



Universiteit  
Leiden  
The Netherlands

## **KiDS-Legacy: cosmological constraints from cosmic shear with the complete Kilo-Degree Survey**

Wright, A.H.; Stözlner, B.; Asgari, M.; Bilicki, M.; Giblin, B.; Heymans, C.; ... ; Zhang, Y.-H.

### **Citation**

Wright, A. H., Stözlner, B., Asgari, M., Bilicki, M., Giblin, B., Heymans, C., ... Zhang, Y. -H. (2025). KiDS-Legacy: cosmological constraints from cosmic shear with the complete Kilo-Degree Survey. *Astronomy And Astrophysics*, 703. doi:10.1051/0004-6361/202554908

Version: Accepted Manuscript

License: [Creative Commons CC BY 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/4288580>

**Note:** To cite this publication please use the final published version (if applicable).

# KiDS-Legacy: Cosmological constraints from cosmic shear with the complete Kilo-Degree Survey<sup>★</sup>

Angus H. Wright<sup>1\*\*</sup>, Benjamin Stözlner<sup>1</sup>, Marika Asgari<sup>2</sup>, Maciej Bilicki<sup>3</sup>, Benjamin Giblin<sup>4</sup>, Catherine Heymans<sup>1,4</sup>, Hendrik Hildebrandt<sup>1</sup>, Henk Hoekstra<sup>5</sup>, Benjamin Joachimi<sup>6</sup>, Konrad Kuijken<sup>5</sup>, Shun-Sheng Li<sup>1,5</sup>, Robert Reischke<sup>1,7</sup>, Maximilian von Wietersheim-Kramsta<sup>8,9</sup>, Mijin Yoon<sup>5</sup>, Pierre Burger<sup>7,10,11</sup>, Nora Elisa Chisari<sup>5,12</sup>, Jelte de Jong<sup>13</sup>, Andrej Dvornik<sup>1</sup>, Christos Georgiou<sup>14</sup>, Joachim Harnois-Déraps<sup>15</sup>, Priyanka Jalan<sup>3</sup>, Anjitha John William<sup>3</sup>, Shahab Joudaki<sup>16,17</sup>, Giorgio Francesco Lesci<sup>18,19</sup>, Laila Linke<sup>20</sup>, Arthur Loureiro<sup>21,22</sup>, Constance Mahony<sup>1,23,24</sup>, Matteo Maturi<sup>25</sup>, Lance Miller<sup>23</sup>, Lauro Moscardini<sup>18,19,26</sup>, Nicola R. Napolitano<sup>27</sup>, Lucas Porth<sup>7</sup>, Mario Radovich<sup>28</sup>, Peter Schneider<sup>7</sup>, Tilman Tröster<sup>29</sup>, Edwin Valentijn<sup>5</sup>, Anna Wittje<sup>1</sup>, Ziang Yan<sup>1</sup>, and Yun-Hao Zhang<sup>4,5</sup>

(Affiliations can be found after the references)

Received 31 March 2025 / Accepted 31 July 2025

## ABSTRACT

We present cosmic shear constraints from the completed Kilo-Degree Survey (KiDS), where the cosmological parameter  $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3} = 0.815^{+0.016}_{-0.021}$  is found to be in agreement ( $0.73\sigma$ ) with results from the *Planck* Legacy cosmic microwave background experiment. The final KiDS footprint spans 1347 square degrees of deep nine-band imaging across the optical and near-infrared (NIR), along with an extra 23-square degrees of KiDS-like calibration observations of deep spectroscopic surveys. Improvements in our redshift distribution estimation methodology, combined with our enhanced calibration data and multi-band image simulations, allowed us to extend our lensed sample out to a photometric redshift of  $z_B \leq 2.0$ . Compared to previous KiDS analyses, the increased survey area and redshift depth results in a  $\sim 32\%$  improvement in constraining power in terms of  $\Sigma_8 \equiv \sigma_8 (\Omega_m/0.3)^\alpha = 0.821^{+0.014}_{-0.016}$ , where  $\alpha = 0.58$  has been optimised to match the revised degeneracy direction of  $\sigma_8$  and  $\Omega_m$  for our current survey at higher redshift. We adopted a new physically motivated intrinsic alignment (IA) model that jointly depends on the galaxy sample's halo mass and spectral type distributions, and which is informed by previous direct alignment measurements. We also marginalised over our uncertainty on the impact of baryon feedback on the non-linear matter power spectrum. Compared to previous KiDS analyses, we conclude that the increase seen in  $S_8$  primarily results from our improved redshift distribution estimation and calibration, as well as a new survey area and improved image reduction. Our companion paper presents a full suite of internal and external consistency tests (including joint constraints with other datasets), finding the KiDS-Legacy dataset to be the most internally robust sample produced by KiDS to date.

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

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>	<b>3</b>	<b>Data and analysis</b>	<b>8</b>
<b>2</b>	<b>Modelling and theoretical framework</b>	<b>3</b>	<b>3.1</b>	KiDS and VIKING	9
2.1	Shear power spectrum	3	3.2	KiDZ	10
2.2	Matter power spectrum	3	3.3	Blinding	10
2.3	Baryonic feedback	4	3.4	Redshift distributions and calibration	10
2.4	Intrinsic alignments	4	3.5	Shape measurement and calibration	10
2.4.1	The non-linear linear alignment (NLA) model	4	<b>4</b>	<b>Null tests</b>	<b>12</b>
2.4.2	Additional redshift dependence (NLA- $z$ )	5	4.1	Paulin-Henriksson et al. (2008) model	12
2.4.3	Restricted TATT model (NLA- $k$ )	5	4.2	Bacon et al. (2003) statistics	13
2.4.4	Galaxy-type and mass-dependent model (NLA- $M$ )	6	4.3	$B$ modes	13
2.5	Cosmic shear summary statistics	6	<b>5</b>	<b>Results</b>	<b>14</b>
2.5.1	2PCFs ( $\xi_{\pm}$ )	6	5.1	Fiducial constraints	14
2.5.2	COSEBIs ( $E_n/B_n$ )	7	5.2	Analysis variations	15
2.5.3	Band powers ( $C_{E,B}$ )	7	5.2.1	Modelling $P(k)$ with CAMB	16
			5.2.2	Gaussian covariance	16
			5.2.3	Iterative covariance	17
			5.2.4	Intrinsic alignment modelling	17
			5.2.5	Impact of observational and systematic nuisance parameters	19
			5.2.6	Alternative redshift distribution calibration	19
			5.2.7	Scale cuts	20

<sup>★</sup> We dedicate this article to the memory of Nick Kaiser. He turned weak lensing into a powerful and practical cosmology probe, inspiring us to conceive the Kilo-Degree Survey. The KiDS project has relied greatly on his brilliant and wide-ranging foundational work.

\*\* awright@astro.rub.de

5.2.8	Removing tomographic redshift bins . . .	20
<b>6</b>	<b>Discussion</b>	<b>21</b>
<b>7</b>	<b>Summary</b>	<b>22</b>
<b>A</b>	<b>Emulating matter power spectra with COSMOPower</b>	<b>26</b>
<b>B</b>	<b>Galaxy sample properties for intrinsic alignment modelling</b>	<b>26</b>
<b>C</b>	<b>Band-powers derivation</b>	<b>27</b>
<b>D</b>	<b>KiDS-Legacy variable depth mocks and covariance validation</b>	<b>27</b>
<b>E</b>	<b><i>B</i>-mode analysis</b>	<b>28</b>
<b>F</b>	<b>Cosmological parameters from <math>\xi_{\pm}</math></b>	<b>30</b>
<b>G</b>	<b>Additional posterior constraints</b>	<b>32</b>
G.1	Whisker diagram in $S_8$ . . . . .	32
G.2	Redshift distributions . . . . .	32
G.3	Cosmological parameters . . . . .	32
<b>H</b>	<b>Tabulated constraints</b>	<b>35</b>
<b>I</b>	<b>Post-unblinding analyses and changes</b>	<b>36</b>
I.1	Comparison between Legacy and KiDS-1000 . . .	36
I.1.1	Analysing KiDS-1000 with Legacy calibration methods . . . . .	37
I.1.2	Updating KiDS-1000 to the Legacy calibration sample . . . . .	37
I.1.3	Legacy analysis in the KiDS-1000 volume . . .	37
I.1.4	Sixth tomographic bin and NLA- $M$ . . . . .	37
I.1.5	Additional Legacy area . . . . .	37
I.2	Expanded scale cuts . . . . .	38
I.3	Iterative covariances . . . . .	38

## 1. Introduction

Our current understanding of the large-scale structure (LSS) of the Universe, obtained over the last three decades thanks to increasingly sophisticated astronomical observations, can be regarded as one of the great successes of modern scientific inquiry. Detailed observations of the LSS probe fundamental physics, since the structures we observe were formed by an interplay of physical effects. These primarily include the self-gravity of matter, cosmic expansion, astrophysical feedback processes, and the repulsive effect of the cosmological constant or dark energy.

When it comes to understanding the growth and evolution of LSS, the cosmological stage is set by observations of the cosmic microwave background (CMB) that reveal the structure of dark and baryonic matter at early times. Together with other probes, including big-bang nucleosynthesis (BBN) and baryon-acoustic oscillations (BAOs), the CMB has been used to establish and tightly constrain the so-called standard model of cosmology,  $\Lambda$ CDM. It is a strikingly simple model based on only a handful of numerical parameters (Planck Collaboration et al. 2020; Qu et al. 2024; Ge et al. 2025) and characterised, in particular, by two components that are at the origin of its name: a cosmological constant ( $\Lambda$ , which dominates the energy landscape at late times) and a cold dark matter (CDM) component that is dominant over baryonic matter by a factor of roughly five.

Observations of the CMB constrain the parameters of this model at high redshift ( $z \sim 1100$ ), when the Universe was approximately 370 thousand years old. However, the model can be subsequently used to predict the properties of the LSS at late times ( $z \lesssim 2$ ; an extrapolation over more than 10 billion years), which can then be tested with low-redshift probes of cosmic structures.

Of the low-redshift probes, the relative prevalence of CDM poses a problem for those that rely on luminous tracers, such as galaxy clustering measurements. Modelling and constraining the relationship between luminous and dark matter (DM), also known as galaxy bias, constitutes one of the major systematic challenges of such measurements. In contrast, observations of the weak gravitational lensing effect of the LSS are intrinsically sensitive to all matter (dark or otherwise), eliminating the systematic uncertainty of galaxy bias. This unique feature has turned such ‘cosmic shear’ experiments into a major cosmological probe that is just entering its golden age. Indeed, it is the combination of cosmic shear with galaxy clustering – and their cross-correlation called galaxy-galaxy lensing (GGL) – that will potentially yield the most precise test of the standard model.

Over the past decade, the comparison between observations of low- $z$  LSS and their CMB-based predictions has revealed a tension in the  $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3}$  parameter. Therefore, it has been the focus of extensive study. All contemporary cosmic shear measurements to date, except one, have yielded values of  $S_8$  that are lower than CMB predictions, indicating a universe with less structure and/or matter. The significance of these measurements varies, but there is a clear tendency towards lower values in the cosmic shear results of the past decade. Only the Deep Lens Survey (DLS; Jee et al. 2016) has reported a value that is higher than that of *Planck*, and even then only slightly. In contrast, lower values have been reported by the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS; Heymans et al. 2013), as well as from previous results from the ongoing ‘stage-III’ surveys by the Kilo-Degree Survey (KiDS; Asgari et al. 2021; Li et al. 2023a), the Dark Energy Survey (DES; Amon et al. 2022; Secco et al. 2022), the Hyper-Suprime Camera survey (HSC; Li et al. 2023c; Dalal et al. 2023), and a joint analysis by Dark Energy Survey and Kilo-Degree Survey Collaboration et al. (2023). Other low- $z$  LSS probes typically also share this trend but again with a varying level of significance (see Abdalla et al. 2022, for an overview).

Cosmic shear probes relatively low redshifts, and relatively small scales compared to other probes of LSS (see e.g. Figure 1 of Broxterman & Kuijken 2024, for a graphical summary). This makes cosmic shear somewhat more sensitive than other low- $z$  probes to non-linear structure formation and astrophysical effects: the latter mainly involve feedback processes driven by supernovae and active galactic nuclei, collectively referred to below as baryonic feedback. It has therefore been suggested that inaccuracies in the modelling of non-linear scales might be the reason for the tension in  $S_8$ , rather than a more fundamental cosmological cause (Yoon et al. 2019; Yoon & Jee 2021; Amon & Efstathiou 2022; Preston et al. 2023).

A definitive answer to the reality of the  $S_8$  tension is within reach. Currently, three stage-III surveys have finished their observations and are preparing their final analyses; one of which we present here. The next generation of surveys, called stage-IV, are just beginning, represented by the ESA *Euclid* mission (Euclid Collaboration et al. 2025), the Legacy Survey of Space and Time (LSST) conducted with the Vera C. Rubin Observatory (Ivezić et al. 2019), the NASA Roman Space Telescope (Spergel et al. 2015), and the Chinese Space Station Optical Sur-

vey (Gong et al. 2019). We note that *Euclid* is expected to yield first cosmologically relevant results in 2026. These stage-IV surveys will soon provide datasets capable of measuring the growth of structure at low  $z$  with unprecedented precision.

New facilities and new instruments invariably bring along new challenges owing to, for instance, the novelty of their hardware. Comparatively, the final (usually called ‘legacy’) releases of a survey typically represent the culmination of years or decades of analysis, understanding, and refinement regarding their survey and the associated data. As such, the final releases of stage-III surveys will provide a crucial insight and reference for subsequent lensing studies of LSS with stage-IV surveys.

Here, we present the cosmic shear analysis of the completed Kilo-Degree Survey (KiDS) based on the fifth and final data release (KiDS-DR5; Wright et al. 2024), dubbed KiDS-Legacy. With extensive wavelength coverage enabled by its unique pairing with its near-infrared (NIR) sister survey VIKING (VISTA Kilo-Degree Infrared Galaxy Survey) and with its superb image quality in the  $r$ -band, KiDS is the stage-III survey that is arguably most suited to simultaneously leverage high-redshift galaxies and accurately calibrate their redshift distributions. Over the past decade, KiDS has been pushing to increasingly high redshifts, while continuously improving the redshift calibration, shear measurements, shear calibration through image simulations, and methodology to accurately extract cosmological constraints from the imaging data.

This paper represents the culmination of these efforts, providing the most robust and precise analysis of KiDS to date. Our dataset is the largest analysed by KiDS thus far (Sect. 3.1), utilising the most extensive spectroscopic redshift compilation prepared for any cosmic shear analysis (Sect. 3.2), and has been calibrated using the most sophisticated simulations produced for this purpose (Sect. 3.4). Improvements in methodology (Sect. 2) have allowed us to access the highest redshifts (with the most tomographic bins) studied for cosmic shear to date (Sect. 3.4). An efficient workflow for this study was enabled by the new CosmoPIPE analysis framework, which we have also made publicly available. We present null tests of our dataset in Sect. 4. The cosmological results are presented in Sect. 5, discussed in Sect. 6, and summarised in Sect. 7. Additionally, we include supplementary information regarding our power spectrum emulation (Appendix A), on the definition of parameters for our fiducial intrinsic alignment model (Appendix B), construction of band power spectra (Appendix C), covariance validation (Appendix D),  $B_n$ -mode exploration (Appendix E), correlation function cosmological analyses (Appendix F), detailed posterior analyses (Appendix G), tables of cosmological constraints (Appendix H), and details of changes made to the analysis and manuscript post-unblinding (Appendix I). Finally, this paper has been released alongside two companion papers: the work of Stölzner et al. (2025) presents an extensive analysis of internal and external consistency tests of the KiDS-Legacy dataset (including joint constraints with other datasets) and the work of Wright et al. (2025) describes the various redshift distribution and bias estimation developments made for KiDS-Legacy.

## 2. Modelling and theoretical framework

The main cosmic shear observable is the ellipticity correlation between pairs of galaxies. The signal depends on the two-point statistics of the shear and galaxy intrinsic alignment fields, which both trace the matter distribution and therefore depend on cosmological information. For our predictions and sampling, we used CosmoSIS, a modular cosmology pipeline described in

Zuntz et al. (2015). We included a number of new and/or modified modules in our analysis, including a CosmoPOWER emulator for non-linear power spectra (Appendix A) and new intrinsic alignment models (Sect. 2.4). In this section, we start with the derivation of cosmic shear power spectra from matter power spectra and then detail our choices of intrinsic alignment models. Next, we introduce our observable two-point statistics, briefly discussing their measurement and prediction methodology. We end this section with a brief review of our analytical covariance modelling.

### 2.1. Shear power spectrum

Our two-point statistic predictions are calculated from shear (cross-)power spectra of galaxy samples  $i$  and  $j$ ,  $C_{GG}^{(ij)}(\ell)$ , where G stands for ‘gravitational’, which distinguishes these power spectra from intrinsic alignment contributions to the two-point signal. Shear power spectra are given by Limber-approximated projections (Kaiser 1992; LoVerde & Afshordi 2008; Kilbinger et al. 2017) taking the form

$$C_{GG}^{(ij)}(\ell) = \int_0^{\chi_{\text{hor}}} d\chi \frac{W_G^{(i)}(\chi) W_G^{(j)}(\chi)}{f_K^2(\chi)} P_{\text{m,nl}}\left(\frac{\ell + 1/2}{f_K(\chi)}, z(\chi)\right), \quad (1)$$

where  $P_{\text{m,nl}}$  is the non-linear matter power spectrum and  $f_K(\chi)$  is the comoving angular diameter distance. Although the integral in Eq. (1) technically runs over the entire line of sight to the horizon,  $\chi_{\text{hor}}$ , in practice, the weak lensing kernel,  $W_G^{(i)}(\chi)$ , limits the integral to the comoving distance of the furthest galaxy in our sample, expressed as

$$W_G^{(i)}(\chi) = \frac{3H_0^2 \Omega_m}{2c^2} \frac{f_K(\chi)}{a(\chi)} \int_\chi^{\chi_{\text{hor}}} d\chi' n_S^{(i)}(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')}, \quad (2)$$

where  $n_S^{(i)}(\chi)$  is the distribution of comoving distances of galaxies in source sample,  $i$ ; in practice, we express this in terms of redshift, namely,  $N(z) dz = n_S(\chi) d\chi$ . Here,  $\Omega_m$  is the total matter density parameter today,  $H_0$  is the Hubble constant,  $c$  is the speed of light, and  $a(\chi)$  is the scale factor normalised to unity today.

### 2.2. Matter power spectrum

In our cosmic shear analysis, we remained mostly sensitive to Fourier scales  $0.1 h/\text{Mpc} \lesssim k \lesssim 1 h/\text{Mpc}$ , with a weighting of scales that depends on the two-point statistic (see Sect. 2.5). These scales are well within the (quasi-) non-linear regime and therefore require an accurate non-linear power spectrum model for their predictions. Currently, there are multiple approaches that can produce non-linear matter power spectra, with varying levels of accuracy, for this range of scales. Here, we followed the most popular approach for cosmic shear analyses, using a method inspired by the halo model and calibrated against simulations. Specifically, we used the Boltzmann solver CAMB (Lewis et al. 2000; Howlett et al. 2012) to estimate the linear matter power spectrum and the augmented halo model ‘HMCODE2020’ (Mead et al. 2021) to derive its non-linear evolution.

HMCODE2020 has a similar level of accuracy as popular emulators trained directly on N-body simulations (2.5%; see e.g. Knabenhans et al. 2023; Aricò et al. 2021) and is twice as accurate as HALOFIT from Takahashi et al. (2012), which has been used in a number of previous cosmic shear analyses (see e.g. Amon et al. 2022; Hikage et al. 2019).

Dark Energy Survey and Kilo-Degree Survey Collaboration et al. (2023) presented a comparison of HMCODE2020 and HALOFIT with EUCLIDEMULATORV2, showing that HMCODE2020 produces unbiased results when analysing mock data created with EUCLIDEMULATORV2; meanwhile HALOFIT tends to overestimate power for lower values of  $S_8$ , resulting in an underestimated  $S_8$  value. Similarly, an earlier version of HMCODE, HMCODE2015 (Mead 2015), which was used in previous KiDS analyses, is  $\sim 30\%$  less accurate than HMCODE2020. This can subsequently produce small differences between cosmological constraints estimated using the two codes: Tröster et al. (2021) demonstrated that the most important difference between these versions of HMCODE is in their treatment of baryon feedback, which results in a slightly higher  $S_8$  using HMCODE2020. Lastly, we note that the parameter values at which HMCODE was fitted to simulations are somewhat more constrained than our chosen priors for  $h^2\Omega_m$  (see Mead et al. 2021, and Table 3). However, given that the fitting functions of HMCODE are based on physically motivated parameters, we did not anticipate any significant issues just beyond this range. Furthermore, we note that our analysis remained unaffected by this choice.

### 2.3. Baryonic feedback

The primary features of the matter power spectrum can be determined by assuming all matter is in the form of (pressureless) CDM, with modifications due to processes involving baryonic matter. At large scales ( $0.05 h/\text{Mpc} \lesssim k \lesssim 0.5 h/\text{Mpc}$ ), BAOs produce oscillations in the power spectrum. Over a wide range of intermediate scales ( $0.1 h/\text{Mpc} \lesssim k \lesssim 10 h/\text{Mpc}$ ), power is suppressed through feedback processes from active galactic nuclei (AGN), reaching a maximum suppression at around  $k \sim 5 h/\text{Mpc}$ . At even smaller scales ( $k \gtrsim 10 h/\text{Mpc}$ ), star formation processes lead to a sharp increase in power (see e.g. Semboloni et al. 2011).

To model baryonic feedback processes, HMCODE2020 performs a modification to the concentration-mass relation of the halo profiles as a function of their gas fraction; this, in turn, depends on the strength of the AGN feedback (parametrised by  $\log(T_{\text{AGN}}) \in [7.1, 8.3]$ ). This feedback model is then calibrated against a suite of hydrodynamical simulations from BAHAMAS (McCarthy et al. 2017), which had been generated using a wide range of effective feedback amplitudes and cosmological models. Recently, Schaye et al. (2023, FLAMINGO) presented a new suite of hydrodynamical simulations, including emulators for generating power spectra (Schaller et al. 2025), as a mechanism for modelling baryonic effects. Moving beyond hydrodynamical simulations, Salcido et al. (2023, SP( $k$ )) presented alternative phenomenological models for the same purpose, while Schneider & Teyssier (2015); Aricò et al. (2020) presented perturbative models that act directly on the outputs of dark-matter only simulations and can be empirically calibrated (Grandis et al. 2024; Schneider et al. 2022). We did not utilise these models for our cosmological analyses here, as these models were published over the course of the analysis of KiDSL Legacy or were not (at time of writing) implemented within our COSMOSIS sampling framework. However, we note that a reanalysis of our data with these models is a natural and straightforward project that will be undertaken in a future publication.

### 2.4. Intrinsic alignments

The intrinsic alignment (IA) of galaxies is correlated with the observed ellipticities of galaxy images, making them indistinguishable from the cosmic shear effect we wish to probe for cosmological parameter estimation. In the limit of small shears and IA, the two-point statistics of galaxy ellipticities are given by

$$C_{\varepsilon\varepsilon}^{(ij)}(\ell) = C_{\text{GG}}^{(ij)}(\ell) + C_{\text{GI}}^{(ij)}(\ell) + C_{\text{IG}}^{(ij)}(\ell) + C_{\text{II}}^{(ij)}(\ell), \quad (3)$$

where  $C_{\text{GG}}^{(ij)}$  is the cosmic shear signal given by Eq. (1). Analogous Limber equations can be derived for IA contributions, which are typically grouped into two categories. First, intrinsic-intrinsic (II) correlations represent the correlated alignment of galaxies that are tidally aligned by the same ambient LSS:

$$C_{\text{II}}^{(ij)}(\ell) = \int_0^{\chi_{\text{hor}}} d\chi \frac{n_S^{(i)}(\chi) n_S^{(j)}(\chi)}{f_K^2(\chi)} P_{\text{II},X}^{(ij)}\left(\frac{\ell+1/2}{f_K(\chi)}, z(\chi)\right). \quad (4)$$

Second, shear-intrinsic (GI) correlations are driven by the mutual gravitational effects of LSS on (i) galaxies within the structures, and (ii) the lensed images of background sources that these structures produce (Hirata & Seljak 2004):

$$C_{\text{GI}}^{(ij)}(\ell) = \int_0^{\chi_{\text{hor}}} d\chi \frac{W_G^{(i)}(\chi) n_S^{(j)}(\chi)}{f_K^2(\chi)} P_{\text{GI},X}^{(ij)}\left(\frac{\ell+1/2}{f_K(\chi)}, z(\chi)\right), \quad (5)$$

with  $C_{\text{IG}}^{(ij)} = C_{\text{GI}}^{(ji)}$ . In Eqs. (4) and (5) the power spectra  $P_{\text{II},X}$  and  $P_{\text{GI},X}$  are given by the chosen IA model. We note that the IA contributions feature a more compact kernel along the line of sight than the lensing kernel in Eq. (1), and hence the Limber approximation is poorer for these signals. However, this effect is not significant for our analysis, as the overall IA contributions to the total signal is small, making up only around 10 percent of the signal in our analysis. Additionally, we are only probing scales  $\ell \gtrsim 50$ , where the Limber approximation is sufficient even for more compact kernels (Leonard et al. 2023).

Since IA are caused by the local matter environment, they have a different redshift scaling than lensing, which however is insufficient to isolate the cosmological signal without substantial information loss (Joachim et al. 2008). Therefore, cosmic shear analyses opt for physically motivated IA models that are inferred jointly with the cosmology (see Lamman et al. 2024, for a review of the field). Guided by this, we implement four IA models in KiDSL-Legacy: a widely used baseline with a single overall amplitude parameter (NLA, Sect. 2.4.1), two minimal extensions of NLA with either an additional redshift (NLA- $z$ , Sect. 2.4.2) or scale dependence (via our restricted TATT model, NLA- $k$ ; Sect. 2.4.3), and a version of NLA that explicitly incorporates the well-established trends of IA with galaxy type and host halo mass (NLA- $M$ , Sect. 2.4.4). The latter model is the fiducial model chosen for our cosmological analyses.

#### 2.4.1. The non-linear linear alignment (NLA) model

The so-called NLA model (Bridle & King 2007) assumes that the intrinsic ellipticity of a galaxy depends linearly on its surrounding tidal quadrupole (Catelan et al. 2001). Moreover, it is commonly assumed that it is the tidal field at some early time during the formation of the galaxy that sets the amplitude of alignment (Hirata & Seljak 2004). This leads to the following

expressions for the IA power spectra,

$$P_{\delta\text{I,NLA}}(k, z) = -A_{\text{IA}} \frac{C_1 \rho_{\text{cr}} \Omega_{\text{m}}}{D(z)} P_{\text{m,nl}}(k, z), \quad (6)$$

$$P_{\text{II,NLA}}(k, z) = A_{\text{IA}}^2 \left( \frac{C_1 \rho_{\text{cr}} \Omega_{\text{m}}}{D(z)} \right)^2 P_{\text{m,nl}}(k, z),$$

where  $C_1 \rho_{\text{cr}} \approx 0.0134$  is a constant and  $D(z)$  is the linear growth factor normalised to unity at  $z = 0$ . The model is empirically extended to non-linear scales by employing the full, non-linear matter power spectrum  $P_{\text{m,nl}}$ , encapsulating a mixture of non-linear gravitational effects as well as alignment processes on the scales of dark matter halos (cf. Schneider & Bridle 2010). The model has a single free parameter, namely, the global, dimensionless amplitude  $A_{\text{IA}}$ , which necessitates the assumption that there are no coherent trends in how galaxy ellipticities respond to their local tidal field within the survey's galaxy population. Since  $A_{\text{IA}}$  is well-constrained by tomographic cosmic shear data, we chose a wide top-hat prior for  $A_{\text{IA}}$  to ensure its posterior is data-driven. All previous KiDS cosmic shear analyses have used the NLA model as their default and, to date, there has been no evidence that this simple prescription is insufficient in representing current-generation datasets (e.g. Dark Energy Survey and Kilo-Degree Survey Collaboration et al. 2023).

#### 2.4.2. Additional redshift dependence (NLA- $z$ )

There is currently no observational evidence of redshift evolution in the amplitude of IA (Joachimi et al. 2011; Fortuna et al. 2021b), but the statistical uncertainties on measurements underpinning this statement are still large, especially when extrapolated to the redshift range of cosmic shear samples. Moreover,  $A_{\text{IA}}$  could also acquire an effective redshift scaling; for example, due to a combination of a luminosity dependence and Malmquist bias. Hence, we tested whether the data prefer  $A_{\text{IA}}$  to evolve over the tomographic bins. Since individual IA amplitudes per tomographic bin are poorly constrained (see the detailed analysis in Samuroff et al. 2019), we chose an extended model that is linear in terms of the scale factor,

$$P_{\delta\text{I,zdep}}^{(i)}(k, z) = - \left( A_{\text{IA}} + B_{\text{IA}} \left[ \frac{\langle a \rangle^{(i)}}{a_{\text{piv}}} - 1 \right] \right) \frac{C_1 \rho_{\text{cr}} \Omega_{\text{m}}}{D(z)} P_{\text{m,nl}}(k, z), \quad (7)$$

$$P_{\text{II,zdep}}^{(ij)}(k, z) = \left( A_{\text{IA}} + B_{\text{IA}} \left[ \frac{\langle a \rangle^{(i)}}{a_{\text{piv}}} - 1 \right] \right) \times \left( A_{\text{IA}} + B_{\text{IA}} \left[ \frac{\langle a \rangle^{(j)}}{a_{\text{piv}}} - 1 \right] \right) \left( \frac{C_1 \rho_{\text{cr}} \Omega_{\text{m}}}{D(z)} \right)^2 P_{\text{m,nl}}(k, z),$$

where  $\langle a \rangle^{(i)}$  is the average scale factor in tomographic bin  $i$ , evaluated using the estimated  $N(z)$ . The parameter  $A_{\text{IA}}$  thus is the IA amplitude at  $a_{\text{piv}} \approx 0.769$ , which corresponds to the commonly chosen pivot of  $z = 0.3$  for IA studies (Joachimi et al. 2011). We set the prior range of the second parameter,  $B_{\text{IA}}$ , by fitting our linear amplitude dependence to the compilation of KiDS and SDSS measurements studied in Fortuna et al. (2021b, see their Figure 10). Increasing the resulting constraint by 50%, to account for the limited coverage of these calibration samples and for potential additional implicit redshift dependencies, we obtain a Gaussian prior on  $B_{\text{IA}}$  with mean  $-3.7$  and standard deviation 4.3. For direct comparisons between amplitudes estimated be-

tween this model and others, we define

$$A_{\text{IA,total}}^{\text{NLA},z,(i)} = \hat{A}_{\text{IA}} + \hat{B}_{\text{IA}} \left( \frac{\langle a \rangle^{(i)}}{a_{\text{piv}}} - 1 \right), \quad (8)$$

where  $\hat{\cdot}$  denotes the best-fitted (i.e. maximum a posteriori) value of each parameter.

#### 2.4.3. Restricted TATT model (NLA- $k$ )

By design, the angular scale dependence of the NLA model is identical to that of the matter power spectrum projected over the redshift distribution of the sample. A natural way to extend this is to consider next-to-linear alignment effects, specifically: (i) a density weighting (Hirata & Seljak 2004), accounting for the fact that alignments are only measured at local overdensities of the matter distribution where galaxies reside; and (ii) quadratic alignments as expected for tidal torquing effects (Crittenden et al. 2001; Catelan et al. 2001).

These contributions are subsumed into the tidal alignment, tidal torquing (TATT) model (Blazek et al. 2019), each with a free amplitude parameter. As can be seen from Figure 1 of Blazek et al. (2019), the density-weighting and quadratic terms have very similar dependence on physical scale. Owing to the scale-mixing of angular statistics, cosmic shear measurements poorly distinguish small and localised scale differences, which leads to pronounced degeneracies among the four TATT parameters (e.g. Samuroff et al. 2019). To avoid these, we implement the full TATT model within the analysis pipeline, but fix all but one of the free parameters that govern the non-linear terms. We treat this model as a more general modification to NLA, to capture potential deviations in scale from the model. We chose to keep the density weighting amplitude term free, thereby adding a single additional free parameter, but note that this choice is fairly inconsequential: given the aforementioned similarity between the scale dependence of the terms, implementing only the density weighting or only the quadratic term would produce a similar end product. The NLA- $k$  model is then expressed as

$$P_{\delta\text{I,scaled}}(k, z) = P_{\delta\text{I,NLA}}(k, z) + A_{\text{IA}} b_{\text{src}} \frac{C_1 \rho_{\text{cr}} \Omega_{\text{m}}}{D(z)} [A_{0|0E}(k, z) + C_{0|0E}(k, z)],$$

$$P_{\text{II,scaled}}(k, z) = P_{\text{II,NLA}}(k, z) + A_{\text{IA}}^2 b_{\text{src}}^2 \left( \frac{C_1 \rho_{\text{cr}} \Omega_{\text{m}}}{D(z)} \right)^2 A_{0E|0E}(k, z) + 2 A_{\text{IA}}^2 b_{\text{src}} \left( \frac{C_1 \rho_{\text{cr}} \Omega_{\text{m}}}{D(z)} \right)^2 [A_{0|0E}(k, z) + C_{0|0E}(k, z)], \quad (9)$$

where the expressions for  $A_{0|0E}$ ,  $C_{0|0E}$ , and  $A_{0E|0E}$  are given in Blazek et al. (2019). We consider these terms in the Limber approximation where only transverse modes contribute, and assume that their redshift evolution is inherited from the linear matter power spectrum. In the original density-weighting term, the free parameter  $b_{\text{src}}$  can be interpreted as an effective galaxy bias of the weak lensing source sample, albeit on fairly non-linear scales. For our model, we chose a top-hat prior in the range  $[-0.5, 1.5]$  to ensure it encompasses NLA ( $b_{\text{src}} = 0$ ) and reasonable values expected for the (unknown) source galaxy bias ( $b_{\text{src}} \sim 1$ ). For comparisons between the alignment amplitude estimated with this model and others, we report the alignment

amplitude of the NLA component of the model from Eq. (9), namely,

$$A_{\text{IA,total}}^{\text{scaledep}} = \hat{A}_{\text{IA}}. \quad (10)$$

#### 2.4.4. Galaxy-type and mass-dependent model (NLA- $M$ )

Two trends in IA strength are firmly established observationally: (i) a dichotomy between late- and early-type galaxies where only the latter have yielded significant detections on cosmological scales (Samuroff et al. 2023; Johnston et al. 2019; Heymans et al. 2013; Georgiou et al. 2025; see also Mandelbaum et al. 2011; Tonegawa et al. 2025; McCullough et al. 2024 for additional examples of null or marginal detections of late-type alignments); and (ii) a scaling of early-type alignment amplitude with host halo mass that is well described by a power-law (Joachimi et al. 2011; Fortuna et al. 2025). Both have solid theoretical underpinning. First, rotationally supported late-type galaxies are expected to align through tidal torquing, which depends quadratically on the tidal quadrupole, whereas spheroid shapes should respond linearly and thus be well described by NLA-like models. Second, Piras et al. (2018) motivated the halo mass dependence through an analytical argument and also reproduced it for simulated dark matter haloes.

We incorporated these IA trends explicitly into the amplitude of the NLA model. We assumed that, in each tomographic bin, only the fraction  $f_r$  of ‘red’ (i.e. early-type) galaxies intrinsically align, whereas the remaining galaxies have zero alignment. The latter assumption is consistent with observations, but within a large statistical uncertainty and involving a substantial extrapolation in sample properties, primarily due to the imperfect relation between galaxy colour and rotational/pressure support (which is the property most-likely to cleanly stratify intrinsic alignment amplitudes). Moreover, we assumed that the red galaxy population IA amplitude has a power-law dependence on the average halo mass  $\langle M_h \rangle$  within a tomographic bin. This leads to

$$P_{\delta\text{I,massdep}}^{(i)}(k, z) = -A_{\text{IA}} f_r^{(i)} \left( \frac{\langle M_h \rangle^{(i)}}{M_{\text{h,piv}}} \right)^\beta \frac{C_1 \rho_{\text{cr}} \Omega_m}{D(z)} P_{\text{m,nl}}(k, z), \quad (11)$$

$$P_{\text{II,massdep}}^{(ij)}(k, z) = A_{\text{IA}}^2 f_r^{(i)} f_r^{(j)} \left( \frac{\langle M_h \rangle^{(i)}}{M_{\text{h,piv}}} \right)^\beta \left( \frac{\langle M_h \rangle^{(j)}}{M_{\text{h,piv}}} \right)^\beta \times \left( \frac{C_1 \rho_{\text{cr}} \Omega_m}{D(z)} \right)^2 P_{\text{m,nl}}(k, z),$$

where we adopt the pivot halo mass  $M_{\text{h,piv}} = 10^{13.5} h^{-1} M_\odot$  from Fortuna et al. (2025). The same authors also fit the power-law dependence to a compilation of early-type IA measurements (see their Figure 4), and we adopt their joint posterior on  $A_{\text{IA}}$  and  $\beta$  as our prior for cosmic shear inference. Their posterior is well approximated by a bivariate Gaussian with  $A_{\text{IA}} = 5.74 \pm 0.29$ ,  $\beta = 0.44 \pm 0.03$ , and a correlation coefficient of  $-0.59$ . Details on how we determine  $f_r$  and halo masses for our sample of cosmic shear sources, and how we treat their associated uncertainties, are provided in Appendix B. For comparisons between the alignment amplitude estimated with this model and others, we define

$$A_{\text{IA,total}}^{\text{massdep,(i)}} = \hat{A}_{\text{IA}} f_r^{(i)} \left( \frac{\langle M_h \rangle^{(i)}}{M_{\text{h,piv}}} \right)^\beta. \quad (12)$$

## 2.5. Cosmic shear summary statistics

In cosmic shear analyses, it is most convenient to measure the signal using shear two-point correlation functions (2PCFs) while the modelling is done through shear power spectra (see Eq. 1). Alternatively, one could take the Fourier transform of the shear field and work in harmonic space, which brings the measurements closer to the theoretical predictions. We chose to use 2PCFs for the measurement, as they shield us from a particular form of  $E$ -/ $B$ -mode mixing that occurs in harmonic space in the presence of the (unavoidable) image masks. In addition, the shot-noise term does not impact 2PCFs (present only in their covariance), as it is only relevant at zero lag. In harmonic space, both shot noise and masking must be modelled and corrected, prior to beginning cosmological analyses. For more details see e.g. Alonso et al. (2019).

Our cosmological analysis here leverages two-point statistics: the complete orthogonal sets of  $E$ -/ $B$ -integrals (COSEBIs, Sect. 2.5.2) and band power spectra (Sect. 2.5.3), constructed using linear combinations of finely-binned 2PCFs (Sect. 2.5.1).

### 2.5.1. 2PCFs ( $\xi_\pm$ )

The shear two-point correlation functions ( $\xi_\pm$ , Kaiser 1992) are functions of the tangential and cross components of shear ( $\gamma_t$  and  $\gamma_\times$ , respectively) defined with respect to the line connecting pairs of galaxies with angular separation  $\theta$  (see e.g. Bartelmann & Schneider 2001),

$$\xi_\pm(\theta) = \langle \gamma_t \gamma_t \rangle(\theta) \pm \langle \gamma_\times \gamma_\times \rangle(\theta). \quad (13)$$

The measured 2PCFs are binned in  $\theta$  to reduce measurement noise and compress the data. In practice, we measure  $\xi_\pm(\theta)$  for 1000 logarithmically spaced  $\theta$ -bins between  $\theta_{\text{min}} = 2'$  and  $\theta_{\text{max}} = 300'$  using TREECORR<sup>1</sup> (Jarvis et al. 2004). The inputs to TREECORR are galaxy catalogues, which include galaxy ellipticities ( $\epsilon_1$  and  $\epsilon_2$ ), galaxy positions (right ascension and declination), and galaxy weights ( $w$ ). The weights used by TREECORR are equal to the recalibrated shape weights (Sect. 3.5) multiplied by the redshift distribution gold-weights (Sect. 3.4). Our  $\xi_\pm(\theta)$  are then estimated using

$$\hat{\xi}_\pm^{(ij)}(\bar{\theta}) = \frac{\sum_{ab} w_a w_b [\epsilon_{t,a}^{\text{obs}} \epsilon_{t,b}^{\text{obs}} \pm \epsilon_{\times,a}^{\text{obs}} \epsilon_{\times,b}^{\text{obs}}] \Delta_{ab}^{(ij)}(\bar{\theta})}{\sum_{ab} w_a w_b (1 + m_a)(1 + m_b) \Delta_{ab}^{(ij)}(\bar{\theta})}, \quad (14)$$

where  $\Delta_{ab}^{(ij)}(\bar{\theta})$  is the bin selection function (which limits the sums to galaxy pairs of separation within the angular bin labelled by  $\bar{\theta}$  and the tomographic bins  $i$  and  $j$ ), and  $\epsilon_t^{\text{obs}}$  and  $\epsilon_\times^{\text{obs}}$  are the projected tangential and radial components of observed ellipticities between pairs of galaxies. The estimator includes the galaxies’ multiplicative shear measurement biases  $m_a$  estimated via image simulations (Sect. 3.4).

Using the flat-sky approximation, we can relate the 2PCFs to the angular (shear) power spectra by Hankel transformations:

$$\xi_\pm^{(ij)}(\theta) = \int_0^\infty \frac{\ell d\ell}{2\pi} [C_{\epsilon\epsilon,E}^{(ij)}(\ell) \pm C_{\epsilon\epsilon,B}^{(ij)}(\ell)] J_{0/4}(\ell\theta), \quad (15)$$

where  $C_{\epsilon\epsilon,E/B}^{(ij)}(\ell)$  are the  $E$ -/ $B$ -mode angular power spectra for redshift bins  $i$  and  $j$ , and  $J_{0/4}(x)$  are the zeroth- and fourth-order Bessel functions of the first kind. We expect cosmic shear to produce only  $E$ -mode signals up to first order (Schneider et al.

<sup>1</sup> We use TREECORR version 4.2.3, with binslop= 1.5.

1998; Hilbert et al. 2009), while data systematic errors can in principle create both  $E$  and  $B$  modes. As a result,  $B$ -modes have long been used as a diagnostic tool in cosmic shear analyses (see e.g. Pen et al. 2002; Hoekstra et al. 2002; Heymans et al. 2005; Kilbinger et al. 2013).

As  $\xi_{\pm}$  mix  $E$  and  $B$  modes, they are not ideal for cosmic shear analysis. In addition, they variably weight a large range of physical scales in each angular bin, increasing sensitivity to scale-dependent effects such as baryonic feedback (which must be marginalised over using imperfect models). Figure 1 shows the range of  $\ell$  probed by the  $\xi_{\pm}$  over a range of angular separations. The  $J_0$  filter in Eq. (15) means that  $\xi_+$  mixes a very wide range of scales ( $1 \lesssim \ell \lesssim 10^4$ ) and  $J_4$  makes  $\xi_-$  particularly sensitive to smaller scales ( $10 \lesssim \ell \lesssim 10^3$ ). This breadth can create a problem when using  $\xi_{\pm}$  for inference, as the power spectra and intrinsic alignment models become increasingly uncertain at smaller scales. Consequently, we expect the cosmological constraints to be possibly affected by a range of systematic effects that may introduce biases and, as such, we do not include  $\xi_{\pm}$  in our primary cosmological analyses. Nevertheless, results for  $\xi_{\pm}$  are provided for completeness and comparison with previous work, in Appendix F.

### 2.5.2. COSEBIs ( $E_n/B_n$ )

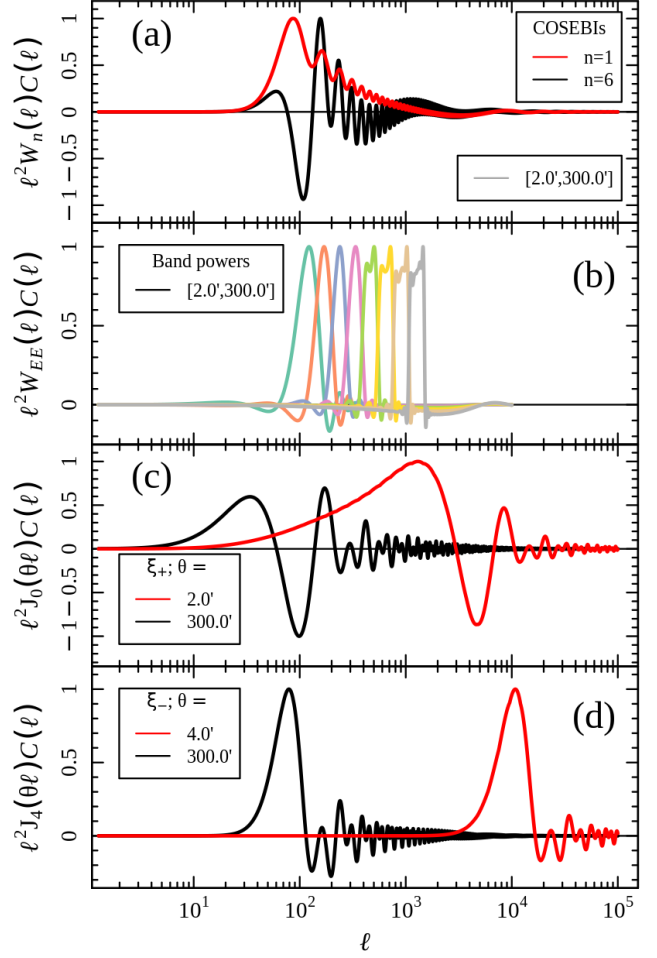
The COSEBIs (Schneider et al. 2010) were originally designed to cleanly separate  $E$ - and  $B$ -modes over a finite  $\theta$ -range. This was achieved through the use of filter functions  $T_{\pm n}(\theta)$ , which form complete bases, creating well-defined COSEBIs  $E$ - and  $B$ -modes ( $E_n/B_n$ ). COSEBIs can be calculated or measured via a linear transformation of  $\xi_{\pm}$ ,

$$\begin{aligned} \hat{E}_n &= \frac{1}{2} \int_{\theta_{\min}}^{\theta_{\max}} d\theta \theta [T_{+n}(\theta) \xi_+(\theta) + T_{-n}(\theta) \xi_-(\theta)], \\ \hat{B}_n &= \frac{1}{2} \int_{\theta_{\min}}^{\theta_{\max}} d\theta \theta [T_{+n}(\theta) \xi_+(\theta) - T_{-n}(\theta) \xi_-(\theta)]. \end{aligned} \quad (16)$$

The COSEBIs modes  $n$  start from 1 and go to infinity. However, the first few COSEBIs are sufficient to capture essentially all the cosmological information in each pair of tomographic bins (Asgari et al. 2012). This means that COSEBIs naturally compress the data into well-defined modes, once  $\theta_{\min}$  and  $\theta_{\max}$  are defined, without any need for subsequent data compression. In previous KiDS analyses (e.g. Asgari et al. 2021), the first five COSEBIs modes were used for cosmological analyses. In this work, however, we add the sixth mode, as we found that (with our increased constraining power, and larger redshift baseline) the sixth mode contains relevant cosmological information and thus improves our constraining power on  $S_8$ . Conversely, the addition of modes 7 – 20 did not result in any improvement of constraining power, returning constraints identical to those measured in the 6 mode case.

We use the relation between COSEBIs and shear power spectra to make our predictions,

$$\begin{aligned} E_n &= \int_0^{\infty} \frac{d\ell \ell}{2\pi} C_{EE}(\ell) W_n(\ell), \\ B_n &= \int_0^{\infty} \frac{d\ell \ell}{2\pi} C_{BB}(\ell) W_n(\ell), \end{aligned} \quad (17)$$



**Fig. 1.** Integrands of the transformation between the angular power spectrum and  $E_n/B_n$  (Eq. 17, panel ‘a’),  $C_{E,B}$  (Eq. C.5, panel ‘b’), and  $\xi_{\pm}$  (Eq. 15, panels ‘c’ and ‘d’). All solid integrands are normalised by their maximum absolute value. For both COSEBIs and 2PCFs, we show the limiting ranges of the statistics ( $n \in \{1, 6\}$  for  $E_n$  and  $\theta \in \{2.0, 300.0\}$  arcmin for  $\xi_{\pm}$ ). For  $C_{E,B}$  we show all eight bins used in KiDS-Legacy, defined between  $\ell = 100$  and  $\ell = 1500$ . Note that, unlike previous KiDS analyses, our new treatment of  $T(\theta)$  means that the  $C_{E,B}$  initially go to zero within their desired  $\ell$  ranges.

where the weight functions  $W_n(\ell)$  are

$$\begin{aligned} W_n(\ell) &= \int_{\theta_{\min}}^{\theta_{\max}} d\theta \theta T_{+n}(\theta) J_0(\ell\theta) \\ &= \int_{\theta_{\min}}^{\theta_{\max}} d\theta \theta T_{-n}(\theta) J_4(\ell\theta). \end{aligned} \quad (18)$$

Figure 1 shows two example integrands from Eq. (17).  $W_n(\ell)$  are highly oscillatory weight functions with a limited range of support over  $\ell$ . This allows us to truncate our integrals in Eq. (17) to a finite range in  $\ell$ , as the  $E_n$  integrand in Eq. (17) is practically zero outside  $10 \lesssim \ell \lesssim 10^3$ .

### 2.5.3. Band powers ( $C_{E,B}$ )

Band power spectra (hereafter ‘band-powers’) refer to binned angular power spectra measured by integrating over  $\xi_{\pm}(\theta)$  (Schneider et al. 2002; van Uitert et al. 2016; Joachimi et al.

2021). Ideally, we want a finite number of bins, labelled as  $l$ , where each bin is only sensitive to a narrow range of  $\ell$ -values between  $\ell_{\text{lo},l}$  and  $\ell_{\text{up},l}$ . We also want the  $E$ - and  $B$ -mode band-powers to be pure, namely, no leakage from  $B$  modes into  $E$  modes and vice versa. For this ideal case, we can express the following:

$$C_{E/B,l}^{(ij)} := \frac{1}{N_l} \int_0^\infty d\ell \ell S_l(\ell) C_{\varepsilon\varepsilon,E/B}^{(ij)}(\ell), \quad (19)$$

for redshift bins  $i$  and  $j$ , where  $S_l(\ell)$  is the response function for bin  $l$ . Here we have normalised the integral with

$$N_l = \int_0^\infty d\ell \ell^{-1} S_l(\ell). \quad (20)$$

In this work, we consider the ideal case of a top-hat response function between  $\ell_{\text{lo},l}$  and  $\ell_{\text{up},l}$  which simplifies the normalisation to

$$N_l = \log(\ell_{\text{up},l}) - \log(\ell_{\text{lo},l}), \quad (21)$$

and produces a band-power that traces  $\ell^2 C(\ell)$  at the logarithmic centre of the bin. We define eight band-powers logarithmically binned between  $\ell_{\text{min}} = 100$  and  $\ell_{\text{max}} = 1500$ . In practice, measuring band-powers with perfect top-hat functions is infeasible, as  $\xi_{\pm}(\theta)$  is only measured for a finite  $\theta$ -range. This has the adverse effect that our band-powers will inevitably mix  $E$  and  $B$  modes<sup>2</sup>. The mixing increases for narrower  $\ell$ -bins and when the available range of  $\theta$ -scales is smaller (Asgari & Schneider 2015). With this definition for band-powers, we forgo perfect  $E$ -/ $B$ -mode separation in favour of more control over the  $\ell$ -values that contribute to our analyses. We present the detailed derivation of our band-power windows  $C_{E,l}$  in Appendix C, and show the integrands of  $C_{E,l}$  in Fig. 1. Note, however, that our integrands behave differently to those in Asgari et al. (2021), due to our different treatment of the edges of  $T(\theta)$  (Appendix C) and due to our different scale-cuts.

## 2.6. Analytic covariance modelling

We presented the covariance modelling for KiDS-Legacy and the `ONECOVARIANCE` code<sup>3</sup> in Reischke et al. (2025). In general, the methodology for the cosmic shear part still follows closely the covariance modelling of KiDS-1000 (Joachimi et al. 2021). Major changes to the previous analysis include: (i) the possibility of including non-uniform source distributions in the mixed term of the covariance from the triplet counts of the survey; (ii) including non-binary masks for the area calculation and survey response; and (iii) offering a more generalised framework for a consistent treatment of all summary statistics used in KiDS-Legacy.

The cosmic shear covariance,  $\mathbf{C}$ , generally consists of four contributions:

$$\mathbf{C} = \mathbf{C}_G + \mathbf{C}_{\text{NG}} + \mathbf{C}_{\text{SSC}} + \mathbf{C}_{\text{mult}}, \quad (22)$$

with the Gaussian contribution  $G$ , the non-Gaussian term  $\text{NG}$  due to non-linear in-survey modes, the super-sample covariance  $\text{SSC}$  from modes larger than the survey, and lastly the contribution due to the uncertainty in the multiplicative shear bias calibration,  $\mathbf{C}_{\text{mult}}$ . Asgari et al. (2021); Secco et al. (2022) included

<sup>2</sup> Other binning options are also explored in the literature, which can alleviate some of the issues with the top-hat response function at the expense of broadening the range of  $\ell$ -values that contribute to each bin (see e.g. Becker et al. 2016).

<sup>3</sup> <https://github.com/rreischke/OneCovariance>

the multiplicative shear bias in the theory prediction instead and marginalised over in the sampling process. If the residual uncertainties on  $m_a$  are small, these two approaches are equivalent. The Gaussian contribution can further be split into three parts:

$$\mathbf{C}_G = \mathbf{C}_{\text{sva}} + \mathbf{C}_{\text{sn}} + \mathbf{C}_{\text{mix}}, \quad (23)$$

describing the sample variance term ‘sva’ due to the finite number of available modes in the survey volume, the pure shot noise term ‘sn’ due to the finite number of sources and their composition, and the mix term ‘mix’.

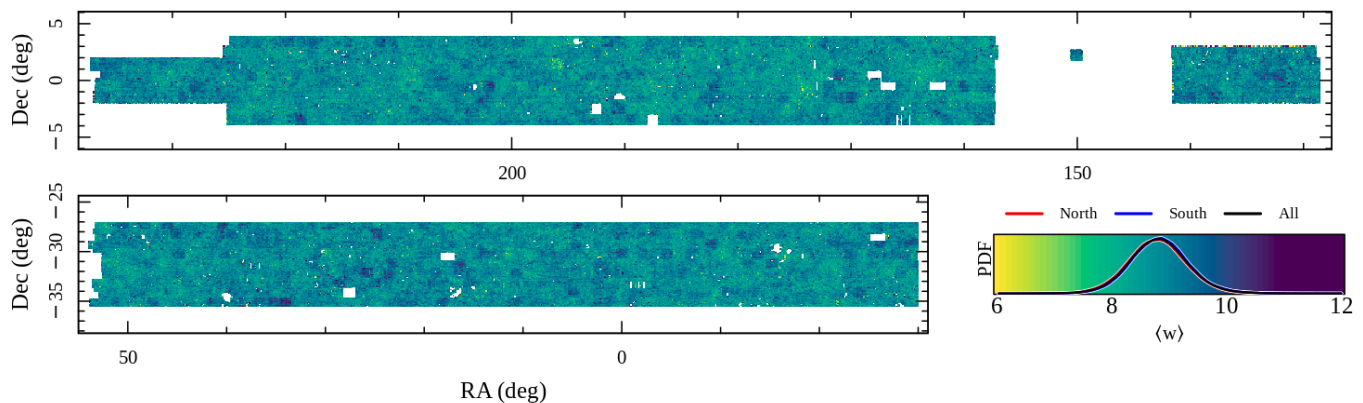
We calculated the covariances for angular power spectra and transform those to the three summary statistics considered. The only exception was the pure noise term, which was directly computed from the weighted number of pairs in the catalogue.  $\text{SSC}$  and  $\text{NG}$  terms were modelled with a halo model approach and the survey variance was directly calculated from the survey mask. Our analytical prescription was tested against GLASS (Tessore et al. 2023) mocks to test the impact of variable depth and the overall agreement. These mocks are discussed in more detail in Appendix D. In Reischke et al. (2025), we showed that the agreement between our analytic covariance and that constructed from GLASS is within 10 percent at large angular separations and below 5 percent for scales below 0.5 degrees. Furthermore, we found that our idealised mix term (which does not account for non-uniformity of the source distribution) did not affect our inference results significantly, and as a result was neglected in our analyses.

## 3. Data and analysis

In this section, we summarise the KiDS-Legacy lensing data and its calibration, the associated image simulations, and our analysis blinding strategy. Our analysis was performed using a new version of the `CosmoPipe` infrastructure presented in Wright et al. (2020). Our fiducial pipeline includes all steps required for a catalogues-to-cosmology analysis, leveraging three primary inputs: a (blinded) shear catalogue (Sect. 3.3), a spectroscopic compilation (Sect. 3.2), and a simulated galaxy lightcone (Sect. 3.4). With these data products, the pipeline performs nine primary functions: pre-processing of the wide-field catalogue, simulated catalogue construction, redshift calibration,  $N(z)$  construction, shear calibration, data vector construction, covariance construction, cosmological inference, and post-processing. Details of each processing step are provided alongside the public code base<sup>4</sup>.

Figure 2 shows the footprint of the KiDS-Legacy sample, split between the two survey patches of KiDS (one in each of the northern and southern Galactic caps). The figure is coloured by the mean effective weight (i.e. including both shape measurement and redshift distribution estimation weights) of sources on sky, which is an approximate tracer of survey properties such as the size of the point-spread function and variable depth (see Appendix D and Yan et al. 2025). The weight is defined as described in Section 7 of Wright et al. (2024), and has the range  $w \in [0, \sigma_{\text{pop}}^{-2}]$ , where  $\sigma_{\text{pop}} = 0.255$  is an approximate intrinsic ellipticity dispersion of all sources in KiDS (without e.g. tomographic or brightness selections) defined for Legacy in Li et al. (2023b).

<sup>4</sup> [github.com/AngusWright/CosmoPipe](https://github.com/AngusWright/CosmoPipe)



**Fig. 2.** Distribution of the mean weight of the KiDS-Legacy source galaxies in on-sky bins ( $0.1 \times 0.1$  degrees). The weights here include both shape measurement weights (Sect. 3.5) and redshift distribution gold weighting (Sect. 3.4).

**Table 1.** Properties of the six KiDS-Legacy tomographic bins.

Tomographic Bin	photo- $z$ Selection	$N_{\text{gold}}$	$n_{\text{eff, gold}}$ [arcmin $^{-2}$ ]	$\sigma_{\epsilon, \text{gold}}$	$\langle z \rangle_{\text{est}}$	$\delta z$ [ $\times 10^{-2}$ ]	$m_{\text{gold}}$ [ $\times 10^{-2}$ ]
1	$0.10 < z_{\text{B}} \leq 0.42$	7 416 371	1.77	0.28	0.335	$-2.6 \pm 0.2 \pm 1.0$	$-2.3 \pm 0.6$
2	$0.42 < z_{\text{B}} \leq 0.58$	7 359 911	1.65	0.27	0.477	$1.4 \pm 0.1 \pm 1.0$	$-1.6 \pm 0.6$
3	$0.58 < z_{\text{B}} \leq 0.71$	6 799 681	1.50	0.29	0.587	$-0.2 \pm 0.2 \pm 1.0$	$-1.1 \pm 0.7$
4	$0.71 < z_{\text{B}} \leq 0.90$	6 880 432	1.46	0.26	0.789	$0.8 \pm 0.1 \pm 1.0$	$2.0 \pm 0.7$
5	$0.90 < z_{\text{B}} \leq 1.14$	6 477 538	1.35	0.28	0.940	$-1.1 \pm 0.2 \pm 1.0$	$3.0 \pm 0.8$
6	$1.14 < z_{\text{B}} \leq 2.00$	5 960 461	1.07	0.30	1.224	$-5.4 \pm 0.4 \pm 1.0$	$4.5 \pm 0.9$
1–6	$0.10 < z_{\text{B}} \leq 2.00$	40 894 394	8.79	0.28	0.680	–	–

**Notes.** All values are computed for the final gold-selected sample of sources, and with the appropriate shape-measurement and gold weights applied. Where relevant values are computed using an effective area of  $967.4 \text{ deg}^2$ . Values of  $\sigma_{\epsilon}$  and  $n_{\text{eff}}$  are computed using Eqs. C.9 and C.12 of Joachimi et al. (2021), respectively. The  $\sigma_{\epsilon}$  values are the ellipticity dispersion averaged over both components.  $m_{\text{gold}}$  is the multiplicative shear measurement bias, averaged over both ellipticity components, estimated from our SKILLS simulations.  $\sigma_{\epsilon}$  values include the appropriate correction for multiplicative shear biases (i.e. the  $m_{\text{gold}}$  column).

### 3.1. KiDS and VIKING

The KiDS-Legacy lensing sample was drawn from the fifth and final data release of the KiDS survey<sup>5</sup>, KiDS-DR5, described in Wright et al. (2024). It includes 1347 square-degree tiles observed in the  $u, g, r$ , and  $i$  filters with the VST (Capaccioli et al. 2005) and OmegaCAM (Kuijken 2011) between 2011 and 2019, as well as the 2009–2018 VIKING survey VISTA/VIRCAM observations (Edge et al. 2013) of the same area in  $Z, Y, J, H$ , and  $K_s$ . Additionally, the dataset contains spectroscopic calibration fields (the ‘KiDZ’ fields, 27 square-degree tiles in total, including four that are part of the main survey).

KiDS-DR5 significantly extends the previous (fourth) data release (Kuijken et al. 2019), which was the basis for the KiDS-1000 analyses (Asgari et al. 2021; Heymans et al. 2021; Tröster et al. 2021). It represents a substantial increase in survey area (by 34%) and in  $i$ -band depth (doubling the exposure time with a second pass). The volume of photometric redshift calibration data has also grown, including photometry for 5 times more spectroscopic sources than in the original KiDS-1000 analysis (Hildebrandt et al. 2021). The deeper  $i$ -band depth and more extensive calibration data has allowed us to push the photometric redshift limit for galaxies in the lensing sample from 1.2 to 2. The combined effect of the increase in area and depth is an increase in the survey volume of a factor of 3.5.

<sup>5</sup> [https://www.eso.org/sci/observing/phase3/news.html#KIDS\\_DR5](https://www.eso.org/sci/observing/phase3/news.html#KIDS_DR5)

Several aspects of the data analysis were changed between DR4 and DR5, as detailed in Wright et al. (2024). The astrometry reference catalogue for the THELI reduction moved from the Sloan Digital Sky Survey (SDSS Ahumada et al. 2020) and the 2-micron All Sky Survey (2MASS Skrutskie et al. 2006) to Gaia DR2 (Gaia Collaboration: Brown et al. 2018), and the astrometric solutions and photometric zero-point determinations in the ASTRO-WISE  $u$ -band reductions were made more robust. Also, the manual and automated masking of artefacts in the images, mostly due to ghost reflections, artificial satellite tracks, and detector instabilities, was revisited and improved for the final release. Most importantly, the weak lensing shape measurement method uses an updated version of *lensfit* (v321), with improvements in the sampling of the ellipticity likelihood surface and of the treatment of residual PSF leakage in the shapes and associated weights (Li et al. 2023b). Details and tests of these new implementations are given in Wright et al. (2024), and in the sections below. A summary of the properties of the KiDS-Legacy sample is given in Table 1, including details of tomographic bin limits, final effective number densities of galaxies (after redshift distribution calibration), and redshift and shape measurement bias parameters.

Lastly, we note that a pernicious astrometric issue was identified during the investigation of our B-mode null tests (Sect. 4.3), which resulted in additional masking of approximately 4% of the survey sources to be used in the KiDS-Legacy analysis. More details are given in Sect. 4.3 and Appendix E. The total area after

masking is  $967.4 \text{ deg}^2$ , with  $n_{\text{eff}} = 8.79$  per arcminute within the photometric redshift limits of KiDS-Legacy.

### 3.2. KiDZ

The KiDZ data consist of VST and VISTA images that target well-studied spectroscopic survey fields, including the Cosmic Evolutions Survey field (COSMOS; Scoville et al. 2007), the DEEP2 Galaxy Redshift Survey (DEEP2; Newman et al. 2013) 02h and 23h fields, the *Chandra* Deep Field South (CDFs), the Visible Multi-object Spectrograph (VIMOS) Very Large Telescope (VLT) Deep Survey (VVDS; Le Fèvre et al. 2005) 14h field, and the VIMOS Public Extragalactic Redshift Survey (VIPERS; Guzzo et al. 2014) W1 and W4 fields (a full list is given in Wright et al. 2024). They were purposely taken under very similar observing conditions, and with the same observing setup, as the main KiDS/VIKING survey and serve as the calibration sample for photometric redshifts of the faint KiDS sources. A total of 126 085 sources with spectroscopic redshifts have nine-band KiDZ photometry. Additional photometric redshift calibration via cross-correlations uses the overlap between KiDS and shallower, wide-angle spectroscopic redshift surveys, principally: SDSS, the Galaxy And Mass Assembly (Driver et al. 2022) survey, 2-degree Field Lensing Survey (Blake et al. 2016), WiggleZ survey (Blake et al. 2008), and the Dark Energy Spectroscopic Instrument (DESI Collaboration et al. 2024).

### 3.3. Blinding

We used the same blinding methodology as the one implemented in Asgari et al. (2021), first presented in Kuijken et al. (2015). This process involves the double-blind generation of three realisations of our shape measurement catalogues. One of these is the true catalogue, while in the other two there are induced systematic differences in measured shapes that result in (up to)  $\pm 2\sigma$  change in the inferred  $S_8$ . The true catalogue is equally likely to be the one that yields the highest, middle, or lowest  $S_8$ . These blinded data vectors were analysed throughout the project until the analyses were finalised, cosmological chains were run, the results written up in this publication, and the manuscripts had undergone internal review within the KiDS collaboration. As such, this manuscript (for example) was written containing results for all three blinds, with only the discussion and conclusion sections written after unblinding.

The authors note (for the reference of future surveys) that this blinding scheme effectively hampered the diagnosis of subtle systematic effects in the dataset (see Appendix E), as we were limited in our ability to directly compare measured shapes between previously constructed catalogues and our KiDS-Legacy catalogues. Such comparisons are extremely useful for the validation of spatially localised systematic patterns in the shape catalogue, which are obfuscated when translated into two-point statistics. This blinding strategy, however, makes such direct comparisons between shape catalogues problematic, as it can also lead to unblinding.

### 3.4. Redshift distributions and calibration

Redshift distributions (hereafter ‘ $N(z)$ ’) used in KiDS-Legacy were estimated and calibrated leveraging both colour-based direct calibration using self-organising maps (SOMs Kohonen 1982) and spatial cross-correlations (CC) between tracer spec-

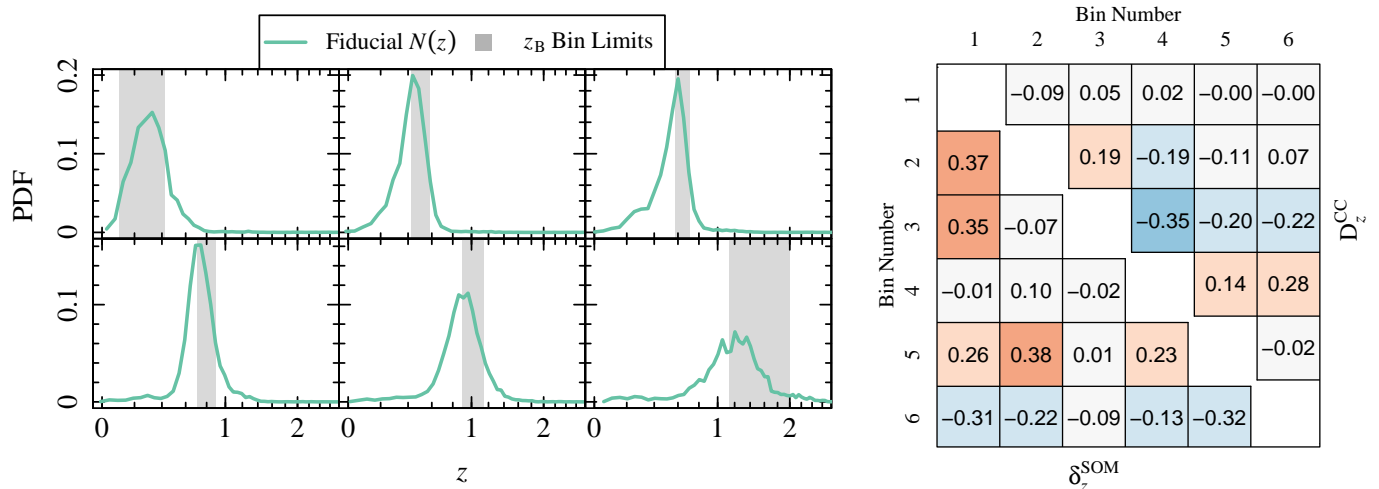
troscopic samples and the KiDS-Legacy sources. For a detailed description of the  $N(z)$  construction and validation process used in this work, we refer to Wright et al. (2025). In brief, the  $N(z)$  here differ in their construction (compared to previous work from KiDS) in three main regards: a considerably larger redshift calibration sample from KiDZ (see Sect. 3.2); a new weighting scheme that replaces the binary ‘gold-class’ with a continuous ‘gold-weight’; and through the optional inclusion of weights for spectroscopic objects, such as our so-called ‘prior-volume weighting’. Gold-weights primarily account for noise in the colour-based association between wide-field sources and their calibrating spectra, and are computed by repeating the gold classification  $N$  times (where we arbitrarily chose  $N = 10$ ). Similarly, our prior-volume weighting applies an a priori weighting to the distribution of calibrating spectra as a function of redshift, to reduce bias caused by colour-redshift degeneracies and the complex redshift selection function of our spectroscopic compilation.

The redshift distribution calibration for KiDS-Legacy made use of two suites of simulations. For the calibration of the SOM redshift distributions (and calibration of source shape measurements, Sect. 3.5), we used the multi-colour SKILLS simulation of Li et al. (2023b). For additional SOM redshift distribution calibration and cross-correlation redshift distribution calibration, we used the MICE2 simulated lightcone (Fosalba et al. 2015a; Crocce et al. 2015; Fosalba et al. 2015b; Carretero et al. 2015; Hoffmann et al. 2015) adapted to KiDS noise levels and sample selections (van den Busch et al. 2020).

Figure 3 shows the fiducial redshift distributions used in this cosmological analysis. These  $N(z)$  include the appropriate  $\delta z$  shifts in each of their mean, as inferred using our simulations and described in Wright et al. (2025). As such, the  $N(z)$  presented are shown as-used in the generation of the theory/models during parameter inference. In Wright et al. (2025) we quantified the impact of a series of different analysis choices for  $N(z)$  estimation and calibration, demonstrating that they have no significant impact on the final  $N(z)$  that we used for cosmological inference. The figure also presents the off-diagonal elements of the correlation matrices for  $\delta z$ , used to construct our correlated  $\delta z$  priors. The lower triangle shows the correlation matrix estimated from our SOM redshift distribution calibration, while the upper triangle shows the correlation matrix for the CC estimated redshift distributions.

### 3.5. Shape measurement and calibration

The measurement and calibration of source shapes in KiDS-Legacy were performed with the latest KiDS methods, which have already been established in the literature. We refer to Li et al. (2023b) for an extensive description of the state-of-the-art for shape measurement and calibration in KiDS. Topics discussed therein include: the *lensfit* algorithm (v321; see also Miller et al. 2007, 2013; Fenech Conti et al. 2017), used for measuring shapes of galaxies (including modelling of the PSF); shape calibration using SKILLS image simulations; and PSF leakage corrections (with minor updates to the PSF leakage estimation as described in Wright et al. 2024). The SKILLS image simulations are multi-band image simulations, constructed specifically to match the observational properties of KiDS-Legacy in all available bands, and featuring realistic clustering and correlations between galaxy properties and environment, allowing for realistic treatment of blending effects. The simulations also feature redshift-dependent shear (MacCrann et al. 2022), although this is determined to be of negligible importance



**Fig. 3.** *Left:*  $N(z)$  used in the fiducial cosmic shear analyses of KiDS-Legacy. Each panel is one tomographic bin whose definition is determined by the grey band in photo- $z$ . The  $N(z)$  have been shifted according to the mean bias estimated using our SKiLLS simulations, making each  $N(z)$  a correct representation of the effective  $N(z)$  that is used in the cosmological analyses. Fiducial biases are provided in Table 1. *Right:*  $\delta z$  correlation matrix for our fiducial  $N(z)$  ( $\delta z^{\text{SOM}}$ , lower triangle) and the correlation matrix inferred from our cross correlation  $N(z)$  on the data ( $D_z^{\text{CC}}$ , upper triangle).

to the computation of multiplicative bias (Li et al. 2023b) and redshift distribution bias (Wright et al. 2025) at the sensitivity of KiDS-Legacy.

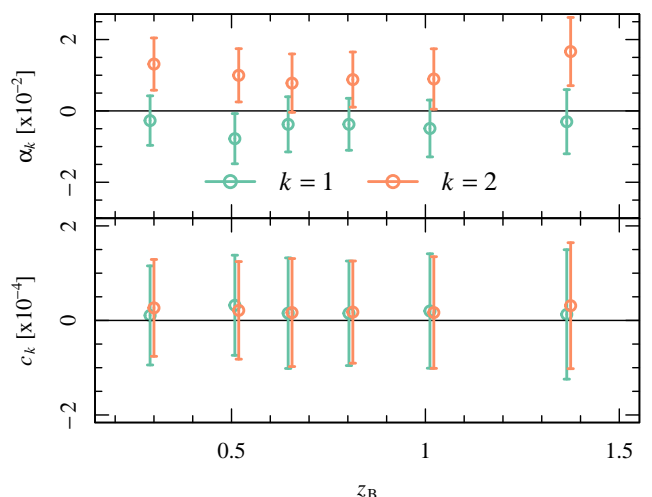
The multiplicative bias estimation in KiDS-Legacy was performed using the same suite of multi-colour image simulations from SKiLLS used for redshift distribution calibration. Redshift distribution calibration utilises many realisations of the sample to estimate uncertainties on tomographic bias parameters (see Wright et al. 2025). Afterwards, these samples were also used to estimate the multiplicative shear measurement bias,  $m$ , per tomographic bin. This means that, for the first time, we have been able to directly estimate the covariance between our redshift distribution bias calibration and our multiplicative shear bias estimation. The resulting covariance matrix (and the corresponding correlation matrix) are functionally block-diagonal: there is no cross covariance between  $m$  and  $\delta z$  estimates.

We checked for the degree of PSF leakage measured in the lensing sample through a direct empirical calculation of the correlation between PSF and galaxy ellipticities. Here, we document the measured PSF contamination in our KiDS-Legacy sample, and the measured residual one- and two-dimensional constant ellipticity terms ( $c$ -terms) in the dataset. We adopted the first-order systematics model from Heymans et al. (2006):

$$\varepsilon_k^{\text{obs}} = (1 + m_k)(\varepsilon_k^{\text{int}} + \gamma_k) + \alpha_k \varepsilon_k^{\text{PSF}} + c_k, \quad k \in \{1, 2\}, \quad (24)$$

where  $k$  denotes the individual ellipticity components,  $\gamma$  is the (true) shear, and the superscripts ‘obs’, ‘int’, and ‘PSF’ refer to a source’s observed, intrinsic, and modelled PSF ellipticity, respectively. In the limit of large numbers of sources averaged over a large sky area, we can assume  $\langle (1 + m_k)(\varepsilon_k^{\text{int}} + \gamma_k) \rangle = 0$ , and therefore model the residual PSF contamination using a simple linear regression between observed source ellipticities (i.e. after recalibration to remove PSF leakage Li et al. 2023b; Wright et al. 2024) and modelled PSF ellipticities.

Figure 4 shows the  $\alpha_k$  and  $c_k$  PSF leakage terms in each ellipticity component of the KiDS-Legacy sample, measured in each tomographic bin. Note that the shape recalibration process used in KiDS-Legacy necessitates that the measured PSF leakage fac-



**Fig. 4.** PSF leakage measured in the KiDS-Legacy lensing sample after recalibration of shapes, computed per tomographic bin and ellipticity component  $k$ , and following the first-order systematics model in Eq. (24).

tors be approximately zero, as the recalibration specifically addresses any residual in this parameter, in bins of source signal-to-noise ratio (SN) and resolution (Li et al. 2023b). The residual  $\alpha_k$  are all negligible, with the  $\alpha_2$  having a slightly larger residual amplitude overall ( $\alpha_2 \approx 0.01$ , roughly three orders of magnitude smaller than that measured before any recalibration effort). Furthermore, we note that the constant term shown in the figure ( $c_k$ ) is not the same as the additive bias which is removed from our galaxy sample. In practice, the additive component of our ellipticities was computed directly (using a weighted mean) from the sample of galaxies per tomographic bin and hemisphere:  $\hat{c}_k = \langle \varepsilon_k \rangle_w$ . The  $\hat{c}$ -terms that were subtracted from our recalibrated shape catalogues prior to computation of correlation functions are provided in Table 2. In contrast, the figure presents the

**Table 2.** Subtracted additive shear per ellipticity component, tomographic bin, and hemisphere.

Hemi	Tomographic Bin	$\hat{c}_1$ ( $\times 10^{-4}$ )	$\hat{c}_2$ ( $\times 10^{-4}$ )
North	1	$3.372 \pm 1.528$	$7.941 \pm 1.442$
	2	$8.852 \pm 1.642$	$5.594 \pm 1.531$
	3	$4.523 \pm 1.747$	$4.533 \pm 1.777$
	4	$4.722 \pm 1.713$	$5.368 \pm 1.665$
	5	$6.658 \pm 1.887$	$5.532 \pm 1.890$
	6	$4.224 \pm 2.252$	$10.26 \pm 2.400$
South	1	$-3.398 \pm 1.626$	$-8.002 \pm 1.572$
	2	$-9.536 \pm 1.519$	$-6.026 \pm 1.590$
	3	$-4.755 \pm 1.835$	$-4.766 \pm 1.731$
	4	$-4.532 \pm 1.653$	$-5.152 \pm 1.594$
	5	$-6.117 \pm 1.910$	$-5.082 \pm 1.834$
	6	$-3.717 \pm 2.151$	$-9.027 \pm 2.282$

constant component of our linear regression (Eq (24)) to all data per tomographic bin after removal of the additive bias. As such it is expected that the  $c_k$ -terms are small, but not necessarily zero.

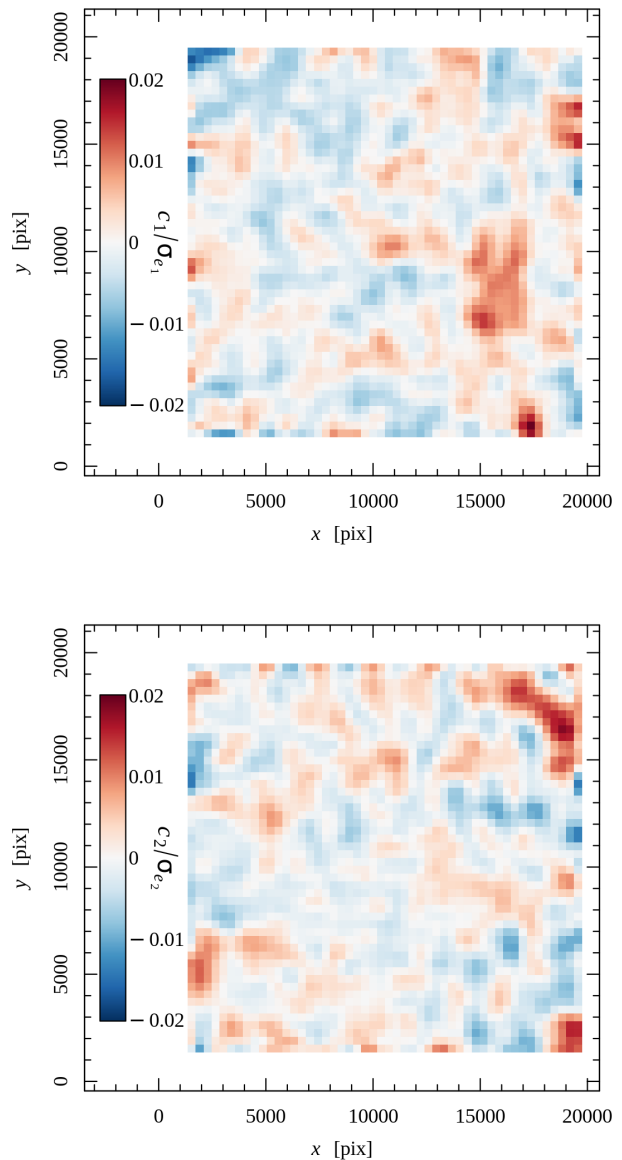
As a further quality test, we computed the non-tomographic 2D  $c$ -terms in camera coordinates per ellipticity component, which is shown in Fig. 5. The 2D  $c$ -terms were calculated as a weighted mean per bin (using both shape- and gold-weights), with the resulting map smoothed with a Gaussian with one-bin standard deviation (approximately 1/2), and the computed  $c$ -term scaled by the (also smoothed) measured ellipticity dispersion in each bin. The residual  $c$ -term can be seen to have amplitudes  $|c_i/\sigma_{\varepsilon_i}| \leq 1\%$ , indicating that these 2D patterns are negligible in the context of our cosmic shear measurements. The primary structure that is visible in  $c_1$  was identified in Hildebrandt et al. (2020) as being due to an electronic effect that introduces an anomalous signal in the read-out direction of one detector on OmegaCAM. However, as shown in Asgari et al. (2021), the amplitude of this signal ( $\lesssim 1\%$  of the ellipticity dispersion) is negligible in the context of our cosmic shear measurements.

#### 4. Null tests

In this section, we outline the various null tests that were completed prior to the unblinding of the cosmological analysis of KiDS-Legacy; the reader interested in cosmological parameter estimates only may wish to skip to Section 5. We performed three primary tests: an analysis of the Paulin-Henriksson et al. (2008) systematics model, which quantifies the impact of PSF modelling residuals on correlation functions; an analysis of Bacon et al. (2003) statistics, which test specifically for contamination of the galaxy shape measurements by the PSF; and an analysis of  $B$  modes, which tests for a systematic signal that cannot be generated by cosmic shear.

##### 4.1. Paulin-Henriksson et al. (2008) model

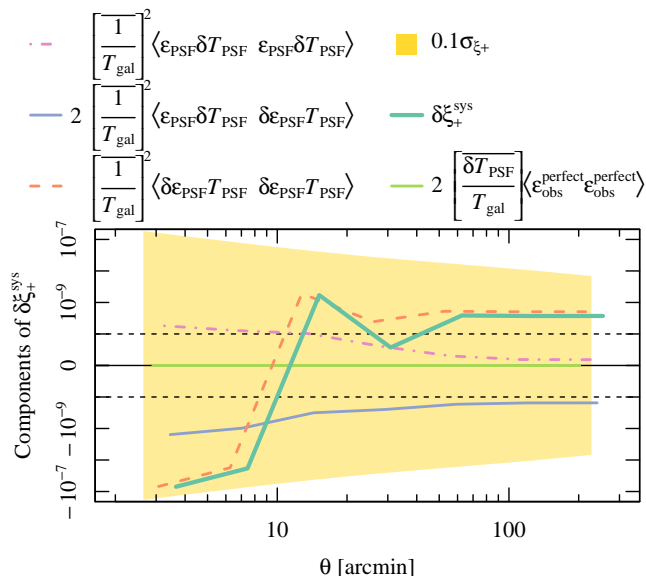
To validate the fidelity of our shape measurement catalogue, we first compute the Paulin-Henriksson et al. (2008) systematics model for KiDS-Legacy, which aims to quantify the impact of PSF modelling residuals on cosmic shear correlation functions. This systematics model, in the context of the shear two-point



**Fig. 5.** Average of each ellipticity component in bins of the focal plane, for sources in all tomographic bins, relative to the (weighted) dispersion of ellipticities in the same bin. The weighted ellipticity dispersion per component over the bins in the focal plane are  $0.287 \pm 0.001$  and  $0.288 \pm 0.001$ , indicating that the observed pattern is not driven by variable ellipticity dispersion across the focal plane. The figures still indicate the presence of the two-dimensional pattern reported in Figure 2 of Hildebrandt et al. (2020).

correlation function, reduces to the following:

$$\begin{aligned}
\langle \varepsilon_{\text{obs}} \varepsilon_{\text{obs}} \rangle &\simeq \left( 1 + 2 \left[ \frac{\delta T_{\text{PSF}}}{T_{\text{gal}}} \right] \right) \langle \varepsilon_{\text{obs}}^{\text{perfect}} \varepsilon_{\text{obs}}^{\text{perfect}} \rangle \\
&+ \left[ \frac{1}{T_{\text{gal}}} \right]^2 \langle (\varepsilon_{\text{PSF}} \delta T_{\text{PSF}}) (\varepsilon_{\text{PSF}} \delta T_{\text{PSF}}) \rangle \\
&+ 2 \left[ \frac{1}{T_{\text{gal}}} \right] \langle (\varepsilon_{\text{PSF}} \delta T_{\text{PSF}}) (\delta \varepsilon_{\text{PSF}} T_{\text{PSF}}) \rangle \\
&+ \left[ \frac{1}{T_{\text{gal}}} \right]^2 \langle (\delta \varepsilon_{\text{PSF}} T_{\text{PSF}}) (\delta \varepsilon_{\text{PSF}} T_{\text{PSF}}) \rangle,
\end{aligned} \tag{25}$$



**Fig. 6.** Contributions to the additive systematic  $\delta\xi_+^{\text{sys}}(\theta)$  from the Paulin-Henriksson et al. (2008) systematics model. The various lines show the four components, with their summation in thick blue-green. The yellow band depicts 10% of the cosmic shear uncertainty on  $\xi_+$ . The figure is linear between the black dashed lines, and logarithmic outside. This demonstration is for tomographic bin six, but all bins show quantitatively similar results.

where we only keep terms that correlate to first order. It should be noted that we use a shorthand notation in Eqs. (25), (26), and (27) as in Section 3.3 of Giblin et al. (2021), so that  $\langle ab \rangle$  denotes  $\xi_{\pm}$ . Here,  $\epsilon_{\text{obs}}$  is the observed ellipticity of a source,  $\epsilon_{\text{obs}}^{\text{perfect}}$  is the noiseless and unbiased measurement of source ellipticity (which we approximate as being equal to the modelled shear:  $\epsilon_{\text{obs}}^{\text{perfect}} = \gamma_{\text{theory}}$ ), and  $T$  is the  $R^2$  object size estimate.

Figure 6 shows the various contributions to the additive systematic,  $\delta\xi_+^{\text{sys}}(\theta)$ , from the Paulin-Henriksson et al. (2008) systematics model. The four terms from Eq. (25), shown in various colours, cause  $\langle\epsilon_{\text{obs}}\epsilon_{\text{obs}}\rangle$  to deviate from  $\langle\epsilon_{\text{perfect}}\epsilon_{\text{perfect}}\rangle$ . The total systematic  $\delta\xi_+^{\text{sys}}$ :

$$\delta\xi_+^{\text{sys}}(\theta) := \langle\epsilon_{\text{obs}}\epsilon_{\text{obs}}\rangle - \left\langle\epsilon_{\text{obs}}^{\text{perfect}}\epsilon_{\text{obs}}^{\text{perfect}}\right\rangle, \quad (26)$$

is given by the summation of these four terms<sup>6</sup> and can be compared to the yellow band which encloses 10% of the uncertainty on the cosmic shear correlation function  $\xi_+$ , from the ONECOVARIANCE (Sect. 2.6). The figure presents the analysis of our sixth tomographic bin (quantitatively similar results are obtained for all other bins), and demonstrates that errors in our PSF modelling are sufficiently small (consistently less than 10% of our correlation function uncertainty) that there is no significant contamination of our two-point statistics from this source of bias.

<sup>6</sup> The terms in Eq. (25) are related to the often-used and so-called ‘ $\rho$ ’-statistics, as recommended by Rowe (2010) and Jarvis et al. (2016). In contrast to the total systematic term  $\delta\xi_+^{\text{sys}}$  (Eq. 26), the systematic requirements on the ‘ $\rho$ ’-statistics, relative to the statistical noise on the measured cosmic shear signal, are unspecified. We therefore chose not to explore the ‘ $\rho$ ’-statistics directly.

As such, we conclude that PSF modelling errors are negligible for the analysis of cosmic shear in KiDS-Legacy.

#### 4.2. Bacon et al. (2003) statistics

We also used the method from Bacon et al. (2003) to infer the degree of PSF leakage into the two-point correlation function measurement using

$$\xi_{\pm}^{\text{sys}} = \frac{\langle\epsilon_{\text{obs}}^{\text{PSF}}\epsilon_{\text{obs}}^{\text{PSF}}\rangle^2}{\langle\epsilon_{\text{PSF}}\epsilon_{\text{PSF}}\rangle}. \quad (27)$$

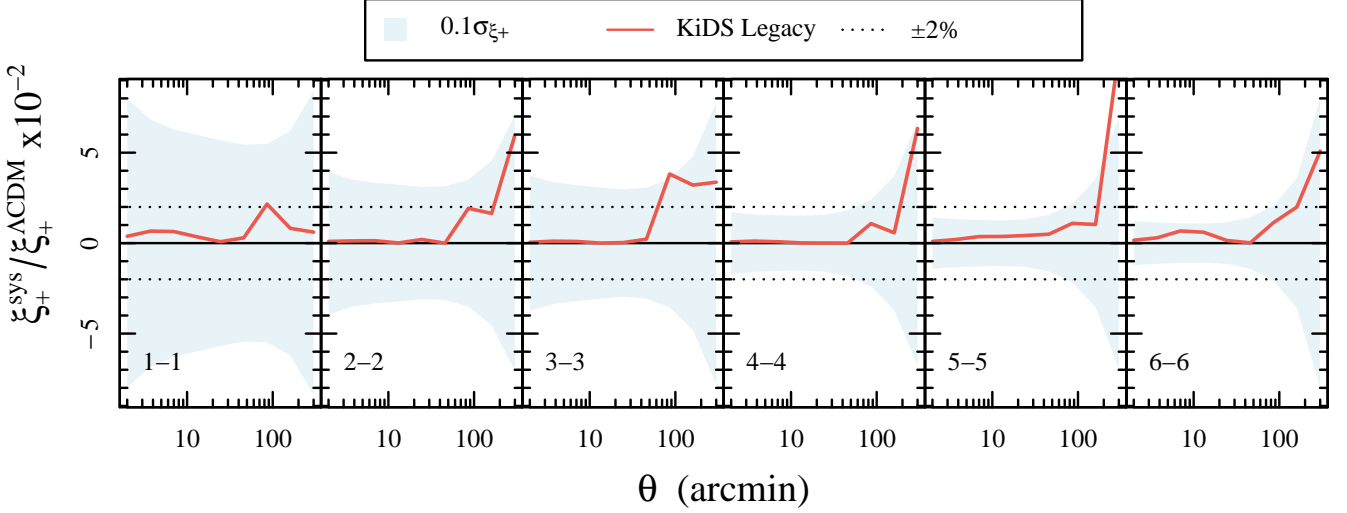
In Fig. 7, we show the value of  $\xi_{\pm}^{\text{sys}}$  in each tomographic bin autocorrelation, as a fraction of the theory prediction of the cosmic shear signal from our best-fit  $\xi_{\pm}^{\text{CDM}}$  cosmology ( $\xi_{\pm}^{\Lambda\text{CDM}}$ ). Results from the cross-correlation bins are quantitatively similar to the autocorrelation. As a reference, we also show 10% of the  $1\sigma$  uncertainty on the cosmic shear signal as reported by the ONECOVARIANCE (Sect. 2.6). The measured fractional PSF contamination is less than the benchmark 10% of the cosmic shear correlation function uncertainty for all but five data points (out of 189 across all tomographic bin combinations). As such, we conclude that the PSF contamination of cosmic shear correlation functions is negligible in KiDS-Legacy, as expected from our PSF leakage analysis (Sect. 3.5).

#### 4.3. B modes

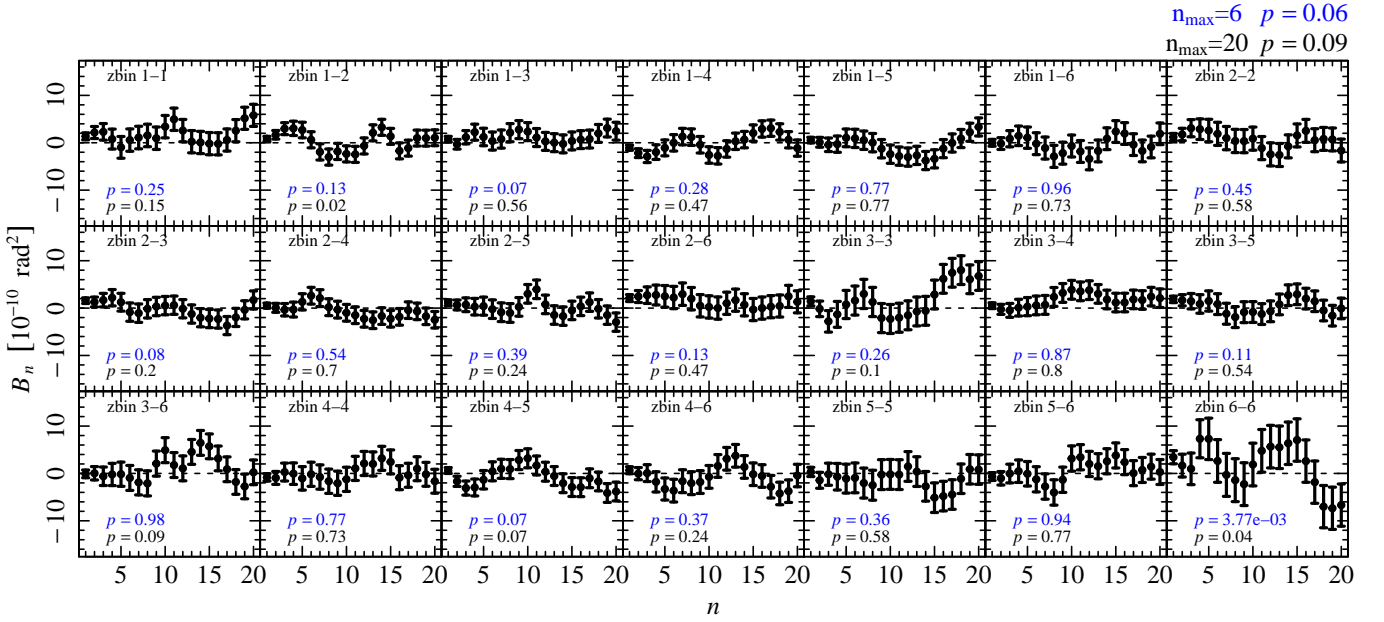
Our third null test examined the significance of *B*-mode signals in the data vector used for cosmological analyses. This test was based on the understanding that, at the sensitivity available to stage-III cosmological imaging surveys, *B*-mode signals of cosmological and astrophysical origin ought to be negligible. As such, any significant detection of *B* modes implies a contamination of the data vector by systematic effects, which may similarly (but undiagnosably) also impact the *E*-modes used for cosmological inference.

We used a tomographic analysis, equivalent to that of the fiducial cosmological analysis, to derive the significance of *B* modes (estimated via the COSEBIs  $B_n$ ) in our data vector. In the early stages of our blinded cosmological analysis, it became clear that our initial KiDS-Legacy sample failed this null test, at high significance. This led to a re-examination of the cosmological sample, which in turn led to a redefinition of the survey footprint to conservatively exclude 46.6 square-degrees with increased astrometric scatter between exposures in our lensing imaging (see Appendix E).

The *B*-mode signal is presented in Fig. 8. The  $p$ -values for the measured *B* modes across all tomographic bin combinations are 0.04 in the case of 6  $B_n$  modes, and 0.09 in the case of 20 modes, and thus both pass our required threshold of  $p > 0.01$  (adopted from previous work within both the KiDS and DES teams). Looking at the individual tomographic bins, only the sixth bin autocorrelation, with 6  $B_n$  modes, has a  $p$ -value below 0.01. However, with 21 tests (one per bin combination) it is expected that one will occasionally find  $p < 0.01$  even with correlated data drawn from the null hypothesis; indeed, this is essentially the calculation that is performed in the computation of the ‘all bins’  $p$ -value, which passes.



**Fig. 7.** Ratio of PSF contamination in the cosmic shear correlation function,  $\xi_+^{\text{sys}}$ , to the predicted best-fit cosmic shear correlation function,  $\xi_+^{\Lambda\text{CDM}}$ . The statistic is shown relative to 10% of the  $1\sigma$  uncertainty on our fiducial  $\xi_+$  correlation functions (blue). We show only tomographic bin auto-correlations here, as the cross-correlation signals are extremely similar to the auto-correlations (i.e. combinations  $i - j$ ;  $\forall j > i$  are quantitatively similar to the  $i - i$  signal).



**Fig. 8.** COSEBIs  $B$  modes measured in KiDS-Legacy. Each panel is annotated with the  $p$ -value of the  $B$ -mode signals for twenty modes (black) and our fiducial six modes (blue). The total  $p$ -value for the full data vector, also for six and twenty modes, is annotated in the upper right corner of the figure. The  $B$ -mode signals are highly correlated across modes within a tomographic bin, so readers are cautioned against so-called ‘ $\chi$ -by-eye’.

## 5. Results

In this section, we present the cosmological results from KiDS-Legacy split between our fiducial results (Sect. 5.1), and variations of analysis and modelling choices (Sect. 5.2). Furthermore, we refer the reader to our companion paper (Stölzner et al. 2025) for detailed analyses of the internal consistency of KiDS-Legacy, and for analysis of the external consistency between KiDS-Legacy and datasets from DES, DESI, Pantheon+ (Brout et al. 2022), and *Planck*.

### 5.1. Fiducial constraints

The fiducial analysis of KiDS-Legacy used two statistics ( $E_n$ ,  $C_E$ ) measured over  $2'0 \leq \theta \leq 300'$ , an NLA- $M$  intrinsic alignment model, correlated  $\delta z$  prior marginalisation, emulated COSMOPower  $P(k)$  trained on CAMB and HMCODE2020, covariances constructed with the ONECOVARIANCE and augmented with the uncertainty in the shear calibration, and sampling performed by NAUTILUS (Lange 2023) within COSMOSIS, using a Gaussian likelihood. We chose the NAUTILUS sampler due to its superior efficiency, robustness of uncertainties, and accurate evidence estimation (see e.g. Lemos et al. 2023; Dark Energy Survey and Kilo-Degree Survey Collaboration et al.

**Table 3.** Fiducial prior ranges used in the analysis of KiDS-Legacy.

Parameter	Prior range
$S_8$	[0.5, 1.0]
$\Omega_c h^2$	[0.051, 0.255]
$\Omega_b h^2$	[0.019, 0.026]
$\Omega_K$	0.0
$n_s$	[0.84, 1.10]
$h$	[0.64, 0.82]
$\log_{10}(T_{\text{AGN}}/K)$	[7.3, 8.3]
$\sum m_\nu$	0.06eV
$w_0$	-1.0
$w_a$	0.0
$\delta z^{(i)}$	$\mathcal{N}(\mu_z, \mathbf{C}_{\delta z})$
$A_{\text{IA}}, b_{\text{IA}}$	$\mathcal{N}(\mu_{\text{IA,Ab}}, \mathbf{C}_{\text{IA,Ab}})$
$\log_{10}(\langle M \rangle^{(i)})$	$\mathcal{N}(\mu_{\text{IA,M}}, \mathbf{C}_{\text{IA,M}})$

**Notes.** Parameters in the first block are inputs to the  $P(k)$  emulator, and all use flat priors denoted by their lower and upper limits [lower, upper]. We sample over five cosmological parameters:  $S_8$ ,  $\Omega_c$  and  $\Omega_b$  (the density parameters for cold dark matter and baryonic matter, respectively);  $h$  (the dimensionless Hubble parameter); and  $n_s$  (the spectral index of the primordial power spectrum). Additional inputs to the power spectrum emulator include four fixed cosmological parameters:  $\Omega_K$  (the energy density parameter associated with spatial curvature);  $w_0$  and  $w_a$  (describing the equation of state of dark energy and its redshift evolution); and  $\sum m_\nu$  (the total mass of all families of neutrinos). We also sample over the baryon feedback parameter from HMCode2020  $\log_{10}(T_{\text{AGN}}/K)$ . Parameters in the second block are implemented within additional COSMOSIS modules, and are used in the intrinsic alignment model ( $A_{\text{IA}}, b_{\text{IA}}$ , and  $\log_{10}(\langle M \rangle^{(i)})$ ) or our redshift distribution bias mitigation ( $\delta z^{(i)}$ ). These sets of parameters are specified with multivariate Gaussian priors denoted by  $\mathcal{N}(\mu_x, \mathbf{C}_x)$ , where  $\mu$  is the vector of prior means and  $\mathbf{C}$  is the related covariance matrix. Values of  $\mu_{\text{IA},x}$  are given in Appendix B. Values of  $\mu_z$ ,  $\text{diag}(\mathbf{C}_{\delta z})$  are given in Table 1, and the correlation matrix of  $\mathbf{C}_{\delta z}$  is shown in Fig. 3.

2023). Priors for the various cosmological and nuisance parameters are given in Table 3. Correlated priors used for our redshift distribution bias parameters are given in Sect. 3.4, and for the IA model are given in Appendix B.

Figure 9 presents the fiducial marginal constraints in  $\sigma_8$ ,  $S_8$ , and  $\Omega_m$ . Marginal constraints over both  $S_8$  and  $\Sigma_8 = \sigma_8(\Omega_m/0.3)^\alpha$  (with the exponent fitted to the  $\Omega_m$  vs.  $\sigma_8$  contour) are presented in Table 4, for three estimation types: marginal mean and standard deviation, marginal mode and 1D 68% highest posterior density interval (HPDI), and maximum a posteriori (MAP) and 68% projected joint-distribution highest posterior density interval (PJ-HPD). Dark Energy Survey and Kilo-Degree Survey Collaboration et al. (2023, Sect. 2.6) present a discussion of the benefits and detriments of each metric. We also report the  $\chi^2$  of our cosmological model evaluated at the MAP, and the model probability to exceed (PTE). The MAP was evaluated using an adaptive Nelder-Mead (downhill simplex) algorithm, as described in Joachimi et al. (2021).

Data vectors for the fiducial  $E_n$  and  $C_E$  are given in Fig. 10, alongside translated posterior distributions (TPDs; see Köhlinger et al. 2019, for the first application of these consistency metrics to cosmic shear datasets) drawn from the fiducial posterior computed for each statistic.

Our fiducial cosmological results, from COSEBIs, estimate a value of  $S_8 = 0.813 \pm 0.018$ , evaluated using the marginal mean and standard deviation. This is the tightest constraint made with our COSEBIs measurements, although all constraints are rela-

tively consistent in their constraining power: the marginal mode and HPDI estimate is  $S_8 = 0.815_{-0.021}^{+0.016}$ , and the maximum a posteriori and PJ-HPD estimate is  $S_8 = 0.811_{-0.015}^{+0.022}$ . The consistency between the measurements is a reflection of the Gaussianity of the marginal posterior, and of the  $N$ -dimensional posterior (in the case of the PJ-HPD). This is despite the necessary smoothing of the posterior distribution that is required of the HPDI computation, which is performed using a Gaussian kernel with standard deviation defined using the method of Sheather & Jones (1991). Testing of the HPDI method indicates that, for our posterior distributions in  $S_8$ , the chosen smoothing kernel is unlikely to introduce additional spurious uncertainty to our cosmological parameter constraints. As such, we expect that the constraints returned from our various summarisation methods to be robust. In the following, all values and uncertainties correspond to the marginal mode and HPDI estimate, unless otherwise specified.

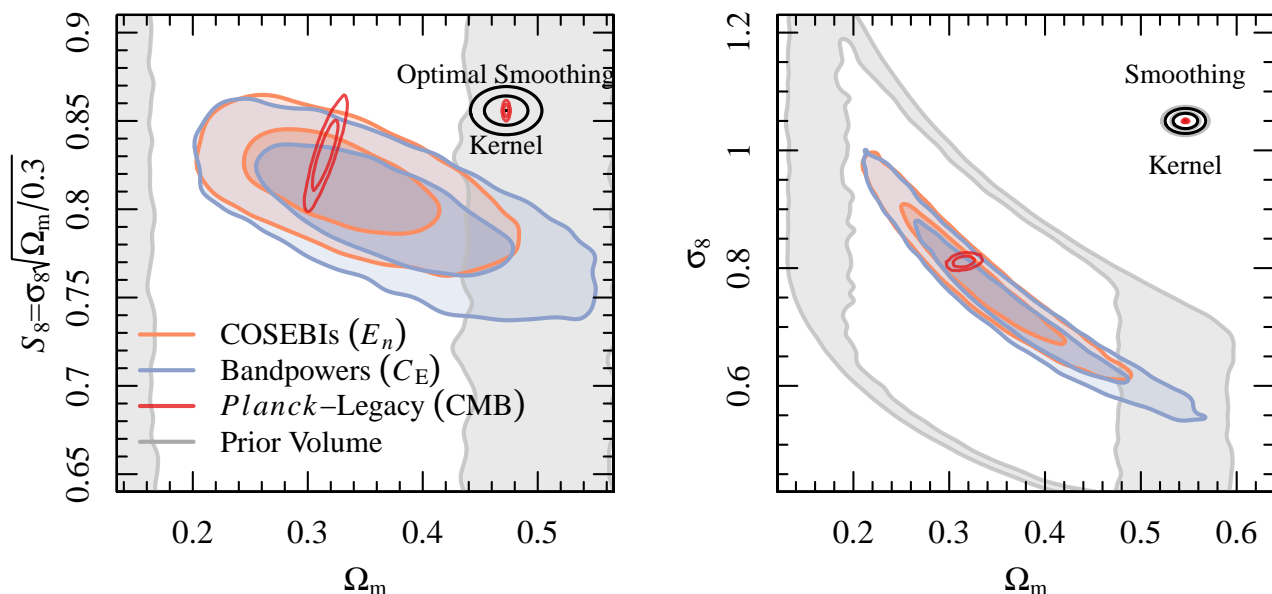
The above discussion about the stability of our summarisation statistics is contrasted by the observed variance between our  $S_8$  constraints from different statistics. The variance in  $S_8$  between our statistics is seemingly large, with our  $C_E$  and  $E_n$  differing by  $0.55\sigma$ . However, examination of the 2D posterior distribution in  $\Omega_m$  and  $S_8$  (Fig. 9) indicates that the optimal degeneracy direction for KiDS-Legacy is somewhat different than the assumed  $\alpha = 0.5$  parametrised by  $S_8$ . This is expected, given the different scales probed by each of our statistics, and has the effect of exacerbating apparent differences between our statistics when marginalising on the  $S_8$  parameter (as translation along the degeneracy can give the false impression of internal inconsistency between final parameter constraints).

When fitting for the optimal  $\alpha$  under the more general  $\Sigma_8 = \sigma_8(\Omega_m/0.3)^\alpha$ , we found the best fitting  $\alpha$  is 0.58 for  $E_n$  and 0.60 for  $C_E$ . These significant deviations from  $\alpha = 0.5$  can cause offsets between derived values of  $S_8$ , as is seen between our  $C_E$  and  $E_n$  constraints. Additionally, the change in degeneracy direction has a significant impact on the apparent constraining power of each of our statistics. For  $C_E$  in particular, we find a significant  $\sim 40\%$  reduction in the relative marginal uncertainty between our  $S_8$  and  $\Sigma_8$  parameters. As such, we opt to preferentially present results in the  $\Sigma_8$  parametrisation in this manuscript, especially when quantifying the impact of analysis choices on recovered constraints. We note, however, that the different  $\alpha$  values between the different statistics mean that  $\Sigma_8$  values recovered using each statistic are not directly comparable to one another.

## 5.2. Analysis variations

In addition to the fiducial constraints, we provide a range of additional cosmological constraints computed with various permutations of our modelling and optimisation choices. Specifically, we present: the impact of the emulation of the matter power spectrum (Sect. 5.2.1), modelling of the signal with only the Gaussian component of the covariance (Sect. 5.2.2), iterative computation of covariances (which require a pre-defined cosmology; Sect. 5.2.3), the impact of different choices of intrinsic alignment models (Sect. 5.2.4), the role of observational systematics (Sect. 5.2.5), the impact of different redshift calibrations (Sect. 5.2.6), the influence of different scale cuts (Sect. 5.2.7), and the impact of individual tomographic bins on our estimated cosmological parameters (Sect. 5.2.8).

Figure 11 presents a whisker diagram of all the various  $\Sigma_8$  (that is, for the optimal  $\Omega_m, \sigma_8$ -degeneracy) results presented in this work. The figure includes our fiducial results in each statistic, with constraints computed using the three methods previ-



**Fig. 9.** Fiducial  $\Omega_m, S_8$  (left), and  $\Omega_m, \sigma_8$  (right) constraints for our fiducial two-point statistics ( $E_n$  and  $C_E$ , orange and purple respectively), compared to the CMB results from *Planck*-Legacy (red). The grey contours outline the marginal distribution of our a priori volume, with (inner)  $1\sigma$  and (outer)  $2\sigma$  boundary levels.

**Table 4.** Constraints on  $S_8$  ( $\alpha = 0.50$ ) and  $\Sigma_8$  ( $\alpha$  free).

Setup	Statistic	$\alpha$	$\chi^2$	dof	PTE	$N_{\text{samp}}^{\text{PJ-HPD}}$	Marginal Mode+HPDI	Max. Apost. +PJ-HPD	Marginal Mean + CI
$S_8$ Fiducial	$E_n$	0.50	127.8	120.5	0.307	18	$0.815^{+0.016}_{-0.021}$	$0.811^{+0.022}_{-0.015}$	$0.813^{+0.018}_{-0.018}$
$S_8$ Fiducial	$C_E$	0.50	151.0	162.5	0.731	6	$0.799^{+0.022}_{-0.025}$	$0.804^{+0.016}_{-0.031}$	$0.797^{+0.023}_{-0.024}$
$\Sigma_8$ Fiducial	$E_n$	0.58	127.8	120.5	0.307	23	$0.821^{+0.014}_{-0.016}$	$0.817^{+0.029}_{-0.006}$	$0.820^{+0.015}_{-0.015}$
$\Sigma_8$ Fiducial	$C_E$	0.60	151.0	162.5	0.731	19	$0.813^{+0.018}_{-0.016}$	$0.822^{+0.017}_{-0.020}$	$0.813^{+0.017}_{-0.017}$

**Notes.**  $\alpha$  in this table refers to the exponent in the equation  $\Sigma_8 = \sigma_8 (\Omega_m/0.3)^\alpha$ . When  $\alpha = 0.5$  (upper section) reported constraints are for  $S_8$ . Otherwise, constraints are for the optimal degeneracy direction for each statistic, thereby showing the maximal constraining power. We report the  $\chi^2$  of the best-fit (i.e. maximum a posteriori) model for each statistic, and the number of effective degrees of freedom (dof, computed as  $N_{\text{data}} - N_{\text{param}}^{\text{eff}}$  where  $N_{\text{param}}^{\text{eff}}$  is estimated as described in Appendix B of Stözlner et al. 2025). We compute the best-fit probability-to-exceed (PTE), which indicates the fraction of possible data vectors, drawn from a generative model described by our best fit, that would be at least as extreme as our observed data vector. We provide three constraints over our primary cosmological parameters per setup: the marginal mode and marginal highest posterior density interval (HPDI), the marginal mean and standard deviation, and the maximum a posteriori point and projected joint highest posterior density interval (PJ-HPD). The latter statistic can be noisy in the limit of few samples  $N_{\text{samp}}^{\text{PJ-HPD}}$  residing within the PJ-HPD, and so we also provide the number of samples in the PJ-HPD region as a guide.

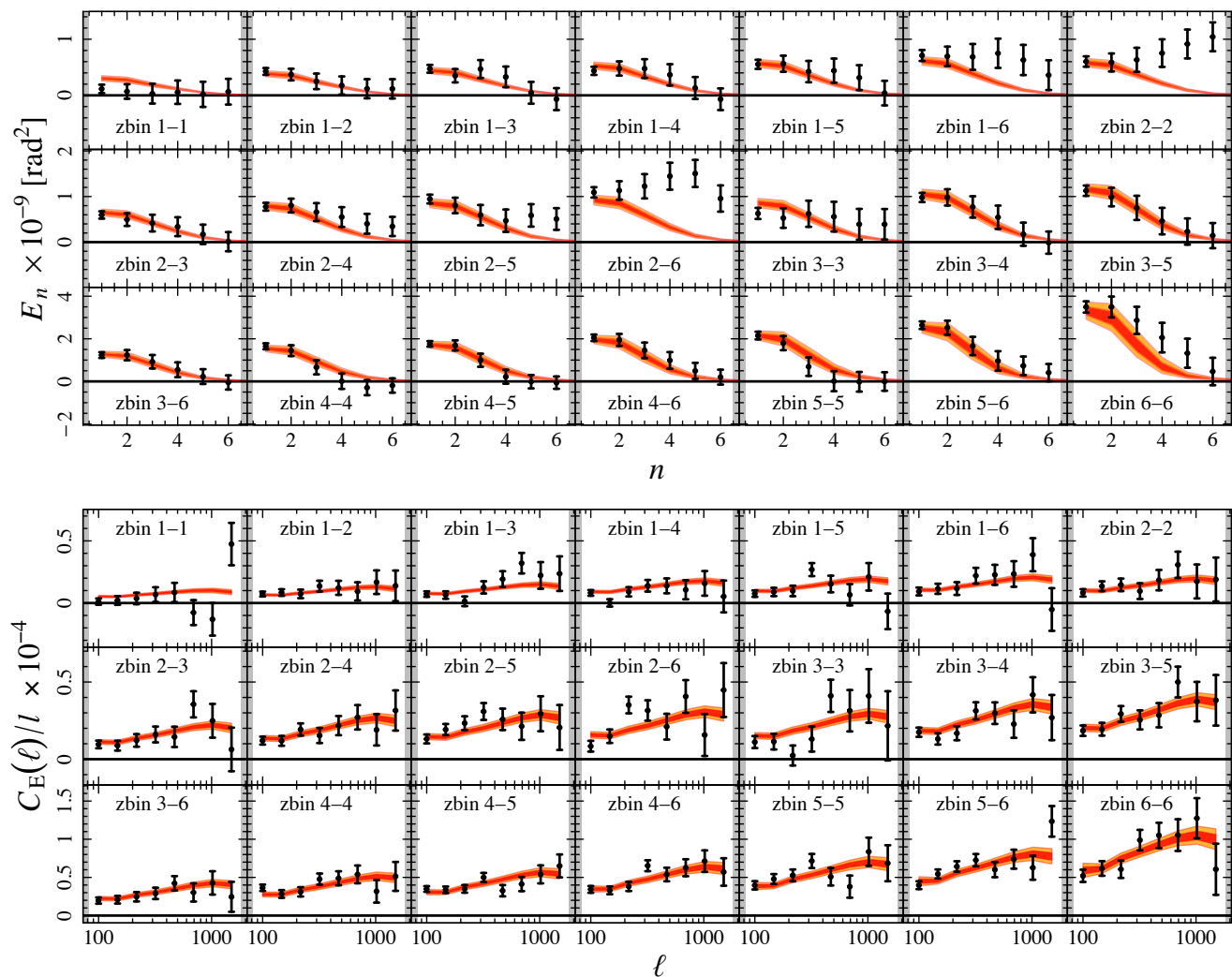
ously described in Table 4. All discussions below refer to our fiducial metric, the marginal mode and HPDI summary, unless stated otherwise. Additionally, we have annotated the figure with a measure of the difference between our various results, computed using the metric  $(\Sigma_8^{\text{fid}} - \Sigma_8^{\text{var}})/\sigma_{\Sigma_8^{\text{var}}}$ . Differences are computed between the  $\Sigma_8$  values of each setup and the fiducial. We also annotate the fiducial constraints with their Hellinger distance (see e.g. Appendix G.1 of Heymans et al. 2021) to the constraints from the public, fiducial *Planck*-Legacy analysis (Planck Collaboration et al. 2020). Finally, we present a similar figure for  $S_8$  in Appendix G, and tables of all constraints are presented in Appendix H.

### 5.2.1. Modelling $P(k)$ with CAMB

We tested the impact of the CosmoPower emulator by re-running our fiducial analysis directly with CAMB and comparing the cosmological constraints between these runs in Fig. 11. We found that, for all statistics, the CAMB posteriors were consistent with those estimated with the emulated power spectrum, showing maximal changes in the marginal value of  $\Sigma_8$  of  $0.03\sigma$ . We consider this level of change to be consistent with chain-to-chain variations, and therefore negligible.

### 5.2.2. Gaussian covariance

We tested the impact of our covariance modelling on the recovered value of  $\Sigma_8$  in Fig. 11, by performing the fiducial measurement with only the Gaussian component of the covariance acti-



**Fig. 10.** Data vector and TPDs for the fiducial analysis of  $E_n$  (upper panel) and  $C_E$  (lower panel). TPDs are shown as polygons spanning the 68<sup>th</sup> (red), and 95<sup>th</sup> (orange) percentiles of the posterior models. The  $E_n$  signals are highly correlated across modes within a tomographic bin, so readers are cautioned against so-called ‘ $\chi$ -by-eye’. The bin 2 autocorrelation signal, for example, is consistent with the best fit model (PTE= 0.09), despite the apparent divergence of the data from the model. The overall PTEs for the full tomographic dataset in each statistic are provided in Table 4.

vated in the `ONECOVARIANCE` code. We found a reduction in the uncertainty on  $\Sigma_8$  of between 12% ( $C_E$ ) and 19% ( $E_n$ ), and no change in the value of  $\Sigma_8$  with respect to the fiducial ( $\Delta\Sigma_8 = 0.02\sigma$ ). This is an indication that, while shape noise is still sufficient to shroud information on non-linear scales, KiDS-Legacy has a non-negligible contribution to its constraining power from these non-Gaussian scales.

### 5.2.3. Iterative covariance

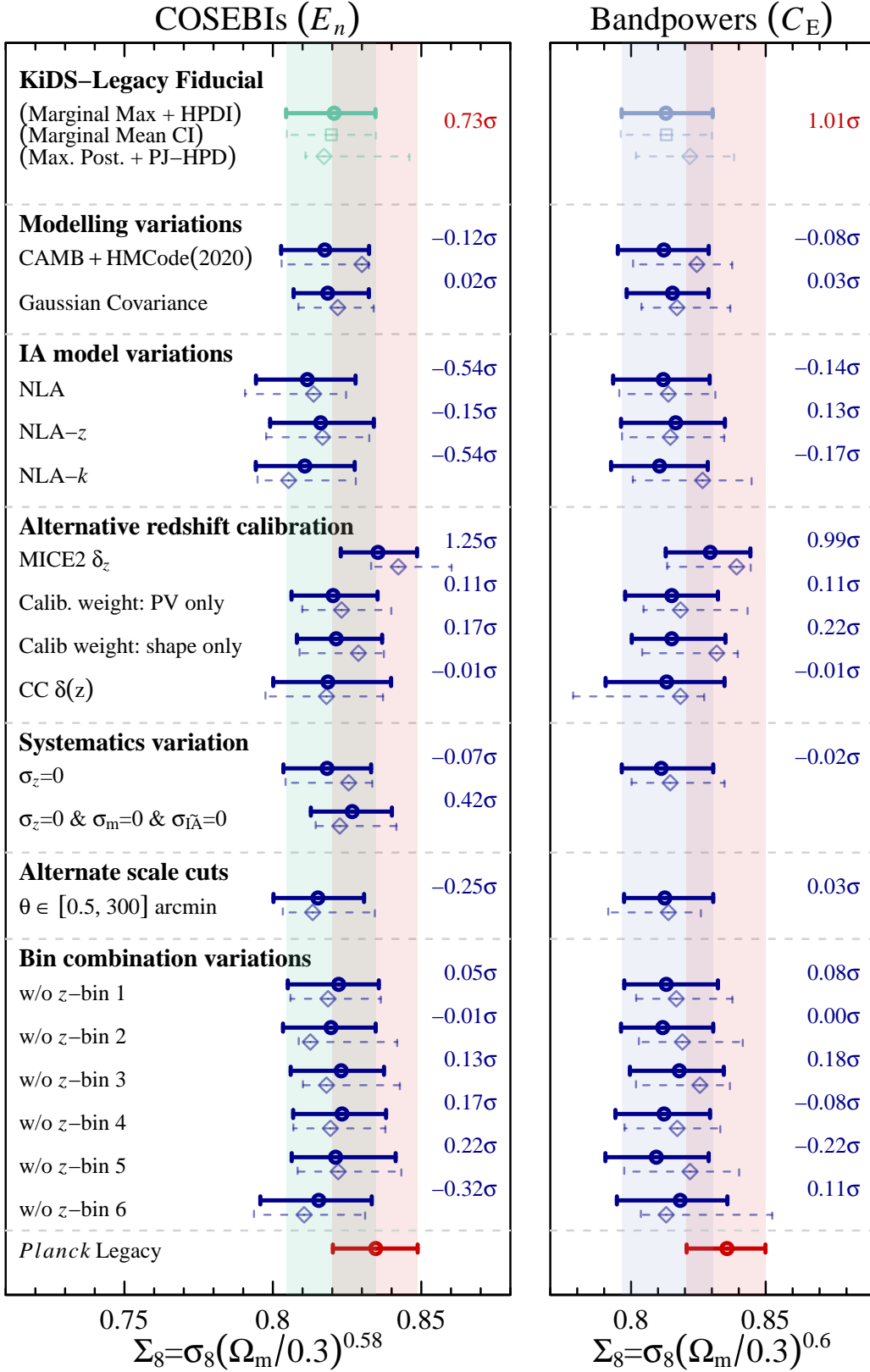
The covariance used in our cosmological inference modelling assumes a fixed cosmology. If this fixed cosmology differs from the best-fit cosmology preferred by the dataset, this can add a source of inconsistency in our fiducial results. Therefore, as with previous KiDS analyses (see e.g. van Uitert et al. 2018), we verified the fidelity of our cosmological inference by recomputing the fiducial sample covariances at the maximum a posteriori cosmology for each chain, and rerun our cosmological inference. This iterative evaluation of the covariance and cosmological parameters do not need to converge after a single iteration, and so we ran this process for many ( $\lesssim 10$ ) iterations. We found that the cosmo-

logical constraints (and constraining power) converged after two iterations, even for our most complex model choices. This is in good agreement with previous investigations into the impact of varying cosmological parameters on covariances in the context of stage-III surveys (Reischke et al. 2017; Kodwani et al. 2019).

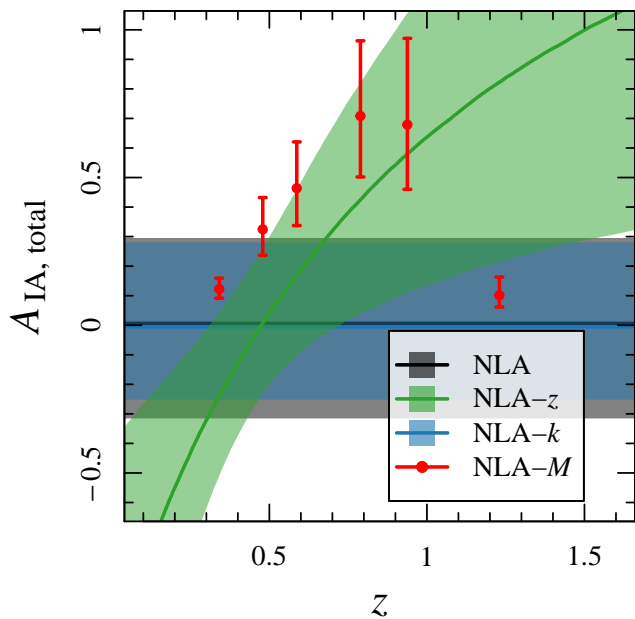
### 5.2.4. Intrinsic alignment modelling

We tested the cosmological parameter constraints recovered when modelling intrinsic alignments with our four different models. The resulting cosmological parameters are presented in Fig. 11.

First, we see that the impact of changing our IA models on the marginal constraints of  $S_8$  and  $\Sigma_8$  are mild. For our  $C_E$  statistic, we found maximal differences of  $0.17\sigma$  across all model variations. Also,  $E_n$  values are more sensitive to our chosen IA model, showing maximal differences of  $0.54\sigma$  (for the NLA and NLA- $k$  models and  $\Sigma_8$ ). Second, we see that changing the IA model has a minor impact on the constraining power of our survey, with systematic changes in constraining power being most evident for the NLA- $z$  model: we see a consistent reduction in



**Fig. 11.** Whisker diagram of  $\Sigma_8$  cosmological results presented in this work. The best-fitting  $\alpha$  used for each statistic is annotated on the x-axis. Each panel shows the compilation of results for one of our two-point statistics, as annotated. In each panel, we show the fiducial constraints in the relevant statistic with coloured points (colour-coded by the statistic), and highlight the fiducial marginal HPDI with a band of the equivalent colour. We also highlight the marginal HPDI in  $\Sigma_8$  from *Planck*-Legacy (shaded red). All variations to the analysis are annotated, with the difference relative to the fiducial result quantified using the metric  $(\Sigma_8^{\text{fid}} - \Sigma_8^{\text{var}})/\sigma_{\Sigma_8^{\text{var}}}$  (listed in blue). The Hellinger tension between the fiducial analyses and *Planck* is annotated in red.



**Fig. 12.** Constraints on the total intrinsic alignment amplitude from the various intrinsic alignment models, per tomographic bin (where appropriate), computed using the  $E_n$  posteriors. Bands and error bars represent the  $1\sigma$  posterior constraint on each model. Our NLA- $M$  model (red) is shown per tomographic bin, with the point positioned at the mean photo- $z$  of the bin.

constraining power (compared to the fiducial NLA- $M$  model) of 12 – 13% for  $\Sigma_8$  across all statistics. For our NLA and NLA- $k$  models, the reduction in constraining power compared to the fiducial is slightly smaller, at  $\sim 5\%$ . This reduced constraining power (despite the fact that e.g. the NLA model has fewer parameters) is attributed to the informative prior applied in our NLA- $M$  model (see Sect. 2.4.4 and Appendix B). Finally, we also note mild changes to our recovered goodness-of-fit under the various IA models. In all cases, the  $\chi^2$  of the modified IA model is consistent with the fiducial (changes in PTE at the order of 1 – 5%).

The stability of our cosmological constraints under variation of the IA modelling provides confidence that any possible misspecification of the IA model is unlikely to introduce bias in our cosmological parameter constraints. Furthermore, we can compare the implied amplitude of IA under each of our models by examining the  $A_{IA, total}$  values (see Eqs. 8, 10, 12). Figure 12 shows the value of  $A_{IA, total}$  inferred from our  $E_n$  analysis, as a function of redshift. All statistics show quantitatively similar behaviour. In Appendix G, we present the posterior distributions in all parameters (Figs. G.2 and G.3) for each IA model.

For our NLA and NLA- $k$  models, there is no explicit dependence on tomographic bin number (or, more accurately, redshift), and thus the IA amplitude is necessarily constant across the redshift axis for these models. Both of these models are consistent with IA amplitudes of zero. A change in IA amplitude with redshift is nominally expected, though, as the combination of Malmquist bias and the mass-scaling of IA amplitudes combine to increase the effective measured IA with increasing redshift, thereby motivating our NLA- $z$  and NLA- $M$  models.

For our fiducial NLA- $M$  model, the IA do change with tomographic bin, due to their varying fraction of passive/red galaxies and their typical mass (blue galaxies have zero IA in this model). In our analysis, the alignment amplitudes initially climb through

bins one to five, but rapidly drop in bin six where there are very few galaxies with an early-type SED (see Appendix B and Table B.1).<sup>7</sup> As such, the NLA- $M$  model represents a functional form that is not accessible to any of our other models. It is therefore interesting to note the similarity between the inferred values of  $A_{IA, total}$  in our NLA- $z$  and NLA- $M$  models. Both show the expected increase in alignment amplitudes with increasing redshift, although the NLA- $z$  model is at slightly lower amplitudes overall. We attribute this difference to the sixth tomographic bin: as the NLA- $z$  model requires that all bins follow the same (monotonic) trend, any reduction in  $A_{IA, total}$  between the fifth and sixth tomographic bins will require a compensatory reduction in alignment amplitude across all bins.

### 5.2.5. Impact of observational and systematic nuisance parameters

To demonstrate the impact of the various nuisance parameters and the relative importance of marginalisation over various priors, we ran a posterior analysis with our intrinsic alignment model uncertainty ( $\sigma_{IA}$ ) and redshift and shear calibration uncertainty ( $\sigma_z, \sigma_m$ ) nuisance parameters set to zero. The associated parameters to these uncertainties were fixed to the centre of their fiducial prior. We repeated the posterior calculation using only the remaining (cosmological and physical) parameters. Figure 11 shows these results. We found that the posterior constraints on  $\Sigma_8$  in the absence of marginalisation over these observational and systematic nuisance parameters are tighter by 10% (20%) for our  $E_n$  ( $C_E$ ) measurements, and are shifted with respect to the fiducial by  $0.28\sigma$  ( $0.14\sigma$ ). This suggests that KiDS-Legacy is statistics-limited.

### 5.2.6. Alternative redshift distribution calibration

We analysed the changes in parameter constraints found under different redshift calibration scenarios. In particular, we demonstrate the constraints found when we switch off shape-measurement weighting of the calibration sample, and when we switch off prior volume weighting of the calibration sample. In both cases, we found that the cosmological constraints are consistent with our fiducial constraints, showing a mild increase of  $\Delta\Sigma_8 \leq 0.17\sigma$ . This is, to some degree, expected: if our redshift distribution estimation changed appreciably when modifying our calibration method, this would also be reflected in our estimated bias parameters  $\delta z$ . As such, this test is particularly relevant for demonstrating possible sensitivity to source galaxies that reside outside the redshift baseline of our SKiLLS simulations. The lack of significant change in our cosmological parameters under changing the redshift calibration methodology suggests that our analysis is not strongly influenced by sources outside the redshift limits of SKiLLS.

Conversely, we also show in Fig. 11 the result of calibrating our redshift distributions using simulations from MICE2. These are the simulations that were used for KiDS-1000  $N(z)$  calibration, but differ here in that we applied our KiDS-Legacy calibration methodology to these simulations. Wright et al. (2025) discuss the accuracy of the redshift distribution bias parameters estimated from MICE2, in particular possible biases due to the MICE2 redshift limits of  $0.1 \lesssim z \lesssim 1.4$ . As our bin six  $N(z)$

<sup>7</sup> This may be a (suboptimal) feature of our SED based classification, which is unable to reliably identify the precursor sources of modern pressure-supported systems using uncertain SED models, especially at high- $z$ .

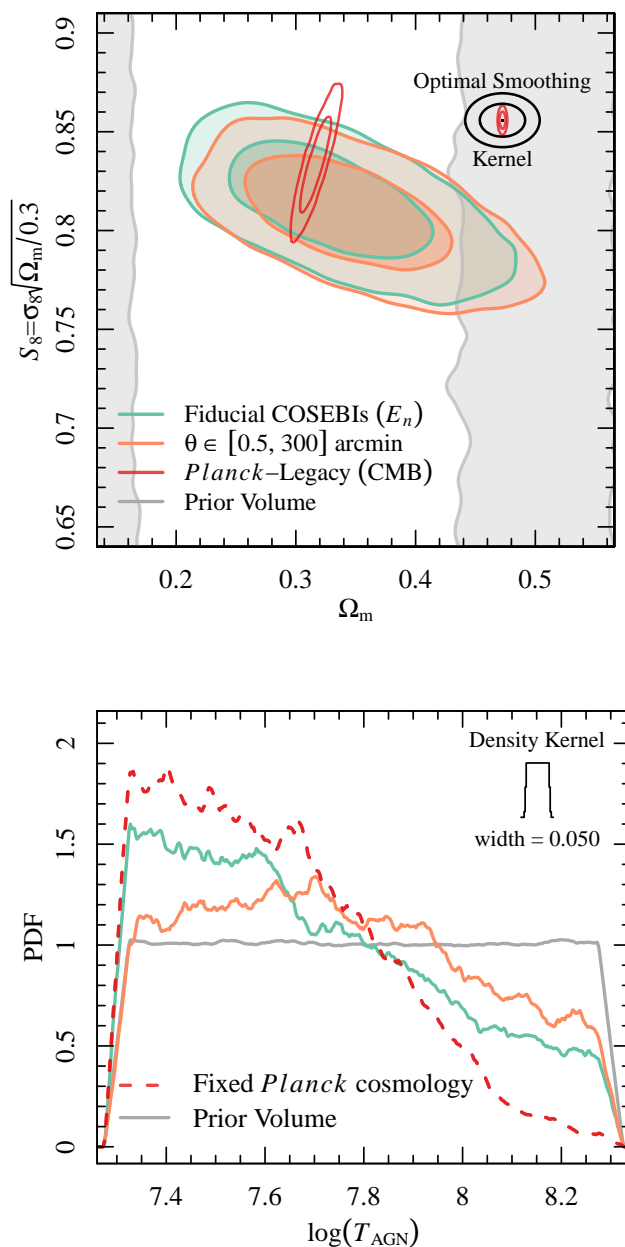
(in particular) extends beyond this upper limit, it is expected that our redshift distribution bias parameters there will be inaccurate. Therefore, we include them here predominantly as a demonstration of the effect of switching from MICE2 to SKiLLS. The maximum difference in the mean redshifts between our fiducial (effective)  $N(z)$  and that from MICE2 is  $\Delta\mu_{z,\text{eff}} = 0.016$ , in bin 3. However, the difference found is coherent between bins 3 – 6, which is therefore expected to cause a shift in  $S_8$ . As seen in Fig. 11, our analysis using  $N(z)$  calibrated with the MICE2 simulations increases the central value of our  $\Sigma_8$  constraints, by up to  $1.25\sigma$  (for  $E_n$ ). We consider these results inferior to the fiducial redshift calibration, though, given the considerably less realistic simulations (compared to SKiLLS), and the smaller redshift baseline that truncates the sixth bin to half of its supposed width.

In addition to our SOM calibrated redshift distributions, we also show results using our redshift distributions calibrated to cross-correlation measurements (Wright et al. 2025). Our cosmological constraints from this analysis are essentially unchanged in their central value for both  $E_n$  and  $C_E$ , showing a  $0.01\sigma$  shift in the central value of  $\Sigma_8$  (computed with the fiducial  $\alpha$  for each statistic). However, the results have lower constraining power, given the larger uncertainty in (particularly) the sixth tomographic bin: uncertainties in the optimal projection of  $\Sigma_8$  are 23% (29%) larger for our CC  $N(z)/\delta(z)$  marginalisation than in the fiducial case for  $E_n$  ( $C_E$ ). Nonetheless, the strong agreement between the estimates made with our CC and SOM  $N(z)$  provides confidence that redshift distribution biases in KiDS-Legacy are under control.

### 5.2.7. Scale cuts

In KiDS-1000, Asgari et al. (2021) presented cosmological constraints using a range of angular scales from  $0.5 \leq \theta \leq 300'$ . Subsequent investigation by Dark Energy Survey and Kilo-Degree Survey Collaboration et al. (2023), however, found that an optimal treatment of baryonic systematic effect mitigation (for  $E_n/B_n$ ) was achieved using  $2.0 \leq \theta \leq 300'$ . For KiDS-1000 (Asgari et al. 2021), this resulted in a shift in the marginal parameter constraints of  $S_8$  at the level of  $0.8\sigma$ . Reanalysis of KiDS-1000 by Li et al. (2023a), with improved shape measurement and calibration from Li et al. (2023b), showed a reduced sensitivity to the choice of scale cut, seeing a shift of  $0.3\sigma$ . We also tested the influence of changing our scale cuts, shown in the marginal projection of  $\Omega_m, S_8$  in Fig. 13, finding that including smaller scales results in a slight ( $0.1\sigma$ ) decrease in the marginal value of  $\Sigma_8$  and  $S_8$ . We also saw a slight change in the preferred value of our baryonic feedback parameter (see Fig. 13) where there is a reduction in posterior probability mass at negligible feedback amplitudes (i.e.  $\log_{10}(T_{\text{AGN}}/K) \approx 7.3$ , consistent with dark-matter only) when adding in smaller scale information. However, this must be interpreted with caution, as both posteriors are formally prior dominated in this parameter. This extra data is expected to be sensitive to baryon feedback, but it is difficult to distinguish from a simple change in the noise realisation. In either case, though, it is clear that the additional scales have at most a minor impact on our cosmological constraints.

We note that the expanded scale cut has a non-negligible effect on the significance of our  $B$  mode null test, causing it to once again fail our  $p > 0.01$  requirement. A full reanalysis of the additional astrometric selections required to reach non-significance in our null test with this expanded scale cut has not been explored here, and so the accuracy of the constraints with these expanded scales is not guaranteed. Nonetheless, the stability of



**Fig. 13.** Comparison of the marginal constraints on  $\Omega_m$  and  $S_8$  (top) and  $\log_{10}(T_{\text{AGN}}/K)$  (bottom) when using a wider range of scales ( $0.5 \leq \theta \leq 300'$ , orange), compared to the fiducial (blue-green). We compare the constraints to the full fiducial *Planck*-Legacy posterior (solid red, left) and the marginal constraint on  $\log_{10}(T_{\text{AGN}}/K)$  when analysing our fiducial data vector with cosmology fixed to the *Planck*-Legacy best-fit and marginalisation over nuisance parameters only (dashed red, right).

the cosmological constraints in the presence of both baryonic feedback contamination and possible systematics contamination is encouraging.

### 5.2.8. Removing tomographic redshift bins

In previous analyses of KiDS, the removal of individual tomographic bins (including all cross-correlations) introduced changes in marginal cosmological constraints of up to  $1.8\sigma$  in  $\Sigma_8$ . (see e.g. Asgari et al. 2021). We performed the same test here, and show the results in Fig. 11. We found that our cosmolog-

ical results are more robust to the removal of individual tomographic bins than previous work from KiDS: we observe a maximal difference in  $\Sigma_8$  when removing individual bins of  $0.24\sigma$  for  $E_n$ ,  $0.17\sigma$  for  $C_E$ . We attribute this decreased sensitivity at least partly to the reduction in relative importance of each tomographic bin (as signal is shared more evenly between the bins using our approximately equal-number tomography, and sixth bin). Furthermore, in our companion paper (Stölzner et al. 2025), we performed a series of internal consistency tests aimed at specifically estimating the constraints from different combinations of tomographic bins, finding that they were all consistent with each other, unlike KiDS-1000. As such, this represents a considerable improvement in internal consistency for KiDS that we attribute to the enhanced redshift calibration samples and methodology, as well as with the general improvements in data quality seen in Legacy compared to previous KiDS data releases.

## 6. Discussion

The results from KiDS-Legacy are consistent with the results of *Planck* Legacy, with a marginal  $\Sigma_8$  consistency of  $0.73\sigma$  for  $E_n$  and  $1.01\sigma$  for  $C_E$ . The constraints from KiDS-Legacy are also of comparable constraining power to those from *Planck* when projected into  $\Sigma_8$ , with *Planck* having a marginal uncertainty of  $\pm 0.015$  compared to Legacy's 0.016.

Compared to previous cosmological constraints from KiDS, Legacy represents a decrease in the marginal tension with *Planck* (computed using the Hellinger distance) of  $2.2\sigma$ , compared to previous work from Li et al. (2023a), who utilised methods and modelling choices that most resemble those used in Legacy. Such a difference in marginal constraints from individual KiDS analyses was deemed acceptable prior to unblinding (indeed, such a judgement call was required prior to unblinding), however was nonetheless surprising to the authors. This prompted an investigation into the underlying drivers of the change in cosmological constraints between the two samples.

A complete description of the investigation into the consistency between KiDS-1000 and KiDS-Legacy is provided in Appendix I. In summary, we found that improvements to our redshift distribution estimation and calibration methodology are responsible for approximately two thirds ( $1.32\sigma$ ) of the reduction in marginal tension with *Planck*, as reported by KiDS-1000 and Legacy. An additional reduction is attributable to a reduction in statistical noise from the sources outside the KiDS-1000 footprint. These results are shown graphically in Fig. I.1.

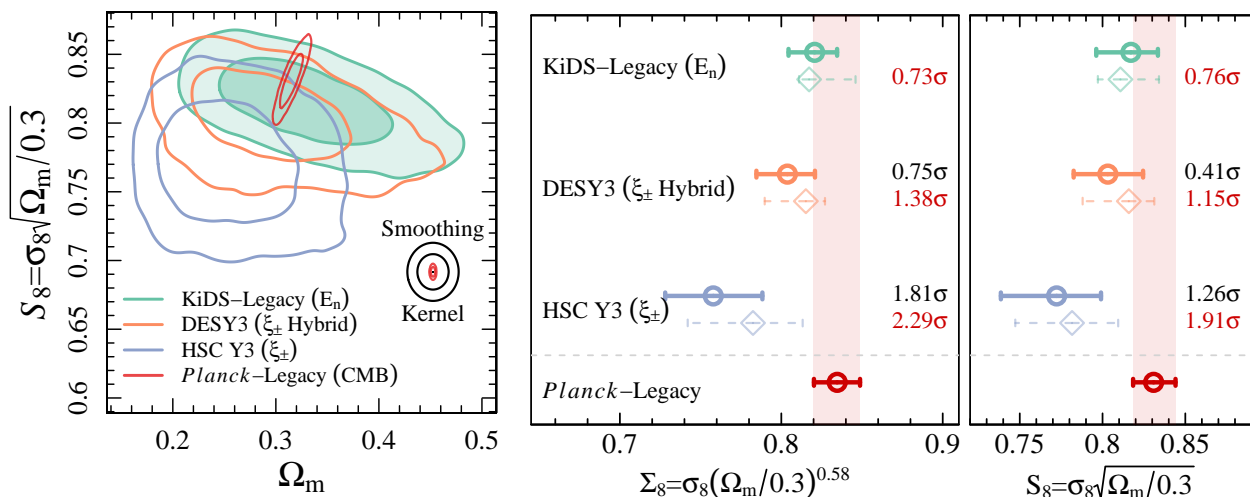
We also explored the consistency between our new Legacy cosmological parameters with recent results from the literature. Figure 14 presents the  $E_n$  cosmological constraints from Legacy in the context of two recent stage-III cosmic shear analyses: DES-Y3 hybrid-pipeline results from Dark Energy Survey and Kilo-Degree Survey Collaboration et al. (2023), and 2PCF measurements from HSC-Y3 (Li et al. 2023c). DES-Y3 is a largely independent analysis (there is a small on-sky overlap), which is also most consistent with Legacy: differences between marginal cosmological parameter constraints are  $0.75\sigma$  in  $\Sigma_8$  and  $0.41\sigma$  in  $S_8$ . Comparisons to the joint analysis of DESY3 and KiDS-1000, or to HSC, are both more covariate (due to area overlap) and show larger differences ( $1.5\sigma$  and  $1.8\sigma$  in  $\Sigma_8$  respectively), but are otherwise consistent. Overall, we found good consistency between all samples presented here, and with *Planck*. To this end, there is no evidence of significant tension in  $S_8$  (or in the more constraining  $\Sigma_8$ ) from KiDS-Legacy. Furthermore, the consistency that we found between KiDS-Legacy and other low-redshift cosmological

probes motivates the joint analysis of cosmological parameters with these surveys, which can be found in our companion paper (Stölzner et al. 2025). Lastly, as mentioned before, we do not include  $\xi_{\pm}$  into our fiducial analysis. However, the constraints and resulting consistency metrics are very similar and is discussed in Appendix F.

We performed an analysis of the effective constraining power provided by KiDS-Legacy over the matter power spectrum  $P(k)$  and the amplitude of baryon feedback. Figure 15 shows our constraint over  $P(k)$  from the fiducial  $E_n$ , estimated by sampling the full posterior of KiDS-Legacy and generating posterior predictive distributions for  $P(k)$  using CAMB and HMCode2020. We show the  $P(k)$  constraint normalised to the mean of all posterior samples, with a  $0.1\sigma$  region shaded around this central reference. This allows us to directly compare our  $P(k)$  constraint to the distribution of  $P(k)$  returned by two power spectrum emulators, BACCO (Aricò et al. 2021) and EUCLIDEMULATOR2 (Euclid Collaboration: Knabenhans et al. 2021), also evaluated at the same posterior samples. We note, though, that these emulators have a reduced range of validity compared to our fiducial priors, particularly in  $n_s$ . Nonetheless, we see that the differences between the emulated power spectra and those returned by CAMB are within the  $0.1\sigma$  credible interval of our posterior constraint on  $P(k)$ , demonstrating that differences between the various  $P(k)$  models (evaluated at the same cosmological and nuisance parameters) are consistently less than 10% of our constraining power, and are thus unlikely to be a significant source of bias in our analysis. With increased constraining power (from e.g. stage-IV surveys), however, the uncertainty in our modelling of  $P(k)$  will become a significant fraction of the final uncertainty.

We also explored the importance of baryon feedback modelling on our cosmological constraints, by examining the posterior on our feedback amplitude parameter  $\log_{10}(T_{\text{AGN}}/K)$  in the context of our  $P(k)$  constraints. We found that the contribution to the power spectrum from our inferred feedback amplitude is negligible for scales  $k \lesssim 0.2 h/\text{Mpc}$ . Above this scale our posterior constraint on the feedback model introduces a 1 – 10% suppression in power (compared to the dark-matter only case), and is consistent with the fiducial baryon feedback model from FLAMINGO at all scales. Above approximately  $k = 1 h/\text{Mpc}$ , our marginal posterior on  $P(k)$  excludes the more extreme FLAMINGO feedback models. Our MAP estimate of  $\log_{10}(T_{\text{AGN}}/K)$  is at the lower boundary limit of the prior, indicating that our best-fitting model prefers negligible baryon feedback amplitudes. Our marginal constraint over  $\log_{10}(T_{\text{AGN}}/K)$  is unconstrained (see Fig. 13), but is nonetheless skewed towards the MAP value.

Finally, given the agreement between our cosmological constraints and those from *Planck*, we investigated the inferred amplitude of baryon feedback with cosmological parameters fixed to those reported by *Planck*. This has the effect of placing a strong informative prior that the cosmological model is as described by *Planck*, and used cosmic shear to marginalise effectively over only observational systematic effects (IA,  $\delta z$ , and baryon feedback). Our posterior constraints over all systematic parameters are essentially unchanged with respect to the fiducial, with the only small change of note being that our baryon feedback posterior skewed slightly more strongly towards the lower prior boundary of  $\log_{10}(T_{\text{AGN}}/K) = 7.3$ , while the MAP remains at the boundary (see Fig. 13). In the fiducial case our marginal posterior on  $\log_{10}(T_{\text{AGN}}/K)$  was unconstrained (using the metric of Asgari et al. 2021, see their Appendix A), however with the informative *Planck* prior we were able to place an upper limit constraint of  $\log_{10}(T_{\text{AGN}}/K) \leq 7.71$  (8.01) at one-sided 68%



**Fig. 14.** Comparison between KiDS-Legacy  $E_n$  cosmological constraints and recent stage-III cosmic shear survey results from the literature. We show the 2D posteriors for each analysis  $\Omega_m, S_8$  (left) as well as the marginal projections in  $\Sigma_8$  ( $\alpha = 0.58$ , middle), and  $S_8$  (right). The 1D marginal ‘whisker’ plots are annotated as previously: circular points represent marginal mode and HPD intervals, while diamonds represent maximum a posteriori points and their associated PJ-HPD. Values are all calculated directly from the chains released by each analysis/survey, and in all cases except Legacy we take the MAP directly from the chain. We see that all surveys presented here are consistent with both KiDS-Legacy (black annotations) and *Planck* (red annotations) using our fiducial consistency metric of  $|\Delta X| \leq 2.6\sigma$  (where  $X$  is either  $\Sigma_8$  or  $S_8$ ), as estimated using the Hellinger tension between marginal mode summaries.

(95%) confidence. As such, we argue that there is little evidence in KiDS-Legacy for significant amplitudes of baryon feedback, even when placing strong informative priors on cosmological parameters.

## 7. Summary

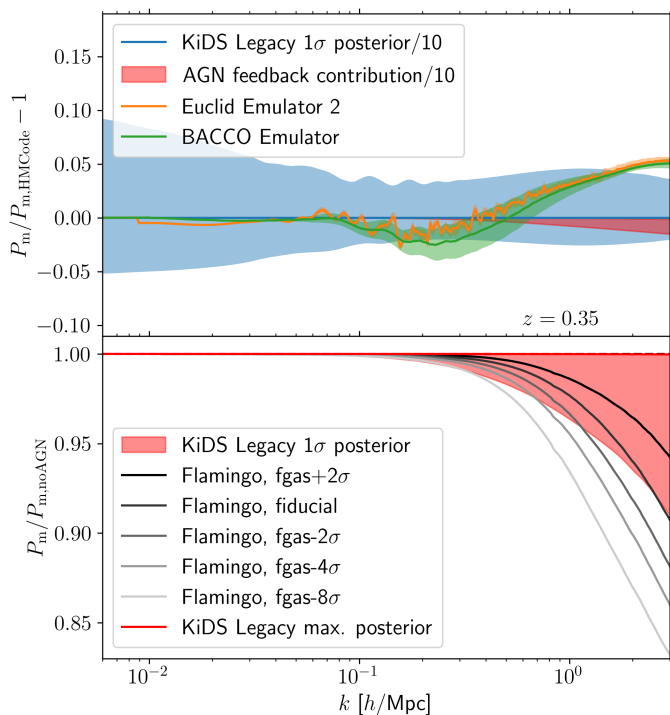
We present a cosmological analysis of the completed Kilo-Degree Survey (KiDS). Utilising data from KiDS Data Release 5 (Wright et al. 2024), we constructed our cosmic shear source sample from 1347 square degrees of optical and near-infrared imaging, containing deeper imaging, new astrometric and photometric calibrations, and cleaner masking of spurious sources and artefacts. Our improved data was then coupled with a greatly expanded spectroscopic calibration dataset named ‘KiDZ’ (consisting of 2.4 – 5.0 times the spectra as used in the analysis of KiDS-1000), along with an updated redshift distribution calibration methodology (‘gold-weight’; see Wright et al. 2025). This allowed us to calibrate sources at higher redshifts than those in any previous KiDS analysis. We defined six tomographic bins, spanning photometric redshifts between  $0.1 < z_B \leq 2.0$ , and calibrated the source redshift distributions with a precision up to six times higher than previously possible. We calibrated the source shapes using multi-band image simulations (Li et al. 2023b), finding that the KiDS-Legacy sample has the lowest systematic contamination (by e.g. PSF leakage) of any KiDS dataset to date. As per standard practices in the field, our cosmological analysis was performed fully blind.

Our fiducial cosmological parameter constraints were estimated with COSEBIs ( $E_n$ ), utilising conservative  $\theta \in [2.0, 300.0]$  arcmin scales and a new physically-motivated intrinsic alignment (IA) model. Our fiducial  $E_n$  measurement constrains  $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3} = 0.815^{+0.016}_{-0.021}$ , which is in full agreement (Hellinger distance of  $0.76\sigma$ ) with the marginal prediction  $S_8 = 0.834^{+0.016}_{-0.016}$  from the *Planck* Legacy cosmic microwave background experiment. Our fiducial measurement represents a  $\sim 15\%$  improvement in constraining power over the previous results from KiDS. However, we also found a significant

change in optimal degeneracy between  $\sigma_8$  and  $\Omega_m$ , with  $\Sigma_8 \equiv \sigma_8 (\Omega_m/0.3)^\alpha$  and  $\alpha = 0.58$  being preferred by our  $E_n$  measurement. We found  $\Sigma_8 = 0.821^{+0.014}_{-0.016}$ , which is a  $\sim 32\%$  improvement in optimal constraining power compared to previous KiDS analyses. In this projection, the corresponding Hellinger distance to *Planck* is  $0.73\sigma$ . We further computed the cosmological parameters using band power spectra ( $C_E$ ) and found full agreement with our fiducial COSEBIs. We performed a number of analysis variations, with our cosmological parameter constraints shown to be consistent across all modelling choices, statistics, tomographic bins, and summary statistics. Our companion paper (Stözlner et al. 2025) presents a full suite of internal and external consistency tests, finding the KiDS-Legacy dataset to be the most internally robust sample produced by KiDS to date.

We tested for systematic bias with respect to previous results from KiDS, finding that most of the differences in parameter constraints can be attributed to improved redshift distribution estimation and calibration, as well as new survey area and improved image reduction. We leveraged our consistency with *Planck* to estimate the preferred baryon feedback amplitude of KiDS-Legacy, allowing us to place an upper limit on  $\log_{10} T_{\text{AGN}} \leq 7.71$  (8.01) at 68% (95%) confidence. As such, the results from KiDS-Legacy paint a harmonious picture of cosmology at the end of stage-III, with full agreement between the cosmological parameters estimated by both low- and high-redshift probes of cosmic structure.

*Acknowledgements.* The authors would like to sincerely thank Matthias Bartelmann, who has acted as the external blinding coordinator for every blinded analysis performed by KiDS in the last 10+ years. AHW and HHi are supported by the Deutsches Zentrum für Luft- und Raumfahrt (DLR), made possible by the Bundesministerium für Wirtschaft und Klimaschutz, under project 50QE2305, and acknowledge funding from the German Science Foundation DFG, via the Collaborative Research Center SFB1491 ‘Cosmic Interacting Matters - From Source to Signal’. BS, CH, & ZY acknowledge support from 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. MA is supported by the UK Science and Technology Facilities Council (STFC) under grant number ST/Y002652/1 and the Royal Society under grant numbers RGS2222268 and ICAR1231094. MB & PJ are supported by the Polish National Science Center through grant no. 2020/38/E/ST9/00395. MB is



**Fig. 15.** *Upper panel:* Estimate of the systematic uncertainty in our modelling of the matter power spectrum using `CAMB` and `HMCode2020`, compared to two  $P(k)$  emulators `BACCO` (green) and `EUCLIDEMULATOR2` (orange). The figure shows the ratio between the power spectrum from `CAMB` and `HMCode2020` to those returned by each emulator, evaluated at 10 000 representative samples from the fiducial  $E_n/B_n$  posterior with a tenth of a standard-deviation. *Lower panel:* Constraints on baryonic feedback suppression from KiDS-Legacy. The marginal HPDI of feedback amplitudes from our fiducial  $E_n/B_n$  is shown in red, relative to a range of hydrodynamical models of baryonic feedback suppression from `FLAMINGO` (Schaye et al. 2023).

also supported by grant no. 2020/39/B/ST9/03494. BG acknowledges support from the UKRI Stephen Hawking Fellowship (grant reference EP/Y017137/1). CH also acknowledges support from the UK Science and Technology Facilities Council (STFC) under grant ST/V000594/1. HHi, CM, RR, & AD are supported by an ERC Consolidator Grant (No. 770935). HHi is also supported by a DFG Heisenberg grant (Hi 1495/5-1). HHo & MY acknowledge support from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program with Grant agreement No. 101053992. BJ acknowledges support by the ERC-selected UKRI Frontier Research Grant EP/Y03015X/1 and by STFC Consolidated Grant ST/V000780/1. KK acknowledges support from the Royal Society and Imperial College. SSL is receiving funding from the programme “Netzwerke 2021”, an initiative of the Ministry of Culture and Science of the State of North-Rhine Westphalia. MvWK acknowledges the support by the UKSA and STFC (grant no. ST/X001075/1). PB acknowledges financial support from the Canadian Space Agency (Grant No. 23EXPROSS1) and the Waterloo Centre for Astrophysics. CG is funded by the MICINN project PID2022-141079NB-C32. JHD acknowledges support from an STFC Ernest Rutherford Fellowship (project reference ST/S004858/1) SJ acknowledges the Ramón y Cajal Fellowship (RYC2022-036431-I) from the Spanish Ministry of Science and the Dennis Sciama Fellowship at the University of Portsmouth. LL is supported by the Austrian Science Fund (FWF) [ESP 357-N]. AL acknowledges support from the research project grant ‘Understanding the Dynamic Universe’ funded by the Knut and Alice Wallenberg Foundation under Dnr KAW 2018.0067. CM acknowledges support from the Beecroft Trust, and the Spanish Ministry of Science under the grant number PID2021-128338NB-I00. LM acknowledges support from STFC grant ST/N000919/1. LMo acknowledges the financial contribution from the grant PRIN-MUR 2022 20227RNLY3 “The concordance cosmological model: stress-tests with galaxy clusters” supported by Next Generation EU and from the grant ASI n. 2024-10-HH.0 ‘Attività scientifiche per la missione Euclid – fase E’. NRN acknowledges financial support from the National Science Foundation of China, Research Fund for Excellent International Scholars (grant n. 12150710511), and from the research grant from China Manned Space Project n. CMS-CSST-2021-A01. LP acknowledges

support from the DLR grant 50QE2002. MR acknowledges financial support from the INAF grant 2022. TT acknowledges funding from the Swiss National Science Foundation under the Ambizione project PZ00P2\_193352. YZ acknowledges the studentship from the UK Science and Technology Facilities Council (STFC). Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 179.A-2004, 177.A-3016, 177.A-3017, 177.A-3018, 298.A-5015. This work was performed in part at Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611.

*Author Contributions:* All authors contributed to the development and writing of this paper. The authorship list is given in three groups: the lead authors (AHW, BS), followed by two alphabetical groups. The first alphabetical group includes those who are key contributors to both the scientific analysis and the data products of this manuscript and release. The second group covers those who have either made a significant contribution to the preparation of data products or to the scientific analyses of KiDS-Legacy.

## References

- Abdalla, E., Abellán, G. F., Aboubrahim, A., et al. 2022, *Journal of High Energy Astrophysics*, 34, 49
- Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, *ApJS*, 249, 3
- Alonso, D., Sanchez, J., Slosar, A., & LSST Dark Energy Science Collaboration. 2019, *MNRAS*, 484, 4127
- Amon, A. & Efstathiou, G. 2022, *MNRAS*, 516, 5355
- Amon, A., Gruen, D., Troxel, M. A., et al. 2022, *Phys. Rev. D*, 105, 023514
- Aricò, G., Angulo, R. E., Contreras, S., et al. 2021, *MNRAS*, 506, 4070
- Aricò, G., Angulo, R. E., Hernández-Monteaugudo, C., et al. 2020, *MNRAS*, 495, 4800
- Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, *A&A*, 645, A104
- Asgari, M. & Schneider, P. 2015, *A&A*, 578, A50
- Asgari, M., Schneider, P., & Simon, P. 2012, *A&A*, 542, A122
- Asgari, M., Tröster, T., Heymans, C., et al. 2020, *A&A*, 634, A127
- Bacon, D. J., Massey, R. J., Refregier, A. R., & Ellis, R. S. 2003, *MNRAS*, 344, 673
- Bartelmann, M. & Schneider, P. 2001, *Phys. Rep.*, 340, 291
- Becker, M. R., Troxel, M. A., MacCrann, N., et al. 2016, *Phys. Rev. D*, 94, 022002
- Benítez, N. 2000, *ApJ*, 536, 571
- Bertin, E. & Arnouts, S. 1996, *A&AS*, 117, 393
- Blake, C., Amon, A., Childress, M., et al. 2016, *MNRAS*, 462, 4240
- Blake, C., Brough, S., Couch, W., et al. 2008, *Astronomy and Geophysics*, 49, 5.19
- Blazek, J. A., MacCrann, N., Troxel, M. A., & Fang, X. 2019, *Phys. Rev. D*, 100, 103506
- Bridle, S. & King, L. 2007, *New Journal of Physics*, 9, 444
- Brout, D., Scolnic, D., Popovic, B., et al. 2022, *ApJ*, 938, 110
- Broxterman, J. C. & Kuijken, K. 2024, *A&A*, 692, A201
- Capaccioli, M., Mancini, D., & Sedmak, G. 2005, *The Messenger*, 120, 10
- Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, *MNRAS*, 447, 646
- Catelan, P., Kamionkowski, M., & Blandford, R. D. 2001, *MNRAS*, 320, L7
- Crittenden, R. G., Natarajan, P., Pen, U.-L., & Theuns, T. 2001, *ApJ*, 559, 552
- Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, *MNRAS*, 453, 1513
- Dalal, R., Li, X., Nicola, A., et al. 2023, *Phys. Rev. D*, 108, 123519
- Dark Energy Survey and Kilo-Degree Survey Collaboration, Abbott, T. M. C., Aguena, M., et al. 2023, *The Open Journal of Astrophysics*, 6, 36
- DESI Collaboration, Adame, A. G., Aguilar, J., et al. 2024, *AJ*, 168, 58
- 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
- Edge, A., Sutherland, W., Kuijken, K., et al. 2013, *The Messenger*, 154, 32
- Euclid Collaboration, Mellier, Y., Abdurro’uf, et al. 2025, *A&A*, 697, A1
- Euclid Collaboration: Knabenhans, M., Stadel, J., Potter, D., et al. 2021, *MNRAS*, 505, 2840
- Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, *MNRAS*, 467, 1627
- Fortuna, M. C., Dvornik, A., Hoekstra, H., et al. 2025, *A&A*, 694, A322
- 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
- Gaia Collaboration: Brown, A. G. A., Vallenari, A., Prusti, T., et al. 2018, *A&A*, 616, A1
- Ge, F., Millea, M., Camphuis, E., et al. 2025, *Phys. Rev. D*, 111, 083534
- Georgiou, C., Chisari, N. E., Bilicki, M., et al. 2025, *A&A*, 699, A252
- Giblin, B., Heymans, C., Asgari, M., et al. 2021, *A&A*, 645, A105

- Gong, Y., Liu, X., Cao, Y., et al. 2019, *ApJ*, 883, 203
- Grandis, S., Aricò, G., Schneider, A., & Linke, L. 2024, *MNRAS*, 528, 4379
- Guzzo, L., Scodreggio, M., Garilli, B., et al. 2014, *A&A*, 566, A108
- Heydenreich, S., Schneider, P., Hildebrandt, H., et al. 2020, *A&A*, 634, A104
- Heymans, C., Brown, M. L., Barden, M., et al. 2005, *MNRAS*, 361, 160
- Heymans, C., Grocutt, E., Heavens, A., et al. 2013, *MNRAS*, 432, 2433
- Heymans, C., Tröster, T., Asgari, M., et al. 2021, *A&A*, 646, A140
- Heymans, C., White, M., Heavens, A., Vale, C., & van Waerbeke, L. 2006, *MNRAS*, 371, 750
- 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
- Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, *A&A*, 633, A69
- Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, *A&A*, 647, A124
- Hirata, C. M. & Seljak, U. 2004, *Phys. Rev. D*, 70, 063526
- Hoekstra, H., Yee, H. K. C., & Gladders, M. D. 2002, *ApJ*, 577, 595
- Hoffmann, K., Bel, J., Gaztañaga, E., et al. 2015, *MNRAS*, 447, 1724
- Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, *J. Cosmology Astropart. Phys.*, 04, 027
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111
- Jarvis, M., Bernstein, G., & Jain, B. 2004, *MNRAS*, 352, 338
- Jarvis, M., Sheldon, E., Zuntz, J., et al. 2016, *MNRAS*, 460, 2245
- Jee, M. J., Tyson, J. A., Hilbert, S., et al. 2016, *ApJ*, 824, 77
- 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
- Johnston, H., Georgiou, C., Joachimi, B., et al. 2019, *A&A*, 624, A30
- Johnston, H., Wright, A. H., Joachimi, B., et al. 2021, *A&A*, 648, A98
- Kaiser, N. 1992, *ApJ*, 388, 272
- Kilbinger, M., Fu, L., Heymans, C., et al. 2013, *MNRAS*, 430, 2200
- Kilbinger, M., Heymans, C., Asgari, M., et al. 2017, *MNRAS*, 472, 2126
- Knabenhans, M., Brinckmann, T., Stadel, J., Schneider, A., & Teyssier, R. 2023, *MNRAS*, 518, 1859
- Kodwani, D., Alonso, D., & Ferreira, P. 2019, *The Open Journal of Astrophysics*, 2, 3
- Köhlinger, F., Joachimi, B., Asgari, M., et al. 2019, *MNRAS*, 484, 3126
- Kohonen, T. 1982, *Biological Cybernetics*, 43, 59
- Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, *MNRAS*, 454, 3500
- Kuijken, K. 2011, *The Messenger*, 146, 8
- Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, *A&A*, 625, A2
- Lamman, C., Tsaprazi, E., Shi, J., et al. 2024, *The Open Journal of Astrophysics*, 7, 14
- Lange, J. U. 2023, *MNRAS*, 525, 3181
- Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, *A&A*, 439, 845
- Lemos, P., Weaverdyck, N., Rollins, R. P., et al. 2023, *MNRAS*, 521, 1184
- Leonard, C. D., Ferreira, T., Fang, X., et al. 2023, *The Open Journal of Astrophysics*, 6, 8
- Lewis, A., Challinor, A., & Lasenby, A. 2000, *ApJ*, 538, 473
- Li, S.-S., Hoekstra, H., Kuijken, K., et al. 2023a, *A&A*, 679, A133
- Li, S.-S., Kuijken, K., Hoekstra, H., et al. 2023b, *A&A*, 670, A100
- Li, X., Zhang, T., Sugiyama, S., et al. 2023c, *Phys. Rev. D*, 108, 123518
- 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., Blake, C., Bridle, S., et al. 2011, *MNRAS*, 410, 844
- Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinckmann, J. 2006, *MNRAS*, 368, 715
- McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, *MNRAS*, 465, 2936
- McCullough, J., Amon, A., Legnani, E., et al. 2024, *arXiv e-prints*, arXiv:2410.22272
- Mead, A. 2015, *HMcode: Halo-model matter power spectrum computation*, Astrophysics Source Code Library, record ascl:1508.001
- Mead, A. J., Brieden, S., Tröster, T., & Heymans, C. 2021, *MNRAS*, 502, 1401
- 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
- Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *ApJS*, 208, 5
- Oehl, V. & Tröster, T. 2024, *arXiv e-prints*, arXiv:2407.08718
- Paulin-Henriksson, S., Amara, A., Voigt, L., Refregier, A., & Bridle, S. L. 2008, *A&A*, 484, 67
- Pen, U.-L., Van Waerbeke, L., & Mellier, Y. 2002, *ApJ*, 567, 31
- Piras, D., Joachimi, B., Schäfer, B. M., et al. 2018, *MNRAS*, 474, 1165
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6
- Preston, C., Amon, A., & Efstathiou, G. 2023, *MNRAS*, 525, 5554
- Qu, F. J., Sherwin, B. D., Madhavacheril, M. S., et al. 2024, *ApJ*, 962, 112
- Reischke, R., Kiessling, A., & Schäfer, B. M. 2017, *MNRAS*, 465, 4016
- Reischke, R., Unruh, S., Asgari, M., et al. 2025, *A&A*, 699, A124
- Rowe, B. 2010, *MNRAS*, 404, 350
- Salcido, J., McCarthy, I. G., Kwan, J., Upadhye, A., & Font, A. S. 2023, *MNRAS*, 523, 2247
- Samuroff, S., Blazek, J., Troxel, M. A., et al. 2019, *MNRAS*, 489, 5453
- Samuroff, S., Mandelbaum, R., Blazek, J., et al. 2023, *MNRAS*, 524, 2195
- Schaller, M., Schaye, J., Kugel, R., Broxterman, J. C., & van Daalen, M. P. 2025, *MNRAS*, 539, 1337
- Schaye, J., Kugel, R., Schaller, M., et al. 2023, *MNRAS*, 526, 4978
- Schneider, A., Giri, S. K., Amodeo, S., & Refregier, A. 2022, *MNRAS*, 514, 3802
- Schneider, A. & Teyssier, R. 2015, *J. Cosmology Astropart. Phys.*, 2015, 049
- Schneider, M. D. & Bridle, S. 2010, *MNRAS*, 402, 2127
- Schneider, P., Eifler, T., & Krause, E. 2010, *A&A*, 520, A116
- Schneider, P., van Waerbeke, L., Jain, B., & Kruse, G. 1998, *MNRAS*, 296, 873
- Schneider, P., van Waerbeke, L., & Mellier, Y. 2002, *A&A*, 389, 729
- 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
- Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M. P., & McCarthy, I. G. 2011, *MNRAS*, 417, 2020
- Sheather, S. J. & Jones, M. C. 1991, *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, pp. 683
- Shuntov, M., McCracken, H. J., Gavazzi, R., et al. 2022, *A&A*, 664, A61
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, *arXiv e-prints*, arXiv:1503.03757
- Spurio Mancini, A., Piras, D., Alsing, J., Joachimi, B., & Hobson, M. P. 2022, *MNRAS*, 511, 1771
- Stözlner, B., Wright, A. H., Asgari, M., et al. 2025, *A&A*, accepted, arXiv:2503.19442
- Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, *ApJ*, 761, 152
- Tessore, N., Loureiro, A., Joachimi, B., von Wietersheim-Kramsta, M., & Jeffrey, N. 2023, *The Open Journal of Astrophysics*, 6, 11
- Tonegawa, M., Okumura, T., & Hayashi, M. 2025, *PASJ*, 77, 389
- Tröster, T., Asgari, M., Blake, C., et al. 2021, *A&A*, 649, A88
- van den Busch, J. L., Hildebrandt, H., Wright, A. H., et al. 2020, *A&A*, 642, A200
- van Uitert, E., Cacciato, M., Hoekstra, H., et al. 2016, *MNRAS*, 459, 3251
- van Uitert, E., Cacciato, M., Hoekstra, H., & Herbonnet, R. 2015, *A&A*, 579, A26
- van Uitert, E., Joachimi, B., Joudaki, S., et al. 2018, *MNRAS*, 476, 4662
- van Wietersheim-Kramsta, M., Lin, K., Tessore, N., et al. 2025, *A&A*, 694, A223
- Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2025, *A&A*, submitted, arXiv:2503.19440
- Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2020, *A&A*, 640, L14
- Wright, A. H., Kuijken, K., Hildebrandt, H., et al. 2024, *A&A*, 686, A170
- Yan, Z., Wright, A. H., Elisa Chisari, N., et al. 2025, *A&A*, 694, A259
- Yoon, M. & Jee, M. J. 2021, *ApJ*, 908, 13
- Yoon, M., Jee, M. J., Tyson, J. A., et al. 2019, *ApJ*, 870, 111
- Zuntz, J., Paterno, M., Jennings, E., et al. 2015, *Astronomy and Computing*, 12, 45

<sup>1</sup> Ruhr University Bochum, Faculty of Physics and Astronomy, Astrophysical Institute (AIRUB), German Centre for Cosmological Lensing, 44780 Bochum, Germany

<sup>2</sup> School of Mathematics, Statistics and Physics, Newcastle University, Herschel Building, NE1 7RU, Newcastle-upon-Tyne, UK

<sup>3</sup> Center for Theoretical Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668 Warsaw, Poland

<sup>4</sup> Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK

<sup>5</sup> Leiden Observatory, Leiden University, P.O.Box 9513, 2300RA Leiden, The Netherlands

<sup>6</sup> Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

<sup>7</sup> Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany

<sup>8</sup> Institute for Computational Cosmology, Ogden Centre for Fundamental Physics - West, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK.

<sup>9</sup> Centre for Extragalactic Astronomy, Ogden Centre for Fundamental Physics - West, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

<sup>10</sup> Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada

<sup>11</sup> Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada

- <sup>12</sup> Institute for Theoretical Physics, Utrecht University, Princetonplein 5, 3584CC Utrecht, The Netherlands.
- <sup>13</sup> Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
- <sup>14</sup> Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
- <sup>15</sup> School of Mathematics, Statistics and Physics, Newcastle University, Herschel Building, NE1 7RU, Newcastle-upon-Tyne, UK
- <sup>16</sup> Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Av. Complutense 40, E-28040 Madrid, Spain
- <sup>17</sup> Institute of Cosmology & Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth, PO1 3FX, United Kingdom
- <sup>18</sup> Dipartimento di Fisica e Astronomia "Augusto Righi" - Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, 40129 Bologna, Italy
- <sup>19</sup> INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
- <sup>20</sup> Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria
- <sup>21</sup> The Oskar Klein Centre, Department of Physics, Stockholm University, AlbaNova University Centre, SE-106 91 Stockholm, Sweden
- <sup>22</sup> Imperial Centre for Inference and Cosmology (ICIC), Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2AZ, UK
- <sup>23</sup> Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, United Kingdom.
- <sup>24</sup> Donostia International Physics Center, Manuel Lardizabal Ibilbidea, 4, 20018 Donostia, Gipuzkoa, Spain.
- <sup>25</sup> Zentrum für Astronomie, Universität Heidelberg, Philosophenweg 12, D-69120 Heidelberg, Germany; Institute for Theoretical Physics, Philosophenweg 16, D-69120 Heidelberg, Germany
- <sup>26</sup> Istituto Nazionale di Fisica Nucleare (INFN) - Sezione di Bologna, viale Berti Pichat 6/2, I-40127 Bologna, Italy
- <sup>27</sup> Department of Physics "E. Pancini" University of Naples Federico II C.U. di Monte Sant'Angelo Via Cintia, 21 ed. 6, 80126 Naples, Italy
- <sup>28</sup> INAF - Osservatorio Astronomico di Padova, via dell'Osservatorio 5, 35122 Padova, Italy
- <sup>29</sup> Institute for Particle Physics and Astrophysics, ETH Zürich, Wolfgang-Pauli-Strasse 27, 8093 Zürich, Switzerland

## Appendix A: Emulating matter power spectra with COSMOPOWER

The most computationally expensive section of our modelling pipeline is the prediction of the matter power spectra with `CAMB`. Fortunately, we can circumvent this time-consuming computation through the use of artificial neural network emulators, which act to provide a direct non-linear mapping between input cosmological parameters and the corresponding output power spectrum. Such an emulator can be created with relative ease using the `COSMOPOWER` framework of Spurio Mancini et al. (2022). The original `COSMOPOWER` emulator was trained to reproduce the linear matter power spectrum,  $P_{m,lin}(k)$ , and the perturbations caused by non-linear evolution,  $P_{m,nl}(k)/P_{m,lin}(k)$ , given a set of relevant input parameters. We updated this process to emulate residuals  $\log_{10}[P(k)/P_{ref}(k)]$  between logarithmic power spectra (for both linear and non-linear spectra) and a single reference power spectrum instead. The reference spectrum was chosen to be near the centre of our sampling/prior volume, which is defined by the ranges of the various parameters in Table 3. These ranges subsequently also define the range of validity of our emulator.

Construction of the emulator requires a large sample of training and testing power-spectra, which were generated by `CAMB`, over the seven-dimensional<sup>8</sup> volume that we wish to sample. For this purpose, we generated 450 000 training and 50 000 testing spectra, produced by a Latin hypercube sampling over our volume.

The consistency between our emulated spectra and the target spectra returned by `CAMB` and `HMCODE2020` is validated in the publicly available emulator repository on GitHub<sup>9</sup>, through computation of the distribution of absolute fractional residuals between the target and estimated spectra for all 50 000 testing samples. We found that our updated method of emulation produces a more uniform error distribution than the original emulation method from Spurio Mancini et al. (2022), across the wide range of scales that we emulate ( $k \in [10^{-5}, 20] h/\text{Mpc}$ ), and generally reduces the maximum error over all scales. Our emulator exhibits a maximal error below 0.1% for 68% of our test samples, in both the linear and non-linear power spectra.

## Appendix B: Galaxy sample properties for intrinsic alignment modelling

For our fiducial galaxy-type and mass-dependent intrinsic alignment (IA) model, we need to measure the average host halo mass  $M_h$  and the red galaxy fraction  $f_r$  in each tomographic bin. These, and relevant intermediate properties, are summarised in Table B.1. We note that, where appropriate, we computed averages as the means over logarithmic quantities as these propagate straightforwardly to the averages of derived quantities in power-law relations. Moreover, all averages were weighted by the shear measurement weight and redshift estimation code weight.

To determine  $f_r$ , we used the output of the photometric redshift estimation from the template-fitting code `BPZ` (Benítez 2000). `BPZ` employs a set of six model templates, ordered in terms of star formation activity, and interpolates linearly between ad-

<sup>8</sup> Four of our parameters were kept fixed to fiducial values in our power spectra: the curvature contribution to the total energy density ( $\Omega_K$ ) was fixed to zero, the sum of the neutrino masses ( $\sum m_\nu$ ) was fixed to its minimum standard value, and the equation-of-state parameters of dark energy ( $w_0, w_a$ ), were fixed at the values for a cosmological constant. See Table 3.

<sup>9</sup> <https://github.com/KiDS-WL/CosmoPowerCosmosis>

**Table B.1.** Early-type galaxy sample properties for the mass-dependent intrinsic alignment model per tomographic bin.

Bin	$\langle z \rangle$	$\log_{10} \left( \frac{\langle L_r \rangle}{L_0} \right)$	$\log_{10} M_{\text{eff}} [h^{-1} M_\odot]$	$f_r$
1	0.34	-1.32	11.69	0.15
2	0.48	-0.72	12.46	0.20
3	0.59	-0.47	12.76	0.17
4	0.79	-0.28	12.93	0.24
5	0.94	-0.14	13.08	0.19
6	1.23	0.01	13.21	0.03

**Notes.** Columns, from left to right: tomographic bin number, mean redshift, logarithmic normalised galaxy luminosity in the  $r$ -band, logarithm of the effective halo mass, and the red galaxy fraction.

acent template to determine a best-fitting spectral energy distribution. The `BPZ` output  $T_B$  encodes the best-fit templates and their admixture in steps of 0.1. We chose  $T_B < 1.9$  to define early-type galaxies, encompassing all galaxies with some contribution of an elliptical galaxy spectrum (template no. 1). We validated that, at least on a matched sample between the KiDS and GAMA (Driver et al. 2011) surveys (i.e. for relatively bright and low-redshift galaxies), this  $T_B$  cut is capable of isolating the red sequence.

We found that in the first five tomographic bins  $f_r$  fluctuates around 0.2, broadly in line with expectations for a flux-limited sample. The sixth bin contains a negligible fraction of early-type galaxies, a feature that we also see in our simulated sixth bin from SKiLLS. The statistical error on  $f_r$  is very small. We repeated the measurement varying  $T_B$  by  $\pm 0.1$  and obtained changes in  $f_r$  of less than 0.02. Therefore, we decided to consider  $f_r$  as fixed in the IA model.

To compute halo masses, we made use of the  $r$ -band luminosities derived as part of the KiDS Data Release 5 (Wright et al. 2024). van Uitert et al. (2015) measured effective halo masses for SDSS (LOWZ and CMASS) early-type galaxies via weak lensing, as a function of luminosity and redshift. We used their reported halo mass results and galaxy properties to construct the following linear relation

$$\log_{10} M_h^{(i)} [h^{-1} M_\odot] = \gamma \left( \frac{1}{1 + \langle z \rangle^{(i)}} - \frac{1}{1 + z_{\text{piv}}} \right) + \delta \log_{10} \left( \frac{\langle L_r \rangle^{(i)}}{L_0} \right) + \Delta M, \quad (\text{B.1})$$

where we chose  $z_{\text{piv}} = 0.3417$ , such that the parameters  $\gamma$  and  $\Delta M$  are uncorrelated to a good approximation. An additional parameter  $\delta$  describes the dependence on the average luminosity  $\langle L_r \rangle$ , normalised by the reference luminosity  $L_0 = 10^{11} (h/0.7)^{-2} L_\odot$ . van Uitert et al. (2015) showed that the luminosity dependence evolves only mildly with redshift, so that we determine  $\delta$  as the inverse variance-weighted average scaling over the four samples they considered. Thus, we obtained  $\gamma = 1.25 \pm 0.33$ ,  $\delta = 1.42 \pm 0.16$ , and  $\Delta M = 13.56 \pm 0.02$ . In line with expectations, both the average luminosity and halo mass increase monotonically by more than an order of magnitude from the first to the sixth tomographic bin (see Table B.1).

Statistical uncertainties on the halo mass are propagated via Monte Carlo sampling from the fit and measurement uncertainties. The standard deviation of  $\log_{10} M_h$  is approximately 0.2 in the first tomographic bin, driven by the extrapolation of Eq. (B.1) to fainter samples, and of order 0.1 in the remaining bins, where

a reduced error due to luminosities more compatible with the range fitted by van Uitert et al. (2015) is gradually traded off with larger uncertainty because of extrapolation to higher redshifts. Halo mass estimates are strongly correlated (correlation coefficient greater than 0.9) between tomographic bins due to the extrapolation from the luminosity and redshift ranges considered by van Uitert et al. (2015). Therefore, we propagated the uncertainty in halo masses into our IA model by assuming a multivariate Gaussian prior with standard deviation of 0.2 in  $\log_{10} M_h$  for all bins and the correlation matrix as obtained from the Monte Carlo error propagation.

Throughout the construction of our astrophysics-informed galaxy IA model, we have not distinguished between central and satellite galaxies, although their alignment mechanisms are believed to be different (Schneider & Bridle 2010). The NLA model assumes a fixed relation between linear and non-linear scales where centrals and satellites tend to dominate IA signals, respectively, although our scale-dependent IA model tests whether the data prefers a different ratio on average over the whole survey. However, the satellite fraction in our early-type galaxy sample is unknown, yet almost certainly evolves with halo mass and redshift (see e.g. Shuntov et al. 2022). For a given mass and redshift, early-type satellite fractions are typically twice as large as those of late types (Mandelbaum et al. 2006), so that the impact of potential IA signals of blue satellites is less of a concern.

Moreover, the satellite fractions in most galaxy samples used in direct IA constraints are also unknown. Fortuna et al. (2025) estimated a satellite fraction of 30% and possibly higher (see also van Uitert et al. 2016) in a subset of the GAMA-KiDS sample and subsequently excluded it from their fit. Johnston et al. (2019) studied alignments amongst and between centrals and satellites. They tentatively found a steeper slope of the red satellite IA signal on small transverse scales. The mass scaling of the IA amplitude in red satellites is qualitatively similar to that of centrals (Fortuna et al. 2021a).

In summary, neither cosmic shear surveys nor direct measurements to calibrate IA currently provide sufficient information to treat the IA signals of satellites and their associated trends separately. We are forced to treat satellite effects as part of an effective small-scale IA model, with any discrepancies in the combined dependencies on scale and galaxy sample properties expected to be comfortably absorbed in the still wide priors. Once surveys reach the statistical power to significantly constrain IA models beyond an overall amplitude, calibration measurements of IA will need to be (re-)analysed separately for central and satellite populations. Halo models (Fortuna et al. 2021a) will be a natural choice to accurately represent different IA mechanisms and scalings on large and small scales.

### Appendix C: Band-powers derivation

To arrive at the correct response functions for our band-powers, we first write  $C_{E/B,l}$  as a function of  $\xi_{\pm}(\theta)$ , in the same manner as for the COSEBIs in Eq. (16),

$$C_{E/B,l} = \frac{\pi}{\mathcal{N}_l} \int_0^{\infty} d\theta \theta T(\theta) \left[ g_+(\theta) \xi_+(\theta) \pm g_-(\theta) \xi_-(\theta) \right], \quad (\text{C.1})$$

where  $g_{\pm}^l(\theta)$  are the band-power filter functions for bin  $l$  and  $T(\theta)$  is a function that limits the range of  $\theta$ -scales that contribute to the integral. If  $\xi_{\pm}$  is known for all scales then  $T(\theta) = 1$  everywhere. We now write  $g_{\pm}^l(\theta)$  in terms of the idealised response function,

$$g_{\pm}^l(\theta) = \int_0^{\infty} d\ell \ell S_l(\ell) J_{0/4}(\ell\theta). \quad (\text{C.2})$$

If  $S_l(\ell)$  is a top-hat then,

$$g_+(\theta) = \frac{1}{\theta^2} \left[ \theta \ell_{\text{up},l} J_1(\theta \ell_{\text{up},l}) - \theta \ell_{\text{lo},l} J_1(\theta \ell_{\text{lo},l}) \right], \quad (\text{C.3})$$

$$g_-(\theta) = \frac{1}{\theta^2} \left[ \mathcal{G}_-(\theta \ell_{\text{up},l}) - \mathcal{G}_-(\theta \ell_{\text{lo},l}) \right],$$

with,

$$\mathcal{G}_-(x) = \left( x - \frac{8}{x} \right) J_1(x) - 8J_2(x). \quad (\text{C.4})$$

Inserting for  $\xi_{\pm}$  from Eq. (15) into Eq. (C.1) we arrive at,

$$C_{E,l}^{(ij)} = \frac{1}{2\mathcal{N}_l} \int_0^{\infty} d\ell \ell \left[ W_{EE}^l(\ell) C_{\varepsilon\varepsilon,E}^{(ij)}(\ell) + W_{EB}^l(\ell) C_{\varepsilon\varepsilon,B}^{(ij)}(\ell) \right], \quad (\text{C.5})$$

$$C_{B,l}^{(ij)} = \frac{1}{2\mathcal{N}_l} \int_0^{\infty} d\ell \ell \left[ W_{BE}^l(\ell) C_{\varepsilon\varepsilon,E}^{(ij)}(\ell) + W_{BB}^l(\ell) C_{\varepsilon\varepsilon,B}^{(ij)}(\ell) \right],$$

where the weight functions are,

$$W_{EE}^l(\ell) = \int_0^{\infty} d\theta \theta T(\theta) \left[ J_0(\ell\theta) g_+(\theta) + J_4(\ell\theta) g_-(\theta) \right], \quad (\text{C.6})$$

$$W_{EB}^l(\ell) = \int_0^{\infty} d\theta \theta T(\theta) \left[ J_0(\ell\theta) g_+(\theta) - J_4(\ell\theta) g_-(\theta) \right],$$

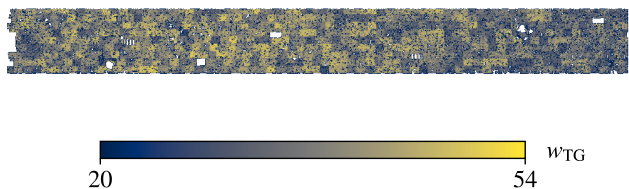
with  $W_{EE}^l(\ell) = W_{BB}^l(\ell)$  and  $W_{EB}^l(\ell) = W_{BE}^l(\ell)$ . Since  $\xi_{\pm}$  are measured between  $\theta_{\text{min}}$  and  $\theta_{\text{max}}$ ,  $T(\theta)$  has to be defined such that no information beyond this range is included. To avoid ringing, we apodise  $T(\theta)$  with a Hahn window and smooth its transitions to zero:

$$T(\theta) = \begin{cases} 0; & x < x_{\text{lo}} - \frac{\Delta_x}{2} \\ \cos^2 \left[ \frac{\pi}{2} \frac{x - (x_{\text{lo}} + \Delta_x/2)}{\Delta_x} \right]; & x_{\text{lo}} - \frac{\Delta_x}{2} \leq x < x_{\text{lo}} + \frac{\Delta_x}{2} \\ 1; & x_{\text{lo}} + \frac{\Delta_x}{2} \leq x < x_{\text{up}} - \frac{\Delta_x}{2} \\ \cos^2 \left[ \frac{\pi}{2} \frac{x - (x_{\text{up}} - \Delta_x/2)}{\Delta_x} \right]; & x_{\text{up}} - \frac{\Delta_x}{2} \leq x < x_{\text{up}} + \frac{\Delta_x}{2} \\ 0; & x \geq x_{\text{up}} + \frac{\Delta_x}{2}. \end{cases} \quad (\text{C.7})$$

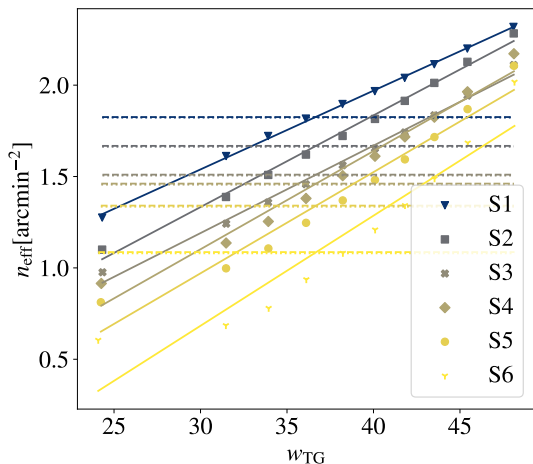
Here  $x = \log(\theta)$ ,  $\Delta_x$  is the log-width of the apodisation,  $x_{\text{lo}} - \Delta_x/2 = \log(\theta_{\text{min}})$  and  $x_{\text{up}} + \Delta_x/2 = \log(\theta_{\text{max}})$ . We require  $\Delta_x > \log(\theta_{\text{up}}/\theta_{\text{lo}})$ , and in practice have used  $\Delta_x = 0.5$  for all KiDS-Legacy  $C_{E,B}$  analyses. We note, also, that this apodisation method is different to how we applied the apodisation for KiDS-1000 analysis, where we instead had  $x_{\text{lo}} = \log(\theta_{\text{min}})$  and  $x_{\text{up}} = \log(\theta_{\text{max}})$ . This change was made as it was deemed more intuitive/appropriate that the band power windows be contained primarily within the requested  $\ell$  ranges after apodisation, rather than before.

### Appendix D: KiDS-Legacy variable depth mocks and covariance validation

Our analysis used an analytical covariance matrix, as summarised in Sect. 2.6. Here, we describe the mocks which were used to validate the covariance matrix (see Reischke et al. 2025) in more detail.



**Fig. D.1.** Cartesian spatial map (corresponding  $N_{\text{side}} = 1024$ , showing a local flat projection of the healpix map) of the TG weights,  $w_{\text{TG}}$ , throughout a single field of our KiDS-Legacy-like mock catalogue. Larger TG weights correspond to a higher local source density.



**Fig. D.2.** Dependence of the galaxy density,  $n_{\text{eff}}$ , on TG weights,  $w_{\text{TG}}$  in the KiDS-Legacy-like mock catalogue. The data points represent the mean  $n_{\text{eff}}$  of ten equi-populated bins in  $w_{\text{TG}}$ . The solid line shows the linear fit to the aforementioned data points of their respective tomographic bin, labelled S1-S6. The dashed horizontal lines show the mean values of  $n_{\text{eff}}$  calculated from the galaxy samples with variable depth per tomographic bin, while the dashed horizontal lines show the values of  $n_{\text{eff}}$  for the respective galaxy samples without any spatial variations in the observational depth.

The log-normal simulations are based on GLASS<sup>10</sup> (Generator for Large-Scale Structure, Tessore et al. 2023) with a detailed description given in von Wietersheim-Kramsta et al. (2025) as well as Section 8.1 in Reischke et al. (2025). The resulting density field realisations, in concentric shells, were subsequently integrated along the line-of-sight weighted by the weak lensing kernel, Eq. (2), to create a total of 4224 realisations of the shear field.

For each realisation, galaxies were sampled, factoring in systematic effects such as the survey footprint and variable depth (as caused by seeing, sky transparency fluctuations, and differences in the number of overlapping pointings) using SALMO<sup>11</sup> (Speedy Acquisition for Lensing and Matter Observables, Joachimi et al. 2021). The inclusion of a sixth tomographic bin in KiDS-Legacy incorporates more faint galaxies in the sample, enhancing the overall magnitude of variable depth and boosting the lensing signal relative to the noise. Although previous analyses suggested that depth variability effects on the signal in KiDS-1000 remain below the noise threshold (Heydenreich et al. 2020), this might not be the case for KiDS-Legacy.

To model variable depth, we rely on organised randoms (ORs, introduced in Johnston et al. 2021; Yan et al. 2025). ORs

use SOMs to generate random samples that reflect the spatial variability caused by a high-dimensional systematic space. The ORs provide Tiaogeng (TG) weights,  $w_{\text{TG}}$ , (see Yan et al. 2025, for a detailed explanation), quantifying the spatial variability of the selection function of the galaxy sample as a function of position on the sky. The ORs robustly trace the variations in depth for KiDS-Legacy. Yan et al. (2025) demonstrated that the TG weights are uncorrelated with  $r$ -band magnitude, photometric redshift, and underlying cosmological density variations. In contrast, they are highly correlated with the magnitude limit of the survey in the  $r$ -band, which is used for both source extraction and shape measurement.

Figure D.1 shows the TG weights for a single field in our KiDS-Legacy-like mock, as a function of position on the simulated sky. Additionally, Fig. D.2 shows the corresponding local source density of sources as a function of TG weights, colour-scaled for the different source bins. The horizontal dashed lines depict the effective homogeneous number density. Solid lines are linear fits to the measured effective number density in ten equi-populated bins of the TG weights. This linear interpolation allows for an effective evaluation of  $n_{\text{eff}}$  at each pixel.

The local density variations seen here also affect the redshift distributions. This issue is addressed by reapplying the photometric redshift calibration using SOMs (see Wright et al. 2025, for details) to each tomographic bin and each of the ten TG weight bins. The resulting redshift distribution then enters the shear maps via the weak lensing kernel.

On each simulation, we measured the shear two-point correlation functions as indicated in Sect. 2.5, from which the sample covariance could be estimated directly. These mocks were then compared to the ONECOVARIANCE code. The results are presented in Reischke et al. (2025) with 10 percent agreement for the variances, very similar to what was found in previous analysis (Joachimi et al. 2021). In particular, we found that the limiting factor in the covariance modelling is not the effect of variable depth but rather the survey geometry on large scales. All these effects, however, do not significantly affect the cosmological results presented in this paper.

## Appendix E: $B$ -mode analysis

During the initial phases of the blinded KiDS-Legacy analysis, our  $B$ -mode null-test analyses (see Sect. 4.3) failed our required significance threshold ( $p \geq 0.01$ ). This led to a dedicated effort to investigate the possible causes of the  $B$ -mode signal in the dataset, starting with the highest level data products and systematically working our way back through the data production and analysis pipelines, to identify the root cause of the  $B$ -mode signal.

As KiDS-1000 did not show significant  $B$  modes (specifically, as analysed by Li et al. 2023a using the same shape-measurement code and analysis choices, such as scale cuts), the initial test in our investigation was to determine whether the observed  $B$ -mode significance in KiDS-Legacy was owing to a decrease in covariance (specifically shot noise) caused by the increased area of KiDS-Legacy over KiDS-1000. We therefore started by analysing the KiDS-Legacy  $B$ -mode significance in the same footprint as KiDS-1000, finding that the KiDS-Legacy data in this footprint still exceeded the  $B$ -mode significance requirement.

Subsequent investigation of the  $B$ -mode origin focussed on PSF contamination (via Paulin-Henriksson statistics; see Sect. 4), one- and two-dimensional  $c$ -terms, covariance construction, and shape measurement biases (as a function of, for ex-

<sup>10</sup> <https://glass.readthedocs.io/stable/>

<sup>11</sup> <https://github.com/Linc-tw/salmo>

ample, galaxy morphology and position on-sky). Each of these tests was performed locally (per square degree tile), in the detector plane (i.e. coadding all sources in the frame of the focal plane), with spatial jackknife subsets of the full survey, and globally (e.g. at the data vector level, and looking for population-level trends). Each of these investigations failed to illuminate the origin of the  $B$ -mode signal. Finally, we investigated the lowest-level possible cause of the anomalous  $B$ -mode signal: systematic failures in the astrometric solution used in the THELI calibration (with Gaia) of  $r$ -band images. The global astrometric accuracy of KiDS-Legacy is presented in Wright et al. (2024), where we demonstrated the accuracy of the THELI (constructed with a Gaia reference) and ASTRO-WISE (constructed with a SDSS and 2MASS reference) catalogues both globally (see their Figures 10, 11, C1, and C2) and as a function of right ascension and declination (RA and Dec; see their Figures 8 and 9), compared to SDSS, 2MASS, and Gaia. In our new investigation, we explored the astrometric accuracy in finer detail (i.e. higher on-sky resolution) and using the individual astrometric solutions of THELI exposures and ASTRO-WISE co-adds.

We computed two test statistics for investigation of the  $B$ -mode signal. The first test statistic is sensitive to the astrometric consistency between THELI co-added images, which are used for source detection, and ASTRO-WISE co-added images, which are used for photometric measurements (and therefore impacts photometric redshift estimation, which is used for tomographic binning). This metric ( $\delta r_{\text{int}}$ ) measured the systematic difference between the centroid of all targets extracted from these images. This was achieved by matching all sources (within  $2''$ ) extracted from ASTRO-WISE and THELI co-added  $r$ -band images. With these matches, we computed the radial separation  $r$  between each source's centroid in the two images and then calculated the median of these separations in  $1 \times 1$  arcminute on-sky bins. The 2D map of median separations was then assigned back to each source.

Our second test statistic was computed between individual exposures of the THELI  $r$ -band images, which are used for shape measurement. This metric ( $\delta r_{\text{exp}}$ ) measures the systematic difference between the centroids of targets extracted from individual THELI exposures, which are point-like and have high S/N. This was achieved by matching sources extracted from ASTRO-WISE (used here purely as a static reference) and the individual THELI exposures (within  $2''$ ). For the ASTRO-WISE static reference, we applied selections on SOURCE EXTRACTOR (Bertin & Arnouts 1996) output variables

$$\text{FLUX\_RADIUS} < 4.0 \quad (\text{E.1})$$

and

$$\text{FLUX\_ISO}/\text{FLUXERR\_ISO} > 10.0, \quad (\text{E.2})$$

to select point-like high-S/N sources. Settings used in the extraction of sources are provided in Table 4 of Wright et al. (2024). For the sources extracted from the individual THELI exposures, we performed the same point source selection (Eq. E.1) and a modified S/N criteria, to account for the reduced depth of the individual exposures compared to the co-add:

$$\text{FLUX\_ISO}/\text{FLUXERR\_ISO} > 10.0/\sqrt{(5)}. \quad (\text{E.3})$$

With these matches, we then computed the radial separation  $r_{ij}$  between all exposures in  $\{i, j \in [1, 5] | i \neq j\}$ , computed the maximum of these separations in  $2 \times 2$  arcminute on-sky bins, and smoothed these maps with a 1 arcminute Gaussian kernel. We

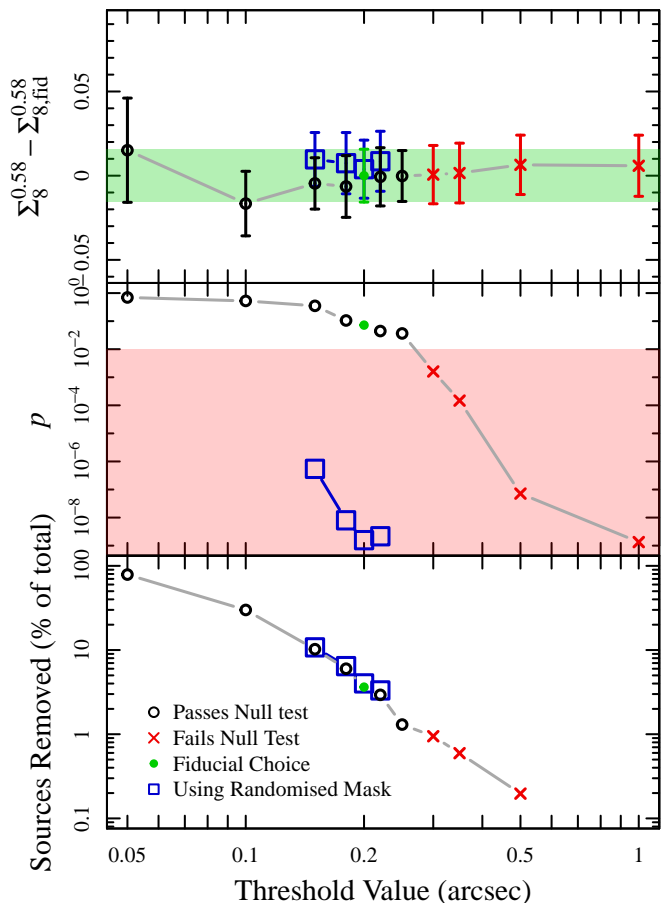
then combined these individual  $15 i \leftrightarrow j$  maps into a final single map, by selecting again the maximum separation per on-sky bin. This resulting map was then once-again smoothed using a 1 arcminute Gaussian kernel.

These maps provided us with two test statistics that could be used as an additional mask for the survey, given specific thresholds in the statistic values. The first metric is relevant for possible failures in photometry: if the astrometric solution between ASTRO-WISE and THELI differs, then the apertures extracted from THELI imaging will be incorrectly placed on the ASTRO-WISE images. The second metric is relevant for possible failures in shape measurement: the shape measurement algorithm *lensfit*, which is used in KiDS, assumes perfect astrometric alignment of exposures within the THELI  $r$ -band imaging. As such, any residual astrometric misalignment will manifest as an erroneously estimated shape, and coherent misalignments may manifest as an  $E$ - or  $B$ -mode cosmic-shear signal.

For our first statistic, the degree of inconsistency between the imaging that can lead to systematic errors in photometry is relatively benign: our photometric measurement code GAAP has a minimum aperture size of  $0''.7$ , driven largely by the fact that our imaging has seeing that is consistently larger than  $0''.6$  (see e.g. Figure 7 of Wright et al. 2024). Systematic errors in aperture centroids at the level of the FWHM and larger will lead to significant flux to be lost from the target aperture (assuming a point-like source), and thus may pose a problem for our photometry. However, there are two additional relevant considerations. Firstly, in our cosmological analyses, all sources that are used must be resolved and thus (in some direction) larger than the PSF FWHM. Secondly, even if significant flux is missed, if the ASTRO-WISE imaging are all well aligned to each other, then the flux will be consistently missed in all bands, thereby keeping the source SED largely consistent. As such, we expect the effect to be rarely relevant, and (when it is relevant) only to have a minor impact on photometric redshifts. Nonetheless, we chose to mask all areas of the survey where our first statistic exceeds the minimum PSF FWHM: that is, where  $\delta r_{\text{int}} > 0''.6$ . This corresponds to approximately  $0.7 \text{ deg}^2$  of removed data over the full KiDS-Legacy footprint.

For our astrometric statistic, there is much less intuition available regarding the level of astrometric error that may be problematic for our analysis. As such, we chose to investigate (while remaining fully blinded; outcomes of the various tests were very similar for all blinds) the impact of various masks (defined by thresholds in  $\delta r_{\text{exp}}$ ) on our null test results, and on our cosmological inference. Figure E.1 shows the results of this investigation. On the  $x$ -axis of each panel in the figure, we show the threshold in our astrometric metric  $\delta r_{\text{exp}}$  that was used to define a mask. This mask was then applied to the full KiDS-Legacy cosmic shear source sample, and the end-to-end cosmological analysis pipeline was run (including  $N(z)$  estimates, shape recalibration, two-point statistics, covariances, and cosmological chains). For each threshold, we were therefore able to report the change in sample size (bottom panel of Fig. E.1; similar to the change in area), the change in the  $B$ -mode null test  $p$ -value (middle panel), and the change in the inferred marginal value of  $\Sigma_8$  (top panel).

Firstly, we see that the null test  $p$ -value passes our pre-defined threshold of  $p = 0.01$  with  $\delta r_{\text{exp}} \lesssim 0''.25$ , corresponding to a masked fraction of sources of at least 1%. Secondly, we see that the value of  $\Sigma_8$  is robust to the application of increasingly conservative thresholds, demonstrating a  $\Delta\Sigma_8 \lesssim 0.5\sigma$  variation when discarding up to  $100 \text{ deg}^2$  of possibly contaminated data and causing the  $B$ -mode significance to decrease to  $p > 0.3$ . In



**Fig. E.1.** Summary of our blinded  $B$ -mode investigation where we measure the COSEBIs  $E_n/B_n$  cosmic shear statistics, masking survey area where our astrometric metric,  $\delta r_{\text{exp}}$ , exceeds a series of threshold values (shown on the  $x$ -axis). The middle panel shows the  $p$ -value that quantifies the likelihood that the  $B$ -mode is consistent with zero. As we reduce the  $\delta r_{\text{exp}}$  threshold, the  $p$ -value increases above  $p \leq 0.01$  (red shaded region), and the masked survey passes the  $B$ -mode null test. As the mask reduces the total number of sources (lower panel), the blue squares show the  $p$ -values from a random mask analysis, demonstrating that the improved  $B$ -mode is not driven by a reduction in area. In the upper panel we find that the astrometric masking has a benign impact on our  $\Sigma_8$  cosmological constraints. The green data point and shaded regions shows the results from our fiducial analysis, demonstrating that failing data-vectors are consistent with our fiducial, all having less than  $0.4\sigma$  changes in  $\Sigma_8$ .

the region around where we first pass the null test, and do not throw away an overly excessive fraction of the source sample (i.e.  $\delta r_{\text{exp}} \in [0''.20, 0''.25]$ ) our recovered marginal value of  $\Sigma_8$  is at most 0.006 (or  $0.36\sigma$ ) smaller than in the case where we perform no masking at all (equivalent to the threshold of  $1''$ ). Thirdly, we can see that the null test  $p$ -value is correlated with the fraction of removed area, which begs the question of whether the improvement in our null test might simply be the result of reduced constraining power related to the smaller sample. To test this possibility, we must construct a new mask that removes the same amount of area from the survey at a given threshold, but which does not trace the underlying astrometric errors that we think are likely responsible for our  $B$ -mode signal. To do this, we constructed a randomised version of our astrometric metric, by randomising the individual masks (defined per square degree

tile) between the 1347 tiles of the survey. Furthermore, we also flip the shuffled masks in both the RA and Dec directions, to remove any possible correlations between the constructed masks and the focal plane. We then repeated the analysis using the randomised mask, and found that the  $B$ -mode signal remained significant (see the blue points in Fig. E.1). As such, we conclude that our astrometric mask is removing a bone fide  $B$ -mode signal from the data vector, and not just removing sensitivity to the  $B$  mode by lowering the constraining power.

As such, from these tests we can see that our astrometric statistic targets our  $B$ -mode signal, and that our recovered cosmology is fairly insensitive to our choice of the mask. We therefore opt to define a fiducial threshold in our test statistic that passes our  $B$ -mode test and makes some intuitive sense. To that end, we define the fiducial mask to be where  $\delta r_{\text{exp}} > 0''.2$ , where  $0''.2$  is chosen as a round number that is smaller than the `THELI` pixel size ( $0''.212$ ). This corresponds to approximately 4% of the full KiDS-Legacy source sample, drawn from a previously unmasked area of 46.6 square degrees.

## Appendix F: Cosmological parameters from $\xi_{\pm}$

Here, we present cosmological parameter constraints from  $\xi_{\pm}$ , for comparison with previous work and external datasets. We followed the KiDS tradition of using nine logarithmic bins in  $\theta$ , but constructed these over the same scale cuts as used by our fiducial  $E_n/B_n$ :  $\theta \in [2, 300]'$ . We also removed the first  $\theta$ -bin from  $\xi_{-}$ , in effect limiting it to  $\theta \in [4, 300]'$ , following previous KiDS analyses. For our theoretical predictions we must also apply this binning scheme, however we note that it is not sufficient to simply integrate  $\xi_{\pm}(\theta)$  in Eq. (15) over the bin width. Instead, we must weight  $\xi_{\pm}(\theta)$  by the weighted number of galaxy pairs,  $N_{\text{pair}}(\theta)$  (see Appendix A of Asgari et al. 2020, for details).

We used the finely binned  $\xi_{\pm}(\theta)$  and its corresponding number of galaxy pairs to make our measurements and our predictions,

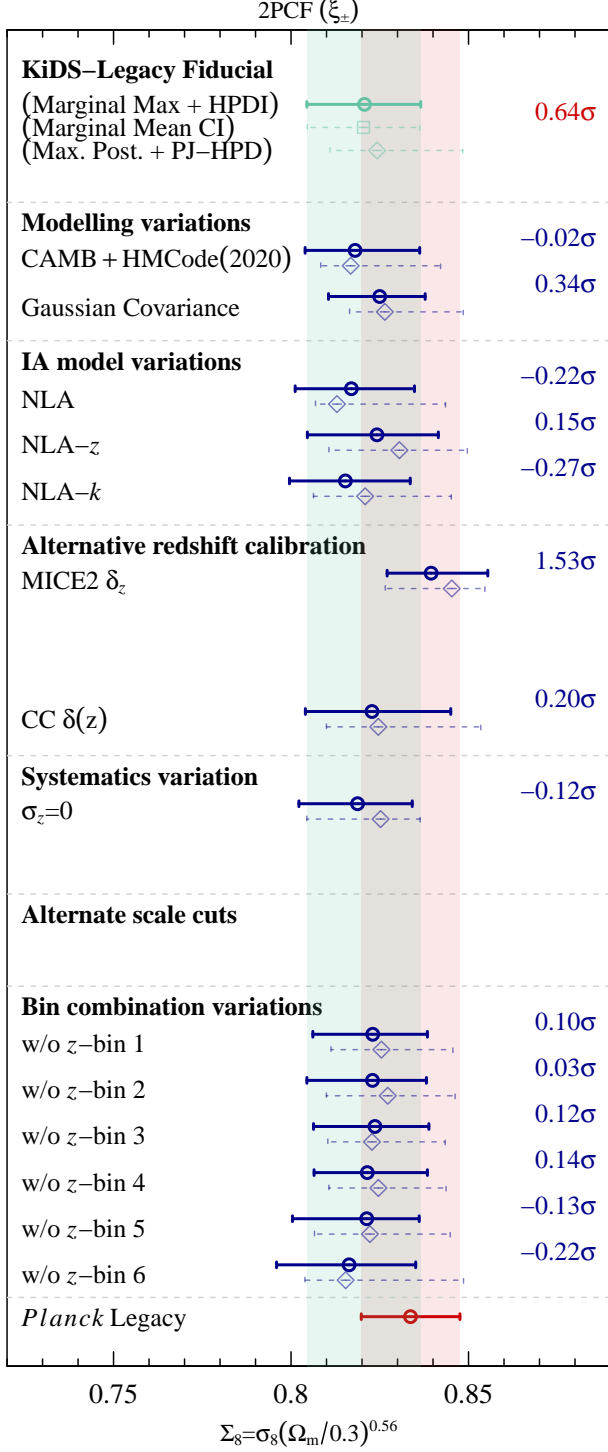
$$\xi_{\pm}^{\text{binned}}(\theta_{\text{bin}}) = \frac{\sum N_{\text{pair}}(\theta) \xi_{\pm}(\theta) \Delta(\theta_{\text{bin}} - \theta)}{\sum N_{\text{pair}}(\theta) \Delta(\theta_{\text{bin}} - \theta)}, \quad (\text{F.1})$$

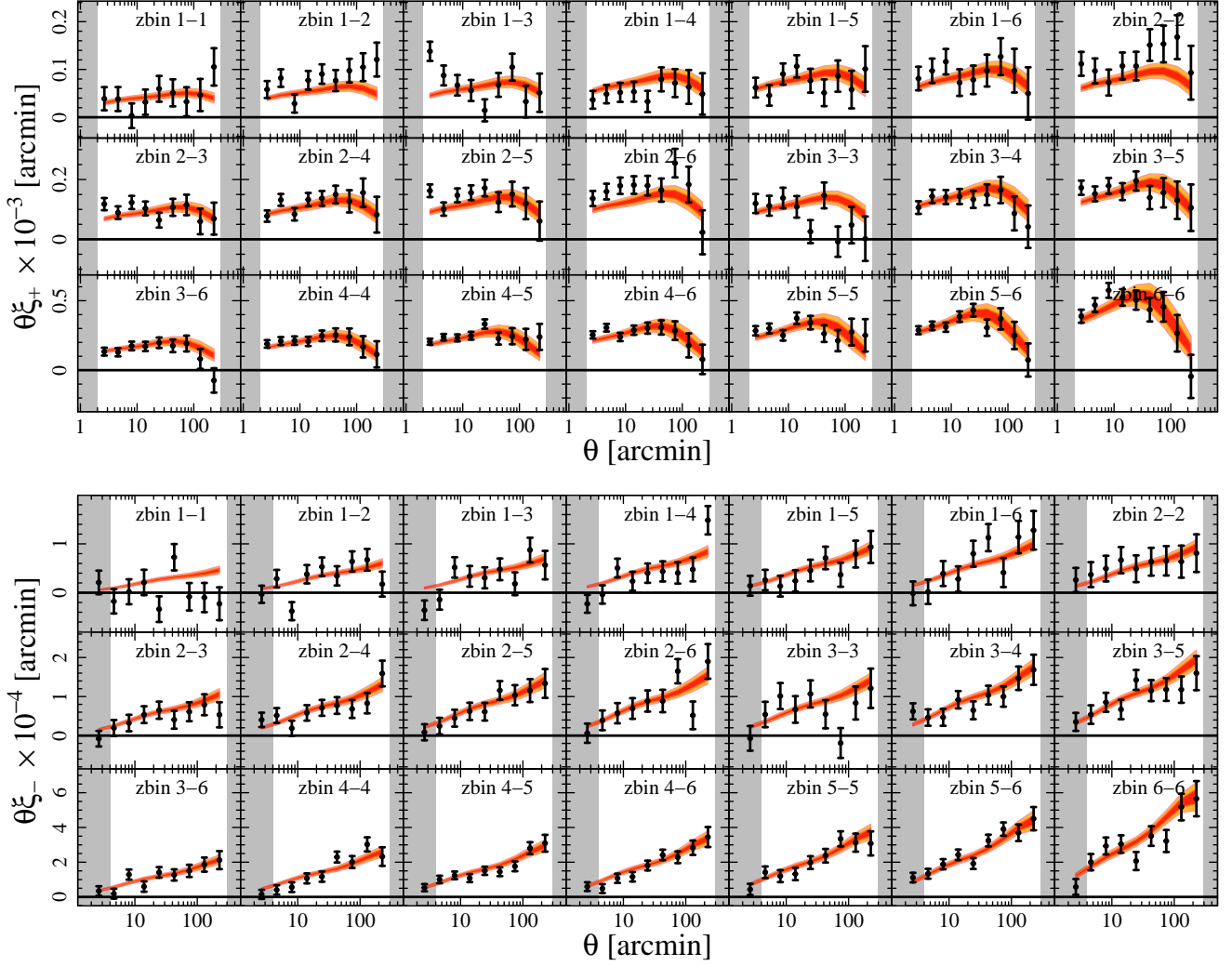
where  $\Delta(\theta_{\text{bin}} - \theta)$  is the binning function, which is equal to 1 if  $\theta$  is within the boundaries of the bin labelled  $\theta_{\text{bin}}$ , and is 0 otherwise. When predicting  $\xi_{\pm}^{\text{binned}}(\theta_{\text{bin}})$ , we assumed that the distribution of the number of pairs and the value of  $\xi_{\pm}(\theta)$  were both essentially constant within each of the 1000 bins, and used the predicted values of  $\xi_{\pm}$  at the centres of the bins from Eq. (15) in Eq. (F.1).

Figure F.1 presents the whisker diagram for our  $\xi_{\pm}$  constraints under all analyses, and Fig. F.2 presents our posterior predictive distributions for  $\xi_{\pm}$ . Despite our concerns regarding the possible systematic contamination of our  $\xi_{\pm}$  signals, our cosmological parameter constraints from  $\xi_{\pm}$  are in good agreement with the results from our other statistics. We note, however, that our posterior predictive distributions (shown in Fig. F.2) are much closer to the significance limit ( $\text{PTE} < 0.01$ ) than our other statistics (see results for the fiducial setup in Table F.1, and for the analysis variations in Table H.1). Finally, we note that our companion paper Stözlner et al. (2025) finds strong evidence for a scale-dependent tension between cosmological parameters modelled with  $\xi_{\pm}$ . Furthermore, Oehl & Tröster (2024) found that the likelihood of  $\xi_{\pm}$  becomes non-Gaussian on large scales. This is evidence that our choice to present constraints with  $\xi_{\pm}$  as less reliable was warranted.

**Table F.1.** Constraints on  $S_8$  ( $\alpha = 0.50$ ) and  $\Sigma_8$  ( $\alpha$  free) for  $\xi_{\pm}$  under our fiducial setup.

	Setup	statistic	$\alpha$	$\chi^2$	dof	PTE	$N_{\text{samp}}^{\text{PJ-HPD}}$	Marginal Mode+HPD	Max. Apost. +PJ-HPD	Marginal Mean + CI
$S_8$	Fiducial	$\xi_{\pm}$	0.50	413.1	351.5	0.013	48	$0.825^{+0.017}_{-0.018}$	$0.827^{+0.022}_{-0.015}$	$0.825^{+0.018}_{-0.017}$
$\Sigma_8$	Fiducial	$\xi_{\pm}$	0.55	413.1	351.5	0.013	11	$0.821^{+0.016}_{-0.016}$	$0.825^{+0.025}_{-0.013}$	$0.821^{+0.016}_{-0.016}$


**Fig. F.1.** Whisker diagram of  $\Sigma_8$  cosmological results from  $\xi_{\pm}$ . The figure is annotated as in Fig. 11.



**Fig. F.2.** Data vector and posterior predictive distributions for the fiducial run of our  $\xi_{\pm}$ . Scales in the grey shaded region are excluded from our fiducial analysis. TPDs are shown for 68<sup>th</sup> (red) and 95<sup>th</sup> (orange) percentiles of the posterior models.

## Appendix G: Additional posterior constraints

In this appendix we provide additional posterior constraints, for various combinations of parameters and model setups.

### G.1. Whisker diagram in $S_8$

In this section we show the  $S_8$  equivalent of our main whisker-diagrams, for comparison with previous work. Figure G.1 shows the constraints for our  $E_n$ ,  $C_E$ , and  $\xi_{\pm}$  statistics, with all analysis variations included. The variability between the constraints is larger in this projection than in the  $\Sigma_8$  projections, due to shifting of the posterior mode along the  $\sigma_8$ ,  $\Omega_m$  degeneracy. Otherwise, we note that all conclusions of this work are unchanged when analysing the  $S_8$  space, rather than the  $\Sigma_8$  space.

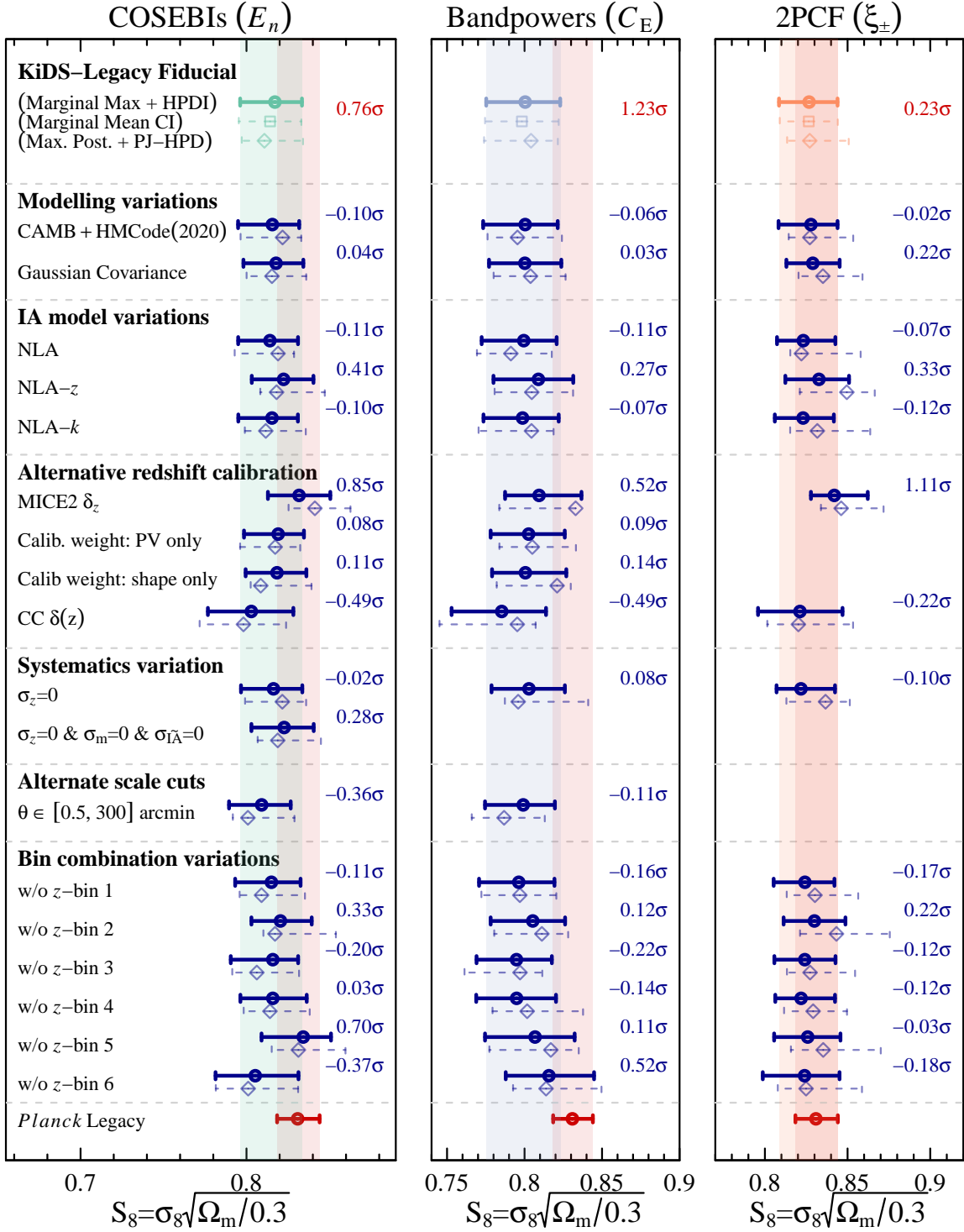
### G.2. Redshift distributions

Figure G.2 shows the posterior constraints on our redshift distribution bias parameters, for chains analysed using various intrinsic alignment models, compared to the prior. The chains are all consistent with the prior, with only the low- $z$  tomographic bins showing any tendency to shift away from the prior. Moreover,

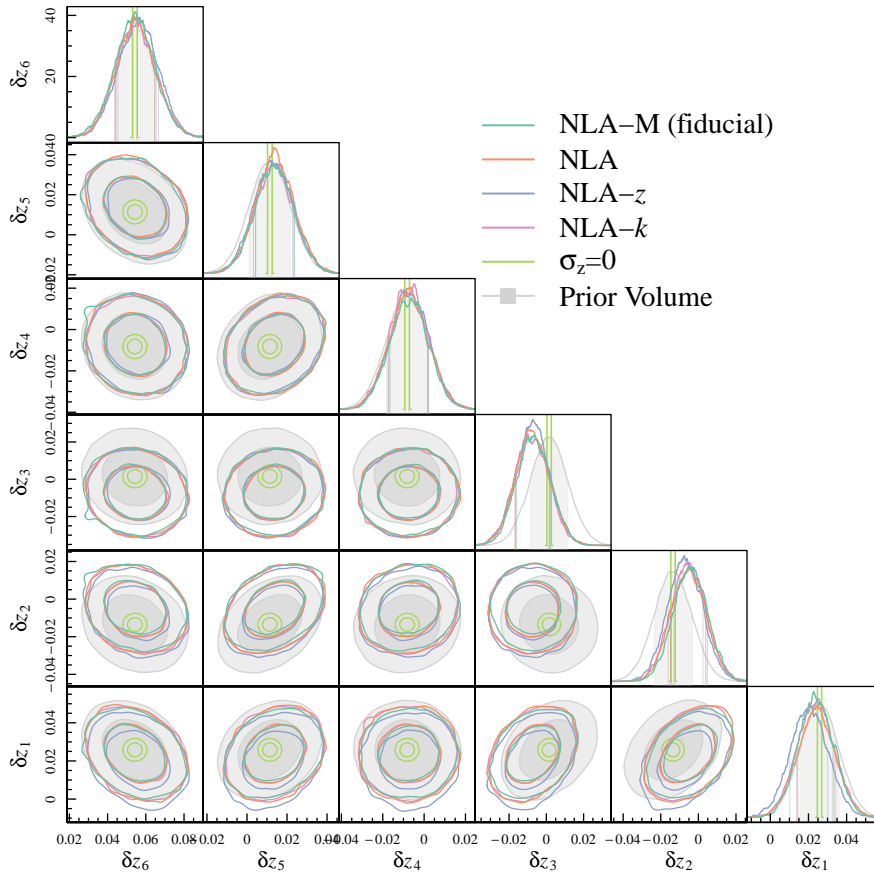
the posteriors are all fully consistent between the various analyses, demonstrating that the choice of intrinsic alignment model does not have a significant impact over the inferred redshift distribution biases.

### G.3. Cosmological parameters

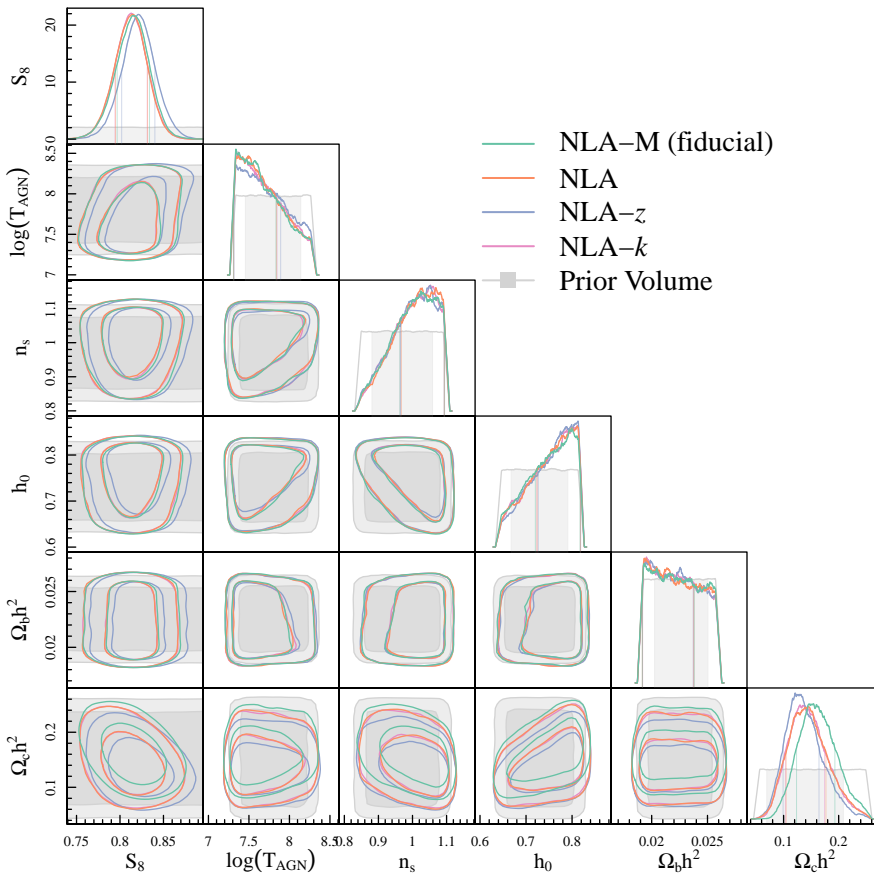
Figure G.3 shows the posterior constraints on our cosmological parameters, again for chains analysed using various intrinsic alignment models, and again compared to the prior. Only  $S_8$  and  $\Omega_m$  can be reliably considered to be constrained, although there are interesting preferences in certain parameters.



**Fig. G.1.** Whisker diagram of  $S_8$  cosmological results presented in this work. Each panel shows the compilation of results for one of our two-point statistics, as annotated. In each panel, we show the fiducial constraints in the relevant statistic with coloured points (colour-coded by the statistic), and highlight the fiducial marginal HPDI with a band of the equivalent colour. We also highlight the marginal HPDI in  $S_8$  from *Planck*-Legacy (shaded red). All variations to our analysis are annotated, and values plotted in dark blue. We annotate each analysis variation with the difference (quantified using the Gaussian approximation to the Hellinger tension) that the resulting marginal mode/HPDI has with respect to the fiducial (dark blue) or *Planck* (red).



**Fig. G.2.** Posterior constraints on our redshift distribution bias parameters for chains utilising our various intrinsic alignment models. The filled contour shows the prior derived from our SKiLLS simulations. Smoothing kernels are matched between all contours (and marginal distributions), and are shown by the non-zero width of the  $\sigma_z = 0$  contours/marginals.



**Fig. G.3.** Posterior constraints on our cosmological parameters for chains utilising our various intrinsic alignment models. The filled contour shows the prior on each cosmological parameter.

## Appendix H: Tabulated constraints

Table H.1. Constraints on  $\Sigma_8$  under our analysis variations.

Setup	Statistic	$\alpha$	$\chi^2$	dof*	PTE	PJ-HPD $N_{\text{samp}}$	Marginal Mode+HPD	Max. Apost. +PJ-HPD	Marginal Mean + CI	
$\Sigma_8$	Iteration 1	$E_n$	0.58	127.3	120.5	0.318	23	$0.821^{+0.014}_{-0.016}$	$0.825^{+0.021}_{-0.014}$	$0.820^{+0.015}_{-0.015}$
$\Sigma_8$	Iteration 1	$C_E$	0.60	151.0	162.5	0.731	18	$0.811^{+0.020}_{-0.014}$	$0.824^{+0.006}_{-0.028}$	$0.813^{+0.017}_{-0.017}$
$\Sigma_8$	Iteration 1	$\xi_{\pm}$	0.55	413.1	351.5	0.013	23	$0.823^{+0.014}_{-0.018}$	$0.829^{+0.013}_{-0.020}$	$0.821^{+0.016}_{-0.016}$
$\Sigma_8$	No Bin 1	$E_n$	0.58	89.8	84.5	0.326	31	$0.823^{+0.014}_{-0.017}$	$0.819^{+0.017}_{-0.013}$	$0.821^{+0.015}_{-0.015}$
$\Sigma_8$	No Bin 1	$C_E$	0.60	102.0	114.5	0.792	74	$0.814^{+0.019}_{-0.015}$	$0.818^{+0.020}_{-0.015}$	$0.815^{+0.017}_{-0.017}$
$\Sigma_8$	No Bin 1	$\xi_{\pm}$	0.56	282.5	249.5	0.074	33	$0.823^{+0.016}_{-0.017}$	$0.825^{+0.021}_{-0.014}$	$0.822^{+0.016}_{-0.016}$
$\Sigma_8$	No Bin 2	$E_n$	0.56	78.5	84.5	0.663	32	$0.820^{+0.015}_{-0.017}$	$0.814^{+0.030}_{-0.004}$	$0.820^{+0.015}_{-0.015}$
$\Sigma_8$	No Bin 2	$C_E$	0.60	116.4	114.5	0.433	49	$0.811^{+0.019}_{-0.016}$	$0.819^{+0.023}_{-0.016}$	$0.813^{+0.017}_{-0.017}$
$\Sigma_8$	No Bin 2	$\xi_{\pm}$	0.55	285.5	249.5	0.058	54	$0.819^{+0.020}_{-0.013}$	$0.829^{+0.019}_{-0.018}$	$0.822^{+0.017}_{-0.017}$
$\Sigma_8$	No Bin 3	$E_n$	0.59	93.6	84.5	0.234	10	$0.824^{+0.015}_{-0.017}$	$0.819^{+0.025}_{-0.008}$	$0.823^{+0.015}_{-0.016}$
$\Sigma_8$	No Bin 3	$C_E$	0.61	113.8	114.5	0.501	25	$0.818^{+0.018}_{-0.017}$	$0.828^{+0.011}_{-0.024}$	$0.818^{+0.017}_{-0.017}$
$\Sigma_8$	No Bin 3	$\xi_{\pm}$	0.56	288.8	249.5	0.044	18	$0.824^{+0.015}_{-0.017}$	$0.823^{+0.021}_{-0.012}$	$0.822^{+0.016}_{-0.016}$
$\Sigma_8$	No Bin 4	$E_n$	0.58	98.6	84.5	0.140	51	$0.824^{+0.015}_{-0.016}$	$0.820^{+0.019}_{-0.012}$	$0.823^{+0.016}_{-0.015}$
$\Sigma_8$	No Bin 4	$C_E$	0.60	109.1	114.5	0.625	5	$0.814^{+0.016}_{-0.019}$	$0.818^{+0.016}_{-0.020}$	$0.812^{+0.017}_{-0.017}$
$\Sigma_8$	No Bin 4	$\xi_{\pm}$	0.56	301.5	249.5	0.013	43	$0.822^{+0.017}_{-0.015}$	$0.825^{+0.019}_{-0.014}$	$0.823^{+0.016}_{-0.016}$
$\Sigma_8$	No Bin 5	$E_n$	0.56	99.6	84.5	0.125	31	$0.823^{+0.019}_{-0.016}$	$0.824^{+0.022}_{-0.014}$	$0.825^{+0.017}_{-0.017}$
$\Sigma_8$	No Bin 5	$C_E$	0.59	102.5	114.5	0.782	7	$0.809^{+0.019}_{-0.019}$	$0.821^{+0.016}_{-0.026}$	$0.808^{+0.019}_{-0.019}$
$\Sigma_8$	No Bin 5	$\xi_{\pm}$	0.56	296.1	249.5	0.023	25	$0.822^{+0.015}_{-0.021}$	$0.823^{+0.023}_{-0.016}$	$0.818^{+0.018}_{-0.018}$
$\Sigma_8$	No Bin 6	$E_n$	0.61	91.5	84.5	0.283	32	$0.817^{+0.019}_{-0.017}$	$0.814^{+0.020}_{-0.016}$	$0.817^{+0.018}_{-0.018}$
$\Sigma_8$	No Bin 6	$C_E$	0.59	105.5	114.5	0.715	4	$0.817^{+0.019}_{-0.023}$	$0.813^{+0.039}_{-0.010}$	$0.815^{+0.021}_{-0.020}$
$\Sigma_8$	No Bin 6	$\xi_{\pm}$	0.57	301.4	249.5	0.014	18	$0.816^{+0.018}_{-0.021}$	$0.814^{+0.033}_{-0.011}$	$0.815^{+0.019}_{-0.019}$
$\Sigma_8$	CAMB	$E_n$	0.58	128.2	120.5	0.299	38	$0.817^{+0.015}_{-0.015}$	$0.830^{+0.002}_{-0.027}$	$0.818^{+0.015}_{-0.015}$
$\Sigma_8$	CAMB	$C_E$	0.60	151.7	162.5	0.718	35	$0.812^{+0.017}_{-0.017}$	$0.825^{+0.013}_{-0.024}$	$0.812^{+0.017}_{-0.017}$
$\Sigma_8$	CAMB	$\xi_{\pm}$	0.56	411.1	351.5	0.016	23	$0.818^{+0.019}_{-0.014}$	$0.818^{+0.025}_{-0.009}$	$0.821^{+0.016}_{-0.016}$
$\Sigma_8$	Gaussian Cov	$E_n$	0.60	127.6	120.5	0.312	19	$0.820^{+0.014}_{-0.011}$	$0.823^{+0.011}_{-0.014}$	$0.821^{+0.013}_{-0.012}$
$\Sigma_8$	Gaussian Cov	$C_E$	0.61	151.9	162.5	0.714	24	$0.817^{+0.014}_{-0.016}$	$0.819^{+0.020}_{-0.013}$	$0.816^{+0.015}_{-0.015}$
$\Sigma_8$	Gaussian Cov	$\xi_{\pm}$	0.56	411.3	351.5	0.015	18	$0.825^{+0.013}_{-0.014}$	$0.826^{+0.022}_{-0.010}$	$0.825^{+0.013}_{-0.013}$
$\Sigma_8$	MICE2 $\delta_z$	$E_n$	0.60	123.2	120.5	0.415	18	$0.836^{+0.013}_{-0.012}$	$0.842^{+0.012}_{-0.015}$	$0.837^{+0.012}_{-0.012}$
$\Sigma_8$	MICE2 $\delta_z$	$C_E$	0.62	150.2	162.5	0.746	6	$0.832^{+0.015}_{-0.016}$	$0.840^{+0.009}_{-0.022}$	$0.832^{+0.015}_{-0.015}$
$\Sigma_8$	MICE2 $\delta_z$	$\xi_{\pm}$	0.57	400.3	351.5	0.037	12	$0.839^{+0.015}_{-0.013}$	$0.845^{+0.008}_{-0.019}$	$0.841^{+0.014}_{-0.014}$
$\Sigma_8$	Calib PV wgt	$E_n$	0.58	137.8	120.5	0.134	26	$0.820^{+0.015}_{-0.014}$	$0.823^{+0.017}_{-0.013}$	$0.821^{+0.014}_{-0.014}$
$\Sigma_8$	Calib PV wgt	$C_E$	0.60	149.7	162.5	0.756	42	$0.815^{+0.017}_{-0.017}$	$0.818^{+0.025}_{-0.014}$	$0.815^{+0.017}_{-0.017}$
$\Sigma_8$	Calib shape wgt	$E_n$	0.58	142.1	120.5	0.087	40	$0.822^{+0.015}_{-0.013}$	$0.829^{+0.008}_{-0.020}$	$0.822^{+0.014}_{-0.014}$
$\Sigma_8$	Calib shape wgt	$C_E$	0.60	150.2	162.5	0.746	39	$0.816^{+0.019}_{-0.016}$	$0.831^{+0.008}_{-0.028}$	$0.816^{+0.017}_{-0.017}$
$\Sigma_8$	CC Delta z	$E_n$	0.63	129.6	120.5	0.269	25	$0.831^{+0.020}_{-0.017}$	$0.831^{+0.018}_{-0.018}$	$0.832^{+0.018}_{-0.018}$
$\Sigma_8$	CC Delta z	$C_E$	0.65	152.4	162.5	0.704	11	$0.827^{+0.022}_{-0.020}$	$0.829^{+0.012}_{-0.032}$	$0.827^{+0.021}_{-0.021}$
$\Sigma_8$	CC Delta z	$\xi_{\pm}$	0.61	415.4	351.5	0.011	27	$0.826^{+0.020}_{-0.018}$	$0.828^{+0.027}_{-0.014}$	$0.827^{+0.019}_{-0.019}$
$\Sigma_8$	[0.5,300] arcmin	$E_n$	0.58	138.9	120.5	0.121	16	$0.818^{+0.014}_{-0.016}$	$0.816^{+0.013}_{-0.017}$	$0.817^{+0.015}_{-0.015}$
$\Sigma_8$	NLA	$E_n$	0.54	128.2	120.5	0.299	63	$0.810^{+0.017}_{-0.015}$	$0.816^{+0.013}_{-0.020}$	$0.811^{+0.016}_{-0.016}$

Continued on next page

Table H.1 continued.

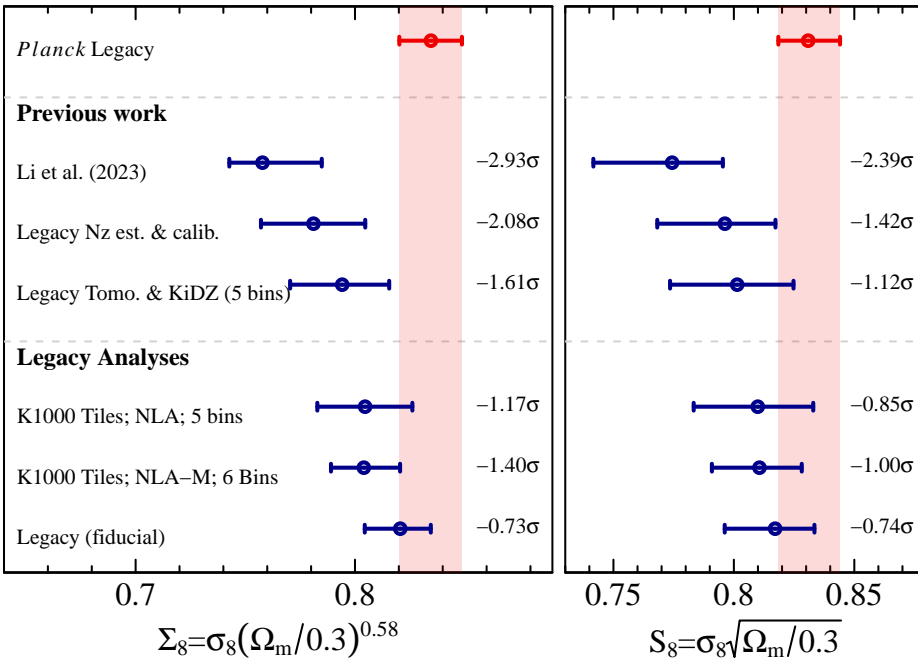
Setup	Statistic	$\alpha$	$\chi^2$	dof*	PTE	$N_{\text{samp}}^{\text{PJ-HPD}}$	Marginal Mode+HPD	Max. Apost. +PJ-HPD	Marginal Mean + CI
$\Sigma_8$ NLA	$C_E$	0.58	153.4	162.5	0.683	66	$0.809^{+0.018}_{-0.018}$	$0.810^{+0.018}_{-0.018}$	$0.808^{+0.018}_{-0.018}$
$\Sigma_8$ NLA	$\xi_{\pm}$	0.53	414.4	351.5	0.012	32	$0.821^{+0.017}_{-0.016}$	$0.817^{+0.033}_{-0.006}$	$0.820^{+0.016}_{-0.017}$
$\Sigma_8$ NLA-z	$E_n$	0.55	126.6	120.5	0.334	69	$0.819^{+0.016}_{-0.018}$	$0.817^{+0.021}_{-0.013}$	$0.819^{+0.017}_{-0.017}$
$\Sigma_8$ NLA-z	$C_E$	0.58	153.4	162.5	0.683	68	$0.817^{+0.016}_{-0.023}$	$0.813^{+0.018}_{-0.020}$	$0.814^{+0.019}_{-0.019}$
$\Sigma_8$ NLA-z	$\xi_{\pm}$	0.54	411.5	351.5	0.015	46	$0.828^{+0.016}_{-0.021}$	$0.837^{+0.018}_{-0.022}$	$0.827^{+0.018}_{-0.018}$
$\Sigma_8$ NLA-k	$E_n$	0.54	127.5	120.5	0.314	69	$0.813^{+0.014}_{-0.018}$	$0.808^{+0.021}_{-0.011}$	$0.811^{+0.016}_{-0.016}$
$\Sigma_8$ NLA-k	$C_E$	0.59	152.5	162.5	0.702	99	$0.810^{+0.016}_{-0.020}$	$0.823^{+0.019}_{-0.025}$	$0.808^{+0.018}_{-0.018}$
$\Sigma_8$ NLA-k	$\xi_{\pm}$	0.53	413.5	351.5	0.013	90	$0.819^{+0.018}_{-0.016}$	$0.826^{+0.027}_{-0.015}$	$0.820^{+0.017}_{-0.017}$
$\Sigma_8$ No Sigma <sub>z</sub>	$E_n$	0.58	131.9	120.5	0.225	40	$0.818^{+0.015}_{-0.014}$	$0.826^{+0.011}_{-0.019}$	$0.819^{+0.015}_{-0.014}$
$\Sigma_8$ No Sigma <sub>z</sub>	$C_E$	0.60	154.1	162.5	0.669	62	$0.811^{+0.019}_{-0.015}$	$0.815^{+0.020}_{-0.014}$	$0.813^{+0.017}_{-0.016}$
$\Sigma_8$ No Sigma <sub>z</sub>	$\xi_{\pm}$	0.56	418.6	351.5	0.008	41	$0.819^{+0.015}_{-0.017}$	$0.826^{+0.011}_{-0.021}$	$0.819^{+0.016}_{-0.016}$
$\Sigma_8$ No Systematics	$E_n$	0.60	133.7	122.5	0.231	135	$0.829^{+0.012}_{-0.015}$	$0.823^{+0.016}_{-0.010}$	$0.827^{+0.013}_{-0.013}$

## Appendix I: Post-unblinding analyses and changes

In this appendix we detail all analyses that were taken, and any changes that were made to the analysis and/or manuscript, after unblinding.

### I.1. Comparison between Legacy and KiDS-1000

As discussed in Sect. 6, KiDS-Legacy presents a decrease in the marginal Hellinger tension with *Planck* of  $2.2\sigma$  for  $\Sigma_8$  ( $1.65\sigma$  for  $S_8$ ). After the unblinding of KiDS-Legacy we investigated the differences between the Legacy and KiDS-1000 samples and analyses, to determine whether the change in cosmological parameters could be attributed to systematic or random variations. To do this, we reanalysed the KiDS-1000 and KiDS-Legacy shape catalogues with a series of analysis variations, which changed the sample used for cosmological inference and/or the analysis choices that were used in the inference. This required us, in each case, to rerun our  $N(z)$  estimation, our correlation function measurement, covariance computation, and posterior inference. We did not, however, remeasure the redshift distribution bias parameters or the multiplicative shape measurement biases for every setup.



**Fig. I.1.** Variation in recovered  $E_n$  cosmological constraints for  $\Sigma_8$  (left) and  $S_8$  (right) when moving between KiDS-1000 and KiDS-Legacy analyses, relative to the constraints from *Planck* (red). Each point is annotated with the Hellinger tension between the marginal constraint and that of *Planck*.

Figure I.1 presents a graphical summary of the changes we found in our recovered cosmological constraints when analysing KiDS-1000 using Legacy analysis choices and calibration methods, and when analysing Legacy using subsets defined by the KiDS-

1000 footprint and photo- $z$  baseline. Our investigation of the consistency between the data from Li et al. (2023a) and Legacy was focussed on two areas: the impact of Legacy analysis choices and calibration methods, and the impact of the new on-sky area available to Legacy. We quantified the effect of each analysis variation by recomputing the Hellinger tensions with respect to *Planck*, as this allows us to circumvent the complex covariance between various analyses of KiDS-1000 and KiDS-Legacy.

As a result of this investigation, we found no evidence of statistically significant systematic biases in the sample and analysis of KiDS-Legacy. We determined that the primary differences between the cosmological results of Li et al. (2023a) and Legacy are attributable to changes to our redshift calibration sample and methodology, and to statistical noise driven by the new area available to KiDS-Legacy outside the KiDS-1000 footprint. We have made no changes to the analysis methodology of Legacy as presented in the main text after this investigation, and thus conclude that this suite of tests do not represent a possible source of bias in our blinded analysis.

### I.1.1. Analysing KiDS-1000 with Legacy calibration methods

We started by investigating the impact of the updated redshift estimation and calibration methods of Wright et al. (2025) on the cosmological constraints of Li et al. (2023a). To do this we reran the entirety of the cosmic shear measurement from KiDS-1000, from raw shape catalogues to cosmological inference, incorporating our new redshift estimation and calibration methodologies (leveraging e.g. gold-weight and simulation construction with sample matching). In this analysis we computed revised redshift bias estimates using the original KiDS-1000 redshift calibration sample, but updating to SKILLS image simulations and gold-weight redshift distribution estimation. The result of this end-to-end reanalysis are provided in Fig. I.1 as ‘Legacy Nz est. & calib.’. We found that the updated redshift distribution calibration method brings the cosmological estimates from the Li et al. (2023a) sample into better agreement with *Planck*, at  $2.08\sigma$  consistency in  $\Sigma_8$  and  $1.42\sigma$  consistency in  $S_8$ . Importantly, both of these measurements satisfy the typical KiDS threshold for consistency of  $2.3\sigma$  (equivalent to a one-sided probability to exceed of 0.01).

### I.1.2. Updating KiDS-1000 to the Legacy calibration sample

We also measured the contribution to the difference between KiDS-1000 and KiDS-Legacy caused by the differences to our analysis choices and calibration sample, including our new tomographic bin limits and our greatly expanded redshift calibration dataset. To do this we again performed a full end-to-end reanalysis of KiDS-1000, this time with our new tomographic bin limits (excluding the sixth bin), and new redshift calibration sample from KiDZ. In this analysis we assume the redshift distribution biases and multiplicative shear biases are the same as the fiducial KiDS-Legacy analysis. We find that the use of the larger KiDZ calibration sample also slightly reduces the offset with *Planck*, to  $1.61\sigma$  consistency in  $\Sigma_8$  and  $1.12\sigma$  consistency in  $S_8$  (‘Legacy Tomo. & KiDZ (5 bins)’). This demonstrates that the KiDS-1000 sample, when analysed with our larger spectroscopic calibration sample and new redshift calibration methods (but without adding higher redshift sources, or using updated imaging, or new photo- $z$ , or additional area) is consistent with the results from *Planck* and with the results from Legacy.

### I.1.3. Legacy analysis in the KiDS-1000 volume

We explored the consistency between the constraints from KiDS-1000 and KiDS-Legacy by analysing the two samples in the same volume, and with the same analysis choices. We were able to do this fairly easily thanks to our analysis from Sect. I.1.2, which homogenised the tomography and  $N(z)$  estimation and calibration between the two samples. By restricting KiDS-Legacy to only the footprint of KiDS-1000, excluding the sixth tomographic bin, and using the NLA IA model, we were able to perform a comparison that is mostly free of statistical fluctuations. In fact there are still differences in the source samples due to changes in source extraction, image masking, source selection, and more, which we assumed add only a small amount of statistical noise to the comparison. We find that the Legacy sample in the same volume as KiDS-1000 is slightly closer to *Planck* than KiDS-1000, with a  $1.17\sigma$  consistency in  $\Sigma_8$  and  $0.85\sigma$  consistency in  $S_8$ . This indicates that the sum total of changes (such as updated imaging calibrations, photo- $z$ , masking, etc.) lead to a  $0.44\sigma$  ( $0.27\sigma$ ) decrease in the Hellinger distance between *Planck* and KiDS; significantly less than the change introduced by our new calibration methods.

### I.1.4. Sixth tomographic bin and NLA- $M$

We further investigated the influence of adding the sixth tomographic bin and our new NLA- $M$  IA model to the analysis of Legacy in the footprint of KiDS-1000 (‘K1000 Tiles; NLA- $M$ ; 6 bins’), finding that these changes caused a significant reduction in our marginal uncertainties (as expected) but did not lead to a significant change in the estimated cosmological parameters. As such, our Hellinger distance with *Planck* increases with these additions, to  $1.40\sigma$  consistency in  $\Sigma_8$  and  $1.00\sigma$  consistency in  $S_8$ .

### I.1.5. Additional Legacy area

Finally, we added the data back into Legacy that resides outside the footprint of KiDS-1000, and recover our fiducial result, which exhibits a  $0.73\sigma$  consistency in  $\Sigma_8$  and  $0.74\sigma$  consistency in  $S_8$ . This demonstrates that there is a non-negligible statistical noise component in the difference between the sources probed within the footprint of KiDS-1000 and over the final KiDS-Legacy area.

### 1.2. Expanded scale cuts

Post unblinding, we discovered a typographical error in the pipeline which performed the measurement of correlation functions with expanded scale cuts (Sect. 5.2.7). This error meant that the pre-unblinding measurements were made with uncalibrated shapes directly from our shape measurement code *lensfit*, rather than with the recalibrated shapes (Sect. 3.5). We reran the expanded scale cut analysis with the correct recalibrated shapes post-unblinding, finding that the consistency between the analyses with fiducial and expanded scale cuts improved from  $0.3\sigma$  to  $0.15\sigma$  with the use of the correct shapes. No conclusions were changed as a result of this error.

### 1.3. Iterative covariances

All fiducial chains were recomputed using our iterative covariance framework post-unblinding. This was a conscious pre-unblinding choice, as the iterative covariances require non-negligible CPU time for the computation of MAP and new covariances. We found that this process had a negligible impact on our constraints (see Sect. 5.2.3), and thus does not represent a possible source of bias in our blinded analysis.