

# Redshift distributions of extragalactic galaxy surveys

*Benjamin Stölzner*

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
**Doctor of Philosophy**  
of  
**University College London.**

Department of Physics and Astronomy  
University College London

February 9, 2023

I, Benjamin Stölzner, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

# Abstract

Measurements of the large-scale structure of the universe are key observables to study fundamental physics. This thesis focuses on the calibration of photometric redshift distributions of extragalactic galaxy surveys and their impact on the analysis of large-scale structure data.

First, I examine the impact of redshift distribution uncertainties on the cosmological inference from weak lensing measurements. The weak gravitational lensing effect, known as cosmic shear, distorts the shape of galaxy images due to the distribution of gravitating matter along the line of sight. Thus, it provides a probe of the matter distribution in the universe. However, modelling the observed cosmic shear signal requires knowledge about the distribution of observed galaxies along the line of sight, which is usually determined through photometric redshifts. I develop a method that accurately propagates residual redshift distribution uncertainties into the weak lensing likelihood and perform a self-calibration of the redshift distribution with cosmic shear data.

Second, I develop a new method to assign photometrically observed galaxies to tomographic redshift bins. The goal is to obtain compact distributions and to reduce the overlap between redshift bins caused by catastrophic outliers in the photometric redshift estimation. This is achieved by combining a self-organising map with a simulated annealing algorithm which optimises the clustering cross-correlation signal between a photometric galaxy catalogue and a spectroscopically observed sample of reference galaxies.

Finally, I perform consistency tests in cosmological analyses. These tests include a study of the consistency between the constraints on cosmological

parameters probed by the five tomographic bins of the Kilo-Degree Survey. Furthermore, I study the internal consistency of the  $\Lambda$ CDM model by dividing the model into regimes: one that describes the evolution of the isotropic background of the universe and one describing matter density perturbations. This model is constrained by cosmic shear, galaxy clustering, and cosmic microwave background measurements.

# Impact Statement

This thesis focusses on the role of photometric redshift distributions in current and future extragalactic galaxy surveys. These surveys provide observations of billions of galaxies which allow us to better understand the fundamental properties of the Universe.

Our work provides a method to accurately propagate residual uncertainties in photometric redshift distributions into the cosmological inference from weak lensing measurements. Photometric redshifts allow us to determine the distances of extragalactic objects, which are required in the modelling of cosmological observables. This is achieved through observations of objects with a set of broad filters that allow us to construct a low-resolution measurement of the spectral energy distribution, from which we infer the redshift of an object. Given the statistical power of upcoming galaxy surveys, the handling of uncertainties in the photometric redshift estimation will be crucial in the analysis of cosmological data. Our work therefore improves on an important aspect of the analysis of upcoming cosmological surveys and is currently being adopted in an upcoming analysis of data from the Kilo-Degree Survey.

Additionally, we develop an optimisation method for the assignment of photometrically observed galaxies to tomographic redshift bins. Cosmological analyses are often performed tomographically, which allows us to study the properties of the Universe in slices along the line of sight. This is usually done by assigning galaxies to bins based on their photometric redshift estimate, which however is subject to catastrophic outliers, where the photometric redshift estimate differs greatly from the true redshift of the galaxy. Our work

provides an optimised assignment scheme, which makes the underlying redshift distributions in each tomographic bin more compact, helping to improve the accuracy of future cosmological surveys.

Finally, we perform consistency tests in cosmological analyses. These tests play an important role in studies of cosmological data, since we require our datasets to be self-consistent in order to infer robust constraints on cosmological parameters. Our work therefore makes an important contribution to the analysis of cosmological data.

The work presented in this thesis has resulted in a peer-reviewed publication (chapter 2; [Stölzner et al., 2021](#)) and a second publication that has been submitted for peer-review (chapter 3; [Stölzner et al., 2023](#)). Additionally, the work presented in section 5.2 has been included in one of the key papers of the analysis of the fourth data release of the Kilo-Degree Survey ([Asgari et al., 2021](#)). Finally, the results of section 5.3 are also part of a peer-reviewed publication ([Ruiz-Zapatero et al., 2021](#)).

# Acknowledgements

First of all, I would like to thank my supervisors Benjamin Joachimi and Andreas Korn for their expert guidance and advice on my work. Thank you for all the discussions and suggestions, both in-person and online, and for giving me the opportunity to join the DESC and KiDS collaborations.

Moreover, I would like to thank the current and former students, postdocs, and staff of the UCL Astrophysics group and the UCL Cosmoparticle Initiative for creating a positive and friendly work environment. In particular, thank you to Catarina, Max, Francesca, Lillian, Kiyam, Davide, and Arthur for all the lunches, dinners, pub visits, trips, and (most importantly!) coffee breaks.

Finally, I want to thank my family for supporting me and making the PhD possible.

# UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server;
- are in submission to a peer-reviewed publication;
- have been published in a peer-reviewed publication, e.g. journal, textbook.

*This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.*

**Paper title:** Self-calibration and robust propagation of photometric redshift distribution uncertainties in weak gravitational lensing

<b>1. For a research manuscript that has already been published</b> (if not yet published, please skip to section 2):	
<b>a) Where was the work published?</b> (e.g. journal name)	Astronomy & Astrophysics
<b>b) Who published the work?</b> (e.g. Elsevier/Oxford University Press):	EDP Sciences
<b>c) When was the work published?</b>	23/06/2021
<b>d) Was the work subject to academic peer review?</b>	Yes
<b>e) Have you retained the copyright for the work?</b>	No
[If no, please seek permission from the relevant publisher and check the box next to the below statement]:	
<input checked="" type="checkbox"/> <i>I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.</i>	
<b>2. For a research manuscript prepared for publication but that has not yet been published</b> (if already published, please skip to section 3):	
<b>a) Has the manuscript been uploaded to a preprint server?</b> (e.g. medRxiv):	<b>If yes, which server?</b> Click or tap here to enter text.
<b>b) Where is the work intended to be published?</b> (e.g. names of journals that you are planning to submit to)	
<b>c) List the manuscript's authors in the intended authorship order:</b>	
<b>d) Stage of publication</b>	

**3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4):**

Benjamin Stölzner carried out the analysis and led the writing of the manuscript. Benjamin Joachimi and Andreas Korn provided feedback and suggestions throughout the project and contributed to the writing of the manuscript. Hendrik Hildebrandt provided the datasets analysed in this work and provided feedback on the analysis. Angus Wright reviewed the paper internally in the KiDS collaboration and provided suggestions on the analysis.

**4. In which chapter(s) of your thesis can this material be found?**

Chapter 2

**5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):**

**Candidate:**



**Date:**

11/8/2022

**Supervisor/  
Senior Author**  
(where  
appropriate):



**Date:**

C11/08/2022

# UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server;
- are in submission to a peer-reviewed publication;
- have been published in a peer-reviewed publication, e.g. journal, textbook.

*This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.*

**Paper title:** Optimising the shape of photometric redshift distributions with clustering cross-correlations

<b>1. For a research manuscript that has already been published</b> (if not yet published, please skip to section 2):		
<b>a) Where was the work published?</b> (e.g. journal name)		
<b>b) Who published the work?</b> (e.g. Elsevier/Oxford University Press):		
<b>c) When was the work published?</b>		
<b>d) Was the work subject to academic peer review?</b>		
<b>e) Have you retained the copyright for the work?</b>		
[If no, please seek permission from the relevant publisher and check the box next to the below statement]:		
<input type="checkbox"/> <i>I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.</i>		
<b>2. For a research manuscript prepared for publication but that has not yet been published</b> (if already published, please skip to section 3):		
<b>a) Has the manuscript been uploaded to a preprint server?</b> (e.g. medRxiv):	Yes	<b>If yes, which server?</b> arXiv
<b>b) Where is the work intended to be published?</b> (e.g. names of journals that you are planning to submit to)	Monthly Notices of the Royal Astronomical Society	
<b>c) List the manuscript's authors in the intended authorship order:</b>	Benjamin Stölzner, Benjamin Joachimi, Andreas Korn, the LSST Dark Energy Science Collaboration	
<b>d) Stage of publication</b>	Undergoing revision after peer review	

**3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4):**

Benjamin Stölzner carried out the analysis and led the writing of the manuscript. Benjamin Joachimi and Andreas Korn provided feedback and suggestions throughout the project and contributed to the writing of the manuscript. The paper was reviewed internally by members of the LSST Dark Energy Science Collaboration.

**4. In which chapter(s) of your thesis can this material be found?**

Chapter 3

**5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):**

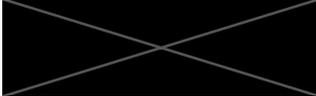
**Candidate:**



**Date:**

11. 8. 2022

**Supervisor/  
Senior Author**  
(where  
appropriate):



**Date:**

C11/08/2022

# Contents

<b>1</b>	<b>Introduction</b>	<b>28</b>
1.1	Homogeneous cosmology . . . . .	28
1.1.1	General Relativity . . . . .	29
1.1.2	The content of the Universe . . . . .	31
1.2	The Large Scale Structure of the Universe . . . . .	35
1.2.1	Matter power spectrum . . . . .	35
1.2.2	Large scale structure probes . . . . .	42
1.2.2.1	Galaxy clustering . . . . .	42
1.2.2.2	Weak gravitational lensing . . . . .	46
1.3	Photometric redshifts . . . . .	53
1.4	The current status of cosmology . . . . .	57
<b>2</b>	<b>Self-calibration and robust propagation of photometric redshift distribution uncertainties in weak gravitational lensing</b>	<b>61</b>
2.1	Introduction . . . . .	62
2.2	Redshift distribution model . . . . .	65
2.3	Theoretical modelling of the cosmic shear signal . . . . .	68
2.3.1	Weak lensing model and likelihood . . . . .	68
2.3.2	Marginal likelihood . . . . .	71
2.4	Redshift distribution self-calibration . . . . .	73
2.5	KV450 likelihood analysis . . . . .	76
2.5.1	Redshift distribution self-calibration . . . . .	79
2.5.2	Marginalisation over nuisance parameters . . . . .	84

2.6	Summary and conclusions . . . . .	89
<b>3</b>	<b>Optimising the shape of photometric redshift distributions with clustering cross-correlations</b>	<b>92</b>
3.1	Introduction . . . . .	93
3.2	Methodology . . . . .	95
3.2.1	Galaxy clustering . . . . .	96
3.2.2	Simulated annealing . . . . .	100
3.2.3	Self-organising maps . . . . .	102
3.3	Results . . . . .	107
3.4	Conclusions . . . . .	113
<b>4</b>	<b>Optimising photometric redshift distributions with realistic reference datasets</b>	<b>117</b>
4.1	SOM-based cross-correlation measurements . . . . .	118
4.2	Probabilistic bin assignment scheme . . . . .	119
4.3	Results for idealised reference samples . . . . .	121
4.4	Results for DESI-like reference samples . . . . .	123
4.5	Conclusions . . . . .	128
<b>5</b>	<b>Internal consistency tests of cosmological data</b>	<b>130</b>
5.1	Bayesian statistics . . . . .	131
5.2	Internal consistency between tomographic bins in KiDS-1000 . .	137
5.2.1	Methodology . . . . .	138
5.2.2	Results . . . . .	142
5.3	Consistency of background and perturbations in the $\Lambda$ CDM model	148
5.3.1	Split cosmological model . . . . .	149
5.3.2	Results . . . . .	151
<b>6</b>	<b>Conclusions</b>	<b>155</b>
	<b>Appendices</b>	<b>161</b>

<b>A</b>	<b>Appendix to Chapter 2</b>	<b>161</b>
A.1	Marginalisation over nuisance parameters . . . . .	161
A.2	Tests of the redshift distribution calibration . . . . .	162
A.2.1	Width of input data histograms . . . . .	162
A.2.2	Redshift distribution calibration with simulations . . . . .	164
A.2.3	Number of Gaussian components . . . . .	166
A.3	Comparison between fiducial KV450 and Gaussian comb . . . . .	167
A.4	Posteriors of cosmological parameter constraints . . . . .	168
<b>B</b>	<b>Appendix to Chapter 3</b>	<b>169</b>
B.1	Photometric redshifts . . . . .	169
B.2	SOM clustering . . . . .	170
B.3	Energy in the simulated annealing optimisation . . . . .	170

# List of Figures

- 1.1 Measurements of the linear matter power spectrum from CMB observations (Planck Collaboration et al., 2020a), weak lensing (Troxel et al., 2018), galaxy clustering (Reid et al., 2010), and the Lyman-alpha forest (Palanque-Delabrouille et al., 2015). The solid black line shows the  $\Lambda$ CDM model prediction for the linear matter power spectrum while the black dotted line shows the impact of non-linear clustering at redshift  $z = 0$ . . . . . 40
- 1.2 Map of galaxies observed by SDSS. Each dot represents a galaxy with the colour representing the g-r colour. Image credit: M. Blanton and SDSS. Source: <https://www.sdss.org/science/orangepie/>, published under a Creative Commons Attribution license (CC-BY). . . . . 43
- 1.3 Geometry of a thin-lens system. The observer is located at position O, the source galaxy at S, and the image of the galaxy appears at I. The distances between the observer and the source galaxy, the observer and the lens, and the lens and the source galaxy are  $D_S$ ,  $D_L$ , and  $D_{LS}$ , respectively. Figure adapted from (Bartelmann and Maturi, 2017). . . . . 48
- 1.4 Transmission functions of the six filters used by the Vera C. Rubin Observatory’s Legacy Survey of Space and Time. Source: <https://github.com/lsst/throughputs> . . . . . 56

- 1.5 Marginalised posterior distribution of  $\sigma_8$  and  $\Omega_m$ , measured by Tröster et al. (2021); Planck Collaboration et al. (2020b); Heymans et al. (2021), and Abbott et al. (2018). Figure adapted from Heymans et al. (2021). . . . . 58
- 2.1 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). . . . . 76
- 2.2 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. 2.3. Orange curves represent the redshift distributions after iterative optimisation of cosmological and nuisance parameters. . . . . 80
- 2.3 Correlation matrix of best-fit comb amplitudes with 30 components per redshift bin. . . . . 81

- 2.4 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. . . . . 83
- 2.5 Marginalised posteriors for  $A_{\text{IA}}$  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 2.2 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. 85
- 2.6 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. 88

- 3.1 Sketch of the optimisation algorithm that reassigns photometric galaxies to alternative redshift bins. We train a self-organising map (SOM) with a high resolution  $R_{\text{original}}$  on the observed colours of galaxies in the photometric sample, from which we infer SOMs with arbitrary resolutions  $R < R_{\text{original}}$ . Additionally, we infer point estimates of the photometric redshift to divide the sample into tomographic bins, so that each SOM node is assigned initially to the most common redshift bin of galaxies in this node. We initialise the simulated annealing algorithm with a starting temperature  $T_{\text{max}}$  and a resolution  $R_{\text{min}}$ , which is coupled to the temperature. In each iteration of the annealing algorithm we select a node of a SOM with the current resolution, which we randomly reassign to a different redshift bin. Measuring the angular cross-correlation between the photometric sample and the reference sample, we calculate the energy of the system from Eq. (3.3). Comparing the change in energy and the current temperature, we determine whether to keep the redshift bin assignment or to restore the previous state using Eq. (3.8). We then decrease the system's temperature and, depending on the temperature, we either keep the current SOM resolution or increase the resolution using the scheme outlined in Fig. 3.2. We iterate through these steps until we reach the given final temperature  $T_{\text{min}}$  and final resolution  $R_{\text{max}}$ . . . . . 97

- 3.2 Sketch of the clustering method that is used to decrease the resolution of a SOM. We combine nodes of a high-resolution SOM, consisting of  $M$  nodes, using the number  $N$  of galaxies that are mapped onto each SOM node and the weight vectors  $\mathbf{x}$ , that connect the two-dimensional SOM data space to the original high dimensional data space. In each step we merge the two SOM nodes with the minimum distance of their corresponding weight vectors and compute the weight vector of the combined node as the average of the two weight vectors, where the number of galaxies in each node act as weights. By iterating the merging process we build a hierarchy of clusters until all SOM nodes are merged into one single node after  $M - 1$  steps. From the hierarchy of nodes we can derive SOMs with any resolutions lower than the resolution of the original SOM. . . . . 106
- 3.3 Illustration of the self-organising map used in the analysis. The SOM consists of 200x200 nodes on a rectangular grid in toroidal geometry and is trained on the observed colours of galaxies in the photometric galaxy sample. Coloured labels indicate the tomographic bin to which galaxies in each node are assigned, with SOM nodes which do not contain any galaxies labelled with ‘-1’. . . . . 109
- 3.4 Left: Initial cross-correlation matrix between bins of the photometric sample, inferred from photometric redshift estimates, and the reference sample. Right: Cross-correlation matrix after optimisation of the bin assignment of the photometric sample. A perfect assignment with noise-free clustering measurements would yield the identity matrix. . . . . 110

- 3.5 Comparison of the initial redshift distribution of each tomographic bin, obtained using the photometric redshift estimate of individual galaxies (dashed lines), and the redshift distribution after simulated annealing (solid lines). Dotted lines indicate the redshift bin edges. . . . . 112
- 3.6 Left: Comparison of the percentage of probability mass of the redshift distributions, shown in Fig. 3.5, within the respective bin range before and after simulated annealing. Right: Comparison of the percentage of the redshift distribution that is located within the tails of the distribution, where we define the tails of the distribution as the region in redshift space that lies more than one bin width outside of the boundaries of a given tomographic bin. . . . . 113
- 4.1 Comparison of the energy as a function of steps between the default bin assignment scheme and the probabilistic bin assignment scheme. . . . . 120
- 4.2 Comparison of the initial redshift distribution (dashed lines) and the final redshift distribution (solid lines) after optimisation of the bin assignment with a SOM-based cross-correlation measurements. Dotted lines indicate the redshift bin edges. . . . 122
- 4.3 Left: Comparison of the percentage of probability mass of the redshift distributions, shown in Fig. 4.2, within the respective bin range before and after optimisation. Right: Comparison of the percentage of the redshift distribution that is located within the tails of the distribution, where we define the tails of the distribution as the region in redshift space that lies more than one bin width outside of the boundaries of a given tomographic bin. . . . . 123

- 4.4 Overview of the redshift distributions of galaxies in the cosmoDC2 catalogue with a DESI-like selection for four samples: Luminous Red Galaxies (LRGs), Emission Line Galaxies (ELGs), high redshift quasars (QSOs), and a magnitude-limited sample. The boundaries of the redshift bins are illustrated by the vertical dotted lines. . . . . 126
- 4.5 Same as Fig. 4.2, but after optimisation with a DESI-like selected reference sample. . . . . 128
- 4.6 Same as Fig. 4.3, but after optimisation with a DESI-like selected reference sample. . . . . 128
- 5.1 Sketch of the significance criterion for the tension between two duplicates of a parameter  $\theta$  modelling the observed data in two regimes. The blue curve shows the posterior distribution of the difference between both instances of the parameter. We evaluate the posterior at the origin, indicated by the vertical red dotted line, and infer the fraction of the distribution with a lower posterior density, illustrated by the grey shaded region. This fraction is identified with the probability mass of a one-dimensional Gaussian distribution outside of an interval  $[-m\sigma, m\sigma]$ , where  $m$  denotes the level of the tension in terms of  $\sigma$ . . . . . 136
- 5.2 Redshift distribution of KiDS-1000 galaxies in five tomographic redshift bins selected via their best-fit photometric redshift  $z_B$ . Figure adapted from [Asgari et al. \(2021\)](#). . . . . 138

- 5.3 Marginalised posterior distributions in the  $S_8$ - $\Omega_m$  plane of the consistency test between tomographic redshift bins. The test duplicates cosmological parameters with one set of parameters modelling the signal in one single bin while the other set models the signal in the remaining bins. The covariance between bins is accounted for via the data covariance matrix. The orange contours refer to the cosmological parameters of the isolated bin and the blue contours show the constraints from the remaining bins. The solid contours show the results from sampling the parameter space with the POLYCHORD sampler with emulation of the matter power spectrum with COSMOPOWER, whereas the dotted contours display posteriors obtained with the MULTINEST sampler and calculation of the matter power spectrum with CLASS. . . . . 143
- 5.4 Marginalised posterior distributions of the difference between parameter duplicates in the consistency test between tomographic redshift bins, shown in Figure 5.3. Each panel presents constraints from an analysis that isolates a single redshift bin. . . . 146
- 5.5 Marginalised posteriors distributions in the  $S_8$ - $\Omega_m$  plane in a consistency test with mock data vectors containing an undetected population of galaxies at high redshifts. The left and right panels corresponds to a mock data vector with 5% and 20% of galaxies at high redshifts, respectively. The cross indicates the input cosmological parameters. . . . . 148

- 5.6 Marginalised posterior distributions for  $S_8$ ,  $\Omega_m$ , and  $h$  in a split cosmological analysis in geometry (orange contours) and growth (blue contours) theory regimes and the fiducial analysis with one set of  $\Lambda$ CDM parameters (green contours). Additionally, the red contours show constraints from Planck ([Planck Collaboration et al., 2020b](#)). The leftmost column shows constraints from weak lensing data only, while the remaining columns show the constraints when adding clustering, Lyman- $\alpha$ , and recombination data, respectively. . . . . 152
- 5.7 Marginalised posterior distributions of the difference between parameter duplicates in the split cosmological analysis in geometry and growth theory regimes, shown in Fig. 5.6. The corresponding tension levels are given in Table 5.5. . . . . 153
- A.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. . 163
- A.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. 165

- A.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  $\sigma_{\text{comb}}$  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. . . . . 166
- A.4 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. . . . . 168
- B.1 Left: Scatter plot of the true redshift of galaxies in the photometric sample and the point estimate of the photometric redshift inferred via SED template fitting. Right: Comparison between the true redshift distribution and the redshift distribution inferred from point estimates of the photometric redshift. . . . . 169
- B.2 Illustration of the SOM used in the analysis at different resolutions. The original SOM, trained on the observed colours of galaxies with a resolution of  $R = 200$ , is illustrated in the top left panel. The remaining panels show the SOM with reduced resolution inferred with the clustering method described in Sect. 3.2.3. The colours represent the mean of the true redshift of galaxies in each node. . . . . 171
- B.3 Evolution of the energy of the simulated annealing algorithm. . . . . 172

# List of Tables

- 2.1 Model parameters and their priors for the KV450 cosmic shear analysis, adopted from [Hildebrandt et al. \(2020\)](#). 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. . . . . 79
- 2.2 Results of the iterative fitting of cosmological and nuisance parameters to the KV450 cosmic shear data and comparison to the fiducial KV450 likelihood. When optimising cosmological parameters, we fit  $\omega_{\text{cdm}}$ ,  $\omega_{\text{b}}$ ,  $A_{\text{s}}$ ,  $n_{\text{s}}$ , and  $h$  as well as the nuisance parameters  $A_{\text{IA}}$ ,  $A_{\text{bary}}$ ,  $\delta c$ , and  $A_{\text{c}}$ . When optimising nuisance parameters, we vary the amplitudes of the Gaussian comb. Results are shown for the two most interesting parameters,  $A_{\text{IA}}$  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. . . . . 81

- 4.1 Redshift bins of a sample of reference galaxies with a DESI-like selection function. The sample is divided into nine tomographic bins between  $0 \leq z < 3$  separately for each of the four classes of objects in the DESI-like catalogue: Luminous Red Galaxies (LRGs), Emission Line Galaxies (ELGs), high redshift quasars (QSOs), and a magnitude-limited sample (MagLim). . . . . 127
- 5.1 Sampling parameters and their priors, adopted from [Asgari et al. \(2021\)](#). The upper section lists cosmological parameters, while the lower section shows nuisance parameters. The third column reports whether the parameter is duplicated in the split analysis between redshift bins. The five shift parameters  $\delta_z$  are correlated through their covariance matrix  $C$  and their means  $\boldsymbol{\mu}$  are fixed to their mean values reported by [Hildebrandt et al. \(2021\)](#). . . . 140
- 5.2 Evidence, Bayes factor, suspiciousness, Bayesian model dimensionality, tension probability, and tension level for a split cosmological analysis of cosmic shear data from KiDS-1000. The split separates individual tomographic redshift bins that are modelled with a separate set of cosmological parameters with cross-correlations taken into account via the data covariance matrix. The bottom row provides the evidence and the Bayesian model dimensionality of the fiducial analysis with one set of cosmological parameters. . . . . 145

5.3	Significance of the tension in a split cosmological analysis of cosmic shear data from KiDS-1000 that separates individual redshift bins. Each part of the data vector is modelled with a separate set of cosmological parameters from which we infer the posterior distribution of the difference between parameter duplicates. The level $m$ of the tension is quantified by determining the fraction of the posterior with lower density than the density at the origin which is identified with the probability mass of an one-dimensional Gaussian distribution outside an $m\sigma$ interval. . . . .	146
5.4	Overview of the observables used in the consistency test between background and perturbations in the $\Lambda$ CDM model and their classification into the two regimes. . . . .	151
5.5	Significance of the tension between the geometry and growth theory regime in a split cosmological analysis. Each regime is modelled with a separate set of cosmological parameters from which we infer the posterior distribution of the difference between parameter duplicates. The level $m$ of the tension is quantified by determining the fraction of the posterior with lower density than the density at the origin which is identified with the probability mass of an one-dimensional Gaussian distribution outside an $m\sigma$ interval. . . . .	154
A.1	Comparison between the fiducial KV450 likelihood and the modified likelihood with redshift distribution parameterised by the Gaussian comb model. 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. . . . .	167

# Chapter 1

## Introduction

### 1.1 Homogeneous cosmology

The current cosmological model is based on two fundamental aspects: Einstein's theory of General Relativity and the cosmological principle. Einstein's theory of General Relativity connects the energy content of the Universe to the structure of spacetime. Thus, it allows us to understand the evolution of the Universe on large scales. The cosmological principle states that the Universe is homogeneous and isotropic when viewed on large scales, i.e. the structure of the Universe looks the same for every observer looking in any direction. However, we require a deviation from uniformity at small scales in order to explain the evolution of structure in the Universe, such as galaxies.

The field of cosmology has made rapid progress in the last 30 years, thanks to large data sets from various instruments. In particular, space-based Cosmic Microwave Background (CMB) experiments such as the Cosmic Background Explorer (COBE, [Mather et al., 1990](#)), the Wilkinson Microwave Anisotropy Probe (WMAP, [Bennett et al., 2003](#)), and most recently Planck ([Planck Collaboration et al., 2014](#)), large galaxy surveys such as the 2dF and 6dF Galaxy Survey ([Colless et al., 2001](#); [Jones et al., 2004](#)), the Sloan Digital Sky Survey (SDSS, [Abazajian et al., 2003](#)), the Kilo-Degree Survey (KiDS, [Kuijken et al., 2015](#)), and the Dark Energy Survey (DES, [Abbott et al., 2018](#)), supernovae observations such as [Riess et al. \(1998\)](#), the Supernova Cosmology

Project (SCP, [Perlmutter et al., 1999](#)), SH0ES ([Riess et al., 2016](#)), and Pantheon ([Scolnic et al., 2018](#)), and many other experiments have conducted valuable tests of the predictions of the  $\Lambda$ CDM cosmological model and have significantly improved our understanding of the Universe.

In this introduction we review the key aspects of the theory of General Relativity and the evolution of large-scale structure in the Universe. Additionally, we introduce photometric redshift estimation techniques and weak gravitational lensing, which are the focus of chapters 2, 3, and 4.

### 1.1.1 General Relativity

In Einstein's theory of General Relativity the spacetime metric  $g_{\mu\nu}$  plays a fundamental role. It connects the coordinates  $x^\mu$  of an observer and the invariant line element  $ds$  by

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu. \quad (1.1)$$

The metric incorporates the effects of gravity and is therefore a function of time and space  $g_{\mu\nu}(x^\mu)$ . In a homogeneous and isotropic Universe, the most general metric is the so-called Friedmann Lemaitre Robertson Walker (FLRW) metric, which is in comoving spherical coordinates given by

$$ds^2 = -dt^2 + a^2(t) \left[ \frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2\theta d\phi^2) \right], \quad (1.2)$$

where we adopt the signature  $(-1, 1, 1, 1)$  of the metric tensor and use units with the speed of light  $c = 1$ . Here, the time coordinate is  $t$ , and the spatial coordinates are given by  $r$ ,  $\theta$ , and  $\phi$ . In this equation, the scale factor  $a(t)$  accounts for the expansion of space, relating comoving and physical coordinates by

$$x_{\text{phys}}^i = a(t) x_{\text{comoving}}^i. \quad (1.3)$$

Note that the scale factor can only depend on time, because if it depended on the spatial coordinates, homogeneity would be violated. The imprint of

spatial curvature is encoded in the factor  $k$ . Due to homogeneity and isotropy the curvature must be the same everywhere at a given time. There are three possible scenarios: for  $k = 0$  the Universe is Euclidean and is called flat. For  $k > 0$  it is positively curved and called a closed Universe. For  $k < 0$  the Universe is negatively curved and therefore called open. We can further restrict the possible values of  $k$  by applying transformations to  $r$  and  $a(t)$ :  $\tilde{r} = r \sqrt{|k|}$  and  $\tilde{a}(t) = a(t) / \sqrt{|k|}$ . After this rescaling, the metric is formally the same as in Eq. (1.2), but with the factor  $k$  restricted to 0, +1, or -1, depending on whether the Universe is flat, closed, or open.

The relation between the metric and the matter and energy content of the Universe is given by the Einstein equation:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \quad (1.4)$$

where  $G_{\mu\nu}$  is the Einstein tensor,  $G$  is the gravitational constant, and  $T_{\mu\nu}$  is the energy-momentum tensor. In this equation,  $\Lambda$  denotes the cosmological constant, which was first introduced by Einstein to give a static solution to the equation under the assumption that  $\Lambda$  is constant in time and space. However, in today's cosmological model it accounts for the accelerated expansion of the Universe caused by Dark Energy. Employing the cosmological principle we find that the energy-momentum tensor takes the form of a perfect fluid, given by:

$$T_{\mu\nu} = (p + \rho)u_\mu u_\nu - pg_{\mu\nu}, \quad (1.5)$$

with the pressure  $p$ , the energy density  $\rho$ , and the four-velocity  $u^\mu$  of the fluid. For a comoving observer the velocity is  $u_\mu = (1, 0, 0, 0)$  and therefore the stress-energy tensor  $T_\nu^\mu$  becomes diagonal:

$$T_\nu^\mu = g^{\mu\alpha} T_{\alpha\nu} = \text{diag}(-\rho, p, p, p), \quad (1.6)$$

where we summed over repeated indices. By inserting the FLRW metric into

the Einstein equation, we find the relation between the scale factor and the energy content of the Universe, which is given by the Friedmann equations (Friedmann, 1922, 1924):

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}, \quad (1.7)$$

$$\left(\frac{\ddot{a}}{a}\right) = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}. \quad (1.8)$$

The first Friedmann equation (1.7) gives a relation between the Hubble parameter  $H(t)$ , which is defined as the relative expansion rate

$$H(t) = \frac{\dot{a}(t)}{a(t)}, \quad (1.9)$$

and the homogeneous energy density of the Universe  $\rho(t)$ , the spatial curvature  $k$  and the cosmological constant  $\Lambda$ . The acceleration equation (1.8) gives information about whether the expansion of the Universe is accelerating or decelerating.

### 1.1.2 The content of the Universe

The energy density  $\rho$  entering the first Friedmann equation (1.7) is the sum over all different components present in the Universe. We distinguish between the ultra-relativistic component (radiation) and the non-relativistic component (matter). The associated energy densities  $\rho_R$  and  $\rho_M$  scale differently with the scale factor due to their equation of state. The Friedmann equation becomes

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}(\rho_R + \rho_M) - \frac{k}{a^2} + \frac{\Lambda}{3}. \quad (1.10)$$

Furthermore, we can define the energy density of curvature and Dark Energy as  $\rho_K = -\frac{3k}{8\pi G a^2}$  and  $\rho_\Lambda = \frac{\Lambda}{8\pi G}$ , respectively. Thus, the Friedmann equation becomes:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}(\rho_R + \rho_M + \rho_K + \rho_\Lambda). \quad (1.11)$$

This equation describes the evolution of the Universe. Given the different evolution with the scale factor, each component might have dominated during different epochs in the history of the Universe. We discuss the four components below.

**Radiation:** This component is comprised of photons from the cosmic microwave background (CMB) and neutrinos. The contribution of radiation to the total energy density at the present time is less than 1%.

**Matter:** This component consists of two types of matter. First, baryonic matter which refers to all visible matter that contributes to approximately 5% of the total energy density of the current Universe ([Planck Collaboration et al., 2020b](#)). In the context of cosmology, the term baryon refers to neutrons, protons, and electrons although in particle physics electrons are correctly classified as leptons. The second type of matter does only interact gravitationally, but not through electromagnetic forces, and is therefore called dark matter. The first evidence of dark matter was found through the velocity dispersion of galaxies in the Coma cluster ([Zwicky, 1933](#)) and through galaxy rotation curves ([Rubin and Ford, 1970](#)), while modern experiments observed a contribution of approximately 26% of dark matter to the total energy density ([Planck Collaboration et al., 2020b](#)).

**Curvature:** This component describes the contribution of spatial curvature, as discussed in section 1.1.1, to the total energy density of the Universe.

**Dark Energy:** This component comprises the remaining 69% of the energy density of the Universe and is indistinguishable from a cosmological constant with current observational capabilities. Therefore, it is a primary focus of upcoming surveys such as Euclid ([Laureijs et al., 2011](#)) and the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST; [LSST Science Collaboration et al., 2009](#)). It was first detected through observations of Type Ia supernovae ([Riess et al., 1998](#); [Perlmutter et al., 1999](#)) and is responsible for the accelerated expansion of the Universe at the present time.

For each component we can derive the evolution of the Hubble parameter

during an epoch in which this component is the dominating one. For radiation and matter we first need to determine how the energy density scales with  $a$ . The first Bianchi identity,  $T_{;\mu}^{\mu\nu} = 0$ , yields

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p) = -3\frac{\dot{a}}{a}(1 + w)\rho, \quad (1.12)$$

where we defined the equation of state parameter of the fluid as  $w = p/\rho$ . This equation tells us how the energy gets diluted due to the expansion of the Universe depending on the equation of state. By integrating Eq. (1.12) we find the scaling of  $\rho$  with respect to the scale factor:

$$\rho \propto a^{-3(1+w)}. \quad (1.13)$$

For each given species we insert the corresponding equation of state parameter in order to infer the scaling of the energy density with respect to the scale factor. Via integration of Eq. (1.11) we find the evolution of the scale factor during the epoch in which one of the components is dominating:

1. **Radiation domination:** For ultra-relativistic matter we know the equation of state from statistical thermodynamics, which is given by  $p = \frac{\rho}{3}$ . Thus, Eq. (1.13) yields

$$\rho \propto a^{-4} \Rightarrow H^2 \propto a^{-4} \Rightarrow a \propto t^{\frac{1}{2}}. \quad (1.14)$$

2. **Matter domination** For non-relativistic matter the kinetic energy is negligible, implying  $p = 0$  and thus  $w = 0$ . Then, Eq. (1.13) leads to

$$\rho \propto a^{-3} \Rightarrow H^2 \propto a^{-3} \Rightarrow a \propto t^{\frac{2}{3}}. \quad (1.15)$$

3. **Curvature domination** For curvature, the Friedmann equation yields

$$H^2 \propto \frac{k}{a^2}. \quad (1.16)$$

Here, we have to distinguish between the three possible values of  $k$ . For  $k = 0$  the Universe is flat and there is no effect of curvature at all. For  $k = +1$  and  $k = -1$  we find

$$H^2 \propto a^{-2} \Rightarrow a \propto t. \quad (1.17)$$

However, for  $k = +1$  and assuming that there is no cosmological constant, the expansion stops at some point due to the scaling proportional to  $a^{-2}$  and the Universe recollapses afterwards.

4. **Cosmological constant domination** The cosmological constant is a term in the Friedmann equation which does not scale with  $a$  and therefore leads to an exponential expansion of the Universe. This can also be inferred by treating the cosmological constant as a species contributing to the cosmological fluid. Comparing the cosmological constant to the energy-momentum tensor (1.6), we see that the cosmological constant is equivalent to a fluid with  $T_{\nu}^{\mu} = -\frac{\Lambda}{8\pi G}\delta_{\nu}^{\mu}$ . This is the energy-momentum tensor of a perfect fluid with equation of state  $\rho = -p = \frac{\Lambda}{8\pi G}$ . Eq. (1.12) then yields  $\dot{\rho} = 0$  and therefore we find

$$H^2 = \text{const} \Rightarrow a \propto e^{\frac{\Lambda t}{3}}, \quad (1.18)$$

which is equivalent to what we read off from the Friedmann equation.

As shown above, the Hubble parameter evolves during the four epochs proportional to  $a^{-4}$ ,  $a^{-3}$ ,  $a^{-2}$ , and  $a^0$ . Due to these scalings and the fact that the scale factor is always increasing, the four components can have potentially dominated the expansion of the Universe from early to late times in the order in which they were listed. However, the existence of these terms in the Friedmann equation does not necessarily imply that each individual component dominated the expansion at some point during the history of the Universe. From the proportionality of the scale factor with respect to time during the four epochs we

infer that the Universe is in decelerated expansion during radiation domination and matter domination, while the rate of expansion stays constant during curvature domination. Only during cosmological constant domination we do find accelerated expansion.

By measuring the energy densities of the contents of the Universe and the value of the Hubble parameter at the present time we can extrapolate  $a(t)$  at any time and thus infer the history of the Universe. Evaluating the Friedmann equation at the present time yields

$$1 = \frac{8\pi G}{3H_0^2}\rho_{R,0} + \frac{8\pi G}{3H_0^2}\rho_{M,0} - \frac{k}{a_0^2 H_0^2} + \frac{\Lambda}{3H_0^2} \quad (1.19)$$

$$= \Omega_R + \Omega_M + \Omega_k + \Omega_\Lambda, \quad (1.20)$$

where we defined the relative contributions of the term  $x$  to the present day expansion by  $\Omega_x$  and used the subscript 0 for quantities evaluated today. The Friedmann equation in terms of these quantities at any time is then given by

$$H^2(a) = H_0^2 \left[ \Omega_R \left( \frac{a_0}{a} \right)^4 + \Omega_M \left( \frac{a_0}{a} \right)^3 + \Omega_k \left( \frac{a_0}{a} \right)^2 + \Omega_\Lambda \right], \quad (1.21)$$

where  $H_0$  refers to the value of the Hubble parameter at the present time. Recent observations taken by the Planck satellite have shown that the spatial curvature of our Universe is very close to zero with  $\Omega_K = 0.001 \pm 0.002$  (Planck Collaboration et al., 2020b) therefore we can neglect  $\Omega_K$  and parameterise the evolution of our Universe by the four parameters  $\Omega_R$ ,  $\Omega_M$ ,  $\Omega_\Lambda$  and  $H_0$ .

## 1.2 The Large Scale Structure of the Universe

### 1.2.1 Matter power spectrum

General Relativity, as discussed in section 1.1, describes the evolution of the homogeneous background of the Universe on large scales. This is consistent with CMB experiments which only find temperature fluctuations of order  $\Delta T/T \sim 10^{-5}$  (Smoot et al., 1992). However, the existence of the so-called

cosmic web, which is comprised of structures such as galaxy clusters, voids, and filaments, requires perturbations from the homogeneous background in order for these structures to be formed. In the cosmological model it is postulated that after the Big Bang the Universe went through a period of accelerated expansion, called inflation. Inflation is capable of explaining the statistical homogeneity of the Universe at large scales and the existence of structures, which originate from quantum fluctuations that grow to cosmological scales during the period of accelerated expansion.

We parameterise perturbations from the homogeneous background via the density contrast

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \quad (1.22)$$

where  $\bar{\rho}(t)$  denotes the mean matter density of the Universe at time  $t$  and  $\rho(\mathbf{x}, t)$  is the matter density at a position  $\mathbf{x}$  and time  $t$ . Assuming the presence of primordial fluctuations after the end of inflation, we describe the amplitude of the fluctuations as a power spectrum  $P(\mathbf{k})$  using

$$(2\pi)^3 P(\mathbf{k}) \delta_D(\mathbf{k} + \mathbf{k}') = \langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle, \quad (1.23)$$

where  $\delta_D$  is the Dirac delta function and the angular brackets denote the ensemble average. Here,  $\delta(\mathbf{k})$  is the Fourier transform of the matter power spectrum which is defined by

$$\delta(\mathbf{k}) = \int e^{-i\mathbf{k}\cdot\mathbf{x}} \delta(\mathbf{x}) d\mathbf{x}. \quad (1.24)$$

We assume that the primordial power spectrum  $P_0$  is only dependent on the amplitude  $k$  of the wave vector. Therefore, we find

$$P_0(k) = A_s \left( \frac{k}{k_0} \right)^{n_s}, \quad (1.25)$$

where  $A_s$  is the amplitude of primordial fluctuations and  $n_s$  denotes the scalar spectral index with  $n_s = 1$  corresponding to a scale-invariant Harrison-Zel'dovich

power spectrum. Additionally,  $k_0$  denotes the pivot scale which is arbitrarily set to  $k_0 = 0.05 \text{ Mpc}^{-1}$ . At later times, the power spectrum is modulated by the transfer function  $T(k)$  that depends on cosmological parameters describing the growth of matter perturbations and by the growth factor  $D(t)$  describing how initial perturbations grow with time. If we assume Gaussian initial conditions, i.e. we assume that the initial perturbations are generated from a Gaussian random field as predicted by the theory of inflation (Guth, 1981), and assuming linear perturbations, the power spectrum can be computed via

$$P(k, t) = A_s \left( \frac{k}{k_0} \right)^{n_s} T^2(k) D^2(t). \quad (1.26)$$

The time and scale dependencies of the power spectrum can be separated because of the homogeneity of the Universe.

The matter power spectrum quantifies the scale dependence of density perturbations which can be measured by various cosmological probes. The shape of the power spectrum is determined by the evolution of perturbations during different times. In general, these can be derived from relativistic perturbation theory which, for the purpose of determining the evolution of cold dark matter, can be approximated by Newtonian perturbation theory. However, the formation of stars, galaxies, and clusters are non-linear processes that cannot be described by linear perturbation theory and that can only be computed numerically. Therefore, the non-linear growth of structures is commonly modelled using numerical simulations.

In Fourier space, the evolution of density perturbations  $\delta$  is given by (Dodelson, 2003)

$$\ddot{\delta} + 2H\dot{\delta} + \left( \frac{c_s^2 k^2}{a^2} - 4\pi G\bar{\rho} \right) \delta = 0, \quad (1.27)$$

where  $c_s$  denotes the sound speed. During radiation domination, relativistic perturbation theory predicts a negligible density contrast, so that Eq. (1.14) and Eq. (1.27) yield

$$\ddot{\delta}_{\text{cdm}} + \frac{\dot{\delta}_{\text{cdm}}}{t} = 0. \quad (1.28)$$

Thus, we find either  $\delta_{\text{cdm}} = \text{const.}$  or  $\delta_{\text{cdm}} \propto \ln t$ , so that matter perturbations either grow slowly or do not grow at all during this era. On the other hand, a matter dominated universe can be parameterised via  $\Omega_{\text{m}} = 1$ , so that Eq. (1.15) and Eq. (1.27) reduce to

$$\ddot{\delta}_{\text{m}} + \frac{4\dot{\delta}_{\text{m}}}{3t} - \frac{2\delta_{\text{m}}}{3t^2} = 0, \quad (1.29)$$

where we used  $c_s^2 = 0$ . Therefore, we find  $\delta_{\text{m}} \propto \frac{1}{t}$  and  $\delta_{\text{m}} \propto t^{2/3}$ , so that perturbations grow with a power law. However, if the Universe does not expand, the second term in Eq. (1.27) vanishes, which leads to an exponential growth of perturbations. Finally, during Dark Energy domination we find  $a \propto \exp(Ht)$  so that Eq. (1.27) yields

$$\ddot{\delta}_{\text{m}} + 2H\dot{\delta}_{\text{m}} = 0. \quad (1.30)$$

Therefore, we find  $\delta_{\text{m}} = \text{const.}$  or  $\delta_{\text{m}} \propto \exp(-2Ht)$ . Thus, Dark Energy domination leads to a suppression of the growth of structure.

So far, we have only considered perturbations that are smaller than the Hubble radius, which is defined by

$$r_{\text{H}}(t_0) = \frac{c}{H_0}. \quad (1.31)$$

The Hubble radius defines a sphere centred on an observer at the present time outside of which objects move away from the observer with a speed faster than the speed of light. Therefore, the Hubble radius is interpreted as the horizon of the observable Universe. Perturbations outside the horizon grow with a different rate than perturbations inside the horizon, which can be derived as follows: We consider two patches of the Universe inside and outside the horizon with the same Hubble rate. The Friedmann equations yield

$$\frac{8\pi G}{3}\rho_{\text{out}} - \frac{\kappa_{\text{out}}}{a^2} = \frac{8\pi G}{3}\rho_{\text{in}}, \quad (1.32)$$

where  $\rho_{\text{out}}$  and  $\rho_{\text{in}}$  denote the density outside and inside the horizon, respectively, and  $\kappa_{\text{out}}$  is the curvature outside the horizon. Additionally, we assumed a flat Universe inside the horizon. We find

$$\frac{\rho_{\text{in}} - \rho_{\text{out}}}{\rho_{\text{out}}} = \frac{3}{8\pi G} \frac{\kappa_{\text{out}}}{a^2 \rho_{\text{out}}}. \quad (1.33)$$

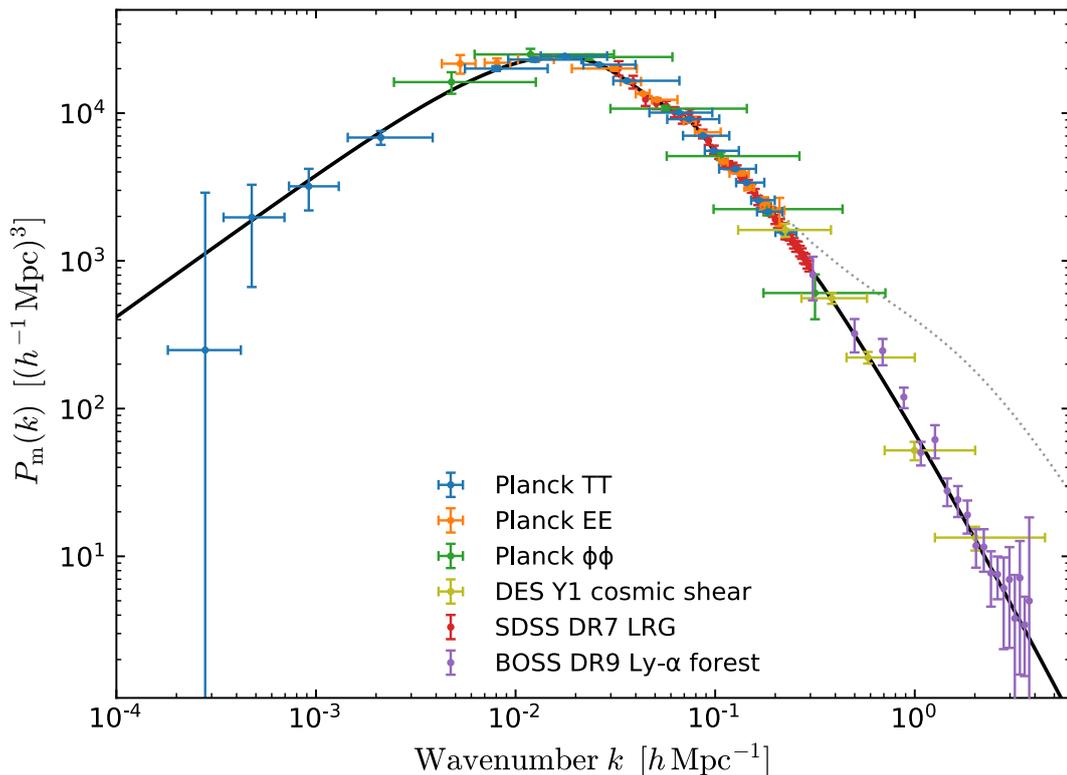
Therefore, perturbations outside the horizon evolve like

$$\delta \propto \frac{1}{a^2 \rho}. \quad (1.34)$$

From this equation we can infer the evolution of perturbations outside the horizon, which grow proportional to  $t$  during radiation domination and proportional to  $t^{2/3}$  during matter domination while being suppressed during Dark Energy domination.

The growth of perturbations then allows us to explain the shape of the matter power spectrum, which is illustrated by the black line in Fig. 1.1. An important quantity characterising the power spectrum is the scale of the matter-radiation equality  $k_{\text{eq}}$  that determines the turnover of the power spectrum. Perturbations on large scales with  $k < k_{\text{eq}}$  enter the horizon during matter domination. Therefore, they grow proportional to  $t$  during radiation domination and proportional to  $t^{2/3}$  when the Universe is matter dominated. On large scales, the matter power spectrum therefore scales like the primordial power spectrum with  $P \propto k^{n_s}$  and  $n_s \approx 1$ . On the other hand, perturbations with  $k > k_{\text{eq}}$  enter the horizon during radiation domination. Therefore, they grow as  $t$  during radiation domination until they enter the horizon, when their growth stops. They then grow again during matter domination. Therefore, small-scale perturbations are suppressed compared to large-scale perturbations and the matter power spectrum is a decreasing function of  $k$  on small scales.

Additionally, Fig. 1.1 shows measurements of the linear matter power spectrum from the CMB observations (Planck Collaboration et al., 2020a), weak lensing (Troxel et al., 2018), galaxy clustering (Reid et al., 2010), and



**Figure 1.1:** Measurements of the linear matter power spectrum from CMB observations (Planck Collaboration et al., 2020a), weak lensing (Troxel et al., 2018), galaxy clustering (Reid et al., 2010), and the Lyman-alpha forest (Palanque-Delabrouille et al., 2015). The solid black line shows the  $\Lambda$ CDM model prediction for the linear matter power spectrum while the black dotted line shows the impact of non-linear clustering at redshift  $z = 0$ .

the Lyman-alpha forest (Palanque-DeLabrouille et al., 2015). However, many experiments do not probe the linear power spectrum directly, but instead require non-linear corrections. In particular, probes that utilise galaxies as tracers of the matter power spectrum in reality do not measure the matter power spectrum, but instead measure the galaxy power spectrum. On large scales, these power spectra can be related via a linear bias model (Kaiser, 1984), but this model breaks down on small scales, where non-linear processes determine the formation of stars and galaxies and the clustering of galaxies. The dark matter distribution on non-linear scales can for example be modelled using the popular halo model formalism, which assumes that all dark matter is distributed in dark matter halos. Galaxies are then formed within the dark

matter halos, so that they trace the distribution of dark matter. In this model, the dark matter distribution is summarised by the halo mass function (Jenkins et al., 2001), which describes the number density of dark matter halos with a given mass at a given redshift, the halo density profile (Navarro et al., 1996), which describes the distribution of mass within dark matter halos, and the halo bias function (Zehavi et al., 2005), which relates dark matter halos to the underlying dark matter distribution. By accounting for non-linear corrections, the galaxy power spectrum can then be related to the matter power spectrum.

By observing the distribution of galaxies on the sky we can probe the matter power spectrum and therefore we can use observations to derive constraints on cosmological parameters using Eq. (1.26). However, galaxies only give insight into the distribution of baryonic matter which acts as a tracer of the underlying dark matter density. A common assumption is a linear bias model (Kaiser, 1984) that relates the visible galaxy overdensity  $\delta_g$  to the total matter density via

$$\delta_g = b_g \delta, \quad (1.35)$$

where  $b_g$  denotes the linear galaxy bias that can in general depend on the scale  $k$  and on galaxy properties such as luminosity, type, colour, stellar mass, and redshift (Fry, 1996; Mann et al., 1998; Tegmark and Peebles, 1998). In this model, the observed galaxy power spectrum is related to the matter power spectrum via

$$P_g(k, z) = b_g^2 P(k, z). \quad (1.36)$$

This relation is utilised in studies of galaxy clustering (see for example Abazajian et al., 2003). However, a linear bias model in general is too simplistic and thus the modelling of the galaxy bias is a crucial task that can potentially induce systematic errors in the inferred constraints on cosmological parameters. Alternatively, measurements of weak gravitational lensing of galaxies is a powerful probe that gives access to the combined distribution of baryonic and dark matter. Therefore, it is an essential part of the cosmological analyses of

upcoming large-scale galaxy surveys such as Euclid and LSST. However, both galaxy clustering and weak gravitational lensing require information about the distribution of galaxies in three dimensions and hence require estimates of the distance of the observed galaxies. In an expanding Universe the wavelength of a photon emitted from an object is stretched proportional to the scale factor, so that the observed wavelength is larger than the emitted wavelength. This increase in the observed wavelength is usually defined in terms of redshift

$$z \equiv \frac{\lambda_o - \lambda_e}{\lambda_e}, \quad (1.37)$$

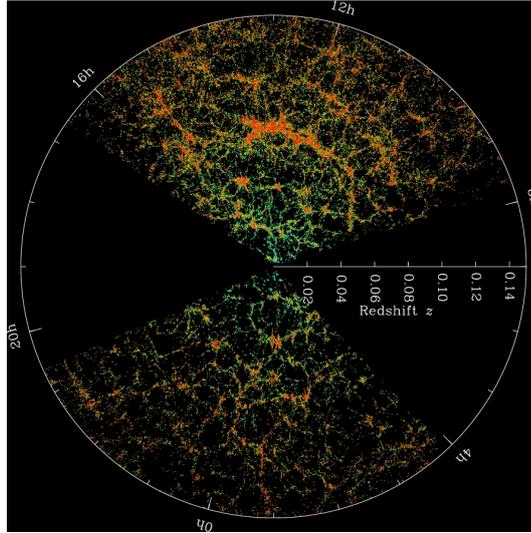
which is used as a measure of the distance to the observed object. In section 1.3 we discuss how we can measure the redshift of large samples of galaxies observed in modern extragalactic galaxy surveys.

## 1.2.2 Large scale structure probes

As discussed in the previous section, the matter content of the Universe is mostly comprised of dark matter, which is not directly observable due to its lack of electromagnetic interactions. Thus, the matter power spectrum can only be probed indirectly through observations of baryonic matter. Although a significant proportion of the baryonic matter is located in intergalactic gas clouds, one of the most powerful probes of the matter spectrum are observations of astrophysical objects, in particular galaxies. These observations give access to the density field of galaxies, which serves as a biased tracer of the underlying dark matter distribution. In this section we review two of the most important probes of the large scale structure of the Universe: galaxy clustering and weak gravitational lensing.

### 1.2.2.1 Galaxy clustering

A direct measurement of the galaxy density field is provided by extragalactic surveys which observe the angular position of galaxies on the sky as well as the distance of galaxies in the form of redshift. Fig. 1.2 shows a map of galaxies observed by SDSS up to a redshift of  $z = 0.15$ . We observe that galaxies are not



**Figure 1.2:** Map of galaxies observed by SDSS. Each dot represents a galaxy with the colour representing the g-r colour. Image credit: M. Blanton and SDSS. Source: <https://www.sdss.org/science/orangepie/>, published under a Creative Commons Attribution license (CC-BY).

distributed randomly, but aggregate in so-called groups and clusters that are gravitationally bound and form structures, called filaments. By analysing these structures we can measure the three-dimensional power spectrum of galaxies,  $P_g(\mathbf{k})$ , which can be related to the matter power spectrum by assuming a bias model as outlined in section 1.2.1.

The most important statistic to characterise the spatial distribution of galaxies is the two-point correlation function, which describes the excess of galaxy pairs within a given volume relative to a random distribution. Considering a point  $\mathbf{x}$  and an average galaxy number density  $\bar{n}$ , the probability of finding a galaxy in a volume element  $dV$  is given by

$$P = \bar{n}dV, \quad (1.38)$$

which is independent of  $\mathbf{x}$  if the Universe is statistically homogeneous. The probability of finding a galaxy at  $\mathbf{x}$  and at the same time finding a second galaxy at location  $\mathbf{y}$  is given by

$$P = (\bar{n}dV)^2 (1 + \xi_g(\mathbf{x}, \mathbf{y})). \quad (1.39)$$

In this equation we introduced the two point correlation function of galaxies,  $\xi_g(\mathbf{x}, \mathbf{y})$ , characterising the probability of finding an excess of galaxies due to clustering. For  $\xi_g > 0$  we find an excess of galaxies in the volume element while for  $\xi_g < 0$  we find a shortage of galaxies. If  $\xi_g = 0$ , galaxies are uncorrelated and the probability of finding galaxy pairs reduces to the product of two random distributions. Analogously, we define the correlation function of the total matter density  $\rho_m$  as

$$\langle \rho_m(\mathbf{x}) \rho_m(\mathbf{y}) \rangle = \bar{\rho}_m^2 (1 + \xi_m(\mathbf{x}, \mathbf{y})), \quad (1.40)$$

where we defined  $\xi_m(\mathbf{x}, \mathbf{y}) = \langle \delta(\mathbf{x}) \delta(\mathbf{y}) \rangle$  and used  $\langle \delta(\mathbf{x}) \rangle = 0$ , which follows from Eq. (1.22). Since we assume the Universe to be homogeneous and isotropic, the correlation function can only depend on the separation of the two coordinates  $r = |\mathbf{x} - \mathbf{y}|$ . Moreover, since we only observe a single Universe, we employ the ergodic hypothesis allowing us to replace the ensemble average with a spatial average, which makes the correlation function accessible in practice.

Observationally, we can calculate the correlation function by counting the number of galaxy pairs of a given separation  $r$  relative to the number of galaxy pairs in a randomly distributed sample of galaxies. One particular estimator of the correlation function is then given by

$$\xi_g(r) = \frac{DD(r) - RR(r)}{RR(r)}, \quad (1.41)$$

where  $DD(r)$  denotes the number of galaxy pairs in a bin of width  $\Delta r$  in  $\left[ r - \frac{\Delta r}{2}, r + \frac{\Delta r}{2} \right]$  and  $RR(r)$  is the number of pairs in a randomly distributed sample. This is the simplest estimator for the two-point correlation function of galaxies. However, alternative estimators have been studied, some of which are preferable to the standard estimator given in Eq. 1.41 due to being less sensitive to noise (see for example [Kerscher et al., 2000](#)). In particular, one of the most commonly used estimators is the Landy-Szalay estimator ([Landy and](#)

Szalay, 1993), which is defined as

$$\xi_g^{\text{LS}}(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)}. \quad (1.42)$$

Here,  $DR(r)$  denotes the number of pair counts among the data and the random sample. In chapter 3 we will apply the Landy-Szalay estimator to calculate the cross-correlation between samples of photometrically and spectroscopically observed galaxies.

Spectroscopic observations of galaxies have shown that in a certain range of separations the galaxy correlation function can be approximated with a power law

$$\xi_g(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \quad (1.43)$$

with a correlation length  $r_0 = 5.05 \pm 0.26h^{-1}\text{Mpc}$  and a slope  $\gamma = 1.67 \pm 0.03$  found by Hawkins et al. (2003) on scales  $0.1h^{-1}\text{Mpc} < r < 12h^{-1}\text{Mpc}$  for galaxies in the 2dF Galaxy Redshift Survey. Eq. 1.43 requires information about the position of galaxies in 3D. However, accurate distance measurements are difficult to obtain. As an alternative, we can measure the clustering of galaxies projected onto two dimensions on the sky. The angular two point correlation  $w(\theta)$  can be obtained from the three-dimensional correlation function via integration along the line of sight (see for example Peebles, 1980). Again assuming the power law defined in Eq. (1.43), we find

$$w(\theta) = A\theta^{1-\gamma}, \quad (1.44)$$

with an amplitude  $A$  that can be derived from the real-space correlation function if the redshift distribution of galaxies is known. Analogous to Eq. (1.41), we can determine the angular correlation function from observations by counting galaxy pairs in bins of angular separation  $\theta$ .

In practice, we can relate both the galaxy correlation function and the matter density correlation function to the corresponding power spectra. Under the assumption that the distribution of galaxies in the Universe traces the

underlying dark matter distribution, we derive constraints on the matter power spectrum and thus constrain cosmological parameters. This, however, proves challenging since it requires an assumption about the relation between galaxy and matter power spectra. A popular choice is the linear bias model given in Eq. (1.36) which, however, in general is dependent on various galaxy properties. For example, [Zehavi et al. \(2005\)](#) found that SDSS galaxies are more clustered the more luminous they are. Moreover, red galaxies show a higher clustering amplitude and a steeper slope than blue galaxies. Therefore, bias parameters of different galaxy samples need to be accounted for in cosmological analyses of galaxy clustering. An option to break the degeneracy between cosmological parameters and galaxy bias is to measure clustering as a function of cosmic time, for example through baryon acoustic oscillations and redshift-space distortions, since we expect the clustering of galaxies to evolve over time due to the growth of structures. Thus, the determination of the redshift of galaxies, discussed in section 1.3, is a vital tool that enables such analyses of galaxy clustering.

### 1.2.2.2 Weak gravitational lensing

The gravitational lensing effect is the deflection of light emitted by distant galaxies caused by gravitational fields. Such gravitational fields are due to massive objects, acting as lenses, distributed between the observer and the observed source galaxy. This effect is commonly categorised into two types. First, if the deflection is strong enough, multiple images of the same galaxy can be identified in the same observation, the galaxy image can be distorted into arcs, or, if the source galaxy is perfectly aligned the lens, the image can appear as a ring. This effect is referred to as strong gravitational lensing. However, the source galaxy needs to be closely aligned with the lens in order to create the strong gravitational lensing effect. Thus, strong gravitational lenses are comparatively rare. The second type of gravitational lensing, called weak gravitational lensing, is a lot more common. This effect manifests itself when the deflection of the light is small, leading to a distortion of the shape and size of the distant galaxy, which is often unobservable by visual inspection of the

galaxy image since we do not know the original size and shape of the source galaxy. Therefore, weak lensing studies require an ensemble of galaxy images in order to statistically measure the distortion of galaxy shapes that can be directly related to the matter power spectrum.

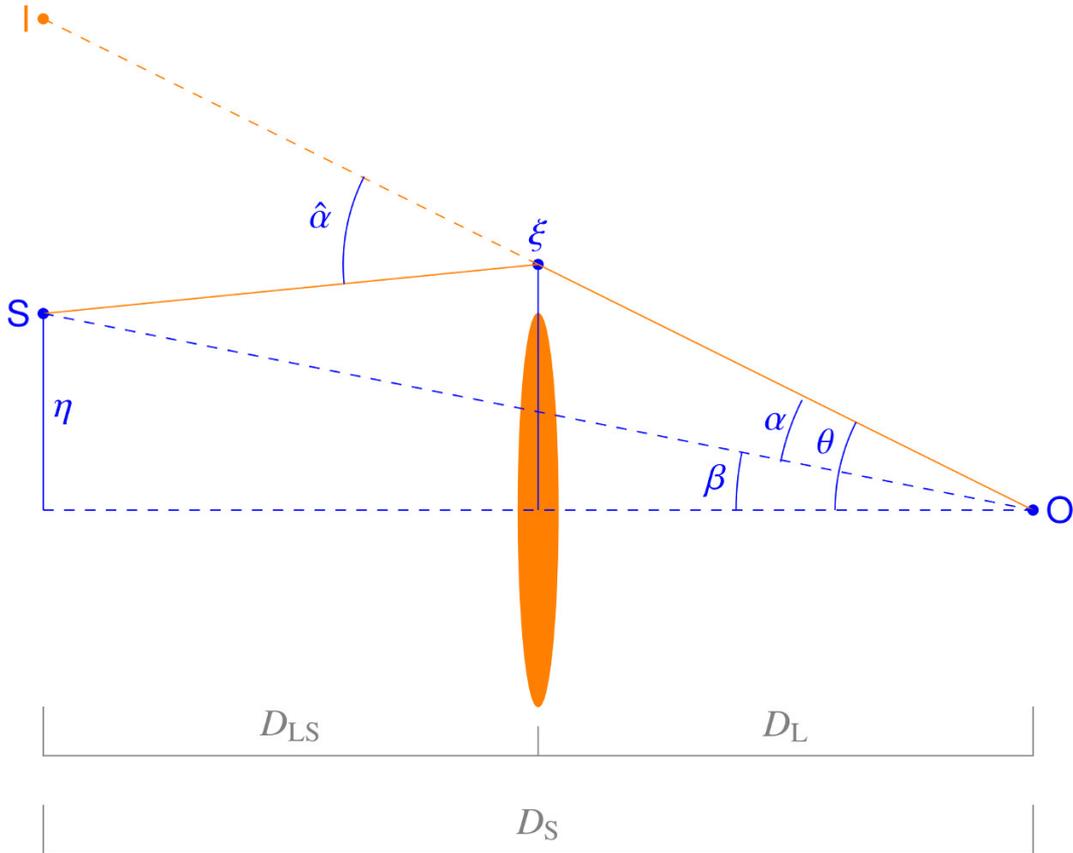
The weak gravitational lensing effect is a powerful probe of cosmology, since it directly maps all gravitating matter along the line of sight and does not require knowledge about the relation between the galaxy and matter distribution. While cosmic shear by the large scale structure was first detected by Kaiser et al. (2000), Bacon et al. (2000), Wittman et al. (2000), and Van Waerbeke et al. (2000), recent cosmological analyses from the three current weak lensing experiments KiDS (Asgari et al., 2021), HSC (Hikage et al., 2019), and DES (Amon et al., 2022) have showcased the potential of cosmic shear measurements as a probe of the cosmological model and therefore it is a primary probe for upcoming galaxy surveys, such as Euclid and LSST.

To model the weak lensing effect we follow Bartelmann and Maturi (2017) and consider the lens system illustrated in Fig. 1.3. We make use of the thin-lens approximation, where the extent of the lens along the line of sight is small compared to the distance between the observer and the source galaxy. In this approximation we can describe light paths as straight lines. The observer is located at position O, the source galaxy at S, and the image of the galaxy appears at I. The distances between the observer and the source galaxy, the observer and the lens, and the lens and the source galaxy are  $D_S$ ,  $D_L$ , and  $D_{LS}$ , respectively. We read off the angular position  $\beta$  of the source from the point of view of the observer

$$\beta = \theta - \alpha, \quad (1.45)$$

where  $\theta$  denotes the angular position of the image from the point of view of the observer and the reduced deflection angle is  $\alpha$ . If we use angular diameter distances  $D$ , we find that the Euclidean relation for the separation  $s$  between two lines enclosing an angle  $\gamma$

$$s = D \cdot \gamma \quad (1.46)$$



**Figure 1.3:** Geometry of a thin-lens system. The observer is located at position O, the source galaxy at S, and the image of the galaxy appears at I. The distances between the observer and the source galaxy, the observer and the lens, and the lens and the source galaxy are  $D_S$ ,  $D_L$ , and  $D_{LS}$ , respectively. Figure adapted from (Bartelmann and Maturi, 2017).

still holds. Therefore, we can write the reduced deflection angle as

$$\alpha \equiv \frac{D_{LS}}{D_S} \hat{\alpha}. \quad (1.47)$$

We employ Fermat's principle stating that light travelling between two points takes the path that minimises travel time, so that the deflection angle is given by

$$\hat{\alpha} = -2 \int \nabla_{\perp} \Phi d\lambda, \quad (1.48)$$

where  $\Phi$  denotes the gravitational potential,  $\nabla_{\perp}$  is the gradient perpendicular to the light ray, and  $d\lambda$  parameterises the distance travelled. Instead of perpendicular distances, in weak lensing we typically consider the position of

galaxies in terms of their angular position  $\boldsymbol{\theta}$  on a sphere. Therefore, we replace the gradient  $\nabla_{\perp} = D_L^{-1} \nabla_{\theta}$ . Using Eqs. (1.47) and (1.48) we find

$$\boldsymbol{\alpha} = \nabla_{\theta} \left( 2 \frac{D_{\text{LS}}}{D_L D_S} \int \Phi dz \right) \equiv \nabla_{\theta} \psi, \quad (1.49)$$

where we defined the lensing potential  $\psi$ .

In general, the assumption of a thin lens does not hold since the light emitted by the source galaxy is continuously distorted by the large-scale structure of the Universe. Therefore, we need to generalise this expression for an extended lens. This can be achieved by using the flat-sky approximation and integrating over comoving distances. Under the assumption of a flat Universe we find

$$\psi(\boldsymbol{\theta}) = 2 \int_0^{\chi_S} \frac{\chi_S - \chi}{\chi_S \chi} \Phi(\chi_{\perp}(\boldsymbol{\theta}), \chi_{\parallel}) d\chi. \quad (1.50)$$

Here,  $\chi_S$  denotes the comoving distance to the source galaxy and the gravitational potential is evaluated at positions  $\chi_{\parallel}$  parallel and  $\chi_{\perp}(\boldsymbol{\theta})$  perpendicular to the line of sight.

Assuming a source much smaller than the typical scale of variations in the deflection angle, we can linearise the lens equation (1.45) using a Taylor approximation. We find

$$\delta\boldsymbol{\beta} \approx \mathcal{A} \delta\boldsymbol{\theta}, \quad (1.51)$$

where  $\delta\boldsymbol{\beta}$  and  $\delta\boldsymbol{\theta}$  denote the extent of the source and the image, respectively and  $\mathcal{A}$  is the Jacobian matrix of the mapping between the source and the image, defined as

$$\mathcal{A}_{ij} = \frac{\partial \beta_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial^2 \psi}{\partial \theta_i \partial \theta_j} \equiv \delta_{ij} - \psi_{ij}. \quad (1.52)$$

Eq. (1.51) states that the mapping between the source and its image is proportional to the curvature of the lensing potential  $\psi_{ij}$ , which in the absence of a lensing potential reduces to an identical mapping. The Jacobian matrix is

often parameterised as

$$\mathcal{A} = \begin{pmatrix} 1 - \kappa & 0 \\ 0 & 1 - \kappa \end{pmatrix} + \begin{pmatrix} -\gamma_1 & -\gamma_2 \\ -\gamma_2 & \gamma_1 \end{pmatrix}, \quad (1.53)$$

with the gravitational shear

$$\gamma = \gamma_1 + i\gamma_2 = \frac{1}{2}(\psi_{11} - \psi_{22}) + i\psi_{12}, \quad (1.54)$$

and the convergence

$$\kappa = \frac{1}{2}(\psi_{11} + \psi_{22}) = \frac{1}{2}\nabla^2\psi. \quad (1.55)$$

Eq. (1.53) illustrates two different effects of the lensing potential acting on a galaxy image. The first term describes a magnification of the source galaxy, while the second term is associated with a change in the shape of the galaxy.

In observations we only measure the separation  $\delta\boldsymbol{\theta}$  of the image. Therefore, we want to infer the separation  $\delta\boldsymbol{\beta}$  of the original source galaxy, which requires the inverse of Eq. (1.51):

$$\delta\boldsymbol{\theta} \approx \mathcal{A}^{-1}\delta\boldsymbol{\beta}. \quad (1.56)$$

The images are typically only slightly distorted so that  $\kappa \ll 1$  and  $|\gamma^2| \ll 1$ . Thus, the Jacobian matrix is invertible and reads

$$\mathcal{A}^{-1} = \mu \begin{pmatrix} 1 - \kappa + \gamma_1 & \gamma_2 \\ \gamma_2 & 1 - \kappa - \gamma_1 \end{pmatrix}, \quad (1.57)$$

where the magnification factor is defined as

$$\mu = \frac{1}{\det\mathcal{A}} = \frac{1}{(1 - \kappa)^2 - \gamma^2} \approx 1 + 2\kappa. \quad (1.58)$$

The distortion of the galaxy image causes the shape of a circular source galaxy to become elliptical. The ellipticity of a galaxy image is typically defined

via

$$\epsilon = \frac{a-b}{a+b}, \quad (1.59)$$

where  $a$  and  $b$  denote the semi-major and semi-minor axes, respectively. However, the observed galaxies in general are non-circular with an intrinsic ellipticity  $\epsilon_s$ . Thus, the observed ellipticity reads

$$\epsilon = \frac{\epsilon_s + g}{1 + \epsilon_s g^*} \approx \epsilon_s + g, \quad (1.60)$$

with the reduced shear

$$g \equiv \frac{\gamma}{1 - \kappa}. \quad (1.61)$$

Assuming that galaxies are randomly oriented, the ensemble average of intrinsic galaxy ellipticities is zero, i.e.  $\langle \epsilon_s \rangle = 0$ . Therefore, for a large galaxy sample, we find

$$\langle \epsilon \rangle = \gamma. \quad (1.62)$$

However, if we average over a large area of the sky we expect the shear signal to vanish since we assume isotropy of the Universe on large scales. Therefore, we typically employ two-point statistics to measure the effects of weak gravitational lensing. To derive the angular power spectrum of cosmic shear, we first transform  $\kappa$ ,  $\gamma_1$ , and  $\gamma_2$ , given in Eq. (1.54) and Eq. (1.55), to Fourier space. We find

$$2\hat{\kappa} = -l^2 \hat{\psi}, \quad 2\hat{\gamma}_1 = -(l_1^2 + l_2^2) \hat{\psi}, \quad \hat{\gamma}_2 = -l_1 l_2 \hat{\psi}, \quad (1.63)$$

with hats denoting Fourier transformed quantities. We calculate

$$4|\hat{\gamma}|^2 = ((l_1^2 - l_2^2)^2 + 4l_1^2 l_2^2) 4|\hat{\psi}|^2 = (l_1^2 + l_2^2)^2 |\hat{\psi}|^2 = 4|\hat{\kappa}|^2. \quad (1.64)$$

Thus, we find equality between the power spectra of convergence and shear. In Fourier space the convergence power spectrum is related to the convergence

two-point correlation function via

$$\langle \hat{\kappa}(\ell)\hat{\kappa}^*(\ell') \rangle = (2\pi)^2 \delta_D^{(2)}(\ell - \ell') C_\kappa(\ell). \quad (1.65)$$

Using  $\kappa = \frac{1}{2}\nabla^2\psi$  and the Poisson equation we find

$$\kappa = 4\pi G \int_0^{\chi_S} \frac{\chi(\chi_S - \chi)}{\chi_S} a^2 \bar{\rho}_m \delta d\chi = \frac{3H_0^2 \Omega_m}{2} \int_0^{\chi_S} \frac{\chi(\chi_S - \chi)}{a\chi_S} \delta d\chi, \quad (1.66)$$

with the density contrast  $\delta$  defined in Eq. (1.22). The second equality follows by comparing Eq. (1.11) and Eq. (1.21), from which we find the mean matter density  $\bar{\rho}_m = 3H_0^2 \Omega_m / 8\pi G a^3$ . Using the Limber approximation (Kaiser, 1992), we find that the angular power spectrum of a quantity  $x(\boldsymbol{\theta})$  is given by

$$C_x(\ell) = \int_0^{\chi_S} \frac{w^2(\chi)}{\chi^2} P_x \left( \frac{\ell + 1/2}{\chi} \right) d\chi, \quad (1.67)$$

where  $P_x$  denotes the 3D power spectrum and  $w(\chi)$  is a weight function. From Eq. (1.66) we identify the weight function

$$w(\chi) = \frac{3H_0^2 \Omega_m}{2} \frac{\chi(\chi_S - \chi)}{a\chi_S}, \quad (1.68)$$

so that the convergence power spectrum, which is equal to the shear power spectrum, becomes

$$C_\gamma(\ell) = C_\kappa(\ell) = \frac{9}{4} H_0^4 \Omega_m^2 \int_0^{\chi_S} \left( \frac{\chi_S - \chi}{a\chi_S} \right)^2 P_\delta \left( \frac{\ell + 1/2}{\chi}, \chi \right) d\chi, \quad (1.69)$$

with the matter power spectrum  $P_\delta$ .

While Eq. (1.69) describes the shear power spectrum of galaxies located in a single source plane at  $\chi_S$ , cosmic shear surveys observe galaxies that are distributed over a long distance range. Additionally, weak lensing analyses are performed tomographically by dividing the observed sample into slices along the line of sight in order to enhance the precision with which cosmological parameters can be determined (Hu, 1999). Thus, we require a modelling of the

cosmic shear signal in each tomographic bin and the cross-correlation signal between bins, which arises since lensing is a cumulative effect. In this case, the cross-correlation power spectrum between two bins  $\alpha$  and  $\beta$  is given by

$$C_{\gamma}^{(\alpha\beta)}(\ell) = \int_0^{\chi_H} \frac{q^{(\alpha)}(\chi)q^{(\beta)}(\chi)}{\chi^2} P_{\delta} \left( \frac{\ell+1/2}{\chi}, \chi \right) d\chi, \quad (1.70)$$

where  $\chi_H$  is the comoving horizon distance. Here, we defined the lensing efficiency by

$$q^{(\alpha)}(\chi) = \frac{3H_0^2\Omega_m}{2} \frac{\chi}{a(\chi)} \int_{\chi}^{\chi_H} n_{\chi}^{(\alpha)}(\chi') \frac{\chi' - \chi}{\chi'} d\chi', \quad (1.71)$$

where  $n_{\chi}^{(\alpha)}(\chi)$  denotes the distribution of sources in redshift bin  $\alpha$ . In chapter 2 we present a flexible model to parameterise redshift distributions of tomographic redshift bins and develop a method to propagate residual uncertainties in the redshift distribution into a weak lensing analysis.

### 1.3 Photometric redshifts

Modern extragalactic galaxy surveys observe millions of objects and aim to construct three-dimensional maps of the Universe from two-dimensional images of galaxies on the sky. To do so, we need to determine the distance of the observed objects. Distances are commonly inferred from the cosmological redshift, which provides a measure of the apparent motion of an object away from the observer due to Hubble's law. A photon with wavelength  $\lambda_e$ , emitted by an object moving away from the observer, loses energy while travelling towards the observer due to the expansion of the Universe. Therefore, it is detected at a longer wavelength  $\lambda_o$ . The redshift of the observed object is defined as the ratio between emitted and observed wavelength:

$$z \equiv \frac{\lambda_o - \lambda_e}{\lambda_e}. \quad (1.72)$$

Since photons travel on geodesics with  $ds^2 = 0$ , we can use Eq. (1.2) to infer a relation between wavelength and scale factor (see for example [Peacock, 1998](#)):

$$\frac{\lambda_o}{\lambda_e} = \frac{a_o}{a_e}, \quad (1.73)$$

where  $a_e$  denotes the scale factor at the time of the emission of the photon. Using Eq. (1.72) and defining  $a_0 \equiv 1$ , we find

$$a(z) = \frac{1}{1+z}. \quad (1.74)$$

We can now express the Hubble parameter, given in Eq. (1.21), as a function of redshift:

$$H^2(z) = H_0^2 \left[ \Omega_R (1+z)^4 + \Omega_M (1+z)^3 + \Omega_k (1+z)^2 + \Omega_\Lambda \right]. \quad (1.75)$$

Finally, assuming a cosmological model, we can use this expression to determine the comoving distance of the observed object, which is defined via

$$d_c(z) = \int_0^z \frac{1}{H(z')} dz', \quad (1.76)$$

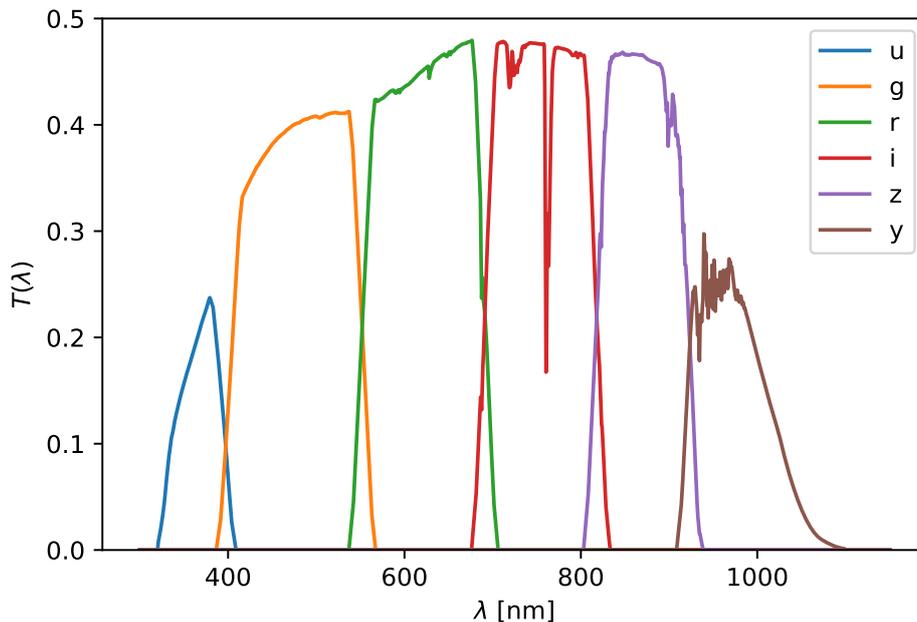
from which we infer the physical distance using Eq. (1.3).

The most accurate redshift measurement of a given object is inferred from spectroscopic observations. By measuring the flux as a function of wavelength through a spectrograph we obtain the spectral energy distribution (SED) of the object. We can then identify features such as the wavelength of emission and absorption lines of atoms and molecules in the observed SED. The rest-frame wavelengths of these features are well known thanks to laboratory experiments and therefore we can employ Eq. (1.72) to infer the redshift. This technique, commonly referred to as spectroscopic redshift, provides redshift measurements with high precision. However, the measurement of the full spectrum of an object with a high signal-to-noise ratio requires a significant amount of telescope time. Therefore, this technique limits the number of objects that can be observed in

a survey. Additionally it limits the depth in magnitude up to which objects can be observed, which is determined by the achievable signal-to-noise ratio. This makes this method infeasible for surveys that prioritise a large number of observed objects.

Instead of spectroscopy, many modern surveys rely on photometry to infer redshifts. As an alternative for measuring the full SED, the spectrum is observed through a set of filters. Each filter has a well-determined transmission function  $T^b(\lambda)$  describing the fraction of the photon flux that passes through filter  $b$  at a given wavelength  $\lambda$ . Fig. 1.4 shows the transmission curves of the six filters (u, g, r, i, z, and y) of the Vera C. Rubin Observatory's Legacy Survey of Space and Time. Given the relatively broad filters, spanning wavelength ranges of  $\Delta\lambda \approx 100 - 200$  nm, this technique provides low-resolution measurements of the spectrum. Thus, it is impossible to identify narrow features like individual emission and absorption lines. However, broad features such as the Balmer break can be identified in photometric observations. This feature is caused by the absorption of photons in star-forming galaxies by ionised hydrogen atoms. Photons with a wavelength below the Balmer limit with rest-frame wavelength  $\lambda_B = 346.6$  nm are absorbed, leading to a suppression of the SED below the Balmer limit. A similar feature can be found at the Lyman limit with rest-frame wavelength  $\lambda_L = 91.2$  nm. These features cause so-called dropouts which are sources that can be observed in one filter but not in the neighbouring one. Depending on the filter in which the dropout is found we can locate the position of the Balmer or Lyman break and hence we can estimate the redshift of the galaxy. Additionally, the broad wavelength range observed in each exposure of the camera leads to a higher signal-to-noise ratio. Thus, the telescope time required to observe a single object is greatly reduced compared to spectroscopic observations. Additionally, the higher signal-to-noise ratio gives access to fainter objects and therefore increases the depth of the survey.

To estimate redshifts from photometry, commonly referred to as photometric redshifts, we can use template-fitting methods. This technique, first



**Figure 1.4:** Transmission functions of the six filters used by the Vera C. Rubin Observatory's Legacy Survey of Space and Time. Source: <https://github.com/lsst/throughputs>

introduced by Baum (1962) and further developed by Puschell et al. (1982), Koo (1985), Loh and Spillar (1986), and Connolly et al. (1995), has become widely used in surveys that require redshifts for a large number of faint sources. The photometric redshift is estimated using a library of theoretical SED templates that describe the normalised flux  $F_t(\lambda, z)$  for a galaxy of type  $t$  shifted to redshift  $z$ . For each template we infer the flux observed through each filter band  $b$  via

$$F_t^b(z) = \int_0^\infty F_t(\lambda, z) T^b(\lambda) d\lambda. \quad (1.77)$$

By maximising the likelihood  $\mathcal{L}_t$  defined by

$$-\log \mathcal{L} \propto \chi_t^2(z) = \sum_t \left( \frac{F_{\text{obs}} - c F_t^b(z)}{\sigma^b} \right)^2, \quad (1.78)$$

where  $\sigma^b$  is the uncertainty on the flux measurement in filter  $b$  and  $c$  is a normalisation constant, we infer the best-fitting redshift and galaxy type.

However, degeneracies between the templates observed through a limited set of filters lead to catastrophic outliers, where the photometric redshift estimate differs greatly from the true redshift of the galaxy.

The maximum likelihood method can be improved using a Bayesian framework, introduced by [Benítez \(2000\)](#) and implemented in the popular code BPZ. Using Bayes's theorem, the probability of a galaxy with colour  $C$  and magnitude  $m$  having redshift  $z$  is given by

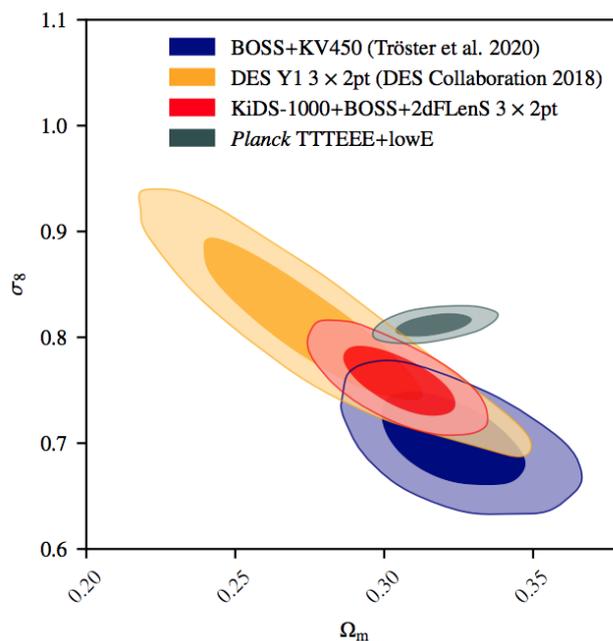
$$p(z, t|C, m) \propto p(z, t|m)p(C|z, t, m), \quad (1.79)$$

where  $p(C|z, t, m)$  is equivalent to the likelihood  $\mathcal{L}$  and  $p(z, t|m)$  denotes the prior probability of a galaxy with magnitude  $m$  having redshift  $z$  and type  $t$ . Thus, this method allows incorporating prior information on the observed galaxies in a survey into the photometric redshift estimation, reducing the amount of catastrophic outliers. Nevertheless, photometric redshift estimates are in general less reliable than their spectroscopic counterparts. However, the ability to observe large numbers of faint objects makes photometric observations attractive for surveys that prioritise the number of observed sources over precise redshift estimates, such as weak lensing studies.

## 1.4 The current status of cosmology

Recent developments in cosmology over the past few decades have led to rapid progress in our understanding of the Universe. Some of the most influential probes are measurements of temperature anisotropies in the CMB, most recently by Planck ([Planck Collaboration et al., 2020b](#)), which are in excellent agreement with a flat  $\Lambda$ CDM cosmological model of a Universe consisting of approximately 5% baryons, 26% cold dark matter, and 69% dark energy. Moreover, large spectroscopic and photometric galaxy surveys, such as SDSS, DES and KiDS, have derived constraints on cosmological parameters with unprecedented precision.

However, recent analyses revealed a tension between CMB observations



**Figure 1.5:** Marginalised posterior distribution of  $\sigma_8$  and  $\Omega_m$ , measured by Tröster et al. (2021); Planck Collaboration et al. (2020b); Heymans et al. (2021), and Abbott et al. (2018). Figure adapted from Heymans et al. (2021).

from Planck and probes of the large-scale structure of the Universe, for example the so-called  $3 \times 2$ pt analysis by Heymans et al. (2021), which combines three cosmological analyses with two-point correlation functions: weak lensing from KiDS, galaxy clustering from the Baryon Oscillation Spectroscopic Survey (BOSS), and galaxy-galaxy lensing from KiDS, BOSS, and the 2-degree Field Lensing Survey (2dFLenS). This tension manifests itself in the structure growth parameter  $S_8 = \sigma_8 \sqrt{\Omega_m/0.3}$  which was measured to be  $S_8 = 0.766^{+0.020}_{-0.014}$ . This value is  $8.3 \pm 2.6\%$  lower than the value measured by Planck. The tension is illustrated in Fig. 1.5 showing the constraints on  $\sigma_8$  and  $\Omega_m$  from Heymans et al. (2021) in red compared to Planck in grey and earlier analyses in blue and yellow. It is currently at a level between 2 and 3  $\sigma$ , which raises the question of whether or not this is a statistical fluctuation or if our current cosmological model is insufficient to correctly describe the evolution of the Universe between early times (probed by the CMB) and late times (probed by galaxy surveys). This remains a field of active research in cosmology.

In addition to the apparent tension between the CMB and galaxy redshift surveys, direct model-independent measurements of the current expansion rate  $H_0$  by [Riess et al. \(2021\)](#) are in  $4.2 \sigma$  tension with the prediction from Planck under the assumption of a  $\Lambda$ CDM cosmological model. Direct measurements of  $H_0$  are obtained from the so-called cosmic distance ladder, which is a multi-step process to determine the distance to extragalactic objects. It relies on observations of Cepheids and type Ia supernovae, which are objects of known intrinsic luminosity, so called standard candles. The first rung of the cosmic distance ladder consists of distance measurements of Cepheids in the Milky Way via the parallax method. These Cepheids act as a calibration sample, for which we know both the distance and the apparent magnitude. Using the relation between the period of variability and the absolute magnitude of Cepheids ([Leavitt and Pickering, 1912](#)), the distance of extragalactic Cepheids can then be obtained from observations of their apparent magnitude and period. Thus, observations of Cepheids in the host galaxies of type Ia supernovae allow for the calibration of the type Ia supernova distances. Since supernovae are visible from great distances, this allows for distance measurements of far galaxies from which the Hubble constant is obtained. This study finds a value of  $H_0 = 73.2 \pm 1.33 \text{kms}^{-1} \text{Mpc}^{-1}$  which is in disagreement with the value inferred by Planck of  $H_0 = 67.37 \pm 0.54 \text{kms}^{-1} \text{Mpc}^{-1}$ . The tension is even more stringent than the aforementioned  $S_8$ -tension and raises the question if we need to consider an extended cosmological model in order to explain the discrepancy between probes of the early and late Universe. In particular, since  $H_0$  measurements via the cosmic distance ladder are model-independent, we require a cosmological model that accelerates the expansion of the Universe at later times in order to make the two measurements compatible. However, other possible explanations are systematic effects in the calibration various quantities in the cosmic distance ladder that bias the inferred  $H_0$  value or statistical fluctuations, although these are less likely given the current level of the tension.

In the next few years, multiple new instruments will come online which will

allow us to constrain cosmology with unprecedented precision. In particular, the Legacy Survey of Space and Time (LSST, [Ivezić et al., 2019](#)) will be conducted at the newly built Vera C. Rubin Observatory located on the peak of Cerro Pachón in southern Chile. This survey will observe billions of objects in six optical filter bands, shown in Fig.1.4, over the course of 10 years. With a 8.4m primary mirror and a field of view of 3.5 deg in diameter it will be able to observe an area of about  $18000 \text{ deg}^2$  of the southern sky every three days. Thus, it will enable various scientific studies, most importantly the study of dark matter and dark energy through measurements of weak lensing, galaxy clustering, and type Ia supernovae.

In addition to the ground-based LSST, space-based telescopes such as ESA's Euclid satellite ([Laureijs et al., 2011](#)) and NASA's Nancy Grace Roman Space Telescope ([Spergel et al., 2015](#)) will be set to launch within the next few years. These instruments will provide observations that are highly complementary to LSST ([Rhodes et al., 2017](#); [Eifler et al., 2021](#)), since they are not affected by atmospheric effects. This allows for higher-resolution images. On the downside, given the limited size of space-based telescopes, they will not be able to observe objects as faint as the ones observed by LSST and therefore will detect fewer objects. With these upcoming instruments we will be able to constrain cosmological parameters with an even higher precision than before and gain insight into the nature of the aforementioned tensions found in current surveys.

## Chapter 2

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

The calibration of the redshift of galaxies via photometric methods is a major source of uncertainty in many cosmological analyses, as discussed in section 1.3. Thus, the propagation of redshift distribution uncertainties in a consistent way into the cosmological analysis plays a crucial role in cosmic shear studies. In this chapter, we reproduce the work on this subject published in [Stölzner et al. \(2021\)](#).

**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 the third data release of the Kilo-Degree Survey combined with the VISTA Kilo-Degree Infrared Galaxy Survey (KV450). 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 of the medians of the five tomographic redshift distributions of up to  $\Delta z \approx 0.02$ , which are however degenerate with an observed decrease of the amplitude of intrinsic galaxy alignments by about 10%.

## 2.1 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 ( $\Lambda$ CDM) 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 subsamples 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 and 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 and 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 and 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 and 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 deg<sup>2</sup>.

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 subsamples. 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. 2.2 the redshift distribution model is described. The theoretical modelling of the cosmic shear signal with analytic marginalisation over nuisance parameters is presented in Sect. 2.3, and the cosmic shear self-calibration method of the redshift distributions is described in Sect. 2.4. Results are presented in Sect. 2.5 and discussed in Sect. 2.6.

## 2.2 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  $\sigma_{\text{comb}}^2$ :

$$n^{(\alpha)}(z) := \sum_{i=1}^{N_z} A_i^\alpha \mathcal{K}(z; z_i, \sigma_{\text{comb}}^2), \quad (2.1)$$

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

$$\mathcal{K}(z; z_i, \sigma_{\text{comb}}^2) = \frac{z}{N(z_i, \sigma_{\text{comb}})} \exp\left\{-\frac{(z - z_i)^2}{2\sigma_{\text{comb}}^2}\right\}, \quad (2.2)$$

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

$$N(z_i, \sigma) = \sqrt{\frac{\pi}{2}} z_i \sigma \operatorname{erfc}\left(-\frac{z_i}{\sqrt{2}\sigma}\right) + \sigma^2 \exp\left\{-\frac{z_i^2}{2\sigma^2}\right\}. \quad (2.3)$$

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

$$\sum_{i=1}^{N_z} A_i^\alpha = 1. \quad (2.4)$$

Using Eq. (2.4), we write the amplitude of the  $N_z$ -th component in terms of the remaining  $N_z - 1$  amplitudes:

$$A_N^\alpha = 1 - \sum_{i=1}^{N_z-1} A_i^\alpha. \quad (2.5)$$

Inserting Eq. (2.5) back into Eq. (2.1), we find

$$n^{(\alpha)}(z) = \mathcal{K}(z; z_{N_z}, \sigma_{\text{comb}}^2) \quad (2.6)$$

$$\begin{aligned} &+ \sum_{i=1}^{N_z-1} A_i^\alpha \left[ \mathcal{K}(z; z_i, \sigma_{\text{comb}}^2) - \mathcal{K}(z; z_{N_z}, \sigma_{\text{comb}}^2) \right] \\ &:= \sum_{i=1}^{N_z} A_i^\alpha n_i(z), \end{aligned} \quad (2.7)$$

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

$$n_i(z) = \mathcal{K}(z; z_i, \sigma_{\text{comb}}^2) - \mathcal{K}(z; z_{N_z}, \sigma_{\text{comb}}^2) \cdot (1 - \delta_{iN_z}). \quad (2.8)$$

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

$$a_i^\alpha := \ln A_i^\alpha \quad (2.9)$$

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

$$\Sigma \quad (2.10)$$

$$\Pr(\{a_i^\alpha\} | \mathbf{d}_{\text{cal}}) \approx \mathcal{N}(a_i^\alpha; a_i^{*\alpha}, \Sigma_{\text{cal}}), \quad (2.11)$$

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

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, 2006](#)). 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, 2006](#)).

## 2.3 Theoretical modelling of the cosmic shear signal

### 2.3.1 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 and 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

$$\xi_{\pm}^{(\alpha\beta)}(\theta) = \frac{1}{2\pi} \int_0^\infty d\ell \ell C_{\text{GG}}^{(\alpha\beta)}(\ell) J_{0,4}(\ell\theta), \quad (2.12)$$

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

$$C_{\text{GG}}^{(\alpha\beta)}(\ell) = \int_0^{\chi_{\text{H}}} d\chi \frac{q^{(\alpha)}(\chi)q^{(\beta)}(\chi)}{f_K^2(\chi)} P_\delta\left(\frac{\ell+1/2}{f_K(\chi)}, \chi\right), \quad (2.13)$$

where  $P_\delta$  is the matter power spectrum and  $f_K$ ,  $\chi$ , and  $\chi_{\text{H}}$  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

$$q^{(\alpha)}(\chi) = \frac{3H_0^2\Omega_m}{2c^2} \frac{f_K(\chi)}{a(\chi)} \int_{\chi}^{\chi_H} d\chi' n_{\chi'}^{(\alpha)}(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')}, \quad (2.14)$$

with  $a(\chi)$  being the scale factor and  $n^\alpha(z)dz = n_{\chi}^{(\alpha)}(\chi)d\chi$  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

$$C_{\text{GG}}^{(\alpha\beta)}(\ell) = \sum_{i,j=1}^{N_z} A_i^\alpha A_j^\beta \int_0^{\chi_H} d\chi \frac{q_i(\chi)q_j(\chi)}{f_K^2(\chi)} P_\delta\left(\frac{\ell+1/2}{f_K(\chi)}, \chi\right) \quad (2.15)$$

$$:= \sum_{i,j=1}^{N_z} A_i^\alpha A_j^\beta c'_{ij}(\ell), \quad (2.16)$$

where we defined  $q_i(\chi)$  as the lensing efficiency of the  $i$ -th component of the redshift distribution model, as defined in Eq. (2.6). 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. (2.12), we can compute the two-point shear correlation function between two tomographic redshift bins via

$$\xi_{\text{GG}}^{(\alpha\beta)}(\theta) = \sum_{i,j=1}^{N_z} A_i^\alpha A_j^\beta \int_0^\infty \frac{d\ell}{2\pi} J_{0/4}(\ell\theta) c'_{ij}(\ell) \quad (2.17)$$

$$:= \sum_{i,j=1}^{N_z} A_i^\alpha A_j^\beta x_{\pm}^{(ij)}(\theta), \quad (2.18)$$

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 and Seljak, 2004):

$$\xi_{\pm} = \xi_{\text{GG}} + \xi_{\text{II}} + \xi_{\text{GI}}. \quad (2.19)$$

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 and Seljak, 2004; Bridle and King, 2007; Joachimi et al., 2011). The contributions of GI and II alignments to the two-point shear correlation function were calculated using Eq. (2.12) with the II and GI angular power spectra:

$$C_{\text{II}}^{(\alpha\beta)} = \int d\chi \frac{n^{(\alpha)}(\chi)n^{(\beta)}(\chi)}{f_K^2(\chi)} P_{\text{II}} \left( \frac{\ell+1/2}{f_K(\chi)}, \chi \right), \quad (2.20)$$

$$C_{\text{GI}}^{(\alpha\beta)} = \int d\chi \frac{q^{(\alpha)}(\chi)n^{(\beta)}(\chi) + n^{(\alpha)}(\chi)q^{(\beta)}(\chi)}{f_K^2(\chi)} P_{\text{GI}} \left( \frac{\ell+1/2}{f_K(\chi)}, \chi \right). \quad (2.21)$$

Again, we used the linear dependence on the redshift distribution to extract the amplitudes of the redshift distribution model in analogy to Eq. (2.18). The power spectra of intrinsic galaxy alignments,  $P_{\text{II}}$  and  $P_{\text{GI}}$ , are related to the matter power spectrum  $P_{\delta}$  via

$$P_{\text{II}}(k, z) = F^2(z)P_{\delta}(k, z) \quad (2.22)$$

$$P_{\text{GI}}(k, z) = F(z)P_{\delta}(k, z), \quad (2.23)$$

with

$$F(z) = -A_{\text{IA}}C_1\rho_{\text{crit}}\frac{\Omega_{\text{m}}}{D_+(z)}. \quad (2.24)$$

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

$$\mathcal{L} = -\frac{1}{2}\chi^2 + \text{const.} = -\frac{1}{2} \sum_{ij} (d_i - m_i) C_{ij}^{-1} (d_j - m_j) + \text{const.}, \quad (2.25)$$

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

$$\mathcal{L} = -\frac{1}{2} \sum_{l,\alpha,\beta,l',\alpha',\beta'} \Delta_l^{(\alpha\beta)} Z_{(l,\alpha,\beta)(l',\alpha',\beta')} \Delta_{l'}^{(\alpha'\beta')} + \text{const.}, \quad (2.26)$$

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  $\mathbf{Z}$  with elements  $Z_{(l,\alpha,\beta)(l',\alpha',\beta')}$ . We have defined

$$\Delta_l^{(\alpha\beta)} \equiv d_l^{(\alpha\beta)} - \sum_{i,j=1}^{N_z} A_i^\alpha A_j^\beta x_\pm^{(ij)}(\theta_l), \quad (2.27)$$

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  $x_\pm^{(ij)}(\theta)$ .

### 2.3.2 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  $\mathbf{p}_{\text{sam}}$  and parameters that we analytically marginalise over by  $\mathbf{p}_{\text{ana}}$ . In particular,  $\mathbf{p}_{\text{sam}}$  includes cosmological parameter as well as nuisance parameters that account for intrinsic alignments, baryon feedback, and additive shear bias. The parameters  $\mathbf{p}_{\text{ana}}$  are the collection of

amplitude parameters  $\{a_i^\alpha\}$ . We obtain

$$\Pr(\mathbf{d}|\mathbf{p}_{\text{sam}}) = \int d^{N_{\text{ana}}} p_{\text{ana}} \Pr(\mathbf{d}|\mathbf{p}_{\text{sam}}, \mathbf{p}_{\text{ana}}) \Pr(\mathbf{p}_{\text{ana}}), \quad (2.28)$$

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. (2.11).

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 and 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,

$$\mathcal{L}_{\text{marg}} \equiv -2 \ln \Pr(\mathbf{d}|\mathbf{p}_{\text{sam}}), \quad (2.29)$$

is then given by [Bridle et al. \(2002\)](#) and [Taylor and Kitching \(2010\)](#):

$$\mathcal{L}_{\text{marg}} = \mathcal{L}_{\text{fid}} - \frac{1}{2} \mathcal{L}'^\tau \left[ \mathcal{L}'' + 2\Sigma_{\text{cal}}^{-1} \right]^{-1} \mathcal{L}' + \ln \det \left( \mathbb{I} + \frac{1}{2} \mathcal{L}'' \Sigma_{\text{cal}} \right), \quad (2.30)$$

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

$$\mathcal{L}_{\text{fid}} \equiv -2 \ln \Pr(\mathbf{d}|\mathbf{p}_{\text{sam}}, \mathbf{p}_{\text{ana}}^*). \quad (2.31)$$

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.1. All of these derivatives are to be evaluated at the best fit of the nuisance parameters  $\mathbf{p}_{\text{ana}}^*$ . The covariance matrix of nuisance parameters, originating from the calibration of the photometric redshift distribution, is given by  $\Sigma_{\text{cal}}$ . For the  $N_{\text{bin}}$  tomographic bins used in

the analysis,  $N_{\text{bin}} \times N_z$  nuisance parameters are marginalised over (modulo those amplitudes fixed by the normalisation of the redshift distribution given in Eq. (2.4)).

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 and 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.

## 2.4 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. 2.3.2. 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 dimensionality of the nuisance parameter space, we adopted an iterative method, which is illustrated in Fig. 2.1:

First, we fitted the Gaussian comb model, defined in Sect. 2.2, to pre-calibrated redshift distribution histograms. This was done by minimising

$$\chi^2 = \sum_{ij} \left( n_i^{\text{data}} - n_i^{\text{model}} \right) C_{ij}^{-1} \left( n_j^{\text{data}} - n_j^{\text{model}} \right), \quad (2.32)$$

where  $n_i^{\text{data}}$  and  $n_i^{\text{model}}$  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  $\Sigma_{\text{cal}}$  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  $\mathbf{p}_{\text{sam}}$  of the standard weak lensing likelihood conditioned on the best-fit parameters  $\mathbf{p}_{\text{ana}}$ . This step is illustrated by the blue arrows in Fig. 2.1.

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

$$\delta\mathbf{p}_{\text{ana}} = -\mathcal{L}' \left[ \mathcal{L}'' + 2\Sigma_{\text{cal}}^{-1} \right]^{-1}, \quad (2.33)$$

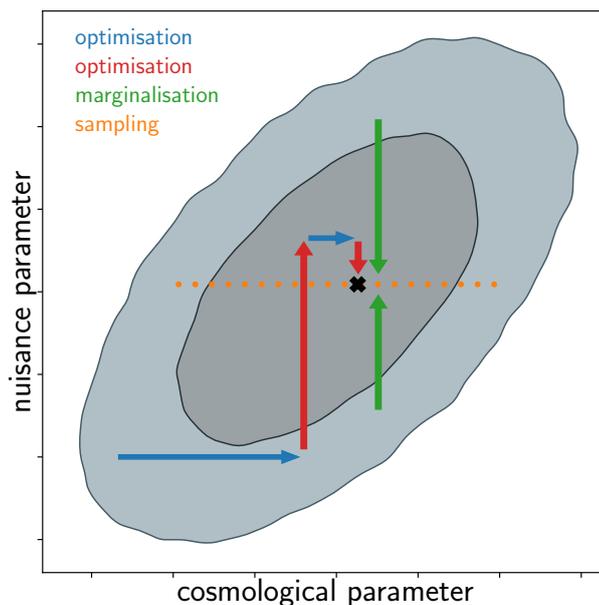
assuming a Gaussian prior on the parameters  $\mathbf{p}_{\text{ana}}$ . Fixing the parameters  $\mathbf{p}_{\text{sam}}$  to the ones found in step 2, we used Newton's method to minimise Eq. (2.33) so that the parameters  $\mathbf{p}_{\text{ana}}$  converge towards the peak of the likelihood. Since 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.<sup>1</sup> The red arrows in Fig. 2.1 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  $\mathbf{p}_{\text{ana}}^*$  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. 2.1.

While it is advantageous to infer the initial values for  $\mathbf{p}_{\text{ana}}$  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

---

<sup>1</sup>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 and Kitching, 2010).



**Figure 2.1:** 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).

minimum of the posterior distribution.

## 2.5 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  $450 \text{ deg}^2$ . 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 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. 2.2 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<sup>2</sup> package ([Audren et al., 2013](#); [Brinckmann and Lesgourgues, 2019](#)). We sampled the likelihood in the MULTINEST<sup>3</sup> mode ([Feroz et al., 2009, 2019](#)) using the python wrapper PYMULTINEST<sup>4</sup> ([Buchner et al., 2014](#)). The matter power spectrum is estimated with the public code CLASS<sup>5</sup> ([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:  $\omega_{\text{CDM}}$ ,  $\omega_{\text{b}}$ ,  $A_{\text{s}}$ ,  $n_{\text{s}}$ , and  $h$ . Additionally, the model includes four nuisance parameters that account for intrinsic alignments ( $A_{\text{IA}}$ ), baryon feedback ( $A_{\text{bary}}$ ), and additive shear bias ( $\delta c$  and  $A_{\text{c}}$ ). 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 2.1. Finally, we adopted the cosmic shear data from [Hildebrandt et al. \(2020\)](#), which consists

---

<sup>2</sup>[https://github.com/brinckmann/montepython\\_public](https://github.com/brinckmann/montepython_public)

<sup>3</sup><https://github.com/farhanferoz/MultiNest>

<sup>4</sup><https://github.com/JohannesBuchner/PyMultiNest>

<sup>5</sup>[https://github.com/lesgourg/class\\_public](https://github.com/lesgourg/class_public)

Parameter	Symbol	Prior
CDM density	$\omega_{\text{CDM}}$	[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	$A_{\text{IA}}$	[-6, 6]
Baryon feedback amplitude	$A_{\text{bary}}$	[2.00, 3.13]
Constant $c$ -term offset	$\delta c$	$0.0000 \pm 0.0002$
2D $c$ -term amplitude	$A_c$	$1.01 \pm 0.13$

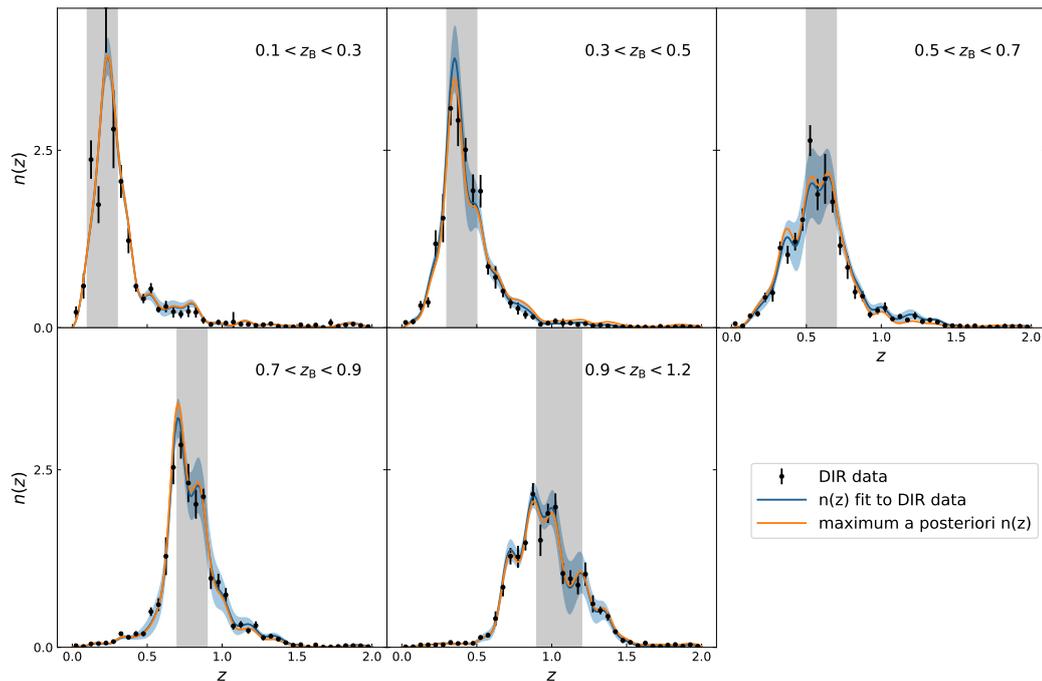
**Table 2.1:** Model parameters and their priors for the KV450 cosmic shear analysis, adopted from [Hildebrandt et al. \(2020\)](#). The first five rows are cosmological parameters, and the remaining rows represent nuisance parameters. Brackets indicate top-hat priors, and values with errors indicate Gaussian priors. We note that, in contrast to [Hildebrandt et al. \(2020\)](#), linear shifts in the mean of the redshift distributions are excluded.

of measurements of the two-point shear correlation functions between the five tomographic redshift bins and the corresponding analytic covariance matrix.

### 2.5.1 Redshift distribution self-calibration

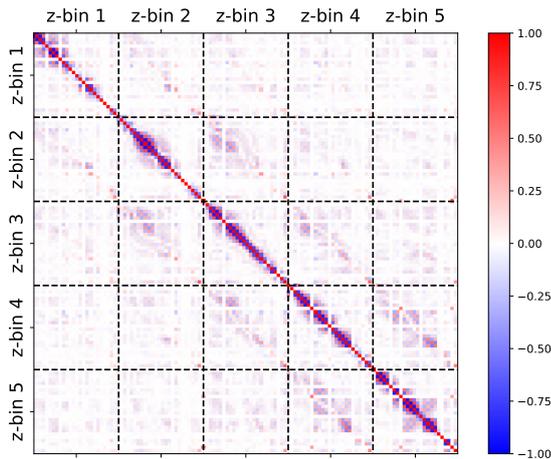
The first step in the calibration of the KV450 redshift distribution was to fit the modified Gaussian mixture model, defined in Sect. 2.2, to the redshift histograms of [Hildebrandt et al. \(2020\)](#), which are pre-calibrated with deep spectroscopic samples. The fit was done simultaneously for all 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 A.2 we perform several tests, comparing different choices for the



**Figure 2.2:** 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. 2.3. Orange curves represent the redshift distributions after iterative optimisation of cosmological and nuisance parameters.

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,  $\sigma_{\text{comb}}$ , 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  $\sigma_{\text{comb}} = 0.067$ , 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. 2.2; 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



**Figure 2.3:** Correlation matrix of best-fit comb amplitudes with 30 components per redshift bin.

	$A_{\text{IA}}$	$S_8$	$\chi^2$
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

**Table 2.2:** Results of the iterative fitting of cosmological and nuisance parameters to the KV450 cosmic shear data and comparison to the fiducial KV450 likelihood. When optimising cosmological parameters, we fit  $\omega_{\text{cdm}}$ ,  $\omega_{\text{b}}$ ,  $A_{\text{s}}$ ,  $n_{\text{s}}$ , and  $h$  as well as the nuisance parameters  $A_{\text{IA}}$ ,  $A_{\text{bary}}$ ,  $\delta c$ , and  $A_{\text{c}}$ . When optimising nuisance parameters, we vary the amplitudes of the Gaussian comb. Results are shown for the two most interesting parameters,  $A_{\text{IA}}$  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.

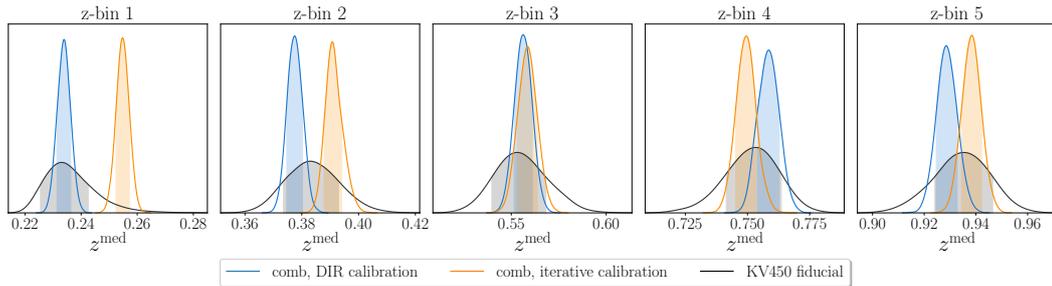
shown in Fig. 2.3.

We proceeded with a further calibration of the redshift distribution using the iterative fitting method of cosmological and nuisance parameters described in Sect. 2.4. The fit result after each step for the two parameters that are mostly constrained by the data, the intrinsic alignment amplitude  $A_{\text{IA}}$ , and the amplitude of matter density fluctuations,  $S_8 = \sigma_8(\Omega_{\text{m}}/0.3)^{0.5}$ , are reported in Table 2.2. 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 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  $N_{\text{dof}} = N_{\text{d}} - N_{\Theta}$ , where  $N_{\text{d}}$  is the size of the data vector 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  $N_{\text{dof}} = 186$ . 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 and Lemos, 2019b](#); [Raveri and Hu, 2019](#); [Joachimi et al., 2021](#)), we refrain from a further interpretation of the goodness of fit.



**Figure 2.4:** 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.

We find a shift in the two most interesting parameters for which the cosmic shear likelihood has the largest constraining power,  $A_{1A}$  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. 2.2.

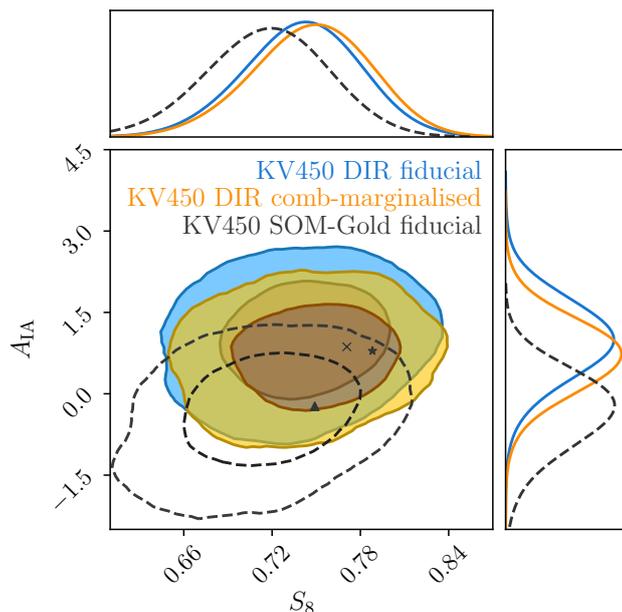
Figure 2.4 shows comparisons of the median of the redshift distribution of each tomographic bin inferred from the original DIR histograms and the Gaussian comb before and after iterative calibration. We chose the median as our summary statistic since the mean of the DIR histograms is less stable with 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. The distribution of the median redshift of the fiducial KV450 redshift histograms is inferred by resampling the redshift histograms from a multivariate Gaussian distribution with the DIR covariance matrix that describes correlations between histogram bins of all tomographic bins. Similarly, the distribution of the median redshift of the Gaussian comb

model is obtained by resampling the comb amplitudes using the covariance matrix of comb amplitudes. 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. However, we would expect that both the redshift histograms and the comb model show similar constraints on the median redshift. The fact that the constraints from the comb model are tighter than the ones from the redshift histograms can be interpreted as a sign that the model is not flexible enough to accurately describe the full shape of the redshift distribution. This can potentially be alleviated by further optimising the parameters that characterise the Gaussian comb model. This is left for future work. 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 intrinsic alignments and the redshift distributions, which is discussed in the following sections.

### 2.5.2 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 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 A.3, show that the two models are in good agreement.

Figure 2.5 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,  $A_{IA}$  and  $S_8$ . The posterior distribution of all remaining parameters is shown in Appendix A.4. We find a slight shift in the posterior towards smaller values of



**Figure 2.5:** Marginalised posteriors for  $A_{IA}$  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 2.2 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.

the intrinsic alignment amplitude and larger values of  $S_8$ . Additionally, Fig. 2.5 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,  $A_{IA}$ , is compatible with  $A_{IA} = 0$ , whereas Hildebrandt et al. (2020) found  $A_{IA} \approx 1$ . 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 2.2). 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  $A_{\text{IA}} = 0.1_{-0.1}^{+0.1}$ . 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.

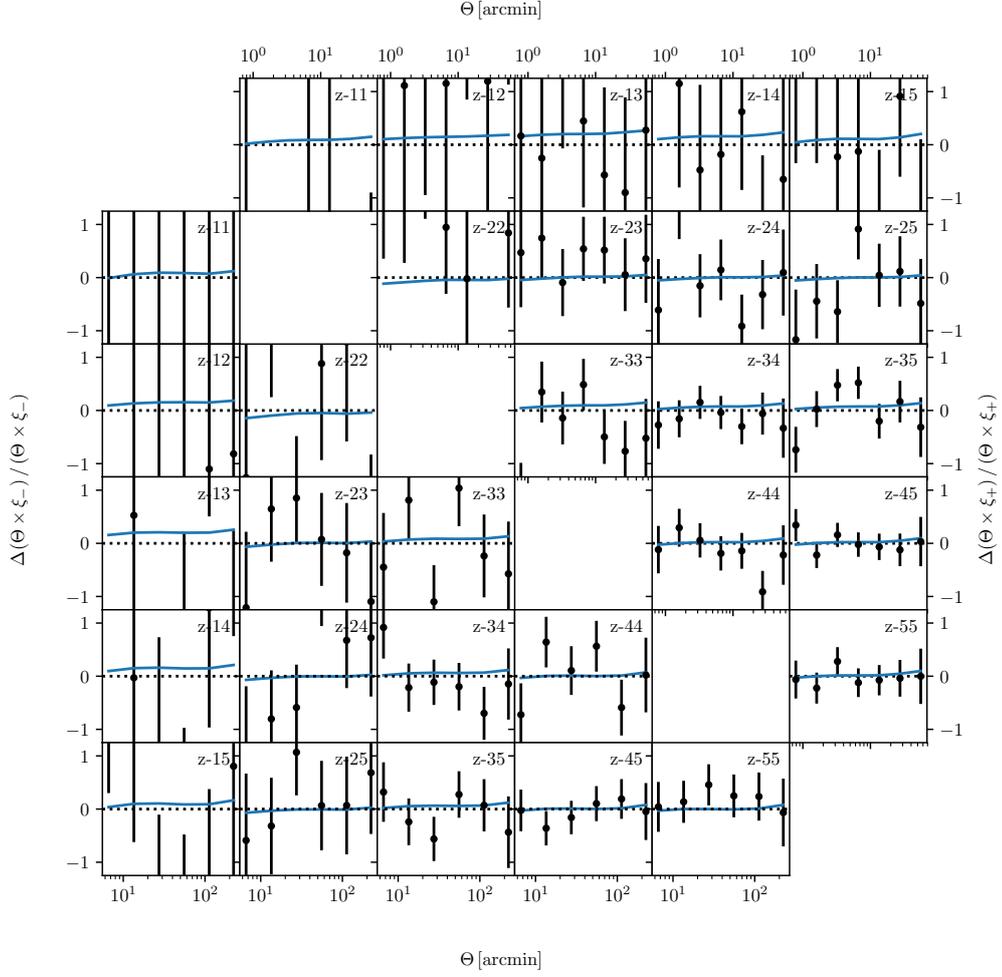
The intrinsic alignment amplitude,  $A_{\text{IA}}$ , is not a cosmological parameter, but instead originates from the modelling of correlations between intrinsic ellipticities of neighbouring galaxies, II, and correlations between intrinsic ellipticities of foreground galaxies and background galaxies, GI. As can be inferred from Eqs. (2.22) and (2.23), the GI term gives a negative contribution to  $\xi_{\pm}$  that is proportional to  $A_{\text{IA}}$ , whereas the II term contributes positively, proportionally to  $A_{\text{IA}}^2$ . 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. 2.5, which nevertheless are still in good agreement.

Furthermore, since the signal-to-noise ratio 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. 2.4. This is illustrated in Fig. 2.6, 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. 2.4.

The black data points in Fig. 2.6 show the relative difference between the observed two-point correlation function and the fiducial best fit. These indicate that the signal-to-noise ratio 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. 2.4 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 and 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



**Figure 2.6:** 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.

between intrinsic alignments and redshift distributions by providing a more accurate redshift distribution calibration.

## 2.6 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 pre-calibrated 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. However, testing the suspected degeneracy between the redshift distribution amplitudes and the amplitude of intrinsic alignments requires a sampling of the likelihood with both redshift distribution amplitudes and the intrinsic alignment amplitude as free parameters. Given the large number of redshift distribution amplitudes such a sampling is computationally unfeasible. 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 and 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 ‘6x2pt’ analyses, which combine all two-point correlation functions between overlapping imaging and spectroscopic surveys.

## Chapter 3

# Optimising the shape of photometric redshift distributions with clustering cross-correlations

Cosmological analyses, such as studies of cosmic shear or galaxy clustering, are usually performed tomographically, which requires the assignment of galaxies to redshift bins. This is usually done using estimates of the photometric redshift of galaxies and thus the inferred redshift bins are subject to catastrophic outliers in the redshift estimation, as discussed in section 1.3. In this chapter, we reproduce the work presented in [Stölzner et al. \(2023\)](#), in which an optimisation method for the assignment of galaxies to redshift bins is developed.

**Abstract:** We present an optimisation method for the assignment of photometric galaxies into a chosen set of redshift bins. This is achieved by combining simulated annealing, an optimisation algorithm inspired by solid-state physics, with an unsupervised machine learning method, a self-organising map (SOM) of the observed colours of galaxies. Starting with a sample of galaxies that is divided into redshift bins based on a photometric redshift point estimate, the simulated annealing algorithm repeatedly reassigns SOM-selected subsamples of galaxies, which are close in colour, to alternative redshift bins. We

optimise the clustering cross-correlation signal between photometric galaxies and a reference sample of galaxies with well-calibrated redshifts. Depending on the effect on the clustering signal, the reassignment is either accepted or rejected. By dynamically increasing the resolution of the SOM, the algorithm eventually converges to a solution that minimises the number of mismatched galaxies in each tomographic redshift bin and thus improves the compactness of their corresponding redshift distribution. This method is demonstrated on the synthetic LSST cosmoDC2 catalogue. We find a significant decrease in the fraction of catastrophic outliers in the redshift distribution in all tomographic bins, most notably in the highest redshift bin with a decrease in the outlier fraction from 57 per cent to 16 per cent.

### 3.1 Introduction

The calibration of the redshift distribution of cosmological surveys plays a crucial role in current studies of cosmology. While spectroscopic observations of galaxies allow for accurate redshift measurements of the source redshift distribution, complete spectroscopic measurements are often infeasible given the large number of observed objects in current and upcoming surveys, such as the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST; [Ivezić et al., 2019](#)) and the European Space Agency’s Euclid survey ([Laureijs et al., 2011](#)). Therefore, surveys often rely on multi-band photometry to determine the redshift of observed objects (see [Salvato et al., 2019](#), for a review). However, photometric methods suffer from systematic biases and catastrophic outliers in the redshift estimation and thus require a sophisticated calibration of the redshift distribution in order to derive robust constraints on cosmology (see for example: [Ma et al., 2006](#); [Huterer et al., 2006](#); [Bernstein and Huterer, 2010](#); [Cunha et al., 2014](#)).

Cosmological analyses, for example studies of weak gravitational lensing by the large-scale structure of the Universe, are often performed tomographically, which allows for the utilisation of information about the evolution of the

Universe. In tomographic cosmic shear analyses the galaxy sample is split into several redshift bins using photometric redshift estimates of individual galaxies. The cosmic shear signal is then estimated by measuring the cross-correlation between the shapes of galaxies in the tomographic bins, which improves constraints on cosmological parameters (Hu, 1999).

A tomographic analysis usually requires two steps. First, the sample of galaxies needs to be divided into redshift bins. This is usually done using galaxy photometry, which are used to estimate the redshift of individual galaxies in the survey, for example via spectral energy distribution (SED) template fitting codes. However, the true redshift distributions of tomographic bins extend beyond the bin boundaries because of noise, systematic biases and catastrophic outliers in the photometric redshift estimation. Therefore, the second step is the calibration of the actual redshift distribution of each tomographic bin which is important when deriving theoretical predictions for the observed weak lensing signal given the sensitivity of the observed signal to the tails of the redshift distribution (Ma et al., 2006). For example, such calibration methods include angular cross-correlation clustering measurements with overlapping spectroscopic reference samples (e.g. Newman, 2008; Matthews and Newman, 2010; Ménard et al., 2013; McQuinn and White, 2013; McLeod et al., 2017; van den Busch et al., 2020; Gatti et al., 2022) and direct calibration methods with spectroscopic subsamples that are, potentially after re-weighting, representative of the full sample (Lima et al., 2008; Bonnett et al., 2016; Hildebrandt et al., 2017). Furthermore, hierarchical Bayesian models that combine photometry measurements of individual galaxies and clustering measurements with tracer populations in a robust way have been used for redshift calibration (Sánchez and Bernstein, 2019; Alarcon et al., 2020). Additionally, the clustering properties of photometric galaxies can be utilised to increase the precision of photometric redshifts (Jasche and Wandelt, 2012). Moreover, self-organising maps (SOMs) can be used to assign galaxies to groups based on their photometry (Masters et al., 2015; Buchs et al., 2019; Wright et al., 2020a; Myles et al., 2021),

which allows one to derive subsamples of galaxies that are fully represented by spectroscopic reference samples.

In this work we develop a calibration method that improves the first step by reducing the number of outliers in tomographic redshift bins. We develop a method that updates the assigned redshift bin of galaxies in a given photometric catalogue that otherwise would be assigned to an incorrect redshift bins if point estimates of the photometric redshift are used to assign galaxies to bins. The goal is to obtain a sample of galaxies that is divided into well-localised redshift bins. This is achieved by combining a self-organising map, which is used to group galaxies of a similar colour into cells, with measurements of clustering cross-correlations. We make use of point estimates of the photometric redshifts of galaxies to divide a galaxy catalogue into tomographic bins and apply a simulated annealing algorithm to reassign (SOM-based) cells of galaxies to alternative redshift bins. The optimisation algorithm utilises measurements of the clustering cross-correlation between the photometric sample and a reference sample with well-calibrated redshift measurements and maximises correlations between photometric and reference bins of the same redshift while minimising correlations between bins that are disjoint in redshift. We demonstrate the method on the synthetic LSST cosmoDC2 catalogue ([Korytov et al., 2019](#)).

The paper is structured as follows: The methods that we use are described in section 2. We show our results in section 3 and finally we discuss our main conclusions in section 4.

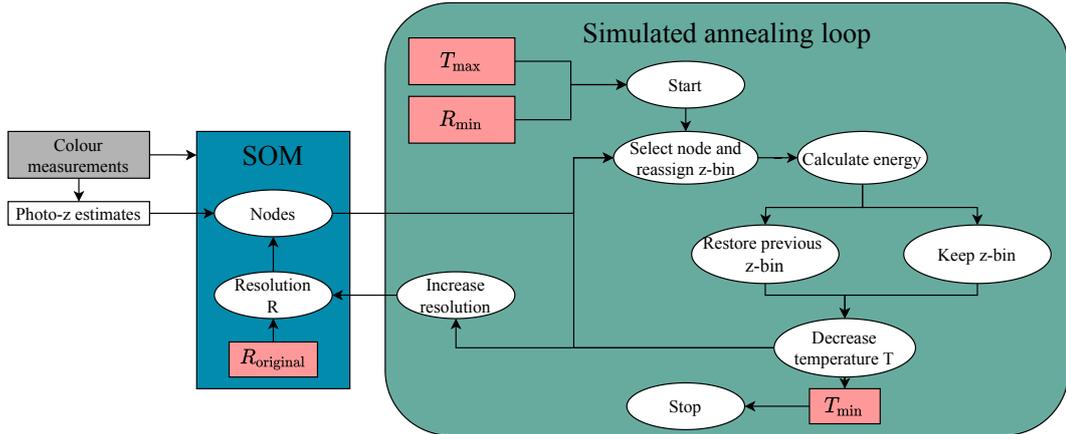
## 3.2 Methodology

In this section we summarise the optimisation method, called SHARPZ, that we use to assign photometric galaxies to tomographic redshift bins, which is illustrated in Fig. 3.1. We start from a catalogue of galaxies that are observed in several photometric bands from which point estimates of the photometric redshift of individual galaxies are estimated. These redshift estimates are used to divide the catalogue into tomographic bins. We aim to minimise mismatches

between the photometric redshift and the true redshift of the catalogue that are caused by imprecise redshift estimates in order to infer tomographic bins that are well-localised within the bin boundaries. We employ an overlapping sample of reference galaxies which is divided into the same tomographic bins using accurate redshift measurements in order to quantify how well the true redshift distribution of the photometric sample resembles the redshift distribution of the well-calibrated reference sample. To do so, we measure the angular cross-correlation between the photometric sample and the reference sample. This measurement relies on the property that galaxies cluster spatially, so that we expect a clustering signal between two overlapping photometric and reference samples, whereas samples that are separated in redshift are expected to show no clustering signal. Further details on the clustering measurements can be found in section 3.2.1. We employ a simulated annealing algorithm, explained in section 3.2.2, to randomly reassign photometric galaxies to a different redshift bin in order to maximise the correlation between overlapping photometric and reference bins while minimising the correlation of bins with no overlap in redshift. However, a reassignment of single galaxies only has a marginal impact on the correlation signal between different samples, since the cross-correlation is a statistical property of large samples of galaxies. Therefore, we additionally employ a self-organising map, which is described in section 3.2.3, to derive sets of galaxies of similar colour. By reassigning a set of galaxies in each step of the simulated annealing algorithm, we achieve a measurable effect on the clustering signal between photometric and reference samples, which allows us to employ the combined clustering signals as an objective function to be maximised by the algorithm.

### 3.2.1 Galaxy clustering

In order to determine how well localised within bin boundaries the true redshift distribution of the tomographic bins is, we employ an additional data set comprised of galaxies for which an accurate redshift measurement is available. This can be obtained for example, through spectroscopic observations of galaxies



**Figure 3.1:** Sketch of the optimisation algorithm that reassigns photometric galaxies to alternative redshift bins. We train a self-organising map (SOM) with a high resolution  $R_{\text{original}}$  on the observed colours of galaxies in the photometric sample, from which we infer SOMs with arbitrary resolutions  $R < R_{\text{original}}$ . Additionally, we infer point estimates of the photometric redshift to divide the sample into tomographic bins, so that each SOM node is assigned initially to the most common redshift bin of galaxies in this node. We initialise the simulated annealing algorithm with a starting temperature  $T_{\text{max}}$  and a resolution  $R_{\text{min}}$ , which is coupled to the temperature. In each iteration of the annealing algorithm we select a node of a SOM with the current resolution, which we randomly reassign to a different redshift bin. Measuring the angular cross-correlation between the photometric sample and the reference sample, we calculate the energy of the system from Eq. (3.3). Comparing the change in energy and the current temperature, we determine whether to keep the redshift bin assignment or to restore the previous state using Eq. (3.8). We then decrease the system’s temperature and, depending on the temperature, we either keep the current SOM resolution or increase the resolution using the scheme outlined in Fig. 3.2. We iterate through these steps until we reach the given final temperature  $T_{\text{min}}$  and final resolution  $R_{\text{max}}$ .

on the same area. Thus, we distinguish between two galaxy samples:

1. A photometric galaxy sample of galaxies, which is comprised of objects that are observed through several optical filters. The photometric measurements of those objects are used to infer redshift estimates via the template fitting code BPZ (Benítez, 2000).
2. A reference sample, which is comprised of objects with precise redshift measurements, for example through spectroscopic observations.

Here, we assume for simplicity that the reference sample is fully representative of the photometric sample. While the method can be applied with a reference sample that only partially overlaps with the photometric sample, as discussed in section 3.4, the impact of an inhomogeneous reference sample will be explored in forthcoming work.

Both the photometric sample and the reference sample are divided into  $N_{\text{bins}}$  redshift bins based on photometric redshift estimates and spectroscopic redshift measurements, respectively. We then measure the two-point correlation function between photometric bins and reference bins using the public code TREECORR (Jarvis et al., 2004). Using the angular positions of galaxies, we compute the cross-correlation  $w_{ij}^{\text{phot-ref}}$  between photometric bin  $i$  and reference bin  $j$  via the Landy-Szalay estimator (Landy and Szalay, 1993) defined in Eq. (1.42). For each bin of angular separation  $\Theta$  the cross-correlation is defined as

$$w_{ij}^{\text{phot-ref}} = \frac{D_i^{\text{phot}} D_j^{\text{ref}} - D_i^{\text{phot}} R_j^{\text{ref}} - R_i^{\text{phot}} D_j^{\text{ref}} + R_i^{\text{phot}} R_j^{\text{ref}}}{R_i^{\text{phot}} R_j^{\text{ref}}}, \quad (3.1)$$

where  $D_i^{\text{phot}} D_j^{\text{ref}}$  denotes the number of observed galaxy pairs of the photometric and reference bins within a single angular bin with range  $\theta \in [0.01^\circ, 0.1^\circ]$ .  $D_i^{\text{phot}} R_j^{\text{ref}}$  and  $R_i^{\text{phot}} D_j^{\text{ref}}$  denote the number of observed galaxy pairs of a random sample with uniform density that follows the geometry of the survey and the photometric and reference bins, respectively. Finally,  $R_i^{\text{phot}} R_j^{\text{ref}}$  denotes the number of galaxy pairs of random samples.

After calculating the cross-correlation between all combinations of photometric and reference bins, we construct the cross-correlation matrix whose elements are defined via

$$\rho_{ij} = \frac{w_{ij}^{\text{phot-ref}}}{w_j^{\text{ref}}}. \quad (3.2)$$

Here,  $w_j^{\text{ref}}$  denotes the auto-correlation of reference bin  $j$ , which serves as a normalisation factor and is calculated by replacing  $D_i^{\text{phot}}$  with the reference sample  $D_j^{\text{ref}}$  in Eq.(3.1).

The correlation matrix defined in Eq. (3.2) acts as a measure of how

well the galaxies in each bin of the photometric sample match the underlying true redshift bin. If the redshifts of photometric galaxies were perfectly determined, the correlation matrix between photometric and reference bins would therefore resemble a diagonal matrix. However, we expect non-zero correlation signals between neighbouring redshift bins that are induced by the large structure at their common boundary. The relative magnitude of these off-diagonal correlation signals is dependent on the width of the redshift bins and therefore we expect these signals to be small given the relatively broad redshift bins considered in this work. Additionally, noise and catastrophic outliers in the redshift estimation lead to mismatches between the photometric redshift estimates and the underlying truth and therefore reduce the correlation signal on the diagonal elements. Consequently, they induce a correlation signal on the off-diagonal elements of the cross-correlation matrix. Therefore, we aim to optimise the correlation matrix with the goal to achieve convergence towards a diagonal matrix, which would indicate an optimal assignment of photometric galaxies to redshift bins. The optimisation algorithm requires an objective function, which we define as the difference between the average elements on the diagonal and the off-diagonal elements of the covariance matrix:

$$E \equiv \frac{1}{N_{\text{bins}}} \sum_i \left( \rho_{ii} - \frac{1}{N_{\text{bins}} - 1} \sum_{i \neq j} \rho_{ij} \right), \quad (3.3)$$

where  $N_{\text{bins}}$  denotes the number of tomographic redshift bins. This equation, which quantifies the diagonality of the matrix, defines the so-called ‘energy’ of the system, which the simulated annealing algorithm maximises in order to optimise the assignment of photometric galaxies into redshift bins. Furthermore, our choice of normalisation ensures that the energy is independent of the total number of tomographic redshift bins.

Future applications of this work include studies of a more realistic setup where the reference sample consist of a collection of spectroscopically observed galaxies which are not necessarily representative of the photometric sample.

In this case, the correlation matrix is dependent on the galaxy bias of the photometric and reference samples. Assuming a linear bias model (Kaiser, 1984), the mean galaxy overdensity is related to the mean matter overdensity via

$$\delta_g = b\delta_m, \quad (3.4)$$

where the bias  $b$  can depend on the scale and on colour, redshift, and morphology of galaxies (Fry, 1996; Mann et al., 1998; Tegmark and Peebles, 1998). For a representative reference sample the cross-correlation between the photometric sample and the reference sample is proportional to the product of the biases (see for example Moessner and Jain, 1998),

$$w_{ij}^{\text{phot-ref}} \propto b_i^{\text{phot}} b_j^{\text{ref}}, \quad (3.5)$$

while for the auto-correlation of the reference sample we find

$$w_i^{\text{ref}} \propto (b_i^{\text{ref}})^2. \quad (3.6)$$

Here, we assumed a redshift-dependent galaxy bias, which, however, is assumed to be constant within the boundaries of the photometric redshift bins. The diagonal elements of Eq. (3.2) for a photometric sample with perfect redshift estimates become

$$\rho_{ii} = \frac{b_i^{\text{phot}}}{b_i^{\text{ref}}}, \quad (3.7)$$

which is equal to one since we assumed a representative reference sample.

### 3.2.2 Simulated annealing

To optimise the assignment of photometric galaxies to tomographic redshift bins we employ a simulated annealing algorithm (Kirkpatrick et al., 1983; Kirkpatrick, 1984; Černý, 1985), which is a technique inspired by the process of heating and cooling metals to reduce their defects and thus maximising the energy of the given system. This method was originally applied in optimisation problems in large discrete parameter spaces and has been generalised to

optimisation problems in continuous parameter spaces (Vanderbilt and Louie, 1984; Corana et al., 1987; Bohachevsky et al., 1986; Press and Teukolsky, 1991). An overview of the algorithm and its applications is provided, for example, in Henderson et al. (2003) and Press et al. (2007). Simulated annealing has been widely used in various fields, including astrophysics. For example, Stoica et al. (2008) used simulated annealing to detect filaments in galaxy catalogues, while Tempel et al. (2014, 2016) applied it as a tool to detect patterns in filaments. Moreover, Chira and Plionis (2019) applied a simulated annealing algorithm to identify patterns in astronomical images. While In contrast to typical optimisation methods, which usually aim to find the exact optimum, the simulated annealing algorithm achieves an approximation of the global optimum (Mitra et al., 1986). In this work, the system is characterised by a set of labels that refer to the redshift bin of each individual galaxy in the photometric sample and the energy of the system is defined in Eq. (3.3). Given the large number of observed galaxies in photometric surveys we deem exact optimisation methods computationally infeasible and thus we employ the simulated annealing algorithm to optimise the sorting of galaxies into redshift bins. Additionally, the algorithm features a method of avoiding local extrema which allows for finding an approximation of the global optimum of the objective function.

The simulated annealing algorithm works as follows: The system is characterised by a set of labels  $l_k$  that denote the redshift bin to which each photometric galaxy  $k$  is assigned. For a given set of labels we use Eq. (3.3) to measure the current energy of the system. Additionally, the system is assigned a temperature  $T$  which is a hyperparameter that decreases exponentially from an initial temperature  $T_{\max}$  to a temperature  $T_{\min}$ . At each iteration of the algorithm the state of the system is altered, i.e. a subset of galaxies is assigned to a different redshift bin, resulting in a change in the system's energy. We then calculate the change in energy  $\Delta E$  which is used in conjunction with the temperature to determine whether the altered state is accepted or rejected. A

value  $\Delta E > 0$  indicates that the altered state provides a better solution to the optimisation problem and therefore the new state is accepted. If  $\Delta E < 0$ , the altered state provides a worse solution to the optimisation problem. However, the algorithm allows for a temporary acceptance of a worse solution in order to be capable of leaving local maxima of the objective function and finding the global solution to the optimisation problem. This is achieved by drawing a random number  $\alpha$  in the interval  $[0, 1]$  and comparing the change in energy to the current temperature of the system by evaluating

$$P = \exp\left(\frac{\Delta E}{T}\right). \quad (3.8)$$

If  $P > \alpha$ , the altered state is accepted and otherwise it is rejected. This allows the algorithm to temporarily explore regions of lower energy that provide a worse solution to the optimisation problem. Since the temperature decreases exponentially, the acceptance probability of a state that worsens the optimisation also decreases over time, so that eventually the algorithm with a high probability only accepts states that provide a better solution to the optimisation problem. Therefore it is important to determine the appropriate setting of the initial and final temperatures,  $T_{\max}$  and  $T_{\min}$ , so that the algorithm starts with a reasonable probability of accepting worse solutions and finishes at a temperature at which only states that provide a better solution are accepted.

### 3.2.3 Self-organising maps

The selection of galaxies that are reassigned to a different redshift bin in each iteration is a crucial step in the simulated annealing algorithm. The energy defined in Eq. (3.3) is dependent on the angular cross-correlation between the photometric and reference sample which we optimise by reassigning galaxies to alternative redshift bins. Thus, it is essential to select a set of galaxies that are reassigned together in order to achieve a measurable effect on the objective function. Additionally, we want to select groups of galaxies that are likely to belong to the same tomographic bin, so that they can be reassigned

to a common redshift bin. As a tool to select groups of galaxies we use a self-organising map that is trained on the colour measurements of individual galaxies in the photometric sample. This allows for the selection of galaxies of a similar colour, which we expect to also be close in redshift.

A self-organising map (SOM; Kohonen, 1990) is a type of artificial neural network that produces a low-dimensional representation of high-dimensional data using an unsupervised learning technique. SOMs have been widely used in the context of photometric redshift estimation in large cosmological surveys. In particular, Geach (2012) first proposed a SOM as a tool for object classification and estimation of photometric redshifts with deep spectroscopic samples which was applied to data from the Cosmological Evolution Survey (COSMOS; Scoville et al., 2007). Additionally, Way and Klose (2012) used a SOM to estimate photometric redshifts of galaxies from SDSS DR7 (Abazajian et al., 2009). SOMs were also used to characterise spectroscopic outliers in SDSS (Fustes et al., 2013) and to derive photometric redshift PDFs (Carrasco Kind and Brunner, 2014; Speagle and Eisenstein, 2017). Recently, SOMs have been used in weak lensing surveys such as KiDS (Wright et al., 2020a; Hildebrandt et al., 2021) and DES (Buchs et al., 2019; Myles et al., 2021) to calibrate photometric redshift distributions of weak lensing source catalogues in various tomographic bins.

In this work, we project a data set containing five colour measurements (u-g, g-r, r-i, i-z, and z-y) onto a two-dimensional space. The map space of the SOM consists of nodes that are arranged on a two-dimensional grid that is usually connected via a rectangular or hexagonal geometry. Furthermore, the topology of the map can be chosen as either planar or toroidal, where the top and bottom as well as the left and right edges are connected to avoid boundary effects. The total number of nodes determines the so-called resolution of the SOM, which determines how well the SOM can separate features in the original data space. In this work, we refer to a map consisting of a rectangular grid of ( $R \times R$ ) nodes as a map of resolution  $R$ . For every SOM node there exists a

weight vector that links the node to a point in the original high-dimensional data space and thus consists of the corresponding colour values. The training process iteratively alters the randomly initialised weight vectors so that they provide a mapping between the SOM nodes and the data set. In each step the Euclidean distance between the weight vectors and a randomly selected data point is computed. The weight vector of the node that is closest to the data point is called the best matching unit (BMU). Additionally, the neighbourhood of the BMU is identified, which consists of all nodes within a given radius around the BMU. All weight vectors inside the neighbourhood are shifted towards the data point by a fraction of their distance to the data point. This fraction is dependent on the distance between weight vector and data point so that the closer a node is to the BMU, the more its weight vector is shifted. This process is repeated for all data points in the training sample. Moreover, the radius of the neighbourhood around the BMU shrinks over time, so that the number of altered weight vectors in each training step also decreases.

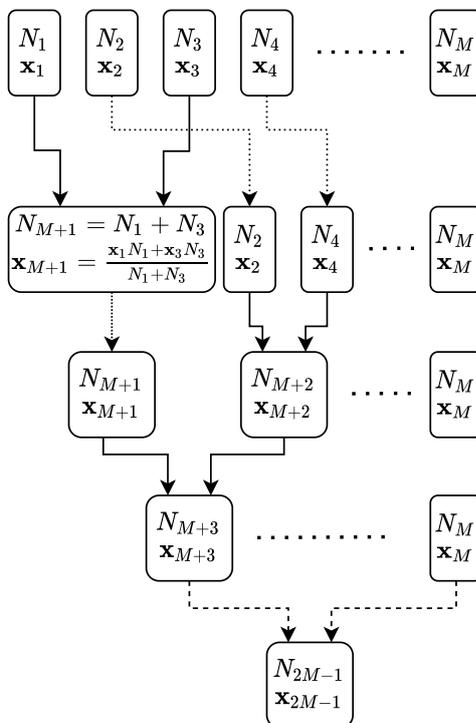
After training, the weight vectors provide a mapping of the galaxy sample onto a two-dimensional space where galaxies of similar colour are mapped close together while dissimilar galaxies are mapped further apart. Galaxies that are mapped onto a specific node then form a set of galaxies which are close in the original colour space. The total number of SOM nodes then dictates how accurate galaxy clusters in the original colour space can be separated.

Since the SOM groups galaxies in cells that are similar in colour space, we expect that these galaxies are also close in redshift. Therefore, we make use of the SOM nodes to select sets of galaxies that are assigned to a different redshift bin in each step of the simulated annealing algorithm. Thus, the resolution of the SOM determines the number of galaxies that are relabelled at a time, which imposes a limit on the accuracy of the resulting final assignment of galaxies to redshift bins. While a low-resolution SOM allows us to relabel more galaxies at a time and thus results in a shorter runtime of the algorithm, a high-resolution SOM gives a more accurate result since it allows for a finer separation of

galaxies in colour space. Thus, it is advantageous to vary the resolution while running the algorithm, starting with a SOM at a low resolution, denoted  $R_{\min}$ , and increasing the resolution over time up to the maximum resolution  $R_{\max}$ . The advantage of this method is that in the beginning, when we expect the fraction of mislabelled galaxies to be highest, we reassign a larger number of galaxies at a time. By increasing the resolution over time we continuously split the SOM nodes into two, which allows for a finer separation of galaxies in colour space so that the accuracy of the final assignment improves. Finally, we stop the algorithm at a resolution  $R_{\max}$ , at which the average number of galaxies per node becomes so small that continuing the relabelling becomes computationally infeasible given the small impact on the energy of the system.

In order to be able to dynamically scale the resolution of the SOM, we train a SOM at a high resolution and apply clustering methods to merge nodes that are close in the original data space. From the hierarchy of merged SOM nodes we can then extract a SOM with a lower number of nodes, which corresponds to a lower resolution. This allows us generate SOMs with any arbitrary resolution lower than the original resolution without training multiple SOMs. Standard hierarchical clustering techniques (see for example [Müllner, 2011](#)) allow for building a hierarchy, where two objects that minimize a given agglomeration criterion are clustered in each step. Thus, this method can be used to iteratively merge the two SOM nodes with the minimum distance between their corresponding weight vectors. However, such clustering methods, applied to the weight vectors of the SOM, do not make use of the information on the number of galaxies in each node. Therefore, we perform the SOM clustering using a weighted method, illustrated in Fig. 3.2, that utilises the additional information on the number of galaxies per node.

We construct clusters in a bottom-up approach by iteratively merging nodes until only a single node is left. Initially, every node of the high-resolution SOM forms its own cluster. We identify the two nodes with the smallest distance between their corresponding weight vectors. Using the number of



**Figure 3.2:** Sketch of the clustering method that is used to decrease the resolution of a SOM. We combine nodes of a high-resolution SOM, consisting of  $M$  nodes, using the number  $N$  of galaxies that are mapped onto each SOM node and the weight vectors  $\mathbf{x}$ , that connect the two-dimensional SOM data space to the original high dimensional data space. In each step we merge the two SOM nodes with the minimum distance of their corresponding weight vectors and compute the weight vector of the combined node as the average of the two weight vectors, where the number of galaxies in each node act as weights. By iterating the merging process we build a hierarchy of clusters until all SOM nodes are merged into one single node after  $M - 1$  steps. From the hierarchy of nodes we can derive SOMs with any resolutions lower than the resolution of the original SOM.

galaxies that are assigned to each node as weights, we compute the weight vector of the combined node by calculating the weighted average of the two weight vectors. Starting with a SOM consisting of  $M$  nodes, we are left with one single node after  $M - 1$  clustering steps. After building the hierarchy of merged SOM nodes we can then infer SOMs at resolutions lower than the resolution of the initial SOM.

### 3.3 Results

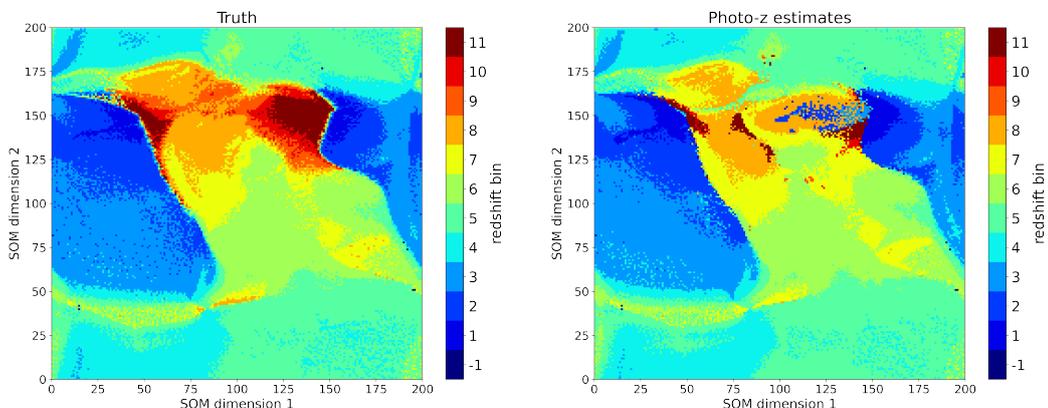
We apply the redshift calibration method to the cosmoDC2 catalogue (Korytov et al., 2019). This is a large synthetic catalogue designed by the LSST Dark Energy Science Collaboration to support the development of analysis pipelines. In particular, we employ a subset of the `cosmoDC2_1.1.4` catalogue, covering about  $58 \text{ deg}^2$  of the sky with a magnitude limit of  $i < 25.3$ , which corresponds to the LSST gold sample selection for weak lensing (LSST Science Collaboration et al., 2009). This catalogue provides colour measurements of approximately  $10^7$  galaxies with redshifts  $0 < z < 3$  in the six LSST filter bands (u,g,r,i,z, and y). The photometric redshift is estimated via the template fitting code BPZ (Benítez, 2000). In Appendix B.1 we provide a comparison between the point estimate of the photometric redshift and the true redshift of galaxies in the catalogue. Based on the photometric redshift estimate we divide the catalogue into ten bins of equal redshift width between  $0 < z < 2$ , where the redshift range of the  $i$ -th bin is defined as  $[0.2(i - 1), 0.2i]$ , and one additional bin with  $z > 2$ . To generate the reference sample of galaxies with well-calibrated redshifts, we draw a random subset of the catalogue that contains 10 per cent of the total galaxies. We assume that we are provided with precise redshift measurements for the galaxies in this subsample and therefore we divide the reference sample into the aforementioned redshift bins using the true simulated redshift. Since in this case the reference sample is perfectly representative of the photometric sample, Eq. (3.7) shows that for an optimal assignment of photometric galaxies the correlation matrix should become close to an identity matrix with small

contributions on off-diagonal elements between neighbouring bins, which are induced by the large-scale structure at their common boundary. However, the resolution of the SOM determines the number of galaxies that are typically grouped into one node. Therefore, it imposes a limit on how well the SOM can separate galaxies by redshift, so that we do not expect the correlation matrix to converge to an exact identity matrix.

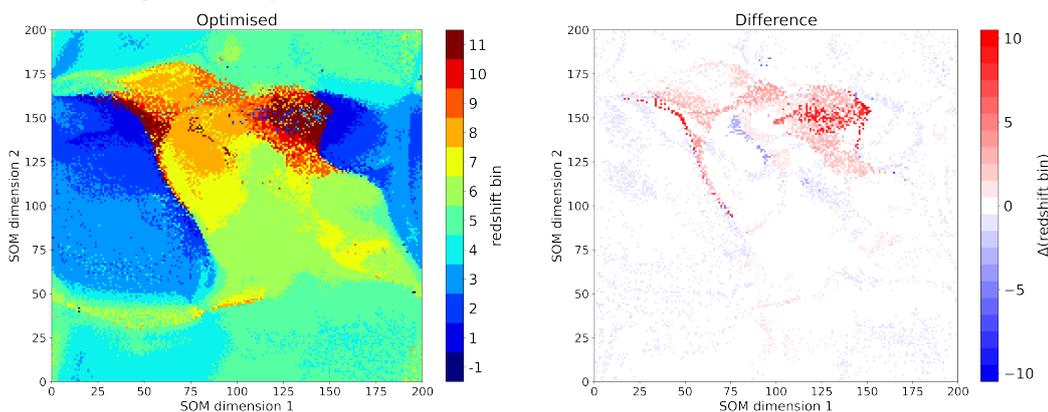
We train a self-organising map and 200x200 nodes on a rectangular grid on the observed colours of galaxies in the photometric sample using the public code SOMOCLU (Wittek et al., 2017) and choose a toroidal geometry to avoid boundary effects. The SOM is illustrated in Fig. 3.3a with colours representing the mean of the true simulated redshift of galaxies in each node. We note that the true redshift is used solely for illustration purposes and is not used in the further analysis of the photometric sample. In Appendix B.2 we provide comparisons between SOMs at different resolutions that are inferred from the high-resolution SOM using the method described in section 3.2.3.

As can be observed in Fig. 3.3a, the SOM achieves a separation of galaxies by redshift by relying purely on the colour information of individual galaxies. Furthermore, we find regions where high-redshift nodes are adjacent to low-redshift nodes, which we presume is where catastrophic errors in the photometric redshift estimate preferentially occur (Masters et al., 2015).

We assign each SOM node to a redshift bin, depending on the most common photometric redshift bin of galaxies in each node. This is illustrated in Fig. 3.3b with colours indicating tomographic redshift bins. We note that a small number of SOM nodes do not contain any galaxies, so that they cannot be assigned to a redshift bin, which is indicated by a label of ‘-1’. Comparing Fig. 3.3b to the true redshifts, shown in Fig. 3.3a, we find that some SOM nodes show a significant mismatch between the true redshift and photometric redshift bin, especially in the region of high-redshift galaxies. This is further emphasised by the left panel of Fig. 3.4, which shows the cross-correlation matrix between the photometric and reference samples. We find a rather low

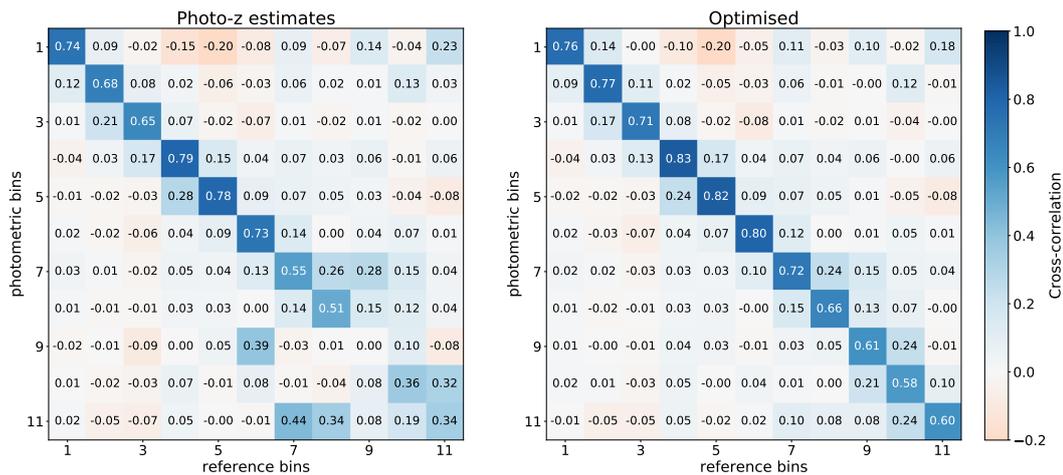


- (a) Colours indicate the photometric redshift bin of each SOM node, based on the true redshift of individual galaxies, which does not enter the optimisation process and is used for illustration purpose of the ideal bin assignment only.
- (b) Colours indicate the initial photometric redshift bins of each SOM node, based on the estimate of the photometric redshift of individual galaxies inferred with BPZ.



- (c) Colours represent the final redshift bin of each SOM node after running the simulated annealing algorithm.
- (d) Colours represent the shift of the redshift bin of each node between panels b and c.

**Figure 3.3:** Illustration of the self-organising map used in the analysis. The SOM consists of 200x200 nodes on a rectangular grid in toroidal geometry and is trained on the observed colours of galaxies in the photometric galaxy sample. Coloured labels indicate the tomographic bin to which galaxies in each node are assigned, with SOM nodes which do not contain any galaxies labelled with ‘-1’.



**Figure 3.4:** Left: Initial cross-correlation matrix between bins of the photometric sample, inferred from photometric redshift estimates, and the reference sample. Right: Cross-correlation matrix after optimisation of the bin assignment of the photometric sample. A perfect assignment with noise-free clustering measurements would yield the identity matrix.

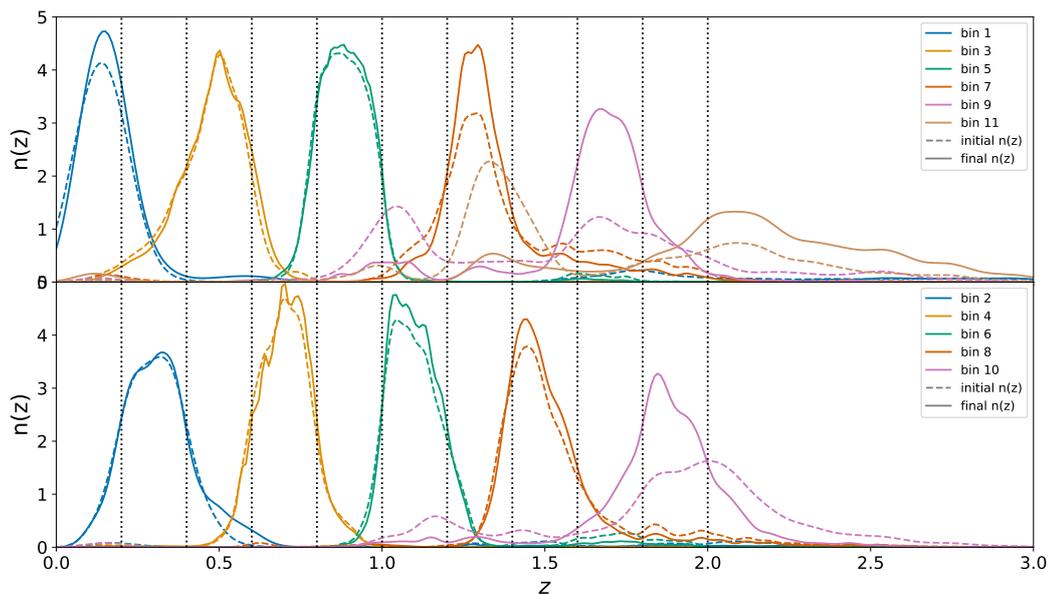
correlation signal on the diagonal for high redshift bins, indicating a mismatch of the redshifts of galaxies in these bins. Looking at the off-diagonal elements, we find high non-zero cross-correlation signals between the photometric and reference samples. This suggests that high-redshift bins are contaminated with low redshift galaxies and vice versa, caused by catastrophic failures in the photometric redshift estimation with BPZ.

We then proceed with re-sorting galaxies in the photometric sample to different tomographic bins using the simulated annealing algorithm described in section 3.2.2. Here, the initial state of the system is the set of redshift bin labels obtained from the photometric redshift estimates illustrated in Fig. 3.3b. We start with an initial SOM resolution of  $R_{\min} = 30$ , obtained with the method outlined in section 3.2.3. The SOM resolution is increased over time until reaching the final resolution of  $R_{\max} = 80$ . This is achieved by coupling the SOM resolution linearly to the temperature of the system which decreases from  $T_{\max}$  to  $T_{\min}$ . The range of the SOM resolution is chosen so that in the initial phase a larger portion of galaxies is relabelled which then decreases with increasing SOM resolution. We determine the maximum temperature such that initially there is a chance of about 50% to accept a worse state for a typical

value of  $\Delta E$  at a resolution of  $R_{\min}$ . The minimum temperature on the other hand is chosen such that the chance of accepting a worse state at resolution  $R_{\max}$  approaches zero. The simulated annealing algorithm returns a modified set of redshift bin labels, where SOM nodes were relabelled to different redshift bins in order to diagonalise the cross-correlation matrix shown in the left panel of Fig. 3.4. The resulting optimised matrix is shown in the right panel of Fig. 3.4. We observe that the algorithm succeeds in reducing the cross-correlation signal between photometric and reference bins on the off-diagonal elements while increasing the auto-correlation signal on the diagonal and thus increasing the energy of the system, defined in Eq. (3.3), from 0.56 to 0.68. In Appendix B.3 we discuss the evolution of the energy during the simulated annealing optimisation.

The resulting redshift bins of each SOM node are illustrated in Fig. 3.3c, which is the equivalent to Fig. 3.3b, but instead of the initial redshift bin labels we show the modified labels returned by the algorithm. A comparison of the two figures shows that the simulated annealing algorithm indeed succeeds in identifying those regions of the SOM where the photometric redshift estimates do not match the true redshift of the galaxies and therefore shifts these nodes towards higher redshift bins. This is further illustrated in Fig. 3.3d, where we show the magnitude of the shift in the redshift label per SOM node with positive numbers indicating a shift towards higher redshift bins and negative numbers indicating a shift towards a lower redshift bin. We find that the most significant changes occur in the aforementioned high-redshift nodes and at the boundaries between high- and low-redshift nodes.

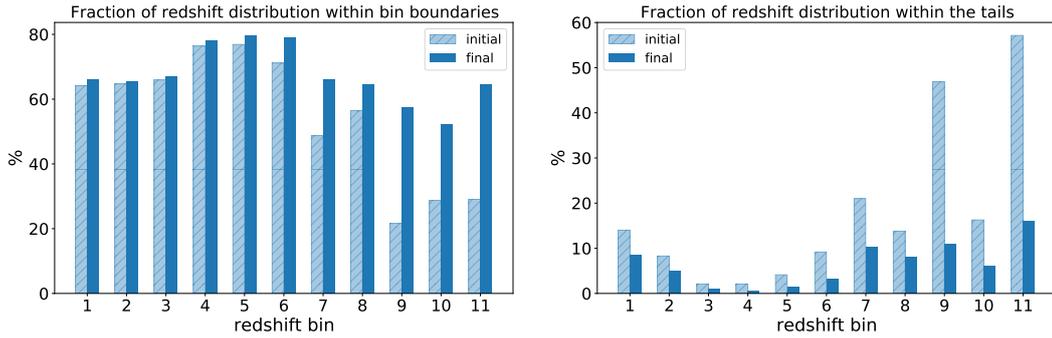
The redshift distributions of the tomographic bins are illustrated in Fig. 3.5 with dashed lines indicating the initial redshift distributions and solid lines representing the distributions after relabelling via simulated annealing. We note that the redshift distributions are inferred from the true underlying redshifts of galaxies assigned to each bin, which are not available in a real observational data set and need to be calibrated separately, e.g. via cross-



**Figure 3.5:** Comparison of the initial redshift distribution of each tomographic bin, obtained using the photometric redshift estimate of individual galaxies (dashed lines), and the redshift distribution after simulated annealing (solid lines). Dotted lines indicate the redshift bin edges.

correlation measurements. We find that the algorithm significantly improves the redshift distributions of high-redshift bins, which initially showed significant deviations from the predefined redshift intervals in the tails of the distribution. The correlation matrix shown in Fig. 3.4 indicates a low-level anticorrelation between photometric bin 1 and reference bins 4 and 5 that remains even after optimisation. However, this feature is not observed in Fig. 3.5 which shows no significant overlap between the redshift distribution of bin 1 with either bin 4 or bin 5. The origin of this feature is unclear.

We quantify the extent to which the redshift distributions lie within the boundaries of the redshift bins in the left panel of Fig. 3.6, comparing the initial and final distribution. We find that the algorithm helps to shift the redshift distribution significantly to lie within the bin boundaries, especially in the higher redshift bins, where we find improvements of about 30 per cent. Additionally, in the right panel of Fig. 3.6 we quantify how much of the redshift distribution is located within the tails of distribution. Here, we define the tails of the distribution as the region in redshift space that lies more than



**Figure 3.6:** Left: Comparison of the percentage of probability mass of the redshift distributions, shown in Fig. 3.5, within the respective bin range before and after simulated annealing. Right: Comparison of the percentage of the redshift distribution that is located within the tails of the distribution, where we define the tails of the distribution as the region in redshift space that lies more than one bin width outside of the boundaries of a given tomographic bin.

one bin width outside of the boundaries of a given tomographic bin. We find a substantial decrease of the tails of the redshift distribution. Again, the biggest improvements are found in the high redshift bins where initially a large percentage of the distribution is located in the tails, which decreases significantly after the simulated annealing. The biggest change is found in bin 11, which initially only contains 32 per cent of the probability mass within the bin boundaries and a large fraction of about 57 per cent in the tails of the distribution. These quantities shift significantly in the final redshift distribution, with about 64 per cent within the bin boundaries and 16 per cent in the tails. Furthermore, we find significant improvements in bin 9, which initially contains about 22 per cent of galaxies within bin boundaries which increases to 57 per cent. The fraction of galaxies in the tails of this redshift bin also decreases significantly from 47 per cent to 11 per cent.

### 3.4 Conclusions

In this paper we presented a method, SHARPZ, to group a sample of galaxies into tomographic redshift bins using estimates of the photometric redshift with subsequent re-sorting using an algorithm that optimises the angular cross-correlation between the photometric galaxy sample and an overlapping

sample of reference galaxies. We utilised a simulated annealing algorithm that reassigns groups of galaxies to redshift bins and determines the effect on the cross-correlation matrix by calculating a measure of the diagonality of the matrix. This was combined with a self-organising map (SOM) that was trained on the colour information of photometric galaxies. The SOM allows choosing sets of galaxies that are reassigned in each step of the simulated annealing. Additionally, the resolution of the SOM was increased over time in order to achieve a greater accuracy of the final resulting photometric redshift bins.

We applied this method to a synthetic catalogue, cosmoDC2, that aims to resemble measurements of the upcoming Vera C. Rubin Observatory’s Legacy Survey of Space and Time. Our results show that the method significantly reduces the fraction of catastrophic outliers in the tails of the redshift distribution in all tomographic bins, most notably in the highest redshift bins where we find improvements by up to 40 per cent. We found that it succeeds in shifting the redshift distributions towards being within the boundaries of the tomographic bins. High redshift bins show the greatest improvements, where the probability mass within the bin boundaries increases up to about 30 per cent, while the improvement in the low redshift bins, whose redshift distributions initially are already quite compact, is smaller with the probability mass increasing by a few percent. Additionally, we found that the method also greatly reduces the amount of the redshift distribution that is located in the tails of the distribution. Again, we find the biggest improvement in the high redshift bins, where the initial performance of the photometric redshift estimates is worst.

The quality of the redshift distributions inferred with our optimisation method depends on the choice of the SOM parameters. Initially, we trained a SOM consisting of 200 x 200 nodes and scaled the resolution from low to high during the optimisation process. The resolution determines the number of galaxies that are reassigned together to an alternative redshift bin, which scales from large numbers of galaxies to smaller numbers. Thus, the resolution imposes a limit on how well the SOM can separate features in colour space

that can be associated with different redshift bins. By choosing a maximum resolution we therefore implicitly put a limit on the quality of the final redshift distribution that can be achieved. The quality of the final redshift distributions can be improved by increasing the maximum resolution at the cost of a longer computation time since a higher resolution implies a selection of fewer galaxies in each step, up to the limit where the algorithm selects individual galaxies. However, at a certain point noise in the clustering observable will limit the observable effect on the objective function. Therefore the method is ultimately limited by both the noise limit of the clustering measurement and the SOM resolution. Additionally, the choice of features on which the SOM is trained influences how well the SOM can separate galaxies of different redshifts. In our analysis we trained the SOM on five galaxy colours, which we found to perform well for the data set considered in this work. Future applications with alternative data sets should however explore alternative sets of training features, for example different colour combinations or the addition of magnitude information, which can help breaking colour-redshift degeneracies.

In our analysis we made the assumption that the reference sample covers the full area observed by the photometric survey, while in a realistic application the reference sample will only have a partial sky overlap with the photometric survey. However, as long as the survey is spatially homogeneous, we can optimise the assignment of galaxies to tomographic bins in the area covered by both the photometric and reference survey and then use the SOM to expand the redshift bin assignments to the full photometric survey. Therefore, a complete overlap of the two samples is not a general requirement of the method. Additionally, the SOM can be used to reproduce the results without re-running the simulated annealing algorithm. Moreover, we can further improve the calculation of the clustering signal by measuring the cross-correlation signal between reference samples and cells of the highest-resolution SOM. For a given assignment of SOM cells to tomographic bins we can then stack the cross-correlation signals of the individual cells using the hierarchical structure of the SOM. In this

way, we obtain the correlation signal between photometric and reference bins without re-calculating the cross-correlation in each step of the optimisation algorithm, which will lead to a decrease in computational cost. This is studied in section 4.1.

Additionally, we made the assumption that the reference sample is fully representative of the photometric sample. In section 4.4 we study how an incomplete reference sample influences the cross-correlation measurements between the photometric and reference sample and how it impacts the quality of the inferred tomographic bins.

While finalising this work, [Zuntz et al. \(2021\)](#) put forward a paper on the optimisation of the tomographic binning for the DESC 3x2pt analysis. In particular, the COMPLEXSOM method utilises a matrix of auto- and cross-power spectra, which is a statistic similar to the one employed in this work. While in this work the redshift bin edges of the galaxy sample are fixed, the COMPLEXSOM method instead optimises the parameters that determine the bin edges.

Our work demonstrates that the optimisation method provides a significant improvement of the redshift distribution of a synthetic survey compared to photometric estimates of the redshift. Therefore it provides a promising complement to existing redshift calibration methods in upcoming surveys. An application to observational data is left for future work.

## Chapter 4

# Optimising photometric redshift distributions with realistic reference datasets

In the previous chapter, we developed an optimisation method for the assignment of galaxies to a set of redshift bins based on multi-band photometry. This method relies on clustering cross-correlation measurements between photometrically observed galaxies and a sample of reference galaxies with known redshifts. It was tested on a simulated catalogue of galaxies, cosmoDC2, from which a sample of reference galaxies that are representative of the full galaxy catalogue was selected. This sample was comprised of 10% of galaxies in the simulated catalogue. However, in an application with real data the reference sample will be comprised of a compilation of spectroscopically observed galaxies which are obtained from overlapping surveys. Thus, the reference sample will not be representative of the photometric sample, since the selection of observed galaxies differs among surveys and is optimised for measurements of baryon acoustic oscillations and redshift-space distortions. Additionally, the density of spectroscopically observed galaxies per square degree is much lower than the density of photometrically observed galaxies, since spectroscopic observations are more time-consuming. Therefore, it is essential to replace the idealised reference sample, on which the optimised bin assignment scheme was developed,

with a more realistic set of reference galaxies which is more similar to data that can be expected from upcoming surveys.

In this chapter, we test and further develop the optimisation scheme with a set of reference galaxies that mimics galaxies observed by the Dark Energy Spectroscopic Instrument (DESI; [DESI Collaboration et al., 2016](#)). Furthermore, we develop a method to calculate the cross-correlation signal between photometric and reference samples that utilises the SOM to avoid the repeated recalculation of the clustering signal, which leads to a decrease in runtime of the optimisation algorithm.

## 4.1 SOM-based cross-correlation measurements

The optimisation method presented in chapter 3 relies on the recalculation of the clustering cross-correlation signal between photometric and reference bins galaxies after the reassignment of a SOM-selected group of photometric galaxies to a different tomographic redshift bin. Hence, the calculation of the cross-correlation signal makes up the majority of the computing time. Here, we develop a method that does not require a recalculation of the cross-correlation. We take advantage of the method of galaxy selection via the SOM. Since we assume that galaxies in each node are close in redshift and thus belong to the same tomographic bin, we can pre-compute the cross-correlation between all SOM-nodes and the set of reference bins. The full cross-correlation signal between two specific photometric and reference bins can then be calculated by stacking the cross-correlation measurements for all SOM nodes belonging to one tomographic bin. The number of observed galaxy pairs between photometric bin  $i$  and reference bin  $j$ , entering the Landy-Szalay estimator in Eq. (3.1), is given by

$$D_i^{\text{phot}} D_j^{\text{ref}} = \sum_{k \in i} D_k^{\text{node}} D_j^{\text{ref}}, \quad (4.1)$$

where  $D_k^{\text{node}} D_j^{\text{ref}}$  is the number of galaxy pairs observed between SOM node  $k$  and reference bin  $j$ . Here, we sum over all nodes  $k$  that are assigned to the  $i$ -th photometric redshift bin. Similarly, the number of galaxy pairs between random samples and photometric and reference bins and the number of galaxy pairs between random samples are given by

$$D_i^{\text{phot}} R_j^{\text{ref}} = \sum_{k \in i} D_k^{\text{node}} R_j^{\text{ref}}; \quad (4.2)$$

$$R_i^{\text{phot}} R_j^{\text{ref}} = \sum_{k \in i} R_k^{\text{node}} D_j^{\text{ref}}; \quad (4.3)$$

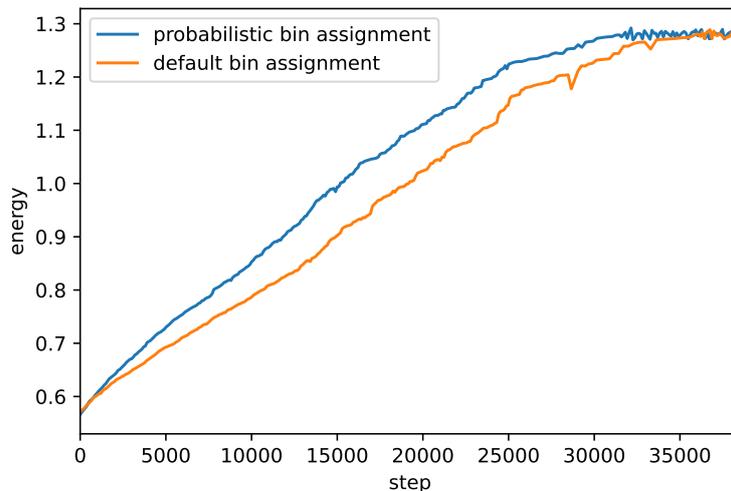
$$R_i^{\text{phot}} R_j^{\text{ref}} = \sum_{k \in i} R_k^{\text{node}} R_j^{\text{ref}}, \quad (4.4)$$

where  $R_k^{\text{node}}$  denotes a random sample with uniform density proportional to the number of galaxies in the  $k$ -th SOM node.

In each step of the optimisation we select a SOM node that is assigned to a different redshift bin, which modifies the elements  $k$  that are summed in Eqs. (4.1), (4.2), (4.3), and (4.4). However,  $D_k^{\text{node}}$  and  $R_k^{\text{node}}$  themselves remain unchanged, so that we can pre-compute and store these quantities. Thus, the calculation of the cross-correlation signal only requires a summation instead of a recalculation of the number of galaxy pairs as was done in chapter 3. This decreases the time required to construct the cross-correlation matrix defined in Eq. (3.2) by a factor of 100. Therefore, it significantly decreases the runtime of the algorithm, which is particularly useful when considering larger surveys since the computing time required to count galaxy pairs scales with the number of objects in the survey.

## 4.2 Probabilistic bin assignment scheme

The default method of reassigning a group of SOM-selected galaxies to alternative redshift bins, which was applied in chapter 3, randomly selects a new redshift bin with equal probability for all alternative redshift bins. In this section we develop a probabilistic assignment scheme in order to increase the convergence rate of the optimisation algorithm. To do so, we make use of the



**Figure 4.1:** Comparison of the energy as a function of steps between the default bin assignment scheme and the probabilistic bin assignment scheme.

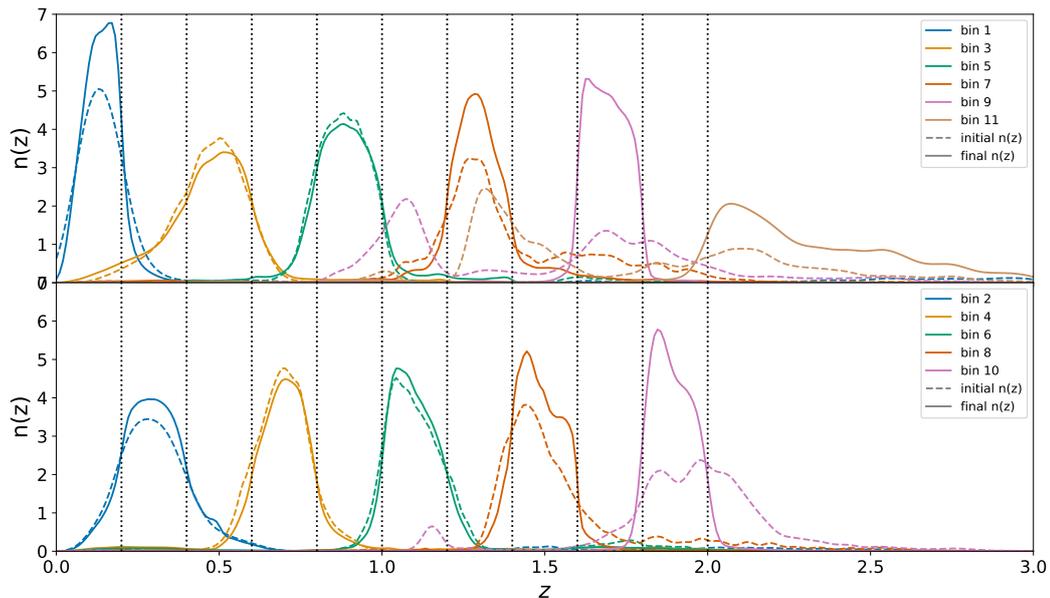
information about the photometric redshift estimate of individual galaxies in each SOM node. For each SOM node, we derive the distribution of photometric redshifts of galaxies. We then assign a selection probability to each redshift bin relative to the number of galaxies with photometric redshift within the bin boundaries. Thus, if the distribution of photometric redshift is spread across multiple bins, we find a higher probability to reassign a SOM node to a more populated bin. This allows the algorithm to converge faster to the optimal solution if there is a fraction of galaxies with photometric redshift close to the true redshift.

In Fig. 4.1, we compare the energy of the system as a function of steps for the default bin assignment scheme and the probabilistic bin assignment scheme. While the optimisation algorithm converges to the same energy with both assignment schemes, the probabilistic bin assignment requires fewer steps. Thus, the probabilistic bin assignment scheme further reduces the computing time of the optimisation algorithm in addition to the reduction in computing time due to the SOM-based cross-correlation measurements.

### 4.3 Results for idealised reference samples

We test the SOM-based cross-correlation measurements on the idealised sample of reference galaxies, which is comprised of 10% of galaxies in the cosmoDC2 catalogue. However, in contrast to the work presented in chapter 3, we employ the full cosmoDC2 catalogue, covering  $440 \text{ deg}^2$  on the sky, instead of a subset of the catalogue. This change is feasible because of the decrease in computing time due to the SOM based cross-correlation measurements. Since the SOM was trained on the colours of galaxies in the subset of the catalogue, we project the remaining galaxies in the full catalogue onto the SOM in order to assign each galaxy to a SOM node. In doing so, we assume that the dataset on which the SOM was trained is representative of the full dataset. This is the case for the cosmoDC2 dataset. However, it could still lead to a mild underperformance of the assignment of galaxies to redshift bins.

During the optimisation process we discovered a tendency of the algorithm to assign small populations of high-redshift galaxies to lower redshift bins, which was most prominent in the third and fifth tomographic bin. This is most likely due to a higher galaxy bias at high redshifts, which allows the algorithm to achieve a higher signal on the diagonal elements of the correlation matrix while only slightly increasing the signal on the off-diagonal, which is still consistent with zero within the noise. To circumvent this issue we implement two modifications in the definition of the energy, which was originally defined in Eq. (3.3). Since the energy is an objective function that is chosen arbitrarily to parameterise how close to diagonal the correlation matrix is, we can modify this definition in order to improve the optimisation. First, we implement two variables  $w_{\text{diag}}$  and  $w_{\text{offdiag}} = 1 - w_{\text{diag}}$ , which parameterise a weight that is applied to the diagonal and off-diagonal elements of the correlation matrix. This allows us to reduce the gain in energy that is achieved by moving galaxies with a high bias to low redshift bins. Additionally, for realistic reference samples we expect the optimal values on the diagonal to differ from unity since the galaxy bias of the reference sample in general is different from the galaxy bias

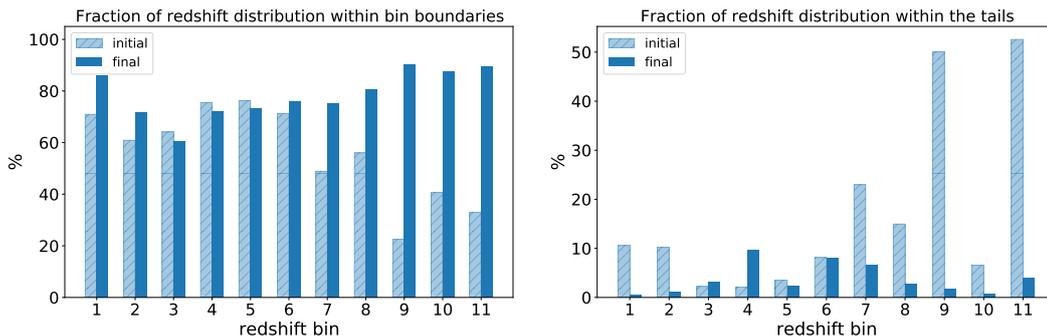


**Figure 4.2:** Comparison of the initial redshift distribution (dashed lines) and the final redshift distribution (solid lines) after optimisation of the bin assignment with a SOM-based cross-correlation measurements. Dotted lines indicate the redshift bin edges.

of the photometric sample, which can increase the gain in energy even further. Thus, downweighting the diagonal values circumvents this issue. Secondly, we take the absolute value of the off-diagonal elements before summation in order to prevent negative entries on the off-diagonal compensating positive entries. With these modifications, the energy reads

$$E \equiv \frac{1}{N_{\text{bins}}} \sum_i \left( w_{\text{diag}} \rho_{ii} - \frac{1 - w_{\text{diag}}}{N_{\text{bins}} - 1} \sum_{i \neq j} |\rho_{ij}| \right). \quad (4.5)$$

The results of the optimisation with  $w_{\text{diag}} = 0.2$  are shown in Fig. 4.2. We note that since the dataset is comprised of the full cosmoDC2 catalogue, the initial redshift distributions differ slightly from the ones shown in Fig. 3.5, which were obtained from a subset of the catalogue. We find a better performance of the optimisation algorithm in bins 1, 9, 10, and 11 which are more localised within their bin boundaries compared to the earlier results presented in Fig. 3.5. Additionally, we find that the outlier rate is significantly reduced, which is most prominent in the high-redshift bins which in Fig. 3.5



**Figure 4.3:** Left: Comparison of the percentage of probability mass of the redshift distributions, shown in Fig. 4.2, within the respective bin range before and after optimisation. Right: Comparison of the percentage of the redshift distribution that is located within the tails of the distribution, where we define the tails of the distribution as the region in redshift space that lies more than one bin width outside of the boundaries of a given tomographic bin.

featured secondary peaks at low redshift that are mostly removed in Fig. 4.2. These observations are illustrated in Fig 4.3, which shows a comparison of the percentage of probability mass of the redshift distributions within the respective bin range before and after optimisation and a comparison of the percentage of the distribution located within the tails. However, we find that the final redshift distributions in bins 3, 4, and 5 are slightly worse than the initial distributions. This is most likely a residual effect of the assignment of high-redshift outliers to low redshift bins. This issue can potentially be removed by further optimising the weighting between elements on the diagonal and the off-diagonal. This is left for future work.

## 4.4 Results for DESI-like reference samples

The Dark Energy Spectroscopic Instrument (DESI) will measure spectra of about 35 million objects which are selected from three photometric surveys, referred to as the DESI Legacy Imaging Surveys (Dey et al., 2019). These surveys completely overlap with DESI’s footprint which covers approximately 14000 deg<sup>2</sup> on sky divided into two contiguous regions in the Northern and Southern Galactic Caps covering an area of 9900 deg<sup>2</sup> and 4000 deg<sup>2</sup>, respectively. Given the overlap between footprints of DESI and LSST on a few thousand square

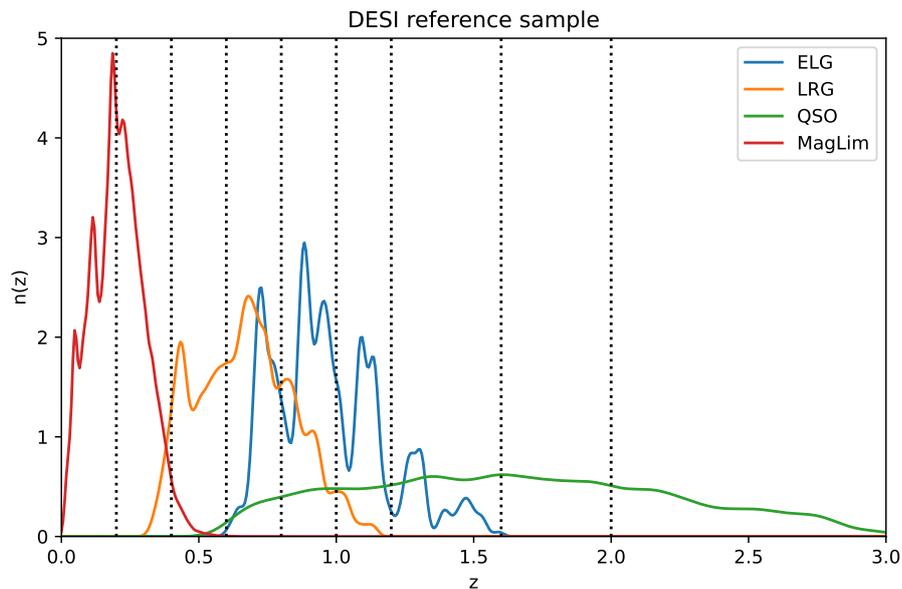
degrees on sky, depending on the observing strategy (DESI Collaboration et al., 2016), its spectroscopic observations of galaxies make it an ideal source of reference galaxies for the characterisation of the redshift distribution of LSST galaxies.

In this section, we explore the optimisation of tomographic bins of LSST with a reference sample mimicking a sample observed by DESI which is available in the cosmoDC2 catalogue. This sample is constructed from the cosmoDC2 catalogue with a DESI-like selection of four different types of objects: Luminous Red Galaxies (LRGs), Emission Line Galaxies (ELGs), high redshift quasars (QSOs), and a magnitude-limited sample (MagLim) approximating the DESI Bright Galaxy Sample. The samples are selected as follows: for QSOs the objects with the highest star formation rate (SFR) and a magnitude cut of  $19.5 < r < 23.4$  are selected so that the number of objects per redshift and per square degree,  $dN_{\text{QSO}}/dz d\Omega$  with  $d\Omega = \sin\theta d\theta d\phi$ , matches the values given in Table 2.3 of DESI Collaboration et al. (2016). Subsequently, ELGs are selected with the same criteria, but with an additional cut on the maximum SFR defined as the minimum SFR of QSOs in order to ensure that the samples are independent. The LRG sample is obtained by selecting objects with the largest stellar mass while also limiting the maximum SFR to the minimum SFR of the ELG sample. Finally, the MagLim sample is selected with a magnitude cut of  $r < 19.5$  which ensures that this sample is independent of the LRG, ELG, and QSO samples. This selection method provides us with four samples of different classes of objects akin to those that will be observed by DESI.

The corresponding redshift distributions of the four reference samples are illustrated in Fig. 4.4. However, the redshift distributions in this figure show some distinct features that would not be expected from a realistic catalogue mimicking the DESI survey. The redshift distributions show dips in the redshift distribution that are most prominent in the distribution of ELGs and which are not present in the redshift distribution given by DESI Collaboration et al. (2016). These dips are caused by unrealistic degeneracies between the redshift

and the SFR of galaxies in the cosmoDC2 catalogue. Similar degeneracies are found between the redshift and the stellar mass. SFRs and stellar masses are used to select galaxies in rather broad histogram bins with  $\Delta z = 0.1$ , given by [DESI Collaboration et al. \(2016\)](#). Therefore, the degeneracies affect the redshift distribution within the histogram bins, creating dips in the distribution that would not be expected from a realistic sample of galaxies. Moreover, since the magnitude-limited sample is explicitly constructed to exclude the other samples, these features also affect the shape of the magnitude-limited sample. Constructing a more realistic sample of DESI galaxies requires a realistic modelling of the SFRs and stellar masses in the catalogue. Additionally, a finer binning of the reference histogram could reduce the dips in the redshift distribution that are picked up by the selection method. However, this is the best approximation of a DESI-like galaxy sample that is readily available in the cosmoDC2 catalogue. Since it is clearly non-representative of the photometric sample studied in chapter 3, we deem this sample acceptable for a first study of the optimisation method with non-representative reference samples and leave the creation of a more realistic DESI-like sample for future work.

The sample of photometric galaxies is comprised of the remaining galaxies in the cosmoDC2 catalogue that are not selected as reference galaxies. We divide the photometric and the reference sample into nine bins between  $0 \leq z < 3$  which are illustrated by the vertical dotted lines. Since the reference sample is comprised of four subsamples of galaxies with different properties, we expect the galaxy bias to be different for each subsample. Therefore, we keep the subsamples separated when dividing the reference sample into redshift bins. Since each subsample covers a different redshift range, we only select the reference bins with a sufficient number of galaxies for the calculation of the clustering cross-correlation with the photometric sample. We construct the correlation matrix between photometric and reference samples with 12 reference bins in total, which in some cases cover the same redshift range. A summary of the properties of the reference bins is given in Table 4.1. In contrast



**Figure 4.4:** Overview of the redshift distributions of galaxies in the cosmoDC2 catalogue with a DESI-like selection for four samples: Luminous Red Galaxies (LRGs), Emission Line Galaxies (ELGs), high redshift quasars (QSOs), and a magnitude-limited sample. The boundaries of the redshift bins are illustrated by the vertical dotted lines.

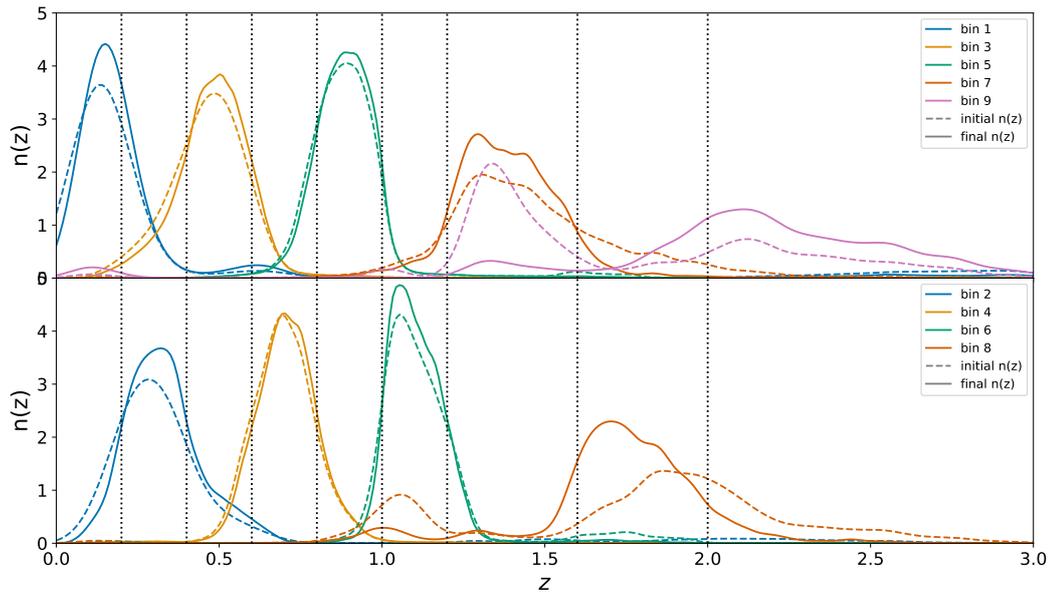
to the idealised reference sample, which was comprised of 10% of galaxies in the cosmoDC2 catalogue, the DESI-like reference sample contains only approximately 1% of galaxies in the cosmoDC2 catalogue. Furthermore, we choose broader bins above  $z = 1.2$  because of the sparsity of high-redshift reference galaxies.

After calculating the cross-correlation between the reference and photometric samples in each SOM node using the method described in section 4.1, we continue by running the optimisation algorithm on the newly constructed cross-correlation matrix. Again, we employ the definition of the energy given in Eq. (4.5) with  $w_{\text{diag}} = 0.2$ . The resulting redshift distributions are shown in Fig 4.5. Similar to the optimisation with the idealised reference sample, we find the greatest improvements in the redshift distributions of high redshift bins. Especially the secondary peaks at low redshift in the initial redshift distributions of bins 8 and 9 are greatly reduced. However, we find an increase in the redshift

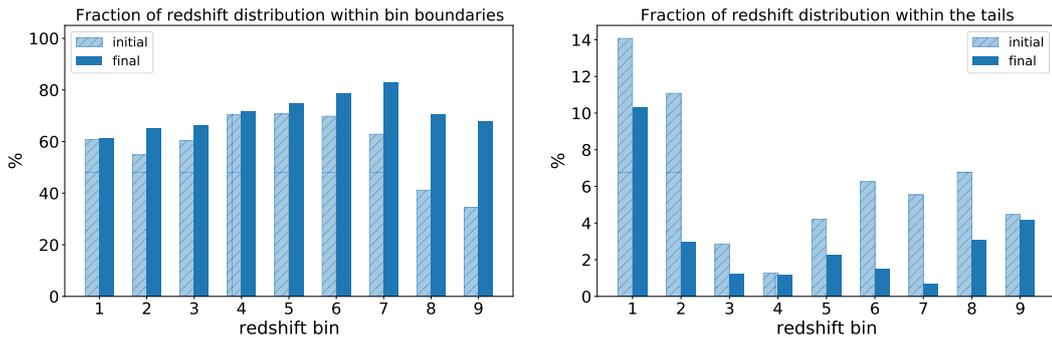
Sample	Redshift	$N_{\text{gal}}$
MagLim	$0.0 \leq z < 0.2$	132775
	$0.2 \leq z < 0.4$	136824
LRG	$0.4 \leq z < 0.6$	68795
	$0.6 \leq z < 0.8$	85809
	$0.8 \leq z < 1.0$	44812
ELG	$0.6 \leq z < 0.8$	111912
	$0.8 \leq z < 1.0$	181733
	$1.0 \leq z < 1.2$	116785
	$1.2 \leq z < 1.5$	63173
QSO	$1.2 \leq z < 1.5$	15019
	$1.5 \leq z < 2.0$	14797
	$2.0 \leq z < 3.0$	18223

**Table 4.1:** Redshift bins of a sample of reference galaxies with a DESI-like selection function. The sample is divided into nine tomographic bins between  $0 \leq z < 3$  separately for each of the four classes of objects in the DESI-like catalogue: Luminous Red Galaxies (LRGs), Emission Line Galaxies (ELGs), high redshift quasars (QSOs), and a magnitude-limited sample (MagLim).

distribution of bin 9 at  $z \approx 0.1$  and in the redshift distribution of the first bin at  $z \approx 0.6$ . These features are most likely due to the aforementioned trade-off between improvements of the signal on the diagonal at the cost of an only small increase on the off-diagonal for small populations with high galaxy bias. This effect is presumably stronger for the DESI-like reference sample which is comprised of several subsamples with different galaxy biases. Nevertheless, we find that this more realistic reference sample is capable of greatly improving the assignment of photometric galaxies to redshift bins so that the true underlying redshift distribution in each tomographic bin is substantially more localised within the bin boundaries. This is illustrated in Fig. 4.6, which again provides a comparison between the percentage of the initial and final redshift distributions localised within the bin boundaries and the tails of the distribution. This figure shows that there is improvement in the redshift distribution of all tomographic bins. On the downside, the DESI-like sample is much more sparser as can be inferred from the third column in Table 4.1. Therefore, we require broader bins



**Figure 4.5:** Same as Fig. 4.2, but after optimisation with a DESI-like selected reference sample.



**Figure 4.6:** Same as Fig. 4.3, but after optimisation with a DESI-like selected reference sample.

at high redshifts in order to obtain reference samples with a sufficiently high density of objects for the cross-correlation measurement.

## 4.5 Conclusions

We conclude that with the updated definition of the energy and the SOM-based method of constructing the cross-correlation matrix, our optimisation method is capable of significantly improving the compactness of the redshift distributions of tomographic bins with a realistic set of reference galaxies. While we require broader bins at high redshift compared to the idealised reference

sample, we still find a significant improvement in the high redshift bins. This optimisation method was tested on a simulated dataset for the upcoming LSST with a selection of reference galaxies resembling the DESI survey. In future work, we plan to apply this method to real data in order to test the performance of the optimised bin assignment scheme. For this purpose, we will employ data from KiDS, which was already extensively used in several works in the context of redshift calibration ([Wright et al., 2020a,b](#); [van den Busch et al., 2020](#); [Hildebrandt et al., 2021](#)). Therefore, it is an ideal dataset for the comparison of the performance of our optimised redshift bin assignment scheme and the resulting constraints on cosmological parameters from cosmic shear with well-established analyses.

## Chapter 5

# Internal consistency tests of cosmological data

Photometric redshift distributions play a crucial role in analyses of cosmological data. As discussed in chapter 3, we can optimise the assignment of galaxies to tomographic redshift bins. However, we still require a careful propagation of residual uncertainties in the redshift distribution into the cosmological inference from weak lensing measurements, for example using the method discussed in chapter 2. Given the importance of redshift distributions in weak lensing analyses, it is essential to assess the internal consistency between redshift bins of the weak lensing dataset in order to test the constraints on cosmological parameters for systematic effects induced by the redshift calibration. Such tests for systematic biases play a crucial role in the ongoing debate about the apparent tension between probes of the early and late Universe, discussed in section 1.4.

On the other hand, if the data itself is self-consistent, the tension between probes of the early and late Universe can be interpreted as a sign of new physical effects beyond the standard  $\Lambda$ CDM cosmological model. Therefore, it is equally important to test the consistency of the  $\Lambda$ CDM model, for example through a split into the homogeneous background and perturbations. If the data prefers a different evolution of the Universe in these two regimes, this can be seen as a sign that we require an extension of the cosmological model to alleviate

the observed tension between probes. In this chapter, we perform two such consistency tests. First, we test the internal consistency between tomographic redshift bins in the KiDS-1000 weak lensing analysis in order to assess whether or not the tomographic redshift bins prefer a different cosmology. Secondly, we perform a test of the flat  $\Lambda$ CDM model by probing the consistency between the homogeneous background and matter perturbations using a compilation of weak lensing, galaxy clustering, Lyman- $\alpha$ , and CMB datasets.

## 5.1 Bayesian statistics

To test the consistency in a cosmological analysis we perform a comparison between the standard cosmological model and an alternative model that is designed to either probe the internal consistency of the data set or the internal consistency of the cosmological model. This is achieved by employing Bayesian statistics. These are based on Bayes' theorem

$$P(\boldsymbol{\theta}|\mathcal{D}) = \frac{P(\mathcal{D}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathcal{D})}, \quad (5.1)$$

which relates the posterior probability  $P(\boldsymbol{\theta}|\mathcal{D})$  for a set of model parameters  $\boldsymbol{\theta}$  given the data  $\mathcal{D}$  to the likelihood function  $P(\mathcal{D}|\boldsymbol{\theta})$  and the prior probability of the model  $P(\boldsymbol{\theta})$ , normalised by the prior probability of the data  $P(\mathcal{D})$ . In this context,  $P(\mathcal{D})$  is called the evidence and can be computed via

$$\mathcal{Z} = \int P(\mathcal{D}|\boldsymbol{\theta})P(\boldsymbol{\theta})d\boldsymbol{\theta}. \quad (5.2)$$

Here, we integrate over the volume of the prior, so that the evidence does not explicitly depend on the model. In practice, we are usually interested in the posterior distribution of cosmological parameters, which are obtained by generating samples of the likelihood from the prior. In this way, we infer constraints on cosmological parameters while omitting the evidence factor. However, for the purpose of model comparison the evidence plays a key role.

The evidence can be interpreted as the probability of generating the data

$\mathcal{D}$  given a specific model with parameters  $\boldsymbol{\theta}$  that are sampled from the prior. In general, a model with a large number of parameters gives a lower evidence than a model with less parameters, unless the more complicated model gives a significantly better fit to the data. Thus, the evidence serves as a tool for a comparison of two models that are supposed to describe the same dataset. Assuming equal prior probabilities, we compare two models A and B using the Bayes factor

$$R = \frac{\mathcal{Z}_A}{\mathcal{Z}_B} = \frac{\int P(\mathcal{D}|\boldsymbol{\theta}_A)P(\boldsymbol{\theta}_A)d\boldsymbol{\theta}_A}{\int P(\mathcal{D}|\boldsymbol{\theta}_B)P(\boldsymbol{\theta}_B)d\boldsymbol{\theta}_B}, \quad (5.3)$$

where  $\boldsymbol{\theta}_A$  and  $\boldsymbol{\theta}_B$  denote a set of model parameters for model A and B, respectively. In general, a value of  $R > 1$  indicates preference for model A, whereas  $R < 1$  indicates preference for model B. However, the Bayes factor can only be interpreted qualitatively and there is no consensus on a threshold that signals a clear preference for a specific model. A common choice is to use Jeffreys' scale (Jeffreys, 1939) to evaluate the Bayes ratio. In this interpretation, a value of  $1 \leq R < 3$  is associated with weak preference for model A,  $3 \leq R < 20$  with definite preference, and  $20 \leq R < 150$  with strong preference. Finally,  $R \geq 150$  is interpreted as very strong preference for model A. However, due to the integration over the prior volume in Eq. (5.2), the evidence is strongly dependent on the prior. In general, a smaller prior volume increases the evidence and therefore changing the prior volume in principle allows for arbitrary changes in the Bayes factor. If the priors are well-motivated, we can rely on the Bayes factor to make a statement about the preferred model. However, in practice we often choose wide, uninformative priors, which makes the Bayes factor a suboptimal metric for model comparison.

The usage of the Bayes ratio as a metric to quantify the tension between the DES Y1 and Planck datasets is discussed in Handley and Lemos (2019a). Considering two independent datasets A and B that are combined at the likelihood level, the Bayes factor is given by

$$R = \int \frac{\mathcal{P}_A \mathcal{P}_B}{\pi} d\theta = \left\langle \frac{\mathcal{P}_B}{\pi} \right\rangle_{\mathcal{P}_A} = \left\langle \frac{\mathcal{P}_A}{\pi} \right\rangle_{\mathcal{P}_B}, \quad (5.4)$$

where  $\mathcal{P}_A$  and  $\mathcal{P}_B$  denote the posterior of the two datasets and  $\pi$  is the shared prior. Thus, [Handley and Lemos \(2019a\)](#) interpret the Bayes ratio as the average over the posterior of one dataset of the ratio between the posterior of the other dataset and the prior. Furthermore, it is pointed out that the Bayes ratio is only dependent on the priors on constrained parameters shared between both likelihoods and not on the priors on unconstrained parameters or additional nuisance parameters. The choice of wide priors, which is often made in order to be uninformative, can hide an existing tension between datasets. On the other hand, choosing a physically reasonable prior that does not significantly change the posterior distribution can provide a lower bound on the Bayes ratio.

In order to circumvent the issue of the prior dependence of the Bayes factor, [Handley and Lemos \(2019a\)](#) and [Lemos et al. \(2020\)](#) propose the so-called suspiciousness as an alternative statistic to quantify the tension between datasets. The suspiciousness is understood as the value of the Bayes factor corresponding to the narrowest prior that does not significantly change the shape of the posterior. To define the suspiciousness, we first consider the Kullback-Leibler (KL) divergence ([Kullback and Leibler, 1951](#))

$$D = \int P(\boldsymbol{\theta}|\mathcal{D}) \ln \left( \frac{P(\boldsymbol{\theta}|\mathcal{D})}{P(\boldsymbol{\theta})} \right) d\boldsymbol{\theta} = \left\langle \ln \left( \frac{P(\boldsymbol{\theta}|\mathcal{D})}{P(\boldsymbol{\theta})} \right) \right\rangle_{P(\boldsymbol{\theta}|\mathcal{D})}, \quad (5.5)$$

where  $\ln \left( \frac{P(\boldsymbol{\theta}|\mathcal{D})}{P(\boldsymbol{\theta})} \right)$  is the Shannon information ([Shannon, 1948](#)). In general, the KL divergence is a measure of the difference between two distributions, while in this case it quantifies the average information provided by the data when going from the prior to the posterior. Since the Shannon information is prior dependent, the KL divergence is also dependent on the prior. The suspiciousness  $S$  is then defined as

$$\ln S = \ln R - \ln I, \quad (5.6)$$

where  $I$  denotes the information ratio

$$\ln I = D_B - D_A. \quad (5.7)$$

For a multivariate Gaussian posterior distribution, [Handley and Lemos \(2019a\)](#) show that both  $\ln R$  and  $\ln I$  are dependent on the prior via an additive term of  $\ln V_\pi$ , where  $V_\pi$  is the volume of the prior. Thus, the dependence on the prior volume in Eq. (5.6) cancels out, so that the suspiciousness is prior independent. Therefore, the suspiciousness can be employed as an alternative metric to quantify the tension between datasets, which is unaffected by changes in the prior as long as they do not affect the shape of the posterior distribution.

In general, large positive values of the suspiciousness indicate agreement between datasets, while large negative values are associated with disagreement. Moreover, [Lemos et al. \(2020\)](#) show that under the assumption of a Gaussian posterior the quantity  $d - 2\log S$  follows a  $\chi^2$  distribution. Here,  $d$  denotes the Bayesian model dimensionality ([Spiegelhalter et al., 2002](#)) quantifying the effective number of constrained parameters. The tension probability  $p$  of the two datasets being this discordant by chance is determined via

$$p = \int_{d-2\log S}^{\infty} \chi_d^2(x) dx, \quad (5.8)$$

where  $\chi_d^2$  denotes the  $d$ -dimensional  $\chi^2$  distribution. While this is only true for Gaussian posteriors, [Handley and Lemos \(2019a\)](#) propose computing the Bayesian model dimensionality of the shared constrained parameters via

$$d = d_A + d_B - d_{AB}, \quad (5.9)$$

where  $d_A$  and  $d_B$  denote the individual model dimensionalities of the two datasets A and B, while  $d_{AB}$  denotes the model dimensionality of the combined dataset. The tension probability can further be cast into a corresponding  $\sigma$  level. A value of  $p < 0.05$  corresponds to a  $2\sigma$  Gaussian standard deviation and

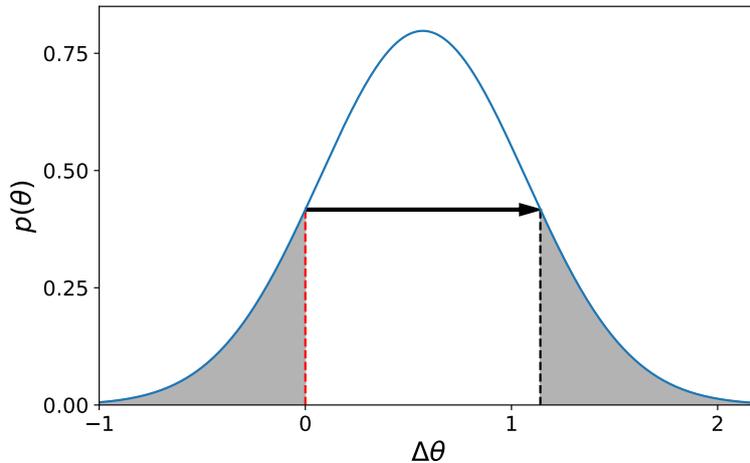
is considered as moderate tension, whereas  $p < 0.003$ , corresponding to  $3\sigma$ , is considered as strong tension. With this definition, the suspiciousness provides a quantitative measurement of the tension between two datasets. However, [Handley and Lemos \(2019a\)](#) point out that for non-Gaussian distributions only overly small values of  $p$  should be regarded with suspicion since  $p$  is only a rough estimate of the tension. In this case, the suspiciousness indicates tension if  $S \ll -\sqrt{d/2}$ , where  $p$  serves as a metric for the difference between the left and the right side of the inequality.

In this chapter, we perform consistency tests that rely on a split of the data vector into two parts which are each modelled independently with separate sets of cosmological parameters. This split of the data vector into two regions is dependent on how we probe the consistency of the given dataset. For example, if the data consists of measurements in several redshift bins, we can test the consistency between redshift bins by splitting the data vector along redshift bins. On the other hand, to probe the consistency of the cosmological model, the data vector can be divided into a region describing the homogeneous background and a region describing matter perturbations. We then test the following hypotheses:

1. There exists one set of cosmological parameters that describe the data.
2. There exist two sets of cosmological parameters that each describe one part of the dataset.

Hence, the first hypothesis refers to the fiducial cosmological analysis of the given dataset, whereas the second hypothesis corresponds to an analysis with a split data vector. By evaluating the aforementioned tension metrics we assess the question of whether or not the data shows a preference for the split model, indicating a tension between the two parts of the data vector.

As an alternative to quantifying the tension in terms of the Bayes factor or the suspiciousness, we consider the posterior distributions of the two sets of cosmological parameters in the second hypothesis. Sampling the likelihood



**Figure 5.1:** Sketch of the significance criterion for the tension between two duplicates of a parameter  $\theta$  modelling the observed data in two regimes. The blue curve shows the posterior distribution of the difference between both instances of the parameter. We evaluate the posterior at the origin, indicated by the vertical red dotted line, and infer the fraction of the distribution with a lower posterior density, illustrated by the grey shaded region. This fraction is identified with the probability mass of a one-dimensional Gaussian distribution outside of an interval  $[-m\sigma, m\sigma]$ , where  $m$  denotes the level of the tension in terms of  $\sigma$ .

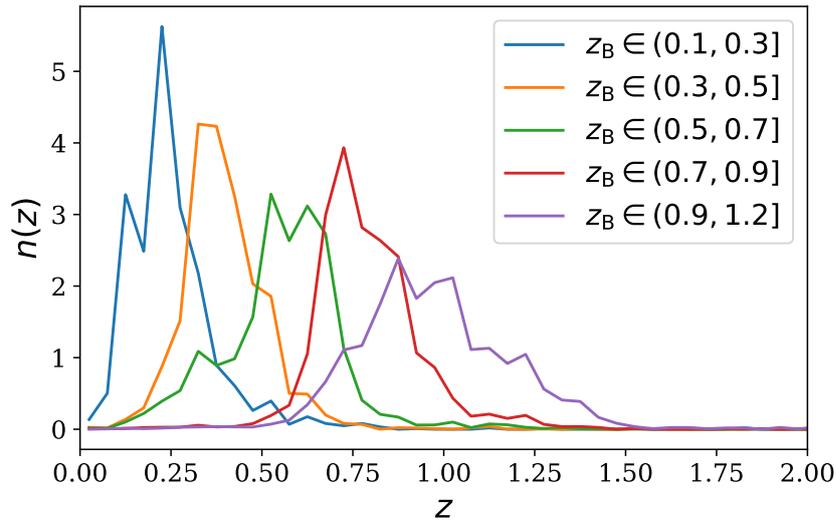
provides us with two posterior distributions for each cosmological parameter. In the absence of tension between the two regimes in which the data is split we expect both posteriors to constrain a similar region in parameter space. However, we expect the two contours to be broader than in the standard analysis, since splitting the data into subsets leads to a loss in constraining power in the two regions. Deviations between the posterior distributions of parameter duplicates are interpreted as a sign of tension within the dataset. Therefore, we quantify the deviation between the two posterior distributions to test whether the data prefers a different cosmology in the two regions.

We follow the methodology of [Köhlinger et al. \(2019\)](#) and derive the posterior distribution of the difference between parameter duplicates. If both parameters are consistent, we expect the posterior distribution of the difference to be centred around the origin. A deviation from the origin indicates an inconsistency between the two parts of the data vector. An example of such

a posterior distribution in one dimension is illustrated in Fig. 5.1, where the blue curve represents the posterior of the difference  $\Delta\theta$  between two instances of a single parameter  $\theta$ . We evaluate the posterior distribution at the origin, indicated by the vertical red dotted line, and determine the fraction of the distribution with lower density, illustrated by the grey shaded region. This fraction then serves as a metric for the tension between the two parameter instances. The lower the fraction the less likely it is that they are in agreement. To infer the  $m$  value of the tension we identify the fraction with the probability mass of an one-dimensional Gaussian distribution outside of the interval  $[-m\sigma, m\sigma]$ .

## 5.2 Internal consistency between tomographic bins in KiDS-1000

The most recent analysis of cosmic shear data from the fourth data release of KiDS (Kuijken et al., 2019) has provided some of the most stringent constraints on the structure growth parameter  $S_8$  to date. Given the importance of these measurements in the ongoing debate about tensions between early and late Universe probes, it is essential to perform a robust cosmological analysis by testing for systematic effects. Asgari et al. (2021) test the consistency of the cosmological analysis by comparing three kinds of two-point statistics that compress the observed cosmic shear data: Complete Orthogonal Sets of E/B-Integrals (COSEBIs; Schneider et al., 2010), band power estimates derived from correlation functions (Schneider et al., 2002; Becker and Rozo, 2016; van Uitert et al., 2018), and two-point correlation functions of cosmic shear. Additionally, the sensitivity of the analysis to the choice of astrophysical and nuisance parameters was tested. In this section we review the internal consistency tests of tomographic redshift bins that were performed as part of the weak lensing analysis of the KiDS-1000 dataset. These results were summarised in section 4.3 and Appendix B.2 of Asgari et al. (2021).



**Figure 5.2:** Redshift distribution of KiDS-1000 galaxies in five tomographic redshift bins selected via their best-fit photometric redshift  $z_B$ . Figure adapted from [Asgari et al. \(2021\)](#).

### 5.2.1 Methodology

The KiDS galaxy sample contains measurements of galaxy magnitudes in nine photometric bands from which photometric redshifts are inferred with the template-fitting code BPZ ([Benítez, 2000](#)). Based on their best-fit photometric redshift estimate  $z_B$ , the observed galaxies are divided into five tomographic bins between  $0.1 < z_B \leq 1.2$ . The underlying redshift distributions are calibrated using catalogues of deep spectroscopic reference galaxies that were re-weighted with a self-organising map ([Hildebrandt et al., 2021](#)). The redshift distributions and the boundaries of the five bins are displayed in Fig. 5.2. As can be seen in this Figure, the underlying redshift distributions of the five tomographic bins overlap significantly due to the imprecise photometric redshift estimates, as discussed in section 1.3.

The simplest possible consistency test between redshift bins can be performed by removing entire redshift bins from the analysis and determining the effect on the constraints on cosmological parameters. In this way, we determine to what extent each redshift bin contributes to the inferred value of  $S_8$ , which is the best constrained parameter in weak lensing analyses. The removal of

redshift bins results in a loss of constraining power causing an inflation of the uncertainty on  $S_8$ . [Asgari et al. \(2021\)](#) quantify the loss of constraining power in terms of the parameter  $\Sigma_8 = \sigma_8(\Omega_m/0.3)^\alpha$ , where  $\alpha = 0.5$  corresponds to the established definition of  $S_8$ . This choice was made because the degeneracy between  $\sigma_8$  and  $\Omega_m$  differs between summary statistics, depending on the angular scales entering the analysis. A loss in constraining power by less than 15% was found when excluding single redshift bins, except for the fifth bin which causes a 60% increase in uncertainty. Similarly, the exclusion of redshift bins causes a change in the numerical value of  $\Sigma_8$  by up to  $1.8\sigma$  when removing the fourth bin and up to  $0.5\sigma$  when removing the fifth bin.

The aforementioned consistency test does not necessarily indicate an inconsistency between redshift bins because the observed signal is expected to be strongest at higher redshifts. Since light emitted from more distant galaxies passes through more structure between the source and the observer, it is more affected by the lensing effect. Thus, the exclusion of high redshift bins from the analysis leads to a significant loss of signal. Therefore, it is desirable to perform a consistency test that does not require the removal of a significant portion of the observed signal. To do so, we compare the constraints on cosmological parameters preferred by the individual bins by simultaneously modelling the cosmic shear signal with two independent sets of cosmological parameters.

We follow the methodology outlined in Section 5.1 and split the data vector into two parts: one that contains cosmic shear measurements in one individual redshift bin and its cross-correlation with the remaining bins and one that contains cosmic shear measurements in the remaining bins and their cross-correlation. Both parts of the data vector are coupled through the data covariance matrix. The signal in both parts can then be modelled independently, for example using two instances of the cosmological code CLASS ([Blas et al., 2011](#)). We run five separate MCMC analyses in which we separate a single redshift bin at a time. This analysis is implemented in the 2COSMOS extension of the parameter estimation code MONTEPYTHON ([Köhlinger et al., 2019](#)).

Parameter		Prior	Dupl.
$\omega_{\text{cdm}}$	Reduced cold dark matter density	[0.051, 0.255]	yes
$\omega_{\text{b}}$	Reduced baryon density	[0.019, 0.026]	yes
$S_8$	Amplitude of matter density fluctuations	[0.1, 1.3]	yes
$h$	Reduced Hubble parameter	[0.64, 0.82]	yes
$n_s$	Spectral index of the prim. power spectrum	[0.84, 1.1]	yes
$A_{\text{IA}}$	Amplitude of intrinsic galaxy alignments	[-6, 6]	yes
$A_{\text{bary}}$	Baryon feedback parameter	[2, 3.13]	yes
$\delta_z$	Shift in the mean of KiDS $n(z)$ distributions	$\mathcal{N}(\boldsymbol{\mu}, C)$	no

**Table 5.1:** Sampling parameters and their priors, adopted from [Asgari et al. \(2021\)](#). The upper section lists cosmological parameters, while the lower section shows nuisance parameters. The third column reports whether the parameter is duplicated in the split analysis between redshift bins. The five shift parameters  $\delta_z$  are correlated through their covariance matrix  $C$  and their means  $\boldsymbol{\mu}$  are fixed to their mean values reported by [Hildebrandt et al. \(2021\)](#).

The sampling parameters and their priors are listed in Table 5.1. We compare results from the split analysis to the ones from the fiducial KiDS analysis with a single cosmological model using the metrics discussed in section 5.1.

In contrast to the analysis we presented in [Asgari et al. \(2021\)](#), we make a few modifications to the MCMC sampling. Previously, we made use of the MULTINEST sampler ([Feroz et al., 2009, 2019](#)). Nested samplers, such as MULTINEST, sample the posterior distribution with a set of so-called live points, which are repeatedly replaced with new points with a larger value of the posterior. After replacement of a live point it is referred to as a dead point. Since randomly generating points with higher posterior values is computationally inefficient, especially in high-dimensional data spaces or when the current set of live points is close to the maximum of the posterior, MULTINEST features an ellipsoidal sampling technique ([Mukherjee et al., 2006](#)). This method calculates a D-dimensional ellipsoid from the current set of live points and generates a new point within the ellipsoid, which is expanded by a factor that is inversely proportional to the efficiency. MULTINEST provides two distinct estimates of the Bayesian evidence. In addition to the traditional method of calculation the evidence from the set of accepted samples, it provides an alternative evidence

estimate based on the so-called importance nested sampling (INS) method. This method makes use of all points drawn during sampling, instead of discarding points that do not lie within a certain iso-likelihood contour. In principle, both estimates of the evidence should agree within their error bars. However, as shown by [Lemos et al. \(2022\)](#), ellipsoidal sampling can lead to biased estimates of both the posterior and the evidence if the efficiency is too high, so that the expansion factor of the ellipsoid is too small. In this case, the ellipsoid does not enclose the tails of the posterior distribution, which are then excluded from the sampling. This issue can be resolved by tuning the efficiency hyperparameter at the cost of a longer runtime of the sampling algorithm. Additionally, the optimal value of the efficiency is unknown a priori, so that tuning the parameter requires running multiple chains. In practice we find a deviation of the two evidence estimates when sampling the posterior distribution with the fiducial MULTINEST settings of [Asgari et al. \(2021\)](#). Since the Bayes factor, defined in Eq. (5.3), requires an estimate of the evidence, it becomes even more unreliable for model comparison because it is not clear which estimate is closer to the truth. Therefore, in [Asgari et al. \(2021\)](#) we reported the Bayes factors for both estimates.

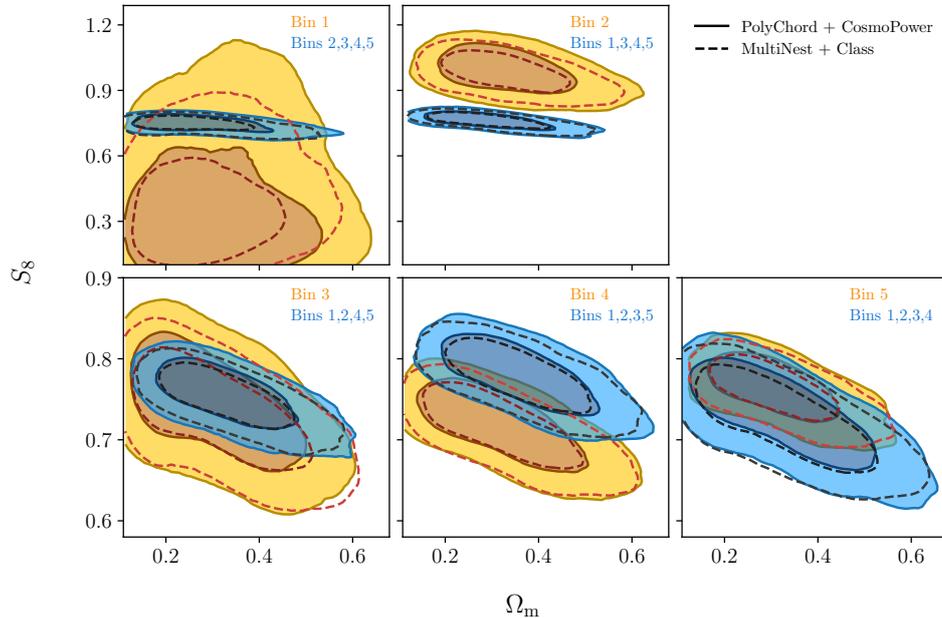
Here, we circumvent this issue by using the POLYCHORD sampler ([Handley et al., 2015a,b](#)), which is an alternative nested sampling code featuring a slice-sampling method ([Neal, 2003; Aitken and Akman, 2013](#)) for the generation of new live points. This method generates new live points from random slices through the parameter space including the current live point. [Lemos et al. \(2022\)](#) find a minimal bias in the evidence estimate from POLYCHORD and show that evidence estimates are more robust with respect to changes in the hyperparameters. However, we find that in our application the runtime of the POLYCHORD sampler is several times slower than the MULTINEST sampler. Therefore, we replace the CLASS calculation of the matter power spectrum with the COSMOPOWER ([Spurio Mancini et al., 2022](#)) emulator, which significantly reduces the computational cost of the MCMC sampler, providing a speed-up

by a factor of up to  $\mathcal{O}(10^4)$ . The emulator is trained on matter power spectra computed with CLASS including non-linear corrections from HMCODE (Mead et al., 2015). The training samples are generated on a grid within the prior ranges reported in Table 5.1 for the five cosmological parameters listed in the upper section and the baryon feedback parameter. We find errors in the emulation of the matter power spectrum of less than 0.01% for redshifts between  $0 \leq z < 5$  and wavenumbers between  $10^{-5} \text{Mpc}^{-1} \leq k < 10 \text{Mpc}^{-1}$ .

### 5.2.2 Results

In Fig. 5.3 we present the marginal posterior distribution from the split analysis using band power spectra in the subspace spanned by  $S_8$  and  $\Omega_m$ . In each panel the orange contour represents constraints from a single redshift bin while the blue contour shows the constraints from the remaining redshift bins. Visually, we find consistency between both sets of cosmological parameters in all bins except for the second one, where there is no overlap between the two contours. Additionally, the fourth bin prefers slightly lower  $S_8$  values, as can be observed from the shift in the orange posterior in the bottom middle panel which, however, still overlaps with the constraints from the remaining bins. Moreover, the dotted contours show the posterior distributions obtained with the MULTINEST sampler with calculation of the matter power spectrum with CLASS, which were analysed in Asgari et al. (2021). We find good consistency between the two samplers with slightly broader contours from the POLYCHORD sampler. This is consistent with Lemos et al. (2022), who found that MULTINEST systematically reports smaller credible intervals than POLYCHORD.

We quantify the significance of the split cosmological model by calculating the Bayes factor and the suspiciousness in comparison to the standard  $\Lambda$ CDM analysis. The results are reported in Table 5.2. The leftmost column provides the evidence  $\mathcal{Z}_{\text{split}}$  of the split analysis, while the next column shows the Bayes factor  $\log R = \log \mathcal{Z}_{\text{fid}} - \log \mathcal{Z}_{\text{split}}$  for the comparison between the fiducial and the split cosmological model. We find preference for the fiducial cosmological



**Figure 5.3:** Marginalised posterior distributions in the  $S_8$ - $\Omega_m$  plane of the consistency test between tomographic redshift bins. The test duplicates cosmological parameters with one set of parameters modelling the signal in one single bin while the other set models the signal in the remaining bins. The covariance between bins is accounted for via the data covariance matrix. The orange contours refer to the cosmological parameters of the isolated bin and the blue contours show the constraints from the remaining bins. The solid contours show the results from sampling the parameter space with the POLYCHORD sampler with emulation of the matter power spectrum with COSMOPOWER, whereas the dotted contours display posteriors obtained with the MULTINEST sampler and calculation of the matter power spectrum with CLASS.

model in all redshift bins except the second one, which shows strong preference for the split model when interpreting with Jeffreys' scale. This is in agreement with the visual inspection of the marginalised posterior distribution in Fig. 5.3, which indicated an inconsistency between the posteriors in the second bin. The remaining columns show the suspiciousness, the Bayesian model dimensionality of the split model, and the tension probability and the corresponding  $\sigma$  value of the tension between redshift bins. Here, the tension probability is calculated from Eq. (5.8) with  $d_{\text{diff}} = d_{\text{split}} - d_{\text{fid}}$ , where  $d_{\text{split}}$  and  $d_{\text{fid}}$  denote the Bayesian model dimensionalities of the split and the fiducial analysis, respectively. In

this analysis, the effective number of constrained parameters is smaller than the number of model parameters due to the usage of informative priors and parameter degeneracies.

As shown by [Joachimi et al. \(2021\)](#), popular methods of determining the effective number of parameters (see for example [Spiegelhalter et al., 2002](#); [Handley and Lemos, 2019b](#)) give biased estimates of the effective number of parameters and therefore we require a more robust estimate that can be obtained by running computationally expensive chains with a large number of realisations of the cosmic shear data vector. Thus, the tension probabilities reported in Table 5.2 must be interpreted with caution. Furthermore, the split analysis of the first redshift bins results in a Bayesian model dimensionality that is lower than the model dimensionality of the fiducial analysis, so that  $d_{\text{diff}}$  becomes negative. Consequently, we cannot calculate the tension probability for this bin. Most likely, the reason for this is that the cosmological parameters are mostly unconstrained by the first redshift bin, because of the low cosmic shear signal at low redshifts. This is illustrated by the orange contour in the top left panel of Fig. 5.3, which mostly corresponds to the prior distribution of  $S_8$  and  $\Omega_m$ . Thus, the limited constraining power of this bin and the shortcomings of our estimation method of the Bayesian model dimensionality result in an unrealistic estimate of  $d_{\text{diff}}$ .

To further assess the tension between the contours of the split analysis in Fig 5.3, we derive the posterior distribution of the difference between duplicated cosmological parameters. The resulting contours are shown in Fig. 5.4 for  $S_8$  and  $\Omega_m$ . Here, a value of  $\Delta S_8 = 0$  and  $\Delta \Omega_m = 0$  corresponds to consistency between the two parameter instances. Any deviation of the contour from the zero point can then be interpreted as a sign of tension. As outlined in Section 5.1, we quantify the tension between the two parameter instances by determining the fraction of the posterior with lower density than the value at the origin. We calculate the tension between the one-dimensional posterior distributions of  $S_8$ ,  $\Omega_m$ , and  $A_{\text{IA}}$ . Additionally, we determine the tension for

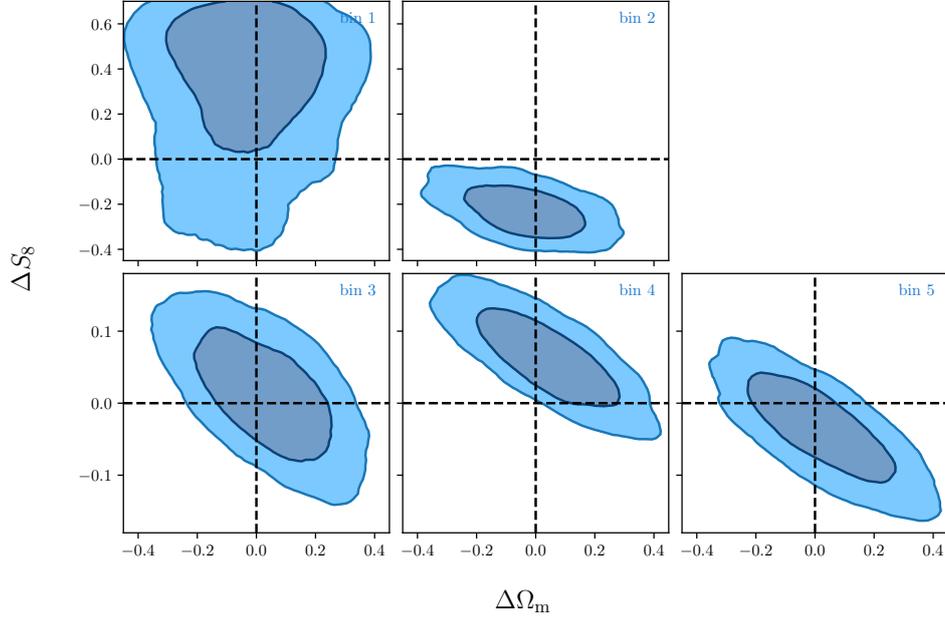
split	$-\ln \mathcal{Z}$	$\log_{10} R$	$\ln S$	$d$	$p$	$N_\sigma$
z-bin 1	$85.1 \pm 0.1$	$0.5 \pm 0.1$	$0.7 \pm 0.1$	$5.2 \pm 0.2$	–	–
z-bin 2	$80.3 \pm 0.2$	$-1.6 \pm 0.1$	$-8.0 \pm 0.1$	$7.5 \pm 0.2$	$(8.2 \pm 3.3) \cdot 10^{-5}$	$4.0 \pm 0.1$
z-bin 3	$88.9 \pm 0.2$	$2.2 \pm 0.1$	$0.7 \pm 0.11$	$8.4 \pm 0.3$	$0.691 \pm 0.069$	$0.4 \pm 0.1$
z-bin 4	$85.3 \pm 0.2$	$0.6 \pm 0.1$	$-3.2 \pm 0.11$	$7.4 \pm 0.2$	$0.011 \pm 0.003$	$2.6 \pm 0.1$
z-bin 5	$88.2 \pm 0.2$	$1.9 \pm 0.1$	$-0.1 \pm 0.11$	$7.7 \pm 0.2$	$0.325 \pm 0.044$	$1.0 \pm 0.1$
fid.	$83.9 \pm 0.1$	–	–	$5.9 \pm 0.2$	–	–

**Table 5.2:** Evidence, Bayes factor, suspiciousness, Bayesian model dimensionality, tension probability, and tension level for a split cosmological analysis of cosmic shear data from KiDS-1000. The split separates individual tomographic redshift bins that are modelled with a separate set of cosmological parameters with cross-correlations taken into account via the data covariance matrix. The bottom row provides the evidence and the Bayesian model dimensionality of the fiducial analysis with one set of cosmological parameters.

two combinations of the three parameters. We include the intrinsic alignment parameter  $A_{IA}$  since this parameter can to some extent mitigate systematic effects, as seen in section 2.5.2 (see also [van Uitert et al., 2018](#); [Efstathiou and Lemos, 2018](#)).

The resulting tension levels are reported in Table 5.3. We find that the tension is strongest in the second redshift bin, reaching  $2.7\sigma$  and  $1.4\sigma$  for  $\Delta(S_8)$  and  $\Delta(A_{IA})$ , respectively. When combining the two parameters we find  $2.5\sigma$ , while adding  $\Delta(\Omega_m)$  increases the tension even further. However,  $\Omega_m$  is relatively poorly constrained by the data, so that the increased tension might just be caused by additional noise fluctuations. This test further corroborates our previous findings of an internal inconsistency between the second redshift bin and the remaining bins. Additionally, we find that the slight shift in  $S_8$  in the fourth bin, observed in the bottom middle panel of Fig. 5.3, corresponds to a  $1.5\sigma$  shift in  $S_8$ .

The apparent tension in the second bin can potentially be explained by a contamination with a small population of high-redshift galaxies in this bin. Such a galaxy population would induce a signal due to more structure being present between the source and the observer. If this population remains undetected when calibrating the redshift distribution, the theoretical prediction



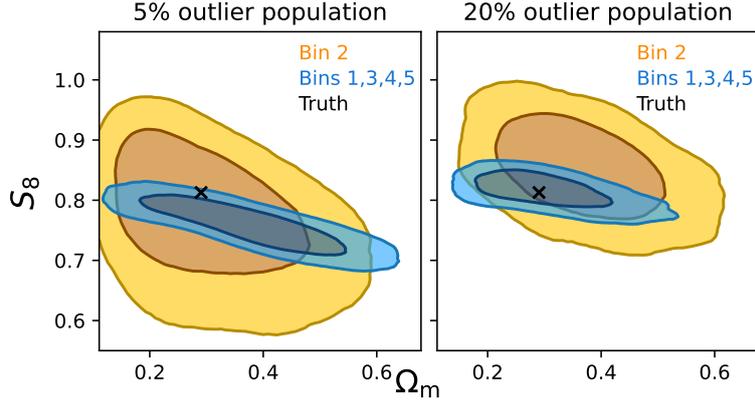
**Figure 5.4:** Marginalised posterior distributions of the difference between parameter duplicates in the consistency test between tomographic redshift bins, shown in Figure 5.3. Each panel presents constraints from an analysis that isolates a single redshift bin.

split	$\Delta(\Omega_m)$	$\Delta(S_8)$	$\Delta(A_{IA})$	$\Delta(S_8, A_{IA})$	$\Delta(\Omega_m, S_8, A_{IA})$
z-bin 1	$0.1\sigma$	$1.6\sigma$	$0.4\sigma$	$0.7\sigma$	$0.4\sigma$
z-bin 2	$0.3\sigma$	$2.7\sigma$	$1.4\sigma$	$2.5\sigma$	$2.9\sigma$
z-bin 3	$0.3\sigma$	$0.0\sigma$	$0.2\sigma$	$0.0\sigma$	$0.0\sigma$
z-bin 4	$0.3\sigma$	$1.5\sigma$	$1.5\sigma$	$1.4\sigma$	$2.0\sigma$
z-bin 5	$0.1\sigma$	$0.6\sigma$	$1.1\sigma$	$0.8\sigma$	$0.4\sigma$

**Table 5.3:** Significance of the tension in a split cosmological analysis of cosmic shear data from KiDS-1000 that separates individual redshift bins. Each part of the data vector is modelled with a separate set of cosmological parameters from which we infer the posterior distribution of the difference between parameter duplicates. The level  $m$  of the tension is quantified by determining the fraction of the posterior with lower density than the density at the origin which is identified with the probability mass of an one-dimensional Gaussian distribution outside an  $m\sigma$  interval.

for the observed signal would be systematically lower. This effect can then be compensated with a higher amplitude of the matter power spectrum by increasing  $S_8$ . To test the effect of such a signal on the posterior distribution of  $S_8$ , we perform an analysis with a mock data vector in which we artificially insert a signal at high redshift. Using a Gaussian distribution centred at  $z = 1.6$  with a width of  $\sigma = 0.05$  we generate two mock data vectors: one that contains 5% of the total number of galaxies at high redshifts in the second bin and one that contains 20% of the total number of galaxies. The former data vector corresponds to a population of galaxies that could be undetected in a realistic application, whereas the latter data vector serves as a cross-check of how a unrealistically large population of high-redshift galaxies would bias the cosmic shear constraints.

The resulting posterior distributions are illustrated in Fig. 5.5, where the left and right panels corresponds to the mock data vector with 5% and 20% of galaxies at high redshifts, respectively. Both contours show a significant difference to the one obtained from real observed data. The bottom middle panel in Fig. 5.3 shows no overlap between the two posterior distributions. However, the bin 2 constraints from mock data overlap completely with the contour originating from the remaining redshift bins. Only when considering an unrealistically large outlier population we observe a slight shift of the posterior distribution of the second bin towards higher values of  $S_8$ . However, this shift is by no means compatible with our findings from real observed data. Obtaining a contour similar to the real observation we would require an even higher outlier population, which is highly unlikely to be misidentified in the redshift calibration. Thus, we conclude that is unlikely that the observed discrepancy in the second bin is caused by an outlier population of galaxies and the source of this tension remains unclear. However, [Asgari et al. \(2021\)](#) found that excluding the first or the second bin has a negligible impact on the final results of the KiDS-1000 cosmic shear analysis, which is why it is not excluded from the fiducial analysis.



**Figure 5.5:** Marginalised posterior distributions in the  $S_8$ - $\Omega_m$  plane in a consistency test with mock data vectors containing an undetected population of galaxies at high redshifts. The left and right panels corresponds to a mock data vector with 5% and 20% of galaxies at high redshifts, respectively. The cross indicates the input cosmological parameters.

### 5.3 Consistency of background and perturbations in the $\Lambda$ CDM model

Recent analyses of cosmological data have indicated tensions in the standard  $\Lambda$ CDM cosmological model. As discussed in section 1.4, measurements of the Hubble parameter  $H_0$  through observations of the cosmic distance ladder (Riess et al., 2021) and CMB predictions (Planck Collaboration et al., 2020b) and measurements of the structure growth parameter  $S_8$  with galaxy surveys (Heymans et al., 2021) and CMB data point towards a tension between probes of the early and late Universe. This tension can have multiple origins. First, if we assume that our cosmological model is complete, the tension might originate from the data itself or its analysis. For example, unknown systematic effects or shortcomings in the modelling of the observables can bias the constraints on cosmological parameters. Second, if there are no systematic biases in the observables, the tension might arise from an incomplete cosmological model that lacks a description of important physical effects beyond  $\Lambda$ CDM. Therefore, it is essential to test each hypothesis. In this section we focus on the possibility of an incomplete cosmological model leading to a discrepancy between the aforementioned probes. This analysis was published in Ruiz-Zapatero et al.

(2021) and is summarised in this section.

### 5.3.1 Split cosmological model

Our model of the Universe consists of a homogeneous background (see section 1.1) that describes the Universe on large scales and perturbations that give rise to matter density fluctuations causing the growth of large scale structure (see section 1.2). By following the methodology of splitting the cosmological model into two regimes that are modelled with independent sets of cosmological parameters, we test the internal consistency of the  $\Lambda$ CDM model. Here, we divide the cosmological model into the homogeneous background, referred to as ‘geometry’, and perturbations which we refer to as ‘growth’. Similar to the study in the previous section we employ the tension metrics discussed in section 5.1 to assess the consistency between the two theory regimes.

We employ several observables: weak gravitational lensing (WL), baryon acoustic oscillations (BAOs), redshift space distortions (RSDs), and probes of the early Universe. In the following we discuss how each observable probes either the geometry or growth regime or a combination of both. A summary is given in Table 5.4.

**Weak lensing:** Weak gravitational lensing probes the matter power spectrum and thus we attribute the power spectrum in Eq. (1.69) to the growth regime. On the other hand, the lensing efficiency, defined in Eq. (1.71), is dependent on comoving distances (under the assumption of a flat Universe). Since the comoving distance is purely related to the homogeneous background, we assign the lensing efficiency to the geometry regime. Finally, the prefactor in Eq. (1.71), originating from the Poisson equation, is assigned to the geometry regime. As probe of weak lensing we again employ cosmic shear measurements from KiDS-1000 (Asgari et al., 2021).

**Baryon acoustic oscillations:** BAOs are fluctuations in the density of baryonic matter and are thus classified as perturbations. However, our main BAO observable, namely the position of the BAO peak, is solely related to the homogeneous background, since the size of the sound horizon, which is the

comoving distance a sound wave could travel between the beginning of the Universe and the decoupling of photons and baryons, is dependent on the matter content of the Universe. Therefore, the BAO observable is assigned to the geometry regime. We employ BAO measurements from the 6dF galaxy survey (Jones et al., 2009) and BOSS DR12 (Alam et al., 2017), which we refer to as ‘clustering’. Additionally, we draw constraints on BAOs from observations of the Lyman- $\alpha$  forest and quasars from eBOSS DR14 (Blomqvist et al., 2019; de Sainte Agathe et al., 2019). These datasets are referred to as ‘Lyman- $\alpha$ ’.

**Redshift space distortions:** RSDs are caused by peculiar velocities of galaxies along the line of sight in addition to the cosmological expansion, leading to a distortion of the spatial distribution of galaxies when viewed as a function of redshift. Since the peculiar motion of galaxies is dependent on the surrounding matter distribution, RSDs are related to the growth theory regime. We employ measurements from BOSS DR12 (Alam et al., 2017) as probe of RSDs which are included in the ‘clustering’ data set.

**Early Universe:** Instead of a full reanalysis of CMB data from Planck, which would be beyond the scope of this work, we employ measurements of  $A_s$ ,  $n_s$ , and  $\theta^*$  as a multivariate Gaussian likelihood obtained by marginalising the Planck TT,TE,EE+lowl+lowE posterior distributions (Planck Collaboration et al., 2020b) over the remaining parameters. While  $A_s$  and  $n_s$  describe the amplitude and the spectral index of the primordial power spectrum and thus are attributed to the growth regime, the parameter  $\theta^*$  provides a measurement of the BAO peak at the time of recombination and is thus assigned to the geometry regime. The two remaining cosmological parameters, namely the reduced baryon density  $\omega_b$  and the reduced cold dark matter density  $\omega_{\text{cdm}}$ , cannot be readily categorised as either geometry or growth since they are both related to density fluctuations as well as the expansion history. Thus, we refrain from including CMB constraints on the matter densities in this analysis. We refer to this dataset as ‘recombination’.

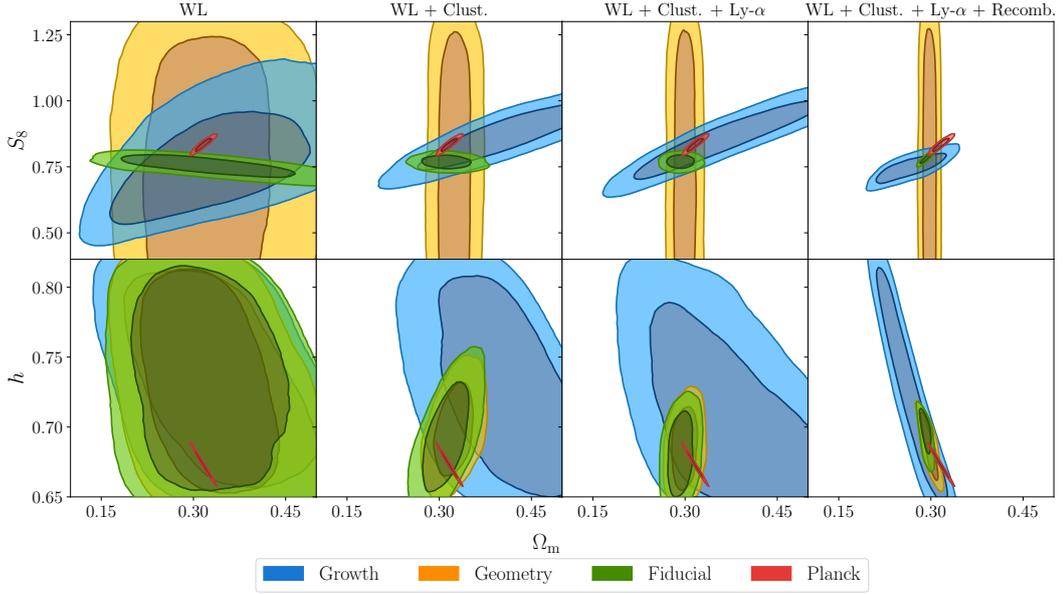
geometry	growth
BAO angular scale	RSD growth rate
WL efficiency	WL matter power spectrum
CMB first acoustic peak position	CMB primordial power spectrum

**Table 5.4:** Overview of the observables used in the consistency test between background and perturbations in the  $\Lambda$ CDM model and their classification into the two regimes.

### 5.3.2 Results

We sample the parameter space using the 2COSMOS extension of MONTEPYTHON (Köhlinger et al., 2019) with matter power spectra inferred from CLASS (Blas et al., 2011), using the MULTINEST sampling method (Feroz et al., 2009, 2019). The sampling parameters and their priors are listed in Table 5.1. The constraints of the split cosmological analysis on  $S_8$ ,  $\Omega_m$ , and  $h$  are illustrated in Fig. 5.6. The orange and blue contours show the marginalised posterior distributions from the geometry and growth regimes, respectively, while the green contour illustrates the constraint from the fiducial analysis with one set of  $\Lambda$ CDM parameters. Additionally, the red contours provide the posterior distribution obtained by Planck Planck Collaboration et al. (2020b) as a comparison. The leftmost column shows the constraints from an analysis of weak lensing data only, while the remaining columns show the constraints when subsequently adding data from clustering, Lyman- $\alpha$ , and recombination.

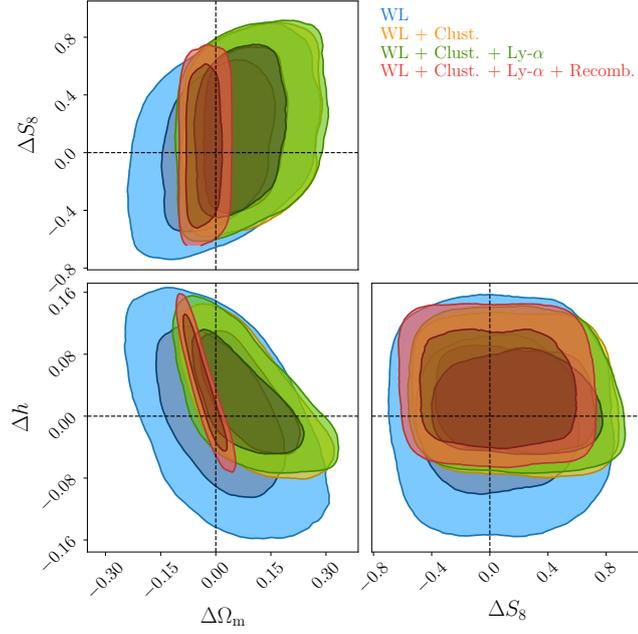
We find that for the various cosmological parameters in the two regimes the constraining power differs among the probes. From the top row of Fig. 5.6 we find that weak lensing only constrains  $\Omega_m$  very weakly. Clustering, Lyman- $\alpha$ , and recombination data provide stronger constraints on  $\Omega_m$  in both regimes, although the constraints on the geometry regime are a lot stronger. This is as expected, since BAO measurements from galaxy clustering and Lyman- $\alpha$  solely probe the homogeneous background. Only RSD measurements from galaxy clustering allow constraining the growth regime, which remains mostly unchanged when adding Lyman- $\alpha$  data since this dataset is not sensitive to the growth regime. Adding recombination data then puts the strongest constraints



**Figure 5.6:** Marginalised posterior distributions for  $S_8$ ,  $\Omega_m$ , and  $h$  in a split cosmological analysis in geometry (orange contours) and growth (blue contours) theory regimes and the fiducial analysis with one set of  $\Lambda$ CDM parameters (green contours). Additionally, the red contours show constraints from Planck (Planck Collaboration et al., 2020b). The leftmost column shows constraints from weak lensing data only, while the remaining columns show the constraints when adding clustering, Lyman- $\alpha$ , and recombination data, respectively.

on the matter density in both regimes. Concerning  $S_8$ , we find that none of the probes constrains the geometry regime, so that the  $S_8$  posterior is driven by the growth regime. However, considering the fiducial analysis, we find even stronger constraints on  $S_8$ , which are most likely driven by the interplay between  $\Omega_m$  and  $S_8$  in the combined cosmological model. Overall, we visually find good consistency between both theory regimes with the posterior distribution from the fiducial analysis residing at the intersection between both contours.

The bottom row of Fig. 5.6 presents the marginalized posterior distribution for  $h$  and  $\Omega_m$ . While the weak lensing observable is not capable of constraining either  $h$  or  $\Omega_m$  significantly, the addition of clustering, Lyman- $\alpha$ , and recombination data allows for constraining  $h$  in the geometry regime, so that the fiducial contour is almost exclusively driven by the geometry regime. Again, we visually find a good agreement between geometry, growth, and the



**Figure 5.7:** Marginalised posterior distributions of the difference between parameter duplicates in the split cosmological analysis in geometry and growth theory regimes, shown in Fig. 5.6. The corresponding tension levels are given in Table 5.5.

fiducial analysis.

To quantify the consistency between geometry and growth regimes we consider the difference between parameter duplicates for  $S_8$ ,  $\Omega_m$ , and  $h$ . The corresponding posterior distributions are presented in Fig. 5.7. Visually, the two regimes seem in good agreement since the contours are centred around the origin, which confirms the observation from Fig. 5.6. The associated tension levels are given in Table 5.5. We find a maximum tension of  $1.3\sigma$  in the  $h - \Omega_m$  plane when combining data from weak lensing and galaxy clustering. All remaining combinations of the three parameters show a lower tension. Thus, we conclude that the data does not favour a split cosmological model and instead shows good agreement with the standard cosmological model.

Comparing the posterior distributions of geometry and growth, shown in Fig. 5.6, to the Planck posteriors, we find that individually the two regimes are in agreement with Planck. However, when combining both regimes in the fiducial analysis we find that Planck prefers a higher  $S_8$  value than the probes

### 5.3. Consistency of background and perturbations in the $\Lambda$ CDM model 154

dataset	$\Delta(\Omega_m)$	$\Delta(S_8)$	$\Delta(h)$	$\Delta(h, \Omega_m)$	$\Delta(S_8, \Omega_m, h)$
WL	$0.0\sigma$	$0.1\sigma$	$0.3\sigma$	$0.0\sigma$	$0.0\sigma$
WL + Clust.	$0.8\sigma$	$0.2\sigma$	$0.2\sigma$	$1.3\sigma$	$1.1\sigma$
WL + Clust. + Ly- $\alpha$	$0.5\sigma$	$0.6\sigma$	$0.1\sigma$	$0.7\sigma$	$0.6\sigma$
WL + Clust. + Ly- $\alpha$ + Recomb.	$1.0\sigma$	$0.5\sigma$	$0.1\sigma$	$0.4\sigma$	$0.2\sigma$

**Table 5.5:** Significance of the tension between the geometry and growth theory regime in a split cosmological analysis. Each regime is modelled with a separate set of cosmological parameters from which we infer the posterior distribution of the difference between parameter duplicates. The level  $m$  of the tension is quantified by determining the fraction of the posterior with lower density than the density at the origin which is identified with the probability mass of an one-dimensional Gaussian distribution outside an  $m\sigma$  interval.

employed in this work. This is consistent with earlier studies of the KiDS-1000 dataset (Asgari et al., 2021; Heymans et al., 2021; Tröster et al., 2021). Thus, we conclude that the tension cannot be attributed to the geometry or growth regime. We note that while studies of KiDS-1000 data have found a tension of up to  $3\sigma$ , the remaining probes employed in this work have been shown to be in agreement with Planck (see for example Sánchez et al., 2017; Handley and Lemos, 2019a; Blomqvist et al., 2019). Therefore, we conclude that the tension observed between  $S_8$  constraints from weak lensing and CMB experiments cannot be explained by a discrepancy between the homogeneous background and perturbations in the flat  $\Lambda$ CDM model since our analysis shows a good consistency between both regimes.

## Chapter 6

# Conclusions

Photometric redshift distributions play a crucial role for precision cosmology. In this thesis we developed a method to accurately propagate residual uncertainties in photometric redshift distributions into the cosmological inference with weak lensing data from KiDS. Additionally, we presented an optimisation method for the assignment of photometric galaxies observed by LSST into a chosen set of tomographic redshift bins, which is not only crucial for weak lensing analyses, but also for studies of galaxy clustering. Finally, we performed consistency tests between redshift bins in a weak lensing analysis and consistency tests between probes of geometry and growth in the  $\Lambda$ CDM model.

In chapter 2, we propagated uncertainties in photometric redshift distributions into the weak lensing analysis of KV450. Weak lensing studies measure the distortion of galaxy images due to the deflection of light to infer the distribution of matter in the Universe. To model the observed signal they require an accurate calibration of the redshift distribution of source galaxies and are thus sensitive to residual uncertainties in the redshift distribution. These uncertainties are usually accounted for via nuisance parameters that model a shift in the mean of the redshift distribution. While this method is sufficient for stage III surveys such as KiDS, we expect variations in the shape of the redshift distribution to become significant for stage IV surveys such as LSST. Additionally, galaxy clustering is more sensitive to the shape of the redshift distribution, so that such analyses benefit from a more complex modelling of

the uncertainties on the redshift distribution.

Instead of modelling a shift in the mean of the redshift distribution, we parameterised the redshift distributions of tomographic redshift bins using a flexible Gaussian mixture model which allows for the parametrisation of arbitrary changes in the shape of the redshift distribution via a set of amplitudes of the Gaussian components. We performed a self-calibration of the redshift distributions via cosmic shear measurements by iteratively fitting cosmological and redshift nuisance parameters arising from the redshift distribution model. We then implemented an analytic marginalisation method over the nuisance parameters which recovered the cosmological posterior. This method is particularly useful for upcoming surveys such as LSST, where we expect higher order moments of the redshift distribution to become more important. Using KV450 data, we found slight shifts in the posterior distributions of the model parameters, which were strongest for the amplitude of intrinsic galaxy alignments. This effect is likely caused by degeneracies between the redshift distribution amplitudes and the intrinsic alignment amplitude, which allows for the mitigation of systematic effects, and therefore we require a tighter prior on intrinsic alignments through external constraints. This result is consistent with earlier studies that also found discrepant values of the intrinsic alignment amplitude (Wright et al., 2020b; Hildebrandt et al., 2020; Fortuna et al., 2021; van Uitert et al., 2018; Efstathiou and Lemos, 2018).

In chapter 3, we optimised the assignment of photometrically observed galaxies into tomographic redshift bins. Modern extragalactic galaxy surveys often rely on photometric observations through a set of filters to infer estimates of the redshifts of individual galaxies. This technique allows for the identification of broad features in the spectral galaxy distribution, from which the redshift is estimated. These estimates of the photometric redshift are commonly employed to assign galaxies to a chosen set of redshift bins. Although this technique allows us to observe large samples of galaxies in a short time, we find catastrophic outliers, where the true redshift of an object is misestimated by a large amount.

Additionally, the limited precision of photometric redshift estimates causes the true underlying redshift distribution of galaxies in each tomographic bin to spread outside the actual bin boundaries.

To reduce the rate of outliers in tomographic redshift bins inferred from photometric redshift estimates, we developed an optimisation method that moves galaxies between redshift bins. We optimised the clustering cross-correlation signal between a sample of photometric galaxies and a reference sample of galaxies with well-calibrated redshifts. This was applied to a simulated data set, cosmoDC2, that mimics the upcoming LSST. Using a reference sample that is representative of the photometric sample, we found a significant decrease in the fraction of outliers in each tomographic bin, which was most prominent at high redshifts.

An important extension of this work is to study how a realistic sample of reference galaxies impacts the assignment of photometric galaxies to tomographic bins. In a realistic application, the sample of galaxies will be provided by a complementary spectroscopic survey that overlaps on sky with the photometric survey. For LSST, such observations can be provided by DESI. However, given that spectroscopic observations are significantly more time-consuming than photometric observations, leading to a smaller sample size of reference galaxies, and because of different observing strategies of the two surveys, the reference sample will not be representative of the photometric sample.

We studied the impact of a non-representative reference sample in chapter 4, where we constructed a DESI-like sample from the cosmoDC2 simulation using selection functions that mimic a galaxy sample observed by DESI. Using four samples of objects that will be observed by DESI, Luminous Red Galaxies, Emission Line Galaxies, high redshift quasars, and a magnitude-limited sample, which were selected from the cosmoDC2 catalogue, we performed the optimisation of redshift bin assignments with a SOM-based method of calculating the clustering cross-correlation between photometric and reference bins. We found a similar performance of the optimisation to what was achieved with

the representative reference sample, although the DESI-like selected sample requires broader bins at high redshifts given the sparsity of observed galaxies and quasars in this redshift range. A future direction of this work is to examine the optimisation method by applying it to real observed data. For this purpose, KiDS provides an ideal dataset since it was already used extensively in studies of the calibration of photometric redshift distributions, which will allow us to assess how the optimised assignment of photometric galaxies to tomographic bins influences the constraints on cosmological parameters. A challenge for the analysis of real data from KiDS are the cross-correlation measurements with a reference dataset which is comprised of data from various spectroscopic surveys covering different regions of the KiDS footprint. We expect the galaxy bias of reference galaxies to be vastly different for each spectroscopic sample, which impacts the metric used to determine the quality of the redshift bin assignments. Therefore, we potentially require to further modify the optimisation algorithm, for example by adjusting the weighting between the diagonal and off-diagonal elements of the cross-correlation matrix.

In chapter 5, we performed consistency tests in the cosmological analysis of KiDS-1000. First, we tested the internal consistency between tomographic redshift bins. This test is of particular importance in order to assess the robustness of the analysis with respect to systematic effects. We performed this test by modelling the cosmic shear signal in specific redshift bins with a second independent set of cosmological parameters. This allowed us to assess the consistency between the constraints on cosmological parameters from individual redshift bins compared to the constraints from the remaining redshift bins. We found good consistency between all redshift bins except for the second one, which shows preference for a higher value of the amplitude of matter density fluctuations,  $S_8$ . While the source of this discrepancy remains unclear and is under active investigation, [Asgari et al. \(2021\)](#) found that this redshift bin has a negligible impact on the final result of the KiDS-1000 cosmic shear analysis.

Secondly, we performed a consistency test of the  $\Lambda$ CDM model. This was achieved by splitting the model into the homogeneous background and perturbations that give rise to matter density fluctuations causing the growth of structure in the Universe. These two regimes were modelled with two independent sets of cosmological parameters. By analysing the consistency between the posterior distribution in the two regimes we tested if the data prefers a different evolution of the universe for the homogeneous background and perturbations, which would indicate an internal tension in the cosmological model. We found good consistency between both sets of parameters, showing that the two regimes are in good agreement. This result is consistent with the recent analysis of galaxy clustering and weak lensing data from DES by [Muir et al. \(2021\)](#), who split  $\Omega_m$  between geometry and growth theory regimes and found no significant disagreement between the two regimes.

The calibration and modelling of photometric redshift distributions play a crucial role for upcoming galaxy surveys, such as LSST and Euclid. Thus, exploiting synergies between the two surveys will greatly benefit cosmological analyses requiring accurate estimates of the redshift of observed objects. In particular, combining observations of galaxies through the six LSST filters with the three near-infrared filters of Euclid will improve the accuracy of the inferred photometric redshifts and reduce the scatter and outlier rate ([Rhodes et al., 2017](#)). A focus of upcoming surveys is a so-called 3x2pt analysis, which combines cosmological analyses with two-point correlation functions of galaxy clustering, weak lensing, and galaxy-galaxy lensing and will provide constraints on the structure growth parameter  $S_8$  with unprecedented precision. Thus, we require the characterisation of the true underlying redshift distribution of observed objects. This can be achieved for example through angular cross-correlation measurements with overlapping reference galaxies with spectroscopic redshift measurements from DESI. Therefore, the combination of datasets from various surveys is one of the most important tasks in future analyses of cosmological data.

The optimisation technique developed in chapter 3 and 4 provides a method of further decreasing the rate of catastrophic outliers after the initial assignment of galaxies to tomographic redshift bins based on photometric measurements. Therefore, it will benefit the 3x2pt analysis which requires compact redshift distributions. Additionally, the marginalisation method over residual redshift distribution uncertainties from the calibration with spectroscopic reference catalogues, developed in chapter 2, can be generalised to the analysis of galaxy clustering and galaxy-galaxy lensing and is therefore applicable in an upcoming 3x2pt analysis of LSST data.

## Appendix A

# Appendix to Chapter 2

### A.1 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. 2.3.2. For the specific case of marginalising over the redshift distribution nuisance parameters, the vector  $\mathcal{L}'$  has elements

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

with

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

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. (2.27),

is denoted by  $\Delta_l^{\alpha\beta}$ . The elements of the Hessian matrix,  $\mathcal{L}''$ , read

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

with

$$\begin{aligned} \frac{\partial^2 \Delta_l^{(\alpha\beta)}}{\partial a_m^\mu \partial a_n^\nu} = & -A_m^\mu A_n^\nu x_\pm^{(mn)}(\theta_l) (\delta_{\alpha\mu} \delta_{\beta\nu} + \delta_{\beta\mu} \delta_{\alpha\nu}) \\ & - \delta_{mn} \delta_{\mu\nu} A_m^\mu \sum_i x_\pm^{(im)}(\theta_l) (\delta_{\alpha\mu} A_i^\beta + \delta_{\beta\mu} A_i^\alpha). \end{aligned} \quad (\text{A.4})$$

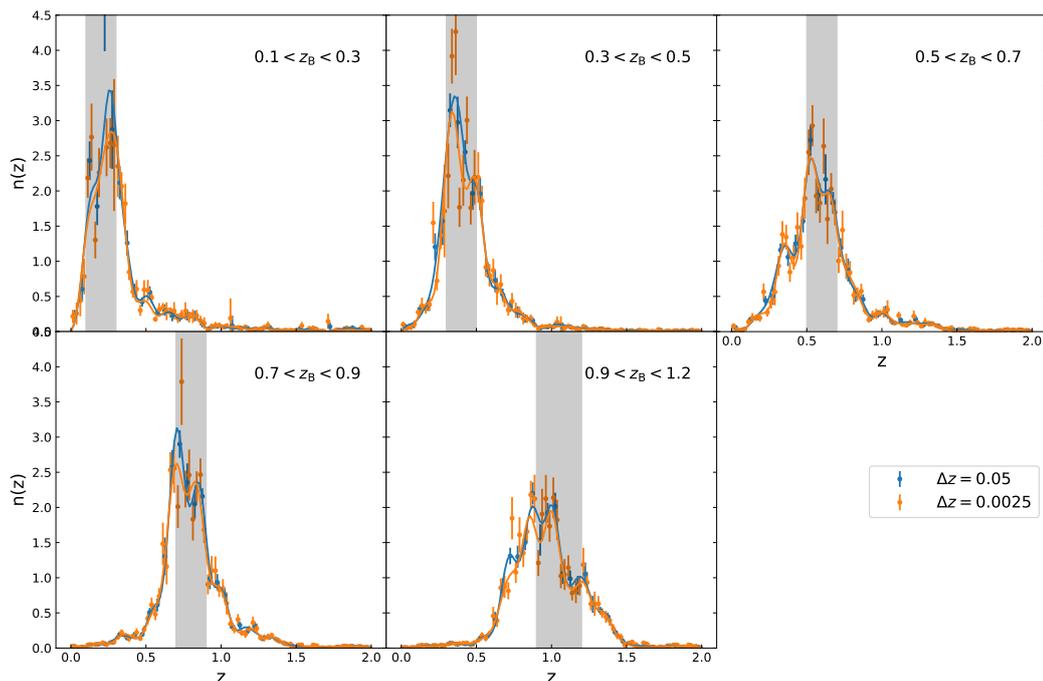
## A.2 Tests of the redshift distribution calibration

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.

### A.2.1 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 A.1 shows a comparison of two fits with 30 Gaussian components;



**Figure A.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  $\Delta z = 0.0025$ , 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.

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  $\Delta z = 0.0025$ . 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  $\chi^2 = 22750$  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. A.1.

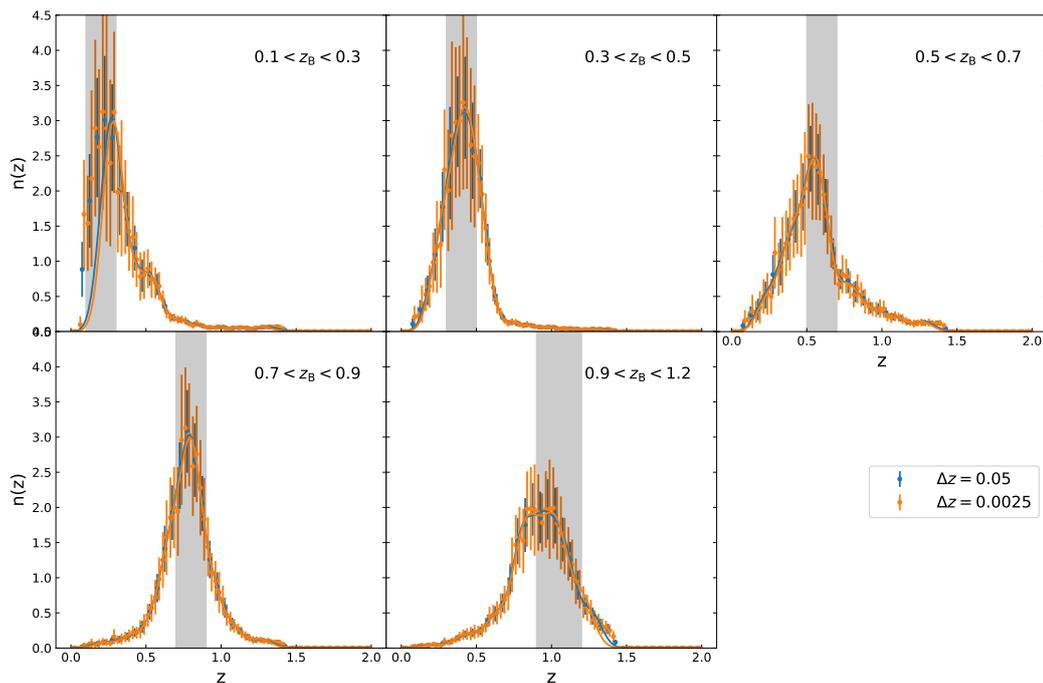
Empirically, we find that rescaling the square root of the covariance  $C_{ij}$  between histogram bins by an additive and multiplicative factor via

$$C'_{ij} = \left(2 \sqrt{C_{ij}} + 0.01\delta_{ij}\right)^2 \quad (\text{A.5})$$

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. 2.2, 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 A.2.2, and leave further investigation of an improved uncertainty quantification for the pre-calibrated redshift distributions for future work.

## A.2.2 Redshift distribution calibration with simulations

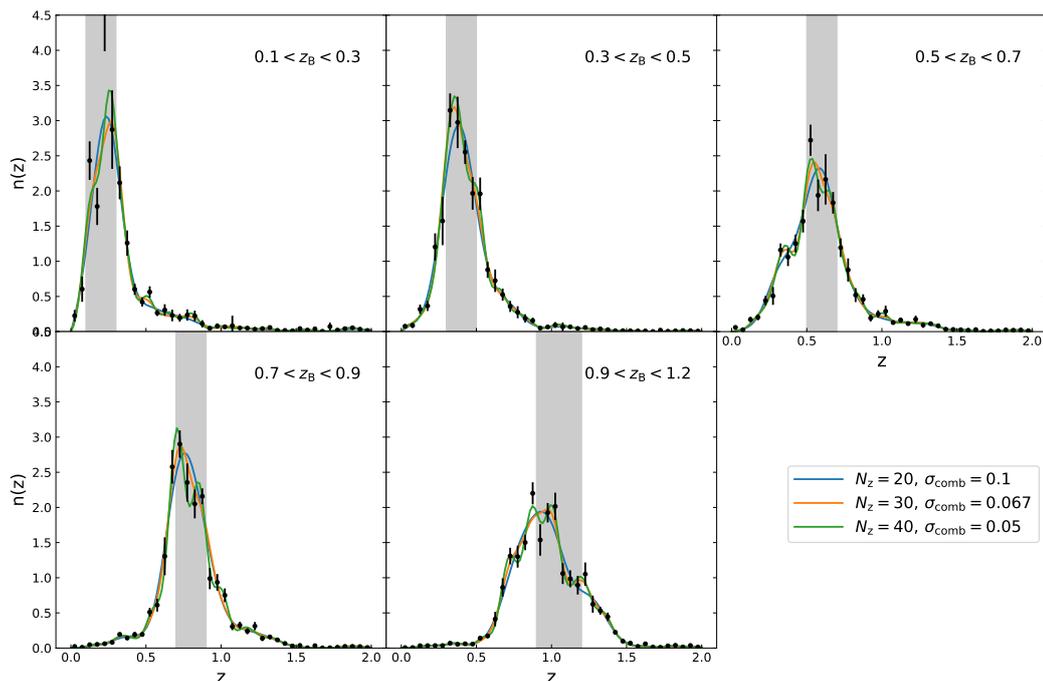
Since in Appendix A.2.1 we found that the Gaussian comb model provides a bad fit to the actual data, which is likely caused by an underestimation of the error bars, we tested our calibration method with simulations. We used redshift distributions that are calibrated with the fiducial DIR method on simulated mock catalogues (van den Busch et al., 2020) based on the MICE simulation (Fosalba et al., 2015a; Crocce et al., 2015; Fosalba et al., 2015b; Carretero et al.,



**Figure A.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.

2015; Hoffmann et al., 2015). Analogous to Appendix A.2.1, we compared two types of data histograms with bin widths of  $\Delta z = 0.05$  and  $\Delta z = 0.025$ . Figure A.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  $\Delta z = 0.025$ .

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.



**Figure A.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  $\sigma_{\text{comb}}$  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.

### A.2.3 Number of Gaussian components

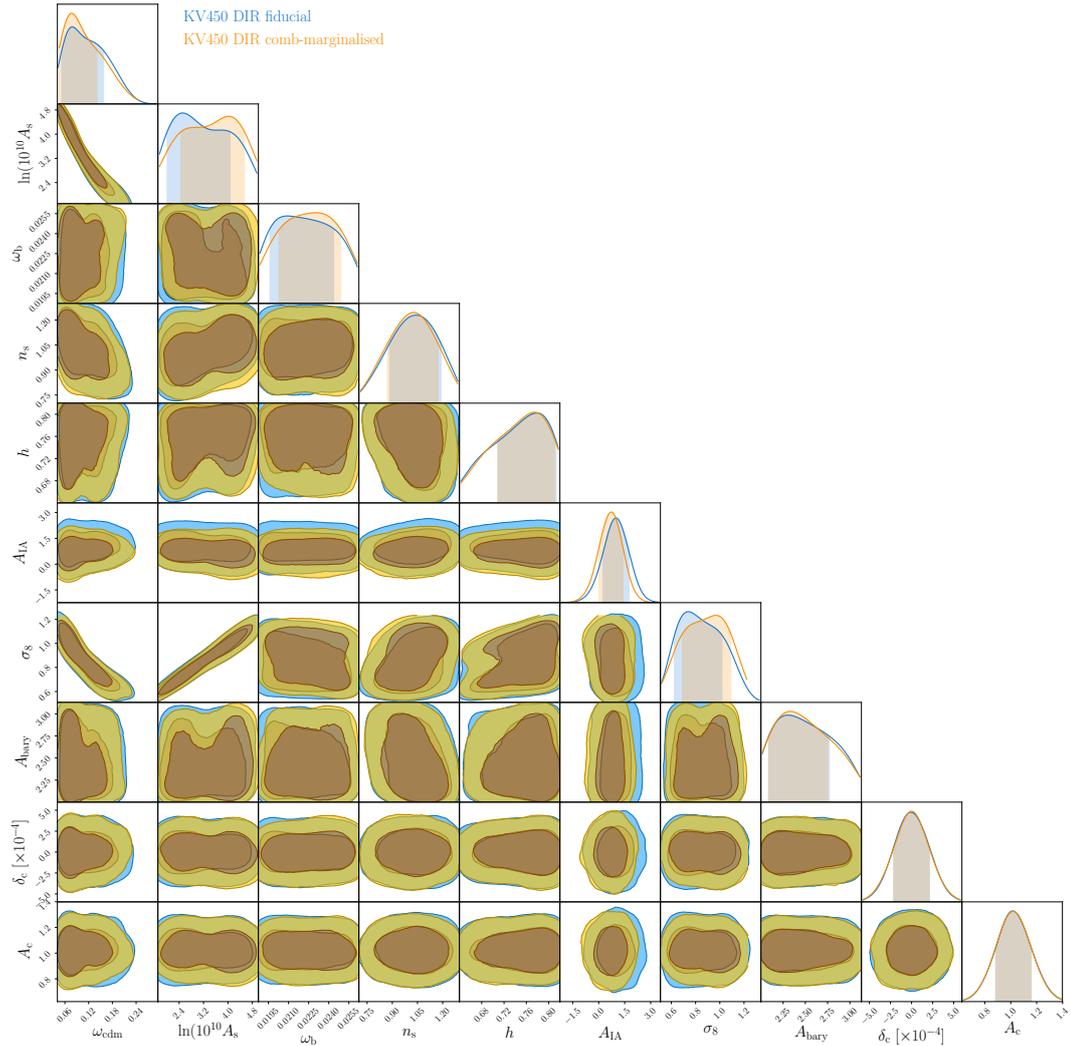
In Fig. A.3 we show a comparison of fits of a Gaussian comb with 20, 30, and 40 components to the pre-calibrated redshift histograms with bin width  $\Delta z = 0.05$ . The width  $\sigma_{\text{comb}}$  of the Gaussians is equal to the separation between each component. We find that variations in the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

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

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

### A.3 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 A.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.



**Figure A.4:** 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.

## A.4 Posteriors of cosmological parameter constraints

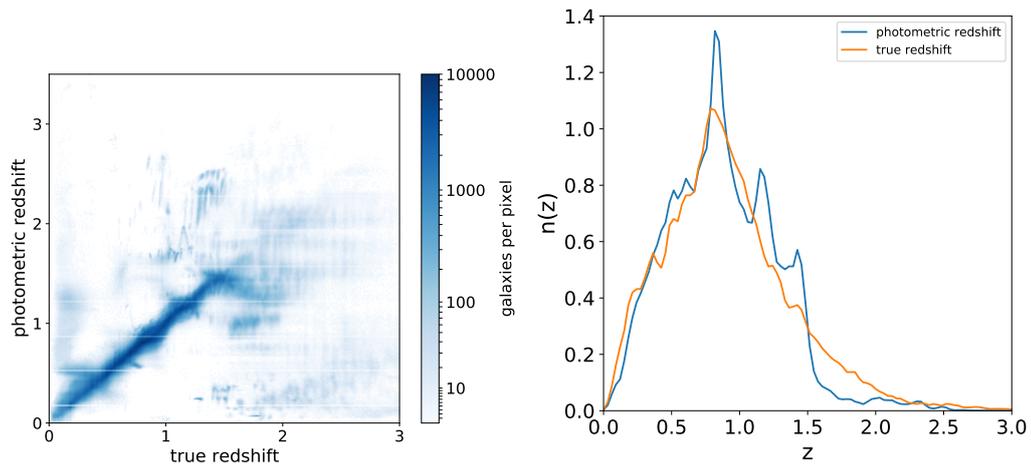
In Fig. A.4 we show marginalised posteriors of cosmological and nuisance parameters. The KV450 likelihood with a Gaussian comb and analytical marginalisation over nuisance parameters is compared to the fiducial KV450 likelihood.

## Appendix B

# Appendix to Chapter 3

### B.1 Photometric redshifts

In this Appendix we compare the true redshifts of galaxies in the synthetic galaxy catalogue with the point estimate of the photometric redshift inferred via SED template fitting. The left panel of Fig. B.1 shows a scatter plot of the true redshift and the photometric redshift. The redshift distributions inferred from the true redshift and the photometric redshift, respectively, are illustrated in the right panel of Fig. B.1.



**Figure B.1:** Left: Scatter plot of the true redshift of galaxies in the photometric sample and the point estimate of the photometric redshift inferred via SED template fitting. Right: Comparison between the true redshift distribution and the redshift distribution inferred from point estimates of the photometric redshift.

## B.2 SOM clustering

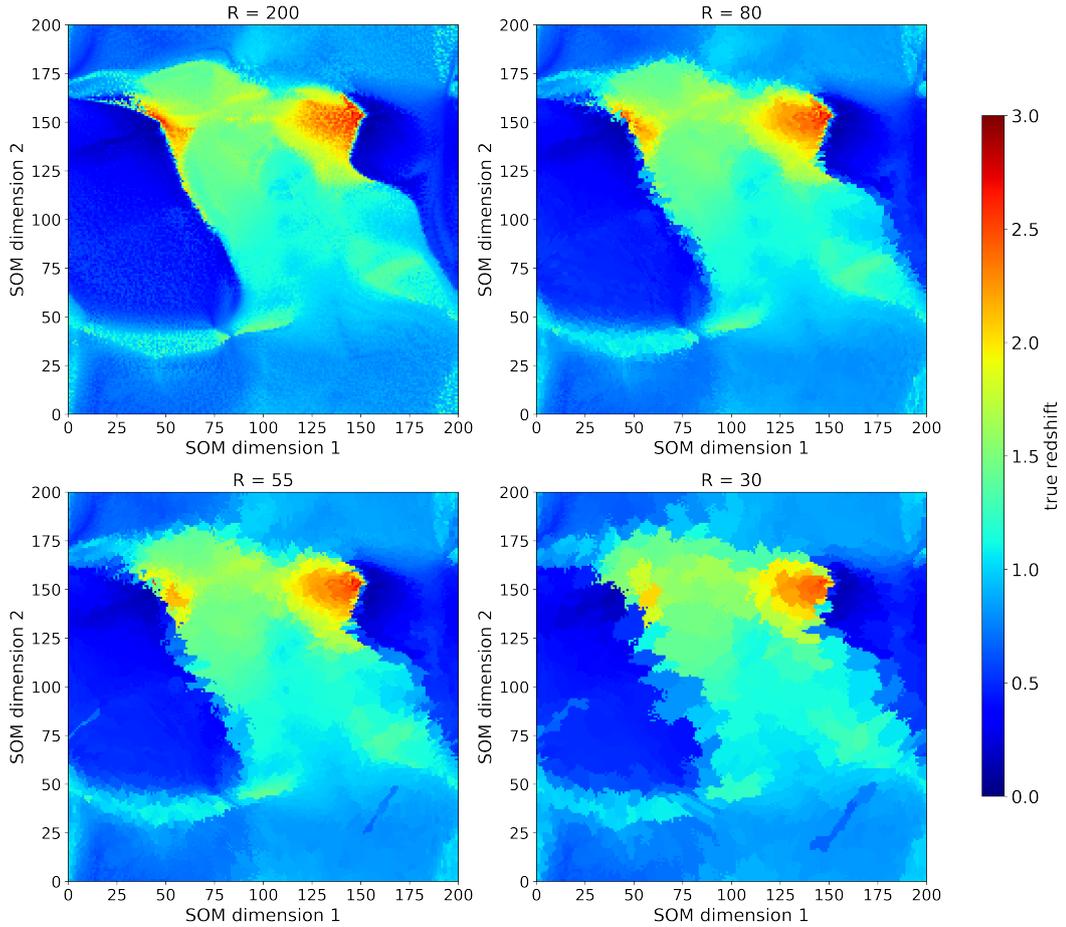
In this Appendix we provide a comparison of SOMs with decreased resolutions derived from a high-dimensional SOM using the hierarchical clustering method described in Section 3.2.3. The original SOM, trained on the observed colours of galaxies with a resolution of  $R = 200$ , is illustrated in the top left panel of Fig. B.2. The remaining panels show SOMs with lower resolutions, inferred from the original SOM via clustering of the weight vectors. The bottom right panel shows the SOM with a resolution of  $R = 30$ , which is the initial resolution from which the simulated annealing algorithm selects groups of galaxies. The top right panel shows the SOM with the final resolution  $R = 80$ , while a SOM with an intermediate resolution of  $R = 55$  is illustrated in the bottom left panel. The colours in each panel represent the mean of the true redshift of galaxies in each node. We note that since the low-resolution SOMs are constructed from the high-resolution SOM with  $R = 200$ , the axes in each panel refer to the index of the high-dimensional SOM.

## B.3 Energy in the simulated annealing optimisation

In this appendix we show the evolution of the energy during the simulated annealing optimisation. In Fig. B.3 we illustrate the energy of the system after six iterations of the algorithm. Each iteration corresponds to a full run of the algorithm with an initial SOM resolution of  $R_{\min} = 30$  and a final resolution  $R_{\max} = 80$  with  $N_{\text{steps}} = 2000$ <sup>1</sup>. We observe that in the first three iterations the algorithm achieves an approximately equal increase in the energy of the system, while the later iterations show smaller increases in the energy, indicating that the algorithm converges towards the maximum energy. The method of consecutively running the algorithm multiple times allows us to explore how many steps in total are needed for the algorithm to converge towards the maximum achievable energy for a given final resolution  $R_{\max}$ . We

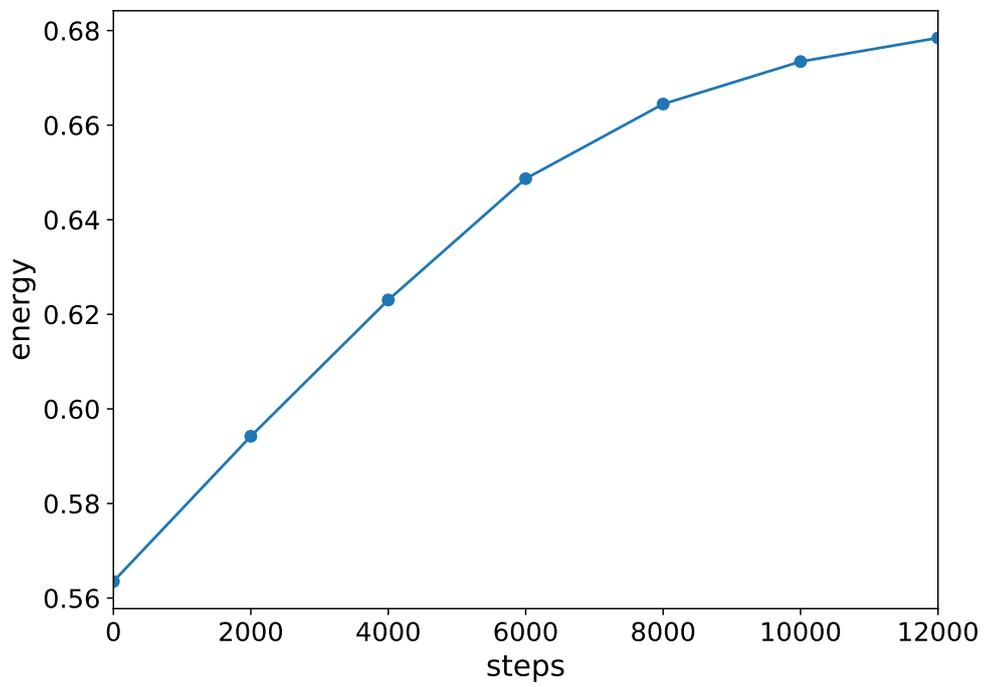
---

<sup>1</sup>The computing time for one step on a 16 core machine is approximately 10 seconds



**Figure B.2:** Illustration of the SOM used in the analysis at different resolutions. The original SOM, trained on the observed colours of galaxies with a resolution of  $R = 200$ , is illustrated in the top left panel. The remaining panels show the SOM with reduced resolution inferred with the clustering method described in Sect. 3.2.3. The colours represent the mean of the true redshift of galaxies in each node.

note that after obtaining the final assignment of galaxies to tomographic bins via simulated annealing, this result can potentially be further improved by re-running the algorithm with an initial resolution of  $R_{\min} = 80$  and an even higher resolution  $R_{\max}$  which can be increased up to the initial resolution of the SOM. However, this comes at the cost of a longer runtime, since higher resolutions imply a selection of fewer galaxies in each step, up to the limit where the algorithm selects individual galaxies. Furthermore, at a certain point noise in the clustering observable will limit the observable effect on the energy of the system.



**Figure B.3:** Evolution of the energy of the simulated annealing algorithm.

# Bibliography

- Abazajian, K. et al. The First Data Release of the Sloan Digital Sky Survey. *AJ*, 126(4):2081–2086, October 2003. doi: [10.1086/378165](https://doi.org/10.1086/378165).
- Abazajian, K.N. et al. The Seventh Data Release of the Sloan Digital Sky Survey. *ApJS*, 182(2):543–558, June 2009. doi: [10.1088/0067-0049/182/2/543](https://doi.org/10.1088/0067-0049/182/2/543).
- Abbott, T.M.C. et al. Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D*, 98(4): 043526, August 2018. doi: [10.1103/PhysRevD.98.043526](https://doi.org/10.1103/PhysRevD.98.043526).
- Aihara, H. et al. The Hyper Suprime-Cam SSP Survey: Overview and survey design. *PASJ*, 70:S4, January 2018. doi: [10.1093/pasj/psx066](https://doi.org/10.1093/pasj/psx066).
- Aitken, S. and Akman, O. Nested sampling for parameter inference in systems biology: Application to an exemplar circadian model. *BMC systems biology*, 7:72, 07 2013. doi: [10.1186/1752-0509-7-72](https://doi.org/10.1186/1752-0509-7-72).
- Alam, S. et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *MNRAS*, 470(3):2617–2652, September 2017. doi: [10.1093/mnras/stx721](https://doi.org/10.1093/mnras/stx721).
- Alarcon, A., Sánchez, C., Bernstein, G.M. and Gaztañaga, E. Redshift inference from the combination of galaxy colours and clustering in a hierarchical Bayesian model - Application to realistic N-body simulations. *MNRAS*, 498(2):2614–2631, October 2020. doi: [10.1093/mnras/staa2478](https://doi.org/10.1093/mnras/staa2478).

- Amara, A. and Réfrégier, A. Optimal surveys for weak-lensing tomography. *MNRAS*, 381(3):1018–1026, November 2007. doi: [10.1111/j.1365-2966.2007.12271.x](https://doi.org/10.1111/j.1365-2966.2007.12271.x).
- Amon, A. et al. Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to data calibration. *Phys. Rev. D*, 105(2):023514, January 2022. doi: [10.1103/PhysRevD.105.023514](https://doi.org/10.1103/PhysRevD.105.023514).
- Asgari, M. et al. KiDS-1000 cosmology: Cosmic shear constraints and comparison between two point statistics. *A&A*, 645:A104, January 2021. doi: [10.1051/0004-6361/202039070](https://doi.org/10.1051/0004-6361/202039070).
- Audren, B., Lesgourgues, J., Benabed, K. and Prunet, S. Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code. *JCAP*, 1302:001, 2013. doi: [10.1088/1475-7516/2013/02/001](https://doi.org/10.1088/1475-7516/2013/02/001).
- Bacon, D.J., Refregier, A.R. and Ellis, R.S. Detection of weak gravitational lensing by large-scale structure. *MNRAS*, 318(2):625–640, October 2000. doi: [10.1046/j.1365-8711.2000.03851.x](https://doi.org/10.1046/j.1365-8711.2000.03851.x).
- Bartelmann, M. and Maturi, M. Weak gravitational lensing. *Scholarpedia*, 12(1):32440, January 2017. doi: [10.4249/scholarpedia.32440](https://doi.org/10.4249/scholarpedia.32440).
- Baum, W.A. Photoelectric Magnitudes and Red-Shifts. In McVittie, G.C., editor, *Problems of Extra-Galactic Research*, volume 15, page 390, January 1962.
- Becker, M.R. and Rozo, E. Fourier band-power E/B-mode estimators for cosmic shear. *MNRAS*, 457(1):304–312, March 2016. doi: [10.1093/mnras/stv3018](https://doi.org/10.1093/mnras/stv3018).
- Benítez, N. Bayesian Photometric Redshift Estimation. *ApJ*, 536(2):571–583, June 2000. doi: [10.1086/308947](https://doi.org/10.1086/308947).

- Benjamin, J. et al. CFHTLenS tomographic weak lensing: quantifying accurate redshift distributions. *MNRAS*, 431(2):1547–1564, May 2013. doi: [10.1093/mnras/stt276](https://doi.org/10.1093/mnras/stt276).
- Bennett, C.L. et al. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results. *ApJS*, 148(1):1–27, September 2003. doi: [10.1086/377253](https://doi.org/10.1086/377253).
- Bernstein, G. and Huterer, D. Catastrophic photometric redshift errors: weak-lensing survey requirements. *MNRAS*, 401(2):1399–1408, January 2010. doi: [10.1111/j.1365-2966.2009.15748.x](https://doi.org/10.1111/j.1365-2966.2009.15748.x).
- Bishop, C.M. *Pattern Recognition and Machine Learning*. Information science and statistics. Springer, 2006. ISBN 9780387310732. doi: [10.1117/1.2819119](https://doi.org/10.1117/1.2819119).
- Blas, D., Lesgourgues, J. and Tram, T. The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. *J. Cosmology Astropart. Phys.*, 2011(7):034, Jul 2011. doi: [10.1088/1475-7516/2011/07/034](https://doi.org/10.1088/1475-7516/2011/07/034).
- Blomqvist, M. et al. Baryon acoustic oscillations from the cross-correlation of Ly $\alpha$  absorption and quasars in eBOSS DR14. *A&A*, 629:A86, September 2019. doi: [10.1051/0004-6361/201935641](https://doi.org/10.1051/0004-6361/201935641).
- Bohachevsky, I.O., Johnson, M.E. and Stein, M.L. Generalized simulated annealing for function optimization. *Technometrics*, 28(3):209–217, 1986. doi: [10.1080/00401706.1986.10488128](https://doi.org/10.1080/00401706.1986.10488128).
- Bonnett, C. et al. Redshift distributions of galaxies in the Dark Energy Survey Science Verification shear catalogue and implications for weak lensing. *Phys. Rev. D*, 94(4):042005, August 2016. doi: [10.1103/PhysRevD.94.042005](https://doi.org/10.1103/PhysRevD.94.042005).
- Bridle, S.L., Crittenden, R., Melchiorri, A., Hobson, M.P., Kneissl, R. and Lasenby, A.N. Analytic marginalization over CMB calibration and beam uncertainty. *MNRAS*, 335(4):1193–1200, October 2002. doi: [10.1046/j.1365-8711.2002.05709.x](https://doi.org/10.1046/j.1365-8711.2002.05709.x).

- Bridle, S. and King, L. Dark energy constraints from cosmic shear power spectra: impact of intrinsic alignments on photometric redshift requirements. *New Journal of Physics*, 9(12):444, December 2007. doi: [10.1088/1367-2630/9/12/444](https://doi.org/10.1088/1367-2630/9/12/444).
- Brinckmann, T. and Lesgourgues, J. MontePython 3: Boosted MCMC sampler and other features. *Physics of the Dark Universe*, 24:100260, March 2019. doi: [10.1016/j.dark.2018.100260](https://doi.org/10.1016/j.dark.2018.100260).
- Buchner, J., Georgakakis, A., Nandra, K., Hsu, L., Rangel, C., Brightman, M., Merloni, A., Salvato, M., Donley, J. and Kocevski, D. X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *A&A*, 564:A125, April 2014. doi: [10.1051/0004-6361/201322971](https://doi.org/10.1051/0004-6361/201322971).
- Buchs, R. et al. Phenotypic redshifts with self-organizing maps: A novel method to characterize redshift distributions of source galaxies for weak lensing. *MNRAS*, 489(1):820–841, October 2019. doi: [10.1093/mnras/stz2162](https://doi.org/10.1093/mnras/stz2162).
- Carrasco Kind, M. and Brunner, R.J. SOMz: photometric redshift PDFs with self-organizing maps and random atlas. *MNRAS*, 438(4):3409–3421, March 2014. doi: [10.1093/mnras/stt2456](https://doi.org/10.1093/mnras/stt2456).
- Carretero, J., Castander, F.J., Gaztañaga, E., Crocce, M. and Fosalba, P. An algorithm to build mock galaxy catalogues using MICE simulations. *MNRAS*, 447(1):646–670, February 2015. doi: [10.1093/mnras/stu2402](https://doi.org/10.1093/mnras/stu2402).
- Černý, V. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *J. Optim. Theory Appl.*, 45(1):41–51, January 1985. ISSN 0022-3239. doi: [10.1007/BF00940812](https://doi.org/10.1007/BF00940812).
- Chira, M. and Plionis, M. A simulated annealing algorithm to quantify patterns in astronomical data. *MNRAS*, 490(4):5904–5920, December 2019. doi: [10.1093/mnras/stz2885](https://doi.org/10.1093/mnras/stz2885).

- Colless, M. et al. The 2dF Galaxy Redshift Survey: spectra and redshifts. *MNRAS*, 328(4):1039–1063, December 2001. doi: [10.1046/j.1365-8711.2001.04902.x](https://doi.org/10.1046/j.1365-8711.2001.04902.x).
- Connolly, A.J., Csabai, I., Szalay, A.S., Koo, D.C., Kron, R.G. and Munn, J.A. Slicing Through Multicolor Space: Galaxy Redshifts from Broadband Photometry. *AJ*, 110:2655, December 1995. doi: [10.1086/117720](https://doi.org/10.1086/117720).
- Corana, A., Marchesi, M., Martini, C. and Ridella, S. Minimizing multimodal functions of continuous variables with the “simulated annealing” algorithm. *ACM Trans. Math. Softw.*, 13(3):262–280, sep 1987. ISSN 0098-3500. doi: [10.1145/29380.29864](https://doi.org/10.1145/29380.29864).
- Crocce, M., Castander, F.J., Gaztañaga, E., Fosalba, P. and Carretero, J. The MICE Grand Challenge lightcone simulation - II. Halo and galaxy catalogues. *MNRAS*, 453(2):1513–1530, October 2015. doi: [10.1093/mnras/stv1708](https://doi.org/10.1093/mnras/stv1708).
- Cunha, C.E., Huterer, D., Lin, H., Busha, M.T. and Wechsler, R.H. Spectroscopic failures in photometric redshift calibration: cosmological biases and survey requirements. *MNRAS*, 444(1):129–146, October 2014. doi: [10.1093/mnras/stu1424](https://doi.org/10.1093/mnras/stu1424).
- de Jong, J.T.A. et al. The first and second data releases of the Kilo-Degree Survey. *A&A*, 582:A62, October 2015. doi: [10.1051/0004-6361/201526601](https://doi.org/10.1051/0004-6361/201526601).
- de Jong, J.T.A. et al. The third data release of the Kilo-Degree Survey and associated data products. *A&A*, 604:A134, August 2017. doi: [10.1051/0004-6361/201730747](https://doi.org/10.1051/0004-6361/201730747).
- de Sainte Agathe, V. et al. Baryon acoustic oscillations at  $z = 2.34$  from the correlations of Ly $\alpha$  absorption in eBOSS DR14. *A&A*, 629:A85, September 2019. doi: [10.1051/0004-6361/201935638](https://doi.org/10.1051/0004-6361/201935638).
- DESI Collaboration et al. The DESI Experiment Part I: Science, Targeting, and Survey Design. *arXiv e-prints*. arXiv:1611.00036, October 2016.

- Dey, A. et al. Overview of the DESI Legacy Imaging Surveys. *AJ*, 157(5):168, May 2019. doi: [10.3847/1538-3881/ab089d](https://doi.org/10.3847/1538-3881/ab089d).
- Dodelson, S. *Modern cosmology*. Academic Press, San Diego, CA, 2003.
- Drlica-Wagner, A. et al. Dark Energy Survey Year 1 Results: The Photometric Data Set for Cosmology. *ApJS*, 235(2):33, April 2018. doi: [10.3847/1538-4365/aab4f5](https://doi.org/10.3847/1538-4365/aab4f5).
- Edge, A., Sutherland, W., Kuijken, K., Driver, S., McMahon, R., Eales, S. and Emerson, J.P. The VISTA Kilo-degree Infrared Galaxy (VIKING) Survey: Bridging the Gap between Low and High Redshift. *The Messenger*, 154: 32–34, December 2013.
- Efstathiou, G. and Lemos, P. Statistical inconsistencies in the KiDS-450 data set. *MNRAS*, 476(1):151–157, May 2018. doi: [10.1093/mnras/sty099](https://doi.org/10.1093/mnras/sty099).
- Eifler, T. et al. Cosmology with the Roman Space Telescope: synergies with the Rubin Observatory Legacy Survey of Space and Time. *MNRAS*, 507(1): 1514–1527, October 2021. doi: [10.1093/mnras/stab533](https://doi.org/10.1093/mnras/stab533).
- Feroz, F., Hobson, M.P. and Bridges, M. MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS*, 398(4): 1601–1614, Oct 2009. doi: [10.1111/j.1365-2966.2009.14548.x](https://doi.org/10.1111/j.1365-2966.2009.14548.x).
- Feroz, F., Hobson, M.P., Cameron, E. and Pettitt, A.N. Importance Nested Sampling and the MultiNest Algorithm. *The Open Journal of Astrophysics*, 2(1):10, Nov 2019. doi: [10.21105/astro.1306.2144](https://doi.org/10.21105/astro.1306.2144).
- Fortuna, M.C., Hoekstra, H., Joachimi, B., Johnston, H., Chisari, N.E., Georgiou, C. and Mahony, C. The halo model as a versatile tool to predict intrinsic alignments. *MNRAS*, 501(2):2983–3002, February 2021. doi: [10.1093/mnras/staa3802](https://doi.org/10.1093/mnras/staa3802).

- Fosalba, P., Crocce, M., Gaztañaga, E. and Castander, F.J. The MICE grand challenge lightcone simulation - I. Dark matter clustering. *MNRAS*, 448(4): 2987–3000, April 2015a. doi: [10.1093/mnras/stv138](https://doi.org/10.1093/mnras/stv138).
- Fosalba, P., Gaztañaga, E., Castander, F.J. and Crocce, M. The MICE Grand Challenge light-cone simulation - III. Galaxy lensing mocks from all-sky lensing maps. *MNRAS*, 447(2):1319–1332, February 2015b. doi: [10.1093/mnras/stu2464](https://doi.org/10.1093/mnras/stu2464).
- Friedmann, A. Über die Krümmung des Raumes. *Zeitschrift für Physik*, 10: 377–386, January 1922. doi: [10.1007/BF01332580](https://doi.org/10.1007/BF01332580).
- Friedmann, A. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift für Physik*, 21(1):326–332, December 1924. doi: [10.1007/BF01328280](https://doi.org/10.1007/BF01328280).
- Fry, J.N. The Evolution of Bias. *ApJ*, 461:L65, April 1996. doi: [10.1086/310006](https://doi.org/10.1086/310006).
- Fustes, D., Manteiga, M., Dafonte, C., Arcay, B., Ulla, A., Smith, K., Borrachero, R. and Sordo, R. An approach to the analysis of SDSS spectroscopic outliers based on self-organizing maps. Designing the outlier analysis software package for the next Gaia survey. *A&A*, 559:A7, November 2013. doi: [10.1051/0004-6361/201321445](https://doi.org/10.1051/0004-6361/201321445).
- Gatti, M. et al. Dark energy survey year 3 results: weak lensing shape catalogue. *MNRAS*, 504(3):4312–4336, July 2021. doi: [10.1093/mnras/stab918](https://doi.org/10.1093/mnras/stab918).
- Gatti, M. et al. Dark Energy Survey Year 3 Results: clustering redshifts - calibration of the weak lensing source redshift distributions with redMaGiC and BOSS/eBOSS. *MNRAS*, 510(1):1223–1247, February 2022. doi: [10.1093/mnras/stab3311](https://doi.org/10.1093/mnras/stab3311).
- Geach, J.E. Unsupervised self-organized mapping: a versatile empirical tool for object selection, classification and redshift estimation in large

- surveys. *MNRAS*, 419(3):2633–2645, January 2012. doi: [10.1111/j.1365-2966.2011.19913.x](https://doi.org/10.1111/j.1365-2966.2011.19913.x).
- Guth, A.H. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23(2):347–356, January 1981. doi: [10.1103/PhysRevD.23.347](https://doi.org/10.1103/PhysRevD.23.347).
- Hadzhiyska, B., Alonso, D., Nicola, A. and Slosar, A. Analytic marginalization of  $N(z)$  uncertainties in tomographic galaxy surveys. *J. Cosmology Astropart. Phys.*, 2020(10):056, October 2020. doi: [10.1088/1475-7516/2020/10/056](https://doi.org/10.1088/1475-7516/2020/10/056).
- Handley, W.J., Hobson, M.P. and Lasenby, A.N. POLYCHORD: next-generation nested sampling. *MNRAS*, 453(4):4384–4398, November 2015a. doi: [10.1093/mnras/stv1911](https://doi.org/10.1093/mnras/stv1911).
- Handley, W.J., Hobson, M.P. and Lasenby, A.N. polychord: nested sampling for cosmology. *MNRAS*, 450:L61–L65, June 2015b. doi: [10.1093/mnrasl/slv047](https://doi.org/10.1093/mnrasl/slv047).
- Handley, W. and Lemos, P. Quantifying tensions in cosmological parameters: Interpreting the DES evidence ratio. *Phys. Rev. D*, 100(4):043504, August 2019a. doi: [10.1103/PhysRevD.100.043504](https://doi.org/10.1103/PhysRevD.100.043504).
- Handley, W. and Lemos, P. Quantifying dimensionality: Bayesian cosmological model complexities. *Phys. Rev. D*, 100(2):023512, July 2019b. doi: [10.1103/PhysRevD.100.023512](https://doi.org/10.1103/PhysRevD.100.023512).
- Hawkins, E. et al. The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe. *MNRAS*, 346(1):78–96, November 2003. doi: [10.1046/j.1365-2966.2003.07063.x](https://doi.org/10.1046/j.1365-2966.2003.07063.x).
- Henderson, D., Jacobson, S.H. and Johnson, A.W. *The Theory and Practice of Simulated Annealing*, pages 287–319. Springer US, Boston, MA, 2003. ISBN 978-0-306-48056-0. doi: [10.1007/0-306-48056-5\\_10](https://doi.org/10.1007/0-306-48056-5_10).

- Heymans, C. et al. KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints. *A&A*, 646:A140, February 2021. doi: [10.1051/0004-6361/202039063](https://doi.org/10.1051/0004-6361/202039063).
- Hikage, C. et al. Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data. *PASJ*, 71(2):43, April 2019. doi: [10.1093/pasj/psz010](https://doi.org/10.1093/pasj/psz010).
- Hildebrandt, H. et al. KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *MNRAS*, 465(2):1454–1498, February 2017. doi: [10.1093/mnras/stw2805](https://doi.org/10.1093/mnras/stw2805).
- Hildebrandt, H. et al. KiDS+VIKING-450: Cosmic shear tomography with optical and infrared data. *A&A*, 633:A69, Jan 2020. doi: [10.1051/0004-6361/201834878](https://doi.org/10.1051/0004-6361/201834878).
- Hildebrandt, H. et al. KiDS-1000 catalogue: Redshift distributions and their calibration. *A&A*, 647:A124, March 2021. doi: [10.1051/0004-6361/202039018](https://doi.org/10.1051/0004-6361/202039018).
- Hirata, C.M. and Seljak, U. Intrinsic alignment-lensing interference as a contaminant of cosmic shear. *Phys. Rev. D*, 70(6):063526, September 2004. doi: [10.1103/PhysRevD.70.063526](https://doi.org/10.1103/PhysRevD.70.063526).
- Hoffmann, K., Bel, J., Gaztañaga, E., Crocce, M., Fosalba, P. and Castander, F.J. Measuring the growth of matter fluctuations with third-order galaxy correlations. *MNRAS*, 447(2):1724–1745, February 2015. doi: [10.1093/mnras/stu2492](https://doi.org/10.1093/mnras/stu2492).
- Hoyle, B. et al. Dark Energy Survey Year 1 Results: redshift distributions of the weak-lensing source galaxies. *MNRAS*, 478(1):592–610, July 2018. doi: [10.1093/mnras/sty957](https://doi.org/10.1093/mnras/sty957).
- Hoyle, B. and Rau, M.M. Self-consistent redshift estimation using correlation functions without a spectroscopic reference sample. *MNRAS*, 485(3):3642–3660, May 2019. doi: [10.1093/mnras/stz502](https://doi.org/10.1093/mnras/stz502).

- Hu, W. Power Spectrum Tomography with Weak Lensing. *ApJ*, 522(1):L21–L24, September 1999. doi: [10.1086/312210](https://doi.org/10.1086/312210).
- Huterer, D., Takada, M., Bernstein, G. and Jain, B. Systematic errors in future weak-lensing surveys: requirements and prospects for self-calibration. *MNRAS*, 366(1):101–114, February 2006. doi: [10.1111/j.1365-2966.2005.09782.x](https://doi.org/10.1111/j.1365-2966.2005.09782.x).
- Ivezić, Ž. et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *ApJ*, 873(2):111, March 2019. doi: [10.3847/1538-4357/ab042c](https://doi.org/10.3847/1538-4357/ab042c).
- Jarvis, M., Bernstein, G. and Jain, B. The skewness of the aperture mass statistic. *MNRAS*, 352(1):338–352, July 2004. doi: [10.1111/j.1365-2966.2004.07926.x](https://doi.org/10.1111/j.1365-2966.2004.07926.x).
- Jasche, J. and Wandelt, B.D. Bayesian inference from photometric redshift surveys. *MNRAS*, 425(2):1042–1056, September 2012. doi: [10.1111/j.1365-2966.2012.21423.x](https://doi.org/10.1111/j.1365-2966.2012.21423.x).
- Jeffreys, H. *The Theory of Probability*. Oxford Classic Texts in the Physical Sciences. 1939. ISBN 978-0-19-850368-2, 978-0-19-853193-7.
- Jenkins, A., Frenk, C.S., White, S.D.M., Colberg, J.M., Cole, S., Evrard, A.E., Couchman, H.M.P. and Yoshida, N. The mass function of dark matter haloes. *MNRAS*, 321(2):372–384, February 2001. doi: [10.1046/j.1365-8711.2001.04029.x](https://doi.org/10.1046/j.1365-8711.2001.04029.x).
- Joachimi, B., Mandelbaum, R., Abdalla, F.B. and Bridle, S.L. Constraints on intrinsic alignment contamination of weak lensing surveys using the MegaZ-LRG sample. *A&A*, 527:A26, March 2011. doi: [10.1051/0004-6361/201015621](https://doi.org/10.1051/0004-6361/201015621).
- Joachimi, B. et al. KiDS-1000 methodology: Modelling and inference for joint weak gravitational lensing and spectroscopic galaxy clustering analysis. *A&A*, 646:A129, February 2021. doi: [10.1051/0004-6361/202038831](https://doi.org/10.1051/0004-6361/202038831).

- Jones, D.H. et al. The 6dF Galaxy Survey: samples, observational techniques and the first data release. *MNRAS*, 355(3):747–763, December 2004. doi: [10.1111/j.1365-2966.2004.08353.x](https://doi.org/10.1111/j.1365-2966.2004.08353.x).
- Jones, D.H. et al. The 6dF Galaxy Survey: final redshift release (DR3) and southern large-scale structures. *MNRAS*, 399(2):683–698, October 2009. doi: [10.1111/j.1365-2966.2009.15338.x](https://doi.org/10.1111/j.1365-2966.2009.15338.x).
- Kaiser, N. On the spatial correlations of Abell clusters. *ApJ*, 284:L9–L12, September 1984. doi: [10.1086/184341](https://doi.org/10.1086/184341).
- Kaiser, N. Weak Gravitational Lensing of Distant Galaxies. *ApJ*, 388:272, April 1992. doi: [10.1086/171151](https://doi.org/10.1086/171151).
- Kaiser, N., Wilson, G. and Luppino, G.A. Large-Scale Cosmic Shear Measurements. *arXiv e-prints*. [astro-ph/0003338](https://arxiv.org/abs/astro-ph/0003338), March 2000.
- Kerscher, M., Szapudi, I. and Szalay, A.S. A Comparison of Estimators for the Two-Point Correlation Function. *ApJ*, 535(1):L13–L16, May 2000. doi: [10.1086/312702](https://doi.org/10.1086/312702).
- Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983. doi: [10.1126/science.220.4598.671](https://doi.org/10.1126/science.220.4598.671).
- Kirkpatrick, S. Optimization by simulated annealing: Quantitative studies. *Journal of Statistical Physics*, 34(5):975–986, March 1984. ISSN 1572-9613. doi: [10.1007/BF01009452](https://doi.org/10.1007/BF01009452).
- Köhlinger, F., Joachimi, B., Asgari, M., Viola, M., Joudaki, S. and Tröster, T. A Bayesian quantification of consistency in correlated data sets. *MNRAS*, 484(3):3126–3153, April 2019. doi: [10.1093/mnras/stz132](https://doi.org/10.1093/mnras/stz132).
- Kohonen, T. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990. doi: [10.1109/5.58325](https://doi.org/10.1109/5.58325).

- Koo, D.C. Optical multicolors : a poor person's Z machine for galaxies. *AJ*, 90:418–440, March 1985. doi: [10.1086/113748](https://doi.org/10.1086/113748).
- Korytov, D. et al. CosmoDC2: A Synthetic Sky Catalog for Dark Energy Science with LSST. *ApJS*, 245(2):26, December 2019. doi: [10.3847/1538-4365/ab510c](https://doi.org/10.3847/1538-4365/ab510c).
- Kuijken, K. et al. The fourth data release of the Kilo-Degree Survey: ugr imaging and nine-band optical-IR photometry over 1000 square degrees. *A&A*, 625:A2, May 2019. doi: [10.1051/0004-6361/201834918](https://doi.org/10.1051/0004-6361/201834918).
- Kuijken, K. et al. Gravitational lensing analysis of the Kilo-Degree Survey. *MNRAS*, 454(4):3500–3532, December 2015. doi: [10.1093/mnras/stv2140](https://doi.org/10.1093/mnras/stv2140).
- Kullback, S. and Leibler, R.A. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79 – 86, 1951. doi: [10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694).
- Landy, S.D. and Szalay, A.S. Bias and Variance of Angular Correlation Functions. *ApJ*, 412:64, July 1993. doi: [10.1086/172900](https://doi.org/10.1086/172900).
- Laureijs, R. et al. Euclid Definition Study Report. *arXiv e-prints*. doi: [arXiv:1110.3193](https://arxiv.org/abs/1110.3193), October 2011.
- Leavitt, H.S. and Pickering, E.C. Periods of 25 Variable Stars in the Small Magellanic Cloud. *Harvard College Observatory Circular*, 173:1–3, March 1912.
- Leistedt, B., Hogg, D.W., Wechsler, R.H. and DeRose, J. Hierarchical Modeling and Statistical Calibration for Photometric Redshifts. *ApJ*, 881(1):80, August 2019. doi: [10.3847/1538-4357/ab2d29](https://doi.org/10.3847/1538-4357/ab2d29).
- Lemos, P. et al. Robust sampling for weak lensing and clustering analyses with the dark energy survey. *MNRAS*, October 2022. doi: [10.1093/mnras/stac2786](https://doi.org/10.1093/mnras/stac2786).

- Lemos, P., Köhlinger, F., Handley, W., Joachimi, B., Whiteway, L. and Lahav, O. Quantifying Suspiciousness within correlated data sets. *MNRAS*, 496(4): 4647–4653, August 2020. doi: [10.1093/mnras/staa1836](https://doi.org/10.1093/mnras/staa1836).
- Lima, M., Cunha, C.E., Oyaizu, H., Frieman, J., Lin, H. and Sheldon, E.S. Estimating the redshift distribution of photometric galaxy samples. *MNRAS*, 390(1):118–130, October 2008. doi: [10.1111/j.1365-2966.2008.13510.x](https://doi.org/10.1111/j.1365-2966.2008.13510.x).
- Loh, E.D. and Spillar, E.J. Photometric Redshifts of Galaxies. *ApJ*, 303:154, April 1986. doi: [10.1086/164062](https://doi.org/10.1086/164062).
- LSST Science Collaboration et al. LSST Science Book, Version 2.0. *arXiv e-prints*. arXiv:0912.0201, December 2009.
- Ma, Z., Hu, W. and Huterer, D. Effects of Photometric Redshift Uncertainties on Weak-Lensing Tomography. *ApJ*, 636(1):21–29, January 2006. doi: [10.1086/497068](https://doi.org/10.1086/497068).
- Mann, R.G., Peacock, J.A. and Heavens, A.F. Eulerian bias and the galaxy density field. *MNRAS*, 293(3):209–221, January 1998. doi: [10.1046/j.1365-8711.1998.01053.x](https://doi.org/10.1046/j.1365-8711.1998.01053.x).
- Masters, D. et al. Mapping the Galaxy Color-Redshift Relation: Optimal Photometric Redshift Calibration Strategies for Cosmology Surveys. *ApJ*, 813(1):53, November 2015. doi: [10.1088/0004-637X/813/1/53](https://doi.org/10.1088/0004-637X/813/1/53).
- Mather, J.C. et al. A Preliminary Measurement of the Cosmic Microwave Background Spectrum by the Cosmic Background Explorer (COBE) Satellite. *ApJ*, 354:L37, May 1990. doi: [10.1086/185717](https://doi.org/10.1086/185717).
- Matthews, D.J. and Newman, J.A. Reconstructing Redshift Distributions with Cross-correlations: Tests and an Optimized Recipe. *ApJ*, 721(1):456–468, September 2010. doi: [10.1088/0004-637X/721/1/456](https://doi.org/10.1088/0004-637X/721/1/456).

- McLeod, M., Balan, S.T. and Abdalla, F.B. A joint analysis for cosmology and photometric redshift calibration using cross-correlations. *MNRAS*, 466(3): 3558–3568, April 2017. doi: [10.1093/mnras/stw2989](https://doi.org/10.1093/mnras/stw2989).
- McQuinn, M. and White, M. On using angular cross-correlations to determine source redshift distributions. *MNRAS*, 433(4):2857–2883, August 2013. doi: [10.1093/mnras/stt914](https://doi.org/10.1093/mnras/stt914).
- Mead, A.J., Peacock, J.A., Heymans, C., Joudaki, S. and Heavens, A.F. An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models. *MNRAS*, 454(2):1958–1975, December 2015. doi: [10.1093/mnras/stv2036](https://doi.org/10.1093/mnras/stv2036).
- Ménard, B., Scranton, R., Schmidt, S., Morrison, C., Jeong, D., Budavari, T. and Rahman, M. Clustering-based redshift estimation: method and application to data. *arXiv e-prints*. arXiv:1303.4722, March 2013.
- Mitra, D., Romeo, F. and Sangiovanni-Vincentelli, A. Convergence and finite-time behavior of simulated annealing. *Advances in Applied Probability*, 18(3):747–771, 1986. doi: [10.2307/1427186](https://doi.org/10.2307/1427186).
- Moessner, R. and Jain, B. Angular cross-correlation of galaxies: a probe of gravitational lensing by large-scale structure. *MNRAS*, 294(1):L18–L24, February 1998. doi: [10.1046/j.1365-8711.1998.01378.x](https://doi.org/10.1046/j.1365-8711.1998.01378.x).
- Muir, J. et al. DES Y1 results: Splitting growth and geometry to test  $\Lambda$  CDM. *Phys. Rev. D*, 103(2):023528, January 2021. doi: [10.1103/PhysRevD.103.023528](https://doi.org/10.1103/PhysRevD.103.023528).
- Mukherjee, P., Parkinson, D. and Liddle, A.R. A Nested Sampling Algorithm for Cosmological Model Selection. *ApJ*, 638(2):L51–L54, February 2006. doi: [10.1086/501068](https://doi.org/10.1086/501068).
- Müllner, D. Modern hierarchical, agglomerative clustering algorithms. *arXiv e-prints*. arXiv:1109.2378, September 2011.

- Myles, J. et al. Dark Energy Survey Year 3 results: redshift calibration of the weak lensing source galaxies. *MNRAS*, 505(3):4249–4277, August 2021. doi: [10.1093/mnras/stab1515](https://doi.org/10.1093/mnras/stab1515).
- Navarro, J.F., Frenk, C.S. and White, S.D.M. The Structure of Cold Dark Matter Halos. *ApJ*, 462:563, May 1996. doi: [10.1086/177173](https://doi.org/10.1086/177173).
- Neal, R.M. Slice sampling. *The Annals of Statistics*, 31(3):705 – 767, 2003. doi: [10.1214/aos/1056562461](https://doi.org/10.1214/aos/1056562461).
- Newman, J.A. Calibrating Redshift Distributions beyond Spectroscopic Limits with Cross-Correlations. *ApJ*, 684(1):88–101, September 2008. doi: [10.1086/589982](https://doi.org/10.1086/589982).
- Palanque-Delabrouille, N. et al. Neutrino masses and cosmology with Lyman-alpha forest power spectrum. *J. Cosmology Astropart. Phys.*, 2015(11):011, November 2015. doi: [10.1088/1475-7516/2015/11/011](https://doi.org/10.1088/1475-7516/2015/11/011).
- Peacock, J.A. *Cosmological Physics*. Cambridge University Press, 1998. doi: [10.1017/CBO9780511804533](https://doi.org/10.1017/CBO9780511804533).
- Peebles, P. *The Large-scale Structure of the Universe*. Princeton Series in Physics. Princeton University Press, 1980. ISBN 978-0-691-08240-0. doi: [10.1515/9780691206714](https://doi.org/10.1515/9780691206714).
- Perlmutter, S. et al. Measurements of  $\Omega$  and  $\Lambda$  from 42 High-Redshift Supernovae. *ApJ*, 517(2):565–586, June 1999. doi: [10.1086/307221](https://doi.org/10.1086/307221).
- Planck Collaboration et al. Planck 2013 results. I. Overview of products and scientific results. *A&A*, 571:A1, November 2014. doi: [10.1051/0004-6361/201321529](https://doi.org/10.1051/0004-6361/201321529).
- Planck Collaboration et al. Planck 2018 results. I. Overview and the cosmological legacy of Planck. *A&A*, 641:A1, September 2020a. doi: [10.1051/0004-6361/201833880](https://doi.org/10.1051/0004-6361/201833880).

- Planck Collaboration et al. Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, September 2020b. doi: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910).
- Press, W.H. and Teukolsky, S.A. Simulated annealing optimization over continuous spaces. *Computers in Physics*, 5(4):426–429, 1991. doi: [10.1063/1.4823002](https://doi.org/10.1063/1.4823002).
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, USA, 3 edition, 2007. ISBN 0521880688.
- Puschell, J.J., Owen, F.N. and Laing, R.A. Near-infrared photometry of distant radio galaxies - Spectral flux distributions and redshift estimates. *ApJ*, 257: L57–L61, June 1982. doi: [10.1086/183808](https://doi.org/10.1086/183808).
- Rau, M.M., Wilson, S. and Mandelbaum, R. Estimating redshift distributions using hierarchical logistic Gaussian processes. *MNRAS*, 491(4):4768–4782, February 2020. doi: [10.1093/mnras/stz3295](https://doi.org/10.1093/mnras/stz3295).
- Raveri, M. and Hu, W. Concordance and discordance in cosmology. *Phys. Rev. D*, 99(4):043506, February 2019. doi: [10.1103/PhysRevD.99.043506](https://doi.org/10.1103/PhysRevD.99.043506).
- Reid, B.A. et al. Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies. *MNRAS*, 404(1):60–85, May 2010. doi: [10.1111/j.1365-2966.2010.16276.x](https://doi.org/10.1111/j.1365-2966.2010.16276.x).
- Rhodes, J. et al. Scientific Synergy between LSST and Euclid. *ApJS*, 233(2): 21, December 2017. doi: [10.3847/1538-4365/aa96b0](https://doi.org/10.3847/1538-4365/aa96b0).
- Riess, A.G. et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116(3):1009–1038, September 1998. doi: [10.1086/300499](https://doi.org/10.1086/300499).
- Riess, A.G. et al. A 2.4% Determination of the Local Value of the Hubble Constant. *ApJ*, 826(1):56, July 2016. doi: [10.3847/0004-637X/826/1/56](https://doi.org/10.3847/0004-637X/826/1/56).

- Riess, A.G., Casertano, S., Yuan, W., Bowers, J.B., Macri, L., Zinn, J.C. and Scolnic, D. Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with  $\Lambda$ CDM. *ApJ*, 908(1):L6, February 2021. doi: [10.3847/2041-8213/abdbaf](https://doi.org/10.3847/2041-8213/abdbaf).
- Rubin, V.C. and Ford, W. Kent, J. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *ApJ*, 159:379, February 1970. doi: [10.1086/150317](https://doi.org/10.1086/150317).
- Ruiz-Zapatero, J. et al. Geometry versus growth. Internal consistency of the flat  $\Lambda$ CDM model with KiDS-1000. *A&A*, 655:A11, November 2021. doi: [10.1051/0004-6361/202141350](https://doi.org/10.1051/0004-6361/202141350).
- Salvato, M., Ilbert, O. and Hoyle, B. The many flavours of photometric redshifts. *Nature Astronomy*, 3:212–222, June 2019. doi: [10.1038/s41550-018-0478-0](https://doi.org/10.1038/s41550-018-0478-0).
- Sánchez, A.G. et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the configuration-space clustering wedges. *MNRAS*, 464(2):1640–1658, January 2017. doi: [10.1093/mnras/stw2443](https://doi.org/10.1093/mnras/stw2443).
- Sánchez, C. and Bernstein, G.M. Redshift inference from the combination of galaxy colours and clustering in a hierarchical Bayesian model. *MNRAS*, 483(2):2801–2813, February 2019. doi: [10.1093/mnras/sty3222](https://doi.org/10.1093/mnras/sty3222).
- Schaan, E., Ferraro, S. and Seljak, U. Photo-z outlier self-calibration in weak lensing surveys. *J. Cosmology Astropart. Phys.*, 2020(12):001, December 2020. doi: [10.1088/1475-7516/2020/12/001](https://doi.org/10.1088/1475-7516/2020/12/001).
- Schneider, P., van Waerbeke, L., Kilbinger, M. and Mellier, Y. Analysis of two-point statistics of cosmic shear. I. Estimators and covariances. *A&A*, 396:1–19, December 2002. doi: [10.1051/0004-6361:20021341](https://doi.org/10.1051/0004-6361:20021341).

- Schneider, P., Eifler, T. and Krause, E. COSEBIs: Extracting the full E-/B-mode information from cosmic shear correlation functions. *A&A*, 520:A116, September 2010. doi: [10.1051/0004-6361/201014235](https://doi.org/10.1051/0004-6361/201014235).
- Scolnic, D.M. et al. The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. *ApJ*, 859(2):101, June 2018. doi: [10.3847/1538-4357/aab9bb](https://doi.org/10.3847/1538-4357/aab9bb).
- Scoville, N. et al. The Cosmic Evolution Survey (COSMOS): Overview. *ApJS*, 172(1):1–8, September 2007. doi: [10.1086/516585](https://doi.org/10.1086/516585).
- Sevilla-Noarbe, I. et al. Dark Energy Survey Year 3 Results: Photometric Data Set for Cosmology. *ApJS*, 254(2):24, June 2021. doi: [10.3847/1538-4365/abeb66](https://doi.org/10.3847/1538-4365/abeb66).
- Shannon, C.E. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi: [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x).
- Smoot, G.F. et al. Structure in the COBE Differential Microwave Radiometer First-Year Maps. *ApJ*, 396:L1, September 1992. doi: [10.1086/186504](https://doi.org/10.1086/186504).
- Speagle, J.S. and Eisenstein, D.J. Deriving photometric redshifts using fuzzy archetypes and self-organizing maps - I. Methodology. *MNRAS*, 469(1):1186–1204, July 2017. doi: [10.1093/mnras/stw1485](https://doi.org/10.1093/mnras/stw1485).
- Spergel, D. et al. Wide-Field Infrared Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report. *arXiv e-prints*. [arXiv:1503.03757](https://arxiv.org/abs/1503.03757), March 2015.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002. doi: <https://doi.org/10.1111/1467-9868.00353>.

- Spurio Mancini, A., Piras, D., Alsing, J., Joachimi, B. and Hobson, M.P. COSMOPOWER: emulating cosmological power spectra for accelerated Bayesian inference from next-generation surveys. *MNRAS*, 511(2):1771–1788, April 2022. doi: [10.1093/mnras/stac064](https://doi.org/10.1093/mnras/stac064).
- Stoica, R.S., Martinez, V.J. and Saar, E. A three dimensional object point process for detection of cosmic filaments. *arXiv e-prints*. arXiv:0809.4358, September 2008. doi: [10.48550/arXiv.0809.4358](https://doi.org/10.48550/arXiv.0809.4358).
- Stözlner, B., Joachimi, B., Korn, A., Hildebrandt, H. and Wright, A.H. Self-calibration and robust propagation of photometric redshift distribution uncertainties in weak gravitational lensing. *A&A*, 650:A148, June 2021. doi: [10.1051/0004-6361/202040130](https://doi.org/10.1051/0004-6361/202040130).
- Stözlner, B., Joachimi, B., Korn, A. and LSST Dark Energy Science Collaboration. Optimizing the shape of photometric redshift distributions with clustering cross-correlations. *MNRAS*, 519(2):2438–2450, February 2023. doi: [10.1093/mnras/stac3630](https://doi.org/10.1093/mnras/stac3630).
- Taylor, A.N. and Kitching, T.D. Analytic methods for cosmological likelihoods. *MNRAS*, 408(2):865–875, Oct 2010. doi: [10.1111/j.1365-2966.2010.17201.x](https://doi.org/10.1111/j.1365-2966.2010.17201.x).
- Tegmark, M. and Peebles, P.J.E. The Time Evolution of Bias. *ApJ*, 500(2): L79–L82, June 1998. doi: [10.1086/311426](https://doi.org/10.1086/311426).
- Tempel, E., Stoica, R.S., Martínez, V.J., Liivamägi, L.J., Castellan, G. and Saar, E. Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the SDSS. *MNRAS*, 438(4):3465–3482, March 2014. doi: [10.1093/mnras/stt2454](https://doi.org/10.1093/mnras/stt2454).
- Tempel, E., Stoica, R.S., Kipper, R. and Saar, E. Bisous model-Detecting filamentary patterns in point processes. *Astronomy and Computing*, 16:17–25, July 2016. doi: [10.1016/j.ascom.2016.03.004](https://doi.org/10.1016/j.ascom.2016.03.004).

- Tröster, T. et al. KiDS-1000 Cosmology: Constraints beyond flat  $\Lambda$ CDM. *A&A*, 649:A88, May 2021. doi: [10.1051/0004-6361/202039805](https://doi.org/10.1051/0004-6361/202039805).
- Troxel, M.A. et al. Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear. *Phys. Rev. D*, 98(4):043528, August 2018. doi: [10.1103/PhysRevD.98.043528](https://doi.org/10.1103/PhysRevD.98.043528).
- van den Busch, J.L., Hildebrandt, H., Wright, A.H., Morrison, C.B., Blake, C., Joachimi, B., Erben, T., Heymans, C., Kuijken, K. and Taylor, E.N. Testing KiDS cross-correlation redshifts with simulations. *A&A*, 642:A200, October 2020. doi: [10.1051/0004-6361/202038835](https://doi.org/10.1051/0004-6361/202038835).
- van Erp, S., Oberski, D.L. and Mulder, J. Shrinkage priors for bayesian penalized regression. *Journal of Mathematical Psychology*, 89:31–50, 2019. ISSN 0022-2496. doi: <https://doi.org/10.1016/j.jmp.2018.12.004>.
- van Uitert, E. et al. KiDS+GAMA: cosmology constraints from a joint analysis of cosmic shear, galaxy-galaxy lensing, and angular clustering. *MNRAS*, 476(4):4662–4689, June 2018. doi: [10.1093/mnras/sty551](https://doi.org/10.1093/mnras/sty551).
- Van Waerbeke, L. et al. Detection of correlated galaxy ellipticities from CFHT data: first evidence for gravitational lensing by large-scale structures. *A&A*, 358:30–44, June 2000.
- Vanderbilt, D. and Louie, S.G. A monte carlo simulated annealing approach to optimization over continuous variables. *Journal of Computational Physics*, 56(2):259–271, 1984. ISSN 0021-9991. doi: [https://doi.org/10.1016/0021-9991\(84\)90095-0](https://doi.org/10.1016/0021-9991(84)90095-0).
- Way, M.J. and Klose, C.D. Can Self-Organizing Maps Accurately Predict Photometric Redshifts? *PASP*, 124(913):274, March 2012. doi: [10.1086/664796](https://doi.org/10.1086/664796).
- Wittek, P., Gao, S.C., Lim, I.S. and Zhao, L. somoclu: An efficient parallel library for self-organizing maps. *Journal of Statistical Software*, 78(9):1–21, 2017. doi: [10.18637/jss.v078.i09](https://doi.org/10.18637/jss.v078.i09).

- Wittman, D.M., Tyson, J.A., Kirkman, D., Dell'Antonio, I. and Bernstein, G. Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales. *Nature*, 405(6783):143–148, May 2000. doi: [10.1038/35012001](https://doi.org/10.1038/35012001).
- Wright, A.H. et al. KiDS+VIKING-450: A new combined optical and near-infrared dataset for cosmology and astrophysics. *A&A*, 632:A34, December 2019. doi: [10.1051/0004-6361/201834879](https://doi.org/10.1051/0004-6361/201834879).
- Wright, A.H., Hildebrandt, H., van den Busch, J.L. and Heymans, C. Photometric redshift calibration with self-organising maps. *A&A*, 637:A100, May 2020a. doi: [10.1051/0004-6361/201936782](https://doi.org/10.1051/0004-6361/201936782).
- Wright, A.H., Hildebrandt, H., van den Busch, J.L., Heymans, C., Joachimi, B., Kannawadi, A. and Kuijken, K. KiDS+VIKING-450: Improved cosmological parameter constraints from redshift calibration with self-organising maps. *A&A*, 640:L14, August 2020b. doi: [10.1051/0004-6361/202038389](https://doi.org/10.1051/0004-6361/202038389).
- Zehavi, I. et al. The Luminosity and Color Dependence of the Galaxy Correlation Function. *ApJ*, 630(1):1–27, September 2005. doi: [10.1086/431891](https://doi.org/10.1086/431891).
- Zhang, P., Pen, U.L. and Bernstein, G. Self-calibration of photometric redshift scatter in weak-lensing surveys. *MNRAS*, 405(1):359–374, June 2010. doi: [10.1111/j.1365-2966.2010.16445.x](https://doi.org/10.1111/j.1365-2966.2010.16445.x).
- Zuntz, J. et al. Dark Energy Survey Year 1 results: weak lensing shape catalogues. *MNRAS*, 481(1):1149–1182, November 2018. doi: [10.1093/mnras/sty2219](https://doi.org/10.1093/mnras/sty2219).
- Zuntz, J. et al. The LSST-DESC 3x2pt Tomography Optimization Challenge. *The Open Journal of Astrophysics*, 4(1):13, October 2021. doi: [10.21105/astro.2108.13418](https://doi.org/10.21105/astro.2108.13418).
- Zwicky, F. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, January 1933.