Astronomy & Astrophysics manuscript no. output ©ESO 2025
March 25, 2025

KiDS-Legacy: Consistency of cosmic shear measurements and joint cosmological constraints with external probes

Benjamin Stölzner[1⋆], Angus H. Wright[1], Marika Asgari[2], Catherine Heymans[1,3], Hendrik Hildebrandt[1], Henk Hoekstra[4], Benjamin Joachimi[5], Konrad Kuijken[4], Shun-Sheng Li[1,4], Constance Mahony[22,23,1], Robert Reischke[1,6], Mijin Yoon[4], Maciej Bilicki[19], Pierre Burger[7,8,6], Nora Elisa Chisari[9,4], Andrej Dvornik[1], Christos Georgiou[21], Benjamin Giblin[3], Joachim Harnois-Déraps[2], Priyanka Jalan[19], Anjitha John William[19], Shahab Joudaki[10,11], Giorgio Francesco Lesci[13,14], Laila Linke[12], Arthur Loureiro[24,25], Matteo Maturi[26,27], Lauro Moscardini[13,14,15], Nicola R. Napolitano[28,29,30], Lucas Porth[6], Mario Radovich[16], Tilman Tröster[20], Maximilian von Wietersheim-Kramsta[17,18], Anna Wittje[1], Ziang Yan[1], and Yun-Hao Zhang[3,4]

(Affiliations can be found after the references)

Received ; accepted

ABSTRACT

We present a cosmic shear consistency analysis of the final data release from the Kilo-Degree Survey (KiDS-Legacy). By adopting three tiers of consistency metrics, we compare cosmological constraints between subsets of the KiDS-Legacy dataset split by redshift, angular scale, galaxy colour and spatial region. We also review a range of two-point cosmic shear statistics. With the data passing all our consistency metric tests, we demonstrate that KiDS-Legacy is the most internally consistent KiDS catalogue to date. In a joint cosmological analysis of KiDS-Legacy and DES Y3 cosmic shear, combined with data from the Pantheon+ Type Ia supernovae compilation and baryon acoustic oscillations from DESI Y1, we find constraints consistent with *Planck* measurements of the cosmic microwave background with $S_8 \equiv \sigma_8 \sqrt{\Omega_\mathrm{m}/0.3} = 0.814^{+0.011}_{-0.012}$ and $\sigma_8 = 0.802^{+0.022}_{-0.018}$.

Key words. gravitational lensing: weak – cosmology: observations – large-scale structure of Universe – cosmological parameters – methods: statistical

Contents

1. Introduction

In the current era of precision cosmology, weak lensing surveys have probed the standard Λ-cold dark matter (ΛCDM) cosmological model to unprecedented precision. The weak gravitational lensing effect causes distortions in the shape of galaxy images, known as cosmic shear. This allows mapping the distribution of gravitating matter along the line of sight, which is sensitive to the shape and amplitude of the matter power spectrum. Since its first detection (Kaiser et al. 2000; Wittman et al. 2000; Van Waerbeke et al. 2000; Bacon et al. 2000), cosmic shear has become a primary cosmological probe for imaging galaxy surveys. Recent analyses from the three current stage-III weak lensing surveys, the ESO Kilo-Degree Survey (KiDS; Kuijken et al. 2015), the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al. 2016), and the Subaru Hyper Suprime Cam Subaru Strategic Program (HSC; Aihara et al. 2018), have showcased the potential of cosmic shear measurements as a probe of the cosmological model, making it a main science driver for upcoming galaxy surveys conducted with the Vera C. Rubin Observatory (Ivezić et al. 2019), the *Euclid* satellite (Euclid Collab-

⋆ stoelzner@astro.rub.de

oration et al. 2024), the Nancy Grace Roman Space Telescope (Spergel et al. 2015), and the Chinese Space Station Telescope (Gong et al. 2019).

Weak lensing studies mainly constrain the structure growth parameter $S_8 = \sigma_8 \sqrt{\Omega_m/0.3}$, which combines the matter density parameter Ω_m and the standard deviation of matter density perturbations in spheres of $8\,h^{-1}$Mpc radius, σ_8. Current cosmic shear measurements have yielded S_8 values that are lower (Asgari et al. 2021; Heymans et al. 2021; Abdalla et al. 2022; Amon et al. 2022; Secco et al. 2022; Dalal et al. 2023; Li et al. 2023b; Dark Energy Survey and Kilo-Degree Survey Collaboration et al. 2023) than values derived from observations of the cosmic microwave background (Planck Collaboration et al. 2020) at a $\sim 2\sigma$ level. However, there is no consensus about whether this feature, which is commonly referred to as the 'S_8 tension', is a result of systematics in the data analysis or theoretical modelling, statistical fluctuations, or effects beyond the standard flat ΛCDM model.

Given the unclear nature of the apparent S_8 tension, recent works have studied probes of the late Universe, finding consistency between independent cosmic shear surveys (Amon & Efstathiou 2022; Amon et al. 2023; Longley et al. 2023; Dark Energy Survey and Kilo-Degree Survey Collaboration et al. 2023). In addition to tests of the consistency between independent data sets, tests of the internal consistency of a given data set are of particular importance in order to rule out systematic effects within one data set as the source of the inconsistency between different data sets (Efstathiou & Lemos 2018; Köhlinger et al. 2019; Raveri et al. 2020; Li et al. 2021).

This work is part of a series of KiDS-Legacy papers. The production process of all data products in the fifth KiDS data release (DR5), including shape measurements and the KiDS-Legacy sample selection, is presented in Wright et al. (2024). The calibration of the photometric redshift distribution is described in Wright et al. (2025a) and multi-band image simulations enabling a joint shear and redshift calibration are presented in Li et al. (2023a). The modelling of the covariance for the three main summary statistics is summarised in Reischke et al. (2024) and the angular clustering of KiDS-Legacy galaxies is analysed in Yan et al. (2025). The fiducial cosmic shear analysis is conducted in Wright et al. (2025b, hereafter W25). In this work, we perform several internal consistency tests of the KiDS-Legacy data, focusing on their impact on cosmological constraints inferred in the cosmic shear analysis. In particular, we split the KiDS-Legacy data set into various subsets based on redshift, spatial region, angular scale, and colour. We then perform a split cosmological analysis by modelling the observed data in each subset with a separate set of cosmological parameters. By evaluating several consistency metrics, we quantify the level of agreement between the data subsets. As established in the previous KiDS-1000 analysis (Asgari et al. 2021), constraints on cosmology are inferred from three different two-point statistics (COSEBIs, band powers, and two-point correlation functions). Here, we quantify the internal consistency between summary statistics.

While current cosmic shear surveys cannot constrain both Ω_m and σ_8 separately, the addition of external data allows for a breaking of the degeneracy. For this purpose we employ data from recent measurements of baryon acoustic oscillations (BAOs), redshift space distortions (RSDs), and Type Ia supernovae (SN Ia), which place tight constraints on the matter density. We perform a consistency test and a joint cosmological analysis of KiDS-Legacy data combined with data from the recent DESI Y1 BAO analysis (Adame et al. 2025), the earlier eBOSS (Alam et al. 2021) BAO and RSD analysis, and the Pantheon+ SN Ia compilation (Scolnic et al. 2022; Brout et al. 2022). Additionally, we conduct a joint analysis with DES Y3 cosmic shear data (Amon et al. 2022; Secco et al. 2022) and quantify the consistency between KiDS-Legacy and *Planck* CMB constraints (Planck Collaboration et al. 2020).

This paper is structured as follows. In Sect. 2 we provide a brief summary of the KiDS-Legacy data and external data employed in this work. In Sect. 3 we review the theoretical model for weak lensing observables and discuss the metrics quantifying the internal consistency of the data. In Sect. 4 we present the results of the internal consistency tests. We provide the results of our joint cosmological constraints with external data in Sect. 5 and conclude in Sect. 6. Appendix A summarises the data properties of the KiDS-Legacy catalogue divided into sub-samples. In Appendix B we provide details about the estimate of the effective number of constrained parameters in our analysis.

2. Data

2.1. KiDS-Legacy

The Kilo-Degree Survey (Kuijken et al. 2015; de Jong et al. 2015, 2017; Kuijken et al. 2019; Wright et al. 2024) is a public survey conducted by the European Southern Observatory with the VLT Survey Telescope (VST). The KiDS and the complementary VISTA Kilo-Degree Infrared Galaxy Survey (VIKING; Edge et al. 2013) combine optical and near-infrared imaging in nine photometric bands. In this work, we analyse weak lensing data from the fifth and final data release (DR5). Here, we provide a brief summary of the survey and refer the reader to Wright et al. (2024) for the details.

The KiDS-DR5 data set consists of imaging data covering an area of 1347 deg^2 on sky divided into two distinct stripes across the celestial equator in the North Galactic Cap and across the South Galactic Pole, respectively. All sources are observed in four optical bands (u, g, r, and i) with the VST as well as in five near-infrared bands (Z, Y, J, H, and K_s) from VIKING. In comparison to the previous data release, KiDS-DR5 features a second pass of i-band observations and an increase of 34% in survey area. The lensing sample is obtained via a masking process, selecting sources with high-quality data in all photometric bands, and applying a sequence of cuts on magnitude, colour, and lensing-related quantities. This sample, dubbed KiDS-Legacy, contains approximately 43 million sources on 967 deg^2 of sky, corresponding to an effective number density $n_{\rm eff} = 8.79\,\mathrm{arcmin}^{-2}$.

Photometric redshift estimates of KiDS-Legacy sources are computed via the Bayesian Photo-z code (BPZ; Benítez 2000). The deeper i-band depth and a significantly larger spectroscopic calibration data set allow for a higher photometric redshift limit of $z_B = 2$ compared to previous KiDS analyses, enabling the addition of another redshift bin. The sources in the KiDS-Legacy lensing sample are divided into six approximately equi-populated bins via their photometric redshift estimates. The redshift distributions of the resulting bins are inferred via a direct calibration with deep spectroscopic surveys using self-organising maps (SOMs; Lima et al. 2008; Masters et al. 2015; Wright et al. 2020a). While this calibration method was utilised in earlier KiDS analyses (Wright et al. 2020b; Hildebrandt et al. 2021), the KiDS-Legacy redshift calibration features several improvements such as the use of one SOM per tomographic bin instead of one overall SOM, a selection of sources via SOM-derived gold weights, and accounting for prior volume effects. Furthermore, the SOM redshift distributions are calibrated with the multicolour SKiLLS simulations (Li et al. 2023a). For a detailed discussion of the improved calibration method, we refer the reader to the KiDS-Legacy redshift calibration manuscript (Wright et al. 2025a).

Shape measurements in KiDS-Legacy were performed with an updated version of the LENSFIT algorithm (Miller et al. 2013; Fenech Conti et al. 2017) and calibrated with the SKiLLS image simulations as established in Li et al. (2023a). The shape measurements were validated with a series of systematics tests as outlined in W25.

2.2. External data

We employ several external data sets featuring measurements of BAOs, RSDs, and SN Ia to quantify the consistency of the KiDS results and infer joint constraints with the KiDS-Legacy cosmic shear data. In this section, we briefly describe the external data sets. In this work, we use publicly available likelihoods, which are implemented in the CosmoSIS standard library [1].

We make use of BAO measurements from the first data release of the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016, 2022) survey. In particular, DESI targets four different classes of extragalactic objects: a bright galaxy sample (BGS; Hahn et al. 2023), luminous red galaxies (LRGs; Zhou et al. 2023), emission line galaxies (ELGs; Raichoor et al. 2023), and quasars (QSOs; Chaussidon et al. 2023). Cosmological results from DESI BAO measurements were presented in Adame et al. (2025). The DESI likelihood provides BAO measurements from the four DESI sub-samples, covering a wide range of redshifts. In particular, we employ measurements from the BGS ($0.1 < z < 0.4$), two LRG samples ($0.4 < z < 0.6$ and $0.6 < z < 0.8$, respectively), an ELG sample ($1.1 < z < 1.6$), a combined LRG and ELG sample ($0.8 < z < 1.1$), a QSO sample ($0.8 < z < 2.1$), and a Lyman-α forest sample ($1.77 < z < 4.16$).

As an alternative to recent BAO measurements from DESI, we employ data from the Sloan Digital Sky Survey's (SDSS) Baryon Oscillation Spectroscopic Survey (BOSS) and Extended Baryon Oscillation Spectroscopic Survey (eBOSS). This data set provides BAO measurements as well as measurements of RSDs. The likelihoods provide constraints from the SDSS DR7 main galaxy sample (Ross et al. 2015; Howlett et al. 2015), BOSS DR12 (Alam et al. 2017), eBOSS DR16 ELGs (Tamone et al. 2020; Raichoor et al. 2021; de Mattia et al. 2021), eBOSS DR16 LRGs (Bautista et al. 2021; Gil-Marín et al. 2020), eBOSS DR16 QSOs (Neveux et al. 2020; Hou et al. 2021) and the eBOSS DR16 Lyman-α forest (du Mas des Bourboux et al. 2020).

Additionally, we employ measurements of Type Ia supernovae from the Pantheon+ compilation (Scolnic et al. 2022). This data set consists of 1701 light curves of 1550 spectroscopically confirmed SN Ia with redshifts $z \in (0.001, 2.26)$. Cosmological constraints from this data set were presented in Brout et al. (2022).

We employ cosmic shear measurements from the DES (The Dark Energy Survey Collaboration 2005; Dark Energy Survey Collaboration et al. 2016; Flaugher et al. 2015). In particular, we make use of the 'KiDS-excised' DES data vector presented in Dark Energy Survey and Kilo-Degree Survey Collaboration et al. (2023, herafter DES+KiDS), which is based on the analysis of the DES Y3 cosmic shear measurements (Amon et al. 2022; Secco et al. 2022) and excludes 8% of DES data in the overlap region between KiDS and DES. Following the methodology of this study, we neglect the cross-covariance between the two surveys, which was shown to be sufficiently small. Furthermore, we adopt the 'ΛCDM-optimised' angular scale cuts of Amon et al. (2022) and Secco et al. (2022).

Finally, we adopt CMB measurements from Planck Collaboration et al. (2020). Here, we make use of the compressed *Planck* likelihood of Prince & Dunkley (2019), which approximates the $\ell < 30$ temperature likelihood by two Gaussian data points and employs the *Planck plik-lite* TTTEEE likelihood for $\ell > 30$. Following the methodology outlined in this work, we impose a Gaussian prior on the optical depth to reionisation τ, which is derived from base ΛCDM parameter constraints from *Planck*.

3. Methodology

3.1. Cosmic shear model

Our cosmic shear analysis pipeline is based on W25 and is implemented in the public CosmoPipe[2] infrastructure. In this section, we summarise the theoretical modelling of the cosmic shear signal. We employ three summary statistics as established in the previous KiDS-1000 analysis (Asgari et al. 2021). In addition to real space shear two-point correlation functions (2PCFs), which are commonly used in cosmic shear studies, we make use of two additional summary statistics which are derived from 2PCF measurements. First, we compute complete orthogonal sets of E/B-integrals (COSEBIs; Schneider et al. 2010; Asgari et al. 2012), which provide a clean separation of E- and B-modes. This is of particular advantage since for current surveys we expect only the E modes to carry the cosmic shear signal to first order, allowing for the B modes to be used as a null test for residual systematics. Second, we employ band power spectra inferred from correlation functions (Schneider et al. 2002; Becker & Rozo 2016; van Uitert et al. 2018). This statistic allows for an approximate separation of E- and B-modes and follows the underlying angular power spectra. In this work, we select COSEBIs as our fiducial statistic when reporting analysis results, following the choice made in W25.

In general, we model the signal for each summary statistic S via a linear transformation of the cosmic shear power spectrum $C_{\epsilon\epsilon}^{(ij)}$:

$$S^{(ij)} = \int_0^\infty \mathrm{d}\ell\, C_{\epsilon\epsilon}^{(ij)}(\ell) W_S(\ell)\,, \tag{1}$$

where $W_S(\ell)$ is a weight function depending on the angular scale and the summary statistic itself. The cosmic shear power spectrum is given by the sum of the gravitational lensing power spectrum (GG), the intrinsic alignment of galaxies (II), and the corresponding cross term (GI):

$$C_{\epsilon\epsilon}^{(ij)}(\ell) = C_{\mathrm{GG}}^{(ij)}(\ell) + C_{\mathrm{II}}^{(ij)}(\ell) + C_{\mathrm{GI}}^{(ij)}(\ell) + C_{\mathrm{IG}}^{(ij)}(\ell)\,. \tag{2}$$

Under the assumption of the extended Limber approximation (Kaiser 1992; LoVerde & Afshordi 2008; Kilbinger et al. 2017) the gravitational lensing power spectrum can be written as

$$C_{\mathrm{GG}}^{(ij)}(\ell) = \int_0^{\chi_\mathrm{H}} \mathrm{d}\chi\, \frac{W_\mathrm{G}^{(i)}(\chi) W_\mathrm{G}^{(j)}(\chi)}{f_\mathrm{K}^2(\chi)} P_{\mathrm{m,nl}}\left(\frac{\ell + 1/2}{f_\mathrm{K}(\chi)}, z(\chi)\right)\,, \tag{3}$$

where $P_{\mathrm{m,nl}}$ denotes the non-linear matter power spectrum and f_K, χ, and χ_H are the comoving angular diameter distance, the

[1] `https://github.com/joezuntz/cosmosis-standard-library/tree/v4.0/likelihood`

[2] `https://github.com/AngusWright/CosmoPipe/tree/KiDSLegacy_CosmicShear`

comoving radial distance, and the comoving horizon distance, respectively. The weak lensing kernel $W_{\rm G}^{(i)}(\chi)$ is given, for example, in eq. 2 in W25. Furthermore, the underlying galaxy sample is divided into tomographic bins using estimates of the photometric redshift, allowing for an increase in constraining power (Hu 1999). Therefore, Eq. (3) refers to the cross cosmic shear signal between combinations of tomographic bins i and j, where the probability distribution of comoving distances of galaxies per bin enters the window function $W_{\rm G}^{(i)}(\chi)$.

In W25 we explore a range of models of the intrinsic alignment (IA) of galaxies, finding no significant impact of the IA model on the S_8 constraints. In this work, we therefore employ the fiducial mass-dependent IA model, dubbed NLA-M. This model extends the non-linear linear alignment model (NLA; Bridle & King 2007), incorporating an alignment of red, early-type galaxies, while assuming zero alignment of blue, late-type galaxies. The fraction of early-type galaxies is inferred by selecting galaxies with spectral type $T_{\rm B} < 1.9$. The spectral type is inferred with BPZ, which uses a set of six model templates of the spectral energy distribution (SED), ordered approximately by star formation activity, and determines the best-fitting SED by interpolating between templates. The cut on $T_{\rm B}$ selects galaxies with contributions of an elliptical galaxy spectrum. We model the IA power spectrum of red galaxies as a power law dependent on the average halo mass within a tomographic bin (see eq. 11 in W25). In this model, we employ two nuisance parameters, $A_{\rm IA}$ and β, which parameterise the IA amplitude and the slope of the IA mass scaling, respectively. We adopt the joint posterior on $A_{\rm IA}$ and β from Fortuna et al. (2025) as prior in our analysis, which we approximate by a bivariate Gaussian distribution. Additionally, we employ a multivariate Gaussian prior on the halo mass per tomographic bin. For a detailed description of the IA model, we refer the reader to section 2.3.4 and appendix B in W25.

3.2. Consistency metrics

To assess the internal consistency of the KiDS-Legacy data set, we follow the methodology of Köhlinger et al. (2019) and subdivide the data in many ways before analysing the subsets jointly. In particular, we apply splits at the data vector level by redshift and angular scale and at the catalogue level by spatial region and colour. Additionally, we construct a joint data vector of different summary statistics. For each split, we analyse the cosmic shear data with two modelling setups:

1. Fiducial cosmological model: one set of parameters models the full data set.
2. Split cosmological model: two sets of parameters model two mutually exclusive (but generally correlated) subsets of the data.

For splits at the data vector-level the first setup is equivalent to the fiducial cosmic shear analysis setup of W25. For splits at the catalogue level and splits by summary statistic we construct a different data vector, whose information content may differ from the fiducial data vector: for example, the red-blue split excludes shape correlations between red and blue galaxies. The second modelling setup features two independent sets of cosmological parameters, which allow us to assess whether or not different subsets of the data prefer different cosmologies. A list of cosmological and nuisance parameters is given in Table 1. The cosmic shear analysis features a number of nuisance parameters which are marginalised over, assuming Gaussian or top-hat priors. Since the posterior distribution of the nuisance parameters is entirely driven by the prior (see W25) we do not duplicate these parameters in the split cosmological analysis, but instead keep them shared between both data subsets. An exception is the colour-based split, as discussed in Sect. 4.2.2.

In practice, we conduct a likelihood analysis for each data split employing both the fiducial model and the split cosmological model. We then evaluate various consistency metrics, testing if the split cosmological model is preferred over the fiducial model. There exists a variety of statistical tools allowing for model comparison and an estimation of the significance of the preference of a specific model. These can be grouped into techniques that compress the full likelihood or posterior into a single summary statistic, parameter-space methods that focus on differences in single or multiple model parameters, and methods that quantify differences in data vector space. In this work, we perform three tiers of consistency tests, following the nomenclature of Köhlinger et al. (2019). We note that only the tier 1 test requires a cosmological analysis with the fiducial model since it performs a model comparison. The remaining tiers focus on the analysis with the split cosmological model in order to quantify the consistency between data subsets.

3.2.1. Tier 1: Evidence-based metric

The first tier of consistency tests includes tests compressing the full likelihood or posterior into a single summary statistic. A widely used example of such a metric is the Bayes ratio, which in logarithmic form is given by the difference between Bayesian evidences

$$\log_{10} R = \log_{10} \mathcal{Z}_{\rm fiducial} - \log_{10} \mathcal{Z}_{\rm split} , \tag{4}$$

where $\mathcal{Z}_{\rm fiducial/split}$ denotes the Bayesian evidence of the two models which, in general, can be computed by integrating the product of the likelihood $\mathcal{L}$ and the prior π

$$\mathcal{Z} = \int {\rm d}\boldsymbol{\theta}\, \mathcal{L}(\boldsymbol{\theta})\pi(\boldsymbol{\theta}) , \tag{5}$$

over the parameters $\boldsymbol{\theta}$. We note that for two independent data sets A and B, the evidence for the split model simplifies to $\mathcal{Z}_{\rm split} = \mathcal{Z}_{\rm A}\mathcal{Z}_{\rm B}$. However, this equality does not hold for correlated data sets, i.e. for most data splits considered in this work. Therefore, the evidence needs to be computed from the joint posterior distribution of the two subsets, modelled with two sets of parameters, and taking the cross-correlation between subsets into account.

In general, values of $\log_{10} R > 0$ corresponds to preference for the fiducial model while $\log_{10} R < 0$ indicates preference for the split model. The Bayes ratio is usually interpreted using Jeffreys' scale (Jeffreys 1939), which provides limits for the degree of preference for a specific model. This scale associates values of $|\log_{10} R| > [\frac{1}{2}, 1, 2]$ with 'substantial', 'strong', and 'decisive' preference for the specific model, respectively, but this lacks a clear motivation and there is no consensus on when to report tension between models. Additionally, the Bayes ratio suffers from a prior dependence which is suboptimal when using wide, uninformative priors. This makes the Bayes ratio particularly suboptimal for the analysis presented in this work, since it involves a duplication of the parameter space and the corresponding prior volume.

To circumvent these issues, Handley & Lemos (2019b) proposed the so-called suspiciousness S:

$$\ln S = \ln R - \ln I , \tag{6}$$

Table 1. Model parameters and their priors.

Type	Parameter	Prior	Duplicated	Description
Cosmological	$\omega_{\rm cdm}$	$\mathcal{U}(0.051, 0.255)$	Yes	Reduced cold dark matter density
	$\omega_{\rm b}$	$\mathcal{U}(0.019, 0.026)$	Yes	Reduced baryon density
	h	$\mathcal{U}(0.64, 0.82)$	Yes	Reduced Hubble parameter
	$n_{\rm s}$	$\mathcal{U}(0.84, 1.1)$	Yes	Spectral index of the primordial power spectrum
	S_8	$\mathcal{U}(0.5, 1.0)$	Yes	Structure growth parameter
Nuisance	$\log_{10} T_{\rm AGN}$	$\mathcal{U}(7.3, 8.3)$	Yes	Baryon feedback parameter
	$A_{\rm IA}$	$\mathcal{N}(5.74, \mathbf{C}_{A_{\rm IA},\beta})$	No / red-blue split only	Amplitude of intrinsic galaxy alignments for red galaxies
	β	$\mathcal{N}(0.44, \mathbf{C}_{A_{\rm IA},\beta})$	No	Slope of the mass scaling of intrinsic galaxy alignments
	$\log_{10} M_{\rm i}$	$\mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_M)$	No	Mean halo mass of early-type galaxies per tomographic bin i
	$\delta_{{\rm z},i}$	$\mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_z)$	No / red-blue split only	Shift of the mean of the redshift distribution per tomographic bin i

Notes. The first two columns provide parameter names and types of the sampling parameters. The third column lists the adopted prior with uniform priors denoted by $\mathcal{U}$ and Gaussian priors denoted by $\mathcal{N}$. The fourth column indicates whether or not the parameter is duplicated in the split cosmological analysis. The fifth column provides a brief parameter description. The priors on $A_{\rm IA}$ and β are centred at the mean of the posterior reported by Fortuna et al. (2025) including a correlated prior with correlation coefficient $r = -0.59$. The prior on $M_{\rm i}$ is inferred from halo mass measurements as discussed in appendix B of W25. The δ_z priors are centred at the mean of the shift of the redshift distribution and correlated through a covariance matrix estimated following the methodology of Wright et al. (2025a).

where the information ratio $\ln I = \mathcal{D}_{\rm split} - \mathcal{D}_{\rm fiducial}$ is defined through the Kullback-Leibler divergence $\mathcal{D}$ (Kullback & Leibler 1951), which quantifies the information gain between the prior and the posterior. The suspiciousness is designed to remove the effect of the prior volume from the Bayes ratio. As discussed in Handley & Lemos (2019b), it is insensitive to the choice of prior as long as the prior does not affect the shape of the posterior distribution. Under the assumption of Gaussian posteriors, a tension probability can be identified via the quantity $d - 2\ln S$. Here, d denotes the difference between the effective number of constrained parameters by the two models

$$d = N_\Theta^{\rm split} - N_\Theta^{\rm fiducial} \,, \tag{7}$$

with the effective number of free parameters constrained by the posterior distribution, N_Θ. We note that the suspiciousness can be rephrased in terms of the expectation value of the log-likelihood (see appendix G.3 in Heymans et al. 2021) and therefore does not strictly require the computation of the evidence.

In a prior-informed analysis with correlated sampling parameters, such as the cosmic shear analysis in this work, this quantity is smaller than the number of free parameters and is non-trivial to determine. There exist several estimators, such as the Bayesian model complexity (BMC, Spiegelhalter et al. 2002) and the Bayesian model dimensionality (BMD, Handley & Lemos 2019a). However, Joachimi et al. (2021) report that commonly used dimensionality measures in general are biased estimators of the effective number of parameters. As an alternative, they proposed an estimation via χ^2 minimisation of a set of mock data vectors, which on average was found to reproduce an unbiased estimate of the true value of N_Θ. Therefore, we adopt this strategy as the fiducial method in the analysis. For comparison, we additionally compute the BMD given by

$$N_\Theta/2 = \langle \ln \mathcal{L}^2 \rangle_P - \langle \ln \mathcal{L} \rangle_P^2 \,, \tag{8}$$

where $\langle\rangle_P$ denotes the average over the posterior. This quantity can be directly obtained as a byproduct of common posterior sampling algorithms, such as MCMC or nested sampling, and therefore is available at no additional computational cost. For Gaussian posteriors, the tension probability inferred from the suspiciousness statistic is then determined by

$$p_{\rm t} = \int_{d-2\ln S}^{\infty} {\rm d}x\, \chi_d^2(x) \,, \tag{9}$$

where $\chi_d^2(x)$ denotes the probability density function of a d-dimensional χ^2-distribution. The corresponding number of sigma can then be inferred from the tension probability via

$$N_{\sigma,S} = \sqrt{2}\, {\rm erf}^{-1}\,(1 - p_{\rm t}) \,. \tag{10}$$

3.2.2. Tier 2: Multi-dimensional parameter metric

The second tier consists of an analysis of the posterior distribution of parameter duplicates in the split cosmological model. Given that the posterior distribution for several sampling parameters is prior-dominated, we restrict the tier 2 test to a subset of parameters $\boldsymbol{\theta}$ that are constrained by the data while marginalising over the remaining parameters $\boldsymbol{\theta}^{\rm marg}$. We calculate the difference

$$\Delta\boldsymbol{\theta} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \tag{11}$$

for each data point in the chain, where $\boldsymbol{\theta}_{1/2}$ denotes the two parameter instances in the subspace of parameters of interest. We then analyse the posterior $P(\Delta\boldsymbol{\theta})$ in the subspace of parameters of interest. The posterior of parameter differences is given by

$$P(\Delta\boldsymbol{\theta}) = \int {\rm d}\boldsymbol{\theta}_1\, P^{\rm marg}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_1 - \Delta\boldsymbol{\theta}) \,, \tag{12}$$

where $P^{\rm marg}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ denotes the posterior distribution marginalised over the remaining (unconstrained) parameters. In the absence of internal tension in the data, we expect $P(\Delta\boldsymbol{\theta})$ to be centred on the origin, while internal inconsistencies may shift the posterior away from the origin. To quantify the deviation of the posterior from zero, we follow the approach of Köhlinger et al. (2019) and model the posterior with a kernel density estimator (KDE). We evaluate the KDE at the origin and determine the volume of the posterior where the probability of a shift is higher than the probability of no shift. Mathematically, this is equivalent to

$$V = \int_{P(\Delta\boldsymbol{\theta})<P(0)} {\rm d}\Delta\boldsymbol{\theta} P(\Delta\boldsymbol{\theta}) \,, \tag{13}$$

while in practice we compute the fraction of (weighted) samples with a posterior value lower than $P(0)$. The significance of the shift $m\sigma$ is computed by identifying V with the probability mass of a one-dimensional Gaussian distribution outside of the interval $[-m\sigma, m\sigma]$. Thus, the tension in levels of sigma is given by

$$m = \sqrt{2}\,\mathrm{erf}^{-1}(1 - V)\ . \tag{14}$$

We note that this approach is mathematically equivalent to the Monte Carlo exact parameter shift method adopted in Raveri et al. (2020) and Dark Energy Survey and Kilo-Degree Survey Collaboration et al. (2023).

3.2.3. Tier 3: Posterior predictive distribution metric

The third tier of consistency tests compares the observed data with predictions generated from the posterior distribution of model parameters. This is usually probed via the posterior predictive distribution (PPD), which describes the distribution of data realisations given the observed data and assuming a particular model. Given a set of observed data $\boldsymbol{d}$ and a model $\mathcal{M}$ the distribution of data realisations $\hat{\boldsymbol{d}}$ is given by

$$\mathrm{P}(\hat{\boldsymbol{d}}|\boldsymbol{d}, \mathcal{M}) = \int \mathrm{d}\boldsymbol{\theta}\, P(\hat{\boldsymbol{d}}|\boldsymbol{\theta}, \mathcal{M}) P(\boldsymbol{\theta}|\boldsymbol{d}, \mathcal{M})\ . \tag{15}$$

By testing whether the observed data is compatible with being drawn from the PPD, the data can be probed for internal inconsistencies. This is typically achieved by drawing data realisations from the PPD and evaluating a test statistic for both the synthetic and the observed data. This allows us to calculate p-values representing the probability of getting a higher test statistic for synthetic data realisations than for the observed data, which serves as a measure of consistency in the data (Doux et al. 2021).

As an alternative to a test of the PPD, Köhlinger et al. (2019) introduced the so-called translated posterior distribution (TPD) as a special case of the PPD, which can be obtained by translating posterior samples back into model predictions. Therefore, it describes the distribution of model predictions given the uncertainty of model parameters. Since the TPD can directly be generated as a byproduct of the sampling process, we adopt the TPD in our consistency test in data space and employ the χ^2 values as test statistic for the internal consistency between subsets of the data. Prior to the consistency analysis, we conducted sensitivity tests on internally inconsistent mock data, which confirmed that our TPD-based consistency metric yields estimates of the significance of the internally inconsistency that is compatible with the estimates inferred with the tier 1 and tier 2 metrics (see Appendix C).

Considering the split cosmological model, we infer the TPD for each set of cosmological parameters, $\boldsymbol{\theta}_{\mathrm{A}}$ and $\boldsymbol{\theta}_{\mathrm{B}}$ consisting of theory predictions for the full data vector, $\boldsymbol{t}(\boldsymbol{\theta}_{\mathrm{A}})$ and $\boldsymbol{t}(\boldsymbol{\theta}_{\mathrm{B}})$. We then quantify to what extent the observed data in one subset $\boldsymbol{d}_{\mathrm{A}}$ is compatible with the TPD inferred from the other subset and vice versa. To do so, we draw a realisation of $\boldsymbol{d}_{\mathrm{A}}$ conditioned on the TPD of subset B for each sample, denoted by $\boldsymbol{d}_{\mathrm{A}}^{\mathrm{sim}}$. Since the conditional distribution of one set of variables conditioned on the other is Gaussian if both sets are jointly Gaussian, the simulated data points are given by a multivariate Gaussian distribution $\mathcal{N}(\mu_{\mathrm{A}}^{\mathrm{sim}}, \mathbf{C}_{\mathrm{A}}^{\mathrm{sim}})$ with (see for example Bishop 2006)

$$\mu_{\mathrm{A}}^{\mathrm{sim}} = \boldsymbol{t}_{\mathrm{A}}(\boldsymbol{\theta}_{\mathrm{B}}) + \mathbf{C}_{\mathrm{AB}}\mathbf{C}_{\mathrm{BB}}^{-1}\left[\boldsymbol{d}_{\mathrm{B}} - \boldsymbol{t}_{\mathrm{B}}(\boldsymbol{\theta}_{\mathrm{B}})\right] \tag{16}$$

$$\mathbf{C}_{\mathrm{A}}^{\mathrm{sim}} = \mathbf{C}_{\mathrm{AA}} - \mathbf{C}_{\mathrm{AB}}\mathbf{C}_{\mathrm{BB}}^{-1}\mathbf{C}_{\mathrm{BA}}\ , \tag{17}$$

where $\mathbf{C}_{\mathrm{AB}}$ denotes the cross-covariance between the two data subsets. For each simulated data vector we then compute the χ^2-statistic given by

$$\chi^2\left[\boldsymbol{d}, \boldsymbol{t}(\boldsymbol{\theta})\right] = \left[\boldsymbol{d} - \boldsymbol{t}(\boldsymbol{\theta})\right]^T \mathbf{C}^{-1}\left[\boldsymbol{d} - \boldsymbol{t}(\boldsymbol{\theta})\right]\ . \tag{18}$$

We quantify the consistency between data regions in terms of the p-value $p(A|B)$, which is given by the fraction of posterior samples with

$$\chi^2\left[\boldsymbol{d}_{\mathrm{A}}^{\mathrm{sim}}, \boldsymbol{t}_{\mathrm{A}}(\boldsymbol{\theta}_{\mathrm{B}})\right] > \chi^2\left[\boldsymbol{d}_{\mathrm{A}}, \boldsymbol{t}_{\mathrm{A}}(\boldsymbol{\theta}_{\mathrm{B}})\right]\ . \tag{19}$$

The p-value quantifies the probability of the data in data subset A being a realisation of the TPD of subset B. Thus, low p-values indicate an internal inconsistency of the data. We note that Doux et al. (2021) show that this method of quantifying consistencies can result in p-values that are biased low if the two posteriors prefer vastly dissimilar regions in parameter space. This can be circumvented by calibrating the p-value with simulated data. In this work, we interpret the TPD based p-value as a conservative metric for the internal consistency since it can exaggerate a potential tension in the data set and reserve a further calibration of the p-value for cases for which it fails to pass our adopted threshold for internal consistency.

4. Results of internal consistency tests

In this section, we present the results of our internal consistency analysis. We adopt the fiducial CosmoPipe pipeline and sample the parameter space via the Nautilus[3] sampler (Lange 2023), which is interfaced with the cosmological parameter estimation code Cosmosis (Zuntz et al. 2015). A list of model parameters is provided in Table 1. As is standard practice in stage-III cosmic shear analyses, we adopt a multivariate Gaussian likelihood. In the split cosmological analysis, we duplicate parameters with uniform priors, while parameters with informative Gaussian priors are not duplicated. An exception is the colour-based split of the catalogue, for which we calibrate separate redshift nuisance parameters and allow for different intrinsic galaxy alignments between the subsets. We model the theoretical prediction for the observed signal in each subset with the two independent sets of cosmological parameters. When comparing theory and data, both subsets are linked through the data covariance matrix, which is computed analytically using the OneCovariance[4] code (Reischke et al. 2024). We note that our covariance model adopts the NLA model of intrinsic alignments as opposed to the fiducial NLA-M model. As was shown by Reischke et al. (2024), however, the contribution of intrinsic alignments has a negligible impact on the covariance of KiDS-Legacy data. In W25, we consider p-values of $p > 0.01$, corresponding to a 2.36σ offset, to be consistent with the null hypothesis for systematics tests. Therefore, we adopt the same threshold for the internal consistency tests conducted in this work.

Before unblinding of the KiDS-Legacy catalogue, we conducted the full consistency analysis on one blind for all three summary statistics. The blinding process, adopted from Kuijken et al. (2015), involves the generation of two additional catalogues with systematic differences in the measured galaxy shapes, which result in up to $\pm 2\sigma$ shifts in the inferred S_8. As the consistency tests are not sensitive to the overall S_8 value

[3] https://github.com/johannesulf/nautilus/tree/v0.7.2

[4] https://github.com/rreischke/OneCovariance/tree/v1.0.0

we do not expect the blinding to have an impact on the internal consistency of the KiDS-Legacy data set. With the exception of the split by angular scales with 2PCFs reported in Sect. 4.1.3, we found no differences between consistency tests for the three statistics. We therefore limit our analysis to COSEBIs, except for the split of the data vector by scale, for which we employ all three summary statistics. When evaluating the consistency in parameter space, we focus on $\Omega_{\rm m}$ and S_8, which are the two parameters that are mostly constrained by cosmic shear data. Prior to the cosmological analysis, we determine the number of constrained parameters N_Θ that is required for the suspiciousness test, as discussed in Sect. 3.2.1. The results of this analysis are presented in Appendix B. In Appendix C, we present a sensitivity analysis with mock realisations of the fiducial data vector, showing that in the absence of internal tension each metric yields a level of consistency that is compatible with typical noise fluctuations in the data.

4.1. Splits at the data vector level

In the case of data vector-level splits, we employ the fiducial KiDS-Legacy cosmic shear data vector, covariance matrix, and the prior on the shift in the mean of the redshift distribution per tomographic bin of W25 when conducting the likelihood analysis.

4.1.1. Redshift bin split

The first split at the data vector level is designed to test the internal consistency between the six tomographic redshift bins. In this way, we probe the data for any errors in our redshift calibration and redshift-dependent modelling effects, such as the impact of baryon feedback or the effect of IA of galaxies, which has a larger relative contribution to the total signal compared to the lensing signal at low redshifts. For each bin, we divide the theory vector into one subset containing the autocorrelation of the specific bin and its cross-correlation with the remaining redshift bins. The second subset consists of all auto-correlations of the remaining redshift bins and their cross-correlation. This split is analogous to the consistency test between KiDS-1000 redshift bins presented in Asgari et al. (2021). A complementary method of testing the consistency between redshift bins is the entire removal of single tomographic bins from the analysis. This test is demonstrated in W25 and is commonly applied in weak lensing studies (see for example Amon et al. 2022; Li et al. 2023b).

The first six panels of Fig. 1 show the marginalised posterior distribution of the split cosmological analysis for the split by redshift bin. Each panel represents the constraints from a single redshift bin and its cross-correlation with the other bins in yellow and the constraints from the auto- and cross-correlations between the remaining bins in red. For reference, we visualize the constraints from the fiducial cosmic shear analysis with the black dashed line. A visual inspection of the contours indicates a good agreement between the split and fiducial analyses. As expected, low redshift bins only yield loose constraints on cosmological parameters, which however are in good agreement with the remaining tomographic bins. The largest shift between contours is observed in the split of the second bin. We note that the previous consistency analysis of KiDS-1000 data (see appendix B in Asgari et al. 2021) also showed the largest discrepancy in the second tomographic bin. However, we emphasise that along with the inclusion of additional survey area and calibration data and the re-reduction of previously released data in KiDS-DR5, the definition of tomographic bins changed from equidistant binning in photometric redshift in KiDS-1000 to equi-populated binning in KiDS-Legacy, which Sipp et al. (2021) recommend as a better choice for the reduction of statistical errors. Therefore, any direct comparison between the consistency analysis of tomographic bins in KiDS-1000 and KiDS-Legacy should be made with caution.

The consistency levels for the redshift bin splits are listed in Table 2. The first two columns show the results of the tier 1 test with evidence-based metrics. The Bayes' ratio indicates preference for the single cosmological model, ranging from 'barely worth mentioning' in bin 2 to 'strong' in bin 6 according to Jeffreys' scale. In terms of the suspiciousness, all bins are found to be in agreement with $N_{\sigma,S} < 1$ except for bins 2 and 4, which show slightly larger values at 1.16σ and 1.39σ, respectively. The good consistency between redshift bins is further confirmed by the tier 2 multi-parameter metric test. All redshift bins show agreement with $N_\sigma \leq 1.39$, which is found when considering a shift in $\Omega_{\rm m}$ in the fifth bin. The final two columns list the consistency level from the tier 3 PPD test in terms of the p-value for data vector predictions for the data subset listed in the second column, inferred from the TPD of the other subset (column 8) and vice versa (column 9). Here, all p-values pass our threshold of $p > 0.01$. Overall, we highlight that all consistency metrics indicate a good internal consistency for the split by redshift bin. In particular, while the previous KiDS-1000 analysis showed an internal inconsistency between redshift bins of up to 3σ, we find the KiDS-Legacy data to be in better internal agreement with consistency levels better than 1.39σ and compatible with typical statistical fluctuations.

4.1.2. Auto-correlation vs cross-correlation

In addition to the redshift bin split, we apply a split of the data vector between auto- and cross-correlation signals of the tomographic bins. This split allows us to probe the data for systematic effects that affect the two types of correlation signals through different processes. In particular, the individual IA contributions to the cosmic shear signal, given in Eq. (2), can be attributed to either the auto- or the cross-correlation signal. The II term is generated through the alignment of physically close galaxies due to tidal forces of the nearby large-scale structure. Therefore, it predominantly affects the autocorrelation of tomographic bins. The GI term on the other hand is induced by the large-scale structure causing both the intrinsic alignment of nearby galaxies and the lensing of distant galaxies, which leads to an anti-correlation between the shapes of galaxy pairs that are separated in redshift. Therefore, this effect manifests itself primarily in the cross-correlation signal between tomographic bins. We divide the theory vector into one subset consisting of the auto-correlation signal of all six tomographic bins and the second subset containing all cross-correlations between bins.

The bottom right panel of Fig. 1 shows the posterior from the consistency test between the autocorrelation group of all tomographic bins in yellow and the cross-correlation group in red. The constraints from the fiducial cosmic shear analysis are visualised with the black dashed line. The consistency metrics, listed in Table 2 signify an agreement between both groups in all tests. This finding is in agreement with the almost indistinguishable posteriors shown in the last panel of Fig. 1. At first glance, it may be surprising that the auto-correlation and cross-correlation parts of the data vector have the same constraining power as each other, and the fiducial data vector. Typically the GI contribution to the cross-correlation signal provides the main information to

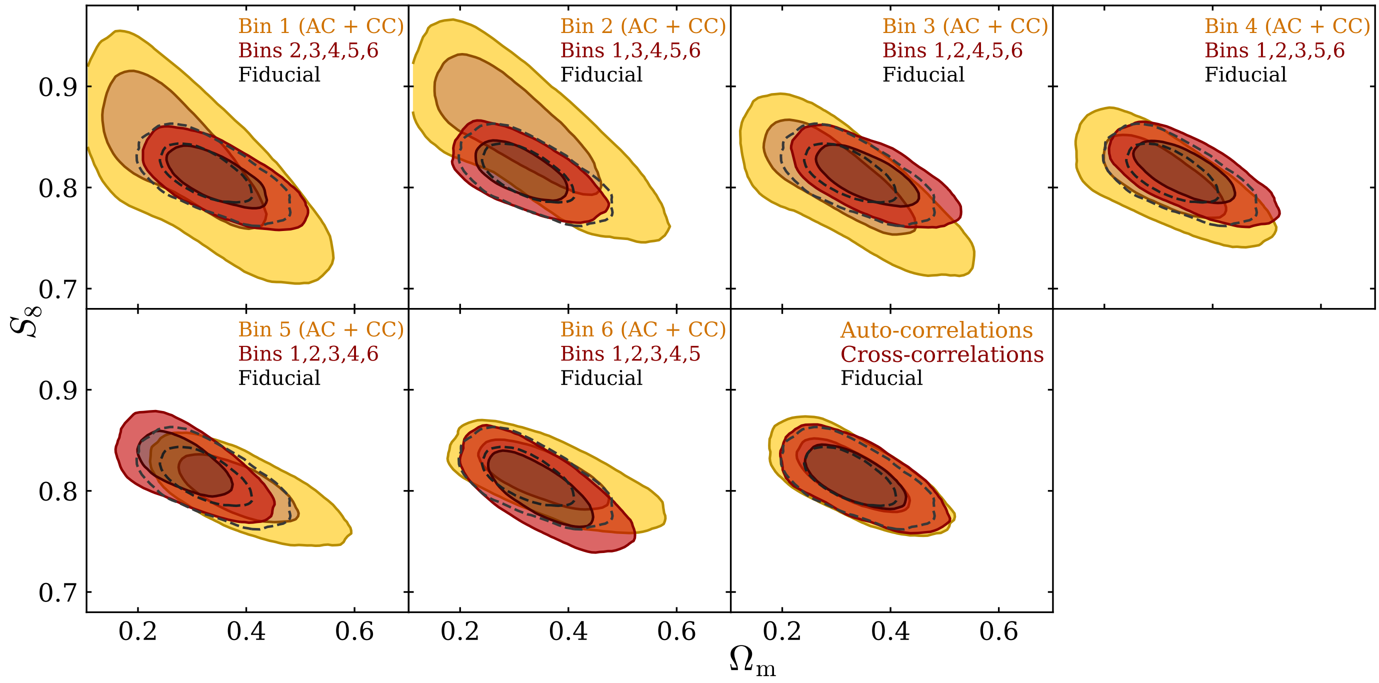

Fig. 1. Posterior distribution of the two instances of cosmological parameters in a split by redshift bin for COSEBIs. The yellow contours show the posterior of parameters modelling one specific redshift bin and its cross-correlation with the other bins, while the red contours show the posterior distribution of the parameters modelling the auto- and cross-correlation signal of the remaining redshift bins. The dashed contours show the fiducial constraints for reference. The final panel presents the posterior distribution in a split between auto-correlations of all redshift bins and their cross-correlations. When running the chains, both regimes are linked through the cross-covariance between redshift bins. The inner and outer contours of the marginalised posteriors correspond to the 68% and 95% credible intervals, respectively.

Table 2. Consistency metrics for data vector level splits of KiDS-Legacy data.

		Tier 1		Tier 2			Tier 3	
Statistic	Split	$\log_{10} R$	$N_{\sigma,S}$	$N_\sigma(\Delta\Omega_\mathrm{m})$	$N_\sigma(\Delta S_8)$	$N_\sigma(\Delta(\Omega_\mathrm{m}, S_8))$	$p(A\|B)$	$p(B\|A)$
COSEBIs	Bin 1	0.91	0.50	0.81	0.61	0.35	0.40	0.49
	Bin 2	0.31	1.16	0.34	1.09	1.34	0.14	0.74
	Bin 3	1.07	0.65	0.66	0.02	0.61	0.72	0.39
	Bin 4	0.89	1.39	0.65	0.22	0.72	0.85	0.23
	Bin 5	1.38	0.87	1.39	0.99	0.84	0.67	0.26
	Bin 6	1.46	0.47	0.09	0.51	0.50	0.43	0.49
	Redshift bin AC vs CC	1.88	0.18	0.07	0.06	0.00	0.23	0.48
	Scale	1.48	0.63	0.07	0.69	0.62	0.37	0.38
Band powers	Scale	0.40	1.49	0.16	1.04	0.84	0.51	0.88
2PCFs	Scale	-2.03	3.69	0.30	1.81	1.63	0.34	0.20

Notes. The first column indicates the summary statistic and the second column lists the split applied to the data vector. The third and fourth columns report the tier 1 evidence-based metric of the Bayes ratio and the tension level inferred from the suspiciousness, respectively. The next three columns present the results of the tier 2 multi-dimensional parameter metric test for Ω_m, S_8, and their combination. The final two columns shows the tension level arising from the tier 3 PPD metric test in terms of the p-value for data vector predictions for the data subset listed in the second column inferred from the PPD of the other subset (column 8) and vice versa (column 9).

constrain the IA amplitude, and as such a cosmic shear analysis that excludes the cross-terms decreases the overall constraining power. The fact that we find the same constraining power can be understood by recognising that the NLA-M intrinsic alignment model parameters in our analysis are prior dominated and also shared between the two data splits with the cross-correlation data informing the auto-correlation IA model. If this was not the case then we would expect to see a degradation in constraining power when analysing the auto- and cross-correlations separately.

4.1.3. Scales

When selecting which scales to include in a cosmic shear analysis there is a trade-off between the desire to maximise signal-to-noise by including as much data as possible, and minimising the impact of unaccounted systematic errors that are expected to contaminate the smallest angular scales. Our fiducial cosmological constraints analyse data collected over the angular range $\theta \in [2', 300']$, and in this section we assess the scale consistency by separating small and large scale data using all three statistics, 2PCFs, band power spectra and COSEBIs.

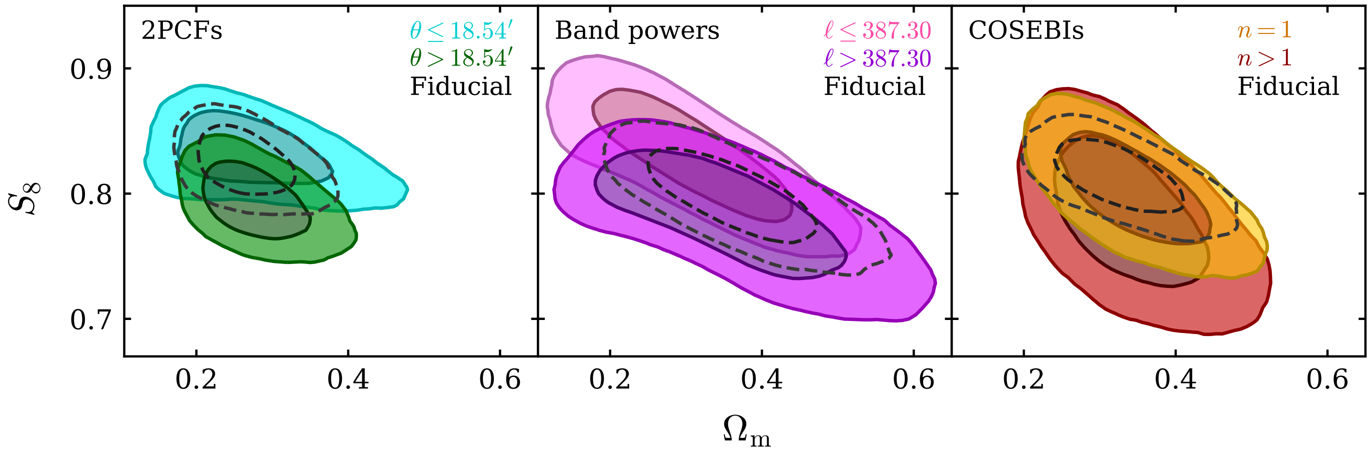

Fig. 2. Posterior distribution of the two instances of cosmological parameters in a split by scale for 2PCFs (left), band powers (middle), and COSEBIs (right) in comparison to the fiducial analysis with each summary statistic, illustrated by the black dashed lines. The inner and outer contours of the marginalised posteriors correspond to the 68% and 95% credible intervals, respectively.

For the 2PCFs, we split the 9 logarithmically spaced angular bins at an angle of $\theta \approx 18.54'$, with the first/second set consisting of four/five angular bins ensuring comparable signal-to-noise in both sub-samples[5]. For this comparison, we find a tension at 3.69σ between small and large scales when considering the suspiciousness test. This can be observed in the left panel of Fig. 2, which shows a preference for lower values of S_8 at large scales and a preference for higher values at small scales, with a 1.81σ shift in S_8 between the two. The PPD-based test, however, concludes that the data vectors are consistent. We note that W25 do not include 2PCFs in the fiducial cosmic shear analysis, only providing 2PCF constraints for completeness and comparison with previous works. This is because the 2PCF is particularly sensitive to the effect of baryon feedback which is challenging to model (see for example Asgari et al. 2020). As we have not optimised the angular scales for a 2PCF KiDS-Legacy analysis (see for example Krause et al. 2021) we conclude that the tension reported by the suspicious tier 1 test is likely to be caused by an insufficient treatment of baryonic effects for the fiducial $\theta \in [2', 300']$ angular scale range. As such we do not expect the cosmological constraints from 2PCFs to be as reliable as those inferred with our fiducial COSEBIs statistic. We further note that on large angular scales, the assumption of a Gaussian likelihood does not hold (see for example Sellentin et al. 2018; Louca & Sellentin 2020; Joachimi et al. 2021; Oehl & Tröster 2024), which particularly affects the ξ_+ correlation function, leading to a potential bias in the inferred S_8.

With our band power spectra we can test the consistency between small and large multipoles in Fourier space (see fig. 1 of W25 for the band power filter functions that are compact in ℓ). We divide the eight logarithmically spaced bands between $\ell \in [100, 1500]$ at a limit of $\ell \approx 387$, creating two subsets consisting of four bands each. The middle panel of Fig. 2 shows the cosmological constraints from this split analysis with the second to last row of Table 2 reporting the suite of consistency metrics. We find agreement between the low and high ℓ band power measurements for all our tests.

It is hard to define a data split for a scale-sensitivity analysis with COSEBIs as each E_n mode is sensitive to a range of ℓ-scales, varying only in the way those scales are combined (see for example fig. 1 of W25). We therefore chose to conduct a data-split analysis of the COSEBI $n = 1$ mode, which carries the majority of the constraining power, versus the other $n = 2$ to $n = 6$ modes. The result is visualised in the right panel of Fig. 2 with the consistency metrics listed in Table 2, demonstrating an agreement between the COSEBIs modes in all tests.

[5] Following previous KiDS analyses, we limit the ξ_- correlation function to $\theta > 4'$, removing the first θ-bin from the analysis.

4.2. Splits at the catalogue level

In this section, we split the data on the catalogue level, analysing the two distinct sections of the KiDS footprint, KiDS-North and KiDS-South, and splitting the galaxy sample by colour. When conducting the likelihood analysis, we construct a joint data vector and covariance matrix of both subsets, doubling the dimensionality of the data vector with respect to the fiducial analysis. In Appendix A we report the data properties of each catalogue split with calibrated redshifts and shear measurements for each sample along with the B-mode signal of each subset which we find to be consistent with the null hypothesis.

4.2.1. North-South split

As discussed in Sect. 2, the KiDS observations were taken on two distinct patches on sky: one at the celestial equator, dubbed KiDS-North, and one at the South Galactic Pole, dubbed KiDS-South. The two patches, which are shown in fig. 2 of W25, cover a similar area on sky with approximately 496 deg^2 of post-masking data in KiDS-North and 472 deg^2 in KiDS-South. Thus, both patches contain a comparable number of sources per tomographic bin. In the fiducial analysis pipeline, we combine independent measurements of the cosmic shear 2PCF per patch into a single measurement, from which we compute the COSEBIs data vector. In the North-South split, we keep the 2PCF measurements in each patch separate in order to test their consistency. By doing so, we obtain one data vector per hemisphere. Since the two patches are separated on sky, we do not expect a cross-correlation signal between patches. Therefore, the combined covariance matrix only consists of two non-zero blocks, each containing the covariance for North and South, respectively.

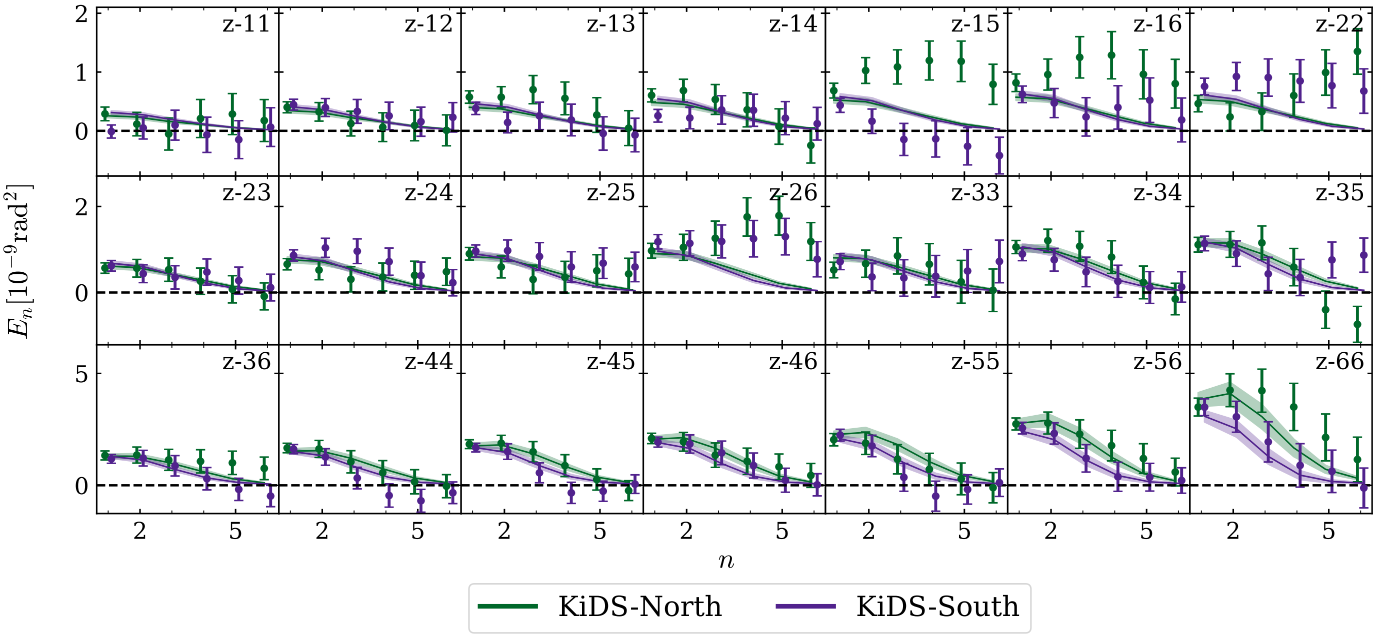

Fig. 3. COSEBIs E-mode measurements and their best-fitting model for a split cosmological analysis of the North-South split catalogue. The green and purple data points show the measurements of the KiDS-North and KiDS-South sample, respectively. The best-fitting theoretical predictions are given by the solid lines, and the 1σ interval of the translated posterior distributions are illustrated by the shaded regions. Each panel represents auto- or cross-correlation between tomographic bins. For visualisation purposes, we display the discrete n modes with an offset on the x-axis.We note that the E-mode signals are highly correlated within a tomographic bin and advise against a so-called 'χ-by-eye'.

Table 3. Consistency metrics for catalogue level splits of KiDS-Legacy data.

	Tier 1		Tier 2			Tier 3	
Split	$\log_{10} R$	$N_{\sigma,S}$	$N_\sigma(\Delta\Omega_\mathrm{m})$	$N_\sigma(\Delta S_8)$	$N_\sigma(\Delta(\Omega_\mathrm{m}, S_8))$	$p(A\|B)$	$p(B\|A)$
North vs South	0.83	0.71	1.09	0.36	0.60	0.36	0.39
Red vs Blue T_B = 3.0	-0.44	2.61	1.08	0.08	0.74	0.76	0.40
Red vs Blue T_B = 1.9	-1.29	2.84	0.70	0.46	0.25	0.64	0.35

Notes. The second and third columns report the Bayes ratio and the tension level inferred from the suspiciousness, respectively. The next three columns present the results of the tier 2 test for Ω_m, S_8, and their combination. The final two columns shows the tension level arising from the tier 3 test in terms of the p-value for data vector predictions for the first subset of the catalogue inferred from the TPD of the other subset (column 7) and vice versa (column 8).

In Fig. 3, we present the COSEBIs data vector per hemisphere along with the corresponding theory prediction from the best-fitting model with uncertainties inferred from the TPDs. As the data properties of the KiDS-North and KiDS-South patches are very similar, in terms of the redshift distributions, multiplicative shear and redshift calibration, and the ellipticity dispersion (see Appendix A), we have chosen to use a single shared set of observational nuisance parameters in our likelihood analysis, shown in the left panel of Fig. 4. Although the KiDS-North patch tends to favour a shift towards lower values of Ω_m compared to our fiducial analysis, both patches are in good agreement with a value of $N_\sigma(\Delta\Omega_\mathrm{m}) = 1.09$ inferred in the tier 2 multi-parameter consistency test. Considering S_8, we find both patches to be in agreement, which is confirmed by the tier 1 evidence and tier 3 PPD tests tabulated in the first row in Table 3. We conclude that the cosmological constraints from observations in KiDS-North and KiDS-South are fully consistent.

4.2.2. Red-blue split

Observational evidence shows that red early-type galaxies intrinsically align, in contrast to blue late-type galaxies, where intrinsic alignments have yet to be detected (Hirata et al. 2007; Joachimi et al. 2011; Heymans et al. 2013; Samuroff et al. 2019; Georgiou et al. 2019; Johnston et al. 2019; Fortuna et al. 2021; Tugendhat et al. 2020; Samuroff et al. 2023; Georgiou et al. 2025). We therefore chose to split the KiDS-Legacy galaxies into a sample of red and blue galaxies to study the impact of intrinsic galaxy alignments on our cosmic shear signal. This also allows us to explore a secondary effect where the more spherical shape of red galaxies changes the populations' ellipticity distribution, leading to possible differences in the shear calibration correction (Kannawadi et al. 2019).

We follow the methodology of Li et al. (2021), dividing the KiDS-Legacy sample into two subsets based on the spectral type T_B reported by the BPZ code. Li et al. (2021) defines blue galaxies via a threshold of $T_\mathrm{B} > 3.0$ to provide a similar constraining power per data subset. This is in contrast to appendix B of W25, where red galaxies are selected with a threshold of $T_\mathrm{B} \leq 1.9$.

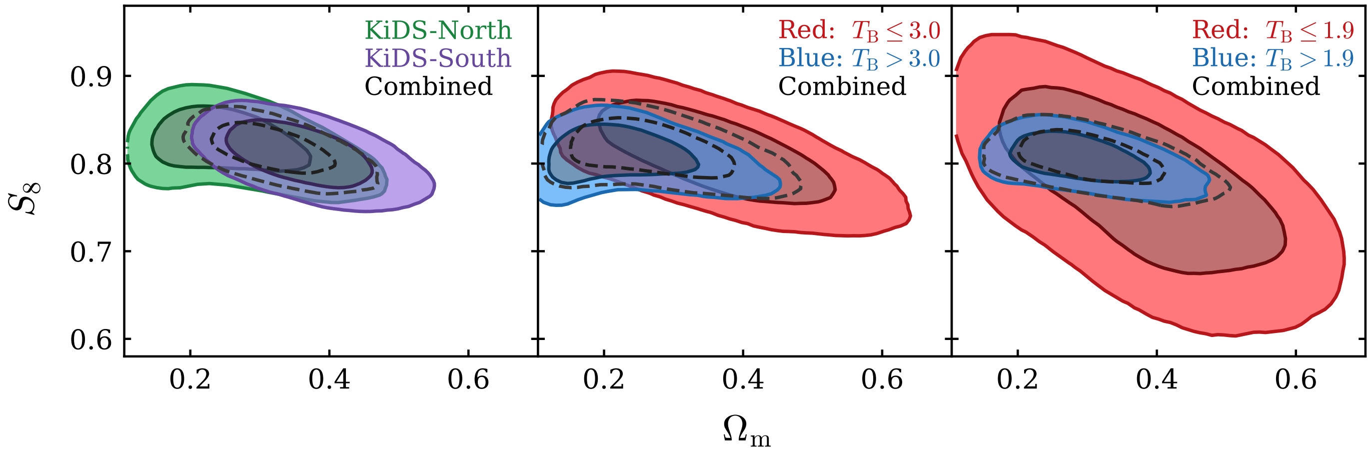

Fig. 4. Posterior distribution of parameter duplicates in the $\Omega_m - S_8$ plane for catalogue level splits for COSEBIs. Left panel: North-South split. Middle panel: Red-blue split defined via a cut on the spectral type of $T_B = 3.0$. Right panel: Red-blue split defined via a cut on the spectral type of $T_B = 1.9$. For reference, the black dashed contours show constraints from the analysis with a single set of parameters modelling both data subsets. The inner and outer contours of the marginalised posteriors correspond to the 68% and 95% credible intervals, respectively.

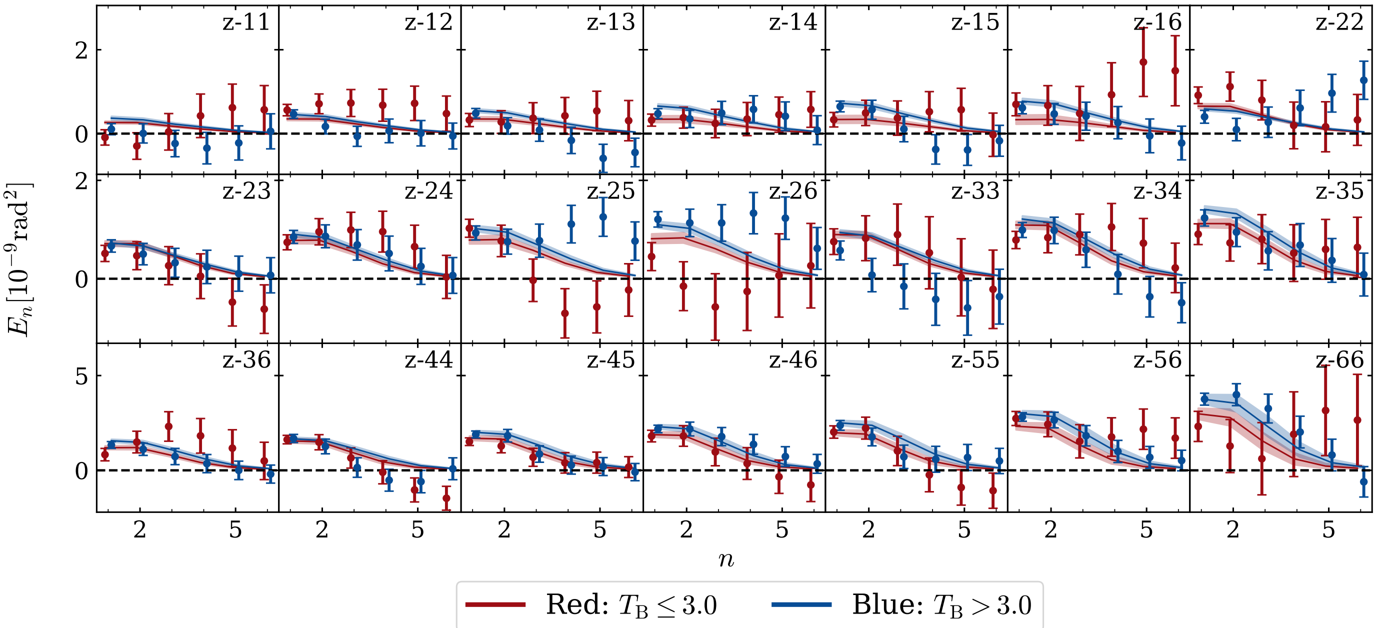

Fig. 5. COSEBIs E-mode measurements and their best-fitting model for a split cosmological analysis of the red-blue split catalogue, defined via a cut on the spectral type $T_B = 3.0$. The red and blue data points show the measurements of the red and blue sample, respectively. The best-fitting theoretical predictions are given by the solid lines, and the 1σ interval of the translated posterior distributions are illustrated by the shaded regions. Each panel represents auto- or cross-correlation between tomographic bins, as indicated by the label in the top right corner. For visualisation purposes, we display the discrete n modes with an offset on the x-axis. We note that the E-mode signals are highly correlated within a tomographic bin and advise against a so-called 'χ-by-eye'.

We present an analysis of each threshold[6], where dividing the sample at a threshold of $T_B = 3.0$ results in the red sample containing approximately one third of the galaxies, with the more stringent cut of $T_B = 1.9$ leaving $\sim 16\%$ of galaxies in the red sample. In Appendix A, we present the redshift distributions of each sample and the separate redshift nuisance parameters that are marginalised over in our analysis. In contrast to the division of galaxies by hemisphere, the red and blue galaxy samples are expected to be highly correlated and these correlations are taken into account via the cross-covariance between cosmic shear measurements of the red and blue samples computed with the OneCovariance code.

Our fiducial cosmic shear analysis adopts the NLA-M IA model which sets any blue galaxy alignment to zero for $T_B > 1.9$. This model is therefore not applicable when considering a split between red and blue galaxies at a threshold of $T_B = 3.0$

[6] We note another alternative, that we do not explore here, is to split a catalogue by colour as proposed by McCullough et al. (2024), applying a SOM-based selection on $r - z$ colour in order to derive a high-purity sample of blue galaxies.

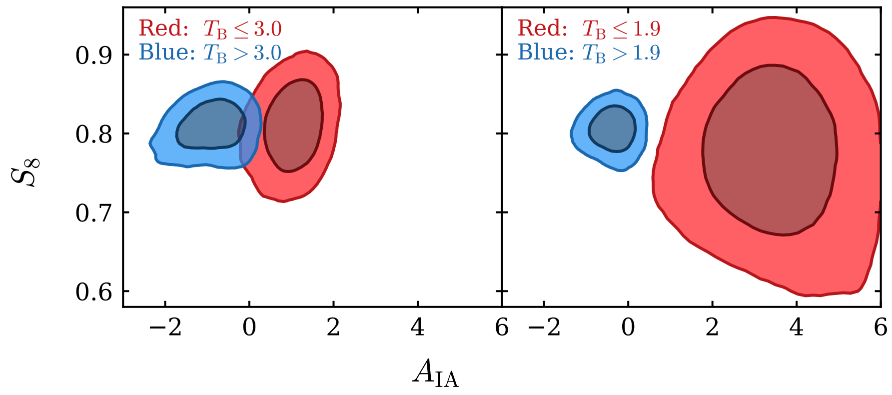

Fig. 6. Constraints on S_8 and $A_{\rm IA}$ for colour-based splits of the catalogue. Left panel: Red-blue split defined via a cut on the spectral type of $T_{\rm B} = 3.0$. Right panel: Red-blue split defined via a cut on the spectral type of $T_{\rm B} = 1.9$. For this threshold, the catalogue only contains very few red galaxies with $z_{\rm B} > 1.14$. Therefore, the red galaxy sample only encompasses the first five tomographic bins. The inner and outer contours of the marginalised posteriors correspond to the 68% and 95% credible intervals, respectively.

and we revert to the standard NLA model in this analysis for both $T_{\rm B}$ thresholds with a separate amplitude parameter $A_{\rm IA}$ for each galaxy sample.

The measured COSEBIs E modes for the colour-based split with a threshold $T_{\rm B} = 3.0$ and the TPDs from the best-fitting theory model are illustrated in Fig. 5. Here, each colour represents measurements of the auto- and cross-correlation between tomographic bins of the given sample. We chose not to include cross-correlation measurements between the red and blue bins as these signals mix contributions from theoretical predictions for the red and blue signals which can not be easily modelled with our current pipeline. The cross-correlation between the red and blue data points is, however, taken into account in our consistency analysis via the joint covariance matrix.

The posterior distribution for both instances of $\Omega_{\rm m}$ and S_8 in the split cosmological analysis are shown in the middle and right panels of Fig. 4. As expected, the red galaxy sample yields weaker constraints on S_8 and $\Omega_{\rm m}$ than the blue sample due to the higher number density of blue galaxies. This is particularly true for the $T_{\rm B} = 1.9$ selected sample where the marginalised posteriors are almost unconstrained by the red galaxy sample. Nevertheless, the cosmological parameter posterior distributions for both red/blue splits are in good agreement. Looking at the consistency metrics listed in Table 3 for the $T_{\rm B} = 3.0$ ($T_{\rm B} = 1.9$) selected samples, we find that the suspiciousness test shows a preference for the split cosmological model at 2.61σ (2.84σ). In contrast, the tier 2 test yields an agreement between parameters with $N_\sigma \leq 1.08$ ($N_\sigma \leq 0.70$) when considering S_8, $\Omega_{\rm m}$, and their combination. This is confirmed by the tier 3 tests, which suggests a good agreement between the TPDs and the observed data in both subsets.

The tier 1 evidence-based preference for the split cosmological model can be explained by the additional freedom in the modelling of intrinsic galaxy alignments. While the combined galaxy sample analysis assumes a shared IA amplitude for both red and blue galaxies, the split analysis models intrinsic alignments with two independent parameters. As shown in Fig. 6, there is a significant difference between the IA amplitudes for the two samples with a marginal mode and the 1D 68% highest posterior density interval (HPDI) of $A_{\rm IA}^{\rm blue} = -0.75^{+0.43}_{-0.62}$ and $A_{\rm IA}^{\rm red} = 1.08^{+0.46}_{-0.44}$ for the $T_{\rm B} = 3.0$ sample split. For the $T_{\rm B} = 1.9$ split, we find the IA amplitude of blue galaxies to be $A_{\rm IA}^{\rm blue} = -0.32^{+0.33}_{-0.36}$, while the red galaxy sample yields $A_{\rm IA}^{\rm red} = 3.32^{+1.13}_{-0.97}$. This corresponds to a difference in their posterior distribution at $N_\sigma(\Delta A_{\rm IA}) = 2.81$ for $T_{\rm B} = 3.0$ and $N_\sigma(\Delta A_{\rm IA}) = 2.57$ for $T_{\rm B} = 1.9$. These results are compatible with the central assumption of the NLA-M model used in our fiducial cosmic shear analysis, which assumes zero alignment of blue galaxies. We therefore do not interpret the tier 1 result as an indication of internal inconsistency given that the physical mechanism behind red and blue galaxy alignment is expected to differ. As the evidence-based tier 1 consistency test compresses the full posterior into a single statistic, the physical difference between red and blue galaxies is reflected as a preference for the split cosmological model. Based on the other consistency metrics reported in Table 3, we can therefore conclude there is internal consistency between our samples of red and blue galaxies.

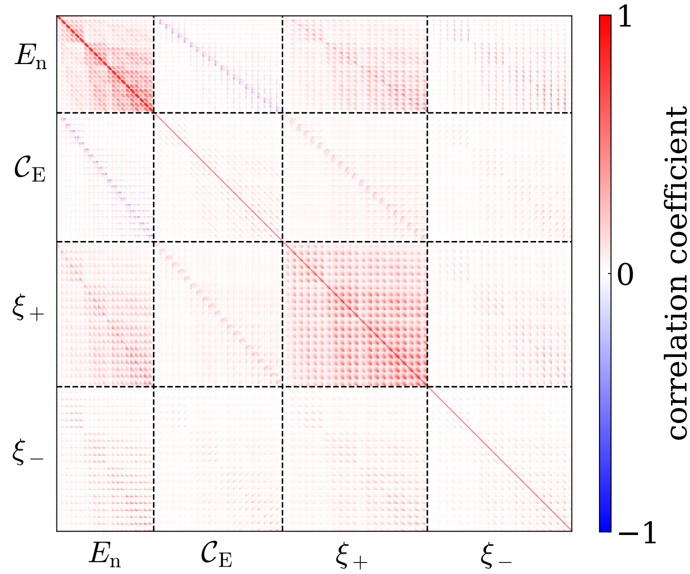

Fig. 7. Correlation matrix between measurements of COSEBIs, band powers, and 2PCFs.

4.3. Consistency between summary statistics

While the two primary summary statistics in the fiducial analysis, COSEBIs and band powers, as well as the additional statistic, binned 2PCFs, are derived from the same cosmic shear 2PCF measurements, they differ in their sensitivity to spatial scales as well as systematic and modelling effects. Thus, we quantify the consistency between all combinations of two summary statistics. Although this does not correspond to a split of the data vector or the catalogue since each statistic originates from the same measurements of the cosmic shear 2PCFs, we can nevertheless use the same methodology to quantify whether or not the statistics

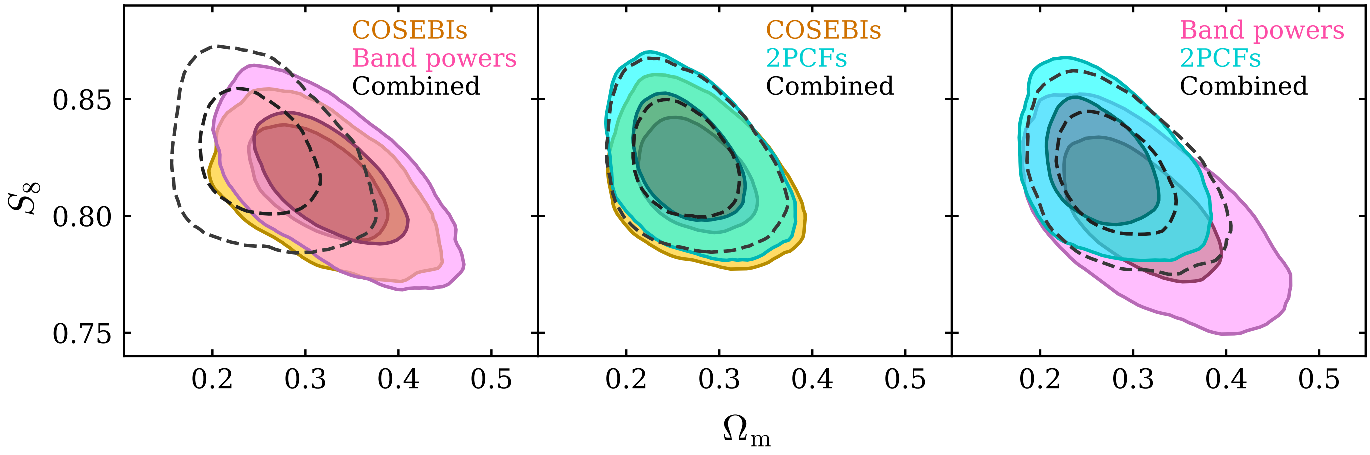

Fig. 8. Posterior distribution of parameter duplicates in the $\Omega_\mathrm{m} - S_8$ plane for split cosmological analyses with two summary statistics. For reference, the black dashed contours show constraints from the analysis with a single set of parameters modelling both data subsets. The inner and outer contours of the marginalised posteriors correspond to the 68% and 95% credible intervals, respectively.

prefer different cosmologies. In practice, we construct three data vectors that each combine measurements of two summary statistics and model the theoretical prediction with one set of parameters per statistic. As depicted in Eq. (1), they differ only in the corresponding weight function, which vary in their sensitivity to different angular scales. Therefore, we expect the statistics to be highly correlated, since the different summary statistics are derived from the same two-point correlation function measurements. Thus, it is particularly important to derive a robust estimate of the covariance between summary statistics, which is enabled by the OneCovariance code. The resulting correlation matrix is displayed in Fig. 7. As discussed in Reischke et al. (2024), the full covariance matrix between all summary statistics can be non-positive definite due to numerical noise. However, the sub-covariance matrices between two summary statistics are still positive definite and invertible.

The marginalised posterior distribution for Ω_m and S_8 are displayed in Fig. 8. Here, the solid contours refer to the analysis with the split cosmological model and the black dashed lines show constraints from the analysis of the combined data vector of two summary statistics with a single set of parameters. For the combination of COSEBIs and 2PCFs as well as band powers and 2PCFs, we find the constraints from the combined analysis to be in agreement with the split analyses. However, for the combination of COSEBIs and band powers we observe a preference for lower values of Ω_m in the combined analysis. We further investigate the origin of this feature by decomposing the likelihood into the contribution from the auto-correlation of COSEBIs and band powers and the contribution from their cross-correlation. This analysis shows that the shift towards low Ω_m is driven by the cross-covariance terms between COSEBIs and band powers. Additionally, we inspect the posterior distribution of the remaining split parameters, which is displayed in Fig. 9. We find that the two instances of the baryon feedback parameter, $\log T_\mathrm{AGN}$, show a preference for different amounts of baryonic feedback. While the COSEBIs posterior peaks at a low value of $\log T_\mathrm{AGN}$, corresponding to a dark matter only scenario, the posterior from band powers tends towards the upper edge of the prior. The origin of this feature most likely lies in their varying response to different scales. In the combined analysis with a single set of parameters, in which both statistics share the same baryonic feedback parameter, this most likely causes the shift of the posterior towards low Ω_m. As a consequence, the evidence-based consistency metric, provided in Table 4, reports a preference for the split cosmological model with $N_{\sigma,S} = 3.78$. In the tier 2 parameter space and tier 3 data space metrics, however, we find both statistics to be in agreement. We note that our default consistency test in parameter space focusses on the two parameters that are mostly constrained by our cosmic shear data, Ω_m and S_8. Considering the apparent discrepancy in the baryon feedback parameter, we compute the significance of the shift in $\log T_\mathrm{AGN}$, which we find to be $N_\sigma(\Delta \log T_\mathrm{AGN}) = 2.54$. Furthermore, we note that the posterior in the combined analysis of COSEBIs and band powers closely resembles the posterior of an analysis with 2PCFs (see appendix E in W25). As can be inferred from the corresponding window functions, the combination of COSEBIs and band powers covers approximately the same range of scales that is probed by 2PCFs, which provides an explanation for the similarity between their posteriors.

For the remaining combinations of summary statistics, our consistency analysis finds an agreement in all tests. Therefore, we conclude that for the parameters of interest we find the three summary statistics to be in agreement. This confirms the result of the fiducial cosmic shear analysis (W25), which reports a good agreement between marginalised S_8 constraints inferred individually with the three statistics. However, we note that a combined analysis of two summary statistics proves to be challenging due to the high degree of correlation between the statistics, as showcased in our combined analysis of COSEBIs and band powers.

5. Combination with external data

Weak lensing on its own only constrains a nearly degenerate combination of σ_8 and Ω_m, which is commonly described in terms of the S_8 parameter, where the width of the degeneracy is reflected in the uncertainty on S_8. Combining our KiDS-Legacy cosmic shear measurements with external data sets allows us to break the degeneracy between the two parameters. In particular, spectroscopic galaxy surveys provide complementary constraints on the expansion history of the Universe through measurements of the BAO feature over a range of redshifts. While the BAO feature resides in the quasi-linear clustering regime,

Table 4. Consistency metrics for the combination of summary statistics.

	Tier 1		Tier 2			Tier 3	
Statistic	$\log_{10} R$	$N_{\sigma,S}$	$N_\sigma(\Delta\Omega_\mathrm{m})$	$N_\sigma(\Delta S_8)$	$N_\sigma(\Delta(\Omega_\mathrm{m}, S_8))$	$p(A\|B)$	$p(B\|A)$
COSEBIS vs Bandpowers	0.09	3.78	0.87	0.06	0.85	0.24	0.85
COSEBIS vs $\xi_\pm$	2.75	0.21	0.24	0.45	0.15	0.45	0.09
Bandpowers vs $\xi_\pm$	1.46	1.31	0.82	1.09	0.57	0.87	0.08

Notes. The second and third columns report the Bayes ratio and the tension level inferred from the suspiciousness, respectively. The next three columns present the results of the tier 2 test for Ω_m, S_8, and their combination. The final two columns shows the tension level arising from the tier 3 test in terms of the p-value for data vector predictions for the first statistic inferred from the TPD of the other statistic (column 7) and vice versa (column 8).

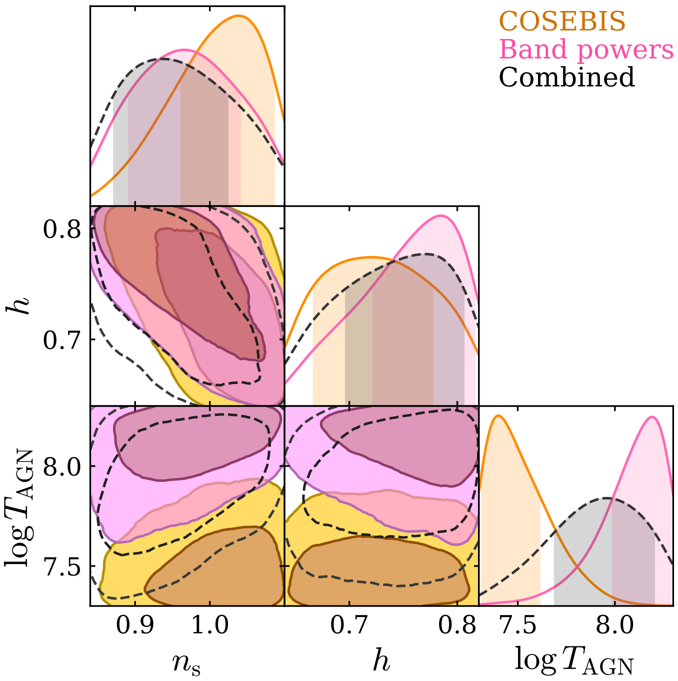

Fig. 9. Constraints on n_s, h, and $\log T_\mathrm{AGN}$ in a split cosmological analysis with COSEBIs (yellow) and band powers (pink). For reference, the black dashed contours show constraints from the analysis with a single set of parameters modelling both data sets. The inner and outer contours of the marginalised posteriors correspond to the 68% and 95% credible intervals, respectively.

RSD measurements probe the growth of structure. We adopt BAO measurements from DESI DR1 (Adame et al. 2025) as well as BAO and RSD measurements from eBOSS DR16 (Alam et al. 2021), which provide a tight constraint on Ω_m. Additionally, observations of SN Ia provide an alternative method of constraining the expansion history of the Universe. This method relies on measurements of the luminosity distances as a function of redshift, which provide an independent constraint on the matter density. Here, we employ SN Ia measurements from the Pantheon+ compilation (Scolnic et al. 2022; Brout et al. 2022).

We treat the KiDS and the external data vectors as independent and therefore compute the joint likelihood by multiplying the individual likelihoods of each experiment. Additionally, we assume independence between the BAO and SN measurements and conduct a joint analysis of KiDS and Pantheon+ data in combination with DESI Y1 BAO and eBOSS DR16, respectively. In Table 5, we quantify the consistency between KiDS-Legacy and the external data sets. As can be observed in Fig. 10, BAO, RSD, and SN Ia measurements put a tight constraint on the matter density. Additionally, BAO measurements constrain the Hubble parameter by incorporating an external calibration of the absolute BAO scale. Since the external data sets are sensitive to parameters that are mostly unconstrained by cosmic shear, we find a good agreement between KiDS-Legacy and DESI, eBOSS, and Pantheon+ in all tests. In the joint analyses, we find the marginal mode and the HPDI to be

$$
\begin{aligned}
\text{KiDS + DESI + Pantheon+:} \quad & S_8 = 0.818^{+0.015}_{-0.014} \\
& \sigma_8 = 0.803^{+0.024}_{-0.021} \\
& \Omega_\mathrm{m} = 0.311^{+0.011}_{-0.012} \\
\text{KiDS + eBOSS + Pantheon+:} \quad & S_8 = 0.819^{+0.014}_{-0.015} \\
& \sigma_8 = 0.798^{+0.023}_{-0.022} \\
& \Omega_\mathrm{m} = 0.315^{+0.012}_{-0.013} \,.
\end{aligned} \tag{20}
$$

The breaking of the σ_8-Ω_m degeneracy results in a reduction in the uncertainty on σ_8 by about 72% compared to the fiducial constraint of W25. In terms of S_8, this corresponds to a 22% uncertainty reduction. We note, however, that the preferred degeneracy direction for KiDS-Legacy in the more general $\Sigma_8 = \sigma_8(\Omega_\mathrm{m}/0.3)^\alpha$ parameterisation differs from the $\alpha = 0.5$ assumed in the definition of S_8, as discussed in section 5.1 in W25. Using the preferred $\alpha = 0.58$, we find $\Sigma_8 = 0.819^{+0.015}_{-0.013}$ in the combined analysis of KiDS-Legacy + DESI Y1 BAO + Pantheon+, which is consistent with the results of W25.

Additionally, we present a joint analysis between cosmic shear measurements of DES and KiDS. This analysis was previously conducted in DES+KiDS, who combined cosmic shear data from DES Y3 and KiDS-1000 data and García-García et al. (2024), who reanalysed DES Y3, KiDS-1000, and HSC DR1 data with a common harmonic-space pipeline. Here, we adopt the 'KiDS-excised' DES Y3 2PCF data vector of DES+KiDS and update the combined analysis with our KiDS-Legacy COSEBIs measurements. We sample the parameter space with our KiDS-Legacy pipeline within our fiducial prior space and adopt the independent Gaussian priors for incorporating the uncertainty on the shear and redshift calibration in DES (see table 1 of DES+KiDS). We note that the DES Y3 measurements excluded the overlap region between both surveys, which are therefore considered to be independent. Following the methodology of DES+KiDS, we adopt independent IA parameters for both surveys. We employ the fiducial NLA-M model for KiDS IA in our combined analysis. Since the prior on the NLA-M parameters is survey-dependent and an application of the NLA-M model to DES data is beyond the scope of this work, we model DES IA with the NLA-z model, which is the fiducial IA model in the Hybrid analysis pipeline of DES+KiDS. However, as shown in W25, changes in the IA modelling only have a minor impact

Table 5. Consistency metrics for the combination of KiDS-Legacy with external data.

	Tier 1		Tier 2			Tier 3	
External data set	$\log_{10} R$	$N_{\sigma,S}$	$N_\sigma(\Delta\Omega_m)$	$N_\sigma(\Delta S_8)$	$N_\sigma(\Delta(\Omega_m, S_8))$	$p(A\|B)$	$p(B\|A)$
DESI Y1 BAO	0.24	0.26	0.45	0.55	0.35	0.46	0.49
eBOSS DR16	0.28	0.29	0.22	0.41	0.19	0.47	0.59
Pantheon+	0.40	0.44	0.29	0.23	0.13	0.47	0.80
DESI Y1 BAO + Pantheon+	0.32	0.45	0.06	0.90	0.10	0.47	0.77
eBOSS DR16 + Pantheon+	0.34	0.44	0.06	0.44	0.05	0.47	0.79
DES Y3	1.14	0.31	0.81	0.58	0.05	0.35	0.30
Planck 2018	0.99	0.77	0.17	0.61	0.15	0.47	0.57

Notes. The third and fourth columns report the Bayes ratio and the tension level inferred from the suspiciousness, respectively. The next three columns present the results of the tier 2 test for Ω_m, S_8, and their combination. The final two columns shows the tension level arising from the tier 3 test in terms of the p-value for data vector predictions for the external data data set listed in the second column inferred from the KiDS-Legacy TPD (column 8) and vice versa (column 9).

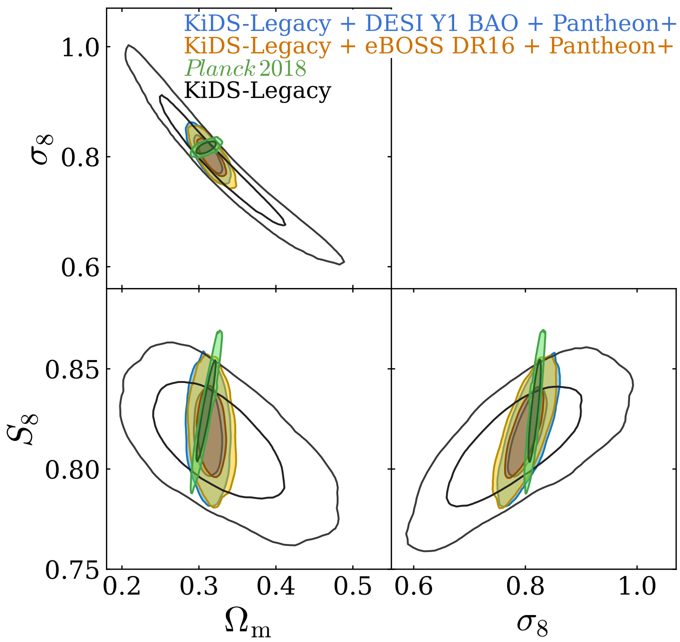

Fig. 10. Marginalised constraints for the joint distributions of Ω_m, σ_8, and S_8 from KiDS-Legacy cosmic shear data (black), its combination with Pantheon+ SN Ia data and DESI Y1 BAO data (blue), and its combination with Pantheon+ SN Ia data and eBOSS DR16 BAO and RSD data (orange). These results can be compared to CMB constraints (green) inferred with the compressed *Planck* likelihood by Prince & Dunkley (2019). The inner and outer contours of the marginalised posteriors correspond to the 68% and 95% credible intervals, respectively.

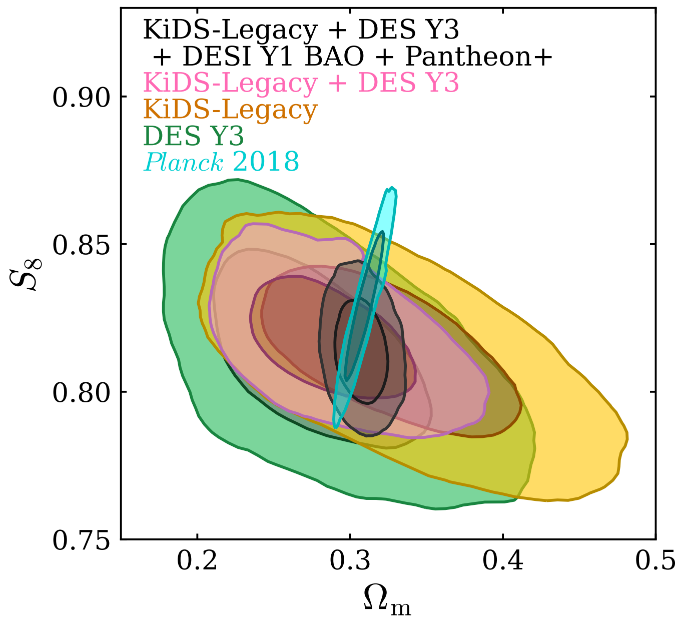

Fig. 11. Marginalised cosmic shear constraints for the joint distribution of Ω_m and S_8 from KiDS-Legacy (yellow), DES-Y3 (green), and their combination (pink). These results can be compared to the CMB posterior inferred with the compressed *Planck* likelihood by Prince & Dunkley (2019). The black contour shows constraints from a joint analysis of KiDS-Legacy with DES Y3, DESI Y1 BAO, and Pantheon+ data. The inner and outer contours of the marginalised posteriors correspond to the 68% and 95% credible intervals, respectively.

on the cosmological constraints in KiDS-Legacy and therefore we do not expect the choice of IA model to make a significant impact on the consistency analysis between KiDS and DES. The respective posteriors in the combined parameter space of S_8 and Ω_m for KiDS-Legacy (yellow), DES Y3 (green), and their combination (pink) are illustrated in Fig. 11 and their consistency is quantified in Table 5. Overall, we find an agreement between both surveys up to $N_\sigma = 0.81$, which is reported by the parameter space test in Ω_m. Thus, we consider the cosmological constraints from both surveys to be consistent, which is in agreement with the earlier study conducted with KiDS-1000 data. The joint analysis of KiDS-Legacy + DES Y3 yields

$$\begin{aligned} S_8 &= 0.818^{+0.012}_{-0.014} \\ \sigma_8 &= 0.838^{+0.059}_{-0.060} \\ \Omega_m &= 0.277^{+0.042}_{-0.028} \,. \end{aligned} \tag{21}$$

Compared to the earlier study of joint KiDS + DES data, this corresponds to a 26% reduction in uncertainty on S_8 and a shift towards higher S_8 by about 1σ, which can be attributed to the preference for higher S_8 in the KiDS-Legacy data set.

The consistency analysis between KiDS-Legacy and various probes of the low-redshift Universe individually signifies a good agreement. Therefore, this enables a joint analysis with our compilation of external data sets. We note that we do not consider a joint analysis with both DESI Y1 BAO and eBOSS DR16 data given the overlapping survey footprints. Therefore, we conduct a joint analysis of KiDS-Legacy with DES Y3, DESI Y1 BAO, and Pantheon+ data, yielding

$$
\begin{aligned}
S_8 &= 0.814^{+0.011}_{-0.012} \\
\sigma_8 &= 0.802^{+0.022}_{-0.018} \\
\Omega_{\rm m} &= 0.307^{+0.011}_{-0.011} \,.
\end{aligned} \tag{22}
$$

This represents a 38% improvement in constraining power on S_8 and a 75% improvement in σ_8 compared to the fiducial constraints with KiDS-Legacy.

In light of the apparent S_8 tension that was reported in earlier cosmic shear studies, we quantify the consistency between our KiDS-Legacy cosmic shear constraints and CMB measurements from *Planck*. We conduct a consistency analysis, assuming independence between both surveys, and quantify the consistency in Table 5. For modelling the CMB power spectra we employ the CosmoPower emulator (Spurio Mancini et al. 2022), which was shown to reproduce the fiducial *Planck* parameter constraints with similar accuracy to common Boltzmann solvers. We find no evidence for a disagreement between KiDS-Legacy cosmic shear measurements and *Planck* CMB data with $N_\sigma = 0.77$ inferred with the suspiciousness statistic. Moreover, we find S_8 to be in agreement at 0.61σ between both surveys. This removes the S_8 tension, which was found to be significant at the $\sim 2\sigma$ level in earlier KiDS studies. Combined analyses of probes of the early Universe, such as CMB measurements, and probes of the late Universe, such as cosmic shear, are commonly employed in studies of extended cosmological models beyond ΛCDM. A detailed study of extended cosmological models is beyond the scope of this work, and we therefore leave the combined analysis of KiDS-Legacy and *Planck* data for a forthcoming publication.

6. Summary and conclusions

In this analysis, we have demonstrated that the KiDS-Legacy cosmic shear data exhibits a high level of internal consistency. We find agreement between the cosmological constraints from: different tomographic redshift bins; the auto- and cross-correlation measurements; the COSEBIs, band power spectra and 2PCF statistics; the North and South regions within the KiDS footprint; red and blue galaxies. Our three tiers of consistency metrics use Bayesian evidence, measured shifts in multi-dimensional parameter space and translated posterior distributions to quantify consistency between different data splits. From our range of consistency tests it is worth highlighting two key results. In a red-blue galaxy split analysis, we confirm the results of previous studies that find strong intrinsic galaxy alignments between red early-type galaxies and no significant alignment in the blue galaxy population. This supports our decision to adopt a new colour-dependent NLA-M model in our primary KiDS-Legacy analysis. In our analysis of the cosmic shear signal measured across the six tomographic redshift bins, we find a $N_\sigma < 1.39$ consistency. This represents a marked improvement over previous KiDS analyses where the second tomographic bin, covering a range of $0.3 < z_{\rm B} \leq 0.5$, was identified as a significant outlier. We credit this improvement to advances in redshift calibration methodology and the enhanced spectroscopic data set adopted for KiDS-Legacy (see Wright et al. 2025a, for details).

In Wright et al. (2025b) we present our fiducial cosmic shear analysis, finding $S_8 = 0.815^{+0.016}_{-0.021}$. We use our three-tier analysis to demonstrate external consistency between this result and BAO constraints[7] from DESI DR1 combined with the Pantheon+ SN Ia compilation. We infer a joint constraint of $S_8 = 0.818^{+0.015}_{-0.014}$, representing a 22% reduction in uncertainty on this parameter over our fiducial result. KiDS-Legacy is also shown to be consistent with cosmic shear data from DES Y3, where a joint analysis of the two surveys finds $S_8 = 0.818^{+0.012}_{-0.014}$. Combining the KiDS, DES, DESI and Pantheon data sets we deliver a 1.4% precision measurement of $S_8 = 0.814^{+0.011}_{-0.012}$. This result is consistent with S_8 measurements of the cosmic microwave background by *Planck*, which is in 0.77σ agreement with KiDS-Legacy.

KiDS-Legacy cosmic shear data exhibits a high level of internal and external consistency resulting from significant improvements in the data reduction and scientific analyses of KiDS since its inception over a decade ago. Upcoming cosmology experiments will be required to pass the three tiers of stringent internal consistency analyses that we have presented here, and as such this analysis provides a useful blueprint for future studies.

Acknowledgements. BS, ZY, and CH 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. AHW is supported by the Deutsches Zentrum für Luft- und Raumfahrt (DLR), made possible by the Bundesministerium für Wirtschaft und Klimaschutz, and acknowledges funding from the German Science Foundation DFG, via the Collaborative Research Center SFB1491 "Cosmic Interacting Matters - From Source to Signal". PB acknowledges financial support from the Canadian Space Agency (Grant No. 23EXPROSS1) and the Waterloo Centre for Astrophysics. LL is supported by the Austrian Science Fund (FWF) [ESP 357-N]. 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. HHo, SSL, and 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. LM 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". MvWK acknowledges the support by the UKSA and STFC (grant no. ST/X001075/1). AD, HHi, CM, and RR are supported by an ERC Consolidator Grant (No. 770935). HHi is supported by a DFG Heisenberg grant (Hi 1495/5-1), the DFG Collaborative Research Center SFB1491, and the DLR project 50QE2305. MB, PJ, and AJW are supported by the Polish National Science Center through grant no. 2020/38/E/ST9/00395. MB is also supported by the Polish National Science Center through grant no. 2020/39/B/ST9/03494. LP acknowledges support from the DLR grant 50QE2002. 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). 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 Northrhine Westphalia. CG is funded by the MICINN project PID2022-141079NB-C32. CH acknowledges support from the UK Science and Technology Facilities Council (STFC) under grant ST/V000594/1. 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 RGSR2222268 and ICAR1231094. CM acknowledges support from the Beecroft Trust, the Spanish Ministry of Science under the grant number PID2021-128338NB-I00. JHD acknowledges support from an STFC Ernest Rutherford Fellowship (project reference ST/S004858/1). MR acknowledges financial support from the INAF grant 2022". 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. 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.

Kilo-Degree Survey: 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.

[7] Similar results are found when combining KiDS-Legacy with BAO and RSD galaxy clustering data from eBOSS DR16, where $S_8 = 0.819^{+0.014}_{-0.015}$.

Dark Energy Spectroscopic Instrument: This research used data obtained with the Dark Energy Spectroscopic Instrument (DESI). DESI construction and operations is managed by the Lawrence Berkeley National Laboratory. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High-Energy Physics, under Contract No. DE–AC02–05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico (CONACYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: www.desi.lbl.gov/collaborating-institutions. The DESI collaboration is honored to be permitted to conduct scientific research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. National Science Foundation, the U.S. Department of Energy, or any of the listed funding agencies.
SDSS-IV: Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is www.sdss4.org. SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard & Smithsonian, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.
Dark Energy Survey: This project used public archival data from the Dark Energy Survey (DES). Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology FacilitiesCouncil of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the OzDES Membership Consortium, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University. Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. BG acknowledges support from the UKRI Stephen Hawking Fellowship (grant reference EP/Y017137/1).
Planck: Based on observations obtained with Planck (http://www.esa.int/Planck), an ESA science mission with instruments and contributions directly funded by ESA Member States, NASA, and Canada.
Software: The figures in this work were created with matplotlib (Hunter 2007) and ChainConsumer (Hinton 2016), making use of the NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), pandas (Wes McKinney 2010), Cosmosis (Zuntz et al. 2015), Nautilus (Lange 2023), camb (Lewis et al. 2000; Howlett et al. 2012), and CosmoPower (Spurio Mancini et al. 2022) software packages.
Author Contributions: All authors contributed to the development and writing of this paper. The authorship list is given in three groups: the lead authors (BS,AHW), 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 since its inception.

References

Abdalla, E., Abellán, G. F., Aboubrahim, A., et al. 2022, Journal of High Energy Astrophysics, 34, 49
Adame, A. G., Aguilar, J., Ahlen, S., et al. 2025, J. Cosmology Astropart. Phys., 2025, 021
Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4
Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617
Alam, S., Aubert, M., Avila, S., et al. 2021, Phys. Rev. D, 103, 083533
Amon, A. & Efstathiou, G. 2022, MNRAS, 516, 5355
Amon, A., Gruen, D., Troxel, M. A., et al. 2022, Phys. Rev. D, 105, 023514
Amon, A., Robertson, N. C., Miyatake, H., et al. 2023, MNRAS, 518, 477
Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, A&A, 645, A104
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., Refregier, A. R., & Ellis, R. S. 2000, MNRAS, 318, 625
Bautista, J. E., Paviot, R., Vargas Magaña, M., et al. 2021, MNRAS, 500, 736
Becker, M. R. & Rozo, E. 2016, MNRAS, 457, 304
Benítez, N. 2000, ApJ, 536, 571
Bishop, C. M. 2006, Pattern Recognition and Machine Learning, Information science and statistics (Springer), 738
Bridle, S. & King, L. 2007, New Journal of Physics, 9, 444
Brout, D., Scolnic, D., Popovic, B., et al. 2022, ApJ, 938, 110
Chaussidon, E., Yèche, C., Palanque-Delabrouille, N., et al. 2023, ApJ, 944, 107
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
Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
de Jong, J. T. A., Verdoes Kleijn, G. A., Boxhoorn, D. R., et al. 2015, A&A, 582, A62
de Jong, J. T. A., Verdoes Kleijn, G. A., Erben, T., et al. 2017, A&A, 604, A134
de Mattia, A., Ruhlmann-Kleider, V., Raichoor, A., et al. 2021, MNRAS, 501, 5616
DESI Collaboration, Abareshi, B., Aguilar, J., et al. 2022, AJ, 164, 207
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv e-prints, arXiv:1611.00036
Doux, C., Baxter, E., Lemos, P., et al. 2021, MNRAS, 503, 2688
du Mas des Bourboux, H., Rich, J., Font-Ribera, A., et al. 2020, ApJ, 901, 153
Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32
Efstathiou, G. & Lemos, P. 2018, MNRAS, 476, 151
Euclid Collaboration, Mellier, Y., Abdurro'uf, et al. 2024, arXiv e-prints, arXiv:2405.13491
Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, MNRAS, 467, 1627
Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150
Fortuna, M. C., Dvornik, A., Hoekstra, H., et al. 2025, A&A, 694, A322
Fortuna, M. C., Hoekstra, H., Joachimi, B., et al. 2021, MNRAS, 501, 2983
García-García, C., Zennaro, M., Aricò, G., Alonso, D., & Angulo, R. E. 2024, J. Cosmology Astropart. Phys., 2024, 024
Georgiou, C., Chisari, N. E., Bilicki, M., et al. 2025, arXiv e-prints, arXiv:2502.09452
Georgiou, C., Johnston, H., Hoekstra, H., et al. 2019, A&A, 622, A90
Gil-Marín, H., Bautista, J. E., Paviot, R., et al. 2020, MNRAS, 498, 2492
Gong, Y., Liu, X., Cao, Y., et al. 2019, ApJ, 883, 203
Hahn, C., Wilson, M. J., Ruiz-Macias, O., et al. 2023, AJ, 165, 253
Handley, W. & Lemos, P. 2019a, Phys. Rev. D, 100, 023512
Handley, W. & Lemos, P. 2019b, Phys. Rev. D, 100, 043504
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357
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
Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, A&A, 647, A124
Hinton, S. R. 2016, The Journal of Open Source Software, 1, 00045

Hirata, C. M., Mandelbaum, R., Ishak, M., et al. 2007, MNRAS, 381, 1197
Hou, J., Sánchez, A. G., Ross, A. J., et al. 2021, MNRAS, 500, 1201
Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, J. Cosmology Astropart. Phys., 2012, 027
Howlett, C., Ross, A. J., Samushia, L., Percival, W. J., & Manera, M. 2015, MNRAS, 449, 848
Hu, W. 1999, ApJ, 522, L21
Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jeffreys, H. 1939, The Theory of Probability, Oxford Classic Texts in the Physical Sciences (OUP Oxford)
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
Johnston, H., Georgiou, C., Joachimi, B., et al. 2019, A&A, 624, A30
Kaiser, N. 1992, ApJ, 388, 272
Kaiser, N., Wilson, G., & Luppino, G. A. 2000, arXiv e-prints, astro
Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, A&A, 624, A92
Kilbinger, M., Heymans, C., Asgari, M., et al. 2017, MNRAS, 472, 2126
Köhlinger, F., Joachimi, B., Asgari, M., et al. 2019, MNRAS, 484, 3126
Krause, E., Fang, X., Pandey, S., et al. 2021, arXiv e-prints, arXiv:2105.13548
Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, A&A, 625, A2
Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, MNRAS, 454, 3500
Kullback, S. & Leibler, R. A. 1951, The Annals of Mathematical Statistics, 22, 79
Lange, J. U. 2023, MNRAS, 525, 3181
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
Li, S.-S., Kuijken, K., Hoekstra, H., et al. 2021, A&A, 646, A175
Li, S.-S., Kuijken, K., Hoekstra, H., et al. 2023a, A&A, 670, A100
Li, X., Zhang, T., Sugiyama, S., et al. 2023b, Phys. Rev. D, 108, 123518
Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, MNRAS, 390, 118
Longley, E. P., Chang, C., Walter, C. W., et al. 2023, MNRAS, 520, 5016
Louca, A. J. & Sellentin, E. 2020, The Open Journal of Astrophysics, 3, 11
LoVerde, M. & Afshordi, N. 2008, Phys. Rev. D, 78, 123506
Masters, D., Capak, P., Stern, D., et al. 2015, ApJ, 813, 53
McCullough, J., Amon, A., Legnani, E., et al. 2024, arXiv e-prints, arXiv:2410.22272
Miller, L., Heymans, C., Kitching, T. D., et al. 2013, MNRAS, 429, 2858
Neveux, R., Burtin, E., de Mattia, A., et al. 2020, MNRAS, 499, 210
Oehl, V. & Tröster, T. 2024, arXiv e-prints, arXiv:2407.08718
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Prince, H. & Dunkley, J. 2019, Phys. Rev. D, 100, 083502
Raichoor, A., de Mattia, A., Ross, A. J., et al. 2021, MNRAS, 500, 3254
Raichoor, A., Moustakas, J., Newman, J. A., et al. 2023, AJ, 165, 126
Raveri, M., Zacharegkas, G., & Hu, W. 2020, Phys. Rev. D, 101, 103527
Reischke, R., Unruh, S., Asgari, M., et al. 2024, arXiv e-prints, arXiv:2410.06962
Ross, A. J., Samushia, L., Howlett, C., et al. 2015, MNRAS, 449, 835
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
Schneider, P., Eifler, T., & Krause, E. 2010, A&A, 520, A116
Schneider, P., van Waerbeke, L., Kilbinger, M., & Mellier, Y. 2002, A&A, 396, 1
Scolnic, D., Brout, D., Carr, A., et al. 2022, ApJ, 938, 113
Secco, L. F., Samuroff, S., Krause, E., et al. 2022, Phys. Rev. D, 105, 023515
Sellentin, E., Heymans, C., & Harnois-Déraps, J. 2018, MNRAS, 477, 4879
Sipp, M., Schäfer, B. M., & Reischke, R. 2021, MNRAS, 501, 683
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, arXiv:1503.03757
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. 2002, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64, 583
Spurio Mancini, A., Piras, D., Alsing, J., Joachimi, B., & Hobson, M. P. 2022, MNRAS, 511, 1771
Tamone, A., Raichoor, A., Zhao, C., et al. 2020, MNRAS, 499, 5527
The Dark Energy Survey Collaboration. 2005, arXiv e-prints, arXiv:0510346
Tugendhat, T. M., Reischke, R., & Schäfer, B. M. 2020, MNRAS, 494, 2969
van Uitert, E., Joachimi, B., Joudaki, S., et al. 2018, MNRAS, 476, 4662
Van Waerbeke, L., Mellier, Y., Erben, T., et al. 2000, A&A, 358, 30
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261
Wes McKinney. 2010, in Proceedings of the 9th Python in Science Conference, ed. Stéfan van der Walt & Jarrod Millman, 56 – 61
Wittman, D. M., Tyson, J. A., Kirkman, D., Dell'Antonio, I., & Bernstein, G. 2000, Nature, 405, 143
Wright, A. H., Hildebrandt, H., Kuijken, K., et al. 2019, A&A, 632, A34
Wright, A. H., Hildebrandt, H., van den Busch, J. L., & Heymans, C. 2020a, A&A, 637, A100
Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2020b, A&A, 640, L14
Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2025a
Wright, A. H., Kuijken, K., Hildebrandt, H., et al. 2024, A&A, 686, A170
Wright, A. H., Stölzner, B., Asgari, M., et al. 2025b
Yan, Z., Wright, A. H., Elisa Chisari, N., et al. 2025, A&A, 694, A259
Zhou, R., Dey, B., Newman, J. A., et al. 2023, AJ, 165, 58
Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45

[1] Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing, 44780 Bochum, Germany
[2] School of Mathematics, Statistics and Physics, Newcastle University, Herschel Building, NE1 7RU, Newcastle-upon-Tyne, United Kingdom
[3] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, United Kingdom
[4] Leiden Observatory, Leiden University, P.O.Box 9513, 2300RA Leiden, The Netherlands
[5] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom
[6] Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany
[7] Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[8] Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[9] Institute for Theoretical Physics, Utrecht University, Princetonplein 5, 3584CC Utrecht, The Netherlands
[10] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Av. Complutense 40, E-28040 Madrid, Spain
[11] Institute of Cosmology & Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth, PO1 3FX, United Kingdom
[12] Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria
[13] Dipartimento di Fisica e Astronomia "Augusto Righi" - Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, I-40129 Bologna, Italy
[14] Istituto Nazionale di Astrofisica (INAF) - Osservatorio di Astrofisica e Scienza dello Spazio (OAS), via Piero Gobetti 93/3, I-40129 Bologna, Italy
[15] Istituto Nazionale di Fisica Nucleare (INFN) - Sezione di Bologna, viale Berti Pichat 6/2, I-40127 Bologna, Italy
[16] INAF - Osservatorio Astronomico di Padova, via dell'Osservatorio 5, 35122 Padova, Italy
[17] Institute for Computational Cosmology, Ogden Centre for Fundament Physics - West, Department of Physics, Durham University, South Road, Durham DH1 3LE, United Kingdom
[18] Centre for Extragalactic Astronomy, Ogden Centre for Fundament Physics - West, Department of Physics, Durham University, South Road, Durham DH1 3LE, United Kingdom
[19] Center for Theoretical Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668 Warsaw, Poland
[20] Institute for Particle Physics and Astrophysics, ETH Zürich, Wolfgang-Pauli-Strasse 27, 8093 Zürich, Switzerland
[21] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
[22] Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, United Kingdom
[23] Donostia International Physics Center, Manuel Lardizabal Ibilbidea, 4, 20018 Donostia, Gipuzkoa, Spain
[24] The Oskar Klein Centre, Department of Physics, Stockholm University, AlbaNova University Centre, SE-106 91 Stockholm, Sweden
[25] Imperial Centre for Inference and Cosmology (ICIC), Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2AZ, United Kingdom
[26] Zentrum für Astronomie, Universität Heidelberg, Philosophenweg 12, D-69120 Heidelberg, Germany
[27] Institute for Theoretical Physics, Philosophenweg 16, D-69120 Heidelberg, Germany
[28] Department of Physics "E. Pancini" University of Naples Federico II C.U. di Monte Sant'Angelo Via Cintia, 21 ed. 6, 80126 Naples, Italy

[29] INAF – Osservatorio Astronomico di Capodimonte, Salita Moiariello 16, I-80131 Napoli, Italy
[30] INFN, Sez. di Napoli, via Cintia, 80126, Napoli, Italy

Appendix A: Catalogue level splits: data properties and B-mode analysis

In this Appendix, we summarise the data properties of the KiDS-Legacy catalogue split into mutually exclusive subsets, which are analysed in Sect. 4.2. Table A.1 lists the redshift range, the fraction of sources with respect to the total number of sources, the effective number density, the ellipticity dispersion, the shift in the mean of the redshift distribution, and the multiplicative shear bias. We provide these data for the fiducial catalogue, a catalogue split by hemisphere, and two catalogues split into samples of red and blue galaxies. The redshift distribution per tomographic bin for each catalogue level split is illustrated in Fig. A.1.

For the split by hemisphere, we find the ellipticity dispersion, the shift in the mean of the redshift distribution per bin, and the multiplicative shear bias to be in agreement at the 1σ level between both patches. Furthermore, the redshift distributions per patch in each tomographic bin show a good agreement between hemispheres.

For the colour-based split, we define red galaxies via a selection on the spectral type $T_{\rm B}$ reported by the BPZ code. We adopt a threshold of $T_{\rm B} \leq 1.9$ from appendix B of W25, which selects objects with contributions from an elliptical galaxy spectrum (template E1). For this split, the red galaxy sample only encompasses the first five tomographic bins since the catalogue only contains very few red galaxies with $z_{\rm B} > 1.14$, making it challenging to calibrate the redshift distribution of the sixth bin. However, given the vanishingly low signal-to-noise ratio of such a sparsely populated bin, we do not expect the exclusion of this bin to make an impact on the analysis. By contrast, Li et al. (2021) previously performed a consistency test between red and blue galaxies with cosmic shear data from the third KiDS data release (KV450, de Jong et al. 2017; Wright et al. 2019), defining blue galaxies via a threshold of $T_{\rm B} \leq 3.0$. This selection additionally encompasses objects with contributions from two spiral galaxy templates (Sbc and Sdc), while objects with contributions of irregular and starburst galaxy spectra (templates Im, SB2, and SB3) are labelled as blue galaxies. Originally, this threshold was chosen in order to ensure a similar constraining power per data subset, although this no longer holds in KiDS-Legacy given the addition of high redshift galaxies, which are predominantly blue.

To first order, we expect cosmic shear to only produce E-mode signals, making B-mode signals negligible for current stage-III surveys. Therefore, B modes are a useful quantity for null-tests for residual systematics in the cosmic shear measurement. Here, we quantify the significance of the B modes for the split catalogues and refer to appendix D in W25 for a discussion of B modes in the fiducial catalogue. We employ COSEBIs as test statistic since it allows for a clean separation of E and B modes. To quantify the significance of the B mode, we compute the χ^2 assuming the null hypothesis and compute the corresponding p-value. Here, the p-value is equal to the probability of producing a B mode that is more significant than the observed signal, assuming that the B mode is randomly drawn from a Gaussian distribution with zero mean.

For each subset of the catalogue, we compute the first six COSEBIs B modes and quantify the significance for the full data vector consisting of all 21 combinations of the six tomographic redshift bins. Additionally, we quantify the B mode significance individually for each tomographic bin combination. The measured B-mode signal for the three catalogue level splits are presented in Figs. A.2, A.3, and A.4. All p-values pass our required threshold of $p > 0.01$, which is the community standard for B-mode tests (Dark Energy Survey and Kilo-Degree Survey Collaboration et al. 2023), for both the individual tomographic bin combination and for the combination of all tomographic bins.

Appendix B: Effective number of constrained parameters

For the calculation of the tension probability with the suspiciousness statistic, we require the difference between the effective number of constrained parameters N_Θ in the fiducial and split cosmological model, as defined in Eq. (7). We follow the methodology of Joachimi et al. (2021) and infer N_Θ from mock realisations of the cosmic shear data vector. Assuming a Planck Collaboration et al. (2020) cosmology and our fiducial data covariance matrix, we generate 1000 realisations of the data vector from a multivariate Gaussian distribution. For each mock realisation, we then maximise the posterior and obtain an estimate of the best-fit χ^2. As discussed in Joachimi et al. (2021), the distribution of $\chi^2_{\rm best}$ is well described by a χ^2-distribution with a number of degrees of freedom $k_{\rm dof} = N_{\rm d} - N_{\Theta,\chi^2}$, where $N_{\rm d}$ denotes the dimensionality of the data vector. We therefore fit the distribution of $\chi^2_{\rm best}$-values from mock data vectors to a χ^2-distribution in order to determine the effective number of degrees of freedom, denoted N_{Θ,χ^2}.

The inferred estimates of N_Θ are summarised in Table B.1 for the fiducial cosmic shear analysis setup with all summary statistics, dubbed '1cosmo', as well as for the split cosmological analysis setups with COSEBIs considered in Sect. 4. For the split between angular scales, we employ band powers and 2PCFs, as discussed in Sect. 4.1.3. Additionally, we infer N_Θ for catalogue- and statistic-level splits in a single cosmological setup. For comparison, we provide the Bayesian model dimensionality $N_{\Theta,\rm BMD}$ computed via Eq. (8). For both estimates, we list the resulting difference in N_Θ between the fiducial cosmological model and the split cosmological model, denoted by d_{χ^2} and $d_{\rm BMD}$. Overall, we find the number of constrained parameters to be higher in the split cosmological model. We attribute this to the doubling of the cosmological parameter space, where both instances of the parameters to which cosmic shear is sensitive are independently constrained.

Appendix C: Sensitivity tests of consistency metrics

In this Appendix, we test the sensitivity of the consistency metrics. In particular, we assess the impact of noise fluctuations in a consistency analysis with internally consistent data. We generate 100 realisations of the fiducial data vector from a multivariate Gaussian distribution assuming a reference cosmology with $S_8 = 0.777$ and conduct consistency analyses for a split analysis of the fifth redshift bin. For this test, we compute the corresponding number of sigma from the tier 3 PPD-based metric via Eq. (10). However, as discussed in Sect. 3.2.3, we emphasise that Doux et al. (2021) show that the PPD metric can result in p-values that are biased towards low values and therefore the level of tension can potentially be overestimated. We find that the tier 1 evidence-based metric yields a tension below 1σ (2σ) for 70% (97%) of the mocks, while the tier 2 multi-dimensional parameter metric and the tier 3 PPD metric yield tension below 1σ (2σ) for 69% (98%), and 65% (99%) of the mocks, respectively. Thus, we conclude that in the absence of internal tension,

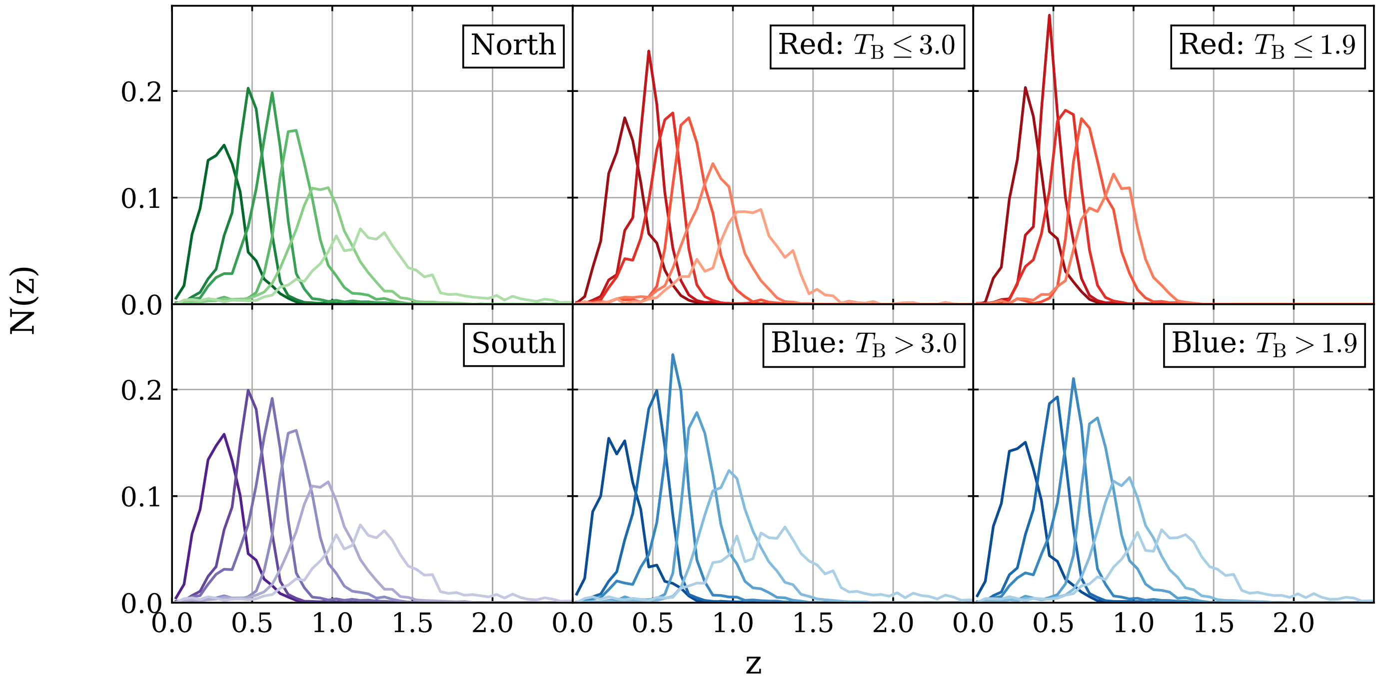

Fig. A.1. Redshift distributions per tomographic bin for catalogue level splits. Left panel: North-South split. Middle panel: Red-blue split defined via a cut on the spectral type of $T_{\rm B} = 3.0$. Right panel: Red-blue split defined via a cut on the spectral type of $T_{\rm B} = 1.9$.

the inferred consistency with each metric individually is compatible with typical noise fluctuations. The distribution of consistency metrics is shown in Fig. C.1. We find the three metrics to be correlated with $\rho_{\rm Tier1,Tier2} = 0.64$, $\rho_{\rm Tier1,Tier3} = 0.22$, and $\rho_{\rm Tier2,Tier3} = 0.15$. Additionally, we test the sensitivity of the consistency metrics to internally inconsistent data by generating noise-free mock data vectors, applying shifts in the input S_8 of the fifth bin by $\Delta S_8 = [0.01, 0.02, 0.03, 0.04]$, respectively. For each mock, we conduct a consistency analysis and display the corresponding metrics as red crosses in Fig. C.1. We find that all three metrics are capable of recovering the input tension in the data, with the estimate of the significance of the internal inconsistency being consistent between each metric.

Table A.1. Data properties per tomographic bin for catalogue level splits.

Setup	Bin	z_B	f_{total} [%]	n_{eff}[arcmin^{-2}]	σ_e	$\delta_z = z_{est} - z_{true}$	m
Fiducial	1	$0.10 < z_B \leq 0.42$	18.1	1.77	0.28	-0.026 ± 0.010	-0.023 ± 0.006
	2	$0.42 < z_B \leq 0.58$	18.0	1.65	0.27	0.013 ± 0.010	-0.016 ± 0.006
	3	$0.58 < z_B \leq 0.71$	16.6	1.50	0.29	-0.002 ± 0.010	-0.011 ± 0.007
	4	$0.71 < z_B \leq 0.90$	16.8	1.46	0.26	0.008 ± 0.010	0.020 ± 0.007
	5	$0.90 < z_B \leq 1.14$	15.8	1.35	0.28	-0.011 ± 0.010	0.030 ± 0.008
	6	$1.14 < z_B \leq 2.00$	14.6	1.07	0.30	-0.054 ± 0.011	0.045 ± 0.009
North	1	$0.10 < z_B \leq 0.42$	9.1	1.74	0.28	-0.025 ± 0.010	-0.024 ± 0.009
	2	$0.42 < z_B \leq 0.58$	9.4	1.67	0.27	0.013 ± 0.010	-0.021 ± 0.009
	3	$0.58 < z_B \leq 0.71$	8.6	1.50	0.29	-0.001 ± 0.010	-0.008 ± 0.011
	4	$0.71 < z_B \leq 0.90$	8.3	1.40	0.26	0.009 ± 0.010	0.016 ± 0.010
	5	$0.90 < z_B \leq 1.14$	7.7	1.27	0.28	-0.011 ± 0.010	0.021 ± 0.011
	6	$1.14 < z_B \leq 2.00$	6.9	0.99	0.30	-0.057 ± 0.011	0.040 ± 0.013
South	1	$0.10 < z_B \leq 0.42$	9.0	1.81	0.28	-0.025 ± 0.010	-0.017 ± 0.010
	2	$0.42 < z_B \leq 0.58$	8.6	1.62	0.27	0.013 ± 0.010	-0.010 ± 0.010
	3	$0.58 < z_B \leq 0.71$	8.1	1.49	0.29	-0.001 ± 0.010	-0.015 ± 0.012
	4	$0.71 < z_B \leq 0.90$	8.5	1.51	0.26	0.008 ± 0.010	0.018 ± 0.011
	5	$0.90 < z_B \leq 1.14$	8.1	1.43	0.28	-0.012 ± 0.010	0.026 ± 0.012
	6	$1.14 < z_B \leq 2.00$	7.6	1.16	0.30	-0.057 ± 0.011	0.047 ± 0.013
Red: $T_B \leq 3.0$	1	$0.10 < z_B \leq 0.42$	6.8	0.73	0.27	-0.008 ± 0.010	-0.031 ± 0.011
	2	$0.42 < z_B \leq 0.58$	6.0	0.60	0.25	-0.003 ± 0.010	-0.038 ± 0.011
	3	$0.58 < z_B \leq 0.71$	6.1	0.62	0.29	-0.016 ± 0.010	-0.043 ± 0.013
	4	$0.71 < z_B \leq 0.90$	6.7	0.62	0.25	-0.003 ± 0.010	-0.017 ± 0.011
	5	$0.90 < z_B \leq 1.14$	6.4	0.58	0.30	-0.014 ± 0.010	-0.043 ± 0.014
	6	$1.14 < z_B \leq 2.00$	2.7	0.21	0.30	-0.081 ± 0.011	-0.032 ± 0.020
Blue: $T_B > 3.0$	1	$0.10 < z_B \leq 0.42$	11.3	1.00	0.28	-0.025 ± 0.010	-0.016 ± 0.007
	2	$0.42 < z_B \leq 0.58$	12.0	1.03	0.28	0.030 ± 0.010	-0.005 ± 0.007
	3	$0.58 < z_B \leq 0.71$	10.6	0.87	0.28	0.019 ± 0.010	0.015 ± 0.009
	4	$0.71 < z_B \leq 0.90$	10.1	0.83	0.27	0.007 ± 0.010	0.047 ± 0.009
	5	$0.90 < z_B \leq 1.14$	9.4	0.76	0.27	-0.023 ± 0.010	0.075 ± 0.010
	6	$1.14 < z_B \leq 2.00$	11.9	0.83	0.30	-0.060 ± 0.011	0.065 ± 0.010
Red: $T_B \leq 1.9$	1	$0.10 < z_B \leq 0.42$	2.6	0.26	0.26	-0.014 ± 0.010	-0.032 ± 0.019
	2	$0.42 < z_B \leq 0.58$	3.5	0.34	0.24	-0.008 ± 0.010	-0.036 ± 0.017
	3	$0.58 < z_B \leq 0.71$	2.9	0.28	0.30	-0.015 ± 0.010	-0.047 ± 0.022
	4	$0.71 < z_B \leq 0.90$	4.0	0.36	0.25	0.001 ± 0.010	-0.016 ± 0.019
	5	$0.90 < z_B \leq 1.14$	3.0	0.25	0.34	-0.005 ± 0.010	-0.086 ± 0.028
	6	$1.14 < z_B \leq 2.00$	0.6	—	—	—	—
Blue: $T_B > 1.9$	1	$0.10 < z_B \leq 0.42$	15.5	1.48	0.28	-0.023 ± 0.010	-0.022 ± 0.006
	2	$0.42 < z_B \leq 0.58$	14.5	1.29	0.28	0.024 ± 0.010	-0.010 ± 0.006
	3	$0.58 < z_B \leq 0.71$	13.7	1.20	0.29	0.007 ± 0.010	-0.001 ± 0.008
	4	$0.71 < z_B \leq 0.90$	12.8	1.09	0.27	0.005 ± 0.010	0.030 ± 0.008
	5	$0.90 < z_B \leq 1.14$	12.8	1.08	0.27	-0.019 ± 0.010	0.044 ± 0.008
	6	$1.14 < z_B \leq 2.00$	14.1	1.03	0.30	-0.062 ± 0.011	0.048 ± 0.009

Notes. We list the index of the tomographic bin, the range in photometric redshift z_B, the fraction of sources with respect to the total number of sources, the effective number density n_{eff}, the ellipticity dispersion σ_e, the shift in the mean of the redshift distribution δ_z, and the multiplicative shear bias m.

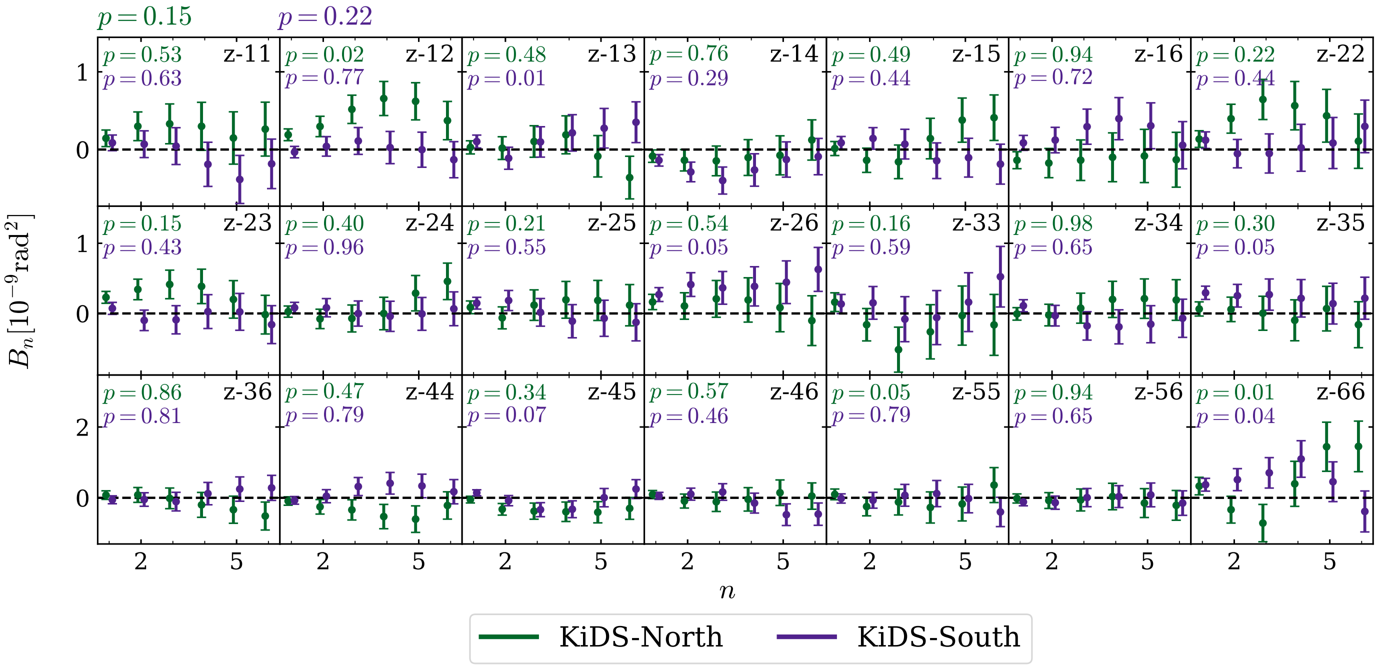

Fig. A.2. COSEBIs B-mode measurements for the North-South split catalogue. The green and purple data points show the measurements of the KiDS-North and KiDS-South sample, respectively. Each panel represents auto- or cross-correlation between tomographic bins, as indicated by the label in the top right corner. The corresponding p-value is denoted in the top left corner of each panel. The p-value of the combined data vector is given in the top left corner of the figure. For visualisation purposes, we display the discrete n modes with an offset on the x-axis. We note that the B-mode signals are highly correlated within a tomographic bin and advise against a so-called 'χ-by-eye'.

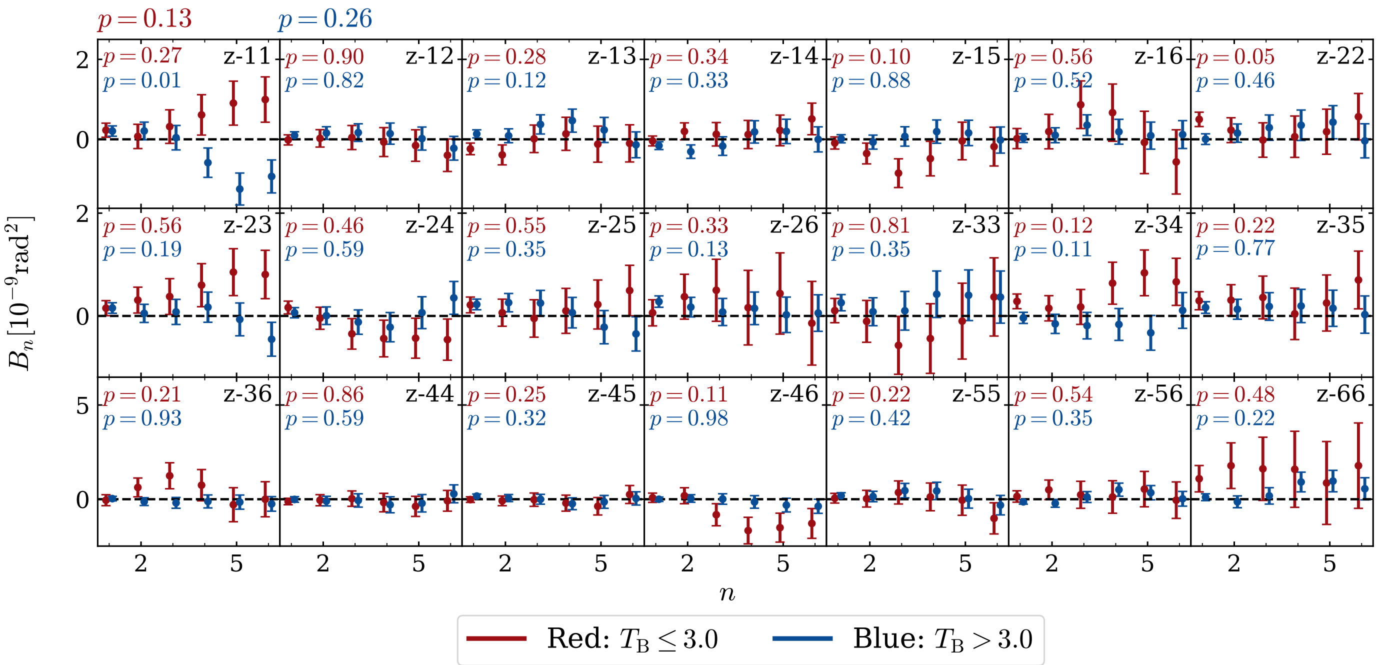

Fig. A.3. Same as Fig. A.2, but for the red-blue split catalogue defined via a cut on the spectral type T_B = 3.0.

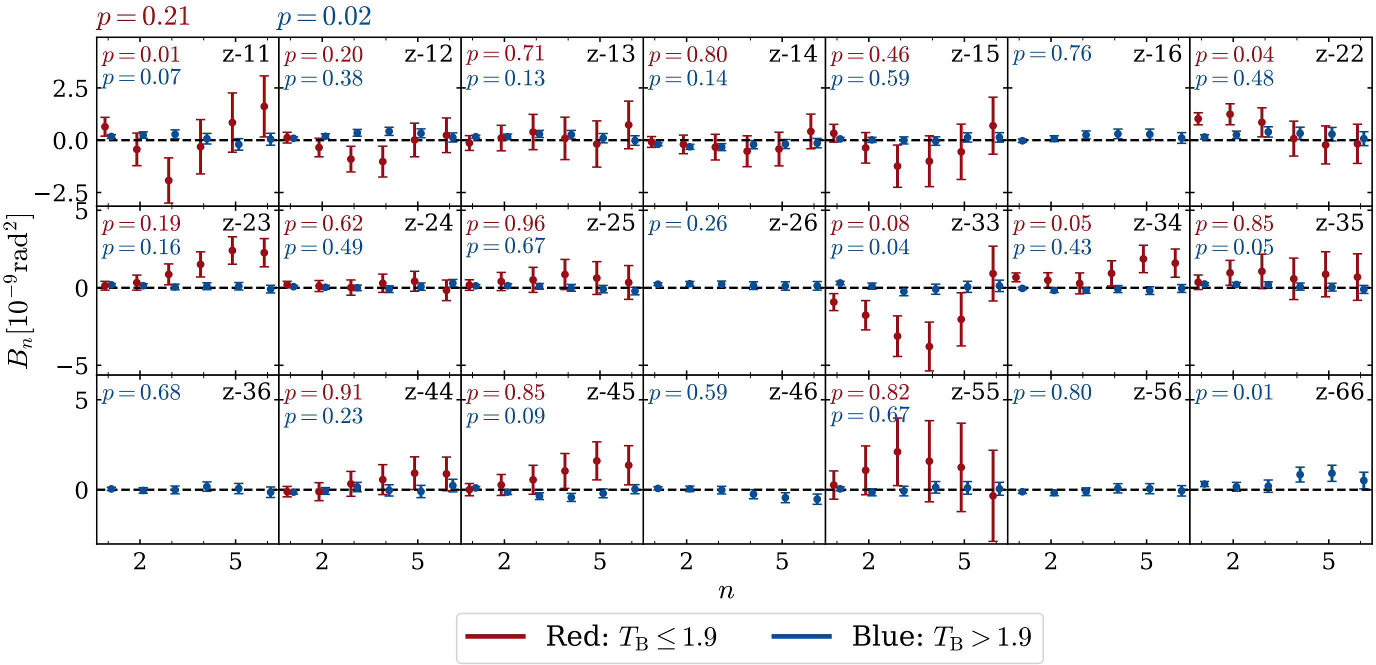

Fig. A.4. Same as Fig. A.2, but for the red-blue split catalogue defined via a cut on the spectral type $T_{\rm B}$ = 1.9. For this threshold, the catalogue only contains very few red galaxies with $z_{\rm B} > 1.14$. Therefore, the red galaxy sample only encompasses the first five tomographic bins.

Table B.1. Effective number of constrained parameters

Split type	Setup	N_{Θ,χ^2}	$N_{\Theta,\rm BMD}$	d_{χ^2}	$d_{\rm BMD}$
	Fiducial - 1cosmo (COSEBIs)	5.60	6.00	-	-
	Fiducial - 1cosmo (Band powers)	5.23	4.16	-	-
	Fiducial - 1cosmo (2PCFs)	7.79	7.79	-	-
Data vector	Bin 1	8.05	7.60	2.45	1.60
	Bin 2	8.20	5.52	2.60	-0.48
	Bin 3	8.48	7.83	2.88	1.83
	Bin 4	8.69	9.05	3.09	3.05
	Bin 5	9.01	8.51	3.41	2.51
	Bin 6	8.85	8.06	3.25	2.06
	Redshift bin AC vs CC	8.53	8.81	2.93	2.81
	$E_{n=1}$ vs $E_{n>1}$	7.35	7.22	1.75	1.22
	Small multipoles vs large multipoles (Band powers)	7.70	5.98	2.47	1.82
	Small scales vs large scales (2PCFs)	10.67	10.35	2.88	2.56
Catalogue	North vs south	9.17	7.45	3.03	1.81
	Red vs blue, $T_{\rm B}$ = 3.0	9.90	7.55	3.63	0.80
	Red vs blue, $T_{\rm B}$ = 1.9	7.03	8.70	3.40	3.80
	North vs south - 1cosmo	6.14	5.64	-	-
	Red vs blue - 1cosmo, $T_{\rm B}$ = 3.0	6.27	6.75	-	-
	Red vs blue - 1cosmo, $T_{\rm B}$ = 1.9	3.63	5.54	-	-
Statistic	COSEBIs vs Band powers	10.22	10.69	3.26	4.59
	COSEBIs vs 2PCFs	9.45	11.16	3.80	3.40
	Band powers vs 2PCFs	8.57	10.01	3.59	3.03
	COSEBIs vs Band powers - 1cosmo	6.96	6.10	-	-
	COSEBIs vs 2PCFs - 1cosmo	5.65	7.76	-	-
	Band powers vs 2PCFs - 1cosmo	4.98	6.98	-	-

Notes. For each analysis setup we provide the effective number of constrained parameters inferred via χ^2 minimisation of mock data vectors (Joachimi et al. 2021) in the third column, and the Bayesian model dimensionality (Handley & Lemos 2019a), inferred directly from the chain, in the fourth column. The first three rows show the results in an analysis with the fiducial setup, dubbed '1cosmo', in which we do not apply any splits to the data. The remaining rows present results when splitting the data into subsets. Here, we report the effective number of constrained parameters with the split cosmological model. For the splits at the catalogue level and the splits by summary statistic we additionally report the results with a shared set of parameters. The final two columns report the difference in the number of constrained parameters between the fiducial cosmological model and the split cosmological model. We employ COSEBIs as the summary statistic except when explicitly stated.

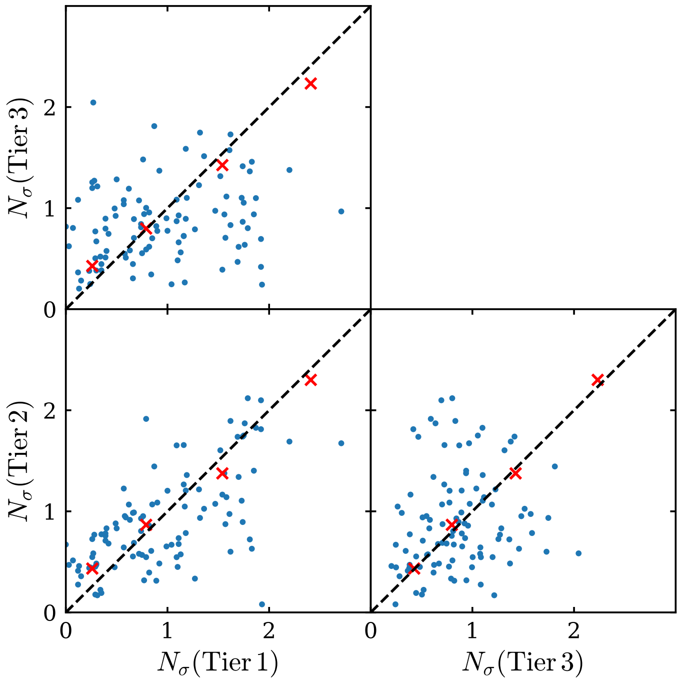

Fig. C.1. Distribution of the tier 1 evidence-based metric, the tier 2 multi-dimensional parameter metric, and the tier 3 PPD metric for 100 mock realisations of the fiducial data vector generated from a multivariate Gaussian distribution assuming a reference cosmology with $S_8 = 0.777$. The red crosses indicate the consistency metrics inferred in analyses with noise-free data vectors with systematic shifts in the input S_8 of the fifth bin by $\Delta S_8 = [0.01, 0.02, 0.03, 0.04]$.