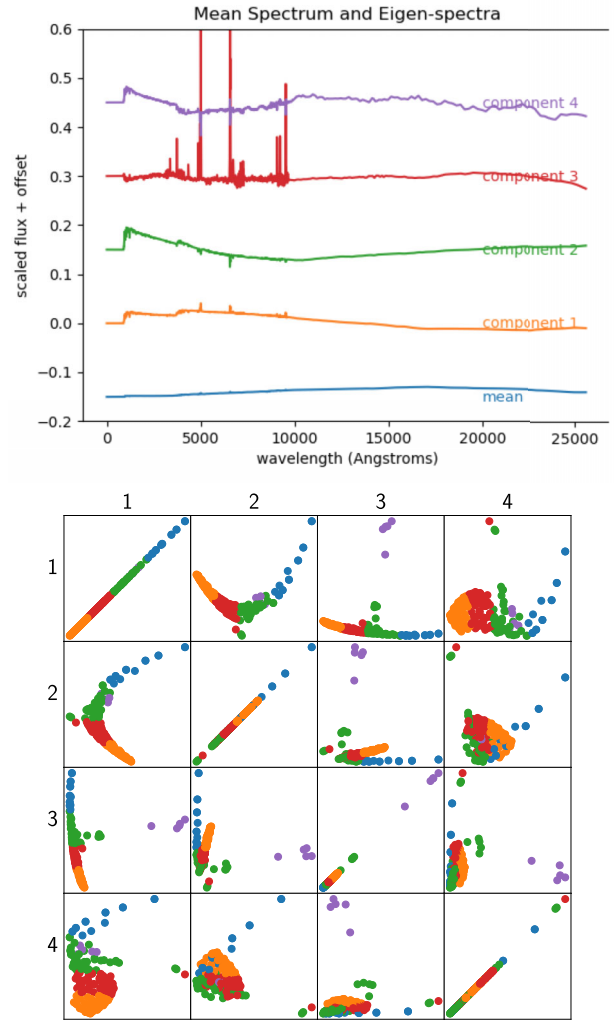


Figure 2. Top: Scatter plot of examples of the first four principal components extracted from a large template set. The flux has been scaled and offset for visualization purposes. Bottom: a 4×4 scatter plot of all templates plotted against component number. The colours indicate the five different sets where the k -means clustering algorithm placed the template.

obvious characteristics of certain templates, such as emission lines, can be seen. In the lower panel, we see the groupings in different colours. This device allows us to sample representative templates without biasing the results towards types that are over-represented in the rather heterogeneous complete template set. The method can be adapted to other template choices, such as those generated from theoretical models, as in the pop-cosmos Bayesian Hierarchical model programme (J. Alsing et al. 2023; S. Thorp et al. 2024; J. Alsing et al. 2024; S. Deger et al. 2025; S. Thorp et al. 2025), or a mixture of the two.

3 TESTS ON SIMULATIONS

To test the validity of the method and the new implementation, sets of simulated redshift distributions were created. We draw $N_{\text{gal}} = 2 \times 10^6$ r -band magnitudes from the r -band distribution of the KV450 survey. Types were drawn uniformly and independently of magnitude from the full set of over 300 templates, so

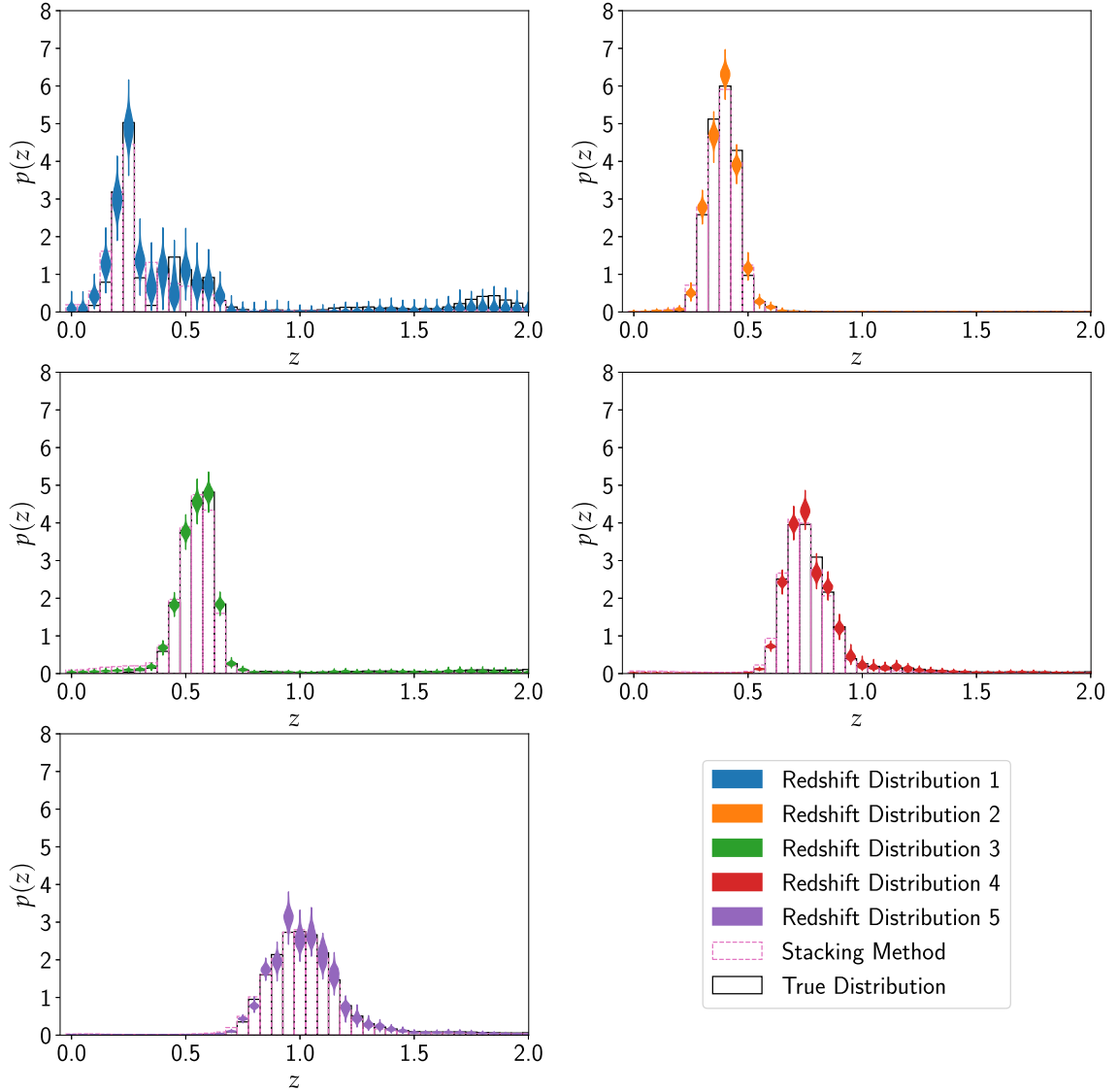


Figure 3. The tomographic redshift distributions generated using the simulated data. The redshift intervals for each tomographic distribution are as follows: $0.1 < z \leq 0.3$, $0.3 < z \leq 0.5$, $0.5 < z \leq 0.7$, $0.7 < z \leq 0.9$, and $0.9 < z \leq 1.2$. The shaded region on each bin (violin plot) indicates the uncertainty associated with the inferred heights of the tomographic redshift distribution. The solid black bars indicate the true distribution while the coloured dashed bars represent the stacked likelihood.

Table 1. The average error in median redshift and variance from the simulated data, using the two analysis methods, namely, the stacked likelihood (SL) and the BHM model of this paper. Note that the BHM performs better in almost every measure.

z_B range	BHM		Stacked	
	Δz_m	$\Delta \sigma_z^2$	Δz_m	$\Delta \sigma_z^2$
Full survey	0.004	-0.003	-0.011	0.035
$0.1 < z_B \leq 0.3$	0.003	0.015	0.004	0.213
$0.3 < z_B \leq 0.5$	-0.002	-0.0005	0.004	0.003
$0.5 < z_B \leq 0.7$	0.004	-0.003	0.014	0.076
$0.7 < z_B \leq 0.9$	0.001	-0.002	0.012	0.035
$0.9 < z_B \leq 1.2$	0.005	-0.003	0.004	0.035

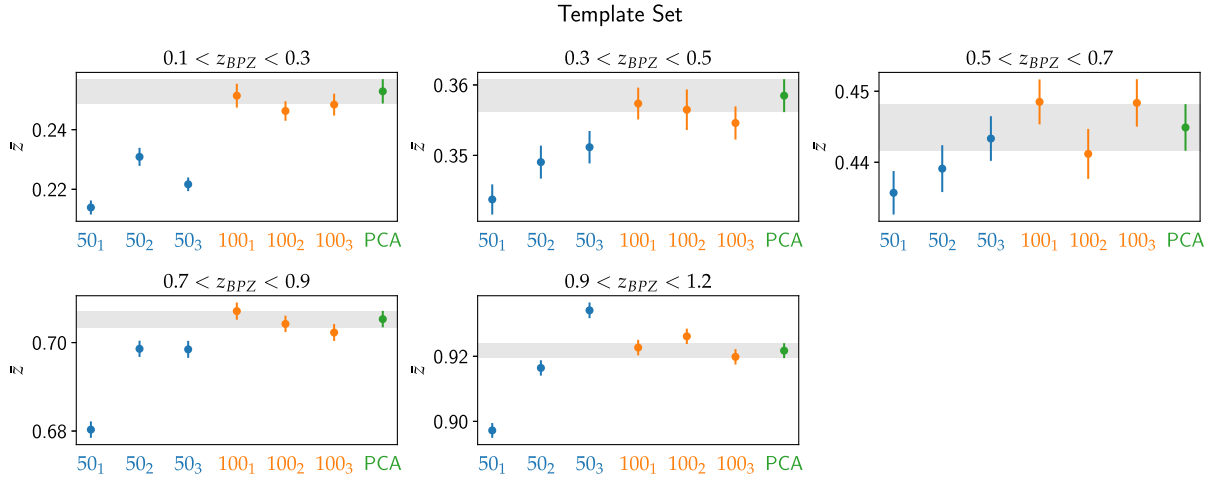


Figure 4. The median tomographic redshifts based on the redshift distributions of the KV450 data and their standard deviations, using 6 random template sets (3 containing 50 templates, denoted by 50_n and 3 containing 100 templates, denoted by 100_n). These sets are compared to the same result from the 50 templates selected in roughly equal numbers from each of the clusters of templates grouped by clustering in the first four principal components space (‘PCA’, in green). Note that 50 random templates gives widely varying median redshifts, whereas with 100 templates, the median redshifts are much more stable. However, the 50 more controlled representative PCA-selected templates yield a median redshift that agrees well with the 100-template random sets, so we can use this smaller, faster PCA set for KV450 analysis.

there is a level of model misspecification in that this is a wider set than the PCA template set used for analysis. The redshifts of the simulated sources were drawn from the following conditional distribution:

$$p(z|m_k) = \frac{4}{5}\mathcal{N}(\mu_{1,k}, \sigma_{1,k}^2) + \frac{1}{5}\mathcal{N}(\mu_{2,k}, \sigma_{2,k}^2) \quad (8)$$

and μ, σ are the means and variances of the Gaussian distributions. Lastly 5 per cent of the simulated population were given redshifts independently of magnitude and type following a uniform distribution, to test the method’s ability to assign correctly sources that are outliers from the main population. We emphasize that this is not meant to be an approximation to the real galaxy distribution, but is a test case of a non-trivial distribution that is not unimodal.

The redshift distribution parameters are both chosen to increase with magnitude bin (m_k) in order to replicate some of the behaviour seen in the BPZ priors. Explicitly, we take

$$\mu_{1,k} = 0.20 + 0.25 m_k, \quad (9)$$

$$\mu_{2,k} = 0.8 + 0.25 m_k, \quad (10)$$

$$\sigma_{1,k}^2 = 0.05 + 0.05 m_k, \quad (11)$$

$$\sigma_{2,k}^2 = 0.18 + 0.02 m_k. \quad (12)$$

when $m_k = [19, 20, 21, 22, 23, 24]$.

To simulate magnitudes and their errors and galaxy weights, we have used distributions from the KV450 data set. Details of the procedure are given in Appendix A.

To test the method, we compare the inferred redshift distributions with the true simulated distributions. A total of 3000 samples were collected on the t, z, m cell allocations of each source. Weights were then applied to each source before summing to obtain the redshift distributions. Fig. 3 compares the stacked likelihood with the BHM, showing good recovery of the input redshift distributions. Visually there is not too much difference, but a quantitative analysis shows that the BHM is a significant improvement over the stacked likelihood, which is expected given that it should provide a principled statistical solution.

We focus on the median redshifts z_m , since they are mostly responsible for the inferred value of σ_8 . We show the differences between the inferred medians and the truth in Table 1, shows that the BHM provides a more accurate value than the stacked likelihood. We also see that the variance is also recovered more accurately.

In more detail, the average χ^2 across the five tomographic bins is calculated to be $\bar{\chi}^2 = 102$ and $\bar{\chi}^2 = 18$ for the stacked likelihood and BHM model, respectively. Thus, the BHM model again shows significant improvement. It should also be noted that both approaches perform worse in the 5th bin (with the stacked likelihood and BHM having $\chi_{5\text{th}}^2 = 308$ and $\chi_{5\text{th}}^2 = 34$, respectively). A possible consequence of this is that the photometric data are less informative at higher redshifts where more filters would be of use.

4 KIDS + VIKING 450 ANALYSIS

4.1 Data

The data used for the cosmological parameter inference and demonstration of the application of this Bayesian hierarchical model were the KiDS + VIKING-450 data set¹ (A. Kannawadi et al. 2019; A. H. Wright et al. 2019). This is a combination of optical (KiDS) and near infrared (VIKING) photometric data. KiDS is a wide angle photometric survey which use the OmegaCAM camera within the Very Large Telescope (VLT) Survey Telescope (J. T. A. Jong et al. 2013). With a target area of $\sim 1500 \text{ deg}^2$, the survey intends to reveal cosmological information through the large scale distribution of matter. Although multiple data releases have taken place (J. T. A. Jong et al. 2015, 2017) KiDS-450 refers to DR3 of which 450 deg^2 of sources in the *ugri* bands have become available.

¹<https://kids.strw.leidenuniv.nl/DR3/kv450data.php>.

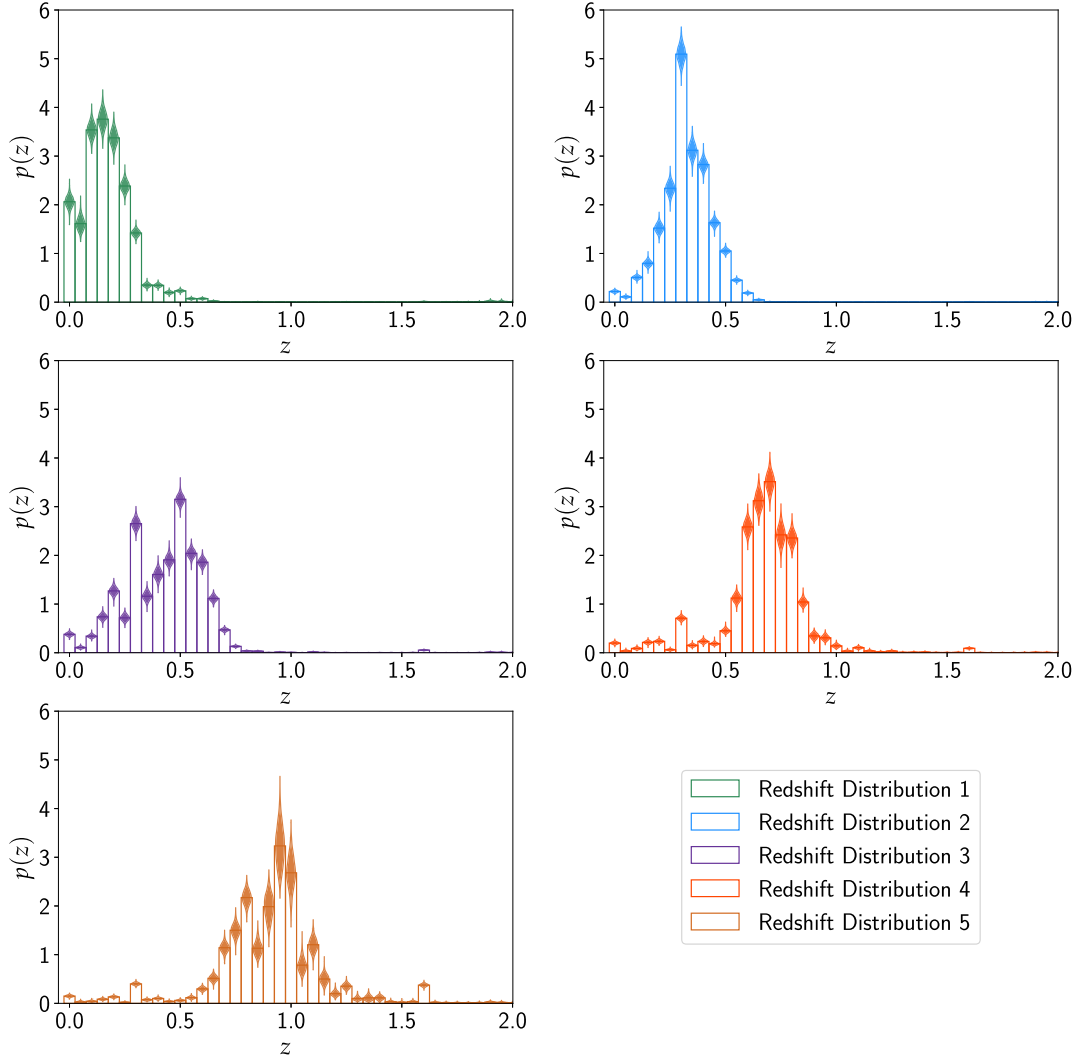


Figure 5. The KV450 tomographic redshift distributions generated using the first set of galaxy samples. The nominal (BPZ) redshift intervals for each tomographic distribution are as follows: $0.1 < z \leq 0.3$, $0.3 < z \leq 0.5$, $0.5 < z \leq 0.7$, $0.7 < z \leq 0.9$, and $0.9 < z \leq 1.2$. The shaded region on each bin (violin plot) indicates the uncertainty associated with the inferred heights of the tomographic redshift distributions. These are then propagated in the likelihood analysis when sampling the full posterior distribution of the cosmological and nuisance parameters.

Infrared data is collected from the VISTA Kilo-degree infrared Galaxy survey which used the Visible and Infrared CAMera (VIR-CAM) on ESO’s 4m VISTA telescope (B. P. Venemans et al. 2015). The planned target is for VIKING to cover $\sim 1350 \text{ deg}^2$ of sky across the *ZYJHK_s* IFR bands. Due to incompleteness in the coverage of VIKING, in order to ensure that the combined data set of KV450 is fully covered in all nine bands, it consequently has a reduced unmasked area of 341 deg^2 .

The specific data used were the Gaussian Aperture and PSF magnitudes (GAaP), a post processing stage of data reduction unchanged from previous releases (K. Kuijken 2008; K. Kuijken et al. 2015). Although other forms of multiband data are available the GAaP magnitude/fluxes are the default used by KiDS. As the details of the photometric data are more relevant to the Bayesian Hierarchical Model than for DIR, a further data filtering was undertaken, following M. Bilicki et al. (2018), to which we refer the reader for the rationale and discussion. Sources that do not include magnitude errors in every band are removed as well as those requiring $\text{GAaP_Flag_ugriZYJHKs} = 0$. This requirement reduced the KV450 area by ~ 5 per cent. Unlike the analysis used

by KiDS for both their DIR and BPZ for full redshift distribution and tomographic bin allocation, we have not used the GAaP magnitudes, rather the GAaP fluxes and errors have been used. These are available in the KiDS data set as ‘FLUX_GAaP_b’ and ‘FLUXERR_GAaP_b’, where ‘b’ labels the 9 filters. The fluxes also need to be corrected before analysis using offset and extinction correction terms:

$$F_{\text{corrected},b} = F_b \times 10^{\frac{2}{5}(\text{extinction}_b - \text{offset}_b)}, \quad (13)$$

where $F_{\text{corrected},b}$ is the final flux and F_b is the flux in the catalogue. A similar modification is made to the flux errors. The extinction term is taken directly from ‘EXTINCTION_b’ column of the data set. The magnitude offset is a per-tile correction. Sources are grouped by the tile of sky in which they were observed, this is found under ‘THELI_NAME’. Sources within the same tile have the same offset which is calculated as the median value of the following quantity:

$$\text{offset}_b = \text{MAG_GAaP_b} + 2.5 \log_{10}(\text{FLUX_GAaP_b}). \quad (14)$$

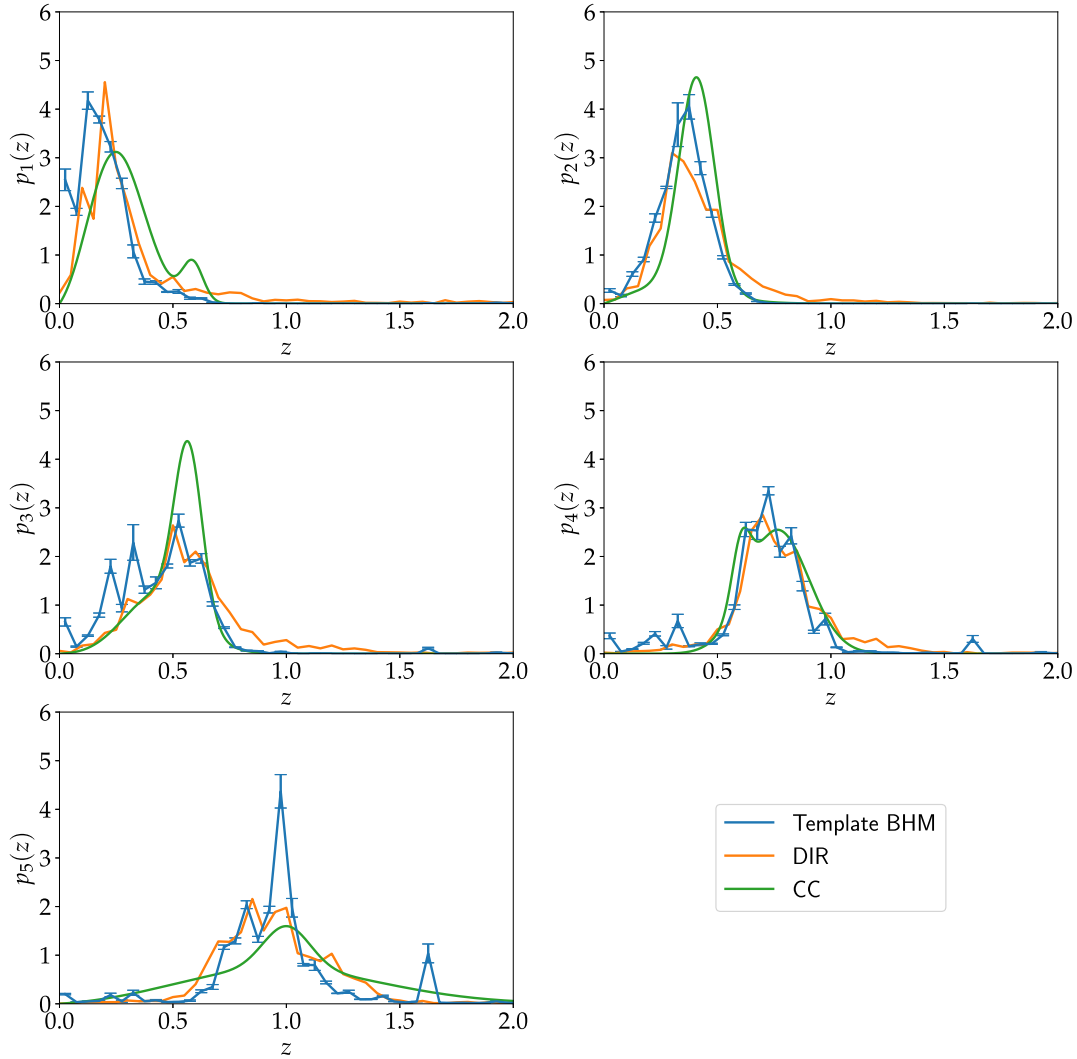


Figure 6. The KV450 BHM tomographic redshift distributions (blue) compared with the DIR and CC methods of H. Hildebrandt et al. (2020). The blue points show the variability in the mean $n(z)$, determined from the different galaxy subsamples. The others are shown without errors.

The median is used since there is typically a small number of undetected or unobserved sources, which don't have individual offsets. The tomographic bin allocation is determined by the BPZ values reported in the original data 'Z_b' and the weights are taken directly from 'recal_weight'. Fifty PCA templates were used for the analysis as described in Section 2.2.

4.2 Template and galaxy subsample sensitivity

To examine the issue of template sensitivity, 3 groups of randomly chosen 50 template sets and 3 groups of randomly chosen 100 template sets were used to produce redshift distributions for the 5 tomographic bins. The median redshift of the distributions was then compared to the median derived from 50 templates drawn using the PCA and k -means clustering technique presented in Section 2.2 to check for consistency. The median redshift was chosen as the comparison statistic due to its influence on the weak lensing power spectrum, since the main sensitivity of cosmological results to the redshift distributions is to the median values. For example, from Fig. 4, we can see that when choosing 50 random

templates, there is significant variability in the median redshift. However, the issue is no longer present when one chooses a set of 100, showing 100 templates to be an adequate number to achieve template insensitivity. However, when compared to the 50 templates present in the PCA analysis, one can see that the PCA 50-template set gives median redshifts that are in good agreement with the 100 template analyses, so for computational efficiency we use the PCA representative 50 template set for the cosmology analysis below.

Applying this method to all galaxies in the KV450 sample would be too slow, so we use a subset of galaxies to infer the redshift distributions. To test sensitivity to the subset used, we randomly pick 10^5 galaxies and infer their tomographic redshift distributions and repeat the experiment four times. We also generate a fifth set by making a 'combined set', concatenating all the four subsamples.

Taking the second half of the thinned MCMC chains generates 5000 and 20 000 $n(z)$ samples for the individual and combined sets, which are then used to sample the cosmological and nuisance model parameters' posterior. We find only minor variations

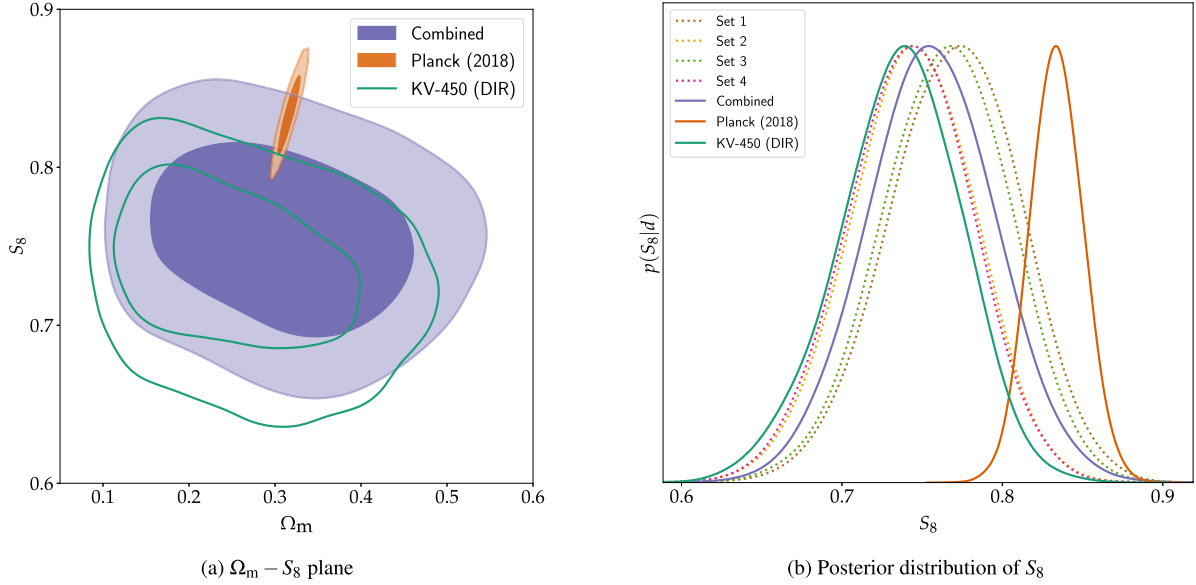


Figure 7. Panel (a): The marginalized posterior distributions (normalized to the same peak height) in the $\Omega_m - S_8$ plane for the KV450 analysis (in purple) and Planck (in orange). The inner and outer contours correspond to the 68 and 95 per cent credible intervals, respectively. Panel (b): The marginal posterior distribution for the S_8 parameter in various experiments. The broken curves correspond to the results obtained when different random subsets of 10^5 galaxies are used to determine $n(z)$. The purple curve shows the results when the sets are combined. The distribution of S_8 , as obtained, by the Planck Collaboration VI (2020), is also plotted in orange.

in the values of the parameters inferred, at the level of 0.01 in S_8 (see Fig. 7), so in Section 5 we report only the results obtained with the combined 20 000 $n(z)$ samples.

Finally, in Appendix C we show the colours of the template set along with the colours of a random subset of the KV450 galaxies. Given the large and variable errors on the photometry, it is difficult to make conclusive statements, but an expanded template set with wider colour coverage, particularly in the infrared, would be required for high-precision application of this method in future analyses.

4.3 KV450 redshift distributions

In Fig. 5 we show a histogram of the combined samples of the tomographic redshift distributions for KV450, with shaded regions indicating the marginal posteriors of the histogram heights.

A comparison with other methods can be made by inspection of appendix C of H. Hildebrandt et al. (2020). We show a comparison with DIR and CC (after fitting with a Gaussian Mixture Model) in Fig. 6. Since the methods are different, we do not expect them to agree in detail, but a comparison with the CC technique is useful to make. If there is a positive CC of the spectroscopic sample with the photometric sample at a given redshift, it is difficult to see how that could arise without some photometric galaxies being at those redshifts, so we might be concerned if the inferred BHM $n(z)$ was zero there. Inspection of the CC curves of H. Hildebrandt et al. (2020) shows no evidence for non-zero CC redshifts where the BHM distributions vanish, except in the highest-redshift bin, where magnification could plausibly introduce correlations.

The main differences are the less prominent high-redshift tail in the Bayesian method, giving rise to the higher σ_8 inferred here.

5 COSMOLOGY RESULTS

We use standard weak lensing theory (see Appendix B), and include systematics in the same way as H. Hildebrandt et al. (2020) so that direct comparison can be made. We show in Fig. 7(a) our $S_8 - \Omega_m$ posterior using the combined set of $n(z)$, and compare with that obtained in Planck Collaboration VI (2020) and with the DIR KV450 analysis of H. Hildebrandt et al. (2020). In Fig. 7(b), we focus only on the 1D marginal posterior distribution of S_8 and compare the results obtained with $n(z)$ obtained from each galaxy subset (1, 2, 3, 4, and from analysis of the combined set of 4×10^5 galaxies) with the Planck posterior. As expected, the posterior due to the combined set is roughly centred around the other posterior distributions for KV450, with the results from redshift distributions based on each random set of 10^5 galaxies showing relatively small sample variance. The residual sample standard deviation of S_8 for the combined sample is approximately 0.006, or a subdominant 0.15σ , estimated approximately from the standard deviation of the MAP values divided by $\sqrt{4}$. This could be reduced with a larger galaxy sample. We find a higher value of $S_8 = 0.756^{+0.039}_{-0.039}$ compared to the original KV450 analysis, reducing the tension with Planck for this data set. The error bar is unchanged, as expected, since we use the same summary statistics and cosmology model, and the propagation of the uncertainties in the redshift distributions adds little to the overall error.

In Fig. 8, for completeness we show the full 1D and 2D marginal posterior distribution of all the cosmological and nuisance parameters. The inner and outer contours are plotted at 68 and 95 per cent credible intervals, respectively. For most parameters, a flat uniform prior is assumed except for δc and A_c for which Gaussian priors are assumed. The mean and 1σ credible intervals are also reported in Table 2, but note that the posteriors for most of the parameters apart from S_8 are prior-dominated.

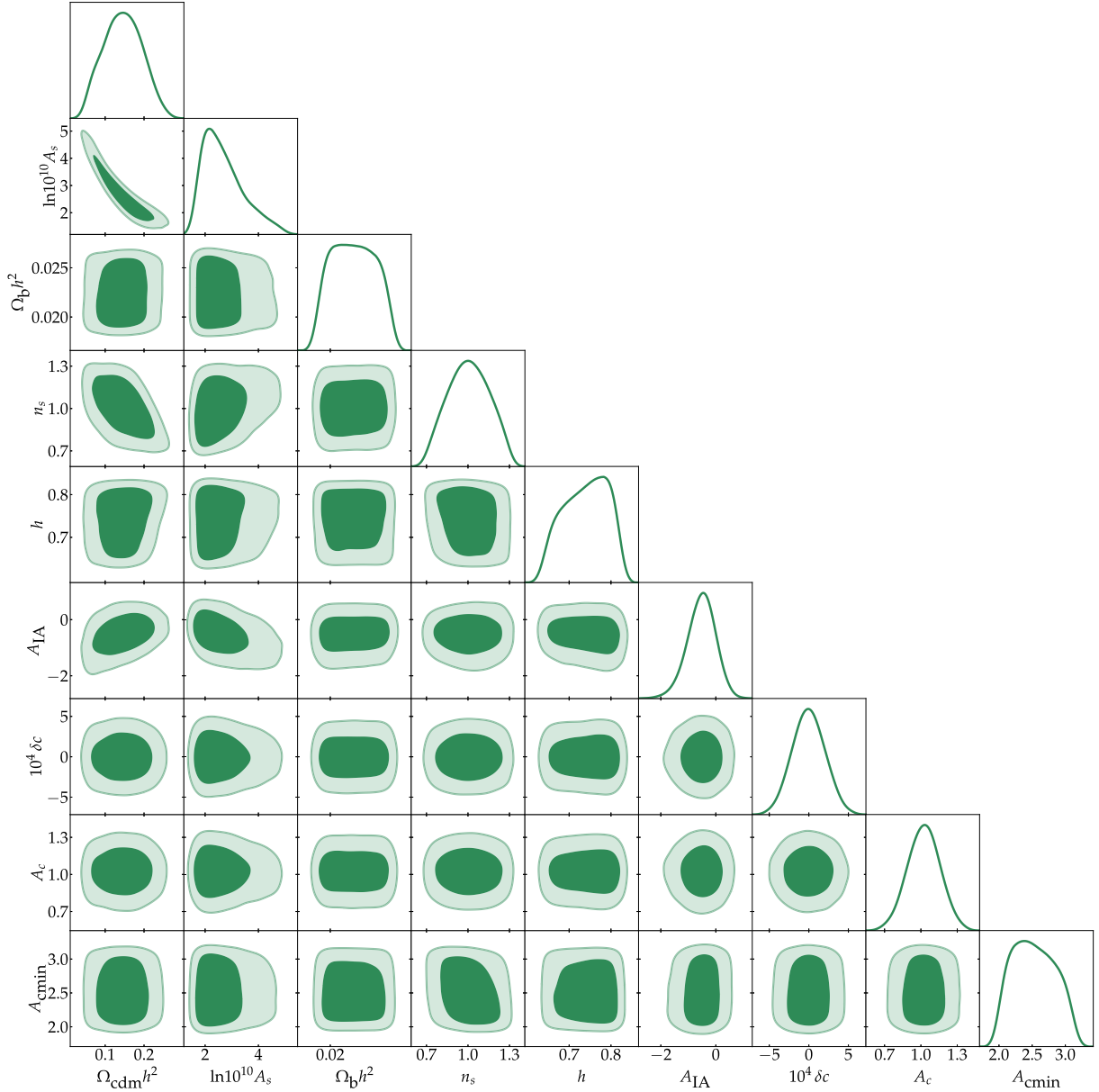


Figure 8. The full marginalized 1D and 2D posterior distribution of both sets of cosmological and nuisance parameters. The inner and outer contours correspond to the 68 and 95 per cent credible intervals. For all parameters, a top-hat prior is assumed except for δc and A_c , for which Gaussian priors are adopted. We refer the reader to Table 2 for further details on the choice of priors and the numerical results with their associated uncertainties.

While most of the results from weak lensing analyses tend to agree with each other, there remains the tension with the value of S_8 from Planck, although in this work, this tension is reduced, from 2.3σ to 1.9σ , with an unchanged error bar.

6 CONCLUSIONS

In this paper, we have adapted and developed the Bayesian method of B. Leistedt et al. (2016) into a practical procedure to infer redshift distributions of galaxies from photometric surveys and a set of templates, by explicitly sampling the redshifts of the galaxies observed in the survey, and applying the method to a real survey in the form of the KV450 data set. We have also introduced a PCA-based method to reduce the large set of templates to a manageable representative sample, and show that this is effective in

providing robust results with around 50 templates. We have investigated sensitivity to the template choices, and to the subset of 10^5 source galaxies used to determine the redshift distributions. Tests on simulated data show that the inferred redshift distributions are more accurate than those from stacked likelihoods.

We have applied this technique in a proof-of-concept study, inferring cosmological parameters from the shear two-point correlation functions from the KV450 data products. We keep the same summary statistics, changing only the redshift distributions, which were originally generated using DIR technique. We marginalize over the $n(z)$ by drawing at random from the samples of $n(z)$ at each likelihood evaluation. This is in contrast to the original KV450 analysis which introduced offsets in the $n(z)$ distributions along the redshift axis as nuisance parameters. Our analysis yields a slightly higher value of $S_8 = 0.756^{+0.039}_{-0.039}$

Table 2. Cosmological parameters from KV450 analysis with the new $n(z)$ redshift distributions determined by the Bayesian Hierarchical model. Note that, with the exception of S_8 , many of the formal errors are determined by the priors chosen.

Parameter	Symbol	Prior	Value
CDM density	$\Omega_{\text{cdm}}h^2$	$\mathcal{U}[0.01, 0.99]$	$0.15^{+0.05}_{-0.05}$
Scalar spectrum amplitude	$\ln 10^{10} A_s$	$\mathcal{U}[1.7, 5.0]$	$2.70^{+0.79}_{-0.74}$
Baryon density	$\Omega_b h^2$	$\mathcal{U}[0.019, 0.0226]$	$0.022^{+0.002}_{-0.002}$
Scalar spectral index	n_s	$\mathcal{U}[0.7, 1.3]$	$1.01^{+0.16}_{-0.15}$
Hubble parameter	h	$\mathcal{U}[0.64, 0.82]$	$0.74^{+0.06}_{-0.06}$
IA amplitude	A_{IA}	$\mathcal{U}[-6.0, 6.0]$	$-0.53^{+0.49}_{-0.49}$
Baryon feedback amplitude	c_{min}	$\mathcal{U}[2.00, 3.13]$	$2.5^{+0.4}_{-0.3}$
Constant c -term offset	$10^4 \delta c$	$\mathcal{U}[-6.0, 6.0]$	$-0.03^{+2.0}_{-2.0}$
2D c -term amplitude	A_c	$\mathcal{U}[0.62, 1.40]$	$1.0^{+0.1}_{-0.1}$
Matter density	Ω_m	Derived	$0.31^{+0.10}_{-0.10}$
$S_8 = \sigma_8 \sqrt{\Omega_m/0.3}$	S_8	Derived	$0.756^{+0.039}_{-0.039}$

compared to the original KV450 analysis, which reports a value of $S_8 = 0.737^{+0.040}_{-0.036}$, reducing the tension with Planck from 2.3σ to 1.9σ .

The tension in S_8 between the weak lensing analyses and the Planck analysis is reduced with the latest KiDS Legacy results (A. H. Wright et al. 2025), with the main change coming from the change in the way the redshift distributions are calculated, highlighting the importance of $n(z)$ in cosmological inference. Without spectroscopic redshifts for all sources, all methods suffer from some limitation, such as incomplete spectroscopic coverage, reliance on theoretical models for galaxy spectra, or, in this case, sensitivity to templates, for which a set with broader colour coverage would be necessary for accurate determination for future catalogues. As perhaps emphasized by the KiDS Legacy results, it is helpful to have a range of techniques available in order to assess the robustness of the cosmological results to the approach taken to determine the redshift distributions. This method gives an alternative approach.

It is worth placing our approach in a broader context. A fully rigorous treatment of cosmic shear would account for the three-dimensional variation of source galaxy number density across the survey volume: the shear field is sampled not uniformly but in proportion to the local galaxy number density, which traces the underlying density field. Consistently accounting for this in both the estimator and the theory prediction would require weighting the lensing kernel to include the effect of the three-dimensional number density field, a level of treatment that is not typically used in current cosmic shear analysis. The standard approach uses a smooth, sky-averaged population-level average redshift distribution. Our approach infers the sample redshift distributions from the observed galaxies, and captures the sky-averaged variation due to clustering. However, the main effect of source clustering is in the angular directions (L. Linke et al. 2025), which is not accounted for in our approach (nor in current analyses), and is an important direction for future work, but note that it is not a major effect, even at the expected precision of Euclid (L. Linke et al. 2025).

ACKNOWLEDGEMENTS

We thank Boris Leistedt and Daniel Mortlock for informative discussion. We also thank Hendrik Hildebrandt and Fabian Köhlinger for the continuous advice regarding the workings of KiDS

data. GTK acknowledges funding through the Science and Technology Facilities Council (STFC). AM is supported through the LSST-DA Catalyst Fellowship project; this publication was thus made possible through the support of Grant 62192 from the John Templeton Foundation to LSST-DA. FL acknowledges financial support from the Agence Nationale de la Recherche (ANR) through grant INFOCW, under reference ANR-23-CE46-0006-01. This analysis is based on data products observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017, 177.A-3018, 179.A-2004, 298.A-5015, and on data products produced by the KiDS consortium.

DATA AVAILABILITY

The data underlying this article are available from the KiDS website at <https://kids.strw.leidenuniv.nl>. The templates used are from <https://www.stsci.edu/dcoe/BPZ/>.

REFERENCES

- Abbott T. M. C. et al., 2022, *Phys. Rev. D*, 105, 023520
 Alsing J., Peiris H., Mortlock D., Leja J., Leistedt B., 2023, *ApJS*, 264, 29
 Alsing J., Thorp S., Deger S., Peiris H. V., Leistedt B., Mortlock D., Leja J., 2024, *ApJS*, 274, 12
 Asgari M. et al., 2021, *A&A*, 645, A104
 Autenrieth M., Wright A. H., Trotta R., van Dyk D. A., Stenning D. C., Joachimi B., 2024, *MNRAS*, 534, 3808
 Bartelmann M., Schneider P., 2001, *Phys. Rep.*, 340, 291
 Benítez N., 2000, *ApJ*, 536, 571
 Bernardeau F., van Waerbeke L., Mellier Y., 1997, *A&A*, 322, 1
 Bilicki M. et al., 2018, *A&A*, 616, A69
 Blas D., Lesgourgues J., Tram T., 2011, *J. Cosmology Astropart. Phys.*, 2011, 034
 Bonnett C. et al., 2016, *Phys. Rev. D*, 94, 042005
 Bridle S., King L., 2007, *New J. Phys.*, 9, 444
 Brinckmann T., Lesgourgues J., 2019, *Phys. Dark Universe*, 24, 100260
 Buchs R. et al., 2019, *MNRAS*, 489, 820
 Choi A. et al., 2016, *MNRAS*, 463, 3737
 Deger S., Peiris H. V., Thorp S., Mortlock D. J., Jagwani G., Alsing J., Leistedt B., Leja J., 2025, preprint ([arXiv:2509.20430](https://arxiv.org/abs/2509.20430))
 Euclid Collaboration, 2019, *A&A*, 627, A59
 Euclid Collaboration, 2024, *A&A*, 684, A138
 Harnois-Déraps J., van Waerbeke L., Viola M., Heymans C., 2015, *MNRAS*, 450, 1212
 Hikage C. et al., 2019, *PASJ*, 71, 43

- Hildebrandt H. et al., 2012, *MNRAS*, 421, 2355
Hildebrandt H. et al., 2020, *A&A*, 633, A69
Hirata C. M., Seljak U., 2004, *Phys. Rev. D*, 70, 063526
Huterer D., Takada M., Bernstein G., Jain B., 2006, *MNRAS*, 366, 101
Ilbert O. et al., 2009, *ApJ*, 690, 1236
Jain B., Seljak U., 1997, *ApJ*, 484, 560
Joachimi B., Mandelbaum R., Abdalla F. B., Bridle S. L., 2011, *A&A*, 527, A26
de Jong J. T. A., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., 2013, *Exp. Astron.*, 35, 25
de Jong J. T. A. et al., 2015, *A&A*, 582, A62
de Jong J. T. A. et al., 2017, *A&A*, 604, A134
Joudaki S. et al., 2020, *A&A*, 638, L1
Kannawadi A. et al., 2019, *A&A*, 624, A92
Kilbinger M., 2015, *Rep. Prog. Phys.*, 78, 086901
Köhlinger F. et al., 2017, *MNRAS*, 471, 4412
Kuijken K., 2008, *A&A*, 482, 1053
Kuijken K. et al., 2015, *MNRAS*, 454, 3500
LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
Le Fèvre O. et al., 2013, *A&A*, 559, A14
Leistedt B., Mortlock D. J., Peiris H. V., 2016, *MNRAS*, 460, 4258
Leistedt B., Alsing J., Peiris H., Mortlock D., Leja J., 2023, *ApJS*, 264, 23
Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, *MNRAS*, 390, 118
Linke L. et al., 2025, *A&A*, 693, A210
Liske J. et al., 2015, *MNRAS*, 452, 2087
LoVerde M., Afshordi N., 2008, *Phys. Rev. D*, 78, 123506
Malz A. L., Hogg D. W., 2022, *ApJ*, 928, 127
Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, *MNRAS*, 454, 1958
Miller L., Kitching T. D., Heymans C., Heavens A. F., van Waerbeke L., 2007, *MNRAS*, 382, 315
Miller L. et al., 2013, *MNRAS*, 429, 2858
Munshi D., Valageas P., van Waerbeke L., Heavens A., 2008, *Phys. Rep.*, 462, 67
Myles J. et al., 2021, *MNRAS*, 505, 4249
Newman J. A., 2008, *ApJ*, 684, 88
Newman J. A. et al., 2015, *Astropart. Phys.*, 63, 81
Padmanabhan N. et al., 2005, *MNRAS*, 359, 237
Planck Collaboration VI, 2020, *A&A*, 641, A6
Rau M. M. et al., 2023, *MNRAS*, 524, 5109
Ruiz-Zapatero J., Hadzhiyska B., Alonso D., Ferreira P. G., García-García C., Mootoovaloo A., 2023, *MNRAS*, 522, 5037
Stözlner B., Joachimi B., Korn A., Hildebrandt H., Wright A. H., 2021, *A&A*, 650, A148
Thorp S., Alsing J., Peiris H. V., Deger S., Mortlock D. J., Leistedt B., Leja J., Loureiro A., 2024, *ApJ*, 975, 145
Thorp S. et al., 2025, *ApJ*, 993, 240
Venemans B. P. et al., 2015, *MNRAS*, 453, 2259
Walcher J., Groves B., Budavári T., Dale D., 2011, *Ap&SS*, 331, 1
Wright A. H. et al., 2019, *A&A*, 632, A34
Wright A. H., Hildebrandt H., van den Busch J. L., Heymans C., 2020, *A&A*, 637, A100
Wright A. H. et al., 2025, *A&A*, 703, A158

APPENDIX A: APPENDIX A: GENERATION OF SIMULATED DATA

For completeness, we include here the details of the generation of the simulated data used for testing.

The KV450 summary statistics are given a weight given by the shape measurement algorithm *lensfit* (L. Miller et al. 2007, 2013), so in order to simulate realistically the error distribution and *lensfit* weight distribution, a neural network (MLPRegressor from *sklearn*) was trained with one hidden layer of 100 nodes, with input fluxes and output errors or *lensfit* weights. The

network was then used to generate simulated flux errors and *lensfit* weights for the simulator. Based on the allocated true z , t , m , estimated fluxes in the 9 filters are then drawn directly from the model given by equations (1) and (2).

Simulated sources were allocated to tomographic bins using the KV450 BPZ pipeline, converting to magnitudes and magnitude errors according to $m_b = -2.5 \log_{10}(\hat{F}_b)$ and error $m_{e,b} = 1.086 \times \sigma_b / \hat{F}_b$. Sources have filters classed as ‘unobserved’ if they fall under one of two conditions, either $\sigma_b > \hat{F}_b$ or $m > b_{\text{lim}}$ where b_{lim} is the magnitude limit for a certain panel of sky. These limits were randomly allocated to sources. Although technically these sources are ‘undetected’, they are classed as unobserved. This methodology is replicated from the original KV450 BPZ results which found that using ‘unobserved’ over ‘undetected’ led to fewer extreme outliers, and is kept to keep consistency with the KiDS analysis, so we can use the same summary statistics. The classification of ‘unobserved’ means setting $m = 99$ and BPZ then treats the filter to have maximum error and no information can be gathered from it. BPZ also takes a M_0 input, which in KV450 data is labelled ‘MAG_AUTO’ and consists of a nonlinear combination of r and i magnitudes and their errors. The M_0 value is used in the BPZ prior. M_0 values were also simulated using a trained neural network where $m_{r,i}$ are the inputs and M_0 the output. Note that this only affects the assignment to tomographic bins; it is not the correct procedure for inference and is not used in the BHM.

The BPZ point estimate \hat{z}_{BPZ} was then used to allocate sources to 5 tomographic bins in the range $z = 0.1$ to 1.2 , with internal boundaries at (0.3, 0.5, 0.7, 0.9). Sources with estimated redshifts outside this range were discarded. The simulated surveys were analysed using the same PCA template set as the main survey.

APPENDIX B: APPENDIX B: WEAK LENSING DATA MODEL

In this Appendix, we review the weak lensing theory used, and define the summary statistics used in the analysis. Since it is relatively standard, we place it here. To effect a direct comparison, we replicate the KV450 analysis, but with a new set of redshift distributions, which we marginalize over. All other data products, such as the data vector and the covariance matrix are the same as those by H. Hildebrandt et al. (2020) to which we refer the reader for further details.

Cosmic shear summary statistics

Here we define the summary statistics used, which are the same as in the original KV450 analysis, to allow a direct comparison of results. KV450 employs the TREECORR code to estimate the shear correlation function between two tomographic bins α and β , that is,

$$\hat{\xi}_{\pm}^{\alpha\beta}(\theta) = \frac{\sum_{pq} w_p w_q \left[\epsilon_t^\alpha(x_p) \epsilon_t^\beta(x_q) \pm \epsilon_\times^\alpha(x_p) \epsilon_\times^\beta(x_q) \right]}{\sum_{pq} w_p w_q}, \quad (\text{B1})$$

where w is the *lensfit* weight, ϵ_t and ϵ_\times correspond to the tangential and cross ellipticities of a galaxy with reference to the vector $x_p - x_q$ between a pair of galaxies p and q . Nine logarithmically spaced bins are used within the interval $\theta \in [0'.5, 300']$. We follow the same procedure as H. Hildebrandt et al. (2020) and use the first seven bins for ξ_+ and last six bins for ξ_- . Since we have five tomographic redshift distributions, we have five auto and ten

cross pairs to tomographic redshift bins, the total number of data points is $(10 + 5) \times (7 + 6) = 195$,

and use the data covariance matrix of H. Hildebrandt et al. (2020).

In Section B, we cover the forward theoretical model and throughout this work, for consistency with the original KV450 analysis we assume a Gaussian likelihood for the data with a fixed covariance matrix.

Theory

To model the data, the two-point correlation functions between bins α and β are used as the forward cosmological model. It is strictly a linear combination of the E- and B-mode power spectra, that is,

$$\xi_{\pm}^{\alpha\beta}(\theta) = \int_0^\infty \frac{\ell}{2\pi} J_{0/4}(\ell\theta) [C_{\ell,\alpha\beta}^{\text{EE}} \pm C_{\ell,\alpha\beta}^{\text{BB}}], \quad (\text{B2})$$

where $J_{0/4}$ are Bessel functions of the first kind. In the absence of systematics, the B-mode power spectra are assumed to be zero and the main contribution is due to the convergence power spectra, $C_{\ell,\alpha\beta}^{\text{EE}} = C_{\ell,\alpha\beta}^{\kappa\kappa}$. Under the Limber approximation (M. LoVerde & N. Afshordi 2008) and assuming a flat universe, the EE power spectrum is given by

$$C_{\ell,\alpha\beta}^{\text{EE}} = \int_0^{X_H} d\chi \frac{w_\alpha(\chi) w_\beta(\chi)}{\chi^2} P_\delta(k; \chi), \quad (\text{B3})$$

where $P_\delta(k; \chi)$ is the non-linear matter power spectrum and χ the comoving distance. The weight function $w_\alpha(\chi)$ for tomographic bin α is

$$w_\alpha(\chi) = A\chi(1+z) \int_\chi^{X_H} d\chi' n_\alpha(\chi') \left(\frac{\chi' - \chi}{\chi'} \right) \quad (\text{B4})$$

and $A = \frac{3H_0^2 \Omega_m}{2c^2}$. The weight functions depend on the redshift distribution, $n_\alpha(z) dz = n_\alpha(\chi) d\chi$, which are here normalized such that

$$\int n(z) dz = 1.$$

In this analysis, we sample five cosmological parameters

$$\Phi = [\Omega_{\text{cdm}} h^2, \ln 10^{10} A_s, \Omega_b h^2, n_s, h],$$

where Ω_{cdm} is the CDM density parameter, A_s is the power spectrum amplitude, Ω_b is the baryon density (so $\Omega_m = \Omega_{\text{cdm}} + \Omega_b$), n_s is the scalar spectral index, and h is the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

We assume two massless neutrinos and one massive neutrino are assumed, with the latter fixed at $m_\nu = 0.06 \text{ eV}$. We also record two derived quantities, namely Ω_m and $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3}$. Sampling is done in `MontePython` (T. Brinckmann & J. Lesgourgues 2019), which also wraps `CLASS` (D. Blas, J. Lesgourgues & T. Tram 2011). Apart from the cosmological parameters, we also have to account for nuisance parameters, which we discuss briefly in the next section.

Systematics

The recovered C_ℓ is a biased tracer of the convergence power spectrum. The model which is commonly used in recent weak lensing analyses incorporates additional terms due to intrinsic alignments, II and GI, that is,

$$C_{\ell,\alpha\beta}^{\text{tot}} = C_{\ell,\alpha\beta}^{\kappa\kappa} + A_{\text{IA}}^2 C_{\ell,\alpha\beta}^{\text{II}} + A_{\text{IA}} C_{\ell,\alpha\beta}^{\text{GI}}. \quad (\text{B5})$$

The II term which arises due to correlation of ellipticities in the local environment, contributes positively to the total lensing signal, while the GI term arises due to correlation between intrinsic ellipticities of foreground and gravitational shear of background galaxies (C. M. Hirata & U. Seljak 2004). The II power spectrum is modelled as

$$C_{\ell,\alpha\beta}^{\text{II}} = \int_0^{X_H} d\chi \frac{n_\alpha(\chi) n_\beta(\chi)}{\chi^2} P_\delta(k; \chi) F^2(\chi), \quad (\text{B6})$$

where $F(\chi) = C_1 \rho_{\text{crit}} \Omega_m / D_+(\chi)$ (S. Bridle & L. King 2007; B. Joachimi et al. 2011). C_1 is a constant given by $5 \times 10^{-14} h^{-2} \text{ M}_\odot^{-1} \text{ Mpc}^3$, $D_+(\chi)$ is the linear growth factor normalized to unity today and ρ_{crit} is the critical density of the Universe today. The GI power spectrum is modelled as

$$C_{\ell,\alpha\beta}^{\text{GI}} = \int_0^{X_H} d\chi \frac{w_\alpha(\chi) n_\beta(\chi) + w_\beta(\chi) n_\alpha(\chi)}{\chi^2} P_\delta(k; \chi) F(k; \chi). \quad (\text{B7})$$

Note that A_{IA} in equation (B5) is a free amplitude parameter, which is marginalized over in the sampling procedure. Taking into account the two effects due to intrinsic alignment, the shear correlation function can be written as

$$\xi_{\pm}^{\text{tot}} = \xi_{\pm}^{\text{EE}} + A_{\text{IA}}^2 \xi_{\pm}^{\text{II}} + A_{\text{IA}} \xi_{\pm}^{\text{GI}}. \quad (\text{B8})$$

Another source of theoretical systematics is baryon feedback. It causes a reduction in the matter power spectrum at small scales. In the analysis of KiDS-450 data (F. Köhlinger et al. 2017), baryon feedback was modelled using a fitting formulae as derived by J. Harnois-Déraps et al. (2015). It is worth highlighting that these processes are not very well understood and hydrodynamical simulations were not strictly carried out with a view to matching them to weak lensing observations, but rather to understand the processes behind galaxy formation and feedback effects.

In this work, following the same approach as in H. Hildebrandt et al. (2020) for a comparative study, we use `HMCode` (A. J. Mead et al. 2015) where the halo bloating parameter, $\eta = 0.98 - 0.12 c_{\text{min}}$ changes the halo density profile. $c_{\text{min}} \in [2.0, 3.13]$ is a free parameter which is marginalized over in the likelihood analysis. A value of $c_{\text{min}} = 2.0$ corresponds to an extreme feedback model (AGN) while $c_{\text{min}} = 3.13$ corresponds to dark matter only.

In addition to the above nuisance parameters, two additional parameters, δc and A_c , were introduced to account for the additive shear bias and position dependent additive bias pattern, respectively. In short, the set of nuisance parameters, which we denote as Ψ is (see H. Hildebrandt et al. (2020) for full details):

$$\Psi = [A_{\text{IA}}, c_{\text{min}}, \delta c, A_c].$$

The new $n(z)$ method developed in this work is then coupled with the KV450 likelihood and the cosmological and nuisance parameters are inferred.

APPENDIX C: APPENDIX C: KV450 AND TEMPLATE COLOURS

In Fig. C1, we show the distribution of KV450 colours and the PCA template colours. The templates are shown at different redshifts roughly spanning the range of KV450 redshifts. The diagonals show the prior distributions of template colours at the indicated redshifts. The median and maximum KV450 galaxy errors are shown in each subfigure. One per cent of colour outlier galaxies are outside the limits for display purposes.

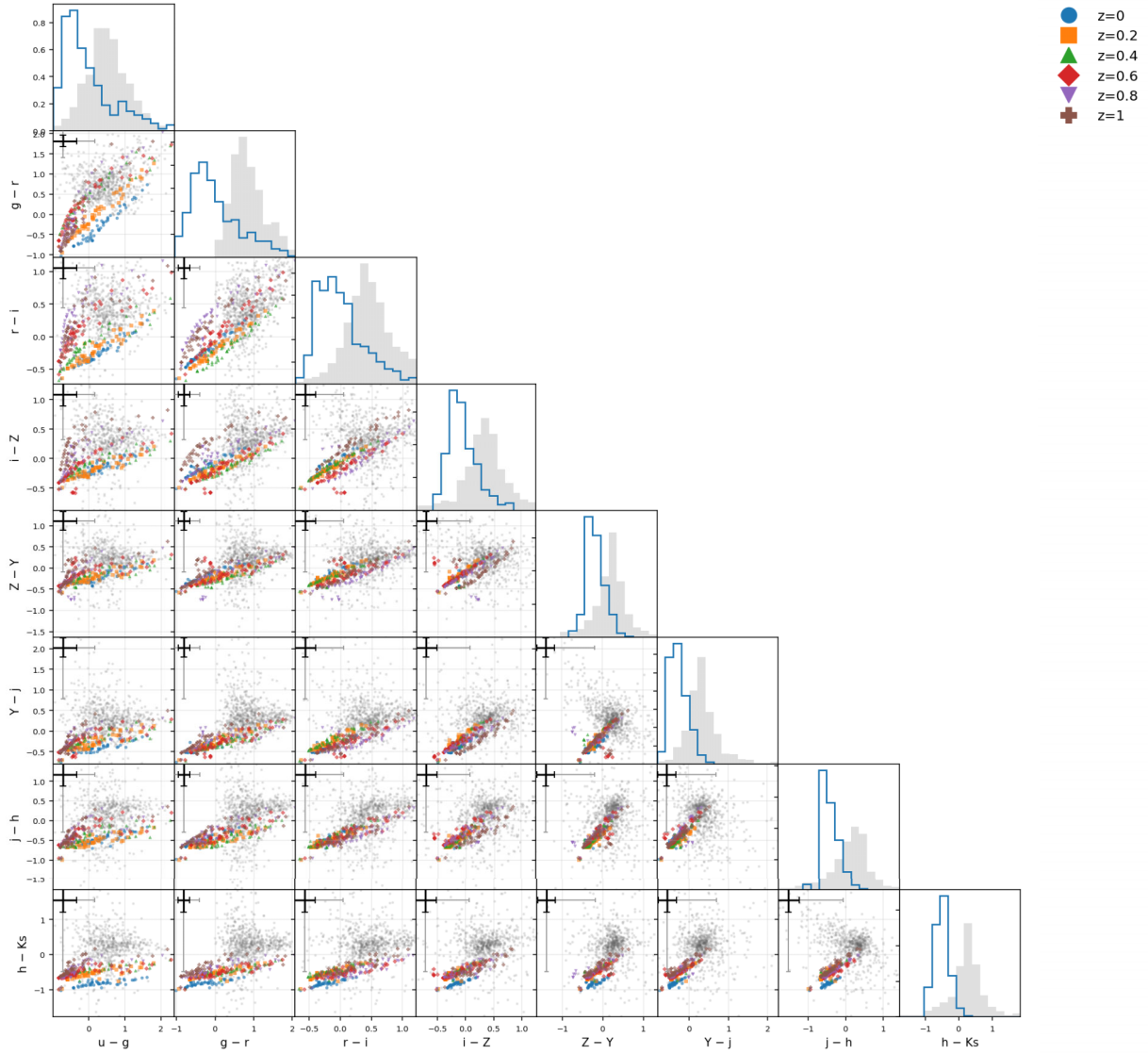


Figure C1. The distribution of colours of a random subset of KV450 galaxies (in grey) and the PCA templates. Median and maximum errors are shown by the error bars.

This paper has been typeset from a $\text{T}_{\text{E}}\text{X}/\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ file prepared by the author.