

KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints

Catherine Heymans^{1,2*}, Tilman Tröster^{1**}, Marika Asgari¹, Chris Blake³, Hendrik Hildebrandt², Benjamin Joachimi⁴, Konrad Kuijken⁵, Chieh-An Lin¹, Ariel G. Sánchez⁶, Jan Luca van den Busch², Angus H. Wright², Alexandra Amon⁷, Maciej Bilicki⁸, Jelte de Jong⁹, Martin Crocce^{10,11}, Andrej Dvornik², Thomas Erben¹², Fedor Getman¹³, Benjamin Giblin¹, Karl Glazebrook³, Henk Hoekstra⁵, Shahab Joudaki¹⁴, Arun Kannawadi^{15,5}, Chris Lidman¹⁶, Fabian Köhlinger², Lance Miller¹⁴, Nicola R. Napolitano¹⁷, David Parkinson¹⁸, Peter Schneider¹², Huan Yuan Shan^{19,20}, and Christian Wolf¹⁶

¹ Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK

² Ruhr-Universität Bochum, Astronomisches Institut, German Centre for Cosmological Lensing (GCCL), Universitätsstr. 150, 44801, Bochum, Germany

³ Centre for Astrophysics & Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia

⁴ Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

⁵ Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, the Netherlands

⁶ Max-Planck-Institut für extraterrestrische Physik, Postfach 1312, Giessenbachstrasse 1, D-85741 Garching, Germany

⁷ Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA

⁸ Center for Theoretical Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668, Warsaw, Poland

⁹ Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, the Netherlands

¹⁰ Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain

¹¹ Institut d'Estudis Espacials de Catalunya (IEEC), Carrer Gran Capita 2, E-08034, Barcelona, Spain

¹² Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany

¹³ INAF - Astronomical Observatory of Capodimonte, Via Moiariello 16, 80131 Napoli, Italy

¹⁴ Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

¹⁵ Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA

¹⁶ Research School of Astronomy and Astrophysics, Australian National University, Canberra ACT 2600, Australia

¹⁷ University of Chinese Academy of Sciences, Beijing 100049, China

¹⁸ Korea Astronomy and Space Science Institute, 776 Daedeokdae-ro, Yuseong-gu, Daejeon 34055, Republic of Korea

¹⁹ Shanghai Astronomical Observatory (SHAO), Nandan Road 80, Shanghai 200030, China

²⁰ University of Chinese Academy of Sciences, Beijing 100049, China

July 31, 2020

ABSTRACT

We present a joint cosmological analysis of weak gravitational lensing observations from the Kilo-Degree Survey (KiDS-1000), with redshift-space galaxy clustering observations from the Baryon Oscillation Spectroscopic Survey (BOSS), and galaxy-galaxy lensing observations from the overlap between KiDS-1000, BOSS and the spectroscopic 2-degree Field Lensing Survey (2dFLenS). This combination of large-scale structure probes breaks the degeneracies between cosmological parameters for individual observables, resulting in a constraint on the structure growth parameter $S_8 = \sigma_8 \sqrt{\Omega_m/0.3} = 0.766^{+0.020}_{-0.014}$, that has the same overall precision as that reported by the full-sky cosmic microwave background observations from *Planck*. The recovered S_8 amplitude is low, however, by $8.3 \pm 2.6\%$ relative to *Planck*. This result builds from a series of KiDS-1000 analyses where we validate our methodology with variable depth mock galaxy surveys, our lensing calibration with image simulations and null-tests, and our optical-to-near-infrared redshift calibration with multi-band mock catalogues and a spectroscopic-photometric clustering analysis. The systematic uncertainties identified by these analyses are folded through as nuisance parameters in our cosmological analysis. Inspecting the offset between the marginalised posterior distributions, we find that the S_8 -difference with *Planck* is driven by a tension in the matter fluctuation amplitude parameter, σ_8 . We quantify the level of agreement between the CMB and our large-scale structure constraints using a series of different metrics, finding differences with a significance ranging between $\sim 3\sigma$, when considering the offset in S_8 , and $\sim 2\sigma$, when considering the full multi-dimensional parameter space.

Key words. gravitational lensing: weak, methods: data analysis, methods: statistical, surveys, cosmology: observations

1. Introduction

Observations of the cosmic microwave background (CMB) have delivered high-precision constraints for the cosmological pa-

rameters of the flat, cold dark matter and cosmological constant model of the Universe (Λ CDM, *Planck* Collaboration et al. 2018). With only six free parameters, this flat Λ CDM model provides an exquisite fit to observations of the anisotropies in the CMB. The same model predicts a range of different ob-

* Catherine Heymans: heymans@roe.ac.uk

** Tilman Tröster: ttr@roe.ac.uk

servables in the present day Universe, including the cosmic expansion rate (Weinberg 1972), and the distribution of, and gravitational lensing by, large-scale structures (Peebles 1980; Bartelmann & Schneider 2001; Eisenstein et al. 2005). In most cases there is agreement between the measured cosmological parameters of the flat Λ CDM model, when comparing those constrained at the CMB epoch with those constrained through a variety of lower-redshift probes (see the discussion in Planck Collaboration et al. 2018, and references therein). Recent improvements in the statistical precision of the lower-redshift probes have, however, revealed some statistically significant differences. Most notably a 4.4σ difference in the value of the Hubble constant, H_0 , has been reported using distance ladder estimates in Riess et al. (2019). If this difference cannot be attributed to systematic errors in either, or both, experiment, this result suggests that the flat Λ CDM model is incomplete.

Many extensions have been proposed to reconcile the observed differences between high- and low-redshift probes (see for example Riess et al. 2016; Poulin et al. 2018; Di Valentino et al. 2020). All, however, require additional components to the cosmological model that move it even further away from the standard model of particle physics, a model that already struggles to motivate the existence of cold dark matter and a cosmological constant. As the statistical power of the observations continues to improve, focus has moved to establishing a full understanding of all systematic errors, and the development of mitigation approaches, in preparation for the high-precision ‘full-sky’ imaging and spectroscopic cosmology surveys of the 2020’s (*Euclid*, Laureijs et al. 2011; LSST Science Collaboration et al. 2009; DESI Collaboration et al. 2016).

We present a multi-probe ‘same-sky’ analysis of the evolution of large-scale structures, using overlapping spectroscopic and optical-to-near-infrared imaging surveys. Our first observable is the weak gravitational lensing of background galaxies by foreground large-scale structures, known as ‘cosmic shear’. Our second observable is the *anisotropic clustering of galaxies* within these large-scale structures, combining measurements of both redshift-space distortions and baryon acoustic oscillations. Our third observable is the weak gravitational lensing of background galaxies by the matter surrounding foreground galaxies, known as ‘galaxy-galaxy lensing’. As these three sets of two-point statistics are analysed simultaneously, this combination of probes is usually referred to as a ‘ $3 \times 2\text{pt}$ ’ analysis.

Each observable in our multi-probe analysis is subject to systematic uncertainties. For a cosmic shear analysis, the observable is a combination of the true cosmological signal with a low-level signal arising from the intrinsic alignment of galaxies, as well as potential residual correlations in the data induced by the atmosphere, telescope and camera. The signal can also be scaled by both shear and photometric redshift measurement calibration errors (see Mandelbaum 2018, and references therein). For a galaxy clustering analysis, the observable is the true cosmological signal modulated by an uncertain, non-linear and evolving, galaxy bias function. This function maps how the galaxies trace the underlying total matter distribution (see Desjacques et al. 2018, and references therein). The cosmological clustering also needs to be accurately distinguished from artificial clustering in the galaxy sample, arising from potentially uncharacterised inhomogeneities in the target selection (see for example Ross et al. 2012). Finally, the galaxy-galaxy lensing analysis is subject to the systematics that impact both the cosmic shear and clustering analyses.

When analysing these observables in combination the different astrophysical and systematic dependencies allow for some degree of self-calibration (Bernstein & Jain 2004; Hu & Jain 2004; Bernstein 2009; Joachimi & Bridle 2010). ‘Same-sky’ surveys, in which imaging for weak lensing observables overlaps with spectroscopy for anisotropic galaxy clustering measurements, also allows for their cross-correlation. Such a survey design therefore presents a robust cosmological tool that can calibrate and mitigate systematic and astrophysical uncertainties through a series of nuisance parameters. In addition to enhanced control over systematics, this combination of probes breaks cosmological parameter degeneracies from each individual probe. For a flat Λ CDM model this leads to significantly tighter constraints on the matter fluctuation amplitude parameter, σ_8 , and the matter density parameter, Ω_m , whilst also decreasing the uncertainty on the recovered dark energy equation of state parameter in extended cosmology scenarios (Hu & Jain 2004; Gaztañaga et al. 2012).

Three variants of a joint ‘ $3 \times 2\text{pt}$ ’ analysis have been conducted to date. van Uitert et al. (2018) present a joint power-spectrum analysis of the Kilo-Degree Survey (KiDS, Kuijken et al. 2015) with the Galaxy And Mass Assembly survey (GAMA, Liske et al. 2015), incorporating projected angular clustering measurements. Joudaki et al. (2018) present a joint analysis of KiDS with the 2-degree Field Lensing Survey (2dFLenS, Blake et al. 2016) and the overlapping area in the Baryon Oscillation Spectroscopic Survey (BOSS, Alam et al. 2015), incorporating redshift-space clustering measurements. Abbott et al. (2018) present a joint real-space lensing-clustering analysis of the Dark Energy Survey (DES Y1, Drlica-Wagner et al. 2018), using a high-quality photometric redshift sample of luminous red galaxies for their projected angular clustering measurements. In all three cases a linear galaxy bias model was adopted.

In this analysis we enhance and build upon the advances of previous ‘ $3 \times 2\text{pt}$ ’ studies. We analyse the most recent KiDS data release (KiDS-1000, Kuijken et al. 2019), more than doubling the survey area from previous KiDS studies. We utilise the full BOSS area and the ‘full-shape’ anisotropic clustering measurements of Sánchez et al. (2017), incorporating information from both redshift-space distortions and the baryon acoustic oscillation as our galaxy clustering probe. We adopt a non-linear evolving galaxy bias model, derived from renormalised perturbation theory (Crocce & Scoccimarro 2006; Chan et al. 2012). We maximise the signal-to-noise in our KiDS-BOSS galaxy-galaxy lensing analysis, by including additional overlapping spectroscopy of BOSS-like galaxies from 2dFLens.

This paper is part of the KiDS-1000 series. The KiDS-1000 photometry and imaging is presented in Kuijken et al. (2019). The core weak lensing data products are presented and validated in Giblin et al. (shear measurements, 2020), and Hildebrandt et al. (redshift measurements, 2020b). Asgari et al. (2020b) conduct the cosmic shear analysis using a range of different two-point statistics, and Joachimi et al. (2020) detail the methodology behind our ‘ $3 \times 2\text{pt}$ ’ analysis, with a particular focus on pipeline validation and accurate covariance matrices. In this analysis we constrain the cosmological parameters of the flat Λ CDM model. A range of different extensions to the Λ CDM model are considered in Tröster et al. (in prep.), including varying dark energy, neutrino mass, spatial curvature and various modified gravity scenarios (Bose et al. 2020).

In this paper we review the data and provide a concise summary of the findings of the KiDS-1000 series of papers in Section 2. We present our joint cosmological constraints in Sec-

tion 3, and conclude in Section 4. Appendices tabulate the galaxy properties (A), the adopted cosmological parameter priors (B), and the cosmological parameter constraints (C). They also discuss: a series of sensitivity tests (D); the expected differences between parameter constraints for overlapping weak lensing surveys (E); a range of different ‘tension’ metrics (F); the redundancy, validation and software review for our pipeline (G); and the minor analysis additions that were included after the analysis was formally unblinded (H).

2. Data and Methodology

2.1. Surveys: KiDS, BOSS and 2dFLens

The Kilo-Degree Survey (KiDS, de Jong et al. 2013), covers 1350 deg^2 split into two fields, one equatorial and one southern. Matched-depth imaging in nine bands spans the optical, $ugri$, through to the near-infrared, $ZYJHK_s$, where the near-infrared imaging was taken as part of the KiDS partner survey VIKING (the VISTA Kilo-degree INfrared Galaxy survey, Edge et al. 2013). High-quality seeing was routinely allocated to the primary KiDS r -band VST-OmegaCAM observations, resulting in a mean r -band seeing of 0.7 arcseconds, with a time-allocated maximum of 0.8 arcseconds. This combination of full-area spatial and wavelength resolution over a thousand square degrees provides a unique weak lensing survey that allows for enhanced control of systematic errors (Giblin et al. 2020; Hildebrandt et al. 2020b). This analysis uses data from the fourth KiDS data release of 1006 deg^2 of imaging, (hence the name KiDS-1000), which has an effective area, after masking, of 777 deg^2 . KiDS is a public survey from the European Southern Observatory, with data products freely accessible through the ESO archive¹.

Giblin et al. (2020) present a series of null-tests to validate the KiDS-1000 shear catalogue in five tomographic bins spanning a photometric redshift range of $0.1 < z_B \leq 1.2$ (see Appendix A for details of the properties of each bin). Meeting their requirement that any systematic detected induces less than a 0.1σ change in the inferred cosmic shear constraints on the clustering cosmological parameter $S_8 = \sigma_8 \sqrt{\Omega_m}/0.3$, they conclude that the shear catalogue is ‘science-ready’, with no significant non-lensing B-mode distortions detected. Kuijken et al. (2015) present the catalogue-level blinding methodology that we adopted to introduce $\pm 2\sigma$ differences in the recovered value of S_8 in order to retain team ignorance over the final cosmological results until all analysis decisions were finalised (for further details on blinding see Appendix H). Hildebrandt et al. (2020b) present the KiDS-1000 photometric redshift calibration. This is determined using the self-organising map (SOM) methodology of Wright et al. (2020a), and is validated with a cross-correlation clustering analysis, following van den Busch et al. (2020). The SOM identifies and excludes any galaxies that are poorly represented in the spectroscopic calibration sample, in terms of their nine-band colours and magnitudes. The resulting ‘gold’ photometric sample, with an accurately calibrated redshift distribution, is then re-simulated in the KiDS image simulations of Kannawadi et al. (2019) in order to determine the shear calibration corrections for each tomographic bin, and an associated uncertainty (see Giblin et al. 2020; Hildebrandt et al. 2020b, for full details).

The Baryon Oscillation Spectroscopic Survey (BOSS, Alam et al. 2015), spans an effective area of 9329 deg^2 , with

spectroscopic redshifts for 1.2 million luminous red galaxies (LRG) in the redshift range $0.2 < z < 0.9$. A range of different statistical analyses of the clustering of BOSS galaxies have been used in combination with CMB measurements, to set tight constraints on extensions to the standard flat Λ CDM model (see Alam et al. 2017; eBOSS Collaboration et al. 2020, and references therein). We adopt the anisotropic clustering measurements of Sánchez et al. (2017) in this multi-probe analysis. BOSS only overlaps with the equatorial stripe of the KiDS survey, with 409 deg^2 of the BOSS survey lying within the KiDS-1000 footprint. BOSS galaxies in this overlapping region are used as lenses in our galaxy-galaxy lensing analysis, with an effective lens number density of $0.031 \text{ arcmin}^{-2}$ (see Appendix A for details). BOSS is a public survey from the third Sloan Digital Sky Survey (York et al. 2000), and we analyse data from the twelfth data release² (DR12, Alam et al. 2015).

The 2-degree Field Lensing Survey (2dFLenS, Blake et al. 2016), spans 731 deg^2 , with spectroscopic redshifts for 70 000 galaxies out to $z < 0.9$. This galaxy redshift survey from the Anglo-Australian Telescope (AAT) was designed to target areas already mapped by weak lensing surveys to facilitate ‘same-sky’ lensing-clustering analyses (Johnson et al. 2017; Amon et al. 2018; Joudaki et al. 2018; Blake et al. 2020). We use data from the 2dFLenS LRG sample that was targeted to match the BOSS-LRG selection, but with sparser sampling. 2dFLenS thus provides an additional sample of BOSS-like galaxies in the KiDS southern stripe where there is 425 deg^2 of overlap within the KiDS-1000 footprint. 2dFLenS galaxies in this overlapping region are used as lenses in our galaxy-galaxy lensing analysis, with an effective lens number density of $0.012 \text{ arcmin}^{-2}$ (see Appendix A for details). 2dFLenS was an AAT Large Program that has been made public³.

2.2. Cosmic shear

The observed cosmic shear angular power spectrum, $C_{\epsilon\epsilon}(\ell)$, measures a combination of the distortions arising from weak gravitational lensing by large-scale structures (labelled with a subscript ‘G’) with a low-level contaminating astrophysical signal arising from the intrinsic alignment of galaxies with the large-scale structures within which they are embedded (labelled with a subscript ‘I’). These contributions can be separated as

$$C_{\epsilon\epsilon}^{(ij)}(\ell) = C_{GG}^{(ij)}(\ell) + C_{GI}^{(ij)}(\ell) + C_{IG}^{(ij)}(\ell) + C_{II}^{(ij)}(\ell), \quad (1)$$

where the indices i and j indicate cross-correlations between the five tomographic source samples. The theoretical power spectra are given by Limber-approximated projections with

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

where $a, b \in \{I, G\}$, $f_K(\chi)$ is the comoving angular diameter distance and χ is the comoving radial distance which runs out to the horizon, χ_{hor} . The weight functions, $W(\chi)$, encode information about how the signal scales with the KiDS-1000 survey depth (see equations 15 and 16 of Joachimi et al. 2020). In the cases of power spectra that include intrinsic ‘I’ terms, the weight function also encodes the intrinsic galaxy alignment model, which we take to be the ‘NLA’ model from Bridle & King (2007). For Stage III surveys like KiDS-1000, this model has been shown to

² BOSS data access: data.sdss.org/sas/dr12/boss/lss/

³ 2dFLenS data access: 2dflens.swin.edu.au/data.html

¹ KiDS data access: kids.strw.leidenuniv.nl/DR4

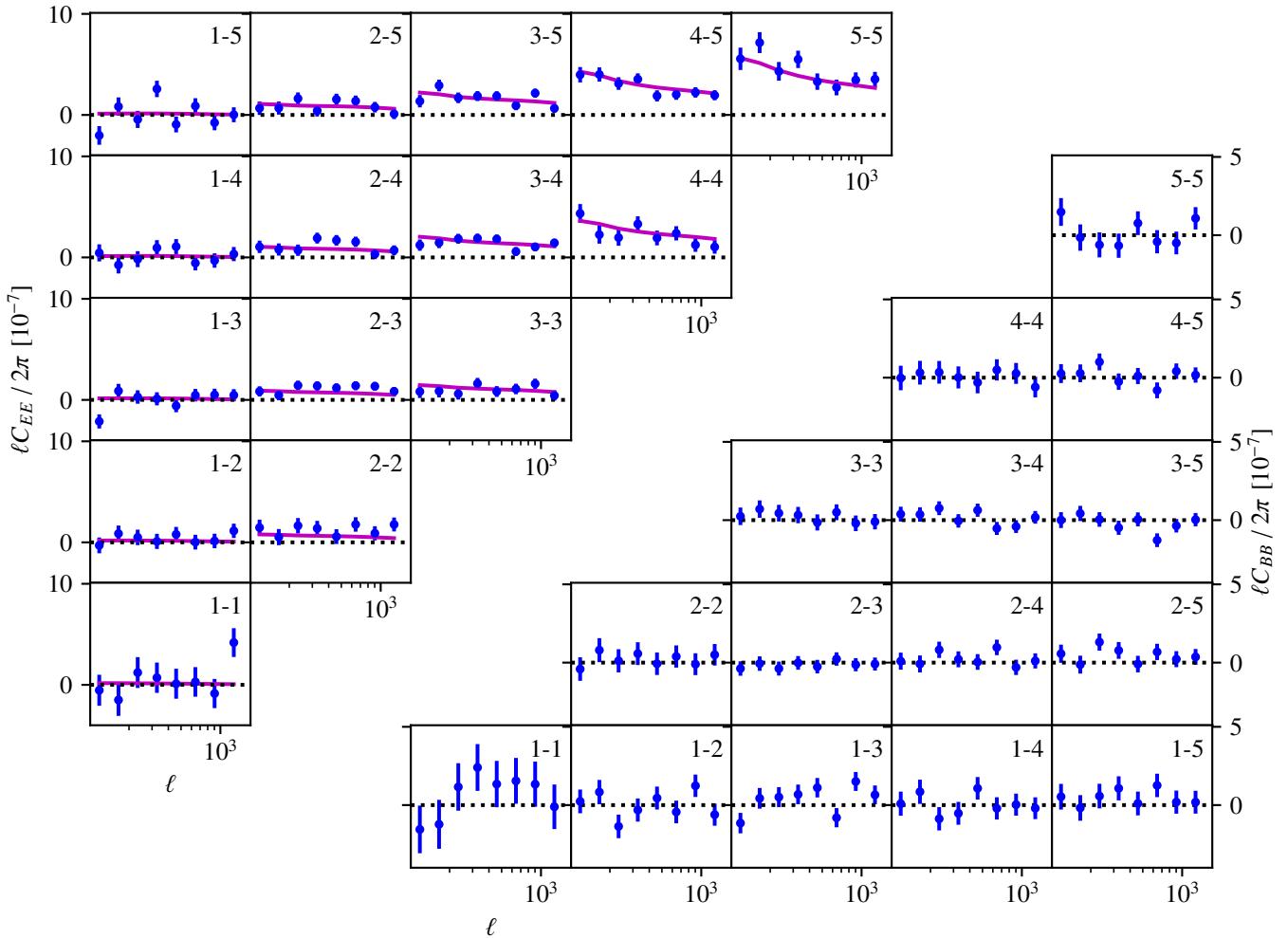


Fig. 1: KiDS-1000 cosmic shear power spectra: Tomographic band powers comparing the E-modes (upper left block) with the best-fit cosmological model from our combined multi-probe analysis. The tomographic bin combination is indicated in the upper right corner of each sub-panel. The null-test B-modes (lower right block - note the reduced ordinate scale), are consistent with zero for both the full data vector, and each bin combination individually. The errors are estimated analytically (Joachimi et al. 2020). See Sect. 3 for a discussion on the goodness-of-fit.

be sufficiently flexible, capturing the likely more complex underlying intrinsic alignment model, without biasing cosmological parameters (Fortuna et al. 2020). The cosmological information for cosmic shear power spectrum is contained in both the geometric weight functions, $W(\chi)$, and in the evolution and shape of the non-linear matter power spectrum, $P_{m,\text{nl}}(k, z)$, which we model using the halo formalism⁴ of Mead et al. (2015). Weak lensing is therefore a very valuable cosmological probe, as it is sensitive to changes in both the distance-redshift relation and to the growth of structures.

⁴ We calculate the non-linear power spectrum using HMCode (Mead et al. 2015), which is incorporated in CAMB (Lewis & Bridle 2002). Joachimi et al. (2020) demonstrate that the Mead et al. (2015) halo model prescription provides a sufficiently accurate model of the non-linear matter power spectrum into the highly non-linear regime through a comparison to weak lensing observables emulated using the N-body CosmicEmu simulations (Heitmann et al. 2014). It also has the added benefit of allowing us to marginalise over our uncertainty on the impact of baryon feedback on the shape of the non-linear total matter power spectrum (Semboloni et al. 2011; Mead et al. 2015, 2020).

We estimate the cosmic shear angular power spectrum through a linear transformation of the real-space two-point shear correlation function (Schneider et al. 2002). This approach circumvents the challenge of accurately determining the survey mask for a direct power spectrum estimate. Joachimi et al. (2020) detail the apodisation advances that we have adopted for the transformation, in addition to the modelling that we use to account for the minor differences between the theoretical expectation of the true angular power spectrum in Eq. (1) and the measured ‘band powers’.

Fig. 1 presents the Asgari et al. (2020b) KiDS-1000 cosmic shear power spectra for the auto- and cross-correlated tomographic bins. Here we have constructed both E-mode (upper left) and B-mode (lower right) band powers in order to isolate any non-lensing B-mode distortions (see equations 17 to 21 of Joachimi et al. 2020). As expected from the analysis of Giblin et al. (2020), the measured B-modes are found to be consistent with zero⁵. The measured E-modes can be compared to

⁵ Giblin et al. (2020) present a ‘COSEBIs’ B-mode analysis following Asgari et al. (2019). The alternative band power B-mode measurement,

the theoretical expectation from Eq. (1), given the best-fit set of cosmological parameters from our multi-probe analysis in Sect. 3.

2.3. Anisotropic galaxy clustering

Galaxy clustering observations probe the 3D non-linear galaxy-galaxy power spectrum and we follow Sánchez et al. (2017) in modelling this quantity based on a perturbation theory approach with

$$P_{gg}(k, z) = \sum_{\alpha, \beta} \alpha \beta P_{\alpha\beta}(k, z) + b_1 \gamma_3^- P_{b_1 \gamma_3^-}(k, z) + P_{\text{noise}}(k, z). \quad (3)$$

Here $\alpha, \beta \in [b_1, b_2, \gamma_2]$, introduce the linear and quadratic bias parameters b_1 and b_2 , in addition to the non-local bias parameters γ_2 and γ_3^- . Each power spectrum term on the right hand side of the equation is given by different convolutions of the linear matter power spectrum in appendix A of Sánchez et al. (2017). In the case of an effective linear galaxy bias model (see for example van Uitert et al. 2018; Abbott et al. 2018) only the b_1 bias parameter is considered to be non-zero and Eq. (3) reduces to $P_{gg}(k, z) = b_1^2 P_{m,nl}^{\text{pert}}(k, z)$, where $P_{m,nl}^{\text{pert}}(k, z)$ is the perturbation theory estimate of the non-linear matter power spectrum which is accurate at the two percent level to $k \lesssim 0.3h \text{ Mpc}^{-1}$ (Sánchez et al. 2017).

Sánchez et al. (2017) present the anisotropic redshift-space correlation function of galaxy clustering with the galaxy pairs separated into three ‘wedges’ equidistant⁶ in μ , where μ is the cosine of the angle between the line of sight and the line connecting the galaxy pairs. As such the 3D correlation function is measured for pairs that are either mainly transverse to the line of sight, mainly parallel to the line of sight, or placed into an intermediate sample between these two cases. The redshift-space correlation function $\xi_{gg}(s, \mu, z)$, where s is the co-moving galaxy-pair separation, is given by

$$\begin{aligned} \xi_{gg}(s, \mu, z) &= \sum_{l=0}^2 L_{2l}(\mu) \frac{(-1)^l (4l+1)}{(2\pi)^2} \int_0^\infty dk k^2 j_{2l}(ks) \\ &\times \int_{-1}^1 d\mu_1 L_{2l}(\mu_1) P_{gg,s}(k, \mu_1, z), \end{aligned} \quad (4)$$

where L_i denotes the Legendre polynomial of degree i , j_i is the spherical Bessel function of order i , and $P_{gg,s}(k, \mu, z)$ is the 3D redshift-space power spectrum that includes the non-linear real-space power spectrum, Eq. (3), and the galaxy-velocity and velocity-velocity power spectrum (see Sánchez et al. 2017, for details, including how the Alcock-Paczynski distortions are accounted for in the modelling). The same model and ‘wedge’ approach was adopted in the Fourier-space anisotropic galaxy clustering analysis of Grieb et al. (2017), finding consistent results.

presented in Fig. 1, is consistent with random noise, finding a p -value of $p = 0.68$ for the full data vector. Here p corresponds to the probability of randomly producing a noisy B-mode that is more significant than the measurements. Inspecting each individual tomographic bin combination, we find that these are also consistent with random noise, with a minimum $p = 0.02$ found for the 1_3 bin combination. A $\sim 2\sigma$ deviation is expected, given the 15 different bin combinations analysed, and we note that the bin combination outlier in this test differs from the $\sim 2\sigma$ deviation bin combination outliers in the two different COSE-BIs analyses, supporting the hypothesis that the measured B-modes are simple noise fluctuations.

⁶ For the three wedges i , the separation in μ is given by $(i-1)/3 < \mu \leq i/3$.

Fig. 2 presents the Sánchez et al. (2017) BOSS-DR12 anisotropic clustering correlation functions in three wedges, and two redshift slices⁷, with $0.2 < z \leq 0.5$, and $0.5 < z \leq 0.75$, for the scales used in this analysis with $20 < s < 160 h^{-1} \text{ Mpc}$. The measured correlation functions can be compared to the theoretical expectation given the best-fit set of cosmological parameters from our joint multi-probe analysis in Sect. 3.

2.4. Galaxy-galaxy lensing

The observed galaxy-galaxy lensing angular power spectra, $C_{ne}(\ell)$, measure a combination of weak lensing distortions around foreground galaxies (labelled with a subscript ‘gG’) with a low-level intrinsic alignment signal arising from the fraction of the source galaxy population that reside physically close to the lenses (labelled with a subscript ‘gI’). We also consider the low-level lensing-induced magnification bias (labelled with a subscript ‘mG’). These three contributions can be separated as

$$C_{ne}^{(ij)}(\ell) = C_{gG}^{(ij)}(\ell) + C_{gI}^{(ij)}(\ell) + C_{mG}^{(ij)}(\ell), \quad (5)$$

where the index i indicates the two lens galaxy samples, and the index j indicates the five tomographic source samples. The theoretical power spectra are given by Limber-approximated projections following Eq. (2), with two key differences. The first is that the lens weight function is replaced by the redshift distribution of the lenses. The second is that the non-linear matter power spectrum is replaced with the non-linear cross power spectrum between the galaxy and matter distribution $P_{gm}(k, z)$ (see equations 24 and 25 of Joachimi et al. 2020, for the full expressions). For the magnification bias power spectrum, $C_{mG}(\ell)$, we refer the reader to appendix B in Joachimi et al. (2020). We find that the inclusion or exclusion of this term has a negligible impact on our cosmological constraints, but we retain it nevertheless.

We adopt the non-linear galaxy bias model from Sánchez et al. (2017), and approximate the non-linear cross power spectrum as

$$\begin{aligned} P_{gm}(k, z) &= b_1 P_{m,nl}(k, z) + \{b_2 \mathcal{F}_{b_2}(k) - \gamma_2 \mathcal{F}_{\gamma_2}(k) \\ &- \gamma_3^- \mathcal{F}_{\gamma_3^-}(k)\} P_{m,lin}^2(k, z). \end{aligned} \quad (6)$$

Here $P_{m,nl}(k, z)$ is the non-linear matter power spectrum modelled using Mead et al. (2015), in contrast to the less accurate perturbation theory estimate $P_{m,nl}^{\text{pert}}(k, z)$ used in Eq. (3). The logarithm of the functions $\mathcal{F}_\alpha(k)$ are second-order polynomial fits that we use to model $P_\alpha(k, z_{\text{ref}})/P_{m,lin}^2(k, z_{\text{ref}})$, the ratio between the different bias terms in the full perturbation theory model in Eq. (3), and the square of the linear matter power spectrum. This approach permits a reasonable extrapolation of the Sánchez et al. (2017) perturbation model into the non-linear regime beyond $k = 0.3h \text{ Mpc}^{-1}$. This is necessary in order to carry out the redshift-weighted projection of the 3D model to estimate the 2D galaxy-galaxy lensing observable, Eq. (2). No matter which ℓ -scales we restrict our analysis to, high- k scales will contribute to all angular scales at some level (Joachimi et al. 2020; Asgari et al. 2020a). This approach also decreases the compute time of the galaxy-galaxy lensing likelihood evaluations, by several orders of magnitude, in comparison to the direct perturbative calculation.

⁷ We do not include the Sánchez et al. (2017) central redshift bin measurements in this analysis. The central bin fully overlaps with the two primary redshift bins, shown in Fig. 2, and was found not to add any significant constraining power.

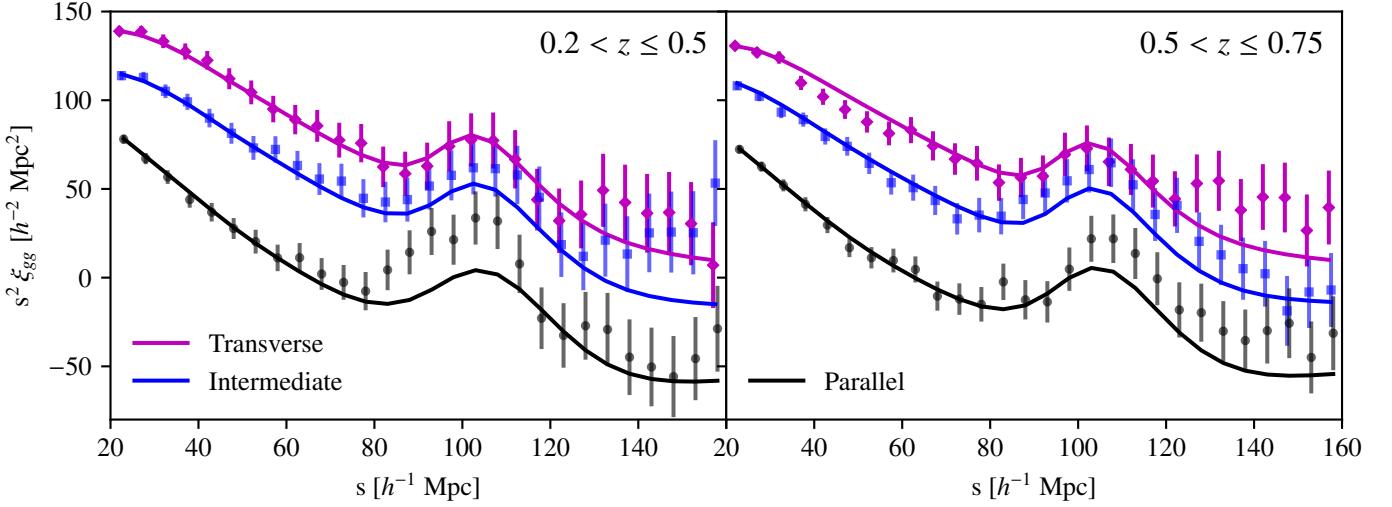


Fig. 2: BOSS-DR12 anisotropic clustering from Sánchez et al. (2017): The transverse (pink), intermediate (blue) and parallel (black) clustering wedges in two redshift bins, compared with the best-fit cosmological model from our combined multi-probe analysis. The errors, estimated from mock BOSS catalogues (Kitaura et al. 2016), are highly correlated, particularly at large scales (see Fig. A.1).

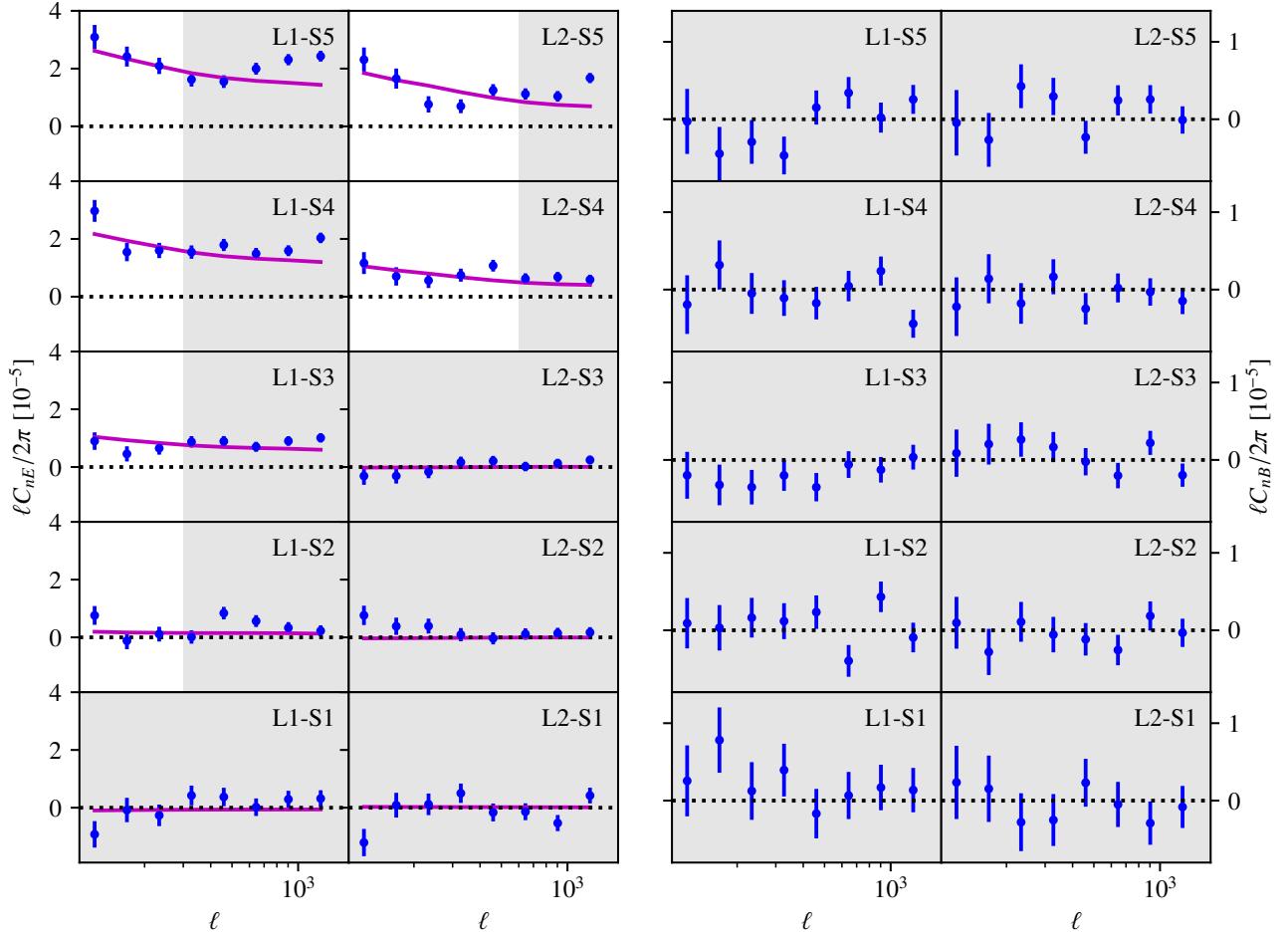


Fig. 3: KiDS-1000 galaxy-galaxy lensing power spectra: Tomographic band powers comparing the E-modes (left block) with the best-fit cosmological model from our combined multi-probe analysis. The tomographic bin combination of BOSS and 2dFLenS lenses (L) with KiDS-1000 sources (S), is indicated in the upper right corner of each sub-panel. Data within grey regions are not included in the cosmological analysis. The null-test B-modes (right block - note the reduced ordinate scale), are consistent with zero for both the full data vector, and each bin combination individually. The errors are estimated analytically (Joachimi et al. 2020).

Fig. 3 presents the KiDS-1000 galaxy-galaxy lensing power spectra, around lenses from the BOSS and 2dFLenS surveys (see Blake et al. 2020, for the real-space KiDS-1000 galaxy-galaxy lensing measurements for BOSS and 2dFLenS separately). Each panel presents the cross-correlation between each of the five different tomographic source bins, denoted ‘S’, with the two different lens bins, denoted ‘L’ (see Table A.1 for details). Here we have constructed both E-mode (left) and B-mode (right) band powers in order to isolate any non-lensing B-mode distortions. As expected from the analysis of Giblin et al. (2020), the measured B-modes are found to be consistent with zero. The measured E-modes can be compared to the theoretical expectation given the best-fit set of cosmological parameters from our joint multi-probe analysis in Sect. 3 in the non-shaded regions.

The shaded regions in Fig. 3 are excluded from our analysis for two reasons. For overlapping lens-source bins (L1 with S1 and L2 with S1 to S3), the intrinsic alignment terms $C_{\text{gl}}(\ell)$ are expected to become significant. This raises the question of the validity of the arguably rudimentary ‘NLA’ intrinsic alignment model when used in combination with a non-linear galaxy bias model (see Blazek et al. 2019, for a self-consistent perturbative approach to both intrinsic alignment and galaxy bias modelling). As these bin combinations carry little cosmological information, we exclude this data from our cosmological inference analysis, using it instead in a redshift-scaling null test of the catalogue in Giblin et al. (2020). For separated lens-source bins we introduce a maximum ℓ -scale beyond which the contributions from scales $k > 0.3h \text{ Mpc}^{-1}$ become significant (see figure 2 in Joachimi et al. 2020). In this regime uncertainties in the extrapolation of the Sánchez et al. (2017) non-linear galaxy bias model into the non-linear regime (Eq. 6) may well render the $C_{\text{ne}}(\ell)$ model invalid. The ℓ -limit depends on the redshift of the lens bin. Fig. 3 therefore serves as an important illustration of the necessity of improving non-linear galaxy bias and non-linear intrinsic alignment modelling for future studies, in order to fully exploit the cosmological signal contained within the galaxy-galaxy lensing observable.

2.5. Multi-probe covariance

Joachimi et al. (2020) present the multi-probe covariance matrix adopted in this study, verified through an analysis of over 20 000 fast full-sky mock galaxy catalogues derived from log-normal random fields. Given that only 4% of the BOSS footprint overlaps with KiDS-1000, in an initial step we validate the approximation that the BOSS anisotropic galaxy clustering observations are uncorrelated with the cosmic shear and galaxy-galaxy lensing observations. By imposing realistic overlapping BOSS and KiDS-1000 footprints in our mock catalogues, we find that cross-correlation, between the projected BOSS-like angular galaxy correlation function, and the KiDS-1000-like weak lensing signals, is less than $\sim 5\%$ of the auto-correlation terms along the diagonal of covariance matrix. With such a low cross-correlation, we can safely assume independence between the clustering and lensing observations, allowing us to adopt the Sánchez et al. (2017) covariance matrix⁸ for the anisotropic galaxy clustering observations, $\xi_{\text{gg}}(s, \mu, z)$, setting the clustering-lensing cross-correlation terms to zero.

The covariance of the two weak lensing observations is calculated analytically, combining terms that model pure Gaussian shape noise, survey sampling variance and the noise-mixing that

occurs between these two components, in addition to higher-order terms that account for mode-mixing between the in-survey modes and between the observed in-survey and the unobserved out-of-survey modes (known as super-sample covariance, Takada & Hu 2013). The covariance also includes a contribution to account for our uncertainty on the multiplicative shear calibration correction (Kannawadi et al. 2019). Joachimi et al. (2020) demonstrate that every term in the covariance is important, each dominating in different regions of the covariance with one exception: non-Gaussian variance between the in-survey modes is always sub-dominant. We therefore review the approximations made in these analytical calculations. Whilst the complex KiDS-1000 mask is fully accounted for in the shape-noise terms, and the super-sample terms, for all other terms it is assumed that the scales we measure are much smaller than any large-scale features in the survey footprint. Furthermore, the survey is assumed to be homogeneous in its depth, which is invalid for any ground-based survey where the survey depth becomes a sensitive function of the observing conditions (Heydenreich et al. 2020). With mock catalogues, we have the freedom to impose complex masks and variable depth to quantify the impact of these effects on the derived covariance, finding differences typically $\lesssim 10\%$, with a maximum difference of $\sim 20\%$. The majority of the differences were found to be driven by the mix-term between the Gaussian shape noise and the Gaussian sampling variance. Through a mock multi-probe data vector inference analysis, Joachimi et al. (2020) demonstrate that these differences in the covariance are not expected to lead to any systematic bias in the recovery of the KiDS-1000 cosmological parameters, nor to any significant differences in the confidence regions of the recovered parameters. We therefore adopt an analytical covariance in our analysis, shown in Fig. A.1, and refer the reader to Joachimi et al. (2020) for further details, where their section 4 presents the mocks, section 5 and appendix E presents the analytical covariance model, and appendix D presents detailed comparisons of the mock and analytical covariance.

2.6. Parameter inference methodology

We use the KiDS Cosmology Analysis Pipeline, KCAP⁹ built from the COSMOSIS analysis framework of Zuntz et al. (2015), adopting the nested sampling algorithm MULTINEST (Feroz & Hobson 2008; Feroz et al. 2009, 2019). The KCAP bespoke modules include: the BOSS wedges likelihood from Sánchez et al. (2017); the band power cosmic shear and galaxy-galaxy lensing likelihood based on Eq. (1) and Eq. (6); tools to permit correlated priors on nuisance parameters; and tools to sample over the clustering parameter $S_8 = \sigma_8 \sqrt{\Omega_m / 0.3}$, a parameter which is typically only derived. Scripts are also provided to derive the best-fit parameter values at the maximum multivariate posterior, denoted MAP (maximum a posteriori), and an associated credible region given by the projected joint highest posterior density region, which we denote PJ-HPD (Joachimi et al. 2020). This concise list of new modules reflect the primary updates in the KiDS-1000 parameter inference methodology compared to previous KiDS analyses, which we discuss in more detail below.

Our $3 \times 2\text{pt}$ model has 20 free parameters, with five to describe flat ΛCDM in addition to fifteen nuisance parameters. Eight of these nuisance parameters describe the galaxy bias

⁸ The BOSS ξ_{gg} covariance is derived from the MD-PATCHY BOSS mock catalogues of Kitaura et al. (2016).

⁹ KCAP will be made public on acceptance of the KiDS-1000 analysis papers. Interested groups can however request early access by e-mail to the lead authors of this paper.

model, with four in each lens redshift bin. The remaining seven allow us to marginalise over our uncertainty on the impact of baryon feedback (one parameter), intrinsic galaxy alignment (one parameter), and the mean of the source redshift distribution in each tomographic bin (five correlated parameters). The priors adopted for each parameter are listed in Appendix B.

Adopted priors are usually survey-specific, with the intention to be uninformative on the parameter that lensing studies are most sensitive to, S_8 . Different prior choices, particularly on the amplitude of the primordial power spectrum of scalar density fluctuations, A_s , have however been shown to lead to non-negligible changes in the derived S_8 parameter (Joudaki et al. 2017; Chang et al. 2019; Joudaki et al. 2020; Asgari et al. 2020c). Joachimi et al. (2020) show that even with wide priors on A_s , the sampling region in the σ_8 - Ω_m plane is significantly truncated at low values of σ_8 and Ω_m , with the potential to introduce a subtle bias towards low values of σ_8 . In this analysis, we address this important issue of implicit informative priors by sampling directly in S_8 . By adopting a very wide S_8 prior, our constraints on S_8 are therefore not impacted by our choice of prior. We note, however, that this approach is expected to lead to a more conservative constraint on S_8 , compared to an analysis that adopts a uniform prior¹⁰ on $\ln A_s$.

We account for the uncertainty in our source redshift distributions using nuisance parameters δ_z^i which modify the mean redshift of each tomographic bin i . By analysing mock KiDS catalogues, Wright et al. (2020a) determined the mean bias per redshift bin μ^i , and also the covariance between the different redshift bins, $C_{\delta z}$. This covariance arises from sampling variance in the spectroscopic training sample, which impacts, to some degree, the redshift calibration of all bins. We therefore adopt the multivariate Gaussian prior $N(\mu; C_{\delta z})$, for the vector δ_z (see section 3 of Hildebrandt et al. 2020b, for details).

Adopting the Bayesian paradigm for inference, we provide our constraints in the form of a series of samples that describe the full posterior distribution¹¹. In Sect. 3 we explore this multi-dimensional posterior in the traditional way, visualising the 2D and 1D marginal posterior distributions for a selection of parameters. In cosmological parameter inference it is standard to also report a point estimate of the one-dimensional marginal posterior distribution with an associated 68% credible interval. It is not always stated, however, how these point estimates and intervals are defined.

We provide two different point estimates for our cosmological parameters, with the first reporting the standard maximum of the marginal distribution, along with a credible interval that encompasses 68% of the marginal highest posterior density, which we denote by M-HPD. For the high-dimensional parameter space of a multi-probe weak lensing analysis, we find that this standard marginalised point estimate leads to a value for S_8 that is lower than the maximum of the multivariate joint posterior, with an offset of up to $\sim 1\sigma$, dependent on which probes are combined (see section 7 of Joachimi et al. 2020). This is not a result of an error in the KCAP inference pipeline. Rather, it is a generic feature of projecting high-dimensional asymmetric distributions into one dimension, prompting the development of an alternative approach to reporting point estimates for cosmological parameters.

¹⁰ For quantitative information about the impact of implicit A_s priors, see table 5 and figure 22 of Joachimi et al. (2020).

¹¹ Our MULTINEST full posterior samples will be publicly released on the acceptance of this paper.

Our fiducial S_8 constraints follow this alternative, reporting the parameter value at the maximum of the joint posterior (MAP), along with a 68% credible interval based on the joint, multi-dimensional highest posterior density region, projected onto the marginal posterior of the S_8 parameter (PJ-HPD). Here we step through the posterior MULTINEST samples, ordered by their decreasing posterior density. For each model parameter we determine the extrema within the n highest posterior samples, and the posterior mass contained within the marginal distribution of each parameter, limited by the extrema values. We iterate, increasing the number of samples analysed, n , until the posterior mass reaches the desired 68% level. The PJ-HPD credible interval is then reported as the parameter extrema at this point n in the sample list, and we repeat the process for each model parameter of interest (see section 6.4 of Joachimi et al. 2020, for further details).

We note that the MAP reported by MULTINEST provides a noisy estimate of the true MAP due to the finite number of samples, and we therefore conduct an optimisation step using the 36 samples with the highest posterior values as starting points. We use both the Nelder & Mead (1965) and Powell (1964) optimisation algorithms, as well as two-step optimisation using both algorithms. While the MAP estimates found in this optimisation step increase the posterior probability by a factor of 2 to 4 compared to the MAP estimated from the MULTINEST samples, they exhibit scatter in parameters constrained by the galaxy clustering likelihood. We suspect this is due to numerical noise in the galaxy clustering likelihood which results in many local minima that inhibit the convergence of the optimisation step. This suspicion is strengthened by the fact that the MAP estimates do not exhibit this scatter for probe combinations that exclude the clustering observable. For this reason, we report the median of the MAP estimates, weighted by their posterior probability, for probes that include the galaxy clustering likelihood, since a global optimisation would be computationally prohibitively expensive. For the other probes the reported MAP is given by the parameter set at the maximum posterior found amongst all estimates (see also Muir et al. 2020, who adopt a similar approach).

We note that the presence of offsets between marginal S_8 constraints and those derived from the full multivariate joint posterior highlights how efforts to accurately quantify tension based solely on one-point estimates should be undertaken with some level of caution. Tension can also be assessed in terms of the overlap between the full posterior distributions (see for example Handley & Lemos 2019; Lemos et al. 2019; Raveri & Hu 2019), which we discuss further in Sect. 3.2 and Appendix F.

3. Results

We present our multi-probe constraints on the cosmological parameters of the flat Λ CDM model in Fig. 4, showing the marginalised posterior distributions for matter fluctuation amplitude parameter, σ_8 , the matter density parameter, Ω_m , and the dimensionless Hubble parameter, h , where the BOSS galaxy clustering constraints (GC: shown blue), break the σ_8 - Ω_m degeneracy in the KiDS-1000 cosmic shear constraints (CS: shown pink), resulting in tight constraints on σ_8 in the combined 3×2 pt analysis (shown red). Reporting the MAP values with PJ-HPD credible intervals for the parameters that we are most sensitive

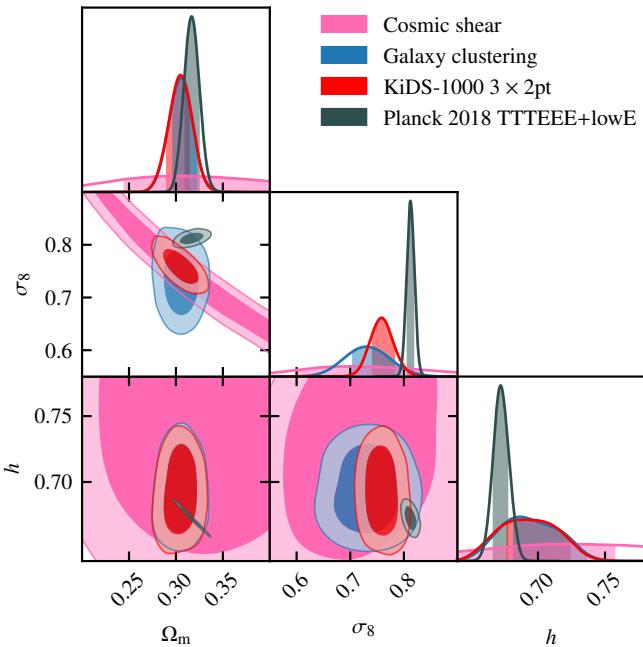


Fig. 4: Marginal multi-probe constraints on the flat Λ CDM cosmological model, for the matter fluctuation amplitude parameter, σ_8 , the matter density parameter, Ω_m , and the dimensionless Hubble parameter, h . The BOSS galaxy clustering constraints (blue), can be compared to the KiDS-1000 cosmic shear constraints (pink), the combined 3×2 pt analysis (red), and CMB constraints from *Planck* Collaboration et al. (2018, grey).

to, we find

$$\begin{aligned} \sigma_8 &= 0.76^{+0.025}_{-0.020} \\ \Omega_m &= 0.305^{+0.010}_{-0.015} \\ S_8 &= 0.766^{+0.020}_{-0.014}. \end{aligned} \quad (7)$$

Our constraints can be compared to the marginalised posterior distributions from *Planck* (shown grey in Fig. 4), finding consistency between the marginalised constraints on Ω_m and h , but an offset in σ_8 , which we discuss in detail in Sect. 3.2.

Tabulated constraints for the full set of cosmological parameters are presented in Appendix C, quoting our fiducial MAP with PJ-HPD credible intervals, along with the marginal posterior mode with M-HPD credible intervals. As discussed in Joachimi et al. (2020), the marginal mode estimate is known to yield systematically low values of S_8 in mock data analyses, as can be seen in Fig. 5, which compares the joint posterior constraints (solid) with the marginal posterior constraints (dashed).

We find good agreement between the different probe combinations and single-probe S_8 constraints, demonstrating internal consistency between the different cosmological probes, in Fig. 5. As forecast by Joachimi et al. (2020), the addition of the galaxy-galaxy lensing observable adds very little constraining power, with similar results found for the full 3×2 pt analysis and the combined cosmic shear and clustering analysis. This is a result of the significant full area of BOSS in comparison to the size of the BOSS-KiDS overlap region, and the fact that our lack of an accurate galaxy bias model on the deeply non-linear scales that weak lensing probes, prohibits the inclusion of large sections of our galaxy-galaxy lensing data vector, shown in Fig. 3.

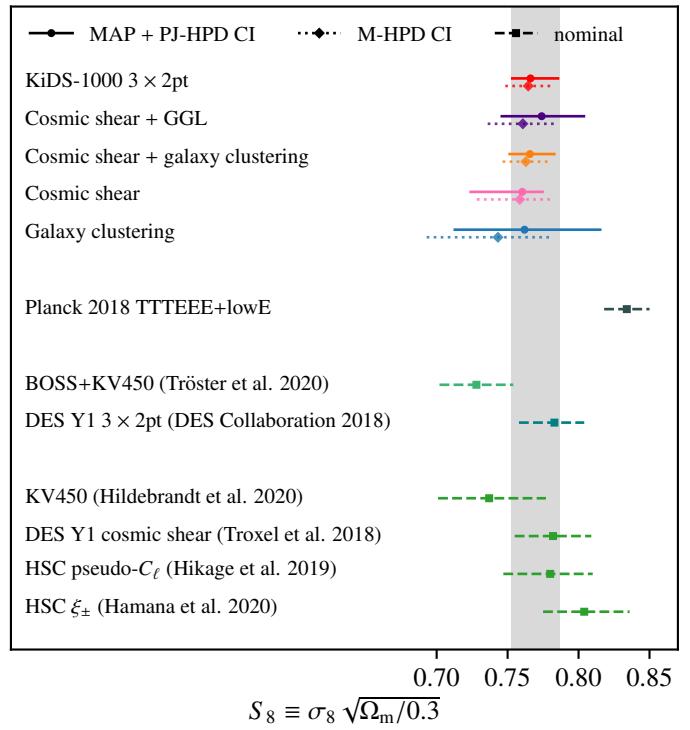


Fig. 5: Constraints on the structure growth parameter $S_8 = \sigma_8 \sqrt{\Omega_m/0.3}$ for different probe combinations: 3×2 pt, cosmic shear with galaxy-galaxy lensing (GGL), and cosmic shear with galaxy clustering, along with the single probe analyses. Our fiducial and preferred MAP with PJ-HPD credible interval (solid) can be compared to the standard, but shifted, marginal posterior mode with M-HPD credible intervals (dotted). Our results can also be compared to weak lensing measurements from the literature, which typically quote the mean of the marginal posterior mode with tail credible intervals (dashed).

The addition of the galaxy-galaxy lensing does, however, serve to moderately tighten constraints on the amplitude of the intrinsic alignment model A_{IA} , as seen in Fig. 6.

Fig. 5 also demonstrates the good agreement between our constraints and weak lensing results from the literature, comparing to cosmic shear-only results from the Hyper Suprime-Cam Strategic Program (HSC, Hikage et al. 2019; Hamana et al. 2020), DES Y1 (Troxel et al. 2018) and an earlier KiDS analysis (KV450 Hildebrandt et al. 2020a), in addition to the previous KV450-BOSS ‘ 2×2 pt’ analysis of Tröster et al. (2020) and the DES Y1 3×2 pt analysis from Abbott et al. (2018). We refer the reader to Asgari et al. (2020b) for a discussion and comparison of different cosmic shear results. In Sect. 3.1 we present a more detailed comparison of our results with 3×2 pt results in the literature.

Fig. 6 displays the marginal posterior distributions for an extended set of cosmological parameters. We find that the allowed range for the linear galaxy bias, b_1 , in each redshift bin (lower two rows), is almost halved with the addition of the weak lensing data. This constraint does not arise, however, from the sensitivity of the galaxy-galaxy lensing observable to galaxy bias (shown to be relatively weak in the CS+GGL contours). Instead, in this analysis, it is a result of the degeneracy breaking in the σ_8 - Ω_m plane, tightening constraints on σ_8 which, for galaxy clus-

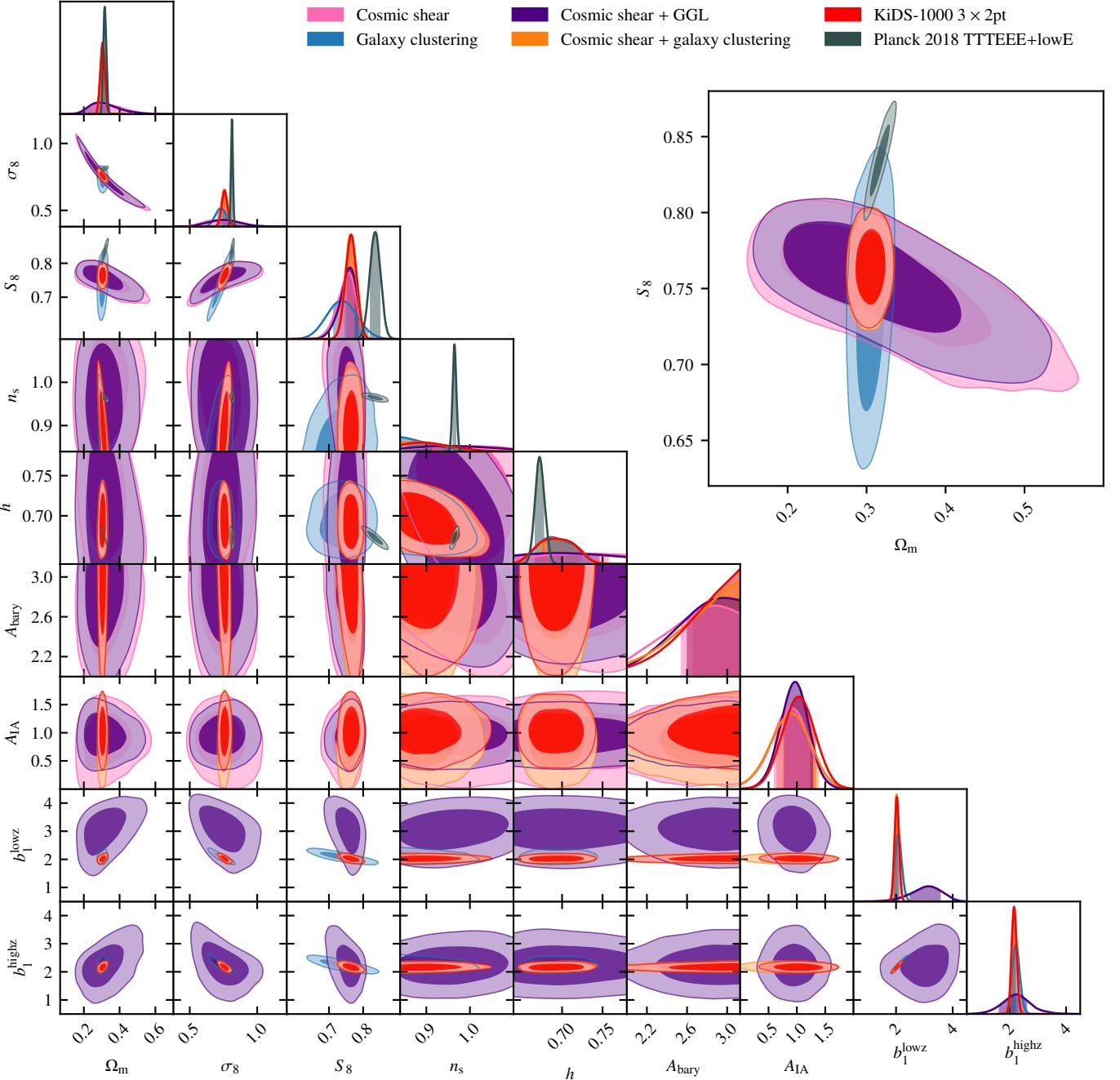


Fig. 6: Marginalised posterior distributions for an extended set of cosmological parameters covering the matter density parameter Ω_m , the matter fluctuation amplitude parameter, σ_8 , the structure growth parameter S_8 , the spectral index n_s , the dimensionless Hubble parameter, h , the baryon feedback amplitude parameter, A_{bary} , the intrinsic alignment amplitude, A_{IA} , and the linear bias parameters for the low and high BOSS redshift bins, b_1 . The KiDS-1000 cosmic shear results (pink), can be compared to the BOSS galaxy clustering results (blue), the combination of cosmic shear with BOSS and 2dFLenS galaxy-galaxy lensing (GGL, purple), and the full $3 \times 2\text{pt}$ analysis (red). The combination of cosmic shear with galaxy clustering (orange) is only distinguishable from the $3 \times 2\text{pt}$ result in the A_{bary} and A_{IA} panels. For parameters constrained by the CMB, we also include constraints from *Planck* Collaboration et al. (2018, grey).

tering, is degenerate with galaxy bias. The improved constraints on galaxy bias do not, however, fold through to improved constraints on h , which the weak lensing data adds very little information to.

For our primary cosmological parameter, S_8 , our constraints are uninformed by our choice of priors. This statement cannot

be made for the other ΛCDM parameters, however, as shown in Fig. 6. The most informative prior that we have introduced to our $3 \times 2\text{pt}$ analysis is on the spectral index, n_s . As noted by Tröster et al. (2020), the BOSS galaxy clustering constraints favour a low value for n_s , where they find $n_s = 0.815 \pm 0.085$. From the Tröster et al. (2020) sensitivity analysis to the adopted

Table 1: The goodness-of-fit of the flat Λ CDM cosmological model to each of the single and joint probe combinations with cosmic shear, galaxy clustering and galaxy-galaxy lensing (GGL).

Probe	χ^2_{MAP}	Data DoF	Model DoF (Joachimi et al. 2020)	PTE	Model DoF (Raveri & Hu 2019)	PTE
Cosmic shear	152.1	120	4.5	0.013	3.0	0.016
Galaxy clustering	167.7	168	—	—	10.6	0.272
Cosmic shear + GGL	178.7	142	8.7	0.005	7.3	0.007
Cosmic shear + galaxy clustering	319.9	288	—	—	11.9	0.036
KiDS-1000 3 \times 2pt	356.2	310	—	—	12.5	0.011

Notes. We list the χ^2 value at the maximum of the posterior, the number of degrees of freedom (DoF) of the data, the effective DoF of the model, and the probability to exceed (PTE) the measured χ^2 value, assuming the total DoF are given by data DoF – model DoF. The effective DoF of the model are estimated following Joachimi et al. (2020) and Raveri & Hu (2019), accounting for the impact of priors and non-linear dependencies between the parameters.

maximum clustering scale, we observe that this preference appears to be driven by the amplitude of the large scale clustering signal with $s > 100 h^{-1}$ Mpc. We note that spurious excess power in this regime could plausibly arise from variations in the stellar density impacting the BOSS galaxy selection function (Ross et al. 2017). Our choice to impose a theoretically motivated informative prior for n_s , as listed in Table B.1, helps to negate this potential systematic effect without degrading the overall goodness-of-fit to the galaxy clustering measurements. Our prior choice is certainly no more informative than the n_s priors that are typically used in weak lensing and clustering analyses (see for example Abbott et al. 2018; eBOSS Collaboration et al. 2020). We recognise, however, that this well-motivated prior choice acts to improve the BOSS-only error on Ω_m by roughly a third, and decrease the BOSS-only best-fitting value for Ω_m and h by $\sim 0.5 \sigma$ (see Fig. B.1). With $< 10\%$ differences on the constraints on S_8 and h , however, and only a $\sim 0.1 \sigma$ difference in the BOSS-only best-fitting value for S_8 , which is consistent with the typical variation between different MULTINEST analyses, we conclude that our prior choice does not impact on our primary S_8 constraints. With the informative or uninformative n_s prior, our constraints on h remain consistent with the Hubble parameter constraints from both Planck Collaboration et al. (2018) and Riess et al. (2019).

Fig. 7 illustrates the results of a series of sensitivity tests, where we explore how our 3×2 pt constraints on S_8 change when: we ignore the impact of baryon feedback (the ‘No baryon’ case), fixing $A_{\text{bary}} = 3.13$, corresponding to the non-linear matter power spectrum for a dark-matter only cosmology; we limit the analysis to a linear galaxy bias model, setting all higher-order bias terms in Eq. (3) to zero, as well as restricting the redshift-space distortion model to a Gaussian velocity distribution; and we remove individual tomographic bins from our weak lensing observables. The systematic offset that arises from the use of a linear-bias model highlights the importance of accurate non-linear galaxy bias modelling in 3×2 pt analyses. This series of tests is dissected further in Appendix D, and complements the detailed KiDS-1000 internal consistency analysis of Asgari et al. (2020b, appendix B), which demonstrates that the change seen with the removal of tomographic bin 4 is fully consistent with expected statistical fluctuations.

Table 1 records the goodness-of-fit for each component in our 3×2 pt analysis, where we report the χ^2 value at the maximum posterior, χ^2_{MAP} (see Sect. 2.6 for a discussion of our optimised MAP-finder). The effective number of degrees of freedom (DoF) does not equate to the standard difference between

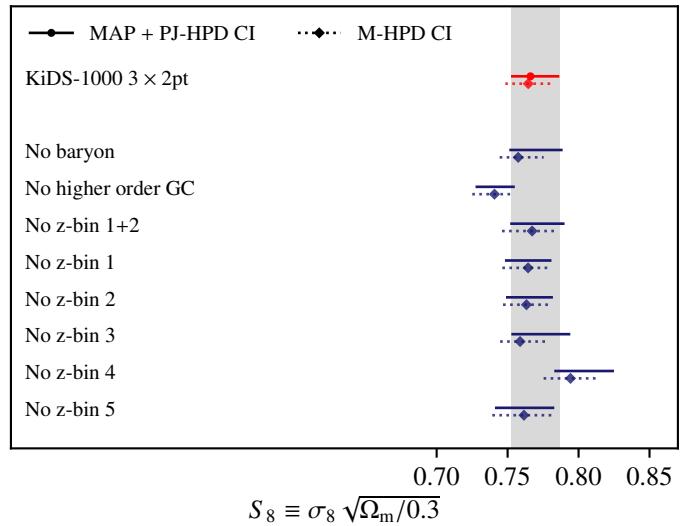


Fig. 7: 3×2 pt constraints on S_8 for a series of sensitivity tests; when we ignore the impact of baryon feedback (the ‘No baryon’ case), limit the analysis to a linear galaxy bias model (the ‘No higher order GC’ case), and remove individual tomographic bins from our weak lensing observables.

the total number of data points (Data DoF) and the total number of model parameters (20 in the case of our 3×2 pt analysis), as a result of the adopted priors and the non-linear dependencies that exist between the model parameters. For some probe combinations we calculate the effective number of degrees of freedom in the model (Model DoF), using the estimator described in section 6.3 of Joachimi et al. (2020). As this approach is computationally expensive, however, we also estimate the Model DoF following Raveri & Hu (2019), recognising that, for the cases explored in Joachimi et al. (2020), this approach results in a slightly lower model DoF.

We find that the goodness-of-fit is excellent for the BOSS galaxy clustering. For all other cases, the goodness-of-fit is certainly acceptable¹², with the probability to exceed the measured χ^2 given by $p \gtrsim 0.01$. We note that the cosmic shear analysis of Asgari et al. (2020b) shows no significant changes in the in-

¹² We define acceptable as the PTE $p \geq 0.001$, which corresponds to less than a $\sim 3\sigma$ event. Abbott et al. (2018) define acceptable as $\chi^2/\text{DoF} < 1.4$. We meet both these requirements.

ferred cosmological parameters when using different two-point statistics which exhibit an excellent goodness-of-fit. As such, we could be subject to an unlucky noise fluctuation that particularly impacts the band power estimator in Eq. (1). Cautiously inspecting Fig. 1, as ‘ χ -by-eye’ is particularly dangerous with correlated data points, we nevertheless note a handful of outlying points, for example the low ℓ -scales in the fifth tomographic bin. We also note that Giblin et al. (2020) document a significant but low-level PSF residual systematic in the KiDS-1000 fourth and fifth tomographic bins that was shown to reduce the overall goodness-of-fit in a cosmic shear analysis, but not bias the recovered cosmological parameters (see the discussion in Amara & Réfrégier 2008). Future work to remove these low-level residual distortions is therefore expected to further improve the goodness-of-fit.

3.1. Comparison with weak lensing surveys

Our results are consistent with weak lensing constraints in the literature. We limit our discussion in this section to published $3 \times 2\text{pt}$ analyses, referring the reader to Asgari et al. (2020b) who discuss how the KiDS-1000 cosmic shear results compare with other weak lensing surveys. We note that direct comparisons of cosmological parameters should be approached with some caution, as the priors adopted by different surveys and analyses are often informative (see section 6.1 in Joachimi et al. 2020). Homogenising priors for cosmic shear analyses, for example, has been shown to lead to different conclusions when assessing inter-survey consistency (Chang et al. 2019; Joudaki et al. 2020; Asgari et al. 2020c).

Abbott et al. (2018) present the first year $3 \times 2\text{pt}$ DES analysis (DES Y1), finding $S_8 = 0.773^{+0.026}_{-0.020}$, where they report the marginal posterior maximum and the tail credible intervals. This is in excellent agreement with our equivalent result, differing by 0.3σ , with the DES-Y1 error being 40% larger than the KiDS-1000-BOSS $3 \times 2\text{pt}$ results. The inclusion of BOSS to our $3 \times 2\text{pt}$ analysis results in tight constraints on Ω_m . This leads to joint KiDS-1000-BOSS constraints on $\sigma_8 = 0.760^{+0.021}_{-0.023}$ that are more than twice as constraining compared to the DES Y1-alone $3 \times 2\text{pt}$ analysis, which found $\sigma_8 = 0.817^{+0.045}_{-0.056}$, as shown in Fig. 8. This comparison serves to highlight the additional power that can be extracted through the combination of spectroscopic and photometric surveys, and the promising future for the planned overlap between the Dark Energy Spectroscopic Instrument survey (DESI Collaboration et al. 2016) and the 4-metre Multi-Object Spectroscopic Telescope (4MOST, Richard et al. 2019), with *Euclid* and the Vera C. Rubin Observatory Legacy Survey of Space and Time (Laureijs et al. 2011; LSST Science Collaboration et al. 2009), in addition to the nearer-term $\sim 1400 \text{ deg}^2$ of overlap between BOSS and the Hyper Suprime-Cam Strategic Program (HSC, Aihara et al. 2019).

van Uitert et al. (2018) and Joudaki et al. (2018) present $3 \times 2\text{pt}$ analyses for the second KiDS weak lensing release (KiDS-450), finding, respectively, $S_8 = 0.800^{+0.029}_{-0.027}$ (KiDS with GAMA) and $S_8 = 0.742 \pm 0.035$ (KiDS with BOSS and 2dFLenS limited to the overlap region). Both results are consistent with our KiDS-1000 results, noting that the increase in our S_8 constraining power, by a factor of ~ 2 in this analysis, is driven by increases in both the KiDS survey area, and the analysed BOSS survey area.

The impact of doubling the KiDS area can be seen by comparing to Tröster et al. (2020), in Fig. 8, who present a joint cosmic shear and galaxy clustering analysis of the KV450 KiDS

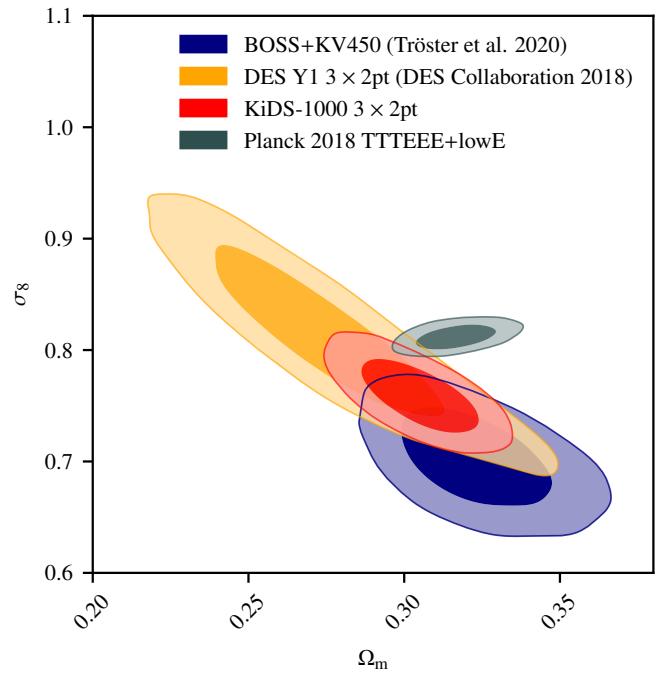


Fig. 8: Marginalised posterior distribution in the σ_8 - Ω_m plane, comparing the $3 \times 2\text{pt}$ analyses from KiDS-1000 and DES Y1 (Abbott et al. 2018) with the CMB constraints from *Planck* Collaboration et al. (2018). The KiDS-1000 $3 \times 2\text{pt}$ result can also be compared to our previous KV450-BOSS analysis from Tröster et al. (2020).

release with the full BOSS area, finding $S_8 = 0.728 \pm 0.026$. The $\sim 40\%$ improvement in constraining power is consistent with expectations from the increased survey area, but a straightforward area-scaling comparison is inappropriate given that KiDS-1000 features improvements in the accuracy of the shear and photometric redshift calibrations, albeit at the expense of a decrease in the effective number density (see Giblin et al. 2020; Hildebrandt et al. 2020b, for details).

The offset in S_8 between the KiDS-1000-BOSS and KV450-BOSS S_8 constraints reflects a number of differences between the two analyses. First, as the S_8 constraints from the $3 \times 2\text{pt}$ analysis are primarily driven by KiDS (see Fig. 6), we expect a reasonable statistical fluctuation in this parameter given the sampling variance arising from the significant increase in the KiDS survey area. Using a simple model analysis in Appendix E, we conclude that we should expect differences, on average, of $|\Delta S_8| = 0.016$, and as such the increase that we find in S_8 between KV450 and KiDS-1000 is consistent with the expectation from simple statistical fluctuations. BOSS primarily constrains Ω_m which is impacted by the choice of prior on n_s . The wider n_s prior adopted in Tröster et al. (2020), favours a slightly higher but less well-constrained value for Ω_m , leading to a slightly lower but less well-constrained value for σ_8 , when combined with cosmic shear (see Appendix B). If we had also chosen an uninformative prior on n_s for our KiDS-1000-BOSS analysis, a decision that we cannot revise post unblinding, this would have likely served to exacerbate any tension with the *Planck* CMB constraints.

3.2. Comparison with *Planck*

In our KiDS-1000-BOSS $3 \times 2\text{pt}$ analysis, we find good agreement with *Planck* for the matter density parameter, Ω_m , and the Hubble parameter, h , (see Fig. 4). The amplitude of matter fluctuations, σ_8 , that we infer from the clustering of galaxies within, and lensing by, the large-scale structure of the low-redshift Universe is lower, however, than that inferred by *Planck*¹³ from the CMB.

To quantify the level of discrepancy in the amplitude of matter fluctuations, we first concentrate on the parameter $S_8 = \sigma_8 \sqrt{\Omega_m / 0.3}$ as it is tightly constrained and only exhibits negligible degeneracies, if at all, with the other cosmological parameters, Ω_m , h , and n_s , as illustrated in Fig. 6. Comparing the reported marginal S_8 constraints, we find S_8 to be $8.3 \pm 2.6\%$ lower than the CMB constraint from *Planck* Collaboration et al. (2018).

We define the widely used S_8 -difference measure

$$\tau = \frac{|\bar{S}_8^{3 \times 2\text{pt}} - \bar{S}_8^{\text{Planck}}|}{\sqrt{\text{Var}[S_8^{3 \times 2\text{pt}}] + \text{Var}[S_8^{\text{Planck}}]}}, \quad (8)$$

where \bar{S}_8 and $\text{Var}[S_8]$ denote the mean and standard deviations of the *Planck* and $3 \times 2\text{pt}$ S_8 posterior distributions. If both distributions are Gaussian, τ can be used to measure how likely it is that the mean of the difference between the distributions is consistent with zero.

Comparing the S_8 posterior distributions between our $3 \times 2\text{pt}$ analysis and the *Planck* `plik_lite_TTTEEE+lowl+lowE` likelihood, we find $\tau = 3.1$, i.e. that there is a 3.1σ difference between the KiDS-1000 and *Planck* constraints. Adopting two tension measures that do not assume Gaussianity of the marginal posterior distributions, both, the ‘Hellinger’ distance and the distribution of the S_8 parameter shifts, indicate 3.1σ difference between our $3 \times 2\text{pt}$ analysis and *Planck* (see Appendices F.1 and F.2 for details). Our result thus continues the general trend of low-redshift probes preferring low amplitudes of matter fluctuations¹⁴ (Heymans et al. 2013; Alam et al. 2017; Abbott et al. 2018; Hikage et al. 2019; Bocquet et al. 2019; Palanque-Delabrouille et al. 2020; Wright et al. 2020b; DES Collaboration et al. 2020). In these cases the reported low S_8 , or σ_8 , constraints are formally statistically consistent with *Planck*, and well below the detection of any anomalies at the 5σ -level. Considering, however, the $\sim 3\sigma$ difference that we have reported, and the overall trend in the literature, we would argue that we are reaching an uncomfortable point when it comes to regarding the S_8 offset as a simple statistical fluke.

Sanchez (2020) pointed out that comparing S_8 between different experiments can be misleading due to the implicit dependence of σ_8 on h . Besides the intrinsic dependence of the amplitude of the matter power spectrum on h , measurements of σ_8 also depend on h through the value of $8 h^{-1}\text{Mpc}$, the radius of the sphere within which the matter fluctuations are measured. In this way, constraints on σ_8 derived from data sets with different posterior distributions on h represent the average of

$\sigma(R)$ over different ranges of scales. We therefore also consider $S_{12} = \sigma_{12} (\Omega_m h^2 / 0.14)^{0.4}$ (Sanchez 2020), where σ_{12}^2 is the variance of the linear matter field at redshift zero in spheres of radius 12 Mpc. We find $S_{12} = 0.754^{+0.015}_{-0.018}$, with the value inferred by *Planck* being $S_{12} = 0.817^{+0.011}_{-0.015}$, a difference of $\tau = 3.0$, in agreement with the S_8 results.

In light of the large parameter spaces that are being considered, we recognise that focussing on a single parameter can paint a simplistic picture of the agreement, or disagreement, between probes. On a fundamental level, the question we wish to answer is whether a single model of the Universe can describe both the CMB as well as the low-redshift large-scale structure of the Universe. Within our Bayesian inference framework, the Bayes factor provides a natural approach to model selection. The two models under consideration are

M_1 : Both our $3 \times 2\text{pt}$ data and *Planck*’s measurements of the CMB are described by a single flat ΛCDM cosmology.

M_2 : The two data sets are described by different cosmologies for the low- and high- redshift Universe, respectively.

The Bayes factor is then

$$R = \frac{P(\mathbf{d}|M_1)P(M_1)}{P(\mathbf{d}|M_2)P(M_2)}, \quad (9)$$

where $P(\mathbf{d}|M_i)$ is the probability of the data \mathbf{d} under model M_i – the Bayesian evidence.

We assume the model priors $P(M_1)$ and $P(M_2)$ to be equal, that is, we make no a-priori assumption on the likelihood of M_1 or M_2 . We use *anesthetic*¹⁵ (Handley 2019) to compute R and find $\ln R = 3.1 \pm 0.3$, which can be interpreted as odds of 23 ± 6 in favour of model M_1 , and consistency between our $3 \times 2\text{pt}$ measurement and *Planck*. Given the dependence of R on the parameter priors (Handley & Lemos 2019), we consider this result with some caution, and review a series of alternative metrics that seek to quantify the tension between the full KiDS-1000 and *Planck* multi-dimensional cosmological parameter constraints. These metrics are summarised in Table 2, and below, with further details provided in Appendix F.

Handley & Lemos (2019) propose the ‘suspiciousness’ statistic S that is based on the Bayes factor, R , but hardened against prior dependences. We find that the probability of observing our measured suspiciousness statistic is 0.08 ± 0.02 , which corresponds to a KiDS-*Planck* tension at the level of $1.8 \pm 0.1\sigma$ (see Appendix F.3 for details).

Raveri & Hu (2019) introduce a number of metrics to quantify the consistency of data sets. We consider their Q_{DMAP} metric, which explores the change in the goodness-of-fit when two data sets are combined. A reduction in the goodness-of-fit can then be translated into a tension metric, finding a KiDS-*Planck* tension at the level of 2.1σ (see Appendix F.4 for details). Their Q_{UDM} metric generalises the notion of parameter differences between posteriors to multiple dimensions, finding KiDS-*Planck* tension at the level of 1.9σ (see Appendix F.5 for details).

Comparing tension metrics, we find a fairly consistent picture. Reviewing tension in terms of the single parameter that our $3 \times 2\text{pt}$ analysis is most sensitive to, S_8 , leads to a $\sim 3\sigma$ tension with *Planck*. Including additional parameters into the tension analysis, parameters which KiDS is mainly insensitive to, serves to effectively dilute the tension, reducing the measure to the $\sim 2\sigma$ level.

¹³ A recent independent Atacama Cosmology Telescope CMB analysis reports $S_8 = 0.830 \pm 0.043$, in agreement with the *Planck* constraint of $S_8 = 0.834 \pm 0.016$ (ACT, Aiola et al. 2020). Our results are fully consistent with the ACT CMB analysis, reflecting the larger uncertainty in the ACT constraints.

¹⁴ Although we note the very recent clustering analysis released from e-BOSS which is fully consistent with Planck (eBOSS Collaboration et al. 2020).

¹⁵ <https://github.com/williamjameshandley/anesthetic>

Table 2: Estimators of the consistency between our fiducial $3 \times 2\text{pt}$ analysis and the *Planck* Collaboration et al. (2018) TT-TEEE+lowE results.

Metric	Reference	Value	PTE	PTE [σ]
τ	Eq. (8)	3.1	—	3.1σ
d_H	Eq. (F.1)	0.95	—	3.1σ
p_S	Eq. (F.3)	0.9981	0.0019	3.1σ
R	Eq. (9)	23 ± 6	—	—
$\ln R$	Eq. (F.5)	3.1 ± 0.3	—	—
$\ln S$	Eq. (F.4)	-2.0 ± 0.1	0.08 ± 0.02	$1.8 \pm 0.1\sigma$
Q_{DMAP}	Eq. (F.10)	9.5	0.037	2.1σ
Q_{UDM}	Eq. (F.11)	7.7	0.054	1.9σ

Notes. The considered tension metrics are: 1) the differences between the means in S_8 , divided by the standard deviations added in quadrature, τ ; 2) the Hellinger distance, d_H , between S_8 posterior distributions (Appendix F.1); 3) the fraction of the S_8 parameter shift distribution that has a higher probability than no shifts (Appendix F.2); 4) the Bayes ratio R between a model that assumes a single cosmology for both our $3 \times 2\text{pt}$ and the *Planck* data, and a model that uses separate cosmologies for the two data sets; 5) the logarithm $\ln R$ of the Bayes ratio; 6) the suspiciousness $\ln S$ (Handley & Lemos 2019); 7 & 8) the concordance/discordance estimators Q_{DMAP} and Q_{UDM} (Raveri & Hu 2019). The columns list the estimator metric considered, the relevant equation reference, the measured value, and the ‘probability-to-exceed’ the measured value (PTE), under the assumption that the two data sets are consistent.

4. Conclusions

In this analysis we have presented constraints on the flat ΛCDM cosmological model by combining observations of gravitational lensing and galaxy clustering to directly probe the evolution and distribution of the large-scale structures in the Universe. Our survey of the $z \lesssim 1$ low-redshift Universe finds a matter distribution that is less clustered, compared to predictions from the best-fitting ΛCDM model to early-Universe CMB observations (*Planck* Collaboration et al. 2018). This tendency for low-redshift probes to favour a smoother matter distribution compared to the CMB expectation has persisted since the first large-scale weak lensing survey (CFHTLenS, Heymans et al. 2013), but the significance of this effect has always been tantalisingly around, or below, the $\sim 3\sigma$ level. It is therefore unclear if these differences are merely a statistical fluctuation, unaccounted for systematic errors, or a sign of interesting new physics.

Our new result does not lead to a resolution in the matter of statistical fluctuations, finding a 3.1σ offset in the structure growth parameter $S_8 = \sigma_8 \sqrt{\Omega_m / 0.3}$ with $S_8 = 0.766^{+0.020}_{-0.014}$. Comparing the marginal S_8 constraints, we find S_8 to be $8.3 \pm 2.6\%$ lower than the CMB constraint from *Planck* Collaboration et al. (2018). For a series of ‘tension’ metrics that quantify differences in terms of the full posterior distributions, we find that the KiDS-1000 and *Planck* results agree at the $\sim 2\sigma$ level. Through our series of image simulation analyses (Kannawadi et al. 2019), catalogue null-tests (Giblin et al. 2020), variable depth mock galaxy survey analyses (Joachimi et al. 2020), optical-to-near-infrared photometric-spectroscopic redshift calibration, validated with mocks (Wright et al. 2020a; van den Busch et al. 2020; Hildebrandt et al. 2020b), internal consistency tests (Asgari et al. 2020b, Fig. 6 and Appendix D), and marginalisation over a series of nuisance parameters that encompass our theoretical and calibration uncertainties (Appendix B), we argue

that we have, however, addressed the question of $3 \times 2\text{pt}$ systematic errors, robustly assessing and accounting for all sources of systematics that are known about in the literature.

The KiDS-1000 cosmic shear constraints are highly complementary to the BOSS galaxy clustering constraints, leading to tight constraints in our joint $3 \times 2\text{pt}$ analysis that are more than twice as constraining for the matter fluctuation amplitude parameter, $\sigma_8 = 0.760^{+0.021}_{-0.023}$, compared to previous $3 \times 2\text{pt}$ analyses. In the future, analysis of the clustering and galaxy-galaxy lensing of photometric samples with very accurate photometric redshifts (see for example Vakili et al. 2019), presents an opportunity for a future alternative KiDS-only $3 \times 2\text{pt}$ photometric analysis, similar to the approach taken in Abbott et al. (2018).

In the next few years, two weak lensing surveys will see first light, with the launch of the *Euclid* satellite and the opening of the Vera C. Rubin Observatory. These observatories will build the first two ‘full-sky’ weak lensing surveys, which are highly complementary in terms of their differing strengths in depth and spatial resolution. Combined with complementary overlapping redshift spectroscopy from DESI, 4MOST and *Euclid*, the multi-probe weak lensing and spectroscopic galaxy clustering methodology, which we have implemented in this analysis, provides a promising route forward for these next generation surveys. We view this $3 \times 2\text{pt}$ approach as just the start of the story, however, looking forward to a future combined analysis of weak lensing and galaxy clustering with both photometric and spectroscopic lenses, a combination which we call a ‘ $6 \times 2\text{pt}$ ’ approach (Bernstein 2009). This would allow for the optimal combination of information from the clustering cross-correlation of spectroscopic and photometric galaxies (Newman 2008), an observable that we currently only use as an independent tool to validate our photometric redshift calibration (Hildebrandt et al. 2020b). Developments in the area of highly non-linear galaxy bias, baryon feedback and intrinsic alignment modelling, along with a sufficiently flexible but tractable redshift distribution model and an accurate ‘ $6 \times 2\text{pt}$ ’ covariance estimate, will all be required in order to realise this long-term goal. The effort will, however, be worthwhile allowing for the implementation of arguably the most robust methodology available to mitigate systematic errors, whilst simultaneously enhancing cosmological parameter constraints.

The ESO-KiDS public survey completed observations in July 2019, spanning 1350 deg^2 . We therefore look forward to the fifth and final KiDS data release, ‘KiDS-Legacy’, along with new results from the concurrent ‘Stage-III’ surveys, DES and HSC, whilst the community prepares for the next exciting chapter of ‘full-sky’ weak lensing surveys.

Acknowledgements. We are completely indebted to Eric Tittley at the IfA for going well beyond the call of duty to save the KiDS-1000 data products after an explosion in our server room destroyed the RAID. We also thank our external blinder Matthias Bartelmann who revealed the key for which of the three catalogues analysed was the true unblinded catalogue on the 9th July 2020, right at the end of the KiDS-1000 study which was submitted to A&A on the 30th July 2020. We also wish to thank the Vera C. Rubin Observatory LSST-DESC Software Review Policy Committee (Camille Avestruz, Matt Becker, Celine Combet, Mike Jarvis, David Kirkby, Joe Zuntz with CH) for their Software Policy document which we followed, to the best of our abilities, during the KiDS-1000 project. Following this policy the software used to carry out the various analyses presented in this paper will be made public on publication of this paper. The figures in this work were created with `matplotlib` (Hunter 2007) and `getdist` (Lewis 2019), making use of the `numpy` (Oliphant 2006) and `scipy` (Jones et al. 2001) software packages. We also made extensive use of the `TREECORR` and `COSMOSIS` software packages and thank Mike Jarvis and Joe Zuntz for their continuing enhancements and clear documentation.

This project has received significant funding from the European Union’s Horizon 2020 research and innovation programme. We thank and acknowledge support from: the European Research Council under grant agreement No. 647112

(CH, TT, MA, CL and BG), No. 770935 (HHI, AHW, JLvdB and AD) and No. 693024 (SJ) in addition to the Marie Skłodowska-Curie grant agreement No. 797794 (TT). We also 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 (CH, FK); the Deutsche Forschungsgemeinschaft Heisenberg grant Hi 1495/5-1, (HHI); the Netherlands Organisation for Scientific Research Vici grant 639.043.512 (AK, HHO) and grant 621.016.402 (JdJ); the Alexander von Humboldt Foundation (KK); the Polish Ministry of Science and Higher Education through grant DIR/WK/2018/12, and the Polish National Science Center through grant no. 2018/30/E/ST9/00698 (MB); the Royal Society through an Enhancement Award RGF/EA/181006 (BG); the Australian Research Council grants DP160102235 and CE17010013 (KG); the Beecroft Trust (SJ); STFC grant ST/N000919/1 (LM); the NSFC of China under grant 11973070, the Shanghai Committee of Science and Technology grant No.19ZR1466600, and the Key Research Program of Frontier Sciences, CAS, Grant No. ZDBS-LY-7013 (HYS).

The results in this paper are based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017, 177.A-3018 and 179.A-2004, and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA and NWO-M grants; Target; the University of Padova, and the University Federico II (Naples). Data processing for VIKING has been contributed by the VISTA Data Flow System at CASU, Cambridge and WFAU, Edinburgh.

The BOSS-related results in this paper have been made possible thanks to SDSS-III. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

The 2dFLenS-related results are based on data acquired through the Australian Astronomical Observatory, under program A/2014B/008. It would not have been possible without the dedicated work of the staff of the AAO in the development and support of the 2dF-AAOmega system, and the running of the AAT.

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

References

- Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, Phys. Rev. D, 98, 043526
- Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 106
- Aiola, S., Calabrese, E., Maurin, L., et al. 2020, arXiv e-prints, arXiv:2007.07288
- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12
- Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617
- Amara, A. & Réfrégier, A. 2008, MNRAS, 391, 228
- Amon, A., Heymans, C., Klaes, D., et al. 2018, MNRAS, 477, 4285
- Asgari, M., Friswell, I., Yoon, M., et al. 2020a, arXiv e-prints, arXiv:2004.07811
- Asgari, M., Heymans, C., Hildebrandt, H., et al. 2019, A&A, 624, A134
- Asgari, M., Lin, C., & KiDS-Collaboration. 2020b, A&A, x, x
- Asgari, M., Tröster, T., Heymans, C., et al. 2020c, A&A, 634, A127
- Bartelmann, M. & Schneider, P. 2001, Phys. Rep., 340, 291
- Begeman, K., Belikov, A. N., Boxhoorn, D. R., & Valentijn, E. A. 2013, Experimental Astronomy, 35, 1
- Beran, R. 1977, The Annals of Statistics, 5, 445
- Bernstein, G. & Jain, B. 2004, ApJ, 600, 17
- Bernstein, G. M. 2009, ApJ, 695, 652
- Blake, C., Amon, A., Asgari, M., et al. 2020, arXiv e-prints, arXiv:2005.14351
- Blake, C., Amon, A., Childress, M., et al. 2016, MNRAS, 462, 4240
- Blazek, J. A., MacCrann, N., Troxel, M. A., & Fang, X. 2019, Phys. Rev. D, 100, 103506
- Bocquet, S., Dietrich, J. P., Schrabback, T., et al. 2019, ApJ, 878, 55
- Bose, B., Cataneo, M., Tröster, T., et al. 2020, arXiv e-prints, arXiv:2005.12184
- Bridle, S. & King, L. 2007, New Journal of Physics, 9, 444
- Brinckmann, T. & Lesgourges, J. 2019, Physics of the Dark Universe, 24, 100260
- Chan, K. C., Scoccimarro, R., & Sheth, R. K. 2012, Phys. Rev. D, 85, 083509
- Chang, C., Wang, M., Dodelson, S., et al. 2019, MNRAS, 482, 3696
- Chisari, N. E., Alonso, D., Krause, E., et al. 2019, ApJS, 242, 2
- Crocce, M. & Scoccimarro, R. 2006, Phys. Rev. D, 73, 063519
- de Jong, J. T. A., Kuijken, K., Applegate, D., et al. 2013, The Messenger, 154, 44
- DES Collaboration, Abbott, T., Aguena, M., et al. 2020, arXiv e-prints, arXiv:2002.11124
- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv e-prints, arXiv:1611.00036
- Desjacques, V., Jeong, D., & Schmidt, F. 2018, Phys. Rep., 733, 1
- Di Valentino, E., Melchiorri, A., Mena, O., & Vagnozzi, S. 2020, Phys. Rev. D, 101, 063502
- Drlica-Wagner, A., Sevilla-Noarbe, I., Rykoff, E. S., et al. 2018, ApJS, 235, 33
- eBOSS Collaboration, Alam, S., Aubert, M., et al. 2020, arXiv e-prints, arXiv:2007.08991
- Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32
- Efstathiou, G. & Lemos, P. 2018, MNRAS, 476, 151
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560
- Erben, T., Hildebrandt, H., Miller, L., et al. 2013, MNRAS, 433, 2545
- Feroz, F. & Hobson, M. P. 2008, MNRAS, 384, 449
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
- Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, The Open Journal of Astrophysics, 2, 10
- Fortuna, M. C., Hoekstra, H., Joachimi, B., et al. 2020, arXiv e-prints, arXiv:2003.02700
- Gaztañaga, E., Erikson, M., Crocce, M., et al. 2012, MNRAS, 422, 2904
- Giblin, B., Heymans, C., Asgari, M., et al. 2020, arXiv e-prints, arXiv:2007.01845
- Grieb, J. N., Sánchez, A. G., Salazar-Albornoz, S., et al. 2017, MNRAS, 467, 2085
- Hamana, T., Shirasaki, M., Miyazaki, S., et al. 2020, PASJ, 72, 16
- Handley, W. 2019, The Journal of Open Source Software, 4, 1414
- Handley, W. & Lemos, P. 2019, Phys. Rev. D, 100, 043504
- Heitmann, K., Lawrence, E., Kwan, J., Habib, S., & Higdon, D. 2014, ApJ, 780, 111
- Heydenreich, S., Schneider, P., Hildebrandt, H., et al. 2020, A&A, 634, A104
- Heymans, C., Grocutt, E., Heavens, A., et al. 2013, MNRAS, 432, 2433
- Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43
- Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020a, A&A, 633, A69
- Hildebrandt, H., Wright, A., & KiDS-Collaboration. 2020b, A&A, x, x
- Hu, W. & Jain, B. 2004, Phys. Rev. D, 70, 043009
- Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, arXiv e-prints, arXiv:1807.06209
- Jain, B. & Seljak, U. 1997, ApJ, 484, 560
- Jarvis, M. 2015, Astrophysics Source Code Library, ascl:1508.007
- Joachimi, B. & Bridle, S. L. 2010, A&A, 523, A1
- Joachimi, B., Lin, C. A., Asgari, M., et al. 2020, arXiv e-prints, arXiv:2007.01844
- Johnson, A., Blake, C., Amon, A., et al. 2017, MNRAS, 465, 4118
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python
- Joudaki, S., Blake, C., Heymans, C., et al. 2017, MNRAS, 465, 2033
- Joudaki, S., Blake, C., Johnson, A., et al. 2018, MNRAS, 474, 4894
- Joudaki, S., Hildebrandt, H., Traykova, D., et al. 2020, A&A, 638, L1
- Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, A&A, 624, A92
- Kilbinger, M., Bonnett, C., & Coupon, J. 2014, Astrophysics Source Code Library, ascl:1402.026
- Kitaura, F.-S., Rodríguez-Torres, S., Chuang, C.-H., et al. 2016, MNRAS, 456, 4156
- Köhlinger, F., Joachimi, B., Asgari, M., et al. 2019, MNRAS, 484, 3126
- 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
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193
- Lemos, P., Köhlinger, F., Handley, W., et al. 2019, arXiv e-prints, arXiv:1910.07820
- Lesgourges, J. 2011, arXiv e-prints, arXiv:1104.2932
- Lewis, A. 2019, arXiv e-prints, arXiv:1910.13970
- Lewis, A. & Bridle, S. 2002, Phys. Rev. D, 66, 103511
- Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
- Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, MNRAS, 452, 2087
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, LSST Science Book, Version 2.0
- Mandelbaum, R. 2018, ARA&A, 56, 393

- Mead, A. J., Peacock, J. A., Heymans, C., Joudaki, S., & Heavens, A. F. 2015, MNRAS, 454, 1958
- Mead, A. J., Tröster, T., Heymans, C., Van Waerbeke, L., & McCarthy, I. G. 2020, arXiv e-prints, arXiv:2005.00009
- Miller, L., Heymans, C., Kitching, T. D., et al. 2013, MNRAS, 429, 2858
- Muir, J., Bernstein, G. M., Huterer, D., et al. 2020, MNRAS, 494, 4454
- Nelder, J. & Mead, R. 1965, "Comput. J.", 7, 308
- Newman, J. A. 2008, ApJ, 684, 88
- Oliphant, T. E. 2006, A guide to NumPy, Vol. 1 (Trelgol Publishing USA)
- Olive, K. A. & Particle Data Group. 2014, Chinese Physics C, 38, 090001
- Palanque-Delabrouille, N., Yèche, C., Schöneberg, N., et al. 2020, J. Cosmology Astropart. Phys., 2020, 038
- Peebles, P. J. E. 1980, The large-scale structure of the universe (Princeton University Press)
- Poulin, V., Boddy, K. K., Bird, S., & Kamionkowski, M. 2018, Phys. Rev. D, 97, 123504
- Powell, M. J. D. 1964, The Computer Journal, 7, 155
- Raveri, M. & Hu, W. 2019, Phys. Rev. D, 99, 043506
- Raveri, M., Zacharekias, G., & Hu, W. 2020, Phys. Rev. D, 101, 103527
- Richard, J., Kneib, J. P., Blake, C., et al. 2019, The Messenger, 175, 50
- Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, ApJ, 876, 85
- Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, ApJ, 826, 56
- Ross, A. J., Beutler, F., Chuang, C.-H., et al. 2017, MNRAS, 464, 1168
- Ross, A. J., Percival, W. J., Sánchez, A. G., et al. 2012, MNRAS, 424, 564
- Sánchez, A. G. 2020, arXiv e-prints, arXiv:2002.07829
- Sánchez, A. G., Scoccimarro, R., Crocce, M., et al. 2017, MNRAS, 464, 1640
- Schneider, P., van Waerbeke, L., Kilbinger, M., & Mellier, Y. 2002, A&A, 396, 1
- Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, ApJ, 859, 101
- Sellentin, E. 2020, MNRAS, 492, 3396
- Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M. P., & McCarthy, I. G. 2011, MNRAS, 417, 2020
- Takada, M. & Hu, W. 2013, Phys. Rev. D, 87, 123504
- Tröster, T., Asgari, M., & the KiDS-Collaboration. in prep., A&A, x, x
- Tröster, T., Sánchez, A. G., Asgari, M., et al. 2020, A&A, 633, L10
- Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018, Phys. Rev. D, 98, 043528
- Vakili, M., Bilicki, M., Hoekstra, H., et al. 2019, MNRAS, 487, 3715
- van Daalen, M. P., Schaye, J., Booth, C. M., & Dalla Vecchia, C. 2011, MNRAS, 415, 3649
- van den Busch, J. L., Hildebrandt, H., Wright, A. H., et al. 2020, arXiv e-prints, arXiv:2007.01846
- van Uitert, E., Joachimi, B., Joudaki, S., et al. 2018, MNRAS, 476, 4662
- Weinberg, S. 1972, Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity (John Wiley & Sons)
- 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, arXiv e-prints, arXiv:2005.04207
- York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579
- Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45

Table A.1: Galaxy properties for the BOSS and 2dFLenS lens ('L') samples and the KiDS-1000 source ('S') samples.

ID	z_{\min}	z_{\max}	mean z	n_{eff}	$\sigma_{\epsilon,i}$	m
KiDS-1000:						
S1	0.1	0.3	0.26	0.62	0.27	-0.009 ± 0.019
S2	0.3	0.5	0.40	1.18	0.26	-0.011 ± 0.020
S3	0.5	0.7	0.56	1.85	0.27	-0.015 ± 0.017
S4	0.7	0.9	0.79	1.26	0.25	0.002 ± 0.012
S5	0.9	1.2	0.98	1.31	0.27	0.007 ± 0.010
BOSS:						
L1	0.2	0.5	0.38	0.014	-	-
L2	0.5	0.75	0.61	0.016	-	-
2dFLenS:						
L1	0.2	0.5	0.36	0.006	-	-
L2	0.5	0.75	0.60	0.006	-	-

Notes. Columns include the bin identifier, ID, the minimum and maximum redshift selection for the bin, $z_{\min/\max}$, which applies to photometric redshifts for the sources S, and spectroscopic redshifts for the lenses L, along with the mean redshift of the bin and the effective galaxy number density, per square arcminute, n_{eff} . For the source bins we also include the measured ellipticity dispersion per component, $\sigma_{\epsilon,i}$, and the shear calibration correction, m , and its uncertainty.

Appendix A: Galaxy properties and the $3 \times 2\text{pt}$ covariance

This Appendix tabulates the properties of the KiDS-1000 tomographic source samples, along with the properties of the BOSS and 2dFLenS lens samples, in Table A.1. We list the spectroscopic redshift selection for the lenses ($z_{\min} < z_s \leq z_{\max}$), and the photometric redshift selection for the sources ($z_{\min} < z_B \leq z_{\max}$), along with the mean redshift of each sample. For the source sample, the true redshift distributions are estimated in Hildebrandt et al. (2020b), using the SOM methodology from Wright et al. (2020a). The shear calibration correction, m , which can also be referred to in the literature as the responsivity, $R = 1 + m$, is listed for each source bin (Kannawadi et al. 2019). The effective number density of lenses and sources defines the number of galaxies per square arcminute in the case of unit weights and, for the sources, unit responsivity (see equations C.11 and C.13 in Joachimi et al. 2020). We also list the effective ellipticity dispersion, $\sigma_{\epsilon,i}$, per ellipticity component, i , for each the weighted and calibrated source galaxy samples (equation C.8 in Joachimi et al. 2020).

Fig. A.1 displays the correlation coefficients of the $3 \times 2\text{pt}$ covariance matrix for the three observables; cosmic shear E-mode power spectra, C_E , galaxy-galaxy lensing E-mode power spectra, C_{ne} , and the anisotropic galaxy clustering in low and high redshift bins, ξ_{gg} (see Sect. 2 for details). The cross-correlation between the two lensing and the clustering observables is set to zero, as mock data analyses showed these correlations to be negligible for the KiDS and BOSS footprints (Joachimi et al. 2020). The Fourier-space lensing observables are shown to be significantly less correlated between ℓ -scales, in comparison to the physical-scale clustering observables.

Appendix B: Parameter priors

This Appendix tabulates the adopted KiDS-1000 priors and sampling parameters in Table B.1. The uniform prior on the dimen-

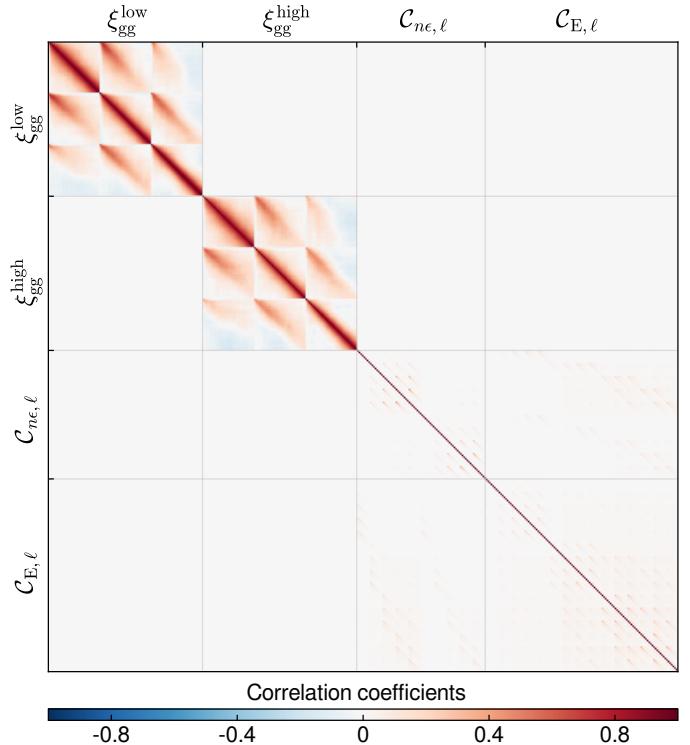


Fig. A.1: Correlation coefficients of the $3 \times 2\text{pt}$ covariance matrix for the cosmic shear, C_E , galaxy-galaxy lensing, C_{ne} and galaxy clustering observables, ξ_{gg} (see Sect. 2 for details). Here the band powers C , are related to the angular power spectrum C , in Eq. (1) and Eq. (3), as $C = \ell^2 C/(2\pi)$.

sionless Hubble constant, h , reflects distance-ladder $\pm 5\sigma$ constraints from Riess et al. (2016), which encompasses the value of h favoured by *Planck* Collaboration et al. (2018). The uniform prior on the baryon density, $\omega_b = \Omega_b h^2$, reflects big bang nucleosynthesis $\pm 5\sigma$ constraints from Olive & Particle Data Group (2014). The uniform prior on the CDM density, $\omega_c = \Omega_c h^2$, reflects Supernova Type Ia $\pm 5\sigma$ constraints on Ω_m from Scolnic et al. (2018) combined with the most extreme allowed values of h and ω_b , given their priors.

As discussed in Sect. 2.6 we choose to sample with an uninformative uniform prior on S_8 to avoid implicit informative priors from a uniform prior on the primordial power spectrum amplitude A_s . We choose a fixed model for the properties of neutrinos, adopting the normal hierarchy at the minimum sum of masses, $\Sigma m_\nu = 0.06$ eV, following *Planck* Collaboration et al. (2018). We consider extended models, including variations on neutrino mass in Tröster et al. (in prep.).

The uniform prior on the scalar spectral index, n_s , reflects a restriction in our likelihood implementation, where the Sánchez et al. (2017) galaxy clustering likelihood becomes prohibitively slow for $n_s > 1.1$. With the upper limit of the top-hat prior fixed by this computational limitation, we choose to symmetrise the prior around the theoretical expectation of $n_s = 0.97$. In Fig. B.1 we demonstrate the impact of this informative prior on the BOSS-only galaxy clustering constraints, highlighting how informative this choice of prior is. We argue that adopting an informative prior is justified however, given our theoretical prior knowledge of the Harrison-Zel'dovich spectrum. Fig. B.1 also helps to illustrate that had we chosen an uninformative prior on n_s for our KiDS-1000-BOSS analysis, a decision taken more

Table B.1: KiDS-1000 sampling parameters and priors.

Parameter	Symbol	Prior
Hubble constant	h	[0.64, 0.82]
Baryon density	ω_b	[0.019, 0.026]
CDM density	ω_c	[0.051, 0.255]
Density fluctuation amp.	S_8	[0.1, 1.3]
Scalar spectral index	n_s	[0.84, 1.1]
Linear galaxy bias	b_1 [2]	[0.5, 9]
Quadratic galaxy bias	b_2 [2]	[-4, 8]
Non-local galaxy bias	γ_3^- [2]	[-8, 8]
Virial velocity parameter	a_{vir} [2]	[0, 12]
Intrinsic alignment amp.	A_{IA}	[-6, 6]
Baryon feedback amp.	A_{bary}	[2, 3.13]
Redshift offsets	δ_z	$\mathcal{N}(\mu; C_{\delta z})$

Notes. Primary cosmological parameters for the flat Λ CDM model are listed in the first section. The second section lists astrophysical nuisance parameters to model galaxy bias (with independent parameters for each of the two BOSS redshift bins as indicated with the bracket [2]), intrinsic galaxy alignments, and baryon feedback. Observational redshift nuisance parameters are listed in the final section. Prior values in square brackets are the limits of the adopted uniform top-hat priors. $\mathcal{N}(\mu; C)$ corresponds to a five dimensional multivariate Gaussian prior with mean μ and covariance $C_{\delta z}$.

than a year before unblinding our analysis, this would have likely served to exacerbate any tension with the Planck CMB constraints.

Turning to astrophysical priors, the galaxy bias parameter top-hat priors on b_1 , b_2 , γ_3^- , (Eq. 3), and on a_{vir} (see the ‘fingers of god’ model in equations 6 to 9 of Joachimi et al. 2020), match those adopted in Tröster et al. (2020), which cover a wider range than those used in Sánchez et al. (2017). The two BOSS redshift slices have independent sets of parameters. Wide uniform priors for the intrinsic alignment parameter A_{IA} are chosen to be uninformative. Uniform priors on the baryon feedback parameter A_{bary} are chosen such that the resulting Mead et al. (2015) model of the non-linear matter power spectrum encompasses both the most aggressive feedback model from the van Daalen et al. (2011) suite of hydrodynamical simulations, along with the dark matter-only case where $A_{\text{bary}} = 3.13$.

There are five additional correlated nuisance parameters, δ_z^i , that model uncertainty in the mean of the source redshift distributions. We adopt a multivariate Gaussian prior for the vector δ_z with a mean $\mu = (0.0001, 0.0021, 0.0129, 0.0110, -0.0060)$, and a covariance, $C_{\delta z}$, as calibrated using mock galaxy catalogues in Wright et al. (2020a). The diagonal terms of $C_{\delta z}$ are typically at the level of $\sim (0.01)^2$, with off-diagonal correlation coefficients ranging between ~ 0.1 and ~ 0.3 (see section 3 and figure 2 of Hildebrandt et al. 2020b, for details).

Appendix C: Parameter constraints

This Appendix tabulates the maximum posterior (MAP) and marginalised constraints on the flat Λ CDM cosmological parameters, in Table C.1, for the different combinations of the three large-scale structure probes considered in this work. For constraints from KiDS-1000 cosmic shear alone, we refer the reader to Asgari et al. (2020b).

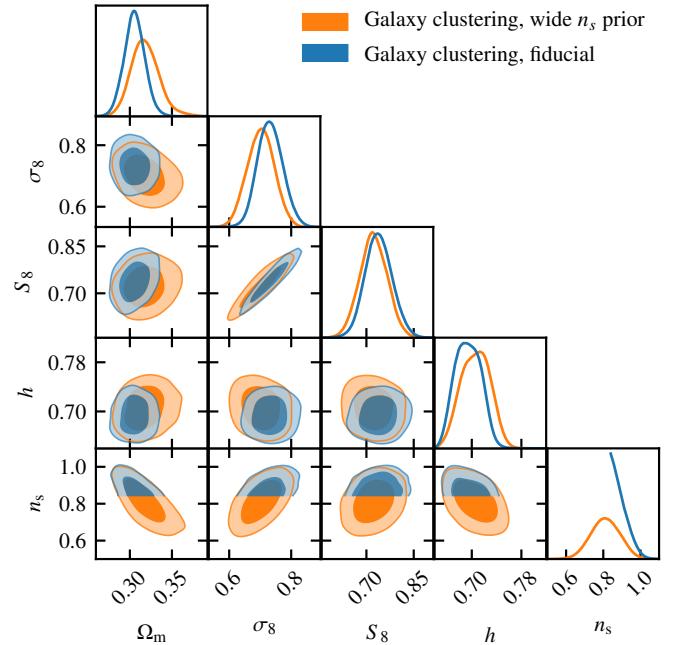


Fig. B.1: The impact of the n_s prior: Comparing marginalised posterior distributions for the BOSS galaxy clustering analysis for our fiducial analysis (blue) with the constraints when adopting an uninformative prior on n_s (orange). Opening the parameter space to arguably unphysical values of n_s favours higher values and weaker constraints on Ω_m . Constraints on S_8 are, however, fairly insensitive to the choice of n_s prior.

Table C.1: Parameter constraints for the probe combinations considered in this work: $3 \times 2\text{pt}$, cosmic shear (CS) and galaxy-galaxy lensing (GGL), cosmic shear and galaxy clustering (GC), and galaxy clustering by itself. We refer the reader to Asgari et al. (2020b) for cosmic shear-only constraints. For each probe, the first column lists the MAP value with the PJ-HPD CI. If the PJ-HPD CI could not be robustly determined, no uncertainty estimate is provided. The second column for each probe lists the maximum of the marginal posterior, together with the marginal HPD CI. Parameter for which the marginal probability at either prior edge exceeds 13% of the peak probability are deemed unconstrained (see appendix A of Asgari et al. 2020b), and are denoted by a dash. Finally, parameters that are not sampled are left blank.

Parameter	$3 \times 2\text{pt}$ (joint)	$3 \times 2\text{pt}$ (marginal)	GC (joint)	GC (marginal)	CS+GGL (joint)	CS+GGL (marginal)	CS+GC (joint)	CS+GC (marginal)
S_8	$0.766^{+0.02}_{-0.014}$	$0.765^{+0.017}_{-0.016}$	$0.762^{+0.054}_{-0.05}$	$0.743^{+0.038}_{-0.05}$	$0.774^{+0.031}_{-0.029}$	$0.761^{+0.024}_{-0.025}$	$0.766^{+0.018}_{-0.015}$	$0.763^{+0.016}_{-0.016}$
$h^2 \Omega_c$	0.123	$0.124^{+0.0096}_{-0.011}$	$0.136^{+0.0065}_{-0.018}$	$0.123^{+0.011}_{-0.009}$	$0.0721^{+0.089}_{-0.016}$	$0.12^{+0.057}_{-0.034}$	$0.126^{+0.075}_{-0.013}$	$0.124^{+0.01}_{-0.011}$
$h^2 \Omega_b$	0.0225	—	$0.0246^{+0.0012}_{-0.0037}$	—	0.0258	—	$0.0219^{+0.0022}_{-0.0022}$	—
h	$0.695^{+0.03}_{-0.019}$	$0.696^{+0.02}_{-0.026}$	$0.716^{+0.02}_{-0.035}$	$0.687^{+0.029}_{-0.018}$	$0.642^{+0.1}_{-1.2e-05}$	—	$0.695^{+0.022}_{-0.023}$	$0.69^{+0.026}_{-0.02}$
n_s	0.901	—	$0.851^{+0.06}_{-0.01}$	—	0.947	—	$0.894^{+0.05}_{-0.042}$	—
Abaray	$3.13^{+1.6e-05}_{-0.47}$	—	—	—	3.13	—	3.13	—
A_{PA}	$1.07^{+0.27}_{-0.31}$	$1.01^{+0.31}_{-0.28}$	$2.8^{+8.8}_{-12} \times 10^{-3}$	$1.41^{+8.4}_{-11} \times 10^{-3}$	$0.972^{+0.25}_{-0.26}$	$0.947^{+0.44}_{-0.31}$	$0.892^{+0.35}_{-0.36}$	$2.45^{+9.5}_{-12} \times 10^{-3}$
δz_1	$1.93^{+11}_{-10} \times 10^{-3}$	$9.09^{+11}_{-9.7} \times 10^{-3}$	—	$0.0102^{+0.014}_{-0.0069}$	$8.75^{+11}_{-8.7} \times 10^{-3}$	$0.0117^{+0.013}_{-0.0085}$	$0.0121^{+0.0095}_{-0.012}$	—
δz_2	$0.0101^{+0.016}_{-0.0068}$	$-0.0204^{+0.01}_{-0.0093}$	—	$-0.0206^{+0.0091}_{-0.0091}$	$-0.0199^{+0.0092}_{-0.0092}$	$-0.013^{+0.0091}_{-0.0091}$	$-0.0119^{+0.0098}_{-0.012}$	$-0.0119^{+0.0098}_{-0.011}$
δz_3	$-0.0207^{+0.0093}_{-0.01}$	—	—	$-0.0138^{+0.0066}_{-0.0066}$	$-0.012^{+0.0062}_{-0.0099}$	$-0.0166^{+0.0098}_{-0.0064}$	$-0.0156^{+0.0077}_{-0.0085}$	—
δz_4	$-0.0143^{+0.0076}_{-0.008}$	$-0.0145^{+0.0082}_{-0.0076}$	$4.54^{+10}_{-8.7} \times 10^{-3}$	$6.06^{+7.3}_{-10} \times 10^{-3}$	$6.85^{+8.7}_{-8.6} \times 10^{-3}$	$6.5^{+12}_{-7.1} \times 10^{-3}$	$6.54^{+9.4}_{-9} \times 10^{-3}$	—
δz_5	$5.23^{+8.3}_{-10} \times 10^{-3}$	—	—	$3.29^{+0.3}_{-0.81}$	$3.14^{+0.55}_{-0.55}$	$1.99^{+0.69}_{-0.67}$	$2.02^{+0.085}_{-0.07}$	—
$b_{1\text{lowz}}$	$2^{+0.79}_{-0.079}$	$2.04^{+0.064}_{-0.093}$	$2.01^{+0.13}_{-0.16}$	$2.09^{+0.11}_{-0.13}$	0.968	—	0.443^{+1}_{-1}	$0.326^{+1}_{-0.07}$
$b_{1\text{lowz}}^2$	$0.3^{+0.62}_{-0.69}$	$0.289^{+0.71}_{-0.61}$	$0.475^{+1.1}_{-1.1}$	$0.624^{+0.96}_{-1}$	$0.752^{+0.62}_{-0.56}$	—	$0.594^{+0.58}_{-0.57}$	$0.853^{+0.53}_{-0.52}$
$b_{2\text{lowz}}$	$0.695^{+0.48}_{-0.54}$	$0.879^{+0.44}_{-0.52}$	$0.688^{+0.59}_{-0.57}$	$0.752^{+0.62}_{-0.56}$	7.87	—	$4.28^{+1.2}_{-1}$	$4.37^{+1.1}_{-1.1}$
$\gamma_{3-}^{\text{lowz}}$	$4.14^{+0.84}_{-0.97}$	$4.24^{+0.86}_{-0.96}$	$4.3^{+1.1}_{-1.2}$	$4.44^{+1.1}_{-1.2}$	—	—	$2.17^{+0.92}_{-0.84}$	$2.17^{+0.92}_{-0.84}$
$a_{\text{vir}}^{\text{lowz}}$	$2.13^{+0.95}_{-0.99}$	$2.17^{+0.93}_{-0.93}$	$2.14^{+0.16}_{-0.16}$	$2.23^{+0.13}_{-0.14}$	$2.3^{+0.33}_{-0.68}$	$2.26^{+0.48}_{-0.52}$	$2.12^{+0.1}_{-0.094}$	$2.12^{+0.1}_{-0.094}$
$b_{1\text{highz}}$	$-1.15^{+0.75}_{-0.41}$	$-1.03^{+0.72}_{-0.46}$	$-0.948^{+2.3}_{-0.75}$	$-0.901^{+2.3}_{-1}$	$5.82^{+2.3}_{-3.2}$	—	$-1.16^{+1.8}_{-0.48}$	$-0.951^{+1.7}_{-0.85}$
$b_{2\text{highz}}$	$0.369^{+0.8}_{-0.79}$	$0.505^{+0.9}_{-0.62}$	$-0.126^{+0.94}_{-1.1}$	$0.452^{+0.7}_{-1.2}$	$8^{+0.0058}_{-5.4}$	—	$-0.0141^{+0.91}_{-0.91}$	$0.412^{+0.8}_{-0.81}$
$\gamma_{3-}^{\text{highz}}$	$1.18^{+1.5}_{-0.97}$	—	$1.84^{+2.4}_{-1.8}$	—	—	$1.32^{+2.7}_{-1.3}$	—	—
Ω_m	$0.305^{+0.01}_{-0.015}$	$0.306^{+0.012}_{-0.013}$	$0.313^{+0.0097}_{-0.017}$	$0.307^{+0.011}_{-0.015}$	$0.239^{+0.11}_{-0.076}$	$0.29^{+0.087}_{-0.07}$	$0.306^{+0.013}_{-0.012}$	$0.307^{+0.012}_{-0.014}$
σ_8	$0.76^{+0.025}_{-0.02}$	$0.759^{+0.02}_{-0.045}$	$0.75^{+0.048}_{-0.047}$	$0.725^{+0.048}_{-0.036}$	$0.867^{+0.2}_{-0.17}$	$0.729^{+0.13}_{-0.098}$	$0.758^{+0.019}_{-0.024}$	—
σ_{12}	$0.743^{+0.03}_{-0.026}$	$0.737^{+0.029}_{-0.027}$	$0.717^{+0.058}_{-0.035}$	$0.707^{+0.046}_{-0.04}$	$0.887^{+0.15}_{-0.23}$	$0.72^{+0.095}_{-0.14}$	$0.734^{+0.03}_{-0.027}$	$0.732^{+0.03}_{-0.024}$
S_{12}	$0.754^{+0.021}_{-0.013}$	$0.754^{+0.015}_{-0.018}$	$0.751^{+0.038}_{-0.054}$	$0.728^{+0.041}_{-0.045}$	$0.77^{+0.067}_{-0.046}$	$0.757^{+0.036}_{-0.05}$	$0.753^{+0.022}_{-0.013}$	$0.754^{+0.014}_{-0.019}$
A_s	$1.85^{+0.25}_{-0.17} \times 10^{-9}$	$1.79^{+0.22}_{-0.18} \times 10^{-9}$	$1.7^{+0.33}_{-0.16} \times 10^{-9}$	$1.64^{+0.27}_{-0.18} \times 10^{-9}$	$5.34^{+5.2}_{-4.3} \times 10^{-9}$	—	$1.81^{+0.27}_{-0.15} \times 10^{-9}$	$1.78^{+0.18}_{-0.18} \times 10^{-9}$
$100\theta_{\text{MC}}$	1.05	$1.05^{+0.011}_{-0.015}$	$1.06^{+0.0082}_{-0.019}$	$1.05^{+0.012}_{-0.013}$	$0.96^{+0.13}_{-0.023}$	$1.09^{+0.039}_{-0.058}$	$1.05^{+0.012}_{-0.016}$	$1.05^{+0.012}_{-0.014}$

Appendix D: Sensitivity tests

In this Appendix we present, with Fig. D.1, the marginalised posterior distributions for a selection of the $3 \times 2\text{pt}$ sensitivity tests explored in Sect. 3. These tests were all shown to recover consistent constraints on S_8 , in Fig. 5. Here we explore these tests in more detail.

We compare our analysis which fully marginalises over our uncertainty in the baryon feedback parameter (red), with our ‘No baryon’ case (sea-green), where $A_{\text{bary}} = 3.13$, corresponding to the non-linear matter power spectrum for a dark-matter only cosmology. Here we find very little difference, as our $3 \times 2\text{pt}$ analysis already favours high values of A_{bary} . Our choice of scales and adoption of the band-power cosmic shear statistic for this analysis also makes us less sensitive to uncertainties in the baryon feedback parameter, compared to a standard two-point correlation function analysis (Asgari et al. 2020c).

The removal of the two highest photometric redshift bins (blue and purple) primarily impacts S_8 . These two bins carry the majority of the signal-to-noise in our analysis, and so it is not surprising that the removal of nearly half the constraining data, in each case, can result in $\sim 1\sigma$ changes in the recovered S_8 (see for example the discussion in Appendix E). We refer the reader to Asgari et al. (2020b) where we present a detailed internal consistency analysis of the cosmic shear signal, following Köhlinger et al. (2019), (see also Efstathiou & Lemos 2018), concluding that the two highest photometric redshift bins are consistent with the full data set. A potential $\sim 3\sigma$ -level flag is, however, raised in Asgari et al. (2020b), over the internal consistency of the second tomographic bin. In our analysis where we remove the two lowest photometric redshift bins (orange), we find that these bins contribute very little to the S_8 constraint, and only serve to tighten the constraints on the intrinsic alignment parameter A_{IA} .

Finally we turn to the galaxy bias test (lime green), where we limit the analysis to a linear galaxy bias model, b_1 , setting all higher-order bias terms in Eq. (3) to zero¹⁶, as well as imposing a Gaussian galaxy velocity distribution by setting a_{vir} to zero. We can see that this biases the recovered cosmological constraints, as the amplitude of the linear galaxy bias increases in an attempt to model the enhanced power on small scales that the non-linear bias induces. The erroneous increase in b_1 leads to a decrease in the recovered S_8 , and also an overall reduction in the goodness-of-fit of the model with $\chi^2_{\text{MAP}} = 379.8$ for ~ 300 degrees of freedom. This result should serve as a point of caution for $3 \times 2\text{pt}$ analyses that adopt an effective linear bias model (see also the discussion in Asgari et al. 2020a), although we note that our analysis is particularly sensitive to the galaxy bias model given the high signal-to-noise BOSS clustering observations that probe physical scales as low as $s_{\min} = 20h^{-1}\text{ Mpc}$.

Appendix E: Expected S_8 differences between partially overlapping weak lensing surveys

In this Appendix we construct a simple model to estimate the expected statistical fluctuation in S_8 constraints from partially overlapping weak lensing surveys, specifically our previous (KV450, Wright et al. 2020b), and current KiDS analyses. Assuming that we only wish to constrain the overall amplitude of the measurement, the inference can be approximated by a lin-

¹⁶ Removing all higher-order galaxy bias terms also requires the uncoupling of b_1 from γ_2 .

ear least-squares problem, where S_8 , the parameter of interest, is Gaussian distributed.

In the following, measurements derived from the KV450 footprint are denoted by an X in the subscript, while measurements on the new area added for KiDS-1000 are denoted with a Y . Measurements derived from the full KiDS-1000 footprint are denoted with a Z . Now, let S_X^{KV450} denote the value of S_8 estimated from the KV450 footprint, using the KV450 methodology; $S_Y^{\text{KiDS-1000}}$ the value estimated from the newly added area using the KiDS-1000 methodology; and $S_Z^{\text{KiDS-1000}}$ the KiDS-1000 value estimated from the full footprint. Approximating the estimates from disjoint footprints as independent, the S_8 measurement from the full KiDS-1000 area is then given by the area-weighted average¹⁷

$$S_Z^{\text{KiDS-1000}} = \frac{A_X S_X^{\text{KiDS-1000}} + A_Y S_Y^{\text{KiDS-1000}}}{A_Z}. \quad (\text{E.1})$$

Here A_X is the effective area of the KV450 footprint, A_Z is the effective area of the KiDS-1000 footprint, and $A_Y = A_Z - A_X$, is the additional area added between the two data releases. The uncertainty $\sigma_X^{\text{KiDS-1000}}$ on the measurement $S_X^{\text{KiDS-1000}}$ is then related to the uncertainty $\sigma_Z^{\text{KiDS-1000}}$ on measurements $S_Z^{\text{KiDS-1000}}$, as

$$\sigma_X^{\text{KiDS-1000}} = \sqrt{\frac{A_Z}{A_X}} \sigma_Z^{\text{KiDS-1000}}, \quad (\text{E.2})$$

and analogously for $\sigma_Y^{\text{KiDS-1000}}$. Note that the uncertainty σ_X^{KV450} differs from $\sigma_X^{\text{KiDS-1000}}$ due to differences in the KiDS-1000 and KV450 methodologies, such as the adopted m -calibration uncertainty.

We define $\Delta = S_Z^{\text{KiDS-1000}} - S_X^{\text{KV450}}$, the offset between the KiDS-1000 and KV450 S_8 measurements. The uncertainty on Δ is then given by

$$\sigma_\Delta^2 = (\sigma_X^{\text{KV450}})^2 + (\sigma_Z^{\text{KiDS-1000}})^2 - 2 \sqrt{\frac{A_X}{A_Z}} \sigma_X^{\text{KV450}} \sigma_Z^{\text{KiDS-1000}}, \quad (\text{E.3})$$

where we have assumed a 100% correlation between the KiDS-1000 and KV450 measurements within the KV450 footprint area X , but approximate the areas X and Y as fully uncorrelated. This approximation neglects the large-scale correlations between X and Y . Given that the majority of the error budget for KiDS-1000 stems from the random shape noise component, which is independent between X and Y , this approximation is sufficient for this toy model. The absolute difference, $|\Delta|$ has the cumulative distribution function, CDF,

$$\text{CDF}(x) = \text{erf}\left(\frac{x}{\sigma_\Delta \sqrt{2}}\right), \quad (\text{E.4})$$

and an expectation value $E[|\Delta|] = \sqrt{2/\pi} \sigma_\Delta$.

We now compare the actual KV450 and KiDS-1000 S_8 constraints, given the effective areas $A_{\text{KV450}} = A_X = 341.3\text{ deg}^2$ and $A_{\text{KiDS-1000}} = A_Z = 777.4\text{ deg}^2$. Using the marginal S_8 constraints from the two-point shear correlation function analysis, comparing Wright et al. (2020b, KV450: $S_8 = 0.716^{+0.043}_{-0.038}$) with Asgari et al. (2020b, KiDS-1000: $S_8 = 0.768^{+0.016}_{-0.020}$), we find $|\Delta| = 0.052 = 1.6\sigma_\Delta$, which can be compared against the expected offset of $E[|\Delta|] = 0.026$. We expect to find an offset of this, or a larger magnitude, 10% of the time.

¹⁷ The amplitude of the weak lensing signal scales roughly with S_8 (Jain & Seljak 1997), with an uncertainty variance that scales with the inverse survey area (Schneider et al. 2002).

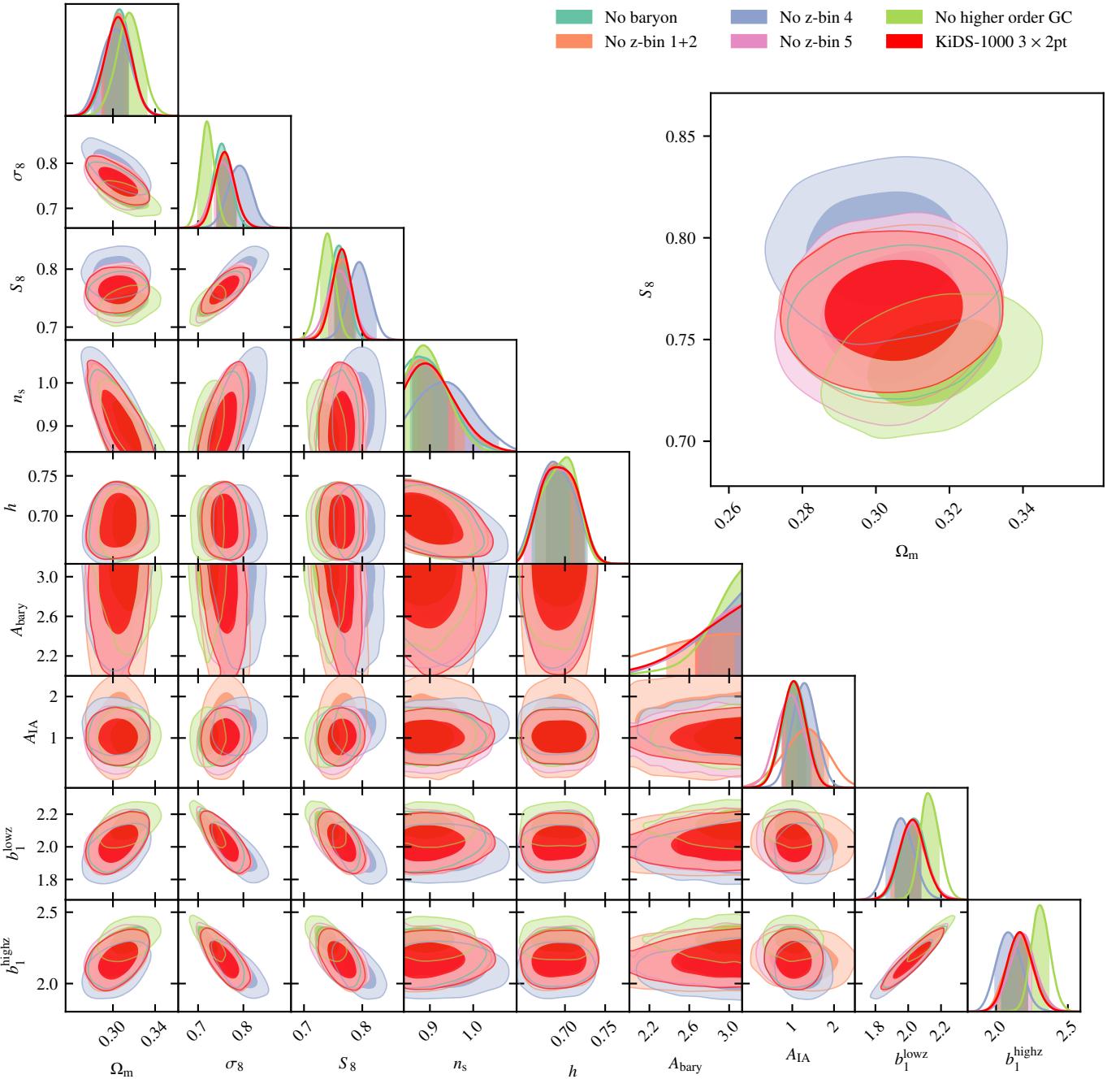


Fig. D.1: Marginalised posterior distributions for the extended set of cosmological parameters shown in Fig. 6, comparing the fiducial $3 \times 2\text{pt}$ analysis (red) to a selection of our sensitivity test analyses where we ignore the impact of baryon feedback (the ‘No baryon’ case, sea-green), limit the analysis to a linear galaxy bias model (the ‘No higher order GC’ case, lime-green), and remove individual tomographic bins from our weak lensing observables (orange, purple and pink).

This simple model analysis is sufficient to conclude that the increase in S_8 that we find between KV450 and KiDS-1000 is consistent with the expectation from simple statistical fluctuations. A complete assessment could be conducted by analysing a reasonable fraction of the 20 000 KiDS mock catalogues from Joachimi et al. (2020), with the KiDS-1000 and KV450 footprints imposed. Unfortunately this is, however, out-of-scope for this analysis, given the significant compute power that would be incurred.

Appendix F: Tension estimators

In this Appendix we review the range of tension estimators employed in Sect. 3.2 to quantify the tension between the cosmological parameter constraints from *Planck* and our KiDS-1000-BOSS $3 \times 2\text{pt}$ analysis. These are in addition to the standard Gaussian offset measure, τ , in Eq. (8), and the Bayes factor, R , in Eq. (9).

Appendix F.1: Hellinger tension measure: d_H

To assess the tension between the marginal S_8 distributions of two experiments, we have employed the widely used Gaussian-approximation measure, τ , given in Eq. (8). As an alternative that does not depend on the Gaussian assumption, we construct a sample-based, one-dimensional tension estimator adopting the Hellinger distance d_H . The Hellinger distance is widely used in statistics as a stable metric in optimisation problems that require the comparison of distributions (e.g. Beran 1977). It is given by

$$d_H^2 [p; q] = \frac{1}{2} \int dx \left[\sqrt{p(x)} - \sqrt{q(x)} \right]^2, \quad (\text{F.1})$$

where the integral runs over the support of the PDFs of the distributions p and q . We choose this distance definition as it is symmetric, avoids PDFs in the denominator and the associated instability for sample estimates, and has a closed-form expression for the case of two Gaussians:

$$d_H^2 \left[\mathcal{N}(\mu_1, \sigma_1^2); \mathcal{N}(\mu_2, \sigma_2^2) \right] = 1 - \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}} \exp \left\{ -\frac{(\mu_1 - \mu_2)^2}{4(\sigma_1^2 + \sigma_2^2)} \right\}, \quad (\text{F.2})$$

where the normal distributions have means $\mu_{1,2}$ and variances $\sigma_{1,2}^2$.

We estimate the Hellinger distance between our samples via a discretised version of Eq. (F.1). The PDFs are built from the samples in two ways: through binning into a histogram (with the number of bins chosen to scale with the square root of the number of samples), and through kernel density estimation (using a Gaussian kernel with the width chosen by Scott's rule). The former approach will in general over-estimate tension as tails are underpopulated due to the discrete sampling, whereas the latter approach under-estimates tension due to the smoothing of the PDFs by the kernel density estimator. Empirically, we find that a simple average of the two approaches yields good accuracy over the few- σ tension range that we are interested in.

We invert Eq. (F.2) to determine the expected difference in the means, $|\mu_1 - \mu_2|$ if the two distributions were Gaussian, for a given value of d_H . To do so, we use the standard deviations estimated from the samples. The resulting difference in the means is inserted into Eq. (8) to obtain a measure of tension in the familiar σ units¹⁸.

Comparing the $3 \times 2\text{pt}$ and *Planck* S_8 distributions, we find a Hellinger distance $d_H = 0.95$, which corresponds to a tension of 3.1σ .

Appendix F.2: Distribution of parameter shifts: p_S

Following Raveri et al. (2020, see also Köhlinger et al. 2019), we consider the distribution $P(\Delta\theta)$ of the parameter shifts $\Delta\theta = \theta_{\text{Planck}} - \theta_{3 \times 2\text{pt} + \text{Planck}}$. The significance of any shift can then be assessed by calculating how much of $P(\Delta\theta)$ is enclosed within the iso-probability surface at $P(0)$:

$$p_S = \int_{P(\Delta\theta) > P(0)} P(\Delta\theta) d\Delta\theta. \quad (\text{F.3})$$

¹⁸ Our Hellinger tension measure is validated by comparing to Eq. (8) for Gaussian random variates with the same number of realisations as the sizes of our posterior samples. We find that systematic deviations in the Hellinger tension are at 0.1σ or less for tension up to 4σ , with a standard deviation of $\pm 0.05\sigma$.

In other words, how much of the distribution of parameter shifts $P(\Delta\theta)$ is more likely than the probability of no shifts. Accurately evaluating the integral in Eq. (F.3) when $P(\Delta\theta)$ is not centred on zero requires very large numbers of samples to sufficiently cover the tails of the distribution. For a single parameter, S_8 , we find $p_S = 0.9981$, with a PTE of 0.0019, or 3.1σ . This is in agreement with the other single-parameter tension estimators (8) and (F.1). For the five shared parameters between our $3 \times 2\text{pt}$ analysis and *Planck* we find lower significances, around 2σ , albeit with very large uncertainties owing to the insufficient sampling of the tails of the parameter shift distribution.

Appendix F.3: Suspiciousness: S

Handley & Lemos (2019) propose the ‘suspiciousness’ statistic S based on the Bayes factor, R , but hardened against prior dependences. They define

$$\ln S = \ln R - \ln I, \quad (\text{F.4})$$

where $\ln R$ is the logarithm of the Bayes factor in Eq. (9):

$$\ln R = \ln Z_{3 \times 2\text{pt} + \text{Planck}} - \ln Z_{3 \times 2\text{pt}} - \ln Z_{\text{Planck}}, \quad (\text{F.5})$$

with the evidence Z_i given by

$$Z = \int \mathcal{L}\pi d\theta, \quad (\text{F.6})$$

the integral of the likelihood \mathcal{L} and prior π over the parameters θ . The information ratio I in Eq. (F.4) is defined as

$$\ln I = \mathcal{D}_{3 \times 2\text{pt}} + \mathcal{D}_{\text{Planck}} - \mathcal{D}_{3 \times 2\text{pt} + \text{Planck}}, \quad (\text{F.7})$$

with \mathcal{D}_i being the Kullback-Leibler divergence between the posterior P and prior for probe i :

$$\begin{aligned} \mathcal{D} &= \int P \ln \frac{P}{\pi} d\theta = \int P \ln \mathcal{L} d\theta - \ln Z \\ &= \langle \ln \mathcal{L} \rangle_P - \ln Z. \end{aligned} \quad (\text{F.8})$$

The second equality follows from Bayes theorem: $P = \mathcal{L}\pi / Z$. Using this definition of \mathcal{D} allows us to rephrase the suspiciousness solely in terms of the expectation values of the log-likelihoods:

$$\begin{aligned} \ln S &= \langle \ln \mathcal{L}_{3 \times 2\text{pt} + \text{Planck}} \rangle_{P_{3 \times 2\text{pt} + \text{Planck}}} - \langle \ln \mathcal{L}_{3 \times 2\text{pt}} \rangle_{P_{3 \times 2\text{pt}}} \\ &\quad - \langle \ln \mathcal{L}_{\text{Planck}} \rangle_{P_{\text{Planck}}}. \end{aligned} \quad (\text{F.9})$$

We find $\ln S = -2.0 \pm 0.1$. For Gaussian posteriors, the quantity $d - 2\ln S$ is distributed as χ_d^2 , where $d = d_{3 \times 2\text{pt}} + d_{\text{Planck}} - d_{3 \times 2\text{pt} + \text{Planck}}$ is the difference in the Bayesian model dimensionalities. This allows us to assign a probability of the observed suspiciousness under the assumption that the two data sets are in concordance. We calculate d , following Handley & Lemos (2019), from the variances of the log-likelihoods. Other estimates of the Bayesian model dimensionality, such as those introduced in Raveri & Hu (2019) yield very similar results. We calculate $d = 3.3 \pm 0.7$ and therefore conclude that the probability of observing our measured suspiciousness statistic is 0.08 ± 0.02 , or $1.8 \pm 0.1\sigma$.

Appendix F.4: Goodness-of-fit change: Q_{DMAP}

Raveri & Hu (2019) introduce the Q_{DMAP} statistic that quantifies how the goodness-of-fit changes when two data sets are combined, based on the difference of the log-likelihoods, $\ln \mathcal{L}(\boldsymbol{\theta}_i^{\text{MAP}})$, at the MAP of the respective posteriors¹⁹,

$$Q_{\text{DMAP}} = 2 \ln \mathcal{L}(\boldsymbol{\theta}_{3 \times 2\text{pt}}^{\text{MAP}}) + 2 \ln \mathcal{L}(\boldsymbol{\theta}_{\text{Planck}}^{\text{MAP}}) - 2 \ln \mathcal{L}(\boldsymbol{\theta}_{3 \times 2\text{pt}+\text{Planck}}^{\text{MAP}}). \quad (\text{F.10})$$

For Gaussian posteriors, the sampling distribution of Q_{DMAP} is χ^2 -distributed with $d_{3 \times 2\text{pt}} + d_{\text{Planck}} - d_{3 \times 2\text{pt}+\text{Planck}}$ degrees of freedom. Here we follow Raveri & Hu (2019) and calculate the model dimensionalities by $d_i = N - \text{tr}[C_{\pi}^{-1} C_P]$, where N is the total number of varied parameters, C_{π} is the prior parameter covariance, and C_P is the posterior parameter covariance. We find consistent results when using the variance of the log-likelihoods, the approach taken for the suspiciousness calculation. We find $Q_{\text{DMAP}} = 9.5$, with $d = 3.6$, corresponding to a probability of 0.037, or 2.1σ .

Appendix F.5: Parameter difference in multiple dimensions: Q_{UDM}

Raveri & Hu (2019) introduce the Q_{UDM} statistic that quantifies the difference between cosmological parameters across multiple dimensions. Here

$$Q_{\text{UDM}} = \Delta \bar{\boldsymbol{\theta}}^T (C_{\text{Planck}} - C_{3 \times 2\text{pt}+\text{Planck}})^{-1} \Delta \bar{\boldsymbol{\theta}}, \quad (\text{F.11})$$

where $\Delta \bar{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}}_{\text{Planck}} - \bar{\boldsymbol{\theta}}_{3 \times 2\text{pt}+\text{Planck}}$ is the shift in the mean of the shared parameters, and C_{Planck} and $C_{3 \times 2\text{pt}+\text{Planck}}$ are the posterior parameter covariances for Planck and $3 \times 2\text{pt} + \text{Planck}$, respectively. The estimator Eq. (F.11) is asymmetric in the first data set if the posteriors are non-Gaussian. In this case, the data set with the more Gaussian posterior should be used, which in our case is Planck . We use `tensiometer`²⁰ to compute Q_{UDM} and its associated number of degrees of freedom. We find $Q_{\text{UDM}} = 7.7$ with $d = 3$. The sampling distribution is approximated by χ_d^2 , such that the probability of the observed value of Q_{UDM} is 0.054, or 1.9σ .

Appendix G: Redundancy, pipeline validation and software review

In this Appendix we briefly review the redundancy in the KiDS-1000 analysis, from pixels through to parameters. Multi-band pixel processing and photometry in the optical has been carried out on the full KiDS-1000 data set using two independent pipelines, ASTROWISE and THELI (Begeman et al. 2013; Erben et al. 2013). This approach allowed us to resolve a range of different issues, primarily in the astrometric solution, for the handful of problem fields that resisted automated processing. We adopt the ASTROWISE reduction for multi-band photometry and the THELI r -band reduction for object detection and weak lensing shape measurement (Kuijken et al. 2019). We consider only one shape measurement technique, `lensfit` (Miller et al. 2013), but this is calibrated using a series of different image simulations that vary the input galaxy properties to assess the

¹⁹ As an interesting aside, we note that the combined quantity $Q_{\text{DMAP}} + 4 \ln S$ corresponds to the twice the deviance information criterion ratio introduced in Joudaki et al. (2017).

²⁰ <https://github.com/mraveri/tensiometer>

sensitivity of the shear estimator to: the presence of blending with unresolved and undetected objects, blending due to enhanced galaxy clustering, photometric redshift selection bias, size-ellipticity correlations in the galaxy properties, varying stellar density, and the choice of smooth or realistic galaxy profiles (see Kannawadi et al. 2019; Giblin et al. 2020, for details).

Our fiducial photometric redshift calibration has been compared to two additional independent calibration approaches, and has been validated on mock photometry catalogues (see Wright et al. 2020a; van den Busch et al. 2020; Hildebrandt et al. 2020b, for details). Using a series of null-tests we have validated the resulting shear-redshift catalogues in Giblin et al. (2020). Our catalogue-to-observables pipeline, based on the two-point correlation function code `TREECORR` (Jarvis 2015), has been validated through an independent analysis using the alternative `ATHENA` package (Kilbinger et al. 2014), and through mock catalogue analysis (Joachimi et al. 2020). Our cosmological inference code `KCAP` with `CAMB` (Lewis et al. 2000), has been validated against the Core Cosmology Library (`CCL`, Chisari et al. 2019) and through the recovery of input parameters in mock data analysis (Joachimi et al. 2020). In Asgari et al. (2020b) we also verify that we produce the same results in our cosmic-shear only analysis using a completely independent inference code based on `MONTEPYTHON` with `CLASS` (Lesgourgues 2011; Brinckmann & Lesgourgues 2019; Köhlinger et al. 2019; Hildebrandt et al. 2020a). Our analysis of BOSS follows Sánchez et al. (2017), which is in good agreement with the numerous independent parallel analyses of the same DR12 data set presented in Alam et al. (2017).

In regards to software review, for the majority of cases we have adopted the ‘four-eye’ approach, collaboratively building code through a git repository, with major updates reviewed through pull requests. This approach has applied throughout the full pixels-to-parameters process, and to all types of software, both our significant tools, analysis code and, for the most part, simple paper support scripts. On publication of this paper, our repositories will become open source for others to use and build upon, but with a caveat for users to recognise that we are not software engineers. Adopting this style of software review for KiDS-1000 has been an extremely beneficial exercise for the team, with many lessons learnt for how to improve our software engineering skills and our approach to open source collaborative coding for future projects.

Appendix H: Post-unblinding analyses

This analysis was carried out ‘blind’, such that our final key result, our constraint on S_8 , was unknown until all analysis choices were fixed. Our blinding strategy creates three versions of the catalogue, where one is the truth, and the other two are modified to introduce up to a $\sim \pm 2\sigma$ deviation in the recovered value of S_8 (Kuijken et al. 2015; Giblin et al. 2020). In contrast to previous KiDS analyses, we set ourselves a challenge to only run our blind data analysis inference once, after fully developing and road-testing the pipeline using mock catalogues and data vectors (Joachimi et al. 2020). Unexpectedly, the only analysis where we failed to meet this challenge were for data vectors that included galaxy clustering. Here a bug in a naming convention in an updated `COSMOSIS-CAMB` interface resulted in $\sim 0.5\sigma$ errors in the recovery of σ_8 in our initial BOSS re-analysis. This was corrected and updated for our fiducial analysis before unblinding.

Our fiducial results, presented in Fig. 4 and tabulated in Appendix C, were carried out for all three blinds using a covariance matrix derived assuming a fiducial cosmology given by the

best-fit parameters from Tröster et al. (2020). We reserved our iterated-covariance analysis, however, for post-unblinding, re-analysing the true data vector with an updated covariance matrix derived adopting the best-fit parameters from our initial blinded run of the true catalogue. This iterative step changed our value for the full $3 \times 2\text{pt}$ analysis S_8 by 0.2σ for the MAP constraint, and by 0.1σ for the marginal constraint. Our tension-consistency analysis with the CMB constraints from Planck was also conducted after unblinding, as we were unblind to Planck throughout the process.

As we are only interested in relative shifts in the sensitivity tests presented in Fig D.1, these were only conducted for a single blind, and re-calculated post-unblinding for the true catalogue. Wishing to be fully transparent in this Appendix, we note that owing to limited resources the sensitivity test was not updated during the blinded phase of the project in order to correct for the COSMOSIS-CAMB interface error which only impacted the BOSS constraints. We argue that this was appropriate given the nature of the test and given that this error has been corrected post-unblinding. Our cosmic shear and galaxy clustering analysis, featured in our internal consistency test in Fig. 5, was also not re-processed blind after the BOSS error was corrected, as both our forecast in Joachimi et al. (2020), and our initial analysis, confirmed that the constraints from our cosmic shear and galaxy clustering analysis were almost identical to constraints from the full $3 \times 2\text{pt}$ analysis, which was correctly analysed for all three blinds.

As reported in Giblin et al. (2020), co-author Kannawadi was unblinded early in the process to permit accurate calibration of the shear measurements in his re-analysis of KiDS-1000-like image simulations. Co-authors Wright and Heymans also wish to record that they independently had a suspicion over which blind was the truth based on the changes that the SOM gold selection made to the blinded KiDS-1000 catalogue effective number density and ellipticity dispersion, compared to the impact the SOM selection made to the KiDS-VIKING-450 catalogue analysed in Wright et al. (2020b). These suspicions were never shared with the rest of the team, nor discussed with each other, and on unblinding were found to both be false, and different! This issue nevertheless highlighted to us the challenges of catalogue-level blinding for successive data releases where little has changed in the core data reduction process. This is particularly relevant if the people working at the coal-face of catalogue production are the same people working on the cosmological analysis. Future KiDS blinding will therefore likely adopt the approach advocated by Sellentin (2020), where the covariance matrix is modified in the analysis.