MNRAS 000, 1–16 (0000) Preprint 20 April 2023 Compiled using MNRAS LATEX style file v3.0

Photometric Redshift Uncertainties in Weak Gravitational Lensing Shear Analysis: Models and Marginalization

Tianqing Zhang[1], Markus Michael Rau[1,2], Rachel Mandelbaum[1], Xiangchong Li[1], Ben Moews[1,3]

[1]McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA.
[2]High Energy Physics Division, Argonne National Laboratory, Lemont, IL 60439, USA
[3]Pittsburgh Supercomputing Center, Carnegie Mellon University & University of Pittsburgh, 300 S Craig St, Pittsburgh, PA 15213, USA.

20 April 2023

ABSTRACT

Recovering credible cosmological parameter constraints in a weak lensing shear analysis requires an accurate model that can be used to marginalize over nuisance parameters describing potential sources of systematic uncertainty, such as the uncertainties on the sample redshift distribution $n(z)$. Due to the challenge of running Markov Chain Monte-Carlo (MCMC) in the high dimensional parameter spaces in which the $n(z)$ uncertainties may be parameterized, it is common practice to simplify the $n(z)$ parameterization or combine MCMC chains that each have a fixed $n(z)$ resampled from the $n(z)$ uncertainties. In this work, we propose a statistically-principled Bayesian resampling approach for marginalizing over the $n(z)$ uncertainty using multiple MCMC chains. We self-consistently compare the new method to existing ones from the literature in the context of a forecasted cosmic shear analysis for the HSC three-year shape catalog, and find that these methods recover statistically consistent errorbars for the cosmological parameter constraints for predicted HSC three-year analysis, implying that using the most computationally efficient of the approaches is appropriate. However, we find that for datasets with the constraining power of the full HSC survey dataset (and, by implication, those upcoming surveys with even tighter constraints), the choice of method for marginalizing over $n(z)$ uncertainty among the several methods from the literature may modify the 1σ uncertainties on $\Omega_m - S_8$ constraints by $\sim$4%, and a careful model selection is needed to ensure credible parameter intervals.

Key words: methods: data analysis; methods: statistical; gravitational lensing: weak

1 INTRODUCTION

Over the past decade, wide-field imaging surveys, e.g., the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al. 2016), the Kilo-Degree Survey (KiDS; de Jong et al. 2017), and the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP; Aihara et al. 2018), became increasingly powerful, reaching fainter magnitudes and larger areas, and employing improved methods for controlling systematic biases and uncertainties (for a review, see Mandelbaum 2018). Future surveys such as the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezić et al. 2019; LSST Science Collaboration et al. 2009), the *Nancy Grace Roman* Space Telescope High Latitude Imaging Survey (Spergel et al. 2015; Akeson et al. 2019) and *Euclid* (Laureijs et al. 2011) will provide even larger data volumes and require more stringent control of systematic errors. With these developments, cosmic shear, the coherent weak gravitational lensing effect on the light from the distant galaxies caused by the large scale structure, becomes one of the most powerful probes to test the standard model of cosmology (Hu 2002; Huterer 2010; Hamana et al. 2020; Asgari et al. 2021; Amon et al. 2022; Secco et al. 2022).

The prevalent method of cosmological parameter analysis based on cosmic shear currently relies on tomographic binning (Hu 1999) and measuring the two-point correlation function (2PCF) of the source galaxy shapes (e.g., Hildebrandt et al. 2020; Hamana et al. 2020; Amon et al. 2022; Secco et al. 2022). For this approach to cosmological analysis, the distribution of the source galaxy distances along the line-of-sight, commonly known as the sample redshift distribution $n(z)$, is an important quantity for forward modeling the auto- or cross-2PCF of cosmic shear within or between tomographic bins, respectively (e.g., Huterer et al. 2006).

Due to the expense of spectroscopic observations for galaxy samples at the depths of current imaging surveys, weak lensing measurements typically rely on multi-band photometric redshifts as their initial source of redshift information, having only limited and typically not representative training samples with spectroscopic redshifts. There two primary categories of photometric redshift estimation methods (for a review, see Salvato et al. 2019) are as follows: (a) template fitting, which is based on finding the best-fit spectral energy distributions (SED) template by fitting to the broad-band photometry; and (b) machine learning methods, which use the training sample to learn a relationship between redshift, photometry, and po-

tentially other information, e.g., morphological parameters. The outputs of these photo-z methods are normally probability density functions for individual galaxies, which we will call $p(z)$.

Deriving the aforementioned sample redshift distributions $n(z)$ based on uncertain and potentially biased individual galaxy $p(z)$ is highly non-trivial (e.g., Malz 2021), as doing so properly requires deconvolution of the uncertainties and correction for any biases. Methods for reconstructing properly calibrated $n(z)$ include direct calibration based on magnitude re-weighting to match a reference sample with known redshifts (DIR; Cunha et al. 2009) and cross-correlating spectroscopic samples and photometric samples (CC; Newman 2008; Sánchez & Bernstein 2019; Rau et al. 2020). Additionally, some methods aim to estimate $n(z)$ directly from photometric observables instead of using photometric redshifts (see, e.g., Lima et al. 2008), with the latter branching into machine learning and related approaches in recent years (Malz et al. 2018; Henghes et al. 2021). More recent work permits the combination of the $p(z)$ with a regularized deconvolution of their uncertainty, in combination with the CC method (Rau et al. 2022) – a method that is being applied in practice by Rau et al., *in prep.* to data from the HSC survey.

Since the cosmic shear signal is sensitive to the sample redshift distribution, it is necessary to carefully model the uncertainties on $n(z)$ and marginalize over them for the current and upcoming surveys (Malz & Hogg 2022). The marginalization is not a trivial task, since the uncertainties on $n(z)$ are often modeled in a high dimensional space, making attempts to run a full MCMC extremely computationally intensive. Therefore, several methods have been used to approximately marginalize over the redshift distribution uncertainties. This includes allowing just a shift in the mean redshift of the $n(z)$ for each tomographic bin, a method that has been adopted in many cosmology analyses (e.g., Hamana et al. 2020; Joudaki et al. 2020; Amon et al. 2022; Secco et al. 2022). In other cases, methods have been developed to marginalize over realistic uncertainties on $n(z)$, for example by (a) combining 750 MCMC chains each run with a different random realization sampled from the prior for $n(z)$ (Hildebrandt et al. 2017), (b) analytically approximating the likelihood function on the redshift nuisance parameters (Hadzhiyska et al. 2020; Stölzner et al. 2021), and (c) ranking $n(z)$ realizations in a lower dimensionality latent space to reduce the number of nuisance parameters (Cordero et al. 2022).

In this study, we develop and apply methodology to systematically compare the performance of methods of $n(z)$ uncertainty marginalization for cosmic shear. Our goal is to quantify tradeoffs such as systematic bias, credible uncertainty estimation, and computational costs. For this purpose, we start by presenting the new resampling approaches for marginalizing over uncertainties in the sample redshift distribution $n(z)$. We apply the new method and compare it with several existing approaches in the literature, in the context of cosmic shear with the three-year HSC shear catalog (HSC Y3; Li et al. 2022). We consider the above-mentioned tradeoffs and make a recommendation for methodology that would be appropriate for cosmology analysis of the HSC Y3 shear catalog.

The structure of this paper is as follows. In Section 2, we provide brief background on $n(z)$ uncertainty modeling and the tomographic 2PCF cosmological analysis of cosmic shear. In Section 3, we outline the approaches we will explore for marginalization over ensemble redshift uncertainties, including the new method and several pre-existing methods in the literature. We also explain the specific setup for the cosmological analysis we use for comparing these methods. In Section 4, we show the results for the cosmological parameter inference using multiple approaches for redshift uncertainty marginalization. In Section 5, we summarize our findings in this paper and discuss their practical implications.

2 BACKGROUND

In this section, we provide the background that motivates this study. In Section 2.1, we introduce the weak lensing shear analysis paradigm of this paper, and describe the modeling and marginalization of redshift distribution uncertainties in previous shear analyses. Section 2.2 describes our flexible parametrization for the sample redshift distribution and discusses our choice of prior on the associated sample redshift distribution model parameters.

2.1 Weak Lensing Shear Analysis

In this work, we discuss marginalization over the $n(z)$ uncertainties in a tomographic weak lensing shear analysis (Hu 1999) based on the two-point correlation function (2PCF;e.g., Peebles 1980; Kaiser et al. 2000; Fu et al. 2008; Huff et al. 2014). In this section, we provide a brief background of this analysis paradigm. We define terms, e.g., the data vector (observable) and its covariance matrix, and the forward model that predicts the theoretical value of the observable given cosmological and nuisance parameters. Among nuisance parameters, we emphasize the parameterization of the redshift distribution uncertainties, which is the focus of this paper.

The goal of the weak lensing shear analysis is to extract information about the cosmological model from the shear 2PCF. The cosmic shear observable that is commonly measured in real space analyses (e.g., Hamana et al. 2020; Joudaki et al. 2020; Amon et al. 2022; Secco et al. 2022) is the correlation functions of the observed galaxy shears $\xi_{\pm}^{ij}(\theta)$, where θ is the angular separation of the galaxies, and i and j are the indices of the tomographic bin pair. The data vector is obtained by concatenating $\xi_{\pm}^{ij}(\theta)$ from different tomographic bin pairs across all angular bins used for the measurement.

The observed data vector $\boldsymbol{D}$ is compared to the theoretical data vector $\boldsymbol{T}$, which is predicted by a forward modeling pipeline that considers the cosmological parameters and systematic biases and uncertainties, e.g., the uncertainties on the redshift distribution, and the intrinsic alignment of galaxy shapes due to gravitational tidal effects (IA; Croft & Metzler 2000; Heavens et al. 2000). The log-likelihood of a model parameter vector $\boldsymbol{\Omega}$ is computed by

$$\log(\mathcal{L}(\boldsymbol{\Omega}|\boldsymbol{D})) = (\boldsymbol{D} - \boldsymbol{T}(\boldsymbol{\Omega}))\boldsymbol{\Sigma}^{-1}(\boldsymbol{D} - \boldsymbol{T}(\boldsymbol{\Omega}))^{T}, \quad (1)$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{D}$. MCMC samplers such as MultiNest (Feroz & Hobson 2008; Feroz et al. 2009, 2019) are used to efficiently sample over the parameter space and provide parameter inferences based on the likelihood in Eq. (1) and the prior information on the parameters.

An important step to forward model the shear-shear 2PCF in tomographic bin pairs is to project the 3-D matter power

spectrum $P(k, z)$ to the 2-D angular shear power spectrum C_ℓ^{ij}. Under the Limber approximation, the angular shear power spectrum (Seljak 1998; Hu 1999) between bins i and j is

$$C_\ell^{ij} = \int \frac{\mathrm{d}\chi}{\chi^2} P(\ell/\chi; z(\chi)) q^i(\chi) q^j(\chi), \tag{2}$$

where $P(\ell/\chi; z(\chi))$ is the matter power spectrum at z. $q^i(\chi)$ and $q^j(\chi)$ are the corresponding lensing efficiency function for tomographic bins i and j. $q^i(\chi)$ is directly determined by the underlying redshift distribution $n^i(z)$:

$$q^i(\chi) = \frac{3\Omega_m H_0^2}{2c^2} \frac{\chi}{a(\chi)} \int_\chi^{\chi_h} \mathrm{d}\chi' n^i(\chi'(z)) \frac{\chi' - \chi}{\chi'}, \tag{3}$$

where Ω_m is the matter density parameter, H_0 is the Hubble constant, χ is the comoving radial distance, a is the scale factor, and c is the speed of light (e.g., Kilbinger 2015; Krause & Eifler 2017). Here we have used the formalism for a flat geometry. We can see that $n^i(z)$ is a key factor determining the angular shear power spectrum, which itself directly determines the shear-shear 2PCF $\xi_\pm^{ij}$ (Bartelmann & Schneider 2001; Joachimi & Bridle 2010). Under the flat-sky approximation, $\xi_\pm^{ij}$ is expressed as

$$\xi_\pm^{ij}(\theta) = \frac{1}{2\pi} \int \mathrm{d}\ell \ell C_\ell^{ij} J_{2\mp2}(\ell\theta), \tag{4}$$

where J_n is the n-th order Bessel function of the first kind. This deep connection between the redshift distribution and the cosmic shear observables is the reason why it is important to marginalize over the uncertainties on $n(z)$ to recover credible cosmological parameter constraints.

The sample redshift distribution $n(z)$ is often modeled as arrays of histogram bin heights $\boldsymbol{\phi}_{\mathrm{nz}}$, as is further described in Sec. 2.2. Since sampling in high dimensional parameter spaces is very computationally expensive, it may not be possible to model the sample redshift distribution uncertainties in every redshift bin that $n(z)$ is estimated on. A majority of previous shear analysis (e.g., Hamana et al. 2020; Joudaki et al. 2020; Amon et al. 2022; Secco et al. 2022) parameterized the redshift distribution of bin i by allowing its mean redshift to shift,

$$n^i(z) = n^i(z - \Delta z^i). \tag{5}$$

With the shift model, the number of free parameters is equal to the number of the tomographic bins. The priors on these parameters are determined by the prior distributions of the calibrated $n(z)$. The shift model tremendously reduces the number of parameters compared to use of all histogram bin heights $\boldsymbol{\phi}_{\mathrm{nz}}$, though it suffers from a limited number of degrees of freedom compared to the realistic $n(z)$ uncertainties. With cosmic shear analysis becoming increasingly systematics-dominated as the statistical uncertainties become smaller, various methods have been introduced to marginalize over a more realistic estimate of the $n(z)$ prior. In Hildebrandt et al. (2020), 750 realizations were drawn from the $n(z)$ prior, after which cosmic shear analyses were run on each realization. The chains were then directly concatenated to derive constraints on the cosmological parameters, including their uncertainties. Stölzner et al. (2021) applied the Laplace approximation to the prior of the redshift parameters and assumed the likelihood function is a multivariate Gaussian, thereby analytically marginalizing over the redshift parameter using the self-calibration algorithm. In Cordero et al. (2022), realizations of $n(z)$ were drawn from the prior distribution, then mapped into a lower-dimensional latent space, within which the likelihood function is smooth.

In this paper, we revisit some of the methods mentioned above to marginalize over the $n(z)$ uncertainties, carrying out tests on mock cosmic shear analyses. We propose a new method of marginalizing over the $n(z)$ uncertainties based on statistical principles. By comparing the new method to other options, we aim to provide the optimal approach for the HSC Y3 cosmic shear analysis.

2.2 Prior Specification on the Sample Redshift Distribution

In this section, we briefly summarize how a prior on the sample redshift distribution was specified. For a discussion on the $n(z)$ inference methodology we refer to Rau et al. (in prep.).

As shown in Eq. (2), the sample redshift distribution enters the modelling of two point functions via the transfer function in Eq. (3). The entire redshift range is subdivided into N_{bins} histogram bins, and the sample redshift distribution in the i-th tomographic bin is parametrized as

$$n^i(z; \boldsymbol{\phi}^i_{\mathrm{nz}}) = \sum_{k=1}^{N_{\mathrm{bins}}} \boldsymbol{\phi}^i_{\mathrm{nz,k}} \mathbf{1}(z \in [z_L^k, z_R^k]), \tag{6}$$

where $[z_L^k, z_R^k]$ denotes the left/right edges of histogram bin k. $\boldsymbol{\phi}^i_{\mathrm{nz,k}}$ is the k-th histogram bin height in the i-th tomographic bin. $\mathbf{1}$ is the indicator function. The distinction between 'histogram bin' and tomographic bin is as follows: the former denotes the bins of the histogram parametrization, the latter denotes the selection bins of the tomography. Eq. 6 defines the histogram heights vector $\boldsymbol{\phi}^i_{\mathrm{nz}}$ as the parameters of a linear basis function model for the sample redshift distribution with tophat basis functions.

The prior $p(\boldsymbol{\phi}^i_{\mathrm{nz}})$, i.e., uncertainties on the sample redshift distribution histogram bin heights in the i-th tomographic bin, is inferred using an extension of the methodology developed in Rau et al. (2022). It combines information from both spatial cross-correlations of a reference sample with spectroscopic redshifts and a sample with photometric redshift information. We reiterate that a future publication will provide more details of the inference methodology (Rau, et al, *in prep.*). The method utilizes the 'S16A CAMIRA-LRG sample' (Ishikawa et al. 2021a), a sample of Luminous Red Galaxies selected using the CAMIRA algorithm (Oguri 2014) from the HSC data observed in the first observing season of 2016, as a reference sample. This choice can be motivated by the accurate photometric redshift estimates that are available for the LRGs (relative to the photometric redshift errors in the full HSC S16A sample), and a sufficiently high number density.

The spatial cross-correlation between the CAMIRA-LRG sample (c) and a photometric sample (p) can be predicted as

$$\boldsymbol{w_{\mathrm{sc}}} \propto \boldsymbol{\phi}_{\mathrm{nz,p}}\, \boldsymbol{b_p}\, \boldsymbol{b_c}\, \boldsymbol{w_{\mathrm{DM}}}\,, \tag{7}$$

where $\boldsymbol{\phi}_{\mathrm{nz,p}}$ denotes the parameters of the sample redshift distribution, $(\boldsymbol{b_p}/\boldsymbol{b_c})$ denote the galaxy-dark matter bias parameters of the (photometric/CAMIRA-LRG) samples in each redshift bin and $\boldsymbol{w_{\mathrm{DM}}}$ denotes the dark-matter contribution to the cross-correlation signal. We present a sim-

plified vector notation, where the elements in Eq. 7 correspond to the cross-correlation measurements in each redshift bin, obtained by measuring the correlation amplitude within a spatial annulus of physical distance as described in Morrison et al. (2017). Using the auto-correlation of the CAMIRA-LRG galaxies the method fits the linear bias model $b_c(z) = b_0(1+z)$, where $b_0 = 1.06 \pm 0.03$, consistent with previous measurements from Ishikawa et al. (2021b). The covariance of the cross-correlation likelihoods is estimated using bootstrap resampling and approximated to be diagonal. This is done for simplicity and can be an inaccurate approximation due to the high correlation of neighboring bins. The method uses THE-WIZZ[1] (Morrison et al. 2017) for the cross-correlation measurements, and selects a scale annulus of $1.5 - 5.0$ Mpc in analogy to Gatti et al. (2021).

We include information from the photometry into the inference by combining the individual galaxy redshift uncertainties of a set of models. Our model set consists of a template fitting code MIZUKI (Tanaka 2015) that defines a likelihood, empirical codes MLZ[2] (Carrasco Kind & Brunner 2013) and EPHOR (Tanaka et al. 2018) that define a conditional probability density function obtained on a training set and FRANKEN-Z[3] (Speagle et al. 2019) that uses a flux-error weighted score function to map training set objects to galaxies in the photometric dataset. We refer to Tanaka et al. (2018) for a summary of the different methodologies that are available to us. We note that the machine learning-based algorithms do not produce likelihoods (unlike SED fitting techniques). However we will treat their estimates as likelihoods within this framework and refer to a future publication for a description of the technical details.

Following the methodology developed in Rau et al. (2022), we infer posteriors of sample redshift distributions as shown in Fig. 1 using information from both the cross-correlation data vector and the photometry of galaxies. The horizontal axis shows the redshift, the vertical the normalized sample redshift distribution. The legend lists the redshift ranges selected on the best fitting redshift derived using the MIZUKI template fitting code that we use to define the tomographic bins. The error contours correspond to the 68% confidence intervals. The aforementioned posteriors of sample redshift distributions constructed using the joint likelihood of spatial cross-correlations and photometry is then used as the prior distribution on the sample redshift distribution in the following analysis. We neglect here the covariance between the spatial cross-correlations and the lensing observables.

In this work, we assume that the uncertainties in the ensemble redshift distribution for the HSC three-year and full analysis do not significantly decrease compared with those for the first-year HSC analysis. Constraints on the sample redshift distribution are limited by (a) practical issues such as the redshift range of the LRG sample and our knowledge of the galaxy-dark matter bias; (b) the model uncertainty between photometric redshift codes, estimated using the COSMOS2015 field (Laigle et al. 2016). The modeling uncertainty is limited by the cosmic variance, and is independent of the survey area, therefore will not decrease for the HSC three-year analysis compared with the first-year analysis. As a result, the redshift uncertainties are expected to decrease much more slowly than the cosmic shear covariance matrix as the survey area grows.

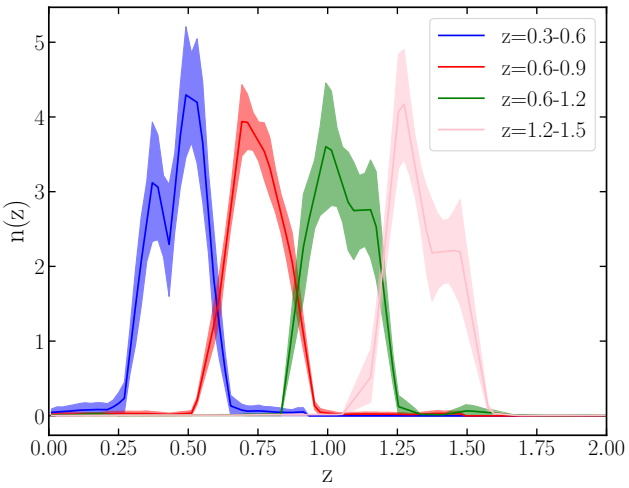

Figure 1. The sample redshift distribution estimated by cross correlation with 4 tomographic bins, for HSC S16A (Rau, et al., *in prep.*) The shaded regions represent the 68% confidence intervals of the distributions.

3 METHODS

In this section, we describe the methods used to carry out this work. In Section 3.1, we describe our parameter inference pipeline, implemented using CosmoSIS (Zuntz et al. 2015). In Section 3.2, we describe the methods to marginalize over the $n(z)$ uncertainties during the cosmological parameter inference. In addition to employing existing approaches from the literature, we also propose a new method for marginalizing $n(z)$ uncertainties for cosmic shear analysis: a statistically accurate formulation for sampling from the $n(z)$ covariance.

The key terminology used for redshift distribution and statistical inference throughout this section, their mathematical symbols, and description are listed in Table 1.

3.1 Cosmological forward modeling

In this section, we describe the cosmic shear forward modeling process, including the cosmological model, the astrophysical model, and other nuisance parameters, for computing the mock data vector and parameter inference. For an initial exploration, we considered a 2-parameter ΛCDM model that only varies Ω_m and σ_8. We then considered a full analysis with 5 ΛCDM parameters, 2 astrophysical nuisance parameters, $N_{\rm tomo} = 4$ multiplicative bias parameters and 2 PSF systematics parameters, for a total of 13 parameters. $N_{\rm tomo} = 4$ additional parameters were added for marginalizing over $n(z)$ uncertainties for the shift model. The modeling pipeline used CosmoSIS (Zuntz et al. 2015), which is a well-tested and validated platform for cosmological inference (e.g., in Abbott et al. 2022).

The cosmological model is described in Section 3.1.1, while

[1] https://github.com/morriscb/The-wiZZ/
[2] https://github.com/mgckind/MLZ
[3] https://github.com/joshspeagle/frankenz

Terminology	Symbol	Description
$n^i(z)$ prior	$P(\boldsymbol{\phi}^i_{\rm nz}\|\boldsymbol{\alpha}^i)$	The prior on the $n^i(z)$ histogram bin heights in the i-th tomographic bin. Specifically, we adapt the posterior in Rau et al. (in prep.) $P(\boldsymbol{\phi}^i_{\rm nz}\|\boldsymbol{\alpha}^i)$ parameterized on the Dirichlet parameter $\boldsymbol{\alpha}^i$ for the i-th tomographic bin, as the prior, which is described in Section 2.2. We sometimes refer to this as $n(z)$ prior, when the tomographic bin is not specified.
Average $n^i(z)$	$\langle\boldsymbol{\phi}^i_{\rm nz}\rangle$	The average histogram bin heights for $n^i(z)$ in the i-th tomographic bin, averaged over 10000 realizations of $\boldsymbol{\phi}^i_{\rm nz}$ sampled from the $n(z)$ prior.
Mean redshift	$\langle z^i\rangle$	The mean redshift of the i-th tomographic bin, calculated by $\langle z^i\rangle = \int zP(\boldsymbol{\phi}^i_{\rm nz}\|\boldsymbol{\alpha}^{\boldsymbol{i}})\mathrm{d}z$, where $P(\boldsymbol{\phi}^i_{\rm nz}\|\boldsymbol{\alpha}^i)$ is the $n^i(z)$ prior of the samples in the i-th tomographic bin.
Data vector	$\boldsymbol{D}$	Shear data vector $\boldsymbol{D} = [\boldsymbol{\xi}^{\boldsymbol{ij}}_+, \boldsymbol{\xi}^{\boldsymbol{ij}}_-]$, where ij is ordered in $[11, 12, 13, 14, 22, 23, 24, 33, 34, 44]$. The generation of data vector is described in Section 3.1.3.
Covariance (matrix)	$\boldsymbol{\Sigma}$	Covariance matrix of the data vector $\boldsymbol{D}$, $\Sigma_{ij} = \langle D_i D_j\rangle$. The covariance matrix used in this work is described in Section 3.1.4.
Inference posterior	$P(\boldsymbol{\Omega}\|\boldsymbol{D})$	The posterior distribution on the cosmological and astrophysical parameters $\boldsymbol{\Omega}$ after marginalizing over the nuisance parameters. In this paper, we specifically consider the $n(z)$ parameters as the nuisance parameters.
Log evidence	$\log(P(\boldsymbol{D}\|\boldsymbol{\phi}_{\rm nz,s}))$	The log-evidence of a particular realization of the $\boldsymbol{\phi}_{\rm nz,s}$, expressed in Eq. (12).
Number of tomographic bins	$N_{\rm tomo}$	The number of tomographic bins, which results in the number of nuisance parameters for the multiplicative bias and shift model. In this work, $N_{\rm tomo} = 4$
Number of resampling $\boldsymbol{\phi}_{\rm nz}$	$N_{\rm sample}$	The number of realizations sampled from the $n(z)$ prior for the direct and Bayesian resampling methods, described in Sec. 3.2.2. For the full analyses in this work, $N_{\rm sample} = 250$.
Number of histogram bins	$N^i_{\rm bins}$	Number of histogram bin heights in the i-th tomographic bins. This is the same as the length of $\boldsymbol{\phi}^i_{\rm nz}$. In this work, $N^i_{\rm bins} = 18(18, 25, 20)$ for tomographic bins 1(2,3,4), respectively.

Table 1. Table of the redshift distribution and statistics terminologies used for quantities throughout Section 3. We also provide a short description of each quantity and the specific values used in this work or a reference to the section where they are described.

the astrophysical and other nuisance parameters are described in Section 3.1.2. The analysis setup (tomographic bins, angular scales, etc.) and mock data vector are described in Section 3.1.3. The sampler and covariance matrices are described in Section 3.1.4.

3.1.1 Cosmological Model

We adopted a ΛCDM cosmological model throughout this work. We computed the linear matter power spectrum using CAMB (Lewis et al. 2000; Lewis & Bridle 2002; Howlett et al. 2012), and the nonlinear matter power spectrum using the updated HALOFIT (Takahashi et al. 2012) from the original version (Smith et al. 2003). The neutrino mass Ω_ν was fixed to zero, since the weak lensing shear is relatively insensitive to it. The cosmological parameters in our model are provided in Table 2, including their fiducial values, priors, and whether they are varied or fixed in our analysis.

3.1.2 Astrophysical and Nuisance Parameters

Throughout the analysis, we used the nonlinear alignment model (NLA Krause & Eifler 2017) to model the intrinsic alignment (IA) signal (see also Hirata & Seljak 2004; Bridle & King 2007, for the development and further extension of the NLA model). In this paper, we adopted the NLA model with an additional term that includes redshift evolution of the alignment amplitude, namely,

$$A(z) = A_{\rm IA}\left[\frac{1+z}{1+z_0}\right]^\eta, \tag{8}$$

where the fiducial values and priors of the parameters $A_{\rm IA}$, η, and z_0 are shown in Table 3. In practice, the redshift evolution parameter may absorb some evolution of the source sample properties with redshift, since intrinsic alignments depend on galaxy properties.

Since the IA model in this work has redshift evolution, the intrinsic alignment model parameters may have some degeneracy with the redshift distribution $n(z)$, which motivates marginalizing over the $n(z)$ uncertainty in the analysis.

We computed the shear-shear angular power spectrum from the matter power spectrum and the input $n(z)$, using the formalism in Section 2.1. We then added the NLA model shear-IA and IA-IA angular power spectrum to the shear-shear angular power spectrum. Next, we included a per-bin multiplicative shear bias into the observed shear power spec-

trum using

$$\hat{C}_\ell^{ij} = (1 + m^i + m^j) C_\ell^{ij}, \tag{9}$$

where m^i and m^j are the multiplicative biases of bins i and j, respectively. We used Eq. (4) to compute the shear-shear correlation function $\xi_\pm^{ij}$.

Finally, we employed a simple model for the additive shear biases at the correlation function level. We included the PSF leakage term α and PSF shape error term β, using the same model as in Hamana et al. (2020). Our model of the shear-shear correlation function with PSF systematics is

$$\xi_\pm^{ij,\mathrm{model}} = \xi_\pm^{ij} + \alpha^2 \langle e_\mathrm{PSF} e_\mathrm{PSF} \rangle + \alpha\beta \langle e_\mathrm{PSF} \delta e_\mathrm{PSF} \rangle + \beta^2 \langle \delta e_\mathrm{PSF} \delta e_\mathrm{PSF} \rangle, \tag{10}$$

where e_PSF and δe_PSF are the PSF shape and the modeling error of the PSF shape, respectively.

Table 3 lists the astrophysical and other nuisance parameters, with their fiducial values, priors, and whether they are varied or fixed in our analysis.

3.1.3 Analysis settings and mock data vector

In this work, we used 4 tomographic bins, resulting in 10 tomographic bin pairs. We adopted the angular binning used in the real-space cosmic shear analysis of the first-year HSC catalog (Hamana et al. 2020), i.e., 9 angular bins between 8.06 arcmin and 50.89 arcmin for ξ_+, and 8 angular bins between 32.11 arcmin and 160.93 arcmin for ξ_-. Our data vector $\boldsymbol{D}$, which includes ξ_+ and ξ_-, has a length of 170.

We generated mock data vectors using the forward modeling pipeline described above. To be able to compare the recovered parameter values with their true values, we did not add noise to the data vectors.

We used the Planck results in Planck Collaboration et al. (2020) for the fiducial cosmological parameters in Table 2. For the IA parameters in Table 3, we adopted typical integer values for the amplitude A_IA and redshift power η, and $z_0 = 0.62$ for the pivot redshift[4] (Troxel et al. 2018; Hamana et al. 2020). We adopted the prior on α and β from Hamana et al. (2020), and set the fiducial values to zero.

Our mock shear data vector was generated by averaging the $\xi_\pm$ over 1000 realizations of $n(z)$ sampled from its prior. Note that the auto-correlation $\xi_\pm^{ii}(\langle n(z) \rangle) \neq \langle \xi_\pm^{ii}(n(z)) \rangle$, with up to 0.75% difference, as is demonstrated in Appendix A. Therefore, we cannot simply use the mean value of the $n(z)$ prior to generate the mock data vector.

3.1.4 Sampler and Covariance Matrices

We sampled the parameter space and estimate the Bayesian evidence using MultiNest (Feroz & Hobson 2008; Feroz et al. 2009, 2019), due to its rapid speed for relatively accurate evidence evaluation in constant efficiency mode[5]. We fixed the efficiency to 0.1, which is the default value for MultiNest, throughout this work. The log-likelihood of the model is computed by Eq (1), with the corresponding covariance matrices.

In this work, we carried out our analyses with two covariance matrices: (a) We estimated the covariance matrix for cosmic shear using the HSC three-year shear catalog. For this purpose, we divided every element in the HSC first-year covariance $\boldsymbol{\Sigma}_{y1}$ (Hamana et al. 2020) by 3, since the survey area is roughly 3 times larger. We denote this covariance matrix as $\boldsymbol{\Sigma}_{y3} = \boldsymbol{\Sigma}_{y1}/3$. (b) We estimated the covariance matrix for cosmic shear with the full HSC survey, which is roughly 10 times the area of the first-year catalog. We denote this covariance as $\boldsymbol{\Sigma}_\mathrm{full} = \boldsymbol{\Sigma}_{y1}/10$. There are several significant limitations of this approximation to the future HSC analyses: (a) We decreased the covariance by a factor of the increase in survey area, without considering that the survey footprint has become considerably more contiguous, so the survey edge effects become less important. (b) We adopted the same angular binning and scale cuts as for the HSC first-year analysis, while those cuts are likely to be different for the upcoming three-year analysis and future analyses.

However, we used the covariance matrix of the $n(z)$ prior from the first-year HSC shape catalog when analyzing the three-year and full data vector. In the real analyses, the covariance of the $n(z)$ for the three-year and full catalogs is likely to decrease. However, it is a systematics-dominated quantity, so its uncertainty will not decrease with area as

Parameter	Fiducial	Prior	2-p	full analysis
σ_8	0.824	$U[0.4, 1.2]$	✓	✓
Ω_b	0.0489	$U[0.03, 0.07]$		✓
n_s	0.967	$U[0.87, 1.07]$		✓
h_0	0.677	$U[0.55, 0.9]$		✓
Ω_m	0.311	$U[0.1, 0.8]$	✓	✓
τ	0.0561	const.		
Ω_ν	0.0	const.		
w	-1.0	const.		
w_a	0.0	const.		

Table 2. Fiducial values and priors of the cosmological parameters used in this paper, along with whether or not they are varied (✓) or not (blank) in the two-parameter (2-p) and full analysis. $U[a, b]$ represents a uniform distribution from a to b.

Parameter	Fiducial	Prior	2-p	full analysis
A_IA	1.0	$U[-5, 5]$		✓
η	0.0	$U[-5, 5]$		✓
z_0	0.62	const.		
m_1	0.0	$\mathcal{N}(0, 0.01)$		✓
m_2	0.0	$\mathcal{N}(0, 0.01)$		✓
m_3	0.0	$\mathcal{N}(0, 0.01)$		✓
m_4	0.0	$\mathcal{N}(0, 0.01)$		✓
α	0.0	$\mathcal{N}(0, 0.01)$		✓
β	0.0	$\mathcal{N}(0, 1.11)$		✓

Table 3. Fiducial values and priors of the astrophysical and nuisance parameters used in this paper, along with whether or not they are varied (✓) or not (blank) in the two-parameter (2-p) and full analysis. $U[a, b]$ represents a uniform distribution from a to b, while $\mathcal{N}(\mu, \sigma)$ represents a Gaussian distribution with mean value μ and standard deviation σ.

[4] We have used $z_0 = 0.62$ for consistency with previous analysis. However, as described in Longley et al. (2022), this choice does not affect the results much; choosing the mean redshift for the HSC survey gives consistent results.

[5] In Lemos et al. (2022), it is shown that varying the efficiency can bias the model evidence for MultiNest, therefore, we fixed the efficiency of MultiNest to eliminate this bias and for its speed over PolyChord (Handley et al. 2015a; Handley et al. 2015b)

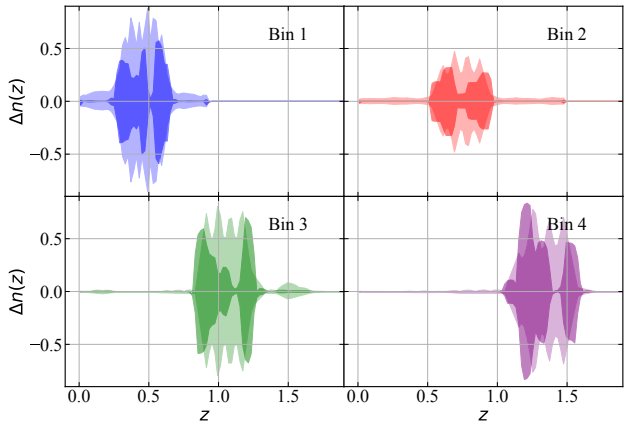

Figure 2. The 68% confidence intervals of the $n^i(z)$ uncertainties of the fiducial $n(z)$ prior (light shaded regions) and the shift model (dark shaded regions). The shift model generates an unrealistic distribution of $n(z)$ uncertainties, underestimating the uncertainty at most redshifts but overestimating it around the edges of bins 2, 3 and 4.

rapidly as does the cosmic shear data vector. Our choice to keep it fixed represents a conservative assumption regarding our ability to understand and control systematic biases and uncertainties in the photometric redshift estimation and the cross-correlation calibration of $n(z)$. As a result of this choice, the impact of $n(z)$ uncertainty on the cosmological parameter constraints gets worse as the dataset grows.

3.2 Marginalizing over $n(z)$ uncertainty

In this section, we introduce the different approaches for marginalizing over uncertainty in the ensemble $n(z)$ that are implemented on the mock cosmic shear analysis described in Section 3.1. In Section 3.2.1, we introduce the shift model's parameterization. In Section 3.2.2, we introduce the resampling approach, i.e., marginalizing over the sample redshift distribution uncertainties by running many chains with different realizations drawn from the $n(z)$ prior. We propose a new technique for weighting the chains when combining them, based on model evidence, motivated by Bayes theorem.

The $n(z)$ prior that is marginalized over in this work is specified by the histogram bin heights $\boldsymbol{\phi}^i_{\rm nz,k}$ at the center redshift of the histogram z_k for tomographic bin i. respectively, modeled by 4 independent Dirichlet distributions. The Dirichlet distributions are parameterized by arrays $\boldsymbol{\alpha}^i$, with length equal to the number of histogram bins in the corresponding tomographic bin, specified in Section 2.2.

3.2.1 Shift Model

The shift model is a simple and approximate model for representing uncertainties in $n(z)$. It allows the sample redshift distribution to shift coherently in redshift space following Eq. (5). It is used to marginalize over $n(z)$ uncertainties in many cosmic shear analysis (e.g., Hildebrandt et al. 2020; Hamana et al. 2020; Amon et al. 2022; Secco et al. 2022). For this model, we use the average histogram bin heights $\langle\boldsymbol{\phi}^i_{\rm nz}\rangle$ as the fiducial redshift distribution, specified in row 2 of Table 1. We let the $\langle\boldsymbol{\phi}^i_{\rm nz}\rangle$ of each tomographic bin shift individually.

Parameter	Fiducial	Prior
Δz^1	0.0	$\mathcal{N}(0, 0.012)$
Δz^2	0.0	$\mathcal{N}(0, 0.01)$
Δz^3	0.0	$\mathcal{N}(0, 0.018)$
Δz^4	0.0	$\mathcal{N}(0, 0.021)$

Table 4. Fiducial values and priors used for the shift model parameterization of $n(z)$ uncertainties. The standard deviation of the Gaussian prior is calculated from the σ of $\langle z^i\rangle$ from 1000 draws from the $n(z)$ prior.

Therefore, using this model involves introducing $N_{\rm tomo} = 4$ nuisance parameters. We determined the prior on the Δz^i by computing the distribution of the mean redshift $\langle z^i\rangle$ of the tomographic bin i over 10000 realizations of histogram bin heights drawn from the $n^i(z)$ prior. We used a Gaussian distribution for the prior, with zero means and standard deviations determined by the distributions of $\langle z^i\rangle$. The priors on the shift parameters for the four tomographic bins are listed in Table 4.

In Fig. 2, we show a comparison of the $n(z)$ uncertainty included by the shift model (in dark shaded regions), versus the total uncertainty of the $n(z)$ prior (in light shaded regions). The uncertainty of the shift model is generated by shifting $\langle n^i(z)\rangle$ with Δz^i sampled from the prior listed in Table 4. Compared to the full $n(z)$ prior, the shift model underestimates the uncertainties at most redshifts, especially for redshifts where $\langle n^i(z)\rangle$ is relatively flat. The shift model also overestimates the uncertainties in the wings of the redshift distribution for some the tomographic bins. At the $n(z)$ level, the shift model is an inaccurate representation of the real uncertainty.

3.2.2 Resampling Approaches

A different approach for marginalizing over the $n(z)$ uncertainty, with fewer approximations, is to sample many realizations of histogram bin heights from the $n(z)$ prior, and run the cosmological parameter estimation process on each realization as if there are no $n(z)$ uncertainties. To incorporate the $n(z)$ uncertainties in the cosmological parameter estimates, the final step is to combine the results from the different MCMC chains. We refer to this approach as the "resampling approach".

In this work, we propose a resampling method that is based on Bayes' theorem to marginalize over the $n(z)$ uncertainty. We start by deriving the posterior on the cosmological and astrophysical parameters, $P(\boldsymbol{\Omega}|\boldsymbol{D})$, after marginalizing over the uncertainty in the $n(z)$ histogram bin heights, which we denoted $\boldsymbol{\phi}_{\rm nz}$. Here $\boldsymbol{D}$ is the observed cosmic shear data vector $\boldsymbol{D} = [\boldsymbol{\xi}_+, \boldsymbol{\xi}_-]$, see row 4 of Table. 1. This posterior is as follows:

$$\begin{aligned} P(\boldsymbol{\Omega}|\boldsymbol{D}) &= \int \mathrm{d}\boldsymbol{\phi}_{\rm nz} P(\boldsymbol{\Omega}, \boldsymbol{\phi}_{\rm nz}|\boldsymbol{D}) = \int \mathrm{d}\boldsymbol{\phi}_{\rm nz} P(\boldsymbol{\Omega}|\boldsymbol{D}, \boldsymbol{\phi}_{\rm nz}) P(\boldsymbol{\phi}_{\rm nz}|\boldsymbol{D}) \\ &= \frac{1}{P(\boldsymbol{D})} \int \mathrm{d}\boldsymbol{\phi}_{\rm nz} P(\boldsymbol{\Omega}|\boldsymbol{D}, \boldsymbol{\phi}_{\rm nz}) P(\boldsymbol{D}|\boldsymbol{\phi}_{\rm nz}) P(\boldsymbol{\phi}_{\rm nz}). \end{aligned} \tag{11}$$

The first line of the equation is based on conditional probability, and the second line is based on Bayes' theorem. Here $P(\boldsymbol{\Omega}|\boldsymbol{D}, \boldsymbol{\phi}_{\rm nz})$ is the posterior on $\boldsymbol{\Omega}$ with a specific realisation of the redshift distribution $\boldsymbol{\phi}_{\rm nz}$. $P(\boldsymbol{\phi}_{\rm nz})$ is the $n^i(z)$ prior, for

which we chose to use $P(\boldsymbol{\phi}_{\rm nz}|\boldsymbol{\alpha})$, the posterior probability distribution for the redshift distribution derived using an extension of the methodology from Rau et al. (2022). $P(\boldsymbol{D}|\boldsymbol{\phi}_{\rm nz})$ is the Bayesian evidence of the data given $\boldsymbol{\phi}_{\rm nz}$, evaluated by integrating the joint conditional probability over $\boldsymbol{\Omega}$,

$$P(\boldsymbol{D}|\boldsymbol{\phi}_{\rm nz}) = \int {\rm d}\boldsymbol{\Omega} P(\boldsymbol{D}|\boldsymbol{\Omega}, \boldsymbol{\phi}_{\rm nz}) P(\boldsymbol{\Omega}). \quad (12)$$

We rely on the MultiNest estimation to the log-evidence, which is shown to have a constant bias from the truth in Lemos et al. (2022), if the efficiency is kept fixed. This is fine for our purpose: the constant bias on the log-evidence results in a constant factor in the evidence, which is normalized out for the Bayesian weight ω_s.

We now describe how we utilize the resampling approach to estimate $P(\boldsymbol{\Omega}|\boldsymbol{D})$ in Eq. (11). We sampled $N_{\rm sample}$ realizations of the redshift distribution $\boldsymbol{\phi}_{\rm nz,s}$, where $s = 1 \ldots N_{\rm sample}$, is the index of a particular realization from the $n(z)$ prior, i.e., $P(\boldsymbol{\phi}_{\rm nz}|\boldsymbol{\alpha})$. We combined the inferred posterior distributions for each one (as represented by the MCMC chains), $P(\boldsymbol{\Omega}|\boldsymbol{D}, \boldsymbol{\phi}_{\rm nz,s})$. By doing so, we effectively evaluated the integral of Eq. (11), which can be written the form of a summation,

$$P(\boldsymbol{\Omega}|\boldsymbol{D}) = \frac{1}{N_{\rm sample} P(D)} \sum_{s=1}^{N} P(\boldsymbol{\Omega}|\boldsymbol{D}, \boldsymbol{\phi}_{\rm nz,s}) P(\boldsymbol{D}|\boldsymbol{\phi}_{\rm nz,s}), \quad (13)$$

where $\boldsymbol{\phi}_{\rm nz,s}$ is the sth sample of the redshift distribution. Based on Eq. (13), we designed a Bayesian weight ω_s for combining the posteriors $P(\boldsymbol{\Omega}|\boldsymbol{D}, \boldsymbol{\phi}_{\rm nz,s})$ that satisfies the following two conditions:

$$\omega_s \propto P(\boldsymbol{D}|\boldsymbol{\phi}_{\rm nz,s}) \quad (14)$$

$$\sum_{s=1}^{N_{\rm sample}} \omega_s = 1. \quad (15)$$

Finally, the marginalized posterior of $\boldsymbol{\Omega}$ from the Bayesian resampling can be expressed as

$$P(\boldsymbol{\Omega}|\boldsymbol{D}) = \sum_{s=1}^{N} P(\boldsymbol{\Omega}|\boldsymbol{D}, \boldsymbol{\phi}_{\rm nz,s}) \omega_s. \quad (16)$$

Note that the constant $1/(N_{\rm sample} P(\boldsymbol{D}))$ in Eq. (11) is absorbed in ω_s since summation of ω_s is normalized to 1. This weight ω_s, which is proportional to the Bayesian evidence shown in Eq (12), preserves Bayes' theorem, effectively down-weighting the $n(z)$ realizations that are not likely to generate the cosmic shear data vector $\boldsymbol{D}$. A similar resampling approach was used in Hildebrandt et al. (2020); however, the MCMC chains were concatenated with equal weights, which does not preserve Bayes' theorem. We therefore call our approach "Bayesian resampling", and call the method from Hildebrandt et al. (2020) "direct resampling", throughout the paper.

In principle, with enough samples of the redshift distribution, the Bayesian resampling approach should accurately marginalize over the full prior on $n(z)$ in the cosmic shear analysis, giving more credible parameter constraints than simplified parameterizations, e.g., the shift model. However, it does have its drawbacks: (a) it is computationally intensive to run the full analysis for $N_{\rm sample}$ times, where $N_{\rm sample}$ is the number of redshift distribution samples, (b) it requires the sample redshift distribution $n(z)$ to have a well-defined probability distribution from which samples can be drawn, which might not be the case for some surveys depending on how they infer the ensemble $n(z)$.

3.2.3 Methods Summary

In this section, we briefly summarize the methods for marginalizing over $n(z)$ uncertainty in this work, including the notation and terminology of the marginalization methods.

- No $n(z)$ Uncertainty: We use the average histogram bin height of the $n(z)$ prior, $\langle \boldsymbol{\phi}^i_{\rm nz} \rangle$, as the sample redshift distribution, without marginalizing over any $n(z)$ uncertainties. This is the baseline that other methods are compared to.
- Direct Resampling: We sample $N_{\rm sample}$ realizations of $\boldsymbol{\phi}_{\rm nz,s}$ from the $n(z)$ prior and run cosmological parameter inference on each realization without explicitly accounting for the evidence of the $\boldsymbol{\phi}_{\rm nz,s}$. The chains for different $\boldsymbol{\phi}_{\rm nz,s}$ are then combined with equal weights, implicitly incorporating the $n(z)$ uncertainties into the resulting parameter constraints.
- Bayesian Resampling: This method begins as does direct resampling, but the chains for different $\boldsymbol{\phi}_{\rm nz,s}$ are weighted by their Bayesian evidence, as described in Section 3.2.2.
- Shift Model: The average histogram bin heights $\langle \boldsymbol{\phi}^i_{\rm nz} \rangle$ is allowed to shift on redshift individually for each tomographic bins, resulting in $N_{\rm tomo} = 4$ nuisance parameters for marginalizing over redshift uncertainty, as described in Section 3.2.1.

3.3 Probability Integral Transformation

In this section, we introduce our validation method for the parameter inference results. We note that validating the probability calibration of inference results is an integral part of testing novel inference methodology. Since the 'true value' of a parameter of interest is viewed in the Bayesian picture as a random variable, the posteriors derived using an inference methodology have to present an accurate estimate of that unknown distribution.

A necessary requirement is that our inference adheres to Bayes theorem, which forms the basis of the statistical test presented in the following. To test this, we perform a statistical test based on the probability integral transformation (PIT; Casella et al. 2002; Schmidt et al. 2020) to test the validity of the inference statistics. We perform PIT on the cumulative density function (CDF) of S_8, as it is the parameter that the cosmic shear constrains most precisely. The true posterior of the inferred S_8 can be yielded by Bayes' theorem:

$$P(S_8|\boldsymbol{D}) = \frac{P(\boldsymbol{D}|S_8) P(S_8)}{P(D)}. \quad (17)$$

We define the CDF of S_8 to be

$$F_{S_8}(S_8) = \int_0^{S_8} {\rm d}S_8' P(S_8'|\boldsymbol{D}). \quad (18)$$

According to the PIT theorem, a random variable Y drawn from the distribution of $F_{S_8}(S_8)$ in Eq. (18), has a range of $[0, 1]$, and the CDF of Y follows

$$F_Y(y) = y, \quad (19)$$

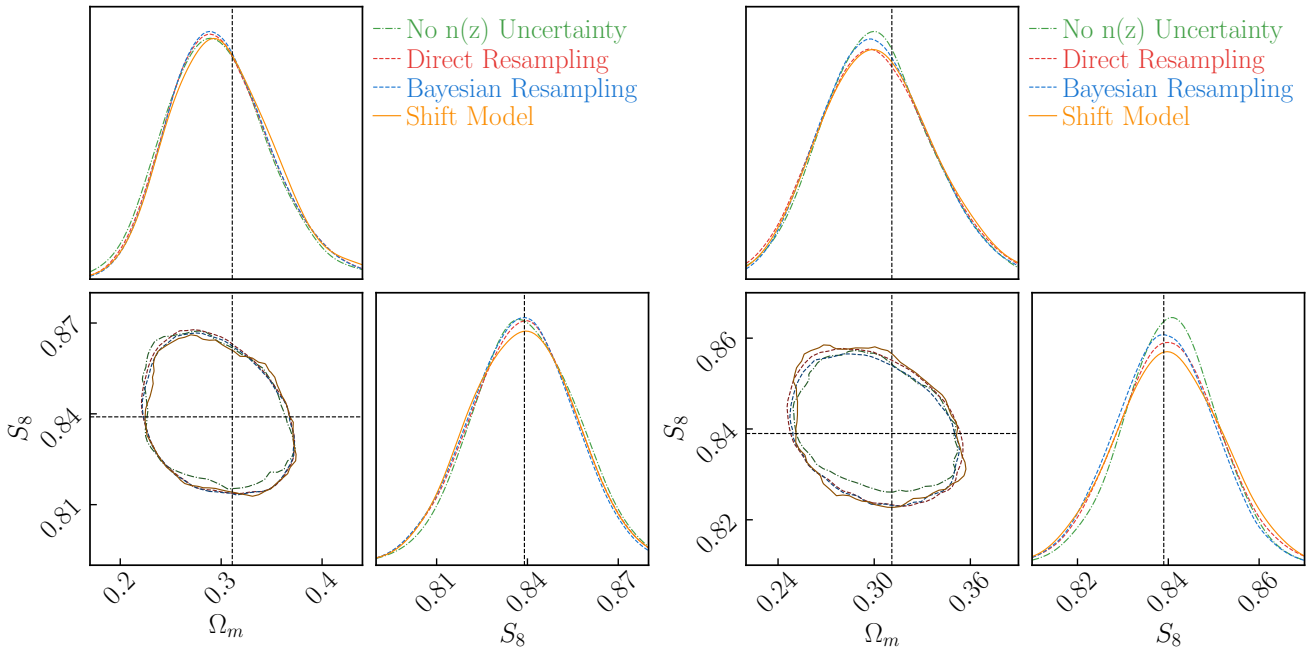

Figure 3. Parameter constraints for the full analyses, with the three-year covariance matrix Σ_{y3} (left), and the full-data covariance matrix $\Sigma_{\rm full}$ (right), with parameters listed in Table 2 and 3, and with $n(z)$ uncertainty marginalized using three different approaches. The green contour shows the results using the average $n(z)$ with no uncertainties, while the red and blue contours show the results using the direct and Bayesian resampling approaches described in Section 3.2.2. The orange contours use the shift model parameterization, with $N_{\rm tomo} = 4$ nuisance shift model parameters, described in Section 3.2.1. The dashed lines are the true values used to generate data vector. This plot is made using CHAINCONSUMER (Hinton 2016)

where y is a specific value of Y between $[0, 1]$.

To test the credibility of our inference pipeline, we estimate the CDF of S_8, namely, $\hat{F}_{S_8}(y)$, by generating pairs of data vectors $\boldsymbol{D}^\mu$ and S_8^μ, where $\mu = 1, 2 \ldots N_{\rm PIT}$, and $N_{\rm PIT} = 50$. For each μ, we sample a pair of $(\Omega_m^\mu, \sigma_8^\mu)$ with the uniform prior $U_{\Omega_m}[0.2, 0.4]$ and $U_{\sigma_8}[0.7, 1.0]$, and compute the corresponding $S_8^\mu = \sigma_8^\mu \sqrt{\Omega_m^\mu / 0.3}$. We first produce a noiseless data vector using the average $n(z)$, and then add a random noise realization generated using $\boldsymbol{\Sigma}_{\rm y3}$. The noisy data vector is denoted $\boldsymbol{D}^\mu$.

We run the full inference pipeline on each pair of $\boldsymbol{D}^\mu$ and S_8^μ, which generates a posterior $P^\mu(S_8|\boldsymbol{D}^\mu)$. For each μ, we estimate

$$\hat{Y}^\mu = F^\mu_{S_8}(S_8^\mu), \tag{20}$$

where $F^\mu_{S_8}$ is the CDF of the S_8 posterior for the μ-th sample. We compare the CDF of $\hat{Y}$ with the expected uniform distribution in Sec. 4.2.

By conducting the PIT test, we are checking that the posterior distribution of the cosmological parameters inferred in the inference pipeline is statistically consistent with the true posterior given by Bayes' theorem. This is a crucial validation test for the results of this work, since our conclusion that compares marginalization methods relies on accurate posterior errorbars of the inferred parameters. Crucially, this test must be done using data vectors with noise added according to the covariance matrix, since that noise is what broadens the parameter distribution that we are trying to infer.

4 RESULTS

In this section, we show the results of forecasting cosmic shear analyses with different marginalization approaches, following the methods outlined in Sec. 3. In Section 4.1, we show results of the full analyses, where 5 cosmological parameters, 2 IA parameters, 4 multiplicative biases, 2 PSF systematics parameters, and any parameters used to parametrize uncertainty in $n(z)$ are jointly fit. In Section 4.2, we show the PIT validation on noisy data vectors. In Section 4.3, we compare the results in this work to that of other work.

4.1 Full analysis

In this section, we show the results of the full cosmic shear analysis on the noiseless mock data vector using the redshift marginalization methods listed in Section 3.2.3. We consider 5 cosmological parameters, listed in Table 2 and explained in Section 3.1.1. Additionally, we consider 2 IA parameters, $N_{\rm tomo} = 4$ multiplicative biases, and 2 PSF systematics parameters, listed in Table 3 and explained in Section 3.1.2.

We ran a baseline analysis with the average $n(z)$ and no marginalization for comparison, and three marginalization approaches: the direct and Bayesian resampling, described in Section 3.2.2, and the shift model, described in Section 3.2.1. For the resampling approach, we ran $N_{\rm sample} = 250$ chains for both $\boldsymbol{\Sigma}_{y3}$ and $\boldsymbol{\Sigma}_{\rm full}$ covariance matrices. There are $N_{\rm tomo} = 4$ nuisance parameters for the shift model, for which the fiducial values and priors are listed in Table 4.

In the top row of Fig. 3, we show the 2-d posterior contours and their 1-d projections on the Ω_m-S_8 plane for all four analyses, for the $\boldsymbol{\Sigma}_{y3}$ covariance (left), and $\boldsymbol{\Sigma}_{\rm full}$ co-

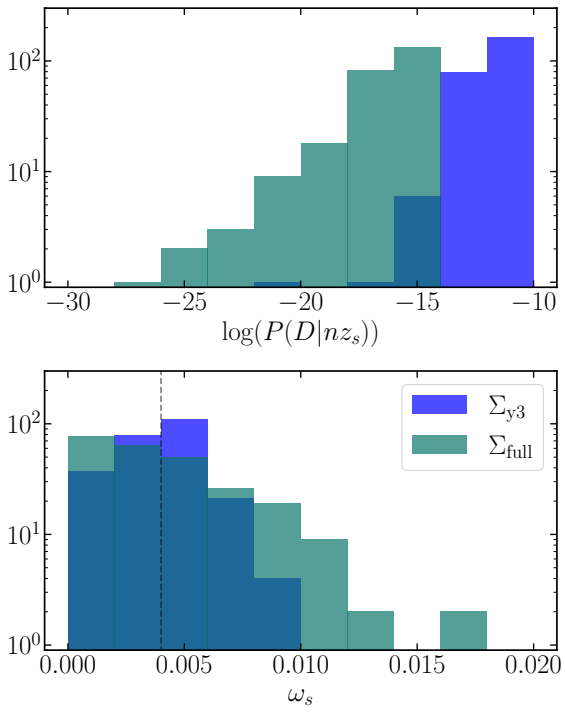

Figure 4. Top row: the distributions of the log-evidence $\log(P(\boldsymbol{D}|\boldsymbol{\phi}_{\rm nz,s}))$, defined in Eq. (12). **Bottom row:** the Bayesian weight, ω_s, defined in Eq. (16) and applied to the chains in the Bayesian resampling approach. The vertical dashed line in the bottom panel is the constant weight applied to each chain in the 'direct resampling' method, $\omega_c = 1/250$. The distributions of log-evidence and Bayesian weight are broader for $\mathbf{\Sigma}_{\rm full}$ than for $\mathbf{\Sigma}_{\rm y3}$, because the same amount of $n(z)$ uncertainty has a larger impact on the more statistically powerful dataset, i.e., the full HSC dataset.

variance (right). For the three-year HSC analyses, the different methods of redshift marginalization do not make a visible difference in the contour plot. However, the contours are visibly different for the future full data set of HSC. For $\mathbf{\Sigma}_{y3}$, the number of resampling for both covariances are $N_{\rm sample,y3} = N_{\rm sample,full} = 250$.

In Fig. 4, we show the distribution of log-evidence $\log(P(\boldsymbol{D}|\boldsymbol{\phi}_{\rm nz,s}))$ and the Bayesian weight ω_s, defined in Eq. (12) and Eq. (16), of the chains in the resampling approach. The direct resampling method applies uniform weights, while the Bayesian resampling method applies the Bayesian weights ω_s. Since the HSC full data-set has a three-times smaller covariance matrix than the three-year data-set, the same $n(z)$ uncertainty causes a more significant scatter in both the log-evidence and Bayesian weight. This means that Bayesian resampling will become increasingly favoured over direct resampling as the dataset becomes more statistically powerful. In practice, the Bayesian resampling approach is assigning more weight to realizations of the $n(z)$ that produce data vectors that are more consistent with the expected one, while down-weighting realizations with less evident $n(z)$.

In Fig. 5, we show the uncertainty for individual cosmological parameters from the full analysis chains in Fig. 3. We used the mean parameter value as the point estimation and the 68% confidence interval as the error bars of the "No $n(z)$ uncertainty" run for the reference. We also show the true value of the parameters in dashed line, as a comparison. For the three-year analyses, shown on the left, marginalizing over the redshift distribution uncertainty does not noticably increase the error bars on either Ω_m and S_8, except when using the "Direct resampling" method. Since the Bayesian resampling method provides a principled approach to incorporation of redshift distribution uncertainties, we take the consistency between that method and the no marginalization method as a sign that the uncertainty in the cosmic shear data vector dominates the uncertainties on cosmological parameters. Therefore, the "Direct resampling" may be introducing spurious uncertainty by failing to down-weight $n(z)$ realizations that are inconsistent with the data vectors, and is not recommended. For the full HSC dataset analyses, shown on the right, we can see that the conclusion of the three-year analyses holds, though the differences between the methods are more visible. The mean posteriors of the Ω_m are systematically lower than the true input value across different methods. We suspect that the banana-shaped $\Omega_m - \sigma_8$ degeneracy that occurs in the full analysis skews the projected distribution of Ω_m to the lower end, which also causes the underestimation of Ω_m in Fig. 7.

We further computed the Figure of Merit (FoM) in the Ω_m-S_8 plane (or Ω_m-S_8-$A_{\rm IA}$ space) to compare the methods, defining the FoM as

$$\mathrm{FoM} = \frac{1}{\sqrt{|\mathbf{F}^{-1}|}}, \tag{21}$$

where $\mathbf{F}$ is the Fisher matrix of $[\Omega_m, S_8]$(or $[\Omega_m, S_8, A_{\rm IA}]$). $\mathbf{F}$ is calculated by taking the inverse of the covariance matrix of $[\Omega_m, S_8]$ (or $[\Omega_m, S_8, A_{\rm IA}]$), approximating the MultiNest posterior as a 2(3)-d Gaussian distribution. This approximation effectively marginalizes over the other parameters that are varied during the parameter inference. The FoM is proportional to the reciprocal of the contour area. In Fig. 6, we plot the FoM of all the marginalization methods, divided by the FoM value of the "No $n(z)$ Uncertainty". The two oranges lines correspond to the full analyses in this section. Unsurprisingly, the direct resampling method provides more conservative parameter constraints compared to the Bayesian resampling method, since it does not downweight the outlier $n(z)$ realizations even though they are unlikely to produce the observed shear data vector. The shift model is slightly conservative for $\mathbf{\Sigma}_{\rm full}$, and slightly optimistic for $\mathbf{\Sigma}_{y3}$, compared to the Bayesian resampling. The errorbars on the FoM values are obtained by bootstrapping the chains. As a cross-check on our errorbars, we also ran 10 chains using the shift model for the Y3 analysis, with different sampling seeds. The errorbars obtained using the standard deviations of the inferred cosmological parameters using these 10 chains is within 5% of those from bootstrapping, which suggests that seeding noise cannot explain the differences in FoM between the methods.

Additionally, we report the ratio of FoMs to the fiducial one in the 3D Ω_m-S_8-$A_{\rm IA}$ space using the full covariance matrix $\mathbf{\Sigma}_{\rm full}$. Since the amplitude of intrinsic alignment is also sensitive to the redshift distribution, different marginalization methods also impact its constraints. The FoM in the Ω_m-S_8-$A_{\rm IA}$ space (purple line) follows the same trend as the orange dashed lines in Figure 6, however, the difference between Bayesian resampling and shift model decreases from 3% to 1% of 1-σ, while the difference between the Bayesian resampling

and direct resampling decreases from 4% to 3.3% of 1-σ[6]. This further strengthens the conclusion that the Bayesian resampling method behaves comparably to the shift model in HSC Y3 cosmic shear analyses, while direct resampling tends to overestimate the uncertainty in the parameter constraints.

Finally, Fig. 6 also shows a FoM comparison for an analysis with only two free cosmological parameters, σ_8 and Ω_m, rather than with all cosmological parameters free. For more details of this analysis, see Appendix B. For this more limited analysis, the direct resampling method overestimates the uncertainties in the (Ω_m, S_8) plane compared to the Bayesian resampling, and therefore is not recommended. The shift model is slightly conservative in this more limited analysis for the full dataset, and slightly optimistic for the three year analysis.

In Fig. 7, we show the 1-d mean posterior points of 250 chains in the resampling approach, run with $\mathbf{\Sigma}_{\rm full}$. The color of the points are coded by the Bayesian weight ω_s of the chain, which is proportional to the model evidence $P(\boldsymbol{D}|\boldsymbol{\phi}_{\rm nz})$. We can see that drawing different samples from the $n(z)$ posterior introduces scatter in the mean values in the Ω_m-S_8 plane, but generally the samples with mean closer to the centre of the cluster receive a higher weight, while the $n(z)$ samples that generate outliers are down-weighted. This plot demonstrates the necessity of considering whether a given $n(z)$ sample is likely to have generated the data vector that we are observing during the resampling process – as is done in the Bayesian resampling approach, but not direct resampling. We notice that there are nz_s samples that generate mean posterior at the centre of the cluster, but receive a very low weight. There are two possible explanations: (a) the realization nz_s has a best-fit data vector that is on average unbiased compared to the mock data vector $\boldsymbol{D}$, but for certain redshifts or θ values there are significant deviations (with opposite signs, so they compensate on average); (b) the best-fit data vector deviated from that for the true cosmological parameters in a way that is compensated by biases in other cosmological parameters besides Ω_m and S_8. The mean values of the Ω_m are systematically lower than the true value of the input, as we explained earlier in this section.

Following the above presentation of the analysis results, we also compare the computational performance of each redshift distribution marginalization method. In Table 5, we show the MultiNest settings used for each method, and the computational expense of the full analysis in CPU-hours. The resampling approaches are two orders of magnitude slower than the shift model. While the Bayesian resampling and shift methods lead to comparable uncertainties, as is shown in Fig. 6, the tremendous computational efficiency of the shift model compared to the Bayesian resampling makes it the recommended choice for the HSC three-year analyses.

For the full HSC three-year cosmic shear analysis, our results suggest that the shift model will produce uncertainties on cosmological parameters that are consistent with the principled Bayesian resampling method to within 3% of 1-σ. Considering that the orders of magnitude difference in computational expense, we recommend the shift model as a well-understood and sufficiently accurate approach for the HSC three-year analysis.

[6] ΔFoM/FoM0 is proportional to $-2\Delta\sigma/\sigma_0$ for two parameters, while ΔFoM/FoM0 is proportional to $-3\Delta\sigma/\sigma_0$ for three parameters, where σ_0 is the $< 68\%$ confidence range of 'no marginalization'

4.2 Inference Validation

In this section, we present the inference validation by performing the probability integral transformation (PIT), as described in Sec. 3.3. We will focus our analysis on the shift model since it represents the simplest methodology that is appropriate for our data as described in the previous sections. While computationally more expensive, we could also perform the same test for the Bayesian and direct Resampling methods. Given that the three aforementioned methods perform similarly in the context of HSC Y3 analysis, we defer a more detailed investigation to future work and concentrate here on the shift model case.

We sample $N_{\rm pit} = 50$ Ω_m-σ_8 pairs, generate a corresponding noisy data vector, obtain the marginalized S_8 posterior from the full inference with shift model, and compute the CDF of the corresponding true S_8 values.

In Fig. 8, we compare the CDF of $\hat{Y}$, the CDF of S_8 evaluated at the true S_8, with the CDF of an expected uniform distribution, shown in the black dashed line. On visual inspection, the estimated CDF follows the expected $y = x$ line nicely. We also conduct an Kolmogorov–Smirnov (K-S) test, which computes the maximum difference between the CDF and the expected CDF. The K-S results is $D = 0.094$, with a p-value of 0.737. This means that $\hat{Y}$ is highly consistent with the uniform distribution, which validates our inference pipeline.

4.3 Literature Comparison

In this section, we compare our Y3 results with the results of marginalizing over $n(z)$ uncertainties in other cosmic shear analysis works. This comparison necessarily excludes the Bayesian resampling approach outlined in this paper, as to the authors' knowledge it has not previously been applied.

In Hildebrandt et al. (2017), direct resampling marginalization is tested using 3 different $n(z)$ uncertainty distributions: weighted direct calibration (DIR; Lima et al. 2008), angular cross-correlation calibration (CC; Newman 2008), and a recalibration of $p(z)$ estimated by BPZ (BOR; Bordoloi et al. 2010). Compared to 'no $n(z)$ marginalization', the DIR, CC, and BOR approaches increase the uncertainty on S_8 by 14%, 90%, and 19% of 1-σ, respectively. In comparison, we find that direct resampling increases the S_8 errorbar by 5.8% of 1-σ for $\mathbf{\Sigma}_{\rm y3}$, and 15.3% of 1-σ for $\mathbf{\Sigma}_{\rm full}$. This smaller increase in uncertainty is likely due to the larger covariance matrix and a tighter $n(z)$ prior for the HSC Y3 analysis.

In Hamana et al. (2020), the shift model is adopted as the fiducial approach to marginalizing over $n(z)$ uncertainty, and is compared with 'no $n(z)$ marginalization'. The uncertainty on S_8 increased by 1.4% of 1-σ after marginalizing over the $n(z)$ uncertainties with the shift model, with a wider prior than the one in this work. In this work, the errorbar on S_8 increased by 2.0% of 1-σ after marginalizing using the shift model. Given that Hamana et al. (2020) has a larger covariance on the shear data vector, as well as a larger prior on the

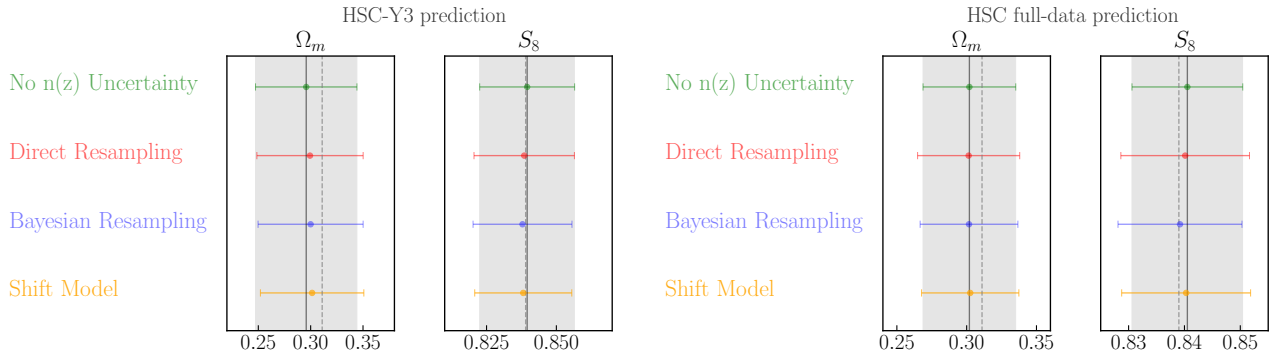

Figure 5. The 68% confidence intervals (bars) and mean values (dots) of the 1-d projection for the $n(z)$ marginalization approaches compared to the 'No $n(z)$ uncertainty' run, for the $\mathbf{\Sigma}_{y3}$ covariance matrix (left), and $\mathbf{\Sigma}_{\rm full}$ covariance matrix (right), for the full analysis. The solid reference lines and the shaded area are the mean values and the 68% confidence intervals of the "No $n(z)$ uncertainty" run. The dashed lines are the parameter truth in Table 2. We find that the mean values of Ω_m are systematically lower than the true input across different methods. The could be caused by the skewness of the posterior distribution.

Method	Live Points	Efficiency	Tolerance	# of chains	CPU-hour/chain	total CPU-hour
No $n(z)$ uncertainty	500	0.1	0.2	1	1.76h*56	98.9h
Direct Resampling	200	0.1	0.2	250	1.05h*28	7350.0h
Bayesian Resampling	200	0.1	0.2	250	1.05h*28	7350.0h
Shift Model	500	0.1	0.2	1	1.77h*56	99.1h

Table 5. The MultiNest settings used in this work, and computational expense for different marginalization methods, for the full analysis using $\mathbf{\Sigma_{y3}}$. The chains are run on Vera, a dedicated server for the McWilliams Center for Cosmology. Each node is equipped with 2 Intel Haswell (E5-2695 v3) CPUs, which have 14 cores per CPU. The resampling approaches, due to the need to run hundreds of individual analyses, are two orders of magnitude slower than the shift model. All chains are ran in constant efficiency mode for more accurate evidence estimation.

shift model parameters, the results should not agree exactly, and there is no reason to believe they are inconsistent.

In Troxel et al. (2018), the shift model is compared with 'no $n(z)$ marginalization'. The prior on the shift model parameters are comparable to this work, while the covariance matrix of Troxel et al. (2018) is smaller than this work. Troxel et al. (2018) found a 4.4% of 1-σ increase in the S_8 uncertainty, which slightly larger than this work. Given the differences in the data covariance for the analyses, the fact that marginalization had a greater impact in Troxel et al. (2018) is consistent with expectations.

In Amon et al. (2022), the shift model is compared with a more sophisticated $n(z)$ marginalization method, called HYPERRANK (Cordero et al. 2022). The shift model is found to be sufficient for cosmic shear analyses for DES Y3, as validated by HYPERRANK. The fact that a current survey found the shift model to be sufficient is consistent with our finding for HSC Y3.

In Stölzner et al. (2021), a self-calibrated method that models the histogram bin heights $\boldsymbol{\phi}_{\rm nz}$ as a series of comb Gaussian functions is used to analytically marginalize over the $n(z)$ uncertainties. The results are compared to the analysis in Wright et al. (2020), which uses a shift model. There are only 1% differences in χ^2 between the results from the self-calibration method and the shift model, though there is a 10% of 1-σ difference in the intrinsic alignment amplitude A_{IA}. This once again shows that the shift model is sufficient for the current generation of cosmic shear analysis for the purpose of cosmological parameter inference, which is consistent with our conclusion.

4.4 Summary of results

Overall, our results show that the shift model is a computationally efficient and credible marginalization method for the HSC three-year analysis. Therefore, we recommend that the HSC three-year analysis adopt the shift model for marginalizing $n(z)$ uncertainty.

For the resampling approaches, we find that using the direct resampling approach consistently results in larger contours compared to the Bayesian resampling, as expected. Therefore, we suggest future cosmological analysis adopt the Bayesian resampling method, if resampling is necessary.

For cosmic shear analyses with a substantial uncertainties on the sample redshift distribution, we recommend comparing any candidate marginalization methods for $n(z)$ with the results from the Bayesian resampling method, as the Bayesian resampling method provide a statistically-principled posterior on the marginalized parameters.

5 CONCLUSIONS

The goal of this work was to understand the performance of methods for incorporating uncertainty in the ensemble redshift distribution in cosmological weak lensing shear analyses, including their impact on computational expense and on the estimated uncertainties on cosmological parameters.

We proposed a statistically-principled method, called Bayesian resampling, for marginalizing over the uncertainties of the sample redshift distribution $n(z)$ in the cosmic shear analysis. By adding a weight proportional to the model evidence of each $n(z)$ realization, Bayesian resampling effectively

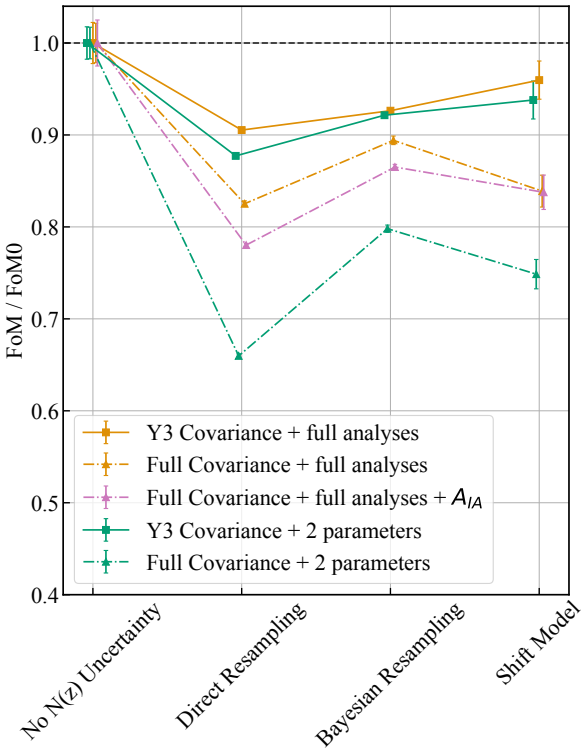

Figure 6. The Figure of Merit (FoM) comparison in the Ω_m-S_8 plane for each marginalization method. The uncertainties on the FoM are computed by bootstrapping the MCMC chains 100 times. All FoMs are divided by the FoM of the "No $n(z)$ uncertainty" run. The purple line shows FoM ratio for the 3D Ω_m-S_8-$A_{\rm IA}$ space, while other lines shows the FoM in the Ω_m-S_8 space. The direct resampling method is clearly the most conservative method of those tested in this work. The shift model shows similar performance to the Bayesian resampling method. The errorbars on the FoM are given by bootstrapping the chains, which matches the errorbars given by running the analysis with different sampling seeds.

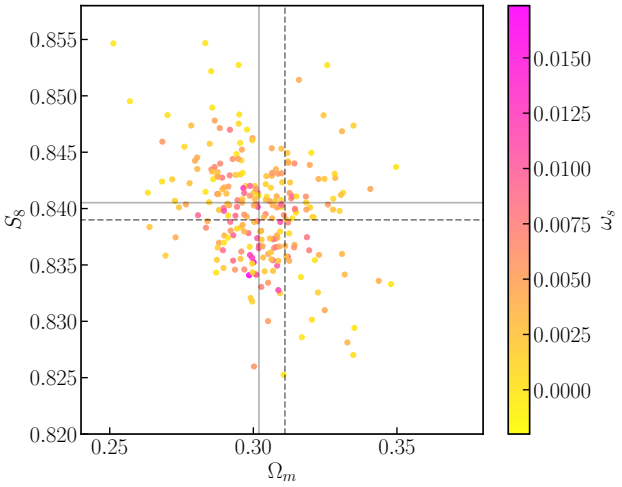

Figure 7. A scatter plot showing the mean posterior values for the $N_{\rm sample} = 250$ chains with different samples drawn from the $n(z)$ prior, analyzed with all 17 parameters (using the shift model) and with $\mathbf{\Sigma}_{\rm full}$. The colors indicate the Bayesian weight ω_s that the chain receives, determined by its evidence $P(\boldsymbol{D}|\boldsymbol{\phi}_{\rm nz})$, defined in Eq. (12). The solid line represents the mean statistics of the 'No $n(z)$ uncertainties' run, while the dashed line represents the true input parameters.

down-weights those realizations that are unlikely to generate the observed cosmic shear data vector. The Bayesian resampling method can be applied to any $n(z)$ uncertainties that can be modeled by a probability distribution, even if such parameterization is at a high dimensionality that makes it impossible to model in MCMC.

We ran mock analyses for the HSC three-year and full-data cosmic shear, with 3 $n(z)$ marginalization methods: (a) the newly developed Bayesian resampling method; (b) the direct resampling, for which the weights of all $n(z)$ realizations are the same; (c) the shift model, the most prevalent parameterization used in cosmic shear analyses. Additionally, we ran analyses without marginalizing over the $n(z)$ as a comparison. Our mock data vector is the average cosmic shear signal from the fiducial cosmology, and its covariance is estimated by reducing the covariance compared to that in Hamana et al. (2020) to account for survey area increases, for the three-year analysis and full analysis correspondingly. Our full theoretical model consists 5 ΛCDM parameters, 2 intrinsic alignment

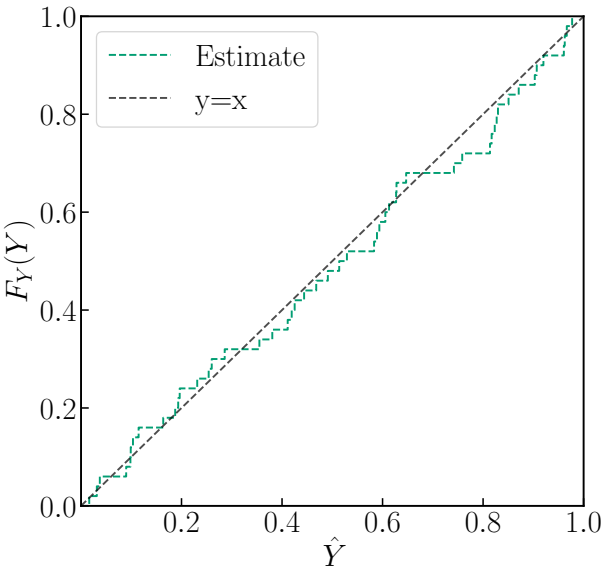

Figure 8. The CDF of $\hat{Y}$, defined in Sec. 3.3. If the statistical inference preserves Bayes' theorem, the CDF of $\hat{Y}$ should follow $y = x$, which is plotted as the black dashed line. This plot shows a high level of consistency between the estimated and expected distribution of $\hat{Y}$.

parameters, 4 multiplicative biases and 2 PSF systematics parameters, plus the 4 redshift parameters when the shift model is adopted.

We compared the 3 marginalization methods and the analysis without marginalization in terms of their impact on the Ω_m-S_8 contours, their 1-d errorbars, the figure of merit (FoM), and computational cost. Here is a high-level summary of how the methods compared to each other.

- Marginalizing over $n(z)$ uncertainties yields larger errorbars on both Ω_m and S_8 for all methods.
- Bayesian resampling yields significant tighter errorbars than direct resampling, implying that the direct resampling is overly-conservative for marginalization.
- The shift model produces consistent errorbars to the Bayesian resampling for HSC Y3. Given that the computational cost for the shift model is $\sim$ 100 times less, it is the recommended method for the upcoming HSC Y3 cosmic shear analyses. For the HSC full analysis, the shift model can yield errorbars that differ by $\sim$ 3% of 1-σ compared to Bayesian resampling, so it is unclear even in that case whether alternative methods are worthwhile.
- Although the differences between the marginalization methods are statistically evident, the visual differences in the parameter constraint contours are not particularly noticeable.

To test the credibility of our inferred posterior probability distributions of cosmological parameters, we conducted the probability interval transformation (PIT) test on noisy data vectors generated with a range of cosmological parameters, to ensure the applicability of our results to real cosmic shear analyses. We sampled 50 pairs of Ω_m-σ_8-$\boldsymbol{D}$, and compare the CDF distribution of S_8 at the true S_8^μ values with a uniform distribution. Our estimated CDF distribution passes the K-S test, thus validating our inference pipeline using the shift model.

These results have a few implications for future cosmic shear analyses. First, our results suggest that the shift model should be compared with Bayesian resampling for specific survey scenarios (statistical constraining power, etc.) to assess whether the shift model performs sufficiently well to be usable, given its far lower computational expense. The shift model is fundamentally a different $n(z)$ uncertainty model from the original $n(z)$ distribution. Second, when using the resampling approach to marginalizing over $n(z)$ uncertainties is necessary for a weak lensing measurement, the Bayesian resampling approach is preferred over direct resampling, because of its consistency with Bayesian statistics. Moreover, Bayesian resampling does require an accurate estimate of the ratio of the Bayesian evidences between realizations of redshift distributions.

There are several caveats in this work. (a) We reach the conclusion that a sophisticated marginalization method is going to be increasingly preferred based on the assumption that lensing measurements become more powerful as survey area increases, but the uncertainty on $n(z)$ is presumed to be systematics dominated.The reason for this assumption is that the $n(z)$ uncertainties are limited by the cosmic variance of the COSMOS2015 field, which we used to assess the modeling uncertainties. If this assumption changes, then the comparison needs to be revisited. This assumption is discussed in detail in Section 2.2. (b) We use the same angular and tomographic binning for the mock analyses in this paper, though the actual analyses of HSC Y3 and full data are likely to have different binning strategies. We also make very simple estimates of the covariance matrices in the mock analyses, ignoring the evolving footprint shape of the HSC survey. (c) The assumption in this work is that we can place a prior on the source redshift distribution that is statistically independent of our data vector. That was a good approximation in this case, for $n(z)$ calibration based on photometry and cross-correlations, and for the data vector involving shear-shear only. However, future analyses with more complex data vectors (e.g., including large-scale structure clustering) and/or $n(z)$ posteriors may violate this assumption in our formalism, which would require additional efforts to take into account.

We conclude by mentioning some avenues for future investigations. First, the cosmic shear data vector is sensitive to the mean redshift of the tomographic bin, which is likely the reason why the shift model is sufficient for current surveys in practice. However, galaxy clustering is sensitive to other statistics of the ensemble redshift distribution, such as its width (e.g., Abbott et al. 2022). Therefore, the validity of the shift model in galaxy-galaxy lensing, clustering and 3x2pt analyses should be directly tested.

Finally, the resampling approach for the $n(z)$ marginalization requires thousands of CPU-hours. Importance sampling methods can be added to the method to reduce the number of realizations needed. However, importance sampling faces other challenges: since $n(z)$ distributions normally are parameterized with high dimensionality, the importance weights are easily dominated by a few samples. It might also be extremely challenging to perform importance sampling on some $n(z)$ priors. It would be valuable to identify solutions to this problem and demonstrate how to effectively accelerate $n(z)$ resampling approaches using importance sampling.

CONTRIBUTORS

TZ developed the mock analysis pipeline, conducted the cosmology inferences and validation tests, and led the writing of the manuscript. MMR proposed the project, advised on the experimental design and analysis, provided and wrote about the sample redshift distribution data, and provided feedbacks on the results. RM advised on the motivation, experimental design and analysis, and edited the manuscript. XL provided implementation guidance on the inference pipeline and feedbacks to the results. BM provided comments and feedback to the results, and edited the manuscript.

Acknowledgments

We thank the anonymous referee for constructive feedback on this work. We thank Alex Malz, Chad Schafer, Andresa Campos, Biwei Dai, Danielle Leonard for their helpful discussion with us.

TZ and RM are supported in part by the Department of Energy grant DE-SC0010118 and in part by a grant from the Simons Foundation (Simons Investigator in Astrophysics, Award ID 620789).

We thank the developers of CosmoSIS, NumPy, and ChainConsumer for making their software openly accessible.

DATA AVAILABILITY

The data vector, covariance matrix and stacked $p(z)$ distribution are available on http://th.nao.ac.jp/MEMBER/hamanatk/HSC16aCSTPCFbugfix/index.html. The sample redshift distribution, the cosmology inference and analysis code will be shared on reasonable request to the authors.

REFERENCES

Abbott T. M. C., et al., 2022, Phys. Rev. D, 105, 023520
Aihara H., et al., 2018, PASJ, 70, S4
Akeson R., et al., 2019, arXiv e-prints, arXiv:1902.05569
Amon A., et al., 2022, Phys. Rev. D, 105, 023514
Asgari M., et al., 2021, A&A, 645, A104
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bordoloi R., Lilly S. J., Amara A., 2010, MNRAS, 406, 881
Bridle S., King L., 2007, New Journal of Physics, 9, 444
Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483
Casella G., Berger R., Company B. P., 2002, Statistical Inference. Duxbury advanced series in statistics and decision sciences, Thomson Learning, https://books.google.com/books?id=0x_vAAAAMAAJ
Cordero J. P., et al., 2022, MNRAS, 511, 2170
Croft R. A. C., Metzler C. A., 2000, ApJ, 545, 561
Cunha C. E., Lima M., Oyaizu H., Frieman J., Lin H., 2009, MNRAS, 396, 2379
Dark Energy Survey Collaboration et al., 2016, MNRAS, 460, 1270
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2019, OJAp, 2, 10
Fu L., et al., 2008, A&A, 479, 9
Gatti M., et al., 2021, MNRAS, 510, 1223
Hadzhiyska B., Alonso D., Nicola A., Slosar A., 2020, J. Cosmology Astropart. Phys., 2020, 056
Hamana T., et al., 2020, PASJ, 72, 16
Handley W. J., Hobson M. P., Lasenby A. N., 2015a, MNRAS Letters, 450, L61
Handley W. J., Hobson M. P., Lasenby A. N., 2015b, MNRAS, 453, 4384
Heavens A., Refregier A., Heymans C., 2000, MNRAS, 319, 649
Henghes B., Pettitt C., Thiyagalingam J., Hey T., Lahav O., 2021, MNRAS, 505, 4847
Hildebrandt H., et al., 2017, MNRAS, 465, 1454
Hildebrandt H., et al., 2020, Astron. Astrophys., 633, A69
Hinton S. R., 2016, The Journal of Open Source Software, 1, 00045
Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526
Howlett C., Lewis A., Hall A., Challinor A., 2012, J. Cosmology Astropart. Phys., 2012, 027
Hu W., 1999, ApJ, 522, L21
Hu W., 2002, Phys. Rev. D, 65, 023003
Huff E. M., Eifler T., Hirata C. M., Mandelbaum R., Schlegel D., Seljak U., 2014, MNRAS, 440, 1322
Huterer D., 2010, General Relativity and Gravitation, 42, 2177
Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101
Ishikawa S., Okumura T., Oguri M., Lin S.-C., 2021a, ApJ, 922, 23
Ishikawa S., Okumura T., Oguri M., Lin S.-C., 2021b, ApJ, 922, 23
Ivezić v. Z., et al., 2019, Astrophys. J., 873, 111
Joachimi B., Bridle S. L., 2010, A&A, 523, A1
Joudaki S., et al., 2020, Astron. Astrophys., 638, L1
Kaiser N., Wilson G., Luppino G. A., 2000, arXiv e-prints, pp astro–ph/0003338
Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901
Krause E., Eifler T., 2017, MNRAS, 470, 2100
LSST Science Collaboration et al., 2009, arXiv e-prints, arXiv:0912.0201
Laigle C., et al., 2016, ApJS, 224, 24
Laureijs R., et al., 2011, arXiv e-prints, arXiv:1110.3193
Lemos P., et al., 2022, arXiv e-prints, arXiv:2202.08233
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Li X., et al., 2022, PASJ, 74, 421
Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, ApJ, 390, 118
Longley E. P., et al., 2022, arXiv e-prints, arXiv:2208.07179
Malz A. I., 2021, Phys. Rev. D, 103, 083502
Malz A. I., Hogg D. W., 2022, ApJ, 928, 127
Malz A. I., Marshall P. J., DeRose J., Graham M. L., Schmidt S. J., and R. W., 2018, ApJ, 156, 35
Mandelbaum R., 2018, Ann. Rev. Astron. Astrophys., 56, 393
Morrison C. B., Hildebrandt H., Schmidt S. J., Baldry I. K., Bilicki M., Choi A., Erben T., Schneider P., 2017, MNRAS, 467, 3576
Newman J. A., 2008, ApJ, 684, 88
Oguri M., 2014, MNRAS, 444, 147
Peebles P. J. E., 1980, The large-scale structure of the universe
Planck Collaboration et al., 2020, A&A, 641, A6
Rau M. M., Wilson S., Mandelbaum R., 2020, MNRAS, 491, 4768
Rau M. M., Morrison C. B., Schmidt S. J., Wilson S., Mandelbaum R., Mao Y. Y., Mao Y. Y., LSST Dark Energy Science Collaboration 2022, MNRAS, 509, 4886
Salvato M., Ilbert O., Hoyle B., 2019, Nature Astronomy, 3, 212
Sánchez C., Bernstein G. M., 2019, MNRAS, 483, 2801
Schmidt S. J., et al., 2020, MNRAS, 499, 1587
Secco L. F., et al., 2022, Phys. Rev. D, 105, 023515
Seljak U., 1998, ApJ, 506, 64
Smith R. E., et al., 2003, MNRAS, 341, 1311
Speagle J. S., et al., 2019, MNRAS, 490, 5658
Spergel D., et al., 2015, arXiv e-prints, arXiv:1503.03757
Stölzner B., Joachimi B., Korn A., Hildebrandt H., Wright A. H., 2021, A&A, 650, A148
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Tanaka M., 2015, ApJ, 801, 20
Tanaka M., et al., 2018, PASJ, 70, S9
Troxel M. A., et al., 2018, Phys. Rev. D, 98, 043528
Wright A. H., Hildebrandt H., van den Busch J. L., Heymans C., Joachimi B., Kannawadi A., Kuijken K., 2020, A&A, 640, L14
Zuntz J., et al., 2015, Astronomy and Computing, 12, 45
de Jong J., et al., 2017, Astron. Astrophys., 604, A134

APPENDIX A: IMPACT OF $\xi_\pm(\langle\phi_{\rm NZ}\rangle)$

In Figure A1, we demonstrate that generating the auto-correlation of the mock data vector using the average $n(z)$ is different from the taking the average of data vectors generated by random draws from the posterior for $n(z)$. Therefore, for this work, in which the conclusion is sensitive to bias at the sub-percentage level, we choose to use $\langle\xi_\pm^{ii}(n(z))\rangle$ from 1000 $n(z)$ samples to generate the mock data vector in Section 3.1.3.

The reason that only the auto-correlation is affected in Figure A1 can be explained by Eqs. (2) and (3). Since $n^i(\chi(z))$ is independent of $n^j(\chi(z))$ if $i \neq j$, the transfer functions $q^i(\chi)$ and $q^j(\chi)$ are thus independent. As a result, when $i \neq j$,

$$\langle C_\ell^{ij}\rangle = \int \frac{{\rm d}\chi}{\chi^2} P(\ell/\chi; z(\chi))\langle q^i(\chi)\rangle\langle q^j(\chi)\rangle. \tag{A1}$$

Notice that Eq. (A1) only holds when $n^i(\chi(z))$ is independent

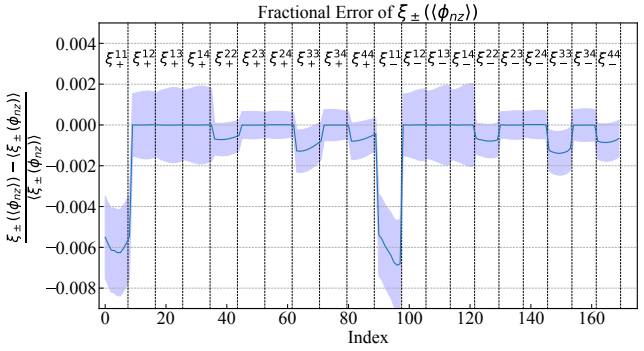

Figure A1. The fractional error of the $\xi_\pm^{ij}(\langle n(z)\rangle)$ compared to $\langle\xi_\pm^{ij}(n(z))\rangle$. The x-axis is the index of the value in the data vector, with 9 angular bins for each tomographic ξ_+^{ij} bin pair, and 8 angular bins for each ξ_-^{ij} bin pair. There is a statistically significant sub-percent difference between the auto-correlations, which shows that taking the average of $n(z)$ does not commute with computing the 2PCF. Therefore $\langle\xi_\pm^{ij}(n(z))\rangle$ should be used to avoid sub-percent bias when this level of precision matters. This fractional difference is largest at low redshift because the uncertainties in the mean redshifts are comparable in magnitude for all bins, but the 2PCF is lower at low redshift.

of $n^j(\chi(z))$. Otherwise, both auto- and cross-correlations in the mock data vector will be affected by using the average $n(z)$. Also note that in the case that some overall source of uncertainty were to lead to correlations between the uncertainties in the redshift distributions for different bins, both auto- and cross-correlations would be affected.

APPENDIX B: TWO-PARAMETER ANALYSES

In this work, we also carried out the cosmological parameter inference for a case where only Ω_m and σ_8, along with $n(z)$ marginalization nuisance parameters, are freed. This scaled-down test is initially designed for testing and sanity-checking our inference and analysis software. The constant values for other cosmological and nuisance parameters, and the priors for Ω_m and σ_8, are listed in Tables 2 and 3.

We carried out cosmological parameter estimation for the three marginalization methods in the 2-parameter cases, along with the "no $n(z)$ uncertainty" run for comparison. The contour plots in the Ω_m and S_8 plane are well-behaved, and the results lead to similar conclusions as for the full analyses, so we do not show them in the paper. The Figure-of-Merit ratio of the three marginalization methods to that of "no $n(z)$ uncertainty" is shown in green lines in Figure 6, and the conclusion based on the 2-parameter cases is similar to ones drawn from the full analyses.