Cosmic tomography with weak gravitational lensing

Li, S.S.

Cosmic tomography with weak gravitational lensing

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 25 oktober 2023
klokke 10:00 uur

door

Shun-Sheng Li

geboren te Zhejiang, China
in 1994

Promotores:	Prof. dr. K.H. Kuijken	
	Prof. dr. H. Hoekstra	
Promotiecommissie:	Prof. dr. I.A.G. Snellen	
	Prof. dr. H.J.A. Röttgering	
	Prof. dr. M. Franx	
	Prof. dr. C. Heymans	(University of Edinburgh)
	Dr. N.E. Chisari	(Universiteit Utrecht)

ISBN: 978-94-6419-936-9

A digital version of this thesis is available at `https://openaccess.leidenuniv.nl`.

Cover: It is a composite of astronomical images, combining real sky images from KiDS with simulated ones from SKiLLS. Interspersed within are artificial elements demonstrating weak lensing shear effects, name tags for KiDS and SKiLLS, and additional text. The images were generated or combined using the public GalSim package. Can you distinguish the real galaxies from the simulated ones?

There is no great chance of observing this phenomenon.
— Albert Einstein (1936)

CONTENTS

1 Introduction **1**
- 1.1 Foundations of the modern cosmology . . . 3
  - 1.1.1 The homogeneous and isotropic Universe . . . 3
  - 1.1.2 Structure evolution . . . 5
  - 1.1.3 Testing the standard cosmological model . . . 10
- 1.2 Weak gravitational lensing . . . 14
  - 1.2.1 Fundamentals of gravitational lensing . . . 14
  - 1.2.2 Galaxy-galaxy lensing . . . 19
  - 1.2.3 Cosmic shear . . . 20
  - 1.2.4 Challenges in weak lensing analysis . . . 22
- 1.3 This thesis . . . 25

2 KiDS+VIKING-450: An internal-consistency test for cosmic shear tomography with a colour-based split of source galaxies **27**
- 2.1 Introduction . . . 28
- 2.2 Data . . . 29
- 2.3 Calibration of redshift distributions . . . 30
- 2.4 Calibration of shape measurements . . . 34
- 2.5 Cosmic shear signal . . . 37
  - 2.5.1 Statistical measures . . . 37
  - 2.5.2 Theoretical modelling . . . 38
- 2.6 Consistency tests . . . 40
  - 2.6.1 Covariance matrix . . . 42
  - 2.6.2 Test setup . . . 43
- 2.7 Results . . . 45
- 2.8 Summary and discussion . . . 46

3 KiDS-Legacy calibration: Unifying shear and redshift calibration with the SKiLLS multi-band image simulations **51**
- 3.1 Introduction . . . 52
- 3.2 Input mock catalogues . . . 54
  - 3.2.1 Galaxies: SURFS-Shark simulations with COSMOS morphology 55
  - 3.2.2 Stars: Point objects with synthetic photometry . . . 60
- 3.3 KiDS+VIKING nine-band image simulations . . . 60

3.3.1 KiDS-like optical images . . . 60
3.3.2 VIKING-like infrared images . . . 65
3.3.3 SKiLLS fiducial setup . . . 66
3.3.4 Photometry and photometric redshifts . . . 67
3.4 Shape measurements with the updated *lens*fit . . . 73
3.4.1 The self-calibration version of *lens*fit . . . 73
3.4.2 Updates for KiDS-Legacy analysis . . . 74
3.4.3 Comparison between KiDS and SKiLLS . . . 77
3.5 Shear biases for the updated *lens*fit . . . 77
3.5.1 Results from the constant shear simulations . . . 79
3.5.2 Impact of blends at different redshifts . . . 83
3.5.3 PSF modelling bias . . . 89
3.5.4 Results . . . 90
3.6 Sensitivity analysis . . . 91
3.6.1 Impact of observational conditions . . . 91
3.6.2 Impact of the input galaxy morphology . . . 93
3.7 Discussion and conclusions . . . 94
3.A An empirical modification to the synthetic photometry . . . 98
3.B Modelling multivariate distributions with vine copulas . . . 99
3.C Transformation of the SDSS filters to the KiDS/VIKING filters . . . 101
3.D Selection criteria for the updated *lens*fit catalogue . . . 102
3.E Building the variable shear field . . . 105

4 KiDS-1000: Cosmology with improved cosmic shear measurements 109
4.1 Introduction . . . 110
4.2 Updated weak lensing shear catalogue . . . 112
4.2.1 Galaxy shapes measured with the updated *lens*fit . . . 112
4.2.2 Validation of the shear estimates . . . 113
4.3 Shear and redshift calibration . . . 118
4.3.1 Calibration . . . 119
4.3.2 Calibration uncertainties . . . 121
4.4 Cosmological inference . . . 122
4.5 Results . . . 126
4.5.1 Fiducial analysis results . . . 126
4.5.2 Impact of shear biases . . . 129
4.5.3 Impact of altering inference setups . . . 130
4.6 Summary . . . 136
4.A Shear bias in two-point statistics . . . 139
4.B SOM redshift calibration . . . 139
4.C Systematic uncertainties from the shear calibration . . . 141
4.C.1 Quantifying residual shear biases with sensitivity analysis . . . 141
4.C.2 Propagating residual shear biases with forward modelling . . . 144
4.D Contour plots for all free parameters . . . 145

5 Weak lensing mass-luminosity scaling relations for galaxy groups: Testing the robustness of the halo model formalism **147**
5.1 Introduction . . . 148
5.2 Data . . . 150
5.2.1 Lenses: GAMA groups . . . 150
5.2.2 Sources: KiDS galaxies . . . 151
5.3 Galaxy-galaxy weak lensing signals . . . 152
5.3.1 ESD measurements . . . 153
5.3.2 Covariance matrix estimation . . . 155
5.4 Halo model and occupation statistics . . . 156
5.4.1 Halo model formalism . . . 157
5.4.2 Model ingredients . . . 158
5.4.3 Halo occupation statistics . . . 160
5.4.4 Model fitting . . . 161
5.5 Results from the fiducial model . . . 161
5.5.1 Stacked halo masses . . . 162
5.5.2 Group luminosity-halo mass relation . . . 162
5.6 Model sensitivity analyses . . . 165
5.6.1 Sensitivity to the incompleteness . . . 165
5.6.2 Sensitivity to the mis-centring parameters . . . 166
5.7 Conclusions . . . 168
5.A Posterior distributions of all free model parameters . . . 170

Bibliography **173**

Nederlandstalige samenvatting **185**

Publications during PhD study **189**

Curriculum vitae **191**

Acknowledgements **193**

CHAPTER 1

Introduction

Approximately 13.8 billion years ago (Planck Collaboration et al. 2020), an event known as the 'Big Bang' marked the inception of the physical Universe. As the Universe continues to expand, its energy density decreases, leading to the separation of the four fundamental forces and the cooling of particles. Eventually, the temperature fell to such a level that the initial density fluctuations could grow by gravitationally attracting surrounding matter, resulting in the formation of gas clouds, stars, galaxies, and the large-scale structures we observe today.

Unveiling this big picture of the history of our Universe, as depicted in Figure 1.1, stands as one of the most significant accomplishments of twentieth-century cosmological research, if not the most significant. The theoretical and observational advancements in the field are truly remarkable. This is especially evident with the establishment of the spatially flat Λ Cold Dark Matter (ΛCDM) model, which hinges on only six parameters and yet successfully describes several key observations with astounding accuracy, including the expansion of the Universe, the measurements of the Cosmic Microwave Background (CMB), and the matter distribution of cosmic large-scale structures.

As we move into the 21st century, the field of cosmology continues to thrive, intertwined with advancements in observational techniques and the subsequent challenges in accurately interpreting increasingly precise measurements. On the one hand, the development of various cosmological probes, facilitated by modern technology, yields unprecedented precision in our measurements. This progress provides a massive opportunity to unravel the mysteries surrounding some critical elements of our current cosmological model, such as the origins of initial density fluctuations and the enigmatic nature of dark matter and dark energy. On the other hand, the technical advancements come with the challenging task of controlling various systematic effects that stem from both observational conditions and astrophysical contamination. Therefore, devising effective strategies to control these systematics, in order to harness the full statistical power of advanced cosmological probes, remains a pivotal topic in 21st-century cosmological studies.

Among the techniques that show promise as cosmological probes but present practical challenges is weak gravitational lensing. According to general relativity, light

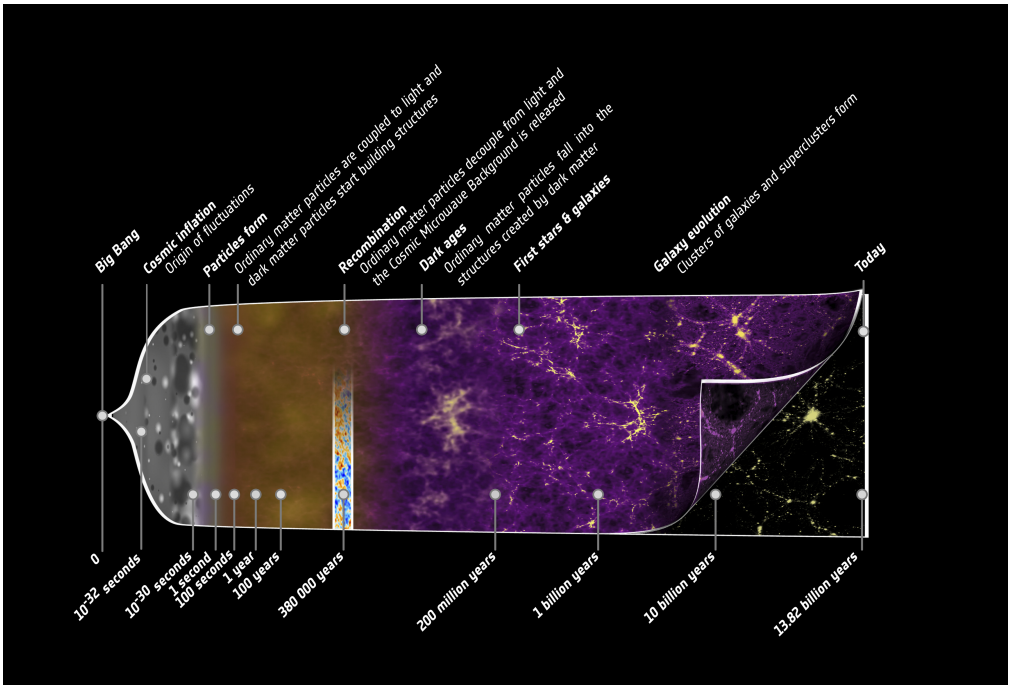

Figure 1.1: An overview of the history of our Universe, from its initial state—characterised by near uniformity, punctuated by small fluctuations—to the complex cosmic structure we observe today. [Credit: ESA – C. Carreau]

from distant galaxies is distorted by the gravitational fields associated with foreground matter. Consequently, measuring these small yet coherent distortions offers a direct probe into the matter distribution in large-scale structures, or those associated with individual massive objects, such as galaxy groups or clusters. However, in practice, robustly measuring these small lensing-induced distortions is difficult due to distortions introduced by observational conditions and instrumental effects. The task is further complicated by the challenges involved in accurately determining the distances to the galaxies. Even with robust weak lensing measurements in hand, extracting cosmological information from these measured signals is non-trivial due to astrophysical effects like the intrinsic alignment of galaxies and baryonic effects.

This thesis focuses on this very topic, with a specific emphasis on the role of weak gravitational lensing in deriving cosmological parameters and investigating dark matter halos. Our primary objective is to enhance the accuracy of weak lensing analyses by improving the calibration of signal measurements and scrutinising the inference pipeline through consistency and sensitivity tests. Additionally, we aim to deepen our understanding of the relationship between dark matter halos and galaxies. This relationship is critical not only for achieving robust cosmological inference but also for enhancing our understanding of galaxy formation and evolution.

1.1 Foundations of the modern cosmology

Before delving into the cosmological applications of weak lensing, it is instructive to review the foundational elements of contemporary cosmological models. For the sake of brevity and specificity, we focus on the current standard cosmological model, namely the ΛCDM model. We begin with an examination of the isotropic and homogeneous average Universe in Sect. 1.1.1, followed by an analysis of the evolution of inhomogeneities at small scales in Sect. 1.1.2. Finally, we conclude with a discussion of key observational tests to the standard cosmological model in Sect. 1.1.3.

1.1.1 The homogeneous and isotropic Universe

The standard cosmological model is built upon two main foundations: the cosmological principle and the validity of general relativity at cosmological scales. The cosmological principle states that, on sufficiently large scales, the average matter distribution is both homogeneous and isotropic. This assumption leads to the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, which defines the geometry of space-time as follows

$$\mathrm{d}s^2 \equiv -g_{\mu\nu}\mathrm{d}x^\mu x^\nu = c^2\mathrm{d}t^2 - a^2(t)\mathrm{d}l^2 \,, \tag{1.1}$$

where c stands for the speed of light. The metric tensor, $g_{\mu\nu}$, establishes the relationship between the coordinate values and the interval $\mathrm{d}s$ in the four-dimensional space-time. The scale factor, $a(t)$, captures the expansion of space. The comoving spatial element, $\mathrm{d}l$, is then defined as

$$\mathrm{d}l^2 \equiv \left[\frac{\mathrm{d}r^2}{1-Kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta\ \mathrm{d}\phi^2)\right] \,, \tag{1.2}$$

where r, θ, and ϕ represent comoving spatial polar coordinates, and K is the curvature parameter. A K value of zero corresponds to a spatially flat Euclidean space, $K > 0$ to a positively curved, spherical space, and $K < 0$ to a negatively curved, hyperbolic space.

The relationship between the space-time geometry and the energy content in the Universe is described by the Einstein field equation:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} \,. \tag{1.3}$$

Here, $R_{\mu\nu}$ and R denote the Ricci tensor and scalar, respectively, which describe the local curvature of space-time. The cosmological constant, Λ, was initially introduced by Einstein to achieve a static Universe, but it is now used to explain the observed accelerating expansion of the Universe. G is the Newtonian gravitational constant, and $T_{\mu\nu}$ represents the energy-momentum tensor.

In the case of a perfect fluid—an apt approximation for the average matter distribution at large scales under the cosmological principle—the energy-momentum tensor is formed by a simple diagonal matrix $T^\mu{}_\nu = \mathrm{diag}(\rho c^2, -P, -P, -P)$, where ρc^2 represents the

energy density, and P denotes pressure. By inserting this energy-momentum tensor and the FLRW metric into the Einstein field equation, we obtain the Friedmann equations, which describe the dynamics of the cosmological expansion:

$$\left[\frac{\dot{a}(t)}{a(t)}\right]^2 = \frac{8\pi G}{3}\rho(t) - \frac{c^2}{a^2(t)}K + \frac{c^2}{3}\Lambda\,, \tag{1.4}$$

and

$$\frac{\ddot{a}(t)}{a(t)} = -\frac{4\pi G}{3}\left(\rho + \frac{3}{c^2}P(t)\right) + \frac{c^2}{3}\Lambda\,. \tag{1.5}$$

The connection between the cosmological constant and the energy content is apparent from these equations. The influence of the cosmological constant is equivalent to an energy component with a constant density and pressure that satisfy $\rho_\Lambda = -P_\Lambda/c^2 = \Lambda c^2/(8\pi G)$. As such, the cosmological constant is also regarded as the simplest form of the mysterious dark energy, which is considered to be driving the observed accelerating expansion of the Universe.

Broadly speaking, the matter components in the Universe can be categorised into two groups: non-relativistic matter, which includes cold dark matter and ordinary baryons, and relativistic matter, consisting of photons and relativistic neutrinos. These two types of matter show different pressure-density relationships. For a perfect fluid, this relationship is characterised by the equation of state parameter w, defined as $w \equiv P/(\rho c^2)$. For non-relativistic matter, the pressure is negligible, corresponding to a $w_\mathrm{m} = 0$. On the other hand, relativistic matter has a $w_\mathrm{rad} = 1/3$. In similar terms, the cosmological constant can be interpreted as a dark energy component with a $w_\Lambda = -1$.

These equations of state enable us to re-formulate the Friedmann equations into the following form:

$$H^2(t) = H_0^2\left[\Omega_{\Lambda,0} + \Omega_{\mathrm{m},0}\,a^{-3}(t) + \Omega_{\mathrm{rad},0}\,a^{-4}(t)\right] - \frac{c^2}{a^2(t)}K\,, \tag{1.6}$$

where the Hubble parameter, $H(t) \equiv \dot{a}(t)/a(t)$, is introduced. Ω_i is defined as the ratio of the density ρ_i to the critical density $\rho_\mathrm{crit}(t) \equiv 3H^2(t)/(8\pi G)$. The subscript 0 indicates values measured at the current epoch ($a(t_0) = 1$).

The Hubble parameter at the current epoch, H_0, is also known as the Hubble constant, and it measures the current rate of cosmological expansion. Historically, the exact value of the Hubble constant has been uncertain, with estimates ranging from $50 - 100$ km s^{-1} Mpc^{-1}. Consequently, it is common to introduce a little h factor, defined as $H_0 = 100h$ km s^{-1} Mpc^{-1}, to absorb the uncertainties in the exact value of H_0. This convention will be used throughout this chapter.

The cosmological parameters in Eq. (1.6) were precisely determined through the *Planck* CMB observations, although some discrepancies persist among results from different cosmological probes (as discussed in Sect. 1.1.3). By combining information from temperature and polarisation maps, as well as lensing reconstruction, and adopting a spatially flat ΛCDM model ($K = 0$), Planck Collaboration et al. (2020) constrained

$H_0 = 67.36 \pm 0.54$ km s^{-1} Mpc^{-1}, $\Omega_{\Lambda,0} = 0.6847 \pm 0.0073$, $\Omega_{\mathrm{m},0} = 0.3153 \pm 0.0073$, and $\Omega_{\mathrm{rad},0} < 10^{-4}$.

1.1.2 Structure evolution

An entirely homogeneous and isotropic Universe would not spontaneously form the complex cosmological structures we observe today. The existence of these large-scale structures, such as groups and clusters of galaxies, implies the presence of initial density fluctuations. These early cosmological inhomogeneities, serving as the seeds of our current observed structures, grew due to gravitational instability. While the origin of these primordial density fluctuations remains a subject of debate — with the current popular theories suggesting quantum fluctuations associated with inflation or topological defects from early-time Universe phase transitions — the evolution of these fluctuations is well-described by the current standard cosmological model. In fact, when these primordial cosmological perturbations are considered as initial conditions, the standard ΛCDM model accurately predicts both the observed anisotropy in the CMB and the matter distribution in the late-time Universe. We briefly review some key aspects of structure formation within the framework of the standard ΛCDM model in this section.

Density fluctuations and two-point statistical measures

Typically, the cosmological density fluctuations are described by a perturbation field, δ, defined as the contrast between the local density and the mean background density as follows:

$$\delta(\boldsymbol{x}) \equiv \frac{\rho(\boldsymbol{x})}{\langle\rho\rangle} - 1 \,. \tag{1.7}$$

Here, $\langle\cdots\rangle$ denotes the averaging operator, and $\boldsymbol{x}$ represents the comoving spatial coordinate. It is also convenient to consider the perturbation field as a superposition of many signal modes, which in a flat comoving geometry, is quantified by the Fourier transform:

$$\delta(\boldsymbol{x}) = \sum_{\boldsymbol{k}} \delta_{\boldsymbol{k}} \exp(i\boldsymbol{k}\cdot\boldsymbol{x}) \,; \qquad \delta_{\boldsymbol{k}} = \frac{1}{V}\int \mathrm{d}^3\boldsymbol{x}\, \exp(-i\boldsymbol{k}\cdot\boldsymbol{x})\, \delta(\boldsymbol{x}) \,, \tag{1.8}$$

where V denotes the volume of a sufficiently large box, inside which the perturbations are assumed to be periodic, and $\boldsymbol{k}$ represents the wavevector. The cosmological principle implies that while the perturbation field characterises inhomogeneities, its statistical properties should exhibit homogeneity and isotropy like the background density field (see Chapter 16 of Peacock 1999 for a detailed discussion). This enables us to simplify calculations by reducing the three-dimensional coordinates of $\boldsymbol{x}$ and the wavevector $\boldsymbol{k}$ to single dimensions, x and k respectively.

Given the stochastic nature of the initial fluctuations, the meaningful predictions about the cosmological structure concern its statistical properties. Thus, it is necessary to introduce some statistical measures of the perturbation field. One of the basic measures

is the two-point correlation function, defined as

$$\xi(r) \equiv \langle \delta(x)\delta(x+r) \rangle \,, \tag{1.9}$$

where r denotes the separation between two spatial points in comoving units. This correlation function quantifies the clumpiness of the density field, and thus can serve as a means to characterise the cosmological structure. An equivalent measure in the Fourier space is the power spectrum, which is the Fourier transform of the two-point correlation function:

$$P(k) \equiv V\langle |\delta_k|^2 \rangle = 4\pi \int \mathrm{d}r\, \xi(r) \frac{\sin(kr)}{kr} r^2 \,, \tag{1.10}$$

In this equation, we integrated the angular coordinates in the 3D Fourier transform (as defined in Eq. 1.8), based on the assumption that the perturbation field is homogeneous and isotropic.

Assuming that the initial fluctuations follow a Gaussian random field, a condition which aligns with current observational data and the predictions of the popular inflation theory, the power spectrum or two-point correlation function entirely captures the statistical properties of the field. However, for non-Gaussian random fields, which could develop at later times due to non-linear dynamics, higher-order correlation functions become necessary. For the sake of brevity, we shall not extend our discussion to these higher-order statistics in this short introduction.

Linear evolution in the early stage

During the early stage of structure evolution, fluctuations are typically small ($\delta_k \ll 1$). In these circumstances, we can employ a perturbative approach to analytically solve the evolution problem, subject to certain approximations. The choice between Newtonian mechanics or a relativistic treatment is influenced by the properties of the matter content and the size of the perturbation. The simplest case involves sub-horizon perturbations—fluctuations on scales much smaller than the horizon size—in pressureless matter, which comprises cold dark matter and non-relativistic baryons.

In this scenario, the Newtonian perturbation theory is applicable. Moreover, if we disregard the entropy perturbation, the evolution of each Fourier mode of these perturbations can be described by the equation:

$$\ddot{\delta}_k + 2H(t)\dot{\delta}_k = 4\pi G\langle \rho \rangle \delta_k \,. \tag{1.11}$$

In this equation, a damping term $2H(t)\dot{\delta}_k$ arises due to the cosmological expansion, resulting in the growth of the fluctuations following a power law in time, rather than an exponential growth. This slow growth rate enables probing the primordial fluctuations using observations from the present epoch, as the evolved matter distribution still retains significant information about the initial state.

Non-linear evolution and gravitational collapse

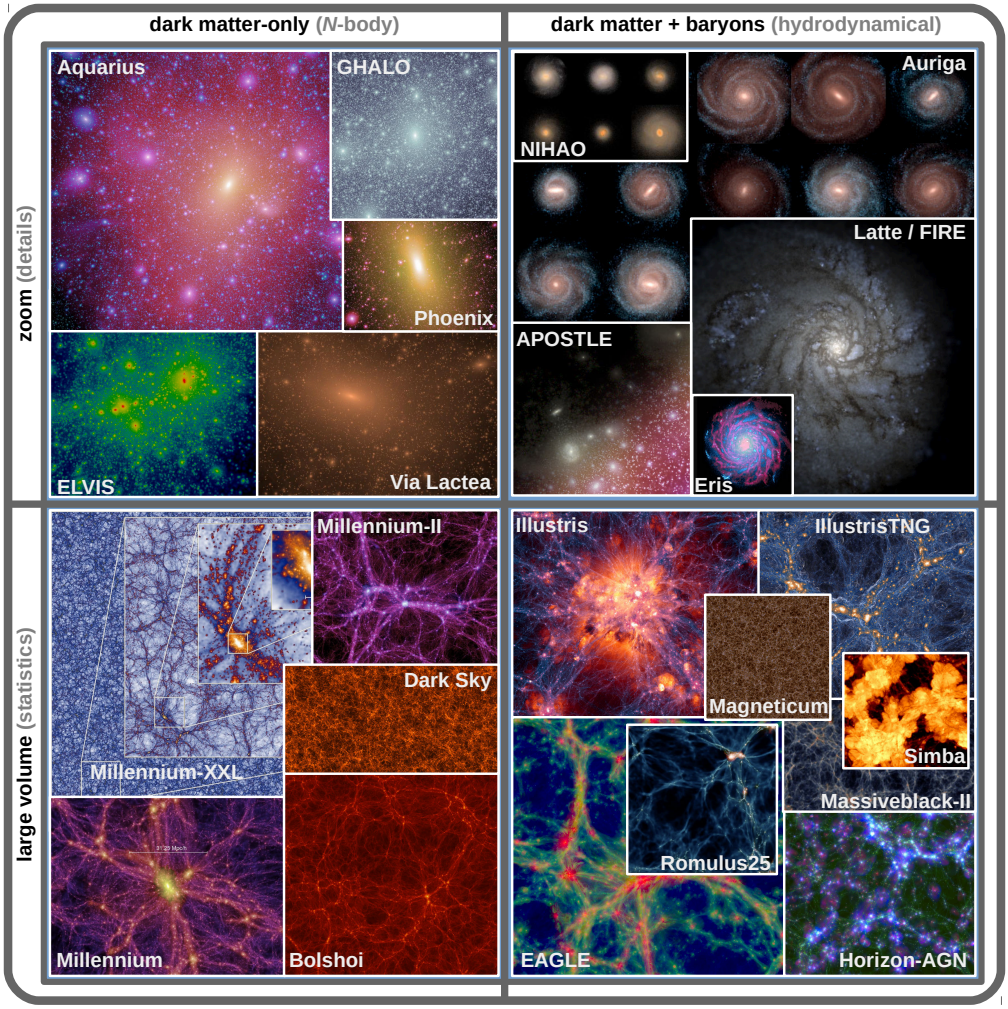

Figure 1.2: Visual representations of some numerical simulations. The simulations are categorised into large-volume simulations, which statistically represent large-scale structures, and zoom-in simulations, focusing on resolving smaller scales. Each of these types is further classified into two classes: dark matter-only N-body simulations, which consider only gravitational effects, and hydrodynamical simulations, which also include non-gravitational processes such as gas cooling and feedback mechanisms. [Credit: Vogelsberger et al. (2020)]

As the fluctuations continue to grow, the linear approximation becomes less and less accurate, and eventually a full non-linear treatment of gravitational collapse is required to effectively study galaxy formation. While the underlying physical principles remain the same, finding analytical solutions becomes increasingly challenging or even impossible

due to the complexity of these non-linear dynamical systems. This is where numerical simulations prove indispensable. Figure 1.2 showcases visual representations from some recent simulations. Such cosmological simulations are essential for understanding the intricate process of structure and galaxy formation, and they serve as key tools in interpreting observations from modern cosmological surveys.

While non-linear dynamics are complex, valuable insights can still be drawn from simplified analytical models. Comparing non-linear results with their linear counterparts can provide a qualitative understanding of the evolution of cosmic structures. A useful example involves the evolution of spherically symmetric, sub-horizon fluctuations in pressureless matter. For these small-scale fluctuations, the Newtonian equation $\ddot{r} = -GM/r^2$ is applicable, where r is the radius of a given mass shell, and M is the mass enclosed within that shell. Assuming that M remains constant over time and that the mass shell initially expands but ultimately collapses due to a negative net energy, we can solve the Newtonian equation to find the times of maximum expansion ($\dot{r} = 0$) and collapse ($r = 0$). Substituting these times into the formula derived from the linear treatment gives corresponding linear perturbation values of approximately 1.06 and 1.69, respectively. While these values do not match the actual amplitude of the fluctuations, they allow us to use linear perturbation theory to predict the timing of the turn-around and collapse.

The final density of the collapsed objects cannot be directly estimated from the aforementioned simplified model because the Newtonian equation diverges when $r \to 0$. This divergence arises because the assumption of a constant M breaks down when r is small and particles move across different mass shells. Ultimately, the kinetic energy from the random motion of particles balances the gravitational potential energy, and the collapsed system reaches virial equilibrium. Consequently, the virial theorem can be used to estimate the final overdensity of the collapsed system. Under the assumption of a matter-dominated Universe, the overdensity of the collapsed system in the spherical collapse model is found to be $\Delta_c = 18\pi^2 \approx 178$. Although this value is dependent on the idealised assumptions of the spherical collapse model, it implies that virialised regions can be effectively described by a sphere with an average density roughly 200 times the mean density of the Universe. This definition of virialised objects continues to be widely used in contemporary studies.

Smoothed perturbation filed and Press-Schechter formalism

A notable application of the aforementioned simple spherical collapse model is the Press-Schechter formalism proposed by Press & Schechter (1974). This method predicts the mass function of collapsed objects from the initial perturbation field, bypassing the need for detailed non-linear dynamic calculations. The basic idea is to evolve the perturbation field using linear theory and connect it to the mass function of collapsed objects. This connection is established by using a smoothed version of the perturbation

field, defined as

$$\Delta(\boldsymbol{x}; R) \equiv \int \mathrm{d}^3\boldsymbol{x}'\ \delta(\boldsymbol{x}')W(\boldsymbol{x}+\boldsymbol{x}'; R)\ ; \qquad \int \mathrm{d}^3\boldsymbol{x}\ W(\boldsymbol{x}; R)\ = 1, \tag{1.12}$$

where $W(\boldsymbol{x}; R)$ is a spherical window function with a characteristic radius R. One of the most commonly used window functions is the top-hat window function defined as

$$W(\boldsymbol{x}; R) = \begin{cases} \left(\frac{4\pi}{3}R^3\right)^{-1} & \text{if } |\boldsymbol{x}| \leq R \\ 0 & \text{otherwise} \end{cases}. \tag{1.13}$$

With this window function, it is straightforward to establish a connection between the characteristic radius and the average mass contained within the window volume: $M = 4\pi R^3\langle\rho\rangle/3$. A similar correlation between the characteristic radius and mass can be established for any chosen window function.

In practice, a top-hat window function with a radius of $8h^{-1}$ Mpc is often used because the variance of the cosmic structure at present day, when smoothed with this function, approximates one. This window function, when applied to smooth the present linear perturbation field—computed by evolving the initial perturbation field to the present day using the linear treatment—yields a root-mean-square measure known as σ_8, which is commonly used to quantify the strength of cosmic density fluctuations.

The Press-Schechter formalism assumes that the probability $\Delta(R) > \delta_\mathrm{c}$, denoted as $\mathcal{P}(>\delta_\mathrm{c})$, is equivalent to the fraction of collapsed objects with a mass greater than M, represented as $F(>M)$. Here, δ_c is the overdensity of collapsed objects inferred from the linear treatment. In the spherical collapse model discussed earlier, $\delta_\mathrm{c} \approx 1.69$. However, this assumption contains an inherent flaw due to the limits of linear treatment, which suggest only the regions initially overdense can result in collapsed objects. Consequently, this leads to a loss of half the total mass when transitioning from $\mathcal{P}(>\delta_\mathrm{c})$ to $F(>M)$.

To remedy this issue, Press & Schechter (1974) assumed without proof that the remaining mass in the underdense regions will be accreted into the collapsed objects, resulting in doubling all collapsed object masses but keeping the shape of the mass function unchanged, thus, $F(>M) = 2\mathcal{P}(>\delta_\mathrm{c})$. In the case of $\Delta(\boldsymbol{x}; R)$ following a Gaussian random field, the resulting Press-Schechter mass function is as follows:

$$n(M)\mathrm{d}M = \sqrt{\frac{2}{\pi}}\frac{\langle\rho\rangle}{M^2}\frac{\delta_\mathrm{c}}{\sigma_M}\exp\left(-\frac{\delta_\mathrm{c}^2}{2\sigma_M^2}\right)\left|\frac{\mathrm{d}\ln\sigma_M}{\mathrm{d}\ln M}\right|\mathrm{d}M\ , \tag{1.14}$$

where σ_M represents the dispersion of the smoothed overdensity field, as filtered by a window function corresponding to mass M.

Despite its reliance on several idealised assumptions and some unjustified ansatz, the Press-Schechter mass function effectively captures the general behaviour of the true mass function. Specifically, it correctly predicts the exponential decrease at the high mass end and the power-law increase at the low mass end. Further refinements of the

Press-Schechter formalism, including the extended Press-Schechter formalism by Bond et al. (1991) and the merger-tree method (e.g. Lacey & Cole 1993), continue to play significant roles in current hierarchical models of galaxy formation and serve as the backbone of modern semi-analytical models of galaxy formation.

1.1.3 Testing the standard cosmological model

Apart from theoretical developments, which primarily focus on parametrising the Universe with cosmological models and understanding structure formation, another crucial aspect of cosmological studies is the use of astronomical observations to test and refine these models. These observational endeavours play a significant role in falsifying established cosmological theories. With the advent of increasingly powerful cosmological surveys and the new observational window opened by the detection of gravitational waves, we can anticipate significant advancements in the coming decades. In this section, we provide a brief overview of some key cosmological probes that have played — and will continue to play — a central role in shaping our current standard cosmological model.

Determining the Hubble constant

The start of observational cosmology can be traced back to the pioneering work of Hubble (1929), who observationally confirmed the expansion of the Universe by measuring the recession of galaxies, as had been theoretically predicted by Lemaître (1927). The slope of the relationship between the distance to a galaxy and its recession velocity is now termed the Hubble constant H_0, which quantifies the current expansion rate of the Universe.

Accurately determining the value of H_0 remains an active area of research in observational cosmology. Figure 1.3 provides a summary of the current constraints on H_0 derived from various indirect and direct methods. The indirect method typically involves using CMB measurements to infer H_0 values based on assumed cosmological models. The direct method currently involves three approaches: the 'standard candle' method, which is based on bright objects with absolute luminosity that can be inferred from non-distance-related measurements; the 'standard siren' method based on gravitational waves, which enables a direct measurement of H_0 without forming any 'cosmic distance ladder'; and time-delay measurements from strongly lensed quasars.

Beyond the impressive precision of the current constraints from these varied methods, the most striking finding is the 4σ to 6σ discrepancy between the results from the indirect method and those from the direct method. This statistically significant H_0 tension triggered extensive interest in exploring potential systematic effects inherent in either method or in refining the current cosmological model (see Di Valentino et al. 2021 for a recent review).

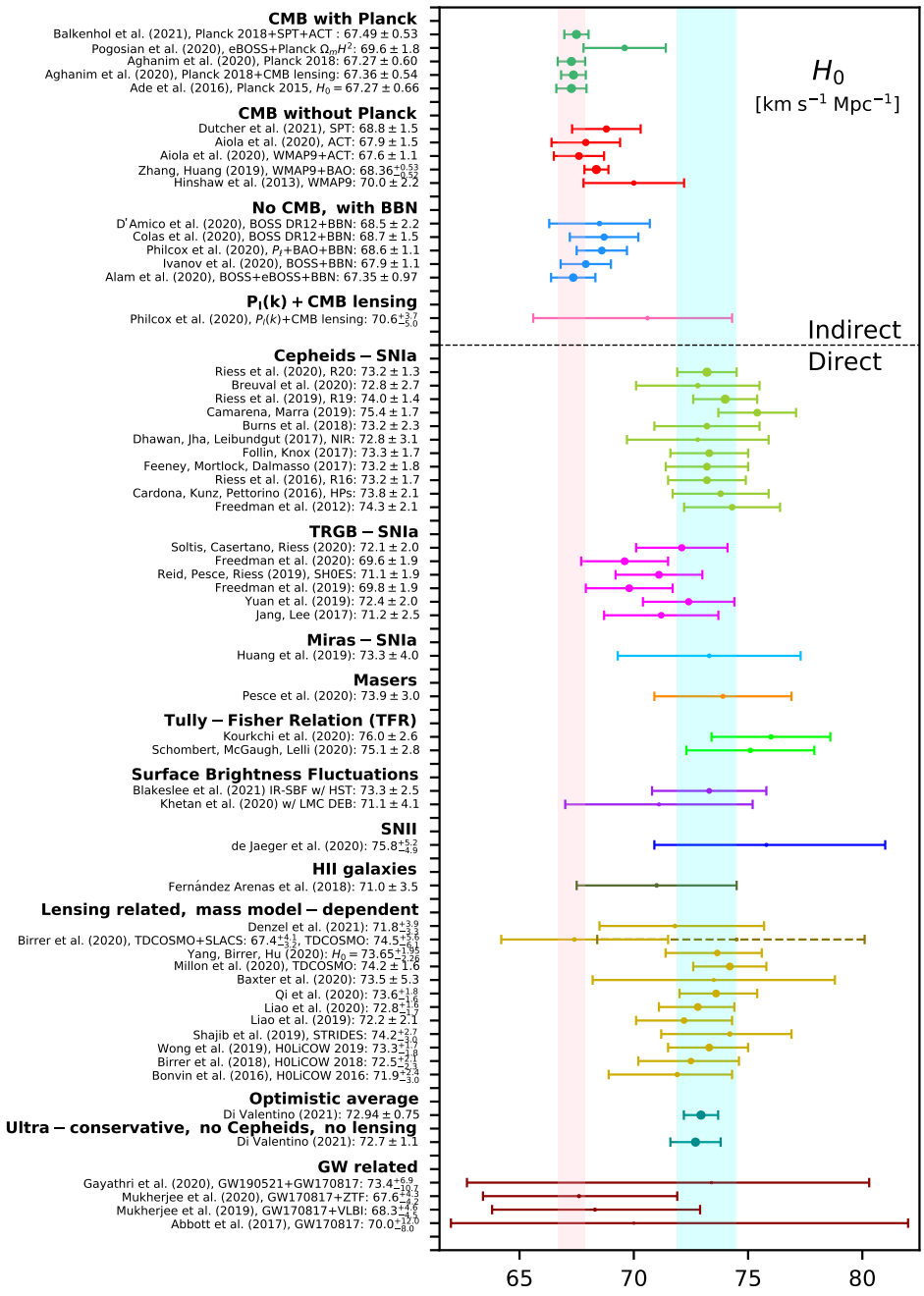

Figure 1.3: A summary plot showing the constraints on H_0 derived from various indirect and direct cosmological probes. The error bars denote the 68% credible intervals. The y-axis labels show the sources of the measurements, which are grouped based on their respective methods. [Credit: Di Valentino et al. (2021)]

Cosmic microwave background

The discovery of the CMB by Penzias & Wilson (1965) served as a crucial piece of evidence in support of the 'Big Bang' theory. The properties of the CMB observed today are consistent with predictions from this theory. According to the 'Big Bang' theory, the early Universe was extremely hot and dense, resulting in ionisation of all atoms and frequent photon scattering. This intense scattering resulted in a blackbody spectrum. As the Universe expanded, its temperature decreased, eventually reaching a point where protons and electrons could combine to form neutral atoms. This 'recombination epoch' rendered the Universe transparent, as neutral atoms could no longer scatter photons via Thomson scattering. This allowed photons to travel freely through space. The photons emitted from this 'last scattering surface' are what we now observe as the CMB, providing invaluable insights into the early stages of the Universe.

The pattern of CMB anisotropies is influenced by all cosmological parameters, making it an extremely powerful tool for constraining these parameters. Currently, the most precise constraints on the majority of cosmological parameters are provided by the *Planck* CMB measurements (Planck Collaboration et al. 2020). The Atacama Cosmology Telescope (Fowler et al. 2007) and South Pole Telescope (Carlstrom et al. 2011) experiments continue these efforts, with a focus on achieving higher angular resolution in CMB measurements and increasing sensitivity in detecting CMB polarisations. Such advances will not only further refine the constraints on cosmological parameters, but also aid in studying distant galaxy clusters and their environments by leveraging the interactions between CMB photons and the hot ionised gas within these clusters, a phenomenon known as the Sunyaev-Zeldovich effect (Sunyaev & Zeldovich 1972).

Measuring large-scale matter distribution

Another major class of observational methods involves the direct measurement of matter distribution in the late-time Universe, using techniques such as galaxy clustering and weak gravitational lensing, with the latter being the primary focus of this thesis. As detailed in Sect. 1.1.2, the evolution of cosmic structure within an expanding Universe encodes valuable information about initial density fluctuations. According to the standard ΛCDM model, the current matter distribution is mainly determined by the initial perturbation field and the horizon scale at the time of matter-radiation equality ($\rho_\mathrm{m} = \rho_\mathrm{rad}$). This horizon scale, as shown by Bardeen et al. (1986) and Sugiyama (1995), is characterised by the parameter $\Omega_\mathrm{m}h$ and to a lesser extent, the baryonic mass fraction. Consequently, observations of the large-scale structure offer direct constraints on the cosmological parameters Ω_m and σ_8. A standard approach to simplify the comparison of cosmological results from different probes is the introduction of a structure growth parameter, $S_8 = \sigma_8\sqrt{\Omega_\mathrm{m}/0.3}$.

Galaxy clustering operates on the assumption that galaxy locations trace the cosmic matter distribution. However, galaxies, being the outcome of complex, non-linear processes, can deviate from this assumption, resulting in a statistical difference between

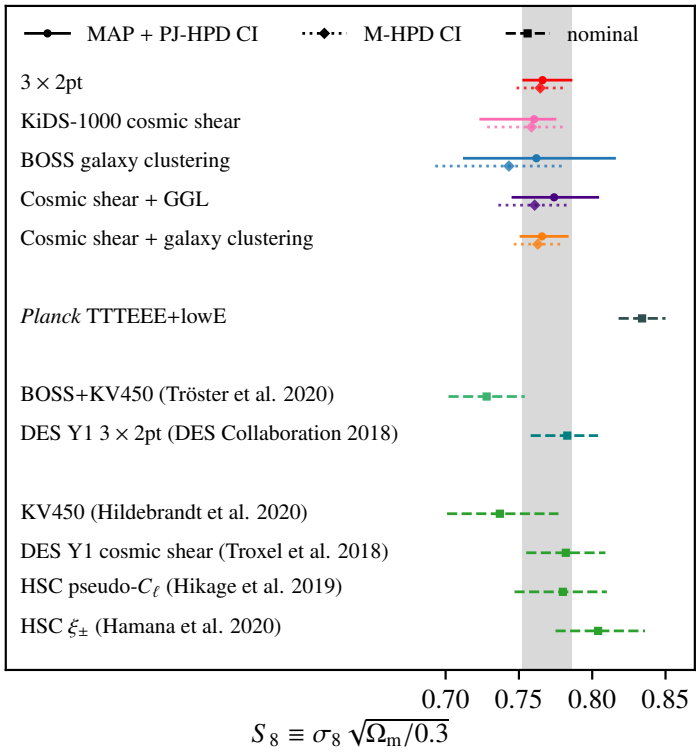

Figure 1.4: Constraints on S_8 from different cosmological probes. The label 3×2pt refers to results from a joint analysis combining three sets of two-point statistics: galaxy position-position correlations (galaxy clustering), galaxy shear-position correlations (galaxy-galaxy lensing), and galaxy shear-shear correlations (cosmic shear). Other labels represent the respective survey teams and the probes used. The error bars correspond to the 68% credible intervals. [Credit: Heymans et al. (2021)]

the galaxy and underlying matter distributions (see Desjacques et al. 2018 for a review). This galaxy bias complicates the interpretation of measured galaxy clustering signals. Furthermore, constructing a 3D galaxy distribution depends on galaxy redshift estimates, which encapsulate both the cosmological expansion and the line-of-sight peculiar velocity, thus introducing contamination from intricate dynamical effects.

On the other hand, the weak lensing effect, which directly correlates with the gravitational field of foreground matter, provides an unbiased trace of the large-scale structure. However, the minute distortions induced by the weak lensing effect are challenging to measure and susceptible to various systematic errors associated with the measurement process. Despite advances in high-quality imaging surveys and innovative shear measurement algorithms, careful control of systematic effects remains essential as survey statistical power increases.

A promising approach entails a joint analysis of these two methods, as they

probe the same underlying large-scale structure of the Universe. This method can break the degeneracy between cosmological parameters found in individual observables, significantly increasing the precision of cosmological parameter constraints. Furthermore, because different observables are affected by different systematic effects, a joint analysis permits some degree of self-calibration, leading to more accurate cosmological results.

Figure 1.4 presents the S_8 constraints from a recent joint analysis by the Kilo-Degree Survey (KiDS, Heymans et al. 2021), comparing results with the *Planck* CMB measurements and other contemporary galaxy surveys such as the Baryon Oscillation Spectroscopic Survey (BOSS, Alam et al. 2015), the Dark Energy Survey (DES, Dark Energy Survey Collaboration et al. 2016), and the Hyper Suprime-Cam (HSC) survey (Aihara et al. 2018). It shows that the precision of the current joint analysis of galaxy clustering and weak lensing is already comparable to the *Planck* CMB measurements. Interestingly, a mild 2σ to 3σ discrepancy exists between the S_8 values measured from large-scale matter distribution and those inferred from CMB measurements. Explaining this mild S_8 tension is more subtle than resolving the more significant H_0 tension. Given the high precision of modern surveys, any slight systematic residuals in any probe could induce such level of difference, underscoring the importance of identifying and controlling measurement and astrophysical systematic errors. More detailed discussions on this aspect will be presented in the following sections and throughout this thesis.

1.2 Weak gravitational lensing

The deflection of light by massive objects due to the curvature of space-time induced by their gravity is a common phenomenon in modern astronomical observations and is well-grounded in the theory of general relativity. Cosmological lensing effects can be broadly divided into two categories depending on the prominence of the lensing effect, as illustrated in Fig. 1.5: strong lensing, which produces substantial arcs or multiple images; and weak lensing, which causes gravitational distortions to the source galaxy images that are not visually detectable and thus require a statistical analysis.

This section provides a concise overview of gravitational lensing theory, emphasising weak lensing effects. We begin with a brief introduction to the fundamental concepts of lensing theory in Sect. 1.2.1, which is followed by discussions on the two primary weak lensing configurations: galaxy-galaxy lensing and cosmic shear in Sect. 1.2.2 and Sect. 1.2.3, respectively. The section concludes with a summary of major challenges related to the measurement and modelling of weak lensing for cosmological applications, presented in Sect. 1.2.4.

1.2.1 Fundamentals of gravitational lensing

For the study of gravitational lensing, we typically deal with lens objects that are much smaller than the Hubble length cH_0^{-1} and have peculiar velocities much smaller than the speed of light. In such cases, the lensing effects can be represented by introducing small

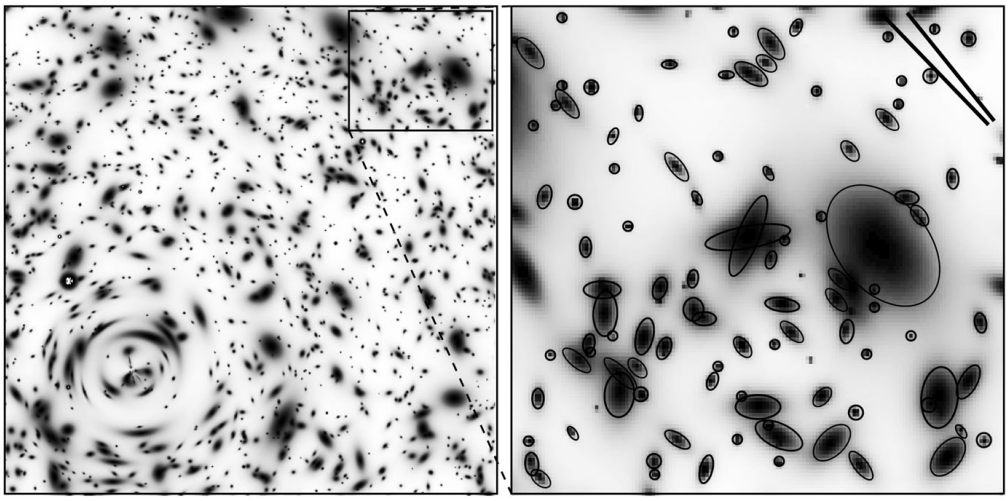

Figure 1.5: Simulated images illustrating the two lensing regimes. The bottom left part of the left panel plot concerns the strong lensing regime, where giant arcs and multiple images are produced. As the radial distance of source galaxies from the lens centre increases, the lensing effect decreases, and eventually lensing-induced distortions become smaller than the intrinsic ellipticity of the source galaxies. The right panel offers a zoomed-in view of this weak lensing regime, where the contours represent galaxy shapes as determined from their second moments. [Credit: Mellier (1999)]

perturbations to the FLRW metric (Eq. 1.1) as

$$\mathrm{d}s^2 = \left(1 + \frac{2\Phi}{c^2}\right) c^2 \mathrm{d}t^2 - \left(1 - \frac{2\Phi}{c^2}\right) a^2(t) \mathrm{d}l^2 \,, \tag{1.15}$$

where Φ is the Newtonian potential, which is assumed to be weak ($\Phi \ll c^2$) and satisfies the Poisson equation:

$$\nabla^2 \Phi = 4\pi G \rho \,. \tag{1.16}$$

Here, ρ represents the mass density of the lens objects, and the gradient operator, ∇, operates on the physical, or proper, coordinates.

This perturbed FLRW metric, in combination with the light geodesic equation ($\mathrm{d}s = 0$), forms the theoretical foundation for solving gravitational lensing problems that meet the weak gravitational field requirements. However, deriving a general analytical solution from these equations is challenging and often requires approximations. Therefore, a more practical approach would be to find specific solutions for individual problems based on their unique conditions. We shall adopt this approach, starting with a simple case known as the geometrically-thin lens to introduce some general concepts associated with gravitational lensing theory. More practical examples concerning the cosmological applications of weak gravitational lensing will be discussed in subsequent sections.

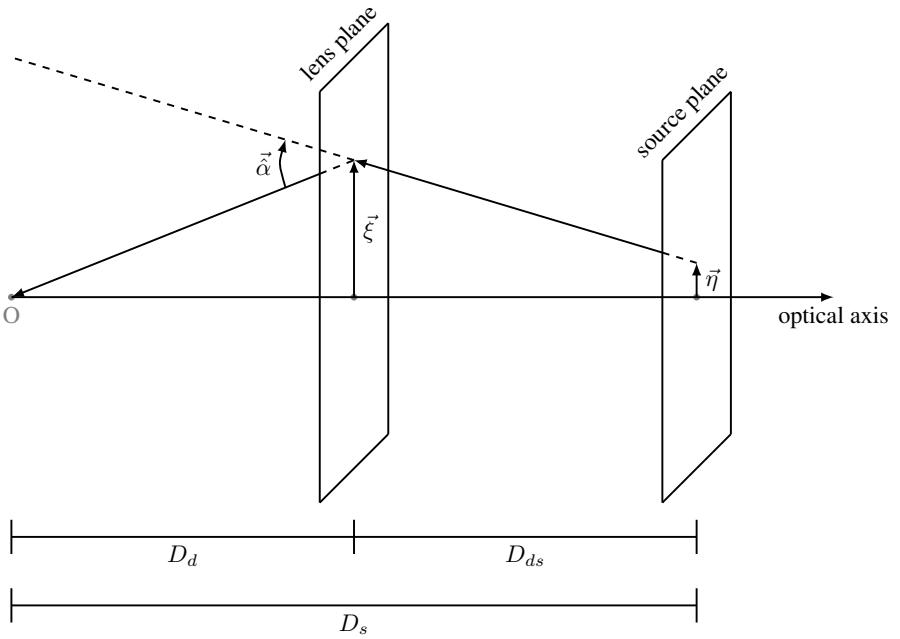

Figure 1.6: A sketch illustrating the geometry of a thin gravitational lensing system.

The deflection angle and lens equation

A geometrically-thin lens corresponds to a lens whose size is much smaller than the distances between the observer, lens, and source. This is a valid assumption for most astrophysical objects, such as galaxies and clusters of galaxies. Under such circumstances, we can use a geometric relationship to describe the effects of gravitational lensing on light propagation, as illustrated in Fig. 1.6. Here, the lens is located at an angular diameter distance $D_{\rm d}$, while the source object is located at an angular diameter distance $D_{\rm s}$. The angular diameter distance between the lens and source is denoted as $D_{\rm ds}$.

Locally, we can construct source and lens planes that are perpendicular to a straight line (referred to as the optical axis in the figure) extending from the observer to the lens, and onwards to infinity. These planes intersect the optical axis at the centres of the source and lens, respectively. Then, the smoothly curved trajectory of light, which is the reality under gravitational lensing, can be approximated by two straight lines that intersect at a point in the lens plane, forming a kink. The magnitude and direction of this kink are described by a deflection angle, $\vec{\hat{\alpha}}$, which quantifies the extent of light deflection due to the gravitational lensing effect.

The simplest lens model under this configuration involves a point mass M serving as the lens. As long as the impact parameter—defined as the distance between the kink and the lens—is much larger than the Schwarzschild radius of the lens ($|\vec{\xi}| \gg R_{\rm s} \equiv 2GMc^{-2}$), the weak gravitational field approximation remains valid, and the deflection angle can

be found as (see e.g. Narayan & Bartelmann 1997):

$$\vec{\hat{\alpha}}_{\rm p}(\vec{\xi}) = \frac{4GM}{c^2\vec{\xi}}. \tag{1.17}$$

An extended lens object can be treated as a collection of point masses. The resulting deflection angle is then calculated as the vector sum of the deflections due to each individual point mass. Each point mass element has a mass $\mathrm{d}M = \rho \mathrm{d}V$, where ρ is the mass density of the lens and $\mathrm{d}V$ is a volume element. If we denote the three-dimensional location of the kink as (ξ_1, ξ_2, r_3) and that of the mass element $\mathrm{d}M$ as (ξ_1', ξ_2', r_3'), then the impact vector is $\vec{\xi} - \vec{\xi}'$, independent of r_3 due to our assumption of a geometrically-thin lens. The total deflection angle can be obtained as (see e.g. Bartelmann & Schneider 2001):

$$\vec{\hat{\alpha}}(\vec{\xi}) = \frac{4G}{c^2}\int \mathrm{d}^2\xi'\ \Sigma(\vec{\xi}')\ \frac{\vec{\xi} - \vec{\xi}'}{|\vec{\xi} - \vec{\xi}'|^2}\ ; \qquad \Sigma(\vec{\xi}) \equiv \int \mathrm{d}r_3\ \rho(\xi_1, \xi_2, r_3)\ , \tag{1.18}$$

where $\Sigma(\vec{\xi})$ is the surface mass density.

Once we determined the deflection angle, mapping from the source plane to the lens plane becomes straightforward using the geometric relations illustrated in Fig. 1.6. Let $\vec{\eta}$ represent the two-dimensional source position in the source plane, we have

$$\vec{\eta} = \frac{D_{\rm s}}{D_{\rm d}}\vec{\xi} - D_{\rm ds}\vec{\hat{\alpha}}(\vec{\xi})\ . \tag{1.19}$$

We can simplify this equation by introducing angular coordinates, defined as $\vec{\beta} \equiv \vec{\eta}/D_{\rm s}$ and $\vec{\theta} \equiv \vec{\xi}/D_{\rm d}$, and a scaled deflection angle

$$\vec{\alpha}(\vec{\theta}) \equiv \frac{D_{\rm ds}}{D_{\rm s}}\vec{\hat{\alpha}}(D_{\rm d}\vec{\theta}) = \frac{1}{\pi}\int \mathrm{d}^2\theta'\ \kappa(\vec{\theta}')\frac{\vec{\theta} - \vec{\theta}'}{|\vec{\theta} - \vec{\theta}'|^2}\ , \tag{1.20}$$

where $\kappa(\vec{\theta}) \equiv \Sigma(D_{\rm d}\vec{\theta})/\Sigma_{\rm cr}$ is the dimensionless surface mass density, often called convergence, with $\Sigma_{\rm cr} \equiv c^2 D_{\rm s}/(4\pi G D_{\rm d} D_{\rm ds})$ representing the critical surface mass density. With these definitions, Eq. (1.19) simplifies to

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta})\ . \tag{1.21}$$

This relation, known as the lens equation, provides the mapping from a sky position in the source plane to a position in the lens plane. In physical terms, this means that a source situated at a position $\vec{\beta}$ in the sky will appear at position $\vec{\theta}$ due to the gravitational lensing effect. Notably, this mapping from the source plane to the lens plane may not be one-to-one. Indeed, Eq. (1.21) can yield more than one solution for certain $\vec{\beta}$. In physical terms, these multiple solutions correspond to multiple images of the same source object being detected by the observer, a phenomenon commonly referred to as

strong lensing.

The Fermat potential and time delay

Equation (1.20) implies the existence of a deflection potential, defined as

$$\psi(\vec{\theta}) \equiv \frac{1}{\pi} \int \mathrm{d}^2\theta' \, \kappa(\vec{\theta}') \ln|\vec{\theta} - \vec{\theta}'| \,, \tag{1.22}$$

which satisfies $\alpha = \nabla_\theta \psi$. Here, ∇_θ represents the two-dimensional gradient operator in respect to θ, acting on the lens plane. This deflection potential can be interpreted as a two-dimensional analogue of the Newtonian potential, given that it also satisfies the two-dimensional Poisson equation:

$$\nabla^2_\theta \psi = 2\kappa = \frac{2\Sigma}{\Sigma_{\mathrm{cr}}} \,. \tag{1.23}$$

Building on this, we can define a Fermat potential

$$\tau(\vec{\theta}; \vec{\beta}) \equiv \frac{1}{2}(\vec{\theta} - \vec{\beta})^2 - \psi(\vec{\theta}) \,, \tag{1.24}$$

and the condition that its gradient $\nabla_\theta \tau$ equals to zero is equivalent to the lens equation given in Eq. (1.21).

As demonstrated by Schneider (1985), the Fermat potential $\tau(\vec{\theta}; \vec{\beta})$, subject to an affine transformation, represents the light travel time from the source to the observer, accounting for the lensing effect. Thus, the lens equation can be interpreted as another manifestation of Fermat's principle, which states that light follows a path where the travel time is stationary, i.e., a local minimum, maximum or saddle point. Given the generality of the Fermat potential and its connection to the light travel time, it proves valuable in understanding the properties of lens systems, including the study of multiple images and time delays in strong lensing (e.g., Blandford & Narayan 1986).

The Jacobian matrix and image distortion

Until now, we have focused on the mapping of a single source position. In reality, source objects like galaxies have an extended shape, and light emitted from different parts of these sources will be deflected differently. This results in distortions in the observed images. In general, quantifying this image distortion involves solving the lens equation for each point within an extended source. However, when a source is relatively small and the lens properties do not change significantly within its angular scale, the image distortion can be linearly represented by the Jacobian matrix:

$$\mathcal{A}(\vec{\theta}) = \frac{\partial \beta}{\partial \theta} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \,, \tag{1.25}$$

where

$$\gamma_1 = \frac{1}{2}\left(\frac{\partial^2\psi}{\partial\theta_1^2} - \frac{\partial^2\psi}{\partial\theta_2^2}\right); \qquad \gamma_2 = \frac{\partial^2\psi}{\partial\theta_1\partial\theta_2}\,, \tag{1.26}$$

represent the two components of the shear parameter $\gamma \equiv \gamma_1 + \mathrm{i}\gamma_2$. The convergence κ, introduced earlier, is connected to the deflection potential through Eq. (1.23).

According to Liouville's theorem, surface brightness is conserved before and after gravitational lensing, assuming no photon emission or absorption occurs during the process. As a result, the Jacobian matrix contains all information about changes in the local shape and flux of the source pre- and post-lensing distortion. To better grasp how the parameters in the Jacobian matrix relate to the distortions of the source image, we can re-formulate the Jacobian matrix as

$$\mathcal{A}(\vec{\theta}) = (1-\kappa)\begin{pmatrix} 1-g_1 & -g_2 \\ -g_2 & 1+g_1 \end{pmatrix}\,, \tag{1.27}$$

by introducing the reduced shear parameter: $g = \gamma/(1-\kappa)$.

With Eq. (1.27), it is evident that the factor $(1-\kappa)$ influences the image isotropically, affecting only the size and apparent flux of the observed image without altering its shape. In the context of weak lensing, since the intrinsic properties of source objects are unknown, this isotropic magnification effect cannot be directly measured for individual lens systems[1]. On the other hand, the anisotropic distortions are entirely determined by the reduced shear g, making it the central quantity in the study of weak gravitational lensing.

1.2.2 Galaxy-galaxy lensing

After establishing the general principles of gravitational lensing, we can now explore some practical examples of weak lensing that are vital in observational cosmology. Our first example involves lens systems where individual massive objects, such as galaxies, or groups/clusters of galaxies, serve as the lens. Such lens systems are commonly referred to as 'galaxy-galaxy lensing' in contemporary weak lensing studies. Considering that a typical galaxy cluster is only a few Mpc in size, while the cosmological distances involved are usually several hundred to a few thousand Mpc, the geometrically-thin lens condition is well satisfied. Thus, the results derived in the previous section can be directly implemented in galaxy-galaxy lensing studies.

The key element in galaxy-galaxy lensing studies involves the correlation between the surface mass density and the lensing effect, as illustrated in Eq. (1.18). This correlation enables a range of applications such as estimating the total lens mass, mapping the matter distribution, and more. Although the convergence κ is directly linked to the surface mass density by its definition, it is challenging to measure in practice. Conversely, the

[1]However, we can statistically measure the magnification effect, for example, by evaluating changes in the local source number density (e.g. Broadhurst et al. 1995; Schneider et al. 2000; Hildebrandt et al. 2009) or in the size-magnitude-redshift relation of galaxies (e.g. Alsing et al. 2015).

shear parameter γ is relatively easier to measure from observed source images. Hence, establishing a connection between convergence and shear is crucial.

Assuming the surface mass distribution of the lens is axisymmetric, this connection can be readily derived from their relations to the deflection potentials (Eqs. 1.23 and 1.26):

$$\gamma_{\rm t}(\theta) = \overline{\kappa(<\theta)} - \kappa(\theta) \ , \tag{1.28}$$

where $\overline{\kappa(<\theta)}$ represents the average value of κ within a radius of θ. We also introduced the tangential shear: $\gamma_{\rm t} \equiv -\gamma_1 \cos(2\varphi) - \gamma_2 \sin(2\varphi)$, where φ represents the relative position angle of the source with respect to the lens centre. Remarkably, this relationship holds for general matter distributions as long as we replace local $\gamma_{\rm t}(\theta)$ and $\kappa(\theta)$ estimations with their azimuthally-averaged counterparts $\langle\gamma_{\rm t}(\theta)\rangle$ and $\langle\kappa(\theta)\rangle$ (Bartelmann 1995).

In practical terms, this implies that we can determine the azimuthally-averaged mass profile of a lens object by measuring the azimuthally-averaged tangential shear, irrespective of whether the density distribution is axisymmetric. This method can be used to map the projected matter distribution of lens objects or to estimate the total mass of the lens, provided the density profile of the lens is available. This application of galaxy-galaxy lensing with real-world data will be the topic of Chapter 5.

1.2.3 Cosmic shear

Another primary application of the weak lensing effect is its use in directly probing the large-scale structure of the late-time Universe. This method, known as 'cosmic shear', considers all matter distributions between the source and the observer as the lens. As mentioned in Sect. 1.1.3, the study of large-scale structures provides powerful constraints on cosmological models, and cosmic shear is one of the few methods capable of such examination. Therefore, cosmic shear plays a central role in modern cosmological surveys.

On the theoretical side, 'the lens' of cosmic shear, spanning the entire space between the source and the observer, breaks the geometrically-thin lens condition and requires an extension of the classical lens theory. Furthermore, given the lack of a concrete lens in this context, cosmic shear measurements depend on statistical analyses to establish a connection with the statistical properties of the cosmic density field. Covering all these aspects with careful derivations would exceed the scope of this introductory section. Therefore, we will only outline the key elements of cosmic shear without detailed derivation, focusing on the second-order statistical measures, namely, the two-point correlation function and power spectrum as defined in Sect. 1.1.2. For a more comprehensive discussion, we refer to Bartelmann & Schneider (2001) and Kilbinger (2015).

Although the geometrically-thin lens condition no longer holds, the assumption of weak gravitational perturbations remains applicable. Thus, the perturbed FLRW metric and the Poisson equation, as defined in Eqs. (1.15) and (1.16), are still valid. Considering $2\Phi/c^2$ as a small parameter, we can identify an effective deflection potential expressed

as

$$\psi_{\rm eff}(\vec{\theta}, w) = \frac{2}{c^2}\int_0^w \mathrm{d}w' \frac{f_K(w-w')}{f_K(w)f_K(w')}\Phi[f_K(w')\vec{\theta},\ w']. \tag{1.29}$$

Here, w represents the comoving distance, $f_K(w)$ is the comoving angular diameter distance, which equals w in a spatially flat Universe, and the integral is performed from the observer to a target source.

Using this definition of $\psi_{\rm eff}$, we can apply Eq. (1.23) to introduce an effective convergence parameter, formulated as

$$\kappa_{\rm eff}(\vec{\theta}, w) = \frac{3H_0^2\Omega_{\rm m,0}}{2c^2}\int_0^w \mathrm{d}w'\ \frac{f_K(w')f_K(w-w')}{f_K(w)}\frac{\delta[f_K(w')\vec{\theta},\ w']}{a(w')}\ , \tag{1.30}$$

where we used the Poisson equation (1.16) and replaced ρ with the overdensity δ, as defined in Eq. (1.7). This equation reveals the correlation between the cosmic shear effect and cosmological parameters. The effective convergence parameter is essentially a line-of-sight integral of the cosmic perturbation field, with the scale factor $a(w')$ and a combination of comoving angular diameter distance factors acting as a weighting function. The combination of the Hubble constant and matter density parameter only influences the amplitude of $\kappa_{\rm eff}(\vec{\theta}, w)$.

In practice, cosmic shear measurements from a collection of source galaxies are averaged to enhance the signal-to-noise ratio. If these sources have a redshift distribution such that $p_z(z)dz = G(w)dw$, we can derive an averaged effective convergence, expressed as

$$\overline{\kappa}_{\rm eff} = \frac{3H_0^2\Omega_{\rm m,0}}{2c^2}\int_0^{w_{\rm H}} \mathrm{d}w'\ \overline{W}(w')f_K(w')\frac{\delta[f_K(w')\vec{\theta},\ w']}{a(w')}\ . \tag{1.31}$$

Here, the weighting function, also known as the lensing efficiency factor, is defined as

$$\overline{W}(w') \equiv \int_{w'}^{w_{\rm H}} \mathrm{d}w\ G(w)\frac{f_K(w-w')}{f_K(w)}\ . \tag{1.32}$$

In these equations, the upper limit of integration corresponds to the horizon distance, equivalent to infinite redshift.

To connect the cosmic shear measurements to the statistical properties of the cosmic overdensity field, we use the power spectrum and apply the Limber's equation, which links the two-point statistical measures of a projected field to its original three-dimensional field (Limber 1953). Following this approach, we can derive

$$P_\kappa(k) = \frac{9H_0^4\Omega_{\rm m,0}^2}{4c^4}\int_0^{w_{\rm H}} \mathrm{d}w\ \frac{\overline{W}^2(w)}{a^2(w)}P_\delta\left(\frac{k}{f_K(w)}, w\right)\ , \tag{1.33}$$

which links the power spectrum of convergence field to the power spectrum of the cosmic matter distribution.

However, as mentioned earlier, convergence is challenging to measure in weak

lensing studies, necessitating a connection to the shear parameter. Fortunately, in the weak lensing limit, the two-point statistical properties of convergence and shear are identical: $P_\gamma = P_\kappa$ (Blandford et al. 1991). This relationship can be straightforwardly demonstrated in Fourier space by considering the relations of convergence and shear to the deflection potential, as defined in Eqs. (1.23) and (1.26). Hence, we can use the observed shear statistics to directly study the cosmic matter distribution, adhering to the same relationship as in Eq. (1.33).

The line-of-sight integral in Eq. (1.31) implies that cosmic shear primarily measures two-dimensional projections of the three-dimensional cosmic matter distributions. To mitigate the signal loss from this integration, contemporary cosmic shear analyses often include an additional layer of correlations along the line of sight by binning galaxies based on their redshifts, a method known as tomographic analysis. This technique has proven highly effective in enhancing the statistical power of cosmic shear for constraining cosmological parameters (e.g. Hu 1999). Furthermore, it enables cosmic shear to constrain properties of dark energy by facilitating measurements of the evolution of cosmic structure (e.g. Huterer 2002). Chapter 4 presents an application of this cosmic shear tomography using real-world data.

1.2.4 Challenges in weak lensing analysis

By now, it should be evident that the theory of weak lensing is well-established and has shown itself to be extremely powerful in studying the matter distribution of both individual objects and large-scale cosmic structures. However, applying it to real-world data presents substantial challenges that warrant further investigation. This is particularly true given the advancements of current and upcoming cosmological surveys, and the intriguing disagreements found among different cosmological probes, as discussed in Sect. 1.1.3. This section provides a brief overview of some key systematic issues encountered during the measurement and modelling of weak lensing. Chapters 2 and 3 will delve deeper into these systematics, exploring their impact and discussing strategies for their control in real-world applications.

Shear measurement and selection biases

The initial challenge in weak lensing studies arises when attempting to measure subtle weak lensing distortions from noisy, faint galaxy images. This complexity is amplified by distortions caused by the point spread function (PSF), resulting from instrumental and observational conditions (e.g. Paulin-Henriksson et al. 2008; Massey et al. 2013), the detection or selection bias during the identification of faint objects (e.g. Hartlap et al. 2011; Chang et al. 2013; Hoekstra et al. 2021), and the blending effects that occur when two or more objects are in close proximity on the sky (e.g. Hoekstra et al. 2015; Dawson et al. 2016; Hoekstra et al. 2017; Mandelbaum et al. 2018; Samuroff et al. 2018; Euclid Collaboration et al. 2019). These factors not only introduce shear measurement bias but also modify the selection function of the source sample, complicating the interpretation of the measured shear signals. Therefore, managing shear measurement and selection

biases is a crucial first step in any weak lensing studies.

Following two decades of development, the state-of-the-art shear measurement methods are capable of controlling measurement biases at a sub-percent level (e.g. Heymans et al. 2006; Massey et al. 2007; Bridle et al. 2010; Kitching et al. 2012; Mandelbaum et al. 2015). In addition, the use of realistic pixel-level image simulations helps address residual biases and account for selection biases (see Chapter 3). This progress paints a promising picture for controlling shear biases in future high-requirement weak lensing surveys. Further investigation into blending effects and the interplay between shear and redshift measurements could still provide benefits (e.g. MacCrann et al. 2022; Li et al. 2023b). However, the shear measurement alone does not appear to pose the major challenge for future surveys. More demanding are tasks such as careful modelling of the PSF and improving the realism of image simulations (e.g. Li et al. 2023a,b). These improvements, related to the shear measurement and calibration, seem to be the more challenging aspects and warrant further study.

Photometric redshift uncertainties

Determining the distances of source galaxies presents a significant challenge for current and future weak lensing surveys. This distance estimation relies on observing the electromagnetic spectral energy distribution (SED) of a galaxy. The cosmological expansion stretches the observed SED towards longer wavelengths by a factor $(1 + z)$, where z is the redshift. Therefore, the redshift can be related to a proper distance, assuming a cosmological model. To measure the stretch of the SED, we need to identify some characteristic features in it, such as emission and absorption lines. These features are easy to identify if the observed wavelength resolution is high, i.e., if we obtain the spectra of galaxies.

However, for deep imaging surveys, the fraction of galaxies with observed spectra is very low. Fortunately, weak lensing analyses do not require precise redshift estimates for each source galaxy. The source redshifts enter the calculation through an integral, as shown in Eq. (1.32). Therefore, as long as the average redshift of the ensemble of source galaxies is accurately estimated, the analysis yields unbiased results. This allows for the use of so-called photometric redshifts, which are less precise.

This method measures the flux of distant galaxies in a few broader wavelength filters, thereby obtaining a sparse sampling of the SED. It estimates the redshift based on broad features such as the Lyman and Balmer breaks. Nevertheless, the raw estimates from this process have large uncertainties that exceed the requirements of current and future weak lensing analyses (e.g. Hildebrandt et al. 2010; Salvato et al. 2019). Therefore, an additional step of redshift calibration, using more precise spectroscopic reference samples, is necessary. However, this calibration process introduces its own biases, for instance caused by incompleteness or non-representative sampling of the samples with spectroscopic redshifts. This results in another layer of residual bias correction that also relies on realistic simulations.

The first-step calibration is currently undergoing rapid development with the rise

of machine learning-based techniques, such as self-organising maps (e.g. Kohonen 1982; Masters et al. 2015). The focus now is on enhancing the spectroscopic reference samples and gaining a better understanding of the transfer function in both samples, which describes how intrinsic photometry maps to measured ones and its correlation with redshifts. This progress is crucial as colour-based redshift calibration methods rely on the assumption that galaxy colour and redshift are closely correlated, and that the spectroscopic calibration sample and the faint photometric target sample share similar underlying properties (e.g. Hildebrandt et al. 2021; Myles et al. 2021; Rau et al. 2022). Simulations can aid this understanding, but the current absence of simulations for the spectroscopic reference sample, and the complexity of the transfer function in spectroscopic measurements, make this a challenging endeavour. Therefore, further explorations in this direction are critical.

For the second step, residual bias correction, realism in simulations is pivotal, much like in shear calibration (e.g. Hoyle et al. 2018; van den Busch et al. 2020; DeRose et al. 2022). Recent advancements in multi-band image simulations appear promising (e.g. Li et al. 2023b). These simulations naturally account for the transfer function in the photometric sample and the blending effect, an aspect previously overlooked in catalogue-level simulations. Furthermore, multi-band image simulations enable a joint calibration of both shear and redshift estimates, leading to a full end-to-end calibration, that is beneficial for both processes. The development of such multi-band image simulations will be the focus of Chapter 3.

Baryonic effects and Intrinsic alignment

Even after achieving robust shear and redshift measurements, interpreting the weak lensing signal, particularly the cosmic shear signal, remains a complex task. This complexity arises from the fact that cosmic shear probes the matter distribution in the late-time Universe, which undergoes non-linear evolutions, as discussed in Sect. 1.1.2. Additionally, measured statistical signals include contamination from the intrinsic alignment (IA) of galaxies, a non-lensing-induced coherent alignment of galaxies, originating from tidal effects between neighbouring galaxies. Therefore, a crucial aspect of weak lensing studies involves enhancing modelling accuracy to account for these non-linear and astrophysical effects.

Regarding non-linear evolution, cosmological numerical simulations offer a promising solution. Advances in simulation algorithms and computational power now make it possible to create gravitational-only N-body simulations that meet the requirements for interpreting future weak lensing measurements (e.g. Heitmann et al. 2009; Nishimichi et al. 2019; Euclid Collaboration et al. 2021). However, baryonic processes, such as feedback from star formation and active galactic nuclei, also alter matter distributions on scales relevant to cosmic shear studies (e.g. van Daalen et al. 2011; Semboloni et al. 2011). These non-gravitational processes cannot be captured by N-body simulations, necessitating hydrodynamic simulations that include baryonic processes. However, constructing hydrodynamic simulations that capture all astrophysical processes is compu-

tationally demanding. More cost-effective alternatives, such as simplified hydrodynamic simulations with fixed subgrid physics or semi-analytical models, require a sufficient understanding of the astrophysical processes that are not included in the original simulations, and this remains an active research area (e.g. Schaye et al. 2023; Kugel et al. 2023).

The issue of IA contamination is notably challenging to resolve due to our currently limited understanding of its properties (see e.g. Joachimi et al. 2015; Troxel & Ishak 2015 for reviews). The prevailing approach introduces empirical models with nuisance parameters to mitigate the IA impact (e.g. Hirata & Seljak 2004; Bridle & King 2007; Blazek et al. 2019). However, these IA parameters were found to correlate with other nuisance parameters, such as those related to redshift calibration (see Chapter 2). This issue is further complicated by the fact that different types of galaxies exhibit varying IA behaviours, which complicates the interpretation of a mixed-colour weak lensing sample (e.g. Johnston et al. 2019; Fortuna et al. 2021b; Samuroff et al. 2022). In this regard, a strategy of splitting the source samples based on their colours would be beneficial, a method we employ in our Chapter 2 study. Moreover, advancements in cosmological simulations allow for the study of IA within these simulations, providing valuable insights into the underlying mechanisms of IA signals and aiding in refining our modelling approach (e.g. Tenneti et al. 2014; Chisari et al. 2015; Hilbert et al. 2017; Kurita et al. 2021).

1.3 This thesis

The rest of this thesis delves into the real-world applications of weak lensing as a cosmological probe. We focus on two main aspects of weak lensing applications: cosmic shear and galaxy-galaxy lensing, as previously introduced. Our goal is to enhance constraints on cosmological parameters and deepen our understanding of the connection between dark matter halos and galaxies. Throughout this work, we devote considerable attention to identifying, testing, and controlling systematic effects that arise during weak lensing measurements and modelling, recognising this aspect as critical for current and future weak lensing studies.

In **Chapter 2**, we conduct an internal consistency test for cosmic shear analysis using data from the Kilo-Degree Survey (KiDS). We split source galaxies based on their colours and assess the robustness of the current KiDS pipeline when handling samples with varied galaxy properties. Our findings confirm that the existing KiDS pipeline effectively accounts for various systematic residuals. However, we identify a correlation between IA parameters and nuisance parameters associated with redshift calibrations, emphasising the need for improved IA models. Notably, the consistency test method we use is insensitive to assumed cosmological models, making it an effective tool for verifying the robustness of cosmic shear analysis before drawing any cosmological conclusions.

Chapter 3 introduces a suite of multi-band image simulations for shear and redshift calibration of KiDS weak lensing analysis. These newly developed simulations enable a

joint calibration of shear and redshift, enhancing the realism and accuracy over previous efforts. By integrating cosmological simulations with high-quality imaging data, we generate a large volume of simulated galaxies, accurately mirroring realistic galaxy properties to an adequate depth. We improve the realism of image simulations by carefully addressing various observational and instrumental effects, including noise background, point spread function, and stellar density. Furthermore, we study blended systems at different redshifts by introducing variable shear fields into the simulations. Our study reveals a correlation between shear and redshift biases due to blending effects, underscoring the importance of joint shear and redshift calibration. We also identify a minor but noteworthy impact of PSF modelling errors on shear bias. After conducting sensitivity tests, we confirm the robustness of the current KiDS fiducial shape measurement algorithm, within the requirements of the KiDS weak lensing analysis. For future weak lensing surveys with stricter requirements, we recommend further studies on blending effects, refinement of PSF modelling, and improvements to shape measurement techniques to reduce their sensitivity to variations in galaxy properties.

In **Chapter 4**, we conduct a cosmic shear analysis, incorporating improvements in cosmic shear measurements and calibrations facilitated by the work conducted in Chapter 3. Moreover, we take into account recent advancements in cosmological inference and investigate the influence of prior information on IA modelling on the final cosmological results. We propose treating the statistical and systematic uncertainties from the shear calibration separately, given their distinct origins. Our final cosmological results align with previous KiDS studies and other weak lensing surveys, showing a tension of approximately 2.3σ with the *Planck* cosmic microwave background constraints on S_8.

Finally, in **Chapter 5**, we investigate the connection between dark matter haloes and galaxies using galaxy-galaxy lensing. We measure galaxy-galaxy lensing signals around galaxy groups identified by the Galaxy And Mass Assembly (GAMA) project using KiDS data. We interpret these statistical lensing signals through the halo model formalism to infer the total halo mass associated with these galaxy groups. By adopting optical observables provided by GAMA, we constrain the scaling relation between the halo mass and the total luminosity of the groups. We assess the robustness of our halo mass estimates by varying the treatment of halo model parameters not well-constrained by the current data. These sensitivity tests help us identify critical model components and guide future improvements. Given the increased statistical power of the measured signals, this exploration and future enhancement of our current model are considered necessary.

CHAPTER 2

KiDS+VIKING-450: An internal-consistency test for cosmic shear tomography with a colour-based split of source galaxies

Abstract

We performed an internal consistency test for the KiDS+VIKING-450 (KV450) cosmic shear analysis, focusing on the colour-based split of source galaxies. Using the same measurements and calibrations across both sub-samples, we evaluated the properties of the shear measurements and the efficiency of the calibration pipelines. On the modelling side, we explored the observational nuisance parameters, particularly those associated with redshift calibration and intrinsic alignments, using a Bayesian analysis equipped with specific test parameters. Our study confirmed that the current nuisance parameters adequately capture residual systematic deviations in the KV450 data, albeit with minor discrepancies observed in the second and third redshift bins. Our results additionally revealed the degeneracy between the apparent amplitude of intrinsic alignments and redshift uncertainties in low redshift bins. Given the relative insensitivity of our test to the assumed cosmological model, it can be implemented in the cosmic shear analysis prior to drawing any cosmological conclusions.

S.-S. Li, K. Kuijken, H. Hoekstra,
H. Hildebrandt, B. Joachimi, and A. Kannawadi
Astronomy & Astrophysics, 646, A175

2.1 Introduction

Cosmic shear, a coherent distortion of distant galaxy shapes due to weak gravitational lensing by large-scale structures, is sensitive to the amplitude of matter density fluctuations (quantified by σ_8, which is the standard deviation of linear-theory density fluctuations in a sphere of radius $8h^{-1}$ Mpc, where $H_0 = 100h$ km s^{-1} Mpc^{-1}) and the mean matter density (Ω_m). As a result, cosmic shear surveys commonly report a derived parameter $S_8 \equiv \sigma_8(\Omega_\mathrm{m}/0.3)^{0.5}$. Alternatively, cosmic microwave background (CMB) measurements can infer local density fluctuations by extrapolating the measured amplitude of temperature fluctuations at recombination, based on a cosmological model. By comparing results from these two distinct probes, we can test cosmological models.

The latest *Planck* Legacy analysis of CMB measurements predicts an $S_8 = 0.832 \pm 0.013$ (68% credible region), assuming the standard Λ cold dark matter (ΛCDM) model (Planck Collaboration et al. 2020). This results is slightly higher than the results from the recent cosmic shear surveys, such as the Dark Energy Survey (DES; Troxel et al. 2018, $S_8 = 0.782^{+0.027}_{-0.027}$), the Hyper Suprime-Cam Subaru Strategic Program (HSC; Hikage et al. 2019, $S_8 = 0.780^{+0.030}_{-0.033}$), and especially the Kilo-Degree Survey (KiDS; Hildebrandt et al. 2020, hereafter H20, $S_8 = 0.737^{+0.040}_{-0.036}$).

It is crucial to consider potential systematic effects associated with observations when interpreting results from different surveys. Internal consistency tests are thus a standard part of any cosmological probe. Cosmic shear studies often base these checks on a split of the estimated two-point shear correlations (Köhlinger et al. 2019; or Sect. 7.4 of H20). By assigning duplicate model parameters to each subset, one can perform theoretical modelling of the reconstructed data vector and assess data consistency by comparing these duplicate parameters. This approach aids in verifying the consistency of a specific sample of source galaxies, but it only tests at a late stage in the analysis. Moreover, doubling cosmological parameters incurs a significant computational cost, which hinders further splitting of the source sample in practice. However, additional splits could be particularly interesting, as systematics may vary among them.

Source galaxy properties present two main challenges to calibration pipelines: shape measurements and redshift estimates. First, different galaxy samples typically exhibit varying ellipticity distributions, with red, early-type galaxies tending to have rounder shapes than their blue, late-type counterparts (Hill et al. 2019; Kannawadi et al. 2019, hereafter K19). This leads to a correlation between shear bias and the underlying galaxy sample, primarily because shape measurements are sensitive to ellipticity distributions. For instance, the *lens*fit algorithm used in the KiDS survey assigns weights to measured ellipticities, resulting in a bias towards intermediate ellipticity values (Fenech Conti et al. 2017). Second, both the accuracy and precision of photometric redshift estimates depend on a galaxy's broad spectral features, such as the Balmer break below 4000 (Salvato et al. 2019). The prominence of these broad spectral features varies by galaxy spectral type. Generally, galaxies with an older stellar population appear red at rest-frame optical wavelengths and exhibit a pronounced 4000 break. The more young stars a galaxy contains, the bluer it appears, causing the Balmer break and other broad spectral features

to be washed out. Consequently, errors in photometric redshifts correlate with galaxy spectral type (Mo et al. 2010).

In this chapter, we consider these sample-related systematic effects, specifically focusing on photometric redshift uncertainty. We divided the KiDS source galaxies into two mutually exclusive sub-samples based on their spectral types and applied the same measurement and calibration pipelines to each sub-sample. By doing this, we explored how sample-related systematics can affect measurements and assessed the effectiveness of calibration pipelines in mitigating these effects. This split also has implications for modelling intrinsic alignments, which must be explicitly accounted for. To quantify consistency, we performed a Bayesian analysis using dedicated test parameters that describe the relative deviations of nuisance parameters between the two sub-samples. By examining their posterior distributions, we can determine if the original settings are sufficient to capture residual biases. The analysis code is publicly available[1].

Our approach complements other studies that assess the consistency of inferred cosmological parameters by removing tomographic bins or by splitting the sample by galaxy type, while marginalising over the corresponding nuisance parameters (Köhlinger et al. 2019; Samuroff et al. 2019). Instead, we focused on a different aspect: keeping cosmological parameters fixed while examining changes in the nuisance parameters. We found that our method can effectively test for inconsistencies in redshift distributions and highlight the degeneracy between redshift uncertainties and apparent intrinsic alignment signals, all while remaining insensitive to cosmological assumptions.

The remainder of this chapter is organised as follows. In Sect. 2.2, we provide a brief overview of the cosmic shear catalogues under consideration. We present the redshift calibration in Sect. 2.3 and the shear bias calibration in Sect. 2.4. Next, we introduce the measurement and modelling of the shear signal in Sect. 2.5. We discuss the covariance matrix and consistency tests in Sect. 2.6. Our main results are presented in Sect. 2.7, and we conclude with a summary in Sect. 2.8.

2.2 Data

Our test is based on the first release of optical and infrared KiDS cosmic shear data, known as KiDS+VIKING-450 (KV450; Wright et al. 2019, hereafter W19). This data set includes four-band optical photometry ($ugri$) from the first three data releases of KiDS (de Jong et al. 2015, 2017) and five-band near-infrared photometry ($ZYJHK_s$) from the overlapping VISTA Kilo-Degree Infrared Galaxy Survey (VIKING, Edge et al. 2013).

Details on the derivation and verification of this cosmic shear catalogue can be found in the main KiDS cosmic shear papers (Hildebrandt et al. 2017; H20) and their companion papers (Fenech Conti et al. 2017; W19). The public catalogue provides all the necessary information for conducting a tomographic cosmic shear analysis. Among the most important columns are the photometric redshifts (photo-zs, or z_B as in the catalogues) and the galaxy shapes (described by two ellipticity components, ϵ_1 and ϵ_2).

[1] https://github.com/lshuns/CosmicShearRB

The z_B values are estimated using the Bayesian photometric redshift code (BPZ, Benítez 2000; Coe et al. 2006) with an improved redshift prior from Raichoor et al. (2014) and the nine-band photometry from W19. The galaxy shapes are measured from the r-band images (median seeing of 0.″7) using the *lens*fit algorithm (Miller et al. 2007; Kitching et al. 2008; Miller et al. 2013) with a self-calibration for noise bias (Fenech Conti et al. 2017).

Throughout this study, we only use sources with valid nine-band photometry (`GAAP_Flag_ugriZYJHKs==0`). This mask reduces the original area by approximately 5% and retains around 13 million objects, which is identical to the choice made by the main KV450 cosmic shear analysis. Following H20, we bin source galaxies into five tomographic bins defined as $0.1 < z_B \leq 0.3$, $0.3 < z_B \leq 0.5$, $0.5 < z_B \leq 0.7$, $0.7 < z_B \leq 0.9$, $0.9 < z_B \leq 1.2$. To check for systematic effects caused by galaxy properties, we further split the whole sample into two sub-samples based on the spectral types of source galaxies. This is achieved by using the T_B values reported by the BPZ code during the photo-z estimation procedure (see Benítez 2000, for a detailed discussion). Briefly, the T_B value is calculated within a Bayesian framework using six templates of galaxy spectra (Coleman et al. 1980; Kinney et al. 1996). We define our two sub-samples as $T_B \leq 3$ (a combination of E1, Sbc, Scd types, labelled as 'red') and $T_B > 3$ (a combination of Im and two starburst types, labelled as 'blue'). This cut is chosen to ensure similar statistical power in the two sub-samples (see Fig. 2.1). Source properties of these two sub-samples are summarised in Table 2.1.

2.3 Calibration of redshift distributions

One of the most challenging tasks in a tomographic cosmic shear study is estimating the source redshift distribution for each tomographic bin. Since these intrinsic redshift distributions vary with galaxy samples, we need to calibrate the photo-z estimates for the two sub-samples separately. We employed the technique from the fiducial KV450 cosmic shear analysis, known as DIR in H20, for this task. This method directly estimates the underlying redshift distributions of a photometric sample using deep spectroscopic redshift (spec-z) catalogues that overlap with the photometric survey. In this section, we briefly discuss our implementation of this method and refer interested readers to the original papers for more details (Lima et al. 2008; Hildebrandt et al. 2017, 2020).

The DIR method necessitates that the calibration sample (the spec-z sample) spans, at least sparsely, the full extent of the multi-band magnitude space covered by the target sample (the photo-z sample), and that the mapping from magnitude space to redshift space is unique. Consequently, the coverage of the spec-z sample is essential for this method's accuracy. We used the same set of spec-z catalogues as employed in the fiducial KV450 cosmic shear analysis. This includes the zCOSMOS survey (Lilly et al. 2009), the DEEP2 survey (Newman et al. 2013), the VIMOS VLT Deep survey (Le Fèvre et al. 2013), the GAMA-G15Deep survey (Kafle et al. 2018), and a combined catalogue provided by ESO in the *Chandra* Deep Field South area. These independent spec-z surveys, with different lines-of-sight and depths, minimise shot noise and sample

Table 2.1: Source information in the two sub-samples.

Sample	Bin	Photo-z range	Total *lens*fit weights	$n_{\rm eff}$ $\left[\mathrm{arcmin}^{-2}\right]$	$\sigma_{\epsilon,i}$	m-bias	Mean($z_{\rm DIR}$)	Median($z_{\rm DIR}$)
$T_{\rm B} \leq 3$	1	$0.1 < z_{\rm B} \leq 0.3$	7 031 963	0.38	0.279	-0.029 ± 0.010	0.351	0.282
(red)	2	$0.3 < z_{\rm B} \leq 0.5$	10 404 223	0.59	0.252	-0.009 ± 0.007	0.430	0.396
	3	$0.5 < z_{\rm B} \leq 0.7$	15 508 696	0.90	0.276	-0.010 ± 0.007	0.546	0.531
	4	$0.7 < z_{\rm B} \leq 0.9$	9 837 460	0.64	0.250	0.008 ± 0.006	0.744	0.732
	5	$0.9 < z_{\rm B} \leq 1.2$	8 466 542	0.59	0.275	0.006 ± 0.008	0.909	0.894
$T_{\rm B} > 3$	1	$0.1 < z_{\rm B} \leq 0.3$	7 269 125	0.42	0.270	-0.004 ± 0.008	0.437	0.244
(blue)	2	$0.3 < z_{\rm B} \leq 0.5$	12 200 673	0.75	0.277	-0.007 ± 0.006	0.573	0.431
	3	$0.5 < z_{\rm B} \leq 0.7$	21 116 034	1.46	0.292	-0.002 ± 0.006	0.791	0.644
	4	$0.7 < z_{\rm B} \leq 0.9$	12 134 896	0.92	0.286	0.026 ± 0.006	0.914	0.842
	5	$0.9 < z_{\rm B} \leq 1.2$	10 207 426	0.87	0.293	0.036 ± 0.009	1.081	1.022

The effective number density $n_{\rm eff}$ is calculated using Eq. (1) of Heymans et al. (2012). The reported ellipticity dispersion is defined as $\sigma_{\epsilon,i} = (\sigma_{\epsilon 1} + \sigma_{\epsilon 2})/2$. The m-bias is defined in Eq. (2.1) and detailed in Sect. 2.4. Uncertainties reported were computed from the dispersion of 50 bootstrap samples. The mean and median of the redshift distributions were obtained from the DIR calibration, detailed in Sect.2.3.

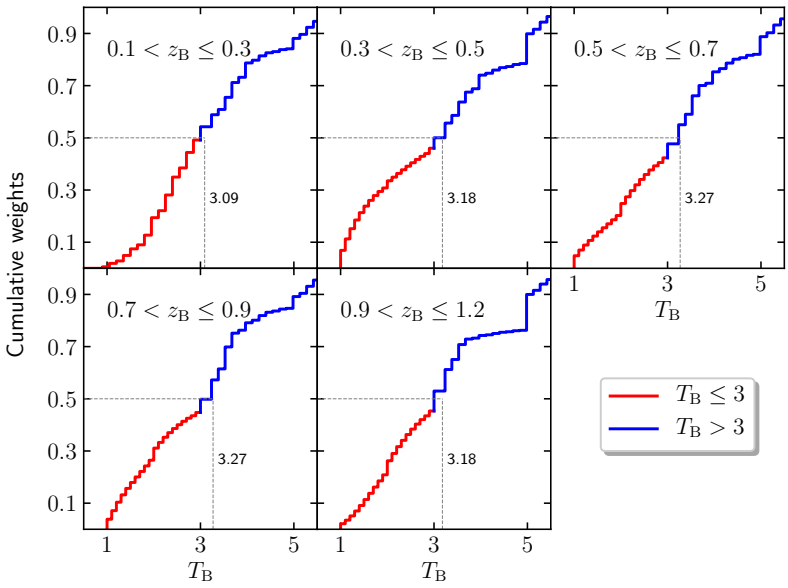

Figure 2.1: Cumulative *lens*fit-weighted distributions of T_B values. The dashed line indicates the ideal equal split in each tomographic bin, which is closely approximated by our split at $T_B = 3$.

variance in the calibration sample.

Since the spec-z catalogues cannot fully represent the photometric sample, it is necessary to weight spec-z objects to ensure an appropriate match between the spectroscopic and photometric distributions. The method, based on a kth nearest neighbour (kNN) approach, is detailed in Sect. 3 of Hildebrandt et al. (2017). In brief, it assigns weights to the spec-z objects by comparing the volume densities of the spec-z and photometric objects in the nine-band magnitude space ($ugriZYJHK_s$). As a result, KiDS+VIKING-like observations are required in the same areas as the aforementioned spec-z surveys. H20 have constructed these photometric observations from multiple sources, depending on the availability of specific data sets in those spec-z survey fields. We adopted the same sample and split it with the same criterion as used for the main KV450 sample to build two representatives of our two sub-samples.

The resulting redshift distributions of the two sub-samples are shown in Fig. 2.2. Also presented are the mean and median differences between these two redshift distributions (see Table 2.1 for separate values). The importance of photo-z calibration is demonstrated by the tails of the DIR redshift distributions compared to the ranges selected by the photo-z cuts (shaded regions). These differences between the DIR results and photo-z estimates are more significant in the red sub-sample, where an overall bias towards overestimating photo-z is shown. This may seem counter-intuitive at first, given the

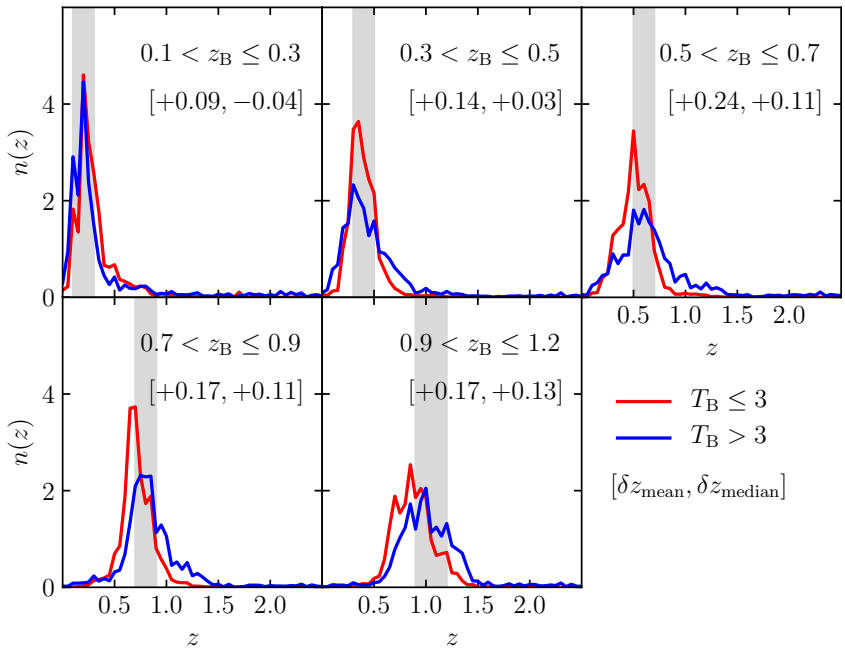

Figure 2.2: Redshift distributions for the two sub-samples, estimated using the DIR technique. Shaded regions correspond to photo-z cuts for the tomographic binning. Mean and median differences were calculated as $\delta z_{\mathrm{mean/median}} = z_{\mathrm{mean/median,blue}} - z_{\mathrm{mean/median,red}}$.

discussion presented in Sect. 2.1, which states that young stars can wash out spectral features for photo-z estimation, resulting in larger errors in bluer galaxies. However, we emphasise that the 'red' sub-sample defined in Sect. 2.2 is not purely red, but also includes Sbc and Scd types (see Sect. 2.2), which could worsen the photo-z estimates. For our purposes, we are interested in the redshift difference between the two sub-samples. As can be seen, the differences are significant, with the median differences as high as ∼0.13 and the mean differences ∼0.24 in certain bins. This level of difference will result in considerably different cosmic shear signals for the two sub-samples (see Sect. 2.5).

In practice, the DIR method is susceptible to various systematic effects, primarily induced by the incompleteness of the spec-z sample, due to selection effects and sample variance in the different spectroscopic surveys that comprise the spec-z catalogue (see Wright et al. 2020a for an updated method that is more robust against such incompleteness). To account for these potential systematic effects, H20 introduced five nuisance parameters δ_{z_i} in their model to allow for linear shifts of the redshift distributions $n_i(z) \rightarrow n_i(z + \delta_{z_i})$ (see Table 2.2). Priors for these parameters are obtained using a spatial bootstrapping approach. In our consistency tests described below, we

focus on an extension of these nuisance parameters to the colour-split sub-samples (see Sect. 2.6).

2.4 Calibration of shape measurements

The shape measurements are susceptible to various biases due to the noise of galaxy images, the complexity of galaxy shapes, selection effects, and so on (see Sect. 2 of K19 for a theoretical discussion). The weak lensing community has conducted several blind challenges to test the performance of shape measurement pipelines (see, e.g. Heymans et al. 2006; Massey et al. 2007; Bridle et al. 2010; Kitching et al. 2012; Mandelbaum et al. 2015). These tests, based on simplified image simulations, are useful for understanding common sources of shear bias but cannot eliminate biases in a specific survey. In particular, differences in selection criteria between surveys affect the shear bias. These residual biases need to be calibrated with dedicated, tailor-made image simulations (see, e.g. Hoekstra et al. 2015). Following Heymans et al. (2006), we quantify these residual biases using a linear parameterisation:

$$g_i^{\rm obs} = (1 + m_i) g_i^{\rm true} + c_i \ , \tag{2.1}$$

where $g_i^{\rm obs}$ and $g_i^{\rm true}$ are the observed and true gravitational shears, respectively, with $i = 1, 2$ referring to the two different components. In practice, we found isotropy in the m results, meaning that $m_1 \approx m_2$, so we simply adopt $m = (m_1 + m_2)/2$.

The two types of biases, m (the multiplicative bias) and c (the additive bias or c-term), have distinct sources and properties. The former is typically determined from image simulations, while the latter can be inferred directly from the data. As demonstrated in K19, shear biases depend not only on the selection function but also on the overall population of galaxies. Consequently, shear calibrations should be performed separately for samples containing different galaxy populations. This was the case for the various tomographic bins in the KV450 analysis and is even more relevant for our split analysis.

We therefore re-estimated the multiplicative biases in the two sub-samples using the COllege simulations (COSMOS-like lensing emulation of ground experiments, K19), which were also employed in the fiducial KV450 cosmic shear analysis. The main features of the COllege simulations are the observation-based input catalogue and the assignment of photometric redshifts. The input catalogue contains information on galaxy morphology and position from *Hubble* Space Telescope observations (Griffith et al. 2012) of the COSMOS field (Scoville et al. 2007). The photometric redshifts of simulated galaxies are assigned by cross-matching the input catalogue to the KiDS catalogue. This setup ensures a high level of realism in the simulated catalogue and allows us to analyse the simulated data using the same pipelines as for the real data. K19 have demonstrated that the simulated catalogue faithfully matches the full KV450 catalogue in all crucial properties, including galaxy shapes, sizes, and positions.

As anticipated, we observed significant differences in galaxy properties between the two sub-samples. We showcase one of these comparisons in Fig. 2.3, which compares

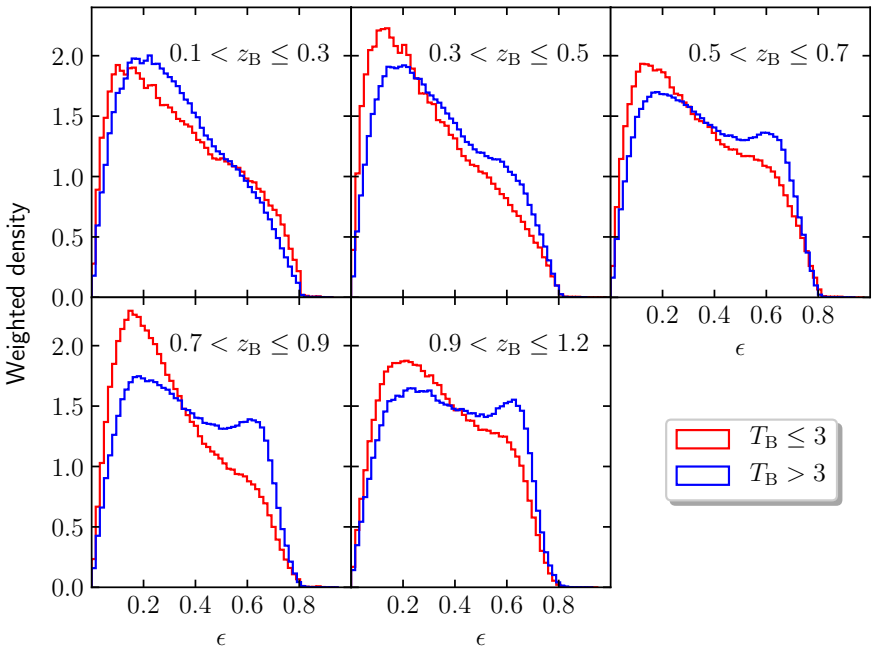

Figure 2.3: Normalised *lens*fit-weighted distributions of ellipticities for galaxies in the two sub-samples. The ellipticity is defined as $\epsilon = \sqrt{\epsilon_1^2 + \epsilon_2^2}$. We note that the differing distributions reflect distinct galaxy populations and suggest different shear biases in the two sub-samples.

the distributions of galaxy ellipticities. As previously mentioned in Sect. 2.1, ellipticity variance is one of the main sources of shape measurement biases (see also Viola et al. 2014) and thus serves as an indicator of the variance in shear biases between the two sub-samples.

Our calibration approach is identical to that used in the fiducial KV450 cosmic shear analysis. It adopts a re-weighting scheme referred to as 'Method C' in Fenech Conti et al. (2017) to account for slight differences between the observations and the simulations. The m value is reported per tomographic bin, using a weighted average of individual galaxies belonging to the corresponding tomographic bin. We direct readers to Sect. 6 of K19 for further details.

We present our estimates of multiplicative biases for the two sub-samples in Fig. 2.4, comparing them with the results from the whole sample. The five sections from top to bottom correspond to the five tomographic bins, ranging from lower to higher redshifts. We noticed some significant differences in the m values, particularly for the higher tomographic bins; these differences mainly arise from the ellipticity distributions shown in Fig. 2.3. However, when considering their impact on the cosmic shear signals, the

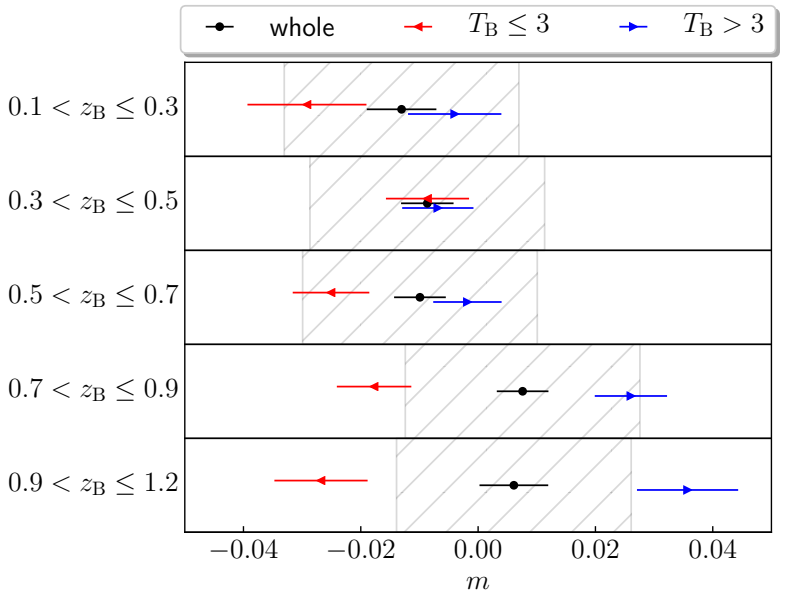

Figure 2.4: Multiplicative biases for the two sub-samples and the whole sample in each tomographic bin. Errors shown were estimated from bootstrapping. The hatched regions indicate the 0.02 error budget adopted by H20.

adjustments induced by these m-value differences are much smaller than those caused by the redshift differences, as demonstrated in Fig. 2.5. Therefore, we assume that residual systematics from the shear calibration are secondary and focus our consistency tests on the redshift calibration.

The treatment of additive bias in the fiducial KV450 cosmic shear analysis is sophisticated (see Sect. 4 of H20 for details). Briefly, the treatment can be summarised in three aspects: First, the value of c_i in each tomographic bin and in each patch is estimated by averaging over the measured galaxy ellipticities. These c_i values are then subtracted from the galaxy ellipticities before the shear correlation functions are calculated (Eq. 2.2). Second, a nuisance parameter δ_c is introduced into the model to account for a potential offset of the empirically determined c_i values. The result from forward-modelling suggests that δ_c is very close to 0 (see Table 2.2). Third, a position-dependent additive bias pattern in the ϵ_1 ellipticity component is introduced to account for an imperfection in the OmegaCAM detector chain. This pattern is publicly available as a supplementary file along with the main cosmic shear catalogues. Furthermore, another nuisance parameter A_c is introduced to allow an overall scaling of this 2D pattern (see Table 2.2).

We primarily adhered to this strategy for the additive bias calibration. We corrected the c-term per tomographic bin and per patch using the same empirical approach mentioned above. We also incorporated the 2D c-term pattern into our models. However, we excluded the two nuisance parameters δ_c and A_c from our model, as they do not have a significant impact on the fit.

2.5 Cosmic shear signal

The cosmic shear signal is encoded in the measured shapes of source galaxies as minor coherent distortions. Therefore, proper statistical measures and models are essential for a cosmic shear study. We delve into these processes in this section: Section 2.5.1 presents the construction of the joint data vector for the two sub-samples using the shear correlation functions; while Section 2.5.2 discusses the modelling of the measured data vector while taking into account various astrophysical and cosmological effects. Our analysis builds on the fiducial cosmic shear analysis of H20 but includes adjustments to accommodate our testing objectives.

2.5.1 Statistical measures

The shear signal is quantified by two-point shear correlation functions. These can be calculated from two tomographic bins i and j as follows

$$\xi_{\pm}^{ij}(\theta) = \frac{\sum_{ab} w_a w_b \left[\epsilon_t^i(\boldsymbol{x}_a)\epsilon_t^j(\boldsymbol{y}_b) \pm \epsilon_\times^i(\boldsymbol{x}_a)\epsilon_\times^j(\boldsymbol{y}_b)\right]}{(1+m^i)(1+m^j)\sum_{ab} w_a w_b}\,, \tag{2.2}$$

where $\epsilon_{t,\times}$ represent the tangential and cross ellipticities concerning the vector $\boldsymbol{x}_a - \boldsymbol{y}_b$ between a pair of galaxies (a, b), and w is the *lens*fit weight. The summation runs over all galaxy pairs within a designated spatial bin $\Delta\theta$ for each $\theta = |\boldsymbol{\theta}_b - \boldsymbol{\theta}_a|$. The multiplicative biases m^i, obtained in Sect. 2.4 for each tomographic bin i, are factored in.

We computed Eq. (2.2) separately for the two sub-samples using the publicly available TreeCorr code[2] (Jarvis et al. 2004). The spatial binning scheme is identical to the one used in H20, featuring nine logarithmically spaced bins within the $[0.''5, 300'']$ interval. We used the first seven bins for ξ_+ and the last six bins for ξ_-. These selection criteria help mitigate the impact of baryonic feedback on small scales and additive shear biases on large scales (for details, see H20). The joint data vector $(\xi_\pm^{\rm blue},\ \xi_\pm^{\rm red})$ that we created through these measurements comprises $(7+6)\times 15\times 2 = 390$ data points.

Our estimates of the data vector are presented in Fig. 2.5, with differences defined as $\Delta\xi_\pm = \xi_\pm^{\rm blue} - \xi_\pm^{\rm red}$. The accompanying error bars were derived from the analytical covariance matrix, as discussed in Sect. 2.6.1. The two sets of data vectors correspond to the results obtained with and without the multiplicative shear calibration. The discrepancy between them is minor, as anticipated given the relatively small m values (refer to Table 2.1). Some bins show non-zero trends, which can be attributed to the differing redshift distributions of the two sub-samples, as depicted in Fig. 2.2. The influence of the redshift distributions in explaining these measurements is further elaborated in the subsequent section.

[2] https://github.com/rmjarvis/TreeCorr

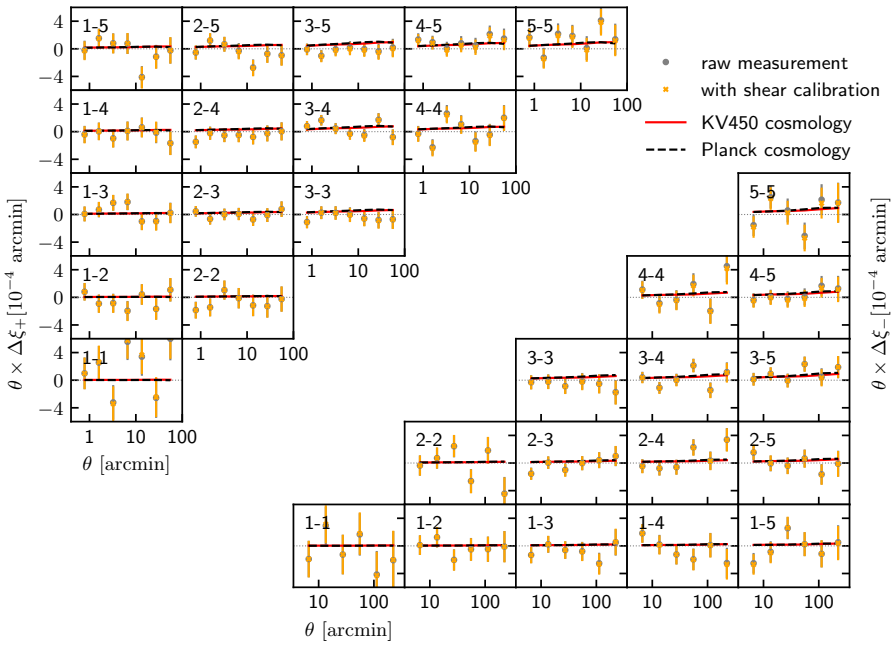

Figure 2.5: Differences between two-point shear correlation functions for the two sub-samples ($\Delta\xi_\pm = \xi_\pm^{\rm blue} - \xi_\pm^{\rm red}$). The error bars are defined as $\sigma_C = \sqrt{C_{\rm b,D} + C_{\rm r,D} - 2C_{\rm br,D}}$, where the subscript 'D' refers to the diagonal of a matrix, and the three unique parts of the complete covariance matrix are denoted as $C_{\rm b}$ for the blue sub-sample, $C_{\rm r}$ for the red sub-sample, and $C_{\rm br}$ for their cross-covariance. These errors closely match the measurement errors reported by the TreeCorr code ($\sigma_{\rm measure}/\sigma_C \gtrsim 0.8$), suggesting that the diagonal elements of the covariance matrix are predominantly influenced by measurement noise. The overall consistency between the two sets of data vectors—with and without shear calibration (represented by orange crosses and black dots, respectively)—implies that the multiplicative bias exerts a minimal effect in this study.

2.5.2 Theoretical modelling

The measured correlation functions $\xi_\pm^{ij}(\theta)$ are connected to the lensing convergence power spectrum $P_\kappa^{ij}(\ell)$ through (see e.g. Bartelmann & Schneider 2001)

$$\xi_\pm^{ij}(\theta) = \frac{1}{2\pi} \int {\rm d}\ell\, \ell P_\kappa^{ij}(\ell) J_{0/4}(\ell\theta)\,, \tag{2.3}$$

where ℓ signifies the angular wavenumber in the Fourier domain. The terms $J_{0/4}(\ell\theta)$ represent Bessel functions of the first kind, where J_0 corresponds to the zeroth-order (applied in the case of ξ_+), and J_4 denotes the fourth-order (used for ξ_-). Using the

Kaiser-Limber approximation (Limber 1953; Kaiser 1992, 1998; Loverde & Afshordi 2008), $P_\kappa^{ij}(\ell)$ is in turn related to the physical matter power spectrum P_δ, via

$$P_\kappa^{ij}(\ell) = \int_0^{\chi_\mathrm{H}} \mathrm{d}\chi \, \frac{q_i(\chi) q_j(\chi)}{[f_\mathrm{K}(\chi)]^2} P_\delta \left(\frac{\ell + 1/2}{f_\mathrm{K}(\chi)}, \chi \right) , \tag{2.4}$$

where χ represents the comoving radial distance, and $f_\mathrm{K}(\chi)$ denotes the comoving angular distance. The integration is carried out up to χ_H, which is the comoving horizon distance. The lensing efficiency $q_i(\chi)$ for a given tomographic bin i can be defined as

$$q_i(\chi) = \frac{3H_0^2\Omega_\mathrm{m}}{2c^2} \frac{f_\mathrm{K}(\chi)}{a(\chi)} \int_\chi^{\chi_\mathrm{H}} \mathrm{d}\chi' \, n_i(\chi') \frac{f_\mathrm{K}(\chi' - \chi)}{f_\mathrm{K}(\chi')} , \tag{2.5}$$

which is dependent on the redshift distribution of galaxies, represented as $n_i(\chi)\mathrm{d}\chi = n_i(z)\mathrm{d}z$, as well as various other cosmological parameters. This implies that variances in the redshift distributions between two sub-samples can lead to differences in their shear signals.

We used the Boltzmann-code CLASS (Blas et al. 2011) to compute the matter power spectrum, incorporating non-linear corrections from HMCode (Mead et al. 2016). In line with H20, we adopted a ΛCDM model comprising five primary cosmological parameters and an additional parameter representing baryonic feedback processes on smaller scales. These parameters include the densities of cold dark matter and baryons (Ω_CDM and Ω_b), the amplitude and the index of the scalar power spectrum ($\ln(10^{10}A_\mathrm{s})$, n_s), the scaled Hubble parameter (h), and the amplitude of the halo mass-concentration relation (B).

For consistency tests, it is not necessary to traverse the entire cosmological parameter space since these parameters remain the same for both sub-samples. Therefore, we set the aforementioned cosmological parameters to two distinct sets of best-fit values derived from KV450 (Hildebrandt et al. 2020) and *Planck* (Planck Collaboration et al. 2020) (see Table 2.2). This approach allows us to simplify our theoretical models while concurrently investigating potential dependencies on cosmology.

The final piece of information necessary for modelling the observed correlation functions is the intrinsic alignment (IA) of galaxies (Troxel & Ishak 2015; Joachimi et al. 2015). In the KiDS analyses, this effect is accounted for by incorporating a 'non-linear linear' IA model into the measured shear signal (Hirata & Seljak 2004; Bridle & King 2007):

$$\hat{\xi}_\pm = \xi_\pm + \xi_\pm^\mathrm{II} + \xi_\pm^\mathrm{GI} , \tag{2.6}$$

where $\hat{\xi}_\pm$ and $\xi_\pm$ represent the measured shear signal and the pure cosmic shear signal, respectively. The IA signals are incorporated in the form of $\xi_\pm^\mathrm{II}$ and $\xi_\pm^\mathrm{GI}$. The term $\xi_\pm^\mathrm{II}$ represents the 'intrinsic-intrinsic' correlation, that is, the correlation between the intrinsic ellipticities of nearby galaxies. The term $\xi_\pm^\mathrm{GI}$ stands for the 'gravitational-intrinsic' correlation, which refers to the correlation between the intrinsic ellipticity of a foreground galaxy and the shear experienced by a background galaxy. These two IA

terms can be calculated using the same formula as given in Eq. (2.3), but with the power spectra defined as

$$P_{\rm II}^{ij}(\ell) = \int_0^{\chi_{\rm H}} d\chi\ F^2(z) \frac{n_i(\chi) n_j(\chi)}{[f_{\rm K}(\chi)]^2} P_\delta\left(\frac{\ell+1/2}{f_{\rm K}(\chi)}, \chi\right) , \tag{2.7}$$

$$P_{\rm GI}^{ij}(\ell) = \int_0^{\chi_{\rm H}} d\chi\ F(z) \frac{q_i(\chi) n_j(\chi) + q_j(\chi) n_i(\chi)}{[f_{\rm K}(\chi)]^2} P_\delta\left(\frac{\ell+1/2}{f_{\rm K}(\chi)}, \chi\right) , \tag{2.8}$$

where

$$F(z) = -A_{\rm IA} C \rho_{\rm crit,0} \frac{\Omega_{\rm m}}{D_+(z)} . \tag{2.9}$$

The normalisation constant is given by $C = 5 \times 10^{-14} h^{-1} M_\odot^{-1} {\rm Mpc}^3$, and $\rho_{\rm crit,0}$ is the critical density today. The linear growth factor $D_+(z)$ is normalised to unity today. Consistent with H20, we disregarded the redshift or luminosity dependence of intrinsic alignments, assigning one nuisance parameter $A_{\rm IA}$ to account for these effects (for a more detailed treatment, see Fortuna et al. 2021a).

Equipped with all the necessary data, we can now forward-model the shear correlation functions. For illustrative purposes, we first maintained all model parameters constant, using the redshift distributions estimated in Sect. 2.3 to predict the combined data vector of the two sub-samples. The outcomes are depicted in Fig. 2.5. Two distinct predictions arise from two separate sets of cosmological parameters: the red solid line represents the KV450 best-fit values, and the black dashed line symbolises the *Planck* best-fit values. All other nuisance parameters align with the best-fit KV450 results as indicated in Table 2.2. Even with this simplified approach, the predicted results generally mirror the trends observed in the data, underscoring the redshift difference as the primary factor for the dissimilar shear correlation functions in the two sub-samples. Another notable characteristic is the resemblance between the two predictions derived from two distinct sets of cosmological parameters. This suggests that our test model is not significantly influenced by the adopted cosmological parameters. However, to accurately assess the goodness of fit and verify the robustness of the pipelines, we require a more meticulous Bayesian analysis, implementing appropriate test models and taking into account correlations between measurements.

2.6 Consistency tests

Quantifying internal consistency is no trivial task, given the correlations between measurements and the challenge of comparing different models. On the one hand, overlooking intrinsic correlations between measurements can lead to unreliable conclusions. As demonstrated by Köhlinger et al. (2019), failing to consider these correlations can confound residual systematics with the overall goodness of fit. On the other hand, null tests based on global summary statistics, such as Bayesian evidence, are practically challenging for high-dimensional models (see e.g. Trotta 2008). Furthermore, varying prior choices between hypotheses can complicate the interpretation of the final

Table 2.2: Model parameters and their best-fit values from KV450 cosmic shear analysis (Hildebrandt et al. 2020) and the *Planck* CMB analysis (Planck Collaboration et al. 2020).

Parameter	KV450	*Planck*	Definition
$\Omega_{\rm CDM}h^2$	0.058	0.120	CDM density today
$\Omega_{\rm b}h^2$	0.022	0.022	Baryon density today
$\ln(10^{10}A_{\rm s})$	4.697	3.045	Scalar spectrum amplitude
$n_{\rm s}$	1.128	0.966	Scalar spectrum index
h	0.780	0.673	Hubble parameter
B	2.189	-	Baryon feedback amplitude
$A_{\rm IA}$	0.494	-	IA amplitude
$\delta_c \times 10^5$	2.576	-	c-term offset
A_c	1.143	-	2D c-term amplitude
δ_{z_1}	−0.006	-	Bin 1 offset
δ_{z_2}	0.001	-	Bin 2 offset
δ_{z_3}	0.026	-	Bin 3 offset
δ_{z_4}	−0.002	-	Bin 4 offset
δ_{z_5}	0.003	-	Bin 5 offset

The first five parameters are the standard cosmological parameters. The remaining parameters are nuisance parameters introduced by Hildebrandt et al. (2020) to account for various effects associated with cosmic shear analysis. The KV450 best-fit values are derived from the primary Monte Carlo Markov Chain, which is publicly available at `http://kids.strw.leidenuniv.nl/cosmicshear2018.php`. The *Planck* best-fit values correspond to the TT, TE, EE+lowE+lensing results with the `Plik` likelihood (Table 1 of Planck Collaboration et al. 2020).

results (Handley & Lemos 2019b; Lemos et al. 2019).

In this section, we address these challenges. First, we construct an analytical covariance matrix to account for all the correlations between measurements (Sect. 2.6.1). Then, we conduct a Bayesian analysis using dedicated test parameters to quantify potential discrepancies between measurements from the two sub-samples (Sect. 2.6.2). Our conclusions are based on the posterior distributions of these test parameters. This approach allows us to strike a balance between model accuracy and simplicity.

The modelling pipeline described below is publicly available[3]. It is a modified

[3]`https://github.com/lshuns/montepython_KV450`

version of the MontePython package (Audren et al. 2013; Brinckmann & Lesgourgues 2018) using the PyMultiNest algorithm (Buchner et al. 2014), which is a Python wrapper of the nested sampling algorithm MultiNest (Feroz et al. 2009). The original MontePython package was used for the KV450 cosmological analysis in H20 and for the consistency tests with a split of data vector (Köhlinger et al. 2019).

2.6.1 Covariance matrix

We estimated the covariance matrix for the joint data vector assembled in Sect. 2.5.1 using the analytical model developed in Hildebrandt et al. (2017), H20, and Joachimi et al. (2021). This analytical approach represents an advancement over traditional numerical or Jackknife methods, offering benefits in managing noise effects and finite survey areas. Here, we provide a brief overview of the key aspects of this analytical approach, while detailed discussions can be found in Sect. 5 of Hildebrandt et al. (2017) and Joachimi et al. (2021).

The analytical model consists of three components: a Gaussian term related to sample variance and shape noise, a non-Gaussian term originating from in-survey modes, and a third term, also non-Gaussian, derived from super-survey modes (known as super-sample covariance, or SSC). The Gaussian term is calculated following Joachimi et al. (2008), with a transfer function taken from Eisenstein & Hu (1998) and non-linear corrections from Takahashi et al. (2012). The information about the sources used is detailed in Table 2.1; it includes the effective galaxy number density ($n_{\rm eff}$) and the weighted ellipticity dispersion ($\sigma_{\epsilon,i}$). The second non-Gaussian term is derived using the formalism proposed by Takada & Hu (2013), incorporating the halo mass function and halo bias from Tinker et al. (2010). The halo profile is described using a Fourier-transform version of the NFW model (Navarro et al. 1996; Scoccimarro et al. 2001), with the concentration-mass relation taken from Duffy et al. (2008). The final SSC term is modelled once again using the formalism from Takada & Hu (2013). The survey footprint is modelled with a HEALPix map (Górski et al. 2005).

The shear calibration outlined in Sect. 2.4 also entails uncertainties. Following the approach used in H20 and Wright et al. (2020b), we considered a systematic uncertainty $\sigma_m = 0.02$ for the multiplicative biases as estimated by K19. This uncertainty is propagated into the covariance matrix using the relation $C_{ij}^{\rm cal} = 4\xi_i^{\rm T}\xi_j^{\rm T}\sigma^2 m + C_{ij}$, where $\xi^{\rm T}$ is the joint data vector predicted with the KV450 best-fit values and the DIR redshift distributions (refer to Sect. 2.3 for more details). The error associated with the additive biases was neglected due to its minimal impact. For a more detailed discussion on this topic, please refer to Appendix D4 of Hildebrandt et al. (2017).

The final correlation matrix for the joint data vector is presented in Fig. 2.6. There are noticeable contributions from the off-diagonal regions, which highlight the significant correlations within each individual sub-sample as well as between the two sub-samples. The importance of potential correlations between parts of a split was previously emphasised in Köhlinger et al. (2019), but here we provide a more direct confirmation. By incorporating the complete covariance matrix into our consistency

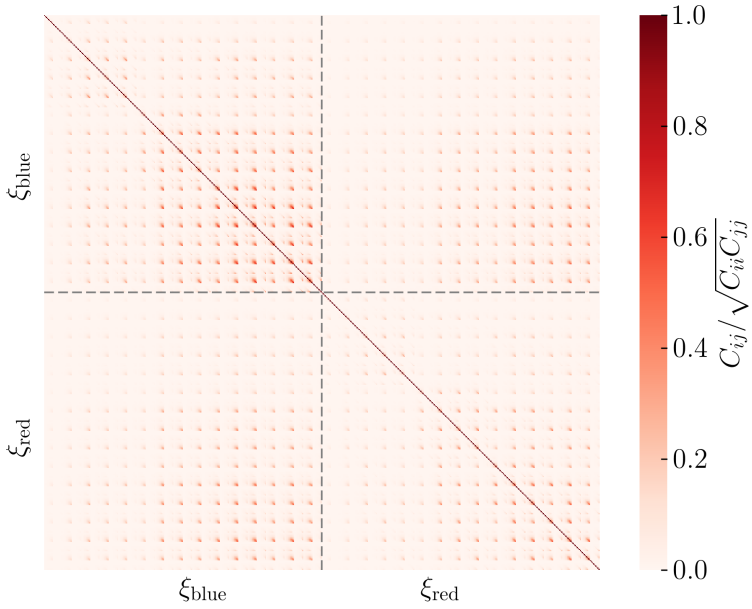

Figure 2.6: Analytical correlation matrix for the joint data vector. To highlight the correlations, the covariance C_{ij} is normalised using the diagonal $\sqrt{C_{ii}C_{jj}}$.

tests, we automatically account for all the data correlations.

We examined the relative contributions of the Gaussian and non-Gaussian terms to the complete covariance matrix. Generally, the Gaussian term dominates the diagonal parts of the covariance matrix, while the non-Gaussian term contributes more significantly to the off-diagonal regions. This pattern is also clearly outlined in Joachimi et al. (2021).

Given that our test model is especially sensitive to the difference $\Delta\xi$ between the two sub-samples, we built the covariance matrix of $\Delta\xi$ as $C_\Delta = C_{\rm blue} + C_{\rm red} - 2C_{\rm cross}$. We then compared this to the covariance matrices of the individual data vectors ($\xi_{\rm blue}$ or $\xi_{\rm red}$). Our findings indicate that the non-Gaussian contributions are significantly diminished in C_Δ, with an overall reduction of $\lesssim 75\%$ compared to $C_{\rm blue}$. The Gaussian contributions are also slightly reduced, primarily in the off-diagonal regions. These reductions in the covariance matrix C_Δ can be explained by the cancellation of sample variance. Hence, we confirmed that our test model is robust against uncertainties in the sample variance and alterations in the cosmological parameters.

2.6.2 Test setup

With the covariance matrix in place, we can explore the parameter space using a Bayesian analysis. Our main goal is to evaluate whether a unified set of nuisance parameters could adequately account for the residual systematics in the two sub-samples. Consequently, we opted to hold all cosmological parameters constant, under the assumption that the two sub-samples should intrinsically have identical values for these parameters.

To substantiate this assumption, we ran an additional test where we allowed cosmological parameters to vary. The results are in an agreement with our fixed-cosmology framework, revealing minimal degeneracy between cosmological parameters and test parameters. Therefore, we maintained our approach of fixing cosmological parameters, simplifying the likelihood function and preventing the unnecessary traversal of a high-dimensional parameter space.

To ensure robustness against any possible residual effects stemming from an inaccurate selection of cosmological parameters, we implemented two different settings. One setting used cosmological parameters derived from the KV450 cosmic shear analysis, while the other employed parameters from the *Planck* CMB results (see Table 2.2). This approach enabled us to factor in potential discrepancies in our choice of cosmological parameters.

Our test model, $\mathcal{H}_1$, contains six test parameters besides the nuisance parameters used in H20. These test parameters are a shift in IA amplitude, denoted as $A_{\rm IA,s}$, and shifts in redshift offsets, represented as $\delta_{z_i,\rm s}$. They are implemented in the two sub-samples as

$$X_{\rm blue/red} = X \pm X_{\rm s} \,, \tag{2.10}$$

where X stands for either the $A_{\rm IA}$ or δ_{z_i} parameters, while $X_{\rm s}$ designates the corresponding test parameters. The blue sub-sample is characterised by the positive sign, while the red sub-sample is associated with the negative sign.

While a discrepancy in the IA signal is anticipated, differences in the redshift offsets should be non-existent if the calibration pipeline is impervious to sample-related systematics. Any non-zero values for $\delta_{z_i,\rm s}$ would indicate the presence of residual systematics that are not fully encompassed by the common nuisance parameters. Therefore, our conclusions are primarily grounded on the posterior distributions of these test parameters.

For control purposes, we also established a base model, $\mathcal{H}_0$, with the same set of nuisance parameters as in H20, to model the joint data vector derived from our two sub-samples. It contains six free nuisance parameters: the amplitude of the IA signal $A_{\rm IA}$ (refer to Sect. 2.5.2) and the redshift offset δ_{z_i} for each tomographic bin i (refer to Sect. 2.3).

However, it is important to note that this assumption is stronger than what is mandated by data consistency. Given that the IA signal is dependent on the galaxy population, it is not expected to be identical across the two sub-samples.

The prior distributions for all free parameters are detailed in Table 2.3. The common nuisance parameters leverage the priors from H20. Specifically, $A_{\rm IA}$ employs a wide flat prior, while δ_{z_i} adopts Gaussian priors, the variance of which is determined using a spatial bootstrapping technique during the redshift calibration process (refer to Sect. 3.2 of H20). For the six new test parameters in the test model $\mathcal{H}_1$, we selected wide, uninformative priors. These choices, as demonstrated in Sect. 2.7, incorporate prior knowledge of redshift uncertainties into the common nuisance parameters, while simultaneously enabling a comprehensive exploration of the test parameters. We emphasise that the

main objective of our test is to ascertain whether the KV450 nuisance parameters are adequate for capturing any residual systematics.

Our testing method, which does not rely on Bayesian evidence to identify tensions, avoids the 'suspiciousness' issue associated with common model-selection methods (Lemos et al. 2019). In this regard, our approach is akin to the second tier of the Bayesian consistency tests proposed by Köhlinger et al. (2019). However, rather than duplicating the cosmological parameters and drawing conclusions from the posterior distributions of differences in these parameters, we concentrate on the nuisance parameters, particularly those tied to redshift calibration.

Another key difference lies in our approach to data splitting. We performed a colour-based split of the source galaxies, repeating measurements and calibrations for the sub-samples, whereas Köhlinger et al. (2019) based their analysis on a split of the measured correlation functions. Consequently, our method is more attuned to potential inconsistencies within the source samples, while their approach offers a more global test of residual systematics and their impact on the ultimate cosmological results. Thus, our test serves as a complementary examination of pipeline robustness alongside theirs.

2.7 Results

The primary outcomes of our consistency tests are depicted in Fig. 2.7, where we present the marginal posterior constraints of the five test parameters, $\delta_{z_i,\mathrm{s}}$, introduced in Sect. 2.6.2. Each of the five sections in the plot corresponds to one of the five tomographic bins. The two sets of results, drawn from the KV450 best-fit cosmology (represented by red lines) and the *Planck* best-fit cosmology (black lines), are in agreement, reinforcing the notion that our test model is not sensitive to the specific choice of cosmological parameters.

All of the values are consistent with zero within approximately 1.5σ, which suggests that the KV450 calibration pipelines are effectively correcting for these sample-related systematics. This finding further implies that there is no need to introduce additional nuisance parameters for the current analysis.

The two tomographic bins that exhibit slightly non-zero differences are the second bin (approximately 1.2σ) and the third bin (approximately 1.3σ). The interpretation of such a level of difference is nuanced, given the statistical power of current data. We emphasise that the $\delta_{z_i,\mathrm{s}}$ parameters we constrained here represent the shifts in the redshift offsets within the two sub-samples. These shifts are expected to be larger than the mean redshift offsets (δ_{z_i}), considering the notable redshift differences between the two sub-samples and the breadth of the DIR redshift distributions (refer to Fig. 2.2).

As demonstrated in Table 2.3, all $\delta_{z_i,\mathrm{s}}$ values are less than the width of the underlying redshift distributions and are approximately zero within the uncertainties. This observation attests to the overall precision of the DIR redshift distributions.

Table 2.3 presents the posterior results for all free parameters, as well as the best-fit χ^2 values for all models. While we do not base our conclusions strictly on the χ^2 test, due to the complex nature of Bayesian models where dimensionality is not directly

defined by the number of free parameters (see, for example, Handley & Lemos 2019a), it is still informative to compare these values. Upon considering the number of free parameters, it appears that the test model $\mathcal{H}_1$ is statistically on par with the control model $\mathcal{H}_0$. This further reinforces our initial conclusion regarding the suitability of the current nuisance parameters in accounting for any remaining systematic errors.

Figure 2.8 shows the contour plot for the test model, with a notable feature being the pronounced degeneracy between $A_{\rm IA,s}$ and $\delta_{z_i,{\rm s}}$ in the lower redshift bins (refer to Fig. 2.8). This correlation primarily contributes to the ambiguity in the test parameters. The complex relationship between the IA signal and redshift uncertainties was also observed in Wright et al. (2020b), where an updated redshift calibration of the KV450 data resulted in a negligible IA amplitude. Our observation underlines the complexity inherent in deciphering the apparent IA signal.

We performed an additional test in which we constrained $\delta_{z_i,{\rm s}} = 0$ in the test model $\mathcal{H}_1$. The result was a significantly positive $A_{\rm IA,s}$ value, suggesting $A_{\rm IA,blue} > A_{\rm IA,red}$. This contradicts established IA studies (see Joachimi et al. 2015 for a review), implying that IA parameters could potentially mask issues with redshift estimates. Therefore, we must exercise caution when interpreting the IA parameters.

To further investigate the influence of the IA parameters in our test model, we executed an additional test, $\mathcal{T}_1$, where $A_{\rm IA,s}$ was held constant at zero. This approach maximises the shifts of the redshift offsets by disregarding the IA difference in the two sub-samples. Even with this conservative estimate, the shifts are $\lesssim 2.1\sigma$ for all redshift bins, with the most significant values still appearing in the third bin (refer to Table 2.3).

2.8 Summary and discussion

We conducted an internal consistency test on the KV450 cosmic shear analysis by dividing source galaxies based on colour, yielding two statistically equivalent sub-samples comprising distinct galaxy populations (refer to Figs. 2.1, 2.2, and 2.3). Uniform measurements and calibrations were applied to these sub-samples, and we evaluated changes in the two-point correlation functions due to known variations in redshift distributions and multiplicative biases (see Fig. 2.5). With cosmological parameters fixed, we used a Bayesian analysis and specific test parameters to scrutinise the internal consistency of observational nuisance parameters, focusing on those related to redshift distributions. We noticed a degeneracy between redshift uncertainties and the inferred IA amplitude for lower redshift bins. However, we found no signs of internal inconsistency in the KV450 data, affirming the adequacy of the current approach—utilising a common set of nuisance parameters to linearly shift redshift distributions—for addressing residual systematics in redshift calibration.

The internal consistency test we introduced is resilient to uncertainties in background cosmology and cosmic variance. It can be employed in upcoming cosmic shear surveys prior to making any cosmological inference. This weak sensitivity to cosmology aligns with the 'shear-ratio' test (Jain & Taylor 2003; Schneider 2016; Unruh et al. 2019), which has been used to verify the precision of redshift distributions in current cosmic

Table 2.3: Priors and posterior results for all models.

Parameter	Prior	KV450			Planck		
		$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{T}_1$	$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{T}_1$
$A_{\rm IA}$	$[-6, 6]$	$1.442^{+0.826}_{-0.898}$	$1.049^{+0.818}_{-0.871}$	$0.976^{+0.776}_{-0.804}$	$1.741^{+0.507}_{-0.533}$	$1.358^{+0.463}_{-0.495}$	$1.340^{+0.466}_{-0.476}$
δ_{z_1}	0.000 ± 0.039	$-0.012^{+0.037}_{-0.037}$	$-0.000^{+0.035}_{-0.038}$	$0.001^{+0.035}_{-0.037}$	$-0.037^{+0.028}_{-0.036}$	$-0.008^{+0.036}_{-0.040}$	$-0.005^{+0.038}_{-0.039}$
δ_{z_2}	0.000 ± 0.023	$-0.006^{+0.019}_{-0.023}$	$-0.001^{+0.022}_{-0.021}$	$-0.000^{+0.021}_{-0.022}$	$-0.011^{+0.019}_{-0.019}$	$-0.003^{+0.020}_{-0.022}$	$-0.002^{+0.021}_{-0.019}$
δ_{z_3}	0.000 ± 0.026	$0.009^{+0.023}_{-0.022}$	$0.006^{+0.022}_{-0.023}$	$0.006^{+0.022}_{-0.026}$	$0.020^{+0.020}_{-0.018}$	$0.019^{+0.020}_{-0.021}$	$0.021^{+0.020}_{-0.020}$
δ_{z_4}	0.000 ± 0.012	$-0.002^{+0.012}_{-0.011}$	$-0.001^{+0.012}_{-0.011}$	$-0.002^{+0.012}_{-0.012}$	$0.003^{+0.011}_{-0.012}$	$0.003^{+0.012}_{-0.012}$	$0.003^{+0.012}_{-0.013}$
δ_{z_5}	0.000 ± 0.011	$0.002^{+0.011}_{-0.011}$	$0.003^{+0.012}_{-0.010}$	$0.002^{+0.011}_{-0.011}$	$0.006^{+0.012}_{-0.011}$	$0.005^{+0.011}_{-0.010}$	$0.006^{+0.011}_{-0.011}$
$A_{\rm IA,s}$	$[-6, 6]$	-	$0.571^{+1.178}_{-1.337}$	-	-	$0.536^{+0.793}_{-0.967}$	-
$\delta_{z_1,{\rm s}}$	$[-0.3, 0.3]$	-	$0.032^{+0.142}_{-0.085}$	$0.079^{+0.076}_{-0.069}$	-	$0.004^{+0.122}_{-0.098}$	$0.072^{+0.057}_{-0.066}$
$\delta_{z_2,{\rm s}}$	$[-0.3, 0.3]$	-	$0.080^{+0.087}_{-0.068}$	$0.116^{+0.048}_{-0.055}$	-	$0.039^{+0.059}_{-0.053}$	$0.069^{+0.032}_{-0.033}$
$\delta_{z_3,{\rm s}}$	$[-0.3, 0.3]$	-	$0.066^{+0.057}_{-0.051}$	$0.087^{+0.037}_{-0.041}$	-	$0.040^{+0.042}_{-0.039}$	$0.060^{+0.027}_{-0.030}$
$\delta_{z_4,{\rm s}}$	$[-0.3, 0.3]$	-	$0.002^{+0.048}_{-0.050}$	$0.014^{+0.044}_{-0.045}$	-	$0.009^{+0.039}_{-0.039}$	$0.019^{+0.037}_{-0.037}$
$\delta_{z_5,{\rm s}}$	$[-0.3, 0.3]$	-	$-0.002^{+0.053}_{-0.051}$	$0.005^{+0.051}_{-0.050}$	-	$0.008^{+0.048}_{-0.046}$	$0.015^{+0.046}_{-0.046}$
$N_{\rm data}$	-	390	390	390	390	390	390
$N_{\rm para}$	-	6	12	11	6	12	11
χ^2	-	366.8	356.4	356.1	357.5	364.5	364.4

The table shows parameters for the test model $\mathcal{H}_1$ (12 free parameters), control model $\mathcal{H}_0$ (6 common parameters), and test setting $\mathcal{T}_1$ (ignoring IA signal differences). The initial six are common nuisance parameters; the subsequent six address potential discrepancies between sub-samples (see Eq. 2.10). Bracketed priors denote top-hat ranges while values with errors indicate Gaussian distributions. Displayed results are mean posterior values with χ^2 values for maximum likelihood.

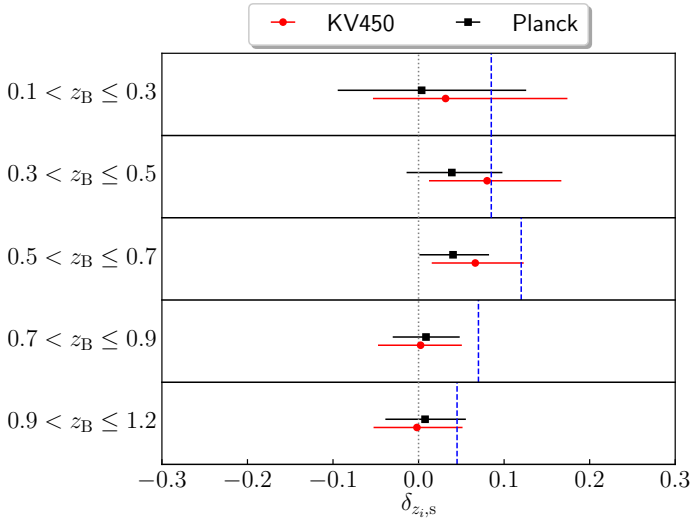

Figure 2.7: Constraints on $\delta_{z_i,\rm s}$ for each tomographic bin under the $\mathcal{H}_1$ model. The errors depicted represent the 68% credible intervals derived from the MCMC analysis. For contextual comparison, the vertical blue lines indicate half of the mean differences between the reconstructed DIR redshift distributions of the two sub-samples (as seen in Fig. 2.2).

shear surveys (Heymans et al. 2012; H20; Giblin et al. 2021). The 'shear-ratio' test, a cross-correlation method, relies on the galaxy-galaxy lensing signals from two or more source samples at varying redshift bins. Consequently, these two tests are sensitive to different systematics, making them complementary tools.

While our discussion primarily focused on redshift calibration, we discovered that the test is also contingent on our assumptions about IA signals (refer to Fig. 2.8). Without an extensive exploration of IA models, our test is already capable of identifying the degeneracy between the IA signals and redshift uncertainties, a concept previously hinted at in other studies(see Sect. 6.6 of Hildebrandt et al. 2017). Recently, Samuroff et al. (2019) conducted an analysis similar to ours, but based on DES data. Their focus was the IA signal and cosmological parameters, while they marginalised over observational nuisance parameters. While this differs from our approach, it links to our test via the IA signals, which both tests examined. They achieved more precise constraints on the IA signals in sub-samples by employing a range of IA models. We can similarly enhance our test model to glean deeper insights into IA signals and their correlation with other nuisance parameters in future cosmic shear data.

Acknowledgements

We thank Shahab Joudaki for carefully reading the manuscript and providing useful comments. SSL is supported by NOVA, the Netherlands Research School for Astronomy.

KK acknowledges support by the Alexander von Humboldt Foundation. HHo acknowledges support from Vici grant 639.043.512, financed by the Netherlands Organisation for Scientific Research (NWO). HHi is supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft (Hi 1495/5-1) as well as an ERC Consolidator Grant (No. 770935). This work is based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 100.A-0613, 102.A-0047, 179.A-2004, 177.A-3016, 177.A-3017, 177.A-3018, 298.A-5015, and on data products produced by the KiDS consortium.

Author Contributions: All authors contributed to the development and writing of this paper. The authorship list is given in two groups: the lead authors (SSL, KK, HHo), followed by an alphabetical group of key contributors to both the scientific analysis and the data products.

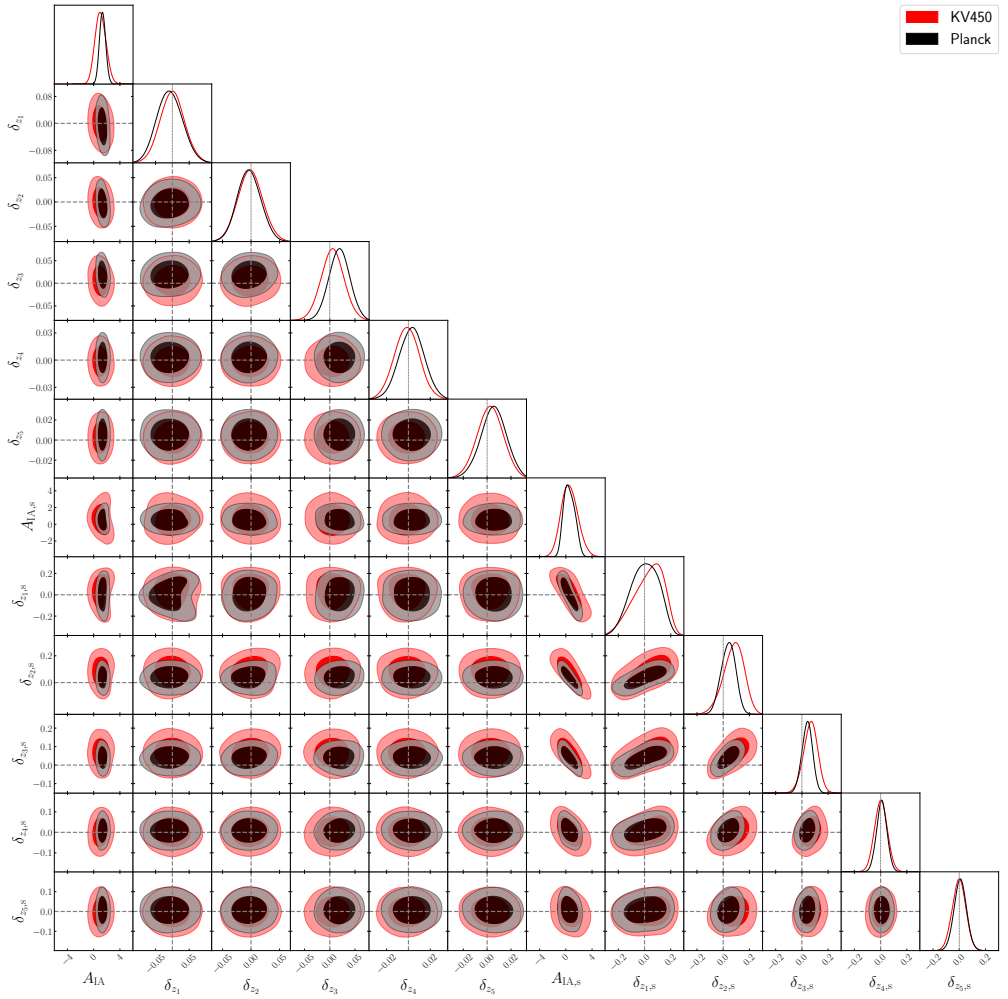

Figure 2.8: Contour plots outlining the 68% and 95% credible regions for all free parameters within the $\mathcal{H}_1$ model. The plotting ranges align with the prescribed prior ranges. Dashed lines demarcate zero values, which would represent the ideal case. The two different colours correlate with the two sets of results derived from the KV450 and *Planck* cosmological values, respectively. The faint degeneracy observed between δz_i, s in the lower redshift bins is an artefact of the substantial degeneracy between $A_{\rm IA,s}$ and δ_{z_i},s. This degeneracy dissipates in the $\mathcal{T}_1$ test setting, wherein $A_{\rm IA,s}$ is held constant at zero.

CHAPTER 3

KiDS-Legacy calibration: Unifying shear and redshift calibration with the SKiLLS multi-band image simulations

Abstract

We present SKiLLS, a suite of multi-band image simulations for the weak lensing analysis of the complete Kilo-Degree Survey (KiDS), dubbed KiDS-Legacy analysis. The resulting catalogues enable joint shear and redshift calibration, enhancing the realism and accuracy over previous efforts. By integrating cosmological simulations with high-quality imaging data, we created a large volume of simulated galaxies, faithfully mirroring the realistic galaxy properties to a sufficient depth. We accounted for point spread function (PSF) variability across CCD images, while also considering variations in stellar density and noise levels between different pointings. Using variable shear fields, we studied blended systems at different redshifts. Our results show a discernible redshift-shear bias correlation, underscoring this subtle yet noticeable higher-order blending effect. Additionally, we detected a minor but noteworthy impact of PSF modelling errors on shear bias. Finally, we conducted sensitivity tests, affirming the robustness of our fiducial shape measurement algorithm, *lens*fit, within the KiDS weak lensing analysis requirement. For future, more stringent weak lensing surveys, we recommend further study into blending effects, refinement of PSF modelling, and improvement of shape measurement techniques to minimise their sensitivity to variations in galaxy properties.

S.-S. Li, K. Kuijken, H. Hoekstra et al.
Astronomy & Astrophysics, 670, A100

3.1 Introduction

Weak gravitational lensing, the small deflection of light rays caused by inhomogeneous matter distributions, is a powerful tool for observational cosmology as an unbiased tracer of gravity (see Bartelmann & Schneider 2001, for a review). It allows us to study the underlying distribution of both baryonic and dark matter (see Refregier 2003; Hoekstra & Jain 2008; Kilbinger 2015, for some reviews). Together with redshift estimates for the sources, the cosmological lensing signal can even quantify the growth of the cosmic structure and infer the properties of dark energy (e.g. Hu 1999; Huterer 2002). Recent weak lensing surveys, including the Kilo-Degree Survey + VISTA Kilo-degree INfrared Galaxy (KiDS+VIKING) survey (de Jong et al. 2013; Edge et al. 2013)[1], the Dark Energy Survey (DES, Dark Energy Survey Collaboration et al. 2016)[2], and the Hyper Suprime-Cam (HSC) survey (Aihara et al. 2018)[3], provided some of the tightest cosmological constraints on the clumpiness of matter in the local Universe (Heymans et al. 2021; Abbott et al. 2022; Hamana et al. 2020). The upcoming so-called Stage IV surveys, such as the ESA *Euclid* space mission (Laureijs et al. 2011)[4], the Rubin Observatory Legacy Survey of Space and Time (LSST, Ivezić et al. 2019)[5], and the NASA *Nancy Grace Roman* space telescope (Spergel et al. 2015)[6], will advance the field significantly by increasing the statistical power of weak lensing measurements by more than an order of magnitude.

While promising, measuring the weak lensing signals to the desired accuracy in practice is demanding (see Mandelbaum 2018, for a recent review). In particular, the observed images of distant galaxies are smeared by the point spread function (PSF) and contain pixel noise, biasing the measurements of galaxy shapes (e.g. Paulin-Henriksson et al. 2008; Massey et al. 2013; Melchior & Viola 2012; Refregier et al. 2012). These issues drove the early development of many shape measurement methods and triggered a series of community-wide blind challenges based on image simulations, including the Shear TEsting Programme (STEP, Heymans et al. 2006; Massey et al. 2007) and the Gravitational LEnsing Accuracy Testing (GREAT, Bridle et al. 2010; Kitching et al. 2012; Mandelbaum et al. 2015). These early efforts illuminated some crucial issues and paved the way to calibrate the systematic biases for an actual survey using image simulations.

Early applications of simulation-based calibration have already demonstrated that the calibration accuracy depends on how well the simulation matches the survey under consideration, especially the observational conditions and the galaxy properties (e.g. Miller et al. 2013; Hoekstra et al. 2015, 2017; Samuroff et al. 2018). Therefore, recent implementations carefully mimic the data processing procedures and use morphological measurements from deep imaging surveys to reproduce the measured galaxy properties

[1] https://kids.strw.leidenuniv.nl
[2] https://darkenergysurvey.org
[3] https://hsc.mtk.nao.ac.jp/ssp/
[4] https://sci.esa.int/web/euclid/
[5] https://www.lsst.org/
[6] https://roman.gsfc.nasa.gov/

for a specific survey (e.g. Mandelbaum 2018; Kannawadi et al. 2019 hereafter K19; MacCrann et al. 2022). Alternately, newer methods, such as the Bayesian Fourier Domain (Bernstein & Armstrong 2014) and METACALIBRATION (Huff & Mandelbaum 2017; Sheldon & Huff 2017), seek an unbiased estimate of the shear either using deeper data as a prior or directly calibrating the measurements using the observed data.

Recent studies highlighted the effect of blending. The blending effect occurs when two or more objects are close together in the image plane, so their light distributions overlap. It introduces biases during both the selection and measurement processes. For example, Hartlap et al. (2011) found that the rejection of recognised blends alters the selection function of the final sample (see also Chang et al. 2013). In some circumstances, blended systems are so close that they appear as single objects. These unrecognised blends increase the shape noise by decreasing the number density and widening the measured ellipticity dispersion (e.g. Dawson et al. 2016; Mandelbaum et al. 2018). Even if the blended objects are below the detection limit, they still introduce correlated noise that affects the detection and measurement of the adjacent bright galaxies (e.g. Hoekstra et al. 2015, 2017; Samuroff et al. 2018), an effect that becomes even more dramatic when the clustering of galaxies is considered (Euclid Collaboration et al. 2019). Given all of these concerns, it is essential for image simulations to contain faint objects and physical clustering features.

More concerns arise when considering a tomographic analysis, which is at the core of current and future weak lensing surveys. From the shear estimate side, the tomographic binning approach introduces further selections that link the shear bias to redshift estimates (K19, MacCrann et al. 2022). From the redshift estimate side, redshift calibration methods need mock photometric catalogues to verify their performance. These mock catalogues must resemble the target data in object selections and photometric measurements, which are challenging to address at the catalogue level (Hoyle et al. 2018; Wright et al. 2020a; van den Busch et al. 2020; DeRose et al. 2022).

All these issues become even more challenging for the KiDS-Legacy analysis, the weak lensing analysis of the complete KiDS. It covers the entire 1350 $\deg^2$ survey area, a ∼35% increase over the latest KiDS release (KiDS-DR4, Kuijken et al. 2019). More importantly, thanks to the deeper i-band observations and dedicated observations in spectroscopic survey fields, the KiDS-Legacy analysis aims to unleash the power of high-redshift samples (up to a redshift of $z\sim2$). The improved statistical power, however, makes a higher demand on the shear and redshift calibrations, including an assessment of the cross-talk between the systematic errors in the shear and redshift estimates.

In this chapter, we present SKiLLS (SURFS-based KiDS-Legacy-Like Simulations), the third generation of image simulations for KiDS following SCHOo*l* (Simulations Code for Heuristic Optimization of *lens*fit, Fenech Conti et al. 2017, hereafter FC17) and COllege (COSMOS-like lensing emulation of ground experiments, K19). By simulating multi-band imaging that includes realistic galaxy evolution and clustering in terms of colour, morphology and number density, SKiLLS allows for the simultaneous measurement of shear and photometric redshifts from the same simulation. This study, therefore, provides the first joint calibration of these two key observables for cosmic

shear analyses. With our approach, we provide a natural solution to address the expected cross-talk between shear and redshift bias, accounting for the impact of blends that carry different shears (Dawson et al. 2016; Mandelbaum et al. 2018; MacCrann et al. 2022). We also release our simulation pipeline, which contains customisable features for general use by other surveys[7].

The remainder of this chapter is structured as follows. In Sect. 3.2, we build input catalogues for image simulations. Then in Sect. 3.3, we detail the creation and processing of the KiDS-like multi-band images, starting from instrumental setups and ending with photometric catalogues. Section 3.4 reviews our fiducial shape measurement algorithm, *lens*fit (Miller et al. 2007; Kitching et al. 2008; Miller et al. 2013), with some improvements introduced for the KiDS-Legacy analysis. The shear calibration results for the updated *lens*fit measurements are presented in Sect. 3.5, and the sensitivity test is conducted in Sect. 3.6. Finally, we conclude in Sect. 3.7.

Throughout the chapter, we define the complex ellipticity of an object as

$$\boldsymbol{\epsilon} \equiv \epsilon_1 + \mathrm{i}\epsilon_2 = \left(\frac{1-q}{1+q}\right) \exp(2\mathrm{i}\phi) \,, \tag{3.1}$$

where q and ϕ denote the axis ratio and the position angle of the major axis, respectively. In terms of the quadrupole moments of the measured surface brightness Q_{ij}, this definition equals

$$\boldsymbol{\epsilon} = \frac{Q_{11} - Q_{22} + 2\mathrm{i}Q_{12}}{Q_{11} + Q_{22} + 2\,(Q_{11}Q_{22} - Q_{12}^2)^{1/2}} \,. \tag{3.2}$$

As stated by Bartelmann & Schneider (2001), this ellipticity definition is convenient because it directly links to the weak lensing shear signal $\boldsymbol{\gamma}$ via the estimator

$$\boldsymbol{\gamma} = \frac{\sum_i w_i \, \boldsymbol{\epsilon}_i}{\sum_i w_i} \,, \tag{3.3}$$

where w_i is a weight assigned per object to account for individual measurement uncertainties[8]. Although the cosmic shear analysis uses higher-order statistical measures, such as the two-point correlation functions (e.g. Kaiser 1992), the simple estimator presented in Eq. (3.3) is commonly used for constraining the shear bias from image simulations (e.g. Heymans et al. 2006).

3.2 Input mock catalogues

To generate mock images, we need input catalogues of galaxies and stars with realistic morphology, photometry and clustering. We detail our procedure for building these catalogues in this section. Section 3.2.1 describes how we create the mock galaxy catalogue by combining deep observations with up-to-date cosmological and galactic

[7] https://github.com/KiDS-WL/MultiBand_ImSim.git

[8] Strictly speaking, the expectation value of the ellipticity is $\boldsymbol{\gamma}/(1-\kappa)$, where κ is the convergence. But as $\kappa \ll 1$ in the weak lensing regime, we can safely neglect this term.

simulations. Section 3.2.2 shows how we generate stellar multi-band magnitude distributions from a population synthesis code.

3.2.1 Galaxies: SURFS-Shark simulations with COSMOS morphology

Our input galaxy catalogue, a compilation of simulations and observations, is crafted to achieve a balance between considerable sample volume and realistic galaxy morphology. In the subsequent parts of this section, we first review the simulation part, which includes the clustering and multi-band photometry. Next, we introduce our specifically designed algorithm, which learns galaxy morphology from observations and applies this knowledge to simulated galaxies — a crucial step towards ensuring accurate shear calibration.

Generating synthetic galaxies from simulations

To jointly calibrate the shear and redshift estimates, we must base the image simulations on wide and deep ($z > 2$) cosmological simulations, where the true redshift is known. In the previous KiDS redshift calibration, van den Busch et al. (2020) used the MICE Grand Challenge (MICE-GC) simulation, an N-body light-cone simulation that covers an octant of the sky (Fosalba et al. 2015a). However, the MICE simulation has a redshift limit of $z{\sim}1.4$, preventing its use for calibrating the high-redshift samples in the KiDS-Legacy analysis (up to $z{\sim}2$). Therefore, we switched to another public N-body simulation from the Synthetic UniveRses For Surveys (SURFS, Elahi et al. 2018).

The SURFS simulation we adopted has a box size of 210 h^{-1}cMpc (cMpc stands for comoving megaparsec), containing 1536^3 particles with a mass of $2.21 \times 10^8\ h^{-1}\mathrm{M}_\odot$, and a softening length of 4.5 h^{-1}ckpc (ckpc stands for comoving kiloparsec). It assumes a ΛCDM cosmology with parameters from Planck Collaboration (2016). The final halo catalogues and merger trees are constructed from 200 snapshots starting at redshift $z = 24$, using the phase-space halo-finder code VELOCIraptor (Cañas et al. 2019; Elahi et al. 2019a) and the halo tree-builder code treefrog (Elahi et al. 2019b). We refer to Lagos et al. (2018) for details on the building and Poulton et al. (2018) for validating the halo catalogues and merger trees.

The galaxy properties, including the star formation history and the metallicity history, are from an open-source semi-analytic model named Shark[9] (Lagos et al. 2018). The model parameters are tuned to reproduce the $z = 0, 1$ and 2 stellar-mass functions (Wright et al. 2018), the $z = 0$ black hole-bulge mass relation (McConnell & Ma 2013) and the mass-size relations at $z = 0$ (Lange et al. 2016). Any other observables are predictions of the model, which also match well with observations (see Lagos et al. 2018 for more details). As for weak lensing calibration, the most crucial property is the redshift evolution of the galaxy number density (e.g. Hoekstra et al. 2017), which we checked in detail in Appendix 3.A and found it to be sufficient for KiDS.

The light cones from the Shark outputs are created using the code stingray

[9] https://github.com/ICRAR/shark

(Chauhan et al. 2019), an improved version of the code used by Obreschkow et al. (2009). It first tiles the simulation boxes together to build a complex 3D field along the line of sight, then draws galaxy properties from the closest available time-step, resulting in spherical shells of identical redshifts. A possible issue would be the same galaxy appearing once in every box but with different intrinsic properties due to cosmic evolution. To avoid this problem, STINGRAY randomises galaxy positions by applying a series of operations consisting of 90 deg rotations, inversions, and continuous translations. We refer to Chauhan et al. (2019) for more details about the light-cone construction.

The final mock-observable sky covers ∼108 deg^2 with minimum repetition of the large-scale structure. The sample variance bias caused by the replicating structure is negligible for our direct shear and photometric redshift calibration. Since we learn galaxy morphology from deep observations, our input galaxy sample is still limited mainly by the observational data we have, which only covers ∼1 deg^2 (see Sect. 3.2.1 for details). We test the robustness of our calibration results against this sample variance bias using the sensitivity analysis detailed in Sect. 3.6.

The multi-band photometry is drawn from a stellar population synthesis technique implemented in the PROSPECT[10] and VIPERFISH[11] packages. PROSPECT (Robotham et al. 2020) is a high-level package combining the commonly used stellar synthesis libraries with physically motivated dust attenuation and re-emission models; while VIPERFISH is a light wrapper to aid the interface with the SHARK outputs. We refer to Lagos et al. (2019) for detailed predictions, validations and a demonstration that the predicted results agree with observations in a broad range of bands from the far-ultraviolet to far-infrared, without any fine-tuning with observations.

For our purpose, we care most about the nine-band photometry covered by the KiDS+VIKING data, so we compared the synthetic near-infrared and optical magnitude distributions to observations from the COSMOS2015 catalogue (Laigle et al. 2016). Figure 3.1 shows the magnitude distributions of eight filters available in both SHARK and COSMOS2015 catalogues, together with an analytical fitting result from Eq. (4) of FC17. The counts in the original simulations are ∼35% lower than the observations with some variation between filters. As this affects the blending level and then the shear bias (Hoekstra et al. 2015, 2017), we calibrated the original synthetic photometry for a better agreement. The technical details are presented in Appendix 3.A. In short, we found that the differences in the magnitude distributions stem from the difference in stellar mass-to-light ratio between the simulations and observations. Therefore, we scaled the original SHARK magnitudes using a modification factor derived from the stellar mass-to-light ratio difference. The modification is the same for all bands, preserving the intrinsic colours of individual galaxies. The modified magnitudes now agree with the observations within ∼3%.

We later noticed that Bravo et al. (2020) proposed a similar fine-tuning method when working with the panchromatic Galaxy And Mass Assembly (GAMA) survey. They used an abundance matching method by comparing the number counts between SHARK

[10] https://github.com/asgr/ProSpect
[11] https://github.com/asgr/Viperfish

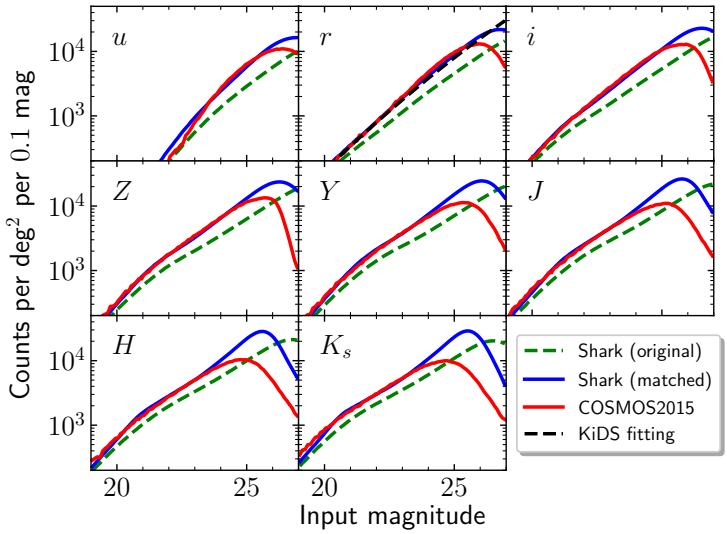

Figure 3.1: Number of galaxies per square degree per 0.1 magnitude in the input apparent magnitudes. The green dashed lines represent data from the original SURFS-SHARK mock catalogue, while the blue solid lines indicate the modified results. The red solid lines correspond to the COSMOS2015 observations, factoring in flags for the UltraVISTA area within the COSMOS field after the removal of saturated objects and problematic areas (yielding a 1.38 square degree effective area, as per Table 7 in Laigle et al. 2016). The black dahsed linne depicts the analytical fitting result in the r-band, derived from FC17. The g-band photometry, absent from the COSMOS2015 catalogue, is not displayed here. We note that the COSMOS2015 catalogue lacks completeness at $K_s \gtrsim 24.5$ (Laigle et al. 2016).

and GAMA after fine binning in redshift and r-band apparent magnitude. They tuned magnitudes for all SHARK galaxies with $r<21.3$ to match the number counts in GAMA. Their modifications are consistent with our results, albeit targeting different magnitude ranges.

Learning galaxy morphology from observations

Simulating galaxies with realistic morphology is essential for accurate shear calibration. Following K19, we represent the galaxy morphology using the Sérsic profile (Sérsic 1963) with three parameters: the effective radius determining the galaxy size (also known as the half-light radius), the Sérsic index describing the concentration of the brightness distribution, and the axis ratio determining the galaxy ellipticity. We learned these structural parameters from deep observations accounting for their mutual correlations and their correlations to galaxy photometry and redshift. Figure 3.2 shows the workflow for the learning algorithm.

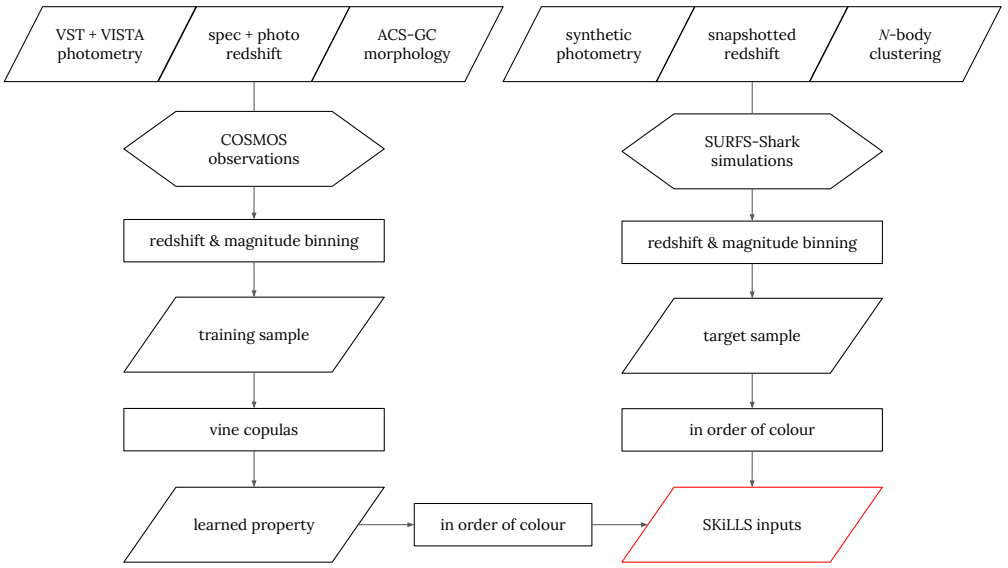

Figure 3.2: Flowchart summarising the algorithm employed in constructing the SKiLLS input mock catalogue. The SKiLLS galaxies inherit the synthetic multi-band photometry and 3D positions from the SURFS-Shark simulations, while their morphology is learned from the observations in the COSMOS field using a vine-copula modelling-based algorithm (see the detailed explanation in Sect. 3.2.1)

3

We start with a 'reference' sample comprising morphology, photometry and redshifts from several deep observations. The structural parameters are adopted from the catalogue produced by Griffith et al. (2012), who fitted Sérsic models to the galaxy images taken by the Advanced Camera for Surveys (ACS) instrument on the *Hubble* Space Telescope (HST). We used their results derived from the COSMOS survey and cleaned the sample by only preserving objects with a good fit (FLAG_GALFIT_HI = 0) and reasonable size (half-light radius between 0.″01 and 10″) to avoid contamination. We note that this catalogue was also used by K19 and proved to be sufficient for KiDS-like simulations.

The r-band photometry is derived from a deep VST-COSMOS catalogue using 24 separate VST observations of the COSMOS field taken from KiDS and the SUpernova Diversity And Rate Evolution (SUDARE) survey (Cappellaro et al. 2015; De Cicco et al. 2019). These observations have a maximum seeing of 0.″82, close to the KiDS r-band image qualities. To ensure consistent measurements, we conducted the stacking and detection processes using the same pipeline as the standard KiDS data processing. The stacked image has an average seeing of 0.″75 and a total exposure time of 42 120 seconds, which is a factor of ∼23 over a standard KiDS observation. The limiting magnitude of the final deep catalogue is more than one magnitude deeper than usual KiDS catalogues. To include colour information, we also used the K_s-band photometry from the COSMOS2015 catalogue (Laigle et al. 2016), as it originates from the UltraVISTA project (McCracken et al. 2012) that shares the same instruments with the VIKING

near-infrared observations.

The redshifts are taken from the catalogue compiled by van den Busch et al. (2022). It contains observations from several spectroscopic and high-quality photometric surveys in the COSMOS field. The spectroscopic redshifts were collected from G10-COSMOS (Davies et al. 2015), DEIMOS (Hasinger et al. 2018), hCOSMOS (Damjanov et al. 2018), VVDS (Le Fèvre et al. 2013), LEGA-C (van der Wel et al. 2016), FMOS-COSMOS (Silverman et al. 2015), VUDS (Le Fèvre et al. 2015), C3R2 (Masters et al. 2017, 2019; Euclid Collaboration et al. 2020; Stanford et al. 2021), DEVILS (Davies et al. 2018) and zCOSMOS (*private comm. from M. Salvato*), while the photometric redshifts were from the PAU survey (Alarcon et al. 2021) and COSMOS2015 (Laigle et al. 2016). For sources with multiple measurements, a specific 'hierarchy' was defined with orders based on the quality of measured redshifts to choose the most reliable redshift estimates (see Appendix A in van den Busch et al. 2022, for details). Given the high quality of the redshift estimates, we treated them as true redshifts.

All catalogues mentioned above overlap in the COSMOS field, so we can combine them by cross-matching objects based on their sky positions. The final reference catalogue has 75 403 galaxies with all the necessary information. It has a limiting magnitude of 27 in the r-band but suffers incompleteness after $m_{\rm r} \gtrsim 24.5$. We verified that the incompleteness at the faint end does not bias the overall morphological distribution by comparing it to measurements from the *Hubble* Ultra Deep Field observations (Coe et al. 2006).

We aim to inherit not only the individual distributions of structural parameters but also their mutual dependence and possible correlations with redshifts and magnitudes. To achieve this goal, we developed a learning algorithm based on a statistical inference technique, dubbed vine-copulas (e.g. Joe 2014; Czado 2019). A brief introduction to the technique is presented in Appendix 3.B. In short, a copula-based method models joint multi-dimensional distributions by separating the dependence between variables from the marginal distributions. It is popular in studies concerning dependence modelling, given its flexibility and reliability. In practice, we first divided galaxies into 30×40 bins based on their redshifts and r-band magnitudes. Each bin contains a similar number of reference galaxies. Then in each bin, we built a data-driven vine-copula model from the measured $r - K_s$ colour and morphological parameters using the public `pyvinecopulib` package[12]. The learned vine-copula model can be sampled to produce an arbitrary number of vectors of parameters from the constrained multi-dimensional distributions. We decided to generate the same number of vectors as the available SHARK galaxies and assign them to the SHARK galaxies in the order of $r - K_s$ colour. This approach allows us to mimic observations from the underlying distributions rather than repeatedly sampling from the measured values.

Figure 3.3 shows the correlations between the magnitude and the two critical structural parameters: half-light radius and ellipticity, in several redshift bins. We see that the learned sample follows the average trends of the reference sample. Figure 3.4 presents two-dimensional contour plots in several magnitude bins to better inspect the

[12] `https://github.com/vinecopulib/pyvinecopulib`

underlying distributions of morphological parameters. We again see agreements in correlations between the size and ellipticity and between the size and concentration, proving that our copula-based algorithm captures the multi-dimensional dependence from the reference sample.

3.2.2 Stars: Point objects with synthetic photometry

We treated stars as perfect point objects. Their multi-band photometry was obtained from the population synthesis code, TRILEGAL (Girardi et al. 2005, with version 1.6 and the default model from its website[13]). We generated six stellar catalogues at galactic coordinates evenly spaced across the KiDS footprint to capture the variation of stellar densities between KiDS tiles. Each catalogue spans 10 deg^2. When simulating a specific tile image covering 1 deg^2, we selected the stellar catalogue whose central pointing is closest to the target tile, then randomly drew ten per cent of stars from that catalogue as the input. Figure 3.5 shows the r-band magnitude distributions of the six stellar catalogues compared to the catalogue used by the COllege simulations. The broader coverage of stellar densities is noticeable, marking one of the improvements in SKiLLS. Also, stars in SKiLLS have nine-band magnitudes consistently predicted from a library of stellar spectra (see Girardi et al. 2005, for details), while in COllege, stars only have r-band magnitudes.

3.3 KiDS+VIKING nine-band image simulations

This section details the creation and processing of the multi-band mock images. We start with the creation of KiDS-like optical images (Sect. 3.3.1) and VIKING-like infrared images (Sect. 3.3.2), then summarise the SKiLLS fiducial setups in Sect. 3.3.3. We end the section with the measurement of colours and photometric redshifts (Sect. 3.3.4).

3.3.1 KiDS-like optical images

Each KiDS pointing consists of four-band optical images taken with the OmegaCAM camera at the VLT Survey Telescope (Kuijken 2011): u, g, r and i. The r-band images are the primary products used for the shear measurement, while the remaining bands are only for photometric measurements. The science array of the OmegaCAM camera has a $\sim 1^\circ \times 1^\circ$ field of view covered by 8×4 CCD images, each of size 2048×4100 pixels with an average resolution of $0\farcs214$. Although the CCDs are mounted as closely as possible, a narrow gap between the neighbouring CCDs is technically inevitable. The average gap sizes between the pixels of neighbouring CCDs are:

- between the long sides of the CCDs: 1.5 mm (100 pixels)
- central gap along the short sides: 0.82 mm (55 pixels)

[13] http://stev.oapd.inaf.it/cgi-bin/trilegal

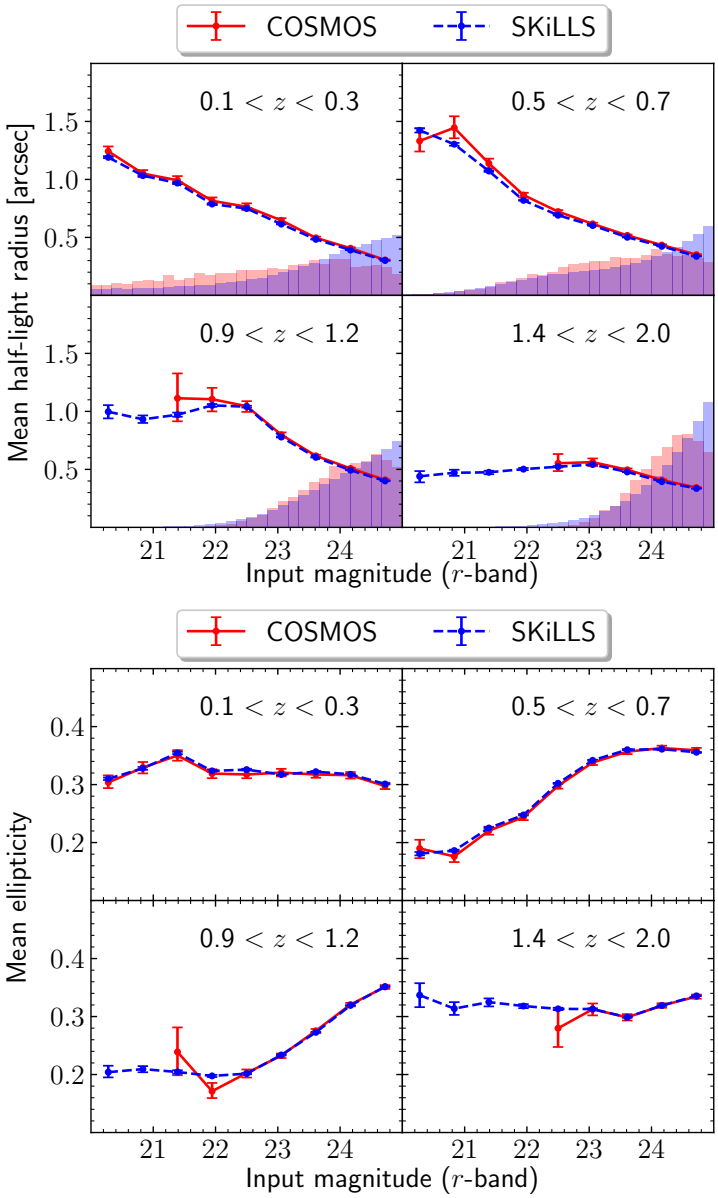

Figure 3.3: Comparison of the overall magnitude-morphology relationships across several redshift bins. The training and target samples are represented by red solid and blue dashed lines, respectively. The top panel illustrates the average half-light radius as a function of r-band magnitude, while the bottom panel depicts the average ellipticity as a function of r-band magnitude. Statistical uncertainties displayed are calculated from 500 bootstrap iterations. The top panel also includes histograms of normalised magnitude distributions, indicating that the bright galaxies at higher redshifts from the simulation contribute minimally to the overall population.

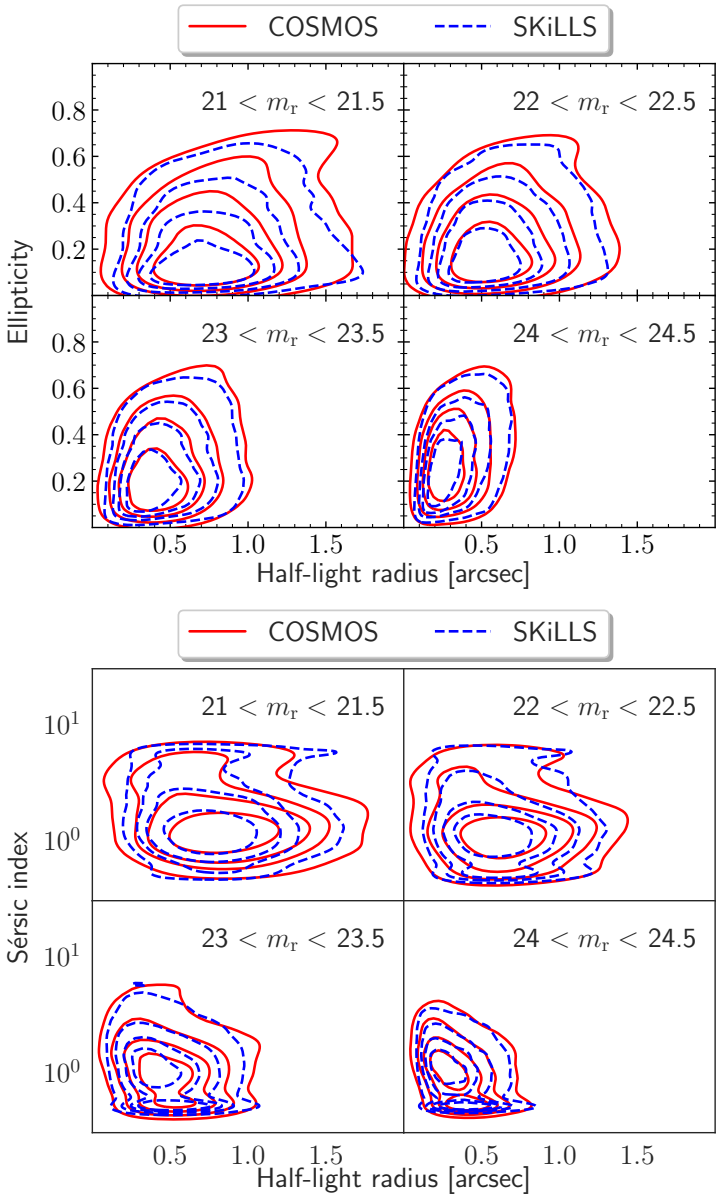

Figure 3.4: Two-dimensional kernel density plots of morphological parameters within distinct magnitude bins. The red solid lines represent the training samples, while the blue dashed lines correspond to the target samples. In the top panel, the correlation between galaxy size and ellipticity is displayed, while the bottom panel elucidates the correlation between galaxy size and Sérsic index. The depicted contour levels correspond to 20%, 40%, 60%, 80%.

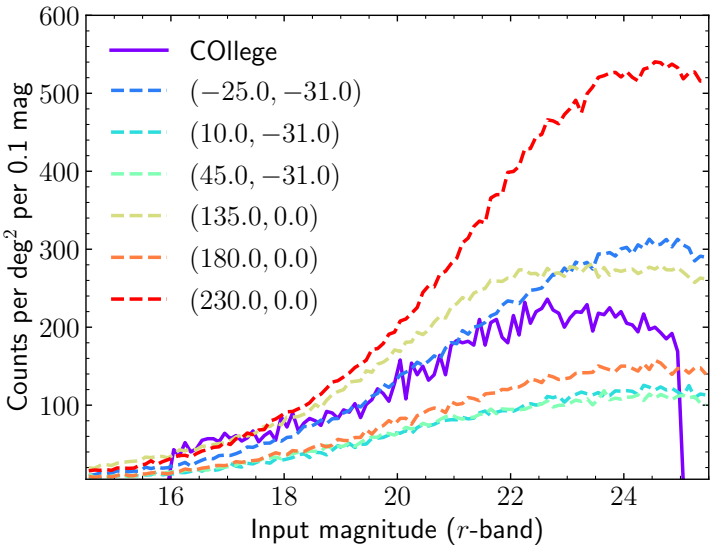

Figure 3.5: The r-band input magnitude distributions for the six stellar catalogues utilised by SKiLLS. Labels represent the pointing centres (RA, Dec), with the exception of 'COllege', which signifies the stellar catalogue employed by K19.

- wide gap along short sides: 5.64 mm (376 pixels)

To avoid 'dead zones' caused by these gaps, each tile image incorporates multiple dithered exposures (five in the g, r and i bands, four in the u band). The dithers form a staircase pattern with steps of 25″ in RA and 85″ in declination to match the gaps between CCDs (de Jong et al. 2013).

KiDS raw observations are processed with two independent pipelines: the Astro-WISE pipeline designed for the photometric measurements (McFarland et al. 2013; de Jong et al. 2015)[14], and the THELI pipeline optimised for the shape measurements (Erben et al. 2005; Schirmer 2013; Kuijken et al. 2015)[15]. While the former is applied to all four-band observations, the latter is only used for the r-band observations, as KiDS only measures galaxy shapes for lensing in the r-band images. The main difference between the Astro-WISE and THELI pipelines is in the co-addition process, where the former resamples all exposures to the same pixel grid with a uniform 0.″20 pixel size, while the latter preserves the original pixels to maintain image fidelity as much as possible.

We kept all these features in mind when generating SKiLLS optical images. We created raw exposures using the GalSim pipeline[16] (Rowe et al. 2015), with galaxies and stars from the mock catalogues described in Sect. 3.2. The underlying canvas mimicked the science array of the OmegaCAM camera, including pixels and gaps.

[14] http://www.astro-wise.org/
[15] https://www.astro.uni-bonn.de/theli/
[16] https://github.com/GalSim-developers/GalSim

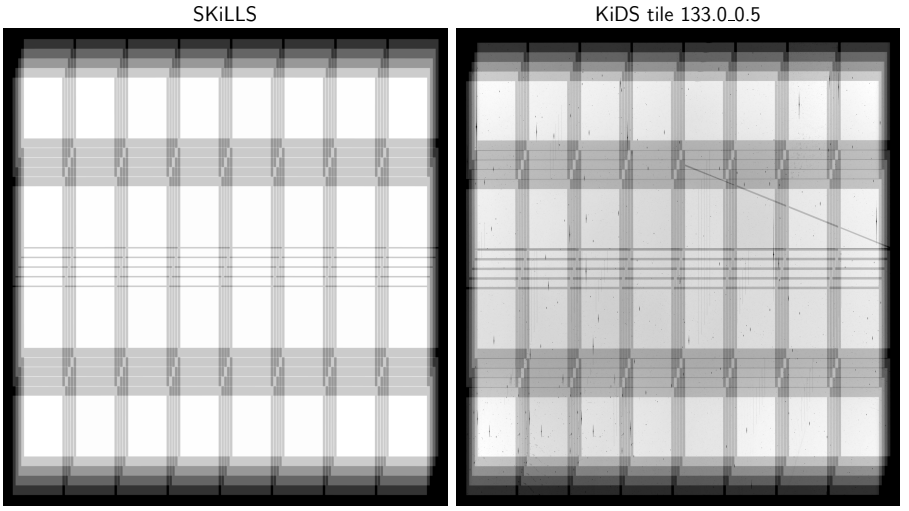

Figure 3.6: Comparison of the THELI weight image generated by SKiLLS (shown in the left panel) with a randomly chosen example from KiDS (depicted in the right panel). The 8×4 CCDs span a sky area of approximately one square degree. The shallow regions are attributed to gaps in individual exposures. A similar level of agreement is also observed in the ASTRO-WISE co-added images.

Galaxies and stars were mapped to the canvas using the gnomonic (TAN) projection of their original sky coordinates. Following the KiDS image processing, we stacked exposures using the SWARP software (Bertin 2010), with the identical setups as in the KiDS pipelines, including ASTRO-WISE-like images re-gridded to a uniform $0\farcs20$ pixel size and THELI-like images preserving the original $0\farcs214$ pixel size. Figure 3.6 compares a co-added THELI weight image from SKiLLS to a randomly selected tile from KiDS. It shows that the SKiLLS images contain the main features of KiDS images, including the gaps and dither patterns, albeit lacking subtle features, such as the inhomogeneous backgrounds between CCDs and masks of satellites.

Besides the image layout, we need information on the pixel noise and point spread function (PSF) to mimic observational conditions. We extracted this information from the fourth public data release of KiDS (KiDS-DR4, or DR4 for short, Kuijken et al. 2019). It has a total of 1006 square-degree survey tiles with stacked $ugri$ images along with their weight maps, masks and source catalogues. We selected a representative sample of 108 tiles and replicated their properties in our image simulations (see Sec. 3.3.3 for details). For the raw pixel noise, we adopted Gaussian distributions with variances estimated from the ASTRO-WISE weight maps corrected with a boost factor of $\sim$1.145 $\left[= (0.214/0.2)^2\right]$ to account for the re-gridding effect. For the PSF, we used two approaches, depending on the different usages of the images.

For the r-band images from which galaxy shapes are measured, we used the position-dependent PSF models for individual exposures. These PSF models, constructed from

well-identified stars, are in the form of two-dimensional polynomial functions and can recover a PSF image in the pixel grid for any given image position (see Miller et al. 2013; Kuijken et al. 2015; Giblin et al. 2021 for details). In practice, we recovered 32 PSF images for each exposure using the centre positions of the CCD images. The recovered PSF images contain modelling uncertainties, which can introduce artificial spikes when being used to simulate bright stars. Therefore, we applied a cosine-tapered window to the original PSF image to suppress the modelling noise at its outskirts. The two edges of the window function are defined at 5 and 10 times the full-width half-maximum (FWHM) of the target PSF to preserve features in the central region as much as possible. With these recovered PSF images, we can treat the 32 CCD images separately using their own PSFs, a significant improvement from the constant PSF used in previous work. The recovered PSF image is also superior to a Moffat profile as it captures more delicate features of complex PSFs, such as ellipticity gradients.

For other optical bands where only photometry is measured, we still adopted the Moffat profile, given that the photometric measurement is insensitive to the detailed profile of PSF. We estimated the Moffat parameters by modelling bright stars identified in the Astro-WISE images. Since the photometry is measured from the stacked images and is less sensitive to the gentle PSF variation within a given tile, we kept the PSF model invariant for all exposures for simplicity. To alleviate the Moffat fitting bias introduced by the pixelisation of CCD images, we applied the first-order correction to the measured Moffat parameters using image simulations. Specifically, we simulated the pixelated PSF image using measured Moffat parameters and then remeasured them with the same fitting code. The difference between the remeasured and input values is the correction factor and is subtracted from the initially measured value. Our test shows that this correction can suppress the original percent-level bias down to a sub-percent level, which is sufficient for our photometry-related purpose.

3.3.2 VIKING-like infrared images

To improve the accuracy of photometric redshifts, KiDS includes near-infrared (NIR) measurements from the VISTA Kilo-degree Infrared Galaxy (VIKING) survey (Edge et al. 2013). The two surveys share an almost identical footprint. We refer to Wright et al. (2019) for details of the VIKING imaging and its usage in KiDS. Briefly, the VIKING data have three levels of products: exposures, paw-prints, and tiles. Given the complex NIR backgrounds, the VIKING survey first takes multiple exposures in quick succession with small jitter steps for reliable estimation of the noisy background. These exposures are then stacked together to create the second level of product: the 'paw-print'. A paw-print still contains gaps between individual detectors, so six paw-prints with a dither pattern are used to produce a contiguous tile image. However, these co-added tiles have non-contiguous PSF patterns caused by the large dithers between successive paw-prints. Therefore, in the KiDS+VIKING analyses, photometry is done on individual paw-prints instead of the co-added tiles. The dither pattern of paw-prints causes multiple flux measurements per source (typically four in the case of the J-band and two in the other

bands). The final flux estimate for each source is a weighted average of the individual measurements with the weights derived from individual flux errors.

Given the complexity of the VIKING observing strategy, we simplified the NIR-band observations in SKiLLS with single images per square degree of KiDS tile. To compensate for the simplified images, we considered the overlap between individual paw-prints when estimating the observational conditions. As we show in Sect. 3.3.4, this simplified approach can still achieve realistic photometry, which is the only important quality we seek from the NIR-band images.

Specifically, we created a 'flat-field image' for each paw-print with the same size and pixel scale. Its pixel value equals the absolute standard deviation of the background pixel values on the corresponding paw-print. For each KiDS pointing, we selected all VIKING paw-prints that overlap in the given one square-degree sky area and stacked their flat-field images with shifts accounting for the different sky pointings of the paw-prints. We took the median pixel value of the co-added flat-field image as the final pixel noise of the corresponding KiDS pointing. In doing so, we captured various overlapping VIKING paw-prints in individual KiDS pointings. Following the typical situations of the KiDS+VIKING data (Wright et al. 2019), we only preserved KiDS pointings with at least two paw-prints in the $ZYHK_s$-bands and at least four paw-prints in the J-band. This requirement reduced the number of pointings from 1006 to 979, which is still plentiful for our purpose. As for the PSF, we employed a constant Moffat profile for each KiDS pointing. The PSF FWHM is a weighted average from overlapping VIKING paw-prints with the weights determined by their noise levels. In order to determine the Moffat concentration index for a given FWHM value, we fitted Moffat profiles to bright stars in some representative paw-prints. The Moffat fitting bias introduced by the pixelisation is corrected using the same method introduced in Sect. 3.3.1. We found the relationship between the Moffat index n and FWHM (arcsec) in VIKING images to be: $\ln(n) = 66.56\ \exp(-6.36\ \mathrm{FWHM}) + 0.90$. This empirical formula is used to pair each FWHM with a unique Moffat index.

3.3.3 SKiLLS fiducial setup

Since we have 108 deg^2 of Shark galaxies as described in Sect. 3.2.1, we selected 108 KiDS pointings for the SKiLLS fiducial run. Figure 3.7 shows the sky locations of the selected 108 tiles along with the 979 KiDS-DR4 tiles that have the nine-band noise and PSF information. Clusters of the selected blocks pair with the six stellar catalogues generated from Trilegal so that SKiLLS captures the stellar density variation across the whole KiDS survey (see Sect. 3.2.2).

Figure 3.8 compares the r-band noise and PSF properties between the SKiLLS selected tiles and all usable KiDS-DR4 tiles. We measured the PSF size and ellipticity using the weighted quadrupole moments with a circular Gaussian window of dispersion 2.5 pixels, the typical galaxy size in the KiDS sample. The PSF size is defined as

$$r_{\mathrm{PSF}} \equiv (Q_{11}Q_{22} - Q_{12}^2)^{1/4}\,, \tag{3.4}$$

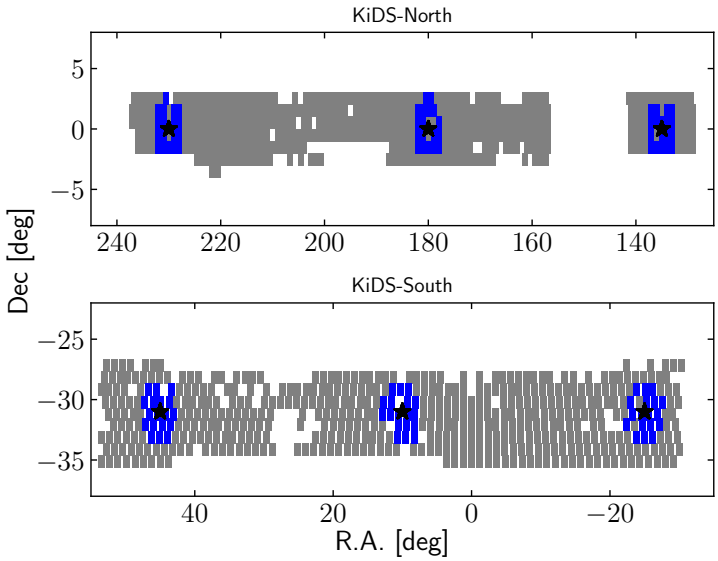

Figure 3.7: Sky distribution of the KiDS-DR4 tiles. The tiles marked in blue are incorporated in the SKILLS fiducial run, accounting for a total of 108 tiles. The grey blocks represent all KiDS-DR4 tiles, amounting to 979, for which nine-band noise and PSF information are available. The black stars signal the centres of the stellar catalogues generated using TRILEGAL (Girardi et al. 2005).

where Q_{ij} are the weighted quadrupole moments, and the PSF ellipticity is defined by Eq. (3.2). Figure 3.8 shows that the selected tiles represent the KiDS-DR4 data well. Because we vary PSF for individual CCD images and exposures, the 108 SKiLLS images cover 17 280 different PSF models, a significant extension of the 65 PSF models used by FC17 and K19. That also explains the smooth distributions of the PSF parameters. Figure 3.9 shows similar comparisons for other bands. Again we see fair agreements across all bands. As KiDS-DR4 already covers ∼75% of the whole survey, we expect a similar agreement to the KiDS-Legacy data. The wide coverage of the noise and PSF properties also makes the SKiLLS results more robust than previous simulations and simplifies sensitivity tests (see Sect. 3.6 for details).

3.3.4 Photometry and photometric redshifts

With the simulated multi-band images, we can measure colours and estimate photometric redshifts (photo-zs) for simulated galaxies using the same tools developed in KiDS with minor adjustments.

For galaxy colours, we used the GAAP (Gaussian Aperture and PSF) pipeline (Kuijken et al. 2015, 2019). It provides accurate multi-band colours by accounting for PSF differences between filters and optimises signal-to-noise ratio (S/N) by down-

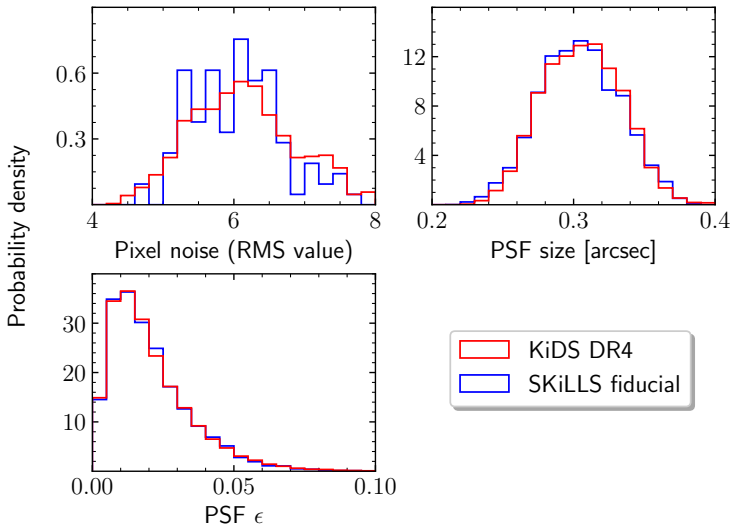

Figure 3.8: Comparison of normalised histograms for pixel noise (top left), PSF size (top right), and PSF ellipticity (bottom left) between KiDS-DR4 (depicted in red) and SKiLLS (in blue) within the r-band images. Both PSF size and ellipticity are measured from the recovered PSF image employing a circular Gaussian window of a sigma value of 2.5 pixels.

weighting the noise-dominated outskirts. The latter is possible because the photo-z estimation only needs the ratio of the fluxes from the same part of a galaxy in the given bands rather than the total light. A prerequisite for the GAAP pipeline is a detection catalogue with source positions and aperture parameters, which we measured from the THELI-like r-band images using the SEXTRACTOR code (Bertin & Arnouts 1996). Once the detection catalogue is ready, we can obtain the list-driven photometry by running the GAAP algorithm on the u, g, r and i ASTRO-WISE-like images and the Z, Y, J, H and K_s simple images. In short, the GAAP method includes three major steps:

1. Homogenising PSFs by convolving the whole image with a spatially variable kernel map modelled from high S/N stars. The resulting image has a simple Gaussian PSF, for which estimating the PSF-independent Gaussian aperture flux is possible. The main side effect is that the convolution process introduces correlated noise between neighbouring pixels, complicating the estimation of measurement uncertainties. GAAP handles this by tracking the noise covariance matrix through the whole process.

2. Defining an elliptical Gaussian aperture function for each source using the size and shape parameters measured by SEXTRACTOR on the r-band detection images. In practice, users must customise the minimum and maximum GAAP aperture sizes

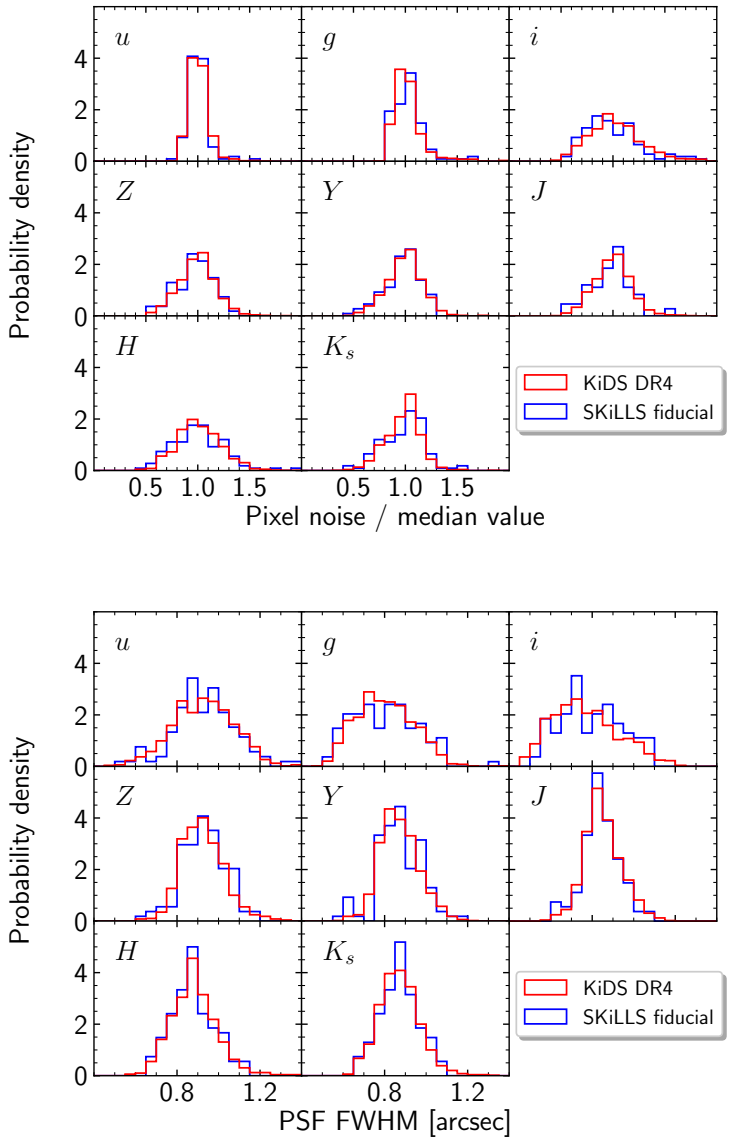

Figure 3.9: Comparison of normalised histograms for pixel noise (top) and PSF FWHM (bottom) across bands used solely for photometry, contrasting KiDS-DR4 (red) with SKiLLS (blue). The equivalent comparisons for the lensing r-band images are illustrated in Fig. 3.8. To facilitate comparison in the same range, pixel noise values are normalised by their median values in each band.

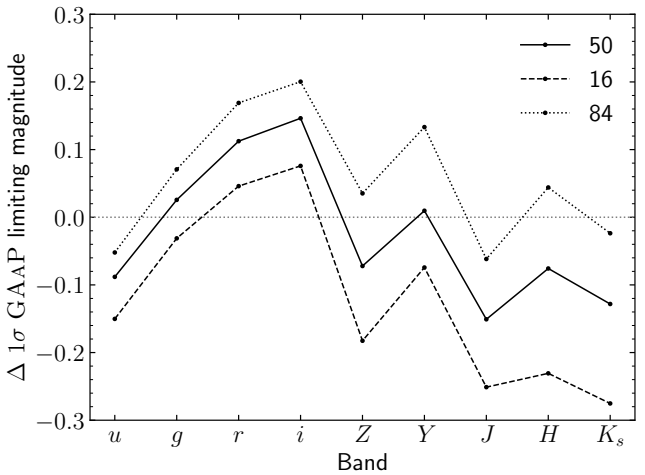

Figure 3.10: Differences in the median 1σ GAAP limiting magnitudes across the nine bands (simulation - data). The three lines represent the 16, 50 and 84 percentiles derived from the 108 tiles incorporated in the SKiLLS fiducial run. The larger scatters in the NIR bands can be partially attributed to the adoption of a simplified simulation strategy.

to balance the S/N and the effect of blending. Following the KiDS fiducial setup, we set the maximum aperture to $2''$ to avoid contamination from neighbouring sources. We conducted two separate runs by setting the minimum aperture to $0''.7$ and $1''.0$. When used as the input for the photo-z estimation, a source-by-source decision was made to optimise the flux errors across the nine bands (see Kuijken et al. 2019 for details).

3. Performing the aperture photometry on the PSF-Gaussianised images for each band using the defined aperture functions. It is worth stressing that GAAP aims to provide robust colours for the high S/N parts of galaxies; it underestimates the total fluxes for extended sources by design.

Figure 3.10 compares the nine-band 1σ GAAP limiting magnitudes between the KiDS-DR4 data and SKiLLS fiducial results. We calculated the median limiting magnitudes for tiles in both KiDS and SKiLLS and then compared their differences. We see a general agreement for all the bands, verifying our noise and PSF modelling. Noticeably, even for the NIR bands where we simplified the VIKING observations with single images, the differences are still tolerable, albeit with larger uncertainties. Figure 3.11 compares the GAAP photometric distributions between the simulation and data. Once again, we see a decent agreement in both magnitude and colour distributions.

For the photo-z estimation, we implemented the public Bayesian Photometric Redshift (BPZ, Benítez 2000) code with the re-calibrated template set from Capak (2004) and the Bayesian redshift prior from Raichoor et al. (2014). We closely followed the

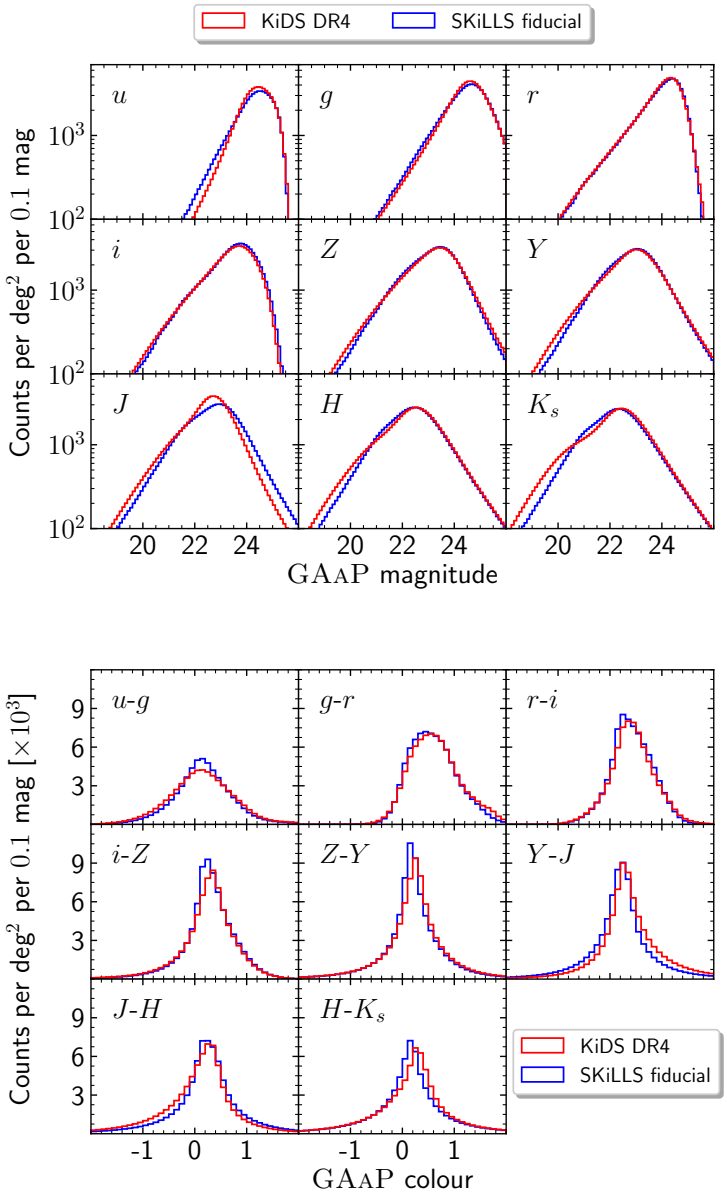

Figure 3.11: Comparison of the GAAP magnitudes (top panel) and colours (bottom panel) for KiDS-DR4 (red) and SKiLLS (blue). The displayed results incorporate all galaxies possessing valid photometric measurements (with the GAAP flags equating to zero across nine bands). No shape-measurement-related filters have been applied at this stage.

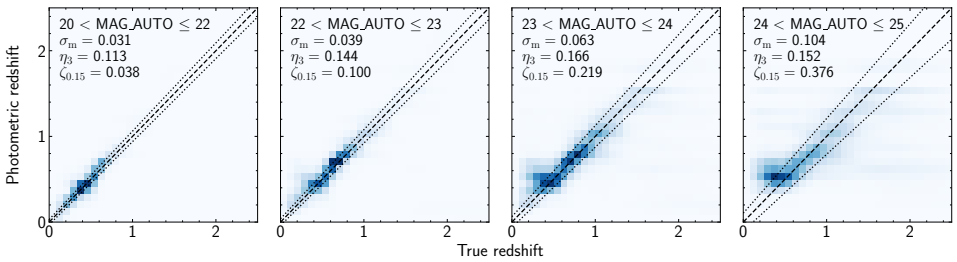

Figure 3.12: Photometric versus true redshifts distributed across several apparent r-band magnitude bins. The annotated statistics are: the normalised median-absolute-deviation (σ_m) of the variable $\Delta z/(1+z)$, the proportion of sources with $|\Delta z/(1+z)| > 3\sigma_m$ (η_3) and the proportion of sources with $|\Delta z/(1+z)| > 0.15$ ($\zeta_{0.15}$). The dashed lines represent the one-to-one correlation, while the dotted lines indicate $|\Delta z/(1+z)| = \sigma_m$.

settings in the KiDS-DR4 analysis (Kuijken et al. 2019) unless it conflicts with the simulation input. For example, we set ZMAX to 2.5, the limiting redshift of SKiLLS galaxies, instead of 7.0 as in the data. We tested the choice of ZMAX in the simulations and found that only 0.1% of the test sample resulted in estimates differing more than 0.1, which means most of the objects have similar photo-z estimates and end up in the same tomographic bins for these two choices. Moreover, the SHARK photometry in the u, g, r, i and Z bands is based on the Sloan Digital Sky Survey (SDSS) photometric system, which is slightly different from the KiDS/VIKING system (Kuijken et al. 2019). We corrected these slight differences in the measured GAAP magnitudes in order to use the KiDS/VIKING filters to run the BPZ code. The detailed procedures and comparisons are described in Appendix 3.C. Overall, the modification is minor and has a negligible impact on the magnitude, colour distributions, and final shear biases. Still, it improves the agreement between the simulation and the data in the photo-z distributions. Unless specified otherwise, we base our fiducial results on the transformed photometry.

Figure 3.12 compares the estimated photo-z to the true redshift from the input SURFS-SHARK simulations in several measured magnitude bins. It shows the photo-z vs. true redshift distributions, along with annotated statistics based on the distributions of $(z_B - z_{true})/(1+z_{true}) \equiv \Delta z/(1+z)$ values. We see the BPZ code works well in SKiLLS and is at the same level as in KiDS (Wright et al. 2019). More detailed verification of the SKiLLS photo-z performance is presented in the companion redshift calibration paper (van den Busch et al., in prep).

As for the redshift calibration, our end-to-end approach, which starts with image simulation followed by object detection, PSF homogenisation, forced multi-band photometry, and photo-z estimation, is a significant improvement compared to previous catalogue-level simulations (e.g. Hoyle et al. 2018; van den Busch et al. 2020; DeRose et al. 2022). The image-simulation-based approach not only yields more realistic observational uncertainties but also naturally accounts for the blending effect, which

is hard to address at the catalogue level. As for the shear calibration, these photo-z estimates are essential for performing tomographic selections (K19). Our approach that directly measures the photo-zs from simulated images accounts for various measurement uncertainties of photo-zs, hence a tomographic selection consistent with how it is done in the data. Moreover, using the same mock catalogue in both shear and redshift calibration unites these two long-separated processes in the KiDS-Legacy analysis.

3.4 Shape measurements with the updated *lens*fit

The primary task of any weak lensing survey is to measure the shapes of galaxy images. Previous KiDS analyses tackled this task using a likelihood-based code, dubbed *lens*fit (Miller et al. 2007; Kitching et al. 2008; Miller et al. 2013). It is the default shape measurement algorithm for the KiDS-Legacy analysis, with some updates described in this section. We test SKiLLS using this updated *lens*fit code[17].

3.4.1 The self-calibration version of *lens*fit

The *lens*fit code, first developed for CFHTLenS (Heymans et al. 2012), follows a Bayesian model-fitting approach. We refer to Miller et al. (2013) for its detailed formalism. In brief, it first performs a joint fit to individual exposures using a PSF-convolved galaxy model, which yields a likelihood distribution of seven parameters: 2D position, flux, scalelength, bulge-to-total flux ratio and complex ellipticity. Then it deduces the ellipticity parameters from the likelihood-weighted mean values by marginalising other parameters with priors as described by Miller et al. (2013). For each ellipticity estimate, an inverse-variance weight is also determined from (Miller et al. 2013)

$$w_i \equiv \left[\frac{\sigma^2_{\epsilon,\,i}\,\epsilon^2_{\rm max}}{2\epsilon^2_{\rm max} - 4\sigma^2_{\epsilon,\,i}} + \sigma^2_{\epsilon,\,\rm pop} \right]^{-1} , \tag{3.5}$$

where $\sigma_{\epsilon,\,i}$ is the uncertainty of the measured ellipticity, $\sigma_{\epsilon,\,\rm pop}$ is the ellipticity dispersion of the galaxy population (intrinsic shape noise), and $\epsilon_{\rm max}$ is the maximum allowed ellipticity in the *lens*fit model-fitting. As for KiDS data, we adopted $\sigma_{\epsilon,\,\rm pop} = 0.253$ and $\epsilon_{\rm max} = 0.804$.

The code has evolved as KiDS progressed. The most significant is a self-calibration scheme for noise bias, as detailed in FC17. The pixel noise in a given image skews the likelihood, which biases the estimate of individual galaxy ellipticities. It is a complex function of the signal-to-noise ratio, galaxy properties and PSF morphology, making it difficult to predict accurately. Thus, *lens*fit conducts an approximate correction using the measurements themselves, that is a self-calibration. The basic idea is to simulate a test galaxy with parameters measured from the first run, then remeasure the test galaxy using the same pipeline. The difference between the remeasured and input values serves as a

[17]Nevertheless, we note that SKiLLS can also calibrate other algorithms, such as the KiDS Metacalibration catalogue (Yoon et al., in prep.).

correction factor for the corresponding parameter. Since its introduction, self-calibration has been a standard part of *lens*fit, given its promising overall performance (Mandelbaum et al. 2015; FC17; K19). We keep this feature for the KiDS-Legacy analysis.

3.4.2 Updates for KiDS-Legacy analysis

A long-standing mystery of all previous *lens*fit analyses has been the presence of a small but significant residual bias in ϵ_2 that is uncorrelated with the PSF and the underlying shear (Miller et al. 2013; Hildebrandt et al. 2016; Giblin et al. 2021). We now understand that this feature arises from an anisotropic error in the original likelihood sampler, which has been corrected in our algorithm. However, we found that this correction inadvertently increases the fraction of residual PSF contamination in the weighted average signal (see the discussion in Giblin et al. 2021). Besides, object selection and galaxy weights are also known to introduce bias (e.g. Kaiser 2000, Bernstein & Jarvis 2002, Hirata & Seljak 2003, Jarvis et al. 2016 and FC17). These selection biases can be more severe than the raw measurement bias and hence cannot be ignored even for a perfect self-calibration measurement algorithm.

FC17 presented a method to isotropise weights using an empirical correction scheme, which has been adopted in previous KiDS studies to mitigate these biases. Unfortunately, we found this approach to be insufficient for the improved *lens*fit algorithm. Furthermore, we found the approach to be sensitive to the sample volume, and therefore hard to apply consistently to the data and simulations. So, we introduce a new empirical correction scheme that mitigates the PSF contamination to the weighted shear signal.

Weight correction

We begin by investigating the impact of PSF leakages on the reported weight. For galaxies of similar surface brightness, those oriented in the same direction as the PSF tend to exhibit a higher integrated signal-to-noise ratio compared to those perpendicular to the PSF. This orientation bias introduces an asymmetry in the measurement variance (the $\sigma^2_{\epsilon,\,i}$ term in Eq. 3.5), which can be quantified using a first-order linear function:

$$S_i = \alpha_S \epsilon_{\mathrm{PSF},\,i,\,\mathrm{proj}} + \mathcal{N}\left[\langle S\rangle,\ \sigma_S\right] \ . \tag{3.6}$$

Here, $S_i \equiv \sigma^2_{\epsilon,\,i}$ signifies the measurement variance, while $\epsilon_{\mathrm{PSF},\,i,\,\mathrm{proj}}$ denotes the scalar projection of the PSF ellipticity along the direction of the galaxy ellipticity. The term α_S quantifies the degree of PSF contamination in the measurement variance. The noise, denoted by $\mathcal{N}\left[\langle S\rangle,\ \sigma_S\right]$, is assumed to follow a Gaussian distribution with a mean of $\langle S\rangle$ and a standard deviation of σ_S.

Following FC17, we estimate the PSF contamination as a function of the integrated signal-to-noise ratio (ν_{SN}) reported by *lens*fit and the resolution, which is defined as

$$\mathcal{R} \equiv \frac{r^2_{\mathrm{PSF}}}{r^2_{\mathrm{PSF}} + r^2_{ab}} \ , \tag{3.7}$$

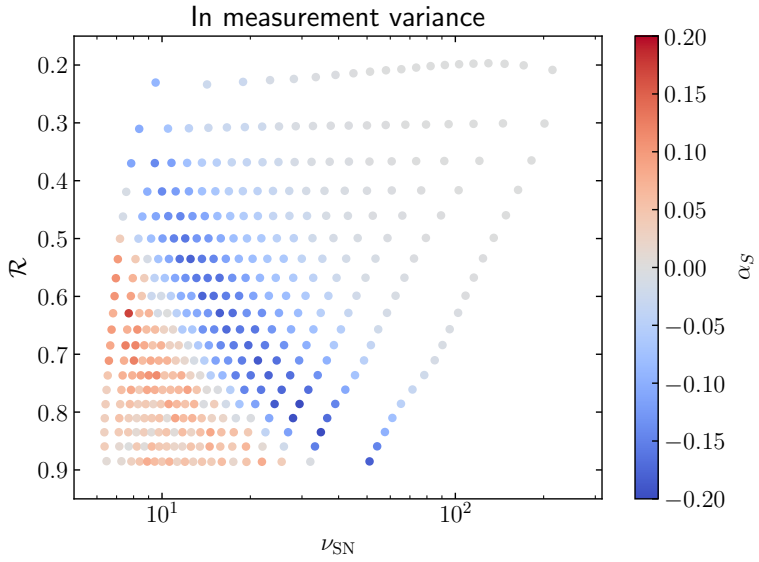

Figure 3.13: PSF leakage in the measurement variance as a function of signal-to-noise ratio and the resolution parameter $\mathcal{R}$. We note that, according to the definition in Eq. (3.7), a larger $\mathcal{R}$ corresponds to a lower resolution.

where $r_{ab} \equiv r_{\rm e}\sqrt{q}$ is the circularised galaxy size with $r_{\rm e}$ and q denoting the scalelength along the major axis and the axis ratio, respectively. The PSF size $r_{\rm PSF}$ is defined by Eq. (3.4). By construction, the resolution $\mathcal{R}$ has a value between 0 and 1, with a larger value corresponding to a more poorly resolved object.

When estimating α_S, we first divide galaxies into an irregular 20×20 grid of $\nu_{\rm SN}$ and $\mathcal{R}$, each containing the same number of objects. Then in each bin, we perform a linear regression using Eq. (3.6) to measure α_S. Figure 3.13 shows the measurements for the KiDS-DR4 re-run with the updated *lens*fit. It demonstrates a clear correlation between the estimated α_S and the $\nu_{\rm SN}$ and $\mathcal{R}$. We derive the corrected measurement variance for individual galaxies through $\sigma^2_{\epsilon,\,i,\,\rm corr} = \sigma^2_{\epsilon,\,i} - \alpha_S \epsilon_{\rm PSF,\,i,\,proj}$, where the value of α_S is determined based on which $\nu_{\rm SN}$-$\mathcal{R}$ bin the target galaxy is assigned to. The corrected *lens*fit weight is then calculated with

$$w_{{\rm corr},\,i} \equiv \left[\frac{\sigma^2_{\epsilon,\,i,\,\rm corr}\,\epsilon^2_{\rm max}}{2\epsilon^2_{\rm max} - 4\sigma^2_{\epsilon,\,i,\,\rm corr}} + \sigma^2_{\epsilon,\,\rm pop}\right]^{-1}, \tag{3.8}$$

following Eq. (3.5). We verified that this approach is sufficient to remove the overall weight bias and is robust against the binning scheme.

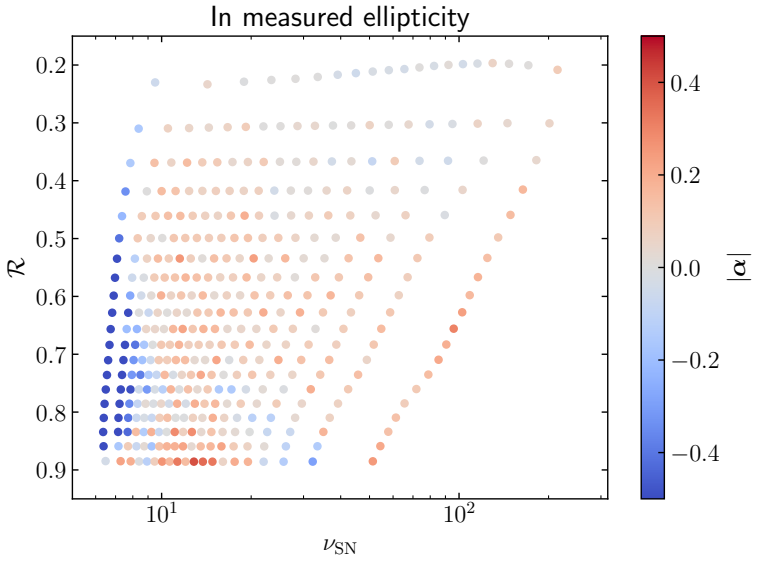

Figure 3.14: PSF leakage in the measured ellipticity after the weight calibration, plotted as a function of signal-to-noise ratio and the resolution parameter $\mathcal{R}$. We note that, according to the definition in Eq. (3.7), a larger $\mathcal{R}$ corresponds to a lower resolution.

3

Ellipticity correction

In addition to the weight bias, there is still some residual PSF leakage in the measured ellipticity because of the residual noise bias and selection effects. To first order, this residual PSF bias can be formulated as

$$\epsilon_{\text{obs},\,i} = \epsilon_{\text{true},\,i} + \boldsymbol{\alpha}\,\epsilon_{\text{PSF},\,i} + \boldsymbol{c} + \mathcal{N}\,[0,\ \sigma_\epsilon]\ , \tag{3.9}$$

where $\epsilon_{\text{obs},\,i}$ is the measured ellipticity, $\epsilon_{\text{true},\,i}$ is the underlying true ellipticity, $\boldsymbol{\alpha}$ is the fraction of the PSF ellipticity $\epsilon_{\text{PSF},\,i}$ that leaks into the measured ellipticity, and $\boldsymbol{c}$ is an additive term uncorrelated with the PSF. $\mathcal{N}\,[0,\ \sigma_\epsilon]$ denotes the noise in individual shape measurements, which are assumed to follow a Gaussian distribution of mean 0 and standard variation σ_ϵ. We note that all parameters in Eq. (3.9) are complex numbers ($\boldsymbol{\alpha} = \alpha_1 + \mathrm{i}\alpha_2$). We focus on the $\boldsymbol{\alpha}$ term, as the $\boldsymbol{c}$ term with the improved likelihood sampler is now small in practice, and the $\mathcal{N}\,[0,\ \sigma_\epsilon]$ vanishes for an ensemble of galaxies.

Like the weight bias correction, we first estimate $\boldsymbol{\alpha}$ in the $20\times\ 20$ grid of $\nu_{\rm SN}$ and $\mathcal{R}$ using a linear regression of Eq. (3.9). Figure 3.14 shows the amplitude of $\boldsymbol{\alpha}$ in the 2D $\nu_{\rm SN}$ and $\mathcal{R}$ plane. We see modest values in most situations, except for the low $\nu_{\rm SN}$ cases, where it drops abruptly to negative values. We confirmed that the negative tail is mainly from the selection effects by measuring the PSF leakage using the input ellipticity in simulations. This non-trivial negative tail prevents us from using the direct correction

approach introduced in the weight bias correction section. Therefore, we propose a hybrid approach, with a fitting procedure for the overall trend and a direct correction for residuals. Specifically, we first fit the measured $\boldsymbol{\alpha}$ as a function of $\nu_{\rm SN}$ and $\mathcal{R}$, using a function of the form

$$\boldsymbol{\alpha}_{\rm p}(\nu_{\rm SN},\ \mathcal{R}) = \boldsymbol{a}_0 + \boldsymbol{a}_1\nu_{\rm SN}^{-2} + \boldsymbol{a}_2\nu_{\rm SN}^{-3} + \boldsymbol{b}_1\mathcal{R} + \boldsymbol{c}_1\mathcal{R}\ \nu_{\rm SN}^{-2}\,, \tag{3.10}$$

whose coefficients are constrained using the weighted mean results from the $20\times\ 20$ grid. Then, we correct the raw measurements of individual galaxies using $\boldsymbol{\epsilon}_{{\rm obs},\,i,\,{\rm tmp}} = \boldsymbol{\epsilon}_{{\rm obs},\,i} - \boldsymbol{\alpha}_{\rm p}(\nu_{{\rm SN},\,i},\ \mathcal{R}_i)\ \boldsymbol{\epsilon}_{{\rm PSF},\,i}$, where the polynomial $\boldsymbol{\alpha}_{\rm p}(\nu_{{\rm SN},\,i},\ \mathcal{R}_i)$ is determined from the target galaxy's $\nu_{{\rm SN},\,i}$ and $\mathcal{R}_i$. After removing the overall trend, we use the corrected $\boldsymbol{\epsilon}_{{\rm obs},\,i,\,{\rm tmp}}$ to measure the residual $\boldsymbol{\alpha}_{\rm r}$, which changes mildly across the 2D $\nu_{\rm SN}$ and $\mathcal{R}$ plane. Therefore, we can conduct the direct correction through $\boldsymbol{\epsilon}_{{\rm obs},\,i,\,{\rm corr}} = \boldsymbol{\epsilon}_{{\rm obs},\,i,\,{\rm tmp}} - \boldsymbol{\alpha}_{\rm r}\ \boldsymbol{\epsilon}_{{\rm PSF},\,i}$, where the values of $\boldsymbol{\alpha}_{\rm r}$ for individual galaxies are determined based on which $\nu_{\rm SN}$-$\mathcal{R}$ bin they are assigned. This two-step approach balances performance and robustness. We verified that the corrected measurements have negligible PSF leakages and the results are robust against the binning scheme.

3.4.3 Comparison between KiDS and SKiLLS

We applied the updated *lens*fit code to KiDS-DR4 and SKiLLS r-band images. The object selections after the measurements are detailed in Appendix 3.D. In short, we largely followed the selection criteria proposed in Hildebrandt et al. (2017), with an additional resolution cut introduced to mitigate the PSF contamination. We applied the same selections to the KiDS data and SKiLLS simulated catalogue to ensure a consistent selection effect, even though SKiLLS does not contain artefacts like asteroids and binary stars.

Figure 3.15 compares the weighted distributions of some critical observables reported by the updated *lens*fit. The SKiLLS results match the KiDS-DR4 data reasonably well. We also checked the properties of the close pairs. Specifically, we show the magnitude difference and the projected distance between close pairs in the measured catalogues. Both properties agree well between the data and simulations, implying SKiLLS has realistic clustering features. These realistic neighbouring properties are essential for an accurate shear calibration, especially when considering the shear interference between blended objects (see Sect. 3.5 for details).

3.5 Shear biases for the updated *lens*fit

The central task of image simulations is to quantify the average shear bias for a selected source sample. This is done by comparing the inferred shear $\gamma_{\rm obs}$, to the input shear $\gamma_{\rm input}$, which have a linear correlation to the first order (Heymans et al. 2006)

$$\gamma_{\rm obs} = (1+\boldsymbol{m})\ \gamma_{\rm input} + \boldsymbol{c}\,, \tag{3.11}$$

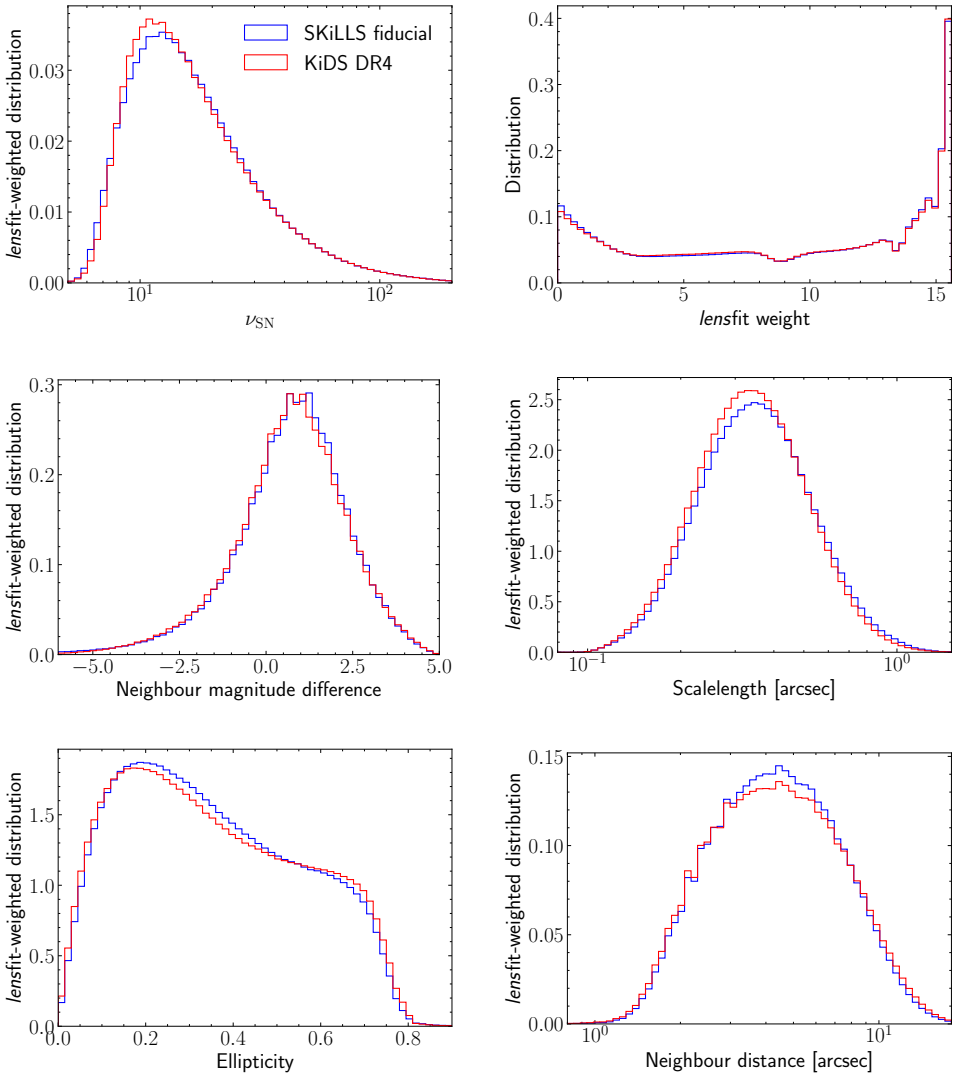

Figure 3.15: Comparison of the updated *lens*fit measurements between KiDS (represented in red) and SKiLLS (in blue). All distributions are normalised using *lens*fit weights, with the exception of the *lens*fit weight distribution itself. Neighbour properties are based on the nearest neighbour identified in the measured catalogue. The magnitude difference is calculated by subtracting the primary target's neighbour magnitude from that of the neighbour. The absence of closely paired galaxies with a distance less than ∼1 arcsecond results from the conservative blending cut used by KiDS (refer to Appendix 3.D). This cut effectively mitigates the most severe blending bias.

where $\boldsymbol{m}$ is known as the multiplicative bias, and $\boldsymbol{c}$ is the additive bias. The simulation-based calibration focuses on the multiplicative bias, as the additive bias is usually corrected empirically (for example, the correction scheme proposed in Sect. 3.4.2). So we use the term 'shear bias' and 'multiplicative bias' interchangeably throughout the paper. We note that all parameters in Eq. (3.11) are in complex forms, such as $\boldsymbol{m} = m_1 + \mathrm{i}m_2$. However, we found m_1 and m_2 to be consistent in our analysis, so unless specified, we only report the amplitude m.

The shear calibration methodology keeps evolving as our understanding of systematics deepens. Early studies demonstrated that the shear bias correlated with galaxy properties and PSFs, especially the signal-to-noise ratio and resolution (e.g. Miller et al. 2013; Hoekstra et al. 2015; Mandelbaum et al. 2018; Samuroff et al. 2018). So the first lesson is to avoid using one averaged result from the whole simulation as a scalar calibration to the entire data unless the simulations perfectly represent the data. A natural procedure then attempts to estimate the shear bias as a function of the galaxy and PSF properties (e.g. Miller et al. 2013; Jarvis et al. 2016). Nevertheless, we can only derive the relation of the bias to the noisy, measured properties, as the true properties are unknown in actual data. FC17 found that the relation derived from the measured properties introduces biases because of the correlations between observed quantities, an effect referred to as the 'calibration selection bias'. So the second lesson is that we should be cautious about object-based shear calibrations that rely on the relation to the noisy properties. That is why the recent simulations try to resemble the data and only provide a mean correction for an ensemble of galaxies (e.g. K19). The latest lesson, stressed by MacCrann et al. (2022), is the interplay between shear estimates of blended objects at different redshifts, a higher-order effect that the traditional constant shear simulations cannot capture. It becomes more important as the precision of surveys improves.

Our shear calibration method builds on all these lessons. We created constant shear simulations following the previous KiDS tomographic calibration method but with improvements to the photo-z estimates by taking advantage of the simulated multi-band images (Sect. 3.5.1). Using additional blending-only variable shear simulations, we applied a correction to account for the interplay between blends containing different shears (Sect. 3.5.2). When testing the PSF modelling algorithm in image simulations, we detected a small but noticeable change of shear bias, which was also corrected in our fiducial results (Sect. 3.5.3).

3.5.1 Results from the constant shear simulations

Our constant shear simulations largely followed FC17 and K19 with some simplifications for better usage of computational resources. Table 3.1 lists the main changes we made compared to our predecessor. Given the 108 deg^2 of unique synthetic galaxies we built in Sect. 3.2, we mimicked 108 KiDS pointings, where we vary the PSF, noise level and stellar density as detailed in Sect. 3.3. To reduce the shape noise, we copied each tile image with galaxies rotated by 90 degrees. We created four sets of constant shear

simulations with input shear: (0.0283, 0.0283), (0.0283, −0.0283), (−0.0283, −0.0283), (−0.0283, 0.0283). The total simulated area is 864 (= $108 \times 4 \times 2$) deg^2, which is equivalent to ∼5170 deg^2 after accounting for the shape noise cancellation (= $864 \times (\sigma_{\epsilon,\mathrm{raw}}/\sigma_{\epsilon,\mathrm{SNC}})^2$, where $\sigma_{\epsilon,\mathrm{raw}}$ and $\sigma_{\epsilon,\mathrm{SNC}}$ denote the weighted dispersion of the mean input ellipticities before and after the shape noise cancellation), which is roughly four times the final KiDS-Legacy area.

For a tomographic analysis, we need to estimate the bias for each redshift bin separately, given that the galaxy properties vary between bins. This requires photo-z estimates for the simulated galaxies. For SKiLLS, we can follow the KiDS processing steps to directly measure photo-zs, thanks to the simulated nine-band images. We conducted the detection from the THELI-like r-band images, the PSF Gaussianisation and forced multi-band photometry using the GAAP pipeline, and the photo-z estimates with the BPZ code (see Sect. 3.3.4 for details). This consistent data processing ensures that SKiLLS embraces realistic photometric properties, marking one of the most significant improvements over the previous image simulations.

Table 3.1: Differences between the COllege (K19) and SKiLLS simulations.

		COllege (K19)	SKiLLS (this work)
Galaxies	Morphology	Sérsic models with parameters taken directly from the HST-ACS measurements (Griffith et al. 2012)	Sérsic models with parameters learned from the HST-ACS measurements (Sect. 3.2.1)
	Photometry	Single-band magnitudes from the Subaru r^+-band observations	Nine-band synthetic magnitudes based on a semi-analytic model (Sect. 3.2.1)
	Depth	Limited by the HST-ACS measurements	Extending to 27th magnitude in the r band
	Position	Based on the observed locations in the COSMOS field	Based on the SURFS N-body simulations (Elahi et al. 2018)
Stars	Photometry	Single-band synthetic magnitudes from the Besançon model (Robin et al. 2003; Czekaj et al. 2014)	Nine-band synthetic magnitudes from the TRILEGAL model (Girardi et al. 2005)
Images	Band	the r-band images only	the full nine-band images

Continued on next page

Table 3.1 continued from previous page

		COllege (K19)	SKiLLS (this work)
	Layout	32 CCDs with even gaps in between	32 CCDs with variable gaps as in the actual camera (Fig. 3.6)
	PSF	13 sets of spatially constant Moffat profiles, with each containing five different models corresponding to the five exposures	108 sets of spatially varying polynomial models, with each containing 5×32 different models
	Noise	One fixed noise level for all tiles	108 different noise levels
	Stack	Only THELI-like stacks for shape measurements	Both THELI-like and ASTRO-WISE-like stacks for shape and photometric measurements, respectively
Measure	Shape	From the self-calibration version of *lens*fit with the weight bias correction of FC17	From the updated *lens*fit with the AlphaRecal method detailed in Sect. 3.4.2
	photo-z	Assigned with the KiDS observations of the COSMOS field	Measured from the simulated nine-band images following the KiDS photometric processing steps (Sect. 3.3.4)
Sample variance		Identical input catalogues of galaxies and stars for all the 13 realisations	Different galaxy catalogues for the 108 realisations and six stellar catalogues for the selected sky blocks (Fig. 3.7)

Continued on next page

Table 3.1 continued from previous page

	COllege (K19)	SKiLLS (this work)
Input shears[a]	Eight sets of constant shears	Four sets of constant shears in the baseline simulations and a variable shear field for the blended objects (Appendix 3.E)
Shape noise cancellation[b]	Each tile has three counterparts with galaxies rotated by 45, 90 and 135 degrees	Each tile has one counterpart with galaxies rotated by 90 degrees
Total simulated area	416 deg^2	864 deg^2 in the constant shear simulations plus 7776 deg^2 of blending-only simulations for the correction of the 'shear interplay' effect (Sect. 3.5.2)

[a] We verified that the four sets of input shears are sufficient to recover the previous results.

[b] Although more rotations suppress shape noise more efficiently (FC17), the selection effects diminish the actual performance of the shape noise cancellation (K19).

As shown in Fig. 3.15, SKiLLS matches KiDS generally well but not perfectly. K19 argued that an accurate estimate of the shear bias must account for any mismatches between the simulations and the target data. Therefore, we followed FC17 and K19 to reweight the simulation estimates using the *lens*fit reported $\nu_{\rm SN}$ and resolution factor $\mathcal{R}$ (Eq. 3.7). Specifically, for each tomographic bin, we first divided simulated galaxies into 20×20 bins of $\nu_{\rm SN}$ and $\mathcal{R}$, each containing equal *lens*fit weight. Then we estimated the multiplicative bias for each $\nu_{\rm SN}$-$\mathcal{R}$ bin using Eq. (3.11). Galaxies in the target data were assigned the bias based on the $\nu_{\rm SN}$-$\mathcal{R}$ bin they fall in, and the final bias for each tomographic bin was the *lens*fit-weighted average of these individual assignments. This procedure ensures the estimated bias accounts for any $\nu_{\rm SN}$ and $\mathcal{R}$ differences between the simulations and the data while also minimising the impact of the calibration selection bias.

Table 3.2 and Figure 3.16 show the multiplicative bias estimates for the KiDS-DR4 re-run with the updated *lens*fit from our constant shear simulations. The quoted errors only contain the statistical uncertainties from the linear fitting. Compared to Table 2 of K19, we reduced the statistical uncertainties by about half because of the larger sky area

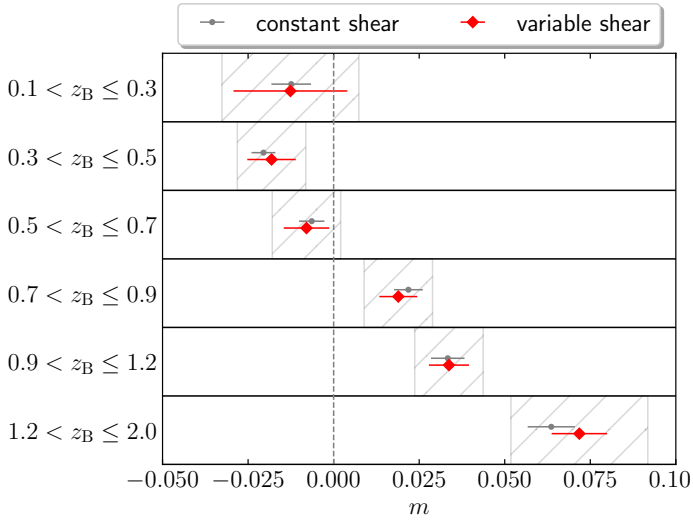

Figure 3.16: Multiplicative bias across tomographic bins for KiDS-DR4, with the updated *lens*fit. The red diamonds represent our final results, including corrections for the shear-interplay effect (Sect. 3.5.2) and PSF modelling bias (Sect. 3.5.3). The grey points, on the other hand, display the raw results from the idealised constant shear simulations (Sect. 3.5.1). The hatched regions denote the proposed nominal error budgets, intended for comparison (refer to Sect. 3.6 for more details).

simulated. Direct comparisons between the calibration values quoted in Table 3.2, cannot be made to those in K19 and Giblin et al. (2021). We updated the shape measurement algorithm *lens*fit and calibrated the raw measurement against PSF contamination in our analysis (see Sect. 3.4.2). These changes modify the effective size and signal-to-noise ratio distribution of the samples and hence the overall calibration in each tomographic bin. Furthermore, Giblin et al. (2021) accounts for the Wright et al. (2020a) 'gold' selection for photometric redshifts, which reduces the effective number density by ∼20%, compared to the sample simulated in this analysis.

3.5.2 Impact of blends at different redshifts

MacCrann et al. (2022) recently highlighted a complication that arises from blended objects at different redshifts, which are, therefore, sheared by different amounts. It stems from the fact that when objects are blended, a shear measurement of one object responds to the shear of the neighbouring object. This higher-order effect, which we refer to as 'shear interplay' through this paper, cannot be captured by the aforementioned constant shear simulations. So, we built an extra suite of variable shear simulations to account for this effect.

Since the shear interplay only happens when objects are blended, we built a blending-only input catalogue for these additional simulations to save some computing time. This blending-only catalogue only contains bright galaxies with bright neighbours, assuming

Table 3.2: Shear bias for the six tomographic bins.

z_B range	Ratio of N_{eff} (blending / whole)	$\Delta\bar{m}_{blending}$	Δm_{PSF}	m_{raw}	m_{final}
$0.1 < z_B \leq 0.3$	0.345	-0.012 ± 0.034	$+0.002 \pm 0.001$	-0.012 ± 0.006	-0.013 ± 0.017
$0.3 < z_B \leq 0.5$	0.332	-0.003 ± 0.014	$+0.004 \pm 0.001$	-0.021 ± 0.004	-0.018 ± 0.007
$0.5 < z_B \leq 0.7$	0.365	-0.021 ± 0.012	$+0.004 \pm 0.001$	-0.006 ± 0.004	-0.008 ± 0.007
$0.7 < z_B \leq 0.9$	0.366	-0.018 ± 0.008	$+0.003 \pm 0.001$	$+0.022 \pm 0.004$	$+0.019 \pm 0.006$
$0.9 < z_B \leq 1.2$	0.370	-0.013 ± 0.007	$+0.005 \pm 0.001$	$+0.033 \pm 0.005$	$+0.034 \pm 0.006$
$1.2 < z_B \leq 2.0$	0.358	$+0.000 \pm 0.008$	$+0.007 \pm 0.002$	$+0.064 \pm 0.007$	$+0.072 \pm 0.008$

The ratio of N_{eff} between the blending-only simulation and the whole simulation is calculated from the measured catalogue with the *lens*fit weight taken into account. The $\Delta\bar{m}_{blending}$ is the mean residual bias introduced by the shear-interplay effect, estimated from the blending-only simulations (see Sect. 3.5.2 for details). The correction to the whole sample should also account for the N_{eff} ratio and the correlation with the signal-to-ratio and resolution (see Sect. 3.5.2 for details). The Δm_{PSF} is the residual bias introduced by the PSF modelling errors (see Sect. 3.5.3 for details). The m_{raw} results are derived from the idealised constant shear simulations (Sect. 3.5.1), and the m_{final} are our final estimates with the corrections for the shear-interplay effect and PSF modelling bias (Sect. 3.5.4). The uncertainties quoted along with individual m values are reported by the linear regression fitting, thus only reflecting the statistical power of SKiLLS simulations. All results are based on the KiDS-DR4 re-run with the updated *lens*fit before any redshift calibration. They only indicate the general performance of the updated *lens*fit.

that the blending effects caused by the faint objects are sufficiently accounted for by our main constant shear simulations, which include galaxies down to magnitude 27. It means we only ignore the higher-order shear-interplay effect from the faint objects, which is valid as long as the excluded faint galaxies are below the measurement limit of the survey. In practice, we selected all galaxies with an input r-band magnitude < 25. The choice of this magnitude cut meets the overall sensitivity of the KiDS survey. We further discarded those isolated galaxies whose nearest neighbour is $4''$ away based on their input positions (see Fig. 3.20). The final selected sample covers $\sim$10% of the entire input catalogue. But after the *lens*fit measurements, this blending-only simulation covers $\sim$35% of the objects measured in the whole simulation (see Table 3.2 for the exact values). The higher fraction in the measured catalogue is because most objects fainter than 25 in the r-band magnitude are not measurable for KiDS.

To properly account for the shear-interplay effect, we need realistic shear fields with proper correlations between the shear and the environment of galaxies. We refer to Appendix 3.E for technical details of our approach to creating such variable shear fields. In short, we considered two primary contributions to the weak lensing signal: the cosmic shear due to the large-scale structure and the tangential shear induced by the foreground objects (also known as the galaxy-galaxy lensing effect). The cosmic shear was learned from the MICE Grand Challenge (MICE-GC) simulation (Fosalba et al. 2015b), whilst the tangential shear was calculated analytically by assuming Navarro-Frenk-White (Navarro et al. 1995) density profiles for the underlying dark matter halos. Figure 3.17 shows the average shear signals as a function of redshift. We see a roughly linear relationship between the mean signals and redshift. On average, the cosmic shear contributes more than the tangential shear. However, we note that the importance of the tangential shear varies between systems depending on the host halo mass of the foreground galaxies.

To increase the constraining power, we used 32 variable shear fields generated from the same learning algorithm but with different choices for the direction of the shear. Specifically, we created four variable shear fields with directions of the cosmic shear that differ by 90°. Then, we made eight copies for each shear field by rotating the final shear by 45° each time. We also created an extra suite of blending-only constant shear simulations to serve as a reference. The final sky area of these additional simulations is 7776 deg^2 ($= 108 \times 36 \times 2$). Except for the input shear, these blending-only simulations use the same pipeline, observational conditions and random seeds as the full simulations detailed in Sect. 3.5.1 so that we can directly correct the constant shear results using the extra bias estimated from these additional simulations.

While estimating the shear bias for constant shear simulations is straightforward by directly conducting the linear least squares fitting to all measurements using Eq. (3.11), given that the input shear values do not depend on the underlying sample. The situation is more complicated for variable shear simulations. The crucial caveat is that the shear bias is now correlated with redshift [$m^{\rm varShear}_{\rm blending}(z_{\rm true})$] due to the shear-interplay effect. Owing to the realistic shear field we built, we can measure $m^{\rm varShear}_{\rm blending}(z_{\rm true})$ directly from simulations by performing the least squares fitting to sub-samples of

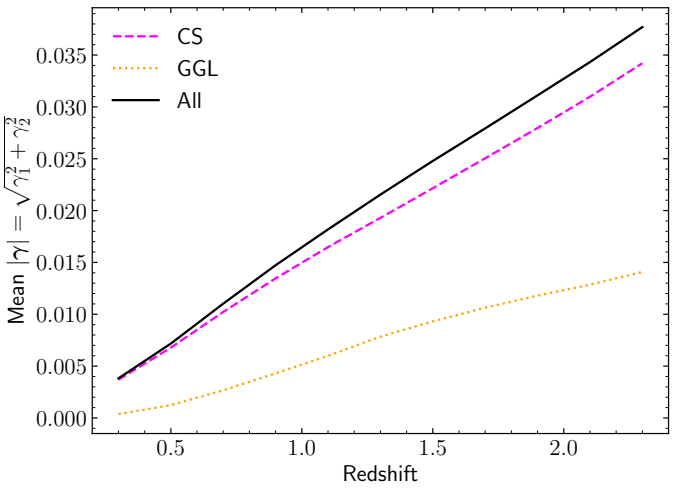

Figure 3.17: Variation of the shear field with respect to redshift. The solid black line represents the average amplitude of the final utilised shears, comprising two components: the cosmic shear (dashed magenta line) and the tangential shear (dotted orange line). For more information, see Appendix 3.E.

galaxies split based on their true redshift. The same approach can also be applied to the blending-only constant shear simulations to get $m^{\text{constShear}}_{\text{blending}}(z_{\text{true}})$; only, in that case, we would expect a negligible correlation with the true redshift, except for some fluctuations stemming from the different signal-to-noise ratios between true redshift bins. Figure 3.18 shows the difference $\Delta m_{\text{blending}}(z_{\text{true}}) \equiv m^{\text{varShear}}_{\text{blending}}(z_{\text{true}}) - m^{\text{constShear}}_{\text{blending}}$, which is a direct measure of the impact of the shear-interplay effect, as the only difference between the simulations is the input shear value. It demonstrates evident residuals that correlate with redshift, indicating the non-trivial impact of the shear-interplay effect. Interestingly, the high-redshift outliers, which have an estimated photo-z much lower than their true redshifts, show the most noticeable residuals across all tomographic bins, implying that the blends with objects from different redshifts are likely responsible for those outliers. This coupling between the photo-z and shear biases in blended systems warrants a dedicated future study.

To correct the raw shear bias derived in Sect. 3.5.1, an average correction $\Delta\bar{m}_{\text{blending}}$ is necessary, which takes into account z_{true} via the equation:

$$\Delta\bar{m}_{\text{blending}} = \int_0^\infty dz_{\text{true}}\; n(z_{\text{true}})\; \Delta m_{\text{blending}}(z_{\text{true}})\;. \tag{3.12}$$

Here, $n(z_{\text{true}})$ signifies the weighted number density related to redshift (as represented by the dashed lines in Fig. 3.18). The averaged results for individual tomographic bins are presented in Table 3.2 and Fig. 3.19.

In practice, the blending fraction, which is associated with the signal-to-noise ratio

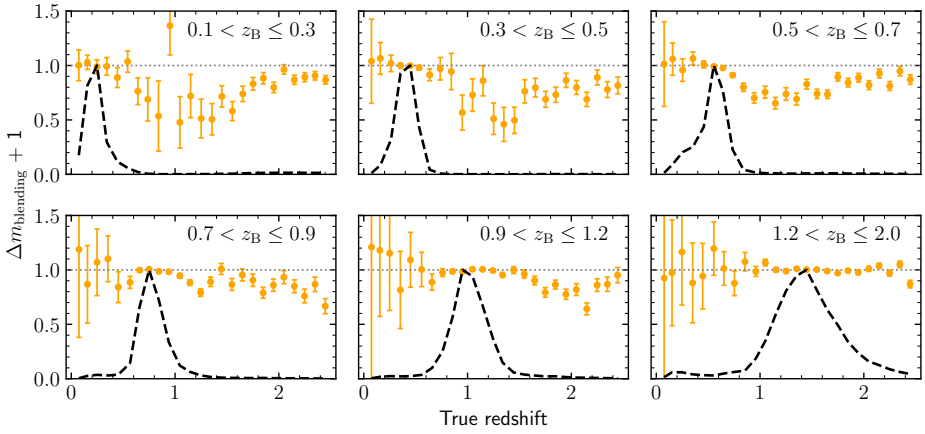

Figure 3.18: Shear bias residuals caused by the shear-interplay effect (orange points) as a function of true redshift, as estimated from the blending-only simulations. The residuals are derived from $\Delta m_{\rm blending} \equiv m^{\rm varShear}_{\rm blending} - m^{\rm constShear}_{\rm blending}$, where $m^{\rm varShear}_{\rm blending}$ refers to the shear bias from the blending-only variable shear simulations, and $m^{\rm constShear}_{\rm blending}$ refers to the shear bias from the blending-only constant shear simulations. The error bars correspond to the fitting uncertainties reported by the linear regression. These uncertainties are influenced by two factors: the number of objects used in the fitting and the amplitude of the input shear value. The dashed lines represent the normalised number density with respect to redshift.

and resolution—much like the bias itself—should also be accounted for. Hence, we implement the correction in each $\nu_{\rm SN}$-$\mathcal{R}$ bin, adhering to the binning approach proposed for reweighting the simulation (refer to Sect. 3.5.1). Specifically, within each $\nu_{\rm SN}$-$\mathcal{R}$ bin, we calculate the average correction $\Delta\bar{m}_{\rm blending}$ and the blending fraction. The blending fraction is measured as the ratio of the effective number counts between the blending-only simulation and the complete simulation. Subsequently, we adjust the raw bias in each $\nu_{\rm SN}$-$\mathcal{R}$ bin with the product of $\Delta\bar{m}_{\rm blending}$ and blending fraction. The final corrected bias is the *lens*fit-weighted average of these adjusted biases. This correction methodology can be easily merged with the reweighting procedure, as they employ the same binning strategy.

Another more direct way to inspect the blending effect is to check the relation between the shear bias and the nearest neighbour distance in the input catalogue. Figure 3.20 demonstrates such estimations for both constant shear and variable shear simulations. We see a clear correlation between the bias and the neighbour distance in both simulations, indicating the significant impact of the blending effect. It also confirms our choice of $4''$ to define blended systems, as we barely see any correlation after this threshold. The other important finding is that the traditional constant shear simulations can already capture the dominant contributions from the blending effect. The higher-order impact

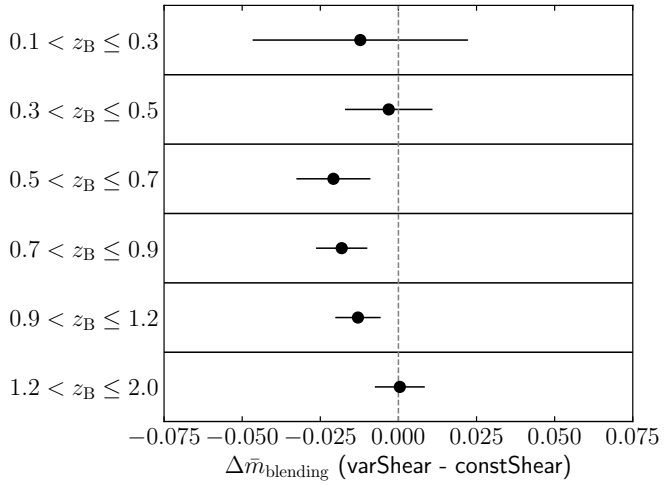

Figure 3.19: Mean residual multiplicative bias induced by the shear-interplay effect, as computed from Eq. (3.12). We emphasise that these results are derived from the blending-only simulations. When applying to the entire sample, we must also account for the blending fraction (the third column of Table 3.2).

we study in this section, shown as the bias difference between the variable shear and constant shear simulations, contributes relatively minor except for the very close blends. The aggressive treatment of the blending in *lens*fit can partially explain this finding, as it throws away most of the recognised blends (Hildebrandt et al. 2017).

We note that our variable shear simulations and the correction methodology differ from those of MacCrann et al. (2022). In their study, the simulated shear changes as a function of redshift, but, per redshift slice, it remains constant across the field of view. The chosen redshift intervals and adjusted shear have no physical meaning in their setups. But they built four sets of simulations by choosing different redshift intervals, so they were able to fit a smooth model to the simulated results, obtaining a continuous redshift-bias relation. In our approach, we computed the variable shear fields using a more physical model that accounts for the shear correlations to both the redshift and clustering of galaxies (see Appendix 3.E). Thanks to these realistic shear fields, we can measure the redshift-bias relation directly from the simulations without additional model fitting procedures. Our direct measurements confirmed the non-trivial impact of the shear-interplay effect (see Fig. 3.18). By design, our method results in large uncertainties for low redshift bins due to the small input shear values. Fortunately, these low redshift bins carry little cosmic shear signals, making the overall downgrade tolerable. Albeit following a different approach, our final result is consistent with MacCrann et al. (2022) finding that the overall correction due to the shear-interplay effect is negligible for the current weak lensing surveys. However, it will potentially impact the next-generation surveys.

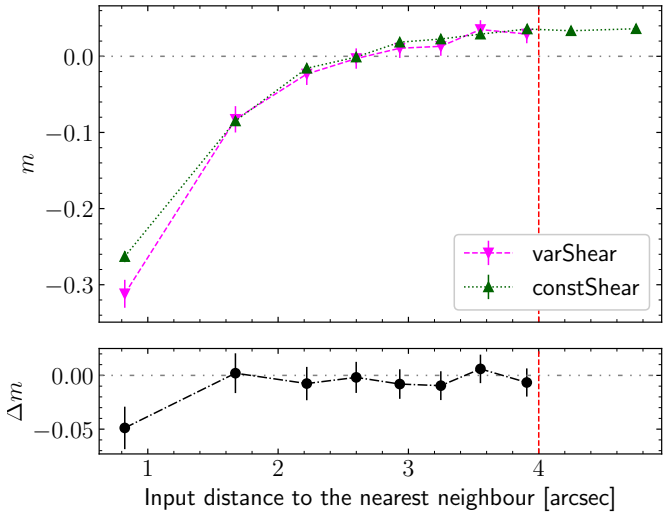

Figure 3.20: Multiplicative bias as a function of the distance to the nearest neighbour. The distance is measured in the input catalogue after exclusion of faint galaxies with r-band input magnitude >25. The values on the x-axis correspond to the weighted average of each sub-sample selected for the estimation of the multiplicative bias shown on the y-axis. The top panel displays the individual biases measured from the blending-only variable shear simulations (magenta points) and the blending-only constant shear simulations (dark green points). The vertical dashed lines denote the threshold we set when creating the blending-only simulations. Two additional dark green points beyond the threshold are calculated from the complete constant shear simulations. The bottom panel shows the difference between these two estimates (varShear - constShear).

3.5.3 PSF modelling bias

So far, we have ignored the PSF modelling errors, given the expected accuracy of PSF models relative to the requirement of the current weak lensing surveys (see e.g. Giblin et al. 2021). We used the input PSF for shape measurements (i.e. assuming perfect PSF modelling). However, as the requirement of systematics becomes more stringent, it becomes necessary to check the impact of PSF modelling errors. This section quantifies this impact by including the PSF modelling procedure in the simulations.

The SKiLLS images have realistic stellar populations and variable PSFs across the field, so we can apply the PSF modelling code directly to the simulated images using similar setups as for the data. We refer to Kuijken et al. (2015) for detailed descriptions of the PSF modelling algorithm used by KiDS. In short, it describes the position-dependent PSFs at the detector resolution using a set of amplitudes on a 48×48 pixel grid. The spatial variation of each pixel value is fitted with a two-dimensional polynomial of order n, with additional flexibility for allowing the lowest order coefficients to differ from CCD to CCD. This extra freedom allows for a more complex PSF variation between CCDs and, in principle, allows for discontinuities in the PSF between adjacent CCDs. When fitting

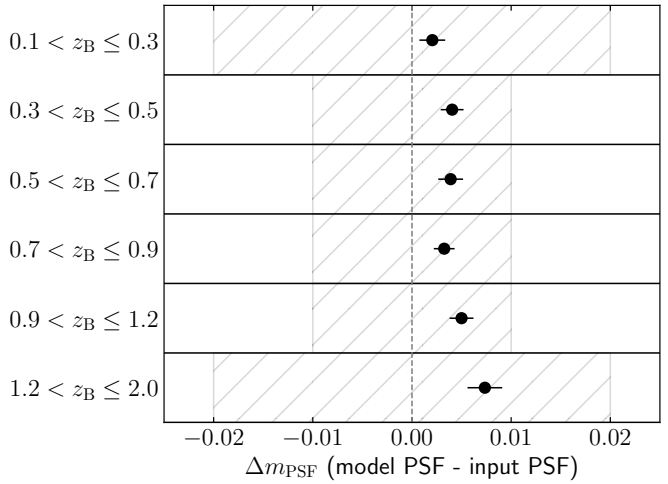

Figure 3.21: Adjustments in multiplicative bias when adopting modelled PSFs instead of the input PSFs. The hatched regions represent the nominal error budgets designated for comparison (refer to Sect. 3.6 for additional details).

to individual stars, the flux and centroid of each star are allowed to change, and a sinc function interpolation is used to align the PSF model with the star position. Following Giblin et al. (2021), we set $n = 4$ and allow the polynomial coefficients up to order 1 to vary between CCDs. We skipped the complicated star-galaxy separation procedure with an implicit assumption that the point-source sample used by KiDS is sufficiently pure as verified using NIR colours in Giblin et al. (2021). Instead, we built a perfect star sample by cross-matching the detected catalogue with the input star catalogue. However, we still applied the same magnitude and signal-to-noise ratio cuts as used in the data to ensure a similar noise level in the modelled stars.

We selected 30 tiles from the available 108 fiducial tiles to test the influence of PSF modelling uncertainty on the multiplicative bias. These selected tiles cover the whole range of the PSF size, including the minimum and maximum. We performed the PSF modelling on the selected tiles and re-ran *lens*fit using the modelled PSFs. Since all the images and detection catalogues are unchanged, the shift of the shear bias directly quantifies the contribution of the PSF modelling errors. Figure 3.21 and Table 3.2 show the shifts for the six tomographic bins. We find the PSF modelling procedure does introduce small yet noticeable biases. Our fiducial results take these additional biases into account.

3.5.4 Results

The final results after accounting for both the shear-interplay effect and PSF modelling errors are listed as $m_{\rm final}$ in Table 3.2 and shown as the red points in Fig. 3.16. Within the current statistical uncertainties, the average shifts due to the shear-interplay effect and

PSF modelling errors are insignificant across all redshift bins, as indicated in Fig. 3.16 between the grey points and the red points. A more noticeable change is the increased uncertainty introduced by the correction of the shear-interplay effect, especially in the low redshift bins where the input shear values are overall small in the variable shear simulations. Our proposed systematic error budgets account for these additional uncertainties (the hatched regions in Fig. 3.16).

3.6 Sensitivity analysis

Given the resemblance between the SKiLLS and KiDS images and the reweighting in the signal-to-noise ratio and $\mathcal{R}$ when estimating the shear biases, it is reasonable to assume that the estimates from SKiLLS can be used to correct the actual measurements. Nevertheless, it is still worth testing the robustness of SKiLLS results and accounting for any potential systematic uncertainties. We start with tests proposed by FC17 and K19 in Sect. 3.6.1. Thanks to the broad coverage of observational conditions in SKiLLS, we can quickly achieve these analyses without dedicated test runs. Additionally, we test how sensitive the *lens*fit results are to the changes in the input galaxy morphology (Sect. 3.6.2). For comparison reasons, we propose some nominal error budgets based on the general performance of SKiLLS and the overall requirements of lensing analyses with KiDS. Specifically, we set an error budget of 0.02 for the first and sixth tomographic bins and 0.01 for the remaining bins. We found these nominal error budgets are conservative enough that our results are robust within them. Nevertheless, we note that these nominal error budgets can be over-conservative for cosmic shear analyses. In which case, we can estimate more accurate systematic uncertainties following other more aggressive approaches proposed by previous KiDS analyses (Giblin et al. 2021; Asgari et al. 2021).

3.6.1 Impact of observational conditions

When developing SKiLLS, we improved most of the critical sources of uncertainty in the previous KiDS simulations. For instance, we based our input galaxy catalogue on N-body simulations, so it has reasonable clustering features and is complete down to 27 in the r-band magnitude. We learned realistic morphologies from observations using a powerful technique, dubbed vine copulas, which captures the multi-dimensional correlations between ellipticities and other galaxy properties. We included six stellar catalogues to account for the varying stellar densities across the survey sky. We covered more variations of the PSF models and background noise levels. Above all, we measured photo-zs directly from the simulated multi-band images to properly account for the correlation between the measurement uncertainties on the redshift and shear estimates. Consequently, most of the sensitivity analyses proposed by FC17 and K19 are either trivial or redundant for the SKiLLS results.

Still, we examine the robustness of the *lens*fit results against some crucial properties by comparing between sub-samples. The basic idea is to split the fiducial simulations into three sub-samples based on a targeted property and examine the consistency of

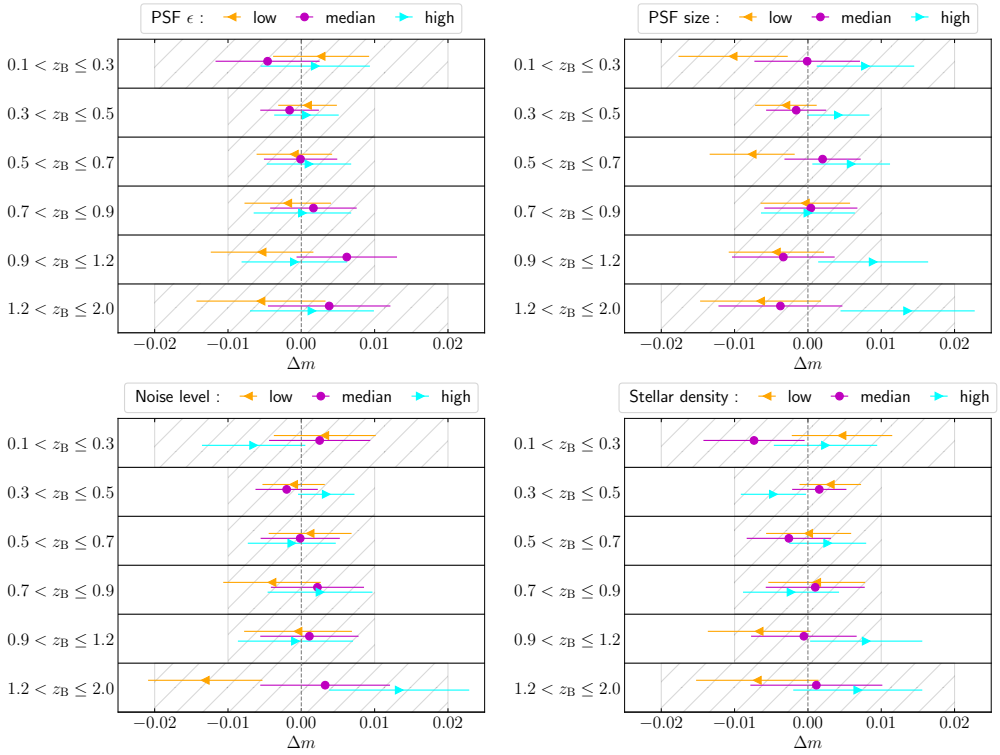

Figure 3.22: Variations in multiplicative bias when the fiducial simulations are divided into three subsets, each reflecting different observational conditions. From left to right and top to bottom, the panels present results when the division is based on PSF ellipticity, PSF size, background noise level in r-band images, and stellar density. The hatched regions denote the proposed nominal error budgets for comparison (refer to Sect. 3.6 for additional details). We note that these shifts represent the maximum possible systematic biases in our results (see Sect. 3.6.1 for further clarification).

their bias estimates to the fiducial results. These sub-samples contain roughly equal numbers of measured objects while covering different ranges of the targeted property. After applying the overall shear correction from the whole sample to the sub-samples, we calculate their residual biases to quantify the impact of the variations of the targeted property. We note that the estimated residuals are not systematic biases in our fiducial results, but they indicate the robustness of the shape measurement algorithm against the tested properties. Ideally, if the simulations fully match the data in the distributions of the targeted property, we would still expect an accurate bias estimate even if the estimated residuals are large. For that account, the estimated residuals are conservative upper limits of the systematic biases in our results.

Figure 3.22 shows the estimated residuals for the variations in four critical properties of the simulated images: the PSF ellipticity, PSF size, background noise level in r-band images, and stellar density. It indicates that our fiducial results are robust within the

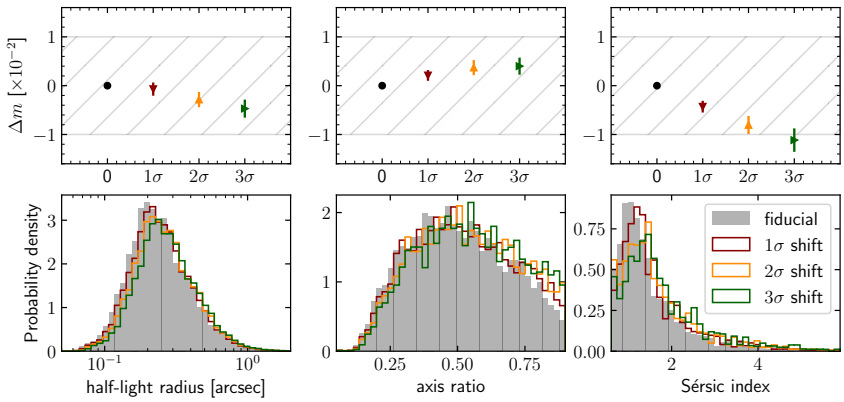

Figure 3.23: Variations in multiplicative bias due to increased morphological parameters of all input galaxies. Top panels show shifts in bias for increased factors of $1 + 1\sigma$ (dark red), $1 + 2\sigma$ (dark orange), and $1 + 3\sigma$ (dark green), where σ indicates median relative uncertainties from Griffith et al. (2012). Hatched regions represent the nominal 0.01 error budget. Bottom panels show normalised histograms of parameter distributions before and after the alterations. The shift of entire distributions depicts extreme cases, as mean shifts across all galaxies would typically be smaller.

nominal error budgets, considering the shifts shown in the plots are the upper limits of possible deviations.

3.6.2 Impact of the input galaxy morphology

We learned the galaxy morphology from Griffith et al. (2012) based on Sérsic models fitted to the HST observations. We have shown that our copula-based learning algorithm captures the properties of the reference sample (see Sect. 3.2.1). However, the reference sample itself contains measurement errors. This section examines how sensitive the *lens*fit measurements are to the changes in the input galaxy morphology.

We focus on the three morphological parameters used to describe the Sérsic profile: the half-light radius, axis ratio and Sérsic index. To get some indication of the overall accuracy of the reference sample, we first checked the fitting uncertainties. We found that the median relative uncertainties for these three parameters are $\lesssim 5\%$, $\lesssim 5\%$ and $\lesssim 10\%$, respectively. We took these values (quoted as σ below) as the benchmark for changing the input galaxy morphology. We built new input catalogues by increasing a certain parameter with 1σ, 2σ and 3σ each time while keeping the other parameters unchanged. We generated test simulations using these new input catalogues and measured the bias difference with respect to the fiducial simulations.

Figure 3.23 presents the test results from 10 tiles of simulations. We find minor residuals in most cases, with the most significant shifts seen when changing the Sérsic index. We note that we shifted all galaxies with the same amount of fractions, resulting

in an overall shift of the whole distribution, as shown in the bottom panels of Fig. 3.23. Given that the entire distribution's uncertainty is much smaller than the individual measurement uncertainties, we are testing the most extreme cases. Hence, the measured residuals only indicate the sensitivity of *lens*fit towards the input galaxy properties but cannot be seen as systematics in our fiducial results. To achieve tighter requirements for future surveys, we will need a shape measurement method that is less susceptible to the galaxy properties, as the fidelity of image simulations will always be limited by the realism of the input galaxy catalogue. For the upcoming KiDS-Legacy analysis, we will, therefore, also explore an alternative method based on the METACALIBRATION technique (Huff & Mandelbaum 2017; Sheldon & Huff 2017), which is expected to be more robust against the galaxy properties (Yoon et al., in prep.).

3.7 Discussion and conclusions

Achieving an unbiased measurement of the ensemble shear signal is crucial for maintaining the integrity of precision cosmology studies using weak lensing surveys. Contemporary shape measurement techniques attained an accuracy level of one percent, or even a fraction thereof. However, as the statistical powers of weak lensing surveys continue to expand, the demand for systematics control intensifies. This resulted in an increased focus on higher-order effects that differ from the shape estimation bias, such as selection bias, PSF modelling errors, and shear-interplay bias. These effects present a challenge to eradicate merely by refining shape measurement algorithms. Alternatively, image simulations offer promising capabilities for calibrating these higher-order effects.

In this chapter, we introduced the third-generation image simulations for the KiDS survey, termed SKiLLS, following SCHO*o*l (FC17) and COllege (K19). These simulations incorporate several substantial enhancements to meet the calibration requirements of the KiDS-Legacy analysis, which utilises an updated *lens*fit. Among the key improvements are the simulation of full nine-band images and the creation of a joint shear-redshift mock catalogue. Balancing the sample volume and realism of galaxy morphology, we combined cosmological simulations with deep imaging observations as input. Additionally, we augmented the image realism by incorporating variations in PSF between CCDs, stellar density, and noise levels between pointings. We closely mirrored the entire KiDS procedure for photometric measurements, encompassing r-band detection, PSF Gaussianisation, forced multi-band photometry, and photo-z estimation. The large volume of simulated galaxies and their realistic photometric properties will enhance not only shear calibration but also redshift calibration, providing a valuable resource for further study (van den Busch et al., in prep.).

We extended our investigations to consider the impact of galaxy blends at varying redshifts by generating realistic shear fields that take into account both redshift and galaxy clustering. We also accounted for PSF modelling errors by implementing the PSF modelling procedures on the image simulations. Additionally, we conducted sensitivity tests, including changing the input galaxy properties, demonstrating the robustness of SKiLLS-calibrated measurements for future KiDS lensing studies. The final shear

calibration results for the updated *lens*fit are outlined in Table 3.2 and Fig. 3.16. Based on our statistical uncertainties and sensitivity tests, we believe that the shear bias estimated from SKiLLS is within the nominal error budget of 0.02 for the first and sixth tomographic bins and 0.01 for the others. Furthermore, our studies provide valuable insights for calibrating future weak lensing surveys.

The success of image simulations heavily depends on the realism of the input galaxy population in terms of photometry, morphology, and clustering. Recent image simulations used high-quality imaging observations as input, but these have limitations in sample volume and depth, which may soon prove inadequate for next-generation weak lensing surveys. While there is an alternative approach that uses input galaxy populations from cosmological simulations, these simulations are currently unable to fully replicate observed galaxy morphology – a crucial feature in image simulations.

In our study, we explored the feasibility of an integrated approach that combines the advantages of cosmological simulations with high-quality imaging observations. We introduced a copula-based learning algorithm designed to mimic and establish a link between observed morphology and synthetic galaxies from cosmological simulations. The results indicate that this hybrid methodology shows promise for future image simulations requiring a substantial volume of galaxies.

Recent studies have already indicated the necessity for shear calibration to consider redshift-related selections. This requires simulating multi-band observations to account for the measurement of photometric redshifts (e.g. K19; MacCrann et al. 2022). We extended this to demonstrate that multi-band image simulations, with a sufficiently large volume of galaxies, not only improve shear calibration but also redshift calibration. By performing the full procedure for photometric measurements, we achieved realistic photometric properties in the mock catalogue. This comprehensive approach advances over previous catalogue-level simulations (e.g. Hoyle et al. 2018; van den Busch et al. 2020; DeRose et al. 2022). In addition, image simulations allow us to examine the impact of blending on redshift estimates, which is challenging to account for at the catalogue level. Given the importance of blending, we believe that integrating shear and redshift calibrations with multi-band image simulations will be crucial for future high-accuracy tomographic analyses.

MacCrann et al. (2022) recently investigated the effects of blended systems in which galaxies experience varying shears, a phenomenon we refer to as 'shear interplay' throughout this paper. We extended their work by creating realistic variable shear fields that account for both the redshift and clustering of galaxies, explicitly including galaxy-galaxy lensing contributions. Although our final results confirmed its relatively minor impact on current weak lensing surveys (see Fig. 3.16), we detected a significant correlation between redshift and shear bias from our blending-only variable shear simulations. This correlation underscores the presence of the shear-interplay effect and its contributions (see Fig. 3.18). Furthermore, we observed that photo-z outliers demonstrate the most pronounced shear interplay, implying a mutual origin of the shear and redshift biases. A focused study is needed to further investigate this correlation in blended systems, as it will become increasingly important for the next generation of

weak lensing surveys.

Typically, image simulations bypass the PSF modelling process, due to the PSF validation conducted in data (see e.g. Giblin et al. 2021). However, thanks to the realistic SKiLLS images, we are able to evaluate the impact of PSF modelling errors by applying the PSF modelling code directly on simulated images. By comparing the shear biases derived from runs with and without PSF modelling, we discerned residual biases from PSF modelling errors, albeit below one percent. Despite being negligible for current requirements, this will be a point of concern for upcoming weak lensing surveys. Therefore, we underscore the necessity of refining the PSF modelling algorithm or incorporating it into image simulations for future surveys.

Lastly, we investigated the sensitivity of our simulation to the characteristics of the input galaxy population. By altering the input values of morphological parameters, we determined that our current standard shape measurement method, *lens*fit, is somewhat sensitive to the input galaxy shapes, but this sensitivity falls within an acceptable range for KiDS analysis. Despite this, we aim to implement an alternative approach based on the Metacalibration technique (Huff & Mandelbaum 2017; Sheldon & Huff 2017) for KiDS-Legacy analysis, which has proven to be more robust against variations in galaxy properties (Yoon et al., in prep.). For future weak lensing surveys, it will be crucial to develop methods less susceptible to galaxy properties, as image simulations may never fully capture the observed galaxy population due to limitations in the input catalogue.

Acknowledgements

We thank Fedor Getman for providing the deep VST-COSMOS catalogue and Arun Kannawadi for reading the manuscript and providing useful comments. We also wish to thank other members of the KiDS-Legacy Calibration Working Group (especially Benjamin Joachimi, Benjamin Stölzner and Anna Wittje) for informative discussions and suggestions through numerous teleconferences. This work used the compute resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University. We acknowledge support from: the Netherlands Research School for Astronomy (SSL); the Royal Society and Imperial College (KK); the Netherlands Organisation for Scientific Research (NWO) under Vici grant 639.043.512 (HHo); and the UK Science and Technology Facilities Council (STFC) under grant ST/N000919/1 (LM) and ST/V000594/1 (CH). We also acknowledge support from the European Research Council (ERC) under grant agreement No. 647112 (CH) and No. 770935 (HHi, JLvdB, AHW); the Max Planck Society and the Alexander von Humboldt Foundation in the framework of the Max Planck-Humboldt Research Award endowed by the Federal Ministry of Education and Research (CH, MY); the Deutsche Forschungsgemeinschaft Heisenberg grant Hi 1495/5-1 (HHi); the Polish National Science Center through grants no. 2020/38/E/ST9/00395, 2018/30/E/ST9 /00698, 2018/31/G/ST9/03388 and 2020/39/B/ST9/03494 (MBi); the Polish Ministry of Science and Higher Education through grant DIR/WK/2018/12 (MBi); the University of Western Australia through a Scholarship for International Research Fees and Ad Hoc Postgraduate

Scholarship (MBr); and the ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D) through project number CE170100013 (CL). The results in this paper are based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs: 088.D-4013, 092.A-0176, 092.D-0370, 094.D-0417, 177.A-3016, 177.A-3017, 177.A-3018 and 179.A-2004, and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA and NWO-M grants; Target; the University of Padova, and the University Federico II (Naples). Contributions to the data processing for VIKING were made by the VISTA Data Flow System at CASU, Cambridge and WFAU, Edinburgh. The SURFS-Shark simulations were produced at the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

Author Contributions: All authors contributed to the development and writing of this paper. The authorship list is given in three groups: the lead authors (SSL, KK, HHo, LM) followed by two alphabetical groups. The first alphabetical group includes those who are key contributors to both the scientific analysis and the data products. The second group covers those who have either made a significant contribution to the data products, or to the scientific analysis.

3.A An empirical modification to the synthetic photometry

We detail the proposed empirical modification of the SHARK photometry in this appendix. It intends to improve the agreement of the magnitude counts between the simulations and observations, which is critical for the redshift and shear calibrations.

We took the COSMOS2015 catalogue as the benchmark under an implicit assumption that the COSMOS field is representative. The COSMOS2015 catalogue is a near-infrared-selected photometric catalogue containing 30-band photometry, precise photometric redshifts and stellar masses for more than half a million objects (Laigle et al. 2016). We note that measurement uncertainties and modelling errors are inevitable for observations, especially for faint objects. Therefore, the COSMOS2015 catalogue cannot, in principle, be treated as the truth. Nevertheless, these uncertainties are tolerable for calibrating a KiDS-like sample. Following this reasoning, we tuned the simulated properties solely based on the COSMOS2015 measurements for the sake of simplicity, but caution any physical interpretation of our modified results.

First of all, we must locate the cause of the discrepancy. As the SHARK free parameters were tuned using the observed stellar mass functions, we would expect the number density of the SHARK galaxies is realistic. This is confirmed by Figure 3.24, where we see a good agreement of the stellar mass distributions between the data and simulations. As a next step, we inspected the stellar mass-to-light ratio ($\Upsilon_\star$), for which took the K_s-band photometry as an indicator of the total luminosity as it is least affected by the dust extinctions. Figure 3.25 shows the comparing results as a function of the stellar mass in several redshift bins. Noticeably, the SHARK $\Upsilon_\star$ is systematically higher than the COSMOS2015 one, especially in the low stellar mass and low redshift ranges. It can, at least partially, explain the discrepancy seen in the magnitude distributions. Fortunately, this $\Upsilon_\star$ difference is easy to calibrate without changing other intrinsic properties, such as the colours, redshifts, and positions.

We, therefore, conducted an empirical modification of the simulated magnitudes to account for the $\Upsilon_\star$ difference. We divided SHARK and COSMOS2015 galaxies into 24×23 evenly spaced small bins based on their redshifts and stellar masses. In each bin, we calculated the median $\Upsilon_\star$ for the SHARK and COSMOS2015 galaxies, separately. To mitigate the observational uncertainties, we only used the COSMOS2015 galaxies with good stellar mass estimations ($\delta \mathrm{M}_\star < 0.15 \mathrm{M}_\star$). For bins that lack observations, we extrapolated $\Upsilon_{\star,\,\mathrm{obs}}$ as a function of $\mathrm{M}_\star$ for each redshift slice. After inspecting the general trend, we found a good fit by combining an exponential descending function in the low $\mathrm{M}_\star$ end and a linear ascending function in the high $\mathrm{M}_\star$ end. From these estimates, we constructed a magnitude modification factor Δmag as

$$\Delta\mathrm{mag} = -2.5 \log_{10} \left(\frac{\mathrm{median}[\Upsilon_{\star,\,\mathrm{SHARK}}]}{\mathrm{median}[\Upsilon_{\star,\,\mathrm{obs}}]} \right) . \tag{3.13}$$

Figure 3.26 demonstrates the estimated Δmag values in the 2D redshift-stellar mass plane. Following the difference seen in Fig. 3.25, substantial modifications happen in the low mass and low redshift bins. Therefore, the magnitude modification reduces the

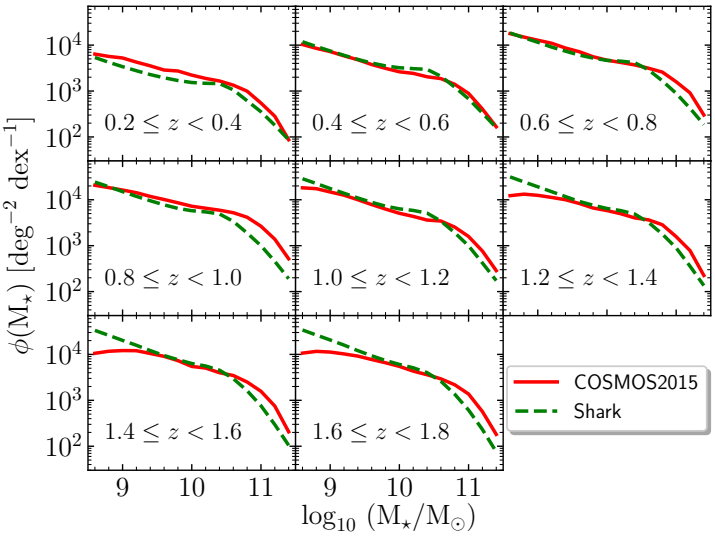

Figure 3.24: Comparison of stellar mass functions. The COSMOS2015 catalogue (represented by red solid lines) utilises the median values derived from the marginalised likelihood distributions. On the other hand, the SHARK catalogue (represented by green dashed lines) operates under the assumption that the total stellar mass is equivalent to the combined stellar masses of the bulge and the disc.

range of magnitudes of SHARK galaxies. We note that the different bands share the same Δmag values, so the colours of individual galaxies are preserved.

3.B Modelling multivariate distributions with vine copulas

We outline some necessary background on the vine-copula modelling in this appendix. For a comprehensive introduction, we refer to Joe (2014) and Czado (2019).

A copula is simply a multivariate cumulative distribution function (CDF) with uniformly distributed margins. The Sklar (1959) theorem states that any d-dimensional CDF $F(\boldsymbol{x})$, with univariate margins $F_1(x_1), ..., F_d(x_d)$, can be described as $F(\boldsymbol{x}) = C_{1,...,d}(F_1(x_1), ..., F_d(x_d))$, where $C_{1,...,d}$ is the corresponding copula function. Therefore, given a joint probability distribution function (PDF) $f(\boldsymbol{x})$ with d-dimensional variables $\boldsymbol{x} = (x_1, ..., x_d)$, we can always find a copula density $c_{1,...,d}$ that is the partial differentiation of the copula $C_{1,...,d}$, such that

$$f(\boldsymbol{x}) = c_{1,...,d}(F_1(x_1), ..., F_d(x_d)) \cdot f_1(x_1) \cdots f_d(x_d) \; . \tag{3.14}$$

It means we can divide the modelling of any joint multi-dimensional PDF into two parts: one for the independent distributions of the individual random variables

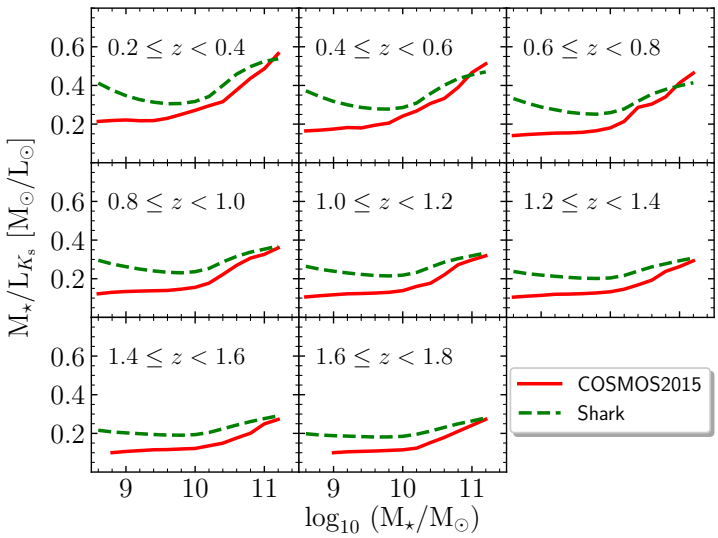

Figure 3.25: K_s-band stellar mass-to-light ratio as a function of stellar mass. The red and green lines correspond to the galaxies in the COSMOS2015 and SHARK catalogues, respectively.

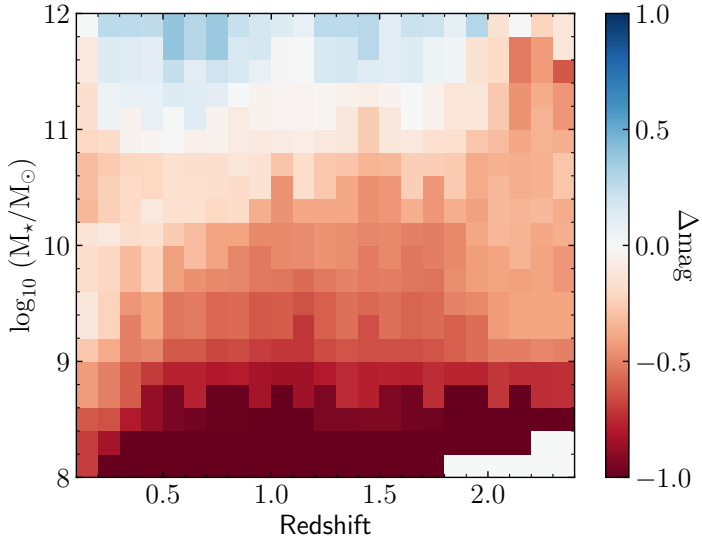

Figure 3.26: Distribution of the magnitude modification factor Δmag in the redshift-stellar mass plane. Negative values are denoted in red, while positive values are denoted in blue. The definition of Δmag is given in Eq. (3.13). For each galaxy, the same Δmag value is added to the apparent magnitudes across all available bands.

$\{f_i(x_i)\}$, and the other for their mutual dependence captured by the copula density $c_{1,...,d}(F_1(x_1), ..., F_d(x_d))$.

The restriction of the classical copula method is that most of the flexible copula families available in the literature are bivariate, making it tricky to deal with high-dimensional distributions. In this aspect, the vine copula method stands out as an effective approach (Bedford & Cooke 2002; Aas et al. 2009). A vine copula is a graphical model organising a set of bivariate copulas, called pair-copulas. The chain rule states that any PDF $f(\boldsymbol{x})$ can be decomposed as

$$f(\boldsymbol{x}) = f(x_d) \cdot f(x_{d-1}|x_d) \cdot f(x_{d-2}|x_{d-1}, x_d) \cdots f(x_1|x_2, ..., x_d) \ , \tag{3.15}$$

with $f(.|.)$ being the conditional PDF. Aas et al. (2009) further states that each term in Eq. (3.15) can be decomposed into an appropriate pair-copula times a conditional marginal density as described by the following general formula

$$f(x|\boldsymbol{v}) = c_{xv_j|\boldsymbol{v}_{-j}}(F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j})) \cdot f(x|\boldsymbol{v}_{-j}) \ , \tag{3.16}$$

where $\boldsymbol{v}$ stands for a d-dimensional vector, v_j is an arbitrary component of $\boldsymbol{v}$, and $\boldsymbol{v}_{-j}$ denotes the v-vector excluding this component. Therefore, the multiple dependence can be captured by a product of pair-copulas acting on underlying conditional probability distributions. Since the decomposition shown in Eq. (3.15) is not unique, there is a significant number of possible pair-copula constructions. These possibilities are organised by the graphical models, that is the vines.

3.C Transformation of the SDSS filters to the KiDS/VIKING filters

This appendix details the transformation of the Sloan Digital Sky Survey (SDSS) photometric system to the KiDS/VIKING system. The SDSS photometric system comprises five colour bands (u, g, r, i, z) that cover wavelengths ranging from ultra-violet at 3000 to near-infrared at 11 000 (Fukugita et al. 1996), whilst the KiDS/VIKING system contains optical filters (u, g, r, i) mounted on the VST OmegaCAM camera (Kuijken 2011) and near-infrared filters (Z, Y, J, H, K_s) mounted on the VISTA infrared camera (González-Fernández et al. 2018). Figure 3.27 compares the filter curves from these two systems. The differences are noticeable, especially for the Z filter, where the KiDS/VIKING system cuts the tail towards long wavelengths. We used the following relation to correct these differences:

$$X_{\mathrm{KiDS/VIKING}} = X_{\mathrm{SDSS}} + j(z_{\mathrm{true}}) \ (X_{\mathrm{SDSS}} - W_{\mathrm{SDSS}}) + h(z_{\mathrm{true}}) \ , \tag{3.17}$$

where X corresponds to the target filter, whilst W is another filter, helping to define the colour. Given the superior depth of the r-band measurement, we picked it as the Y filter whenever possible. When the r band is the target filter, we chose the g band as the Y filter. The coefficients $j(z_{\mathrm{true}})$ and $h(z_{\mathrm{true}})$ are correlated with the redshift, for which we

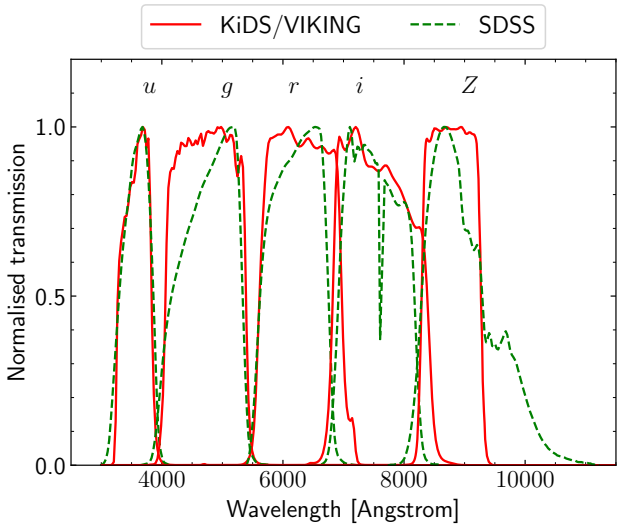

Figure 3.27: Comparison of normalised transmission curves. The curves represent the $ugriZ$ filters in the SDSS photometric system (green dashed lines) and the KiDS/VIKING system (red solid lines).

took values from the ProSpect web-portal[18]. For the redshift, we used the true redshift from the input SURFS-Shark simulations.

As for the SKiLLS measured photometry, we need to correct six measurements: the five u-, g-, r-, i-, Z-band magnitudes measured in the Astro-WISE images (MAG_GAAP_X) and the r-band magnitudes measured in the THELI images (MAG_AUTO). There is no need to correct the remaining $YJHK_s$ bands as SKiLLS also uses VISTA filters for them. Figure 3.28 shows the distributions of the magnitude modification as a function of the initially measured magnitude. The modifications are generally small, especially for the u and g bands. Even for the r and Z bands with the most significant differences, the majority of objects has a modification $\lesssim 0.05$. Accordingly, the changes in the overall magnitude and colour distributions are negligible. Still, we get a better agreement with the data in the photo-z distributions after transforming to the KiDS/VIKING filters, as shown in Fig. 3.29.

3.D Selection criteria for the updated *lens*fit catalogue

This appendix details all selections we propose to the updated *lens*fit shear catalogue. Most of the selection criteria were taken from earlier KiDS analyses, documented in Hildebrandt et al. (2017). These include:

1. Several *lens*fit `fitclass` cuts to discard:

   (a) objects without sufficient data, for example, those fall near the image edge

[18] `https://transformcalc.icrar.org/`

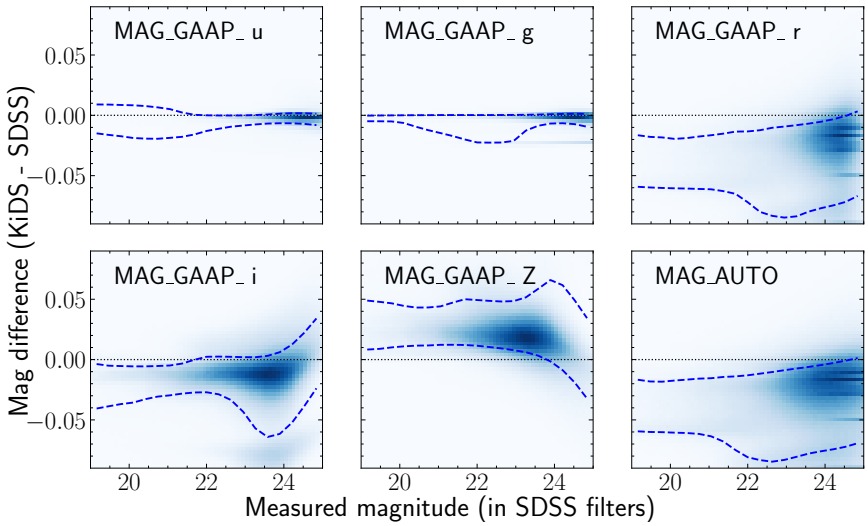

Figure 3.28: Joint distributions of the initially measured magnitude and the magnitude modifications. The dashed lines indicate the 16 and 84 percentiles. 'MAG_GAAP_X' magnitudes are those measured by the GAAP in the ASTRO-WISE images, while 'MAG_AUTO' represents magnitudes measured by SEXTRACTOR in the *r*-band THELI images (refer to Sect. 3.3 for more details).

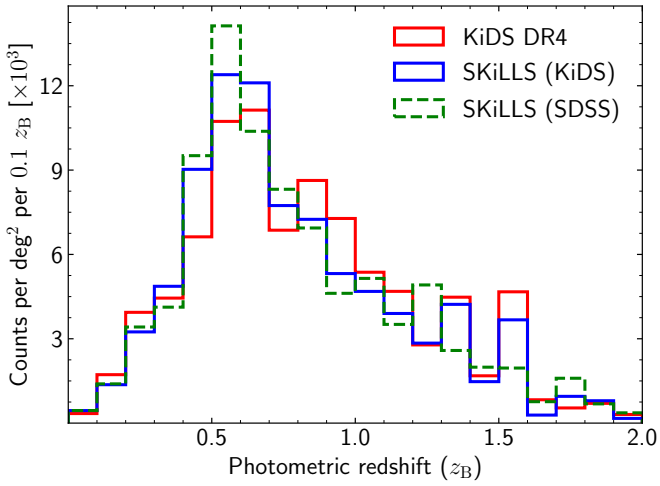

Figure 3.29: Distributions of the photometric redshift estimates. The KiDS-DR4 results are represented by the red histogram. The initial measurements in the SDSS filters are shown by the green histogram, while the blue histogram presents results corrected to the KiDS/VIKING filters. Most of the improvement is visible around $z_B \sim 0.55$ and 1.55.

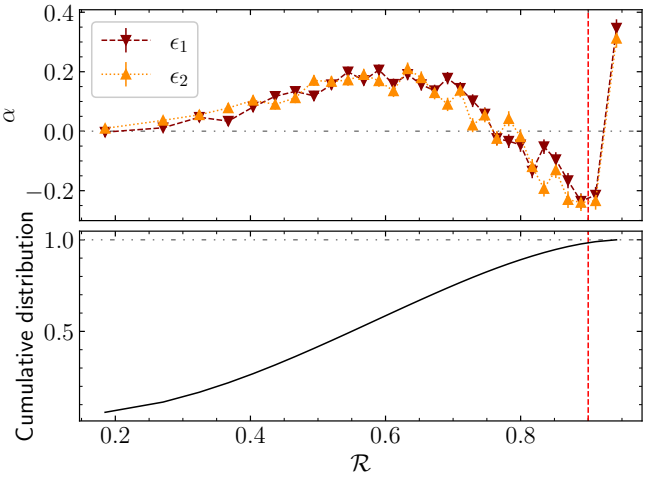

Figure 3.30: PSF leakage and effective number density based on the resolution factor. Top panel: PSF leakage; Bottom panel: effective cumulative distribution. Measurements for weighted average ellipticity ϵ_1 (dark-red triangle) and ϵ_2 (dark-orange triangle) were taken before PSF contamination correction. The vertical red dashed line indicates the proposed resolution cut ($\mathcal{R} < 0.9$), resulting in a 2% loss in effective number density.

or a defect (`fitclass` = −1),

(b) objects classified as duplicates (`fitclass` = −10),

(c) objects poorly fitted by the given bulge plus disc galaxy model (`fitclass` = −4),

(d) objects identified as stars and star-like point sources (`fitclass` = 1 and 2),

(e) objects whose fitted centroid is more than 4 pixels away from the input centroid (`fitclass` = −7),

(f) objects that are unmeasurable, usually because of being too faint (`fitclass` = −3).

2. A magnitude cut to remove bright objects (`MAG_AUTO` > 20).

3. A contamination radius cut to mitigate blending effects (`contamination_radius` > 4.25 pixels)

4. Removing asteroids based on the object colours (`MAG_GAAP_g` − `MAG_GAAP_r` ≤ 1.5 or `MAG_GAAP_i` − `MAG_GAAP_r` ≤ 1.5).

5. Removing unresolved binary stars by requiring objects with ellipticity > 0.8 to have a measured scalelength

$$\geq 0.5 \times 10^{(24.2-\texttt{MAG\_GAAP\_r})/3.5} \text{ pixels} .$$

6. A non-zero weight cut using the weight bias corrected weight (Sect. 3.4.2).

7. A resolution cut to remove poorly resolved objects ($\mathcal{R} < 0.9$).

The resolution cut is a new criterion proposed in this work. When developing our empirical correction method for the PSF contamination (Sect. 3.4.2), we noticed that objects with poor resolution contain very high PSF leakages, as demonstrated in Fig. 3.30. These poor-resolution outliers contribute little to the effective number density but introduce significant bias. So we propose a new selection using the resolution factor defined in Eq. (3.7). We found the proposed cut of $\mathcal{R} < 0.9$ can remove most outliers while only decreasing the effective number density by ∼2 per cent.

3.E Building the variable shear field

In this appendix, we detail the creation of a realistic shear field accounting for the shear dependence on the redshift and clustering of galaxies. We considered the two main contributions to the weak lensing signals: the cosmic shear from the large-scale structure, and the tangential shear from the foreground objects (also known as the galaxy-galaxy lensing effect).

We split the blending-only sample into two classes based on their relative line-of-sight distances to their brightest neighbours. Those more distant than their brightest neighbours are referred as the background galaxies, whilst the remaining are the foreground galaxies. This classification is necessary to quantify the shear correlations within the blended systems. We found a roughly equal number of foreground and background galaxies in our sample.

For the cosmic shear effect, we learned it from the galaxy lensing mocks associated with the MICE Grand Challenge (MICE-GC) simulation (Fosalba et al. 2015b). The MICE-GC simulation is a large volume N-body light-cone simulation developed by the Marenostrum Institut de Ciències de l'Espai (MICE) collaboration (Fosalba et al. 2015a). It contains $\sim 6.9 \times 10^{10}$ dark matter particles with a mass of $\sim 2.9 \times 10^{10}\ h^{-1}\mathrm{M}_{\odot}$ and a softening length of 50 h^{-1}kpc, in a box of 3072 h^{-1}Mpc aside. The simulation starts at $z_{\mathrm{i}} = 100$ and produces the light-cone in 265 steps from $z = 1.4$ to 0. It builds halo catalogues using the Friends-of-Friends algorithm (Crocce et al. 2015), and subsequently populates galaxies using halo occupation distribution recipes along with the subhalo abundance matching technique (Carretero et al. 2015). The construction of all-sky lensing maps follows the Onion Universe approach, which reaches a sub-arcminute spatial resolution up to $z = 1.4$ (Fosalba et al. 2015b). Here we used the second version of the catalogue, named MICECAT2, from the CosmoHub web-portal (Carretero et al. 2017; Tallada et al. 2020)[19].

Following the building of the blending-only sample for SKiLLS, we selected blended objects and classified foreground and background galaxies for MICECAT2 under the same conditions expect for the magnitude cut. We first estimated the relationship between

[19] https://cosmohub.pic.es/

the mean cosmic shear amplitude and redshifts by averaging individual shear values of galaxies in redshift bins defined with a width of 0.1. These redshift-dependent mean amplitudes are good approximations for cosmic shears experienced by the foreground galaxies. It is more intricate to get proper cosmic shears for the background galaxies. Because of the overlapping line-of-sights of the blended objects, we expect the cosmic shear experienced by the background galaxy ($\gamma_{\rm B}$) to correlate with that in its neighbour ($\gamma_{\rm F}$). Based on our tests, the correlation can be described by a linear formula

$$\gamma_{\rm B}(z_{\rm B}, z_{\rm F}) = A(z_{\rm B}, z_{\rm F}) \cdot \gamma_{\rm F} + \gamma_{\rm I}(z_{\rm B}, z_{\rm F}) \,, \tag{3.18}$$

with the scaling factor

$$A(z_{\rm B}, z_{\rm F}) \equiv \frac{D_{\rm c,B} - 0.5 D_{\rm c,F}}{D_{\rm c,B}} \cdot \frac{D_{\rm c,F}}{D_{\rm c,F} - 0.5 D_{\rm c,F}} \,, \tag{3.19}$$

and an offset $\gamma_{\rm I}(z_{\rm B}, z_{\rm F}) = \mathcal{N}\left[0,\ \sigma_{\rm I}(z_{\rm B}, z_{\rm F})\right]$ following the Gaussian distribution with a mean of zero and variance depending on redshifts of both galaxies. The $D_{\rm c,B}$ and $D_{\rm c,F}$ denote the comoving distances to the background galaxy and its neighbour, respectively. The scaling factor A reflects the geometrical relation between the blended objects; whilst the offset $\gamma_{\rm I}$ specifies contributions from the intermediate structures between blended galaxies. We estimated the redshift-dependent variance of $\gamma_{\rm I}$ again from MICECAT2 by measuring the dispersion of $\gamma_{\rm B} - A \cdot \gamma_{\rm F}$ in each redshift bin. Because the MICECAT2 stops at $z = 1.4$, we linearly extrapolated measured values to $z = 2.5$, which is the limit of SKiLLS. Figure 3.31 shows the learned cosmic shear as a function of redshift. The black solid line indicates the mean amplitude of the $\gamma_{\rm F}$ component; whilst the coloured lines present the dispersion of the $\gamma_{\rm I}$ component. It illustrates that the linear extrapolation captures the general trends towards the high redshift for both components.

We note that MICECAT2 assumes a ΛCDM cosmology with parameters from the Wilkinson Microwave Anisotropy Probe five-year data (WMAP5, Dunkley et al. 2009), whilst our base SURFS-Shark simulation uses cosmological parameters from Planck Collaboration (2016). Therefore, the cosmic shear field we learned from MICECAT2 does not necessarily match the galaxy mock we are using. But, since the current calibration still adopts one-point statistics (see Eq. 3.11), our calibration results are robust against detailed galaxy populations or underlying cosmologies and even more so to the higher-order correlation between galaxy populations and cosmology. We defer the proper treatment using a ray-tracing approach with consistent properties from the underlying cosmological simulations to future studies.

Besides the cosmic shear, a background galaxy also suffers from the tangential shear induced by the host dark matter halo of its neighbour. We calculated this effect analytically by assuming Navarro-Frenk-White (NFW) density profiles for dark matter halos presented in the SURFS-Shark simulation. The NFW profile, proposed by Navarro et al. (1995), is the most popular analytical model for dark matter halos, given its ability to describe the radial matter distribution of dark matter halos over a wide range

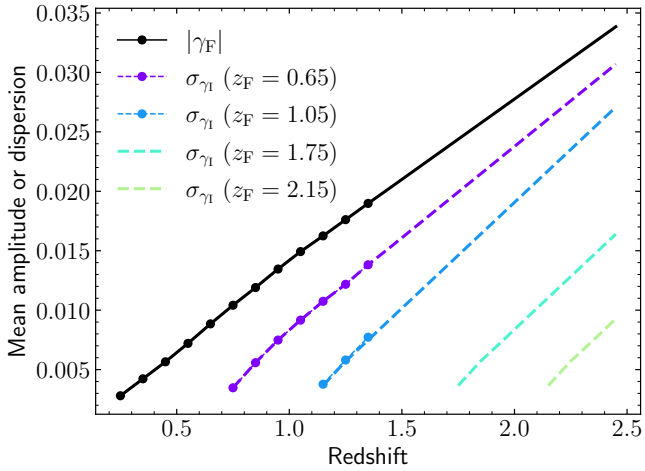

Figure 3.31: Cosmic shear signals learned from MICECAT2 (as described in Eq. 3.18). The black solid line and points represent the mean amplitude of the γ_F component, while the coloured lines and points denote the γ_I dispersion for various redshifts of the foreground galaxies. The points are direct measurements from MICECAT2, while the lines represent linear extrapolations.

of masses (Navarro et al. 1996, 1997). Its mass density is described by the formula

$$\rho(r) = \frac{\rho_{\rm cr}\,\delta_c}{(r/r_{\rm s})(1+r/r_{\rm s})^2}\,, \tag{3.20}$$

where δ_c and $r_{\rm s}$ are two free parameters known as the characteristic overdensity and the scale radius, respectively. We set the normalisation to the critical density at the redshift of the halo $\rho_{\rm cr} \equiv 3H^2(z)/(8\pi G)$ with $H(z)$ the Hubble parameter at that same redshift and G the gravitational constant. With the definition of the virial radius, $r_{\rm 200c}$, the radius inside which the mean mass density of the halo equals $200\rho_{\rm cr}$, we can construct a so-called concentration parameter $c \equiv r_{\rm 200c}/r_{\rm s}$ and relate it to δ_c through

$$\delta_c = \frac{200}{3}\,\frac{c^3}{\ln(1+c) - c/(1+c)}\,. \tag{3.21}$$

In practice, we used `mvir_subhalo`, the virial mass of the subhalo from the SURFS-Shark simulation[20], to calculate the virial radius for each lens. For the concentration parameter, we adopted the concentration–mass relation from Duffy et al. (2008)

$$c = 7.85\,\left(\frac{M_{\rm vir}}{2\times10^{12}\ h^{-1}{\rm M}_\odot}\right)^{-0.081}(1+z)^{-0.71}\,. \tag{3.22}$$

[20] `https://shark-sam.readthedocs.io/en/latest/output_files.html`

We note that Eq. (3.22) is estimated from N-body simulations based on a WMAP5 cosmology (Komatsu et al. 2009), which has slightly different parameter values from the Planck Collaboration (2016) cosmology used by the SURFS simulations. Nevertheless, the weak-lensing shear amplitude is dominated by the enclosed mass of the lens but has minor sensitivity to the concentration (e.g., Viola et al. 2015). Therefore, we ignored any potential WMAP5-to-Planck cosmology correction to Eq. (3.22).

Recognising the spherically symmetric feature of the NFW profile, we can derive the radial-dependent tangential shear as (Bartelmann 1996; Wright & Brainerd 2000):

$$\gamma_{\rm t}(x) = \frac{\rho_{\rm cr}\,\delta_c\,r_{\rm s}}{\Sigma_{\rm cr}} g(x)\ , \tag{3.23}$$

where $x \equiv R_{\rm FB}/r_{\rm s}$ is a dimensionless radial distance factor defined as the ratio of $R_{\rm FB}$, the projected radial separation between the lens and the source, to the scale radius of the lens. The critical surface mass density

$$\Sigma_{\rm cr} \equiv \frac{c^2}{4\pi G}\frac{D_{\rm a,B}}{D_{\rm a,F}\,D_{\rm a,FB}} \tag{3.24}$$

is a geometric term depending on the angular diameter distances to the source $D_{\rm a,B}$, to the lens $D_{\rm a,F}$ and between the lens and the source $D_{\rm a,FB}$. The radial dependence of the shear is captured by the function $g(x)$ as

$$\begin{aligned} g(x) = {} & \frac{4}{x^2}\ln\left(\frac{x}{2}\right) \\ & + \begin{cases} \dfrac{2}{1-x^2} + \dfrac{8-12x^2}{x^2(1-x^2)^{3/2}}\ \mathrm{arctanh}\sqrt{\dfrac{1-x}{1+x}} & (x<1) \\ 10/3 & (x=1) \\ \dfrac{2}{1-x^2} + \dfrac{12x^2-8}{x^2(x^2-1)^{3/2}}\ \arctan\sqrt{\dfrac{x-1}{1+x}} & (x>1) \end{cases} . \end{aligned}$$

With all these ingredients in hand, we can now assign galaxy a specific shear value based on its redshift and neighbouring conditions. In summary, those identified as foreground galaxies only contain the redshift-dependent mean amplitude $\gamma_{\rm F}(z_{\rm F})$, whilst the background galaxies combine the cosmic shear from Eq. (3.18) and the tangential shear from Eq. (3.23). This treatment accounts for not only the redshift-shear dependence but also the correlations between the blended objects.

CHAPTER 4

KiDS-1000: Cosmology with improved cosmic shear measurements

Abstract

We present refined cosmological parameter constraints derived from a cosmic shear analysis of the fourth data release from the Kilo-Degree Survey (KiDS-1000). Our refinements are driven by enhanced galaxy shape measurements using an updated version of the *lens*fit code, and improved shear calibration achieved with a newly developed suite of multi-band image simulations. Additionally, we incorporate recent advancements in cosmological inference from the joint Dark Energy Survey Year 3 and KiDS-1000 cosmic shear analysis. Assuming a spatially flat standard cosmological model, we constrain $S_8 \equiv \sigma_8(\Omega_\mathrm{m}/0.3)^{0.5} = 0.776^{+0.029+0.002}_{-0.027-0.003}$, where the second set of uncertainties accounts for the systematic uncertainties within the shear calibration. These systematic uncertainties stem from minor deviations from realism in the image simulations and the sensitivity of the shear measurement algorithm to the morphology of the galaxy sample. Despite these changes, our results align with previous KiDS studies and other weak lensing surveys, and find a $\sim 2.3\sigma$ level of tension with the *Planck* cosmic microwave background constraints on S_8.

S.-S. Li, H. Hoekstra, K. Kuijken et al.
Astronomy & Astrophysics, submitted

4.1 Introduction

Weak gravitational lensing by large-scale structure, also known as cosmic shear, is a powerful technique for studying the matter distribution in the Universe without assuming a specific correlation between dark and baryonic matter (e.g. Blandford et al. 1991; Miralda-Escude 1991; Kaiser 1992)[1]. Owing to its remarkable potential in exploring the cosmic matter distribution, cosmic shear analysis gained popularity since its first detection over twenty years ago (Bacon et al. 2000; Kaiser et al. 2000; Van Waerbeke et al. 2000; Wittman et al. 2000). When distance information for source galaxies is also known, we can differentiate them along the line of sight and perform a tomographic analysis, which entails reconstructing the three-dimensional matter distribution from multiple two-dimensional projections. This tomographic cosmic shear analysis is especially effective for constraining dark energy properties, as it sheds light on the evolution of cosmic structures (e.g. Hu 1999; Huterer 2002).

Recent surveys, such as the Kilo-Degree Survey (KiDS, de Jong et al. 2013), the Dark Energy Survey (DES, Dark Energy Survey Collaboration et al. 2016), and the Hyper Suprime-Cam (HSC) survey (Aihara et al. 2018), primarily focus on constraining the amplitude of matter density fluctuations. Conventionally, this quantity is characterised by the parameter $S_8 \equiv \sigma_8(\Omega_\mathrm{m}/0.3)^{0.5}$, where Ω_m is the matter density parameter and σ_8 is the standard deviation of matter density fluctuations in spheres of radius $8h^{-1}$ Mpc, computed using linear theory, where the Hubble constant $H_0 = 100h$ km s^{-1} Mpc^{-1}. Interestingly, the S_8 values derived from these weak lensing surveys are consistently lower than those predicted by cosmic microwave background (CMB) observations from the *Planck* satellite.

Specifically, the latest cosmic shear analyses from KiDS ($0.759^{+0.024}_{-0.021}$, Asgari et al. 2021, A21 hereafter), DES ($0.759^{+0.025}_{-0.023}$, Amon et al. 2022; Secco et al. 2022), and HSC ($0.769^{+0.031}_{-0.034}$, Li et al. 2023c; $0.776^{+0.032}_{-0.033}$, Dalal et al. 2023) provide S_8 values that are roughly 2σ lower than the *Planck* predictions (0.832 ± 0.013, Planck Collaboration et al. 2020) based on the standard spatially flat Λ cold dark matter (ΛCDM) cosmological model. Most recently, a joint cosmic shear analysis of the DES Y3 and KiDS-1000 by the two survey teams (DES and KiDS Collaboration et al. 2023, DK23 hereafter) yields an S_8 constraint of $0.790^{+0.018}_{-0.014}$, which is closer to the *Planck* results, but still shows a level of 1.7σ difference. This mild difference in the S_8 constraints between the weak lensing surveys and CMB observations triggered extensive discussions from various perspectives, encompassing potential systematic errors in the data (e.g. Efstathiou & Lemos 2018; Köhlinger et al. 2019), the influence of the baryonic physics (e.g. Schneider et al. 2002; Amon & Efstathiou 2022; Preston et al. 2023), and a potential deviation from the standard ΛCDM model (see Perivolaropoulos & Skara 2022 for a recent review).

Here, we focus on the control of systematics in the cosmic shear analysis, particularly

[1]However, with increasing precision in weak lensing observations, the impact of baryonic processes, such as radiative cooling and feedback from star formation and active galactic nuclei, on the observed matter distribution can no longer be ignored for small-scale structures (e.g. van Daalen et al. 2011; Semboloni et al. 2011).

those arising during the KiDS shear measurement process. Measuring lensing-induced shear from noisy pixelised galaxy images is a challenging task, complicated further by distortions caused by the point spread function (PSF) resulting from instrumental and observational conditions, as well as blending effects that arise when two or more objects are close on the sky (see Mandelbaum 2018 for a review). These factors can introduce significant measurement biases (e.g. Paulin-Henriksson et al. 2008; Melchior & Viola 2012; Refregier et al. 2012; Massey et al. 2013; Dawson et al. 2016; Euclid Collaboration et al. 2019) and alter the selection function of the source sample, leading to selection bias (e.g. Hartlap et al. 2011; Chang et al. 2013; Hoekstra et al. 2021). Therefore, obtaining unbiased shear measurements relies on careful calibration, which can be performed using either pixel-level image simulations (e.g. Miller et al. 2013; Hoekstra et al. 2015; Fenech Conti et al. 2017, FC17 hereafter; Samuroff et al. 2018; Mandelbaum et al. 2018) or the data themselves (e.g. Huff & Mandelbaum 2017; Sheldon & Huff 2017; Sheldon et al. 2020).

Additionally, in the case of large-area imaging surveys, determining the distance information for individual source galaxies depends on redshifts derived from broad-band photometric observations. These photometric redshift estimates, which are subject to significant uncertainty, require careful calibration using spectroscopic reference samples (e.g. Hoyle et al. 2018; Tanaka et al. 2018; Hildebrandt et al. 2021). Furthermore, recent studies showed that the blending of source images results in the coupling of shear and redshift biases (e.g. MacCrann et al. 2022; Li et al. 2023b, L23 hereafter). Consequently, a joint calibration of these two estimates becomes essential, necessitating the use of multi-band image simulations for future cosmic shear analyses.

In light of all these concerns, we implemented several improvements to the cosmic shear measurements in KiDS, as detailed in L23. We enhanced the accuracy of the galaxy shape measurements by using an upgraded version of the *lens*fit code (Miller et al. 2007, 2013; Kitching et al. 2008), complemented by an empirical correction scheme that reduces PSF contamination. More notably, in L23 we introduced SKiLLS (SURFS-based KiDS-Legacy-Like Simulations), a suite of multi-band image simulations that enables joint calibration of shear and redshift estimates. This is an important element for the forthcoming weak lensing analysis of the complete KiDS survey, known as the KiDS-Legacy analysis (Wright et al. in prep.).

In this chapter, we take an intermediate step towards the forthcoming KiDS-Legacy analysis by applying the improvements from L23 to a cosmic shear analysis based on the fourth data release of KiDS (KiDS-1000, A21). In contrast to previous KiDS cosmic shear analyses, which used shear calibration methods developed in FC17 and Kannawadi et al. (2019, K19 hereafter) based on single-band image simulations, the current analysis adopted SKiLLS, marking the first instance of multi-band image simulations being used for KiDS cosmic shear analysis[2]. We also incorporated recent advancements in cosmological inference and updated the current cosmological parameter constraints from KiDS. In particular, we updated the code for the non-linear evolution of the matter

[2]K19 did attempt to assign photo-z estimates from data to simulations, but the actual photo-z measurements were not simulated.

power spectrum calculation from HMCODE to the latest HMCODE-2020 version (Mead et al. 2021). We also investigated the impact of the intrinsic alignment model by incorporating amplitude priors inspired by Fortuna et al. (2021a).

The remainder of this chapter is structured as follows. In Sect. 4.2, we introduce and validate the updated KiDS shear catalogue, followed by the shear and redshift calibration in Sect. 4.3. We describe our cosmological inference method in Sect. 4.4 and present the results in Sect. 4.5. Finally, we summarise the results in Sect. 4.6.

4.2 Updated weak lensing shear catalogue

Our shear catalogue is based on the KiDS-ESO-DR4 data release (Kuijken et al. 2019), which combines optical observations in the *ugri* bands from KiDS using the ESO VLT Survey Telescope (de Jong et al. 2013) and near-infrared observations in the $ZYJHK_s$ bands from the ESO VISTA Kilo-degree INfrared Galaxy (VIKING) survey using the VISTA telescope (Edge et al. 2013). The data set covers 1006 deg^2 survey tiles and includes nine-band photometry measured using the GAAP technique (Kuijken et al. 2015). The photometric redshifts (photo-zs) for individual source galaxies were estimated using the BPZ code (Benítez 2000). After masking, the effective area of the data set in the CCD pixel frame is 777.4 deg^2 (Giblin et al. 2021). To perform the cosmic shear analysis, we divided the source sample into five tomographic bins based on the BPZ estimates (z_B). The first four bins have a spacing of $\Delta z_\mathrm{B} = 0.2$ in the range $0.1 < z_\mathrm{B} \leq 0.9$, while the fifth bin covers the range $0.9 < z_\mathrm{B} \leq 1.2$, following the previous KiDS cosmic shear analyses.

4.2.1 Galaxy shapes measured with the updated *lens*fit

When preparing the shear measurements for the upcoming data release of KiDS, we upgraded the *lens*fit code (Miller et al. 2007, 2013; Kitching et al. 2008) from version 309*c* to version 321 (see L23 for details). The latest version includes a correction to an anisotropic error in the original likelihood sampler, which previously caused a small yet noticeable residual bias that was not related to the PSF or underlying shear (Miller et al. 2013; Hildebrandt et al. 2016; Giblin et al. 2021). We used the new code to re-measure the galaxy shapes in the KiDS-ESO-DR4 data set, resulting in a new shear catalogue. Throughout the paper, we refer to the new shear catalogue as KiDS-1000-v2 to distinguish it from the previous KiDS-1000(-v1) shear catalogue.

The raw measurements from the *lens*fit code suffer from biases primarily due to the PSF anisotropy, but also because of the object selection and weighting scheme. To address these biases, FC17 introduced an empirical correction scheme to isotropise the original measurement weights, which was used in previous KiDS studies (see also K19). However, this approach is insufficient for the current version of the *lens*fit code. Furthermore, L23 found that the method was susceptible to variations in the sample size, posing challenges for consistent application to both data and simulations.

Therefore, a new correction scheme was introduced by L23 that modifies both the

measured ellipticities and weights to ensure the average PSF leakage, defined as the fraction of the PSF ellipticity leaking into the shear estimator, is negligible in each tomographic bin. For further details, we direct readers to L23. In summary, the new correction scheme first isotropises the measurement weights, then adjusts the measured ellipticities to eliminate any remaining noise bias and selection effects. We note that this correction scheme is not designed to refine the shape measurements of individual galaxies; rather, it aims to ensure that the collectively weighted shear signal is robust against PSF leakage. In this paper, we applied this newly developed empirical correction to the KiDS-1000-v2 shear catalogue.

4.2.2 Validation of the shear estimates

In order to use the weak lensing shear catalogue for cosmological inference, it is crucial to first verify the accuracy of the shear estimation and ensure that the residual contamination from systematic effects is within the acceptable level for scientific analysis. To achieve this, Giblin et al. (2021) proposed a series of null-tests to assess the robustness of the KiDS-1000-v1 shear catalogue. With the updated galaxy shape measurements in the KiDS-1000-v2 catalogue, it is necessary to repeat some of these tests to confirm the reliability of the new catalogue.

As the KiDS-1000-v2 catalogue updates only the galaxy shape measurements while maintaining the established photometry and PSF models, we did not repeat tests related to photometry and PSF modelling. We started by examining the PSF leakage in the weighted *lens*fit shear estimator, using the first-order systematics model proposed by Heymans et al. (2006). This model takes the form (Giblin et al. 2021)

$$\epsilon_k^{\rm obs} = (1+m_k)(\epsilon_k^{\rm int}+\gamma_k) + \alpha_k \epsilon_k^{\rm PSF} + c_k \ , \quad [k=1,2] \ , \tag{4.1}$$

where $\epsilon^{\rm obs}$ denotes the measured galaxy ellipticity, m is the multiplicative shear bias[3], $\epsilon^{\rm int}$ refers to the intrinsic galaxy ellipticity, γ stands for the cosmic shear signal (which is the parameter of interest), α is the PSF leakage factor, and c is an additive term comprising residual biases unrelated to the PSF or underlying shear. The subscript $k = 1, 2$ denotes the two ellipticity components. We note that we did not include PSF modelling errors in Eq. (4.1), as we used the same PSF model as Giblin et al. (2021), who had already confirmed its accuracy. Assuming that $(\epsilon_k^{\rm int} + \gamma_k)$ averages to zero for a large galaxy sample (a property validated with the KiDS data; see, for example, Sect. 3 in Giblin et al. 2021), we can determine the α and c parameters from the data using a simple linear regression method.

Figure 4.1 presents the measured PSF leakage α and the additive term c for the KiDS-1000-v2 catalogue, alongside the measurements from the KiDS-1000-v1 catalogues for comparison. As expected, the KiDS-1000-v2 catalogue exhibits a mean α-term consistent with zero for all redshift bins, owing to the empirical correction scheme

[3]Throughout this paper, we interchangeably use 'multiplicative bias' and 'shear bias', as our simulation-based shear calibration only addresses this parameter. Conversely, PSF leakage and the additive term are empirically corrected.

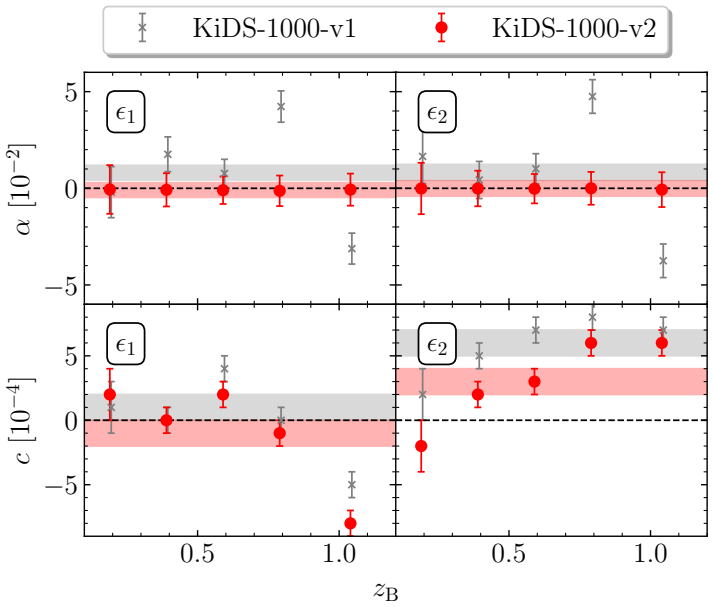

Figure 4.1: PSF contamination α (top panels) and additive term c (bottom panels) as a function of tomographic bin labelled by the central z_B value. The measurements are obtained from a weighted linear fitting using Eq. (4.1). The red points represent measurements from the KiDS-1000-v2 catalogue, while the grey points show the measurements from the KiDS-1000-v1 catalogue. The red/grey coloured bars correspond to results from the entire sample without tomographic binning.

outlined in Sect. 4.2.1 (see also Sect. 4 in L23). The upgraded *lens*fit code has reduced the overall c_2-term by half, reaching a level of $c_2 \sim (3 \pm 1) \times 10^{-4}$ for the entire sample. However, despite this improvement, the c term has not been eliminated, particularly in distant tomographic bins where a small but noticeable c term still persists, which was not seen in the simulations.

To correct for these residual small additive c-terms, we used the same empirical correction method as in previous KiDS analyses. Specifically, we subtracted the weighted average ellipticity from the observed ellipticity for each redshift bin as $\epsilon^{\rm obs}_{\rm corr} = \epsilon^{\rm obs} - \overline{\epsilon^{\rm obs}}$. Nevertheless, we caution that subtracting the mean c-term does not guarantee the removal of all additive biases, especially when detector-level effects, such as 'charge transfer inefficiency' (e.g. Rhodes et al. 2007; Massey 2010) and 'pixel bounce' (e.g. Toyozumi & Ashley 2005), can introduce position-dependent bias patterns. Although we have detected such effects in KiDS data (Hildebrandt et al. 2020; Giblin et al. 2021), their level does not affect the current cosmic shear analysis. More specifically, Asgari et al. (2019) demonstrated that even if current detector-level effects were increased by a factor of ten, they would not cause significant bias for KiDS-like analyses.

The cosmic shear signal is conventionally measured using the two-point shear

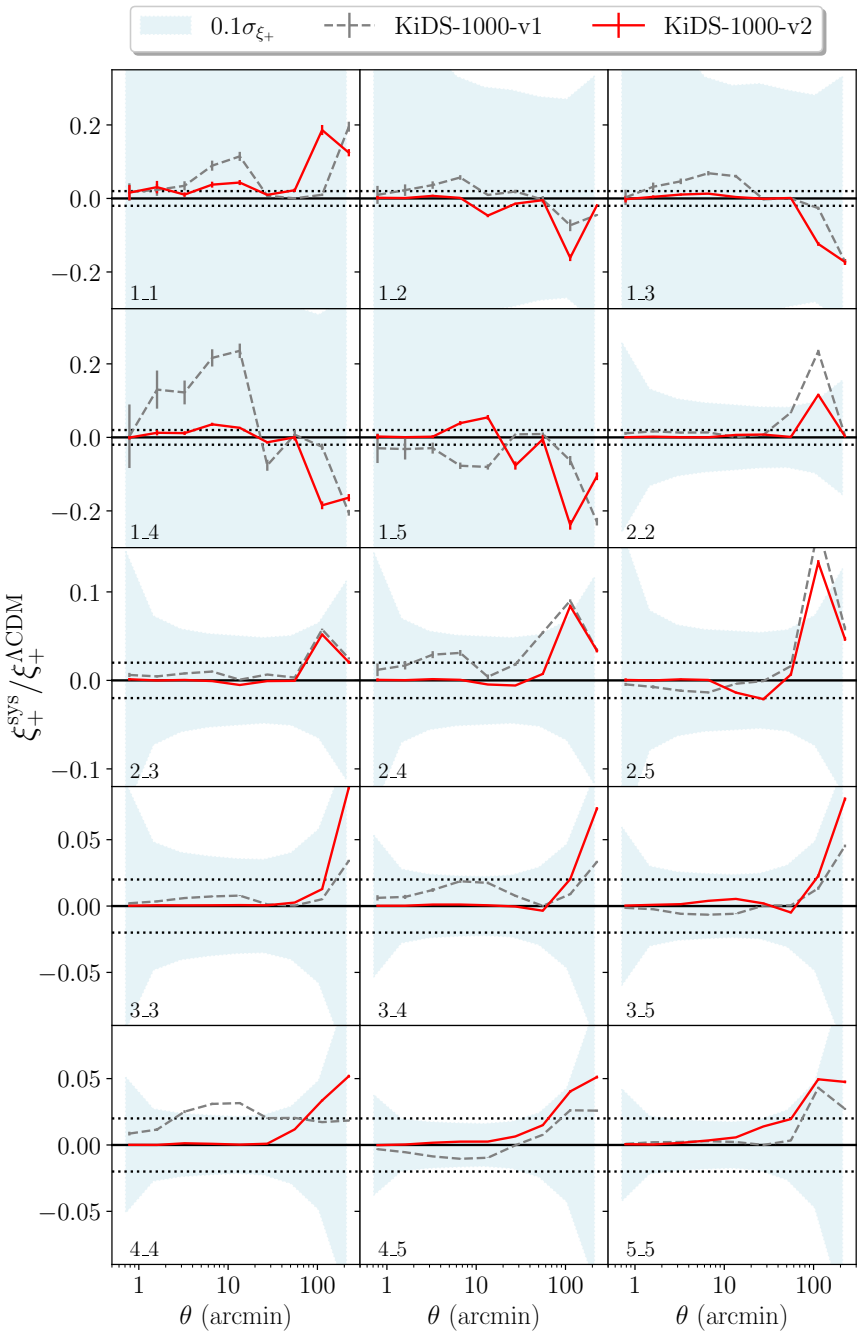

Figure 4.2: Ratio of the PSF contamination $\xi_+^{\rm sys}$, computed using Eq. (4.3), to the predicted amplitude of the cosmic shear signal $\xi_+^{\Lambda\rm CDM}$ across all 15 tomographic bin combinations. The red lines represent results derived from the KiDS-1000-v2 catalogue, while the grey lines correspond to those obtained from the KiDS-1000-v1 catalogue. The blue shaded regions denote ±10% of the standard deviation of the measured cosmic shear signal, determined from the covariance matrix based on the statistics from the KiDS-1000-v2 catalogue.

correlation function, defined as[4]

$$\hat{\xi}^{ij}_{\pm}(\theta) = \frac{\sum_{ab} w_a w_b \left[\epsilon^i_t(\boldsymbol{x}_a)\epsilon^j_t(\boldsymbol{y}_b) \pm \epsilon^i_\times(\boldsymbol{x}_a)\epsilon^j_\times(\boldsymbol{y}_b)\right]}{\sum_{ab} w_a w_b}\,, \tag{4.2}$$

where θ represents the separation angle between a pair of galaxies (a, b), the tangential and cross ellipticities $\epsilon_{t,\times}$ are computed with respect to the vector $\boldsymbol{x}_a - \boldsymbol{y}_b$ that connects the galaxy pair, and the associated measurement weight is denoted by w. Therefore, it is crucial to examine the systematics in the two-point statistics. Following the method of Bacon et al. (2003), we estimate the PSF leakage into the two-point correlation function measurement using

$$\xi^{\rm sys}_{\pm} = \frac{\langle \epsilon^{\rm obs}\epsilon^{\rm PSF}\rangle^2}{\langle \epsilon^{\rm PSF}\epsilon^{\rm PSF}\rangle}\,, \tag{4.3}$$

where the $\langle\cdot\rangle$ represents the correlation function.

In Fig. 4.2, we present the ratio of the measured $\xi^{\rm sys}_+$ to the theoretical predictions of the cosmic shear signal. The blue shaded region denotes $\pm 10\%$ of the standard deviation of the cosmic shear signal, extracted from the analytical covariance. This covariance is calculated using an independent implementation of the methodology of Joachimi et al. (2021), and it incorporates the sample statistics of the updated catalogue. We compare the results from the KiDS-1000-v2 catalogue with those from the KiDS-1000-v1 catalogue. We observe general improvements, particularly in the high-redshift bins, where the PSF contamination is now negligible. The only exceptions are found in some large-scale bins ($\theta > 60$ arcmin), where the expected fiducial cosmic shear signal is relatively small and overwhelmed by high statistical noise.

To leading order, the weak lensing effect introduces only curl-free gradient distortions (E-mode signal), which makes the curl distortions (B-mode signal) a useful null-test for residual systematics in the shear measurement[5]. Following the convention of KiDS (Hildebrandt et al. 2017; Giblin et al. 2021), we use the complete orthogonal sets of E/B-integrals (COSEBIs, Schneider et al. 2010) to measure the B-mode signal. The COSEBIs provide an optimal E/B separation by combining different angular scales from the $\hat{\xi}_\pm$ measurements.

Figure 4.3 presents the measured B-mode signals for all combinations of tomographic bins in our analysis, alongside the B-mode measurements from the KiDS-1000-v1 catalogue for comparison. To enable direct comparison, we used the same scale range of $(0.'5, 300')$ as in Giblin et al. (2021) for calculating the COSEBIs B-mode[6]. Assuming

[4]In this study, all measurements of the two-point shear correlation function are conducted using the TreeCorr code (Jarvis et al. 2004; Jarvis 2015).

[5]Some higher-order effects from lensing, such as source redshift clustering (e.g. Schneider et al. 2002), and intrinsic alignment of nearby galaxies (e.g. Troxel & Ishak 2015; Joachimi et al. 2015) can also introduce B-mode signals. However, these contributions are expected to be negligible for current weak lensing surveys (e.g. Hilbert et al. 2009)

[6]We also evaluated an alternate scale range of $(2', 300')$, consistent with our fiducial cosmic shear analysis. As anticipated, the B-mode signal was more negligible in this scenario due to reduced small-scale contamination.

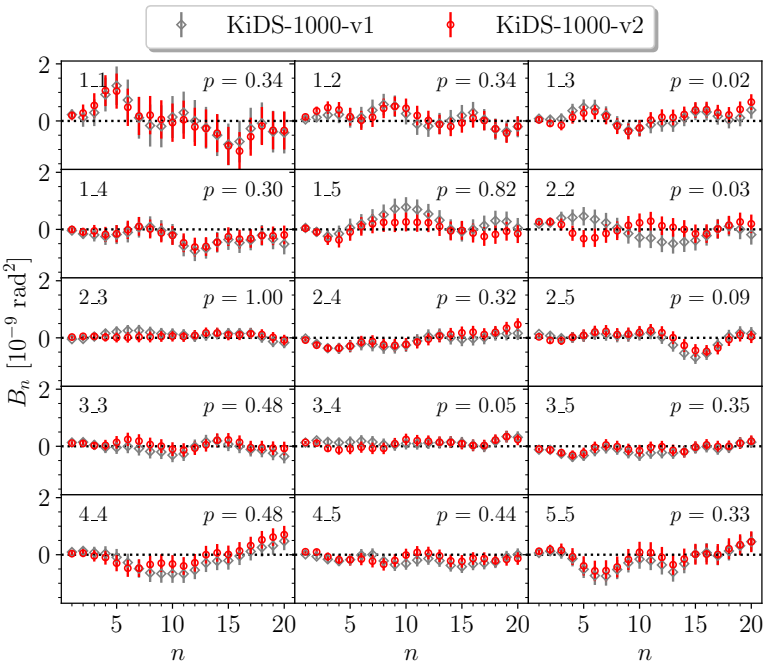

Figure 4.3: Measurements of the B-mode signals using COSEBIs for the KiDS-1000-v2 catalogue (red points) compared to the KiDS-1000-v1 catalogue (grey points). The error bars originate from the diagonal of an analytical covariance matrix, accounting solely for measurement noise. For the KiDS-1000-v2 catalogue, we re-calculated the covariance using the method introduced by Joachimi et al. (2021), incorporating the updated statistics. The p-values for the KiDS-1000-v2 catalogue, shown in the top right corner of each panel, were calculated with 20 degrees of freedom, which corresponds to the number of modes used in each correlation.

a null signal, we computed the p-value for each B-mode measurement, setting the degrees of freedom equal to the number of modes in each measurement ($n = 20$). The covariance matrix, accounting only for shot noise, was estimated using an analytical model from Joachimi et al. (2021) applied to the updated catalogue. It is noteworthy that our covariance matrix differs from the one used in Giblin et al. (2021). This is due to the changes in sample statistics resulting from the updated shape measurement code and redshift calibration relative to the KiDS-1000-v1 catalogue used in Giblin et al. (2021). Most diagonal entries in our matrix show reduced uncertainties, ranging from a level of per cent to ten per cent. Therefore, if the absolute systematic levels are comparable between the two catalogues, our test would likely show a slight increase in the final p-values compared to those in Giblin et al. (2021). As indicated in the top right corner of each panel, the estimated p-values suggest that the measured B-mode signals align with a null signal across all bin combinations. The lowest p-value, $p = 0.02$, was found in the cross-correlation between the first and third tomographic bins.

After conducting all these tests, we can conclude that the KiDS-1000-v2 catalogue has reduced systematics when compared to the results from the KiDS-1000-v1 catalogue. These improvements are largely attributed to the updated version of the *lens*fit code, as well as the implementation of a new empirical correction scheme for PSF contamination. These results give us the confidence to use the updated catalogue for cosmological inference.

4.3 Shear and redshift calibration

The main improvement in our calibration comes from the use of SKiLLS multi-band image simulations, as developed in L23. These simulations fuse cosmological simulations with high-quality observational data to create mock galaxies with photometric and morphological properties closely resembling real-world galaxies. The observational data used by SKiLLS, drawn from the catalogue of Griffith et al. (2012), is identical to that used in K19. In L23, we developed a vine-copula-based algorithm that learns the measured morphological parameters from this catalogue and assigns them to the SURFS-Shark mock galaxies (Elahi et al. 2018; Lagos et al. 2018). We verified that the learning procedure maintains the observed multi-dimensional correlations between morphological parameters, magnitude, and redshifts. Nevertheless, both the observed catalogue from Griffith et al. (2012) and the learning algorithm possess inherent limitations, resulting in unavoidable uncertainties in our simulation input catalogue. These uncertainties are addressed in our shear calibration in Sect. 4.3.2.

To create KiDS+VIKING-like nine-band images, SKiLLS replicated the instrumental and observational conditions of 108 representative tiles selected from six sky pointings evenly distributed across the KiDS-DR4 footprint. The star catalogue was generated for each sky pointing using the TRILEGAL population synthesis code (Girardi et al. 2005) to account for the variation in stellar densities across the footprint. For the primary r-band images, on which the galaxy shapes were measured, SKiLLS included the correlated pixel noise introduced by the stacking process and the PSF variation between CCD images.

On the data processing side, SKiLLS followed the entire KiDS procedure, including object detection, PSF homogenisation, forced multi-band photometry, photo-z estimation, and shape measurements. The end result is a self-consistent joint shear-redshift mock catalogue that matches KiDS observations in both shear and redshift estimates. By taking this end-to-end approach, we accounted for photo-z-related selection effects in our shear bias estimation and enabled redshift calibration using the same mock catalogue. While our current analysis focuses on the improvement in shear calibration, it represents an intermediate step towards the KiDS-Legacy analysis, which will implement joint shear and redshift calibrations facilitated by the SKiLLS mock catalogue.

4.3.1 Calibration

To correct for shear bias in our measurements, we followed the method used in previous KiDS studies (FC17, K19). We applied an average shear bias correction factor, denoted as $\overline{m^i}$, to each tomographic bin i. This factor was calculated by averaging the individual m values of all sources within the bin, with each individual m value obtained using Eq. (4.1). In order to better align the simulations with the target data, we adhered to KiDS conventions by re-weighting the simulation estimates using the *lens*fit reported model signal-to-noise ratio and resolution, which is defined as the ratio of the PSF size to the measured galaxy size. More information on the re-weighting procedure can be found in Sec. 5.1 of L23.

Although the averaging method addresses the noise in individual source's m estimation, it does not account for correlations involving shear bias. Thus, we have $\langle[1+m^i(\theta')][1+m^j(\theta'+\theta)]\rangle = (1+\overline{m^i})(1+\overline{m^j})$, with θ and θ' representing different separation angles between galaxy pairs. To test this assumption, we directly measured $\langle[1+m^i(\theta')][1+m^j(\theta'+\theta)]\rangle$ from image simulations and compared it to $(1+\overline{m^i})(1+\overline{m^j})$. Further details on this test can be found in Appendix 4.A. In summary, we found a negligible difference between the two estimators, a result that falls well within the current KiDS requirements. This validates the assumption for the KiDS analysis.

Given that the updated galaxy shape measurements also lead to changes in the sample selection function, it is necessary to repeat the redshift calibration for the KiDS-1000-v2 catalogue, even though our primary focus is to improve shear calibration. To quantify the changes in galaxy samples introduced by the modifications in shape measurements from the KiDS-1000-v1 to KiDS-1000-v2 catalogues, we compared their effective number densities before applying any redshift calibration. The observed percentage differences in each tomographic bin, from low to high redshift bins, are -1.8%, -0.4%, 0.2%, 1.3%, and 3.2%. Here, negative values indicate a decrease in density from the v1 to the v2 catalogue, while positive values signify an increase. These differences are largely attributed to changes in the weighting scheme from *lens*fit version 309*c* to version 321, as well as the implementation of the new empirical correction scheme for PSF leakage, as discussed in Sect. 4.3 and in L23. For this, we employed a methodology identical to the one used by Wright et al. (2020b), Hildebrandt et al. (2021) and van den Busch et al. (2022) (vdB22 hereafter). It is based on a direct calibration method (Lima et al. 2008) implemented with a self-organising map (SOM, Kohonen 1982; Masters et al. 2015). More information on our implementation is provided in Appendix 4.B, while Wright et al. (2020b), Hildebrandt et al. (2021) and vdB22 offer more comprehensive discussions.

The SOM-based redshift calibration method uses a 'gold selection' criterion to filter out sources that are not represented in the spectroscopic reference sample (see Appendix 4.B). However, this process influences shear biases as it alters the selection function of the final sample. To ensure a consistent estimation of shear biases, we created the SKiLLS-gold catalogue by mimicking this quality control on the SKiLLS mock catalogue, using the same SOM trained by the spectroscopic reference sample as the real

Table 4.1: Data properties for the KiDS-1000-v2 catalogue.

Bin	$n_{\rm eff}$ [arcmin^{-2}]	$\sigma_{\epsilon,i}$	$\delta z = z_{\rm est} - z_{\rm true}$	$m_{\rm raw}$	$m_{\rm final}$	σ_m
1	0.68	0.27	0.000 ± 0.0096	−0.023	−0.021	0.019
2	1.30	0.26	0.002 ± 0.0114	−0.025	−0.023	0.008
3	1.97	0.28	0.013 ± 0.0116	−0.013	−0.015	0.007
4	1.39	0.27	0.011 ± 0.0084	0.018	0.015	0.006
5	1.35	0.29	-0.006 ± 0.0097	0.032	0.031	0.006

Comparable summary statistics for the KiDS-1000-v1 catalogue can be found in Table 1 of A21. We note that the differences in summary statistics between our work and A21 stem from both the updated *lens*fit code and the enhanced redshift calibration outlined in vdB22. The effective number density $n_{\rm eff}$ and the ellipticity dispersion per ellipticity component $\sigma_{\epsilon,i}$ are calculated using the formulae provided in Appendix C of Joachimi et al. (2021). The $n_{\rm eff}$ values in this table are derived from an effective area of 777.4 square degrees in the CCD pixel frame, making them directly comparable to the values in Table 1 of A21. The correlated Gaussian redshift priors are based on the differences between the estimated and true redshifts, $\delta z = z_{\rm est} - z_{\rm true}$, as reported in vdB22. The priors are denoted as $\mu_i \pm \sigma_i$, where μ_i represents the mean shift and σ_i corresponds to the square root of the covariance matrix's diagonal elements. The $m_{\rm raw}$ results are derived from idealised constant shear simulations, while the $m_{\rm final}$ results, our fiducial outcomes, include corrections for the shear-interplay effect and PSF modelling bias. Statistical uncertainties, determined by the simulation volume, are directly computed from the fiducial simulations and denoted as σ_m.

data. We derived the appropriate shear bias correction factors from this SKiLLS-gold catalogue for individual tomographic bins, and present these values in Table 4.1. It is worth noting that the shear bias estimates presented in this work differ slightly from those in L23, which did not include the gold selection procedure. Despite this, the differences in the estimated shear biases are relatively minor across all tomographic bins, with the first tomographic bin showing the most noticeable change of 0.008.

Our fiducial results, $m_{\rm final}$, account for the impact of PSF modelling uncertainties and the 'shear interplay' effect, which occurs when galaxies from different redshifts are blended together. For more details on these effects, we refer to L23 and MacCrann et al. (2022). Additionally, we provide the idealised $m_{\rm raw}$ results, which do not consider these higher-order effects. By comparing the cosmological constraints obtained from these two cases, we aim to evaluate the robustness of previous KiDS results with respect to these higher-order effects, which were not taken into account in the earlier shear calibration (FC17, K19).

4.3.2 Calibration uncertainties

Systematic uncertainties arising from redshift and shear calibrations can propagate into cosmological analyses, potentially leading to biased results. Therefore, it is crucial to adequately address these uncertainties in the analysis. In this section, we outline our approach to managing these calibration uncertainties.

The uncertainties in redshift calibration were addressed by introducing an offset parameter for the estimated mean redshift of galaxies in each tomographic bin. This offset parameter, described as correlated Gaussian priors, serves as a first-order correction to both the statistical and systematic uncertainties associated with redshift calibration. Table 4.1 lists the exact values for these parameters, which we obtained from vdB22 and Hildebrandt et al. (2021). They determined these prior values using spectroscopic and KiDS-like mock data generated by van den Busch et al. (2020). We consider the current priors to be conservative enough to account for any potential changes in the redshift biases from KiDS-1000-v1 to KiDS-1000-v2, given that both catalogues use the same photometric estimates. However, for the forthcoming KiDS-Legacy analysis, we plan to re-estimate these values based on the new SKiLLS mock data.

We improved our approach to handling uncertainties related to the shear calibration. In L23, nominal uncertainties were proposed for each tomographic bin based on sensitivity analyses. This aimed to ensure the robustness of the shear calibration within the specified uncertainties, but at the cost of reducing statistical power. In this work, we aim to improve this approach by separately accounting for the statistical and systematic uncertainties within the shear calibration.

The statistical uncertainties, as presented in Table 4.1, are computed directly from simulations and are limited only by the volume of the simulations, which can be increased with more computing resources[7]. These uncertainties are also easily propagated into the covariance matrix for cosmological inference. Although increasing the simulation volume could, in principle, reduce these uncertainties, we found that the current values already comfortably meet the KiDS requirements, thus further efforts in this direction were considered not necessary.

If the SKiLLS simulations perfectly match KiDS data, these statistical uncertainties would be the only contribution to the final uncertainty from the shear calibration. However, since our simulations are not a perfect replica of the real observations, residual shear biases may still be present in the data even after calibration. These biases, referred to as systematic uncertainties, are typically the primary source of error in shear calibration. Increasing the simulation volume cannot improve these uncertainties as they are determined by the realism of the image simulations. The level of these uncertainties can only be roughly estimated through sensitivity analyses.

Since the systematic residual shear biases directly scale the data vector, accurately quantifying their impact using the covariance matrix is challenging. Therefore, we use a forward modelling approach to capture the impact of these systematic uncertainties. Instead of incorporating these uncertainties into the covariance matrix, we examine how

[7]However, the finite volume of the input galaxy sample prevents an indefinite increase.

the final estimates of the cosmological parameters change due to the shift in signals caused by the systematic residual shear biases. This forward modelling approach can be easily implemented using simple optimisation algorithms since the shift is small, and the covariance remains unchanged. More details on how to determine residual shear biases and implement the forward modelling approach are provided in Appendix 4.C.

4.4 Cosmological inference

The cosmological inference in this study largely aligns with the approach used in the KiDS-1000-v1 analyses (A21; vdB22), with minor modifications primarily influenced by the recent joint DES Y3+KiDS-1000 cosmic shear analysis (DK23). In this section, we outline the configurations and reasoning behind these choices in our fiducial analysis. For certain notable changes, we also conduct extended analysis runs with different configurations to evaluate the impact of these modifications. Our analysis code is publicly accessible[8].

Our code builds upon the `Cat_to_Obs_K1000_P1`[9] and the KiDS Cosmology Analysis Pipeline (KCAP)[10] infrastructure, as developed in Giblin et al. (2021), Joachimi et al. (2021), A21, Heymans et al. (2021) and Tröster et al. (2021). The `Cat_to_Obs_K1000_P1` pipeline converts KiDS shear and redshift measurements into various second-order statistics, with the assistance of the TreeCorr code (Jarvis et al. 2004; Jarvis 2015). Meanwhile, KCAP estimates cosmological parameters using the CosmoSIS framework, which bridges the likelihood function calculation pipelines with MCMC samplers (Zuntz et al. 2015).

We measure the shear field using Complete Orthogonal Sets of E/B-Integrals (COSEBIs, Schneider et al. 2010). As reported by Asgari et al. (2020), COSEBIs offer enhanced robustness against small-scale effects on the shear power spectrum, which primarily stem from complex baryon feedback. Furthermore, we account for baryon feedback when modelling the matter-matter power spectrum, using HMCODE-2020 (Mead et al. 2021) within the CAMB framework with the version 1.4.0 (Lewis et al. 2000; Howlett et al. 2012).

HMCODE-2020, an updated version of HMCODE (Mead et al. 2015, 2016), models the non-linear matter-matter power spectrum, incorporating the influence of baryon feedback through an enhanced halo-model formalism. This updated model is empirically calibrated using hydrodynamical simulations, following a more physically informed approach. Unlike its predecessor calibrated with OWLS hydrodynamical simulations (van Daalen et al. 2011), this newer version uses the updated BAHAMAS hydrodynamical simulations for calibration (McCarthy et al. 2017). These simulations, in turn, are calibrated to reproduce the observed galaxy stellar mass function and the hot gas mass fractions of groups and clusters. This calibration ensures that the simulation accurately reflects the impact of feedback on the overall distribution of matter (refer to McCarthy et al. 2017 for

[8] https://github.com/lshuns/CSK1000LF321
[9] https://github.com/KiDS-WL/Cat_to_Obs_K1000_P1
[10] https://github.com/KiDS-WL/kcap

further details). Furthermore, HMCODE-2020 improves the modelling of baryon-acoustic oscillation damping and massive neutrino treatment, achieving an improved accuracy of 2.5% (compared to the previous version's 5%) for scales $k < 10h\,\mathrm{Mpc}^{-1}$ and redshifts $z < 2$ (Mead et al. 2021).

The model incorporates a single-parameter variant, T_{AGN}, representing the heating temperature of active galactic nuclei (AGN). Higher T_{AGN} values correspond to more intense AGN feedback, leading to a lower observed matter power spectrum. Following DK23, we use a uniform prior on $\log_{10}(T_{\mathrm{AGN}})$ ranging from 7.3 to 8.0. This choice is motivated by the findings from the BAHAMAS hydrodynamical simulations (McCarthy et al. 2017; van Daalen et al. 2020).

Given the characteristics of COSEBIs and the implementation of the HMCODE, the KiDS-1000-v1 analyses included small-scale measurements down to $\theta_{\mathrm{min}} = 0.'5$. This strategy was, however, re-evaluated in DK23, which suggested more stringent scale cuts for the KiDS COSEBIs data vector, determined by the baryon feedback mitigation strategy proposed by Krause et al. (2021). Following this recommendation, we apply a scale cut of $\theta_{\mathrm{min}} = 2'$ in our fiducial analysis.

We use the non-linear linear alignment (NLA) model to describe the intrinsic alignment (IA) of galaxies. This model combines the linear alignment model with a non-linear power spectrum and contains a single free parameter A_{IA} to describe the amplitude of IA signals (Hirata & Seljak 2004; Bridle & King 2007). It is also common to include a power law, with an index denoted as η_{IA}, to capture potential redshift evolution of the IA strength. To distinguish it from the redshift-independent NLA model, we refer to this variant as the NLA-z model.

In line with previous KiDS analyses, we take the redshift-independent NLA model as our fiducial choice since introducing η_{IA} has a minimal effect on the primary S_8 constraint (A21), and current direct observations of IA signals show little evidence of substantial redshift evolution (e.g., Joachimi et al. 2011; Singh et al. 2015; Johnston et al. 2019; Fortuna et al. 2021b; Samuroff et al. 2022). However, Fortuna et al. (2021a) suggests that the selection of galaxy samples resulting from the redshift binning may introduce a detectable redshift variation in the IA signal, although its impact remains negligible for current weak lensing analyses. To assess the impact of η_{IA} on our results, we perform an extended run using the NLA-z model, following the same prior selection as in DK23.

The KiDS-1000-v1 analyses adopted a broad and uninformative prior for A_{IA}, ranging from $[-6, 6]$, considering that the data can constrain it and that an incorrect informative prior could bias the final cosmological results. Although uncertainties regarding IA signals remain large, recent developments in the field have improved our knowledge of the expected IA signal strength. For instance, Fortuna et al. (2021a) used a halo model formalism, incorporating results from the latest direct IA measurements, and predicted $A_{\mathrm{IA}} = 0.44 \pm 0.13$ for the redshift-independent NLA model targeted for KiDS-like mixed-colour lensing samples[11]. This prediction aligns well with the

[11] Fortuna et al. (2021a) also examined the NLA-z model under similar conditions, but found the fits were predominantly driven by the low-redshift bins, resulting in less accurate recovery of large-scale alignments

constraints from recent cosmic shear analyses (A21; Secco et al. 2022; Li et al. 2023c; Dalal et al. 2023). Moreover, recent studies revealed that other nuisance parameters in such analyses, especially those related to redshift calibration uncertainties, can result in misleading $A_{\rm IA}$ values (Hikage et al. 2019; Wright et al. 2020b; Li et al. 2021; Fischbacher et al. 2023).

Given these considerations, we consider it is necessary to explore the prior for the $A_{\rm IA}$ parameter. As an initial step towards a fully informed $A_{\rm IA}$ approach, we begin by simply narrowing the previously broad prior, leaving a more comprehensive exploration of the IA model setups for the forthcoming KiDS-Legacy analysis. In our fiducial analysis, we choose a flat yet narrower prior of $[-0.2, 1.1]$, which corresponds to the 5σ credible region of predictions by Fortuna et al. (2021a). We note that our new prior will not significantly impact the sampling results, provided that the final posterior distributions fall within the set prior range. For comparison purposes, we also conduct a test run using the wider $[-6, 6]$ prior.

Sampling the high-dimensional posterior distribution is a challenging task. In the KiDS-1000-v1 analyses, an ellipsoidal nested sampling algorithm, MultiNest (Feroz et al. 2009), was used. However, recent studies demonstrated that MultiNest systematically underestimates the 68% credible intervals for S_8 by about 10% in current weak lensing analyses (Lemos et al. 2023; DK23; Li et al. 2023c). A promising alternative is the sliced nested sampling algorithm, PolyChord (Handley et al. 2015a,b), which provides more accurate estimates of parameter uncertainties. However, PolyChord is almost five times slower than MultiNest. As a result, we opt to use PolyChord for our main analysis, while retaining MultiNest for our testing purposes.

When presenting point estimates and associated uncertainties for parameter constraints, we adhere to the recommendations of Joachimi et al. (2021). We derive our best-fit point estimates from the parameter values at the maximum of the joint posterior (MAP). Given that the MAP reported by the sampling code can be affected by noise due to the finite number of samples, we enhance the precision of the MAP by conducting an additional local optimisation step. This process initiates from the MAP reported by the sampling code and utilises the Nelder-Mead minimisation method (Nelder & Mead 1965), a method also employed by A21. To represent uncertainties linked to these estimates, we compute the 68% credible interval based on the joint, multi-dimensional highest posterior density region, projected onto the marginal posterior of the parameter of interest (PJ-HPD). This hybrid approach is more robust against projection effects stemming from high-dimensional asymmetric posterior distributions than traditional 1D marginal summary statistics (refer to Sect. 6 in Joachimi et al. 2021 for a comprehensive discussion). To facilitate comparison with results from other surveys, we also provide constraints based on the traditional mean and maximum of the 1D marginal posterior, along with their respective 68% credible intervals.

It is worth noting that, as systematic uncertainties from shear calibration are excluded in the construction of our covariance matrix (see Sect. 4.3.2), the uncertainties derived from the main sampling chains do not fully account for the true uncertainties. To

at high redshifts.

Table 4.2: Fiducial model parameters and their priors.

Parameter	Symbol	Prior
Density fluctuation amp.	S_8	[0.1, 1.3]
CDM density	ω_c	[0.051, 0.255]
Baryon density	ω_b	[0.019, 0.026]
Hubble constant	h	[0.64, 0.82]
Scalar spectral index	n_s	[0.84, 1.1]
AGN heating temperature	$\log_{10}(T_{AGN}[K])$	[7.3, 8.0]
Intrinsic alignment amp.	A_{IA}	[−0.2, 1.1]
Redshift offsets	δ_z	$\mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\sigma^2})$

The first section lists the primary cosmological parameters describing the ΛCDM model assumed, while the second section contains nuisance parameters related to baryon feedback, intrinsic alignments, and redshift biases. The values in square brackets indicate the limits of top-hat priors. The notation $\mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\sigma^2})$ refers to a normal prior with mean $\boldsymbol{\mu}$ and (co-)variance $\boldsymbol{\sigma^2}$, as specified in Table 4.1.

compensate for the additional uncertainties arising from residual shear biases, we employ a forward modelling approach. This method involves shifting the data vector and subsequently the likelihood, based on the estimated residual shear biases, followed by recalculating the MAP. As the adjustment is minor and the covariance matrix remains static, it is not necessary to re-sample the posterior distribution. Instead, we simply need to repeat the previously mentioned local optimisation step. Starting with the original MAP and using the updated likelihood, we can determine the new MAP corresponding to each shift in the data vector. The variation in these MAP estimates represents additional uncertainties introduced by the systematic uncertainties arising from shear calibration. Further details on this process can be found in Appendix 4.C.

Table 4.2 summarises the model parameters and their priors as used in our fiducial analysis. These parameters can be broadly classified into two categories: the first category includes five cosmological parameters, which describe the spatially flat ΛCDM model we employ. We fix the sum of the neutrino masses to a value of 0.06 eV c^{-2}, where c is the speed of light. This choice is based on Hildebrandt et al. (2020)'s finding of the negligible influence of neutrinos on cosmic shear analyses. The second category encompasses three nuisance parameters, accounting for astrophysical and measurement uncertainties as previously discussed. We note that all parameters, with the exception of T_{AGN} and A_{IA}, retain the same priors as those used in the KiDS-1000-v1 cosmic shear analyses. The T_{AGN} parameter replaces the previous baryon feedback amplitude parameter associated with the preceding version of HMCODE, while the A_{IA} parameter adopts a narrower prior for reasons previously discussed.

4.5 Results

In this section, we present our cosmological parameter constraints and evaluate the robustness of our findings against a variety of systematic uncertainties. We begin by presenting the outcomes from our fiducial analysis in Sect. 4.5.1. We then assess the impact of shear biases in Sect. 4.5.2, by quantifying the shifts in final constraints resulting from different shear bias scenarios. This highlights the main development of our work. Additionally, since we implemented several changes to the cosmological inference pipeline, we evaluate the effects of these adjustments by comparing results from multiple setup variations in Sect. 4.5.3.

4.5.1 Fiducial analysis results

Our fiducial model has a total of twelve free parameters: five are cosmological parameters specifying the spatially flat ΛCDM model with a fixed total neutrino mass, and the remaining seven are nuisance parameters addressing astrophysical and redshift calibration uncertainties, as detailed in Sect. 4.4. However, not all of these parameter are constrained by the cosmic shear analysis. In this section, we focus on the primary parameters that our analysis constrains. Meanwhile, the posterior distributions for all free parameters are displayed as contour plots in Appendix 4.D for reference.

Table 4.3 provides the point estimates along with their corresponding 68% credible intervals for the primary parameter as constrained by our fiducial analysis using the PolyChord sampling code. We display results using three summary statistics: MAP and PJ-HPD, the mean of the 1D marginal posterior, and the maximum of the 1D marginal. As discussed in DK23, each of these approaches has its own advantages and limitations. Specifically, the accurate determination of MAP and PJ-HPD can be challenging, while marginal constraints for multi-dimensional posteriors are prone to projection effects. Aligning with the KiDS convention, we choose the MAP and PJ-HPD constraints as our headline results, but caution against direct comparisons with results from other surveys that might use different summary statistics. The uncertainties we report include additional contributions from the systematic uncertainties associated with our shear calibration, as detailed in Sect. 4.5.2. These additional uncertainties are overall small compared to the main sampling uncertainties, so when plotting the posterior distributions or conducting extended runs for test purposes, we do not incorporate these uncertainties.

Figure 4.4 shows the projected 2D posterior distributions for the parameters Ω_m and S_8, as derived from our fiducial setups employing PolyChord and MultiNest. We see that MultiNest results yield a roughly 10% narrower width of the posterior distribution compared to PolyChord, aligning with previous findings (Lemos et al. 2023; DK23; Li et al. 2023c). However, as expected, the results from the two sampling codes show consistency in terms of best-fit values. In addition, we compare these results with those from the cosmic microwave background (CMB) analysis by the *Planck* satellite, using their baseline ΛCDM chains with the `Plik` likelihood from their most recent *Planck*-2018 results (Planck Collaboration et al. 2020). An offset is evident between

Table 4.3: Primary parameter constraints from our fiducial analysis, based on the KiDS-1000-v2 catalogue, as determined using the PolyChord sampling code.

Parameter	MAP & PJ-HPD	Marginal	
		Mean	Maximum
S_8	$0.776^{+0.029+0.002}_{-0.027-0.003}$	$0.765^{+0.029}_{-0.023}$	$0.769^{+0.027}_{-0.029}$
Ω_m	$0.259^{+0.115+0.001}_{-0.064-0.001}$	$0.302^{+0.062}_{-0.115}$	$0.273^{+0.102}_{-0.088}$
σ_8	$0.835^{+0.151+0.002}_{-0.158-0.003}$	$0.791^{+0.124}_{-0.163}$	$0.752^{+0.173}_{-0.129}$
A_IA	$0.348^{+0.350+0.009}_{-0.322-0.011}$	$0.400^{+0.330}_{-0.339}$	$0.397^{+0.340}_{-0.346}$

Our headline results, based on the MAP & PJ-HPD statistics, include additional uncertainties that account for systematic uncertainties within the shear calibration. These uncertainties, originating from minor deviations from realism in the image simulations and the shear measurement algorithm's sensitivity to the morphology of the galaxy sample, are estimated using a forward modelling approach (as detailed in Sect. 4.5.2). On the other hand, the statistical uncertainties within the shear calibration, determined by the simulation volume, are folded into the main uncertainties through their inclusion in the covariance matrix used for the cosmological inference. The mean-marginal is determined through `postprocess` within CosmoSIS using the default settings (Zuntz et al. 2015); while the max-marginal is calculated using the `ChainConsumer` with the settings of statistics='max' and kde=1.0 (Hinton 2016). The indicated uncertainties correspond to the 68% credible intervals.

our cosmic shear results and those from *Planck*-2018. Adopting the Hellinger distance tension metric (Beran 1977; Heymans et al. 2021; DK23), we detected a 2.35σ tension in the constrained S_8 values. For the constrained parameter set (S_8, Ω_m), a similar level of tension, 2.30σ, was found using the Monte Carlo exact parameter shift method (Raveri et al. 2020; DK23).

Figure 4.5 presents our primary S_8 constraints and compares them with those from other contemporary cosmic shear surveys and the *Planck* CMB analysis. For ease of comparison, we show all three summary statistics for our fiducial results, while for other surveys, we display their headline values, as per their preferred summary statistics. Overall, our results align well with those from all major contemporary cosmic shear surveys.

We note that our fiducial analysis pipeline is similar to the DK23 Hybrid pipeline with one notable difference: while DK23 included a free neutrino parameter, we kept the total neutrino mass fixed. DK23 showed that this additional degree of freedom in the cosmological parameter space can slightly increase the projected marginal S_8 values relative to an analysis with a fixed neutrino mass. However, since we refer to their

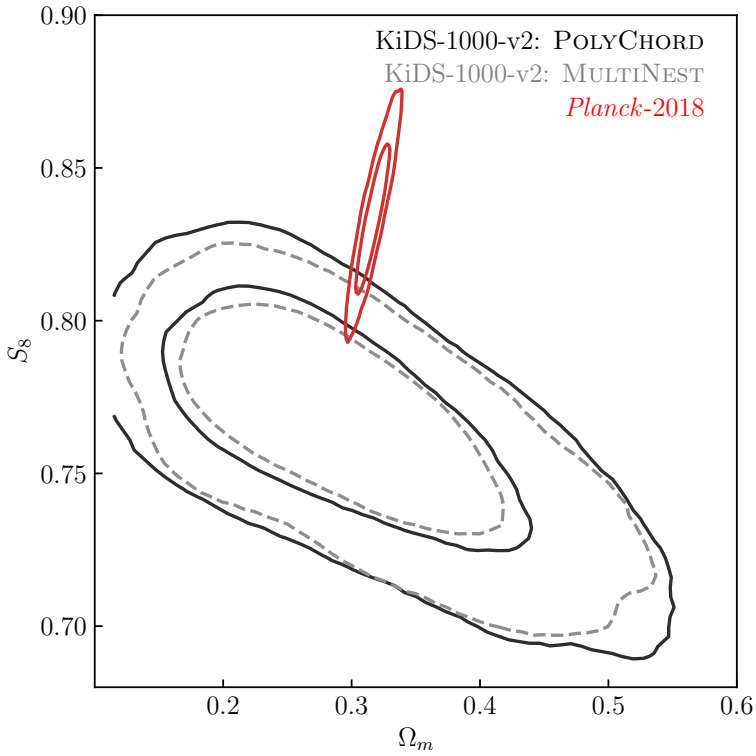

Figure 4.4: Comparison of projected 2D posterior distributions for the parameters Ω_m and S_8 as derived from our fiducial setups using two sampling codes: PolyChord (solid black line) and MultiNest (dashed grey line), against the *Planck*-2018 results (solid red line). The contours correspond to the 68% and 95% credible intervals and are smoothed using a Gaussian Kernel Density Estimation (KDE) with a bandwidth scaled by a factor of 1.5, made possible by the `ChainConsumer` package (Hinton 2016).

MAP & PJ-HPD results in Fig. 4.5, the comparison should not be influenced by these projection effects (for more details, refer to the discussion in DK23).

It is interesting to note that our fiducial results align almost identically with the KiDS-1000-v1 re-analysis conducted by DK23, which used the A21 redshift calibration. This alignment arises from a balance of several effects in our analysis. Our improved shear calibration tends to increase S_8, while the enhanced vdB22 redshift calibration tends to lower it. Moreover, our analysis does not show a significant increase in S_8 when introducing scale cuts, as seen in the KiDS-1000-v1 Hybrid analysis. This helps reconcile the minor difference between our results and those of A21. We explore these changes in more detail in the following sections.

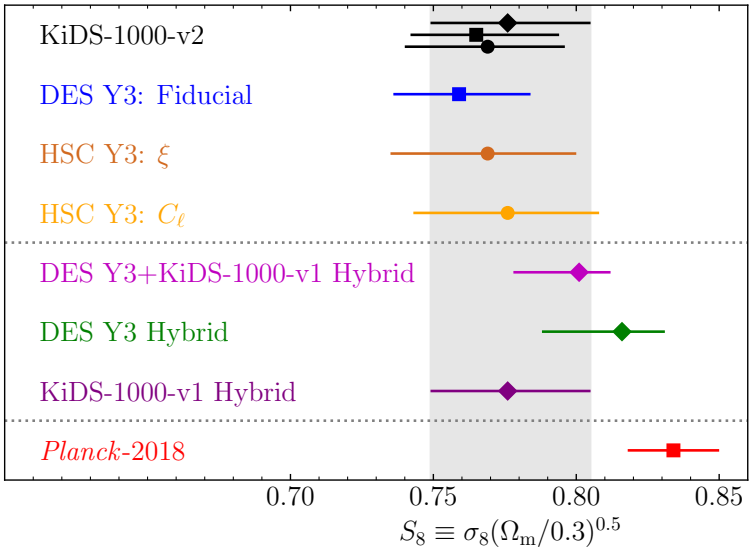

Figure 4.5: Marginalised constraints on S_8 derived from our fiducial analysis with PolyChord, compared with those from other contemporary cosmic shear surveys and the *Planck* CMB analysis. Three sections, separated by dotted horizontal lines, indicate results of different origins. The first section includes results from individual cosmic shear surveys with their own analysis pipelines. The second section presents results from a collaborative effort between DES and KiDS teams, which built a hybrid pipeline for analysing the data from both groups (DK23). The final section displays results from the *Planck* CMB analysis. Different labels are used for different statistical methods: the diamond represents results using the MAP & PJ-HPD statistics, the square denotes the mean-marginal statistics, and the circle shows the maximum-marginal statistics. The error bars correspond to the 68% credible intervals.

4.5.2 Impact of shear biases

The primary aims of this study are to assess the impact of higher-order shear biases on the final parameter constraints and to develop a methodology for effectively addressing shear calibration uncertainties. Both of these aims can be achieved by examining the shifts in the constrained cosmological parameters resulting from different shear bias scenarios. As discussed in Sect. 4.4 and Appendix 4.C, the residual shear biases have only a minor effect on the measured data vector. This allows us to determine the shifts in the best-fit values of the constrained parameters using a local minimisation algorithm, such as the Nelder-Mead method (Nelder & Mead 1965). These shifts in the best-fit values indicate the additional uncertainties stemming from systematic uncertainties in shear calibration.

Figure 4.6 shows shifts in our primary S_8 constraints for different residual shear bias scenarios. For comparison, we also include a shaded region denoting different levels

of PJ-HPD credible intervals, as derived from our fiducial PolyChord chain. Apart from the extreme case where no shear calibration is applied, all other residual shear bias scenarios result in shifts less than 10 per cent of the initial sampling uncertainties. Notably, neglecting the higher-order correction for the shear-interplay effect and uncertainties in PSF modelling results in a negligible shift of only -0.03σ. This finding reinforces the reliability of previous KiDS cosmic shear analyses, which did not consider these higher-order effects.

The S_8 shifts, resulting from the input morphology test simulations, indicate additional systematic uncertainties within our shear bias calibration. The generation of these test simulations is detailed in Appendix 4.C. Briefly, we generated six sets of test simulations, where the input values of three morphological parameters of the adopted Sérsic profile - the half-light radius (labelled as 'size' in the figure), axis ratio (labelled as 'q'), and Sérsic index (labelled as 'n') - were shifted up and down. We observe that shifts in the input galaxy axis ratio lead to the most significant changes in S_8: a -0.10σ shift for increased input axis ratio and a $+0.06\sigma$ shift for decreased input axis ratio. This behaviour aligns with our expectations for the *lens*fit code employed in our analysis. As it incorporates prior information on measured galaxy ellipticities during its Bayesian fitting process, it is more sensitive to changes in the distributions of sample ellipticities.

These S_8 shifts, obtained from the test simulations, provide a quantitative measure of the potential impact of inaccuracies in the input morphology and the sensitivity of the *lens*fit code to the underlying sample morphology distributions. When presenting the S_8 constraints, we account for these systematic uncertainties by including the maximum shifts into the reported uncertainties. In other words, we consider the shifts corresponding to the changes in input axis ratio (represented as dashed lines in Fig. 4.6), from the six sets of test simulations, as additional systematic uncertainties. These are reported alongside the original statistical uncertainties from the main sampling chain. It should be noted that these additional systematic uncertainties are specific to the SKiLLS image simulations and the *lens*fit shape measurement code used in our analysis. To reduce these uncertainties, future advancements in shear measurements should focus on improving the realism of image simulations and enhancing the robustness of the shear measurement algorithm.

4.5.3 Impact of altering inference setups

Although our main updates revolve around the shear measurement and calibration, we have also implemented several modifications to the cosmological inference pipeline, drawing upon recent developments from DK23. As such, it is beneficial to conduct some extended runs with various setup configurations.

For these test runs, we employ MultiNest as our sampling code, as it operates approximately five times faster than PolyChord, but at the cost of underestimating the width of the posterior distributions and thus the reported uncertainties by about 10%. However, the best-fit values from MultiNest are not biased (as evident in Fig. 4.4). Thus, comparisons made using MultiNest will yield conservative but unbiased results.

Table 4.4: Point estimates for S_8 from different inference setups.

Setups	MAP & PJ-HPD			Mean marginal		Maximum marginal	
	χ^2	S_8	ΔS_8	S_8	ΔS_8	S_8	ΔS_8
Fiducial: PolyChord	62.7	$0.776^{+0.029+0.002}_{-0.027-0.003}$	-	$0.765^{+0.029}_{-0.023}$	-	$0.769^{+0.027}_{-0.029}$	-
Fiducial: MultiNest	62.7	$0.774^{+0.027}_{-0.023}$	0.00σ	$0.765^{+0.028}_{-0.017}$	0.00σ	$0.771^{+0.023}_{-0.027}$	0.00σ
NLA: $A_{\rm IA}$: $[-6, 6]$	62.7	$0.774^{+0.034}_{-0.020}$	0.02σ	$0.766^{+0.028}_{-0.020}$	0.03σ	$0.770^{+0.023}_{-0.026}$	-0.01σ
NLA-z: $A_{\rm IA}, \eta_{\rm IA}$: $[-5, 5]$	62.7	$0.776^{+0.022}_{-0.025}$	0.10σ	$0.765^{+0.027}_{-0.020}$	0.00σ	$0.770^{+0.023}_{-0.025}$	-0.05σ
Scales: $0'.5 < \theta < 300'$	64.5	$0.769^{+0.014}_{-0.036}$	-0.17σ	$0.755^{+0.029}_{-0.022}$	-0.40σ	$0.763^{+0.022}_{-0.030}$	-0.31σ
KiDS-1000-v1 setups	64.9	$0.771^{+0.017}_{-0.029}$	-0.13σ	$0.755^{+0.026}_{-0.020}$	-0.44σ	$0.759^{+0.022}_{-0.024}$	-0.50σ

The 'Fiducial: PolyChord' denotes our headline results, which are the same as those presented in Table 4.3. The 'Fiducial: MultiNest' represents the same parameter setup as our fiducial analysis, but employs the MultiNest sampling code. We use this as the reference to assess test results because all test runs utilise the MultiNest code for increased speed. When comparing the 'Fiducial: MultiNest' results with the primary PolyChord results, we can conclude that the best-fit values obtained from MultiNest are unbiased. The relative shift in S_8, denoted as ΔS_8, is calculated by comparing the best-fit values from the test runs to the reference 'Fiducial: MultiNest' run. The ΔS_8 values are expressed as a fraction of σ, which signifies the standard deviation of estimates from the test run. We calculate ΔS_8 for different summary statistics separately for consistency. For MAP & PJ-HPD results, we also present the best-fit χ^2 values. For comparison, the best-fit χ^2 values from A21 and vdB22 are 82.2 and 63.2, respectively.

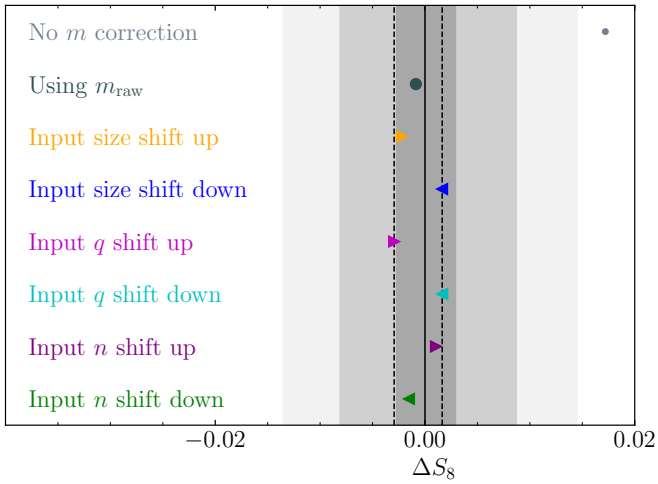

Figure 4.6: Shifts in best-fit values of S_8 under different residual shear bias scenarios. The shift, ΔS_8, is calculated as $\Delta S_8 = S_8^{\rm test} - S_8^{\rm fiducial}$, where $S_8^{\rm test}$ represents the best-fit values in the test scenarios determined by a local minimisation method using the best-fit values from the fiducial analysis ($S_8^{\rm fiducial}$) as a starting point. The grey shaded regions represent different percentiles of the credible intervals derived from our fiducial PolyChord run. From the innermost to the outermost region, these percentiles are 6.8%, 20.4%, and 34%, corresponding to 0.1, 0.3, and 0.5 fractions of the reported sampling uncertainties. The dashed lines display the maximum shifts encountered in the six sets of morphology test simulations. These maximum shifts are used as the additional uncertainties in the reported best-fit values to account for the systematic uncertainties arising from shear calibration.

4

Priors for the NLA model

We begin by testing the prior for the NLA model. As discussed in Sect. 4.4, our fiducial analysis implemented a redshift-independent NLA model with a narrow flat prior for the amplitude parameter $A_{\rm IA}$. This model, motivated by the work of Fortuna et al. (2021a), serves as an alternative to the uninformative broad prior previously used. To investigate the impact of this change on our final results, we performed two additional runs: one employing a redshift-independent NLA model with a broad $A_{\rm IA}$ prior ranging from $[-6, 6]$, in line with KiDS-1000-v1 analyses, and another allowing for a redshift-dependent IA amplitude, i.e., the NLA-z variant. The redshift evolution is modelled using a power-law of the form $[(1+z)/(1+0.62)]^{\eta_{\rm IA}}$, with priors of $[-5, 5]$ for both $A_{\rm IA}$ and $\eta_{\rm IA}$, in line with DK23.

Figure 4.7 presents a comparison of the posterior distributions obtained from the different NLA prior setups, and Table 4.4 lists the point estimates for the critical S_8 parameter. We see consistent constraints on S_8 across all setups. The constrained $A_{\rm IA}$ under our narrower prior setup also aligns with those from the broad priors, albeit

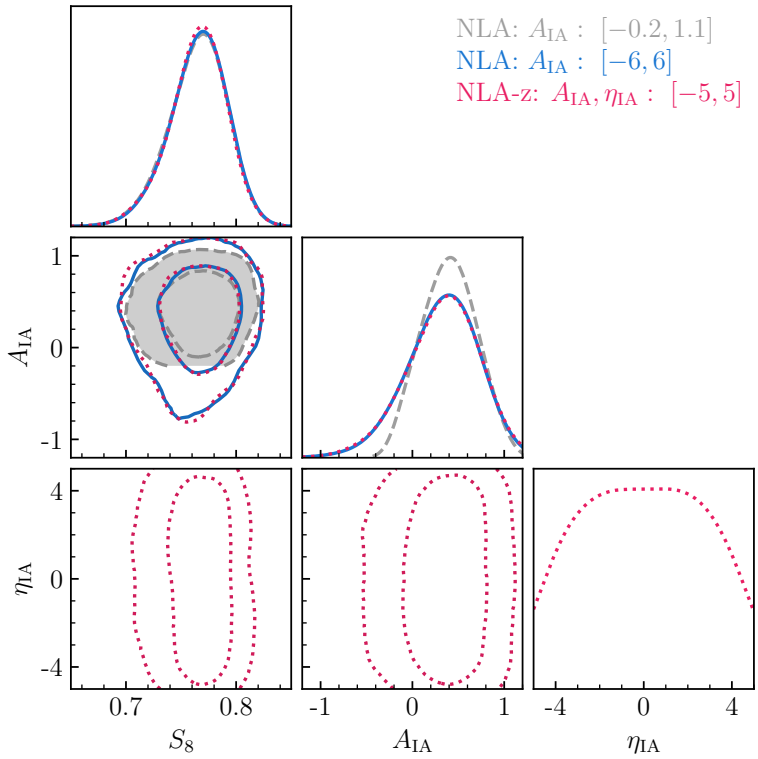

Figure 4.7: Comparison of projected posterior distributions for the parameters S_8, $A_{\rm IA}$, and $\eta_{\rm IA}$, derived from three different NLA prior setups. The contours correspond to the 68% and 95% credible intervals and are smoothed using Gaussian KDE with a bandwidth scaled by a factor of 1.5.

spanning a narrower range due to the constrained prior range, validating the prior range used in our fiducial analysis. Additionally, we observe that the $\eta_{\rm IA}$ parameter is not constrained by the data, suggesting that the use of the NLA-z model may not be necessary for current weak lensing analyses.

Different scale cuts

In our fiducial analysis, we adopted a scale cut for the measured data vectors, ranging from $2'$ to $300'$, as suggested by DK23. This is a change from the KiDS-1000-v1 analyses, which used a range of $0.'5 < \theta < 300'$. A re-analysis of KiDS-1000-v1 with this new scale cut by DK23 led to a $0.7 - 0.8\sigma$ increase in the S_8 constraint. Using mock analyses, they found that this offset could arise from noise fluctuations 23% of the time.

In light of the updates to our shear measurement, we revisited this test. Interestingly, we observed a smaller difference between the two scale cuts than what was reported by DK23. Specifically, we observed shifts of -0.17σ, -0.40σ, and -0.31σ, corresponding

to the MAP & PJ-HPD, mean marginal, and maximum marginal summary statistics, respectively (refer to Table 4.4 for exact values).

We attribute this increased robustness against small scale fluctuations to our improved empirical corrections of the PSF leakages into shear measurement. This is supported by Figs. 4.2 and 4.3, where we see that the shear signals measured from the KiDS-1000-v2 catalogues exhibit overall smaller systematic errors. We note that Giblin et al. (2021) performed a mock test using the two-point correlation function and identified a change of less than 0.1σ in the S_8 constraints when the detected PSF residuals were incorporated into the KiDS-1000-v1 mock data. Nevertheless, it is plausible that these systematic effects have a more significant influence on COSEBIs, given their use of more sophisticated weighting functions (Schneider et al. 2010). To quantify the improvements brought about by the updated shear measurements regarding the robustness of the COSEBIs, a similar mock analysis based on the COSEBIs statistic is warranted. We consider this an important topic for future study. For the current analysis, the test results simply affirm the robustness of our primary S_8 constraints.

KiDS-1000-v1 setups

To draw a direct comparison with the KiDS-1000-v1 results and evaluate the impact of our improved shear measurements and calibration, we performed a test run using the same inference pipeline and parameter priors as in the KiDS-1000-v1 analyses conducted by A21 and vdB22. The differences compared to our fiducial analysis setup include: measurements from scales of 0.′5 to 300′, use of the older version of HMCODE, sampling with the MULTINEST code, and a broad $A_{\rm IA}$ prior ranging from $[-6, 6]$. As shown in Fig. 4.8, our test results are well-aligned with the outcomes of the analyses by A21 and vdB22. Notably, our new results show an increase in the S_8 value relative to vdB22, bringing it closer to the result obtained by A21.

We re-emphasise that our redshift calibration aligns with that of vdB22, who expanded the redshift calibration sample to more than double the size used by A21 (see Appendix 4.B for details). This means that our redshift-related selection function closely mirrors that used in the vdB22 sample. However, due to changes in the weighting and selection scheme between the KiDS-1000-v2 catalogue and the KiDS-1000-v1 catalogue, our sample cannot be considered as directly comparable to theirs.

To provide a more quantitative understanding of the sample differences among the three analyses, we compared the effective number density of the source sample in our analysis to those used in A21 and vdB22. The differences for each tomographic bin are 9.6%, 9.8%, 6.1%, 10.6%, and 2.8% when compared to A21; and −1.8%, −1.3%, −0.7%, 0.7%, and 3% when compared to vdB22. Here, positive values signify an increase, while negative values denote a decrease. The differences between our catalogue and that of A21 stem from both shear measurement and redshift calibration, whereas the difference between ours and that of vdB22 arises mainly from the shear measurement, as we used the same SOM for the 'gold' selection (see Appendix 4.B). As such, comparing our results directly with those of vdB22 can provide clearer insights into the impact of

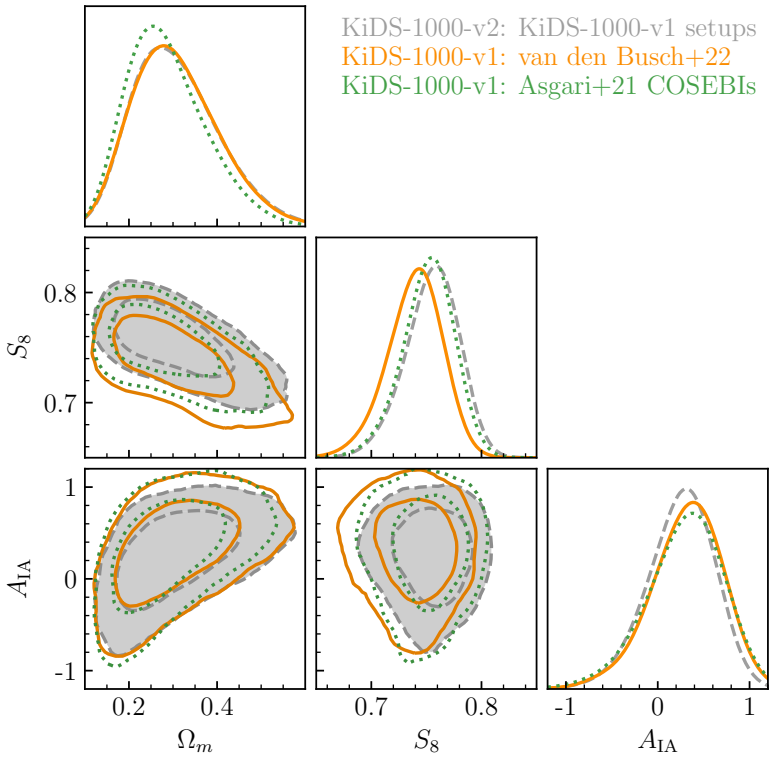

Figure 4.8: Comparison of projected posterior distributions for parameters $\Omega_{\rm m}$, S_8, and $A_{\rm IA}$ from our analysis (grey dashed lines) based on the KiDS-1000-v2 catalogue, to those from vdB22 (orange solid lines) and A21 (green dotted lines), both of which are based on the KiDS-1000-v1 catalogues. The cosmological inference pipeline and parameter priors are identical across all three analyses presented here. In terms of measurements, vdB22 and A21 use the same shear measurements and calibration, while vdB22 and our analysis share the same redshift calibration. The contours correspond to the 68% and 95% credible intervals and are smoothed using Gaussian KDE with a bandwidth scaled by a factor of 1.5.

our improvements in shear measurements. It is also worth noting that the increased effective number density in high redshift bins compared to vdB22 is largely due to the increased weighting of faint objects in the updated version of *lens*fit code. However, this comes at the cost of increased sample ellipticity dispersion, with a maximum increase of 6% found in the fifth bin. These subtle differences in the source catalogues change the noise properties of the samples. Consequently, even with perfect calibration in each study, we would not expect to derive identical cosmological constraints from each analysis.

4.6 Summary

We conducted a cosmic shear analysis using the KiDS-1000-v2 catalogue, which is an updated version of the public KiDS-1000(-v1) catalogue with respect to shear measurements and calibration. Under the assumption of a spatially flat ΛCDM cosmological model, we derived constraints on $S_8 = 0.776^{+0.029+0.002}_{-0.027-0.003}$ based on the MAP & PJ-HPD summary statistics. The second set of uncertainties were incorporated to account for the systematic uncertainties within our shear calibration. The mean-marginal and maximum-marginal values obtained from the same sampling chain are $0.765^{+0.029}_{-0.023}$ and $0.769^{+0.027}_{-0.029}$, respectively. Our results are consistent with earlier results from KiDS-1000-v1 and other contemporary weak lensing surveys, but show a $\sim 2.3\sigma$ level of tension with the *Planck* cosmic microwave background constraints.

The main improvements in our analysis, relative to the KiDS-1000-v1 cosmic shear analyses, are attributed to the enhanced cosmic shear measurement and calibration. These enhancements were achieved through the updated version of the *lens*fit shape measurement code, a new empirical correction scheme for PSF contamination, and the newly developed SKiLLS multi-band image simulations, as detailed in L23. We verified the reliability of the new measurement via a series of catalogue-level null tests proposed by Giblin et al. (2021). The results indicate that the KiDS-1000-v2 catalogue shows overall better control over measurement systematics compared to the KiDS-1000-v1 catalogues. This improvement in reducing measurement systematics assists in reducing noise in small scale measurements, thereby enhancing the robustness of our cosmological parameter constraints against varying scale cut choices.

Our methodology for shear calibration largely aligns with the one detailed in L23, where we account for higher-order blending effects that arise when galaxies from different redshifts are blended, as well as the uncertainties in PSF modelling. However, when comparing the outcomes from the shear calibration with and without these higher-order adjustments, we found that these effects have a negligible impact on the present weak lensing analysis, a conclusion that is in line with the findings of Amon et al. (2022).

We recommend treating the statistical and systematic uncertainties from the shear calibration separately, given their distinct origins. The statistical uncertainties, which are determined by the simulation volume, can be reduced and are readily incorporated into the covariance matrix used for cosmological inference. On the other hand, systematic uncertainties, associated with the realism of image simulations and sensitivity of the shape measurement algorithm, can be more effectively addressed when considered as residual shear biases post-calibration. Assuming these residual shear biases are small, a forward modelling approach, combined with a local minimisation method, can be used to estimate their impact on the final parameter constraints. In our analysis, these additional systematic uncertainties contribute roughly 8% of the final uncertainty on S_8. However, ongoing efforts to enhance shear measurement and calibration, such as increasing the realism of image simulations through Monte-Carlo Control Loops (Refregier & Amara 2014) and leveraging new techniques like Metacalibration/Metadetection (Huff & Mandelbaum 2017; Sheldon & Huff 2017; Sheldon et al. 2020; Hoekstra et al. 2021) to

improve measurement robustness against underlying sample properties, may well lead to a reduction in these additional systematic uncertainties.

In our fiducial analysis, we opted for a redshift-independent NLA model with a narrow flat prior for the IA amplitude parameter, $A_{\rm IA}$, motivated by the work of Fortuna et al. (2021a). However, we also investigated two alternative scenarios: one with a broad $A_{\rm IA}$ prior for the redshift-independent NLA model, echoing the KiDS-1000-v1 analysis by A21, and the other, the NLA-z variant, allowing for redshift evolution of the IA amplitude, as per the recent joint DES Y3+KiDS-1000 cosmic shear analysis (DK23). In all three scenarios, we found fully consistent constraints for S_8 and $A_{\rm IA}$, indicating that the impact of the variations among these scenarios is negligible. To better understand the IA signals and their impact on cosmic shear analyses, future tests need to implement more substantial variations in IA models, for instance, the halo model formalism introduced by Fortuna et al. (2021a). Such exploration would not only enhance our understanding of the measured IA signals, but also help mitigate correlations between nuisance parameters, thereby improving the precision of future cosmic shear analyses.

Acknowledgements

We acknowledge support from: the Netherlands Research School for Astronomy (SSL); the Netherlands Organisation for Scientific Research (NWO) under Vici grant 639.043.512 (HHo); the Royal Society and Imperial College (KK); the Polish National Science Center through grants no. 2020/38/E/ST9/00395, 2018/30/E/ ST9/00698, 2018/31/G/ST9/03388 and 2020/39/B/ST9/03494 (MB); the Polish Ministry of Science and Higher Education through grant DIR/WK/2018/12 (MB); the Royal Society through an Enhancement Award (RGF/EA/181006) and the Royal Society of Edinburgh for support through the Saltire Early Career Fellowship with ref. number 1914 (BG); the European Research Council (ERC) under Grant number 647112 (CH) and Consolidator Grant number 770935 (HHi, JLvdB, AHW, RR); the Max Planck Society and the Alexander von Humboldt Foundation in the framework of the Max Planck-Humboldt Research Award endowed by the Federal Ministry of Education and Research (CH); the UK Science and Technology Facilities Council (STFC) under grant ST/V000594/1 (CH) and ST/V000780/1 (BJ); Heisenberg grant of the Deutsche Forschungsgemeinschaft grant Hi 1495/5-1 (HHi); CMS-CSST-2021-A01 and CMS-CSST-2021-A04, NSFC of China under grant 11973070 (HYS); Key Research Program of Frontier Sciences, CAS, Grant No. ZDBS-LY-7013 (HYS); Program of Shanghai Academic/Technology Research Leader (HYS). The results in this paper are based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs: 088.D-4013, 092.A-0176, 092.D-0370, 094.D-0417, 177.A-3016, 177.A3017, 177.A-3018 and 179.A-2004, and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA and NWO-M grants; Target; the University of Padova, and the University Federico II (Naples). Contributions to the data processing for VIKING were made by the VISTA

Data Flow System at CASU, Cambridge and WFAU, Edinburgh.
Author Contributions: All authors contributed to the development and writing of this paper. The authorship list is given in three groups: the lead authors (SSL, HHo, KK) followed by two alphabetical groups. The first alphabetical group includes those who are key contributors to both the scientific analysis and the data products. The second group covers those who have either made a significant contribution to the data products, or to the scientific analysis.

4.A Shear bias in two-point statistics

When calibrating the shear measurements in the two-point correlation function, it is usually assumed that the correlations involving the shear bias can be ignored, which includes correlations between different tomographic and spatial angular bins. This simplification leads to the following relationship between the true correlation function of cosmic shear in tomographic bins i, j, denoted as ξ^{ij}, and the measured signal $\hat{\xi}^{ij}$:

$$\begin{aligned}\hat{\xi}^{ij}(\theta) &= \langle \hat{\gamma}^i(\theta')\, \hat{\gamma}^j(\theta'+\theta)\rangle \\ &= \langle [1+m^i(\theta')]\, [1+m^j(\theta'+\theta)]\, \gamma^i(\theta')\, \gamma^j(\theta'+\theta)\rangle \\ &= \langle [1+m^i(\theta')]\, [1+m^j(\theta'+\theta)]\rangle\, \xi^{ij}(\theta) \\ &\simeq (1+\overline{m^i})\, (1+\overline{m^j})\, \xi^{ij}(\theta)\,,\end{aligned} \tag{4.4}$$

where $\overline{m^i}$ is estimated by averaging over all sources in a given tomographic bin i, and we use $\langle\cdot\rangle$ to denote the correlation function. We also assumed that the shear bias is independent of the underlying shear to simplify the equation. The result of Eq. (4.4) allows us to average the multiplicative biases over all the galaxies in a given tomographic bin to mitigate the individual noisy bias estimation.

However, in principle, the shear bias can be scale dependent due to spatial fluctuations in source density (e.g. Samuroff et al. 2018). With SKiLLS, we can directly examine these correlations by measuring the shear bias in the two-point estimators. We measured the shear correlation function in the SKiLLS mock catalogue using Eq. (4.2). Since we know the true $\xi_+^{ij}(\theta) = \gamma^2_{\rm input}$ in simulations, where $\gamma_{\rm input}$ is the amplitude of the constant input shear, we can estimate the shear bias in the two-point correlation function directly by comparing the measured $\hat{\xi}_+^{ij}$ to the input ξ_+^{ij} following Eq. (4.4).

Figure 4.9 shows the difference between the shear biases with and without considering its correlations, defined as $\Delta m_\xi \equiv \langle [1+m^i(\theta')]\, [1+m^j(\theta'+\theta)]\rangle - (1+\overline{m^i})\, (1+\overline{m^j})$. It shows that the difference is negligible across all scales and tomographic bins, in agreement with the statistical uncertainties of our shear calibration, which are represented by the shaded regions. These findings confirm that we can neglect the correlations between shear biases in the current KiDS weak lensing analysis.

4.B SOM redshift calibration

This appendix provides information on the redshift calibration reference sample and SOM configurations used in our analysis. For a more comprehensive overview and validation of the SOM redshift calibration method in the KiDS analysis, we refer to Wright et al. (2020b), Hildebrandt et al. (2021) and vdB22.

We employed the fiducial spectroscopic sample described in vdB22 as our calibration reference sample. This sample comprises spectroscopic redshift estimates (spec-zs) from various spectroscopic surveys that overlap with KiDS fields, enabling us to assign KiDS photometric measurements to objects in the reference sample. In cases where an

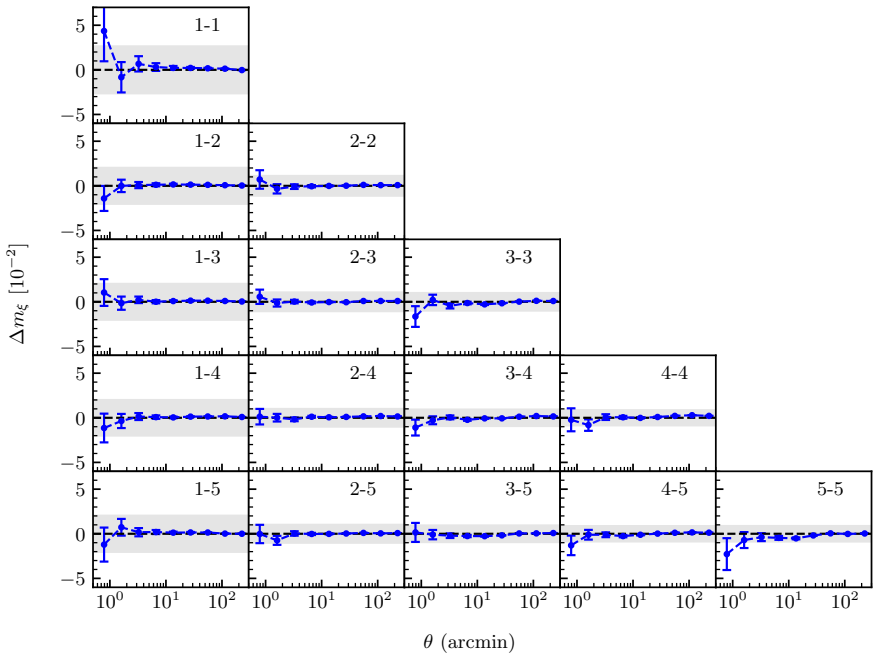

Figure 4.9: Two-point correlations between the multiplicative shear biases. The correlation is estimated as $\Delta m_\xi \equiv \hat{\xi}_+^{ij}/\gamma^2_{\rm input} - (1+\overline{m^i})\ (1+\overline{m^j})$. The 15 panels represent the different combinations of the five redshift bins utilised in our cosmic shear analysis. The shaded regions within each panel denote the statistical uncertainties of our shear calibration for each tomographic bin, as outlined in Table 4.1.

4

object had multiple spectroscopic measurements, vdB22 defined a specific hierarchy to select the most reliable redshift estimates based on the quality of the measurements. For further details on the adopted spectroscopic samples and the compilation procedure, readers are referred to Appendix A of vdB22.

For our calibration, we used a 101×101 hexagonal SOM trained on the r-band magnitude and 36 colours derived from the PSF-matched, list-driven nine-band $ugriZYJHK_s$ photometry from the KiDS+VIKING surveys. This SOM is identical to the fiducial SOM constructed in vdB22. We segregated the reference and target samples into the trained SOM cells separately for each tomographic bin, allowing us to create comparable groupings between the spectroscopic and photometric sources in each bin. During this process, we further categorised the original SOM cells using a hierarchical cluster analysis implemented by the 'hclust' function within the R Stats Package[12] to increase the number of galaxies per grouping. We adopted the same number of clusters per

[12] https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/hclust

bin (4000, 2200, 2800, 4200, and 2000) as Wright et al. 2020b, who determined these numbers using simulations produced by van den Busch et al. (2020).

To mitigate the effects of photometric noise and the incompleteness of the reference sample, we applied an additional selection step to the SOM groupings. We excluded any grouping where the mean spectroscopic redshift of the reference sample $\overline{z_{\rm spec}}$ and the mean photometric redshift of the target sample $\overline{z_{\rm B}}$ exhibited a significant discrepancy, defined as $|\overline{z_{\rm spec}} - \overline{z_{\rm B}}| > 5\sigma_{\rm mad}$. Here, $\sigma_{\rm mad}$ represents the normalised median absolute deviation of all SOM groupings, which we calculated to be 0.122 in our case. This step allowed us to define the KiDS 'gold' sample, which we used to compute the redshift distributions and perform the cosmic shear analysis.

4.C Systematic uncertainties from the shear calibration

In this appendix, we outline our approach to address the systematic uncertainties arising from shear calibration. Our methodology involves two primary steps: In Sect. 4.C.1, we quantify the potential residual biases after implementing our simulation-based shear calibration. In Sect. 4.C.2, we propagate these systematic uncertainties into the final uncertainties of the estimated cosmological parameters.

We propose a separate accounting of the shear calibration uncertainties, as it is considered more accurate and informative than the traditional approach, which uses nominal shear calibration uncertainties that are deliberately overestimated to encompass potential systematic uncertainties arising from shear calibration. Our approach clearly illustrates the extent to which the final cosmological parameters of interest are influenced by these systematic uncertainties from shear calibration.

Furthermore, as mentioned in Sect. 4.3.2, these systematic uncertainties have fundamentally different origins from the statistical uncertainties incorporated in the covariance matrix. They represent the fundamental limitations of current simulation-based shear calibration methods. The limitations inherent in these systematic uncertainties cannot be eliminated by merely increasing the scale of image simulations. However they can be mitigated by empirically enhancing the realism of the image simulations, for example, using the Monte-Carlo Control Loop method (Refregier & Amara 2014), or by improving the robustness of the shear measurement algorithm, such as the METACALIBRATION/METADETECTION method (Huff & Mandelbaum 2017; Sheldon & Huff 2017; Sheldon et al. 2020; Hoekstra et al. 2021).

4.C.1 Quantifying residual shear biases with sensitivity analysis

Residual biases may persist after simulation-based shear calibration due to imperfect alignment between simulations and data, as elucidated by K19. These discrepancies pose challenges for shear calibration methods dependent on image simulations and underscore the need for re-weighting simulations to more closely align with the data. However, given that intrinsic galaxy properties in real data are unknown, this re-weighting process relies on noisy measured properties, rendering it vulnerable to calibration selection

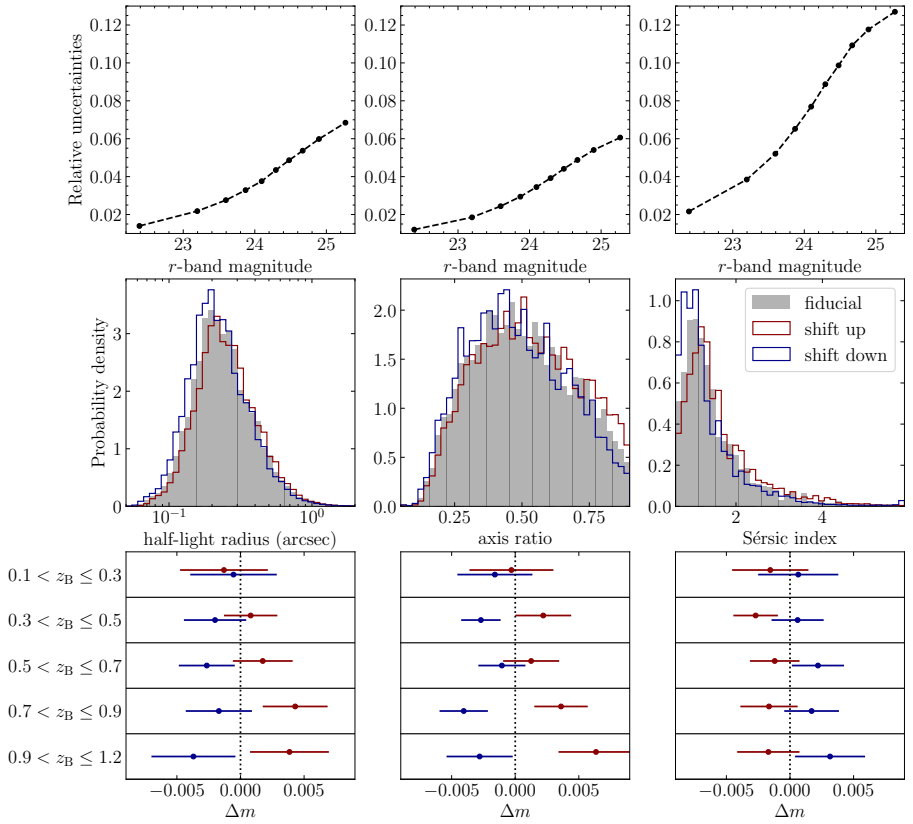

Figure 4.10: Changes in input morphological parameter values for assessing residual biases after simulation-based shear calibration. From left to right, the order is half-light radius, axis ratio, and Sérsic index. *Top panels*: The relationship between relative fitting uncertainties and the r-band magnitude, as measured from the catalogues of Griffith et al. (2012). The values shown are calculated using the median values of the binned samples. *Middle panels*: The overall distributions of input morphological parameters. Comparisons are drawn among the fiducial (grey shades), the test sample with input values increased by an amount corresponding to the relative uncertainties (red lines), and the test sample with input values decreased by the same amount of relative uncertainties (blue lines). *Bottom panels*: The shifts in multiplicative biases in tomographic bins resulting from changes in input morphological parameter values. Both fiducial and test simulations are re-weighted using the same data set, with the method detailed in Sect. 4.3.1.

biases as discussed by FC17. The uncertainties linked with the measured properties cause galaxies to be intermixed among defined bins, leading to the up-weighting or down-weighting of certain galaxies. As a result, even if the re-weighted sample aligns with the data in terms of the distribution of measured properties, it does not ensure identicality in terms of intrinsic properties. In other words, shear biases can still vary between two samples with identical distributions of apparent measured properties. Our aim is to quantify these residual biases and incorporate them into the final uncertainties of cosmological parameters.

The SKiLLS multi-band image simulations used in this analysis incorporate several enhancements, informed by insights gathered from previous KiDS simulation studies (FC17, K19). These improvements include: reproducing variations in star density, PSF, and noise background across the KiDS footprint; incorporating faint galaxies down to an r-band magnitude of 27 to account for correlated noise from undetected objects (e.g. Hoekstra et al. 2017); including realistic clustering from N-body simulations to address blending effects (e.g. K19); and adopting an end-to-end approach for photo-z estimation to account for photo-z measurement uncertainties. These improvements augment the robustness of the shear biases estimated from SKiLLS against various observational conditions.

In an investigation on the propagation of observational biases in shear surveys, Kitching et al. (2019) demonstrated that the measured shear power spectrum is, to first order, predominantly influenced by the mean of the multiplicative bias field across a survey. This suggests that if the shear bias estimated from simulations accurately reflects the mean value of the targeted sample, the shear calibration will be robust enough for KiDS-like cosmic shear analyses. Therefore, we conclude that potential residual biases related to observational conditions have negligible influence on our shear calibration, and we focus on systematic uncertainties arising from galaxy morphology uncertainties, specifically the assumed Sérsic profile and its parameters derived from *Hubble* Space Telescope observations (Griffith et al. 2012). For a model-fitting shape measurement code like *lens*fit, these galaxy morphology uncertainties are the main sources of residual shear biases after implementing the simulation-based shear calibration.

The deviation from the Sérsic profile is challenging to address for the current SKiLLS simulations, as our copula-based learning algorithm requires a parameterised model for its application. However, the Sérsic model has been validated as sufficient for KiDS-like analyses by K19, who used the same morphology catalogue as our work. Thus, we focus on the measurement uncertainties of the Sérsic parameters: half-light radius, axis ratio, and Sérsic index. We first examined the fitting uncertainties reported by Griffith et al. (2012) to assess the accuracy of these parameters in our input catalogue. We found that the median relative uncertainties for these parameters are a smooth function of galaxy magnitude, as shown in the top panels of Fig. 4.10. This allows us to capture these correlations through simple linear interpolation.

We interpreted these relative uncertainties as indicators of the systematic uncertainties in our input morphology. We assumed the most extreme scenarios, in which these measured statistical uncertainties are all caused by a coherent bias in the same direction.

Consequently, we adjusted all galaxies in our input sample in the same direction, with the amplitude of the adjustment determined based on their r-band magnitude using a simple linear interpolation of the measured median correlations. We examined shifts towards both larger and smaller values and considered the three Sérsic parameters separately. This resulted in six test simulations corresponding to the six different sets of variations in input morphology parameter values. The input parameter distributions for these test simulations, as shown in the middle panels of Fig. 4.10, are compared to the distributions of the fiducial simulations. A clear shift of the entire distribution is evident, suggesting that our test simulations represent the most extreme scenarios in which the measured statistical uncertainties are coherently biased in the same direction, a situation that is unlikely in reality. Therefore, the residual biases we identified from these test simulations provide a conservative estimate.

We applied the same data analysis procedures to the test simulations as we did to the fiducial simulations, including shear and redshift estimates. We also followed the same re-weighting procedure for the test simulations as for the fiducial simulations, ensuring that the calibration selection biases are also captured. The differences in shear biases between these test simulations and our fiducial simulation are illustrated in the bottom panels of Fig. 4.10. The small differences indicate that the residual shear biases, after implementing our fiducial shear bias calibration, are insignificant.

4.C.2 Propagating residual shear biases with forward modelling

Accurately incorporating the systematic uncertainties from shear calibration into the covariance matrix presents a challenge, as residual shear biases directly scale the data vector, as shown in Eq. (4.4). A more direct approach is to assess the shift in the measured shear signal caused by the residual shear biases and evaluate how these data vector shifts influence the constrained cosmological parameters. Given the minor residual shear biases illustrated in Fig. 4.10 and the unchanged covariance, it is not necessary to reiterate the sampling of the posterior distributions for each shift. Instead, we can implement a local minimisation algorithm to find nearby best-fit values for each shift, using starting points from the fiducial sampling chain. The range of these new best-fit values, each associated with a shift, indicates the additional systematic uncertainties introduced by the residual shear biases.

This approach naturally integrates with our existing cosmological inference method, as outlined in Sect. 4.4, which already requires an additional local optimisation step to refine the best-fit values identified by the sampling code. We simply replicated this optimisation step, using the original best-fit value as the starting point and the shifted likelihood to determine the best-fit values associated with various alterations in measured signals. The variability in these test best-fit values provides an expanded credible region for the inferred parameters, thereby representing the systematic uncertainties from shear calibration. We included these additional uncertainties when presenting the point estimates of our primary parameters (see Sect. 4.5.2 for details).

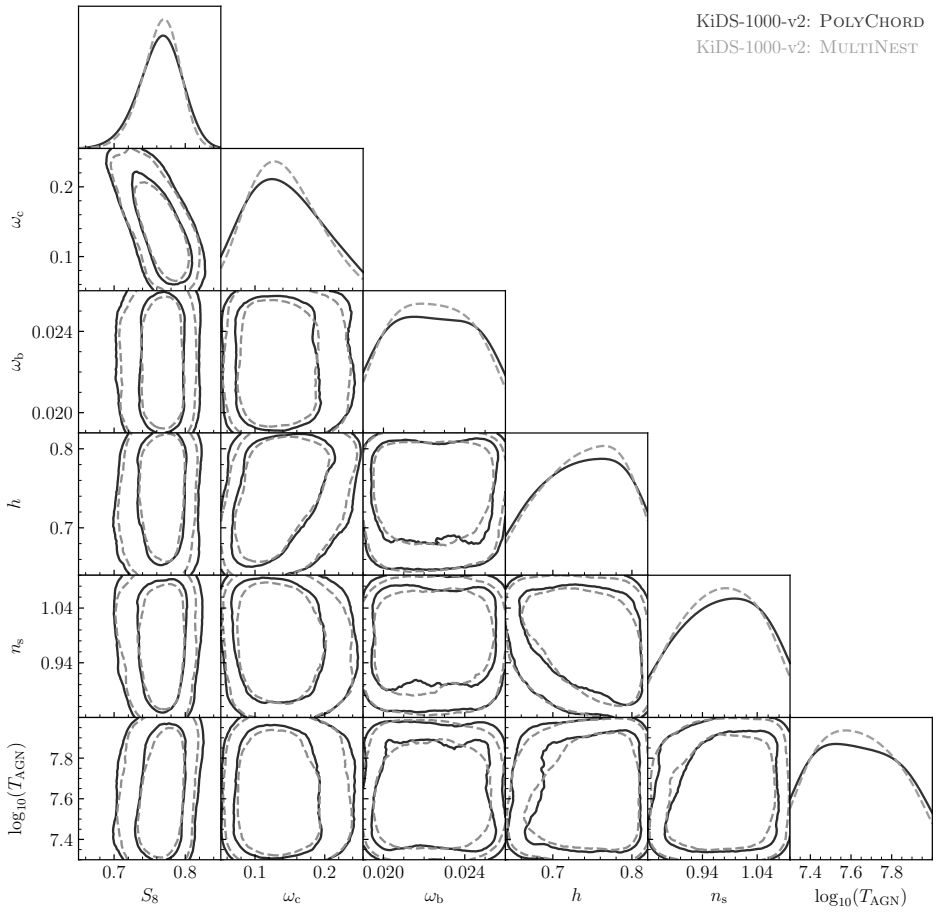

Figure 4.11: Posterior distributions of cosmological and baryonic parameters from our fiducial analyses, as generated by PolyChord (black solid lines) and MultiNest (grey dashed lines). The contours represent the 68% and 95% credible intervals and are smoothed using a Gaussian KDE with a bandwidth scaled by a factor of 1.5. We note that S_8 is the only parameter that our data robustly constrains.

4.D Contour plots for all free parameters

In this appendix, we provide two supplementary contour plots that display the posterior distributions of all twelve free parameters from our fiducial analyses, as produced by both the PolyChord and MultiNest sampling codes. The overall concordance between the results generated by PolyChord and MultiNest is evident.

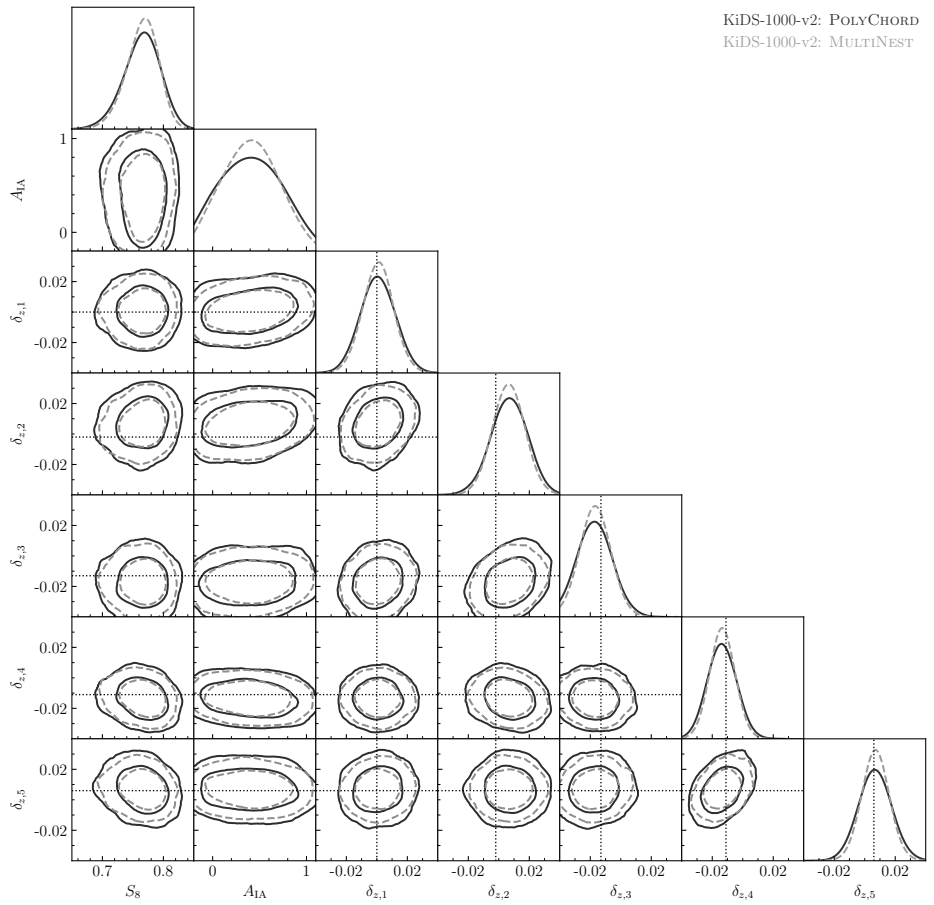

Figure 4.12: Posterior distributions of S_8 and nuisance parameters from our fiducial analyses, as generated by PolyChord (black solid lines) and MultiNest (grey dashed lines). The contours represent the 68% and 95% credible intervals and are smoothed using a Gaussian KDE with a bandwidth scaled by a factor of 1.5. We note that the Gaussian priors we have set, as outlined in Table 4.2, strongly influence the redshift offset parameters. The dotted lines represent the central values of these Gaussian priors.

CHAPTER 5

Weak lensing mass-luminosity scaling relations for galaxy groups: Testing the robustness of the halo model formalism

Abstract

Understanding the relation between baryonic observables and dark matter haloes is crucial for studying galaxy formation and evolution, and for deriving accurate cosmological constraints from galaxy surveys. In this chapter, we study this galaxy-halo connection by conducting a galaxy-galaxy weak lensing analysis on galaxy groups identified by the Galaxy and Mass Assembly survey, using galaxy shape catalogues from the fourth data release of the Kilo-Degree Survey. We interpreted the measured signals using a halo model formalism, employing a conditional luminosity function to describe the connection between halos and galaxies. Consistent with previous work, we found that the halo mass scales with the total r-band luminosity of the group according to a power law, with a slope of $1.25^{+0.12}_{-0.10}$. To assess the robustness of the current halo model formalism, we conducted sensitivity analyses. These revealed that the treatment of mis-centring in our model might introduce biases into the scaling relation constraints, which could exceed acceptable limits for future, more precise weak lensing measurements. To remedy this, it would be important to leverage insights from large-volume hydrodynamic simulations to develop a more physical model.

5.1 Introduction

According to the current standard model of cosmology, galaxies form within cold dark matter haloes, which originate from small initial density perturbations amplified by gravitational instability. This framework predicts a strong correlation between galaxy properties and the properties of their host dark matter haloes (see Wechsler & Tinker 2018, for a review). Dark matter haloes dominate the local gravitational potential, impacting matter clustering and providing the environment for the formation and evolution of galaxies (e.g. Blumenthal et al. 1984; Davis et al. 1985). On the flip side, various baryonic processes associated with galaxy formation, particularly the energetic feedback processes from supernovae (SNe) and active galactic nuclei (AGN), reshape the matter distribution on small scales (e.g. van Daalen et al. 2011; Hellwing et al. 2016; Chisari et al. 2018; van Daalen et al. 2020). Therefore, obtaining an accurate and precise understanding of the galaxy-halo connection is not only essential for studying galaxy formation and evolution, but also crucial for ensuring the accuracy of cosmological constraints derived from observations of large-scale structures (e.g. Semboloni et al. 2011; Schneider et al. 2020; Castro et al. 2021; Debackere et al. 2021).

Given that dark matter haloes typically host multiple galaxies, catalogues of galaxy groups and clusters are important in studying the galaxy-halo connection. Although massive galaxy clusters serve as a powerful tool for constraining cosmological models (see Allen et al. 2011, for a review), they are relatively rare and represent extreme conditions. On the other hand, galaxy groups, which host the majority of present-day galaxies and a significant portion of baryonic matter, offer a more representative view of galaxy formation (e.g. Robotham et al. 2011). They also contribute significantly to the cosmic shear signal (e.g. Semboloni et al. 2011; Debackere et al. 2020). Moreover, the gravitational binding energy of galaxy groups is comparable to the energy released by feedback processes from SNe and AGN, making galaxy groups particularly valuable for studying the impact of baryonic feedback (e.g. McCarthy et al. 2010; Kettula et al. 2015).

However, the robust identification of galaxy groups is a complex task that requires spectroscopic surveys with high spatial and redshift completeness. The Galaxy and Mass Assembly project (GAMA, Driver et al. 2011) represents one such effort. Achieving a 95 per cent spectroscopic completeness down to the r-band magnitude of 19.65 and covering approximately 250 deg^2 of sky area, GAMA currently offers the highest available redshift density over such an extensive area (Driver et al. 2022). As a result, it produced a reliable galaxy group catalogue with accurate estimations of group properties (Robotham et al. 2011).

The next challenge lies in measuring the dark matter properties of galaxy groups. This complexity becomes evident even when estimating basic properties like the mass of the dark matter haloes. For massive galaxy clusters, the X-ray measurement of the intracluster medium is commonly used for estimating the masses of individual clusters, under the assumption of hydrostatic equilibrium (see Ettori et al. 2013, for a review). However, when this method is applied to galaxy groups, its effectiveness diminishes

significantly due to their faint X-ray signals (e.g. Eckmiller et al. 2011; Pop et al. 2022; Bahar et al. 2022). Furthermore, various baryonic processes, including cooling, star formation, and feedback processes, can cause deviation from hydrostatic equilibrium, biasing the mass estimation that are based on this technique (e.g. Rasia et al. 2006; Biffi et al. 2016; Barnes et al. 2021; Logan et al. 2022).

Weak gravitational lensing provides an alternative approach to directly determining halo mass (e.g. Tyson et al. 1990; Hoekstra et al. 2001; Mandelbaum et al. 2006; Leauthaud et al. 2010; Hudson et al. 2015; Zacharegkas et al. 2022). It measures the subtle yet coherent distortions in the shapes of background galaxies, caused by the gravitational field of a foreground lens (see Bartelmann & Schneider 2001, for a review). These distortions directly trace the matter distributions in the foreground lenses, enabling the inference of total halo mass without requiring assumptions about their dynamical state.

However, the weak lensing signals produced by individual galaxy groups have low signal-to-noise ratios, which limits the precision of mass determinations for individual groups. To overcome this, we typically use an averaged measure from a collection of galaxy groups, which are selected and stacked based on narrowly binned observable properties. This method enables us to create a statistical description of the scaling relation between these observable properties and the masses of dark matter haloes (e.g. Viola et al. 2015; Rana et al. 2022).

To interpret stacked weak lensing measurements, we need a robust statistical model. The halo model combined with halo occupation statistics offers such a theoretical framework (e.g. Seljak 2000; Cooray & Sheth 2002; van den Bosch et al. 2013; Berlind & Weinberg 2002; Yang et al. 2003; Vale & Ostriker 2004; Cooray 2006). This approach statistically describes the properties of dark matter haloes and how galaxies inhabit them, thus creating a link between the underlying matter distributions and the statistical measures of weak lensing signals. In practice, the halo model contains several theoretically motivated or empirically required components, which are not always well constrained by the data. Combined with the intricate interplay and degeneracy among various parameters, this makes the selection of appropriate priors and interpretation of the parameters for the halo model a non-trivial task.

In this chapter, we study the scaling relation between group luminosity and halo mass, and assess the robustness of the current halo model formalism. We measure the weak lensing signals around galaxy groups identified by the GAMA survey, using the galaxy shape catalogue from the Kilo-Degree Survey (KiDS, de Jong et al. 2013; Kuijken et al. 2015). With the complete coverage of the three equatorial GAMA fields provided by the fourth data release of KiDS (KiDS-DR4, Kuijken et al. 2019; Giblin et al. 2021), our analysis has nearly double the sample volume compared to Viola et al. (2015). Moreover, we update the empirical modelling approach used by previous studies by adopting a more theoretically driven conditional luminosity function (e.g. Yang et al. 2003; Vale & Ostriker 2004; Cooray 2006). We evaluate the robustness of our current halo model framework by investigating its sensitivity to various adjustments in model setups. These sensitivity tests help identify crucial model ingredients and guide future

improvements. Considering the increased statistical power of the measured signals, such exploration and future enhancement of our current model are deemed necessary.

The rest of this chapter is structured as follows: Section 5.2 provides an overview of the data used in our analysis. Section 5.3 describes the measurements of weak lensing signals and the associated covariance matrix. The current modelling approach is introduced in Sect. 5.4, and the results derived from it are presented in Sect. 5.5. The sensitivity tests are discussed in Sect. 5.6. Finally, we conclude the chapter in Sect. 5.7. Throughout this chapter, we adopt a spatially flat Λ cold dark matter (ΛCDM) model, with parameters constrained by Planck Collaboration et al. (2020), for any calculations that require a cosmological framework: $\Omega_{\rm m} = 0.3158$, $\sigma_8 = 0.8120$, $n_{\rm s} = 0.96605$, $\Omega_{\rm b} = 0.04939$, and $h = 0.6732$. When we report values that are dependent on Hubble's constant, H_0, we include this dependency via $h = H_0/100$ km s^{-1} Mpc^{-1}, which facilitates comparison of results derived from different surveys and simulations.

5.2 Data

The data for our analysis originate from two surveys: the GAMA survey, which produced the galaxy group catalogue including baryonic observables such as stellar mass and luminosity; and the KiDS survey, which provided the shape measurements of background galaxies. In this section, we provide a concise overview of the catalogues used in our study. For further technical details, we direct interested readers to the relevant data release papers.

5.2.1 Lenses: GAMA groups

GAMA is a high-density, high-completeness spectroscopic survey conducted using the AAOmega instrument on the Anglo-Australian Telescope (Driver et al. 2011). Our analysis involves data from three equatorial fields of the GAMA II phase (G09, G12, G15), each covering a sky area of 60 square degrees (Liske et al. 2015). The GAMA data in these fields have a spectroscopic completeness of approximately 98 per cent for galaxies within the observed magnitude limit of $r < 19.8$. In particular, we use three key GAMA products: the G^{3}C group catalogue[1] (version 10, Robotham et al. 2011), the `StellarMassesLambdar` catalogue[2] (version 24, Taylor et al. 2011), and the random catalogue[3] (version 2, Farrow et al. 2015).

The G^{3}C group catalogue (version 10) consists of 26 194 groups identified using a friends-of-friends (FoF) algorithm. This method establishes connections between galaxies based on their three-dimensional separations. The algorithm has been validated on mock catalogues derived from semi-analytic simulations and has been found robust against outliers and linking errors. In our study, we only consider groups with a minimum of five identified members to reduce the impact of interlopers (Robotham et al. 2011).

[1] www.gama-survey.org/dr4/schema/dmu.php?id=115
[2] www.gama-survey.org/dr4/schema/dmu.php?id=1010
[3] www.gama-survey.org/dr4/schema/dmu.php?id=109

After applying this selection, we are left with a total of 2752 groups.

In our analysis, we consider the Brightest Cluster/group Galaxy (BCG) as the central galaxy. Another commonly used method for selecting the central galaxy involves iteratively removing group members that are furthest from the group's centre of light. However, Robotham et al. (2011) found that, for groups with more than five members, this iterative procedure converges on the BCG 95% of the time. For weak lensing analyses, the subtle difference between these two methods becomes even more statistically negligible, as illustrated in Appendix A of Viola et al. (2015).

The stellar masses of the BCGs are obtained from the `StellarMassesLambdar` catalogue (version 24). It estimates galaxy stellar masses using stellar population synthesis models from Bruzual & Charlot (2003), assuming a Chabrier (2003) initial mass function. The model fits are applied over a fixed rest-frame wavelength range (300 – 11000) using matched aperture photometry derived from the Lambda Adaptive Multi-Band Deblending Algorithm in R (LAMBDAR, Wright et al. 2016). We utilise the `logmstar` value from the catalogue, which represents the total mass of all luminous material and remnants, but excluding mass recycled back into the interstellar medium. We do not correct the aperture-photometry-based stellar mass, as the flux-based scaling factor has been removed from the latest GAMA stellar mass catalogue, and not all galaxies in the GAMA survey have an accurate total flux estimation.

The GAMA random catalogue (version 2) is employed in our analysis to quantify additive shear biases in the weak lensing signals. This catalogue comprises randomly distributed points, designed to reflect the same selection function as the main spectroscopic survey. For our analysis, we randomly select 1 million points from this catalogue for each of the GAMA fields under consideration. Since these random points do not correspond to any actual matter distribution in the field, performing equivalent weak lensing measurements around them illuminates potential systematic errors in our process.

5.2.2 Sources: KiDS galaxies

KiDS is a wide-field imaging survey, specifically designed to measure weak gravitational lensing effects (de Jong et al. 2013; Kuijken et al. 2015). It spans 1350 square degrees of the sky, with optical images in the $ugri$ bands taken from the ESO VLT Survey Telescope. Among these, the r-band images, offering the highest imaging quality, are used for measuring galaxy shapes. In collaboration with the VISTA Kilo-degree INfrared Galaxy survey (VIKING, Edge et al. 2013) using the nearby ESO VISTA telescope, the KiDS shear catalogue also incorporates photometry from five $ZYJHK_s$ near-infrared bands. This additional data significantly enhances the accuracy of photometric redshift estimates.

For our analysis, we employ the public KiDS-1000 shear catalogue from the fourth data release of KiDS (Kuijken et al. 2019; Giblin et al. 2021). This catalogue fully covers the three equatorial fields of GAMA, as illustrated in Fig. 5.1. Thanks to this complete coverage, we are now able to measure weak lensing signals around all 2752 selected GAMA groups, approximately doubling the number used in previous similar

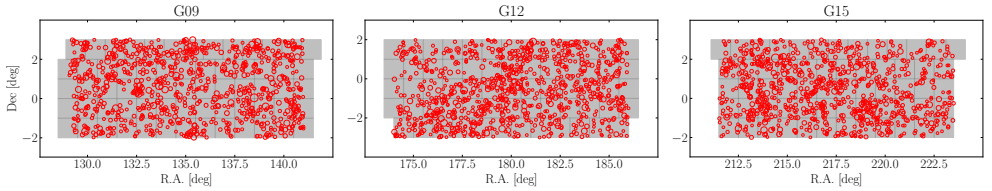

Figure 5.1: KiDS-DR4 coverage across the three equatorial GAMA fields (G09, G12, G15). The grey boxes represent KiDS tile images, each covering 1 square degree. The red circles indicate the selected GAMA groups, each consisting of at least five members. The size of these circles corresponds to the logarithm of the group richness. With the KiDS-DR4 data, we have achieved complete coverage of weak lensing measurements across the three equatorial GAMA fields.

analyses by Viola et al. (2015) and Rana et al. (2022).

The galaxy shapes in the KiDS-1000 catalogue are measured using the *lens*fit code (Miller et al. 2013; Fenech Conti et al. 2017) applied to the r-band images. These shear measurements are further calibrated using image simulations developed by Kannawadi et al. (2019). The photometric redshift estimates are derived from nine-band photometry using the BPZ code (Benítez 2000), and are calibrated with a spectroscopic reference sample as detailed in Hildebrandt et al. (2021).

5.3 Galaxy-galaxy weak lensing signals

The lensing effect introduces coherent tangential distortions in the apparent shapes of background galaxies. These distortions, known as the tangential shear, $\gamma_{\rm t}$, correlate with the projected mass density contrast of the foreground lens[4] (e.g. Bartelmann & Schneider 2001):

$$\Delta\Sigma(R) \equiv \bar{\Sigma}(\leq R) - \Sigma(R) = \Sigma_{\rm cr}\gamma_{\rm t}(R) \,, \tag{5.1}$$

where the mass contrast, $\Delta\Sigma(R)$, is also commonly referred to as the excess surface density (ESD). The $\Sigma(R)$ represents the local surface mass density at a projected comoving separation, R, between the lens and source, while $\bar{\Sigma}(\leq R)$ denotes the mean surface density within this radius. $\Sigma_{\rm cr}$, known as the critical surface density, serves as a measure of lensing efficiency and is defined as

$$\Sigma_{\rm cr} \equiv \frac{c^2}{4\pi G}\frac{D(z_{\rm s})}{D(z_{\rm l})D(z_{\rm l},\, z_{\rm s})} \,, \tag{5.2}$$

where G and c denote the gravitational constant and the speed of light, respectively. $D(z)$ is the angular diameter distance at redshift z.

[4]Throughout this work, we do not distinguish between the original shear γ and the reduced shear $g \equiv \gamma/(1-\kappa)$, given that the convergence κ is much less than one in the weak lensing regime.

Therefore, by measuring the ESD profile for foreground objects, we can infer their masses, assuming a certain density profile for the object. In this section, we detail how we estimate ESD for the selected GAMA galaxy groups from the KiDS shear measurements (Sect. 5.3.1) and the corresponding covariance matrix necessary for modelling (Sect. 5.3.2).

5.3.1 ESD measurements

We estimated the tangential shear by calculating the azimuthal average of the tangential projection, $\epsilon_{\rm t}$, of the *lens*fit measured ellipticities of the KiDS source galaxies. This is defined as

$$\begin{bmatrix} \epsilon_{\rm t} \\ \epsilon_{\times} \end{bmatrix} \equiv \begin{bmatrix} -\cos(2\phi) & -\sin(2\phi) \\ \sin(2\phi) & -\cos(2\phi) \end{bmatrix} \cdot \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} , \tag{5.3}$$

where ϕ denotes the relative position angle of the source in relation to the lens. The azimuthal average of the cross projection, denoted as $\epsilon_{\times}$, can serve as a useful indicator of potential systematic contamination, given that the lensing effect only induces tangential shear to the leading order.

To account for both measurement and geometric effects, a weight was assigned to each lens-source pair during the computation of the azimuthal average. This weight is given by

$$w_{\rm ls} \equiv w_{\rm s}\, \tilde{\Sigma}_{\rm cr,l}^{-2} , \tag{5.4}$$

where $w_{\rm s}$ is the *lens*fit weight, which accounts for shape measurement uncertainties, and $\tilde{\Sigma}_{\rm cr,l}$ is the 'effective critical surface density'. The latter is used to down-weight lens-source pairs that are close in redshift and thus carry fewer lensing signals. This 'effective critical surface density' was calculated for each lens, by integrating the redshift distribution of the source galaxies behind the given lens to statistically represent the source distance. This approach is designed to mitigate the potential impact of uncertainties in the source redshift estimates. Following Eq. (5.2), the 'effective critical surface density' was calculated as

$$\tilde{\Sigma}_{\rm cr,l}^{-1} = \frac{4\pi G}{c^2} D(z_{\rm l}) \int_{z_{\rm l}+\delta_z}^{\infty} {\rm d}z_{\rm s}\, \frac{D(z_{\rm l},\, z_{\rm s})}{D(z_{\rm s})}\, n(z_{\rm s}) , \tag{5.5}$$

where the source redshift distribution $n(z_{\rm s})$ was determined from a deep spectroscopic reference catalogue that was re-weighted to closely match the KiDS-1000 sources (see Hildebrandt et al. 2021 for details). Following Dvornik et al. (2017), we introduced a redshift difference threshold, $\delta_z = 0.2$, to mitigate contamination from group members. This redshift cutoff, $z_{\rm s} > z_{\rm l} + \delta_z$, was applied to the source galaxies involved in the calculation as well as to the reference spectroscopic sample used for determining the source redshift distribution.

The median velocity dispersion of the GAMA galaxy groups used in our study is $\sim$300 km s^{-1}, which is not massive enough for measuring lensing signals from individual groups. Therefore, we used a stacking process to boost the signal-to-noise

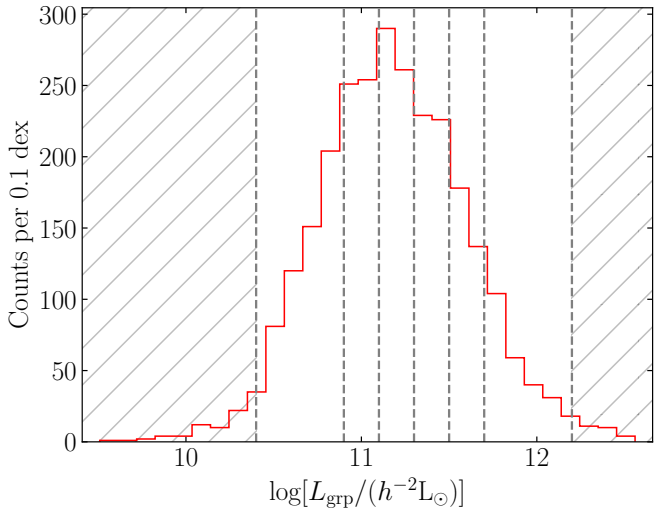

Figure 5.2: Distribution of the group total r-band luminosity underlying our binning strategy for the measurement of stacked ESD profiles. The vertical lines represent the boundaries of the bins, with their corresponding values detailed in Table 5.1. Objects falling within the hatched regions are excluded from our stacked analyses.

ratio. Following Eq. (5.1) and considering the weighting scheme mentioned above, the stacked ESD profile for an ensemble of galaxy groups can be estimated as

$$\Delta\Sigma(R) = \left[\frac{\sum_{\rm ls} w_{\rm ls}\, \epsilon_{\rm t}\, \tilde{\Sigma}_{\rm cr,l}}{\sum_{\rm ls} w_{\rm ls}}\right] \frac{1}{1+K}\,, \tag{5.6}$$

where the correction

$$K = \frac{\sum_{\rm ls} w_{\rm ls}\, m_{\rm s}}{\sum_{\rm ls} w_{\rm ls}}\,, \tag{5.7}$$

accounts for the multiplicative shear biases in the *lens*fit shape measurements. The correction factor was derived from image simulations developed by Kannawadi et al. (2019), who estimated an average m value for each redshift bin used in the KiDS-1000 cosmic shear analyses (Giblin et al. 2021; Asgari et al. 2021). We directly assigned these average m values to the galaxies used in our analysis based on their redshift, ignoring potential deviations in the galaxy size and signal-to-noise ratio distribution between the samples selected for our analysis and those used in the KiDS-1000 cosmic shear analyses. The overall correction factor is small, with K approximately equal to 0.001, and remains independent of the angular scale at which it is computed.

The additive shear biases were addressed by conducting weak lensing measurements around one million random points selected from the GAMA random catalogue (version 2, Farrow et al. 2015). These additive biases are both scale-dependent, with substantial biases on scales larger than $1h^{-1}$Mpc, and patch-dependent (see Appendix A of Dvornik

Table 5.1: Summary of the binning limits, number of groups, mean redshift of the groups, and the mean stellar mass of the BCGs for each bin used in the stacked ESD measurements.

Observable	Range	$N_{\rm groups}$	$z_{\rm mean}$	$M^{\rm BCG}_{\star,\rm mean}$
$L_{\rm grp}$	(10.4, 10.9]	628	0.12	10.59
	(10.9, 11.1]	477	0.17	10.79
	(11.1, 11.3]	528	0.21	10.91
	(11.3, 11.5]	432	0.26	11.00
	(11.5, 11.7]	312	0.29	11.08
	(11.7, 12.2]	267	0.32	11.18

The units for luminosity and stellar mass are given as $\log(L_{\rm grp}/[h^{-2}{\rm L}_\odot])$ and $\log(M_\star/[h^{-2}{\rm M}_\odot])$, respectively.

et al. 2017). Thus, we performed the correction separately for three GAMA patches (G9, G12, and G15). The overall corrections remain minor, with values at the sub-percent level, attributable to the complete coverage of the GAMA fields by the KiDS observations.

In our analysis, we divided the GAMA groups into six bins, based on their total r-band luminosity ($L_{\rm grp}$). We set lower and upper limits to exclude the tails of the distributions, as demonstrated in Fig. 5.2. This approach helps to mitigate group detection effects and prevents the estimates from being overly influenced by a small number of outliers within each bin. The boundaries for each bin were chosen to maintain similar signal-to-noise ratios across all bins in the measured ESDs. The statistical details of our defined bins are outlined in Table 5.1.

We measured the ESD profiles over 10 logarithmically-spaced radial bins, within a range of 0.03 to 2 h^{-1}Mpc. This range was determined by taking into account the signal-to-noise ratio at larger separations and the blending effects at smaller ones. The measured ESD profiles are illustrated in Fig. 5.3. The additive and multiplicative shear biases present in the original measurements were corrected using methods discussed earlier. The overall signal-to-noise ratios for the ESD measurements were found to be approximately 27.

5.3.2 Covariance matrix estimation

The ellipticity measurements of source galaxies can be used multiple times for $\Delta\Sigma(R)$ estimates across various radial bins and lenses, leading to correlations between the stacked ESD estimates. To account for these correlations in our modelling, we employed

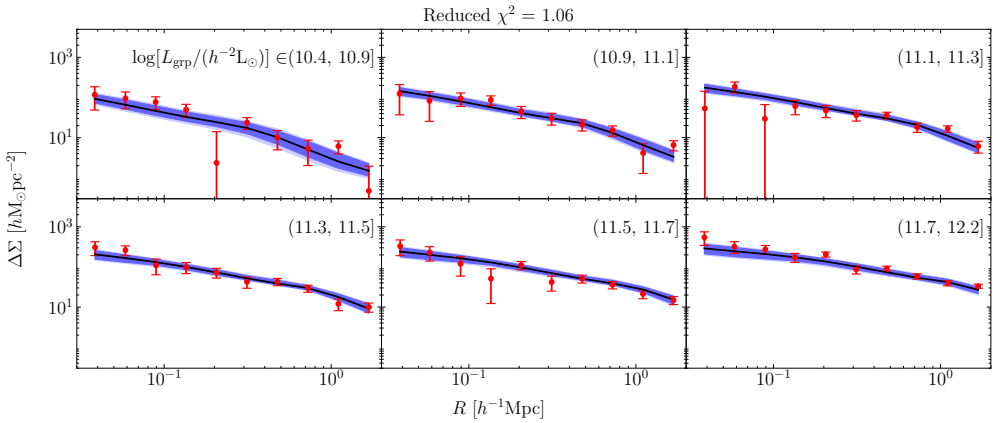

Figure 5.3: Excess Surface Density (ESD) profiles for selected GAMA groups, binned into six groups according to their total r-band luminosity ($L_{\rm grp}$). The error bars correspond to the square root of the covariance matrix's diagonal elements. The black lines depict the best-fit results obtained from our halo model, as detailed in Sect. 5.4. The dark and light blue shaded regions represent the 68% and 95% credible intervals of the fitting, respectively. The figure title shows the reduced χ^2 value of the best-fit results, calculated assuming 54 degrees of freedom (six independent fitting parameters to 60 data points).

the covariance matrix estimation technique as developed by Viola et al. (2015). This approach, proven valid in previous KiDS+GAMA analyses (e.g. Sifón et al. 2015; Brouwer et al. 2016), takes into account the shape noise of source galaxies and incorporates information about the survey geometry. However, it does not account for cosmic variance. Viola et al. (2015) demonstrated that this simplification is valid for measurements conducted within the range of $R \leq 2h^{-1}$Mpc, which is consistent with the range we adopt for our analysis.

5.4 Halo model and occupation statistics

From a statistical perspective, the projected mass surface density, $\Sigma(R)$, of an ensemble of lenses is related to the galaxy-matter cross-correlation function, $\xi_{\rm gm}(r)$, through the Abel transform:

$$\Sigma(R) = 2\,\bar{\rho}_{\rm m} \int_R^\infty {\rm d}r\ [\xi_{\rm gm}(r) + 1]\ \frac{r}{\sqrt{r^2 - R^2}}\,, \tag{5.8}$$

where we assume that the stacked density profile of the lenses is spherically symmetric. This is a valid assumption, given that our measurements are derived from stacking data from multiple lenses with varied orientations. Therefore, we can interpret the measured $\Delta\Sigma(R)$ signals if we have a model to describe $\xi_{\rm gm}(r)$. The halo model, complemented by halo occupation statistics, offers such a theoretical framework (e.g. Seljak 2000;

Cooray & Sheth 2002; Peacock & Smith 2000; Berlind & Weinberg 2002; Yang et al. 2003; van den Bosch et al. 2013).

In this section, we detail how we employ this framework to interpret our stacked ESD measurements. We begin with a concise overview of the halo model formalism in Sect. 5.4.1, largely adhering to the notations used by van den Bosch et al. (2013) and van Uitert et al. (2016). Then, we specify our choice of the model ingredients in Sect. 5.4.2. Our approach to halo occupation statistics is detailed in Sect. 5.4.3. Finally, we outline our fitting procedure in Sect. 5.4.4.

5.4.1 Halo model formalism

The halo model assumes that all dark matter resides within virialised haloes, the sizes of which are determined by a chosen overdensity threshold. In line with conventions in weak lensing studies, we define this threshold such that the average density within a certain radius is 200 times the mean density of the Universe, denoted as $\bar{\rho}_\mathrm{m}$. As a result, the mass of a specific halo can be formulated as

$$M_\mathrm{h} = \frac{4\pi}{3}\, 200\, \bar{\rho}_\mathrm{m}\, r_{200}^3 \,. \tag{5.9}$$

Using a formalism that encapsulates the internal density profile of these haloes, such as the widely used Navarro-Frenk-White (NFW) profile (Navarro et al. 1997), we can describe the matter-matter power spectrum of the Universe, using a theoretical approach based on the correlations between dark matter haloes (e.g. van den Bosch et al. 2013).

With a statistical understanding of how galaxies populate dark matter haloes, often referred to as halo occupation statistics, we can extend the halo model framework to include calculations for both the galaxy-galaxy power spectrum and the galaxy-matter power spectrum. Using the notation of van den Bosch et al. (2013) and van Uitert et al. (2016), we can concisely express the three types of power spectra as

$$P_{xy}(k) = P_{xy}^\mathrm{1h}(k) + P_{xy}^\mathrm{2h}(k) \,, \tag{5.10}$$

where

$$P_{xy}^\mathrm{1h}(k) = \int \mathrm{d}M_\mathrm{h}\, \mathcal{H}_x(k,\, M_\mathrm{h})\, \mathcal{H}_y(k,\, M_\mathrm{h})\, n_\mathrm{h}(M_\mathrm{h}) \,, \tag{5.11}$$

and

$$\begin{aligned} P_{xy}^\mathrm{2h}(k) = &\int \mathrm{d}M_{\mathrm{h},1}\, \mathcal{H}_x(k,\, M_{\mathrm{h},1})\, n_\mathrm{h}(M_{\mathrm{h},1}) \\ &\int \mathrm{d}M_{\mathrm{h},2}\, \mathcal{H}_y(k,\, M_{\mathrm{h},2})\, n_\mathrm{h}(M_{\mathrm{h},1})\, P_\mathrm{h}(k|M_{\mathrm{h},1}, M_{\mathrm{h},2}) \,. \end{aligned} \tag{5.12}$$

The terms $P_{xy}^\mathrm{1h}(k)$ and $P_{xy}^\mathrm{2h}(k)$, known as the one-halo and two-halo terms, describe the correlations within a single halo and between different haloes, respectively. Here, k represents the wavenumber, $n_\mathrm{h}(M_\mathrm{h})$ denotes the halo mass function, and $P_\mathrm{h}(k|M_{\mathrm{h},1}, M_{\mathrm{h},2})$

stands for the halo power spectrum associated with the mass values $M_{\rm h,1}$ and $M_{\rm h,2}$. This spectrum serves as a biased tracer of the linear dark matter power spectrum $P_{\rm m}^{\rm lin}(k)$:

$$P_{\rm h}(k|M_{\rm h,1}, M_{\rm h,2}) = b_{\rm h}(M_{\rm h,1})\ b_{\rm h}(M_{\rm h,2})\ P_{\rm m}^{\rm lin}(k)\ , \tag{5.13}$$

where $b_{\rm h}(M_{\rm h})$ refers to the large-scale halo bias. The subscripts, x and y, either refer to 'g' for galaxies or 'm' for matter, corresponding to different forms of $\mathcal{H}_x(k,\ M_{\rm h})$:

$$\mathcal{H}_{\rm m}(k,\ M_{\rm h}) \equiv \frac{M_{\rm h}}{\bar{\rho}_{\rm m}}\ \tilde{u}_{\rm m}(k|M_{\rm h})\ , \tag{5.14}$$

or

$$\mathcal{H}_{\rm g}(k,\ M_{\rm h}) \equiv \frac{\langle n_{\rm g}|M_{\rm h}\rangle}{\bar{n}_{\rm g}}\ \tilde{u}_{\rm g}(k|M_{\rm h})\ . \tag{5.15}$$

In these equations, $\tilde{u}_{\rm m}(k|M_{\rm h})$ describes the normalised density profile of dark matter haloes in Fourier space. The term $\tilde{u}_{\rm g}(k|M_{\rm h})$ describes the galaxy distributions within the halo, and its form depends on the types of galaxies. The term $\langle n_{\rm g}|M_{\rm h}\rangle$ refers to the average number of galaxies that reside in a halo of mass $M_{\rm h}$, i.e., the halo occupation distribution (HOD), and $\bar{n}_{\rm g}$ is the average number of galaxies integrated over all halo masses, given by

$$\bar{n}_{\rm g} = \int {\rm d}M_{\rm h}\ \langle n_{\rm g}|M_{\rm h}\rangle\ n_{\rm h}(M_{\rm h})\ . \tag{5.16}$$

The desired galaxy-matter cross-correlation function, $\xi_{\rm gm}$, is simply a Fourier transform of the galaxy-dark matter power spectrum $P_{\rm gm}(k)$:

$$\xi_{\rm gm}(r) = \frac{1}{2\pi^2}\int_0^\infty {\rm d}k\ P_{\rm gm}(k)\ \frac{\sin(kr)}{kr}\ k^2\ . \tag{5.17}$$

5.4.2 Model ingredients

For the internal density distribution of dark matter haloes, we adopted the Navarro-Frenk-White (NFW, Navarro et al. 1997) profile, with a mass-concentration relation from Duffy et al. (2008):

$$c_{\rm m} = f_{\rm c} \times 10.14\ \left(\frac{M_{\rm h}}{2\times10^{12}\ h^{-1}{\rm M}_\odot}\right)^{-0.081}\ (1+z_{\rm l})^{-1.01}\ , \tag{5.18}$$

where $f_{\rm c}$ is a scaling parameter, which we allowed to vary during the model fitting. The redshift dependence in this equation was derived from the results of a high-resolution N-body simulation conducted under the concordance ΛCDM cosmology (Bullock et al. 2001). Although more complex redshift dependencies are anticipated theoretically, they predominantly apply to redshifts greater than one (e.g. Muñoz-Cuartas et al. 2011), exceeding the highest lens redshift in our study.

To account for the mass contribution from central galaxies residing in the innermost region of the dark matter halo, we incorporated a point mass into the NFW density

profile. This mass was set to the mean stellar mass of the selected central galaxies for each stacked bin (see Table 5.1). Considering the scales of our ESD measurements, which range from 0.03 to 2 h^{-1}Mpc, our analysis is not sensitive to the detailed matter distributions within the innermost part of the dark matter halo.

For the halo mass function and the halo bias, we employed the calibrated fitting functions from Tinker et al. (2010), who derived parameters from a series of cosmological N-body simulations within the framework of the ΛCDM cosmology. Given that we fit the ESD profiles up to 2.0 h^{-1}Mpc, and that the halo bias only enters our calculations via the two-halo term (as shown in Eq. 5.12), our analysis is not sensitive to the precise form of the halo bias function given the current statistical uncertainties of the observed signals.

If the central galaxy (or in our case, the BCG) resides exactly at the centre of its host halo, the $\tilde{u}_{\rm g}(k|M_{\rm h})$ term shown in Eq. (5.15) would consistently equate to one. However, previous studies showed that BCGs often do not perfectly trace the centre of their host dark matter halos (e.g. Skibba et al. 2011; George et al. 2012; De Propris et al. 2021; Ahad et al. 2023). We statistically modelled this mis-centring following Viola et al. (2015) (also see Oguri & Takada 2011; More et al. 2015), with an equation:

$$\tilde{u}_{\rm g}(k|M_{\rm h}) = 1 - p_{\rm off} + p_{\rm off} \times \exp\left[-\frac{1}{2}\, k^2\, (r_{\rm s}\mathcal{R}_{\rm off})^2\right] , \tag{5.19}$$

where $r_{\rm s}$ represents the scale radius of the halo, as described by the NFW profile. This model assumes that a fraction $p_{\rm off}$ of BCGs is mis-centred, with the normalised radial distribution of these mis-centred galaxies relative to the true halo centre following a Gaussian distribution with a width of $r_{\rm s}\mathcal{R}_{\rm off}$. It is clear that setting either $p_{\rm off}$ or $\mathcal{R}_{\rm off}$ to zero results in a model without mis-centring. In our current fiducial model, we treated both $p_{\rm off}$ and $\mathcal{R}_{\rm off}$ as uninformative free parameters.

Although the incorporation of these two mis-centring parameters seems physically well-motivated, the data might not fully comply with the idealised assumptions of this model, which imply an isotropic random mis-identification of group centres leading to a Gaussian radial distribution of mis-centred galaxies, and also presume perfect group identification. For example, Ahad et al. (2023) found that line-of-sight projections, which result in a discrepancy between the projected and intrinsic luminosity, account for approximately half of the identified mis-centred groups in their simulations. Furthermore, the GAMA group-finding algorithm is susceptible to aggregation and fragmentation effects (see Appendix A of Jakobs et al. 2018). Aggregation refers to the phenomenon where two smaller groups, located in close proximity and linked by an intermediate galaxy, are identified as a single larger group. On the other hand, fragmentation occurs when a single, intrinsically large group is identified as several smaller groups. According to Ahad et al. (2023), the aggregation effects cause mis-centring in roughly 5% of their sample. Further complicating the issue is the high degeneracy observed between the mis-centring parameters and the concentration scaling parameter, $f_{\rm c}$ (see Sect. 5.1.2 of Viola et al. 2015). All these factors make the practical handling and interpretation of the mis-centring parameters challenging. We assess the robustness of the constrained halo

masses against various setups of mis-centring parameters to guide future improvements, in Sect. 5.6.

5.4.3 Halo occupation statistics

To extend the halo model framework for interpreting the galaxy-matter power spectrum used in the analysis of measured ESD signals (Eqs. 5.8 and 5.17), we require a formalism to describe the term $\langle n_{\rm g}|M_{\rm h}\rangle$ in Eq. (5.15). This term, often referred to as the halo occupation statistics or the halo occupation distribution (HOD), can be addressed in two ways. One approach involves directly constructing models to represent the average number of galaxies or groups associated with a dark matter halo of mass $M_{\rm h}$ (e.g. Berlind & Weinberg 2002). Alternatively, we can use the conditional stellar mass or luminosity function (CLF) to express the number density of galaxies or groups, based on their stellar mass or luminosity, given a halo mass (e.g. Yang et al. 2003). In our analysis, we opt for the latter methodology due to its direct connection to the relation between baryonic observable properties and halo mass, which is the focus of our study. It is worth noting that our CLF-based HOD model differs from the previous analysis by Viola et al. (2015), who defined the HOD directly based on the average number of groups as a function of halo mass.

In our fiducial model, we adopted a log-normal distribution to describe the group luminosity distribution given a halo mass. This has proven to be a validated CLF for central galaxy properties (e.g. Yang et al. 2008; Cacciato et al. 2009; van den Bosch et al. 2013; van Uitert et al. 2016):

$$\Phi(L_{\rm grp}|M_{\rm h}) = \frac{1}{\sqrt{2\pi}\,\ln(10)\,\sigma_L\,L_{\rm grp}}\,\exp\left[-\frac{(\log L_{\rm grp}-\log L_{\rm grp,m})^2}{2\sigma_L^2}\right]. \tag{5.20}$$

This function consists of two free parameters: the scatter σ_L and the mean $L_{\rm grp,m}$ for a given halo mass $M_{\rm h}$. In our current model, we considered σ_L as a halo mass-independent free parameter, following van Uitert et al. (2016). For the mean $L_{\rm grp,m}$, we assumed a power-law scaling relation between $L_{\rm grp,m}$ and the halo mass $M_{\rm h}$:

$$\frac{L_{\rm grp,m}}{10^{11.5}\,h^{-2}{\rm L}_\odot} = C_L\,\left(\frac{M_{\rm h}}{10^{14}\,h^{-1}{\rm M}_\odot}\right)^{\alpha_L}, \tag{5.21}$$

where C_L denotes the amplitude and α_L is the index. Both are free parameters in our fitting process.

Under the assumption of sample completeness, which is valid given the high completeness of the GAMA survey and our exclusion of distribution tails (see Fig. 5.2), we can calculate the mean number of groups per specific observable bin as follows:

$$\langle n_{\rm g}|M_{\rm h}\rangle = \int_{L_{\rm grp,min}}^{L_{\rm grp,max}} {\rm d}L_{\rm grp}\,\Phi(L_{\rm grp}|M_{\rm h})\,, \tag{5.22}$$

Table 5.2: Free parameters in our fiducial model, including their priors and derived constraints.

Parameter	Prior	Constraints
f_c	[0.2, 3]	$1.23^{+0.30}_{-0.26}$
p_{off}	[0, 1]	$0.50^{+0.08}_{-0.10}$
$\mathcal{R}_{off}$	[0, 3.5]	$3.13^{+0.27}_{-0.48}$
σ_L	[0.01, 2]	$0.07^{+0.07}_{-0.05}$
C_L	[0, 3]	$0.98^{+0.12}_{-0.14}$
α_L	[0, 5]	$0.80^{+0.07}_{-0.07}$

The parameters can be categorised into two main groups. The first group contains the standard halo model components (Sect. 5.4.2), while the second group includes parameters associated with the halo occupation statistics (Sect. 5.4.3). The reported constraints correspond to the marginalised median values of the sampling, with the uncertainties indicating the marginalised 68% credible intervals.

where $L_{grp,min}$ and $L_{grp,max}$ denote the lower and upper boundaries of the binning, respectively, as detailed in Table 5.1. We assess the impact of potential sample incompleteness in Sect. 5.6.

5.4.4 Model fitting

We fitted the aforementioned halo model to the stacked ESD measurements and sampled the posterior space using the `emcee` code (Foreman-Mackey et al. 2013), which implements the affine invariant Markov Chain Monte Carlo (MCMC) ensemble sampler (Goodman & Weare 2010). The convergence of the MCMC chains is evaluated using the integrated autocorrelation time (e.g. Goodman & Weare 2010). For the halo model calculations requiring lens redshifts, we used the mean redshifts from each stacked bin, as presented in Table 5.1. Our fiducial model incorporates six free parameters, all assigned broad, uninformative priors. The ranges for these priors are detailed in Table 5.2.

5.5 Results from the fiducial model

The best-fit results for the ESD profiles, along with their 68% and 95% credible intervals, are illustrated in Fig. 5.3. The reduced best-fit χ^2 value for our current fiducial model is approximately 1.06, assuming independence among the six free parameters in the model. This suggests a reasonable fit of our fiducial model to the data. The constraints

on the model parameters are presented in Table 5.2 as median values derived from the marginalised distributions, accompanied by their uncertainties representing the 68% credible intervals. The corresponding posterior distributions of these parameters are visualised as contour plots in Appendix 5.A.

5.5.1 Stacked halo masses

While our CLF-based halo model offers a direct estimation of the scaling relation between stacked properties and halo mass, via Eq. (5.21), it is instructive to estimate the average halo mass $\langle M_\mathrm{h} \rangle$ for each stacked bin. We can calculate this using the following equation:

$$\langle M_\mathrm{h} \rangle = \int \mathrm{d}M_\mathrm{h}\, \mathcal{P}(M_\mathrm{h}) M_\mathrm{h} \,, \tag{5.23}$$

where

$$\mathcal{P}(M_\mathrm{h}) = \int_{L_\mathrm{grp,min}}^{L_\mathrm{grp,max}} \mathrm{d}L_\mathrm{grp}\, \mathcal{P}(M_\mathrm{h}|L_\mathrm{grp}) \,, \tag{5.24}$$

is an integral of the conditional halo mass distribution $\mathcal{P}(M_\mathrm{h}|L_\mathrm{grp})$ for a given observable value of L_grp, which can be derived from the components of the halo model using Bayes' theorem:

$$\mathcal{P}(M_\mathrm{h}|L_\mathrm{grp}) = \frac{\Phi(L_\mathrm{grp}|M_\mathrm{h})\; n_\mathrm{h}(M_\mathrm{h})}{\bar{n}_\mathrm{g}} \,, \tag{5.25}$$

where $\Phi(L_\mathrm{grp}|M_\mathrm{h})$ is the CLF presented in Eq. (5.20), n_h is the halo mass function from Tinker et al. (2010), and $\bar{n}_\mathrm{g}$ is the average number of galaxies, as defined in Eq. (5.16).

Figure 5.4 shows the inferred halo mass distributions and the corresponding average halo mass for each stacked bin. To demonstrate the scatter in the sampled halo mass distributions, we randomly selected 100 sets of model parameter values from the sample space and plotted the corresponding distributions. The vertical lines in the figure represent the median values of the estimated average halo masses, with the shaded regions indicating the 68% credible intervals. Most of the sampled distributions exhibit well-defined peaks around the median values of the estimated average halo masses. As we move towards bins with higher observable values, the overall distribution shifts towards larger halo mass values. This scaling relation is quantified by the two parameters presented in the power-law scaling function of Eq. (5.21).

5.5.2 Group luminosity-halo mass relation

One key advantage of our CLF-based halo mass model, in contrast to the model employed by Viola et al. (2015), is its ability to directly constrain the scaling relation between the stacked properties and halo mass during the main sampling procedure. Given that the scaling relation is a part of our CLF-based model (Eq. 5.21), it obviates the need for an additional linear regression process performed on the estimated mean halo masses and stacked observable values, which can be challenging due to correlated measurement errors and intrinsic scatter in these averaged properties (e.g. Kelly 2007).

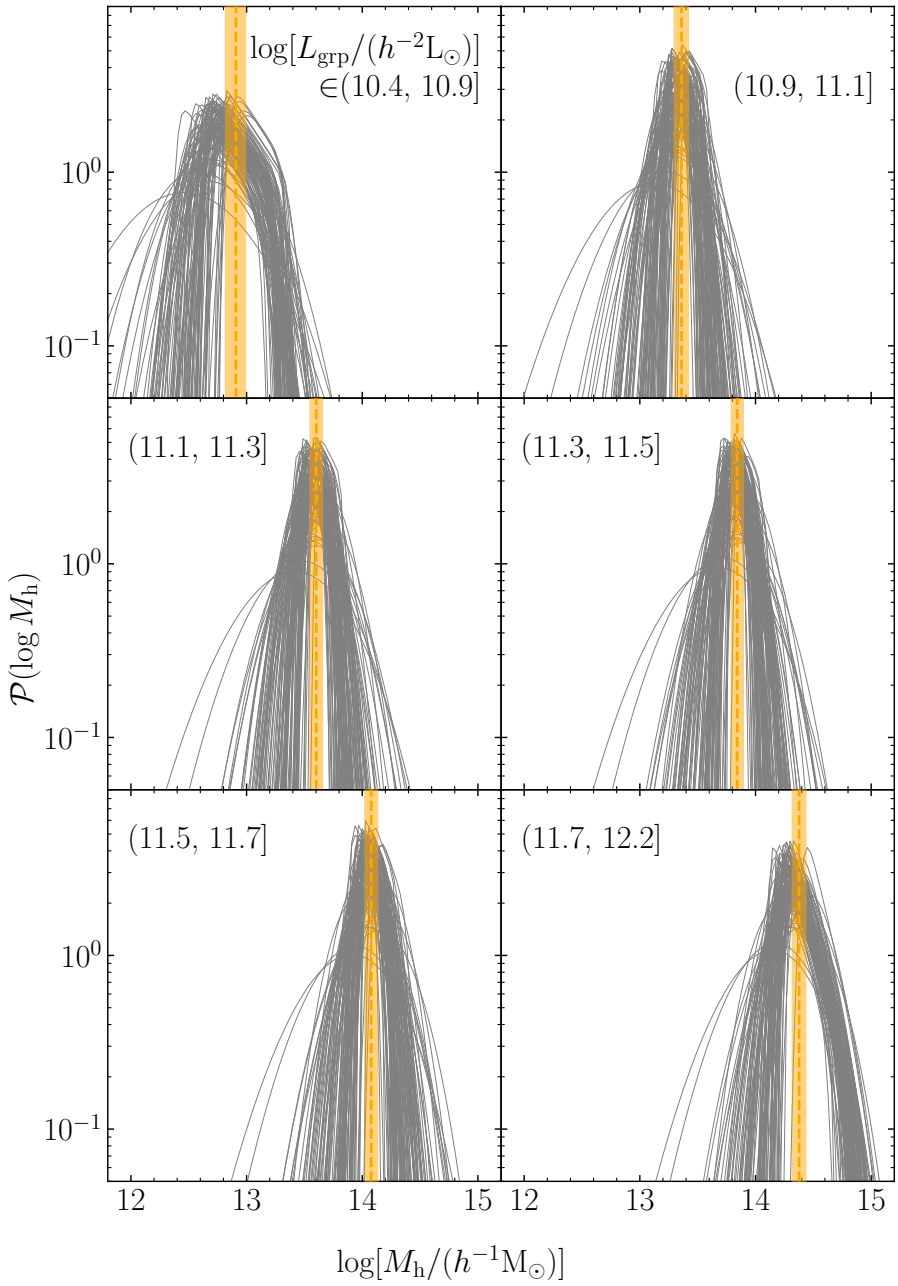

Figure 5.4: Halo mass distributions and average halo masses for each stacked bin. Grey lines illustrate the distributions derived from 100 randomly selected sets of parameter values from the sampling. Orange dashed vertical lines denote the median values of the estimated average halo masses, and the shaded regions represent the corresponding 68% credible intervals. These average halo masses were calculated from the sampled halo mass distributions, according to Eq. (5.23).

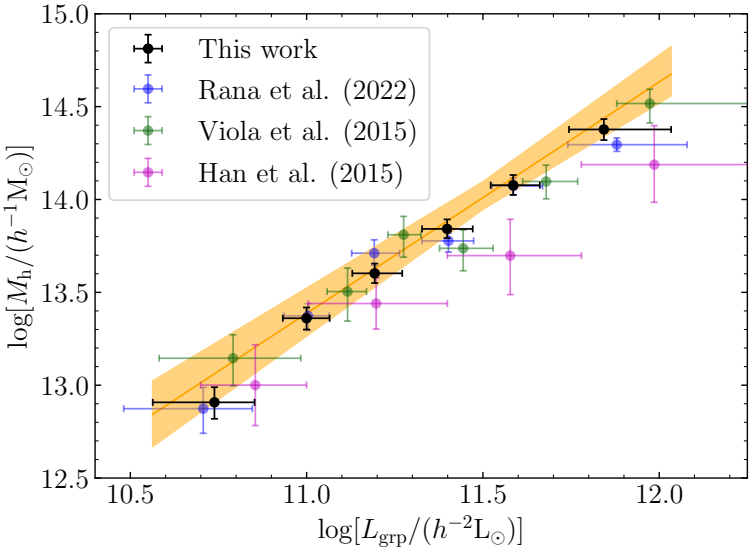

Figure 5.5: Relation between the r-band luminosity of galaxy groups and their halo masses, as determined from our weak lensing analysis. The orange line represents results derived from the median values of the estimated parameters, with the shaded regions illustrating the corresponding 68% credible intervals; the parameter values are provided in Table 5.2. The black points represent the average halo masses calculated in Sect. 5.5.1, with the error bars denoting the 68% credible intervals. The corresponding L_{grp} values for these points are derived as the median values of each stacked bin, with error bars denoting the 16th and 84th percentiles of the distribution within the bin. Results from previous weak lensing analyses based on HSC data (Rana et al. 2022), the first KiDS weak lensing data (Viola et al. 2015), and SDSS data (Han et al. 2015) are represented by blue, green, and magenta points, respectively. We note that our scaling relation is directly extracted from our CLF-based halo model, rather than fitted the estimated mean halo mass values, so the orange line is not expected to align perfectly with the black points, due to different marginalisations of the high-dimensional parameter space.

Figure 5.5 presents the scaling relation between the r-band luminosity of galaxy groups and their halo masses, as estimated from our current halo model. It also compares these results with mean halo masses estimated in Sect. 5.5.1 and from previous weak lensing analyses of the GAMA group sample, albeit with different source samples. These analyses include recent studies based on weak lensing measurements from the Subaru Hyper Suprime-Cam (HSC) survey (Rana et al. 2022), the first KiDS weak lensing data (Viola et al. 2015), and the Sloan Digital Sky Survey (SDSS, Han et al. 2015). Considering the differences in data sets and modelling procedures, the remarkable consistency between the various analyses confirms the overall validity of our new CLF-based halo model for galaxy group studies.

5.6 Model sensitivity analyses

While our current halo model generally offers a good fit to the data, as evidenced by a reduced χ^2 value of 1.06 for the best-fit model, the physical interpretation of the constrained parameters can be complex. This complexity arises primarily from the idealised assumptions inherent in our model, which may not always align with the realities of the data, particularly given the complexity of the group-finding procedure not accounted for in our current idealised model. Furthermore, the degeneracy among model parameters could introduce bias into the estimation of certain parameters when derived from projected marginalisations. These factors complicate the interpretation of our estimated model parameters and prompt questions about the robustness of our scaling relation estimates. Acknowledging these challenges, we perform sensitivity analyses in this section, focusing on two key areas of uncertainty: selection effects (Sect. 5.6.1) and mis-centring parameters (Sect. 5.6.2). These tests serve to assess not only the robustness of our current model but also to provide valuable insights for potential improvements in our future modelling efforts.

5.6.1 Sensitivity to the incompleteness

An important assumption in our CLF-based halo model is that the analysed sample is complete, implying that every halo with a specific halo mass M_h has a detected central galaxy. This assumption facilitates the integration of Eq. (5.22) to estimate the average number of galaxies for a given halo mass. Even though this is a reasonable assumption for the highly complete GAMA sample, it is worth investigating potential sample incompleteness. To address this, we introduced an incompleteness function into the integration to account for cases where certain haloes might not include a detected central galaxy:

$$\langle n_\mathrm{g}|M_\mathrm{h}\rangle = \int_{L_\mathrm{grp,min}}^{L_\mathrm{grp,max}} \mathrm{d}L_\mathrm{grp}\ \Phi(L_\mathrm{grp}|M_\mathrm{h})\ F_\mathrm{I}(M_\mathrm{h})\ , \tag{5.26}$$

where the incompleteness function $F_\mathrm{I}(M_\mathrm{h})$ is defined as:

$$F_\mathrm{I}(M_\mathrm{h}) \equiv \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{\log M_\mathrm{h}-\log M_\mathrm{erf}}{\sigma_\mathrm{erf}}\right)\right]\ , \tag{5.27}$$

with $\mathrm{erf}(M_\mathrm{h})$ being an error function with a characteristic mass of M_erf and scatter σ_erf. This incompleteness model assumes that the sample incompleteness increases or decreases monotonically with respect to the halo mass, an approach similar to those used by van Uitert et al. (2016) and Tinker et al. (2013).

This test model comprises eight free parameters. Assuming parameter independence, we obtained a reduced χ^2 value of 1.04, which marginally surpasses the 1.06 derived from our fiducial six-parameter model. Upon visualising the fitted ESD profiles, we observed that most improvements occurred in the lowest luminosity bin, which is

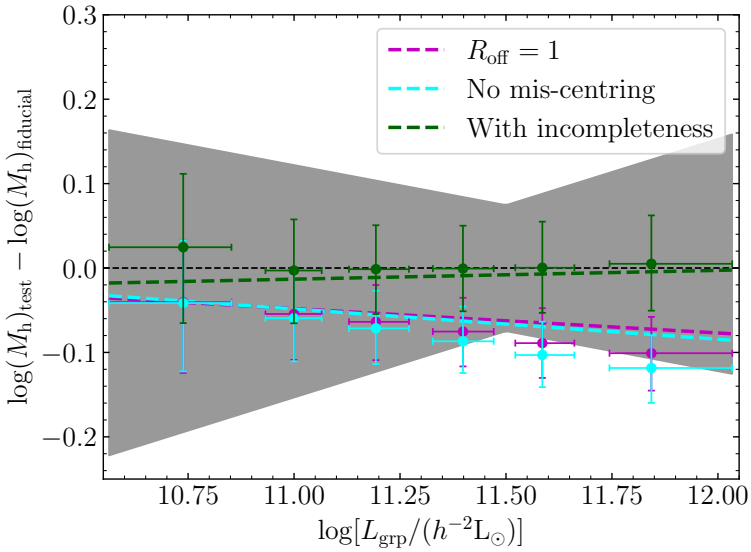

Figure 5.6: Shifts in the constrained average halo masses and scaling relation resulting from our sensitivity analyses concerning sample incompleteness and mis-centring. The shaded regions represent the 68% credible regions of the fiducial model's constraints on the scaling relation between the halo mass and group r-band luminosity. The points indicate the shifts in the constrained average halo masses, while the lines denote the shifts in the constrained scaling relations. The colour green corresponds to results from the incompleteness test (Sect. 5.6.1), whereas magenta and cyan represent results from the two mis-centring-related test models (Sect. 5.6.2). Results from the test model with fixed $p_{\rm off} = 1$ are not shown due to the lack of convergence.

expected for a flux-limited survey. However, when examining the posterior distributions, we found that the two incompleteness parameters are not constrained by the current data. Furthermore, the average halo mass estimation and the scaling relation inferred from this more complicated model closely align with the results from the fiducial model, as illustrated in Fig. 5.6. Following the principle of Occam's Razor, we conclude that introducing two more incompleteness parameters into our current model is not necessary.

5.6.2 Sensitivity to the mis-centring parameters

As previously noted in Sect. 5.4.2, the causes of mis-centring in a group sample are more intricate than what the idealised Gaussian distribution-based mis-centring model accounts for. The simplicity of our model complicates the interpretation of the estimated mis-centring parameters. This complication is further exacerbated by their degeneracy with the concentration parameter, as demonstrated in the projected contour plot of Fig. 5.7. Moreover, the posterior distribution of $\mathcal{R}_{\rm off}$ shows a tendency to lean towards the higher

limit of our already broad priors, suggesting that the simple Gaussian distribution model has difficulty describing the complex mis-centring in our data. This raises questions regarding its effect on our halo mass estimates and scaling relation constraints.

Therefore, we performed sensitivity analyses by varying the configurations of the mis-centring parameters within the framework of the current model. The exploration of more realistic mis-centring models will be left to our future analyses. We tested three alternate setups of the mis-centring parameters: fixing $p_{\rm off} = 1$ and only varying $\mathcal{R}_{\rm off}$ to examine if a single Gaussian distribution suffices for describing the central galaxy position distribution of the entire sample; fixing $\mathcal{R}_{\rm off} = 1$ and only varying $p_{\rm off}$ to explore if altering the mis-centred fraction alone can adequately describe the mis-centring scenario; and setting both $p_{\rm off}$ and $\mathcal{R}_{\rm off}$ to zero to ascertain whether the mis-centring effect can be ignored in the current analysis.

We first checked the posterior distributions of test runs and compared them to those from the fiducial model, as shown in Fig. 5.7. Except for the case where we fix $p_{\rm off} = 1$, the other two alternatives yield well-constrained model parameters. The lack of convergence in the case with fixed $p_{\rm off} = 1$ indicates that a single Gaussian distribution fails to describe the central galaxy position distribution of the entire sample, suggesting that most central galaxies remain close to the group's true centre and thus follow a much sharper distribution. We also found that changes in the mis-centring parameters affect the constraints on the concentration parameter $f_{\rm c}$ due to their degeneracy. Specifically, downplaying or disregarding mis-centring effects drives $f_{\rm c}$ towards lower values, warranting caution in interpreting the concentration parameter constrained in the current model, a finding consistent with Viola et al. (2015).

Next, we examined the fitted ESD profiles and the best-fit χ^2 values. The test models yielded slightly sub-optimal fits, with reduced χ^2 values of 1.34 and 1.35 for the cases with $\mathcal{R}_{\rm off} = 1$ and without mis-centring, respectively. Upon visually inspecting the fitted ESD profiles, we found that the noticeable degradation in fit occurs in the small separation measurements in the higher $L_{\rm grp}$ bins. These results imply that a model lacking proper consideration for the mis-centring of selected central galaxies would struggle to accurately describe the innermost regions of the measured ESD signals.

Finally, we investigated the impact of these test models on the estimations of average halo mass and scaling relations. As illustrated in Fig. 5.6, alterations in the setups of the mis-centring parameters affect both constraints, particularly at the high mass end. Interestingly, the scaling relation, directly constrained by our CLF-based halo model, demonstrates a greater robustness against changes in the mis-centring parameters. The most significant shift remains within the 68% credible regions of the fiducial constraints. In contrast, the average halo mass exhibits shifts that slightly exceed the 68% credible regions of the fiducial constraints. This outcome lends additional credence to our approach of directly constraining the scaling relations from the model, as opposed to fitting them to the measured average values.

These tests, which focus on the mis-centring parameters, underscore the importance of properly accounting for mis-centring within the group sample. Our current fiducial model appears to outperform all test models due to its better fit to the measured signals.

However, the impact on the scaling relation constraints from these model variations is concerning for upcoming, more precise measurements, and underscores the need for the development of a more sophisticated mis-centring model. We plan to address this in our future analyses, using hydrodynamic simulations as a tool for model refinement.

5.7 Conclusions

We conducted a galaxy-galaxy weak lensing analysis using the KiDS-1000 data, aiming to constrain the scaling relation between the group luminosity and their halo masses for galaxy groups identified by GAMA. The size of our sample is nearly double of that used in the similar work conducted by Viola et al. (2015). Furthermore, we updated their empirical Halo Occupation Distribution (HOD) model with a more physically motivated Conditional Luminosity Function (CLF, Yang et al. 2003; Vale & Ostriker 2004; Cooray 2006). Our new CLF-based halo model incorporates six free parameters, compared to the eleven free parameters in the previous model. Most importantly, it allows for direct sampling of the scaling relation parameters during the main modelling procedure, thereby bypassing the need for an additional regression process based on the average halo mass estimates as required in previous studies. Despite having considerably fewer free parameters, our current model still provides a good fit to the Excess Surface Density (ESD) measurements, with a best-fit reduced χ^2 value of 1.06.

Our current fiducial model yields a power-law relation between the halo mass and the r-band luminosity of the group:

$$\frac{M_\mathrm{h}}{10^{14}\ h^{-1}\mathrm{M}_\odot} = \left(1.02^{+0.22}_{-0.14}\right)\left(\frac{L_\mathrm{grp}}{10^{11.5}\ h^{-2}\mathrm{L}_\odot}\right)^{1.25^{+0.12}_{-0.10}} . \tag{5.28}$$

These results align with previous findings from Viola et al. (2015), who reported a normalisation of 0.95 ± 0.14 and a power-law index of 1.16 ± 0.13, and Rana et al. (2022), who reported a normalisation of 0.81 ± 0.04 and a power-law index of 1.01 ± 0.07.

While our overall sample size is larger than that used by Viola et al. (2015), the uncertainties of our final constraints are comparable to theirs. This is largely because we applied a more stringent scale cut to alleviate blending effects on small scales—we used a scale cut of 0.03 h^{-1}Mpc compared to their 0.02 h^{-1}Mpc. Additionally, we excluded the tails of the L_grp distributions to mitigate potential group detection effects, as shown in Fig. 5.2. In this sense, we consciously traded some statistical power for increased robustness. Crucially, our sensitivity tests revealed that uncertainties in certain elements of our current halo model do have an impact on our results, although the impact is within acceptable limits given the current measurement uncertainties. This underscores the necessity for further model refinement to prepare for future analyses with more precise measurements.

Most importantly, we acknowledge the limitations of the current model in accurately representing the innermost regions of the ESD measurements. These small-scale measurements exhibit strong sensitivity to both the mass-concentration and mis-centring

parameters. Moreover, these parameters display similar impacts on the measured ESD signals, leading to a degeneracy between them in the current model, a finding also reported by Viola et al. (2015). Our sensitivity tests further revealed that alterations to the configuration of the mis-centring parameters can impact the constraints on the scaling relation to a degree that is only marginally acceptable for the current analysis. Given the complex origins of mis-centring in a galaxy group sample, as noted in Sect. 5.4.2 and also in Ahad et al. (2023) and Jakobs et al. (2018), we believe the development of a more realistic mis-centring model is one of the crucial aspects for future analyses.

Such advancement can be achieved through the analysis of galaxy-galaxy lensing measurements and modelling within the context of cosmological simulations. With the advent of new large-volume hydrodynamic simulations (e.g. Schaye et al. 2023; Kugel et al. 2023), we are well-positioned to refine the current modelling framework. This can be accomplished by replacing certain uninformative or idealised model elements with treatments that are more informed by simulations. During these simulation-based explorations, it is essential to ensure that data selection effects are accurately replicated when constructing mock data from simulations (e.g. Jakobs et al. 2018). We plan to pursue this direction in our future analyses.

5.A Posterior distributions of all free model parameters

In this appendix, we present contour plots displaying the posterior distributions of all free parameters derived from our fiducial model (Sect. 5.4), along with two test models related to the mis-centring parameters (Sect. 5.6.2). We did not show results from the test model with fixed $p_{\rm off} = 1$ due to its lack of convergence. Likewise, we did not include the results from the incompleteness test as it yields constraints almost identical to those of our fiducial model in the same parameters, and the two new incompleteness parameters are not well-constrained.

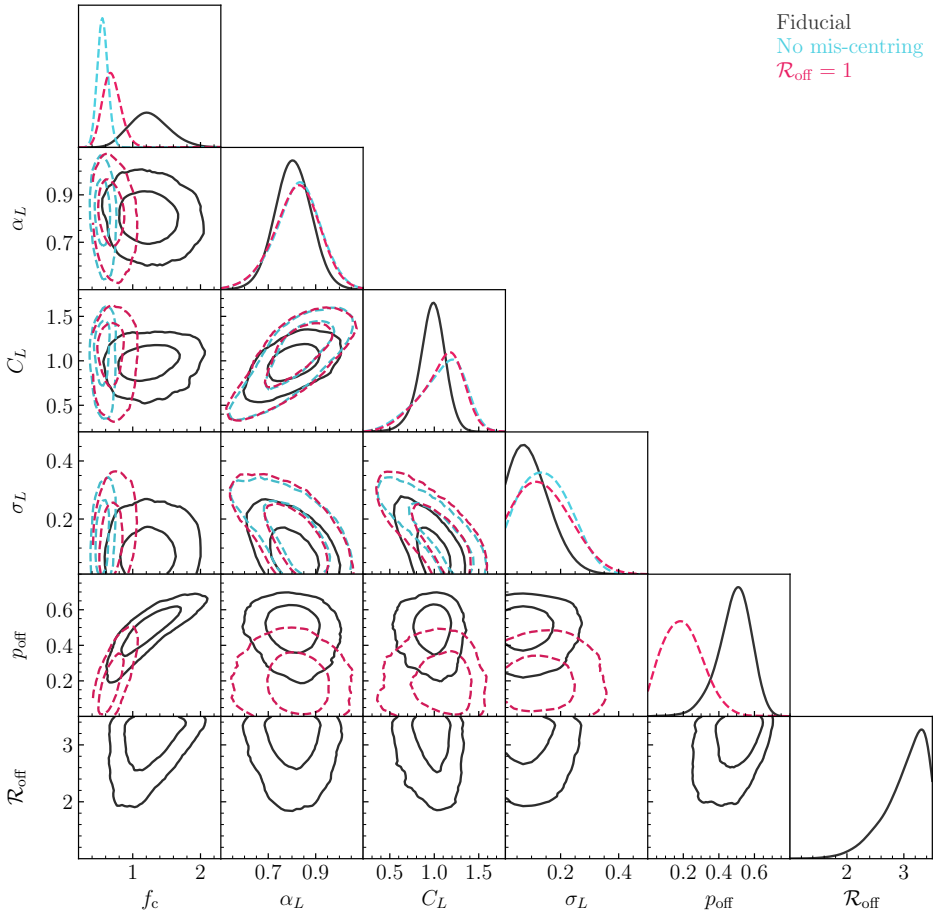

Figure 5.7: Posterior distributions of all free parameters derived from our fiducial model (Sect. 5.4), along with two test models related to the mis-centring parameters (Sect. 5.6.2). The black solid lines signify results from the fiducial six-parameter model, while the magenta and cyan dashed lines represent results from the two mis-centring-related test models. The contours denote the 68% and 95% credible intervals and have been smoothed using a Gaussian kernel density estimation with a bandwidth scaled by a factor of 1. These plots were generated using the ChainConsumer package (Hinton 2016).

Bibliography

Aas, K., Czado, C., Frigessi, A., & Bakken, H. 2009, Insurance: Mathematics and economics, 44, 182

Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2022, Phys. Rev. D, 105, 023520

Ahad, S. L., Bahé, Y. M., & Hoekstra, H. 2023, MNRAS, 518, 3685

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4

Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12

Alarcon, A., Gaztanaga, E., Eriksen, M., et al. 2021, MNRAS, 501, 6103

Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, ARA&A, 49, 409

Alsing, J., Kirk, D., Heavens, A., & Jaffe, A. H. 2015, MNRAS, 452, 1202

Amon, A. & Efstathiou, G. 2022, MNRAS, 516, 5355

Amon, A., Gruen, D., Troxel, M. A., et al. 2022, Phys. Rev. D, 105, 023514

Asgari, M., Heymans, C., Hildebrandt, H., et al. 2019, A&A, 624, A134

Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, A&A, 645, A104

Asgari, M., Tröster, T., Heymans, C., et al. 2020, A&A, 634, A127

Audren, B., Lesgourgues, J., Benabed, K., & Prunet, S. 2013, J. Cosmology Astropart. Phys., 2013, 001

Bacon, D. J., Massey, R. J., Refregier, A. R., & Ellis, R. S. 2003, MNRAS, 344, 673

Bacon, D. J., Refregier, A. R., & Ellis, R. S. 2000, MNRAS, 318, 625

Bahar, Y. E., Bulbul, E., Clerc, N., et al. 2022, A&A, 661, A7

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15

Barnes, D. J., Vogelsberger, M., Pearce, F. A., et al. 2021, MNRAS, 506, 2533

Bartelmann, M. 1995, A&A, 303, 643

Bartelmann, M. 1996, A&A, 313, 697

Bartelmann, M. & Schneider, P. 2001, Phys. Rep., 340, 291

Bedford, T. & Cooke, R. M. 2002, The Annals of Statistics, 30, 1031

Benítez, N. 2000, ApJ, 536, 571

Beran, R. 1977, The Annals of Statistics, 5, 445

Berlind, A. A. & Weinberg, D. H. 2002, ApJ, 575, 587

Bernstein, G. M. & Armstrong, R. 2014, MNRAS, 438, 1880

Bernstein, G. M. & Jarvis, M. 2002, AJ, 123, 583

Bertin, E. 2010, SWarp: Resampling and Co-adding FITS Images Together

Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393

Biffi, V., Borgani, S., Murante, G., et al. 2016, ApJ, 827, 112

Blandford, R. & Narayan, R. 1986, ApJ, 310, 568

Blandford, R. D., Saust, A. B., Brainerd, T. G., & Villumsen, J. V. 1991, MNRAS, 251, 600

Blas, D., Lesgourgues, J., & Tram, T. 2011, J. Cosmology Astropart. Phys., 2011, 034

Blazek, J. A., MacCrann, N., Troxel, M. A., & Fang, X. 2019, Phys. Rev. D, 100, 103506

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440

Bravo, M., Lagos, C. d. P., Robotham, A. S. G., Bellstedt, S., & Obreschkow, D. 2020, MNRAS, 497, 3026

Bridle, S., Balan, S. T., Bethge, M., et al. 2010, MNRAS, 405, 2044

Bridle, S. & King, L. 2007, New Journal of Physics, 9, 444

Brinckmann, T. & Lesgourgues, J. 2018, arXiv e-prints, arXiv:1804.07261

Broadhurst, T. J., Taylor, A. N., & Peacock, J. A. 1995, ApJ, 438, 49

Brouwer, M. M., Cacciato, M., Dvornik, A., et al. 2016, MNRAS, 462, 4451

Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000

Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125

Bullock, J. S., Kolatt, T. S., Sigad, Y., et al. 2001, MNRAS, 321, 559

Cañas, R., Elahi, P. J., Welker, C., et al. 2019, MNRAS, 482, 2039

Cacciato, M., van den Bosch, F. C., More, S., et al. 2009, MNRAS, 394, 929

Capak, P. L. 2004, PhD thesis, UNIVERSITY OF HAWAI'I

Cappellaro, E., Botticella, M. T., Pignata, G., et al. 2015, A&A, 584, A62

Carlstrom, J. E., Ade, P. A. R., Aird, K. A., et al. 2011, PASP, 123, 568

Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, MNRAS, 447, 646

Carretero, J., Tallada, P., Casals, J., et al. 2017, in Proceedings of the European Physical Society Conference on High Energy Physics. 5-12 July, 488

Castro, T., Borgani, S., Dolag, K., et al. 2021, MNRAS, 500, 2316

Chabrier, G. 2003, PASP, 115, 763

Chang, C., Jarvis, M., Jain, B., et al. 2013, MNRAS, 434, 2121

Chauhan, G., Lagos, C. d. P., Obreschkow, D., et al. 2019, MNRAS, 488, 5898

Chisari, N., Codis, S., Laigle, C., et al. 2015, MNRAS, 454, 2736

Chisari, N. E., Richardson, M. L. A., Devriendt, J., et al. 2018, MNRAS, 480, 3962

Coe, D., Benítez, N., Sánchez, S. F., et al. 2006, AJ, 132, 926

Coleman, G. D., Wu, C. C., & Weedman, D. W. 1980, ApJS, 43, 393

Cooray, A. 2006, MNRAS, 365, 842

Cooray, A. & Sheth, R. 2002, Phys. Rep., 372, 1

Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, MNRAS, 453, 1513

Czado, C. 2019, Lecture notes in statistics, Vol. 222, Analyzing Dependent Data with Vine Copulas: A Practical Guide with R (Cham: Springer International Publishing AG)

Czekaj, M. A., Robin, A. C., Figueras, F., Luri, X., & Haywood, M. 2014, A&A, 564, A102

Dalal, R., Li, X., Nicola, A., et al. 2023, arXiv e-prints, arXiv:2304.00701

Damjanov, I., Zahid, H. J., Geller, M. J., Fabricant, D. G., & Hwang, H. S. 2018, ApJS, 234, 21

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270

Davies, L. J. M., Driver, S. P., Robotham, A. S. G., et al. 2015, MNRAS, 447, 1014

Davies, L. J. M., Robotham, A. S. G., Driver, S. P., et al. 2018, MNRAS, 480, 768

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371

Dawson, W. A., Schneider, M. D., Tyson, J. A., & Jee, M. J. 2016, ApJ, 816, 11

De Cicco, D., Paolillo, M., Falocco, S., et al. 2019, A&A, 627, A33

de Jong, J. T. A., Verdoes Kleijn, G. A., Boxhoorn, D. R., et al. 2015, A&A, 582, A62

de Jong, J. T. A., Verdoes Kleijn, G. A., Erben, T., et al. 2017, A&A, 604, A134

de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, Experimental Astronomy, 35, 25

De Propris, R., West, M. J., Andrade-Santos, F., et al. 2021, MNRAS, 500, 310

Debackere, S. N. B., Schaye, J., & Hoekstra, H. 2020, MNRAS, 492, 2285

Debackere, S. N. B., Schaye, J., & Hoekstra, H. 2021, MNRAS, 505, 593

DeRose, J., Wechsler, R. H., Becker, M. R., et al. 2022, Phys. Rev. D, 105, 123520

DES and KiDS Collaboration, :, Abbott, T. M. C., et al. 2023, arXiv e-prints, arXiv:2305.17173

Desjacques, V., Jeong, D., & Schmidt, F. 2018, Phys. Rep., 733, 1

Di Valentino, E., Mena, O., Pan, S., et al. 2021, Classical and Quantum Gravity, 38, 153001

Driver, S. P., Bellstedt, S., Robotham, A. S. G., et al. 2022, MNRAS, 513, 439

Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, MNRAS, 413, 971

Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, MNRAS, 390, L64

Dunkley, J., Komatsu, E., Nolta, M. R., et al. 2009, ApJS, 180, 306

Dvornik, A., Cacciato, M., Kuijken, K., et al. 2017, MNRAS, 468, 3251

Eckmiller, H. J., Hudson, D. S., & Reiprich, T. H. 2011, A&A, 535, A105

Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32

Efstathiou, G. & Lemos, P. 2018, MNRAS, 476, 151

Eisenstein, D. J. & Hu, W. 1998, ApJ, 496, 605

Elahi, P. J., Cañas, R., Poulton, R. J. J., et al. 2019a, PASA, 36, e021

Elahi, P. J., Poulton, R. J. J., Tobar, R. J., et al. 2019b, PASA, 36, e028

Elahi, P. J., Welker, C., Power, C., et al. 2018, MNRAS, 475, 5338

Erben, T., Schirmer, M., Dietrich, J. P., et al. 2005, Astronomische Nachrichten, 326, 432

Ettori, S., Donnarumma, A., Pointecouteau, E., et al. 2013, Space Sci. Rev., 177, 119

Euclid Collaboration, Guglielmo, V., Saglia, R., et al. 2020, A&A, 642, A192

Euclid Collaboration, Knabenhans, M., Stadel, J., et al. 2021, MNRAS, 505, 2840

Euclid Collaboration, Martinet, N., Schrabback, T., et al. 2019, A&A, 627, A59

Farrow, D. J., Cole, S., Norberg, P., et al. 2015, MNRAS, 454, 2120

Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, MNRAS, 467, 1627

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601

Fischbacher, S., Kacprzak, T., Blazek, J., & Refregier, A. 2023, J. Cosmology Astropart. Phys., 2023, 033

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306

Fortuna, M. C., Hoekstra, H., Joachimi, B., et al. 2021a, MNRAS, 501, 2983

Fortuna, M. C., Hoekstra, H., Johnston, H., et al. 2021b, A&A, 654, A76

Fosalba, P., Crocce, M., Gaztañaga, E., & Castander, F. J. 2015a, MNRAS, 448, 2987

Fosalba, P., Gaztañaga, E., Castander, F. J., & Crocce, M. 2015b, MNRAS, 447, 1319

Fowler, J. W., Niemack, M. D., Dicker, S. R., et al. 2007, Appl. Opt., 46, 3444

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

George, M. R., Leauthaud, A., Bundy, K., et al. 2012, ApJ, 757, 2

Giblin, B., Heymans, C., Asgari, M., et al. 2021, A&A, 645, A105

Girardi, L., Groenewegen, M. A. T., Hatziminaoglou, E., & da Costa, L. 2005, A&A, 436, 895

González-Fernández, C., Hodgkin, S. T., Irwin, M. J., et al. 2018, MNRAS, 474, 5459

Goodman, J. & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759

Griffith, R. L., Cooper, M. C., Newman, J. A., et al. 2012, ApJS, 200, 9

Hamana, T., Shirasaki, M., Miyazaki, S., et al. 2020, PASJ, 72, 16

Han, J., Eke, V. R., Frenk, C. S., et al. 2015, MNRAS, 446, 1356

Handley, W. & Lemos, P. 2019a, Phys. Rev. D, 100, 023512

Handley, W. & Lemos, P. 2019b, Phys. Rev. D, 100, 043504

Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015a, MNRAS, 450, L61

Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015b, MNRAS, 453, 4384

Hartlap, J., Hilbert, S., Schneider, P., & Hildebrandt, H. 2011, A&A, 528, A51

Hasinger, G., Capak, P., Salvato, M., et al. 2018, ApJ, 858, 77

Heitmann, K., Higdon, D., White, M., et al. 2009, ApJ, 705, 156

Hellwing, W. A., Schaller, M., Frenk, C. S., et al. 2016, MNRAS, 461, L11

Heymans, C., Tröster, T., Asgari, M., et al. 2021, A&A, 646, A140

Heymans, C., Van Waerbeke, L., Bacon, D., et al. 2006, MNRAS, 368, 1323

Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, MNRAS, 427, 146

Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43

Hilbert, S., Hartlap, J., White, S. D. M., & Schneider, P. 2009, A&A, 499, 31

Hilbert, S., Xu, D., Schneider, P., et al. 2017, MNRAS, 468, 790

Hildebrandt, H., Arnouts, S., Capak, P., et al. 2010, A&A, 523, A31

Hildebrandt, H., Choi, A., Heymans, C., et al. 2016, MNRAS, 463, 635

Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69

Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, A&A, 647, A124

Hildebrandt, H., van Waerbeke, L., & Erben, T. 2009, A&A, 507, 683

Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454

Hill, A. R., van der Wel, A., Franx, M., et al. 2019, ApJ, 871, 76

Hinton, S. R. 2016, The Journal of Open Source Software, 1, 00045

Hirata, C. & Seljak, U. 2003, MNRAS, 343, 459

Hirata, C. M. & Seljak, U. 2004, Phys. Rev. D, 70, 063526

Hoekstra, H., Franx, M., Kuijken, K., et al. 2001, ApJ, 548, L5

Hoekstra, H., Herbonnet, R., Muzzin, A., et al. 2015, MNRAS, 449, 685

Hoekstra, H. & Jain, B. 2008, Annual Review of Nuclear and Particle Science, 58, 99

Hoekstra, H., Kannawadi, A., & Kitching, T. D. 2021, A&A, 646, A124

Hoekstra, H., Viola, M., & Herbonnet, R. 2017, MNRAS, 468, 3295

Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, J. Cosmology Astropart. Phys., 2012, 027

Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, MNRAS, 478, 592

Hu, W. 1999, ApJ, 522, L21

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Hudson, M. J., Gillis, B. R., Coupon, J., et al. 2015, MNRAS, 447, 298

Huff, E. & Mandelbaum, R. 2017, arXiv e-prints, arXiv:1702.02600

Huterer, D. 2002, Phys. Rev. D, 65, 063001

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111

Jain, B. & Taylor, A. 2003, Phys. Rev. Lett., 91, 141302

Jakobs, A., Viola, M., McCarthy, I., et al. 2018, MNRAS, 480, 3338

Jarvis, M. 2015, TreeCorr: Two-point correlation functions, Astrophysics Source Code Library, record ascl:1508.007

Jarvis, M., Bernstein, G., & Jain, B. 2004, MNRAS, 352, 338

Jarvis, M., Sheldon, E., Zuntz, J., et al. 2016, MNRAS, 460, 2245

Joachimi, B., Cacciato, M., Kitching, T. D., et al. 2015, Space Sci. Rev., 193, 1

Joachimi, B., Lin, C. A., Asgari, M., et al. 2021, A&A, 646, A129

Joachimi, B., Mandelbaum, R., Abdalla, F. B., & Bridle, S. L. 2011, A&A, 527, A26

Joachimi, B., Schneider, P., & Eifler, T. 2008, A&A, 477, 43

Joe, H. 2014, Dependence modeling with copulas (CRC press)

Johnston, H., Georgiou, C., Joachimi, B., et al. 2019, A&A, 624, A30

Kafle, P. R., Robotham, A. S. G., Driver, S. P., et al. 2018, MNRAS, 479, 3746

Kaiser, N. 1992, ApJ, 388, 272

Kaiser, N. 1998, ApJ, 498, 26

Kaiser, N. 2000, ApJ, 537, 555

Kaiser, N., Wilson, G., & Luppino, G. A. 2000, Large-Scale Cosmic Shear Measurements

Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, A&A, 624, A92

Kelly, B. C. 2007, ApJ, 665, 1489

Kettula, K., Giodini, S., van Uitert, E., et al. 2015, MNRAS, 451, 1460

Kilbinger, M. 2015, Reports on Progress in Physics, 78, 086901

Kinney, A. L., Calzetti, D., Bohlin, R. C., et al. 1996, ApJ, 467, 38

Kitching, T. D., Balan, S. T., Bridle, S., et al. 2012, MNRAS, 423, 3163

Kitching, T. D., Miller, L., Heymans, C. E., van Waerbeke, L., & Heavens, A. F. 2008, MNRAS, 390, 149

Kitching, T. D., Paykari, P., Hoekstra, H., & Cropper, M. 2019, The Open Journal of Astrophysics, 2, 5

Köhlinger, F., Joachimi, B., Asgari, M., et al. 2019, MNRAS, 484, 3126

Kohonen, T. 1982, Biological cybernetics, 43, 59

Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330

Krause, E., Fang, X., Pandey, S., et al. 2021, arXiv e-prints, arXiv:2105.13548

Kugel, R., Schaye, J., Schaller, M., et al. 2023, arXiv e-prints, arXiv:2306.05492

Kuijken, K. 2011, The Messenger, 146, 8

Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, A&A, 625, A2

Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, MNRAS, 454, 3500

Kurita, T., Takada, M., Nishimichi, T., et al. 2021, MNRAS, 501, 833

Lacey, C. & Cole, S. 1993, MNRAS, 262, 627

Lagos, C. d. P., Robotham, A. S. G., Trayford, J. W., et al. 2019, MNRAS, 489, 4196

Lagos, C. d. P., Tobar, R. J., Robotham, A. S. G., et al. 2018, MNRAS, 481, 3573

Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24

Lange, R., Moffett, A. J., Driver, S. P., et al. 2016, MNRAS, 462, 1470

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193

Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14

Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, A&A, 576, A79

Leauthaud, A., Finoguenov, A., Kneib, J.-P., et al. 2010, ApJ, 709, 97

Lemaître, G. 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Lemos, P., Köhlinger, F., Handley, W., et al. 2019, arXiv e-prints, arXiv:1910.07820

Lemos, P., Weaverdyck, N., Rollins, R. P., et al. 2023, MNRAS, 521, 1184

Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473

Li, S.-S., Hoekstra, H., Kuijken, K., et al. 2023a, arXiv e-prints, arXiv:2306.11124

Li, S.-S., Kuijken, K., Hoekstra, H., et al. 2021, A&A, 646, A175

Li, S.-S., Kuijken, K., Hoekstra, H., et al. 2023b, A&A, 670, A100

Li, X., Zhang, T., Sugiyama, S., et al. 2023c, arXiv e-prints, arXiv:2304.00702

Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, ApJS, 184, 218

Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, MNRAS, 390, 118

Limber, D. N. 1953, ApJ, 117, 134

Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, MNRAS, 452, 2087

Logan, C. H. A., Maughan, B. J., Diaferio, A., et al. 2022, A&A, 665, A124

Loverde, M. & Afshordi, N. 2008, Phys. Rev. D, 78, 123506

MacCrann, N., Becker, M. R., McCullough, J., et al. 2022, MNRAS, 509, 3371

Mandelbaum, R. 2018, ARA&A, 56, 393

Mandelbaum, R., Lanusse, F., Leauthaud, A., et al. 2018, MNRAS, 481, 3170

Mandelbaum, R., Rowe, B., Armstrong, R., et al. 2015, MNRAS, 450, 2963

Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, MNRAS, 368, 715

Massey, R. 2010, MNRAS, 409, L109

Massey, R., Heymans, C., Bergé, J., et al. 2007, MNRAS, 376, 13

Massey, R., Hoekstra, H., Kitching, T., et al. 2013, MNRAS, 429, 661

Masters, D., Capak, P., Stern, D., et al. 2015, ApJ, 813, 53

Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2017, ApJ, 841, 111

Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, ApJ, 877, 81

McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, MNRAS, 465, 2936

McCarthy, I. G., Schaye, J., Ponman, T. J., et al. 2010, MNRAS, 406, 822

McConnell, N. J. & Ma, C.-P. 2013, ApJ, 764, 184

McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156

McFarland, J. P., Verdoes-Kleijn, G., Sikkema, G., et al. 2013, Experimental Astronomy, 35, 45

Mead, A. J., Brieden, S., Tröster, T., & Heymans, C. 2021, MNRAS, 502, 1401

Mead, A. J., Heymans, C., Lombriser, L., et al. 2016, MNRAS, 459, 1468

Mead, A. J., Peacock, J. A., Heymans, C., Joudaki, S., & Heavens, A. F. 2015, MNRAS, 454, 1958

Melchior, P. & Viola, M. 2012, MNRAS, 424, 2757

Mellier, Y. 1999, ARA&A, 37, 127

Miller, L., Heymans, C., Kitching, T. D., et al. 2013, MNRAS, 429, 2858

Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, MNRAS, 382, 315

Miralda-Escude, J. 1991, ApJ, 380, 1

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution

More, S., Miyatake, H., Mandelbaum, R., et al. 2015, ApJ, 806, 2

Muñoz-Cuartas, J. C., Macciò, A. V., Gottlöber, S., & Dutton, A. A. 2011, MNRAS, 411, 584

Myles, J., Alarcon, A., Amon, A., et al. 2021, MNRAS, 505, 4249

Narayan, R. & Bartelmann, M. 1997, Lectures on Gravitational Lensing

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1995, MNRAS, 275, 720

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493

Nelder, J. A. & Mead, R. 1965, The Computer Journal, 7, 308

Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5

Nishimichi, T., Takada, M., Takahashi, R., et al. 2019, ApJ, 884, 29

Obreschkow, D., Klöckner, H. R., Heywood, I., Levrier, F., & Rawlings, S. 2009, ApJ, 703, 1890

Oguri, M. & Takada, M. 2011, Phys. Rev. D, 83, 023008

Paulin-Henriksson, S., Amara, A., Voigt, L., Refregier, A., & Bridle, S. L. 2008, A&A, 484, 67

Peacock, J. A. 1999, Cosmological Physics

Peacock, J. A. & Smith, R. E. 2000, MNRAS, 318, 1144

Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419

Perivolaropoulos, L. & Skara, F. 2022, New A Rev., 95, 101659

Planck Collaboration. 2016, A&A, 594, A13

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6

Pop, A.-R., Hernquist, L., Nagai, D., et al. 2022, arXiv e-prints, arXiv:2205.11528

Poulton, R. J. J., Robotham, A. S. G., Power, C., & Elahi, P. J. 2018, PASA, 35, e042

Press, W. H. & Schechter, P. 1974, ApJ, 187, 425

Preston, C., Amon, A., & Efstathiou, G. 2023, arXiv e-prints, arXiv:2305.09827

Raichoor, A., Mei, S., Erben, T., et al. 2014, ApJ, 797, 102

Rana, D., More, S., Miyatake, H., et al. 2022, MNRAS, 510, 5408

Rasia, E., Ettori, S., Moscardini, L., et al. 2006, MNRAS, 369, 2013

Rau, M. M., Dalal, R., Zhang, T., et al. 2022, arXiv e-prints, arXiv:2211.16516

Raveri, M., Zacharegkas, G., & Hu, W. 2020, Phys. Rev. D, 101, 103527

Refregier, A. 2003, ARA&A, 41, 645

Refregier, A. & Amara, A. 2014, Physics of the Dark Universe, 3, 1

Refregier, A., Kacprzak, T., Amara, A., Bridle, S., & Rowe, B. 2012, MNRAS, 425, 1951

Rhodes, J. D., Massey, R. J., Albert, J., et al. 2007, ApJS, 172, 203

Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523

Robotham, A. S. G., Bellstedt, S., Lagos, C. d. P., et al. 2020, MNRAS, 495, 905

Robotham, A. S. G., Norberg, P., Driver, S. P., et al. 2011, MNRAS, 416, 2640

Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, Astronomy and Computing, 10, 121

Salvato, M., Ilbert, O., & Hoyle, B. 2019, Nature Astronomy, 3, 212

Samuroff, S., Blazek, J., Troxel, M. A., et al. 2019, MNRAS, 489, 5453

Samuroff, S., Bridle, S. L., Zuntz, J., et al. 2018, MNRAS, 475, 4524

Samuroff, S., Mandelbaum, R., Blazek, J., et al. 2022, arXiv e-prints, arXiv:2212.11319

Schaye, J., Kugel, R., Schaller, M., et al. 2023, arXiv e-prints, arXiv:2306.04024

Schirmer, M. 2013, ApJS, 209, 21

Schneider, A., Stoira, N., Refregier, A., et al. 2020, J. Cosmology Astropart. Phys., 2020, 019

Schneider, P. 1985, A&A, 143, 413

Schneider, P. 2016, A&A, 592, L6

Schneider, P., Eifler, T., & Krause, E. 2010, A&A, 520, A116

Schneider, P., King, L., & Erben, T. 2000, A&A, 353, 41

Schneider, P., van Waerbeke, L., & Mellier, Y. 2002, A&A, 389, 729

Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, ApJ, 546, 20

Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1

Secco, L. F., Samuroff, S., Krause, E., et al. 2022, Phys. Rev. D, 105, 023515

Seljak, U. 2000, MNRAS, 318, 203

Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M. P., & McCarthy, I. G. 2011, MNRAS, 417, 2020

Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Sheldon, E. S., Becker, M. R., MacCrann, N., & Jarvis, M. 2020, ApJ, 902, 138

Sheldon, E. S. & Huff, E. M. 2017, ApJ, 841, 24

Sifón, C., Cacciato, M., Hoekstra, H., et al. 2015, MNRAS, 454, 3938

Silverman, J. D., Kashino, D., Sanders, D., et al. 2015, ApJS, 220, 12

Singh, S., Mandelbaum, R., & More, S. 2015, MNRAS, 450, 2195

Skibba, R. A., van den Bosch, F. C., Yang, X., et al. 2011, MNRAS, 410, 417

Sklar, M. 1959, Publ. inst. statist. univ. Paris, 8, 229

Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, arXiv:1503.03757

Stanford, S. A., Masters, D., Darvish, B., et al. 2021, ApJS, 256, 9

Sugiyama, N. 1995, ApJS, 100, 281

Sunyaev, R. A. & Zeldovich, Y. B. 1972, Comments on Astrophysics and Space Physics, 4, 173

Takada, M. & Hu, W. 2013, Phys. Rev. D, 87, 123504

Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, ApJ, 761, 152

Tallada, P., Carretero, J., Casals, J., et al. 2020, Astronomy and Computing, 32, 100391

Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2018, PASJ, 70, S9

Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, MNRAS, 418, 1587

Tenneti, A., Mandelbaum, R., Di Matteo, T., Feng, Y., & Khandai, N. 2014, MNRAS, 441, 470

Tinker, J. L., Leauthaud, A., Bundy, K., et al. 2013, ApJ, 778, 93

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878

Toyozumi, H. & Ashley, M. C. B. 2005, PASA, 22, 257

Tröster, T., Asgari, M., Blake, C., et al. 2021, A&A, 649, A88

Trotta, R. 2008, Contemporary Physics, 49, 71

Troxel, M. A. & Ishak, M. 2015, Phys. Rep., 558, 1

Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018, Phys. Rev. D, 98, 043528

Tyson, J. A., Valdes, F., & Wenk, R. A. 1990, ApJ, 349, L1

Unruh, S., Schneider, P., & Hilbert, S. 2019, A&A, 623, A94

Vale, A. & Ostriker, J. P. 2004, MNRAS, 353, 189

van Daalen, M. P., McCarthy, I. G., & Schaye, J. 2020, MNRAS, 491, 2424

van Daalen, M. P., Schaye, J., Booth, C. M., & Dalla Vecchia, C. 2011, MNRAS, 415, 3649

van den Bosch, F. C., More, S., Cacciato, M., Mo, H., & Yang, X. 2013, MNRAS, 430, 725

van den Busch, J. L., Hildebrandt, H., Wright, A. H., et al. 2020, A&A, 642, A200

van den Busch, J. L., Wright, A. H., Hildebrandt, H., et al. 2022, A&A, 664, A170

van der Wel, A., Noeske, K., Bezanson, R., et al. 2016, ApJS, 223, 29

van Uitert, E., Cacciato, M., Hoekstra, H., et al. 2016, MNRAS, 459, 3251

Van Waerbeke, L., Mellier, Y., Erben, T., et al. 2000, A&A, 358, 30

Viola, M., Cacciato, M., Brouwer, M., et al. 2015, MNRAS, 452, 3529

Viola, M., Kitching, T. D., & Joachimi, B. 2014, MNRAS, 439, 1909

Vogelsberger, M., Marinacci, F., Torrey, P., & Puchwein, E. 2020, Nature Reviews Physics, 2, 42

Wechsler, R. H. & Tinker, J. L. 2018, ARA&A, 56, 435

Wittman, D. M., Tyson, J. A., Kirkman, D., Dell'Antonio, I., & Bernstein, G. 2000, Nature, 405, 143

Wright, A. H., Driver, S. P., & Robotham, A. S. G. 2018, MNRAS, 480, 3491

Wright, A. H., Hildebrandt, H., Kuijken, K., et al. 2019, A&A, 632, A34

Wright, A. H., Hildebrandt, H., van den Busch, J. L., & Heymans, C. 2020a, A&A, 637, A100

Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2020b, A&A, 640, L14

Wright, A. H., Robotham, A. S. G., Bourne, N., et al. 2016, MNRAS, 460, 765

Wright, C. O. & Brainerd, T. G. 2000, ApJ, 534, 34

Yang, X., Mo, H. J., & van den Bosch, F. C. 2003, MNRAS, 339, 1057

Yang, X., Mo, H. J., & van den Bosch, F. C. 2008, ApJ, 676, 248

Zacharegkas, G., Chang, C., Prat, J., et al. 2022, MNRAS, 509, 3119

Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45

Nederlandstalige samenvatting

Het vaststellen van de fundamentele natuurkunde wetten die het ontstaan en de ontwikkeling van het universum beschrijven, is een van de belangrijkste prestaties van het kosmologisch onderzoek uit de twintigste eeuw. De theoretische en observationele vooruitgangen die in het veld zijn geboekt, zijn werkelijk exceptioneel. Het standaard kosmologische model, het ongekromde Λ 'Cold Dark Matter' (ΛCDM) model, dat slechts van zes vrije parameters afhangt, kan met indrukwekkende nauwkeurigheid diverse kosmologische sleutelobservaties voorspellen, zoals de uitdijing van het heelal, de metingen van de kosmische achtergrondstraling, en de verdeling van materie binnen de grootschalige structuren van het universum.

Ook in de 21e eeuw blijft het vakgebied van de kosmologie zich ontwikkelen. Vooruitgangen in observationele technieken gaan gepaard met verbeteringen in de modellen om de steeds nauwkeurigere metingen accuraat te interpreteren. De verbeteringen van verschillende kosmologische metingen, die mogelijk gemaakt worden door moderne technologieën, leiden tot een ongekende precisie in onze voorspellingen. Deze vooruitgang biedt een unieke kans om de mysteries rondom sommige essentiële elementen van het huidige kosmologische model, zoals de oorsprong van initiële dichtheidsfluctuaties en raadselachtige aard van donkere materie en energie, te ontrafelen. Echter, de technische ontwikkelingen gaan gepaard met de uitdagende taak om diverse systematische effecten, die te maken hebben met bijvoorbeeld observationele omstandigheden, onder controle te houden. Het implementeren van strategieën om deze systematische effecten meester te zijn, en zo het volledige potentieel van de metingen te benutten, is en blijft een centraal onderwerp binnen de kosmologie van de 21e eeuw.

Een veelbelovende maar uitdagende methode om kosmologische modellen mee te testen is de zwakke gravitatielenswerking. Volgens de algemene relativiteitstheorie wordt het licht van verre sterrenstelsels afgebogen door de zwaartekrachtvelden die veroorzaakt worden door de materie in het universum. Het meten van kleine maar coherente vervormingen van de afbeeldingen van sterrenstelsels geeft ons hierdoor een directe kijk op de verdeling van materie in de grootschalige structuren van het universum, of in individuele massieve objecten, zoals groepen van sterrenstelsels of clusters. Echter, in de praktijk wordt het robuust meten van deze door lenswerking geïnduceerde vervormingen bemoeilijkt door soortgelijke vervormingen die veroorzaakt kunnen worden door meetomstandigheden en instrumentatie. De metingen worden verder bemoeilijkt doordat de analyse sterk afhangt van de afstanden tot de sterrenstelsels, iets wat in de praktijk lastig nauwkeurig te meten valt voor grote groepen sterrenstelsels.

Zelfs met accurate metingen van de zwakke lenswerking, is het extraheren van de kosmologische informatie uit deze signalen niet triviaal. De interferentie wordt bovendien bemoeilijkt door de intrinsieke oplijning van sterrenstelsels en door de precieze invloed van supernova's en superzware zwarte gaten op de verdeling van materie in het universum.

Deze thesis richt zich op de bijdrage van verschillende systematische effecten binnen zwakke zwaartekrachtlenswerking analyses. In het bijzonder kijken we naar de invloed van deze effecten op het afleiden van kosmologische parameters en het onderzoeken van de eigenschappen van donkere materie halo's. Ons primaire doel is om de nauwkeurigheid en kallibratie van de zwakke gravitatielens analyses te verbeteren door de inferentiestappen kritisch te onderzoeken met behulp van consistentieen gevoeligheidstests. Daarnaast streven we ernaar om ons begrip van de relatie tussen donkere materie halo's en hun sterrenstelsels te verbeteren. Deze relatie is niet alleen cruciaal voor accurate kosmologische inferenties, maar ook voor het vergroten van ons begrip wat betreft de vorming en evolutie van sterrenstelsels.

In **Hoofdstuk 2** voeren we een consistentietest uit voor de zwakke zwaartekracht-lenswerking analyse met Kilo-Degree Survey (KiDS) metingen. We verdelen de achterliggende sterrenstelsels op basis van hun kleuren en we beoordelen de robuustheid van de huidige KiDS resultaten ten opzichte van het verwerken van subgroepen van sterrenstelsels met verschillende eigenschappen. Onze resultaten bevestigen dat de huidige KiDS analyse op een correcte manier rekening houdt met systematische onzekerheden. Echter, we identificeren een correlatie tussen IA-parameters en parameters die horen bij de roodverschuivingskalibratie. Deze correlatie onderschrijft de noodzaak voor verbeterde IA-modellen voor toekomstige analyses. We vinden dat de consistentietest die we gebruiken ongevoelig is voor het aangenomen kosmologische model, waardoor deze bijzonder geschikt is om de robuustheid van zwaartekrachtlenseffect analyses te verzekeren voordat de metingen gebruikt worden om enige kosmologische conclusies te trekken.

Hoofdstuk 3 introduceert een reeks multi-band beeldsimulaties voor de kalibratie van de KiDS zwakke gravitatielensanalyse. Deze nieuwe simulaties maken een gezamenlijke kalibratie van het gravitatielenseffect en de roodverschuiving metingen mogelijk, wat de correctheid en nauwkeurigheid ten opzichte van eerdere analyses vergroot. Door de simulaties te combineren met hoogwaardige afbeeldingen van verre sterrenstelsels, genereren we een grote set aan gesimuleerde sterrenstelsels met eigenschappen die overeenkomen met de metingen. We verbeteren de accuraatheid van de afbeeldingen door zorgvuldig verschillende observationele en instrumentele effecten mee te nemen, zoals achtergrondruis, de punt-spreidingsfunctie (PSF), en de sterdichtheid.

Verder bestuderen we meng-effecten op verschillende roodverschuivingen door variabele zwakke zwaartekrachtlens velden in de simulaties te introduceren. Onze studie laat een correlatie tussen het gravitatielenseffect en roodverschuivingsfouten zien als gevolg van meng-effecten. Deze correlatie geeft het belang van de gezamenlijke kalibratie aan. We identificeren ook een kleine maar niet verwaarloosbare invloed van PSF-modelleringsfouten op de gravitatielenseffect metingen. Met behulp van gevoeligheidstests bevestigen we de robuustheid van het vormmetingsalgoritme binnen

de vereisten van de huidge KiDS zwakke gravitatielensanalyse. Toekomstige zwakke zwaartekrachtlensonderzoeken, met strengere vereisten, raden we aan om verdere studies naar meng-effecten, verbeteringen van PSF-modellering, en vormmetingstechnieken uit te voeren om de gevoeligheid voor verschillen in sterrenstelseleigenschappen te verminderen.

In **Hoofdstuk 4** voeren we een zwakke gravitatielensanalyse uit, waarbij we de verbeteringen meenemen die we geïntroduceerd hebben in Hoofdstuk 3. Bovendien houden we rekening met recente vooruitgangen in kosmologische inferenties en onderzoeken we de invloed van IA-modellering op de uiteindelijke kosmologische resultaten. We stellen voor om de statistische en systematische onzekerheden van de kalibratie, gezien hun verschillende oorsprong, afzonderlijk te behandelen. Onze uiteindelijke kosmologische resultaten komen overeen met eerdere analyses van KiDS metingen en die van andere zwakke gravitatielensonderzoeken, waarbij een discrepantie van ongeveer 2.3σ op S_8 wordt gemeten ten opzichte van de Planck kosmische microgolfachtergrondfluctuaties resultaten.

Ten slotte onderzoeken we in **Hoofdstuk 5** de relatie tussen donkere materie halo's en hun sterrenstelsels met behulp van de het zwakke lenswerkingeffect. We meten signalen rond sterrenstelsgroepen die gemeten zijn binnen het Galaxy And Mass Assembly (GAMA) project. We analyseren de zwaartekrachtlens metingen met behulp van van het halo model formalisme. Hiermee bepalen we de totale halo massa van de sterrenstelsgroepen. Door optische GAMA metingen te gebruiken, leggen we de schalingsrelatie tussen de halo massa en de helderheid van de sterrenstelselgroepen vast. We beoordelen de robuustheid van onze metingen door de behandeling van verschillende parameters te variëren. De gevoeligheidstests helpen ons bij het identificeren van kritieke modelcomponenten en zullen als handvatten gebruikt kunnen worden voor toekomstige verbeteringen van het model. Gezien de verhoogde sensitiviteit van toekomstige metingen, zijn verbeteringen van het model noodzakelijk.

Publications during PhD study

Refereed publications

2. *KiDS-Legacy calibration: Unifying shear and redshift calibration with the SKiLLS multi-band image simulations*
   Li, S.-S.; Kuijken, K.; Hoekstra, H.; Miller, L.; Heymans, C.; Hildebrandt, H.; van den Busch, J.L.; Wright, A.H.; Yoon, M.; Bilicki, M.; Bravo, M.; Lagos, C.P.; 2023, A&A, 670, A100 (**Chapter 3**)

1. *KiDS+VIKING-450: An internal-consistency test for cosmic shear tomography with a colour-based split of source galaxies*
   Li, S.-S.; Kuijken, K.; Hoekstra, H.; Hildebrandt, H.; Joachimi, B.; Kannawadi, A.; 2021, A&A, 646, A175 (**Chapter 2**)

Submitted publications

3. *KiDS-1000: Cosmology with improved cosmic shear measurements*
   Li, S.-S.; Hoekstra, H.; Kuijken, K.; Asgari, M.; Bilicki, M.; Giblin, B.; Heymans, C.; Hildebrandt, H.; Joachimi, B.; Miller, L.; van den Busch, J.L.; Wright, A.H.; Kannawadi, A.; Reischke, R.; Shan, H.; submitted to A&A, arXiv:2306.11124 (**Chapter 4**)

2. *DES Y3 + KiDS-1000: Consistent cosmology combining cosmic shear surveys*
   Dark Energy Survey and Kilo-Degree Survey Collaborations; submitted to the Open Journal of Astrophysics, arXiv:2305.17173

1. *Strong lensing selection effects*
   Sonnenfeld, A.; **Li, S.-S.**; Despali, G.; Shajib, A.J.; Taylor, E.N.; submitted to A&A, arXiv:2301.13230

Curriculum vitae

Born on a snowy day in 1994, I took my first steps in a small town named Dongyang in Zhejiang, China. My parents often recall how I was a quiet nerd, keen on calculations but with a poor memory. The disparity between my language and math scores on my school report cards echoed their observations. My ever-growing curiosity about nature ultimately drove me to pursue a career in scientific research.

In pursuit of higher education, I left my hometown after 18 years and joined Nanjing University, renowned for having the oldest Astronomy department in China. The rigorous courses in mathematics, physics, and astronomy clarified many of the theories that piqued my curiosity during high school. It was a reaffirmation that I was on the right path. Additionally, I participated in some early research training programs offered by various institutes. A particularly formative experience was the two-month 'UWA Research Training Program' at the University of Western Australia in Perth. This was my first immersion into a different culture, which enriched both my research and language skills and broadened my worldview. It was this journey that solidified my aspiration to study in a diverse international environment.

My deep dive into research kicked off in Beijing during the final year of my Bachelor's program. Guided by Prof. Shude Mao, I started to explore the world of gravitational lensing, especially in relation to gravitational waves. Our first paper on the topic has since collected over a hundred citations. Yet, beyond the recognition, the sheer joy and fulfilment I derived from the process truly fueled my passion. I completed my Bachelor's degree in Astronomy and proceeded with a Master's in Astrophysics, deepening my expertise in gravitational lensing, spanning strong to microlensing, and expanding from gravitational waves to planetary exploration.

Fulfilling my aspiration to study abroad, I relocated to the picturesque town of Leiden in the Netherlands for my PhD studies. Under the guidance of Profs. Koen Kuijken and Henk Hoekstra, I delved into the fascinating realm of gravitational lensing within cosmology. My involvement in international collaborations, notably the Kilo-Degree Survey (KiDS) and the *Euclid* space mission, greatly enriched my academic journey. With steadfast support from my supervisors and the KiDS team, we developed the third-generation image simulations for KiDS, dubbed 'SKiLLS'. This naming convention follows its predecessors - 'SCHOol' and COllege', symbolising both the development of KiDS and our shared personal growth. Beyond my primary research, I contributed to Master's courses on 'Gravitational lensing' and 'Large scale structure and galaxy formation', and helped mentor two Master's students through their research projects.

The global pandemic did throw a wrench into my PhD journey, especially affecting collaborations and connections with colleagues. Virtual interactions became our norm for over two years. However, I was fortunate to have some trips both before and after the pandemic, forging lasting memories. These include visits to institutions in the UK and Germany and delivering oral presentations at international conferences.

Come November 2023, I will embark on a new adventure as a KIPAC Rubin postdoctoral fellow at Stanford. I eagerly await the challenges and discoveries that lie ahead, especially with the Legacy Survey of Space and Time by the Vera C. Rubin Observatory.

Acknowledgements

Words always fall short when it comes to acknowledgements, especially given the immense support I have received from countless remarkable individuals throughout my PhD journey. I am deeply grateful to have crossed paths with such talented and generous souls, each imparting invaluable lessons and insights. Foremost among them are my two supervisors. Many students would consider themselves fortunate to be mentored by just one of you, yet I had the rare privilege of learning from both. Our engaging discussions have been instrumental in shaping my academic and personal trajectory. Without your unwavering patience, guidance, and support—even in the face of my sometimes naive questions and broken English—I wouldn't be where I am today. I hope our paths keep crossing, permitting me to draw upon your wisdom and advice throughout my career.

The colleagues from the Leiden weak lensing and cosmological simulations groups have been another indispensable source of knowledge and guidance. I cherish our conversations during group meetings, lunches, and conferences. Engaging with each of you, discussing science, and learning about our diverse cultures have been truly enriching experiences.

A significant portion of my knowledge can be attributed to the exceptional KiDS team. The fusion of science and camaraderie we experienced during the KiDS busy weeks will forever resonate with me. Collaborations of this nature are priceless, and I am convinced that every international research group should embrace a similar approach! A special mention goes to the KiDS-Legacy Calibration team. Our collective effort birthed 'SKiLLS', which stands as one of my most valuable experiences.

I would like to thank the administrative and support staff at the Leiden Observatory. Every interaction with you was imbued with warmth and efficiency, significantly simplifying my life and work at the institute. The dedicated IT team, in particular, stands out; you are the finest technical support group I have ever encountered. I am thankful for all the tasks handled upon my request and, more so, for your proactive behind-the-scenes contributions.

I am also indebted to my fellow PhD mates, particularly the 'Class of 2019'. We journeyed through the same challenging phases, supporting one another. The memories we created during off-campus hangouts, game nights, and dinners, both pre- and post-pandemic, will always hold a special place in my heart. I hope for our paths to intertwine once more in the future.

Beyond academia, I owe a debt of gratitude to my girlfriend and her remarkable circle of friends who became a major part of my social life. Your company at countless

memorable events has been one of the highlights of my PhD journey. You have added so much joy and warmth to my time here.

In conclusion, my heartfelt gratitude goes to my family and friends back home. They made countless sacrifices, especially given my prolonged absences. Their enduring faith and encouragement have been my pillars of strength, dispelling my moments of doubt and lighting the way to clarity. I will forever treasure the warmth of home.