# Self-calibration and robust propagation of photometric redshift distribution uncertainties in weak gravitational lensing

B. Stölzner[1], B. Joachimi[1], A. Korn[1], H. Hildebrandt[2], and A. H. Wright[2]

[1] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
  e-mail: `benjamin.stolzner.18@ucl.ac.uk`
[2] Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing, 44780 Bochum, Germany

**ABSTRACT**

We present a method that accurately propagates residual uncertainties in photometric redshift distributions into the cosmological inference from weak lensing measurements. The redshift distributions of tomographic redshift bins are parameterised using a flexible modified Gaussian mixture model. We fit this model to pre-calibrated redshift distributions and implement an analytic marginalisation over the potentially several hundred redshift nuisance parameters in the weak lensing likelihood, which is demonstrated to accurately recover the cosmological posterior. By iteratively fitting cosmological and nuisance parameters arising from the redshift distribution model, we perform a self-calibration of the redshift distributions via the tomographic cosmic shear measurements. Our method is applied to KV450 data, which comprises a combination of the third data release of the Kilo-Degree Survey and the VISTA Kilo-Degree Infrared Galaxy Survey. We find constraints on cosmological parameters that are in very good agreement with the fiducial KV450 cosmic shear analysis and investigate the effects of the more flexible model on the self-calibrated redshift distributions. We observe posterior shifts in the medians of the five tomographic redshift distributions of up to $\Delta z \approx 0.02$, which are, however, degenerate with an observed decrease in the amplitude of intrinsic galaxy alignments of about 10%.

**Key words.** gravitational lensing: weak – cosmology: observations – galaxies: photometry – surveys – methods: analytical

## 1. Introduction

Weak gravitational lensing by the large-scale structure of the Universe, known as cosmic shear, is a powerful probe of cosmology. Rapid progress is being made in this field thanks to current and upcoming dedicated surveys such as the Dark Energy Survey (DES; Drlica-Wagner et al. 2018; Zuntz et al. 2018; Sevilla-Noarbe et al. 2021; Gatti et al. 2021), the Subaru Hyper Suprime-Cam (HSC; Aihara et al. 2018; Hikage et al. 2019), and the European Southern Observatory (ESO) Kilo-Degree Survey (KiDS; Kuijken et al. 2019; Asgari et al. 2021). These surveys allow us to test the predictions of the standard Lambda cold dark matter (ΛCDM) cosmological model by constraining the matter density and the amplitude of matter density fluctuations to unprecedented precision.

The main observables of weak lensing experiments are distortions of the ellipticities of background galaxies. Due to the weak signal and the impact of noise on the ellipticity measurement, this effect is measured statistically from large samples of galaxies. In order to model the theoretical prediction of the observed signal, an accurate calibration of the source redshift distribution is required. Given the large number of sources in a typical weak lensing survey, a complete spectroscopic redshift measurement is infeasible, and therefore the redshift is estimated from photometry (see Salvato et al. 2019 for a review).

Several methods of photometric redshift calibration have been developed, such as direct calibration with spectroscopic sub-samples that are, potentially after re-weighting, representative of the full sample (Lima et al. 2008; Bonnett et al. 2016; Hildebrandt et al. 2017) and angular cross-correlation clustering measurements with spectroscopic reference samples that overlap in redshift (e.g. Newman 2008; Matthews & Newman 2010; Ménard et al. 2013; McLeod et al. 2017). These methods can be merged using hierarchical Bayesian models that combine photometry measurements of individual galaxies and clustering measurements with tracer populations in a robust way (Sánchez & Bernstein 2019; Alarcon et al. 2020). Furthermore, the redshift distribution in weak lensing surveys can be self-calibrated to some extent from the data themselves (Zhang et al. 2010; Benjamin et al. 2013; Schaan et al. 2020). However, it is not only crucial to adopt a calibration method that estimates the true redshift distribution as precisely as possible, but also to choose a model that is flexible enough to describe the redshift distribution accurately. Such a model then allows us to propagate uncertainties in the redshift distribution, which arise from the calibration, into the actual cosmic shear analysis.

Examples of such flexible redshift distribution models are Gaussian mixture models (Hoyle & Rau 2019; Leistedt et al. 2019) and hierarchical logistic Gaussian processes (Rau et al. 2020), which are applied to calibrate redshift distributions of galaxy samples via cross-correlation clustering measurements with overlapping spectroscopic samples. Gaussian processes are non-parametric, that is, they are not limited by a functional form, and therefore they fulfil the condition of being able to accurately fit the redshift distribution. However, since the fit parameters of the Gaussian process are non-linear, implementing the Gaussian process in the weak lensing likelihood (with fit parameters acting as nuisance parameters) and subsequent marginalisation requires a carefully chosen kernel that needs to be adapted to the redshift distribution. As an alternative to Gaussian processes, the

redshift distribution can be parameterised using linear basis function models with a fixed number of parameters, so that we can readily apply an analytic marginalisation over nuisance parameters.

It is common to parameterise the uncertainty on the redshift distribution using a shift in the mean of the distribution (Hildebrandt et al. 2020, 2021; Hikage et al. 2019; Abbott et al. 2018; Hoyle et al. 2018), which captures the effect of uncertainties in the redshift distribution on the weak lensing analysis to the first order (Amara & Réfrégier 2007). However, with larger surveys and decreasing statistical uncertainties, the contribution of higher orders will become important (Wright et al. 2020a). Furthermore, this parameterisation has the disadvantage of introducing probability weights at negative redshift values. Therefore, it is particularly interesting to adopt redshift distribution models that capture arbitrary variations in the distribution.

In this paper we present such a flexible redshift distribution model with linear fit parameters, as well as a technique that provides an analytic marginalisation over nuisance parameters that originate from the redshift distribution calibration. We parameterise the redshift distribution of samples of galaxies as a 'comb', that is, a modified Gaussian mixture model with fixed, equidistant separation between components, identical variance, and a fixed number of components. The amplitudes of each Gaussian component serve as fit parameters in the redshift distribution calibration.

We implement this redshift distribution model in the weak lensing likelihood. Since the model is linear in the fit parameters, we can analytically marginalise over the fitted amplitudes. The advantage of this procedure is that we can use a large number of components to fit the redshift distribution, which gives the model enough flexibility to fit a potentially complex redshift distribution. At the same time, we do not increase the total number of free sampling parameters of the likelihood, so that it is still feasible to sample the likelihood without a significant increase in runtime. Additionally, the marginalisation method allows us to propagate correlations between all fit parameters of the redshift distribution into the likelihood. Thus, we incorporate the correlation between the redshift distributions of tomographic bins, which are induced by the calibration method, into the cosmic shear analysis. We then demonstrate our approach on the KiDS+VIKING (KV450) dataset comprising the ESO KiDS (Kuijken et al. 2015, 2019; de Jong et al. 2015, 2017) and the fully overlapping VISTA Kilo-Degree Infrared Galaxy Survey (VIKING; Edge et al. 2013) on a survey area of $450\,\mathrm{deg}^2$.

In order to allow the cosmic shear measurement to self-calibrate the redshift distribution, we adopt a two-step calibration method. First, we fit the comb model to the redshift histograms of Hildebrandt et al. (2020), which were calibrated with deep spectroscopic sub-samples. Second, we apply an iterative fitting method of both the cosmological and nuisance parameters originating from the redshift calibration. The best-fit nuisance parameters then represent a model of the redshift distribution that is calibrated with both deep spectroscopic catalogues and cosmic shear data. In contrast to the fiducial analysis of Hildebrandt et al. (2020), this method takes the full variability in the redshift distributions into account. When sampling the weak lensing likelihood, we then marginalise analytically over the set of best-fit nuisance parameters.

The paper is structured as follows: In Sect. 2 the redshift distribution model is described. The theoretical modelling of the cosmic shear signal with analytic marginalisation over nuisance parameters is presented in Sect. 3, and the cosmic shear self-calibration method of the redshift distributions is described

in Sect. 4. Results are presented in Sect. 5 and discussed in Sect. 6.

## 2. Redshift distribution model

We modelled the redshift distribution, $n^{(\alpha)}(z)$, of each tomographic bin, $\alpha$, as a comb, that is, a slightly modified Gaussian mixture with $N_z$ components per bin, with fixed, equidistant separation in redshift between the components, and with identical variance $\sigma_{\mathrm{comb}}^2$:

$$n^{(\alpha)}(z) := \sum_{i=1}^{N_z} A_i^\alpha \; \mathcal{K}\left(z; z_i, \sigma_{\mathrm{comb}}^2\right), \tag{1}$$

where the only free parameters to be fitted are the amplitudes $A_i^\alpha$. The model is linear in the amplitudes, which allows us to apply an analytic marginalisation over nuisance parameters when sampling the weak lensing likelihood. We chose to model the normalised 'teeth' of the comb as

$$\mathcal{K}\left(z; z_i, \sigma_{\mathrm{comb}}^2\right) = \frac{z}{N(z_i, \sigma_{\mathrm{comb}})} \; \exp\left\{-\frac{(z - z_i)^2}{2\sigma_{\mathrm{comb}}^2}\right\}, \tag{2}$$

with normalisation over the interval $[0, \infty]$:

$$N(z_i, \sigma) = \sqrt{\frac{\pi}{2}} \, z_i \, \sigma \, \mathrm{erfc}\left(-\frac{z_i}{\sqrt{2}\sigma}\right) + \sigma^2 \exp\left\{-\frac{z_i^2}{2\sigma^2}\right\}. \tag{3}$$

While this method does not depend on a particular choice of $\mathcal{K}$, this form has the advantage of ensuring $n^{(\alpha)}(0) = 0$. The redshift distribution is normalised so that

$$\sum_{i=1}^{N_z} A_i^\alpha = 1. \tag{4}$$

Using Eq. (4), we write the amplitude of the $N_z$th component in terms of the remaining $N_z - 1$ amplitudes:

$$A_N^\alpha = 1 - \sum_{i=1}^{N_z-1} A_i^\alpha. \tag{5}$$

Inserting Eq. (5) back into Eq. (1), we find

$$n^{(\alpha)}(z) = \mathcal{K}\left(z; z_{N_z}, \sigma_{\mathrm{comb}}^2\right) \tag{6}$$

$$+ \sum_{i=1}^{N_z-1} A_i^\alpha \left[\mathcal{K}\left(z; z_i, \sigma_{\mathrm{comb}}^2\right) - \mathcal{K}\left(z; z_{N_z}, \sigma_{\mathrm{comb}}^2\right)\right]$$

$$:= \sum_{i=1}^{N_z} A_i^\alpha n_i(z), \tag{7}$$

where we redefined the amplitude $A_{N_z}^\alpha \equiv 1$ and

$$n_i(z) = \mathcal{K}\left(z; z_i, \sigma_{\mathrm{comb}}^2\right) - \mathcal{K}\left(z; z_{N_z}, \sigma_{\mathrm{comb}}^2\right) \cdot \left(1 - \delta_{iN_z}\right). \tag{8}$$

Since the amplitudes should be positive, it is convenient to define

$$a_i^\alpha := \ln A_i^\alpha \tag{9}$$

as the actual fit parameters. The final result of the redshift calibration procedure with data $\boldsymbol{d}_{\mathrm{cal}}$ is then

$$\Pr\left(\{a_i^\alpha\} | \boldsymbol{d}_{\mathrm{cal}}\right) \approx \mathcal{N}\left(a_i^\alpha; a^{*\alpha}_i, \Sigma_{\mathrm{cal}}\right), \tag{10}$$

where the posterior is approximated by a multivariate Gaussian distribution with best-fit $a_i^{*\alpha}$ and covariance $\Sigma_{\rm cal}$.

The model developed in this section is a particular example of a linear basis function model, which is a class of models that involve linear combinations of fixed non-linear functions of the input variables (see for example Bishop 2016). While the linear dependence on the model parameters simplifies the analysis of this class of models, it requires the choice of an appropriate number of basis function components. In this work we determine the number of components by repeatedly performing a fit to the observed data with a varying number of components and selecting the model that provides the best fit to the observed data. Alternatively, a common approach in regression problems is to turn to Bayesian frameworks, which provide methods of determining the model complexity. A Bayesian framework requires the specification of a prior on the model parameters, which can work similarly to penalty terms in regularised least-squares regression. In particular, so-called shrinkage priors (see van Erp et al. 2019 for a review) are used to reduce the size of coefficient estimates by shrinking them towards zero. Variables that correspond to coefficients that are shrunk exactly to zero drop out of the model. Therefore, assuming a shrinkage prior provides a method to reduce the dimensionality of a given model. Furthermore, the linear model can be related to Gaussian process models when imposing Gaussian priors on the basis function amplitudes (see for example Bishop 2016).

## 3. Theoretical modelling of the cosmic shear signal

### 3.1. Weak lensing model and likelihood

It is standard practice to use two-point statistics of the gravitational shear as summary statistics in weak lensing studies. In this paper we employ the two-point shear correlation function between two tomographic bins. However, it is straightforward to apply the formalism to other two-point statistics, such as Complete Orthogonal Sets of E/B-Integrals (COSEBIs; Schneider et al. 2010) and band power estimates derived from the correlation functions (Schneider et al. 2002; Becker & Rozo 2016; van Uitert et al. 2018), since they are all linear functionals of the cosmic shear angular power spectrum.

The two-point correlation function between two tomographic bins, $\alpha$ and $\beta$, is defined via

$$\xi_\pm^{(\alpha\beta)}(\theta) = \frac{1}{2\pi} \int_0^\infty {\rm d}\ell\, \ell C_{\rm GG}^{(\alpha\beta)}(\ell) J_{0,4}(\ell\theta), \tag{11}$$

where $J_{0,4}(\ell\theta)$ are Bessel functions of the first kind and $C_{\rm GG}^{(\alpha\beta)}(\ell)$ is the angular weak lensing convergence power spectrum. Using the Limber approximation, the angular power spectrum reads (Kaiser 1992)

$$C_{\rm GG}^{(\alpha\beta)}(\ell) = \int_0^{\chi_{\rm H}} {\rm d}\chi \frac{q^{(\alpha)}(\chi)q^{(\beta)}(\chi)}{f_K^2(\chi)} P_\delta\left(\frac{\ell + 1/2}{f_K(\chi)}, \chi\right), \tag{12}$$

where $P_\delta$ is the matter power spectrum and $f_K$, $\chi$, and $\chi_{\rm H}$ are the co-moving angular diameter distance, the co-moving radial distance, and the co-moving horizon distance, respectively. The lensing efficiency is given by

$$q^{(\alpha)}(\chi) = \frac{3H_0^2\Omega_{\rm m}}{2c^2} \frac{f_K(\chi)}{a(\chi)} \int_\chi^{\chi_{\rm H}} {\rm d}\chi' n_\chi^{(\alpha)}(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')}, \tag{13}$$

with $a(\chi)$ being the scale factor and $n^\alpha(z){\rm d}z = n_\chi^{(\alpha)}(\chi){\rm d}\chi$ being the distribution of galaxies in redshift bin $\alpha$. Since $q^{(\alpha)}$ is a linear functional of the corresponding redshift distribution, it is

straightforward to extract the amplitudes of the redshift distribution model. We find

$$C_{\rm GG}^{(\alpha\beta)}(\ell) = \sum_{i,j=1}^{N_z} A_i^\alpha A_j^\beta \int_0^{\chi_{\rm H}} {\rm d}\chi \frac{q_i(\chi)q_j(\chi)}{f_K^2(\chi)} P_\delta\left(\frac{\ell + 1/2}{f_K(\chi)}, \chi\right) \tag{14}$$

$$:= \sum_{i,j=1}^{N_z} A_i^\alpha A_j^\beta\, c'_{ij}(\ell), \tag{15}$$

where we defined $q_i(\chi)$ as the lensing efficiency of the $i$th component of the redshift distribution model, as defined in Eq. (6). In the final equality we defined $c'_{ij}(\ell)$, which is the angular weak lensing power spectrum for two Gaussian mixture components at $z_i$ and $z_j$, as redshift distributions. Using Eq. (11), we can compute the two-point shear correlation function between two tomographic redshift bins via

$$\xi_{\rm GG}^{(\alpha\beta)}(\theta) = \sum_{i,j=1}^{N_z} A_i^\alpha A_j^\beta \int_0^\infty \frac{{\rm d}\ell\,\ell}{2\pi} J_{0/4}(\ell\theta)\, c'_{ij}(\ell) \tag{16}$$

$$:= \sum_{i,j=1}^{N_z} A_i^\alpha A_j^\beta x_\pm^{(ij)}(\theta), \tag{17}$$

where we defined the two-point correlation function of two Gaussian comb components, $x_\pm^{(ij)}(\theta)$.

The observed weak lensing signal does not correspond to $\xi_\pm^{(\alpha\beta)}$ directly, but is contaminated by correlations between intrinsic ellipticities of neighbouring galaxies, II, and correlations between intrinsic ellipticities of foreground galaxies and background galaxies, GI (Hirata & Seljak 2004):

$$\xi_\pm = \xi_{\rm GG} + \xi_{\rm II} + \xi_{\rm GI}. \tag{18}$$

We followed the method presented in Hildebrandt et al. (2017) to model the effects of intrinsic galaxy alignments using a 'non-linear linear' model (Hirata & Seljak 2004; Bridle & King 2007; Joachimi et al. 2011). The contributions of GI and II alignments to the two-point shear correlation function were calculated using Eq. (11) with the II and GI angular power spectra:

$$C_{\rm II}^{(\alpha\beta)} = \int {\rm d}\chi \frac{n^{(\alpha)}(\chi)n^{(\beta)}(\chi)}{f_K^2(\chi)} P_{\rm II}\left(\frac{\ell + 1/2}{f_K(\chi)}, \chi\right), \tag{19}$$

$$C_{\rm GI}^{(\alpha\beta)} = \int {\rm d}\chi \frac{q^{(\alpha)}(\chi)n^{(\beta)}(\chi) + n^{(\alpha)}(\chi)q^{(\beta)}(\chi)}{f_K^2(\chi)} P_{\rm GI}\left(\frac{\ell + 1/2}{f_K(\chi)}, \chi\right). \tag{20}$$

Again, we used the linear dependence on the redshift distribution to extract the amplitudes of the redshift distribution model in analogy to Eq. (17). The power spectra of intrinsic galaxy alignments, $P_{\rm II}$ and $P_{\rm GI}$, are related to the matter power spectrum $P_\delta$ via

$$P_{\rm II}(k, z) = F^2(z)P_\delta(k, z) \tag{21}$$

$$P_{\rm GI}(k, z) = F(z)P_\delta(k, z), \tag{22}$$

with

$$F(z) = -A_{\rm IA}C_1\rho_{\rm crit}\frac{\Omega_{\rm m}}{D_+(z)}. \tag{23}$$

Here, $D_+(z)$ denotes the linear growth factor, $\rho_{\rm crit}$ is the critical density at redshift $z = 0$, and $C_1$ is a fixed normalisation constant that is set such that $C_1\rho_{\rm crit} = 0.0134$ (Joachimi et al. 2011). The

redshift-independent amplitude of intrinsic alignments, $A_{IA}$, is left as the only free parameter, which is implemented as a sampled nuisance parameter in the weak lensing likelihood.

In general, the Gaussian log-likelihood is defined as

$$\mathcal{L} = -\frac{1}{2}\chi^2 + \text{const.} = \sum_{ij}(d_i - m_i)\,C_{ij}^{-1}\left(d_j - m_j\right) + \text{const.}, \quad (24)$$

where $d_i$ and $m_i$ denote the observed data and the model prediction, respectively, with the inverse covariance $C_{ij}^{-1}$. Thus, the weak lensing log-likelihood reads

$$\mathcal{L} = -\frac{1}{2}\sum_{l,\alpha,\beta,l',\alpha',\beta'} \Delta_l^{(\alpha\beta)}\, Z_{(l,\alpha,\beta)(l',\alpha',\beta')}\, \Delta_{l'}^{(\alpha'\beta')} + \text{const.}, \quad (25)$$

where the indices $\alpha$ and $\beta$ run over all unique combinations of tomographic redshift bins. The two-point correlation function is analysed in $\theta$-bins that are denoted by $l$. The inverse covariance, which is assumed to be cosmology independent, is given by $Z$ with elements $Z_{(l,\alpha,\beta)(l',\alpha',\beta')}$. We have defined

$$\Delta_l^{(\alpha\beta)} \equiv d_l^{(\alpha\beta)} - \sum_{i,j=1}^{N_z} A_i^\alpha A_j^\beta\, x_\pm^{(ij)}(\theta_l), \quad (26)$$

where $d_l$ denotes the element of the observed data vector in $\theta$-bin $l$ at angular scale $\theta_l$ and the indices $i$ and $j$ count over all possible combinations of components of the redshift distribution model. We note that all cosmology dependence is in the $x_\pm^{(ij)}(\theta)$.

### 3.2. Marginal likelihood

The goal is to analytically derive the likelihood of a weak lensing experiment, marginalised over the potentially large number of nuisance parameters originating from the redshift calibration. We denote the parameters over which we sample the posterior distribution by $\boldsymbol{p}_{sam}$ and parameters that we analytically marginalise over by $\boldsymbol{p}_{ana}$. In particular, $\boldsymbol{p}_{sam}$ includes cosmological parameter as well as nuisance parameters that account for intrinsic alignments, baryon feedback, and additive shear bias. The parameters $\boldsymbol{p}_{ana}$ are the collection of amplitude parameters $\{a_i^\alpha\}$. We obtain

$$\Pr(\boldsymbol{d}|\boldsymbol{p}_{sam}) = \int \mathrm{d}^{N_{ana}} p_{ana}\, \Pr(\boldsymbol{d}|\boldsymbol{p}_{sam},\boldsymbol{p}_{ana})\Pr(\boldsymbol{p}_{ana}), \quad (27)$$

where the prior on analytically marginalised parameters is given, in this case, by the posterior of the fit to the redshift distribution defined in Eq. (10).

In the following we assume that the overall weak lensing likelihood is Gaussian. Moreover, we apply a Laplace approximation to the posterior in the sub-space spanned by the redshift nuisance parameters, that is, we effectively assume the posterior to be well represented by a Gaussian in this regime. As shown by Taylor & Kitching (2010), we can always maximise the likelihood for non-Gaussian distributions so that the assumption of Gaussianity locally around the peak of the likelihood is justified. The marginalised log-likelihood,

$$\mathcal{L}_{marg} \equiv -2\ln\Pr(\boldsymbol{d}|\boldsymbol{p}_{sam}), \quad (28)$$

is then given by Bridle et al. (2002) and Taylor & Kitching (2010):

$$\mathcal{L}_{marg} = \mathcal{L}_{fid} - \frac{1}{2}\mathcal{L}'^\tau\left[\mathcal{L}'' + 2\Sigma_{cal}^{-1}\right]^{-1}\mathcal{L}'$$
$$+ \ln\det\left(\mathbb{I} + \frac{1}{2}\mathcal{L}''\Sigma_{cal}\right), \quad (29)$$

where $\mathbb{I}$ denotes the identity matrix and $\mathcal{L}_{fid}$ is the log-likelihood evaluated at the best fit of the nuisance parameters,

$$\mathcal{L}_{fid} \equiv -2\ln\Pr(\boldsymbol{d}|\boldsymbol{p}_{sam},\boldsymbol{p}_{ana}^*). \quad (30)$$

The vector of derivatives of the log-likelihood with respect to the nuisance parameters $a_i^\alpha$ is denoted by $\mathcal{L}'$, and the Hessian matrix of second derivatives with respect to the nuisance parameters is denoted by $\mathcal{L}''$. Analytic expressions of these quantities are given in Appendix A. All of these derivatives are to be evaluated at the best fit of the nuisance parameters $\boldsymbol{p}_{ana}^*$. The covariance matrix of nuisance parameters, originating from the calibration of the photometric redshift distribution, is given by $\Sigma_{cal}$. For the $N_{bin}$ tomographic bins used in the analysis, $N_{bin} \times N_z$ nuisance parameters are marginalised over (modulo those amplitudes fixed by the normalisation of the redshift distribution given in Eq. (4)).

To test the validity of the approximate marginalised likelihood, we could perform a short initial Markov chain Monte Carlo (MCMC) analysis of the full likelihood, as proposed by Taylor & Kitching (2010). This method would allow us to identify potential non-Gaussianities. Any non-Gaussian parameters could then be removed from the analytic marginalisation and instead numerically marginalised over via MCMC. The downside of this method is that the initial MCMC run is computationally expensive, especially when the number of nuisance parameters is large. As an alternative to a full MCMC, we could instead sample the likelihood with a reduced set of nuisance parameters in order to validate the approximations made in the marginalised likelihood. By selecting different sets of nuisance parameters, for example those describing the tails of the redshift distribution, we could probe the likelihood in different regions of the parameter space. This would allow us to gradually test the assumption that the posterior distribution in the sub-space spanned by the nuisance parameters can be approximated by a Gaussian.

## 4. Redshift distribution self-calibration

It is standard practice to include nuisance parameters $\delta z_i$ in the weak lensing likelihood (Abbott et al. 2018; Hikage et al. 2019; Asgari et al. 2021), which linearly shift the whole redshift distribution of each tomographic bin with a prior that is derived from the calibration with external datasets. When sampling the likelihood, this then allows for a self-calibration of the redshift distribution with cosmic shear measurements through a shift in the mean of the redshift distributions within the allowed prior range.

In this work we replaced the shifts in the mean with the amplitudes of the comb components. Thus, the model can accommodate more complex variations in the redshift distributions. This, however, comes at the cost of an increase in the number of nuisance parameters from 5 to $5 \times N_z$, where the number of Gaussian components per bin, $N_z$, is typically of order 30. Given the dimensionality of the new nuisance parameter space, a sampling of nuisance parameters via MCMC methods becomes computationally prohibitive. Thus, we marginalised analytically over the uncertainties on the fitted amplitudes, as outlined in Sect. 3.2. By doing so, we lose the ability, however, to self-calibrate the redshift distributions with cosmic shear data since the amplitudes no longer appear as free parameters in the likelihood. In order to retain the calibration of the redshift distribution with cosmic shear data, we performed an additional calibration step.

Our goal is to find the best fit in the combined parameter space of cosmological and nuisance parameters. Given the high
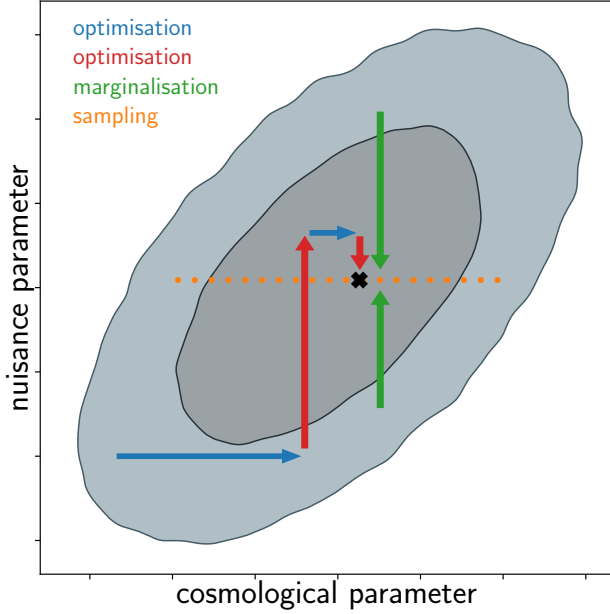
**Fig. 1.** Sketch of the iterative fitting method used to determine the best fit in the combined parameter space of cosmological and nuisance parameters. We alternate between optimising cosmological parameters (numerically; blue arrows), keeping nuisance parameters fixed, and optimising nuisance parameters (using Newton's method; red arrows), keeping cosmological parameters fixed. After several iterations we achieve convergence to the best fit in the combined parameter space. After optimising the likelihood, we set the amplitudes of the Gaussian comb to the best-fit parameters and proceed with sampling the likelihood in cosmological parameter space (dotted orange line) while analytically marginalising over nuisance parameters (green arrows).

dimensionality of the nuisance parameter space, we adopted an iterative method, which is illustrated in Fig. 1:

First, we fitted the Gaussian comb model, defined in Sect. 2, to pre-calibrated redshift distribution histograms. This was done by minimising

$$\chi^2 = \sum_{ij} \left( n_i^{\mathrm{data}} - n_i^{\mathrm{model}} \right) C_{ij}^{-1} \left( n_j^{\mathrm{data}} - n_j^{\mathrm{model}} \right), \tag{31}$$

where $n_i^{\mathrm{data}}$ and $n_i^{\mathrm{model}}$ are the observed and modelled histogram amplitudes in bin $i$, respectively, and $C_{ij}^{-1}$ denotes the inverse covariance matrix between the histogram bins in all five tomographic redshift bins. We estimated the uncertainties on the fit parameters by resampling the data vector using a multivariate Gaussian distribution, from which we calculated the covariance matrix $\Sigma_{\mathrm{cal}}$ of the fit parameters.

Second, we fixed the amplitudes of the Gaussian comb to the best-fit parameters found in the previous step. We then ran a non-linear optimiser to find the best-fit parameters $\boldsymbol{p}_{\mathrm{sam}}$ of the standard weak lensing likelihood conditioned on the best-fit parameters $\boldsymbol{p}_{\mathrm{ana}}$. This step is illustrated by the blue arrows in Fig. 1.

Third, for fixed parameters $\boldsymbol{p}_{\mathrm{sam}}$, the displacement from the peak of the likelihood in the sub-space of parameters $\boldsymbol{p}_{\mathrm{ana}}$ is given by (Taylor & Kitching 2010):

$$\delta\boldsymbol{p}_{\mathrm{ana}} = -\mathcal{L}' \left[ \mathcal{L}'' + 2\Sigma_{\mathrm{cal}}^{-1} \right]^{-1}, \tag{32}$$

assuming a Gaussian prior on the parameters $\boldsymbol{p}_{\mathrm{ana}}$. Fixing the parameters $\boldsymbol{p}_{\mathrm{sam}}$ to the ones found in step 2, we used Newton's method to minimise Eq. (32) so that the parameters $\boldsymbol{p}_{\mathrm{ana}}$ converge towards the peak of the likelihood. Since the constraints

on these parameters, which describe the redshift distribution, are dominated by the external priors through the original calibration, we anticipated the correction by the Newton step to be small[1]. The red arrows in Fig. 1 represent this calibration step.

By iterating over steps 2 and 3, we expected small corrections of both sets of parameters towards their best-fit values in the combined parameter space. The best-fit parameters $\boldsymbol{p}_{\mathrm{ana}}^*$ then represent the redshift distributions of each tomographic bin, calibrated using both spectroscopic catalogues and the actual cosmic shear measurements. After calibrating the redshift distributions, we set the amplitudes of the Gaussian comb in the weak lensing likelihood to the best-fit parameters and proceeded with the sampling of the likelihood in cosmological parameter space with pre-marginalised redshift distribution parameters. The sampling of the weak lensing likelihood is illustrated by the green arrows in Fig. 1.

While it is advantageous to infer the initial values for $\boldsymbol{p}_{\mathrm{ana}}$ from a prior redshift calibration, the optimisation scheme itself is expected to be valid for any initial values. We tested for the existence of local minima in the posterior distribution by performing the optimisation for various choices of the initial values. We find that each optimisation converges towards consistent parameter values, which shows that the optimisation method succeeds in finding the global minimum of the posterior distribution.

## 5. KV450 likelihood analysis

We used data from the ESO KiDS (Kuijken et al. 2015, 2019; de Jong et al. 2015, 2017) and the fully overlapping VIKING survey (Edge et al. 2013). This dataset, dubbed KV450, combines optical and near-infrared data on a survey area of 450 deg². The photometric redshift calibration is greatly improved compared to the earlier KiDS dataset (Hildebrandt et al. 2017) thanks to the addition of five near-infrared bands from VIKING that complement the four optical bands from KiDS. These additional bands improve the accuracy of photometric redshifts, which are used to define the tomographic bins. The fiducial technique of redshift calibration in KV450 utilised a weighted direct calibration, dubbed DIR, of five tomographic bins with deep spectroscopic catalogues. Uncertainties on the redshift distribution are estimated by a spatial bootstrapping method (Hildebrandt et al. 2020). The robustness of the photometric redshift calibration has been tested by excluding certain catalogues from the calibration sample as well as using alternative calibration samples. Additionally, the angular cross-correlation between KV450 galaxies and spectroscopic calibration samples has been studied as an alternative to the fiducial direct weighted calibration.

Our analysis is based on the fiducial KV450 cosmic shear analysis presented in Hildebrandt et al. (2020), in which the combined KiDS+VIKING dataset (Wright et al. 2019) is binned into five tomographic redshift bins based on their most probable Bayesian redshift, $z_B$, inferred with the photo-$z$ code BPZ (Benítez 2000). Four bins of width $\Delta z = 0.2$ in the range $0.1 < z_B \leq 0.9$ and a fifth bin with $0.9 < z_B \leq 1.2$ are defined and calibrated using the aforementioned direct calibration method. The estimated redshift distribution is then used to model the two-point shear correlation function, and constraints on

---

[1] For a likelihood that is close to Gaussian, we can find the maximum in one step. However, even if the initial redshift distribution is substantially different from the true underlying distribution, so that the likelihood at the initial values of the fit parameters is non-Gaussian, we can use Newton's method to iterate towards the peak (Taylor & Kitching 2010).

cosmological parameters are derived via sampling of the weak lensing likelihood.

Self-organising maps (SOMs) have recently been proposed as a method to mitigate systematic biases arising from the redshift calibration, by assigning galaxies to groups based on their photometry (Buchs et al. 2019; Wright et al. 2020a; Masters et al. 2015). This method allows samples of galaxies to be constrained such that they are fully represented by spectroscopic reference samples. It was recently applied to the KV450 (Wright et al. 2020b) and KiDS-1000 (Hildebrandt et al. 2021; Asgari et al. 2021) datasets. However, in those works the uncertainties on the redshift distributions are parameterised in terms of shifts in the mean of the redshift distributions with a prior that parameterises correlations between the redshift distributions of tomographic bins. This prior is inferred from simulations (Wright et al. 2020a; Hildebrandt et al. 2021; van den Busch et al. 2020). A spatial bootstrapping was not performed, and as such an estimate of the full covariance of the redshift distribution is not available. In this work we therefore reverted to the fiducial KV450 dataset, for which such an estimation of the full covariance of the redshift distribution is available, and we leave the application to more recent KiDS datasets to future work.

In this work we calibrated the redshift distribution by fitting the Gaussian comb model defined in Sect. 2 to the redshift distribution histograms of Hildebrandt et al. (2020). Additionally, we extended the KV450 likelihood code originally used in the fiducial analysis of Hildebrandt et al. (2020) by implementing the analytic marginalisation over nuisance parameters. The original likelihood is publicly available in the MONTEPYTHON[2] package (Audren et al. 2013; Brinckmann & Lesgourgues 2019). We sampled the likelihood in the MULTINEST[3] mode (Feroz et al. 2009, 2019) using the python wrapper PYMULTINEST[4] (Buchner et al. 2014). The matter power spectrum is estimated with the public code CLASS[5] (Blas et al. 2011) with non-linear corrections from HMCODE (Mead et al. 2015).

We adopted the cosmological model from Hildebrandt et al. (2020), that is, a flat $\Lambda$CDM cosmology with five parameters: $\omega_{CDM}$, $\omega_b$, $A_s$, $n_s$, and $h$. Additionally, the model includes four nuisance parameters that account for intrinsic alignments ($A_{IA}$), baryon feedback ($A_{bary}$), and additive shear bias ($\delta c$ and $A_c$). We note that, in contrast to the fiducial KV450 analysis, we did not include linear shifts $\delta z_i$ in the mean redshift in each tomographic bin as nuisance parameters since variations in the redshift distributions are taken into account by the amplitudes of the Gaussian comb model with analytic marginalisation over the corresponding uncertainties. Our choices of priors for the nine cosmological and nuisance parameters are identical to the ones used in Hildebrandt et al. (2020) and are reported in Table 1. Finally, we adopted the cosmic shear data from Hildebrandt et al. (2020), which consists of measurements of the two-point shear correlation functions between the five tomographic redshift bins and the corresponding analytic covariance matrix.

## 5.1. Redshift distribution self-calibration

The first step in the calibration of the KV450 redshift distribution was to fit the modified Gaussian mixture model, defined in

---

**Table 1.** Model parameters and their priors for the KV450 cosmic shear analysis, adopted from Hildebrandt et al. (2020).

| Parameter | Symbol | Prior |
|---|---|---|
| CDM density | $\omega_{CDM}$ | $[0.01, 0.99]$ |
| Scalar spectrum amplitude | $\ln(10^{10} A_s)$ | $[1.7, 5.0]$ |
| Baryon density | $\omega_b$ | $[0.019, 0.026]$ |
| Scalar spectral index | $n_s$ | $[0.7, 1.3]$ |
| Hubble parameter | $h$ | $[0.64, 0.82]$ |
| Intrinsic alignment amplitude | $A_{IA}$ | $[-6, 6]$ |
| Baryon feedback amplitude | $A_{bary}$ | $[2.00, 3.13]$ |
| Constant $c$-term offset | $\delta c$ | $0.0000 \pm 0.0002$ |
| 2D $c$-term amplitude | $A_c$ | $1.01 \pm 0.13$ |

**Notes.** The first five rows are cosmological parameters, and the remaining rows represent nuisance parameters. Brackets indicate top-hat priors, and values with errors indicate Gaussian priors. We note that, in contrast to Hildebrandt et al. (2020), linear shifts in the mean of the redshift distributions are excluded.

Sect. 2, to the redshift histograms of Hildebrandt et al. (2020), which are pre-calibrated with deep spectroscopic samples. The fit was done simultaneously for all five tomographic redshift bins in order to account for the correlations between bins. We performed the fit using two different input data histograms: the fiducial histograms with a bin width of $\Delta z = 0.05$ and histograms with a smaller bin width of $\Delta z = 0.025$. Although both histograms trace the same underlying redshift distribution of galaxies in each tomographic bin, their biases and variances will generally be different. By fitting the comb model to the two types of histograms, we tested how the redshift distribution calibration is affected by noise.

Moreover, we were free to choose the number of Gaussian components of the redshift distribution model and the variance of each component. In Appendix B we perform several tests, comparing different choices for the aforementioned free parameters, and address their impact on the cosmological analysis. In particular, as a rule of thumb for the width of the Gaussian component, $\sigma_{comb}$, we limited ourselves to values that ensure an overlap of two to three components at each point in redshift space. We find that the analysis is robust with respect to these choices. In this section we report our fiducial result using a model with $N_z = 30$ equidistant components between $0 \leq z \leq 2$ and a variance of $\sigma_{comb} = 0.067$, which is equal to the separation between the mean redshift of each component. This model was fitted to the redshift histograms with a bin width of $\Delta z = 0.05$.

The best-fit model is illustrated as blue curves in Fig. 2; the shaded regions indicate the uncertainties on the redshift distributions, which are derived from the diagonal elements of the covariance matrix of fit parameters. The correlation matrix of fit parameters is shown in Fig. 3.

We proceeded with a further calibration of the redshift distribution using the iterative fitting method of cosmological and nuisance parameters described in Sect. 4. The fit result after each step for the two parameters that are mostly constrained by the data, the intrinsic alignment amplitude $A_{IA}$, and the amplitude of matter density fluctuations, $S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$, are reported in Table 2. We note that in this analysis, $S_8$ is a derived parameter that is inferred from CLASS.
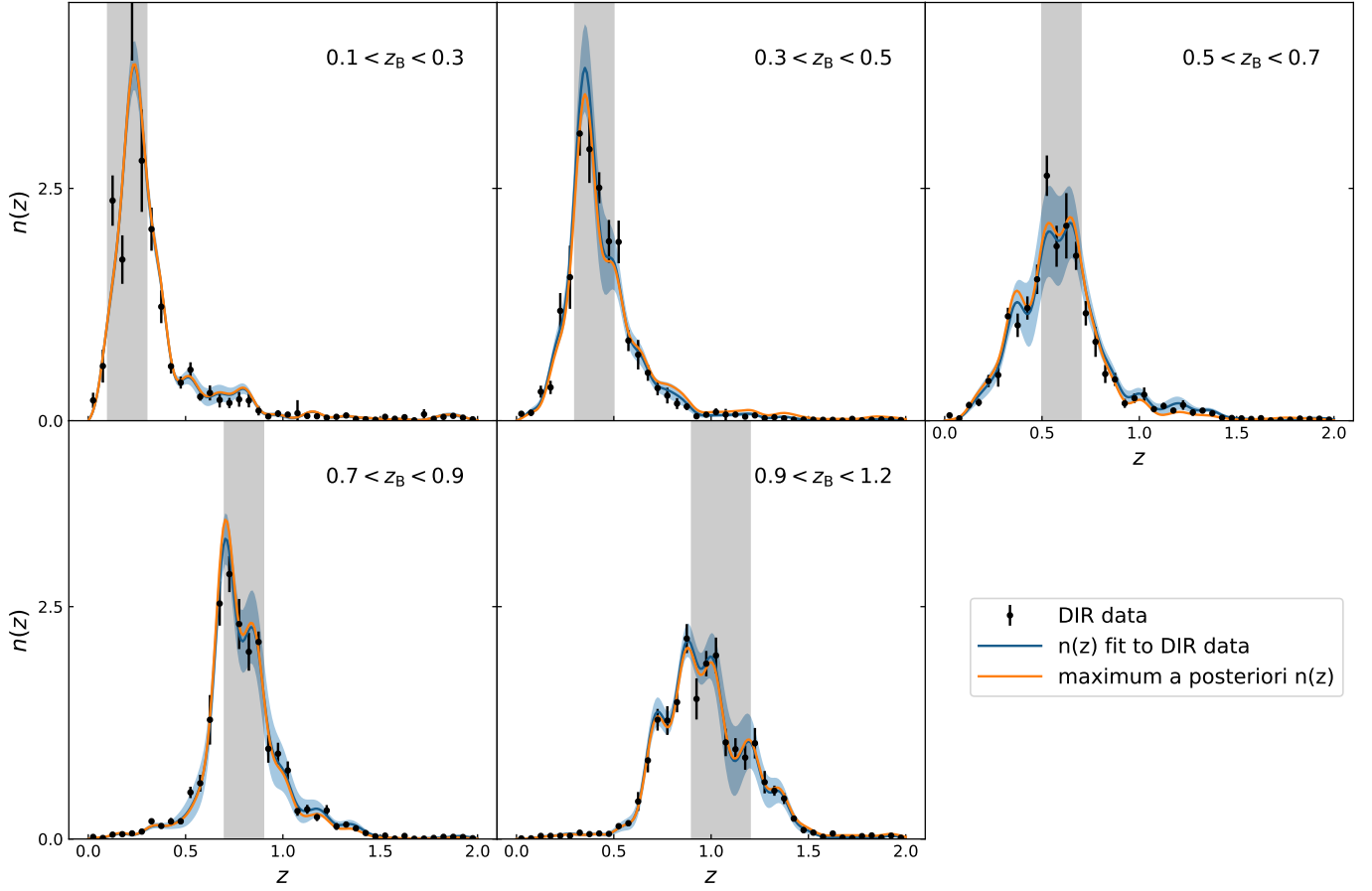
The iterative optimisation method shows a fast convergence to the best fit in the full parameter space of cosmological and nuisance parameters after only two cosmology optimisation steps and one redshift nuisance parameter optimisation. This was

**Fig. 2.** Fit results of a Gaussian mixture with 30 components to the redshift distribution in five tomographic redshift bins. Blue curves indicate redshift distributions fitted to the pre-calibrated DIR redshift histograms, shown in black. Shaded regions indicate the uncertainties on the redshift distributions derived from the diagonal elements of the correlation matrix of fit parameters, shown in Fig. 3. Orange curves represent the redshift distributions after iterative optimisation of cosmological and nuisance parameters.
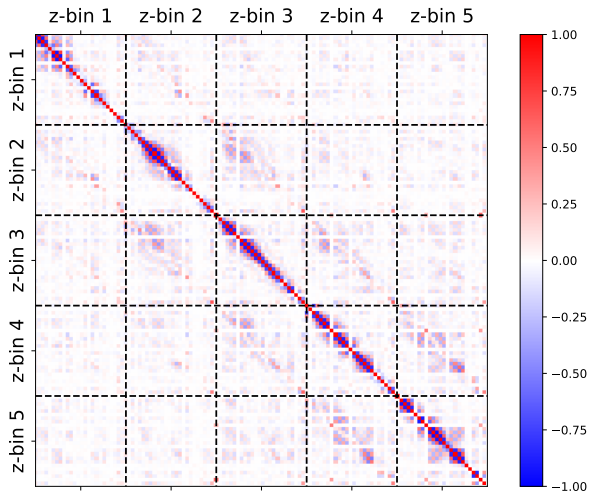


**Fig. 3.** Correlation matrix of best-fit comb amplitudes with 30 components per redshift bin.
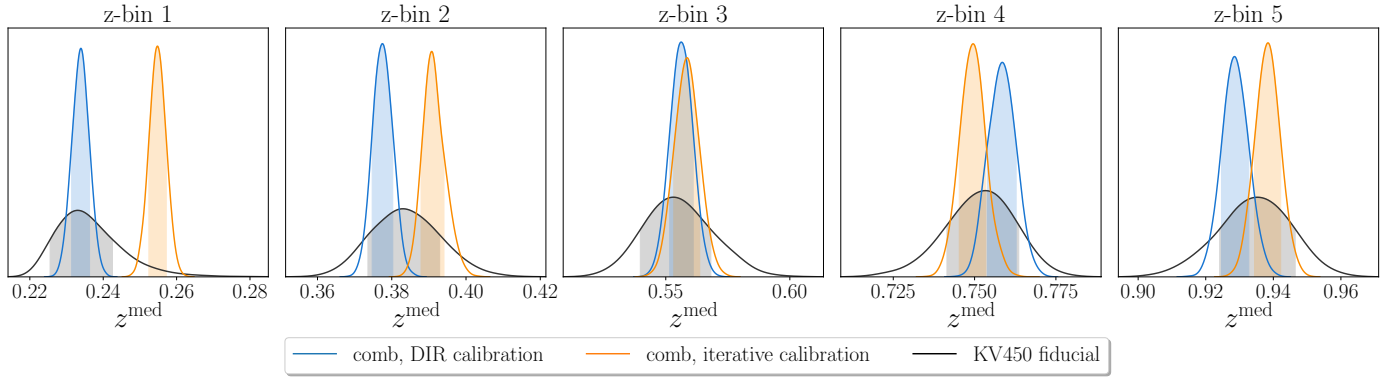
**Table 2.** Results of the iterative fitting of cosmological and nuisance parameters to the KV450 cosmic shear data and comparison to the fiducial KV450 likelihood.

|  | $A_{\mathrm{IA}}$ | $S_8$ | $\chi^2$ |
|---|---|---|---|
| Fiducial KV450 likelihood | 0.8656 | 0.7708 | 179.88 |
| 1. Cosmology optimisation | 0.7353 | 0.7768 | 180.67 |
| 2. Nuisance optimisation | – | – | 179.11 |
| 3. Cosmology optimisation | 0.7903 | 0.7882 | 178.62 |

**Notes.** When optimising cosmological parameters, we fit $\omega_{\mathrm{cdm}}$, $\omega_{\mathrm{b}}$, $A_{\mathrm{s}}$, $n_{\mathrm{s}}$, and $h$ as well as the nuisance parameters $A_{\mathrm{IA}}$, $A_{\mathrm{bary}}$, $\delta c$, and $A_c$. When optimising nuisance parameters, we vary the amplitudes of the Gaussian comb. Results are shown for the two most interesting parameters, $A_{\mathrm{IA}}$ and $S_8$, for which the cosmic shear likelihood has the largest constraining power. We find convergence after three iterations of the calibration, which results in a better fit to the cosmic shear data compared to the fiducial analysis.

unsurprising since we started from an already well-calibrated redshift distribution and as such expected only small corrections from the Newton optimisation step.

Using the best-fit $\chi^2$ values as a measure of goodness of fit, we find that with $\chi^2 = 178.62$ our model provides an improvement in $\chi^2$ of roughly 1% with respect to the fiducial KV450

model with a value of $\chi^2 = 179.88$. While in the present analysis allowing for a full variation in redshift distribution only gives a slight improvement compared to a linear shift in the mean, this method could become more relevant for future analyses with increased precision.

It is common practice to assess the goodness of fit by making the assumption that the $\chi^2$ statistic follows a $\chi^2$ distribution with $N_{\mathrm{d.o.f.}} = N_{\mathrm{d}} - N_{\Theta}$, where $N_{\mathrm{d}}$ is the size of the data vector

z-bin 1    z-bin 2    z-bin 3    z-bin 4    z-bin 5

— comb, DIR calibration    — comb, iterative calibration    — KV450 fiducial

**Fig. 4.** Posterior distribution of the median redshift of each tomographic redshift bin, inferred by drawing realisations of the Gaussian comb amplitudes from a multivariate Gaussian distribution. Black curves indicate the median redshift of the KV450 redshift histograms calibrated using the fiducial DIR method. The blue curves show the median redshift of the Gaussian comb that is fitted to the DIR histograms. The orange curves represent the median redshift of the Gaussian comb after iterative self-calibration with cosmic shear measurements.

and $N_\Theta$ is the number of sampling parameters. However, this assumption is only valid under the condition that the data are normally distributed, the model is linearly dependent on the sampling parameters, and there is no informative prior on the parameter ranges (see for instance Joachimi et al. 2021). In general, these conditions are not met in cosmological analyses, and this is particularly true for this work since we assumed a Gaussian prior on the amplitudes of the Gaussian comb, which is inferred from the redshift distribution calibration. Therefore, the naive estimation of the number of degrees of freedom is a poor estimation of the true effective number of degrees of freedom since we added a large number of strongly correlated nuisance parameters with informative priors. For a conservative estimate of the number of degrees of freedom in our model, we can assume the nuisance parameters to be essentially fixed by the prior and therefore do not count them as sampling parameters, which leads to $N_{\rm d.o.f.} = 186$. While a more robust estimate of the effective number of degrees of freedom can be inferred from mocks or posterior predictive data realisations (Spiegelhalter et al. 2002; Handley & Lemos 2019; Raveri & Hu 2019; Joachimi et al. 2021), we refrain from a further interpretation of the goodness of fit.

We find a shift in the two most interesting parameters for which the cosmic shear likelihood has the largest constraining power, $A_{\rm IA}$ and $S_8$, compared to the fiducial KV450 analysis. These shifts are further investigated in the following section, where we sample the weak lensing likelihood and derive marginalised posteriors of cosmological parameters. The resulting redshift distributions after iterative self-calibration are illustrated as orange curves in Fig. 2.

Figure 4 shows comparisons of the median of the redshift distribution of each tomographic bin inferred from the original DIR histograms and the Gaussian comb before and after iterative calibration. We chose the median as our summary statistic since the mean of the DIR histograms is less stable with respect to variations in the cutoff redshift at the high-redshift tail of the distribution, which is most likely caused by the underestimation of the error bars in the DIR method. The median, on the other hand, is less sensitive to the choice of the cutoff redshift. We find that the fit of the Gaussian comb yields constraints on the median redshift that are about 50% tighter relative to the DIR histograms. Additionally, we observe that the shift in the median after iterative self-calibration is largest in the first two redshift bins and less significant in the three higher redshift bins. This is most likely caused by degeneracies between the amplitude of
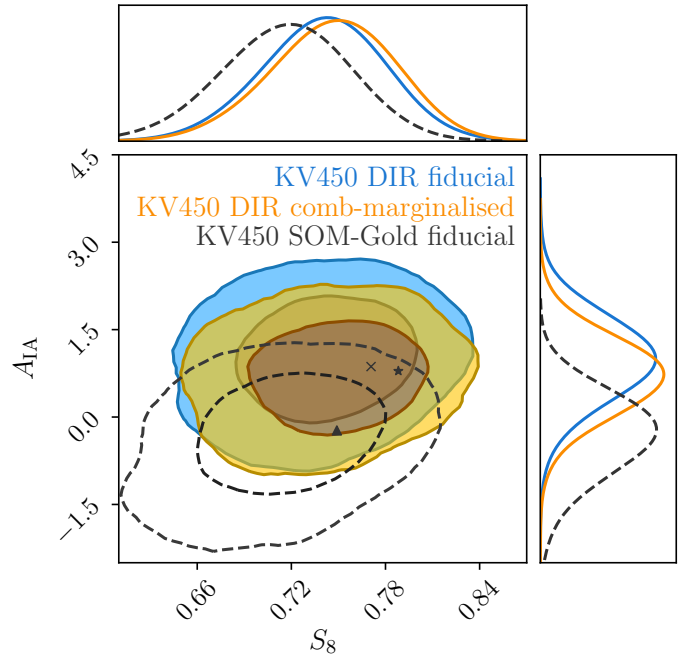


**Fig. 5.** Marginalised posteriors for $A_{\rm IA}$ and $S_8$. The orange contours present the results from the KV450 likelihood with a self-calibrated Gaussian comb and analytical marginalisation over nuisance parameters, while the blue contours refer to the fiducial KV450 constraints. The star indicates the best-fit values from Table 2 for the KV450 likelihood with a Gaussian comb, and the cross indicates the best-fit values for the fiducial KV450 likelihood. The dashed contour shows the posterior distribution from the KV450 'gold' sample (Wright et al. 2020b), which is constructed by removing photometric sources that are not directly represented by the overlapping spectroscopic reference samples using SOMs. Therefore, this contour is inferred from a different sample of galaxies with a different redshift distribution.

intrinsic alignments and the redshift distributions, which is discussed in the following sections.

### 5.2. Marginalisation over nuisance parameters

Using the redshift distribution calibrated in the previous section, we sampled the weak lensing likelihood in cosmological parameter space with analytical marginalisation over the uncertainty
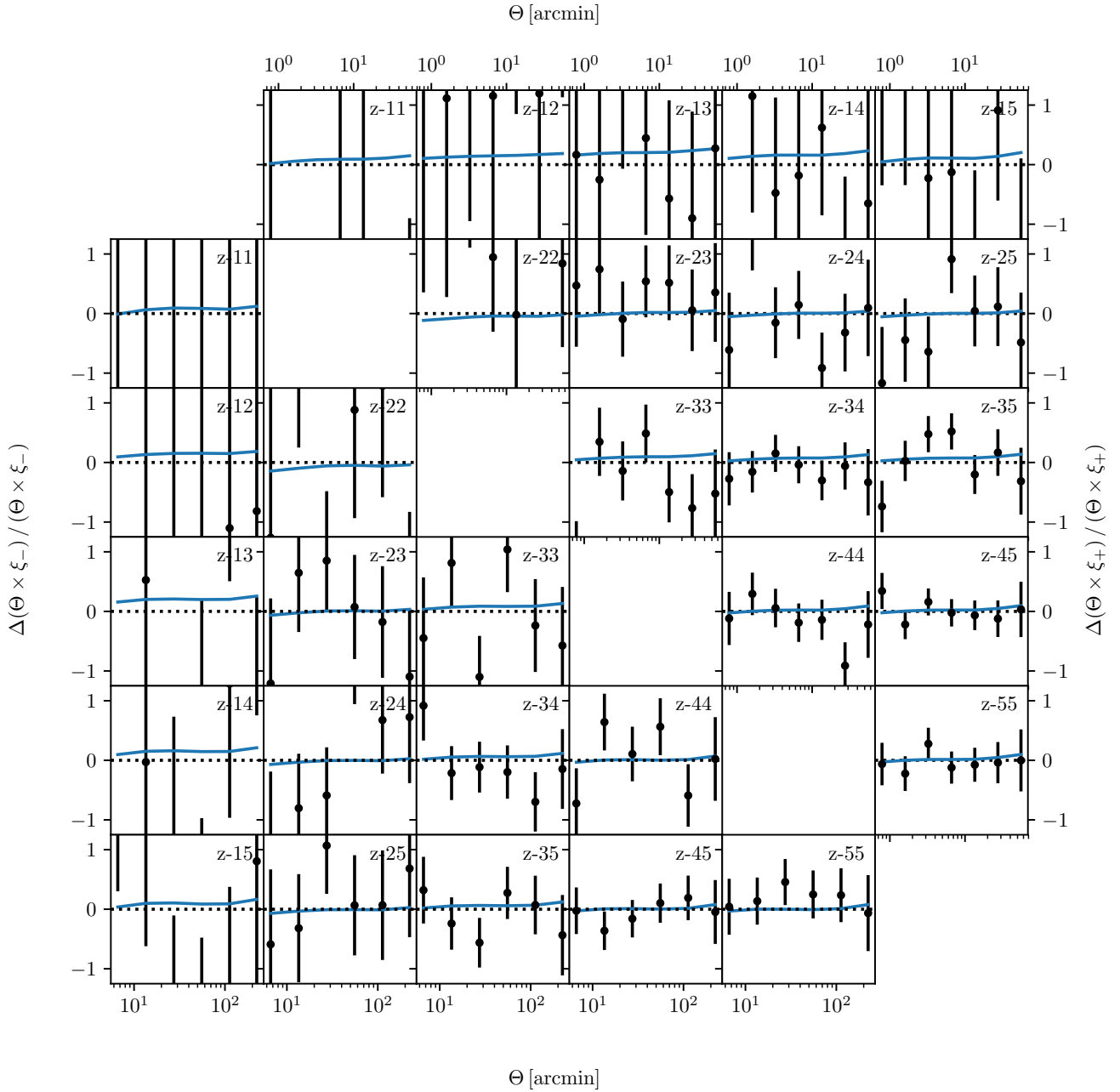
**Fig. 6.** Relative difference between the best-fit KV450 two-point shear correlation functions of the fiducial likelihood and the best fit after the iterative calibration of cosmological and nuisance parameters: $\xi_+$ (*upper right*) and $\xi_-$ (*lower left*). Black data points illustrate the relative difference between the observed two-point shear correlation functions and the best fit of the fiducial likelihood.

on the amplitudes of the fitted redshift distribution. Prior to the sampling of the marginalised likelihood, we tested whether we could reproduce the result of the fiducial KV450 analysis by sampling the likelihood with the comb model, but without applying the marginalisation over nuisance parameters. The results of this consistency test, discussed in Appendix C, show that the two models are in good agreement.

Figure 5 illustrates the results, comparing (i) the KV450 likelihood with a Gaussian comb and analytical marginalisation over nuisance parameters with (ii) the fiducial KV450 likelihood. We show marginalised posteriors and best-fit values for the two parameters that are fully constrained with KV450 data, $A_{\rm IA}$ and $S_8$. The posterior distribution of all remaining parameters is shown in Appendix D. We find a slight shift in

the posterior towards smaller values of the intrinsic alignment amplitude and larger values of $S_8$. Additionally, Fig. 5 shows the posterior distribution for the KV450 'gold' sample, which is derived using SOMs (Wright et al. 2020b). We emphasise that Wright et al. (2020b) use a different selection of the photometric sample by removing photometric sources that are not directly represented by the overlapping spectroscopic reference samples. Thus, the redshift distributions of the fiducial KV450 sample and the KV450 gold sample are not comparable.

The constraint from the KV450 gold sample on the intrinsic alignment amplitude, $A_{\rm IA}$, is compatible with $A_{\rm IA} = 0$, whereas Hildebrandt et al. (2020) found $A_{\rm IA} \approx 1$. However, these results are still consistent within their errors, as discussed by Wright et al. (2020b). The iterative self-calibration of the

redshift distribution performed in this work leads to a decrease in the intrinsic alignment amplitude of about 10% (see Table 2). Thus, we find a trend similar to that found by Wright et al. (2020b), although the change in the intrinsic alignment amplitude is not as strong. Recent studies of intrinsic alignments have also found results that are in disagreement with the fiducial KV450 analysis, such as Fortuna et al. (2021), who predict $A_{IA} = 0.1^{+0.1}_{-0.1}$. Since the constraints on the intrinsic alignment amplitude differ between analyses and the role of intrinsic alignments is a subject of active research, it is worth investigating how this parameter can influence the theoretical prediction of the cosmic shear signal.

The intrinsic alignment amplitude, $A_{IA}$, is not a cosmological parameter, but instead originates from the modelling of correlations between intrinsic ellipticities of neighbouring galaxies, II, and correlations between intrinsic ellipticities of foreground galaxies and background galaxies, GI. As can be inferred from Eqs. (21) and (22), the GI term gives a negative contribution to $\xi_\pm$ that is proportional to $A_{IA}$, whereas the II term contributes positively, proportionally to $A_{IA}^2$. Thus, a shift in the redshift distribution can (at least to some extent) be counteracted by a shift in the intrinsic alignment amplitude, so that overall we find a good fit to the observed cosmic shear two-point correlation function. This effect is a possible explanation for the observed shift in the contours in Fig. 5, which nevertheless are still in good agreement.

Furthermore, since the signal-to-noise ratio (S/N) is lower in the low redshift bins compared to the high redshift bins, the relative contribution of intrinsic alignments is stronger in the low redshift bins. This can explain the larger shift in the median of the redshift distribution in the first two redshift bins that is observed in Fig. 4. This is illustrated in Fig. 6, which shows the relative difference between the best-fit KV450 two-point shear correlation functions of the fiducial likelihood and the best fit after iterative calibration of cosmological and nuisance parameters. We find a relative difference of the best-fit curves of up to 20%, which is largest in the first redshift bin. This is compatible with the observed shift in the median of the redshift distribution of the first redshift bin shown in Fig. 4.

The black data points in Fig. 6 show the relative difference between the observed two-point correlation function and the fiducial best fit. These indicate that the S/N in this bin is very low, so that the shift in the posterior redshift distribution does not have a significant impact on the overall best-fit likelihood value of the combined fit. In the second bin we find the smallest shift in the best-fit curve, although Fig. 4 shows a significant shift in the median of the redshift distribution in this bin. This is an indication that the shift in the intrinsic alignment amplitude towards a lower value possibly mitigates the effect of the shifted redshift distribution, so that the effect on the likelihood value is minimal. From these observations, we conclude that the intrinsic alignment parameter $A_{IA}$ does not solely measure the amplitude of intrinsic alignments, but instead picks up contributions from systematic shifts in the redshift distribution due to the degeneracy between the parameters. Variations between the constraints on the intrinsic alignment parameter were also reported by Wright et al. (2020b), who found differences of up to $|\Delta A_{IA}| \sim 1.0\sigma$ between analyses. However, since Wright et al. (2020b) used a different galaxy sample, the intrinsic alignment amplitude could be intrinsically different. Furthermore, the effect of the intrinsic alignment parameter mitigating systematic effects has been studied recently in other works, such as van Uitert et al. (2018) and Efstathiou & Lemos (2018).

We conclude that our method provides constraints on cosmological parameters that are compatible with the fiducial KV450 analysis while taking all photometric redshift uncertainties into account. Our model provides a slightly better fit of the redshift distribution to the cosmic shear data since the large number of model parameters allows the model to reflect small variations in the redshift distribution. However, we suspect that variations in the redshift distribution can be correlated with variations in the intrinsic alignment amplitude, and therefore a tighter prior on the intrinsic alignment parameter through external constraints is required. Our approach can help reduce the degeneracy between intrinsic alignments and redshift distributions by providing a more accurate redshift distribution calibration.

## 6. Summary and conclusions

In this paper we developed a method to model photometric redshift distributions of galaxy samples with strong correlations between tomographic bins using a modified Gaussian mixture model. We have shown that photometric redshift uncertainties arising from the calibration of the redshift distribution can be accurately propagated to the weak lensing likelihood via an analytic marginalisation over the model parameters. This allowed us to use a fairly complex model of the redshift distribution without an increase in the number of sampling parameters in the weak lensing likelihood. Additionally, we developed an iterative method to fit cosmological and nuisance parameters in order to perform a self-calibration of the redshift distribution with cosmic shear data.

We applied these methods to the public KiDS+VIKING-450 (KV450) cosmic shear data. We fitted the modified Gaussian mixture model to the fiducial KV450 redshift distributions in five tomographic bins that were calibrated with deep spectroscopic surveys and implemented the marginalisation method in the public KV450 likelihood code. We performed the iterative fitting and found fast convergence to the best fit in the combined space of cosmological and nuisance parameters. Next, we sampled the weak lensing likelihood using the redshift distribution that was calibrated in the previous step and derived constraints on cosmological parameters.

We found that our model can fit complex redshift distributions thanks to the tunable number of model parameters. Since we marginalise analytically over nuisance parameters, the large number of redshift nuisance parameters does not increase the runtime of the posterior sampling. Our model provides a slightly better fit to the data compared to the fiducial KV450 likelihood since the fiducial likelihood only allows a shift in the mean of the redshift distribution of each bin and thus requires a precalibrated redshift distribution that closely resembles the true underlying distribution. Given the large uncertainties of photometric redshift calibration methods in general, a complex model that can reflect the uncertainties is advantageous. Therefore, with decreasing statistical uncertainties and increasing survey data, the method presented in this paper is particularly useful for upcoming surveys, where we expect higher order moments of the redshift distribution uncertainty to become increasingly important.

The marginalised posterior distributions of the remaining model parameters are in agreement with the fiducial KV450 analysis. However, we found slight shifts in the posterior constraints on the model parameters, which are strongest for the amplitude of intrinsic alignments, $A_{IA}$. We suspect that these shifts are caused by degeneracies between the redshift distribution amplitudes and the intrinsic alignment amplitude, so that

a shift in the redshift distribution can be compensated by a shift in the intrinsic alignment amplitude. This mitigation of systematic effects by the intrinsic alignment parameter is likely to be the reason for the relatively large shift in the median of the redshift distribution in the second redshift bin that we found after the iterative calibration of model parameters. Thus, to get unbiased constraints on the redshift distribution, we require a tighter prior on the intrinsic alignment parameter through external constraints. This will ensure that systematic effects are not absorbed by the intrinsic alignment parameter. This result is consistent with earlier works, such as Wright et al. (2020b), Hildebrandt et al. (2020), Fortuna et al. (2021), van Uitert et al. (2018), and Efstathiou & Lemos (2018), which also found discrepant values of the intrinsic alignment amplitude and studied systematic effects on intrinsic alignments. Thus, this work further emphasises the necessity of an accurate modelling of intrinsic alignments.

While finalising this work, Hadzhiyska et al. (2020) put forward a paper on the analytic marginalisation of redshift distribution uncertainties applied to galaxy clustering measurements from the HSC first data release. Their marginalisation method results in a modified data covariance matrix that downweights modes of the data vector that are sensitive to variations in the redshift distribution. This approach also allowed them to take the full shape of the redshift distribution into account. However, since this method directly modifies the data covariance matrix, it is unclear if it allows for a self-calibration of the redshift distribution with cosmic shear measurements.

The method presented in this paper is not only applicable to cosmic shear analyses, but can also be adapted to other probes, such as galaxy-galaxy lensing and galaxy clustering. Therefore, it can especially be used in future joint '6 × 2 pt' analyses, which combine all two-point correlation functions between overlapping imaging and spectroscopic surveys.

# References

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, Phys. Rev. D, 98, 043526
Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4
Alarcon, A., Sánchez, C., Bernstein, G. M., & Gaztañaga, E. 2020, MNRAS, 498, 2614
Amara, A., & Réfrégier, A. 2007, MNRAS, 381, 1018
Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, A&A, 645, A104
Audren, B., Lesgourgues, J., Benabed, K., & Prunet, S. 2013, JCAP, 1302, 001
Becker, M. R., & Rozo, E. 2016, MNRAS, 457, 304
Benítez, N. 2000, ApJ, 536, 571
Benjamin, J., Van Waerbeke, L., Heymans, C., et al. 2013, MNRAS, 431, 1547
Bishop, C. 2016, Pattern Recognition and Machine Learning, Information Science and Statistics (New York: Springer)
Blas, D., Lesgourgues, J., & Tram, T. 2011, JCAP, 2011, 034
Bonnett, C., Troxel, M. A., Hartley, W., et al. 2016, Phys. Rev. D, 94, 042005
Bridle, S., & King, L. 2007, New J. Phys., 9, 444
Bridle, S. L., Crittenden, R., Melchiorri, A., et al. 2002, MNRAS, 335, 1193
Brinckmann, T., & Lesgourgues, J. 2019, Phys. Dark Univ., 24, 100260
Buchner, J., Georgakis, A., Nandra, K., et al. 2014, A&A, 564, A125
Buchs, R., Davis, C., Gruen, D., et al. 2019, MNRAS, 489, 820
Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, MNRAS, 447, 646
Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, MNRAS, 453, 1513
de Jong, J. T. A., Verdoes Kleijn, G. A., Boxhoorn, D. R., et al. 2015, A&A, 582, A62
de Jong, J. T. A., Verdoes Kleijn, G. A., Erben, T., et al. 2017, A&A, 604, A134
Drlica-Wagner, A., Sevilla-Noarbe, I., Rykoff, E. S., et al. 2018, ApJS, 235, 33
Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32
Efstathiou, G., & Lemos, P. 2018, MNRAS, 476, 151
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, Open J. Astrophys., 2, 10
Fortuna, M. C., Hoekstra, H., Joachimi, B., et al. 2021, MNRAS, 501, 2983
Fosalba, P., Crocce, M., Gaztañaga, E., & Castander, F. J. 2015a, MNRAS, 448, 2987
Fosalba, P., Gaztañaga, E., Castander, F. J., & Crocce, M. 2015b, MNRAS, 447, 1319
Gatti, M., Sheldon, E., Amon, A., et al. 2021, MNRAS, 504, 4312
Hadzhiyska, B., Alonso, D., Nicola, A., & Slosar, A. 2020, JCAP, 2020, 056
Handley, W., & Lemos, P. 2019, Phys. Rev. D, 100, 023512
Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43
Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454
Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69
Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, A&A, 647, A124
Hirata, C. M., & Seljak, U. 2004, Phys. Rev. D, 70, 063526
Hoffmann, K., Bel, J., Gaztañaga, E., et al. 2015, MNRAS, 447, 1724
Hoyle, B., & Rau, M. M. 2019, MNRAS, 485, 3642
Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, MNRAS, 478, 592
Joachimi, B., Mandelbaum, R., Abdalla, F. B., & Bridle, S. L. 2011, A&A, 527, A26
Joachimi, B., Lin, C. A., Asgari, M., et al. 2021, A&A, 646, A129
Kaiser, N. 1992, ApJ, 388, 272
Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, MNRAS, 454, 3500
Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, A&A, 625, A2
Leistedt, B., Hogg, D. W., Wechsler, R. H., & DeRose, J. 2019, ApJ, 881, 80
Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, MNRAS, 390, 118
Masters, D., Capak, P., Stern, D., et al. 2015, ApJ, 813, 53
Matthews, D. J., & Newman, J. A. 2010, ApJ, 721, 456
McLeod, M., Balan, S. T., & Abdalla, F. B. 2017, MNRAS, 466, 3558
Mead, A. J., Peacock, J. A., Heymans, C., Joudaki, S., & Heavens, A. F. 2015, MNRAS, 454, 1958
Ménard, B., Scranton, R., Schmidt, S., et al. 2013, ArXiv e-prints [arXiv:1303.4722]
Newman, J. A. 2008, ApJ, 684, 88
Rau, M. M., Wilson, S., & Mandelbaum, R. 2020, MNRAS, 491, 4768
Raveri, M., & Hu, W. 2019, Phys. Rev. D, 99, 043506
Salvato, M., Ilbert, O., & Hoyle, B. 2019, Nat. Astron., 3, 212
Sánchez, C., & Bernstein, G. M. 2019, MNRAS, 483, 2801
Schaan, E., Ferraro, S., & Seljak, U. 2020, JCAP, 2020, 001
Schneider, P., van Waerbeke, L., Kilbinger, M., & Mellier, Y. 2002, A&A, 396, 1
Schneider, P., Eifler, T., & Krause, E. 2010, A&A, 520, A116
Sevilla-Noarbe, I., Bechtol, K., Carrasco Kind, M., et al. 2021, ApJS, 254, 24
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. 2002, J. R. Stat. Soc.: Series B (Stat. Methodol.), 64, 583
Taylor, A. N., & Kitching, T. D. 2010, MNRAS, 408, 865
van den Busch, J. L., Hildebrandt, H., Wright, A. H., et al. 2020, A&A, 642, A200
van Erp, S., Oberski, D. L., & Mulder, J. 2019, J. Math. Psychol., 89, 31
van Uitert, E., Joachimi, B., Joudaki, S., et al. 2018, MNRAS, 476, 4662
Wright, A. H., Hildebrandt, H., Kuijken, K., et al. 2019, A&A, 632, A34
Wright, A. H., Hildebrandt, H., van den Busch, J. L., & Heymans, C. 2020a, A&A, 637, A100
Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2020b, A&A, 640, L14
Zhang, P., Pen, U.-L., & Bernstein, G. 2010, MNRAS, 405, 359
Zuntz, J., Sheldon, E., Samuroff, S., et al. 2018, MNRAS, 481, 1149

# Appendix A: Marginalisation over nuisance parameters

In this appendix we provide the analytic expressions for the vector of derivatives of the log-likelihood with respect to the nuisance parameters $a_i^\alpha$ of the redshift distribution and the Hessian matrix of second derivatives with respect to the nuisance parameters. These quantities enter the calculation of the log-likelihood marginalised over the nuisance parameters described in Sect. 3.2. For the specific case of marginalising over the redshift distribution nuisance parameters, the vector $\mathcal{L}'$ has elements

$$
\frac{\partial \mathcal{L}}{\partial a_m^\mu} = \sum_{l,l',\alpha,\beta,\alpha',\beta'} \left\{ \frac{\partial \Delta_l^{(\alpha\beta)}}{\partial a_m^\mu} Z_{(l,\alpha,\beta)(l',\alpha',\beta')} \Delta_{l'}^{(\alpha'\beta')} \right.
$$
$$
\left. + \Delta_l^{(\alpha\beta)} Z_{(l,\alpha,\beta)(l',\alpha',\delta)} \frac{\partial \Delta_{l'}^{(\alpha'\beta')}}{\partial a_m^\mu} \right\}, \tag{A.1}
$$

with

$$
\frac{\partial \Delta_l^{(\alpha\beta)}}{\partial a_m^\mu} = -A_m^\mu \sum_i x_\pm^{(im)}(\theta_l) \left( \delta_{\alpha\mu} A_i^\beta + \delta_{\beta\mu} A_i^\alpha \right), \tag{A.2}
$$

where $\delta_{\alpha\beta}$ denotes the Kronecker delta symbol. The indices $\alpha$ and $\beta$ run over all unique combinations of tomographic redshift bins. The two-point shear correlation function of two Gaussian comb components, $i$ and $j$, in $\theta$-bin $l$ is denoted by $x_\pm^{(ij)}(\theta_l)$, and the inverse data covariance is given by $Z$. The difference between the observed and predicted signals, as defined in Eq. (26), is denoted by $\Delta_l^{\alpha\beta}$. The elements of the Hessian matrix, $\mathcal{L}''$, read

$$
\frac{\partial^2 \mathcal{L}}{\partial a_m^\mu \partial a_n^\nu} = \sum_{l,l',\alpha,\beta,\alpha',\beta'} \left\{ \frac{\partial \Delta_l^{(\alpha\beta)}}{\partial a_m^\mu} Z_{(l,\alpha,\beta)(l',\alpha',\beta')} \frac{\partial \Delta_{l'}^{(\alpha'\beta')}}{\partial a_n^\nu} \right.
$$
$$
+ \frac{\partial \Delta_l^{(\alpha\beta)}}{\partial a_n^\nu} Z_{(l,\alpha,\beta)(l',\alpha',\beta')} \frac{\partial \Delta_{l'}^{(\alpha'\beta')}}{\partial a_m^\mu}
$$
$$
+ \frac{\partial^2 \Delta_l^{(\alpha\beta)}}{\partial a_m^\mu \partial a_n^\nu} Z_{(l,\alpha,\beta)(l',\alpha',\beta')} \Delta_{l'}^{(\alpha'\beta')}
$$
$$
\left. + \Delta_l^{(\alpha\beta)} Z_{(l,\alpha,\beta)(l',\alpha',\beta')} \frac{\partial^2 \Delta_{l'}^{(\alpha'\beta')}}{\partial a_m^\mu \partial a_n^\nu} \right\}, \tag{A.3}
$$

with

$$
\frac{\partial^2 \Delta_l^{(\alpha\beta)}}{\partial a_m^\mu \partial a_n^\nu} = -A_m^\mu A_n^\nu x_\pm^{(mn)}(\theta_l) \left( \delta_{\alpha\mu}\delta_{\beta\nu} + \delta_{\beta\mu}\delta_{\alpha\nu} \right)
$$
$$
- \delta_{mn}\delta_{\mu\nu} A_m^\mu \sum_i x_\pm^{(im)}(\theta_l) \left( \delta_{\alpha\mu} A_i^\beta + \delta_{\beta\mu} A_i^\alpha \right). \tag{A.4}
$$

# Appendix B: Tests of the redshift distribution calibration

In this appendix we test to what extent the number of Gaussian components of the comb model affects the fit to the pre-calibrated redshift histograms. Additionally, we test the stability of the fit results when changing the number of bins of input data histograms and test the calibration method on simulations.

## B.1. Width of input data histograms

The redshift distributions of Hildebrandt et al. (2020), calibrated with the fiducial DIR method, consist of histograms with bin

width $\Delta z = 0.05$ for each tomographic bin and a covariance matrix that links all five tomographic bins. To test the sensitivity of the fit with respect to the input data, we performed fits using a second set of histograms with a smaller bin width of $\Delta z = 0.025$ that were calibrated with the same method.

Figure B.1 shows a comparison of two fits with 30 Gaussian components; the blue lines represent a fit to the histograms with bin width $\Delta z = 0.05$, and the orange lines represent a fit to the histograms with bin width $\Delta z = 0.025$. We note that the error bars correspond to the diagonal elements of the covariance matrix of the data histograms. The fit of redshift distributions, however, is performed using the full covariance matrix.

By visually inspecting the fitted redshift distributions, we observe some deviations between the two curves, which are, however, already present in the input data. Although the two histograms are supposed to represent the same source redshift distribution, they show some fluctuations (especially in the peaks of the distributions), which have an impact on the fitted curves. More importantly, however, we find goodness of fit values of $\chi^2 = 4500$ and $\chi^2 = 22\,750$ for 50 and 250 degrees of freedom, respectively. This indicates a bad fit of the model to the data regardless of which data histogram is used. To find the cause of the bad fit, we repeated the fit using only the diagonal elements of the covariance matrix, which reduces the goodness of fit values to $\chi^2 = 470$ and $\chi^2 = 1400$ for 50 and 250 degrees of freedom, respectively. This test shows that the bad fit is to a great extent caused by the off-diagonal elements of the covariance matrix. However, excluding the off-diagonal elements does not lead to an acceptable goodness of fit value. We suspect that the uncertainties on the pre-calibrated redshift distributions are underestimated, which would explain the discrepancies between the blue and orange data points shown in Fig. B.1.

Empirically, we find that rescaling the square root of the covariance $C_{ij}$ between histogram bins by an additive and multiplicative factor via

$$
C_{ij}' = \left( 2\sqrt{C_{ij}} + 0.01\delta_{ij} \right)^2 \tag{B.1}
$$

leads to $\chi^2 = 80$ for 50 degrees of freedom. With this rescaling, the widths of the posterior distributions of the median redshift for both the redshift histograms and the Gaussian comb, shown in Fig. 2, inflate by approximately the same factor. Therefore, we assume that a potential underestimation of the error bars impacts the fiducial analysis and the analysis presented in this paper in the same way. We stress that the quality of the fit does not have a significant impact on the main analysis of this paper, as shown in Appendix B.2, and leave further investigation of an improved uncertainty quantification for the pre-calibrated redshift distributions for future work.

## B.2. Redshift distribution calibration with simulations

Since in Appendix B.1 we found that the Gaussian comb model provides a bad fit to the actual data, which is likely caused by an underestimation of the error bars, we tested our calibration method with simulations. We used redshift distributions that are calibrated with the fiducial DIR method on simulated mock catalogues (van den Busch et al. 2020) based on the MICE simulation (Fosalba et al. 2015a,b; Crocce et al. 2015; Carretero et al. 2015; Hoffmann et al. 2015). Analogous to Appendix B.1, we compared two types of data histograms with bin widths of $\Delta z = 0.05$ and $\Delta z = 0.025$. Figure B.2 shows a comparison of the two fits with 30 Gaussian components, where blue lines represent a fit to the histograms with bin width $\Delta z = 0.05$ and orange lines represent a fit to the histograms with bin width $\Delta z = 0.025$.
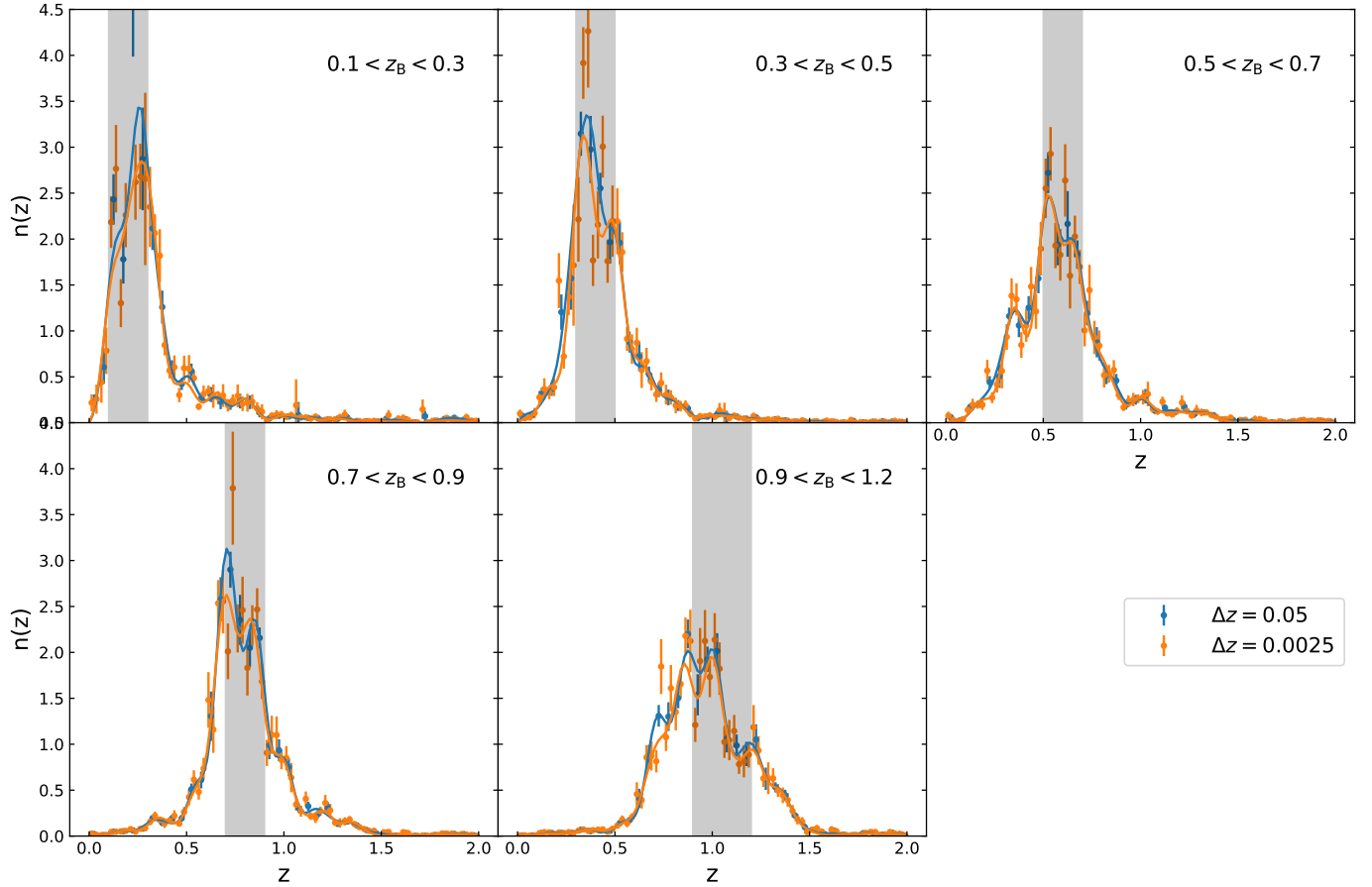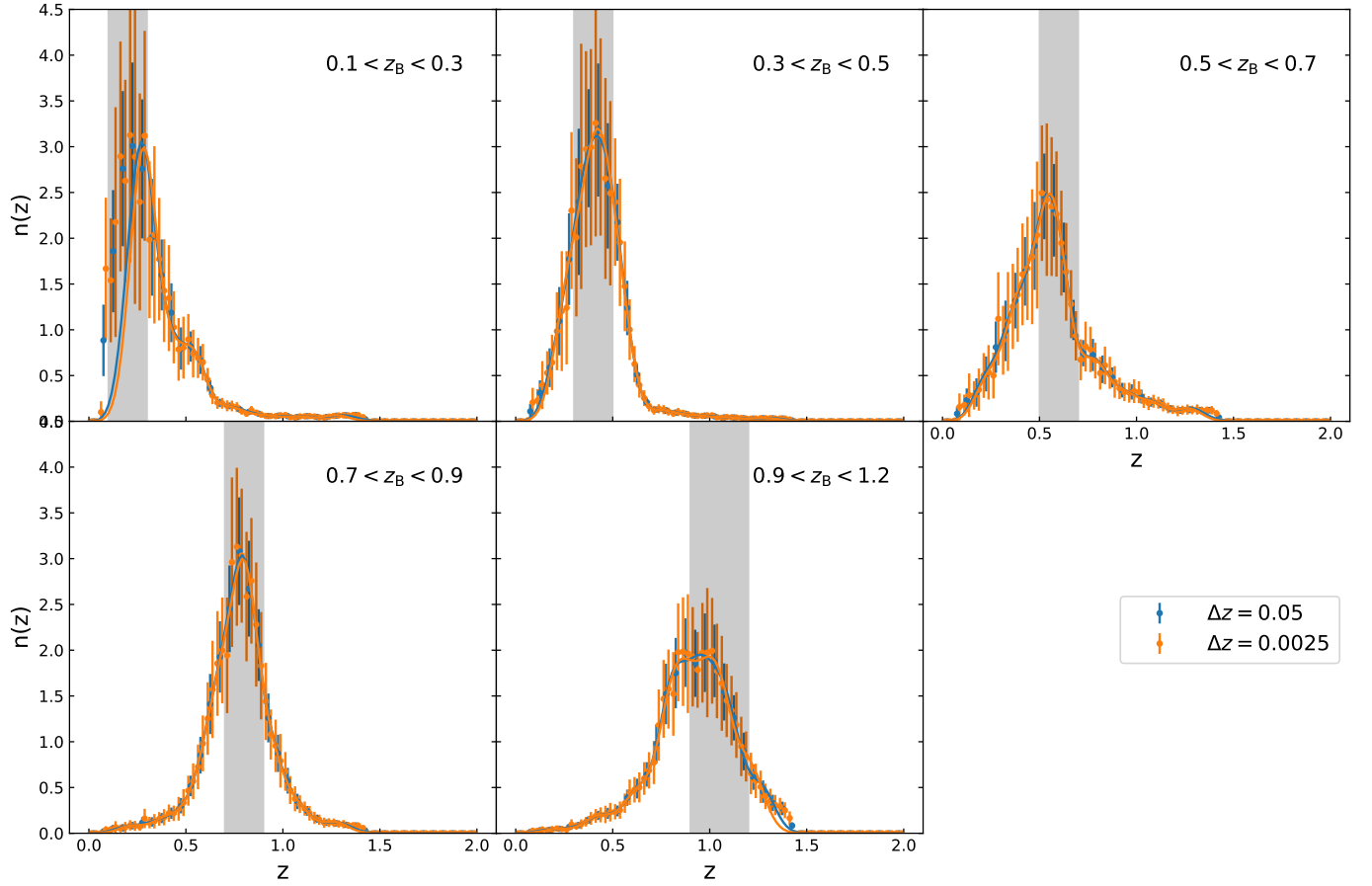
**Fig. B.1.** Comparison of the Gaussian comb with 30 components fitted to two different pre-calibrated histograms. The blue and orange points show histograms with bin widths of $\Delta z = 0.05$ and $\Delta z = 0.025$, respectively. The error bars correspond to the diagonal elements of the covariance matrix. The lines represent the Gaussian comb with 30 components fitted to the data histograms. We note that when fitting the redshift distribution, the full covariance matrix of the data histogram is taken into account.

We find that with goodness of fit values of $\chi^2 = 75$ and $\chi^2 = 320$ for 50 and 250 degrees of freedom, respectively, our Gaussian comb model fits the data reasonably well. Moreover, both the data histograms and the corresponding fitted redshift distributions are in excellent agreement. We conclude that the Gaussian comb model is capable of accurately describing the redshift distribution. The worse goodness of fit when fitting real data is likely due to the presence of noise and an underestimation of the uncertainties.

### B.3. Number of Gaussian components

In Fig. B.3 we show a comparison of fits of a Gaussian comb with 20, 30, and 40 components to the pre-calibrated redshift histograms with bin width $\Delta z = 0.05$. The width $\sigma_{comb}$ of the Gaussians is equal to the separation between each component. We find that variations in the number of comb components have a marginal impact on the redshift distribution, with changes of the median of order $\Delta z^{med} = 0.001$.

**Fig. B.2.** Comparison of the Gaussian comb with 30 components fitted to pre-calibrated histograms from the MICE simulation. The blue and orange points show histograms with bin widths of $\Delta z = 0.05$ and $\Delta z = 0.025$, respectively. The error bars correspond to the diagonal elements of the covariance matrix. The lines represent the Gaussian comb with 30 components fitted to the data histograms. We note that when fitting the redshift distribution, the full covariance matrix of the data histogram is taken into account.
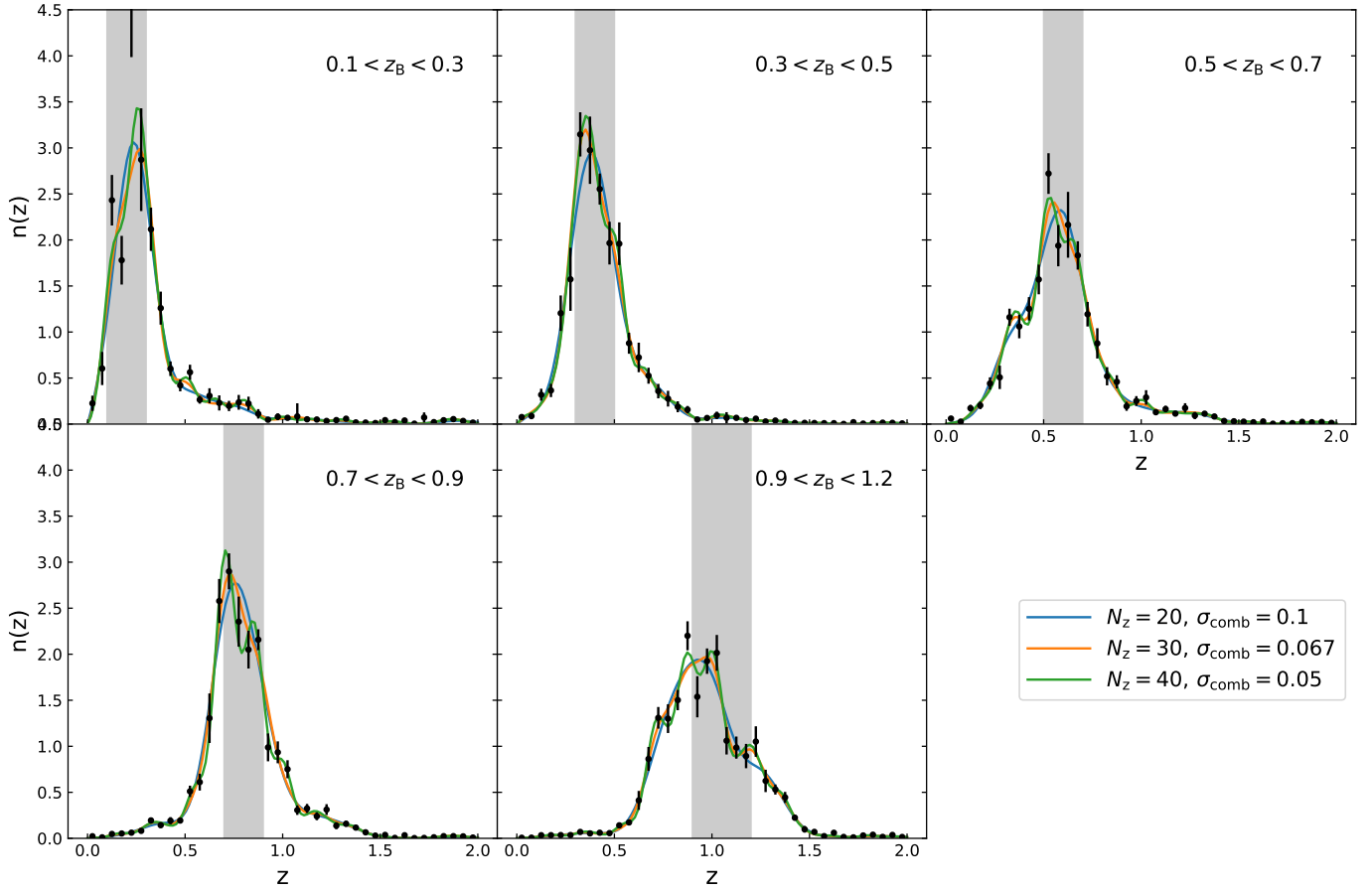
**Fig. B.3.** Comparison of a Gaussian comb with 20, 30, and 40 components fitted to a pre-calibrated histogram with bin width $\Delta z = 0.05$. The width $\sigma_{comb}$ of the Gaussians is equal to the separation between each component. Data points are shown in black, with error bars corresponding to the diagonal elements of the covariance matrix. Blue, orange, and green lines represent the Gaussian combs with 20, 30, and 40 components, respectively. We note that when fitting the redshift distribution, the full covariance matrix of the data histogram is taken into account.

## Appendix C: Comparison between the fiducial KV450 likelihood and the modified likelihood with Gaussian comb

In order to test if the fitted redshift distribution is capable of reproducing the results of the fiducial KV450 analysis, we sampled the likelihood using the Gaussian comb model as the parameterisation of the redshift distribution, but without marginalisation over the uncertainties on the nuisance parameters. To be able to compare the two likelihoods, we fixed the nuisance parameters $\delta z_i$ of the fiducial KV450 likelihood. The results of these fits are presented Table C.1, which shows the mean posterior values of cosmological and nuisance parameters. We find that constraints from both setups are fully consistent, and therefore we conclude that our Gaussian comb model can be used as an alternative to the fiducial redshift distributions.

**Table C.1.** Comparison between the fiducial KV450 likelihood and the modified likelihood with redshift distribution parameterised by the Gaussian comb model.

| Parameter | Fiducial KV450 | KV450 with Gaussian comb |
|---|---|---|
| $\omega_{cdm}$ | $0.112^{+0.029}_{-0.060}$ | $0.112^{+0.046}_{-0.060}$ |
| $\ln 10^{10}A_s$ | $3.30 \pm 0.92$ | $3.34 \pm 0.92$ |
| $\omega_b$ | $0.0223 \pm 0.0021$ | $0.0222^{+0.0018}_{-0.0025}$ |
| $n_s$ | $1.03^{+0.15}_{-0.13}$ | $1.01 \pm 0.13$ |
| $h$ | $0.749^{+0.067}_{-0.028}$ | $0.746^{+0.062}_{-0.033}$ |
| $A_{IA}$ | $0.89^{+0.64}_{-0.58}$ | $0.87^{+0.64}_{-0.58}$ |
| $c_{min}$ | $2.50^{+0.22}_{-0.45}$ | $2.49^{+0.23}_{-0.40}$ |
| $\delta c$ | $0.00000 \pm 0.00019$ | $0.00000 \pm 0.00019$ |
| $A_c$ | $1.03 \pm 0.12$ | $1.02 \pm 0.12$ |
| $\Omega_m$ | $0.242^{+0.052}_{-0.11}$ | $0.242^{+0.055}_{-0.11}$ |
| $\sigma_8$ | $0.86^{+0.18}_{-0.20}$ | $0.87 \pm 0.17$ |
| $S_8$ | $0.746^{+0.029}_{-0.028}$ | $0.748^{+0.029}_{-0.03}$ |

**Notes.** Reported are the mean posterior values and the 68% confidence intervals. The first five lines are cosmological parameters, and the remaining lines represent nuisance parameters.

# Appendix D: Posteriors of cosmological parameter constraints

In Fig. D.1 we show marginalised posteriors of cosmological and nuisance parameters. The KV450 likelihood with a Gaussian comb and analytical marginalisation over nuisance parameters is compared to the fiducial KV450 likelihood.
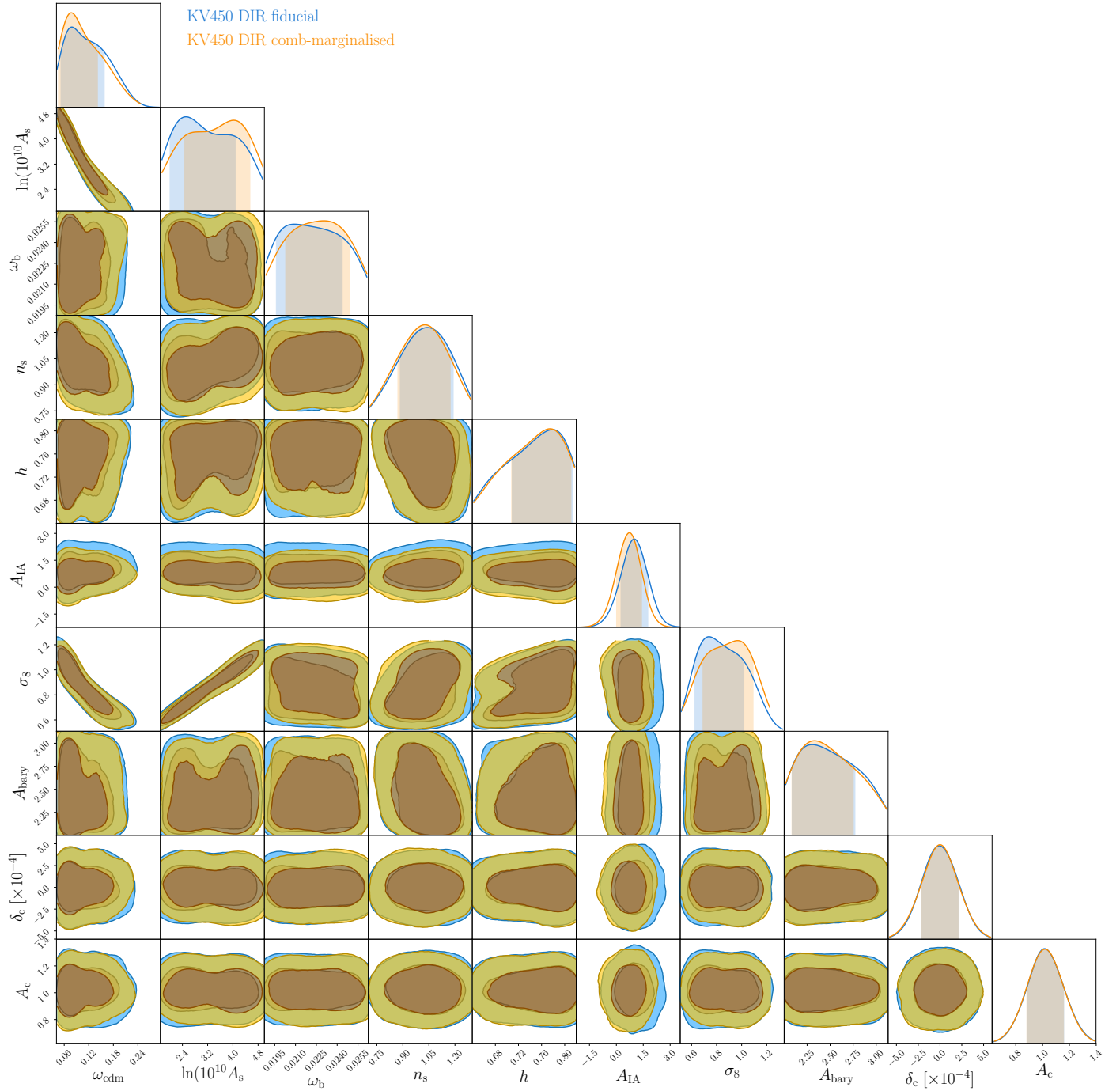


**Fig. D.1.** Marginalised posteriors for all parameters of the KV450 likelihood. Blue contours present the results from the KV450 likelihood with a Gaussian comb and analytical marginalisation over nuisance parameters, while the orange contours refer to the fiducial KV450 constraints.